UNIQUENESS OF WEIGHTS FOR RECURRENT NETS

Francesca Albertini
Universita' di Padova, Dipartimento di Matematica,
Via Belzoni 7, 35100 Padova, Italy,
Eduardo D. Sontag
Rutgers University, Department of Mathematics,
New Brunswick, NJ 08903, U.S.A.
E-mail: albertini@pdmat1.unipd.it, sontag@hilbert.rutgers.edu

Keywords: recurrent networks, identifiability, observability.

1. Introduction

We study recurrent neural networks evolving either in discrete or in continuous time. The dynamics of such systems can be described by a set of either difference or differential equations of the following type (to simplify notations, we use the superscripts "+" and "." to denote time-shift and time-derivative respectively, and we omit the time arguments t):

$$\begin{array}{lcl} x^+ \text{ (or } \dot{x}) & = & \vec{\sigma}(Ax + Bu) \\ y & = & Cx, \end{array} \tag{1}$$

with $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, $C \in \mathbb{R}^{p\times n}$, and where $\vec{\sigma}(x) = (\sigma(x_1), \ldots, \sigma(x_n))$, with σ a function from $\mathbb{R}$ to itself. (For the continuous-time case, we always assume that the function σ is globally Lipschitz, so that the existence and uniqueness of solutions for the differential equation is guaranteed.)

We will consider the *parameter identifiability* problem, which asks about the possibility of recovering the entries of the matrices A, B, and C from the input/output map $u(\cdot) \mapsto y(\cdot)$ of the system. This question has been already addressed in [1] and in [2] (and, for feedforward nets, in [6] and in [4]) where it is proved that, under appropriate minimality assumptions, the *zero-initial state* i/o behavior determines, up to a small number of symmetries, the weights of the model. In this paper we establish the same result for *arbitrary-initial state* i/o maps. Moreover, we show that, for a generic subclass of these models, the minimality conditions needed in order for the results to hold are exactly the observability conditions found in the recent paper [3]. It is interesting to notice that, inside this subclass, these observability conditions are also necessary for identifiability.

2. The activation function σ

Given any map $\sigma : \mathbb{R} \to \mathbb{R}$, we say that σ satisfies the *independence property* ("**IP**" from now on) if, for every positive integer l, nonzero real numbers $b_1, \ldots, b_l$, and real numbers $\beta_1, \ldots, \beta_l$ such that $(b_i, \beta_i) \neq \pm(b_j, \beta_j)\ \ \forall i \neq j$, it must hold that the functions $1\,,\ \sigma(b_1 x + \beta_1)\,,\ \ldots\,,\ \sigma(b_l x + \beta_l)$ are linearly independent. The function $\sigma : \mathbb{R} \to \mathbb{R}$ which appears in our model will always assumed to satisfy property **IP** and to be odd. Given a matrix M, we denote by M_i the $i-$th row of M. For any two positive integer

$n,\ m$, we let:

$$\mathcal{B}_{n,m} \;=\; \left\{ B \in \mathbb{R}^{n\times m} \;\middle|\; \begin{array}{l} B_i \neq 0 \;\; \forall\, i = 1,\ldots,n \\ B_i \neq \pm B_j \;\; \forall\, i \neq j \end{array} \right\}.$$

3. Statement of the main results

Assume we are given a σ-system $\Sigma \equiv (A,B,C)_\sigma$ (either in discrete or in continuous time) initialized at a given state x_0 (we will write $\Sigma \equiv (A,B,C,x_0)_\sigma$ to denote the σ-system Σ together with the initial state x_0). Then we can associate to Σ an i/o map as follows. In the discrete-time case, for any sequence of inputs $u_1,\ldots,u_k$ a sequence of outputs $y_0 = Cx_0, y_1 \ldots, y_k$ is generated. In the continuous-time case for any control function $u(\cdot) : [0,T] \to \mathbb{R}^m$ (which we assume to be measurable essentially bounded), after solving the differential equation with initial state x_0 and denoting its solution with $x(t)$, we have a corresponding output function $y(t) = Cx(t)$. Thus, in both cases, to the given initialized system $\Sigma \equiv (A,B,C,x_0)_\sigma$ we associate an i/o map:

$$\lambda_{\Sigma,x_0} : \left\{ \begin{array}{ccccl} (u_1,\ldots,u_k) & \to & (y_0,\ldots,y_k) & & \text{discrete} \\ u(\cdot) & \to & y(\cdot) & & \text{continuous.} \end{array} \right.$$

We say that two initialized σ-system $\Sigma \equiv (A,B,C,x_0)_\sigma$, and $\tilde\Sigma \equiv (\tilde A,\tilde B,\tilde C,\tilde x_0)_\sigma$ are *i/o equivalent* if $p = \tilde p$, $m = \tilde m$, and $\lambda_{\Sigma,x_0} = \lambda_{\tilde\Sigma,\tilde x_0}$. We will study the problem of determining when two given initialized σ-systems are i/o equivalent.

Let:

$$\Lambda^n \;=\; \left\{ T \in \ \mathrm{Gl}\ (n) \;\middle|\; T = PQ \ \text{ with } \begin{array}{l} P = \ \text{ permutation matrix },\\ Q = \mathrm{Diag}\,(\beta(1),\ldots,\beta(n)),\ \beta(i) = \pm 1. \end{array} \right\}.$$

Definition 3.1 We say that two initialized σ-systems $\Sigma_i = (D_i, A_i, B_i, C_i, x_i)_\sigma$, for $i = 1,2$ are (sign-permutation) *equivalent* if $n_1 = n_2 = n$, and if there exists a matrix $T \in \Lambda^n$ such that:

$$A_2 = T^{-1}A_1T,\ C_2 = C_1T,\ B_2 = T^{-1}B_1,\ x_2 = T^{-1}x_1.$$

Since σ was assumed to be odd, it is easy to see that two equivalent σ-systems are i/o equivalent. Our aim is to show that *generically* also the converse holds. We say that a subset of $\mathbb{R}^p$ is *generic* if it is nonempty and if its complement is the set of zeroes of a finite number of polynomials in p variables. By a generic subset of σ-systems we mean a generic subset of $\mathbb{R}^{n^2+nm+np}$ when we identify the set of all σ-systems with $\mathbb{R}^{n^2+nm+np}$.

We first let $\mathcal{S}$ to be the subset of σ-systems (1) for which the matrix B is in $\mathcal{B}_{n,m}$. Notice that $\mathcal{S}$ is a generic subset of σ-systems.

We say that a subspace V of $\mathbb{R}^n$ is a *coordinate subspace* if it is of the type:

$$V \;=\; \text{span}\ \{ e_{i_1},\ldots,e_{i_k} \},$$

where e_{i_j} are the vectors of the canonical basis in $\mathbb{R}^n$. For any pair of matrices (A,C) we denote by $\mathcal{O}_\mathrm{c}(A,C)$ the largest A-invariant coordinate subspace included in $\ker C$. Next Remark presents a simple procedure to compute $\mathcal{O}_\mathrm{c}(A,C)$.

Remark 3.2 For any σ-system Σ, we let:

$$\begin{array}{rcl} I_0(\Sigma) & = & \{\, i \mid \exists\, j \in \{1,\ldots,p\} \;\text{ and }\; c_{ji} \neq 0\,\}, \\ I_k(\Sigma) & = & \{\, i \mid \exists\, l \in I_{k-1}(\Sigma) \;\text{ with }\; a_{li} \neq 0\,\}, \end{array}$$

and

$$\begin{array}{rcl} I(\Sigma) & = & \bigcup_{k\geq 0} I_k(\Sigma), \\ I^{\mathrm{c}}(\Sigma) & = & \{1,\ldots,n\} \setminus I(\Sigma)\,. \end{array}$$

Moreover, for each subset $J \subseteq \{1,\ldots,n\}$, let V_J be the following coordinate subspace:

$$V_J = \;\text{span}\,\{\, e_j \mid j \in J\,\}\,.$$

With these notations we have:

Proposition 3.3 $\mathcal{O}_{\mathrm{c}}(A,C) = V_{I^{\mathrm{c}}(\Sigma)}$.

A system Σ is said to be *observable* if for each two distinct initial states there exists some control that gives different output when Σ is started at those states (for a precise statement, both for discrete and continuous time systems, see e.g. [5]). Given these definitions, the following result holds (see [3], Theorem 1):

Proposition 3.4 A system $\Sigma \in \mathcal{S}$ is observable if and only if $\ker A \cap \ker C = \mathcal{O}_{\mathrm{c}}(A,C) = 0$.

Now, we can state our main results, which will be proved in a later section.

Theorem 1 *Assume that σ is an odd map which satisfies property* **IP***, (and, for the continuous-time case, it is globally Lipschitz). Let $\Sigma \equiv (A,B,C,x_0)_\sigma$ and $\tilde{\Sigma} \equiv (\tilde{A},\tilde{B},\tilde{C},\tilde{x}_0)_\sigma$ be two observable σ-systems in $\mathcal{S}$. Then, these systems are i/o equivalent if and only if they are equivalent.*

Remark 3.5 Notice that, inside the class $\mathcal{S}$, the observability condition is also necessary in order to guarantee the implication

$$\text{i/o equivalence} \quad \Rightarrow \quad \text{equivalence.}$$

More precisely, if $\mathcal{S}_0 \subseteq \mathcal{S}$ is any class of systems so that this implication holds for any pair of systems in $\mathcal{S}_0$ and initial states, then every system in $\mathcal{S}_0$ must be observable. This is proved as follows. Assume that $\Sigma \equiv (A,B,C)_\sigma$ is not observable. By definition of observability, there exist then two *distinct* states x_1 and x_2 which are not distinguishable. This means precisely that $\Sigma_1 \equiv (A,B,C,x_1)_\sigma$ and $\Sigma_2 \equiv (A,B,C,x_2)_\sigma$ are i/o equivalent. If the above implication would hold, then the systems must be equivalent. So there is a $T \in \Lambda^n$ such that $TB = B$ and $Tx_2 = x_1$. But the fact that $B \in \mathcal{B}_{n,m}$ implies that T is the identity. (In other words, the action of Λ^n on $\mathcal{S}$ is a free group action.) Thus $x_1 = x_2$, a contradiction.

Remark 3.6 It is also interesting to notice that, if a σ-system is not observable because $\mathcal{O}_c(A,C) \neq 0$, a reduction by unobservability is possible. It is only necessary to note that one can induce a dynamics on the quotient space $\mathbb{R}^n/\mathcal{O}_c(A,C)$, which can be done because the subspace $\mathcal{O}_c(A,C)$ is an A-invariant coordinate subspace. More precisely, the construction is as follows.

Let $\Sigma \in \mathcal{S}$. If $\mathcal{O}_c(A,C) \neq 0$, then after if necessary reordering variables, the equations for Σ take the following block form. The first block of variables corresponds to a basis of $\mathcal{O}_c(A,C)$, and has size $n_1 =$ dimension of $\mathcal{O}_c(A,C)$; the second set of variables has size $n_2 = n - n_1$.

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

Then $(A,B,C,(x_1,x_2))_\sigma$ and $\Sigma' \equiv (A_3,B_2,C_2,x_2)_\sigma$ are i/o-equivalent, for any initial state $x = (x_1,x_2)$ of the original system, and the second system has lower dimension. The system Σ' is again in $\mathcal{S}$. Moreover, Σ' is observable if $\ker A \cap \ker C = 0$. Indeed, the choice of variables insures that $\mathcal{O}_c(A_3,C_2) = 0$. Furthermore,

$$\operatorname{rank} \begin{pmatrix} A_1 & A_2 \\ 0 & A_3 \\ 0 & C_2 \end{pmatrix} = n$$

implies

$$\operatorname{rank} \begin{pmatrix} 0 & A_3 \\ 0 & C_2 \end{pmatrix} = n_2$$

so the conclusion $\ker A_3 \cap \ker C_2 = 0$ follows from $\ker A \cap \ker C = 0$.

References

[1] Albertini, F., and E.D. Sontag, "For neural networks, function determines form," *Neural Networks*, to appear. Summary in: "For neural networks, function determines form," *Proc. IEEE Conf. Decision and Control, Tucson, Dec. 1992*, IEEE Publications, 1992, pp. 26-31.

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

[3] Albertini, F., and E.D. Sontag, "State observability in recurrent neural networks," *System and Control Letters*, to appear.

[4] C. Fefferman, "Reconstructing a neural net from its output," preprint, Princeton University, 1993.

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

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