Selective Positive–Negative Feedback Produces the Winner-Take-All Competition in Recurrent Neural Networks

Shuai Li, Bo Liu, and Yangming Li

***Abstract*— The winner-take-all (WTA) competition is widely observed in both inanimate and biological media and society. Many mathematical models are proposed to describe the phenomena discovered in different fields. These models are capable of demonstrating the WTA competition. However, they are often very complicated due to the compromise with experimental realities in the particular fields; it is often difficult to explain the underlying mechanism of such a competition from the perspective of feedback based on those sophisticate models. In this paper, we make steps in that direction and present a simple model, which produces the WTA competition by taking advantage of selective positive–negative feedback through the interaction of neurons via *p*-norm. Compared to existing models, this model has an explicit explanation of the competition mechanism. The ultimate convergence behavior of this model is proven analytically. The convergence rate is discussed and simulations are conducted in both static and dynamic competition scenarios. Both theoretical and numerical results validate the effectiveness of the dynamic equation in describing the nonlinear phenomena of WTA competition.**

***Index Terms*— Competition, nonlinear, recurrent neural networks, selective positive–negative feedback, winner-take-all (WTA).**

I. Introduction

WINNER-TAKE-ALL (WTA) refers to the phenomenon in which agents in a group compete with each other for activation and only the one with the highest input stays active while all the others get deactivated. It widely exists in nature and society: for most plants, the main central stem, which only appears slightly stronger than the other (side) stems at the very beginning of the plant development, grows more and more strongly and eventually dominates over the others [1]. It has been observed in society that, once a firm gets ahead, it is more likely to become better and better over time while the others will fall further behind [2]. Neuroscientists find that the contrast gain in the visual systems comes from a WTA competition among overlapping neurons [3]. Other examples of the WTA competition include decision making in the cortex [4], [5], animal behaviors [6], cell fate competition [7], [8], etc.

There are various mathematical models presented to describe the WTA competition phenomena. The N species Lotka–Volterra model is often used to model the competitive interaction between species. Under elaborately selected parameters, the N species Lotka–Volterra model is able to demonstrate the WTA competition. This model is applied in [9] and [10] to generate the WTA competition. In the field of computational neuroscience, the FitzHugh–Nagumo model is often used to describe the dynamic interaction of neurons. In some situations, the neurons interact with each other in a WTA manner and the winner spikes. Inspired by this fact, this describing model is in turn applied to generate the WTA behavior [11]–[13]. In [11], the authors show that the model outputs oscillate under a set of parameter setups, and the oscillatory amplitude of the winner is greater than the spiking threshold while the amplitude of the losers are much less than the threshold. In [13], theoretical analysis on the stability and convergence of a large-scale WTA network is conducted by using nonlinear contraction theory. In addition, the authors show that the proposed network is stable for a range of parameters. In [14]–[17], the WTA problem is solved by modeling it as an optimization problem. In [14], a combinatorial optimization solver is proposed to solve the problem. In [15], the problem is modeled as a convex quadratic programming problem and a recurrent neural network developed for solving constrained quadratic programming is applied to solve it. Following the same problem formulation, the neural network proposed in [15] is simplified in [16] by tailoring the structure and taking advantage of the nonlinearity provided by a saturation function used in the model. In [17], a one-layer recurrent neural network is developed to solve the WTA competition by modeling the problem as a constrained linear programming. Although the optimization-based approach solves the problem accurately, operations such as saturation function, matrix multiplication of the state vector, etc. are often necessary in the iterations to approach the desired solution and thus are often computationally intensive. In addition, the resulting dynamics is often complicated and are often difficult to explain the WTA mechanism from its dynamic equations.

Although many models have been proposed to explain and generate the WTA behavior [9]–[21], these models are often very complicated due to the compromise with experimental

Manuscript received February 25, 2012; revised November 24, 2012; accepted November 24, 2012. Date of publication December 21, 2012; date of current version January 11, 2012.

S. Li is with the Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ 07030 USA (e-mail: lshuai@stevens.edu).

B. Liu is with the Department of Computer Science, University of Massachusetts, Amherst, MA 01003 USA (e-mail: boliu@cs.umass.edu).

Y. Li is with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109 USA (e-mail: ymli@umich.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TNNLS.2012.2230451

realities in the particular fields. Consequently, the essence of the WTA competition may be embedded in the interaction dynamics of those models, but difficult to tell from the sophisticated dynamic equations. Motivated by this, we develop a simple neural network model to solve the problem. The proposed model has a star communication topology between neurons and is scalable to situations with a large number of competitors. The model is described by an ordinary equation with the space dimension equal to the number of competitors. In addition, compared with the model using the Euclidean norm for global information exchange, the proposed model extends the results to the more general p-norm cases. Moreover, the proposed model demonstrates different robustness and convergence speed for different choice of parameters and thus allows the user to choose a set of parameters for better performance in applications.

The remainder of this paper is organized as follows. In Section II, preliminaries on p-norm and system stability are provided. In Section III, the analytical model is presented and the underlying competition mechanism is explained from a selective positive–negative feedback perspective. In Section IV, the convergence results are proved in theory. In Section V, a phenomenon relevant to the one-sided competition versus the closely matched competition is discussed and explained by using the proposed model. In Section VI, simulation examples are given to show the effectiveness of the proposed model. This paper is concluded in Section VII.

II. Preliminaries

In this section, we present some useful preliminaries for p-norm and system stability. We first present preliminaries on p-norm.

For an n-dimensional vector $x = [x_1, x_2, \ldots, x_n]^T$ with $x_i \in \mathbb{R}$ for $i = 1, 2, \ldots, n$, its p-norm, denoted as $\|x\|_p$, is defined as follows:

$$\|x\|_p = (|x_1|^p + |x_2|^p + \cdots + |x_n|^p)^{\frac{1}{p}} \tag{1}$$

where $p > 0$.

For $\|x\|_p$, the following partial derivative results hold:

$$\begin{aligned}\frac{\partial \|x\|_p^p}{\partial x_i} &= \frac{\partial(|x_1|^p + |x_2|^p + \cdots + |x_n|^p)}{\partial x_i}\\ &= \frac{\partial |x_i|^p}{\partial x_i}\\ &= \frac{\partial |x_i|^p}{\partial x_i}\\ &= p|x_i|^{p-1}\mathrm{sgn}(x_i)\end{aligned} \tag{2}$$

where sgn$(\cdot)$ is the sign function defined as

$$\mathrm{sgn}(u) = \begin{cases} 1, & \text{if } u > 0\\ 0, & \text{if } u = 0\\ -1, & \text{if } u < 0\end{cases} \tag{3}$$

with $u \in \mathbb{R}$.

Based on the partial derivative of $\|x\|_p$ shown in (2), the gradient of $(1/p)\|x\|_p^p$ can be obtained as

$$\nabla \frac{1}{p}\|x\|_p^p = \mathrm{sig}^{p-1}(x) \tag{4}$$

for $p > 0$ with the operator "$\mathrm{sig}^k(\cdot)$" defined as

$$\mathrm{sig}^k(x) = [|x_1|^k \mathrm{sgn}(x_1), |x_2|^k \mathrm{sgn}(x_2), \ldots, |x_n|^k \mathrm{sgn}(x_n)]^T \tag{5}$$

where $x = [x_1, x_2, \ldots, x_n]^T$. According to this definition, we can directly obtain

$$x^T \mathrm{sig}^k(x) = \|x\|_{k+1}^{k+1}. \tag{6}$$

The following inequalities hold for the estimation of p-norms for different p values:

$$\|x\|_p \le \|x\|_r \le n^{\frac{1}{r}-\frac{1}{p}}\|x\|_p \tag{7}$$

where $p > r > 0$, and n represents the dimension of the vector x.

The following results will be used later as tools for convergence analysis.

Definition 1 ([22]): A continuous function $\alpha : [0, a) \to [0, \infty)$ is said to belong to class $\mathcal{K}$ if it is strictly increasing and $\alpha(0) = 0$. It is said to belong to class $\mathcal{K}_\infty$ if $a = \infty$ and $\alpha(r) \to \infty$ as $r \to \infty$.

Lemma 1 ([22]): Let $\mathbb{D} \subset \mathbb{R}^n$ be a domain that contains the origin, and $V : [0, \infty) \times \mathbb{D} \to \mathbb{R}$ be a continuous differentiable function such that

$$\alpha_1(\|x\|) \le V(t, x) \le \alpha_2(\|x\|) \tag{8}$$

$$\dot{V} = \frac{\partial V}{\partial t} + \frac{\partial V}{\partial x} f(t, x) \le -W(x) \qquad \forall \|x\| \ge \mu > 0 \tag{9}$$

$\forall t \ge 0$ and $\forall x \in \mathbb{D}$, where α_1 and α_2 are class $\mathcal{K}$ functions and $W(x)$ is a continuous positive definite function. Take $r > 0$ such that $\mathbb{B}_r \subset \mathbb{D}$ and suppose that $\mu < \alpha_2^{-1}(\alpha_1(r))$. Then, for every initial state $x(t_0)$, satisfying $\|x(t_0)\| \le \alpha_2^{-1}(\alpha_1(r))$, there is $T \ge 0$ (dependent on $x(t_0)$ and μ) such that the solution of $\dot{x} = f(t, x)$ satisfies

$$\|x(t)\| \le \alpha_1^{-1}(\alpha_2(\mu)) \qquad \forall t \ge t_0 + T. \tag{10}$$

Moreover, if $\mathbb{D} = \mathbb{R}^n$ and α_1 belongs to class $\mathcal{K}_\infty$, then the result (15) holds for any initial state $x(t_0)$, with no restriction on how large μ is.

The following Lemma is also useful for the analysis of the ultimate behavior of a dynamic system.

Lemma 2 ([23]): Let $\Omega \subset \mathbb{D}$ be a compact set that is positively invariant with respect to $\dot{x} = f(x)$. Let $V : \mathbb{D} \to \mathbb{R}$ be a C^1-function such that $\dot{V}(x) \le 0$ on Ω. Let $\mathbb{E}$ be the set of all points in Ω such that $\dot{V}(x) = 0$. Let $\mathbb{M}$ be the largest invariant set in $\mathbb{E}$. Then, every solution starting in Ω approaches $\mathbb{M}$ as $t \to \infty$.

The mapping V in Lemma 2 is not necessary to be positive definite, which is a major difference from the Lyapunov function in conventional stability analysis of dynamic systems [23]. Instead, V is required to be a continuous differentiable function in Lemma 2, which is much looser than the requirement of positive definiteness.

III. WTA Neural Network

A. Neural Network-Based WTA Problem

In this paper, we are concerned with a neural network-based approach to find the winner in a group of competitors.

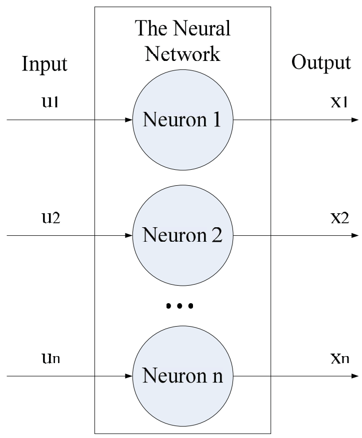


Fig. 1. Input–output block diagram of the proposed model.

Concretely, we want to find $i^* = \text{argmax}\{u_1, u_2, \ldots, u_n\}$ for the input vector $u = [u_1, u_2, \ldots, u_n]^T \in \mathbb{R}^n$ with $u_i \in \mathbb{R}$ by using neural networks, i.e., to find the winner among $u_1, u_2, \ldots, u_n$ by neural networks.

B. Neurodynamics

Inspired by the normalized recurrent neural network [24] and the use of the general norm on modeling the power of signals [25], we propose a recurrent neural network with a general p-norm as the regulation term for the WTA competition. The proposed model has the following dynamic for the ith neuron in a group of totally n neurons:

$$\dot{x}_i = c_0(u_i - c_1\|x\|_{p+1}^{p+1})|x_i|^p \text{sgn}(x_i) \tag{11}$$

where $x_i \in \mathbb{R}$ denotes the state of the i neuron, $u_i \in \mathbb{R}$ is the input and $u_i \geq 0$, $u_i \neq u_j$ for $i \neq j$, $p \in \mathbb{R}$, $p \geq 0$, $\|x\|_{p+1}$ is the $(p+1)$-norm of the state vector $x = [x_1, x_2, \ldots, x_n]^T$, $c_0 \in \mathbb{R}$, and $c_0 > 0$ and $c_1 \in \mathbb{R}$, $c_1 > 0$ are both constant.

The dynamic (11) can be written into the following compact form by stacking up the state for all neurons:

$$\dot{x} = c_0\big(u \circ \text{sig}^p(x) - c_1\|x\|_{p+1}^{p+1}\text{sig}^p(x)\big) \tag{12}$$

where $x = [x_1, x_2, \ldots, x_n]^T$, $u = [u_1, u_2, \ldots, u_n]^T$, the operator "$\circ$" represents the multiplication component-wise, i.e., $u \circ x = [u_1x_1, u_2x_2, \ldots, u_nx_n]^T$.

Remark 1: As shown in Fig. 1, the proposed neural network can be regarded as a black box. The ith neuron in the network receives input u_i and outputs x_i through the dynamic interactions with other neurons. As will be proved in Section IV, with the proposed model (11), the winner neuron $i^* = \text{argmax}\{u_1, u_2, \ldots, u_n\}$ can be identified by checking whether $\lim_{t\to\infty} x_i(t) = 0$ (if $\lim_{t\to\infty} x_i(t) \neq 0$, $i = i^*$ and otherwise, $i \neq i^*$).

Remark 2: The neurodynamics described by (11) is connected in a star topology (see Fig. 2). As can be observed from (11), the ith neuron is only connected to the central node, which computes the p-norm of the whole network state values.

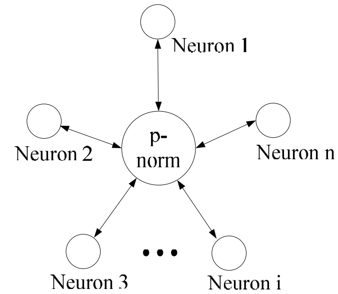


Fig. 2. Star topology diagram of the proposed model.

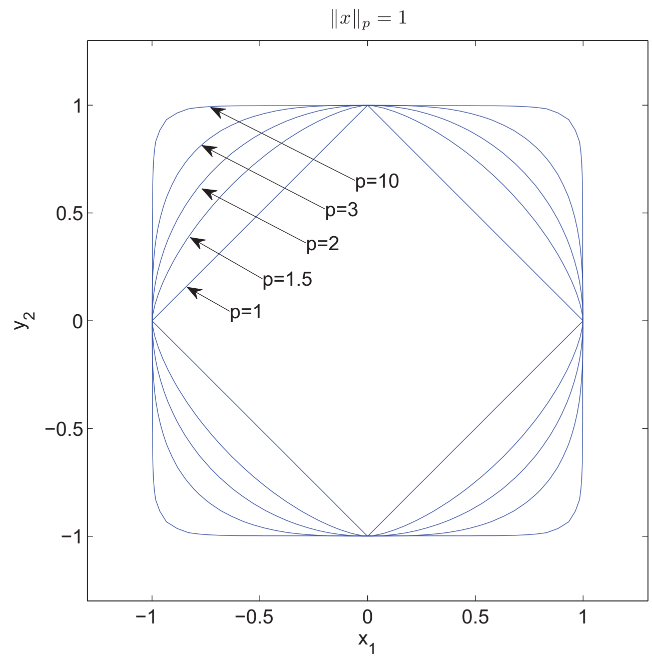


Fig. 3. Level set for $\|x\|_p = 1$ in 2-D space with different value of p.

The information exchange between neurons comes indirectly from their direct interaction with the central node.

Remark 3: Choosing the Euclidean norm, which corresponds to the special case of (11) by choosing $p = 1$, the proposed model reduces to the following in vector form:

$$\dot{x} = c_0(u \circ x - c_1\|x\|^2 x) \tag{13}$$

where $\|\cdot\|$ represents the Euclidean norm. In other words, the proposed model is a generalization from (13), which uses the Euclidean norm to the general p-norm scheme. Note that this generalization is not trivial as the p-norm function $y = \|x\|_p$ corresponds to different level sets (see Fig. 3) and thus leads to completely different dynamic evolution of x in (11).

Remark 4: Particularly for $p = 0$, the proposed model (11) reduces to the following in vector form

$$\dot{x} = c_0\big(u \circ \text{sgn}(x) - c_1\|x\|_1\text{sgn}(x)\big) \tag{14}$$

where $\text{sgn}(x) = [\text{sgn}(x_1), \text{sgn}(x_1), \ldots, \text{sgn}(x_n)]^T$ for $x = [x_1, x_2, \ldots, x_n]^T$ with $\text{sgn}(\cdot)$ being the sign for scalar entries. Note that this is a typical recurrent neural network with hard-limiting activation function and is often able to demonstrate a finite-time convergence as shown in [26] and [27]. In addition, it is noteworthy that the global information term $\|x\|_p^p$ reduces to $\|x\|_1$ in (14), which is the norm widely used in machine learning for data sparsification due to its ability to approximate the cardinality $\|x\|_0$ [28]–[30]. It will be an interesting topic to investigate the protocol from the perspective of sparse optimization.

IV. Convergence Results

In this section, theoretical results on the dynamic system (11) are presented. The rigorous proof of the main results needs the uses of LaSalle's invariant set principle [23], [31], local stability analysis, and the ultimate boundedness theory [22].

With Lemma 1, we are able to prove the following lemma for our main result.

Lemma 3: There exists $T \geq 0$ (dependent on $x(t_0)$ and μ) such that the solution of the neuron dynamic (12) satisfies

$$\|x(t)\| \leq \mu \qquad \forall t \geq t_0 + T \tag{15}$$

where $\mu = \mu_0(u_{\max} + \mu_1/c_1)^{(1/p+1)}$ with $\mu_1 > 0$ being any positive constant, $\mu_0 = \max\{n^{(1/2)-(1/p+1)}, 1\}$, and $u_{\max} = \max\{u_1, u_2, \ldots, u_n\}$.

Proof: We prove the result by following the framework of Lemma 1. Let $\mathbb{D} = \mathbb{R}^n$, $V = \frac{1}{2}x^T x$ and $\alpha_1(\|x\|) = \alpha_2(\|x\|) = \frac{1}{2}\|x\|^2 = V$. For V, we have

$$\begin{aligned}\dot{V} &= x^T\dot{x}\\ &= c_0 x^T\left(u \circ \mathrm{sig}^p(x) - c_1\|x\|_{p+1}^{p+1}\mathrm{sig}^p(x)\right)\\ &= c_0 x^T\left(\mathrm{diag}(u)\mathrm{sig}^p(x) - c_1\|x\|_{p+1}^{p+1}\mathrm{sig}^p(x)\right)\\ &\leq c_0(u_{\max} - c_1\|x\|_{p+1}^{p+1})x^T\mathrm{sig}^p(x)\\ &= c_0(u_{\max} - c_1\|x\|_{p+1}^{p+1})\|x\|_{p+1}^{p+1}.\end{aligned} \tag{16}$$

Note that $\mathrm{diag}(u)$ is a diagonal matrix and its largest eigenvalue is $u_{\max}$, from which the last inequality in (16) is obtained. According to the inequalities given in (7), we have

$$\|x\| \leq \begin{cases} n^{(\frac{1}{2}-\frac{1}{p+1})}\|x\|_{p+1}, & \text{for } p > 1\\ \|x\|_{p+1}, & \text{for } p \leq 1\end{cases}$$

where n represents the dimension of x. Inequality (17) is equivalent to

$$\|x\| \leq \mu_0\|x\|_{p+1} \tag{17}$$

where $\mu_0 = \max\{n^{(1/2)-(1/p+1)}, 1\}$. Together with (16), we have,

$$\dot{V} \leq c_0\left(u_{\max} - c_1\left(\frac{\|x\|}{\mu_0}\right)^{p+1}\right)\|x\|_{p+1}^{p+1}. \tag{18}$$

With this result, it is sufficient for $\left(u_{\max} - c_1(\|x\|/\mu_0)^{p+1}\right) \leq -\mu_1 \leq 0$, i.e., $\|x\| \geq \mu_0((u_{\max} + \mu_1)/(c_1))^{1/p+1}$, to guarantee the following inequality:

$$\dot{V} \leq -c_0\mu_1\|x\|_{p+1}^{p+1} \tag{19}$$

with $\mu_1 > 0$. These results fall into the framework of Lemma 1 by choosing $\mu = \mu_0((u_{\max} + \mu_1)/(c_1))^{1/p+1}$ and $W(x) = c_0\mu_1\|x\|_{p+1}^{p+1}$ in (9). Therefore, we conclude, according to Lemma 1, that

$$\|x\| \leq \alpha_1^{-1}(\alpha_2(\mu)) = \mu = \mu_0\left(\frac{u_{\max} + \mu_1}{c_1}\right)^{\frac{1}{p+1}} \quad \forall t > t_0 + T \tag{20}$$

for any initialization of x. This completes the proof. ■

This lemma reveals the state of the dynamic model (12) is ultimately bounded inside a compact super ball in $\mathbb{R}^n$ with radius μ. In other words, this super ball is positively invariant with respect the system dynamic (12). With this result on hand, we can confine our analysis in this super ball for further investigation of the system behaviors by applying LaSalle's invariant set principle.

Theorem 1: The solution of the system involving n dynamic neurons with the ith neuron described by (11) globally approaches 0 for $i \neq i^*$ and approaches $(u_{i^*}/c_1)^{1/p+1}e_{i^*}$ (or $-(u_{i^*}/c_1)^{1/p+1}e_{i^*}$) for $i = i^*$ as $t \to \infty$, provided any initialization with the initial value of the i^* neuron positive (or negative), where i^* denotes the label of the winner, i.e., $i^* = \mathrm{argmax}\{u_1, u_2, \ldots, u_n\}$.

Proof: There are three steps for the proof. The first step is to prove that the state variable ultimately converges to a set consisting of a limit number of points, the second step proves there are only two single points among the candidates that are stable, and the third step gives the initial conditions to decide which stable equilibrium point the system will converge to.

Step 1: According to Lemma 3, the state variable x in the system dynamic (12) is ultimately bounded by a compact super ball in $\mathbb{R}^n$ with radius μ implying this super ball is positively invariant with respect the system dynamic (12) and the super ball $\{x \in \mathbb{R}^n | \|x\| \leq \mu\}$ is qualified to be the set Ω in Lemma 2.

Let $V = V_1 + V_2$, with

$$V_1 = -\frac{1}{p+1}\sum_{i=1}^{n} u_i|x_i|^{p+1} \tag{21}$$

$$V_2 = \frac{c_1}{2(p+1)}\|x\|_{p+1}^{2p+2}. \tag{22}$$

Apparently, V is a C^1-function. For V_1, we have

$$\begin{aligned}\dot{V}_1 &= -\sum_{i=1}^{n} u_i|x_i|^p\mathrm{sgn}(x_i)\dot{x}_i\\ &= -\left(\mathrm{sig}^p(x)\right)^T\mathrm{diag}(u)\dot{x}.\end{aligned} \tag{23}$$

For V_2, we have

$$\begin{aligned}\dot{V}_2 &= \frac{c_1}{2(p+1)}\frac{d(\|x\|_{p+1}^{p+1})^2}{dt}\\ &= \frac{c_1}{(p+1)}\|x\|_{p+1}^{p+1}\frac{d(\|x\|_{p+1}^{p+1})}{dt}\\ &= \frac{c_1}{(p+1)}\|x\|_{p+1}^{p+1}(\nabla\|x\|_{p+1}^{p+1})^T\dot{x}\\ &= \frac{c_1}{(p+1)}\|x\|_{p+1}^{p+1}(p+1)\left(\mathrm{sig}^p(x)\right)^T\dot{x}\\ &= c_1\|x\|_{p+1}^{p+1}\left(\mathrm{sig}^p(x)\right)^T\dot{x}.\end{aligned} \tag{24}$$

Accordingly

$$\begin{aligned}\dot{V} &= \dot{V}_1 + \dot{V}_2\\ &= -\left(\left(\mathrm{sig}^p(x)\right)^T \mathrm{diag}(u) - c_1\|x\|_{p+1}^{p+1}\left(\mathrm{sig}^p(x)\right)^T\right)\dot{x}\\ &= -c_0\left\|u \circ \mathrm{sig}^p(x) - c_1\|x\|_{p+1}^{p+1}\mathrm{sig}^p(x)\right\|^2\\ &\le 0. \end{aligned} \tag{25}$$

According to the expression of $\dot{V}$ obtained in (25), we find $\mathrm{diag}(u)\mathrm{sig}^p(x) = c_1\|x\|_{p+1}^{p+1}\mathrm{sig}^p(x)$ by letting $\dot{V} = 0$. Note that $\mathrm{diag}(u)\mathrm{sig}^p(x) = c_1\|x\|_{p+1}^{p+1}\mathrm{sig}^p(x)$ is an eigenvector equation relative to the matrix $\mathrm{diag}(u)$ and the vector $\mathrm{sig}^p(x)$. Note that the eigenvalue and eigenvector pairs of the diagonal matrix $\mathrm{diag}(u)$ are u_i and ke_i for $i = 1, 2, \ldots, n$, with $k \in \mathbb{R}$ a scaling constant and e_i denoting an n-dimensional vector with the ith component 1 and all the other components 0. Therefore, the solution for $\mathrm{diag}(u)\mathrm{sig}^p(x) = c_1\|x\|_{p+1}^{p+1}\mathrm{sig}^p(x)$ is the solution of the two equations $c_1\|x\|_{p+1}^{p+1} = u_i$ and $\mathrm{sig}^p(x) = ke_i$ for $i = 1, 2, \ldots, n$ (i.e., $x_e = \pm(u_i/c_1)^{1/p+1}e_i$ by solving the two equations) and the trivial solution $x_e = 0$.

Define the set $\mathbb{M} = \{0, \pm(\frac{u_i}{c_1})^{\frac{1}{p+1}}e_i \,\mathrm{for}\, i = 1, 2, \ldots, n\}$. According to Lemma 2, every solution starting in $\boldsymbol{\Omega} = \{x \in \mathbb{R}^n | \|x\| \le \mu\}$ approaches $\mathbb{M}$ as $t \to \infty$. Together with the fact proven in Lemma 3 that every solution stays in $\boldsymbol{\Omega}$ ultimately, we conclude that every solution with any initialization approaches $\mathbb{M}$ as $t \to \infty$.

Step 2: The first step in this proof shows there are several candidate fixed points to stay for the dynamic system. In this step, we show that all those fixed points in $\mathbb{M}$ are unstable except for the one corresponding to the winner, i.e., $x = \pm(u_{k^*}/c_1)^{1/p+1}e_{k^*}$, where $k^* = \mathrm{argmax}\{u_1, u_2, \ldots, u_n\}$. To show the instability of some equilibrium points, we only need to show that there exists a streamline starting from that equilibrium point to elsewhere, which is equivalent to the fact that there exists a streamline from a nonequilibrium point to the equilibrium point for the new dynamic system with time t replaced by $-t$ (note that for an autonomous system, replacing t with $-t$ means that the initial state of the original system is identical to the ultimate state of the new system). Following this idea, we consider the following auxiliary system with reversed time direction:

$$\dot{x}_i = -c_0(u_i - c_1\|x\|_{p+1}^{p+1})|x_i|^p \mathrm{sgn}(x_i) \tag{26}$$

and we need to show that there exists a state x_0, the streamline of (26) starting from which ends up at the equilibrium point x_e.

For $x_e = 0$, we choose $x_0 = ke_1$, where $k > 0$ is a small positive constant and e_1 denotes an n-dimensional vector with the first component 1 and all the other components 0. Clearly, x_j for $j = 2, 3, \ldots, n$ starting from $x_0 = ke_1$ for the auxiliary system (26) stays at $x_j = 0$ in values since $\dot{x}_j = 0$ for them, while $\dot{x}_1 = -c_0(u_1 - c_1\|x\|_{p+1}^{p+1})|x_1|^p\mathrm{sgn}(x_1) < 0$ for $x_1 > 0$ and small enough k, which means x_1 keeps reducing to zero. Therefore, we conclude that $x_e = 0$ is unstable.

For $x_e = (u_i/c_1)^{1/p+1}e_i$ with $i \neq i^*$ (i^* denotes the winner neuron), we choose $x_0 = x_e + ke_{i*}$ with $k > 0$ being a constant to test the convergence. For $j \neq i^*$, the value of x_j of the auxiliary system (26) starting from $x_0 = x_e + ke_{i*}$ stays at $x_j = x_{ej}$ in values since $u_j|x_j|^p\mathrm{sgn}(x_j) = \|x\|_{p+1}^{p+1}|x_j|^p\mathrm{sgn}(x_j)$ (i.e., $\dot{x}_j = 0$). For $j = i^*$, $\dot{x}_{i^*} = -c_0(u_{i^*} - c_1\|x_e + ke_{i^*}\|_{p+1}^{p+1})|k|^p\mathrm{sgn}(k)$ at $x_j = k$. Note that $x_e = (u_i/c_1)^{1/p+1}e_i$ implies $c_1\|x_e + ke_{i^*}\|_{p+1}^{p+1} = u_i < u_{i^*}$. In addition, $\|x_e + ke_{i^*}\|_{p+1}^{p+1} \approx \|x_e\|_{p+1}^{p+1}$ for small enough $k > 0$. Accordingly, $\dot{x}_{i^*} < 0$ for small enough $k > 0$. Therefore, we conclude that $x_e = (u_i/c_1)^{1/p+1}e_i$ is unstable.

It is worth noting that for $i = i^*$, $x_0 = x_e + ke_i$. Note that $x_e + ke_i = \left((u_i/c_1)^{1/p+1} + 1\right)e_i$ and $\|x_e + ke_i\|_{p+1}^{p+1} = \left((u_i/c_1)^{1/p+1} + 1\right)^{p+1} > \left((u_i/c_1)^{1/p+1}\right)^{p+1} = \|x_e\|_{p+1}^{p+1}$ for any $k > 0$. Also, computing $p + 1$ norm on both sides of $x_e = (u_i/c_1)^{1/p+1}e_i$ generates $u_i = c_1\|x_e\|_{p+1}^{p+1}$. Together with $\|x_e + ke_i\|_{p+1}^{p+1} > \|x_e\|_{p+1}^{p+1}$, we get $u_i - c_1\|x_e + ke_i\|_{p+1}^{p+1} < 0$ for $k > 0$. Also, note that x_{ei} dominates over k for small enough $k > 0$ in $\mathrm{sgn}(x_{ei} + k)$ and $|x_{ei} + k|$; we thus conclude $\dot{x}_i = -c_0(u_i - c_1\|x_e + ke_i\|_{p+1}^{p+1})|x_{ei} + k|^p\mathrm{sgn}(x_{ei} + k) > 0$ for small enough positive constant k (recall $x_{ei} > 0$), which is different from the cases with $i \neq i^*$.

The instability of $x_e = -(u_i/c_1)^{1/p+1}e_i$ with $i \neq i^*$ (i^* denotes the winner neuron) can be similarly proved and thus omitted.

Step 3: Actually, we can conclude that the steady-state value is $(u_{i^*}/c_1)^{1/p+1}e_{i^*}$ if the initial state of the winner is positive, while it is $-(u_{i^*}/c_1)^{1/p+1}e_{i^*}$ if the initial value is negative by noting that $\dot{x}_{i^*} = 0$ when $x_{i^*} = 0$ in (11) for $i = i^*$, which means the state value x_{i^*} will never cross the critical value $x_i^* = 0$.

In summary, we conclude that every solution approaches $x_e = (u_{i^*}/c_1)^{1/p+1}e_{i^*}$ (or $x_e = -(u_{i^*}/c_1)^{1/p+1}e_{i^*}$) ultimately, provided any initialization with the initial value of the i^* neuron positive (or negative), where $i^* = \mathrm{argmax}\{u_1, u_2, \ldots, u_n\}$ and e_{i^*} being an n-dimensional vector with the i^*th component 1 and all the other components 0. Entrywise, the solution approaches $x_i = 0$ for $i \neq i^*$ and $x_{i^*} = \pm(u_{i^*}/c_1)^{1/p+1}$, which completes the proof. ■

V. Discussion on One-Sided Competition Versus Closely Matched Competition

In this section, we provide a comparison of an one-sided competition and a closely matched competition using the proposed model.

Imagine a football game. There are only two neurons, and the competition happens between the two teams. If one team has an overwhelming strength, it may demonstrate an obvious advantage over its opponent in a very early stage, whereas it often takes a relatively long time for a fierce competition between two closely matched teams to demonstrate a clear win or loss. Analogously, we may expect to observe a fast convergence in the WTA competition, where a distinct advantage for one neuron over others exists while a slow convergence for closely matched competitions. Theoretically speaking, this expectation corresponds to the statement that the convergence rate of the WTA competition has a strong dependence on the comparisons of the input value of the winner and the input values of the others. This phenomenon can be explained by (11).

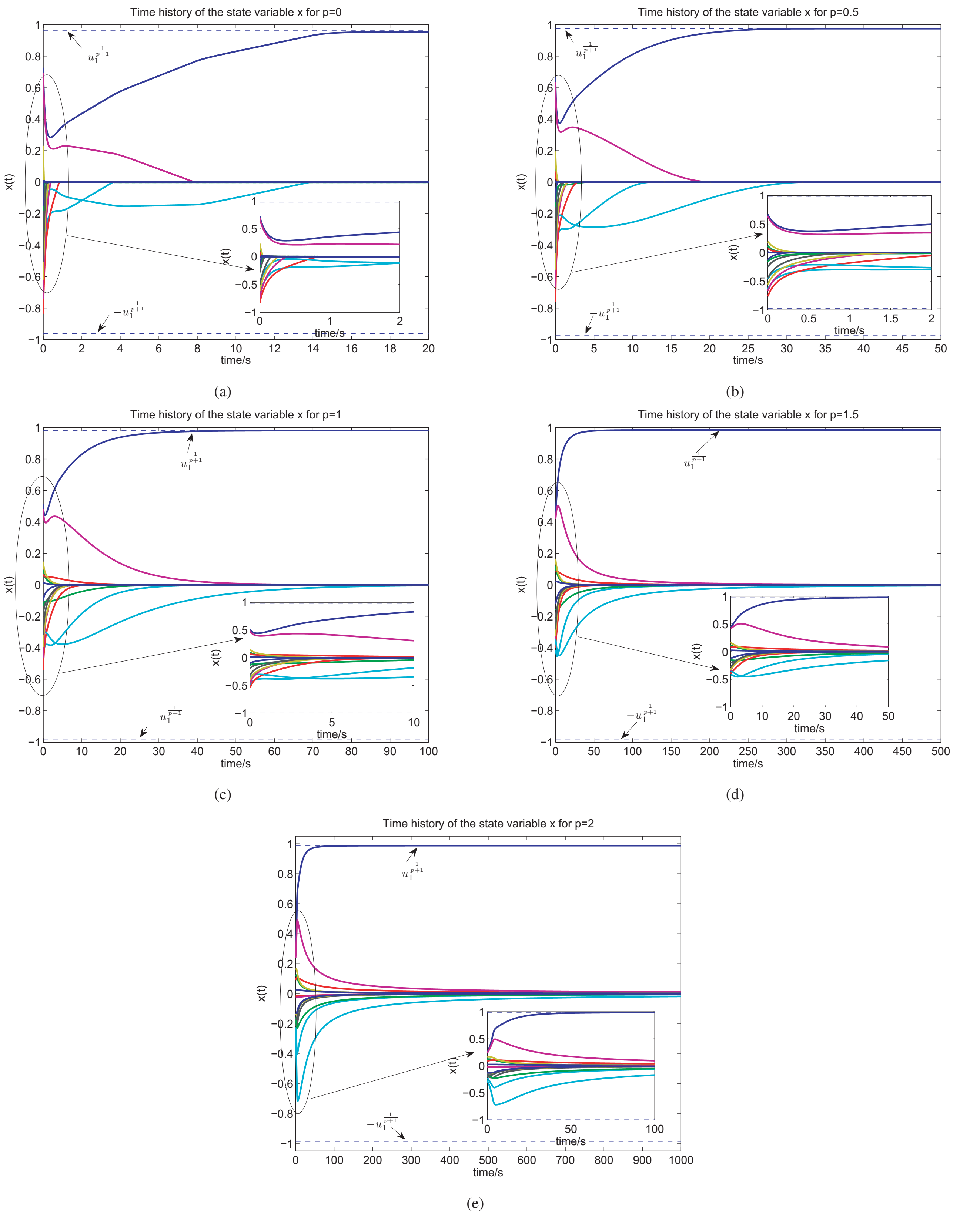


Fig. 4. Comparisons of the neural state trajectories in the static competition scenario with 15 neurons under $p = 0$, $p = 0.5$, $p = 1$, $p = 1.5$, $p = 2$, and $p = 2.5$.

For simplicity, we consider the case with parameter $c_0 = c_1 = p = 1$ in (11)

$$\dot{x} = u \circ x - \|x\|^2 x \tag{27}$$

where $\|x\|$ represents the Euclidean norm of the vector x. As there exists a strong nonlinearity in (27), it is difficult to analyze the convergence rate directly. Nevertheless, we can approximately analyze the convergence rate by considering

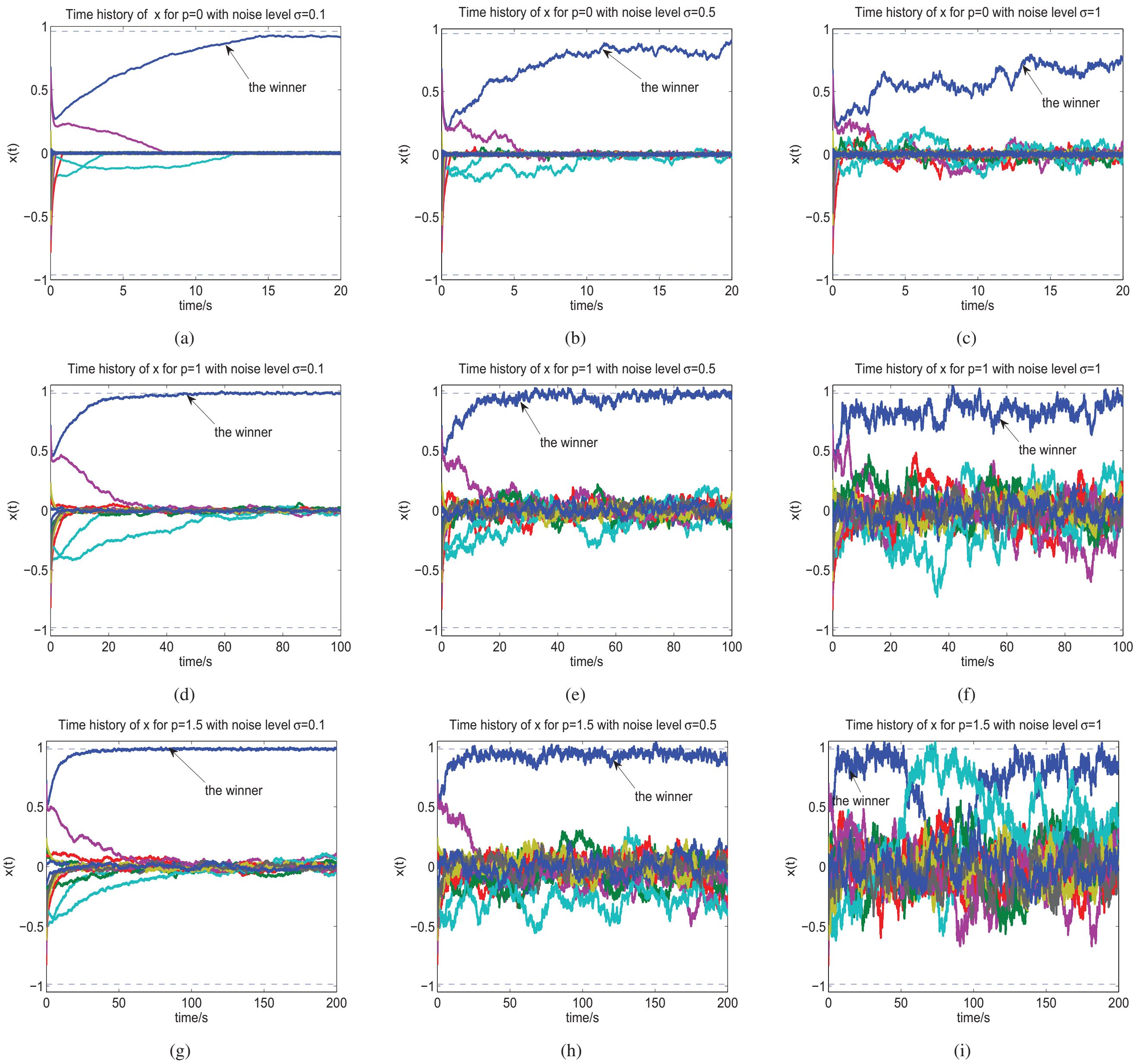


Fig. 5. Comparisons of the neural state trajectories of the static competition scenario in presence of additive noise under the norm parameter $p = 0$, $p = 1$, and $p = 1.5$ and noise level $\sigma = 0.1, \sigma = 0.5$, and $\sigma = 1$.

its linearization about the equilibrium point. According to Theorem 1, the stable equilibrium point is $x_e = \pm(u_{i^*}/c_1)^{1/p+1}e_{i^*} = \pm\sqrt{u_{i^*}}e_{i^*}$, where i^* denotes the label of the winner and e_{i^*} denotes an n-dimensional vector with the i^*th element being 1 and all the other elements being zeros. The linearized system around this fixed point is

$$\dot{x} = \big(\mathrm{diag}(u) - 2x_e x_e^T - \|x_e\|^2 I_n\big)x \tag{28}$$

where I_n is an $n \times n$ identity matrix. The system matrix of the above system is a diagonal matrix and its jth diagonal component, which is also its jth eigenvalue, is $(u_j - u_{i^*})$ for $j \neq i^*$ and $-2u_{i^*}$ for $j = i^*$. The linear system (28) has all eigenvalues negative and its convergence rate is determined by the largest eigenvalue $2u_{i^*}$. In other words, (27) has an approximate convergence rate $2u_{i^*}$.

VI. Simulation Examples

In this section, simulations are provided to illustrate the WTA competition phenomenon generated by the neural dynamic (11). We consider two scenarios: one is static competition, i.e., the input u is constant, and the other is dynamic competition, i.e., the input u is time-varying.

A. Static Competition

1) Simulation Setup: For the static competition problem, we consider time-invariant signals as the input. In the simulation, we consider a problem with $n = 15$ neurons. The input u is randomly generated between 0 and 1, which is $u = [0.9619, 0.0046, 0.7749, 0.8173, 0.8687, 0.0844, 0.3998, 0.2599, 0.8001, 0.4314, 0.9106, 0.1818, 0.2638, 0.1455, 0.1361]$, and the state is randomly initialized between -1 and 1, which is

$x(0) = [0.7386, 0.1594, 0.0997, -0.7101, 0.7061, 0.2441, -0.2981, 0.0265, -0.1964, -0.8481, -0.5202, -0.7534, -0.6322, -0.5201, -0.1655]$. In the simulation, we choose $c_0 = c_1 = 1$.

2) Convergence: Fig. 4 shows the evolution of state values of all neurons with time under different choice of the parameter p. From this figure, it can be observed that only a single state (corresponds to the first neuron, which has the largest value in u) has a nonzero value eventually and all the other state values are suppressed to zero. Also, the value of x_1 approaches ${u_5}^{1/p+1}$ (note that we choose $c_1 = 1$ in this simulation example), which is consistent with the claim made in Theorem 1 since the initial value $x_1(0) > 0$.

3) Convergence Speed: As can be observed in Fig. 4, it takes about 14 s for the model to converge for $p = 0$, 30 s for $p = 0.5$, 80 s for $p = 1$, 350 s for $p = 1.5$, and more than 1000 s for $p = 2$, which implies that a faster convergence can be obtained by choosing a smaller p in the proposed model for $p \geq 0$.

4) Robustness Against Additive Noise: In the real implementation of the proposed model, additive noise resulting from the computation error, quantization error, system noise, etc., may enter the input channel. In this situation, the steady-state value of 0 for the loser and $\pm(u_{i^*}/c_1)^{1/p+1}$ for the winner cannot be reached accurately. However, it can be expected that the neural states still converge to the vicinity of the desired values when the additive noise is within certain level. The readers are referred to [32]–[35] for theoretical investigations of the robustness of a general recurrent neural network against the uncertainties in delay, disturbance, etc. In this part, we explore such a property of the proposed model by simulation and compare the robustness of the proposed model under different choices of the parameter p.

The noise-polluted neural network model considered in this part writes for the ith neuron as

$$\dot{x}_i = c_0 \left(u_i - c_1 \|x\|_{p+1}^{p+1} \right) |x_i|^p \mathrm{sgn}(x_i) + \upsilon_i \tag{29}$$

where υ_i is a Guassian white noise with zero mean and σ^2 variance and is independent of υ_j for $i \neq j$. In the simulation, we choose three different noise levels, i.e., $\sigma = 0.1$, 0.5, and 1, to evaluate the performance of the proposed model. Fig. 5 plots the evolution of the neural state trajectories in presence of additive noise under the norm parameter $p = 0$, $p = 1$, $p = 1.5$, and noise level $\sigma = 0.1$, $\sigma = 0.5$, and $\sigma = 1$. From this figure, we can observe that the neural states are still able to converge to the vicinity of the desired value in the presence of additive noise with a small value of σ, which reveals the robustness of the proposed model to noisy inputs. With the increase of σ, the neural state value becomes more and more noisy. For the same level of noise (i.e., the same σ), it can be observed from Fig. 5 that the proposed model with a smaller p is less sensitive to the influence of the additive noise. Particularly for $\sigma = 1$, as shown in Fig. 5, the winner can still demonstrate a clear difference from the losers in state value for the cases with $p = 0$ and $p = 1$, while the state values almost mix together for the case with $p = 1.5$.

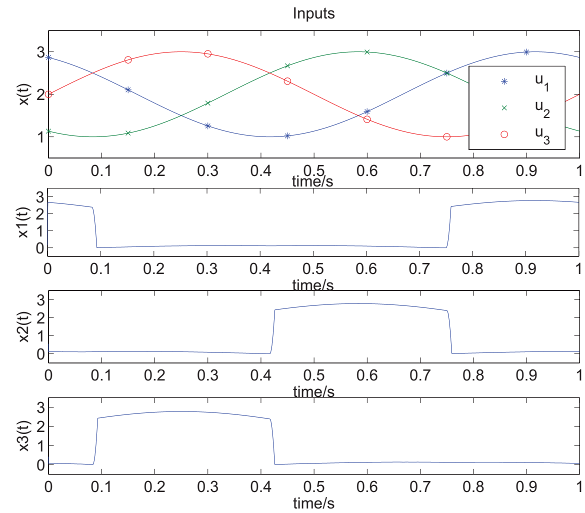


Fig. 6. Simulation results for the dynamic WTA competition.

B. Dynamic Competition

In this part, we consider the scenario with time-varying inputs. For the dynamic system (11), the convergence can be accelerated by choosing a large scaling factor c_0 and a small value of p for $p \geq 0$. The resulting fast response allows the computation of the winner in real time with time-varying input $u(t)$. In the simulation, we choose $c_0 = 10^4$, $c_1 = 1$, $p = 0$, and consider $n = 3$ neurons with input $u_i(t) = 1 + \sin(2\pi t + (2\pi i/3))$ for $i = 1, 2, 3$, respectively. The initial state values are randomly generated between 0 and 1. The four input signals and the absolute value of the state variables are plotted in Fig. 6. From this figure, we can see that the system can successfully find the winner in real time. Note that, according to the Theorem 1, the output value of the winner is u_{i^*} for $p = 0$ (recall that the state values are initialized greater than zero in this simulation), which is equal to the value of the input.

VII. Conclusion

In this paper, a recurrent neural network was proposed to explain and generate the WTA competition. In contrast to existing models, this dynamic equation features a simple expression and extends the case with Euclidean norm term for neural interaction to the more general p-norm cases. The fact that the state value of the winner converges to be active while the others get deactivated was proven theoretically. The convergence rate was discussed based on a local approximation. Simulations with both static inputs and dynamic inputs were performed. Convergence speed and robustness of the proposed model against additive noise were also explored by simulation. The results validates the effectiveness of the proposed model.

Acknowledgment

The authors would like to thank Leonard for bringing us deep insight into the model.

REFERENCES

[1] E. A. Dun, B. J. Ferguson, and C. A. Beveridge, "Apical dominance and shoot branching. divergent opinions or divergent mechanisms?" *Plant Physiol.*, vol. 142, no. 3, pp. 812–819, 2006.

[2] R. H. Frank and P. J. Cook, *The Winner-Take-All Society: Why the Few at the Top Get so Much More Than the Rest of Us*. London, U.K.: Penguin Books, 2010.

[3] D. K. Lee, L. Itti, C. Koch, and J. Braun, "Attention activates winner-take-all competition among visual filters," *Nature Neurosci.*, vol. 2, no. 4, pp. 375–81, 1999.

[4] L. Clark, R. Cools, and T. W. Robbins, "The neuropsychology of ventral prefrontal cortex: Decision-making and reversal learning," *Brain Cognit.*, vol. 55, no. 1, pp. 41–53, 2004.

[5] S. Kurt, A. Deutscher, J.M. Crook, F.W. Ohl, E. Budinger, C.K. Moeller, H. Scheich, and H. Schulze, "Auditory cortical contrast enhancing by global winner-take-all inhibitory interactions," *PLoS ONE*, vol. 3, no. 3, p. 12, 2008.

[6] M. Enquist and S. Ghirlanda, *Neural Networks and Animal Behavior*. Princeton, NJ: Princeton Univ. Press, Aug. 2005.

[7] D. Dubnau, "Winner takes all in a race for cell fate," *Molecular Syst. Biol.*, vol. 7, p. 558, Dec. 2011.

[8] L. Almeida, M. Idiart, and J. E. Lisman, "A second function of gamma frequency oscillations: An E%-max winner-take-all mechanism selects which cells fire," *J. Neurosci.*, vol. 29, no. 23, pp. 7497–7503, 2009.

[9] C. Benkert and D. Z. Anderson, "Controlled competitive dynamics in a photorefractive ring oscillator: 'Winner-takes-all' and the 'voting-paradox' dynamics," *Phys. Rev. A*, vol. 44, no. 7, pp. 4633–4638, 1991.

[10] H. G. Emilio, C. Lopez, S. Pigolotti, and K. H. Andersen, "Species competition: Coexistence, exclusion and clustering," *Philosoph. Trans. Royal Soc. A, Math. Phys. Eng. Sci.*, vol. 367, no. 3, pp. 3183–3195, 2008.

[11] W. Wang and J. E. Slotine, "Fast computation with neural oscillators," *Neurocomputing*, vol. 69, nos. 16–18, pp. 2320–2326, 2006.

[12] M. Oster, R. Douglas, and S. Liu, "Computation with spikes in a winner-take-all network," *Neural Comput.*, vol. 21, pp. 2437–2465, Sep. 2009.

[13] U. Rutishauser, R. J. Douglas, and J. E. Slotine, "Collective stability of networks of winner-take-all circuits," *Neural Comput.*, vol. 23, no. 3, pp. 735–773, 2011.

[14] Z. Xu, H. Jin, K. S. Leung, Y. Leung, and C. K. Wong, "An automata network for performing combinatorial optimization," *Neurocomputing*, vol. 47, nos. 1–4, pp. 59–83, 2002.

[15] S. Liu and J. Wang, "A simplified dual neural network for quadratic programming with its KWTA application," *IEEE Trans. Neural Netw.*, vol. 17, no. 6, pp. 1500–1510, Nov. 2006.

[16] X. Hu and J. Wang, "An improved dual neural network for solving a class of quadratic programming problems and its κ-winners-take-all application," *IEEE Trans. Neural Netw.*, vol. 19, no. 12, pp. 2022–2031, Dec. 2008.

[17] Q. Liu and J. Wang, "A one-layer recurrent neural network with a discontinuous activation function for linear programming," *Neural Comput.*, vol. 20, no. 5, pp. 1366–1383, 2008.

[18] A. L. Yuille and N. M. Grzywacz, "A winner-take-all mechanism based on presynaptic inhibition feedback," *Neural Comput.*, vol. 1, pp. 334–347, Sep. 1989.

[19] A. Simon and J. Winder, "A model for biological winner-take-all neural competition employing inhibitory modulation of NMDA-mediated excitatory gain," *Neurocomputing*, vols. 26–27, pp. 587–592, Jun. 1999.

[20] Y. Fang, M. A. Cohen, and T. G. Kincaid, "Dynamic analysis of a general class of winner-take-all competitive neural networks," *IEEE Trans. Neural Netw.*, vol. 21, no. 5, pp. 771–783, May 2010.

[21] J. P. F. Sum, C. S. Leung, P. K. S. Tam, G. H. Young, W. K. Kan, and L. W. Chan, "Analysis for a class of winner-take-all model," *IEEE Trans. Neural Netw.*, vol. 10, no. 1, pp. 64–71, Jan. 1999.

[22] H. Khalil, *Nonlinear Systems*. Englewood Cliffs, NJ: Prentice-Hall, Jan. 2002.

[23] A. Isidori, *Nonlinear Control Systems*, 3rd ed. New York: Springer-Verlag, 1995.

[24] D. P. Mandic and J. A. Chambers, "A normalised real time recurrent learning algorithm," *Signal Process.*, vol. 80, no. 9, pp. 1909–1916, 2000.

[25] D. R. Mandic, E. V. Papoulis, and C. G. Boukis, "A normalized mixed-norm adaptive filtering algorithm robust under impulsive noise interference," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 6. Apr. 2003, pp. 333–336.

[26] Q. Liu and J. Wang, "Finite-time convergent recurrent neural network with a hard-limiting activation function for constrained optimization with piecewise-linear objective functions," *IEEE Trans. Neural Netw.*, vol. 22, no. 4, pp. 601–613, Apr. 2011.

[27] Q. Liu and J. Wang, "A one-layer recurrent neural network with a discontinuous hard-limiting activation function for quadratic programming," *IEEE Trans. Neural Netw.*, vol. 19, no. 4, pp. 558–570, Apr. 2008.

[28] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Royal Stat. Soc., Ser. B*, vol. 58, no. 1, pp. 267–288, 1994.

[29] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York: Cambridge Univ. Press, 2004.

[30] T. R. Diethe, "Sparse machine learning methods with applications in multivariate signal processing," Ph.D. dissertation, Dept. Compt. Sci., University College London, London, U.K., 2010.

[31] J. P. LaSalle and S. Lefschetz, *Stability by Liapunov's Direct Method with Applications*. New York: Academic, 1973.

[32] Y. Shen and J. Wang, "Robustness analysis of global exponential stability of recurrent neural networks in the presence of time delays and random disturbances," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 1, pp. 87–96, Jan. 2012.

[33] Y. Xiao, Y. Liu, C. Leung, J. P. Sum, and K. Ho, "Analysis on the convergence time of dual neural network-based," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 4, pp. 676–682, Apr. 2012.

[34] B. Ammar, F. Cherif, and A. M. Alimi, "Existence and uniqueness of pseudo almost-periodic solutions of recurrent neural networks with time-varying coefficients and mixed delays," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 1, pp. 109–118, Jan. 2012.

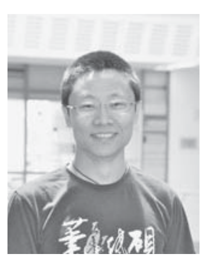
[35] Z. Wang, H. Zhang, and B. Jiang, "LMI-based approach for global asymptotic stability analysis of recurrent neural networks with various delays and structures," *IEEE Trans. Neural Netw.*, vol. 22, no. 7, pp. 1032–1045, Jul. 2011.

Shuai Li received the B.E. degree in precision mechanical engineering from the Hefei University of Technology, Hefei, China, and the M.E. degree in automatic control engineering from the University of Science and Technology of China, Hefei. He is currently pursuing the Ph.D. degree in electrical engineering with the Stevens Institute of Technology, Hoboken, NJ.

He is on the editorial board of the *International Journal of Distributed Sensor Networks*. His current research interests include dynamic neural networks, wireless sensor networks, robotic networks, machine learning, and other dynamic problems defined on a graph.

Bo Liu received the M.E. degree in computer engineering from the Stevens Institute of Technology, Hoboken, NJ. He is currently pursuing the Ph.D. degree in computer science with the University of Massachusetts, Amherst.

His current research interests include machine learning, data mining, graphical model, kernel approaches, and stochastic optimization.



Yangming Li received the B.E. and M.E. degrees in computer science from the Hefei University of Technology, Hefei, China, and the Ph.D. degree in automatic control engineering from the University of Science and Technology of China, Hefei.

He is a Post-Doctoral Fellow with the APRIL Laboratory, University of Michigan, Ann Arbor, and a Research Associate with the Robot Sensor & Human-Machine Interaction Laboratory, Institute of Intelligent Machines, Chinese Academy of Sciences, Beijing, China. His current research interests include robotics, including machine perception, tracking, object classification, navigation, and planning and map building.