Batch Gradient Method with Smoothing $L_{1/2}$ Regularization and Momentum for Pi-sigma Networks

Kh. Sh. Mohamed[1, 2], Y. Sh. Mohammed [3, 4, 5], A. A. Elzain[3, 6], Kh. M. Makin [3, 6], and Elnoor. A. A. Noh [4, 7]

[1]School of Mathematical Sciences, Dalian University of Technology, Dalian 116024, PR China
[2]Mathematical Department, College of Science, Dalanj University, Dalanj, Sudan

[3]Department of Physics, College of Science & Art, Qassim University, Oklat Al- Skoor, P.O.Box: 111, Saudi Arabia.

[4]Physics Department, College of Education, Dalanj University, Dalanj, Sudan

[5]Africa City for Technology, Khartoum, Sudan

[6]Department of Physics, University of Kassala, Kassala, P.O.Box: 266, Sudan

[7]Department of Chemistry, College of Art & Science, Albaha University, Baljurashi, Saudi Arabia

Abstract: ***In this paper, we study the Batch gradient method with Smoothing $L_{1/2}$ Regularization and Momentum for Pi-sigma Networks, assuming that the training samples are permuted stochastically in each cycle of iteration. The usual $L_{1/2}$ Regularization term involves absolute value and is not differentiable at the origin, which typically causes oscillation of the gradient method of the error function during the training. However, using the Smoothing approximation techniques, the deficiency of the norm $L_{1/2}$ Regularization term can be addressed. Corresponding convergence results of Smoothing $L_{1/2}$ Regularization are proved, that is, the weak convergence result is proved under the uniformly boundedness assumption of the activation function and its derivatives.***

Keywords: Pi-sigma Network, Batch Gradient Method, smoothing $L_{1/2}$ Regularization, Momentum, boundedness, convergence

1. Introduction

In the neural networks research field, The pi-sigma neural networks (PSNN) [1] is a type of feed forward polynomial neural network (FPNN) and it's the most popular models. These networks are known to provide inherently more successful mapping capability than traditional feed forward networks and to improve the learning efficiency. The neural networks consisting of the PSNN modules are widely used for classification and approximation problems. PSNN use product neurons are only composed of the product of inputs, and the number of weights required increases combinatorially with the dimension of the inputs [2-5]. To speed up and stabilize the training iteration procedure, a momentum term is often added to the increment formula for the weights so that the new weight updating rule becomes a combination of the present gradient of the error function and the previous weight updating increment [6-10]. Convergence of batch gradient method with momentum has been discussed in [11-13], under the condition that the error function is quadratic with respect to the weights. This condition was then relaxed in [14], where a gradient method with adaptive momentum for two-layer FNNs is considered and the online gradient method with momentum and its convergence for FNNs readers can refer to literature [9, 10, 15-17], in which some adaptive momentum terms are proposed and the related convergence results are established.

However, in the procedure of training FNN with SSE, the weights sometimes become very large and over-fitting tends to occur.A standard technique to prevent over- fitting is regularization, in which an extra term that penalizes large weights are added to the conventional error function [18-23]. Briefly to overview the main points of this paper, the origin modify the using smooth $L_{1\backslash 2}$ regularization in the extra term acts as a brute-force to drive unnecessary weights to zero and to prevent the weights from taking too large in the training processing into networks.

A commonly used regularization term is the squared penalty [18, 24], a term proportional to the magnitude of the network weights and many experiments have shown that to be provides a way to control the magnitude of the weight.

The regularization methods are recently developed as feasible approaches to solve the problem such that variable selection problem in machine learning. In general, the regularization methods take the form [25].

$$min = \left\{\frac{1}{n}\sum_{i=1}^{n} \iota\left(y_i, f(x_i)\right) + \lambda\|\omega\|_p^p\right\} \quad (1)$$

where $\iota(.,.)$ $l\left(y_{i,}f(x_i)\right)$ is a loss function, $\{(x_i, y_i)_{i=1}^{n}\}$ is a data set, λ is the regularization parameter, and $\|\omega\|_p^p$ p- norm of function f and is normally taken as the norm of the coefficient of linear model. Almost all the existing learning algorithms can be considered as a special form of this regularization framework. The best subset selection, namely, the L_0 penalty, along with the traditional model selection criteria such as AIC and BIC For example, when $p= 0$ [26, 27]. The L_0 regularization can be understood as a penalized least squares with penalty $\|\omega\|_0$, in which the parameter λ functions as balancing the two objective terms. The complexity of the model is proportional with the number of variable selection, and solving the model generally is intractable, particularly when N is large (It is NP-hard, see [28]). It is well known that the L_1 regularization has a very close relationship with the model Lasso and Basis Pursuit, two independent works of Tibshirani [29], and that of Chen, Donoho, and Saunders [24]. The L_1 regularization problem can be transformed into an equivalent convex quadratic optimization problem, and therefore, can be very efficiently solved. It can also result in sparse solution of the considered problem, with a promise that, under some mild conditions, the resultant solution coincides with one of the solutions of L_0 regularization [30-32]. L_1 regularization has been widely used to discourage the weights from taking large values [33, 34]. However, $p > 0$ may not lead to the sparsity-promoting property, so $p \in [0, 1]$ are required [35].

The $L_{1\backslash 2}$ regularization and propose a novel successfully applied in variable selection and feature extraction problems in high dimensional and massive data analysis. Recently, Xu et al. [36] justified that the sparsity-promotion ability of the $L_{1\backslash 2}$ problem was strongest among the L_p minimization problems with all $p \in [1/2, 1)$ and similar in $p \in (0, 1/2]$. So the $L_{1\backslash 2}$ problem can be taken as a representative of L_p $(0 < p < 1)$ problems. However, as proved by Ge et al. [37], finding the global minimal value of the $L_{1\backslash 2}$ problem was still strongly NP-hard. Finally, we propose the following $L_{1\backslash 2}$ Regularization:

$$\tilde{\beta}_{L_{1/2}} = \arg\min\left\{\frac{1}{n}\sum_{i=1}^{n}(\beta^T x_i - y_i)^2 + \lambda\|\beta\|_{L_{1/2}}\right\} \quad (2)$$

where $\lambda \geq 0$ is an appropriate regularization parameter and

$$\|\beta\|_{L_{1/2}} = \sum_{i=1}^{s}|\beta_i|^{1/2}$$

The $L_{1/2}$ regularization is a nonconvex and non-Lipschitz problem. Due to the existence of the term $|\beta_i|^{1/2}$, the objective function is even not directionally differentiable at a point with some $\beta_i = 0$, which makes the problem is very difficult to solve. Existing numerical methods that are very efficient for solving smooth problem could not be used directly. One possible way to develop numerical methods for solving equ. (2) is to smoothing the term $|\beta_i|^{1/2}$ using some smoothing function. However, it is easy to see that the derivative of the smoothing function will be unbounded and differentiable. Consequently, it is not desirable that the smoothing function based numerical methods could work well. Recently, the $L_{1/2}$ regularization has been successfully applied in [38] proposed a constrained optimization reformulation to the unconstrained $L_{1/2}$ regularization problem. The reformulation is to minimizing a smooth function subject to some quadratic constraints and nonnegative constraints such as the generalized gradient (GG) and recurrent neural network (RNN) methods shown as [39, 40].

In this paper, we study the deterministic convergence of the batch gradient method with both momentum term and smoothing $L_{1/2}$ regularization term. Note that the usual $L_{1/2}$ regularization is not smooth at the origin, which makes the problem is very difficult to solve. To overcome this drawback using smoothing function, so it is easy to see that the derivative of the smoothing function will be unbounded and differentiable.

The rest of this paper is organized as follows. In the next section the batch gradient method with smoothing $L_{1/2}$ regularization penalty term and momentum is described for training PSNN model. In section 3, the convergence results of batch gradient method with smoothing $L_{1/2}$ regularization and momentum are presented. The detailed proofs of the main results are gathered in section 4. Finally, we conclude this paper in section 5.

2. Batch gradient method with smoothing $L_{1\backslash 2}$ regularization and momentum (BGMSRM)

2.1. Batch gradient method with $L_{1\backslash 2}$ regularization and momentum

Consider a three-layer network consisting of P input node, N hidden nodes, and one output nodes. Suppose that by $\omega_k = (\omega_{k1}, \dots, \omega_{kp})^T \in R^p$ be the weight vector between the input units and the hidden unit $(k = 1, 2, \dots, N)$. To simplify the presentation, we write all the weight parameters in a compact form $\omega = (\omega_1^T, \dots, \omega_N^T) \in R^{NP}$. the weights on the connections between the product node and the summation node are fixed to 1. We have included a special input unit $\xi_P = -1$, corresponding to the biases ω_{np} with fixed value 1. The topological structure of PSNN is shown in Fig. 1.

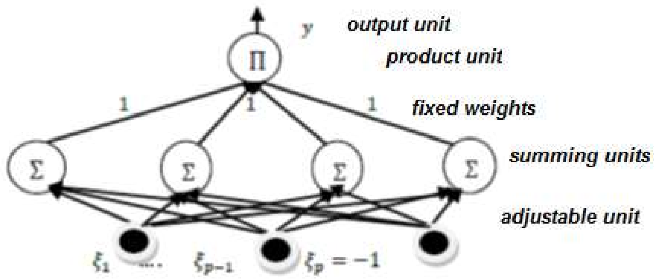

Figure 1: PSNN structure with a single output

suppose that $\{x^j, o^j\}_{j=1}^J \subset R^p \times R$ is a given set of training samples, where $\xi^j = (\xi_1^j, \xi_2^j, \dots \xi_p^j) \in \mathbb{R}^p$. Let $g: \mathbb{R} \to \mathbb{R}$ be a transfer function for the hidden and output node, which is typically, but not necessarily, a sigmoid function. for an input $x \in \mathbb{R}^p$, the output of of the network can be written as

$$y = g\left(\prod_{i=1}^{N}\left(\sum_{k=1}^{p}(\omega_{ik}x_i)\right)\right) = g\left(\prod_{i=1}^{N}(\omega_i.x)\right) \quad (3)$$

For simplicity, we write $g_j(t) = \frac{1}{2}(o^j - g(t))^2$. Define the error function takes the form

$$\check{E}(\omega) = \frac{1}{2}\sum_{j=1}^{J}\left(o^j - g\left(\prod_{i=1}^{N}(\omega_i.\xi^j)\right)\right)^2$$

$$= \sum_{j=1}^{J} g_j\left(\prod_{i=1}^{N}(\omega_i.\xi^j)\right) \quad (4)$$

Its gradient with respect to the weight vector ω_k $(k = 1, 2, 3, \dots N)$ is

$$\breve{E}_{\omega_k}(\omega) = \sum_{j=1}^{J} g_j'\left(\prod_{i=1}^{N}(\omega_i.\xi^j)\right)\prod_{\substack{i=1\\ i\neq n}}^{N}(\omega_i.\xi^j)\,\xi^j$$

By adding $L_{1/2}$ regularization penalty term, the modified cost error function takes the form

$$E(\omega) = \tilde{E}(\omega) + \lambda\sum_{k=1}^{N}|\omega_k|^{1/2} \quad (5)$$

where $\lambda > 0$ is a penalty coefficient.

The gradient of error function by adding $L_{1/2}$ regularization penalty term with respect to ω_k $(k = 1, 2, 3, \dots, N)$. i.e.,

$$E_{\omega_k}(\omega) = \breve{E}_{\omega_k}(\omega) + \frac{\lambda\, sgn\,(\omega_k)}{2\,|\omega_k|^{1/2}} \quad (6)$$

Given an initial weight $\omega^0 \epsilon R^p$, the batch gradient method with $L_{1\backslash 2}$ update the weights iteratively by

$$\omega_k^{mJ+j} = \omega_k^{mJ+j-1} - \eta_m \Delta_j^m \omega_k^{mJ+j}, \quad (7)$$

where $m = 0, 1, 2, \dots;\ j = 1, 2, \dots, J$;

$$\Delta_j^m \omega_k^{mJ+j} = -\eta_m\left[\breve{E}_{\omega_k^{mJ+j-1}}(\omega) + \frac{\lambda\, sgn\,(\omega_k^{mJ+j-1})}{2\,J\,\left|\omega_k^{mJ+j-1}\right|^{1/2}}\right] \quad (8)$$

where $\lambda > 0$ represents the learning rate in the $k-th$ training epoch. To speed up and stabilize the training iteration procedure, a momentum term is often added to the increment rule equ. (8) and this gives the batch gradient method with $L_{1\backslash 2}$ regularization and momentum, i.e.,

$$\Delta_j^m \omega_k^{mJ+j} = -\eta_m\left[\breve{E}_{\omega_k^{mJ+j-1}}(\omega) + \frac{\lambda\, sgn\,(\omega_k^{mJ+j-1})}{2\,J\,\left|\omega_k^{mJ+j-1}\right|^{1/2}}\right] + \alpha_{mj}\,\Delta\omega_k^{mJ+j-1} \quad (9)$$

where α_{mj} is the momentum coefficient with respect to the term $\Delta\omega_k^{mJ+j-1}$.

2.2 Smoothing $L_{1\backslash 2}$ regularization and momentum

A modified $L_{1\backslash 2}$ regularization term is proposed by smoothing the usual one at the origin, resulting in the following error function with a smoothing $L_{1\backslash 2}$ regularization penalty term:

$$E(\omega) = \tilde{E}(\omega) + \lambda\sum_{k=1}^{N}|\omega_k|^{1/2} \quad (10)$$

where $f(x)$ is a smooth function that approximates $|x|$. for definiteness and simplicity, we choose $f(x)$ as a piecewise polynomial function:

$$f(x) = \begin{cases} |x| \; if \; |x| \geq a \\ -\frac{1}{8a^3}x^4 + \frac{3}{4a}x^2 + \frac{3}{8} \; if \; |x| < a \end{cases} \quad (11)$$

where a is a small positive constant. Then It is easy to get

$$f(x) \in \left[\frac{3}{8}a, +\infty\right), f'^{(x)} \in [-1, 1], f''^{(x)} \in \left[0, \frac{3}{2a}\right] \quad (12)$$

The gradient of the error function can be written as in equ. (4) with

$$E_{\omega_k}(\omega) = \breve{E}_{\omega_k}(\omega) + \frac{\lambda\, f'(\omega_k)}{2J\, f\,(\omega_k)^{1/2}} \quad (13)$$

where $\lambda > 0$ is a penalty parameter and $k = 1, 2, 3, \dots, N$.

Given an initial weight W^0, the batch gradient method with $L_{1\backslash 2}$ regularization penalty update the weights $\{W^m\}$ iteratively by

$$\omega_k^{mJ+j} = \omega_k^{mJ+j-1} - \eta_m \Delta_j^m \omega_k^{mJ+j}, \quad (14)$$

where $m = 0, 1, 2, \dots;\ j = 1, 2, \dots, J$; and

$$\Delta_j^m \omega_k^{mJ+j} = -\eta_m\left[\breve{E}_{\omega_k^{mJ+j-1}}(\omega) + \frac{\lambda\, f'(\omega_k^{mJ+j-1})}{2J\, f(\omega_k^{mJ+j-1})^{1/2}}\right] \quad (15)$$

where $\lambda > 0$ is a penalty parameter and $k = 1, 2, 3, \dots, N$. To speed up and stabilize the training iteration procedure, a momentum term is often added to the increment rule equ. (15) and this gives the batch gradient method with $L_{1\backslash 2}$ regularization and momentum, i.e.,

$$\Delta_j^m \omega_k^{mJ+j} = -\eta_m\left[\breve{E}_{\omega_k^{mJ+j-1}}(\omega) + \frac{\lambda\, f'(\omega_k^{mJ+j-1})}{2J\, f(\omega_k^{mJ+j-1})^{1/2}}\right] + \alpha_{mj}\,\Delta_j^m \omega_k^{mJ+j-1} \quad (16)$$

where α_{mj} is the momentum coefficient with respect to the term $\Delta_j^m \omega_k^{mJ+j-1}$.

For the sake of description, we denote

$$p^{m,i,j} = -\eta_m\left[\breve{E}_{\omega_k^{mJ+j-1}}(\omega) + \frac{\lambda\, f'(\omega_k^{mJ+j-1})}{2J\, f(\omega_k^{mJ+j-1})^{1/2}}\right] \quad (17)$$

Particularly when $i = 1$ denote

$$p^{m,j} = -\eta_m\left[\breve{E}_{\omega_k^{mJ}}(\omega) + \frac{\lambda\, f'(\omega_k^{mJ})}{2J\, f(\omega_k^{mJ})^{1/2}}\right] \quad (18)$$

Then there holds

$$E_{\omega_k^{mJ}}(\omega) = \sum_{j=1}^{J} p^{m,j} \quad (19)$$

and the learning rule equ. (16) becomes

$$\Delta_j^m \omega_k^{mJ+j-1} = -\eta_m p^{m,J,J} + \Delta_j^m \omega_k^{mJ+j-1},$$
$$m = 0, 1, \dots; j = 1, 2, \dots. J \quad (20)$$

In this work, by choosing an initial $\eta_0 \in (0, 1]$ and positive constant β, we inductively determine η_m in equ. (18) by (cf.[41])

$$\frac{1}{\eta_{m+1}} = \frac{1}{\eta_m} + \beta, m = 0, 1, 2 \dots\dots (21)$$

It is easy to get from equ. (20) that $\eta_m = -\eta_0/(1 + m\beta\eta_0)$ for $m = 0, 1, \dots$ hence there hold $\eta_m = o(1\backslash m)$ and for $\eta_m \to 0\ as\ m \to \infty$, and for the momentum coefficients α_{mj} in equ. (19), then we choose them by the rule

$$\alpha_{mj} = \begin{cases} \frac{\eta_m^2 \|p^{mJ}\|}{\left\|\Delta\omega_k^{mJ+j-1}\right\|} \; if \left\|\Delta\omega_k^{mJ+j-1}\right\| \neq 0 \\ 0 \; else \end{cases} \quad (22)$$

3. Convergence Results

In this section we present some convergence theorems of the Batch gradient method with smoothing $L_{1\backslash 2}$ regularization with momentum term in equ. (14). To analyze the convergence of the algorithm, we need the following assumptions.

Assumption (A1) $\left|g_j'(t)\right|$ and $\left|g_j''(t)\right|$ $(j = 1, 2, \dots., J)$ are uniformly bounded, i.e., there is a constant $C > 0$ such that $\left|g_j'(t)\right| \leq C$ and $\left|g_j''(t)\right| \leq C$.

Assumption (A2) inequality (75) valid, and β and η_0 in equ. (19) satify $\beta > max\{1, \beta_0\}$ and $0 < \eta_0 \leq min\{1, 1/\beta_0 - 1/\beta\}$.

Assumption (A3) the set $\Omega_0 \in \{w \in \Omega : E_w(w) = 0\}$ Contains finite points, where Ω is closed bounded region such that $\{\omega^m\} \subset \Omega$.

Theorem 1 。 Let the error function be given by equ. (10), let Assumptions (A1) and (A2) be satisfied, and let the weight $\{\omega^m\}$ be generated by the algorithm equ. (14). Then there holds

$$E(\omega^{(m+1)J}) \le E(\omega^{mJ}), m = 0, 1, \dots\dots$$

Theorem 2. Under the same assumption of Theorem 1, the weight sequence $\{\omega^m\}$ generated by equ. (16) is uniformly bounded.

Theorem 3. Let the error function be given by equ. (10), and let the weight $\{\omega^m\}$ be updated by equ. (16). if Assumption (A1) and (A2) are valid, then there holds the following weak convergence result

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

Furthermore, if Assumption (A4) is also valid, we have the strong convergence

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

4. Proofs

For convenient, we use the following notation:

$$r_k^{m,j} = p^{m,j,j} - p^{m,j}, m = 0, 1, 2, \dots, j = 1, 2, \dots, J, 1 \le k \le N \quad (23)$$

$$v_k^{m,j} = \omega_k^{(m+1)J} - \omega_k^{mJ}, m = 0, 1, 2, .. \quad (24)$$

$$v_k^{m,j} = \sum_{i=1}^{j-1}\left(\alpha_{m,i}\,\Delta_j^m\omega_k^{mJ+j-1} - \eta_m p^{m,i,i}\right), 1 \le k \le N, 1 \le j \le J, m = 0, 1, 2, \dots \quad (25)$$

Then, by the error function equ. (10) we have

$$E(\omega^{(m+1)J}) = \sum_{j=1}^{J} g_j\left(\prod_{i=1}^{N}(\omega_i^{(m+1)J}.\xi^j)\right) + \lambda\sum_{k=1}^{N} f(\omega_k^{(m+1)J})^{\frac{1}{2}} \quad (26)$$

$$E(\omega^{mJ}) = \sum_{j=1}^{J} g_j\left(\prod_{i=1}^{N}(\omega_i^{mJ}.\xi^j)\right) + \lambda\sum_{k=1}^{N} f(\omega_k^{mJ})^{\frac{1}{2}} \quad (27)$$

***Lemma* 5.** *Let $\{\eta_m\}$ be given by equ. (21). There hold*

$$0 < \eta_m < \eta_{m+1} \le 1, m = 1, 2, \dots\dots \quad (28)$$

$$\frac{\tau}{m} < \eta_m < \frac{\rho}{m}, \tau = \frac{\eta_0}{1+\eta_0\beta}, \rho = \frac{1}{\beta}, m = 1, 2, \dots.. \quad (29)$$

Proof. This lemma is easy to validate by virtue of equ. (21) and $\eta_0 \in (0, 1]$.

***Lemma* 6.** *If Assumption (A1) is valid, for $\{\eta_m\}$ satisfies equ. (21), there holds*

$$\sum_{j=1}^{j}\|r_{m,j}\| \le C_2\eta_m\sum_{j=1}^{J}\|p^{m,j}\| \quad (30)$$

$$\|v_k^{m,j}\| \le \left\|\sum_{j=1}^{J} p^{m,j}\right\| + C_3\eta_m^2\sum_{j=1}^{J}\|p^{m,j}\| \quad (31)$$

$$\|v_k^{m,j}\|^2 \le C_4\eta_m^2\sum_{j=1}^{J}\|p^{m,j}\|^2 \quad (32)$$

where C_2, C_3 *and* C_4 are positive constants and $m = 0, 1, 2, \dots; j = 1, 2, \dots J$.

Proof. By Assumption (A2), (A2), equ. (30) and Cauchy-Schwartz, we have

$$\left|\prod_{i=1}^{N}(\omega_i^{mJ+j}.\xi^j) - \prod_{i=1}^{N}(\omega_i^{mJ}.\xi^j)\right| \le \left|\prod_{i=1}^{N-1}(\omega_i^{mJ+j}.\xi^j)\right|\left|(\omega_N^{mJ+j} - \omega_N^{mJ})\xi^j\right| + \left|\prod_{i=1}^{N-2}(\omega_i^{mJ+j}.\xi^j)(\omega_N^{mJ}.\xi^j)\right|\left|(\omega_{N-1}^{mJ+j} - \omega_{N-1}^{mJ})\xi^j\right| + \dots..+\left|\prod_{i=1}^{N}(\omega_i^{mJ}.\xi^j)\right|\left|(\omega_1^{mJ+j} - \omega_1^{mJ})\xi^j\right| \le C_5\left(\sum_{i=1}^{j-1}2\|p^{m,i}\| + \sum_{i=1}^{j-1}\|r_k^{i,m}\|\right) \quad (33)$$

where $C_5 = C^N(1 \le j \le J, 1 \le k \le N, m = 0, 1, 2, \dots)$. Similarly, easy to get

$$\left|\prod_{\substack{i=1\\i\ne k}}^{N}(\omega_i^{mJ+j}.\xi^j) - \prod_{\substack{i=1\\i\ne k}}^{N}(\omega_i^{mJ}.\xi^j)\right| \le C_6\left(\sum_{i=1}^{j-1}2\|p^{m,i}\| + \sum_{i=1}^{j-1}\|r_k^{i,m}\|\right) \quad (34)$$

where $C_6 = C^{N-1}$ $(1 \le j \le J, 1 \le k \le N, m = 0, 1, 2, \dots)$. combination equ. (22), equ. (23) and $0 < \eta_m \le 1$ gives

$$\|\omega_k^{mJ+j-1} - \omega_k^{mJ}\| \le \sum_{i=1}^{j-1}\left(\alpha_{m,i}\,\|\Delta_j^m\omega_k^{mJ+j-1}\| + \eta_m\|p^{m,i}\| + \eta_m\|r_{m,i}\|\right) \le \sum_{i=1}^{j-1}\left(\eta_m^2\,\|p^{m,i}\| + \eta_m\|p^{m,i}\| + \eta_m\|r_k^{i,m}\|\right) \le \eta_m\left(\sum_{i=1}^{j-1}2\|p^{m,i}\| + \sum_{i=1}^{j-1}\|r_k^{i,m}\|\right), j = 2, 3, \dots\dots J \quad (35)$$

By Assumption (A1), (A2), equ. (15), equ. (22), equ. (33), equ. (34), equ. (35) and differential mean value theorem, we have

$$\|r_k^{m,j}\| = \left\|\eta_m g_j''\left(\prod_{i=1}^{N}(\omega_i^{mJ+j-1}.\xi^j)\right)\prod_{\substack{i=1\\i\ne k}}^{N}(\omega_i^{mJ+j-1}.\xi^j)\xi^j - \eta_m g_j'\left(\prod_{i=1}^{N}(\omega_i^{mJ}.\xi^j)\right)\prod_{\substack{i=1\\i\ne k}}^{N}(\omega_i^{mJ}.\xi^j)\xi^j + \frac{\lambda}{2J}\left(\frac{f'(\omega_k^{mJ+j-1})}{f(\omega_k^{mJ+j-1})^{1/2}} - \frac{f'(\omega_k^{mJ})}{f(\omega_k^{mJ})^{1/2}}\right)\right\|$$

$$\le \left\|\eta_m g_j''(t_{j,m})\prod_{\substack{i=1\\i\ne k}}^{N}(\omega_i^{mJ+j-1}.\xi^j)\left(\prod_{i=1}^{N}(\omega_i^{mJ+j-1}.\xi^j) - \prod_{i=1}^{N}(\omega_i^{mJ}.\xi^j)\right)\xi^j\right\|$$

$$+\left\|\eta_m g_j'\left(\prod_{i=1}^{N}(\omega_i^{mJ}.\xi^j)\right)\left(\prod_{\substack{i=1\\i\neq k}}^{N}(\omega_i^{mJ+j-1}.\xi^j) - \prod_{\substack{i=1\\i\neq k}}^{N}(\omega_i^{mJ}.\xi^j)\right)\xi^j\right\| + \frac{\lambda}{2J}F''(t_{n,k,m,j})\|\omega_k^{mJ+j-1}-\omega_k^{mJ}\|$$

$$\leq (C_5C^{N+1}+C_6C^2+\lambda M)\,\eta_m\left(\sum_{i=1}^{j-1}2\|p^{m,i}\| + \sum_{i=1}^{j-1}\|r_k^{i,m}\|\right)$$

$$\leq C_7\eta_m\sum_{i=1}^{j-1}\left(2\,\|p^{m,i}\| + \|r_k^{i,m}\|\right), j=2,3,\ldots.J \quad (36)$$

where $t_{j,m}, \tilde{t}_{n,m,j}$ and $t_{i,k,m,j}$ are suitable constants. and $C_7 = C_5C^{N+1}+C_6C^2+\lambda M, j=2,3,\ldots.,J$.

Note that for any $j=1$, denotation functions equ. (18) and equ. (23) imply

$$\|r_k^{1,m}\| = 0 \quad (37)$$

This together with equ. (36) we get

$$\|r_k^{2,m}\| \leq C_7\eta_m(2\,\|p^{m,1}\| + \|r_k^{1,m}\|) = 2C_7\eta_m\|p^{m,1}\| \quad (38)$$

And

$$\|r_k^{2,m}\| \leq C_7\eta_m(2\,\|p^{m,2}\| + \|r_k^{2,m}\| + 2\,\|p^{m,1}\|) = 2C_7(1+C_7)\eta_m(\|p^{m,2}\| + \|p^{m,1}\|) \quad (39)$$

Applying an induction on $\|r_k^{j,m}\|$, we have for $2\leq j\leq J$

$$\|r_k^{j,m}\| \leq 2C_7(1+C_7)^{j-2}\eta_m\sum_{i=1}^{j-1}\|p^{m,i}\| \quad (40)$$

A sum of $j=1,2,\ldots.J$ yields equ. (30) in Lemma 6 immediately:

$$\sum_{j=1}^{J}\|r_k^{j,m}\| = \sum_{j=2}^{J}\|r_k^{j,m}\| \leq C_2\eta_m\sum_{j=1}^{J}\|p^{mj}\| \quad (41)$$

where $C_2 = 2C_7\sum_{j=1}^{J}(1+C_7)^{j-2}$.

Next, we prove equ. (31). In view of equ. (24) and equ. (25), we have

$$v_k^{m,j} = \sum_{j=1}^{J}\left(\alpha_{mj}\,\Delta\omega_k^{mJ+j-1} - \eta_m P^{m\,j} - \eta_m r_k^{j,m}\right) \quad (42)$$

Setting $C_3=(1+C_2)$ and using equ. (24) and equ. (41), there holds

$$\|v_k^{m,j}\| \leq \eta_m\left\|\sum_{j=1}^{J}P^{m\,j}\right\| + \eta_m\sum_{j=1}^{J}\|r_k^{j,m}\| + \sum_{j=1}^{J}\alpha_{mj}\|\Delta\omega_k^{mJ+j-1}\|$$

$$\leq \eta_m\left\|\sum_{j=1}^{J}P^{m\,j}\right\| + C_2\eta_m^2\sum_{j=1}^{J}\|P^{m\,j}\| + \eta_m^2\sum_{j=1}^{J}\|P^{m\,j}\|$$

$$= \eta_m\left\|\sum_{j=1}^{J}P^{m\,j}\right\| + C_3\eta_m^2\sum_{j=1}^{J}\|P^{m\,j}\| \quad (43)$$

Finally, we prove equ. (32) by virtue of equ. (43).

Again using $0<\eta_m\leq 1$, the estimation equ. (43) can be rewritten as

$$\|v_k^{m,j}\| \leq (1+C_3)\,\eta_m\sum_{j=1}^{J}\|P^{m\,j}\| \quad (44)$$

Squaring two sides of equ. (44) and applying Cauchy-Schwartz inequality, we have

$$\|v_k^{m,j}\|^2 \leq (1+C_3)^2\eta_m^2\left(\sum_{j=1}^{J}\|P^{m\,j}\|\right)^2 \leq C_4\eta_m^2\sum_{j=1}^{J}\|p^{mj}\|^2 \quad (45)$$

where $C_4 = J(1+C_3)^2$. the proof it is completed.

Lemma 7.

If Assumption $(A1)$ *is valid and* η_m *satisfies* equ. (21) *thus hold*

$$\left|\left(-\eta_m\sum_{j=1}^{J}p^{m,j}\right).\left(\sum_{j=1}^{J}r_{m,j}\right)\right| \leq C_8\eta_m^2\sum_{j=1}^{J}\|p^{m,i}\|^2 \quad (46)$$

$$\left|\left(\sum_{j=1}^{J}p^{m,j}\right).\left(\sum_{j=1}^{J}\alpha_{mj}\,\Delta\omega_k^{mJ+j-1}\right)\right| \leq J\eta_m^2\sum_{j=1}^{J}\|p^{m,j}\|^2 \quad (47)$$

Proof.

It is similar to the proof of Lemma 3 in [43] and thus omitted. To prove the monotonicity of $E(\omega^{mJ})$, first we need to give an estimation of the difference $E(\omega^{(m+1)J}) - E(\omega^{mJ})$. shown in the following lemma.

Lemma 8.

There is a positive constant γ independent of m such that

$$E(\omega^{(m+1)J}) \leq E(\omega^{mJ}) - \eta_m\left\|\sum_{j=1}^{J}p^{m,j}\right\|^2 + \gamma\eta_m^2\sum_{j=1}^{J}\|p^{m,j}\|^2 \quad (48)$$

Proof

Using Taylor's formula to first and second orders for

$$\prod_{i=1}^{N}(\omega_i^{(m+1)J}.\xi^j) = \prod_{i=1}^{N}(\omega_i^{mJ}.\xi^j) - \sum_{k=1}^{N}\left(\prod_{\substack{i=1\\i\neq k}}^{N}(\omega_i^{mJ+j}.\xi^j)(\omega_k^{(m+1)J} - \omega_k^{mJ})\xi^j\right)$$

$$+\frac{1}{2}\sum_{\substack{n_1,n_2=1\\ n_1\neq n_2}}^{N}\left(\prod_{\substack{i=1\\ i\neq k_1,k_2}}^{N} t_{i,m,j}\right)[(\omega_{k_1}^{(m+1)J}-\omega_{k_1}^{mJ})\xi^j][(\omega_{k_2}^{(m+1)J}-\omega_{k_2}^{mJ})\xi^j] \quad (49)$$

where $t_{i,m,j}\in\mathbb{R}$ is on the line segment between $\omega_i^m.\xi^j$ and $\omega_i^{m+1}.\xi^j$. Again applying the Taylor expansion and noting equ. (25) and equ. (49), we have

$$g_j\left(\prod_{i=1}^{N}(\omega_i^{(m+1)J}.\xi^j)\right)=\left(\prod_{i=1}^{N}(\omega_i^{mJ}.\xi^j)\right)+g_j'\left(\prod_{i=1}^{N}(\omega_i^{mJ}.\xi^j)\right)\sum_{k=1}^{N}\left(\prod_{\substack{i=1\\ i\neq k}}^{N}(\omega_i^{mJ+j}.\xi^j)(\omega_k^{(m+1)J}-\omega_k^{mJ})\xi^j\right)$$

$$+\frac{1}{2}\sum_{\substack{k_1,k_2=1\\ k_1\neq k_2}}^{N}\left(\prod_{\substack{i=1\\ i\neq n_1,n_2}}^{N} t_{i,m,j}\right)[(\omega_{k_1}^{(m+1)J}-\omega_{k_1}^{mJ})\xi^j][(\omega_{k_2}^{(m+1)J}-\omega_{k_2}^{mJ})\xi^j]$$

$$+\frac{1}{2}g_j''(t_{i,m})\left(\prod_{i=1}^{N}(\omega_i^{(m+1)J}.\xi^j)-\prod_{i=1}^{N}(\omega_i^{mJ}.\xi^j)\right)^2 \quad (50)$$

where $t_{i,k,m}\in\mathbb{R}$ is on the line segment between $\omega_i^m.\xi^j$ and $\omega_i^{m+1}.\xi^j$. for $j=2,3,..,J$.

By using equs. (15), (26) - (28), and Taylor expansion, we get

$$E(\omega^{(m+1)J})=E(\omega^{mJ})+\sum_{j=1}^{J}g_j'\left(\prod_{i=1}^{N}(\omega_i^{mJ}.\xi^j)\right)\sum_{k=1}^{N}\left(\prod_{\substack{i=1\\ i\neq k}}^{N}(\omega_i^{mJ+j}.\xi^j)(\omega_k^{(m+1)J}-\omega_k^{mJ})\xi^j\right)$$

$$+\frac{\lambda}{2}\sum_{k=1}^{N}\sum_{j=1}^{J}\left(\frac{f'(\omega_k^{mJ})}{f\,(\omega_k^{mJ})^{1/2}}+F''(t_{j,k,m})v_k^{m,j}\right)v_k^{m,j}+\sigma_1+\sigma_2 \quad (51)$$

where $t_{i,m}, t_{j,k,m}\in\mathbb{R}$ line segment between $\omega_i^{(m+1)J}.\xi^j$ and $\omega_i^{mJ}.\xi^j$, $\sigma_1=\frac{1}{2}\sum_{\substack{n_1,n_2=1\\ n_1\neq n_2}}^{N}\left(\prod_{\substack{i=1\\ i\neq k_1,k_2}}^{N} t_{i,m,j}\right)[(\omega_{k_1}^{(m+1)J}-\omega_{k_1}^{mJ})\xi^j][(\omega_{k_2}^{(m+1)J}-\omega_{k_2}^{mJ})\xi^j]$ and $\sigma_2=\frac{1}{2}g_j''(t_{i,m}')(\prod_{i=1}^{N}(\omega_i^{(m+1)J}.\xi^j)-\prod_{i=1}^{N}(\omega_i^{mJ}.\xi^j))^2$.

Noticing by equ. (23) and equ. (24), easily to get

$$p^{m,j,j}=p^{m,j}+[p^{m,j,j}-p^{m,j}]$$
$$=p^{m,j}+\left[\left(\sum_{j=1}^{N}g_j'\left(\prod_{i=1}^{N}(\omega_i^{mJ+j-1}.\xi^j)\right)\prod_{\substack{i=1\\ i\neq k}}^{N}(\omega_i^{mJ+j-1}.\xi^j)\xi^j-\sum_{j=1}^{N}g_j'\left(\prod_{i=1}^{N}(\omega_i^{mJ}.\xi^j)\right)\prod_{\substack{i=1\\ i\neq k}}^{N}(\omega_i^{mJ}.\xi^j)\,\xi^j\right)\right.$$
$$\left.+\lambda\left(\frac{f'(\omega_k^{mJ+j-1})}{2J\,f(\omega_k^{mJ+j-1})^{1/2}}-\frac{f'(\omega_k^{mJ})}{2J\,f(\omega_k^{mJ})^{1/2}}\right)\right] \quad (52)$$

It holds

$$E(\omega^{(m+1)J})-E(\omega^{mJ})$$
$$=\sum_{j=1}^{J}\left(g_j'\left(\prod_{i=1}^{N}(\omega_i^{mJ}.\xi^j)\right)\prod_{\substack{i=1\\ i\neq k}}^{N}(\omega_i^{mJ+j}.\xi^j)\xi^j+\frac{\lambda f'(\omega_k^{mJ})}{2\,J\,f\,(\omega_k^{mJ})^{1/2}}\right)\left(\sum_{j=1}^{J}(\alpha_{m,j}\,\Delta\omega_k^{mJ+j-1}-\eta_m p^{m,j,j})\right)$$
$$+\frac{\lambda}{2}\sum_{k=1}^{N}F''(t_{j,k,m})(v_k^{m,j})^2+\sigma_1+\sigma_2$$
$$=\sum_{j=1}^{J}\left(g_j'\left(\prod_{i=1}^{N}(\omega_i^{mJ}.\xi^j)\right)\prod_{\substack{i=1\\ i\neq k}}^{N}(\omega_i^{mJ+j}.\xi^j)\xi^j+\frac{\lambda f'(\omega_k^{mJ})}{2\,f\,(\omega_k^{mJ})^{1/2}}\right)\left(\sum_{j=1}^{J}\alpha_{m,j}\,\Delta\omega_k^{mJ+j-1}-\eta_m\sum_{j=1}^{J}p^{m,j}\right.$$
$$-\eta_m\left(\left(\sum_{j=1}^{N}g_j'\left(\prod_{i=1}^{N}(\omega_i^{mJ+j-1}.\xi^j)\right)\prod_{\substack{i=1\\ i\neq k}}^{N}(\omega_i^{mJ+j}.\xi^j)\xi^j-\sum_{j=1}^{N}\delta_j\left(\prod_{i=1}^{N}(\omega_i^{mJ}.\xi^j)\right)\prod_{\substack{i=1\\ i\neq n}}^{N}(\omega_i^{mJ}.\xi^j)\,\xi^j\right)\right.$$
$$\left.\left.+\lambda\left(\frac{f'(\omega_k^{mJ+j-1})}{2J\,f(\omega_k^{mJ+j-1})^{1/2}}-\frac{f'(\omega_k^{mJ})}{2J\,f(\omega_k^{mJ})^{1/2}}\right)\right)\right)$$
$$+\frac{\lambda}{2}\sum_{k=1}^{N}F''(t_{j,k,m})(v_k^{m,j})^2+\sigma_1+\sigma_2$$
$$=\left(\sum_{j=1}^{J}p^{m,j}\right).\left(\sum_{j=1}^{J}\alpha_{m,j}\,\Delta\omega_k^{mJ+j-1}-\eta_m\sum_{j=1}^{J}p^{m,j}-\eta_m\sum_{j=1}^{J}p^{m,j,j}+\eta_m\sum_{j=1}^{J}p^{m,j}\right)$$
$$=\left(\sum_{j=1}^{J}p^{m,j}\right)\left(\sum_{j=1}^{J}\alpha_{m,j}\,\Delta\omega_k^{mJ+j-1}-\eta_m\sum_{j=1}^{J}p^{m,j}-\eta_m\sum_{j=1}^{J}(p^{m,j,j}+p^{m,j})\right)$$
$$+\frac{\lambda}{2}\sum_{k=1}^{N}F''(t_{j,k,m})(v_k^{m,j})^2+\sigma_1+\sigma_2$$

where $t_{j,k,m}$ lies in between $\omega_i^{mJ}.x^j$ and $\omega_i^{(m+1)J}.x^j$, and from equ. (25), equ. (28) and equ. (45), $M=\frac{\sqrt{6}}{\sqrt{a^3}}$, and $F(x)\equiv(f(x))^{\frac{1}{2}}$. Note that

$$F'(x) = \frac{f'(x)}{2\sqrt{f(x)}}$$

$$F''(x) = \frac{2f''(x).f(x) - [f'(x)]^2}{4[f(x)]^{\frac{3}{2}}}$$

$$\leq \frac{f''(x)}{2\sqrt{f(x)}} \leq \frac{\sqrt{6}}{2\sqrt{a^3}}$$

By using equ. (32) and Lemma 5 for $1 \leq j \leq J, 1 \leq k \leq N, m = 0, 1, 2, \ldots$, and Cauchy-Schwartz Theorem, we have

$$\frac{\lambda}{2}\sum_{k=1}^{N} F''(t_{j,k,m})(v_k^{m,j})^2 \leq \lambda M C_4 \eta_m^2 \sum_{j=1}^{J}\|p^{m,j}\|^2$$

$$\leq C_9 \sum_{j=1}^{J}\|p^{m,j}\|^2 \quad (54)$$

where $C_9 = \lambda MJ(1 + C_3)^2$.

by Assumption (A1), (A2), equ. (30), equ. (33) and Cauchy–Schwarz for $j = 2, 3, \ldots\ldots$, we get

$$|\rho_1| = \frac{1}{2}\left|\sum_{j=1}^{J} g_j''(t_{i,m})\left(\prod_{i=1}^{N}(\omega_i^{(m+1)J}.\xi^j) - \prod_{i=1}^{N}(\omega_i^{mJ}.\xi^j)\right)^2\right|$$

$$\leq \frac{1}{2}JCC_5^2\,(2 + C_2\eta_0)^2\eta_m^2 \sum_{j=1}^{J}\|p^{m,i}\|^2$$

$$\leq C_{10}\sum_{j=1}^{J}\|p^{m,j}\|^2 \quad (55)$$

where $C_{10} = \frac{1}{2}JCC_5^2\,(2 + C_2\eta_m)^2$.

Using Assumption (A1), (A2), equ. (32), $0 \leq \eta_m \leq 1$ and Cauchy–Schwarz for $j = 2, 3, \ldots\ldots$, we have

$$|\rho_2| \leq \frac{1}{2}\left|\sum_{j=1}^{J}\sum_{\substack{k_1,k_2=1 \\ k_1\neq k_2}}^{N}\left((v_{k_1}^{m,j})\xi^j\right)\left((v_{k_2}^{m,j})\xi^j\right)\right|$$

$$\leq \frac{1}{2}C^{(N+1)}J\sum_{\substack{k_1,k_2=1 \\ k_1\neq k_2}}^{N}\|v_{k_1}^{m,j}\|.\|v_{k_2}^{m,j}\|$$

$$\leq \frac{1}{2}C^{(N+1)}J(1-N)(1+C_3)^2\,\eta_m^2\sum_{j=1}^{J}\left(\sum_{j=1}^{J}\|p^{m,j}\|\right)^2$$

$$\leq C_{11}\sum_{j=1}^{J}\|p^{m,j}\|^2 \quad (56)$$

where $C_{11} = \frac{1}{2}C^{(N+1)}J(1-N)(1+C_3)^2\,\eta_m^2$. Set

$$\gamma = J + C_8 + C_9 + C_{10} + C_{11} \quad (57)$$

Obviously, γ is a positive constant independent of the iteration m.

Substituting **Lemma 8** and **7**, equs. (54)-(57) into equ. (53) immediately, i.e.,

$$E(\omega^{(m+1)J}) - E(\omega^{mJ})$$

$$\leq -\eta_m\left\|\sum_{j=1}^{J}p^{m,j}\right\|^2$$

$$+ (J + C_8 + C_9 + C_{10}$$

$$+ C_{11})\eta_m^2\sum_{j=1}^{J}\|p^{m,j}\|^2$$

$$= -\eta_m\left\|\sum_{j=1}^{J}p^{m,j}\right\|^2 + \gamma\eta_m^2\sum_{j=1}^{J}\|p^{m,j}\|^2 \quad (58)$$

The proof it is completed.

Lemma 9.

Let Assumptions (A1) - (A3) be satisfied if there holds

$$\left\|\sum_{j=1}^{J}p^{m,j}\right\|^2 \geq \gamma\eta_m\sum_{j=1}^{J}\|p^{m,j}\|^2, m = 1, 2, \ldots. (59)$$

then

$$\left\|\sum_{j=1}^{J}p^{m+1,j}\right\|^2 \geq \gamma\eta_{m+1}\sum_{j=1}^{J}\|p^{m+1,j}\|^2, m = 1, 2, \ldots. (60)$$

Proof.

By the mean value theorem

$$p^{m+1,J} - p^{m,J}$$

$$= \left(g_j'\left(\prod_{i=1}^{N}(\omega_i^{mJ+j-1}.\xi^j)\right)\prod_{\substack{i=1 \\ i\neq k}}^{N}(\omega_i^{mJ+j-1}.\xi^j)\xi^j\right.$$

$$\left. - \; g_j'\left(\prod_{i=1}^{N}(\omega_i^{mJ}.\xi^j)\right)\prod_{\substack{i=1 \\ i\neq k}}^{N}(\omega_i^{mJ}.\xi^j)\,\xi^j\right)$$

$$+ \lambda\left(\frac{f'(\omega_k^{mJ+j-1})}{2J\,f(\omega_k^{mJ+j-1})^{1/2}} - \frac{f'(\omega_k^{mJ})}{2J\,f(\omega_k^{mJ})^{1/2}}\right)$$

$$= g_j''(t_{j,m})\prod_{i=1}^{N}(\omega_i^{mJ+j-1}.\xi^j)\left(\prod_{i=1}^{N}(\omega_i^{mJ+j-1}.\xi^j)\right.$$

$$\left. - \prod_{i=1}^{N}(\omega_i^{mJ}.\xi^j)\right)\xi^j$$

$$+ \frac{\lambda}{2J}F''(t_{i,k,m,j})(\omega_k^{mJ+j-1} - \omega_k^{mJ}) \quad (61)$$

where $t_{j,m} \in \mathbb{R}$ lies on the segment between $\omega_i^{mJ}.\xi^j$ and $\omega_i^{(m+1)J}.\xi^j$

Applying the triangle inequality to equ. (61) and using equ. (36) and equ. (44), we have

$$\|p^{(m+1)J}\| \leq \|p^{mJ}\|$$

$$+ C_5C^{N+1}\eta_m\left(\sum_{j=1}^{J}2\|p^{m,i}\| + \sum_{j=1}^{J}\|r_k^{i,m}\|\right)$$

$$+ \lambda M\,(1 + C_3)\,\eta_m\sum_{j=1}^{J}\|P^{m\,j}\|$$

$$\leq \|p^{mJ}\| + C_5C^{N+1}\eta_m\left(\sum_{j=1}^{J}2\|p^{m,i}\| + C_2\eta_m\sum_{j=1}^{J}\|p^{m,j}\|\right)$$

$$+ \lambda M\,(1 + C_3)\,\eta_m\sum_{j=1}^{J}\|P^{m\,j}\|$$

$$\leq \|p^{mJ}\| + [C_5 C^{N+1}(2 + C_2\eta_m) + \lambda M\,(1 + C_3)]\,\eta_m \sum_{j=1}^{J} \|P^{m\,j}\|$$

$$\leq \|p^{mJ}\| + C_{13}\,\eta_m \sum_{j=1}^{J} \|P^{m\,j}\| \quad (62)$$

where $C_{13} = C_5 C^{N+1}(2 + C_2\eta_m) + \lambda M\,(1 + C_3)$.
Thus

$$\sum_{j=1}^{J} \|p^{m+1,j}\|^2 \leq \sum_{j=1}^{J} \|p^{m,j}\|^2 + 2C_{13}\,\eta_m \sum_{j=1}^{J} \|p^{mJ}\|^2 + JC_{13}^2\eta_m^2 \sum_{j=1}^{J} \|p^{mJ}\|^2$$

$$\leq [1 + C_{14}\eta_m(1 + \eta_m)] \sum_{j=1}^{J} \|p^{m,j}\|^2 \quad (63)$$

where $C_{14} = max\{2JC_{13}J^2C_{13}^2\}$. Further, this together with equ. (61) yields

$$\frac{1}{\gamma\eta_m}\left\|\sum_{j=1}^{J} p^{m,j}\right\|^2 \leq \sum_{j=1}^{J} \|p^{mJ}\|^2 \leq \frac{1}{1 + C_{14}\eta_m(1 + \eta_m)} \sum_{j=1}^{J} \|p^{m+1,j}\|^2$$

$$\left\|\sum_{j=1}^{J} p^{m,j}\right\|^2 \geq \gamma\eta_m \sum_{j=1}^{J} \|p^{mJ}\|^2 \leq \frac{\gamma\eta_m}{1 + C_{14}\eta_m(1 + \eta_m)} \sum_{j=1}^{J} \|p^{m+1,j}\|^2 \quad (64)$$

On the other hand, it follows from equ. (61) and $0 \leq \eta_m \leq \eta_0 \leq 1$ that

$$\left(\sum_{j=1}^{J} \|p^{mJ}\|\right)^2 \leq J \sum_{j=1}^{J} \|p^{mJ}\|^2 \leq \frac{1}{\gamma\eta_m}\left\|\sum_{j=1}^{J} p^{m,j}\right\|^2 \quad (65)$$

and

$$\eta_m \sum_{j=1}^{J} \|p^{mJ}\| \leq \sqrt{\frac{J\eta_m}{\gamma}}\left\|\sum_{j=1}^{J} p^{m,j}\right\| \leq \sqrt{\frac{J}{\gamma}}\left\|\sum_{j=1}^{J} p^{m,j}\right\| \quad (66)$$

A combination of equ. (31), $0 \leq \eta_m \leq 1$ and equ. (66) leads to

$$\|v_k^{m,j}\| \leq \eta_m \left\|\sum_{j=1}^{J} p^{m,j}\right\| + C_2\eta_m^2 \sum_{j=1}^{J} \|p^{m,j}\| \leq \left(1 + C_2\sqrt{\frac{J}{\gamma}}\right)\eta_m \left\|\sum_{j=1}^{J} p^{m,j}\right\| \quad (67)$$

For equ. (59), we obtain by the triangle inequality, equ. (36) and equ. (67), that

$$\sum_{j=1}^{J} p^{m+1,j} = \sum_{j=1}^{J} p^{m,j} + JC_5 C^{N+1}\eta_m(2 + C_2\eta_m) \sum_{j=1}^{J} \|P^{m\,j}\| + \lambda M(v_k^{m,j})$$

$$\left\|\sum_{j=1}^{J} p^{m+1,j}\right\| \geq \left\|\sum_{j=1}^{J} p^{m,j}\right\| - \lambda M\|v_k^{m,j}\| \geq \left(1 - \lambda M\left(1 + C_2\sqrt{\frac{J}{\gamma}}\right)\eta_m\right)\left\|\sum_{j=1}^{J} p^{m,j}\right\| \quad (68)$$

It can be easily verified that for any positive $x \geq y - z$, then $x^2 \geq y^2 - 2yz$ (69)
Applying equ. (69) to equ. (68) and noting equ. (64), there holds

$$\left\|\sum_{j=1}^{J} p^{m+1,j}\right\|^2 \geq (1 + C_{15}\eta_m)\left\|\sum_{j=1}^{J} p^{m,j}\right\|^2 \geq \frac{\gamma\eta_m(1 + C_9\eta_m)}{1 + C_{14}\eta_m} \sum_{j=1}^{J} \|p^{m+1,j}\|^2 \quad (70)$$

Where $C_{15} = \lambda M\left(1 + C_2\sqrt{J/\gamma}\right)$.
Obviously, if

$$\frac{\gamma\eta_m(1 + C_9\eta_m)}{1 + C_{14}\eta_m(1 + \eta_m)} \geq \gamma\eta_{m+1} \quad (71)$$

From this easy to get equ. (60) is proved. Hence we need only to verify equ. (71) under the assumptions presented in ***Lemma 9***.
Then we substituting equ. (21) into equ.(71), we get

$$\beta - C_8 - C_9 \geq (C_8 + \beta C_9)\eta_m \quad (72)$$

Furthermore, if η_0 and β in equ. (21) satisfy the conditions in Assumption (A2), there holds

$$0 \leq \eta_m \leq \eta_0 \leq \frac{1}{\beta_0} - \frac{1}{\beta} = \frac{\beta - \beta_0}{\beta\beta_0} \quad (73)$$

That is equ. (70).
Since equ. (72) and equ. (71) are identical, the inequality equ. (71) is thus proved. Hence, the inequality equ. (60) also has been proved.
The next two Lemmas will be used to prove our convergence results. Their proofs are omitted since they are quite similar to those of Lemma 3.5 in [41] and theorem 3.5.10 in [43], respectively

Lemma 10.
Suppose that the series $\sum_{n=1}^{\infty} a_n^2/n < \infty$, *that* $a_n > 0$ *for* $n = 1, 2, \ldots$.. and that there exists a constant $\mu > 0$ such that $|a_{n+1} - a_n| < \mu/n$, $n = 1, 2, \ldots$ then, we have $\lim_{n\to\infty} a_n = 0$.

Lemma 11.
Let $F: \Phi \subset R^p \to R\,(p \geq 1)$ be continuous for a bounded closed region Φ. if the set $\Phi_0 = \{x \in \Phi: F_x(x) = 0\}$ has finite points and the sequence $\{x_n\} \in \Phi$ satisfy:
(1) $\lim_{n\to\infty} \|F_x(x_n)\| = 0$,
(2) $\lim_{n\to\infty} \|x_{n-1} - x_n\| = 0$.
Then, there exists $x^* \in \Phi_0$ such that $\lim_{n\to\infty} x_n = x^*$
Now we are ready to prove the main theorems.

Proof of Theorem 1.

In virtue of equ. (48), if for any nonnegative integer m

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Combining this with equ. (98) yields

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

Hence

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

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, "$L_{1/2}$ 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.