TRANSPORT IN RESERVOIR COMPUTING*

G MANJUNATH† AND JUAN-PABLO ORTEGA‡

Abstract. Reservoir computing systems are essentially dynamical systems influenced by an exogenous input. Such systems are extensively used in biologically inspired information processing, and are the state-of-the art techniques for several machine learning tasks. If the statistics of the response or output of the system depends discontinuously on the distribution of the inputs, a fundamental challenge arises in applications where inherent changes in the input stochastic source or noise is expected. This problem can be experimentally demonstrated by showing that altering input statistics can drastically affect the statistics of the response. We solve this instability problem by providing sufficient conditions under which both the marginals and the joint distributions of the response depend continuously on that of the input. To prove our results, we establish the existence of an invariant measure and show that its dependence on the input process is continuous when the processes are endowed with the Wasserstein distance. The main tool in these developments is the characterization of those invariant measures as fixed points of naturally defined Foias operators that appear in this context and which are examined extensively in the paper. These fixed points are obtained by imposing a newly introduced stochastic state contractivity on the driven system that is readily verifiable in examples. Stochastic state contractivity can be satisfied by systems that are not state-contractive, which is a need typically evoked to guarantee the echo state property in reservoir computing. As a result, it may actually be satisfied even if the echo state property is missing.

Key words. driven dynamical systems, reservoir computing, recurrent neural network, echo state property, unique solution property, Frobenius-Perron operator, Foias operator, transport, invariant measure, stochastic contraction, Wasserstein distance.

MSC codes. 62M45, 68Q32, 65P40, 60G10

1. Introduction.

Transport in dynamical systems is studied at both microscopic and macroscopical levels. On the one hand, at the microscopic level, if one is interested in the motion of a particle in a fluid, and the particle is assumed to be so light that it can do nothing but follow the liquid, then the motion of the fluid totally determines the fate of the particle. In particular, the different dynamical properties of the fluid can create transport barriers for the particle (see, for instance, [2, 42]) trapping its trajectory in a subset of the phase space. When the motion of individual trajectories is not possible due to the inherent loss of predictability that is typical in very general classes of dynamical systems, transport at the macroscopic level is useful. In such a macroscopic study, one aims to make predictions regarding the longtime evolution of ensembles of trajectories and the term transport refers to the properties of the time evolution of measures [26]. Loosely stated, such transport does not describe anymore the motion of a particle but instead describes how mass "accumulates" over a period of time. In the language of dynamical systems, such transport concerns the evolution of an ensemble of initial conditions. Such macroscopic transport can yield very simple measure dynamics even when the underlying microscopic dynamics is very

*Submitted to the editors DATE.

Funding: GM acknowledges support from the National Research Foundation of South Africa under Grant UID 115940 as well as the hospitality of the Division of Mathematical Sciences of the Nanyang Technological University which funded the visit in September 2022 during which this paper was written. JPO acknowledges partial financial support coming from the Swiss National Science Foundation (grant number 200021_175801/1). We thank Lukas Gonon and Lyudmila Grigoryeva for interesting exchanges that inspired some of the results in this paper.

†Department of Mathematics & Applied Mathematics, University of Pretoria (manjunath.gandhi@up.ac.za,).

‡Division of Mathematical Sciences, Nanyang Technological University (Juan-Pablo.Ortega@ntu.edu.sg.

complex.

An important case takes place when the simplified macroscopic dynamics exhibits a single limit. This feature is related to certain statistical stability which means that a swarm of initial conditions with different initial distributions/densities can all converge (typically in the L^1 norm) to an asymptotic distribution/density which is known as an **invariant measure/density** [26]. The invariant measure/density sheds light on how the asymptotic states of the system get distributed. In the case of time-independent or autonomous systems, the time evolution of measures/densities is well-studied using the **Frobenius-Perron operator** (see, for instance, [26] and Figure 2). Non-autonomous extensions of these results present serious challenges since the input affects the time-evolution of the measures at each time step. These difficulties are even more pronounced whenever other natural metrics like, for instance, the Wasserstein distance are used in the space of measures. In order to visualize why this is so, recall that for autonomous systems the Wasserstein distance intuitively corresponds to the effort in moving a mount of mass that is not disturbed during transportation; for a nonautonomous system, it would correspond to the effort of moving a mount of mass that can also be disarranged during transportation.

In this work, we consider a class of time-dependent dynamical systems that arise in the field of systems theory and, more specifically, in **reservoir computing (RC)** [21, 22, 30, 29, 28, 40]. RC uses input-output systems that are defined with the help of a **driven** or **state-space system**, that is, a continuous function $g : U \times X \to X$ on an **input space** U and a **state space** X (both are metric spaces). The main difference between general driven systems and RC is that for the latter and for supervised machine learning applications, the function g is not trained but (partially) randomly generated, and the corresponding input-output system is obtained out of a (functionally simple) observation equation of the states. Various families of reservoir computing systems have been shown to exhibit universal approximation properties [18, 17, 15, 8, 16, 14]. A **solution** of g for a given bi-infinite input $\{u_n\}_{n\in\mathbb{Z}}$ is a bi-infinite sequence $\{x_n\}_{n\in\mathbb{Z}}$ whenever the equality $x_{n+1} = g(u_n, x_n)$ is satisfied for all $n \in \mathbb{Z}$. The terms x_n of the solution are referred to as state values or reservoir states in the RC context.

If for each input $\{u_n\}_{n\in\mathbb{Z}}$ there exists exactly one solution $\{x_n\}_{n\in\mathbb{Z}}$, then g is said to have the **echo state property (ESP)** [21] or the **unique solution property (USP)** [32]. Often in practice, only a class of inputs is considered when formulating the ESP, like for instance the one coming from the realizations of a stochastic process. In this work, we place ourselves in a setup that does not necessarily imply the ESP for all possible inputs, as it has been empirically demonstrated, for instance across the echo state networks (ESNs) literature [21, 22], that the performance of ESNs is sometimes enhanced when the reservoir dynamics does not have this property [28].

From a qualitative dynamics point of view, the unique solution property is equivalent (at least for compactly driven systems) to the fact that for repeated runs of the RC system with a given input sequence and using different initial conditions in the state space, the resulting state sequences get closer and closer to a solution when the RC runs are longer and longer. This property is called the **uniform attracting property (UAP)** [32, Definition 3] and amounts to the system exhibiting what is called a **uniform point attractor**. In [32, Theorem 1] it has been shown that the UAP is equivalent to the USP.

In order to develop some intuition about these facts, we will use as an example echo state networks (ESNs), a family of RC systems that has been profusely used in applications. ESNs are determined by a state map $g : \mathbb{R}^d \times \mathbb{R}^N \to \mathbb{R}^N$ defined as

$$g(u, x) = \overline{\tanh}(Cu + \alpha Ax), \tag{1.1}$$

where A and C are called the reservoir (or connectivity) and input matrices, respectively, which have appropriate dimensions and, in the RC context, are randomly generated. The symbol $\overline{\tanh}$ denotes the nonlinear activation function tanh applied in a component-wise manner. Notice that when a large-amplitude input is used in this RC system, the activation function tanh is saturated because the reservoir neurons become highly stimulated, the tanh quenches strongly and, as a consequence, the initial condition is forgotten. On the other hand, for small-amplitude inputs, such "washing out" qualities may be lost. For instance, a constant zero input, is a good candidate for the USP to cease to hold. In practice, however, the relevant input range frequently contains zero. More specifically, all that is often mentioned about permissible inputs is their range, which consequently yields non-typical inputs like the constant-zero signal as an allowable input, that must hence be accounted for when determining the USP. Much has been studied regarding the intricacies associated to the USP and its dynamical implications. For instance, in [33], the USP with respect to an input has been considered and, in particular, it has been shown that even if the USP with respect to all inputs does not hold, the USP can still hold with probability 1 when some given stationary ergodic process is considered as the input. However, vital questions remain unanswered.

This paper studies how measures are transported in the RC framework with the goal of answering the following fundamental questions: **(i)** If input sequences originate from a stationary stochastic process, how would the output reservoir states get asymptotically distributed? In particular, is there an invariant distribution/measure available for the output states of RC systems? **(ii)** When such an invariant measure exists, does it depend continuously on the input stationary measure when a natural measure of dissimilarity between probability measures like, for instance, the Wasserstein distance is employed? The Wasserstein distance is a particularly appropriate choice first because of its good analytical properties but, more importantly, because of its relation with optimal transport (see [41, 37] and references therein).

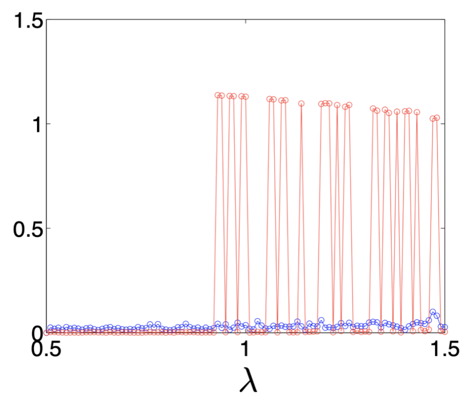

FIG. 1. *Consider a driven system that is a recurrent neural network (RNN) of the type* (1.1) *with input and reservoir matrices randomly initialized and with a reservoir matrix set to having a spectral radius of* 1.5. *This condition makes the RNN incompatible with the USP [21]. Input is fed into the RNN that follows an exponential distribution with parameter* λ. *In blue we depict the Wasserstein-1 distance (introduced later on) between the exponential distributions with parameters* λ *and* $\lambda + \epsilon$ *(with* $\epsilon = 0.01$*) against* λ. *In red is the corresponding Wasserstein-1 distance observed between the distributions of the states of a single neuron observed. The distances are calculated using 3500 data points. Larger values on the vertical axis point to the discontinuity of the distribution of the states of the neuron as* λ *is varied.*

These questions are relevant in two contexts: **(a)** Any future notion of stochastic reservoir computing where the stochasticity of the reservoir is controlled by the input requires the properties **(i)** or **(ii)**. **(b)** The echo state property or the USP in the RC framework has been found to yield certain stability properties; for instance, the USP guarantees an input-related stability [31] that implies that close-by input sequences lead to close-by reservoir state sequences. In this context, it is natural to ask what other hypotheses may be required so that close-by input stationary distributions/measures lead to close-by invariant reservoir stationary distributions/measures. At this point, it is important to emphasize that in the absence of the USP, such robustness is actually not available in general. Indeed, as we numerically illustrate in Figure 1 with a ESN of the type introduced in (1.1) that does not have the USP, large variations in the distributions of the reservoir states can be obtained even when the input distribution is varied slightly.

In this work we propose to deal with the questions **(i)** and **(ii)** by providing sufficient conditions under which these properties are satisfied and that, in particular, guarantee that behaviors similar to the one depicted in Figure 1 cannot occur. Since we intend to ensure that these features are available even when the USP is not satisfied, we consider a notion in which the resulting reservoir state sequences can be separated from each other possibly with non-vanishing probability. More precisely, given a probability measure θ on the input space U and $0 < c < 1$, we say that the driven system g is a (θ, c)**-stochastic contraction** when

$$\int_U d_X(g(u,x), g(u,y))\, d\theta(u) \leq c\, d_X(x,y) \quad \text{holds for all } x, y \in X. \tag{1.2}$$

We stress that contractivity, a requirement commonly invoked to ensure the echo state property (called USP here) in reservoir computing, can be satisfied by systems that are not state-contractive, and may even be satisfied in the absence of the echo state property (see Remark 4.4). Similar conditions have been formulated in the literature, mostly in the L^p context to, for example, prove the stability of functional autoregressive models (see [10, Chapter 6] and references therein). In our setup, and since the questions **(i)** and **(ii)** above concern distributions of solution sequences in X rather than just the values in X, we shall introduce later on in Definition 3.1 a system $\mathbf{G}$ in sequence space induced by g, and use the above contraction property to handle **(i)** and **(ii)** in Theorems 4.5 and 4.14. Regarding the time evolution of measures, in the context of autonomous systems the Frobenius-Perron operator [26] has been traditionally used. In our non-autonomous setup, we utilize one of its generalizations, namely the the so-called **Foias operator**, which was used, for instance, in [26, Chapter 12] to analyze systems driven by IID noise and studied mostly on spaces equipped with the L^1 norm (see Figure 2 for an illustration). Also, while transport problems in autonomous dynamical systems are completely solved when the exit times from each measurable set is known [42], with RC systems such questions have not yet been addressed. In that respect, this paper can be viewed as the first step to answer the question of the convergence towards a stationary distribution or invariant measure and, more generally, whether we can reliably use the statistical information of the reservoir for information processing.

The statistical properties of nonautonomous dynamical systems have been studied before. The idea of "loss of memory" of initial densities was introduced in [36] that is a generalization of the notion of decay of correlations (e.g., [27, 43]). The phrase "loss of memory" comes from the assumption that all relevant densities are drawn to the same moving target in the space of densities in an impartial manner. In particular, the authors in [36] deal with a family of piecewise expanding maps on a compact interval with uniform expanding properties and show that if ρ_0 and $\hat{\rho}_0$ are two arbitrary initial densities with respect to a reference measure m, then their time-evolutions ρ_t and $\hat{\rho}_t$ get closer to each other with exponential speed, that is, $\int |\rho_t - \hat{\rho}_t| dm < Ce^{-\alpha t}$.

Other extensions or generalizations in different settings concerning the speed of loss of memory in results can be found in the literature. Among them, we can refer to results for piecewise expanding maps in [20], and for random composition of two-dimensional Anosov diffeomorphisms in [39]. More specifically, [39] consider a weaker convergence notion where it is shown that $|\int f \circ T_{1,n} d\mu_1 - f \circ T_{1,n} d\mu_2| \leq Ce^{-\alpha n}$, where f is a Hölder observable and $T_{1,n}$ denotes the composition of n maps. A variation of convergence called "conditional memory loss" was established in [34] for expanding Lasota-Yorke maps interval maps. Although there are variations of such results we refer to Stenlund in [39] who makes the general statement: "Much of the statistical theory of stationary dynamical systems can be carried over to sufficiently chaotic non-stationary systems." So to fill this relative sparseness for statistical properties of nonautonomous systems, weaker hypotheses with less expanding properties as in the mixing or chaotic situations need to be considered. The work in [1] pertains to results for a class of maps on a compact interval that have a neutral fixed point, while the work in [3] relaxes the topological transitivity conditions on an interval. Our work here also relaxes on the global expanding hypotheses since the composition of maps induced when an input sequence drives a driven system can be allowed to have only some local expanding properties which the stochastic contraction property permits.

The general organization of the paper is as follows: Section 2 introduces the setup in relation with driven systems and their associated Foias operators. In particular, Proposition 2.3 determines when such an operator is well-defined as a map between Wasserstein spaces. Section 3 contains a detailed account on how a driven system g naturally induces another driven system $\mathbf{G}$ in sequence space. The sequence space representation is important since it provides additional tools for the characterization of the solutions of a driven system. Section 4 is the core of the paper and contains the main results. More specifically, we prove in this section the **existence and uniqueness of invariant measures for the Foias operators** in both the state (Theorem 4.14) and sequence spaces (Theorem 4.15), as well as the **continuity of their dependences on the input process**. The main tool to achieve this is Banach's Fixed Point Theorem, which requires two conditions, namely contractivity and continuity, which will be implied for the Foias operators in both the state and sequence spaces by **conditions that are readily verifiable and exclusively formulated for the driven system g defined in the state space**. Indeed, most of the developments in that section consist in showing that the contractivity and continuity conditions imposed on g translate to similar properties at the level of the Foias operators in both the state and sequence spaces. The contractivity question is mostly treated in Subsection 4.1 where it is shown that the newly introduced notion of **stochastic state contractivity** (1.2) for the driven system g, ensures that the Foias operators in state and in sequence spaces are also contractive with respect to the Wasserstein distance (see Figure 3 for a summary of the implications between contractivity in different spaces). We emphasize that stochastic state contractivity is less restrictive than the standard state contractivity condition evoked to ensure the USP. All the continuity questions are contained in Subsection 4.2 (see Figure 4 for a summary of the different continuity implications). Subsection 4.3 contains the two main results Theorem 4.14 and Theorem 4.15. Section 5 concludes the paper.

2. Preliminaries.

As we already mentioned in the introduction, we place ourselves in the context of **driven systems** induced by a function $g : U \times X \to X$ which has as domain the metric **input space** (U, d_U) and the metric **state space** (X, d_X). We say that a bi-infinite **output sequence** $\mathbf{x} = \{x_n\}_{n\in\mathbb{Z}} \in X^{\mathbb{Z}}$ is **compatible** or is a **solution** for the **input sequence** $\mathbf{u} = \{u_n\}_{n\in\mathbb{Z}} \in U^{\mathbb{Z}}$ when the following identity is satisfied

$$x_{n+1} = g(u_n, x_n), \quad \text{for all } n \in \mathbb{Z}. \tag{2.1}$$

The driven system g has the **unique solution property (USP)** if for each input sequence there is a unique output sequence that is compatible with it. In the reservoir computing framework, the USP is usually referred to as the echo state property (ESP) and is often ensured by imposing various contraction properties. The USP guarantees the existence of a unique causal and time-invariant **filter** $U_g : U^{\mathbb{Z}} \to X^{\mathbb{Z}}$ (see [6] or [19] for detailed definitions) which is characterized by the relation

$$U_g(\mathbf{u})_{n+1} = g\left(u_n, U_g(\mathbf{u})_n\right), \quad \text{for all } n \in \mathbb{Z}. \tag{2.2}$$

We shall show in our work that in the presence of stochastic inputs, even if the USP condition is not satisfied by the driven system, we can still use the **Foias operator** to associate to it a *continuous input-output system in the space of stochastic processes.*

Wasserstein distances. We recall some relevant definitions next. Suppose that (Y, d_Y) is a **Polish space** (that is, it is complete and separable) and denote by $P(Y)$ the set of Borel probability measures. Let $p \in [0, \infty)$ and define the **Wasserstein space** $P_p(Y)$ of order p as

$$P_p(Y) := \left\{ \mu \in P(Y) \;\middle|\; \int_Y d_Y(y_0, y)^p d\mu(y) < +\infty \right\},$$

where $y_0 \in Y$ is arbitrary since it can be shown that this definition does not depend on the point y_0. This space can be made into a Polish space by using the **Wasserstein-p distance** $W_p : P_p(Y) \times P_p(Y) \to [0, \infty)$ (see [41, Theorem 6.18])

$$\begin{aligned} W_p(\mu,\nu) &= \left(\inf_{\pi\in\Pi(\mu,\nu)} \int_{Y\times Y} d_Y(x_1,x_2)^p d\pi(x_1,x_2) \right)^{1/p} \\ &= \inf_{\pi\in\Pi(\mu,\nu)} \left\{ \left(\mathrm{E}\left[d_Y(X_1,X_2)^p\right]\right)^{1/p} \mid \mathrm{law}(X_1) = \mu,\ \mathrm{law}(X_2) = \nu \right\}, \end{aligned} \tag{2.3}$$

where $\Pi(\mu, \nu)$ is the set of all joint Borel probability measures on $Y \times Y$ whose marginals are μ and ν, that is, $\mu(A) = \pi(A \times Y)$ and $\nu(A) = \pi(Y \times A)$ for all regular Borel subsets $A \subset Y$. The pair of random variables (X_1, X_2) in the second line of the definition are defined on $Y \times Y$.

All the work in this paper will be conducted using the Wasserstein-1 distance W_1 that in the sequel will be denoted just as W. The **Kantorovich-Rubinshtein** duality formula (see [23, 25] or [9, Theorem 11.8.2]) provides an alternative expression for the Wasserstein-1 distance, namely

$$W(\mu,\nu) = \sup_{f\in \mathrm{Lip}_1(Y,\mathbb{R})} \left\{ \int_Y f d\mu - \int_Y f d\nu \right\}, \tag{2.4}$$

where $\mathrm{Lip}_1(Y, \mathbb{R})$ denotes the set of all real-valued functions f on Y so that $|f(y_1) - f(y_2)| \leq d_Y(y_1, y_2)$, for all $y_1, y_2 \in Y$ or, equivalently, the set of real-valued Lipschitz continuous functions with Lipschitz constant smaller or equal to one.

An important fact that we will use repeatedly is that W metrizes the weak convergence in $P_1(Y)$ (see [41, Theorem 6.9]). More specifically, we say that the sequence

$\{\mu_n\}_{n\in\mathbb{N}}$ of measures in $P_1(Y)$ **converges weakly** to $\mu \in P_1(Y)$, whenever for any continuous real-valued function f such that $|f(y)| \leq C(1 + d(y_0, y))$, for some $C \in \mathbb{R}$, and some (and then any) $y_0 \in Y$, we have that

$$\int_Y f(y) d\mu_n(y) \to \int_Y f(y) d\mu(y). \tag{2.5}$$

We say that W metrizes the weak convergence in $P_1(Y)$ because the statement $\{\mu_n\}_{n\in\mathbb{N}}$ converges weakly to $\mu \in P_1(Y)$ is equivalent to $W(\mu_n, \mu) \to 0$. See [41, Definition 6.8] for other characterizations of weak convergence in $P_1(Y)$.

We refer to the Chapter 6 of the monograph [41] for the central role of the Wasserstein metric in the study of optimal transport. In particular, unlike the total variation norm or the Kullback-Leibler divergence, it helps in comparing measures that are not absolutely continuous with respect to each other, a situation that often arises in practice while considering empirical distributions.

The Frobenius-Perron and the Foias operators. Using the terminology in [26, Chapter 12] and in the same setup as in the previous paragraph, we can associate to each Lipschitz continuous discrete-time dynamical system on Y a natural operator $P_f : P_1(Y) \to P_1(Y)$ that describes how probability distributions on Y are mapped by the dynamical system. More specifically, let $f \in \mathrm{Lip}(Y)$ be a Lipschitz continuous self-map of Y. The **Frobenius-Perron operator** $P_f : P_1(Y) \to P_1(Y)$ associated to f is defined by $P_f(\mu) = f_*\mu$, with $f_*\mu$ the pushed-forward measure of $\mu \in P_1(Y)$ by f given by $f_*\mu(A) = \mu\left(f^{-1}(A)\right)$, for any Borel subset $A \subset Y$. The Lipschitz condition on f implies that $P_f(\mu) \in P_1(Y)$. Notice that *if d_Y is a bounded metric, then the Frobenius-Perron operator $P_f : P_1(Y) \to P_1(Y)$ is defined for any measurable self-map* f of Y and *if Y is compact then P_f is defined for any continuous map f*. Additionally, if $\mathrm{Lip}_c(Y)$ denotes the space of c-Lipschitz continuous dynamical systems on Y and $f \in \mathrm{Lip}_c(Y)$, then $P_f \in \mathrm{Lip}_c(P_1(Y))$ when $P_1(Y)$ is endowed with the Wasserstein-1 distance.

The notion of Frobenius-Perron operator for a dynamical system can be extended to driven systems g of the type introduced in (2.1), in which case is called the **Foias operator** (see [26, Definition 12.4.2]).

Definition 2.1. *Let $g : U \times X \to X$ be a measurable driven system that has as domain the input (U, d_U) and state (X, d_X) spaces that are assumed to be Polish spaces. The* **Foias operator** *$P_g : P(U) \times P_1(X) \to P_1(X)$ associated to g is defined by*

$$P_g(\theta, \mu) = \int_U g_{u\,*}\mu \, d\theta(u), \tag{2.6}$$

where $g_u : X \to X$ is defined by $g_u(x) = g(u, x)$, for all $u \in U$. The term measurable means that the preimage $g^{-1}(A)$ by g of any Borel subset $A \subset X$ of X is a Borel subset of $U \times X$. The equality that defines $P_g(\theta, \mu) \in P_1(X)$ in (2.6) is an abbreviation for the measure that for any Borel subset $A \subset X$ takes the value

$$P_g(\theta, \mu)(A) = \int_X \left(\int_U \mathbf{1}_A \left(g(u, x) \right) \, d\theta(u) \right) d\mu(x), \tag{2.7}$$

with $\mathbf{1}_A : X \to \{0, 1\}$ the indicator function of A.

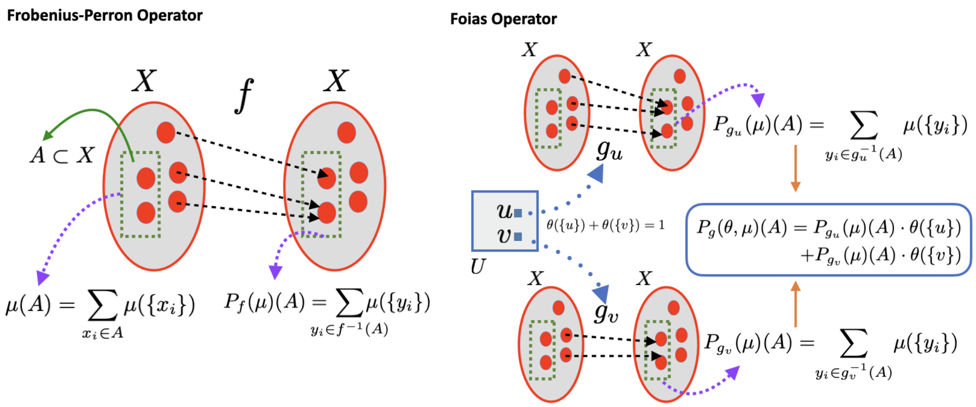

FIG. 2. *An illustration of Frobenius-Perron and the Foias operators. The Frobenius-Perron operator pushes forward measures using a dynamical system. The Foias operator is a non-autonomous generalization that integrates the push-forwards given by a driven system with respect to a measure in the input space.*

Remark 2.2. It is easy to show using Fubini's Theorem that the Foias operator coincides with the push-forward map $g_* : P(U \times X) \longrightarrow P(X)$ when restricted to product measures in $P(U \times X)$ of the form $\tau(C \times A) = \theta(C) \cdot \mu(A)$, with C and A Borel sets in U and X, respectively, and $\theta \in P(U)$, $\mu \in P_1(X)$. In other words, the Foias operator P_g coincides with the push-forward map g_* of a driven system g when applied to independent random variables in U and X (that is, the ones that have laws θ and μ). Indeed, for any such measure and any Borel set $A \subset X$, Fubini's theorem guarantees that:

$$
\begin{aligned}
g_*\tau(A) &= \int_X \mathbf{1}_A(x)\, d\left(g_*\tau\right)(x) = \int_{U\times X} \mathbf{1}_A(g(u,x))\, d\tau(u,x) \\
&= \int_{U\times X} \mathbf{1}_A(g(u,x))\, d\theta(u) d\mu(x) = \int_X \left(\int_U \mathbf{1}_A\left(g(u,x)\right)\, d\theta(u)\right) d\mu(x) \\
&= P_g(\theta,\mu)(A).
\end{aligned} \tag{2.8}
$$

The following result provides conditions that ensure that the Foias operator is well-defined and that, in particular, maps into $P_1(X)$. Most of the time in this paper we shall be working under the hypothesis in part **(i)**.

PROPOSITION 2.3. *In the setup of the previous definition, the Foias operator is well-defined under any of the following hypotheses:*
(i) *d_X is a bounded metric and g is measurable.*
(ii) *The input U and state X spaces are compact and g is continuous.*
(iii) *The input U space is compact, g is continuous, and the maps g_u are all Lipschitz continuous with constants c_u such that $\sup_{u\in U}\{c_u\} = c < +\infty$.*

Proof.
All that it needs to be shown is that for an element $x_0 \in X$ (and hence for any) and $(\theta,\mu) \in P(U) \times P_1(X)$, the integral

$$\int_X d_X(x,x_0) dP_g(\theta,\mu)(x) = \int_X \int_U d_X(g_u(x),x_0) d\theta(u) d\mu(x)$$

is finite in the presence of any of the three hypothesis in the statement. It is clearly the case when d_X is a bounded metric. Under the hypotheses in **(ii)**, the continuous function $d_X(g_u(x), x_0)$ reaches a maximum $M > 0$ at a point (u', x') and hence

$$\int_X d_X(x, x_0) dP_g(\theta, \mu)(x) = \int_X \int_U d_X(g_u(x), x_0) d\theta(u) d\mu(x)$$
$$\leq M \int_X \int_U d\theta(u) d\mu(x) = M < +\infty.$$

Regarding part **(iii)** we shall show that $\int_X d_X(x, g_{u_0}(x_0)) dP_g(\theta, \mu)(x) < +\infty$ for some fixed $u_0 \in U$. By the triangle inequality and the Lipschitz condition

$$d_X(g_u(x), g_{u_0}(x_0)) \leq d_X(g_u(x), g_u(x_0)) + d_X(g_u(x_0), g_{u_0}(x_0)) \leq c d_X(x, x_0) + M,$$

where M is the maximum of the function $d_X(g_u(x_0), g_{u_0}(x_0))$ thought of as a continuous function of the variable u on the compact set U. Consequently,

$$\int_X d_X(x, g_{u_0}(x_0)) dP_g(\theta, \mu)(x) \leq c \int_Y d_X(x, x_0) d\mu(x) + M < +\infty,$$

since $\mu \in P_1(X)$. ■

3. Driven systems in sequence spaces. In this section we study how driven systems induce natural maps between input and output sequence spaces that will be used later on in the paper.

Sequence spaces. We saw in the previous section how driven systems that satisfy the USP naturally induce input / output systems between the corresponding sequence spaces and that is why it is important to look into their mathematical properties. First of all, given a topological space Y, we denote the space of bi-infinite and left semi-infinite countable Cartesian products by

$$\overline{Y} = \prod_{i=-\infty}^{\infty} Z_i \quad \text{and} \quad \overleftarrow{Y} = \prod_{i=-\infty}^{-1} Z_i, \quad \text{respectively, where} \quad Z_i = Y,$$

and equipped with the product topology. Alternatively, we can write $\overleftarrow{Y} = Y^{\mathbb{Z}^-}$, where $\mathbb{Z}^-$ (respectively $\mathbb{Z}_-$) denotes the set of strictly negative integer numbers (respectively $\mathbb{Z}^- \cup \{0\}$). Note that there is a natural projection $\pi_{\overleftarrow{Y}} : \overline{Y} \longrightarrow \overleftarrow{Y}$ that extracts from each bi-infinite sequence its left semi-infinite part. We also note that if Y is a Polish space then so are $\overline{Y}$ and $\overleftarrow{Y}$ since we are considering countable products. Additionally, if d_Y is a metric that makes Y complete and $\mathbf{w} \in \mathbb{R}^{\mathbb{N}}$ is a weighting sequence (zero-limit strictly decreasing sequence with $w_1 = 1$) then

$$d_{\overleftarrow{Y}}(\mathbf{x}, \mathbf{y}) = \sup_{i \in \mathbb{N}} \left\{ w_i \overline{d_Y}(x_{-i}, y_{-i}) \right\} \tag{3.1}$$

induces the product topology on $\overleftarrow{Y}$ (which does not depend on $\mathbf{w}$) and makes it into a Polish space (see [17, Theorem 2.6] for a proof). The symbol $\overline{d_Y}$ denotes the standard bounded metric in Y defined by $\overline{d_Y}(x, y) = \min\{d_Y(x, y), 1\}$. The distance $d_{\overleftarrow{Y}}$ can be used to define a corresponding Wasserstein-1 space $P_1(\overleftarrow{Y})$ and an associated Wasserstein distance on it that makes it in turn into a Polish space. This metric can be easily extended to bi-infinite sequences $\overline{Y}$. Indeed, one first formulates a metric similar to (3.1) for the space $\overrightarrow{Y}$ of semi-infinite sequences towards $+\infty$. Then, we write $\overline{Y}$ as the Cartesian product of $\overleftarrow{Y}$ and $\overrightarrow{Y}$, and we finally put together those two metrics by taking their maximum, which yields a metric $d_{\overline{Y}}$ for $\overline{Y}$.

Regarding notation, elements in sequence spaces will be written in bold and their entries in normal font. For any $t \in \mathbb{Z}^-$, we define **time delay** operator $T_{-t} : \overleftarrow{Y} \to \overleftarrow{Y}$ by $T_{-t}(\mathbf{y})_s = \mathbf{y}_{s+t}$. This definition can be extended to the case in which $\overleftarrow{Y}$ is replaced by $\overline{Y}$, in which case one can also consider the case $t \in \mathbb{Z}$.

Driven systems in sequence spaces. We now introduce driven systems in sequence spaces induced by the original driven system $g : U \times X \to X$. These objects will be central to the next developments in the paper. It is worth recalling at this point that when dealing with driven systems that have the USP one can make an identification between their solutions in the spaces $\overleftarrow{X}$ and $\overline{X}$ of left semi-infinite and bi-infinite sequences, respectively. More specifically, a bi-infinite solution $\mathbf{x} \in \overline{X}$ of g for the input $\mathbf{u} \in \overline{U}$ determines a unique left semi-infinite sequence $\overleftarrow{\mathbf{x}} = \pi_{\overleftarrow{X}}(\mathbf{x}) \in \overleftarrow{X}$ that we shall also call solution of g for the input $\mathbf{u} \in \overline{U}$. Analogously, given the input $\mathbf{u} \in \overline{U}$ and $\overleftarrow{\mathbf{x}} \in \overleftarrow{X}$ such that $\overleftarrow{x}_{n+1} = g(u_n, \overleftarrow{x}_n)$, for all $n \in \mathbb{Z}^-$, the same recursion determines an element $\mathbf{x} \in \overline{X}$ that solves g for $\mathbf{u}$ and that is necessarily the unique solution, as g has the USP by hypothesis. In the next pages, we shall use these two spaces interchangeably without making a distinction between $\mathbf{x}$ and $\overleftarrow{\mathbf{x}}$; the space to which these sequences belong will be either explicitly specified or determined by the context.

Definition 3.1. *Let $g : U \times X \to X$ be a driven system. We define its* **extension** $\mathbf{G} : \overleftarrow{U} \times \overleftarrow{X} \to \overleftarrow{X}$ **to sequence space** *by*

$$\mathbf{G}(\mathbf{u}, \mathbf{x}) = (\ldots, g(u_{-2}, x_{-2}), g(u_{-1}, x_{-1})). \tag{3.2}$$

Observe first that the semi-infinite solutions of the system associated to a driven system $g : U \times X \to X$ are exactly the fixed points of the map $T_1 \circ \mathbf{G}$ with T_1 the one-lag delay map. More specifically, $\mathbf{x} \in \overleftarrow{X}$ is a solution for the input $\mathbf{u} \in \overleftarrow{U}$ if and only if

$$T_1 \circ \mathbf{G}(\mathbf{u}, \mathbf{x}) = \mathbf{x}. \tag{3.3}$$

In what follows, we focus on the driven system associated to $\mathbf{G}$ in sequence space, its solutions, and their relation with those of the original driven system g. Given a sequence (of sequences) $\{\mathbf{u}_n\}_{n\in\mathbb{Z}}$ with elements in $\overleftarrow{U}$, we say that the sequence (of sequences) $\{\mathbf{x}_n\}_{n\in\mathbb{Z}}$ with elements in $\overleftarrow{X}$ is a **solution of G** for the input $\{\mathbf{u}_n\}_{n\in\mathbb{Z}}$ when

$$\mathbf{x}_{n+1} = \mathbf{G}(\mathbf{u}_n, \mathbf{x}_n), \quad \text{for all } n \in \mathbb{Z}. \tag{3.4}$$

In the following paragraphs we discuss the relation between the solutions of the driven systems associated to g and $\mathbf{G}$ and, more explicitly, the equivalence of the USP for the two systems. This statement is shown under two different sets of hypotheses. In Proposition 3.2 this is done under the assumption g and $\mathbf{G}$ have the existence of solutions property, that is, for every input they both have at least one solution. Subsequently, in Proposition 3.5 a similar equivalence is proved without the need for the existence of solutions hypothesis in $\mathbf{G}$ that, in exchange allows us to conclude the

USP for that map only for a class of inputs that we call time-folded sequences (see Definition 3.3).

Proposition 3.2. *Let $g : U \times X \to X$ be a driven system and let $\mathbf{G} : \overleftarrow{U} \times \overleftarrow{X} \to \overleftarrow{X}$ be its extension to sequence space. Suppose that these systems are such that for every input there exists at least one solution. Then, g has the USP if and only if $\mathbf{G}$ has the USP.*

Proof. We first prove that if g has the USP then so does $\mathbf{G}$. Assume that $\mathbf{G}$ does not have the USP. This means that there exists an input $\{\mathbf{u}_n\}_{n\in\mathbb{Z}}$ with elements in $\overleftarrow{U}$ for which there are two distinct solutions $\{\mathbf{x}_n\}_{n\in\mathbb{Z}}, \{\mathbf{y}_n\}_{n\in\mathbb{Z}}$ with elements in $\overleftarrow{X}$. This implies that $(\mathbf{x}_m)_k \neq (\mathbf{y}_m)_k$, for some $k \in \mathbb{Z}^-$ and $m \in \mathbb{Z}$. Now, since $\{\mathbf{x}_n\}_{n\in\mathbb{Z}}$ and $\{\mathbf{y}_n\}_{n\in\mathbb{Z}}$ are both solutions for the $\mathbf{G}$-system (3.4), they satisfy that: $g((\mathbf{u}_n)_j, (\mathbf{x}_n)_j) = (\mathbf{x}_{n+1})_j$ and $g((\mathbf{u}_n)_j, (\mathbf{y}_n)_j) = (\mathbf{y}_{n+1})_j$, for all $j \in \mathbb{Z}^-$ and all $n \in \mathbb{Z}$. In particular, since $(\mathbf{x}_m)_k \neq (\mathbf{y}_m)_k$, this implies that there are two different solutions $\{(\mathbf{x}_n)_k)\}_{n\in\mathbb{Z}}$ and $\{(\mathbf{y}_n)_k)\}_{n\in\mathbb{Z}}$ of the g-system for the same input $\{(\mathbf{u}_n)_k)\}_{n\in\mathbb{Z}}$ which contradicts the hypothesis that g has the USP.

Next, we show that if $\mathbf{G}$ has the USP then g has the USP. By contradiction, suppose that g does not have the USP and let $\mathbf{x}, \mathbf{y} \in \overline{X}$ be two distinct solutions for the same input $\mathbf{u} \in \overline{U}$. Define the input $\{\mathbf{u}_n\}_{n\in\mathbb{Z}}$ with elements in $\overleftarrow{U}$ by $(\mathbf{u}_n)_j = u_n$, for all $j \in \mathbb{Z}^-$. Also, define the sequences $\{\mathbf{x}_n\}_{n\in\mathbb{Z}}, \{\mathbf{y}_n\}_{n\in\mathbb{Z}}$ with elements in $\overleftarrow{X}$ by $(\mathbf{x}_n)_j = x_n$ and $(\mathbf{y}_n)_j = y_n$, for all $j \in \mathbb{Z}^-$. Clearly, by definition of $\mathbf{G}$ we have that $\mathbf{G}(\mathbf{u}_n, \mathbf{x}_n) = \mathbf{x}_{n+1}$ and $\mathbf{G}(\mathbf{u}_n, \mathbf{y}_n) = \mathbf{y}_{n+1}$, for all $n \in \mathbb{Z}$. This implies there are two different solutions of $\mathbf{G}$ for the same input, which implies $\mathbf{G}$ does not have the USP. ∎

The existence of solutions hypotheses on g and $\mathbf{G}$ in the previous proposition can be ensured under very general hypotheses. For instance, if the state space X is compact and convex, it can be shown [17, Theorem 3.1**(i)**] that the g and the $\mathbf{G}$-systems have solutions for any input. This is a consequence of Schauder's Fixed Point Theorem (see [38, Theorem 7.1, page 75]) when the product topology is used in the corresponding sequence spaces. An extension of this result to a non-compact framework can be found in [19, Theorem 7**(ii)**]. It is worth emphasizing that much like autonomous systems defined on unbounded spaces exhibit interesting dynamics on bounded invariant sets, the relevant dynamics of many non-autonomous systems induced by driven systems is contained in bounded absorbing sets [24]. This all implies that the hypothesis in the previous proposition on the existence of solutions is in practice not a strong one.

Having said all this, another equivalence result for the equivalence of the USP for g and $\mathbf{G}$ can be formulated in which there is no need to invoke an *a priori* knowledge on the existence of solutions for them. The price to pay for this added generality is the restriction the inputs for the $\mathbf{G}$ system to what we call **time-folded** inputs, a notion that we introduce in the next definition.

Definition 3.3. *Let $\{\mathbf{x}_n\}_{n\in\mathbb{Z}}$ be a sequence of elements in the space $\overleftarrow{X}$ of left semi-infinite sequences in X. We say that the sequence (of sequences) $\{\mathbf{x}_n\}_{n\in\mathbb{Z}}$ is* **time-folded** *whenever*

$$T_{-t}\mathbf{x}_n = \mathbf{x}_{n+t}, \quad \text{for all } t \in \mathbb{Z}_- \text{ and } n \in \mathbb{Z}. \tag{3.5}$$

The time delay operator $T_{-t} : \overleftarrow{X} \longrightarrow \overleftarrow{X}$, $t \in \mathbb{Z}_-$, in the definition is the one that was already introduced at the end of Section 2, in view of which, the time-folding relation

(3.5) can be rewritten as

$$(\mathbf{x}_n)_{s+t} = (\mathbf{x}_{n+t})_s, \quad \text{for all } t, s \in \mathbb{Z}_- \text{ and } n \in \mathbb{Z}.$$

The following lemma shows that time-folded sequences in $\overleftarrow{X}$ have a very simple structure and that all their terms can be constructed out of a single element in $\overline{X}$.

LEMMA 3.4. *Let $\{\mathbf{x}_n\}_{n\in\mathbb{Z}}$ be a time-folded sequence of elements in $\overleftarrow{X}$. Then, there exists a unique sequence $\mathbf{x}^0 \in \overline{X}$ such that*

$$\mathbf{x}_n = \mathbf{x}^0_{(-\infty,n]}, \quad \text{for all } n \in \mathbb{Z}, \tag{3.6}$$

where $\mathbf{x}^0_{(-\infty,n]} \in \overleftarrow{X}$ is defined by $\mathbf{x}^0_{(-\infty,n]} = (\ldots, x^0_{n-1}, x^0_n)$ or, equivalently, by

$$\left(\mathbf{x}^0_{(-\infty,n]}\right)_j = x^0_{n+j+1}, \quad \text{for all } j \in \mathbb{Z}^-. \tag{3.7}$$

Proof. Let $\mathbf{x}^0 \in \overline{X}$ be the sequence defined by

$$x^0_n = (\mathbf{x}_n)_{-1}, \quad \text{for all } n \in \mathbb{Z}. \tag{3.8}$$

We now verify that the invariance condition of $\{\mathbf{x}_n\}_{n\in\mathbb{Z}}$ implies the relation (3.6). Indeed, for any $j \in \mathbb{Z}^-$ and $n \in \mathbb{Z}$ we have that

$$(\mathbf{x}_n)_j = \left(T_{-(j+1)}\mathbf{x}_n\right)_{-1} = (\mathbf{x}_{n+j+1})_{-1} = x^0_{n+j+1} = \left(\mathbf{x}^0_{(-\infty,n]}\right)_j,$$

and hence $\mathbf{x}_n = \mathbf{x}^0_{(-\infty,n]}$, for all $n \in \mathbb{Z}$, as required. In the previous expression, the first equality follows from the definition of the time delay operator, the second one follows from the time-folding hypothesis, the third one from (3.8), and the last one from (3.7). ■

PROPOSITION 3.5. *Let $g : U \times X \to X$ be a driven system and let $\mathbf{G} : \overleftarrow{U} \times \overleftarrow{X} \to \overleftarrow{X}$ be its extension to sequence space. Then $\mathbf{x} = \{x_n\}_{n\in\mathbb{Z}}$ is a solution of g for the input $\mathbf{u} = \{u_n\}_{n\in\mathbb{Z}}$ if and only if the sequence $\{\mathbf{x}_{(-\infty,n]}\}_{n\in\mathbb{Z}}$ in $\overleftarrow{X}$ is a solution of $\mathbf{G}$ for the input sequence $\{\mathbf{u}_{(-\infty,n]}\}_{n\in\mathbb{Z}}$ in $\overleftarrow{U}$, where $\mathbf{x}_{(-\infty,n]}$ and $\mathbf{u}_{(-\infty,n]}$ are defined as in* (3.7)*. Consequently, the driven system g has the USP if and only if the induced map $\mathbf{G}$ in sequence space has the USP when restricted to time-folded inputs.*

Proof. Suppose first that $\mathbf{x} = \{x_n\}_{n\in\mathbb{Z}}$ is a solution of g for the input $\mathbf{u} = \{u_n\}_{n\in\mathbb{Z}}$, that is, $x_{j+1} = g(u_j, x_j)$ for all $j \in \mathbb{Z}$. We now show that $\{\mathbf{x}_{(-\infty,n]}\}_{n\in\mathbb{Z}}$ is a solution of $\mathbf{G}$ for the input $\{\mathbf{u}_{(-\infty,n]}\}_{n\in\mathbb{Z}}$. To prove this, we just need to verify $\mathbf{G}(\mathbf{u}_{(-\infty,n]}, \mathbf{x}_{(-\infty,n]}) = \mathbf{x}_{(-\infty,n+1])}$ for all $n \in \mathbb{Z}$. By the definition of $\mathbf{G}$, $\mathbf{G}(\mathbf{u}_{(-\infty,n]}, \mathbf{x}_{(-\infty,n]}) = (\ldots, g(u_{n-1}, x_{n-1}), g(u_n, x_n))$ for any $n \in \mathbb{Z}$. Since $x_{j+1} = g(u_j, x_j)$ for all $j \in \mathbb{Z}$, we have that $\mathbf{G}(\mathbf{u}_{(-\infty,n]}, \mathbf{x}_{(-\infty,n]}) = (\ldots, x_n, x_{n+1}) = \mathbf{x}_{(-\infty,n+1]}$, as required.

Conversely, suppose that $\{\mathbf{x}_{(-\infty,n]}\}_{n\in\mathbb{Z}}$ is a solution of the extension $\mathbf{G}$, for some input sequence $\{\mathbf{u}_{(-\infty,n]}\}_{n\in\mathbb{Z}}$ in $\overleftarrow{U}$, that is, $\mathbf{G}(\mathbf{u}_{(-\infty,n]}, \mathbf{x}_{(-\infty,n]}) = \mathbf{x}_{(-\infty,n+1]}$, for all $n \in \mathbb{Z}$. By the definition of $\mathbf{G}$, $\mathbf{G}(\mathbf{u}_{(-\infty,n]}, \mathbf{x}_{(-\infty,n]}) = (\ldots, g(u_{n-1}, x_{n-1}), g(u_n, x_n))$. Therefore, $\mathbf{x}_{(-\infty,n+1]} = (\ldots, g(u_{n-1}, x_{n-1}), g(u_n, x_n))$, which implies $x_{j+1} = g(u_j, x_j)$, for all $j \leq n$. Since n is arbitrary, we can conclude that the sequence $\mathbf{x} = \{x_n\}_{n\in\mathbb{Z}}$ is a solution of g for the input $\mathbf{u} = \{u_n\}_{n\in\mathbb{Z}}$.

Finally, since the sequences $\mathbf{u} = \{u_n\}_{n\in\mathbb{Z}} \in \overline{U}$ and $\mathbf{x} = \{x_n\}_{n\in\mathbb{Z}} \in \overline{X}$ uniquely determine the time-folded sequences $\{\mathbf{u}_{(-\infty,n]}\}_{n\in\mathbb{Z}}$ and $\{\mathbf{x}_{(-\infty,n]}\}_{n\in\mathbb{Z}}$, respectively, and vice-versa, the two implications that we just proved imply that g has the USP if and only if $\mathbf{G}$ has the USP when restricted to time-folded inputs. ■

Remark 3.6. If in the last statement in the previous proposition we drop the restriction to time-folded inputs, the claim is in general false (unless we add the existence of solutions property as a hypothesis as we did in Proposition 3.2). More specifically, even if g has the USP, the system in sequence space induced by the corresponding $\mathbf{G}$ may not have that property. As an example, consider the one-dimensional linear system $g(u,x) = ax + u$, $|a| < 1$, for which $X = U = \mathbb{R}$ and the sequence of (non-time-folded) input sequences given by $(\mathbf{u}_n)_t = n^2/t^2$, $n \in \mathbb{Z}$, $t \in \mathbb{Z}^-$. Note that for any $n \in \mathbb{Z}$, the system induced by g has a unique solution $\mathbf{x}_n \in \overleftarrow{X}$ associated given by $(\mathbf{x}_n)_t = n^2 \sum_{j=0}^{\infty} \frac{a^j}{(t-j)^2}$, $n \in \mathbb{Z}$, $t \in \mathbb{Z}^-$. On the other hand, the solutions $\{\overline{\mathbf{x}}_n\}_{n\in\mathbb{Z}}$ for the $\mathbf{G}$ system in sequence space that have $\{\mathbf{u}_n\}_{n\in\mathbb{Z}}$ as input satisfy that $\overline{\mathbf{x}}_{n+1} = \mathbf{G}(\overline{\mathbf{x}}_n, \mathbf{u}_n)$, for all $n \in \mathbb{Z}$ or, equivalently, that $a\,(\overline{\mathbf{x}}_n)_t + (\mathbf{u}_n)_t = (\overline{\mathbf{x}}_{n+1})_t$, for all $n \in \mathbb{Z}$ and $t \in \mathbb{Z}^-$. This relation implies that if a solution $\{\overline{\mathbf{x}}_n\}_{n\in\mathbb{Z}}$ exists, it must satisfy that $(\overline{\mathbf{x}}_n)_t = \frac{1}{t^2}\sum_{j=0}^{\infty} a^j (n-j)^2$. Since this series is divergent, we can conclude that the system induced by $\mathbf{G}$ does not hence have the USP for these inputs.

4. Stochastic contractions and invariant measures for driven systems. We place ourselves in this section in a setup similar to the one in Definition 2.1 which ensures the existence of a well-defined Foias operator by using, for instance, the hypotheses that we introduced in Proposition 2.3. The main goal in the following pages is proving the **existence** and **uniqueness** of invariant measures for the Foias operators in both the state and sequence spaces and the **continuity** of their dependences on the input process. The main tool to achieve the results requires two conditions, namely contractivity and continuity. These two conditions will be treated for the maps of interest in two different subsections.

We start by formally introducing the notion of **stochastic state contraction** which will be at the core of our developments and whose importance is given by the fact that it is, in general, less restrictive than the standard contractivity conditions evoked to ensure that (2.1) has the unique solution property (see, for instance, [21, 18, 17]). This concept was already informally discussed in the introduction.

DEFINITION 4.1. *Let $g : U \times X \to X$ be a measurable driven system that has as domain the input (U, d_U) and state (X, d_X) spaces that are assumed to be Polish spaces. Given $\theta \in P(U)$ and $0 < c < 1$, we say that g is a (θ, c)-**contraction** when*

$$\int_U d_X(g_u(x), g_u(y))\, d\theta(u) \leq c\, d_X(x,y) \quad \text{holds for all } x, y \in X. \tag{4.1}$$

We emphasize that given a fixed driven system$g$ the constant c in (4.1) depends in general on the input process $\theta \in P(U)$. This leads us to define the **optimal** or **best contraction constant** for a given input process $\theta \in P(U)$ as

$$c_\theta = \inf\{c \in (0,1) \text{ such that (4.1) holds}\}. \tag{4.2}$$

This condition is satisfied by many parametric models commonly used in time series analysis, as seen in the examples below that have been explored also in [13] using a different approach.

Example 4.2 (**VARMA process with time-varying coefficients).** *Suppose* $\mathbf{X} = (X_t)_{t\in\mathbb{Z}}$, *with* $X_t \in \mathbb{R}^N$, *is a vector autoregressive process of first order with time-varying coefficients on* $\mathbb{R}^N$ *endowed with the Euclidean metric, which we write as:*

$$X_t = A(u_{t-1})X_{t-1} + f(u_{t-1}), \quad t \in \mathbb{Z}, \tag{4.3}$$

where $f : \mathbb{R}^d \to \mathbb{R}^N$ *is a measurable map and* $\mathbf{u} = (u_t)_{t\in\mathbb{Z}} \sim \mathrm{IID}$ *with* $u_t \in \mathbb{R}^d$, *that is,* $\mathbf{u}$ *is a* $\mathbb{R}^d$*-valued sequence of independent and identically distributed random variables. The matrix* $A(u) \in \mathbb{M}_N$ *is assumed to satisfy that* $\mathrm{E}\left[\|A(u)\|\right] < 1$, *where* $\|A(u)\|$ *denotes the operator norm with respect to the Euclidean metric in* $\mathbb{R}^N$ *(recall that* $\|A(u)\| = \sigma_{\max}(A)$*, the top singular eigenvalue of* A*).*

The recursions (4.3) *can be encoded as a driven system of the form* (2.1) *by defining* $g : \mathbb{R}^d \times \mathbb{R}^N \to \mathbb{R}^N$ *as* $g(u,x) = A(u)x + f(u)$. *Let now* $\theta \in P(U)$ *be the law of the components of* $\mathbf{u}$. *Then,* g *is a* $(\theta, \mathrm{E}\left[\|A(u)\|\right])$*-contraction. Indeed, in this case:*

$$\begin{aligned}\int_U d_X(g_u(X), g_u(Y))\, d\theta(u) &= \int_U \|A(u)(X-Y)\|\ d\theta(u)\\ &\leq \int_U \|A(u)\|\,\|X-Y\|\ d\theta(u) \leq \mathrm{E}\left[\|A(u)\|\right]\ d_X(X,Y).\end{aligned}$$

We emphasize that the condition $\mathrm{E}\left[\|A(u)\|\right] < 1$ *is in general less restrictive than* $\|A(u)\| < 1$, *for all* $u \in U$, *which would be the standard condition evoked to ensure that* (4.3) *has the unique solution property (see, for instance, [17, Theorem 3.1]).*

Example 4.3. **GARCH process** We now consider a particular case of the model introduced in (4.3) which is extensively used to describe and eventually to forecast the volatility of financial time series, namely the generalized autoregressive conditional heterostedastic (GARCH) family [11, 4, 12]. The GARCH(1,1) model given by the following equations:

$$r_t = \sigma_t u_{t-1}, \quad u_t \sim \mathrm{IID}(0,1), \quad t \in \mathbb{Z} \tag{4.4}$$

$$\sigma_t^2 = \omega + \alpha r_{t-1}^2 + \beta\sigma_{t-1}^2, \quad t \in \mathbb{Z} \tag{4.5}$$

with parameters that satisfy $\alpha, \beta, \omega \geq 0$, $\alpha + \beta < 1$. We now show that the GARCH(1,1) process in (4.4)-(4.5) falls in the framework introduced in the previous example. Indeed, define

$$X_t = \begin{pmatrix} r_t^2 \\ \sigma_t^2 \end{pmatrix}, \quad f(u_t) := \begin{pmatrix} \omega u_t^2 \\ \omega \end{pmatrix}, \quad A(u_t) = \begin{pmatrix} \alpha u_t^2 & \beta u_t^2 \\ \alpha & \beta \end{pmatrix}, \quad t \in \mathbb{Z}.$$

It is easy to verify that with this choice one has $\mathrm{E}[\|A(u)\|] = \mathrm{E}[\alpha u^2 + \beta] = \alpha + \beta < 1$. In this case it is particularly obvious that the condition $\mathrm{E}[\|A(u)\|] < 1$ is much less restrictive than $\|A(u)\| = \alpha u^2 + \beta < 1$, for all $u \in U$, which is false, for instance when the innovations u_t are not bounded.

Remark 4.4 (**Stochastic contractivity without the USP).** Consider $g : [0,1] \times [0,1] \to [0,1]$ defined by $g(u,x) = ux$, where subsets of $\mathbb{R}$ are endowed with standard Euclidean metric. The system does not have the USP since for an input sequence comprising of just ones, every constant sequence contained in $[0,1]$ is a solution for that input. On the other hand, if $\theta = \delta_0$, that is, θ is a point measure at 0, we observe that $\int_U d_X(g_u(x), g_u(y))\, d\theta(u) = 0 < c\, d_X(x,y)$, for every value of $c > 0$.

4.1. Stochastic contractivity and contractive Foias operators. This subsection shows that the stochastic contractivity of a driven system guarantees that its Foias operator is a contraction with respect to the Wasserstein distance (see Theorem 4.5). Moreover, we also spell out conditions, mostly the stationarity of the input process, which guarantee that the corresponding driven system in sequence space and its Foias operator are also a contraction (see Propositions 4.7 and 4.6). All these different implications are summarized in Figure 3.

Stochastic state contractivity yields contractive Foias operators. The following results shows that if a driven system is stochastic state contractive with respect to a fixed measure in the input space, then the corresponding Foias operator has the same property.

THEOREM 4.5. *Let $g : U \times X \to X$ be a measurable driven system that has as domain the input (U, d_U) and state (X, d_X) spaces that are assumed to be Polish. Fix $\theta \in P(U)$ and suppose that g is a (θ, c)-contraction with $0 < c < 1$. If g has a well defined Foias operator $P_g : P(U) \times P_1(X) \to P_1(X)$ then it is necessarily a c-contraction with respect to the Wasserstein-1 distance on the second entry, that is,*

$$W(P_g(\theta, \mu_1), P_g(\theta, \mu_2)) \leq c\, W(\mu_1, \mu_2), \text{ for any } \mu_1, \mu_2 \in P_1(X). \tag{4.6}$$

Proof. First of all, given $f \in \mathrm{Lip}_1(X, \mathbb{R})$, define the function

$$r_f(x) = \frac{1}{c} \int_U f(g_u(x))\, d\theta(u).$$

It is easy to show that $r_f \in \mathrm{Lip}_1(X, \mathbb{R})$. Indeed, for any $x, y \in X$,

$$\begin{aligned}
|r_f(x) - r_f(y)| &= \left| \frac{1}{c} \int_U f(g_u(x))\, d\theta(u) - \frac{1}{c} \int_U f(g_u(y))\, d\theta(u) \right| \\
&\leq \frac{1}{c} \int_U |f(g_u(x)) - f(g_u(y))|\, d\theta(u) \leq \frac{1}{c} \int_U d_X(g_u(x),\, g_u(y)) d\theta(u) \leq d_X(x, y).
\end{aligned}$$

We now establish (4.6). Take $\mu_1, \mu_2 \in P_1(X)$ arbitrary and let $\tau_1 = \theta \times \mu_1, \tau_2 = \theta \times \mu_2 \in P(U \times X)$ be the product measures. Then, using the characterization in (2.8) of the Foias operator, we have:

$$\begin{aligned}
W(P_g(\theta, \mu_1), P_g(\theta, \mu_2)) &= \sup_{f \in \mathrm{Lip}_1(X, \mathbb{R})} \left(\int_X f(x) d\, P_g(\theta, \mu_1)(x) - \int_X f(x) d\, P_g(\theta, \mu_2)(x) \right) \\
&= \sup_{f \in \mathrm{Lip}_1(X, \mathbb{R})} \left(\int_X f(x)\, d\, (g_* \tau_1(x)) - \int_X f(x)\, d\, (g_* \tau_1(x)) \right) \\
&= \sup_{f \in \mathrm{Lip}_1(X, \mathbb{R})} \left(\int_{U \times X} f \circ g(u, x)\, d\, (\tau_1(u, x) - \tau_2(u, x)) \right) \\
&= \sup_{f \in \mathrm{Lip}_1(X, \mathbb{R})} \left(\int_U \int_X f(g_u(x)) d\Big(\mu_1(x) - \mu_2(x)\Big) d\theta(u) \right), \\
&= \sup_{f \in \mathrm{Lip}_1(X, \mathbb{R})} \left(\int_X \Big(\int_U f(g_u(x)) d\theta(u) \Big) d\Big(\mu_1(x) - \mu_2(x)\Big) \right) \\
&= c \sup_{f \in \mathrm{Lip}_1(X, \mathbb{R})} \left(\int_X r_f(x) d\Big(\mu_1(x) - \mu_2(x)\Big) \right) \leq c\, W(\mu_1, \mu_2),
\end{aligned}$$

where Fubini's theorem was used in the fourth and the fifth equalities, which is available because $f \in \mathrm{Lip}_1(X, \mathbb{R})$. The last inequality follows from the fact that $r_f \in \mathrm{Lip}_1(X, \mathbb{R})$. ◻

The conclusion in the previous theorem can be immediately applied to induced driven systems in sequence spaces, in which case, the metric d_X in the contractivity condition (4.1) has to be replaced by a bounded weighted metric $d_{\overleftarrow{X}}$ of the type that we introduced in (3.1). We shall see later on in Proposition 4.7 that the stochastic contractivity in sequence space is naturally inherited under very general hypotheses from a stochastic contractivity hypothesis for the original system $g : U \times X \to X$.

PROPOSITION 4.6. *Let $g : U \times X \to X$ be a measurable driven system with Polish input and output spaces and let $\mathbf{G} : \overleftarrow{U} \times \overleftarrow{X} \to \overleftarrow{X}$ be the induced driven system in sequence space defined in* (3.2). *The induced system has a well-defined Foias operator $P_{\mathbf{G}} : P(\overleftarrow{U}) \times P_1(\overleftarrow{X}) \to P_1(\overleftarrow{X})$ associated with it. Moreover, let $\Theta \in P(\overleftarrow{U})$ and suppose that $\mathbf{G}$ is a (Θ, c)-stochastic contraction, then $P_{\mathbf{G}}(\Theta, \cdot)$ is also a c-contraction with the Wasserstein-1 metric.*

Proof. The operator $P_{\mathbf{G}} : P(\overleftarrow{U}) \times P_1(\overleftarrow{X}) \to P_1(\overleftarrow{X})$ is well-defined because of the boundedness of the metric $d_{\overleftarrow{X}}$ in (3.1) and part **(i)** of Proposition 2.3. We recall that this metric induces the product topology and makes $\overleftarrow{X}$ into a Polish space. With this in mind, the contractivity claim is a direct corollary of Theorem 4.5 that is obtained by replacing g by $\mathbf{G}$. ◻

Contractive driven systems and their counterparts in sequence spaces. There are situations in which the previous corollary exhibits a special significance, namely, when the contractivity hypothesis on $\mathbf{G}$ can be obtained out of a contractivity hypothesis on the driven system g that generates it. In the next result we show that is the case when, for instance, X is bounded and the input process is stationary.

PROPOSITION 4.7. *Let $g : U \times X \to X$ be a measurable driven system with Polish input and output spaces and let $\mathbf{G} : \overleftarrow{U} \times \overleftarrow{X} \to \overleftarrow{X}$ be the induced driven system in sequence space defined in* (3.2). *Additionally, suppose that (X, d_X) is bounded and let $\Theta \in P(\overleftarrow{U})$ be the law of a stationary process with time-independent marginals $\theta \in P(U)$. Then, if g is a (θ, c)-contraction then $\mathbf{G}$ is also a (Θ, c)-contraction. More specifically, when in $\overleftarrow{X}$ we consider any weighted metric $d_{\overleftarrow{X}}$ of the type introduced in* (3.1)*, we have that:*

$$\int_{\overleftarrow{U}} d_{\overleftarrow{X}}(\mathbf{G}(\mathbf{u}, \mathbf{x}), \mathbf{G}(\mathbf{u}, \mathbf{y}))\, d\Theta(\mathbf{u}) \leq c\, d_{\overleftarrow{X}}(\mathbf{x}, \mathbf{y}), \quad \textit{for all } \mathbf{x}, \mathbf{y} \in \overleftarrow{X}. \tag{4.7}$$

Proof. First of all, the boundedness hypothesis on X allows us to define a weighted metric (3.1) in $\overleftarrow{X}$ without using the bounded metric $\overline{d_X}$ and by replacing it in the definition just by d_X. More explicitly, given the weighted sequence $\mathbf{w}$, the expression

$$d_{\overleftarrow{X}}(\mathbf{x}, \mathbf{y}) = \sup_{i \in \mathbb{N}} \{w_i d_X(x_{-i}, y_{-i})\} \tag{4.8}$$

defines a metric on $\overleftarrow{X}$ that induces the product topology.

Now, since g is a (θ, c)-contraction, we have that for each $t \in \mathbb{Z}^-$ and each

$\mathbf{x}, \mathbf{y} \in \overleftarrow{X}$

$$w_{-t} \int_U d_X(g_u(x_t), g_u(y_t))\, d\theta(u) \leq c w_{-t}\, d_X(x_t, y_t) \leq c \sup_{s \in \mathbb{Z}^-} \{w_{-s}\, d_X(x_s, y_s)\} = c\, d_{\overleftarrow{X}}(\mathbf{x}, \mathbf{y}). \tag{4.9}$$

Now, in order to prove the claim (4.7), define for fixed $\mathbf{x}, \mathbf{y} \in \overleftarrow{X}$ and $t \in \mathbb{Z}^-$ the sequence of functions

$$\begin{array}{rccl} f_t : & \overleftarrow{U} & \longrightarrow & \mathbb{R} \\ & \mathbf{u} & \longmapsto & \max_{s \in \{t, t+1, \ldots, -1\}} \{w_{-s} d_X(g_{u_s}(x_s), g_{u_s}(y_s))\}. \end{array}$$

It is clear that

$$f_t(\mathbf{u}) \leq d_{\overleftarrow{X}}(\mathbf{G}(\mathbf{u}, \mathbf{x}), \mathbf{G}(\mathbf{u}, \mathbf{y})) \text{ and that } \lim_{t \to -\infty} f_t(\mathbf{u}) = d_{\overleftarrow{X}}(\mathbf{G}(\mathbf{u}, \mathbf{x}), \mathbf{G}(\mathbf{u}, \mathbf{y})).$$

Additionally, $f_{t-1}(\mathbf{u}) \geq f_t(\mathbf{u})$, for any $t \in \mathbb{Z}^-$, and hence the Monotone Convergence Theorem allows us to conclude that

$$\int_{\overleftarrow{U}} d_{\overleftarrow{X}}(\mathbf{G}(\mathbf{u}, \mathbf{x}), \mathbf{G}(\mathbf{u}, \mathbf{y}))\, d\Theta(\mathbf{u}) = \lim_{t \to -\infty} \int_{\overleftarrow{U}} f_t(\mathbf{u})\, d\Theta(\mathbf{u}). \tag{4.10}$$

We now bound $\int_{\overleftarrow{U}} f_t(\mathbf{u})\, d\Theta(\mathbf{u})$ by defining, for any $s \in \{t, t+1, \ldots, -1\}$, the set $A_s \subset \overleftarrow{U}$ given by

$$A_s = \left\{ \mathbf{u} \in \overleftarrow{U} \mid f_t(\mathbf{u}) = w_{-s} d_X(\mathbf{G}(\mathbf{u}, \mathbf{x})_s, \mathbf{G}(\mathbf{u}, \mathbf{y})_s) \right\}.$$

In other words, A_s is the measurable subset of $\overleftarrow{U}$ for which the maximum that defines the map f_t is realized for the index s. Using this notation and the inequality in (4.9), we can write:

$$\begin{aligned} \int_{\overleftarrow{U}} f_t(\mathbf{u})\, d\Theta(\mathbf{u}) &= \sum_{s=t}^{-1} \int_{A_s} f_t(\mathbf{u})\, d\Theta(\mathbf{u}) = \sum_{s=t}^{-1} \int_{A_s} w_{-s} d_X(\mathbf{G}(\mathbf{u}, \mathbf{x})_s, \mathbf{G}(\mathbf{u}, \mathbf{y})_s)\, d\Theta(\mathbf{u}) \\ &= \sum_{s=t}^{-1} \int_{A_s} w_{-s} d_X(g_{u_s}(x_s), g_{u_s}(y_s))\, d\Theta(\mathbf{u}) \leq \sum_{s=t}^{-1} \int_{A_s} c d_{\overleftarrow{X}}(\mathbf{x}, \mathbf{y})\ d\Theta(\mathbf{u}) \\ &= c d_{\overleftarrow{X}}(\mathbf{x}, \mathbf{y}) \sum_{s=t}^{-1} \int_{A_s} d\Theta(\mathbf{u}) = c d_{\overleftarrow{X}}(\mathbf{x}, \mathbf{y}), \end{aligned}$$

that is,

$$\int_{\overleftarrow{U}} f_t(\mathbf{u})\, d\Theta(\mathbf{u}) \leq c d_{\overleftarrow{X}}(\mathbf{x}, \mathbf{y}), \quad \text{for all } t \in \mathbb{Z}^-.$$

Consequently, by (4.10):

$$\int_{\overleftarrow{U}} d_{\overleftarrow{X}}(\mathbf{G}(\mathbf{u}, \mathbf{x}), \mathbf{G}(\mathbf{u}, \mathbf{y}))\, d\Theta(\mathbf{u}) = \lim_{t \to -\infty} \int_{\overleftarrow{U}} f_t(\mathbf{u})\, d\Theta(\mathbf{u}) \leq c d_{\overleftarrow{X}}(\mathbf{x}, \mathbf{y}),$$

as required. ■

4.2. Continuity of driven systems and their Foias operators. Given a driven system $g : U \times X \to X$ between Polish spaces for which the Foias operator P_g exists (see, for instance, the conditions in Proposition 2.3) it is natural to study if the continuity of the driven system induces the same property in P_g. More specifically, if we consider the restriction $P_g : P_1(U) \times P_1(X) \to P_1(X)$ to input measures in $P_1(U)$, then both the domain and the target of P_g are endowed with the Wasserstein-1 metric and hence the continuity of this map can be studied. The following uniform continuity hypothesis will be needed in the sequel.

DEFINITION 4.8. *Let $g : U \times X \to X$ be a driven system with Polish input and output spaces. We say that g is* **uniformly continuous on the first entry** *if for any $\epsilon > 0$ there exists $\delta(\epsilon) > 0$ such that if $d_U(u, v) < \delta(\epsilon)$ then $d_X(g(u, x), g(v, x)) < \epsilon$, for all $x \in X$. This definition is extended to the* **uniform continuity on the second entry** *in a straightforward manner.*

THEOREM 4.9. *Let $g : U \times X \to X$ be a measurable driven system with Polish input and output spaces and (X, d_X) compact. If g is uniformly continuous on the first entry, then the corresponding Foias operator $P_g : P_1(U) \times P_1(X) \to P_1(X)$ is continuous on the first entry, that is, the maps $P_g(\cdot, \mu) : P_1(U) \to P_1(X)$ are all continuous for any $\mu \in P_1(X)$.*

Before we present the proof we introduce the following lemma.

LEMMA 4.10. *Let $g : U \times X \to X$ be a measurable driven system that is uniformly continuous on the first entry, with Polish input and output spaces, and (X, d_X) compact. Let $f : X \to \mathbb{R}$ be a continuous (and hence uniformly continuous) function. Given $\mu \in P(X)$, the map $s_f : U \to \mathbb{R}$ defined by*

$$s_f(u) = \int_X f(g(u, x))d\mu(x) \tag{4.11}$$

is uniformly continuous and bounded.

Proof of Lemma 4.10. By the compactness of X and the continuity of f, there exists $M > 0$ such that $|f(x)| \leq M$ for all $x \in M$. This implies that for any $u \in U$:

$$|s_f(u)| = \left| \int_X f(g(u, x))d\mu(x) \right| \leq \int_X M d\mu(x) = M,$$

which proves that s_f is bounded. We now prove that s_f is uniformly continuous. Let $\epsilon > 0$ and let $\delta(\epsilon) > 0$ be the scalar that by the uniform continuity of f implies that

$$|f(x) - f(y)| < \epsilon \text{ whenever } d_X(x, y) < \delta(\epsilon). \tag{4.12}$$

Let now $\delta'(\epsilon) > 0$ be the element that by the uniform continuity of g on the first entry guarantees that

$$d_X(g(u, x), g(v, x)) < \delta(\epsilon) \text{ for any } x \in X \text{ and whenever } d_U(u, v) < \delta'(\epsilon). \tag{4.13}$$

Now, if $u, v \in U$ are such that $d_U(u, v) < \delta'(\epsilon)$ then:

$$\begin{aligned} |s_f(u) - s_f(v)| &= \left| \int_X (f(g(u, x)) - f(g(v, x)))\, d\mu(x) \right| \\ &\leq \int_X |(f(g(u, x)) - f(g(v, x)))|\, d\mu(x) < \epsilon, \end{aligned}$$

where the last inequality is a direct consequence of (4.13) and (4.12), which proves the uniform continuity of s_f. ■

Proof of Theorem 4.9. We will proceed by using the fact that the Wasserstein distance metrizes the weak convergence as characterized in (2.5). Let $\mu \in P_1(X)$ be arbitrary but fixed and let $\{\theta_n\}_{n\in\mathbb{N}}$ be a convergent sequence in $P_1(U)$, that is, there exists $\theta \in P_1(U)$ such that

$$\lim_{n\to\infty} W(\theta_n, \theta) = 0. \tag{4.14}$$

The continuity property that we are interested in is guaranteed if $\lim_{n\to\infty} W(P_g(\theta_n,\mu), P_g(\theta,\mu)) = 0$, which by (2.5) is established if for any continuous function $f : X \to \mathbb{R}$ that satisfies that $|f(x)| \leq C(1 + d_X(x, x_0))$, we have that

$$\lim_{n\to\infty} \int_X f(x)\, dP_g(\theta_n, \mu)(x) = \int_X f(x)\, dP_g(\theta, \mu)(x). \tag{4.15}$$

Using the notation introduced in Lemma 4.10 we rewrite

$$\begin{aligned}\int_X f(x)\, dP_g(\theta_n, \mu)(x) &= \int_X \int_U f(g(u,x))\, d\theta_n(u)\, d\mu(x)\\ &= \int_U \int_X f(g(u,x))\, d\mu(x)\, d\theta_n(u) = \int_U s_f(u)\, d\theta_n(u).\end{aligned}$$

Consequently, the equality (4.15) holds whenever

$$\lim_{n\to\infty} \int_U s_f(u)\, d\theta_n(u) = \int_U s_f(u)\, d\theta(u),$$

which is the case because by (4.14) and by (2.5) we have that $\lim_{n\to\infty} \int_U h(u)\, d\theta_n(u) = \int_U h(u)\, d\theta(u)$, for all $h : U \to \mathbb{R}$ continuous such that $|h(x)| < C(1 + d_U(u_0, u))$, for some $C > 0$. The map s_f has that property because due to Lemma 4.10 it is continuous and bounded by some constant $M > 0$, and hence $|s_f(x)| < M \leq M(1 + d_U(u_0, u))$. ■

The following result extends the continuity statement in the previous theorem to the induced Foias operator $P_{\mathbf{G}} : P(\overleftarrow{U}) \times P_1(\overleftarrow{X}) \to P_1(\overleftarrow{X})$ on the sequence space.

COROLLARY 4.11. *Let $g : U \times X \to X$ be a measurable driven system with Polish input and output spaces, and X compact. Let $\mathbf{G} : \overleftarrow{U} \times \overleftarrow{X} \to \overleftarrow{X}$ be the induced driven system in sequence space as defined in* (3.2) *and assume that* $\mathbf{G}$ *is uniformly continuous on the first entry. Then, the corresponding Foias operator $P_{\mathbf{G}} : P_1(\overleftarrow{U}) \times P_1(\overleftarrow{X}) \to P_1(\overleftarrow{X})$ is continuous on the first entry.*

Proof. It can be obtained in a straightforward manner from Theorem 4.9 by replacing in its statement the driven system g by $\mathbf{G}$, which inherits measurability from g. Note that if U and X are Polish then so are $\overleftarrow{U}$ and $\overleftarrow{X}$. Moreover, $\overleftarrow{X}$ is also compact because of Tychonoff's Theorem [35, Theorem 37.3]. Finally, the induced Foias operator $P_{\mathbf{G}} : P(\overleftarrow{U}) \times P_1(\overleftarrow{X}) \to P_1(\overleftarrow{X})$ is guaranteed to be well-defined by part **(i)** in Proposition 2.3 and by the boundedness of the metric (3.1) on the product space. ■

In order to apply Corollary 4.11 on $\mathbf{G}$, we shall now establish conditions on g in the next corollary which guarantee the uniform continuity of $\mathbf{G}$ on the first entry.

Before we state it, we define some general properties that a metric generating the product topology can possess.

DEFINITION 4.12. *Let* (Y, d_Y) *be a metric space and let* $d_{\overleftarrow{Y}}$ *be a metric that generates the product topology.*

(i) *We say that* $d_{\overleftarrow{Y}}$ *is a* **uniform-product metric** *if for any* $\epsilon > 0$ *there exists a* $\delta > 0$ *such that* $d_{\overleftarrow{Y}}(\mathbf{y}, \mathbf{z}) < \epsilon$ *whenever* $d_Y(y_{-i}, z_{-i}) < \delta$*, for all* $i \geq 1$ *and for all* $\mathbf{y}, \mathbf{z} \in \overleftarrow{Y}$*.*

(ii) *We say that* $d_{\overleftarrow{Y}}$ *is a* **uniform-factor metric** *if for any* $\beta > 0$ *there exists an* $\alpha > 0$ *such that* $d_Y(y_{-i}, z_{-i}) < \beta$ *whenever* $d_{\overleftarrow{Y}}(\mathbf{y}, \mathbf{z}) < \alpha$*, for all* $i \leq -1$ *and for all* $\mathbf{y}, \mathbf{z} \in \overleftarrow{Y}$*.*

It can be readily verified that any metric of the form (3.1) is both a uniform-factor and a uniform-product metric.

COROLLARY 4.13. *Let* $g : U \times X \to X$ *be a driven system with Polish input and output spaces. Suppose that* $d_{\overleftarrow{U}}$ *is a uniform-factor metric, that* $d_{\overleftarrow{X}}$ *is simultaneously a uniform-product and a uniform-product metric, and that* g *is uniformly continuous on the first entry. Then the extension* $\mathbf{G}$ *of* g *to sequence space defined in* (3.2) *is also uniformly continuous on the first entry. Additionally, if* g *is uniformly continuous, then* $\mathbf{G}$ *is also uniformly continuous.*

Proof. We first show that $\mathbf{G}$ is uniformly continuous when g is uniformly continuous. Then the proof of the uniform continuity of $\mathbf{G}$ on the first entry follows from the uniform continuity of g on the first entry easily. We proceed in three steps:

Step 1. Fix $\epsilon > 0$. Since $d_{\overleftarrow{X}}$ is a uniform-product metric, we can find a $\tau > 0$ such that if $d_X(\mathbf{G}(\mathbf{u}, \mathbf{x})_i, \mathbf{G}(\mathbf{v}, \mathbf{y})_i) < \tau$ for all $i \leq -1$, then $d_{\overleftarrow{X}}(\mathbf{G}(\mathbf{u}, \mathbf{x}), \mathbf{G}(\mathbf{v}, \mathbf{y})) < \epsilon$. Fix any such $\tau > 0$.

Step 2. When g is uniformly continuous, we can find a $\gamma > 0$ independent of u_i, v_i, x_i, y_i and independent of $i \leq -1$, so that $d_X(g(u_i, x_i), g(v_i, y_i)) < \tau$ whenever $d_U(u_i, v_i) < \gamma$ and $d_X(x_i, y_i) < \gamma$.

Step 3. Since $d_{\overleftarrow{U}}$ and $d_{\overleftarrow{X}}$ are uniform-factor metrics, given γ there exits a $\delta > 0$ so that if $d_{\overleftarrow{U}}(\mathbf{u}, \mathbf{v}) < \delta$ and $d_{\overleftarrow{X}}(\mathbf{x}, \mathbf{y}) < \delta$ then $d_U(u_i, v_i) < \gamma$ and $d_X(x_i, y_i) < \gamma$, for all $i \leq -1$. This implies in particular that $d_X(x_i, y_i) < \gamma$, for all $i \leq -1$, whenever $d_{\overleftarrow{U} \times \overleftarrow{X}}((\mathbf{u}, \mathbf{x}), (\mathbf{v}, \mathbf{y})) < \delta$.

Hence, we have from the above three steps that if $d_{\overleftarrow{U} \times \overleftarrow{X}}((\mathbf{u}, \mathbf{x}), (\mathbf{v}, \mathbf{y})) < \delta$, we then necessarily have that $d_{\overleftarrow{X}}(\mathbf{G}(\mathbf{u}, \mathbf{x}), \mathbf{G}(\mathbf{v}, \mathbf{y}) < \epsilon$.

In particular, when g is only uniformly continuous on the first entry, we can set $\mathbf{x} = \mathbf{y}$ in the steps above to obtain the implication $d_{\overleftarrow{U}}(\mathbf{u}, \mathbf{v}) < \delta \Longrightarrow d_{\overleftarrow{X}}(\mathbf{G}(\mathbf{u}, \mathbf{x}), \mathbf{G}(\mathbf{v}, \mathbf{x}) < \epsilon$. ■

The implications about the continuity of the different maps that we have proved in this subsection is summarized in Figure 4.

4.3. The fixed points of the Foias operator.

We do not know if the Foias operator P_g of the driven system is jointly continuous, that is, continuous on $P(\overleftarrow{U}) \times P_1(\overleftarrow{X})$ when it is equipped with the product topology. However, the importance of the contractivity and continuity results in the previous two subsections lies in the fact that they can be used in conjunction to prove the continuity of the dependence of the fixed points of P_g on the input process. The following theorem provides a specific statement in this direction for a driven system

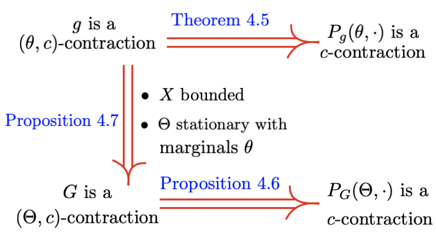

FIG. 3.
Implications between contractivity in different spaces.

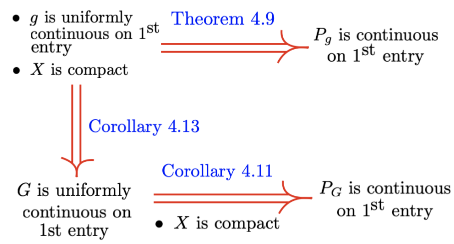

FIG. 4.
Implications of continuity between the different maps.

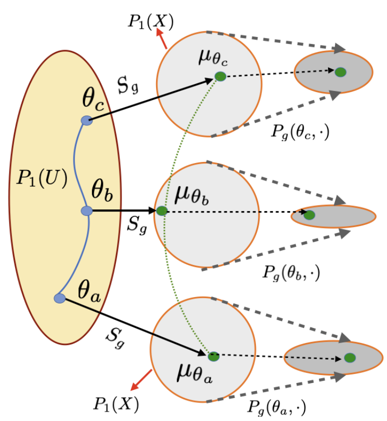

FIG. 5. *Graphical illustration of the map S_g, its continuity, and of its relation with the fixed points of the Foias operator.*

g and its Foias operator P_g. We generalize later on this result in Theorem 4.15 to the driven system $\mathbf{G}$ in sequence space its Foias operator $P_{\mathbf{G}}$.

THEOREM 4.14 (**Fixed points of P_g**). *Consider a measurable driven system $g : U \times X \to X$ with Polish input and output spaces and (X, d_X) compact. Assume that for each $\theta \in P(U)$ the map g is a (θ, c_θ)-stochastic contraction, with c_θ as in (4.2), and that one of the following hypotheses holds true:*

(i) *There exists a constant $c_0 \in (0,1)$ such that $0 < c_\theta < c_0 < 1$ for all $\theta \in P_1(U)$ and g is uniformly continuous on the first entry.*

(ii) *g is uniformly continuous.*

Then, for any $\theta \in P_1(U)$ there exists a unique $\mu_\theta \in P_1(X)$ which is a fixed point of the map $P_g(\theta, \cdot) : P_1(X) \to P_1(X)$ and, moreover, the map $S_g : P_1(U) \to P_1(X)$ that assigns $\theta \mapsto \mu_\theta$, is continuous.

Proof. Notice first that part **(i)** of Proposition 2.3 implies, together with the compactness of X that the Foias map $P_g : P(U) \times P_1(X) \to P_1(X)$ is well-defined. Fix now $\theta \in P(U)$ and recall that by Theorem 4.5, the (θ, c_θ)-stochastic contractivity hypothesis on g implies that $P_g(\theta, \cdot) : P_1(X) \to P_1(X)$ is a c_θ-contraction with respect to the Wasserstein distance. Since the completeness of (X, d_X) implies that of $(P_1(X), W)$ by [41, Theorem 6.18], Banach's Fixed Point Theorem implies the existence of a unique $\mu_\theta \in P_1(X)$ such that $P_g(\theta, \mu_\theta) = \mu_\theta$ for each $\theta \in P(U)$, as well as the existence of the map $S_g : P(U) \to P_1(X)$ that assigns $\theta \mapsto \mu_\theta$.

The remainder of the proof is dedicated to showing that the restriction $S_g : P_1(U) \to P_1(X)$ is continuous in the presence of the hypotheses in **(i)** or **(ii)**.

Assume first that **(i)** holds and suppose that we have a sequence $\{\theta_n\}_{n\in\mathbb{N}} \subset P_1(U)$ such that $\lim_{n\to\infty} \theta_n = \theta \in P_1(U)$. It suffices to show that $\lim_{n\to\infty} \mu_{\theta_n} = \mu_\theta$. Let c_{θ_n} be the the best contraction constant associated to θ_n, for any $n \in \mathbb{N}$, as defined in (4.2), that is,

$$c_{\theta_n} = \inf \left\{ c \in (0,1) \middle| \int_U d_X(g_u(x), g_u(y)) d\theta_n(u) \leq c\, d_X(x,y),\ \forall x, y \in X \right\}. \tag{4.16}$$

By Theorem 4.5, we have

$$W(P_g(\theta_n, \mu_\theta), P_g(\theta_n, \mu_{\theta_n})) \leq c_{\theta_n} W(\mu_\theta, \mu_{\theta_n}). \tag{4.17}$$

We hence have that

$$\begin{aligned} W(\mu_\theta, \mu_{\theta_n}) &\leq W(\mu_\theta, P_g(\theta_n, \mu_\theta)) + W(P_g(\theta_n, \mu_\theta), \mu_{\theta_n}) && \text{(by the triangle inequality)} \\ &= W(\mu_\theta, P_g(\theta_n, \mu_\theta)) + W(P_g(\theta_n, \mu_\theta), P_g(\theta_n, \mu_{\theta_n})) && (\text{since } \mu_{\theta_n} = P_g(\theta_n, \mu_{\theta_n})) \end{aligned}$$

Hence, by (4.17),

$$W(\mu_\theta, \mu_{\theta_n}) \leq \frac{1}{1 - c_{\theta_n}} W(\mu_\theta, P_g(\theta_n, \mu_\theta)). \tag{4.18}$$

Notice now that the hypotheses in point **(i)** and Theorem 4.9 guarantee that the maps $P_g(\cdot, \mu) : P_1(U) \to P_1(X)$ are all continuous, for any $\mu \in P_1(U)$, and hence as $n \to \infty$, $W(\mu_\theta, P_g(\theta_n, \mu_\theta)) \to W(\mu_\theta, P_g(\theta, \mu_\theta)) = W(\mu_\theta, \mu_\theta) = 0$. In addition, since $\limsup_{n\to\infty} c_{\theta_n} \leq c_0 < 1$, then $\frac{1}{1-c_n}$ is bounded above, and hence from (4.18), $\lim_{n\to\infty} W(\mu_\theta, \mu_{\theta_n}) = 0$, as required.

We next consider the case **(ii)**. If $\limsup_{n\to\infty} c_{\theta_n} = c_0 < 1$, then it reduces to the result in part **(i)**. Suppose hence that $\limsup_{n\to\infty} c_{\theta_n} = 1$. For a fixed $x, y \in X$, the map $h_{x,y} : U \to \mathbb{R}$ given by $h_{x,y}(u) = d_X(g_u(x), g_u(y))$ is bounded by some constant $M > 0$ since g is uniformly continuous and X is compact. Thus $|h_{x,y}(u)| < M \leq M(1 + d_U(u, u_0))$. Given that the sequence $\{\theta_n\}_{n\in\mathbb{N}} \subset P_1(U)$ is such that $\lim_{n\to\infty} \theta_n = \theta \in P_1(U)$, the characterization of the weak convergence (2.5) implies that

$$\lim_{n\to\infty} \int_U d_X(g_u(x), g_u(y))\, d\theta_n(u) = \int_U d_X(g_u(x), g_u(y))\, d\theta(u) \quad \text{for all} \quad x, y \in X. \tag{4.19}$$

Consider now the sequence of functions $f_n : X \times X \to \mathbb{R}$, $n \in \mathbb{N}$, defined by $f_n(x,y) := \int_U d_X(g_u(x), g_u(y))\, d\theta_n(u)$. We shall now show that $\{f_n\}_{n\in\mathbb{N}}$ is equicontinuous when in $X \times X$ we consider the product metric

$$d_{X\times X}((x,y),(a,b)) = \max \{d_X(x,a), d_X(y,b)\}.$$

Notice first that for any $(x,y), (a,b) \in X \times X$:

$$\begin{aligned} |f_n(a,b) - f_n(x,y)| &= \left| \int_U (d_X(g_u(a), g_u(b)) - d_X(g_u(x), g_u(y))\, d\theta_n(u) \right| \\ &\leq \int_U \left| \, (d_X(g_u(a), g_u(b)) - d_X(g_u(x), g_u(y)) \, \right| \, d\theta_n(u). \end{aligned} \tag{4.20}$$

The compactness of X implies that $d_X : X \times X \to \mathbb{R}$ is uniformly continuous, and hence for any $\epsilon > 0$ there exists $\delta_{d_X}(\epsilon) > 0$ such that for any $(x, y), (a, b) \in X \times X$ such that $d_{X\times X}((x, y), (a, b)) < \delta_{d_X}(\epsilon)$ we have that $|d_X(x, y) - d_X(a, b)| < \epsilon$. Since the uniform continuity of g implies its uniform continuity on the second entry, for any $\epsilon > 0$ there exists $\delta_g(\epsilon) > 0$ such that $d_X(g_u(x), g_u(y)) < \epsilon$ for any $u \in U$ and any $x, y \in X$ such that $d_X(x, y) < \delta_g(\epsilon)$.

By the definition of the product metric $d_{X\times X}$, if $d_{X\times X}((x, y), (a, b)) < \delta_g(\delta_{d_X}(\epsilon))$ then $d_X(x, a) < \delta_g(\delta_{d_X}(\epsilon))$ and $d_X(y, b) < \delta_g(\delta_{d_X}(\epsilon))$ and hence $d_X(g_u(x), g_u(a)) < \delta_{d_X}(\epsilon)$ and $d_X(g_u(y), g_u(b)) < \delta_{d_X}(\epsilon)$, for any $u \in U$. This implies that

$$d_{X\times X}((g_u(x), g_u(y)), (g_u(a), g_u(b))) < \delta_{d_X}(\epsilon)$$

and hence that

$$|d_X(g_u(a), g_u(b)) - d_X(g_u(x), g_u(y))| < \epsilon \quad \text{for all} \quad u \in U.$$

Since θ_n is a probability measure, from (4.20), we have that $|f_n(a, b) - f_n(x, y)| < \epsilon$, which proves the equicontinuity of $\{f_n\}$.

Let now $\{c_{\theta_{n_j}}\}$ be a subsequence of $\{c_{\theta_n}\}$ such that $\lim_{j\to\infty} c_{\theta_{n_j}} = 1$. By (4.16) we can find a sequence $\{(x_{n_j}, y_{n_j})\}$ in $X \times X$ such that $f_{n_j}(x_{n_j}, y_{n_j}) = c_{\theta_{n_j}} d_X(x_{n_j}, y_{n_j})$ for all $j \in \mathbb{N}$. Without loss of generality we assume $\{(x_{n_j}, y_{n_j})\}$ converges to some point $(x_0, y_0) \in X \times X$.

Next, we note that $\{f_{n_j}\}$ converges point-wise to

$$f(x, y) = \int_U d_X(g_u(x), g_u(y))\, d\theta(u)$$

by (4.19) and by the Arzela-Ascoli theorem we have a convergent subsequence of $\{f_{n_j}\}$. Hence, without loss of generality assume that $\{f_{n_j}\}$ itself converges uniformly to f. Since f is continuous:

$$f(x_0, y_0) = \lim_{j\to\infty} f_{n_j}(x_{n_j}, y_{n_j}) = \lim_{j\to\infty} c_{\theta_{n_j}} d_X(x_{n_j}, y_{n_j}) = d_X(x_0, y_0)$$

which contradicts that $\int_U d(g_u(x_0), g_u(y_0))\, d\theta(u) < c_\theta d_X(x_0, y_0)$ for some $0 < c_\theta < 1$. Hence, $\limsup_{j\to\infty} c_{\theta_{n_j}} < 1$ necessarily and the theorem is proven. ∎

The goal of our last theorem is showing that the conclusions about the existence of fixed points of the maps $P_g(\theta, \cdot)$ and their continuous dependence on θ that we proved in Theorem 4.14 can be extended to $P_{\mathbf{G}}(\Theta, \cdot)$ by using hypotheses that are exclusively formulated in terms of g, provided that the inputs Θ are stationary. We hence denote as

$$P_S(\overleftarrow{U}) = \left\{ \Theta \in P_1(\overleftarrow{U}) \mid T_{1*}\Theta = \Theta \right\}$$

the set of stationary input processes in $P_1(\overleftarrow{U})$. In the second part of the statement, we shall prove that in the presence of the unique solution property (and hence when (3.3) has unique fixed points), the fixed points of $P_{\mathbf{G}}(\Theta, \cdot)$ can be characterized using the push-forward of the filter U_g that is available in that case.

Theorem 4.15 (**Fixed points of $P_{\mathbf{G}}$**). *Let $g : U \times X \to X$ be a measurable driven system with Polish input and output spaces and (X, d_X) compact. Let* $\mathbf{G}$:

$\overleftarrow{U} \times \overleftarrow{X} \to \overleftarrow{X}$ *be the induced driven system in sequence space defined in* (3.2)*. Assume now that for any element* $\Theta \in P_S(\overleftarrow{U})$ *with marginal time-independent laws* $\theta \in P_1(U)$*, the map* g *is a* (θ, c_θ)*-stochastic contraction and that one of the hypotheses* **(i)** *or* **(ii)** *in Theorem 4.14 are satisfied for* g*. Then,*

(i) *For any* $\Theta \in P_S(\overleftarrow{U})$ *there exists a unique* $M_\Theta \in P_S(\overleftarrow{X})$ *which is a fixed point of the map* $P_{\mathbf{G}}(\Theta, \cdot) : P_1(\overleftarrow{X}) \to P_1(\overleftarrow{X})$ *and, moreover, the map* $S_{\mathbf{G}} : P_S(\overleftarrow{U}) \to P_S(\overleftarrow{X})$ *that assigns* $\Theta \mapsto M_\Theta$*, is continuous when the domain and the image are endowed with the Wasserstein-1 distance.*

(ii) *When* g *has the unique solution property and a unique measurable, causal, and time-invariant filter* $U_g : \overleftarrow{U} \longrightarrow \overleftarrow{X}$ *can be associated to it using* (2.2)*, we have that*

$$S_{\mathbf{G}}(\Theta) = U_{g\,*}\Theta, \quad \text{for all } \Theta \in P_S(\overleftarrow{U}). \tag{4.21}$$

Proof. **(i)** This part is proved by mimicking the proof of Theorem 4.14, where the driven system g is replaced by $\mathbf{G}$ and the Foias map P_g by $P_{\mathbf{G}}$. In order to achieve that, we have first to show that our hypotheses on g about contractivity and uniform continuity in Theorem 4.14 translate into analog conditions for $\mathbf{G}$. We recall, first of all, that since U is Polish and X is Polish and compact, then so are $\overleftarrow{U}$ and $\overleftarrow{X}$ with the product topology induced by any of the metrics $d_{\overleftarrow{U}}$ and $d_{\overleftarrow{X}}$ introduced in (3.1). This fact allows us in particular to define the Wasserstein-1 metrics on $P_S(\overleftarrow{U})$ and $P_1(\overleftarrow{X})$. The hypothesis on the (θ, c_θ)-stochastic contractivity of g and the stationarity of Θ imply by Propositions 4.7 and 4.6 that $\mathbf{G}$ is a (Θ, c_θ)-stochastic contraction and that $P_{\mathbf{G}}(\Theta, \cdot)$ is a c_θ-contraction. Additionally, recall that by Corollary 4.13, if g is uniformly continuous or continuous on the first entry (as in the hypotheses in Theorem 4.14) then so is $\mathbf{G}$. Given all these facts, the proof of Theorem 4.14 can be reproduced in our setup for $\mathbf{G}$ and $P_{\mathbf{G}}$ in order to obtain all the claims in part **(i)** except for the time-stationarity of $S_{\mathbf{G}}(\Theta)$ that we postpone to the end of the proof.

(ii) Using the uniqueness property of the map $S_{\mathbf{G}}$ that was established in part **(i)**, it suffices to verify that

$$P_{\mathbf{G}}(\Theta, U_{g\,*}\Theta) = U_{g\,*}\Theta, \quad \text{for any } \Theta \in P_S(\overleftarrow{U}), \tag{4.22}$$

in order to prove the equality (4.21). We first recall that, as we pointed out in (3.3), the filter $U_g : \overleftarrow{U} \longrightarrow \overleftarrow{X}$ is the unique solution of the relation

$$T_1 \circ \mathbf{G}(\mathbf{u}, U_g(\mathbf{u})) = U_g(\mathbf{u}), \quad \text{for all } \mathbf{u} \in \overleftarrow{U}. \tag{4.23}$$

By the definition of $\mathbf{G}$ it is easy to see that

$$T_1 \circ \mathbf{G} = \mathbf{G} \circ T_1, \tag{4.24}$$

which, together with the uniqueness property in (4.23) implies that U_g is necessarily T_1-equivariant, that is,

$$T_1 \circ U_g = U_g \circ T_1. \tag{4.25}$$

These relations imply that (4.23) can be rewritten as

$$\mathbf{G} \circ (T_1 \times U_g \circ T_1) = U_g. \tag{4.26}$$

This expression implies that for any time-invariant $\Theta \in P_S(\overleftarrow{U})$ (which hence satisfies ${T_1}_*\Theta = \Theta$), we have that

$$\mathbf{G}_* \left(\Theta, {U_g}_*\Theta\right) = {U_g}_*\Theta. \tag{4.27}$$

We now observe that the relation (2.8) that was proved in Remark 2.2 for g and P_g can be extended to $\mathbf{G}$ and $P_{\mathbf{G}}$. More specifically, if we consider in the space $\overleftarrow{U} \times \overleftarrow{X}$ the product measure determined by the laws of Θ and ${U_g}_*\Theta$, then

$$P_{\mathbf{G}} \left(\Theta, {U_g}_*\Theta\right) = \mathbf{G}_* \left(\Theta, {U_g}_*\Theta\right) = {U_g}_*\Theta,$$

which proves (4.22), as required.

We conclude the proof by showing the last statement in part **(i)** that was left to be proved, namely, the stationarity of $S_{\mathbf{G}}(\Theta)$. We see now that this is a consequence of the uniqueness of $S_{\mathbf{G}}(\Theta)$ as a fixed point of $P_{\mathbf{G}}(\Theta, \cdot)$ and of the T_1-equivariance of $\mathbf{G}$ in (4.24). Indeed, on the one hand $S_{\mathbf{G}}(\Theta)$ satisfies that

$$P_{\mathbf{G}}(\Theta, S_{\mathbf{G}}(\Theta)) = S_{\mathbf{G}}(\Theta). \tag{4.28}$$

If we now apply ${T_1}_*$ on both sides of (4.28), use again the relation (2.8) for $\mathbf{G}$ and $P_{\mathbf{G}}$, and the equivariance (4.24), we have that

$$\begin{aligned} {T_1}_* S_{\mathbf{G}}(\Theta) &= {T_1}_* P_{\mathbf{G}}(\Theta, S_{\mathbf{G}}(\Theta)) = {T_1}_* \mathbf{G}_*(\Theta, S_{\mathbf{G}}(\Theta)) \\ &= \mathbf{G}_*({T_1}_*\Theta, {T_1}_* S_{\mathbf{G}}(\Theta)) = \mathbf{G}_*(\Theta, {T_1}_* S_{\mathbf{G}}(\Theta)) = P_{\mathbf{G}}(\Theta, {T_1}_* S_{\mathbf{G}}(\Theta)). \end{aligned} \tag{4.29}$$

This equality shows that ${T_1}_* S_{\mathbf{G}}(\Theta)$ is also a fixed point of $P_{\mathbf{G}}(\Theta, \cdot)$, but since that point is unique we necessarily have that ${T_1}_* S_{\mathbf{G}}(\Theta) = S_{\mathbf{G}}(\Theta)$ and hence $S_{\mathbf{G}}(\Theta) \in P_S(\overleftarrow{X})$, as required. ∎

We refer the reader to the examples of convergence with respect to the Wasserstein distances in Example 4.16 next.

Example 4.16 (**The unique solution of the VARMA and GARCH processes).** In the examples 4.2 and 4.3 above we saw that the conditions $\mathrm{E}\left[\|A(u)\|\right] < 1$ and $\alpha + \beta < 1$ guarantee the stochastic contractivity of the VARMA model with time-dependent coefficients and of the GARCH(1,1) model, respectively. We saw that these conditions are vastly less restrictive than enforcing the standard contractivity of the state map that defines these models. Using now Theorem 4.15 we can conclude that both models have a unique stationary solution that corresponds to the fixed points of their respective associated Foias operators. In the case of VARMA, the solution process is

$$X_t = f(u_{t-1}) + \sum_{k=1}^{\infty} A(u_{t-1})A(u_{t-2}) \cdots A(u_{t-k}) f(u_{t-k-1}), \quad t \in \mathbb{Z}^-,$$

and for GARCH(1,1) it can be written as $r_t = \sqrt{h_t} u_{t-1}$, where

$$h_t = \left\{ 1 + \sum_{i=1}^{\infty} a(u_{t-2}) \cdots a(u_{t-i-1}) \right\} \omega, \quad a(u) := \alpha u^2 + \beta, \quad t \in \mathbb{Z}^-.$$

When using the standard approach in time series analysis it is proved that these series convege almost surely (see [7], [5, Theorem 1.1], or [12]). Theorem 4.15 shows that this convergence takes also place with respect to the Wasserstein distance.

5. Conclusions. In this paper we have provided conditions that guarantee the **existence** and **uniqueness** of **asymptotically invariant measures** for driven systems and we have proved that their dependence on the input process is **continuous** when the set of input and output processes are endowed with the Wasserstein distance. These conditions ensure that the invariant measures are robust to changes in the input stochastic source.

These results have been obtained by proving the existence and uniqueness of fixed points of the associated **Foias operators**, which have been profusely studied in the paper in both the state and sequence spaces. This has been achieved by using Banach's Fixed Point Theorem in the context of Foias operators by imposing readily verifiable contractivity and continuity hypotheses that are exclusively formulated for the driven system g defined in the state space. The most important condition is a newly introduced notion of **stochastic state contractivity** for the driven system g, ensures that the Foias operators in state and in sequence spaces are also contractive with respect to the Wasserstein distance. Stochastic state contractivity is less restrictive than the standard state contractivity condition evoked to ensure the USP. In a future work we hope to answer more in depth the intriguing question as to how the echo state property with respect to all typical trajectories is related to the stochastic contraction property that was profusely used in this paper.

In a forthcoming paper we shall study the embedding properties of the mappings S_g and $S_{\mathbf{G}}$ and, in particular, their potential injectivity properties so, that two inputs with different distributions are discriminated in the response. Progress in that direction would make the techniques introduced in this paper very valuable in the construction of dynamical classification tasks.

Acknowledgments: The authors are indebted to two anonymous referees for carefully reviewing the paper and for their suggestions, which have made the presentation more engaging.

REFERENCES

[1] R. Aimino, H. Hu, M. Nicol, and S. Vaienti, *Polynomial loss of memory for maps of the interval with a neutral fixed point*, Discrete and Continuous Dynamical Systems, 35 (2014), pp. 793–806.

[2] H. Aref, *The development of chaotic advection*, Physics of Fluids, 14 (2002), pp. 1315–1325.

[3] J. Atnip, G. Froyland, C. González-Tokman, and S. Vaienti, *Equilibrium states for non-transitive random open and closed dynamical systems*, Ergodic Theory and Dynamical Systems, (2022), pp. 1–23.

[4] T. Bollerslev, *Generalized autoregressive conditional heteroskedasticity*, Journal of Econometrics, 31 (1986), pp. 307–327.

[5] P. Bougerol and N. Picard, *Strict Stationarity of Generalized Autoregressive Processes*, The Annals of Probability, (1992).

[6] S. Boyd and L. Chua, *Fading memory and the problem of approximating nonlinear operators with Volterra series*, IEEE Transactions on Circuits and Systems, 32 (1985), pp. 1150–1161.

[7] A. Brandt, *The stochastic equation Yn +1=AnYn + Bn with stationary coefficients*, Advances in Applied Probability, 18 (1986), pp. 211–220.

[8] C. Cuchiero, L. Gonon, L. Grigoryeva, J.-P. Ortega, and J. Teichmann, *Discrete-time signatures and randomness in reservoir computing*, IEEE Transactions on Neural Networks and Learning Systems, 10.1109/TN (2021), pp. 1–10.

[9] R. M. Dudley, *Real analysis and probability*, CRC Press, 2018.

[10] M. Duflo, *Random Iterative Models*, Springer-Verlag Berlin Heidelberg, 1997, https://doi.org/10.2307/2669576.

[11] R. F. Engle, *Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation*, Econometrica, 50 (1982), pp. 987–1007.

[12] C. Francq and J.-M. Zakoian, *GARCH Models: Structure, Statistical Inference and Financial Applications*, Wiley, 2010.

[13] L. Gonon, L. Grigoryeva, and J.-P. Ortega, *Risk bounds for reservoir computing*, Journal of Machine Learning Research, 21 (2020), pp. 1–61.
[14] L. Gonon, L. Grigoryeva, and J.-P. Ortega, *Approximation error estimates for random neural networks and reservoir systems*, To appear in the Annals of Applied Probability, (2022).
[15] L. Gonon and J.-P. Ortega, *Reservoir computing universality with stochastic inputs*, IEEE Transactions on Neural Networks and Learning Systems, 31 (2020), pp. 100–112, https://arxiv.org/abs/1807.02621.
[16] L. Gonon and J.-P. Ortega, *Fading memory echo state networks are universal*, Neural Networks, 138 (2021), pp. 10–13.
[17] L. Grigoryeva and J.-P. Ortega, *Echo state networks are universal*, Neural Networks, 108 (2018), pp. 495–508, https://arxiv.org/abs//arxiv.org/abs/1806.00797.
[18] L. Grigoryeva and J.-P. Ortega, *Universal discrete-time reservoir computers with stochastic inputs and linear readouts using non-homogeneous state-affine systems*, Journal of Machine Learning Research, 19 (2018), pp. 1–40, http://arxiv.org/abs/1712.00754, https://arxiv.org/abs/1712.00754.
[19] L. Grigoryeva and J.-P. Ortega, *Differentiable reservoir computing*, Journal of Machine Learning Research, 20 (2019), pp. 1–62.
[20] C. Gupta, W. Ott, and A. Török, *Memory loss for time-dependent piecewise expanding systems in higher dimension*, Mathematical Research Letters, 20 (2013), pp. 141–161.
[21] H. Jaeger, *The 'echo state' approach to analysing and training recurrent neural networks with an erratum note*, tech. report, German National Research Center for Information Technology, 2010.
[22] H. Jaeger and H. Haas, *Harnessing Nonlinearity: Predicting Chaotic Systems and Saving Energy in Wireless Communication*, Science, 304 (2004), pp. 78–80, https://doi.org/10.1126/science.1091277.
[23] L. V. Kantorovich and S. G. Rubinshtein, *On a space of totally additive functions*, Vestnik of the St. Petersburg University: Mathematics, 13 (1958), pp. 52–59.
[24] P. E. Kloeden and M. Rasmussen, *Nonautonomous Dynamical Systems*, American Mathematical Society, 2010, https://doi.org/10.3934/dcdsb.2015.20.703, http://www.aimsciences.org/journals/displayArticlesnew.jsp?paperID=10767.
[25] A. S. Kravchenko, *Completeness of the space of separable measures in the Kantorovich-Rubinshtein metric*, Siberian Mathematical Journal, 47 (2006), pp. 68–76.
[26] A. Lasota and M. C. Mackey, *Chaos, Fractals, and Noise: Stochastic Aspects of Dynamics*, vol. 97, Springer Science & Business Media, 1998.
[27] C. Liverani, *Decay of correlations*, Annals of Mathematics, 142 (1995), pp. 239–301.
[28] M. Lukoševičius and H. Jaeger, *Reservoir computing approaches to recurrent neural network training*, Computer Science Review, 3 (2009), pp. 127–149.
[29] W. Maass, *Liquid state machines: motivation, theory, and applications*, in Computability In Context: Computation and Logic in the Real World, S. S. Barry Cooper and A. Sorbi, eds., 2011, ch. 8, pp. 275–296.
[30] W. Maass, T. Natschläger, and H. Markram, *Real-time computing without stable states: a new framework for neural computation based on perturbations*, Neural Computation, 14 (2002), pp. 2531–2560.
[31] G. Manjunath, *Stability and memory-loss go hand-in-hand: three results in dynamics & computation*, To appear in Proceedings of the Royal Society London Ser. A Math. Phys. Eng. Sci., (2020), pp. 1–25, https://doi.org/10.1098/rspa.2020.0563, http://arxiv.org/abs/2001.00766, https://arxiv.org/abs/2001.00766.
[32] G. Manjunath, *Embedding information onto a dynamical system*, Nonlinearity, 35 (2022), p. 1131.
[33] G. Manjunath and H. Jaeger, *Echo state property linked to an input: exploring a fundamental characteristic of recurrent neural networks*, Neural Computation, 25 (2013), pp. 671–696, https://doi.org/10.1162/NECO_a_00411, https://arxiv.org/abs/1309.2848v1.
[34] A. Mohapatra and W. Ott, *Memory loss for nonequilibrium open dynamical systems*, Discrete Contin. Dyn. Syst, 34 (2014), pp. 3747–3759.
[35] J. Munkres, *Topology*, Pearson, second ed., 2014.
[36] W. Ott, L.-S. Young, and M. Stenlund, *Memory loss for time-dependent dynamical systems*, arXiv preprint arXiv:1210.0109, (2012).
[37] V. M. Panaretos and Y. Zemel, *An Invitation to Statistics in Wasserstein Space*, Springer Nature, 2020.
[38] J. H. Shapiro, *A Fixed-Point Farrago*, Springer International Publishing Switzerland, 2016.
[39] M. Stenlund, *Non-stationary compositions of anosov diffeomorphisms*, Nonlinearity, 24

(2011), p. 2991.

[40] G. Tanaka, T. Yamane, J. B. Héroux, R. Nakane, N. Kanazawa, S. Takeda, H. Numata, D. Nakano, and A. Hirose, *Recent advances in physical reservoir computing: A review*, Neural Networks, 115 (2019), pp. 100–123, https://doi.org/10.1016/j.neunet.2019.03.005, https://doi.org/10.1016/j.neunet.2019.03.005.

[41] C. Villani, *Optimal Transport: Old and New*, Springer, 2009.

[42] S. Wiggins, *Chaotic transport in dynamical systems*, vol. 2, Springer Science & Business Media, 2013.

[43] L.-S. Young, *Statistical properties of dynamical systems with some hyperbolicity*, Annals of Mathematics, (1998), pp. 585–650.