A Result on Global Convergence in Finite Time for Nonsmooth Neural Networks

M. Forti, M. Grazzini, P. Nistri and L. Pancioni
Department of Information Engineering
University of Siena
Via Roma, 56 - 53100 Siena, Italy
E-mail: {forti, grazzini, pnistri, pancioni}@dii.unisi.it

***Abstract*— The paper considers a large class of additive neural networks where the neuron activations are modeled by discontinuous functions or by non-Lipschitz functions. A result is established guaranteeing that the state solutions and output solutions of the neural network are globally convergent in finite time toward a unique equilibrium point. The obtained result, which generalizes previous results on convergence in finite time in the literature, is of interest for designing neural networks aimed at solving global optimization problems in real time.**

I. Introduction

In recent years, neural networks with discontinuous neuron activations or with non-Lipschitz neuron activations, have been found useful for addressing several interesting engineering tasks.

In [1], additive neural networks with jump discontinuities have been applied to the solution of global optimization problems. By exploiting the sliding modes in the neural network dynamics, conditions for *global convergence in finite time* toward a unique equilibrium point have been established in [1, Th. 4]. Global convergence is important, in that it prevents a network from the risk of getting stuck at some local minimum of the energy function, see e.g. [2]–[7], and references therein. The property of global convergence in finite time is even more desirable when the minimum must be computed in real time [8], [9]. We stress that such a property cannot be displayed by smooth dynamical systems, since in that case there can be only asymptotic convergence toward an equilibrium point.

Another related field of application for discontinuous neural networks concerns the implementation of analog devices for solving linear and nonlinear programming problems [10]. The neural network in [10] makes use of constraint neurons modeled by ideal diodes with a vertical segment in the conducting region. On this basis the network is able to implement an *exact penalty method* where the circuit equilibrium points coincide with the constrained critical points of the objective function. This permits the network to compute the exact constrained optimal solution for interesting classes of programming problems.

In [11], a classical recurrent neural network as in [12] has been augmented with a few simple discontinuous neuron activations, such as binary threshold functions. It is shown that the use of these discontinuous activations permits 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. A fundamental observation made in [11], 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.

In this paper, we consider the class of additive neural networks introduced in [1], with the goal to obtain conditions of more general applicability for ensuring global convergence in finite time of the state and output solutions. More specifically, while the conditions for convergence in finite time in [1] require that all neuron activations be discontinuous at the equilibrium point, the conditions in this paper are applicable both to discontinuous neuron activations, and to continuous neuron activations that are non-Lipschitz at the equilibrium point.

Notation. Let $x = (x_1, \dots, x_n)^\top, y = (y_1, \dots, y_n)^\top \in \mathbb{R}^n$, where the symbol $\top$ means the transpose, be two given column vectors. By $\langle x, y\rangle = \sum_{i=1}^n x_i y_i$ we denote the usual scalar product in $\mathbb{R}^n$ between x and y. By $\|x\| = \langle x, x\rangle^{1/2} = (\sum_{i=1}^n x_i^2)^{1/2}$ we mean 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 $\overline{\mathrm{co}}\,(Q)$ we denote the closure of the convex hull of set $Q \subset \mathbb{R}^n$.

II. Preliminaries

Here, we report a number of definitions and properties which are needed in the development [13], [14].

A function $f : \mathbb{R}^n \to \mathbb{R}^m$ 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 $f : \mathbb{R}^n \to \mathbb{R}^n$ 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 the Clarke's generalized gradient of f at point x, as follows

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

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 $\partial f : \mathbb{R}^n \multimap \mathbb{R}^n$ is a set-valued map that associates to any $x \in \mathbb{R}^n$ a non-empty compact convex subset $\partial f(x) \subset \mathbb{R}^n$.

The following class of matrices will play a key role in the paper.

Definition 1: Matrix $A \in \mathbb{R}^{n\times n}$ is said to be Lyapunov Diagonally Stable (LDS), if there exists a positive definite diagonal matrix $\alpha = \mathrm{diag}(\alpha_1, \ldots, \alpha_n)$, such that $(1/2)(\alpha A + A^\top \alpha)$ is positive definite.

III. Neural Network Model

Let us consider a class of additive neural networks whose dynamics is governed by the system of differential equations with *discontinuous* right-hand side

$$\dot{x} = Bx + Tg(x) + I \tag{1}$$

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

Henceforth, we suppose that g belongs to the next class of mappings.

Definition 2: We say that $g \in \mathbb{G}$ if and only if, for $i = 1, \ldots, n$, $g_i : \mathbb{R} \to \mathbb{R}$ is a bounded, non-decreasing, and piecewise continuous function on $\mathbb{R}$ (this means that g_i has at most points of jump discontinuity, where there exist finite right and left limits, and that g_i has a finite number of discontinuities in any compact interval of $\mathbb{R}$). ■

We remark that the class of functions in Definition 2 includes activations g_i with jump discontinuities, and also continuous activations g_i that are not locally Lipschitz.

A function $x : [t_a, t_b] \to \mathbb{R}^n$ is a solution in the sense of Filippov of (1), with initial condition $x(t_a) = x_0 \in \mathbb{R}^n$, if the following hold [15]: $x(t)$ 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

$$\dot{x}(t) \in Bx(t) + T\,\overline{\mathrm{co}}\,[g(x(t))] + I \tag{2}$$

where $\overline{\mathrm{co}}\,[g(x)] = (\overline{\mathrm{co}}\,[g_1(x_1)], \ldots, \overline{\mathrm{co}}\,[g_n(x_n)])^\top \subset \mathbb{R}^n$ and $\overline{\mathrm{co}}\,[g_i(x_i)] = [g_i(x_i^-), g_i(x_i^+)] \subset \mathbb{R}$ 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. Clearly, the inclusion (2) 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 (1), and suppose that $\det T \neq 0$. Then, for a.a. $t \in [t_a, t_b]$ we have

$$\dot{x}(t) = Bx(t) + T\gamma(t) + I \tag{3}$$

where

$$\gamma(t) = T^{-1}(\dot{x}(t) - Bx(t) - I) \in \overline{\mathrm{co}}\,[g(x(t))] \tag{4}$$

is the *output solution* of (1) corresponding to $x(t)$. It turns out that γ is a bounded measurable function that is uniquely defined, up to a set with measure zero of the interval $[t_a, t_b]$, by the *state* solution x.

By an *equilibrium point* (EP) $\xi \in \mathbb{R}^n$ of (1) we mean a stationary solution $x(t) = \xi$, $t \geq 0$, of (1). Of course, $\xi \in \mathbb{R}^n$ is an EP of (1) if and only if ξ satisfies the following algebraic inclusion

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

Let ξ be an EP of (N). Then, from (4) it turns out that

$$\eta = T^{-1}(-B\xi - I) \in \overline{\mathrm{co}}\,[g(\xi)] \tag{5}$$

is the unique *output equilibrium point* (OEP) of (N) corresponding to ξ.

In [1] it has been shown that if $g \in \mathbb{G}$, then for any $x_0 \in \mathbb{R}^n$ there is at least a solution $x(t)$ of (1) with initial condition $x(0) = x_0$, which is bounded and hence defined for all $t \geq 0$. Furthermore, there exists at least an EP $\xi \in \mathbb{R}^n$ of (1) with corresponding OEP η.

To address global convergence of the solutions of (1), we will find it useful in the next section to consider the standard change of variables $z = x - \xi$, where ξ is an EP of (1). This transforms (1) into the differential equation

$$\dot{z} = Bz + TG(z) \tag{6}$$

where $G(z) = g(z + \xi) - \eta$. If $g \in \mathbb{G}$, then $G \in \mathbb{G}$. Note that (6) has an EP, and a corresponding OEP, which are both located at 0. If $z(t)$, $t \geq 0$, is a solution of (6), then we shall denote by

$$\gamma^\circ(t) = T^{-1}(\dot{z}(t) - Bz(t)) \in \overline{\mathrm{co}}\,[G(z(t))] \tag{7}$$

the output solution of (6) corresponding to $z(t)$, which is defined for a.a. $t \geq 0$.

IV. Global Convergence in Finite Time

Suppose that $-T \in$ LDS and $g \in \mathbb{G}$. Under these assumptions, it has been proved in [1, Th. 4] that all solutions of (1) are globally convergent toward a unique EP, while the corresponding output solutions of (1) are globally convergent in measure toward the corresponding unique OEP. In the next theorem, which is the main result in this paper, we make a suitable assumption on the location of the EP of (1) with respect to the points where the activations are discontinuous or non-Lipschitz. On this basis, the theorem establishes a result on global convergence in finite time toward the EP and the OEP, respectively, of the state and output solutions of (1).

Theorem 1: Suppose that $-T \in$ LDS and that $g \in \mathbb{G}$. Let $\theta_{\mathrm{D}} = \{i \in \{1, \ldots, n\} : G_i$ is discontinuous at $z_i = 0\}$ and $\theta_{\mathrm{C}} = \{1, \ldots, n\} \backslash \theta_{\mathrm{D}}$. Suppose that for any $i \in \theta_{\mathrm{D}}$ we have $G_i(0^+) > 0$ and $G_i(0^-) < 0$. Furthermore, for any $i \in \theta_{\mathrm{C}}$ there exist $\delta_i, k_i, K_i > 0$, $\sigma_i \in (0,1)$ and $\Sigma_i \in [0,1)$, such that

$$k_i|\rho|^{\sigma_i} \leq |G_i(\rho)| \leq K_i|\rho|^{\Sigma_i}, \quad |\rho| < \delta_i \tag{8}$$

and

$$\mu_{\mathrm{M}} = \max_{i\in\theta_{\mathrm{C}}} \frac{2\sigma_i}{1 + \Sigma_i} < 1. \tag{9}$$

Let $z(t)$, $t \geq 0$, be any solution, and let $\gamma^{\circ}(t)$ be the corresponding output solution of (1), which is defined in (7) for a.a. $t \geq 0$. Then, there exists $+\infty > t_{\phi} > 0$, such that we have $V(z(t)) = 0$ and $z(t) = 0$ for all $t \geq t_{\phi}$, while $\gamma^{\circ}(t) = 0$ for a.a. $t \geq t_{\phi}$. This means that $V(z(t))$, $z(t)$, and $\gamma^{\circ}(t)$ converge to zero in finite time t_{ϕ}. ■

The left inequality in (8) means that, for any $i \in \theta_{\mathrm{C}}$, G_i is continuous but non-Lipschitz at 0, since it grows at least as $k_i|\rho|^{\sigma_i}$, where $\sigma_i \in (0,1)$, in a neighborhood of 0. It is not difficult to verify that the following classes of continuous non-Lipschitz functions G_i satisfy assumptions (8) and (9) of Theorem 1:

1) for any $i \in \theta_{\mathrm{C}}$, (8) is satisfied by G_i with $\sigma_i < 1/2$;
2) for any $i \in \theta_{\mathrm{C}}$, G_i is defined by

$$G_i(\rho) = k_i \mathrm{sgn}(\rho)|\rho|^{\sigma_i}$$

where $\sigma_i \in (0,1)$ and $k_i > 0$.

Theorem 1 is an extension of the previous result on global convergence in finite time given by [1, Th. 4]. In fact, Theorem 1 can be applied to the case where the neuron activations G_i are either discontinuous at 0, or they are non-Lipschitz functions in a neighborhood of 0. Instead, [1, Th. 4] requires that all neuron activations be discontinuous at 0.

Finally, we observe that the proof of Theorem 1 yields a quantitative estimate of the finite convergence time t_{ϕ}, see (17).

Proof of Theorem 1. We need the following notations. Let

$$k_{\mathrm{m}} = \min_{i\in\theta_C}\{k_i\} > 0, \quad \Sigma_{\mathrm{m}} = \min_{i\in\theta_C}\{\Sigma_i\} \geq 0.$$

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

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

Finally, we let

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

Now, consider for (6) the (candidate) Lyapunov function

$$V(z) = \sum_{i=1}^{n} \frac{1}{b_i} z_i^2 + 2c \sum_{i=1}^{n} \alpha_i \int_0^{z_i} G_i(\rho)\, d\rho \qquad (10)$$

where $c > 0$ is a constant and α is as in Definition 1. Let

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

Consider the function $v(t) = V(z(t))$, $t \geq 0$. Since $z(t)$ and $v(t)$ are absolutely continuous for $t \geq 0$, then both functions $z(t)$ and $v(t)$ are differentiable for a.a. $t \in [0, +\infty)$ and we have, on the basis of the 'chain rule' given in [1, Prop. 6],

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

Moreover, by evaluating the scalar product, it is proved in [1, App. IV] that

$$\begin{aligned}\dot{v}(t) \quad \leq \quad & -\|z(t)\|^2 - \|B^{-1}\dot{z}(t)\|^2 - \lambda\|\gamma^{\circ}(t)\|^2 \\ & +2cz^{\top}(t)\alpha B\gamma^{\circ}(t)\end{aligned} \qquad (11)$$

for some $\lambda > 0$.

Since the trajectory $z(t)$ converges to 0 as $t \to +\infty$ [1, Th. 2], we can find an instant t_{δ} such that $z(t) \in [-\delta, \delta]^n$ for all $t > t_{\delta}$. Let $t > t_{\delta}$ and suppose that $v(t)$ and $z(t)$ are differentiable at t. From the inequality $\delta \leq 1$, we have $|z_i(t)|^{p_2} \leq |z_i(t)|^{p_1} \leq 1$, for any $p_1, p_2 > 0$ such that $p_1 \leq p_2$. Let $P(t) = \{i \in 1, \ldots, n : z_i(t) = 0\}$, $\tilde{\theta}_C(t) = \theta_C \backslash P(t)$ and $\tilde{\theta}_D(t) = \theta_D \backslash P(t)$. For any $i \in P(t)$, we have $|\gamma_i^{\circ}(t)| \geq 0 = z_i(t)$, while for any $i \in \tilde{\theta}_C(t)$ we have $|\gamma_i^{\circ}(t)| = |G_i(z_i(t))| \geq k_i |z_i(t)|^{\sigma_i}$. Moreover, for any $i \in \tilde{\theta}_D(t)$ we obtain $m_i \leq |\gamma_i^{\circ}(t)| = |G_i(z_i(t))| \leq K_i$.

Since each G_i is monotone non-decreasing and $0 \in \overline{\mathrm{co}}\,[G_i(0)]$, it follows that

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

and from (11) we obtain

$$\begin{aligned}\dot{v}(t) &\leq -\lambda\|\gamma^{\circ}(t)\|^2 = -\lambda\sum_{i=1}^{n}|\gamma_i^{\circ}(t)|^2 \\ &\leq -\lambda k_{\mathrm{m}}^2\left(\sum_{i\in\tilde{\theta}_C(t)}|z_i(t)|^{2\sigma_i} + \sum_{i\in\tilde{\theta}_D(t)}\frac{m_i^2}{k_{\mathrm{m}}^2}\right).\end{aligned} \qquad (12)$$

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

$$\begin{aligned}v^{\mu}(t) \quad \leq \quad & \left(\sum_{i\notin P(t)}\frac{|z_i(t)|^2}{b_i} + 2c\sum_{i\in\tilde{\theta}_C(t)}\alpha_i K_i\frac{|z_i(t)|^{1+\Sigma_i}}{1+\Sigma_i}\right. \\ & \left. +2c\sum_{i\in\tilde{\theta}_D(t)}\alpha_i K_i|z_i(t)|\right)^{\mu}.\end{aligned}$$

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

$$\begin{aligned}v^{\mu}(t) \leq & \sum_{i\notin P(t)}\frac{|z_i(t)|^{2\mu}}{b_{\mathrm{m}}^{\mu}} + (2c)^{\mu}\left(\sum_{i\in\tilde{\theta}_C(t)}\left(\frac{\alpha_i K_i}{1+\Sigma_i}\right)^{\mu}\right. \\ & \left.\cdot|z_i(t)|^{\mu(1+\Sigma_i)} + \sum_{i\in\tilde{\theta}_D(t)}(\alpha_i K_i)^{\mu}|z_i(t)|^{\mu}\right).\end{aligned}$$

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

and $|z_i(t)|^{2\mu} \leq |z_i(t)|^{\mu}$. Therefore,

$$v^{\mu}(t) \leq \left(\frac{1}{b_{\rm m}^{\mu}} + \left(\frac{2c\alpha_{\rm M}K_{\rm M}}{1+\Sigma_{\rm m}}\right)^{\mu}\right) \sum_{i \in \tilde{\theta}_C(t)} |z_i(t)|^{\mu(1+\Sigma_i)} + \left(\frac{1}{b_{\rm m}^{\mu}} + (2c\alpha_{\rm M}K_{\rm M})^{\mu}\right) \sum_{i \in \tilde{\theta}_D(t)} |z_i(t)|^{\mu}.$$

If $\mu \in [\mu_{\rm M}, 1)$, then from (9) we obtain $\mu(1+\Sigma_i) \geq 2\sigma_i$ and hence

$$v^{\mu}(t) \leq \left(\frac{1}{b_{\rm m}^{\mu}} + (2c\alpha_{\rm M}K_{\rm M})^{\mu}\right) \cdot \left(\sum_{i \in \tilde{\theta}_C(t)} |z_i(t)|^{2\sigma_i} + \sum_{i \in \tilde{\theta}_D(t)} |z_i(t)|^{\mu}\right) \quad (13)$$

where we have taken into account that $(2c\alpha_{\rm M}K_{\rm M})/(1+\Sigma_{\rm m}) \leq 2c\alpha_{\rm M}K_{\rm M}$.

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

a) $m_i^2/k_{\rm m}^2 \geq 1$. In this case, since $|z_i(t)| \leq \delta \leq 1$ we have

$$|z_i(t)|^{\mu} \leq 1 \leq \frac{m_i^2}{k_{\rm m}^2}.$$

b) $m_i^2/k_{\rm m}^2 < 1$. Then, by the definition of δ we have

$$|z_i(t)|^{\mu} \leq \delta^{\mu} \leq \left(\frac{m_i}{k_{\rm m}}\right)^{\frac{2\mu}{\mu_{\rm M}}} = \left(\frac{m_i^2}{k_{\rm m}^2}\right)^{\frac{\mu}{\mu_{\rm M}}} \leq \frac{m_i^2}{k_{\rm m}^2}$$

where we have considered that $\mu/\mu_{\rm M} \geq 1$.

Therefore, from (13) we have

$$v^{\mu}(t) \leq \left(\frac{1}{b_{\rm m}^{\mu}} + (2c\alpha_{\rm M}K_{\rm M})^{\mu}\right) \cdot \left(\sum_{i \in \tilde{\theta}_C(t)} |z_i(t)|^{2\sigma_i} + \sum_{i \in \tilde{\theta}_D(t)} \frac{m_i^2}{k_{\rm m}^2}\right). \quad (14)$$

Equations (12) and (14) thus yield

$$\dot{v}(t) \leq -\lambda k_{\rm m}^2 \left(\sum_{i \in \tilde{\theta}_C(t)} |z_i(t)|^{2\sigma_i} + \sum_{i \in \tilde{\theta}_D(t)} \frac{m_i^2}{k_{\rm m}^2}\right) \leq -\frac{\lambda k_{\rm m}^2}{\frac{1}{b_{\rm m}^{\mu}} + (2c\alpha_{\rm M}K_{\rm M})^{\mu}} v^{\mu}(t). \quad (15)$$

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

$$\dot{v}(t) \leq -Qv^{\mu}(t) \leq 0 \quad (16)$$

where

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

Now, we prove that there exists a time instant $+\infty > t' > t_{\delta}$ such that $v(t') = 0$. Suppose, for the purpose of contradiction, that $v(t) > 0$ for every $t > t_{\delta}$. Then,

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

for a.a. $t > t_{\delta}$ and hence, integrating both sides of the inequality between t_{δ} and t, we obtain

$$v^{1-\mu}(t) \leq v^{1-\mu}(t_{\delta}) - Q(1-\mu)(t - t_{\delta})$$

for each $t > t_{\delta}$. Then, we have $v(t) < 0$ for

$$t > t_{\phi} = t_{\delta} + \frac{v^{1-\mu}(t_{\delta})}{Q(1-\mu)} \quad (17)$$

thus contradicting the previous hypothesis. The same argument also shows that $t' \leq t_{\phi}$. It remains to prove that $v(t) = 0$ for every $t > t'$. In fact, if $v(\tau) > 0$ for some $\tau > t'$, then there exists $t \in (t', \tau)$ such that $\dot{v}(t) > 0$, which contradicts (16). In conclusion, we have $v(t) = 0$ for every $t \geq t_{\phi}$, and this implies that $z(t) = 0$ for $t \geq t_{\phi}$. Moreover, we have $\gamma^{o}(t) = 0$ for a.a. $t \geq t_{\phi}$, see (7). ■

V. Conclusion

In the design of analog neural networks for solving optimization problems in real time, it is often desirable that the state and output solutions of the neural network be globally convergent in finite time toward a unique equilibrium point. The paper has shown that global convergence in finite time can be achieved not only for networks with discontinuous neuron activations, but also for certain classes of networks possessing continuous non-Lipschitz neuron activations.

References

[1] M. Forti and P. Nistri, "Global convergence of neural networks with discontinuous neuron activations," *IEEE Trans. Circuits Syst. I*, vol. 50, no. 11, pp. 1421–1435, November 2003.

[2] M. Hirsch, "Convergent activation dynamics in continuous time networks," *Neural Networks*, vol. 2, pp. 331–349, 1989.

[3] 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, no. 7, pp. 354–366, July 1995.

[4] 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, no. 11, pp. 1308–1317, November 2001.

[5] S. Arik, "Global robust stability of delayed neural networks," *IEEE Trans. Circuits Syst. I*, vol. 50, no. 1, pp. 156–160, January 2003.

[6] S. Hu and J. Wang, "Absolute exponential stability of a class of continuous-time recurrent neural networks," *IEEE Trans. Neural Networks*, vol. 14, no. 1, pp. 35–45, January 2003.

[7] J. Cao and J. Wang, "Global asymptotic and robust stability of recurrent neural networks with time delays," *IEEE Trans. Circuits Syst. I*, vol. 52, no. 2, pp. 417–426, February 2005.

[8] 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. Automatic Control*, vol. 44, pp. 1095–2006, November 1999.

[9] L. V. Ferreira, E. Kaszkurewicz, and A. Bhaya, "Solving systems of linear equations via gradient systems with discontinuous right hand sides: application to LS-SVM," *IEEE Trans. Neural Networks*, vol. 16, no. 2, pp. 501–505, March 2005.

[10] M. Forti, P. Nistri, and M. Quincampoix, "Generalized neural network for nonsmooth nonlinear programming problems," *IEEE Trans. Circuits Syst. I*, vol. 51, no. 9, pp. 1741–1754, September 2004.

[11] R. Gavaldá and H. Siegelmann, "Discontinuities in recurrent neural networks," *Neural Computation*, vol. 11, pp. 715–745, April 1999.

[12] H. T. Siegelmann and E. D. Sontag, "Analog computation via neural networks," *Theoretical Computer Science*, vol. 131, pp. 331–360, 1994.

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

[14] D. Hershkowitz, "Recent directions in matrix stability," *Linear Algebra Appl.*, vol. 171, pp. 161–186, 1992.

[15] A. F. Filippov, "Differential equations with discontinuous right-hand side," *Transl. American Math. Soc.*, vol. 42, pp. 199–231, 1964.