$$\left\|\sum_{j=1}^{J} p^{m,j}\right\|^2 \geq \gamma\eta_m \sum_{j=1}^{J}\left\|p^{m,j}\right\|^2 \quad (74)$$

Then theorem 1 is proved.

For $m=0$, if the left side of equ. (74) is zero, then equ. (11), $E_{\omega_{ik}}(\omega^0) = \sum_{j=1}^{J} p^{0,j} = 0$. Hence, we have already reached a local minimum of the error function, and the iteration can be terminated. Otherwise, if

$E_{\omega_{ik}}(\omega^0) = \sum_{j=1}^{J} p^{0,j} \neq 0$, then we choose $\eta_0 > 0$ such that

$$\left\|\sum_{j=1}^{J} p^{0,j}\right\|^2 \geq \gamma\eta_0 \sum_{j=1}^{J}\left\|p^{0,j}\right\|^2 \quad (75)$$

Recalling *Lemma 9*, we know that inequality equ. (69) holds for all nonnegative. Hence, the monotonicity of the error sequence $\{E(\omega^{mJ})\}$ is proved.

Proof of theorem 2.

By using equ. (8) and theorem 1, we have

$$\lambda\left|\omega_k^{mJ}\right|^{\frac{1}{2}} \leq \mathrm{E}(\omega^{\mathrm{mJ}}) \leq \cdots \leq \mathrm{E}(\omega^{\mathrm{mJ}}), \mathrm{m} = 0,1,2 \ldots\ldots\ldots \quad (76)$$

Thus

$$\left\|\omega_k^{mJ}\right\| \leq \frac{1}{\lambda^2}\left(E(\omega^0)\right)^2 \equiv M_0, m = 0,1,2,\ldots\ldots\ldots \quad (77)$$

This together with the definition of C_7 in equ. (36) indicates

$$\left\|p^{m,j}\right\| \leq \mathrm{C}_7 + \frac{\lambda}{2J} M_0, \mathrm{j} = 1,2,\ldots..\mathrm{J} \quad (78)$$

A combination of equ. (16), equ. (18), equ. (22), $0 \leq \eta_m \leq 1$ and equ. (78) yields

$$\begin{aligned}\left\|\omega_k^{mJ+1}\right\| &= \left\|\omega_k^{mJ} - \eta_m p^{m,j} + \alpha_{m,1}\,\Delta_j^m \omega_k^{mJ}\right\| \\ &\leq \left\|\omega_k^{mJ}\right\| + \left\|\eta_m p^{m,j}\right\| + \left\|+\eta_m^2 p^{m,j}\right\| \\ &\leq M_0 + 2\left(\mathrm{C}_7 + \frac{\lambda}{2\mathrm{J}} M_0\right) \equiv \mathrm{M}_1, \mathrm{m} \\ &= 0,1,2,\ldots \quad (79)\end{aligned}$$

Similarly, we obtain

$$\begin{aligned}\left\|\omega_k^{mJ+2}\right\| &\leq \left\|\omega_k^{mJ+1}\right\| + \left\|\eta_m p^{m,j+1}\right\| + \left\|+\eta_m^2 p^{m,j+1}\right\| \\ &\leq M_1 + 2\left(\mathrm{C}_7 + \frac{\lambda}{2\mathrm{J}} M_1\right) \equiv \mathrm{M}_2\end{aligned}$$

and there are integers M_j $(3 \leq j \leq J)$ such that

$$\left\|\omega_k^{mJ+1}\right\| \leq M_j, m = 0,1,2,\ldots..; j = 3,4,\ldots\ldots.J \quad (81)$$

Setting $M = \max\{M_0, M_1, \ldots\ldots., M_J\}$, Equ. equ. (75) and equ. (76) lead to

$$\left\|\omega_k^{mJ+1}\right\| = M, m = 0,1,2,\ldots\ldots; j = 1,2,\ldots..J \quad (82)$$

Note that the constant M is independent of m and j. the boundedness of the weights $\{\omega^i\}$ is thus proved.

Proof of theorem 3.

Denote

$$\sigma^m \leq -\eta_m \left\|\sum_{j=1}^{J} p^{m,j}\right\|^2 - \gamma\eta_m^2 \sum_{j=1}^{J}\left\|p^{m,j}\right\|^2 \quad (83)$$

By the proof of the theorem 1 and $\eta_m > 0$, it holds $\sigma^m \geq 0$ for all $m = 0,1,\ldots.$

In view of *Lemma 8* and theorem 1, we write

$$\begin{aligned}E\left(\omega^{(m+1)J}\right) &\leq E(\omega^{mJ}) - \sigma^m \leq \cdots. \\ &\leq E(\omega^{mJ}) - \sum_{i=0}^{m}\sigma^i \quad (84)\end{aligned}$$

Note that $\left(\omega^{(m+1)J}\right) \geq 0$ for any $m \geq 0$.

Setting $m \to \infty$, we get

$$\sum_{m=0}^{\infty}\sigma^m \leq E(\omega^0) \leq \infty. \quad (85)$$

A combination of equ. (78) and equ. (29) gives

$$\begin{aligned}\sum_{m=0}^{\infty}\left(\gamma\eta_m^2 \sum_{j=1}^{J}\left\|p^{m,j}\right\|^2\right) &\leq C_{16}\sum_{m=0}^{\infty}\eta_m^2 \\ &< \rho^2 C_{16}\sum_{m=0}^{\infty}\frac{1}{m^2} < \infty \quad (86)\end{aligned}$$

where $C_{16} = \gamma J\left(\mathrm{C}_7 + \frac{\lambda}{2\mathrm{J}} M_0\right)^2$. A combination of equ. (84) and equ. (85) yields

$$\sum_{m=0}^{\infty}\eta_m \left\|\sum_{j=1}^{J} p^{m,j}\right\|^2 < \infty \quad (87)$$

Thus, for any unit vector $e \in \mathbb{R}^n$ with $\|e\| = 1$, there holds

$$\begin{aligned}\sum_{m=1}^{\infty}\frac{1}{m}\|E(\omega^{mJ}).e\,\|^2 &\leq \sum_{m=1}^{\infty}\frac{1}{m}\|E(\omega^{mJ}).e\,\|^2 \\ &< \frac{1}{\tau}\sum_{m=0}^{\infty}\eta_m \left\|\sum_{j=1}^{J} p^{m,j}\right\|^2 < \infty \quad (\mathbf{88})\end{aligned}$$

In addition, by using equ. (44) and equ. (78) such that

$$\begin{aligned}\left\|v_i^{m,j}\right\| &\leq (1 + C_2)\,\eta_m \sum_{j=1}^{J}\left\|P^{m\,j}\right\| < \frac{C_{18}}{m}, m \\ &= 0,1,2\ldots..(89)\end{aligned}$$

This together with equ. (19) and equ. (61) gives

$$\begin{aligned}\left|E_\omega\left(\omega^{(m+1)J}\right).e - E_\omega(\omega^{mJ}).e\right| &\leq C_7\|e\|\left\|v_i^{m,j}\right\| \\ &< \frac{C_{19}}{m}, (90)\end{aligned}$$

where $C_{19} = C_7 C_{17}$. The combination of equs. (88) - (90) and Lemma 10 gives

$$\lim_{m\to\infty}|E_\omega(\omega^{mJ}).e\,| = 0. \quad (91)$$

Since $\boldsymbol{e}$ is arbitrary in $\mathbb{R}^{\boldsymbol{n}}$, we have

$$\lim_{m\to\infty}|E_\omega(\omega^{mJ})\,| = 0. \quad (92)$$

Similarly as equ. (90), there is $C_{20} > 0$ such that for all $j = 1,2,\ldots,J-1$

$$\left|E_\omega\left(\omega^{(m+1)J}\right).e - E_\omega(\omega^{mJ}).e\right| < \frac{C_{20}}{m} \quad (93)$$

Thus

$$\begin{aligned}\left|E_\omega(\omega^{mJ+j}).e\right| &\leq \left|E_\omega(\omega^{mJ}).e\right| \\ &\leq \left|E_\omega(\omega^{mJ}).e\right| + \left|E_\omega(\omega^{mJ+j}).e\right| \\ &\quad - \left|E_\omega(\omega^{mJ}).e\right| \\ &< \left|E_\omega(\omega^{mJ}).e\right| + \frac{C_{21}}{m} \to 0\ (n \to \infty) \quad (94)\end{aligned}$$

Again by the arbitrariness of equ. (92), we have

$$\lim_{m\to\infty}\left|E_\omega(\omega^{mJ+j})\,\right| = 0.\ j = 1,2,\ldots\ldots,J-1 \quad (95)$$

Noticing the non-negativeness of the sequences $\{E_\omega(\omega^{mJ+j})\}$ for $j = 1,2,\ldots\ldots,J$, we concluded by using equ. (92) and equ. (95) that

$$\lim_{m\to\infty}\|E_\omega(\omega^m)\| = 0, \quad (96)$$

Next, we prove the strong convergence. By using equ. (89), we have

$$\lim_{m\to\infty}\left\|\omega_k^{(m+1)J} - \omega_k^{mJ}\right\| = \lim_{m\to\infty}\left\|v_k^{m,j}\right\| = 0 \quad (97)$$

Recalling *Lemma 11* and noting equ. (92), equ. (97) and assumption (A4) there exists $\omega^* \in \Omega_0$ such that

$$\lim_{m\to\infty} \omega^{mJ} = \omega^*, \|E_\omega(\omega^*)\| = 0 \quad (98)$$

Note that for $j = 1, 2, \ldots\ldots, J$, there is $C_{22} > 0$ such that

$$\|\omega_k^{mJ+j} - \omega_k^{mJ}\| = \sum_{i=1}^{J} \|\Delta_j^m \omega_k^{mJ+j}\| = \sum_{i=1}^{J} \|\alpha_{m,i}\, \Delta_j^m \omega_k^{mJ+i-1} - \eta_m p^{m,i,i}\| \le C_{23}\eta_m \to 0 \quad (99)$$

Combining this with equ. (98) yields

$$\lim_{m\to\infty} \|\omega_k^{mJ+j} - \omega_k^*\| = 0, j = 1, 2, \ldots\ldots J \quad (100)$$

Hence

$$\lim_{m\to\infty} \omega^m = \omega^*, \|E_\omega(\omega^*)\| = 0 \quad (101)$$

which completes the proof.

5. Conclusions

In this paper, the propose of $L_{1\backslash 2}$ regularization penalty with smoothing term and momentum introduced into the batch gradient learning algorithm is calculated and a convergence of weak and strong theorem and boundedness are provide when it is used for PSNN. As shown using the same usefully lemmas we prove the monotonicity and then the boundedness of the synaptic weights and the gradient of error sequence convergence to zero as training iteration successfully.

Acknowledgment

We gratefully acknowledge Dalanj University and Dalian University of Technology for supporting this research. Special thanks to Prof. Dr. Wei Wu and Dr. Yan Liu for their kind helps during the period of the research.

References

[1] 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, pp. 13- 18, 1991.
[2] A.J. Hussaina, and P. Liatsib, "Recurrent pi –sigma networks for DPCM image coding", Neurocomputing, 55, pp. 363 – 382, 2002.
[3] Y. Shin, and J. Ghosh, "Approximation of multivariate functions using ridge polynomial networks". International Joint Conference on Neural Networks, 2, pp. 380 – 385, 1992.
[4] M. Shinha, K. Kumar, and P.K. Kalra, "Some new neural network architectures with improved learning schemes", Soft Computing, 4, pp. 214 – 223, 2004.
[5] 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, pp. 346 – 350, 2005..
[6] D.E. Rumelhart, and McClelland, "PDP Research Group, Parallel Distributed Processing – Explorations in the Microstructure of Cognition", MIT Press. MA, 1986.
[7] D.E. Rumelhart, G.E. Hinton, and R.J. Williams, "Learning representations by back- propagating errors", Nature, 323, pp. 533 – 536, 1986.
[8] E. Istook, and T. Martinez, "Improved back- propagating learning in neural networks with windowed momentum", Int. J. Neural Syst., 12 (3 – 4), pp. 303 – 318, 2002.
[9] G. Qin, M. R. Varley, and T. J. Terrell, "Accelerated training of back- propagation networks by using adaptive momentum step", IEE Electron. Lett., 28 (4), pp. 377 – 379, 1992.
[10] L.W. Chan, and F. Fallside, "An adaptive training algorithm for back- propagation networks", Comput. Speech Lang., 2, pp. 205 – 218, 1987.
[11] N.M. Zhang, W. Wu, and G.F. Zheng, "Convergence of gradient method with momentum for two-layer feedforward neural networks", IEEE Trans Neural Netw, 17(2), pp. 522–525, 2006
[12] A. Bhaya, and E. Kaszkurewicz, "Steepest descent with momentum for quadratic functions is a version of the conjugate gradient method", Neural Networks, 17, pp. 65 – 71, 2004.
[13] M. Torii, and M. T. Hagan, "Stability of steepest descent with momentum for quadratic functions", IEEE Trans Neural Netw, 13(3), pp. 752–756, 2002.
[14] N. M. Zhang, W. Wu, and G. F. Zheng, "Convergence of gradient method with momentum for two-layer feed forward neural networks", IEEE Trans Neural Netw, 17(2), pp. 522–525, 2006.
[15] H. M. Shao, and G. F. Zheng, "A new BP algorithm with adaptive momentum for FNNs training", WRI Glob Congr Intell Syst., 4, pp. 16 – 20, 2009.
[16] Shao, Hongmei. and Zheng, Gaofeng, "Convergence analysis of a back-propagation algorithm with adaptive momentum", Neurocomputing, 74, pp. 749 – 752, 2011.
[17] X. Yu, N. K. Loh, and W. C. Miller, "A new acceleration technique for the back- propagation algorithm", in: IEEE International Conference on Neural Networks, 3, pp. 1157–1161, 1993
[18] G. E. Hinton, "Connectionist learning procedures", Artif. Intell., 40, pp.185–234, 1989.
[19] R. Setiono, "A penalty-function approach for pruning feedforward neural networks", Neural Networks, 9, pp. 185–204, 1997.
[20] P. L. Bartlett, "For valid generalization, the size of the weights is more important than the size of the network", in: Advancesin Neural Information Processing Systems, 9, pp. 134–140, 1997.
[21] R. Reed, "Pruning algorithms- a survey", IEEE Trans. Neural Networks, 8, pp. 185–204, 1997.
[22] S. Geman, E. Bienenstock, and R. Doursat, "Neural networks and the bias/variance dilemma", Neural Comput., 4, pp. 1–58, 1992. R. Tibshirani, "Regression shrinkage and selection via the lasso", *J. Royal Stat. Soc. Ser. B*, 58(1), pp. 267–288, 1996.
[23] D. L. Donoho, "Neighborly polytopes and the sparse solution of underdetermined systems of linear equations", Stat. Dept. Stanford Univ. Stanford, CA, Tech. Rep. 4, 2005.
[24] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit", *SIAM J. Sci. Comput.*, 20 (1), pp. 33–61, 1998.
[25] X. Y. Chang, Z. B. Xu, H. Zhang, J. J. Wang, and Y. Liang, "Robust regularization theory based on $L_p(0 <$

$p < 1$) regularization: the asymptotic distribution and variable selection consistence of solutions", Sci. China, 40, pp.985–998, 2010.
[26] H. Akaike, "Information theory and an extension of the maximum likelihood principle", In: B.N.PETROV and F. CSAKI, eds. Second International Symposium on Information Theory. Budapest: Akademiai Kiado, 267-281, 1973.
[27] G. Schwarz, "Estimating the dimension of a model. The Annals of Statistics", 6, pp. 461 – 464, 1978.
[28] B. K. Natraajan, "Sparse approximation to linear systems", *SIAM J. Comput.*, 24(2), pp. 227–234, 1995.
[29] M. Ishikawa, "Structural learning with for getting", Neural Networks, 9(3), pp. 509–521, 1996.
[30] D. L. Donoho, "High-dimensional centrally symmetric polytopes with neighborliness proportional to dimension", *Discrete Comput. Geometry*, 35(4), pp. 617–652, 2006.
[31] Xu, Zongben., Chang, Xiangyu., Xu, Fengmin, and Zhang Hai, "regularization: A Thresholding representation Theory and a Fast Solver", IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARING SYSTEMS, 23(7), 2012.
[32] S. McLoone, and G. Irwin, "Improving neural network training solutions using regularisation", Neurocomputing, 37, pp.71–90, 2001.
[33] K. Saito, and S. Nakano, "Second-order learning algorithm with squared penalty term", Neural Comput., 12, pp.709–729, 2000.
[34] Z. B. Xu, "Data modeling: visual psychology approach and L_1 regularization theory", in: Proceedings of the International Congress of Mathematicians, India, pp. 3151–3184, 2010.
[35] Z. Xu, X. Chang, F. Xu, and H. Zhang, "$L_{½}$ regularization: an iterative half thresholding algorithm", IEEE Transactions on Neural Networks and Learning Systems, 23(7), pp. 1013–1027, 2012.
[36] D. Ge, X. Jiang, and Y. Ye, "A note on the complexity of Lp minimization", Mathematical programming, 129 (2), pp. 285–299, 2011.
[37] D. Li, X. Zhang, and L. Wu, "Constrained optimization reformulation to $L_{1\backslash 2}$ regularization and an interior-point method", 2013, manuscript.
[38] M. Forti, P. Nistri, and M. Quincampoix, "Generalized neural network for nonsmooth nonlinear programming problems", IEEE Transactions on Circuits and Systems I, 51(9), pp. 1741–1754, 2004.
[39] S. Haykin, "Neural networks: a comprehensive foundation", (2nd ed.). Beijing: Tsinghua University Press, Prentice Hall, Beijing, 2001.
[40] Q. S. Liu, and J. Wang, "A one-layer recurrent neural network for constrained nonsmooth optimization", IEEE Transactions on Systems, Man, and Cybernetics, Part B, 41(5), pp. 1323–1333, 2011.
[41] W. Wu, G. R. Feng, and X. Li, "Training multilayer perceptrons via minimization of sum of ridge functions", Advances in Computational Mathematics, 17, pp. 331 – 347, 2002
[42] N. M. Zhang, "Deterministic convergence of an online gradient method with momentum", Lecture Notes in Computer Science, 4113, pp. 94 – 105, 2007
[43] Y. X. Yuan, and W. Y. Sun, "Optimization Theory and Methods", Science Press, Beijing, 2001

Authors Profiles

Khidir Shaib Mohamed (PhD student in Applied Mathematics) received the B.S. in Mathematics from Dalanj University – Dalanj – Sudan (2006) and M.S. in Applied Mathematics from Jilin University – Changchun – China (2011). He work as a lecturer of mathematics at College of Science – Dalanj University since (2011). Now he is a PhD student in Applied Mathematics at School of Mathematical Sciences, Dalian University of Technology, Dalian – China.

Yousif Shoaib Mohammed (Assistant Professor of Computational Physics) received the B.S. in Physics from Khartoum University – Oudurman – Sudan (1994) and High Diploma in Solar Physics from Sudan University of Science and Technology – Khartoum – Sudan (1997) and M.S. in Computational Physics (Solid – Magnetism) from Jordan University – Amman – Jordan and PhD in Computational Physics (Solid – Magnetism – Semi Conductors) from Jilin University – Changchun – China (2010). He worked at Dalanj University since 1994 up to 2013 and worked as Researcher at Africa City of Technology – Khartoum – Sudan since 2012. then from 2013 up to now at Qassim University – Kingdom of Saudi Arabia.

Abd Elmoniem Ahmed Elzain (Associate Professor in Applied Radiation Physics and researcher) received the B.S. in Physics from Kassala University – Kassala - Sudan (1996) and High Diploma of Physics from Gezira University – Madani – Sudan (1997) and M.S. in Applied Radiation Physics from Yarmouk University – Erbid – Jordan (2000) and PhD in Applied Radiation Physics from Kassala University – Kassala – Sudan (2006). He worked at Kassala University since 1996 up to 2010 then from 2010 up to now at Qassim University – Kingdom of Saudi Arabia.

Khalid Masoud Makin (Assistant Professor of Mathematics) received the B.S. in Mathematics from Kassala University – Kassala - Sudan (1996) and M.S. in Mathematics from Sudan University of Science and Technology – Khartoum – Sudan (1999) and PhD in Mathematics from Al-Nilean University – Khartoum – Sudan (2007). He worked at Kassala University since 1996 up to 2010 then from 2010 up to now at Qassim University – Kingdom of Saudi Arabia.

Elnoor Abaker Abdrhman Noh (Associate Professor of Physical Chemistry) received the B.S. in Physics from Khartoum University – Oudurman – Sudan (1991) and M.S. in Physical Chemistry (Corrosion) from Yarmouk University – Erbid – Jordan (1998) and PhD in Physical Chemistry (Computational) from North East Normal University – Changchun – China (2005). He worked at Dalanj University since 1992 up to 2011 then from 2011 up to now at Albaha University – Kingdom of Saudi Arabia.