IDENTIFIABILITY OF DISCRETE-TIME NEURAL NETWORKS[1]

Francesca Albertini (*)
Eduardo D. Sontag
Department of Mathematics Rutgers University, New Brunswick, NJ 08903
E-mail: albertin@hilbert.rutgers.edu, sontag@hilbert.rutgers.edu
(*) Also: Universita' di Padova, Dipartimento di Matematica,
via Belzoni 7, 35100 Padova, Italy.

Key words: Neural networks, input/output behavior, identification from input/output data, control systems

ABSTRACT

This paper shows that from input/output measurements one can identify the structure of discrete-time feedback neural networks. The results presented here provide an analog of the continuous-time results given in a previous paper by the authors. The mathematical details, as well as the assumptions that must be made on the activation functions, are considerably different from those in the continuous-time case.

1 Introduction

We study here *discrete-time* recurrent neural networks, continuing the research that we started in [1] for the continuous-time case. Such networks are nonlinear controlled systems of the following special form:

$$\begin{cases} x(t+1) &= Dx(t) + \vec{\sigma}(Ax(t) + Bu(t)) \\ y(t) &= Cx(t) \quad t \in \mathbb{Z} \end{cases} \tag{1}$$

for some matrices $A \in R^{n\times n}$, $D \in R^{n\times n}$, $B \in R^{n\times m}$, and $C \in R^{p\times n}$. We assume that the matrix D is a diagonal matrix. The notation $\vec{\sigma}$ stands for the diagonal mapping

$$\vec{\sigma}: \mathbb{R}^n \to \mathbb{R}^n; \quad \vec{\sigma}(\sum_{i=1}^{n} a_i e_i) := \sum_{i=1}^{n} \sigma(a_i) e_i \quad \text{for all } a_1, \cdots, a_n \in \mathbb{R}\,, \tag{2}$$

where $\{e_1, \cdots, e_n\}$ is the canonical basis in $\mathbb{R}^n$ and $\sigma: \mathbb{R} \to \mathbb{R}$ is a fixed function, usually called the "activation function". We will assume that σ is an odd function.

The equations define a discrete-time systems whose state space, input-value space, and output-value space are $\mathbb{R}^n$, $\mathbb{R}^m$, and $\mathbb{R}^p$ respectively, in the standard language of control theory. We will call such a system a *σ-system*, and denote it by $\Sigma = (D, A, B, C)_\sigma$. The equations are interpreted as describing the evolution of real-valued variables $x_i, i = 1, \ldots, n$, each of which represents the internal state of a "neuron" or scalar processor at the time t. Each of the $u_i, i = 1, \ldots, m$, is an external input signal, also depending on time, and similarly each y_i, $i = 1, \ldots, p$ denotes an output value. The coefficients of the matrices A, B, C, D are often called the "weights" in the neural network literature. The function σ characterizes how each neuron responds to its aggregate input.

[1]This research was supported in part by US Air Force Grant AFOSR-91-0343, and in part by an INDAM (Istituto Nazionale di Alta Matematica F. Severi, Italy) fellowship.

For motivation about the use of such models, we refer the reader to the introduction of [1]; the interest in the discrete case, and relationships with continuous-time models, are analized in the recent paper [2].

As in [1], we consider the parameter identifiability problem. That is, we wish to know if the i/o behavior transforming inputs to output signals uniquely determines A, B, C, D. Under rather minimal assumptions on the activation map σ, we will prove that this is indeed the case, up to sign reversals of all incoming and outgoing weights at some units and a possible reordering of the variables. The result is essentially the same as that given in [1] for continuous time, and is totally different from what happens in linear systems theory, where the only uniqueness is up to a change of basis in the state space.

There are many similarities with the results in [1], both in the statement of the problem and in the form of the results being proved. However, mathematically the techniques used in the proofs are quite different; for instance, for analytic and odd σ, our assumptions will amount to requiring that the function is not a polynomial, but in the continuous-time case we could simply assume that it was not linear.

1.1 Main Assumptions on σ

We assume given an infinitely differentiable function $\sigma : \mathbb{R} \to \mathbb{R}$ which has the following properties:

$$\sigma \text{ is an odd function; i.e. } \sigma(x) = -\sigma(-x), \tag{3}$$

$$\sigma'(0) \neq 0, \tag{4}$$

$$\sigma^{(\alpha)}(0) \neq 0 \text{ for infinitely many } \alpha\text{ 's}. \tag{5}$$

We will see later (c.f. Remark 1.2) that for our purposes we may always assume that the function σ satisfies the following equation

$$\sigma'(0) = 1, \tag{6}$$

instead of equation (4). It is easy to see that if σ satisfies our assumptions, then the following holds:

$$\boxed{\text{If } \sigma(ax) = a\sigma(x) \text{ for all } x \text{ in a neighbourhood of 0 then } a \in \{\pm 1,\ 0\}.} \tag{P}$$

At various parts we will also impose the following condition on the function σ,

$$\boxed{\sigma'(x) \to 0 \text{ as } |x| \to \infty.} \tag{$A1$}$$

1.2 Systems

We consider σ-systems as in equation (1).

Remark 1.1 Note that different triples of matrices (D, A, B) may define the same function $Dx + \vec{\sigma}(Ax + Bu)$. However, for the type of functions σ we are dealing with this ambiguity will in general not happen. Indeed, assume that

$$(D_1, A_1, B_1, C)_\sigma = (D_2, A_2, B_2, C)_\sigma.$$

Then, it is possible to prove (see Remark 3.3 in [1] for the proof) that the following properties hold:

(a) if $D_1 = D_2$ then $A_1 = A_2$ and $B_1 = B_2$;

(b) if σ satisfies (A1) then $D_1 = D_2$ (and hence also $A_1 = A_2$ and $B_1 = B_2$).

Assume that we are given a fix σ-system, and this system is started at the state

$$x(0) \;=\; 0\,.$$

Then for any sequence of input values $u_1, \ldots, u_k$, a sequence of output signals is generated. In this manner, we can associated to each σ-system, $\Sigma = (D, A, B, C)_\sigma$, an *input-output map*

$$\lambda_\Sigma : (u_1, \ldots, u_k) \mapsto (y_1, \ldots, y_k). \tag{7}$$

Two systems Σ_1 and Σ_2 are *i/o equivalent* if $p_1 = p_2$, $m_1 = m_2$, and

$$\lambda_{\Sigma_1} = \lambda_{\Sigma_2}\,.$$

We will focus our attention on the following question: when are two given σ-systems Σ_1, Σ_2 i/o equivalent?

Remark 1.2 We now explain why $\sigma'(0) \neq 0$ can be replaced by the stronger assumption that $\sigma'(0) = 1$ without loss of generality. Let $\sigma : \mathbb{R} \to \mathbb{R}$ be a differentiable function which satisfies our equations (3), (4), and (5), and let $a = \sigma'(0) \neq 0$. Consider the function $\tilde{\sigma} : \mathbb{R} \to \mathbb{R}$ defined by: $\tilde{\sigma}(x) = \sigma(x/a)$. Then also $\tilde{\sigma}$ satisfies our basic equations, and moreover $\tilde{\sigma}'(0) = 1$.

Now, if $\Sigma = (D, A, B, C)_\sigma$ is a σ-system, we may define the new system having $\tilde{\Sigma} = (D, \tilde{A}, \tilde{B}, C)_{\tilde{\sigma}}$ to be the $\tilde{\sigma}$-system with $\tilde{A} = aA$ and $\tilde{B} = aB$. It is clear that $\forall\; u \in \mathbb{R}^m$:

$$Dx + \vec{\sigma}(Ax + Bu) \;=\; Dx + \vec{\tilde{\sigma}}(\tilde{A}x + \tilde{B}u).$$

Thus for any sequences of control values $u_i,\ i = 1, \ldots, k$, the corresponding output sequecences in Σ and in $\tilde{\Sigma}$ will be the same.

Thus, if Σ_1 and Σ_2 are to σ-systems, and we construct $\tilde{\Sigma}_1$ and $\tilde{\Sigma}_2$ in this manner, we have that Σ_1 and Σ_2 are i/o equivalent if and only if $\tilde{\Sigma}_1$ and $\tilde{\Sigma}_2$ are i/o equivalent. Since our interest is in establishing the existence of various linear equations relating A_1 and A_2, B_1 and B_2, and so forth, and since these equations are not changed under multiplication by a scalar, it is clear that we can assume, without loss of generality, that $\sigma'(0) = 1$. So, from now on, when we consider a differentiable function σ, we implicit assume that equation (6) holds instead of equation (4).

2 Equivalence

We fix a function $\sigma : \mathbb{R} \to \mathbb{R}$. For now we only assume that σ is an odd differentiable function. Let π be any permutation of $\{1, \cdots, n\}$ and P be the permutation matrix which represents π; i.e. $P_{i,j} = \delta_{j,\pi(i)}$, where $\delta_{i,k}$ is the Kronecker delta. It is easy to see that: $\vec{\sigma}(Px) = P\vec{\sigma}(x)$ for all $x \in \mathbb{R}^n$.

Moreover, let Q be any diagonal matrix with ± 1 on the diagonal. Then clearly, we also have that: $\vec{\sigma}(Qx) = Q\vec{\sigma}(x)$ for all $x \in \mathbb{R}^n$. Thus if a matrix T is of the form PQ or QP, where P and Q are as above, we have:

$$\vec{\sigma}(Tx) \;=\; T\vec{\sigma}(x) \qquad \text{for all} \quad x \in \mathbb{R}^n. \tag{8}$$

It can be proved that, if σ satisfies our assumptions, also the converse holds (see Lemma 4.2 in [1]). We let

$$\Lambda^n \;=\; \{\; T \mid T = PQ \;\;\text{or}\;\; T = QP \;\}.$$

So Λ^n is the set of those matrices for which equation (8) holds. Notice that Λ^n has cardinality $2^n n!$.

Definition 2.1 Let $\Sigma_1 = (D_1, A_1, B_1, C_1)_\sigma$, $\Sigma_2 = (D_2, A_2, B_2, C_2)_\sigma$ be two σ-systems, and n_1, n_2 be the dimensions of the state spaces in Σ_1, Σ_2 respectively. We say that Σ_1 and Σ_2 are *equivalent* if $n_1 = n_2 = n$, and if there exists an invertible matrix $T \in \Lambda^n$ such that:

$$A_2 = T^{-1} A_1 T,$$

$$D_2 = T^{-1} D_1 T,$$

$$C_2 = C_1 T,$$

$$B_2 = T^{-1} B_1.$$

Given the previous definition, the next property is trivially proved using equation (8).

Proposition 2.2 Let $\Sigma_1 = (D_1, A_1, B_1, C_1)_\sigma$, $\Sigma_2 = (D_2, A_2, B_2, C_2)_\sigma$ be two σ-systems. If they are equivalent then they are also i/o equivalent.

Notice that, clearly this Proposition holds even if the matrix D is not diagonal. Our goal will be to establish that, for a generic subclass of σ-systems, the previous condition is also necessary.

3 Technical Results

In this section we prove some technical facts which will be used later. Here, we denote by σ any differentiable function from $\mathbb{R}$ to itself. The following Lemma is well-known and easy to prove:

Lemma 3.1 Let k be any positive integer, and let $\rho_1, \ldots, \rho_k$ be k positive, distinct real numbers. If the function σ satisfies equation (5), then the functions $\sigma(\rho_1 x), \ldots, \sigma(\rho_k x)$ are linearly independent.

Given any two vectors $v^1, v^2 \in \mathbb{R}^q$, we denote by $< v^1, v^2 >$ their scalar product; i.e. $< v^1, v^2 >= \sum_{i=1}^q v_i^1 v_i^2$.

Lemma 3.2 Let k, q be two positive integers, and $v^1, \ldots, v^k$ be k non-zero vectors in $\mathbb{R}^q$ so that $v^i \neq \pm v^j$ for all $i \neq j$. Assume that the function σ satisfies equation (5), and it is either an odd function or an even function. If

$$\sum_{i=1}^k c_i \sigma(< v^i, u >) \;=\; 0 \quad \forall\, u \in \mathbb{R}^q, \tag{9}$$

then $c_i = 0$ for all $i = 1, \ldots, k$.

Proof. First we define an equivalence relation on the vectors v^i as follows. We say that v^i is equivalent to v^j, and we write $v^i \approx v^j$, if and only if there exists a constant $k_{ij} \in \mathbb{R}$ such that:

$$v^i \;=\; k_{ij} v^j.$$

Note that since $v^i \neq \pm v^j$ for all $i \neq j$, also

$$|k_{ij}| \neq |k_{lj}| \quad \forall\, i \neq l,\ \forall\, j. \tag{10}$$

It is clear that $\approx$ is an equivalence relation, we decompose the set $\{1, \ldots, k\}$ into equivalence classes for $\approx$:

$$\{\ 1, \ldots, k\ \} \ = \ \bigcup_{i=1}^{l} F_i \qquad \text{where if } \ i_1,\, i_2 \in F_i \ \text{ then } \ v^{i_1} \approx v^{i_2}.$$

Consider any fixed equivalence class F_i. Pick any $F_j \neq F_i$. Then there exists a vector $d \in \mathbb{R}^q$ such that:

$$\begin{array}{lll} d \not\perp v^{i_l} & \text{for all} & i_l \in F_i, \\ d \perp v^{j_l} & \text{for all} & j_l \in F_j. \end{array}$$

Taking the derivative of (9) with respect to u_p, we have:

$$\sum_{r=1}^{k} c_r \sigma'(< v^r, u >) v_p^r \ = \ 0 \qquad \forall\ p = 1, \ldots, q.$$

Thus we also have:

$$\sum_{p=1}^{q} [\sum_{r=1}^{k} c_r \sigma'(< v^r, u >) v_p^r] d_p \ = \ 0,$$

which, by interchanging the sums, implies:

$$\sum_{r=1}^{k} c_r \sigma'(< v^r, u >) < v^r, d >= \ 0.$$

By the way we have chosen the vector d, the previous equation implies:

$$\sum_{\substack{r \in \{1, \ldots, k\} \\ r \notin F_j}} c_r \delta_r \sigma'(< v^r, u >) \ = \ 0, \tag{11}$$

where $\delta_r =< v^r, u >$ is nonzero for each $r \in F_i$.

We may now choose another $F_k \neq F_i$, $k \neq j$, and repeat the above argument, starting with (11) instead of (9). This leads to an equation involving second derivatives of σ. Iterating with each possible F_j, we conclude:

$$\sum_{r \in F_i} c_r \tilde{\delta}_r \sigma^{(l-1)}(< v^r, u >) \ = \ 0, \tag{12}$$

where $\tilde{\delta}_r \neq 0$. Fix an index $p \in F_i$. Since $v^p \neq 0$ there exists a j such that $v_j^p \neq 0$. Moreover, since all the vectors in F_i are proportional to v^p, we have that $v_j^r \neq 0$ for all $r \in F_i$. Let $u \in \mathbb{R}^q$ be the vector whose entries are all zero except for the ith one, which we assume to be equal to $\alpha \in \mathbb{R}$. Evaluating equation (12) at this u we have:

$$\sum_{r \in F_i} c_r \tilde{\delta}_r \sigma^{(l-1)}(k_{rp} v_j^r \alpha) \ = \ 0 \quad \forall\ \alpha \in \mathbb{R};$$

which is equivalent to:

$$\sum_{r\in F_i} c_r\tilde{\delta}_r\sigma^{(l-1)}(k_{rp}\beta) \;=\; 0 \quad \forall\, \beta \in \mathbb{R}. \tag{13}$$

Since σ is either an odd or an even function, so is $\sigma^{(l-1)}$; thus we can assume, without loss of generality, $k_{rp} > 0$, for all $r \in F_i$. Also, (10) gives that they are all distinct. So Lemma 3.1 applies, and we have:

$$c_r\tilde{\delta}_r = 0 \quad \forall\, r \in F_i.$$

Since $\tilde{\delta}_r \neq 0$, the previous equation implies $c_r = 0$ for all $r \in F_i$. Since i was arbitrary, our statement follows. ∎

3.1 Basic Identities

Let $\Sigma = (D, A, B, C)_\sigma$ be a σ-system. For any sequence of input values $u_1, \dots, u_k$, we let:

$$\begin{cases} x_0 &= 0 \\ x_i &= Dx_{i-1} + \vec{\sigma}(Ax_{i-1} + Bu_i) \quad i = 1, \dots, k. \end{cases} \tag{14}$$

If $v \in \mathbb{R}^p$ is any vector, we will denote by $\hat{\sigma}(v)$ the following diagonal matrix:

$$\hat{\sigma}(v) = \mathrm{Diag}\,(\sigma'(v_1), \dots, \sigma'(v_p)). \tag{15}$$

Lemma 3.3 Let $\Sigma = (D, A, B, C)_\sigma$ be a σ-system and pick $u_i \in \mathbb{R}^m$, $i = 1, \cdots, k$. Let x_i be the states defined in equation (14), thought-of as functions of $u_1, \dots, u_i$. Then, for all $j = 1, \dots, m$, and for all $i = 1, \dots, k$, we have:

$$\frac{\partial}{\partial u_{1j}} x_i = [D + \hat{\sigma}(Ax_{i-1} + Bu_i)A] \cdots [D + \hat{\sigma}(Ax_1 + Bu_2)A]\hat{\sigma}(Bu_1)B^j \tag{16}$$

where B^j denote the j-th column of the matrix B.

Proof. We will prove the statement by induction on i.

Since, for $l = 1, \dots, n$, $x_{1l} = \sigma(\sum_{i=1}^m b_{ki}u_{1i})$, we have:

$$\frac{\partial}{\partial u_{1j}} x_{1l} = \sigma'(\sum_{i=1}^m b_{ki}u_{1l})b_{kj};$$

which implies

$$\frac{\partial}{\partial u_{1j}} x_1 = \hat{\sigma}(Bu_1)B^j,$$

as desired. Notice that:

$$\begin{aligned} \frac{\partial}{\partial u_{1j}} x_i &= \frac{\partial}{\partial u_{1j}}[Dx_{i-1} + \vec{\sigma}(Ax_{i-1} + Bu_i)] \\ &= [D + \hat{\sigma}(Ax_{i-1} + Bu_i)A]\frac{\partial}{\partial u_{1j}} x_{i-1}. \end{aligned}$$

So, by the inductive assumption, our statement follows. ∎

Next Corollary follows easily from equation (16).

Corollary 3.4 Let $\Sigma = (D, A, B, C)_\sigma$ and $\tilde{\Sigma} = (\tilde{D}, \tilde{A}, \tilde{B}, \tilde{C})_\sigma$ be two σ-systems. Pick any sequences of control values $(u_1, \ldots, u_k)$, and define x_i, and $\tilde{x}_i$, for $i = 1, \ldots, k$, as in equation (14). If the two systems are i/o equivalent, we have that:

$$\begin{array}{l} C[D + \hat{\sigma}(Ax_{k-1} + Bu_k)A] \cdots [D + \hat{\sigma}(Ax_1 + Bu_2)A]\hat{\sigma}(Bu_1)B = \\ \tilde{C}[\tilde{D} + \hat{\sigma}(\tilde{A}x_{k-1} + \tilde{B}u_k)\tilde{A}] \cdots [\tilde{D} + \hat{\sigma}(\tilde{A}x_1 + \tilde{B}u_2)\tilde{A}]\hat{\sigma}(\tilde{B}u_1)\tilde{B}. \end{array} \tag{17}$$

Remark 3.5 Let Σ and $\tilde{\Sigma}$ be as before. Evaluating equation (17) at the sequence of control values $u_i = 0$, for all $i = 1, \ldots, k$, we have:

$$C(D + A)^{k-1}B = \tilde{C}(\tilde{D} + \tilde{A})^{k-1}\tilde{B}. \tag{18}$$

This says that, if two σ-systems are i/o equivalent, then their undelying linear systems are also i/o equivalent.

4 Main Results

Next we will show that the sufficient condition stated in Proposition 2.2 is also necessary. This will hold for generic systems, in a sense to be discussed. Recall that σ is assumed to satisfy (3), (4), and (5).

If W is a matrix, we will denote by W_i the i-th row of W. First, we let

$$\mathcal{F}_{m,p} := \left\{ (B, C) \;\middle|\; \begin{array}{l} B \in \mathbb{R}^{n\times m},\ C \in \mathbb{R}^{p\times n}, \quad \text{for some} \quad n \geq 1, \\ \forall\, i,\ \exists\, j \quad \text{such that} \quad c_{ji} \neq 0 \\ B_i \neq 0, \quad \text{and} \quad B_i \neq \pm B_j \ \forall i \neq j \end{array} \right\} \tag{19}$$

Notice that, when $p = 1$, the second condition says that all the components of the row vector C are nonzero. Moreover, when $m = 1$, the third condition says that all the components of the vector B are nonzero and they have different absolute values.

Proposition 4.1 Let $\Sigma = (D, A, B, C)_\sigma$ and $\tilde{\Sigma} = (\tilde{D}, \tilde{A}, \tilde{B}, \tilde{C})_\sigma$ be two σ-systems; and let n, $\tilde{n}$ be the dimensions of the state spaces of Σ and of $\tilde{\Sigma}$ respectively. Assume that (B, C) and $(\tilde{B}, \tilde{C})$ are in $\mathcal{F}_{m,p}$. If Σ and $\tilde{\Sigma}$ are i/o equivalent then $n = \tilde{n}$, and there exists an unique matrix $T \in \Lambda^n$ such that:

$$\begin{aligned} \tilde{C} &= CT, \\ \tilde{B} &= T^{-1}B. \end{aligned} \tag{20}$$

Proof. If Σ and $\tilde{\Sigma}$ are i/o equivalent, then for all $u \in \mathbb{R}^m$, we must have:

$$C\vec{\sigma}(Bu) = \tilde{C}\vec{\sigma}(\tilde{B}u),$$

that is,

$$\sum_{i=1}^{n} c_{ji}\sigma(< B_i, u >) = \sum_{i=1}^{\tilde{n}} \tilde{c}_{ji}\sigma(< \tilde{B}_i, u >) \quad \forall\, u \in \mathbb{R}^m, \quad \text{and} \quad \forall\, j = 1, \ldots, p. \tag{21}$$

First we want to prove that for each $k = 1, \ldots, n$, there exists an unique index $l(k) \in \{1, \ldots, \tilde{n}\}$ such that:

$$B_k = \beta(k)\tilde{B}_{l(k)} \quad \text{with} \quad \beta(k) = \pm 1. \tag{22}$$

Fix an arbitrary $k = 1, \ldots, n$. Then, by the assumption on the matrix C, there exists an index j such that $c_{jk} \neq 0$. Consider equation (21) for this particular j. Notice that we can rewrite equation (21) as:

$$c_{jk}\sigma(< B_k, u >) + \sum_{\substack{i=1 \\ i \neq k}}^{n} c_{ji}\sigma(< B_i, u >) - \sum_{i=1}^{\tilde{n}} \tilde{c}_{ji}\sigma(< \tilde{B}_i, u >) \; = \; 0. \tag{23}$$

The previous equation is of the same form as equation (9). Moreover, since $c_{jk} \neq 0$, by Lemma 3.2 and by the assumptions on the matrices B and $\tilde{B}$, we can conclude that there exist two indexes i and $l(i)$ such that:

$$B_i \; = \; \beta(i)\tilde{B}_{l(i)} \quad \text{with} \;\; \beta(i) = \pm 1.$$

If $i = k$ then equation (22) holds. If $i \neq k$, then, since σ is odd, we can rewrite equation (23) as:

$$\begin{gathered}(c_{ji} - \beta(i)\tilde{c}_{jl(i)})\sigma(< B_i, u >) + c_{jk}\sigma(< B_k, u >) + \\ \sum_{\substack{r=1 \\ r \neq k,\, r \neq i}}^{n} c_{jr}\sigma(< B_r, u >) - \sum_{\substack{r=1 \\ r \neq l(i)}}^{\tilde{n}} \tilde{c}_{jr}\sigma(< \tilde{B}_r, u >) \; = \; 0.\end{gathered} \tag{24}$$

Now we repeat the above argument, starting with (24) instead of (23). Thus, we will be able to collect two other terms j, $l(j)$ as above, where j is necessarily different from i since $B_i \neq \pm\tilde{B}_r$ for all $r \neq l(i)$. After a finite number of steps, necessarily the pair k, $l(k)$ will be collected, and so (22) is proved.

Notice that equation (22) implies, in particular, that $n \leq \tilde{n}$. By symmetry, we also have $\tilde{n} \leq n$; so we can conclude:

$$n = \tilde{n}.$$

Using equation (22), we can rewrite equation (21) as:

$$\sum_{i=1}^{n}(c_{ji} - \beta(i)\tilde{c}_{jl(i)})\sigma(< B_i, u >) = 0 \quad \forall\; u \in \mathbb{R}^m. \tag{25}$$

Now, we apply again Lemma 3.2 to equation (25), and we get:

$$c_{ji} \; = \; \beta(i)\tilde{c}_{jl(i)}, \quad \forall\; i = 1, \ldots, m. \tag{26}$$

Let now T^{-1} be the matrix PQ, where P is the permutation matrix representing $l(i)$, i.e. $Pe_i = e_{l(i)}$, and $Q = \mathrm{Diag}\,(\beta(1), \ldots, \beta(n))$. Then, it is easy to verify that equations (22) and (26) say that:

$$\tilde{C} \; = \; CT, \quad \text{and} \quad \tilde{B} \; = \; T^{-1}B.$$

Thus, since $T \in \Lambda^n$, our statement is proved. ∎

Notice that the conclusions of the previous Proposition do not depend on the matrices A, D, and $\tilde{A}, \tilde{D}$. Thus, in particular, the previous statement holds for σ-systems where D is an arbitrary matrix, not necessary a diagonal one. However the fact that D is diagonal is crucial in the proof of the next Proposition.

Proposition 4.2 Assume that Σ and $\tilde{\Sigma}$ satisfy the same assumptions as in Proposition 4.1. If the function σ satisfies also assumption (A1), then, for the matrix $T \in \Lambda^n$ found in Proposition 4.1, the following equation holds:

$$\tilde{D} \;=\; T^{-1} D T. \tag{27}$$

Proof. Let $D = \text{Diag}\,(d_1, \ldots, d_n)$, and $\tilde{D} = \text{Diag}\,(\tilde{d}_1, \ldots, \tilde{d}_n)$. Because $\tilde{D}$ is diagonal and since $T \in \Lambda^n$,

$$(T\tilde{D}T^{-1})_{ij} \;=\; \tilde{d}_{l(i)}\delta_{ij},$$

where with $l(i)$ we have denoted the permutation found in the previous Proposition. Thus to prove our statement, we need to see that:

$$d_i \;=\; \tilde{d}_{l(i)} \quad \forall\, i = 1, \ldots, n. \tag{28}$$

The assumptions $B_i \neq 0$, and $\tilde{B}_i \neq 0$ for all $i = 1, \ldots, n$, guarante that there exists some $u_2 \in \mathbb{R}^m$ such that:

$$(Bu_2)_i \neq 0, \;\; \text{and} \;\; (\tilde{B}u_2)_i \neq 0 \quad \forall\, i = 1, \ldots, n. \tag{29}$$

from Corollary 3.4 (with $k = 2$), we have, for each $\alpha \in \mathbb{R}$:

$$C[D + \hat{\sigma}(A\vec{\sigma}(Bu) + \alpha Bu_2 A]\hat{\sigma}(Bu)B \;=\; \tilde{C}[\tilde{D} + \hat{\sigma}(\tilde{A}\vec{\sigma}(\tilde{B}u) + \alpha\tilde{B}u_2)\tilde{A}]\hat{\sigma}(\tilde{B}u)\tilde{B}.$$

Taking the limit as $\alpha \to \infty$ in the previous equation, since σ satisfies assumption (A1), and by equation (29), we have:

$$CD\hat{\sigma}(Bu)B \;=\; \tilde{C}\tilde{D}\hat{\sigma}(\tilde{B}u)\tilde{B}, \quad \forall\, u \in \mathbb{R}^m.$$

We rewrite the previous equation as:

$$\sum_{k=1}^{n} c_{jk} d_k b_{ki} \sigma'(< B_k, u >) \;=\; \sum_{k=1}^{n} \tilde{c}_{jk} \tilde{d}_k \tilde{b}_{ki} \sigma'(< \tilde{B}_k, u >), \tag{30}$$

where B_i and $\tilde{B}_i$ are the i-th rows of the matrices B and $\tilde{B}$ respectively; the previous equation holds for all $j = 1, \ldots, p$, and for all $i = 1, \ldots, m$. Let T be the matrix found in Proposition 4.1. Using equation (22), and the fact that σ' is an even function, we can rewrite equation (30) as:

$$\sum_{k=1}^{n} (c_{jk} d_k b_{ki} - \tilde{c}_{jl(k)} \tilde{d}_{l(k)} \tilde{b}_{l(k)i}) \sigma'(< B_k, u >) \;=\; 0, \quad \forall\, u \in \mathbb{R}^m.$$

Now by applying Lemma 3.2 to the previous equation, we get that for all $j = 1, \ldots, p$, for all $k = 1, \ldots, n$, and for all $i = 1, \ldots, m$, the following equation holds:

$$c_{jk} d_k b_{ki} \;=\; \tilde{c}_{jl(k)} \tilde{d}_{l(k)} \tilde{b}_{l(k)i}.$$

Fix an index k. Since $B_k \neq 0$, there exists an index i such that $b_{ki} \neq 0$, which, in particular implies also $\tilde{b}_{l(k)i} = \beta(k) b_{ki} \neq 0$. Moreover, there also exists an index j such that $c_{jk} \neq 0$, which implies that $\tilde{c}_{jl(k)} = \beta(k) c_{jk} \neq 0$. Thus from the previous equation we have:

$$d_k \;=\; \tilde{d}_{l(k)} \quad \forall\, k = 1, \ldots, n,$$

as desired. ∎

The next Theorem proves that for a generic subclass of σ-systems the condition of Proposition 2.2 is also necessary. By a *generic* subset of $\mathbb{R}^N$ we mean a nonempty subset of $\mathbb{R}^N$ whose complement is the set of zeroes of a finite number of polynomials in N variables.

We let:

$$\tilde{\mathcal{S}}_{n,m,p} \;=\; \left\{ \; (D,A,B,C)_\sigma \;\middle|\; \begin{array}{l} (B,C)\in\mathcal{F}_{m,p}, \\ (A+D,B) \;\; \text{controllable pair.} \end{array} \right\}.$$

Notice that $\tilde{\mathcal{S}}_{n,m,p}$ is a generic subset of the set of all σ-systems, when we identify the latter with $\mathbb{R}^{n^2+n+np+mn}$.

Theorem 1 *Let* $\Sigma = (D,A,B,C)_\sigma \in \tilde{\mathcal{S}}_{n,m,p}$ *and* $\tilde{\Sigma} = (\tilde{D},\tilde{A},\tilde{B},\tilde{C})_\sigma \in \tilde{\mathcal{S}}_{\tilde{n},m,p}$ *be two σ-systems. Assume that the function σ satisfies assumption (A1) (as well as (3), (4), and (5)). Then the two systems are i/o equivalent if and only if they are equivalent.*

Proof. The sufficiency part is given by Proposition 2.2, thus we need only to prove the necessity part. from Propositions 4.1 and 4.2, we already know that $n=\tilde{n}$, and there exists a matrix $T\in\Lambda^n$ such that:

$$\tilde{C}=CT,\;\; \tilde{B}=T^{-1}B, \;\;\text{and}\;\; \tilde{D}=T^{-1}DT.$$

So to prove equivalence, we need only to see that:

$$\tilde{A} \;=\; T^{-1}AT. \tag{31}$$

By evaluating equation (17) at $u_1=\ldots=u_{k-1}=0$, and $u_k=u$, we have:

$$C(D+\hat{\sigma}(Bu)A)(D+A)^{k-2}B \;=\; \tilde{C}(\tilde{D}+\hat{\sigma}(\tilde{B}u)\tilde{A})(\tilde{D}+\tilde{A})^{k-2}\tilde{B} \;\;\forall\, k\ge 2.$$

Taking limits as in the previous proof, since $B_i\neq 0$, and $\tilde{B}_i\neq 0$ for all $i=1,\ldots,n$, and σ satisfies assumption (A1), the previous equation implies, in particular, that:

$$CD(D+A)^{k-2}B \;=\; \tilde{C}\tilde{D}(\tilde{D}+\tilde{A})^{k-2}\tilde{B},$$

and therefore also:

$$C\hat{\sigma}(Bu)A(D+A)^{k-2}B \;=\; \tilde{C}\hat{\sigma}(\tilde{B}u)\tilde{A}(\tilde{D}+\tilde{A})^{k-2}\tilde{B} \quad \forall\, k\ge 2,\; \forall\, u\in\mathbb{R}^m. \tag{32}$$

Claim: If for some matrices $M,\ N\in\mathbb{R}^{n\times m}$, we have

$$C\hat{\sigma}(Bu)M \;=\; \tilde{C}\hat{\sigma}(\tilde{B}u)N, \tag{33}$$

for all $u\in\mathbb{R}^m$, then $M=TN$.

We will prove this Claim below; first we show how to derive equation (31) from it.

By applying the Claim to equation (32), we have:

$$A(D+A)^{k-2}B \;=\; T\tilde{A}(\tilde{D}+\tilde{A})^{k-2}\tilde{B} \quad \forall\, k\ge 2. \tag{34}$$

Now, we prove by induction on r, the following equation:

$$(\tilde{D}+\tilde{A})^r T^{-1}B \;=\; T^{-1}(D+A)^r B. \tag{35}$$

If $r = 0$ there is nothing to prove, so we assume $r > 0$. We have:

$$(\tilde{D} + \tilde{A})^r T^{-1} B = [\tilde{D}(\tilde{D} + \tilde{A})^{r-1} T^{-1} B + \tilde{A}(\tilde{D} + \tilde{A})^{r-1} T^{-1} B].$$

Using equation (34) and the inductive assumption we have:

$$(\tilde{D} + \tilde{A})^r T^{-1} B = [\tilde{D} T^{-1} (D + A)^{r-1} B + T^{-1} A (D + A)^{r-1} B].$$

Since $\tilde{D} T^{-1} = T^{-1} D$, the previous equation gives us the inductive step.

Combining now equation (34) with equation (35), we have:

$$A(A + D)^{k-2} B = T \tilde{A} T^{-1} (A + D)^{k-2} B \quad \forall\ k \geq 2.$$

Since the pair $(A + D, B)$ is controllable, the previous equation implies (31), as desired.

To complete our proof, now we establish the Claim.

Equation (33) says:

$$\sum_{k=1}^{n} c_{jk} m_{ki} \sigma'(< B_k, u >) \ = \ \sum_{k=1}^{n} \tilde{c}_{jk} n_{ki} \sigma'(< \tilde{B}_k, u >),$$

for all $j = 1, \ldots, p$, and all $i = 1, \ldots, m$. Notice that this equation is of the same type as equation (30). Thus arguing as in that case, we have:

$$c_{jk} m_{ki} \ = \ \tilde{c}_{jl(k)} n_{l(k)i} \quad \forall\ j,\ k,\ i.$$

Fix two arbitrary indexes k, i. By the assumption on the matrix C, there exists j such that $c_{jk} \neq 0$, which, in turn, implies $\tilde{c}_{jl(k)} = \beta(k) c_{jk} \neq 0$. So, using the previous equation, we have:

$$m_{ki} \ = \ \beta(k) n_{l(k)i} \quad \forall\ k,\ i;$$

which is equivalent to $M = TN$, as desired. ∎

References

[1] Albertini, F., and E.D. Sontag, "For neural networks, function determines form," *Proc. IEEE Conf. Decision and Control, Tucson, Dec. 1992*, IEEE Publications, 1992, to appear. Expanded version with same title, Report SYCON-92-03, Rutgers Center for Systems and Control; available from SCAD database (ftp ftp.utdallas.edu, login anonymous, dir pub/scad/preprints/syconreports), or by e-mail from sontag@control.rutgers.edu.

[2] Wang, X., and E.K. Blum, "Discrete-time versus continuous-time models of neural networks," *J. Comp. Syst. Sci.* **45**(1992): 1-19.