American Journal of Neural Networks and Applications
2016; 2(1): 1-5
http://www.sciencepublishinggroup.com/j/ajnna
doi: 10.11648/j.ajnna.20160201.11
ISSN: 2469-7400(Print); ISSN: 2469-7419(Online)

Convergence of Online Gradient Method for Pi-sigma Neural Networks with Inner-penalty Terms

Kh. Sh. Mohamed[1, 2], Xiong Yan[3], Y. Sh. Mohammed[4, 5], Abd-Elmoniem A. Elzain[5, 6], Habtamu Z. A.[2], Abdrhaman M. Adam[2]

[1]Mathematical Department, College of Science, Dalanj University, Dalanj, Sudan
[2]School of Mathematical Sciences, Dalian University of Technology, Dalian, China
[3]School of Science, Liaoning University of Science & Technology, Anshan, China
[4]Physics Department, College of Education, Dalanj University, Dalanj, Sudan
[5]Department of Physics, College of Science & Art, Qassim University, Oklat Al- Skoor, Saudi Arabia
[6]Department Department of Physics, University of Kassala, Kassala, Sudan

Email address:

khshm7@yahoo.com (Kh. Sh. Mohamed), xy-zhxw@163.com (Xiong Yan), yshm@yahoo.com (Y. Sh. Mohammed), Abdelmoniem1@yahoo.com (Abd-Elmoniem A. E.), habtamuz@mail.dlut.edu.cn(Habtamu Z. A), abdelrhaman013@yahoo.com(A. M. Adam)

To cite this article:

Kh. Sh. Mohamed, Xiong Yan, Y. Sh. Mohammed, Abd-Elmoniem A. Elzain, Habtamu Z. A., Abdrhaman M. Adam. Convergence of Online Gradient Method for Pi-sigma Neural Networks with Inner-penalty Terms. *American Journal of Neural Networks and Applications*. Vol. 2, No. 1, 2016, pp. 1-5. doi: 10.11648/j.ajnna.20160201.11

Received: March 14, 2016; **Accepted**: March 30, 2016; **Published**: May 10, 2016

Abstract: This paper investigates an online gradient method with inner- penalty for a novel feed forward network it is called pi-sigma network. This network utilizes product cells as the output units to indirectly incorporate the capabilities of higher-order networks while using a fewer number of weights and processing units. Penalty term methods have been widely used to improve the generalization performance of feed forward neural networks and to control the magnitude of the network weights. The monotonicity of the error function and weight boundedness with inner- penalty term and both weak and strong convergence theorems in the training iteration are proved.

Keyword: Convergence, Pi-sigma Network, Online Gradient Method, Inner-penalty, Boundedness

1. Introduction

A novel higher order feedforward polynomial neural network is known to provide inherently more powerful mapping abilities than traditional feed forward neural network called the pi-sigma network (PSN) [2]. This network utilizes product cells as the output units to indirectly incorporate the capabilities of higher-order networks while using a fewer number of weights and processing units. The neural networks consisting of the PSN modules has been used effectively in pattern classification and approximation problems [1, 4, 10, 11].There are two ways of training to updating weight: The first approach, batch (offline) training[18],the weights are modified after each training pattern is presented to the network. Second approach, online training, the weights updating immediately after each training sample is fed see [13]. The penalty term is often introduced into the network training algorithms has been widely used so as to control the magnitude of the weights and to improve the generalization performance of the network [6, 8], here the generalization performance refers to the capacity of a neural network to give correct outputs for untrained data. Specially cause, in the second approach the training weights updating become very large and over-fitting tends to occur, by adding the penalty term in into the cost function, when use second approach has been successfully application see [3, 7, 12, 14], which acts as a brute-force to drive unnecessary weights to zero and to prevent the weights from taking too large in the training process. In the work area of penalty term at the same of the inner-penalty term (IP), which have worked to reduce the magnitude of the network weights with efficiency improve the generalization performance of the network [5, 9, 17]. In this paper, we prove the (strong and weak) convergence of the online gradient with inner penalty and the monotonicity of the error function and the weight sequence are uniformly bounded during the training procedure with inner-penalty.

The rest of this paper is organized as follows. The neural network structure and the online gradient method with inner-penalty are described in Section 2. The preliminary lemmas are disruption in Section 3. The convergence results are presented and the rigorous proofs of the main results are provided in Section 4. Finally, in Section 5 we conclusions this study.

2. PSN-IPAlgorithm

PSN is a higher order feed forward polynomial neural network consisting of a single hidden layer. The hidden layer has summing units where as the output layer has product units. PSN, which has a three-layer network consisting of p input units, N summation units, and 1 product layers. Let $\omega_k = \left(\omega_{k1}, \omega_{k2} \dots, \omega_{kp}\right)^T \in \mathbb{R}^P (1 \le k \le N)$ the weight vectors connecting the input and summing units, and write $\omega = (\omega_1^T, \omega_2^T, \dots, \omega_N^T) \in \mathbb{R}^{Np}$. We have included a special input unit ξ_P, corresponding to the biases ω_{kP}, with fixed value-1. The structure of PSN is shown in Figure 1.

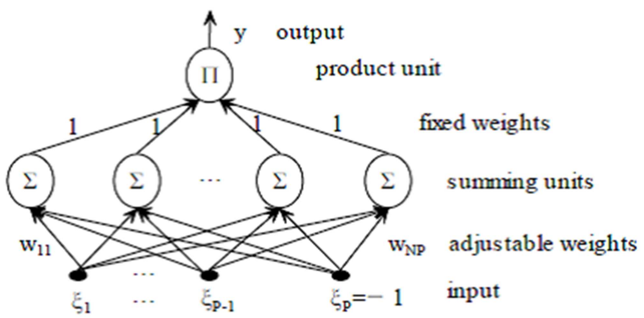

Figure 1. *PSN structure with a single output*

Where $\xi^j = \left(\xi_1^j, \dots, \xi_{P-1}^j, \xi_P^j\right) \in \mathbb{R}^p$ and $\xi_P^j \equiv -1$. Assume $g: \mathbb{R} \to \mathbb{R}$ is a given activation function. For an input $\xi^j \in \mathbb{R}^P$, the output of the network is

$$y = \mathrm{g}(\textstyle\prod_{i=1}^N (\omega_i \cdot \xi^j)) \tag{1}$$

The network supplied with a given set of training samples $\{\xi^j, o^j\}_{j=1}^J \subset \mathbb{R}^P \times \mathbb{R}$. The error function with a inner penalty given by

$$E(\omega) = \frac{1}{2}\sum_{j=1}^{J}(o^j - \mathrm{g}(\prod_{i=1}^{N}(\omega_i \cdot \xi^j)))^2 + \frac{\lambda}{2}\sum_{j=1}^{J}\sum_{k=1}^{N}(\omega_k \cdot \xi^j)^2$$

$$= \textstyle\sum_{j=1}^J \mathrm{g_j}(\prod_{i=1}^N (\omega_i \cdot \xi^j)) + \frac{\lambda}{2}\sum_{j=1}^J \sum_{k=1}^N (\omega_k \cdot \xi^j)^2 \tag{2}$$

Where $\lambda > 0$ is a inner penalty coefficient and $\mathrm{g_j}(t) = \frac{1}{2}(o^j - \mathrm{g}(t))^2$The gradient function is given by

$$E_{\omega_k}(\omega) = \textstyle\sum_{j=1}^J [\mathrm{g_j'}(\prod_{i=1}^N (\omega_i \cdot \xi^j)) \prod_{\substack{i=1 \\ i\ne k}}^N (\omega_i \cdot \xi^j) + \lambda(\omega_k \cdot \xi^j)]\xi^j \tag{3}$$

Given an initial weight ω^0, the online method with inner penalty updates them iteratively by the form

$$\omega_k^{m+1} = \omega_k^m + \Delta\omega_k^m, \quad m = 0,1, \dots. \tag{4}$$

$$\Delta\omega_k^m = [\mathrm{g_j'}(\textstyle\prod_{i=1}^N (\omega_i^m \cdot \xi^j)) \prod_{\substack{i=1 \\ i\ne k}}^N (\omega_i^m \cdot \xi^j) + \lambda(\omega_k^m \cdot \xi^j)]\, \xi^j \tag{5}$$

Where $\eta > 0$ is the learning rate in the mth training cycle.We denote by$\|\cdot\|$ the usual Euclidean norm and the corresponding derived matrix norm and the following Assumption is imposed throughout this paper.

Assumption 1.

$$|\mathrm{g_j}(t)|, |\mathrm{g_j'}(t)|, |\mathrm{g_j''}(t)| \le M_1, \forall t \in \mathbb{R}, 1,1 \le j \le J$$

Assumption 2.

$$\|\xi^j\| \& |\omega_k^m \cdot \xi^j| < M_1, 1 \le k \le N, 1 \le j \le Jm = 0,1, \dots$$

Assumption 3.

The learning rateη and penalty parameter $\lambda$are chosen to satisfy the condition: $0 \le \eta \le 1/(\lambda\widetilde{M} + M)$

Assumption 4.

$\{\omega_k^m\}_{m=0,1,\cdots}$ are contained in bounded closed region $\Theta \subset \mathbb{R}^{NP}$, and there are exist points in set $\Theta_0 = \{\omega \in \Theta | E_\omega(\omega) = 0\}$.

3. Preliminary Lemmas

The next lemma present the montonicity of the sequence $\{E(\omega)\}$. It is essential for the proof of weakly convergence of PSN with penalty, presented in the following Theorems. For sake of description, we denote

$$r_{k,j}^m = \Delta\omega_k^{m+1} - \Delta\omega_k^m \tag{6}$$

$$\psi_j^{m+1} = \textstyle\prod_{i=1}^N (\omega_i^{m+1} \cdot \xi^j) \tag{7}$$

$$\varphi_{k,j}^{m+1} = \textstyle\prod_{\substack{i=1 \\ i\ne k}}^N (\omega_i^{m+1} \cdot \xi^j) \tag{8}$$

$$1 \le j \le J, 1 \le i \le N, m = 0,1, \cdots$$

To begin with, first we present a few lemmas as preparation to prove Theorems

*Lemma 1.*Let Assumption 1~2 are valid, there hold

$$(i)\|\psi_j^{m+1} - \psi_j^m\| \le M_2 \textstyle\sum_{k=1}^N \|\Delta\omega_k^m\| \tag{9}$$

$$(ii)\|\varphi_{k,j}^{m+1} - \varphi_{k,j}^m\| \le \widetilde{M}_2 \textstyle\sum_{k=1}^N \|\Delta\omega_k^m\| \tag{10}$$

Proof .By Assumption 2and Cauchy- Schwartz inequality, we have

$$|\psi_j^{m+1} - \psi_j^m| \le \left|\prod_{i=1}^{N-1}(\omega_i^{m+1} \cdot \xi^j)\right| |(\omega_N^{m+1} - \omega_N^m)\xi^j|$$

$$+\left|\prod_{i=1}^{N-2}(\omega_i^{m+1}\cdot\xi^j)\,(\omega_N^m\cdot\xi^j)\right|(\omega_{N-1}^{m+1}-\omega_{N-1}^m)\xi^j$$

$$+\cdots+\left|\prod_{\mathrm{i}=2}^{N}(\omega_i^m\cdot\xi^j)\right|\left|(\omega_1^{m+1}-\omega_1^m)\xi^j\right|$$

$$\le M_1^{N-1}\|\xi^j\|\sum_{k=1}^{N}\|\Delta\omega_k^m\|\le M_1^N\sum_{k=1}^{N}\|\Delta\omega_k^m\|$$

$$\le M_2\textstyle\sum_{k=1}^{N}\|\Delta\omega_k^m\| \tag{11}$$

where $M_2=M_1^N$, $1\le j\le J$, $1\le k\le N$, $m=0,1,\cdots$. Similarly, we get

$$\|\varphi_{k,j}^{m+1}-\varphi_{k,j}^m\|\le\widetilde{M}_2\textstyle\sum_{k=1}^{N}\|\Delta\omega_k^m\| \tag{12}$$

Here, $\widetilde{M}_2=M_1^{N-1}$. This is completes the proof.

*Lemma 2.*Suppose Assumptions 1~3 are satisfied, and the weight sequence$\{\omega_k^m\}_{m=0,1,\cdots}$is generated by (4) ~ (5), then

$$E(\omega_k^{m+1})-E(\omega_k^m)$$

$$\le-\left(\tfrac{1}{\eta}-\lambda-M_3-M_4-M_5\right)\textstyle\sum_{k=1}^{N}\|\Delta w_k^m\|^2 \tag{13}$$

Proof. Applying Taylor's formula to extend $\mathrm{g_j}(\psi_j^{m+1})$ at (ψ_j^m), we have

$$\mathrm{g_j}(\psi_j^{m+1})-\mathrm{g_j}(\psi_j^m)=\mathrm{g_j'}\left(\psi_j^m\right)\left(\varphi_{k,j}^m\right)\xi^j\sum_{k=1}^{N}(\omega_k^{m+1}-\omega_k^m)$$

$$+\frac{1}{2}\mathrm{g}''(t_1)\left(\psi_j^{m+1}-\psi_j^m\right)^2$$

$$+\tfrac{1}{2}\textstyle\sum_{\substack{k_1,k_2=1\\k_1\neq k_2}}^{N}\left(\prod_{\substack{i=1\\i\neq k_1,k_2}}^{N}t_2\right)(\omega_{k_1}^{m+1}-\omega_{k_1}^m)(\omega_{k_2}^{m+1}-\omega_{k_2}^m)(\xi^j)^2 \tag{14}$$

where $t_1,t_2\in\mathbb{R}$ are on the line segment between ψ_j^{m+1} and ψ_j^m. After dealing with (14) by accumulation $\mathrm{g_j}(\psi_j^{m+1})$ for $1\le j\le J$, we obtain from (2), (4), (5) and Taylor's formula we obtain

$$E(\omega_k^{m+1})-E(\omega_k^m)=\sum_{j=1}^{J}[\mathrm{g_j}(\psi_j^{m+1})-\mathrm{g_j}(\psi_j^m)]$$

$$+\frac{\lambda}{2}\sum_{j=1}^{J}\sum_{k=1}^{N}[(\omega_k^{m+1}\cdot\xi^j)^2-(\omega_k^m\cdot\xi^j)^2]$$

$$=\sum_{j=1}^{J}[\mathrm{g_j'}(\psi_j^m)(\varphi_{k,j}^m)\xi^j+\lambda(\omega_k^m\cdot\xi^j)\xi^j](\Delta\omega_k^m)$$

$$+\lambda\sum_{j=1}^{J}\sum_{k=1}^{N}(\Delta\omega_k^m\cdot\xi^j)^2+\Im_1+\Im_2+\Im_3+\Im_4$$

$$=-\tfrac{1}{\eta}\textstyle\sum_{k=1}^{N}\|\Delta\omega_k^m\|^2+\Im_1+\Im_2+\Im_3+\Im_4 \tag{15}$$

where

$$|\Im_1|\le\tfrac{1}{\eta}\textstyle\sum_{j=1}^{J}(\Delta\omega_k^m)\cdot\sum_{k=1}^{N}\left(r_{k,j}^m\right) \tag{16}$$

$$|\Im_2|\le\tfrac{1}{2}\textstyle\sum_{j=1}^{J}\mathrm{g}''(t_1)\left(\psi_j^{m+1}-\psi_j^m\right)^2 \tag{17}$$

$$|\Im_3|\le\tfrac{1}{2}\textstyle\sum_{\mathrm{j}=1}^{\mathrm{J}}\mathrm{g}'(\psi_j^m)\sum_{\substack{k_1,k_2=1\\k_1\neq k_2}}^{N}\left(\prod_{\substack{i=1\\i\neq k_1,k_2}}^{N}t_2\right)(\Delta\omega_{k_1}^m\cdot\Delta\omega_{k_2}^m)\left(\xi^j\right)^2 \tag{18}$$

$$|\Im_4|\le\tfrac{\lambda}{2}\textstyle\sum_{j=1}^{J}\sum_{k=1}^{N}(\Delta\omega_k^m\cdot\xi^j)^2 \tag{19}$$

By Assumption 1, (4) ~ (5), Lemma 1and the mean value theorem gives

$$\|r_{k,j}^m\|=\|\mathrm{g_j'}(\psi_j^{m+1})\varphi_{k,j}^{m+1}-\mathrm{g_j'}(\psi_j^m)\varphi_{k,j}^m\|\|\xi^j\|$$

$$+\lambda\|\omega_k^{m+1}-\omega_k^m\|\|\xi^j\|^2$$

$$\le\|\mathrm{g_j''}(t_1)(\psi_j^{m+1}-\psi_j^m)(\varphi_{k,j}^{m+1})\xi^j\|$$

$$+\|\mathrm{g_j'}(\psi_j^m)(\varphi_{k,j}^{m+1}-\varphi_{k,j}^m)\|+\lambda\|\omega_k^{m+1}-\omega_k^m\|\|\xi^j\|^2$$

$$\le(M_2M_1^{N+1}+\widetilde{M}_2M_1^2+\lambda M_1^2)\textstyle\sum_{k=1}^{N}\|\Delta\omega_k^m\|^2 \tag{20}$$

Thus with (16) gives

$$|\Im_1|\le\tfrac{1}{\eta}\textstyle\sum_{k=1}^{N}\|\Delta\omega_k^m\cdot r_{k,j}^m\|$$

$$\le M_3\textstyle\sum_{k=1}^{N}\|\Delta\omega_k^m\|^2 \tag{21}$$

Here, $M_3=M_2M_1^{N+1}+\widetilde{M}_2M_1^2+\lambda M_1^2+1$. By (4), (9) in Lemma 1 and Cauchy- Schwartz inequality, we have

$$|\Im_2|\le\frac{1}{2}M_1\left|\psi_j^{m+1}-\psi_j^m\right|^2$$

$$\le(\tfrac{1}{2}M_2^2+\lambda M_3)M_1\textstyle\sum_{k=1}^{N}\|\Delta\omega_k^m\|^2$$

$$\le M_4\textstyle\sum_{k=1}^{N}\|\Delta\omega_k^m\|^2 \tag{22}$$

where $M_4=\frac{1}{2}M_1M_2^2+\lambda M_1^2$. It follows from Assumption 1~2, (2) and Taylor's formula, we obtain

$$|\Im_3|\le\frac{1}{2}M_1^{N+1}\sum_{j=1}^{J}\sum_{\substack{k_1,k_2=1\\k_1\neq k_2}}^{N}\|\Delta\omega_{k_1}^m\cdot\Delta\omega_{k_2}^m\|$$

$$\le\frac{1}{2}M_1^{N+1}J(N-1)\sum_{j=1}^{J}\sum_{k=1}^{N}\|\Delta\omega_{k_1}^m\cdot\Delta\omega_{k_2}^m\|$$

$$\le M_5\textstyle\sum_{k=1}^{N}\|\Delta\omega_k^m\|^2 \tag{23}$$

where $M_5=\frac{1}{2}M_1^{N+1}J(N-1)$. By Assumption (2) and (4) leads

$$|\Im_4|\le\lambda\|\xi^j\|^2\sum_{k=1}^{N}\|\Delta\omega_k^m\|^2\le\frac{1}{2}\lambda M_1^2\sum_{k=1}^{N}\|\Delta\omega_k^m\|^2$$

$$\le\lambda\widetilde{M}\textstyle\sum_{k=1}^{N}\|\Delta\omega_k^m\|^2 \tag{24}$$

where$\widetilde{M} = \frac{1}{2}M_1^2$. Collate (20) ~(24) into (15) gives

$$E(\omega_k^{m+1}) - E(\omega_k^m)$$
$$\leq -(\frac{1}{\eta} - \lambda - M_3 - M_4 - M_5)\sum_{k=1}^N \|\Delta w_k^m\|^2 \quad (25)$$

This is completes the proof.

*Lemma 3.*Suppose that $F: \mathbb{R}^K \to \mathbb{R}$ is continuous and differentiable on a compact set $\breve{D} \subset \mathbb{R}^K$ and that $\Theta = \{\mathcal{Z} \in \breve{D} | \nabla h(\mathcal{Z}) = 0\}$ has only finite number of point. If a sequence $\{\mathcal{Z}^m\}_{m=1}^\infty \in \breve{D}$ satisfies then $\lim_{m\to\infty}\|\mathcal{Z}^{m+1} - \mathcal{Z}^m\| = 0$, $\lim_{m\to\infty}\|\nabla h(\mathcal{Z}^m)\| = 0$. Then there exists a point $\mathcal{Z}^* \in \Theta$ such that $\lim_{m\to\infty} \mathcal{Z}^m = \mathcal{Z}^*$.

Proof. This result is basically the same as Theorem 14.1.5 in [16], and the detailed proof is thus omitted.

4. Convergence Theorems

Now, we can elucidate and proofs the convergence theorems, which we needed

Theorem 1.(Monotonicity): Let Assumption 1~3 are valid and the weight sequence$\{\omega_k^m\}_{m=0,1,\cdots}$be generated by (4) ~ (5),then

$$E(\omega_k^{m+1}) - E(\omega_k^m), m = 0,1,\cdots \quad (26)$$

Proof. Let

$$M = M_3 + M_4 + M_5 \quad (27)$$

By Assumption 3,which satisfies

$$0 < \eta < \frac{1}{\lambda\widetilde{M} + M} \quad (28)$$

Thus with (28) and Lemma 2, we have

$$E(\omega_k^{m+1}) - E(\omega_k^m) \leq -(\frac{1}{\eta} - \lambda\widetilde{M} - M)\sum_{k=1}^N \|\Delta\omega_k^m\|^2$$
$$\leq 0 \quad (29)$$

This completes the proof of the Theorem 1.

Theorem 2. (Boundednss): Suppose that Assumption of Theorem 1 are valid, the weight sequence$\{\omega_k^m\}_{m=0,1,\cdots}$be generated by (4) ~ (5) are uniformly bounded.

Proof. By Assumption 1, and Theorem 1, we have

$$E(\omega_k^m) - E(\omega_k^{m-1}) \leq \cdots \leq E(\omega_k^0)$$
$$= \sum_{j=1}^J g_j(\psi_j^0) + \frac{\lambda}{2}\sum_{j=1}^J\sum_{k=1}^N(\omega_k^0 \cdot \xi^j)^2 \leq C \quad (30)$$

and

$$C = JM_1(1 + \frac{\lambda}{2}\sum_{k=1}^N\|\omega_k^0\|^2) \quad (31)$$

From (2), (30) gives

$$\lambda(\omega_k^m \cdot \xi^j)^2 \leq 2E(\omega_k^m) \leq 2C \quad, j = 1,2,\dots,J \quad (32)$$

By (4) ~ (5), we have

$$\omega_k^m = \omega_k^0 - \eta\sum_{t=1}^{m-1}\sum_{k=1}^N[g_j'(\psi_j^t)\varphi_{k,j}^t + \lambda(\omega_k^t \cdot \xi^j)]\xi^j \quad (33)$$

Let the second part of above equation be ω_{k1}^m, Denote $\mathbb{R}_1 = span\{\xi^1, \xi^2, .., \xi^J\} \subset \mathbb{R}^n$ and $\mathbb{R}_2 = \mathbb{R}_1^\perp$ be the orthogonal complement space of $\mathbb{R}_1$. Denote the second part of (33) by ω_{i1}^m, obviously $\omega_{k1}^m \in \mathbb{R}_1$. we divide ω_k^0 into $\omega_k^0 = \omega_{k1}^0 + \omega_{k2}^0$, where $\omega_{k1}^0 \in \mathbb{R}_1$ and $\omega_{k2}^0 \in \mathbb{R}_2$. Then $\omega_k^m = (\omega_{k1}^0 + \omega_{k1}^m) \oplus \omega_{k2}^0 = \widetilde{\omega}_{k1}^m \oplus \omega_{k2}^0$. Applying this to (33) we have

$$|d_t| \coloneqq |\widetilde{\omega}_{k1}^m \cdot \xi^t| = |\omega_k^m \cdot \xi^j| \leq \sqrt{\frac{2C}{\lambda}}, \quad t = 1,2,..,T \quad (34)$$

Suppose $\{\xi^{i_1}, \xi^{i_2}, .., \xi^{i_T}\}(i_t \in \{1,\dots,J\}, t = 1,2,\dots,T)$ is a base of the space $\mathbb{R}_1$. There are $a_t \in \mathbb{R}$ $(t = 1,2,\dots,T)$ such that $\widetilde{\omega}_{k1}^m = a_1\xi^{i_1} + \cdots + a_T\xi^{i_T}$. Then $(a_1\xi^{i_1} + \cdots + a_T\xi^{i_T}) \cdot \xi^{i_t} = d_t, t = 1, \dots, T$. we get

$$\begin{pmatrix} \xi^{j_1} .\xi^{j_1} \dots & \xi^{j_t} .\xi^{j_1} \\ \vdots \ \vdots \ \vdots & \vdots \ \vdots \\ \xi^{j_1}\ \xi^{j_t} \dots & \xi^{j_t} .\xi^{j_t} \end{pmatrix}\begin{pmatrix} a_1 \\ \vdots \\ a_T \end{pmatrix} = \begin{pmatrix} d_1 \\ \vdots \\ d_T \end{pmatrix} \quad (35)$$

Is a base, the coefficient determinant equal to zero, and the system of the linear equations has a unique solution. Assume that the coefficient determinant equals to. Then the solution is as follows

$$L = \begin{vmatrix} \xi^{j_1}.\ \xi^{j_1}\ \dots\xi^{j_{t-1}}.\xi^{j_1}\boldsymbol{d_1}\xi^{j_{t-1}}.\xi^{j_1}\dots\xi^{j_1}.\xi^{j_1} \\ \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \\ \xi^{j_1}\ \xi^{j_t}\ \dots\xi^{j_{t-1}}.\xi^{j_t}\boldsymbol{d_t}\xi^{j_{t-1}}.\xi^{j_t}\dots\xi^{j_1}\ \xi^{j_T} \end{vmatrix}$$

Then the solution is as follows

$$a_t = L \cdot D^{-1} \quad (36)$$

Let the maximum absolute value of all the sub-determinant with rank $(T-1)$ of the coefficient determinant is D', then $|a_t| \leq |D'| \cdot |D^{-1}| \cdot \sum_{t=0}^T|d_t|$. By (34) we have $|a_t| \leq |D'| \cdot |D^{-1}| \cdot T \cdot \sqrt{\frac{2C}{\lambda}}$. $t = 1,2,\dots,T$. Denote $\tilde{C} = \max_{1\leq t\leq T}\|\xi^{j_1}\|$, then

$$\|\widetilde{\omega}_{k1}^m\| = \|a_1\xi^{j_1} + \cdots + a_T\xi^{j_T}\|$$
$$\leq |D' \cdot D^{-1}| \cdot \tilde{C} \cdot T^2 \cdot \sqrt{\frac{2C}{\lambda}} \quad (37)$$

That is $\widetilde{\omega}_{k1}^m$ are bounded uniformly bounded. So from (29), we know ω_k^m are uniformly bounded. In all, we get$\{\omega_k^m\}_{m=0,1,\cdots}$are uniformly bounded, i.e., there exist a bounded closed region $D \subset \mathbb{R}^k$ such that $\{\omega_k^m\} \subset D$.

Theorem 3. (Weak convergence): Suppose that Assumption 1~3 are valid and the weight sequence $\{\omega_k^m\}_{m=0,1,\cdots}$ be generated by (4) ~ (5), then

$$\lim_{m\to\infty}\|E_{\omega_k}(\omega_k^m)\| = 0, \quad (38)$$

Furthermore, if Assumption 4 is also valid, we have the strong convergence: There exists $\omega^* \in \Theta_0$ such that

$$\lim_{m\to\infty}\omega_k^m = \omega_k^* \quad (39)$$

Proof .By (28) and setting $\beta > 0$ such that $\beta = \frac{1}{\eta} - \lambda\widetilde{M} - M$, we have

$$E(\omega_k^{m+1}) \leq E(\omega_k^m) - \beta \sum_{k=1}^{N} \|\Delta\omega_k^m\|^2$$

$$\leq \cdots \leq E(\omega_k^0) - \beta \sum_{i=0}^{m} \sum_{k=1}^{N} \|\Delta\omega_k^m\|^2 \quad (40)$$

Since $E(\omega_k^{m+1}) > 0$ for any $m = 0,1,\cdots$. We set $m \to \infty$

$$\sum_{i=0}^{m} \sum_{k=1}^{N} \|\Delta\omega_k^m\|^2 \leq \frac{E(\omega_k^0)}{\beta} < \infty \quad (41)$$

Combining (3) ~ (5), immediately gives the weak convergence result:

$$\lim_{m\to\infty} \left\|E_{\omega_k}(\omega_k^m)\right\| = 0, m = 0,1,\cdots \quad (42)$$

Next we prove the strong convergence it follows from (4)~(5) and (42) that leads

$$\lim_{m\to\infty} \|\Delta\omega_k^m\| = 0, \quad 0 \leq k \leq N \quad (43)$$

Note that the error function $E(\omega^m)$ defined in (2) is continuously differentiable. By (43), Assumptions 4 and Lemma 3, immediately get the desired result. This completes the proof.

5. Conclusion

Through our study of this paper, the monotoncity of the error function $E(\omega^m)$in formula (2) and the weight sequence boundedness $\{\omega_k^m\}_{m=0,1,\cdots}$ via formula (4) ~ (5) for the online gradient method with inner-penalty are presented, under those condition both weakly and strongly convergence theorems are proved.

Acknowledgment

We gratefully acknowledge Dalanj University for supporting this research. And our thanks to the anonymous reviewers and the editors for their previous helpful comments and valuable suggestion for their kind helps during the period of the research.

References

[1] A J Hussaina and P Liatsisb, Recurrent pi-sigma networks for DPCM image coding. Neurocomputing, 55(2002) 363-382.

[2] Y Shin and J Ghosh, The pi-sigma network: An efficient higher-order neural network for pattern classification and function approximation. International Joint Conference on Neural Networks, 1(1991) 13–18.

[3] P L Bartlett, For valid generalization, the size of the weights is more important than the size of the network, Advances in Neural Information Processing Systems 9 (1997) 134–140.

[4] L J Jiang, F Xu and S R Piao, Application of pi-sigma neural network to real-time classification of seafloor sediments. Applied Acoustics, 24(2005) 346–350.

[5] R Reed, Pruning algorithms-a survey. IEEE Transactions on Neural Networks 8 (1997) 185–204.

[6] G Hinton, Connectionist learning procedures, Artificial Intelligence 40(1989)185-243.

[7] S Geman, E Bienenstock, R Doursat, Neural networks and the bias/variance dilemma, Neural Computation 4 (1992) 1–58.

[8] S Loone and G Irwin, Improving neural network training solutions using regularisation,Neurocomputing37(2001)71-90.

[9] A S Weigend, D E Rumelhart and B A Huberman, Generalization by weight-elimination applied to currency exchange rate prediction. Proc. Intl Joint Conf. on Neural Networks 1(Seatle, 19916) 837-841.

[10] Y Shin and J Ghosh, Approximation of multivariate functions using ridge polynomial networks, International Joint Conference on Neural Networks 2 (1992) 380-385.

[11] M Sinha, K Kumar and P K Kalra, Some new neural network architectures with improved learning schemes. Soft Computing, 4 (2000) 214-223.

[12] R Setiono, A penalty-function approach for pruning feed forward neural networks, Neural Networks 9 (1997) 185–204.

[13] W Wu and Y S Xu, Deterministic convergence of an online gradient method for neural networks, Journal of Computational and Applied Mathematics 144 (1-2) (2002) 335-347.

[14] H S Zhang and W Wu, Boundedness and convergence of online gradient method with penalty for linear output feed forward neural networks, Neural Process Letters 29 (2009) 205–212.

[15] H F Lu, W Wu, C Zhang and X Yan, Convergence of Gradient Descent Algorithm for Pi- Sigma Neural Networks, Journal of Information and Computational Science 3: 3 (2006) 503-509.

[16] YX Yuan and WY Sun, Optimization Theory and Methods, Science Press, Beijing, 2001.

[17] J Kong and W Wu, Online gradient methods with a punishing term for neural networks. Northeast Math. J 1736(2001) 371-378.

[18] W Wu, G R Feng and X Z Li, Training multiple perceptrons via minimization of sum of ridge functions, Advances in Computational Mathematics 17 (2002) 331-347.