



Realization theory of recurrent neural networks and rational systems

Thibault Defourneau, Mihaly Petreczky

► To cite this version:

Thibault Defourneau, Mihaly Petreczky. Realization theory of recurrent neural networks and rational systems. 58th IEEE Conference on Decision and Control (CDC), Dec 2019, Nice, France. 10.1109/CDC40024.2019.9029326 . hal-02398581

HAL Id: hal-02398581

<https://hal.science/hal-02398581>

Submitted on 31 Dec 2020

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 networks and rational systems

Thibault Defourneau, Mihaly Petreczky

Abstract—In this paper, we show that, under mild assumptions, input-output behavior of a continuous-time recurrent neural network (RNN) can be represented by a rational or polynomial non-linear system. The assumptions concern the activation function of RNNs, and they are satisfied by many classical activation functions. We also present an algorithm for constructing the polynomial and rational system. This embedding of RNNs into rational systems can be useful for stability, identifiability, and realization theory for RNNs, as these problems have been studied for polynomial/rational systems. In particular, we use this embedding for deriving necessary conditions for realizability of an input-output map by RNN, and for deriving sufficient conditions for minimality of an RNN.

I. INTRODUCTION

One of the challenges in machine learning is to provide a mathematical theory for analyzing learning algorithms. Recently, there has been a surge of interest in the use of neural networks, leading to the emergence of the field of deep learning. One of the most widespread models used in deep learning are recurrent neural networks (RNNs). RNNs can be seen as non-linear dynamical systems equipped with an internal state, input and output. Learning such an RNN from data is equivalent to estimating the parameters of the RNN, viewed as a dynamical system. That is, learning algorithms for RNNs correspond to system identification algorithms, and developing a mathematical theory for learning RNNs is equivalent to developing system identification for RNNs. There is a rich literature on system identification, in particular on system identification for linear systems [15]. Note that linear dynamical systems are a particular class of RNNs.

One of the principal building blocks of system identification theory for linear systems is realization theory. Realization theory can be viewed as an attempt to solve an idealized system identification problem, where there is infinite data, not modelling error, etc. In general, the aim of *realization theory* is to understand the relationship between an observed behavior and dynamical systems producing this observed behavior. For the particular case of RNNs, the main questions of realization theory can be stated as follows:

- 1) Which class of observed behaviors (input-output maps) can be represented by an RNN ?

Centre de Recherche en Informatique, Signal et Automatique de Lille (CRISTAL), CNRS, Ecole Centrale, Université de Lille, Avenue Carl Gauss, 59650 Villeneuve-d'Ascq, France (e-mail: thibault.defourneau@univ-lille.fr, mihaly.petreczky@centralelille.fr)

This work was partially funded by: (1) CPER Data project, which is co-financed by European Union with the financial support of European Regional Development Fund (ERDF), French State and the French Region of Hauts-de-France, (2) by CNRS project PEPS blanc BayesRealForRNN, and (3) by Agence Nationale pour la Recherche Project ROCC-SYS under the Grant agreement ANR-14-C27-0008

- 2) How can we characterize minimal RNNs (RNNs of the least complexity) representing a certain observed behavior ? What is the appropriate definition of minimality (smallest number of neurons, etc.) for RNNs, are minimal RNNs related by some transformation ?
- 3) Is there a constructive procedure for constructing a RNN representation from input-output behavior which can be proven to be mathematically correct ?

For linear systems, realization theory [12], [24] has been useful for system identification, for example it helped to address identifiability, canonical forms and gave rise to subspace identification algorithms. We expect that realization theory of RNNs will lead to similarly useful results for the latter dynamical systems.

In order to develop realization theory of RNNs, we embed RNNs into the class of rational systems, and then we use realization theory of rational/polynomial systems in order to derive new results on realization theory of RNNs.

In this paper, we consider RNNs in continuous-time, in order to avoid some technical difficulties encountered in the discrete-time case. Note that discrete-time RNNs can be viewed as discretizations of continuous-time RNNs.

In this paper we assume that the high-order derivatives of the activation function satisfy a polynomial equation. Several widely-used activation functions have this property.

- We show that an input-output map can be realized by a RNN, only if it can be realized by a rational system, i.e. a non-linear system defined by vector fields and readout maps which are fractions of polynomials. We present an explicit construction of such a rational system.
- We present a necessary condition for existence of a realization by RNNs, using results from realization theory of rational systems. This necessary condition is a generalization of the well-known rank condition for Hankel matrices of linear systems.
- We formulate sufficient conditions for observability/reachability/minimality of RNNs, using existing realization theory for rational systems [19], [16], [4], [22].

Note that RNNs could be viewed as analytic systems and one could try to apply realization theory of analytic systems [11], [10], [9]. However, realization theory for analytic systems is not computationally effective. In contrast, the conditions for observability/reachability/minimality of RNNs which are derived in this paper are computationally effective and they are less restrictive than those which can be obtained by viewing RNNs as analytic systems [11], [10], [9].

To the best of our knowledge, the results of the paper are new. RNNs have been widely used in the machine

learning literature, both in discrete-time and continuous-time, [25], [26]. Observability of RNNs was studied in [3], controllability in [21] and minimality in [2]. This paper was inspired by [3], [21], [2], but in contrast to [3], [21], [2], we do not use any assumption on the structure of the weight matrices, except for observability issues. This means that the results of this paper can be applied even when the results of [3], [21], [2] are not applicable.

In Section II we present the basic notation and terminology. In Section III we present the construction of the rational system which realizes the same input-output map as an RNN. In Section IV, we use the results of Section III to derive necessary conditions for existence of a realization by RNN. Finally, in Section V we use the results of Section III to derive sufficient conditions for minimality of RNNs. An extended version of this paper containing detailed proofs can be found in the technical report [8].

II. BASIC DEFINITIONS

In this section, we fix the notation used in the paper and we recall some algebraic tools necessary for this paper. Then we recall the definition of RNNs and of rational systems.

A. Preliminaries

We use the standard terminology and notation from commutative algebra and algebraic geometry see [14], [23], [6]. In particular, by $\mathbb{R}[X_1, \dots, X_n]$ we denote the algebra of real polynomials in n variables and by $R(X_1, \dots, X_n)$ we denote the quotient field of $\mathbb{R}[X_1, \dots, X_n]$. We refer to the element of $R(X_1, \dots, X_n)$ as rational functions in n variables.

B. Recurrent neural networks.

Below we define formally what we mean by recurrent neural networks in continuous-time. We will follow the notation of [3], [21], [2].

Definition 1: A recurrent neural network, abbreviated by RNN, with input-space $\mathcal{U} \subset \mathbb{R}^m$ and output-space \mathbb{R}^p , is a dynamical system

$$\Sigma : \begin{cases} \dot{x}(t) = \vec{\sigma}(Ax(t) + Bu(t)), & x(0) = x_0 \\ y(t) = Cx(t) \end{cases} \quad (1)$$

where

- $\sigma : \mathbb{R} \rightarrow \mathbb{R}$ is a continuous globally Lipschitz scalar function. It is called the *activation function*.
- $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times m}$ and $C \in \mathbb{R}^{p \times n}$ are matrices, called *weight matrices*,
- the map $\vec{\sigma} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is defined by

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

- $u(t)$ is an input, $x(t) \in \mathbb{R}^n$ is the state and $y(t)$ is the output at time t , and $x_0 \in \mathbb{R}^n$ is the initial-state.

We denote such a system by $\Sigma = (A, B, C, \mathcal{U}, \sigma, x_0)$.

If we take σ the identity map in Definition 1, it is the same as a linear system in control theory. Next we define formally what we mean by a solution and an input-output map of a system $\Sigma = (A, B, C, \mathcal{U}, \sigma, x_0)$. To this end, in the

sequel, we denote by $PC([0; +\infty[, X)$ the set of piecewise-continuous functions from $[0; +\infty[$ to $X \subseteq \mathbb{R}^k$, $k > 0$.

Definition 2: A triple (x, u, y) is a *solution* of an RNN $\Sigma = (A, B, C, \mathcal{U}, \sigma, x_0)$ if $u \in PC([0; +\infty[, \mathcal{U})$, $x : [0; +\infty[\rightarrow \mathbb{R}^n$, $y : [0; +\infty[\rightarrow \mathbb{R}^p$, x is absolutely continuous, and (1) holds.

Remark 1: As the activation function σ is globally Lipschitz, it follows that for every $u \in PC([0; +\infty[, \mathcal{U})$, there exist unique x, y such that (x, u, y) is a solution of Σ .

In this paper, we focus on solutions (x, u, y) of an RNN such that u is *piecewise constant*.

Notation 1 (\mathcal{U}_{pc}): We denote by \mathcal{U}_{pc} the set of all piecewise-constant functions of the form $u : [0; +\infty[\rightarrow \mathcal{U}$.

Definition 3: Let $p : \mathcal{U}_{pc} \rightarrow PC([0; +\infty[, \mathbb{R}^p)$ be an input-output map, and let $\Sigma = (A, B, C, \mathcal{U}, \sigma, x_0)$ be an RNN. Σ is said to be a *realization* of the input-output map p if for every $u \in \mathcal{U}_{pc}$, the unique solution (x, u, y) , $x(0) = x_0$ of Σ is such that $p(u) = y$.

C. Rational and polynomial systems.

Informally, rational respectively polynomial systems are control systems in continuous time, whose differential equations and readout maps are rational functions, i.e. they are fractions of polynomials, respectively polynomial functions.

Definition 4 (Polynomial and rational systems): A rational system with input-space $\mathcal{U} \subset \mathbb{R}^m$, state-space \mathbb{R}^n and output-space \mathbb{R}^p is a dynamical system as

$$\mathcal{R} : \begin{cases} \dot{v}_i(t) = \frac{P_{i,u(t)}(v(t))}{Q_{i,u(t)}(v(t))}, & i = 1, \dots, n, & v(0) = v_0 \\ y_k(t) = \frac{h_{k,1}(v(t))}{h_{k,2}(v(t))}, & k = 1, \dots, p \end{cases} \quad (2)$$

where

- $u(t)$ is the input, $v(t) = (v_1(t), \dots, v_n(t))^T$ is the state and $y(t) = (y_1(t), \dots, y_p(t))^T$ is the output at time t . Moreover $v_0 \in \mathbb{R}^n$ is the initial state;
- $h_{k,1}, h_{k,2}$, $k = 1, \dots, p$ are non-zero polynomials in n variables and for all $\alpha \in \mathcal{U}$, $i = 1, \dots, n$, $P_{i,\alpha}, Q_{i,\alpha}$ are polynomials in n variables, $Q_{i,\alpha}$ is non-zero.

We will identify the rational system \mathcal{R} with the tuple $(\{P_{i,\alpha}, Q_{i,\alpha}\}_{i=1, \dots, n, \alpha \in \mathcal{U}}, \{h_{k,1}, h_{k,2}\}_{k=1, \dots, p}, \mathcal{U}, v_0)$. We will say that \mathcal{R} is *polynomial*, if $h_{k,2} = 1$, $Q_{i,\alpha} = 1$ for all $k = 1, \dots, p$, $i = 1, \dots, n$, $\alpha \in \mathcal{U}$.

Note that the existence and uniqueness of a state trajectory of a rational system requires some care, as in our definition we did not exclude the possibility that $x(t)$ passes through the zero set of a denominator $Q_{i,\alpha}$.

Definition 5: A triplet (v, u, y) is a solution of a rational system \mathcal{R} of the form (2), if the input $u : [0; +\infty[\rightarrow \mathcal{U}$ is piecewise constant, the state $v : [0; +\infty[\rightarrow \mathbb{R}^n$ is absolutely continuous, the output $y : [0; +\infty[\rightarrow \mathbb{R}^p$ is piecewise continuous, and they satisfy

$$\begin{aligned} \dot{v}_i(t) Q_{i,u(t)}(v(t)) &= P_{i,u(t)}(v(t)) \\ y_k(t) h_{k,2}(v(t)) &= h_{k,1}(v(t)), \end{aligned} \quad (3)$$

for $1 \leq i \leq n$, and $1 \leq k \leq p$, where $v(t) = (v_1(t), \dots, v_n(t))^T$ is the state and $y(t) = (y_1(t), \dots, y_p(t))^T$.

By uniqueness of a solution of an analytic differential equation, for any initial state $v_0 \in \mathbb{R}^n$ such that $Q_{i,u(t)}(v_0) \neq 0$ and $h_{k,1}(v_0) \neq 0$, there exist at most one solution (v, u, y) of \mathcal{R} such that $v(0) = v_0$. In particular, if \mathcal{R} is a polynomial system, then for any initial state $v_0 \in \mathbb{R}^n$ there exist at most one solution (v, u, y) of \mathcal{R} such that $v(0) = v_0$.

Definition 6: Let $p : \mathcal{U}_{pc} \rightarrow PC([0; +\infty[, \mathbb{R}^p)$ be an input-output map, and let \mathcal{R} be a rational system of the form (2). We say that \mathcal{R} realizes p , if for any $u \in \mathcal{U}_{pc}$ there exists a solution (v, u, y) of \mathcal{R} such that $v(0) = v_0$ and $p(u) = y$.

III. REPRESENTING RNNs BY RATIONAL SYSTEMS.

In this section we show that an RNN realization of a given input-output map imply the existence of a rational system which is a realization of the same map. Moreover, we present the construction of such a rational realization. We introduce the following assumption.

Assumption 1 (A1): The function $\sigma : \mathbb{R} \rightarrow \mathbb{R}$ is analytic, and there exist an integer $N > 0$ and N analytic functions $\xi_1, \dots, \xi_N : \mathbb{R} \rightarrow \mathbb{R}$ such that

$$\begin{cases} \sigma V_0(\xi_1, \dots, \xi_N) = U_0(\xi_1, \dots, \xi_N) \\ \dot{\xi}_i V_i(\xi_1, \dots, \xi_N) = U_i(\xi_1, \dots, \xi_N), \text{ if } 1 \leq i \leq N \end{cases} \quad (4)$$

where U_k, V_k are polynomials in N variables.

Assumption (A1) involves existence of analytic functions $\{\xi_i\}_{i=1}^N$ and hence it is not easy to check it. In fact, Assumption (A1) can be replaced by the following hypothesis, which involves only derivatives of the activation function σ .

Assumption 2 (A2): The function $\sigma : \mathbb{R} \rightarrow \mathbb{R}$ is analytic, and there exist an integer $N > 0$ and a no-zero polynomial Q in $N + 1$ variables, such that

$$Q(\sigma, \sigma^{(1)}, \dots, \sigma^{(N)}) = 0 \quad (5)$$

where $\sigma^{(i)}$ denotes the i -th derivative of σ .

Lemma 1 (Equivalence of (A1) and (A2)): A function σ satisfies Assumption (A1) if and only if it satisfies Assumption (A2).

The proof of Lemma 1 is presented in [8].

Example 1: Let consider an analytic RNN of a response map for which the activation function σ is the hyperbolic tangent th , or the sigmoid function S given below:

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

These functions are *analytic* and satisfy a differential polynomial equation, more precisely the hyperbolic tangent verifies $y^{(1)} = 1 - y^2$ with $y(0) = 0$, and the sigmoid function satisfies $y^{(1)} = y(1 - y)$ with $y(0) = \frac{1}{2}$. It follows that Assumption (A2) holds.

Then we restrict the set of input maps, by supposing the following assumption.

Assumption 3 (Finite input set): In the rest of the paper we assume that $\mathcal{U} \subset \mathbb{R}^m$ is a **finite** set.

Notation 2: We denote by $|\mathcal{U}|$ the cardinality of \mathcal{U} . We set $\mathcal{U} = \{\alpha_1, \dots, \alpha_K\}$, where $\alpha_i \in \mathbb{R}^m$, and $\alpha_i \neq \alpha_j$ if $i \neq j$. In that case, we have $|\mathcal{U}| = K$.

The assumption that \mathcal{U} is finite is not an overly restrictive one, and it is satisfied in many potential applications.

Next we present the main theorem of this paper.

Theorem 1 (Embedding RNNs into rational systems):

Let $\sigma : \mathbb{R} \rightarrow \mathbb{R}$ be a globally Lipschitz function which satisfies (A1) and assume that \mathcal{U} is finite. Consider an input-output map $p : \mathcal{U}_{pc} \rightarrow PC([0; +\infty[, \mathbb{R}^p)$. If Σ is a RNN with activation function σ and input space \mathcal{U} , and Σ is a realization of p , then there exists a rational system which is a realization of p .

The proof of Theorem 1 is presented in [8]. The proof of the theorem relies on defining a rational system associated with the RNN. In order to avoid excessive notation, in the sequel we identify the sum of fraction of multi-variable polynomials $\sum_{k=1}^N \frac{P_k}{Q_k}$ with the fraction obtained by bringing all summands to the same denominator, i.e., $\frac{\sum_{k=1}^N P_k \prod_{r=1, r \neq k}^N Q_r}{\prod_{k=1}^N Q_k}$.

Definition 7: Let Σ be an RNN, whose activation function satisfies (A1), and assume that $\xi = (\xi_1, \dots, \xi_N)^T$ satisfies (4). Define the *rational system* $\mathcal{R}(\Sigma)$ associated with the RNN $\Sigma = (A, B, C, \mathcal{U}, \sigma, x_0)$, $A = (a_{i,j})_{i,j=1}^n$, $C = (c_{k,i})_{k=1, \dots, p, i=1, \dots, n}$ as follows:

$$\forall i = 1, \dots, N, \quad j = 1, \dots, n, \quad \alpha \in \mathcal{U} :$$

$$\dot{v}_{i,j,\alpha}(t) = \frac{U_i(v_{j,\alpha}(t))}{V_i(v_{j,\alpha}(t))} \left\{ \sum_{l=1}^n a_{j,l} \frac{U_0(v_{l,\beta}(t))}{V_0(v_{l,\beta}(t))} \right\} \quad \text{if } u(t) = \beta$$

$$v_{j,\alpha}(t) = (v_{1,j,\alpha}(t), \dots, v_{N,j,\alpha}(t)),$$

$$v_{j,\alpha}(0) = \xi(e_j^T (Ax_0 + B\alpha))$$

$$\dot{x}_j(t) = \frac{U_0(v_{j,\beta}(t))}{V_0(v_{j,\beta}(t))} \quad \text{if } u(t) = \beta, \quad \text{and } x_j(0) = e_j^T x_0,$$

$$y_k(t) = \sum_{i=1}^n c_{k,i} x_i(t), \quad k = 1, \dots, p$$

Remark 2: If Assumption (A1) is satisfied with polynomial equations, i.e. $V_i = 1$, $i = 0, \dots, N$, like in examples from Example 1, then $\mathcal{R}(\Sigma)$ is a polynomial system.

The proof of Theorem 1 relies on the following simple result, which is interesting on its own right. Let $\xi = (\xi_1, \dots, \xi_n)$ is from (4), and define $F : \mathbb{R}^n \rightarrow \mathbb{R}^L$, as $F(x) = (z_1, \dots, z_{nN|\mathcal{U}|}, x^T)^T$, where $z_{\phi(i,j,\alpha)} = \xi_i(e_j^T (Ax + B\alpha))$ and for any $i = 1, \dots, N, j = 1, \dots, n, \alpha \in \mathcal{U}$, $\phi(i, j, \alpha) = N \cdot |\mathcal{U}| \cdot (j-1) + N \cdot (r-1) + i$ if $\alpha = \alpha_r$, $r = 1, \dots, K$.

Lemma 2: If (x, u, y) is a solution of the RNN $\Sigma = (A, B, C, \mathcal{U}, \sigma, x_0)$, then (v, u, y) , with $v(t) = F(x(t))$ for $t \geq 0$, is a solution of $\mathcal{R}(\Sigma)$.

The proof of Lemma 2 is presented in [8].

IV. EXISTENCE OF AN RNN REALIZATION

Theorem 1 allows us to formulate a necessary condition for realizability of an input-output map by an RNN, using conditions of [19, Theorem 5.16] for existence of a realization by a rational system. In order to present this condition, we need to introduce additional notation and terminology.

Definition 8: An input-output map $p : \mathcal{U}_{pc} \rightarrow PC([0; +\infty[, \mathbb{R}^p)$ is *causal* if, for all $t \geq 0$ and for all

$u, v \in \mathcal{U}_{pc}$ such that $\forall s \in [0, t] : u(s) = v(s)$, it holds that $\forall s \in [0, t] : p(u)(s) = p(v)(s)$.

In other words, p is causal, if $p(u)(t)$ depends only on the values of u on the interval $[0, t]$. If p is the input-output map of a control system, then causality must necessarily hold.

Definition 9: An input-output map $p : \mathcal{U}_{pc} \rightarrow PC([0; +\infty[, \mathbb{R}^p)$ is *analytic* if, for all $k \in \{1, \dots, p\}$ and for all $\alpha_1, \dots, \alpha_l \in \mathcal{U}$, $l > 0$, the function $\phi_{p,k,\alpha_1,\dots,\alpha_l} : ([0, +\infty[)^l \rightarrow \mathbb{R}$ is analytic, where

$$\begin{aligned} \phi_{p,k,\alpha_1,\dots,\alpha_l}(t_1, \dots, t_l) &= p_k(u_{t_1,\dots,t_l}^{\alpha_1,\dots,\alpha_l})(T_l) \\ u_{t_1,\dots,t_l}^{\alpha_1,\dots,\alpha_l}(t) &= \begin{cases} \alpha_i & \text{if } t \in [T_{i-1}, T_i[, \quad i = 1, \dots, l \\ \alpha_l & \text{if } t \geq T_l \end{cases} \\ T_0 &= 0, \quad T_i = \sum_{j=1}^i t_j, \quad i = 1, 2, \dots, l. \end{aligned} \quad (6)$$

where p_k is the k -th component of p , i.e. $p(u) = (p_1(u), \dots, p_p(u))^T$ for all $u \in \mathcal{U}_{pc}$.

Notation 3: We denote by $\mathcal{A}(\mathcal{U}_{pc})$ the set of causal analytic input-output maps of the form $p : \mathcal{U}_{pc} \rightarrow PC([0; +\infty[, \mathbb{R})$.

Note that the set $\mathcal{A}(\mathcal{U}_{pc})$ forms an algebra over the field of real numbers with the usual point-wise addition, multiplication and multiplication by scalar. It can be shown that the algebra $\mathcal{A}(\mathcal{U}_{pc})$ is isomorphic to the ring of functions $\mathcal{A}(\widetilde{\mathcal{U}_{pc}} \rightarrow \mathbb{R})$ defined in [19, Definition 4.3], therefore it is an integral domain.

Definition 10: Let $\varphi \in \mathcal{A}(\mathcal{U}_{pc})$ and define the *derivative* $D_\alpha \varphi$ of φ along $\alpha \in \mathcal{U}$ as the function $D_\alpha \varphi : \mathcal{U}_{pc} \rightarrow PC([0; +\infty[, \mathbb{R})$, such that for all $u \in \mathcal{U}_{pc}$, for all $t \geq 0$,

$$\begin{aligned} (D_\alpha \varphi(u))(t) &= \frac{d}{ds} \left(\varphi(u_\alpha)(t+s) \right) \Big|_{s=0}, \\ u_\alpha(\tau) &= \begin{cases} u(\tau) & \tau \in [0, t[\\ \alpha & \tau > t \end{cases} \end{aligned}$$

It is easy to see that $D_\alpha \varphi$ is also causal and analytic, and hence $D_\alpha \varphi$ belongs to $\mathcal{A}(\mathcal{U}_{pc})$. Now we define the observation algebra of an input-output map.

Definition 11: Let $p : \mathcal{U}_{pc} \rightarrow PC([0; +\infty[, \mathbb{R}^p)$ be an analytic and causal input-output map. The *observation algebra* of p , denoted by $\mathcal{A}_{obs}(p)$, is the smallest sub-algebra of the algebra $\mathcal{A}(\mathcal{U}_{pc})$ such that the following holds.

- Consider the components $p_k : \mathcal{U}_{pc} \rightarrow PC([0; +\infty[, \mathbb{R})$, $k = 1, \dots, p$ of p , i.e., $\forall u \in \mathcal{U}_{pc} : p(u) = (p_1(u), \dots, p_p(u))^T$. For every $k = 1, \dots, p$, $p_k \in \mathcal{A}_{obs}(p)$.
- For every $g \in \mathcal{A}_{obs}(p)$, $D_\alpha g \in \mathcal{A}_{obs}(p)$, $\alpha \in \mathcal{U}$, i.e., $\mathcal{A}_{obs}(p)$ is closed under taking derivatives D_α , $\alpha \in \mathcal{U}$.

We call the *observation field*, denoted by $\mathcal{Q}_{obs}(p)$, the field of fractions of $\mathcal{A}_{obs}(p)$.

If an input-output map $p : \mathcal{U}_{pc} \rightarrow PC([0; +\infty[, \mathbb{R}^p)$ is realized by an RNN $\Sigma = (A, B, C, \mathcal{U}, \sigma, x_0)$, with σ satisfying Assumption (A1), then the rational system $\mathcal{R}(\Sigma)$, given in Definition 7, also realizes the input-output map p by Theorem 1 and Lemma 1. Thus the transcendence degree of the observation algebra $\mathcal{A}_{obs}(p)$ is finite by [19, Theorem 5.16]. We are now in the position to state a

necessary condition for existence of an RNN realization, which summarizes the arguments above.

Theorem 2 (Existence of an RNN: necessary condition): The input-output map $p : \mathcal{U}_{pc} \rightarrow PC([0; +\infty[, \mathbb{R}^p)$ has a realization by an RNN whose activation function satisfies Assumption (A1) only if p is causal, analytic and $\text{trdeg } \mathcal{A}_{obs}(p) < +\infty$.

V. MINIMALITY, REACHABILITY AND OBSERVABILITY

In this section, we first provide sufficient conditions for minimality of RNNs. We also provide a Hankel-rank like condition for minimality of the RNNs. Finally we relate reachability and observability properties for RNNs with the corresponding properties of rational systems.

A. Sufficient conditions for minimality of RNNs

Define the *dimension* of a RNN Σ of the form (1) (respectively rational system \mathcal{R} of the form (2)) as the number n of its state variables, and we denote the dimension of Σ (respectively \mathcal{R}) by $\dim(\Sigma)$ (respectively $\dim(\mathcal{R})$). Note that our definition of dimension of a rational system coincides with that of [20, Definition 13]. We say that RNN Σ with activation function σ is a σ -*minimal* realization (respectively the rational system \mathcal{R} is a *minimal realization*) of an input-output map p , if Σ (respectively \mathcal{R}) is a realization of p and there exists no RNN Σ' with activation function σ (respectively no rational system \mathcal{R}'), such that Σ' (respectively \mathcal{R}') is a realization of p and $\dim(\Sigma') < \dim(\Sigma)$ (respectively $\dim(\mathcal{R}') < \dim(\mathcal{R})$).

Lemma 3: Let $\Sigma = (A, B, C, \sigma, x_0)$ be an RNN, whose activation function σ satisfies (A1) and assume that Σ is a realization of the input-output map p . If the rational system $\mathcal{R}(\Sigma)$, given in Definition 7, is a minimal realization of p , then Σ is a minimal RNN realization of p .

The proof of Lemma 3 is presented in [8].

Unfortunately, in most of the cases, $\mathcal{R}(\Sigma)$ will not be minimal. Intuitively, this has to do with the fact that the states $x_1(t), \dots, x_n(t)$ of $\mathcal{R}(\Sigma)$ are integrals of the other states, leading to lack of observability if $n > 1$ and $x(0)$ is chosen so that $Cx(0) = 0$. In order to remedy this problem, we define *auxiliary rational systems* associated with RNNs.

Definition 12: Let $\Sigma = (A, B, C, \mathcal{U}, \sigma, x_0)$ be an RNN, whose activation function satisfies (A1), assume that $\xi = (\xi_1, \dots, \xi_N)$ satisfies (4). Define the *auxiliary rational system* $\mathcal{R}_{aux}(\Sigma)$ associated with the RNN Σ , $A = (a_{i,j})_{i,j=1}^n$, $C = (c_{k,i})_{k=1,\dots,p, i=1,\dots,n}$ as follows:

$$\forall i = 1, \dots, N, \quad j = 1, \dots, n, \quad \alpha \in \mathcal{U} :$$

$$\dot{v}_{i,j,\alpha}(t) = \frac{U_i(v_{j,\alpha}(t))}{V_i(v_{j,\alpha}(t))} \left\{ \sum_{l=1}^n a_{j,l} \frac{U_0(v_{l,\beta}(t))}{V_0(v_{l,\beta}(t))} \right\}, \quad \text{if } u(t) = \beta$$

$$v_{j,\alpha}(t) = (v_{1,j,\alpha}(t), \dots, v_{n,j,\alpha}(t))^T,$$

$$v_{j,\alpha}(0) = \xi(e_j^T (Ax_0 + B\alpha)),$$

$$y_{k,\alpha}(t) = \sum_{i=1}^n c_{k,i} \frac{U_0(v_{i,\alpha}(t))}{V_0(v_{i,\alpha}(t))}, \quad k = 1, \dots, p$$

Remark 3 (Polynomial $\mathcal{R}_{aux}(\Sigma)$): If Assumption (A1) is satisfied with polynomial equations, i.e. $V_i = 1$, $i = 0, \dots, N$, like in examples from Example 1, then $\mathcal{R}_{aux}(\Sigma)$ is a polynomial system.

Note that the rational system $\mathcal{R}_{aux}(\Sigma)$ does not realize the input-output map p_{Σ, x_0} of the RNN Σ , but it realizes an input-output map which is constructed from derivatives of p . More precisely, let $p : \mathcal{U}_{pc} \rightarrow PC([0; +\infty[, \mathbb{R}^p])$ be an input-output map realized by the RNN Σ . Define the input-output map $\hat{p} : \mathcal{U}_{pc} \rightarrow PC([0; +\infty[, \mathbb{R}^{p|U|})$ as follows:

$$\forall u \in \mathcal{U}_{pc}, \quad \hat{p}(u) = (D_{\alpha_1} p(u), \dots, D_{\alpha_K} p(u))^T, \quad (7)$$

where, $D_{\alpha} p(u) = (D_{\alpha} p_1(u), \dots, D_{\alpha} p_p(u))^T$, $\alpha \in \mathcal{U}$, $D_{\alpha} p_k(u)$ is defined in Definition 10 and $p(u) = (p_1(u), \dots, p_p(u))^T$.

Lemma 4: Let $\Sigma = (A, B, C, \mathcal{U}, \sigma, x_0)$ be an RNN and let $p : \mathcal{U}_{pc} \rightarrow PC([0; +\infty[, \mathbb{R}^p])$ be an input-output map. If the RNN Σ realizes p , then the rational system $\mathcal{R}_{aux}(\Sigma)$ realizes the input-output map \hat{p} defined in (7). The proof of Lemma 4 is presented in [8].

Lemma 5: Let $\Sigma = (A, B, C, \mathcal{U}, \sigma, x_0)$ be an RNN and let p be an input-output map realized by Σ . If $\mathcal{R}_{aux}(\Sigma)$ is a minimal realization of \hat{p} , then the RNN Σ is σ -minimal realization of p .

The proof of Lemma 5 is presented in [8].

It is known from [16], [17] that *algebraic, rational and semi-algebraic observability* and *algebraic reachability* characterize minimality of rational systems and these properties can be checked using methods of computational algebra [18]. In particular, we can derive sufficient conditions for the minimality of the RNN Σ using these reachability and observability concepts for rational systems. To this end, below the notions of algebraic reachability and algebraic/semi-algebraic observability for rational systems. Let \mathcal{R} be a rational system of the form (2). Define the set $R_{\mathcal{R}}(v_0)$ of reachable states of \mathcal{R} as the set formed by $v(t)$ such that $t \geq 0$ and (v, u, y) is a solution of \mathcal{R} with $v(0) = v_0$. The system \mathcal{R} is said to be *algebraically reachable*, if there is no non-trivial polynomial which is zero on $R_{\mathcal{R}}(v_0)$. The system \mathcal{R} is called *accessible*, if $R_{\mathcal{R}}(v_0)$ contains an open subset of \mathbb{R}^n . It is clear that accessibility implies algebraic reachability.

Let \mathcal{R} be a rational system of the form (2). Recall from [16, Definition 3.19] or from [5, Definition 4] that *observation algebra* of \mathcal{R} , denoted by $\mathcal{A}_{obs}(\mathcal{R})$, is the smallest sub-algebra of the field of rational functions $R(X_1, \dots, X_n)$ which contains $\frac{h_{k,1}}{h_{k,2}}$, $k = 1, \dots, p$ and which is closed under taking the formal Lie derivatives with respect to the formal vector fields $f_{\alpha} = \sum_{i=1}^n \frac{P_{i,\alpha}}{Q_{i,\alpha}} \frac{\partial}{\partial X_i}$. If \mathcal{R} is polynomial, i.e. $Q_{i,\alpha} = 1$, $i = 1, \dots, n$, $h_{k,2} = 1$, $k = 1, \dots, p$, then $\mathcal{A}_{obs}(\mathcal{R})$ is the sub-algebra of the ring of polynomials $R[X_1, \dots, X_n]$. Following [16] we say that the rational system \mathcal{R} is *algebraically observable*, if $\mathcal{A}_{obs}(\mathcal{R}) = R[X_1, \dots, X_n]$. Following [17] that \mathcal{R} is *semi-algebraically observable* if $\text{trdeg}(\mathcal{A}_{obs}(\mathcal{R})) = n$. We will say that \mathcal{R} is *observable*, if for every two distinct initial states v_0'', v_0' there exists solutions (v'', u, y'') and (v', u, y') of \mathcal{R} such that

$v''(0) = v_0'', v'(0) = v_0'$, and $y'' \neq y'$. It is easy to see that algebraic observability implies semi-algebraic observability. Moreover, for polynomial systems algebraic observability implies observability [4].

Lemma 6 (Sufficient conditions): If one of the conditions below holds, then Σ is σ -minimal realization of p :

- $\mathcal{R}_{aux}(\Sigma)$ is semi-algebraically observable and algebraically reachable.
- $\mathcal{R}_{aux}(\Sigma)$ is polynomial, it is algebraically observable and algebraically reachable.
- $\mathcal{R}_{aux}(\Sigma)$ is polynomial, it is algebraically observable and accessible.

The above lemma is then a direct consequence of Lemma 5 and [20, Proposition 6], its detailed proof is presented in [8].

B. A Hankel-rank like condition for minimality of RNNs

Recall that a linear system is a minimal realization of its input-output map, if and only if the dimension of this system equals the rank of the Hankel-matrix of the input-output map. We would like to formulate a similar result, where the role of the rank of the Hankel-matrix is played by the observation algebra $\mathcal{A}_{obs}(p)$. To this end, recall from [20, Lemma 1, Theorem 4] that $\mathcal{R}(\Sigma)$ is minimal if and only if $\dim(\mathcal{R}(\Sigma)) = \text{trdeg} \mathcal{A}_{obs}(p)$. In a similar manner, if $\mathcal{A}_{obs}(\hat{p})$ is the observation algebra of the input-output map \hat{p} defined in (7), then $\mathcal{R}_{aux}(\Sigma)$ is minimal if and only if $\dim(\mathcal{R}_{aux}(\Sigma)) = \text{trdeg} \mathcal{A}_{obs}(\hat{p})$.

Lemma 7: Assume that σ satisfies (A1), and let the RNN Σ be a realization of the input-output map p . If one of the following conditions hold for $n = \dim(\Sigma)$: **(1)** $\text{trdeg} \mathcal{A}_{obs}(p) = n(1 + |U|N)$, or **(2)** $\text{trdeg} \mathcal{A}_{obs}(\hat{p}) = n|U|N$, then Σ is a σ -minimal realization of p .

The proof of Lemma 7 is presented in [8].

C. Some aspects of reachability and observability of RNNs

One may wonder how restrictive the conditions of Lemma 6 are, and how they relate to accessibility/reachability and observability of the RNN Σ studied in [3], [21], [1]. In order to present this relationship more precisely, we introduce the following terminology. Define the reachable set of an RNN $\Sigma = (A, B, C, \mathcal{U}, \sigma, x_0)$

$$R_{\Sigma}(x_0) = \{x(t) \mid t \geq 0, (x, u, y) \text{ is a solution of } \Sigma, x(0) = x_0\}.$$

We will say that Σ is *accessible*, if $R_{\Sigma}(x_0)$ contains an open subset of \mathbb{R}^n , we say that Σ is *algebraically reachable* if there is no non-trivial polynomial which is zero on $R_{\Sigma}(x_0)$. We say that Σ is *span-reachable*, if the linear span of the elements $R_{\Sigma}(x_0)$ is \mathbb{R}^n , i.e. Σ is reachable if there exist no linear function which is zero on $R_{\Sigma}(x_0)$. Clearly, if Σ is accessible, then it is algebraically reachable, and if Σ is algebraically reachable, then it is span-reachable. We say that the RNN Σ is *weakly observable* if for every initial state $\hat{x} \in \mathbb{R}^n$ there is an open subset V of \mathbb{R}^n such that $\hat{x} \in V$ and for every $\hat{x} \neq \bar{x} \in V$, there exist solution (x, u, y) and (x', u, y') of Σ , with $x(0) = \hat{x}$ and $x'(0) = \bar{x}$, such that

$y \neq y'$. Then the RNN Σ is *observable* if for every initial state $\hat{x} \in \mathbb{R}^n$, $V = \mathbb{R}^n$ in the latter definition.

Lemma 8: Let $\Sigma = (A, B, C, \mathcal{U}, \sigma, x_0)$ be an RNN.

- If $\mathcal{R}_{aux}(\Sigma)$ is algebraically reachable, then Σ is span-reachable. In particular, if $\mathcal{R}_{aux}(\Sigma)$ is accessible, then Σ is span-reachable.
- If $\mathcal{R}_{aux}(\Sigma)$ is polynomial, and it is observable, and if the function σ is invertible and $\text{Ker}(A)$ is trivial, then Σ is observable. In particular, if \mathcal{R}_{aux} is algebraically observable, then Σ is observable.
- If $\mathcal{R}_{aux}(\Sigma)$ is polynomial, and it is semi-algebraically observable, and if the function σ is invertible and $\text{Ker}(A)$ is trivial, then Σ is weakly observable.

The proof of Lemma 8 is presented in [8]. Note that invertability of σ holds for many commonly used activation functions, see Example 1. Observe that accessibility, and algebraic / semi-algebraically observability conditions for rational / polynomial systems can be checked by using methods of computer algebra [18].

From [11, Theorem 1.12] it follows that if the RNN Σ is accessible and weakly observable, then it is a minimal realization of its input-output map p . From the comparison between the conditions of Lemma 6 with those of Lemma 8 it is clear that minimality of $\mathcal{R}_{aux}(\Sigma)$ is a much weaker condition than accessibility and weak observability of Σ . This suggests that using realization theory of rational systems is likely to yield more useful results for RNNs than using realization theory of general analytic systems.

Recall from [3, Theorem 1] that a necessary condition for observability of $\Sigma = (A, B, C, \mathcal{U}, \sigma, x_0)$ is that the largest A -invariant coordinate subspace of Σ included in $\text{Ker}(C)$ is trivial. More precisely, following [3] we say that a vector subspace V of \mathbb{R}^n is a *coordinate subspace* if it is spanned by some vectors from the canonical basis of \mathbb{R}^n , i.e. there exists an integer $s > 0$ and integers $i_1, \dots, i_s \in \{1, \dots, n\}$ such that V is spanned by e_{i_1}, \dots, e_{i_s} , where (e_1, \dots, e_n) denotes the canonical basis of \mathbb{R}^n . Let $\mathcal{O}_c(A, C)$ be the largest coordinate subspace which is A -invariant and contained in $\text{Ker}(C)$.

Lemma 9: If $\mathcal{R}_{aux}(\Sigma)$ is polynomial and it is semi-algebraically observable, then there exists no non-trivial coordinate subspace which is A -invariant and contained in $\text{Ker}(C)$, i.e. $\mathcal{O}_c(A, C) = \{0\}$.

The proof of Lemma 9 is presented in [8].

By [3, Theorem 1] $\mathcal{O}_c(A, C) = \{0\}$, it is also sufficient if $\ker(C) \cap \ker(A) = \{0\}$, the activation function σ satisfies only the IPP property, given in [3] for example, and if B verifies a condition on its rows. But here we do not need the latter hypothesis on Σ .

VI. CONCLUSIONS

We have shown that input-output maps of a large class of recurrent neural networks can be represented by rational/polynomial systems, and we used this fact to derive necessary and sufficient conditions for existence of a realization by a recurrent neural network and its minimality. Future research will be directed towards deriving a more

complete realization theory of recurrent neural network and for using the results of realization theory for analyzing machine learning algorithms.

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 & Control Letters* 22(1994): 235-244.
- [4] Z. Bartoszewicz, Minimal polynomial realizations, *Mathematics of control, signals, and systems*, 1988, pages 227-237, volume 1
- [5] Z. Bartoszewicz, Rational systems and observation fields, *Systems & Control Letters* 9, 1987, pages 379-386
- [6] D. Cox, J. Little, and D. OShea, *Ideals, varieties, and algorithms: An introduction to computational algebraic geometry and commutative algebra*, Springer, third edition, 2007.
- [7] P. DAlessandro, A. Isidori, and A. Ruberti. Realization and structure theory of bilinear dynamical systems. *SIAM J. Control and Optimization*, 517535, 1974.
- [8] T. Defourneau, M. Petreczky, and L. Hetel, Realization theory of recurrent neural networks and rational systems, arXiv preprint, arXiv:1903.05609
- [9] R. Hermann, and A. J. Krener, Nonlinear controllability and observability, *IEEE Transactions on automatic control*, 1977, pages 728-740, volume 22
- [10] A. Isidori, *Nonlinear control systems*, Springer, third edition, 2013
- [11] 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 331. D. Reidel Publishing Company, Dordrecht, 1986.
- [12] T. Kailath, *Linear Systems*, Prentice-Hall, New Jersey, 1979, ISBN 978-0-13-536961-6
- [13] R.E. Kalman. Mathematical description of linear dynamical systems. *SIAM J. Control and Optimization*, 1(2): 152159, 1963
- [14] E. Kunz. Introduction to commutative algebra and algebraic geometry. Birkhauser, Boston, 1985.
- [15] L. Ljung, *System identification: theory for the user (second edition)*, Prentice-Hall, Englewood Cliffs, NJ, 1999
- [16] 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
- [17] J. Němcová, M. Petreczky and J.H. van Schuppen, Realization theory of Nash systems, *SIAM J. Control & Optimization*, 2013, pages 3386-3414, volume 51
- [18] 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
- [19] 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
- [20] J. Němcová, and J.H. van Schuppen, Realization theory for rational systems: Minimal rational realizations, *Acta Applicandae Mathematica*, 2010, pages 605-626, volume 110
- [21] Y. Qiao, E.D. Sontag, Further results on controllability of recurrent neural networks, *Systems & Control Letters* 36, 1999, pages 121-129
- [22] Y. Wang and E.D. Sontag, *Algebraic differential equations and rational control systems*, *SIAM J. Control Optim.*, 30(5):11261149, 1992.
- [23] O. Zariski and P. Samuel, *Commutative algebra I, II.*, Springer, 1958.
- [24] A. Lindquist and G. Picci, *Linear Stochastic Systems*, ser. Series in Contemporary Mathematics. Springer-Verlag Berlin Heidelberg, 2015, vol. 1.
- [25] I. Sustkever, Training Recurrent Neural Networks, PhD thesis, University of Toronto, 2013.
- [26] A. Caterini and D. E. Chang, Deep Neural Networks in a Mathematical Framework, Springer, *SpringerBriefs in Computer Science*, 2018.