International Journal of Mathematics and
Computer Science, **17**(2022), no. 3, 1061–1077

New Self-scaling Quasi-Newton methods for unconstrained optimization

Issam A. R. Moghrabi[1], Basim A. Hassan[2], Aadil Askar[3]

[1]Department of Accounting and M.I.S.
College of Business Administration
Gulf University for Science and Technology
West Meshref, Kuwait

[2]Department of Mathematics
The University of Mosul
Mosul, Iraq

[3]Department of Self Development Skills
King Saud University
Riyadh, Saudi Arabia

email: moughrabi.i@gust.edu.kw, basimah@uomosul.edu.iq,
aadil@ksu.edu.sa

(Received January 30, 2022, Accepted March 4, 2022)

Abstract

The quasi-Newton technique is one of the most well-known iterative solutions for unconstrained optimization problems. The quasi-Newton techniques are known for their high accuracy and rapid convergence. In this paper, we derive four self-scaling BFGS methods. The Wolfe line search criteria define the step size selection. The method of global convergence is demonstrated when the objective function is uniformly convex. Preliminary computational tests on a set of 30 unconstrained optimization test functions suggest that this technique is more efficient and resilient than implementing the non-scaled version of the BFGS. In terms of iteration count and function/gradient

Keywords and phrases: Unconstrained optimization, nonlinear optimization, Quasi-Newton methods, Self-scaling quasi-Newton Broyden family, machine learning.
AMS (MOS) Subject Classifications: 65K10.
ISSN 1814-0432, 2022, http://ijmcs.future-in-tech.net

evaluations, comparative findings reveal that the suggested strategy is computationally efficient.

1 Introduction

Consider the unconstrained nonlinear optimization problem

$$\min f(x), \text{ where } f : R_n \to R.$$

In a significant number of research articles, the quasi-Newton approaches have been shown to be successful in addressing the aforesaid problem. The Newton method, which is based on the second order Taylor series approximation, requires that the matrix of second order derivatives be computed for each iteration. It is usually better to estimate the Hessian matrix (or its inverse) using a symmetric positive definite matrix using an efficient technique rather than computing it explicitly. The central idea of this study is to approximate the Hessian with a symmetric positive definite matrix that is updated on an iteration basis. The gradient of the objective function $fatx_k$ is indicated as g_k for the present iteration k and the square positive-definite matrix B_k approximates $Q(x_k)$, the true Hessian of f. Traditionally, the Secant equation is given as:

$$B_{(k+1)}s_k = y_k, \tag{1.1}$$

for $y_k = g_{k+1} - g_k$ and $s_k = x_{k+1} - x_k$ must be satisfied by the new updated Hessian matrix approximation.

The Newton technique, which is based on the second order Taylor series approximation, entails computing the Hessian matrix of second order derivatives at each step, as we all know. In reality, it is frequently preferable to approximate the Hessian matrix (or its inverse) with a symmetric positive definite matrix using an efficient approach rather than computing it exactly. Davidon [7] was the first to propose using a symmetric positive definite matrix to approximate the Hessian. Quasi-Newton techniques are a class of methods that approach the Newton method by using a symmetric positive definite estimate of the Hessian or the inverse Hessian instead of the precise value. Let B_k represent a Hessian approximation at the kth iteration.

So, the update formula, which uses the information available at the kth iteration to create a matrix $B_{(k+1)}$ from the current matrix B_k, is a significant component in the implementation of the quasi-Newton approach since it affects the numerical behavior of the techniques. It is intended that further improvements would come as near to the true Hessian as feasible. There are several well-known updates of B_k that satisfy the traditional quasi-Newton equation in the literature, but very few have been able to compete numerically with the well-known Broyden Fletcher Goldfarb Shanno (BFGS) formula defined by:

$$B_{(k+1)} = B_k + \frac{y_k y_k^T}{s_k^T y_k} - \frac{B_k s_k s_k^t B_k}{s_k^T B_k s_k}. \tag{1.2}$$

Equation (1.2) satisfies relation (1.1).

Quasi-Newton methods proceed iteratively to determine a new solution approximation using

$$x_{k+1} = x_k + \alpha_k d_k, \tag{1.3}$$

where d_k represents the computed search direction vector at iteration k. The search direction is computed as the solution for

$$B_k d_k = -g_k \tag{1.4}$$

Alternatively, many implementations would rather use

$$B_k d_k = -g_k \; or \; d_k = -H_k g_k. \tag{1.5}$$

To avoid the cost of solving a system of linear equations, other implementations would update factors of the Hessian approximation B_k, instead.

The step length α_k in (1.4) is obtained by solving

$$\alpha_k = \frac{-g_k^T d_k}{d_k^T Q d_k}, \tag{1.6}$$

where, in practice, the solution is obtained by means of some line search methodology to satisfy conditions like the following Powell-Wolfe relations [26]

$$f(x_{k+1}) \leq f(x_k) + 10^{-5} s_k^T g_k \tag{1.7}$$

and

$$s_k^T g_{k+1} \geq 0.9 \; p_k^T g_k. \tag{1.8}$$

The BFGS approach is the most successful of all the quasi-Newton methods, according to the numerical experiment. However, the global convergence for general functions is still an open problem even if the method is established to be convergent (global and superlinear) for convex problems [9], [10], [17].

As a result, it is worth looking into the possibility of deriving a new quasi-Newton approach that not only has global convergence but also outperforms the BFGS method in terms of computational cost. The quest for a more efficient quasi-Newton approach than the BFGS method began in the 1970s and continues today. In reality, the majority of academics are interested in implementing novel methods derived from the original quasi-Newton equation. However, relatively few attempts have been made to revisit the original quasi-Newton equation [13],[15], [19], [29]. Dai et al. [6] provided an example to show how the traditional BFGS method might fail when dealing with non-convex functions that need inexact line search. Li and Fukushima [17] updated the conventional Secant equation (1.1) to create a modified BFGS approach that is globally convergent without requiring the goal function f is convex. The promising results prompted continued study into modified Secant procedures in the hopes of improving the quality of the produced Hessian updates and hence the numerical performance of quasi-Newton methods. This paper is organized as follows:
In the next section, we introduce some methods that addressed modifications to the Secant equation. In Section 3, we derive a novel Secant-like method using a quadratic model of the objective function. The global convergence for the method is then proven, under certain assumptions in Section 4. In Section 5, we present numerical results. Finally, we conclude our paper in Section 6 by a discussion and conclusions.

2 Self-scaling methods

Self-scaling variable-metric algorithms were proposed in the 1970s, and they proved to be significantly more efficient than previous approaches. Oren and Luenberger [23], Oren and Spedicato [24], Shanno and Phua [28] developed these techniques for minimizing an unconstrained nonlinear function in a number of articles. Although the BFGS formula has some important theoretical and numerical qualities, when applied to ill-conditioned situations it frequently suffers. To address this flaw, different strategies for scaling B_k before computing B_{k+1} have been proposed. We will look at a few of these strategies, bearing in mind that the presented strategies may be used not only with the BFGS method, but also with other Broyden family methods [5]. The advantageous qualities of the BFGS are preserved since these changes are not necessarily utilized when x_k and B_k are sufficiently close, in a sense, to a solution x^* and the corresponding Hessian, respectively. Several studies have looked into ways to increase the 'quality' of the Hessian approximation matrix by revisiting the traditional secant relation (1.1) to employ determined gradient, iteration difference vectors, and conveniently accessible function evaluations in some sort of a scaling process. (see, for instance, [2], [5],[7], [8], [12],[14], [22], [25],[27].

The self-scaling BFGS techniques have been devised, originally suggested and investigated for the minimization of quadratic functions, in order to improve the performances of the BFGS approach. Before updating the Hessian approximation B_k, Oren and Luenberger [23] scaled it; that is, they substituted B_k with $\beta_k B_k$, where β_k is a self-scaling factor determined to lower the condition number of R_k when applied to a quadratic function with Hessian G, where

$$R_k = G^{\frac{1}{2}} H_k G^{\frac{1}{2}}, \tag{2.9}$$

and H_k is the current inverse Hessian approximation.

The updating is done in such a way that

$$B_{k+1} s_k = \rho_k y_k. \tag{2.10}$$

When $\rho_k = 1$, this condition reduces to the Secant equation (1.1). G^{-1} is constant for a quadratic function and satisfies $s_k = G^{-1} y_k$ for any corresponding s_k and y_k; plainly, the goal of an adopted updating formula is that the matrix H_k tends to the inverse Hessian, $G^{-1}(x_k)$, for a general objective function. If f is a quadratic and exact line searches are carried out, $H_{n+1} = G^{-1}$ after n iterations. One known choice of ρ_k in (2.13) is [4]

$$\rho_k = \frac{s_k^T y_k}{4 s_k^T g_{k+1} + 2 s_k^T g_k - 6(f_{k+1} - f_k)}. \tag{2.11}$$

The search direction is initially calculated using the self-scaling memoryless BFGS approximation of the minimization function's Hessian as a solution of a linear algebraic problem. The initial self-scaling BFGS technique is described by the matrix $B_0 = \rho_0 I$, where ρ_0 is some positive scalar. The choice of the scaling parameter ρ_0 has a significant impact on the efficiency of the self-scaling BFGS technique. The goal of scaling the BFGS approximations to the Hessian of the reducing function is to enhance the distribution of the eigenvalues of the update matrix and hence the numerical stability. Oren and Luenberger provided the most essential equations for computing the scaling parameter ρ_k in the case of self-scaling BFGS update for quadratic objective functions [23]. The choice they used is given by

$$\rho_k = \frac{s_k^T y_k}{s_k^T s_k}. \tag{2.12}$$

Also, Oren and Spedicato suggested

$$\rho_k = \frac{y_k^T y_k}{s_k^T y_k}. \tag{2.13}$$

Nocedal and Yuan [23] considered the scaling

$$\rho_k = \frac{s_k^T y_k}{s_k^T B_k s_k}. \tag{2.14}$$

Nocedal [22] gave a detailed examination of the scaled BFGS technique using the scaling parameter proposed in [24]. When applied to a basic quadratic function of only two variables and employing inexact line search, they obtained very poor numerical results compared to the standard BFGS approach. This agrees with Shanno and Phua's findings [28]. Al-Baali [1] offered a minor improvement to this scaled BFGS technique for which he reported some results on the method's global and superlinear convergence. The rationale behind this improvement resulting from applying an initial scaling was that if $k = 1$, the eigenvalues of the Hessian approximation B_{k+1} could be decreased and, therefore, huge eigenvalues could be avoided.

The authors in [30] considered

$$\beta_k = \frac{s_k^T y_k}{4s_k^T g_{k+1} + 2s_k^T g_k - 6(f_{k+1} - f_k)}. \tag{2.15}$$

There were several additional modified BFGS algorithms proposed. Biggs [4] and Yuan et al. [31] produced some modified BFGS algorithms and established their global convergence properties using various function interpolation conditions. The goal of their strategy was to scale the BFGS updating formula's third term. The improved BFGS approach in [31] combines the gradient and function values information into a single step. Liao [18] proposed another self-scaling modified BFGS approach. Two positive scaling parameters were incorporated in this technique, which scale the second and third parts of the BFGS updating formula, correcting the eigenvalues of B_k better than the original unscaled BFGS.

The techniques discussed above lend themselves to a wide spectrum of applications such as machine learning, fluid mechanics, deep learning, solution of nonlinear equations, solution of differential equations and others. Another possible venue of applications is in Human Performance Technology (HPT). HPT mainly relies on the Performance Improvement (PI) characteristics of computing systems that in turn rely on logical decisions enabled by proposed algorithms. While many perceive PI synonymous to HPT, PI guides to measuring performance and structure elements within results-oriented systems [27]. Moreover, PI helps to extend the scope of instructional design (ID) since it employs systems approach to address performance opportunities and problems. A cross-sectional qualitative study [3] reveals that the use of mobile Electronic Performance Support Systems (EPSS) resulted in an increase in work performances and work efficiencies of its mobile users. Although EPSS gives satisfactory results to the users, utilization and efficient usage of systems resources demands strategic planning, which implies optimization of system resources. Optimization in this context can be achieved by assigning priorities to tasks based on techniques of mathematical optimization. Mathematical Optimization (MO) experts often carry on practices for optimization of process in industries. Implementing decision support tools to operational, tactical, and strategic planning and developmental processes fundamentally need optimized solutions. Due to COVID-19 pandemic, ongoing wave of digitalization also has had major impacts on the educational sector that needs optimization techniques to ensure efficient usage of system resources available online.

Next, we consider new self-scaling methods driven by the reported success of the methods outlined earlier in this study, which is spurred by a similar notion of integrating more of the data available on each iteration that the typical quasi-Newton methods often leave unexploited. A quadratic model of the objective function is used to propose self-scaling BFGS updates that satisfy (2.10). We then verify the new method's global convergence feature on generic functions under some reasonable assumptions.

3 A new self-scaling modified quasi-Newton method

To proceed with our derivation of the new self-scaling quasi-Newton method, we need the following convex quadratic model:

$$f_{k+1} = f_k + s_k^T g_k + \frac{1}{2} s_k^T Q(x_k) s_k, \tag{3.16}$$

for which

$$\nabla f_{k+1} = g_k + Q(x_k)s_k. \tag{3.17}$$

For exact line search, the optimal step size α_k along the direction d_k is given by (1.6).

From (1.6) and (3.17), we obtain

$$s_k^T Q(x_k)s_k = 2s_k^T y_k + 2(f_{k+1} - f_k), \tag{3.18}$$

thus yielding

$$Qs_k = \frac{2s_k^T y_k + 2(f_{k+1} - f_k)}{s_k^T y_k} y_k. \tag{3.19}$$

A reasonable approximation to the Hessian matrix of (3.16) is a sequence of positive definite matrices B_{k+1} which are required to satisfy

$$B_{k+1}s_k = \frac{2s_k^T y_k + 2(f_{k+1} - f_k)}{s_k^T y_k} y_k. \tag{3.20}$$

The relation in (3.20) can be expressed as a scaled version of (1.1) as

$$B_{k+1}s_k = \rho_k^{BA1} y_k, \tag{3.21}$$

where

$$\rho_k^{BA1} = \frac{2s_k^T y_k + 2(f_{k+1} - f_k)}{s_k^T y_k}, \tag{3.22}$$

termed as the $BA1$ method. On the other hand, using equation (3.18) in (3.16), we obtain

$$B_{k+1}s_k = \frac{f_k - f_{k+1} + \frac{1}{2}s_k^T y_k}{s_k^T y_k} y_k. \tag{3.23}$$

The above relation is similar to the scaling in (3.21) and is denoted by ρ_k^{BA2}. Thus,

$$B_{k+1}s_k = \rho_k^{BA2} y_k, \tag{3.24}$$

where

$$\rho_k^{BA2} = \frac{f_k - f_{k+1} + 1/(2s_k^T y_k)}{s_k^T y_k}. \tag{3.25}$$

However, if exact line search is assumed, then (3.22) and (3.25) reduce to

$$\rho_k^{BA3} = \frac{2\alpha_k g_k^T g_k + 2(f_{k+1} - f_k)}{s_k^T y_k} \, and \rho_k^{BA4} = \frac{f_k - f_{k+1} + 1/2\alpha_k g_k^T g_k}{s_k^T y_k} \tag{3.26}$$

respectively.

Outline of the new Algorithm

1. Choose some starting point $x_0 \in R^n$
2. Provide an initial symmetric positive-definite matrix estimate to the Hessian matrix H_0 (usually $H_0 = I$)
3. $k = 0$ (the iteration count)
4. Compute $g_0 = g(x_0)$
5. Repeat
   (a) Let $d_k = -H_k g_k$
   (b) Compute the step size along the search direction d_k: namely α_k, such that conditions (1.7) and (1.8) are satisfied;
   (c) $x_{k+1} = x_k + \alpha_k d_k$
   (d) compute $s_k = x_{k+1} - x_k$ and $y_k = g_{k+1} - g_k$
   (e) If $s_k^T y_k > 0$, then
   $H_{k+1} = H_k - \frac{H_k y_k s_k^T + s_k y_k^T H_k}{s_k^T y_k} + \left[\frac{1}{\rho_k^{BA1}} + \frac{y_k^T H_k y_k}{s_k^T y_k}\right] \frac{s_k s_k^T}{s_k^T y_k},;$
   (f) $k = k + 1$
6. until the stopping criteria stop1 are satisfied (see Section 5).

4 Global convergence property of the new algorithm

In this Section, the global convergence of the new methods will be established by proving the following property

$$\liminf_{k \Rightarrow \infty} \|g_k\| = 0 \tag{4.27}$$

The BFGS update generates conjugate gradient search directions provided that the objective function is quadratic and exact line searches are employed. To prove the same for the new updates, we first introduce a useful lemma which can be proved in a similar way to the proof of Lemma 5.1 in [33].

Lemma 1
If the BFGS algorithm is applied to the quadratic function

$$f(x) = \frac{1}{2}x^T Gx + b^T x, \tag{4.28}$$

using the same starting point x_0 and initial symmetric positive definite matrix H_0 , then

$$H_k g^\star = H_0 g^\star, \tag{4.29}$$

where g^* is the gradient at the minimum. The detailed proof can be found in [20]. Powell [25] established convergence for the standard conjugate gradient method by proving they satisfy

$$\liminf_{k \Rightarrow \infty} \|g_k\| = 0, \tag{4.30}$$

if the level set $x : f(x) \leq f(x_\star)$ is bounded and α_k is defined so that $g_{k+1}^T d_k = 0, k \geq 1$ holds for all k. The following theorem is used to show that the new methods generate the same search directions as the standard Conjugate Gradient method and hence global convergence follows.

Theorem 1
Assume that $f(x)$ is a quadratic function defined in (4.28) and that the line searches are exact: if H_k is any symmetric positive definite matrix and for the new updating formula

$$H_{k+1} = H_k - \frac{H_k y_k s_k^T + s_k y_k^T H_k}{s_k^T y_k} + \left[\frac{1}{\rho_k^{BA1}} + \frac{y_k^T H_k y_k}{s_k^T y_k}\right] \frac{s_k s_k^T}{s_k^T y_k}, \tag{4.31}$$

where $\rho_k^B A1$ defined in (3.22), then the search direction

$$d_{k+1}^{new} = -H_{k+1}^{new} g_{k+1} \tag{4.32}$$

is identical to the Hestenes/Stiefel Conjugate Gradient direction $d_C G$, defined as [16]

$$d_{k+1}^{CG} = -g_{k+1} + \frac{y_k^T g_{k+1}}{y_k^T d_k} d_k \text{ for } k \geq 1 \tag{4.33}$$

Proof
The update formula (4.31) can be expressed as

$$H_{k+1} = H_k - \frac{H_k y_k s_k^T}{s_k^T y_k} - \frac{s_k y_k^T H_k}{s_k^T y_k} + \left[\frac{1}{\rho_k^{BA1}} + \frac{y_k^T H_k y_k}{s_k^T y_k}\right] \frac{s_k s_k^T}{s_k^T y_k}$$

Now,

$$d_{k+1}^{new} = -H_k g_{k+1} + \frac{H_k y_k s_k^T g_{k+1}}{s_k^T y_k} + \frac{y_k^T H_k g_{k+1}}{s_k^T y_k} \frac{s_{k-1}}{\rho_k^{BA2}} \frac{s_k^T g_{k+1}}{s_k^T y_k} s_k - \frac{y_k^T H_k y_k s_k^T g_{k+1}}{(s_k^T y_k)^2} s_k \tag{4.34}$$

using the property $s_k^T g_{k+1} = 0$ quoted earlier which holds for line searches we get

$$d_{k+1}^{new} = -H_k g_{k+1} + \frac{y_k^T H_k g_{k+1}}{s_k^T y_k} s_k \tag{4.35}$$

The vector g_{k+1} can be substituted for $H_k g_{k+1}$ by using lemma 1.

Therefore,

$$d_{k+1}^{new} = -g_{k+1} + \frac{y_k^T g_{k+1}}{s_k^T y_k} s_k. \tag{4.36}$$

We also know that d^{BFGS} and d^{CG} are identical [20], and d^{new} is identical to d^{BFGS} with exact line searches. Hence equation (4.36) can be expressed as

$$d_{k+1}^{new} = -g_{k+1} + \frac{y_k^T g_{k+1}}{d_k^T y_k} d_k = d_{k+1}^{CG}. \tag{4.37}$$

Consequently, the proof is complete.

Upon applying lemma 1 to the sequence $\{B_k\}_{k \in K}$, there exist a_1 and a_2 such that (4.32) follows as k tends to infinity. This concludes the proof.

5 Numerical Results

This section details the numerical performance of the derived algorithms. We evaluate the proposed algorithms' numerical efficiency against the traditional BFGS formula on a collection of 30 test problems ranging in size from 2 to 1000. The test functions can be found in the CUTEst library [12] and the set offered in [22]. Table 1 lists the collection of the used test functions. The code was written in MATLAB 4.4 and tested on a Vaio Laptop running Windows XP with a 2.30 GHz processor and 4 GB of RAM. In all of our testing, we employed the strong Powell-Wolfe line search conditions (1.7) and (1.8).

The algorithms termination criteria used are defined as:

If $|f(x_k)| > 10^{-5}$, let $stop = \frac{|f(x_k) - f(x_{k+1})|}{|f(x_k)|}$;

Otherwise, let $stop = |f(x_k) - f(x_{k+1})|$. For every problem, if $\|g_k\| < \varepsilon$ or $stop < 10^{-5}$ is satisfied, the algorithm terminates. Table 1 summarizes the numerical results where NI, NF and NG denote the number of iterations,

		BFGS		ρ_k^{BA1}		ρ_k^{BA2}		ρ_k^{BA3}		ρ_k^{BA4}	
P. No.	n	IN	NF	IN	NF	IN	NF	IN	NF	IN	NF
Froth	2	9	26	10	28	7	20	10	32	11	34
Badscp	2	43	166	7	45	38	146	5	39	5	39
Badscb	2	3	30	3	30	3	30	3	30	3	30
Beale	2	15	50	4	16	12	42	15	76	12	43
Jensam	2	2	27	2	27	2	27	2	27	2	27
Bard	3	16	54	11	79	17	53	9	77	22	69
Gauss	3	2	4	2	4	2	4	2	4	2	4
Gulf	3	2	27	2	27	2	27	2	27	2	27
Box	3	2	27	2	27	2	27	2	27	2	27
Wood	4	19	61	8	49	18	59	25	102	24	85
Kowosb	4	21	65	5	36	22	67	16	94	25	76
Bd	4	17	54	9	52	13	43	34	131	33	127
Osb1	5	2	27	2	27	2	27	2	27	2	27
Biggs	6	25	72	19	79	26	76	6	40	19	82
Osb2	11	3	31	3	31	3	31	3	31	3	31
Singx	400	64	209	23	101	40	130	69	255	54	196
Pen1	400	2	27	2	27	2	27	2	27	2	27
Pen2	200	2	5	2	5	2	5	2	5	2	5
Vardim	100	2	27	2	27	2	27	2	27	2	27
Trig	500	9	33	19	103	10	34	15	72	10	35
Bv	500	2	4	2	4	2	4	2	4	2	4
Ie	500	6	16	7	42	7	19	7	14	7	19
Band	500	57	281	4	12	13	81	4	12	13	71
Lin	500	2	4	2	4	2	4	2	4	2	4
Lin1	500	3	7	3	7	3	7	3	7	3	7
Lin0	500	3	7	3	7	3	7	3	7	3	7
Total		333	1341	158	896	255	1024	247	1198	267	1130

Table 1: Comparison of different BFGS-algorithms with different test functions and different dimensions.

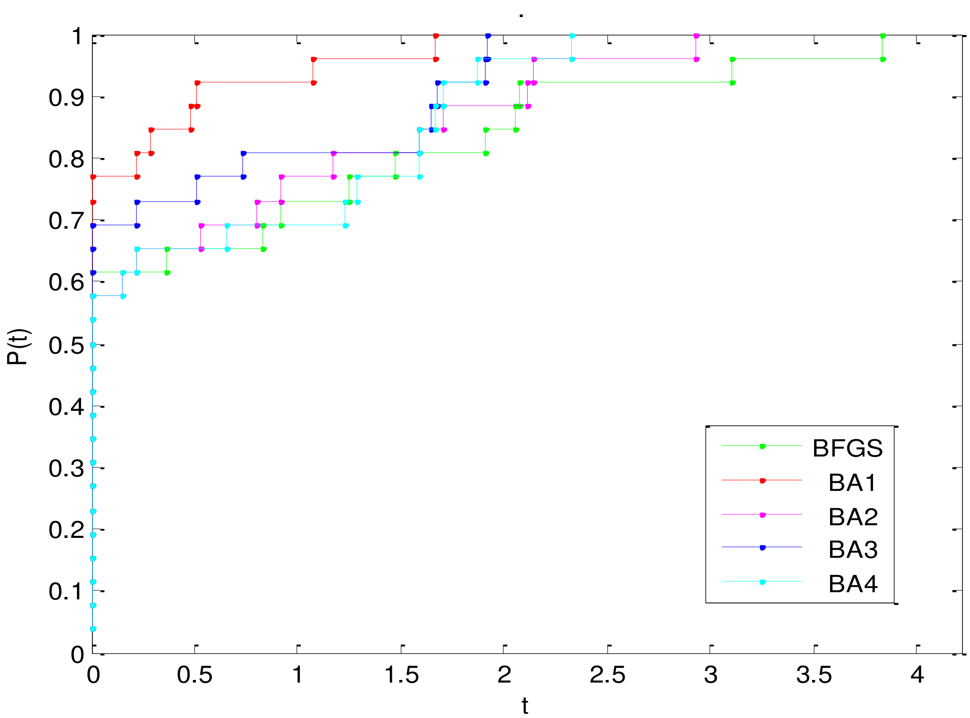

Figure 1: Performance based on the number of iterations.

function and gradient evaluations, respectively. The parameter n denotes the dimension of the test problem.

The graphical results of the test issues in terms of the overall number of iterations are shown in Figures 1 and 2. That is, the proportion p() of problems for which each strategy is within a factor of the optimal number of iterations for each approach is depicted. The top curve depicts the method that performs better in terms of iteration count that are within a factor of the best number of such count. Figures 1 and 2 indicates that ρ_k^{BA1} surpasses the other four methods, making it the best performer.

6 Discussion and Conclusions

In this paper, we have considered new self-scaling options for the Broyden family of updates though the focus of this research was on the most successful member; namely, the BFGS formula. We have shown that the new derivations are numerically successful. The approaches presented here enhance the performance of the traditional Secant algorithms and provide further attesta-

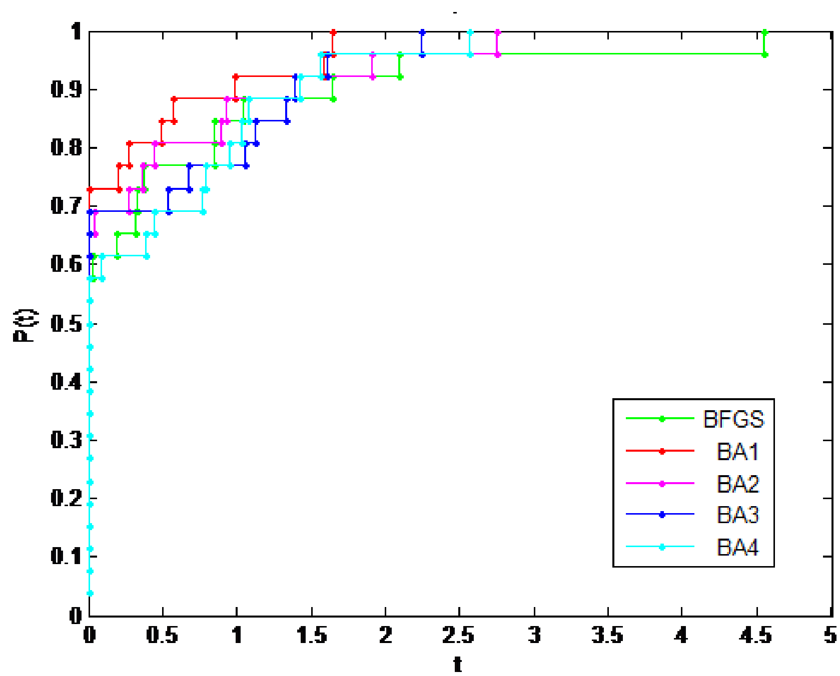

Figure 2: Performance based function evaluation.

tion to the viability of the self-scaling techniques. The test results presented in Table 1 and depicted in Figures 1 and 2 show that all the four scaling methods examined outperform the standard BFGS. On average, around 68 percent of the test functions were minimized with a smaller number of iterations and evaluations. The convergence of the derived methods was addressed for quadratic functions. The primary insight from this research is that scaling the terms of the conventional BFGS update may result in more efficient algorithms than the regular BFGS algorithm. Choosing the values for the scaling factors, on the other hand, is a difficult process. Another method is to scale the terms of the standard BFGS update during a subset of iterations, such as the first few iterations. Another interesting proposal in the same line of attempts to enhance the BFGS technique is to scale some of the terms of the BFGS update rather than the whole update [21]. In any case, the BFGS quasi-Newton techniques continue to amaze, and there is always space for improvement in terms of numerical performance.

References

[1] Mehiddin Al-Baali, Global and superlinear convergence of a restricted class of self-scaling methods with inexact line searches for convex functions, Computational Optimization and Applications, **9,** no. 2, (1998), 191–203.

[2] Mehiddin Al-Baali, Lucio Grandinetti, On the Behavior of Damped Quasi-Newton Methods for Unconstrained Optimization, Iranian Journal of Operations Research, **3,** no. 1, (2012), 1–10.

[3] Aadil Askar, Interactive ebooks as a tool of mobile learning for digital-natives in higher education: Interactivity, preferences, and ownership, Educational Technology Research and Development **60,** no. 1, (2014), 7–13.

[4] M. C. Biggs, Minimization algorithms making use of non-quadratic properties of the objective function, IMA Journal of Applied Mathematics, **8,** no. 3, (1971), 315–327.

[5] Charles George Broyden, The convergence of a class of double-rank minimization algorithms 1, general considerations, IMA Journal of Applied Mathematics, **6,** no. 1, (1970), 76–90.

[6] Yuhong Dai, Jinyun Yuan, Ya-Xiang Yuan. Modified two-point step-size gradient methods for unconstrained optimization, Computational Optimization and Applications, **22,** no. 1, (2002), 103–109.

[7] William C. Davidon, Variable metric method for minimization, SIAM Journal on Optimization, **1,** no. 1 (1991), 1–17.

[8] R. Dehghani, M. M. Hosseini, N. Bidabadi, The modified quasi-Newton methods for solving unconstrained optimization problems, International Journal of Numerical Modeling: Electronic Networks, Devices and Fields **32,** no. 1, (2019), e2459.

[9] John E. Dennis Jr., Robert B. Schnabel, Least change secant updates for quasi-Newton methods, Siam Review, **21,** no. 4, (1979), 443–459.

[10] Roger Fletcher, An overview of unconstrained optimization, Algorithms for Continuous Optimization, (1994).

[11] J. A. Ford, I. A. Moghrabi, Alternative parameter choices for multi-step quasi-Newton methods, Optimization Methods and Software, **2,** nos. 3-4, (1993), 357–370.

[12] J. A. Ford, I. A. R. Moghrabi, Multi-step quasi-Newton methods for optimization, Journal of Computational and Applied Mathematics **50,** nos. 1-3, (1994), 305–323.

[13] Basim A. Hassan, M. Ghada, A New Quasi-Newton Equation on the Gradient Methods for Optimization Minimization Problem, Indonesian Journal of Electrical Engineering and Computer Science, **19,** no. 2, (2020), 737–744.

[14] Basim A. Hassan, A new type of quasi-Newton updating formulas based on the new quasi-Newton equation, Numerical Algebra, Control & Optimization, **10,** no. 2, (2020), 227.

[15] Basim A. Hassan, A new type of quasi-Newton updating formulas based on the new quasi-Newton equation, Numerical Algebra, Control & Optimization, **10,** no. 2, (2020), 227.

[16] Magnus Rudolph Hestenes, Eduard Stiefel, Methods of conjugate gradients for solving linear systems, 1952.

[17] Dong-Hui Li, Masao Fukushima, A modified BFGS method and its global convergence in nonconvex minimization, Journal of Computational and Applied Mathematics, **129,** nos. 1-2, (2001), 15–35.

[18] Aiping Liao, Modifying the BFGS method, Operations research letters, **20,** no. 4, (1997), 171–177.

[19] Issam A. R. Moghrabi, New two-step conjugate gradient method for unconstrained optimization, International Journal of Applied Mathematics, **33,** no. 5, (2020), 853.

[20] Larry Nazareth, A relationship between the BFGS and conjugate gradient algorithms and its implications for new algorithms, SIAM Journal on Numerical Analysis, **16,** no. 5, (1979), 794–800.

[21] Jorge Nocedal, Ya-xiang Yuan, Analysis of a self-scaling quasi-Newton method, Mathematical Programming, **61,** no. 1, (1993), 19–37.

[22] Jorge Nocedal, Theory of algorithms for unconstrained optimization, Acta Numerica, **1,** no. 1 (1992), 199–242.

[23] Shmuel S. Oren, David G. Luenberger, Self-scaling variable metric (ssvm) algorithms: Part i: Criteria and sufficient conditions for scaling a class of algorithms, Management Science, **20,** no. 5, (1974), 845–862.

[24] Shmuel S. Oren, Emilio Spedicato, Optimal conditioning of self-scaling variable metric algorithms, Mathematical Programming, **10,** no. 1, (1976), 70–90.

[25] Michael J. D. Powell, Nonconvex minimization calculations and the conjugate gradient method, In Numerical analysis, Springer, Berlin, Heidelberg, 1984, 122–141.

[26] N. Y. Deng, Z. F. Li, Some global convergence properties of a conic-variable metric algorithm for minimization with inexact line searches. Optimization Methods and Software, **5,** no. 1, (1995), 105–122.

[27] Rita C. Richey (ed.), Encyclopedia of terminology for educational communications and technology, Springer, New York, 2013.

[28] David F. Shanno, Kang-Hoh Phua, Matrix conditioning and nonlinear optimization, Mathematical Programming, **14,** no. 1, (1978), 149–160.

[29] Zengxin Wei, Guoyin Li, Liqun Qi, New quasi-Newton methods for unconstrained optimization problems, Applied Mathematics and Computation, **175,** no. 2, (2006), 1156–1188.

[30] Wenyu Sun, Ya-Xiang Yuan, Optimization theory and methods: nonlinear programming, **1,** Springer Science & Business Media, 2006.

[31] Gonglin Yuan, Zengxin Wei, Yanlin Wu, Modified limited memory BFGS method with nonmonotone line search for unconstrained optimization, Journal of the Korean Mathematical Society, **47,** no. 4, (2010), 767–788.

[32] J. Z. hang, N. Y. Deng, L. H. Chen, New quasi-Newton equation and related methods for unconstrained optimization, Journal of Optimization Theory and Applications, **102,** no. 1, (1999), 147–167.

[33] Zhiwei Tony Qin, The relationships between CG, BFGS, and two limited-memory algorithms, Furman University Electronic Journal of Undergraduate Mathematics, **12,** no. 1, (2007), 5–20.