Front. Electr. Electron. Eng. 2012, 7(1): 32–48
DOI 10.1007/s11460-012-0186-y

REVIEW ARTICLE

Wenlian LU, Tianping CHEN, Bo LIU, Xiangnan HE

Dynamical behaviors of recurrently connected neural networks and linearly coupled networks with discontinuous right-hand sides

Abstract The aim of this paper is to provide a systematic review on the framework to analyze dynamics in recurrently connected neural networks with discontinuous right-hand sides with a focus on the authors' works in the past three years. The concept of the Filippov solution is employed to define the solution of the neural network systems by transforming them to differential inclusions. The theory of viability provides a tool to study the existence and uniqueness of the solution and the Lyapunov function (functional) approach is used to investigate the global stability and synchronization. More precisely, we prove that the diagonal-dominant conditions guarantee the existence, uniqueness, and stability of a general class of integro-differential equations with (almost) periodic self-inhibitions, interconnection weights, inputs, and delays. This model is rather general and includes the well-known Hopfield neural networks, Cohen-Grossberg neural networks, and cellular neural networks as special cases. We extend the absolute stability analysis of gradient-like neural network model by relaxing the analytic constraints so that they can be employed to solve optimization problem with non-smooth cost functions. Furthermore, we study the global synchronization problem of a class of linearly coupled neural network with discontinuous right-hand sides.

Keywords delayed integro-differential system, discontinuous activation, almost periodic function, non-smooth cost function, complete synchronization

Received October 8, 2011; accepted November 25, 2011

Wenlian LU (✉), Tianping CHEN, Bo LIU, Xiangnan HE
School of Mathematical Sciences, Fudan University, Shanghai 200433, China
E-mail: wenlian@fudan.edu.cn

1 Introduction

It is well known that recurrently connected neural networks (RCNNs), proposed by Cohen and Grossberg [1], Hopfield [2,3], and Chua and Yang [4,5], have been extensively studied in both theory and applications. They have been successfully applied in signal processing, pattern recognition, and associative memories, especially in static image treatment. The key point of success of an algorithm depends on whether the dynamical flow converges to a given equilibrium or manifold. Therefore, dynamical behavior analysis of neural networks is the first step for the desired applications.

In Ref. [1], the authors proposed a competitive and cooperative network to generate self-organized and self-adaptive neural networks, which can be modeled as an ODE system:

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = A_i(x_i)\Big[-d_i(x_i) + \sum_{j=1}^{n} t_{ij} g_j(x_j) + I_i\Big], \quad i = 1, 2, \ldots, n, \tag{1}$$

which is known as the Cohen-Grossberg neural networks (CGNNs) and widely used in pattern recognition, signal processing, and associative memory. Here, $x_i(t)$ denotes the state variable of the ith neuron, $d_i(\cdot)$ represents the self-inhibition function with which the ith neuron will reset its potential to the resting state in isolations when disconnected from the network, t_{ij} denotes the strength of the jth neuron on the ith neuron, $g_i(\cdot)$ denotes the activation function of the ith neuron, I_i denotes the external input to the ith neuron, and $A_i(\cdot)$ denotes amplification function of the ith neuron.

References [2,3] developed a computing method using recurrent network based on energy functions, which is known as the Hopfield neural networks (HNNs):

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = -d_i x_i + \sum_{j=1}^{n} t_{ij} g_j(x_j) + I_i, \quad i = 1, 2, \ldots, n. \tag{2}$$

HNN is characterized as the activation function $g_j(\cdot)$ is a sigmoid function, for example,

$$g_j(s) = \frac{1}{1 + \exp(-\lambda_j s)}, \tag{3}$$

when the interconnections $T = [t_{ij}]_{i,j=1}^{n}$ is symmetric, i.e., $t_{ij} = t_{ji}$. This model can be used to search the minimum points of the quadratic objective function $\sum_{i,j=1}^{n} w_{ij} v_i v_j$ over the discrete set $\{0, 1\}^n$. Thus, the following energy function

$$L(y) = \sum_{i=1}^{n} \left[d_i \int_0^{y_i} g_i^{-1}(\rho)\mathrm{d}\rho - I_i y_i - \frac{1}{2}\sum_{j=1}^{n} t_{ij} y_i y_j \right]$$

is a Lyapunov function for Eq. (2). It can be seen that Eq. (2) has the following gradient-like form:

$$\begin{cases} \dot{x} = -Dx - \partial f(y) + I, \\ y = g(x), \end{cases} \tag{4}$$

with $f(y) = -(1/2)\sum_{i,j=1}^{n} t_{ij} y_i y_j$. Here, $D = \mathrm{diag}[d_1, d_2, \ldots, d_n]$ and $g(y) = [g_1(y_1), g_2(y_2), \ldots, g_n(y_n)]^\top$. This model is named as analytic neural networks (ANNs) [6], since $f(\cdot)$ was assumed to be analytic when the general model was proposed.

Also, in Refs. [4,5], Chua and Yang proposed a locally connected network model named as cellular neural network (CNN), which can be described by

$$\begin{aligned} \dot{x}_i &= -d_i x_i + \sum_{i=1}^{n} t_{ij} y_j + I_i, \\ y_i &= s(x_i), \ i = 1, 2, \ldots, n, \end{aligned} \tag{5}$$

where the activation function $s(\rho) = (|\rho + 1| - |\rho - 1|)/2$ is the linear saturation function. With assuming that $a_{ij} = a_{ji}$ for all $i, j = 1, 2, \ldots, n$, it can be seen that Eq. (5) has the form (24) below by letting $f(y) = -(1/2)\sum_{i,j=1}^{n} t_{ij} y_i y_j$ as well. Therefore, it can be seen that Eq. (4) is a unified form of HNN and CNN with symmetry interconnections. However, when they are asymmetric, Eq. (4) is not valid any more.

On the other hand, in practice, time delays inevitably occur due to the finite switching speed of the amplifiers and communication time. Moreover, to process moving images, one must introduce time delays in the signals transmitted among cells [7]. Neural networks with time delays have much more complicated dynamics. Furthermore, research of delayed neural networks with varying self-inhibitions, interconnection weights, and inputs is another important issue, because, in real world, self-inhibitions, interconnection weights, and inputs should vary as time varies. Thus, we are to study the following **asymmetric** delayed neural networks with a general form:

$$\begin{aligned} \frac{\mathrm{d}x_i(t)}{\mathrm{d}t} &= A_i(x_i(t))\Bigg[-d_i(t)x_i(t) + \sum_{j=1}^{n} a_{ij}(t) g_j(x_j(t)) \\ &\quad + \sum_{j=1}^{n} \int_0^{\infty} g_j(x_j(t-s)) d_s K_{ij}(t,s) + I_i(t) \Bigg], \\ &\qquad i = 1, 2, \ldots, n, \end{aligned} \tag{6}$$

where $d_s K_{ij}(t,s)$, $i, j = 1, 2, \ldots, n$, are Lebesgue-Stieltjes measures with respect to s, which denotes the delayed terms. If $a_i(s) = 1$ for all i, it takes form:

$$\begin{aligned} \frac{\mathrm{d}x_i(t)}{\mathrm{d}t} &= -d_i(t)x_i(t) + \sum_{j=1}^{n} a_{ij}(t) g_j(x_j(t)) \\ &\quad + \sum_{j=1}^{n} \int_0^{\infty} g_j(x_j(t-s)) d_s K_{ij}(t,s) + I_i(t), \\ &\qquad i = 1, 2, \ldots, n, \end{aligned} \tag{7}$$

which is firstly introduced in Ref. [8] with asymmetric interconnections.

In the past decades, the analysis of the dynamics of neural network systems has attracted a large amount of interests from diverse research fields. A huge number of papers have arisen to give all kinds of sufficient conditions to verify their global convergence. For details, we refer the interested readers to a book chapter [9] and the references therein.

However, most of the works were based on the assumption that the activation functions are continuous even globally Lipschitz. As mentioned in Ref. [10], a brief review on some common neural network models reveals that neural networks with discontinuous activations are of importance and do frequently arise in practice. For example, consider the classical HNNs with graded response neurons [2]. The standard assumption is that the activations used are in high-gain limit, where they closely approach discontinuous and comparator functions. As shown in Refs. [2,3], the high-gain hypothesis is crucial to make negligible the connection to the neural network energy function of the term depending on neuron self inhibitions, and to favor binary output formation. For example, the activation function $g_i(\cdot)$ is selected as the sign function $\mathrm{sign}(s)$.

Also, a conceptually analogous model based on hard comparators is also used to describe the discrete-time neural networks in Ref. [11]. Another important example is the neural networks introduced in Ref. [12] to solve linear and nonlinear programming problems. Those networks exploit constrained neurons with a diode-like input-output activations. Again, in order to satisfy the constraints, the diodes are required to possess a very high slope in the conducting region, i.e., they should approximate the discontinuous characteristic of an ideal diode. When dealing with dynamical systems possessing high-slope nonlinear elements, it is often of advantage to model them with a system of differential equations with discontinuous right-hand side, rather than studying the case where the slope is high but of finite value [13].

In this paper, for the discontinuous activations, we make following assumption:

$\mathcal{B}_1$: $g_i(\cdot)$ is nondecreasing and local Lipschitzian, except on a set of isolated points $\{\rho_k^i\}$. More precisely, for each $i = 1, 2, \ldots, n$, $g_i(\cdot)$ is monotonically nondecreasing and continuous, except on a set of isolated points $\{\rho_k^i\}$, where the right and left limits $g_i^+(\rho_k^i)$ and $g_i^-(\rho_k^i)$ satisfy $g_i^+(\rho_k^i) > g_i^-(\rho_k^i)$; in each compact set of $\mathbb{R}$, $g_i(\cdot)$ has only finite number of discontinuities; moreover, denote the set of discontinuities by order $\{\rho_k^i: \ \rho_{k+1}^i > \rho_k^i, \ k \in Z\}$ and there exist positive constants $G_{i,k} > 0$, $i = 1, 2, \ldots, n$, $k \in Z$ such that $|g_i(\xi) - g_i(\zeta)| \leqslant G_{i,k}|\xi - \zeta|$ holds for all $\xi, \zeta \in (\rho_k^i, \rho_{k+1}^i)$.

In Refs. [14,15], for the ANN, the Łojasiewicz gradient inequality reveals a fundamental relation between gradient vector fields and the difference of the values of the function, which provides a powerful tool to analyze the gradient dynamical systems. This inequality demands that the cost function be analytic. However, in many situations, the cost function may not be analytic or smooth. For example, the distance $f(y) = d(y, \bar{y})$ is not smooth at the point $y = \bar{y}$, where $\bar{y}$ is a given reference point and $d(\cdot, \cdot)$ is the metric in a given metric space. In addition, the activation function is not always analytic or strictly monotone increasing. For example, the saturation function is piecewise analytic and monotonely increasing but not strictly increasing. Therefore, it is natural and necessary to investigate a generalized dynamical system model that includes non-smooth cost function as well as piecewise analytic and non-strictly increasing activation functions.

In this paper, we focus on several results given in Refs. [16–19]. The main goal of this article is to present a cohesive overview of the key approach for the analysis of differential systems with discontinuous right-hand sides. This approach applies to diverse sorts recurrently connected neural networks, including periodic or almost periodic delayed neural networks, gradient-like neural system, and coupled neural networks. As a unified framework, we introduce the concepts of Filippov solution and the theory of differential inclusions, and then divide the analysis into two steps:

- Existence of the Cauchy problem of the differential inclusions;
- Stability analysis via the Lyapunov functional approach.

By this fashion, we study the stability of a class of differential-integral equations with discontinuous activations, almost periodic coefficients and delays. We propose a novel class of gradient-like neural systems without the analyticity restrictions. Thus, it can be used to solve a class of optimization problems with non-smooth cost functions. In addition, we consider synchronization problem of linearly coupled differential systems that has discontinuity at their right-hands.

We organize this chapter as follows. We briefly introduce the Filippov theory in Sect. 2 and the viability of differential inclusions in Sect. 3. In Sect. 4, we study the periodicity and almost periodicity. In Sect. 5, we investigate the class of gradient-like systems with non-smooth cost functions. In Sect. 6, we discuss synchronization in network of coupled neural models with discontinuous right-hand sides. We present discussion and reviews of literature on this topic and compare them with the results in Sect. 7.

We present the notations in the following:

- $A^\top$ for a matrix A denotes the transpose of A.
- $\|\cdot\|$ denotes the norm of a vector in some sense. In particular, $\|v\|_2$ for a vector $v = [v_1, v_2, \ldots, v_n]^\top$ denotes the 2-norm by the way $\|v\|_2 = \sqrt{\sum_{i=1}^n |v_i|^2}$; $\|v\|_{\{\xi,\infty\}}$ for some positive vector $\xi = (\xi_1, \xi_2, \ldots, \xi_n)^\top$ is denoted by $\|v\|_{\{\xi,\infty\}} = \max_i \xi_i^{-1}|v_i|$; $\|v\|_{\xi,1}$ is denoted by $\|v\|_{\{\xi,1\}} = \sum_{i=1}^n \xi_i|v_i|$; $\|v\|_1 = \sum_{i=1}^n |v_i|$. The norm of a matrix is induced by the definition of the norm of vectors.
- $C([a,b], \mathbb{R}^n)$ denotes the class of continuous functions from $[a,b]$ to $\mathbb{R}^n$. For each $x(\cdot) \in C([a,b], \mathbb{R}^n)$, its norm is denoted by $\|x(\cdot)\| = \max_{t\in[a,b]} \|x(t)\|$ for some vector norm $\|\cdot\|$.
- $\mathbb{R}_+^n$ denotes the first orthant, $\mathbb{R}_+^n = \{x = (x_1, x_2, \ldots, x_n)^\top: \ x_i > 0, \ \forall\ i = 1, 2, \ldots, n\}$.
- sign$(\cdot)$ denotes the signature function.

2 Filippov solution and differential inclusions

When the activations $g_i(\cdot)$, $i = 1, 2, \ldots, n$, are discontinuous in Eq. (2) or Eqs. (6) and (7), or the cost function in Eq. (24) is non-smooth, the right-hand side of the underlying system is not continuous any more. The theory of ordinary differential equations does not work.

Consider the following system:

$$\frac{\mathrm{d}x}{\mathrm{d}t} = f(x), \tag{8}$$

where $f(\cdot)$ is not continuous. Reference [20] proposed the following definition of the solution for the system (8) by transforming it into differential inclusions [21].

Definition 1 A set-value map defined as

$$\phi(x) = \bigcap_{\delta>0} \bigcap_{\mu(N)=0} \overline{co}\Big[f(\overline{\mathcal{O}}(x,\delta) - N)\Big], \tag{9}$$

where $\overline{co}(E)$ is the closure of the convex hull of some set E, $\overline{\mathcal{O}}(x,\delta) = \{y \in \mathbb{R}^n : \|y - x\| \leqslant \delta\}$, and $\mu(N)$ is the

Lebesgue measure of the set N. A solution of the Cauchy problem for Eq. (8) with initial condition $x(0) = x_0$ is an absolutely continuous function $x(t)$, $t \in [0,T)$, which satisfies: $x(0) = x_0$, and differential inclusion:

$$\frac{\mathrm{d}x}{\mathrm{d}t} \in \phi(x), \qquad a.e.\ \ t \in [0,T). \tag{10}$$

Here, $\phi(x)$ is a *set-valued map*, which is defined as that, for each point x, there corresponds a non-empty set $F(x) \subset \mathbb{R}^n$. For more details about set-valued map, please refer to Ref. [22]. *Differential inclusions* mean that there is an absolutely continuous function $x(t)$ such that Eq. (10) holds for almost every t. More details can be found in Ref. [21].

In case that the cost function is non-smooth, which implies that the differential of f may be discontinuous, we can directly use the definition of Filippov solution to redefine the solution for system (4). First of all, we should redefine the differential of f at those non-smooth points. Here, for a strictly continuous function $f : \mathbb{R}^n \to \mathbb{R}$, the *Clarke's generalized gradient* of f at $x \in \mathbb{R}^n$, which can always be used to handle gradient flow on non-smooth functions, can be written as

$$\partial f = \{p \in \mathbb{R}^n :\ f^o(x,v) \geqslant \langle p, v\rangle,\ \forall\ v \in \mathbb{R}^n\}.$$

Thus, system (4) takes the form of differential inclusions:

$$\dot{x} \in -Dx - \partial f(y) + I, \quad y = g(x). \tag{11}$$

For the delayed asymmetric neural networks (6) and (7), we should consider the delayed version of the Filippov solution. In Refs. [21,23,24], the following functional differential inclusion with memory were proposed:

$$\frac{\mathrm{d}x}{\mathrm{d}t}(t) \in F(t, A(t)x), \tag{12}$$

where $F\ : \mathbb{R} \times C([-\tau,0],\mathbb{R}^n) \mapsto \mathbb{R}^n$ is a given set-value map, and

$$[A(t)x](\theta) = x_t(\theta) = x(t+\theta). \tag{13}$$

Inspired by these works, in this paper, we denote $\overline{co}[g_i(s)] = [g_i^-(s), g_i^+(s)]$ and $\overline{co}[g(x)] = \overline{co}[g_1(x_1)] \times \overline{co}[g_2(x_2)] \times \cdots \times \overline{co}[g_n(x_n)]$, where $\times$ denotes the Cartesian product. Thus, we can define solution of the system (6) in the Filippov sense as follows.

Definition 2 For a continuous function $\phi(\theta) = [\phi_1(\theta), \phi_2(\theta), \ldots, \phi_n(\theta)]^\top$ and a measurable function $\lambda(\theta) = [\lambda_1(\theta), \lambda_2(\theta), \ldots, \lambda_n(\theta)]^\top \in \overline{co}[g(\phi(\theta))]$ for almost all $\theta \in (-\infty, 0]$, an absolute continuous function $x(t) = x(t,\phi,\lambda) = [x_1(t), x_2(t), \ldots, x_n(t)]^\top$ associated with a measurable function $\gamma(t) = [\gamma_1(t), \gamma_2(t), \ldots, \gamma_n(t)]^\top$ is said to be a solution of the Cauchy problem for the system (6) on $[0,T)$ (T might be ∞) with initial value $(\phi(\theta), \lambda(\theta))$, $\theta \in (-\infty, 0]$, if

$$\begin{cases} \dfrac{\mathrm{d}x_i(t)}{\mathrm{d}t} = A_i(x_i(t))\Big[-d_i(t)x_i(t) + \displaystyle\sum_{j=1}^n a_{ij}(t)\gamma_j(t) \\ \qquad + \displaystyle\int_0^\infty \gamma_j(t-s)d_sK_{ij}(t,s) + I_i(t)\Big], \\ \qquad\qquad\qquad\qquad a.e.\ t \in [0,T), \\ \gamma_i(t) \in \overline{co}[g_i(x_i(t))], \qquad a.e.\ t \in [0,T), \\ x_i(\theta) = \phi_i(\theta), \qquad \theta \in (-\infty, 0], \\ \gamma_i(\theta) = \lambda_i(\theta), \qquad a.e.\ \theta \in (-\infty, 0], \end{cases} \tag{14}$$

holds for all $i = 1, 2, \ldots, n$.

The solution of the system (22) below can be defined by the same way.

3 Viability

The first question arises when we write down the differential inclusions to replace the original differential equations: whether do these differential inclusions have solutions and whether are the solutions unique?

For any $f \in \Xi$, by the theory given in Ref. [25], the following system:

$$\begin{cases} \dfrac{\mathrm{d}u_i^f}{\mathrm{d}t}(t) = A_i(x_i)\Big[-d_i(t)u_i^f(t) + \displaystyle\sum_{j=1}^n a_{ij}(t)\sigma_j^f(t) \\ \qquad + \displaystyle\sum_{j=1}^n \int_0^\infty \sigma_j^f(t-s)d_sK_{ij}(t,s) + I_i(t)\Big], \\ u_i^f(\theta) = \phi_i(\theta), \quad \theta \in (-\infty, 0], \\ \sigma_i^f(\theta) = \begin{cases} \lambda_i(\theta), & \theta \leqslant 0, \\ f_i(u_i^f(\theta)), & \theta > 0, \end{cases} \quad i = 1, 2, \ldots, n, \end{cases} \tag{15}$$

admits a unique solution $u_f(t) = [u_1(t), u_2(t), \ldots, u_n(t)]^\top$ on $[0,T)$, where T might be ∞.

We make following assumption for the initial value:

$\mathcal{B}_2$: The initial condition satisfies that $\phi(\theta) \in C((-\infty,0],\mathbb{R}^n)$ is bounded and $\lambda(\theta)$ is measurable and essentially bounded.

First, we prove that the solutions $u^f(t)$ are uniformly bounded with respect to $f \in \Xi$.

Lemma 1 *Suppose that the assumptions $\mathcal{B}_{1,2}$ hold. If there exist constants $\xi_i > 0$, $i = 1, 2, \ldots, n$, and $\delta > 0$ such that $d_i(t) \geqslant \delta$ and*

$$\xi_i a_{ii}(t) + \sum_{j=1, j\neq i}^n \xi_j |a_{ji}(t)| + \sum_{j=1}^n \xi_j \int_0^\infty e^{\delta s} |d\bar{K}_{ji}(s)| < 0 \tag{16}$$

hold for all $t \geqslant 0$ and $i = 1, 2, \ldots, n$, then the solutions $u^f(t)$ are uniformly bounded with respect to $f \in \Xi$, i.e., there exists $M = M(\phi,\lambda) > 0$, which is independent of $f \in \Xi$, such that $\|u^f(t)\|_{\{\xi,1\}} \leqslant M$ holds for all $f \in \Xi$ and $t \geqslant 0$. Consequently, the existence interval of $u^f(t)$ can be extended to $[0,\infty)$.

Now, for any sequence

$$\left\{g^m(x) = (g_1^m(x_1), g_2^m(x_2), \ldots, g_n^m(x_n))^\top\right\}_{m\in\mathbb{N}} \in \Xi$$

satisfying

$$\lim_{m\to\infty} d_{\rm H}(Graph(g^m(K)), \overline{co}[g(K)]) = 0, \quad \text{for all } K \subset \mathbb{R}^n, \tag{17}$$

where $d_{\rm H}(\cdot,\cdot)$ denotes the Hausdorff metric of $\mathbb{R}^n$, we construct a sequence of delayed systems with high-slope continuous activations as follows:

$$\frac{{\rm d}u_i^m(t)}{{\rm d}t} = -d_i(t)u_i^m(t) + \sum_{j=1}^n a_{ij}(t)\sigma_j^m(t) + \sum_{j=1}^n \int_0^\infty \sigma_j^m(t-s)d_sK_{ij}(t,s) + I_i(t), \quad i = 1,2,\ldots,n, \tag{18}$$

where $u_i^m(\theta) = \phi_i(\theta)$, $\theta \in (-\infty, 0]$, and

$$\sigma_j^m(\theta) = \begin{cases} \lambda_j(\theta), & \theta \leqslant 0, \\ g_j^m(u_j(\theta)), & \theta > 0. \end{cases}$$

For instance, let $\{\rho_{k,i}\}$ be the set of discontinuous points of $g_i(\cdot)$. Pick a strictly decreasing sequence $\{\delta_{k,i,m}\}$ with $\lim_{m\to\infty} \delta_{k,i,m} = 0$ and define $I_{k,i,m} = [\rho_{k,i} - \delta_{k,i,m}, \rho_{k,i} + \delta_{k,i,m}]$ such that for every $k_1 \neq k_2$, $I_{k_1,i,m} \bigcap I_{k_2,i,m} = \emptyset$ hold. Then, we define functions $g_i^m(\cdot)$ as follows:

$$g_i^m(s) = \begin{cases} g_i(s), & s \notin \bigcup_{k\in\mathbb{Z}} I_{k,i,m}, \\ \dfrac{g_i(\rho_{k,i}+\delta_{k,i,m}) - g_i(\rho_{k,i}-\delta_{k,i,m})}{2\delta_{k,i,m}} \cdot [s - \rho_{k,i} - \delta_{k,i,m}] + g_i(\rho_{k,i}+\delta_{k,i,m}), & s \in I_{k,i,m}. \end{cases}$$

It can be seen that the sequence $\{g^m(\cdot)\}_{m\in\mathbb{N}} \subset \Xi$ satisfies condition (17).

Second, we will point out that the solution sequence of the system sequence (18) converges to a solution of the system (6) in the Filippov's sense.

Lemma 2 *Suppose that the assumptions $\mathcal{B}_{1,2}$ hold. If the condition* (16), *then for each initial value pair* (ϕ,λ), *the system* (6) *has a solution in the Filippov's sense on the whole time interval* $[0,\infty)$.

4 Global convergence of neural networks with discontinuous activations

In this section, we present several results on the stability of neural networks with discontinuous activations. In particular, we consider the dynamics of Eqs. (7) and (6). For Eq. (7), we concentrate on the almost periodic dynamics when the parameters are all almost periodic. In Eq. (6), we focus on its nonnegative dynamics, namely, the trajectory stays in the first orthant $\mathbb{R}_+^n$ for all time.

4.1 Almost periodic dynamics of Hopfield neural networks with discontinuous activations

We firstly consider the system (7) rewritten as follows:

$$\frac{{\rm d}x_i(t)}{{\rm d}t} = -d_i(t)x_i(t) + \sum_{j=1}^n a_{ij}(t)g_j(x_j(t)) + \sum_{j=1}^n \int_0^\infty g_j(x_j(t-s))d_sK_{ij}(t,s) + I_i(t), \quad i = 1,2,\ldots,n, \tag{19}$$

with the almost periodic parameters.

$\mathcal{B}_3$: $d_i(t)$ and $a_{ij}(t)$ are all continuous functions, $i,j = 1,2,\ldots,n$ and $d_i(t) \geqslant \delta > 0$, $a_{ii}(t) < 0$ hold for all $i = 1,2,\ldots,n$ and $t \in \mathbb{R}$; for any $s \in \mathbb{R}$, the Lebesgue-Stieltjes measures $d_sK_{ij}(t,s) : t \mapsto d_sK_{ij}(t,s)$ are continuous, i.e., $\lim_{h\to 0}\int_0^\infty |d_sK_{ij}(t+h,s) - d_sK_{ij}(t,s)| = 0$ holds for all $i,j = 1,2,\ldots,n$ and $t \in \mathbb{R}$, and $d_sK_{ij}(t,s)$ is dominated by some Lebesgue-Stieltjes $d\bar{K}_{ij}(s)$ independent of t satisfying $\int_0^\infty e^{\delta s}|d\bar{K}_{ij}(s)| < +\infty$ for all $i,j = 1,2,\ldots,n$ and some $\delta > 0$. Here, the domination means $|d_sK_{ij}(t,s)| \leqslant |d\bar{K}_{ij}(s)|$ i.e., $\int_0^\infty f(s)|d_sK_{ij}(t,s)| \leqslant \int_0^\infty f(s)|d\bar{K}_{ij}(s)|$ holds for all $t \geqslant 0$ and any nonnegative measurable function $f(\cdot)$; moreover, $d_i(t)$, $a_{ij}(t)$, $I_i(t)$, and $d_sK_{ij}(t,s)$ all possess almost periodic property, i.e., for any $\epsilon > 0$, there exists $l = l(\epsilon)$ such that for any interval $[\alpha, \alpha+l]$, there exists $\omega \in [\alpha, \alpha+l]$ such that

$$\begin{aligned} &|d_i(t+\omega) - d_i(t)| < \epsilon, \\ &|a_{ij}(t+\omega) - a_{ij}(t)| < \epsilon, \\ &|I_i(t+\omega) - I_i(t)| < \epsilon, \\ &\int_0^\infty |d_sK_{ij}(t+\omega,s) - d_sK_{ij}(t,s)| < \epsilon \end{aligned}$$

hold for all $i,j = 1,2,\ldots,n$ and $t \in \mathbb{R}$.

We study the almost periodicity of the delayed neural networks. The main results stated below was proven in Ref. [16].

Theorem 1 *Suppose the assumptions $\mathcal{B}_{1,2,3}$ are satisfied. If there exist constants $\xi_i > 0$, $i = 1,2,\ldots,n$, and $\delta > 0$ such that $d_i(t) \geqslant \delta$ and*

$$\xi_i a_{ii}(t) + \sum_{j=1, j\neq i}^n \xi_j |a_{ji}(t)| + \sum_{j=1}^n \xi_j \int_0^\infty e^{\delta s}|d\bar{K}_{ji}(s)| < 0 \tag{20}$$

hold for all $t \geqslant 0$ and $i = 1,2,\ldots,n$, then

1) *For every initial value* (ϕ,λ), *the system* (19) *has a unique solution in the sense of Eq.* (14);

2) *There exists a unique almost periodic solution* $x^*(t)$ *for the system* (19)*, which is globally exponentially stable, i.e., for any other solution* $x(t)$ *with the initial condition* (ϕ, λ)*, there exists a constant* $M = M(\phi, \lambda) > 0$ *such that*

$$\|x(t) - x^*(t)\|_{\{\xi,1\}} \leqslant Me^{-\delta t}$$

holds for all $t \geqslant 0$.

Since any periodic function can be regarded as an almost-periodic function, all the results apply to periodic case. Now, replacing assumption $\mathcal{B}_3$ with $\mathcal{B}_3^*$:

$\mathcal{B}_3^*$: $d_i(t)$ and $a_{ij}(t)$ are all continuous functions, $i, j = 1, 2, \ldots, n$ and $d_i(t) \geqslant \delta > 0$, $a_{ii}(t) < 0$ hold for all $i = 1, 2, \ldots, n$ and $t \in \mathbb{R}$; for any $s \in \mathbb{R}$, the Lebesgue-Stieltjes measures $d_sK_{ij}(t, s) : t \mapsto d_sK_{ij}(t, s)$ are continuous, i.e., $\lim_{h\to 0} \int_0^\infty |d_sK_{ij}(t + h, s) - d_sK_{ij}(t, s)| = 0$ holds for all $i, j = 1, 2, \ldots, n$ and $t \in \mathbb{R}$, and $d_sK_{ij}(t, s)$ is dominated by some Lebesgue-Stieltjes $d\bar{K}_{ij}(s)$ independent of t satisfying $\int_0^\infty e^{\delta s}|d\bar{K}_{ij}(s)| < +\infty$ for all $i, j = 1, 2, \ldots, n$ and some $\delta > 0$. Here, the domination means $|d_sK_{ij}(t, s)| \leqslant |d\bar{K}_{ij}(s)|$ i.e., $\int_0^\infty f(s)|d_sK_{ij}(t, s)| \leqslant \int_0^\infty f(s)|d\bar{K}_{ij}(s)|$ holds for all $t \geqslant 0$ and any nonnegative measurable function $f(\cdot)$; moreover, $d_i(t)$, $a_{ij}(t)$, $I_i(t)$, and $d_sK_{ij}(t, s)$ are all periodic with the same period ω, i.e.,

$$\begin{aligned}
&d_i(t + \omega) = d_i(t),\\
&a_{ij}(t + \omega) = a_{ij}(t),\\
&I_i(t + \omega) = I_i(t),\\
&\int_0^\infty |d_sK_{ij}(t + \omega, s)| = d_sK_{ij}(t, s)
\end{aligned}$$

hold for all $i, j = 1, 2, \ldots, n$ and $t \in \mathbb{R}$.

Then, we have

Corollary 1 *Suppose that the discontinuous activations satisfy the assumptions* $\mathcal{B}_{1,2}$ *and* $\mathcal{B}_3^*$*. If there exist positive constants* ξ_i*,* $i = 1, 2, \ldots, n$*, and* $\delta > 0$ *such that* $d_i(t) \geqslant \delta$ *and*

$$\xi_i a_{ii}(t) + \sum_{j=1, j\neq i}^n \xi_j |a_{ji}(t)| + \sum_{j=1}^n \xi_j \int_0^\infty e^{\delta s}|d\bar{K}_{ji}(s)| < 0$$

hold for all $t \geqslant 0$ *and* $i = 1, 2, \ldots, n$*, then*

1) *For each initial data with assumption* A_3*, the system* (19) *has a unique solution in the sense of Eq.* (14)*;*
2) *There exists a unique periodic solution* $x^*(t)$ *for system* (19)*, which is globally exponentially stable.*

Furthermore, a constant can be regarded as a periodic function with any period. Therefore, for the following delayed system:

$$\frac{\mathrm{d}x_i(t)}{\mathrm{d}t} = -d_i x_i(t) + \sum_{j=1}^n a_{ij} g_j(x_j(t)) + \sum_{j=1}^n \int_0^\infty g_j(x_j(t - s))d_sK_{ij}(s) + I_i, \quad i = 1, 2, \ldots, n, \tag{21}$$

where parameters are all constants. Then, we have

Corollary 2 *Suppose that the discontinuous activations satisfy the assumption* $\mathcal{B}_{1,2}$*. If there exist positive constants* ξ_i*,* $i = 1, 2, \ldots, n$*, and* $\delta > 0$ *such that* $d_i \geqslant \delta$ *and*

$$\xi_i a_{ii} + \sum_{j=1, j\neq i}^n \xi_j |a_{ji}| + \sum_{j=1}^n \xi_j \int_0^\infty e^{\delta s}|d\bar{K}_{ji}(s)| \leqslant 0$$

hold for all $t \geqslant 0$ *and* $i = 1, 2, \ldots, n$*, then*

1) *For each initial data with the assumption, the system* (21) *has a unique solution in the sense of Eq.* (14)*;*
2) *The system* (21) *has a unique equilibrium* x^**, which is globally exponentially stable.*

4.2 Nonnegative almost periodic dynamics of delayed Cohen-Grossberg neural networks with discontinuous activations

In the previous subsection, we consider global stability of a general class of delayed HNNs with discontinuous activations. Here, $x(t) \in \mathbb{R}^n$. In the past decades, a lot of literature have been involved in stability analysis of CGNNs that can be regarded as special cases of Eq. (6). In most of these papers, it is assumed that amplifier functions $A_i(\cdot)$ are always **positive**, even greater than some positive numbers $A_i(\cdot) \geqslant \underline{A}_i > 0$. Hence, the dynamical analysis overlaps that of Hopfield models. See Refs. [26–28] for example. However, in the original papers [1,29,30], the authors proposed this model as a kind of competitive-cooperation dynamical system for decision rules, pattern formation, and parallel memory storage. Hereby, each state of neuron x_i might be the population size, activity, or concentration, etc., of the ith species in the system, which is always nonnegative for all time. To guarantee the positivity of the states, one should assume $A_i(\rho) > 0$ for all $\rho > 0$ and $A_i(0) = 0$ for all $i = 1, 2, \ldots, n$. It is clear that this subset of CGNNs includes the famous Volterra-Lotka competitive-cooperation equations, which can be formalized as follows:

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = A_i x_i\left(I_i - \sum_{j=1}^n a_{ij}x_j\right), \quad i = 1, 2, \ldots, n,$$

with letting $A_i(\rho) = A_i\rho$, for all $\rho > 0$ and given $A_i > 0$, and $g_i(\rho) = \rho$, $i = 1, 2, \ldots, n$.

The aim of this subsection is to study the nonnegative periodic dynamical behaviors of the delayed CGNN system (6), rewritten as follows:

$$\dot{x}_i = A_i(x_i)\Big[- d_i(t)x_i(t) + \sum_{j=1}^{n} a_{ij}(t)g_j(x_j(t)) + \sum_{j=1}^{n}\int_0^{\infty} g_j(x_j(t-s))d_sK_{ij}(t,s) + I_i(t)\Big], \quad i = 1,2,\ldots,n, \quad (22)$$

with discontinuous activations, periodic coefficients, and without assuming the strictly positivity of the amplifier functions. $d_sK_{ij}(t,s)$, $i,j = 1,2,\ldots,n$, are Lebesgue-Stieltjes measures with respect to s for any $t \in \mathbb{R}$. Hereby, we focus our study of the dynamical behaviors on the first orthant $\mathbb{R}^n_+$ and consider all trajectories initiated in the first orthant $\mathbb{R}^n_+$ instead of the whole space $\mathbb{R}^n$.

Consider the system of differential equation (22). The amplifier functions $A_i(\cdot)$ is assumed to satisfy:

$\mathcal{B}_4$: For all $i = 1,2,\ldots,n$, $A_i(s)$ is continuous and for $s \geqslant 0$ with $A_i(s) > 0$ for $s > 0$ and $A_i(0) = 0$, and

$$\int_0^{\epsilon} \frac{ds}{A_i(s)} = +\infty, \quad i = 1,2,\ldots,n,$$

where ϵ is an arbitrary positive number.

The activation functions can be of discontinuity and assumed to satisfy $\mathcal{B}_1$. The time-varying coefficients, including the self-inhibitions, interconnection coefficients, and the external inputs, are all almost periodic, i.e., satisfying $\mathcal{B}_2$.

The solution of the Cauchy problem of the delayed differential system (22) according to positive initial conditions $\phi_i(\theta) > 0$, $\theta \in (-\infty, 0]$, still in the sense defined in Eq. (14). Thus, suppose that the solution of Eq. (22) can exist to $T = +\infty$, the asymptotical stability of a given nonnegative solution $x^*(t)$ (if existing) is also in the first orthant.

Definition 3 $x^*(t)$ is said to be asymptotically $\mathbb{R}^n_+$-stable if for any positive bounded initial conditions $\phi_i(\theta)$ and measurable essentially bounded functions $\lambda_i(\theta)$ for $\theta \in [-\infty, 0]$ and $i = 1,2,\ldots,n$, the solution $x(t)$ in the sense of Eq. (22) satisfies

$$\lim_{t\to\infty} |x_i(t) - x_i^*(t)| = 0, \quad i = 1,2,\ldots,n.$$

The results in this subsection can be derived from the same fashion in Refs. [16] and [18] with minor modifications. The diagonal dominant conditions are the key for the existence and its stability of the almost periodic solution of the delayed system (22).

The assumption $\mathcal{B}_4$ guarantees that each solution trajectory (if existing) of the system (22) is positive if the initial data are positive. By Lemma 2 in Ref. [18], we have

Lemma 3 (*Positivity*) *Under the assumptions $\mathcal{B}_{1,3,4}$, if initial values are positive, i.e., $x_i(0) > 0$ for all $i = 1,2,\ldots,n$, then each solution of the system (22) is positive within the duration time.*

The following theorem gives the viability for the system (22).

Theorem 2 [18] *Under the assumptions $\mathcal{B}_{1,2,3,4}$, for each positive initial conditions, the system (22) admits one positive bounded solution of which the duration time interval is $[0, +\infty)$.*

This theorem can be derived directly from by Ref. [18].

With diagonal dominant condition, we obtained the existence of a nonnegative periodic solution as well as its $\mathbb{R}^n_+$-stability.

Theorem 3 *Under the assumptions $\mathcal{B}_{1,2,3,4}$, if there exist positive constants $\xi_1, \xi_2, \ldots, \xi_n$ and $\delta > 0$ such that for any $t > 0$, $d_i(t) \geqslant \delta$ and the following conditions are satisfied:*

$$\xi_i a_{ii}(t) + \sum_{j\neq i}^{n} \xi_j a_{ji}(t) + \sum_{j=1}^{n} \xi_j \int_0^{\infty} e^{\delta s}|d_s\hat{K}_{ij}(t,s)| < 0, \quad i = 1,2,\ldots,n,$$

then, the system (22) has a nonnegative almost periodic solution and this solution is globally asymptotically stable.

This theorem can be proven by a combination of the approaches in Refs. [16] and [18].

The following corollary is a direct consequence of previous theorem and comes from Ref. [18].

Corollary 3 *Suppose that the discontinuous activations satisfy assumptions $\mathcal{B}_{1,2,4}$ and $\mathcal{B}_3^*$. If there exist positive constants ξ_i, $i = 1,2,\ldots,n$, and $\delta > 0$ such that $d_i(t) \geqslant \delta$ and*

$$\xi_i a_{ii}(t) + \sum_{j=1, j\neq i}^{n} \xi_j |a_{ji}(t)| + \sum_{j=1}^{n} \xi_j \int_0^{\infty} e^{\delta s}|d\bar{K}_{ji}(s)| < 0$$

hold for all $t \geqslant 0$ and $i = 1,2,\ldots,n$, then for each initial data with assumption A_3, the system (22) has a nonnegative periodic solution $x^*(t)$ that is globally exponentially stable.

Furthermore, a constant can be regarded as a periodic function with any period. We have the following result, which also comes from Ref. [18].

Corollary 4 *Consider the following delayed system:*

$$\frac{\mathrm{d}x_i(t)}{\mathrm{d}t} = A_i(x_i)\Big[- d_i x_i(t) + \sum_{j=1}^{n} a_{ij} g_j(x_j(t)) + \sum_{j=1}^{n}\int_0^{\infty} g_j(x_j(t-s))d_sK_{ij}(s) + I_i\Big], \quad i = 1,2,\ldots,n. \quad (23)$$

Suppose that the discontinuous activations satisfy the assumption $\mathcal{B}_{1,3,4}$. If there exist positive constants ξ_i,

$i = 1, 2, \ldots, n$, and $\delta > 0$ such that $d_i \geqslant \delta$ and

$$\xi_i a_{ii} + \sum_{j=1, j\neq i}^{n} \xi_j |a_{ji}| + \sum_{j=1}^{n} \xi_j \int_0^\infty e^{\delta s} |d\bar{K}_{ji}(s)| \leqslant 0$$

hold for all $t \geqslant 0$ and $i = 1, 2, \ldots, n$, then the system (23) has a unique nonnegative equilibrium x^* that is globally exponentially stable.

5 Convergence of a class of gradient-like systems with non-smooth cost functions

This section is concerned with the convergence of the following dynamical system:

$$\begin{cases} \dot{x} \in -Dx - \partial f(y) + \theta, \\ y = g(x), \end{cases} \tag{24}$$

where $x = [x_1, x_2, \ldots, x_n]^\top \in \mathbb{R}^n$ is the state vector, $D = \mathrm{diag}[d_1, d_2, \ldots, d_n]$ with $d_i > 0$, $i = 1, 2, \ldots, n$, $f(y) : \mathbb{R}^n \to \mathbb{R}$ is a cost function, $y = g(x) = [y_1, y_2, \ldots, y_n]^\top$ denotes the output vector with $y_i = g_i(x_i)$, $g_i(\cdot)$ is a nonlinear activation function, and $\theta = [\theta_1, \theta_2, \ldots, \theta_n]^\top \in \mathbb{R}^n$ is a constant input vector.

Inspired by the ideas in Refs. [6,31,32], in this section, we analyze the convergence dynamics of the model (24). Here, we introduce some definitions and lemmas on non-smooth analysis, variational analysis, and differential inclusions. We also give some mathematical description of the generalized neural network model to be studied.

Via the Clarke gradient, the generalized model for non-smooth cost function $f(\cdot)$ can be formulated as follows:

$$\begin{cases} \dot{x} \in -Dx - \partial f(y) + \theta, \\ y = g(x) = [g_1(x_1), g_2(x_2), \ldots, g_n(x_n)]^\top, \end{cases} \tag{25}$$

where $x = [x_1, x_2, \ldots, x_n]^\top \in \mathbb{R}^n$ denotes the state variable vector, $D = \mathrm{diag}[d_1, d_2, \ldots, d_n]$ is a positive diagonal matrix with $d_i > 0$, $i = 1, 2, \ldots, n$, function $f(\cdot)$ is a cost function, $y = [y_1, y_2, \ldots, y_n]^\top$ denotes the output variable vector, $g(\cdot)$ is the activation function to obtain the outputs, and $\theta = [\theta_1, \theta_2, \ldots, \theta_n]^\top$ denotes the input vector with $\theta_i \in \mathbb{R}$.

Here, we list some sets for the cost $f(\cdot)$ below.

1) $\mathcal{A}_1 = \{f(\cdot)$: f is strictly continuous and regular in $\mathbb{R}^n\}$;
2) $\mathcal{A}_2 = \{f(\cdot)$: f is subanalytic in $\mathbb{R}^n\}$;
3) $\mathcal{A}_3 = \{f(\cdot)$: f is ϕ-convex in $\mathbb{R}^n$ for some continuous function ϕ: $\mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^3 \to \mathbb{R}^+\}$.

Note that $\mathcal{A}_3 \subset \mathcal{A}_1$. We denote by $\mathcal{D}$ the set of all positive diagonal matrices.

Definition 4 (Absolute stability [6,10]) For a given cost function class $\mathcal{F}$, an activation function set $\mathcal{G}$, a diagonal matrix set $\mathcal{D}$, and an input set $\theta \subset \mathbb{R}^n$, we say that 1) the state of the system (25) is absolutely stable with respect to $(\mathcal{F}, \mathcal{G}, \mathcal{D}, \mathcal{E})$ if for any $f(\cdot) \in \mathcal{F}$, $g \in \mathcal{G}$, $D \in \mathcal{D}$, and $\theta \in \mathcal{E}$, each state trajectory $x(t)$ of the system (25) converges to certain $\bar{x} \in \mathbb{R}^n$, i.e., $\lim_{t\to\infty} x(t) = \bar{x}$; 2) the output of the system (25) is absolutely stable with respect to $(\mathcal{F}, \mathcal{G}, \mathcal{D}, \mathcal{E})$ if for any $f(\cdot) \in \mathcal{F}$, $g(\cdot) \in \mathcal{G}$, $D \in \mathcal{D}$, and $\theta \in \mathcal{E}$, each output trajectory $y(t)$ of the system (25) converges to certain $\bar{y} \in \mathbb{R}^n$, i.e., $\lim_{t\to\infty} y(t) = \bar{y}$.

Following two typical activation functions often used in neural network models. One is the sigmoid function: $g(\rho) = 1/(1 + \exp(-\rho))$ and the other is the saturation function: $s(\rho) = (|\rho + 1| - |\rho - 1|)/2$. In this paper, we provide a generalized class of activation functions, which can be described as a piecewise analytic, increasing, and bounded function class $\mathcal{G}$, including these two functions.

Definition 5 A function $h(z) = [h_1(z_1), h_2(z_2), \ldots, h_n(z_n)]^\top : \mathbb{R}^n \to \mathbb{R}^n$ is said to belong to the vector-valued function class $\mathcal{G}(n)$ of n-dimension if the following conditions are satisfied for $i = 1, 2, \ldots, n$:

1) $h_i(z)$ is bounded, continuous, and increasing;
2) There exist at most countable interval divisions for $\mathbb{R}$ by ordered numbers $\cdots < \rho_k^i < \rho_{k+1}^i < \cdots$ which can be infinity, where $I_k^i = (\rho_k^i, \rho_{k+1}^i)$ is an open interval or one-side infinite interval such that $h_i|_{I_k^i}$ is either analytic with $h_i'(z) > 0$ on $z \in [\rho_k^i, \rho_{k+1}^i]$, or a constant; there exists a locally finite such interval division of $\mathbb{R}$.

One can see that the sigmoid function: $h(z) = 1/(1 + \exp(-z))$, which is strictly increasing and analytic, and the saturation function: $s(\rho) = (|\rho+1| - |\rho-1|)/2$, both belong to the class $\mathcal{G}(1)$. Also, we can define some special subclass of $\mathcal{G}(n)$.

Definition 6 A function $h(\cdot)$ is said to belong to $\mathcal{G}_1(n)$ if $h(\cdot) \in \mathcal{G}(n)$ and $h_k(\cdot)$ is C^1 continuous with $h_k'(z) > 0$ for all $u \in \mathrm{dom}(h_k)$ and $k = 1, 2, \ldots, n$.

Definition 7 A function $h(\cdot)$ is said to belong to $\mathcal{G}_2(n)$ if $h(\cdot) \in \mathcal{G}_1(n)$ and $h_k(\cdot)$ is C^2 continuous for all $k = 1, 2, \ldots, n$.

It is clear that $h(z) = [h_1(z_1), h_2(z_2), \ldots, h_n(z_n)]^\top$ belongs to $\mathcal{G}(n)$ $(\mathcal{G}_{1,2}(n))$ if and only if each $h_i(z_i)$ is a scalar function of $\mathcal{G}(1)$ $(\mathcal{G}_{1,2}(1))$, respectively. One can see that the sigmoid activation function belongs to $\mathcal{G}_2(1)$. For $h(\cdot) \in \mathcal{G}(1)$, we divide all indices of the interval set $\{I_i\}$ into two subsets $\mathcal{I}_1$ and $\mathcal{I}_2$ such that $i \in \mathcal{I}_1$ implies that $h|_{I_i}$ is analytic with positive first-order derivative and $i \in \mathcal{I}_2$ implies that $h|_{I_i}$ is a constant. We also define $\mathcal{J}_1 = h(\{I_i : i \in \mathcal{I}_1\}) = \bigcup_{i\in\mathcal{I}_1} (h(\rho_i), h(\rho_{i+1}))$ which is a union of a number of open intervals, $\mathcal{J}_2 = h(\{I_i : i \in \mathcal{I}_2\}) = \bigcup_{i\in\mathcal{I}_2} h(\rho_i)$ which is a union of a number of isolated points, and $\mathcal{J}_3 = \bigcup_{i\in\mathcal{I}_1} \mathrm{Bd}(h(I_i)) - \mathcal{J}_2$. Thus, $\mathcal{J}_k$, $k = 1, 2, 3$, compose a disjoint splitting of $\mathcal{R}(h)$.

Due to the monotonicity and boundedness, we define the inverse function of $h(\cdot) \in \mathcal{G}(1)$ as a set-valued function:

$$h^{-1}(u) = \begin{cases} h^{-1}(u), & u \in \mathcal{J}_1 \bigcap \mathcal{J}_3, \\ I_i, & u \in \mathcal{J}_2 \text{ such that } u = h(I_i) \\ & \quad \text{for some } i \in \mathcal{I}_2. \end{cases}$$

It can be seen that the $h^{-1}(\cdot)$ is increasing, i.e., $x_1 > x_2$ holds for each $y_1 \in h(x_1)$ and $y_2 \in h(x_2)$ with $y_1 > y_2$; and on each open interval $h(I_k)$, $k \in \mathcal{I}_1$, $h^{-1}(\cdot)$ is also analytic (Theorem 1.5.3 in Ref. [33]).

The existence of the solution of a Cauchy problem of Eq. (24) is guaranteed by the following theorem given in Ref. [17].

Theorem 4 (*Theorem* 1 *in Ref.* [17]) *For each* $f \in A_1 \bigcap \mathcal{A}_2$, $g \in \mathcal{G}(n)$, $D \in \mathcal{D}$, *and* $\theta \in \mathbb{R}^n$, *the system* (25) *has at least one solution* $x(t, x_0)$ *initiated at* $x(0) = x_0$ *for any* $x_0 \in \mathbb{R}^n$ *and defined on* $[0, +\infty)$.

The following theorem tells that the system (25) has at least one equilibrium.

Theorem 5 (*Theorem* 2 *in Ref.* [17]) *For each* $f \in \mathcal{A}_1 \bigcap \mathcal{A}_2$, $g \in \mathcal{G}(n)$, $D \in \mathcal{D}$, *and* $\theta \in \mathbb{R}^n$, *the system* (25) *has at least one equilibrium* x^*, *i.e.,* $0 \in F(x^*) + \theta$ *holds.*

Theorem 4 indicates that there exists a solution but does not guarantee the uniqueness or continuous dependence. However, in the case that $g \in \mathcal{G}_2(n)$ and $f(\cdot)$ is ϕ-convex for some positive continuous function ϕ, then both uniqueness and continuous dependence of the solution is given by the following theorem.

Theorem 6 (*Theorem* 3 *in Ref.* [17]) *For any* $g(\cdot) \in \mathcal{G}_2(n)$, $f(\cdot) \in \mathcal{A}_2 \bigcap \mathcal{A}_3$, $D \in \mathcal{D}$, *and* $\theta \in \mathbb{R}^n$, *the system* (25) *has the unique solution for each initial condition* $x_0 \in \mathbb{R}^n$ *and this solution continuously depends on the initial condition.*

5.1 State convergence

The following theorem states the absolute stability of the state of the neural system.

Theorem 7 (*Theorem* 4 *in Ref.* [17]) (*Absolute stability of state*) *The system* (25) *is absolutely stable with respect to* $(\mathcal{A}_1 \bigcap \mathcal{A}_2, \mathcal{G}_1(n), \mathcal{D}, \mathbb{R}^n)$, *i.e., for any* $f(\cdot) \in \mathcal{A}_1 \bigcap \mathcal{A}_2$, $g(\cdot) \in \mathcal{G}_1(n)$, $D \in \mathcal{D}$, *and* $\theta \in \mathbb{R}^n$, *each trajectory* $x(t)$ *possesses some* $\bar{x} \in \mathbb{R}^n$ *such that* $\lim_{t\to\infty} x(t) = \bar{x}$.

Since $g(\cdot)$ is continuous, the output $y(t)$ must converge, i.e., $\lim_{t\to\infty} y(t) = g(\bar{x})$. If uniqueness and continuous dependence can be guaranteed, we can conclude that $x(t)$ converges to some equilibrium of the system (25).

Proposition 1 *For any* $f \in \mathcal{A}_2 \bigcap \mathcal{A}_3$, $g(\cdot) \in \mathcal{G}_2(n)$, $D \in \mathcal{D}$, *and* $\theta \in \mathbb{R}^n$, *the limit* $\bar{x}$ *of the trajectory* $x(t)$ *of the system* (24) *satisfies that* $\bar{x}$ *is an equilibrium of* $F(x) + \theta$.

In fact, according to the uniqueness and continuous dependence of the solution for the system (25) by Theorem 6, we have $x(\bar{x}, \tau) = \lim_{t\to\infty} x(x(x_0, t), \tau) = \lim_{t\to\infty} x(x_0, t + \tau) = \bar{x}$. So, $\mathrm{d}x(\bar{x}, \tau)/\mathrm{d}\tau = 0$ for all $\tau \geqslant 0$. Letting $\tau = 0$, we have $0 \in F(\bar{x}) + \theta$.

With the same arguments as in the proof of Theorem 4.7 in Ref. [32], we similarly give the convergence rate as follows.

Proposition 2 *Under the assumptions in Theorem* 7, *let* ϑ *be the Łojasiewicz exponent near* $\bar{x}$. *There exist* $c_{1,2} > 0$, $T_{1,2,3} > 0$, *and* $k_1 > 0$ *such that*

1) *If* $\vartheta \in (1/2, 1)$, *then* $|x(t) - \bar{x}| \leqslant c_1(t+1)^{\frac{1-\vartheta}{2\vartheta-1}}$ *holds for all* $t \geqslant T_1$;
2) *If* $\vartheta = 1/2$, *then* $|x(t) - \bar{x}| \leqslant c_2 \exp(-k_1 t)$ *holds for all* $t \geqslant T_2$;
3) *If* $\vartheta \in [0, 1/2)$, *then* $x(t)$ *converges in finite time, i.e.,* $x(t) = \bar{x}$ *for all* $t \geqslant T_3$.

The following result is a direct consequence from Proposition 2 and the "perturbation idea" in Ref. [34].

Corollary 5 1) *If* $f(\cdot)$ *is* C^2 *near* $\bar{x}$, $g(\cdot) \in \mathcal{G}_2(n)$, $D \in \mathcal{D}$, *and* $\nabla F(\bar{x})$ *is nonsingular, then* $x(t)$ *converges exponentially; additionally, if* $f(\cdot) \in C^2(\mathbb{R}^n, \mathbb{R})$ *and* $g(\cdot) \in \mathcal{G}_2(n)$, *then for almost all* $\theta \in \mathbb{R}^n$, *any solution* $x(t)$ *of the following system:*

$$\dot{x} = -Dx - \nabla f(y) + \theta, \ y = g(x), \tag{26}$$

converges exponentially;

2) *Under the assumptions made in Theorem* 7, *if there exist some* $\delta_0 > 0$ *and* $\delta > 0$ *such that* $m(F(x) + \theta) > \delta_0$ *holds for all* $x \in B_\delta(\bar{x}) \backslash \{\bar{x}\}$, *then* $x(t)$ *converges in finite time.*

5.2 Output convergence

In this subsection, we analyze the convergence behavior of the system (25). In case that the activation function belongs to the class $\mathcal{G}(n)$, we give the following theorem on output convergence.

Theorem 8 (*Theorem* 5 *in Ref.* [17]) (*Absolute stability of output*) *The output* $y(t)$ *of the system* (25) *is absolutely stable with respect to* $(\mathcal{A}_1 \bigcap \mathcal{A}_2, \mathcal{G}(n), \mathcal{D}, \mathbb{R}^n)$, *i.e., for any* $f(\cdot) \in \mathcal{A}_1 \bigcap \mathcal{A}_2$, $g(\cdot) \in \mathcal{G}(n)$, $D \in \mathcal{D}$, *and* $\theta \in \mathbb{R}^n$, *each output trajectory* $y(t)$ *possesses* $\bar{y} \in \mathbb{R}^n$ *such that* $\lim_{t\to\infty} y(t) = \bar{y}$. *Moreover, if either* 1) $\partial f(y)$ *is a singleton at* $\bar{y}$, *or* 2) $\bar{y}_i \in \mathcal{J}_1^i \bigcup \mathcal{J}_3^i$ *holds for all* $i = 1, 2, \ldots, n$, *then the state* $x(t)$ *also converges.*

In fact, if $\partial f(\bar{y})$ is not a singlet set and $\bar{y}_i \in \mathcal{J}_2^i$ holds for some i, we cannot guarantee the convergence of the state, $x(t)$, as indicated by the following counterexample of one-dimensional system:

$$\begin{cases} \dot{x} \in -x - \partial f(y), \\ f(y) = 2|y - 1|, \\ y = g(x), \\ x(0) = 0. \end{cases} \tag{27}$$

That is,

$$\dot{x} \in -x + 2v(1 - g(x)),$$

where $v(\cdot)$ is a set-valued map as the convex closure of the sign function:

$$v(\rho) = \begin{cases} 1, & \rho > 0, \\ [-1, 1], & \rho = 0, \\ -1, & \rho < 0, \end{cases} \text{ and } g(x) = \frac{|x+1| - |x-1|}{2}.$$

Then, at the beginning, $x(0) = 0 \in (-1, 1)$, it becomes $\dot{x} = -x + 2$, which implies the solution is $x(t) = -2\exp(-t) + 2$ which increases; at $t = \ln 2$, $x(t) = 1$, then the equation becomes $\dot{x} \in -x + 2\zeta(t)$ where $\zeta(t) \in [-1, 1]$, which implies the solution becomes $x(t) = \exp(-(t - \ln 2))[1 + 2\int_0^{t-\ln 2}\exp(\tau)\zeta(\tau)\mathrm{d}\tau]$. Letting $\zeta(\tau) \geqslant 1/2$, we have

$$x(t) \geqslant \exp(-t + \ln 2)\left(1 + \int_0^{t-\ln 2}\exp(\tau)\mathrm{d}\tau\right) = 1$$

for all $t \geqslant \ln 2$.

Therefore, one can verify that the following:

$$x(t) = \begin{cases} -2\exp(-t) + 2, & t \leqslant \ln 2, \\ \exp(-t + \ln 2)[1 + 2\int_0^{t-\ln 2}\exp(\tau)\zeta(\tau)\mathrm{d}\tau], & \\ & t > \ln 2, \end{cases}$$

is one solution of the system (27), where $\zeta(\tau)$ can be any measurable function with $\zeta(\tau) \in [1/2, 1]$. It can be seen that the solution is not asymptotically stable if $\zeta(\tau) = \sin\tau/4 + 3/4$. Also, noting that $f(y)$ is a convex function, one can see that for cellular-type neural networks, a ϕ-convex, even actually convex function, is unable to guarantee the uniqueness. The non-uniqueness of the solution of differential equations (inclusions) actually induces a set-valued action, which means that the system can reach multiple targets from same initial state.

Let

$$L(y) = f(y) + \sum_{i=1}^{n} d_i \int_0^{y_i} g_i^{-1}(\rho)\mathrm{d}\rho - \theta^\top y.$$

Then, we have

Proposition 3 1) *If $\bar{y}_i \in \mathcal{J}_1^i$ for all $i = 1, 2, \ldots, n$, $L(y)$ is C^2 near $\bar{y}$, $D \in \mathcal{D}$, and $\nabla^2 L(\bar{y})$ is nonsingular, then $y(t)$ converges to $\bar{y}$ exponentially;*

2) *Under the assumptions made in Theorem 8, and $\bar{y} \in \Phi_\xi$, if there exist $\delta_0 > 0$ and $\delta > 0$ such that for each ξ' with $\Phi_{\xi'} \supset \Phi_\xi$, $m(\partial_{y^{\xi'}} L(y)) > \delta$ for all $y \in B_\delta(\bar{y}) \bigcap \Phi_{\xi'}$ with $y \neq \bar{y}$, then $y(t)$ converges to $\bar{y}$ in finite time.*

This proposition can be proven by the similar arguments as in Corollary 5.

5.3 Two examples: Sigmoid and saturation activation functions

The following two typical activation functions: sigmoid function $g(u) = 1/(1 + \exp(-u))$ is frequently used in the HNNs [2,3], and saturation function $s(u) = (|u + 1| - |u - 1|)/2$ is used in the CNNs [4,5].

Consider the following system:

$$\dot{x} \in -Dx - \partial f(y) + \theta,\ y = g(\Gamma x),$$

where $\Gamma = \mathrm{diag}[\gamma_1, \gamma_2, \ldots \gamma_n]$ with $\gamma_i > 0$, $i = 1, 2, \ldots, n$, being the scaling parameter corresponding to the ith variable and $g(\cdot)$ is either a sigmoid function or a saturation function. Note that the sigmoid function belongs to the class $\mathcal{G}_2(n)$ and the saturation function belongs to $\mathcal{G}(n)$ but does not belong to $\mathcal{G}_1(n)$. Therefore, by the results given in the previous subsection, we have

Theorem 9 (*see Theorem* 6 *in Ref.* [17])

1) *If $g(\cdot)$ is the sigmoid function, $f(\cdot) \in \mathcal{A}_1 \bigcap \mathcal{A}_2$, $D \in \mathcal{D}$, and $\theta \in \mathbb{R}^n$, then each state trajectory $x(t)$ is convergent;*
2) *If $g(\cdot)$ is the saturation function, $f(\cdot) \in \mathcal{A}_1 \bigcap \mathcal{A}_2$, $D \in \mathcal{D}$, and $\theta \in \mathbb{R}^n$, then each output trajectory $y(t)$ is convergent.*

Furthermore, the convergence rate can be obtained as consequences of the results above.

5.4 Application: Simple examples of seeking local minimum point of non-smooth function

Here, we give a simple example to illustrate the minimum-seeking capability of the following special model of the system (25) with a scaling parameter λ:

$$\begin{cases} \dot{x} \in -x - \partial f(y) + \theta, \\ y = g(\lambda x) = [g_1(\lambda x_1), g_2(\lambda x_2) \ldots, g_n(\lambda x_n)]^\top, \end{cases} \tag{28}$$

for certain non-smooth objective function $f(\cdot)$ in the discrete state space $\{0, 1\}^n$, where each $g_i(\cdot)$ is a sigmoid function or a saturation function. The minimum-seeking problem can be formulated as follows:

$$\begin{cases} \text{minimize } f(y_1, y_2, \ldots, y_n), \\ \text{subject to } y_i \in \{0, 1\},\ i = 1, 2, \ldots, n. \end{cases} \tag{29}$$

A local minimum point of the problem (29) is defined by the *Hamming neighborhood*. That is, for any $y \in \{0, 1\}^n$, its Hamming neighborhood is denoted by a set of n elements, $N(y) = \{z :\ z = [y_1, y_2, \ldots, 1 - y_i, y_{i+1}, \ldots, y_n]^\top$, for all $i = 1, 2, \ldots, n\}$, i.e., all elements lie at a Hamming distance of one to y: $\{y :\ f(y) = \min[f(y), f(z), z \in N(y)]\}$.

Let $y(t, \lambda)$ be the trajectory of the output of the system (28) and $\bar{y}(\lambda) = \lim_{t\to\infty} y(t, \lambda)$. Suppose $\{\lambda_j\}$ is such a sequence that $\lambda_j \nearrow \infty$ and $\bar{y}(\lambda_j)$ converges to some $\bar{y}$. In Ref. [34], it is proven that $\bar{y}$ is a local minimum points for a class of multilinear objective function $f(\cdot)$ if θ is selected sufficiently small and avoids in a set with zero measure. The key of the method lies in two respects. First, only such equilibria, which converge to $\{0,1\}^n$ as $\lambda \to \infty$, can be stable. This implies that for almost all initial data and sufficiently large λ, $y(\lambda, t)$ converges into a small neighborhood of objective set $\{0,1\}^n$. If $\bar{y} \in \{0,1\}^n$, avoiding θ in some set of zero measure, the parity condition is satisfied:

$$[\partial f(\bar{y})]_i < 0 \text{ if } \bar{y}_i = 1, \text{ and } [\partial f(\bar{y})]_i > 0 \text{ if } \bar{y}_i = 0. \quad (30)$$

Second, the parity condition (30) of such stable equilibrium actually implies local minimum.

We hope to apply this method to the case that objective functions could be non-smooth. Unfortunately, careful investigation suggests that it be impossible to succeed for all regular, strictly continuous, and subanalytic functions. Hence, we try to find the class of objective functions such that the model (28) can solve the problem (29). It can be seen that the second condition of the parity condition can be extended to a little larger class of functions.

Proposition 4 *If f is a lower-C^2 function, then the parity condition* (30) *implies a local minimum on the set* $\{0,1\}^n$.

For proof, see Ref. [16].

We consider a special objective function $f(y) = p(y) + |y - y^*|_1$, where $p(y) = \sum_{i,j=1}^{n} a_{ij} y_i y_j$ is a quadratic polynomial and y^* is a reference point. One can see that $f(y)$ is a regular, strictly continuous, subanalytic, and non-smooth function. Since $\rho^2 = \rho$ holds for all $\rho \in \{0,1\}$, we can rewrite $p(y)$ as a multi-affine function $\acute{p}(y) = \sum_{i=1}^{n} a_{ii} y_i + \sum_{i=1} \sum_{j>i} 2a_{ij} y_i y_j$. Therefore, $f(\cdot)$ can be regarded as a sum of a multi-affine function $\acute{p}$ and a L^1 distance between y and y^*. Then, the differential model to solve this optimization problem can be written as

$$\dot{x}_i \in -x_i - a_{ii} - 2\sum_{j\neq i} a_{ij} y_j - \text{sign}(y_i - y_i^*) + \theta_i,$$
$$y_i = g_i(\lambda x_i), \ i = 1, 2, \ldots, n. \quad (31)$$

Then, similar to the discussion in Ref. [34], we have

Proposition 5 *Suppose that*

1) *either $y_i^* \notin [0,1]$ holds for all $i = 1, 2, \ldots, n$;*
2) *or if $y_i^* \in [0,1]$ holds for some index i, then $a_{ii} > -2\sum_{j>i, a_{ij}<0} a_{ij} + 1$ or $a_{ii} < -2\sum_{j>i, a_{ij}>0} a_{ij} - 1$ holds.*

Select each activation function as the sigmoid function $g_i(\rho) = 1/(1 + \exp(-\rho))$ or the saturation function $g_i(\rho) = (|\rho + 0.5| - |\rho - 0.5| + 1)/2$. Then, there exists $\epsilon > 0$ such that for any $|\theta| < \epsilon$, except for a set with zero measure, for almost all initial conditions, $\bar{y}(\lambda)$ is near a local minimum point of the problem (29) *for a sufficiently large value of λ.*

In fact, it is not difficult to see that either the condition 1) or 2) implies that only $\bar{y} \in \{0,1\}$ could be the limit of any convergent sequence of equilibria of the system (31) $\bar{y}(\lambda_j)$ avoiding θ in a set of zero measure. According to Proposition 4, this proposition can be concluded.

As an illustration, to minimize $f(y) = a_{11}y_1^2 + a_{22}y_2^2 + 2a_{12}y_1y_2 + |y_1 - y_1^*| + |y_2 - y_2^*|$ over $\{0,1\}^2$, we use the following differential system:

$$\begin{cases} \dot{x}_1 \in -x_1 - a_{11} - 2a_{12}y_2 - \text{sign}(y_1 - y_1^*) + \theta_1, \\ \dot{x}_2 \in -x_2 - a_{22} - 2a_{12}y_1 - \text{sign}(y_2 - y_2^*) + \theta_2, \end{cases}$$
$$y_i = g_i(\lambda x_i), \ i = 1, 2, \quad (32)$$

where $v(\cdot)$ is the convex closure of the sign function. First, let $A = \begin{bmatrix} -1.5 & 0.2 \\ 0.2 & 1.1 \end{bmatrix}$ and $y^* = [0.1, 0.7]^\top$. Figure 1 indicates that the terminal limit $\bar{y}(\lambda)$ converge to a local minimum point $[1,0]^\top$ as λ goes to the infinity.

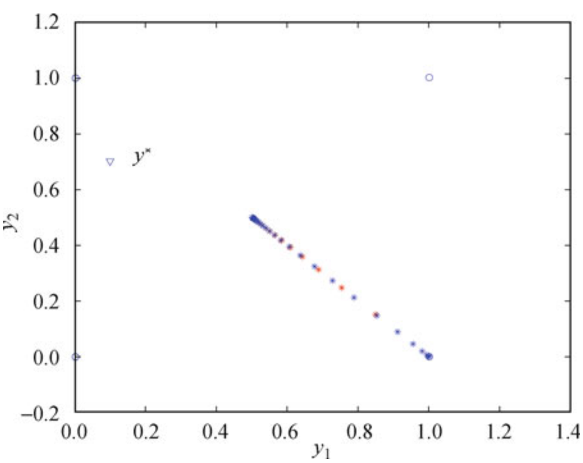


Fig. 1 The limits $\bar{y}(\lambda)$ converge to a local minimum point $[1,0]^\top$. We select a random θ with $|\theta| < 0.001$, $A = [-1.5\ 0.2;\ 0.2\ 1.1]$, $y^* = [0.1, 0.7]^\top$, random initial data, and $g(\cdot)$ as the sigmoid function (plotted in blue ∗) or saturation function (plotted in red ∗). The reference point y^* is targeted by a triangular and the vertices of the cube $\{0,1\}^n$ are indicated by o. λ is selected from 10^{-2} to 10^2.

Next, let $A = \begin{bmatrix} 1.0965 & -1.0787 \\ -1.0787 & 1.2535 \end{bmatrix}$ and $y^* = [-0.1, 1.2]^\top$. It can be seen that the objective function has two local minimum points: $[0,0]^\top$ and $[1,1]^\top$. Figure 2 indicates that the terminal limits $\bar{y}(\lambda)$ converge to such two local minimum points as λ goes to the infinity. This also indicates that the solution may approach the local minima instead of the global one.

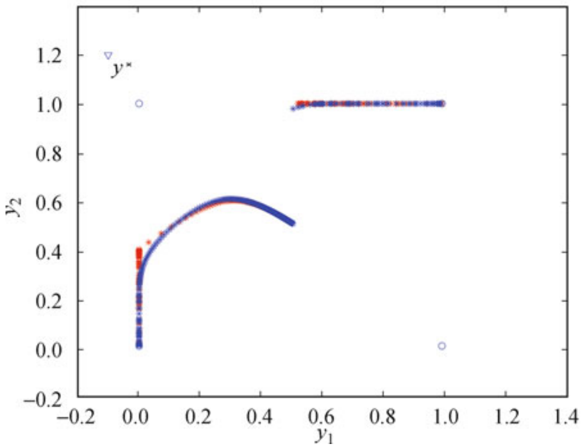


Fig. 2 The limits $\bar{y}(\lambda)$ converge to two local minimum points $[0,0]^\top$ and $[1,1]^\top$. We select a random θ with $|\theta| < 0.001$, $A = [1.0965\ \ -1.0787; -1.0787\ \ 1.2535]$, $y^* = [-0.1, 1.2]^\top$, random initial data, and $g(\cdot)$ as the sigmoid function (plotted in blue $*$) or saturation function (plotted in red $*$). The reference point y^* is targeted by a triangular and the vertices of the cube $\{0,1\}^n$ are indicated by o. λ is selected from 10^{-2} to 10^2.

6 Synchronization of linearly coupled neural networks with discontinuous right-hand sides

In this section, we discuss synchronization of linearly coupled systems with discontinuous right-hand sides.

Consider the following differential systems:

$$\frac{\mathrm{d}x^i(t)}{\mathrm{d}t} = f(x^i(t), t) + \sum_{j=1}^{N} a_{ij}\Gamma x^j(t), \quad i = 1, 2, \ldots, N, \tag{33}$$

where $x^i(t) \in \mathbb{R}^n$ denotes the state variable vector of the neuron i, f: $\mathbb{R}^n \times \mathbb{R}_+ \mapsto \mathbb{R}^n$ represents the right-hand of the node dynamics of the uncoupled system, and $a_{ij} \geqslant 0$ for $i, j = 1, 2, \ldots, N$ with $i \neq j$ denote the interaction between the two nodes by the manner that $a_{ij} > 0$ indicates that there is a directed edge from the node j to node i, $a_{ii} = -\sum_{j=1, j\neq i}^{N} a_{ij}$ for $i = 1, 2, \ldots, N$, and Γ represents the inner coupling configuration. In particular, picking $\Gamma = \mathrm{diag}[\gamma_1, \gamma_2, \ldots, \gamma_n]$. In this case, if $a_{ij} > 0$, then $\gamma_k \neq 0$ implies that the node j is linked to i by the kth component.

Mathematically, *complete synchronization* (*synchronization* for simplicity) with respect to the network (33) can be defined as

Definition 8 The network (33) of coupled nodes can achieve complete synchronization if

$$\lim_{t\to+\infty} \|x^i(t) - x^j(t)\| = 0, \quad i, j = 1, 2, \ldots, N.$$

In literature concerned with synchronization analysis of coupled differential systems, the following *QUAD* condition is assumed.

Definition 9 (QUAD condition or decreasing condition) A function f is said to satisfy the QUAD condition or decreasing condition if there exist some positive definite diagonal matrix P and a diagonal matrix Δ such that

$$(x-y)^\top P\big\{[f(x,t) - f(y,t)] - \Delta(x-y)\big\} \leqslant -\epsilon(x-y)^\top(x-y)$$

holds for some $\epsilon > 0$ and any $x, y \in \mathbb{R}^n$ with $x \neq y$.

We define the *weak-QUAD* condition as follows:

Definition 10 A function $f(x,t)$ is said to satisfy the weak-QUAD condition if there exist some positive definite diagonal matrix P and a diagonal matrix Δ such that

$$(x-y)^\top P\big\{[f(x,t) - f(y,t)] - \Delta(x-y)\big\} \leqslant 0$$

holds for any $x, y \in \mathbb{R}^n$ with $x \neq y$.

Based on the weak-QUAD condition, we can define a class of semi-QUAD functions for discontinuous functions as follows:

Definition 11 (Definition 7 in Ref. [19]) A discontinuous function $f(x,t)$ is called a semi-QUAD function if it satisfies the condition λ, and there is a continuous function $g(x,t)$, which satisfies the weak-QUAD condition such that

$$\lim_{t\to\infty}[f(x,t) - g(x,t)] = 0, \quad \forall\ x \in \mathbb{R}^n.$$

Furthermore, let $h(t) = \sup_{x\in\mathbb{R}^n} |f(t,x) - g(t,x)|$, then $h(t)$ is bounded and

$$\int_0^\infty h(t) < +\infty.$$

A simple example of the semi-QUAD functions is

$$f(x,t) = x + e^{-t}\mathrm{sign}(x),$$

where the corresponding limit function is $g(x,t) = x$, and $h(t) = e^{-t}$.

To conceive a distance from the collective states of the coupled system (33) to the synchronization subspace:

$$S = \{[x^{1\top}, x^{2\top}, \ldots, x^{N\top}] \in \mathbb{R}^{nN} :\ x^i = x^j,\ \forall\ i \neq j\},$$

we introduce certain structural matrices.

Definition 12 [35] Let $M_{m\times n}(\mathbb{R})$ be the class of $m\times n$ real matrices, and $T_1(k,K)$ be the set of matrices with entries in $M_{k\times k}(\mathbb{R})$ such that the sum of the entries in each row is equal to K for some $K \in M_{k\times k}(\mathbb{R})$. Let $M_2(k)$ be the set of matrices M with entries in $\{\alpha I_k : \alpha \in \mathbb{R}\}$ such that each row of M contains zeros and exactly one αI_k and one $-\alpha I_k$ for some nonzero α, and for any pair of indices i and j, there exist indices $i_1, i_2, \ldots, i_l$ with $i_1 = i$ and $i_l = j$ such that for all $1 \leqslant q < l$, $M_{p,i_q} \neq 0$ and $M_{p,i_{q+1}} \neq 0$ for some integer p.

In the following, we assume that the discontinuous region of f is composed of a series of smooth hypersurfaces of dimension m $(m < n)$, denoted by S_i, $i = 1, 2, 3, \ldots$ (may be empty or infinite). Suppose that $S_i = \{(y, t) \in \mathbb{R}^n \times \mathbb{R}_+ |\ \varphi^i(y, t) = 0\}$, with $\varphi^i(y, t) \in C^1(\mathbb{R}^n \times \mathbb{R}_+)$. Then, the continuous region of f is a series of connected regions divided by the S_i.

Let G_i^+ (G_i^-) be a connected region such that

1) $f(x, t)$ is continuous on G_i^+ (G_i^-);
2) $S_i \subset \partial G_i^+$ $(\partial G_i^-) \subset \bigcup_k S_k$;
3) $\varphi_i(x, t) > 0$ (< 0) on G_i^+ (G_i^-).

With these assumptions and notations, for each i, G_i^+ and G_i^- are two different regions with S_i as their common boundary. For the convenience of the later use, in the following, we summarize the properties of f.

Definition 13 Let $F(x, t) = \mathcal{K}[f](x, t)$ be the set valued map generated by f, then f is said to satisfy the condition λ if the following conditions are satisfied:

1) $F(x, t)$ satisfies the basic conditions;
2) For any $x \in \mathbb{R}^n$ and $t \in \mathbb{R}_+$, $\|F(x, t)\| \leqslant a(t)\|x\| + b(t)$, where $a(t)$, $b(t)$ are functions defined on $\mathbb{R}_+$ which are integrable on any finite interval of t;
3) For any $x, y \in \mathbb{R}^n$ and $t \in \mathbb{R}_+$, $\|F(x, t) - F(y, t)\| \leqslant h(\|x - y\|)$, where h: $\mathbb{R}_+ \mapsto \mathbb{R}_+$ is continuous and nondecreasing;
4) $B(S_i, \delta)$ and $B(S_j, \delta)$ have no intersection for all $i \neq j$ and some $\delta > 0$;
5) On the closure of each G_i^+ and each G_i^-, there exists a continuous extension of f;
6) On each S_i, $\nabla\varphi^i(y, t) \neq 0$ for almost all t, where $\nabla\varphi^i(y, t) = \left[\dfrac{\partial\varphi^i}{\partial y_1}, \dfrac{\partial\varphi^i}{\partial y_2}, \ldots, \dfrac{\partial\varphi^i}{\partial y_n}\right]$.

In case that f is continuous, we always assume that f satisfies the condition λ with empty set of discontinuities and $F(x, t) = \mathcal{K}[f](x, t) = f(x, t)$.

Following a similar arguments in Ref. [23], it is not difficult to verify that if f satisfies condition λ, then the solution of the coupled networks exists and can be extended to $\mathbb{R}_+$ for any initial value at $t = 0$. Under the assumption that f satisfies condition λ, we can define the solution of system (33) by the following differential inclusion:

$$\frac{\mathrm{d}x^i}{\mathrm{d}t} \in F(x^i(t), t) + \sum_{j=1}^{N} a_{ij}\Gamma x^j(t), \quad i = 1, 2, \ldots, N, \quad (34)$$

where $F(x, t) = \mathcal{K}[f](x, t)$.

The following results guarantee the synchronization of the coupled system (33).

Theorem 10 (*Theorem* 1 *in Ref.* [19]) *Suppose that $f(x, t)$ satisfies condition λ. If for any given initial value $\tilde{x}_0 \in \mathbb{R}^{nN}$, letting $\tilde{x}(t) = [x^1(t)^\top, x^2(t)^\top, \ldots, x^N(t)^\top]^\top$ be the solution from $\tilde{x}_0$, there exists an $(N-1) \times N$ matrix $M \in M_2(n)$, an $N \otimes N$ matrix $T \in T_1(n, K)$, and a positive definite matrix $P \in M_{n(N-1)\times n(N-1)}(\mathbb{R})$ such that the symmetric part of $PS(M, T + A \otimes \Gamma)$ is negative definite and*

$$\sup_{t\geqslant 0}\left\{\int_0^t \sup_{\xi_s\in\tilde{F}(\tilde{x}(s),s)} \left\{\tilde{x}(s)^\top M^\top PM\big[\xi(s) - T\tilde{x}(s)\big]\right\}\mathrm{d}s\right\} < +\infty, \quad (35)$$

where $\tilde{F}(\tilde{x}, t) = \{[\xi^1(t)^\top, \xi^2(t)^\top, \ldots, \xi^N(t)^\top]^\top,\ \xi^i(t) \in F(x^i, t),\ i = 1, 2, \ldots, N\}$, then we have

$$\lim_{t\to+\infty} \|x^i(t) - x^j(t)\| = 0, \quad i, j = 1, 2, \ldots, N.$$

In general, it is difficult to verify the conditions proposed in Theorem 10 directly, since $\sup_{\xi(t)\in\tilde{F}(\tilde{x},t)} \left\{\tilde{x}^\top M^\top PM\big[\xi(t) - T\tilde{x}\big]\right\} \leqslant 0$ may not hold globally.

In the following, we derive two corollaries as direct consequences from Theorem 10. These corollaries will provide some more testable but tighter conditions.

In particular, let

$$M = \begin{bmatrix} -I_n & I_n & & 0 \\ \vdots & & \ddots & \\ -I_n & 0 & & I_n \end{bmatrix}, \quad \Gamma = \begin{bmatrix} I_{n_1} & 0 \\ 0 & 0 \end{bmatrix}$$

with $n_1 \leqslant n$, $P = I_N \otimes I_n$ and $T = I_N \otimes \Delta$ with $\Delta = \begin{bmatrix} LI_{n_1} & 0 \\ 0 & -\alpha I_{n-n_1} \end{bmatrix}$ for some constants $L \in \mathbb{R}$ and $\alpha > 0$. Then, we have

$$S(M, T + A \otimes \Gamma) = (LI_{N-1} + B) \otimes \Gamma - \alpha I_{N-1} \otimes \begin{bmatrix} 0 & 0 \\ 0 & I_{n-n_1} \end{bmatrix},$$

where $B = [b_{ij}]_{N-1\times N-1}$ with $b_{ij} = a_{i+1,j+1} - a_{1,j+1}$.

As consequences of Theorem 10, we have

Corollary 6 *Suppose that $f(x, t)$ satisfies condition λ, Δ is defined as above. If there exist $C \in \mathbb{R}$, $L < -\lambda_{\max}(B^S)$ and $\alpha > 0$ such that*

$$\sup_{t\geqslant 0}\left\{\int_0^t \sup_{\substack{\xi^1(s)\in F(x^1(s),s)\\ \xi^i(s)\in F(x^i(s),s)}} \left\{[\xi^i(s) - \xi^1(s) - \Delta(x^i(s) - x^1(s))]^\top[x^i(s) - x^1(s)]\right\}\mathrm{d}s\right\} \leqslant C$$

holds for all $i = 2, 3, \ldots, N$, then the coupled system (33) synchronizes, i.e., $\lim\limits_{t\to+\infty} \|x^i(t) - x^j(t)\| = 0$, $\forall\ i, j = 1, 2, \ldots, N$.

Let Δ and $\xi(t)$ be defined as in Corollary 6 and denote

$$\tilde{T}_i = \big\{t \in \mathbb{R}_+ \big|\ \sup\{[\xi^i(t) - \xi^1(t) - \Delta(x^i(t) - x^1(t))]^\top \cdot[x^i(t) - x^1(t)]\} > 0\big\}, \quad (36)$$

and $T_i = \mathbb{R}_+\backslash\tilde{T}_i$. It can be verified that T_i and $\tilde{T}_i$ are measurable for each i. Then, from Corollary 6, we have

Corollary 7 *Suppose that $f(x,t)$ satisfies condition λ. Given an initial value $\tilde{x}_0 \in \mathbb{R}^{nN}$, if the trajectory $[x^1(t)^\top, x^2(t)^\top, \ldots, x^N(t)^\top]^\top$ satisfies that $\|x^i(t) - x^1(t)\|$ is uniformly bounded (the bound depends on the initial value but is uniform for all $t \in \mathbb{R}_+$) and $\mu(\tilde{T}_i) < +\infty$ for $i = 2, 3, \ldots, N$, then system (33) synchronizes.*

Now, we discuss semi-QUAD functions, and we have the following corollary from Theorem 10.

Corollary 8 *Let $f(x,t)$ be a semi-QUAD function whose limit function is $g(x,t)$ such that $g(x,t)$ satisfies the weak-QUAD condition for some $\overline{P}$, Δ. If there exists an $(N-1)\times N$ matrix $M \in M_2(n)$ such that the symmetric part of $PS(M, I_N \otimes \overline{P} + A \otimes \Gamma)$ is negative definite, and the trajectories of system (33) is bounded, then system (33) reaches complete synchronization.*

At last, we investigate networks of linearly coupled networks with nonidentical external inputs.

Consider the following networks of linearly coupled dynamical systems:

$$\frac{\mathrm{d}x^i(t)}{\mathrm{d}t} = f(x^i(t), t) + \sum_{j=1}^{n} a_{ij}\Gamma x^j(t) + p_i(t), \quad i = 1, 2, \ldots, N, \quad (37)$$

where f: $\mathbb{R}^n \times \mathbb{R} \mapsto \mathbb{R}^n$ is a continuous map satisfying the QUAD condition for some $\overline{P}$, Δ, $p_i(t)$ is an integrable function of t, which is not necessarily continuous. We have the following corollary.

Corollary 9 *Suppose $f(x,t)$ satisfies the QUAD condition for some $\overline{P}$, Δ. If for any initial value the trajectories of system (37) are bounded, and there exists an $(N-1)\times N$ matrix $M \in M_2(n)$ such that the symmetric part of $(I_{N-1} \otimes \overline{P})S(M, I_N \otimes \Delta + A \otimes \Gamma)$ is negative definite and*

$$\int_0^{+\infty} |p_i(t) - p_j(t)|\mathrm{d}t < +\infty, \quad \forall i \neq j,$$

then, the system (37) synchronizes.

As applications of the theoretical results, we present a scheme to synchronize a class of switching systems via coupling the switching-driving signals. Consider N switching systems with K switching dynamics:

$$\dot{y}^i = f_k(y^i, t), \quad \text{if} \quad \xi^i \in \Omega_k, \ k = 1, 2, \ldots, K. \quad (38)$$

Here, ξ^i denotes the switching signal of the ith system, which can be described by a differential equation:

$$\dot{\xi}^i = g(\xi^i, t), \quad i = 1, 2, \ldots, N,$$

where Ω_k, $k = 1, 2, \ldots, K$, denote domains on the space of the signals. We present a coupling scheme over a graph $\mathcal{G}$ with the Laplacian $A = [a_{ij}]_{i,j=1}^N$ via only coupling their switching-driving signals as follows:

$$\begin{cases} \dot{\xi}^i = g(\xi^i, t) + \sum_{j=1}^N a_{ij}\xi^j, \\ \dot{y}^i = f_k(y^i, t), \quad \text{if } \xi_k \in \Omega_k, \ k = 1, 2, \ldots, K, \end{cases} \quad (39)$$

where $i = 1, 2, \ldots, N$.

From Corollary 7, we can conclude that if the trajectories are uniformly bounded and the sum of the time intervals on which the signals are located in different region of Ω_k's is finite, the system (39) synchronizes.

In the following, we present a simple example to illustrate this synchronization scheme.

Consider a network of N all-to-all coupled three-dimensional individual dynamical systems that has the form (38) where

$$f_1(x,t) = \cos t,$$

and

$$f_k(x,t) = \begin{cases} -\frac{3}{2}x_k + |\sin t|x_{k+1} + |\cos t|x_{k+2} - 1, & x_1 < 0, \\ -\frac{3}{2}x_k + |\sin t|x_{k+1} + |\cos t|x_{k+2} + 1, & x_1 > 0, \end{cases}$$

for $k = 2, 3$, and $x_4 = x_2$, $x_5 = x_3$.

It is clear that, under the scheme (39), the inner coupling matrix $\Gamma = \mathrm{diag}[1, 0, 0]$. We can show that the conditions proposed in Corollary 7 are satisfied. Therefore, the coupled networks (39) synchronizes.

7 Discussion and conclusions

The stability of neural networks has always been a focus in the last two decades. Among them, the neural networks with discontinuous activations have been attracting increasing interests recently. In the pioneering works [10,28,36,37], the stability of the equilibrium of neural networks with discontinuous activations were investigated. Periodicity of neural networks with discontinuous activations was firstly studied in Ref. [38] to our best knowledge. Other works include Refs. [39–41] and others. In comparison, the model (6) in this paper is more general and includes HNNs, CGNNs and CNNs as special cases. The almost periodicity we considered is also more general than periodicity and fixed point attractor.

Different from the existing literature concerned with the stability of CGNN model, here we consider the positive dynamics that is more related to the original model proposed by Cohen and Grossberg [1]. In Refs. [42,43], the authors studied the global stability of nonnegative equilibrium or nonnegative periodic solution of the CGNNs with assuming $A_i(\rho) > 0$ for $\rho > 0$ and $A_i(0) = 0$, $i = 1, 2, \ldots, n$. To our best knowledge, Refs. [1,30] provided the pioneering study on the dynamics of such neural network model with assuming $A_i(\rho) > 0$ for all $\rho > 0$ and $A_i(0) = 0$ for all $i = 1, 2, \ldots, n$ and without considering any time delay. As far as we know, few works investigated the nonnegative dynamics of CGNNs with discontinuous activations that was done in this paper.

As far as we know, this is the first paper using the celebrated Łojasiewicz inequality for the stability analysis of neural networks. From then on, Forti et al. published several papers using the Łojasiewicz inequality to characterize the absolute stability of generalized neural networks for quadratic programming [44] and standard CNNs [31]. In these works, the concept of stability is quite different from other literature dealing with this issue because in this case the equilibria are not unique, even not countable. They obtained the stability without knowledge of equilibria as needed in other references. In this paper, we release the restriction of the analyticity by sub-analyticity. We use the recent result that was extended in Refs. [32,45] of this inequality to the non-smooth case, which made it possible to study the gradient inclusion with non-smooth cost functions. The similar idea was also used to solve linear and quadratic programming [44,46,47].

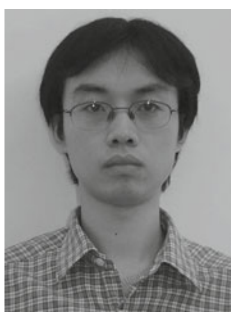
Synchronization, which means that the dynamics of nodes share the same time-spatial property, can be found in a wide variety of research fields, such as biological systems, chemistry, nonlinear optics, meteorology, etc. Till now, in most existing works concerning the model (33), the right-hand side is companied with the continuous assumption, i.e., the right-hand side of the uncoupled system is continuous or even Lipschitz continuous. To the best of our knowledge, the few exceptions include Refs. [48–50], in which discontinuous functions appear. In Ref. [48], discontinuity arises due to the temporal switching of the coupling. In Refs. [49,50], the authors studied the master-slave synchronization for a class of discontinuous bi-modal piecewise affine (PWA) systems. Both can be considered as some kind of switching systems. In Ref. [48], switching occurs in the coupling, while in Refs. [49,50], it lied in the uncoupled individual system. It should be noted that, in these papers, the QUAD condition still be required in discussion of synchronization, though the systems are discontinuous, which are relaxed in the current paper.

Acknowledgements This project was supported by the National Natural Science Foundation of China (Grant No. 60974015), Shanghai Guidance of Science and Technology (SGST) (No. 09DZ2272900), the Foundation for the Author of National Excellent Doctoral Dissertation of China (No. 200921), and also by Shanghai Rising-Star Program (No. 11QA1400400).

References

1. Cohen M A, Grossberg S. Absolute stability of global pattern formation and parallel memory storage by competitive neural networks. IEEE Transactions on Systems, Man, and Cybernetics, 1983, 13(5): 815–826
2. Hopfield J J. Neural networks and physical systems with emergent collective computational abilities. Proceedings of the National Academy of Sciences of the United States of America, 1982, 79(8): 2554–2558
3. Hopfield J J. Neurons with graded response have collective computational properties like those of two-stage neurons. Proceedings of the National Academy of Sciences of the United States of America, 1984, 81(10): 3088–3092
4. Chua L O, Yang L. Cellular neural networks: Theory. IEEE Transactions on Circuits and Systems, 1988, 35(10): 1257–1272
5. Chua L O, Yang L. Cellular neural networks: Applications. IEEE Transactions on Circuits and Systems, 1988, 35(10): 1273–1290
6. Forti M, Tesi A. Absolute stability of analytic neural networks: An approach based on finite trajectory length. IEEE Transactions on Circuits and Systems I: Regular Papers, 2004, 51(12): 2460–2469
7. Civalleri P P, Gilli L M, Pandolfi L. On stability of cellular neural networks with delay. IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications, 1993, 40(3): 157–165
8. Chen T P, Lu W L, Chen G R. Dynamical behaviors of a large class of general delayed neural networks. Neural Computation, 2005, 17(4): 949–968
9. Lu W L, Chen T P. Global convergent dynamics of delayed neural networks. In: Atay F M, ed. Complex Time-Delay Systems, Understanding Complex Systems, Chapter 7. Springer-Verlag, 2010
10. Forti M, Nistri P. Global convergence of neural networks with discontinuous neuron activations. IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications, 2003, 50(11): 1421–1435
11. Harrer H, Nossek J A, Stelzl R. An analog implementation of discrete-time cellular neural networks. IEEE Transactions on Neural Networks, 1992, 3(3): 466–476
12. Kennedy M P, Chua L O. Neural networks for nonlinear programming. IEEE Transactions on Circuits and Systems, 1988, 35(5): 554–562
13. Utkin V I. Variable structure systems with sliding modes. IEEE Transactions on Automatic Control, 1977, 22(2): 212–222
14. Łojasiewicz S. Une proprieté topologique des sous-ensembles analytiques réels, Colloques internationaux du C.N.R.S. Les Équations aux Dériveés Partielles, 1963, 117: 87–89
15. Łojasiewicz S. Sur la géométrie semi- et sous-analytique. Annales de L'Institut Fourier, 1993, 43(5): 1575–1595
16. Lu W L, Chen T P. Almost periodic dynamics of a class of delayed neural networks with discontinuous activations. Neural Computation, 2008, 20(4): 1065–1090
17. Lu W L, Wang J. Convergence analysis of a class of nonsmooth gradient systems. IEEE Transactions on Circuits and Systems I: Regular Papers, 2008, 55(11): 3514–3527
18. He X N, Lu W L, Chen T P. Nonnegative periodic dynamics of delayed Cohen-Grossberg neural networks with discontinuous activations. Neurocomputing, 2010, 73(13–15): 2765–2772

19. Liu B, Lu W L, Chen T P. New conditions on synchronization of networks of linearly coupled dynamical systems with non-Lipschitz right-hand sides. Neural Networks, 2012, 25(1): 5–13
20. Filippov A F. Classical solution of differential equations with multivalued right-hand side. SIAM Journal on Control, 1967, 5(4): 609–621
21. Aubin J P, Cellina A. Differential Inclusions. Berlin: Springer-Verlag, 1984
22. Aubin J P, Frankowska H. Set-Valued Analysis. Boston: Birkhauser, 1990
23. Aubin J P. Viability Theory. Boston: Birhauser, 1991
24. Haddad G. Monotine viable trajectories for functional differetial inclusions. Journal of Differential Equations, 1981, 42(1): 1–24
25. Hale J. Theory of Functional Differential Equations. New York: Springer-Verlag, 1977
26. Chen T P, Rong L B. Robust global exponential stability of Cohen-Grossberg neural networks with time delays. IEEE Transactions on Neural Networks, 2004, 15(1): 203–206
27. Lu W L, Chen T P. New conditions on global stability of Cohen-Grossberg neural networks. Neural Computation, 2003, 15(5): 1173–1189
28. Lu W L, Chen T P. Dynamical behaviors of Cohen-Grossberg neural networks with discontinuous activation functions. Neural Networks, 2005, 18(3): 231–242
29. Grossberg S. Biological competition: Decision rules, pattern formation, and oscillations. Proceedings of the National Academy of Sciences of the United States of America, 1980, 77(4): 2338–2342
30. Grossberg S. Nonlinear neural networks: Principles, mechanisms, and architectures. Neural Networks, 1988, 1(1): 17–61
31. Forti M, Tesi A. The Łojasiewicz exponent at equilibrium point of a standard CNN is 1/2. International Journal of Bifurcation and Chaos in Applied Sciences and Engineering, 2006, 16(8): 2191–2205
32. Bolte J, Daniilidis A, Lewis A. The Łojasiewicz inequality for nonsmooth functions with applications to subgradient dynamical systems. SIAM Journal on Optimization, 2007, 17(4): 1205–1223
33. Krantz S G, Parks H R. A Primer of Real Analytic Functions. 2nd ed. Boston: Birkhäuser, 2002
34. Vidyasagar M. Minimum-seeking properties of analog neural networks with multilinear objective functions. IEEE Transactions on Automatic Control, 1995, 40(8): 1359–1375
35. Wu C W, Chua L O. Synchronization in an array of linearly coupled dynamical systems. IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications, 1995, 42(8): 430–447
36. Forti M, Nistri P, Papini D. Global exponential stability and global convergence in finite time of delayed neural networks with infinite gain. IEEE Transactions on Neural Networks, 2005, 16(6): 1449–1463
37. Lu W L, Chen T P. Dynamical behaviors of delayed neural network systems with discontinuous activation functions. Neural Computation, 2006, 18(3): 683–708
38. Papini D, Taddei V. Global exponential stability of the periodic solution of the delayed neural networks with discontinuous activations. Physics Letters A, 2005, 343(1–3): 117–128
39. Wang Y, Zuo Y, Huang L, Li C. Global robust stability of delayed neural networks with discontinuous activation functions. IET Control Theory & Applications, 2008, 2(7): 543–553
40. Liu X, Cao J. On periodic solutions of neural networks via differential inclusions. Neural Networks, 2009, 22(4): 329–334
41. Wang L. Multistability of almost periodic solutions of neural networks with discontinuous activation functions. In: Proceedings of the Third International Workshop on Advanced Computational Intelligence. 2010, 16–20
42. Lin W, Chen T P. Positive periodic solutions of delayed periodic Lotka-Volterra systems. Physics Letters A, 2005, 334(4): 273–287
43. Lu W L, Chen T P. R_+^n-global stability of a Cohen-Grossberg neural network system with nonnegative equilibria. Neural Networks, 2007, 20(6): 714–722
44. Forti M, Nistri P, Quincampoix M. Convergence of neural networks for programming problems via a nonsmooth Łojasiewicz inequality. IEEE Transactions on Neural Networks, 2006, 17(6): 1471–1486
45. Bolte J, Daniilidis A, Lewis A. A nonsmooth Morse-Sard theorem for subanalytic functions. Journal of Mathematical Analysis and Applications, 2006, 321(2): 729–740
46. Liu Q, Wang J. A one-layer recurrent neural network with a discontinuous hard-limiting activation function for quadratic programming. IEEE Transactions on Neural Networks, 2008, 19(4): 558–570
47. Liu Q, Wang J. A one-layer recurrent neural network with a discontinuous activation function for linear programming. Neural Computation, 2008, 20(5): 1366–1383
48. Belykh I, Belykh V, Hasler M. Blinking model and synchronization in small-world networks with a time-varying coupling. Physica D: Nonlinear Phenomena, 2004, 195(1–2): 188–206
49. Pavlov A, Pogromsky A, van de Wouw N, Nijmeijer H. On convergence properties of piecewise affine systems. International Journal of Control, 2007, 80(8): 1233–1247
50. van de Wouw N, Pavlov A. Tracking and synchronisation for a class of PWA systems. Automatica, 2008, 44(11): 2909–2915

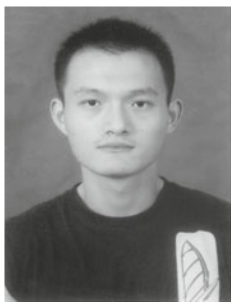


Wenlian LU received the B.S. degree in mathematics and the Ph.D. degree in applied mathematics from Fudan University, Shanghai, China, in 2000 and 2005, respectively. He was a postdoctoral researcher in the Max Planck Institute for Mathematics, Leipzig, Germany, from 2005 to 2007. From 2007 to 2011, he was an associate professor in the School of Mathematical Sciences and a senior member of the Center for Computational Systems Biology, Fudan Univer-

sity. Now, he is a professor in the School of Mathematical Sciences, Fudan University. He is an author and co-author of more than 40 papers in neural network, applied mathematics, and physics. His journal papers received more than 800 other citations from the SCI expanded and have been cited more than 2000 times from Google Scholar. He is currently an associate editor for Neurocomputing and a guest associated editor for International Journal of Bifurcation and Chaos. He is also a recipient of the Second Prize of National Natural Sciences Award of Shanghai in 2008 and the Young Researcher Award of the Asian-Pacific Neural Network Assembly in 2011. His current research interests include neural networks, nonlinear dynamical systems, and complex systems.

Tianping CHEN is a professor in the School of Mathematical Sciences, Fudan University, Shanghai, China. His current research interests include harmonic analysis, approximation theory, neural networks, signal processing, dynamical systems, and complex networks. He is Associate Editor for IEEE Transactions on Neural Networks. He is also a recipient of several awards, including the Second Prize of National Natural Sciences Award of China in 2002, the Outstanding Paper Award of IEEE Transactions on Neural Networks in 1997, and the Best Paper Award of the Japanese Neural Network Society in 1997.

Bo LIU received his B.A. degree in biotechnology from Zhejiang University, Hangzhou, China, and completed his Ph.D. degree in applied mathematics at Fudan University, Shanghai, China, in 2010 studying under Tianping Chen. Now he holds a postdoctoral position with School of Life Sciences, Fudan University. He has published more than 10 papers in complex networks and dynamical systems. His research interests include synchronization and consensus problems in complex networks, dynamical systems, and computational and system biology.



Xiangnan HE received the B.S. degree in Department of Mathematics from Zhejiang University, Hangzhou, China, in 2007. Now he is in Fudan University, Shanghai, China, as a Ph.D. candidate at Center for Computational Systems Biology. His research interests include neural networks, nonlinear dynamical systems, complex systems, and mathematics.