Mathematics and Statistics 11(1): 191-198, 2023
DOI: 10.13189/ms.2023.110122 http://www.hrpub.org

A New Quasi-Newton Method with PCG Method for Nonlinear Optimization Problems

Bayda Ghanim Fathi, Alaa Luqman Ibrahim*

Department of Mathematics, Faculty of Science, University of Zakho, Iraq

Received October 3, 2022; Revised December 19, 2022; Accepted January 16, 2023

Cite This Paper in the Following Citation Styles

(a): *[1] Bayda Ghanim Fathi, Alaa Luqman Ibrahim , " A New Quasi-Newton Method with PCG Method for Nonlinear Optimization Problems," Mathematics and Statistics, Vol. 11, No. 1, pp. 191 - 198, 2023. DOI: 10.13189/ms.2023.110122.*

(b): *Bayda Ghanim Fathi, Alaa Luqman Ibrahim (2023). A New Quasi-Newton Method with PCG Method for Nonlinear Optimization Problems. Mathematics and Statistics, 11(1), 191 - 198. DOI: 10.13189/ms.2023.110122.*

Abstract The major stationary iterative method used to solve nonlinear optimization problems is the quasi-Newton (QN) method. Symmetric Rank-One (SR1) is a method in the quasi-Newton family. This algorithm converges towards the true Hessian fast and has computational advantages for sparse or partially separable problems [1]. Thus, investigating the efficiency of the SR1 algorithm is significant. It's possible that the matrix generated by SR1 update won't always be positive. The denominator may also vanish or become zero. To overcome the drawbacks of the SR1 method, resulting in better performance than the standard SR1 method, in this work, we derive a new vector y_k^* depending on the Barzilai-Borwein step size to obtain a new SR1 method. Then using this updating formula with preconditioning conjugate gradient (PCG) method is presented. With the aid of inexact line search procedure by strong Wolfe conditions, the new SR1 method is proposed and its performance is evaluated in comparison to the conventional SR1 method. It is proven that the updated matrix of the new SR1 method, H_{k+1}^{new}, is symmetric matrix and positive definite matrix, given H_k is initialized to identity matrix. In this study, the proposed method solved 13 problems effectively in terms of the number of iterations (NI) and the number of function evaluations (NF). Regarding NF, the new SR1 method also outperformed the classic SR1 method. The proposed method is shown to be more efficient in solving relatively large-scale problems (5,000 variables) compared to the original method. From the numerical results, the proposed method turned out to be significantly faster, effective and suitable for solving large dimension nonlinear equations

Keywords Unconstrained Optimization, Preconditioned Conjugate Gradient Methods, Quasi-Newton Methods, Variable Metric Methods, Line Search Method, Strong Wolfe Conditions

1. Introduction

We deal with a class of methods for local unconstrained minimization problems, i.e., for finding a local minimum point $x^* \in \mathbb{R}^n$ such that

$$f(x^*) = \min_{x\in\mathbb{R}^n} f(x), \tag{1.1}$$

where $f: \mathbb{R}^n \to \mathbb{R}$ is a twice continuously differentiable objective function and x is an n-dimensional vector space. Only the first derivative of the objective function f is employed to solve the non-linear problems of unconstrained optimization using the conjugate gradient (CG) method and (QN) method. They are frequently utilized in applications when just the first derivative is known or higher derivatives are difficult to calculate. The behavior of a given approach may be of more relevance in some applications than the optimal solution itself. The methods are applicable to a wide range of fields; some examples from various fields are shown below.

In [2], a developed update QN methods is utilizing the unique structure of the problem to address distributed parameter estimation problems in the field of electromagnetics. Another use for these methods is optical tomography [3]. These two techniques can also be utilized in neural network programs; for examples, see [4–5]. The approaches can also be applied to radiation therapy optimization [6].

Preconditioning is a crucial way for creating an effective conjugate gradient solver for difficult problems in scientific computing [7]. That means, the (PCG) is the most effective method for solving large and complicated problems in unconstrained optimization.

In this study, the PCG methods discussed depend on both conjugate gradient and quasi-Newton methods, so a review of their properties is given. Emphasis falls on their application to quadratic functions, since this is the basis for development of algorithms in the following section.

Given that the preconditioned conjugate gradient methods addressed in this study are based on both the conjugate gradient and quasi-Newton methods, a review of their characteristics is provided. As the foundation for developing algorithms in the following section, the emphasis is on their application to quadratic functions.

1.1. The Conjugate Gradient Method (CG)

The conjugate-gradient approach will be discussed in this section. It is a low storage algorithm that was first described in a publication by Hestenes and Stiefel [8]. When n is large, the CG approach can be used to track the optimal solution x^* of an unconstrained optimization problem (1.1) [9]. This method was inspired by the issue of a quadratic function

$$f(x) = \frac{1}{2}x^T Gx + b^T x, \tag{1.2}$$

where G is the Hessian matrix that is symmetric matrix and positive definite matrix of size $n \times n$ and b is a vector of length n. The search directions of this method are generated without the need to store the matrix G [10]. Note that the gradients of the CG method are mutually orthogonal, $g_{k+1}^T g_j = 0$ for all $j = 1, \dots, k$, and $k = 1, \dots$ [11]. The first step we define the direction d_1 is performed along the steepest descent direction $-g_1$. There after search direction d_{k+1} is a linear combination of the steepest descent $-g_{k+1}$ and the previous search direction d_k scaled with a parameter β_k, i.e.

$$d_{k+1} = -g_{k+1} + \beta_k d_k, \tag{1.3}$$

where β_k is a scalar determined with the requirement that d_k and d_{k+1} must be conjugate with respect to the G Hessian matrix for problem (1.2) [11]. Below are the three most used formulas for β_k:

$$\beta_k^{HS} = \frac{g_{k+1}^T y_k}{d_k^T y_k}, \quad \text{(Hestenes-Stiefel, 1952)}. \tag{1.4a}$$

$$\beta_k^{FR} = \frac{\|g_{k+1}\|^2}{\|g_k\|^2}, \quad \text{(Fletcher-Reeves, 1964)} \tag{1.4b}$$

$$\beta_k^{PR} = \frac{g_{k+1}^T y_k}{\|g_k\|^2}. \quad \text{(Polack-Ribiere, 1969)} \tag{1.4c}$$

There are constructed examples where β_k^{PR} and β_k^{HS} formulations struggle greatly. However, computational tests indicate that PR and HS both outperform FR [9].

The Wolfe conditions are frequently used as the foundation for the termination conditions for the CG line search. The standard Wolfe conditions are

$$f(x_k + \alpha_k d_k) \le f(x_k) + \rho \alpha_k g_k^T d_k, \tag{1.5a}$$

$$g(x_k + \alpha_k d_k)^T d_k \ge \sigma g_k^T d_k, \tag{1.5b}$$

where d_k is a descent direction and the constants are within the period $0 < \rho < \sigma < 1$ [12]. For some CG algorithms, the strong versions of the Wolfe conditions, (1.5a), and

$$|g(x_k + \alpha_k d_k)^T d_k| \le \sigma g_k^T d_k, \tag{1.5c}$$

with $\sigma < 1$, are required to guarantee convergence and enhance stability, for more details see [13].

1.2. The Quasi-newton Methods

The QN methods, also known as variable metric methods, provide an important family of widely applicable methods for solving smooth unconstrained problems. To minimize a function f in (1.1) based on the real vector $x$of n-dimensional, these algorithms defined

$$x_{k+1} = x_k - \alpha_k H_k g_k, \tag{1.6}$$

where x_k is the kth approximation to the minimum point, g_k is the gradient of f at x_k, H_k is an n ×n matrix that approximates the inverse Hessian of f at x_k and α_k is a positive step size parameter whose value is selected according to a rule which depends on the specific method. The approximation H_k is based on information about the inverse Hessian that is deduced from observations of previous gradients. This approximation is usually updated after each repetition.

Davidon [14] proposed the first quasi-Newton approach of this type. This technique, known as the SR1 algorithm, uses a symmetric H_k and a line search to choose α_k in order to minimize f along the line $x_k - \alpha_k H_k g_k$. The required QN algorithm has three key characteristics: 1- each H_k matrix is positive definite; 2- the directions of search are identical to those of the CG method for quadratic problems if $H_k = I$ (Fletcher and Reeves [15]); and 3- once more for a quadratic problem, the kth approximation H_k is identically equal to the inverse Hessian. It is believed that these three characteristics underpin the successful convergence characteristics the approach frequently exhibits.

Many contributions are taken into consideration when deriving new updating formulae possessing some or all of the three properties of the algorithm mentioned above, see [16-20].

At each iteration, the new approximation H_{k+1} is selected to take account for the new curvature information which is done by satisfying the quasi-Newton condition ($H_{k+1} y_k = v_k$). The quasi-Newton condition can be satisfied by an endless number of rank-two updates. The modifications to the Broyden (1970) one-parameter class are now being taken into consideration. The matrix H_{k+1} is defined by

$$H_{k+1} = H_k - \frac{H_k y_k y_k{}^T H_k}{y_k{}^T H_k y_k} + \frac{v_k v_k{}^T}{v_k{}^T y_k} + \theta_k v_k v_k{}^T, \tag{1.7a}$$

where

$$v_k = ({y_k}^T H_k y_k)^{\frac{1}{2}} \left(\frac{v_k}{{v_k}^T y_k} - \frac{H_k y_k}{{y_k}^T H_k y_k}\right). \quad (1.7b)$$

Different choices of the scalar parameter θ_k define different updates. This family is known as the Broyden family [21] or as the one-parameter family of updates [22]. It is also referred to as the Broyden family in [23] and [24]. It is easy to verify that any update from this class satisfies the quasi-Newton condition. Moreover, if H_k is symmetric (like the actual inverse Hessian matrix), H_{k+1} will also be symmetric; this property is called hereditary symmetry. The (SR1) update method belongs to the Broyden family, provided by the formula

$$H_{k+1} = H_k + U_k,$$

where $U_k = \frac{(v_k - H_k y_k)(v_k - H_k y_k)^T}{{y_k}^T (v_k - H_k y_k)}$ and it is called the correction matrix.

It was originally discovered by Davidon [25], according to [26]. One fact about SR1 is that even if H_k is positive definite matrix, the update matrix H_{k+1} may not have this property. The possibility of steps when no update satisfies the secant condition is a serious drawback. As long as the denominator is different from zero the method proceeds with a unique rank-one update. If $H_k y_k = v_k$ the only update that satisfies the secant condition is $U_k = 0$, such that the same H_k matrix can be used for another iteration. The failure occurs when $H_k y_k = v_k$ and ${y_k}^T (v_k - H_k y_k) = 0$ at the same iteration.

1.3. Preconditioned Conjugate Gradient (PCG) Method

The PCG methods are modifications of the CG method which were first developed by Axelsson (1974) and used for solving systems of linear equations. The idea of preconditioning has been extended directly to nonlinear problems [12].

Conjugate gradient type methods are only practical methods for problems where the number of variables is large. This is because storage for only a few n-vectors is required to solve an n-variable problem. The conventional CG approach does not always work, however preconditioning with the use of a suitable matrix can speed up convergence by transforming the variables while preserving the fundamental characteristics of the method. The preconditioning matrices are based on the inverse Hessian approximations generated in a quasi-Newton method.

In PCG approaches, a matrix transformation is applied to the newly created conjugate gradient direction. Therefore, the main idea behind this method is the use of the quasi-Newton updates to accelerate the CG methods.

Applying the CG technique with a variable preconditioner that is updated using a quasi-Newton formula is the way for accelerating CG methods with quasi-Newton matrices. Nazareth proposed this procedure, which is known as the variable metric conjugate gradient method (1979).

Both the QN and CG approaches have advantages and disadvantages when applied to generic functions. In 1972, McCormick and Ritter demonstrated that, in general, the QN methods converge faster than the CG methods (and necessitate fewer function evaluations) [28].

The high-speed computer storage must contain n(n+1)/2 places because QN methods often create a series of symmetric positive-definite matrices of order n. However, QN methods often create a series of symmetric positive-definite matrices of order n, therefore n(n+1)/2 locations in the high-speed computer storage are required. As n grows, n(n+1)/2 must grow to be too large, making the variable metric approaches ineffective (only with virtual memory). Therefore, the precondition conjugate gradient method (PCG), a new class of CG methods, has been devised. Preconditioning is the process of transforming a problem so that its Hessian has clustered eigenvalues and is well-conditioned. The PCG method's goal is to maintain order n storage needs while enhancing convergence qualities; for more information, see [29] and [30].

1.3.1. Algorithm of the Preconditioned Conjugate Gradient (PCG)

Step 1: Select $x_1 \epsilon \mathbb{R}^n$. Set $k = 1$ and $d_1 = -H_1 g_1$ where $H_1 = I$ and g_1 is the gradient of f.

Step 2: Compute the step length α_k by using a line search procedure which guarantees the Wolfe conditions to be satisfied, and set

$$x_{k+1} = x_k + \alpha_k d_k.$$

Step 3: If $g_1 = 0$ then stop, else compute H_{k+1} and copute the new direction d_{k+1} as

$$d_{k+1} = -H_{k+1} g_{k+1} + \beta_k d_k, \quad (1.8)$$

set k = k + 1 and go to Step 2.

It can be noticed that by setting H_k = I for any iteration k, the popular (unpreconditioned) Nonlinear Conjugate Gradient (NCG) method is obtained. The parameter β_k can be chosen in a variety of ways. For PNCG algorithm, the most recurrent choice is the following:

$$\beta_k^{HS} = \frac{y_k^T H_{k+1} g_{k+1}}{d_k^T y_k}.$$

We recall that with respect to other gradient methods, a more accurate line search procedure is required to determine the step length α_k in a PCG algorithm. This is due to the presence of the term $\beta_k d_k$ in (1.8). The latter fact motivates the use of the strong Wolfe conditions to compute the step length α_k, which also guarantees that $v_k^T y_k > 0$ for any k.

The remainder of this paper is as follows: in Section 2, the theoretical characteristics of the SR1 preconditioner sequences and a matrix preconditioner update created are discussed. A new SR1 quasi newton method is driven in Section 3, together with the theoretical characteristics of the preconditioned matrices, and a matrix version of the preconditioner update is developed by using the proposed

SR1. Several numerical findings that support the computational benefit offered by the good performance of the new SR1 update to speed up the PCG technique are shown in Section 4. The numerical findings are then displayed in Section 5.

2. Derivation of the Modified SR1 Method

In this section, we will concentrate on obtaining the modified SR1 algorithm for the preconditioned conjugate gradient method. A new vector is suggested as follows:

Let $y_k^* = \frac{1}{\alpha_k^{BB}} y_k - \frac{\theta}{\alpha_k^{BB}} y_k$, where $\theta \in (0,1)$ an $\alpha_k^{BB} = \frac{v_k^T v_k}{y_k^T v_k}$ is the Barzilai-Borwein step size; see [31] for more information. That means

$$y_k^* = (1-\theta)\frac{y_k^T v_k}{v_k^T v_k} y_k, \tag{2.1}$$

Let's now assume that the quasi-Newton equation is represented by the following equation:

$$H_{k+1}^{New}\, y_k^* = v_k. \tag{2.2}$$

In the SR1 formula, the correction term is symmetric and has the form $\alpha_k L_k {L_k}^T$ where $\alpha_k \in R$ and $L_k \in R^n$. Therefore, the update equation is

$$H_{k+1}^{New} = H_k + \alpha_k L_k {L_k}^T, \tag{2.3}$$

where $H_{k+1}^{New} \in R^{n\times n}$.

By multiplying both sides of the above equation by y_k^* from the wright and using (2.2), we get

$$H_{k+1}^{New} y_k^* = H_k y_k^* + \alpha_k L_k {L_k}^T y_k^* = v_k.$$

Since ${L_k}^T y_k^*$ is a scalar, we obtain

$$v_k - H_k y_k^* = (\alpha_k {L_k}^T y_k^*) L_k, \tag{2.4}$$

which implies that $L_k = \frac{v_k - H_k y_k^*}{\alpha_k {L_k}^T y_k^*}$. That means

$$\alpha_k L_k {L_k}^T = \frac{(v_k - H_k y_k^*)(v_k - H_k y_k^*)^T}{\alpha_k ({L_k}^T y_k^*)^2}. \tag{2.5}$$

Multiplying (2.4) by ${y_k^*}^T$, we achieve

$${y_k^*}^T (v_k - H_k y_k^*) = (\alpha_k {L_k}^T y_k^*) {y_k^*}^T L_k.$$

Since α_k is a scalar. Also by using the property of inner product, we have

$${y_k^*}^T (v_k - H_k y_k^*) = \alpha_k \left({L_k}^T y_k^*\right)^2. \tag{2.6}$$

Substituting (2.6) in (2.5), we get

$$\alpha_k L_k {L_k}^T = \frac{(v_k - H_k y_k^*)(v_k - H_k y_k^*)^T}{{y_k^*}^T (v_k - H_k y_k^*)}. \tag{2.7}$$

Therefore

$$H_{k+1}^{New} = H_k + \frac{(v_k - H_k y_k^*)(v_k - H_k y_k^*)^T}{{y_k^*}^T (v_k - H_k y_k^*)}. \tag{2.8}$$

This is the new SR1 update matrix. Below is the modified SR1 algorithm using the PCG-method.

2.1. Algorithm of the Modified SR1 with the PCG-Method

Step 1: Set $k = 0$, select an initial point x_0, a real symmetric positive definite $H_0 = I$ and

$$\varepsilon = 1 \times 10^{-5}.$$

Step 2: Calculate the gradient g_k of a function $f(x_k)$.

Step 3: Compute $d_k = -H_k g_k$.

Step 4: Determine $\alpha_k > 0$ by line search which satisfies the strong Wolfe condition.

Step 5: Set $v_k = \alpha_k d_k$ and $x_{k+1} = x_k + v_k$.

Step 6: Compute $g_{k+1,}$ if $\|g_{k+1}\| < \varepsilon$, then stop.

Else

Go to Step (7).

Step 7: Calculate y_k^* as given in (2.1) and update H_{k+1}^{New} using (2.8).

Step 8: Compute the direction $d_{k+1} = -H_{k+1}^{New} g_{k+1} + \frac{{y_k^*}^T H_{k+1}^{New} g_{k+1}}{d_k^T y_k^*} d_k$.

Step 9: If $|g_k^T g_{k+1}| \geq 0.2\|g_{k+1}\|^2$ go to step (3).

Else

Set $k = k + 1$ and repeat from Step (4).

Theorem 1: If the new SR1 method in (2.8) is applied to the quadratic with Hessian $G = G^T$, then

$$H_{k+1}^{new} y_k^* = v_k, k \geq 0.$$

Proof: Multiplying both sides of equation (2.8) by y_k^* from the right, we have:

$$H_{k+1}^{new} y_k^* = H_k y_k^* + \frac{(v_k - H_k y_k^*)(v_k - H_k y_k^*)^T y_k^*}{{y_k^*}^T (v_k - H_k y_k^*)}.$$

Since $(v_k - H_k y_k^*)^T y_k^*$ and also ${y_k^*}^T (v_k - H_k y_k^*)$ are scalars.

So, we have $H_{k+1}^{new} y_k^* = H_k y_k^* + v_k - H_k y_k^*$.

Then, $H_{k+1}^{new} y_k^* = v_k$. Hence, the proof is complete.

Theorem 2: If H_k^{New} is a positive definite matrix, then theH_{k+1}^{New} which is generated by equation (2.8) is also a positive definite matrix.

Proof: Multiplying both sides of (2.8) by y_k^* from the right and by ${y_k^*}^T$ from the left, we get

$${y_k^*}^T H_{k+1}^{new} y_k^* = {y_k^*}^T H_k y_k^* + \frac{{y_k^*}^T (v_k - H_k y_k^*)(v_k - H_k y_k^*)^T y_k^*}{{y_k^*}^T (v_k - H_k y_k^*)}.$$

After simplifying using algebraic operations, we get

$${y_k^*}^T H_{k+1}^{new} y_k^* = {y_k^*}^T v_k.$$

Or

$${y_k^*}^T H_{k+1}^{new} y_k^* = (1-\theta)\frac{(y_k^T v_k)^2}{v_k^T v_k}.$$

Because $\theta \in (0,1)$, it is clearly that the right-hand side of the above equation is a positive number. This completes the proof.

3. Numerical Results

The purpose of this section is to evaluate how well the updated SR1 method has been implemented. On a concatenation of test problems for unconstrained nonlinear optimization derived from [32], several computational experiments are conducted. All programs are written in the FORTRAN95 language to demonstrate the usage and effectiveness of the suggested approach with various dimensions of $4 \leq n \leq 5000$.

Table 1 compares the performance of the new SR1 approach and the original SR1 method using NI and NF for 13 distinct non-linear functions. The modified SR1 technique with the precondition CG method's rate of improvement is shown in Table 2, which demonstrates its superiority to the standard approach in terms of the NI and NF.

Table 1. Comparison between the standard SR1 method and the new SR1 method.

Test Function	n	$Standard\ SR1$		New SR1	
		NI	NF	NI	NF
Miele (1, 2, 2,…)	4	34	329	21	298
	10	34	329	21	298
	50	41	182831	28	328
	100	47	182999	28	328
	500	53	183098	34	359
	1000	53	183098	34	359
	5000	65	189124	47	424
G-Cantrel (1,2, 2,…)	4	36	253	17	80
	10	36	235	17	80
	50	43	331	19	96
	100	43	331	22	132
	500	60	496	31	228
	1000	60	554	35	272
	5000	72	616	52	464
Cubic (-1.2, 1, …)	4	15	48	12	35
	10	15	48	13	37
	50	14	48	13	37
	100	16	61	13	37
	500	17	56	13	37
	1000	16	50	13	37
	5000	16	178	14	39
Beal (0, 0, …)	4	11	29	11	27
	10	11	29	11	27
	50	12	31	12	29
	100	12	31	12	29
	500	12	31	12	29
	1000	12	31	12	29
	5000	12	31	12	29
Powell (3, -1, 0, 1, …)	4	30	80	29	82
	10	30	93	30	84
	50	31	95	30	84
	100	32	97	30	84
	500	33	97	30	84
	1000	33	97	30	84
	5000	33	97	31	86
Powell3 (0, 1, 2, …)	4	14	35	14	32
	10	14	35	15	34
	50	15	37	15	34
	100	15	37	15	34
	500	15	37	15	34
	1000	15	37	15	34
	5000	16	40	15	34

Table 1 Continued

G-Wolfe (-1, …)	4	11	24	11	24
	10	25	51	25	51
	50	42	85	42	85
	100	44	89	44	89
	500	47	95	47	95
	1000	50	101	49	99
	5000	106	294	105	241
G-Edger (1, 0, …)	4	5	14	5	14
	10	5	14	5	14
	50	5	14	5	14
	100	5	14	5	14
	500	5	14	5	14
	1000	6	16	6	16
	5000	6	16	6	16
G-Wood (-3, -1, -3, -1, …)	4	20	50	19	49
	10	22	54	19	49
	50	23	57	19	49
	100	23	57	19	49
	500	23	57	19	49
	1000	23	57	21	53
	5000	23	57	21	53
TRI (1, . . .)	4	3	7	3	7
	10	9	19	9	19
	50	44	89	44	89
	100	71	142	71	142
	500	188	377	188	377
	1000	277	555	277	555
	5000	660	1321	660	1321
Rosen (-1.2, 1, …)	4	30	89	30	84
	10	31	95	30	84
	50	31	89	30	84
	100	31	89	30	84
	500	32	93	30	84
	1000	34	97	30	84
	5000	37	10	30	84
Recip (2, 5, 1, …)	4	5	18	5	16
	10	5	18	5	16
	50	5	18	5	16
	100	5	18	5	16
	500	5	18	5	16
	1000	5	18	5	16
	5000	6	20	6	18
Fred (30,3, …)	4	8	24	8	23
	10	8	24	8	23
	50	8	24	8	23
	100	8	24	8	23
	500	8	24	8	23
	1000	8	24	8	23
	5000	8	24	8	23
Total		**3273**	**930838**	**2944**	**9538**

Table 2. Relative efficiency of the between algorithms

Tools	$Standard\ SR1$	New SR1
NI	100%	89.948 %
NF	100%	1.0247 %

This table demonstrates that the proposed strategy improves NI by 10.052% and NF by 98.9753%. In comparison to the normal SR1 approach, the modified SR1 method has generally improved by 54.51365%.

4. Conclusions

Since they have high theoretical and practical convergence features, quasi-Newton or variable metric methods are well known and frequently employed in combination with unconstrained optimization. Fast convergence, algorithmic simplicity, stability, low memory requirements, and a reliable solution estimate are requirements for a good approach. All of these conditions are met by the preconditioned conjugate gradient method, making it an efficient iterative technique.

A modification of the SR1 algorithm to drive a new preconditioning matrix is suggested. The proof of its positive definiteness is achieved. Moreover, it has been showed that the matrix of New SR1, H_{k+1}^{new}, satisfied the QN-condition (or the secant equation). Numerical results showed that our suggested method gives an effective numerical result in practically. It is shown from the tables that the performance of the SR1 quasi-Newton method can be significantly improved using preconditioned conjugate gradient (PCG) method with the modified SR1 method.

REFERENCES

[1] Byrd, R.H., Khalfan, H.F., Schnabel, R.B., Analysis of a symmetric rank-one trust region method, SIAM J. Optim, 6, 1025–1039, 1996.

[2] E. Haber. Quasi-Newton methods for large-scale electromagnetic inverse problems. Inverse Problems, 21(1):305–333, 2005.

[3] A. D. Klose and A. H. Hielscher. Quasi-Newton methods in optical tomographic image reconstruction. Inverse Problems, 19(2):387–409, 2003.

[4] J. W. Denton and M. S. Hung. A comparison of nonlinear optimization methods for supervised learning in multilayer feedforward neural networks. European Journal of Operational Research, 93:358–368, 1996.

[5] A. L. Ibrahim and M. G. Mohammed, "A new three-term conjugate gradient method for training neural networks with global convergence" Indonesian Journal of Electrical Engineering and Computer Science, Vol. 28, No. 1, October 2022, pp. 547~554. DOI: http://doi.org/10.11591/ijeecs.v28.i1.pp551-558

[6] F. Carlsson and A. Forsgren. Iterative regularization in intensity-modulated radiation therapy optimization. Medical Physics, 33(1):225–234, 2006.

[7] M. Benzi, Preconditioning Techniques for Large Linear Systems: A Survey, Journal of Computational Physics, 182, pp. 418-477, 2002.

[8] M. R. Hestenes and E. Stiefel. Methods of conjugate gradients for solving linear systems. J. Research Nat. Bur. Standards, 49:409–436, 1953.

[9] S. G. Nash and A. Sofer. Linear and Nonlinear Programming. McGraw- Hill, New York, 1996.

[10] D. G. Luenberger. Linear and Nonlinear Programming. Second ed., Addison-Wesley, New York, 1989.

[11] J. Nocedal. Numerical Optimization. Springer, New York, 1999.

[12] P. Wolfe, Convergence conditions for ascent methods. II: Some corrections, SIAM Review, 13, pp. 185-188, 1971.

[13] Y. H. Dai, Analyses of conjugate gradient methods, Ph.D. thesis, Institute of Computational Mathematics and Scientific/Engineering Computing, Chinese Academy of Sciences, 1997.

[14] W. C. DAVIDON, "Variable Metric Method for Minimization," A.E.C. Research and De- velopment Rep. ANL-5990 (Rev.), 1959.

[15] R. I. Fletcher, and C. M. Reeves, "Function Minimization by Conjugate Gradients," Comp. J., Vol. 7, pp. 149-154, 1964.

[16] C. G. BROYDEN, "Quasi-Newton Methods and Their Application to Function Minimization," Maths. of Comp., Vol. 21, pp. 368-381, 1967.

[17] C. G. BROYDEN "The Convergence of a Class of Double-Rank Minimization Algorithms 1. General Considerations," J.I.M.A., Vol. 6, pp. 76-90, 1970.

[18] D. GOLDFARB, "A Family of Variable Metric Methods Derived by Variational Means," M1laths. of Comp., Vol. 24, pp. 23-26, 1970.

[19] O. Dasril, G. Kh. Wen, N. B. Bujang and Sh. N Salahudin. "New approach on global optimization problems based on meta-heuristic algorithm and quasi-Newton method ", Vol. 12, No. 5, October 2022, pp. 5182~5190. DOI: 10.11591/ijece.v12i5.pp5182-5190

[20] Shareef, S. G. Ibrahim, A. L. & Yaseen, Z. T, "A new self-scaling variable metric (DFP) method for unconstrained optimization problems", General Letters in Mathematics, 9 (1), 46- 52, 2020.

[21] P. E. Gill, W. Murray, and M. H. Wright. Practical Optimization. Academic Press, London, 2003.

[22] D. G. Luenberger. Linear and Nonlinear Programming. Second ed., Addison-Wesley, New York, 1989.

[23] R. Fletcher. Practical Methods of Optimization. Second ed., John Wiley & Sons, Chichester,1987.

[24] W. C. Davidon. Variable metric methods for minimization. Argonne National Lab Report (Argonne, IL), 1959.

[25] J. Nocedal. Theory of algorithms for unconstrained

optimization. In Acta numerica, Acta Numer., pages 199–242. Cambridge Univ. Press, Cambridge, 1992.

[26] Axelsson, O., On Preconditioning and Convergence acceleration in sparse matrix problems, CERN Data Handling Division Report, 74-1, (1974).

[27] Nazareth, L.,"A Relationship Between the BFGS and Conjugate Gradient Algorithms and Its Implications for New Algorithms", SIAM J. Numer. Anal., 16, 1979.

[28] McCormick, G. and Ritter, K., Methods of Conjugate Directions Versus Quasi-Newton Methods", Math. Prog., 3, 1972.

[29] Buckley, A.G., "A Combined Conjugate Gradient Quasi-Newton Minimization Algorithm", Math. Prog., 15,1978a.

[30] Buckley, A. and Lenir, A.,"QN-Like Variable Storage Conjugate Gradients", Math. Prog., 27, 1983.

[31] Barzilai, J. and Borwein, J.M.," Tow point step size gradient methods", IMA J. Numer. Anal., 8, 141-148, 1988.

[32] N. Andrei, "An unconstrained optimization test functions collection," Adv. Model. Optim, vol. 10, no. 1, pp. 147–161, 2008.