Complete Controllability of Continuous-Time Recurrent Neural Networks*

Eduardo Sontag
Dept. of Mathematics, Rutgers University
New Brunswick, NJ 08903
sontag@control.rutgers.edu

Héctor Sussmann
Dept. of Mathematics, Rutgers University
New Brunswick, NJ 08903
sussmann@hamilton.rutgers.edu

Abstract

This paper presents a characterization of controllability for the class of control systems commonly called (continuous-time) recurrent neural networks. The characterization involves a simple condition on the input matrix, and is proved when the activation function is the hyperbolic tangent.

1 Introduction

This paper continues the study of system-theoretic properties of recurrent networks. Assume given a locally Lipschitz map $\sigma : \mathbb{R} \to \mathbb{R}$. By an n-dimensional, m-input *(recurrent) σ-net* we mean a continuous-time control system of the form

$$\dot{x}(t) \;=\; \vec{\sigma}^{(n)}\left(Ax(t) + Bu(t)\right) , \tag{1}$$

where $A \in \mathbb{R}^{n\times n}$ and $B \in \mathbb{R}^{n\times m}$. Here, for each map $\sigma : \mathbb{R} \to \mathbb{R}$ and each positive integer n, we use $\vec{\sigma}^{(n)}$ to denote the diagonal mapping

$$\vec{\sigma}^{(n)} \,:\, \mathbb{R}^n \to \mathbb{R}^n \,:\, \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \mapsto \begin{pmatrix} \sigma(x_1) \\ \vdots \\ \sigma(x_n) \end{pmatrix} . \tag{2}$$

(Sometimes one includes, in addition, an observation or measurement function $y = Cx$, but this paper will not deal with observation issues.) The spaces $\mathbb{R}^m$ and $\mathbb{R}^n$ are called respectively the input-value space and the state space of the net. Observe that the linear systems customarily studied in control theory are precisely the σ-nets for which σ is the identity function. Our main result will be for the special case of the hyperbolic tangent $\sigma = \tanh$, the "sigmoid" function used in neural nets theoretical as well as experimental work.

*Supported in part by US Air Force Grant AFOSR-94-0293

In the area of neural networks, one interprets the vector equations for x in (1) as representing the evolution of an ensemble of n "neurons," where each coordinate x_i of x is a real-valued variable which represents the internal state of the ith neuron, and each coordinate $u_i, i = 1, \ldots, m$ of u is an external input signal. The coefficients A_{ij}, B_{ij} denote the weights, intensities, or "synaptic strengths," of the various connections. The transformation $\sigma : \mathbb{R} \to \mathbb{R}$ is called the "activation function". Systems of this type have been employed in areas as varied as digital signal processing (see for instance [6], [5], [10]), control (see e.g. [14], [18], [11], [17]), the design of associative memories ("Hopfield nets"), language inference, and sequence extrapolation for time series prediction. Special purpose chips are being built to implement recurrent nets directly in hardware; for instance, Hitachi's Wafer Scale Integration chips have been designed to implement Hopfield nets with over 500 neurons and 30,000 synaptic connections.

In past work we have studied, for such models, questions of parameter identifiability ([3]), observability ([4]), system approximation ([13]), computability ([12]), parameter reconstruction ([9]), and sample complexity for learning and generalization ([7], [8]).

Here we focus on problems of controllability. The fundamental contribution in this area was a recent paper by Albertini and Dai Pra, cf. [1]. This paper dealt with the study of the *forward accessibility* property. Recall that a system such as (1) is said to be forward accessible if from each initial state it is possible to reach, by using appropriate inputs $u(\cdot)$, an open subset of the state space $\mathbb{R}^n$. Albertini and Dai Pra showed that forward accessibility holds provided that

- the "independence property" (cf. [16], [2]) holds for σ, and
- B is in a certain class $\mathbf{B}_{n,m}$ of matrices which was introduced in [2] and [4] and is reviewed below.

(For the special but most important case of single input systems, $m = 1$, the condition $B \in \mathbf{B}_{n,m}$ means that the entries of the vector B are all nonzero and have different absolute values. The independence property asserts that distinct dilations and translates of σ must be linearly independent.)

An extremely surprising aspect of this result is that for matrices $B \in \mathbf{B}_{n,m}$, accessibility holds independently of the choice of the matrix A. This is in sharp contrast with linear systems, for which the only B's so that the system $\dot{x} = Ax + Bu$ is accessible no matter what A is are those B's of full rank n. Thus we were motivated to ask if the same condition $B \in \mathbf{B}_{n,m}$ which Albertini and Dai Pra used to guarantee forward accessibility also ensures *controllability*. Controllability means that from each initial state it is possible to reach, by using appropriate inputs $u(\cdot)$, the entire state space, not just some — potentially very small — open subset. Naturally, this is a much more interesting property. Surprisingly, the answer turns out to be yes, for the standard sigmoidal function $\sigma = \tanh$ which is ubiquitous in neural network practice. Adding to the unexpected developments, it turns out that the use of tanh is essential: there are other functions σ for which the independence property holds, so that in particular a σ-net is accessible for every $B \in \mathbf{B}_{n,m}$, but controllability fails for some $B \in \mathbf{B}_{n,m}$. We provide a counterexample with $\sigma = \arctan$.

1.1 Definitions and Statements of the Main Results

For any measurable (essentially) bounded control $u : [0, T] \to \mathbb{R}^m$ and any state ξ we use $\phi(t, \xi, u)$ to denote the solution $x(t)$ of (1) having initial condition $x(0) = \xi$. The function $x(\cdot)$ is defined on some maximal subinterval of $[0, T]$, and if σ is globally Lipschitz, which is the case with our main example given below, then it is defined on the entire interval. Given $\xi, \zeta \in \mathbb{R}^n$, we say that ξ can be steered, or controlled, to ζ if there is some $T \geq 0$ and some control u on $[0, T]$ such that the solution is defined for all $t \in [0, T]$ and $\phi(T, \xi, u) = \zeta$. The system (1) is *controllable* if every $\xi \in \mathbb{R}^n$ can be steered to every $\zeta \in \mathbb{R}^n$. (See [15] for generalities and basic facts about control systems.)

For each pair of positive integers n and m, we let

$$\mathbf{B}_{n,m} := \{B \in \mathbb{R}^{n\times m}, (\forall\, i) \ \mathrm{row}_i(B) \neq 0 \ \text{ and } \ (\forall\, i \neq j) \ \mathrm{row}_i(B) \neq \pm \mathrm{row}_j(B)\}$$

where $\mathrm{row}_i(\cdot)$ denotes the ith row of the given matrix. In the special case $m = 1$, a vector $b \in \mathbf{B}_{n,1}$ if and only if all its entries are nonzero and have different absolute values. The complement of $\mathbf{B}_{n,m}$ is an algebraic subset of $\mathbb{R}^{n\times m}$, so $\mathbf{B}_{n,m}$ is generic in every possible sense (full measure, open dense).

We let Σ be the class of maps $\sigma : \mathbb{R} \to \mathbb{R}$ which are locally Lipschitz and have the following properties:

1. σ is an odd function, i.e. $\sigma(-r) = -\sigma(r)$ for all $r \in \mathbb{R}$;
2. $\sigma_\infty = \lim_{s\to+\infty} \sigma(s)$ exists and is > 0;
3. $\sigma(r) < \sigma_\infty$ for all $r \in \mathbb{R}$;
4. for each $a, b \in \mathbb{R}$, $b > 1$,
$$\lim_{s\to+\infty} \frac{\sigma_\infty - \sigma(a + bs)}{\sigma_\infty - \sigma(s)} = 0\,. \tag{3}$$

Remark 1.1 The requirement that σ be odd is merely imposed for convenience, and can be weakened in many of the results. The most critical assumption on elements of the class Σ is the last one. This is a nontrivial requirement; note for instance that the function $\sigma = \arctan$ does *not* satisfy it, since the limit is in that case $1/b$. □

We now list our main results; proofs are given in later sections.

Lemma 1.2 The function $\tanh \in \Sigma$.

Theorem 1 *Assume that $\sigma \in \Sigma$, $B \in \mathbf{B}_{n,m}$, and A is arbitrary. Then the system (1) is controllable.*

A trivial converse of Theorem 1 is as follows:

Lemma 1.3 If $B \in \mathbb{R}^{n\times m}$ and the odd function $\sigma : \mathbb{R} \to \mathbb{R}$ are such that for all $A \in \mathbb{R}^{n\times n}$ the system (1) is controllable, then $B \in \mathbf{B}_{n,m}$.

However, we give also the following example to show that when $B \notin \mathbf{B}_{n,m}$, it may still be the case that the system (1) is controllable, provided A is appropriately chosen.

Lemma 1.4 The following system in dimension two:

$$\begin{aligned} \dot{x} &= \sigma(y) \\ \dot{y} &= \sigma(u) \end{aligned}$$

where $\sigma(s) = \tanh s$, is controllable.

Observe that $B = \mathrm{col}(0,1) \notin \mathbf{B}_{n,1}$, and

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$

but with $A = 0$ the system would not be controllable.

Finally, one may ask what happens if $\sigma \notin \Sigma$ but σ is still a "sigmoidal" type function, with a graph qualitatively similar to that of tanh. Specifically, one may consider the nonlinearity $\sigma = \arctan$, which has also appeared, albeit much less often, in the neural network literature. We have the following counterexample.

Proposition 1.5 Let $\sigma = \arctan$. Then the 4-dimensional, single-input system

$$\begin{aligned} \dot{x}_1 &= \sigma(x_1 + x_2 + x_3 + x_4 + 2u) \\ \dot{x}_2 &= \sigma(x_1 + x_2 + x_3 + x_4 + 12u) \\ \dot{x}_3 &= \sigma(-3u) \\ \dot{x}_4 &= \sigma(-4u) \end{aligned}$$

is not controllable.

Observe that this system is forward accessible, because it satisfies the conditions in [1]. Indeed, the nonlinearity $\sigma = \arctan$ satisfies the "independence property" (cf. [2]), and clearly the matrix $B = \mathrm{col}(2, 12, -3, -4)$ belongs to $\mathbf{B}_{n,1}$.

2 Proof of the Main Results

We first prove Lemma 1.2.

Proof. The first three properties are clear, with $\sigma_\infty = 1$, so we need to prove the limit property (3). Note that $\theta(x) = (1 + e^{-x})^{-1}$ satisfies

$$\frac{\theta(r)}{\theta(t)} = \theta(r) + e^{r-t}\theta(-r) \tag{4}$$

for all r, t, and that $1 - \tanh x = 2\theta(-2x)$ for all $x \in \mathbb{R}$. Thus

$$\frac{1 - \tanh(a + bs)}{1 - \tanh s} = \underbrace{\theta(-2a - 2bs)}_{\to 0} + e^{-2a} \underbrace{e^{2(1-b)s}}_{\to 0} \underbrace{\theta(2a + 2bs)}_{\to 1} \to 0$$

as desired. ∎

2.1 A Result on Convex Hulls

Let σ be a map $\mathbb{R} \to \mathbb{R}$. For each vector $\mathbf{a} \in \mathbb{R}^n$ and matrix $B \in \mathbb{R}^{n \times m}$, we write

$$S_{\mathbf{a},B} := \left\{ \vec{\sigma}^{(n)}(\mathbf{a} + Bu)\,,\; u \in \mathbb{R}^m \right\}.$$

We use $\mathrm{int}(S)$ and $\mathrm{co}(S)$ to denote, respectively, the interior and the convex hull of a set S.

Lemma 2.1 Pick $\sigma \in \Sigma$, $B \in \mathbf{B}_{n,m}$, and arbitrary $\mathbf{a} \in \mathbb{R}^n$. Then $0 \in \operatorname{int}(\operatorname{co}(S_{\mathbf{a},B}))$.

Proof. Since $B \in \mathbf{B}_{n,m}$, there is some $\mathbf{u} \in \mathbb{R}^m$ such that the numbers $b_i := \mathrm{row}_i(B)\mathbf{u}$ are all nonzero and have distinct absolute values. (Because the set of u's that satisfy at least one of the equations $\mathrm{row}_i(B)u = 0$, $\mathrm{row}_i(B)u + \mathrm{row}_j(B)u = 0$, or $\mathrm{row}_i(B)u - \mathrm{row}_j(B)u = 0$, is a finite union of hyperplanes in $\mathbb{R}^m$.) As $S_{\mathbf{a},B\mathbf{u}} \subseteq S_{\mathbf{a},B}$, it is enough to show the result for $B\mathbf{u}$ instead of B. So we assume from now on that $B = \mathrm{col}(b_1, \ldots, b_n)$ and the b_i are all nonzero and have distinct absolute values.

Assume by way of contradiction that $0 \notin \operatorname{int}(\operatorname{co}(S_{\mathbf{a},B}))$. Using a separating hyperplane, we know that there is a nonzero vector $c = (c_1, \ldots, c_n)$ such that $c\vec{\sigma}^{(n)}(\mathbf{a} + Bu) \geq 0$ for all $u \in \mathbb{R}$. Writing $\mathbf{a} = \mathrm{col}(a_1, \ldots, a_n)$, this means that

$$\sum_{i=1}^{n} c_i \sigma(a_i + b_i u) \;\geq\; 0 \quad \forall\, u \in \mathbb{R}\,. \tag{5}$$

We now prove that such an inequality cannot hold, if $\sigma \in \Sigma$ and the b_i are nonzero and have distinct absolute values, unless all the c_i are equal to 0. Since σ is odd, we may assume that each $b_i > 0$, since any term $c_i\sigma(a_i + b_i u)$ with $b_i < 0$ can be rewritten as $(-c_i)\sigma(-a_i + (-b_i)u)$. Thus, reordering if needed, we assume that $0 < b_1 < \ldots < b_n$. Finally, dropping all those terms in the sum for which $c_i = 0$, we may assume that $c_1 \neq 0$. Taking the limit in (5) as $u \to -\infty$ we obtain $\sum_{i=1}^{n} c_i(-\sigma_\infty) \geq 0$. So may rewrite (5) as

$$\sum_{i=1}^{n} c_i \left(\sigma_\infty - \sigma(a_i + b_i u)\right) \;\leq\; 0 \quad \forall\, u \in \mathbb{R}\,. \tag{6}$$

Therefore

$$c_1 \;+\; \sum_{i=2}^{n} c_i \frac{\sigma_\infty - \sigma(a_i + b_i u)}{\sigma_\infty - \sigma(a_1 + b_1 u)} \;\leq\; 0 \quad \forall\, u \in \mathbb{R}\,. \tag{7}$$

If we prove that each term in the sum converges to zero as $u \to +\infty$ then it will follow that $c_1 \leq 0$. But this fact follows from property (3) (applied with $a = a_i - b_i a_1/b_1$, $b = b_i/b_1$, and noting that $s = a_1 + b_1 u \to \infty$ as $u \to \infty$).

If we take instead the limit in (5) as $u \to +\infty$, we find that $\sum_{i=1}^{n} c_i\sigma_\infty \geq 0$. We may therefore also rewrite (5) in the form:

$$\sum_{i=1}^{n} c_i \left(\sigma_\infty + \sigma(a_i + b_i u)\right) \;\geq\; 0 \quad \forall\, u \in \mathbb{R}\,. \tag{8}$$

Letting $v = -u$ and $a_i' = -a_i$, and using that σ is odd,

$$c_1 \;+\; \sum_{i=2}^{n} c_i \frac{\sigma_\infty - \sigma(a_i' + b_i v)}{\sigma_\infty - \sigma(a_1' + b_1 v)} \;\geq\; 0 \quad \forall\, v \in \mathbb{R}\,. \tag{9}$$

Taking the limit as $v \to +\infty$ and appealing again to property (3), we conclude that also $c_1 \geq 0$. Thus $c_1 = 0$, contradicting the assumption made earlier. ∎

2.2 A Local Controllability Lemma

Lemma 2.2 Let Ω be an open subset of $\mathbb{R}^n$, let U be a finite set, and let $f : \Omega \times U \mapsto \mathbb{R}^n$ be a map such that $f(x, u)$ is continuous in x for each $u \in U$. Let Σ be the control system $\dot{x} = f(x, u)$. Let $x_0 \in \Omega$ be a point such that 0 is an interior point of the convex hull of the set of vectors $\{f(x_0, u) : u \in U\}$. Then there exists a neighborhood W of x_0, contained in Ω, such that for every $x \in W$ there are trajectories of Σ going from x to x_0 and from x_0 to x.

Proof. We may obviously assume, without loss of generality, that $x_0 = 0$. Let

$$S = \{v \in \mathbb{R}^n : \|v\| = 1\}\,.$$

Define a function $\varphi : \Omega \mapsto \mathbb{R}$ by letting

$$\varphi(x) = \max\{\psi(v, x) : v \in S\}\,,$$

where

$$\psi(v, x) = \min\{\langle v, f(x, u)\rangle : u \in U\}\,.$$

The function ψ is a minimum of a finite collection of continuous functions, so ψ is continuous. Then φ is well-defined and continuous as well.

Since 0 is an interior point of the convex hull of the set $\{f(0, u) : u \in U\}$, we can find a $\delta > 0$ such that for every $v \in S$ the vector $-4\delta v$ is a convex combination of the $f(0, u)$, $u \in U$. Given any $v \in S$, the number -4δ is equal to $\langle v, -4\delta v\rangle$, which is a convex combination of the numbers $\langle v, f(0, u)\rangle$, $u \in U$. So at least one of these numbers is $\leq -4\delta$. So $\psi(v, 0) \leq -4\delta$. Since this is true for all $v \in S$, we conclude that $\varphi(0) \leq -4\delta$. Since φ is continuous, there exists $\alpha > 0$ such that, if $B = \{x \in \mathbb{R}^n : \|x\| \leq \alpha\}$, then $B \subseteq \Omega$ and $\varphi(x) \leq -2\delta$ whenever $x \in B$.

Now fix a point $\bar{x} \in B$, and let $\mathcal{T}$ be the set of all triples (I, η, ξ) such that (i) I is an interval, (ii) $I \subseteq [0, \infty[$, (iii) $0 \in I$, (iv) $\eta : I \to U$ is measurable, (v) $\xi : I \to \Omega$ is a trajectory of Σ corresponding to the control η (that is, ξ is absolutely continuous on $[0, T]$ for every $T \in I$, and $\dot{\xi}(t) = f(\xi(t), \eta(t))$ for almost all $t \in I$), (vi) $\xi(0) = \bar{x}$, and (vii) the derivative of the function $t \mapsto \|\xi(t)\|$ is $\leq -\delta$ for almost every $t \in I$.

If $(I, \eta, \xi) \in \mathcal{T}$, then ξ is locally absolutely continuous on I, by the definition of trajectory, so the map $t \mapsto \|\xi(t)\|$ is also locally absolutely continuous. Then condition (vii) implies that the inequality $\|\xi(t)\| \leq \|\xi(0)\| - \delta t$ holds for all $t \in I$. Therefore $\|\xi(t)\| \leq \alpha$ for all $t \in I$, so ξ is entirely contained in B. It follows that ξ is Lipschitz (with Lipschitz constant $C = \max\{C(u) : u \in U\}$, where $C(u) = \sup\{\|f(x, u)\| : x \in B\}$). Moreover, the interval I must be bounded, since $t \in I$ implies $t \leq \frac{\alpha}{\delta}$. Therefore, if I is not compact then the map ξ extends to the closure $\bar{I}$ of I, which is compact. It is clear that the extension is also a trajectory of Σ.

Order $\mathcal{T}$ in the obvious way, by letting $(I_1, \eta_1, \xi_1) \preceq (I_2, \eta_2, \xi_2)$ iff $I_1 \subseteq I_2$ and η_1, ξ_1 are the restrictions to I_1 of η_2, ξ_2. It is then obvious that every totally ordered nonempty subset of $\mathcal{T}$ has an upper bound in $\mathcal{T}$. So $\mathcal{T}$ has a maximal element (I, η, ξ) by Zorn's lemma. In view of the Lipschitz property discussed in the previous paragraph, $I = [0, T]$ for some T.

We now show that $\xi(T) = 0$. Suppose this is not true. Let $\xi(T) = \hat{x}$, so $\hat{x} \in B$ and $\|\hat{x}\| > 0$. Since $\varphi(\hat{x}) \leq -2\delta$, we have $\psi(\hat{v}, \hat{x}) \leq -2\delta$, where $\hat{v} = -\frac{\hat{x}}{\|\hat{x}\|}$. So there exists $u \in U$ such that $\langle \hat{v}, f(\hat{x}, u)\rangle \leq -2\delta$. Let $\xi' : [T, T + \beta] \to \Omega$ be an integral curve of the vector field $x \mapsto f(x, u)$, such that $\xi'(T) = \hat{x}$. (Such a curve exists, for some $\beta > 0$, by the existence theorem for ordinary differential equations with a continuous right-hand side.) The function $x \mapsto \|x\|$ is smooth on a neighborhood of $\hat{x}$, and its derivative along the curve ξ' is given by

$$\rho(t) = \frac{d}{dt}\Big(\|\xi'(t)\| \Big) = \langle v(t), f(\xi(t), u)\rangle ,$$

where

$$v(t) = \frac{\xi'(t)}{\|\xi'(t)\|} .$$

(Here we have used the chain rule, together with the facts that the gradient of $x \mapsto \|x\|$ is $\frac{x}{\|x\|}$ and the derivative of $t \mapsto \xi'(t)$ is $t \mapsto f(\xi'(t), u)$.) It is clear that $v(t)$ is continuous as a function of t, and $v(T) = \hat{v}$. Since $\rho(T) = \langle \hat{v}, f(\hat{x}, u)\rangle \leq -2\delta$, we can assume, by taking β small enough, that $\rho(t) \leq -\delta$ for $t \in [T, T+\beta]$. So, if we let $\tilde{I} = [0, T+\beta]$, we can extend η to a control $\tilde{\eta} : \tilde{I} \to U$ by letting $\tilde{\eta}(t) = u$ for $t \in]T, T+\beta]$, and define $\tilde{\xi} : \tilde{I} \to \Omega$ to be the curve whose restrictions to $[0,T]$ and $[T, T+\beta]$ are ξ, ξ'. It is clear that the triple $(\tilde{I}, \tilde{\eta}, \tilde{\xi})$ is in $\mathcal{T}$, and $(I, \eta, \xi) \prec (\tilde{I}, \tilde{\eta}, \tilde{\xi})$ but $(I, \eta, \xi) \neq (\tilde{I}, \tilde{\eta}, \tilde{\xi})$. This contradicts the maximality of (I, η, ξ). This contradiction proves that $\xi(T) = 0$, as stated.

We have thus shown that for an arbitrary point $\bar{x} \in B$ there is a trajectory of Σ that goes from $\hat{x}$ to 0. We may also apply this argument to the reversed system $\dot{x} = -f(x, u)$, whose trajectories are those of Σ run backwards in time. Indeed, if the convex hull of the set $\{f(x_0, u) : u \in U\}$ contains zero in its interior, then the same is true for the convex hull of $\{-f(x_0, u) : u \in U\}$. Thus, we find some other ball B' with the property that every $x \in B'$ can be reached from x_0 by a trajectory of Σ. Then the neighborhood $W = B \cap B'$ has the desired property. ∎

2.3 Proof of Theorem 1

We first show that if $x_0 \in \mathbb{R}^n$ then the set $\mathcal{R}(x_0)$ of points reachable from x_0 is open. To see this, let $\bar{x} \in \mathcal{R}(x_0)$. By Lemma 2.1, with $\mathbf{a} = A\bar{x}$, there exists a finite subset U of $\mathbb{R}^n$ such that 0 is an interior point of the convex hull of the set of vectors $\{\vec{\sigma}^{(n)}(A\bar{x} + Bu)\, , \, u \in U\}$. If we define $f(x, u) = \vec{\sigma}^{(n)}(Ax + Bu)$ for $u \in U$, Lemma 2.2 tells us that there is a neighborhood W of $\bar{x}$ all whose points are reachable from $\bar{x}$ by trajectories of $\dot{x} = f(x, u)$. Clearly, every trajectory of $\dot{x} = f(x, u)$ is a trajectory of the system (1). So $W \subseteq \mathcal{R}(x_0)$. This proves that $\mathcal{R}(x_0)$ is open.

We now prove that $\mathcal{R}(x_0)$ is closed. To show this, pick $\bar{x}$ in the closure of $\mathcal{R}(x_0)$. Applying Lemmas 2.1 and 2.2, we find a neighborhood W of $\bar{x}$ such that every $x \in W$ can be steered to $\bar{x}$ by a trajectory of (1). Since $W \cap \mathcal{R}(x_0) \neq \emptyset$, we can find a point $\hat{x}$ which is reachable from x_0 and can be steered to $\bar{x}$. Therefore $\bar{x} \in \mathcal{R}(x_0)$. So $\mathcal{R}(x_0)$ is closed, as stated.

Since $\mathbb{R}^n$ is connected, and $\mathcal{R}(x_0)$ is open, closed, and nonempty (because $x_0 \in \mathcal{R}(x_0)$), we conclude that $\mathcal{R}(x_0) = \mathbb{R}^n$. Since x_0 is an arbitrary point of $\mathbb{R}^n$, the system (1) is controllable, and our proof is complete. ∎

2.4 Proof of Lemma 1.3

Take any $B \in \mathbb{R}^{n\times m}$ and any odd function $\sigma : \mathbb{R} \to \mathbb{R}$. Assume that $B \notin \mathbf{B}_{n,m}$. We show that the system (1) is not controllable when $A = 0$. Note that the system equations are now $\dot{x} = \vec{\sigma}^{(n)}(Bu)$. Since $B \notin \mathbf{B}_{n,m}$, either one of its rows is zero or some two rows are equal up to a sign change. Assume first that the ith row is zero. Then the equation for the ith coordinate of x is $\dot{x}_i = 0$; this implies that $x_i(t)$ is constant along all trajectories and hence the system cannot be controllable. Assume instead that $\mathrm{row}_i(B) = \mathrm{row}_j(B)$ for some $i \neq j$. In that case, $\dot{x}_i - \dot{x}_j = \sigma(\mathrm{row}_i(B)u) - \sigma(\mathrm{row}_j(B)u) = 0$, no matter what the control $u(\cdot)$ is, so the quantity $x_i(t) - x_j(t)$ is constant along trajectories, again contradicting controllability. Finally, if $\mathrm{row}_i(B) = -\mathrm{row}_j(B)$, then since σ is odd we have that $x_i(t) + x_j(t)$ is constant along trajectories, and once more the system cannot be controllable. ∎

3 Proof of Lemma 1.4

Here we prove Lemma 1.4, which is an example to illustrate that the controllability of system (1) depends on the form of A, when $B \notin \mathbf{B}_{n,m}$. We will prove controllability for more general σ than the specific example *tanh* of interest. The assumptions made here (much more than necessary, but enough to make the proof almost trivial) are as follows: $\sigma : \mathbb{R} \to (-1, 1)$ is locally Lipschitz, odd, strictly increasing, onto, and there exists $\sigma'(0) = c \neq 0$.

Proof. There is a neighborhood $\mathcal{U}$ of $(x, y) = (0, 0)$ such that each pair of states in $\mathcal{U}$ can be steered to each other (local controllability about $x = y = 0$, because the linearized system is $\dot{x} = cy, \dot{y} = cu$).

We claim that for each $\xi \in \mathbb{R}^2$ there is some $\zeta \in \mathcal{U}$ which ξ can be steered into. Once this claim is proved, controllability follows. Indeed, assume given any two states ξ_1 and ξ_2. Then there is a control u_1 steering ξ_1 to some $\zeta_1 \in \mathcal{U}$, by the claimed property. Similarly, there is a control u_3 which steers some $\zeta_2 \in \mathcal{U}$ to ξ_2. (Proof: we must show that ξ_2 can be controlled to some $\zeta_2 \in \mathcal{U}$ with respect now to the time-reversed system $\dot{x} = -\sigma(y), \dot{y} = -\sigma(u)$, cf. Lemma 2.6.8 in [15]. With the new variable $z := -y$, the equations become $\dot{x} = \sigma(z), \dot{z} = \sigma(u)$, which coincide with those of the original system.) Finally, there is a control u_2 taking ζ_1 to ζ_2. Then the concatenation of u_1, u_2, u_3 steers ξ_1 to ξ_2.

To prove the claim, it is sufficient to exhibit a continuous function $k : \mathbb{R}^2 \to \mathbb{R}$ with the property that every trajectory of $\dot{x} = y$, $\dot{y} = \sigma(k(x, y))$ is defined on $[0, \infty)$ and converges to zero (k is a continuous feedback stabilizer). We take $k(x, y) := \sigma^{-1}(-\sigma(x)/2 - \sigma(y)/2)$. Thus, it suffices to show that the system

$$\begin{aligned} \dot{x} &= \sigma(y) \\ \dot{y} &= -\frac{1}{2}\sigma(x) - \frac{1}{2}\sigma(y) \end{aligned}$$

is globally asymptotically stable. For this we take the Lyapunov function

$$V(x, y) \ := \ \frac{1}{2}\int_0^x \sigma(s)\, ds \ + \ \int_0^y \sigma(s)\, ds\,.$$

This function is positive definite and proper, because of the assumptions made on σ. Its derivative along trajectories is $-[\sigma(y)]^2/2 \leq 0$ and only vanishes along trajectories for which $y(t) \equiv 0$. Along such trajectories $0 \equiv \dot{y} = (-1/2)\sigma(x)$ implies that also $x \equiv 0$. By the LaSalle invariance principle, the system is indeed globally asymptotically stable. ∎

4 Proof of Proposition 1.5

Lemma 4.1 Suppose that $\sigma : \mathbb{R} \to \mathbb{R}$ is any continuous function, and $b_i, i = 1,2,3,4$ are real numbers, so that

$$\sigma(a + b_1\mu) + \sigma(a + b_2\mu) + \sigma(b_3\mu) + \sigma(b_4\mu) > 0 \qquad \forall\, \mu \in \mathbb{R},\, \forall\, a \geq 1\,. \tag{10}$$

Consider any measurable $u : [0,T] \to \mathbb{R}$ and assume that $(x_1(\cdot), x_2(\cdot), x_3(\cdot), x_4(\cdot))$ is an (absolutely continuous) solution of the system

$$\begin{aligned}
\dot{x}_1 &= \sigma(x_1 + x_2 + x_3 + x_4 + b_1 u) \\
\dot{x}_2 &= \sigma(x_1 + x_2 + x_3 + x_4 + b_2 u) \\
\dot{x}_3 &= \sigma(b_3 u) \\
\dot{x}_4 &= \sigma(b_4 u)
\end{aligned}$$

defined on $[0,T]$. Along this solution, let

$$\alpha(t) := x_1(t) + x_2(t) + x_3(t) + x_4(t)\,.$$

Then, if $\alpha(0) > 1$ it must hold that $\alpha(t) > 1$ for all $t \in [0,T]$.

Proof. If the conclusion is false, there is a t_0 so that $\alpha(t_0) = 1$ and $\alpha(t) > 1$ for all $t \in [0, t_0)$. But Property (10) says that $\alpha'(t) > 0$ for almost all $t \in [0, t_0)$, which means that $\alpha(t_0) \geq \alpha(0) > 1$, a contradiction. ∎

Note that in this case the system cannot be completely controllable; for instance, the initial state $(1,1,1,1)$ ($\alpha(0) = 4$) cannot be controlled to the origin ($\alpha(0) = 0$).

The Lemma given below says that Property (10) is satisfied when $\sigma = \arctan$, $b_1 = 2$, $b_2 = 12$, $b_3 = -3$, and $b_4 = -4$. This establishes the validity of Proposition 1.5.

Lemma 4.2 The function

$$f_a(\mu) = \arctan(a + 2\mu) + \arctan(a + 12\mu) - \arctan(3\mu) - \arctan(4\mu)$$

is positive for all $\mu \in \mathbb{R}$ and all $a \geq 1$.

Proof. Since $f_a(u)$ is increasing as a function of a, it suffices to prove the result for the special case $a = 1$. We write $f = f_1$. Note that

$$q(\mu) f'(\mu) = -\mu p(\mu)$$

where

$$q(\mu) = (1 + 2\mu + 2\mu^2)(1 + 9\mu^2)(1 + 16\mu^2)(1 + 12\mu + 72\mu^2)$$

is always positive, and

$$p(\mu) = 74 + 511\mu + 1752\mu^2 + 6132\mu^3 + 10656\mu^4\,.$$

The polynomial p has just two real roots, both in the interval $I = [-0.32, -0.26]$ (the roots are approximately at $-.3136, -.2657$), and it is positive outside I. It follows that $f'(\mu) > 0$ if $\mu < 0, \mu \notin I$ and $f'(\mu) \leq 0$ if $\mu \geq 0$, which together with

$$\lim_{\mu \to \infty} \mu^2 f(\mu) \;=\; \lim_{\mu \to -\infty} \mu^2 f(\mu) \;=\; \frac{37}{144} \;>\; 0\,,$$

and $f(-0.26) > 0$ implies that $f(\mu) > 0$ whenever $\mu \notin I$. Thus we only need to prove that $f > 0$ on I. Write $f(\mu) = f_1(\mu) - f_2(\mu)$, where

$$f_1(\mu) = \arctan(1 + 2\mu) + \arctan(1 + 12\mu)\,.$$

Note that both f_1 and f_2 are strictly increasing functions. For $\mu \in I$, $f_1(\mu) \geq f_1(-0.32) > -0.9$ and $f_2(\mu) \leq f_2(-0.26) < -1.4$, so indeed $f_1(\mu) - f_2(\mu) > -0.9 + 1.4 > 0$ on this interval. ∎

References

[1] Albertini, F., and Dai Pra, P., "Forward accessibility for recurrent neural networks," *IEEE Trans. Automat. Control* **40** (1995): 1962-1968

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

[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, "State observability in recurrent neural networks," *Systems & Control Letters* **22**(1994): 235-244.

[5] Back, A.D., and A.C. Tsoi, "FIR and IIR synapses, a new neural network architecture for time-series modeling," *Neural Computation* **3** (1991): 375–385.

[6] Bengio, Y., *Neural Networks for Speech and Sequence Recognition*, Thompson Computer Press, Boston, 1996.

[7] Dasgupta, B., and E.D. Sontag, "Sample complexity for learning recurrent perceptron mappings," *IEEE Trans. Inform. Theory* **42** (1996): 1479-1487.

[8] Koiran, P., and E.D. Sontag, "Vapnik-Chervonenkis dimension of recurrent neural networks," submitted.

[9] Koplon, R., and E.D. Sontag, "Using Fourier-neural recurrent networks to fit sequential input/output data," *Neurocomputing*, to appear.

[10] 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*, Zürich, pp. 197-200, July 1990.

[11] Polycarpou, M.M., and P.A. Ioannou, "Neural networks and on-line approximators for adaptive control," in *Proc. Seventh Yale Workshop on Adaptive and Learning Systems*, pp. 93-798, Yale University, 1992.

[12] Siegelmann, H.T., and E.D. Sontag, "On the computational power of neural nets," *J. Comp. Syst. Sci.* **50**(1995): 132-150.

[13] Sontag, E.D., "Neural nets as systems models and controllers," in *Proc. Seventh Yale Workshop on Adaptive and Learning Systems*, pp. 73-79, Yale University, 1992.

[14] Sontag, E.D., "Neural networks for control," in *Essays on Control: Perspectives in the Theory and its Applications* (H.L. Trentelman and J.C. Willems, eds.), Birkhauser, Boston, 1993, pp. 339-380.

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

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

[17] Zbikowski, R., "Lie algebra of recurrent neural networks and identifiability," *Proc. Amer. Automatic Control Conference*, San Francisco, 1993, pp. 2900-2901.

[18] Zbikowski, R., and K.J. Hunt, eds., *Neural Adaptive Control Technology* World Scientific Publishing, 1996.