Observability of discrete-time recurrent neural networks coupled with linear systems

Francesca Albertini
Dipartimento di Matematica
Pura e Applicata
Università di Padova
Via Belzoni 7, 35131 Padova
Italy
albertini@pdmat1.unipd.it

Paolo Dai Pra
Dipartimento di Matematica
Pura e Applicata
Università di Padova
Via Belzoni 7, 35131 Padova
Italy
daipra@pdmat1.unipd.it

Keywords: observability, recurrent neural networks, linear systems.

Abstract

We give an algebraic characterization of observability for a class of models obtained by coupling recurrent neural networks with linear systems.

1 Introduction

In a recent series of papers [9, 2, 3, 4, 5, 1] some system theoretic properties of recurrent neural networks have been analyzed (see [6] for an introduction on neural computation). Recurrent neural networks, that are used in many applications such as speech processing and learning, have the mathematically appealing feature of providing a class of 'semilinear' models for which one might expect that the theory is easier and closer to the one of linear systems than is the case of general nonlinear models. In particular, for such systems observability has been shown to be equivalent to a simple algebraic condition on the parameters of the network.

In this paper we address the problem of determining observability conditions for discrete-time systems obtained by "coupling" a recurrent neural network with a linear system. More precisely, we deal with systems of the form:

$$\begin{array}{rcl} x_1(t+1) & = & \vec{\sigma}(A_{11}x_1(t) + A_{12}x_2(t) + B_1u(t)) \\ x_2(t+1) & = & A_{21}x_1(t) + A_{22}x_2(t) + B_2u(t) \\ y(t) & = & Cx(t) = C_1x_1(t) + C_2x_2(t) \end{array} \qquad (1)$$

where $x_1 \in \mathbb{R}^k$, $x_2 \in \mathbb{R}^{n-k}$, $u \in \mathbb{R}^m$, $y \in \mathbb{R}^p$ and $A_{11}, A_{12}, A_{21}, A_{22}, B_1, B_2, C_1, C_2$ are matrices of appropriate dimensions. Moreover, for $v \in \mathbb{R}^k$,

$$\vec{\sigma}(v) = (\sigma(v_1), \ldots, \sigma(v_k))$$

where $\sigma : \mathbb{R} \to \mathbb{R}$ is some assigned nonlinear function.

These models include *interconnections* between linear systems and recurrent neural networks (for the definition of interconnection of two systems see [8]). As observed in [7] (Sec. 6.4, where the identification question was addressed), in many applications it is natural to assume that some components of the system evolve driven by linear equations, while the rest of the system is nonlinear, and modeled by recurrent neural networks. Unlike in [7], we admit interaction between the linear and the nonlinear components of the system.

Two states $x, z \in \mathbb{R}^n$ are said to be *indistinguishable* (we write $x \sim z$) if, for every input sequence $(u(t))_{t\geq 0}$, the solutions of (1) corresponding to initial conditions $x(0) = x$ and $x(0) = z$ give rise to the same output sequence $(y(t))_{t\geq 0}$. When the output sequence is the same up to time $t = d$ we say that x, z are indistinguishable in d-steps, and we write $x \sim^d z$. A system is said to be observable if $x \sim z$ implies $x = z$.

Our aim in this paper is to characterize observability of systems of type (1) through simple algebraic conditions on the matrices of the model. Similarly to [4], this will be done under nonlinearity assumptions on the function σ and nodegenericity conditions on the matrix B_1.

In section 2 we give the precise assumptions on σ and B_1, and state our main theorem (1), together with some clarifying examples. Section 3 is devoted to the proof of this result.

2 Statement of main result

The observability result we give in this paper will be proved under the following assumptions on the function σ and the matrix B_1.

Definition 2.1 A function $\sigma : \mathbb{R} \to \mathbb{R}$ is said to satisfy the *independence property* (**IP**) if for every $l > 0$, any nonzero real numbers $b_1, b_2, \ldots, b_l$ such that $b_i \neq \pm b_j$ for $i \neq j$, and any real numbers $\beta_1, \beta_2, \ldots, \beta_l$ the functions of $\xi \in \mathbb{R}$

$$1\,,\ \sigma(b_1\xi+\beta_1)\,,\ldots,\ \sigma(b_l\xi+\beta_l)$$

are linearly independent.

A criterion for showing that a function satisfies the **IP** can be found in [5]. In particular, the main examples in neural networks, i.e. $\sigma(x) = \tanh(x)$ and $\sigma(x) = \arctan(x)$, satisfy the **IP** . In the rest of the paper we denote by $\mathcal{S}$ the class of those systems of type (1) satisfying the following properties:
a) σ satisfies the **IP** ;
b) $B_1 \in \mathcal{B}_{k,m}$, where

$$\mathcal{B}_{k,m} = \left\{ M \in \mathbb{R}^{k\times m} : \begin{array}{ll} M^i \neq 0 & \forall\, i \\ M^i \neq \pm M^j & \forall\, i,j,\ i\neq j \end{array} \right\}$$

(M^i denotes the i-th row of M).

Let $e^{(i)}$, $i = 1,\ldots,k$, denote the canonical basis elements in $\mathbb{R}^k$. A subspace V of $\mathbb{R}^k$ will be called a *coordinate subspace* if $V = 0$ or V has the form

$$V = \text{ span } \{e^{(i_1)},\ldots,e^{(i_j)}\}.$$

From now on $V \subset \mathbb{R}^k$, $W \subset \mathbb{R}^{n-k}$ denote the *maximal* pair of subspaces satisfying the following properties:

P1. V is a coordinate subspace, $V \subset \ker C_1$, $A_{11}V \subset V$;

P2. $W \subset \ker C_2$, $A_{22}W \subset W$;

P3. $A_{21}V \subset W$;

P4. $A_{12}W \subset V$.

Notice that since if (V', W') and (V'', W'') satisfy P1-P4, then so do $(V' + V'', W' + W'')$, such *maximal* subspaces (under the order "$\subset$") exist.

In what follows we let $A_1 = [A_{11}, A_{12}] \in \mathbb{R}^{k\times n}$ and $A_2 = [A_{21}, A_{22}] \in \mathbb{R}^{(n-k)\times n}$. The main result of this paper is the following theorem.

Theorem 1 *Let $x, z \in \mathbb{R}^n$ denote two initial states for a system $\Sigma \in \mathcal{S}$. Then $x \sim z$ if and only if*

$$x - z \in \ker C\ ,\ A_1(x - z) \in V\ ,\ A_2(x - z) \in W. \quad (2)$$

In particular Σ is observable if and only if

$$\ker C \cap A_1^{-1}V \cap A_2^{-1}W = 0$$

("$\cdot^{-1}$" denotes inverse image).

The observability condition in Theorem 1 becomes effective as soon as one has an algorithm to compute V and W. Such an algorithm is given in next Section; it will be the most convenient for proving the Theorem, but not necessarily the fastest one.

In order to see Theorem 1 "at work", we consider some special cases.

Example 2.2 Assume $k = 0$. Thus $\mathcal{S}$ is the class of the (n, m, p)-linear systems. Here there is no V, and W is the largest A_{22}-invariant subspace of $\ker C_2$. The observability conditions given by Theorem 1 is

$$\ker C_2 \cap A_{22}^{-1}W = 0$$

that can be easily shown to be equivalent to

$$W = 0$$

which is the usual observability condition for linear systems.

Example 2.3 Assume $k = n$. Then $\mathcal{S}$ is the class of recurrent neural networks studied in [4]. In this case there is no W, and V is the largest A_{11}-invariant coordinate subspace contained in $\ker C_1$. The observability condition given by Theorem 1 is:

$$\ker C_1 \cap A_{11}^{-1}V = 0,$$

which was found in [4].

Example 2.4 Assume that C_1 has no zero columns. Then there is no nonzero coordinate subspace contained in $\ker C_1$; so $V = 0$. Therefore W is the largest A_{22}-invariant subspace contained in $\ker C_2$ and $\ker A_{12}$.

It follows that $x \sim z$ if and only if $x - z \in \ker C_1 \cap \ker A_1$, and A_2x, A_2z are indistinguishable for the linear systems (A_{22}, C_2), and (A_{22}, A_{12}).

3 Proof of Theorem 1

For a given matrix D, we denote by I_D the set

$$I_D = \{\, i \,|\ \text{the } \ i\text{-th column of } D \text{ is zero } \}.$$

In particular, for a coordinate subspace

$$V = \text{ span } \{\, e^{(i_1)},\ldots,e^{(i_j)}\,\},$$

$V \subseteq \ker D$ if and only if all $i_k \in I_D$. The following technical fact describes the way the **IP** of σ will be used in the rest of the proof.

Lemma 3.1 Assume that $D \in \mathbb{R}^{q\times k}$, $B_1 \in \mathcal{B}_{k,m}$, σ satisfies the **IP** . Then the following two properties are equivalent for each pair of vectors ξ, η:

1. $\xi_j = \eta_j$ for all $j \notin I_D$,
2. $D\sigma(\xi + B_1u) = D\sigma(\eta + B_1u)$ for all $u \in \mathbb{R}^m$.

The proof of this lemma can be found in [4].

We now construct an increasing sequence of indexes J_d, and a decreasing sequence of subspaces $V_d \subseteq \mathbb{R}^k$, $W_d \subseteq \mathbb{R}^{n-k}$, for $d \geq 1$, where V_d is a coordinate subspace. Let:

$$\begin{array}{rcl} J_1 &=& \{1,\ldots,k\} \setminus I_{C_1}; \\ V_1 &=& \text{ span } \{e^{(j)} \,|\, j \notin J_1\}; \\ W_1 &=& \ker C_2; \end{array}$$

and, for $d > 1$, let:

$$J_{d+1} \;=\; J_1 \cup \{ i \,|\, \exists\, j \in J_s \;\text{ s. t. }\; (A_{11})^{ji} \neq 0 \,\}$$
$$\cup \{ i \,|\, \exists\, j, \exists 0 \leq l \leq d-1 \;\text{ such that }\; (C_2 A_{22}^l A_{21})^{ji} \neq 0 \,\}$$
$$\cup \cup_{s=1}^{d-1} \{ i \,|\, \exists\, j \in J_s \;\text{ such that }\; (A_{12} A_{22}^{d-s-1} A_{21})^{ji} \neq 0 \,\}$$
$$V_{d+1} \;=\; \text{span}\, \{ e^{(j)} \,|\, j \notin J_{d+1} \,\}$$
$$W_{d+1} \;=\; \left\{ w \;\middle|\; \begin{array}{l} A_{22}^l w \in \ker C_2 \;\text{ for }\; 0 \leq l \leq d, \\ A_{12} A_{22}^{d-s} w \in V_s \;\text{ for }\; 1 \leq s \leq d \end{array} \right\}.$$

Since the sequence (V_d, W_d) is decreasing, than it becomes stationary after a finite number of steps. The proof of Theorem 1 is an immediate consequence of the following two facts. Recall that with (V, W) we denote the maximal pair satisfying the conditions P1-P4 given in the previous section.

Proposition 3.2 The following identities hold:

$$V = \cap_{d \geq 1} V_d, \qquad W = \cap_{d \geq 1} W_d.$$

Proposition 3.3 The following properties are equivalent:
i) $x \sim^d z$;
ii) $x - z \in \ker C$; $A_1 x - A_1 z \in V_d$; $A_2 x - A_2 z \in W_d$.

Proof of Proposition 3.2
Define:

$$J_\infty = \cap_{d \geq 1} J_d; \quad V_\infty = \cap_{d \geq 1} V_d; \quad W_\infty = \cap_{d \geq 1} W_d.$$

Thus $V_\infty =$ span $\{ e^{(j)} \,|\, j \notin J_\infty \}$. By using the recursive definition of J_d, V_d, and W_d and the fact that there exists $\bar{d}$ such that $V_d = V_\infty$, $W_d = W_\infty$ for $d \geq \bar{d}$, the following properties are easily seen:

$$\{ \, i \,|\, \exists\, j \in J_\infty \;\text{ so that }\; (A_{11})^{ji} \neq 0 \,\} \subseteq J_\infty. \tag{3}$$

$$\{ i \,|\, \exists\, j, \exists l \geq 0 \;\text{ so that }\; (C_2 A_{22}^l A_{21})^{ji} \neq 0 \,\} \subseteq J_\infty. \tag{4}$$

$$\{ i \,|\, \exists\, j \in J_\infty, \exists l \geq 0 \;\text{ so that }\; (A_{12} A_{22}^l A_{21})^{ji} \neq 0 \,\} \subseteq J_\infty. \tag{5}$$

$$W \subseteq \left\{ w \;\middle|\; \begin{array}{l} A_{22}^l w \in \ker C_2 \;\text{ for }\; l \geq 0, \\ A_{12} A_{22}^l w \in V_\infty \;\text{ for }\; l \geq 0 \end{array} \right\}. \tag{6}$$

We now prove $V_\infty \subseteq V$ and $W_\infty \subseteq W$ by showing that V_∞ and W_∞ satisfy properties P1-P4 in the definition on V, W.

P1. The only thing which is not obvious is the A_{11}-invariance of V_∞. So let $v \in V_\infty$. By (3), we have $v_i = 0$ if $\exists j \in J_\infty$ such that $(A_{11})^{ji} \neq 0$. Thus, for $j \in J_\infty$:

$$(A_{11} v)_j = \sum_{i=1}^{k} A_{11}^{ji} v_i = 0$$

and, therefore, $A_{11} v \in V_\infty$.

P2. Let $w \in W_\infty$. We show that $A_{22} w \in W_d$ for every $d \geq 1$. The fact that $A_{22}^l (A_{22} w) \in \ker C_2$ for $0 \leq l \leq d$ follows clearly by (6). Moreover, for $1 \leq s \leq d$, it follows from (6) that:

$$A_{12} A_{22}^{d-s} (A_{22} w) \in V_\infty \subseteq V_s.$$

P3. Let $v \in V_\infty$. We have to show that $A_{21} v \in W_d$ for all $d \geq 1$. Using the same argument used to get P1, it is easy to see that from (4) we get:

$$C_2 A_{22}^l A_{21} v = 0 \quad \forall\, l \geq 0,$$

and from (5) we have:

$$A_{12} A_{22}^l A_{21} v \in V_\infty \subseteq V_s \quad \forall\, s \geq 1.$$

This implies that $A_{21} v \in W_d$ for all d, as desired.

P4. This is immediate by equation (6).

To complete the proof, we have to show that $V \subseteq V_d$, and $W \subseteq W_d$ for all $d \geq 1$. We do this by induction. For $d = 1$ this is clear.

Now let $e^{(i)} \in V$. We show that $e^{(i)} \in V_{d+1}$, i.e. $i \notin J_{d+1}$.

- Suppose i is such that there exists $j \in J_d$ with $(A_{11})^{ji} \neq 0$. Then we have

  $$(A_{11} e^{(i)})_j = (A_{11})^{ji} \neq 0 \quad \Rightarrow \quad A_{11} e^{(i)} \in V_d.$$

  This is impossible since $A_{11} V \subset V$ and, by inductive assumption, $V \subset V_d$.

- Suppose i is such that there exist j and $0 \leq l \leq d-1$ with $(C_2 A_{22}^l A_{21})^{ji} \neq 0$. As before, this implies

  $$C_2 A_{22}^l A_{21} e^{(i)} \neq 0$$

  that is impossible, since $A_{21} V \subset W$ and $W \subset \ker (C_2 A_{22}^l)$ for every $l \geq 0$.

- Suppose i is such that there exist $0 \leq s \leq d-1$ and $j \in J_s$ with $(A_{12} A_{22}^{d-s} A_{21})^{ji} \neq 0$. This implies

  $$A_{12} A_{22}^{d-s} A_{21} e^{(i)} \notin V_s.$$

  This is impossible since $A_{22}^{d-s} A_{21} V \subset V$, $A_{12} W \subset W$ and, by inductive assumption, $V \subset V_s$.

Thus we have shown that $e^{(i)} \in V_{d+1}$.

Now let $w \in W$. We show that $w \in W_{d+1}$.

- The condition $A_{22}^l w \in \ker C_2$ for $0 \leq l \leq d$ follows from $W \subset \ker C_2$ and $A_{22} W \subset W$.

- The condition $A_{12} A_{22}^{d-s} w \subset V_s$ for $1 \leq s \leq d$ follows from $A_{22} W \subset W$, $A_{12} W \subset V$ and the fact that, by inductive assumption, $V \subset V_s$.

Proof of Proposition 3.3.
First introduce the following notation; for $x \in \mathbb{R}^n$ we let

$$x^+(u) = \begin{pmatrix} C_1\sigma(A_1x + B_1u) \\ C_2(A_2x + B_2u) \end{pmatrix}.$$

We proceed by induction on d. For $d = 1$:

$$x \sim^1 z \;\Leftrightarrow\; \left\{ \begin{array}{rcl} Cx & = & Cz \\ Cx^+(u) & = & Cz^+(u) \;\; \forall u \end{array} \right.$$

$$\Leftrightarrow \left\{ \begin{array}{rcl} x - z & \in & \ker C \\ C_1\sigma(A_1x + B_1u) & = & C_1\sigma(A_1z + B_1u) \\ C_2(A_2x + B_2u) & = & C_2(A_2z + B_2u). \end{array} \right.$$

By Lemma 3.1 this is equivalent to

$$\left\{ \begin{array}{rcll} x - z & \in & \ker C & \\ (A_1x)_i & = & (A_1z)_i & \text{for} \;\; i \notin I_{C_1} \\ A_2x - A_2z & \in & \ker C_2 & \end{array} \right.$$

$$\Leftrightarrow \left\{ \begin{array}{rcl} x - z & \in & \ker C \\ A_1(x - z) & \in & V_1 \\ A_2(x - z) & \in & W_1. \end{array} \right.$$

For the inductive step, observe that

$$x \sim^{d+1} z \quad \Leftrightarrow \quad x^+(u) \sim^d z^+(u) \;\; \forall u \quad \text{and} \quad Cx = Cz$$

$$\Leftrightarrow \left\{ \begin{array}{rcl} x - z & \in & \ker C \\ Cx^+(u) & = & Cz^+(u) \;\; \forall u \\ A_1x^+(u) - A_1z^+(u) & \in & V_d \;\; \forall u \\ A_2x^+(u) - A_2z^+(u) & \in & W_d \;\; \forall u \end{array} \right.$$

$$\Leftrightarrow \left\{ \begin{array}{rcl} x - z & \in & \ker C \\ A_1(x - z) & \in & V_1 \\ A_2(x - z) & \in & W_1 \\ A_1x^+(u) - A_1z^+(u) & \in & V_d \;\; \forall u \\ A_2x^+(u) - A_2z^+(u) & \in & W_d \;\; \forall u. \end{array} \right.$$

Thus, it is enough to show that for all u

$$\begin{array}{c} A_1x^+(u) - A_1z^+(u) \in V_d \\ A_2x^+(u) - A_2z^+(u) \in W_d \end{array} \;\Leftrightarrow\; \begin{array}{c} A_1(x - z) \in V_{d+1} \\ A_2(x - z) \in W_{d+1} \end{array} \tag{7}$$

The l.h.s. of (7) means:

- $\forall j \in J_d$

  $$[A_{11}\sigma(A_1x + B_1u) + A_{12}(A_2x + B_2u)]_j =$$

  $$[A_{11}\sigma(A_1z + B_1u) + A_{12}(A_2z + B_2u)]_j.$$

- $\forall 0 \le l \le d - 1$

  $$C_2A_{22}^l[A_{21}\sigma(A_1x + B_1u) + A_{22}(A_2x + B_2u)] =$$

  $$= C_2A_{22}^l[A_{21}\sigma(A_1z + B_1u) + A_{22}(A_2z + B_2u)].$$

- $\forall j \in J_s$, $1 \le s \le d - 1$

  $$[A_{12}A_{22}^{d-s-1}[A_{21}\sigma(A_1x+B_1u)+A_{22}(A_2x+B_2u)]]_j =$$

  $$= [A_{12}A_{22}^{d-s-1}[A_{21}\sigma(A_1z+B_1u)+A_{22}(A_2z+B_2u)]]_j.$$

which is equivalent to, by lemma 3.1,

- $(A_1x)_i = (A_1z)_i$ if there exists $j \in J_d$ with $A_{11}^{ji} \neq 0$;
- $A_{12}(A_2x - A_2z) \in V_d$;
- $(A_1x)_i = (A_1z)_i$ if $\exists j$ and $\exists 0 \le l \le d$ with $(C_2A_{22}^lA_{21})^{ji} \neq 0$;
- $C_2A_{22}^{l+1}(A_2x - A_2z) = 0 \;\forall 0 \le l \le d - 1$;
- $(A_1x)_i = (A_1z)_i$ if $\exists 1 \le s \le d-1$, and $\exists j \in J_s$ with $(A_{12}A_{22}^{d-s-1}A_{21})^{ji} \neq 0$;
- $A_{12}A_{22}^{d-s}(A_2x - A_2z) \in V_s \;\; \forall 1 \le s \le d - 1$.

These conditions are easily shown to be equivalent to:

$$A_1(x - z) \in V_{d+1}, \qquad A_2(x - z) \in W_{d+1}.$$

References

[1] Albertini F., and Dai Pra P., "Forward accessibility for recurrent neural networks" submitted for pubblication.

[2] Albertini, F., and E.D. Sontag, "For neural networks, function determines form," *Neural Networks*, 6 (1993):975-990.

[3] Albertini, F., and E.D. Sontag, "Identifiability of discrete-time neural networks" *Proc. European Control Conference*, Groningen, July, 1993:460-465.

[4] Albertini, F., and E.D. Sontag, "State observability in recurrent neural networks" *System and Control Letters*, 22, No. 4, 1994.

[5] Albertini, F., E.D. Sontag, and V. Maillot, "Uniqueness of weights for neural networks," *Artificial Neural Networks with Applications in Speech and Vision*, R. Mammone ed., Chapman and Hall, London, 1993, pp. 115-125.

[6] Hertz, J., A. Krogh, and R.G. Palmer, *Introduction to the Theory of Neural Computation*, Addison-Wesley, Redwood City, 1991.

[7] Koplon, R.B., *Linear systems with constrained outputs and transitions*, Ph.D. Thesis, Rutgers University, 1994.

[8] Sontag, E.D., *Mathematical Control Theory: Deterministic Finite Dimensional Systems*, Springer, New York, 1990.

[9] Sussmann, H.J., "Uniqueness of the weights for minimal feedforward nets with a given input-output map," *Neural Networks* **5**(1992): 589-593.