



HAL
open science

Realization Theory Of Recurrent Neural ODEs Using Polynomial System Embeddings

Martin Gonzalez, Thibault Defourneau, Hatem Hajri, Mihaly Petreczky

► **To cite this version:**

Martin Gonzalez, Thibault Defourneau, Hatem Hajri, Mihaly Petreczky. Realization Theory Of Recurrent Neural ODEs Using Polynomial System Embeddings. 2022. hal-03782520

HAL Id: hal-03782520

<https://hal.science/hal-03782520v1>

Preprint submitted on 21 Sep 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Realization Theory Of Recurrent Neural ODEs Using Polynomial System Embeddings

Martin Gonzalez^{a,*}, Thibault Defourneau^{b,1}, Hatem Hajri^a, Mihaly Petreczky^c

^aInstitut de Recherche Technologique SystemX, Palaiseau, France.

^bTrinov, 75001 Paris, France.

^cCentre de Recherche en Informatique, Signal et Automatique de Lille, UMR CNRS 9189, France.

Abstract

In this paper we show that neural ODE analogs of recurrent (ODE-RNN) and Long Short-Term Memory (ODE-LSTM) networks can be algorithmically embedded into the class of polynomial systems. This embedding preserves input-output behavior and can suitably be extended to other neural DE architectures. We then use realization theory of polynomial systems to provide necessary conditions for an input-output map to be realizable by an ODE-LSTM and sufficient conditions for minimality of such systems. These results represent the first steps towards realization theory of recurrent neural ODE architectures, which is expected to be useful for model reduction and learning algorithm analysis of recurrent neural ODEs.

Keywords: Realization theory, Neural ODEs, Recurrent Neural Networks, Long Short-Term Memory, System Identification.

Introduction

Long Short-Term Memory networks (LSTMs) represent a generalization of recurrent neural networks (RNNs) widely used in text analysis tasks such as grammar correction and next word prediction. They were introduced [26], and have been studied intensively ever since, due to their effectiveness for learning long-term dependencies in comparison with RNN algorithms and its variants (see [25] and [7] for details).

More recently, the introduction of neural ODEs [11] and other implicit network architectures [4, 43, 46, 33, 15, 51, 53, 35] has opened the door to new machine learning paradigms tightly related to dynamical system modeling techniques. For instance, neural ODEs can be seen as continuous-time dynamical systems and as infinitesimally connected ResNets while the latter can be seen as Euler discretizations of neural ODEs. As such, they exploit a rich available theory on both sides, offering memory efficiency whilst their recurrent analogs have the ability of handling irregular data and are suitable for tackling generative problems and time series (particularly in physics), becoming relevant to both modern machine learning and traditional mathematical modeling.

In this paper we make the first steps towards developing realization theory of recurrent neural ODE architectures. We focus on the present article on neural ODE analogs of RNNs and LSTMs and we aim at characterizing those input-output maps which can be represented by these systems and understanding the minimal size of such systems sufficient to be able to represent a given input-output map.

The motivation for studying realization theory for neural ODEs is that learning algorithms for such systems from data correspond to system identification algorithms. Realization theory is central in system identification as it can be viewed as an attempt to solve a system identification problem through idealized qualitative analysis, where there is infinite data and no modelling error. For linear systems, realization theory [30, 36] allowed to address identifiability, canonical forms and gave rise to subspace identification algorithms.

In order to make the discussion more precise, let us determine an ambient class of dynamical systems containing ODE-RNNs, ODE-LSTMs² and polynomial systems as some of its subclasses, and having all desired properties needed for our study. For a subset $Z \subset \mathbb{R}^k$, $k \geq 1$, we denote:

- \mathcal{Z} (resp. \mathcal{Z}_{pc} , \mathcal{Z}_{ac}) the set of continuous (resp. piecewise-continuous, resp. absolutely continuous) functions from $[0; +\infty[$ to Z . Denote by \mathcal{Z}_{pco} the set of all piecewise-constant functions from $[0, +\infty[$ to Z which are constant starting from a certain point, i.e., $h \in \mathcal{Z}_{pco}$, if h is piecewise-constant and there exists $T_h \geq 0$ such that the restriction of h to $[T_h, +\infty[$ is constant.
- \mathcal{Z}^f the set of functions from $[0, T]$ to Z , for some $T > 0$. We let the reader combine this with the above notations.

The ambient class \mathcal{F} of dynamical systems we will consider in this paper is described by differential equations of the form

$$\begin{cases} \dot{x}(t) = f(x(t), u(t)) \\ y(t) = g(x(t)) \end{cases} \quad t \geq 0, \quad (1)$$

with initial condition $x(0) = x_0 \in \mathbb{R}^n$, where

²These will be specified by a general class of two-scale ODE-based RNNs where multiple continuous-time memory architectures arise.

*Corresponding author. *E-mail address:* martin.gonzalez@irt-systemx.fr

¹The work presented in this paper was carried out while T. Defourneau was employed by Centre de Recherche en Informatique, Signal et Automatique de Lille.

- $(x, u, y) \in \mathcal{X}_{ac} \times \mathcal{U}_{pc} \times \mathcal{S}_{pc}$ ³, where $X = \mathbb{R}^n, S = \mathbb{R}^p$ denote respectively the state and output spaces and $U = \{\alpha_1, \dots, \alpha_K\} \subset \mathbb{R}^m$ is a finite input space with cardinality K ;
- $f : X \times U \rightarrow X$ is analytic on its first argument and $g : X \rightarrow S$ is analytic.

We will identify these systems with tuples of the form $\Sigma = (f, g, x_0)$. The triple (m, n, p) will be called the format of Σ .

On the one hand, polynomial systems are a subclass of such systems and many methods of computational algebra can be used to determine qualitative properties of such systems such as observability, reachability and minimality. On the other hand, one can think of ODE-RNNs and ODE-LSTMs as subclasses of \mathcal{F} which can be parameterized according to some class of learning weight functions $\theta(t)$, which will be assumed to be constant for simplicity. As such, under mild assumptions, we can associate polynomial systems to large classes of ODE-RNNs and ODE-LSTMs and, by doing so, infer such qualitative properties on these classes. More specifically,

- We show that an i-o map can be realized by an ODE-RNN or an ODE-LSTM, only if it can be realized by a polynomial system, i.e. a non-linear system defined by vector fields and readout maps which are polynomials. We present an explicit algorithmic construction of such a polynomial system.
- We infer sufficient conditions for minimality/observability/reachability/accessibility of ODE-RNNs and ODE-LSTMs from the properties of their associated polynomial systems [41, 38, 5, 48].
- We present a necessary condition for existence of a realization by ODE-RNNs and ODE-LSTMs, using results from realization theory of polynomial systems. This necessary condition is a generalization of the well-known rank condition for Hankel matrices of linear systems.

Note that elements in \mathcal{F} could be viewed as analytic systems for which there is an existing realization theory [29, 27, 24]. However, as analytical functions do not have a finite representation, this approach is not computationally effective: there are no algorithms for checking minimality, deciding equivalence of two systems neither transforming a system to a minimal one. Nevertheless, seen as polynomial systems, computer algebra tools can be used to address these issues [40]. In addition, since polynomial systems have much more algebraic structures than analytic systems and the conditions for minimality/observability/reachability studied here are less restrictive than those which can be obtained by the analytic approach.

Related work: To the best of our knowledge, the results of the paper are new. Observability, controllability and minimality of ODE-RNNs were investigated in [3, 44, 2], but no

results on existence of a realization were provided, and the results of [3, 44, 2] used certain assumptions on the weights of the ODE-RNNs. In contrast to [3, 44, 2], in this paper we consider ODE-LSTMs and we address the issue of existence of a realization by ODE-LSTM. Moreover, the technique used in this paper is completely different from that of [3, 44, 2]. Rational embeddings and elements of realization theory of a subclass of ODE-RNNs were considered in [14]. In comparison to [14], the main novelty is that in this paper we consider both ODE-RNNs and ODE-LSTMs and that detailed proofs and examples are provided. That is, the current paper extends the results of [14].

1. Preliminaries

We denote $\mathbb{R}[X_1, \dots, X_n]$ the algebra of real polynomials in n variables and denote $\mathbb{R}(X_1, \dots, X_n)$ its quotient field, whose elements are rational functions in n variables. If R is an integral domain over \mathbb{R} then the transcendence degree $\text{trdeg} R$ of R over \mathbb{R} is defined as the transcendence degree over \mathbb{R} of the field F of fractions of R and it equals the greatest number of algebraically independent elements of F over \mathbb{R} . Let n, m be two integers and let $\mathbb{R}[X_1, \dots, X_n; \mathbb{R}^m]$ be the set of tuples (P_1, \dots, P_m) whose m components are polynomials in n variables.

Denote $C^\omega(\mathbb{R})$ the algebra of real analytic functions over \mathbb{R} and denote $\sigma^{(i)}$ the i -th derivative of $\sigma \in C^\omega(\mathbb{R})$. We denote $C_1^\omega(\mathbb{R})$ the subset of $C^\omega(\mathbb{R})$ of those analytic functions satisfying

$$\sigma^{(1)} = P(\sigma), \quad (2)$$

for some $P \in \mathbb{R}[X]$. In this paper we will consider only activation functions of this sort. In particular, the hyperbolic tangent and the logistic functions

$$\forall x \in \mathbb{R}, \quad th(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}, \quad S(x) = \frac{1}{1 + e^{-x}}$$

are both elements of $C_1^\omega(\mathbb{R})$ as they are the unique solutions of the differential equations

$$th'(x) = 1 - th^2(x) \quad S'(x) = S(x) - S^2(x)$$

with initial conditions $th(0) = 0$ and $S(0) = \frac{1}{2}$.

For a map $\sigma : \mathbb{R} \rightarrow \mathbb{R}$ denote $\vec{\sigma} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ the map defined by

$$\vec{\sigma} : (x_1, \dots, x_n)^T \mapsto (\sigma(x_1), \dots, \sigma(x_n))^T \quad (3)$$

Let \odot be the Hadamard product $(A \odot B)_{ij} := (A)_{ij}(B)_{ij}$ where A, B matrices of same dimension. In particular, for $P, Q, R \in \mathbb{R}^n$, the expression $P \odot Q = R$ is equivalent to $P_i Q_i = R_i$, for $i \in [n] := \{1, \dots, n\}$.

1.1. Polynomial systems

Define the subclass \mathcal{F}_0 of \mathcal{F} consisting of those systems described by

$$\begin{cases} \dot{x}(t) = P_{u(t)}(x(t)) \\ y(t) = g(x(t)), \end{cases} \quad t \geq 0, \quad (4)$$

³Allowing $u \in \mathcal{U}_{pc}$ represents a relatively small technical difficulty that we avoid here for clarity.

with initial condition $x(0) = x_0 \in \mathbb{R}^n$, where $P_{u(t)} \in \mathbb{R}[X_1, \dots, X_n; \mathbb{R}^n]$ for $u(t) \in U$ and $g \in \mathbb{R}[X_1, \dots, X_n; \mathbb{R}^p]$. These are polynomial systems and will be identified with tuples $\mathcal{P} = (\{P_{u(t)}\}_{u(t) \in U}, g, x_0)$ and have at most one solution $(x, u, y) \in \mathcal{X}_{ac} \times \mathcal{U}_{pco} \times \mathcal{S}_{pc}$ given an initial state x_0 .

Definition 1.1. We say that a polynomial system $\mathcal{P} \in \mathcal{F}_0$ is a polynomial embedding of a system $\Sigma \in \mathcal{F}$ if for any solution $(x, u, y) \in \mathcal{X}_{ac} \times \mathcal{U}_{pco} \times \mathcal{S}_{pc}$ of Σ , there is a continuous injection F such that $(F(x), u, y) \in F(\mathcal{X}_{ac}) \times \mathcal{U}_{pco} \times \mathcal{S}_{pc}$ is a solution of \mathcal{P} . We will denote such systems by $\mathcal{P}(\Sigma)$.

1.2. The algebra of input-output maps

In this section we introduce causal analytic i-o maps and verify that under some assumptions their observation algebras are well-defined. For the latter, we make use of some technical definitions (see [38, Definitions 4.2, 4.3]), allowing us to define derivations and show that the ring of input-output maps forms an integral domain structure.

To this end, we remark that any element u of \mathcal{U}_{pco} is completely determined by determined by tuples $t_1, \dots, t_l \in [0, T], \alpha_1, \dots, \alpha_l \in U$ for some $l \geq 1$ such that:

$$u_{t_1, \dots, t_l}^{\alpha_1, \dots, \alpha_l}(t) = \begin{cases} \alpha_i & \text{if } t \in [T_{i-1}, T_i[, i \in [l] \\ \alpha_l & \text{if } t \geq T_l \end{cases}$$

for some interval decomposition

$$T_0 = 0, T_i = \sum_{j=1}^i t_j, i \in [l] \quad T = T_l.$$

Now, for $p : \mathcal{U}_{pco} \rightarrow \mathcal{S}_{pc}$, each component of $p(u) := (p_1(u), \dots, p_p(u))$ is of the form $p_k(u_{t_1, \dots, t_l}^{\alpha_1, \dots, \alpha_l})$. This remark will be used to define the class of causal and analytic input-output maps. In turn, any input-output map realized by a system from \mathcal{F} belongs to this latter class.

Definition 1.2. A map $p : \mathcal{U}_{pco} \rightarrow \mathcal{S}_{pc}$ is

- causal if, $\forall u, v \in \mathcal{U}_{pco}, t \geq 0$, we have:
 $(u(s) = v(s), \forall s \in [0, t]) \Rightarrow (p(u)(s) = p(v)(s), \forall s \in [0, t]).$
- analytic if $\forall k \in [p], \alpha_1, \dots, \alpha_l \in U, l > 0$, the following function is analytic

$$\begin{aligned} \phi_{p,k,\alpha_1, \dots, \alpha_l} : ([0, +\infty[)^l &\rightarrow \mathbb{R} \\ (t_1, \dots, t_l) &\mapsto p_k(u_{t_1, \dots, t_l}^{\alpha_1, \dots, \alpha_l})(T_l). \end{aligned}$$

Set $S^0 = \mathbb{R}$. Denote $\mathcal{A}(\mathcal{U}_{pco})$ the set of causal analytic maps $p : \mathcal{U}_{pco} \rightarrow \mathcal{S}_{pc}^0$. It is naturally a \mathbb{R} -algebra.

For our purposes, we need to define a derivation operation on $\mathcal{A}(\mathcal{U}_{pco})$. To this end, we will use the following observation: for each $\alpha \in U, t > 0$ and $u \in \mathcal{U}_{pco}$, we can construct an element $u_{t,\alpha} \in \mathcal{U}_{pco}$ by setting

$$u_{t,\alpha}(\tau) = \begin{cases} u(\tau) & \tau \in [0, t[\\ \alpha & \tau \geq t. \end{cases}$$

Then for all $\alpha \in U$, we define the map

$$D_\alpha : \mathcal{A}(\mathcal{U}_{pco}) \rightarrow \mathcal{A}(\mathcal{U}_{pco}) \quad \varphi \mapsto D_\alpha(\varphi)$$

given, for all $u \in \mathcal{U}_{pco}, t \geq 0$ by,

$$(D_\alpha \varphi(u))(t) = \frac{d}{ds} (\varphi(u_{t,\alpha})(t+s)) \Big|_{s=0}.$$

The map D_α is a well-defined derivation.

Next, we will argue that $\mathcal{A}(\mathcal{U}_{pco})$ is an integral domain. To this end notice that the set \mathcal{U}_{pc}^f of piecewise constant functions over finite intervals is closed by interval truncation, concatenation and piecewise time dilatation. Hence, \mathcal{U}_{pc}^f is a set of admissible inputs in the sense of [38, Definition 4.1]. Consequently, we can use the definition of analytic functions in the sense of [38, Definition 4.3]. Let us denote by $\mathcal{A}(\mathcal{U}_{pc}^f, \mathbb{R})$ the set of analytic functions $\mathcal{U}_{pc}^f \rightarrow \mathbb{R}$ in the sense of [38, Definition 4.3]. From [38, Theorem 4.4] it follows that the ring $\mathcal{A}(\mathcal{U}_{pc}^f, \mathbb{R})$ is an integral domain. Below we will present an \mathbb{R} -algebra isomorphism between $\mathcal{A}(\mathcal{U}_{pco})$ and $\mathcal{A}(\mathcal{U}_{pc}^f, \mathbb{R})$. The existence of such an isomorphism then implies that $\mathcal{A}(\mathcal{U}_{pco})$ is also an integral domain. In order to define this isomorphism, we observe that for each $v : [0, T_v] \rightarrow U$ in \mathcal{U}_{pc}^f we can construct a unique element $u_v \in \mathcal{U}_{pco}$ by setting

$$u_v(t) = \begin{cases} v(t) & t \in [0, T_v[\\ v(T_v) & t \geq T_v. \end{cases}$$

Let us define the map

$$\begin{aligned} \ell : \mathcal{A}(\mathcal{U}_{pco}) &\longrightarrow \mathcal{A}(\mathcal{U}_{pc}^f, \mathbb{R}) \\ (\varphi : \mathcal{U}_{pco} \rightarrow \mathcal{S}_{pc}^0) &\longmapsto (\tilde{\varphi} : \mathcal{U}_{pc}^f \rightarrow \mathbb{R}) \end{aligned}$$

given, for each function $v : [0, T_v] \rightarrow U$ in \mathcal{U}_{pc}^f , by $\tilde{\varphi}(v) = \varphi(u_v)(T_v)$.

It then follows ℓ is an \mathbb{R} -algebra isomorphism and hence $\mathcal{A}(\mathcal{U}_{pco})$ is an integral domain.

The above discussion allow us to introduce the following.

Definition 1.3. Let $p : \mathcal{U}_{pco} \rightarrow \mathcal{S}_{pc}$ be analytic and causal. The observation algebra $\mathcal{A}_{obs}(p)$ of p is the smallest sub-algebra $\mathcal{A}(\mathcal{U}_{pco})$ containing each $p_k \in \mathcal{A}_{obs}(p)$ and closed under D_α , for all $\alpha \in U$. The field of fractions $\mathcal{Q}_{obs}(p)$ of $\mathcal{A}_{obs}(p)$ will be called observation field of p and we denote $trdeg \mathcal{A}_{obs}(p)$ the transcendence degree of $\mathcal{A}_{obs}(p)$ over \mathbb{R} .

Definition 1.4. Let $\Sigma \in \mathcal{F}$ be a system with initial state x_0 . It is called a (piecewise constant) realisation of a map $p : \mathcal{U}_{pco} \rightarrow \mathcal{S}_{pc}$ if for all $u \in \mathcal{U}_{pco}$ the unique solution (x, u, y) of Σ such that $x(0) = x_0$ satisfies $p(u) \equiv y$.

Remark 1.5. If a system Σ realizes an i-o map $p : \mathcal{U}_{pco} \rightarrow \mathcal{S}_{pc}$, then the polynomial embedding $\mathcal{P}(\Sigma)$, when it exists, also realizes p .

1.3. Minimality, reachability and observability of polynomial systems

Let $\Sigma \in \mathcal{F}$ be a system of format (m, n, p) as in the above subsection. The dimension $dim(\Sigma)$ of Σ is the dimension of its state-space.

A polynomial system \mathcal{P} realizing an i-o map p is *minimal* if there is no polynomial system \mathcal{P}' realizing p such that $\dim(\mathcal{P}') < \dim(\mathcal{P})$. Define the set of reachable states of a polynomial system \mathcal{P} as:

$$\mathbf{R}_{\mathcal{P}}(v_0) = \{v(t) \mid t \geq 0, (v, u, y) \text{ is a solution of } \mathcal{P}, v(0) = v_0\}$$

and recall from [6, Definition 4] that its *observation algebra* $\mathcal{A}_{obs}(\mathcal{P})$ is the smallest sub-algebra of the ring $R[X_1, \dots, X_n]$ which contains h_k , $k \in [p]$ and which is closed under taking the formal Lie derivatives with respect to the formal vector fields $f_\alpha = \sum_{i=1}^n P_{i,\alpha} \frac{\partial}{\partial X_i}$. Its fraction field $\mathcal{Q}_{obs}(\mathcal{P})$ will be called its *observation field*. Finally, \mathcal{P} is *minimal* if $\dim(\mathcal{P}) = \text{trdeg} \mathcal{A}_{obs}(\mathcal{P})$ (see [42, Lemma 1, Theorem 4] for details). Notice that the other implication is true for rational systems, but not for polynomial ones. A polynomial system \mathcal{P} is

- algebraically reachable, if there is no non-trivial polynomial which is zero on $\mathbf{R}_{\mathcal{P}}(v_0)$;
- accessible, if $\mathbf{R}_{\mathcal{P}}(v_0)$ contains an open subset of \mathbb{R}^n ;
- algebraically observable, if $\mathcal{A}_{obs}(\mathcal{P}) = R[X_1, \dots, X_n]$ (see [38]);
- semi-algebraically observable if $\text{trdeg}(\mathcal{A}_{obs}(\mathcal{P})) = n$ (see [39]);
- observable, if for every two distinct initial states v_0, v'_0 there exists solutions (v, u, y) and (v', u, y') of \mathcal{P} such that $v(0) = v_0$, $v'(0) = v'_0$, and $y \neq y'$.

Accessibility implies algebraic reachability and algebraic observability implies semi-algebraic observability, and semi-algebraic observability implies observability. A polynomial system \mathcal{P} is minimal if it is algebraically reachable and algebraically observable (see [42, Theorem 4] and [5] for details). Notice that the other implication is true for rational systems but not for polynomial ones. Algebraic, rational and semi-algebraic observability and algebraic reachability of polynomial systems can be checked using methods of computational algebra [40].

Define the reachable set of a system Σ in \mathcal{F} of format (m, n, p) by

$$\mathbf{R}_{\Sigma}(s_0) = \{s(t) \mid t \geq 0, (s, u, y) \text{ is a solution of } \Sigma, s(0) = s_0\}.$$

We will say that Σ is

- accessible, if $\mathbf{R}_{\Sigma}(s_0)$ contains an open subset of \mathbb{R}^n ;
- algebraically reachable if there is no non-trivial polynomial which is zero on $\mathbf{R}_{\Sigma}(s_0)$;
- span-reachable, if the linear span of the elements $\mathbf{R}_{\Sigma}(s_0)$ is \mathbb{R}^n ;
- reachable if there exist no linear function which is zero on $\mathbf{R}_{\Sigma}(s_0)$;
- weakly observable if for every initial state $\hat{s} \in \mathbb{R}^n$ there is an open subset V of \mathbb{R}^n such that $\hat{s} \in V$ and for every $\hat{s} \neq \bar{s} \in V$, there exist solution (s, u, y) and (s', u, y') of Σ , with $s(0) = \hat{s}$ and $s'(0) = \bar{s}$, such that $y \neq y'$;

- observable if for every initial state $\hat{s} \in \mathbb{R}^n$, $V = \mathbb{R}^n$ in the latter definition.

Accessibility implies algebraic reachability which in turn implies span-reachability. Observability implies weak observability. Finally, if the system Σ realizes an i-o map p , is accessible and weakly observable, then it is minimal dimensional among all the systems from \mathcal{F} realizing p (see [29, Theorem 1.12]).

2. Realization theory of dynamical neural networks

2.1. Recurrent neural nets and LSTM embeddings

Let us denote by $\mathcal{F}_1 \subset \mathcal{F}$ the class of systems described by differential equations

$$\Sigma : \begin{cases} \dot{x}(t) = \vec{\sigma}(Ax(t) + Bu(t)) \\ y(t) = Cx(t) \end{cases} \quad t \geq 0, \quad (5)$$

with initial condition $x(0) = x_0 \in \mathbb{R}^n$ and where

- $\sigma \in C_1^\omega(\mathbb{R})$ is Lipschitz continuous,
- $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times m}$ and $C \in \mathbb{R}^{p \times n}$ are matrices.

Definition 2.1. An element of \mathcal{F}_1 will be called a recurrent neural ODE (ODE-RNN). We will identify such systems with tuples $\Sigma = (A, B, C, \sigma, x_0)$ and the triple (m, n, p) will be its format.

Let us denote by \mathcal{F}_2 the subclass of \mathcal{F} of systems described by differential equations of the form

$$\Sigma : \begin{cases} \dot{x}(t) = \mathfrak{U}^0 x(t) + g^2(t) \odot x(t) + g^3(t) \odot g^1(t) \\ \dot{z}(t) = g^4(t) \\ y(t) = Cs(t) \end{cases} \quad t \geq 0, \quad (6)$$

with initial condition $s(0) = s_0 = (x_0^T, z_0^T)^T \in \mathbb{R}^{2n}$, where

- $g^i(t) := \vec{\sigma}_i(\mathfrak{U}^i h(t) + \mathfrak{B}^i u(t) + b^i)$, for $i \in [4]$,
- $u(t) \in \mathbb{R}^m$, $s(t) = (x(t)^T, z(t)^T)^T \in \mathbb{R}^{2n}$, $y(t) \in \mathbb{R}^p$,
- $h(t) = z(t) \odot \vec{\sigma}_5^2(x(t))$,
- $\sigma = \{\sigma_1, \sigma_2, \sigma_3, \sigma_4, \sigma_5\} \subset C_1^\omega(\mathbb{R})$ are all Lipschitz continuous,
- $\mathfrak{U} = \{\mathfrak{U}^0, \mathfrak{U}^1, \mathfrak{U}^2, \mathfrak{U}^3, \mathfrak{U}^4\} \subset \mathbb{R}^{n \times n}$,
- $\mathfrak{B} = \{\mathfrak{B}^1, \mathfrak{B}^2, \mathfrak{B}^3, \mathfrak{B}^4\} \subset \mathbb{R}^{n \times m}$, and $C \in \mathbb{R}^{p \times 2n}$,
- $b = \{b^1, b^2, b^3, b^4\} \subset \mathbb{R}^n$.

Definition 2.2. An element of \mathcal{F}_2 will be called a long short-term neural ODE (ODE-LSTM). We identify such systems with tuples $\Sigma = (\mathfrak{U}, \mathfrak{B}, b, C, \sigma, s_0)$, set the triple $(m, 2n, p)$ to be its format and define the (ordered) set σ to be its activation.

We will restrict our attention to solutions of systems Σ in \mathcal{F}_1 or in \mathcal{F}_2 of the form $(x, u, y) \in \mathcal{X}_{ac} \times \mathcal{U}_{pco} \times \mathcal{S}_{pc}$ and we recall that activations σ are assumed to be in $C_1^\omega(\mathbb{R})$ so that global existence and uniqueness of solutions is verified and determined by u and for some initial value (see [14] for details). Notice that the Euler discretization of such systems correspond exactly to residual RNNs and residual LSTMs (see 4.1 for details).

Theorem 2.3. *All ODE-RNNs and ODE-LSTMs in \mathcal{F}_1 and \mathcal{F}_2 have polynomial embeddings. Moreover, if an i-o map p has a realization by a ODE-LSTM, or a ODE-RNN then p is causal, analytic and $\text{trdeg } A_{obs}(p) < +\infty$.*

The proof of this theorem will be done in Subsection 3.1.

We highlight the fact that our definitions of polynomial embeddings are completely explicit and can be easily implemented algorithmically. Theorem 2.3 allows to infer qualitative properties on neural nets induced by properties of their polynomial embeddings as we will show in the next section.

2.2. Qualitative properties of ODE-LSTMs

A system Σ in \mathcal{F}_2 with given activation σ realizing an i-o map p is said to be σ -minimal there exists no Σ' in \mathcal{F}_2 with activation σ , such that Σ' is a realization of p and $\dim(\Sigma') < \dim(\Sigma)$.

Lemma 2.4. Assume that an i-o map p is realized by a system Σ in \mathcal{F}_2 with given activation function σ . If one of the conditions below holds, then Σ is a σ -minimal realization of p :

1. $\mathcal{P}(\Sigma)$ is a minimal realization of p ,
2. $\text{trdeg } A_{obs}(p) = \dim(\Sigma)$,
3. $\mathcal{P}(\Sigma)$ is semi-algebraically observable and algebraically reachable,
4. $\mathcal{P}(\Sigma)$ is algebraically observable and accessible.

Proof. The second point comes from the first and from [42, Proposition 6]. The rest of the proof is straightforward. \square

Proposition 2.5. *Let Σ be a system in \mathcal{F}_2 .*

1. *If $\mathcal{P}(\Sigma)$ is accessible, then Σ is also accessible.*
2. *If $\mathcal{P}(\Sigma)$ is algebraically reachable, then Σ is span-reachable. In particular, if $\mathcal{P}(\Sigma)$ is accessible, then Σ is span-reachable.*
3. *If $\mathcal{P}(\Sigma)$ is observable, then Σ is observable. In particular, if $\mathcal{P}(\Sigma)$ is algebraically observable, then Σ is observable.*
4. *If $\mathcal{P}(\Sigma)$ is semi-algebraically observable, then Σ is weakly observable.*

The proof of the above proposition will be done in Subsection 3.2. Notice that a similar result was established for \mathcal{F}_1 in [14] with the help of an auxiliary polynomial embedding.

Note that the variables involved in the polynomial embedding $\mathcal{P}(\Sigma)$ can be naturally reordered, or simply reduced, following the expressions of the activation functions σ of Σ .

Accessibility and algebraic/semi-algebraic observability conditions for rational/ polynomial systems can be checked by using methods of computer algebra [40]. In contrast, for checking accessibility and (weak) observability of an ODE-LSTM

the only systematic tools are the rank conditions [24, Theorems 2.2, 2.5, 3.1, 3.5] or [27, Corollaries 2.2.5, 2.3.5], which are not computational effective manner for analytic σ . Notice that minimality of $\mathcal{P}(\Sigma)$ is a much weaker condition than accessibility and weak observability of Σ . This suggests that using realization theory of polynomial systems is likely to yield more useful results for ODE-LSTMs than using realization theory of general analytic systems.

3. Proofs

3.1. Proof of Theorem 2.3

Let us prove that all ODE-RNN have polynomial embeddings. Let $\Sigma = (A, B, C, \sigma, x_0)$ be an ODE-RNN with format (m, n, p) . Denote $L = Kn + n$ and consider the bijection

$$\begin{aligned} \phi : [K] \times [n] &\longrightarrow [Kn] \\ (r, j) &\longmapsto \phi(r, j) := r + K(j - 1) \end{aligned}$$

For L formal symbols X_γ , let us write $X_{\phi(j,r)}$ for the unique index $\gamma \in [Kn]$ such that $\phi(j, r) = \gamma$.

Then one can construct an associated polynomial system $\mathcal{P}(\Sigma) = (\{P_\alpha\}_\alpha, h, v_0)$ with

- $P_\alpha \in \mathbb{R}[X_1, \dots, X_L; \mathbb{R}^{Kn}]$
- $h \in \mathbb{R}[X_1, \dots, X_L; \mathbb{R}^p]$
- $v_0 \in \mathbb{R}^L$.

as follows:

$$\begin{aligned} P_{\phi(j,r),u(t)} &= \tilde{P}(X_{j,r}) \left(\sum_{k=1}^n a_{j,k} \tilde{P}_0(X_{k,u(t)}) \right) \\ P_{Kn+j,r} &= \tilde{P}_0(X_{j,r}), \\ h_k &= \sum_{j=1}^n c_{k,j} X_{j+Kn} \\ (v_0)_{\phi(j,r)} &= \sigma(e_j^T (Ax_0 + B\alpha_r)), \\ (v_0)_{Kn+j} &= e_j^T x_0. \end{aligned}$$

where $k \in [p]$, $j \in [n]$, $r \in [K]$ and $u(t)$ ranges the K - components of $\{P_\alpha\}_\alpha$ and where \tilde{P}, \tilde{P}_0 are polynomials in one variables defining a polynomial system, which by [14, Lemma 1] is equivalent to Assumption (2). By [14, Lemma 2], if (x, u, y) is a solution of a ODE-RNN Σ , then $(F(x), u, y)$ is a solution of $\mathcal{P}(\Sigma)$ where $F : \mathbb{R}^n \rightarrow \mathbb{R}^L$ is given by $F(x) = (z_1, \dots, z_{nK}, x^T)^T$, where $z_{\phi(j,\alpha)} = \sigma(e_j^T (Ax + B\alpha))$.

Let us now prove that all ODE-LSTMs have polynomial embeddings. Let $\Sigma = (\mathcal{U}, \mathcal{B}, b, C, \sigma, s_0)$ be a ODE-LSTM with format $(m, 2n, p)$, denote $U = \{\alpha_1, \dots, \alpha_K\}$ its input space and $s(t) = (x(t)^T, z(t)^T)^T$ its state trajectory. Consider the ordering

$$\begin{aligned} \phi : [4] \times [n] \times [K] &\longrightarrow [4nK], \\ (l, j, r) &\longmapsto \phi(l, j, r) := j + n(l - 1) + 4n(r - 1). \end{aligned}$$

Recall the discussion after (6) the function $h : [0, +\infty[\rightarrow \mathbb{R}^n$, i.e., $h(t)_j = z_j(t)\sigma_5(x_j(t))$, $j = 1, 2, \dots, n$. In particular, there

exists a function $\chi : \mathbb{R}^{2n} \rightarrow \mathbb{R}^n$, $\chi(s(t)) = h(t)$. Clearly, we can defined a function $R^l : U \times \mathbb{R}^{2n} \rightarrow \mathbb{R}^n$ such that for $\alpha \in U$, $R(\alpha, s)$ is polynomial in s and

$$R^l(\alpha, s(t)) = \mathfrak{U}^l \underbrace{\chi(s(t))}_{h(t)} + \mathfrak{W}^l \alpha + b^l, \quad l \in [4].$$

By writing $\zeta_{\phi(l,j,r)}(s(t)) = \sigma_l(e_j^T(R^l(\alpha_r, s(t))))$ we can define a map $F : \mathbb{R}^{2n} \rightarrow \mathbb{R}^{4nK+3n}$ such that

$$F(s(t)) = (\zeta_1(s(t)), \dots, \zeta_{4nK}(s(t)), \vec{\sigma}_5^T(x(t)), s(t)).$$

Define $v(t) = F(s(t))$. Its coordinates are explicitly written as

$$\begin{cases} v_{\phi(l,j,r)}(t) = \sigma_l(e_j^T(R^l(\alpha_r, s(t)))) , \\ v_{4nK+j}(t) = \sigma_5(x_j(t)), \quad j \in [n] \\ v_{4nK+n+k}(t) = s_k(t) \quad k \in [2n] \end{cases} \quad (7)$$

Let us ease the notation and write

- $v_{\phi(l,j,r)}(t) = v_{(l-1)n+j,r}(t)$,
- $v_{\alpha_r}(t) = (v_{1,r}(t), \dots, v_{4n,r}(t))$,
- $v_{4nK+*}(t) = (v_{4nK+1}(t), \dots, v_{4nK+n}(t))$,
- $v(t) = (v_{\alpha_1}(t), \dots, v_{\alpha_K}(t), v_{4nK+*}(t), x(t)^T, z(t)^T)^T$
- $x_j(t) = v_{4nK+n+j}(t)$ and $z_j(t) = v_{4nK+2n+j}(t)$,

Set $\mathfrak{U}^l = (\mathfrak{U}_{i,j}^l)_{i,j=1}^n$, $0 \leq l \leq 4$, and $\sigma_k^{(1)} = P_k(\sigma_k)$, $k \in [5]$. For each $r = 1, 2, \dots, K$, define the polynomials $Q_{j,r}$, $Q_{5n+j,r}$, $Q_{ln+j,r}$, $j = 1, \dots, n$, $l \in [4]$, as follows:

- $Q_{j,r}(v(t)) = \sum_{i=1}^n \mathfrak{U}_{j,i}^0 x_i(t) + v_{n+j,r}(t)x_j(t) + v_{2n+j,r}(t)v_{j,r}(t)$
- $Q_{5n+j,r}(v(t)) = Q_{j,r}(v(t))P_5(v_{4nK+j}(t))$
- for $l \in [4]$,

$$Q_{ln+j,r}(v(t)) = P_l(v_{\phi(l,j,r)}(t)) \times \sum_{i=1}^n \left(\mathfrak{U}_{j,i}^l (v_{\phi(4,i,r)}(t)v_{4nK+i}(t) + z_j(t)Q_{i,r}(v(t))P_5(v_{4nK+i}(t))) \right)$$

We set $v(t)$ to be the state variables of the polynomial system $\mathcal{P}(\Sigma) = (\{Q_u\}_u, C, v_0)$ defined, for $l \in [4]$, $j \in [n]$, by

$$\dot{x}_j(t) = Q_{j,u(t)}(v(t)), \quad x_j(0) = e_j^T x_0,$$

$$\dot{z}_j(t) = v_{3n+j,u(t)}(t), \quad z_j(0) = e_j^T z_0$$

$$\dot{v}_{(l-1)n+j,u(t)}(t) = Q_{ln+j,u(t)}(v(t)),$$

$$\dot{v}_{4nK+j}(t) = Q_{5n+j,u(t)}(v(t))$$

$$v_{\phi(l,j,r)}(0) = \sigma_l(e_j^T(R^l(\alpha_r, s(0)))) ,$$

$$v_{4nK+j}(0) = \sigma_5(e_j^T x_0)$$

$$h(0) = z_0 \odot \vec{\sigma}_5^T(x_0),$$

$$y(t) = Cs(t)$$

with initial state $v_0 = v(0)$.

Lemma 3.1. Let $v(t) = F(s(t))$. If (s, u, y) is a solution of an ODE-LSTM Σ , then (v, u, y) is a solution of $\mathcal{P}(\Sigma)$.

Proof. Let (s, u, y) be a solution of Σ . To show that (v, u, y) is a solution of $\mathcal{P}(\Sigma)$, it suffices to prove that v satisfies the differential equation (4). For $u(t) \in U$, we calculate the first time-derivative of $x_j(t)$. Following (6), we know that we have

$$\begin{aligned} & \dot{x}_j(t) \\ &= \sum_{i=1}^n \mathfrak{U}_{j,i}^0 x_i(t) + g_j^2(t)x_j(t) \\ & \quad + g_j^2(t)\sigma_1(e_j^T(R^1(u(t), t))) \\ &= \sum_{i=1}^n \mathfrak{U}_{j,i}^0 x_i(t) + \sigma_2(e_j^T(R^2(u(t), t)))x_j(t) \\ & \quad + \sigma_3(e_j^T(R^3(u(t), t)))\sigma_1(e_j^T(R^1(u(t), t))) \\ &= Q_{j,u(t)}(v(t)), \end{aligned}$$

as desired. Now notice that $\dot{z}_j(t) = v_{\phi(4,j,r)}(t)$, where r, t are such that $u(t) = \alpha_r \in U$. Then we get

$$\begin{aligned} \dot{v}_{4nK+j}(t) &= \frac{d}{dt}(\sigma_5(x_j(t))) \\ &= \dot{x}_j(t)P_5(\sigma_5(x_j(t))) \\ &= Q_{j,u(t)}(v(t))P_5(v_{4nK+j}(t)). \end{aligned}$$

Now take $l \in [4]$. We obtain

$$\begin{aligned} \dot{v}_{\phi(l,j,r)}(t) &= \frac{d}{dt}(\sigma_l(e_j^T(R^l(\alpha_r, s(t)))) \\ &= P_l(v_{\phi(l,j,r)}(t)) \left(\sum_{i=1}^n \mathfrak{U}_{j,i}^l \dot{h}(t) \right) \\ &= P_l(v_{\phi(l,j,r)}(t)) \\ & \quad \left(\sum_{i=1}^n \mathfrak{U}_{j,i}^l (\dot{z}_i(t)\sigma_5(x_i(t)) + z_i(t)\dot{x}_i(t)P_5(\sigma_5(x_i(t)))) \right) \\ &= P_l(v_{\phi(l,j,r)}(t)) \left(\sum_{i=1}^n \mathfrak{U}_{j,i}^l (v_{3n+i,u(t)}(t)v_{4nK+i}(t) \right. \\ & \quad \left. + z_i(t)Q_{i,u(t)}(v(t))P_5(v_{4nK+i}(t))) \right) \\ &= Q_{ln+j,u(t)}(v(t)). \end{aligned}$$

This completes the proof of Lemma 3.1. \square

We observe that the map F is a smooth map, in particular it is a continuous map and Lemma 3.1 showed that, for all $u \in \mathcal{U}_{pco}$, if (s, u, y) is a solution of Σ , then (v, u, y) is a solution of $\mathcal{P}(\Sigma)$, with

$$v(t) = F(s(t)), \quad \forall t \geq 0, \quad (8)$$

where $s(t) = (x(t)^T, z(t)^T)^T \in \mathbb{R}^{2n}$.

Finally, if an i-o map $p : \mathcal{U}_{pco} \rightarrow \mathcal{S}_{pc}$ is realized by a system Σ in \mathcal{F}_1 or \mathcal{F}_2 , then the polynomial embedding $\mathcal{P}(\Sigma)$ also realizes p by [14, Lemma 2] and Lemma 3.1. Thus p is causal and analytic and $\text{trdeg } \mathcal{A}_{obs}(p) < \infty$ (by [5, Theorem 3], [41, Theorem 5.16]).

This concludes the proof of Theorem 2.3.

3.2. Proof of Proposition 2.5

First, we prove that, if $\mathcal{P}(\Sigma)$ is accessible, then Σ is also accessible. By definition of the map F from (7), we obtain

$$F(R_\Sigma(s_0)) = R_{\mathcal{P}(\Sigma)}(v_0), \quad (9)$$

where $R_\Sigma(s_0)$ and $R_{\mathcal{P}(\Sigma)}(v_0)$ are respectively the reachable set of Σ and of $\mathcal{P}(\Sigma)$, and $v_0 = \nu(0)$. Now suppose that the polynomial system $\mathcal{P}(\Sigma)$ is accessible, i.e. there exists a non-empty open set O included in $R_{\mathcal{P}(\Sigma)}(v_0)$. Thus $F^{-1}(O)$ is a non-empty open set (because F is a continuous map) included in $R_\Sigma(s_0)$, so that Σ is accessible.

Next, we prove that, if $\mathcal{P}(\Sigma)$ is algebraically reachable, then Σ is span-reachable. Suppose that $\mathcal{P}(\Sigma)$ is algebraically reachable, i.e. there is no non-trivial polynomial which vanishes on the reachable set $R_{\mathcal{P}(\Sigma)}(v_0)$. Take $u \in \mathcal{U}_{pco}$ such that (s, u, y) a solution of Σ , with $s(t) = (x(t)^T, z(t)^T)^T \in \mathbb{R}^{2n}$, for $t \geq 0$. Consider (v, u, y) the solution of $\mathcal{P}(\Sigma)$ obtained by Lemma 3.1. Assume that Σ is not span-reachable, i.e. there exist reals $\lambda_1, \dots, \lambda_{2n}$ such that

$$\sum_{j=1}^n \lambda_j x_j(t) + \sum_{j=1}^n \lambda_{n+j} z_j(t) = 0.$$

Then taking the first derivative of the above equation gives

$$\sum_{j=1}^n \lambda_j \mathcal{Q}_{j,u(t)}(v(t)) + \sum_{j=1}^n \lambda_{n+j} v_{3n+j,u(t)}(t) = 0,$$

which is a contradiction, because there exists at least one non-trivial polynomial (given by the above equation) vanishing on the reachable set $R_{\mathcal{P}(\Sigma)}(v_0)$.

Next, we prove that, if $\mathcal{P}(\Sigma)$ is observable, the Σ is also observable. Take $s_0, s'_0 \in \mathbb{R}^{2n}$ two initial states of Σ such that $s_0 \neq s'_0$. Thus we have $v_0 = F(s_0) \neq F(s'_0) = v'_0$, because the map F , defined in (7), is injective. As $\mathcal{P}(\Sigma)$ is observable, there exist solutions (v, u, y) and (v', u, y') of $\mathcal{P}(\Sigma)$ such that $v(0) = v_0$ and $v'(0) = v'_0$, and $y \neq y'$. By Lemma 3.1 and by (9), there exist solutions (s, u, y) and (s', u, y') of Σ with $s(0) = s_0$ and $s'(0) = s'_0$, satisfying

$$\forall t \geq 0, \quad F(s(t)) = v(t), \quad \text{and} \quad F(s'(t)) = v'(t), \quad (10)$$

and $y \neq y'$. Thus Σ is observable, as desired. Now if $\mathcal{P}(\Sigma)$ is algebraically reachable, then it is observable, by [5, Proposition 3]. Then Σ is observable, by the above arguments.

Finally, we prove that, if $\mathcal{P}(\Sigma)$ is semi-algebraically observable, then Σ is weakly observable. By [39, Proposition 4.20, Corollary 4.22], the polynomial system $\mathcal{P}(\Sigma)$ is weakly observable. Let $s_0 \in \mathbb{R}^{2n}$ be an initial state of Σ , and let $v_0 = F(s_0)$. As $\mathcal{P}(\Sigma)$ is weakly observable, there exist a open set V with $v_0 \in V$ such that, for all $v'_0 \neq v_0 \in V$, there exist solutions (v, u, y) and (v', u, y') satisfying $v(0) = v_0$, $v'(0) = v'_0$, and $y \neq y'$. We set $U = F^{-1}(V)$ which is an open set (because F is a continuous map) such that $s_0 \in U$. Take $s'_0 \neq s_0 \in U$, and set $v'_0 = F(s'_0) \in V$. By injectivity of F , $v'_0 \neq v_0 \in V$. Thus we can find solutions (v, u, y) and (v', u, y') of $\mathcal{P}(\Sigma)$ as above. Then there exist solutions (s, u, y) and (s', u, y') of Σ such that (10) holds. We know that $y \neq y'$. Thus Σ is weakly observable.

4. Examples

Linear systems are a particular case of ODE-LSTMs, by taking σ_1 the identity map, $\sigma_2 = \sigma_4 = 0$ (the constant functions equal to 0), $\sigma_3 = 1$ (the constant function equal to 1), $b^1 \in \mathbb{R}^n$ the trivial vector and $C \in \mathbb{R}^{p \times 2n}$ a matrix of the form $(\tilde{C}, 0)$ with $\tilde{C} \in \mathbb{R}^{p \times n}$. Also ODE-RNNs are particular cases of ODE-LSTMs by taking $\sigma_2, \sigma_4 = 0$, $\sigma_3 = 1$, σ_1 to be any non-constant continuous globally Lipschitz function, and taking b_1 to be trivial.

Remark 4.1. For a suitable choice of C and U^0 , and taking $\sigma_3 = \sigma_4 = \sigma_2$, the Euler discretization of an ODE-LSTM in \mathcal{F}_2 is given by

$$\Sigma_{\text{discretized}} : \begin{cases} x(k+1) = x(k) + f(k) \odot x(k) + i(k) \odot g^1(k) \\ f(k) = \vec{\sigma}_2(\mathcal{U}^2 h(k) + \mathbb{B}^2 u(k) + b^2) \\ i(k) = \vec{\sigma}_3(\mathcal{U}^3 h(k) + \mathbb{B}^3 u(k) + b^3) \\ z(k+1) = z(k) + \vec{\sigma}_2(\mathcal{U}^4 h(k) + \mathbb{B}^4 u(k) + b^4) \\ h(k) = z(k) \odot \vec{\sigma}_5(x(k)) \\ x(0) = x_0, \quad z_0 = z(0) \\ y(k) = z(k). \end{cases}$$

This is closely related to LSTM networks defined in [18], where, at the k th step, $x(k)$, $f(k)$, $i(k)$ are usually called respectively the cell, the forget gate and the input gate and $u(k)$, $z(k)$ respectively the input and the output. The presence of a skip connection makes this actually a *residual* LSTM which, being the discretization of an ODE, enjoys of (gradient) stable dynamics, contrary to vanilla LSTMs. In addition, our construction should be readily applicable to LEM networks defined in [47] which are also presented as discretized two-gated recurrent neural ODEs and to State-Space models [21, 20, 19].

Example 4.2. Let us exhibit an ODE-LSTM whose polynomial embedding is minimal. Set $\mathcal{U} = \{u\} \subset \mathbb{R}$ and consider:

$$\Sigma : \begin{cases} \dot{x}(t) = \sigma(x(t)z(t) + u)x(t) \\ \dot{z}(t) = 0 \\ x(0) = 0, \quad z(0) = a, \quad \text{with } a \neq 0 \\ y(t) = x(t), \end{cases} \quad (11)$$

where σ is the sigmoid function. Here $\sigma_1 = \sigma_3 = \sigma_4 = 0$, σ_5 is the identity, $U^2 = 1$, $W^2 = 1$, b^2 is the zero vector and $C = (1, 0) \in \mathbb{R}^{1 \times 2}$. We can rewrite (11) as:

$$\Sigma : \begin{cases} \dot{x}(t) = \sigma(ax(t) + u)x(t), \\ x(0) = 0, \\ y(t) = x(t). \end{cases} \quad (12)$$

Then, as $m(\sigma) = 2$, $\mathcal{P}(\Sigma)$ is given by

$$\mathcal{P}(\Sigma) : \begin{cases} \dot{v}_{1,u} = av_2 v_{1,u} (1 - v_{1,u}), \\ \dot{v}_2 = v_{1,u} v_2, \\ v_{1,u}(0) = \sigma(u), \quad v_2(0) = 0, \\ y(t) = h(v(t)) = v_2(t). \end{cases}$$

where $v(t) = (v_{1,u}(t), v_2(t))^T$, $v_{1,u}(t) = \sigma(ax(t) + u)$ and $v_2(t) = x(t)$, for $t \geq 0$. It is clear that $v_{1,u}$ and v_2 belong to the observation field $\mathcal{Q}_{obs}(\mathcal{P}(\Sigma))$.

Now we prove that $\dim(\mathcal{P}(\Sigma)) = 2$. Observe that

$$\forall t \geq 0, \quad \frac{\dot{v}_{1,u}(t)}{1 - v_{1,u}(t)} = av_2(t),$$

and notice that $v_{1,u}(t) < 1$, because the sigmoid function takes value in $]0; 1[$. By taking the primitives of both sides of the above equation, there is $c \in \mathbb{R}$ such that

$$\forall t \geq 0, \quad \ln(1 - v_{1,u}(t)) = av_2(t) + c,$$

which implies that

$$\forall t \geq 0, \quad v_{1,u}(t) = 1 - Ke^{av_2(t)}, \text{ with } K = e^c.$$

Thus, there is no non-trivial polynomial which vanishes on the set of reachable states of $\mathcal{P}(\Sigma)$, i.e. $\mathcal{P}(\Sigma)$ is algebraically reachable. We conclude that $\dim(\mathcal{P}(\Sigma)) = 2 = m(\sigma)$ which implies that Σ is minimal.

In what follows we set $\mathcal{U} = \{\alpha_1, \alpha_2\} \subset \mathbb{R}$ with $\alpha_1 \neq \alpha_2$.

Example 4.3. Let us exhibit a reduction method of polynomial ODE-LSTM embeddings. Consider:

$$\Sigma : \begin{cases} \dot{x}(t) = \sigma(h(t) + u(t))x(t) + \sigma(h(t) + u(t)), \\ \dot{z}(t) = \sigma(h(t) + u(t)), \\ h(t) = x(t)z(t), \\ x(0) = 0, z(0) = 0, \\ y(t) = z(t), \end{cases}$$

where we take $\sigma_1 = \sigma_2 = \sigma_4 = \sigma$ to be the sigmoid function, $\sigma_3 = 1$ and σ_5 the identity map. Here $n(\sigma) = 2$, $k(\sigma) = 1$ so that $m(\sigma) = 4$. For $k = 1, 2$ and $t \geq 0$, using Lemma 3.1, let us set $v_{1,\alpha_k}(t) = \sigma_1(x(t)z(t) + \alpha_k) = \sigma(x(t)z(t) + \alpha_k) = v_{2,\alpha_k}(t) = v_{4,\alpha_k}(t)$, $v_{3,\alpha_k}(t) = 1$, and $v_5(t) = x(t)$. Thus, $\mathcal{P}(\Sigma)$ is simply given by

$$\begin{cases} \dot{v}_{1,\alpha_1} = (v_{1,u(t)}v_2v_3 + v_{1,u(t)}v_3 + v_{1,u(t)}v_2)v_{1,\alpha_1}(1 - v_{1,\alpha_1}), \\ \dot{v}_{1,\alpha_2} = (v_{1,u(t)}v_2v_3 + v_{1,u(t)}v_3 + v_{1,u(t)}v_2)v_{1,\alpha_2}(1 - v_{1,\alpha_2}), \\ \dot{v}_2 = v_{1,u(t)}v_2 + v_{1,u(t)}, \\ \dot{v}_3 = v_{1,u(t)}, \\ v_{1,\alpha_k}(0) = \sigma(\alpha_k), v_2(0) = v_3(0) = 0, \text{ for } k = 1, 2, \\ y(t) = v_3(t), \end{cases}$$

where we set $x(t) = v_2(t)$ and $z(t) = v_3(t)$.

Example 4.4. Let us now exhibit an accessible ODE-LSTM whose polynomial embedding is not accessible:

$$\Sigma : \begin{cases} \dot{x}(t) = \sigma(h(t) + u(t))^2, \\ \dot{z}(t) = \sigma(h(t) + u(t)), \\ h(t) = x(t)z(t), \\ s_0 = (x(0), z(0)) = (0, 0)^T, \\ y(t) = z(t). \end{cases} \quad (13)$$

In this case, $n(\sigma) = 2$. We first prove that Σ is accessible. Let $f_{\alpha_1}, f_{\alpha_2} : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ be vector fields generated by Σ . We denote by $\mathcal{L}_\Sigma(s_0)$ the smallest Lie algebra containing $f_{\alpha_1}, f_{\alpha_2}$ and closed by Lie brackets. We then have

$$\begin{cases} \frac{1}{\sigma(\alpha_2)}f_{\alpha_2}(s_0) - \frac{1}{\sigma(\alpha_1)}f_{\alpha_1}(s_0) = (\sigma(\alpha_2) - \sigma(\alpha_1))(1, 0)^T \in \mathbb{R}^2, \\ \frac{1}{\sigma(\alpha_2)^2}f_{\alpha_2}(s_0) - \frac{1}{\sigma(\alpha_1)^2}f_{\alpha_1}(s_0) = \frac{\sigma(\alpha_1) - \sigma(\alpha_2)}{\sigma(\alpha_1)\sigma(\alpha_2)}(0, 1)^T \in \mathbb{R}^2. \end{cases}$$

As $\sigma(\alpha_1), \sigma(\alpha_2) > 0$ and $\sigma(\alpha_1) \neq \sigma(\alpha_2)$ because $\alpha_1 \neq \alpha_2$ and σ is bijective and takes values in $]0; 1[$, then $\dim \mathcal{L}_\Sigma(s_0) = 2 = n(\sigma)$. By [28, Theorem 3.10], Σ is accessible. Now $\mathcal{P}(\Sigma)$ is given by

$$\begin{cases} \dot{v}_{1,\alpha_1} = (v_{1,u(t)}^2v_3 + v_{1,u(t)}v_2)v_{1,\alpha_1}(1 - v_{1,\alpha_1}), \\ \dot{v}_{1,\alpha_2} = (v_{1,u(t)}^2v_3 + v_{1,u(t)}v_2)v_{1,\alpha_2}(1 - v_{1,\alpha_2}), \\ \dot{v}_2 = v_{1,u(t)}^2, \\ \dot{v}_3 = v_{1,u(t)}, \\ v_{1,\alpha_k}(0) = \sigma(\alpha_k), v_2(0) = v_3(0) = 0, \text{ for } k = 1, 2, \\ y(t) = v_3(t), \end{cases}$$

where, for $t \geq 0, \alpha \in \mathcal{U}$, we set $v_{1,\alpha}(t) = \sigma(x(t)z(t) + \alpha)$, $v_2(t) = x(t)$ and $v_3(t) = z(t)$. We denote $v_0 = (\sigma(\alpha_1), \sigma(\alpha_2), 0, 0)^T \in \mathbb{R}^4$ the initial state of $\mathcal{P}(\Sigma)$ and let $g_{\alpha_1}, g_{\alpha_2} : \mathbb{R}^4 \rightarrow \mathbb{R}^4$ be vector fields generated by the polynomial system $\mathcal{P}(\Sigma)$. We denote $\mathcal{L}_{\mathcal{P}(\Sigma)}(v_0)$ the smallest Lie algebra containing $g_{\alpha_1}, g_{\alpha_2}$ and closed by Lie brackets. It is easy to prove that $(g_{\alpha_1}(v_0), g_{\alpha_2}(v_0))$ is linearly independent, so that $2 \leq \dim \mathcal{L}_{\mathcal{P}(\Sigma)}(v_0)$.

Example 4.5. Let us exhibit an ODE-LSTM Σ which, seen as an analytic system, is both accessible and weakly observable and thus is minimal:

$$\Sigma : \begin{cases} \dot{x}(t) = \sigma(h(t) + u(t))x(t) + \sigma(h(t) + u(t))^2, \\ \dot{z}(t) = \sigma(h(t) + u(t)), \\ h(t) = x(t)z(t), \\ x(0) = 0, z(0) = 0, \\ y(t) = z(t), \end{cases} \quad (14)$$

where $\sigma_1 = \sigma_2 = \sigma_3 = \sigma_4 = \sigma$ is the sigmoid function and σ_5 the identity map. Here $n(\sigma) = 2$ and $\mathcal{P}(\Sigma)$ is given by:

$$\begin{cases} \dot{v}_{1,\alpha_1} = (v_{1,u(t)}v_2v_3 + v_{1,u(t)}^2v_3 + v_{1,u(t)}v_2)v_{1,\alpha_1}(1 - v_{1,\alpha_1}), \\ \dot{v}_{1,\alpha_2} = (v_{1,u(t)}v_2v_3 + v_{1,u(t)}^2v_3 + v_{1,u(t)}v_2)v_{1,\alpha_2}(1 - v_{1,\alpha_2}), \\ \dot{v}_2 = v_{1,u(t)}v_2 + v_{1,u(t)}^2, \\ \dot{v}_3 = v_{1,u(t)}, \\ v_{1,\alpha_1}(0) = \sigma(\alpha_1), v_{1,\alpha_2}(0) = \sigma(\alpha_2), v_2(0) = v_3(0) = 0, \\ y(t) = v_3(t). \end{cases}$$

Now, for $t \geq 0$, set $v_{1,\alpha_k}(t) = \sigma(x(t)z(t) + \alpha_k)$, where $k = 1, 2$, and set $v_2(t) = x(t)$ and $v_3(t) = z(t)$.

Notice that equations (13) hold in this case, so $\dim \mathcal{L}_\Sigma(s_0) = 2 = n(\sigma)$. Then, by [28, Theorem 3.10], we conclude that Σ is accessible (i.e. its reachable set from $s_0 = (0, 0)^T$ contains a non-empty open set).

Now let's prove that Σ is weakly observable by proving that $\mathcal{P}(\Sigma)$ is semi-algebraically observable. Denote $g_{\alpha_1}, g_{\alpha_2} : \mathbb{R}^4 \rightarrow \mathbb{R}^4$ the vector fields generated by $\mathcal{P}(\Sigma)$, and L_{g_α} the Lie derivative operator along g_α , for $\alpha \in \mathcal{U}$. The output map of $\mathcal{P}(\Sigma)$

is $h = v_3$. It is then clear that v_3 belongs to the observation algebra of $\mathcal{P}(\Sigma)$. Moreover we have

$$L_{g_{\alpha_1}} h = v_{1,\alpha_1}, \quad L_{g_{\alpha_2}} h = v_{1,\alpha_2},$$

which shows that $v_{1,\alpha_1}, v_{1,\alpha_2}$ also belong to the observation algebra $\mathcal{A}_{obs}(\mathcal{P}(\Sigma))$ of $\mathcal{P}(\Sigma)$ as the latter algebra is closed under Lie derivatives along $g_{\alpha_1}, g_{\alpha_2}$. Thus, in $\mathcal{Q}_{obs}(\mathcal{P}(\Sigma))$ we have

$$v_2 = \frac{\frac{L_{g_{\alpha_2}} L_{g_{\alpha_1}} h}{L_{g_{\alpha_1}} h(1-L_{g_{\alpha_1}})} - h(L_{g_{\alpha_2}} h)^2}{h(L_{g_{\alpha_2}} h) + L_{g_{\alpha_2}} h}.$$

This proves that $\mathcal{P}(\Sigma)$ is semi-algebraically observable. Thus Σ is weakly observable by Lemma 2.5. As Σ is seen as an analytic system and is accessible and weakly observable, it is then minimal.

5. Conclusions and perspectives

We have shown that i-o maps realized by large classes of recurrent neural ODEs (namely ODE-RNNs and ODE-LSTMs) can be represented by polynomial systems, and we used this fact to derive necessary and sufficient conditions for the existence of realizations by such systems and their minimality. Future research will be directed towards improving these results to derive a complete realization theory for ODE-LSTMs and apply them to formulating theoretical guarantees for learning ODE-LSTMs.

Acknowledgments

The first author thanks Pierre Marion for useful comments which helped improving the clarity of the exposition. This work has been supported by the French government under the "France 2030" program, as part of the SystemX Technological Research Institute.

References

[1] F. Albertini, P. Dai Pra, Forward accessibility for recurrent neural networks, *IEEE Transactions on Automatic Control*, 1995, pages 1962-1968, volume 40.

[2] F. Albertini, E.D. Sontag, For Neural Networks, Function Determines Form, *Neural Networks* 6(1993): 975-990.

[3] F. Albertini and E.D. Sontag, State observability in recurrent neural networks, *Systems f3 Control Letters* 22(1994): 235-244.

[4] S. Bai, J. Z. Kolter, V. Koltun, Deep equilibrium models. *Advances in Neural Information Processing Systems*, 32, 2019.

[5] Z. Bartoszewicz, Minimal polynomial realizations, *Mathematics of control, signals, and systems*, 1988, pages 227-237, volume 1.

[6] Z. Bartoszewicz, Rational systems and observation fields, *Systems & Control Letters* 9, 1987, pages 379-386.

[7] Y. Bengio, P. Simard, and P. Frasconi, Learning long-term dependencies with gradient descent is difficult, *IEEE Transactions on neural networks*, 5(2), pages 157-166, 1994.

[8] S. P. Bhat and D. S. Bernstein, Geometric homogeneity with applications to finite time stability, *Mathematics of Control, Signals and Systems*, vol. 17, pp. 101 – 127, 2005.

[9] J. Bochnak, M. Coste, and M.-F. Roy, *Real algebraic geometry*, Springer-Verlag, Berlin Heidelberg, 1998.

[10] A. Caterini and D. E. Chang, Deep Neural Networks in a Mathematical Framework, Springer, *SpringerBriefs in Computer Science*, 2018.

[11] R.T. Chen, Y. Rubanova, J. Bettencourt and D. Duvenaud, *Neural ordinary differential equations*. Advances in neural information processing systems, 31, 2018.

[12] D. Cox, J. Little, and D. O'Shea, *Ideals, varieties, and algorithms: An introduction to computational algebraic geometry and commutative algebra*, Springer, third edition, 2007.

[13] P. D'Alessandro, A. Isidori and A. Ruberti, Realization and structure theory of bilinear dynamical systems. *SIAM J. Control and Optimization*, 517–535, 1974.

[14] T. Defourneau, M. Petreczky, Realization theory of recurrent neural networks and rational systems, IEEE 58th Conference on Decision and Control (CDC), 2019.

[15] A. Fermanian, P. Marion, J.-P. Vert and G. Biau, *Framing RNN as a kernel method: A neural ODE approach* Advances in Neural Information Processing Systems 34, 2021.

[16] F.A. Gers, N. Schraudolph, and J. Schmidhuber, Learning precise timing with LSTM recurrent networks, *Journal of Machine Learning Research*, 3, pages 115-143, 2002.

[17] F.A. Gers, and J. Schmidhuber, LSTM recurrent networks learn simple context free and context sensitive languages, *IEEE Transactions on Neural Networks*, 12(6), pages 1333-1340, 2001.

[18] F.A. Gers, and J. Schmidhuber, Recurrent nets that time and count, *Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks. IJCNN 2000. Neural Computing: New Challenges and Perspectives for the New Millennium*, volume 3, 2000.

[19] Goel, Karan and Gu, Albert and Donahue, Chris and Ré, Christopher *It's Raw! Audio Generation with State-Space Models*, International Conference on Machine Learning (ICML), 2022.

[20] Gu, Albert and Goel, Karan and Ré, Christopher, *Efficiently Modeling Long Sequences with Structured State Spaces*, ICLR, 2022.

[21] Gu, Albert and Johnson, Isys and Goel, Karan and Saab, Khaled and Dao, Tri and Rudra, Atri and Ré, Christopher, *Combining Recurrent, Convolutional, and Continuous-time Models with Linear State-Space Layers*, Advances in neural information processing systems, 34, 2021.

[22] J. Hanson, M. Raginsky, E. Sontag *Learning Recurrent Neural Net Models of Nonlinear Systems*, Proceedings of the 3rd Conference on Learning for Dynamics and Control, PMLR 144, 2021.

[23] K. He, X.Zhang, S. Ren and J. Sun, Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, 2016.

[24] R. Hermann, and A. J. Krener, Nonlinear controllability and observability, *IEEE Transactions on automatic control*, 1977, pages 728-740, volume 22.

[25] S. Hochreiter, *Untersuchungen zu dynamischen neuronalen Netzen*, Institut für Informatik Technische Universität München, Germany, Ph.D thesis under the direction of W. Brauer, 1991.

[26] S. Hochreiter, and J. Schmidhuber, Long short term memory, *Neural Computation*, 9(8), pages 1735-1780, 1997.

[27] A. Isidori, *Nonlinear control systems*, Springer, third edition, 2013.

[28] B. Jakubczyk, *Introduction to geometric nonlinear control ; controllability and Lie bracket*, Lectures given at the Summer School on Mathematical Control Theory, Trieste 3-28 Septembre 2001, International Atomic Energy Agency (IAEA), Volumn 38, 2002.

[29] B. Jakubczyk, Realization theory for nonlinear systems: three approaches. In M. Fliess and M. Hazenwinkel, editors, *Algebraic and Geometric Methods in Nonlinear Control Theory*, pages 3–31. D. Reidel Publishing Company, Dordrecht, 1986.

[30] T. Kailath, *Linear Systems*, Prentice-Hall, New Jersey, 1979, ISBN 978-0-13-536961-6.

[31] R.E. Kalman, Mathematical description of linear dynamical systems. *SIAM J. Control and Optimization*, 1(2): 152–159, 1963.

[32] R.E. Kalman, *On minimal partial realization of a linear input-output map*, Topics in Mathematical Systems Theory, 1969.

[33] P. Kidger, J. Morrill, J. Foster and T. Lyons, *Neural Controlled Differential Equations for Irregular Time Series*, Advances in Neural Information Processing Systems, 33, 2020.

[34] E. Kunz, *Introduction to commutative algebra and algebraic geometry*, Birkhäuser, Boston, 1985.

[35] Z. Li, N. B. Kovachki, K. Azizzadenesheli, B. Liu, K. Bhattacharya, A. Stuart and A. Anandkumar *Fourier Neural Operator for Parametric Partial Differential Equations*. International Conference on Learning Representations, 2021.

- [36] A. Lindquist and G. Picci, *Linear Stochastic Systems*, Series in Contemporary Mathematics. Springer-Verlag Berlin Heidelberg, 2015, vol. 1.
- [37] L. Ljung, *System identification: theory for the user (second edition)*, Preice-Hall, Englewood Cliffs, NJ, 1999.
- [38] J. Němcová, *Rational Systems in Control and System Theory*, Centrum Wiskunde & Informatica (CWI), Amsterdam, Ph.D. thesis under the direction of Jan H. van Schuppen, 2009.
- [39] J. Němcová, M. Petreczky and J.H. van Schuppen, *Realization theory of Nash systems*, *SIAM J. Control & Optimization*, 2013, pages 3386-3414, volume 51.
- [40] J. Němcová, M. Petreczky, J. H. van Schuppen, Observability reduction algorithm for rational systems, *IEEE Conference on Decision and Control (CDC)*, 2016, pages 5738-5743.
- [41] J. Němcová, and J.H. van Schuppen, Realization theory for rational systems: The existence of rational realizations, *SIAM J. Control Optim.*, 2009, pages 2840-2856, volume 48.
- [42] J. Němcová, and J.H. van Schuppen, Realization theory for rational systems: Minimal rational realizations, *Acta Applicandae Mathematica*, 2010, pages 605-626, volume 110.
- [43] A. Pal, A. Edelman and Ch. Rackauckas, *Mixing Implicit and Explicit Deep Learning with Skip DEQs and Infinite Time Neural ODEs (Continuous DEQs)*, arXiv preprint arXiv:2201.12240, 2022.
- [44] Y. Qiao, E.D. Sontag, Further results on controllability of recurrent neural networks, *Systems & Control Letters* 36, 1999, pages 121-129.
- [45] L. Rosier, Homogeneous Lyapunov function for homogeneous continuous vector field, *Systems & Control Letters*, vol. 19, pp. 467 - 473, 1992.
- [46] Y. Rubanova, R.T.Q. Chen, D.K. Duvenaud, *Latent ordinary differential equations for irregularly-sampled time series* Advances in neural information processing systems 32, 2019.
- [47] T. K. Rusch, S. Mishra, N. B. Erichson and M. W. Mahoney, *Long Expressive Memory for Sequence Modeling*, International Conference on Learning Representations, 2022.
- [48] Y. Wang and E.D. Sontag, *Algebraic differential equations and rational control systems*, *SIAM J. Control Optim.*, 30(5):1126–1149, 1992.
- [49] E.D. Sontag, Y. Wang, and A. Megretski, Input classes for identification of bilinear systems, *IEEE Transactions Autom. Control*, 2009, pages 195–207, volume 54.
- [50] I. Sustkever, Training Recurrent Neural Networks, PhD thesis, University of Toronto, 2013.
- [51] B. Tzen, M. Raginsky, Neural stochastic differential equations: Deep latent gaussian models in the diffusion limit. arXiv preprint arXiv:1905.09883, 2019.
- [52] J.M. Van den Hof, System theory and system identification of compartmental systems, PhD thesis, University of Groningen, 1996.
- [53] W. Xu, R.T. Chen, X. Li and D. Duvenaud, *Infinitely Deep Bayesian Neural Networks with Stochastic Differential Equations*. International Conference on Artificial Intelligence and Statistics, 2022.
- [54] O. Zariski and P. Samuel, *Commutative algebra I, II.*, Springer, 1958.