FORWARD ACCESSIBILITY FOR RECURRENT NEURAL NETWORKS

Francesca Albertini, Paolo Dai Pra

Abstract— **We give an algebraic characterization of forward accessibility for recurrent neural networks.**

Keywords— **recurrent networks, controllability, forward accessibility.**

I. Introduction.

In this paper we deal with control systems which evolve either in discrete or in continuous time and whose dynamics is given by:

$$\begin{array}{rcl} x(t+1) \text{ (or } \dot{x}(t)) & = & \vec{\sigma}(Ax(t)+Bu(t)) \\ x(0) & = & x_0 \end{array} \qquad (1)$$

where $x(t) \in \mathbb{R}^n$, $u(t) \in \mathbb{R}^m$, $A \in \mathbb{R}^{n\times n}$, and $B \in \mathbb{R}^{n\times m}$. Moreover

$$\vec{\sigma}(x) = (\sigma(x_1), \ldots, \sigma(x_n))$$

where $\sigma : \mathbb{R} \to \mathbb{R}$ is some assigned nonlinear function. For continuous-time models we always assume that the control function $u(\cdot)$ is measurable essentially bounded and the map σ is at least locally Lipschitz, so that the differential equation has an unique local solution. Systems of this type, often together with a linear output equation $y(t) = Cx(t)$, are called *recurrent neural networks*.

Recurrent neural networks are often used as identification models or as prototype dynamic controllers in many applications such as speech processing, signal processing (see [5]), and control (see [6]). A typical problem in these applications consists in determining the 'weights' of the network (i.e. the entries of the matrices in (1)) that provide the best fit to some training data, minimizing a given error functional. For an extensive introduction on neural computation and other applications of neural networks we refer to [4].

The purpose of this paper is to continue the analysis of the system-theoretic properties of these models. In [1], [2], and [3] characterizations of identifiability and observability have been given. More precisely, under nonlinearity conditions on $\sigma(\cdot)$ and non-degenericity conditions on the matrix B, necessary and sufficient conditions for identifiability and observability are given, in terms of algebraic assumptions on the matrices A, B, C. In this paper we give similar results concerning controllability of a system of type (1).

The study of these system-theoretic properties is important from a pure mathematical point of view. In fact, neural networks are a very natural generalizations of linear systems. They provide 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. Besides this mathematical interest the study of these basic properties is also motivated by the applications of neural networks. As we said before, usually recurrent networks are used as models whose parameters must be fit to input/output data, minimizing a cost function; to perform this procedure some algorithms are used (such as gradient descent). In this contest many numerical and theoretical issues arise. One question that can be asked is about the possibility of having different networks which give the same behavior; among these models one would like to choose the one with the best features, like, for example, least dimension (minimality: we say that a network is minimal if no other networks have the same input/output behavior with state space having strictly lower dimension). Connections between minimality, controllability and observability are well known for linear systems, but only weaker results can be given for general smoooth nonlinear models (see [8]). In [2], and in the present paper we contrubite to giving sharper results in this direction for recurrent neural networks.

Universita' di Padova, Dipartimento di Matematica Pura e Applicata, Via Belzoni 7, 35100 Padova, Italy. E-mail: albertin@hilbert.unipd.it, daipra@pdmat1.unipd.it, fax:39-49-8758596

A system is said to be *controllable* if for every two states x_1, x_2 there is a sequence of controls $u(0), \cdots, u(k)$ in the discrete time case, or a control function $u(t)$ for $t \in [0, T]$ in the continuous-time case, which steers x_1 to x_2. This notion is quite strong, and is usually very hard to study for nonlinear systems. We therefore restrict our study to the weaker notion of *forward accessibility*: a control system is said to be forward accessible if, for every initial state x_0, the set of points to which x_0 can be steered contains an open subset of the state-space. In particular, this property implies that the system is not confined in a submanifold of the state space with positive codimension. Accessibility is a natural and relevant property of nonlinear control systems. Unlike for linear systems, we will see that the accessibility conditions for models of type (1), and their relations with the minimality of the network, differ substantially from discrete to continuous time.

We first looked (section II) at the discrete-time case. In this context, our main result is the following: under some non-degenericity conditions on $\sigma(\cdot)$ and B, a system of type (1) is forward accessible if and only if rank $[A, B] = n$. Notice that this rank condition is weaker than controllability for the linear system ($\sigma(x) = x$), which is rank $[A - \lambda I, B] = n$ for every $\lambda \in \mathbb{C}$. The assumptions on $\sigma(\cdot)$ and B we need to make to get this result are stronger than the ones in [1], [2], and [3]; however for the choice $\sigma(x) = \tanh(x)$, such assumptions are satisfied for B in a simple dense open subset of $\mathbb{R}^{n\times m}$ (see section II-C).

This first part is organized as follows. In section II-A we prove our main result for single-input systems ($m = 1$); the extension to the general case is given in section II-B. Section II-C is devoted to the discussion of the assumptions on $\sigma(\cdot)$ and B. Section II-D gives a different kind of sufficient condition for forward accessibility using a weaker condition on the map σ and adding a non-degenericity condition on the pair A, B.

In the second part of the paper (section III) we deal with continuous-time dynamics where the characterization of forward accessibility is much easier.

II. Forward accessibility for discrete-time models

In this first part we deal with discrete-time dynamics. From now on, we assume that a model of type (1) evolving in discrete-time is given, i.e., a system where the dynamics is given by the difference equation:

$$\begin{array}{rcl} x(t+1) & = & \vec{\sigma}(Ax(t)+Bu(t)) \\ x(0) & = & x_0. \end{array} \qquad (2)$$

A. *Forward accessibility in the non-degenerate case for single input models.*

In this section we restrict our attention to single input systems of type (2), i.e. where the dynamics is given by:

$$x(t+1) = \vec{\sigma}(Ax(t)+bu(t)) \qquad (3)$$

with $b \in \mathbb{R}^n$, $u(t) \in \mathbb{R}$. Our aim is to give necessary and sufficient conditions for these models to be forward accessible. These conditions will be subject to a non degenericity assumption that will be given as a joint property of the activation function σ and the vector b.

Definition II.1: We say that the function σ and the vector b satisfy the *n-independence property* (n-**IP**) if the following conditions hold:

1. σ is differentiable, $\sigma'(x) \neq 0$ for all $x \in \mathbb{R}$;
2. $b_i \neq 0$ for all $i = 1, \ldots, n$;
3. for $1 \leq k \leq n$ let O_k be the set of all the subsets of $\{1, \ldots, n\}$ of cardinality k, and $a_1, \ldots, a_n$ arbitrary real numbers. Then the functions $\{f_I : I \in O_k\}$ with
$$f_I(x) = \prod_{i \in I} \sigma'(a_i + b_i x),$$
are linearly independent.

Remark II.2: The n-**IP** property and some weaker properties used later in the paper, are strong and presumably not necessary for our results. Those assumptions, in particular, exclude linear systems ($\sigma(x) = x$). Our techniques do not allow to fill the gap to linear systems, since we strongly use nonlinearity of $\sigma(\cdot)$.

In section II-C we discuss more this property.

Now, we state the main result of this paper.

Theorem 1: Let Σ be a system of type (3), such that σ and b satisfy the n-**IP** property. Then Σ is forward accessible if and only if
$$\operatorname{rank}[A, b] = n.$$
Notice that for $n = 1$ the result is trivial, thus we assume $n \geq 2$. The proof of Theorem 1 is preceded by some technical lemmas. The leading idea is standard: we look for conditions that guarantee the input-state map $(u(1), \ldots, u(n)) \to x(n)$ has a non-degenerate Jacobian, for some value of the control sequence. This is, of course, a sufficient conditions for forward accessibility. The computation of the Jacobian matrix of the input-state map reveals a nontrivial algebraic structure.

Consider an initial state x_0, a control sequence $u(0), \ldots, u(n-1)$, and let $x(0) = x_0$, $x(i) = \vec{\sigma}(Ax(i-1) + bu(i-1))$ for $i = 1, \ldots, n$. The state $x(n)$ is a function of $u(0), \ldots, u(n-1)$, and we let W_n be the Jacobian matrix $\partial_{u(0),\ldots,u(n-1)} x(n)$. For each $x \in \mathbb{R}^n$ we denote by $\hat{\sigma}(x)$ the diagonal matrix whose i-th element is $\sigma'(x_i)$, and we define, for $i = 1, \ldots, n$:
$$D(i) = \hat{\sigma}(Ax(i-1) + bu(i-1)). \tag{4}$$

Lemma II.3: Let define, for $i = 1, \ldots, n-2$
$$g(i) = AD(n-2)AD(n-3)\cdots AD(i)b, \tag{5}$$
and $g(n-1) = g(n) = b$. Then
$$\begin{aligned} W_n \;=\; & D(n)[AD(n-1)g(1), AD(n-1)g(2), \ldots, \\ & AD(n-1)g(n-1), g(n)]. \end{aligned} \tag{6}$$

Proof: To get our result, we prove by induction on $l \geq 1$, that for all $0 \leq k \leq l-1$,
$$\partial_{u(k)} x(l) = D(l)A \cdots AD(k+1)b.$$
Let $l = 1$, then $x(1) = \vec{\sigma}(Ax_0 + bu(0))$. This implies that $\partial_{u(0)}(x(1))_i = \sigma'((Ax_0)_i + b_i u(0))b_i$; so we have $\partial_{u(0)} x(1) = \hat{\sigma}(Ax_0 + bu(0))b$, as desired.

Assume now that $l > 1$, then we have:
$$\begin{aligned} \partial_{u(k)} x(l) &= \partial_{u(k)}[\sigma(Ax(l-1) + bu(l-1))] \\ &= \begin{cases} D(l)b & \text{if } k = l-1, \\ D(l)A\partial_{u(k)} x(l-1) & \text{if } k < l-1, \end{cases} \end{aligned}$$
which, by inductive assumption, implies the thesis. ■

Due to assumption 1. on $\sigma'(\cdot)$, it is clear that the matrix $D(n)$ is non-singular. Therefore W_n is singular if and only if the matrix
$$\begin{aligned} Z_n \;=\; & [AD(n-1)g(1), AD(n-1)g(2), \ldots, \\ & AD(n-1)g(n-1), g(n)] \end{aligned}$$
is singular. In the next lemma we compute the determinant of Z_n (notice that Z_n does not depend on $u(n-1)$, which appears only in $D(n)$).

Lemma II.4: For any control sequence $u(0), \ldots, u(n-2)$ we have:
$$\det Z_n = \sum_{i=1}^{n} \det A(i) \det G(i) \prod_{j=1, j\neq i}^{n} \sigma'(y_j + b_j u(n-2)),$$
where $y_j = (Ax(n-2))_j$, and
$$A(i)_{l,k} = \begin{cases} a_{l,k} & \text{if } k < i \\ a_{l,k+1} & \text{if } i \leq k < n \\ b_l & \text{if } k = n, \end{cases}$$
$$G(i)_{l,k} = \begin{cases} (AD(n-2)\cdots AD(k)b)_l & \text{for } l < i \\ (AD(n-2)\cdots AD(k)b)_{l+1} & \text{for } l \geq i, \end{cases}$$
with, for $n = 2$, $G(1) = b_2$, and $G(2) = b_1$. Notice that $A(i) \in \mathbb{R}^{n\times n}$ and $G(i) \in \mathbb{R}^{(n-1)\times(n-1)}$.

Proof: We denote by S_n the set of permutations on $\{1, \cdots, n\}$.
$$\det Z_n = \sum_{\pi \in S_n} (\operatorname{sgn} \pi) \prod_{i=1}^{n-1}$$
$$\left(\sum_{k=1}^{n} a_{\pi(i),k} \sigma'(y_k + b_k u(n-2)) g_k(i) \right) g_{\pi(n)}(n) =$$
$$= \sum_{k_1,\cdots,k_{n-1}} \left(\sum_{\pi \in S_n} (\operatorname{sgn} \pi) a_{\pi(1),k_1} \cdots a_{\pi(n-1),k_{n-1}} g_{\pi(n)}(n) \right)$$
$$\prod_{j=1}^{n-1} g_{k_j}(j) \prod_{j=1}^{n-1} \sigma'(y_{k_j} + b_{k_j} u(n-2)).$$
Letting
$$\phi(j, i) = \begin{cases} j & \text{if } j < i \\ j+1 & \text{if } j \geq i, \end{cases}$$
we have:
$$\det Z_n = \sum_{i=1}^{n} \left[\sum_{\tau \in S_{n-1}} \right.$$
$$[\sum_{\pi \in S_n} (\operatorname{sgn} \pi) a_{\pi(1),\tau(\phi(1,i))} \cdots a_{\pi(n-1),\tau(\phi(n-1,i))} g_{\pi(n)}(n)]$$
$$\left. \prod_{j=1}^{n-1} g_{\tau(\phi(j,i))}(j) \right] \prod_{j=1, j\neq i}^{n-1} \sigma'(y_j + b_j u(n-2)) =$$

$$= \sum_{i=1}^{n} \det A(i) \det G(i) \prod_{j=1, j\neq i}^{n} \sigma'(y_j + b_j u(n-2)).$$

■

The expression for $\det Z_n$ given in lemma II.4 suggests how the proof of Theorem 1 should go. In fact, by the n-**IP** property, the functions

$$\{ \prod_{j\neq i} \sigma'(y_j + b_j u(n-2)) \; : \; i = 1, \ldots, n \}$$

are linearly independent. Therefore, in order for $\det Z_n$ to be zero for every input-sequence it must be $\det A(i) \det G(i) = 0$ for $i = 1, \ldots, n$. The relation between $\det A(i)$ and $\det G(i)$ is analyzed in the following lemmas.

Lemma II.5: For $k = 0, 1, \ldots, n-2$ we let

$$i^{(k)} = \{ i_1^{(k)}, \ldots, i_k^{(k)} \}$$

be an ordered subset of $\{1, \ldots, n\}$, and we define $G(i^{(k)}) \in \mathbb{R}^{(n-k)\times(n-k)}$ by

$$G(i^{(k)})_{r,s} \; = \; (AD(n-k-1) \cdots AD(s)b)_{\phi(r, i^{(k)})}$$

where $\phi(r, i^{(k)}) = r + |\{j \, : \, i_j^{(k)} \leq r\}|$ (for $k = 0$ we define $G(\emptyset)_{r,n} = b_r$ and for $k = n-1$, $G(i^{(n-1)}) = b_i$ with $i \notin i^{(n-1)}$).

Then, for $0 \leq k \leq n-2$, we have:

$$\det G(i^{(k)}) = \sum_{i^{(k+1)}} \det A(i^{(k)}, i^{(k+1)}) \det G(i^{(k+1)})$$

$$\prod_{i \notin i^{(k+1)}} \sigma'((Ax(n-k-2))_j + b_j u(n-k-2))$$

where $A(i^{(k)}, i^{(k+1)}) \in \mathbb{R}^{(n-k)\times(n-k)}$ is obtained from A by:

- removing the rows $i_1^{(k)}, \ldots, i_k^{(k)}$;
- removing the columns $i_1^{(k+1)}, \ldots, i_{k+1}^{(k+1)}$;
- adding, as a last column, the b without the components in $i^{(k)}$.

Proof: Observe that for $k = 0$, $G(\emptyset) = Z_n$. So lemma II.5 for $k = 0$ coincides with lemma II.4. The proof for the other values of k follows the same lines, and we omit it. ■

Next lemma is a simple technical fact which will be used later; its proof is easily established.

Lemma II.6: Let V be a vector space, and $v_1, \ldots, v_p, w$ be $p+1$ vectors in V, with $w \neq 0$. Assume that for all $i = 1, \ldots, p$ the vectors $\{v_j, \, w \, | j \neq i\}$ are linearly dependent. Then also the vectors $\{v_1, \ldots, v_p\}$ are linearly dependent.

Now we can prove our main result.

Proof of Theorem 1. First assume $\operatorname{rank}[A, b] < n$. Then let V be the range of the map:

$$\begin{array}{ccc} \mathbb{R}^{n+1} & \rightarrow & \mathbb{R}^n \\ (x, u) & \mapsto & Ax + bu. \end{array}$$

We have $\dim V = \operatorname{rank}[A, b] < n$. The reachable set from any point is contained in $\vec{\sigma}(V)$, which, clearly, does not contain any open set.

Now assume that $\operatorname{rank}[A, b] = n$. We will show that $\det Z_n \neq 0$ for some $u(0), \ldots, u(n-2)$, and this implies forward accessibility. By the way of contradiction, suppose $\det Z_n = 0$ for every $u(0), \ldots, u(n-2)$. As we have already observed, it follows that, for every $i = 1, \ldots, n$, and every input sequence $u(0), \ldots, u(n-3)$:

$$\det A(i) \det G(i) \; = \; 0.$$

Notice that $A(i)$ is a constant matrix, while $G(i)$ is a function of $u(0), \ldots, u(n-3)$. Suppose that there is $i \in \{1, \ldots, n\}$ such that $\det A(i) \neq 0$. Then, necessarily, $\det G(i) = 0$ for every $u(0), \ldots, u(n-3)$. We expand $\det G(i)$ using lemma II.5:

$$\det G(i) \; = \; \sum_{i^{(2)}} \det A(i, i^{(2)}) \det G(i^{(2)})$$

$$\prod_{j \notin i^{(2)}} \sigma'((Ax(n-3))_j + b_j u(n-3)).$$

By the n-**IP** property, we have that, for every $i^{(2)}$ and every $u(0), \ldots, u(n-3)$,

$$\det A(i, i^{(2)}) \det G(i^{(2)}) = 0.$$

Now we argue as before. We assume that there is an $i^{(2)}$ such that $\det A(i^{(2)}) \neq 0$, and therefore $\det G(i^{(2)}) = 0$, and we expand $\det G(i^{(2)})$ using lemma II.5. We proceed in this way until we find an $i^{(k)}$ such that $\det G(i^{(k)}) = 0$ and $\det A(i^{(k)}, i^{(k+1)}) = 0$ for every $i^{(k+1)}$. Notice that this must be true for some $k \leq n-2$, since, otherwise, there would be some $i^{(n-1)}$ with $\det G(i^{(n-1)}) = 0$; this is impossible since $G(i^{(n-1)}) = b_j \neq 0$, with $\{j\} = \{1, \ldots, n\} \setminus i^{(n-1)}$.

So let $i^{(k)}$ be such that $\det A(i^{(k)}, i^{(k+1)}) = 0$ for every $i^{(k+1)}$. We show that this implies that, for every i, $\det A(i) = 0$, so that the above procedure indeed stops at the first step. The determinant $\det A(i)$ can be expanded with respect to the $n-k$ rows in $i^{(n-k)} = \{1, \ldots, n\} \setminus i^{(k)}$. All complementary matrices in this expansion are either of the form $A(i^{(k)}, i^{(k+1)})$, whose determinant is zero, or are submatrices of A of dimension $n-k$, whose determinant is again zero by lemma II.6. Therefore $\det A(i) = 0$ for $i = 1, \ldots, n$, which means that b and any $n-1$ columns of A are linearly dependent. Then, again by lemma II.6, also the n columns of A are linearly dependent; thus we have that $\operatorname{rank}[A, b] < n$, which contradicts our assumption. ■

Remark II.7: The n-**IP** property in definition II.9 is stronger than a similar property used in [2] for proving identifiability results for recurrent neural networks (see section II-C for the precise definition). In particular, it follows from [2] (Theorem 1), that any *observable* recurrent network satisfying the n-**IP** property is *minimal*. As a consequence of Theorem 1, these minimal networks are not necessarily forward accessible, since observability does not imply $\operatorname{rank}[A, b] = n$ (see [3]).

Remark II.8: Assume that $\operatorname{rank}[A, b] < n$, and let $V = [A, b](\mathbb{R}^{n+1})$. Then, except for the initial condition, the state in the dynamics (3) is confined in $\vec{\sigma}(V)$. Suppose also, for simplicity, $x(0) = 0$, and $\sigma(0) = 0$. Then the control system with state space V, control space $\mathbb{R}$ and dynamics:

$$\begin{array}{rcl} v(t+1) & = & A\vec{\sigma}(v(t)) + Bu(t), \quad v(0) = 0, \\ y(t) & = & C\vec{\sigma}(v(t)), \end{array} \tag{7}$$

has the same input/output behavior of (3) with output equation $y(t) = Cx(t)$. The state space of (7) has lower dimension than the one in (3), but (7) is not a neural network. Thus this does not contradict the minimality result given in [2]. We believe, however, that it is of some interest the fact that, supposing the n-**IP** , either a network is forward accessible or a state-space dimension reduction can be easily performed as in (7).

B. Multiple input systems.

In this section we extend Theorem 1 to systems of the type:

$$x(t+1) = \vec{\sigma}(Ax(t) + Bu(t)), \tag{8}$$

where $u(t) \in \mathbb{R}^m$ and $B \in \mathbb{R}^{n\times m}$. Although notations become rather heavy, the extension is conceptually straightforward; not all the details will be given.

Definition II.9: We say that the function σ and the matrix B satisfy the n-**IP** property, if

1. σ is differentiable, $\sigma'(x) \neq 0$ for all $x \in \mathbb{R}$;
2. denote by b_i, $i = 1, \dots, n$, the rows of the matrix B; then $b_i \neq 0$ for all $i = 1, \dots, n$;
3. for $1 \leq k \leq n$ let O_k be the set of all the subsets of $\{1, \dots, n\}$ of cardinality k, and $a_1, \dots, a_n$ arbitrary real numbers. Then the functions $\{f_I : I \in O_k\}$, $f_I : \mathbb{R}^m \to \mathbb{R}$ given by:
$$f_I(u) = \prod_{i\in I} \sigma'(a_i + b_i u),$$
are linearly independent.

Theorem 2: Let Σ be a system of type (8), such that σ and B satisfy the n-**IP** property. Then Σ is forward accessible if and only if

$$\text{rank}\,[A, B] = n. \tag{9}$$

Sketch of the proof. The necessity of condition (9) is verified as in Theorem 1. For the converse , we compute the Jacobian matrix

$$W_n = \nabla_{u(0),\dots,u(n-1)} x_n \in \mathbb{R}^{n\times nm}.$$

As in Section II-A we define:

$$D(i) = \hat{\sigma}(Ax(i-1) + Bu(i-1)) \in \mathbb{R}^{n\times n},$$

and

$$g(i) = AD(n-2)AD(n-3)\cdots AD(i)B \in \mathbb{R}^{n\times m},$$

with $g(n-1) = g(n) = B$. We repeat the proof of lemma II.3 and we get:

$$W_n = D(n)[AD(n-1)g(1), \cdots, AD(n-1)g(n-1), g(n)]$$
$$:= D(n)Z_n.$$

Since $D(n)$ is nonsingular, all we have to show is that if (9) holds, then $\text{rank}\, Z_n = n$ for some sequence $u(0), \dots, u(n-2)$ (again $u(n-1)$ appears only in $D(n)$). In what follows we denote by a^i, b^j, the columns of A and B respectively. If $\text{rank}\,[A, B] = n$, then there are indexes $\lambda_1, \dots, \lambda_r, \mu_{r+1}, \dots, \mu_n$ such that the $n \times n$ matrix:

$$C(\lambda, \mu) = [a^{\lambda_1}, \dots, a^{\lambda_r}, b^{\mu_{r+1}}, \dots, b^{\mu_n}]$$

is nonsingular. As in Theorem 1, using Lemma II.6, one show that it is not restrictive to assume that $C(\lambda, \mu)$ has at least one column taken from the columns of B.

Now let $\nu_1, \cdots, \nu_r$ and $\rho_1 \leq \rho_2 \leq \cdots \leq \rho_r$ be elements of $\{1, 2, \cdots, n\}$, and define

$$\begin{aligned} Z_n^{\nu,\rho,\mu} \;=\; & [AD(n-1)g^{\nu_1}(\rho_1), \cdots, \\ & AD(n-1)g^{\nu_r}(\rho_r), b^{\mu_{r+1}}, \cdots, b^{\mu_n}]. \end{aligned}$$

Moreover, for $k = \{k_1, k_2, \cdots, k_r\} \subset \{1, \cdots, n\}$ with $k_1 < k_2 < \cdots < k_r$ we let

$$G(k, \nu, \rho) = \begin{pmatrix} g_{k_1}^{\nu_1}(\rho_1) & \cdots & g_{k_1}^{\nu_r}(\rho_r) \\ \vdots & & \vdots \\ g_{k_r}^{\nu_1}(\rho_1) & \cdots & g_{k_r}^{\nu_r}(\rho_r) \end{pmatrix} \in \mathbb{R}^{r\times r}.$$

Then, similarly to Lemma II.4, we get

$$\det Z_n^{\nu,\rho,\mu} = \sum_k \det C(k,\mu) \det G(k,\nu,\rho) \prod_{j\notin k} \sigma'(y_j + b_j u(n-2))$$

with $y_j = (Ax(n-2))_j$. By using the n-**IP** we have that if Σ is not forward accessible then

$$\det C(k,\mu) \det G(k,\nu,\rho) = 0 \tag{10}$$

for every k, ν, ρ and any sequence $u(1), \cdots, u(n-3)$. At this point one has to prove an analog of Lemma II.5, to give a recursive formula for $\det G(k,\nu,\rho)$. It is not hard to see that, if $\rho_r = n-1$, then the structure of the matrix $G(k,\nu,\rho)$ is similar to the one of $Z_n^{\nu,\rho,\mu}$, and the recursion can be performed. As in the proof of Theorem 1 such recursion has to stop, in the sense that at a certain step one sees that a family of square submatrices of $C(k,\mu)$ must be degenerate, and this implies that $C(k,\mu)$ itself is degenerate (in this part of the argument the arbitrariness of ρ in (10) is used). We therefore obtain $\det C(k,\mu) = 0$ for every k, which is false since $\det C(\lambda,\mu) \neq 0$.

∎

*C. Some remarks on the n-**IP** .*

The n-**IP** property, as stated in Definition II.1, is given as a joint property of $\sigma(\cdot)$ and b. This is not, indeed, what is desirable in applications, since usually σ is a given elementary function. So, for a given activation function σ, we would like to give simple and hopefully "generic" sufficient conditions of b to guarantee that $\sigma(\cdot)$ and b satisfy the n-**IP** . A related problem has appeared in [1], [2], and [3] where, to obtain observability and identifiability of systems of type (2), a property weaker than the n-**IP** has been used. Such property, that will be called **IP**, is obtained from Definition II.1 replacing 3. with

3'. for $a_1, \cdots, a_n$ arbitrary real numbers the functions $\{\sigma(a_i + b_i x) : i = 1, \cdots, n\}$ are linearly independent.

The extension of this property to the multiple input case is similarly obtained from Definition II.9. It has been shown (see [3]) that, for a wide class of activation functions (e.g. $\arctan(\cdot)$ and $\tanh(\cdot)$), the **IP** is implied by the following "genericity" condition on b:

$$b_i \neq 0 \;\text{ and }\; |b_i| \neq |b_j| \;\text{ for every }\; i, j = 1, \cdots, n\;,\; i \neq j. \tag{11}$$

The question is whether, for some commonly used activation function σ, condition (11) implies the n-**IP** too. We first give an example where the answer is negative, which shows that the n-**IP** is strictly stronger than the **IP** .

Suppose $\sigma(x) = \arctan(x)$, $n \geq 4$ and b arbitrary. If the n-**IP** were satisfied, then the functions

$$\sigma'(b_1u)\sigma'(b_2u)\sigma'(b_3u),\;\; \sigma'(b_1u)\sigma'(b_3u)\sigma'(b_4u),$$
$$\sigma'(b_1u)\sigma'(b_2u)\sigma'(b_4u),\;\; \sigma'(b_2u)\sigma'(b_3u)\sigma'(b_4u)$$

would be linearly independent. By multiplying a linear combination of those functions by $[\prod_{i=1}^4 \sigma'(b_iu)]^{-1}$ we get that the functions $\frac{1}{\sigma'(b_iu)}$, $i = 1, \cdots, 4$ are linearly independent. But this is impossible, since $\frac{1}{\sigma'(b_iu)} = 1 + b_i^2u^2$, and 4 polynomials of degree 2 cannot be linearly independent.

The example above seems to be exceptional, since the pathology is caused by the fact that $\sigma'(x)$ is a rational function. Another typical activation function, $\sigma(x) = \tanh(x)$, appears to have a better behavior, in the sense that reasonable conditions on b can be given so that tanh and b satisfy the n-**IP** . To see this let us take vectors $a, b \in \mathbb{R}^n$. We first observe that to show that the functions

$$\{\prod_{i\in I}\sigma'(a_i+b_iu) : I \in O_{n-k}\}$$

are linearly independent is equivalent to showing that

$$\{\prod_{i\in I}\frac{1}{\sigma'(a_i+b_iu)} : I \in O_k\}$$

are linearly independent. In other words we have to give conditions on b so that the functions

$$\{\prod_{i\in I}\cosh^2(a_i+b_iu) : I \in O_k\}$$

are linearly independent, for every $1 \le k \le n-1$. Notice that, for $I \in O_k$ (we write $|I| = k$) and letting $\alpha_i = 2a_i, \beta_i = 2b_i$ we get

$$\begin{aligned}\prod_{i\in I}\cosh^2(a_i+b_iu) &= \frac{1}{2^k}\prod_{i\in I}(1+\cosh(\alpha_i+\beta_iu))\\ &= \frac{1}{2^k}\sum_{J\subset I}\prod_{i\in J}\cosh(\alpha_i+\beta_iu)\\ &= \frac{1}{2^k}\sum_{J\in I}\frac{1}{2^{|J|}}\sum_{\lambda\in\{-1,1\}^J}\cosh(\lambda^T\alpha+\lambda^T\beta u).\end{aligned}$$

Thus, for given real numbers c_I, $I \in O_k$, we have

$$\begin{aligned}&2^k\sum_{I\in O_k}c_I\prod_{i\in I}\cosh^2(a_i+b_iu)\\ &= \sum_{I\in O_k}\sum_{J\subset I}\frac{1}{2^{|J|}}c_I\sum_{\lambda\in\{-1,1\}^J}\cosh(\lambda^T\alpha+\lambda^T\beta u)\\ &= \sum_{J:|J|\le k}\left[\frac{1}{2^{|J|}}\sum_{|I|=k, I\supset J}c_I\right]\sum_{\lambda\in\{-1,1\}^J}\cosh(\lambda^T\alpha+\lambda^T\beta u)\\ &= \sum_{\lambda\in\{-1,0,1\}^n, |\lambda|\le k}\left[\frac{1}{2^{|\lambda|}}\sum_{|I|=k, I\supset \text{ supp } \lambda}c_I\right]\cosh(\lambda^T\alpha+\lambda^T\beta u)\end{aligned} \tag{12}$$

where $|\lambda| = \sum_1^n |\lambda_i|$ and $\text{supp}\lambda = \{i \in \{1, \cdots, n\} : \lambda_i \ne 0\}$. Now we denote by Λ_k the set of those $\lambda \in \{-1, 0, 1\}^n$ such that $|\lambda| = k$, modulo the equivalence relation $\lambda \sim \lambda'$ if $\lambda = \pm\lambda'$, and $\Lambda = \cup_{k=0}^n \Lambda_k = \{-1, 0, 1\}^n/\sim$. We also let g_b be the function $\Lambda \to \mathbb{R}^+$ defined by $g(\lambda) = |\lambda^T b|$.

Proposition II.10: A sufficient condition for tanh and b to satisfy the n-**IP** is that, for every $1 \le k \le n-1$ and every $I \in O_k$, there is a λ_I with $\text{supp } \lambda_I = I$ such that g_b is 1-1 on the set $\Gamma_k = \{\lambda_I : I \in O_k\}$, and

$$g_b(\Gamma_k) \cap g_b((\cup_{h=0}^k \Lambda_h) \setminus \Gamma_k) = \emptyset.$$

In particular this is true if g_b is 1-1 on all Λ.

Proof: This is a straightforward consequence of (12) and the fact that if $h_1, \cdots, h_p$ are real numbers and $k_1, \cdots, k_p$ are positive distinct real numbers, then the functions $\{\cosh(h_i + k_iu) : i = 1, \cdots, p\}$ are linearly independent. ■

We now make some comments on the assumptions in Proposition II.10. First we notice that the assumption on b is "generic", i.e. is satisfied for b in the complement of an analytic subset of $\mathbb{R}^n$ (in particular b may vary in an open dense subset of $\mathbb{R}^n$). Moreover the stronger condition that g_b is 1-1 on all Λ is a natural generalization of (11), that says that g_b is 1-1 on $\Lambda_0 \cup \Lambda_1$. All these conditions, however, appear stronger than what is needed for the n-**IP** . We indeed conjecture that, for every dimension n, conditions (11) implies the n-**IP** for tanh and b. This is trivial to check for $n = 2, 3$ and we could make it for $n = 4$, although we do not show details here.

D. Another sufficient condition

In this section we present another sufficient condition for forward accessibility, using a weaker condition on the map σ but adding a new non-degenericity condition on the pair A, B. We deal with the multiple-input case directly (see equation (8)).

Definition II.11: We say that the function σ and the matrix B satisfy the n-**WIP** (n-*weak independence property*) if:

1. σ is differentiable, $\sigma'(x) \ne 0$ for all $x \in \mathbb{R}$;
2. denote by b_i, $i = 1, \ldots, n$, the rows of the matrix B; then $b_i \ne 0$ for all $i = 1, \ldots, n$;
3. let $a_1, \ldots, a_n$ be arbitrary real numbers, then the functions from $\mathbb{R}^m$ to $\mathbb{R}$ $(\sigma'(a_i+b_iu))^{-1}$, for $i = 1, \ldots, n$, are linearly independent.

Remark II.12: Notice that the n-**WIP** is weaker than the n-**IP** given by definition II.9, in fact the third requirement of the previous definition is exactly the third requirement of definition II.9 for the case $k = n-1$ (see also discussion in section II-C).

We now have the following Theorem:

Theorem 3: Let Σ be a system of type (8), such that σ and B satisfy the n-**WIP** . If there exists a matrix $H \in \mathbb{R}^{m\times n}$ such that:

(a) the matrix $(A + BH)$ is invertible,

(b) the rows of the matrix $[(A + BH)^{-1}B]$ are all non-zero,

then Σ is forward accessible.

Before giving the proof of this Theorem, we make some comparison between this statement and Theorem 2. It is easy to see that condition (a) of the previous Theorem is equivalent to $\text{rank}\,[A, B] = n$, thus (a) is also a necessary condition for forward accessibility. So Theorem 3 adds a new non-degenericity condition on A, B (condition (b)), and guarantees forward accessibility with weaker assumption on σ. Moreover it is interesting to notice that for the single-input case condition (b) is independent on H. In fact the following claim holds:

Claim 1: Let $h, k \in \mathbb{R}^{n\times 1}$ be such that $A + bh^t$ and $A + bk^t$ are invertible. Then

$$((A + bk^t)^{-1}b)_i \ne 0 \quad \forall\, i \quad \iff \quad ((A + bh^t)^{-1}b)_i \ne 0 \quad \forall\, i.$$

Proof: It is not restrictive to assume $k = 0$. Let $w = A^{-1}b$ and $v = (A + bh^t)^{-1}b$, to get our claim it is sufficient to prove that there exists $\lambda \ne 0$ such that $v = \lambda w$. Since $A + bh^t$ is invertible, we have that $h^tw \ne -1$, otherwise $(A + bh^t)w = b - b = 0$. Thus we may let:

$$\lambda = \frac{1}{1 + h^tw}.$$

Let $v' = \lambda w$ then $(A+bh^t)v' = \lambda(Aw+bh^tw) = \lambda b(1+h^tw) = b$. So we may conclude $v' = v$, as desired. ■

Proof of Theorem 3. Given Σ and H as in the statement, let:

$$A' = A + BH,$$

and Σ' be the system whose dynamics is given by:

$$x(t+1) = \vec{\sigma}(A'x(t) + Bv(t)).$$

Notice that we can transform Σ into Σ' (and viceversa) by using the feedback $u(t) = Hx(t)+v(t)$ (and $v(t) = u(t)-Hx(t)$). Thus Σ is forward accessible if and only if Σ' is forward accessible. So, without loss of generality, we may assume that the matrix A is itself invertible and that the matrix $A^{-1}B$ has all nonzero rows.

It is easy to see that if the pair A, B satisfies these conditions then we can find a vector $\bar{u} \in \mathbb{R}^m$ satisfying:

- $(B\bar{u})_i \neq 0$ for all $i = 1, \dots, n$;
- $[(A^{-1}B)\bar{u}]_i \neq 0$ for all $i = 1, \dots, n$.

So, by taking controls of the form $u = \alpha\bar{u}$ with $\alpha \in \mathbb{R}$, it is enough to prove the statement for the single-input case, assuming the matrix A invertible, and the vectors b and $A^{-1}b$ with all nonzero components.

Let x_0 be any initial state and $u(0), \dots, u(n-1)$ a control sequence. Using notations similar to those of lemma II.3, we let:

$$\begin{aligned} x(0) &= x_0, \quad x(i) = \vec{\sigma}(Ax(i-1) + bu(i-1)), \ i = 1, \dots, n; \\ D(i) &= \hat{\sigma}(Ax(i-1) + bu(i-1)), \ i = 1, \dots, n; \\ g_k(i) &= AD(n-k) \cdots AD(i)b, \quad \text{for} \ \ k = 2, \dots, n-1, \\ & \qquad\qquad\qquad\qquad\qquad \text{and} \ \ i = 1, \dots, n-k. \end{aligned}$$

Moreover we let:

$$\begin{aligned} H(2) &= [AD(1)b, b], \\ H(k) &= [AD(k-1)g_{n-k+2}(1), \cdots, \\ & \qquad AD(k-1)g_{n-k+2}(k-2), AD(k-1)b, b], \end{aligned}$$

for $k = 3, \dots, n$. Notice that $H(k)$ is a matrix in $\mathbb{R}^{n \times k}$. If W_n denotes the Jacobian matrix $\partial_{u(0),\dots,u(n-1)}x(n)$, then lemma II.3 says that:

$$W_n = D(n)H(n).$$

Claim 2: For $k = 3, \dots, n$, if $\operatorname{rank} H(k) < k$ for every $u(0), \dots, u(k-2)$, then $\operatorname{rank} H(k-1) < k-1$ for every $u(0), \dots, u(k-3)$.

Before proving the claim we see how our Theorem follows from this claim.

Assume, by the way of contradiction, that Σ is not forward accessible. Then, since $D(n)$ is invertible, we must have $\operatorname{rank} H(n) < n$ for all $u(0), \dots, u(n-2)$. Thus, by applying recursively the previous claim, we get $\operatorname{rank} H(2) < 2$ for all $u(0)$. Since A and $D(1)$ are invertible matrices, this rank condition is equivalent to:

$$\operatorname{rank}[b, D(1)^{-1}A^{-1}b] < 2, \quad \forall\, u(0).$$

Denoting by $v = A^{-1}b$, the previous condition means that:

$$b_j \frac{1}{\sigma'[(Ax(0))_i + b_i u(0)]} v_i - b_i \frac{1}{\sigma'[(Ax(0))_j + b_j u(0)]} v_j = 0,$$

for all $u(0)$ and all $1 \leq i, j \leq n$. Since all components of the vectors b and v are nonzero by assumption, the previous equation (when $i \neq j$) contradicts the fact that $1/\sigma'$ is n-**WIP** ; so we may conclude that Σ is forward accessible.

Now we give the proof of the previous claim. If $\operatorname{rank} H(k) < k$ for all $u(0), \dots, u(k-2)$, then:

$$\operatorname{rank}[g_{n-k+2}(1), \cdots, g_{n-k+2}(k-2), b, D(k-1)^{-1}A^{-1}b] < k$$

for all $u(0), \dots, u(k-2)$. Letting $v(i) = g_{n-k+2}(i)$, for $i = 1, \dots, k-2$, $v(k-1) = b$, and $v(k) = A^{-1}b$, the previous equation becomes:

$$\operatorname{rank}[v(1), \dots, v(k-1), D(k-1)^{-1}v(k)] < k,$$

for all $u(0), \dots, u(k-2)$. Notice that $[D(k-1)^{-1}v(k)]_i = \frac{1}{\sigma'([Ax(k-2)]_i + b_i u(k-2))} v(k)_i$. Fix a sequence $u(0), \dots, u(k-3)$, and choose $i_1 < \dots < i_k$ indices in $1, \dots, n$. Then we have that the determinant of the submatrix of $H(k)$ corresponding to these rows has to be identically zero for all $u(k-2)$. By expanding the determinant with respect to the last column, and denoting by M_{i_j} the minor corresponding to the i_j-row, we have:

$$(-1)^k \sum_{j=1}^{k} M_{i_j} v(k)_{i_j} \frac{1}{\sigma'([Ax(k-2)]_{i_j} + b_{i_j} u(k-2))} = 0$$

for every $u(k-2) \in \mathbb{R}$. This equation together with the facts that the map $1/\sigma'$ is n-**WIP** , $v(k)_i \neq 0$ for all $i = 1, \dots, n$, and M_{i_j}, $x(k-2)$ are independent on $u(k-2)$, gives us that:

$$M_{i_j} = 0 \quad \forall\, i_j.$$

So we may conclude that:

$$\operatorname{rank}[v(1), \dots, v(k-1)] < k-1 \quad \forall\, u(0), \dots, u(k-3). \tag{13}$$

Equation (13) proves our claim, since

$$v(i) = AD(k-2)g_{n-k+3}(i), \quad \text{for} \quad i = 1, \dots, k-3,$$

$v(k-2) = AD(k-2)b$, and $v(k-1) = b$, thus $H(k-1) = [v(1), \dots, v(k-1)]$. ∎

III. Forward accessibility for continuous-time models.

In this Section we discuss forward accessibility for recurrent neural networks evolving in continuous-time, i.e. systems evolving according to the equation:

$$\begin{aligned} \dot{x}(t) &= \vec{\sigma}(Ax(t) + Bu(t)) \\ x(0) &= x_0 \end{aligned} \tag{14}$$

where A, B and $\vec{\sigma}$ are as in (1). We will show that such models are forward accessible as soon as σ and B satisfy the weaker condition **IP** (see section II-C), with no restriction on the pair A, B.

In what follows we let X_u be the vector field defined by

$$X_u(x) := \vec{\sigma}(Ax + Bu).$$

Lemma III.1: Let $x_0 \in \mathbb{R}^n$ and suppose σ and B satisfy the **IP** . Then there are controls $u_1, \cdots, u_n$ such that the vectors $X_{u_1}(x_0), \cdots, X_{u_n}(x_0)$ are linearly independent.

Proof: Denote by $v := Ax_0$, by b_i the i-th row of the matrix B, and let $M(u_1, \dots, u_n)$ be the matrix whose columns are the vectors $X_{u_i}(x_0) = \vec{\sigma}(v + Bu_i)$. Saying that there exist $u_1, \dots, u_n$ such that the vectors X_{u_i}, $i = 1, \dots, n$, are linearly independent, it is equivalent to say that there exist $u_1, \dots, u_n$ such that $\det M(u_1, \dots, u_n) \neq 0$.

We now give the proof of the lemma by induction on n. The case $n = 1$ is obvious; let $n > 1$, then we have:

$$\det M(u_1, \ldots, u_n) =$$

$$= \sum_{\pi \in S_n} \text{sgn } (\pi) \sigma(v_{\pi(1)} + b_{\pi(1)} u_1) \cdots \sigma(v_{\pi(n)} + b_{\pi(n)} u_n)$$

$$\sum_{i=1}^{n} \Big(\sum_{\pi \in S_n, \pi(n)=i} \text{sgn } (\pi) \sigma(v_{\pi(1)} + b_{\pi(1)} u_1) \cdots$$

$$\cdots \sigma(v_{\pi(n-1)} + b_{\pi(n-1)} u_{n-1}) \Big) \sigma(v_i + b_i u_n).$$

If $\det M(u_1, \ldots, u_n) = 0$ for every $u_1, \ldots, u_n$, by the **IP** we must have

$$0 = \sum_{\pi \in S_n, \pi(n)=i} \text{sgn } (\pi) \sigma(v_{\pi(1)} + b_{\pi(1)} u_1) \cdots$$

$$\cdots \sigma(v_{\pi(n-1)} + b_{\pi(n-1)} u_{n-1}) =$$

$$= (-1)^i \sum_{\tau \in S_{n-1}} \text{sgn } (\tau) \sigma(w_{\tau(1)} + \tilde{b}_{\tau(1)} u_1) \cdots$$

$$\cdots \sigma(w_{\tau(n-1)} + \tilde{b}_{\tau(n-1)} u_{n-1}),$$

where

$$w_j = \begin{cases} v_j & \text{if } j < i \\ v_{j+1} & \text{if } j \geq i \end{cases} \quad \tilde{b}_j = \begin{cases} b_j & \text{if } j < i \\ b_{j+1} & \text{if } j \geq i \end{cases}$$

By inductive assumption we get that this last expression can not be zero for all $u_1 \ldots, u_{n-1}$, and so also $\det M(u_1, \ldots, u_n)$ can not be identically zero. ■

Given the vector fields X_u, it is known that if Lie $\{X_u \mid u \in \mathbb{R}^m\}$ has full rank at x then the system is forward accessible from x (see [7]). This result together with the previous lemma, gives the following:

Theorem 4: Let Σ be a model of type (14), assume that σ and B satisfy the **IP** . Then Σ is forward accessible.

Remark III.2: From the previous theorem, together with Theorem 1 in [2], we have that a system of type (14), with a linear output equation $y(t) = Cx(t)$, and satisfying the **IP** , if it is observable is both minimal and forward accessible.

References

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

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

[3] Albertini, F., and E.D. Sontag, "State observability in recurrent neural networks," *System and Control Letters*, 22, 4(1994): 235-244.

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

[5] Matthews, M., "A state-space approach to adaptive nonlinear filtering using recurrent neural networks," *Proc. 1990 IASTED Symp. on Artificial Intelligence Applications and Neural Networks*, Zurig, July 1990, pp:197-200.

[6] Polycarpou, M.M., and P.A. Ioannou, "Neural networks and on-line approximators for adaptive control ," *Proc. Seventh Yale Workshop on Adaptive and Learning Systems*, Yale University, 1992.

[7] Sussmann, H.J., "Orbits of families of vector fields and integrability of distributions" *Trans. A.M.S.* 180 (1973):171-188.

[8] Sussmann, H.J., "Existence and Uniqueness of Minimal Realization of Nonlinear Systems" *Mathematical Systems Theory*, 10, 1977:263-284.