Recurrent neural networks coupled with linear systems: observability in continuous and discrete time

Francesca Albertini
and
Paolo Dai Pra
Universita' di Padova
Dipartimento di Matematica
Via Belzoni 7, 35131 Padova
Italy

Abstract

We give necessary and sufficient conditions for observability of a class of recurrent neural networks having a subsystem where the activation function is the identity. An algorithm for computing all pairs of indistinguishable states is also given.

1 Introduction

Since the pioneering works of McCulloch and Pitts [10] and Hebb [8], neural networks have attracted the interest of researchers from many areas of sciences (for a review on the subject see [13] and [9]). In particular, recurrent neural networks are used as a computational tool in many applications, and their effectiveness and reliability have been widely recognized. Only recently, however, these models have been studied from a system-theoretic point of view. Since a recurrent neural network provides a state-space description of a nonlinear input-output map, questions about identifiability ([3, 4, 7, 6]), observability ([5]) and controllability ([1]) arise naturally. It turns out that the abovementioned properties can be expressed in terms of simple algebraic conditions on the "weights" of the network.

In this paper we continue the study, initiated in [2], of the observability of systems obtained by "coupling" a recurrent neural network with a linear system. As observed in [11] (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 [11], we admit interaction between the linear and the nonlinear components of the system. More precisely, denoting by "+" the time shift in discrete time and by "$\cdot$" the time derivative in continuous time, we deal with systems of the form

$$\begin{array}{rcl} x_1^+ \ \ (\text{ or } \ \dot{x}_1) & = & \vec{\sigma}(A_{11}x_1 + A_{12}x_2 + B_1 u) \\ x_2^+ \ \ (\text{ or } \ \dot{x}_2) & = & A_{21}x_1 + A_{22}x_2 + B_2 u \\ y & = & Cx \, = \, C_1 x_1 + C_2 x_2 \end{array} \tag{1}$$

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. For the continuous time case, to have uniqueness and local existence for the solution of (1), we require $\sigma(\cdot)$ to be locally Lipschitz, and we only admit *input signals* $(u(t))_{t\geq 0}$ that are measurable and locally essentially bounded.

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 $(y(t))_{t\geq 0}$. In discrete time, 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 $x = z$.

The problem of determining whether a system is observable, besides being interesting in itself, is particularly relevant in the context of neural networks since it is strictly related to identifiability and minimality of the network (see [6]). Such connection is well established for recurrent neural networks ($k = n$ in (1)), and the extension to "coupled" systems is presently under investigation.

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 [5], this will be done under nonlinearity assumptions on the function σ and nodegenericity conditions on the matrix B_1. The main result of this paper has already been proved in [2] for discrete time systems; the proof given there, which is *constructive* in the sense that provides an algorithm to check observability, does not seem to be extendible to continuous time. The proof given here is less constructive, but works in continuous time as well as in discrete time. Moreover, the observability condition for systems of type (1) is the same in discrete and continuous time, and so in both cases the algorithm introduced in [2] (see Section 2) provides all pairs of indistinguishable states (notice the analogy with linear systems).

The paper is organized as follows. In section 2 we give the precise assumptions on σ and B_1, and state our main theorem (Theorem 1), together with an algorithm to check observability (Proposition 2.4) and some clarifying examples. Section 3 is devoted to the proof of Theorem 1 and in Section 4 we give the proof of Proposition 2.4.

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, i.e. if $a_0 + \sum_{i=1}^{l} a_i\sigma(b_i\xi + \beta_i) = 0$ for all $\xi \in \mathbb{R}$, then $a_i = 0$ for $i = 0, \ldots, l$.

A criterion for showing that a function satisfies the **IP** can be found in [7]. 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

$$\mathcal{B}_{k,m} = \left\{ M \in \mathbb{R}^{k\times m} : \begin{array}{ll} M^i \neq 0 & \forall\, i = 1, \ldots, k \\ M^i \neq \pm M^j & \forall\, i, j = 1, \ldots, k\,,\ i \neq j \end{array} \right\}$$

(M^i denotes the i-th row of M). Conditions a) and b) are used in the proof of Theorem 1 below. Their rough meaning is: a) the system is fully nonlinear; b) the control is 'nontrivially' effecting all components of the network.

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

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

For a given matrix D, we let I_D be the set $I_D = \{\, i \mid \text{ the } \ i \text{ -th column of } D \text{ is zero } \}$. Notice that, for a coordinate subspace $V = \text{ span } \{\, e^{(i_1)}, \ldots, e^{(i_j)}\,\}$, $V \subseteq \ker D$ if and only if all $i_k \in I_D$.

Definition 2.2 We denote by $V \subset \mathbb{R}^k$, and $W \subset \mathbb{R}^{n-k}$ 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. The notation V, W will be kept for the rest of the paper.

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*

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

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 is in terms of the two subspaces V and W. The crucial point is that these subspaces can be determined by an algorithm which consists in solving a finite number of linear algebraic equations. We now give a description of this algorithm.

We inductively 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:

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

Remark 2.3 The sequence $(V_d, W_d)_{d\geq 1}$ is decreasing, and so it becomes stationary after a finite number of steps. It is possible to give a condition for the termination of this algorithm. Suppose that we obtain a stationary string $V_s = V_{s+1} = \ldots = V_{s+n-k}$ of lenght $n-k+1$. Then, by using the definition of V_d and W_d, and by applying the Hamilton-Cayley Theorem, one shows that $V_d = V_s$ for all $d \geq s$ and that $W_d = W_{s+n-k}$ for all $d \geq s+n-k$. It follows that the sequence $(V_d, W_d)_{d\geq 1}$ becomes stationary after at most $(n-k+1)k$ steps, for $k \geq 1$, or n steps, for $k = 0$.

It turns out that the sequence (V_d, W_d) stabilizes exactly at (V, W), as stated in the following Proposition.

Proposition 2.4 The following identities hold:

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

The proof of this Proposition can be found in [2]; however, since it presents a *constructive* way to compute the subspaces V and W it is also given in this work (Section 4). For discrete-time systems the meaning of the subspaces V_d, W_d is clarified by the following Proposition.

Proposition 2.5 Let Σ be a discrete time system in $\mathcal{S}$. Then the following properties are equivalent:
i) $x \sim^d z$;
ii) $x - z \in \ker C$; $A_1 x - A_1 z \in V_d$; $A_2 x - A_2 z \in W_d$.

For the proof of this result we refer to [2]. We conclude this Section by showing the application of Theorem 1 to some special cases.

Example 2.6 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 condition 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 2.7 Assume $k = n$. Then $\mathcal{S}$ is the class of recurrent neural networks studied in [5]. 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 [5].

Example 2.8 Assume $A_{12} = 0$ and $A_{21} = 0$. Thus the linear and the nonlinear dynamics are decoupled (as in [11]). By what observed in the two previous examples, we get that observability of the whole system is equivalent to the two systems, characterized by $(\sigma, A_{11}, B_1, C_1)$ and (A_{22}, B_2, C_2) respectively, being both observable. The separate observability is clearly necessary for observability of the combined system, but the sufficiency is not an obvious fact. For instance, if the two components were both linear, then such "separation property" would be false, in general.

Example 2.9 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 $A_2 x$, $A_2 z$ are indistinguishable for the linear systems (A_{22}, C_2), and (A_{22}, A_{12}).

3 Proof of Theorem 1

The condition $\Sigma \in \mathcal{S}$ enters in the proof of Theorem 1 through the following technical fact whose proof can be found in [5].

Lemma 3.1 Assume that $D \in \mathbb{R}^{q \times l}$, $B_1 \in \mathcal{B}_{l,m}$, σ satisfies the **IP** . Then the following two properties are equivalent for each ξ, $\eta, \in \mathbb{R}^l$, and each α, $\beta \in \mathbb{R}^q$:

1. $\xi_j = \eta_j$ for all $j \notin I_D$, $\alpha = \beta$,
2. $D\vec{\sigma}(\xi + B_1 u) + \alpha = D\vec{\sigma}(\eta + B_1 u) + \beta$ for all $u \in \mathbb{R}^m$.

We now introduce some useful notations. Given $x \in \mathbb{R}^n$ and $u \in \mathbb{R}^m$, we first let, for discrete-time models,$x^+(u)$ be the state reached from x using the control value u. For continuous-time models, if $v(t)$ is the control function constantly equal to u, we denote by $x_u(t)$ the corresponding trajectory ; notice that $x_u(t)$ is certainly defined on an interval of the form $[0, \epsilon_u)$. Sometimes we will deal with two trajectories of this type starting at two different initial states; in this case by $[0, \epsilon_u)$ we mean the interval in which both trajectories are defined.

For any two pairs of states $(x, z), (x', z') \in \mathbb{R}^n \times \mathbb{R}^n$, we write

$$(x, z) \rightsquigarrow (x', z')$$

if, for discrete-time, we can find an input sequence $u_1, \ldots, u_p$, for some $p \geq 0$, which steers the state x (resp., z) to x' (resp., z'). For continuous-time, we require that there exists some control function $u(t) : [0, T] \to \mathbb{R}^m$, such that, it is possible to solve the differential equation (1) starting at x (resp z), for the the entire interval $[0, T]$, and at time T the state x' (resp. z') is reached. With this terminology, two states $(x, z) \in \mathbb{R}^n \times \mathbb{R}^n$ are distinguishable if and only if there is some pair $(x', z') \in \mathbb{R}^n$ such that $(x, z) \rightsquigarrow (x', z')$ and $Cx' \neq Cz'$.

In what follows the proofs for the discrete-time case are quite similar to the proofs for continuous-time models and simpler, thus they are only skected.

Lemma 3.2 Let $H_1 \in \mathbb{R}^{q\times k}$, $H_2 \in \mathbb{R}^{q\times(n-k)}$, with $q \geq 1$, $1 \leq i \leq q$, and $x, z \in \mathbb{R}^n$.

(a)
- For continuous-time models, if $(H_1 A_1 x_u(t))_i = (H_1 A_1 z_u(t))_i$ for all $u \in \mathbb{R}^m$ and all $t \in [0, \epsilon_u)$ then we have:
$$\begin{array}{l} (H_1 A_{12} A_2 x)_i = (H_1 A_{12} A_2 z)_i, \\ (H_1 A_{11})_{ij} = 0 \quad \text{for all } j \text{ such that} \quad (A_1 x)_j \neq (A_1 z)_j. \end{array} \tag{3}$$
- For discrete-time models, if $(H_1 A_1 x^+(u))_i = (H_1 A_1 z^+(u))_i$ for all $u \in \mathbb{R}^m$ then the same conclusions, as in equation (3), hold.

(b)
- For continuous-time models, if $(H_2 A_2 x_u(t))_i = (H_2 A_2 z_u(t))_i$ for all $u \in \mathbb{R}^m$ and all $t \in [0, \epsilon_u)$ then we have:
$$\begin{array}{l} (H_2 A_{22} A_2 x)_i = (H_2 A_{22} A_2 z)_i, \\ (H_2 A_{21})_{ij} = 0 \quad \text{for all } j \text{ such that} \quad (A_1 x)_j \neq (A_1 z)_j. \end{array} \tag{4}$$
- For discrete-time models, if $(H_2 A_2 x^+(u))_i = (H_2 A_2 z^+(u))_i$ for all $u \in \mathbb{R}^m$ then the same conclusions, as in equation (4), hold.

Proof. (a) Suppose we are in the continuous-time case, and fix $u \in \mathbb{R}^m$. If $(H_1 A_1 x_u(t))_i = (H_1 A_1 z_u(t))_i$ for all $t \in [0, \epsilon_u)$ then $(H_1 A_1 \dot{x}_u(t))_i|_{t=0} = (H_1 A_1 \dot{z}_u(t))_i|_{t=0}$. The previous equation means:

$$\begin{array}{l} \sum_{j=1}^p (H_1 A_{11})_{ij} \sigma((A_1 x)_j + (B_1 u)_j) + (H_1 A_{12} A_2 x)_i = \\ \sum_{j=1}^p (H_1 A_{11})_{ij} \sigma((A_1 z)_j + (B_1 u)_j) + (H_1 A_{12} A_2 z)_i. \end{array} \tag{5}$$

Since this equation holds for all $u \in \mathbb{R}^m$, applying Lemma 3.1 our conclusions follow.

The proof for the discrete-time case is the same since $(H_1 A_1 x^+(u))_i = (H_1 A_1 z^+(u))_i$, for all $u \in \mathbb{R}^m$, implies directly that equations (5) hold.

(b) The proof of this statement is similar to the proof of the previous one. In fact, using the same arguments as in (a), one easily sees that, for both cases, the assumptions imply:

$$\begin{array}{l} \sum_{j=1}^p (H_2 A_{21})_{ij} \sigma((A_1 x)_j + (B_1 u)_j) + (H_2 A_{22} A_2 x)_i = \\ \sum_{j=1}^p (H_2 A_{21})_{ij} \sigma((A_1 z)_j + (B_1 u)_j) + (H_2 A_{22} A_2 z)_i. \end{array}$$

Now it is sufficient to use again Lemma 3.1 to conclude. ∎

Lemma 3.3 If $x \sim z$, then $C_2 A_2 x = C_2 A_2 z$ and $(A_1 x)_i = (A_1 z)_i$ for all $i \notin I_{C_1}$

Proof. If $x \sim z$ then, for all $u \in \mathbb{R}^m$, we have $Cx^+(u) = Cz^+(u)$, for discrete-time, or $Cx_u(t) = Cz_u(t)$, for $t \in [0, \epsilon_u)$, for continuous-time. In both cases, this implies:

$$C_1 \vec{\sigma}(A_1 x + B_1 u) + C_2 A_2 x + C_2 B_2 u = C_1 \vec{\sigma}(A_1 z + B_1 u) + C_2 A_2 z + C_2 B_2 u$$

for all $u \in \mathbb{R}^m$. Applying Lemma 3.1 our conclusions follow. ∎

Lemma 3.4 If $x \sim z$ then for all $q \geq 0$ we have:

1. $C_2 A_{22}^q A_2 x = C_2 A_{22}^q A_2 z$
2. $(C_2 A_{22}^q A_{21})_{ij} = 0$ for all i and all j such that $(A_1 x)_j \neq (A_1 z)_j$.

Proof. 1. Induction on $q \geq 0$. The case $q = 0$ is the first conclusion of Lemma 3.3; thus we may assume that 1. holds for q and for all indistinguishable pairs. First assume we are dealing with continuous-time models. Notice that if $x \sim z$, then $x_u(t) = z_u(t)$ for all $u \in \mathbb{R}^m$ and all $u \in [0, \epsilon_u)$. Thus, by the inductive assumption, we have

$$C_2 A_{22}^q A_2 x_u(t) = C_2 A_{22}^q A_2 z_u(t). \tag{6}$$

Now we can apply Lemma 3.2 (part (b), first equality, with $H_2 = C_2 A_{22}^q$) and we get

$$C_2 A_{22}^{q+1} A_2 x = C_2 A_{22}^{q+1} A_2 z$$

as desired. The proof in discrete time is the same, with $x^+(u)$ and $z^+(u)$ replacing $x_u(t)$ and $z_u(t)$.
2. We again apply Lemma 3.2 (part (b), second equality) to equation (6), and we obtain

$$(C_2 A_{22}^q A_{21})_{ij} = 0 \quad \text{for all } j \text{ such that } \ (A_1 x)_j \neq (A_1 x)_j.$$

Similarly in discrete time. ∎

Lemma 3.5 Let $1 \le i \le k$, and $x,\, z \in \mathbb{R}^n$; assume that for all $x',\, z' \in \mathbb{R}^n$ such that $(x,z) \rightsquigarrow (x',z')$ we have:

$$(A_1 x')_i = (A z')_i. \tag{7}$$

Then, for all $q \ge 0$,:

(a) $(A_{12} A_{22}^q A_2 \xi)_i = (A_{12} A_{22}^q A_2 \zeta)_i$ for all ξ, ζ such that $(x,z) \rightsquigarrow (\xi,\zeta)$;

(b) $(A_{12} A_{22}^q A_{21})_{ij} = 0$ for all j such that there exists a pair ξ, ζ such that $(x,z) \rightsquigarrow (\xi,\zeta)$, and $(A_1 \xi)_j \neq (A_1 \zeta)_j$.

Proof.
(a) We will prove this part by induction on q. Let ξ, ζ such that $(x,z) \rightsquigarrow (\xi,\zeta)$. If $q = 0$ equation (7) implies that, for all $u \in \mathbb{R}^m$, and for all $t \in [0, \epsilon_u)$, $(A_1 \xi_u(t))_i = (A_1 \zeta_u(t))_i$, for continuous-time models, and $(A_1 \xi^+(u))_i = (A_1 \zeta^+(u))_i$, for discrete-time models. In any case, applying Lemma 3.2 part (a) (with $H_1 = I$), we get:

$$(A_{12} A_2 \xi)_i = (A_{12} A_2 \zeta)_i$$

as desired. Assume now that the statement is true for q, this in particular implies that:

$$\begin{array}{ll} (A_{12} A_{22}^q A_2 \xi_u(t))_i = (A_{12} A_{22}^q A_2 \zeta_u(t))_i & \text{for cont. time} \\ (A_{12} A_{22}^q A_2 \xi^+(u))_i = (A_{12} A_{22}^q A_2 \zeta^+(u))_i & \text{for discr. time,} \end{array} \tag{8}$$

and these equations hold for all $u \in \mathbb{R}^m$, and for all $t \in [0, \epsilon_u)$. From Lemma 3.2 part (b) (with $H_2 = A_{12} A_{22}^q$), the inductive step follows.
(b) Fix j such that there exists a pair ξ, ζ such that $(x,z) \rightsquigarrow (\xi,\zeta)$, and $(A_1 \xi)_j \neq (A_1 \zeta)_j$. Notice that both equations in (8) hold for this pair ξ, ζ, thus the second conclusion of Lemma 3.2 part (b) gives:

$$(A_{12} A_{22}^q A_{21})_{ij} = 0$$

as desired. ∎

Since if two states x, z are indinstinguishable it is clear that $x - z \in \ker C$, next Proposition shows that the conditions stated in (2) are necessary for indistinguishability.

Proposition 3.6 If $x \sim z$ then $A_1(x-z) \in V$ and $A_2(x-z) \in W$.

Proof. Given $x \sim z$ we let:

$$J := \{\, i \,|\, \exists\, (x',z'), \ \text{ with } \ (x,z) \rightsquigarrow (x',z'), \ \text{ and } \ (A_1 x')_i \neq (A_1 z')_i \,\}.$$

Then we define:

$$\hat{V} := \ \text{span} \ \{\, e^{(i)} \,|\, i \in J \,\},$$

where $e^{(i)}$ are the vectors of the canonical base in $\mathbb{R}^n$, and

$$\hat{W} := \ \text{the largest } A_{22}\text{-invariant subspace, contained in } \ker C_2, \text{ such that } A_{12}\hat{W} \subseteq \hat{V}.$$

We will prove that:

(i) $\hat{V}$ is A_{11}-invariant, $\hat{V} \subseteq \ker C_1$, and $A_{21}\hat{V} \subseteq \hat{W}$.

(ii) $\tilde{W} = \{\, A_{22}^q (A_2 x' - A_2 z') \,|\, q \ge 0,\, (x,z) \rightsquigarrow (x',z') \} \subseteq \hat{W}$.

From these facts we get our statement since:

- $A_1(x-z) \in \hat{V}$, by definition,
- $A_2(x-z) \in \hat{W}$, by (ii),
- $\hat{V} \subseteq V$, and $\hat{W} \subseteq W$, by (i).

Thus we need to prove (i), and (ii).

(i)

- A_{11}-*invariance.*
  $\hat{V}$ is A_{11} invariant if and only if $(A_{11})_{ij} = 0$ for all i, j such that $j \in J$, and $i \notin J$. Now fix $j \in J$ and $i \notin J$. Then there exists (ξ, ζ) such that $(x, z) \rightsquigarrow (\xi, \zeta)$ and $(A_1\xi)_j \neq (A_1\zeta)_j$. Since $i \notin J$ we have

  $$\begin{array}{rcl} (A_1\xi_u(t))_i & = & (A_1\zeta_u(t))_i \\ (A_1\xi^+(u))_i & = & (A_1\zeta^+(u))_i \end{array} \quad \forall u \in \mathbb{R}^m, t \in [0, \epsilon_u)$$

  in continuous and discrete time respectively. Now we apply Lemma 3.2 (part (a), with $H_1 = I$) and we get

  $$(A_{11})_{iq} = 0 \ \ \forall\, q \ \text{ such that } \ (A_1\xi)_q \neq (A_1\zeta)_q.$$

  In particular $(A_{11})_{ij} = 0$ as desired.
- $\hat{V} \subset \ker C_1$.
  If $x \sim z$ and $(x, z) \rightsquigarrow (x', z')$ then $x' \sim z'$. Thus, Lemma 3.3 yields $J \subset I_{C_1}$ from which the conclusion follows.
- $A_{21}\hat{V} \subset \hat{W}$.
  It is sufficient to prove

  $$\left\{ \begin{array}{l} C_2 A_{22}^q A_{21} e^{(j)} = 0 \\ A_{12} A_{22}^q A_{21} e^{(j)} \in \hat{V} \end{array} \right.$$

  for all $j \in J$, $q \geq 0$ or, equivalently,

  $$\left\{ \begin{array}{ll} (C_2 A_{22}^q A_{21})_{ij} = 0 & \forall j \in J, \forall i \\ (A_{12} A_{22}^q A_{21})_{ij} = 0 & \forall i \notin J, \forall j \in J. \end{array} \right. \tag{9}$$

  The first equality in (9) is easily obtained by applying part (b) of Lemma 3.4. The second one follows from part (b) of Lemma 3.5 after having observed that if $i \notin J$ then $(A_1x')_i = (A_1z')_i$ for all (x', z') such that $(x, z) \rightsquigarrow (x', z')$, and if $j \in J$ there exists (x', z') such that $(x, z) \rightsquigarrow (x', z')$ and $(A_1x')_j \neq (A_1z')_j$.

(ii). $\tilde{W}$ is by definition A_{22}-invariant. Thus to prove that $\tilde{W} \subset \hat{W}$ we need to show that $\tilde{W} \subset \ker C_2$ and $A_{12}\tilde{W} \subset \hat{V}$. This amounts to establish the following identities:

$$C_2 A_{22}^q A_2 x' = C_2 A_{22}^q A_2 z' \ \ \forall (x', z') \ \text{ such that } \ (x, z) \rightsquigarrow (x', z') \tag{10}$$

$$A_{12} A_{22}^q A_2 (x' - z') \in \hat{V} \ \ \forall (x', z') \ \text{ such that } \ (x, z) \rightsquigarrow (x', z'). \tag{11}$$

Since $(x, z) \rightsquigarrow (x', z')$ implies $x' \sim z'$, (10) is just part (a) of Lemma 3.4. Moreover (11) is equivalent to

$$(A_{12} A_{22}^q A_2 x')_i = (A_{12} A_{22}^q A_2 z')_i \ \ \forall i \notin J$$

that follows from part (a) of Lemma 3.5. ∎

Now we prove Theorem 1.

Proof of Theorem 1. As observed before, Proposition 3.6 shows that the conditions stated in (2) are necessary for indistinguishability; thus we only need to prove the sufficiency part.

Assume first that we are dealing with the discrete-time case. We show that if $x, z \in \mathbb{R}^n$ satisfy the conditions in (2) then also $x^+(u), z^+(u)$ satisfy the same conditions for all $u \in \mathbb{R}^m$. This fact clearly implies that $x \sim z$. The following implications hold:

$$\begin{array}{l} A_1(x-z) \in V \Rightarrow x_1^+(u) - z_1^+(u) \in V \\ A_2(x-z) \in W \Rightarrow x_2^+(u) - z_2^+(u) \in W. \end{array} \tag{12}$$

The first implication follows from the fact that V is a coordinate subspace (for this type of subspace if $\alpha \in V$ then also $\vec{\sigma}(\alpha) \in V$), while the second is easy. Since $V \subseteq \ker C_1$, and $W \subseteq \ker C_2$, (12) yields $x^+(u) - z^+(u) \in \ker C$. Moreover V is A_{11}-invariant, and $A_{12}W \subseteq V$, and so we have:

$$A_1(x^+(u) - z^+(u)) = A_{11}(x_1^+(u) - z_1^+(u)) + A_{12}(x_2^+(u) - z_2^+(u)) \in V.$$

Finally, A_{22}-invariance of W, and the fact that $A_{21}V \subseteq W$, gives:

$$A_2(x^+(u) - z^+(u)) = A_{12}(x_1^+(u) - z_1^+(u)) + A_{22}(x_2^+(u) - z_2^+(u)) \in W.$$

Thus also the pair $(x^+(u), z^+(u))$ satifies equation (2), as desired.

Assume now that we are dealing with the continuous-time models. For a fixed but arbitrary input signal $(u(t))_{t\geq 0}$, let $x(t)$, $z(t)$ denote the corresponding solutions of (1), associated to initial conditions $x(0)$, $z(0)$. The pair $(x(t), z(t))$ solves the differential equation in $\mathbb{R}^{2n}$

$$\begin{pmatrix} \dot{x} \\ \dot{z} \end{pmatrix} = F(x, z) \tag{13}$$

where

$$F(x,z) \;=\; \begin{pmatrix} \vec{\sigma}(A_1x + B_1u) \\ A_2x + B_2u \\ \vec{\sigma}(A_1z + B_1u) \\ A_2z + B_2u \end{pmatrix}.$$

Let $Z = \{(x,z) \in \mathbb{R}^{2n} : x - z \in \ker C, A_1(x-z) \in V, A_2(x-z) \in W\}$. In the proof for the discrete time case we showed that if $(x,z) \in Z$ then $F(x,z) \in Z$. Thus Z is stable for the flow of (13), i.e. if $(x(0), z(0)) \in Z$ then $(x(t), z(t)) \in Z$. Since $(u(t))_{t\geq 0}$ is arbitrary and $Z \subset \ker C$ this completes the proof.

4 Proof of Proposition 2.2

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:

$$\{\, i \,|\, \exists\, j \in J_\infty \ \text{ so that } \ (A_{11})_{ji} \neq 0 \,\} \subseteq J_\infty. \tag{14}$$

$$\{ i \,|\, \exists\, j, \exists l \geq 0 \ \text{ so that } \ (C_2 A_{22}^l A_{21})_{ji} \neq 0 \,\} \subseteq J_\infty. \tag{15}$$

$$\{ i \,|\, \exists\, j \in J_\infty, \exists l \geq 0 \ \text{ so that } \ (A_{12} A_{22}^l A_{21})_{ji} \neq 0 \,\} \subseteq J_\infty. \tag{16}$$

$$W \subset \{\, w \,|\, A_{22}^l w \in \ker C_2 \ \text{ for } \ l \geq 0, A_{12} A_{22}^l w \in V_\infty \ \text{ for } \ l \geq 0 \,\}. \tag{17}$$

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 (14), 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 (17). Moreover, for $1 \leq s \leq d$, it follows from (17) 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 (15) we get:

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

and from (16) 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 (17).

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, $V \subset V_s$.

References

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

[2] Albertini F., and Dai Pra P., "Observability for discrete time recurrent neural networks coupled with linear systems", submitted for ECC '95.

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

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

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

[6] Albertini, F., and E.D. Sontag, "Uniqueness of weights for Recurrent Nets," *Proc. Mathematical Theory of Networks and Systems*, Regensburg, August, 1993, to appear.

[7] 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: 115-125.

[8] Hebb, D.O., *The organization of behavior*, Wiley, New York, 1949.

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

[10] McCulloch, W.S., and W. Pitts, "A logical calculus of the ideas immanent in nervous activity", *Bull.Math.Biophys.* 5 (1943): 115-133.

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

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

[13] Sontag, E.D., *Some topics in neural networks and control*, Siemens Corporate Research, Inc., Report LS93-02, 1993.

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