Parametric Sensitivity and Scalability of *k*-Winners-Take-All Networks with a Single State Variable and Infinity-Gain Activation Functions*

Jun Wang[1,2] and Zhishan Guo[1]

[1] Department of Mechanical and Automation Engineering,
The Chinese University of Hong Kong, Shatin, N.T., Hong Kong
[2] Department of Computer Science and Engineering,
Shanghai Jiao Tong University, Minhang, Shanghai, China
{jwang,zsguo}@mae.cuhk.edu.hk

Abstract. In recent years, several k-winners-take-all (kWTA) neural networks were developed based on a quadratic programming formulation. In particular, a continuous-time kWTA network with a single state variable and its discrete-time counterpart were developed recently. These kWTA networks have proven properties of global convergence and simple architectures. Starting with problem formulations, this paper reviews related existing kWTA networks and extends the existing kWTA networks with piecewise linear activation functions to the ones with high-gain activation functions. The paper then presents experimental results of the continuous-time and discrete-time kWTA networks with infinity-gain activation functions. The results show that the kWTA networks are parametrically robust and dimensionally scalable in terms of problem size and convergence rate.

Keywords: K winners-take-all, recurrent neural networks, optimization, scalability, parametric sensitivity.

1 Introduction

As a generalization of the winner-take-all (WTA) operation [1], the k-winners-take-all (kWTA) operation selects the k largest inputs out of n inputs ($1 \leq k \leq n$). kWTA has been shown to be a computationally powerful operation compared with standard neural network models of threshold logic gates [2]. In addition, it has numerous applications in k-neighborhood classification, k-means clustering, sorting, machine learning, and data mining; e.g., decoding [1], image processing [3], computer vision [4],[5], feature extraction [6], and associative memories [7], mobile robot navigation [8], etc.

* The work described in this paper was supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project no. CUHK417608E).

L. Zhang, J. Kwok, and B.-L. Lu (Eds.): ISNN 2010, Part I, LNCS 6063, pp. 77–85, 2010.

When the number of inputs is large or a selection process has to be operated in real time, parallel algorithms and hardware implementation are desirable. In the past twenty years, many WTA and kWTA networks have been developed; e.g., [9]-[25]. In particular, a continuous-time kWTA network with a single state variable and guaranteed global convergence [23] was developed based on a quadratic programming formulation in [20]. It has the simplest architectural complexity due to its single state variable. A discrete-time counterpart was developed recently [25].

In this paper, we will present the simulation results of these kWTA networks to show their parametric robustness and dimension scalability in terms of convergence, which will validate and supplement previous theoretical results for designing the kWTA networks.

2 Problem Formulations

Generally, the kWTA operation can be defined or encoded as the following binary function

$$x_i = f(u_i) = \begin{cases} 1, \text{ if } u_i \in \{k \text{ largest elements of } u\}, \\ 0, \text{ otherwise}, \end{cases} \tag{1}$$

where $u = (u_1, u_2, \ldots, u_n)^T$ is the input vector and $x = (x_1, x_2, \ldots, x_n)^T$ is the output vector.

The kWTA solution can be determined by solving the following linear integer programming problem:

$$\begin{aligned} &\text{maximize } \sum_{i=1}^{n} u_i x_i \quad \text{or} \quad \text{minimize } -\sum_{i=1}^{n} u_i x_i, \\ &\text{subject to } \sum_{i=1}^{n} x_i = k, \\ &\qquad x_i \in \{0, 1\}, \quad i = 1, 2, \ldots, n. \end{aligned} \tag{2}$$

According to its total modularity property [26], the above linear integer programming problem is equivalent to the following linear programming problems if the kth and $(k+1)$th largest elements of u are different (i.e., the solution is unique):

$$\begin{aligned} &\text{minimize } \; -\sum_{i=1}^{n} u_i x_i \quad \text{or} \quad -u^T x, \\ &\text{subject to } \sum_{i=1}^{n} x_i = k, \\ &\qquad 0 \le x_i \le 1, \quad i = 1, 2, \ldots, n. \end{aligned} \tag{3}$$

It was proven in [20] that the kWTA problem is equivalent to the following quadratic integer programming problems:

$$\begin{aligned} &\text{minimize } \; \tfrac{a}{2} \sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} u_i x_i \text{ or } \tfrac{a}{2} x^T x - u^T x, \\ &\text{subject to } \sum_{i=1}^{n} x_i = k, \\ &\qquad x_i \in \{0, 1\}, \quad i = 1, 2, \ldots, n. \end{aligned} \tag{4}$$

where a is a positive constant.

Furthermore, according to [20], if the kth and $(k+1)$th largest elements of u (denoted as $\bar{u}_k$ and $\bar{u}_{k+1}$, respectively) are different, the quadratic integer programming problems is equivalent to the following quadratic programming problems:

$$\begin{aligned} &\text{minimize} \;\; \tfrac{a}{2}\sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} u_i x_i \text{ or } \tfrac{a}{2}x^T x - u^T x, \\ &\text{subject to} \sum_{i=1}^{n} x_i = k, \\ &\qquad\qquad 0 \le x_i \le 1, \quad i = 1, 2, \ldots, n, \end{aligned} \tag{5}$$

where $a \le \bar{u}_k - \bar{u}_{k+1}$ represents the input resolution.

3 Model Descriptions

A kWTA neural network was tailored from the simplified dual network [20]:

$$\text{state equation} \;\; \epsilon\frac{dv}{dt} = -Mv + g(Mv - v + s) - s, \tag{6}$$

$$\text{output equation} \;\; x = Mv + s, \tag{7}$$

where $\epsilon > 0$ is a scaling constant, $v \in \Re^n$ is the state vector, $M = (I_n - ee^T/n)/a$, $s = Mu + (k/n)e$, e is a column vector with all entries being ones, and $g(\cdot)$ is the piecewise linear activation function defined as:

$$g(x_i) = \begin{cases} 0, & x_i < 0_i, \\ x_i, & 0 \le x_i \le 1, \\ 1, & x_i > 1. \end{cases} \tag{8}$$

A kWTA network with a single state variable was developed based on the improved dual neural network [23] with the following following equations:

$$\text{state equation} \;\; \epsilon\frac{dz}{dt} = -\sum_{i=1}^{n} x_i + k, \tag{9}$$

$$\text{output equation} \;\; x_i = g\left(z + \frac{v_i}{a}\right), \quad i = 1, \cdots, n, \tag{10}$$

where $z \in \Re$ is the state variable. The exactly same model was reinvented one year after in [24].

Recently, a discrete-time counterpart of the above kWTA network [23] was presented [25],[24]..

$$\text{state equation} \;\; z(m+1) = z(m) - \beta(\sum_{i=1}^{n} x_i(m) - k), \tag{11}$$

$$\text{output equation} \;\; x_i(m) = g\left(z(m) + \frac{v_i}{a}\right), \quad i = 1, \cdots, n, \tag{12}$$

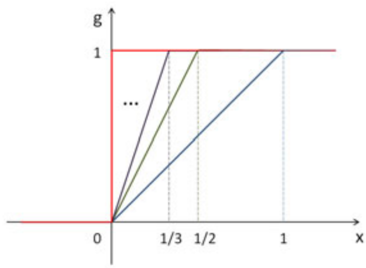

Fig. 1. The piecewise linear activation function with a positive gain parameter ξ

where $\beta > 0$ is the step size. It was proven that the discrete-time kWTA network is globally convergent if $0 < \beta < 2/n$ [25] or $0 < \beta < 2/\sqrt{n}$ [24].

As the output variables are supposed to be binary, the piecewise linear activation function $g(\cdot)$ can be extended with a positive gain parameter ξ as shown in(13) and Figure 1, where ξ is the slope of the linear part:

$$g(x_i) = \begin{cases} 0,\ x_i < 0_i, \\ \xi,\ 0 \leq x_i \leq 1/\xi, \\ 1,\ x_i > 1/\xi. \end{cases} \tag{13}$$

When the gain parameter approaches to positive infinity (i.e., $\xi \to +\infty$), the activation function becomes identical with that in the one of one-layer neural networks with hard-limiting (step) activation function in [22]. In the next section, it will be shown that the kWTA networks with positive infinity gain parameter ξ can still work well in kWTA operations.

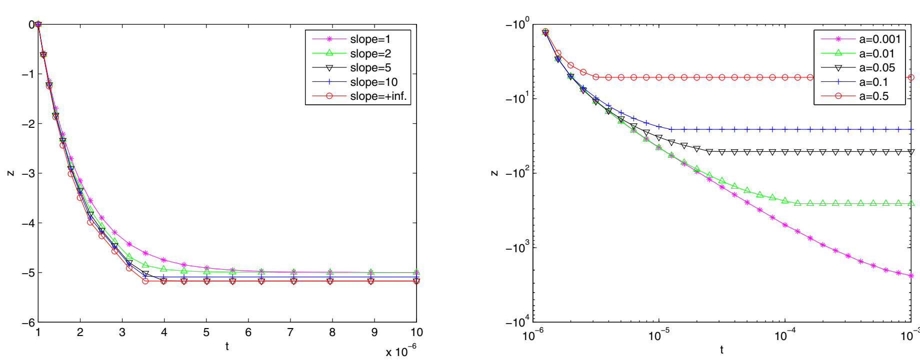

Fig. 2. Stable state variable $z(t)$ with respect to various different values of parameters ξ and a in the continuous-time kWTA network (9)

4 Simulation Results

In this section, the results of extensive simulations will be discussed with respect to various parameters of the kTWA networks (e.g., ξ, a, ϵ, $z(0)$, and n). In all simulations, unless otherwise specified, the inputs are set as $u_i \in \{1, 2, \ldots, n\}$ and the parameters are set as $n = 10, k = 5, a = 0.5, z(0) = 0, \epsilon = 10^{-6}, \xi \to +\infty$.

In Fig. 2, the transient behaviors of the state variable $z(t)$ of the continuous-time kWTA network in (9) are depicted with respect to increasing gain parameter ξ from 1 to $+\infty$ in the left subplot and various values of parameter a ranging from 0.001 to 0.5 in the right subplot where both axes are in a logarithmic scale. The results show that the convergence is faster when the gain parameter ξ is larger or a is smaller. In Fig. 3, the transient behaviors of the state variable $z(t)$ of the continuous-time kWTA network in (9) are depicted with respect to 30

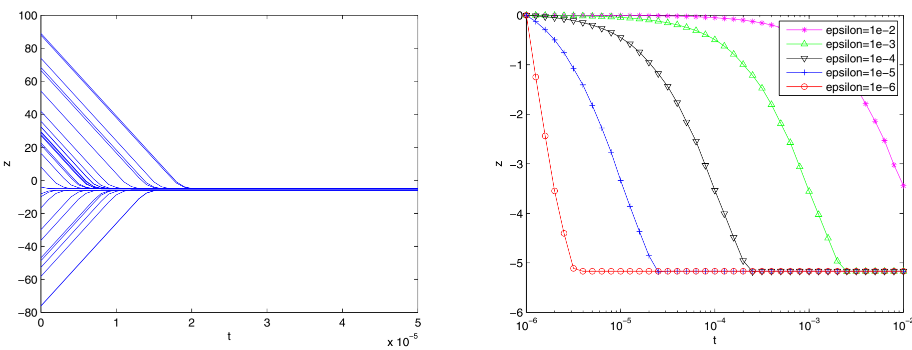

Fig. 3. Stable state variable $z(t)$ with respect to various different values of initial state $z(0)$ and scaling constant ϵ in the continuous-time kWTA network (9)

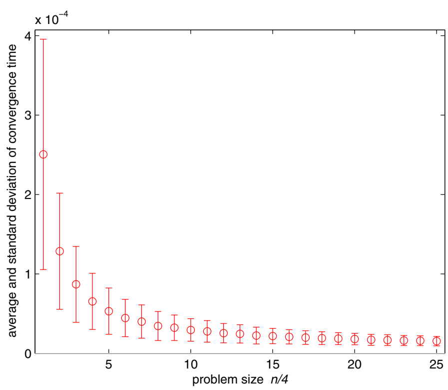

Fig. 4. Average and deviation of convergence time in the continuous-time kWTA network (9) with 1000 random inputs u of increasing size n, where $k = n/2$

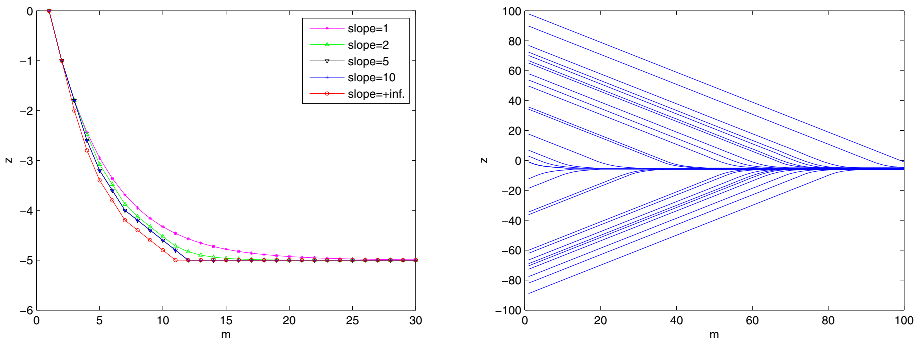

Fig. 5. Stable state variable $z(m)$ with respect to various different values of gain parameter ξ and initial state $z(0)$ in the discrete-time kWTA network (11)

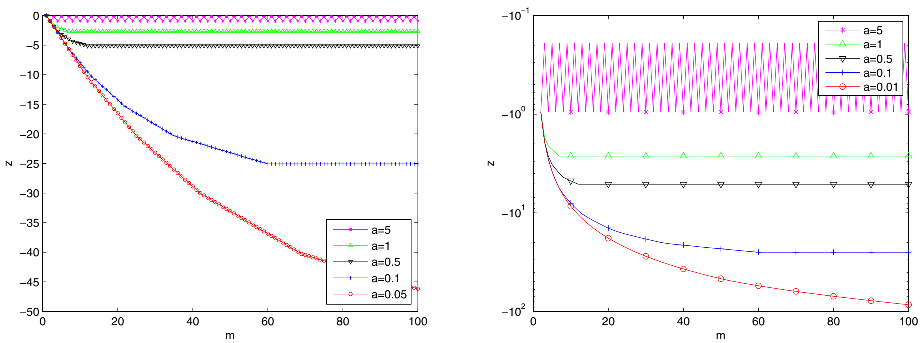

Fig. 6. Stable state variable z with respect to various different values of resolution parameter a in the discrete-time kWTA network (11)

random initial states $z(0)$ drawn from [-100,100] under the uniform distribution in the left subplot and decreasing scaling constant ϵ from 10^{-2} to 10^{-6} in the right subplot where the horizontal axis is in a logarithmic scale. It is obvious that the state variable z is globally stable from any initial state $z(0)$ and its convergence is faster when the scaling constant ϵ is smaller. In Fig. 4, the Monte Carlo simulation results are shown 1000 random inputs $u_i \in \{1, 2, \ldots, n\}$ with increasing sizes (i.e., $n = 4, 8, \ldots, 100$). It is interesting to see that the average and standard deviation of convergence time of the state variable z decrease as the problem size n increases, which reinforces the results in [23]. In particular, the averaged convergence time approaches to a stable value about 15 microseconds for large kWTA problems.

Similar to the results for the continuous-time kWTA network (9) and (10), the simulation results for the discrete-time kWTA network (11) and (12) are shown in Figs. 5-8. Specifically, Fig. 5 depicts the transient behaviors of the state variable $z(m)$ of the discrete-time kWTA network in (11) with respect

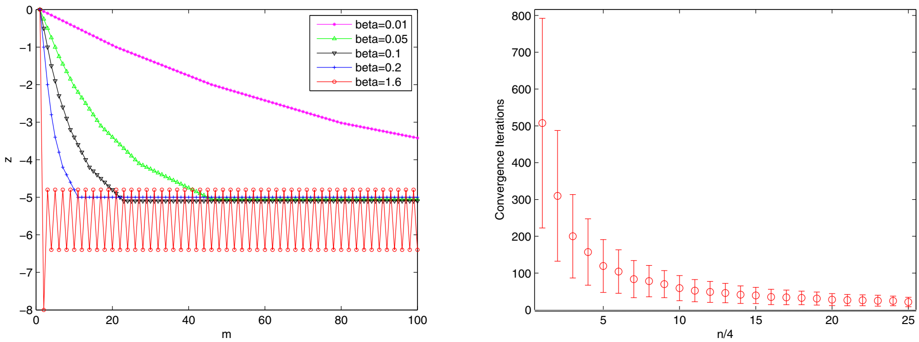

Fig. 7. (i) Stable state variable z with respect to various different values of step size parameter β in the discrete-time kWTA network (11). (ii) Average and deviation of the number of iterations for convergence in the discrete-time kWTA network (11) with 1000 random inputs of increasing size n, where $k = n/2$.

to increasing gain parameter ξ from 1 to $+\infty$ and 30 random initial states $z(0)$ drawn from [-100,100] under the uniform distribution, which show that the convergence is faster when the gain parameter ξ is larger and global stability of the state variable. In Fig. 6, the transient behaviors of state variable $z(m)$ of the discrete-time kWTA network are depicted with respect to various values of parameter a ranging from 0.01 to 5, where the vertical axis for z is in a logarithmic scale in the right subplot. The results show that the convergence of the state variable z is faster when a is larger, which is contrary to that in continuous-time kWTA network shown in Fig. 2 (right subplot). In addition, when a is too big, the state of the discrete-time network becomes unstable. In Fig. 7, the transient behaviors of the state variable $z(m)$ of the discrete-time kWTA network are depicted with respect to various values of its step size parameter β ranging from 0.01 to 1.6. It is not surprising that the convergence is faster when β is bigger and an oscillation occurs when β is too big. In Fig. 8, the Monte Carlo simulation results are shown 1000 random inputs $u_i \in \{1, 2, \ldots, n\}$ with increasing sizes (i.e., $n = 4, 8, \ldots, 100$) for the discrete-time network. It can be observed that, as the problem size n increases, the average number of iterations needed for the of state variable z converges to a stable value 21.50, but its deviation decreases.

5 Concluding Remarks

In this paper, extensive simulation results are reported on the global state stability of the continuous-time and discrete-time k-winners-take-all (kWTA) networks with a single state variable under various parametric configurations. In particular, it is demonstrated that these kWTA networks work faster with high-gain activation functions. In addition, the scalability of the kWTA networks is also shown in terms of input dimension and convergence time. The theoretical justifications of the high-gain activation function in the kWTA networks will be published in a separate paper.

References

1. Majani, E., Erlanson, R., Abu-Mostafa, Y.: On the k-winners-take-all network. In: Touretzky, D.S. (ed.) Advances in Neural Information Processing Systems, vol. 1, pp. 634–642. Morgan-Kaufmann, San Mateo (1989)
2. Maass, W.: On the computational power of winner-take-all. Neural Comput. 12, 2519–2535 (2000)
3. Fish, A., Akselrod, D., Yadid-Pecht, O.: High precision image centroid computation via an adaptive k-winner-take-all circuit in conjunction with a dynamic element matching algorithm for star tracking applications. Analog Integrated Circuits and Signal Processing 39, 251–266 (2004)
4. Marr, D., Poggio, T.: Cooperative computation of stereo disparity. Science 194(4262), 283–287 (1976)
5. Itti, L., Koch, C., Niebur, E.: A model of saliency-based visual attention for rapid scene analysis. IEEE Trans. Pattern Analysis and Machine Intelligence 20, 1254–1259 (1998)
6. Yuille, A., Geiger, D.: Winner-take-all networks. In: The Handbook of Brain Theory and Neural Networks, 2nd edn., pp. 1228–1231. MIT Press, Cambridge (2003)
7. Pouliquen, P.O., Andreou, A.G., Strohbehn, K.: Winner-takes-all associative memory: a hamming distance vector quantizer. Analog Integrated Circuits and Signal Processing 13, 211–222 (1997)
8. DeSouza, G.N., Kak, A.C.: Vision for mobile robot navigation: a survey. IEEE Trans. Pattern Analysis and Machine Intelligence 24, 237–267 (2002)
9. Wolfe, W., Mathis, D., Anderson, C., Rothman, J., Gottler, M., Brady, G., Walker, R., Duane, G., Algghband, G.: K-winner networks. IEEE Trans. Neural Networks 2, 310–315 (1991)
10. Dempsey, G.L., McVey, E.S.: Circuit implementation of a peak detector neural network. IEEE Trans. Circuits Syst. II 40, 585–591 (1993)
11. Wang, J.: Analogue winner-take-all neural networks for determining maximum and minimum signals. Int. J. Electron. 77, 355–367 (1994)
12. Urahama, K., Nagao, T.: K-winners-take-all circuit with o(n) complexity. IEEE Trans. Neural Networks 6(3), 776–778 (1995)
13. Wu, J.L., Tseng, Y.H.: On a constant-time, low-complexity winner-take-all neural network. IEEE Trans. Computers 44, 601–604 (1995)
14. Yen, J., Guo, J., Chen, H.: A new k-winners-take-all neural network and its array architecture. IEEE Trans. Neural Networks 9(5), 901–912 (1998)
15. Sekerkiran, B., Cilingiroglu, U.: A CMOS k-winners-take-all circuit with O(N) complexity. IEEE Trans. Circuits and Systems 46, 1–5 (1999)
16. Sum, J.P.F., Leung, C.S., Tam, P.K.S., Young, G.H., Kan, W.K., Chan, L.W.: Analysis for a class of winner-take-all model. IEEE Trans. Neural Netw. 10, 64–71 (1999)
17. Calvert, B.A., Marinov, C.: Another k-winners-take-all analog neural network. IEEE Trans. Neural Netw. 11, 829–838 (2000)
18. Marinov, C., Calvert, B.: Performance analysis for a k-winners-take-all analog neural network: basic theory. IEEE Trans. Neural Networks 14(4), 766–780 (2003)

19. Marinov, C.A., Hopfield, J.J.: Stable computational dynamics for a class of circuits with $O(N)$ interconnections capable of KWTA and rank extractions. IEEE Trans. Circuits Syst. I 52, 949–959 (2005)
20. Liu, S., Wang, J.: A simplified dual neural network for quadratic programming with its KWTA application. IEEE Trans. Neural Netw. 17(6), 1500–1510 (2006)
21. Hu, X., Wang, J.: Design of general projection neural networks for solving monotone linear variational inequalities and linear and quadratic optimization problems. IEEE Tran. Systems, Man and Cybernetics - Part B 37, 1414–1421 (2007)
22. Liu, Q., Wang, J.: Two k-winners-take-all networks with discontinuous activation functions. Neural Networks 21(2-3), 406–413 (2008)
23. Hu, X., Wang, J.: An improved dual neural network for solving a class of quadratic programming problems and its k-winners-take-all application. IEEE Trans. Neural Networks 19, 2022–2031 (2008)
24. Xia, Y., Sun, C.: A novel neural dynamical approach to convex quadratic program and its efficient applications. Neural Networks 22(10), 1463–1470 (2009)
25. Liu, Q., Cao, J., Liang, J.: A discrete-time recurrent neural network with one neuron for k-winners-take-all operation. In: Yu, W., He, H., Zhang, N. (eds.) ISNN 2009. LNCS, vol. 5551, pp. 272–278. Springer, Heidelberg (2009)
26. Bazaraa, M.S., Jarvis, J.J., Sherali, H.D.: Linear Programming and Network Flows. Wiley, New York (1990)