Mediterr. J. Math. (2018) 15:11
https://doi.org/10.1007/s00009-017-1060-1
1660-5446/18/010001-13
*published online* December 28, 2017
© Springer International Publishing AG, part of Springer Nature 2017

Mediterranean Journal of Mathematics

# Global Convergence Property of Scaled Two-Step BFGS Method

Fahimeh Biglari, Ali Ebadian and Mohammadreza Foroutan

**Abstract.** This paper is aimed to extend the scheme of self scaling, appropriate for the quasi-Newton methods, to the two-step quasi-Newton methods. The scaling scheme has been performed during the main approach of updating the current Hessian approximation and prior to the computation of the next quasi-Newton direction whenever necessary. Global convergence property of the new method is explored on uniformly convex functions with the standard Wolfe line search. Preliminary numerical testing has been performed showing that this technique improves the performance of the two-step method substantially.

**Mathematics Subject Classification.** 90C53.

**Keywords.** Nonlinear programming, multi-step quasi-Newton methods, convergence analysis, self-scaling scheme, BFGS method.

## 1. Introduction

This paper is concerned with the minimization of an objective function $f : \mathbb{R}^n \to \mathbb{R}$

$$\min f(x)$$

where $x \in \mathbb{R}^n$ is a real vector and $f$ is a nonlinear and twice continuously differentiable objective function. If we denote the gradient of $f$ at iterative point $x_k$ by $g_k$ and the Hessian approximation of $f$ at $x_k$ by $B_k$, then quasi-Newton algorithms proceed with the basic iteration

$$x_{k+1} = x_k - \alpha_k B_k^{-1} g_k,$$

where $\alpha_k$ is the steplength. In the BFGS method, the updated Hessian approximation is given by

$$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{y_k^T s_k}, \tag{1.1}$$

where $s_k$ and $y_k$ are given by the expressions $s_k = x_{k+1} - x_k$ and $y_k = g_{k+1} - g_k$, respectively, and the secant equation

$$B_{k+1}s_k = y_k \tag{1.2}$$

is satisfied. Furthermore, $p_k = -B_k^{-1}g_k$ is the quasi-Newton search direction, see [20]. The BFGS method preserves the positive definiteness of the updated matrices $B_k$ if the curvature condition $s_k^T y_k > 0$ holds. This condition is guaranteed if $\alpha_k$ is chosen such that the following strong Wolfe conditions hold

$$f(x_k + \alpha_k p_k) \leq f(x_k) + \sigma_0 \alpha_k g_k^T p_k, \tag{1.3}$$

$$|g_{k+1}^T p_k| \leq \sigma_1 |g_k^T p_k|, \tag{1.4}$$

where $p_k = -B_k^{-1}g_k$ and $0 < \sigma_0 < \frac{1}{2} < \sigma_1 < 1$, and $\sigma_0$ and $\sigma_1$ are constants. Moreover, the BFGS update is scale-invariant if the initial Hessian approximations $B_0$ are chosen appropriately, see [20].

Constant scaling of the quasi-Newton methods, known as self-scaling methods, has been received much attention by several authors. For example, see Oren [21], Oren and Luenberger [22], Oren and Spedicato [23], Spedicato [29], Dennis and Wolkowicz [11], Shanno [26], Shanno and Phua [28], Al-Baali [2,3,5,7] and references there in. The scheme of constant scaling is intended to improve the eigenvalue distribution or to reduce the condition number of the $B_k + 1$ matrix. In this way the performance of the method is accelerated. General formulation of scaling the single step Hessian updates given by (1.1), is as follows.

$$B_{k+1} = \lambda_k \left( B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} \right) + \frac{y_k y_k^T}{y_k^T s_k},$$

where $\lambda_k$ is the scaling parameter. If $\lambda_k = 1$, this family reduces to the unscaled BFGS update. Several expressions have been formulated for the parameter $\lambda_k$ in the mentioned literatures above, and the convergence behaviour of the corresponding methods have been analyzed. Numerical experiments exposing substantial improvement of some of these approaches with respect to the standard Broyden families were also reported, see [2]. Al-Baali [1] considered the following relation for the direct BFGS updates

$$\det(B_{k+1}) = \frac{1}{b_k}\det(B_k),$$

where

$$b_k = \frac{s_k^T B_k s_k}{s_k^T y_k},$$

proposed by Spedicato [30], and suggest some estimates for the size of the eigenvalues of $B_k$ at each iteration to recognize the case of large eigenvalues. He argued that when $b_k$ is small, the determinant increases, implying that, some of the eigenvalues of $B_k$ increase, though some may also decrease. Thus, in this case, the BFGS updating formula has a strong self-correcting property with respect to the determinant, see [9,25]. But, when $b_k$ is sufficiently large, the BFGS formula has slight corrections to $B_k$, although the determinant

decreases. This fact provided a motivation for Al-Baali's choice for the scaling parameter as $\min(1/b_k, 1)$ to cope with the case where $b_k > 1$. In the case of dealing with the inverse Hessian approximations, inequality $b_k > 1$ is equivalent to the following inequality:

$$h_k = \frac{y_k^T H_k y_k}{s_k^T y_k} < 1,$$

see [4]. Therefore, the scaling parameter is selected as follows:

$$\lambda_k = \max\left(\frac{1}{h_k}, 1\right), \tag{1.5}$$

which satisfies condition $\lambda_k \geq 1$. In this way, the constant scaling of the inverse Hessian update is given by

$$H_{k+1} = \lambda_k \left( H_k - \frac{H_k y_k s_k^T + s_k y_k^T H_k}{s_k^T y_k} + \left( \frac{1}{\lambda_k} + \frac{y_k^T H_k y_k}{s_k^T y_k} \right) \frac{s_k s_k^T}{s_k^T y_k} \right).$$

In the following section, we recall the two-step BFGS method and describe the generalization of the scaling technique for such methods. The global and local convergence result of the self-scaling two-step BFGS method is presented in Sect. 3. In Sect. 4, we will describe the detailed algorithm of scaled two-step BFGS, Algorithm 4.1, and we compare the Dolan and Moré [12] performance profile of the new algorithm with ordinary BFGS method, DBFGS by Al-baali and Grandinetti [6] and F21 by Ford and Moghrabi [15]. The conclusion is finally outlined in Sect. 5.

## 2. Generalized Self Scaling Technique for the Two-Step BFGS Method

In this section, the technique of self-scaling proposed for quasi-Newton methods is generalized to the two-step BFGS method. We aim to improve the performance of the method by altering the eigenvalue distribution of the updated matrices. We begin our description of this approach by briefly reviewing the multi-step BFGS method. Multi-step quasi-Newton methods proposed by Ford and Moghrabi [14–18], are characterized by updating the existing Hessian approximation (or its inverse) by means of data deriving from the $m$ most recent iterations. In these methods, the iteration points are interpolated by a curve $\mathbb{X} \equiv \{x(t)\}$ in its Lagrangian form, where $t \in \mathbb{R}$ and $x(t)$ is an interpolating polynomial of degree $m$ satisfying

$$x(t_i) = x_{k-m+i+1}, \qquad \text{for } i = 0, 1, \ldots, m,$$

for given values $\{t_i\}_{i=0}^m$. The basic idea behind these methods is that by means of the chain rule to the vector function $g(x(t))$ it is obtained

$$G(x(t))\frac{\mathrm{d}x(t)}{\mathrm{d}t} = \frac{\mathrm{d}g(x(t))}{\mathrm{d}t}, \tag{2.1}$$

where $G(x) = \nabla^2 f(x)$. The Eq. (2.1) can be used in the construction of the required approximation $B_{k+1}$ of the Hessian $G(x_{k+1})$, by requiring such

approximation to satisfy a relation of the form

$$B_{k+1}r_k = w_k. \tag{2.2}$$

Here, the vectors $r_k$ and $w_k$ are given by

$$r_k = \sum_{j=0}^{m-1} s_{k-j} \left\{ \sum_{i=m-j}^{m} L'_k(t_m) \right\}, \qquad w_k = \sum_{j=0}^{m-1} y_{k-j} \left\{ \sum_{i=m-j}^{m} L'_k(t_m) \right\},$$

where $L_j(t)$ is the $j^{th}$ Lagrange polynomial of degree $m$ associated with $\{t_i\}_{i=0}^{m}$. We note that the standard quasi-Newton methods correspond to $m = 1$, deriving from the step just completed. For this case, the path $\mathbb{X}$ is taken to be the straight line which interpolates the two most recent iterates.

Typical value of the memory parameter $m$, considered in the literature, is $m = 2$. Such methods are called two-step methods. In this case, the matrix $B_{k+1}$ is required to satisfy a simple adaptation of the secant Eq. (2.2).

$$B_{k+1}(s_k - \gamma_k s_{k-1}) = y_k - \gamma_k y_{k-1},$$

where $\gamma_k$ is a scalar that is determined by the parameters defining the curve $\mathbb{X}$, and different choices of this parameter lead to different two-step methods.

One of the most effective two-step quasi-Newton methods for updating $B_k$ is given by

$$B_{k+1} = B_k - \frac{B_k r_k r_k^T B_k}{r_k^T B_k r_k} + \frac{w_k w_k^T}{r_k^T w_k}. \tag{2.3}$$

In (2.3),

$$r_k = s_k - \frac{\delta_k^2}{1 + 2\delta_k} s_{k-1}, \tag{2.4}$$

$$w_k = y_k - \frac{\delta_k^2}{1 + 2\delta_k} y_{k-1}, \tag{2.5}$$

where the quantity $\delta_k$ is defined by the ratio

$$\delta_k = \frac{t_2 - t_1}{t_1 - t_0}. \tag{2.6}$$

In which

$$t_2 = 0, \tag{2.7}$$

$$t_1 = -\{s_k^T B_k s_k\}^{\frac{1}{2}} = -\{-\alpha_k s_k^T g_k\}^{\frac{1}{2}}, \tag{2.8}$$

$$\begin{aligned} t_0 &= -\{(s_k + s_{k-1})^T B_k (s_k + s_{k-1})\}^{\frac{1}{2}} \\ &\approx -\{-\alpha_k s_k^T g_k + 2 s_k^T y_{k-1} + s_{k-1}^T y_{k-1}\}^{\frac{1}{2}}. \end{aligned} \tag{2.9}$$

The Eq. (2.9) is satisfied exactly if two-step method is alternated with the standard single-step approach on successive iterations. In addition, to ensure the generation of positive-definite updates, the following requirement is imposed

$$r_k^T w_k > 10^{-4} \|r_k\| \|w_k\|. \tag{2.10}$$

If this condition on $r_k^T w_k$ is not satisfied, the use of the unit-spaced approach ($\delta_k = 1$) or single-step iteration are possible, for more details see [15].

Assuming that $H_k = B_k^{-1}$ is the inverse Hessian approximation, the inverse update of (2.3) is given by

$$H_{k+1} = H_k - \frac{H_k w_k r_k^T + r_k w_k^T H_k}{r_k^T w_k} + \left(1 + \frac{w_k^T H_k w_k}{r_k^T w_k}\right) \frac{r_k r_k^T}{r_k^T w_k}, \tag{2.11}$$

and the condition

$$H_{k+1} w_k = r_k,$$

on the new inverse Hessian approximation $H_{k+1}$ is as a replacement for the condition imposed by the secant equation.

Now, we apply the technique of scaling in the framework of the two-step method, and we compare the performance of the resulting algorithm in Sect. 4. In light of the similar arguments led to (1.5), scaling of the modified two-step method is required whenever

$$h_k = \frac{w_k^T H_k w_k}{r_k^T w_k} < 1. \tag{2.12}$$

Assuming that $\lambda_k = 1/\min(h_k, 1)$, the scaled modified two-step update is given by:

$$H_{k+1} = \lambda_k \left( H_k - \frac{H_k w_k r_k^T + r_k w_k^T H_k}{r_k^T w_k} + \left(\frac{1}{\lambda_k} + \frac{w_k^T H_k w_k}{r_k^T w_k}\right) \frac{r_k r_k^T}{r_k^T w_k} \right). \tag{2.13}$$

Finally, we let the scaled two-step method be alternated with the standard single-step approach on successive iterations so that the Eq. (2.9) to be satisfied exactly. In more detail, the first and the second iteration will perform as the single-step BFGS method. From the third iteration, two-step method will be performed if the condition (2.10) is satisfied. In this case, the next iteration will be the single-step iteration, while the ordinary secant Eq. (1.2) will be satisfied exactly. In other case, as long as the condition (2.10) is not satisfied, the single-step approach will be used.

## 3. Convergence Analysis

In this section, we establish the global convergence property of the scaled two-step method. We will combine the convergence results of Nocedal and Wright [7] for using two tools of the trace and determinant in the convergence analysis, and Al-Baali [2] for self-scaling technique. To study the convergence analysis of the method, we consider the following direct Hessian update formula

$$B_{k+1} = \lambda_k \left( B_k - \frac{B_k r_k r_k^T B_k}{r_k^T B_k r_k} \right) + \frac{w_k w_k^T}{w_k^T r_k}, \tag{3.1}$$

where $\lambda_k = \min(1/b_k, 1)$ is the scaling factor in which $b_k = r_k^T B_k r_k / w_k^T r_k$. Furthermore, we impose the additional conditions that

$$y_k^T s_k + \gamma_k^2 y_{k-1}^T s_{k-1} - \gamma_k (y_{k-1}^T s_k + y_k^T s_{k-1}) > \nu_1, \text{ and } |\delta_k + \frac{1}{2}| \geq \nu_2 \tag{3.2}$$

when

$$\gamma_k (y_{k-1}^T s_k + y_k^T s_{k-1}) > 0,$$

where $\gamma_k$ denotes the ratio $\delta_k^2/(1 + 2\delta_k)$, and $\nu_1$ and $\nu_2$ are small positive constants. More precisely, we concern with the following algorithm.

**Algorithm 3.1.** *Given initial point $x_0$, Hessian approximation $B_0$, convergence tolerance $\epsilon > 0$, small positive constants $\nu_1$ and $\nu_2$;*
*$k \leftarrow 0$;*
*while $\|g_k\| > \epsilon$;*
*Compute search direction $B_k p_k = -g_k$;*
*Choose a steplength $\alpha_k$ along $p_k$ such that the strong Wolfe conditions (1.3) and (1.4) are satisfied, and set $x_{k+1} = x_k + \alpha_k p_k$;*
*Define $s_k = x_{k+1} - x_k$ and $y_k = g_{k+1} - g_k$;*
*If a single-step iteration is being executed,*
*then set $r_k = s_k$ and $w_k = y_k$;*
*else*
*compute $\{t_j\}_{j=0}^2$ and $\delta_k$, from (2.6)–(2.9);*
*evaluate $\gamma_k = \delta_k^2/(1 + 2\delta_k)$;*
*If the condition $\gamma_k(y_{k-1}^T s_k + y_k^T s_{k-1}) \leq 0$ or the condition (3.2) is satisfied then Compute $r_k$ and $w_k$, defined by (2.4) and (2.5);*
*If ($r_k^T w_k \leq 10^{-4}\|r_k\|\|w_k\|$ or $|\delta_k| > \delta_{max}$ or $|\delta_k + \frac{1}{2}| < \nu_2$) and $(y_{k-1}^T s_k + y_k^T s_{k-1}) \leq 0$ or (3.2) is satisfied with $\delta_k = 1$*
*then recompute $r_k$ and $w_k$ using (2.4) and (2.5) with $\delta_k = 1$.*
*If $r_k^T w_k \leq 10^{-4}\|r_k\|\|w_k\|$ again,*
*then set $r_k = s_k$ and $w_k = y_k$.*
*Update $B_k$ by means of (3.1);*
*$k \leftarrow k + 1$;*
*end(while)*

Now, we state our assumptions about the objective function. We denote the Euclidean vector or matrix norm by $\|.\|$. We also assume that the scaling parameter is chosen, as in Al-Baali [7], such that

$$\nu_3 \leq \lambda_k \leq 1, \tag{3.3}$$

where $\nu_3$ is a positive constant.

**Assumption 3.1.** (a) *The objective function $f$ is twice continuously differentiable.*
(b) *The level set $\Omega = \{x \in \mathbb{R}^n : f(x) \leq f_(x_0)\}$ is bounded convex set, and there exist positive constants $m$ and $M$ such that*

$$m\|z\|^2 \leq z^T G(x) z \leq M\|z\|^2, \tag{3.4}$$

*for all* $z \in \mathbb{R}^n$ *and all* $x \in \Omega$.

It is noted that the condition (3.4) implies the following consequences

$$\frac{y_k^T s_k}{s_k^T s_k} \geq m, \qquad \frac{y_k^T y_k}{y_k^T s_k} \leq M, \qquad \text{and} \qquad \|y_k\| \leq M\|s_k\|, \tag{3.5}$$

see [7,20]. Moreover, from the boundedness of the level set $\Omega$, there is a positive constant $\bar{M}$ such that

$$\|x\| \leq \bar{M} \tag{3.6}$$

for all $x \in \Omega$. The following theorem is expressed to study the global convergence of the method satisfying Assumption 3.1 and the conditions (1.3) and (1.4).

**Theorem 3.2.** *Let* $B_0$ *be any symmetric positive definite initial matrix, and let* $x_0$ *be a starting point, for which Assumption 3.1 is satisfied. Then, the sequence* $\{x_k\}$ *generated by Algorithm 3.1 converges to the minimizer* $x^*$ *of* $f$.

*Proof.* We define

$$m_k = \frac{w_k^T r_k}{r_k^T r_k}, \qquad \text{and} \qquad M_k = \frac{w_k^T w_k}{w_k^T r_k}, \tag{3.7}$$

and show that there are positive constants $m_1$ and $M_1$ such that

$$m_k \geq m_1, \qquad \text{and} \qquad M_k \leq M_1. \tag{3.8}$$

First, we consider the case $\gamma_k(y_{k-1}^T s_k + y_k^T s_{k-1}) \leq 0$. Using (3.5), we have

$$\begin{aligned} M_k = \frac{w_k^T w_k}{w_k^T r_k} &\leq \frac{\|y_k^T\|^2 + 2|\gamma_k|\|y_k^T\|\|y_{k-1}\| + \gamma_k^2\|y_{k-1}^T\|}{\frac{\|y_k\|^2}{M} + \gamma_k^2\frac{\|y_{k-1}\|^2}{M}} \\ &\leq \frac{(\|y_k^T\| + |\gamma_k|\|y_{k-1}\|)^2}{\frac{1}{M}(\|y_k\|^2 + \gamma_k^2\|y_{k-1}\|^2)} \leq 2M. \end{aligned} \tag{3.9}$$

And

$$\begin{aligned} m_k = \frac{w_k^T r_k}{r_k^T r_k} &\geq \frac{m\|s_k\|^2 + m\gamma_k^2\|s_{k-1}\|^2}{\|s_k\|^2 + 2|\gamma_k|\|s_k\|\|s_{k-1}\| + \gamma_k^2\|s_{k-1}\|^2} \\ &\geq \frac{m(\|s_k\|^2 + \gamma_k^2\|s_{k-1}\|^2)}{(\|s_k\| + |\gamma_k|\|s_{k-1}\|)^2} \geq \frac{m}{2}. \end{aligned} \tag{3.10}$$

For this case, we see from (3.9) and (3.10) that the inequalities (3.8) are satisfied with $m_1 = \frac{m}{2}$ and $M_1 = 2M$. In the case, $\gamma_k(y_{k-1}^T s_k + y_k^T s_{k-1}) > 0$,

we have from the condition (3.2), (3.5) and (3.6) that

$$\begin{aligned} M_k = \frac{w_k^T w_k}{w_k^T r_k} &\leq \frac{\|y_k\|^2 + 2|\gamma_k|\|y_k\|\|y_{k-1}\| + \gamma_k^2\|y_{k-1}\|^2}{y_k^T s_k + \gamma_k^2 y_{k-1}^T s_{k-1} - \gamma_k(y_{k-1}^T s_k + y_k^T s_{k-1})} \\ &\leq \frac{1}{\nu_1}(\|y_k\| + |\gamma_k|\|y_{k-1}\|)^2 \leq \frac{M^2}{\nu_1}\left(\|s_k\| + |\frac{\delta^2}{1+2\delta}|\|s_{k-1}\|\right)^2 \\ &\leq \frac{M^2}{\nu_1}\left(\|x_{k+1}\| + \|x_k\| + \frac{\delta_{\max}^2}{2\nu_2}(\|x_k\| + \|x_{k-1}\|)\right)^2 \\ &\leq \frac{4\bar{M}^2 M^2}{\nu_1}\left(1 + \frac{\delta_{\max}^2}{2\nu_2}\right)^2. \end{aligned} \tag{3.11}$$

And

$$\begin{aligned} m_k = \frac{w_k^T r_k}{r_k^T r_k} &\geq \frac{y_k^T s_k + \gamma_k^2 y_{k-1}^T s_{k-1} - \gamma_k(y_{k-1}^T s_k + s_{k-1}^T y_k)}{\|s_k\|^2 + 2|\gamma_k|\|s_k\|\|s_{k-1}\| + \gamma_k^2\|s_{k-1}\|^2} \\ &\geq \frac{\nu_1}{(\|s_k\| + |\gamma_k|\|s_{k-1}\|)^2} \geq \frac{\nu_1}{4\bar{M}^2\left(1 + \frac{\delta_{\max}^2}{2\nu_2}\right)^2}. \end{aligned} \tag{3.12}$$

Hence, the inequalities (3.8) are satisfied with $m_1 = \nu_1 / \left(4\bar{M}^2\left(1 + \frac{\delta_{\max}^2}{2\nu_2}\right)^2\right)$ and $M_1 = \frac{4\bar{M}^2 M^2}{\nu_1}\left(1 + \frac{\delta_{\max}^2}{2\nu_2}\right)^2$ in this case. Subsequently, using (3.8) and the arguments in [20], the following inequality holds.

$$\text{trace}(B_{k+1}) = \text{trace}(B_k) - \frac{\|B_k r_k\|^2}{r_k^T B_k r_k} + \frac{\|w_k\|^2}{w_k^T r_k} \leq \text{trace}(B_k) + M_1. \tag{3.13}$$

Likewise, we can have a simple expression for the determinant as well, see [20,24].

$$\det(B_{k+1}) = \det(B_k)\frac{r_k^T w_k}{r_k^T B_k r_k}\lambda_k^{n-1}.$$

By (3.8), (3.3) and the fact that the largest eigenvalue of $B_k$ is also less than $\text{trace}(B_k)$, we have

$$\det(B_{k+1}) = \det(B_k)\frac{r_k^T w_k}{r_k^T r_k}\frac{r_k^T r_k}{r_k^T B_k r_k}\lambda_k^{n-1} \geq \det(B_k)\frac{m_1}{2\text{trace}(B_k)}\nu_3^{n-1}. \tag{3.14}$$

Since (3.13) and the determinant inequality (3.14) hold, then the rest of the proof can be proceeded from Theorem 3.1. of [10]. □

To analyze the rate of convergence of the Algorithm 3.1, we further assume that the Hessian matrix $G(x)$ is Lipschitz continuous at $x^*$, that is, there exists a positive constant $L$ such that

$$\|G(x) - G(x^*)\| \leq L\|x - x^*\|, \tag{3.15}$$

for all $x$ near $x^*$.

**Theorem 3.3.** *Suppose that $f$ is twice continuously differentiable and that the iterates generated by the Algorithm 3.1 converge to a minimizer $x^*$ at which Assumption 3.1 and (3.15) hold. Assume that the line search scheme for finding $\alpha_k$ starts with testing $\alpha_k = 1$. Then the sequence $x_k$ generated by Algorithm 3.1 converges to $x^*$ at a superlinear rate.*

*Proof.* Using expressions (3.13) and (3.12), the result follows from Theorem (3.2), its proof, and Theorem 4.1. of Byrd et al. [10]. □

## 4. Numerical Results

In this section, we study the performance of the scaled two-step algorithm, denoted SF21 in this paper. We give an algorithm description of the method in updating for inverse Hessian approximation. As it was described in Sect. 2, scaling of the updated matrices is required when the eigenvalues of the inverse Hessian estimates are large. Hence, scaling procedure is employed whenever the inequality (2.12) is hold.

**Algorithm 4.1.** *(Scaled Modified two-step algorithm, denoted by SF21)*
*Given initial point $x_0$, inverse Hessian approximation $H_0$, convergence tolerance $\epsilon > 0$ small positive constants $\nu_1$ and $\nu_2$;*
$k \leftarrow 0$;
*while a convergence test does not hold;*
*Compute the search direction $p_k = -H_k g_k$;*
*Choose a steplength $\alpha_k$ along $p_k$ such that the strong Wolfe conditions (1.3) and (1.4) are satisfied: and set*

$$x_{k+1} = x_k + \alpha_k p_k;$$

*Define $s_k = x_{k+1} - x_k$ and $y_k = g_{k+1} - g_k$;*
*If a single-step iteration is being executed,*
*then set $r_k = s_k$ and $w_k = y_k$;*
*else*
*compute $\{t_j\}_{j=0}^{2}$ and $\delta_k$, from (2.6)–(2.9);*
*evaluate $\gamma_k = \delta_k^2/(1 + 2\delta_k)$;*
*If the condition $\gamma_k(y_{k-1}^T s_k + y_k^T s_{k-1}) \leq 0$ or the condition (3.2) is satisfied then Compute $r_k$ and $w_k$, defined by (2.4) and (2.5);*
*If $(r_k^T w_k \leq 10^{-4}\|r_k\|\|w_k\|$ or $|\delta_k| > \delta_{max}$ or $|\delta_k + \frac{1}{2}| < \nu_2)$ and $(y_{k-1}^T s_k + y_k^T s_{k-1}) \leq 0$ or (3.2) is satisfied with $\delta_k = 1$*
*then recompute $r_k$ and $w_k$ using (2.4) and (2.5) with $\delta_k = 1$.*
*If $r_k^T w_k \leq 10^{-4}\|r_k\|\|w_k\|$ again,*
*then set $r_k = s_k$ and $w_k = y_k$.*
*Update $H_k$ by means of (2.13);*
$k \leftarrow k + 1$;
*end(while)*

We compare this method with

- $\text{BFGS}_0$, standard BFGS algorithm [13],
- BFGS algorithm by Shanno and Phua [27],

- F21, two-step algorithm by Ford and Moghrabi [15],
- DBFGS, damped BFGS algorithm by Al-Baali and Grandinetti [6].

For the $\text{BFGS}_0$, and DBFGS methods, we let the initial Hessian approximation be defined by $H_0 = I$, and for the methods included BFGS, F21 and SF21 the starting matrix after computing the first step but before performing the first quasi-Newton update, is scaled by setting

$$H_0 = \frac{y_k^T s_k}{y_k^T y_k} I.$$

All codes are written in Fortran 77 in double precision and implement the same stopping criterion

$$\|g\|_2 \leq 10^{-5} * \max(1, \|x\|). \tag{4.1}$$

The process is also stopped if the number of function evaluations reaches 5000. All algorithms use exactly the same line search strategy, based on quadratic and cubic interpolations. Line searches are terminated when the strong Wolfe conditions (1.3) and (1.4), with $\sigma_0 = 0.0001$ and $\sigma_1 = 0.9$ are satisfied (see e.g., Fletcher, [13]). For the two-step methods, the safeguarding parameter $\delta_{\max}$ was set to the value 22.0, as suggested by Ford and Moghrabi [15]. Furthermore, the values of $\nu_1 = 10^{-13}$ and $\nu_2 = \nu_3 = 10^{-4}$ are chosen. It is noted that two-step algorithms F21 and SF21 are sensitive to the selected value of $\nu_1$.

We examine 73 test problems (the same were used by Al-Baali, [3]) which are selected from Andrei [8] with the standard starting points. For each test problem, 10 runs with dimension ranging from 4 to 500 are performed. Our metric for assessing the performance of the algorithms is the reduction on the number of function-gradient evaluations to solve each problem. For all the methods, every gradient evaluation was counted as equal to one function value calculation.

In this comparative study, the fraction $P$ of problems is plotted for any given method within a factor $\tau$ of the best number of function-gradient evaluations, based on the performance profiles of Dolan and Moré [12].

It is observed from Figure 1 that the SF21 is always the top performer for all values of $\tau$. Hence, the use of the scaled modified two-step enables a considerable computational saving on most problems with respect to $\text{BFGS}_0$, DBFGS, BFGS and F21 methods, in terms of number of function-gradient calls.

Despite some occasional good results generated by SF21 for the set of measurements being considered here, this method fails to satisfy the stopping criterion (4.1) in 16 out of 730 test problems. Meanwhile, the number of failures for the $\text{BFGS}_0$ method is 49, for DBFGS, this number is 36, for BFGS, this number is 24, for F21, this number is 34. From this point of view, SF21 is favorably more stable method.

It is observed that SF21 update formula can be employed within the framework of the two-step strategy without increase in the linear algebra cost. This means that SF21 can be built very inexpensively and allows to obtain an overall computational saving.

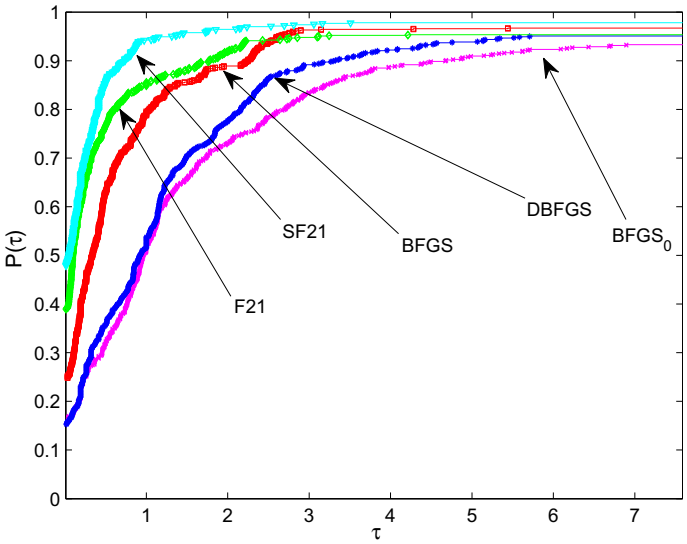


Figure 1. performance profiles based on function and gradient evaluations

## 5. Conclusion

This paper has been devoted to the extension of scaling techniques to the two-step quasi-Newton method. The basic idea of the scaling factor, considered in this paper, is due to Al-Baali [1], which has been employed on the context of extra updates for BFGS method to improve the direct Hessian approximation. In the frame of the two-step quasi-Newton method, the scaling factor has been set to be the maximum of the multiplier of Spedicato and one. This provides a preconditioner that is capable to improve the eigenvalue quantities of the two-step update at current step. Therefore, this scheme of scaling has been performed at the iterations in which the specified criterion is satisfied. This strategy suggests great reduction in the overall iteration cost of scaled two-step algorithm.

The major advantage of our method, SF21, is that it is able to upgrade the efficiency properties such as number of iterations and number of function-gradient calls without causing significant additional storage demand and no additional considerable computational cost, with respect to the ordinary two-step methods. Our experiments showed that the scaled two-step method is superior to the BFGS, DBFGS and F21 methods.

## References

[1] Al-Baali, M.: Extra updates for the BFGS method. Optim. Method. Softw. **13**, 159–179 (2000)

[2] Al-Baali, M.: Global and superlinear convergence of a restricted class of self-scaling methods with inexact line search, for convex functions. Comput. Optim. Appl. **9**, 191–203 (1998)

[3] Al-Baali, M.: New initial Hessian approximations for the limited memory BFGS method for large scale optimization. J. Fac. Sci., UAE University **1**(4), 167–175 (1995)

[4] Al-Baali, M.: Variational quasi-Newton methods for unconstrained optimization. J. Optim. Theory Appl. **7**(l), 127–143 (1993)

[5] Al-Baali, M., Conforti, D., Musmanno, R.: Computational experiments with scaled initial Hessian approximation for the Broyden family methods. Optim. A. Oper. Res. **48**, 375–389 (2000)

[6] Al-Baali, M., Grandinetti, L.: Improved damped quasi-Newton methods for unconstrained optimization. In: Fourth Asian conference on nonlinear analysis and optimization, (NAO-Asia), Taipei, Taiwan, Aug 5–9 (2014)

[7] Al-Baali, M., Khalfan, H.F.: A combined class of self-scaling and modified quasi-Newton methods. Comput. Optim. Appl. **52**(2), 393–408 (2012)

[8] Andrei, N.: An unconstrained optimization test functions collection. Adv. Model. Optim. **10**, 147–161 (2008)

[9] Byrd, R.H., Liu, D.C., Nocedal, J.: On the behaviour of Broyden's class of quasi-Newton methods. SIAM J. Optim. **2**, 533–557 (1992)

[10] Byrd, R.H., Nocedal, J., Yuan, Y.: Global convergence of a class of quasi-Newton methods on convex problems. SIAM J. Numer. Anal. **24**, 1171–1190 (1987)

[11] Dennis, J.E., Wolkowicz, H.: Sizing and least change secant methods. Technical Report, Department of Mathematical Sciences, Rice University (Houston,TX) (1991)

[12] Dolan, E.D., Moré, J.J.: Benchmarking optimization software with performance profiles. Math. Program. **91**(2), 201–203 (2002)

[13] Fletcher, R.: Practical Methods of Optimization, 2nd edn. Wiley, NewYork (1987)

[14] Ford, J.A.: Implicit updates in multistep quasi-Newton methods. Comput. Math. Appl. **42**, 1083–1091 (2001)

[15] Ford, J.A., Moghrabi, I.A.: Alternating multi-step quasi-Newton methods for unconstrained optimization. J. Compt. Appl. Math. **82**, 105–116 (1997)

[16] Ford, J.A., Moghrabi, I.A.: Multi-step quasi-Newton methods for optimization. J. Comput. Appl. Math. **50**, 305–323 (1994)

[17] Ford, J.A., Moghrabi, I.A.: Alternative parameter choices for multi-step quasi-Newton methods. Optim. Methods Softw. **2**, 357–370 (1993)

[18] Ford, J.A., Moghrabi, I.A.: Further investigation of multi-step quasi-Newton methods. Sci. Iran. **1**, 327–334 (1995)

[19] Foroutan, M., Ebadian, A., Biglari, F.: Existence and approximate solutions of nonlinear boundary value problems on the half line. Int. J. Numer. Meth. Heat Fluid Flow (2017) **(submitted)**

[20] Nocedal, J., Wright, S.J.: Springer Series in Operations Research. Numerical Optimization, 2nd edn. Springer, New York (2006)

[21] Oren, S.S.: On the selection of parameters in self-scaling variable metric algorithms. Math. Program. **3**, 351–367 (1974)

[22] Oren, S.S., Luenberger, D.G.: Self-scaling variable metric algorithm. Part I. Manag. Sci. **20**, 845–862 (1974)

[23] Oren, S.S., Spedicato, E.: Optimal conditioning of self-scaling variable metric algorithms. Math. Program. **10**, 70–90 (1976)

[24] Pearson, J.D.: Variable metric methods of minimization. Comput. J. **12**, 171–178 (1969)

[25] Powell, M.J.D.: How bad are the BFGS and DFP methods when the objective function is quadratic? Math. Program. **34**, 34–47 (1986)

[26] Shanno, D.F.: Conditioning of quasi-Newton methods for function minimization. Math. Comput. **24**, 647–656 (1970)

[27] Shanno, D.F.: Conditioning of quasi-Newton methods for function minimization. Math. Comput. **24**, 647–656 (1970)

[28] Shanno, D.F., Phua, K.H.: Matrix conditioning and nonlinear optimization. Math. Program. **14**, 149–160 (1978)

[29] Spedicato, E.: Computational experience with quasi-Newton algorithms for minimization problems of moderately large size. Report CISE-N-175, CISE Documentation Service, Segrate (Milano) (1975)

[30] Spedicato, E.: Stability of Huang's update for the conjugate gradient method. J. Optim. Theory Appl. **11**, 469–479 (1973)

Fahimeh Biglari
Department of Mathematics
Faculty of Science
Urmia University of Technology
Urmia
Iran
e-mail: beiglari.f@gmail.com

Ali Ebadian and Mohammadreza Foroutan
Department of Mathematics
Payame Noor University
Tehran P.O. Box 19395-3697
Iran
e-mail: ebadian.ali@gmail.com

Mohammadreza Foroutan
e-mail: foroutan_mohammadreza@yahoo.com

Received: July 16, 2017.

Revised: December 7, 2017.

Accepted: December 21, 2017.