Available online at www.sciencedirect.com
SCIENCE DIRECT®
Physica D 214 (2006) 88–99
PHYSICA D
www.elsevier.com/locate/physd

# Generalized Lyapunov approach for convergence of neural networks with discontinuous or non-Lipschitz activations

M. Forti*, M. Grazzini, P. Nistri, L. Pancioni

*Dipartimento di Ingegneria dell'Informazione, Università di Siena, v. Roma, 56 - 53100 Siena, Italy*

Received 31 August 2005; accepted 22 December 2005

Communicated by A. Doelman

**Abstract**

The paper considers a class of additive neural networks where the neuron activations are modeled by discontinuous functions or by continuous non-Lipschitz functions. Some tools are developed which enable us to apply a Lyapunov-like approach to differential equations with discontinuous right-hand side modeling the neural network dynamics. The tools include a chain rule for computing the time derivative along the neural network solutions of a nondifferentiable Lyapunov function, and a comparison principle for this time derivative, which yields conditions for exponential convergence or convergence in finite time. By means of the Lyapunov-like approach, a general result is proved on global exponential convergence toward a unique equilibrium point of the neural network solutions. Moreover, new results on global convergence in finite time are established, which are applicable to neuron activations with jump discontinuities, or neuron activations modeled by means of continuous (non-Lipschitz) Hölder functions.


*Keywords:* Discontinuous neural networks; Global exponential stability; Convergence in finite time; Lyapunov approach; Generalized gradient; $M$-matrices and $H$-matrices

## 1. Introduction

It clearly emerges, in the recent literature, that there is an increasing interest in the use of neural network models with discontinuous or non-Lipschitz neuron activations. Next, we highlight the crucial role played by such nonsmooth activations, by discussing some applications taken from the field of continuous-time neural optimization solvers and the field of computational models based on analog recurrent neural networks.

In [1], a class of additive neural networks possessing activations with jump discontinuities has been introduced, and the issue of global convergence toward a unique equilibrium point has been addressed. We recall that global convergence is important in the solution of optimization problems in real time, since it prevents a network from the risk of getting stuck at some local minimum of the energy function, see e.g. [2–14], and references therein. By exploiting the sliding modes in the neural network dynamics, which are intrinsically related to the presence of jump discontinuities, conditions for *global convergence in finite time* have been established in [1, Theorem 4]. As is discussed also in [15], the property of convergence in finite time is especially desirable in the design of real time neural optimization solvers, and it cannot be achieved in smooth dynamical systems, since in that case there can be only asymptotic convergence toward an equilibrium point.

In a recent paper [16], a generalized nonlinear programming circuit with a neural-like architecture has been introduced to solve a large class of constrained optimization problems. As in a penalty method, the dynamics is described by a gradient system of an energy function, which is the sum of the objective function and a barrier function preventing solutions from staying outside the feasibility region. One crucial feature is that, by using constraint neurons modeled by ideal diodes with a vertical segment in the conducting region, the circuit in [16] can implement an *exact penalty method* where the circuit

* Corresponding author. Tel.: +39 0577 233601; fax: +39 0577 233602.
*E-mail address:* forti@dii.unisi.it (M. Forti).

0167-2789/$ - see front matter 
doi:10.1016/j.physd.2005.12.006

equilibrium points coincide with the constrained critical points of the objective function. This property enables the circuit to compute the exact constrained optimal solution for interesting classes of programming problems.

We remark that the results in [16] are in agreement with the general theory in [17], according to which an exact penalty method requires the use of a nondifferentiable barrier function. Therefore, a neural network solver is necessarily characterized by constraint nonlinearities with vertical straight segments, which correspond to the generalized gradient of the nondifferentiable barrier function. Along this line we can also interpret the design procedure, which has been devised in [18], for a class of neural networks with discontinuous neuron activations aimed at solving linear programming problems via a generalized gradient system based on an exact penalty function method. A further related case concerns the discontinuous neural networks proposed in the recent paper [19], which can be used for finding the exact solution of underdetermined linear systems of algebraic equations, as well as those arising in least squares problems for support vector machines.

Finally, we mention the class of analog recurrent neural networks proposed in [20], where a classical recurrent neural network [21] is augmented with a few simple discontinuous neuron activations (e.g., binary threshold functions). It is shown that the use of these discontinuous activations permits us to significantly *increase the computation power*, by enabling operations as products or divisions on the network inputs, or the implementation of other more complex recursive functions. Another fundamental observation made in [20] is that the same computation power can be achieved by simulating the discontinuity with a “clear enough discontinuity”, i.e., by replacing a discontinuous function with a continuous function with a *non-Lipschitz* part, as for example a square root function. Roughly speaking, this is due to the fact that discontinuous functions and non-Lipschitz functions share the peculiar property that even small variations of the neuron state are able to produce significant changes in the neuron output.

In the first part of this paper (Sections 2 and 3), we introduce some tools which enable to apply a generalized Lyapunov-like approach to a class of differential equations with discontinuous right-hand side. The tools include a general chain rule for computing the time derivative along the solutions of a nondifferentiable Lyapunov function, and a comparison principle which gives conditions for exponential convergence, or convergence in finite time. The second part of the paper (Sections 4–6) is aimed at demonstrating the applicability of the generalized Lyapunov approach. To this end, we exploit the approach for obtaining further results on global convergence for the class of additive neural networks with discontinuous neuron activations introduced in [1]. More specifically, a general condition ensuring global exponential convergence toward a unique equilibrium point, with a known convergence rate, is established in the case of neuron activations with jump discontinuities. Furthermore, new conditions for convergence in finite time, with a quantitative estimate of convergence time, are obtained for a class of continuous non-Lipschitz neuron activations modeled by means of Hölder functions. This class includes the square root functions, and other classes of non-Lipschitz functions of potential interest for the applications. The paper ends with some concluding remarks (Section 7).

The results on global convergence are obtained by assuming that the neural network interconnection matrix is Lyapunov Diagonally Stable or it is modeled by an $M$-matrix or an $H$-matrix. Such classes of matrices have been frequently considered in the literature to address global convergence of neural networks in the standard case of Lipschitz continuous neuron activations, see, e.g., [5–7,9] and their references. The concept of $M$-matrices and $H$-matrices is also at the basis of well established techniques for the qualitative analysis of large-scale interconnected dynamical systems [22,23]. The reader is referred to [22,24,25] for a general treatment on the quoted classes of matrices and a discussion of some of their fundamental applications in engineering and physics.

**Notation.** Consider the column vectors $x = (x_1, \dots, x_n)^\top \in \mathbb{R}^n$ and $y = (y_1, \dots, y_n)^\top \in \mathbb{R}^n$, where the symbol $\top$ means the transpose. By $\langle x, y\rangle = x^\top y = \sum_{i=1}^n x_i y_i$ we mean the scalar product of $x, y$, while $\|x\| = \langle x, x\rangle^{1/2} = (\sum_{i=1}^n x_i^2)^{1/2}$ denotes the Euclidean norm of $x$. Let $A \in \mathbb{R}^{n\times n}$ be a square matrix. We denote by $A^\top$ the transpose of $A$, and by $A^{-1}$ the inverse of $A$. Finally, by $\overline{\text{co}}\,(Q)$ we denote the closure of the convex hull of set $Q \subset \mathbb{R}^n$.

## 2. Preliminaries

In this section, we report a number of definitions and properties concerning nonsmooth analysis, which are needed in the development. The reader is referred to [26, Ch. 2] for a thorough treatment.

### 2.1. Locally Lipschitz and regular functions

A function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be Lipschitz near $x \in \mathbb{R}^n$ if there exist $\ell, \epsilon > 0$, such that we have $|f(x_2) - f(x_1)| \le \ell\|x_2 - x_1\|$, for all $x_1, x_2 \in \mathbb{R}^n$ satisfying $\|x_1 - x\| < \epsilon$ and $\|x_2 - x\| < \epsilon$. If $f$ is Lipschitz near any point $x \in \mathbb{R}^n$, then $f$ is said to be *locally Lipschitz* in $\mathbb{R}^n$.

Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is locally Lipschitz in $\mathbb{R}^n$. Then, $f$ is differentiable for almost all (a.a.) $x \in \mathbb{R}^n$ (in the sense of Lebesgue measure). Moreover, for any $x \in \mathbb{R}^n$ we can define Clarke’s generalized gradient of $f$ at point $x$, as follows

$$\partial f(x) = \overline{\text{co}}\left\{\lim_{n\to\infty} \nabla f(x_n) : x_n \to x, x_n \notin N, x_n \notin \Omega\right\}$$

where $\Omega \subset \mathbb{R}^n$ is the set of points where $f$ is not differentiable, and $N \subset \mathbb{R}^n$ is an arbitrary set with measure zero. It can be proved that $\partial f(x) : \mathbb{R}^n \multimap \mathbb{R}^n$ is a set-valued map that associates to any $x \in \mathbb{R}^n$ a non-empty compact convex subset $\partial f(x) \subset \mathbb{R}^n$.

A function $f : \mathbb{R}^n \to \mathbb{R}$, which is locally Lipschitz near $x \in \mathbb{R}^n$, is said to be *regular* at $x$ if the following holds. For all directions $v \in \mathbb{R}^n$, there exists the usual one-sided directional derivative

$$f'(x, v) = \lim_{\rho\downarrow 0} \frac{f(x + \rho v) - f(x)}{\rho}$$

and we have $f'(x, v) = f^0(x, v)$, where

$$f^0(x, v) = \limsup_{\substack{y \to x \\ \rho \downarrow 0}} \frac{f(y + \rho v) - f(y)}{\rho}$$

is the generalized directional derivative of $f$ at $x$ in the direction $v$. The function $f$ is said to be regular in $\mathbb{R}^n$, if it is regular for any $x \in \mathbb{R}^n$.

Note that regular functions admit the directional derivative for all directions $v \in \mathbb{R}^n$, although the derivative may be different for different directions. If a function is continuously differentiable on $\mathbb{R}^n$, then it is regular on $\mathbb{R}^n$, but the converse is not true (e.g., $f(x) = |x| : \mathbb{R} \to \mathbb{R}$). A useful property is that a locally Lipschitz and convex function in $\mathbb{R}^n$ is also regular in $\mathbb{R}^n$ (see [26, Proposition 2.3.6]).

Finally, we observe that if $f : \mathbb{R}^n \to \mathbb{R}$ is locally Lipschitz in $\mathbb{R}^n$, then by virtue of [26, Theorem 2.5.1] the previous definition of Clarke's generalized gradient of $f$ at $x \in \mathbb{R}^n$ is equivalent to the following one

$$\partial f(x) = \{\zeta \in \mathbb{R}^n : f^0(x, v) \geq \langle \zeta, v \rangle \text{ for all } v \in \mathbb{R}^n\}.$$

Moreover, for any $v \in \mathbb{R}^n$ we have [26, Proposition 2.1.2]

$$f^0(x, v) = \max\{\langle \zeta, v \rangle : \zeta \in \partial f(x)\}.$$

### 2.2. *Hölder functions*

Here, we define a class of continuous functions which is of interest for the applications in this paper.

Given $\lambda \in (0, 1]$, $f : \mathbb{R}^n \to \mathbb{R}$ is said to be a $\lambda$-Hölder function near $x \in \mathbb{R}^n$ if there exist $\ell, \epsilon > 0$, such that we have

$$|f(x_2) - f(x_1)| \leq \ell \|x_2 - x_1\|^\lambda$$

for all $x_1, x_2 \in \mathbb{R}^n$ satisfying $\|x_1 - x\| < \epsilon$ and $\|x_2 - x\| < \epsilon$.

When $\lambda = 1$, the definition reduces to that of a function that is Lipschitz near $x$, however when $\lambda \in (0, 1)$ a $\lambda$-Hölder function near $x$ is continuous at $x$ but in general fails to be Lipschitz near $x$.

## 3. Generalized Lyapunov approach

The goal of this section is to develop a Lyapunov-like approach for studying global convergence of the solutions to differential equations with discontinuous right-hand side modeling a class of neural networks. We recall that, according to Filippov's theory, a solution to a differential equation with discontinuous right-hand side is an absolutely continuous function that satisfies a suitable differential inclusion associated to the differential equation [27].

There are two main ingredients in the approach here described: (1) a chain rule for computing the time derivative of a nondifferentiable Lyapunov function along the solutions, which is applicable in the nonsmooth setting considered in this paper; (2) a comparison principle for this time derivative, which gives sufficient conditions for ensuring the desired convergence properties.

To be specific, we consider (candidate) Lyapunov functions $V$ with the following properties.

**Assumption 1.** Function $V(x) : \mathbb{R}^n \to \mathbb{R}$ is:

(i) Regular in $\mathbb{R}^n$;
(ii) positive definite, i.e., we have $V(x) > 0$ for $x \neq 0$, and $V(0) = 0$;
(iii) radially unbounded, i.e., $V(x) \to +\infty$ as $\|x\| \to +\infty$.

Note that a Lyapunov function $V$ as in Assumption 1 is not necessarily differentiable. Suppose that $x(t) : [0, +\infty) \to \mathbb{R}^n$ is absolutely continuous on any compact interval of $[0, +\infty)$. The next property gives a chain rule for computing the time derivative of the composed function $V(x(t)) : [0, +\infty) \to \mathbb{R}$.

**Property 1** (*Chain Rule*). *Suppose that $V$ satisfies Assumption 1, and that $x(t) : [0, +\infty) \to \mathbb{R}^n$ is absolutely continuous on any compact interval of $[0, +\infty)$. Then, $x(t)$ and $V(x(t)) : [0, +\infty) \to \mathbb{R}$ are differentiable for a.a. $t \in [0, +\infty)$, and we have*

$$\frac{\mathrm{d}}{\mathrm{d}t} V(x(t)) = \langle \zeta, \dot{x}(t) \rangle \quad \forall \zeta \in \partial V(x(t)).$$

**Proof.** Since $V(x(t))$ is the composition of a locally Lipschitz function $V(x)$ and an absolutely continuous function $x(t)$, it follows that $V(x(t))$ is absolutely continuous on any compact interval of $[0, +\infty)$. Hence, $V(x(t))$ and $x(t)$ are differentiable for a.a. $t \geq 0$.

Consider an instant $t \in [0, +\infty)$ such that $x(t)$ and $V(x(t))$ are differentiable at $t$. Then, $x(t + h) = x(t) + \dot{x}(t)h + o(h)$, where $o(h)/h \to 0$ as $h \to 0$, and taking into account that $V$ is locally Lipschitz in $\mathbb{R}^n$ we obtain

$$\begin{aligned} &\left| \frac{V(x(t) + \dot{x}(t)h + o(h)) - V(x(t)) - V(x(t) + \dot{x}(t)h) + V(x(t))}{h} \right| \\ &\quad = \left| \frac{V(x(t) + \dot{x}(t)h + o(h)) - V(x(t) + \dot{x}(t)h)}{h} \right| \leq \ell \left| \frac{o(h)}{h} \right| \end{aligned}$$

for some $\ell > 0$. Then,

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} V(x(t)) &= \lim_{h \to 0} \frac{V(x(t + h)) - V(x(t))}{h} \\ &= \lim_{h \to 0} \frac{V(x(t) + \dot{x}(t)h) - V(x(t))}{h}. \end{aligned}$$

Because of the regularity of $V$, by letting $h \to 0+$ we obtain

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} V(x(t)) &= V'(x(t), \dot{x}(t)) \\ &= V^0(x(t), \dot{x}(t)) \\ &= \max\{\langle \zeta, \dot{x}(t) \rangle, \zeta \in \partial V(x(t))\}. \end{aligned}$$

Similarly, by letting $h \to 0-$ and taking into account again that $V$ is regular we obtain

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} V(x(t)) &= -V'(x(t), -\dot{x}(t)) \\ &= -V^0(x(t), -\dot{x}(t)) \\ &= V_0(x(t), \dot{x}(t)) \end{aligned}$$

where

$$\begin{aligned} V_0(x(t), \dot{x}(t)) &= \liminf_{\substack{y \to x(t) \\ h \downarrow 0}} \frac{V(y + h\dot{x}(t)) - V(y)}{h} \\ &= \min\{\langle \zeta, \dot{x}(t) \rangle, \zeta \in \partial V(x(t))\}. \end{aligned}$$

Thus, we have shown that

$$\frac{\mathrm{d}}{\mathrm{d}t}V(x(t)) = \langle \zeta, \dot{x}(t)\rangle \quad \forall \zeta \in \partial V(x(t)). \quad \square$$

The property that follows gives a general condition for convergence.

**Property 2** (*Convergence*). *Suppose that $V$ satisfies Assumption 1, and that $x(t) : [0, +\infty) \to \mathbb{R}^n$ is absolutely continuous on any compact interval of $[0, +\infty)$. Let $v(t) = V(x(t))$, and suppose that there exists a continuous function $\Upsilon : (0, +\infty) \to \mathbb{R}$, with $\Upsilon(\sigma) > 0$ for $\sigma \in (0, +\infty)$, such that we have*

$$\dot{v}(t) \leq -\Upsilon(v(t)) \tag{1}$$

*for all $t \geq 0$ such that $v(t) > 0$ and $v(t)$ is differentiable at $t$. Then,*

$$\lim_{t\to+\infty} v(t) = 0$$

*and*

$$\lim_{t\to+\infty} x(t) = 0.$$

**Proof.** We begin by observing that if $v(t_0) = 0$ for some $t_0 \geq 0$, then we have $v(t) = 0$ for all $t > t_0$. Otherwise, there exists $T > t_0$ such that $v(T) > 0$. Let

$$t_s = \sup\{t \in [t_0, T] : v(t) = 0\}.$$

Hence, we have $t_s < T$, $v(t_s) = 0$ and $v(t) > 0$ for all $t \in (t_s, T]$. From the absolute continuity of $v(t)$, it follows that

$$0 < v(T) = v(t_s) + \int_{t_s}^{T} \dot{v}(\tau)\mathrm{d}\tau = \int_{t_s}^{T} \dot{v}(\tau)\mathrm{d}\tau$$

which contradicts the fact that $\dot{v}(t) \leq -\Upsilon(v(t)) < 0$ for a.a. $t \in (t_s, T]$.

Then, consider the case where $v(t) > 0$ for all $t \geq 0$. Being $\dot{v}(t) \leq 0$ for a.a. $t \geq 0$, it follows that $v(t)$ is non-increasing for $t \geq 0$, hence there exists $\varepsilon \geq 0$ such that $v(t) \to \varepsilon$ as $t \to +\infty$, and $v(t) \geq \varepsilon$ for $t \geq 0$. Suppose, for contradiction, that $\varepsilon > 0$ and let $M = \min_{\sigma \in [\varepsilon, v(0)]} \{\Upsilon(\sigma)\}$. Due to the assumptions on $\Upsilon$, we have $M > 0$. Therefore, for a.a. $t > 0$ we obtain $\dot{v}(t) \leq -\Upsilon(v(t)) \leq -M$, hence $v(t) \leq v(0) - Mt < \varepsilon$ for sufficiently large $t$, which is a contradiction. Then, we conclude that $\varepsilon = 0$, i.e., $v(t) \to 0$ as $t \to +\infty$. Since $V$ is positive definite and radially unbounded, this implies that $x(t) \to 0$ as $t \to +\infty$. $\square$

The next properties address exponential convergence and convergence in finite time.

**Property 3** (*Exponential Convergence*). *Suppose that the assumptions of Property 2 are satisfied, and that in particular $\Upsilon(v) = av$, for all $v \in (0, +\infty)$, where $a > 0$. Then, we have*

$$0 \leq v(t) \leq v(0)\mathrm{e}^{-at}, \quad t \geq 0 \tag{2}$$

*i.e., $v(t)$ converges exponentially to 0 with convergence rate $a$.*

*If, in addition, there exist $c > 0$ and $\vartheta > 0$ such that*

$$0 \leq c\,\|x\|^{\vartheta} \leq V(x) \tag{3}$$

*for all $x \in \mathbb{R}^n$, then we have*

$$\|x(t)\| \leq \left(\frac{v(0)}{c}\right)^{\frac{1}{\vartheta}} \mathrm{e}^{-\frac{a}{\vartheta}t}, \quad t \geq 0 \tag{4}$$

*i.e., $x(t)$ converges exponentially to $x = 0$ with convergence rate $a/\vartheta$.*

**Proof.** If $v(t) = 0$ for some $t \geq 0$, then the result of the property is obvious. Consider an instant $t$ such that $v(t) > 0$. By recalling that $v(t)$ is non-increasing for $t \geq 0$, and accounting for the assumptions on $\Upsilon$ in Property 2, we have

$$\begin{aligned} -\int_0^t \frac{\dot{v}(\tau)}{\Upsilon(v(\tau))}\mathrm{d}\tau &= \int_{v(t)}^{v(0)} \frac{1}{\Upsilon(\sigma)}\mathrm{d}\sigma \\ &= \int_{v(t)}^{v(0)} \frac{\mathrm{d}\sigma}{a\sigma} = \frac{1}{a}\ln\frac{v(0)}{v(t)} \end{aligned}$$

and, due to (1),

$$-\int_0^t \frac{\dot{v}(\tau)}{\Upsilon(v(\tau))}\mathrm{d}\tau \geq t.$$

Hence,

$$\frac{1}{a}\ln\frac{v(0)}{v(t)} \geq t$$

which yields $v(t) \leq v(0)\mathrm{e}^{-at}$.

If condition (3) is satisfied, then for all $t \geq 0$ we have

$$c\,\|x(t)\|^{\vartheta} \leq v(0)\mathrm{e}^{-at}$$

from which the result on exponential convergence of $x(t)$ to 0 immediately follows. $\square$

**Property 4** (*Convergence in Finite Time*). *Suppose that the assumptions of Property 2 are satisfied, and that $\Upsilon$ satisfies the condition*

$$\int_0^{v(0)} \frac{1}{\Upsilon(\sigma)}\mathrm{d}\sigma = t_\phi < +\infty. \tag{5}$$

*Then, we have*

$$v(t) = 0, \quad t \geq t_\phi$$

*and*

$$x(t) = 0, \quad t \geq t_\phi$$

*i.e., $v(t)$ converges to 0, and $x(t)$ converges to $x = 0$, in finite time $t_\phi$.*

*Next, $t_\phi$ is explicitly evaluated for two particular functions $\Upsilon$ satisfying condition (5).*

(a) *If $\Upsilon(\sigma) = k > 0$, for all $\sigma \in (0, +\infty)$, then*

$$t_\phi = \frac{v(0)}{k}.$$

(b) *If* $\Upsilon(\sigma) = Q\sigma^{\mu}$, *for all* $\sigma \in (0, +\infty)$, *where* $\mu \in (0, 1)$ *and* $Q > 0$, *then*

$$t_{\phi} = \frac{v^{1-\mu}(0)}{Q(1-\mu)}.$$

**Proof.** Suppose, for contradiction, that $v(t)$ does not converge to zero in finite time. In such a case, we would have $v(t) > 0$ for all $t > 0$. Let $t_g > t_{\phi}$, where $t_{\phi}$ is defined in (5). Then, accounting for (1) we obtain

$$t_{\phi} = \int_0^{v(0)} \frac{1}{\Upsilon(\sigma)} \mathrm{d}\sigma \geq \int_{v(t_g)}^{v(0)} \frac{1}{\Upsilon(\sigma)} \mathrm{d}\sigma = -\int_0^{t_g} \frac{\dot{v}(\tau)}{\Upsilon(v(\tau))} \mathrm{d}\tau \geq t_g$$

i.e., a contradiction.

Then, $v(t)$ reaches 0 at an instant $t_r \leq t_{\phi}$, and by an argument as in the proof of Property 2, it follows that $v(t) = 0$, and hence $x(t) = 0$, for $t \geq t_r$.

The results given in (a) and (b) are simply obtained by evaluating $t_{\phi}$ in (5) for the considered functions $\Upsilon$. □

Property 3 generalizes to nonsmooth dynamical systems an analogous comparison result which has been widely employed to prove exponential convergence in the context of smooth neural networks defined by Lipschitz-continuous vector fields [6,7,9,28]. Condition (5) in Property 4 represents a general Nagumo-type condition for convergence in finite time. This condition includes as special cases previous conditions in the literature for ensuring convergence in finite time of nonsmooth dynamical systems. For example, the result on convergence in finite time given in [29, Theorem 2] is based on an assumption as in the special case of condition (5) where $\Upsilon(\sigma) = k > 0$, see point (a) of Property 4. A condition as in (a) of the same property has been already employed in the context of neural networks to prove convergence in finite time of discontinuous dynamical systems for solving linear programming problems [16,18].[1] A further application of condition (a) of Property 4 concerns the result on convergence in finite time given in [1, Theorem 4], for the state and output solutions of a class of neural networks possessing neuron activations with jump discontinuities. To the authors' knowledge, a condition as in point (b) of Property 4, where $\Upsilon(\sigma) = Q\sigma^{\mu}$, has not been considered up to now to address convergence in finite time of neural networks.

## 4. Neural network model

We consider additive neural networks whose dynamics is described by the system of differential equations with *discontinuous* right-hand side

$$\dot{x} = Bx + Tg(x) + I \tag{6}$$

where $x \in \mathbb{R}^n$ is the vector of neuron state variables, $B = \mathrm{diag}(-b_1, \ldots, -b_n) \in \mathbb{R}^{n\times n}$ is a diagonal matrix with diagonal entries $-b_i < 0$, $i = 1, \ldots, n$, modeling the neuron self-inhibitions, and $I \in \mathbb{R}^n$ is the vector of neuron biasing inputs. Moreover, the entries of matrix $T \in \mathbb{R}^{n\times n}$ are the neuron interconnections, while the components $g_i$ of the diagonal mapping $g(x) = (g_1(x_1), \ldots, g_n(x_n))^{\top} : \mathbb{R}^n \to \mathbb{R}^n$ are the neuron activations.

As in [1], the neuron activations are modeled with the next class of discontinuous functions.

**Definition 1.** We say that $g \in \mathbb{D}$ if and only if, for $i = 1, \ldots, n$, $g_i : \mathbb{R} \to \mathbb{R}$ is a bounded non-decreasing piecewise continuous function. The last property means that $g_i$ is continuous in $\mathbb{R}$ except for countably many points of discontinuity where there exist finite right and left limits, furthermore $g_i$ has a finite number of discontinuities in any compact interval of $\mathbb{R}$.

We consider for (6) solutions in the sense of Filippov [27]. Filippov's solutions are known to be uniform approximations of solutions of actual systems that possess nonlinearities with very high slope [30,31]. Due to this reason, Filippov's theory has become a standard tool in several applications to control problems and problems in mechanics involving the presence of nonsmooth friction.

Let $g \in \mathbb{D}$. A function $x(t)$, $t \in [t_a, t_b]$, is a solution of (6) in the sense of Filippov, with initial condition $x(t_a) = x_0 \in \mathbb{R}^n$, if the following hold [27]: $x(t)$ is absolutely continuous on $[t_a, t_b]$, $x(t_a) = x_0$, and for a.a. $t \in [t_a, t_b]$, $x(t)$ satisfies the differential inclusion

$$\dot{x}(t) \in Bx(t) + T\,\overline{\mathrm{co}}\,[g(x(t))] + I \tag{7}$$

where we have let $\overline{\mathrm{co}}\,[g(x)] = (\overline{\mathrm{co}}\,[g_1(x_1)], \ldots, \overline{\mathrm{co}}\,[g_n(x_n)])^{\top} : \mathbb{R}^n \multimap \mathbb{R}^n$, and

$$\overline{\mathrm{co}}\,[g_i(x_i)] = [g_i(x_i^-), g_i(x_i^+)] \tag{8}$$

for $i = 1, \ldots, n$. Note that $\overline{\mathrm{co}}\,[g_i(x_i)]$ is an interval with non-empty interior when $g_i$ is discontinuous at $x_i$, while $\overline{\mathrm{co}}\,[g_i(x_i)] = \{g(x_i)\}$ is a singleton when $g_i$ is continuous at $x_i$.

It is pointed out that the neuron activations $g_i$ are not necessarily defined at their points of discontinuity. A solution $x(t)$ of (6) in the sense of Filippov corresponds to a solution of the differential inclusion (7) where, at each point of discontinuity, $g_i$ is defined as the interval $\overline{\mathrm{co}}\,[g_i(x_i)] = [g_i(x_i^-), g_i(x_i^+)]$. From a geometrical viewpoint, this means that the inclusion (7) is simply obtained by filling in the jump discontinuities of $g_i$.

Let $x(t)$, $t \in [t_a, t_b]$, be a solution of (6), and suppose $\det T \neq 0$. Then, for a.a. $t \in [t_a, t_b]$ we have

$$\dot{x}(t) = Bx(t) + T\gamma(t) + I \tag{9}$$

[1] It is worth noticing that in the quoted papers [16,18,29], a condition as in point (a) of Property 4 was used to prove convergence in finite time to a set of points where for some Lyapunov-like function $V$ we have $V(x) \leq 0$. The result in Property 4 can be easily extended to encompass this more general situation, by simply removing the hypothesis that $V$ is positive definite (cf. (ii) of Assumption 1). We have preferred to state Property 4 in the more restrictive case where $V$ is positive definite, since in the paper we are interested in the application to prove global convergence in finite time toward a unique equilibrium point for the solutions of a class of neural networks.

where

$$\gamma(t) = T^{-1}(\dot{x}(t) - Bx(t) - I) \in \overline{\text{co}}\,[g(x(t))] \quad (10)$$

is the *output solution* of (6) corresponding to $x(t)$. It can be easily verified that $\gamma(t)$ is a bounded measurable function, which is uniquely defined by the *state* solution $x(t)$ for a.a. $t \in [t_a, t_b]$.

An *equilibrium point* (EP) $\xi \in \mathbb{R}^n$ of (6) is a stationary solution $x(t) = \xi$, $t \geq 0$, of (6). Clearly, $\xi \in \mathbb{R}^n$ is an EP of (6) if and only if $\xi$ satisfies the following algebraic inclusion

$$0 \in B\xi + T\overline{\text{co}}\,[g(\xi)] + I.$$

Let $\xi$ be an EP of (N). Then, from (10) it turns out that

$$\eta = T^{-1}(-B\xi - I) \in \overline{\text{co}}\,[g(\xi)] \quad (11)$$

is the *output equilibrium point* (OEP) of (N) corresponding to $\xi$.

From previous results [1], we have that if $g \in \mathbb{D}$ then for any $x_0 \in \mathbb{R}^n$ there is at least a solution $x(t)$ of (6) with initial condition $x(0) = x_0$, which is defined and bounded for $t \geq 0$. Moreover, there exists at least an EP $\xi \in \mathbb{R}^n$ of (6).

The goal of this paper is to demonstrate the applicability of the generalized Lyapunov approach in Section 3, by obtaining stronger global convergence results for the neural networks (6), with respect to those previously established in [1]. More specifically, Property 3 is used to prove a new result on *global exponential convergence* of the state solutions of (6) (Section 5), while Property 4 is exploited to obtain new results on *global convergence in finite time* for the state and output solutions of (6) (Section 6).

## 5. Global exponential convergence

In this section, we address global exponential convergence toward a unique EP of the state solutions of (6).

Let $\xi$ be an EP of (6), with corresponding OEP $\eta$. We find it useful to consider the change of variables $z = x - \xi$, which transforms (6) into the differential equation

$$\dot{z} = Bz + TG(z) \quad (12)$$

where $G(z) = g(z+\xi) - \eta$. If $g \in \mathbb{D}$, then we also have $G \in \mathbb{D}$. Note that (12) has an EP, and a corresponding OEP, which are both located at the origin. If $z(t)$, $t \geq 0$, is a solution of (12), then we denote by

$$\gamma^{\text{o}}(t) = T^{-1}(\dot{z}(t) - Bz(t)) \in \overline{\text{co}}\,[G(z(t))] \quad (13)$$

the output solution of (12) corresponding to $z(t)$, which is defined for a.a. $t \geq 0$.

**Definition 2** (*[24]*). We say that matrix $A \in \mathbb{R}^{n\times n}$ is Lyapunov Diagonally Stable (LDS), if there exists a positive definite diagonal matrix $\alpha = \text{diag}(\alpha_1, \ldots, \alpha_n)$, such that $(1/2)(\alpha A + A^\top \alpha)$ is positive definite.

Suppose that $-T \in$ LDS and, as in [1] consider for (12) the (candidate) Lyapunov function

$$V(z) = \sum_{i=1}^{n} \frac{1}{b_i} z_i^2 + 2c \sum_{i=1}^{n} \alpha_i \int_0^{z_i} G_i(\rho)\,\mathrm{d}\rho \quad (14)$$

where $c > 0$ is a constant, and $\alpha = \text{diag}(\alpha_1, \ldots, \alpha_n)$ is a positive definite diagonal matrix such that $(1/2)(\alpha(-T) + (-T)^\top \alpha)$ is positive definite. Note that $V$ is a locally Lipschitz and convex function on $\mathbb{R}^n$, hence it is also regular in $\mathbb{R}^n$ and it satisfies Assumption 1. Also note that, due to the jump discontinuities of $G_i$, $V$ is not differentiable.

The main result in this section is as follows.

**Theorem 1.** *Suppose that $g \in \mathbb{D}$ and that $-T \in$ LDS. Let $z(t)$, $t \geq 0$, be any solution of (12), and $v(t) = V(z(t))$, $t \geq 0$. Then, we have*

$$\dot{v}(t) \leq -b_{\text{m}} v(t), \quad \textit{for a.a. } t \geq 0$$

*where $b_{\text{m}} = \min_{i=1,\ldots,m}\{b_i\} > 0$, and hence*

$$0 \leq v(t) \leq v(0)\,\mathrm{e}^{-b_{\text{m}} t}, \quad t \geq 0$$

*i.e., $v(t)$ converges exponentially to 0 with convergence rate $b_{\text{m}}$.*

*Furthermore, we have*

$$\|z(t)\| \leq \sqrt{b_{\text{M}} v(0)}\,\mathrm{e}^{-\frac{b_{\text{m}}}{2}t}, \quad t \geq 0 \quad (15)$$

*where $b_{\text{M}} = \max_{i=1,\ldots,m}\{b_i\} > 0$, i.e., $z(t)$ is exponentially convergent to 0 with convergence rate $b_{\text{m}}/2$.*

Before giving the proof, we note that Theorem 1 improves the result in [1, Theorem 2], where under the same assumptions only convergence of the state solutions was proved, moreover no estimate of convergence rate was obtained. Theorem 1 is in agreement with previous results for neural networks possessing Lipschitz continuous neuron activations, where global exponential convergence with the same convergence rate was established [6,7,9,28].

**Proof of Theorem 1.** We start by observing that, since each $G_i$ is monotone non-decreasing and $0 \in \overline{\text{co}}\,[G_i(0)]$, it easily follows that for any $z_i \in \mathbb{R}$ we have

$$0 \leq \int_0^{z_i} G_i(\rho)\mathrm{d}\rho \leq z_i \zeta_i \quad \forall \zeta_i \in \overline{\text{co}}\,[G_i(z_i)]. \quad (16)$$

Let $z(t)$, $t \geq 0$, be any solution of (12), hence $z(t)$ is absolutely continuous on any compact interval of $[0, +\infty)$. Since $V$ is regular at any $z(t)$, it is possible to apply the chain rule of Property 1 in order to obtain that for a.a. $t \geq 0$ we have

$$\dot{v}(t) = \langle \zeta, \dot{z}(t) \rangle \quad \forall \zeta \in \partial V(z(t)). \quad (17)$$

Below, we extend the argument used in the proof of [1, Theorem 2], in order to prove by means of Property 3 that $z(t)$ is exponentially convergent to 0.

By evaluating the scalar product in (17), it has been proved in [1, App. IV] that there exists $\lambda > 0$ such that for a.a. $t \geq 0$ we have

$$\begin{aligned}\dot{v}(t) &\leq -\|z(t)\|^2 - \|B^{-1}\dot{z}(t)\|^2 - \lambda\|\gamma^{\text{o}}(t)\|^2 \\ &\quad + 2cz^\top(t)\alpha B\gamma^{\text{o}}(t) \\ &\leq -\|z(t)\|^2 + 2cz^\top(t)\alpha B\gamma^{\text{o}}(t).\end{aligned} \quad (18)$$

Now, note that from (16) it follows that

$$2cz^\top(t)\alpha B\gamma^{\mathrm{o}}(t) = -2c\sum_{i=1}^{n}\alpha_i b_i z_i(t)\gamma_i^{\mathrm{o}}(t)$$
$$\leq -2c\sum_{i=1}^{n}\alpha_i b_i\int_0^{z_i(t)} G_i(\rho)\mathrm{d}\rho \leq 0. \qquad (19)$$

Hence, we obtain

$$\begin{aligned}\dot{v}(t) &\leq -\sum_{i=1}^{n} z_i^2(t) - 2c\sum_{i=1}^{n}\alpha_i b_i z_i(t)\gamma_i^{\mathrm{o}}(t)\\ &\leq -b_{\mathrm{m}}\left(\sum_{i=1}^{n}\frac{1}{b_i}z_i^2(t) + 2c\sum_{i=1}^{n}\alpha_i\int_0^{z_i(t)} G_i(\rho)\mathrm{d}\rho\right)\\ &= -b_{\mathrm{m}}v(t)\end{aligned}$$

for a.a. $t \geq 0$. By applying Property 3, we conclude that $v(t) \leq v(0)\mathrm{e}^{-b_{\mathrm{m}}t}$ for all $t \geq 0$.

Since $\alpha_i > 0$, and the graph of $G_i$ is contained in the first and third quadrant, we have

$$\alpha_i\int_0^{z_i} G_i(\rho)\mathrm{d}\rho \geq 0$$

for all $z_i \in R$. Therefore,

$$\begin{aligned}V(z) &= \sum_{i=1}^{n}\frac{1}{b_i}z_i^2 + 2c\sum_{i}^{n}\alpha_i\int_0^{z_i} G_i(\rho)\mathrm{d}\rho\\ &\geq \sum_{i=1}^{n}\frac{1}{b_i}z_i^2 \geq \frac{1}{b_{\mathrm{M}}}\|z\|^2.\end{aligned}$$

The result on exponential convergence of $x(t)$ to 0 thus follows from (4) of Property 3. □

## 6. Global convergence in finite time

This section is divided in two parts. First, we establish a theorem on global convergence in finite time for LDS interconnection matrices, which is valid for a class of non-Lipschitz neuron activations (Section 6.1). Then, we prove a second result on convergence in finite time (Section 6.2), which requires the stronger assumption that the interconnection matrix is an $H$-matrix (see Definition 4), but is applicable to a larger class of non-Lipschitz neuron activations.

### *6.1. LDS-matrices*

Suppose that $-T \in$ LDS, and consider for (12) the Lyapunov function $V$ defined in (14). Let

$$\theta_D = \{i \in \{1,\ldots,n\} : G_i \text{ is discontinuous at } z_i = 0\}$$

and $\theta_C = \{1,\ldots,n\}\setminus\theta_D$.

In the next theorem we establish a result on global convergence in finite time of the state and output solutions of (12).

**Theorem 2.** *Suppose that $g \in \mathbb{D}$, and that $-T \in$ LDS. Moreover, suppose that for any $i \in \theta_D$ we have $G_i(0^+) > 0$ and $G_i(0^-) < 0$, while for any $i \in \theta_C$ there exist $\delta_i, K_i, K_i^+ > 0$, $\mu_i \in (0,1)$ and $\mu_i^+ \in [0,1)$, such that*

$$K_i|\rho|^{\mu_i} \leq |G_i(\rho)| \leq K_i^+|\rho|^{\mu_i^+}, \quad |\rho| < \delta_i. \qquad (20)$$

*Furthermore, suppose that*

$$\mu_{\mathrm{M}} = \max_{i\in\theta_C}\left\{\frac{2\mu_i}{1+\mu_i^+}\right\} < 1. \qquad (21)$$

*Let $z(t)$, $t \geq 0$, be any solution of (12), and $v(t) = V(z(t))$, $t \geq 0$. Moreover, let $\gamma^{\mathrm{o}}(t)$, for a.a. $t \geq 0$, be the output solution of (12) corresponding to $z(t)$. Then, there exists $t_\delta > 0$ such that we have*

$$\dot{v}(t) \leq -Qv^{\mu}(t), \quad \textit{for a.a. } t > t_\delta$$

*where $\mu \in [\mu_{\mathrm{M}}, 1)$ and $Q > 0$. As a consequence, we have*

$$\begin{aligned}v(t) &= 0, \quad \forall t \geq t_\phi\\ z(t) &= 0, \quad \forall t \geq t_\phi\\ \gamma^{\mathrm{o}}(t) &= 0, \quad \textit{for a.a. } t \geq t_\phi\end{aligned}$$

*where*

$$t_\phi = t_\delta + \frac{v^{1-\mu}(t_\delta)}{Q(1-\mu)}$$

*i.e., $v(t)$, $z(t)$, and $\gamma^{\mathrm{o}}(t)$ converge to zero in finite time $t_\phi$.*

Prior to the proof, we give some remarks illustrating the result in the theorem.

**Remarks.** 1. Theorem 2 is an extension of the result on global convergence in finite time given in [1, Theorem 4]. Indeed, Theorem 2 can be applied to the case where the neuron activations $G_i$ are either discontinuous at 0, or they are modeled by non-Lipschitz functions in a neighborhood of 0. Instead, [1, Theorem 4] requires that all neuron activations be discontinuous at 0. Note that the left inequality in (20) means that $G_i$ grows at least as a $\mu_i$-Hölder function, with $\mu_i \in (0,1)$, in a neighborhood of 0. Finally, it is observed that the proof of Theorem 2 relies on condition (b) in Property 4, while that of [1, Theorem 4] was based on a condition analogous to that in (a) of Property 4. The proof of Theorem 2 also yields a simple quantitative estimate of the finite convergence time $t_\phi$, see (27).

2. It can be easily checked that the following classes of continuous non-Lipschitz functions $G_i$ satisfy assumptions (20) and (21) of Theorem 3:
   (1) For any $i \in \theta_C$, (20) is satisfied by $G_i$ with $\mu_i < 1/2$;
   (2) for any $i \in \theta_C$, $G_i$ is a $\mu_i$-Holder function defined as $G_i(\rho) = k_i\,\mathrm{sgn}(\rho)|\rho|^{\mu_i}$ where $\mu_i \in (0,1)$ and $k_i > 0$.

**Proof of Theorem 2.** We need the following additional notations. Since $-T \in$ LDS, there exists $\alpha = \mathrm{diag}(\alpha_1,\ldots,\alpha_n)$, where $\alpha_i > 0$ for $i = 1,\ldots,n$, such that $(1/2)(\alpha(-T) + (-T)^\top\alpha)$ is positive definite. Let

$$\alpha_{\mathrm{M}} = \max_{i\in\theta_C}\{\alpha_i\} > 0.$$

We also define

$$K_{\mathrm{m}} = \min_{i\in\theta_C}\{K_i\} > 0, \quad \mu_{\mathrm{m}}^{+} = \min_{i\in\theta_C}\{\mu_i^{+}\} \geq 0.$$

For any $i \in \theta_D$, we let

$$m_i = \min\{-G_i(0^-), G_i(0^+)\} > 0.$$

Since $G_i$ has a finite number of discontinuities in any compact interval of $\mathbb{R}$, for any $i \in \theta_D$ there exists $\delta_i \in (0,1]$ such that $G_i$ is a continuous function in $[-\delta_i, 0) \cup (0, \delta_i]$. For any $i \in \theta_D$ we define $K_i^{+} = \sup_{\rho\in[-\delta_i,\delta_i]}\{|G_i(\rho)|\} = \max\{-G_i(-\delta_i), G_i(\delta_i)\} > 0$, and we let

$$K_{\mathrm{M}}^{+} = \max\{K_1^{+}, \ldots, K_n^{+}\} > 0.$$

Finally, we let

$$\delta = \min\left\{1, \min\{\delta_1, \ldots, \delta_n\}, \min_{i\in\theta_D}\left\{\left(\frac{m_i}{K_{\mathrm{m}}}\right)^{\frac{2}{\mu_{\mathrm{M}}}}\right\}\right\} > 0.$$

We are now in a position to address the theorem proof. It is seen from (15) that for

$$t > t_\delta = \frac{2}{b_{\mathrm{m}}}\ln\left(\frac{\sqrt{b_{\mathrm{M}}v(0)}}{\delta}\right) \tag{22}$$

we have $z(t) \in [-\delta, \delta]^n$. Since by definition $\delta \in (0,1]$, then for $t > t_\delta$ we also have $|z_i(t)|^{p_2} \leq |z_i(t)|^{p_1} \leq 1$, for any $p_1, p_2 > 0$ such that $p_1 \leq p_2$.

For $t > t_\delta$, let

$$P(t) = \{i \in 1, \ldots, n : z_i(t) = 0\}, \quad \theta_C^{\sharp}(t) = \theta_C \setminus P(t),$$
$$\theta_D^{\sharp}(t) = \theta_D \setminus P(t).$$

For any $i \in P(t)$, we have $\left|\gamma_i^{\mathrm{o}}(t)\right| \geq 0 = z_i(t)$, while for any $i \in \theta_C^{\sharp}(t)$ we have $\left|\gamma_i^{\mathrm{o}}(t)\right| = |G_i(z_i(t))| \geq K_i|z_i(t)|^{\mu_i}$. Moreover, for any $i \in \theta_D^{\sharp}(t)$ we obtain $m_i \leq \left|\gamma_i^{\mathrm{o}}(t)\right| = |G_i(z_i(t))| \leq K_i^{+}$.

Suppose that $v(t)$ is differentiable at $t > t_\delta$. Then, from (18) and (19) we have

$$\begin{aligned}
\dot{v}(t) &\leq -\lambda\left\|\gamma^{\mathrm{o}}(t)\right\|^2 = -\lambda\sum_{i=1}^{n}\left|\gamma_i^{\mathrm{o}}(t)\right|^2 \\
&\leq -\lambda\left(\sum_{i\in\theta_C^{\sharp}(t)}K_i^2|z_i(t)|^{2\mu_i} + \sum_{i\in\theta_D^{\sharp}(t)}m_i^2\right) \\
&\leq -\lambda\left(K_{\mathrm{m}}^2\sum_{i\in\theta_C^{\sharp}(t)}|z_i(t)|^{2\mu_i} + \sum_{i\in\theta_D^{\sharp}(t)}m_i^2\right) \\
&= -\lambda K_{\mathrm{m}}^2\left(\sum_{i\in\theta_C^{\sharp}(t)}|z_i(t)|^{2\mu_i} + \sum_{i\in\theta_D^{\sharp}(t)}\frac{m_i^2}{K_{\mathrm{m}}^2}\right).
\end{aligned} \tag{23}$$

Let $\mu \in (0,1)$. Since $z_i(t) \leq \delta_i$, we obtain

$$\begin{aligned}
v^{\mu}(t) &= \left(-z^{\top}(t)B^{-1}z(t) + 2c\sum_{i=1}^{n}\alpha_i\int_0^{z_i(t)}G_i(\rho)\,\mathrm{d}\rho\right)^{\mu} \\
&= \left(\sum_{i\notin P(t)}\frac{|z_i(t)|^2}{b_i} + 2c\sum_{i\notin P(t)}\alpha_i\int_0^{z_i(t)}G_i(\rho)\,\mathrm{d}\rho\right)^{\mu} \\
&\leq \left(\sum_{i\notin P(t)}\frac{|z_i(t)|^2}{b_i} + 2c\sum_{i\in\theta_C^{\sharp}(t)}\alpha_i\int_0^{|z_i(t)|}K_i^{+}\rho^{\mu_i^{+}}\,\mathrm{d}\rho\right. \\
&\quad\left. + 2c\sum_{i\in\theta_D^{\sharp}(t)}\alpha_i\int_0^{|z_i(t)|}K_i^{+}\,\mathrm{d}\rho\right)^{\mu} \\
&= \left(\sum_{i\notin P(t)}\frac{|z_i(t)|^2}{b_i} + 2c\sum_{i\in\theta_C^{\sharp}(t)}\alpha_iK_i^{+}\frac{|z_i(t)|^{1+\mu_i^{+}}}{1+\mu_i^{+}}\right. \\
&\quad\left. + 2c\sum_{i\in\theta_D^{\sharp}(t)}\alpha_iK_i^{+}|z_i(t)|\right)^{\mu}.
\end{aligned}$$

Since $(a+b)^{\mu} \leq a^{\mu} + b^{\mu}$ for $a, b \geq 0$ and $\mu \in (0,1)$, we have

$$\begin{aligned}
v^{\mu}(t) &\leq \sum_{i\notin P(t)}\frac{|z_i(t)|^{2\mu}}{b_{\mathrm{m}}^{\mu}} \\
&\quad + (2c)^{\mu}\left(\sum_{i\in\theta_C^{\sharp}(t)}\left(\frac{\alpha_iK_i^{+}}{1+\mu_i^{+}}\right)^{\mu}|z_i(t)|^{\mu(1+\mu_i^{+})}\right. \\
&\quad\left. + \sum_{i\in\theta_D^{\sharp}(t)}(\alpha_iK_i^{+})^{\mu}|z_i(t)|^{\mu}\right).
\end{aligned}$$

Recall that $\mu_i^{+} < 1$, hence $2\mu > \mu\left(1+\mu_i^{+}\right)$ and being $|z_i(t)| \leq \delta \leq 1$, it follows that $|z_i(t)|^{2\mu} \leq |z_i(t)|^{\mu(1+\mu_i^{+})}$ and $|z_i(t)|^{2\mu} \leq |z_i(t)|^{\mu}$. Therefore,

$$\begin{aligned}
v^{\mu}(t) &\leq \sum_{i\in\theta_C^{\sharp}(t)}\frac{|z_i(t)|^{\mu(1+\mu_i^{+})}}{b_{\mathrm{m}}^{\mu}} + \sum_{i\in\theta_D^{\sharp}(t)}\frac{|z_i(t)|^{\mu}}{b_{\mathrm{m}}^{\mu}} \\
&\quad + \left(\frac{2c\alpha_{\mathrm{M}}K_{\mathrm{M}}^{+}}{1+\mu_{\mathrm{m}}^{+}}\right)^{\mu}\sum_{i\in\theta_C^{\sharp}(t)}|z_i(t)|^{\mu(1+\mu_i^{+})} \\
&\quad + (2c\alpha_{\mathrm{M}}K_{\mathrm{M}}^{+})^{\mu}\sum_{i\in\theta_D^{\sharp}(t)}|z_i(t)|^{\mu} \\
&\leq \left(\frac{1}{b_{\mathrm{m}}^{\mu}} + \left(\frac{2c\alpha_{\mathrm{M}}K_{\mathrm{M}}^{+}}{1+\mu_{\mathrm{m}}^{+}}\right)^{\mu}\right)\sum_{i\in\theta_C^{\sharp}(t)}|z_i(t)|^{\mu(1+\mu_i^{+})} \\
&\quad + \left(\frac{1}{b_{\mathrm{m}}^{\mu}} + \left(2c\alpha_{\mathrm{M}}K_{\mathrm{M}}^{+}\right)^{\mu}\right)\sum_{i\in\theta_D^{\sharp}(t)}|z_i(t)|^{\mu}.
\end{aligned}$$

If $\mu \in [\mu_{\mathrm{M}}, 1)$, then from (21) it follows that $\mu(1+\mu_i^{+}) \geq 2\mu_i$ and hence

$$v^{\mu}(t) \leq \left(\frac{1}{b_{\mathrm{m}}^{\mu}} + \left(\frac{2c\alpha_{\mathrm{M}}K_{\mathrm{M}}^{+}}{1+\mu_{\mathrm{m}}^{+}}\right)^{\mu}\right)\sum_{i\in\theta_C^{\sharp}(t)}|z_i(t)|^{2\mu_i}$$

$$
\begin{aligned}
&+\left(\frac{1}{b_{\mathrm{m}}^{\mu}}+\left(2c\alpha_{\mathrm{M}}K_{\mathrm{M}}^{+}\right)^{\mu}\right)\sum_{i\,\in\,\theta_D^{\sharp}(t)}|z_i(t)|^{\mu}\\
&\le\left(\frac{1}{b_{\mathrm{m}}^{\mu}}+\left(2c\alpha_{\mathrm{M}}K_{\mathrm{M}}^{+}\right)^{\mu}\right)\\
&\quad\times\left(\sum_{i\,\in\,\theta_C^{\sharp}(t)}|z_i(t)|^{2\mu_i}+\sum_{i\,\in\,\theta_D^{\sharp}(t)}|z_i(t)|^{\mu}\right)
\end{aligned}\tag{24}
$$

where we have taken into account that $(2c\alpha_{\mathrm{M}}K_{\mathrm{M}}^{+})/(1+\mu_{\mathrm{m}}^{+})\le 2c\alpha_{\mathrm{M}}K_{\mathrm{M}}^{+}$.

Now, let us show that for any $i\in\theta_D$ we have $|z_i(t)|^{\mu}\le m_i^2/K_{\mathrm{m}}^2$. There are the following two possibilities.

(a) $m_i^2/K_{\mathrm{m}}^2\ge 1$. In this case, since $|z_i(t)|\le\delta\le 1$ we have

$$|z_i(t)|^{\mu}\le 1\le\frac{m_i^2}{K_{\mathrm{m}}^2}.$$

(b) $m_i^2/K_{\mathrm{m}}^2<1$. Then, by the definition of $\delta$ we have

$$|z_i(t)|^{\mu}\le\delta^{\mu}\le\left(\frac{m_i}{K_{\mathrm{m}}}\right)^{\frac{2\mu}{\mu_{\mathrm{M}}}}=\left(\frac{m_i^2}{K_{\mathrm{m}}^2}\right)^{\frac{\mu}{\mu_{\mathrm{M}}}}\le\frac{m_i^2}{K_{\mathrm{m}}^2}$$

where we have considered that $\mu/\mu_{\mathrm{M}}\ge 1$.

Therefore, from (24) we have

$$
\begin{aligned}
v^{\mu}(t)&\le\left(\frac{1}{b_{\mathrm{m}}^{\mu}}+\left(2c\alpha_{\mathrm{M}}K_{\mathrm{M}}^{+}\right)^{\mu}\right)\\
&\quad\times\left(\sum_{i\,\in\,\theta_C^{\sharp}(t)}|z_i(t)|^{2\mu_i}+\sum_{i\,\in\,\theta_D^{\sharp}(t)}\frac{m_i^2}{K_{\mathrm{m}}^2}\right).
\end{aligned}\tag{25}
$$

Eqs. (23) and (25) thus yield

$$
\begin{aligned}
\dot v(t)&\le-\lambda K_{\mathrm{m}}^2\left(\sum_{i\,\in\,\theta_C^{\sharp}(t)}|z_i(t)|^{2\mu_i}+\sum_{i\,\in\,\theta_D^{\sharp}(t)}\frac{m_i^2}{K_{\mathrm{m}}^2}\right)\\
&=-\frac{\lambda K_{\mathrm{m}}^2}{\frac{1}{b_{\mathrm{m}}^{\mu}}+\left(2c\alpha_{\mathrm{M}}K_{\mathrm{M}}^{+}\right)^{\mu}}\left(\frac{1}{b_{\mathrm{m}}^{\mu}}+\left(2c\alpha_{\mathrm{M}}K_{\mathrm{M}}^{+}\right)^{\mu}\right)\\
&\quad\times\left(\sum_{i\,\in\,\theta_C^{\sharp}(t)}|z_i(t)|^{2\mu_i}+\sum_{i\,\in\,\theta_D^{\sharp}(t)}\frac{m_i^2}{K_{\mathrm{m}}^2}\right)\\
&\le-\frac{\lambda K_{\mathrm{m}}^2}{\frac{1}{b_{\mathrm{m}}^{\mu}}+\left(2c\alpha_{\mathrm{M}}K_{\mathrm{M}}^{+}\right)^{\mu}}\,v^{\mu}(t).
\end{aligned}\tag{26}
$$

In conclusion, we have shown that for a.a. $t>t_\delta$ we have

$$\dot v(t)\le -Qv^{\mu}(t)$$

where

$$Q=\frac{\lambda K_{\mathrm{m}}^2}{\frac{1}{b_{\mathrm{m}}^{\mu}}+\left(2c\alpha_{\mathrm{M}}K_{\mathrm{M}}^{+}\right)^{\mu}}>0.$$

By applying the result in Point (b) of Property 4, we obtain that $v(t)=0$ and $z(t)=0$ for $t\ge t_\phi$, where

$$t_\phi=t_\delta+\frac{v^{1-\mu}(t_\delta)}{Q(1-\mu)}\tag{27}$$

and $t_\delta$ is given in (22). Finally, (13) implies that $\gamma^{\mathrm{o}}(t)=0$ for a.a. $t\ge t_\phi$. $\square$

### 6.2. H-matrices

In this section, we suppose that the neuron interconnection matrix belongs to a subclass of the LDS interconnection matrices defined by means of the so called $H$-matrices.

**Definition 3** (*[24]*). We say that matrix $A\in\mathbb{R}^{n\times n}$ is an $M$-matrix, if and only if we have $A_{ij}\le 0$ for each $i\ne j$, and all successive principal minors of $A$ are positive.

**Definition 4** (*[24]*). We say that matrix $A\in\mathbb{R}^{n\times n}$ is an $H$-matrix if and only if the comparison matrix of $A$, which is defined as

$$[\mathcal{M}(A)]_{ij}=\begin{cases}|A_{ii}|, & i=j\\ -|A_{ij}|, & i\ne j\end{cases}\tag{28}$$

is an $M$-matrix.

Suppose that $-T$ is an $H$-matrix such that $T_{ii}<0$ for $i=1,\dots,n$. Then, $-T\in$ LDS [24]. Since $\mathcal{M}(-T)$ is an $M$-matrix, there exists a positive definite diagonal matrix $D=\mathrm{diag}(d_1,\dots,d_n)$ such that $D\mathcal{M}(-T)$ is strictly column-sum dominant, i.e., we have

$$-d_jT_{jj}-\sum_{\substack{i=1\\ i\ne j}}^{n}d_i|T_{ij}|>0,\qquad j=1,\dots,n.$$

Let us consider for (12) the (candidate) Lyapunov function

$$W(z)=\sum_{i=1}^{n}d_i|z_i|\tag{29}$$

which is positive definite and radially unbounded. Since $W$ is locally Lipschitz and convex in $\mathbb{R}^n$, then $W$ regular in $\mathbb{R}^n$ and hence satisfies Assumption 1. Note that $W$ is not differentiable.

The following holds.

**Theorem 3.** *Suppose that $g\in\mathbb{D}$, and that $-T$ is an $H$-matrix such that $T_{ii}<0$ for $i=1,\dots,n$. Moreover, suppose that for any $i\in\theta_D$ we have $G_i(0^+)>0$ and $G_i(0^-)<0$, while for any $i\in\theta_C$ there exist $\delta_i$, $K_i>0$, and $\mu_i\in(0,1)$, such that*

$$K_i|\rho|^{\mu_i}\le|G_i(\rho)|,\quad |\rho|<\delta_i.\tag{30}$$

*Let $z(t)$, $t\ge 0$, be any solution of (12), and $w(t)=W(z(t))$, $t\ge 0$. Moreover, let $\gamma^{\mathrm{o}}(t)$, for a.a. $t\ge 0$, be the output solution of (12) corresponding to $z(t)$. Then, there exists $t_\delta>0$ such that we have*

$$\dot w(t)\le -Qw^{\mu}(t),\quad \textit{for a.a. } t>t_\delta$$

*where*

$$\mu = \max_{i \in \theta_C} \{\mu_i\} \in (0, 1)$$

*and* $Q > 0$. *As a consequence, we have*

$$\begin{aligned} w(t) &= 0, \quad \forall t \geq t_\phi \\ z(t) &= 0, \quad \forall t \geq t_\phi \\ \gamma^{\mathrm{o}}(t) &= 0, \quad \textit{for a.a. } t \geq t_\phi \end{aligned}$$

*where*

$$t_\phi = t_\delta + \frac{w^{1-\mu}(t_\delta)}{Q(1-\mu)}$$

*i.e.,* $w(t)$, $z(t)$, *and* $\gamma^{\mathrm{o}}(t)$ *converge to zero in finite time* $t_\phi$.

Again, prior to the theorem proof we report some observations illustrating the result.

**Remarks.** 1. Condition (30) of Theorem 3 is less restrictive than conditions (20) and (21) in Theorem 3, and essentially means that for all $i \in \theta_C$ each $G_i$ should grow in a small neighborhood of zero at least as a $\mu_i$-Hölder function, where $\mu_i \in (0, 1)$. As was already noted before, the hypothesis $-T$ is an $H$-matrix such that $T_{ii} < 0$ for $i = 1, \ldots, n$, is more restrictive than the hypothesis $-T \in$ LDS made in Theorem 2. However, it is known that there are large classes of matrices of interest in the neural network applications, such as those modeling cooperative neural networks, for which the two concepts of $H$-matrices and LDS matrices coincide [24].

2. The theorem proof gives a simple quantitative estimate of the finite convergence time $t_\phi$, see relation (34) below.

**Proof of Theorem 3.** As in the proof of Theorem 2, we let

$$m_i = \min\{G_i(0^+), -G_i(0^-)\}$$

for each $i \in \theta_D$, and

$$K_{\mathrm{m}} = \min_{i \in \theta_C}\{K_i\}, \quad \mu = \max_{i \in \theta_C}\{\mu_i\}.$$

Note that $\mu \in (0, 1)$.

Let $D = \mathrm{diag}(d_1, \ldots, d_n)$, with $d_i > 0$, $i = 1, \ldots, n$, be such that $D\mathcal{M}(-T)$ is strictly column-sum dominant. We let

$$d_{\mathrm{M}} = \max\{d_1, \ldots, d_n\} > 0.$$

Moreover, for each $j \in \{1, \ldots, n\}$ define

$$\Delta T_j = -d_j T_{jj} - \sum_{\substack{i=1 \\ i \neq j}}^{n} d_i \left|T_{ij}\right| > 0$$

and

$$\Delta T_{\mathrm{m}} = \min\{\Delta T_1, \ldots, \Delta T_n\} > 0.$$

Finally, we let

$$\delta = \min\left\{1, \min\{\delta_1, \ldots, \delta_n\}, \min_{i \in \theta_D}\left\{\left(\frac{m_i}{K_{\mathrm{m}}}\right)^{\frac{1}{\mu}}\right\}\right\} > 0.$$

Again, from (15) it follows that for

$$t > t_\delta = \frac{2}{b_{\mathrm{m}}} \ln\left(\frac{\sqrt{b_{\mathrm{M}} w(0)}}{\delta}\right) \tag{31}$$

we have $z(t) \in [-\delta, \delta]^n$.

Taking into account that $z(t)$ and $w(t)$ are differentiable for a.a. $t > t_\delta$, consider $t > t_\delta$ at which both $z(t)$ and $w(t)$ are differentiable, and note that on the basis of (13) the output $\gamma^{\mathrm{o}}$ is well defined at $t$. Thus $\gamma^{\mathrm{o}}$ is defined for a.a. $t > t_\delta$. Let

$$\begin{aligned} N(t) &= \left\{i \in 1, \ldots, n : \gamma_i^{\mathrm{o}}(t) = 0\right\} \\ P(t) &= \{i \in 1, \ldots, n : z_i(t) = 0\}. \end{aligned}$$

For each $z \in R^n$, we have for (29)

$$\partial W(z) = D\overline{\mathrm{co}}\,[\mathrm{sign}(z)]$$

where $\mathrm{sign}(z) = (\mathrm{sign}(z_1), \ldots, \mathrm{sign}(z_n))^\top$. Since as was already noticed $z(t)$ is Lipschitz near $t$, and $W$ is regular at $z(t)$, it is possible to apply the chain rule in Property 1 to obtain that for a.a. $t \geq t_\delta$

$$\dot{w}(t) = \langle \zeta, \dot{z}(t) \rangle \quad \forall \zeta \in \partial W(z(t)).$$

If we let for $j = 1, \ldots, n$ and for a.a. $t \geq t_\delta$

$$u_j(t) = \begin{cases} \mathrm{sign}(z_j(t)) & \text{if } z_j(t) \neq 0 \\ \mathrm{sign}(\gamma_j^{\mathrm{o}}(t)) & \text{if } z_j(t) = 0 \end{cases}$$

then we have $Du(t) = D(u_1(t), \ldots, u_n(t))^\top \in \partial W(z(t))$, hence

$$\begin{aligned} \dot{w}(t) &= u(t)^\top D\dot{z}(t) = u(t)^\top D(Bz(t) + T\gamma^{\mathrm{o}}(t)) \\ &= -\sum_{i \notin P(t)} d_i b_i |z_i(t)| + u(t)^\top DT\gamma^{\mathrm{o}}(t) \\ &\leq -u(t)^\top D(-T)\gamma^{\mathrm{o}}(t) \end{aligned} \tag{32}$$

where we have considered that when $i \notin P(t)$ we have $u_i(t) = \mathrm{sign}(z_i(t))$ and $z_i(t) = |z_i(t)|\,\mathrm{sign}(z_i(t))$, moreover $z_i(t) = 0$ when $i \in P(t)$.

Now, note that if $j \notin N(t)$ we have $\gamma_j^{\mathrm{o}}(t) = |\gamma_j^{\mathrm{o}}(t)|\,\mathrm{sign}(\gamma_j^{\mathrm{o}}(t))$ and then

$$\begin{aligned} u(t)^\top D(-T)\gamma^{\mathrm{o}}(t) &= \sum_{i=1}^{n}\sum_{j=1}^{n} \left(-d_i T_{ij}\right) u_i(t)\gamma_j^{\mathrm{o}}(t) \\ &= \sum_{j=1}^{n} \gamma_j^{\mathrm{o}}(t) \sum_{i=1}^{n} \left(-d_i T_{ij}\right) u_i(t) \\ &= \sum_{j \notin N(t)} \left|\gamma_j^{\mathrm{o}}(t)\right| \Bigg(-d_j T_{jj}\, u_j(t)\,\mathrm{sign}(\gamma_j^{\mathrm{o}}(t)) \\ &\quad + \sum_{\substack{i=1 \\ i \neq j}}^{n} \left(-d_i T_{ij}\right) u_i(t)\,\mathrm{sign}(\gamma_j^{\mathrm{o}}(t))\Bigg) \\ &\geq \sum_{j \in P(t) \setminus N(t)} \left|\gamma_j^{\mathrm{o}}(t)\right| \left(-d_j T_{jj}\,\mathrm{sign}^2(\gamma_j^{\mathrm{o}}(t)) - \sum_{\substack{i=1 \\ i \neq j}}^{n} d_i \left|T_{ij}\right|\right) \end{aligned}$$

$$+ \sum_{j \notin P(t) \cup N(t)} \left|\gamma_j^{\mathrm{o}}(t)\right| \left( - d_j T_{jj} \operatorname{sign}(z_j(t)) \operatorname{sign}(\gamma_j^{\mathrm{o}}(t)) - \sum_{\substack{i=1 \\ i \neq j}}^{n} d_i \left|T_{ij}\right| \right).$$

For any $j \notin N(t)$ we have $\operatorname{sign}^2(\gamma_j^{\mathrm{o}}(t)) = 1$, while for $j \notin P(t) \cup N(t)$ we obtain $z_j(t) \neq 0 \neq \gamma_j^{\mathrm{o}}(t)$, $\operatorname{sign}(z_j(t)) = \operatorname{sign}(\gamma_j^{\mathrm{o}}(t)) \neq 0$ and so $\operatorname{sign}(z_j(t)) \operatorname{sign}(\gamma_j^{\mathrm{o}}(t)) = 1$. Therefore,

$$\begin{aligned}
&u(t)^\top D(-T)\gamma^{\mathrm{o}}(t) \\
&\geq \sum_{j \notin (P(t) \cup N(t))} \left|\gamma_j^{\mathrm{o}}(t)\right| \left( -d_j T_{jj} - \sum_{\substack{i=1 \\ i \neq j}}^{n} d_i \left|T_{ij}\right| \right) \\
&\geq \sum_{j \in \theta_C \setminus (P(t) \cup N(t))} \Delta T_j \left|G_j(z_j(t))\right| + \sum_{j \in \theta_D \setminus (P(t) \cup N(t))} \Delta T_j m_j \\
&\geq \Delta T_{\mathrm{m}} \left( \sum_{j \in \theta_C \setminus (P(t) \cup N(t))} K_j \left|z_j(t)\right|^{\mu_j} + \sum_{j \in \theta_D \setminus (P(t) \cup N(t))} m_j \right).
\end{aligned}$$

Now, note that $K_j \left|z_j(t)\right|^{\mu_j} \geq K_{\mathrm{m}} \left|z_j(t)\right|^{\mu}$ for each $j \in \theta_C$. If instead $j \in \theta_D$, we have $z_j(t) \leq \delta \leq (m_j/K_{\mathrm{m}})^{1/\mu}$, hence $z_j(t)^\mu \leq \delta^\mu \leq m_j/K_{\mathrm{m}}$ and $K_{\mathrm{m}} z_j(t)^\mu \leq m_j$. Under our assumptions on the functions $z_j \to G_j(z_j)$, $j = 1, \ldots, n$, for $j \in P(t) \cup N(t)$ we have $z_j(t) = 0$, hence

$$\begin{aligned}
&u(t)^\top D(-T)\gamma^{\mathrm{o}}(t) \geq \Delta T_{\mathrm{m}} K_{\mathrm{m}} \\
&\quad \times \left( \sum_{j \in \theta_C \setminus (P(t) \cup N(t))} \left|z_j(t)\right|^\mu + \sum_{j \in \theta_D \setminus (P(t) \cup N(t))} \left|z_j(t)\right|^\mu \right) \\
&= \Delta T_{\mathrm{m}} K_{\mathrm{m}} \left( \sum_{j=1}^{n} \left|z_j(t)\right|^\mu \right) \geq \Delta T_{\mathrm{m}} K_{\mathrm{m}} \left( \sum_{j=1}^{n} \left|z_j(t)\right| \right)^\mu \\
&\geq \frac{\Delta T_{\mathrm{m}} K_{\mathrm{m}}}{d_{\mathrm{M}}^\mu} \left( \sum_{j=1}^{n} d_j \left|z_j(t)\right| \right)^\mu = Q w^\mu(t) \qquad (33)
\end{aligned}$$

where we have let

$$Q = \frac{\Delta T_{\mathrm{m}} K_{\mathrm{m}}}{d_{\mathrm{M}}^\mu} > 0.$$

By substituting (33) in (32), we conclude that we have

$$\dot{w}(t) \leq -Q w^\mu(t)$$

for a.a. $t \geq t_\delta$, where $\mu \in (0, 1)$.

By applying the result in Point (b) of Property 4, it follows that we have $w(t) = 0$ and $z(t) = 0$ for $t \geq t_\phi$, where

$$t_\phi = t_\delta + \frac{w^{1-\mu}(t_\delta)}{Q(1-\mu)} \qquad (34)$$

and $t_\delta$ is given in (31). Finally, by exploiting (13) we also obtain that $\gamma^{\mathrm{o}}(t) = 0$ for a.a. $t \geq t_\phi$. □

## 7. Conclusion

The paper has proved results on global exponential convergence toward a unique equilibrium point, and global convergence in finite time, for a class of additive neural networks possessing discontinuous neuron activations or continuous non-Lipschitz neuron activations. The results are of potential interest in view of the neural network applications for solving global optimization problems in real time, where global convergence toward an equilibrium point, fast convergence speed and the ability to quantitatively estimate the convergence time, are of crucial importance. The results have been proved by means of a generalized Lyapunov-like approach, which has been developed in the paper, and is suitable for addressing convergence of nonsmooth dynamical systems described by differential equations with discontinuous right-hand side.

An important open question is whether the results on global convergence here obtained may be extended to more general neural network models incorporating the presence of a delay in the neuron interconnections. This topic goes beyond the scope of the present paper and will constitute a challenging issue for future investigations.

## References

[1] M. Forti, P. Nistri, Global convergence of neural networks with discontinuous neuron activations, IEEE Trans. Circuits Syst. I 50 (11) (2003) 1421–1435.
[2] M. Hirsch, Convergent activation dynamics in continuous time networks, Neural Netw. 2 (1989) 331–349.
[3] C.M. Marcus, R.M. Westervelt, Stability of analog neural network with delay, Phys. Rev. A 39 (1989) 347–359.
[4] E. Kaszkurewicz, A. Bhaya, On a class of globally stable neural circuits, IEEE Trans. Circuits Syst. I 41 (1994) 171–174.
[5] M. Forti, A. Tesi, New conditions for global stability of neural networks with application to linear and quadratic programming problems, IEEE Trans. Circuits Syst. I 42 (7) (1995) 354–366.
[6] X.-B. Liang, J. Wang, An additive diagonal stability condition for absolute stability of a general class of neural networks, IEEE Trans. Circuits Syst. I 48 (11) (2001) 1308–1317.
[7] X.-B. Liang, J. Si, Global exponential stability of neural networks with globally lipschitz continuous activations and its application to linear variational inequality problem, IEEE Trans. Neural Netw. 12 (2) (2001) 349–359.
[8] S. Arik, Global robust stability of delayed neural networks, IEEE Trans. Circuits Syst. I 50 (1) (2003) 156–160.
[9] S. Hu, J. Wang, Absolute exponential stability of a class of continuous-time recurrent neural networks, IEEE Trans. Neural Netw. 14 (1) (2003) 35–45.
[10] J. Cao, J. Wang, Absolute exponential stability of recurrent neural networks with Lipschitz-continuous activation functions and time delays, Neural Netw. 17 (2004) 379–390.
[11] J. Cao, J. Lianga, J. Lamb, Exponential stability of high-order bidirectional associative memory neural networks with time delays, Invent. Math. 136 (1999) 75–87.
[12] H. Qi, L. Qi, Deriving sufficient conditions for global asymptotic stability of delayed neural networks via nonsmooth analysis, IEEE Trans. Neural Netw. 15 (2004) 99–109.
[13] E. Yucel, S. Arik, New exponential stability results for delayed neural networks with time varying delays, Physica D 191 (2004) 314–322.

[14] X. Liao, C. Li, An LMI approach to asymptotical stability of multi-delayed neural networks, Physica D 200 (2005) 139–155.

[15] S.H. Żak, V. Upatising, S. Hui, Solving linear programming problems with neural networks: A comparative study, IEEE Trans. Neural Netw. 6 (1995) 94–104.

[16] M. Forti, P. Nistri, M. Quincampoix, Generalized neural network for nonsmooth nonlinear programming problems, IEEE Trans. Circuits Syst. I 51 (9) (2004) 1741–1754.

[17] D.P. Bertsekas, Necessary and sufficient conditions for a penalty method to be exact, Math. Program. 9 (1975) 87–99.

[18] E.K.P. Chong, S. Hui, S.H. Żak, An analysis of a class of neural networks for solving linear programming problems, IEEE Trans. Automat. Control 44 (1999) 1095–2006.

[19] L.V. Ferreira, E. Kaszkurewicz, A. Bhaya, Solving systems of linear equations via gradient systems with discontinuous right hand sides: Application to LS-SVM, IEEE Trans. Neural Netw. 16 (2) (2005) 501–505.

[20] R. Gavaldá, H.T. Siegelmann, Discontinuities in recurrent neural networks, Neural Comput. 11 (1999) 715–745.

[21] H.T. Siegelmann, E.D. Sontag, Analog computation via neural networks, Theoret. Comput. Sci. 131 (1994) 331–360.

[22] A.N. Michel, R.K. Miller, Qualitative Analysis of Large Scale Dynamical Systems, Academic, New York, 1977.

[23] A.N. Michel, J.A. Farrell, W. Porod, Qualitative analysis of neural networks, IEEE Trans. Circuits Syst. 36 (2) (1989) 229–243.

[24] D. Hershkowitz, Recent directions in matrix stability, Linear Algebr. Appl. 171 (1992) 161–186.

[25] A. Berman, R.J. Plemmons, Nonnegative Matrices in the Mathematical Sciences, Academic, New York, 1979.

[26] F.H. Clarke, Optimization and Non-Smooth Analysis, John Wiley & Sons, New York, 1983.

[27] A.F. Filippov, Differential equations with discontinuous right-hand side, Transl. American Math. Soc. 42 (1964) 199–231.

[28] X.-B. Liang, J. Wang, Absolute exponential stability of neural networks with a general class of activation functions, IEEE Trans. Circuits Syst. I 47 (8) (2000) 1258–1263.

[29] B.E. Paden, S.S. Sastry, Calculus for computing Filippov's differential inclusion with application to the variable structure control of robot manipulator, IEEE Trans. Circuits Syst. 34 (1987) 73–82.

[30] V.I. Utkin, Sliding Modes and Their Application in Variable Structure Systems, MIR Publishers, Moscow, 1978.

[31] J.P. Aubin, A. Cellina, Differential Inclusions, Springer-Verlag, Berlin, 1984.