

Reservoir Computing with Applications to Time Series Forecasting

Lyudmila Grigoryeva

University of Konstanz, Germany

Summer School
University of Vienna
Vienna, 2020

Universality

- Grigoryeva, L. and Ortega, J.-P. 2018.
Universal discrete-time reservoir computers with stochastic inputs and linear readouts using non-homogeneous state-affine systems.
Journal of Machine Learning Research, 19(24), 1-40.
- Grigoryeva, L. and Ortega, J.-P. 2018.
Echo state networks are universal.
Neural Networks, 108, 495-508.
- Gonon, L. and Ortega, J.-P. 2019.
Reservoir computing universality with stochastic inputs.
IEEE Transactions on Neural Networks and Learning Systems, 31(1), 100-112.

Generalization and approximation errors for RC

- Gonon, L., Grigoryeva, L., and Ortega, J.-P. 2019.
Risk bounds for reservoir computing. *Preprint*
- Gonon, L., Grigoryeva, L., and Ortega, J.-P. 2020.
Approximation bounds for random neural networks and reservoir systems.
Preprint

Properties of Reservoir Computing

- Grigoryeva, L. and Ortega, J.-P. 2019.
Differentiable reservoir computing.
Journal of Machine Learning Research, 20(179), 1-62.
- Gonon, L., Grigoryeva, L., and Ortega, J.-P. 2020.
Memory and forecasting capacities of nonlinear recurrent networks.
To appear in *Physica D*.

Link to signature processes, dimensionality reduction, applications

- Cuchiero, C., Gonon, L., Grigoryeva, L., Ortega, J.-P., and Teichmann, J. 2020. **Discrete-time signatures and randomness in reservoir computing.** *Preprint* will be made available in mid September
- Grigoryeva, L. and Ortega, J.-P. 2020. **Dimension reduction in recurrent networks by canonicalization.** *Preprint*
- Gonon, L., Grigoryeva, L., Kukharenko, O., and Ortega, J.-P. **Forecasting realized variances with reservoir computing.** *Preprint* will be made available by the end of October

Input/Output systems

Definition

Let $V_{\mathcal{Z}} \subset \mathcal{Z}^{\mathcal{T}}$ and $V_{\mathcal{Y}} \subset \mathcal{Y}^{\mathcal{T}}$ be two subsets of the spaces of sequences $\mathcal{Z}^{\mathcal{T}}$ and $\mathcal{Y}^{\mathcal{T}}$ with values in the sets $\mathcal{Z}$ and $\mathcal{Y}$, respectively, indexed by the elements in $\mathcal{T} \subset \mathbb{R}$. An **Input/Output (IO)** system is a map $U : V_{\mathcal{Z}} \subset \mathcal{Z}^{\mathcal{T}} \longrightarrow V_{\mathcal{Y}} \subset \mathcal{Y}^{\mathcal{T}}$, where $\mathcal{Z}$ is called the **input space**, $\mathcal{Y}$ is the **output space**, and $\mathcal{T} \subset \mathbb{R}$ is the **temporal index**.

We take $\mathcal{T}$ as a discrete set and adopt discrete-time point of view.

If $\mathcal{T}$ is a compact subset of $\mathbb{R}$, we say that we are in a **finite-time** setup. IO systems for which $\mathcal{T}$ is semi-infinite towards $-\infty$ ($\mathcal{T} = \mathbb{Z}_{-}, \mathbb{Z}, \mathbb{R}_{-}, \mathbb{R}$) will be generically referred to as **filters**, while when $\mathcal{T}$ is semi-infinite exclusively towards $+\infty$ ($\mathcal{T} = \mathbb{Z}_{+}, \mathbb{R}_{+}$) we will talk about **controlled dynamical systems**. Our work takes place for discrete-time filters. IO systems that are not finite-memory will be referred to as infinite-memory. We work with a rich class of infinite-memory systems called fading memory, which is of much relevance in the modeling of many phenomena and for which recent inputs are more relevant than those far in the past. We shall see that infinite-memory filters in that class can be uniformly approximated by finite-memory ones.

Example. A linear IO system

Filters and functionals

Filters $U : V_z \subset (\mathbb{R}^d)^{\mathbb{Z}} \longrightarrow V_y \subset (\mathbb{R}^m)^{\mathbb{Z}}$ and functionals $H : (\mathbb{R}^d)^{\mathbb{Z}} \longrightarrow \mathbb{R}^m$:

- **Causal filter:** for any two elements $\mathbf{z}^1, \mathbf{z}^2 \in (\mathbb{R}^d)^{\mathbb{Z}}$ that satisfy that $\mathbf{z}^1_\tau = \mathbf{z}^2_\tau$ for all $\tau \leq t$, for a given $t \in \mathbb{Z}_-$, we have that $U(\mathbf{z}^1)_t = U(\mathbf{z}^2)_t$.
- **Time-invariant filter:** for any $\tau \in \mathbb{Z}$ define the **time delay** operator $T_\tau : (\mathbb{R}^d)^{\mathbb{Z}} \longrightarrow (\mathbb{R}^d)^{\mathbb{Z}}$ by $T_\tau(\mathbf{z})_t := \mathbf{z}_{t-\tau}$, for any $t \in \mathbb{Z}$. The subset $V \subset (\mathbb{R}^d)^{\mathbb{Z}_-}$ is called time-invariant when $T_\tau(V) = V$, for all $\tau \in \mathbb{Z}$. U is TI when V_z and V_y are TI and $T_\tau \circ U = U \circ T_\tau$, for all $\tau \in \mathbb{Z}$.
- Bijection between causal time-invariant filters and functionals on $V_{z_-} := \mathbb{P}_{\mathbb{Z}_-}(V_z)$ associated to each other:

$$U \longrightarrow H_U(\mathbf{z}) := U(\mathbf{z}^e)_0, \text{ for any } \mathbf{z} \in V_{z_-}$$

$$H \longrightarrow U_H(\mathbf{z})_t := H((\mathbb{P}_{\mathbb{Z}_-} \circ T_{-t})(\mathbf{z})), \text{ for any } t \in \mathbb{Z}$$

where $\mathbb{P}_{\mathbb{Z}_-}$ is the natural projection. It is easy to verify that:

$$H_{U_H} = H, \text{ for any functional } H : V_{z_-} \longrightarrow \mathbb{R}^m,$$

$$U_{H_U} = U, \text{ for any causal TI filter } U : V_z \longrightarrow (\mathbb{R}^m)^{\mathbb{Z}}.$$

- Let $H_1, H_2 : V_{z_-} \longrightarrow \mathbb{R}^m$, $\lambda \in \mathbb{R}$, then $U_{H_1+\lambda H_2}(\mathbf{z}) = U_{H_1}(\mathbf{z}) + \lambda U_{H_2}(\mathbf{z})$, for any $\mathbf{z} \in V_{z_-}$ (the same for functionals)

Discrete-time inputs and outputs

- Sequence spaces $\ell_-^p(\mathbb{R}^d) := \left\{ \mathbf{z} \in (\mathbb{R}^d)^{\mathbb{Z}_-} \mid \|\mathbf{z}\|_p < \infty \right\}$
 - for $1 \leq p < \infty$

$$\|\mathbf{z}\|_p := \left(\sum_{t \in \mathbb{Z}_-} \|\mathbf{z}_t\|^p \right)^{1/p} \quad (1)$$

- for $p = \infty$

$$\|\mathbf{z}\|_\infty := \sup_{t \in \mathbb{Z}_-} \{\|\mathbf{z}_t\|\} \quad (2)$$

The pair $(\ell_-^p(\mathbb{R}^d), \|\cdot\|_p)$, $1 \leq p < \infty$ forms a separable Banach space.

- Weighted sequence spaces $\ell_-^{p,w}(\mathbb{R}^d) := \left\{ \mathbf{z} \in (\mathbb{R}^d)^{\mathbb{Z}_-} \mid \|\mathbf{z}\|_{p,w} < \infty \right\}$, where the weighted p -norms in $(\mathbb{R}^d)^{\mathbb{Z}_-}$ associated to some weighting sequence $w : \mathbb{N} \rightarrow (0, 1]$ (strictly decreasing, limit 0, $w_0 = 1$) are given by

- for $1 \leq p < \infty$

$$\|\mathbf{z}\|_{p,w} := \left(\sum_{t \in \mathbb{Z}_-} \|\mathbf{z}_t\|^p w_{-t} \right)^{1/p} \quad (3)$$

- $p = \infty$

$$\|\mathbf{z}\|_{\infty,w} := \sup_{t \in \mathbb{Z}_-} \{\|\mathbf{z}_t\| w_{-t}\} \quad (4)$$

Decay ratios

The **decay ratio** D_w and the **inverse decay ratio** L_w of w :

$$D_w := \sup_{t \in \mathbb{N}} \left\{ \frac{w_{t+1}}{w_t} \right\} \quad \text{and} \quad L_w := \sup_{t \in \mathbb{N}} \left\{ \frac{w_t}{w_{t+1}} \right\}. \quad (5)$$

Notice that $0 < w_{t+1}/w_t < 1$, for all $t \in \mathbb{N}$, and
 $1 < w_0/w_1 \leq \sup_{t \in \mathbb{N}} \{w_t/w_{t+1}\} = L_w$. Consequently:

$$1 < L_w \leq \infty \quad \text{and analogously,} \quad 0 < D_w \leq 1.$$

Provide a geometric bound for (di/con)vergence speed of w and $w^{-1} := (1/w_t)_{t \in \mathbb{N}}$, resp.

Proposition (Properties of the decay ratios)

Let w be a weighting sequence.

- (i) *For any $t \in \mathbb{N}$, $w_t \leq D_w^t$ and $1/w_t \leq L_w^t$.*
- (ii) *If $D_w < 1$, then $\sum_{t \in \mathbb{N}} w_t \leq \frac{1}{1-D_w}$.*
- (iii) *$L_w D_w \geq 1$.*

Example. Decay ratios

The fading memory property

- We want filters for which the inputs in the far past do not count.
- We encode the **fading memory property (FMP)** as a continuity property:

Definition

We say that CTI filter $U : V_d \subset \ell_-^w(\mathbb{R}^d) \rightarrow V_m \subset \ell_-^w(\mathbb{R}^m)$ has the **fading memory property (FMP)** w.r.t. w if $U : (V_d, \|\cdot\|_w) \rightarrow (V_m, \|\cdot\|_w)$ is continuous, that is for any $\mathbf{z} \in V_d$, any $\epsilon > 0$ there exists $\delta(\epsilon, \mathbf{z}) > 0$ such that

$$\|\bar{\mathbf{z}} - \mathbf{z}\|_w < \delta(\epsilon, \mathbf{z}) \implies \|U(\bar{\mathbf{z}}) - U(\mathbf{z})\|_w < \epsilon, \text{ for any } \bar{\mathbf{z}} \in V_d.$$

- Relations: finite memory $\implies$ FMP, continuity $\not\Rightarrow$ FMP (peak-hold filter), continuity of functional $\not\Rightarrow$ continuity of the associated filter, FMP functional $\not\Rightarrow$ filter is FMP, FMP/continuous filter $\implies$ FMP/continuous functional
- Convenient property: unique steady-state [BC85], or input forgetting, or uniform FMP (uniform continuity w.r.t. topologies induced by w -norms in V_d, V_m).

The following result provides sufficient conditions for the inheritance of the fading memory property and the continuity from functionals to their corresponding filters.

Proposition (Inheritance)

Let w be a weighting sequence and let $H : V_d \subset \ell_-^w(\mathbb{R}^d) \longrightarrow \mathbb{R}^m$ be a functional that has the fading memory property with respect to w and is additionally Lipschitz continuous with constant c_H . Suppose that V_d contains a sequence $\mathbf{z}^0$ such that $U_H(\mathbf{z}^0) \in \ell_-^w(\mathbb{R}^m)$.

Suppose additionally that one of the two following conditions hold

$$(i) \quad R_w := \sup_{s,t \in \mathbb{N}} \left\{ \frac{w_t w_s}{w_{t+s}} \right\} < \infty,$$

$$(ii) \quad \mathcal{L}_w := (L_w^{-t})_{t \in \mathbb{Z}_-} \in \ell_-^w(\mathbb{R}).$$

Then $U_H : V_d \subset \ell_-^w(\mathbb{R}^d) \longrightarrow \ell_-^w(\mathbb{R}^m)$ also has the fading memory property, it is Lipschitz, and $R_w c_H$ (in case (i) holds) or $\|\mathcal{L}_w\|_w c_H$ (in case (ii) is satisfied), respectively, is a Lipschitz constant of U_H . The same conclusion holds for continuous functionals $H : V_d \subset \ell_-^\infty(\mathbb{R}^d) \longrightarrow \mathbb{R}^m$ where the conditions (i) or (ii) are not needed.

In the following statement, for any $\mathbf{u} \in (\mathbb{R}^d)^{\mathbb{Z}^-}$ and any $\mathbf{z} \in (\mathbb{R}^d)^{\mathbb{N}^+}$, the symbol $\mathbf{uz}_t^1 \in (\mathbb{R}^d)^{\mathbb{Z}^-}$, $t \in \mathbb{N}^+$, denotes the concatenation of the vector $\mathbf{u}$ with the truncated vector $\mathbf{z}_t^1 := (\mathbf{z}_1, \dots, \mathbf{z}_t)$ obtained out of $\mathbf{z}$.

Theorem (FMP functionals forget inputs)

Let w be a weighting sequence such that $D_w < 1$ and let

$$U : V_d \subset (\ell_-^w(\mathbb{R}^d), \|\cdot\|_w) \longrightarrow V_m \subset (\ell_-^w(\mathbb{R}^m), \|\cdot\|_w)$$

be a CTI uniform FMP filter and let $H_U : V_d \longrightarrow \mathbb{R}^m$ be the corresponding functional. Then for any $\mathbf{u}, \mathbf{v} \in V_d$ and for any $\mathbf{z} \in (\mathbb{R}^d)^{\mathbb{N}^+}$

$$\lim_{t \rightarrow \infty} \|H_U(\mathbf{uz}_t^1) - H_U(\mathbf{vz}_t^1)\| = 0. \quad (6)$$

Additionally, if V_d is a bounded set in $\ell_-^w(\mathbb{R}^d)$, then there exists a monotonously decreasing sequence w^U with zero limit such that for any $\mathbf{u}, \mathbf{v} \in V_d$ and $\mathbf{z} \in (\mathbb{R}^d)^{\mathbb{N}^+}$ that satisfies (6) it holds that

$$\|H_U(\mathbf{uz}_t^1) - H_U(\mathbf{vz}_t^1)\| \leq w_t^U, \quad \text{for any } t \in \mathbb{Z}_+. \quad (7)$$

Choice of setup for results in this session

Consider **uniformly bounded** families of sequences, that is, subsets of $(\mathbb{R}^d)^{\mathbb{Z}_-}$ of the form

$$K_M := \left\{ \mathbf{z} \in (\mathbb{R}^d)^{\mathbb{Z}_-} \mid \|\mathbf{z}_t\| \leq M \text{ for all } t \in \mathbb{Z}_- \right\}, \quad \text{for some } M > 0. \quad (8)$$

One has that

$$K_M = \overline{B_{\|\cdot\|_\infty}(\mathbf{0}, M)} \subset \ell_-^\infty(\mathbb{R}^d) \subset \ell_-^w(\mathbb{R}^d) \subset (\mathbb{R}^d)^{\mathbb{Z}_-},$$

for any $M > 0$ and any weighting sequence w .

Remark

Advantage: the relative topologies induced on the sets of type K_M by both the product topology in $(\mathbb{R}^d)^{\mathbb{Z}_-}$ and the norm topology of $(\ell_-^w(\mathbb{R}^d), \|\cdot\|_w)$ coincide, for any weighting sequence w .

Advantages

- FMP $\implies$ uniform FMP (due to the fact that $(K_M, \|\cdot\|_w) \subset (\ell_-^w(\mathbb{R}^d), \|\cdot\|_w)$ is a compact, complete, and convex).
- functional $H : K_M \longrightarrow \mathbb{R}^m$ with FMP w.r.t. some $w \implies$ FMP w.r.t. any other w
- filter U_H associated to a FMP functional $H : K_M \longrightarrow \overline{B_{\|\cdot\|}(\mathbf{0}, L)}$ has FMP (notice that there exists $L > 0$ s.t. $H(K_M) \subset \overline{B_{\|\cdot\|}(\mathbf{0}, L)}$)
- filter $U : K_M \longrightarrow \ell_-^w(\mathbb{R}^m)$ with FMP w.r.t. some $w \implies$ FMP w.r.t. any other w

Spaces of filters and functionals

Consider the vector spaces $\mathbb{F}_{K_M}^{FMP}$ and $\mathbb{H}_{K_M}^{FMP}$ formed by the FMP filters and functionals of the form $U : K_M \rightarrow \ell_-^w(\mathbb{R}^m)$ and $H : K_M \rightarrow \mathbb{R}^m$ for some weighting sequence w , respectively. These spaces can be made into normed by

$$\|U\|_{\infty, w} := \sup_{z \in K_M} \{\|U(z)\|_w\} \quad \text{and} \quad \|H\|_{\infty} := \sup_{z \in K_M} \{\|H(z)\|\}.$$

- The compactness of K_M implies that these two norms are finite.
- In the presence of uniformly bounded inputs, the normed spaces $(\mathbb{F}_{K_M}^{FMP}, \|\cdot\|_{\infty, w})$ and $(\mathbb{H}_{K_M}^{FMP}, \|\cdot\|_{\infty})$ made of filters and functionals, respectively, are isometrically isomorphic.

State space systems

- $\mathcal{Z}$ input space, $\mathcal{X}$ state space, $\mathcal{Y}$ output space
- (F, h) reservoir system
- $F : \mathcal{X} \times \mathcal{Z} \rightarrow \mathcal{X}$ state map
- $h : \mathcal{X} \rightarrow \mathcal{Y}$ readout map

Sequences $\mathbf{z} \in \mathcal{Z}^{\mathbb{Z}}$, $\mathbf{y} \in \mathcal{Y}^{\mathbb{Z}}$, and $\mathbf{x} \in \mathcal{X}^{\mathbb{Z}}$ satisfy the **state-space equations** associated to (F, h) , whenever, for any $t \in \mathbb{Z}$, it holds that

$$\begin{cases} \mathbf{x}_t = F(\mathbf{x}_{t-1}, \mathbf{z}_t), \\ \mathbf{y}_t = h(\mathbf{x}_t), \end{cases} \quad (9)$$

where the first equation is called the **state equation** and the second one is called the **observation equation**.

Is this reasonable? Synchronization

Consider an invertible dynamical system on the manifold M defined by $\phi \in \text{Diff}^1(M)$ and let $\omega \in C^1(M, \mathbb{R}^d)$ be an observation map. Denote:

$$\mathbf{y}_{t+1} = \phi(\mathbf{y}_t) \quad \text{and} \quad \mathbf{z}_t = \omega(\mathbf{y}_t).$$

Consider now the state-space system driven by the dynamical system observations

$$\mathbf{x}_{t+1} = F(\mathbf{x}_t, \mathbf{z}_{t+1}).$$

The state-space system is likely to produce a representation for the dynamical system when there exists a **generalized synchronization**, that is, a map $H \in C^1(M, \mathbb{R}^N)$, such that

$$\mathbf{x}_t = H(\mathbf{y}_t), \quad \text{for all } t.$$

Synchronization Theorem

Definition (System flow)

Consider the system defined by $F : \mathbb{R}^N \times \mathbb{R}^d \longrightarrow \mathbb{R}^N$. The associated system flow $\mathcal{F}_t : \mathbb{R}^N \times (\mathbb{R}^d)^t \longrightarrow \mathbb{R}^N$ is the system that maps any initial condition $\mathbf{x}_0 \in \mathbb{R}^N$ and input sequence $\mathbf{Z}_t := \{\mathbf{z}_1, \dots, \mathbf{z}_t\}$ to the value $\mathcal{F}_t(\mathbf{x}_0, \mathbf{Z}_t) \in \mathbb{R}^N$ obtained by applying the state-equation t -times, with initial condition $\mathbf{x}_0 \in \mathbb{R}^N$ and input $\mathbf{Z}_t$

Theorem ([KP96, HOY97])

A unique generalized synchronization exists if and only if for any $\mathbf{x}_0, \bar{\mathbf{x}}_0 \in \mathbb{R}^N$ and any $\mathbf{y}_0 \in \mathbf{M}$

$$\lim_{t \rightarrow \infty} \|\mathcal{F}_t(\mathbf{x}_0, \mathbf{Z}_t^{\mathbf{y}_0}) - \mathcal{F}_t(\bar{\mathbf{x}}_0, \mathbf{Z}_t^{\mathbf{y}_0})\| = 0,$$

with $\mathbf{Z}_t^{\mathbf{y}_0} := \{\omega(\phi(\mathbf{y}_0)), \omega(\phi^2(\mathbf{y}_0)), \dots, \omega(\phi^t(\mathbf{y}_0))\}$.

- The hypothesis of the theorem implies that this definition does not depend on $\mathbf{x}_0$.
- It is a consequence called the **state forgetting property** of a continuity feature that will be at the center of many of our developments: the **fading memory property**.
- The synchronization map H may be "wild" enough to change the attractor information dimension of the response with respect to the driving dynamical system [HOY97].
- We shall device reservoir systems that yield, in the spirit of Takens' Theorem, C^1 synchronization maps.

Remark

*Finding state-space representations for input/output systems is related to the so called **realization problem** in systems and control theory.*

*State-space representations of input/output systems are sometimes referred to as **internal representations** because they involve the use of **internal, latent, or state variables**.*

These latent variables are sometimes just mathematical instruments or, in other occasions, represent non observable parts of a system whose values can be partially inferred out of observed values; this is the goal of Kalman filter-related techniques [DK12, Sär13]. In contrast with the internal representations, filters are referred to as the **external representation** of an input/output system.

Remark

The state-space representations of an input/output system have the advantage of being recursively computable and provide simultaneously a model and an algorithm.

Examples: Linear state space system

Examples: Time-invariant finite-memory filters

Let $U : (\mathbb{R}^d)^{\mathbb{Z}^-} \longrightarrow (\mathbb{R}^m)^{\mathbb{Z}^-}$ be the time-invariant filter determined by the map $h : (\mathbb{R}^d)^\tau \longrightarrow \mathbb{R}^m$, $\tau \in \mathbb{N}^+$, that is,

$$U(\mathbf{z})_t = h(\mathbf{z}_{t-\tau+1}, \dots, \mathbf{z}_t).$$

This filter can be equivalently written as a linear state-space system of the type (??) with a nonlinear readout via

$$A = \begin{pmatrix} \mathbb{O}_{d(\tau-1),d} & \mathbb{I}_{d(\tau-1)} \\ \mathbb{O}_{d,d} & \mathbb{O}_{d,d(\tau-1)} \end{pmatrix} \quad \text{and} \quad C = \begin{pmatrix} \mathbb{O}_{d(\tau-1),d} \\ \mathbb{I}_d \end{pmatrix} \quad (10)$$

and using h as the readout. When h is polynomial, the family made of state-space systems of this type has universal approximation properties in the category of filters with the fading memory property. This will show that any fading memory filter can be uniformly approximated by a finite-memory one.

Examples: Echo State Network (ESN)

State-affine systems (SAS)

Take two polynomials $p(z) \in \mathbb{M}_{N,N}[z]$ and $q(z) \in \mathbb{M}_{N,1}[z]$ on the variable z with matrix coefficients, that is

$$p(z) := A_0 + zA_1 + z^2A_2 + \cdots + z^{n_1}A_{n_1},$$

$$q(z) := B_0 + zB_1 + z^2B_2 + \cdots + z^{n_2}B_{n_2}$$

The **non-homogeneous state-affine system (SAS)** associated to p, q and $\mathbf{W}$ is the reservoir system determined by the state-space transformation:

$$\begin{cases} \mathbf{x}_t = p(z_t)\mathbf{x}_{t-1} + q(z_t), \end{cases} \quad (11)$$

$$\begin{cases} y_t = \mathbf{W}^\top \mathbf{x}_t. \end{cases} \quad (12)$$

Examples: State-Affine Systems (SAS)

Let $\mathbb{M}_{N,M}[\mathbf{z}]$ be the space of polynomials on $\mathbf{z} \in \mathbb{R}^d$ with matrix coefficients in $\mathbb{M}_{N,M}$, that is, the set of elements p of the form

$$p = \sum_{\alpha \in V_p} \mathbf{z}^\alpha A_\alpha,$$

with $V_p \subset \mathbb{N}^d$ a finite subset and $A_\alpha \in \mathbb{M}_{N,M}$ the matrix coefficients. A **state-affine system** is given by

$$\begin{cases} \mathbf{x}_t = p(\mathbf{z}_t)\mathbf{x}_{t-1} + q(\mathbf{z}_t), \\ \mathbf{y}_t = W\mathbf{x}_t, \end{cases} \quad (13)$$

with $p \in \mathbb{M}_{N,N}[\mathbf{z}]$, $q \in \mathbb{M}_{N,d}[\mathbf{z}]$ and $W \in \mathbb{M}_{m,N}$. Note that the linear system is a particular case of this specification that can be obtained by taking as p a polynomial of degree zero and q a polynomial of degree one.

$\text{Nil}_n^k[\mathbf{z}] \subset \mathbb{M}_n[\mathbf{z}]$ is the set of nilpotent $\mathbb{M}_n$ -valued polynomials on $\mathbf{z}$ of index k , that is, $p(\mathbf{z}) \in \text{Nil}_n^k[\mathbf{z}]$ whenever k is the smallest natural number for which $p(\mathbf{z})^k = \mathbf{0}$, for all $\mathbf{z} \in \mathbb{R}$. $\text{Nil}[\mathbf{z}]$ is the set of matrix-valued nilpotent polynomials on $\mathbf{z}$ of any order and any index.

Echo state property (ESP)

The link between state-space systems and input/output systems is given by the following property.

Definition

Let $V_{\mathcal{Z}} \subset \mathcal{Z}^{\mathbb{Z}}$, $V_{\mathcal{X}} \subset \mathcal{X}^{\mathbb{Z}}$, and $V_{\mathcal{Y}} \subset \mathcal{Y}^{\mathbb{Z}}$ be subsets of the input, state, and output spaces, respectively.

- Let F be a state-space map. We say that F has the $(V_{\mathcal{Z}}, V_{\mathcal{X}})$ -**echo state property (ESP)** whenever for each $\mathbf{z} \in V_{\mathcal{Z}}$ there is a unique $\mathbf{x} \in V_{\mathcal{X}}$ such that the first equation in (9) holds.
- Let (F, h) be a state-space system. We say that (F, h) has the $(V_{\mathcal{Z}}, V_{\mathcal{Y}})$ -**echo state property (ESP)** whenever for each $\mathbf{z} \in V_{\mathcal{Z}}$ there is a unique $\mathbf{y} \in V_{\mathcal{Y}}$ such that the equations (9) hold.

Remark

If F has the $(V_{\mathcal{Z}}, V_{\mathcal{X}})$ echo state property then (F, h) has the $(V_{\mathcal{Z}}, V_{\mathcal{Y}})$ -echo state property with

$$V_{\mathcal{Y}} := \left\{ \mathbf{y} \in \mathcal{Y}^{\mathbb{Z}} \mid \mathbf{y}_t = h(\mathbf{x}_t), t \in \mathbb{Z}, (\mathbf{x}_t)_{t \in \mathbb{Z}} \in V_{\mathcal{X}} \right\}. \quad (14)$$

The converse situation does not hold in general.

Reservoir filters

The reservoir system

$$\begin{cases} \mathbf{x}_t = F(\mathbf{x}_{t-1}, \mathbf{z}_t), \end{cases} \quad (15)$$

$$\begin{cases} y_t = h(\mathbf{x}_t), \end{cases} \quad (16)$$

determines a filter when the following existence and uniqueness property holds (**echo state property** [Jae10, YJK12]): for each $\mathbf{z} \in (D_d)^\mathbb{Z}$ there exists a unique $\mathbf{x} \in (D_N)^\mathbb{Z}$ such that for each $t \in \mathbb{Z}$, the relation (15) holds.

- The **state filter** $U^F : (D_d)^\mathbb{Z} \longrightarrow (D_N)^\mathbb{Z}$ is determined by $U^F(\mathbf{z})_t := \mathbf{x}_t \in D_N$
- The **reservoir filter** $U_h^F : (D_d)^\mathbb{Z} \longrightarrow \mathbb{R}^\mathbb{Z}$ is determined by the entire reservoir system, that is, $U_h^F(\mathbf{z})_t := h(U^F(\mathbf{z})_t) = y_t$.

The filters U^F and U_h^F are causal by construction and are necessarily time-invariant [GO18b]. We can hence associate to U_h^F a **reservoir functional** $H_h^F : (D_d)^\mathbb{Z} \longrightarrow \mathbb{R}$ determined by $H_h^F := H_{U_h^F}$.

ESP and FMP for state contracting maps

Under what conditions state systems induced by state maps that are contracting on the state variable have the echo state property and the fading memory properties?

Theorem (The ESP and FMP for contractive state systems)

Let $d, N \in \mathbb{N}$, $D_d \subset \mathbb{R}^d$, $D_N \subset \mathbb{R}^N$ and let $F : D_N \times D_d \rightarrow D_N$ be a continuous reservoir map. Assume that F is a r -contraction on the first entry. Let w be a weighting sequence such that $L_w < \infty$ and let $V_d \subset (D_d)^{\mathbb{Z}^-} \cap \ell_-^w(\mathbb{R}^d)$ be a time invariant set. Suppose that one of the following hypotheses holds:

- (i) D_N is a compact subset of $\mathbb{R}^N$
- (ii) $V_N := (D_N)^{\mathbb{Z}^-} \cap \ell_-^w(\mathbb{R}^N)$ is a complete subset of $(\ell_-^w(\mathbb{R}^N), \|\cdot\|_w)$, F is Lipschitz continuous, and there exists a solution $(\mathbf{x}^0, \mathbf{z}^0) \in V_N \times V_d$ for the system associated to F .

In both cases, if

$$rL_w < 1, \quad (17)$$

then the system associated to F has the (V_d, V_N) -ESP and FMP. This statement also holds true under the hypotheses in part (ii) if $\ell_-^w(\mathbb{R}^N)$ is replaced by $\ell_-^\infty(\mathbb{R}^N)$, in which case condition (17) is not needed and the resulting filter is continuous.

The FMP of the filter associated to a reservoir map propagates to the FMP of the filter of the full reservoir system if the readout map is continuous.

Corollary

Let $M, L > 0$, let $K_M \subset (\mathbb{R}^d)^{\mathbb{Z}^-}$ and $K_L \subset (\mathbb{R}^N)^{\mathbb{Z}^-}$ be subsets of uniformly bounded sequences, and let $F : \overline{B_{\|\cdot\|}}(\mathbf{0}, L) \times \overline{B_{\|\cdot\|}}(\mathbf{0}, M) \longrightarrow \overline{B_{\|\cdot\|}}(\mathbf{0}, L)$ be a continuous state map. Assume, additionally, that F is a contraction on the first entry with constant $0 < r < 1$. Then, the state system associated to F has the echo state property. Moreover, this system has a unique causal and time-invariant filter $U^F : K_M \longrightarrow K_L$ associated that has the fading memory property with respect to any weighting sequence w .

Examples: Linear State Space Systems

In this case the condition (17) reads

$$\|A\|_2 L_w < 1 \quad (18)$$

and the reservoir filter $U_h^F : \ell_-^w(\mathbb{R}^d) \rightarrow \ell_-^w(\mathbb{R}^m)$ is given by

$$U_h^F(\mathbf{z})_t = h \left(\sum_{j=0}^{\infty} A^j C \mathbf{z}_{t-j} \right). \quad (19)$$

The statement (19) is proved by showing, first, that for any $\mathbf{z} \in \ell_-^w(\mathbb{R}^d)$ the sequence $(\mathbf{S}^n)_{n \in \mathbb{N}^+}$ with $\mathbf{S}^n \in (\mathbb{R}^N)^{\mathbb{Z}_-}$ given by $\mathbf{S}_t^n := \sum_{j=0}^n A^j C \mathbf{z}_{t-j}$ is a Cauchy sequence in $\ell_-^w(\mathbb{R}^N)$. Indeed, for any $n \in \mathbb{N}^+$

$$\begin{aligned} \|\mathbf{S}^n\|_w &= \sup_{t \in \mathbb{Z}_-} \{\|\mathbf{S}_t^n\|_w\} \leq \sup_{t \in \mathbb{Z}_-} \left\{ \left(\sum_{j=0}^n \|A\|_2^j \|C\|_2 \|\mathbf{z}_{t-j}\|_w \frac{w_{t-j}}{w_{t-j}} \right) w_{-t} \right\} \\ &\leq \sum_{j=0}^n \|A\|_2^j \|C\|_2 L_w^j \|\mathbf{z}\|_w = \|C\|_2 \|\mathbf{z}\|_w \frac{1 - (\|A\|_2 L_w)^{n+1}}{1 - \|A\|_2 L_w} < \infty, \end{aligned}$$

which proves that $\mathbf{S}^n \in \ell_-^w(\mathbb{R}^N)$.

Example: LSS systems continued

Analogously, it is easy to see that for any $n, m \in \mathbb{N}^+$, $n < m$

$$\begin{aligned}\|\mathbf{S}^n - \mathbf{S}^m\|_w &\leq \|C\|_2 \|\mathbf{z}\|_w \sum_{j=n+1}^m (\|A\|_2 L_w)^j \leq \|C\|_2 \|\mathbf{z}\|_w \sum_{j=n+1}^{\infty} (\|A\|_2 L_w)^j \\ &= \|C\|_2 \|\mathbf{z}\|_w \frac{(\|A\|_2 L_w)^{n+1}}{1 - \|A\|_2 L_w},\end{aligned}$$

which tends to zero as $n \rightarrow \infty$ proving that $(\mathbf{S}^n)_{n \in \mathbb{N}^+}$ is a Cauchy sequence in $\ell_-^w(\mathbb{R}^N)$. As $\ell_-^w(\mathbb{R}^N)$ is a Banach space, $(\mathbf{S}^n)_{n \in \mathbb{N}^+}$ converges to an element $\mathbf{S} \in \ell_-^w(\mathbb{R}^N)$ given by $\mathbf{S}_t := \sum_{j=0}^{\infty} A^j C \mathbf{z}_{t-j}$, $t \in \mathbb{Z}_-$.

Given that the terms of this sequence satisfy the recursion $\mathbf{S}_t = A \mathbf{S}_{t-1} + C \mathbf{z}_t$ and that Theorem (slide 31) guarantees that F has the $(\ell_-^w(\mathbb{R}^d), \ell_-^w(\mathbb{R}^N))$ -echo-state property, we necessarily have that

$$U^F(\mathbf{z})_t = \sum_{j=0}^{\infty} A^j C \mathbf{z}_{t-j}, \quad t \in \mathbb{Z}_-$$

and therefore (19) follows.

Echo State Networks

Suppose that $\sigma : \mathbb{R} \rightarrow [-1, 1]$, σ is non-decreasing, $\lim_{x \rightarrow \infty} \sigma(x) = 1$, $\lim_{x \rightarrow -\infty} \sigma(x) = -1$, and suppose that $L_\sigma := \sup_{x \in \mathbb{R}} \{|\sigma'(x)|\} < \infty$. The condition (17) reads

$$\|A\|_2 L_\sigma L_w < 1.$$

No explicit expression for the filter $U_w^F : \ell_-^w(\mathbb{R}^d) \rightarrow \ell_-^w(\mathbb{R}^m)$ is available.

SAS

Let the reservoir map is given by $F(\mathbf{x}, \mathbf{z}) = p(\mathbf{z})\mathbf{x} + q(\mathbf{z})$ with

$$\begin{aligned} p(\mathbf{z}) &\in \mathbb{M}_{N,N}[\mathbf{z}] \quad \text{or} \quad \text{Trig}_{N,N}[\mathbf{z}], \\ q(\mathbf{z}) &\in \mathbb{M}_{N,d}[\mathbf{z}] \quad \text{or} \quad \text{Trig}_{N,d}[\mathbf{z}]. \end{aligned}$$

Suppose that $F : \mathbb{R}^N \times D_d \longrightarrow \mathbb{R}^N$ with D_d bounded, so that

$$M_p := \sup_{\mathbf{z} \in D_d} \{\|p(\mathbf{z})\|\} < \infty, \quad M_q := \sup_{\mathbf{z} \in D_d} \|q(\mathbf{z})\| < \infty.$$

The condition (17) reads in this case $M_p L_w < 1$ and the reservoir filter $U_W^F : \ell^w(\mathbb{R}^d) \cap (D_d)^{\mathbb{Z}_-} \longrightarrow \ell_-^w(\mathbb{R}^m)$ is given by

$$U_W^F(\mathbf{z})_t = W \left(\sum_{j=0}^{\infty} p(\mathbf{z}_t) p(\mathbf{z}_{t-1}) \cdots p(\mathbf{z}_{t-j+1}) q(\mathbf{z}_{t-j}) \right). \quad (20)$$

The proof that the filter (20) indeed satisfies the state-space system in (13) is carried out by mimicking the case of linear reservoir systems and by noticing that since D_d is bounded and $q : D_d \longrightarrow \mathbb{M}_{N,d}$ is polynomial, then it is necessarily Lipschitz with a constant $L_q > 0$.

The goal - prove that some of the families we introduced are universal approximants.

In the deterministic part we work in the fading memory category with uniformly bounded inputs.

In that case, universality is shown by proving the density of the relevant families with respect to the topology induced by the norm in the spaces described earlier.

Universality results in the deterministic setup

Tools: The Stone-Weierstrass theorem for polynomial subalgebras of real-valued functions defined on compact metric spaces.

Approach: One needs to prove that filters form polynomial algebras.

Theorem (Universality of families of state-space filters)

Let $K_M \subset (\mathbb{R}^d)^{\mathbb{Z}_-}$, I an index set, and let

$$\mathcal{R}_I := \{H_{h_i}^{F_i} : K_M \longrightarrow \mathbb{R}^m \mid h_i \in C^0(D_{N_i}, \mathbb{R}^m), F_i : D_{N_i} \times \overline{B_{\|\cdot\|}}(0, M) \rightarrow D_{N_i}, i \in I, N_i \in \mathbb{N}^+, \\ D_{N_i} \subset \mathbb{R}^{N_i}, H_{h_i}^{F_i} \text{ continuous, } F_i \text{ has the } (K_M, \ell^\infty(\mathbb{R}^{N_i}))\text{-ESP}\}.$$

Let $\mathcal{A}(\mathcal{R}_I)$ be the algebra generated by $\mathcal{R}_I$ using the Hadamard product $\odot$ introduced as

$$H_{h_i}^{F_i} \odot H_{h_j}^{F_j} := (H_{h_i,1}^{F_i} \cdot H_{h_j,1}^{F_j}, \dots, H_{h_i,m}^{F_i} \cdot H_{h_j,m}^{F_j}).$$

Then:

- (i) $\mathcal{A}(\mathcal{R}_I)$ consists of continuous state-space functionals.
- (ii) If $\mathcal{A}(\mathcal{R}_I)$ contains the constant functionals and separates the points in K_M then it is dense in $(\mathbb{H}_{K_M}^{FMP}, \|\cdot\|_\infty)$ that is,

$$\overline{\mathcal{A}(\mathcal{R}_I)} = \mathbb{H}_{K_M}^{FMP}.$$

Hence, given any functional H with the fading memory property and any $\epsilon > 0$ there exists a state-space functional $H_h^F \in \mathcal{A}(\mathcal{R}_I)$ such that

$$\|H - H_h^F\|_\infty < \epsilon.$$

Proof

Corollary

(State-space systems with polynomial readouts are universal) Let $M > 0$, $m \in \mathbb{N}^+$, and $K_M \subset (\mathbb{R}^d)^\mathbb{Z}$, and let

$$\mathcal{R} := \left\{ H_h^F : K_M \rightarrow \mathbb{R}^m \mid F : D_N \times \overline{B_{\|\cdot\|}(\mathbf{0}, M)} \rightarrow D_N, D_N \subset \mathbb{R}^N, h \in \text{Pol}(\mathbb{R}^N, \mathbb{R}^m) \right\}.$$

Then,

$$\mathcal{A}(\mathcal{R}) = \mathcal{R} \quad \text{and} \quad \overline{\mathcal{R}} = \mathbb{H}_{K_M}^{FMP}. \quad (21)$$

Corollary

(Universality of linear systems with polynomial readouts) Let $M > 0$, $m \in \mathbb{N}^+$, $0 < \epsilon < 1$ arbitrary but fixed. Let $K_M \subset (\mathbb{R}^d)^{\mathbb{Z}}$ and define the family

$$\mathcal{L} := \left\{ H_h^{A,C} : K_M \rightarrow \mathbb{R}^m \mid A \in \mathbb{M}_N, \|A\|_2 < 1 - \epsilon, C \in \mathbb{M}_{N,d}, h \in \text{Pol}(\mathbb{R}^N, \mathbb{R}^m) \right\}.$$

Then, $\mathcal{L}$ is dense in $\mathbb{H}_{K_M}^{\text{FMP}}$. The same result holds, if we replace $\mathcal{L}$ by the families $\mathcal{N}$ and $\mathcal{D}$ that contain only nilpotent and diagonal connectivity matrices A , respectively.

Corollary

Universality of the state affine system family *Let $M > 0, m, R \in \mathbb{N}^+$, arbitrary but fixed. Define:*

$$S := \left\{ H_W^{p,q} : K_M \longrightarrow \mathbb{R}^m \mid p \in \mathcal{SNil}_N^k[\mathbf{z}], q \in \mathbb{M}_{N,d}[\mathbf{z}], N \in \mathbb{N} \right\}$$

Then S is dense in $\mathbb{H}_{K_M}^{FMP}$.

Theorem

Universality of the echo state family with invertible activation

functions *Let $M > 0$ and $\epsilon > 0$ arbitrary but fixed and let*

$U : K_M \rightarrow (\mathbb{R}^m)^{\mathbb{Z}_-}$ be a causal and time invariant filter that has the fading memory property. Then, there exists a echo state network

$$\begin{cases} \mathbf{x}_t = \sigma(\mathbf{A}\mathbf{x}_{t-1} + \mathbf{C}\mathbf{z}_t + \zeta) \\ \mathbf{y}_t = \mathbf{W}\mathbf{x}_t \end{cases}$$

with invertible activation $\sigma : \mathbb{R} \rightarrow \mathbb{R}$ that has the echo state property and the fading memory property and whose associated filter

$U_{ESN} : K_M \rightarrow (\mathbb{R}^m)^{\mathbb{Z}_-}$ satisfies that

$$\|U - U_{ESN}\|_{\infty} < \epsilon.$$

References I



S. Boyd and L. Chua.

Fading memory and the problem of approximating nonlinear operators with Volterra series.
IEEE Transactions on Circuits and Systems, 32(11):1150–1161, nov 1985.



Ngai Hang Chan, Wilfredo Palma, and Others.

State space modeling of long-memory processes.
The Annals of Statistics, 26(2):719–740, 1998.



J. Durbin and Siem Jan Koopman.

Time Series Analysis by State Space Methods.
Oxford University Press, 2012.



Lukas Gonon and Juan-Pablo Ortega.

Reservoir computing universality with stochastic inputs.
IEEE Transactions on Neural Networks and Learning Systems, 2018.



Lyudmila Grigoryeva and Juan-Pablo Ortega.

Echo state networks are universal.
Neural Networks, 108:495–508, 2018.



Lyudmila Grigoryeva and Juan-Pablo Ortega.

Universal discrete-time reservoir computers with stochastic inputs and linear readouts using non-homogeneous state-affine systems.
Journal of Machine Learning Research, 19(24):1–40, 2018.



J. R M Hosking.

Fractional differencing.
Biometrika, 1981.

References II



Brian R. Hunt, Edward Ott, and James A. Yorke.

Differentiable generalized synchronization of chaos.

Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics, 55(4):4029–4034, 1997.



Herbert Jaeger.

The 'echo state' approach to analysing and training recurrent neural networks with an erratum note.

Technical report, German National Research Center for Information Technology, 2010.



Lj Kocarev and U Parlitz.

Generalized synchronization, predictability, and equivalence of unidirectionally coupled dynamical systems.

Physical Review Letters, 76(11):1816–1819, 1996.



Irwin W. Sandberg.

Notes of fading-memory conditions.

Circuits, Systems, and Signal Processing, 22(1):43–55, 2003.



Simo Särkkä.

Bayesian Filtering and Smoothing.

Cambridge University Press, 2013.



Izzet B Yildiz, Herbert Jaeger, and Stefan J Kiebel.

Re-visiting the echo state property.

Neural Networks, 35:1–9, nov 2012.