

On the Identification of Recurrent Neural Nets

Dietmar Trummer, Manfred Deistler ¹

Institute of Econometrics, Operations Research and System Theory,

Vienna University of Technology,

Argentinierstr. 8, A-1040 Wien,

Tel.: ++43 +1 58801 11911, Fax.: ++43 +1 58801 11999,

email: Dietmar.Trummer@tuwien.ac.at

Manfred.Deistler@tuwien.ac.at

Abstract

In this paper observational equivalence for so-called Jordan networks, which are a special class of recurrent networks, is analysed. We show this type of neural nets to belong to a wider class of mixed networks and use the description of observational equivalence available for the latter class for obtaining the respective results for the first class.

nonlinear systems, recurrent neural nets, structure theory, identification

1 Introduction

This paper is concerned with analysing observational equivalence for a special class of recurrent neural nets, called Jordan nets. These results can be viewed as a module in a still incomplete theory of semi-nonparametric identification of recurrent neural nets, analogous to the existing theory for the linear case, described in Hannan and Deistler (1988). This paper is very much in the line of the work of Albertini and Dai Pra (1995), Albertini (1993), Albertini and Sontag (1993) (see also Dörfler and Deistler (1998)). In the semi-nonparametric approach, the identification problem may be decomposed into the following three modules: Structure theory, where an idealized problem is analysed, as we commence from the input-output map rather than from data, estimation for given state dimension and estimation of the state dimension.

For our analysis of observational equivalence for Jordan nets we embed this class in the class of so-called mixed nets.

Jordan networks are of the form

$$x_{t+1} = Ax_t + By_t \quad (1)$$

$$y_t = H\bar{\sigma}[Cx_t + Du_t] \quad (2)$$

¹Support by the Austrian 'Fonds zur Förderung der wissenschaftlichen Forschung', Projects P-11213-MAT and P-14438-INF is gratefully acknowledged.

where $x_t \in \mathbb{R}^n$ is the state at time t , $u_t \in \mathbb{R}^m$ is the input, $y_t \in \mathbb{R}^p$ is the output and

$$\begin{aligned} A &\in \mathbb{R}^{n \times n}, B \in \mathbb{R}^{n \times p}, C \in \mathbb{R}^{l \times n}, D \in \mathbb{R}^{l \times m}, \\ H &\in \mathbb{R}^{p \times l} \end{aligned}$$

are the parameter matrices. The nonlinear function $\bar{\sigma} : \mathbb{R}^l \rightarrow \mathbb{R}^l$ is defined as $\bar{\sigma}(x) = (\sigma(x_1), \dots, \sigma(x_l))'$, where σ is the so-called activation function (i.e. an odd and strictly monotonically increasing function, typically (but not necessarily) bounded) and is assumed to be fixed and to satisfy the so-called independence property throughout the paper (we then call σ admissible). We call a matrix D admissible if for all rows d_i of D the following conditions hold:

$$d_i \neq 0 \text{ and } d_i \neq \pm d_j; \quad i \neq j.$$

From a system theoretic point of view the way the Jordan network is formulated, is rather strange, since the state is influenced by the past state and the past output rather than the past input. The idea stems from speech recognition, where the phenomenon of a long influence of past outputs is observed. This quite plausible when we realize that the pronunciation of a certain vowel can be influenced by the preceding four or five consonants (Jordan (1992), Jordan (1986)). So the idea was to feed back the output rather than the nonlinear transformation of the state.

As can be seen from a straight forward calculation the Jordan network can also be written as

$$y_t = H\bar{\sigma}[CA^t x_0 + \sum_{j=1}^t CA^{j-1} B y_{t-j} + D u_t]. \quad (3)$$

Results analogous to those described in the paper have been obtained for so-called Elman networks (Elman (1984), Elman (1991)) e. g. in Albertini and Dai Pra (1995) and Dörfler and Deistler (1998).

One of our main aims is to give a description of the classes of observational equivalence for Jordan nets.

This is done by analysing the transformations between the parameters of so-called mixed networks (see below) and Jordan networks and by using the corresponding results obtained for mixed networks in Albertini and Dai Pra (1995), Albertini (1993) and Albertini and Sontag (1993).

2 Structure Theory for Mixed networks

An extension of the Elman net, the so-called mixed network, has been investigated for instance in Albertini and Dai Pra (1995). This mixed network is of the form:

$$x_{t+1}^1 = \bar{\sigma}[A_{11}x_t^1 + A_{12}x_t^2 + B_1u_t] \quad (4)$$

$$x_{t+1}^2 = A_{21}x_t^1 + A_{22}x_t^2 + B_2u_t \quad (5)$$

$$y_t = C_1x_t^1 + C_2x_t^2 \quad (6)$$

where $x_t^1 \in \mathbb{R}^{n_1}$, $x_t^2 \in \mathbb{R}^{n_2}$ are the states and where the parameter matrices have the corresponding dimensions:

$$\begin{aligned} A_{11} &\in \mathbb{R}^{n_1 \times n_1}, A_{12} \in \mathbb{R}^{n_1 \times n_2}, A_{21} \in \mathbb{R}^{n_2 \times n_1}, \\ A_{22} &\in \mathbb{R}^{n_2 \times n_2}, B_1 \in \mathbb{R}^{n_1 \times m_1}, B_2 \in \mathbb{R}^{n_2 \times m_2}, \\ C_1 &\in \mathbb{R}^{p \times n_1}, C_2 \in \mathbb{R}^{p \times n_2}. \end{aligned}$$

The structure theory of mixed networks is in a certain sense a combination of that of linear and of nonlinear state space systems.

Consider the mixed networks as defined above. We arrange the parameters and the initial values into block matrices using the following self explaining notation:

$$\mathbf{A} = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}, \mathbf{B} = \begin{pmatrix} B_1 \\ B_2 \end{pmatrix}, \quad (7)$$

$$\mathbf{C} = (C_1, C_2), \mathbf{x}_0 = \begin{pmatrix} x_0^1 \\ x_0^2 \end{pmatrix}, \quad (8)$$

$$\mathbf{n} = n_1 + n_2$$

A system $(\mathbf{A}, \mathbf{B}, \mathbf{C})$ is called *admissible*, if B_1 is admissible. In the following we always assume the systems to be admissible.

1. We start with following definitions:

- For given $(\mathbf{A}, \mathbf{B}, \mathbf{C})$ and given initial value \mathbf{x}_0 (i.e. for a given initialised system $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{x}_0)$) the *input/output (I/O) map* is the map $\lambda(\mathbf{x}_0) : (\mathbb{R}^m)^{\mathbb{N}} \rightarrow (\mathbb{R}^p)^{\mathbb{N}}$ attaching the outputs $(y_t | t \in \mathbb{N})$ to the inputs $(u_t | t \in \mathbb{N})$.
- Two states $\mathbf{x}_0, \bar{\mathbf{x}}_0$ are called *indistinguishable* for the system $(\mathbf{A}, \mathbf{B}, \mathbf{C})$ if $\lambda(\mathbf{x}_0) = \lambda(\bar{\mathbf{x}}_0)$ holds.

- The system $(\mathbf{A}, \mathbf{B}, \mathbf{C})$ is said to be *observable* if there are no two different indistinguishable states.
- Two initialised systems $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{x}_0)_{\mathbf{n}}$ and $(\bar{\mathbf{A}}, \bar{\mathbf{B}}, \bar{\mathbf{C}}, \bar{\mathbf{x}}_0)_{\bar{\mathbf{n}}}$ (the \mathbf{n} and $\bar{\mathbf{n}}$ denote the corresponding state dimensions) are called *observationally equivalent*, if their I/O - maps are equal. We then use the symbol $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{x}_0)_{\mathbf{n}} \sim (\bar{\mathbf{A}}, \bar{\mathbf{B}}, \bar{\mathbf{C}}, \bar{\mathbf{x}}_0)_{\bar{\mathbf{n}}}$.
- A system $(\mathbf{A}, \mathbf{B}, \mathbf{C})$ with state dimension \mathbf{n} is called *minimal*, if $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{x}_0) \sim (\bar{\mathbf{A}}, \bar{\mathbf{B}}, \bar{\mathbf{C}}, \bar{\mathbf{x}}_0)_{\bar{\mathbf{n}}}$ implies $\mathbf{n} \leq \bar{\mathbf{n}}$, and this holds for all \mathbf{x}_0 .

First we describe the set of indistinguishable states: A subspace $V \subset \mathbb{R}^n$ is called a *coordinate subspace* if it is generated by elements of the canonical basis $\{e_1, \dots, e_n\}$. We then define W as

$$W = \mathbf{A}^{-1}(V_1 \oplus V_2) \cap \ker \mathbf{C} \quad (9)$$

where $\mathbf{A}^{-1}(V_1 \oplus V_2)$ is the preimage of $(V_1 \oplus V_2)$ under \mathbf{A} and where $V_1 \subseteq \mathbb{R}^{n_1}$ and $V_2 \subseteq \mathbb{R}^{n_2}$ are defined as the maximal pair of subspaces such that the following conditions hold:

- $V_1 \subseteq \ker C_1$, $A_{11} V_1 \subseteq V_1$
and V_1 is a coordinate subspace;
- $V_2 \subseteq \ker C_2$, $A_{22} V_2 \subseteq V_2$;
- $A_{21} V_1 \subseteq V_2$;
- $A_{12} V_2 \subseteq V_1$.

As can be shown, x_1 and x_2 are indistinguishable if and only if $x_1 - x_2 \in W$. For this reason, W is called the *unobservable* subspace.

2. Two matrices (A, B) are called *controllable* if they satisfy the ‘‘linear’’ controllability condition:

$$\text{rk} (B, AB, A^2B, \dots, A^{n-1}B) = n \quad (10)$$

where $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times p}$.

We list the most important results for mixed networks now:

Observability: It can be shown that the system $(\mathbf{A}, \mathbf{B}, \mathbf{C})$ is observable (i.e. $W = 0$) if and only if

$$\ker \mathbf{A} \cap \ker \mathbf{C} = 0 \quad \text{and} \quad V_1 = V_2 = 0 \quad (11)$$

hold.

3. Minimality: The system $(\mathbf{A}, \mathbf{B}, \mathbf{C})_{\mathbf{n}}$ is minimal if and only if

$$V_1 = V_2 = 0 \text{ and } (A_{22}, (B_2, A_{21})) \text{ is controllable.} \quad (12)$$

4. Let G_{n_1, n_2} be defined as the set of matrices

$$\mathbf{T} = \begin{pmatrix} S & 0 \\ 0 & T \end{pmatrix}, \text{ with } \mathbf{T} \in \mathbb{R}^{(n_1+n_2) \times (n_1+n_2)} \\ \text{and } S \in \mathbb{R}^{n_1 \times n_1}, T \in \mathbb{R}^{n_2 \times n_2}, \quad (13)$$

where T is regular and S is a permutation matrix with sign changes (i.e. in each row and column one entry is equal to ± 1 and all other entries are equal to 0).

Let $(\mathbf{A}, \mathbf{B}, \mathbf{C})$ and $(\overline{\mathbf{A}}, \overline{\mathbf{B}}, \overline{\mathbf{C}})$ be two minimal systems, which fulfill the condition $\ker \mathbf{A} \cap \ker \mathbf{C} = 0$ (and thus are observable). Then the initialised systems $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{x}_0)$ and $(\overline{\mathbf{A}}, \overline{\mathbf{B}}, \overline{\mathbf{C}}, \overline{\mathbf{x}}_0)$ are observationally equivalent if and only if the following conditions hold:

$$n_1 = \overline{n}_1, n_2 = \overline{n}_2, \quad (14)$$

there exists a matrix $\mathbf{T} \in G_{n_1, n_2}$ such that

$$\overline{\mathbf{A}} = \mathbf{T}^{-1} \mathbf{A} \mathbf{T}, \overline{\mathbf{B}} = \mathbf{T}^{-1} \mathbf{B} \quad (15)$$

$$\overline{\mathbf{C}} = \mathbf{C} \mathbf{T}, \overline{\mathbf{x}}_0 = \mathbf{T}^{-1} \mathbf{x}_0. \quad (16)$$

Conversely, let $(\mathbf{A}, \mathbf{B}, \mathbf{C})$ be a minimal system. If for all starting values \mathbf{x}_0 all observationally equivalent initialised systems have either larger dimension or are related by a transformation via a matrix \mathbf{T} such that (15), (16) hold, then $(\mathbf{A}, \mathbf{B}, \mathbf{C})$ is observable.

Theorem 1: Let the system $(\mathbf{A}, \mathbf{B}, \mathbf{C})$ (compare (7), (8)) be minimal (i.e. let $V_1 = V_2 = 0$ and the controllability condition hold) with initial value \mathbf{x}_0 . The class of initialised systems, which are observationally equivalent to $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{x}_0)$ is given by $(\overline{\mathbf{A}}, \overline{\mathbf{B}}, \overline{\mathbf{C}}, \overline{\mathbf{x}}_0)$, where

$$\exists \mathbf{T} \in G_{l, n} : \\ \overline{\mathbf{A}} = \mathbf{T}^{-1} \mathbf{A} \mathbf{T}, \overline{\mathbf{B}} = \mathbf{T}^{-1} \mathbf{B}, \overline{\mathbf{C}} = \mathbf{C} \mathbf{T}, \quad (17)$$

$$\overline{\mathbf{T} \overline{\mathbf{x}}_0} = \hat{\mathbf{x}}_0 \in \mathbf{x}_0 + \ker \mathbf{A} \cap \ker \mathbf{C}. \quad (18)$$

Proof: It follows from the proof of the characterisation of the equivalence classes in Albertini and Dai Pra (1995) that the class of observationally equivalent systems is given by (17), where $\mathbf{T} \overline{\mathbf{x}}_0 = \hat{\mathbf{x}}_0$ and where $\hat{\mathbf{x}}_0$ and \mathbf{x}_0 are indistinguishable for $(\mathbf{A}, \mathbf{B}, \mathbf{C})$. This is implied by the fact that the condition $\ker \mathbf{A} \cap \ker \mathbf{C} = \{0\}$ has to be used only for proving minimality (see also Albertini and Dai Pra (1995), Remark 3.5.8).

Therefore it suffices to show: $\hat{\mathbf{x}}_0 - \mathbf{x}_0 \in \ker \mathbf{A} \cap \ker \mathbf{C} \iff \hat{\mathbf{x}}_0$ and \mathbf{x}_0 are indistinguishable for $(\mathbf{A}, \mathbf{B}, \mathbf{C})$.

Sufficiency of the condition above follows by direct calculation:

$$\begin{aligned} 0 &= \mathbf{C}(\mathbf{x}_0 - \hat{\mathbf{x}}_0) = \mathbf{C} \mathbf{x}_0 - \mathbf{C} \hat{\mathbf{x}}_0 = \\ &C_1 x_0^1 + C_2 x_0^2 - C_1 \hat{x}_0^1 - C_2 \hat{x}_0^2 \implies y_0 = \hat{y}_0. \\ 0 &= \mathbf{A}(\mathbf{x}_0 - \hat{\mathbf{x}}_0) = \mathbf{A} \mathbf{x}_0 - \mathbf{A} \hat{\mathbf{x}}_0 \implies \\ &A_{i1} x_0^1 + A_{i2} x_0^2 - A_{i1} \hat{x}_0^1 - A_{i2} \hat{x}_0^2 \\ &\implies x_1^i = \hat{x}_1^i \quad (i = 1, 2) \implies y_1 = \hat{y}_1. \end{aligned}$$

Hence we have $y_t = \hat{y}_t, \forall t \geq 1$. To see necessity, we use indistinguishability:

$$\begin{aligned} y_0 &= \hat{y}_0 \iff C_1 x_0^1 + C_2 x_0^2 - C_1 \hat{x}_0^1 - C_2 \hat{x}_0^2 = \\ &\mathbf{C}(\mathbf{x}_0 - \hat{\mathbf{x}}_0) = 0 \iff \mathbf{x}_0 - \hat{\mathbf{x}}_0 \in \ker \mathbf{C}. \\ \mathbf{x}_1 &= \hat{\mathbf{x}}_1 \implies A_{i1} x_0^1 + A_{i2} x_0^2 - A_{i1} \hat{x}_0^1 - A_{i2} \hat{x}_0^2 \\ &(i = 1, 2) \iff \mathbf{x}_0 - \hat{\mathbf{x}}_0 \in \ker \mathbf{A}. \quad \square \end{aligned}$$

Below we also will analyse the boundary of the parameter space, where for Jordan nets at least one of the conditions in (11), (12) is not satisfied.

3 Mixed network representation and observationally equivalence for the Jordan Network

In this section we use the results for mixed networks described in the previous section to obtain results on the structure of equivalence classes for Jordan nets. This can be done by reformulating Jordan nets as special mixed networks:

Consider the system (1),(2) and define an additional state $z_t := \vec{\sigma}[C x_t + D u_t]$. By calculating z_{t+1} and inserting the state equation (1) we finally get

$$z_{t+1} = \vec{\sigma}[C A x_t + C B H z_t + D u_{t+1}] \quad (19)$$

$$x_{t+1} = A x_t + B H z_t \quad (20)$$

$$y_t = H z_t \quad (21)$$

The corresponding state dimensions are: $z_t \in \mathbb{R}^l, x_t \in \mathbb{R}^n$, therefore we define $n_1 = l, n_2 = n, \mathbf{n} = l + n$. Note that now there is also an additional initial value z_0 , which is defined by $z_0 = \vec{\sigma}[C x_0 + D u_0]$. As can be seen in eq. (19), the input sequence u_t has to be shifted.

The Jordan network can be represented as a mixed network with system block matrices of the following form:

$$\mathbf{A} = \begin{pmatrix} C B H & C A \\ B H & A \end{pmatrix}, \mathbf{B} = \begin{pmatrix} D \\ 0 \end{pmatrix}, \quad (22)$$

$$\mathbf{C} = (H, 0), \mathbf{x}_0 = \begin{pmatrix} z_0 \\ x_0 \end{pmatrix}. \quad (23)$$

Our aim is now to describe the classes of observationally equivalent Jordan networks, using the corresponding results for mixed networks. First for given $(\mathbf{A}, \mathbf{B}, \mathbf{C})$ we consider the set of corresponding Jordan nets. From (22) and (23), we see that A, D and H are unique. The same holds for the matrices B, C if A has rank n and H has rank p .

In the next step we embed the Jordan net into the mixed net by the mapping attaching the parameters of the Jordan neural net to those of the mixed net.

This embedding $i : \mathbb{R}^{n^2 + np + ln + lm + pl} \rightarrow \mathbb{R}^{(l+n)^2 + (l+n)m + p(l+n)}$ is defined by (see (22), (23)):

$$i(A, B, C, D, H) = \left(\begin{pmatrix} CBH & CA \\ BH & A \end{pmatrix}, \begin{pmatrix} D \\ 0 \end{pmatrix}, (H, 0) \right)$$

The mapping i is generically (under the condition that A and H have full rank) injective and relatively to the image of the embedding generically homeomorphic.

A Jordan network is called admissible, if D is admissible. Since by the embedding D corresponds to B_1 , a Jordan net is admissible if and only if also the corresponding mixed network is admissible. Throughout we further assume admissibility of the Jordan network. By J_1 we denote the set of all admissible Jordan nets, which fulfill $\text{rk } A = n$, $\text{rk } H = p$, where $\text{rk } A$ denotes the the rank of A .

Note that on J_1 the embedding i is homeomorphic and that J_1 is open and dense in $\mathbb{R}^{n^2 + np + ln + lm + pl}$, the set of all Jordan nets.

For a Jordan net the the conditions defining the unobservable subspace are of the following form:

- (i) $V_1 \subseteq \ker H$, $CBH V_1 \subseteq V_1$
and V_1 is a coordinate subspace;
- (ii) $A V_2 \subseteq V_2$;
- (iii) $BH V_1 \subseteq V_2$;
- (iv) $CA V_2 \subseteq V_1$.

Note that (i) obviously implies (iii).

A Jordan net is minimal, if and only if $V_1 = V_2 = 0$ holds and in addition $i(A_{22}, (B_2, A_{21})) = (A, (0, BH))$ is controllable. Minimality is preserved by the embedding and whenever a minimal system $(\mathbf{A}, \mathbf{B}, \mathbf{C})$ is an element of $i(J_1)$ then also the whole equivalence class of minimal systems is contained in $i(J_1)$.

If therefore the system $(\mathbf{A}, \mathbf{B}, \mathbf{C})$ is minimal, the set of observationally equivalent systems is given by the transformation via a matrix $\mathbf{T} \in G_{l,n}$. By applying i^{-1} we obtain the characterisation of the class of observationally equivalent systems for Jordan networks from (15) and (16):

$$\begin{aligned} \bar{A} &= T^{-1}AT, & \bar{B} &= T^{-1}B, & \bar{C} &= S^{-1}CT, \\ & & \bar{D} &= S^{-1}D, & \bar{H} &= HS. \end{aligned}$$

Thus the following result holds:

Theorem 2: Two different minimal Jordan networks are observationally equivalent if and only if the two corresponding minimal mixed networks are observationally equivalent.

This result may be of use to restrict the parameter spaces to certain subsets of \mathbb{R}^k for $k = n^2 + np + ln + lm + pl$.

In the following we discuss the conditions for observability and minimality:

1. Note that $w_1 \in \ker H$ and $w_2 \in \ker A$ implies

$$\begin{pmatrix} w_1 \\ w_2 \end{pmatrix} \in \ker \mathbf{A} \cap \ker \mathbf{C}. \quad (24)$$

As $\ker H \neq 0$ always holds for $p < l$, (24) implies that observability (see (11)) will not hold generically for $p < l$ even under minimality (although $\ker A = 0$ holds generically).

If $\ker \mathbf{A} \cap \ker \mathbf{C} \neq 0$, the class of observationally equivalent systems is given by Theorem 1.

2. As far as minimality is concerned, it can be shown (by an argument completely analogous to that showing that a pair of matrices (A, B) is generically controllable) that the set of Jordan systems satisfying the controllability condition $\text{rk}(A, BH) = n$ is also open and dense in J_1 .

We now consider the condition $V_1 = V_2 = 0$ and we first observe that generically $V_1 = 0$ holds. Indeed, in order to be a coordinate subspace, at least one entry of all elements of V_1 has to be zero (This holds, since the case $\text{rank } H = 0$ is excluded). But this can only be the case, if certain elements of H are also zero. We further observe that in J_1 , the condition (ii) does not restrict V_2 at all, whereas condition (iv) does (generically, i.e. in the case $V_1 = 0$) restrict V_2 to be the set of all x such that $Ax \in \ker C$. However, this does not even generically imply $V_2 = 0$, if $l < n$ (Note that C is a $(l \times n)$ matrix).

Remark: We want to analyse if the use of the state dimension l , which has been “artificially“ introduced, is essential (note that the condition $\ker H = \{0\}$ is not fulfilled in the case $p < l$). A small example shows that (while all other dimensions are kept fixed) a certain input/output behaviour can only be modeled with systems with a sufficient large state dimension l .

Let $n = p = m = 1$ and $u_t = 1 \forall t = 1, \dots, T$. Assume that $\sigma(x) = 0$ holds only for $x = 0$ (this just makes the example more simple). Let the output sequence be given by $y_0 \neq 0$, $y_1 = y_2 = 0$, $y_3 \neq 0$, where the nonzero elements may be arbitrary. Then there is no Jordan net with $l = 1$ having this I/O behaviour. This is of course not the case with e.g. $l = 3$.

One could also ask, if the definition of l as state dimension could be avoided by using a sufficiently large n . Of course, the state dimension l is implicitly given by the definition of the Jordan network. Therefore one can equivalently ask, if the matrix H is redundant. The following consideration shows that this is not the case.

Let $(\mathbf{A}, \mathbf{B}, \mathbf{C})_{\mathbf{n}}$ be a minimal mixed network. Then another mixed network $(\overline{\mathbf{A}}, \overline{\mathbf{B}}, \overline{\mathbf{C}})_{\overline{\mathbf{n}}}$ is I/O- equivalent, if $\mathbf{n} = \overline{\mathbf{n}}$ (compare (14)) and the parameter matrices are transformed via \mathbf{T} ; otherwise $\mathbf{n} > \overline{\mathbf{n}}$ must hold. It is mentioned in remark 3.4.6 in Albertini and Dai Pra (1995) that $\mathbf{n} > \overline{\mathbf{n}}$ is indeed equivalent to $n_1 > \overline{n}_1$ und $n_2 > \overline{n}_2$, which for the embedded Jordan network implies $l > \overline{l}$. Thus a small l cannot be compensated by a large n .

4 Summary

The conditions under which the set of observationally equivalent systems for a given minimal Jordan network can be described via a transformation \mathbf{T} have been found by observing that the Jordan network can be represented as a special kind of mixed network. Therefore the transformations defining the parameters of the class of observationally equivalent systems are the same also for Jordan nets. However, as these results are valid for minimal systems, they hold generically only under certain restrictions on the state dimensions. It has to be analysed, how this influences the estimation of Jordan systems.

Additionally, some questions concerning the complete controllability of Jordan nets and the properties of the data-to-parameter mapping (e.g. uniqueness of the parameters from finite amount of data) need further investigation.

References

- Albertini, F. and E. Sontag (1993). For neural networks, function determines form. *Neural Networks* **6**, –.
- Albertini, F. and P. Dai Pra (1995). Recurrent neural networks: Identification and other system theoretic properties. *Neural Network Systems Techniques and Applications* **3**, 1–41.
- Albertini (1993). Controllability of Discrete-Time Nonlinear Systems and some related Topics. PhD thesis. New Brunswick, New Jersey.
- Dörfler, M. and M. Deistler (1998). A structure theory for identification of recurrent neural networks, part 1. In: *Proceedings of the 4th IFAC Nonlinear Control Systems Design Symposium* (H.J.C.Huijberts, H.Nijmeijer, A.J. Van der Schaft and J.M.A. Scherpen, Eds.).
- Elman, J. L. (1984). Structured representations and connectionist models. In: *6th Annual Proceedings of the Cognitive Science Society*. pp. 17–25.
- Elman, J. L. (1991). Distributed representations, simple recurrent networks, and grammatical structure. *Machine Learning* **7**, 195–226.
- Hannan, E. J. and M. Deistler (1988). *The Statistical Theory of Linear Systems*. John Wiley. New York.
- Jordan, M. I. (1986). Attractor dynamics and parallelism in a connectionist sequential machine. Lawrence Erlbaum. pp. 531–546.
- Jordan, M. I. (1992). Constrained supervised learning. *J. of Mathematical Psychology* **36**, 396–425.