Non-essentiality of Reservoir's Fading Memory for Universality of Reservoir Computing

Shuhei Sugiura, Ryo Ariizumi, *Member, IEEE*, Toru Asai, *Member, IEEE*, and Shun-ichi Azuma, *Senior Member, IEEE*

***Abstract*—This paper describes a novel sufficient condition concerning approximations with reservoir computing. Recently, reservoir computing using a physical system as the reservoir has attracted attention. Because many physical systems are modeled as state-space systems, it is necessary to guarantee the approximations given by reservoirs represented as nonlinear state-space systems. There are two problems with existing approaches: a reservoir must have a property called fading memory and must be represented as a set of maps between input and output signals on the bi-infinite-time interval. These two conditions are too strict for reservoirs represented as nonlinear state-space systems as they require the reservoir to have a unique equilibrium state for the zero input. This paper proposes an approach that employs operators from right-infinite-time inputs to right-infinite-time outputs. Furthermore, we develop a novel extension of the Stone–Weierstrass theorem to handle discontinuous functions. To apply the extended theorem, we define functionals corresponding to operators and introduce a metric on the domain of the functionals. The resulting sufficient condition does not require the reservoir to have fading memory or continuity with respect to inputs and time. Therefore, our result guarantees the approximations with very common reservoirs and provides a rationale for physical reservoir computing. We present an example of a physical reservoir without fading memory. With the example reservoir, the reservoir computing model successfully approximates NARMA10, a benchmark task for time series predictions.**

***Index Terms*—Machine learning, neural network, nonlinear dynamical system, reservoir computing.**

I. Introduction

RESERVOIR computing (RC) is a machine learning method for dynamical system approximation and time-series analysis. The concept of RC is derived from recurrent neural networks (RNNs) [1], [2] and is based on the idea of reducing computational costs by updating only certain parameters during the training process [3]–[7]. An RC model consists of a dynamical system called a reservoir and a static function called a readout. The output of an RC model is obtained by mapping the reservoir output by the readout. If the reservoir is sufficiently complex, the RC model can approximate a system by adjusting only the readout parameters [4].

In general, a randomly generated RNN is used as the reservoir. However, because the reservoir is fixed during the training process, a physical system that cannot be adjusted may also be used as the reservoir. Recently, physical RC, which uses a physical system as the reservoir, has received considerable attention [8]–[12]. This is expected to provide a superior physical reservoir to a RNN implemented on a general-purpose computer in terms of processing speed and energy consumption [8], [13]. Because many physical systems are modeled as state-space systems, it is necessary to guarantee the approximations given by reservoirs represented as nonlinear state-space systems.

The universality of RC models lies in their ability to approximate an arbitrary system with arbitrary accuracy. Grigoryeva et al. [14] guaranteed the universality of an echo state network (ESN) [3], which is an RC model composed of a RNN reservoir and a linear readout. Gonon et al. [15] guaranteed the universality of the three classes of RC models for stochastic inputs (one composed of a linear reservoir and a polynomial readout, one composed of a state-affine reservoir and a linear readout, and an ESN). These results relate to specific RC models and are not applicable to RC models with general nonlinear reservoirs. Maass et al. [4] dealt with a more general reservoir represented by a set of operators between inputs and outputs defined on the bi-infinite-time (BIT) interval $\mathbb{R}$. We call these BIT operators. They showed that two properties, namely fading memory and the separation property, are sufficient reservoir conditions for the universality of an RC model with a polynomial readout.

Fading memory is a property of a dynamical system. Although its definition varies across different studies [4], [16]–[18], it can basically be stated as follows: if two inputs are close enough to each other for a sufficiently long time, the outputs for these inputs become arbitrarily close to each other. In other words, fading memory is the special continuity of the output with respect to "recent" inputs. We call a system or operator approximated by RC a "target." Fading memory is important for uniformly approximating the target over an infinite-time interval. In many cases, fading memory is required not only for the reservoir, but also for the target [4], [14]. This is because the uniform approximation of discontinuous functions is very difficult. In contrast, Gonon et al. [15] did not require fading memory in either the target or the reservoir because they evaluated the approximation error in terms of the p-norm ($p < \infty$) instead of the uniform norm.

When applying the result of Maass et al. [4] to reservoirs represented as nonlinear state-space systems, two problems

This work was supported by JSPS KAKENHI under Grant No. JP22K04027 and by "Innovation inspired by Nature" Research Support Program, SEKISUI CHEMICAL CO., LTD.

Manuscript received April 19, 2021; revised August 16, 2021.

are encountered. The first is that fading memory is too strict for a general nonlinear state-space system because, in many cases, it is violated by multiple equilibrium states for the zero input. For example, suppose that two inputs converge to 0 and the state trajectories for these inputs converge to two different equilibrium states. The difference between the two inputs will then converge to 0, but the difference between the outputs may not, which violates fading memory.

The second problem is that a general nonlinear state-space system cannot be represented as BIT operators. To represent a state-space system as well-defined BIT operators, it is important to assume that the system determines a unique bounded state trajectory for a BIT input. A state-space system satisfying this assumption is said to have the echo state property (ESP) [3]. In the discrete-time case, fading memory is derived from ESP [3]. Therefore, ESP causes the same problem as fading memory, i.e., the requirement of a unique equilibrium. This is also confirmed in the continuous-time case as follows: if the system has multiple equilibria, state trajectories remaining at each equilibrium exist for the zero input, which violates the ESP.

This paper proposes a novel universality result of RC with a polynomial readout that avoids these two problems. Instead of BIT operators, we use right-infinite-time (RIT) operators to express a reservoir. An RIT operator is a map from RIT inputs to RIT outputs and can be defined by the input–output relationship of a state-space system with an initial state. Therefore, unlike BIT operators, the uniqueness of outputs is automatically satisfied, independent of ESP. We guarantee universality by converting the RIT operators into functionals and applying the Stone–Weierstrass theorem. Unlike BIT operators, this conversion does not require the time invariance of RIT operators. To apply the Stone–Weierstrass theorem, we define a metric on the functional domain, which is a set of inputs defined on time intervals of various lengths.

Representation by the RIT operators does not change the fact that fading memory is required for a reservoir. To avoid this problem, we extend the Stone–Weierstrass theorem to handle discontinuous functionals, because functional continuity is deeply related to the fading memory of the operators. Our main theorem guaranteeing the universality of RC is a corollary of the extended theorem. As the condition of our main theorem, we propose a novel property of reservoirs named the neighborhood separation property. The neighborhood separation property does not require the operators in a reservoir to have the fading memory property.

For RC with a polynomial readout, we guarantee universality for the uniform approximation of an operator with fading memory. Our contribution is to deal with the following cases:

(i) the reservoir is represented by a set of RIT operators;

(ii) operators in the reservoir do not have fading memory.

These contributions introduce many advantages to the use of a reservoir represented by a state-space system. First, (i) allows the reservoir to vary with time. Second, the combination of (i) and (ii) allows the reservoir to have multiple equilibrium states for the zero input. Furthermore, the novel reservoir condition, i.e., the neighborhood separation property, which we propose to achieve (ii), does not even require the reservoir output to be continuous with respect to time or the input. This allows discontinuous output functions. Therefore, our result can be applied to very common reservoirs and is particularly significant in the field of physical RC, in which reservoirs are not easily adjustable.

The structure of this paper is as follows. Section II discusses the previous work of Maass et al [4], RC in the form of RIT operators, the conversion of RIT operators into functionals, and the metric on the domain of functionals. Section III introduces the extension of the Stone–Weierstrass theorem, presents a novel sufficient reservoir condition, and provides an example of a physical reservoir without fading memory. Section IV concludes the paper.

Notation: Let $\mathbb{R}_+ = [0, \infty)$ and $\mathbb{R}_- = (-\infty, 0]$. For any $i \in \mathbb{N}$, we define a norm $\|\cdot\|$ on $\mathbb{R}^i$ as the Euclidean norm. For any functions f_1, f_2 from a domain E to $\mathbb{R}$, we define the sum $f_1 + f_2$ and the product $f_1 f_2$ of f_1 and f_2 as

$$\begin{aligned} (f_1 + f_2)(x) &= f_1(x) + f_2(x), \\ (f_1 f_2)(x) &= f_1(x) f_2(x) \qquad (x \in E). \end{aligned} \tag{1}$$

For any $f : E \to \mathbb{R}$ and $c \in \mathbb{R}$, we define the product cf of f and c as

$$(cf_1)(x) = cf_1(x) \qquad (x \in E). \tag{2}$$

For any $f_1, \ldots, f_i : E \to \mathbb{R}$, we define a function $(f_1, \ldots, f_i) : E \to \mathbb{R}^i$ as

$$(f_1, \ldots, f_i)(x) = (f_1(x), \ldots, f_i(x)) \qquad (x \in E). \tag{3}$$

For $\delta > 0$, we define the δ-neighborhood $N_\delta(x) \subset E$ of $x \in E$ as

$$N_\delta(x) = \{x' \in E \,|\, d(x, x') < \delta\}, \tag{4}$$

where $d : E \times E \to \mathbb{R}_+$ is a metric on E. For any $t \geq 0$ and a function v on a real interval, we denote the restriction of v to the interval $[-t, 0]$ as $v^{[t]}$.

II. RC Represented by Right-Infinite-Time Operators

A. Reservoir Computing Model

For a transition function $\phi : \mathbb{R}^r \times \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^r$ and an output function $\psi : \mathbb{R}^r \times \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^m$, we define a continuous-time n-input m-output state-space system (ϕ, ψ) as follows:

$$\begin{aligned} \dot{x}(t) &= \phi(x(t), u(t), t), \\ y(t) &= \psi(x(t), u(t), t), \end{aligned} \tag{5}$$

where $x(t) \in \mathbb{R}^r$, $u(t) \in \mathbb{R}^n$, $y(t) \in \mathbb{R}^m$ are the state, input, and output at time t, respectively. For a solution of (5) to exist, we assume that ϕ is locally Lipschitz continuous on $\mathbb{R}^r \times \mathbb{R}^n \times \mathbb{R}$. We consider a RC model with system (ϕ, ψ) and a static function $p : \mathbb{R}^m \to \mathbb{R}$ as the reservoir and the readout, respectively. The output of the RC model at time t is defined as $\hat{y}(t) = p(y(t))$.

In supervised learning, the RC model is used to approximate a given target system. RC provides an approximation by training only the readout p, which is the static part of the model. Hence, RC has the advantage of low computational cost for training. We discuss the approximations of single-output systems, but the results can easily be extended to multiple-output systems because an m^*-output system can be treated as m^* single-output systems.

In this paper, the RC model with reservoir (ϕ, ψ) is said to be universal if, for any target system, there is a polynomial readout p such that the RC model approximates the target. Additionally, if an RC model with a reservoir is universal, then we say the reservoir is universal. In this paper, we propose a novel sufficient condition for the reservoir to be universal.

B. Previous Work

For comparison with our result, we explain the work of Maass et al. [4] relating to inputs and outputs defined on $\mathbb{R}$. Let $A \subset \mathbb{R}^n$ be compact and $K \geq 0$. We define a set U^B of BIT inputs as follows:

$$\begin{aligned} U^B = \{u : \mathbb{R} \to A \mid \forall t_1, t_2 \in \mathbb{R}, \\ \|u(t_1) - u(t_2)\| \leq K\,|t_1 - t_2|\}\,. \end{aligned} \tag{6}$$

Let Y^B be the set of functions from $\mathbb{R}$ to $\mathbb{R}$. Maass et al. [4] represent the reservoir and the target as operators from U^B to Y^B. We call these "BIT operators" because they are maps between signals defined on the BIT interval $\mathbb{R}$. For an operator F and an input signal u, we write the output signal and its value at time t as Fu and $Fu(t)$, respectively. Maass et al. [4] assume that BIT operators have two properties called time invariance and fading memory. Time invariance is defined as follows. An operator $F : U^B \to Y^B$ is said to be time-invariant if the following holds:

$$\begin{aligned} &\forall t \in \mathbb{R}, \forall u_1, u_2 \in U^B, (\forall \tau \in \mathbb{R}, u_2(\tau) = u_1(\tau - t)) \\ &\Rightarrow (\forall \tau \in \mathbb{R}, Fu_2(\tau) = Fu_1(\tau - t))\,. \end{aligned} \tag{7}$$

Time invariance means that a temporal shift of the input also shifts the output. Fading memory is defined as follows:

Definition 1. *A BIT operator $F : U^B \to Y^B$ is said to have fading memory if the following holds:*

$$\begin{aligned} &\forall u_1 \in U^B, \forall \varepsilon > 0, \exists \delta > 0, \exists T > 0, \forall u_2 \in U^B, \\ &\max_{\tau \in [-T, 0]} \|u_1(\tau) - u_2(\tau)\| < \delta \Rightarrow |Fu_1(0) - Fu_2(0)| < \varepsilon. \end{aligned} \tag{8}$$

Fading memory means that if two input signals are close shortly before time 0, the outputs at time 0 are also close, independent of the distant past.

In the work of Maass et al. [4], an m-output reservoir is represented as a set $\mathbb{F}$ of m operators. To be universal, the reservoir $\mathbb{F}$ must have the following separation property:

Definition 2. *Let a reservoir $\mathbb{F}$ be a set of operators from U^B to Y^B. The reservoir $\mathbb{F}$ is said to have the separation property if $\mathbb{F}$ satisfies the following:*

$$\begin{aligned} &\forall u_1, u_2 \in U^B, \exists F \in \mathbb{F}, \\ &(\exists \tau \leq 0, u_1(\tau) \neq u_2(\tau)) \Rightarrow Fu_1(0) \neq Fu_2(0). \end{aligned} \tag{9}$$

The separation property means that the reservoir gives different outputs to different inputs. Suppose that the reservoir does not have the separation property, i.e., the reservoir returns the same output to two different inputs. Then, the RC model cannot approximate a target that returns different outputs to those inputs. Hence, the separation property is necessary to achieve universality. Note that (8) and (9) are the conditions for the output at time 0 because time invariance is assumed, and time 0 is chosen as a representative time.

The result of Maass et al. [4] can be described as follows:

Theorem 1 ([4])**.** *Let a reservoir $\mathbb{F}$ be a set of time-invariant operators from U^B to Y^B with fading memory. Suppose that $\mathbb{F}$ has the separation property. Then, for any time-invariant operator $F^* : U^B \to Y^B$ with fading memory and for any $\varepsilon > 0$, there exist $i \in \mathbb{N}$, $F_1, \ldots, F_i \in \mathbb{F}$, and a polynomial $p : \mathbb{R}^i \to \mathbb{R}$ that satisfy the following:*

$$\forall u \in U^B, \forall t \in \mathbb{R}, |F^*u(t) - p(F_1u(t), \ldots, F_iu(t))| < \varepsilon. \tag{10}$$

The natural number i is used to express that, even if the cardinal number of reservoir outputs $m = |\mathbb{F}|$ is infinite, approximation is possible with a finite number of elements in $\mathbb{F}$. If m is finite, we can set $i = m$. Therefore, Theorem 1 means that, for any target F^*, there is some readout p such that the RC model $\hat{F}u(t) = p(F_1u(t), \ldots, F_mu(t))$ approximates F^*, i.e., the reservoir $\mathbb{F}$ is universal.

To apply Theorem 1 to the reservoir (ϕ, ψ), we must represent the input–output relationship of the reservoir by operators. For this representation, the following is important:

Definition 3. *A system (ϕ, ψ) with a compact state set $X \subset \mathbb{R}^r$ is said to have ESP if, for any BIT input $\in U^B$, there is a unique state trajectory $x : \mathbb{R} \to X$ satisfying (5) for any $t \in \mathbb{R}$.*

This definition is a continuous-time version of the ESP described in [3]. A simple example of a system with ESP is the case where $\phi(x, u, t) = -x + u$ $(x, u, t \in \mathbb{R})$. Then, the bounded state trajectory for $u \in U^B$ is uniquely defined as

$$x(t) = \int_{-\infty}^{t} \exp(\tau - t) u(\tau) d\tau \quad (t \in \mathbb{R})\,. \tag{11}$$

Suppose that the system (ϕ, ψ) with a compact state set $X \subset \mathbb{R}^r$ has ESP. This means that the output for any given input is unique. Therefore, the input–output relationship of the system (ϕ, ψ) is represented by m BIT operators $F_1, \ldots, F_m : U^B \to Y^B$ defined as follows:

$$\begin{aligned} &F_i : u \mapsto y_i \qquad \left(u \in U^B, i \in \{1, \ldots, m\}\right), \\ &(y_1(t), \ldots, y_m(t)) = \psi(x(t), u(t)) \qquad (t \in \mathbb{R})\,, \end{aligned} \tag{12}$$

where $x : \mathbb{R} \to X$ is a state trajectory satisfying (5) for any $t \in \mathbb{R}$. Theorem 1 is applied to the reservoir (ϕ, ψ) by setting $\mathbb{F} = \{F_1, \ldots, F_m\}$. Similarly, the target F^* can be defined by a state-space system.

C. Right-Infinite Time Operators

As in the previous subsection, to apply the result of Maass et al. [4], we need to represent the reservoir by BIT operators. Without ESP, the reservoir may not be represented by well-defined operators because it can have multiple state trajectories and outputs for a single input. However, ESP is too strict for the use of a general nonlinear state-space system as a reservoir. For example, in the case of a time-invariant system, ESP requires a unique equilibrium state $x \in X$ that satisfies $\phi(x, 0) = 0$. If the reservoir has multiple equilibrium states, the state trajectories remaining at each equilibrium state exist for the zero input, which violates the ESP.

For the uniqueness of outputs, we fix the state of the system (ϕ, ψ) at time 0 and deal with the input–output relationship after time 0. This input–output relationship is represented by RIT operators, which are maps between signals defined on $\mathbb{R}_+$. We define a set U of RIT inputs as

$$\begin{aligned} U = \{u : \mathbb{R}_+ \to A \mid \forall t_1, t_2 \geq 0,& \\ \|u(t_1) - u(t_2)\| \leq K\,|t_1 - t_2|\}\,.& \end{aligned} \tag{13}$$

Let Y be the set of functions from $\mathbb{R}_+$ to $\mathbb{R}$. Let $x_0 \in \mathbb{R}^r$ be the state of the system (ϕ, ψ) at time 0. The input–output relationship of the system (ϕ, ψ) is represented by m RIT operators $F_1, \dots, F_m : U \to Y$ defined as follows:

$$\begin{aligned} &F_i : u \mapsto y_i \qquad (u \in U, i \in \{1, \dots, m\})\,, \\ &(y_1(t), \dots, y_m(t)) = \psi(x(t), u(t)) \qquad (t \in \mathbb{R}_+)\,, \end{aligned} \tag{14}$$

where $x : \mathbb{R}_+ \to \mathbb{R}^r$ is a state trajectory satisfying $x(0) = x_0$ and (5) for any $t \in \mathbb{R}_+$.

We represent the reservoir as a set $\mathbb{F}$ of RIT operators. If the reservoir is the system (ϕ, ψ), we set $\mathbb{F} = \{F_1, \dots, F_m\}$, where $F_1, \dots, F_m$ is defined in (14). The universality of a reservoir is defined as follows:

Definition 4. *Let a reservoir $\mathbb{F}$ be a set of operators from U to Y, and let $\mathbb{F}^*$ be a set of target operators from U to Y. The reservoir $\mathbb{F}$ is said to be universal for uniform approximations of an operator in $\mathbb{F}^*$ if, for any operator $F^* \in \mathbb{F}^*$ and $\varepsilon > 0$, there exist $i \in \mathbb{N}$, $F_1, \dots, F_i \in \mathbb{F}$, and a polynomial $p : \mathbb{R}^i \to \mathbb{R}$ that satisfy*

$$\forall u \in U, \forall t \geq 0, |F^*u(t) - p(F_1u(t), \dots, F_iu(t))| < \varepsilon. \tag{15}$$

The universality property means that any operator output can be approximated by a polynomial of the reservoir output.

D. Functionals

To guarantee reservoir universality, we use the following theorem:

Theorem 2 (Stone–Weierstrass theorem [19])**.** *Let E be a compact metric space and $\mathscr{F}$ be a set of continuous functions from E to $\mathbb{R}$. Suppose that, for any distinct x_1, $x_2 \in E$, there is some $f \in \mathscr{F}$ that satisfies $f(x_1) \neq f(x_2)$. Then, for any continuous function $f^* : E \to \mathbb{R}$ and for any $\varepsilon > 0$, there exist $i \in \mathbb{N}$, $f_1, \dots, f_i \in \mathscr{F}$, and a polynomial $p : \mathbb{R}^i \to \mathbb{R}$ that satisfy*

$$\forall x \in E, |f^*(x) - p(f_1(x), \dots, f_i(x))| < \varepsilon. \tag{16}$$

Equations (15) and (16) are similar, but $f_1, \dots, f_i$ are maps to real numbers, whereas $F_1, \dots, F_i$ are maps to functions. Hence, we convert operators into maps from functions to real numbers, which we call functionals.

A functional is a map from finite-length input signals to real numbers. We define a set V of left-infinite-time (LIT) inputs as follows:

$$\begin{aligned} V = \{v : \mathbb{R}_- \to A \mid \forall t_1, t_2 \leq 0,& \\ \|v(t_1) - v(t_2)\| \leq K\,|t_1 - t_2|\}\,.& \end{aligned} \tag{17}$$

The domain of the functional is defined as follows:

$$V^{\mathrm{res}} = \left\{ v^{[t]} \;\mid v \in V, t \geq 0 \right\}, \tag{18}$$

where $v^{[t]}$ is the restriction of v noted in the introduction.

To convert operators into functionals, the operators need to be causal. Causality is defined as follows:

Definition 5. *An RIT operator $F : U \to Y$ said to be causal if F satisfies*

$$\begin{aligned} &\forall u_1, u_2 \in U, \forall t \geq 0, \\ &(\forall \tau \in [0, t]\,, u_1(\tau) = u_2(\tau)) \Rightarrow Fu_1(t) = Fu_2(t). \end{aligned} \tag{19}$$

Causality means that the outputs are unaffected by future inputs. Hence, operators representing the system (ϕ, ψ) are causal. If an operator $F : U \to Y$ is causal, the value of the input $u \in U$ on $[0, t]$ uniquely defines $Fu(t)$. This means that a causal operator maps a finite-length input signal to a real number like a functional. In this paper, we consider only causal operators.

To explain the correspondence between operators and functionals, we define some maps. Let $\lambda : V^{\mathrm{res}} \to \mathbb{R}_+$ be a map returning the domain length for an input. For example, $\lambda(v) = 1$ for $v : [-1, 0] \to A$. We define a map $\sigma : V^{\mathrm{res}} \to U$ as follows:

$$\begin{aligned} &\sigma : v \mapsto u \qquad (v \in V^{\mathrm{res}})\,, \\ &u(\tau) = \begin{cases} v(\tau - t) & (0 \leq \tau < t)\,, \\ v(0) & (\tau \geq t)\,, \end{cases} \end{aligned} \tag{20}$$

where $t = \lambda(v)$. The map σ shifts an input so that it starts at time 0 and continuously extends it onto $\mathbb{R}_+$. We define a map $S : V^{\mathrm{res}} \to U \times \mathbb{R}_+$ as follows:

$$S(v) = (\sigma(v), \lambda(v)) \qquad (v \in V^{\mathrm{res}}) \tag{21}$$

The upper part of Fig. 1 represents the map S. In this subsection, we change the domain and range of an operator $F : U \to Y$ as follows:

$$\begin{array}{cccccc} F: & U & \times & \mathbb{R}_+ & \to & \mathbb{R} \\ & \Psi & & \Psi & & \Psi \\ & u & \times & t & \mapsto & Fu(t) \end{array}\,. \tag{22}$$

This is because Y is a set of functions from $\mathbb{R}_+$ to $\mathbb{R}$. The correspondence between operators and functionals is defined as follows:

Definition 6. *A causal RIT operator $F : U \to Y$ and a functional $f : V^{\mathrm{res}} \to \mathbb{R}$ are said to correspond to each other if the following holds:*

$$f = F \circ S. \tag{23}$$

The following proposition holds for (23):

Proposition 1. *For any functional $f : V^{\mathrm{res}} \to \mathbb{R}$, there is a unique causal RIT operator $F : U \to Y$ satisfying (23).*

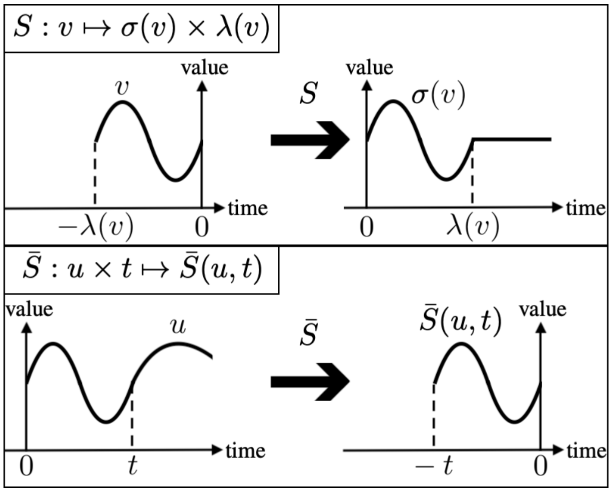


Fig. 1. Example of correspondences by maps σ and $\bar{S}$ in the case of the input dimension $m = 1$.

Proof of Proposition 1. Let $f : V^{\mathrm{res}} \to \mathbb{R}$ be an arbitrary functional. First, we prove that there is some causal operator $F : U \to Y$ satisfying (23). Let $\bar{S} : U \times \mathbb{R}_+ \to V^{\mathrm{res}}$ be the left inverse of S defined as follows:

$$\begin{aligned} &\bar{S} : (u,t) \mapsto v \qquad (u \in U, t \geq 0)\,, \\ &v(\tau) = u(\tau + t) \qquad (-t \leq \tau \leq 0)\,. \end{aligned} \tag{24}$$

As shown in the lower part of Fig. 1, the map $\bar{S}$ shifts the input u to the left by t and restricts the domain to $[-t, 0]$. The following operator $F : U \to Y$ satisfies (23):

$$F = f \circ \bar{S}. \tag{25}$$

The operator F in (25) is causal because $\bar{S}(u,t)$ depends only on the value of u on $[0,t]$.

Next, we prove that there is a unique causal operator $F : U \to Y$ satisfying (23). Suppose that the causal operators $F, F' : U \to Y$ satisfy (23). As shown in Fig. 1, the map $\bar{S}$ is not the right inverse of S, i.e., $S \circ \bar{S} \neq \mathrm{id}_{U \times \mathbb{R}_+}$. However, $F \circ S \circ \bar{S} = F$ holds because F is causal. Composing (23) with $\bar{S}$, we have (25). The same holds for F'. Hence, we have

$$F = F' = f \circ \bar{S}, \tag{26}$$

which proves Proposition 1. □

Proposition 1 means that (23) defines bijection from operators to functionals. The proof of Proposition 1 implies that (23) and (25) are equivalent. Hence, (25) defines the inverse of the bijection defined by (23). Unlike BIT operators, RIT operators do not need time invariance to correspond to the functional. In the first place, RIT operators do not have the concept of time invariance. Using this correspondence between operators and functionals, we apply Theorem 2 and guarantee reservoir universality.

E. Metric

Theorem 2 requires the functionals to have the properties of continuity and domain compactness. Therefore, a metric on V^{res} is needed. A simple method for defining the distance between two inputs of different lengths is to extend the shorter input with zero inputs to match the length of the longer one and use the uniform norm [18]. However, this method cannot distinguish between two zero inputs of different lengths. Hence, this method requires the system (ϕ, ψ) to return a constant output for the zero input. This means that either the state trajectory for the zero input must remain at the initial state x_0 or the output function ψ must be constant. We propose a novel metric that eliminates the need for this assumption.

For $w : \mathbb{R}_+ \to (0,1]$ and $t \geq 0$, we define the weighted norm $\|v\|_w$ of $v : [-t, 0] \to \mathbb{R}^n$ as

$$\|v\|_w = \sup_{\tau \in [-t,0]} \|v(\tau)\| \, w(-\tau). \tag{27}$$

For w and $\theta : \mathbb{R}_+ \to \mathbb{R}_+$, we define the distance $d(v_1, v_2)$ between v_1, $v_2 \in V^{\mathrm{res}}$ as

$$d(v_1, v_2) = \left\| v_1^{[t_{\min}]} - v_2^{[t_{\min}]} \right\|_w + |\theta(t_1) - \theta(t_2)|\,, \tag{28}$$

where $t_1 = \lambda(v_1)$, $t_2 = \lambda(v_2)$, and $t_{\min} = \min\{t_1, t_2\}$. We assume the following for the pair (w, θ):

Assumption 1.

(i) $w : \mathbb{R}_+ \to (0,1]$ *is a non-increasing function that is integrable on* $[0, \infty)$.

(ii) $\theta : \mathbb{R}_+ \to \mathbb{R}_+$ *is a strictly increasing, bounded, and continuous function.*

(iii) *For any* v_1, $v_2 \in V$, $t_1 \geq 0$, *and* $t_2 \geq t_1$,

$$\left\| v_1^{[t_2]} - v_2^{[t_2]} \right\|_w \leq \left\| v_1^{[t_1]} - v_2^{[t_1]} \right\|_w + 2\left[\theta(t_2) - \theta(t_1)\right]. \tag{29}$$

The first term of (28) compares the inputs on the intersection of their domains. From condition (i) of Assumption 1, the function w assigns greater weight to the difference in the newer part of the inputs. The second term of (28) compares the length of the two inputs via the function θ. From condition (ii) of Assumption 1, when the two inputs are longer, the distance caused by the length difference is smaller. Inequality (29) in condition (iii) is the transformation of the following triangle inequality between $v_1^{[t_1]}$, $v_2^{[t_2]}$, and $v_1^{[t_2]}$:

$$d\Big(v_1^{[t_2]}, v_2^{[t_2]}\Big) \leq d\Big(v_1^{[t_1]}, v_2^{[t_2]}\Big) + d\Big(v_1^{[t_1]}, v_1^{[t_2]}\Big). \tag{30}$$

The triangle inequality is necessary for d to be a metric.

The following two propositions hold:

Proposition 2. *The map* $d : V^{\mathrm{res}} \times V^{\mathrm{res}} \to \mathbb{R}_+$ *is a metric on* V^{res}.

Proposition 3. *Suppose that a pair* (w', θ') *satisfies Assumption 1. Let* d' *be another metric on* V^{res} *defined by* (w', θ') *as in (28). Then, metrics* d *and* d' *on* V^{res} *are equivalent, i.e., for any* $\varepsilon > 0$ *and* $v_1 \in V^{\mathrm{res}}$, *there is some* $\delta > 0$ *that satisfies the following for any* $v_2 \in V^{\mathrm{res}}$:

$$d(v_1, v_2) < \delta \Rightarrow d'(v_1, v_2) < \varepsilon, \tag{31a}$$

$$d'(v_1, v_2) < \delta \Rightarrow d(v_1, v_2) < \varepsilon. \tag{31b}$$

The proofs are given in Appendices A and B. From Proposition 3, the distances defined by any pair (w, θ) satisfying

Assumption 1 are equivalent. Therefore, the continuity of functionals on V^{res} and the compactness of sets in V^{res} do not depend on the selection of (w, θ).

We prove that there is a pair (w, θ) that satisfies Assumption 1. Clearly, there is some w that satisfies condition (i). Let $w : \mathbb{R}_+ \to (0, 1]$ satisfy condition (i). We define $\theta : \mathbb{R}_+ \to \mathbb{R}_+$ as

$$\begin{aligned} \theta(t) &= K \int_0^t \bar{w}(\tau) d\tau, \\ \bar{w}(\tau) &= \begin{cases} w(0) & (0 \leq \tau < T), \\ w(\tau - T) & (\tau \geq T), \end{cases} \end{aligned} \tag{32}$$

where $T = M/(2K)$ and $M \geq 0$ is the maximum difference between the two input values defined as

$$M = \max_{a_1, a_2 \in A} \|a_1 - a_2\|. \tag{33}$$

Because w is integrable on $\mathbb{R}_+$ and returns a positive value, θ satisfies condition (ii). Next, we prove that the pair (w, θ) satisfies condition (iii). Because the Lipschitz constant of $v_1 - v_2$ is $2K$ or less, if $t_2 - t_1 \leq T$, then

$$\left\| v_1^{[t_2]} - v_2^{[t_2]} \right\|_w - \left\| v_1^{[t_1]} - v_2^{[t_1]} \right\|_w \leq 2K (t_2 - t_1) w(t_1). \tag{34}$$

Because $\bar{w}(\tau) \geq w(t_1)$ $(t_1 \leq \tau \leq t_2)$, we have

$$2[\theta(t_2) - \theta(t_1)] = 2 \int_{t_1}^{t_2} K \bar{w}(\tau) d\tau \geq 2K (t_2 - t_1) w(t_1). \tag{35}$$

These statements imply (29).

If $t_2 - t_1 > T$, then (33) implies

$$\left\| v_1^{[t_2]} - v_2^{[t_2]} \right\|_w - \left\| v_1^{[t_1]} - v_2^{[t_1]} \right\|_w \leq M w(t_1). \tag{36}$$

Because $\bar{w}(\tau) \geq w(t_1)$ $(t_1 \leq \tau \leq t_1 + T)$, we have

$$\begin{aligned} 2[\theta(t_2) - \theta(t_1)] &= 2 \int_{t_1}^{t_2} K \bar{w}(\tau) d\tau \\ &> 2 \int_{t_1}^{t_1 + T} K \bar{w}(\tau) d\tau \\ &\geq 2KTw(t_1) = Mw(t_1). \end{aligned} \tag{37}$$

These statements imply (29). Therefore, the pair (w, θ) satisfies condition (iii). For example, the following pair (w, θ) satisfying Assumption 1 is obtained from (32):

$$\begin{aligned} w(t) &= \exp(-t) \qquad (t \geq 0), \\ \theta(t) &= \begin{cases} Kt & (0 \leq t < T), \\ K(T + 1 - \exp(T - t)) & (t \geq T). \end{cases} \end{aligned} \tag{38}$$

In the next subsection, we deal with functional continuity and domain compactness with respect to the metric d discussed in this subsection.

F. Completion

To apply Theorem 2, the domain V^{res} of functions must be compact. However, the domain V^{res} of functionals is not compact because, for any $v \in V$, an element $v^{[t]}$ of V^{res} does not converge on V^{res} as $t \to \infty$. Because $v^{[t]}$ is the restriction of $v \in V$ to $[-t, 0]$, it is natural to define $v^{[t]} \to v$ as $t \to \infty$. Therefore, we complete V^{res} by summing V. Note that $V^{\mathrm{res}} \cap V = \emptyset$ because V^{res} is the set of functions on a finite interval, but V is the set of functions on $\mathbb{R}_-$.

We extend the metric d onto V. We define the distance between $v_1 \in V^{\mathrm{res}}$ and $v_2 \in V$ as

$$d(v_1, v_2) = \left\| v_1 - v_2^{[t_1]} \right\|_w + \theta(\infty) - \theta(t_1), \tag{39}$$

where $\theta(\infty) = \lim_{t \to \infty} \theta(t)$ and $t_1 = \lambda(v_1)$. We define the distance between v_1, $v_2 \in V$ as

$$d(v_1, v_2) = \|v_1 - v_2\|_w, \tag{40}$$

where the weighted norm $\|v\|_w$ of $v \in V$ is defined as

$$\|v\|_w = \sup_{\tau \leq 0} \|v(\tau)\| w(-\tau). \tag{41}$$

By the above extension of d, Propositions 2 and 3 hold on $V^{\mathrm{res}} \cup V$. This can be proved in the same way by defining $\lambda(v) = \infty$ and $v^{[\infty]} = v$ for any $v \in V$. Hereafter, we use d as the metric on $V^{\mathrm{res}} \cup V$. The following two propositions hold:

Proposition 4. *The set V is compact.*

Proposition 4 is proved in the same way as Lemma 1 of [16].

Proposition 5. *The set $V^{\mathrm{res}} \cup V$ is compact.*

The proof is given in Appendix C. For any $v \in V$, a function $v^{[t]} \in V^{\mathrm{res}}$ converges to v as $t \to \infty$. Therefore, from Proposition 5, the set $V^{\mathrm{res}} \cup V$ is a compact metric space of accumulation points of V^{res}.

We define the value of the functional $f : V^{\mathrm{res}} \to \mathbb{R}$ on V added to the domain by completion as follows:

$$f(v) = \lim_{i \to \infty} f\left(v^{[i]}\right) \qquad (v \in V). \tag{42}$$

In this subsection, let an operator $F : U \to Y$ correspond to the functional f. As explained in the remainder of this subsection, it is important for the convergence of the limit in (42) that the operator F has fading memory defined as follows:

Definition 7. *A RIT operator $F : U \to Y$ is said to have fading memory if F satisfies*

$$\begin{aligned} &\forall \varepsilon > 0, \exists \delta > 0, \exists T \geq 0, \forall u_1, u_2 \in U, \forall t_1, t_2 \geq T, \\ &\max_{\tau \in [-T, 0]} \|u_1(\tau + t_1) - u_2(\tau + t_2)\| < \delta \\ &\Rightarrow |Fu_1(t_1) - Fu_2(t_2)| < \varepsilon. \end{aligned} \tag{43}$$

The difference from the fading memory (8) of BIT operators is that (43) is the condition for outputs at arbitrary times t_1 and t_2. This is because RIT operators do not have the concept of time invariance, and time 0 cannot represent any other time.

The following assumptions about the operator F are closely related to the continuity of the functional f.

Assumption 2.

(i) $\forall \varepsilon > 0, \forall u_1 \in U, \forall t \geq 0, \exists \delta > 0, \forall u_2 \in U,$ $\max_{\tau \in [0, t]} \|u_1(\tau) - u_2(\tau)\| < \delta \Rightarrow |Fu_1(t) - Fu_2(t)| < \varepsilon.$

(ii) *The image $F(U)$ is equicontinuous.*

(iii) *The operator F has fading memory.*

Condition (i) means that $Fu(t)$ is continuous with respect to the value of u on $[0,t]$. This implies that $Fu(t)$ does not depend on the value of u after t. Hence, causality is derived from condition (i). Then, the following two propositions hold:

Proposition 6. *If the operator F satisfies conditions* (i) *and* (ii) *of Assumption 2, the functional $f : V^{\mathrm{res}} \to \mathbb{R}$ is continuous.*

Proposition 7. *If the operator F satisfies condition* (iii) *of Assumption 2, i.e., F has fading memory, the value of f on V can be defined as (42), and $f : V^{\mathrm{res}} \cup V \to \mathbb{R}$ is continuous at any $v \in V$.*

The proofs are given in Appendices D and E. From Proposition 7, if F has fading memory, we can continuously define an output of f for a LIT input $v \in V$ as a limit of outputs for a finite-length input in V^{res}. A continuous functional from LIT inputs to real numbers can be converted into a time-invariant BIT operator with fading memory [16]. Therefore, an RIT operator with fading memory defines the time-invariant BIT operator with fading memory via the functional. In the remainder of this paper, let the domain of functionals corresponding to an operator with fading memory be $V^{\mathrm{res}} \cup V$ and the value of the functionals on V be defined by (42).

From Propositions 6 and 7, if the operator F satisfies Assumption 2, the functional f is continuous on the compact set $V^{\mathrm{res}} \cup V$, i.e., f is uniformly continuous. The converse is also true, as shown by the following proposition:

Proposition 8. *If and only if the operator $F : U \to Y$ satisfies Assumption 2, the functional $f : V^{\mathrm{res}} \to \mathbb{R}$ is uniformly continuous.*

The proof is given in Appendix F.

G. Application of Stone–Weierstrass Theorem

We define the separation property of a reservoir as follows:

Definition 8. *Let a reservoir $\mathbb{F}$ be a set of operators from U to Y with fading memory, and let $\mathscr{F}$ be a set of functionals from $V^{\mathrm{res}} \cup V$ to $\mathbb{R}$ corresponding to each operator in $\mathbb{F}$. The reservoir $\mathbb{F}$ is said to have the separation property if, for any distinct v_1, $v_2 \in V^{\mathrm{res}} \cup V$, there is some $f \in \mathscr{F}$ that satisfies $f(v_1) \neq f(v_2)$.*

Unlike Definition 2, this separation property is expressed by functionals, but the meaning is the same, i.e., different inputs give different outputs. Let $\mathbb{F}^*$ be the set of operators $F : U \to Y$ satisfying Assumption 2. Then, the following theorem guarantees reservoir universality in Definition 4:

Theorem 3. *Suppose that a reservoir $\mathbb{F} \subset \mathbb{F}^*$ has the separation property. Then, the reservoir $\mathbb{F}$ is universal for uniform approximations of an operator in $\mathbb{F}^*$.*

Theorem 3 gives the same result as Theorem 1 for a reservoir represented by RIT operators.

Proof of Theorem 3. Let $\mathscr{F}$ be a set of functionals from $V^{\mathrm{res}} \cup V$ to $\mathbb{R}$ corresponding to each operator in $\mathbb{F}$ and $f^* : V^{\mathrm{res}} \cup V \to \mathbb{R}$ be a functional corresponding to F^*. From Proposition 5, the set $V^{\mathrm{res}} \cup V$ is compact. From Propositions 6 and 7, functionals in $\mathscr{F}$ and f^* are continuous on $V^{\mathrm{res}} \cup V$. Because $\mathbb{F}$ has the separation property, for any distinct v_1, $v_2 \in V^{\mathrm{res}} \cup V$, there is some $f \in \mathscr{F}$ that satisfies $f(v_1) \neq f(v_2)$. Therefore, from Theorem 2, there exist $i \in \mathbb{N}$, $f_1, \ldots, f_i \in \mathscr{F}$, and a polynomial $p : \mathbb{R}^i \to \mathbb{R}$ that satisfy

$$\forall v \in V^{\mathrm{res}} \cup V, |f^*(v) - p(f_1(v), \ldots, f_i(v))| < \varepsilon. \tag{44}$$

From (25), we can obtain (15) by substituting $v = \bar{S}(u,t) \in V^{\mathrm{res}}$ into (44), which proves Theorem 3. □

III. Reservoir without Fading Memory

A. Extension of the Stone–Weierstrass Theorem

Theorem 3 is still not sufficient to allow us to use a general nonlinear state-space system as a reservoir. This is because Theorem 3 requires the operators in a reservoir to have fading memory (condition (iii) in Assumption 2). For the fading memory of operators, it is still important that a system (ϕ, ψ) has a unique equilibrium state. For example, we consider the following time-invariant case:

$$\begin{aligned} \phi(x,u) &= -x\,(x+1)\,(x-1) + u, \\ \psi(x,u) &= x \qquad (x \in \mathbb{R}, u \in \mathbb{R}). \end{aligned} \tag{45}$$

This system has two stable equilibria and one unstable equilibrium. Let the state at time 0 be $x_0 = 0$ and the RIT operator representing the system be F. We show that F does not have fading memory. We define two inputs u_1 and u_2 as follows:

$$u_1(t) = \exp(-t), \quad u_2(t) = -\exp(-t) \quad (t \geq 0). \tag{46}$$

The state trajectories for u_1 and u_2 converge to stable equilibria 1 and -1, respectively. Hence, we have

$$\lim_{t\to\infty} |Fu_1(t) - Fu_2(t)| = 2. \tag{47}$$

Let $\varepsilon < 2$ be a positive number. Because $u_1(t) - u_2(t)$ converges to 0 as $t \to \infty$, for any $\delta > 0$ and $T \geq 0$, there is some $t \geq T$ satisfying the following:

$$\begin{aligned} &\max_{\tau\in[-T,0]} \|u_1(\tau+t) - u_2(\tau+t)\| < \delta \\ &\wedge |Fu_1(t) - Fu_2(t)| \geq \varepsilon. \end{aligned} \tag{48}$$

This violates the fading memory of F.

Theorem 3 is proved using Theorem 2. Because Theorem 2 requires continuity of functionals, operators in the reservoir must satisfy Assumption 2, including fading memory. To guarantee the universality of a reservoir composed of operators without fading memory, we extend Theorem 2 as follows to handle discontinuous functionals:

Theorem 4. *Let E be a compact metric space and $\mathscr{F}$ be a set of bounded functions from E to $\mathbb{R}$. Suppose that the following holds for any distinct x_1, $x_2 \in E$:*

$$\begin{aligned} &\exists \delta > 0, \exists i \in \mathbb{N}, \exists f_1, \ldots, f_i \in \mathscr{F}, \\ &\overline{(f_1,\ldots,f_i)\,(N_\delta(x_1))} \cap \overline{(f_1,\ldots,f_i)\,(N_\delta(x_2))} = \emptyset. \end{aligned} \tag{49}$$

Then, for any continuous function $f^*: E \to \mathbb{R}$ and for any $\varepsilon > 0$, there exist $i \in \mathbb{N}$, $f_1, \ldots, f_i \in \mathscr{F}$, and a polynomial $p : \mathbb{R}^i \to \mathbb{R}$ that satisfy the following:

$$\forall x \in E, |f^*(x) - p(f_1(x), \ldots, f_i(x))| < \varepsilon. \tag{50}$$

The function $(f_1, \ldots, f_i)$ and the neighborhood N_δ are defined in the introduction.

Proof of Theorem 4. Let $B(E)$ and $C(E)$ be the sets of bounded and continuous functions from E to $\mathbb{R}$, respectively. Because E is compact, $C(E) \subset B(E)$ holds. Let $\mathbb{R}[\mathscr{F}] \subset B(E)$ be the polynomial ring in $\mathscr{F}$ over the field $\mathbb{R}$, i.e., $\mathbb{R}[\mathscr{F}]$ is the minimum set satisfying

(i) $\mathscr{F} \subset \mathbb{R}[\mathscr{F}]$ and $e \in \mathbb{R}[\mathscr{F}]$, where $e \in C(E)$ is a constant function that returns a value of 1.
(ii) $\forall f_1, f_2 \in \mathbb{R}[\mathscr{F}], \forall c \in \mathbb{R}$,
$f_1 + f_2 \in \mathbb{R}[\mathscr{F}], f_1 f_2 \in \mathbb{R}[\mathscr{F}], c f_1 \in \mathbb{R}[\mathscr{F}]$.

We define the norm of $f \in B(E)$ as

$$\|f\| = \sup_{x \in E} |f(x)| . \tag{51}$$

We define the distance between f_1, $f_2 \in B(E)$ as $\|f_1 - f_2\|$. The following holds for the closure $\overline{\mathbb{R}[\mathscr{F}]}$ of $\mathbb{R}[\mathscr{F}]$.

$$\begin{aligned} &\forall f_1, f_2 \in \overline{\mathbb{R}[\mathscr{F}]}, \forall c \in \mathbb{R}, \\ &f_1 + f_2 \in \overline{\mathbb{R}[\mathscr{F}]}, f_1 f_2 \in \overline{\mathbb{R}[\mathscr{F}]}, c f_1 \in \overline{\mathbb{R}[\mathscr{F}]}. \end{aligned} \tag{52}$$

The result of Theorem 4 is equivalent to $C(E) \subset \overline{\mathbb{R}[\mathscr{F}]}$.

Let f^* be an arbitrary function in $C(E)$. Theorem 4 is proved using the following three lemmas.

Lemma 1. *For $f_1, \ldots, f_i \in \overline{\mathbb{R}[\mathscr{F}]}$, we define $f_{\max} = \max\{f_1, \ldots, f_i\}$ and $f_{\min} = \min\{f_1, \ldots, f_i\}$ as follows:*

$$\begin{aligned} &f_{\max}(x) = \max\{f_1(x), \ldots, f_i(x)\}, \\ &f_{\min}(x) = \min\{f_1(x), \ldots, f_i(x)\} \qquad (x \in E). \end{aligned} \tag{53}$$

Then, $f_{\max}$, $f_{\min} \in \overline{\mathbb{R}[\mathscr{F}]}$.

For the proof of Lemma 1, see p.138 of [19].

Lemma 2. *For any a, $b \in E$, there is some $f \in \overline{\mathbb{R}[\mathscr{F}]}$ that is continuous at a and b and satisfies $f(a) = f^*(a)$ and $f(b) = f^*(b)$.*

Proof of Lemma 2. If $a = b$, a constant function $f = f^*(a)e \in \mathbb{R}[\mathscr{F}]$ satisfies the condition. If $a \neq b$, from the condition of Theorem 4, there exist $\delta > 0$, $i \in \mathbb{N}$, and $f_1, \ldots, f_i \in \mathscr{F}$ that satisfy

$$\overline{(f_1, \ldots, f_i)(N_\delta(a))} \cap \overline{(f_1, \ldots, f_i)(N_\delta(b))} = \emptyset. \tag{54}$$

We define $Y \subset \mathbb{R}^i$ as

$$Y = (f_1, \ldots, f_i)(E). \tag{55}$$

Because $f_1, \ldots, f_i$ is bounded, $\overline{Y}$ is compact. We define Y_a, $Y_b \subset Y \subset \mathbb{R}^i$ as

$$\begin{aligned} Y_a &= (f_1, \ldots, f_i)(N_\delta(a)), \\ Y_b &= (f_1, \ldots, f_i)(N_\delta(b)). \end{aligned} \tag{56}$$

Because $\overline{Y_a} \cap \overline{Y_b} = \emptyset$, Urysohn's lemma implies that there is a continuous function $g : \overline{Y} \to \mathbb{R}$ satisfying

$$\begin{aligned} g(y) &= f^*(a) \quad (y \in Y_a), \\ g(y) &= f^*(b) \quad (y \in Y_b). \end{aligned} \tag{57}$$

We define the function $f \in B(E)$ as

$$f(x) = g((f_1, \ldots, f_i)(x)). \tag{58}$$

From (56) and (57), the function f is continuous at a and b and satisfies $f(a) = f^*(a)$ and $f(b) = f^*(b)$.

We now prove that $f \in \overline{\mathbb{R}[\mathscr{F}]}$. Let ε be an arbitrary positive number. Because g is continuous on the compact set $\overline{Y} \subset \mathbb{R}^i$, Theorem 2 implies that there is a polynomial $q : \mathbb{R}^i \to \mathbb{R}$ that satisfies

$$\forall y \in \overline{Y}, |g(y) - q(y)| < \varepsilon. \tag{59}$$

We define $f' \in \mathbb{R}[\mathscr{F}]$ as

$$f'(x) = q((f_1, \ldots, f_i)(x)). \tag{60}$$

From (59), $\|f - f'\| < \varepsilon$ holds. Therefore, $f \in \overline{\mathbb{R}[\mathscr{F}]}$. □

Lemma 3. *For any $a \in E$ and $\varepsilon > 0$, there is some $f \in \overline{\mathbb{R}[\mathscr{F}]}$ that is continuous at a and satisfies $f(x) < f^*(x) + \varepsilon$ $(x \in E)$ and $f(a) = f^*(a)$.*

Proof of Lemma 3. From Lemma 2, for any $b \in E$, there is some $f_b \in \overline{\mathbb{R}[\mathscr{F}]}$ that is continuous at a and b and satisfies $f_b(a) = f^*(a)$ and $f_b(b) = f^*(b)$. Because f^* and f_b are continuous at b, there is some $\delta_b > 0$ that satisfies $f_b(x) < f^*(x) + \varepsilon$ $(x \in N_{\delta_b}(b))$. The family $\{N_{\delta_b}(b) \,|b \in E\}$ is an open cover of the compact set E. Therefore, there exist $i \in \mathbb{N}$ and $b_1, \ldots, b_i \in E$ such that $\left\{N_{\delta_{b_1}}(b_1), \ldots, N_{\delta_{b_i}}(b_i)\right\}$ covers E. The function $f = \min\{f_{b_1}, \ldots, f_{b_i}\}$ is continuous at a and satisfies $f(x) < f^*(x) + \varepsilon$ $(x \in E)$ and $f(a) = f^*(a)$. From Lemma 1, $f \in \overline{\mathbb{R}[\mathscr{F}]}$. □

Theorem 4 is proved using the above lemmas. From Lemma 3, for any $a \in E$ and $\varepsilon > 0$, there is some $f_a \in \overline{\mathbb{R}[\mathscr{F}]}$ that is continuous at a and satisfies $f_a(x) < f^*(x) + \varepsilon$ $(x \in E)$ and $f_a(a) = f^*(a)$. Because f^* and f_a are continuous at a, there is some $\delta_a > 0$ that satisfies $f_a(x) > f^*(x) - \varepsilon$ $(x \in N_{\delta_a}(a))$. The family $\{N_{\delta_a}(a) \,|a \in E\}$ is an open cover of the compact set E. Therefore, there exist $i \in \mathbb{N}$ and $a_1, \ldots, a_i \in E$ such that $\left\{N_{\delta_{a_1}}(a_1), \ldots, N_{\delta_{a_i}}(a_i)\right\}$ covers E. The function $f = \max\{f_{a_1}, \ldots, f_{a_i}\}$ satisfies $f^*(x) - \varepsilon < f(x) < f^*(x) + \varepsilon$ $(x \in E)$, i.e., $\|f^* - f\| < \varepsilon$. From Lemma 1, $f \in \overline{\mathbb{R}[\mathscr{F}]}$.

From the above, for any $f^* \in C(E)$ and $\varepsilon > 0$, there is some $f \in \overline{\mathbb{R}[\mathscr{F}]}$ that satisfies $\|f^* - f\| < \varepsilon$. Therefore, $C(E)$ is included by the closure of $\overline{\mathbb{R}[\mathscr{F}]}$, i.e., $\overline{\mathbb{R}[\mathscr{F}]}$ itself, which proves Theorem 4. □

B. Universality

To state the main theorem in this section, we explain some properties of operators. An operator $F : U \to Y$ is said to be bounded if the image $F(U)$ is uniformly bounded.

Lemma 4. *If an operator $F : U \to Y$ satisfies Assumption 2, F is bounded.*

Proof of Lemma 4. The statement is clear from Propositions 6 and 7, together with the compactness of $V^{\mathrm{res}} \cup V$. □

The following introduces a novel concept on a set of operators:

Definition 9. *Let $\mathbb{F}$ be a set of causal operators from U to Y, and let $\mathscr{F}$ be a set of functionals from V^{res} to $\mathbb{R}$ corresponding to each operator in $\mathbb{F}$. The set $\mathbb{F}$ is said to have the neighborhood separation property (NSP) if the following holds for any distinct v_1, $v_2 \in V^{\mathrm{res}} \cup V$:*

$$\begin{aligned} &\exists \delta > 0, \exists i \in \mathbb{N}, \exists f_1, \ldots, f_i \in \mathscr{F}, \\ &\overline{(f_1, \ldots, f_i)\,(N_\delta(v_1) \cap V^{\mathrm{res}})} \\ &\cap \overline{(f_1, \ldots, f_i)\,(N_\delta(v_2) \cap V^{\mathrm{res}})} = \emptyset. \end{aligned} \tag{61}$$

The set $N_\delta(v)$ is a δ-neighborhood of $v \in V^{\mathrm{res}} \cup V$, but only its intersection with V^{res} is used. Therefore, (61) does not depend on the value of $f \in \mathscr{F}$ on V. Hence, the NSP can be defined for $\mathbb{F}$, even if there is some $f \in \mathscr{F}$ without fading memory. The separation property of Definition 8 guarantees that distinct points correspond to different values. In contrast, the NSP of Definition 9 guarantees that the images of neighborhoods of distinct points are disjoint from each other. The NSP is a stronger condition than the separation property. However, the following lemma holds for the NSP:

Lemma 5. *Suppose that a set $\mathbb{F}$ of operators from U to Y satisfying Assumption 2 has the separation property. Then, the set $\mathbb{F}$ has the NSP.*

Proof of Lemma 5. Let $\mathscr{F}$ be a set of functionals from $V^{\mathrm{res}} \cup V$ to $\mathbb{R}$ corresponding to each operator in $\mathbb{F}$. Because $\mathbb{F}$ has the separation property, for any distinct v_1, $v_2 \in V^{\mathrm{res}} \cup V$, there is some $f \in \mathscr{F}$ that satisfies $f(v_1) \neq f(v_2)$. Because the operator corresponding to f satisfies Assumption 2, Proposition 8 implies that the functional f is continuous. Hence, we have

$$\exists \delta > 0, \overline{f(N_\delta(v_1) \cap V^{\mathrm{res}})} \cap \overline{f(N_\delta(v_2) \cap V^{\mathrm{res}})} = \emptyset, \tag{62}$$

which proves Lemma 5. □

Therefore, the NSP is a weaker condition than the combination of the separation property and Assumption 2.

The main theorem in Section III is as follows:

Theorem 5. *Suppose that a set $\mathbb{F}$ of bounded and causal operators from U to Y has the NSP. Then, the reservoir $\mathbb{F}$ is universal for uniform approximations of an operator in $\mathbb{F}^*$.*

In Theorem 3, the reservoir $\mathbb{F}$ must have the separation property, and operators in $\mathbb{F}$ must satisfy Assumption 2. In Theorem 5, the reservoir $\mathbb{F}$ must have the NSP, and operators in $\mathbb{F}$ must be bounded. From Lemmas 4 and 5, the condition of Theorem 5 is obtained from that of Theorem 3. Moreover, not only does the condition of Theorem 3 explicitly not include the three conditions of Assumption 2, but it does not actually require them. Therefore, Theorem 5 requires the operators in the reservoir to have neither fading memory nor continuity with respect to inputs and time.

Proof of Theorem 5. Let $\mathscr{F}$ be a set of functionals from V^{res} to $\mathbb{R}$ corresponding to each operator in $\mathbb{F}$, and let $f^* : V^{\mathrm{res}} \cup V \to \mathbb{R}$ be a functional corresponding to F^*. From Proposition 5, the set $V^{\mathrm{res}} \cup V$ is compact. Because the operators in $\mathbb{F}$ are bounded, the functionals in $\mathscr{F}$ are also bounded. From Propositions 6 and 7, the functional f^* is continuous on $V^{\mathrm{res}} \cup V$. Because the operators in $\mathbb{F}$ do not always have fading memory, we cannot define the value of $f \in \mathscr{F}$ on V by (42). Therefore, we define the value as

$$f(v) = 0 \qquad (v \in V)\,. \tag{63}$$

For $v \in V^{\mathrm{res}} \cup V$, we define a functional $g_v : V^{\mathrm{res}} \cup V \to \mathbb{R}$ as

$$g_v(v') = \begin{cases} d(v, v') + 1 & (v' \in V)\,, \\ 0 & (v' \in V^{\mathrm{res}})\,. \end{cases} \tag{64}$$

Because the functional g_v is continuous on the compact set V, g_v is bounded on $V^{\mathrm{res}} \cup V$. We define a set $\mathscr{G}$ of bounded functionals as

$$\mathscr{G} = \{g_v \,|\, v \in V^{\mathrm{res}} \cup V\}\,. \tag{65}$$

Lemma 6. *For any distinct v_1, $v_2 \in V^{\mathrm{res}} \cup V$, the following holds:*

$$\begin{aligned} &\exists \delta > 0, \exists i \in \mathbb{N}, \exists f_1, \ldots, f_i \in \mathscr{F} \cup \mathscr{G}, \\ &\overline{(f_1, \ldots, f_i)\,(N_\delta(v_1))} \cap \overline{(f_1, \ldots, f_i)\,(N_\delta(v_2))} = \emptyset. \end{aligned} \tag{66}$$

Proof of Lemma 6. For $\delta > 0$, we define sets $V_1(\delta)$, $V_1^{\mathrm{res}}(\delta)$, $V_2(\delta)$, and $V_2^{\mathrm{res}}(\delta)$ as

$$\begin{aligned} V_1(\delta) &= N_\delta(v_1) \cap V, \quad V_1^{\mathrm{res}}(\delta) = N_\delta(v_1) \cap V^{\mathrm{res}}, \\ V_2(\delta) &= N_\delta(v_2) \cap V, \quad V_2^{\mathrm{res}}(\delta) = N_\delta(v_2) \cap V^{\mathrm{res}}. \end{aligned} \tag{67}$$

Because $\mathbb{F}$ has the NSP, we have

$$\begin{aligned} &\exists \delta > 0, \exists i \in \mathbb{N}, \exists f_1, \ldots, f_i \in \mathscr{F}, \\ &\overline{(f_1, \ldots, f_i)\,(V_1^{\mathrm{res}})} \cap \overline{(f_1, \ldots, f_i)\,(V_2^{\mathrm{res}})} = \emptyset. \end{aligned} \tag{68}$$

We need to prove that $g_{v_1} \in \mathscr{G}$ satisfies

$$\exists \delta > 0, \overline{g_{v_1}(V_1(\delta))} \cap \overline{g_{v_1}(V_2(\delta))} = \emptyset. \tag{69}$$

If v_1, $v_2 \in V$, (69) holds because the restriction of g_{v_1} to V is continuous and satisfies $g_{v_1}(v_1) \neq g_{v_1}(v_2)$. If $v_1 \in V^{\mathrm{res}}$, we have $V_1(\delta) = \emptyset$ for $\delta < \theta(\infty) - \theta(t_1)$. Hence, (69) holds. If $v_2 \in V^{\mathrm{res}}$, (69) can be shown in the same way.

From the definition of g_{v_1}, for any $\delta > 0$, we have

$$\begin{aligned} &g_{v_1}(V_1^{\mathrm{res}}(\delta)) = g_{v_1}(V_2^{\mathrm{res}}(\delta)) = \{0\}\,, \\ &g_{v_1}(V_1(\delta)), g_{v_1}(V_2(\delta)) \subset [1, \infty)\,. \end{aligned} \tag{70}$$

Therefore, we have

$$\begin{aligned} \overline{g_{v_1}(V_1^{\mathrm{res}}(\delta))} \cap \overline{g_{v_1}(V_2(\delta))} = \emptyset, \\ \overline{g_{v_1}(V_1(\delta))} \cap \overline{g_{v_1}(V_2^{\mathrm{res}}(\delta))} = \emptyset. \end{aligned} \tag{71}$$

From (68), (69), and (71), there is some $\delta > 0$ satisfying

$$\overline{(f_1, \ldots, f_i, g_{v_1})\,(N_\delta(v_1))} \cap \overline{(f_1, \ldots, f_i, g_{v_1})\,(N_\delta(v_2))} = \emptyset. \tag{72}$$

□

From Theorem 4 and Lemma 6, there exist $j \in \mathbb{N}$, $f_1, \ldots, f_j \in \mathscr{F} \cup \mathscr{G}$, and a polynomial $p' : \mathbb{R}^j \to \mathbb{R}$ that satisfy

$$\forall v \in V^{\mathrm{res}} \cup V, |f^*(v) - p'(f_1(v), \ldots, f_j(v))| < \varepsilon. \tag{73}$$

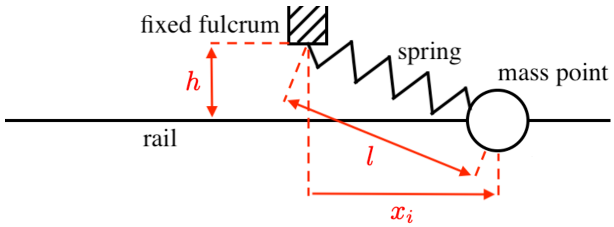


Fig. 2. Structure of the ith subsystem of the reservoir.

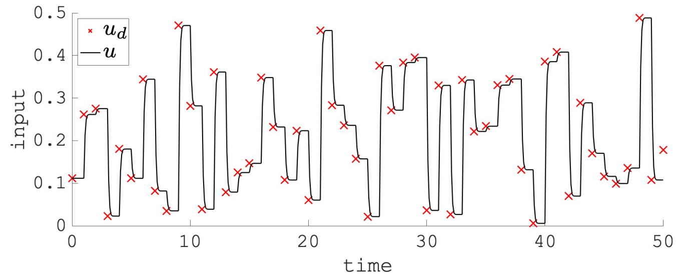


Fig. 3. Discrete-time input u_d and continuous-time input u for test.

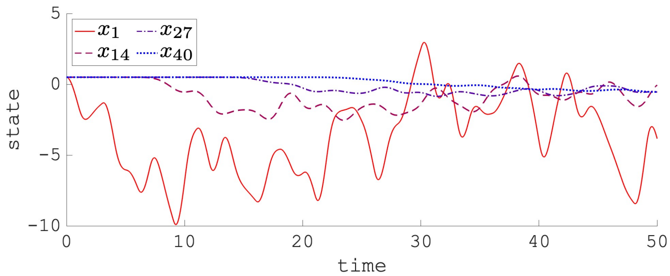


Fig. 4. Reservoir states x_1, x_{14}, x_{27}, x_{40} for test input.

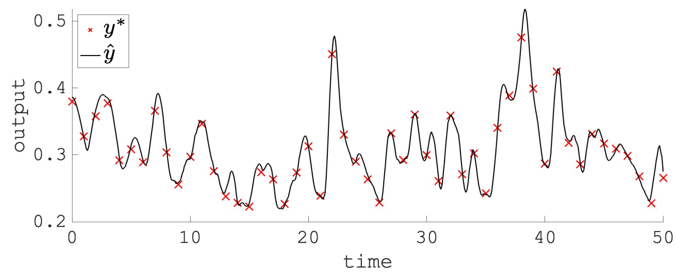


Fig. 5. Target output y^* and RC model output $\hat{y}$ for test input.

Let i be the number of functionals contained in $\mathscr{F}$ among $f_1, \ldots, f_j$. Without loss of generality, we can assume that $f_1, \ldots, f_i \in \mathscr{F}$ and $f_{i+1}, \ldots, f_j \in \mathscr{G}$. We define a polynomial $p : \mathbb{R}^i \to \mathbb{R}$ as

$$p(x_1, \ldots, x_i) = p'(x_1, \ldots, x_i, 0, \ldots, 0) \quad (x_1, \ldots, x_i \in \mathbb{R}) \,. \tag{74}$$

For $v \in V^{\mathrm{res}}$, $f_{i+1}(v), \ldots, f_j(v) = 0$ holds. Therefore, from (25), we obtain (15) by substituting $v = \bar{S}(u,t) \in V^{\mathrm{res}}$ and (74) into (73). Theorem 5 has been proved. □

C. Example

We present an example of a successful reservoir without fading memory. Consider a reservoir that consists of 40 subsystems, where the ith subsystem is defined as follows for $1 \le i \le 40$:

$$\begin{aligned}
\dot{x}_i &= v_i, \\
\dot{v}_i &= \frac{1}{m}\left[k_a x_i \left(\frac{l}{\sqrt{x_i^2 + h^2}} - 1\right) - c_i v_i + F_i\right], \\
F_i &= \begin{cases} k_b\left(x_2 - x_1\right) + 40u - 10 & (i = 1), \\ k_b\left(x_{i+1} + x_{i-1} - 2x_i\right) & (2 \le i \le 39), \\ k_b\left(x_{39} - x_{40}\right) & (i = 40), \end{cases}
\end{aligned} \tag{75}$$

where x_i, $v_i \in \mathbb{R}$ are the states, $u \in [0, 0.5]$ is the input, and h, l, m, c_i, k_a, $k_b \in \mathbb{R}$ are positive constants. We define the state vectors x, $v \in \mathbb{R}^{40}$ and the reservoir output $y \in \mathbb{R}^{40}$ as follows:

$$x = \left(x_1, \ldots, x_{40}\right), \quad v = \left(v_1, \ldots, v_{40}\right), \quad y = x. \tag{76}$$

Each subsystem is a spring–mass–damper system. As shown in Fig. 2, a spring with a natural length l and coefficient k_a connects a point of mass m to a fixed fulcrum. The mass point moves on a straight rail at a distance h from the fixed point while subject to viscous friction c_i. The states x_i and v_i are the position and velocity of the mass point. The mass point in the ith subsystem is connected to the mass points in the $(i-1)$th and $(i+1)$th subsystems by a spring with a constant k_b. The variable F_i is the force that a mass point receives from other mass points and the input.

We set the constants as follows:

$$\begin{aligned}
&h = 1.2,\ l = 1.3,\ m = 1,\ k_a = 0.05,\ k_b = 3, \\
&c_i = \begin{cases} 0.3 & (1 \le i \le 39), \\ 2 & (i = 40). \end{cases}
\end{aligned} \tag{77}$$

The spring in the ith subsystem has a natural length of l when $x_i^2 = l^2 - h^2 = 1/4$. Therefore, for the zero input, $(x, v) = (\mathbf{1}/2, 0)$ and $(x, v) = (-\mathbf{1}/2, 0)$ are stable equilibrium states, where $\mathbf{1} = (1, \ldots, 1) \in \mathbb{R}^{40}$. Hence, the reservoir does not have fading memory, like the example in Section III.A. We set the initial state of the reservoir as $x(0) = \mathbf{1}/2$ and $v(0) = 0$.

We use a neural network as the readout $p : \mathbb{R}^{40} \to \mathbb{R}$. Because the neural network is a continuous function, it can be replaced by polynomials. The neural network p has three hidden layers with 60 nodes. As activation functions, we use rectified linear units (ReLU) in the hidden layers and the identity function in the output layer. The output of the RC model is defined as follows:

$$\hat{y} = p(y). \tag{78}$$

As the approximation target, we use the following discrete-time system:

$$\begin{aligned}
y^*(t+1) =& 0.3x_1^*(t) + 0.05x_1^*(t)\left[\sum_{i=1}^{10} x_i^*(t)\right] \\
&+ 1.5z_9^*(t)u_d(t) + 0.1, \\
x_i^*(t+1) =& \begin{cases} y^*(t+1) & (i=1) \\ x_{i-1}^*(t) & (2 \le i \le 10) \end{cases}, \\
z_i^*(t+1) =& \begin{cases} u_d(t) & (i=1) \\ z_{i-1}^*(t) & (2 \le i \le 9) \end{cases},
\end{aligned} \tag{79}$$

where $x_i^*(t) \in \mathbb{R}$ $(1 \le i \le 10)$ and $z_i^*(t) \in \mathbb{R}$ $(1 \le i \le 9)$ are the states, and $u_d(t) \in [0, 0.5]$ and $y^*(t) \in \mathbb{R}$ are the input and output, respectively. The states x_i^* and z_i^* are the memory of the past output and input, respectively. We set the initial state of the target (79) as $x_i^*(0) = 0.3798$ $(1 \le i \le 10)$ and $z_i^*(0) = 0.25$ $(1 \le i \le 9)$. The number 0.3798 is the limit of the output y^* for a constant input of 0.25. The target (79) proposed by Atiya et al. [20] is generally called NARMA10 and is widely used as a benchmark task for time series predictions.

For the training process, we generate an independent and identically distributed input sequence $u_d(t)$ $(t \in \mathbb{Z}_+)$ sampled from the uniform distribution on $[0, 0.5]$, where $\mathbb{Z}_+ = \{0, 1, 2, \ldots\}$. We define the continuous-time input $u : \mathbb{R}_+ \to [0, 0.5]$ to the reservoir as follows:

$$\begin{aligned}
u(t) &= \begin{cases} u_d(0) & (0 \le t < 1), \\ u_d(\lfloor t \rfloor) + \alpha(t) & (1 \le t), \end{cases} \\
\alpha(t) &= (u_d(\lfloor t \rfloor - 1) - u_d(\lfloor t \rfloor)) \exp\left(10\frac{t - \lfloor t \rfloor}{t - \lfloor t \rfloor - 1}\right),
\end{aligned} \tag{80}$$

where $\lfloor \cdot \rfloor$ is the floor function. As shown in Fig. 3, the continuous-time input u is a step function-like form defined by the discrete-time input u_d. We do not use the step function itself because it is not continuous. Let $\hat{y} : \mathbb{R}_+ \to \mathbb{R}$ and $y^* : \mathbb{Z}_+ \to \mathbb{R}$ be the output of the RC model and target for the input u and u_d, respectively. We train the readout p to minimize the following loss function:

$$L(p, u_d) = \frac{1}{N}\sum_{t=0}^{N} \left(\hat{y}(t) - y(t)^*\right)^2, \tag{81}$$

where $N = 50000$. We test the trained RC model using another input u_d generated in the same way as that used in the training process. The inputs, reservoir states, and outputs in the test are shown in Figs. 3–5, respectively. The value of the loss function L is 2.5×10^{-4} in the test. From Fig. 5, the value of the loss function L is small compared with the range of outputs and so the approximation is successful.

IV. Conclusion

For RC with a polynomial readout, we have guaranteed the universality of a uniform approximation of an operator with fading memory in the case where the reservoir is represented by a set of RIT operators without fading memory. To achieve this, we converted the RIT operators into functionals and applied the extended Stone–Weierstrass theorem. For the application of the Stone–Weierstrass theorem, we defined a metric on the functional domain, which is a set of inputs defined on time intervals of various lengths. We have proposed a novel sufficient reservoir condition, NSP, which requires neither the fading memory of a reservoir nor the continuity of the reservoir output with respect to inputs and time. Therefore, our result can be applied to very common reservoirs, such as nonlinear time-variant state-space systems with multiple equilibria and a discontinuous output function. This is particularly significant in the field of physical RC, in which reservoirs are not easily adjustable. We presented an example of a physical reservoir without fading memory from which the RC model successfully approximates NARMA10, a benchmark task for time series prediction.

Appendix A
Proof of Proposition 2

Let v_1, v_2, and v_3 be arbitrary elements of V^{res}. Clearly d satisfies

$$d(v_1, v_2) = d(v_2, v_1). \tag{82}$$

Therefore, without loss of generality, we can assume that times $t_1 = \lambda(v_1)$, $t_2 = \lambda(v_2)$, $t_3 = \lambda(v_3) \in \mathbb{R}_+$ satisfy $t_1 \le t_2 \le t_3$.

We prove that d satisfies

$$d(v_1, v_2) = 0 \Leftrightarrow v_1 = v_2. \tag{83}$$

In (83), it is clear that the left-hand side is obtained from the right-hand side. We prove that the right-hand side is obtained from the left-hand side by considering the contraposition. If $v_1 \ne v_2$, either $t_1 \ne t_2$ or $\exists \tau \in [-t_1, 0]$ such that $v_1(\tau) \ne v_2(\tau)$, or both. If $t_1 \ne t_2$ holds, then because θ is strictly increasing, we have

$$d(v_1, v_2) \ge \theta(t_2) - \theta(t_1) > 0. \tag{84}$$

If there is some $\tau \in [-t_1, 0]$ satisfying $v_1(\tau) \ne v_2(\tau)$, we have

$$d(v_1, v_2) \ge \left\| v_1 - v_2^{[t_1]} \right\|_w > 0. \tag{85}$$

Therefore, d satisfies (83).

We prove that d satisfies the following triangle inequalities:

$$d(v_1, v_2) + d(v_2, v_3) \ge d(v_3, v_1), \tag{86a}$$

$$d(v_2, v_3) + d(v_3, v_1) \ge d(v_1, v_2), \tag{86b}$$

$$d(v_3, v_1) + d(v_1, v_2) \ge d(v_2, v_3). \tag{86c}$$

The weighted norm $\|\cdot\|_w$ satisfies

$$\left\| v_2 - v_3^{[t_2]} \right\|_w \ge \left\| v_2^{[t_1]} - v_3^{[t_1]} \right\|_w. \tag{87}$$

Therefore, (86a) is proved because

$$\begin{aligned}
& d(v_1, v_2) + d(v_2, v_3) \\
&= \left\| v_1 - v_2^{[t_1]} \right\|_w + \theta(t_2) - \theta(t_1) \\
&\quad + \left\| v_2 - v_3^{[t_2]} \right\|_w + \theta(t_3) - \theta(t_2) \\
&\ge \left\| v_1 - v_2^{[t_1]} \right\|_w + \left\| v_2^{[t_1]} - v_3^{[t_1]} \right\|_w + \theta(t_3) - \theta(t_1) \\
&\ge \left\| v_3^{[t_1]} - v_1 \right\|_w + \theta(t_3) - \theta(t_1) = d(v_3, v_1).
\end{aligned} \tag{88}$$

Inequality (86b) is proved in the same way as (86a). From the condition (iii) of Assumption 1, (86c) is proved because

$$\begin{aligned}&d(v_3,v_1)+d(v_1,v_2)\\&=\left\|v_3^{[t_1]}-v_1\right\|_w+\theta(t_3)-\theta(t_1)\\&\quad+\left\|v_1-v_2^{[t_1]}\right\|_w+\theta(t_2)-\theta(t_1)\\&\geq\left\|v_2^{[t_1]}-v_3^{[t_1]}\right\|_w+2\left[\theta(t_2)-\theta(t_1)\right]+\theta(t_3)-\theta(t_2)\\&\geq\left\|v_2-v_3^{[t_2]}\right\|_w+\theta(t_3)-\theta(t_2)=d(v_2,v_3).\end{aligned}\tag{89}$$

From (82), (83), and (86), the map d is a metric on V^{res}.

Appendix B
Proof of Proposition 3

Propositions (31a) and (31b) can be proved in the same way. Hence, we only prove (31a) here. To prove (31a), we need to show that

$$\begin{aligned}&\forall\varepsilon>0,\forall v_1\in V^{\mathrm{res}},\exists\delta>0,\forall v_2\in V^{\mathrm{res}},\\&\left\|v_1^{[t_{\min}]}-v_2^{[t_{\min}]}\right\|_w<\delta\Rightarrow\left\|v_1^{[t_{\min}]}-v_2^{[t_{\min}]}\right\|_{w'}<\frac{\varepsilon}{2},\end{aligned}\tag{90}$$

and

$$\begin{aligned}&\forall\varepsilon>0,\forall v_1\in V^{\mathrm{res}},\exists\delta>0,\forall v_2\in V^{\mathrm{res}},\\&\left|\theta(t_1)-\theta(t_2)\right|<\delta\Rightarrow\left|\theta'(t_1)-\theta'(t_2)\right|<\frac{\varepsilon}{2},\end{aligned}\tag{91}$$

where $t_1=\lambda(v_1)$, $t_2=\lambda(v_2)$, and $t_{\min}=\min\{t_1,t_2\}$.

First, we prove (90). Because $\lim_{t\to\infty}w'(t)=0$, for any $\varepsilon>0$, there is some $T\geq 0$ satisfying $Mw'(T)<\varepsilon/2$, where $M\geq 0$ is defined in (33). We define a real number $R=w'(0)/w(T)$. Let v_2 be an arbitrary element of V^{res}, and let $v=v_1^{[t_{\min}]}-v_2^{[t_{\min}]}$. Suppose $t_{\min}\leq T$. Because w and w' are decreasing functions, $w(T)\leq w(-\tau)$ and $w'(-\tau)\leq w'(0)$ holds for any $\tau\in[-t_{\min},0]$. Hence, we have

$$\begin{aligned}\|v\|_{w'}&\leq\sup_{\tau\in[-t_{\min},0]}v(\tau)w'(0)\\&=R\sup_{\tau\in[-t_{\min},0]}v(\tau)w(T)\\&\leq R\|v\|_w.\end{aligned}\tag{92}$$

Therefore, $\delta<\varepsilon/(2R)$ satisfies (90).

Suppose $t_{\min}>T$. From the definition of the norm $\|\cdot\|_{w'}$, we have

$$\|v\|_{w'}=\max\left\{\left\|v^{[T]}\right\|_{w'},\sup_{\tau\in[-t_{\min},-T]}\|v(\tau)\|\,w'(-\tau)\right\}.\tag{93}$$

In the same way as (92), the following holds for the first term inside the max operator of (93):

$$\left\|v^{[T]}\right\|_{w'}\leq R\left\|v^{[T]}\right\|_w\leq R\|v\|_w.\tag{94}$$

From the definition of M and T, the following holds for the second term inside the max operator of (93):

$$\sup_{\tau\in[-t_{\min},-T]}\|v(\tau)\|\,w'(-\tau)\leq Mw'(T)<\frac{\varepsilon}{2}.\tag{95}$$

Therefore, $\delta<\varepsilon/(2R)$ satisfies (90).

Next, we prove (91). We define $\theta'(\infty)=\lim_{t\to\infty}\theta'(t)$. Because the function $\theta:\mathbb{R}_+\to[0,\theta(\infty))$ is continuous and strictly increasing, it has the continuous inverse $\theta^{-1}:[0,\theta(\infty))\to\mathbb{R}_+$. Because the function $\theta':\mathbb{R}_+\to[0,\theta'(\infty))$ is also continuous, the composition $\theta'\circ\theta^{-1}:[0,\theta(\infty))\to[0,\theta'(\infty))$ is continuous, i.e.,

$$\begin{aligned}&\forall\varepsilon>0,\forall\alpha_1\in[0,\theta(\infty)),\exists\delta>0,\forall\alpha_2\in[0,\theta(\infty)),\\&|\alpha_1-\alpha_2|<\delta\Rightarrow\left|\theta'\circ\theta^{-1}(\alpha_1)-\theta'\circ\theta^{-1}(\alpha_2)\right|<\frac{\varepsilon}{2}.\end{aligned}\tag{96}$$

Considering $\alpha_1=\theta(t_1)$ and $\alpha_2=\theta(t_2)$, (91) holds.

From (90) and (91), we have

$$\begin{aligned}&d(v_1,v_2)=\left\|v_1^{[t_{\min}]}-v_2^{[t_{\min}]}\right\|_w+|\theta(t_1)-\theta(t_2)|<\delta\\&\Rightarrow d'(v_1,v_2)=\left\|v_1^{[t_{\min}]}-v_2^{[t_{\min}]}\right\|_{w'}+|\theta'(t_1)-\theta'(t_2)|<\varepsilon,\end{aligned}\tag{97}$$

which proves Proposition 3.

Appendix C
Proof of Proposition 5

We prove that any infinite sequence $(v_i)_{i\in\mathbb{N}}\in V^{\mathrm{res}}\cup V$ contains a subsequence converging on $V^{\mathrm{res}}\cup V$. We define $\mathbb{N}_{a1}$ and $\mathbb{N}_{a2}\subset\mathbb{N}$ as

$$\mathbb{N}_{a1}=\{i\in\mathbb{N}\,|v_i\in V\},\quad\mathbb{N}_{a2}=\{i\in\mathbb{N}\,|v_i\in V^{\mathrm{res}}\}.\tag{98}$$

At least one of the subsequences $(v_i)_{i\in\mathbb{N}_{a1}}$ and $(v_i)_{i\in\mathbb{N}_{a2}}$ of $(v_i)_{i\in\mathbb{N}}$ is an infinite sequence. If the sequence $(v_i)_{i\in\mathbb{N}_{a1}}\in V$ is infinite, there is a subsequence of $(v_i)_{i\in\mathbb{N}_{a1}}$ converging on V because V is compact.

Suppose that the subsequence $(v_i)_{i\in\mathbb{N}_{a2}}\in V^{\mathrm{res}}$ is infinite. We define the sequence $(t_i)_{i\in\mathbb{N}_{a2}}\in\mathbb{R}_+$ as $t_i=\lambda(v_i)$ and the sequence $(\tilde{v}_i)_{i\in\mathbb{N}_{a2}}\in V$ as

$$\tilde{v}_i(\tau)=\begin{cases}v_i(\tau) & (-t_i\leq\tau\leq 0),\\ v_i(-t_i) & (\tau<-t_i).\end{cases}\tag{99}$$

Because the set V is compact, there exist $\mathbb{N}_b\subset\mathbb{N}_{a2}$ and $\tilde{v}\in V$ that satisfy

$$\lim_{i\to\infty,i\in\mathbb{N}_b}d(\tilde{v},\tilde{v}_i)=0.\tag{100}$$

If $(t_i)_{i\in\mathbb{N}_b}$ is not bounded, we have

$$\exists\mathbb{N}_{c1}\subset\mathbb{N}_b,\lim_{i\to\infty,i\in\mathbb{N}_{c1}}t_i=\infty.\tag{101}$$

If $(t_i)_{i\in\mathbb{N}_b}$ is bounded, we have

$$\exists\mathbb{N}_{c2}\subset\mathbb{N}_b,\exists t\geq 0,\lim_{i\to\infty,i\in\mathbb{N}_{c2}}t_i=t.\tag{102}$$

We prove that if (101) holds, the subsequence $(v_i)_{i\in\mathbb{N}_{c1}}\in V^{\mathrm{res}}$ converges to $\tilde{v}\in V$. From the definition of $\tilde{v}_i$, $v_i(\tau)=\tilde{v}_i(\tau)$ $(\tau\in[-t_i,0])$ holds. Hence, for any $t'\leq t_i$,

$$\left\|\tilde{v}^{[t']}-v_i^{[t']}\right\|_w=\left\|\tilde{v}^{[t']}-\tilde{v}_i^{[t']}\right\|_w\leq\|\tilde{v}-\tilde{v}_i\|_w.\tag{103}$$

From (103), the distance $d(\tilde{v},v_i)$ satisfies

$$\begin{aligned}d(\tilde{v},v_i)&=\left\|\tilde{v}^{[t_i]}-v_i\right\|_w+\theta(\infty)-\theta(t_i)\\&\leq\|\tilde{v}-\tilde{v}_i\|_w+\theta(\infty)-\theta(t_i)\\&=d(\tilde{v},\tilde{v}_i)+\theta(\infty)-\theta(t_i).\end{aligned}\tag{104}$$

If $i \to \infty$ and $i \in \mathbb{N}_{c1}$, $\tilde{v}_i \to \tilde{v}$ and $t_i \to \infty$ hold. Therefore, $d(\tilde{v}, v_i)$ converges to 0.

We prove that if (102) holds, the subsequence $(v_i)_{i\in\mathbb{N}_{c2}} \in V^{\mathrm{res}}$ converges to $\tilde{v}^{[t]} \in V^{\mathrm{res}}$. From (103), the distance $d\left(\tilde{v}^{[t]}, v_i\right)$ satisfies the following:

$$\begin{aligned} d\left(\tilde{v}^{[t]}, v_i\right) &= \left\|\tilde{v}^{[t_{\min}]} - v_i^{[t_{\min}]}\right\|_w + |\theta(t) - \theta(t_i)| \\ &\le \|\tilde{v} - \tilde{v}_i\|_w + |\theta(t) - \theta(t_i)| \\ &= d(\tilde{v}, \tilde{v}_i) + |\theta(t) - \theta(t_i)|, \end{aligned} \tag{105}$$

where $t_{\min} = \min\{t, t_i\}$. If $i \to \infty$ and $i \in \mathbb{N}_{c2}$, then $\tilde{v}_i \to \tilde{v}$ and $t_i \to t$ hold. Therefore, $d\left(\tilde{v}^{[t]}, v_i\right)$ converges to 0.

From the above, any infinite sequence $(v_i)_{i\in\mathbb{N}} \in V^{\mathrm{res}} \cup V$ contains a subsequence converging on $V^{\mathrm{res}} \cup V$, which proves Proposition 5.

Appendix D
Proof of Proposition 6

Let v_1 be an arbitrary element of V^{res}. We prove that f is continuous at v_1. Let ε be an arbitrary positive number. From condition (ii) of Assumption 2, we have

$$\begin{aligned} &\exists \delta_1 > 0, \forall v_2 \in V^{\mathrm{res}}, \\ &|t_1 - t_2| < \delta_1 \Rightarrow |Fu_2(t_1) - Fu_2(t_2)| < \frac{\varepsilon}{2}, \end{aligned} \tag{106}$$

where $u_2 = \sigma(v_2)$, $t_1 = \lambda(v_1)$, and $t_2 = \lambda(v_2)$. From condition (i) of Assumption 2, we have

$$\begin{aligned} &\exists \delta_2 > 0, \forall v_2 \in V^{\mathrm{res}}, \\ &\max_{\tau\in[0,t_1]} \|u_1(\tau) - u_2(\tau)\| < \delta_2 \Rightarrow |Fu_1(t_1) - Fu_2(t_1)| < \frac{\varepsilon}{2}, \end{aligned} \tag{107}$$

where $u_1 = \sigma(v_1)$. Because the Lipschitz constants of u_1 and u_2 are K or less, for any $u_2 \in U$, we have

$$\begin{aligned} &\max_{\tau\in[0,t_1]} \|u_1(\tau) - u_2(\tau)\| \\ \le &\max_{\tau\in[-t_{\min},0]} \|u_1(\tau + t_1) - u_2(\tau + t_2)\| + K|t_1 - t_2| \\ \le &\frac{1}{w(t_1)} \left\|v_1^{[t_{\min}]} - v_2^{[t_{\min}]}\right\|_w + K|t_1 - t_2|, \end{aligned} \tag{108}$$

where $t_{\min} = \min\{t_1, t_2\}$. Because the function θ is strictly increasing and continuous, we have

$$\begin{aligned} &\exists \delta_3 > 0, \forall t_2 \ge 0, \\ &|\theta(t_1) - \theta(t_2)| < \delta_3 \Rightarrow |t_1 - t_2| < \min\left\{\delta_1, \frac{\delta_2}{2K}\right\}. \end{aligned} \tag{109}$$

Let us define $\delta > 0$ as

$$\delta = \min\left\{\frac{w(t_1)}{2}\delta_2, \delta_3\right\}. \tag{110}$$

Then, from (108) and (109), we have

$$\begin{aligned} &d(v_1, v_2) < \delta \\ &\Rightarrow \left(|t_1 - t_2| < \delta_1, \max_{\tau\in[0,t_1]} \|u_1(\tau) - u_2(\tau)\| < \delta_2\right). \end{aligned} \tag{111}$$

Therefore, from (106) and (107), we have

$$d(v_1, v_2) < \delta \Rightarrow |Fu_1(t_1) - Fu_2(t_2)| < \varepsilon. \tag{112}$$

Because $Fu_1(t_1) = f(v_1)$ and $Fu_2(t_2) = f(v_2)$, f is continuous at v_1, which proves Proposition 6.

Appendix E
Proof of Proposition 7

The fading memory of F is equivalent to

$$\begin{aligned} &\forall \varepsilon > 0, \exists \delta > 0, \exists T \ge 0, \forall v_1, v_2 \in V^{\mathrm{res}}, \\ &\left(\lambda(v_1), \lambda(v_2) \ge T, \max_{\tau\in[-T,0]} \|v_1(\tau) - v_2(\tau)\| < \delta\right) \\ &\Rightarrow |f(v_1) - f(v_2)| < \varepsilon. \end{aligned} \tag{113}$$

We use the following lemma:

Lemma 7. *Any $v \in V$ and sequences $(v_i)_{i\in\mathbb{N}}$, $(v'_i)_{i\in\mathbb{N}} \in V^{\mathrm{res}}$ converging to v satisfy*

$$\lim_{i\to\infty} |f(v_i) - f(v'_i)| = 0. \tag{114}$$

Proof of Lemma 7. Let the sequences $(t_i)_{i\in\mathbb{N}}$ and $(t'_i)_{i\in\mathbb{N}} \in \mathbb{R}_+$ be defined as $t_i = \lambda(v_i)$ and $t'_i = \lambda(v'_i)$, respectively. Let ε be an arbitrary positive number. Because F has fading memory, (113) implies

$$\begin{aligned} &\exists \delta > 0, \exists T \ge 0, \forall i \in \mathbb{N}, \\ &\left(t_i, t'_i \ge T, \max_{\tau\in[-T,0]} \|v_i(\tau) - v'_i(\tau)\| < \delta\right) \\ &\Rightarrow |f(v_i) - f(v'_i)| < \varepsilon. \end{aligned} \tag{115}$$

Because $\theta(\infty) - \theta(t_i) \le d(v, v_i) \to 0$ as $i \to \infty$, we have $t_i \to \infty$ as $i \to \infty$. In the same way, we have $t'_i \to \infty$ as $i \to \infty$. Hence, we have

$$\exists k \in \mathbb{N}, \forall i > k, t_i, t'_i \ge T. \tag{116}$$

Because $\left\|v^{[t_i]} - v_i\right\|_w < d(v, v_i) \to 0$ as $i \to \infty$, we have the following as $i \to \infty$ with $i > k$:

$$\begin{aligned} \max_{\tau\in[-T,0]} \|v(\tau) - v_i(\tau)\| &\le \frac{1}{w(T)} \left\|v^{[T]} - v_i^{[T]}\right\|_w \\ &\le \frac{1}{w(T)} \left\|v^{[t_i]} - v_i\right\|_w \to 0. \end{aligned} \tag{117}$$

In the same way, we have

$$\max_{\tau\in[-T,0]} \|v(\tau) - v'_i(\tau)\| \to 0 \qquad (i > k, i \to \infty). \tag{118}$$

From (117), (118), and the triangle inequality, we have

$$\max_{\tau\in[-T,0]} \|v_i(\tau) - v'_i(\tau)\| \to 0 \qquad (i > k, i \to \infty). \tag{119}$$

From (115), (116), and (119), we have

$$\forall \varepsilon > 0, \exists k \in \mathbb{N}, \forall i > k, |f(v_i) - f(v'_i)| < \varepsilon, \tag{120}$$

which proves Lemma 7. $\square$

To prove Proposition 7 using Lemma 7, we first prove that the right-hand side of (42) converges. We have $\lambda\left(v^{[i]}\right) = i$ and

$$\forall T \ge 0, \forall i, j \ge T, \max_{\tau\in[-T,0]} \left\|v^{[i]}(\tau) - v^{[j]}(\tau)\right\| = 0. \tag{121}$$

Because F has fading memory, (113) leads to

$$\forall \varepsilon > 0, \exists T \ge 0, \forall i, j \ge T, \left|f\left(v^{[i]}\right) - f\left(v^{[j]}\right)\right| < \varepsilon. \tag{122}$$

Therefore, $\left(f\left(v^{[i]}\right)\right)_{i\in\mathbb{N}}$ is a Cauchy sequence in $\mathbb{R}$ and converges.

Next, we prove that if the value of f on V is defined as (42), f is continuous on V. Let v be an arbitrary element of V and $(v_i)_{i\in\mathbb{N}} \in V^{\mathrm{res}} \cup V$ be an arbitrary sequence converging to v. If $v_i \in V$, there is some $t_i \in \mathbb{R}_+$ satisfying

$$d\Big(v_i, v_i^{[t_i]}\Big) < 2^{-i}, \quad \Big|f(v_i) - f\Big(v_i^{[t_i]}\Big)\Big| < 2^{-i}. \tag{123}$$

Let us define a sequence $(v_i')_{i\in\mathbb{N}} \in V^{\mathrm{res}}$ as

$$v_i' = \begin{cases} v_i & (v_i \in V^{\mathrm{res}}), \\ v_i^{[t_i]} & (v_i \in V). \end{cases} \tag{124}$$

We have the following as $i \to \infty$:

$$d(v, v_i') \leq d(v, v_i) + d(v_i, v_i') < d(v, v_i) + 2^{-i} \to 0, \tag{125}$$

$$|f(v_i) - f(v_i')| < 2^{-i} \to 0. \tag{126}$$

From (125), $v_i' \in V^{\mathrm{res}}$ converges to v like $v^{[i]} \in V^{\mathrm{res}}$. Hence From Lemma 7 and (42), $f(v_i') \to f(v)$ holds. Therefore, from (126), $f(v_i) \to f(v)$ holds, which proves Proposition 7.

Appendix F
Proof of Proposition 8

From Propositions 6 and 7, if F satisfies Assumption 2, the value of f on V can be defined as (42) and f is continuous on the compact set $V^{\mathrm{res}} \cup V$. Hence, the functional f is uniformly continuous on $V^{\mathrm{res}} \cup V$.

We now prove that if f is uniformly continuous on V^{res}, F satisfies Assumption 2. Let ε be an arbitrary positive number. There is some $\gamma > 0$ satisfying

$$\forall v_1, v_2 \in V^{\mathrm{res}}, d(v_1, v_2) < \gamma \Rightarrow |f(v_1) - f(v_2)| < \varepsilon. \tag{127}$$

First, we prove condition (i) of Assumption 2. Let u_1 be an arbitrary element of U and $t \geq 0$ be an arbitrary time. Because w is a map to $(0, 1]$, for any $u_2 \in U$, we have

$$d(v_1, v_2) \leq \max_{\tau\in[0,t]} \|u_1(\tau) - u_2(\tau)\|, \tag{128}$$

where $v_1 = \bar{S}(u_1, t)$ and $v_2 = \bar{S}(u_2, t)$. Therefore, from $f(v_1) = Fu_1(t)$, $f(v_2) = Fu_2(t)$, and (127), condition (i) holds.

Next, we prove condition (ii) of Assumption 2. Let $t_1 \geq 0$ be an arbitrary time. Because θ is continuous, there is some $\delta > 0$ satisfying

$$\forall t_2 \geq 0, |t_1 - t_2| < \delta \Rightarrow K|t_1 - t_2| + |\theta(t_1) - \theta(t_2)| < \gamma. \tag{129}$$

As the Lipschitz constant of elements in U is K or less, for any $u \in U$ and $t_2 \geq 0$ satisfying $|t_1 - t_2| < \delta$, we have

$$\begin{aligned} d(v_1, v_2) =& \sup_{\tau\in[-t_{\min},0]} \|u(\tau + t_1) - u(\tau + t_2)\| w(-\tau) \\ &+ |\theta(t_1) - \theta(t_2)| \\ \leq& K|t_1 - t_2| + |\theta(t_1) - \theta(t_2)| < \gamma, \end{aligned} \tag{130}$$

where $v_1 = \bar{S}(u, t_1)$, $v_2 = \bar{S}(u, t_2)$, and $t_{\min} = \min\{t_1, t_2\}$. Therefore, from $f(v_1) = Fu(t_1)$, $f(v_2) = Fu(t_2)$, and (127), condition (ii) holds.

Finally, we prove condition (iii) of Assumption 2. Let $\delta = \gamma/3$. Let $T \geq 0$ satisfy $|\theta(\infty) - \theta(T)| < \gamma/3$. Suppose that $u_1,\, u_2 \in U$ and $t_1,\, t_2 \geq T$ satisfy

$$\max_{\tau\in[-T,0]} \|u_1(\tau + t_1) - u_2(\tau + t_2)\| < \delta. \tag{131}$$

Then, from condition (iii) of Assumption 1, we have

$$\begin{aligned} d(v_1, v_2) &= \left\|v_1^{[t_{\min}]} - v_2^{[t_{\min}]}\right\|_w + [\theta(t_{\max}) - \theta(t_{\min})] \\ &\leq \left\|v_1^{[T]} - v_2^{[T]}\right\|_w + 2[\theta(t_{\min}) - \theta(T)] \\ &\quad + [\theta(t_{\max}) - \theta(t_{\min})] \\ &< \max_{\tau\in[-T,0]} \|u_1(\tau + t_1) - u_2(\tau + t_2)\| \\ &\quad + 2[\theta(\infty) - \theta(T)] < \gamma, \end{aligned} \tag{132}$$

where $v_1 = \bar{S}(u_1, t_1)$, $v_2 = \bar{S}(u_2, t_2)$, $t_{\max} = \max\{t_1, t_2\}$, and $t_{\min} = \min\{t_1, t_2\}$. Therefore, from $f(v_1) = Fu_1(t_1)$, $f(v_2) = Fu_2(t_2)$, and (127), condition (iii) holds, which proves Proposition 8.

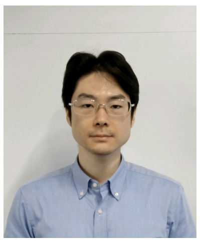
Acknowledgment

We thank Stuart Jenkinson, PhD, from Edanz (https://jp.edanz.com/ac) for editing a draft of this manuscript.

References

[1] R.J. Williams and D. Zipser, "A learning algorithm for continually running fully recurrent neural networks," *Neural Computation*, vol. 1, no. 2, pp. 270–280, 1989.
[2] P.J. Werbos, "Backpropagation through time: what it does and how to do it," *Proceedings of the IEEE*, vol. 78, no. 10, pp. 1550–1560, 1990.
[3] H. Jaeger, "The "echo state" approach to analysing and training recurrent neural networks – with an erratum note," *German National Research Center for Information Technology GMD Technical Report*, 148.34, 2001.
[4] W. Maass, T. Natschläger and H. Markram, "Real–time computing without stable states: a new framework for neural computation based on perturbations," *Neural Computation*, vol. 14, no. 11, pp. 2531–2560, 2002.
[5] J.J. Steil, "Backpropagation-decorrelation: online recurrent learning with O (N) complexity," *In 2004 IEEE international joint conference on neural networks* vol. 2, pp. 843–848, 2004.
[6] D. Verstraeten, B. Schrauwen, M. D'Haene and D. Stroobandt, "An experimental unification of reservoir computing methods," *Neural Networks*, vol. 20, no. 3, pp. 391–403, 2007.
[7] M. Lukoševičius and H. Jaeger, "Reservoir computing approaches to recurrent neural network training," *Computer Science Review*, vol. 3, no. 3, pp. 127–149, 2009.
[8] G. Tanaka et al., "Recent advances in physical reservoir computing: A review," *Neural Networks*, vol. 115, pp. 100–123, 2019.
[9] J.S. Friedman, "Unsupervised learning & reservoir computing leveraging analog spintronic phenomena," *IEEE 16th Nanotechnology Materials and Devices Conference*, pp. 1–2, 2021.
[10] F. Stelzer, A. Röhm, K. Lüdge and S. Yanchuk, "Performance boost of time-delay reservoir computing by non-resonant clock cycle," *Neural Networks*, vol. 124, pp. 158–169, 2020.
[11] M.C. Soriano et al., "Delay-based reservoir computing: noise effects in a combined analog and digital implementation," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 26, no. 2, pp. 388–393, 2014.
[12] J. Dong, M. Rafayelyan, F. Krzakala and S. Gigan, "Optical reservoir computing using multiple light scattering for chaotic systems prediction," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 26, no. 1, pp. 1–12, 2020.
[13] K. Nakajima and I. Fischer, *Reservoir Computing*, Springer, 2021.
[14] L. Grigoryeva and J.P. Ortega, "Echo state networks are universal," *Neural Networks*, vol. 108, pp. 495–508, 2018.
[15] L. Gonon and J.P. Ortega, "Reservoir Computing Universality With Stochastic Inputs," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 1, pp. 100–112, 2020.

[16] S. Boyd and L.O. Chua, "Fading memory and the problem of approximating nonlinear operators with Volterra series," *IEEE Trans. on Circuits and Systems*, vol. 32, no. 11, pp. 1150–1161, 1985.
[17] I.W. Sandberg, "Notes on fading-memory conditions," *Circuits, Systems and Signal Processing*, vol. 22, no. 1, pp. 43–55, 2003.
[18] M.B. Matthews, "Approximating nonlinear fading-memory operators using neural network models," *Circuits, Systems and Signal Processing*, vol. 12, no. 2, pp. 279–307, 1993.
[19] J. Dieudonne, *Foundations of Modern Analysis*, Academic Press, 1969.
[20] A.F. Atiya and A.G. Parlos, "New Results on Recurrent Network Training: Unifying the Algorithms and Accelerating Convergence," *IEEE Transactions on Neural Networks*, vol. 11, no. 3, pp. 697–709, 2000.

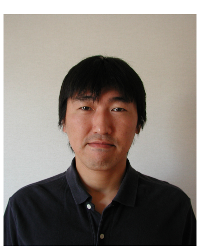


Shuhei Sugiura received his B.E. and M.E. degrees from Nagoya University, Nagoya, Japan, in 2017 and 2020, respectively. He is currently pursuing his Ph.D. degree at Nagoya University.

Ryo Ariizumi (Member, IEEE) received the B.E., M.E., and Ph.D. degrees from Kyoto University, Kyoto, Japan, in 2010, 2012, and 2015, respectively.

He was a Research Fellow with the Japan Society for the Promotion of Science working at Kyoto University from 2014 to 2015, and an Assistant Professor at Nagoya University, Nagoya, Japan, from 2015 to 2023. He is currently an Associate Professor at the Tokyo University of Agriculture and Technology, Tokyo, Japan. His research interests include the control of redundant robots and the optimization of robotic systems.

Dr. Ariizumi has received awards, including the IEEE Robotics and Automation Society Japan Chapter Young Award (IROS2014) in 2014 and the Best Paper Award from the Robotics Society of Japan (RSJ) in 2018.



Toru Asai (Member, IEEE) received his B.E., M.E., and Ph.D. degrees from Tokyo Institute of Technology, Tokyo, Japan, in 1991, 1993, and 1996, respectively.

He worked as a Research Fellow of JSPS between 1996 and 1998. In 1999, he joined the faculty of Osaka University. He is an Associate Professor of the sub-department of Mechatronics between 2015 and 2016 and the Department of Mechanical Systems Engineering since 2017, Nagoya University. His research interests include robust control, switching control, parameter estimation, and industrial applications.

Shun-ichi Azuma (Senior Member, IEEE) received his B.S. from Hiroshima University in 1999 and his M.S. and Ph.D. from the Tokyo Institute of Technology in 2001 and 2004, respectively.

He was a research fellow of the Japan Society for the Promotion of Science from 2004 to 2005, an assistant professor and associate professor at the Graduate School of Informatics, Kyoto University from 2005 to 2011 and from 2011 to 2017, and a professor at the Graduate School of Engineering, Nagoya University from 2017 to 2022. He is currently a professor at the Graduate School of Informatics, Kyoto University.

He served as Associate Editor of IEEE Transactions on Control of Network Systems from 2013 to 2019 and IEEE Control Systems Letters from 2016 to 2019. He currently serves as Associate Editors of IEEE CSS Conference Editorial Board from 2011, Automatica from 2014, Nonlinear Analysis: Hybrid Systems from 2017, International Journal of Control, Automation and Systems from 2018, and IEEE Transactions on Automatic Control from 2019. His research interests include analysis and control of network systems and hybrid systems.