Recurrent Neural Networks: Identification and other System Theoretic Properties

Francesca Albertini [1]
and
Paolo Dai Pra
Dipartimento di Matematica Pura e Applicata
Università di Padova
Via Belzoni 7,
35131 Padova, Italy
albertin@math.unipd.it
daipra@math.unipd.it
Phone : +39 (0)49 8275966
Fax : +39 (0)49 8758596

[1]Also: Istituto di Ingegneria Gestionale, Viale X Giugno 22, 36100, Vicenza, Italy.

Contents

Chapter 1

Introduction

In recent years neural networks have become a widely used tool for both modeling and computational purposes. From plasma control to image or sound processing, from associative memories to digital control, neural networks techniques have been appreciated for their effectiveness and their relatively simple implementability. We refer the reader to [16, 17] (and references therein) for seminal works, and to [26, 9, 32] for more recent reviews on the subject.

The rigorous analysis of neural networks models has attracted less interest that their applications, and so it has proceeded at a slower pace. There are, of course, some exceptions. In the context of associative memories, for instance, a rather sophisticated study can be found in [31, 10].

The purpose of this work is to present up-to-date results on a dynamical version of neural networks, namely on *Recurrent Neural Networks*. Recurrent neural networks are control dynamical systems that, in continuous time, are described by a system of differential equations of the form

$$\left\{ \begin{array}{rcl} \dot{x} & = & \vec{\sigma}(Ax + Bu) \\ \\ y & = & Cx \end{array} \right. \tag{1}$$

where $x \in \mathbb{R}^n$, $u \in \mathbb{R}^m$, $y \in \mathbb{R}^p$ and A, B, C are matrices of appropriate dimension. The function $\vec{\sigma} : \mathbb{R}^n \to \mathbb{R}^n$ is defined by $\vec{\sigma}(x) = (\sigma(x_1), \ldots, \sigma(x_n))$, where σ is an assigned nonlinear function. In (1), u represents the input signal, or *control*, and y represent the output signal, or *observation*. Systems of type (1) are commonly used as *realizations* for nonlinear input/output behaviors; the entries of the matrices A, B, C, often called *weights* of the networks, are determined on the basis of empirical data using some appropriate *best fitting* criterion. A brief survey on the use of recurrent neural networks as models for nonlinear systems can be found in [24]. Models of these types are used in many different areas; see for example [9, 18] and [21, 26] for signal processing and control applications respectively.

The system theoretic study of recurrent neural networks was initiated in [30, 6, 7], and continued in [8, 2]. The purpose of this study is twofold.

- Systems of type (1) provide a class of "semilinear" systems, whose behavior may be expected to exhibit some similarity with linear systems, that correspond to the choice σ = identity. It is therefore natural the attempt of characterizing the standard system theoretic properties - observability, controllability, identifiability, detectability,... - in terms of algebraic equations for the matrices A, B, C, as for linear systems. One may expect that the nonlinearity introduced by a "typical" function σ induces some "chaotic" disorder in the system, whose overall effect can be described in simple terms. Surprisingly enough, this picture turns out to be correct, under quite reasonable assumptions on the model.

- In modeling nonlinear input/output behaviors by recurrent neural networks, it is desirable to avoid redoudances. More precisely, one would like to work with a recurrent neural network that is minimal; this means that the same input/output behavior cannot be produced by a recurrent neural network with lower state space dimension ($= n$ in (1)). This leads to a classical identifiability problem: characterize minimal recurrent neural networks in terms of their weights. Moreover, by analogy with linear systems, questions on relationships between minimality, observability and controllability naturally arise.

There are many other interesting issues related to recurrent neural networks, such as computability, parameter reconstruction, stability properties, memory capacity, sample complexity for learning and others which will not be addressed in this work. Some references on these subjects are for example [23, 13, 19, 22, 11, 12].

The analysis of recurrent neural networks, both in discrete and continuous time, is the subject of Chapter 2. After stating some assumptions on the model (Sections 2.1 and 2.2), we give necessary and sufficient conditions for a system of type (1) to be observable, i.e. for the initial state of the system to be determined by the input/output behavior. Such conditions can be checked by performing a finite number of linear operations on the entries of the matrices A, C. Moreover, it is shown that observability of the corresponding linear system (σ = identity) implies observability of the network, while the opposite is false.

In Section 2.4 we turn to the problem of characterizing identifiability and minimality of recurrent neural networks. These notions are shown to be essentially equivalent to observability. Moreover, for non-observable systems, it is shown that through an observability reduction similar to the one for linear systems (see e.g. [27]), one can lower the state-space dimension without changing the input/output map.

Unlike for linear systems, it may appear that controllability does not play any role in minimality. However, this is not the case. The above observability and minimality results are proved under "genericity" assumptions on the function σ and the control matrix B. In Section 2.5 we show, for continuous time systems, that these assumptions imply forward accessibility, which means that by choosing different controls u we can steer the state of the system to any point of a nonempty open set. Forward accessibility for a discrete-time recurrent neural network is a more difficult problem, and not completely understood. Known results on the subject are summarized in Section 2.5.

Systems of type (1), although very flexible, may be not the best choice in some contexts. In many engineering applications nonlinear systems arise as perturbations of linear ones. This happens, for instance, when a linear and a nonlinear system are *interconnected* (see [27], Chapter 6). There is a natural way of generalizing systems of type (1) to include interconnections with linear systems. We consider control systems that, in continuous time, take the form

$$\left\{\begin{array}{rcl} \dot{x}_1 & = & \vec{\sigma}(A^{11}x_1 + A^{12}x_2 + B^1 u) \\ \\ \dot{x}_2 & = & A^{21}x_1 + A^{22}x_2 + B_2 u \\ \\ y & = & C^1 x_1 + C^2 x_2 \end{array}\right. \tag{2}$$

where $x_1 \in \mathbb{R}^{n_1}$, $x_2 \in \mathbb{R}^{n_2}$, $u \in \mathbb{R}^m$, $y \in \mathbb{R}^p$, and σ is a given nonlinear function. In Chapter 3 we present a complete analysis of observability, identifiability and minimality for such systems, that will be called *Mixed Networks.* The results on observability can also be found in [1, 3], while results concerning identifiability and minimality, that are technically the most demanding, are original contributions of this work.

It is relevant to point out that the main point in understanding minimality of both systems of type (1) and (2), consists in determining the *symmetry group* of minimal systems, i.e. the linear transformations of the state-space that do not affect the type of the system and its input output behavior. It is well known that the symmetry group of a minimal linear system of dimension n is $GL(n)$, the group of all invertible linear transformations. It turns out that minimal recurrent neural networks have a finite symmetry group, while mixed networks have an infinite symmetry group, but much smaller than $GL(n_1+n_2)$. This reduction of the symmetry group is not surprising, since the nonlinearity of σ prevents linear symmetries. What is remarkable is that the knowledge of linear symmetries is enough to understand observability and identifiability.

Chapter 4 presents some open problems related to the properties of recurrent neural networks and mixed networks discussed in this work.

We are both indebted to Prof. E.Sontag, who has lead us into the subject of recurrent neural networks, and has contributed to developing many of the techniques used in this work.

Chapter 2

Recurrent neural networks

2.1 The model

In this Chapter we consider Recurrent Neural Networks evolving either in discrete or continuous time. We use the superscript "+" to denote time shift (discrete time) or time derivative (continuous time). The basic models we deal with are that in which the dynamics are assigned by the following difference or differential equation

$$\begin{array}{rcl} x^+ & = & \vec{\sigma}(Ax + Bu) \\ y & = & Cx \end{array} \tag{3}$$

with $x \in \mathbb{R}^n$, $u \in \mathbb{R}^m$, $y \in \mathbb{R}^p$, $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$ and $C \in \mathbb{R}^{p\times n}$. Moreover $\vec{\sigma}(x) = (\sigma(x_1), \ldots, \sigma(x_n))$, with σ a given *odd* function from $\mathbb{R}$ to itself. Clearly, if σ is the identity, the model in (3) is just a standard linear system. If σ is a nonlinear function, then systems of type (3) are called (single layer) Recurrent Neural Networks (RNN), and the function σ is referred to as the *activation function.* For continuous time models, we always assume that the activation function σ is, at least, locally Lipschitz, and the control map $u(\cdot)$ is locally essentially bounded, so the differential equation in (3) has an unique local solution.

The system theoretic analysis of RNN's will be carried on under two basic conditions, involving the activation function σ and the *control matrix* B. The first condition is a nonlinearity requirement on σ.

Definition 2.1.1 A function $\sigma : \mathbb{R} \to \mathbb{R}$ is an *admissible* activation function if, for all $N \in \mathbb{N}$ and all pairs $(a_1, b_1), \ldots, (a_N, b_N) \in \mathbb{R}^2$ such that $b_i \neq 0$ and $(a_i, b_i) \neq \pm(a_j, b_j)$ for all $i \neq j$, the functions $\xi \to \sigma(a_i + b_i\xi)$, $i = 1, \ldots, N$, and the constant 1 are linearly independent.

Note that no polynomial is an admissible activation function. Sufficient conditions and example of admissible activation functions will be given in Section 2.2.

Next condition is a controllability-type requirement on the matrix B.

Definition 2.1.2 A matrix $B \in \mathbb{R}^{n\times m}$ is an *admissible* control matrix if it has no zero row, and there are no two rows that are equal or opposite.

The above assumption is equivalent to say that, whatever the matrix A is, and provided σ is an admissible activation function, the orbit of the control system

$$x^+ = \vec{\sigma}(Ax + Bu) \tag{4}$$

is not confined in any subspace of positive codimension. In fact, the following statement is easy to prove.

Proposition 2.1.3 *If σ is an admissible activation function, then B is an admissible control matrix if and only if $span\{\vec{\sigma}(Bu) : u \in \mathbb{R}^m\} = \mathbb{R}^n$ or, equivalently, for all $x \in \mathbb{R}^n$ $span\{\vec{\sigma}(x + Bu) : u \in \mathbb{R}^m\} = \mathbb{R}^n$.*

The relation between the admissibility of B and controllability of the system (3) will be the subject of Section 2.5.

Definition 2.1.4 A system of type (3) is said to be an *admissible* RNN if both the activation function σ and the control matrix B are admissible.

2.2 Examples of admissible activation functions

In this Section we state and prove two criteria for verifying admissibility of an activation function. For the first one see also [5].

Proposition 2.2.1 *Suppose σ is a real analytic function, that can be extended to a complex analytic function on a strip $\{z \in \mathbb{C} : |Im(z)| \leq c\} \setminus \{z_0, \bar{z}_0\}$, where $|z_0| = c$ and $z_0, \bar{z}_0$ are singularities (poles or essential singularities). Then σ is admissible.*

Proof. Let $(a_1, b_1), \ldots, (a_N, b_N)$ be such that $b_i \neq 0$ and $(a_i, b_i) \neq \pm(a_j, b_j)$ for all $i \neq j$. Suppose $|b_1| \geq |b_2| \geq \cdots \geq |b_N|$, and let $c_1, \ldots, c_N \in \mathbb{R}$ be such that

$$\sum_{i=1}^{N} c_i \sigma(a_i + b_i \xi) = 0 \tag{5}$$

for all $\xi \in \mathbb{R}$. Since σ is odd, we can assume without loss of generality that $b_i > 0$ for all i. Note that the identity (5) extends to $\{\xi \in \mathbb{C} : |Im(\xi)| \leq \frac{c}{b_1}\} \setminus \{\frac{z_0 - a_i}{b_i}, \frac{\bar{z}_0 - a_i}{b_i} : i = 1, \ldots, k\}$ where $k \leq n$ and $b_1 = \cdots = b_k$. Now, let $\xi_n \in \mathbb{C}$ be a sequence with $|Im(\xi_n)| < \frac{c}{b_1}$, such that $\xi_n \to \frac{z_0 - a_1}{b_1}$ and $|\sigma(a_1 + b_1 \xi_n)| \to +\infty$ as $n \to \infty$. Note that, for every $i > 1$

$$\lim_{n \to \infty} \sigma(a_i + b_i \xi_n) = \sigma(a_i + b_i(\frac{z_0 - a_1}{b_1})) \in \mathbb{C}. \tag{6}$$

Thus, dividing expression (5) by $\sigma(a_1 + b_1 \xi)$ and evaluating at $\xi = \xi_n$, we get:

$$c_1 + \sum_{i=2}^{N} c_i \frac{\sigma(a_i + b_i \xi_n)}{\sigma(a_1 + b_1 \xi_n)} = 0.$$

Letting $n \to \infty$, we conclude $c_1 = 0$. By repeating the the same argument we get $c_i = 0$ for all i, which completes the proof. ∎

It is very easy to show that two standard examples of activation functions, namely $\sigma(x) = \tanh(x)$ and $\sigma(x) = \arctan(x)$, satisfy the assumptions of Proposition 2.2.1, and so are admissible.

The above criterion is based on the behavior of the activation function near a complex singularity. There are natural candidates for activation function that are either non-analytic or entire, so Proposition 2.2.1 does not apply. We give here a further criterion which requires a suitable behavior of the activation function at infinity.

Proposition 2.2.2 *Let σ be an odd, bounded function such that, for M large enough, its restriction to $[M, +\infty)$ is strictly increasing and has decreasing derivative. Define*

$$f(x) = \lim_{\xi \to \infty} \sigma(\xi) - \sigma(x), \tag{7}$$

and assume

$$\lim_{x \to \infty} \frac{f(x)}{f'(x)} = 0. \tag{8}$$

Then σ is an admissible activation function.

Proof. Suppose

$$\sum_{i=1}^{N} c_i \sigma(a_i + b_i \xi) \equiv 0, \tag{9}$$

with $b_i \neq 0$ and $(a_i, b_i) \neq \pm(a_j, b_j)$ for all $i \neq j$. As before, we may assume $b_i > 0$ for all i. We order increasingly the pairs (a_i, b_i) according to the order relation

$$(a, b) > (a', b') \quad \text{if} \quad (b > b') \quad \text{or} \quad (b = b' \text{ and } a > a'). \tag{10}$$

So we assume $(a_1, b_1) < \cdots < (a_N, b_N)$. Letting $\xi \to \infty$ in (9), we get $\sum_i c_i = 0$. This implies

$$\sum_{i=1}^{N} c_i f(a_i + b_i \xi) \equiv 0. \tag{11}$$

We now divide the expression in (11) by $f(a_1 + b_1 \xi)$, and let $\xi \to \infty$. In this way, if we can show that

$$\lim_{\xi \to \infty} \frac{f(a_i + b_i \xi)}{f(a_1 + b_1 \xi)} = 0 \tag{12}$$

for every $i \geq 2$, then we get $c_1 = 0$ and, by iterating the argument, $c_i = 0$ for all i's.

Notice that $a_i + b_i \xi = a_1 + b_1 \xi + (a_i - a_1) + (b_i - b_1)\xi$, so for some $c > 0$, and for ξ sufficiently large, we have $a_i + b_i \xi > a_1 + b_1 \xi + c$. Thus, to prove (12) it is enough to show that, for all $c > 0$,

$$\lim_{x \to \infty} \frac{f(x + c)}{f(x)} = 0. \tag{13}$$

By assumption, f' is increasing for x large, and so

$$f(x + c) \leq f(x) + f'(x + c)c \tag{14}$$

for x sufficiently large. Thus

$$\lim_{x \to \infty} \frac{f(x)}{f(x + c)} \geq \lim_{x \to \infty} \Big(1 - \frac{f'(x + c)}{f(x + c)} c\Big) = +\infty \tag{15}$$

which completes the proof. ∎

Examples of functions satisfying the criterion in Proposition 2.2.2 are

$$\sigma(x) = \mathrm{sgn}(x)[1 - e^{-x^2}] \tag{16}$$

and

$$\sigma(x) = \int_0^x e^{-t^2} dt. \tag{17}$$

2.3 State observability

In general, let Σ be a given system, with dynamics:

$$\begin{array}{rcl} x^+ & = & f(x,u) \\ y & = & h(x) \end{array} \tag{18}$$

where $x \in \mathbb{R}^n$, $u \in \mathbb{R}^m$, $y \in \mathbb{R}^p$; again, the superscript "+" denotes time shift (discrete time) or time derivative (continuous time). Assume, for continuous time models, that the map f is, at least, locally Lipschitz. Moreover let $x_0 \in \mathbb{R}^n$ be a fixed initial state, then to the pair (Σ, x_0) we associate an input/output map λ_{Σ,x_0} as follows.

- For discrete time models, to any input sequence $u_1, \ldots, u_k \in \mathbb{R}^m$, λ_{Σ,x_0} associates the output sequence $y_0 = h(x_0), y_1, \ldots, y_k \in \mathbb{R}^p$ generated by solving (18), with controls u_i $(i = 1, \ldots, k)$, and initial condition x_0.
- For continuous time models, to any input map $u : [0,T] \to \mathbb{R}^m$, which is, at least, locally essentially bounded, we first let $x(t)$ be the solution of the differential equation in (18) with control $u(\cdot)$ an initial condition x_0, and we denote by $[0, \epsilon_u)$ the maximal interval on which this solution is defined. Then, to the input map $u(\cdot)$, λ_{Σ,x_0} associates the output function $y(t) = h(x(t))$, $t \in [0, \epsilon_u)$.

Definition 2.3.1 We say that two states $x_0, x_0' \in \mathbb{R}^n$ are *indistinguishable* for the system Σ if $\lambda_{\Sigma,x_0} = \lambda_{\Sigma,x_0'}$. The system Σ is said to be *observable* if no two different states are indistinguishable.

The above notion is quite standard in system theory. In the case of linear systems, observability is well understood, and the following Theorem holds.

Theorem 2.3.2 *Suppose $\sigma(x) = x$ in (3). Denote by V the largest subspace of $\mathbb{R}^n$ that is A-invariant (i.e. $AV \subset V$) and contained in $\ker C$. Then two states x, z are indistinguishable if and only if $x - z \in V$. In particular, the system is observable if and only if $V = \{0\}$.*

Among the various equivalent observability conditions for linear systems, the one above is the most suitable for comparison with the result we will obtain for Recurrent Neural Networks. It should be remarked that, for general nonlinear systems, rather weak observability results are known (see e.g. [15, 20, 4]).

Before stating the main result of this section we introduce a simple notion, that plays a significant role in this work.

Definition 2.3.3 A subspace $V \subset \mathbb{R}^n$ is called a *coordinate subspace* if it is generated by elements of the canonical basis $\{e_1, \ldots, e_n\}$.

The relevance of coordinate subspaces in the theory of RNN's comes essentially from the fact that they are the only subspaces satisfying the following property, for σ and B admissible:

$$x, z \in \mathbb{R}^n,\ x - z \in V \implies \vec{\sigma}(x + Bu) - \vec{\sigma}(z + Bu) \in V \quad \forall u \in \mathbb{R}^m. \tag{19}$$

This "invariance" property of coordinate subspaces is responsible for some of the linear theory to survive.

Theorem 2.3.4 *Consider an admissible system of type (3). Let $V \subset \mathbb{R}^n$ be the largest subspace such that*
i) $AV \subset V$ and $V \subset \ker C$;
ii) AV is a coordinate subspace.
Then $x, z \in \mathbb{R}^n$ are indistinguishable if and only if $x - z \in V$. In particular, the system is observable if and only if $V = \{0\}$.

Theorem 2.3.4 is a special case of Theorem 3.2.2, whose proof is given in Section 3.3 (see also [8]). We devote the rest of this section to comments and examples.

First of all note that the *largest* subspace satisfying i) and ii) is well defined. Indeed, if V_1, V_2 satisfy i) and ii), so does $V_1 + V_2$. Moreover, if $\hat{V}$ is the largest subspace for which i) holds, then $x - z \in V$ if and only if x, z are indistinguishable for the linear system (A, C). It follows that observability of the linear system (A, C) implies observability of the RNN (3).

We now show that the observabilty condition in Theorem 2.3.4 can be efficiently checked by a simple algorithm. For a given matrix D let I_D denote the set of those indexes i such that the i-th column of D is zero. Define, recursively, the following sequence of subsets of $\{1, 2, \ldots, n\}$:

$$\begin{aligned} J_0 &= \{1, 2, \ldots, n\} \setminus I_C \\ J_{d+1} &= J_0 \cup \{i : \exists j \in J_d \text{ such that } A_{ij} \neq 0\}. \end{aligned} \tag{20}$$

Note that the sequence J_d is increasing, and stabilizes after at most n steps. Now let

$$O_c(A, C) = span\{e_j : j \notin J_\infty\} \tag{21}$$

with $J_\infty = \cap_d J_d$.

Proposition 2.3.5 *The subspace V in Theorem 2.3.4 is given by*

$$V = \ker C \cap A^{-1}\Big(O_c(A, C)\Big). \tag{22}$$

In particular, the system is observable if and only if

$$\ker A \cap \ker C = O_c(A, C) = \{0\}. \tag{23}$$

Proof. It is not hard to see that $O_c(A, C)$ is the largest coordinate subspace contained in $\ker C$ and A-stable. Thus, the r.h.s. of (22) satisfies conditions i) and ii) of Theorem 2.3.4, and thus it is contained in V. For the opposite inclusion just observe that, by definition, $AV \subset O_c(A, C)$. ∎

It is worth noting, and quite easy to prove, that $x - z \in V$ *implies* indistinguishability of x and z, for *any* system of type (3). Admissibility of the system guarantees that the condition is also necessary. We illustrate with some examples how that whole picture change if we drop some admissibility requirements.

Example 2.3.6 Let $\sigma(\cdot)$ be any periodic smooth function of period τ; clearly such a function cannot be admissible. Consider the following system, with $n = 2$ and $p = m = 1$:

$$\begin{aligned} x^+ &= \vec{\sigma}(x + bu) \\ y &= x_1 - x_2, \end{aligned} \tag{24}$$

where b is any admissible 2×1 control matrix.

It is easily checked that the observability conditions in Theorem 2.3.4 are satisfied. However the system is not observable. Indeed, we consider $\bar{x} = (\tau, \tau)$. Then $C\bar{x} = 0$, and, since σ is periodic of period τ, it is easy to see that both for the discrete-time, and for the continuous-time cases, $\bar{x}$ is indistinguishable from 0.

Example 2.3.7 Assume that $\sigma(x) = x^3$, which is not admissible. Consider a system of type (3), with this function σ, $n = 2$, $m = p = 1$, and matrices A, B, and C as follows:

$$\begin{aligned} A &= \begin{pmatrix} 4 & 0 \\ 0 & 1 \end{pmatrix} \\ B &= \begin{pmatrix} 1 \\ 2 \end{pmatrix} \\ C &= (-8, 1) \end{aligned} \tag{25}$$

Notice that also this system satisfies the observability conditions in Theorem 2.3.4, but it is not observable. In fact, let $W = \{(a, 8a) : a \in \mathbb{R}\}$. It is easy to check that if $x \in W$, then $\vec{\sigma}(Ax + Bu) \in W$ for every $u \in \mathbb{R}$. This implies that W is invariant for the dynamics and, being contained in $\ker C$, it contains vectors that are indistinguishable from 0.

In next example we show that the implication "observability of the linear system (A, C)" $\Rightarrow$ "observability of the RNN" may fail if B is not admissible, even though σ is admissible. Note that for linear systems the control matrix B does not influence observability.

Example 2.3.8 Suppose σ is a strictly increasing admissible activation function. Pick any two nonzero real values x_1, x_2 in the image of σ such that:

$$x_1\sigma^{-1}(x_2) \neq x_2\sigma^{-1}(x_1)\,. \tag{26}$$

Such values always exist for nonlinear σ. Consider the discrete-time system with $n = 2$, $p = 1$, and $B = 0$:

$$\begin{aligned} A &= \begin{pmatrix} \frac{\sigma^{-1}(x_1)}{x_1} & 0 \\ 0 & \frac{\sigma^{-1}(x_2)}{x_2} \end{pmatrix} \\ C &= (x_2, -x_1) \end{aligned} \tag{27}$$

Given (26) it is easy to see that the pair (A, C) is observable; however the nonlinear system is not. In fact, the state $x = (x_1, x_2)$ is an equilibrium state and $Cx = 0$, so it is indistinguishable from zero.

2.4 Identifiability and minimality

Recurrent neural networks possess a very simple group of symmetries. Let us consider a model Σ of type (3), and suppose we exchange two components of x. Define $z = Sx$, where S is the operator that exchange two components. Noting that $S = S^{-1}$ and $\vec{\sigma} \circ S = S \circ \vec{\sigma}$ it is easily seen that the following equations hold:

$$\Sigma' = \begin{cases} z^+ &= \vec{\sigma}(A'z + B'u) \\ \\ y &= C'z \end{cases} \tag{28}$$

with $A' = SAS^{-1}$, $B' = SB$, $C' = CS^{-1}$, as for linear systems. Thus we say that the system (28) with initial condition $z_0 = Sx_0$ is *input/output (i/o) equivalent* to (3) with initial condition x_0; this means that $\lambda_{\Sigma,x_0} = \lambda_{\Sigma',z_0}$.

The above argument can be repeated by considering compositions of exchanges, i.e. by letting S be any permutation operator. Since σ is odd, the same apply to operators S that change the sign to some components. Permutation and sign changes generate a finite group that we denote by G_n. We have thus shown that G_n is a symmetry group for any system of type (3), i.e. transformation of the state by an element of G_n gives rise to an equivalent system which is still a RNN.

In what follows a RNN will be identified with the quadruple $\Sigma_n = (\sigma, A, B, C)$, where the index n denotes the dimension of the state space. Moreover, with (Σ_n, x_0), $x_0 \in \mathbb{R}^n$, we will denote the RNN $\Sigma_n = (\sigma, A, B, C)$ together with the initial state x_0; and we will call the pair (Σ_n, x_0) an *initialized RNN.*

We will write

$$(\Sigma_n, x_0) \sim (\Sigma'_{n'}, x'_0)$$

to denote that the two initialized RNN's (Σ_n, x_0), and $(\Sigma'_{n'}, x'_0)$ are input/output equivalent, i.e. $\lambda_{\Sigma_n, x_0} = \lambda_{\Sigma'_{n'}, x'_0}$.

Definition 2.4.1 Let $\Sigma_n = (\sigma, A, B, C)$ and $\Sigma'_{n'} = (\sigma, A', B', C')$ be two RNN's. The two initialized RNN's (Σ_n, x_0) and $(\Sigma'_{n'}, x'_0)$ are said to be *equivalent* if $n = n'$ and there exists $S \in G_n$ such that $A' = SAS^{-1}$, $B' = SB$, $C' = CS^{-1}$, and $x'_0 = Sx_0$.

Definition 2.4.2 An admissible RNN $\Sigma_n = (\sigma, A, B, C)$ is said to be *identifiable* if the following condition holds: for every initial state $x_0 \in \mathbb{R}^n$, and for every admissible initialized RNN $(\Sigma'_{n'} = (\sigma, A', B', C'), x'_0)$ such that (Σ_n, x_0) and $(\Sigma'_{n'}, x'_0)$ are i/o equivalent, then either $n' > n$ or (Σ_n, x_0) and $(\Sigma'_{n'}, x'_0)$ are equivalent.

Note that the notions in Definitions 2.4.1 and 2.4.2 can be given for linear systems (σ = identity) by replacing G_n with $GL(n) = \{n \times n \text{ invertible matrices}\}$ (and cutting the admissibility assumption). We recall the following well known result.

Theorem 2.4.3 *A linear system (A, B, C) is identifiable if and only if is observable and controllable.*

We do not insist here on the notion of controllability, that will be discussed in next section. Remarkably enough, a formally similar result holds for both continuous time and discrete time RNN's.

Theorem 2.4.4 *An admissible RNN is identifiable if and only if it is observable.*

Theorem 2.4.4 is a special case of Theorem 3.4.3, so we do not give its proof here. The "only if" part is not difficult to see. In fact, if a RNN Σ_n is not observable, then there exists two different states x_1, x_2 which are indistinguishable, thus $(\Sigma_n, x_1) \sim (\Sigma_n, x_2)$. Moreover, since the matrix B is admissible, the only matrix $S \in G_n$ such that $B = SB$ is the identity, thus the two initialized RNN's (Σ_n, x_1), (Σ_n, x_2) can not be equivalent.

In comparing Theorems 2.4.3 and 2.4.4 it is seen that no controllability assumption is required in Theorem 2.4.4. In a certain sense, the role of controllability is played by admissibility. Actual connections between admissibility and controllability will be studied in Section 2.5.

We have seen in Section 2.3 that observability of an admissible RNN is equivalent to

$$\ker A \cap \ker C = O_c(A, C) = \{0\}. \tag{29}$$

Suppose $O_c(A,C) \neq \{0\}$, and let $n_1 = \dim O_c(A,C)$. Then, up to a permutation of the elements of the canonical basis, the system (3) can be rewritten in the form

$$\begin{aligned} x_1^+ &= \vec{\sigma}(A_1x_1 + A_2x_2 + B_1u) \\ x_2^+ &= \vec{\sigma}(A_3x_2 + B_2u) \\ y &= C_2x_2 \end{aligned} \tag{30}$$

with $x_1 \in \mathbb{R}^{n_1}$, $x_2 \in \mathbb{R}^{n-n_1}$. It follows that the system $(\Sigma_n = (\sigma, A, B, C))$, with initial state $(x_1, x_2) \in \mathbb{R}^n$ is i/o equivalent to $(\Sigma_{n_2} = (\sigma, A_3, B_2, C_2))$ with initial state $x_2 \in \mathbb{R}^{n-n_1}$. Thus we have performed a reduction in the dimension of the state space. On the other hand, if $O_c(A,C) = \{0\}$, then it will follow from the proof of Theorem 2.4.4 that for $(\Sigma'_{n'} = (\sigma, A', B', C'), x'_0)$ to be i/o equivalent to (Σ_n, x_0) it must be $n' \geq n$. Moreover $n = n'$ if and only if there exists $S \in G_n$ such that $A' = SAS^{-1}$, $B' = SB$, $C' = CS^{-1}$, and x_0, $S^{-1}x'_0$ are indistinguishable for Σ_n. We summarize these remarks as follows.

Definition 2.4.5 An admissible RNN Σ_n is said to be *minimal* if for all $x_0 \in \mathbb{R}^n$, and for any admissible initialized RNN $(\Sigma'_{n'}, x'_0)$ such that (Σ_n, x_0) and $(\Sigma'_{n'}, x'_0)$ are i/o equivalent then it must be $n' \geq n$.

Proposition 2.4.6 *An admissible RNN is minimal if and only if $O_c(A,C) = \{0\}$.*

Notice that clearly if an admissible RNN is identifiable then it is also minimal, while the converse implication may be false, as shown in the next example.

Example 2.4.7 Let $n = 2$, and $m = p = 1$. Consider any RNN $\Sigma_2 = (\sigma, A, B, C)$ where σ is any admissible activation function, and $B = (b_1, b_2)^T$ is any admissible control matrix (i.e. $0 \neq |b_1| \neq |b_2| \neq 0$). Moreover, let $C = (c_1, c_2)$, with $c_i \neq 0$, $i = 1, 2$, and A be the zero matrix.

Then, for this model, $O_c(A,C) = \{0\}$, since C has all nonzero columns, thus by Proposition 2.4.6, it is minimal. On the other hand, this model is not identifiable, since $\ker A \cap \ker C \neq 0$, and so it is not observable.

2.5 Controllability and forward accessibility

A RNN Σ is said to be *controllable* if for every two states x_1, $x_2 \in \mathbb{R}^n$ there exists a sequence of controls $u_1, \ldots, u_k \in \mathbb{R}^m$, for discrete time models, or a control map $u(\cdot) : [0, T] \to \mathbb{R}^m$, for continuous time models, which steers x_1 to x_2. In general, for nonlinear models, the notion of controllability is very difficult to characterize; thus, the weaker notion of forward accessibility is often studied. Let $x_0 \in \mathbb{R}^n$ be a state; Σ is said to be *forward accessible* from x_0, if the set of points that can be reached from x_0, using arbitrary controls, contains an open subset of the state space. A model is said to be forward accessible if it is forward accessible from any state. Even if much weaker than controllability, forward accessibility is an important property. In particular, it implies that from any state, the forward orbit does not lay in a submanifold of the state space with positive codimension.

Unlike the other properties that we have discussed before (observability, identifiability, and minimality) whose characterization was the same for both dynamics, discrete and continuous time, the characterization of controllability and forward accessibility is quite different for the two dynamics.

2.5.1 Controllability and forward accessibility for continuous time RNN's

Let Σ be a continuous time RNN, i.e. the dynamics are given by the differential equation:

$$\dot{x}(t) \;=\; \vec{\sigma}(Ax(t) + Bu(t)), \quad t \in \mathbb{R}. \tag{31}$$

Let X_u be the vector field defined by:

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

Given the vector fields X_u, let $\operatorname{Lie}\{X_u \,|\, u \in \mathbb{R}^m\}$ be the Lie algebra generated by these vector fields. It is known that if $\operatorname{Lie}\{X_u \,|\, u \in \mathbb{R}^m\}$ has full rank at x then the system is forward accessible from x (see [29]). This result together with Proposition 2.1.3 (which says that the $\operatorname{span}\{X_u\}_{u\in\mathbb{R}^m}$ has already full dimension at each $x \in \mathbb{R}^n$), gives the following:

Theorem 2.5.1 *Let Σ be an admissible RNN evolving in continuous time, then Σ is forward accessible.*

Remark 2.5.2 In a recent paper (see [28]) E. Sontag and H. Sussmann proved that if Σ is an admissible RNN, and the activation function σ satisfies some extra assumptions, then Σ is indeed controllable. The extra requirement on σ is fulfilled, for example, when $\sigma = \tanh$. It is quite surprising that these RNN's are controllable for any matrix A.

2.5.2 Controllability and forward accessibility for discrete time RNN's

Let Σ be a discrete time RNN, i.e. the dynamics are given by the difference equation:

$$x(t+1) \;=\; \vec{\sigma}(Ax(t) + Bu(t)), \quad t \in \mathbb{Z}. \tag{32}$$

Unlike in continuous time, characterizing controllability of discrete time models is quite difficult. We will not give the proofs of the results presented in this section since they are very long technical, and we refer the reader to [2].

It is not difficult to see that, in this case, the admissibility assumption is not enough to guarantee forward accessibility. In fact, let $p = \operatorname{rank}[A, B]$, and $V = [A, B]\left(\mathbb{R}^{n+m}\right)$. Then, except possibly for the initial condition, the reachable set from any point is contained in $\vec{\sigma}\,(V)$. If $p < n$ then, clearly, $\vec{\sigma}\,(V)$ does not contain any open set. So a necessary conditions for forward accessibility is $p = n$. The result stated below proves the sufficiency of this condition provided that the control matrix B and the activation function σ satisfy a new condition which is, in some sense, stronger than admissibility.

Definition 2.5.3 We say that the activation function σ and the control matrix B are **n**-*admissible*, if they satisfy the following conditions.

1. σ is differentiable and $\sigma'(x) \neq 0$ for all $x \in \mathbb{R}$.

2. Denote by b_i, $i = 1, \ldots, n$, the rows of the matrix B; then $b_i \neq 0$ for all i.

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 let $a_1, \ldots, a_n$ 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.

A RNN Σ is **n**-*admissible* if its activation function σ and its control matrix B are **n**-admissible.

Remark 2.5.4 Notice that, in the **n**-admissibility property we do not require that no two rows of the matrix B are equal or opposite. Moreover the linear independence on the functions f_I's is asked only for a fixed n. On the other hand, since the functions f_I's involve products of the functions $\sigma'(a_i + b_i u)$, this third requirement is, indeed, quite strong. Notice that the special case $k = 1$ is equivalent to ask that the functions $\sigma(a_i + b_i u)$'s and the constant 1 are linearly independent.

Theorem 2.5.5 *Let Σ be a* **n**-*admissible RNN evolving in discrete time. Then Σ is forward accessible if and only if*

$$\text{rank}\,[A, B] \;=\; n. \tag{33}$$

Remark 2.5.6 The **n**-admissibility property, as stated in Definition 2.5.3, 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. However, it is possible to prove that when $\sigma = \tanh$ then σ and B are **n**-admissible for all matrices B in a "generic" subset of $\mathbb{R}^{n\times m}$, i.e. for B in the complement of an analytic subset of $\mathbb{R}^{n\times m}$ (in particular B may vary in an open dense subset of $\mathbb{R}^{n\times m}$). For more discussion and precise statements on this subject see [2] (section C.).

Next Theorem states another sufficient condition for forward accessibility, using a weaker condition on the map σ but adding a new condition on the pair A, B.

Definition 2.5.7 We say that the activation function σ and the control matrix B are *weakly* **n**-*admissible*, if they satisfy the following conditions.

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

A RNN Σ is *weakly* **n**-*admissible* if its activation function σ and its control matrix B are weakly **n**-admissible.

Remark 2.5.8 Notice that this condition is weaker than the one given in Definition 2.5.3; in fact the third requirement of Definition 2.5.7 is exactly the third requirement of Definition 2.5.3 for the case $k = n - 1$. It is not hard to show that if B is an admissible control matrix then the activation function tanh together with the matrix B are weakly **n**-admissible. However the same would be false for the activation function arctan, for $n \geq 4$.

Theorem 2.5.9 *Let Σ be a weakly* **n**-*admissible RNN evolving in discrete time. 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.

It is easy to see that condition (a) of the previous Theorem is equivalent to rank $[A, B] = n$, thus (a) is also a necessary condition for forward accessibility. So Theorem 2.5.9 adds a new condition on A, B (condition (b)), and guarantees forward accessibility with weaker assumption on σ.

Remark 2.5.10 It is interesting to notice that for the single-input case condition (b) is independent on H. In fact the following fact holds:

> Let $h, k \in \mathbb{R}^{n\times 1}$ be such that $A + bh^t$ and $A + bk^t$ are invertible. Then
>
> $$((A + bk^t)^{-1}b)_i \neq 0 \ \ \forall\, i \quad \Longleftrightarrow \quad ((A + bh^t)^{-1}b)_i \neq 0 \ \ \forall\, i.$$

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

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

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

Chapter 3

Mixed networks

3.1 The model

In this Chapter we consider models obtained by "coupling" a recurrent neural network with a linear system; we called these type of systems *Mixed Networks*. Also for these systems, we consider both models evolving in discrete and continuous time. Again, the superscript "+" will denote time shift (discrete time) or time derivative (continuous time). A Mixed Network (MN) is a system whose dynamics are described by equations of the form:

$$\begin{array}{rcl} x_1^+ & = & \vec{\sigma}(A^{11}x_1 + A^{12}x_2 + B^1 u) \\ x_2^+ & = & A^{21}x_1 + A^{22}x_2 + B^2 u \\ y & = & C^1 x_1 + C^2 x_2 \end{array} \tag{34}$$

with $x_1 \in \mathbb{R}^{n_1}$, $x_2 \in \mathbb{R}^{n_2}$, $u \in \mathbb{R}^m$, $y \in \mathbb{R}^p$, and A^{11}, A^{12}, A^{21}, A^{22}, B^1, B^2, C^1, and C^2 are matrices of appropriate dimensions. We let $n = n_1 + n_2$, and

$$A = \left[\begin{array}{cc} A^{11} & A^{12} \\ A^{21} & A^{22} \end{array} \right], \quad B = \left[\begin{array}{c} B^1 \\ B^2 \end{array} \right], \quad C = \left[C^1, C^2 \right].$$

As for RNN's, we assume that the activation function $\sigma : \mathbb{R} \to \mathbb{R}$ is an odd map. For continuous time models, we always assume that the activation function is, at least, locally Lipschitz, and the control maps are locally essentially bounded; thus local existence and uniqueness of the solutions of the differential equation in (34) are guaranteed.

As for RNN's, the system theoretic analysis for MN's will be carried on for a suitable subclass.

Definition 3.1.1 A system of type (34) is said to be an *admissible* MN if both the activation function σ and the control matrix B^1 are admissible.

For the definitions of an admissible activation function and an admissible matrix see Section 2.1. Notice that the admissibility conditions involveonly the first block of equations in (34), which evolves nonlinearly.

3.2 State observability

In this section we present the state observability result for MN's. For the definitions of state observability and coordinate subspace see the corresponding Section 2.3.

Definition 3.2.1 Consider an admissible MN of type (34). Let $W \subseteq \mathbb{R}^n$ be the maximal subspace (maximal with respect to set inclusion) such that:

i) $AW \subseteq W$, and $W \subseteq \ker C$;

ii) $AW = V_1 \oplus V_2$, where $V_1 \subseteq \mathbb{R}^{n_1}$, $V_2 \subseteq \mathbb{R}^{n_2}$, and V_1 is a coordinate subspace.

We called W the *unobservable subspace.*

It is clear that the unobservable subspace is well-defined, since if W_1, and W_2 are subspaces both satisfying the previous i), and ii), then also $W_1 + W_2$ does.

Theorem 3.2.2 *Let Σ be an admissible MN of type (34), and $W \subseteq \mathbb{R}^n$ be its unobservable subspace. Then $x, z \in \mathbb{R}^n$ are indistinguishable if and only if $x - z \in W$. In particular, Σ is observable if and only if $W = \{0\}$.*

Theorem 3.2.2 will be restated in a different form in Theorem 3.2.5, whose proof is postponed to Section 3.3.

Remark 3.2.3 Notice that if Σ is a RNN (i.e. $n_2 = 0$), then the previous Theorem yields the observability result given in Theorem 2.3.4. On the other hand, if Σ is a linear model (i.e. $n_1 = 0$), then Theorem 3.2.2 gives the usual linear observability result stated in Theorem 2.3.2.

As done for RNN's, we present a simple algorithm to check efficiently the observability condition given by Theorem 3.2.2. First we give a useful characterization of the unobservable subspace W.

Proposition 3.2.4 *Consider an admissible MN of type (34). Let $V_1 \subseteq \mathbb{R}^{n_1}$ and $V_2 \subseteq \mathbb{R}^{n_2}$ be the maximal pair of subspaces (maximal with respect to set inclusion) such that:*

P1. V_1 *is a coordinate subspace,* $V_1 \subset \ker C_1$, $A_{11}V_1 \subseteq V_1$;

P2. $V_2 \subset \ker C_2$, $A_{22}V_2 \subseteq V_2$;

P3. $A_{21}V_1 \subseteq V_2$;

P4. $A_{12}V_2 \subseteq V_1$.

Then the unobservable subspace W is given by

$$W = A^{-1}(V_1 \oplus V_2) \cap \ker C.$$

Proof. First we prove that if $V_1 \subseteq \mathbb{R}^{n_1}$ and $V_2 \subseteq \mathbb{R}^{n_2}$ is any pair of subspaces satisfying properties $P1 - P4$ then $W = A^{-1}(V_1 \oplus V_2) \cap \ker C$ satisfies properties i) and ii) of Definition 3.2.1.

i) Clearly $W \subseteq \ker C$. To see that $AW \subseteq W$ we argue as follows. $AW \subseteq V_1 \oplus V_2 \subseteq \ker C$, since $V_1 \subseteq \ker C_1$, and $V_2 \subseteq \ker C_2$. On the other hand, since $A_{11}V_1 \subseteq V_1$, $A_{22}V_2 \subseteq V_2$, and properties $P3$, $P4$ hold, one gets that $A(V_1 \oplus V_2) \subseteq V_1 \oplus V_2$. Thus, if $x \in W \subseteq A^{-1}(V_1 \oplus V_2)$, then $Ax \in V_1 \oplus V_2$, which implies that $A(Ax) \in A(V_1 \oplus V_2) \subseteq V_1 \oplus V_2$. So $Ax \in A^{-1}(V_1 \oplus V_2) \cap \ker C$, as desired.

ii) We will prove that $AW = V_1 \oplus V_2$. We need only to establish that $V_1 \oplus V_2 \subseteq AW$, being the other inclusion obvious. Notice that $V_1 \oplus V_2 \subseteq \ker C$ and, since $A(V_1 \oplus V_2) \subseteq V_1 \oplus V_2$, we get also $V_1 \oplus V_2 \subseteq A^{-1}(V_1 \oplus V_2)$, and so $V_1 \oplus V_2 \subseteq AW$.

It is not difficult to see that if $W \subseteq \mathbb{R}^n$ is any subspace which satisfies properties i), and ii) of Definition 3.2.1, then the two subspaces $V_1 \subseteq \mathbb{R}^{n_1}$ and $V_2 \subseteq \mathbb{R}^{n_2}$ such that $AW = V_1 \oplus V_2$, satisfy properties $P1 - P4$.

Now, by maximality of both the unobservable subspace W and the pair V_1, V_2, the conclusion follows. ∎

If $V_1 \subseteq \mathbb{R}^{n_1}$ and $V_2 \subseteq \mathbb{R}^{n_2}$ are the two subspaces defined in Proposition 3.2.4, then one has:

$$W = 0 \quad \Leftrightarrow \quad \ker A \cap \ker C = 0, \; V_1 = 0, \; V_2 = 0. \tag{35}$$

In fact, $W = A^{-1}(V_1 \oplus V_2) \cap \ker C$, thus if $W = 0$ then $\ker A \cap \ker C \subseteq W = 0$. Moreover, since $V_1 \oplus V_2 \subseteq A^{-1}(V_1 \oplus V_2) \cap \ker C$, we also have that $V_1 = 0$ and $V_2 = 0$. On the other hand, if $V_1 = 0$ and $V_2 = 0$, then $W = A^{-1}(V_1 \oplus V_2) \cap \ker C = \ker A \cap \ker C = 0$.

Using (35), Theorem 3.2.2 can be rewritten as follows.

Theorem 3.2.5 *Let Σ be an admissible MN of type (34), and $V_1 \subseteq \mathbb{R}^{n_1}$, $V_2 \subseteq \mathbb{R}^{n_2}$ be the two subspaces defined in Proposition 3.2.4. Then $x, z \in \mathbb{R}^n$ are indistinguishable if and only if $x - z \in \ker C$ and $A(x-z) \in V_1 \oplus V_2$. In particular, Σ is observable if and only if*

$$\ker A \cap \ker C = 0, \quad \text{and} \quad V_1 = 0, \; V_2 = 0. \tag{36}$$

Before proving this result, we present an algorithm to compute the subspaces V_1 and V_2, which consists in solving a finite number of linear algebraic equations.

Inductively, we define an increasing sequence of indexes J_d, and two decreasing sequences of subspaces $V_1^d \subseteq \mathbb{R}^{n_1}$, $V_2^d \subseteq \mathbb{R}^{n_2}$, for $d \geq 1$, where V_1^d is a coordinate subspace. Recall that for a given matrix D, we denote by I_D the set of indexes i such that the i-th column of D is zero. Let:

$$\begin{array}{rcl} J_1 & = & \{1, \ldots, k\} \setminus I_{C_1}; \\ V_1^1 & = & \text{span}\,\{e_j \mid j \notin J_1\}; \\ V_2^1 & = & \ker C_2; \end{array}$$

and, for $d > 1$, let:

$$\begin{array}{rcl} J_{d+1} & = & J_1 \cup \{i \mid \exists j \in J_d \text{ such that } A^{11}_{ji} \neq 0\} \\ & & \cup \{i \mid \exists j, \exists 0 \leq l \leq d-1 \text{ such that } (C^2 (A^{22})^l A^{21})_{ji} \neq 0\} \\ & & \cup \cup_{s=1}^{d-1} \{i \mid \exists j \in J_s \text{ such that } (A^{12}(A^{22})^{d-s-1} A^{21})_{ji} \neq 0\} \\ & & \\ V_1^{d+1} & = & \text{span}\,\{e_j \mid j \notin J_{d+1}\} \\ & & \\ V_2^{d+1} & = & \{w \mid (A^{22})^l w \in \ker C^2 \text{ for } 0 \leq l \leq d, A^{12}(A^{22})^{d-s} w \in V_1^s \text{ for } 1 \leq s \leq d\}. \end{array}$$

Remark 3.2.6 It is easy to show that the two sequences V_1^d, and V_2^d for $d \geq 1$ are both decreasing; thus they must become stationary after a finite number of steps. One can find conditions that guarantee the termination of the previous algorithm. Assume that a stationary string $V_1^s = V_1^{s+1} = \ldots = V_1^{s+n_1}$ of length $n_1 + 1$ is obtained. Then, using both the definitions of V_1^d and V_2^d, and applying the Hamilton-Cayley Theorem, one proves that $V_1^d = V_1^s$ for all $d \geq s$ and that $V_2^d = V_2^{s+n_2}$ for all $d \geq s + n_2$. Thus the two sequences V_1^d, and V_2^d become stationary after at most $(n_2+1)n_1$ steps, for $n_1 \geq 1$, or n steps, for $n_1 = 0$.

The two sequences V_1^d, and V_2^d stabilize exactly at V_1, and V_2 as stated in Proposition 3.2.8. Moreover, for MN which evolves in discrete time, the previous subspaces V_1^d, and V_2^d have a precise meaning, as stated next. For discrete time models, given any x, $z \in \mathbb{R}^n$ and any $0 < d \in \mathbb{N}$, we say that x and z are indistinguishable in d-steps if any output sequence from x or z is the same up to time d.

Proposition 3.2.7 *Let Σ be a discrete time admissible MN. Then the following properties are equivalent:*
i) x and z are indistinguishable in d-steps,
ii) $x - z \in \ker C$; $A(x-z) \in (V_1^d \oplus V_2^d)$.

For the proof of this Proposition we refer to [1].

Proposition 3.2.8 *Let $V_1 \subseteq \mathbb{R}^{n_1}$, and $V_2 \subseteq \mathbb{R}^{n_2}$ be the two subspaces defined in Proposition 3.2.4, and V_1^d, and V_2^d be the two sequences of subspaces determined by the algorithm described above. Then the following identities hold:*

$$V_1 = \cap_{d\geq 1} V_1^d, \qquad V_2 = \cap_{d\geq 1} V_2^d. \tag{37}$$

Proof. Let:

$$J_\infty = \cap_{d\geq 1} J_d; \quad V_1^\infty = \cap_{d\geq 1} V_1^d; \quad V_2^\infty = \cap_{d\geq 1} V_2^d.$$

Thus $V_1^\infty = \text{ span } \{e_j \,|\, j \notin J_\infty\}$. The following four properties can be easily proved by using the recursive definition of J_d, V_1^d, and V_2^d and the fact that, for some $\bar{d}$, it holds that $V_1^d = V_1^\infty$, $V_2^d = V_2^\infty$ for $d \geq \bar{d}$.

$$\{\, i \,|\, \exists\, j \in J_\infty \text{ so that } A_{ji}^{11} \neq 0 \,\} \subseteq J_\infty. \tag{38}$$

$$\{ i \,|\, \exists\, j, \exists l \geq 0 \text{ so that } (C^2 (A^{22})^l A^{21})_{ji} \neq 0 \,\} \subseteq J_\infty. \tag{39}$$

$$\{ i \,|\, \exists\, j \in J_\infty, \exists l \geq 0 \text{ so that } (A^{12} (A^{22})^l A^{21})_{ji} \neq 0 \,\} \subseteq J_\infty. \tag{40}$$

$$V_2^\infty \subset \{\, w \,|\, (A^{22})^l w \in \ker C^2 \text{ for } l \geq 0, A^{12}(A^{22})^l w \in V_1^\infty \text{ for } l \geq 0 \,\}. \tag{41}$$

We first prove that the pair V_1^∞ and V_2^∞ satisfies properties P1-P4 of Proposition 3.2.4.

P1. The only non trivial fact is the A^{11}-invariance of V_1^∞. Let $v \in V_1^\infty$. From (38), we get that $v_i = 0$ for all i such that $\exists j \in J_\infty$ with $A_{ji}^{11} \neq 0$. So, for $j \in J_\infty$:

$$(A^{11} v)_j = \sum_{i=1}^{n_1} A_{ji}^{11} v_i = 0$$

and, therefore, $A^{11} v \in V_1^\infty$.

P2. Let $w \in V_2^\infty$. We show that $A^{22} w \in V_2^d$ for every $d \geq 1$. From (41), we first see that $(A^{22})^l (A^{22} w) \in \ker C^2$ for $0 \leq l \leq d$. Moreover, for $1 \leq s \leq d$, again from (41), we have that:

$$A^{12} (A^{22})^{d-s} (A^{22} w) \in V_1^\infty \subseteq V_1^s.$$

P3. Let $v \in V_1^\infty$. We need to prove that $A^{21} v \in V_2^d$ for all $d \geq 1$. Using similar arguments as used to get P1, one can see that (39) yields:

$$C^2 (A^{22})^l A^{21} v = 0 \quad \forall\, l \geq 0,$$

and (40) yields:

$$A^{12} (A^{22})^l A^{21} v \in V_1^\infty \subseteq V_1^s \quad \forall\, s \geq 1.$$

This implies that $A^{21} v \in V_2^d$ for all d, as desired.

P4. This is an immediate consequence of (41).

Thus we have that $V_1^\infty \subseteq V_1$, and $V_2^\infty \subseteq V_2$. To conclude we need to show that also the converse inclusions hold. We will prove, by induction, that $V_1 \subseteq V_1^d$, and $V_2 \subseteq V_2^d$ for all $d \geq 1$. The case $d = 1$ is obvious. Now let $e_i \in V_1$. We show that $e_i \in V_1^{d+1}$, 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
$$(A^{11}e_i)_j = A^{11}_{ji} \neq 0 \qquad \Rightarrow \qquad A_{11}e_i \notin V_1^d.$$
This is impossible since $A^{11}V_1 \subset V_1$ and, by inductive assumption, $V_1 \subset V_1^d$.

- 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 $A^{21}V_1 \subseteq V_2$ and $V_2 \subseteq \ker(C^2(A^{22})^l)$ 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
$$A^{12}(A^{22})^{d-s}A^{21}e_i \notin V_1^s.$$
This is impossible since $(A^{22})^{d-s}A^{21}V_1 \subseteq V_1$, $A^{12}V_2 \subseteq V_2$ and, by inductive assumption, $V_1 \subseteq V_1^s$.

Thus we have shown that $e_i \in V_1^{d+1}$.

Let now $w \in V_2$. We need to prove that $w \in V_2^{d+1}$.

- Since $V_2 \subseteq \ker C^2$, and $A^{22}V_2 \subseteq V_2$, it holds that $(A^{22})^l w \in \ker C^2$ for $0 \leq l \leq d$.

- Since $A^{22}V_2 \subseteq V_2$, $A^{12}V_2 \subseteq V_1$, and, by inductive assumption, $V_1 \subseteq V_1^s$ for all $1 \leq s \leq d$, it holds that $A^{12}(A^{22})^{d-s}w \subset V_1^s$, again for $1 \leq s \leq d$.

∎

We conclude with two examples.

Example 3.2.9 Assume $A_{12} = 0$ and $A_{21} = 0$. Thus the linear and the nonlinear dynamics are decoupled (as in [14], Chapter 6). By what observed in Remarks 3.2.3, we get that observability of the whole system is equivalent to the observability of the RNN characterized by the matrices A_{11}, B_1, and C_1 and of the linear models given by the matrices A_{22}, B_2, and C_2. 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 3.2.10 Assume that C_1 has no zero columns. Then there is no nonzero coordinate subspace contained in $\ker C_1$; so $V_1 = 0$. Therefore V_2 is the largest A_{22}-invariant subspace contained in $\ker C_2$ and $\ker A_{12}$. It follows that the MN is observable if and only $\ker C \cap \ker A = 0$, and the two linear systems with matrices (A_{22}, C_2), and (A_{22}, A_{12}) are observable.

3.3 Proofs on observability

Throughout this section, even if not specifically stated, all the models of MN's we will be dealing with are assumed to be admissible.

A basic ingredient in the proof of the observability result is the following technical fact, which also explains how the admissibility assumptions on the activation map σ and the control matrix B^1 are used (see also [8]). Recall that, for a given matrix D, I_D denotes the set of index i such that the i-th column of D is zero.

Lemma 3.3.1 *Assume that σ is an admissible activation function, $B^1 \in \mathbb{R}^{l\times m}$ is an admissible control matrix, and $D \in \mathbb{R}^{q\times l}$ is any matrix. Then the following two properties are equivalent for each ξ, $\eta, \in \mathbb{R}^l$, and each α, $\beta \in \mathbb{R}^q$:*

1. *$\xi_i = \eta_i$ for all $i \notin I_D$, $\alpha = \beta$,*
2. *$D\vec{\sigma}(\xi + B^1 u) + \alpha = D\vec{\sigma}(\eta + B^1 u) + \beta$ for all $u \in \mathbb{R}^m$.*

Proof. We need only to prove that 2. implies 1., being the other implication obvious. Since B^1 is admissible, there exists $\bar{v} \in \mathbb{R}^m$ such that $b_i = (B^1\bar{v})_i \neq 0$ for all $i = 1, \dots, n$ and $|b_i| \neq |b_j|$ if $i \neq j$. To see this it is enough to notice that the equations $(B^1u)_i = 0$ and $(B^1u)_i = \pm(B^1u)_j$ define a finite number of hyperplanes in $\mathbb{R}^m$, thus to get $\bar{v}$, we only have to avoid their union.

Now, for any $t \in \mathbb{R}$, we consider $t\bar{v} \in \mathbb{R}^m$, and we rewrite (2) as:

$$\sum_{i=1}^{n} D_{li}\sigma\left(\xi_i + b_i t\right) - \sum_{i=1}^{n} D_{li}\sigma\left(\eta_i + b_i t\right) + (\alpha_l - \beta_l) \;=\; 0,\;\; l = 1, \dots, q, \quad \text{and} \quad t \in \mathbb{R}. \qquad (42)$$

Assume that 1. is false. If $\xi_i = \eta_i$ for all $i \notin I_D$, then the first difference in (42) is always zero. Thus if there exists $\bar{l}$ such that $\alpha_{\bar{l}} \neq \beta_{\bar{l}}$, (42) does not hold for this $\bar{l}$, contradicting 2. On the other hand, if there exists $\bar{i} \notin I_D$ such that $\xi_{\bar{i}} \neq \eta_{\bar{i}}$, then by definition of I_D, there exists $\bar{l}$ such that $D_{\bar{l}\bar{i}} \neq 0$. Since in (42) all the pairs for which $\xi_i = \eta_i$ cancel out, we may assume without loss of generality that $\xi_i \neq \eta_i$ for all i (notice that $\xi_{\bar{i}} \neq \eta_{\bar{i}}$, thus not all the pairs cancel). Consider now equation (42) for $l = \bar{l}$. The pairs (ξ_i, b_i), and (η_i, b_i) are all different, thus admissibility of σ implies that (42) can not hold for all $t \in \mathbb{R}$, contradicting 2. ∎

First, we introduce some useful notations. Given $x \in \mathbb{R}^n$ and $u \in \mathbb{R}^m$, for discrete time MN's we denote by $x^+(u)$ the state reached from x using the control value u. For continuous time MN's, 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 $[0, \epsilon_u)$, and it is differentiable on this interval. When dealing with two trajectories of this type starting at two different initial states, by $[0, \epsilon_u)$ we mean the interval in which both trajectories are defined.

Given 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 $u_1, \dots, u_p$, 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 (34) starting at x (resp z), for the entire interval $[0, T]$, and at time T the state x' (resp. z')

is reached. Using 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, proofs in discrete time will be similar to proofs in continuous time, so they will only be sketched. To simplify notations, we write:

$$A^1 = \left(A^{11}, A^{12}\right), \qquad A^2 = \left(A^{21}, A^{22}\right).$$

Lemma 3.3.2 *Let $H^1 \in \mathbb{R}^{q\times n_1}$, $H^2 \in \mathbb{R}^{q\times n_2}$, with $q \geq 1$, $1 \leq i \leq q$, and $x, z \in \mathbb{R}^n$.*

(a) – *For continuous time MN's, if $(H^1A^1x_u(t))_i = (H^1A^1z_u(t))_i$ for all $u \in \mathbb{R}^m$ and all $t \in [0, \epsilon_u)$ then we have:*

$$\begin{array}{l}(H^1A^{12}A^2x)_i = (H^1A^{12}A^2z)_i,\\ (H^1A^{11})_{ij} = 0 \quad \text{for all } j \text{ such that} \quad (A^1x)_j \neq (A^1z)_j.\end{array} \tag{43}$$

– *For discrete time MN's, if $(H^1A^1x^+(u))_i = (H^1A^1z^+(u))_i$ for all $u \in \mathbb{R}^m$, then the same conclusions hold.*

(b) – *For continuous time MN's, if $(H^2A^2x_u(t))_i = (H^2A^2z_u(t))_i$ for all $u \in \mathbb{R}^m$ and all $t \in [0, \epsilon_u)$ then we have:*

$$\begin{array}{l}(H^2A^{22}A^2x)_i = (H^2A^{22}A^2z)_i,\\ (H^2A^{21})_{ij} = 0 \quad \text{for all } j \text{ such that} \quad (A^1x)_j \neq (A^1z)_j.\end{array} \tag{44}$$

– *For discrete time MN's, if $(H^2A^2x^+(u))_i = (H^2A^2z^+(u))_i$ for all $u \in \mathbb{R}^m$, then the same conclusions hold.*

Proof. (a) Suppose we are in the continuous time case, and fix $u \in \mathbb{R}^m$. If $(H^1A^1x_u(t))_i = (H^1A^1z_u(t))_i$ for all $t \in [0, \epsilon_u)$ then $(H^1A^1\dot{x}_u(t))_i|_{t=0} = (H^1A^1\dot{z}_u(t))_i|_{t=0}$. This equation reads:

$$\begin{array}{l}\sum_{j=1}^p (H^1A^{11})_{ij}\sigma((A^1x)_j + (B^1u)_j) + (H^1A^{12}A^2x)_i = \\ \sum_{j=1}^p (H^1A^{11})_{ij}\sigma((A^1z)_j + (B^1u)_j) + (H^1A^{12}A^2z)_i.\end{array} \tag{45}$$

Since (45) holds for all $u \in \mathbb{R}^m$, both equalities in (43) follow directly by applying Lemma 3.3.1.

The proof for the discrete time case is the same since the assumption $(H^1A^1x^+(u))_i = (H^1A^1z^+(u))_i$, for all $u \in \mathbb{R}^m$, implies directly equations (45).

(b) This statement is proved similarly. In fact, by the same arguments as in (a), one easily sees that, for both continuous and discrete time dynamics:

$$\begin{array}{l}\sum_{j=1}^p (H^2A^{21})_{ij}\sigma((A^1x)_j + (B^1u)_j) + (H^2A^{22}A^2x)_i = \\ \sum_{j=1}^p (H^2A^{21})_{ij}\sigma((A^1z)_j + (B^1u)_j) + (H^2A^{22}A^2z)_i,\end{array}$$

for all $u \in \mathbb{R}^m$. So, again, to conclude it is sufficient to use Lemma 3.3.1. ∎

Lemma 3.3.3 *If x, $z \in \mathbb{R}^n$ are indistinguishable, then $C^2A^2x = C^2A^2z$ and $(A^1x)_i = (A^1z)_i$ for all $i \notin I_{C^1}$*

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

$$C^1\vec{\sigma}(A^1x + B^1u) + C^2A^2x + C^2B^2u = C^1\vec{\sigma}(A^1z + B^1u) + C^2A^2z + C^2B^2u$$

for all $u \in \mathbb{R}^m$. From Lemma 3.3.1 our conclusions follow. ∎

Lemma 3.3.4 *If x, $z \in \mathbb{R}^n$ are indistinguishable, 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 $(A^1x)_j \neq (A^1z)_j$.*

Proof. 1. We prove this statement by induction on $q \geq 0$. The case $q = 0$ is the first conclusion of Lemma 3.3.3. Assume that 1. holds for q and for all indistinguishable pairs. We deal first with continuous time models. Notice that since x and z are indistinguishable, then so are $x_u(t)$ and $z_u(t)$ for all $u \in \mathbb{R}^m$, and all $t \in [0, \epsilon_u)$. So, by the inductive assumption, we get

$$C^2(A^{22})^q A^2 x_u(t) = C^2(A^{22})^q A^2 z_u(t). \tag{46}$$

Now, by applying Lemma 3.3.2 (part (b), first equality, with $H^2 = C^2(A^{22})^q$), we get:

$$C^2(A^{22})^{q+1} A^2 x = C^2(A^{22})^{q+1} A^2 z$$

as desired. For discrete time MN's, the proof is the same, after replacing $x_u(t)$ and $z_u(t)$, with $x^+(u)$ and $z^+(u)$.
2. Again we apply Lemma 3.3.2 (part (b), second equality) to equation (46), to conclude

$$(C^2(A^{22})^q A^{21})_{ij} = 0 \text{ for all } j \text{ such that } (A^1x)_j \neq (A^1x)_j.$$

Similarly for discrete time dynamics. ∎

Lemma 3.3.5 *Let $1 \leq i \leq n_1$, and x, $z \in \mathbb{R}^n$. Assume that for any x', $z' \in \mathbb{R}^n$ such that $(x, z) \rightsquigarrow (x', z')$, we have:*

$$(A^1x')_i = (A^1z')_i. \tag{47}$$

Then, for all $q \geq 0$, we have:

(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) By induction on q. Fix any ξ, ζ such that $(x, z) \rightsquigarrow (\xi, \zeta)$. If $q = 0$ equation (47) says that, for continuous time MN's, $(A^1\xi_u(t))_i = (A^1\zeta_u(t))_i$, for all $u \in \mathbb{R}^m$ and for all $t \in [0, \epsilon_u)$; for discrete time ones we get $(A^1\xi^+(u))_i = (A^1\zeta^+(u))_i$, again for all $u \in \mathbb{R}^m$. In any case, we can apply Lemma 3.3.2 (a) with $H^1 = I$, and we have:

$$(A^{12}A^2\xi)_i = (A^{12}A^2\zeta)_i,$$

as desired. Now, assume the statement true for q. Thus, in particular, we get:

$$\begin{array}{ll} (A^{12}(A^{22})^q A^2\xi_u(t))_i = (A^{12}(A^{22})^q A^2\zeta_u(t))_i & \text{for cont. time} \\ (A^{12}(A^{22})^q A^2\xi^+(u))_i = (A^{12}(A^{22})^q A^2\zeta^+(u))_i & \text{for discr. time,} \end{array} \tag{48}$$

and these equations hold for all $u \in \mathbb{R}^m$, and for all $t \in [0, \epsilon_u)$. Again, the inductive step easily follows by applying Lemma 3.3.2 (b) with $H^2 = A^{12}(A^{22})^q$.
(b) Fix any j such that there exists a pair ξ, ζ with $(x, z) \rightsquigarrow (\xi, \zeta)$, and $(A^1\xi)_j \neq (A^1\zeta)_j$. Notice that for this pair ξ, ζ equation (48) hold. Thus, we apply again Lemma 3.3.2 (b) with $H^2 = A^{12}(A^{22})^q$, and we get:

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

as desired. ∎

Next Proposition proves the necessity of the observability conditions stated in Theorem 3.2.5.

Proposition 3.3.6 *Let Σ be an admissible MN, and $V_1 \subseteq \mathbb{R}^{n_1}$, $V_2 \subseteq \mathbb{R}^{n_2}$ be the two subspaces defined in Proposition 3.2.4. If $x, z \in \mathbb{R}^n$ are indistinguishable then $x - z \in \ker C$, and $A(x-z) \in V_1 \oplus V_2$.*

Proof. Since indistinguishability of x, z obviously implies $x - z \in \ker C$, we need only to prove that $A^1(x-z) \in V_1$ and $A^2(x-z) \in V_2$. Let:

$$J := \{\, i \,|\, \exists\, (x', z'), \ \text{ with } \ (x,z) \rightsquigarrow (x', z'), \ \text{ and } \ (A^1x')_i \neq (A^1z')_i \,\}.$$

Then we define:

$$\hat{V}_1 := \ \text{span} \ \{\, e_i \,|\, i \in J \,\},$$

where e_i are the vectors of the canonical base in $\mathbb{R}^n$, and

$$\hat{V}_2 := \ \text{the largest } A^{22}\text{-invariant subspace, contained in } \ker C^2, \text{ such that } A^{12}\hat{V}_2 \subseteq \hat{V}_1.$$

Notice that $\hat{V}_1$ is a coordinate subspace. Next we will establish that:

(i) $\hat{V}_1$ is A^{11}-invariant, $\hat{V}_1 \subseteq \ker C^1$, and $A^{21}\hat{V}_1 \subseteq \hat{V}_2$.

(ii) $\tilde{V}_2 = \{\, (A^{22})^q(A^2x' - A^2z') \,|\, q \geq 0,\ (x,z) \rightsquigarrow (x', z')\} \subseteq \hat{V}_2$.

Our conclusion follows from these statements since:

- $A^1(x-z) \in \hat{V}_1$, by definition of J;
- $A^2(x-z) \in \hat{V}_2$, by (ii);
- $\hat{V}_1 \subseteq V_1$, and $\hat{V}_2 \subseteq V_2$, by (i).

Now, we prove (i), and (ii).

(i)

- A^{11}-*invariance.*

  Since $\hat{V}_1$ is a coordinate subspace, proving A^{11} invariance is equivalent to see that $A^{11}_{ij} = 0$ for all i, j such that $j \in J$, and $i \notin J$. 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, for all $u \in \mathbb{R}^m$:

  $$\begin{array}{rcl} (A^1\xi_u(t))_i & = & (A^1\zeta_u(t))_i, \quad \forall t \in [0, \epsilon_u) \\ (A^1\xi^+(u))_i & = & (A^1\zeta^+(u))_i \end{array}$$

  in continuous and discrete time respectively. Now by applying Lemma 3.3.2 (a) with $H^1 = I$, 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.

- $\hat{V}_1 \subset \ker C^1$.

  If the pair x, z is indistinguishable, then so it is any pair x', z' such that $(x,z) \rightsquigarrow (x', z')$. Thus, the conclusion follows by observing that Lemma 3.3.3 implies $J \subset I_{C^1}$.

- $A^{21}\hat{V}_1 \subset \hat{V}_2$.

  It is sufficient to prove
$$\begin{cases} C^2(A^{22})^q A^{21} e_j = 0 \\ A^{12}(A^{22})^q A^{21} e_j \in \hat{V}_1 \end{cases}$$
  for all $j \in J$, $q \geq 0$ or, equivalently,
$$\begin{cases} \left(C^2(A^{22})^q A^{21}\right)_{ij} = 0 & \forall i,\ \forall j \in J, \\ \left(A^{12}(A^{22})^q A^{21}\right)_{ij} = 0 & \forall i \notin J, \forall j \in J. \end{cases} \tag{49}$$
  The first equality in (49) is easily obtained by applying part (b) of Lemma 3.3.4. The second one follows from part (b) of Lemma 3.3.5 after having observed that

  I) if $i \notin J$ then $(A^1 x')_i = (A^1 z')_i$ for all (x', z') such that $(x, z) \rightsquigarrow (x', z')$,

  II) if $j \in J$ then there exists (x', z') such that $(x, z) \rightsquigarrow (x', z')$ and $(A^1 x')_j \neq (A^1 z')_j$.

(ii). $\tilde{V}_2$ is by definition A^{22}-invariant. Thus to prove that $\tilde{V}_2 \subset \hat{V}_2$ we need to show that $\tilde{V}_2 \subset \ker C^2$ and $A^{12}\tilde{V}_2 \subset \hat{V}_1$. This amounts to establish the following identities:
$$C^2(A^{22})^q A^2 x' = C^2(A^{22})^q A^2 z' \ \ \forall (x', z') \ \text{ such that } \ (x, z) \rightsquigarrow (x', z') \tag{50}$$
$$A^{12}(A^{22})^q A^2 (x' - z') \in \hat{V}_1 \ \ \forall (x', z') \ \text{ such that } \ (x, z) \rightsquigarrow (x', z'). \tag{51}$$
Since $(x, z) \rightsquigarrow (x', z')$ implies that also the pair x', z' is indistinguishable, (50) is just part 1. of Lemma 3.3.4. Moreover (51) is equivalent to
$$(A^{12}(A^{22})^q A^2 x')_i = (A^{12}(A^{22})^q A^2 z')_i \ \ \forall i \notin J$$
which follows from part (a) of Lemma 3.3.5. ∎

Now we prove Theorem 3.2.5.

Proof of Theorem 3.2.5. Necessity is proved in Proposition 3.3.6, thus we only need to prove sufficiency.

Assume first that we are dealing with discrete time MN. We will prove that if $x, z \in \mathbb{R}^n$ satisfy the indistinguishability conditions of Theorem (3.2.5), then, for all $u \in \mathbb{R}^m$, also $x^+(u), z^+(u)$ satisfy the same conditions. This fact will clearly imply that x, z are indistinguishable.

First notice that the following implications hold:
$$\begin{array}{l} A^1(x - z) \in V_1 \Rightarrow x_1^+(u) - z_1^+(u) \in V_1 \\ A^2(x - z) \in V_2 \Rightarrow x_2^+(u) - z_2^+(u) \in V_2. \end{array} \tag{52}$$
Both implications are easily proved using the properties of V_1, and V_2, and the fact that V_1 is a coordinate subspace, and so if $\alpha \in V_1$ then also $\vec{\sigma}(\alpha) \in V_1$. Since $V_1 \subseteq \ker C^1$, and $V_2 \subseteq \ker C^2$, (52) yields $x^+(u) - z^+(u) \in \ker C$. Moreover, A^{11}-invariance of V_1 and the fact that $A^{12}V_2 \subseteq V_1$, implies:
$$A^1(x^+(u) - z^+(u)) = A^{11}(x_1^+(u) - z_1^+(u)) + A^{12}(x_2^+(u) - z_2^+(u)) \in V_1,$$
while, A^{22}-invariance of V_2 and the fact that $A^{21}V_1 \subseteq V_2$, gives:
$$A^2(x^+(u) - z^+(u)) = A^{12}(x_1^+(u) - z_1^+(u)) + A^{22}(x_2^+(u) - z_2^+(u)) \in V_2.$$

Thus $A(x^+(u) - z^+(u)) \in V_1 \oplus V_2$, as desired.

Now we deal with continuous time MN. For a fixed but arbitrary input signal $(u(t))_{t\geq 0}$, let $x(t)$, $z(t)$ denote the corresponding solutions of (34), associated to initial conditions $x(0)$, $z(0)$. The pair $(x(t), z(t))$ solves the differential equation in $\mathbb{R}^{2n}$

$$\begin{pmatrix} \dot{x} \\ \dot{z} \end{pmatrix} = F(x, z) \tag{53}$$

where

$$F(x,z) \;=\; \begin{pmatrix} \vec{\sigma}(A^1 x + B^1 u) \\ A^2 x + B^2 u \\ \vec{\sigma}(A^1 z + B^1 u) \\ A^2 z + B^2 u \end{pmatrix}.$$

Let $Z = \{(x, z) \in \mathbb{R}^{2n} : x - z \in \ker C, A^1(x - z) \in V_1, A^2(x - z) \in V_2\}$. 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 (53), i.e. if $(x(0), z(0)) \in Z$ then $(x(t), z(t)) \in Z$. Being $(u(t))_{t\geq 0}$ arbitrary and, since $(x, z) \in Z \Rightarrow x - z \in \ker C$, the proof is easily completed.

3.4 Identifiability and minimality

As done for RNN's (see Section 2.4), we identify a MN with the quadruple $\Sigma_{n_1,n_2} = (\sigma, A, B, C)$, where n_1, and n_2 are the dimensions of the nonlinear and linear block respectively. As usual we let $n = n_1 + n_2$ and:

$$A = \begin{bmatrix} A^{11} & A^{12} \\ A^{21} & A^{22} \end{bmatrix}, \quad B = \begin{bmatrix} B^1 \\ B^2 \end{bmatrix}, \quad C = \begin{bmatrix} C^1, C^2 \end{bmatrix}.$$

Moreover, with (Σ_{n_1,n_2}, x_0), $x_0 \in \mathbb{R}^n$, we will denote the MN $\Sigma_n = (\sigma, A, B, C)$ together with the initial state x_0; and we will call the pair (Σ_{n_1,n_2}, x_0) an *initialized MN*.

Our goal is to determine the group of symmetries which leaves the i/o behavior of the initialized MN unchanged. In Section 2.4, we have seen that for an admissible RNN this group of symmetries is finite, and it coincides with the group G_n generated by permutations and sign changes. For MN's, since one block of the system behaves linearly, one expects that this group of symmetries will not be finite. Let $\mathcal{G}_n$ be the following set of invertible matrices:

$$\mathcal{G}_n \;=\; \left\{ T \in \mathbb{R}^{n\times n} \;\middle|\; T = \begin{pmatrix} T_1 & 0 \\ 0 & T_2 \end{pmatrix} \quad \text{where} \quad \begin{matrix} T_1 \in G_{n_1} \\ T_2 \in GL(n_2) \end{matrix} \right\}$$

It is easy to see that if $T \in \mathcal{G}_n$, and we let

$$\tilde{A} = T^{-1}AT, \quad \tilde{B} = T^{-1}B, \quad \tilde{C} = CT, \quad \text{and} \quad \tilde{x}_0 = T^{-1}x_0,$$

then the two initialized MN's $(\Sigma_{n_1,n_2} = (\sigma, A, B, C), x_0)$, and $(\tilde{\Sigma}_{n_1,n_2} = (\sigma, \tilde{A}, \tilde{B}, \tilde{C}), \tilde{x}_0)$ have the same input/output behavior.

It is interesting to notice that this implication holds without any assumption on the two MN's, except the fact that the activation function σ is odd. Next we will see that if the MN's are admissible, observable, and satisfy a controllability assumption, then there are no other symmetries. We will write

$$(\Sigma_{n_1,n_2}, x_0) \sim (\tilde{\Sigma}_{\tilde{n}_1,\tilde{n}_2}, \tilde{x}_0)$$

to denote that the two initialized MN's (Σ_{n_1,n_2}, x_0), and $(\tilde{\Sigma}_{\tilde{n}_1,\tilde{n}_2}, \tilde{x}_0)$ are input/output equivalent, i.e. $\lambda_{\Sigma_{n_1,n_2},x_0} = \lambda_{\tilde{\Sigma}_{\tilde{n}_1,\tilde{n}_2},\tilde{x}_0}$ (where $\lambda_{\Sigma_{n_1,n_2},x_0}$ represents the i/o map, see Section 2.3).

Definition 3.4.1 Let $\Sigma_{n_1,n_2} = (\sigma, A, B, C)$ and $\tilde{\Sigma}_{\tilde{n}_1,n_2} = (\sigma, \tilde{A}, \tilde{B}, \tilde{C})$ be two MN's. The two initialized MN's (Σ_{n_1,n_2}, x_0) and $(\tilde{\Sigma}_{\tilde{n}_1,n_2}, \tilde{x}_0)$ are said to be *equivalent* if $n_1 = \tilde{n}_1$, $n_2 = \tilde{n}_2$, and there exists $T \in \mathcal{G}_n$ such that $\tilde{A} = T^{-1}AT$, $\tilde{B} = T^{-1}B$, $\tilde{C} = CT$, and $\tilde{x}_0 = T^{-1}x_0$.

Thus, from the above discussion, if two initialized MN's are equivalent, then they are also i/o equivalent.

Definition 3.4.2 An admissible MN, $\Sigma_{n_1,n_2} = (\sigma, A, B, C)$ is said to be *identifiable* if the following condition holds: for every initial state $x_0 \in \mathbb{R}^n$, and for every admissible initialized MN $(\tilde{\Sigma}_{\tilde{n}_1,\tilde{n}_2} = (\sigma, \tilde{A}, \tilde{B}, \tilde{C}), \tilde{x}_0)$ such that (Σ_{n_1,n_2}, x_0) and $(\tilde{\Sigma}_{\tilde{n}_1,\tilde{n}_2}, \tilde{x}_0)$ are i/o equivalent, then either $\tilde{n} > n$ or (Σ_{n_1,n_2}, x_0) and $(\tilde{\Sigma}_{\tilde{n}_1,\tilde{n}_2}, \tilde{x}_0)$ are equivalent.

Note that these definitions correspond to Definitions 2.4.1 and 2.4.2 given for RNN's. Given two matrices $M \in \mathbb{R}^{p_1 \times p_1}$ and $N \in \mathbb{R}^{p_1 \times p_2}$, we say that the pair (M, N) is controllable if it satisfied the linear Kalman controllability condition (i.e. $\operatorname{rank}[N, MN, \ldots, M^{p_1-1}N] = p_1$).

Theorem 3.4.3 *Let $\Sigma_{n_1,n_2} = (\sigma, A, B, C)$ be an admissible MN. Then Σ_{n_1,n_2} is identifiable if and only if Σ_{n_1,n_2} is observable, and the pair of matrices $(A^{22}, (B^2, A^{21}))$ is controllable.*

The proof of this theorem is given in Section 3.5.

Remark 3.4.4 It is clear that if Σ_{n_1,n_2} is a RNN (i.e. $n_2 = 0$) then the previous Theorem becomes Theorem 2.4.4, which states the identifiability result for RNN. On the other hand for $n_1 = 0$, i.e. Σ_{n_1,n_2} linear, we recover the linear identifiability result stated in Theorem 2.4.3.

Definition 3.4.5 An admissible MN Σ_{n_1,n_2} is said to be *minimal* if for every initial state $x_0 \in \mathbb{R}^n$, and for every admissible initialized MN, $(\tilde{\Sigma}_{\tilde{n}_1,\tilde{n}_2}, \tilde{x}_0)$ such that (Σ_{n_1,n_2}, x_0) and $(\tilde{\Sigma}_{\tilde{n}_1,\tilde{n}_2}, \tilde{x}_0)$ are i/o equivalent, then it must be $\tilde{n} \geq n$.

Remark 3.4.6 It would be reasonable, in Definition 3.4.2, to replace the inequality "$\tilde{n} > n$" with the statement "$\tilde{n}_1 \geq n_1$, $\tilde{n}_2 \geq n_2$, where at least one inequality is strict". Analogously, in Definition 3.4.5, "$\tilde{n} \geq n$" could be replaced by "$\tilde{n}_1 \geq n_1$, $\tilde{n}_2 \geq n_2$". These modified definitions are not logically equivalent to the ones we gave. However, they are indeed equivalent, and this fact will be a byproduct of the proof of Theorem 3.4.3 and Proposition 3.4.7. This means, for instance, that if Σ_{n_1,n_2} has the same i/o behavior of $\tilde{\Sigma}_{\tilde{n}_1,\tilde{n}_2}$ and $n \leq \tilde{n}$, then, necessarily, $n_1 \leq \tilde{n}_1$ and $n_2 \leq \tilde{n}_2$ (of course, we always assume the systems to be admissible).

As observed in Section 2.4 for RNN's, it is obvious that if an admissible MN Σ_{n_1,n_2} is identifiable then it is also minimal, while the converse implication may be false, as shown in Example 2.4.7.

Next Proposition (whose proof will be given in Section 3.5) presents necessary and sufficient conditions for minimality.

Proposition 3.4.7 *Let $\Sigma_{n_1,n_2} = (\sigma, A, B, C)$ be an admissible MN, and let $V_1 \subseteq \mathbb{R}^{n_1}$ and $V_2 \subseteq \mathbb{R}^{n_2}$ be the two subspaces defined in Proposition 3.2.4. Then Σ_{n_1,n_2} is minimal if and only if*

$$V_1 \oplus V_2 = 0 \quad \text{and the pair} \quad \left(A^{22}, \left(B^2, A^{21}\right)\right) \quad \text{is controllable .} \tag{54}$$

Remark 3.4.8 Proposition 3.4.7 will be proved later; however the necessity of the conditions in (54) is not difficult to establish, as shown next.

(a) Assume that $V_1 \oplus V_2 \neq 0$, and let $r_1 = \dim V_1$, $r_2 = \dim V_2$, and $q_i = n_i - r_i$ $(i = 1, 2)$. Without loss of generality, we may assume that that $V_1 =$ span $\{e_1, \ldots, e_{r_1}\}$, where $e_j \in \mathbb{R}^{n_1}$ are the elements of the canonical basis. Thus, we must have (in what follows, we specify only the dimensions of some particular matrices, meaning that all the others have the appropriate dimensions):

$$A^{11}V_1 \subseteq V_1 \;\Rightarrow\; A^{11} = \begin{pmatrix} H^{11} & H^{12} \\ 0 & H^{22} \end{pmatrix}, \quad \text{with } \; H^{22} \in \mathbb{R}^{q_1 \times q_1},$$

$$V_1 \subseteq \ker C^1 \;\Rightarrow\; C^1 = \left(0,\, D^1\right), \quad \text{with } \; D^1 \in \mathbb{R}^{p \times q_1}.$$

Using the linear theory, we also know that there exists a matrix $T_2 \in GL(n_2)$ such that:

$$T_2^{-1} A^{22} T_2 = \begin{pmatrix} L^{11} & 0 \\ L^{21} & L^{22} \end{pmatrix} \quad C^2 T_2 = \left(D^2,\, 0\right),$$

where $L^{11} \in \mathbb{R}^{q_2 \times q_2}$, and $D^2 \in \mathbb{R}^{p \times q_2}$. Moreover, we have:

$$A^{21}V_1 \subseteq V_2 \;\Rightarrow\; T_2^{-1}A^{21} = \begin{pmatrix} 0 & K^{12} \\ K^{21} & K^{22} \end{pmatrix}, \quad \text{with } \; K^{22} \in \mathbb{R}^{q_2 \times q_1},$$

$$A^{12}V_2 \subseteq V_1 \;\Rightarrow\; A^{12}T_2 = \begin{pmatrix} M^{11} & 0 \\ M^{21} & M^{22} \end{pmatrix}, \quad \text{with } \; M^{22} \in \mathbb{R}^{q_1 \times q_2}.$$

Now, if we denote, for $i = 1, 2$, $B^i = (B^{i1}, B^{i2})^T$, with $B^{12} \in \mathbb{R}^{q_1 \times m}$ and $B^{21} \in \mathbb{R}^{q_2 \times m}$, we may rewrite the dynamics of Σ_{n_1,n_2} as:

$$\begin{cases} z_1^+ &= \vec{\sigma}\left(H^{11}z_1 + H^{12}z_2 + M^{11}w_1 + M^{12}w_2 + B^{11}u\right) \\ z_2^+ &= \vec{\sigma}\left(H^{22}z_2 + M^{11}w_1 + B^{12}u\right) \\ w_1^+ &= K^{12}z_2 + L^{11}w_1 + B^{21}u \\ w_2^+ &= K^{21}z_1 + K^{22}z_2 + L^{21}w_1 + L^{22}w_2 + B^{22}u \\ & \\ y &= D^1 z_2 + D^2 w_1 \end{cases}$$

where $z_2 \in \mathbb{R}^{q_1}$ and $w_1 \in \mathbb{R}^{q_2}$. Thus, since the two blocks of variables z_1 and w_2 do not effect the other two blocks and the output, it is clear that $(\Sigma_{n_1,n_2}, (z_1, z_2, w_1, w_2))$ is i/o equivalent to

$$\left(\Sigma_{q_1+q_2} = \left(\sigma, \begin{pmatrix} H^{22} & M^{11} \\ K^{22} & L^{11} \end{pmatrix}, \begin{pmatrix} B^{12} \\ B^{21} \end{pmatrix} \left(D^1, D^2\right)\right), (z_2, w_1)\right).$$

Thus we have performed a reduction in the dimension of the state space since $q_1 + q_2 < n$, and so Σ_{n_1,n_2} can not be minimal.

(b) Now, assume that the pair $(A^{22}, [B^2, A^{21}])$ is not controllable. Let p_1 be the rank of

$$\left(\left(B^2, A^{21}\right), A^{22}\left(B^2, A^{21}\right), \ldots, \left(A^{22}\right)^{n_2-1}\left(B^2, A^{21}\right)\right),$$

and $p_2 = n_2 - p_1$. Then, by the linear theory, there exists an invertible matrix $T^2 \in GL(n_2)$ such that:

$$(T^2)^{-1}A^{22}T^2 = \begin{pmatrix} H^1 & H^2 \\ 0 & H^3 \end{pmatrix}, \; (T^2)^{-1}A^{21} = \begin{pmatrix} K \\ 0 \end{pmatrix}, \; (T^2)^{-1}B^2 = \begin{pmatrix} M \\ 0 \end{pmatrix},$$

where $H^1 \in \mathbb{R}^{p_1 \times p_1}$, $H^2 \in \mathbb{R}^{p_1 \times p_2}$, $H^3 \in \mathbb{R}^{p_2 \times p_2}$, $K\mathbb{R}^{p_1 \times n_2}$, $M \in \mathbb{R}^{p_1 \times m}$. Now, let $A^{12} = (N^1, N^2)$, and $C^2 = (L^1, L^2)$, where the matrices N^1, and L^1 represent the first p_1 columns. Consider the MN $\tilde{\Sigma}_{n_1+p_1}$ given by the matrices:

$$\tilde{A} = \begin{pmatrix} A^{11} & N^1 \\ K & H^1 \end{pmatrix}, \; \tilde{B} = \begin{pmatrix} B^1 \\ M \end{pmatrix}, \; \tilde{C} = \left(C^1, L^1\right).$$

It is easy to see that $(\Sigma_{n_1,n_2}, 0) \sim (\tilde{\Sigma}_{n_1+p_1}, 0)$. Since $n_1 + p_1 < n$, again, Σ_{n_1,n_2} is not minimal.

3.5 Proofs on identifiability and minimality

Now we introduce some useful notations. In what follows, we assume that two initialized admissible MN's, (Σ_{n_1,n_2}, x_0) and $(\tilde{\Sigma}_{\tilde{n}_1,\tilde{n}_2}, \tilde{x}_0)$, both evolving either in continuous or discrete time, are given. We let $A^1 = (A^{11}, A^{12})$ and $A^2 = (A^{21}, A^{22})$, similarly for $\tilde{A}^1$ and $\tilde{A}^2$.

For discrete time models, for any $k \geq 1$ and any $u_1, \ldots, u_k \in \mathbb{R}^m$, we denote by $x_k[u_1, \ldots, u_k]$ (resp. $\tilde{x}_k[u_1, \ldots, u_k]$) the state we reach from x_0 (resp. $\tilde{x}_0$) using the control sequence $u_1, \ldots, u_k$. Moreover we let:

$$y_k[u_1, \ldots, u_k] = Cx_k[u_1, \ldots, u_k] \quad \text{and} \quad \tilde{y}_k[u_1, \ldots, u_k] = \tilde{C}\tilde{x}_k[u_1, \ldots, u_k].$$

For continuous time model, for any $w(\cdot) : [0, T_w] \to \mathbb{R}^m$, we let $x_w(t)$ and $\tilde{x}_w(t)$, for $t \in [0, \epsilon_w)$, be the two trajectories of Σ_{n_1,n_2} and $\tilde{\Sigma}_n$ starting at x_0 and $\tilde{x}_0$ respectively. Here, with $\epsilon_w > 0$, we denote the maximal constant such that both trajectories $x_w(t)$ and $\tilde{x}_w(t)$ are defined on the interval $[0, \epsilon_w)$. Again, by $y_w(t)$ and $\tilde{y}_w(t)$, for $t \in [0, \epsilon_w)$, we denote the two corresponding output signals (i.e. $y_w(t) = Cx_w(t)$, and $\tilde{y}_w(t) = \tilde{C}\tilde{x}_w(t)$).

For any vector $v \in \mathbb{R}^n$, with the superscript 1 (resp. 2) we denote the first n_1 (resp. the second n_2) block of coordinates. Similarly for $\tilde{v} \in \mathbb{R}^{\tilde{n}}$.

We will denote by $W \subseteq \mathbb{R}^n$ (resp. $\tilde{W} \subseteq \mathbb{R}^{\tilde{n}}$) the unobservable subspace of Σ_{n_1,n_2} (resp. $\tilde{\Sigma}_{\tilde{n}_1,\tilde{n}_2}$). Moreover with $V_1 \subseteq \mathbb{R}^{n_1}$ and $V_2 \subseteq \mathbb{R}^{n_2}$ (resp. $\tilde{V}_1 \subseteq \mathbb{R}^{\tilde{n}_1}$ and $\tilde{V}_2 \subseteq \mathbb{R}^{\tilde{n}_2}$) we will denote the two subspaces defined in Proposition 3.2.4 for Σ_{n_1,n_2} (resp. $\tilde{\Sigma}_{\tilde{n}_1,\tilde{n}_2}$). Recall that it holds $W = A^{-1}(V_1 \oplus V_2) \cap \ker C$.

Now we establish some preliminary results. First we state a technical fact, which gives the idea on how the admissibility assumption is going to be used in the proof of the identifiability result. For a given matrix D, I_D denotes the set of indexes i such that the i-th column of D is zero, and I_D^c its complement.

Lemma 3.5.1 *Assume that the following matrices and vectors are given: $B \in \mathbb{R}^{n \times m}$, $\tilde{B} \in \mathbb{R}^{\tilde{n} \times m}$, $C \in \mathbb{R}^{p \times n}$, $\tilde{C} \in \mathbb{R}^{p \times \tilde{n}}$, $D, \tilde{D} \in \mathbb{R}^{p \times m}$, $a \in \mathbb{R}^n$, $\tilde{a} \in \mathbb{R}^{\tilde{n}}$, and $e, \tilde{e} \in \mathbb{R}^p$. Moreover, assume that σ is any admissible activation function, that both B and $\tilde{B}$ are admissible control matrices, and that for all $u \in \mathbb{R}^m$ the following equality holds:*

$$C\vec{\sigma}\,(a + Bu) + Du + e \;=\; \tilde{C}\vec{\sigma}\left(\tilde{a} + \tilde{B}u\right) + \tilde{D}u + \tilde{e}. \tag{55}$$

Then we have:

(a) $e = \tilde{e}$.

(b) $D = \tilde{D}$.

(c) $|I_C^c| = |I_{\tilde{C}}^c|$, *and for any $l \in I_C^c$ there exists $\pi(l) \in I_{\tilde{C}}^c$ and $\beta(l) = \pm 1$, such that*

(c.1) $B_{li} = \beta(l)\tilde{B}_{\pi(l)i}$ *for all* $i \in \{1, \ldots, m\}$*;*

(c.2) $a_l = \beta(l)\tilde{a}_{\pi(l)}$.

(c.3) $C_{il} = \beta(l)\tilde{C}_{i\pi(l)}$ *for all* $i \in \{1, \ldots, p\}$.

Moreover the map π *is injective.*

Proof. Since B and $\tilde{B}$ are admissible control matrices, there exists $\bar{v} \in R^m$ such that:

$$\begin{array}{llll} b_i = (B\bar{v})_i \neq 0 & \forall\, i = 1, \ldots, n, & |b_i| \neq |b_j| & \text{for all } \; i \neq j; \\ \tilde{b}_i = (\tilde{B}\bar{v})_i \neq 0 & \forall\, i = 1, \ldots, n, & |\tilde{b}_i| \neq |\tilde{b}_j| & \text{for all } \; i \neq j. \end{array} \tag{56}$$

Letting $Z \subset \mathbb{R}^m$ to be the set of all vectors $v \in \mathbb{R}^m$ for which (56) holds, we have that Z is a dense subset of $\mathbb{R}^m$. Fix any $\bar{v} \in Z$, then by rewriting equation (55) for $u = x\bar{v}$, with $x \in \mathbb{R}$, and using the notations in (56), we get:

$$\sum_{j=1}^{n} C_{ij}\sigma\left(a_j + b_j x\right) + (D\bar{v})_i x + e_i \;=\; \sum_{j=1}^{\tilde{n}} \tilde{C}_{ij}\sigma\left(\tilde{a}_j + \tilde{b}_j x\right) + (\tilde{D}\bar{v})_i x + \tilde{e}_i, \tag{57}$$

for all $x \in \mathbb{R}$, and all $i \in 1, \ldots, p$. Fix any $i \in \{1, \ldots, p\}$. After possibly some cancellation, equation 57 is of the type:

$$\sum_{j=1}^{r} \gamma_{ij}\sigma\left(\alpha_j + \beta_j x\right) + \left((D\bar{v})_i - (\tilde{D}\bar{v})_i\right) x + (e_i - \tilde{e}_i) \;=\; 0, \quad \forall\, x \in \mathbb{R}.$$

Since σ is admissible, we immediately get:

$$\begin{aligned} e_i - \tilde{e}_i \;&=\; 0, \\ (D\bar{v})_i - (\tilde{D}\bar{v})_i \;&=\; 0. \end{aligned}$$

The first of these equation implies (a), the second implies (b), since it holds for every $\bar{v} \in Z$, and Z is dense.

Now, since (a) and (b) have been proved, we may rewrite (57), as:

$$\sum_{j=1}^{n} C_{ij}\sigma\left(a_j + b_j x\right) \;=\; \sum_{j=1}^{\tilde{n}} \tilde{C}_{ij}\sigma\left(\tilde{a}_j + \tilde{b}_j x\right). \tag{58}$$

Fix any $\bar{l} \in I_C^c$; then there exists $\bar{i} \in \{1, \ldots, p\}$, such that $C_{\bar{i}\bar{l}} \neq 0$. Consider equation (58) for this particular $\bar{i}$. The terms for which $C_{\bar{i}j} = 0$ or $\tilde{C}_{\bar{i}j} = 0$ will cancel (however not all of them will cancel since $C_{\bar{i}\bar{l}} \neq 0$); thus will we remain with an equation of the type:

$$\sum_{p=1}^{r} C_{\bar{i}j_p}\sigma\left(a_{j_p} + b_{j_p} x\right) - \sum_{p=1}^{\tilde{r}} \tilde{C}_{\bar{i}j_p}\sigma\left(\tilde{a}_{j_p} + \tilde{b}_{j_p} x\right) = 0, \; \forall\, x \in \mathbb{R}, \tag{59}$$

for some $r \leq n$, and some $\tilde{r} \leq \tilde{n}$. Since σ is admissible, and the b_{j_p}'s (resp. $\tilde{b}_{j_p}$) have different absolute values, there must exists two indexes j_{p_1}, $j_{\pi(p_1)}$, and $\beta(p_1) = \pm 1$ such that:

$$(a_{j_{p_1}}, b_{j_{p_1}}) \;=\; \beta(p_1)(\tilde{a}_{j_{\pi(p_1)}}, \tilde{b}_{j_{\pi(p_1)}}).$$

So we have:

$$\left(C_{\bar{i}j_{p_1}} - \beta(p_1)\tilde{C}_{\bar{i}j_{\pi(p_1)}}\right)\sigma(a_{j_{p_1}} + b_{j_{p_1}} x) + \sum_{p=1, p\neq p_1}^{r} C_{\bar{i}j_p}\sigma\left(a_{j_p} + b_{j_p} x\right) -$$

$$- \sum_{p=1 p \neq \pi(p_1)}^{\tilde{r}} \tilde{C}_{\bar{i} j_p} \sigma\left(\tilde{a}_{j_p} + \tilde{b}_{j_p} x\right) = 0.$$

Now, by repeating the same arguments, we will find another index j_{p_2}, a corresponding index $j_{\pi(p_2)}$ with $\pi(p_2) \neq \pi(p_1)$, and $\beta(p_2) = \pm 1$. Notice that necessarily $p_2 \neq p_1$, since, otherwise, $|\tilde{b}_{j_{\pi(p_1)}}| = |\tilde{b}_{j_{\pi(p_2)}}|$ would contraddict (56). Thus we will collect together two more terms. Going on with the same arguments, we must have that $r = \tilde{r}$, and after r steps, we end up with and equation of the type:

$$\sum_{p=1}^{r} \left(C_{\bar{i} j_p} - \beta(p) \tilde{C}_{\bar{i} j_{\pi(p)}}\right) \sigma\left(a_{j_p} + b_{j_p} x\right) \; = \; 0 \; \forall x \in \mathbb{R}.$$

Again, by the admissibility of σ, we also have $\left(C_{\bar{i} j_p} - \beta(p) \tilde{C}_{\bar{i} j_{\pi(p)}}\right) = 0$.

Thus, in particular, since $C_{\bar{i}\bar{l}} \neq 0$, we have shown that there exists $\pi(\bar{l})$, and $\beta(\bar{l})$ such that:

$$\begin{array}{rcl} (a_{\bar{l}}, b_{\bar{l}}) & = & \beta(\bar{l})(\tilde{a}_{\pi(\bar{l})}, \tilde{b}_{\pi(\bar{l})}) \\ C_{\bar{i}\bar{l}} & = & \beta(\bar{l}) \tilde{C}_{\bar{i}\pi(\bar{l})}. \end{array} \tag{60}$$

Notice that, if given this $\bar{l} \in I_C^c$, we would have chosen a different index $\hat{i}$ such that $C_{\hat{i}\bar{l}} \neq 0$, then we would have ended up with corresponding $\hat{\pi}(\bar{l})$ and $\hat{\beta}(\bar{l})$. However, since the $(\tilde{b}_j)$'s have all different absolute values, it must be, $\pi(\bar{l}) = \hat{\pi}(\bar{l})$, and $\beta(\bar{l}) = \hat{\beta}(\bar{l})$ (see (60)). Thus we have shown that:

$$\forall\, l \in I_C^c, \;\; \exists\, \pi(l) \in I_{\tilde{C}}^c, \;\; \beta(l) = \pm 1, \;\; \text{such that} \;\; \left\{ \begin{array}{ll} (i) & (a_l, b_l) = \beta(l)(\tilde{a}_{\pi(l)}), \tilde{b}_{\pi(l)}), \\ (ii) & C_{il} = \beta(l) C_{i\pi(l)} \;\; \text{if} \;\; C_{il} \neq 0. \end{array} \right. \tag{61}$$

This implies $|I_C^c| \leq |I_{\tilde{C}}^c|$. By symmetry, we conclude that $|I_C^c| = |I_{\tilde{C}}^c|$. From (61) (i), we get directly that (c.2) holds. Again from (61) (i), we also have:

$$(B\bar{v})_l = \beta(l) \left(\tilde{B}\bar{v}\right)_{\pi(l)},$$

for all $\bar{v} \in Z$. Since Z is dense, this implies (c.1). Moreover (61) (ii) proves (c.3) for those C_{il} different from zero. On the other hand if $C_{il} = 0$ then, necessarily $\tilde{C}_{i\pi(l)}$ has to be zero also. Otherwise, one repeats the argument above exchanging C with $\tilde{C}$, and finds an index $\lambda(\pi(l))$ such that

$$|\tilde{C}_{i\pi(l)}| = |C_{i\lambda(\pi(l))}|, \quad |\tilde{b}_{\pi(l)}| = |b_{\lambda(\pi(l))}|.$$

In particular, $\lambda(\pi(l)) \neq l$, since $C_{il} = 0$ and $C_{i\lambda(\pi(l))} \neq 0$. But then $|b_{\lambda(\pi(l))}| = |\tilde{b}_{\pi(l)}| = |b_l|$, which is impossible since the b_i's have all different absolute values.

The injectivity of the map π is also a consequence of $|b_i| \neq |b_j|$ for all $i \neq j$. ∎

Lemma 3.5.2 *Let* (Σ_{n_1,n_2}, x_0) *and* $(\tilde{\Sigma}_{\tilde{n}_1,\tilde{n}_2}, \tilde{x}_0)$ *be two initialized MN's,* $H_1 \in \mathbb{R}^{q \times n_1}$*,* $H_2 \in \mathbb{R}^{q \times n_2}$*,* $\tilde{H}_1 \in \mathbb{R}^{q \times \tilde{n}_1}$*, and* $\tilde{H}_2 \in \mathbb{R}^{q \times \tilde{n}_2}$*.*

- *For continuous time models, if for all* $w : [0, T_w] \to \mathbb{R}^m$*, and for all* $t \in [0, \epsilon_w)$*, we have:*

$$H_1 x_w^1(t) + H_2 x_w^2(t) = \tilde{H}_1 \tilde{x}_w^1(t) + \tilde{H}_2 \tilde{x}_w^2(t), \tag{62}$$

then, for all $u \in \mathbb{R}^m$*:*

$$\begin{array}{l} H_1 \vec{\sigma}\left(A^1 x_w(t) + B^1 u\right) + H_2 A^2 x_w(t) + H_2 B^2 u = \\ \tilde{H}_1 \vec{\sigma}\left(\tilde{A}^1 \tilde{x}_w(t) + \tilde{B}^1 u\right) + \tilde{H}_2 A^2 \tilde{x}_w(t) + \tilde{H}_2 \tilde{B}^2 u. \end{array} \tag{63}$$

- *For discrete time models, if for all $r \geq 0$, and for all $u_1, \ldots, u_r \in \mathbb{R}^m$, we have:*

$$H_1 x^1[u_1, \ldots, u_r] + H_2 x^2[u_1, \ldots, u_r] = \tilde{H}_1 \tilde{x}^1[u_1, \ldots, u_r] + \tilde{H}_2 \tilde{x}^2[u_1, \ldots, u_r], \tag{64}$$

(when $r = 0$, it is meant that the previous equality holds for the two initial states) then, for all $u \in \mathbb{R}^m$:

$$\begin{aligned} &H_1 \vec{\sigma}\left(A^1 x[u_1, \ldots, u_r] + B^1 u\right) + H_2 A^2 x[u_1, \ldots, u_r] + H_2 B^2 u = \\ &\tilde{H}_1 \vec{\sigma}\left(\tilde{A}^1 \tilde{x}[u_1, \ldots, u_r] + \tilde{B}^1 u\right) + \tilde{H}_2 A^2 \tilde{x}[u_1, \ldots, u_r] + \tilde{H}_2 \tilde{B}^2 u. \end{aligned} \tag{65}$$

Proof. Being the discrete time case obvious, we just prove the continuous time statement. Fix any $\bar{t} \in [0, \epsilon_w)$. For any $u \in \mathbb{R}^m$, let $w_u : [0, \bar{t}+1] \to \mathbb{R}^m$ be the control map defined by $w_u(t) = w(t)$ if $t \in [0, \bar{t}]$, and $w_u = u$ if $t \in [\bar{t}, \bar{t}+1]$. Then the two trajectories $x_{w_u}(t)$ and $\tilde{x}_{w_u}(t)$ are defined on an interval of the type $[0, \bar{t}+\epsilon)$ and are differentiable for any $t \in (\bar{t}, \bar{t}+\epsilon)$. Since equation (62) holds for all $t \in (0, \bar{t}+\epsilon)$, we have, for all $t \in (\bar{t}, \bar{t}+\epsilon)$,

$$H_1 \dot{x}^1_w(t) + H_2 \dot{x}^2_w(t) = \tilde{H}_1 \dot{\tilde{x}}^1_w(t) + \tilde{H}_2 \dot{\tilde{x}}^2_w(t),$$

Now, by taking the limit as $t \to \bar{t}^+$, we get (63), as desired. ∎

Lemma 3.5.3 *If $(\Sigma_{n_1,n_2}, x_0) \sim (\tilde{\Sigma}_{\tilde{n}_1,\tilde{n}_2}, \tilde{x}_0)$, then, for all $l \geq 0$, we have:*

(a) $C^2(A^{22})^l B^2 = \tilde{C}^2(\tilde{A}^{22})^l \tilde{B}^2$;

(b) $|I^c_{C^2(A^{22})^l A^{21}}| = |I^c_{\tilde{C}^2(\tilde{A}^{22})^l \tilde{A}^{21}}|$, *and for all $i \in I^c_{C^2(A^{22})^l A^{21}}$ there exists $\pi(i) \in I^c_{\tilde{C}^2(\tilde{A}^{22})^l \tilde{A}^{21}}$ and $\beta(i) = \pm 1$, such that:*

$$\left(C^2(A^{22})^l A^{21}\right)_{ji} = \beta(i)\left(\tilde{C}^2(\tilde{A}^{22})^l \tilde{A}^{21}\right)_{j\pi(i)},$$

for all $j \in \{1, \ldots, p\}$;

(c1) for continuous time models, for all $w : [0, T_w] \to \mathbb{R}^m$, and for all $t \in [0, \epsilon_w)$, we have:

$$C^2(A^{22})^l A^{21} x^1_w(t) + C^2(A^{22})^{l+1} x^2_w(t) = \tilde{C}^2(\tilde{A}^{22})^l \tilde{A}^{21} \tilde{x}^1_w(t) + \tilde{C}^2(\tilde{A}^{22})^{l+1} \tilde{x}^2_w(t);$$

(c2) for discrete time models, for all $r \geq 0$, for all $u_1, \ldots, u_r \in \mathbb{R}^m$, we have:

$$C^2(A^{22})^l A^{21} x^1[u_1, \ldots, u_r] + C^2(A^{22})^{l+1} x^2[u_1, \ldots, u_r] =$$

$$\tilde{C}^2(\tilde{A}^{22})^l \tilde{A}^{21} \tilde{x}^1[u_1, \ldots, u_r] + \tilde{C}^2(\tilde{A}^{22})^{l+1} \tilde{x}^2[u_1, \ldots, u_r].$$

(When $r = 0$, it is meant that the previous equality holds for the two initial states.)

Proof. Assume that we are dealing with continuous time MN's. We first prove, by induction on $l \geq 0$, statements (a) and (c1). Assume $l = 0$. Since $(\Sigma_{n_1,n_2}, x_0) \sim (\tilde{\Sigma}_{\tilde{n}_1,\tilde{n}_2}, \tilde{x}_0)$, then for all $w : [0, T_w] \to \mathbb{R}^m$, and for all $t \in [0, \epsilon_w)$, we have:

$$C^1 x^1_w(t) + C^2 x^2_w(t) = y_{w_u}(t) = \tilde{y}_{w_u}(t) = \tilde{C}^1 \tilde{x}^1_w(t) + \tilde{C}^2 \tilde{x}^2_w(t).$$

From Lemma 3.5.2 this implies:

$$\begin{aligned} &C^1 \vec{\sigma}\left(A^{11} x^1_w(t) + A^{12} x^2_w(t) + B^1 u\right) + C^2 A^{21} x^1_w(t) + C^2 A^{22} x^2_w(t) + C^2 B^2 u = \\ &\tilde{C}^1 \vec{\sigma}\left(\tilde{A}^{11} \tilde{x}^1_w(t) + \tilde{A}^{12} \tilde{x}^2_w(t) + \tilde{B}^1 u\right) + \tilde{C}^2 \tilde{A}^{21} \tilde{x}^1_w(t) + \tilde{C}^2 \tilde{A}^{22} \tilde{x}^2_w(t) + \tilde{C}^2 \tilde{B}^2 u \end{aligned} \tag{66}$$

Now, by (66) and by applying Lemma 3.5.1 (parts (a) and (b)) we immediately get conclusions (a) and (c1) when $l = 0$. The proof of the inductive step follows the same lines. By inductive assumption, we have:

$$C^2(A^{22})^l A^{21} x_w^1(t) + C^2(A^{22})^{l+1} x_w^2(t) = \tilde{C}^2(\tilde{A}^{22})^l \tilde{A}^{21} \tilde{x}_w^1(t) + \tilde{C}^2(\tilde{A}^{22})^{l+1} \tilde{x}_w^2(t),$$

for all $t \in [0, \epsilon_w)$. By applying again Lemma 3.5.2, we get:

$$\begin{array}{l} C^2(A^{22})^l A^{21} \vec{\sigma}\left(A^{11} x_w^1(t) + A^{12} x_w^2(t) + B^1 u\right) + C^2(A^{22})^{l+1}\left(A^2 x_w(t) + B^2 u\right) = \\ \tilde{C}^2(\tilde{A}^{22})^l \tilde{A}^{21} \vec{\sigma}\left(\tilde{A}^{11} \tilde{x}_w^1(t) + \tilde{A}^{12} \tilde{x}_w^2(t)\right) + + \tilde{C}^2(\tilde{A}^{22})^{l+1}\left(\tilde{A}^2 \tilde{x}_w(t) + \tilde{B}^2 u\right). \end{array}$$

By applying Lemma 3.5.1 (part (a) and (b)), the previous equation gives both (a) and (c1) for the case $l + 1$. Moreover, from part (c) of the same Lemma 3.5.1, also (b) must holds.

The proof for the discrete time dynamics is very similar and simpler. The idea is to establish, again by induction on $l \geq 0$, statement (a) and (c2) first. We only sketch the case $l = 0$. Since $(\Sigma_{n_1,n_2}, x_0) \sim (\tilde{\Sigma}_{\tilde{n}_1,\tilde{n}_2}, \tilde{x}_0)$, given any sequence $u_1, \ldots, u_r, u$, we must have:

$$y[u_1, \ldots, u_r, u] \;=\; \tilde{y}[u_1, \ldots, u_r, u],$$

which is the same as:

$$\begin{array}{l} C^1 \vec{\sigma}\left(A^1 x[u_1, \ldots, u_r] + B^1 u\right) + C^2 A^2 x[u_1, \ldots, u_r] + C^2 B^2 u = \\ \tilde{C}^1 \vec{\sigma}\left(\tilde{A}^1 \tilde{x}[u_1, \ldots, u_r] + \tilde{B}^1 u\right) + \tilde{C}^2 \tilde{A}^2 \tilde{x}[u_1, \ldots, u_r] + \tilde{C}^2 \tilde{B}^2 u \end{array}$$

Again, since the previous holds for every $u \in \mathbb{R}^m$, Lemma 3.5.1 gives conclusion (a) and (c2) for the case $l = 0$. ∎

We omit the proof of the next Lemma, whose conclusions may be established by induction, using the same arguments as in the previous proof.

Lemma 3.5.4 *Assume that $(\Sigma_{n_1,n_2}, x_0) \sim (\tilde{\Sigma}_{\tilde{n}_1,\tilde{n}_2}, \tilde{x}_0)$, and that there exists $i \in \{1, \ldots, n_1\}$, $\beta(i) = \pm 1$, and $\pi(i) \in \{1, \ldots, \tilde{n}_1\}$ such that:*

- *for the continuous time dynamics, for all $w(\cdot) \in [0, T_w] \to \mathbb{R}^m$, for all $t \in [0, \epsilon_w)$ the following holds*

$$\left[A^{11} x_w^1(t) + A^{12} x_w^2(t)\right]_i = \beta(i)\left[\tilde{A}^{11} \tilde{x}_w^1(t) + \tilde{A}^{12} \tilde{x}_w^2(t)\right]_{\pi(i)},$$

- *for discrete time dynamics, for all $r \geq 1$, for all $u_1, \ldots, u_r \in \mathbb{R}^m$, the following holds:*

$$\left[A^{11} x^1[u_1, \ldots, u_r] + A^{12} x^2[u_1, \ldots, u_r]\right]_i = \beta(i)\left[\tilde{A}^{11} \tilde{x}^1[u_1, \ldots, u_r] + \tilde{A}^{12} \tilde{x}^2[u_1, \ldots, u_r]\right]_{\pi(i)}.$$

Then, for all $l \geq 0$, we have:

(a) $\left[A^{12}(A^{22})^l B^2\right]_{ij} = \beta(i)\left[\tilde{A}^{12}(\tilde{A}^{22})^l \tilde{B}^2\right]_{\pi(i)j}$, $j \in \{1, \ldots, m\}$,

(b1) for continuous time dynamics, for all $w(\cdot) \in [0, T_w] \to \mathbb{R}^m$, for all $t \in [0, \epsilon_w)$:
$\left[A^{12}(A^{22})^l A^{21} x_w^1(t) + A^{12}(A^{22})^{l+1} x_w^2(t)\right]_i =$
$\beta(i)\left[\tilde{A}^{12}(\tilde{A}^{22})^l \tilde{A}^{21} \tilde{x}_w^1(t) + \tilde{A}^{12}(\tilde{A}^{22})^{l+1} \tilde{x}_w^2(t)\right]_{\pi(i)}$;

(b2) for discrete time dynamics, for all $r \geq 1$*, for all* $u_1, \ldots, u_r \in \mathbb{R}^m$*:*
$\Big[A^{12}(A^{22})^l A^{21} x^1[u_1, \ldots, u_r] + A^{12}(A^{22})^{l+1} x^2[u_1, \ldots, u_r]\Big]_i =$
$\beta(i) \Big[\tilde{A}^{12}(\tilde{A}^{22})^l \tilde{A}^{21} \tilde{x}^1[u_1, \ldots, u_r] + \tilde{A}^{12}(\tilde{A}^{22})^{l+1} \tilde{x}^2[u_1, \ldots, u_r]\Big]_{\pi(i)}.$

Lemma 3.5.5 *If* $(\Sigma_{n_1,n_2}, x_0) \sim (\tilde{\Sigma}_{\tilde{n}_1,\tilde{n}_2}, \tilde{x}_0)$ *and* $V_1 = 0$*, then* $n_1 \leq \tilde{n}_1$*, and for all* $i \in \{1, \ldots, n_1\}$ *there exists* $\pi(i) \in \{1, \ldots, \tilde{n}_1\}$ *and* $\beta(i) = \pm 1$ *such that:*

(a) $B^1_{ij} = \beta(i) \tilde{B}^1_{\pi(i)j}$ *for all* $j \in 1, \ldots, m$*;*

(b) $C^1_{ji} = \beta(i) \tilde{C}^1_{j\pi(i)}$ *for all* $j \in 1, \ldots, p$*;*

(c1) for continuous time dynamics, for all $w(\cdot) \in [0, T_w] \to \mathbb{R}^m$*, for all* $t \in [0, \epsilon_w)$ *the following holds*

$$\Big[A^{11} x^1_w(t) + A^{12} x^2_w(t)\Big]_i = \beta(i) \Big[\tilde{A}^{11} \tilde{x}^1_w(t) + \tilde{A}^{12} \tilde{x}^2_w(t)\Big]_{\pi(i)};$$

(c2) for discrete time dynamics, for all $r \geq 1$*, for all* $u_1, \ldots, u_r \in \mathbb{R}^m$*, the following holds:*

$$\Big[A^{11} x^1[u_1, \ldots, u_r] + A^{12} x^2[u_1, \ldots, u_r]\Big]_i =$$

$$\beta(i) \Big[\tilde{A}^{11} \tilde{x}^1[u_1, \ldots, u_r] + \tilde{A}^{12} \tilde{x}^2[u_1, \ldots, u_r]\Big]_{\pi(i)}.$$

Moreover the map π *is injective.*

Proof. Assume that we are dealing with continuous time MN's (the proof for the discrete time case is very similar and thus omitted). Since $V_1 = 0$, then, letting J_d to denote the set of indexes defined in Section 3.2, we have that for any $i \in \{1, \ldots, n_1\}$ there exists $d \geq 1$ such that $i \in J_d$. We first prove (a) and (c1) by induction on the first index $d \geq 1$ such that $i \in J_d$.

Assume that $d = 1$, i.e. $i \in J_1$. Thus, by definition, there exists $1 \leq l \leq p$ such that $C^1_{li} \neq 0$. Since $(\Sigma_{n_1,n_2}, x_0) \sim (\tilde{\Sigma}_{\tilde{n}_1,\tilde{n}_2}, \tilde{x}_0)$, then for all $w : [0, T_w] \to \mathbb{R}^m$, and for all $t \in [0, \epsilon_w)$, we have:

$$C^1 x^1_w(t) + C^2 x^2_w(t) = y_{w_u}(t) = \tilde{y}_{w_u}(t) = \tilde{C}^1 \tilde{x}^1_w(t) + \tilde{C}^2 \tilde{x}^2_w(t),$$

from Lemma 3.5.2 this implies:

$$C^1 \vec{\sigma}\left(A^1 x_w(t) + B^1 u\right) + C^2 A^2 x_w(t) + C^2 B^2 u = \tilde{C}^1 \vec{\sigma}\left(\tilde{A}^1 \tilde{x}_w(t) + \tilde{B}^1 u\right) + \tilde{C}^2 \tilde{A}^2 \tilde{x}_w(t) + \tilde{C}^2 \tilde{B}^2 u. \tag{67}$$

Since $C^1_{li} \neq 0$, we have $i \in I^c_{C^1}$, thus, by Lemma 3.5.1, we know that there exist $\pi(i) \in \{1, \ldots, \tilde{n}_1\}$, and $\beta(i) = \pm 1$, such that:

$$B^1_{ij} = \beta(i) \tilde{B}^1_{\pi(i)j},$$

for all $j \in 1, \ldots, m$, and, for all $w(\cdot) \in [0, T_w] \to \mathbb{R}^m$, for all $t \in [0, \epsilon_w)$:

$$\Big[A^{11} x^1_w(t) + A^{12} x^2_w(t)\Big]_i = \beta(i) \Big[\tilde{A}^{11} \tilde{x}^1_w(t) + \tilde{A}^{12} \tilde{x}^2_w(t)\Big]_{\pi(i)}.$$

Now suppose that $i \in J_{d+1}$, for $d > 0$. If $i \in J_1$, then there is nothing to prove; otherwise we are in one of the following three cases:

1. there exists $j \in J_d$, and $A^{11}_{ji} \neq 0$;
2. there exists $1 \leq j \leq p$ and $0 \leq l \leq d-1$ such that $\left(C^2(A^{22})^l A^{21}\right)_{ji} \neq 0$;
3. there exists $j \in J_s$, for $1 \leq s \leq d-1$ such that $\left(A^{12}(A^{22})^{d-s-1}A^{21}\right)_{ji} \neq 0$.

We will prove the three cases separately.
1. Since $j \in J_d$, then, in particular, (c1) holds for this j. Thus we have, for all $w(\cdot) \in [0, T_w] \to \mathbb{R}^m$, for all $t \in [0, \epsilon_w)$:

$$\left[A^{11}x^1_w(t) + A^{12}x^2_w(t)\right]_j = \beta(j)\left[\tilde{A}^{11}\tilde{x}^1_w(t) + \tilde{A}^{12}\tilde{x}^2_w(t)\right]_{\pi(j)}.$$

Now, by applying Lemma 3.5.2 to this equation, we get, that for all $u \in \mathbb{R}^m$:

$$\begin{array}{l}\left[A^{11}\vec{\sigma}\left(A^1x_w(t) + B^1u\right) + A^{12}A^2x_w(t) + A^{12}B^2u\right]_j = \\ \beta(j)\left[\tilde{A}^{11}\vec{\sigma}\left(\tilde{A}^1x_w(t) + \tilde{B}^1u\right) + \tilde{A}^{12}\tilde{A}^2x_w(t) + \tilde{A}^{12}\tilde{B}^2u\right]_{\pi(j)}.\end{array}$$

Now, since $A^{11}_{ji} \neq 0$, we have $i \in I^c_{A^{11}}$; by applying Lemma 3.5.1 to the previous equation, we get that there exists $\pi(i) \in \{1, \ldots, \tilde{n}_1\}$, and $\beta(i) = \pm 1$, such that (a) and (c1) holds for this index i.
2. From Lemma 3.5.3 (c1), we have that for all $w : [0, T_w] \to \mathbb{R}^m$, and for all $t \in [0, \epsilon_w)$, the following equality holds:

$$C^2(A^{22})^l A^{21}x^1_w(t) + C^2(A^{22})^{l+1}x^2_w(t) = \tilde{C}^2(\tilde{A}^{22})^l\tilde{A}^{21}\tilde{x}^1_w(t) + \tilde{C}^2(\tilde{A}^{22})^{l+1}\tilde{x}^2_w(t).$$

By applying Lemma 3.5.2 to the previous equation we get:

$$\begin{array}{l}C^2(A^{22})^l A^{21}\vec{\sigma}\left(A^1x_w(t) + B^1u\right) C^2(A^{22})^{l+1}\left(A^2x_w(t) + B^2u\right) = \\ \tilde{C}^2(\tilde{A}^{22})^l\tilde{A}^{21}\vec{\sigma}\left(\tilde{A}^1\tilde{x}_w(t) + \tilde{B}^1u\right) + \tilde{C}^2(\tilde{A}^{22})^{l+1}\left(\tilde{A}^2\tilde{x}_w(t) + \tilde{B}^2u\right).\end{array} \tag{68}$$

Since $\left(C^2(A^{22})^l A^{21}\right)_{ji} \neq 0$, again by applying Lemma 3.5.1, we get that there exist $\pi(i) \in \{1, \ldots, \tilde{n}_1\}$, and $\beta(i) = \pm 1$, such that (a) and (c1) holds for this index i.
3. Since $j \in J_s$, then, in particular, (c1) holds for this j. Thus we have, for all $w(\cdot) \in [0, T_w] \to \mathbb{R}^m$, for all $t \in [0, \epsilon_w)$:

$$\left[A^{11}x^1_w(t) + A^{12}x^2_w(t)\right]_j = \beta(j)\left[\tilde{A}^{11}\tilde{x}^1_w(t) + \tilde{A}^{12}\tilde{x}^2_w(t)\right]_{\pi(j)}.$$

Thus Lemma 3.5.4 applies, and from conclusion (b1) of this Lemma we get, for all $l \geq 0$:

$$\begin{array}{l}\left[A^{12}(A^{22})^l A^{21}x^1_w(t) + A^{12}(A^{22})^{l+1}x^2_w(t)\right]_j = \\ \beta(j)\left[\tilde{A}^{12}(\tilde{A}^{22})^l\tilde{A}^{21}\tilde{x}^1_w(t) + \tilde{A}^{12}(\tilde{A}^{22})^{l+1}\tilde{x}^2_w(t)\right]_{\pi(j)}.\end{array}$$

Now, letting $l = d - s - 1$, since $\left(A^{12}(A^{22})^l A^{21}\right)_{ji} \neq 0$, we may proceed as before and by applying first lemma 3.5.2, and then Lemma 3.5.1, (a) and (c1) follows also for this index i.

To conclude the proof, we need to establish (b). Using (a) and (c1), we may rewrite, for each $j \in \{1, \ldots, p\}$, equation (67) as:

$$\sum_{i=1}^{n_1}\left(C^1_{ji} - \beta(i)\tilde{C}^1_{j\pi(i)}\right)\vec{\sigma}\left((A^1x_w(t))_i + (B^1u)_i\right) + (C^2A^2x_w(t))_j =$$

$$\sum_{l\neq\pi(i)}\tilde{C}^1_{jl}\vec{\sigma}\left((\tilde{A}^1\tilde{x}_w(t))_l + (\tilde{B}^1u)_l\right) + (\tilde{C}^2\tilde{A}^2\tilde{x}_w(t))_j.$$

Since σ is admissible, the previous equality implies (b).

Notice that the injectivity of π is an obvious consequence of the injectivity of the corresponding map in Lemma 3.5.1. ∎

Given two matrices $M \in \mathbb{R}^{p_1\times p_1}$ and $N \in \mathbb{R}^{p_2\times p_1}$, we say that the pair (M, N) is observable if it satisfied the linear Kalman observability condition, i.e.:

$$\text{rank}\begin{pmatrix} N \\ NM \\ \vdots \\ NM^{p_1-1} \end{pmatrix} = p_1.$$

Lemma 3.5.6 *Let $\Sigma_{n_1,n_2} = (\sigma, A, B, C)$ be an admissible MN. If $V_2 = 0$ then the pair*

$$\left(\begin{pmatrix} C^2 \\ A^{12} \end{pmatrix}, A^{22}\right)$$

is observable.

Proof. Assume, by contradiction, that the above pair is not observable. Then there would exists a non-zero A^{22}-invariant subspace $W_2 \subseteq \mathbb{R}^{n_2}$ which is also contained in $\ker C^2$, and in $\ker A^{12}$. Thus, clearly $W_2 \subseteq V_2$ contradicting the assumption that $V_2 = 0$. ∎

Now we are ready to prove the identification result.

Proof of Theorem 3.4.3

Let $\Sigma_{n_1,n_2} = (\sigma, A, B, C)$ be an admissible MN.

Necessity We prove this part by contradiction.

First assume that Σ_{n_1,n_2} is not observable. Then there exist two different states x_1, $x_2 \in \mathbb{R}^n$ which give the same input/output behavior, thus the two initialized MN's (Σ_{n_1,n_2}, x_1), and (Σ_{n_1,n_2}, x_2) are input/output equivalent. Next we will see that they can't be equivalent. Assume that there is a matrix $T \in \mathcal{G}_n$ (see Definition 3.4.1) such that:

$$A = T^{-1}AT,\ B = T^{-1}B,\ C = CT, \ \text{ and } \ x_1 = T^{-1}x_2.$$

Then, since B^1 is an admissible control matrix, one has that T must be of the type:

$$T = \begin{pmatrix} I & 0 \\ 0 & T^2 \end{pmatrix}.$$

On the other hand, the matrix T^2, must satisfy:

$$A^{12} = T^2A^{12},\ C^2 = T^2C^2,\ A^{22} = T^{-1}A^{22}T,$$

which, by Lemma 3.5.6, implies that also $T^2 = I$. Thus $T = I$, and so we must have $x_1 = x_2$, which is a contradiction. So, in this case, Σ_{n_1,n_2} is not identifiable.

Now, if the pair $(A^{22}, [B^2, A^{21}])$ is not controllable, we have already seen in Remark 3.4.8 that the MN Σ_{n_1,n_2} is not minimal, since it is possible to perform a state space reduction, and so, in particular, Σ_{n_1,n_2} is not identifiable.

Sufficiency We prove this part only for continuous time dynamics, the proof in discrete time being the same after the obvious changes of notations.

Let $x_0 \in \mathbb{R}^n$; consider another admissible initialized MN $(\tilde{\Sigma}_{\tilde{n}_1,\tilde{n}_2}, \tilde{x}_0)$, such that

$$(\Sigma_{n_1,n_2}, x_0) \sim (\tilde{\Sigma}_{\tilde{n}_1,\tilde{n}_2}, \tilde{x}_0).$$

Since Σ_{n_1,n_2} is observable, in particular, $V_1 = 0$, thus, by Lemma 3.5.5, $n_1 \le \tilde{n}_1$, and for all $i \in \{1,\dots,n_1\}$ there exists $\pi(i) \in \{1,\dots,\tilde{n}_1\}$ and $\beta(i) = \pm 1$ such that the conclusions (a), (b) and (c1) of Lemma 3.5.5 hold. Now let $\lambda : \{0,\dots,\tilde{n}_1\} \to \{0,\dots,\tilde{n}_1\}$ be the permutation given by $\lambda(i) = \pi(i)$, if $1 \le i \le n_1$, $\lambda(i) = i$ otherwise. Notice that the map λ is indeed a permutation since the map π is injective (see Lemma 3.5.5). Then, let $T_1 = PD \in G_{\tilde{n}_1}$, where P is the permutation matrix representing λ, and $D = \mathrm{Diag}(\beta(1),\dots,\beta(n_1),1,\dots,1)$. Then, (a), (b), and (c1) of Lemma 3.5.5 can be rephrased as:

$$T_1\tilde{B}^1 = \begin{pmatrix} B^1 \\ \tilde{B}^{11} \end{pmatrix} \tag{69}$$

$$\tilde{C}^1 T_1 = \left(C^1, \tilde{C}^{11}\right) \tag{70}$$

and, for all $w : [0,T_w] \to \mathbb{R}^m$, for all $t \in [0,\epsilon_w)$, we also have:

$$\left[A^{11}x_w^1(t) + A^{12}x_w^2(t)\right]_i = \left[T_1\tilde{A}^{11}\tilde{x}_w^1(t) + T_1\tilde{A}^{12}\tilde{x}_w^2(t)\right]_i, \quad i \in \{1,\dots,n_1\}. \tag{71}$$

Thus, Lemma 3.5.4 applies and we get:

$$\left[A^{12}(A^{22})^l B^2\right]_{ij} = \left[T_1\tilde{A}^{12}(\tilde{A}^{22})^l\tilde{B}^2\right]_{ij}, \; i \in \{1,\dots,n_1\}, j \in \{1,\dots,m\}, \tag{72}$$

and

$$\begin{array}{l}\left(A^{12}(A^{22})^l A^{21}x_w^1(t) + A^{12}(A^{22})^{l+1}x_w^2(t)\right)_i = \\ \left(T_1\tilde{A}^{12}(\tilde{A}^{22})^l\tilde{A}^{21}\tilde{x}_w^1(t) + T_1\tilde{A}^{12}(\tilde{A}^{22})^{l+1}\tilde{x}_w^2(t)\right)_i\end{array} \tag{73}$$

By applying Lemma 3.5.2, from equation (73), we get:

$$\left[A^{12}(A^{22})^l A^{21}\vec{\sigma}\left(A^1x_w(t) + B^1u\right) + A^{12}(A^{22})^{l+1}\left(A^2x_w(t) + B^2u\right)\right]_i =$$

$$\left[T_1\tilde{A}^{12}(\tilde{A}^{22})^l\tilde{A}^{21}\vec{\sigma}\left(\tilde{A}^1x_w(t) + \tilde{B}^1u\right) + T_1\tilde{A}^{12}(\tilde{A}^{22})^{l+1}\left(\tilde{A}^2\tilde{x}_w(t) + \tilde{B}^2u\right)\right]_i =$$

$$\left[T_1\tilde{A}^{12}(\tilde{A}^{22})^l\tilde{A}^{21}T_1^{-1}\vec{\sigma}\left(A^1x_w(t)\left(B^1,\tilde{B}^{11}\right)^T u\right) + T_1\tilde{A}^{12}(\tilde{A}^{22})^{l+1}\left(\tilde{A}^2\tilde{x}_w(t) + \tilde{B}^2u\right)\right]_i,$$

where to get this last equality we have used equations (71) and (69). Now, since the function σ and the matrix B^1 are both admissible, by Lemma 3.5.1, we conclude:

$$\left(A^{12}(A^{22})^l A^{21}\right)_{ij} = \left(T_1\tilde{A}^{12}(\tilde{A}^{22})^l\tilde{A}^{21}T_1^{-1}\right)_{ij}, \; \forall i,j \in \{1,\dots,n_1\}. \tag{74}$$

By Lemma 3.5.3, we also get, for all $l \ge 0$,:

$$C^2(A^{22})^l B^2 = \tilde{C}^2(\tilde{A}^{22})^l\tilde{B}^2. \tag{75}$$

and

$$\left(C^2(A^{22})^l A^{21}\right)_{ij} = \left(\tilde{C}^2(\tilde{A}^{22})^l\tilde{A}^{21}T_1^{-1}\right)_{ij} \;\; \forall i \in \{1,\dots,p\}, \; \forall j \in \{1,\dots,n_1\}. \tag{76}$$

Let:

$$\tilde{A}^{21}T_1^{-1} = \left(\tilde{M}^{21}, \tilde{N}^{21}\right),$$

where $\tilde{M}^{21} \in \mathbb{R}^{\tilde{n}_2 \times n_1}$, and

$$T_1\tilde{A}^{12} = \begin{pmatrix} \tilde{H}^{12} \\ \tilde{K}^{12} \end{pmatrix}$$

where $\tilde{H}^{12} \in \mathbb{R}^{n_1 \times \tilde{n}_2}$ Then equations (72), (74), (75), and (76), say that the two linear models,

$$\left(\left(\begin{array}{c} C^2 \\ A^{12} \end{array} \right), A^{22}, \left(B^2, A^{21} \right) \right), \quad \left(\left(\begin{array}{c} \tilde{C}^2 \\ \tilde{H}^{12} \end{array} \right), \tilde{A}^{22}, \left(\tilde{B}^2, \tilde{M}^{21} \right) \right), \tag{77}$$

are input/output equivalent. Since Σ_{n_1,n_2} is observable, we have $V_2 = 0$. So, by Lemma 3.5.6, the pair $\left(\left(\begin{array}{c} C^2 \\ A^{12} \end{array} \right), A^{22} \right)$ is observable. Moreover, the pair $(A^{22}, (B^2, A^{21}))$ is controllable, by assumption, and thus, by the linear theory (see Theorem 2.4.3), we get $n_2 \leq \tilde{n}_2$.

So, in conclusion, $n = n_1 + n_2 \leq \tilde{n}_1 + \tilde{n}_2 = \tilde{n}$.

Remark 3.5.7 Notice that, up to this point we have only used the i/o equivalence of the two MN's together with the facts that $V_1 = 0$, $V_2 = 0$, and the pair $(A^{22}, (B^2, A^{21}))$ is controllable.

Now, we must show that, if $n = \tilde{n}$ then (Σ_{n_1,n_2}, x_0) is equivalent to $(\tilde{\Sigma}_{\tilde{n}_1,\tilde{n}_2}, \tilde{x}_0)$. Notice that, from what we have seen before, necessarily, $n_1 = \tilde{n}_1$ and $n_2 = \tilde{n}_2$. Moreover, we must have:

$$\begin{array}{rclcrcl} \tilde{B}^1 & = & T_1^{-1} B^1 & \qquad & \tilde{A}^{21} T_1^{-1} & = & \tilde{M}^{21} \\ \tilde{C}^1 & = & C^1 T_1 & & T_1 \tilde{A}^{12} & = & \tilde{H}^{12}. \end{array} \tag{78}$$

On the other hand, by the linear theory (again Theorem 2.4.3), there must exists $T_2 \in GL(n_2)$, such that:

$$\left(\begin{array}{c} \tilde{C}^2 \\ \tilde{H}^{12} \end{array} \right) = \left(\begin{array}{c} C^2 \\ A^{12} \end{array} \right) T^2, \quad \tilde{A}^{22} = T_2^{-1} A^{22} T_2, \quad \left(\tilde{B}^2 \tilde{M}^{21} \right) = T_2^{-1} \left(B^2 A^{21} \right). \tag{79}$$

By applying Lemma 3.5.2 to equation (71) , we also have:

$$\left[A^{11} \vec{\sigma} \left(A^1 x_w(t) + B^1 u \right) + A^{12} \left(A^2 x_w(t) + B^2 u \right) \right]_i =$$
$$\left[T_1 \tilde{A}^{11} T_1^{-1} \vec{\sigma} \left(A^1 x_w(t) + B^1 u \right) + T_1 \tilde{A}^{12} \left(\tilde{A}^2 \tilde{x}_w(t) + \tilde{B}^2 u \right) \right]_i$$

Clearly this equality implies:

$$A^{11} = T_1 \tilde{A}^{11} T_1^{-1}. \tag{80}$$

Now let:

$$T = \left(\begin{array}{cc} T_1 & 0 \\ 0 & T_2 \end{array} \right).$$

Clearly $T \in \mathcal{G}_n$, and by (78), (79), and (80), we have:

$$\tilde{C} = CT, \ \tilde{A} = T^{-1} A T, \ \tilde{B} = T^{-1} B. \tag{81}$$

So, to complete the proof, we only need to show that $\tilde{x}_0 = T^{-1} x_0$. Let $x_1 = T \tilde{x}_0$. Since $(\Sigma_{n_1,n_2}, x_0) \sim (\tilde{\Sigma}_{\tilde{n}_1,\tilde{n}_2}, \tilde{x}_0)$, and (81) holds, we get that x_0 is indistinguishable from x_1. On the other hand, since Σ_{n_1,n_2} is observable, we conclude $T \tilde{x}_0 = x_1 = x_0$, as desired.

Proof of Proposition 3.4.7
We have already seen in Remark 3.4.8 that the conditions in (54) are necessary for minimality. On the other hand, the sufficiency follows directly from the previous proof. In fact, as observed in Remark 3.5.7, to get that $\tilde{n} \leq n$ one needs only the assumptions given by (54).

Remark 3.5.8 It is interesting to notice that the conditions in (54) guarantees, if $n = \tilde{n}$, the existence of a matrix $T \in \mathcal{G}_n$ which gives the similarity of the two set of matrices; however, if we do not have the whole observability condition (i.e. if $\ker A \cap \ker C \neq 0$), then $T \tilde{x}_0$ and x_0 may be different (see also Example 2.4.7).

Chapter 4

Some Open Problems

Recurrent Neural Networks have proved to be a quite stimulating object for system theoretic study. With the results of this work, we may say that minimality in both RNN's and MN's is well understood, under the assumption of admissibility. On the other hand, there are still many interesting questions concerning system theoretic properties of these networks that remain open.

1. While admissibility of the activation function is an acceptable assumption, the admissibility requirement on the control matrix appear to be an undesirable constrain. Is it conceivable, for instance, to be able to reduce the dimension of a RNN by using a non-admissible control matrix? First of all, by dropping the admissibility assumption in the control matrix, the symmetry group of a RNN may considerably increase; this is due to the fact that if B is not admissible then there are non-coordinate subspaces that are stable for $\sigma \circ B$.

   It appears (we are presently working on this problem) that observability for a RNN can be characterized without requiring admissibility of the control matrix. On the other hand, identifiability seems to be much harder. For identifiability, in fact, we compare two systems with control matrices B_1, B_2, so that the maps $\sigma \circ B_1$ and $\sigma \circ B_2$ may have different stable subspaces.

2. Dropping the admissibility assumption for the control matrix B would allow to deal with *multiple layer* neural networks. A multiple, layer neural network is a *cascade* of neural networks such that the output of one network is the input of the successive one. Thus, a double layer RNN would be a system of the type

$$\left\{ \begin{array}{lcl} x_1^+ & = & \vec{\sigma}(A_1 x_1 + B_1 u) \\ x_2^+ & = & \vec{\sigma}(A_2 x_2 + B_2 C_1 x_1) \\ y & = & C_2 x_2 \end{array} \right.$$

   that is still a RNN but with the non admissible control matrix $\begin{pmatrix} B_1 \\ 0 \end{pmatrix}$. By analogy with what is known for feedforward networks (see [25]), we expect that double layer RNN's would better approximate certain nonlinear i/o behaviors than admissible RNN's. The identifiability and minimality problem is particularly relevant in this context.

3. As we mentioned earlier, complete controllability for continuous time RNN's with suitable activation functions has been recently proved. We believe that the techniques

could be used to deal with MN's in continuous time. It seems, however, that controllability in discrete time is much harder, and makes a quite challenging theoretical problem.

4. Forward accessibility in discrete-time has been proved under strong assumptions on the activation function and the control matrix; these assumptions are hard to check, even for systems with relatively low dimension. Improvements in this direction are quite desirable.

5. It appears that some results in this work come from "linear" properties of $\vec{\sigma}$: invariant subspaces, linear transformations that commute with $\vec{\sigma}$, ... We believe it is worthwhile to investigate whether similar arguments can be applied to more general vector fields, that are not produced by scalar functions applied componentwise.

Bibliography

[1] Albertini, F., and Dai Pra P., "Observability of discrete-time recurrent neural networks coupled with linear systems," *Proc. European Control Conference*, Rome, Sept. 5-8, 1995, pg. 1586-1589.

[2] Albertini, F., and Dai Pra P., "Forward accessibility for recurrent neural networks," *IEEE Transaction on Automatic Control*, November,1995, vol.40, (no.11), pg. 1962-8.

[3] Albertini, F., and Dai Pra P., "Recurrent neural networks coupled with linear systems: observability in continuous and discrete time," accepted by *Systems and Control Letters*, 27, No. 2, 1996 pg. 109-116.

[4] Albertini, F., and D'Alessandro D., "Remarks on the Observability of Discrete-Time Nonlinear Systems," *Proc. IFIP, System Modeling and Optimization*, Prague, July 10-14, 1995.

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

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

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

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

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

[10] Bovier, A., Gayrard, V., "Lower bounds on the memory capacity of the dilute Hopfield model", *Cellular automata and cooperative systems*, (Les Houches, 1992), 55–66, NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci., 396, Kluwer Acad. Publ., Dordrecht, 1993.

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

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

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

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

[15] Hermann R. and Krener A.J., “Nonlinear controllability and observability”, *IEEE Trans. Automat. Control.*, **AC-22** (1977), 728-740.

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

[17] Hunt K.J., Sbarbaro D., Zbikowski R., and Gawthrop P.J., “Neural Networks for control systems: A survey”, *Automatica*, 28 (1992): 1083-1122.

[18] Matthews M. “A state-space approach to adaptive nonlinear filtering using recurrent neural networks”, in *Proc. 1990 IASTED Symp. on Artificial Intelligence Applications and Neural Networks*, Zurich, July, 1990: 197-200.

[19] Michel A.N., Farrell J.A. and Porod W, “Qualitative analysis of neural networks”, *IEEE Trans. Circuit and Sys.*, 36 (1989):229-243.

[20] Nijmeijer H., “Observability of autonomous discrete time non-linear systems: a geometric approach”, *Int.J.Control*, **36** (1982), 867-874.

[21] Polycarpou M.M. and Ioannou P.A. “Neural Networks and on-line approximators for adaptive control”, in *Proc. Seventh Yale Workshop on Adaptive and Learning Systems*, pp. 93-98, Yale University, 1992.

[22] Romanik K., “Approximate testing and learnability”, in *Proc. Fifth ACM Workshop on Computational Learning Theory*, Pittsburg, July, 1992.

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

[24] 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.

[25] Sontag E.D., “Feedback stabilization using two-hidden-layer nets”, in *IEEE Trans. Neural Networks* 3, 1992, 981-990.

[26] 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.

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

[28] Sontag E.D. and Sussmann, H.J., “Complete Controllability of Continuous-Time Recurrent Neural Networks”, submitted.

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

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

[31] Talagrand, M., “Resultats rigoureux pour le modle de Hopfield”, (French) *C. R. Acad. Sci. Paris Sr. I Math. 321*(1995), no. 1, 109–112.

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