Global Convergence of Neural Networks With Discontinuous Neuron Activations

M. Forti and P. Nistri

***Abstract*—The paper introduces a general class of neural networks where the neuron activations are modeled by discontinuous functions. The neural networks have an additive interconnecting structure and they include as particular cases the Hopfield neural networks (HNNs), and the standard cellular neural networks (CNNs), in the limiting situation where the HNNs and CNNs possess neurons with infinite gain. Conditions are derived which ensure the existence of a unique equilibrium point, and a unique output equilibrium point, which are globally attractive for the *state* and the *output* trajectories of the neural network, respectively. These conditions, which are applicable to general nonsymmetric neural networks, are based on the concept of Lyapunov diagonally-stable neuron interconnection matrices, and they can be thought of as a generalization to the discontinuous case of previous results established for neural networks possessing smooth neuron activations. Moreover, by suitably exploiting the presence of sliding modes, entirely new conditions are obtained which ensure global convergence in *finite time*, where the convergence time can be easily estimated on the basis of the relevant neural-network parameters. The analysis in the paper employs results from the theory of differential equations with discontinuous right-hand side as introduced by Filippov. In particular, global convergence is addressed by using a Lyapunov-like approach based on the concept of monotone trajectories of a differential inclusion.**

***Index Terms*—Convergence in finite time, discontinuous neuron activations, Filippov solutions, generalized gradient, global convergence, neural networks.**

NOTATION

$\mathbb{R}^n$	Real n space.
A	$= [A_{ij}] \in \mathbb{R}^{n\times n}$, Square matrix.
A'	Transpose of A.
A^{-1}	Inverse of A.
A^S	$= (1/2)(A + A')$, Symmetric part of A.
$\ker A$	Kernel of A.
$\Lambda_m\{A\}$	Minimum eigenvalue of the symmetric matrix A.
$\Lambda_M\{A\}$	Maximum eigenvalue of the symmetric matrix A.
$\|A\|$	Matrix norm of A.
$\|A\|_2$	$= [\Lambda_M\{A'A\}]^{1/2}$, 2 norm of A.
α	$= \mathrm{diag}(\alpha_1, \cdots, \alpha_n) \in \mathbb{R}^{n\times n}$ diagonal matrix with diagonal entries α_i, $i = 1, \cdots, n$.
x	$= (x_1, \cdots, x_n)' \in \mathbb{R}^n$, Column vector.
$\|x\|$	Vector norm of x.
$\|x\|_2$	$= [\sum_{i=1}^n x_i^2]^{1/2}$, Euclidean norm of x.
$\overline{E}$	Closure of set $E \subset \mathbb{R}^n$.
$\mathcal{B}(x, r)$	$= \{y \in \mathbb{R}^n : \|y - x\|_2 < r\}$, Ball with center $x \in \mathbb{R}^n$ and radius r.
$\mathrm{int}(E)$	Interior of E.
∂E	Boundary of E.
$\mathcal{C}_{\mathbb{R}^n} E$	Complementary set of E with respect to $\mathbb{R}^n$.
$\mathrm{co}(E)$	Convex hull of E.
$K[E]$	$= \overline{\mathrm{co}}(E)$, Closure of the convex hull of E.
$\mu(E)$	Lebesgue measure of E.
$\emptyset$	Empty set.
$\nabla V(x)$	Gradient of $V(x)$: $\mathbb{R}^n \to \mathbb{R}$.
$\partial V(x)$	Clarke's generalized gradient of $V(x)$: $\mathbb{R}^n \to \mathbb{R}$.

Manuscript received December 6, 2002; revised April 7, 2003. This paper was recommended by Associate Editor D. Leenaerts.

The authors are with the Dipartimento di Ingegneria dell'Informazione, Università di Siena, 56-53100 Siena, Italy (e-mail: forti@dii.unisi.it; pnistri@dii.unisi.it).

Digital Object Identifier 10.1109/TCSI.2003.818614

I. INTRODUCTION

A BRIEF overview on some common neural-network models easily reveals that neural networks described by differential equations with a discontinuous right-hand side are of importance and do frequently arise in practice. Let us consider, for example, the classical Hopfield neural networks (HNNs) with graded response neurons [1]. The standard assumption is that the activations be employed in the high-gain limit where they closely approach a discontinuous hard comparator function. As shown by Hopfield [1], [2], the high-gain hypothesis is crucial to make negligible the contribution to the neural-network energy function of the term depending on the neuron self inhibitions, and to favor binary output formation. A conceptually analogous model based on hard comparators is used also to describe the discrete-time cellular neural networks (CNNs) [3]. Another important example concerns the class of neural networks introduced by Kennedy and Chua [4] to solve linear and nonlinear programming problems. Those networks exploit constraint neurons with a diode-like input–output activations. Again, in order to guarantee satisfaction of the constraints, the diodes are required to possess a very high slope in the conducting region, i.e., they should approximate the discontinuous characteristic of an ideal diode [5].

When dealing with dynamical systems possessing high-slope nonlinear elements it is often advantageous to model them with a system of differential equations with discontinuous right-hand side, rather than studying the case where the slope is high but of finite value [6]. The main advantage of analyzing the ideal discontinuous case is that such analysis is usually able to give a clear picture of the salient features of motion, such as the presence of *sliding modes*, i.e., the possibility that trajectories be confined for some time intervals on discontinuity surfaces. Another intriguing phenomenon which is peculiar to discontinuous systems concerns the possibility that trajectories converge toward an equilibrium point in *finite time* [7], [8]. This is basically in contrast with the case of smooth dynamical systems, where there can only be asymptotic convergence toward the equilibrium.

In recent literature, one of the most investigated dynamical issues has been to obtain conditions ensuring that a neural network possesses a unique equilibrium point which is globally attractive for the trajectories, see, e.g., [9]–[16], and references therein. The property of global convergence prevents a neural network from the risk of getting stuck at some local minimum of the energy function. Said another way, a globally attractive neural network is well suited to solve global optimization problems in real time, as has been demonstrated in a number of practical applications reported in the literature [12], [14], [17].

All the quoted results on global convergence concern neural networks where the neuron activations are modeled by Lipschitz continuous functions. Thus, they leave open the issue of global convergence in the limiting ideal case of discontinuous neuron activations. As was noted before, this issue gained interest since discontinuous neural networks are frequently encountered in practice. Moreover, the analysis of the ideal discontinuous case might reveal the existence of phenomena, such as the presence of sliding modes and the global convergence in finite time, which are known to be potentially useful for the design of real-time optimization solvers [8], [18], [19].

On the basis of the above discussion, in this paper, we introduce a general class of neural networks which possess discontinuous neuron activations (Section II). The neural networks have an additive interconnecting structure and they include as particular cases the HNNs [1], and the CNNs [20], in the limiting situation where the HNNs and CNNs are modeled by neurons with infinite gain. The paper establishes conditions for the existence of a unique equilibrium point, and a unique output equilibrium point (Section III), and then addresses the issue of global convergence toward the unique equilibrium of the state and the output trajectories (Section IV). The conditions on global convergence, which are applicable to general nonsymmetric neuron interconnection matrices, are based on the concept of Lyapunov diagonally-stable (LDS) neuron interconnection matrices, and they generalize to the discontinuous case previous results for neural networks possessing smooth neuron activations [12]. Moreover, by suitably exploiting the presence of sliding modes, we establish entirely new conditions ensuring global convergence in finite time, where the convergence time can be easily estimated on the basis of the relevant neural-network parameters (Section V). The main conclusions drawn in the paper are summarized in Section VI.

The existing literature reports a few other investigations on discontinuous neural networks, which pertain to a different application context, or to different neural architectures. A significant case is that of HNNs where neurons are modeled by a hard discontinuous comparator function [21]. Different from the case addressed in this paper, the analysis in [21] is valid for symmetric neural networks which possess multiple equilibrium points located in saturation regions, i.e., networks useful to implement content addressable memories. Moreover, at the basis of those results there are conditions which rule out the presence of sliding modes (cf. [21, Theorem 4.6]). Another recent paper has introduced a special neural-like architecture for solving linear programming problems [18] (see also [19]). The architecture substantially differs from the additive neural networks considered in the present study. Moreover, the networks in [18] are designed as gradient systems of a suitable energy function, while it is known that additive neural networks of the Hopfield type are gradient systems only under the restrictive assumption of symmetry of the neuron interconnection matrix [12], [22].

To study the class of discontinuous neural networks in this paper we use concepts from the theory of differential equations with discontinuous right-hand side as introduced by Filippov [23]. This theory has become a standard mathematical tool in a number of engineering applications. In particular, global attractivity of the neural network is addressed via a Lyapunov like approach based on the concept of monotone trajectories of a differential inclusion, see e.g., [7, ch. 5].

A. Preliminaries

Consider a set $E \subset \mathbb{R}^n$. The smallest convex set containing E is said to be the convex hull of E, and is denoted by $\mathrm{co}(E)$. Such a set $\mathrm{co}(E)$ always exists and is the intersection of all convex sets containing E. The following properties hold: $\mathrm{co}(aE) = a\mathrm{co}(E)$ for any $a \in \mathbb{R}$ and $E \subset \mathbb{R}^n$; $\mathrm{co}(E_1 + E_2) = \mathrm{co}(E_1) + \mathrm{co}(E_2)$, for any $E_1, E_2 \subset \mathbb{R}^n$. The closure of the convex hull of E is the intersection of all closed convex sets containing E, and is denoted by $\overline{\mathrm{co}}(E)$, or simply by $K[E]$.

Suppose that to each point x of a set $E \subset \mathbb{R}^n$ there corresponds a nonempty set $F(x) \subset \mathbb{R}^n$. Then, $x \multimap F(x)$ is a set-valued map from $E \multimap \mathbb{R}^n$. A set-valued map F with nonempty values is said to be upper semicontinuous at $x_0 \in E$ if, for any open set $\mathcal{N}$ containing $F(x_0)$, there exists a neighborhood $\mathcal{M}$ of x_0 such that $F(\mathcal{M}) \subset \mathcal{N}$. If E is closed, F has nonempty closed values, and F is bounded in a neighborhood of each point $x \in E$, then F is upper semicontinuous on E if and only if its graph $\{(x, y) \in E \times \mathbb{R}^n : y \in F(x)\}$ is closed.

Let $V\colon \mathbb{R}^n \to \mathbb{R}$ be a locally Lipschitz continuous function. The Clarke's generalized gradient of V at $x \in \mathbb{R}^n$ [24] is defined by

$$\partial V(x) = K\left[\lim \nabla V(x_i) : x_i \to x, x_i \notin \Omega_V \cup \mathcal{N}\right]$$

where $\Omega_V \subset \mathbb{R}^n$ is the set of Lebesgue measure zero where ∇V does not exist, and $\mathcal{N} \subset \mathbb{R}^n$ is an arbitrary set with measure zero.

The next definitions on matrices will be employed in the paper.

Definition 1 [25]: Matrix $A \in \mathbb{R}^{n\times n}$ is said to belong to the class P if and only if all the principal minors of A are positive.

Definition 2 [26]: Matrix $A \in \mathbb{R}^{n\times n}$ is said to be LDS if and only if there exists a diagonal and positive definite matrix $\alpha \in \mathbb{R}^{n\times n}$ such that the symmetric part of αA

$$[\alpha A]^S = \frac{1}{2}(\alpha A + A'\alpha)$$

is positive definite.

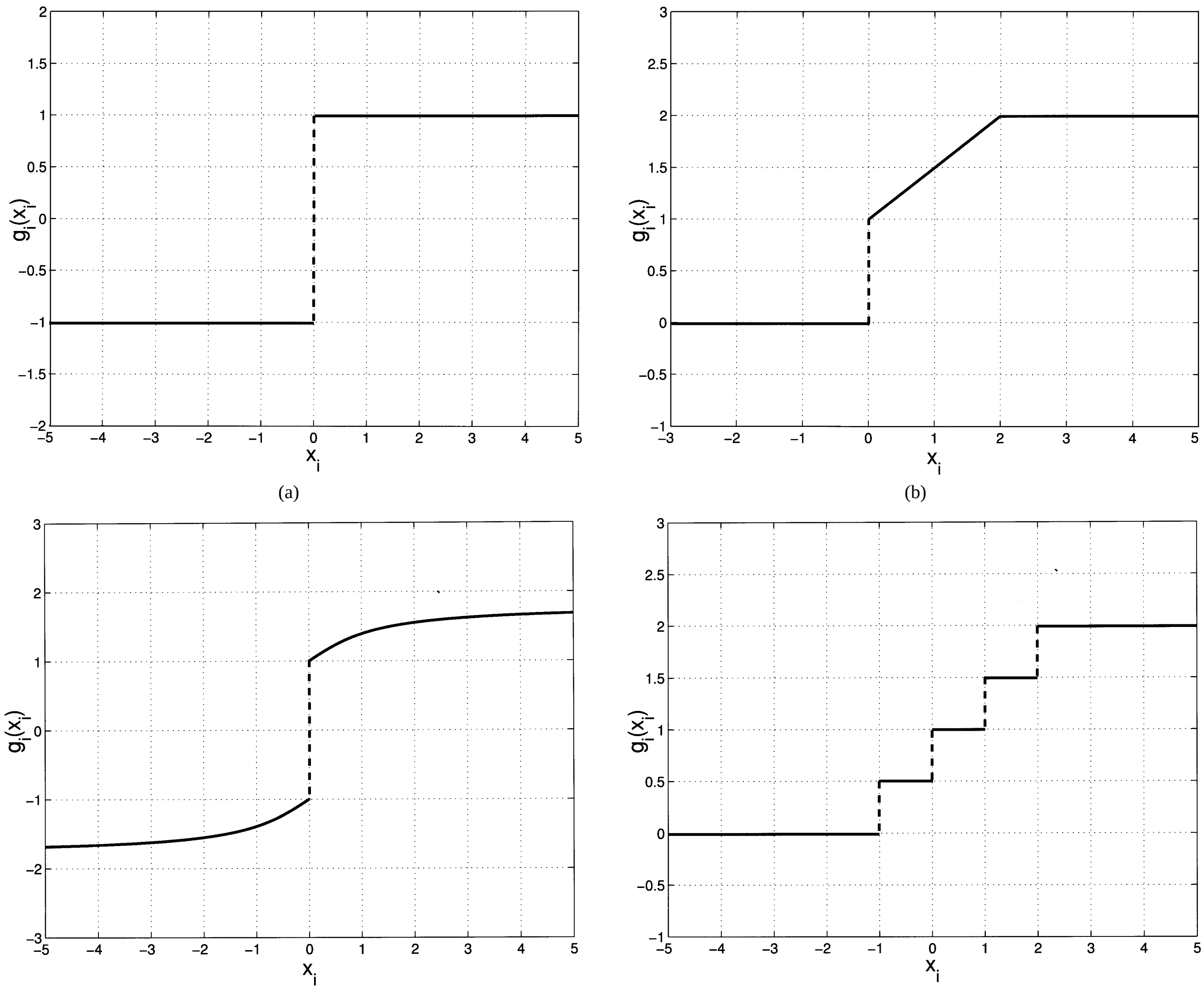


Fig. 1. Examples of discontinuous functions in the class $\mathcal{G}_{\mathcal{D}}$: (a) hard comparator (signum) function; (b) piecewise linear monotonically nondecreasing function; (c) discontinuous strictly increasing function; and (d) multilevel function.

II. Neural-Network Model and Problem Formulation

This section introduces the neural-network model with discontinuous neuron activations which is dealt with in the paper (Section II-A). Then, the global convergence problem addressed in the paper is discussed (Section II-B).

A. Neural-Network Model

We consider neural networks described by the system of differential equations

$$\dot{x} = f(x) = Bx + Tg(x) + I \tag{N1}$$

where f is a vector-valued function, $x = (x_1, \cdots, x_n)' \in \mathbb{R}^n$ is the vector of the neuron states, $B = \mathrm{diag}(-b_1, \cdots, -b_n) \in \mathbb{R}^{n \times n}$ is a diagonal matrix such that $b_i > 0, i = 1, \cdots, n$, model the neuron self-inhibitions, $T \in \mathbb{R}^{n \times n}$ is a matrix whose entries represent the synaptic neuron interconnections, and $I \in \mathbb{R}^n$ is a vector of constant neuron inputs.

The diagonal mapping $g(x) = (g_1(x_1), \cdots, g_n(x_n))'$ has components g_i that model the nonlinear input-output activations of the neurons. In the paper, we assume that g belongs to the following class of discontinuous functions.

Definition 3 (Function Class $\mathcal{G}_{\mathcal{D}}$): We say that $g \in \mathcal{G}_{\mathcal{D}}$ if and only if, for $i = 1, \cdots, n$, g_i satisfies the following assumptions.

a) g_i is piecewise continuous, i.e., g_i is continuous in $\mathbb{R}$ except a countable set of points of discontinuity, ρ_k, where there exist finite right and left limits, $g_i(\rho_k^+)$ and $g_i(\rho_k^-)$, respectively, with $g_i(\rho_k^+) > g_i(\rho_k^-)$; moreover, g_i has a finite number of discontinuities on any compact interval of $\mathbb{R}$.
b) g_i is bounded.
c) g_i is nondecreasing, i.e., for any ρ_a and ρ_b such that $\rho_a > \rho_b$ and g_i is continuous at ρ_a and ρ_b, it results $g_i(\rho_a) \geq g_i(\rho_b)$.

Note from Definition 3 that each function g_i is undefined at the points of discontinuity ρ_k. The class of discontinuous functions $\mathcal{G}_{\mathcal{D}}$ includes a number of neuron activations of interest for the applications. For example, the standard hard comparator function $g_i(x_i) = \mathrm{sgn}(x_i)$ ($\mathrm{sgn}(x_i) = 1$ if $x_i > 0$, $\mathrm{sgn}(x_i) = -1$ if $x_i < 0$, and $\mathrm{sgn}(0)$ undefined), and the multilevel activations, see Fig. 1.

We remark that in the special case where $g_i(x_i) = \mathrm{sgn}(x_i)$, $i = 1, \cdots, n$, we may consider model (N1) as the limit of a HNN where the maximum gain of the neuron activations tends to infinity. It is of importance to note that in the actual applications, the Hopfield networks are indeed usually employed in such a "high-gain" limit situation [1]. Moreover, the discrete-time CNNs introduced in [3] exploit similar neurons with infinite gain.

Since for $g \in \mathcal{G}_\mathcal{D}$ the right-hand side of (N1) is a discontinuous function of the state, x, it is needed to explain what is meant by a solution of a Cauchy problem associated to (N1). A possible definition, which we shall adopt in this paper, is that of Filippov [23].

Definition 4 [23]: Let us consider the set-valued map ϕ: $\mathbb{R}^n \multimap \mathbb{R}^n$ defined as

$$\phi(x) = \bigcap_{\epsilon>0} \bigcap_{\mu(\mathcal{N})=0} K\left[f\left(\mathcal{B}(x,\epsilon) \setminus \mathcal{N}\right)\right]$$

where $\mathcal{N}$ is an arbitrary set with measure zero. A solution of (N1) on an interval $[t_0, t_1]$, $t_0 \le t_1 \le +\infty$, with initial condition $x(t_0) = x_0$, is an absolutely continuous function $x(t)$ defined on $[t_0, t_1]$, such that $x(t_0) = x_0$, and for almost all (a.a.) $t \in [t_0, t_1]$, $x(t)$ satisfies the differential inclusion $\dot{x}(t) \in \phi(x(t))$.

The usefulness in view of the engineering applications of the concept of solutions in the sense of Filippov derives from the fact that they are good approximations of solutions of actual systems with high-gain nonlinearities [6]–[8].

Under the stated assumptions on g, i.e., $g \in \mathcal{G}_\mathcal{D}$, by accounting for the properties of the convex hull, it is easy to verify that $\phi(x) = K[Bx + Tg(x) + I] = Bx + TK[g(x)] + I$, where $K[g(x)] = (K[g_1(x_1)], \cdots, K[g_n(x_n)])'$, and for $i = 1, \cdots, n$

$$K\left[g_i(x_i)\right] = \left[g_i\left(x_i^-\right), g_i\left(x_i^+\right)\right]. \tag{1}$$

Therefore, it follows that the set-valued map $x \multimap \phi(x)$ has nonempty compact convex values. Moreover, it is upper semicontinuous (see [23, p. 67, Lemma 1]) and hence it is measurable. By the measurable selection theorem (see [27, p. 308, Th. 8.1.3]) we have that if $x = x(t)$ is a solution of (N1) on $[t_0, t_1]$, where $t_0 < t_1 \le +\infty$, then, there exists a bounded measurable function $\gamma(t)$ such that for a.a. $t \in [t_0, t_1]$, it results in

$$\gamma(t) \in K\left[g\left(x(t)\right)\right] \quad \dot{x}(t) = Bx(t) + T\gamma(t) + I. \tag{2}$$

Function $\gamma(t)$ represents the neural-network output on $[t_0, t_1]$.

Property 1: Suppose that $g \in \mathcal{G}_\mathcal{D}$ and $\det T \neq 0$. Let $x(t)$ be a solution of (N1) on $[t_0, t_1]$, where $t_0 < t_1 \le +\infty$. Then: a) the neural-network output $\gamma(t)$ in (2) is a bounded measurable function that is uniquely defined by $x(t)$ up to a set of measure zero in $[t_0, t_1]$; b) $\gamma(t)$ is defined and continuous for all $t \in [t_0, t_1]$ if and only if $x(t)$ is continuously differentiable for all $t \in [t_0, t_1]$.

Proof: a) If $\det T \neq 0$, from (2) we get $\gamma(t) = T^{-1}(\dot{x}(t) - Bx(t) - I)$. Since $x(t)$ is absolutely continuous, $x(t)$ is differentiable for a.a. $t \in [t_0, t_1]$. Hence, we have the result. b) Obvious. ■

The next result addresses the existence of solutions for (N1).

Property 2: Suppose that $g \in \mathcal{G}_\mathcal{D}$. Then, for any $x_0 \in \mathbb{R}^n$ there is at least a local solution $x(t)$ of (N1) with initial condition $x(0) = x_0$. Furthermore, any solution is bounded and hence defined on $[0, +\infty)$.

Proof: The local existence of a solution $x(t)$ to (N1) on $[0, t_0]$, $t_0 > 0$, with $x(0) = x_0$, is a straightforward consequence of [23, p. 77, Th. 1]. Moreover, since g is bounded on $\mathbb{R}^n$ and hence also $K[g(x)]$ is bounded on $\mathbb{R}^n$, it is seen that if $\tilde{R} > 0$ is sufficiently large, then it results $\mathrm{sgn}(v_i) = \mathrm{sgn}(-b_i x_i)$, $i = 1, \cdots, n$, for $\|x\|_2 \ge \tilde{R}$ and for any $v \in \phi(x) = Bx + TK[g(x)] + I$. Said another way, ϕ points toward the interior on the boundary of a sufficiently large sphere. Therefore, if we let $\overline{R} = \max\{\|x_0\|_2, \tilde{R}\}$, it follows that $\|x(t)\|_2 \le \overline{R}$ on $[0, +\infty)$. This means that $x(t)$ is bounded and hence defined on $[0, +\infty)$. ■

By an *equilibrium point* (EP) $e \in \mathbb{R}^n$ of (N1), we mean a constant solution of (N1), $x(t) = e$, $t \in [0, +\infty)$. Clearly, e is an EP of (N1) if and only if [7]

$$0 \in \phi(e) = Be + TK\left[g(e)\right] + I. \tag{3}$$

If e is an EP of (N1), from (3) it turns out that there exists a vector $\eta \in \mathbb{R}^n$ such that

$$\eta \in K\left[g(e)\right] \quad 0 = Be + T\eta + I. \tag{4}$$

We say that η is an *output equilibrium point* (OEP) of (N1) corresponding to e.

Proposition 1: Suppose that $g \in \mathcal{G}_\mathcal{D}$ and $\det T \neq 0$. Let $e \in \mathbb{R}^n$ be an EP of (N1). Then, there exists a unique OEP η of (N1) corresponding to e.

Proof: Let e be an EP of (N1), and η be a corresponding OEP of (N1) as defined in (4). Then, $\eta = -T^{-1}(Be + I) \in K[g(e)]$ turns out to be the unique OEP of (N1) corresponding to e. ■

When $\det T = 0$ the results given in Proposition 1 and Property 1 are in general false, as it is illustrated in the next example.

Example 1: Let us consider the second-order neural network

$$\begin{cases} \dot{x}_1 = -x_1 - \mathrm{sgn}(x_1) - \mathrm{sgn}(x_2) \\ \dot{x}_2 = -x_2 - \mathrm{sgn}(x_1) - \mathrm{sgn}(x_2) \end{cases}$$

where $\mathrm{sgn}(\rho) = 1$ for $\rho > 0$, $\mathrm{sgn}(\rho) = -1$ for $\rho < 0$, and $\mathrm{sgn}(0)$ undefined. It results $\det T = 0$, and from (1) we have $\phi_i(0) = [-2, 2]$, $i = 1, 2$, since $K[\mathrm{sgn}(0)] = [-1, 1]$. Hence, $0 \in \phi(0)$, i.e., on the basis of (3), $e = 0$ is an EP of the neural network. Moreover, it can be verified that the equilibrium point is unique. However, the infinitely many vectors $\eta = (\eta_1, \eta_2)'$ with $\eta_1 = -\eta_2$, and $\eta_1 \in [-1, 1]$, are OEPs of (N1) corresponding to the EP $e = 0$.

Now, consider the equilibrium solution $x(t) = e = 0$, $t \ge 0$, and the corresponding output $\gamma(t) = \eta = 0$, which satisfies (2). Then, also $\eta + \hat{\gamma}(t)$, where $\hat{\gamma}(t)$ is an arbitrary measurable function on $[0, +\infty)$ such that $\hat{\gamma}(t) \in \ker T \cap K[g(0)] = \{(y_1, y_2)' \in \mathbb{R}^2 : |y_1|, |y_2| \le 1, y_1 = -y_2\}$, satisfies (2). This implies that the neural-network output is not uniquely defined by the neuron state.[1] ■

[1]This situation is not acceptable, since in this paper we are interested in conditions ensuring convergence of the neural-network output.

B. Problem Formulation

The goal of this paper is to find conditions on the neuron interconnection matrix T, which ensure that (N1) has a unique EP e, and a unique corresponding OEP η. Moreover, we are interested in the case where e and η are globally attractive for the neural-network state and output trajectories, respectively.

It is of importance to remark that while global attractivity of an EP e can be defined in the usual way (cf. Definition 5 below), there are difficulties to define an analogous concept for the OEP η. In fact, the measurable function $\gamma(t)$, which represents the neuron outputs (cf. (2)), is in general defined only for a.a. $t \in [0, +\infty)$, so that it is not always possible to consider its limit as $t \to +\infty$, as for a continuous function. To overcome this difficulty, we introduce a weaker notion of limit, which is applicable to any measurable function $\gamma(t)$, namely the limit in measure of $\gamma(t)$ (see Definition 6 below). Then, we show that this notion is as useful as the usual concept of limit in view of the neural-network applications. Moreover, we point out situations of practical interest where the standard concepts of limit and global attractivity can be employed also for the outputs $\gamma(t)$. Finally, an important case is singled out, i.e., that where there is global convergence of the state and the output trajectories in *finite time*.

Let us recall the standard definition of global attractivity of an EP of (N1).

Definition 5: The EP e of (N1) is said to be globally attractive (GA) if, for each $x_0 \in \mathbb{R}^n$, any trajectory $x(t)$ of (N1), $t \in [0, +\infty)$, with $x(0) = x_0$, is such that $\lim_{t\to+\infty} x(t) = e$.

Let us now address the definition of global attractivity of an OEP of (N1). To this end, we introduce the following concepts.

Definition 6: Let $\gamma(t)\colon [0, +\infty) \to \mathbb{R}^n$ be a measurable function, and $\eta \in \mathbb{R}^n$. We will say that η is the limit in measure of $\gamma(t)$ as $t \to +\infty$, and we write $\eta = \mu\lim_{t\to+\infty} \gamma(t)$, if

$$\forall \epsilon > 0 \quad \exists t_\epsilon > 0 \text{ such that } \mu\{t \in [t_\epsilon, +\infty) : \|\gamma(t) - \eta\| > \epsilon\} < \epsilon.$$

Moreover, we will say that $\gamma^\star \in \mathbb{R}^n$ is an almost cluster point of $\gamma(t)$ as $t \to +\infty$ if

$$\forall \epsilon > 0 \quad \mu\{t \in [0, +\infty) : \|\gamma(t) - \gamma^\star\| \le \epsilon\} = +\infty.$$

The following result establishes a link between the concept of limit in measure and almost cluster point, which will be employed in the paper.

Proposition 2: Suppose that $\gamma(t)\colon [0, +\infty) \to C$ is a measurable function, and $C \subset \mathbb{R}^n$ is a compact set. If $\gamma(t)$ has a unique almost cluster point $\gamma^\star \in C$ as $t \to +\infty$, then $\mu\lim_{t\to+\infty} \gamma(t) = \gamma^\star$.

Proof: See Appendix I. ■

On the basis of the previous considerations, we give the next definition.

Definition 7: Let η be an OEP of (N1) corresponding to the EP e of (N1). The OEP η is said to be μ-globally attractive (μ-GA) if, for each $x_0 \in \mathbb{R}^n$, any trajectory $x(t)$ of (N1), $t \in [0, +\infty)$, with $x(0) = x_0$, is such that $\gamma(t)$ in (2) verifies $\mu\lim_{t\to+\infty} \gamma(t) = \eta$.

The above definition of μ-GA can be applied to general cases where the neural-network output $\gamma(t)$ is only measurable. In this paper we will encounter situations where we can employ the standard concept of global attractivity also for the outputs $\gamma(t)$. A typical case is that where $\gamma(t)$ turns out to be defined and continuous for all $t \ge \tilde{t} > 0$, for some $\tilde{t} > 0$ (cf. (b) of Property 1). To deal with these situations, we introduce the next definition.

Definition 8: Let η be an OEP of (N1) corresponding to the EP e of (N1). The OEP η is said to be Globally Attractive (GA) if, for each $x_0 \in \mathbb{R}^n$, any trajectory $x(t)$ of (N1), $t \in [0, +\infty)$, with $x(0) = x_0$, is such that $\gamma(t)$ in (2) verifies the next properties.

a) $\gamma(t)$ is defined and continuous for all $t \ge \tilde{t} > 0$, for some $\tilde{t} > 0$.

b) it results $\lim_{t\to+\infty} \gamma(t) = \eta$.

Finally, with the next definition, we single out a situation where there is global convergence in finite time.

Definition 9: The neural network (N1) is said to be globally convergent in finite time if the following conditions hold.

1) (N1) has a unique EP e and a unique corresponding OEP η.
2) for each $x_0 \in \mathbb{R}^n$, any trajectory $x(t)$ of (N1), $t \in [0, +\infty)$, with $x(0) = x_0$, is such that it results $x(t) = e$, for $t \ge \tilde{t}$, for some $\tilde{t} > 0$. Moreover, $\gamma(t)$ in (2) verifies $\gamma(t) = \eta$ for $t \ge \tilde{t}$.

Remarks:

1. From Definition 9, it is seen that if (N1) is globally convergent in finite time, then it has a GA EP e and a corresponding GA OEP η. However, the notion of convergence in finite time is much stronger than global attractivity, since it implies that the state $x(t)$ and outputs $\gamma(t)$ become exactly equal to their limits after some finite time has elapsed. Convergence in finite time represents a peculiar feature of systems described by vector fields which are discontinuous in the state [6], [8], [18]. We recall that on the contrary for systems described by smooth vector fields there can be only asymptotic convergence of trajectories toward the limit.

2. It is of interest to discuss in some more details the engineering significance of the μ-limit as given in Definition 6. To this end, let us point out the following result.

Property 3: Suppose that $\gamma(t)\colon [0, +\infty) \to \mathbb{R}^n$ is a measurable function such that $\mu\lim_{t\to+\infty} \gamma(t) = \eta \in \mathbb{R}^n$. Let us consider the system of first-order differential equations

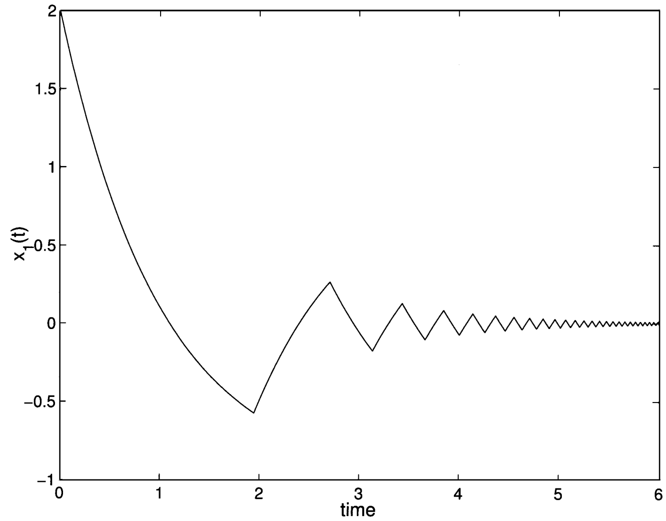
$$\dot{y} = -\frac{1}{\tau}y - \frac{1}{\tau}\gamma(t) \tag{5}$$

where $y \in \mathbb{R}^n$ and τ is a positive constant. Then, any solution $y(t)$ of (5) satisfies

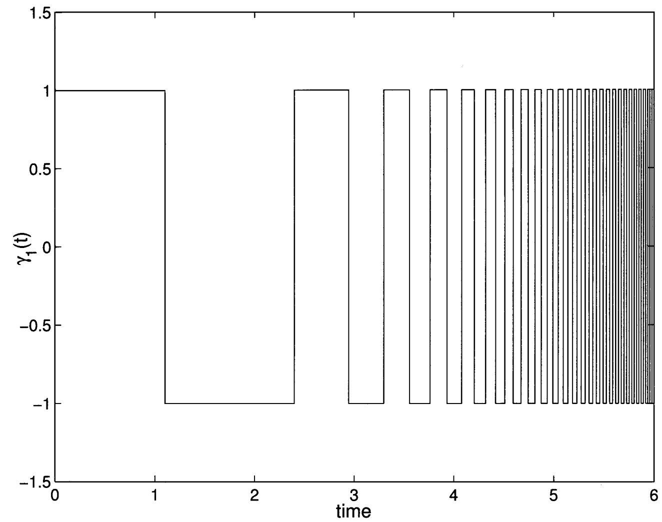
$$\lim_{t\to+\infty} y(t) = \eta.$$

Proof: See Appendix II. ■

Property 3 has a simple interpretation, namely, if the neuron output $\gamma_i(t) \in \mathbb{R}$ is passed through a first-order low-pass filter with transfer function $(1/\tau)(1/(s+1/\tau))$, then the filter output $y_i(t)$ admits η_i as an ordinary limit. In summary, though weaker than the ordinary notion of limit, the μ-limit is actually almost as useful for the engineering applications of neural networks. ■



(a)



(b)

Fig. 2. (a) Time-domain evolution of the state variable $x_1(t)$. (b) Evolution of the corresponding output $\gamma_1(t)$, for the neural network of Example 2.

3. In the standard case considered in the literature where the neuron activations g are continuous and monotone functions, it is easy to see that global attractivity of an EP e also implies global attractivity of the output $g(e)$. Unfortunately, this property is no longer valid for the class of discontinuous activations $g \in \mathcal{G}_{\mathcal{D}}$, as shown in the next Example 2. Therefore, for discontinuous activations it is needed to address separately both global convergence of the state variables and the output variables.

Example 2: Let us consider the second-order neural network

$$\begin{cases} \dot{x}_1 = -x_1 - \operatorname{sgn}(x_2) \\ \dot{x}_2 = -x_2 + \operatorname{sgn}(x_1). \end{cases}$$

The neural network possesses a unique EP $e = 0$ and a unique corresponding OEP $\eta = 0$. By explicitly solving the equations, it is found that for each initial condition $x(0) \neq 0$, the trajectories $x(t)$ are such that $x(t) \to e = 0$ as $t \to +\infty$, i.e., e is GA. This is confirmed by numerical simulations of the neural network obtained with MATLAB, see Fig. 2(a). However, $x_1(t)$ and $x_2(t)$ tend to 0 while indefinitely oscillating between negative and positive values. Therefore, the corresponding neuron outputs $\gamma_i(t)$ indefinitely oscillate between the neuron saturation levels $+1$ and -1. It is also straightforward to see that the outputs have $+1$ and -1 as almost cluster points as $t \to +\infty$, and $\gamma(t)$ does not admit a μ-limit as $t \to +\infty$, i.e., the OEP $\eta = 0$ is not GA, nor μ-GA [see Fig. 2(b)]. Note in particular that $\eta = 0$ is not μ-GA even though $e = 0$ is GA and $\det T \neq 0$. ■

III. Existence and Uniqueness of Equilibrium Point

In this section, we address the existence and uniqueness of the EP and the corresponding OEP of (N1). These properties are a prerequisite for the global attractivity of the EP and OEP of (N1), which is studied in Section IV and Section V. Moreover, we study how the EP and OEP of (N1) depend upon the neural-network input.

We begin by addressing the existence issue.

Property 4: Suppose that $g \in \mathcal{G}_{\mathcal{D}}$. Then, there exists at least an EP e of (N1), and a corresponding OEP η of (N1).

Proof: It suffices to observe that the set-valued map $\tilde{\phi}(x) = -B^{-1}\{TK[g(x)] + I\}$: $\mathbb{R}^n \multimap \mathbb{R}^n$ is upper semicontinuous with nonempty compact convex values. Moreover, since g is bounded and hence $K[g(x)]$ is bounded, it follows that $\tilde{\phi}$ maps a ball centered at the origin, and of sufficiently large radius, into itself. Therefore, from Kakutani's fixed point theorem (see, e.g., [7, p. 85, Corollary 1]), it follows that there exists at least one fixed point of $\tilde{\phi}$, i.e., a point $e \in \mathbb{R}^n$ such that $e \in \tilde{\phi}(e)$, which also represents an EP of (N1). Finally, from (4), we have the existence of an OEP η of (N1) corresponding to e. ■

The main result in this section is given in the next theorem.

Theorem 1: Suppose that $-T \in P$. Then, for any neuron activations $g \in \mathcal{G}_{\mathcal{D}}$, any diagonal matrix B, and any input vector $I \in \mathbb{R}^n$, (N1) has a unique EP e and a unique corresponding OEP η.

Proof: If the OEP η of (N1) is unique, then this is true also for the EP e of (N1). In fact, we have $e = -B^{-1}(T\eta + I)$. Then, it suffices to study the uniqueness of the OEP.

Suppose, for contradiction, that there exist two distinct OEPs η^1 and η^2 of (N1). By definition, there are two EPs e^1 and e^2 of (N1) such that $\eta^1 \in K[g(e^1)]$, $\eta^2 \in K[g(e^2)]$, and $0 = Be^1 + T\eta^1 + I$, $0 = Be^2 + T\eta^2 + I$. Hence

$$-B(e^1 - e^2) - T(\eta^1 - \eta^2) = 0. \tag{6}$$

Now suppose, without loss of generality, that it results $\eta_i^1 \neq \eta_i^2$ for $i = 1, \cdots, m$, and $\eta_i^1 = \eta_i^2$ for $i = m+1, \cdots, n$, where $1 \le m \le n$. Let us introduce the vectors $\eta^{ja} = (\eta_1^j, \cdots, \eta_m^j)' \in \mathbb{R}^m$ and $\eta^{jb} = (\eta_{m+1}^j, \cdots, \eta_n^j)' \in \mathbb{R}^{n-m}$, $j = 1, 2$. In a similar way we can define $e^{ja} \in \mathbb{R}^m$ and $e^{jb} \in \mathbb{R}^{n-m}$. Let us also introduce the matrices $B^{aa} = \operatorname{diag}(-b_1, \cdots, -b_m)$ and $B^{bb} = \operatorname{diag}(-b_{m+1}, \cdots, -b_n)$, and subdivide in an analogous way T into the sub-matrices $T^{aa} \in \mathbb{R}^{m \times m}$, $T^{ab} \in \mathbb{R}^{m \times (n-m)}$, $T^{ba} \in \mathbb{R}^{(n-m) \times m}$, and $T^{bb} \in \mathbb{R}^{(n-m) \times (n-m)}$.

By substituting in (6), and taking into account that $\eta^{1b} - \eta^{2b} = 0$, we get

$$\begin{cases} -B^{aa}(e^{1a} - e^{2a}) - T^{aa}(\eta^{1a} - \eta^{2a}) = 0 \\ -B^{bb}(e^{1b} - e^{2b}) - T^{ba}(\eta^{1a} - \eta^{2a}) = 0. \end{cases} \tag{7}$$

Consider any $i \in \{1, \cdots, m\}$, hence $\eta_i^{1a} \neq \eta_i^{2a}$. Since functions g_i are nondecreasing (see (c) of Definition 3), it immediately follows that $H_i = (e_i^{1a} - e_i^{2a})/(\eta_i^{1a} - \eta_i^{2a}) \geq 0$. Therefore, the first equation of system (7) yields

$$-(B^{aa}H^{aa} + T^{aa})(\eta^{1a} - \eta^{2a}) = 0$$

where $H^{aa} = \operatorname{diag}(H_1, \cdots, H_m)$.

Since $-T \in P$, and $-B^{aa}H^{aa}$ is a diagonal matrix with nonnegative diagonal entries, we also have $-(B^{aa}H^{aa} + T^{aa}) \in P$ [25]. Then, in particular, $\det(B^{aa}H^{aa} + T^{aa}) \neq 0$, so that $\eta^{1a} = \eta^{2a}$, which is a contradiction. ■

Suppose that $-T \in P$ and $g \in \mathcal{G_D}$. Due to Theorem 1, we can define two maps, $S_e\colon \mathbb{R}^n \to \mathbb{R}^n$ and $S_\eta\colon \mathbb{R}^n \to \mathbb{R}^n$, as follows

$$S_e(I) = e \quad S_\eta(I) = \eta.$$

Namely, maps S_e and S_η associate to the neural-network input, I, the unique EP e of (N1), and the unique corresponding OEP η of (N1), respectively.

Moreover, let us consider the map $\Sigma\colon \mathbb{R}^n \to \mathbb{R}^n \times \mathbb{R}^n$ defined as

$$\Sigma(I) = (S_e(I), S_\eta(I)) = (e, \eta). \tag{8}$$

Since $S_\eta(I) = \eta \in K[g(e)]$, we have $\Sigma(I) \in U$, where

$$U = \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^n : x \in \mathbb{R}^n, y \in K[g(x)]\} \tag{9}$$

is the graph of the set-valued map $K[g]$ that models the neuron activations.

The following property holds.

Property 5: Suppose that $-T \in P$ and $g \in \mathcal{G_D}$. Then, Σ is a homeomorphism from $\mathbb{R}^n$ onto U, i.e., Σ is injective, Σ and the inverse Σ^{-1} are continuous, and $\Sigma(\mathbb{R}^n) = U$.

Proof: See Appendix III. ■

Property 5 states that both the EP e and the corresponding OEP η of (N1) depend continuously upon the neural-network input I. Moreover, the pairs constituted by e and η describe the graph U of the set-valued map $K[g]$ when I varies on $\mathbb{R}^n$, i.e., $\Sigma(I)$ is a global parameterization of U.

IV. Global Convergence of Neural Network

In this section, we use an approach based on the concept of Lyapunov functions to address global convergence of the state trajectories and global convergence in measure of the output trajectories of the neural network.

The main result in this section is as follows.

Theorem 2: Suppose that $-T \in$ LDS. Then, for any neuron activations $g \in \mathcal{G_D}$, any diagonal matrix B, and any input vector $I \in \mathbb{R}^n$, (N1) has a unique EP e which is GA, and a unique corresponding OEP η which is μ-GA.

Theorem 2 has wide applicability, indeed it includes classes of interconnection matrices T of theoretical as well as practical interest in the neural-network field [12], such as symmetric and negative definite matrices and skew-symmetric matrices with negative diagonal entries. Moreover, cooperative neural networks where $-T$ is an M-matrix, and cooperative-competitive neural networks where $-T$ is an H-matrix with positive diagonal elements, are such that $-T \in$ LDS.[2]

It is also worth to remark that the LDS condition in Theorem 2 is robust, namely, if $-T \in$ LDS, then the perturbed interconnection matrix $-(T + \Delta T)$ is still LDS, for sufficiently small $\|\Delta T\|$. The allowed $\|\Delta T\|$ which preserves the LDS condition can be quantitatively estimated by using techniques such as those described in [12]. This robustness property of global convergence is highly desirable in view of the applications. In fact, due to tolerances, in the actual electronic implementation it is not possible to exactly realize the nominal matrix T.[3]

Proof of Theorem 2: The hypothesis $-T \in$ LDS implies $-T \in P$. Let e be the unique EP of (N1), and η be the unique corresponding OEP of (N1) (see Theorem 1), i.e., $\eta \in K[g(e)]$ and $0 = Be + T\eta + I$. The change of variable $z = x - e$ translates e into the origin. The neural network is described in terms of variables z by $\dot{z} \in Bz + TK[g(z+e)] + Be + I = Bz + TK[g(z+e)] - T\eta$, hence

$$\dot{z} \in Bz + TK\left[G(z)\right] \tag{N2}$$

where

$$G(z) = g(z + e) - \eta. \tag{10}$$

Proposition 3: Suppose that $g \in \mathcal{G_D}$. Then, G as defined in (10) satisfies the following properties.

i) $G \in \mathcal{G_D}$.

ii) $0 \in K[G(0)]$.

iii) Let $\Gamma = \operatorname{diag}(\Gamma_1, \cdots, \Gamma_n)$ be a diagonal matrix with $\Gamma_i > 0$, $i = 1, \cdots, n$. Then

$$z'\Gamma\tilde{\gamma} \geq 0 \qquad \forall \tilde{\gamma} \in K\left[G(z)\right]. \tag{11}$$

Proof of Proposition 3: (i) Obvious. (ii) We have $K[G(0)] = K[g(e)] - \eta$. Since $\eta \in K[g(e)]$, we conclude that $0 \in K[G(0)]$. (iii) It is an immediate consequence of the fact that the graph of $K[G_i]$, $i = 1, \cdots, n$, belongs to the first and third quadrant of the space $\mathbb{R}^2$. ■

From (ii) of Proposition 3 it immediately follows that (N2) has a unique EP $\tilde{e} = 0$ and a unique corresponding OEP $\tilde{\eta} = 0$. Moreover, if $z(t)$ is a solution of (N2) on $[t_0, t_1]$, $t_0 < t_1 \leq +\infty$, then there exists a bounded measurable function $\tilde{\gamma}(t)$ such that for a.a. $t \in [t_0, t_1]$ it results in

$$\tilde{\gamma}(t) \in K\left[G\left(z(t)\right)\right] \quad \dot{z}(t) = Bz(t) + T\tilde{\gamma}(t). \tag{12}$$

To proceed in the proof of Theorem 2, we use the concept of monotone trajectories of (N2) (see [7, p. 293, Definition 1]). Since $-T \in$ LDS, there exists some diagonal matrix $\alpha = \operatorname{diag}(\alpha_1, \cdots, \alpha_n)$, with $\alpha_i > 0$, $i = 1, \cdots, n$, such that $[\alpha(-T)]^S$ is positive definite. Consider the (candidate)

[2] We refer the reader to [26] for a review on the most commonly employed classes of LDS matrices, [12] for numerical techniques to verify the LDS condition, and [13] for a review on results on global convergence of other classes of neural networks involving the concept of LDS matrices.

[3] To understand the importance of such a robustness issue, note for comparison that there are other classes of neural networks for which the convergence properties are instead not preserved when T is perturbed [28], [29].

Lyapunov function, $V\colon \mathbb{R}^n \to \mathbb{R}$, of the Lur'e Postnikov type (cf. [12])

$$V(z) = -z'B^{-1}z + 2c\sum_{i=1}^{n}\alpha_i\int_0^{z_i} G_i(\rho)d\rho \qquad (13)$$

where

$$c > \frac{1}{2\lambda_m}\|B^{-1}T\|_2^2 > 0 \qquad (14)$$

and

$$\lambda_m = \Lambda_m\left\{[\alpha(-T)]^S\right\} > 0. \qquad (15)$$

Note that under condition (14), being $\|B^{-1}T\|_2^2 = \Lambda_M\{(B^{-1}T)'B^{-1}T\}$, it follows that

$$\lambda = \Lambda_m\left\{2c[\alpha(-T)]^S - (B^{-1}T)'(B^{-1}T)\right\} > 0. \qquad (16)$$

Since $G \in \mathcal{G}_\mathcal{D}$ (see (i) of Proposition 3), V turns out to be positive definite, i.e., $V(0) = 0$ and $V(z) > 0$ if $z \neq 0$. Consider also the positive definite function $W\colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$

$$W(z, v) = \|z\|_2^2 + \|B^{-1}v\|_2^2. \qquad (17)$$

The following holds.

Proposition 4: Suppose that $-T \in$ LDS and $g \in \mathcal{G}_\mathcal{D}$. Then, each trajectory $z(t)$ of (N2) is monotone for $t \in [0, +\infty)$ with respect to function V defined in (13) and function W defined in (17), i.e.,

$$\forall s \geq t \geq 0 \quad V(z(s)) - V(z(t)) + \int_t^s W(z(\sigma), \dot{z}(\sigma))\, d\sigma \leq 0.$$

Proof of Proposition 4: The proof is an immediate consequence of the next basic lemma.

Lemma 1: Suppose that $-T \in$ LDS and $g \in \mathcal{G}_\mathcal{D}$. Then, for a.a. $t \in [0, +\infty)$, V satisfies the Lyapunov property

$$\dot{V}(z(t)) \leq -\|z(t)\|_2^2 - \left\|B^{-1}\dot{z}(t)\right\|_2^2 - \lambda\|\tilde{\gamma}(t)\|_2^2 \leq 0 \quad (18)$$

where $\lambda > 0$ is defined in (16) and $\tilde{\gamma}(t)$ is defined in (12).

Proof of Lemma 1: See Appendix IV. ■

To prove Proposition 4 it suffices to note that, on the basis of Lemma 1, we have

$$\dot{V}(z(t)) \leq -\|z(t)\|_2^2 - \left\|B^{-1}\dot{z}(t)\right\|_2^2 = -W(z(t), \dot{z}(t))$$

for a.a. $t \in [0, +\infty)$, and integrate this inequality on the interval $[t, s]$. ■

By exploiting the result in Proposition 4 we are in a position to complete the proof of Theorem 2. It is seen that function V in (13) is positive definite and radially unbounded ($V(z) \to +\infty$ as $\|z\| \to +\infty$). Moreover, from Lemma 1, $\dot{V}(z(t))$ is negative definite, i.e., $\dot{V}(0) = 0$ and $\dot{V}(z(t)) < 0$ for $z(t) \neq 0$. Therefore, the EP $\tilde{e} = 0$ of (N2) is GA, hence the EP e of (N1) is GA.

Then, we address convergence of $\tilde{\gamma}(t)$. From Proposition 4 and [7, p. 311, Th. 3] it turns out that the almost cluster points $v^\star$ of $\dot{z}(t)$, as $t \to +\infty$, satisfy the equation $W(0, v^\star) = \|B^{-1}v^\star\| = 0$. Therefore, $v^\star = 0$ is the unique almost cluster point of $\dot{z}(t)$, as $t \to +\infty$, corresponding to the EP $\tilde{e} = 0$ of (N2).

Now, let us prove that $\tilde{\eta} = 0$ is the unique almost cluster point of the outputs $\tilde{\gamma}(t)$. To this end, consider for $t \geq 0$ any solution $z(t)$ of (N2). From Proposition 4, $z(t)$ is monotone with respect to V and W. Moreover, $\dot{z}(t) = Bz(t) + T\tilde{\gamma}(t)$ for a.a. $t \in [0, +\infty)$, where $\tilde{\gamma}(t) \in K[G(z(t))]$, or equivalently

$$T^{-1}\dot{z}(t) - T^{-1}Bz(t) = \tilde{\gamma}(t)$$

for a.a. $t \in [0, +\infty)$. It is known that $z(t) \to \tilde{e} = 0$ for $t \to +\infty$, and for any $\epsilon > 0$, $\mu\{t \in [0, +\infty) : \|\dot{z}(t)\| \leq \epsilon\} = +\infty$. Then, it is easily seen that $T^{-1}v^\star = 0 = \tilde{\eta}$ is the unique almost cluster point of $\tilde{\gamma}(t)$ as $t \to +\infty$. From Proposition 2, it follows that $\tilde{\eta} = 0$ is μ-GA. Therefore, also the OEP η of (N1) is μ-GA. ■

V. Stronger Global Convergence Properties of Neural Network

Under the hypotheses $-T \in$ LDS and $g \in \mathcal{G}_\mathcal{D}$, Theorem 2 ensures GA of the EP e of (N1), and μ-GA of the corresponding OEP η of (N1), for *all* neural-network inputs $I \in \mathbb{R}^n$. The goal of this section is to show that for *almost all* inputs I it is possible to establish the standard property of GA, as in Definition 8, also for the OEP η (Section V-A). Moreover, we can find well characterized sets of inputs I, whose measure is nonzero, for which there is global convergence of (N1) in *finite time* to e and η, as in Definition 9 (Section V-B).

A. Global Attractivity of the Output Equilibrium Point

Suppose that $-T \in$ LDS and $g \in \mathcal{G}_\mathcal{D}$. For a given input vector I to the neural network, we denote henceforth by $e = (e_1, \cdots, e_n)'$ the unique EP of (N1), and by $\eta = (\eta_1, \cdots, \eta_n)'$ the unique corresponding OEP of (N1) (cf. Theorem 1). Let

$$\Theta^D(e) = \{i \in \{1, \cdots, n\} : g_i \text{ is discontinuous at } x_i = e_i\} \qquad (19)$$

$$\Theta^C(e) = \{i \in \{1, \cdots, n\} : g_i \text{ is continuous at } x_i = e_i\}. \qquad (20)$$

Note that if $i \in \Theta^D(e)$, then $K[g(e_i)] = [g(e_i^-), g(e_i^+)]$ is an interval with a nonempty interior $\mathrm{int}(K[g(e_i)])$. Instead, if $i \in \Theta^C(e)$, $K[g(e_i)] = \{g(e_i)\}$ is a singleton.

Next, we define two sets of inputs I which play a crucial role to establish GA of η.

Definition 10: Suppose that $-T \in$ LDS and $g \in \mathcal{G}_\mathcal{D}$. Let

$$\mathcal{I}_{\mathrm{GA}} = \left\{I \in \mathbb{R}^n : \forall i = 1, \cdots, n, i \in \Theta^C \text{ or } i \in \Theta^D(e) \text{ and } \eta_i \in \mathrm{int}\left(K\left[g_i(e_i)\right]\right) = \left(g_i\left(e_i^-\right), g_i\left(e_i^+\right)\right)\right\}$$

$$\mathcal{I}'_{\mathrm{GA}} = \mathcal{C}_{\mathbb{R}^n}\mathcal{I}_{\mathrm{GA}} = \left\{I \in \mathbb{R}^n : \Theta^D(e) \neq \emptyset \text{ and } \exists i \in \Theta^D(e) \text{ such that } \eta_i \in \partial K\left[g_i(e_i)\right] = \left\{g_i(e_i^-), g_i(e_i^+)\right\}\right\}.$$

When $I \in \mathcal{I}_{\mathrm{GA}}$, one of the situations illustrated in Fig. 3(a), (b) is verified by e_i and η_i, for all $i = 1, \cdots, n$. On the contrary, when $I \in \mathcal{I}'_{\mathrm{GA}}$, for some $i \in \Theta^D(e)$ we have the situation

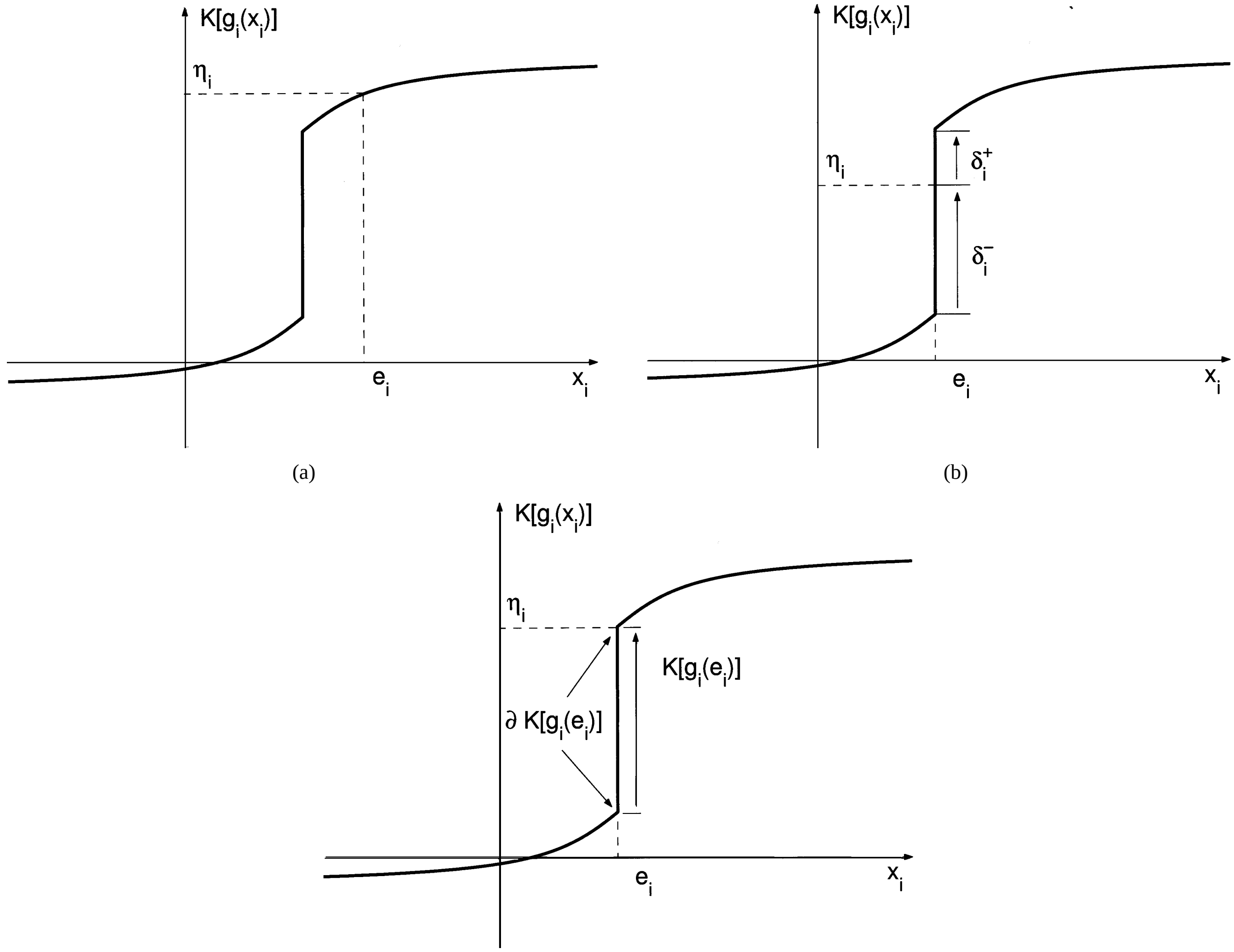


Fig. 3. Geometrical interpretation of the sets of inputs $\mathcal{I}_{\rm GA}$ and $\mathcal{I}'_{\rm GA}$. (a) $I \in \mathcal{I}_{\rm GA}$ and $i \in \Theta^C(e)$. (b) $I \in \mathcal{I}_{\rm GA}$ and $i \in \Theta^D(e)$. (c) $I \in \mathcal{I}'_{\rm GA}$ and $i \in \Theta^D(e)$.

represented in Fig. 3(c). Note in particular from Fig. 3(b) that if $I \in \mathcal{I}_{\rm GA}$ and $i \in \Theta^D(e)$, then, it results in

$$g_i\left(e_i^+\right) - \eta_i = G_i(0^+) = \delta_i^+ > 0 \qquad (21)$$

$$\eta_i - g_i\left(e_i^-\right) = -\,G_i(0^-) = \delta_i^- > 0 \qquad (22)$$

where functions G_i are the components of G defined in (10).

It is intuitively clear that the situation of Fig. 3(c) is not generic. This idea is made precise by the next property.

Property 6: Suppose that $-T \in$ LDS and $g \in \mathcal{G}_{\mathcal{D}}$. Then, set $\mathcal{I}_{\rm GA}$ is open. Moreover, $\partial\mathcal{I}_{\rm GA} = \mathcal{I}'_{\rm GA}$ and $\mu(\mathcal{I}'_{\rm GA}) = 0$.

Proof: See Appendix V. ■

The main result on global attractivity of the OEP of the neural network is as follows.

Theorem 3: Suppose that $-T \in$ LDS and $g \in \mathcal{G}_{\mathcal{D}}$. Then, for any input vector $I \in \mathcal{I}_{\rm GA}$ (i.e., for almost all vectors $I \in \mathbb{R}^n$), (N1) has a unique EP e which is GA, and a unique corresponding OEP η which is GA.

Proof: It suffices to prove that the OEP $\tilde{\eta} = 0$ is GA for (N2). We note these facts. Suppose that $-T \in$ LDS and $g \in \mathcal{G}_{\mathcal{D}}$, and consider the neural networks (N1) and (N2). From Theorem 2, it is known that the EP $\tilde{e} = 0$ of (N2) is GA, hence any trajectory $z(t)$ of (N2) satisfies $z(t) \to 0$ as $t \to +\infty$. Suppose that $i \in \Theta^C(e)$, i.e., g_i is continuous at e_i and G_i is continuous at 0. Then, being g_i and G_i piecewise continuous, $z_i(t) \to 0$ as $t \to +\infty$ also implies that $\tilde{\gamma}_i(t)$ in (12) is defined and continuous for $t > \tilde{t}_i$, for some $\tilde{t}_i > 0$ (see Property 1), and $\tilde{\gamma}_i(t) \to 0$ as $t \to +\infty$. To prove GA of the OEP $\tilde{\eta} = 0$ of (N2) we thus need to show that $\tilde{\gamma}_i(t)$ is defined and continuous for $t > \tilde{t}_i$, for some $\tilde{t}_i > 0$, and $\tilde{\gamma}_i(t) \to 0$ as $t \to +\infty$, also for any $i \in \Theta^d(e)$. To this end, we consider three different cases.

i) $\Theta^D(e) = \emptyset$. From Theorem 2, it is known that $z(t) \to 0$ as $t \to +\infty$. Since G_i, $i = 1, \cdots, n$, are continuous at 0, we immediately conclude that $\tilde{\gamma}(t)$ is defined and continuous for $t > \tilde{t}$, for some $\tilde{t} > 0$, and $\tilde{\gamma}(t) \to 0$ as $t \to +\infty$.

ii) $\Theta^D(e) \neq \emptyset$; $\Theta^D(e) \neq \{1, \cdots, n\}$. Suppose, without loss of generality, that $\Theta^D(e) = \{1, \cdots, p\}$ and $\Theta^C(e) = \{p+1, \cdots, n\}$, where $1 \le p < n$. Let $z^D = (z_1, \cdots, z_p)'$, $z^C = (z_{p+1}, \cdots, z_n)'$, and subdivide vectors and matrices that define (N2), accordingly.

From (N2), we get

$$\dot{z}^D(t) = B^{DD} z^D(t) + T^{DD}\tilde{\gamma}^D(t) + T^{DC}\tilde{\gamma}^C(t) \qquad \text{(ND)}$$

where $\tilde{\gamma}^D(t) = (\tilde{\gamma}_1(t), \cdots, \tilde{\gamma}_p(t))' \in K[G^D(z^D(t))]$. Moreover, $\tilde{\gamma}^C(t) = (\tilde{\gamma}_{p+1}(t), \cdots, \tilde{\gamma}_n(t))'$ is defined and continuous for $t > \tilde{t}_C$, for some $\tilde{t}_C > 0$, and $\tilde{\gamma}^C(t) \to 0$ as $t \to +\infty$.

Let us consider for (ND) the (candidate) Lyapunov function

$$V_D(z^D) = -(z^D)'(B^{DD})^{-1}z^D + 2c_D \sum_{i \in \Theta^D(e)} \alpha_i \int_0^{z_i} G_i(\rho)d\rho$$

where

$$c_D > \frac{1}{2\lambda_{mD}} \left\|(B^{DD})^{-1}T^{DD}\right\|_2^2 > 0$$

and

$$\lambda_{mD} = \Lambda_m \left\{ \left[\alpha^{DD}(-T^{DD})\right]^S \right\}$$

with $\alpha^{DD} = \mathrm{diag}(\alpha_1, \cdots, \alpha_p)$. Under the assumption $-T \in$ LDS, it can be easily verified that $[\alpha^{DD}(-T^{DD})]^S$ is positive definite, hence $\lambda_{mD} > 0$.

The following lemma holds.

Lemma 2: Suppose that $-T \in$ LDS and $g \in \mathcal{G}_\mathcal{D}$. Furthermore, suppose that $I \in \mathcal{I}_{\mathrm{GA}}$ and that $\Theta^D(e) \neq \emptyset$, $\Theta^D(e) \neq \{1, \cdots, n\}$. Then, there exists $\bar{t} > 0$ and a constant $\Delta_D > 0$ such that for a.a. $t \geq \bar{t}$ it results $\dot{V}_D(z^D(t)) < -\Delta_D < 0$ for $z^D(t) \neq 0$.

Proof: See Appendix VI. ■

By integrating between $\bar{t}$ and $t > \bar{t}$ the inequality for $\dot{V}_D(z^D(t))$ of Lemma 2, we conclude that for $t > \tilde{t}_D$, where $\tilde{t}_D = V_D(z^D(\bar{t}))/\Delta_D > 0$, we have $V_D(z^D(t)) = 0$. Hence, $z^D(t) = 0$ for $t > \tilde{t}_D$.

Then, for $t > \tilde{t}_D$ system (ND) reduces to $0 = T^{DD}\tilde{\gamma}^D(t) + T^{DC}\tilde{\gamma}^C(t)$, where $T^{DC}\tilde{\gamma}^C(t)$ is defined and continuous for $t > \tilde{t}_D$, and $T^{DC}\tilde{\gamma}^C(t) \to 0$ as $t \to +\infty$. Moreover, since $-T \in$ LDS, it follows that T^{DD} is nonsingular, hence $\tilde{\gamma}^D(t) = -(T^{DD})^{-1}T^{DC}\tilde{\gamma}^C(t)$ is defined and continuous for $t > \tilde{t}_D$, and $\tilde{\gamma}^D(t) \to 0$ as $t \to +\infty$.

iii) $\Theta^D(e) = \{1, \cdots, n\}$. From Lemma 1 we obtain $\dot{V}(z(t)) \leq -\lambda\|\tilde{\gamma}(t)\|_2^2$. Taking into account (21) and (22), for $z^D(t) \neq 0$ we have $\|\tilde{\gamma}(t)\|_2 \geq \Delta$, where

$$\delta = \min_{i \in \{1, \cdots, n\}} \left\{ \min\left\{\delta_i^+, \delta_i^-\right\}\right\} > 0.$$

Therefore, it results $V(z(t)) = 0$ and $z(t) = 0$ for $t \geq \tilde{t}$, where

$$\tilde{t} = \frac{V\left(z(0)\right)}{\lambda\delta^2}. \tag{23}$$

From (N2) we get $0 = T\tilde{\gamma}(t)$. Hence, $\tilde{\gamma}(t) = 0$ for $t \geq \tilde{t}$, i.e., $z(t)$ and $\tilde{\gamma}(t)$ become 0 in finite time. ■

B. Global Convergence in Finite Time

In this section, we establish a condition ensuring that the neural network (N1) is globally convergent in finite time.

Let

$$\Delta_g = \{x = (x_1, \cdots, x_n)' \in \mathbb{R}^n : \forall i = 1, \cdots, n, g_i \text{ is discontinuous at } x_i\}. \tag{24}$$

Note that $\Delta_g \neq \emptyset$ if and only if, for all $i = 1, \cdots, n$, g_i has at least a point of discontinuity. The next set of inputs I plays a crucial role in the study of global convergence of (N1) in finite time.

Definition 11: Suppose that $-T \in$ LDS and $g \in \mathcal{G}_\mathcal{D}$. Moreover, assume that $\Delta_g \neq \emptyset$. Let

$$\mathcal{I}_{\mathrm{FT}} = \{I \in \mathbb{R}^n : I \in \mathcal{I}_{\mathrm{GA}} \text{ and } e \in \Delta_g\}$$

where e is the unique EP of (N1) and $\mathcal{I}_{\mathrm{GA}}$ is given in Definition 10.

Set $\mathcal{I}_{\mathrm{FT}}$ corresponds to vectors $I \in \mathcal{I}_{\mathrm{GA}}$ such that, for all $i = 1, \cdots, n$, the component g_i of g is discontinuous at $x_i = e_i$. That is, the situation illustrated in Fig. 3(b) is verified for $i = 1, \cdots, n$. This set can be explicitly characterized, as shown in the next result.

Property 7: Suppose that $-T \in$ LDS and $g \in \mathcal{G}_\mathcal{D}$. Moreover, assume that $\Delta_g \neq \emptyset$. Then, $\mathcal{I}_{\mathrm{FT}}$ is a nonempty open set which can be represented as follows:

$$\mathcal{I}_{\mathrm{FT}} = \bigcup_{x^d \in \Delta_g} \tilde{\mathcal{I}}_{\mathrm{FT}}(x^d)$$

where, for $x^d \in \Delta_g$

$$\tilde{\mathcal{I}}_{\mathrm{FT}}(x^d) = \left\{ I \in \mathbb{R}^n : I = -Bx^d - T\eta \text{ with } \eta \in \mathrm{int}\left(K\left[g(x^d)\right]\right) \right\}$$

is the interior of the parallelotope in $\mathbb{R}^n$ defined by the 2^n vertexes

$$v^i = -Bx^d - Ty^i \quad y^i = \left(g_1\left(x_1^{d\pm}\right), \cdots, g_n\left(x_n^{d\pm}\right)\right)'.$$

Proof: The property immediately follows from the fact that matrix T defines a nonsingular linear operator. ■

In the particular interesting case where the neurons activations are modeled by $g_i(x_i) = \mathrm{sgn}(x_i)$, $i = 1, \cdots, n$, as in the Hopfield model, we have $\Delta_g = \{0\}$ and $\mathcal{I}_{\mathrm{FT}} = \tilde{\mathcal{I}}_{\mathrm{FT}}(0)$, where $\tilde{\mathcal{I}}_{\mathrm{FT}}(0)$ is the interior of the parallelotope in $\mathbb{R}^n$ defined by the 2^n vertexes

$$v^{Hi} = -Ty^{Hi} \quad y^{Hi} = (\pm 1, \cdots, \pm 1)'.$$

Example 3: Let us consider the second-order neural network

$$\begin{cases} \dot{x}_1 = -x_1 - \frac{1}{4}\mathrm{sgn}(x_1) - \mathrm{sgn}(x_2) + I_1 \\ \dot{x}_2 = -x_2 + \mathrm{sgn}(x_1) - \frac{1}{4}\mathrm{sgn}(x_2) + I_2 \end{cases} \tag{25}$$

which is obtained from the network of Example 2 by adding neuron self-connections equal to $-1/4$ and a generic input I. By choosing $\alpha = \mathrm{diag}(1, 1)$, we obtain $[\alpha(-T)]^S = \mathrm{diag}(1/4, 1/4)$, i.e., $-T \in$ LDS. It turns out that $\mathcal{I}_{\mathrm{FT}} = \tilde{\mathcal{I}}_{\mathrm{FT}}(0)$ is the interior of the parallelotope in $\mathbb{R}^2$ with vertexes $(-5/4, 3/4)'$, $(5/4, -3/4)'$, $(3/4, 5/4)'$, and $(-3/4, -5/4)'$. ■

The following theorem holds.

Theorem 4: Suppose that $-T \in$ LDS and $g \in \mathcal{G}_\mathcal{D}$. Furthermore, suppose that $\Delta_g \neq \emptyset$ (hence $\mathcal{I}_{\mathrm{FT}} \neq \emptyset$), and $I \in \mathcal{I}_{\mathrm{FT}}$. Then, (N1) is globally convergent in finite time. Moreover, the convergence time of a trajectory $x(t)$ of (N1), which is defined as

$$t_c = \inf\left\{\tilde{t} > 0 : x(t) = e,\ \gamma(t) = \eta,\ f\text{for } t > \tilde{t}\right\}$$

satisfies

$$t_c \leq t_e = \frac{1}{\lambda_m \delta^2} \sum_{i=1}^{n} \alpha_i \int_0^{x_i(0)-e_i} G_i(\rho) d\rho \tag{26}$$

where

$$\lambda_m = \Lambda_m \left\{ [\alpha(-T)]^S \right\} > 0 \tag{27}$$

$$\delta = \min_{i \in \{1, \cdots, n\}} \left\{ \min\left\{\delta_i^+, \delta_i^-\right\}\right\} > 0 \tag{28}$$

and for $i = 1, \cdots, n$

$$g_i\left(e_i^+\right) - \eta_i = \delta_i^+ > 0 \tag{29}$$

$$\eta_i - g_i\left(e_i^-\right) = \delta_i^- > 0. \tag{30}$$

To the authors' knowledge, Theorem 4 is the first general result on global convergence in finite time for the additive neural-network model (N1). It is pointed out, from (26), that the estimated convergence time t_e is influenced by the neuron interconnection matrix T, through the eigenvalue λ_m, and by the location of the OEP η within $K[g(e)]$, through parameter δ (cf. (29) and (30)). Moreover, t_e depends on the distance of the initial condition $x(0)$ from the EP e. The distance corresponds to the integral term at the right-hand side of (26), which also represents the energy part of V related to the neuron nonlinearities G_i, $i = 1, \cdots, n$ (cf. (13)).

We observe that since G satisfies (iii) of Proposition 3, and G_i are bounded on $\mathbb{R}$, i.e., $|G_i(\rho)| \le M$, for some $M > 0$, and any $i = 1, \cdots, n$, we obtain from (26) the explicit estimate

$$t_c \le t_e \le \frac{M}{\lambda_m \delta^2} \sum_{i=1}^{n} \alpha_i \left| x_i(0) - e_i \right|.$$

Finally, when the neurons activations are modeled by $g_i(x_i) = \mathrm{sgn}(x_i)$, $i = 1, \cdots, n$, as in the Hopfield model, by taking into account that $\int_0^{x_i} \mathrm{sgn}(\rho) d\rho = |x_i|$, (26) reduces to the simple formula

$$t_c \le t_e = \frac{1}{\lambda_m \delta^2} \sum_{i=1}^{n} \alpha_i \left| x_i(0) - e_i \right|. \qquad (31)$$

Proof of Theorem 4: For $I \in \mathcal{I}_{\mathrm{FT}}$, we have $\Theta^D(e) = \{1, \cdots, n\}$. Hence, global convergence in finite time of (N1) can be proved as in point (iii) of the proof of Theorem 3.

Let us now verify the estimate of the convergence time in (26). From (23), we have

$$t_c \le \frac{V\left(z(0)\right)}{\lambda \delta^2}$$

where V is defined in (13), λ in (16), and δ in (28). Moreover, constant $c > 0$ satisfies (14).

In accordance to (14), we can choose for any $h > 0$

$$c = \frac{1}{2}(1+h)\frac{\|B^{-1}T\|_2^2}{\lambda_m}$$

where $\lambda_m = \Lambda_m\{[\alpha(-T)]^S\}$. Let $\lambda_M = \|B^{-1}T\|_2^2 = \Lambda_M\{(B^{-1}T)'B^{-1}T\}$. Therefore, from (16) we get $\lambda \ge 2c\lambda_m - \lambda_M = (1+h)\lambda_M - \lambda_M = h\lambda_M$. Hence, by taking into account (13), we have

$$\begin{aligned} t_c &\le \frac{V\left(z(0)\right)}{h\delta^2 \lambda_M} \\ &= \frac{1}{h\delta^2\lambda_M} \sum_{i=1}^{n} \frac{1}{b_i} z_i^2(0) + \frac{1+h}{h} \frac{1}{\delta^2 \lambda_m} \sum_{i=1}^{n} \alpha_i \int_0^{z_i(0)} G_i(\rho) d\rho. \end{aligned}$$

Since the last equation is valid for any $h > 0$, by taking the limit as $h \to +\infty$ we obtain

$$t_c \le t_e = \frac{1}{\lambda_m \delta^2} \sum_{i=1}^{n} \alpha_i \int_0^{z_i(0)} G_i(\rho) d\rho.$$

Hence, the result in (26) follows by recalling that $z_i(0) = x_i(0) - e_i$, $i = 1, \cdots, n$. ■

The next examples illustrate the result in Theorem 4.

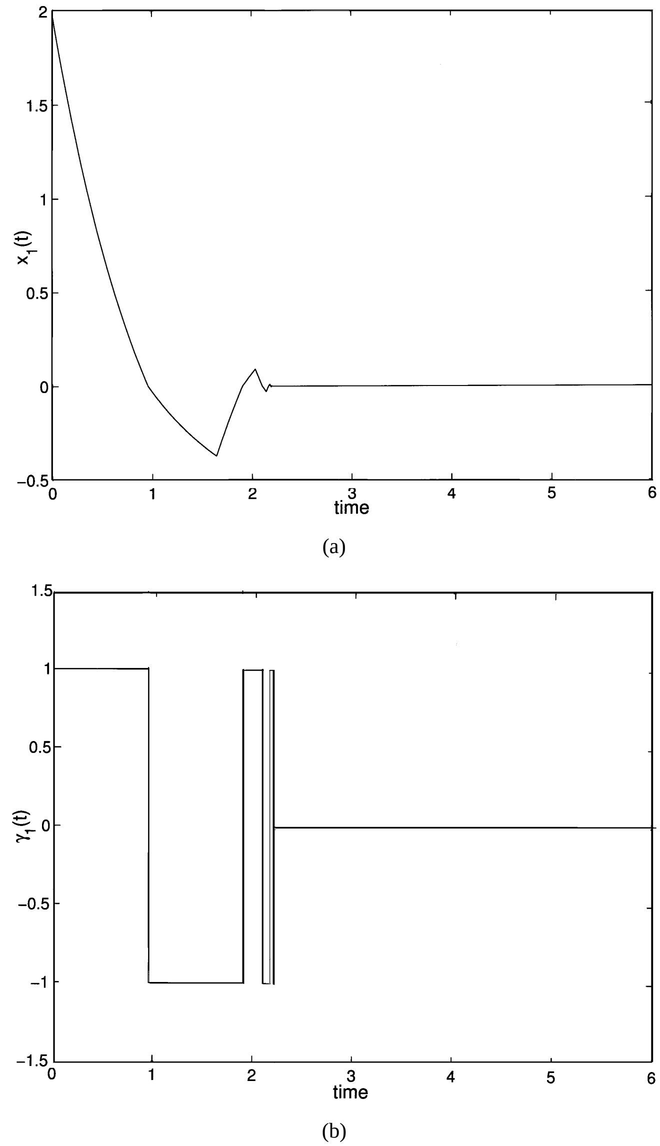


Fig. 4. (a) Time-domain evolution of the state variable $x_1(t)$. (b) Evolution of the corresponding output $\gamma_1(t)$, for the neural network of Example 4.

Example 4: Let us consider the second-order neural network

$$\begin{cases} \dot{x}_1 = -x_1 - \frac{1}{4}\mathrm{sgn}(x_1) - \mathrm{sgn}(x_2) \\ \dot{x}_2 = -x_2 + \mathrm{sgn}(x_1) - \frac{1}{4}\mathrm{sgn}(x_2) \end{cases} \qquad (32)$$

which is obtained from the network of Example 2 by adding neuron self connections equal to $-1/4$.

The neural network has a unique EP $e = 0$ and a unique corresponding OEP $\eta = 0$. As noticed in Example 3, by choosing $\alpha = \mathrm{diag}(1, 1)$, we have $[\alpha(-T)]^S = \mathrm{diag}(1/4, 1/4)$, i.e., $-T \in$ LDS. Moreover, from Example 3, it is seen that $I = 0 \in \mathcal{I}_{\mathrm{FT}}$. Therefore, the conditions of Theorem 4 hold and (32) is globally convergent in finite time.

Fig. 4 depicts the state trajectory $x_1(t)$, and the corresponding output $\gamma_1(t)$, for a solution of (32) starting at $x(0) = (2, 2)'$, as obtained through computer simulations with MATLAB. It is seen that the oscillations rapidly extinguish and $x_1(t)$, $\gamma_1(t)$ converge to their limit in finite time. This is in contrast with Example 2, where $x_1(t)$ and $\gamma_1(t)$ indefinitely oscillate as $t \to +\infty$ (cf. Fig. 2).

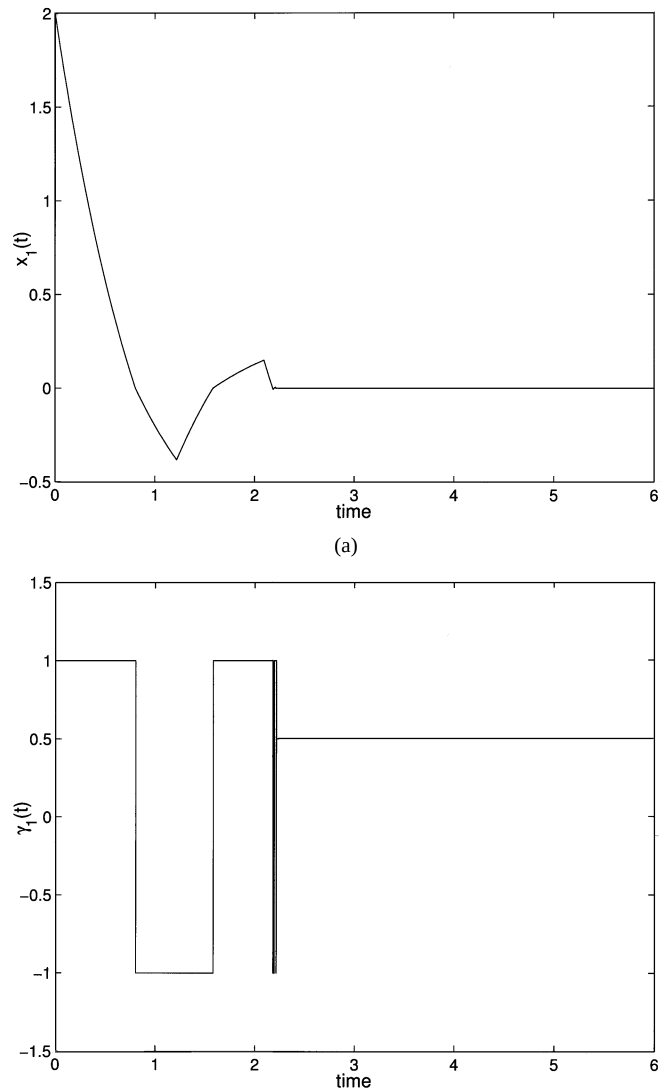


Fig. 5. (a) Time-domain evolution of the state variable $x_1(t)$. (b) Evolution of the corresponding output $\gamma_1(t)$, for the neural network of Example 5.

Let us now verify the estimate (26) (or (31)) of the convergence time. It results $\lambda_m = 1/4$ and $\delta = 1$, hence from (31) we get $t_c \leq t_e = 4\{|x_1(0)| + |x_2(0)|\}$. For the trajectory considered in Fig. 4, which corresponds to $x(0) = (2,2)'$, we obtain $t_e = 16$. From Fig. 4, it is seen that the actual convergence time t_c is less than 2.5, in agreement with (31).[4] ∎

Example 5: In the final example, we add a nonzero input to the neural network (32), i.e., we consider

$$\begin{cases} \dot{x}_1 = -x_1 - \frac{1}{4}\mathrm{sgn}(x_1) - \mathrm{sgn}(x_2) - \frac{3}{8} \\ \dot{x}_2 = -x_2 + \mathrm{sgn}(x_1) - \frac{1}{4}\mathrm{sgn}(x_2) - \frac{5}{8}. \end{cases} \quad (33)$$

From Example 3, $I = (-3/8, -5/8)' \in \mathcal{I}_{\mathrm{FT}}$. From (1), it is easy to verify that $\phi_1(0) = [-13/8, 7/8]$, $\phi_2(0) = [-15/8, 5/8]$. Hence, $0 \in \phi(0)$, i.e., $e = 0$ is still an EP of the neural network (cf. (3)). The EP $e = 0$ is unique, however the nonzero input has shifted the unique corresponding OEP to $\eta = (1/2, -1/2)'$. The hypotheses of Theorem 4 are satisfied and (33) is globally convergent in finite time. Now it results $\delta = 1/2$, hence the estimated convergence time of a trajectory of (33) starting at $x(0)$ is twice that of a trajectory of (32) with the same initial condition. As an example, for a trajectory starting at $x_0 = (2,2)'$, we have $t_c \leq t_e = 32$. Once more, these results are confirmed by numerical simulations, see Fig. 5. ∎

[4]We also see that the estimate t_e of t_c in (26) (or (31)) is quite conservative in this example. This is due to the fact that some terms in the expression of the time derivative of the energy V, as it was obtained in Lemma 1, were not accounted for to arrive at (26). Taking into account such terms would render the expression of the estimated convergence time more complex and less straightforward to apply with respect to (26).

VI. Conclusion

The paper has performed a thorough analysis of global convergence for a large class of neural networks where the neuron activations are modeled by discontinuous functions.

It has been shown that the condition of LDS neuron interconnection matrices ensures global convergence of both the state and the output trajectories of the neural network toward a unique equilibrium point. These results represent a generalization to the discontinuous case of previous results obtained for neural networks described by smooth vector fields.

Moreover, entirely new conditions have been obtained which guarantee global convergence of the state and the output trajectories in *finite time*. Explicit and easy-to-apply formulas have been derived to evaluate the finite convergence time of each trajectory. The conditions involve the LDS concept and, in addition, suitable assumptions on the location of the equilibrium point, with respect to the discontinuities of the neuron activations, which admit a clear geometrical interpretation.

Future work aims at exploiting the results in this paper, notably the presence of sliding modes, and the conditions for global convergence in finite time, to efficiently solve classes of optimization problems arising in practical applications.

Appendix I
Proof of Proposition 2

We will employ the next known result.

Proposition 5 [7, p. 313, Th. 4]: Suppose that $\gamma(t)$: $[0,+\infty) \to C$ is a measurable function, and $C \subset \mathbb{R}^n$ is a compact set. Then, $\gamma(t)$ has at least an almost cluster point $\gamma^\star \in C$ as $t \to +\infty$.

Observe that if $\gamma^\star$ is the unique almost cluster point of $\gamma(t)$ as $t \to +\infty$, then

$$\forall \epsilon > 0 \quad \mu\{t \in [0,+\infty) : \|\gamma(t) - \gamma^\star\| > \epsilon\} < +\infty. \quad (34)$$

To prove (34), suppose for purposes of contradiction that there exists $\epsilon_0 > 0$ such that $\mu\{t \in [0,+\infty) : \|\gamma(t) - \gamma^\star\| > \epsilon_0\} = +\infty$. Let $E_0 = \{t \in [0,+\infty) : \|\gamma(t) - \gamma^\star\| > \epsilon_0\}$, and consider function $\hat{\gamma}$: $E_0 \to C$, which is defined as the restriction of γ to E_0, where $\mu(E_0) = +\infty$. Then, by using an argument as in the proof of [7, p. 313, Th. 4], it turns out that $\hat{\gamma}$ has an almost cluster point $\hat{\gamma}^\star$ different from $\gamma^\star$. This contradicts the uniqueness of the almost cluster point.

Now, we show that (34) implies $\gamma^\star = \mu\lim_{t\to+\infty} \gamma(t)$. Once more suppose for purposes of contradiction that there exists $\epsilon_0 > 0$ such that for any $\tilde{t} \geq 0$ it results $\mu\{t \in [\tilde{t},+\infty) : \|\gamma(t) - \gamma^\star\| > \epsilon_0\} \geq \epsilon_0$. Pick $\tilde{t}_0 = 0$ and let $E_0 = \{t \in [0,+\infty) :$

$\|\gamma(t) - \gamma^\star\| > \epsilon_0\}$. Then, $\epsilon_0 \leq \mu(E_0)$. Therefore, due to the definition of Lebesgue measure of a set [30], there exists a compact set $C_0 \subset E_0$ such that $\mu(C_0) > \epsilon_0/2$. Let $a_0 = \max C_0$, $\tilde{t}_1 > a_0$, and $E_1 = \{t \in [\tilde{t}_1, +\infty) : \|\gamma(t) - \gamma^\star\| > \epsilon_0\}$. Again, we have $\epsilon_0 \leq \mu(E_1)$. Hence, there is a compact set $C_1 \subset E_1 \subset E_0$ such that $\mu(C_1) > \epsilon_0/2$. By induction, we can construct a sequence of compact sets $\{C_i\}_{i=1}^{+\infty}$ such that $C_i \subset E_0$, $C_i \cap C_j = \emptyset$ for each $i \neq j$, and $\mu(C_i) > \epsilon_0/2$ for each i.

In conclusion, $\mu(E_0) > \mu(\cup_{i=1}^{+\infty} C_i) = \Sigma_{i=1}^{+\infty} \mu(C_i) = +\infty$, which contradicts (34) ∎

APPENDIX II
PROOF OF PROPERTY 3

For simplicity we verify the property for $\tau = 1$. For $t \geq 0$ we have

$$y(t) = e^{-t} y(0) + \int_0^t \gamma(u) e^{-(t-u)} du = e^{-t} y(0) + \Upsilon(t).$$

Therefore, it suffices to show that $\lim_{t\to+\infty} \Upsilon(t) = \eta$. To this end, recall that by hypothesis, for all $\epsilon > 0$, there exists $\tilde{t}_\epsilon > 0$ such that $\mu\{t \in [\tilde{t}_\epsilon, +\infty) : \|\gamma(t) - \eta\| > \epsilon\} < \epsilon$. For $t > \tilde{t}_\epsilon$ function $\|\Upsilon(t) - \eta(1 - e^{-t})\| = \| \int_0^t (\gamma(u) - \eta) e^{-(t-u)} du\|$ satisfies

$$\left\|\Upsilon(t) - \eta(1 - e^{-t})\right\| \leq \int_0^{\tilde{t}_\epsilon} \|\gamma(u) - \eta\| \, e^{-(t-u)} du + \int_{\tilde{t}_\epsilon}^t \|\gamma(u) - \eta\| \, e^{-(t-u)} du.$$

Since $\gamma(u)$ and also $\gamma(u) - \eta$ take their values in a compact set, we have for some $\tilde{c} > 0$, $\int_0^{\tilde{t}_\epsilon} \|\gamma(u) - \eta\| e^{-(t-u)} du \leq \tilde{c} e^{-t}(e^{\tilde{t}_\epsilon} - 1)$. Moreover, $\int_{\tilde{t}_\epsilon}^t \|\gamma(t) - \eta\| e^{-(t-u)} du \leq \int_{E_\epsilon^-(t)} \|\gamma(t) - \eta\| e^{-(t-u)} du + \int_{E_\epsilon^+(t)} \|\gamma(t) - \eta\| e^{-(t-u)} du$, where $E_\epsilon^-(t) = \{u \in [\tilde{t}_\epsilon, t) : \|\gamma(u) - \eta\| \leq \epsilon\}$, $E_\epsilon^+(t) = \{u \in [\tilde{t}_\epsilon, t) : \|\gamma(u) - \eta\| > \epsilon\}$, and $\mu(E_\epsilon^+(t)) < \epsilon$ for all $t \geq \tilde{t}_\epsilon$.

Since $\tilde{t}_\epsilon \leq u < t$, and $\|\gamma(u) - \eta\| < \tilde{c}$, $\int_{\tilde{t}_\epsilon}^t \|\gamma(u) - \eta\| e^{-(t-u)} du \leq \epsilon \int_{E_\epsilon^-(t)} e^{-(t-u)} du + \int_{E_\epsilon^+(t)} \|\gamma(u) - \eta\| du \leq \epsilon(1 - e^{-(t-\tilde{t}_\epsilon)}) + \tilde{c}\epsilon$. In conclusion, for each $t \geq \tilde{t}_\epsilon$ it results

$$\left\|\Upsilon(t) - \eta(1 - e^{-t})\right\| < \tilde{c} e^{-t}(e^{\tilde{t}_\epsilon} - 1) + (\tilde{c} + 1)\epsilon.$$

By taking the limit as $t \to +\infty$, we obtain $\lim_{t\to+\infty} \|\Upsilon(t) - \eta\| < (\tilde{c}+1)\epsilon$. Since ϵ is an arbitrary positive number, it follows that $\lim_{t\to+\infty} \Upsilon(t) = \eta$. ∎

APPENDIX III
PROOF OF PROPERTY 5

We begin by showing that maps S_e and S_η are continuous on $\mathbb{R}^n$. Let $I^j \in \mathbb{R}^n$, $S_e(I^j) = e^j$ and $S_\eta(I^j) = \eta^j$, $j = 1$, 2. Suppose, without loss of generality, that it results $\eta_i^1 \neq \eta_i^2$ for $i = 1, \cdots, m$, and $\eta_i^1 = \eta_i^2$ for $i = m+1, \cdots, n$, where $1 \leq m \leq n$. According to the argument and the notations in the proof of Theorem 1, let us consider the system

$$\begin{cases} -B^{aa}(e^{1a} - e^{2a}) - T^{aa}(\eta^{1a} - \eta^{2a}) = I^{1a} - I^{2a} \\ -B^{bb}(e^{1b} - e^{2b}) - T^{ba}(\eta^{1a} - \eta^{2a}) = I^{1b} - I^{2b}. \end{cases}$$

Since functions g_i, $i = 1, \cdots, m$, are nondecreasing, we have $e_i^{1a} - e_i^{2a} = H_i(\eta_i^{1a} - \eta_i^{2a})$ for some $H_i \geq 0$, and $i = 1, \cdots, m$. Therefore, recalling that $\det(B^{aa} H^{aa} + T^{aa}) \neq 0$, it results $\eta^{1a} - \eta^{2a} = -(B^{aa} H^{aa} + T^{aa})^{-1}(I^{1a} - I^{2a})$, where $H^{aa} = \mathrm{diag}(H_1, \cdots, H_m)$. This, together with the fact that $\eta^{1b} = \eta^{2b}$, proves the continuity of S_η. Furthermore, from the equation $e^1 - e^2 = -B^{-1}[T(\eta^1 - \eta^2) + (I^1 - I^2)]$ it follows also the continuity of S_e.

Now, let us prove that map Σ is injective. In fact, let $I^1, I^2 \in \mathbb{R}^n$ with $I^1 \neq I^2$ and assume, for purposes of contradiction, that $\Sigma(I^1) = \Sigma(I^2)$. In this case it results $Be^1 + T\eta^1 = -I^1$ and $Be^1 + T\eta^1 = -I^2$. Hence, $I^1 = I^2$, a contradiction.

Furthermore, from the continuity of Σ, and from the equation $B(e^1 - e^2) + T(\eta^1 - \eta^2) = I^2 - I^1$, where $I^j = \Sigma^{-1}(S_e(I^j), S_\eta(I^j))$, $j = 1$, 2, the continuity of Σ^{-1} immediately follows. Therefore, Σ is a homeomorphism from $\mathbb{R}^n$ to $V = \{\Sigma(I) \in \mathbb{R}^n \times \mathbb{R}^n, I \in \mathbb{R}^n\}$.

Finally, let us show that $V = U$. In fact, let $(x, y) \in U$ and $-I = Bx + Ty$, and consider $\Sigma(I) = (S_e(I), S_\eta(I)) = (e, \eta)$. We want to prove that $(e, \eta) = (x, y)$. To this end, note that we have $B(x - e) + T(y - \eta) = 0$. Therefore, by the same argument as in the proof of Theorem 1, it is an easy matter to show that it results $x = e$ and $y = \eta$. ∎

APPENDIX IV
PROOF OF LEMMA 1

The next technical result is crucial for computing $\dot{V}(z(t))$ and prove Lemma 1.

Proposition 6 (Chain rule): Suppose that $g \in \mathcal{G}_\mathcal{D}$. Consider function V as defined in (13), and let $z(t)$, $t \in [0, +\infty)$, be any solution of (N2). Then, $V(z(t))$ is differentiable for a.a. $t \geq 0$ and it results

$$\dot{V}(z(t)) = \xi'(t)\dot{z}(t) \qquad \forall \xi(t) \in \partial V(z(t)).$$

Proof of Proposition 6: Since $G \in \mathcal{G}_\mathcal{D}$ (cf. (i) of Proposition 3), it follows that V is a convex, locally Lipschitz continuous function. Then, ∇V is defined almost everywhere in $\mathbb{R}^n$ and it is locally bounded. Moreover, any solution $z(t)$ of (12) is locally Lipschitz, since $z(t)$ is locally bounded. This implies that $z(t)$ is strictly differentiable for a.a. $t \geq 0$ [24, Proposition 2.2.4], and V is regular at any $z \in \mathbb{R}^n$ [24, Proposition 2.3.6]. Therefore, the result in the proposition follows by applying [24, Proposition 2.3.9-(iii)]. ∎

To explicitly evaluate $\dot{V}(z(t))$, note that it results from (13)

$$\partial V(z) = -2B^{-1}z + 2c\alpha K[G(z)].$$

Moreover, from Proposition 6 we have for a.a. $t \geq 0$

$$\dot{V}(z(t)) = \xi'(t)\dot{z}(t) \qquad \forall \xi(t) \in \partial V(z(t)).$$

Let us choose

$$\xi(t) = -2B^{-1}z(t) + 2c\alpha\tilde{\gamma}(t) \in \partial V(z(t))$$

where $\tilde{\gamma}(t) \in K[G(z(t))]$ is defined in (12). Then

$$\dot{V}(z(t)) = \left[-2B^{-1}z(t) + 2c\alpha\tilde{\gamma}(t)\right]' [Bz(t) + T\tilde{\gamma}(t)]$$

with $\tilde{\gamma}(t) \in K[G(z(t))]$. Therefore, we obtain

$$\begin{aligned}\dot{V}(z(t)) =& -2z'(t)z(t) - 2z'(t)[B^{-1}T - c\alpha B]\tilde{\gamma}(t)\\ &- 2c\tilde{\gamma}'(t)\,[\alpha(-T)]^S\,\tilde{\gamma}(t)\\ =& -2z'(t)z(t) - 2z'(t)B^{-1}T\tilde{\gamma}(t)\\ &- 2c\tilde{\gamma}'(t)\,[\alpha(-T)]^S\,\tilde{\gamma}(t) + 2cz'(t)\alpha B\tilde{\gamma}(t)\end{aligned}$$

where $2cz'(t)\alpha B\tilde{\gamma}(t) \le 0$ on the basis of (iii) of Proposition 3.

By adding and subtracting $\tilde{\gamma}'(t)[(B^{-1}T)'(B^{-1}T)]\tilde{\gamma}(t)$, and accounting for (16), it results

$$\begin{aligned}\dot{V}(z(t)) =& -2z'(t)z(t) - 2z'(t)B^{-1}T\tilde{\gamma}(t)\\ &-\tilde{\gamma}'(t)\left[(B^{-1}T)'(B^{-1}T)\right]\tilde{\gamma}(t)\\ &-\tilde{\gamma}'(t)\left\{2c\,[\alpha(-T)]^S - (B^{-1}T)'(B^{-1}T)\right\}\tilde{\gamma}(t)\\ &+2cz'(t)\alpha B\tilde{\gamma}(t)\\ =& -\|z(t)\|_2^2 - \left\|z(t) + B^{-1}T\tilde{\gamma}(t)\right\|_2^2\\ &-\tilde{\gamma}'(t)\left\{2c\,[\alpha(-T)]^S - (B^{-1}T)'(B^{-1}T)\right\}\tilde{\gamma}(t)\\ &+2cz'(t)\alpha B\tilde{\gamma}(t)\\ \le& -\|z(t)\|_2^2 - \left\|B^{-1}\left(Bz(t) + T\tilde{\gamma}(t)\right)\right\|_2^2 - \lambda\,\|\tilde{\gamma}(t)\|_2^2\\ =& -\|z(t)\|_2^2 - \left\|B^{-1}\dot{z}(t)\right\|_2^2 - \lambda\,\|\tilde{\gamma}(t)\|_2^2.\end{aligned}$$

■

APPENDIX V
PROOF OF PROPERTY 6

By the continuity of the map Σ defined in (8), it follows that set $\mathcal{I}_{\rm GA}$ is open.

Then, we show that $\partial\mathcal{I}_{\rm GA} = \mathcal{I}'_{\rm GA}$. Let $I^0 \in \mathcal{I}'_{\rm GA}$, hence $I^0 \notin \mathcal{I}_{\rm GA}$. We begin by proving that any neighborhood $\mathcal{V}$ of I^0 contains some $I \in \mathcal{I}_{\rm GA}$, and then $I^0 \in \partial\mathcal{I}_{\rm GA}$, i.e., $\mathcal{I}'_{\rm GA} \subset \partial\mathcal{I}_{\rm GA}$. Since Σ is a homeomorphism from $\mathbb{R}^n$ onto the graph U of $K[g]$ (see Property 5), it sends open sets into open sets, hence $\Sigma(\mathcal{V})$ is a neighborhood of $\Sigma(I^0) = (e^0, \eta^0)$ contained in the graph of $K[g]$. Therefore, there exists $I \in \mathcal{V}$ such that for any $i = 1, \cdots, n$ it results $i \in \Theta^C(e)$ with $e_i \neq e_i^0$, or $i \in \Theta^D(e)$ with $\eta_i \in {\rm int}K[g(e_i)]$ and $e_i = e_i^0$. Thus, $I \in \mathcal{I}_{\rm GA}$. Furthermore, $\partial\mathcal{I}_{\rm GA} \subset \mathcal{I}'_{\rm GA}$. In fact, let $I^0 \in \partial\mathcal{I}_{\rm GA}$. Since $I^0 \notin \mathcal{I}_{\rm GA}$ and $\mathcal{I}_{\rm GA}$ is open, we have $I^0 \in \mathcal{I}'_{\rm GA}$.

Finally, it remains to show that $\mu(\mathcal{I}'_{\rm GA}) = 0$. To this end, we observe that if we drop a row and a corresponding column of the matrices B and T, and the corresponding component of the vector I in (N1), we obtain a system describing the dynamical behavior of the remaining $n-1$ neurons. Such a system satisfies the same assumptions as the original one. Thus, in particular, if we denote by $\hat{\Sigma}$ the map corresponding to the map Σ for the original system (N1) (cf. (8)), we have that $\hat{\Sigma}$ is a homeomorphism and $\hat{\Sigma}(\mathbb{R}^{n-1}) = \hat{U}$, where $\hat{U}$ is the graph of the set-valued map $K[\hat{G}]$ which is obtained by eliminating the activation of the suppressed neuron (cf. Property 5).

Suppose now, without loss of generality, that $(e_1, \eta_1) = (0, 0)$, $0 = G_1(0^-)$ or $0 = G_1(0^+)$, where the pair (e_1, η_1) denotes the first component of (e, η) and ${\rm int}([G_1(0^-), G_1(0^+)]) \neq \emptyset$. Denote by $\hat{B}$ and $\hat{T}$ the matrices obtained by suppressing the first row and first column of B and T respectively.

Then, consider the sets

$$\mathcal{I}_1 = \left\{I \in \mathbb{R}^n : I = -Be - T\eta, (e,\eta) = ((0,\hat{e}), (0,\hat{\eta})), (\hat{e},\hat{\eta}) \in \hat{U}\right\} \subset \mathcal{I}'_{\rm GA}$$

$$\hat{\mathcal{I}}_1 = \left\{\hat{I} \in \mathbb{R}^{n-1} : \hat{I} = -\hat{B}\hat{e} - \hat{T}\hat{\eta}, (\hat{e},\hat{\eta}) \in \hat{U}\right\}.$$

For any $I \in \mathcal{I}_1$ we have $I = (-\Sigma_{k=2}^n T_{1k}\hat{\eta}_k, -\hat{B}\hat{e} - \hat{T}\hat{\eta})'$, where $\hat{\eta} = (\hat{\eta}_2, \cdots, \hat{\eta}_n)'$. Furthermore, one can easily see that $\sum_{k=2}^n T_{1k}\hat{\eta}_k = \langle u, \hat{T}\hat{\eta}\rangle$, where $\langle\cdot,\cdot\rangle$ denotes the scalar product associated to the Euclidean norm and u is the unique solution of the algebraic system $\hat{T}u = \hat{T}_1$, with $\hat{T}_1 = (T_{12}, \cdots, T_{1n})'$. Therefore, if we denote by Σ_1 the map from $\hat{U}$ onto $\mathcal{I}_1$ given by $I = \Sigma_1(\hat{e}, \hat{\eta})$, where $I = (\langle u, \hat{T}\hat{\eta}\rangle, \hat{I})'$, it is straightforward to prove that it is a homeomorphism.

In conclusion

$$\mathcal{I}_1 = \Sigma_1\left(\hat{\Sigma}(\mathbb{R}^{n-1})\right) = \hat{\Sigma}_1(\mathbb{R}^{n-1})$$

where $\hat{\Sigma}_1 = \Sigma_1 \circ \hat{\Sigma}$ is a homeomorphism. Hence, $\mathcal{I}_1$ is a $(n-1)$-dimensional manifold in $\mathbb{R}^n$ with global parameterization $\hat{\Sigma}_1$. Therefore, $\mu(\mathcal{I}_1) = 0$.

Finally, since the number of discontinuities of map G is countable, the set $\partial\mathcal{I}_{\rm GA}$ is contained in a countable number of sets of measure zero and so $\partial\mathcal{I}_{\rm GA}$ has measure zero. ■

APPENDIX VI
PROOF OF LEMMA 2

By an argument analogous to that in the proof of Lemma 1 (see Appendix IV), we get

$$\begin{aligned}\dot{V}_D\left(z^D(t)\right) \le& -\left\|z^D(t)\right\|_2^2 - \left\|(B^{DD})^{-1}\dot{z}^D(t)\right\|_2^2 - \lambda_{mD}\left\|\tilde{\gamma}^D(t)\right\|_2^2\\ &+\left[-2(B^{DD})^{-1}z^D(t) + 2c\alpha^{DD}\tilde{\gamma}^D(t)\right]' T^{DC}\tilde{\gamma}^C(t)\end{aligned}$$

for a.a. $t \ge 0$, where $\tilde{\gamma}^D(t) \in K[G^D(z^D(t))]$. Then

$$\dot{V}_D\left(z^D(t)\right) \le -\lambda_{mD}\left\|\tilde{\gamma}^D(t)\right\|_2^2 + \xi(t)$$

where $\xi(t) = [-2(B^{DD})^{-1}z^D(t) + 2c\alpha^{DD}\tilde{\gamma}^D(t)]'T^{DC}\tilde{\gamma}^C(t) \to 0$ as $t \to +\infty$, since $\tilde{\gamma}^C(t) \to 0$ as $t \to +\infty$, and $\tilde{\gamma}^D(t)$, $z^D(t)$ are bounded.

Suppose that $z^D(t) \neq 0$. Let us define $\delta_D = \min_{i\in\Theta^D(e)}\{\min\{\delta_i^+, \delta_i^-\}\} > 0$ (cf. (21) and (22)). By noting that $\tilde{\gamma}^D(t) \in K[G^D(z^D(t))]$ and that functions g_i, $i = 1, \cdots, n$, are nondecreasing, we obtain $\|\tilde{\gamma}^D(t)\|_2 \ge \delta_D$. Hence

$$\dot{V}_D\left(z^D(t)\right) \le -\lambda_{mD}\delta_D^2 + \xi(t).$$

Since $\xi(t) \to 0$ as $t \to +\infty$, there exists $\bar{t} > 0$ such that for $t > \bar{t}$ and $z^D(t) \neq 0$ it results in

$$\dot{V}_D\left(z^D(t)\right) < -\frac{1}{2}\lambda_{mD}\delta_D^2 = -\Delta_D < 0.$$

■

ACKNOWLEDGMENT

The authors wish to thank the anonymous reviewers for constructive comments which helped to improve the presentation of the paper.

REFERENCES

[1] J. J. Hopfield, "Neurons with graded response have collective computational properties like those of two-state neurons," *Proc. Nat. Acad. Sci.*, vol. 81, pp. 3088–3092, 1984.
[2] D. W. Tank and J. J. Hopfield, "Simple 'neural' optimization networks: an A/D converter, signal decision circuit, and a linear programming circuit," *IEEE Trans. Circuits Syst.*, vol. 33, pp. 533–541, May 1986.
[3] H. Harrer, J. A. Nossek, and R. Stelzl, "An analog implementation of discrete-time cellular neural networks," *IEEE Trans. Neural Networks*, vol. 3, pp. 466–476, May 1992.
[4] M. P. Kennedy and L. O. Chua, "Neural networks for nonlinear programming," *IEEE Trans. Circuits Syst. I*, vol. 35, pp. 554–562, May 1988.
[5] L. O. Chua, C. A. Desoer, and E. S. Kuh, *Linear and Nonlinear Circuits*. New York: McGraw-Hill, 1987.
[6] V. I. Utkin, *Sliding Modes and Their Application in Variable Structure Systems*. Moscow, U.S.S.R.: MIR, 1978.
[7] J. P. Aubin and A. Cellina, *Differential Inclusions*. Berlin, Germany: Springer-Verlag, 1984.
[8] B. E. Paden and S. S. Sastry, "Calculus for computing Filippov's differential inclusion with application to the variable structure control of robot manipulator," *IEEE Trans. Circuits Syst.*, vol. 34, pp. 73–82, Jan. 1987.
[9] T. Roska, C. W. Wu, and L. O. Chua, "Stability of cellular neural networks with dominant nonlinear and delay-type templates," *IEEE Trans. Circuits Syst. I*, vol. 40, pp. 270–271, Apr. 1993.
[10] E. Kaszkurewicz and A. Bhaya, "On a class of globally stable neural circuits," *IEEE Trans. Circuits Syst. I*, vol. 41, pp. 171–174, Feb. 1994.
[11] M. Forti, "On global asymptotic stability of a class of nonlinear systems arising in neural network theory," *J. Different. Eq.*, vol. 113, pp. 246–264, 1994.
[12] M. Forti and A. Tesi, "New conditions for global stability of neural networks with application to linear and quadratic programming problems," *IEEE Trans. Circuits Syst. I*, vol. 42, pp. 354–366, July 1995.
[13] X. Liang and J. Wang, "An additive diagonal stability condition for absolute stability of a general class of neural networks," *IEEE Trans. Circuits Syst. I*, vol. 48, pp. 1308–1317, Nov. 2001.
[14] Y. Xia and J. Wang, "A projection neural network and its application to constrained optimization problems," *IEEE Trans. Circuits Syst. I*, vol. 49, pp. 447–458, Apr. 2002.
[15] S. Arik and V. Tavsanoglu, "On the global asymptotic stability of delayed cellular neural networks," *IEEE Trans. Circuits Syst. I*, vol. 47, pp. 571–574, Apr. 2000.
[16] J. Cao, "A set of stability criteria for delayed cellular neural networks," *IEEE Trans. Circuits Syst. I*, vol. 48, pp. 494–498, Apr. 2001.
[17] X. Liang, "A recurrent neural network for nonlinear continuously differentiable optimization over a compact convex subset," *IEEE Trans. Neural Networks*, vol. 12, pp. 1487–1490, Nov. 2001.
[18] E. K. P. Chong, S. Hui, and S. H. Żak, "An analysis of a class of neural networks for solving linear programming problems," *IEEE Trans. Automat. Contr.*, vol. 44, pp. 1095–2006, Nov. 1999.
[19] M. P. Glazos, S. Hui, and S. H. Żak, "Sliding modes in solving convex programming problems," *SIAM J. Control Optim.*, vol. 36, pp. 680–697, Mar. 1998.
[20] L. O. Chua and L. Yang, "Cellular neural networks: theory," *IEEE Trans. Circuits Syst.*, vol. 35, pp. 1257–1272, 1988.
[21] J.-H. Li, A. N. Michel, and W. Porod, "Analysis and synthesis of a class of neural networks: variable structure systems with infinite gain," *IEEE Trans. Circuits Syst.*, vol. 36, pp. 713–731, May 1989.
[22] M. Hirsch, "Convergent activation dynamics in continuous time networks," *Neural Networks*, vol. 2, pp. 331–349, 1989.
[23] A. F. Filippov, *Differential Equations With Discontinuous Right-Hand Side. Mathematics and its Applications (Soviet Series)*. Boston, MA: Kluwer Academic, 1988.
[24] F. H. Clarke, *Optimization and Non-Smooth Analysis*. New York: Wiley, 1969.
[25] M. Fiedler and V. Pták, "Some generalization of positive definiteness and monotonicity," *Numer. Math.*, vol. 9, pp. 162–173, 1966.
[26] D. Hershkowitz, "Recent directions in matrix stability," *Linear Alg. Appl.*, vol. 171, pp. 161–186, 1992.
[27] J. P. Aubin and H. Frankowska, *Set-Valued Analysis*. Boston, MA: Birkauser, 1990.
[28] M. Di Marco, M. Forti, and A. Tesi, "Bifurcations and oscillatory behavior in a class of competitive cellular neural networks," *Int. J. Bifurcation Chaos*, vol. 10, pp. 1267–1293, June 2000.
[29] ——, "Existence and characterization of limit cycles in nearly symmetric neural networks," *IEEE Trans. Circuits Syst. I*, vol. 49, pp. 979–992, July 2002.
[30] H. L. Royden, *Real Analysis*. New York: MacMillan, 1988.

M. Forti received the laurea degree in electronic engineering from the University of Florence, Florence, Italy, in 1988.

From 1991 to 1998, he was an Assistant Professor in applied mathematics and network theory in the Electronic Engineering Department, University of Florence. In 1998, he joined the Department of Information Engineering of the University of Siena, Siena, Italy, where he is currently an Associate Professor of Electrical Engineering. His main research interests are in the field of nonlinear circuits and systems, with emphasis on the qualitative analysis and stability of circuits modeling artificial neural networks. His research activity also includes some aspects of electromagnetic compatibility.

Dr. Forti presently serves as an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS—I: FUNDAMENTAL THEORY AND APPLICATIONS and the IEEE TRANSACTIONS ON NEURAL NETWORKS.

P. Nistri was born in Rignano sull'Arno, Florence, Italy, in 1948. He received the doctoral degree in mathematics from the University of Florence, Florence, Italy, in 1972.

From 1974 to 1979, he was an Assistant Professor at the University of Calabria, Cosenza, Italy. In 1979, he moved to the Engineering Faculty of the University of Florence, where, from 1982 to 1998, he was an Associate Professor. Since 1998, he is at the Engineering Faculty at the University of Siena, Siena, Italy, where he is a Full Professor of Mathematical Analysis. His main scientific interests are in the field of dynamical systems, mathematical control theory, differential inclusions and topological methods. He is the author of around 100 scientific publications, and the organizer and coordinator of several international scientific activities.

Dr. Nistri is an Associate Editor of several international mathematical journals and serves as referee for many of them.