A Block Version of the BNS Limited-Memory Variable Metric Method for Unconstrained Minimization

Vlček, Jan
2016

Institute of Computer Science
Academy of Sciences of the Czech Republic

A block version of the BNS limited-memory variable metric method for unconstrained minimization.

Jan Vlček[a], Ladislav Lukšan[a,b]

[a]Institute of Computer Science, Academy of Sciences of the Czech Republic,
Pod vodárenskou věží 2, 182 07 Prague 8, Czech Republic and
[b]Technical University of Liberec, Hálkova 6, 461 17 Liberec, Czech Republic

Technical report No. V 1232

August 2016

Pod Vodárenskou věží 2, 182 07 Prague 8 phone: +420 266 052 083, fax: +420 286 585 789,
e-mail:luksan@cs.cas.cz, vlcek@cs.cas.cz

Institute of Computer Science
Academy of Sciences of the Czech Republic

A block version of the BNS limited-memory variable metric method for unconstrained minimization.[1]

Jan Vlček[a], Ladislav Lukšan[a, b]

[a]Institute of Computer Science, Academy of Sciences of the Czech Republic,
Pod vodárenskou věží 2, 182 07 Prague 8, Czech Republic and
[b]Technical University of Liberec, Hálkova 6, 461 17 Liberec, Czech Republic

Technical report No. V 1232

August 2016

Abstract:

A block version of the BFGS variable metric update formula and its modifications are investigated. In spite of the fact that this formula satisfies the quasi-Newton conditions with all used difference vectors and that the improvement of convergence is the best one in some sense for quadratic objective functions, for general functions it does not guarantee that the corresponding direction vectors are descent. To overcome this difficulty, but at the same time utilize the advantageous properties of the block BFGS update, a block version of the limited-memory variable metric BNS method for large scale unconstrained optimization is proposed. Global convergence of the algorithm is established for convex sufficiently smooth functions. Numerical experiments demonstrate the efficiency of the new method.

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

[1]This work was supported by the Grant Agency of the Czech Republic, project No. 13-06684S, and the Institute of Computer Science of the AS CR (RVO: 67985807).

1 Introduction

In this report we propose a block version of the widely used BNS method, see [1], 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 [6], [12]. 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 $x_{k+1} = x_k + s_k$, $s_k = t_k d_k$, $k \geq 0$, where usually the direction vector $d_k \in \mathcal{R}^N$ is $d_k = -H_k g_k$, $g_k = \nabla f(x_k)$, with a symmetric positive definite matrix H_k and where stepsize $t_k > 0$ is chosen in such a way that

$$f_{k+1} - f_k \leq \varepsilon_1 t_k g_k^T d_k, \qquad g_{k+1}^T d_k \geq \varepsilon_2 g_k^T d_k, \qquad k \geq 0 \tag{1.1}$$

(the Wolfe line search conditions, see e.g. [15]), where $0 < \varepsilon_1 < 1/2$, $\varepsilon_1 < \varepsilon_2 < 1$, $f_k = f(x_k)$; typically H_0 is a multiple of I and H_{k+1} is obtained from H_k by a VM update to satisfy the QN condition (secant equation)

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

(see [6], [12]), where $y_k = g_{k+1} - g_k$, $k \geq 0$. For $k \geq 0$ we denote

$$B_k = H_k^{-1}, \qquad b_k = s_k^T y_k,$$

(note that $b_k > 0$ for $g_k \neq 0$ by (1.1)). To simplify the notation we frequently omit index k and replace index $k+1$ by symbol $+$ and index $k-1$ by symbol $-$.

Among VM methods, the BFGS method, see [6], [12], [15], belongs to the most efficient; the update formula preserves positive definite VM matrices and can be written in the following quasi-product form

$$H_+ = (1/b) s s^T + \Big(I - (1/b) s y^T\Big) H \Big(I - (1/b) y s^T\Big). \tag{1.3}$$

The BFGS method can be easily modified for large-scale optimization; the BNS and L-BFGS (see [8], [14], [9] - subroutine PLIS) methods represent its well-known limited-memory adaptations. In every iteration we recurrently update matrix $\zeta_k I$, $\zeta_k > 0$, (without forming an approximation of the inverse Hessian matrix explicitly) by the BFGS method, using m couples of vectors $(s_{k-\tilde{m}}, y_{k-\tilde{m}}), \ldots, (s_k, y_k)$ successively, where

$$\tilde{m} = \min(k, \hat{m}-1), \quad m = \tilde{m} + 1, \quad k \geq 0 \tag{1.4}$$

and $\hat{m} > 1$ is a given parameter. In case of the BNS method, matrix H_+ can be expressed either in the form, see [1],

$$H_+ = \zeta I + \Big[S, \zeta Y\Big] \begin{bmatrix} U^{-T}(D + \zeta Y^T Y) U^{-1} & -U^{-T} \\ -U^{-1} & 0 \end{bmatrix} \begin{bmatrix} S^T \\ \zeta Y^T \end{bmatrix},$$

where $S_k = [s_{k-\tilde{m}}, \ldots, s_k]$, $Y_k = [y_{k-\tilde{m}}, \ldots, y_k]$, $D_k = \mathrm{diag}[b_{k-\tilde{m}}, \ldots, b_k]$, $(U_k)_{i,j} = (S_k^T Y_k)_{i,j}$ for $i \leq j$, $(U_k)_{i,j} = 0$ otherwise (an upper triangular matrix), $k \geq 0$, or in the form, also given in [1]

$$H_+ = S U^{-T} D U^{-1} S^T + \zeta \Big(I - S U^{-T} Y^T\Big)\Big(I - Y U^{-1} S^T\Big), \tag{1.5}$$

thus direction vector can be calculated efficiently without computing of H_+, see [1].

For S^TY nonsingular and any $\bar{H} \in \mathcal{R}^{N\times N}$, the BFGS update formula (1.3) can be easily generalized to the following block version

$$H_+ = S(S^TY)^{-1}S^T + P_S^T\bar{H}P_S\,, \quad P_S = I - Y(S^TY)^{-1}S^T, \tag{1.6}$$

which satisfies the QN conditions $H_+Y = S$, i.e. for the whole block of stored difference vectors. This generalization of the BFGS update of $\bar{H}$ was derived by Schnabel [16] for S^TY and $\bar{H}$ symmetric positive definite, using a variational approach, and by Hu and Storey [7] for quadratic functions, using corrections for the exact line search. Both in [16] and in [7], some modifications of matrices Y (and also S in [7]) are proposed with intent to replace S^TY by a symmetric positive definite matrix. Note that these modifications disturb the QN conditions from previous iterations.

Formula (1.6) is not directly applicable to general functions, since it does not guarantee that the corresponding direction vectors are descent. To overcome this difficulty and at the same time utilize the advantageous properties of the block BFGS update in limited-memory context, in each iteration we determine $n \ge 1$ and split matrices S and Y in such a way that $S = [S_{[1]}, \dots, S_{[n]}]$, $Y = [Y_{[1]}, \dots, Y_{[n]}]$, where all blocks $S_{[i]}^TY_{[i]}$ are positive definite, i.e. matrices $S_{[i]}^TY_{[i]} + Y_{[i]}^TS_{[i]}$ are symmetric positive definite, $i = 1, \dots, n$. Afterwards we replace the BNS formula (1.5) by n successive updates of an initial VM matrix H_I (ζI for the BNS method (1.5)) using a modification of the block BFGS update (1.6) with matrices $S_{[i]}$, $Y_{[i]}$, $i = 1, \dots, n$, instead of S, Y (the block BNS method, see Section 4). Obviously, for $n = m$ we obtain the BNS method. The question how to form suitable blocks $S_{[i]}, Y_{[i]}$ will be discussed in Section 5. Numerical results indicate that this approach can improve results significantly compared to the BNS and L-BFGS method.

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

In Section 2 we derive the block BFGS update for general functions, present its properties and modifications and show some similarities to the corrected BFGS update, see [18] and [17]. 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 investigate the block BNS method and derive a convenient formula similar to (1.5) to represent the resultant VM matrix and a related formula for efficient calculation of the direction vector. The corresponding 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 will denote by $\|\cdot\|_F$ the Frobenius matrix norm, by $\|\cdot\|$ the spectral matrix norm, by $|\cdot|$ the size of both scalars and vectors (the Euclidean vector norm) and by $[A]_{n_1}^{n_2}$ the principal submatrix of A with both row and column indices of entries from n_1 to n_2.

2 The block BFGS update

Using a variational approach, we will derive the block BFGS update (1.6) with $\bar{H} = (1/2)(H + H^T)$ for general functions, investigate its generalized form and show some connections with methods based on vector corrections.

2.1 Derivation and basic properties

To derive the basic variant of the block BFGS update, given by Theorem 2.2, we utilize Theorem 2.1, which is a block version (with S, Y instead of s, y) of Corollary 2.3 in [3].

Lemma 2.1. *Suppose that matrix $J \in \mathcal{R}^{N\times m}$ has a full rank, $u \in \mathcal{R}^m$ and $x^* = J(J^TJ)^{-1}u$. Then x^* is the unique solution to $\min_{x\in\mathcal{R}^N} |x|$ s.t. $J^Tx = u$.*

Proof. Obviously $J^Tx^* = u$. Let $x' = x^* + v$ and $J^Tx' = u$ for some $v \in \mathcal{R}^N$. Then $J^Tv = 0$, thus $|x'|^2 = u^T(J^TJ)^{-1}u + |v|^2$, which yields the desired conclusion. □

Theorem 2.1. *Let $S, Y \in \mathcal{R}^{N\times m}$, $A, W_L, W_R \in \mathcal{R}^{N\times N}$, W_L, W_R nonsingular and let matrix Y have a full rank. Then the unique solution to*

$$\min_{A_N\in\mathcal{R}^{N\times N}} ||W_L^{-1}(A_N - A)W_R^{-1}||_F \quad s.t. \quad A_NY = S \tag{2.1}$$

is

$$A_N = AP_V + S(V^TY)^{-1}V^T, \quad V = W_R^TW_RY, \quad P_V = I - Y(V^TY)^{-1}V^T. \tag{2.2}$$

Proof. We denote $\Omega = W_L^{-1}(A_N - A)W_R^{-1} \triangleq [\omega_1, \ldots, \omega_N]^T$ and $J = W_RY$. Since $J^T\Omega^T = (\Omega W_RY)^T = (A_NY - AY)^TW_L^{-T}$, problem (2.1) can be rewritten

$$\min_{\omega_i\in\mathcal{R}^N} \sum_{i=1}^{N} |\omega_i|^2 \quad s.t. \quad J^T\Omega^T = (S - AY)^TW_L^{-T}.$$

Denoting $[u_1, \ldots, u_N] = (S - AY)^TW_L^{-T}$, this can be broken up into N disjoint problems

$$\min_{\omega_i\in\mathcal{R}^N} |\omega_i|^2 \quad s.t. \quad J^T\omega_i = u_i, \quad i = 1, \ldots, N.$$

Using Lemma 2.1 (J has obviously full rank), we get $\Omega^T = J(J^TJ)^{-1}(S - AY)^TW_L^{-T}$, i.e.

$$\begin{aligned} W_L^{-1}(A_N - A)W_R^{-1} &= \Omega = W_L^{-1}(S - AY)(J^TJ)^{-1}J^T, \\ A_N - A &= (S - AY)(J^TJ)^{-1}J^TW_R, \end{aligned}$$

which gives (2.2) by $J^TW_R = V^T$ and $J^TJ = V^TY$. □

Since matrix A_N is meant as an approximation of the inverse Hessian matrix, thus near to a symmetric matrix, and since the nearest symmetric matrix to any matrix M in Frobenius norm is $\frac{1}{2}(M+M^T)$ by Lemma 2.2, which is Lemma 4.1 in [3], we will construct matrix A^* satisfying $A^*Y = S$ nearest to the subspace of symmetric matrices in $\mathcal{R}^{N\times N}$. Following the approach used in [3], we will find $\lim_{i\to\infty} A_i$, where in view of Theorem 2.1

$$A_0 = AP_V + S(V^TY)^{-1}V^T, \quad A_{i+1} = (1/2)(A_i + A_i^T)P_V + S(V^TY)^{-1}V^T, \quad i = 0, 1, \ldots \tag{2.3}$$

Lemma 2.2. *Let $M \in \mathcal{R}^{N\times N}$. Then matrix $M_S = \frac{1}{2}(M + M^T)$ is the unique solution to*

$$\min_{M_S\in\mathcal{R}^{N\times N}} ||M_S - M||_F \quad s.t. \quad M_S = M_S^T.$$

Theorem 2.2. *Let the assumptions of Theorem 2.1 be satisfied and sequence $\{A_i\}_{i=0}^{\infty}$ be defined by (2.3). Then*

$$\lim_{i\to\infty} A_i = (1/2)P_V^T(A + A^T)P_V + V(V^TY)^{-T}S^TP_V + S(V^TY)^{-1}V^T \triangleq A^*. \tag{2.4}$$

Moreover, if $T \in \mathcal{R}^{m\times m}$ is nonsingular and $V = ST$, we obtain the block BFGS update (1.6) with $H_+ = A^*$, $\bar{H} = (1/2)(A + A^T)$.

Proof. First we prove

$$A_i = (1/2^i) Z + A^*, \qquad Z = V(V^TY)^{-T}(A^TY - S)^T P_V, \tag{2.5}$$

$i = 1, 2, \dots,$ by induction. For $i = 1$ it is true, since from (2.3) we get

$$\begin{aligned} A_1 - S(V^TY)^{-1}V^T &= \tfrac{1}{2}\Big(A_0 + A_0^T\Big)P_V = \tfrac{1}{2}\Big(AP_V + P_V^T A^T + V(V^TY)^{-T}S^T\Big)P_V \\ &= \tfrac{1}{2}\Big(I - P_V^T\Big)AP_V + \tfrac{1}{2}P_V^T(A + A^T)P_V + \tfrac{1}{2}V(V^TY)^{-T}S^TP_V \\ &= \tfrac{1}{2}V(V^TY)^{-T}(A^TY - S)^TP_V + V(V^TY)^{-T}S^TP_V + \tfrac{1}{2}P_V^T(A + A^T)P_V \end{aligned}$$

by $V^TP_V = 0$, $P_V^2 = P_V$ and $I - P_V^T = V(V^TY)^{-T}Y^T$.

Suppose that (2.5) is true for some $i \geq 1$. By $V^TP_V = 0$ and $P_V^2 = P_V$ we obtain

$$(A^*)^TP_V = \frac{1}{2}P_V^T(A + A^T)P_V + V(V^TY)^{-T}S^TP_V = A^* - S(V^TY)^{-1}V^T = A^*P_V$$

and $ZP_V = Z$, $Z^TP_V = 0$, which by (2.3) and (2.5) yields

$$A_{i+1} = \frac{1}{2}(A_i + A_i^T)P_V + S(V^TY)^{-1}V^T = \frac{1}{2^{i+1}}Z + (A^* - S(V^TY)^{-1}V^T) + S(V^TY)^{-1}V^T,$$

i.e. (2.5) is true for $i+1$, which completes the induction. Consequently, this implies (2.4).

Finally, let $V = ST$. Then $P_V = I - Y(T^TS^TY)^{-1}T^TS^T = I - Y(S^TY)^{-1}S^T = P_S$, $S^TP_V = 0$ and $A^* = \frac{1}{2}P_S^T(A + A^T)P_S + S(S^TY)^{-1}S^T$. □

In the sequel, we give some properties of the block BFGS update, similar to the well-known properties of the standard BFGS update. We will investigate the generalized form of (1.6)

$$H_+ = S(S^TY\,C)^{-1}S^T + \Big(I - S(S^TY)^{-T}Y^T\Big)\,\bar{H}\,\Big(I - Y(S^TY)^{-1}S^T\Big), \tag{2.6}$$

where we consider any nonsingular matrices $\bar{H} \in \mathcal{R}^{N\times N}$ and $S^TY, C \in \mathcal{R}^{m\times m}$. First we prove the following lemmas.

Lemma 2.3. *Let $W_i \in \mathcal{R}^{\mu\times\nu}$, $\mu > 0, \nu > 0$, $i = 1, \dots, 4$, and $W_4^TW_3 = I$. Then*

$$\det\Big(I + W_1W_2^T - W_3W_4^T\Big) = \det\Big(W_2^TW_3\Big) \,.\, \det\Big(W_4^TW_1\Big). \tag{2.7}$$

Proof. Denoting $\alpha = \det\Big(I + W_1W_2^T - W_3W_4^T\Big)$, we can write

$$\begin{vmatrix} I & W_2^T & 0 \\ -W_1 & I & W_3 \\ 0 & W_4^T & I \end{vmatrix} = \begin{vmatrix} I & W_2^T & 0 \\ 0 & I + W_1W_2^T & W_3 \\ 0 & W_4^T & I \end{vmatrix} = \begin{vmatrix} I & W_2^T & 0 \\ 0 & I + W_1W_2^T - W_3W_4^T & W_3 \\ 0 & 0 & I \end{vmatrix} = \alpha.$$

The initial determinant on the left can be rewritten in another way

$$\alpha = \begin{vmatrix} I & W_2^T & 0 \\ -W_1 & I & W_3 \\ W_4^TW_1 & 0 & I - W_4^TW_3 \end{vmatrix} = \begin{vmatrix} I & W_2^T & 0 \\ -W_1 & I & W_3 \\ W_4^TW_1 & 0 & 0 \end{vmatrix} = \begin{vmatrix} I & W_2^T & -W_2^TW_3 \\ -W_1 & I & 0 \\ W_4^TW_1 & 0 & 0 \end{vmatrix}$$

by $W_4^TW_3 = I$. To obtain the desired result, we interchange the third block column of the last determinant, multiplied by -1, and the first block column. □

Lemma 2.4. *Let matrix $A \in \mathcal{R}^{N\times N}$ be positive definite. Then A is nonsingular and matrix A^{-1} is also positive definite.*

Proof. Obviously, A is nonsingular. Let $q \in \mathcal{R}^N$, $q \neq 0$, $p = A^{-1}q$. Then $q^TA^{-1}q = p^TA^Tp = p^TAp > 0$. □

Theorem 2.3. *Let matrices S^TY and C be nonsingular and let matrix H_+ be given by (2.6). Then $H_+Y = SC^{-1}$ and*

(a) if we replace matrices S, Y in (2.6) by ST_S, YT_Y with $T_S, T_Y \in \mathcal{R}^{m\times m}$ nonsingular, then the corresponding matrix H_+ can be also written in the form (2.6) with C replaced by $T_Y\, C\, T_S^{-1}$,

(b) for $\bar{H}, H_+$ and $S^T\bar{B}S$ nonsingular and $\bar{B} = \bar{H}^{-1}$, matrix $B_+ = H_+^{-1}$ is given by

$$B_+ = \bar{B} - \bar{B}S(S^T\bar{B}S)^{-1}S^T\bar{B} + YC(S^TY)^{-T}Y^T, \tag{2.8}$$

(c) for $\bar{H}, H_+$ and $S^T\bar{B}S$ nonsingular, the determinant of B_+ is

$$\det B_+ = \det \bar{B} \,.\, \det(S^TY\,C)/\det(S^T\bar{B}S). \tag{2.9}$$

(d) for $\bar{H}$ and $S^TY\,C$ positive definite, also matrix H_+ is positive definite.

Proof. (a) We simply replace S, Y by ST_S, YT_Y in (2.6) and rewrite the relation.

(b) Denoting $B'_+ = \bar{B} - \bar{B}S(S^T\bar{B}S)^{-1}S^T\bar{B} + YC(S^TY)^{-T}Y^T$, we have $B'_+S = YC$, thus we get from (2.6)

$$\begin{aligned} B'_+H_+ &= YC(S^TY\,C)^{-1}S^T + \left(B'_+ - YC(S^TY)^{-T}Y^T\right)\bar{H}\left(I - Y(S^TY)^{-1}S^T\right) \\ &= Y(S^TY)^{-1}S^T + \left(I - \bar{B}S(S^T\bar{B}S)^{-1}S^T\right)\left(I - Y(S^TY)^{-1}S^T\right) \\ &= I - \bar{B}S(S^T\bar{B}S)^{-1}S^T + \bar{B}S(S^T\bar{B}S)^{-1}S^TY(S^TY)^{-1}S^T = I. \end{aligned}$$

(c) Using (2.8) and Lemma 2.3 with $W_1 = \bar{H}YC$, $W_2^T = (S^TY)^{-T}Y^T$, $W_3 = S(S^T\bar{B}S)^{-1}$, $W_4^T = S^T\bar{B}$, we get

$$\begin{aligned} \det B_+ &= \det \bar{B} \,.\, \det\left(I - S(S^T\bar{B}S)^{-1}S^T\bar{B} + \bar{H}YC(S^TY)^{-T}Y^T\right) \\ &= \det \bar{B} \,.\, \det\left(I - W_3W_4^T + W_1W_2^T\right) = \det \bar{B} \,.\, \det\left((S^T\bar{B}S)^{-1}\right).\, \det(S^TY\,C)\,. \end{aligned}$$

(d) Let $q \in \mathcal{R}^N$, $q \neq 0$. If $S^Tq \neq 0$, then $q^TH_+q \geq q^TS(S^TY\,C)^{-1}S^Tq > 0$ by Lemma 2.4, otherwise $q^TH_+q = q^T\bar{H}q > 0$. □

Corollary 2.1. *Let matrices S^TY and $\bar{H}$ be nonsingular, $\bar{H}$ symmetric, $\bar{B} = \bar{H}^{-1}$, and let matrices H_+ given by (2.6) with $C{=}I$ (i.e. by (1.6)) and $S^T\bar{B}S$ be nonsingular. Then*

$$\begin{aligned} \left(\frac{1}{2}\left(H_+ + H_+^T\right)\right)^{-1} &= \bar{B} - \bar{B}S\left(S^T\bar{B}S\right)^{-1}S^T\bar{B} + Y\left(\frac{1}{2}\left(S^TY + Y^TS\right)\right)^{-1}Y^T, &(2.10) \\ \frac{1}{2}\left(B_+ + B_+^T\right) &= \bar{B} - \bar{B}S\left(S^T\bar{B}S\right)^{-1}S^T\bar{B} + \frac{1}{2}Y\left((S^TY)^{-1} + (Y^TS)^{-1}\right)Y^T, &(2.11) \\ \det\left(\frac{1}{2}\left(H_+ + H_+^T\right)\right)^{-1} &= \det \bar{B} \,.\, \det\left(\frac{1}{2}\left((S^TY)^{-1} + (Y^TS)^{-1}\right)\right)^{-1} \Big/ \det(S^T\bar{B}S)\,, &(2.12) \\ \det \frac{1}{2}\left(B_+ + B_+^T\right) &= \det \bar{B} \,.\, \det \frac{1}{2}\left(S^TY + Y^TS\right)/\det(S^T\bar{B}S)\,. &(2.13) \end{aligned}$$

Proof. From (1.6) we obtain $\frac{1}{2}(H_+ + H_+^T) = \frac{1}{2}S((S^TY)^{-1} + (Y^TS)^{-1})S^T + P_S^T\bar{H}P_S$, which can be written in the form (2.6) with $C = (\frac{1}{2}(I + (Y^TS)^{-1}S^TY))^{-1}$ and H_+ replaced by $\frac{1}{2}(H_+ + H_+^T)$. Using Theorem 2.3 (b)–(c), we get (2.10)–(2.12). Since (2.11) can be written in the form (2.8) with $C = \frac{1}{2}(I + (S^TY)^{-1}Y^TS)$ and B_+ replaced by $\frac{1}{2}(B_+ + B_+^T)$, Theorem 2.3 (c) yields (2.13). □

2.2 Connections with methods based on vector corrections

The following lemma shows some relations between the block BFGS update in the form (2.6) and the repeated BFGS update with modified difference vectors.

Lemma 2.5. *Let* $S \triangleq [\check{S}, s]$, $Y \triangleq [\check{Y}, y]$, *matrices* $\check{S}^T\check{Y}$, $\check{T}_S, \check{T}_Y \in \mathcal{R}^{\tilde{m}\times\tilde{m}}$ *be nonsingular,* $\check{C}=\check{T}_Y\check{T}_S^{-1}$, $\check{P}=I-\check{Y}(\check{S}^T\check{Y})^{-1}\check{S}^T$, $\tilde{s}=\check{P}^Ts$, $\tilde{y}=\check{P}y$, $\tilde{b}=\tilde{s}^T\tilde{y}\neq 0$ *and matrix* H_+ *be given by*

$$H_+ = (1/\tilde{b})\,\tilde{s}\tilde{s}^T + \tilde{P}^T\check{H}\tilde{P}, \quad \tilde{P} = I-(1/\tilde{b})\,\tilde{y}\tilde{s}^T, \quad \check{H} = \check{S}(\check{S}^T\check{Y}\,\check{C})^{-1}\check{S}^T + \check{P}^T\bar{H}\check{P}. \tag{2.14}$$

Then matrix H_+ *can be written in the form (2.6) with* $C=T_YT_S^{-1}$, *where*

$$T_S = \begin{bmatrix} \check{T}_S & -(\check{S}^T\check{Y})^{-T}\check{Y}^Ts \\ & 1 \end{bmatrix}, \quad T_Y = \begin{bmatrix} \check{T}_Y & -(\check{S}^T\check{Y})^{-1}\check{S}^Ty \\ & 1 \end{bmatrix} \tag{2.15}$$

(the upper block triangular matrices), and $\check{S}^TB_+\tilde{s} = \check{S}^T\check{H}^{-T}\tilde{s} = 0$ *holds. Moreover, if matrices* $\bar{H}$ *and* $\check{S}^T\check{Y}\,\check{C}$ *are symmetric, then also* $\check{H}$, H_+ *and* $S^TY\,C$ *are symmetric.*

Proof. Setting $\tilde{S}=S\,T_S, \tilde{Y}=Y\,T_Y$, we obtain $\tilde{S}=[\check{S}\check{T}_S, s-\check{S}(\check{S}^T\check{Y})^{-T}\check{Y}^Ts\,]=[\check{S}\check{T}_S, \check{P}^Ts\,]=[\check{S}\check{T}_S, \tilde{s}\,]$ and similarly $\tilde{Y}=[\check{Y}\check{T}_Y, \check{P}y\,]=[\check{Y}\check{T}_Y, \tilde{y}\,]$, which yields

$$\tilde{S}^T\tilde{Y} = \begin{bmatrix} \check{T}_S^T\check{S}^T\check{Y}\check{T}_Y & \check{T}_S^T\check{S}^T\check{P}y \\ s^T\check{P}\check{Y}\check{T}_Y & \tilde{b} \end{bmatrix} = \begin{bmatrix} \check{T}_S^T\check{S}^T\check{Y}\check{T}_Y & 0 \\ 0 & \tilde{b} \end{bmatrix} \tag{2.16}$$

by $\check{P}^T\check{S} = \check{P}\check{Y} = 0$. Using (2.16), we get

$$\tilde{S}(\tilde{S}^T\tilde{Y})^{-1}\tilde{S}^T = \check{S}(\check{S}^T\check{Y}\check{C})^{-1}\check{S}^T + \frac{1}{\tilde{b}}\tilde{s}\tilde{s}^T, \quad \tilde{Y}(\tilde{S}^T\tilde{Y})^{-1}\tilde{S}^T = \check{Y}(\check{S}^T\check{Y})^{-1}\check{S}^T + \frac{1}{\tilde{b}}\tilde{y}\tilde{s}^T. \tag{2.17}$$

From (2.14) we obtain successively

$$\begin{aligned} H_+ &= (1/\tilde{b})\,\tilde{s}\tilde{s}^T + \tilde{P}^T\check{H}\tilde{P} \;=\; (1/\tilde{b})\,\tilde{s}\tilde{s}^T + \tilde{P}^T\check{S}(\check{S}^T\check{Y}\,\check{C})^{-1}\check{S}^T\tilde{P} + \tilde{P}^T\check{P}^T\bar{H}\check{P}\tilde{P} \\ &= (1/\tilde{b})\tilde{s}\tilde{s}^T + \check{S}(\check{S}^T\check{Y}\,\check{C})^{-1}\check{S}^T + \left(I-(1/\tilde{b})\tilde{s}\tilde{y}^T\right)\check{P}^T\bar{H}\check{P}\left(I-(1/\tilde{b})\tilde{y}\tilde{s}^T\right) \\ &= (1/\tilde{b})\tilde{s}\tilde{s}^T + \check{S}(\check{S}^T\check{Y}\,\check{C})^{-1}\check{S}^T + \left(\check{P}^T-(1/\tilde{b})\tilde{s}\tilde{y}^T\check{P}^T\right)\bar{H}\left(\check{P}-(1/\tilde{b})\check{P}\tilde{y}\tilde{s}^T\right) \\ &= \check{S}(\check{S}^T\check{Y}\check{C})^{-1}\check{S}^T + \frac{1}{\tilde{b}}\tilde{s}\tilde{s}^T + \left(I-\check{S}(\check{S}^T\check{Y})^{-T}\check{Y}^T - \frac{1}{\tilde{b}}\tilde{s}\tilde{y}^T\right)\bar{H}\left(I-\check{Y}(\check{S}^T\check{Y})^{-1}\check{S}^T - \frac{1}{\tilde{b}}\tilde{y}\tilde{s}^T\right) \end{aligned}$$

by $\check{P}^2=\check{P}$ and $\check{P}^T\check{S}=0$, which yields $\tilde{P}^T\check{S}=\check{S}-(1/\tilde{b})\tilde{s}\tilde{y}^T\check{P}^T\check{S}=\check{S}$. Using (2.17), we have

$$H_+ = \tilde{S}(\tilde{S}^T\tilde{Y})^{-1}\tilde{S}^T + \left(I-\tilde{S}(\tilde{S}^T\tilde{Y})^{-T}\tilde{Y}^T\right)\bar{H}\left(I-\tilde{Y}(\tilde{S}^T\tilde{Y})^{-1}\tilde{S}^T\right), \tag{2.18}$$

which can be written in the form (2.6) with $C=T_YT_S^{-1}$ by Theorem 2.3*(a)*.

Since $H_+\tilde{y}=\tilde{s}$ and $\check{H}\check{Y}=\check{S}\,\check{C}^{-1}$, i.e. $\check{S}=\check{H}\check{Y}\check{C}$ by (2.14), we have $\check{S}^TB_+\tilde{s}=\check{S}^T\tilde{y}=\check{S}^T\check{P}y=0$ and $\check{S}^T\check{H}^{-T}\tilde{s}=\check{C}^T\check{Y}^T\tilde{s}=\check{C}^T\check{Y}^T\check{P}^Ts=0$ by $\check{P}^T\check{S}=\check{P}\check{Y}=0$.

If matrices $\bar{H}$ and $\check{S}^T\check{Y}\,\check{C}$ are symmetric, then also matrices $\check{H}$ and $\check{T}_S^T\check{S}^T\check{Y}\check{T}_Y$ are symmetric by (2.14) and $\check{T}_S^T\check{S}^T\check{Y}\check{T}_Y=\check{T}_S^T(\check{S}^T\check{Y}\,\check{C})\check{T}_S$ holds, which yields the symmetry of matrices H_+, $\tilde{S}^T\tilde{Y}$ and $S^TY\,C$ by (2.14), (2.16) and equality $S^TY\,C=T_S^{-T}(\tilde{S}^T\tilde{Y})\,T_S^{-1}$. $\square$

In view of relations $\check{S}^TB_+\tilde{s} = \check{S}^T\check{H}^{-T}\tilde{s} = 0$, we can regard transformations $s \to \tilde{s} = \check{P}^Ts = s - \check{S}(\check{S}^T\check{Y})^{-1}\check{Y}^Ts$, $y \to \tilde{y} = \check{P}y = y - \check{Y}(\check{S}^T\check{Y})^{-1}\check{S}^Ty$ (or transformations $S \to \tilde{S}$, $Y \to \tilde{Y}$) in Lemma 2.5 as corrections from previous iterations for conjugacy, which shows some connections with methods [18] and [17], where similar corrections are also used.

Although variational characterizations of such corrections are significant mainly for quadratic functions, see Section 3, the following theorem indicates that we can expect good properties of the block BFGS update also for functions similar to quadratic.

Theorem 2.4. *Let* $S \triangleq [\check{S}, s]$, $Y \triangleq [\check{Y}, y]$, $\ddot{s} = s + \check{S}\sigma$, $\ddot{y} = y + \check{Y}\sigma$, $\sigma \in \mathcal{R}^{\tilde{m}}$, $\tilde{m} \geq 1$, $\tilde{s} = \check{P}^Ts$, $\tilde{y} = \check{P}\,y$, $\check{P} = I - \check{Y}(\check{S}^T\check{Y})^{-1}\check{S}^T$, $\ddot{b} = \ddot{s}^T\ddot{y}$, $\tilde{b} = \tilde{s}^T\tilde{y}$ *and matrix* S^TY *be symmetric positive definite. Then* $\ddot{b} \geq \tilde{b} = s^T\tilde{y} > 0$ *for any* $\sigma \in \mathcal{R}^{\tilde{m}}$. *Moreover, let* $\check{H}$ *be any* *nonsingular matrix satisfying* $\check{H}\check{Y} = \check{S}$ *and* $\ddot{a} = \ddot{y}^T \check{H} \ddot{y}$, $\tilde{a} = \tilde{y}^T \check{H} \tilde{y}$. *If we define matrix* $\ddot{H}_+$ *by*

$$\ddot{H}_+ = (1/\ddot{s}^T\ddot{y})\,\ddot{s}\ddot{s}^T + \Big(I - (1/\ddot{s}^T\ddot{y})\,\ddot{s}\ddot{y}^T\Big)\check{H}\Big(I - (1/\ddot{s}^T\ddot{y})\,\ddot{y}\ddot{s}^T\Big) \tag{2.19}$$

and if a symmetric positive definite matrix $\bar{G}$ *satisfying* $\bar{G}S = Y$ *is given, then within* $\sigma \in \mathcal{R}^{\tilde{m}}$ *we have* $\bar{G}\ddot{s} = \ddot{y}$, $\ddot{a} \ge \tilde{a}$ *and*

$$\|\bar{G}^{1/2}\ddot{H}_+\bar{G}^{1/2} - I\|_F^2 = (1 - \ddot{a}/\ddot{b})^2 - 2\,|\,\bar{G}^{1/2}(\ddot{s} - \check{H}\ddot{y})\,|^2/\ddot{b} + \|\bar{G}^{1/2}\check{H}\bar{G}^{1/2} - I\|_F^2; \tag{2.20}$$

this value is minimized by the choice $\ddot{s} = \tilde{s}$, $\ddot{y} = \tilde{y}$; *for this choice and* $\check{H}$ *given by (2.14) (with* $\check{C} = I$*), matrices* $\ddot{H}_+$ *and* H_+ *given by (1.6) are identical.*

Proof. From $\tilde{s} = \check{P}^T s$ and $\tilde{y} = \check{P}\,y$ we obtain $\tilde{b} = s^T\tilde{y}$ by $\check{P}^2 = \check{P}$, which gives

$$\tilde{b} = b - s^T\check{Y}(\check{S}^T\check{Y})^{-1}\check{S}^T y\,. \tag{2.21}$$

From $\ddot{s} = s + \check{S}\sigma$ and $\ddot{y} = y + \check{Y}\sigma$ we get $\ddot{b} = b + 2y^T\check{S}\sigma + \sigma^T\check{S}^T\check{Y}\sigma$, which can be written

$$\ddot{b} = b - y^T\check{S}(\check{S}^T\check{Y})^{-1}\check{S}^T y + \Big(\sigma + (\check{S}^T\check{Y})^{-1}\check{S}^T y\Big)^T \check{S}^T\check{Y}\Big(\sigma + (\check{S}^T\check{Y})^{-1}\check{S}^T y\Big)\,. \tag{2.22}$$

Since matrices $S^TY, \check{S}^T\check{Y}$ are symmetric positive definite by assumption, we have $\tilde{b} > 0$ by Theorem 2.22 in [5] and $\check{S}^T y = \check{Y}^T s$. Comparing (2.22) and (2.21), we can see that always $\ddot{b} \ge \tilde{b}$ holds.

Let $\bar{G}S = Y$ with $\bar{G}$ symmetric positive definite. Then obviously $\bar{G}\ddot{s} = \ddot{y}$ and $\bar{G}\tilde{s} = \tilde{y}$. Denoting $w = \bar{G}^{1/2}\ddot{s}$, $\tilde{w} = \bar{G}^{1/2}\tilde{s}$, $W = \bar{G}^{1/2}\check{H}\bar{G}^{1/2}$, $W_+ = \bar{G}^{1/2}H_+\bar{G}^{1/2}$ and $M = I - W$, we have $|w|^2 = \ddot{b} \ge \tilde{b} = |\tilde{w}|^2 > 0$ and (2.19) can be written in the form

$$W_+ = (1/|w|^2)ww^T + PWP = I - PMP, \qquad P = I - (1/|w|^2)ww^T, \tag{2.23}$$

by $\bar{G}\ddot{s} = \ddot{y}$ and $P^2 = P$. In view of the fact that the trace of a product of two square matrices is independent of the order of the multiplication, from (2.23) we obtain

$$\begin{aligned} \|I - W_+\|_F^2 &= \|PMP\|_F^2 = \mathrm{Tr}(PMPM) = \mathrm{Tr}\Big(\Big[M - (1/|w|^2)ww^TM\Big]^2\Big) \\ &= \|M\|_F^2 - \mathrm{Tr}\Big(ww^TM^2 + Mww^TM - \Big[w^TMw/|w|^2\Big]\,ww^TM\Big)/|w|^2 \\ &= \|M\|_F^2 - 2|Mw|^2/|w|^2 + (w^TMw)^2/|w|^4, \end{aligned} \tag{2.24}$$

i.e. (2.20) by $Mw = \bar{G}^{1/2}(\ddot{s} - \check{H}\ddot{y})$ and $w^TMw = \ddot{b} - \ddot{a}$. In view of $\check{H}\check{Y} = \check{S}$ by assumption and in view of $s^T\check{Y} = y^T\check{S}$ by symmetry of S^TY, values $|Mw|$ and w^TMw are independent of σ, as we can see from

$$\begin{aligned} \ddot{s} - \check{H}\ddot{y} &= s + \check{S}\sigma - \check{H}y - \check{H}\check{Y}\sigma = s - \check{H}y\,, \\ \ddot{b} - \ddot{a} &= (\ddot{s} - \check{H}\ddot{y})^T\ddot{y} = (s - \check{H}y)^T(y + \check{Y}\sigma) = s^Ty - y^T\check{H}y + s^T\check{Y}\sigma - y^T\check{H}\check{Y}\sigma\,. \end{aligned}$$

In view of (2.24) we can write $\|I - W_+\|_F^2 = \varphi(|\tilde{w}|^2/|w|^2)$, where function

$$\varphi(\xi) = \xi^2(\tilde{w}^TM\tilde{w})^2/|\tilde{w}|^4 - 2\,\xi|M\tilde{w}|^2/|\tilde{w}|^2 + \|M\|_F^2 \tag{2.25}$$

is nonincreasing on $[0,1]$, since $\varphi'(\xi)/2 = \xi(\tilde{w}^TM\tilde{w})^2/|\tilde{w}|^4 - |M\tilde{w}|^2/|\tilde{w}|^2 \le 0$ for $\xi \in [0,1]$ by the Schwarz inequality. Therefore value $\|I - W_+\|_F^2$ is minimized by the choice $\ddot{s} = \tilde{s}$, $\ddot{y} = \tilde{y}$, which gives $|w| = |\tilde{w}|$, i.e. maximizes $|\tilde{w}|/|w|$. For this choice and matrix $\check{H}$ given by (2.14) with $\check{C} = I$, matrices $\ddot{H}_+$ and H_+ given by (1.6) are identical by Lemma 2.5, where for $\check{C} = I$ and S^TY symmetric we have $T_S = T_Y$, thus $C = I$.

The rest follows immediately from $\ddot{a} = (\ddot{a} - \ddot{b}) + \ddot{b} = (\tilde{a} - \tilde{b}) + \ddot{b} \ge (\tilde{a} - \tilde{b}) + \tilde{b}$. □

Seemingly, in accordance with Theorem 2.4, the block BFGS update should be advantageous in case that matrix S^TY is positive definite and near to symmetric (e.g. near to a local minimum). 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 block BFGS update, which for S^TY symmetric corresponds to update (2.19) of $\check{H}$ with

$$\ddot{s} = \tilde{s}, \qquad \ddot{y} = \tilde{y} \tag{2.26}$$

by Lemma 2.5, also the standard BFGS update of $\check{H}$, i.e. (2.19) with

$$\ddot{s} = s, \qquad \ddot{y} = y, \tag{2.27}$$

or a special update of $\check{H}$, i.e. (2.19) with

$$\ddot{s} = s - (s^Ty_-/b_-)s_-, \qquad \ddot{y} = y - (y^Ts_-/b_-)y_-, \tag{2.28}$$

which can be more robust than the block BFGS update. In Section 4 we show how it can be used within the block BNS method. The question how to choose a suitable update will be discussed in Section 5. For functions similar to quadratic, the choice (2.28) can also be characterized variationally:

Theorem 2.5. *Let* $\ddot{S} = [s_-, s]$, $\ddot{Y} = [y_-, y]$, $\hat{s} = s-(s^Ty_-/b_-)s_-$, $\hat{y} = y-(y^Ts_-/b_-)y_-$, $\ddot{s}=s-\alpha s_-$, $\ddot{y}=y-\alpha y_-$, $\alpha\in\mathcal{R}$, $\hat{b}=\hat{s}^T\hat{y}$, $\ddot{b}=\ddot{s}^T\ddot{y}$. *Then* $\hat{b}=s^T\hat{y}$*; if matrix* $\ddot{S}^T\ddot{Y}$ *is symmetric positive definite, then* $\ddot{b}\geq\hat{b}>0$ *for any* $\alpha\in\mathcal{R}$*. Moreover, let* $\check{H}$ *be any nonsingular matrix satisfying* $\check{H}\ddot{Y}=\ddot{S}$ *and* $\ddot{a} = \ddot{y}^T\check{H}\ddot{y}$*. If we define matrix* $\ddot{H}_+$ *by (2.19) and a symmetric positive definite matrix* $\bar{G}$ *satisfying* $\bar{G}\ddot{S}=\ddot{Y}$ *is given, then within* $\alpha\in\mathcal{R}$ *relations* $\bar{G}\ddot{s}=\ddot{y}$ *and (2.20) hold. Besides, values* $\ddot{a}$ *and (2.20) are minimized by the choice* $\ddot{s}= \hat{s}$, $\ddot{y} = \hat{y}$.

Proof. We have $\hat{s}^T\hat{y} = s^T\hat{y} - (s^Ty_-/b_-)s_-^T[y-(s_-^Ty/b_-)y_-] = s^T\hat{y}$. If matrix $\ddot{S}^T\ddot{Y}$ is symmetric positive definite, then $s^Ty_-=y^Ts_-$, value $\ddot{b} = b - 2\alpha s^Ty_- + \alpha^2 b_-$ is minimized by $\alpha = s^Ty_-/b_-$, i.e. by $\ddot{s}= \hat{s}$, $\ddot{y} = \hat{y}$ and the minimum value is $\hat{b} = s^T\hat{y} = b - s^Ty_-\, s_-^Ty/b_-$ with $\hat{b} > 0$ by Theorem 2.22 in [5].

Let $\bar{G}\ddot{S} = \ddot{Y}$ with $\bar{G}$ symmetric positive definite. Setting $\sigma = (0,\dots,0,-\alpha)^T$ and replacing $\tilde{s}$ by $\hat{s}$, $\tilde{y}$ by $\hat{y}$, $\tilde{b}$ by $\hat{b}$ and $\tilde{a}$ by $\hat{y}^T\check{H}\hat{y}$, we can proceed in the same way as in the proof of Theorem 2.4. □

3 Results for quadratic functions

In this section we suppose that f is a quadratic function with a symmetric positive definite Hessian G (thus $GS = Y$ and matrix $S^TY = S^TGS$ is symmetric) and show optimality of the block BFGS method and a role of unit stepsizes, which are very frequent, not only for quadratic functions. Here we consider only the G-conjugacy of vectors.

The following theorem shows that the block BFGS update gives the best improvement of convergence in some sense for linearly independent direction vectors.

Theorem 3.1. *Let* f *be quadratic function* $f(x) = \frac{1}{2}(x-\bar{x})^TG(x-\bar{x})$, $\bar{x} \in \mathcal{R}^N$*, with a symmetric positive definite matrix* G*, the columns of matrix* S *be linearly independent and let* $\hat{S}_i = [s_{k-\tilde{m}},\dots,s_i]$, $\hat{Y}_i = [y_{k-\tilde{m}},\dots,y_i]$, $\hat{P}_i = I-\hat{Y}_i(\hat{S}_i^T\hat{Y}_i)^{-1}\hat{S}_i^T$, $i = k-\tilde{m},\dots,k$, $\ddot{s}_i=s_i+\hat{S}_{i-1}\sigma_{i-1}$, $\ddot{y}_i=y_i+\hat{Y}_{i-1}\sigma_{i-1}$, $\sigma_{i-1}\in\mathcal{R}^{i-1}$, $\tilde{s}_i=\hat{P}_{i-1}^Ts_i$, $\tilde{y}_i=\hat{P}_{i-1}y_i$, $i=k-\tilde{m}+1,\dots,k$, $\ddot{s}_{k-\tilde{m}} = \tilde{s}_{k-\tilde{m}} = s_{k-\tilde{m}}$, $\ddot{y}_{k-\tilde{m}} = \tilde{y}_{k-\tilde{m}} = y_{k-\tilde{m}}$*. Then matrices* $\hat{S}_i^T\hat{Y}_i$ *are symmetric positive* *definite and* $\ddot{s}_i^T\ddot{y}_i \ge \tilde{s}_i^T\tilde{y}_i > 0$, $i = k-\tilde{m},\dots,k$. *Moreover, if matrix* $\bar{H}$ *is symmetric positive definite and if we define matrix* H_+ *by (1.6) and matrix* $\ddot{H}_+ = \ddot{H}_{k+1}$ *by* $\ddot{H}_{k-\tilde{m}} = \bar{H}$ *and*

$$\ddot{H}_{i+1} = (1/\ddot{s}_i^T\ddot{y}_i)\,\ddot{s}_i\ddot{s}_i^T + \Big(I-(1/\ddot{s}_i^T\ddot{y}_i)\,\ddot{s}_i\ddot{y}_i^T\Big)\ddot{H}_i\Big(I-(1/\ddot{s}_i^T\ddot{y}_i)\,\ddot{y}_i\ddot{s}_i^T\Big), \tag{3.1}$$

$i = k-\tilde{m},\dots,k$, *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 the choice* $\ddot{s}_i = \tilde{s}_i$, $\ddot{y}_i = \tilde{y}_i$, $i = k-\tilde{m}+1,\dots,k$.

Proof. Since the columns of S are linearly independent, matrices $\hat{S}_i^T\hat{Y}_i = \hat{S}_i^T G\hat{S}_i$, $i = k-\tilde{m},\dots,k$, are symmetric positive definite and we can set $\hat{H}_{i+1} = \hat{S}_i(\hat{S}_i^T\hat{Y}_i)^{-1}\hat{S}_i^T + \hat{P}_i^T\bar{H}\hat{P}_i$, $i = k-\tilde{m},\dots,k$. Using successively Theorem 2.4 with $\bar{G} = G$ and $\hat{S}_i$, $\hat{Y}_i$, $\hat{H}_i$, $\hat{H}_{i+1}$ instead of S, Y, $\check{H}$, H_+, $i = k-\tilde{m}+1,\dots,k$, we get that values $\|G^{1/2}\ddot{H}_{i+1}G^{1/2}-I\|_F$ are minimized and matrices $\ddot{H}_{i+1}$ and $\hat{H}_{i+1}$ are identical and symmetric positive definite for the choice $\ddot{s}_i = \tilde{s}_i$, $\ddot{y}_i = \tilde{y}_i$, $i = k-\tilde{m}+1,\dots,k$, when $\ddot{H}_{k+1} = \hat{H}_{k+1} = H_+$. □

In Section 2 we mentioned the similarity to the methods based on corrections from previous iterations for conjugacy. The following theorem, similar to Theorem 3.3 in [18], shows that in two successive iterations with VM matrices H, H_+ obtained by the block BFGS updates, the only unit stepsize is sufficient to have all stored direction vectors from previous iterations conjugate with vector s_+.

Theorem 3.2. *Let* f *be a quadratic function* $f(x) = \frac{1}{2}(x-\bar{x})^T G(x-\bar{x})$, $\bar{x} \in \mathcal{R}^N$, *with a symmetric positive definite matrix* G, $S \triangleq [\check{S}, s]$, $Y \triangleq [\check{Y}, y]$, H, H_+ *be symmetric positive definite matrices satisfying* $H\check{Y} = \check{S}$, $H_+\check{Y} = \check{S}$, $d = -Hg$, $d_+ = -H_+g_+$ *and let* $t = 1$, *i.e.* $s = d$. *Then* $\check{S}^T y_+ = \check{Y}^T s_+ = 0$, *i.e. all columns of* $\check{S}$ *are conjugate with vector* s_+.

Proof. In view of $\check{S}^T y = \check{S}^T Gs = \check{Y}^T s$ we obtain

$$-\check{Y}^T d_+ = -\check{S}^T B_+ d_+ = \check{S}^T g_+ = \check{S}^T(y+g) = \check{Y}^T s + \check{S}^T g = \check{Y}^T(s+Hg) = 0,$$

which immediately gives $\check{Y}^T s_+ = \check{S}^T G s_+ = \check{S}^T y_+ = 0$. □

Vectors $\check{S}^T y_+, \check{Y}^T s_+$ from the preceding iteration are used for functions near to quadratic in the process of the suitable update formula selection, see Section 5.

4 The block BNS method

In this section we will derive some representations of matrix H_+ which generalize the BNS formula (1.5). For this purpose, we split matrices S, Y in such a way that $S = [S_{[1]},\dots,S_{[n]}]$, $Y = [Y_{[1]},\dots,Y_{[n]}]$, $n \ge 1$, with all blocks $S_{[i]}^T Y_{[i]}$ positive definite ($S_{[i]}^T Y_{[i]} + Y_{[i]}^T S_{[i]}$ symmetric positive definite), $i = 1,\dots,n$, and use the theory in Section 2 for matrices $S_{[i]}, Y_{[i]}$ instead of S, Y. We consider arbitrary nonsingular matrices $H_I, C_{[i]}$, although only the choice $H_I = \zeta I$, $\zeta > 0$, $C_{[i]} = I$, $i = 1,\dots,n$, is used in our numerical experiments.

To construct matrix H_+, in view of Theorem 2.2 we set $H_{[1]} = H_I$, $H_+ = H_{[n+1]}$, where

$$H_{[i+1]} = S_{[i]}(S_{[i]}^T Y_{[i]} C_{[i]})^{-1} S_{[i]}^T + \frac{1}{2} P_{[i]}^T\Big(H_{[i]} + H_{[i]}^T\Big)P_{[i]}, \quad P_{[i]} = I - Y_{[i]}(S_{[i]}^T Y_{[i]})^{-1} S_{[i]}^T, \tag{4.1}$$

for $S_{[i]}^T Y_{[i]} C_{[i]}$ nonsingular, $i = 1, \dots, n$. Note that matrices $H_{[i]}$, $i=1,\dots,n+1$, have only a theoretical significance and are not formed explicitly and that here we denote by U_i a different matrix than in Section 1.

In the process of splitting matrices S, Y, we start with matrices $S_{[n]}, Y_{[n]}$ to have maximum of the latest QN conditions satisfied. Thus to test positive definiteness of blocks $S_{[i]}^T Y_{[i]}$, $i = n, \dots, 1$, we use a factorization arranged in reverse order compared to the usual LU factorization, see the following lemma, which converts the problem of factorization to the same problem of a smaller dimension, and Section 5 for details.

Lemma 4.1. *Suppose that $A, R, L \in \mathcal{R}^{\mu\times\mu}$, $\mu>0$, $u, v \in \mathcal{R}^{\mu}$, $\alpha\in\mathcal{R}$, $\alpha\neq 0$,*

$$\bar{A} = \begin{bmatrix} A & u \\ v^T & \alpha \end{bmatrix}, \quad \bar{R} = \begin{bmatrix} R & u \\ & \alpha \end{bmatrix}, \quad \bar{L} = \begin{bmatrix} L & \\ (1/\alpha)\, v^T & 1 \end{bmatrix}. \tag{4.2}$$

Then to get $\bar{A}=\bar{R}\bar{L}$, 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.

Further, if $\det[\bar{A}]_i^{\mu+1}\neq 0$, $i=1,\dots,\mu+1$, then we can continue in this way repeatedly, i.e. the whole factorization process is well defined, and the result factorization is unique.

Proof. Let $A - (1/\alpha)uv^T = RL$. Using relations for $\bar{R}, \bar{L}$ in (4.2), we obtain

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

Using Theorem 2.22 in [5], we get (a). Let matrix $\bar{A}$ be positive definite. Then also $\bar{A}^{-1}$ is positive definite by Lemma 2.4, obviously together with all its principal submatrices. Similarly we deduce that $\alpha > 0$ (principal submatrix of $\bar{A}$). Since matrix $A - (1/\alpha)uv^T$ (Schur complement of entry α in $\bar{A}$) is the inverse of a principal submatrix of $\bar{A}^{-1}$ by Theorem 1.23 in [5], it is positive definite by Lemma 2.4. Finally, the existence and uniqueness of the factorization under the conditions above follows from Theorem 1.24 in [5], considering the rows and columns of matrices $\bar{A}, \bar{R}, \bar{L}$ arranged in reverse order. $\square$

The following lemma generalizes the approach used in the proof of Theorem 2.2 in [1].

Lemma 4.2. *Let $\mu, \nu > 0$, $S_L, Y_L \in \mathcal{R}^{N\times\mu}$, $S_R, Y_R \in \mathcal{R}^{N\times\nu}$, $S_C = [S_L, S_R]$, $Y_C = [Y_L, Y_R]$, $U_L, E_L \in \mathcal{R}^{\mu\times\mu}$, $C_R \in \mathcal{R}^{\nu\times\nu}$, $\bar{H}_I \in \mathcal{R}^{N\times N}$, $U_L, S_R^T Y_R$ and C_R nonsingular,*

$$H_L = S_L U_L^{-T} E_L U_L^{-1} S_L^T + (I - S_L U_L^{-1} Y_L^T)\bar{H}_I(I - Y_L U_L^{-1} S_L^T)\,, \tag{4.3}$$

$$H_C = S_R(S_R^T Y_R C_R)^{-1} S_R^T + P_R^T H_L P_R, \quad P_R = I - Y_R(S_R^T Y_R)^{-1} S_R^T\,. \tag{4.4}$$

Then matrix H_C can be written in the form

$$H_C = S_C U_C^{-T} E_C U_C^{-1} S_C^T + (I - S_C U_C^{-T} Y_C^T)\bar{H}_I(I - Y_C U_C^{-1} S_C^T), \tag{4.5}$$

where

$$U_C = \begin{bmatrix} U_L & S_L^T Y_R \\ & S_R^T Y_R \end{bmatrix}, \qquad E_C = \begin{bmatrix} E_L & \\ & Y_R^T S_R C_R^{-1} \end{bmatrix} \tag{4.6}$$

(matrix U_C is upper block triangular, E_C block diagonal).

Proof. From (4.3) – (4.4) we obtain

$$H_C = S_R(S_R^T Y_R C_R)^{-1} S_R^T + P_R^T S_L U_L^{-T} E_L U_L^{-1} S_L^T P_R + K^T \bar{H}_I K, \tag{4.7}$$

where

$$\begin{aligned} K &= \left(I - Y_L U_L^{-1} S_L^T\right)\left(I - Y_R (S_R^T Y_R)^{-1} S_R^T\right) \\ &= I - Y_L U_L^{-1} S_L^T - Y_R (S_R^T Y_R)^{-1} S_R^T + Y_L U_L^{-1} S_L^T Y_R (S_R^T Y_R)^{-1} S_R^T \\ &= I - \left[Y_L, Y_R\right] \begin{bmatrix} U_L^{-1} & -U_L^{-1} S_L^T Y_R (S_R^T Y_R)^{-1} \\ & (S_R^T Y_R)^{-1} \end{bmatrix} \begin{bmatrix} S_L^T \\ S_R^T \end{bmatrix} = I - Y_C\, U_C^{-1} S_C^T . \end{aligned}$$

Using this representation of U_C^{-1}, we obtain $[\,I \;\; 0\,]\, U_C^{-1} = [\,U_L^{-1}, -U_L^{-1} S_L^T Y_R (S_R^T Y_R)^{-1}]$, therefore

$$U_L^{-1} S_L^T P_R = U_L^{-1} S_L^T - U_L^{-1} S_L^T Y_R (S_R^T Y_R)^{-1} S_R^T = [\,I \;\; 0\,]\, U_C^{-1} S_C^T \tag{4.8}$$

by (4.7). Similarly $[\,0 \;\; I\,]\, U_C^{-1} = [\,0, (S_R^T Y_R)^{-1}]$, i.e. $(S_R^T Y_R)^{-1} S_R^T = [\,0 \;\; I\,]\, U_C^{-1} S_C^T$, thus

$$\begin{aligned} S_R (S_R^T Y_R C_R)^{-1} S_R^T &= S_R\, (S_R^T Y_R)^{-T} Y_R^T S_R C_R^{-1} (S_R^T Y_R)^{-1} S_R^T \\ &= S_C\, U_C^{-T} [\,0 \;\; I\,]^T Y_R^T S_R\, C_R^{-1}\, [\,0 \;\; I\,]\, U_C^{-1} S_C^T \\ &= S_C U_C^{-T} \begin{bmatrix} 0 & \\ & Y_R^T S_R C_R^{-1} \end{bmatrix} U_C^{-1} S_C^T . \end{aligned} \tag{4.9}$$

To get (4.5), it suffices to use (4.6) – (4.9) together with $K = I - Y_C\, U_C^{-1} S_C^T$. □

The following theorem describes a basic version of the block BNS method.

Theorem 4.1. *Let $S = [S_{[1]}, \dots, S_{[n]}]$, $Y = [Y_{[1]}, \dots, Y_{[n]}]$, $n \ge 1$, $\mathcal{S}_i = [S_{[1]}, \dots, S_{[i]}]$, $\mathcal{Y}_i = [Y_{[1]}, \dots, Y_{[i]}]$, matrices $S_{[i]}^T Y_{[i]}\, C_{[i]}$ be nonsingular, matrices $H_{[i+1]}$ be given by (4.1), $i = 1, \dots, n$, and $H_{[1]} = H_I$. Then*

$$H_{[i+1]} = \mathcal{S}_i U_i^{-T} E_i U_i^{-1} \mathcal{S}_i^T + \frac{1}{2}\left(I - \mathcal{S}_i U_i^{-T} \mathcal{Y}_i^T\right)\left(H_I + H_I^T\right)\left(I - \mathcal{Y}_i U_i^{-1} \mathcal{S}_i^T\right), \tag{4.10}$$

where (the upper block triangular matrix)

$$U_i = \begin{bmatrix} S_{[1]}^T Y_{[1]} & \cdots & S_{[1]}^T Y_{[i-1]} & S_{[1]}^T Y_{[i]} \\ & \ddots & \vdots & \vdots \\ & & S_{[i-1]}^T Y_{[i-1]} & S_{[i-1]}^T Y_{[i]} \\ & & & S_{[i]}^T Y_{[i]} \end{bmatrix}, \tag{4.11}$$

$$E_i = \operatorname{diag}\left[\frac{1}{2}\left(\Sigma_1 + \Sigma_1^T\right), \dots, \frac{1}{2}\left(\Sigma_{i-1} + \Sigma_{i-1}^T\right), \Sigma_i\right], \quad \Sigma_j = Y_{[j]}^T S_{[j]} C_{[j]}^{-1}, \tag{4.12}$$

$i, j = 1, \dots, n$.

Proof. We will proceed by induction on i. For $i = 1$, update (4.1) can be written

$$H_{[2]} = S_{[1]} (S_{[1]}^T Y_{[1]})^{-T} (Y_{[1]}^T S_{[1]}\, C_{[1]}^{-1}) (S_{[1]}^T Y_{[1]})^{-1} S_{[1]}^T + \frac{1}{2} P_{[1]}^T \left(H_I + H_I^T\right) P_{[1]} ,$$

i.e. (4.10) with $U_1 = S_{[1]}^T Y_{[1]}$, $E_1 = Y_{[1]}^T S_{[1]}\, C_{[1]}^{-1} = \Sigma_1$.

Suppose that (4.10) – (4.12) hold for some $i < n$ and set $\bar{H}_{[i+1]} = \frac{1}{2}(H_{[i+1]} + H_{[i+1]}^T)$ and

$$H_{[i+2]} = S_{[i+1]} (S_{[i+1]}^T Y_{[i+1]}\, C_{[i+1]})^{-1} S_{[i+1]}^T + P_{[i+1]}^T \bar{H}_{[i+1]} P_{[i+1]} \tag{4.13}$$

in view of (4.1). Since $\bar{H}_{[i+1]}$ can be written in the form (4.10) with E_i replaced by $\bar{E}_i = \frac{1}{2}(E_i + E_i^T)$, we can use Lemma 4.2 with $S_L = \mathcal{S}_i$, $Y_L = \mathcal{Y}_i$, $S_R = S_{[i+1]}$, $Y_R = Y_{[i+1]}$, $C_R = C_{[i+1]}$, $S_C = \mathcal{S}_{i+1}$, $Y_C = \mathcal{Y}_{i+1}$, $U_L = U_i$, $E_L = \bar{E}_i$, $\bar{H}_I = \frac{1}{2}(H_I + H_I^T)$, $H_L = \bar{H}_{[i+1]}$, $H_C = H_{[i+2]}$. Denoting $E_{i+1} = \operatorname{diag}\left[\bar{E}_i\, , \Sigma_{i+1}\right]$, we obtain (4.10) with $H_{[i+1]}$, $\mathcal{S}_i$, $\mathcal{Y}_i$, U_i, E_i replaced by $H_{[i+2]}, \mathcal{S}_{i+1}, \mathcal{Y}_{i+1}, U_{i+1}, E_{i+1}$ and the induction is established with $i+1$ replacing i. □

Similar representations of H_+ can be derived also for update (2.19) with the choice (2.28), which we sometimes use instead of the last update (4.1), see Section 5.

Corollary 4.1. *Let* $H_{[1]}=H_I$, $n\geq 1$, $S=[S_{[1]},\ldots,S_{[n]}]\triangleq[\check S,s]$, $Y=[Y_{[1]},\ldots,Y_{[n]}]\triangleq[\check Y,y]$, $S_{[n]}\triangleq[\check S_{[n]},s]$, $Y_{[n]}\triangleq[\check Y_{[n]},y]$, $\hat s=s-\alpha s_-$, $\hat y=y-\beta y_-$, $\alpha=s^Ty_-/b_-$, $\beta=y^Ts_-/b_-$, $\hat s^T\hat y\neq 0$, $\hat S=[\check S,\hat s]$, $\hat Y=[\check Y,\hat y]$, *matrices* $S_{[i]}^TY_{[i]}C_{[i]}$, $i=1,\ldots,n-1$, $\check S_{[n]}^T\check Y_{[n]}\check C_{[n]}$ *be nonsingular, matrices* $H_{[i]}$, $i=2,\ldots,n$, *be given by update (4.1) and matrix* H_+ *by*

$$H_+=(1/\hat s^T\hat y)\,\hat s\hat s^T+\Big(I-(1/\hat s^T\hat y)\,\hat s\hat y^T\Big)\check H\Big(I-(1/\hat s^T\hat y)\,\hat y\hat s^T\Big), \tag{4.14}$$

$$\check H=\check S_{[n]}\Big(\check S_{[n]}^T\check Y_{[n]}\check C_{[n]}\Big)^{-1}\check S_{[n]}^T+\frac{1}{2}\check P_{[n]}^T\Big(H_{[n]}+H_{[n]}^T\Big)\check P_{[n]},\quad \check P_{[n]}=I-\check Y_{[n]}\Big(\check S_{[n]}^T\check Y_{[n]}\Big)^{-1}\check S_{[n]}^T. \tag{4.15}$$

Then

$$\begin{aligned} H_+ &= \hat S\hat U^{-T}\hat E\hat U^{-1}\hat S^T+\tfrac12\Big(I-\hat S\hat U^{-T}\hat Y^T\Big)\Big(H_I+H_I^T\Big)\Big(I-\hat Y\hat U^{-1}\hat S^T\Big) && (4.16)\\ &= S\tilde U^{-T}\tilde E\tilde U^{-1}S^T+\tfrac12\Big(I-S\tilde U^{-T}Y^T\Big)\Big(H_I+H_I^T\Big)\Big(I-Y\tilde U^{-1}S^T\Big), && (4.17)\end{aligned}$$

where

$$\hat U=\begin{bmatrix}\check U & \check S^T\hat y\\ & s^T\hat y\end{bmatrix},\quad \hat E=\begin{bmatrix}\check E & \\ & s^T\hat y\end{bmatrix},\quad \tilde U=\begin{bmatrix}\check U & \check S^Ty\\ \alpha\tilde u^T & s^Ty\end{bmatrix},\quad \tilde E=\begin{bmatrix}\check E & \beta\tilde w\\ \beta\tilde v^T & \kappa\end{bmatrix}, \tag{4.18}$$

$$\check U=\begin{bmatrix} S_{[1]}^TY_{[1]} & \ldots & S_{[1]}^TY_{[n-1]} & S_{[1]}^T\check Y_{[n]}\\ & \ddots & \vdots & \vdots\\ & & S_{[n-1]}^TY_{[n-1]} & S_{[n-1]}^T\check Y_{[n]}\\ & & & \check S_{[n]}^T\check Y_{[n]}\end{bmatrix},\quad \check E=\begin{bmatrix}\frac12(\Sigma_1+\Sigma_1^T) & & \\ & \ddots & & \\ & & \frac12(\Sigma_{n-1}+\Sigma_{n-1}^T) & \\ & & & \check\Sigma_n\end{bmatrix} \tag{4.19}$$

(matrices $\hat U,\check U$ *are upper block triangular,* $\hat E,\check E$ *block diagonal),* $\Sigma_i=Y_{[i]}^TS_{[i]}C_{[i]}^{-1}$, $i=1,\ldots,n-1$, $\check\Sigma_n=\check Y_{[n]}^T\check S_{[n]}\check C_{[n]}^{-1}$, $\tilde u^T=s_-^T\check Y_{[n]}$ *is the last row of* $\check U$, $\tilde v^T$ *the last row of* $\check E$, $\tilde w$ *the last column of* $\check E$ *and* $\kappa=\beta^2\tilde v_{m-1}+s^T\hat y$. *If* $\check C_{[n]}=I$ *then* $\tilde w=\tilde u$, $\tilde v$ *is the last column of* $\operatorname{diag}[S_{[1]}^TY_{[1]},\ldots,S_{[n-1]}^TY_{[n-1]},\check S_{[n]}^T\check Y_{[n]}]$ *and* $\kappa=b+\beta(\beta-\alpha)b_-$.

Proof. We have $\hat s^T\hat y=s^T\hat y$ by Theorem 2.5. Using Theorem 4.1 for updates (4.1), $i=1,\ldots,n-1$ (i.e. for updates (4.1), $i=1,\ldots,n$, with $S_{[n]},Y_{[n]}$ replaced by $\check S_{[n]},\check Y_{[n]}$ or with $S=\mathcal S_n, Y=\mathcal Y_n$ replaced by $\check S,\check Y$), followed by (4.15), we get

$$\check H=\check S\,\check U^{-T}\check E\,\check U^{-1}\check S^T+\frac12\Big(I-\check S\,\check U^{-T}\check Y^T\Big)\Big(H_I+H_I^T\Big)\Big(I-\check Y\check U^{-1}\check S^T\Big) \tag{4.20}$$

and to prove (4.16), it suffices to use Lemma 4.2 for update (4.14) of $\check H$, i.e. with $S_L=\check S$, $Y_L=\check Y$, $S_R=\hat s$, $Y_R=\hat y$, $C_R=1$, $S_C=\hat S$, $Y_C=\hat Y$, $U_L=\check U$, $E_L=\check E$, $\bar H_I=\frac12(H_I+H_I^T)$, $H_L=\check H$, $H_C=H_+$.

Since we can write $\hat S=[\check S,s-\alpha s_-]=ST_S$, $\hat Y=[\check Y,y-\beta y_-]=YT_Y$, where

$$T_S=\operatorname{diag}\left[I,\begin{bmatrix}1 & -\alpha\\ 0 & 1\end{bmatrix}\right]\in\mathcal R^{m\times m},\quad T_Y=\operatorname{diag}\left[I,\begin{bmatrix}1 & -\beta\\ 0 & 1\end{bmatrix}\right]\in\mathcal R^{m\times m}, \tag{4.21}$$

(4.16) yields (4.17) with $\tilde U=T_S^{-T}\hat UT_Y^{-1}$, $\tilde E=T_Y^{-T}\hat ET_Y^{-1}$. After rearrangement we obtain

$$\begin{aligned}\tilde U &= T_S^{-T}\begin{bmatrix}\check U & \check S^T\hat y\\ & s^T\hat y\end{bmatrix}\begin{bmatrix}I & & \\ & 1 & \beta\\ & & 1\end{bmatrix}=\begin{bmatrix}I & & \\ & 1 & \\ & \alpha & 1\end{bmatrix}\begin{bmatrix}\check U & \beta\check S^Ty_-+\check S^T\hat y\\ & s^T\hat y\end{bmatrix}=\begin{bmatrix}\check U & \check S^Ty\\ \alpha\tilde u^T & s^Ty\end{bmatrix},\\ \tilde E &= T_Y^{-T}\begin{bmatrix}\check E & \\ & s^T\hat y\end{bmatrix}\begin{bmatrix}I & & \\ & 1 & \beta\\ & & 1\end{bmatrix}=\begin{bmatrix}I & & \\ & 1 & \\ & \beta & 1\end{bmatrix}\begin{bmatrix}\check E & \beta\tilde w\\ & s^T\hat y\end{bmatrix}=\begin{bmatrix}\check E & \beta\tilde w\\ \beta\tilde v^T & \beta^2\tilde v_{m-1}+s^T\hat y\end{bmatrix}\end{aligned}$$

by $\beta\check S^Ty_-+\check S^T\hat y=\check S^Ty$, $\alpha s_-^Ty+s^T\hat y=\alpha\beta b_-+s^Ty-\alpha\beta b_-=s^Ty$ and $\tilde v_{m-1}=\tilde w_{m-1}$, where obviously $\tilde w=\tilde u$ and $\tilde v_{m-1}=b_-$ for $\check C_{[n]}=I$. □

To estimate the benefit of the block BFGS update in Section 5, we use value $\tilde a=\tilde y^T\check H\,\tilde y$, $\tilde y=\check P_{[n]}y$ (see Theorem 2.4 and relation (4.15)), which can be calculated with a negligible increase in the number of arithmetic operations:

Corollary 4.2. *Let* $H_{[1]} = H_I$, $n \geq 1$, $S = [S_{[1]}, \ldots, S_{[n]}] \triangleq [\check{S}, s]$, $Y = [Y_{[1]}, \ldots, Y_{[n]}] \triangleq [\check{Y}, y]$, $S_{[n]} \triangleq [\check{S}_{[n]}, s]$, $Y_{[n]} \triangleq [\check{Y}_{[n]}, y]$, *matrices* $S_{[i]}^T Y_{[i]} C_{[i]}$ *be nonsingular, matrices* $H_{[i]}$ *be given by update (4.1),* $i = 2, \ldots, n$, *with* $H_I = \zeta I$, $\zeta > 0$, *matrices* $\check{H}$, $\check{P}_{[n]}$, $\check{U}, \check{E}$ *by (4.15) and (4.19) and let* $\tilde{y} = \check{P}_{[n]} y$, $\tilde{a} = \tilde{y}^T \check{H}\, \tilde{y}$. *Then*

$$\tilde{a} = \zeta |y|^2 + y^T \check{S}\, \check{U}^{-T} \Big(\bar{E} + \zeta \check{Y}^T \check{Y} \Big) \check{U}^{-1} \check{S}^T y - 2\, \zeta y^T \check{S}\, \check{U}^{-T} \check{Y}^T y \,, \tag{4.22}$$

where

$$\bar{E} \;=\; \operatorname{diag}\Big[\tfrac{1}{2} \Big(\Sigma_1 + \Sigma_1^T \Big), \ldots, \tfrac{1}{2} \Big(\Sigma_{n-1} + \Sigma_{n-1}^T \Big), 0 \Big] \,, \quad \Sigma_i = Y_{[i]}^T S_{[i]} C_{[i]}^{-1}, \tag{4.23}$$

$i = 1, \ldots, n-1$, *and the dimension of the null matrix is equal to* $\dim(\check{Y}_{[n]}^T \check{S}_{[n]} \check{C}_{[n]}^{-1})$.

Proof. In the same way as in the proof of Theorem 4.1 we get (4.20). Furthermore, since $\check{S}_{[n]}^T \check{P}_{[n]} = 0$ and $\check{P}_{[n]}^2 = \check{P}_{[n]}$, from (4.15) we obtain

$$\check{P}_{[n]}^T \check{H}\, \check{P}_{[n]} = \check{H} - \check{S}_{[n]} \Big(\check{S}_{[n]}^T \check{Y}_{[n]} \check{C}_{[n]} \Big)^{-1} \check{S}_{[n]}^T \,. \tag{4.24}$$

In a similar way as in the proof of Lemma 4.2 (relation (4.9)) we prove (the dimension of the null principal submatrix is equal to $\dim \Sigma_1 + \ldots + \dim \Sigma_{n-1}$)

$$\check{S}_{[n]} \Big(\check{S}_{[n]}^T \check{Y}_{[n]} \check{C}_{[n]} \Big)^{-1} \check{S}_{[n]}^T = \check{S}\, \check{U}^{-T} \begin{bmatrix} 0 & \\ & \check{Y}_{[n]}^T \check{S}_{[n]} \check{C}_{[n]}^{-1} \end{bmatrix} \check{U}^{-1} \check{S}^T \,;$$

for $H_I = \zeta I$, $\zeta > 0$, this together with (4.19) – (4.20) and (4.24) immediately gives

$$\check{P}_{[n]}^T \check{H}\, \check{P}_{[n]} = \check{S} \check{U}^{-T} \bar{E} \check{U}^{-1} \check{S}^T + \zeta \Big(I - \check{S} \check{U}^{-T} \check{Y}^T \Big) \Big(I - \check{Y} \check{U}^{-1} \check{S}^T \Big), \tag{4.25}$$

and subsequently yields (4.22) by $\tilde{a} = y^T \Big(\check{P}_{[n]}^T \check{H} \check{P}_{[n]} \Big) y$. □

Using representation (4.10) or (4.17), the direction vector and an auxiliary vector $Y^T H_+ g_+$ (see Section 5) can be calculated effectively, similarly as for the BNS method, see [1]. E.g. for $H = \zeta I$ and matrix $H_+ = H_{[n+1]}$ given by (4.10) we have (omitting index n)

$$-H_+ g_+ = -\zeta g_+ - S \Big[U^{-T} \Big((E + \zeta Y^T Y) U^{-1} S^T g_+ - \zeta Y^T g_+ \Big) \Big] + Y \Big[\zeta U^{-1} S^T g_+ \Big], \tag{4.26}$$

$$Y^T H_+ g_+ = \zeta Y^T g_+ + Y^T S \Big[U^{-T} \Big((E + \zeta Y^T Y) U^{-1} S^T g_+ - \zeta Y^T g_+ \Big) \Big] - Y^T Y \Big[\zeta U^{-1} S^T g_+ \Big], \tag{4.27}$$

where in brackets we multiply by low-order matrices. Similarly for H_+ given by (4.17)

$$-H_+ g_+ = -\zeta g_+ - S \Big[\tilde{U}^{-T} \Big((\tilde{E} + \zeta Y^T Y) \tilde{U}^{-1} S^T g_+ - \zeta Y^T g_+ \Big) \Big] + Y \Big[\zeta \tilde{U}^{-1} S^T g_+ \Big]; \tag{4.28}$$

from this we easily obtain the corresponding representation of $Y^T H_+ g_+$.

In comparison with the BNS method, here $U, \tilde{U}$ are not triangular matrices generally, which can complicate calculations. Using factorization $S_{[i]}^T Y_{[i]} = R_{[i]} L_{[i]}$, $i = 1, \ldots, n$, where $R_{[i]}$ and $L_{[i]}^T$ are upper triangular matrices, and denoting $L_D = \operatorname{diag}[\, L_{[1]}, \ldots, L_{[n]} \,]$, $\mathcal{E} = L_D^{-T} (E + \zeta Y^T Y)$, we can set $U = U_T\, L_D$, where $U_T = U\, L_D^{-1}$ and L_D^T are upper triangular matrices, and rewrite (4.26) and $\mathcal{E}$ in the form

$$-H_+ g_+ = -\zeta g_+ - S \Big[U_T^{-T} \Big(\mathcal{E} L_D^{-1} U_T^{-1} S^T g_+ - \zeta L_D^{-T} Y^T g_+ \Big) \Big] + Y \Big[\zeta L_D^{-1} U_T^{-1} S^T g_+ \Big], \tag{4.29}$$

$$\mathcal{E} = \operatorname{diag}\Big[\frac{1}{2} \Big(R_{[1]}^T + L_{[1]}^{-T} S_{[1]}^T Y_{[1]} \Big), \ldots, \frac{1}{2} \Big(R_{[n-1]}^T + L_{[n-1]}^{-T} S_{[n-1]}^T Y_{[n-1]} \Big), R_{[n]}^T \Big] + \zeta L_D^{-T} Y^T Y. \tag{4.30}$$

In case of matrix $\tilde{U}$ we can proceed similarly. If we denote

$$\hat{U}_{[n]} = \begin{bmatrix} \check{S}_{[n]}^T \check{Y}_{[n]} & \check{S}^T \hat{y} \\ & s^T \hat{y} \end{bmatrix}, \quad \tilde{U}_{[n]} = \begin{bmatrix} \check{S}_{[n]}^T \check{Y}_{[n]} & \check{S}^T y \\ \alpha \tilde{u}^T & s^T y \end{bmatrix}, \quad \tilde{E}_{[n]} = \begin{bmatrix} \check{\Sigma}_n & \beta \tilde{w} \\ \beta \tilde{v}^T & \kappa \end{bmatrix} \tag{4.31}$$

(submatrices of $\hat{U}, \tilde{U}, \tilde{E}$ in (4.18)), we can see that for $S_{[n]}^T Y_{[n]}$ positive definite (thus $s^T \hat{y} = b - s^T y_- \, y^T s_- / b_- > 0$ and $\check{S}_{[n]}^T \check{Y}_{[n]}$ positive definite) a RL factorization exists for $\hat{U}_{[n]}$ by Lemma 4.1, because all its principal minors are obviously nonzero. Since they do not change by adding to a row (column) a multiple of another row (column), we can also factorize matrix $\tilde{U}_{[n]}$ and write $\tilde{U}_{[n]} = \tilde{R}_{[n]} \tilde{L}_{[n]}$, where $\tilde{R}_{[n]}, \tilde{L}_{[n]}^T$ are upper triangular matrices. Denoting $\tilde{L}_D = \mathrm{diag}[\, L_{[1]}, \ldots, L_{[n-1]}, \tilde{L}_{[n]} \,]$, $\tilde{\mathcal{E}} = \tilde{L}_D^{-T} (\tilde{E} + \zeta Y^T Y)$, we can set $\tilde{U} = \tilde{U}_T \tilde{L}_D$, where $\tilde{U}_T = \tilde{U} \tilde{L}_D^{-1}$ and $\tilde{L}_D^T$ are upper triangular matrices, and rewrite (4.28) and $\tilde{\mathcal{E}}$:

$$-H_+ g_+ = -\zeta g_+ - S\Big[\tilde{U}_T^{-T}\Big(\tilde{\mathcal{E}} \tilde{L}_D^{-1} \tilde{U}_T^{-1} S^T g_+ - \zeta \tilde{L}_D^{-T} Y^T g_+\Big)\Big] + Y\Big[\zeta \tilde{L}_D^{-1} \tilde{U}_T^{-1} S^T g_+\Big], \tag{4.32}$$

$$\tilde{\mathcal{E}} = \mathrm{diag}\Big[\frac{1}{2}\Big(R_{[1]}^T + L_{[1]}^{-T} S_{[1]}^T Y_{[1]}\Big), \ldots, \frac{1}{2}\Big(R_{[n-1]}^T + L_{[n-1]}^{-T} S_{[n-1]}^T Y_{[n-1]}\Big), \tilde{L}_{[n]}^{-T} \tilde{E}_{[n]}\Big] + \zeta \tilde{L}_D^{-T} Y^T Y. \tag{4.33}$$

Our experiments indicate, that this approach can also improve numerical results.

5 Implementation

Using results from the previous sections and assuming that $C_{[1]} = \ldots = C_{[n]} = \check{C}_{[n]} = I$ and $H_I = \zeta I$, $\zeta = b / y^T y > 0$, we will propose a suitable splitting of matrices S, Y, $S = [S_{[1]}, \ldots, S_{[n]}] = [\,\check{S}, s]$, $Y = [Y_{[1]}, \ldots, Y_{[n]}] = [\,\check{Y}, y]$, $n \in [1, m]$ and describe the corresponding algorithm. As we mentioned in Section 4, at first we form the submatrix $S_{[n]}^T Y_{[n]}$ to have maximum of the latest QN conditions satisfied.

In this connection, from now on we denote a set of indices j of vectors s_j, y_j which form matrices $S_{[i]}, Y_{[i]}$ by $\mathcal{I}_i$, a number of column of these matrices by $m_i \geq 1$, $i = 1, \ldots, n$, and a set of indices j of vectors s_j, y_j which correspond to entries of matrix $[S^T Y]_{\underline{\nu}}^{\overline{\nu}}$ (see Section 1), $1 \leq \underline{\nu} \leq \overline{\nu} \leq m$, by $\mathcal{I}_{\underline{\nu}}^{\overline{\nu}}$. Obviously, $\sum_{i=1}^n m_i = m$.

In accordance with the theory in Sections 2, 3 we should use the block BFGS update whenever an objective function is close to a quadratic function (e.g. near to a local minimum). Taking this into consideration, we find such positive definite (to have direction vectors descent) submatrices $S_{[i]}^T Y_{[i]}$ of the largest order, for which $\Delta_i \leq \delta_1$ for $i = n$, $\Delta_i \leq \delta_2$ otherwise, where numbers $\Delta_i = \max_{j_1, j_2 \in \mathcal{I}_i} \{(s_{j_1}^T y_{j_2} - s_{j_2}^T y_{j_1})^2 / (b_{j_2} b_{j_1})\}$ (zero for quadratic functions), can serve as a measure of the deviation from a quadratic function, $i = n, \ldots, 1$.

On the other hand, the use of this update can deteriorate stability, which is most noticeable in case of the last block $S_{[n]}^T Y_{[n]}$ if it is almost symmetric, i.e. $\Delta_i < \delta_3$. Therefore to select the suitable choice from (2.26)–(2.28) for such a block, we estimate the benefit of the block BFGS update in comparison with the corresponding BFGS updates, see below. If we regard this benefit as sufficient or if $m_n \leq 2$, we always use the choice (2.26), otherwise we denote $a_{i,j} = (S_{[n]}^T Y_{[n]})_{i,j}$, $i, j = 1, \ldots, m_n$ and calculate value

$$\theta = \sum_{i=1}^{m_n - 2} \sqrt{|a_{i,m_n} a_{m_n,i}|} / b \tag{5.1}$$

(this formula was chosen empirically), which can be also regarded as an estimate of the deviation f from a quadratic function and is equal to zero for quadratic function if $t_- = 1$, see Theorem 3.2. Subsequently, we use the choice (2.26) for $\theta < \delta_4$, (2.27) for $\theta > \delta_5$ or $\ddot{s}^T\ddot{y} > \delta_6$ and (2.28) otherwise, see Algorithm 5.1 and Procedure 5.3 for details.

It follows from the proof of Theorem 2.4 that $\|\bar{G}^{1/2}\ddot{H}_+\bar{G}^{1/2} - I\|_F^2 = \varphi(\xi)$, $\xi = \tilde{b}/\ddot{b} \in$ (0,1] for S^TY symmetric positive definite, where quadratic function φ given by (2.25) is nonincreasing on $[0,1]$, all its coefficients are independent of $\sigma \in \mathcal{R}^m$, $\varphi(0) = \|M\|_F^2$ corresponds to $\check{H}$, $\varphi(\tilde{b}/b)$ to the standard BFGS update of $\check{H}$, i.e. (2.19) with the choice (2.27) and $\varphi(1)$ to the block BFGS update of $\check{H}$, i.e. (2.19) with the choice (2.26). Although we cannot calculate either $\varphi(\xi)$ or $\varphi'(\xi)$, the following lemma shows that the ratio $b/\tilde{b}$ and a suitable estimate of the decrease of φ on $[\tilde{b}/b, 1]$ can be considered as good indicators of the benefit of the block BFGS update for S^TY near to symmetric.

Lemma 5.1. *Let we denote quantities $\tilde{a}, \tilde{b}$ as in Theorem 2.4, $\tilde{w}, M$ as in the proof of Theorem 2.4, $\xi_1 = \tilde{b}/b \in (0,1]$ and let function $\varphi(\xi)$ be given by (2.25). Then*

$$\varphi(\xi_1) - \varphi(1) \;\geq\; (1 - \tilde{a}/\tilde{b})^2(1 - \xi_1)^2, \tag{5.2}$$

$$[\varphi(0) - \varphi(\xi_1)]/[\varphi(0) - \varphi(1)] \;\leq\; \xi_1(2 - \xi_1). \tag{5.3}$$

Proof. Quadratic function (2.25) can be written in the form

$$\varphi(\xi) = \bar{c}\xi^2 - 2\bar{d}\xi + \|M\|_F^2, \quad \bar{c} = (\tilde{w}^TM\tilde{w}/|\tilde{w}|^2)^2 = (1-\tilde{a}/\tilde{b})^2, \quad \bar{d} = |M\tilde{w}|^2/\tilde{w}|^2. \tag{5.4}$$

Since $\bar{c} \leq \bar{d}$ by the Schwarz inequality, we obtain

$$\varphi(\xi_1) - \varphi(1) = \bar{c}(\xi_1^2 - 1) + 2\bar{d}(1 - \xi_1) \geq \bar{c}(1 - \xi_1)^2.$$

Denoting $\psi(t) = (t\xi_1 - \bar{c}\xi_1^2)/(t - \bar{c})$, $t \neq \bar{c}$, we have

$$[\varphi(0)-\varphi(\xi_1)]/[\varphi(0)-\varphi(1)] = (2\bar{d}\xi_1 - \bar{c}\xi_1^2)/(2\bar{d}-\bar{c}) = \psi(2\bar{d}) \leq \psi(2\bar{c}) = \xi_1(2-\xi_1)$$

by $\psi'(t) = \bar{c}(\xi_1^2 - \xi_1)/(t - \bar{c})^2 \leq 0$. □

Both values $\tilde{a}$, $\tilde{b}$ can be calculated efficiently (with a negligible increase in the number of arithmetic operations): to calculate $\tilde{b}$, by analogy with (2.21) we use formula

$$\tilde{b} = b - s^T\check{Y}_{[n]}(\check{S}_{[n]}^T\check{Y}_{[n]})^{-1}\check{S}_{[n]}^T y \tag{5.5}$$

(for the proof of (2.21) we need not the symmetry of S^TY) and $\tilde{a}$ can be calculated by (4.22). Since we need this value while we create blocks $S_{[n]}, Y_{[n]}$ and thus we have not blocks $S_{[i]}, Y_{[i]}$, $i < n$, created yet (see Algorithm 5.1), we will calculate only an estimate of this value, assuming that all matrices $S_{[i]}, Y_{[i]}$, $i < n$, have one column, i.e. that matrix $\check{H}$ given by (4.20) is calculated by the BNS method, see Section 1. In view of Lemma 5.1 we regard the benefit of the block BFGS update as sufficient, if $(1 - \tilde{b}/b)|1 - \tilde{a}/\tilde{b}| > 1$ together with $b/\tilde{b} > 1.5$ or if $b/\tilde{b} > 50$ (this criterion was found empirically).

To improve the readability of the main algorithm, we first present three auxiliary procedures. Procedure 5.1 serves for updating of basic matrices S^TY, Y^TY, similar to the algorithm given in [1] for updating of matrices D, U, Y^TY in (1.5). In comparison with the standard BNS method, where the upper triangular matrix U is used, we need the whole matrix S^TY here, therefore we use an additional vector $\check{Y}^Ts = -t\,\check{Y}^THg$, see also Algorithm 5.1. Note that the number of arithmetic operations is approximately the same as for the corresponding algorithm in [1]. We present the whole procedure for completeness, although some parts of steps (ii), (iii) are contained in Step 1 of Algorithm 5.1.

Procedure 5.1 *(Updating of basic matrices)*

Given: $t > 0$, matrices $\check{S}$, $\check{Y}$, $\check{S}^T\check{Y}$, $\check{Y}^T\check{Y}$ and vectors s, y, g_+, $\check{S}^T g$, $\check{Y}^T g$, $\check{Y}^T H g$.

(i): Set $S := [\,\check{S}, s]$, $Y := [\,\check{Y}, y]$.

(ii): Compute $S^T g_+ = [\check{S}^T g_+, s^T g_+]$, $Y^T g_+ = [\check{Y}^T g_+, y^T g_+]$, $\check{Y}^T s = -t\,\check{Y}^T H g$.

(iii): Compute $\check{S}^T y = \check{S}^T g_+ - \check{S}^T g$, $\check{Y}^T y = \check{Y}^T g_+ - \check{Y}^T g$, $s^T y$, $y^T y$.

(iv): Set $S^T Y := \begin{bmatrix} \check{S}^T\check{Y} & \check{S}^T y \\ s^T\check{Y} & s^T y \end{bmatrix}$, $Y^T Y := \begin{bmatrix} \check{Y}^T\check{Y} & \check{Y}^T y \\ y^T\check{Y} & y^T y \end{bmatrix}$ and return.

Procedure 5.2, based on Lemma 4.1, is used for seeking out of the positive definite bottom-right-corner submatrix of $[S^T Y]_{\underline{\nu}}^{\overline{\nu}}$ of a maximum order (with $i_D = 0$) and for its RL factorization (with $i_D = 1$), see Procedure 5.3.

Procedure 5.2 *(RL factorization of A)*

Given: A factorization indicator i_D, a global convergence parameter $\varepsilon_D \in (0,1)$, indices bounds $\underline{\nu}$, $\overline{\nu}$, $\underline{\nu} \le \overline{\nu}$, matrix $[S^T Y]_{\underline{\nu}}^{\overline{\nu}} \triangleq A$.

(i): If $i_D = 0$ set $A := A + A^T$. Set $\tilde{\nu} = \overline{\nu} - \underline{\nu} + 1$, $\hat{\nu} := \tilde{\nu}$.

(ii): If $i_D = 0$ and $A_{\hat{\nu},\hat{\nu}} \le \varepsilon_D \mathrm{Tr} A$ set $\underline{\nu} := \min(\underline{\nu} + \hat{\nu}, \overline{\nu})$ and go to *(iv). If* $\hat{\nu} = 1$ *go to (iv).*

(iii): Set $A_{\hat{\nu},j} := A_{\hat{\nu},j}/A_{\hat{\nu},\hat{\nu}}$, $j = 1,\dots,\hat{\nu}-1$. Set $A_{i,j} := A_{i,j} - A_{i,\hat{\nu}} A_{\hat{\nu},j}$, $i = 1,\dots,\hat{\nu}-1$, $j = 1,\dots,\hat{\nu}-1$. Set $\hat{\nu} := \hat{\nu} - 1$ and go to *(ii).*

(iv): If $i_D = 0$ return. Set $L_{i,j} := A_{i,j}$ for $1 \le j < i \le \tilde{\nu}$, $L_{i,j} := 1$ for $1 \le j = i \le \tilde{\nu}$, $R_{i,j} := A_{i,j}$ for $1 \le i \le j \le \tilde{\nu}$, $L_{i,j} := R_{i,j} := 0$ otherwise. Return.

The following Procedure 5.3 is used for formation and factorization of blocks $S_{[i]}^T Y_{[i]}$, $i = 1,\dots,n$ and selection of the suitable update formula. Note that to simplify updating with the choice (2.27), we merely create block $S_{[n]}^T Y_{[n]}$ of order 1, see step *(v).*

Procedure 5.3 *(Block generation)*

Given: Symmetry tolerances $\delta_1, \delta_2, \delta_3$, update-type tolerances $\delta_4, \delta_5, \delta_6$, $\delta_i > 0$, $i = 1,\dots,6$.

(i): Set $\delta := \delta_1$, an indices upper bound $\overline{\nu} := m$, an auxiliary block index $i_B := 1$ and an update-type ((2.26)−(2.28)) indicator $i_U := 0$.

(ii): Find a minimum indices bound $\underline{\nu}$ such that $\max_{j_1,j_2 \in \mathcal{I}_{\underline{\nu}}^{\overline{\nu}}} \left\{ (s_{j_1}^T y_{j_2} - s_{j_2}^T y_{j_1})^2 / (b_{j_2} b_{j_1}) \right\} \le \delta$.

(iii): Using Procedure 5.2 with $i_D = 0$, possibly correct the indices lower bound $\underline{\nu}$. If $m \le 3$ or $\overline{\nu} < m$ or $\overline{\nu} - \underline{\nu} \le 2$ or $\max_{j_1,j_2 \in \mathcal{I}_{\underline{\nu}}^{\overline{\nu}}} \left\{ (s_{j_1}^T y_{j_2} - s_{j_2}^T y_{j_1})^2 / (b_{j_2} b_{j_1}) \right\} > \delta_3$ go to *(v).*

(iv): Compute θ by (5.1), $\tilde{a} = \tilde{a}_C$ by (4.22), $\tilde{b}$ by (5.5) and $\hat{b} = b - s^T y_- \, s_-^T y / b_-$. If $(\,(1 - \tilde{b}/b)|1 - \tilde{a}/\tilde{b}| > 1$ and $b/\tilde{b} > 1.5)$ or $b/\tilde{b} > 50$ or $\theta < \delta_4$ then go to *(v).* If $\theta > \delta_5$ or $\hat{b}/b > \delta_6$ set $i_U := 1$, otherwise set $i_U := 2$.

(v): If $i_U = 1$ set $\underline{\nu} := \overline{\nu}$ and $i_U := 0$. Set $A_{i_B} := [S^T Y]_{\underline{\nu}}^{\overline{\nu}}$. If $i_U = 2$ and $\overline{\nu} = m$ denote by A_{i_B} matrix $\tilde{U}_{[n]}$ in (4.31). Using Procedure 5.2 with $i_D = 1$, find matrices $R_{i_B} = R$, $L_{i_B} = L$ such that $A_{i_B} := R_{i_B} L_{i_B}$. Set $\overline{\nu} := \underline{\nu} - 1$. If $\overline{\nu} \ge 1$ set $\delta := \delta_2$, $i_B := i_B + 1$ and go to *(ii).*

(vi): Set $n := i_B$, $S_{[i]}^T Y_{[i]} := A_{n-i+1}$, $R_{[i]} := R_{n-i+1}$, $L_{[i]} := L_{n-i+1}$, $i = 1, \dots, n$. If $i_U = 2$ set $\tilde{R}_{[n]} := R_{[n]}$, $\tilde{L}_{[n]} := L_{[n]}$. Return.

We now state the method in details. For simplicity, we omit stopping criteria and a contingent restart when some computed direction vector is not descent.

Algorithm 5.1

Data: A maximum number $\hat{m} > 1$ of columns S, Y, line search parameters ε_1, ε_2, $0 < \varepsilon_1 < 1/2$, $\varepsilon_1 < \varepsilon_2 < 1$, tolerance parameters $\delta_1, \dots, \delta_6$, $\delta_i > 0$, $i \in \{1, \dots, 6\}$, $\delta_4 < \delta_5$, 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 (1.1), $g_{k+1} = \nabla f(x_{k+1})$, $s_k = t_k d_k$, $y_k = g_{k+1} - g_k$, $b_k = s_k^T y_k$, $\zeta_k = b_k / y_k^T y_k$. If $k = 0$ set $S_k = [s_k]$, $Y_k = [y_k]$, $S_k^T Y_k = [s_k^T y_k]$, $Y_k^T Y_k = [y_k^T y_k]$, compute $S_k^T g_{k+1}$, $Y_k^T g_{k+1}$ and go to Step 4.

Step 2: Basic matrices updating. Using Procedure 5.1, form matrices S_k, Y_k, $S_k^T Y_k$, $Y_k^T Y_k$.

Step 3: Block generation and factorization. Using Procedure 5.3, find a number of blocks n and an update indicator i_U and form and factorize positive definite blocks $S_{[i]}^T Y_{[i]} = R_{[i]} L_{[i]}$, $i = n, \dots, 1$. Form matrices $U = U_n$ by (4.11), $L_D = \mathrm{diag}[\, L_{[1]}, \dots, L_{[n]} \,]$, $\mathcal{E}$ by (4.30) and $U_T := U L_D^{-1}$ for $i_U = 0$ or $\tilde{U}$ by (4.18), $\tilde{L}_D = \mathrm{diag}[\, L_{[1]}, \dots, L_{[n-1]}, \tilde{L}_{[n]} \,]$, $\tilde{\mathcal{E}}$ by (4.33) and $\tilde{U}_T := \tilde{U} \tilde{L}_D^{-1}$ for $i_U = 2$.

Step 4: Direction vector. Compute $d_{k+1} = -H_{k+1} g_{k+1}$ and an auxiliary vector $Y_k H_{k+1} g_{k+1}$ by (4.29) for $i_U = 0$ or by (4.32) for $i_U = 2$. Set $k := k + 1$. If $k \ge \hat{m}$ 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

In this section, we establish global convergence of Algorithm 5.1. The following assumption and lemma are presented in [17].

Assumption 6.1. *The objective function $f : \mathcal{R}^N \to \mathcal{R}$ is bounded from below and uniformly convex with bounded second-order derivatives (i.e. $0 < \underline{G} \le \underline{\lambda}(G(x)) \le \overline{\lambda}(G(x)) \le \overline{G} < \infty$, $x \in \mathcal{R}^N$, where $\underline{\lambda}(G(x))$ and $\overline{\lambda}(G(x))$ are the lowest and the greatest eigenvalues of the Hessian matrix $G(x)$).*

Lemma 6.1. *Let objective function f satisfy Assumption 6.1. Then $\underline{G} \le |y|^2/b \le \overline{G}$ and $b/|s|^2 \ge \underline{G}$.*

Lemma 6.2. *Let matrix $A_1 \in \mathcal{R}^{\mu \times \mu}$, $\mu > 0$, be positive semidefinite, matrix $A_2 \in \mathcal{R}^{\mu \times \mu}$ symmetric positive semidefinite. Then $0 \le \mathrm{Tr}(A_1 A_2) \le \mathrm{Tr}\, A_1 \, \mathrm{Tr}\, A_2$. Moreover, if A_2 is symmetric positive definite, then $\mathrm{Tr}(A_1 A_2^{-1}) \le \mathrm{Tr}\, A_1 \, (\mathrm{Tr}\, A_2)^{\mu-1} / \det A_2$.*

Proof. We can write $A_2 = Q \Lambda Q^T$ with Q orthogonal and Λ diagonal with $\Lambda_{ii} \ge 0$, $i = 1, \dots, \mu$, thus $\mathrm{Tr}(A_1 A_2) = \mathrm{Tr}(A_1 Q \Lambda Q^T) = \mathrm{Tr}(K \Lambda)$, where matrix $K = Q^T A_1 Q$ is obviously positive semidefinite, which immediately yields $K_{ii} \geq 0$, $i = 1, \ldots, \mu$. Since $\mathrm{Tr}(A_1 A_2) = \mathrm{Tr}(K\Lambda) = \sum_{i=1}^{\mu} K_{ii}\Lambda_{ii}$, we get $0 \leq \mathrm{Tr}(A_1 A_2) \leq \mathrm{Tr}\, K\ \mathrm{Tr}\, \Lambda = \mathrm{Tr}\, A_1\ \mathrm{Tr}\, A_2$.

If A_2 is symmetric positive definite, all eigenvalues Λ_{ii} of matrix A_2 satisfy $\Lambda_{ii} \geq \det A_2/(\mathrm{Tr}\, A_2)^{\mu-1}$, $i = 1, \ldots, \mu$, which yields

$$\mathrm{Tr}(A_1 A_2^{-1}) = \mathrm{Tr}(A_1 Q \Lambda^{-1} Q^T) = \mathrm{Tr}(K\Lambda^{-1}) = \sum_{i=1}^{\mu} K_{ii}\Lambda_{ii}^{-1} \leq \Big[(\mathrm{Tr}\, A_2)^{\mu-1}/\det A_2\Big]\, \mathrm{Tr}\, A_1$$

in view of $\sum_{i=1}^{\mu} K_{ii} = \mathrm{Tr}\, K = \mathrm{Tr}\, A_1$. □

Lemma 6.3. *Let matrices* $A_1, A_2 \in \mathcal{R}^{\mu\times\mu}$, $\mu > 0$, A_2 *nonsingular. Then* $(\mathrm{Tr}(A_1 A_2^{-1}))^2 \leq \mu\, \mathrm{Tr}(A_1^T A_1)\, (\mathrm{Tr}(A_2^T A_2))^{\mu-1}/(\det A_2)^2$.

Proof. For any $A \in \mathcal{R}^{\mu\times\mu}$ we have

$$(\mathrm{Tr}\, A)^2 = \Big(\sum_{i=1}^{\mu}(1.A_{ii})\Big)^2 \leq \mu \sum_{i=1}^{\mu} A_{ii}^2 \leq \mu \sum_{i=1}^{\mu}\sum_{j=1}^{\mu} A_{ij}^2 = \mu\, \mathrm{Tr}(A^T A)$$

by the Schwarz inequality and the assertion follows from Lemma 6.2 in view of

$$\Big(\mathrm{Tr}(A_1 A_2^{-1})\Big)^2 \leq \mu\, \mathrm{Tr}(A_2^{-T} A_1^T A_1 A_2^{-1}) = \mu\, \mathrm{Tr}\Big((A_1^T A_1)(A_2^T A_2)^{-1}\Big).$$ □

Lemma 6.4. *Let matrix* $A \in \mathcal{R}^{\mu\times\mu}$, $\mu > 0$, *be positive definite. Then* $\det \frac{1}{2}(A + A^T) \leq \det A$.

Proof. We will proceed by induction on μ. The result is true for $\mu = 1$. Let it be true for all positive definite matrices of some order $\mu \geq 1$, let $u, v \in \mathcal{R}^{\mu}$ and matrix $\bar{A} = \begin{bmatrix} A & u \\ v^T & \alpha \end{bmatrix}$ be positive definite. Then

$$\begin{vmatrix} A & u \\ v^T & \alpha \end{vmatrix} = \begin{vmatrix} A - uv^T/\alpha & u \\ 0^T & \alpha \end{vmatrix},$$

i.e. $\det \bar{A} = \alpha \det(A - uv^T/\alpha)$, where $\alpha > 0$ and matrix $A - uv^T/\alpha$ is positive definite by Lemma 4.1. This also implies

$$\det \frac{1}{2}(\bar{A} + \bar{A}^T) = \alpha \det\Big(\frac{1}{2}(A + A^T) - ww^T/\alpha\Big), \tag{6.1}$$

where $w = \frac{1}{2}(u + v)$ and matrix $\frac{1}{2}(A + A^T) - ww^T/\alpha$ is symmetric positive definite. Using the induction hypothesis and identity $\det(K + qq^T) = (1 + q^T K^{-1} q) \det K$ (K nonsingular matrix, q vector), which for K positive definite yields

$$\det(K + qq^T) \geq \det K, \tag{6.2}$$

we get

$$\begin{aligned} \det \bar{A} &= \alpha\ \det(A - uv^T/\alpha) \geq \alpha\ \det \tfrac{1}{2}\Big(A - uv^T/\alpha + A^T - vu^T/\alpha\Big) \\ &= \alpha\ \det\Big(\tfrac{1}{2}(A + A^T) - ww^T/\alpha + (u - v)(u - v)^T/(4\alpha)\Big) \\ &\geq \alpha\ \det\Big(\tfrac{1}{2}(A + A^T) - ww^T/\alpha\Big) = \det \tfrac{1}{2}(\bar{A} + \bar{A}^T) \end{aligned}$$

and the induction is established with $\mu + 1$ replacing μ. □

Lemma 6.5. *Let* $A \in \mathcal{R}^{\mu\times\mu}$, $w \in \mathcal{R}^{\mu}$, $\mu, \delta > 0$, *matrix* $\bar{A} = \begin{bmatrix} \alpha & w \\ w^T & A \end{bmatrix}$ *be symmetric positive definite and* $\det \bar{A} \geq \delta(\mathrm{Tr}\, \bar{A})^{\mu+1}$. *Then* $\det A > \delta(\mathrm{Tr}\, A)^{\mu}$.

Proof. Matrices $A - ww^T/\alpha$, A are symmetric positive definite and $\alpha > 0$ by Lemma 4.1, thus $\mathrm{Tr}\, \bar{A} > \alpha$, $\mathrm{Tr}\, \bar{A} > \mathrm{Tr}\, A$. Using (6.2) and (6.1), we obtain

$$\det A \geq \det(A - ww^T/\alpha) = (\det \bar{A})/\alpha > \delta(\mathrm{Tr}\, \bar{A})^{\mu+1}/\mathrm{Tr}\, \bar{A} = \delta(\mathrm{Tr}\, \bar{A})^{\mu} > \delta(\mathrm{Tr}\, A)^{\mu}.$$ □

Theorem 6.1. *Let objective function f satisfy Assumption 6.1. Then, Algorithm 5.1 generates a sequence $\{g_k\}$ that either satisfies $\lim_{k\to\infty}|g_k|=0$ or terminates with $g_k=0$ for some k.*

Proof. Procedure 5.2 with $i_D = 0$ de facto computes RL factorization of matrices $A_{[i]} = \frac{1}{2}(S_{[i]}^T Y_{[i]} + Y_{[i]}^T S_{[i]})$, where L has unit diagonal entries. For $m_i > 1$ (the number of columns of matrices $S_{[i]}, Y_{[i]}$) all diagonal entries of R are greater than $\varepsilon_D \mathrm{Tr}\, S_{[i]}^T Y_{[i]} = \varepsilon_D \mathrm{Tr} A_{[i]}$ in view of step *(ii)* of Procedure 5.2 and for $m_i = 1$ the only entry of R is $\mathrm{Tr} A_{[i]} > \varepsilon_D \mathrm{Tr} A_{[i]}$ by $\varepsilon_D < 1$, thus for $i=1,\dots,n$ and $k \ge 0$ by Lemma 6.4 we have

$$\det\left(\frac{1}{2}\left((S_{[i]}^T Y_{[i]})^{-1} + (Y_{[i]}^T S_{[i]})^{-1}\right)\right)^{-1} \ge \det S_{[i]}^T Y_{[i]} \ge \det A_{[i]} \ge \left(\varepsilon_D \mathrm{Tr}\, A_{[i]}\right)^{m_i}. \tag{6.3}$$

We assume that $C_{[i]} = I$, $i=1,\dots,n$, (see Sectin 5) and denote $\bar{H}_{[i]} = \frac{1}{2}(H_{[i]} + H_{[i]}^T)$, $B_{[i]} = H_{[i]}^{-1}$, $\bar{B}_{[i]} = \bar{H}_{[i]}^{-1}$, $\bar{B}_k = (\frac{1}{2}(H_k + H_k^T))^{-1}$, $\tilde{B}_k = \frac{1}{2}(B_k + B_k^T)$, $i=1,\dots,n+1$, $k \ge 0$. Since in all iterations we choose $H_{[1]} = \zeta_k I$, $\zeta_k = b_k/|y_k|^2$, i.e. $\bar{B}_{[1]} = (|y_k|^2/b_k) I$, Lemma 6.1 gives

$$\mathrm{Tr}\, \bar{B}_{[1]} = (|y_k|^2/b_k)\,\mathrm{Tr}\, I \le N\overline{G}, \quad \det \bar{B}_{[1]} = (|y_k|^2/b_k)^N \ge \underline{G}^N, \quad k \ge 0. \tag{6.4}$$

(i) Suppose first that $i_U = 0$ (i.e. in the kth iteration for all blocks $S_{[i]}^T Y_{[i]}$ we use the block BFGS update, i.e. set $H_+ = H_{k+1} = H_{[n+1]}$, where matrices $H_{[i+1]}$ are given by (4.1), $i=1,\dots,n$). By Corollary 2.1, Theorem 2.3 (b)-(c) and (6.3), updates (4.1) yield

$$\begin{aligned} \bar{B}_{[i+1]} &= \bar{B}_{[i]} - \bar{B}_{[i]} S_{[i]} (S_{[i]}^T \bar{B}_{[i]} S_{[i]})^{-1} S_{[i]}^T \bar{B}_{[i]} + Y_{[i]} A_{[i]}^{-1} Y_{[i]}^T\,, & (6.5)\\ B_{[i+1]} &= \bar{B}_{[i]} - \bar{B}_{[i]} S_{[i]} (S_{[i]}^T \bar{B}_{[i]} S_{[i]})^{-1} S_{[i]}^T \bar{B}_{[i]} + Y_{[i]} (Y_{[i]}^T S_{[i]})^{-1} Y_{[i]}^T\,, & (6.6)\\ \det \bar{B}_{[i+1]} &\ge \det \frac{1}{2}(B_{[i+1]} + B_{[i+1]}^T) = \det \bar{B}_{[i]} \,\det A_{[i]} \Big/ \det(S_{[i]}^T \bar{B}_{[i]} S_{[i]})\,, & (6.7) \end{aligned}$$

$i=1,\dots,n$, where matrices $S_{[i]}^T \bar{B}_{[i]} S_{[i]}$ are symmetric positive definite by Theorem 2.3 (d), since Algorithm 5.1 generates all blocks $S_{[i]}^T Y_{[i]}$ positive definite by Lemma 4.1 and thus all columns of matrices $S_{[i]}, Y_{[i]}$, $i=1,\dots,n$, are linearly independent.

Relation (6.5), Lemma 6.2, relation (6.3) and Lemma 6.1 give

$$\begin{aligned} \mathrm{Tr}\, \bar{B}_{[i+1]} - \mathrm{Tr}\, \bar{B}_{[i]} &\le \mathrm{Tr}(Y_{[i]}^T Y_{[i]} A_{[i]}^{-1}) \le \mathrm{Tr}\, Y_{[i]}^T Y_{[i]} (\mathrm{Tr} A_{[i]})^{m_i - 1} / (\varepsilon_D \mathrm{Tr} A_{[i]})^{m_i} \\ &= \varepsilon_D^{-m_i}\, \mathrm{Tr}\, Y_{[i]}^T Y_{[i]} / \mathrm{Tr}\, A_{[i]} \le \textstyle\sum_{j\in\mathcal{I}_i} (|y_j|^2/b_j)/\varepsilon_D^{m_i} \le m_i\, \overline{G}/\varepsilon_D^{m_i}, \end{aligned} \tag{6.8}$$

$i=1,\dots,n$. Using (6.4), in view of $\varepsilon_D < 1$ and $\sum_{i=1}^n m_i = m$ this yields

$$\mathrm{Tr}\, \bar{B}_{[i]} \le (N + m/\varepsilon_D^m)\, \overline{G} \triangleq \Theta_0, \; i=1,\dots,n+1, \quad \mathrm{Tr}\, \bar{B}_{k+1} = \mathrm{Tr}\, \bar{B}_{[n+1]} \le \Theta_0, \quad k>0. \tag{6.9}$$

Since $\mathrm{Tr}\, B_{[n+1]} - \mathrm{Tr}\, \bar{B}_{[n]} \le \mathrm{Tr}\Big(Y_{[n]}^T Y_{[n]} (Y_{[n]}^T S_{[n]})^{-1}\Big)$ by (6.6), Lemmas 6.1–6.3 and (6.3) give

$$\begin{aligned} \mathrm{Tr}\, B_{[n+1]} - \mathrm{Tr}\, \bar{B}_{[n]} &\le \sqrt{m_n \mathrm{Tr}\,(Y_{[n]}^T Y_{[n]})^2 \Big[\mathrm{Tr}(S_{[n]}^T S_{[n]})\, \mathrm{Tr}(Y_{[n]}^T Y_{[n]})\Big]^{m_n - 1}} \Big/ (\varepsilon_D \mathrm{Tr} A_{[n]})^{m_n} \\ &\le \frac{\sqrt{m_n}}{\varepsilon_D^{m_n}} \sum_{j\in\mathcal{I}_n} \frac{|y_j|^2}{b_j} \Big[\sum_{j\in\mathcal{I}_n} \frac{|s_j|^2}{b_j} \sum_{j\in\mathcal{I}_n} \frac{|y_j|^2}{b_j}\Big]^{\frac{m_n - 1}{2}} \le \frac{m_n^{3/2}}{\varepsilon_D^{m_n}}\, \overline{G} \Big(m_n \frac{\overline{G}}{\underline{G}}\Big)^{m_n}, \end{aligned}$$

which by (6.4) and (6.9) yields

$$\mathrm{Tr}\, B_{k+1} = \mathrm{Tr}\, B_{[n+1]} \le \Theta_0 + (m^2/\varepsilon_D^m)(m\overline{G}/\underline{G})^m\, \overline{G} \triangleq \Theta_1 > \Theta_0, \quad k > 0. \tag{6.10}$$

Since $(\det A)^{1/\mu} \le (1/\mu) \mathrm{Tr}\, A$ for $A \in \mathcal{R}^{\mu\times\mu}$ symmetric positive definite, $\mu > 0$, we have $(\det(S_{[i]}^T \bar{B}_{[i]} S_{[i]}))^{1/m_i} \le \mathrm{Tr}\,(S_{[i]}^T \bar{B}_{[i]} S_{[i]})/m_i$ and relations (6.7) and (6.3), Lemma 6.2, relation (6.9) and Lemma 6.1 give

$$\Big(\frac{\det \bar{B}_{[i+1]}}{\det \bar{B}_{[i]}}\Big)^{\frac{1}{m_i}} \ge \Big(\frac{\det \frac{1}{2}(B_{[i+1]}+B_{[i+1]}^T)}{\det \bar{B}_{[i]}}\Big)^{\frac{1}{m_i}} \ge \frac{m_i(\det A_{[i]})^{\frac{1}{m_i}}}{\mathrm{Tr}(S_{[i]}^T S_{[i]}\,\bar{B}_{[i]})} \ge \frac{m_i(\varepsilon_D \mathrm{Tr}\, A_{[i]})}{\mathrm{Tr} S_{[i]}^T S_{[i]}\,.\mathrm{Tr}\bar{B}_{[i]}}$$
$$\ge \frac{m_i\, \varepsilon_D \mathrm{Tr}\, A_{[i]}}{\Theta_0 \mathrm{Tr}\, S_{[i]}^T S_{[i]}} = \frac{m_i\, \varepsilon_D}{\Theta_0}\, \frac{\sum_{j\in\mathcal{I}_i} b_j}{\sum_{j\in\mathcal{I}_i} |s_j|^2} \ge \frac{m_i\, \varepsilon_D/\Theta_0}{\sum_{j\in\mathcal{I}_i}|s_j|^2/b_j} \ge \frac{\varepsilon_D\, \underline{G}}{\Theta_0}\,, \tag{6.11}$$

$i=1,\dots,n$. Using (6.4), this yields

$$\det \bar{B}_{[n]} \ge \underline{G}^N (\varepsilon_D\, \underline{G}\,/\Theta_0)^{m-m_n}, \tag{6.12}$$
$$\det \tilde{B}_{k+1} = \det \frac{1}{2}(B_{[n+1]}+B_{[n+1]}^T) \ge \underline{G}^N(\varepsilon_D\, \underline{G}\,/\Theta_0)^m \triangleq \Theta_2\,, \quad k>0. \tag{6.13}$$

(ii) Let $i_U=2$ in the kth iteration, i.e. for blocks $S_{[i]}^T Y_{[i]}$, $i=1,\dots,n-1$, we use the block BFGS update (thus also $\mathrm{Tr}\, \bar{B}_{[n]} \le \Theta_0$ (see (6.9)) and (6.12) hold) and for block $S_{[n]}^T Y_{[n]}$ update (4.14)–(4.15) with $\check{C}_{[n]} = I$ and $\hat{s}=\ddot{s}$, $\hat{y}=\ddot{y}$ given by (2.28). Denoting $\check{B}=\check{H}^{-1}$ (positive definite by Theorem 2.3 (d)), $\tilde{B}_A = \frac{1}{2}(\check{B}+\check{B}^T)$, $\bar{H}_A=\frac{1}{2}(\check{H}+\check{H}^T)$, $\bar{B}_A = \bar{H}_A^{-1}$, $\hat{P}=I-(1/\hat{s}^T\hat{y})\hat{y}\hat{s}^T$ and $\check{A}_{[n]}=\frac{1}{2}(\check{S}_{[n]}^T\check{Y}_{[n]}+\check{Y}_{[n]}^T\check{S}_{[n]})$, from (4.15) we obtain

$$\check{B} \;=\; \bar{B}_{[n]} - \bar{B}_{[n]}\check{S}_{[n]}(\check{S}_{[n]}^T\bar{B}_{[n]}\check{S}_{[n]})^{-1}\check{S}_{[n]}^T\bar{B}_{[n]} + \check{Y}_{[n]}(\check{S}_{[n]}^T\check{Y}_{[n]})^{-T}\check{Y}_{[n]}^T\,, \tag{6.14}$$
$$\tilde{B}_A \;=\; \bar{B}_{[n]} - \bar{B}_{[n]}\check{S}_{[n]}(\check{S}_{[n]}^T\bar{B}_{[n]}\check{S}_{[n]})^{-1}\check{S}_{[n]}^T\bar{B}_{[n]} + \check{Y}_{[n]}\check{A}_{[n]}^{-1}\check{Y}_{[n]}^T\,, \tag{6.15}$$
$$\det \tilde{B}_A \;=\; \det \bar{B}_{[n]}\,.\,\det \check{A}_{[n]}/\det(\check{S}_{[n]}^T\bar{B}_{[n]}\check{S}_{[n]}) \tag{6.16}$$

by Theorem 2.3 and Corollary 2.1. In the same way as (6.9) and (6.13) we get

$$\mathrm{Tr}\, \bar{B}_A \le \Theta_0 < \Theta_1, \quad \mathrm{Tr}\, \tilde{B}_A = \mathrm{Tr}\, \check{B} \le \Theta_1, \quad \det \tilde{B}_A \ge \Theta_2. \tag{6.17}$$

Denoting $u=\check{B}\hat{s}/\sqrt{\hat{s}^T\check{B}\hat{s}} = \check{B}\hat{s}/\sqrt{\hat{s}^T\tilde{B}_A\hat{s}}$, $v=\check{B}^T\hat{s}/\sqrt{\hat{s}^T\check{B}\hat{s}} = \check{B}^T\hat{s}/\sqrt{\hat{s}^T\tilde{B}_A\hat{s}}$, we obtain

$$B_{k+1} \;=\; \check{B} - (1/\hat{s}^T\check{B}\hat{s})\check{B}\hat{s}\hat{s}^T\check{B} + (1/\hat{s}^T\hat{y})\hat{y}\hat{y}^T = \check{B} - uv^T + (1/\hat{s}^T\hat{y})\hat{y}\hat{y}^T, \tag{6.18}$$
$$\tilde{B}_{k+1} \;=\; \tilde{B}_A - (1/\hat{s}^T\tilde{B}_A\hat{s})\tilde{B}_A\hat{s}\hat{s}^T\tilde{B}_A + (1/\hat{s}^T\hat{y})\hat{y}\hat{y}^T + (1/4)(u-v)(u-v)^T, \tag{6.19}$$
$$\bar{B}_{k+1} \;=\; ((1/\hat{s}^T\hat{y})\hat{s}\hat{s}^T + \hat{P}^T\bar{H}_A\hat{P})^{-1} = \bar{B}_A - (1/\hat{s}^T\bar{B}_A\hat{s})\bar{B}_A\hat{s}\hat{s}^T\bar{B}_A + (1/\hat{s}^T\hat{y})\hat{y}\hat{y}^T, \tag{6.20}$$

by (4.14), Theorem 2.3 and relations $2(uv^T+vu^T)=(u+v)(u+v)^T-(u-v)(u-v)^T$ and $\frac{1}{2}(u+v)=(1/\hat{s}^T\tilde{B}_A\hat{s})\tilde{B}_A\hat{s}$. Setting $\bar{u}=\bar{B}_A^{-1/2}u$, $\bar{v}=\bar{B}_A^{-1/2}v$, we get

$$-2\,u^Tv \le |u|^2+|v|^2 = \bar{u}^T\bar{B}_A\bar{u}+\bar{v}^T\bar{B}_A\bar{v} \le 2\mathrm{Tr}\,\bar{B}_A \le 2\Theta_0 \tag{6.21}$$

by $\bar{u}^T\bar{u} = u^T\bar{H}_A u = u^T\check{H}u = 1 = \bar{v}^T\bar{v}$ and (6.17). Using (6.2) with $q=\frac{1}{2}(u-v)$, $K=\tilde{B}_A-(1/\hat{s}^T\tilde{B}_A\hat{s})\tilde{B}_A\hat{s}\hat{s}^T\tilde{B}_A+(1/\hat{s}^T\hat{y})\hat{y}\hat{y}^T$, (6.19) and Theorem 2.3, we obtain

$$\det \tilde{B}_{k+1} \ge \det K = (\det \tilde{B}_A)\,\hat{s}^T\hat{y}/\hat{s}^T\tilde{B}_A\hat{s}. \tag{6.22}$$

From $\hat{y}=\bar{P}y$, $\hat{s}=\bar{P}^Ts$, where $\bar{P}=I-(1/b_-)\,y_-s_-^T$, we have

$$|\hat{y}| \le \|\bar{P}\|\,|y| = |y|(|s_-|\,|y_-|/b_-) \le |y|\sqrt{\overline{G}/\underline{G}}\,, \quad |\hat{s}| \le \|\bar{P}^T\|\,|s| \le |s|\sqrt{\overline{G}/\underline{G}} \tag{6.23}$$

by Lemma 6.1. Further, by Theorem 2.5 we have $\hat{s}^T\hat{y} = s^T\hat{y} = b - s^Ty_-\,s_-^Ty/b_-$. Applying Lemma 6.5 repeatedly (m_n-2) times to inequality $\det A_{[n]} \ge \varepsilon_D^{m_n}(\mathrm{Tr}\, A_{[n]})^{m_n}$ (see (6.3)), we have $\det\frac{1}{2}([s_-,s]^T[y_-,y]+[y_-,y]^T[s_-,s]) > \varepsilon_D^{m_n}(b_-+b)^2$. Using Lemma 6.4, we get

$$\hat{s}^T\hat{y} = \frac{1}{b_-}\begin{vmatrix} b_- & s_-^Ty \\ s^Ty_- & b\end{vmatrix} \ge \frac{1}{b_-}\begin{vmatrix} b_- & (s_-^Ty+s^Ty_-)/2 \\ (s_-^Ty+s^Ty_-)/2 & b\end{vmatrix} > \varepsilon_D^{m_n}\frac{(b_-+b)^2}{b_-} > \varepsilon_D^m b. \tag{6.24}$$

Since matrix $\tilde{B}_A$ is symmetric positive definite, from (6.17)–(6.24) we obtain

$$\mathrm{Tr}\, B_{k+1} = \mathrm{Tr}\, \check{B} - u^T v + \frac{|\hat{y}|^2}{\hat{s}^T\hat{y}} < 2\Theta_1 + \frac{|y|^2\, \overline{G}}{\varepsilon_D^m\, b\, \underline{G}} \le 2\Theta_1 + \frac{\overline{G}^{\,2}}{\varepsilon_D^m \underline{G}} \triangleq \Theta_3, \quad \mathrm{Tr}\, \bar{B}_{k+1} < \Theta_3, \tag{6.25}$$

$$\det \tilde{B}_{k+1} \ge (\det \tilde{B}_A) \frac{\hat{s}^T\hat{s}}{\hat{s}^T\tilde{B}_A\hat{s}} \frac{\hat{s}^T\hat{y}}{\hat{s}^T\hat{s}} > \frac{\Theta_2}{\Theta_1} \frac{\varepsilon_D^m\, b\, \underline{G}}{|s|^2\overline{G}} \ge \Theta_2 \frac{\varepsilon_D^m \underline{G}^2}{\Theta_1 \overline{G}} \triangleq \Theta_4, \tag{6.26}$$

$k > 0$, with $\Theta_3 > \Theta_1$ and $\Theta_4 < \Theta_2$, by Lemma 6.1, (6.13), $\varepsilon_D < 1$, $\underline{G} \le \overline{G}$ and (6.9) – (6.10).

(iii) The lowest eigenvalue $\underline{\lambda}(\tilde{B}_k)$ of $\tilde{B}_k$ satisfies $\underline{\lambda}(\tilde{B}_k) \ge \det \tilde{B}_k / (\mathrm{Tr}\, B_k)^{N-1}$ by $\mathrm{Tr}\, \tilde{B}_k = \mathrm{Tr}\, B_k$, $k \ge 0$. Setting $q_k = \bar{H}_k^{1/2} g_k$, from (6.9) – (6.10), (6.13) and (6.25) – (6.26) we get

$$\frac{(s_k^T g_k)^2}{|s_k|^2 |g_k|^2} = \frac{s_k^T B_k s_k}{s_k^T s_k} \frac{g_k^T H_k g_k}{g_k^T g_k} = \frac{s_k^T \tilde{B}_k s_k}{s_k^T s_k} \frac{q_k^T q_k}{q_k^T \bar{B}_k q_k} \ge \frac{\det \tilde{B}_k}{(\mathrm{Tr}\, B_k)^{N-1}} \frac{1}{\mathrm{Tr}\, \bar{B}_k} > \frac{\Theta_4}{\Theta_3^N}, \quad k > 1,$$

which implies $\lim_{k\to\infty} |g_k| = 0$, see Theorem 3.2 in [15] and relations (3.17) – (3.18) ibid. □

7 Numerical experiments

In this section, we compare our results with the results obtained by the L-BFGS method, see [8], [14], by the BNS method [1] and by our best limited-memory methods based on vector corrections, see [18], [17], using the following collections of test problems:

- Test 11 from [11] (55 chosen problems, computed repeatedly ten times for a better comparison), which are problems from CUTE collection [2], some of them modified; used N are given in Table 1, where the modified problems are marked with '*',
- Test 25 from [10] (68 chosen problems), $N = 10000$.

Problem	N	Problem	N	Problem	N	Problem	N
ARWHEAD	5000	DIXMAANI	3000	EXTROSNB	1000	NONDIA	5000
BDQRTIC	5000	DIXMAANJ	3000	FLETCBV3*	1000	NONDQUAR	5000
BROYDN7D	2000	DIXMAANK	3000	FLETCBV2	1000	PENALTY3	1000
BRYBND	5000	DIXMAANL	3000	FLETCHCR	1000	POWELLSG	5000
CHAINWOO	1000	DIXMAANM	3000	FMINSRF2	5625	SCHMVETT	5000
COSINE	5000	DIXMAANN	3000	FREUROTH	5000	SINQUAD	5000
CRAGGLVY	5000	DIXMAANO	3000	GENHUMPS	1000	SPARSINE	1000
CURLY10	1000	DIXMAANP	3000	GENROSE	1000	SPARSQUR	1000
CURLY20	1000	DQRTIC	5000	INDEF*	1000	SPMSRTLS	4999
CURLY30	1000	EDENSCH	5000	LIARWHD	5000	SROSENBR	5000
DIXMAANE	3000	EG2	1000	MOREBV*	5000	TOINTGSS	5000
DIXMAANF	3000	ENGVAL1	5000	NCB20*	1010	TQUARTIC*	5000
DIXMAANG	3000	CHNROSNB*	1000	NCB20B*	1000	WOODS	4000
DIXMAANH	3000	ERRINROS*	1000	NONCVXU2	1000		

Table 1: Dimensions for Test 11 – modified CUTE collection.

The source texts and the reports corresponding to these test collections can be downloaded from the web page `www.cs.cas.cz/luksan/test.html`.

All methods are implemented in the optimization software system UFO, described in [13] and introduced in `www.cs.cas.cz/luksan/ufo.html`. We have used $m = 5$, $\delta_1 = 10^{-2}$,

$\delta_2 = 10^{-1}$, $\delta_3 = 10^{-13}$, $\delta_4 = 10^{-10}$, $\delta_5 = 10^{-3}$, $\delta_6 = 0.5$, $\varepsilon_D = 10^{-6}$, $\varepsilon_1 = 10^{-4}$, $\varepsilon_2 = 0.8$ and the final precision $\|g(x^\star)\|_\infty \le 10^{-6}$.

Table 2 contains the total number of function and also gradient evaluations (NFV) and the total computational time in seconds (Time).

Method	Test 11		Test 25	
	NFV	Time	NFV	Time
L-BFGS	80539	13.941	501651	574.59
BNS	78704	14.344	517186	661.66
Alg. 4.1 in [17]	64395	13.038	319565	420.00
Alg. 4.2 in [18], $n=4$	63987	13.063	309650	415.27
Alg. 5.1	65228	12.211	371830	468.19

Table 2: Comparison of the selected methods.

For a better demonstration of both the efficiency and the reliability, we compare selected optimization methods by using performance profiles introduced in [4]. The performance profile $\rho_M(\tau)$ is defined by the formula

$$\rho_M(\tau) = \frac{\text{number of problems where } \log_2(\tau_{P,M}) \le \tau}{\text{total number of problems}}$$

with $\tau \ge 0$, where $\tau_{P,M}$ is the performance ratio of the number of function evaluations (or the time) required to solve problem P by method M to the lowest number of function evaluations (or the time) required to solve problem P. The ratio $\tau_{P,M}$ is set to infinity (or some large number) if method M fails to solve problem P.

The value of $\rho_M(\tau)$ at $\tau = 0$ gives the percentage of test problems for which the method M is the best and the value for τ large enough is the percentage of test problems that method M can solve. The relative efficiency and reliability of each method can be directly seen from the performance profiles: the higher is the particular curve, the better is the corresponding method. Figures 7.1-4, based on results in Table 2, reveal the performance profiles for tested methods graphically.

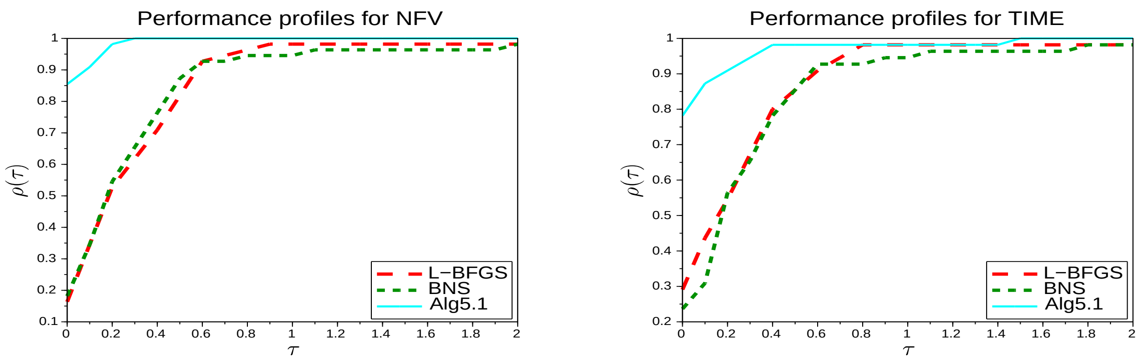

Figure 7.1: Comparison of $\rho_M(\tau)$ for Test 11 and various methods.

Figures 7.1-2 demonstrate the efficiency of our method in comparison with the BNS and the L-BFGS methods and from Figures 7.3-4 we can see that the numerical results for the new method and the results for our methods [18], [17] are comparable.

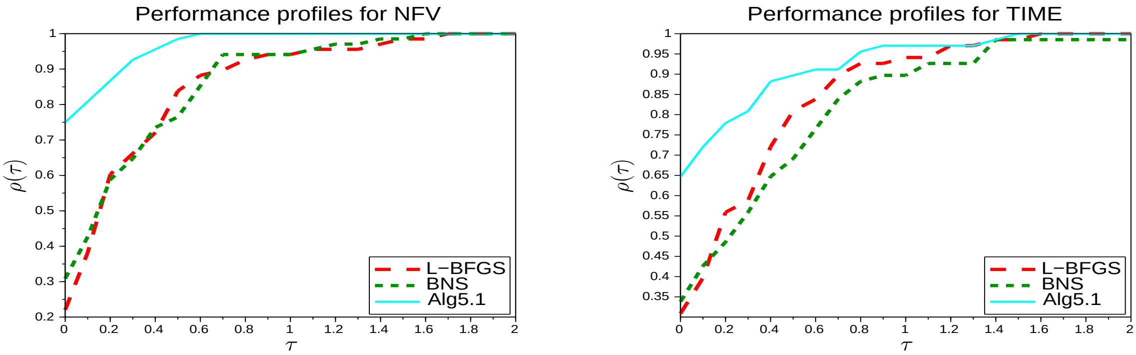

Figure 7.2: Comparison of $\rho_M(\tau)$ for Test 25 and various methods.

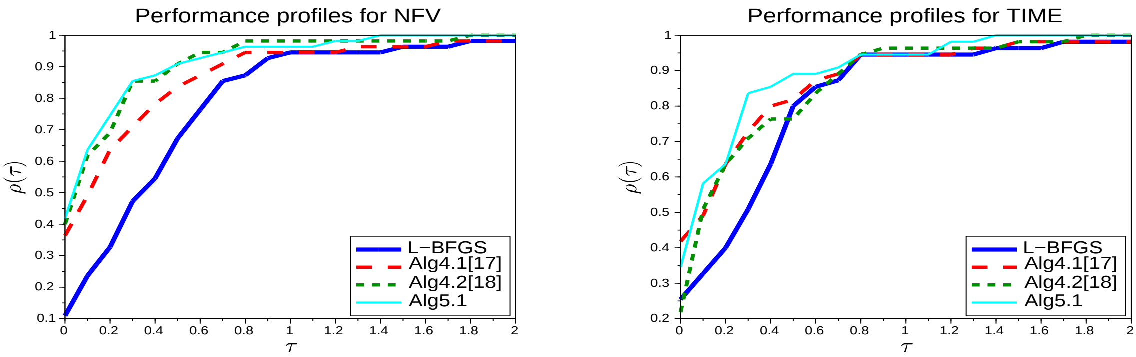

Figure 7.3: Comparison of $\rho_M(\tau)$ for Test 11 and various methods.

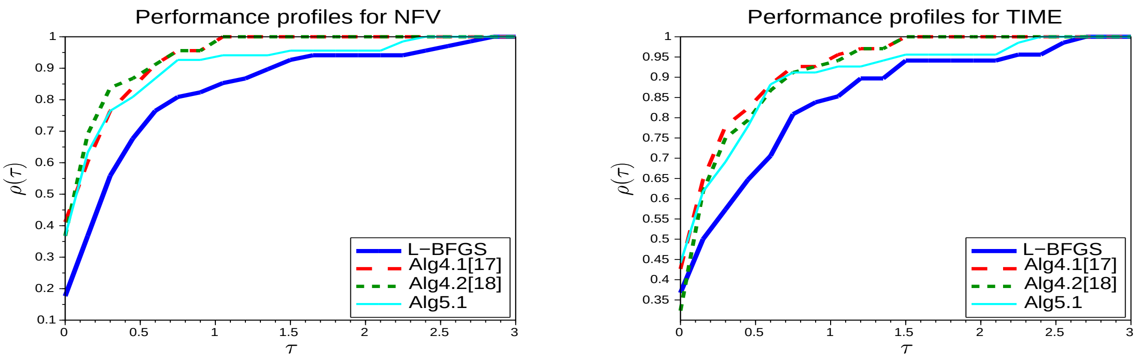

Figure 7.4: Comparison of $\rho_M(\tau)$ for Test 25 and various methods.

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 ([18], [17]).

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.

References

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

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

[3] J.E. Dennis Jr., R.B. Schnabel: Least change secant updates for quasi-Newton methods. SIAM Review 21 (1979), 443-459.

[4] E.D. Dolan, J.J. Moré: *Benchmarking optimization software with performance profiles*, Mathematical Programming 91 (2002) 201-213.

[5] M. Fiedler: *Special matrices and their applications in numerical mathematics*, Martinus Nijhoff Publishers, Dordrecht, The Netherlands, 1986.

[6] R. Fletcher: *Practical Methods of Optimization*, John Wiley & Sons, Chichester, 1987.

[7] Y.F. Hu, C. Storey: *Motivating Quasi-Newton Updates by Preconditioned Conjugate Gradient Methods*, Math. Report A 150, Department of Mathematical Sciences, Loughborough University of Technology, England, 1991.

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

[9] L. Lukšan, C. Matonoha, J. Vlček: *Algorithm 896: LSA - Algorithms for Large-Scale Optimization*, ACM Transactions on Mathematical Software 36 (2009) 16:1-16:29.

[10] L. Lukšan, C. Matonoha, J. Vlček: *Sparse Test Problems for Unconstrained Optimization*, Report V-1064, Prague, ICS AS CR, 2010.

[11] L. Lukšan, C. Matonoha, J. Vlček: *Modified CUTE Problems for Sparse Unconstrained Optimization*, Report V-1081, Prague, ICS AS CR, 2010.

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

[13] L. Lukšan, M. Tůma, C. Matonoha, J. Vlček, N. Ramešová, M. Šiška, J. Hartman: *UFO 2014. Interactive System for Universal Functional Optimization*, Report V-1218, Prague, ICS AS CR, 2014.

[14] J. Nocedal: *Updating quasi-Newton matrices with limited storage*, Math. Comp. 35 (1980) 773-782.

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

[16] R.B. Schnabel: *Quasi-Newton Methods Using Multiple Secant Equations*, Technical Report CU-CS-247-83, Department of Computer Science, University of Colorado at Boulder, USA, 1983.

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

[18] J. Vlček, L. Lukšan: *A modified limited-memory BNS method for unconstrained minimization based on conjugate directions idea*, Optimization Methods & Software 30 (2015) 616-633.