PANM 18

Jan Vlček; Ladislav Lukšan

A generalized limited-memory BNS method based on the block BFGS update

In: Jan Chleboun and Pavel Kůs and Petr Přikryl and Karel Segeth and Jakub Šístek and Tomáš Vejchodský (eds.): Programs and Algorithms of Numerical Mathematics, Proceedings of Seminar. Janov nad Nisou, June 19-24, 2016. Institute of Mathematics CAS, Prague, 2017. pp. 164–171.

Persistent URL: http://dml.cz/dmlcz/703010

Programs and Algorithms of Numerical Mathematics 18
J. Chleboun, P. Kůs, P. Přikryl, K. Segeth, J. Šístek, T. Vejchodský (Eds.)
Institute of Mathematics CAS, Prague 2017

A GENERALIZED LIMITED-MEMORY BNS METHOD BASED ON THE BLOCK BFGS UPDATE

Jan Vlček[1], Ladislav Lukšan[1,2]
[1]Institute of Computer Science, Czech Academy of Sciences,
Pod Vodárenskou věží 2, 182 07 Prague 8, Czech Republic
vlcek@cs.cas.cz, luksan@cs.cas.cz
[2]Technical University of Liberec, Hálkova 6, 461 17 Liberec, Czech Republic

Abstract: A block version of the BFGS variable metric update formula is investigated. It satisfies the quasi-Newton conditions with all used difference vectors and gives the best improvement of convergence in some sense for quadratic objective functions, but it does not guarantee that the direction vectors are descent for general functions. To overcome this difficulty and utilize the advantageous properties of the block BFGS update, a block version of the limited-memory BNS method for large scale unconstrained optimization is proposed. The algorithm is globally convergent for convex sufficiently smooth functions and our numerical experiments indicate its efficiency.

Keywords: Unconstrained minimization, block variable metric methods, limited-memory methods, the BFGS update, global convergence, numerical results

MSC: 65K10

1. Introduction

In this contribution we propose a block version of the widely used BNS method, see [3], for large scale unconstrained optimization

$$\min f(x): \ x \in \mathcal{R}^N,$$

where it is assumed that the problem function $f : \mathcal{R}^N \to \mathcal{R}$ is differentiable.

The BNS method belongs to the variable metric (VM) or quasi-Newton (QN) line search iterative methods, see [9], [11]. They start with an initial point $x_0 \in \mathcal{R}^N$ and generate iterations $x_{k+1} \in \mathcal{R}^N$ by the process $x_{k+1} = x_k + s_k$, $s_k = t_k d_k$, $k \ge 0$, where usually the direction vector $d_k \in \mathcal{R}^N$ is $d_k = -H_k g_k$, matrix H_k is symmetric positive definite and a stepsize $t_k > 0$ is chosen in such a way that

$$f_{k+1} - f_k \le \varepsilon_1 t_k g_k^T d_k, \qquad g_{k+1}^T d_k \ge \varepsilon_2 g_k^T d_k, \qquad k \ge 0 \tag{1}$$

(the Wolfe line search conditions, see [11]), where $0 < \varepsilon_1 < 1/2$, $\varepsilon_1 < \varepsilon_2 < 1$, $f_k = f(x_k)$, $g_k = \nabla f(x_k)$. Typically, H_0 is a multiple of I and H_{k+1} is obtained from

DOI: 10.21136/panm.2016.19

H_k by a VM update to satisfy the QN condition (see [9]) $H_{k+1}y_k = s_k$, $y_k = g_{k+1} - g_k$, $k \ge 0$.

Among VM methods, the BFGS method, see [9], [11], belongs to the most efficient. It preserves positive definite VM matrices and can be written in the form

$$H_+ = (1/b)ss^T + \big(I - (1/b)sy^T\big)H\big(I - (1/b)ys^T\big), \quad b = s^T y, \tag{2}$$

$b > 0$ by (1). Note that for simplification we often omit index k and replace indices $k+1$, $k-1$ by symbols $+, -$, respectively. The BNS and L-BFGS (see [5], [6] – subroutine PLIS) methods represent its well-known limited-memory adaptations (for large-scale optimization). In every iteration we repeatedly update an initial approximation of the inverse Hessian matrix $\zeta_k I$, $\zeta_k > 0$, by the BFGS method, using m couples of vectors $(s_{k-\tilde{m}}, y_{k-\tilde{m}}), \ldots, (s_k, y_k)$ successively (without forming approximations of the inverse Hessian matrix explicitly), where $\tilde{m} = \min(k, \hat{m} - 1)$, $m = \tilde{m} + 1$ and $\hat{m} > 1$ is a given parameter. In the case of the BNS method, matrix H_+ can be expressed in the form, see [3],

$$H_+ = SU^{-T}DU^{-1}S^T + \zeta\Big(I - SU^{-T}Y^T\Big)\Big(I - YU^{-1}S^T\Big), \tag{3}$$

where for $k \ge 0$ we denote $S_k = [s_{k-\tilde{m}}, \ldots, s_k]$, $Y_k = [y_{k-\tilde{m}}, \ldots, y_k]$, $(U_k)_{i,j} = (S_k^T Y_k)_{i,j}$ for $i \le j$, $(U_k)_{i,j} = 0$ otherwise (an upper triangular matrix), $D_k = \mathrm{diag}[b_{k-\tilde{m}}, \ldots, b_k]$.

For $S^T Y$ nonsingular and any $H \in \mathcal{R}^{N \times N}$, the BFGS update formula (2) can be easily generalized to the following block version

$$H_+ = S(S^TY)^{-1}S^T + \big(I - S(S^TY)^{-T}Y^T\big)\,\bar{H}\,\big(I - Y(S^TY)^{-1}S^T\big)\,,\ \bar{H} = \frac{1}{2}(H + H^T), \tag{4}$$

which satisfies the QN conditions $H_+ Y = S$ (for the whole block of stored difference vectors) and was derived in [12] and [4] for $S^T Y, H$ symmetric positive definite.

Formula (4) is not directly applicable to general functions, since it does not guarantee that the corresponding direction vectors are descent. Thus we split matrices S and Y in such a way that $S = [S_{[1]}, \ldots, S_{[n]}]$, $Y = [Y_{[1]}, \ldots, Y_{[n]}]$, with all blocks $S_{[i]}^T Y_{[i]}$ positive definite, i.e. matrices $S_{[i]}^T Y_{[i]} + Y_{[i]}^T S_{[i]}$ symmetric positive definite, $i = 1, \ldots, n$. Then we replace the BNS formula (3) by n successive updates of an initial matrix ζI using a modification of the block BFGS update (4) with matrices $S_{[i]}, Y_{[i]}$, $i = 1, \ldots, n$, instead of S, Y. Obviously, for $n = m$ we obtain the BNS method.

In Section 2 we derive the block BFGS update, investigate its properties and show some similarities to the VM methods, based on the corrected BFGS updates, see the limited-memory BFGS method [13] – [16]. In Section 3 we focus on quadratic functions and show optimality of the block BFGS method and a role of unit stepsizes. In Section 4 we present the block BNS method and derive a convenient formula similar to (3) to represent the resultant VM matrix. The simplified algorithm is described in Section 5. Global convergence of the algorithm is established in Section 6 and numerical results are reported in Section 7. We refer to report [17] for details and proofs of assertions, here we briefly present only the main results.

We will denote the Frobenius matrix norm by $\|\cdot\|_F$.

2. The block BFGS update

Using the following theorem, update (4) can be derived for general functions:

Theorem 1. *Let $W_L, W_R \in \mathcal{R}^{N\times N}$ be nonsingular, matrix Y have a full rank, matrix H_+ be given by (4) and matrices A_{i+1}, $i=-1,0,\dots,$ be the unique solution to*

$$\min_{A_{i+1}\in\mathcal{R}^{N\times N}} \left\|W_L^{-1}(A_{i+1}-\bar{A}_i)W_R^{-1}\right\|_F \ \ s.t.\ A_{i+1}Y=S,\ \ \bar{A}_{-1}=H,\ \bar{A}_i=\frac{1}{2}(A_i+A_i^T),$$

$i\ge 0$. Then for $W_R^TW_RY=ST_S$, T_S square nonsingular, we have $\lim_{i\to\infty}A_i=H_+$.

The new update has similar interesting properties as the standard BFGS update.

Theorem 2. *Let matrix H_+ be given by (4), matrices $S^TY, \bar{H}, H_+, S^T\bar{B}S, T\in\mathcal{R}^{m\times m}$ nonsingular, $\bar{B}=\bar{H}^{-1}$, $B_+=H_+^{-1}$. Then (also for nonsymmetric $\bar{H}$)*

- *(a) matrix H_+ is invariant under the transformation $S\to ST$, $Y\to YT$,*
- *(b) $B_+=\bar{B}-\bar{B}S(S^T\bar{B}S)^{-1}S^T\bar{B}+Y(S^TY)^{-T}Y^T$,*
- *(c) $\det B_+=\det\bar{B}\,.\det(S^TY)/\det(S^T\bar{B}S)$,*
- *(d) for H and S^TY positive definite, also matrix H_+ is positive definite,*
- *(e) for $S\triangleq[\check{S},s], Y\triangleq[\check{Y},y], S^TY, \check{S}^T\check{Y}$ symmetric nonsingular, $\check{P}=I-\check{Y}(\check{S}^T\check{Y})^{-1}\check{S}^T$, $\tilde{s}=\check{P}^Ts$ and $\tilde{y}=\check{P}y$ we have $\tilde{s}^T\tilde{y}\triangleq\tilde{b}\ne 0$, $H_+=(1/\tilde{b})\,\tilde{s}\tilde{s}^T+\tilde{P}^T\check{H}\tilde{P}$, $\tilde{P}=I-(1/\tilde{b})\tilde{y}\tilde{s}^T$, $\check{H}=\check{S}(\check{S}^T\check{Y})^{-1}\check{S}^T+\check{P}^T\bar{H}\check{P}$; besides, $\check{S}^TB_+\tilde{s}=\check{S}^T\check{H}^{-1}\tilde{s}=0$ holds.*

Theorem 2*(e)* shows some connections with our methods [13] – [16] based on vector corrections for conjugacy. The following theorem indicates that we can expect good properties of the block BFGS update also for functions similar to quadratic.

Theorem 3. *Let matrices $\check{S},\check{Y},\check{P},\check{H},H_+$ and vectors $\tilde{s},\tilde{y}$ have the same meaning as in Theorem 2(e), S^TY be symmetric positive definite, $\ddot{s}=s+\check{S}\sigma$, $\ddot{y}=y+\check{Y}\sigma$, $\sigma\in\mathcal{R}^{\tilde{m}}$, $\tilde{m}\ge 1$. Then $\ddot{b}=\ddot{s}^T\ddot{y}\ge\tilde{s}^T\tilde{y}>0$. Moreover, if matrix $\bar{H}$ is nonsingular, $\ddot{P}=I-(1/\ddot{b})\ddot{y}\,\ddot{s}^T$, $\ddot{H}_+=(1/\ddot{b})\ddot{s}\,\ddot{s}^T+\ddot{P}^T\check{H}\ddot{P}$ and $\hat{G}$ is any symmetric positive definite matrix satisfying $\hat{G}S=Y$, then function $\varphi(\sigma)=\|\hat{G}^{\frac{1}{2}}\ddot{H}_+\hat{G}^{\frac{1}{2}}-I\|_F$ is minimized and $\ddot{H}_+=H_+$ holds for $\sigma=-(\check{S}^T\check{Y})^{-1}\check{Y}^Ts$, when $\ddot{s}=\tilde{s},\ddot{y}=\tilde{y}$.*

Paradoxically, the standard BFGS update often gives better results if S^TY is almost symmetric and the Hessian matrix is ill-conditioned. Therefore we will use, in addition to the choice $\ddot{s} = \tilde{s}, \ddot{y} = \tilde{y}$, also the choice $\ddot{s} = s, \ddot{y} = y$, which corresponds to the standard BFGS update of $\check{H}$ and can be easily realized by means of blocks of order one, or a special choice $\ddot{s} = s - (s^T y_-/b_-)s_-$, $\ddot{y} = y - (y^T s_-/b_-)y_-$, which can be more robust than the block BFGS update, see [17] for details.

3. Results for quadratic functions

Compared to the BNS method, the block BFGS update gives the best improvement of convergence in some sense for linearly independent direction vectors:

Theorem 4. *Let $f(x) = \frac{1}{2}(x - \bar{x})^T G(x - \bar{x})$, $\bar{x} \in \mathcal{R}^N$, with a symmetric positive definite matrix G, let all columns of S be linearly independent, $\underline{k} = k - \tilde{m}$, $\hat{S}_i = [s_{\underline{k}}, \ldots, s_i]$, $\hat{Y}_i = [y_{\underline{k}}, \ldots, y_i]$, $\hat{P}_i = I - \hat{Y}_i(\hat{S}_i^T \hat{Y}_i)^{-1}\hat{S}_i^T$, $i = \underline{k}, \ldots, k$, $\ddot{s}_{\underline{k}} = \tilde{s}_{\underline{k}} = s_{\underline{k}}$, $\ddot{y}_{\underline{k}} = \tilde{y}_{\underline{k}} = y_{\underline{k}}$, $\ddot{s}_i = s_i + \hat{S}_{i-1}\sigma_i$, $\ddot{y}_i = y_i + \hat{Y}_{i-1}\sigma_i$, $\sigma_i \in \mathcal{R}^{i-1}$, $\tilde{s}_i = \hat{P}_{i-1}^T s_i$, $\tilde{y}_i = \hat{P}_{i-1} y_i$, $i = \underline{k} + 1, \ldots, k$. Then matrices $\hat{S}_i^T \hat{Y}_i$ are symmetric positive definite and $\ddot{s}_i^T \ddot{y}_i \ge \tilde{s}_i^T \tilde{y}_i > 0$, $i = \underline{k}, \ldots, k$.*

Moreover, let $\bar{H}$ be symmetric positive definite, H_+ be given by (4) and $\ddot{H}_{k+1}$ by

$$\ddot{H}_{\underline{k}} = \bar{H}, \quad \ddot{H}_{i+1} = (1/\ddot{s}_i^T \ddot{y}_i)\, \ddot{s}_i \ddot{s}_i^T + \ddot{P}_i^T \ddot{H}_i \ddot{P}_i, \quad \ddot{P}_i = I - (1/\ddot{s}_i^T \ddot{y}_i)\, \ddot{y}_i \ddot{s}_i^T,$$

$i = \underline{k}, \ldots, k$. Then value $\|G^{1/2}\ddot{H}_+ G^{1/2} - I\|_F$ is minimized and matrices $\ddot{H}_+$ and H_+ are identical and symmetric positive definite for $\ddot{s}_i = \tilde{s}_i$, $\ddot{y}_i = \tilde{y}_i$, $i = \underline{k} + 1, \ldots, k$.

Furthermore, similarly to Theorem 3.3 in [16], we get (see Theorem 3.2 in [17]) that if one stepsize t is unit in two successive iterations with matrices H, H_+ obtained by the block BFGS updates, all stored direction vectors from previous iterations are conjugate with vector s_+; thus if all steps are unit, all matrices S^TY are tridiagonal.

4. The block BNS method

Using Lemma 1, we split matrices S, Y in such a way that $S = [S_{[1]}, \ldots, S_{[n]}]$, $Y = [Y_{[1]}, \ldots, Y_{[n]}]$, $n \ge 1$, with all blocks $S_{[i]}^T Y_{[i]}$ positive definite ($S_{[i]}^T Y_{[i]} + Y_{[i]}^T S_{[i]}$ symmetric positive definite), $i = 1, \ldots, n$, and use the theory in Section 2 for matrices $S_{[i]}, Y_{[i]}$ instead of S, Y. We use the RL factorization instead of the LU one, since we start with the submatrices of S, Y which contain their latest columns to have maximum of the latest QN conditions satisfied. The following lemma converts the problem of factorization to the same problem of a smaller dimension. A generalization of the standard BNS formula (3) is given by Theorem 5.

Lemma 1. *Suppose that $A, R, L \in \mathcal{R}^{\mu\times\mu}$, $\mu > 0$, $u, v \in \mathcal{R}^\mu$, $\alpha \in \mathcal{R}$, $\alpha \ne 0$,*

$$\bar{A} = \begin{bmatrix} A & u \\ v^T & \alpha \end{bmatrix}, \quad \bar{R} = \begin{bmatrix} R & u \\ & \alpha \end{bmatrix}, \quad \bar{L} = \begin{bmatrix} L & \\ (1/\alpha)\, v^T & 1 \end{bmatrix}.$$

Then to have $\bar{A} = \bar{R}\bar{L}$, it suffices to find R, L satisfying $A - (1/\alpha)uv^T = RL$. Moreover, *(a) if $u = v$ then matrix $\bar{A}$ is symmetric positive definite if and only if both $\alpha > 0$ and matrix $A - (1/\alpha)uv^T$ is symmetric positive definite,*

(b) if matrix $\bar{A}$ is positive definite, then $\alpha > 0$ and $A - (1/\alpha)uv^T$ is positive definite.

Theorem 5. *Let $\zeta > 0, H_{[1]} = \zeta I, S = [S_{[1]}, \dots, S_{[n]}]$, $Y = [Y_{[1]}, \dots, Y_{[n]}]$, $S_{[i]}^T Y_{[i]}$ nonsingular, $P_{[i]} = I - Y_{[i]}(S_{[i]}^T Y_{[i]})^{-1} S_{[i]}^T$, $H_{[i+1]} = S_{[i]}(S_{[i]}^T Y_{[i]})^{-1} S_{[i]}^T + \frac{1}{2} P_{[i]}^T (H_{[i]} + H_{[i]}^T) P_{[i]}$, $\Sigma_i = Y_{[i]}^T S_{[i]}$, $1 \le i \le n$, $H_+ = H_{[n+1]}$. Then ($\tilde{U}$ is an upper block triangular matrix)*

$$H_+ = S\tilde{U}^{-T} E \tilde{U}^{-1} S^T + \zeta \big(I - S\tilde{U}^{-T} Y^T\big)\big(I - Y\tilde{U}^{-1} S^T\big)\,, \tag{5}$$

$$E = \operatorname{diag}\big[(1/2)(\Sigma_1 + \Sigma_1^T), \dots, (1/2)(\Sigma_{n-1} + \Sigma_{n-1}^T), \Sigma_n\big]\,, \tag{6}$$

$$\tilde{U} = \begin{bmatrix} S_{[1]}^T Y_{[1]} & \dots & S_{[1]}^T Y_{[n-1]} & S_{[1]}^T Y_{[n]} \\ & \ddots & \vdots & \vdots \\ & & S_{[n-1]}^T Y_{[n-1]} & S_{[n-1]}^T Y_{[n]} \\ & & & S_{[n]}^T Y_{[n]} \end{bmatrix}. \tag{7}$$

Although matrix H_+ is unsymmetric generally, we use the usual direction vector $d_+ = -H_+ g_+$, such that $z^* = x_+ + d_+$ satisfies $g(z^*) = 0$, $g(z) = g_+ + H_+^{-1}(z - x_+)$ (a linear model for gradients which respects the QN conditions); for ill-conditioned problems we usually obtained better results than e.g. with vector $\bar{d}_+ = -(1/2)(H_+ + H_+^T) g_+$.

5. Implementation

Although we need not the symmetry of H_+ to establish global convergence, for better efficiency we also want to have all submatrices $S_{[i]}^T Y_{[i]}$ sufficiently near to symmetric. Since the block BFGS update can deteriorate stability, we sometimes do not use this update for the last block $S_{[n]}^T Y_{[n]}$, see Section 2 and [17] for details.

Algorithm 5.1 (simplified)

Data: A maximum number $\hat{m} > 1$ of columns of matrices S, Y, line search parameters and a global convergence parameter $\varepsilon_D \in (0, 1)$.

Step 0: Initiation. Choose starting point $x_0 \in \mathcal{R}^N$, define starting matrix $H_0 = I$ and direction vector $d_0 = -g_0$ and initiate iteration counter k to zero.

Step 1: Line search. Compute $x_{k+1} = x_k + t_k d_k$, where t_k satisfies (1), $g_{k+1} = \nabla f(x_{k+1})$, $s_k = t_k d_k$, $y_k = g_{k+1} - g_k$, $b_k = s_k^T y_k$, $\zeta_k = b_k / y_k^T y_k$. If $k = 0$ set $S_k = [s_k]$, $Y_k = [y_k]$, $S_k^T Y_k = [b_k]$, $Y_k^T Y_k = [\, y_k^T y_k]$, compute $S_k^T g_{k+1}, Y_k^T g_{k+1}$ and go to Step 4.

Step 2: Matrix updates. Compute $\check{Y}_k^T s_k = -t_k \check{Y}_k^T H_k g_k$ and form basic matrices $S_k := [\check{S}_k, s_k]$, $Y_k := [\check{Y}_k, y_k]$, $S_k^T Y_k := \begin{bmatrix} \check{S}_k^T \check{Y}_k & \check{S}_k^T y_k \\ s_k^T \check{Y}_k & s_k^T y_k \end{bmatrix}$, $Y_k^T Y_k := \begin{bmatrix} \check{Y}_k^T \check{Y}_k & \check{Y}_k^T y_k \\ y_k^T \check{Y}_k & y_k^T y_k \end{bmatrix}$.

Step 3: *Block factorization.* Create and factorize positive definite blocks $S_{[i]}^T Y_{[i]} = R_{[i]} L_{[i]}$ and $S_{[i]}^T Y_{[i]} + Y_{[i]}^T S_{[i]} = \bar{R}_{[i]} \bar{L}_{[i]}$ with unit diagonal entries of $L_{[i]}, \bar{L}_{[i]}$ and with diagonal entries of $\bar{R}_{[i]}$ greater than $\varepsilon_D \operatorname{Tr} S_{[i]}^T Y_{[i]}$, $i = n, \ldots, 1$, where number $n \geq 1$ is determined during this process.

Step 4: *Direction vector.* Compute $d_{k+1} = -H_{k+1} g_{k+1}$ by the block BNS method and an auxiliary vector $Y_k H_{k+1} g_{k+1}$. Set $k := k+1$. If $k \geq \hat{m}$ delete the first column of S_{k-1}, Y_{k-1} and the first row and column of $S_{k-1}^T Y_{k-1}, Y_{k-1}^T Y_{k-1}$ to form matrices $\check{S}_k, \check{Y}_k, \check{S}_k^T \check{Y}_k, \check{Y}_k^T \check{Y}_k$. Go to Step 1.

6. Global convergence

Assumption 1. *The objective function $f : \mathcal{R}^N \to \mathcal{R}$ is bounded from below and uniformly convex with bounded second-order derivatives (i.e. $0 < \underline{G} \leq \underline{\lambda}(G(x)) \leq \overline{\lambda}(G(x)) \leq \overline{G} < \infty$, $x \in \mathcal{R}^N$, where $\underline{\lambda}(G(x))$ and $\overline{\lambda}(G(x))$ are the lowest and the greatest eigenvalues of the Hessian matrix $G(x)$).*

Theorem 6. *If the objective function f satisfies Assumption 1, Algorithm 5.1 generates a sequence $\{g_k\}$ that satisfies $\lim_{k\to\infty} \|g_k\| = 0$ or terminates with $g_k = 0$ for some k.*

The proof of this theorem is based on Theorem 2 and some inequalities for nonsymmetric positive definite matrices, see [17].

7. Numerical experiments

We compare our results with the results obtained by the L-BFGS method [5] and the BNS method [3], all implemented in the system UFO [10], using the following collections of test problems:

- **Test 11** – 55 modified problems [8] from CUTE collection [2] with various dimensions N from 1000 to 5000 (prescribed for the given problem),
- **Test 12** – 73 problems from the collection [1], $N = 1000$, 2000 and 5000,
- **Test 25** – problems from the collection [7], 70 problems for $N = 1000$, 69 of them for $N = 2000$ and $N = 5000$.

The source texts and the reports corresponding to these test collections can be downloaded from the web page `camo.ici.ro/neculai/ansoft.htm` (Test 12) and from `www.cs.cas.cz/luksan/test.html` (Tests 11 and 25).

Method	Test 11	Test 12, $N =$			Test 25, $N =$		
	$N \leq 5000$	1000	2000	5000	1000	2000	5000
L-BFGS	79575	26526	41348	76703	125838	189248	445820
BNS	76463	25575	42227	76667	121297	179829	436457
Alg. 5.1	59858	21583	32425	56299	100334	151035	310684
Alg 5.1 as % of BNS	78	84	77	73	83	84	71

Table 1. Comparison of the total number of function evaluation.

Method	Test 11	Test 12, $N =$			Test 25, $N =$		
	$N \leq 5000$	1000	2000	5000	1000	2000	5000
L-BFGS	11.02	1.70	6.07	23.27	10.83	35.99	207.87
BNS	9.77	1.43	5.88	21.59	9.65	31.51	190.10
Alg. 5.1	7.46	1.23	4.80	16.16	7.88	26.74	135.47
Alg 5.1 as % of BNS	76	86	82	75	82	85	71

Table 2. Comparison of the total computational time in seconds.

We have used $\hat{m} = 5$, $\varepsilon_D = 10^{-6}$ and the final precision $\|g(x^\star)\|_\infty \leq 10^{-6}$. In the last row of Tables 1-2 we give the values for Algorithm 5.1 expressed as percentages of the corresponding values for the BNS method.

8. Conclusions

In this contribution, we derive a block version of the BFGS variable metric update formula for general functions and show some its positive properties and similarities to approaches based on vector corrections ([13] – [16]).

In spite of the fact that this formula does not guarantee that the corresponding direction vectors are descent, we propose the block BNS method for large scale unconstrained optimization, which utilizes the advantageous properties of the block BFGS update and is globally convergent.

Numerical results indicate that the block approach can improve unconstrained large-scale minimization results significantly compared with the frequently used L-BFGS and the BNS methods.

Acknowledgements

This work was supported by the Grant Agency of the Czech Republic, project No. 13-06684S, and the Institute of Computer Science of the CAS (RVO: 67985807). We thank the anonymous referee for inspiring suggestions.

References

[1] Andrei N.: An unconstrained optimization test functions collection. Advanced Modeling and Optimization **10** (2008), 147–161.

[2] Bongartz I., Conn A. R., Gould N., and Toint P. L.: CUTE: constrained and unconstrained testing environment. ACM Transactions on Math. Software **21** (1995), 123–160.

[3] Byrd R. H., Nocedal J., and Schnabel R. B.: Representation of quasi-Newton matrices and their use in limited memory methods. Math. Programming **63** (1994), 129–156.

[4] Hu Y.F. and Storey C.: Motivating Quasi-Newton updates by preconditioned conjugate gradient methods, Math. Report A 150, Department of Mathematical Sciences, Loughborough University of Technology, England, 1991.

[5] Liu D. C.and Nocedal J.: On the limited memory BFGS method for large scale optimization. Math. Prog. **45** (1989), 503–528.

[6] Lukšan L., Matonoha C., and Vlček J.: Algorithm 896: LSA — algorithms for large-scale optimization. ACM Trans. Math. Software **36** (2009), 16:1–16:29.

[7] Lukšan L., Matonoha C., and Vlček J.: Sparse test problems for unconstrained optimization. Report V-1064, ICS AS CR, Prague, 2010.

[8] Lukšan L., Matonoha C., and Vlček J.: Modified CUTE problems for sparse unconstrained optimization. Report V-1081, ICS AS CR, Prague, 2010.

[9] Lukšan L., Spedicato E.: Variable metric methods for unconstrained optimization and nonlinear least squares. J. Comput. Appl. Math. **124** (2000), 61–95.

[10] Lukšan L., Tůma M., Vlček J., Ramešová N., Šiška M., Hartman J., and Matonoha C.: UFO 2014. Interactive system for universal functional optimization. Report V-1218, ICS AS CR, Prague, 2014, `www.cs.cas.cz/luksan/ufo.html`.

[11] Nocedal J. and Wright S. J.: *Numerical optimization*. Springer-Verlag, New York, 1999.

[12] Schnabel R. B.: Quasi-Newton methods using multiple secant equations. Technical Report CU-CS-247-83, Department of Computer Science, University of Colorado at Boulder, USA, 1983.

[13] Vlček J. and Lukšan L.: A conjugate directions approach to improve the limited-memory BFGS method. Appl. Math. Comput. **219** (2012), 800–809.

[14] Vlček J. and Lukšan L.: Modifications of the limited-memory BFGS method based on the idea of conjugate directions. Proc. of Semin. PANM 16, Dolní Maxov, 2012.

[15] Vlček J. and Lukšan L.: A modified limited-memory BNS method for unconstrained minimization derived from the conjugate directions idea. Proc. of Semin. PANM 17, Dolní Maxov, 2014.

[16] Vlček J. and Lukšan L.: A modified limited-memory BNS method for unconstrained minimization based on conjugate directions idea. Optim. Methods Softw. **30** (2015), 616–633.

[17] Vlček J., Lukšan L.: A block version of the BNS limited-memory variable metric method for unconstrained minimization. Report V-1232, ICS AS CR, Prague, 2016.