



Global convergence of neural networks with mixed time-varying delays and discontinuous neuron activations [☆]

Jun Liu ^a, Xinzhi Liu ^{a,*}, Wei-Chau Xie ^b

^a Department of Applied Mathematics, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1

^b Department of Civil and Environmental Engineering, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1

ARTICLE INFO

Article history:

Received 4 March 2010
Received in revised form 16 May 2011
Accepted 17 August 2011
Available online 25 August 2011

Keywords:

Global convergence
Global exponential stability
Neural networks
Time-varying delays
Distributed delays
Discontinuous neuron activations

ABSTRACT

In this paper, we investigate the dynamical behavior of a class of delayed neural networks with *discontinuous neuron activations* and general mixed time-delays involving both time-varying delays and distributed delays. Due to the presence of time-varying delays and distributed delays, the step-by-step construction of local solutions cannot be applied. This difficulty can be overcome by constructing a sequence of solutions to delayed dynamical systems with high-slope activations and show that this sequence converges to a desired Filippov solution of the discontinuous delayed neural networks. We then derive two sets of sufficient conditions for the global exponential stability and convergence of the neural networks, in terms of linear matrix inequalities (LMIs) and *M*-matrix properties (equivalently, some diagonally dominant conditions), respectively. Convergence behavior of both the neuron state and the neuron output are discussed. The obtained results extend previous work on global stability of delayed neural networks with Lipschitz continuous neuron activations, and neural networks with discontinuous neuron activations and only constant delays.

© 2011 Elsevier Inc. All rights reserved.

1. Introduction

A neural network (or artificial neural network) is a mathematical or computational model inspired by the structural and functional aspects of the vast network of neurons in the human brain. Neural networks process information and perform computing tasks by connecting a large number of simple computing units, called (artificial) neurons.

A neural network model can be regarded as a class of nonlinear signal-flow graphs [16]. Each neuron takes weighted inputs from other connected neurons and produces an output through an activation function (often nonlinear). Therefore, neural networks are in fact a class of nonlinear dynamical systems due to the nonlinearity of activations. On the other hand, the computation developed using neural networks is essentially a self-adaptive distributed method based on a certain learning algorithm. The key point for the success of such an algorithm depends on whether or not the states of neurons converge to a given equilibrium or manifold. For example, if we apply a neural network to solve an optimization problem, then the global convergence of the network corresponds to the global optimal solution of the optimization problem. In this sense, the study on dynamics of neural networks is important for the desired applications. Most of the previous results on dynamics of neural networks have been focused on neural networks with Lipschitz continuous neuron activations.

[☆] Research supported by NSERC Canada.

* Corresponding author. Tel.: +1 519 8884567x36007; fax: +1 519 746 4319.

E-mail addresses: j49liu@uwaterloo.ca (J. Liu), xzliu@uwaterloo.ca (X. Liu), xie@uwaterloo.ca (W.-C. Xie).

The original work by Forti and Nistri [8] (see also [28]) has stimulated some interesting recent work (e.g. [4,8–10,28–30,32,36–39,41]) on global stability of the equilibrium for neural networks with *discontinuous* neuron activations, which provide an ideal model for the case in which the gain of the neuron amplifiers is very high [8] and are of importance and frequently encountered in many applications [18,19,25]. For example, if we consider the classical Hopfield model, where usually it is assumed that the amplifiers have high-gain, the sigmoidal neuron activations would approach a discontinuous hard-comparator function [18,25]. Here, the high-gain hypothesis is crucial to make negligible the contribution to the network energy function of the term depending on the neuron self-inhibitions, and to favor binary output formation, as in a hard-comparator function. Moreover, neural networks were used to solve linear and nonlinear programming problems, where the networks exploited constraint neurons with diode-like input–output activations [23]. In order to guarantee that the constraints are satisfied, the diodes are required to have a very high slope in the conducting region, i.e. they should approximate the discontinuous characteristic of an ideal diode [5]. From a practical point of view, it is often advantageous to model systems possessing high-slope nonlinearities with differential equations with discontinuous right-hand side, rather than studying the case where the slope is high but of finite value [34]. Moreover, from a dynamical system point of view, the discontinuous model reveals many interesting dynamical behaviors that are not captured by the continuous model, e.g. presence of sliding modes along discontinuity surfaces and convergence in finite time [8,9].

In practice, *time-delays* are inevitable in the applications of neural networks due to the finite switching speed of amplifiers and communication time [6,31]. Moreover, in some particular applications, time-delays are deliberately introduced. For example, to process moving images, one must introduce time-delays in the signals transmitted among the cells [6]. Time-delay neural networks can also capture the dynamic nature of speech to achieve superior phoneme recognition performance using standard error back-propagation [22,35]. However, time-delays in a neural network can generally cause problems in stability and also make the stability analysis more involved [1,21,27]. The presence of switching delay in a high-gain neuron amplifier is a particularly harmful source of potential instability [9]. A significant example of this concerns a 3-bit analog-to-digital converter implemented via a Hopfield neural network with high-gain neuron amplifiers [33]. While the network without delay converges due to the symmetry of the neuron interconnection matrix, a more realistic model of the neuron amplifiers has to take into account the presence of a delay in their response. Depending on the properties of delays, there is an onset of limit cycle where the outputs oscillate indefinitely between binary values [33]. This example highlights the need to obtain conditions that can rule out such undesirable oscillations, which are intrinsically related to the combination of high-gain nonlinearities and a delay in the amplifier response [9].

Interesting results have been published on delayed neural networks with discontinuous neuron activations [9,29,30,36,37,39,41]. Forti et al. [9] investigated a class of delayed neuron networks with an arbitrary constant delay and discontinuous activations. Sufficient conditions for global exponential stability and global convergence in finite time of the delayed neural networks are given in terms of M -matrix conditions. Lu and Chen [29] studied dynamical behavior of delayed neural networks with discontinuous neuron activations. Under an easy-to-check assumption in terms of linear matrix inequalities, the authors derived global existence of solutions (viability), existence of equilibrium point, and global asymptotic stability of the delayed neural networks. Wang et al. [36] investigated dynamical behavior of the delayed Hopfield neural networks, where the neuron activation functions are assumed to be discontinuous and non-monotonic. Taking the uncertainties into account, Wang et al. [37] and Zuo et al. [41] investigated the robust stability of delayed neural networks with discontinuous neuron activations.

However, almost all of the above mentioned results deal with only a single constant time-delay. Forti et al. pointed out that it would be interesting to investigate discontinuous neural networks with more general delays, such as time-varying or distributed ones. Actually, while quite often a single constant delay is assumed for simplicity, time-delays encountered in practice are generally not uniform and constant. This motivates us to consider more general types of delays, such as time-varying and distributed ones, which are generally more complex and therefore more difficult to deal with. One of the difficulties encountered here is that the step-by-step construction of local solutions to the discontinuous equations with delays is no longer valid [9] if the delays are time-varying and distributed. More recently, Lu and Chen [30] not only investigated almost periodic dynamics of a general class of neural networks described by delayed integro-differential equations, they also showed that, by constructing a sequence of solutions to delayed dynamical systems with high-slope activations, one can obtain a solution in the sense of Filippov to discontinuous equations with general distributed delays.

The main contribution of this paper is to investigate global stability of delayed neural networks with discontinuous neuron activations and mixed time-varying delays and distributed delays. Inspired by a recent work [30], we find that the difficulty of the existence of solutions for neural networks with discontinuous activations and general mixed time-delays can be overcome, which enables us to further investigate global stability. We then investigate global stability and convergence of this general discontinuous delayed neural network model. Two sets of sufficient conditions are established, one in terms of linear matrix inequalities (LMIs) and the other in terms of M -matrix type conditions. These results extend previous work on global stability of delayed neural networks with Lipschitz continuous neuron activations, and neural networks with discontinuous neuron activations and only constant delays. While global convergence of the networks means the ability to compute the exact global minimum of the underlying energy function and makes these networks desirable for solving global optimization problems, the newly developed paradigm of delayed neural networks with discontinuous neuron activations enable us to apply the results to more difficult tasks, such as non-smooth optimization problems and controller design for systems with non-smooth actuator nonlinearities, and to deal with very general types of delays encountered in applications.

The structure of the paper is outlined as follows. Section 2 discusses the neural network model studied in this paper and present some preliminaries. Section 3 presents the main results on global stability and convergence, in which Section 3.1 provides sufficient conditions in terms of linear matrix inequalities and Section 3.2 gives *M*-matrix type conditions. Section 4 gives two examples to demonstrate the main results. Section 5 provides concluding remarks and possible topics for future work.

2. Neural network model and preliminaries

Consider a class of neural networks described by the system of differential equations

$$\frac{dx_i(t)}{dt} = -d_i x_i(t) + \sum_{j=1}^n a_{ij} g_j(x_j(t)) + \sum_{j=1}^n b_{ij} g_j(x_j(t - \tau_{ij}(t))) + \sum_{j=1}^n c_{ij} \int_0^\infty g_j(x_j(t-s)) p_{ij}(s) ds + I_i, \quad i = 1, \dots, n, \quad t \geq 0, \tag{2.1}$$

where x_i is the state variable of the i th neuron; $d_i > 0$ is the self-inhibition of the i th neuron; a_{ij} is the connection strength of the j th neuron on the i th neuron; b_{ij} and c_{ij} are the delayed feedbacks of the j th neuron on the i th neuron, with time-varying delay and distributed delay, respectively; $\tau_{ij}(t)$ are the time-varying delays; $p_{ij}(s)$ are the probability kernels of the distributed delays; $g_i : \mathbb{R} \rightarrow \mathbb{R}$ represents the neuron input–output activation of the i th neuron; and I_i denotes the external input to the i th neuron. In matrix form, we can write $D = \text{diag}(d_1, \dots, d_n)$, $A = (a_{ij})_{i,j=1}^n$, $B = (b_{ij})_{i,j=1}^n$, $C = (c_{ij})_{i,j=1}^n$, and $I = (I_1, \dots, I_n)^T$. The dynamics of system (2.1) is illustrated in the block diagram shown in Fig. 1, which corresponds to the matrix formation of (2.1). The presence of delayed feedbacks is clearly visible in Fig. 1.

Throughout this paper, we suppose that the activation functions g_i belong to the class \mathcal{G} , where \mathcal{G} denotes the class of functions from \mathbb{R} to \mathbb{R} which are monotonically nondecreasing and have at most a finite number of jump discontinuities in every bounded interval.

Moreover, we assume, throughout this paper, that the *time-varying delays* and the *distributed delays* satisfy that, for $i, j = 1, \dots, n$, $\tau_{ij}(t)$ are continuous functions from $[0, \infty)$ to $[0, \infty)$ such that $0 \leq \tau_{ij}(t) \leq r_{ij} < \infty$, where r_{ij} are some non-negative constants, and each kernel function $p_{ij}(\cdot)$ is a measurable function from $[0, \infty)$ to $[0, \infty)$ with $\int_0^\infty p_{ij}(s) ds = 1$.

Before we introduce the concept of *Filippov solution* for system (2.1), we present some preliminary definitions for later use. Suppose $E \subset \mathbb{R}^n$. Then $x \rightarrow F(x)$ is called a *set-valued map* from $E \rightarrow \mathbb{R}^n$, if for each point $x \in E$, there exists a nonempty set $F(x) \subset \mathbb{R}^n$. A set-valued map F with nonempty values is said to be *upper semicontinuous* at $x_0 \in E$, if for any open set N containing $F(x_0)$, there exists a neighborhood M of x_0 such that $F(M) \subset N$. The map $F(x)$ is said to have a *closed (convex, compact) image* if for each $x \in E$, $F(x)$ is closed (convex, compact). Let $x \in \mathbb{R}^n$ be a vector and $A \in \mathbb{R}^{n \times n}$ be a matrix. We use $\|x\|$ to denote the Euclidean norm of x and $\|A\|$ to denote the matrix norm induced by the Euclidean norm.

Now, for $x \in \mathbb{R}^n$ and $g_i \in \mathcal{G}, i = 1, \dots, n$, we denote by $g(x) = (g_1(x_1), \dots, g_n(x_n))^T$, a diagonal mapping, and denote

$$K[g(x)] = (K[g_1(x_1)], \dots, K[g_n(x_n)]),$$

where $K[g_i(x_i)] = [g_i(x_i^-), g_i(x_i^+)]$.

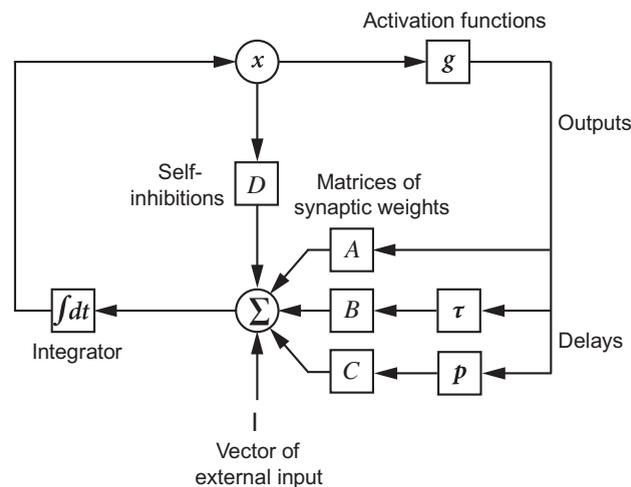


Fig. 1. Block diagram of the neural dynamical system described by Eq. (2.1). Here x denotes the neuron state. Time-varying delays represented by τ and distributed delays denoted by p are both shown in this diagram.

Definition 2.1 (Filippov solutions). Suppose that $\phi = (\phi_1, \dots, \phi_n)^T$ is a continuous function from $(-\infty, 0]$ to \mathbb{R}^n and ψ is a measurable function $\psi = (\psi_1, \dots, \psi_n)^T$ from $(-\infty, 0]$ to \mathbb{R}^n such that $\psi(s) \in K[g(\phi(s))]$ almost everywhere (a.e.) on $(-\infty, 0]$. We say that $x(t) = (x_1(t), \dots, x_n(t))^T$, a function from $(-\infty, T]$ to \mathbb{R}^n , is a *solution* to the initial value problem for (2.1) on $[0, T]$, with initial data (ϕ, ψ) , if

- (i) x is continuous on $(-\infty, T]$ and absolutely continuous on $[0, T]$;
- (ii) there exists a measurable function $\gamma = (\gamma_1, \dots, \gamma_n)^T$ from $(-\infty, T]$ to \mathbb{R}^n such that $\gamma(s) \in K[g(x(s))]$ a.e. on $[0, T]$ and

$$\frac{dx_i(t)}{dt} = -d_i x_i(t) + \sum_{j=1}^n a_{ij} \gamma_j(t) + \sum_{j=1}^n b_{ij} \gamma_j(t - \tau_{ij}(t)) + \sum_{j=1}^n c_{ij} \int_0^\infty \gamma_j(t-s) p_{ij}(s) ds + I_i, \tag{2.2}$$

holds a.e. on $[0, T]$, for all $i = 1, \dots, n$;

- (iii) $x(s) = \phi(s)$ and $\gamma(s) = \psi(s)$ hold for all $s \in (-\infty, 0]$.

Any function as in (2.2) is called an *output* associated with the solution x .

Throughout this paper, the initial functions ϕ and ψ (as described in Definition 2.1) satisfy the following: ϕ is a bounded continuous function from $(-\infty, 0]$ to \mathbb{R}^n and ψ is an essentially bounded measurable function from $(-\infty, 0]$ to \mathbb{R}^n such that $\psi(s) \in K[g(\phi(s))]$ a.e. on $(-\infty, 0]$.

Definition 2.2 (Equilibrium point). We say that a vector $\xi \in \mathbb{R}^n$ is an *equilibrium point* of (2.1) if there exists $\eta \in K[g(\xi)]$ such that

$$0 = -D\xi + (A + B + C)\eta + I,$$

where the vector η is called an *output equilibrium point* corresponding to the equilibrium point ξ .

Remark 2.1. If there exists an equilibrium point of (2.1), without loss of generality, we can assume that 0 is an equilibrium point of system (2.1) with $I = 0$, and 0 is a corresponding output equilibrium point, i.e. $0 \in K[g(0)]$. Actually, if ξ is an equilibrium point of (2.1) and η is the corresponding output equilibrium point, we can always define $y(t) = x(t) - \xi$ and $\mu(t) = \gamma(t) - \eta$, and investigate the resulting system about $y(t)$.

The following lemma provides a useful sufficient condition to check for the existence of an equilibrium point for system (2.1).

Lemma 2.1 (Existence of equilibrium [28]). If $-S$ is a Lyapunov diagonally stable matrix, i.e. there exists a positive definite diagonal matrix P such that $-PS - S^T P$ is positive definite, then there exist $\xi \in \mathbb{R}^n$ and $\eta \in K[g(\xi)]$ such that

$$0 = -D\xi + S\eta + I,$$

where g, D , and I are the same as in system (2.1).

Proof. This important lemma was proved by Lu and Chen [28], using an equilibrium theorem [2]. For the sake of completeness, we present in Appendix A a different and somewhat more direct proof of Lemma 2.1 based on a nonlinear alternative for set-valued maps [12]. □

As an immediate consequence of Lemma 2.1, it is seen that, if $-(A + B + C)$ is Lyapunov diagonally stable, then there exists an equilibrium point for system (2.1).

3. Global stability and convergence

In this section, we present the main results of this paper, which include two sets of sufficient conditions for the global stability and convergence of both the state and the output of system (2.1), in terms of linear matrix inequalities in Section 3.1, and M -matrix type conditions in Section 3.2.

3.1. Sufficient conditions by LMIs

We propose the following linear matrix inequality based on the coefficient matrices A, B , and C of (2.1):

$$Z = \begin{bmatrix} -PA - A^T P - Q - R & -PB & -PC \\ -B^T P & (1 - \rho)Q & 0 \\ -C^T P & 0 & R \end{bmatrix} > 0, \tag{3.1}$$

where P is a positive diagonal matrix, and Q and R are positive definite symmetric matrices.

Remark 3.1. Linear matrix inequalities such as (3.1) can be solved numerically very efficiently using the MATLAB LMI Control toolbox [11]. The role of the inequality proposed in (3.1) is twofold. First, under this assumption, we can show that the matrix $-(A + B + C)$ is Lyapunov diagonally stable, which by Lemma 2.1 implies that there exists an equilibrium point for system (2.1). Second, inequality (3.1) guarantees that we can construct an exponentially decaying Lyapunov function for system (2.1). These will become clearer as we state and prove the main results of this subsection.

The main results of this subsection state that, if the linear matrix inequality (3.1) holds, then there exists an equilibrium point of (2.1) and the equilibrium point is exponentially stable.

Theorem 3.1. *If (3.1) is satisfied, then system (2.1) has an equilibrium point.*

Proof. The matrix inequality (3.1) implies that

$$\begin{bmatrix} -PA - A^T P - Q - R & -PB & -PC \\ -B^T P & Q & 0 \\ -C^T P & 0 & R \end{bmatrix} > 0,$$

which, by Schur's complement (see, e.g. [3]), is equivalent to

$$-PA - A^T P > [PB \quad PC] \begin{bmatrix} Q^{-1} & 0 \\ 0 & R^{-1} \end{bmatrix} \begin{bmatrix} B^T P \\ C^T P \end{bmatrix} + Q + R = PBQ^{-1}B^T P + Q + PCR^{-1}C^T P + R.$$

Since

$$\left[Q^{-\frac{1}{2}} B^T P - Q^{\frac{1}{2}} \right]^T \left[Q^{-\frac{1}{2}} B^T P - Q^{\frac{1}{2}} \right] \geq 0,$$

and

$$\left[R^{-\frac{1}{2}} B^T P - R^{\frac{1}{2}} \right]^T \left[R^{-\frac{1}{2}} B^T P - R^{\frac{1}{2}} \right] \geq 0,$$

we have

$$PBQ^{-1}B^T P + Q \geq PB + B^T P,$$

and

$$PCR^{-1}C^T P + R \geq PC + C^T P.$$

Therefore,

$$-P(A + B + C) - (A + B + C)^T P > 0,$$

i.e. $-(A + B + C)$ is Lyapunov diagonally stable. It follows from Lemma 2.1 that there exists an equilibrium point for system (2.1). The theorem is proved. \square

Note that, in view of Theorem 3.1 and Remark 2.1, we may suppose that $I = 0$ and $0 \in K[g(0)]$, by using a translation $y(t) = x(t) - \xi$ and $\mu(t) = \gamma(t) - \eta$, if necessary. Moreover, in order to write system (2.1) in matrix form and derive stability conditions in terms of linear matrix inequalities, we further assume that $\tau_{ij}(t) \equiv \tau(t)$ and $p_{ij}(t) \equiv p(t)$, for $i, j = 1, \dots, n$, where $0 \leq \tau(t) \leq r < \infty$, $\tau'(t) \leq \rho < 1$, and $\int_0^\infty p(s) ds = 1$, with both r and ρ as non-negative constants. Therefore, a solution $x(t)$ to system (2.1) in the sense of Definition 2.1 satisfies

$$\dot{x}(t) = -Dx(t) + A\gamma(t) + B\gamma(t - \tau(t)) + C \int_0^\infty \gamma(t - s)p(s) ds, \quad (3.2)$$

where $\gamma(t)$ is an output associated with x as in Definition 2.1.

We have the following result on the global stability of system (2.1).

Theorem 3.2. *If (3.1) is satisfied, then the following statements are true.*

- (i) *There exists a solution to system (2.1) on $[0, \infty)$ for any initial data.*
- (ii) *There exist a unique equilibrium point ξ and a unique corresponding output equilibrium point η of (2.1).*
- (iii) *Let $x(t)$ be a solution to system (2.1) on $[0, \infty)$ and $\gamma(t)$ be an associated output. There exist positive constants $M > 0$ and $c > 0$ such that*

$$\|x(t) - \xi\| \leq Me^{-ct}, \quad t \geq 0,$$

i.e. the equilibrium point ξ is globally exponentially stable, and $\gamma(t)$ converge to the output equilibrium point η in measure, i.e. $\lambda - \lim_{t \rightarrow \infty} \gamma(t) = \eta$, where λ is the Lebesgue measure.

The following lemma on linear matrix inequalities is used in the proof of [Theorem 3.2](#) to show that the function $V(t)$ constructed there is exponentially decaying under the matrix inequality (3.1).

Lemma 3.1. *If the matrix inequality (3.1) holds, then there exist a positive constant $\varepsilon < \min_{i \in \{1,2,\dots,n\}} d_i$, a positive diagonal matrix \widehat{P} , and positive definite symmetric matrices \widehat{Q} and \widehat{R} such that*

$$H = \begin{bmatrix} -2D + \varepsilon E_n & A & B & C \\ A^T & \Xi & \widehat{P}B & \widehat{P}C \\ B^T & B^T \widehat{P} & -(1 - \rho)\widehat{Q} & 0 \\ C^T & C^T \widehat{P} & 0 & -\widehat{R} \end{bmatrix} < 0, \tag{3.3}$$

where $\Xi = \widehat{P}A + A^T \widehat{P} + e^{\varepsilon r} \widehat{Q} + e^{\varepsilon r} \widehat{R} + \varepsilon E_n$ and E_n is the n -dimensional identity matrix.

Proof. See [Appendix B](#). □

The following lemma generalizes a useful integral inequality [13], and will be used in the proof of [Theorem 3.2](#) to deal with the distributed delay terms of (2.1).

Lemma 3.2. *Let u be a measurable function defined on $[0, \infty)$ and R be a positive definite symmetric matrix. Suppose that $p : [0, \infty) \rightarrow [0, \infty)$ is measurable and $\int_0^\infty p(s)ds = 1$. We have*

$$\int_0^\infty u^T(s)Ru(s)p(s)ds \geq \left(\int_0^\infty u(s)p(s)ds \right)^T R \int_0^\infty u(s)p(s)ds.$$

Proof. See [Appendix C](#). □

Now we are ready to present the proof of [Theorem 3.2](#).

Proof of Theorem 3.2. For proof of part (i), see [Appendix D](#). The existence of an equilibrium point and an output equilibrium point follows from [Theorem 3.1](#) and the uniqueness follows from part (iii). It remains to show part (iii). Let $x(t)$ be a solution to (2.1) on $[0, \infty)$ and $\gamma(t)$ is an associated output. According to [Remark 2.1](#) and the remark after [Theorem 3.1](#), we can assume both $\xi = 0$, $\eta = 0$, and $x(t)$ obeys (3.2). Consider

$$V(t) = e^{\varepsilon t} x^T(t)x(t) + 2 \sum_{i=1}^n e^{\varepsilon t} \widehat{P}_i \int_0^{x_i(t)} g_i(s)ds + \int_{t-\tau(t)}^t \gamma^T(s)\widehat{Q}\gamma(s)e^{\varepsilon(s+r)}ds + \int_0^\infty \int_{t-s}^t \gamma^T(\theta)\widehat{R}\gamma(\theta)p(s)d\theta ds, \quad t \geq 0, \tag{3.4}$$

where ε , \widehat{P} , \widehat{Q} , and \widehat{R} are given by [Lemma 3.1](#). Here $V(t)$ plays the role of a Lyapunov function candidate, which will be used to bound both the norm of the state $x(t)$ and the output $\gamma(t)$. The main step of proof is to show that $V(t)$ is exponentially decaying under the linear matrix inequality (3.1). Differentiating $V(t)$ according to (3.2) and the generalized chain rule (see [7]; see also [9] or [28] for details), we obtain

$$\begin{aligned} \frac{dV(t)}{dt} &= \varepsilon e^{\varepsilon t} x^T(t)x(t) + 2e^{\varepsilon t} x^T(t) \left[-Dx(t) + A\gamma(t) + B\gamma(t - \tau(t)) + C \int_0^\infty \gamma(t-s)p(s)ds \right] \\ &\quad + 2e^{\varepsilon t} \gamma(t)\widehat{P} \left[-Dx(t) + A\gamma(t) + B\gamma(t - \tau(t)) + C \int_0^\infty \gamma(t-s)p(s)ds \right] + 2\varepsilon e^{\varepsilon t} \sum_i^n \widehat{P}_i \int_0^{x_i(t)} g_i(s)ds + e^{\varepsilon(t+r)} \gamma^T(t)\widehat{Q}\gamma(t) \\ &\quad - e^{\varepsilon(t-\tau(t)+r)} (1 - \tau'(t)) \gamma^T(t - \tau(t))\widehat{Q}\gamma(t - \tau(t)) + e^{\varepsilon t} \int_0^\infty \gamma^T(t)\widehat{R}\gamma(t)p(s)ds - e^{\varepsilon t} \int_0^\infty \gamma^T(t-s)\widehat{R}\gamma(t-s)p(s)ds. \end{aligned}$$

By $\varepsilon < \min_{i \in \{1,2,\dots,n\}} d_i$, $0 \in K[g_i(0)]$, and the monotone property of g_i , we have

$$\varepsilon \int_0^{x_i(t)} g_i(s)ds \leq \varepsilon x_i(t)\gamma_i(t) \leq d_i x_i(t)\gamma_i(t).$$

Using $\tau'(t) \leq \rho < 1$ and [Lemma 3.2](#), we have

$$\frac{dV(t)}{dt} \leq e^{\varepsilon t} z^T H z - \varepsilon \gamma^T(t)\gamma(t) \leq -\varepsilon \gamma^T(t)\gamma(t) \leq 0, \tag{3.5}$$

where H is a negative definite matrix given by [Lemma 3.1](#) and

$$z = \begin{bmatrix} x^T(t) & \gamma^T(t) & \gamma^T(t - \tau(t)) & \int_0^\infty \gamma^T(t-s)p(s)ds \end{bmatrix}^T.$$

It follows that $V(t)$ is nondecreasing on $[0, \infty)$ and

$$V(t) \leq V(0), \quad t \geq 0. \quad (3.6)$$

In view of the definition of $V(t)$ in (3.4) and the fact that all the terms in $V(t)$ are non-negative, we have

$$V(t) \geq e^{\epsilon t} x^T(t)x(t), \quad t \geq 0. \quad (3.7)$$

Combining (3.6) and (3.7) gives

$$\|x(t)\| \leq \sqrt{V(t)}e^{-\frac{\epsilon}{2}t} \leq \sqrt{V(0)}e^{-\frac{\epsilon}{2}t}, \quad t \geq 0.$$

We proceed to show the convergence of output. From (3.5), we have

$$V(t) - V(0) \leq -\epsilon \int_0^t \|\gamma(s)\|^2 ds.$$

Since $V(t) \geq 0$ for all $t \geq 0$, we have

$$\int_0^\infty \|\gamma(s)\|^2 ds \leq \frac{1}{\epsilon} V(0).$$

For any $\epsilon_0 > 0$, let $E_{\epsilon_0} = \{t \in [0, \infty) : \|\gamma(t)\| \geq \epsilon_0\}$. Then

$$\frac{V(0)}{\epsilon} \geq \int_0^\infty \|\gamma(s)\|^2 ds \geq \int_{E_{\epsilon_0}} \|\gamma(s)\|^2 ds \geq \epsilon_0^2 \lambda(E_{\epsilon_0}),$$

where $\lambda(\cdot)$ is the Lebesgue measure. It follows that $\lambda(E_{\epsilon_0}) < \infty$. Therefore, $\gamma(t)$ converges to 0 in measure [8, Proposition 2], i.e. $\lambda - \lim_{t \rightarrow \infty} \gamma(t) = 0$. The proof is complete. \square

Remark 3.2. If there are no distributed delays (i.e. $C = 0$) and the time-varying delay reduces to a constant delay (i.e. $\tau(t) = \tau$ and $\rho = 0$), then it can be seen that Theorems 3.1 and 3.2 include some known results in the literature [29, Theorems 1, 4, 5] as corollaries.

3.2. Sufficient conditions by M -matrix

The M -matrix type conditions allow us to deal with more general types of mixed delays, i.e. τ_{ij} and p_{ij} can be non-identical. We assume that there exist a positive constant $\alpha > 0$ and constants $\rho_{ij} \in [0, 1)$ such that

$$\int_0^\infty e^{\alpha s} p_{ij}(s) ds < \infty, \quad \tau'_{ij}(t) \leq \rho_{ij} < 1, \quad (3.8)$$

for all $i, j = 1, \dots, n$ and $t \geq 0$.

In order to propose some M -matrix type conditions on the coefficient matrices A , B , and C of (2.1), we first introduce the following notations. We define $\mathcal{M}(A) = (\hat{a}_{ij})_{i,j=1}^n$, where $\hat{a}_{ij} = |a_{ii}|$ for $i = j$ and $\hat{a}_{ij} = -|a_{ij}|$ for $i \neq j$, i.e. $\mathcal{M}(A)$ is the comparison matrix of A (see [20]). Let $|C|$ denote the matrix obtained by taking entrywise absolute values of the entries in C . Define $\hat{B} = (\hat{b}_{ij})_{i,j=1}^n$, where $\hat{b}_{ij} = \frac{|b_{ij}|}{1 - \rho_{ij}}$.

Assumption 3.1 (*M -matrix condition*). The matrix $\Theta = \mathcal{M}(A) - \hat{B} - |C|$ is an M -matrix, i.e. all successive principal minors of Θ are positive, and $a_{ii} < 0$ for $i = 1, \dots, n$.

The main results of this subsection state that, if the above condition holds, then there exists an equilibrium point of (2.1) and the equilibrium point is exponentially stable.

According to the theory of M -matrices (see [20]), Assumption 3.1 is equivalent to that there exists a vector $\beta = (\beta_1, \dots, \beta_n) > 0$ such that, for $i = 1, \dots, n$, we have

$$\beta_i a_{ii} + \sum_{\substack{j=1 \\ j \neq i}}^n \beta_j |a_{ji}| + \sum_{j=1}^n \frac{\beta_j |b_{ji}|}{1 - \rho_{ij}} + \sum_{j=1}^n \beta_j |c_{ji}| < 0. \quad (3.9)$$

In view of (3.8) and (3.9), if Assumption 3.1 holds, we can choose a constant $\delta \in (0, \min_{i \in \{1, 2, \dots, n\}} d_i)$ sufficiently small such that the following holds for $i = 1, \dots, n$.

$$\beta_i a_{ii} + \delta + \sum_{\substack{j=1 \\ j \neq i}}^n \beta_j |a_{ji}| + \sum_{j=1}^n \beta_j |b_{ji}| \frac{e^{\delta \tau_{ij}}}{1 - \rho_{ij}} + \sum_{j=1}^n \beta_j |c_{ji}| \int_0^\infty e^{\delta s} p_{ji}(s) ds < 0. \quad (3.10)$$

Actually, if we take the limit of the left-hand side of (3.10) as $\delta \rightarrow 0$, we get exactly the left-hand side of (3.9), since

$$\delta \rightarrow 0, \quad e^{\delta \tau_{ij}} \rightarrow 1, \quad \int_0^\infty e^{\delta s} p_{ji}(s) ds \rightarrow 1,$$

where the last limit follows from the Lebesgue dominated convergence theorem in view of (3.8) and the fact that $\int_0^\infty p_{ij}(s)ds = 1$. While it is equivalent to Assumption 3.1 and inequality (3.9), inequality (3.10) is the one to be used in the proofs of the following results. The idea of proof is similar to that of Theorems 3.1 and 3.2. We shall construct a Lyapunov function candidate $V(t)$ and show that $V(t)$ is exponentially decaying under inequality (3.10) (or, equivalently, Assumption 3.1).

Theorem 3.3. *If Assumption 3.1 is satisfied, then system (2.1) has an equilibrium point.*

Proof. Since $0 < 1 - \rho_{ij} \leq 1$, inequality (3.9) implies that

$$\beta_i a_{ii} + \sum_{\substack{j=1 \\ j \neq i}}^n \beta_j |a_{ji}| + \sum_{j=1}^n \beta_j (|b_{ji}| + |c_{ji}|) < 0, \quad i = 1, \dots, n,$$

which, by $a_{ii} < 0$, implies that

$$\beta_i |a_{ii} + b_{ii} + c_{ii}| \geq \beta_i (-a_{ii} - |b_{ii}| - |c_{ii}|) > \sum_{\substack{j=1 \\ j \neq i}}^n \beta_j (|a_{ji}| + |b_{ji}| + |c_{ji}|) \geq \sum_{\substack{j=1 \\ j \neq i}}^n \beta_j |a_{ji} + b_{ji} + c_{ji}|, \quad i = 1, \dots, n.$$

This shows that the matrix $-(A + B + C)$ is an H -matrix, i.e. the comparison matrix of $-(A + B + C)$ is an M -matrix. Since the diagonal entries of $-(A + B + C)$, given by $-a_{ii} - b_{ii} - c_{ii}$, are positive, we can conclude that $-(A + B + C)$ is Lyapunov diagonally stable [20]. It follows from Lemma 2.1 that there exists an equilibrium point for system (2.1). The proof is complete. \square

Theorem 3.4. *If Assumptions 3.1 is satisfied, then the same statements as in Theorem 3.2 hold.*

Proof of Theorem 3.4. For proof of part (i), see Appendix D. The existence of an equilibrium point and an output equilibrium point follows from Theorem 3.3 and the uniqueness follows from part (iii). It remains to show part (iii). Let $x(t)$ be a solution to (2.1) on $[0, \infty)$ and $\gamma(t)$ is an associated output. Same as before, we can assume both $\xi = 0, \eta = 0$, and $x(t)$ obeys (2.1) with $l = 0$. Consider

$$V(t) = \sum_{i=1}^n \beta_i |x_i(t)| e^{\delta t} + \sum_{i,j=1}^n \frac{\beta_i |b_{ij}|}{1 - \rho_{ij}} \int_{t-\tau_{ij}(t)}^t |\gamma_j(s)| e^{\delta(s+r_{ij})} ds + \sum_{i,j=1}^n \beta_i |c_{ij}| \int_0^\infty \int_{t-s}^t |\gamma_j(\theta)| e^{\delta(s+\theta)} d\theta ds, \quad t \geq 0,$$

where β and δ are from (3.10). Differentiating $V(t)$ gives

$$\begin{aligned} \frac{dV(t)}{dt} &= \sum_{i=1}^n \delta \beta_i |x_i(t)| e^{\delta t} \\ &+ \sum_{i=1}^n \beta_i e^{\delta t} \text{sign}(x_i(t)) \left(-d_i x_i(t) + \sum_{j=1}^n a_{ij} \gamma_j(t) + \sum_{j=1}^n b_{ij} g_j(x_j(t - \tau_{ij}(t))) + \sum_{j=1}^n c_{ij} \int_0^\infty g_j(x_j(t-s)) p_{ij}(s) ds \right) + \sum_{i,j=1}^n \\ &\times \frac{\beta_i |b_{ij}| e^{\delta t}}{1 - \rho_{ij}} \left[|\gamma_j(t)| e^{\delta r_{ij}} - (1 - \tau'(t)) |\gamma_j(t - \tau_{ij}(t))| \right] + \sum_{i,j=1}^n \beta_j |c_{ij}| e^{\delta t} \int_0^\infty \left[|\gamma_j(t)| e^{\delta s} - |\gamma_j(t-s)| \right] ds \\ &\leq \sum_{i=1}^n \beta_i |x_i(t)| e^{\delta t} (-d_i + \delta) + \sum_{i=1}^n e^{\delta t} |\gamma_i(t)| \left(\beta_i a_{ii} + \sum_{\substack{j=1 \\ j \neq i}}^n \beta_j |a_{ji}| + \sum_{j=1}^n \beta_j |b_{ji}| \frac{e^{\delta r_{ij}}}{1 - \rho_{ij}} + \sum_{j=1}^n \beta_j |c_{ji}| \int_0^\infty e^{\delta s} p_{ji}(s) ds \right). \end{aligned}$$

By the inequality in Remark 3.10, it follows that

$$\frac{dV(t)}{dt} \leq -\delta \sum_{i=1}^n |\gamma_i(t)| \leq 0. \tag{3.11}$$

Therefore, $V(t) \leq V(0)$ and

$$\sum_{i=1}^n \beta_i |x_i(t)| \leq V(0) e^{-\delta t}, \quad t \geq 0.$$

Since $\beta > 0$ and all norms in \mathbb{R}^n are equivalent, there exists a positive constant $c_1 > 0$ such that

$$\|x(t)\| \leq c_1 V(0) e^{-\delta t}, \quad t \geq 0.$$

We proceed to show the convergence of output. From (3.11) and that all norms in \mathbb{R}^n are equivalent, we can find a constant c_2 such that

$$V(t) - V(0) \leq -c_2 \int_0^t \|\gamma(s)\| ds.$$

The rest of the proof is similar to that of Theorem 3.2. \square

Remark 3.3. If there are no distributed delays (i.e. $C = 0$) and the time-varying delays reduce to a constant delay (i.e. $\tau_{ij}(t) = \tau$ and $\rho_{ij} = 0$), then Assumption 3.1 becomes that $\Theta = \mathcal{M}(A) - |B|$ is an M -matrix and $a_{ii} < 0, i = 1, \dots, n$, which is exactly the same assumption as proposed by Forti et al. [9] for global convergence of (constantly) delayed neural networks with discontinuous neuron activations. Therefore, the convergence results by Forti et al. [9] can be regarded as corollaries of Theorems 3.3 and 3.4.

4. Examples

Example 4.1. Consider the second-order neural network (2.1) with $D = \text{diag}(0.01, 0.01)$ and

$$A = \begin{bmatrix} -1.5 & -0.1 \\ 0.1 & -1.5 \end{bmatrix}, \quad B = \begin{bmatrix} 0.4 & -0.4 \\ 0.35 & -0.35 \end{bmatrix}.$$

Suppose $C = 0$ (i.e. no distributed delays). Let $\tau_{11}(t) = \tau_{12}(t) = 5 - \sin(t/4)$, $\tau_{21}(t) = \tau_{22}(t) = 5 - \cos(t/4)$, $I = [6 \ 10]^T$, and $g_1(s) = g_2(s) = s + \text{sign}(s)$. Hence $\tau'_{ij}(t) \leq \rho = 1/4$. It can be easily verified that Assumption 3.1 is satisfied. Consider the IVP of (2.1) with initial conditions $\phi(t) = [5\cos(10t), -5\cos(10t)]$ for $t \in [-6, 0]$, and $\psi(t) = [g_1(5\cos(10t)), g_2(-5\cos(10t))]$ for $t \in [-6, 0]$. Fig. 2 shows that both the simulated state $x(t)$ and output $\gamma(t)$ converges to the unique equilibrium point, which is in accordance with the conclusions of Theorem 3.4.

Example 4.2. Consider Example 4.1 with distributed delays given by

$$C = \begin{bmatrix} -0.2 & -0.1 \\ 0.1 & -0.25 \end{bmatrix}, \quad p_{ij}(s) = \begin{cases} \frac{1}{6}, & s \in [0, 6], \\ 0, & s \in (6, \infty). \end{cases}$$

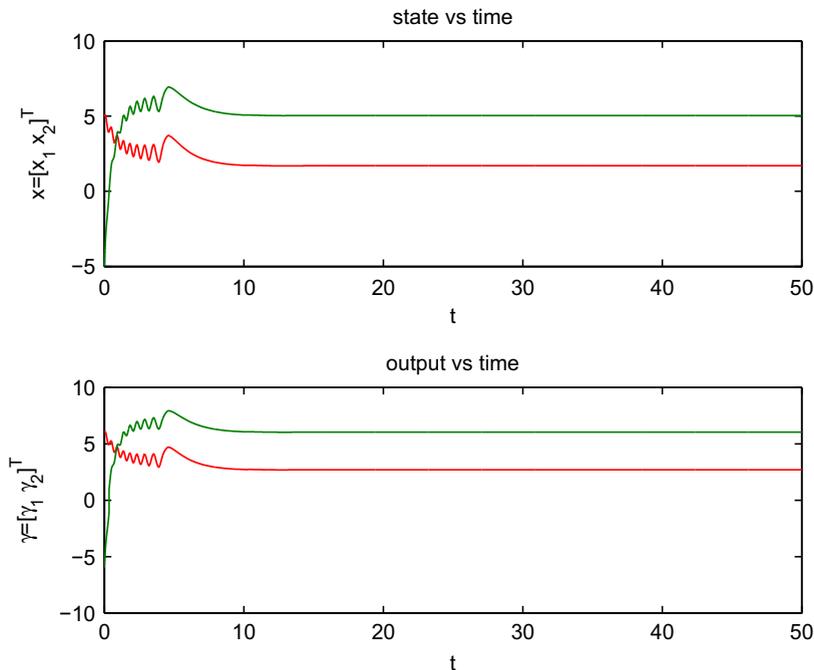


Fig. 2. Simulation results for Example 4.1.

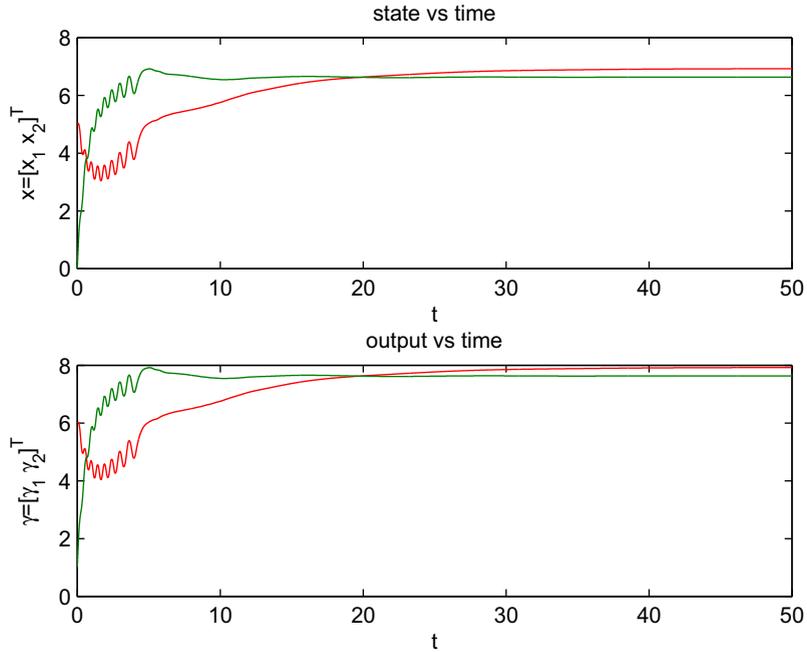


Fig. 3. Simulation results for Example 4.2.

Let $\tau_{ij}(t) = \tau(t) = 5 - \sin(t/2)$, $I = [6 \ 10]^T$, and $g_1(s) = g_2(s) = s + \text{sign}(s)$. Hence $\tau'_{ij}(t) \leq \rho = 1/2$. It can be seen that Assumption 3.1 (M -matrix condition) no longer holds. However, using the MATLAB LMI control toolbox [11], we can check that the matrix inequality (3.1) is satisfied. Under the same initial conditions as in Example 4.1, Fig. 3 shows that the simulated state $x(t)$ and the output $\gamma(t)$ converge to the unique equilibrium point $\xi = [6.85126.6327]^T$ and the unique output equilibrium point $\eta = g(\xi) = [7.85127.6327]$, respectively, which is in accordance with the conclusions of Theorem 3.2. Here the convergence of the output is in the usual sense since $g(x)$ is continuous at $\xi = [6.85126.6327]^T$. With input $I = [0.65 - 1.2]^T$, however, it can be shown that the unique equilibrium is $\xi = [00]^T$ and the unique output equilibrium $\eta = [1 - 1]^T$. In this case $g(x)$ is

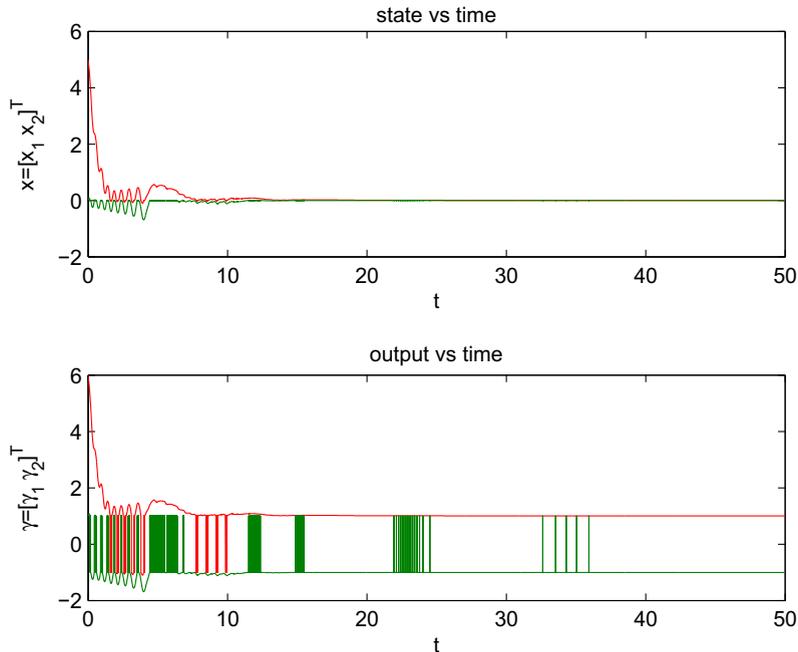


Fig. 4. Simulation results for Example 4.2.

discontinuous at $\xi = [00]^T$ and the convergence of output to η is in the measure sense. Fig. 4 shows the corresponding simulations.

5. Conclusions

In this paper, we have investigated the dynamical behavior of a class of neural networks with mixed time-delays and discontinuous neuron activations. Both time-varying delays and distributed delays have been considered, while almost all of the recent work on neural networks with discontinuous neuron activations considered only constant delays. We have established two sets of sufficient conditions for both the global exponential stability of the neuron state and global convergence of the neuron output. These results extend previous work on global stability of delayed neural networks with Lipschitz continuous neuron activations, and neural networks with discontinuous neuron activations and only constant delays.

We conclude by pointing out a few questions not answered in this paper. Important notions of convergence in finite time for both the neuron state and neuron output have been investigated and interesting results have been reported [8,9]. One of the key points is that, after the neuron state converges to the equilibrium point in finite time, the output can be shown to obey an algebraic memoryless equation [9]. As for neuron networks with time-varying delays and distributed delays, the relation can get significantly more involved. In addition, uniqueness of solution (both the neuron state and output) is closely related to the convergence in finite time [9]. In this paper, uniqueness of solution is not discussed. However, we have shown that, under the theorem conditions, all solutions will converge to the unique equilibrium point.

Furthermore, it remains an interesting topic whether the model of discontinuous neural networks can be extended to a stochastic one, so that the stochastic model can cover the situation with random noise. Noise can arise from many sources and can be treated differently. For example, if the noise is vanishing at the system equilibrium and the intensity is sufficiently small, then global stability properties, such as almost sure and p th moment stability, could still be possible. With non-vanishing noise, one can still study certain stochastic convergence properties (in terms of probability density functions) analogous to ultimate boundedness in the deterministic setting. If the input is affected by noise, one may not be able to have the same (global) stability properties as obtained in the current paper. Particularly, a noisy input can affect the equilibrium of the system, and therefore, even if the system still converges, it may converge to an inaccurate state if the neural network model were used for computation. To reduce noisy effects so that the neural network can converge to an accurate result, noisy reduction and/or feature selection methods can be applied to preprocess the input data [26]. The particular challenge in extending the discontinuous model to a stochastic setting is that one has to develop the theory of stochastic differential equations with discontinuous right-hand sides, possibly following some earlier work [24] on this topic.

Acknowledgment

The research for this paper was supported by the Natural Sciences and Engineering Research Council of Canada. The authors are grateful to the anonymous reviewers for their valuable comments and suggestions, which have greatly helped to improve the quality of this paper.

Appendix A. Proof of Lemma 2.1

The proof we present here is based on the following version of nonlinear alternative for set-valued maps.

Lemma A.1 (Nonlinear alternative [12]). *Let C be a convex set in a normed linear space, and let $U \subset C$ be open with $0 \in U$. Let $\mathcal{K}(\bar{U}, 2^C)$ denote the set of nonempty compact convex upper semicontinuous set-valued maps from the closure of U to C . Then each $F \in \mathcal{K}(\bar{U}, 2^C)$ has at least one of the following properties:*

- (a) F has a fixed point in \bar{U} , i.e. there exists some $x_0 \in \bar{U}$ such that $x_0 \in F(x_0)$;
- (b) there exists $x \in \partial U$ (the boundary of U) and $\bar{\lambda} \in (0, 1)$ such that $x \in \bar{\lambda}F(x)$.

Now we are ready to prove Lemma 2.1.

Proof of Lemma 2.1. It is easy to see that the conclusion of Lemma 2.1 is equivalent to the statement that there exists $\xi \in \mathbb{R}^n$ such that

$$\xi \in D^{-1}\{SK[g(\xi)] + I\}, \quad (\text{A.1})$$

i.e. the set-valued map $G(x)$, defined by $G(x) = D^{-1}\{S K[g(x)] + I\}$ for $x \in \mathbb{R}^n$, has a fixed point. Consider two cases:

- (i) Suppose that G , as a set-valued map from \mathbb{R}^n to \mathbb{R}^n , is bounded. Therefore, there exists an open ball $B(0, r)$, centered at 0 and with radius $r > 0$, such that $G(x) \subset B(0, r)$ for all $x \in \mathbb{R}^n$. We let $U = B(0, r)$. It is clear that the second alternative in Lemma A.1 does not hold for G and U , and, therefore, $G(x)$ has a fixed point by Lemma A.1.

(ii) Suppose that G is unbounded. It follows that g must be unbounded. Since g_i are class \mathcal{G} functions, there must exist $i \in \{1, \dots, n\}$ such that

$$\text{either } \lim_{s \rightarrow \infty} g_i(s) = \infty \text{ or } \lim_{s \rightarrow \infty} g_i(s) = -\infty.$$

Without loss of generality, we shall assume that $\lim_{s \rightarrow \infty} g_i(s) = \infty$. Let

$$a = \Lambda_{\min}(-PS - S^T P) > 0, \quad b = \frac{2\|PI\|^2}{a} > 0,$$

where $\Lambda_{\min}(\cdot)$ represents the minimum real eigenvalue of a matrix. Without loss of generality, we can assume $0 \in K[g(0)]$. Actually, if this is not the case, we can define $\tilde{g} = g - \tilde{\eta}$, where $\tilde{\eta} \in g(0)$, and $\tilde{I} = I + S\tilde{\eta}$. Then, we have $0 \in K[\tilde{g}(0)]$, and \tilde{g} still satisfies $\lim_{s \rightarrow \infty} g_i(s) = \infty$ and $G(x) = D^{-1}\{SK[\tilde{g}(x)] + \tilde{I}\}$. Now, since $0 \in K[g(0)]$ and g_i is in class \mathcal{G} , it follows that

$$\eta^T P D x \geq 0, \quad \forall x \in \mathbb{R}^n, \quad \eta \in K[g(x)]. \tag{A.2}$$

Since $\lim_{s \rightarrow \infty} g_i(s) = \infty$, it follows that there exists $c > 0$ such that $x_i \geq c$ implies $g_i(x_i^-) > \sqrt{\frac{2b}{a}}$. Let $U = \{x \in \mathbb{R}^n : x_i < c\}$. We proceed to show that, with this choice of U , the second alternative of Lemma A.1 does not hold for G . We show this by contradiction. Suppose there exists $x \in \partial U$, i.e. $x_i = c$, and $\bar{\lambda} \in (0, 1)$ such that $x \in \bar{\lambda}G(x)$, i.e. there exists $\eta \in K[g(x)]$ such that

$$0 = -Dx + \bar{\lambda}(S\eta + I). \tag{A.3}$$

It follows from (A.2) and (A.3) that

$$0 = 2\eta^T[-Dx + \bar{\lambda}(S\eta + I)] \leq -\bar{\lambda}\eta^T(-PS - S^T P)\eta + 2\bar{\lambda}\eta^T P I \leq -a\bar{\lambda}\|\eta\|^2 + \frac{a}{2}\bar{\lambda}\|\eta\|^2 + \frac{2}{a}\bar{\lambda}\|PI\|^2 \leq \bar{\lambda}\left(-\frac{a}{2}\|\eta\|^2 + b\right), \tag{A.4}$$

where $\eta_i \in [g_i(x_i^-), g_i(x_i^+)]$. By the choice of c , it follows that $-\frac{a}{2}\|\eta\|^2 + b < 0$, which contradicts (A.4). Therefore, $G(x)$ has a fixed point by Lemma A.1.

The proof is now complete. \square

Appendix B. Proof of Lemma 3.1

Choose $x, y, z, w \in \mathbb{R}^n$ and $\hat{P} = \theta P, \hat{Q} = \theta Q, \hat{R} = \theta R$, where $P, Q,$ and R are from the matrix inequality (3.1) and $\theta > 0$, together with $\varepsilon > 0$ in (3.3), is to be determined later in the proof. We have

$$\begin{aligned} [x^T \ y^T \ z^T \ w^T]H[x^T \ y^T \ z^T \ w^T]^T &= -2x^T D x + \varepsilon x^T x + 2x^T A y + 2x^T B z + 2x^T C w + \theta y^T (P A + A^T P) y + \theta e^{\varepsilon r} y^T Q y \\ &\quad + \theta e^{\varepsilon r} y^T R y + \varepsilon y^T y + 2\theta y^T P B z + 2\theta y^T P C w - (1 - \rho)\theta z^T Q z - \theta w^T R w \\ &= -\theta[y^T \ z^T \ w^T]Z[y^T \ z^T \ w^T]^T - 2x^T D x + \varepsilon x^T x + 2x^T A y + 2x^T B z + 2x^T C w \\ &\quad + \theta(e^{\varepsilon r} - 1)y^T Q y + \theta(e^{\varepsilon r} - 1)y^T R y + \varepsilon y^T y, \end{aligned}$$

where Z is from (3.1). Let $\alpha = \Lambda_{\min}(Z)$, the minimum real eigenvalue of Z , and $d = \min_{i \in \{1, 2, \dots, n\}} d_i$, the above equation gives

$$\begin{aligned} [x^T \ y^T \ z^T \ w^T]H[x^T \ y^T \ z^T \ w^T]^T &\leq -\theta\alpha(y^T y + z^T z + w^T w) - (2d - 4\varepsilon)x^T x \\ &\quad + \left\{ \theta(e^{\varepsilon r} - 1)\|Q\| + \theta(e^{\varepsilon r} - 1)\|R\| + \varepsilon + \varepsilon^{-1}\|A\|^2 \right\} y^T y + \varepsilon^{-1}\|B\|^2 z^T z \\ &\quad + \varepsilon^{-1}\|C\|^2 w^T w. \end{aligned} \tag{B.1}$$

Now choose the positive constant $\varepsilon < d$ (as required by Lemma 3.1) sufficiently small such that

$$2d - 4\varepsilon > 0, \quad (e^{\varepsilon r} - 1)\|Q\| < \alpha, \quad (e^{\varepsilon r} - 1)\|R\| < \alpha.$$

Fix this choice of $\varepsilon > 0$ and let $\theta > 0$ be sufficiently large such that

$$\theta\alpha > \theta(e^{\varepsilon r} - 1)\|Q\| + \theta(e^{\varepsilon r} - 1)\|R\| + \varepsilon + \varepsilon^{-1}\|A\|^2,$$

and

$$\theta\alpha > \varepsilon^{-1}\|B\|^2, \quad \theta\alpha > \varepsilon^{-1}\|C\|^2.$$

Then inequality (B.1) shows that H is negative definite. Lemma 3.1 is proved.

Appendix C. Proof of Lemma 3.2

Applying the Cauchy–Schwartz inequality, we have

$$\begin{aligned} \left(\int_0^\infty u(s)p(s)ds \right)^T R \int_0^\infty u(s)p(s)ds &= \left\| \int_0^\infty R^{\frac{1}{2}}u(s)\sqrt{p(s)}\sqrt{p(s)}ds \right\|^2 \leq \int_0^\infty \|R^{\frac{1}{2}}u(s)\sqrt{p(s)}\|^2 ds \int_0^\infty p(s)ds \\ &= \int_0^\infty u^T(s)Ru(s)p(s)ds. \end{aligned}$$

Lemma 3.2 is proved.

Appendix D. Local Existence of Solutions (Sketch of Proof for Part (i) of Theorems 3.2 and 3.4)

Following the idea of Haddad [14], Lu and Chen [29,30] showed that, by constructing a sequence of delay differential equations with high-slope right-hand sides, one can obtain a Filippov solution of delayed neural networks with discontinuous activations. The same approach enables us to obtain a solution to system (2.1) in the sense of Definition 2.1, i.e. one can first construct a sequence of delay differential equations with high-slope right-hand sides and then prove that the solutions approach a solution of (2.1) in the sense of Definition 2.1. The approach is essentially similar to that by Lu and Chen [29,30], except that here we consider both time-varying delays and distributed delays. We shall only sketch the main steps and interested readers can refer to the papers by Lu and Chen [29,30] for more details.

Sketch of proof for part (i) of Theorems 3.2 and 3.4 Throughout the proof, we shall let $g_i, i = 1, 2, \dots, n$, be fixed class \mathcal{G} functions and ϕ and ψ be initial functions satisfying the assumptions in Definition 2.1. The norms of ϕ and ψ are given by the supremum and essential supremum norm, respectively.

Step 1 (construction of high-slope functions): Let $\{Q_{k,i}\}$ be the set of discontinuous points of g_i . Pick a strictly decreasing $\{\delta_{k,i,m}\}$ with $\lim_{m \rightarrow \infty} \delta_{k,i,m} = 0$, and let $J_{k,i,m} = [Q_{k,i} - \delta_{k,i,m}, Q_{k,i} + \delta_{k,i,m}]$ be intervals such that $J_{k_1,i,m} \cap J_{k_2,i,m} = \emptyset$ for $k_1 \neq k_2$. Define $g_i^m(\cdot)$ by letting $g_i^m(s) = g_i(s)$ if $s \notin J_{k,i,m}$ for any k , and

$$g_i^m(s) = \frac{g_i(+)-g_i(-)}{2\delta_{k,i,m}}(s - Q_{k,i} - \delta_{k,i,m}) + g_i(+),$$

where, for the sake of simplicity, $g_i(+)=g_i(Q_{k,i} + \delta_{k,i,m})$ and $g_i(-)=g_i(Q_{k,i} - \delta_{k,i,m})$. We can observe the following properties of the sequence of functions $\{g^m(\cdot)\}$:

- (i) each function $g^m : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is a diagonal mapping, i.e. $g^m(x) = [g_1^m(x_1), \dots, g_n^m(x_n)]^T$, and g_i^m is nondecreasing and continuous, for $i = 1, \dots, n$;
- (ii) for each compact set $W \subset \mathbb{R}^n$, there exists a constant $C = C(W) > 0$, independent of the choice of g^m in the sequence, such that $|g_i^m(x)| \leq C$ holds for all $x \in W$ and $i = 1, \dots, n$;
- (iii) we have

$$\lim_{m \rightarrow \infty} d_H(\text{Graph}(g^m(S)), \text{Graph}(K[g(S)])) = 0,$$

for all $S \subset \mathbb{R}^n$, where $\text{Graph}(F(E))$ is defined by

$$\text{Graph}(F(E)) = \{(x, y) \mid x \in E, y \in F(x)\},$$

for a set-valued map F and a subset $E \in \mathbb{R}^n$, and $d_H(A, B)$ denotes the Hausdorff metric of two sets in a Euclidean space defined by

$$d_H(A, B) = \sup_{x \in A} \inf_{y \in B} \|x - y\|.$$

Step 2 (local solutions to DDEs with high-slope RHS): for each g^m , consider the following system of delay differential equations

$$\begin{aligned} \frac{dy_i(t)}{dt} &= -d_i y_i(t) + \sum_{j=1}^n a_{ij} g_j^m(y_j(t)) + \sum_{j=1}^n b_{ij} y_j g_j^m(y_j(t - \tau_{ij}(t))) + \sum_{j=1}^n c_{ij} \int_0^\infty g_j^m(y_j(t-s)) p_{ij}(s) ds + I_i, \quad i \\ &= 1, \dots, n, \quad t \geq 0, \end{aligned} \tag{B.2}$$

with $y_i(s) = \phi_i(s)$ for all $s \in (-\infty, 0]$. Following the classical theory of functional differential equations (see [15,17]) and based on the properties observed above for the sequence $\{g^m\}$, there exists an $\alpha > 0$ such that, for any g^m in the sequence, there is a solution to (B.2) on $[0, \alpha]$. Therefore, we obtain a sequence of solutions $\{y^m(t)\}$ to (B.2) on $[0, \alpha]$.

Step 3 (boundedness of solutions to (B.2) and continuation): using a similar Lyapunov functional approach as in the proof of Theorems 3.2 and 3.4, under the matrix inequality (3.1) and Assumption 3.1, respectively, one can show that the solutions

$\{y^m(t)\}$ are uniformly bounded and hence they can be extended to the interval $[0, \infty)$ [15]. Therefore, we obtain a sequence of solutions $\{y^m(t)\}$ to (B.2) on $[0, \infty)$ which are uniformly bounded (the bound is independent of m).

Step 4 (convergence to a Filippov solution [30, Lemma 2]): Here, the main difference is that we have time-varying delays. Using Mazur's convexity theorem [40] and property (iii) of the sequence $\{g^m(\cdot)\}$ (see step 1), one can obtain a sequence of functions, which is a convex combination of the original sequence $\{y^m(t)\}$, converges to a solution of (2.1) in the sense of Definition 2.1 (for details, e.g. how the output $\gamma(t)$ is constructed, see [30]).

References

- [1] C.K. Ahn, Passive learning and input-to-state stability of switched Hopfield neural networks with time-delay, *Information Sciences* 180 (2010) 4582–4594.
- [2] J.P. Aubin, H. Frankowska, *Set-Valued Analysis*, Birkhauser, Boston, 1990.
- [3] S. Boyd, L. Ghaoui, E. Laurent, V. Balakrishnan, *Linear Matrix Inequalities in System and Control Theory*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, 1994.
- [4] X. Chen, Q. Song, Global exponential stability of the periodic solution of delayed Cohen–Grossberg neural networks with discontinuous activations, *Neurocomputing* 73 (2010) 3097–3104.
- [5] L.O. Chua, C.A. Desoer, E.S. Kuh, *Linear and Nonlinear Circuits*, Macgraw-Hill, New York, 1987.
- [6] P.P. Civalleri, M. Gilli, L. Pandolfi, On stability of cellular neural networks with delay, *IEEE Transactions on Circuits and Systems I – Fundamental Theory and Applications* 40 (1993) 157–165.
- [7] F.H. Clarke, *Optimization and Nonsmooth Analysis*, Wiley, New York, 1983.
- [8] M. Forti, P. Nistri, Global convergence of neural networks with discontinuous neuron activations, *IEEE Transactions on Circuits and Systems I – Fundamental Theory and Applications* 50 (2003) 1421–1435.
- [9] M. Forti, P. Nistri, D. Papini, Global exponential stability and global convergence of delayed neural networks with infinite gains, *IEEE Transactions on Neural Networks* 16 (2005) 1449–1463.
- [10] M. Forti, M. Grazzini, P. Nistri, L. Pancioni, Generalised Lyapunov approach for convergence of neural networks with discontinuous or non-Lipshitz activations, *Physica D* 214 (2006) 88–99.
- [11] P. Gahinet, A. Nemirovski, A.J. Laub, M. Chilali, *LMI Control Toolbox – For Use with MATLAB*, The Math Works, Inc., 1995.
- [12] A. Granas, J. Dugundji, *Fixed Point Theory*, Springer-Verlag, New York, 2003.
- [13] K. Gu, An integral inequality in the stability problem of time-delay systems, in: *Proceedings of IEEE Conference Decision Control*, Sydney, Australia, Dec. 2000, pp. 2805–2810.
- [14] G. Haddad, Monotone viable trajectories for functional-differential inclusions, *Journal of Differential Equations* 42 (1981) 1–24.
- [15] J.K. Hale, S.M. Verduyn Lunel, *Introduction to Functional Differential Equations*, Springer-Verlag, New York, 1993.
- [16] S. Haykin, *Neural Networks: A Comprehensive Foundation*, Prentice Hall, NJ, 1994.
- [17] Y. Hino, S. Murakami, T. Naito, *Functional Differential Equations with Infinite Delay*, Springer-Verlag, Berlin, 1991.
- [18] J.J. Hopfield, Neurons with graded response have collective computational properties like those of two-state neurons, *Proceedings of the Academy of Natural Sciences* 81 (1984) 3088–3092.
- [19] J.J. Hopfield, D.W. Tank, Computing with neural circuits: a model, *Science* 233 (1986) 625–633.
- [20] R.A. Horn, C.R. Johnson, *Topics in Matrix Analysis*, Cambridge University Press, Cambridge (UK), 1994.
- [21] C. Huang, Y. He, L. Huang, W. Zhu, p th moment stability analysis of stochastic recurrent neural networks with time-varying delays, *Information Sciences* 178 (2008) 2194–2203.
- [22] M. Jiang, G. Gielen, B. Deng, X. Zhu, A fast learning algorithm for time-delay neural networks, *Information Sciences* 148 (2002) 27–39.
- [23] M.P. Kennedy, L.O. Chua, Neural networks for nonlinear programming, *IEEE Transactions on Circuits and Systems I – Fundamental Theory and Applications* 35 (1988) 554–562.
- [24] A.A. Levakov, Existence theorems for solutions of stochastic differential equations with discontinuous right-hand sides. (Russian), *Differentsial'nye Uravneniya* 36 (2000) 47–53, 1. translation in *Differential Equations* 36 (2000), 56–63.
- [25] J.H. Li, A.N. Michel, W. Porod, Analysis and synthesis of a class of neural networks: variable structure systems with infinite gains, *IEEE Transactions on Circuits and Systems* 36 (1989) 713–731.
- [26] Y. Li, S.C.K. Shiu, S.K. Pal, Combining feature reduction and case selection in building CBR classifiers, *IEEE Transactions on Knowledge and Data Engineering* 18 (3) (2006) 415–429.
- [27] Y. Li, H. Wu, Global stability analysis in Cohen–Grossberg neural networks with delays and inverse Hölder neuron activation functions, *Information Sciences* 180 (2010) 4022–4030.
- [28] W. Lu, T. Chen, Dynamical behaviors of Cohen–Grossberg neural networks with discontinuous activation functions, *Neural Networks* 18 (2005) 231–242.
- [29] W. Lu, T. Chen, Dynamical behaviors of delayed neural network systems with discontinuous activation functions, *Neural Computing* 18 (2006) 683–708.
- [30] W. Lu, T. Chen, Almost periodic dynamics of a class of delayed neural networks with discontinuous activations, *Neural Computing* 20 (2008) 1065–1090.
- [31] C.M. Marcus, R.M. Westervelt, Stability of analog neural networks with delay, *Physical Review A* 39 (1989) 347–359.
- [32] D. Papini, V. Taddei, Global exponential stability of the periodic solution of a delayed neural network with discontinuous activations, *Physical Review A* 343 (2005) 117–128.
- [33] M.J.S. Smith, C.L. Portmann, Practical design and analysis of a simple “neural” optimization circuit, *IEEE Transactions on Circuits and Systems* 36 (1989) 42–50.
- [34] V.I. Utkin, *Sliding Modes and Their Applications in Variable Structure Systems*, MIR Publishers, Moscow, 1978.
- [35] A. Waibel, T. Hanazawa, G. Hinton, K. Shikano, K.J. Lang, Phoneme recognition using time-delay neural networks, *IEEE Transactions on Acoustics, Speech and Signal Processing* 37 (2002) 328–339.
- [36] J. Wang, L. Huang, Z. Guo, Dynamical behavior of delayed Hopfield neural networks with discontinuous activations, *Applied Mathematical Modelling* 33 (2009) 1793–1802.
- [37] Y. Wang, Y. Zuo, L. Huang, C. Li, Global robust stability of delayed neural networks with discontinuous activation functions, *IET Control Theory and Applications* 2 (2008) 543–553.
- [38] H. Wu, Global stability analysis of a general class of discontinuous neural networks with linear growth activation functions, *Information Sciences* 179 (2009) 3432–3441.
- [39] X. Wu, Y. Wang, L. Huang, Y. Zuo, Robust exponential stability criterion for uncertain neural networks with discontinuous activation functions and time-varying delays, *Neurocomputing* 73 (2010) 1265–1271.
- [40] K. Yosida, *Functional Analysis*, Springer-Verlag, Berlin, 1980.
- [41] Y. Zuo, Y. Wang, L. Huang, Z. Wang, X. Liu, X. Wu, Robust stability criterion for delayed neural networks with discontinuous activation functions, *Neural Processing Letters* 29 (2009) 29–44.