A Double Parameter Scaled Modified Broyden-Fletcher-Goldfarb-Shanno Method for Unconstrained Optimization

Neculai ANDREI[1,2]

[1]Research Institute for Informatics, Center for Advanced Modeling and Optimization,
8-10 Averescu Avenue, Bucharest 1, Romania

[2]Academy of Romanian Scientists, 54, Splaiul Independenţei, Bucharest 5, Romania
nandrei@ici.ro

Abstract: In this paper the first two terms of the modified BFGS method given by Yuan and Wei [Comput. Optim. Appl., 47:237-255 (2010)] are scaled with a positive parameter, while the third one is scaled with another positive parameter. The first parameter is determined to cluster the eigenvalues of the modified BFGS matrix. The second one is computed as a preconditioner to the Hessian of the minimizing function combined with minimization of the conjugacy condition from the conjugate gradient methods in order to shift the large eigenvalues to the left. In this method the stepsize is determined by the Wolfe line search conditions. The global convergence is proved in very general conditions, without assuming the convexity of the minimizing function, using only the trace and the determinant of the scaled modified BFGS matrix. The preliminary computational experiments on a set of 80 unconstrained optimization test functions with a medium number of variables show that this algorithm is more efficient and more robust that the Yuan and Wei's modified BFGS update, as well as some other scaled modified BFGS methods we present in this paper, including the double parameter scaled BFGS method by Andrei [Jour. Comput. and App. Math. 332:26-44 (2018)].

Keywords: Unconstrained optimization, Modified BFGS method, Scaled BFGS method, Trace, Determinant, Global convergence, Numerical comparisons.

Mathematics Subject Classification (2010) 49M7. 49M10. 65K05. 90C30

1. Introduction

For solving the unconstrained optimization problem:

$$\min f(x), \tag{1.1}$$

where $f: R^n \to R$ is a twice continuous differentiable function, the well known quasi-Newton BFGS method introduced by Broyden [7], Fletcher [12], Goldfarb [13] and Shanno [23] run as follows. Starting with an initial point $x_0 \in R^n$ and an initial approximation $B_0 \in R^{n\times n}$ to the Hessian of function f, symmetric and positive definite, this method generates the sequence $\{x_k\}$:

$$x_{k+1} = x_k + \alpha_k d_k, \tag{1.2}$$

$k = 0,1,...$, where $d_k \in R^n$ is the BFGS search direction computed as solution of the linear algebraic system:

$$B_k d_k = -g_k, \tag{1.3}$$

and g_k is the gradient $\nabla f(x_k)$ of f at x_k. The matrix B_k in (1.3), known as the BFGS approximation to the Hessian $\nabla^2 f(x_k)$ of f at x_k, is computed by the classical formula:

$$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{y_k^T s_k}, \tag{1.4}$$

$k = 0,1,...$, where $s_k = x_{k+1} - x_k$ and $y_k = g_{k+1} - g_k$. Observe that the first two terms in (1.4) depend by s_k and the approximation B_k. Only the third one is dependent by y_k. An important property of the BFGS formula (1.4), which is called the *standard BFGS updating*, is that B_{k+1} inherits the positive definiteness of B_k if $y_k^T s_k > 0$. The condition $y_k^T s_k > 0$ holds if the stepsize α_k in (1.2) is determined by the Wolfe line search conditions [26, 27]:

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

$$g(x_k + \alpha_k d_k)^T d_k \ge \rho g(x_k)^T d_k, \tag{1.6}$$

where the positive constants σ and ρ satisfy $0 < \sigma < \rho < 1$. Since B_k is positive definite, the search direction d_k generated by (1.3) is a descent direction of f at x_k, no matter whether the Hessian is positive definite or not.

To improve the numerical behavior and to accelerate the convergence of the standard BFGS method, *the modified BFGS* and *the scaled BFGS methods* have been introduced. The modified BFGS methods are based on modifying the vector y_k in such a way to improve the approximation of the Hessian of the minimizing function

https://doi.org/10.24846/v27i2y201801

and to establish the global and the superlinear convergence for general convex functions. The scaled BFGS methods, on the other hand, use the scaling the terms of the BFGS updating formula (1.4). Both these variants of the BFGS method generate efficient algorithms able to solve a large diversity of unconstrained optimization problems of different dimensions and complexities.

In [4] a double parameter scaled BFGS method for unconstrained optimization has been presented, where the first two terms of the standard BFGS update (1.4) are scaled by a positive parameter selected in such a way to cluster the eigenvalues of B_{k+1} while the third one is scaled by another positive parameter chosen to reduce the large eigenvalues of B_{k+1}. In this paper we consider a *combination* of the modified BFGS methods and of the scaled BFGS updates. In other words, our approach is as follows: firstly the standard BFGS method is modified by modifying the vector y_k, and secondly this modified BFGS update is scaled. This is motivated by the fact that the modified BFGS method (with vector y_k modified) is more efficient and more robust versus the well known standard BFGS update (see for example [14, 15, 25, 29 and 30]). Therefore, we are interested to see whether scaling the modified BFGS methods, i.e. scaling the BFGS with vector y_k modified, lead us to more efficient algorithms. Section 2 presents the main modified BFGS methods where the vector y_k is modified in different ways. To be self-contained, the main scaled BFGS updates are presented in Section 3 of this paper (see also [4]). Historically it has been noticed that the BFGS update tends to produce updates with large eigenvalues. Therefore, in Section 4 we develop a double parameter scaled variant of the Yuan and Wei [30] modified BFGS update, which improves the structure of eigenvalues. To cluster the eigenvalues, the first two terms of the modified BFGS update are scaled with a positive parameter. On the other hand, to shift the large eigenvalues to the left, the third term is also scaled with another positive parameter. In Section 5 the global convergence of this variant of double parameter scaled modified BFGS method is proved in vey general conditions without assuming the convexity of the minimizing function and using only the trace and the determinant of the double parameter scaled modified BFGS matrix. Finally, in Section 6 intensive numerical results and comparisons for a collection of 80 medium size (100 variables) unconstrained optimization problems of different complexities taken from [1] are presented. We have the computational evidence that our double parameter scaled modified BFGS method is more efficient and more robust versus some other scaled BFGS methods.

2. Some modified BFGS methods

2.1 Li and Fukushima BFGS updating (Li and Fukushima [14]). In order to obtain the global convergence of the BFGS method for nonconvex minimizing functions, Li and Fukushima [14] introduced a slight modification of the standard BFGS update. In [14] the modified BFGS is computed as:

$$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{z_k z_k^T}{s_k^T z_k}, \tag{2.1}$$

where

$$z_k = y_k + (\max\{0, -y_k^T s_k / \|s_k\|^2\} + \phi(\|g_k\|))s_k$$

and function $\phi: R \to R$ satisfies the following conditions: (i) $\phi(t) > 0$ for all $t > 0$; (ii) $\phi(t) = 0$ if and only if $t = 0$; (iii) $\phi(t)$ is bounded if t is in a bounded set. Observe that by definition of z_k we have that $s_k^T z_k \geq \max\{s_k^T y_k, \phi(\|g_k\|)\|s_k\|^2\} > 0$. This is sufficient to guarantee the positive definiteness of B_{k+1} as long as B_k is positive definite. Therefore (2.1) is well defined. Li and Fukushima consider $\phi(t) = \mu t$, where $\mu > 0$ is a constant.

2.2 Another Li and Fukushima BFGS updating (Li and Fukushima [15]). This modified BFGS, introduced in [15], is defined as:

$$B_{k+1} = \begin{cases} B_k - \dfrac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \dfrac{z_k z_k^T}{s_k^T z_k}, & \text{if } \dfrac{s_k^T z_k}{\|s_k\|^2} \geq \phi(\|g_k\|), \\ B_k, & \text{otherwise,} \end{cases} \tag{2.2}$$

where z_k and ϕ are the same as in the first updating formula of Li and Fukushima. For this modified BFGS formula $s_k^T z_k > 0$ is satisfied, and therefore B_{k+1} inherits the positive definiteness of B_k along the iterations. For these two modified BFGS updating formulas, Li and Fukushima proved their global convergence and the superlinear convergence.

2.3 Wei, Yu, Yuan and Lian BFGS updating (Wei, Yu, Yuan and Lian [25]). In order to get a better approximation of the Hessian of the minimizing function, based on the results given by Li, Tang and Wei [16], Wei, Yu, Yuan and Lian [25] proposed a modified BFGS method which contains not only the gradient value information but also the function values at the current and the previous steps. Their updating formula is defined as follows:

$$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{\tilde{y}_k \tilde{y}_k^T}{s_k^T \tilde{y}_k}, \tag{2.3}$$

where

$$\tilde{y}_k = y_k + \frac{\rho_k}{\|s_k\|^2} s_k, \tag{2.4}$$

$$\rho_k = 2[f(x_k) - f(x_k + \alpha_k d_k)] + (g(x_k + \alpha_k d_k) + g(x_k))^T s_k. \tag{2.5}$$

The numerical experiments show that this modified BFGS method is more efficient that the standard BFGS method (see [25]).

2.4 Yuan and Wei BFGS updating (Yuan and Wei [29, 30]). On the other hand, Yuan and Wei [30] suggested another modified BFGS method for which the global and the superlinear convergence have been established. The corresponding updating formula is:

$$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{\overline{y}_k \overline{y}_k^T}{s_k^T \overline{y}_k}, \tag{2.6}$$

where

$$\overline{y}_k = y_k + \frac{\max\{\rho_k, 0\}}{\|s_k\|^2} s_k, \tag{2.7}$$

and ρ_k is given by (2.5). This modification can ensure that B_{k+1} inherits the positive definiteness of B_k for general convex functions.

3. Some scaled BFGS methods

For the paper to be self-contained, this section presents some scaled BFGS methods taken from [4]. A scaled BFGS update with a multitude of variants was given by:

$$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \gamma_k \frac{y_k y_k^T}{y_k^T s_k}, \tag{3.1}$$

where $\gamma_k > 0$ is the scaling parameter. For the scaling parameter γ_k in (3.1) some values have been proposed in literature, as follows (see [4]).

3.1 Scaled BFGS with different interpolation conditions (Biggs [5, 6] and Yuan [28]).

The value of the scaling parameter γ_k proposed by Biggs [5, 6] is:

$$\gamma_k = \frac{6}{y_k^T s_k}(f(x_k) - f(x_{k+1}) + s_k^T g_{k+1}) - 2. \tag{3.2}$$

In the same line of research, Yuan [28] suggested the following value for the scaling parameter

$$\gamma_k = \frac{2}{y_k^T s_k}(f(x_k) - f(x_{k+1}) + s_k^T g_{k+1}). \tag{3.3}$$

For general nonlinear functions, the inexact line search does not involve the positivity of γ_k. In these cases Yuan restricted γ_k in the interval $[0.01,100]$ and proved the global convergence of this variant of the scaled BFGS method.

3.2 Spectral scaled BFGS (Cheng and Li [10]). Another scaled BFGS method was introduced by Cheng and Li [10]. In this update the scaling parameter γ_k in (3.1) is computed as

$$\gamma_k = \frac{y_k^T s_k}{\|y_k\|^2}, \tag{3.4}$$

obtained as solution of the problem:

$$\min \|s_k - \gamma_k y_k\|^2.$$

3.3 Scaled BFGS with diagonal preconditioning and conjugacy condition (Andrei [2]). Andrei [2] introduced another scaled BFGS update given by (3.1), in which the scaling parameter γ_k is computed in an adaptive manner as:

$$\gamma_k = \min\left\{\frac{y_k^T s_k}{\|y_k\|^2 + \beta_k}, 1\right\}, \tag{3.5}$$

where $\beta_k > 0$ for all $k = 0,1,\ldots$. Intensive numerical experiments showed that this scaled BFGS algorithm with $\beta_k = |s_k^T g_{k+1}|$ is the best one, being more efficient and more robust versus the standard BFGS algorithm as well as versus some other scaled BFGS algorithms.

3.4 Scaling the first two terms of the BFGS update with a parameter (Oren and Luenberger [20] and Nocedal and Yuan [19]). This scaled BFGS update is different from (3.1) and is defined as:

$$B_{k+1} = \delta_k \left[B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} \right] + \frac{y_k y_k^T}{y_k^T s_k}, \tag{3.6}$$

where δ_k is a positive parameter. Concerning the selection of δ_k in (3.6) Oren and Luenberger [20] suggested $\delta_k = y_k^T s_k / s_k^T B_k s_k$ being one of the best, as it simplifies the analysis.

3.5 A double parameter scaled BFGS (Andrei [4]). Motivated by the idea of changing the structure of the eigenvalues of the BFGS approximation to the Hessian matrix (1.4), a *double parameter scaled BFGS method* has been suggested by Andrei [4], In this method the updating of the approximation Hessian matrix B_{k+1} is computed as:

$$B_{k+1} = \delta_k \left[B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} \right] + \gamma_k \frac{y_k y_k^T}{y_k^T s_k}, \tag{3.7}$$

where δ_k and γ_k are positive parameters with the following values:

$$\gamma_k = \min\left\{ \frac{y_k^T s_k}{\|y_k\|^2 + \left| s_k^T g_{k+1} \right|}, 1 \right\}, \tag{3.8}$$

$$\delta_k = \frac{n - \gamma_k \dfrac{\|y_k\|^2}{y_k^T s_k}}{n - \dfrac{\|B_k s_k\|^2}{s_k^T B_k s_k}}, \tag{3.9}$$

where γ_k is given by (3.8).

3.6 Scaling the last terms of the BFGS update with two positive parameters (Liao [17]). In another avenue of research, Liao [17] introduced the two parameter scaled BFGS method:

$$B_{k+1} = B_k - \delta_k \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \gamma_k \frac{y_k y_k^T}{y_k^T s_k} \tag{3.10}$$

where the parameters scaling the terms in the BFGS update are computed in an adaptive way subject to the values of a positive parameter τ_k as:

$$(\delta_k, \gamma_k) = \begin{cases} \left(\dfrac{s_k^T B_k s_k}{s_k^T B_k s_k + y_k^T s_k}, \dfrac{y_k^T s_k}{s_k^T B_k s_k + y_k^T s_k} \right), \\ (\tau_k, 1), \end{cases} \quad \begin{array}{c} \text{if } \dfrac{s_k^T B_k s_k}{s_k^T B_k s_k + y_k^T s_k} \geq \tau_k, \\ \text{otherwise,} \end{array} \tag{3.11}$$

where $0 < \tau_k < 1$.

4. A double parameter scaled modified BFGS method

There are many possibilities to define a scaled modified BFGS method. The idea is to scale the modified BFGS methods presented in Section 2 using the scaling procedures described in Section 3. In this paper we consider a scaling of the Yuan and Wei [30] modified BFGS given by (2.6) and (2.7) using the double parameter scaling BFGS procedure by Andrei [4] defined by (3.7)-(3.9). Concerning the modified BFGS method (2.6), two important tools in the analysis of its properties and of its convergence are the trace and the determinant. The trace of a matrix is the sum of its eigenvalues. The determinant of a matrix is the product of its eigenvalues. By direct computation from (2.6) we obtain:

$$tr(B_{k+1}) = tr(B_k) - \frac{\|B_k s_k\|^2}{s_k^T B_k s_k} + \frac{\|\bar{y}_k\|^2}{\bar{y}_k^T s_k}. \tag{4.1}$$

On the other hand, as in [4] we have:

$$\det(B_{k+1}) = \det(B_k) \frac{\bar{y}_k^T s_k}{s_k^T B_k s_k}. \tag{4.2}$$

In practical implementation the search direction is computed as

$$d_{k+1} = -H_{k+1} g_{k+1}, \tag{4.3}$$

where H_{k+1} is the BFGS approximation to the inverse of B_{k+1} given by (2.6). With a little algebra, using the rank-one Sherman-Morrison-Woodbury formula twice, from (2.6) we get:

$$\begin{aligned} H_{k+1} = H_k &- \frac{H_k \bar{y}_k s_k^T + s_k \bar{y}_k^T H_k}{\bar{y}_k^T s_k} \\ &+ \left(1 + \frac{\bar{y}_k^T H_k \bar{y}_k}{\bar{y}_k^T s_k} \right) \frac{s_k s_k^T}{\bar{y}_k^T s_k}. \end{aligned} \tag{4.4}$$

The study given in [3] emphasized that the efficiency of the BFGS method is strongly dependent on the structure of the eigenvalues of the approximation to the Hessian matrix. Powell [21] and Byrd, Liu and Nocedal [8] point out that the BFGS method suffers more from the large eigenvalues than from the small ones, i.e. the BFGS update tends to produce updates with large eigenvalues. Now, we see that the second term on the right hand side of (4.1) being negative,

it produces a shift of the eigenvalues of the modified B_{k+1} given by (2.6) to the left, and this is independent whether the BFGS update is modified or not. Therefore, by properly scaling the second term on the right hand side of (2.6), the scaled modified BFGS method is able to correct large eigenvalues. On the other hand, the third term on the right hand side of (4.1) being positive it produces a shift of the eigenvalues of the modified B_{k+1} to the right. But,

$$\|\bar{y}_k\|^2 = \|y_k\|^2 + \frac{2\max\{\rho_k,0)}{\|s_k\|^2} y_k^T s_k + \frac{(\max\{\rho_k,0))^2}{\|s_k\|^2},$$

and

$$\bar{y}_k^T s_k = y_k^T s_k + \max\{\rho_k,0\} \geq y_k^T s_k.$$

Therefore, the third term on the right hand side of (4.1) can be bounded as:

$$\frac{\|\bar{y}_k\|^2}{s_k^T \bar{y}_k} \leq \frac{\|y_k\|^2}{s_k^T y_k} + \frac{\max\{\rho_k,0\}}{\|s_k\|^2}\left(2 + \frac{\max\{\rho_k,0\}}{\|s_k\|^2 s_k^T y_k}\right).$$

For convex functions, for all k, $\|y_k\|^2 / s_k^T y_k \leq M$, where M is a positive constant [18]. If the above bound is large, then the modified update B_{k+1} (2.6) may have large eigenvalues. Therefore, a correction of the structure of the eigenvalues of the modified B_{k+1} (2.6) by Yuan and Wei [30] can be achieved by scaling the corresponding terms in (2.6) and this is the main motivation for which we introduce the *scaled modified BFGS methods*. Therefore, in order to change the structure of the eigenvalues of the modified BFGS approximation to the Hessian matrix given by (2.6) in this section we propose a *double parameter scaled modified BFGS method* in which the updating of the approximation Hessian matrix B_{k+1} is computed as:

$$B_{k+1} = \delta_k\left[B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k}\right] + \gamma_k \frac{\bar{y}_k \bar{y}_k^T}{s_k^T \bar{y}_k}, \qquad (4.5)$$

where δ_k and γ_k are positive parameters and $\bar{y}_k$ is determined as in (2.7). Observe that this scaled modified BFGS update contains both the gradient value information and the function values in two successive points. In our scaled modified BFGS method given by (4.5) the parameter δ_k is selected to cluster the eigenvalues of B_{k+1} and γ_k is determined to reduce the large eigenvalues of B_{k+1}, thus hoping to obtain a better distribution of the eigenvalues.

Using the rank-one Sherman-Morrison-Woodbury update formula twice, from (4.5) we get $H_{k+1} = B_{k+1}^{-1}$, where

$$H_{k+1} = \frac{1}{\delta_k}\left[H_k - \frac{H_k \bar{y}_k s_k^T + s_k \bar{y}_k^T H_k}{\bar{y}_k^T s_k} + \left(\frac{\delta_k}{\gamma_k} + \frac{\bar{y}_k^T H_k \bar{y}_k}{\bar{y}_k^T s_k}\right)\frac{s_k s_k^T}{\bar{y}_k^T s_k}\right]. \qquad (4.6)$$

is the approximation to the inverse Hessian. In our algorithm the stepsize α_k in (1.2) is determined by the Wolfe line search (1.5) and (1.6). Therefore $s_k^T y_k > 0$. If $\rho_k > 0$, then we have

$$s_k^T \bar{y}_k = s_k^T\left(y_k + \frac{\rho_k}{\|s_k\|^2} s_k\right) = s_k^T y_k + \rho_k > s_k^T y_k > 0.$$

Proposition 4.1. *If the stepsize α_k is determined by the Wolfe line search (1.5) and (1.6), B_k is positive definite and $\gamma_k > 0$, then B_{k+1} given by (4.5) is also positive definite.*

Proof Using the symmetry and the positivity of B_k, we have

$$(s_k^T B_k z)^2 \leq (s_k^T B_k s_k)(z^T B_k z),$$

with equality if $z = 0$ or $s_k = 0$. On the other hand, by the Wolfe line search (1.5) and (1.6) we have that $y_k^T s_k > 0$. Therefore, using the above inequality we get:

$$z^T B_{k+1} z = \delta_k z^T B_k z - \delta_k \frac{z^T B_k s_k s_k^T B_k z}{s_k^T B_k s_k} + \gamma_k \frac{z^T \bar{y}_k \bar{y}_k^T z}{\bar{y}_k^T s_k} = \delta_k z^T B_k z - \delta_k \frac{(z^T B_k s_k)^2}{s_k^T B_k s_k} + \gamma_k \frac{(z^T \bar{y}_k)^2}{\bar{y}_k^T s_k} \geq \gamma_k \frac{(z^T \bar{y}_k)^2}{\bar{y}_k^T s_k} > 0,$$

for any nonzero z. ■

The corresponding scaled BFGS algorithm can be presented as follows.

The scaled modified BFGS algorithm – SMBFGS

1. Initialization. Choose an initial point $x_0 \in R^n$ and an initial positive definite matrix H_0. Choose the constants σ, ρ with $0 < \sigma < \rho < 1$, and $\varepsilon > 0$ sufficiently small. Compute $g_0 = \nabla f(x_0)$. Set $d_0 = -g_0$. Set $k = 0$.
2. Test a criterion for stopping the iterations. For example, if $\|g_k\| \le \varepsilon$, then stop the iterations. Otherwise, continue with step 3.
3. Compute the stepsize $\alpha_k > 0$ satisfying the Wolfe line search conditions (1.5) and (1.6).
4. Compute $x_{k+1} = x_k + \alpha_k d_k$, $f_{k+1} = f(x_{k+1})$ and $g_{k+1} = \nabla f(x_{k+1})$. Set $s_k = x_{k+1} - x_k$, $y_k = g_{k+1} - g_k$.
5. Compute the scaling factors δ_k and γ_k.
6. Compute $\bar{y}_k$ as in (2.7).
7. Update the inverse Hessian H_k using (4.6).
8. Compute the search direction as $d_{k+1} = -H_{k+1} g_{k+1}$.
9. Set $k = k + 1$ and continue with step 2. ■

If $\delta_k = 1$, $\gamma_k = 1$ and $\bar{y}_k = y_k$ for all $k = 0,1,...$, then the above algorithm is exactly the standard BFGS algorithm. *For different values of the parameters δ_k and γ_k in (4.5) (or (4.6)), different scaled modified BFGS algorithms are obtained.* Therefore, in order to implement the SMBFGS algorithm, some procedures for δ_k and γ_k in step 5 must be defined. A variant of SMBFGS, we consider in this paper, inspired from [4], is as follows. Since the scaled BFGS with diagonal preconditioning and conjugacy condition is the best one (see [2]), in our scaled modified BFGS algorithm γ_k is computed as:

$$\gamma_k = \min\left\{\frac{\bar{y}_k^T s_k}{\|\bar{y}_k\|^2 + |s_k^T g_{k+1}|}, 1\right\}, \tag{4.7}$$

for all $k = 0,1,\ldots$. Concerning the parameter δ_k, the idea taken from the linear conjugate gradient methods (see also [4]), is to select it in such a way that the eigenvalues of B_{k+1} to be clustered. Since the trace of a matrix is the sum of its eigenvalues, the parameter δ_k is selected in such a way that the trace of B_{k+1} given by (4.5) to be equal to n. Therefore, considering the trace of B_{k+1} given by (4.5), from the equation $tr(B_{k+1}) = n$ we obtain:

$$\delta_k = \frac{n - \gamma_k \dfrac{\|\bar{y}_k\|^2}{\bar{y}_k^T s_k}}{n - \dfrac{\|B_k s_k\|^2}{s_k^T B_k s_k}}, \tag{4.8}$$

where γ_k is given by (4.7).

Proposition 4.2. *Let δ_k be computed as in (4.8). Then, for any $k = 0,1,\ldots$, δ_k is positive and close to 1.*

Proof Observe that along the iterations $|s_k^T g_{k+1}| \to 0$. Hence, $\|\bar{y}_k\|^2 / (\|\bar{y}_k\|^2 + |s_k^T g_{k+1}|)$ is close to 1. On the other hand, B_k is symmetric and positive definite. Therefore, it has real and positive eigenvalues: $\lambda_1,\ldots,\lambda_n$. Since B_k is nonsingular and $tr(B_k) = n$, it follows that for any $i = 1,\ldots,n$, $\lambda_i > 0$ such that $\sum_{i=1}^n \lambda_i = n$. But $\|B_0 s_0\|^2 = s_0^T B_0 s_0$. For k sufficiently large, $0 < \|B_k s_k\|^2 < 1$ and $0 < s_k^T B_k s_k < 1$. Since $\|B_k s_k\|^2$ and $s_k^T B_k s_k$ are approximately of the same order of magnitude, it follows that $n >> \|B_k s_k\|^2 / s_k^T B_k s_k$. Therefore, we have $n >> \gamma_k \|\bar{y}_k\|^2 / \bar{y}_k^T s_k$ and $n > \|B_k s_k\|^2 / s_k^T B_k s_k$, i.e. for any $k = 0,1,\ldots$, δ_k is positive and close to 1. Observe that the bigger n is, the closer to 1 δ_k is. ■

5. Global convergence of SMBFGS

The convergence analysis of SMBFGS is based on the same principles as those presented by Andrei [4] (see also Li and Fukushima [14] and Byrd and Nocedal [9]).

Proposition 5.1. *Let δ_k be computed as in (4.8) for $k = 0,1,\ldots$. Then, there are the positive constants $0 < \delta < \Delta$ such that for any $j = 0,1,\ldots,k$,*

$$\delta < \delta_k \delta_{k-1} \cdots \delta_j < \Delta. \tag{5.1}$$

Proof From Proposition 4.2 it follows that δ_k is close to 1 for any $k = 0,1,\ldots$. As a consequence, there are the positive constants $0 < \delta < \Delta$ such that any product of the form $\delta_k \delta_{k-1} \cdots \delta_j$, for any $j = 0,1,\ldots$, is bounded as in (5.1). ■

Proposition 5.2. *Consider the scaled modified* B_{k+1} *given by (4.5), where* γ_k *and* δ_k *are computed as in (4.7) and (4.8), respectively. Then*

$$tr(B_{k+1}) \le \Delta tr(B_0) + (\Delta k + 1) \tag{5.2}$$

and

$$\sum_{i=0}^{k} \frac{\|B_i s_i\|^2}{s_i^T B_i s_i} \le \frac{\Delta}{\delta}(tr(B_0) + k) + \frac{1}{\delta}. \tag{5.3}$$

Proof Observe that

$$\begin{aligned}
tr(B_{k+1}) &= \delta_k tr(B_k) - \delta_k \frac{\|B_k s_k\|^2}{s_k^T B_k s_k} + \gamma_k \frac{\|\overline{y}_k\|^2}{\overline{y}_k^T s_k} \\
&= \delta_k \left(\delta_{k-1} tr(B_{k-1}) - \delta_{k-1} \frac{\|B_{k-1} s_{k-1}\|^2}{s_{k-1}^T B_{k-1} s_{k-1}} \right. \\
&\left. + \gamma_{k-1} \frac{\|\overline{y}_{k-1}\|^2}{\overline{y}_{k-1}^T s_{k-1}} \right) - \delta_k \frac{\|B_k s_k\|^2}{s_k^T B_k s_k} + \gamma_k \frac{\|\overline{y}_k\|^2}{\overline{y}_k^T s_k} \\
&= \ldots \\
&= \delta_k \delta_{k-1} \cdots \delta_0 tr(B_0) \\
&- \delta_k \delta_{k-1} \cdots \delta_0 \frac{\|B_0 s_0\|^2}{s_0^T B_0 s_0} + \delta_k \delta_{k-1} \cdots \delta_1 \gamma_0 \frac{\|\overline{y}_0\|^2}{\overline{y}_0^T s_0} \\
&- \delta_k \delta_{k-1} \cdots \delta_1 \frac{\|B_1 s_1\|^2}{s_1^T B_1 s_1} + \delta_k \delta_{k-1} \cdots \delta_2 \gamma_1 \frac{\|\overline{y}_1\|^2}{\overline{y}_1^T s_1} \\
&\ldots \\
&- \delta_k \delta_{k-1} \frac{\|B_{k-1} s_{k-1}\|^2}{s_{k-1}^T B_{k-1} s_{k-1}} + \delta_k \gamma_{k-1} \frac{\|\overline{y}_{k-1}\|^2}{\overline{y}_{k-1}^T s_{k-1}} \\
&- \delta_k \frac{\|B_k s_k\|^2}{s_k^T B_k s_k} + \gamma_k \frac{\|\overline{y}_k\|^2}{\overline{y}_k^T s_k}.
\end{aligned} \tag{5.4}$$

But, from (4.7), for any $i = 0,\ldots,k$,

$$\begin{aligned}
\gamma_i \frac{\|\overline{y}_i\|^2}{\overline{y}_i^T s_i} &= \frac{\overline{y}_i^T s_i}{\|\overline{y}_i\|^2 + |s_i^T g_{i+1}|} \frac{\|\overline{y}_i\|^2}{\overline{y}_i^T s_i} \\
&= \frac{\|\overline{y}_i\|^2}{\|\overline{y}_i\|^2 + |s_i^T g_{i+1}|} \le 1.
\end{aligned}$$

Therefore, since by Proposition 5.1 there are the positive constants $0 < \delta < \Delta$ such that for any $j = 0,1,\ldots,k$, $\delta < \delta_k \delta_{k-1} \cdots \delta_j < \Delta$, it follows that

$$\begin{aligned}
tr(B_{k+1}) &\le \Delta tr(B_0) - \sum_{i=0}^{k} \delta \frac{\|B_i s_i\|^2}{s_i^T B_i s_i} \\
&+ \sum_{k=1}^{k} \Delta + 1 \le \Delta tr(B_0) + \Delta k + 1.
\end{aligned} \tag{5.5}$$

From (5.5) we get (5.2).

Since B_{k+1} is positive definite, $tr(B_{k+1}) > 0$. Therefore (5.3) is true. ■

Proposition 5.3. *If for all* k, $\gamma_k \ge m$, *where* $m > 0$ *is a constant, and* $\delta_k \ge \theta$, *where* $\theta > 0$ *is a constant, then there is a constant* $c > 0$ *such that for all* k *sufficiently large:*

$$\prod_{i=0}^{k} \alpha_i \ge c^k. \tag{5.6}$$

Proof The determinant of the scaled modified B_{k+1} given by (4.5) is as follows:

$$\begin{aligned}
\det(B_{k+1}) &= \det\left(\delta_k B_k \left(I - \frac{s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{\gamma_k}{\delta_k} \frac{B_k^{-1} \overline{y}_k \overline{y}_k^T}{\overline{y}_k^T s_k} \right) \right) \\
&= \det(\delta_k B_k) \det\left(I - s_k \frac{(B_k s_k)^T}{s_k^T B_k s_k} + \frac{\gamma_k}{\delta_k} (B_k^{-1} \overline{y}_k) \frac{\overline{y}_k^T}{\overline{y}_k^T s_k} \right) \\
&= \delta_k^n \det(B_k) \frac{\gamma_k}{\delta_k} \frac{\overline{y}_k^T s_k}{s_k^T B_k s_k}.
\end{aligned} \tag{5.7}$$

Therefore,

$$\begin{aligned}
\det(B_{k+1}) &= \delta_k^{n-1} \gamma_k \frac{\overline{y}_k^T s_k}{s_k^T B_k s_k} \det(B_k) \\
&= \left(\delta_k^{n-1} \gamma_k \frac{\overline{y}_k^T s_k}{s_k^T B_k s_k} \right) \left(\delta_{k-1}^{n-1} \gamma_{k-1} \frac{\overline{y}_{k-1}^T s_{k-1}}{s_{k-1}^T B_{k-1} s_{k-1}} \right) \det(B_{k-1}) \\
&= \left(\prod_{i=0}^{k} \delta_i^{n-1} \gamma_i \frac{\overline{y}_i^T s_i}{s_i^T B_i s_i} \right) \det(B_0).
\end{aligned} \tag{5.8}$$

But, for all i, $s_i^T B_i s_i \le -\alpha_i s_i^T g_i$ and $\overline{y}_i^T s_i \ge -(1-\rho) s_i^T g_i + \rho_i \ge -(1-\rho) s_i^T g_i$. Besides, for all i, $\gamma_i \ge m$ and $\delta_i \ge \theta$. Therefore,

$$\begin{aligned}
\det(B_{k+1}) &\ge \det(B_0) \prod_{i=0}^{k} \theta^{n-1} m \frac{1-\rho}{\alpha_i} \\
&= \det(B_0) \left(\theta^{n-1} \right)^{k+1} m^{k+1} (1-\rho)^{k+1} \prod_{i=0}^{k} \frac{1}{\alpha_i}.
\end{aligned} \tag{5.9}$$

Since $\det(B_{k+1}) \le \left(\frac{1}{n} tr(B_{k+1})\right)^n$, by using Proposition 5.2, we get

$$\det(B_{k+1}) \le \left(\frac{1}{n}\left(\Delta tr(B_0) + \Delta k + 1\right)\right)^n.$$

Therefore,

$$\prod_{i=0}^{k} \alpha_i \ge \frac{\det(B_0)\theta^{(n-1)(k+1)} m^{k+1}(1-\rho)^{k+1}}{\det(B_{k+1})} \ge \frac{\det(B_0)\theta^{(n-1)(k+1)} m^{k+1}(1-\rho)^{k+1}}{\left(\frac{1}{n}\left(\Delta tr(B_0) + \Delta k + 1\right)\right)^n}. \quad (5.10)$$

When k is sufficiently large, (5.10) implies (5.6). ■

Theorem 5.1. *Let $\{x_k\}$ be generated by the algorithm SMBFGS. Then*

$$\liminf_{k\to\infty} \|g_k\| = 0. \quad (5.11)$$

Proof Assume that $\|g_k\| > \Gamma > 0$, for all k. Observe that $B_k s_k = \alpha_k B_k d_k$. Since f is bounded from below, from the first Wolfe condition (1.5) we have $\sum_{k=0}^{\infty} (-s_k^T g_k) < \infty$. Therefore,

$$\infty > \sum_{k=0}^{\infty} (-s_k^T g_k) = \sum_{k=0}^{\infty} \frac{1}{\alpha_k} s_k^T B_k s_k = \sum_{k=0}^{\infty} \frac{\|g_k\|}{\|B_k s_k\|} s_k^T B_k s_k$$

$$= \sum_{k=0}^{\infty} \frac{s_k^T B_k s_k}{\|B_k s_k\|} \|g_k\| \frac{\|B_k s_k\|}{\|B_k s_k\|} = \sum_{k=0}^{\infty} \frac{s_k^T B_k s_k}{\|B_k s_k\|} \|g_k\| \frac{\alpha_k \|g_k\|}{\|B_k s_k\|}$$

$$= \sum_{k=0}^{\infty} \|g_k\|^2 \alpha_k \frac{s_k^T B_k s_k}{\|B_k s_k\|^2} \ge \Gamma^2 \sum_{k=0}^{\infty} \alpha_k \frac{s_k^T B_k s_k}{\|B_k s_k\|^2}. \quad (5.12)$$

Now, from the geometric inequality, for any $\Omega > 0$ there exists an integer $k_0 > 0$ such that for any positive integer q we have:

$$q\left[\prod_{k=k_0+1}^{k_0+q} \alpha_k \frac{s_k^T B_k s_k}{\|B_k s_k\|^2}\right]^{1/q} \le \sum_{k=k_0+1}^{k_0+q} \alpha_k \frac{s_k^T B_k s_k}{\|B_k s_k\|^2} \le \Omega \quad (5.13)$$

Hence,

$$\left[\prod_{k=k_0+1}^{k_0+q} \alpha_k\right]^{1/q} \le \frac{\Omega}{q}\left[\prod_{k=k_0+1}^{k_0+q} \frac{\|B_k s_k\|^2}{s_k^T B_k s_k}\right]^{1/q}$$

$$\le \frac{\Omega}{q^2} \sum_{k=k_0+1}^{k_0+q} \frac{\|B_k s_k\|^2}{s_k^T B_k s_k} \le \frac{\Omega}{q^2} \sum_{k=0}^{k_0+q} \frac{\|B_k s_k\|^2}{s_k^T B_k s_k}$$

$$\le \frac{\Omega}{q^2}\left(\frac{\Delta}{\delta} tr(B_0) + \frac{\Delta}{\delta}(k_0+q) + \frac{1}{\delta}\right), \quad (5.14)$$

where the last inequality follows from Proposition 5.2. Now, considering $q \to \infty$, we get a contradiction because of Proposition 5.3 which shows that the left-hand side of the above inequality (5.14) is greater than a positive constant. Therefore, (5.11) is true. ■

6. Numerical results and comparisons

In this section we present some numerical results with a Fortran implementation of the double parameter scaled modified SMBFGS algorithms. In all these algorithms the double parameter scaled modified BFGS is computed as in (4.5). The search direction is computed as in (4.3) where H_{k+1} is updated as in (4.6). The algorithm SMBFGS is particularized as follows: SMBFGS1 (SMBFGS with $\gamma_k = 1$ and $\delta_k = 1$, i.e. the standard modified BFGS in which B_{k+1} is computed as in (2.6) and $\bar{y}_k$ is as in (2.7)); SMBFGSD (SMBFGS with γ_k given by (4.7) and δ_k given by (4.8)); SMBFGSA (SMBFGS with γ_k given by (4.7) and $\delta_k = 1$, i.e. the scaled modified BFGS method given by Andrei [2]); SMBFGSB (SMBFGS with $\gamma_k = 6(f(x_k) - f(x_{k+1}) + s_k^T g_{k+1}) / \bar{y}_k^T s_k - 2$ and $\delta_k = 1$, i.e. the scaled modified method by Biggs [5, 6]); SMBFGSC (SMBFGS with $\gamma_k = \bar{y}_k^T s_k / \|\bar{y}_k\|^2$ and $\delta_k = 1$, i.e. the scaled modified BFGS method given by Cheng and Li [10]); MNOYA (SMBFGS with $\gamma_k = 1$ and $\delta_k = \bar{y}_k^T s_k / s_k^T B_k s_k$ i.e. the scaled modified BFGS of Nocedal and Yuan [19] given by (3.6) with y_k replaced by $\bar{y}_k$); SMBFGSY (SMBFGS with $\delta_k = 1$ and $\gamma_k = 2(f(x_k) - f(x_{k+1}) + s_k^T g_{k+1}) / \bar{y}_k^T s_k$, i.e. the scaled modified BFGS method given by Yuan [28]).

All the algorithms used in these numerical experiments implement the Wolfe line search conditions with $\sigma = 0.8$ and $\rho = 0.0001$. The iterations are stopped if the inequality $\|g_k\|_\infty \le 10^{-5}$ is satisfied, where $\|.\|_\infty$ is the maximum absolute component of a vector or if the number of iterations exceeds 1000. In all the algorithms, for all the problems, the initial matrix $H_0 = I$, i.e. the identity matrix. For each scaled modified method, except the method of Liao given by (3.10) and (3.11), where y_k is replaced by $\bar{y}_k$, in order

to get the search direction we use the inverse updating formula (4.6). For the scaled modified BFGS methods by Biggs [5, 6] and Yuan [28], γ_k is restricted in the interval [0.01, 100]. Besides, at the very first iteration of these methods the scaling is not applied. All the codes were written in double precision Fortran and compiled with f77 (default compiler settings) on an Intel Pentium 4, 1.8GHz workstation. In the following, we present the numerical experiments by considering a number of 80 unconstrained optimization test problems of medium size ($n=100$ variables), described in [1] solved with all seven scaled modified BFGS algorithms. The comparisons of the algorithms are given in the following context. Let f_i^{ALG1} and f_i^{ALG2} be the optimal value found by ALG1 and ALG2 for problem $i=1,\ldots,80$, respectively. We say that, in the particular problem i, the performance of ALG1 was better than the performance of ALG2 if:

$$\left|f_i^{ALG1}-f_i^{ALG2}\right|<10^{-3} \tag{6.1}$$

and the number of iterations (#iter), or the number of function-gradient evaluations (#fg), or the CPU time of ALG1 was less than the number of iterations, or the number of function-gradient evaluations, or the CPU time corresponding to ALG2, respectively.

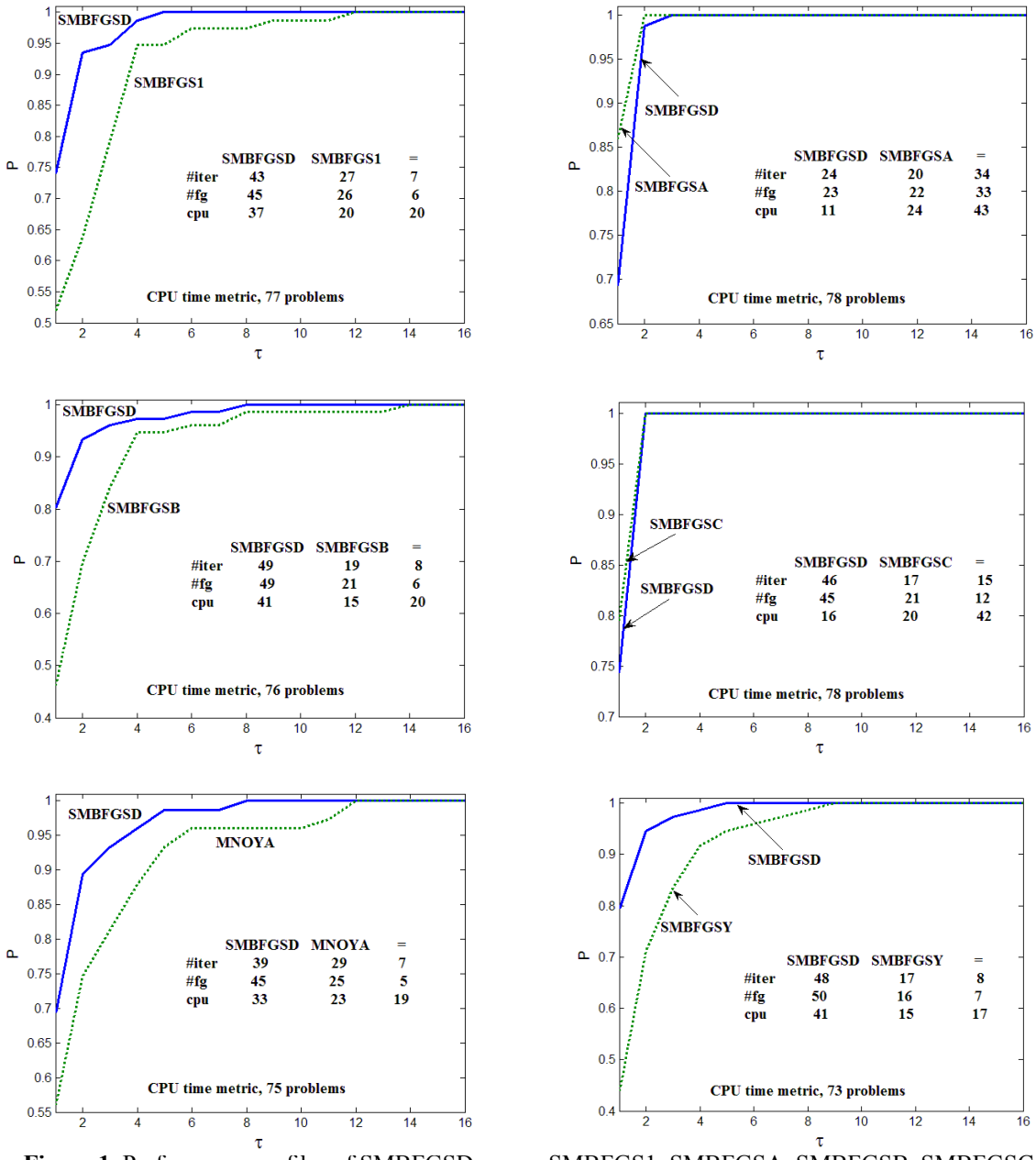


Figure 1. Performance profiles of SMBFGSD versus SMBFGS1, SMBFGSA, SMBFGSB, SMBFGSC, MNOYA and SMBFGSY

Table 1 presents a global overview of the algorithms by showing the total number of iterations (*itert*), the total number of function and its gradient evaluations (*fgt*) and the total CPU computing time (*cput*) for solving 80 unconstrained optimization problems used in this numerical study. Subject to the number of iterations, the double parameter scaled modified SMBFGSD is the best algorithm for solving all 80 unconstrained optimization problems. Subject to the CPU computing time, SMBFGSC is the fastest algorithm. Close to this there are SMBFGSA and SMBFGSD.

Table 1. Characteristics of SMBFGS algorithms for solving 80 unconstrained optimization test problems

	itert	*fgt*	*cput*
SMBFGS1	10114	68963	11.06
SMBFGSD	6735	51758	8.09
SMBFGSA	7083	46826	8.05
SMBFGSB	9826	66109	11.16
SMBFGSC	6936	52770	7.83
MNOYA	11102	60968	12.60
SMBFGSY	12140	80823	13.78

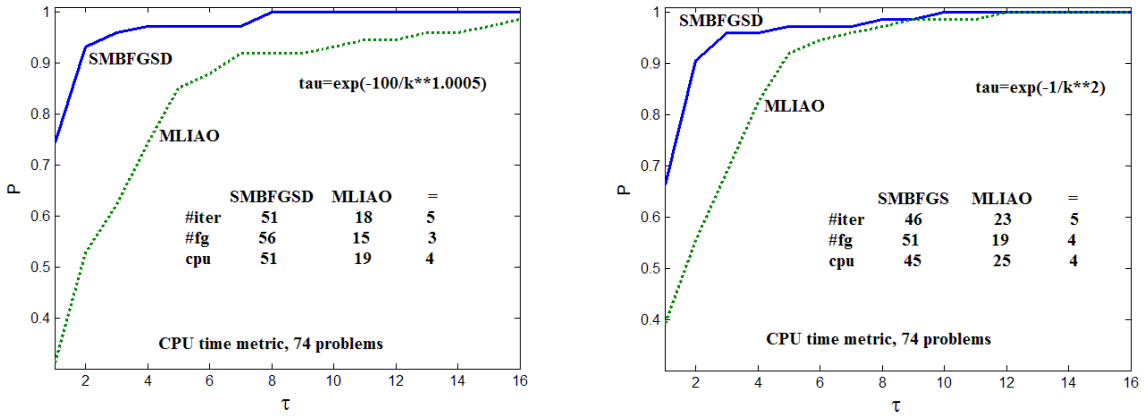
In the first set of numerical experiments we compare SMBFGSD versus SMBFGS1, SMBFGSA, SMBFGSB, SMBFGSC, MNOYA and SMBFGSY. Figure 1 presents the Dolan and Moré's performance profiles [11] of these algorithms for this set of unconstrained optimization test problems based on the CPU time metric. The table inside the figures represents the performances of the algorithms for $\tau = 1$, i.e. the efficiency of the algorithms. For example, when comparing SMBFGSD versus SMBFGS1 (standard modified BFGS algorithm: $\delta_k = 1$ and $\gamma_k = 1$), subject to the number of iterations, we see that SMBFGSD was better in 43 problems (i.e. it achieved the minimum number of iterations in 43 problems), SMBFGS1 was better in 27 problems. Both of them achieved the same number of iterations in 7 problems, etc. Out of 80 problems considered in this set of numerical experiments only for 77 does the criterion (6.1) holds.

Concerning the scaled modified BFGS update of Nocedal and Yuan (MNOYA) given by (4.5) with $\gamma_k = 1$ and $\delta_k = \bar{y}_k^T s_k / s_k^T B_k s_k$, scaling of the first two terms of B_{k+1} matrix leads to disappointing numerical results. This is consistent with the analysis given by Nocedal and Yuan [19] and Shanno and Phua [24] in their study on scaling the BFGS update. On the other hand, in our study on the double parameter scaled modified BFGS algorithm SMBFGSD we emphasize that both parameters γ_k and δ_k are important in the economy of the algorithm: δ_k is computed to cluster the eigenvalues of B_{k+1} and γ_k is responsible for shifting the large eigenvalues to the left. These are the main reasons why SMBFGSD has better performances than MNOYA.

In the second set of numerical experiments we compare the SMBFGSD algorithm versus the modified BFGS update by Liao (MLIAO). The scaled modified BFGS algorithm by Liao is defined by (3.10) and (3.11) where y_k is replaced by $\bar{y}_k$ given by (2.7). Figure 2a presents the Dolan and Moré performance profiles of SMBFGSD versus modified LIAO (MLIAO) with $\tau_k = \exp(-100/k^{1.0005})$. Figure 2b presents the performance profiles of SMBFGSD versus modified LIAO (MLIAO) with $\tau_k = \exp(-1/k^2)$. We observed that if τ_k is small, like in the MLIAO algorithm with $\tau_k = \exp(-100/k^{1.0005})$, then the algorithm takes $\delta_k = s_k^T B_k s_k / (s_k^T B_k s_k + \bar{y}_k^T s_k)$ and $\gamma_k = \bar{y}_k^T s_k / (s_k^T B_k s_k + \bar{y}_k^T s_k)$, as specified in (3.11). On the other hand, if τ_k is relatively large, like in the MLIAO algorithm with $\tau_k = \exp(-1/k^2)$, then the algorithm selects $\delta_k = \tau_k$ and $\gamma_k = 1$, as recommended by (3.11). Since in the modified LIAO algorithm the search direction d_{k+1} is computed as solution of the system $B_{k+1} d_{k+1} = -g_{k+1}$, we generated a Fortran version of the SMBFGSD code where the search direction is computed as solution of the system $B_{k+1} d_{k+1} = -g_{k+1}$ to compare it with the modified LIAO algorithm. From Figure 2 we see that SMBFGSD is top performer versus the modified LIAO and the difference is significant subject to the efficiency and robustness of the algorithms. Since these codes use the same Wolfe line search and the same stopping criterion, they differ only in their choice of the search direction. Again, observe that the numerical results with the modified LIAO algorithm are disappointing. This is because in the modified LIAO the modified BFGS update is obtained by a simple symmetrization procedure from a rank one update (see [17]).

7. Conclusions

In this paper we combined two important approaches for improving the numerical performances of the quasi-Newton BFGS method: the modified BFGS method and the scaled BFGS method. The modified BFGS method consists of modifying the vector y_k in such a way to improve the approximation of the Hessian of the minimizing function [14, 15, 25, 29, 30]. Scaling methods are more diversified and concentrate on scaling the terms of the BFGS update, as presented by: Biggs [5, 6], Yuan [28], Cheng and Li [10],



(a) (b)

Figure 2. Performance profile of SMBFGSD versus two variants of MLIAO

Andrei [2, 4], Nocedal and Yuan [19] and Liao [17]. In our development we scaled the modified BFGS update by Yuan and Wei [30] using the double parameter scaling procedure by Andrei [4]. The idea of scaling the Yuan and Wei's modified BFGS method is to cluster the eigenvalues of the corresponding iteration matrix. For this, the first two terms of the modified BFGS update are scaled with a positive parameter determined to cluster the eigenvalues, while the third term is scaled with another positive parameter to shift to the left the large eigenvalues. Thus we obtained a *double parameter scaled modified BFGS update* for which we proved the global convergence in very general conditions without assuming the convexity of the minimizing function, using only the determinant and the trace of the updating modified scaled matrix.

Intensive numerical experiments of these scaled modified BFGS algorithms on a collection of 80 unconstrained optimization test problems of different complexities [1, 22] showed that the double parameter scaled modified BFGS method is more efficient and more robust versus other scaled modified BFGS methods discussed in this paper.

The conclusion of this study is that scaling the modified BFGS methods improves the numerical performances of the algorithms, leading us to new efficient and robust algorithms for unconstrained optimization. In our double parameter scaled modified BFGS method we noticed that scaling the first two terms of the modified BFGS update to cluster the eigenvalues has a major effect on the numerical performances of the modified BFGS algorithm. However, the most important is scaling the third term of the modified BFGS update. The scaling of the third term will push down to the left the eigenvalues of the modified BFGS update matrix, thus obtaining a better structure of the eigenvalues than the one of the modified BFGS.

It is important to notice that in our development we scaled the modified BFGS method of Yuan and Wei [30], but another line or research is to scale the modified BFGS update of Wei, Yu, Yuan and Lian [25]. In this way other scaled modified BFGS algorithms are obtained which can be compared versus our scaled modified BFGS algorithms considered in this paper.

REFERENCES

1. Andrei, N. (2008). An unconstrained optimization test functions collection, *Advanced Modeling and Optimization – An Electronic International Journal*, *10*, 147-161.

2. Andrei, N. (2017). An adaptive scaled BFGS method for unconstrained optimization, *Numerical Algorithms*, *77*(2), 413-432.

3. Andrei, N. (2017). Eigenvalues versus singular values study in conjugate gradient algorithms for large-scale unconstrained optimization, *Optimization Methods and Software*, *32*, 534-551.

4. Andrei, N. (2018). A double parameter scaled BFGS method for unconstrained optimization, *Journal of Computational and Applied Mathematics*, *332*, 26-44.

5. Biggs, M. C. (1971). Minimization algorithms making use of non-quadratic properties of the objective function, *Journal of the Institute of Mathematics and Its Applications*, *8*, 315-327.

6. Biggs, M. C. (1973). A note on minimization algorithms making use of non-quadratic properties of the objective function, *Journal of the Institute of Mathematics and Its Applications*, *12*, 337-338.

7. Broyden, C. G. (1970). The convergence of a class of double-rank minimization algorithms. I. General considerations, *J. Inst. Math. Appl.*, *6*, 76-90.

8. Byrd, R. H., Liu, D. C. & Nocedal, J. (1992). On the behavior of Broyden's class of quasi-Newton methods, *SIAM J. Optim.*, *2*, 533-557.

9. Byrd, R. H. & Nocedal, J. (1989). A tool for the analysis of quasi-Newton methods with application to unconstrained optimization, *SIAM J. Numer. Anal.*, *26*, 727-739.

10. Cheng, W. Y. & Li, D. H. (2010). Spectral scaling BFGS method, *Journal of Optimization Theory and Applications*, *146*, 305-319.

11. Dolan, E. D. & Moré, J. J. (2002). Benchmarking optimization software with performance profiles, *Mathematical Programming*, *91*, 201-213.

12. Fletcher, R. (1970). A new approach to variable metric algorithms, *The Computer Journal*, *13*, 317-322.

13. Goldfarb, D. (1970). A family of variable metric methods derived by variation mean, *Mathematics of Computation*, *23*, 23-26.

14. Li, D. & Fukushima, M. (2001). A modified BFGS method and its global convergence in nonconvex minimization, *J. Comput. Appl. Math.*, *129*, 15-35.

15. Li, D. & Fukushima, M. (2001). On the global convergence of the BFGS method for nonconvex unconstrained optimization problems, *SIAM J. Optim.*, *11*, 1054-1064.

16. Li, G., Tang, C. & Wei, Z. (2007). New conjugacy condition and related new conjugate gradient methods for unconstrained optimization problems, *J. Comput. Appl. Math.*, *202*, 532-539.

17. Liao, A. (1997). Modifying BFGS method, *Operations Research Letters*, *20*, 171-177.

18. Nocedal, J. (1992). Theory of algorithms for unconstrained optimization, *Acta Numerica*, *1*, 199-242.

19. Nocedal, J. & Yuan, Y. X. (1993). Analysis of self-scaling quasi-Newton method, *Mathematical Programming*, *61*, 19-37.

20. Oren, S. S. & Luenberger, D. G. (1976). Self-scaling variable metric (SSVM) algorithms, Part I: Criteria and sufficient conditions for scaling a class of algorithms, *Management Science*, *20*, 845-862.

21. Powell, M. J. D. (1987). Updating conjugate directions by the BFGS formula, *Mathematical Programming*, *38*, 693-726.

22. Rădulescu, I. C. (2015). Rezolvarea unor probleme de optimizare multi-obiectiv bazată pe algoritmi evolutive, *Revista Română de Informatică* şi *Automatică*, *25*(2), 39-48.

23. Shanno, D. F. (1970). Conditioning of quasi-Newton methods for function minimization, *Mathematics of Computation*, *24*, 647-656.

24. Shanno, D. F. & Phua, K. H. (1978). Matrix conditioning and nonlinear optimization, *Mathematical Programming*, *14*, 149-160.

25. Wei, Z., Yu, G., Yuan, G. & Lian, Z. (2004). The superlinear convergence of a modified BFGS-type method for unconstrained optimization, *Comput. Optim. Appl.*, *29*, 315-332.

26. Wolfe, P. (1969). Convergence conditions for ascent methods, *SIAM Review*, *11*, 226-235.

27. Wolfe, P. (1971). Convergence conditions for ascent methods. II: Some corrections, *SIAM Review*, *13*, 185-188.

28. Yuan, Y. X. (1991). A modified BFGS algorithm for unconstrained optimization, *IMA Journal Numerical Analysis*, *11*, 325-332.

29. Yuan, G. & Wei, Z. (2008). The superlinear convergence analysis of a nonmonotone BFGS algorithm on convex objective functions, *Acta Math. Sinica, Eng. Ser.*, *24*, 35-42.

30. Yuan, G. & Wei, Z. (2010). Convergence analysis for a modified BFGS method on convex minimization, *Comput. Optim. Appl.*, *47*, 237-255.