národní
úložiště
šedé
literatury

Two limited-memory optimization methods with minimum violation of the previous quasi-Newton equations

Vlček, Jan

2020

Dostupný z http://www.nusl.cz/ntk/nusl-432144

Dílo je chráněno podle autorského zákona č. 121/2000 Sb.

Tento dokument byl stažen z Národního úložiště šedé literatury (NUŠL).

Datum stažení: 01.05.2024

Další dokumenty můžete najít prostřednictvím vyhledávacího rozhraní nusl.cz .

Institute of Computer Science
Academy of Sciences of the Czech Republic

Two limited-memory optimization methods with minimum violation of the previous quasi-Newton equations

Jan Vlček[a], Ladislav Lukšan[a, b]

[a]Institute of Computer Science of the Czech Academy of Sciences,
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 1280

September 2020

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

Two limited-memory optimization methods with minimum violation of the previous quasi-Newton equations[1]

Jan Vlček[a], Ladislav Lukšan[a,b]

[a]Institute of Computer Science of the Czech Academy of Sciences,
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 1280

September 2020

Abstract:

Limited-memory variable metric methods based on the well-known BFGS update are widely used for large scale optimization. The block version of the BFGS update, derived by Schnabel (1983), Hu and Storey (1991) and Vlček and Lukšan (2019), satisfies the quasi-Newton equations with all used difference vectors and for quadratic objective functions gives the best improvement of convergence in some sense, but the corresponding direction vectors are not descent directions generally. To guarantee the descent property of direction vectors and simultaneously violate the quasi-Newton equations as little as possible in some sense, two methods based on the block BFGS update are proposed. They can be advantageously combined with methods based on vector corrections for conjugacy (Vlček and Lukšan, 2015). Global convergence of the proposed algorithm is established for convex and sufficiently smooth functions. Numerical experiments demonstrate the efficiency of the new methods.

Keywords:
Unconstrained minimization, variable metric methods, limited-memory methods, variationally derived methods, global convergence, numerical results

[1]This work was supported by the Institute of Computer Science of the CAS (RVO: 67985807).

1 Introduction

In this report we assume that the problem function $f : \mathcal{R}^N \to \mathcal{R}$ is differentiable and propose two new limited-memory variable metric (VM) or quasi-Newton (QN) methods for large scale unconstrained optimization

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

which are based on a block version of the widely used BFGS update, see below.

The VM methods [7, 14] are iterative. They start with an initial point $x_0 \in \mathcal{R}^N$ and generate iterations $x_{k+1} \in \mathcal{R}^N$ by the process $x_{k+1} = x_k + s_k$, $s_k = t_k d_k$, $k \ge 0$, where d_k is the direction vector, $t_k > 0$ the stepsize, chosen regularly in such a way that

$$f(x_+) - f(x) \le \varepsilon_1 \, t \, g^T d, \qquad g_+^T d \ge \varepsilon_2 \, g^T d \tag{1.1}$$

(the Wolfe line search conditions [17]), $0 < \varepsilon_1 < 1/2$, $\varepsilon_1 < \varepsilon_2 < 1$ and $g_k = \nabla f(x_k)$. Usually $d_k = -H_k g_k$ with a symmetric positive VM matrix $H_k \triangleq B_k^{-1}$; typically H_0 is a multiple of I and H_{k+1} is obtained from H_k to satisfy the QN (secant) equation (condition)

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

with the difference vectors s_k, $y_k = g_{k+1} - g_k$. Among VM methods, the BFGS method [7, 14, 17] belongs to the most efficient ones and can be easily modified for large scale optimization; the L-BFGS [10, 16] and BNS [3] methods represent its well-known limited-memory adaptations. We refer to Section 2 for a brief description of these methods.

To incorporate more past information to the update formula, the block version of the BFGS update was derived in [18] for symmetric positive definite VM matrices, using a variational approach, in [8] for quadratic functions, using corrections for the exact line search, and recently in [21] for general functions, using a block variant of the approach in [4]. This update satisfies the QN equations with all used difference vectors and brings the best improvement of convergence in some sense [20, 21] for quadratic objective functions. Nevertheless, it does not guarantee that the corresponding direction vectors are descent directions for general functions.

Using the block BFGS update in a generalized form [21], in Section 3 we derive two new updates variationally with direction vectors that satisfy the descent property. Naturally, some QN equations with the used difference vectors can be violated. To utilize the advantageous properties of the block BFGS updates, we look for a solution to the following constrained optimization problem: to find such an update which satisfies the QN equations with several latest difference vectors and violates the QN equations from previous iterations as little as possible in some sense. In Section 3.1 we show that this problem can be converted equivalently to a problem without equality constraints and with a smaller number of variables. We also derive an expression for the solution of this simpler problem. Our first new update is based on this expression.

Besides, in such optimization problems we can use further constraints, derived from properties of the block BNS method [21]. Then the resultant second new update has a simple implementation and some interesting properties, see Section 3.2.

Our experiments indicate that the efficiency of the methods based on these two updates is similar. Moreover, it can be significantly improved by combination with methods based on vector corrections for conjugacy which use previous difference vectors and store the corrected difference vectors [19, 20, 22]. A main reason for this improvement probably consists in the fact that these vectors commonly contain cumulative information from previous iterations.

In Section 4 we combine these corrections with the corrections that use the subsequent difference vectors (see Section 4.1) to satisfy the QN equations with both the corrected and original, uncorrected latest difference vectors, see Section 4.2. For our new updates, these special corrections appear to give slightly better numerical results than the others.

The application to the limited-memory VM methods and the corresponding algorithm are described in Section 5. Global convergence of the algorithm is established in Section 6 and numerical results are reported in Section 7.

To simplify notation we frequently omit the index k and replace the index $k+1$ by the symbol $+$ and the index $k-1$ by the symbol $-$.

We will denote by $\|\cdot\|_F$ the Frobenius matrix norm, by $\|\cdot\|$ the spectral matrix norm and by $|\cdot|$ the size of both scalars and vectors (the Euclidean vector norm).

2 The L-BFGS and BNS methods

In this section we briefly describe the limited-memory VM methods L-BFGS [10, 16], implemented as subroutine PLIS in [11], and BNS [3]. These methods are based on the BFGS update formula, mentioned in Section 1, which preserves the positive definiteness of H and can be written in the following quasi-product form

$$H_+ = (1/b)ss^T + \Big(I - (1/b)sy^T\Big)H\Big(I - (1/b)ys^T\Big), \quad b = s^Ty \tag{2.1}$$

($b > 0$ for $g \neq 0$ by (1.1)). To modify the BFGS method for large scale optimization, we choose $H_k^I \in \mathcal{R}^{N\times N}$ in every iteration (usually $H_k^I = \zeta_k I$, $\zeta_k > 0$) and recurrently update H_k^I (without forming an approximation of the inverse Hessian matrix explicitly) by the BFGS formula, using m pairs of vectors $(s_{k-\tilde{m}}, y_{k-\tilde{m}}), \ldots, (s_k, y_k)$ successively, where

$$\tilde{m} = \min[k, \hat{m}-1], \quad m = \tilde{m}+1, \quad k \geq 0 \tag{2.2}$$

and $\hat{m} > 1$ is a given parameter. We use the BNS update to guarantee global convergence of Algorithm 5.1 if some conditions for our new methods are not satisfied, see Section 5. Instead of the famous compact form [3], we use it in the form (also given in [3])

$$H_+ = SR^{-T}DR^{-1}S^T + \Big(I - SR^{-T}Y^T\Big)H^I\Big(I - YR^{-1}S^T\Big), \tag{2.3}$$

where $S_k = [s_{k-\tilde{m}}, \ldots, s_k]$, $Y_k = [y_{k-\tilde{m}}, \ldots, y_k]$, $D_k = \mathrm{diag}[b_{k-\tilde{m}}, \ldots, b_k]$, $(R_k)_{i,j} = (S_k^T Y_k)_{i,j}$ for $i \leq j$, $(R_k)_{i,j} = 0$ otherwise (an upper triangular matrix), $k \geq 0$. We can see that for $H^I = \zeta I$ the direction vector $-H_+g_+$ (and subsequently an auxiliary vector $Y^TH_+g_+$, see Section 5) can be calculated efficiently (without computing of H_+ explicitly) by

$$-H_+g_+ = -\zeta g_+ - S\Big[R^{-T}\Big((D+\zeta Y^TY)R^{-1}S^Tg_+ - \zeta Y^Tg_+\Big)\Big] + Y\Big[\zeta R^{-1}S^Tg_+\Big], \tag{2.4}$$

$$Y^TH_+g_+ = \zeta Y^Tg_+ + Y^TS\Big[R^{-T}\Big((D+\zeta Y^TY)R^{-1}S^Tg_+ - \zeta Y^Tg_+\Big)\Big] - Y^TY\Big[\zeta R^{-1}S^Tg_+\Big], \tag{2.5}$$

where in the square brackets we multiply by low-order matrices.

3 Two variationally derived VM updates

The basic variant of the block BFGS update derived in [21] for general functions is

$$H_+ = S\,A^{-1}S^T + \left(I - SA^{-T}Y^T\right) H^I \left(I - YA^{-1}S^T\right), \quad A = S^TY, \tag{3.1}$$

where $H^I \in \mathcal{R}^{N\times N}$ is symmetric and $A \in \mathcal{R}^{m\times m}$ arbitrary nonsingular. The matrix H_+ is the nearest matrix to H^I in some sense that satisfies the QN equation $H_+Y = S$, however the direction vector $d_+ = -H_+g_+$ is not a descent direction generally.

Replacing S by $S\Theta$, $\Theta \in \mathcal{R}^{m\times m}$ nonsingular, in the same way as (3.1) we can derive the generalized form

$$H_+ = S\,XS^T + \left(I - SA^{-T}Y^T\right) H^I \left(I - YA^{-1}S^T\right), \quad X = \Theta A^{-1}, \tag{3.2}$$

which satisfies the violated (for $\Theta \neq I$) QN equations $H_+Y = S(XA) = S\Theta$. Note that the infinitely many times repeated BNS update, investigated in [22], has the same form (with a special symmetric positive definite matrix X).

To express the violation of the QN equations $H_+Y = S$ for H_+ nonsingular, we will use the matrix

$$\Delta = (H_+Y - S)^T B_+ (H_+Y - S) = (B_+S - Y)^T H_+ (B_+S - Y), \quad B_+ = H_+^{-1}. \tag{3.3}$$

Obviously, for H_+ symmetric positive definite we can write $\operatorname{Tr}\Delta = ||B_+^{1/2}(H_+Y - S)||_F^2$. The following lemma gives another expression of Δ, which uses only low-order matrices.

Lemma 3.1. *Let H_+ be given by (3.2) with X symmetric and XA nonsingular. Then*

$$\Delta = A^TXA + X^{-1} - A - A^T. \tag{3.4}$$

Proof. Since (3.2) implies $H_+Y = S(XA)$ and $Y(XA)^{-1} = B_+S$, the desired conclusion follows from $\Delta = Y^TH_+Y + S^TB_+S - A - A^T$. □

Besides (1.2), more latest QN equations can be satisfied, if some lower-right-corner principal submatrix of A is symmetric (see Lemmas 3.2 and 3.4), which can be achieved e.g. by means of vector corrections for conjugacy (see Section 4).

According to our intention, see Section 1, first we will look for such a symmetric positive definite matrix X, which minimizes $\operatorname{Tr}\Delta$ in some sense, see below, and for which update (3.2) satisfies the QN equations with several latest difference vectors.

3.1 The first new update

To derive our first update of the form (3.2), we will solve the problem

$$\min_{X\in\mathcal{S}_m} \operatorname{Tr}\Delta \quad s.t. \quad H_+Y_{(2)} = S_{(2)}, \tag{3.5}$$

where we split S, Y in such a way that $S = [S_{(1)}, S_{(2)}]$, $Y = [Y_{(1)}, Y_{(2)}]$, $S_{(2)}, Y_{(2)} \in \mathcal{R}^{N\times\mu}$, $0<\mu<m$, and by $\mathcal{S}_i$ we denote the set of symmetric positive definite matrices of order i.

The following lemma and theorem show that the QN equations in (3.5) can be replaced by some conditions for structure of X and that it will be sufficient to look for some submatrix of X, which minimizes the trace of some submatrix of Δ.

Lemma 3.2. *Let $S = [S_{(1)}, S_{(2)}]$, $Y = [Y_{(1)}, Y_{(2)}]$, $S_{(2)}, Y_{(2)} \in \mathcal{R}^{N\times\mu}$, $0 < \mu < m$, let*

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} = \begin{bmatrix} S_{(1)}^T Y_{(1)} & S_{(1)}^T Y_{(2)} \\ S_{(2)}^T Y_{(1)} & S_{(2)}^T Y_{(2)} \end{bmatrix} \tag{3.6}$$

be nonsingular and a symmetric matrix $X \in \mathcal{R}^{m\times m}$ be partitioned in the same way. Let A_{22} be symmetric nonsingular and H_+ be given by (3.2). Then $H_+Y_{(2)} = S_{(2)}$ if and only if

$$X_{12} = X_{21}^T = -X_{11}A_{12}A_{22}^{-1}, \quad X_{22} = A_{22}^{-1} + A_{22}^{-1}A_{12}^T X_{11}A_{12}A_{22}^{-1}. \tag{3.7}$$

Proof. In view of (3.6) we have

$$XA = \begin{bmatrix} X_{11}A_{11} + X_{12}A_{21} & X_{11}A_{12} + X_{12}A_{22} \\ X_{12}^T A_{11} + X_{22}A_{21} & X_{12}^T A_{12} + X_{22}A_{22} \end{bmatrix} \triangleq \begin{bmatrix} (XA)_{11} & (XA)_{12} \\ (XA)_{21} & (XA)_{22} \end{bmatrix}, \tag{3.8}$$

thus (3.7) is equivalent to $(XA)_{12} = 0$, $(XA)_{22} = I$ and further to $H_+Y_{(2)} = S_{(2)}$ by $H_+Y = S(XA)$ and linear independence of columns of S, which follows from $\det A \neq 0$. □

Theorem 3.1. *Let the assumptions of Lemma 3.2 be satisfied, $C = A_{11} - A_{12}A_{22}^{-1}A_{21}$, X_{11} be nonsingular and (3.7) hold. Then X and C are nonsingular, $H_+Y_{(2)} = S_{(2)}$ and*

$$XA = \begin{bmatrix} X_{11}C & 0 \\ A_{22}^{-1}(A_{21} - A_{12}^T X_{11}C) & I \end{bmatrix}, \quad X^{-1} = \begin{bmatrix} X_{11}^{-1} + A_{12}A_{22}^{-1}A_{12}^T & A_{12} \\ A_{12}^T & A_{22} \end{bmatrix}, \tag{3.9}$$

$$\Delta = \begin{bmatrix} C^T X_{11}C + X_{11}^{-1} - C - C^T + (A_{12} - A_{21}^T)A_{22}^{-1}(A_{12}^T - A_{21}) & 0 \\ 0 & 0 \end{bmatrix}. \tag{3.10}$$

Moreover, if $A_{22} \in \mathcal{S}_\mu$ then we get $X \in \mathcal{S}_m$ if and only if $X_{11} \in \mathcal{S}_{m-\mu}$.

Proof. Theorem 1.4.2 in [6] yields $\det A = \det C . \det A_{22}$, therefore $\det C \neq 0$. As in the proof of Lemma 3.2 we get (3.8), $H_+Y_{(2)} = S_{(2)}$, $(XA)_{12} = 0$ and $(XA)_{22} = I$. Since

$$X_{22} - X_{21}X_{11}^{-1}X_{12} = A_{22}^{-1} + A_{22}^{-1}A_{12}^T X_{11}A_{12}A_{22}^{-1} - A_{22}^{-1}A_{12}^T X_{11}A_{12}A_{22}^{-1} = A_{22}^{-1}$$

by (3.7), we similarly obtain $\det X = \det A_{22}^{-1} . \det X_{11}$, thus X is nonsingular and all assumptions of Lemma 3.1 are satisfied. From (3.8) and (3.7) we derive

$$\begin{aligned} (XA)_{11} &= X_{11}(A_{11} - A_{12}A_{22}^{-1}A_{21}) = X_{11}C, \\ (XA)_{21} &= -A_{22}^{-1}A_{12}^T X_{11}A_{11} + (A_{22}^{-1} + A_{22}^{-1}A_{12}^T X_{11}A_{12}A_{22}^{-1})A_{21} \\ &= A_{22}^{-1}(A_{21} - A_{12}^T X_{11}(A_{11} - A_{12}A_{22}^{-1}A_{21}))\,. \end{aligned}$$

This gives the first relation in (3.9) and further

$$A^T XA - A^T = A^T(XA - I) = \begin{bmatrix} A_{11}^T & A_{21}^T \\ A_{12}^T & A_{22} \end{bmatrix} \begin{bmatrix} X_{11}C - I & 0 \\ A_{22}^{-1}(A_{21} - A_{12}^T X_{11}C) & 0 \end{bmatrix},$$

which implies

$$(A^T XA - A^T)_{12} = 0, \quad (A^T XA - A^T)_{22} = 0, \quad (A^T XA - A^T)_{21} = A_{21} - A_{12}^T, \tag{3.11}$$

$$\begin{aligned} (A^T XA - A^T)_{11} &= A_{11}^T X_{11}C - A_{11}^T + A_{21}^T A_{22}^{-1}A_{21} - A_{21}^T A_{22}^{-1}A_{12}^T X_{11}C \\ &= C^T X_{11}C - A_{11}^T + A_{21}^T A_{22}^{-1}A_{21}. \end{aligned} \tag{3.12}$$

Similarly, we can express

$$X^{-1} - A = A((XA)^{-1} - I) = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} C^{-1}X_{11}^{-1} - I & 0 \\ -A_{22}^{-1}(A_{21} - A_{12}^T X_{11} C)\, C^{-1}X_{11}^{-1} & 0 \end{bmatrix}$$

by the first relation in (3.9), which yields

$$\begin{aligned} (X^{-1} - A)_{12} &= 0, \quad (X^{-1} - A)_{22} = 0, \quad (X^{-1} - A)_{21} = -A_{21} + A_{12}^T, & (3.13) \\ (X^{-1} - A)_{11} &= A_{11}C^{-1}X_{11}^{-1} - A_{11} - A_{12}A_{22}^{-1}A_{21}C^{-1}X_{11}^{-1} + A_{12}A_{22}^{-1}A_{12}^T \\ &= C(C^{-1}X_{11}^{-1}) - A_{11} + A_{12}A_{22}^{-1}A_{12}^T, & (3.14) \end{aligned}$$

i.e. the second relation in (3.9). Considering that

$$\begin{aligned} (A_{12} - A_{21}^T)A_{22}^{-1}(A_{12}^T - A_{21}) &= A_{12}A_{22}^{-1}A_{12}^T - A_{12}A_{22}^{-1}A_{21} - A_{21}^T A_{22}^{-1}A_{12}^T + A_{21}^T A_{22}^{-1}A_{21} \\ &= A_{12}A_{22}^{-1}A_{12}^T + A_{21}^T A_{22}^{-1}A_{21} - A_{11} - A_{11}^T + C + C^T, \end{aligned}$$

we obtain (3.10) from Lemma 3.1, (3.12) and (3.14). Moreover, for $A_{22} \in \mathcal{S}_\mu$ we have $X_{11} \in \mathcal{S}_{m-\mu}$ if and only if $X \in \mathcal{S}_m$ by the second relation in (3.9), since the Schur complement of A_{22} in X^{-1} is X_{11}^{-1}. □

For $A_{22} \in \mathcal{S}_\mu$ and A nonsingular, by Theorem 3.1 we can equivalently solve the problem

$$\min_{X_{11} \in \mathcal{S}_{m-\mu}} \operatorname{Tr}(C^T X_{11} C + X_{11}^{-1} - C - C^T) \tag{3.15}$$

instead of (3.5). A solution to this problem is given by the following theorem.

Theorem 3.2. *Let C be nonsingular. Then $X_{11} = (CC^T)^{-1/2}$ is the unique solution to (3.15) and for this X_{11}, the matrices X_{11}^{-1}, $C^T X_{11} C$ are similar.*

Proof. For $X_{11} = (CC^T)^{-1/2}$, the matrices X_{11}^{-1}, $C^T X_{11} C = C^{-1}(CC^T)X_{11}C = C^{-1}X_{11}^{-1}C$ are similar. Let $J = (CC^T)^{1/2}$. Obviously, $J \in \mathcal{S}_{m-\mu}$. Since the trace of a product of two square matrices is independent of the order of multiplication, for any $X_{11} \in \mathcal{S}_{m-\mu}$ we have

$$\begin{aligned} \operatorname{Tr}(C^T X_{11} C + X_{11}^{-1}) &= \operatorname{Tr}(X_{11}CC^T + X_{11}^{-1}) = \operatorname{Tr}\left(X_{11}^{1/2}(J^2 + X_{11}^{-2})X_{11}^{1/2}\right) \\ &= \operatorname{Tr}\left(X_{11}^{1/2}(J^2 - JX_{11}^{-1} - X_{11}^{-1}J + X_{11}^{-2})X_{11}^{1/2} + 2J\right) \\ &= \operatorname{Tr}\left(X_{11}^{1/2}\left(J - X_{11}^{-1}\right)^2 X_{11}^{1/2} + 2J\right), \end{aligned}$$

which completes the proof by $X_{11}^{1/2} \in \mathcal{S}_{m-\mu}$ and symmetry of $J - X_{11}^{-1}$. □

Note that for $X_{11} = (CC^T)^{-1/2}$, the matrix $X_{11}C$ is orthogonal, as we can see from

$$(X_{11}C)^{-1} = C^{-1}X_{11}^{-2}X_{11} = C^{-1}(CC^T)X_{11} = (X_{11}C)^T.$$

3.2 The second new update

Our second new update of the form (3.2) utilizes some of properties of the block BNS update [21] to minimize not only $\operatorname{Tr}\Delta$, but also $\|\Delta\|_F$. We consider the update

$$H_+^R = S_R X_R S_R^T + (I - S_R A_R^{-T} Y_R^T) H^I (I - Y_R A_R^{-1} S_R^T), \tag{3.16}$$

where $S_R, Y_R \in \mathcal{R}^{N\times m-1}$ are the matrices S, Y without the first column, $A_R = S_R^T Y_R$ is nonsingular and X_R is symmetric. The block BNS update of H^I, where one of the diagonal blocks is $S_R^T Y_R$ (of order $m-1$), can have the form

$$H_+^{BBNS} = S_R X_R S_R^T + (I - S_R A_R^{-T} Y_R^T) H^A (I - Y_R A_R^{-1} S_R^T)$$

with some suitable $H^A \in \mathcal{R}^{N\times N}$ (see Lemma 6 in [21]). The matrix H_+^{BBNS} has the property $H_+^{BBNS} Y_R = S_R(X_R A_R) = H_+^R Y_R$. By analogy with this, we will look for $X \in \mathcal{S}_m$ which minimizes both $\operatorname{Tr}\Delta$ and $\|\Delta\|_F$ subject to $H_+ Y_R = H_+^R Y_R$. For such X we will show that $XA \in \mathcal{T}_m$, where $\mathcal{T}_i$ is the set of lower triangular matrices of order i.

Theorem 3.3. *Let $H^I \in \mathcal{R}^{N\times N}$ be a symmetric positive definite matrix, $m > 1$ and $S_R, Y_R \in \mathcal{R}^{N\times m-1}$ the matrices S, Y without the first column. Let $A_R = S_R^T Y_R$ be nonsingular, let some $X_R \in \mathcal{S}_{m-1}$ satisfying $X_R A_R \in \mathcal{T}_{m-1}$ be firmly given, H_+^R be given by (3.16) and H_+ by (3.2) with A nonsingular and X symmetric. Then for*

$$A = \begin{bmatrix} \alpha & u^T \\ v & A_R \end{bmatrix}, \quad X = \begin{bmatrix} \xi & \chi^T \\ \chi & X_P \end{bmatrix}, \quad p = A_R^{-T} u, \quad c = \alpha - p^T v, \tag{3.17}$$

$\alpha, \xi \in \mathcal{R}$, we have $c \neq 0$. Further for both $i = 1$ and $i = 2$, the unique solution to

$$\min_{X \in \mathcal{S}_m} \varphi_i(\Delta) \quad s.t. \quad H_+ Y_R = H_+^R Y_R, \quad \varphi_1(\Delta) = \operatorname{Tr}\Delta, \quad \varphi_2(\Delta) = \|\Delta\|_F, \tag{3.18}$$

satisfies $XA \in \mathcal{T}_m$ and is given by the choice $\xi = 1/|c|$, $\chi = -\xi p$ and $X_P = X_R + \xi p p^T$.

Proof. Since $\det A = (\alpha - u^T A_R^{-1} v) \det A_R = c \det A_R$, we get $c \neq 0$. Further,

$$XA = \begin{bmatrix} \alpha\xi + \chi^T v & \xi u^T + \chi^T A_R \\ \alpha\chi + X_P v & \chi u^T + X_P A_R \end{bmatrix} \tag{3.19}$$

by (3.17). The terms $H_+ Y_R, H_+^R Y_R$ in (3.18) can be equivalently rewritten as

$$H_+ Y_R = H_+ Y \begin{bmatrix} 0^T \\ I \end{bmatrix} = S(XA) \begin{bmatrix} 0^T \\ I \end{bmatrix}, \quad H_+^R Y_R = S_R(X_R A_R) = S \begin{bmatrix} 0^T \\ X_R A_R \end{bmatrix}.$$

By this, (3.19) and linear independence of columns of S (which follows from $\det A \neq 0$), the equality $H_+ Y_R = H_+^R Y_R$ is equivalent to

$$\xi u^T + \chi^T A_R = 0^T, \quad \chi u^T + X_P A_R = X_R A_R\,. \tag{3.20}$$

For $X \in \mathcal{S}_m$, we obviously have $H_+ \in \mathcal{S}_N$ and $\xi > 0$ and (3.20) is equivalent to

$$\chi = -\xi p, \quad X_P = X_R + \xi p p^T.$$

In view of this, we can rewrite the first column of XA in (3.19) in the following way:

$$\alpha\xi + \chi^T v = \xi c, \quad \alpha\chi + X_P v = -\alpha\xi p + X_R v + \xi(p^T v)p = X_R v - \xi c p \triangleq z.$$

By this and (3.19) - (3.20) we obtain

$$XA = \begin{bmatrix} \xi c & 0^T \\ z & X_R A_R \end{bmatrix}, \quad (XA)^{-1} = \begin{bmatrix} 1/(\xi c) & 0^T \\ -A_R^{-1} X_R^{-1} z/(\xi c) & A_R^{-1} X_R^{-1} \end{bmatrix}. \tag{3.21}$$

In view of Lemma 3.1, (3.21) and $X^{-1} = A(XA)^{-1}$ we can write Δ in the form

$$\Delta = \begin{bmatrix} \alpha & v^T \\ u & A_R^T \end{bmatrix} \begin{bmatrix} \xi c & 0^T \\ z & X_R A_R \end{bmatrix} + \begin{bmatrix} \alpha & u^T \\ v & A_R \end{bmatrix} \begin{bmatrix} 1/(\xi c) & 0^T \\ -A_R^{-1} X_R^{-1} z/(\xi c) & A_R^{-1} X_R^{-1} \end{bmatrix} - A - A^T$$
$$= \begin{bmatrix} \varrho - 2\alpha & v^T X_R A_R + u^T A_R^{-1} X_R^{-1} - u^T - v^T \\ A_R^T X_R v + X_R^{-1} A_R^{-T} u - u - v & A_R^T X_R A_R + X_R^{-1} - A_R - A_R^T \end{bmatrix},$$

where

$$\begin{aligned}
\varrho &= \left[\alpha, v^T\right] \cdot \begin{bmatrix} \xi c \\ z \end{bmatrix} + \left[\alpha, u^T\right] \cdot \begin{bmatrix} 1/(\xi c) \\ -A_R^{-1} X_R^{-1} z/(\xi c) \end{bmatrix} \\
&= \alpha(\xi c) + v^T z + \alpha/(\xi c) - u^T A_R^{-1} X_R^{-1} z/(\xi c) \\
&= \alpha\big(\xi c + 1/(\xi c)\big) + \big(v^T - p^T X_R^{-1}/(\xi c)\big)(X_R v - \xi c p) \\
&= c\big(\xi c + 1/(\xi c)\big) + v^T X_R v + p^T X_R^{-1} p
\end{aligned}$$

by $p = A_R^{-T} u$, $z = X_R v - \xi c p$ and $c = \alpha - p^T v$. Obviously, we get minimum ϱ for $\xi = 1/|c|$. Since all entries of Δ except for $\varrho - 2\alpha$ are independent of ξ and $\varrho - 2\alpha \ge 0$ by $H_+ \in \mathcal{S}_N$ and (3.3), solutions to both problems (3.18) are the same with $X \in \mathcal{S}_m$ by (3.17), $\xi > 0$ and $X_P - \chi\chi^T/\xi = X_R \in \mathcal{S}_{m-1}$. Finally, $XA \in \mathcal{T}_m$ by (3.21) and $X_R A_R \in \mathcal{T}_{m-1}$. □

To define our second new update (3.2) for $m \ge 2$, we start with $m = 2$ and $X_R = A_R^{-1}$ symmetric positive definite of unit order, which corresponds to the BFGS update (2.1). Then we apply Theorem 3.3 iteratively, substituting A, X from the previous iteration for A_R, X_R. We use this update only if all $c > 0$ (otherwise the resultant VM method appears to be inefficient). Then the final matrix XA is obviously lower triangular with unit diagonal entries, thus the QN equation $H_+ y = s$ is satisfied by $H_+ Y = S(XA)$; additional QN equations can be satisfied similarly as for the first new update, see Lemma 3.4.

The following lemma shows that such a matrix X, determined uniquely by the repeated application of Theorem 3.3, can also be constructed in a different way.

Lemma 3.3. *Let $X \in \mathcal{S}_m$ and $A \in \mathcal{R}^{m\times m}$ be nonsingular with the factorization $A = UL$, where U is upper triangular, L lower triangular and U, L have the same main diagonals. Then XA is lower triangular with unit diagonal entries if and only if $X = U^{-T}U^{-1}$.*

Proof. If $X = U^{-T}U^{-1}$, the matrix $XA = U^{-T}U^{-1}UL = U^{-T}L$ is obviously lower triangular and has unit diagonal entries in view of the assumption that U, L have the same main diagonals.

Let $X \in \mathcal{S}_m$ and $XA \triangleq K$ is lower triangular with unit diagonal entries. Writing $X = R_X^T R_X$ with R_X upper triangular, we have $K = R_X^T R_X UL$. This yields

$$R_X U = R_X^{-T} K L^{-1} \triangleq E, \tag{3.22}$$

i.e. the upper triangular matrix $R_X U$ is equal to the lower triangular matrix $R_X^{-T} K L^{-1}$. Thus E is a diagonal matrix, $R_X = EU^{-1}$ and (3.22) implies $E = (EU^{-1})^{-T} K L^{-1} = E^{-T} U^T K L^{-1}$, or $E^2 = U^T K L^{-1}$. Similarly as above we deduce that this diagonal matrix has unit diagonal entries, i.e. $E^2 = I$ and $X = R_X^T R_X = (U^{-T}E)(EU^{-1}) = U^{-T}U^{-1}$. □

The following lemma shows that QN equations $H_+ Y_{(2)} = S_{(2)}$ can be satisfied under similar conditions as for the first new update.

Lemma 3.4. *Let the assumptions of Lemmas 3.2 and 3.3 be satisfied and* $X = U^{-T}U^{-1}$*. Then* $H_+Y_{(2)} = S_{(2)}$*.*

Proof. In view of Lemma 3.2 it suffices to prove (3.7). Suppose that

$$U = \begin{bmatrix} U_{11} & U_{12} \\ 0 & U_{22} \end{bmatrix}, \quad L = \begin{bmatrix} L_{11} & 0 \\ L_{21} & L_{22} \end{bmatrix} \tag{3.23}$$

with $U_{22}, L_{22} \in \mathcal{R}^{\mu\times\mu}$. Since A_{22} is symmetric nonsingular, from $A = UL$ and (3.23) we get first $A_{22} = U_{22}L_{22} = L_{22}^TU_{22}^T$. Thus the lower triangular matrix $L_{22}U_{22}^{-T}$ is equal to the upper triangular matrix $U_{22}^{-1}L_{22}^T$, where U_{22}, L_{22} have the same main diagonals. This obviously yields $L_{22}U_{22}^{-T} = I$, i.e. $L_{22} = U_{22}^T$.

Further, $A = UL$ implies $A_{12} = U_{12}L_{22}$, or $U_{12} = A_{12}L_{22}^{-1}$, thus (3.23) gives

$$U^{-1} = \begin{bmatrix} U_{11}^{-1} & -U_{11}^{-1}A_{12}A_{22}^{-1} \\ 0 & U_{22}^{-1} \end{bmatrix}, \quad U^{-T} = \begin{bmatrix} U_{11}^{-T} & 0 \\ -A_{22}^{-1}A_{12}^TU_{11}^{-T} & U_{22}^{-T} \end{bmatrix} \tag{3.24}$$

by $L_{22}^{-1}U_{22}^{-1} = A_{22}^{-1}$. From $X = U^{-T}U^{-1}$ we obtain $X_{11} = U_{11}^{-T}U_{11}^{-1}$ and subsequently

$$X = \begin{bmatrix} X_{11} & -X_{11}A_{12}A_{22}^{-1} \\ -A_{22}^{-1}A_{12}^TX_{11} & U_{22}^{-T}U_{22}^{-1} + A_{22}^{-1}A_{12}^TX_{11}A_{12}A_{22}^{-1} \end{bmatrix}, \tag{3.25}$$

thus (3.7) is satisfied by $U_{22}^{-T}U_{22}^{-1} = L_{22}^{-1}U_{22}^{-1} = A_{22}^{-1}$. □

The second new update has the following property, relevant namely in connection with the corrections for conjugacy which use subsequent difference vectors, see Section 4.1.

Lemma 3.5. *Let the assumptions of Lemma 3.3 be satisfied and* $X = U^{-T}U^{-1}$*. Then the matrix* H_+ *given by (3.2) is invariant under the transformation* $S \to S\,T_S$, $Y \to YT_Y$ *for arbitrary lower triangular matrices* $T_S, T_Y \in \mathcal{R}^{m\times m}$ *with the same main diagonals.*

Proof. Let $\tilde S = S\,T_S$, $\tilde Y = Y\,T_Y$, $\tilde A = \tilde S^T\tilde Y$. Then $\tilde A = T_S^TAT_Y = \tilde U\tilde L$, where $\tilde U = T_S^TU$ and $\tilde L = LT_Y$ satisfy the same assumptions as U, L in Lemma 3.3. Therefore corresponding $\tilde X$, determined uniquely by $\tilde X = \tilde U^{-T}\tilde U^{-1} = T_S^{-1}XT_S^{-T}$, see comments before Lemma 3.3, satisfies $\tilde S\tilde X\tilde S^T = SXS^T$. Further, $\tilde Y\tilde A^{-1}\tilde S^T = YA^{-1}S^T$, which completes the proof. □

4 Combination with vector corrections for conjugacy

It was shown in [19, 20, 22] that the performance of limited-memory VM methods can often be improved by corrections for conjugacy which use previous difference vectors to correct the vectors s, y. These corrections can be written in the form $s \to \hat s = s + S_C\sigma$, $y \to \hat y = y + Y_C\eta$, where S_C, Y_C are matrices with some selected columns of S, Y without s, y. The vectors σ, η are chosen so that $\hat y^TS_C = \hat s^TY_C = 0$, with potential modifications if $\hat b = \hat s^T\hat y$ is negative or too small. As we mentioned in Section 1, corresponding $\hat s, \hat y$ can be stored and used in next iterations instead of original s, y. However, this usually requires additional arithmetic operations.

Note that for quadratic functions, $[S_C, s] = S$, $[Y_C, y] = Y$ and the BFGS udate, these corrections represent the best improvement of convergence in some sense under some conditions [20] and that for the theory in this section it is not significant, whether columns of S_C, Y_C were corrected in previous iterations.

Although for one correction vector the improvement of numerical results can often be substantial [19], for our two new methods the benefit of additional correction vectors appears to be questionable. Therefore in this report we will consider only one correction vector. For this type of corrections and $S_C = [s_-]$, $Y_C = [y_-]$, we can write

$$\hat{s} = s - \frac{s^T y_-}{b_-} s_- = \Big(I - \frac{1}{b_-} s_- y_-^T \Big) s, \quad \hat{y} = y - \frac{s_-^T y}{b_-} y_- = \Big(I - \frac{1}{b_-} y_- s_-^T \Big) y, \tag{4.1}$$

which satisfy (conjugacy of $\hat{s}, s_-$ with respect to B and B_+ for $H_+ \hat{y} = \hat{s}$, $H y_- = s_-$)

$$\hat{s}^T y_- = s_-^T \hat{y} = 0. \tag{4.2}$$

These corrections appear to improve efficiency only if $\hat{s}^T \hat{y}/b$ is positive and not too small and if $s^T y_-, s_-^T y$ are not too different.

If we replace s, y by $\hat{s}, \hat{y}$ and construct some new update with the corrections (4.1), we get $H_+[y_-, \hat{y}] = [s_-, \hat{s}]$ by $\hat{s}^T y_- = \hat{y}^T s_- = 0$ and Theorem 3.1 for the first update or by Lemma 3.4 for the second update, both with $S_{(2)} = [s_-, \hat{s}]$ and $Y_{(2)} = [y_-, \hat{y}]$. Thus

$$H_+ y - s = H_+ \hat{y} + (s_-^T y / b_-) s_- - s = \Big((s_-^T y - s^T y_-)/b_- \Big) s_-. \tag{4.3}$$

To satisfy the QN equations with both the corrected and uncorrected (original) latest difference vectors, in Section 4.2 we combine these corrections with the following type of corrections, which use subsequent difference vectors to correct previous columns of S, Y.

4.1 Corrections for conjugacy which use subsequent difference vectors

Corrections $S \to S\,T_S$, $Y \to Y T_Y$, where $T_S, T_Y \in \mathcal{R}^{m \times m}$ are lower triangular matrices with the same main diagonals, represent another type of corrections for conjugacy. They are implicit for our second new update, which is invariant under these transformations, see Lemma 3.5. If the factorization $A = UL$ exists with U, L satisfying the assumptions of Lemma 3.3, we can e.g. simply get $(ST_S)^T (YT_Y) = I$ setting $T_S = U^{-T}$, $T_Y = L^{-1}$.

To correct s_-, y_- in this way, we can use the following formulas

$$\tilde{s}_- = \Big(I - (1/b) s y^T \Big) s_-, \quad \tilde{y}_- = \Big(I - (1/b) y s^T \Big) y_-, \tag{4.4}$$

analogous to (4.1), for $\tilde{s}_-^T \tilde{y}_-$ positive and not too small (note that $\tilde{s}_-^T \tilde{y}_- = \hat{s}^T \hat{y}\, (b_-/b)$). These vectors satisfy

$$\tilde{s}_-^T y = s^T \tilde{y}_- = 0\,. \tag{4.5}$$

The second new update can also be understood as a precorrected BNS update:

Lemma 4.1. *Let the assumptions of Lemma 3.3 be satisfied and $T_Y = L^{-1} D_U$, where D_U is a diagonal matrix with the same main diagonal as U (or L). Then the BFGS update (2.3) with YT_Y instead of Y represents the update (3.2) with $X = U^{-T} U^{-1}$.*

Proof. Let $A = UL$, where U, L have the same main diagonals, and $\tilde{Y} = Y\,T_Y$. Then $S^T \tilde{Y} = A(L^{-1} D_U) = U D_U$, which is an upper triangular matrix with the main diagonal D_U^2. Substituting $U D_U$, $\tilde{Y} = Y L^{-1} D_U$, D_U^2 for R, Y, D into (2.3), we get (3.2) with $X = U^{-T} U^{-1}$ by

$$(UD_U)^{-T} D_U^2 (UD_U)^{-1} = U^{-T} U^{-1}, \quad (UD_U)^{-T} (Y L^{-1} D_U)^T = A^{-T} Y^T. \qquad \square$$

4.2 Corrections which use previous and subsequent difference vectors

In this section we show, how the both of the QN equations $H_+\hat{y} = \hat{s}$ and $H_+y = s$ can be satisfied, using modified corrections (4.1) and (4.4). In this connection, from now on the vectors $\hat{s}, \hat{y}, \tilde{s}_-, \tilde{y}_-$ will have a different meaning than up to now.

For the new type of corrections, which correct both $s, y \to \hat{s}, \hat{y}$ and $s_-, y_- \to \tilde{s}_-, \tilde{y}_-$, we replace the (conjugacy) properties (4.2), (4.5)) by $\hat{s}^T\tilde{y}_- = \tilde{s}_-^T\hat{y} = 0$. Considering advantageous properties of the corrections (4.1), see [19, 20], we want to preserve them partially. By analogy with the BFGS update (see comments after Lemma 2.2 in [19]), we consider the satisfaction of $\hat{s}^Ty_- = 0$ for the new corrected vectors $\hat{s}, \hat{y}$ to be more important than $s_-^T\hat{y} = 0$, which is confirmed by our numerical experiments.

Therefore we will not correct y_-, define $\hat{s}, \hat{y}$ by modified (4.1) with s_- replaced by a corrected vector $\tilde{s}_-$ and simultaneously define $\tilde{s}_-$ by (4.4) with s, y replaced by $\hat{s}, \hat{y}$:

$$\hat{s} = \Big(I - (1/\tilde{s}_-^Ty_-)\tilde{s}_-y_-^T\Big)s, \quad \hat{y} = \Big(I - (1/\tilde{s}_-^Ty_-)y_-\tilde{s}_-^T\Big)y, \quad \tilde{s}_- = \Big(I - (1/\hat{s}^T\hat{y})\hat{s}\hat{y}^T\Big)s_-, \tag{4.6}$$

where $\tilde{s}_-^Ty_- = b_-$ by $\hat{s}^Ty_- = 0$. Obviously, these relations do not determine $\hat{s}, \hat{y}, \tilde{s}_-$ uniquely. Similarly as (4.3) we can obtain $H_+y - s = \Big((\tilde{s}_-^Ty - s^Ty_-)/b_-\Big)\tilde{s}_-$. Thus to have $H_+y = s$, the transformation $s_- \to \tilde{s}_-$ should satisfy the symmetry property $\tilde{s}_-^Ty = s^Ty_-$ (it can be proved that then $\hat{s}, \hat{y}, \tilde{s}_-$, if well defined, are determined by (4.6) uniquely).

The following theorem gives all corrected quantities, describes their properties and shows that all relations (4.6) are satisfied. Note that there is no need to compute $\tilde{s}_-$ explicitly; for the true handling of corrected quantities, see Section 5.

Theorem 4.1. *Let* $\alpha = s^Ty_-/b_-$, $\bar{b} = b - \alpha\, s_-^Ty \neq 0$, $\hat{s} = (s - \alpha s_-)(b - \alpha^2 b_-)/\bar{b}$, $\hat{y} = y - \alpha\, y_-$, $\hat{b} = \hat{s}^T\hat{y} \neq 0$, $\tilde{s}_- = P^Ts_-$, $P = I - (1/\hat{s}^T\hat{y})\,\hat{y}\hat{s}^T$, $\gamma = s_-^Ty - s^Ty_-$. *Then*

(a) $\hat{s}^Ty_- = \tilde{s}_-^T\hat{y} = 0$, $b - \alpha^2 b_- = \hat{b}$, $\tilde{s}_-^Ty_- = b_-$, $\tilde{s}_-^Ty = s^Ty_-$,

(b) $\hat{y} = \Big(I - (1/b_-)y_-\tilde{s}_-^T\Big)y$, $\hat{s} = s - \alpha\tilde{s}_- = \Big(I - (1/b_-)\tilde{s}_-y_-^T\Big)s$.

Moreover, let H_+ be the first or second new update of the form (3.2) with A nonsingular and s, y, s_- replaced by $\hat{s}, \hat{y}, \tilde{s}_-$. If X and H_+ are symmetric positive definite, then

(c) $H_+\hat{y} = \hat{s}$, $H_+y_- = \tilde{s}_-$, $\hat{b} > 0$, $H_+y = s$.

Proof. (a) We get $\hat{s}^Ty_- = 0$ in view of $\alpha = s^Ty_-/b_-$, further $\tilde{s}_-^T\hat{y} = s_-^TP\hat{y} = 0$ and

$$\hat{b} = \hat{s}^T\hat{y} = \hat{s}^T(y - \alpha\, y_-) = \hat{s}^Ty = (b - \alpha\, s_-^Ty)(b - \alpha^2 b_-)/\bar{b} = b - \alpha^2 b_-\,. \tag{4.7}$$

Using again $\hat{s}^Ty_- = 0$, we obtain $\tilde{s}_-^Ty_- = s_-^TPy_- = s_-^Ty_- = b_-$, which yields

$$s_- - \tilde{s}_- = s_- - P^Ts_- = s_- - \Big(I - (1/\hat{b})\hat{s}\hat{y}^T\Big)s_- = (s_-^T\hat{y}/\hat{b})\hat{s} = (\gamma/\hat{b})\hat{s}\,. \tag{4.8}$$

In view of $\hat{s}^Ty = \hat{b}$, see (4.7), from (4.8) we get $\tilde{s}_-^Ty = s_-^Ty - \gamma = s^Ty_- = \alpha b_-$.

(b) The last relation implies $\Big(I - (1/b_-)y_-\tilde{s}_-^T\Big)y = y - \alpha y_- = \hat{y}$. In view of (4.8) the definition of $\hat{s}$ yields

$$s - \alpha s_- = \frac{\bar{b}}{\hat{b}}\,\hat{s} = \frac{b - \alpha^2 b_- - \alpha(s_-^Ty - \alpha b_-)}{\hat{b}}\,\hat{s} = \hat{s} - \alpha\frac{\gamma}{\hat{b}}\,\hat{s} = \hat{s} - \alpha(s_- - \tilde{s}_-), \tag{4.9}$$

i.e. $s = \hat{s} + \alpha\tilde{s}_-$, therefore $\hat{s} = s - \alpha\tilde{s}_- = s - (s^T y_-/b_-)\tilde{s}_- = \Big(I - (1/b_-)\tilde{s}_- y_-^T\Big)s$.

(c) From $\hat{s}^T y_- = \tilde{s}_-^T \hat{y} = 0$ and $\tilde{s}_-^T y_- = b_-$, see (a), we get $H_+[y_-, \hat{y}] = [\tilde{s}_-, \hat{s}]$ by Theorem 3.1 for the first update or by Lemma 3.4 for the second update, both with $S_{(2)} = [\tilde{s}_-, \hat{s}]$, $Y_{(2)} = [y_-, \hat{y}]$. This yields $\hat{b} = \hat{s}^T\hat{y} = \hat{y}^T H_+ \hat{y} > 0$ by assumption and further

$$s - \alpha\tilde{s}_- = \hat{s} = H_+\hat{y} = H_+(y - \alpha y_-) = H_+ y - \alpha\tilde{s}_-$$

by (b), i.e. $H_+ y = s$. □

5 Implementation

In this section we assume that $H^I = \zeta I$, $\zeta = s^T y/y^T y > 0$, and implement two new VM methods, applying the two new updates (Sections 3.1, 3.2) to difference vectors with possible corrections described in Theorem 4.1. To improve efficiency, the corrections $s \to \hat{s}, y \to \hat{y}$ are performed before updating and $\hat{s}, \hat{y}$ are stored and used instead of s, y. To indicate that all columns of S, Y except for the latest (i.e. s, y) can be possibly corrected in previous iterations, we will write $\hat{s}_i, \hat{y}_i$ (and subsequently $\hat{b}_i$), $i < k$. To unify notation, after possible corrections (see Step 4 of Algorithm 5.1) we will often write $\hat{s}_i, \hat{y}_i, \hat{b}_i$ instead of s_i, y_i, b_i also for $i = k$. Nevertheless, if we want to stress that we mean the original, uncorrected vectors, we will write $\mathring{s}_i, \mathring{y}_i, \mathring{b}_i$ for any i.

The vector $\tilde{s}_-$ is used only with the first new update and only together with corrections of s, y and need not be computed. It suffices to consider the matrix $ST \triangleq \tilde{S}$ instead of S (and subsequently $\tilde{A} = \tilde{S}^T Y = T^T A, \tilde{C}, \tilde{X}$ instead of A, C, X), where T is the (low-order) identity matrix, except for $T_{m,m-1} = -\gamma/\hat{b}$ in view of $\tilde{s}_- = s_- - (\gamma/\hat{b})\hat{s}$, see (4.8); obviously $\det T = 1$.

The matrix H_+ modified in this way can be written in the following form

$$\begin{aligned} H_+ &= \tilde{S}\tilde{X}\tilde{S}^T + \zeta\Big(I - \tilde{S}\tilde{A}^{-T}Y^T\Big)\Big(I - Y\tilde{A}^{-1}\tilde{S}^T\Big) \\ &= S\bar{X}S^T + \zeta\Big(I - SA^{-T}Y^T\Big)\Big(I - YA^{-1}S^T\Big), \quad \bar{X} = T\tilde{X}T^T, \end{aligned} \tag{5.1}$$

i.e. in the form (3.2) with $H^I = \zeta I$ and X replaced by $\bar{X}$. Setting $T = I$ and $\bar{X} = X$ in case that we do not correct $\hat{s}_- \to \tilde{s}_-$, we will use the matrices $\tilde{S}, \tilde{A}, \tilde{C}, \tilde{X}$ instead of S, A, C, X and the update (5.1) instead of (3.2) for the both new updates. Note that the matrix $\tilde{A}_{22}$ is diagonal by Theorem 4.1(a).

In view of (5.1), the direction vector $-H_+g_+$ and an auxiliary vector $Y^T H_+ g_+$ (see comments to Procedure 5.2 below) can be calculated efficiently by

$$-H_+g_+ = -\zeta g_+ - S\Big[\Big(\bar{X} + \zeta A^{-T}Y^TYA^{-1}\Big)S^Tg_+ - \zeta A^{-T}Y^Tg_+\Big] + Y\Big[\zeta A^{-1}S^Tg_+\Big], \tag{5.2}$$

$$Y^TH_+g_+ = \zeta Y^Tg_+ + Y^TS\Big[\Big(\bar{X} + \zeta A^{-T}Y^TYA^{-1}\Big)S^Tg_+ - \zeta A^{-T}Y^Tg_+\Big] - Y^TY\Big[\zeta A^{-1}S^Tg_+\Big]. \tag{5.3}$$

For the second new update with $\bar{X} = X = U^{-T}U^{-1}$, we use (5.2) and (5.3) in the form

$$-H_+g_+ = -\zeta g_+ - S\Big[U^{-T}\Big((I + \zeta L^{-T}Y^TYL^{-1})q - \zeta L^{-T}Y^Tg_+\Big)\Big] + Y\Big[\zeta L^{-1}q\Big], \tag{5.4}$$

$$Y^TH_+g_+ = \zeta Y^Tg_+ + Y^TS\Big[U^{-T}\Big((I + \zeta L^{-T}Y^TYL^{-1})q - \zeta L^{-T}Y^Tg_+\Big)\Big] - Y^TY\Big[\zeta L^{-1}q\Big], \tag{5.5}$$

where $q = U^{-1}S^Tg_+$.

All corrections described in Section 4 and also both new updates appear to be more efficient than the L-BFGS (or BNS) method only if A is sufficiently close to a symmetric matrix. Denoting $A = [a_{ij}]_{i,j=1}^m$, we use the values

$$\bar\Delta = \sum_{1\le i<j\le m} (a_{ij}-a_{ji})^2/(a_{ii}a_{jj}), \quad \bar\delta = (\hat s_-^T y - s^T \hat y_-)^2/(b\hat b_-) \tag{5.6}$$

as measures of the deviation. Similarly as for the BFGS method with corrected vectors (to have VM matrices symmetric positive definite, see e.g. [17]), we require $\hat b > 0$ and $\bar b > 0$ (we can readily verify that $\bar b$ corresponds to $\hat s^T \hat y$ for the corrections (4.1)). Besides, as in [22] we can deduce that too small $\bar b, \hat b$ can deteriorate stability. Thus we do not correct s, y if $\bar\delta > \delta_1$ or $\bar b < \delta_2 b$ (all $\delta_i \in (0,1)$); this implies $\hat b > 0$ by Lemma 6.3. Due to our proof of global convergence, we also do not correct if $\max[|\hat s_-|/|\mathring{s}_-|, |\hat y_-|/|\mathring{y}_-|] > \theta$, $\theta > 1$.

Due to $\bar b = \hat b - \alpha\gamma$, see (4.9), i.e. $1 - \bar b/\hat b = \alpha\gamma/\hat b$, and $|\tilde s_- - \hat s_-|/|\hat s| = |\gamma|/\hat b$, see (4.8), in case of the first new update we also do not correct if $(\alpha\gamma/\hat b)^2 > \delta_3$ or $(\gamma/\hat b)^2 > \delta_4$ ($\tilde s_-$ is too different from $\hat s_-$). Similarly for the second new update, if $(\alpha\gamma/\hat b)^2 > \delta_5$.

We use both new updates only for $\bar\Delta \le \delta_6$. For the first new update, if all diagonal entries of $\tilde A_{22}$ are greater than $\varepsilon_D \operatorname{Tr} A$, $\varepsilon_D \in (0,1)$, we calculate $\tilde C$, using Theorem 3.1 with $\tilde S, \tilde A, \tilde C$ instead of S, A, C (recall that all $b_i > 0$ for $g_i \ne 0$ by (1.1)). Then, if all diagonal entries of $\tilde C$ are greater than $\delta_7 \operatorname{Tr} A$ and all eigenvalues of $\tilde C \tilde C^T$ divided by $1 + \|\tilde A_{22}^{-1} \tilde A_{21}\|_F^2$ (to guarantee global convergence, see Section 6) are greater than or equal to $\varepsilon_E \operatorname{Tr} A$, $\varepsilon_E \in (0,1)$, we calculate $\tilde X$, using Theorem 3.2 and (3.7) with $\tilde S, \tilde A, \tilde C, \tilde X$ instead of S, A, C, X. For the second new update, we factorize $A = UL$, where U, L have the same main diagonals, and calculate $X = U^{-T} U^{-1}$, see Lemma 3.3. If some condition is not satisfied or if the factorization fails, we use the BNS method, see Section 2.

For calculation of eigenvalues and eigenvectors of low-order symmetric matrices, the well-known Jacobi iteration method (e.g. [6]) appears to be efficient. It is interesting that the increase in computational time for one iteration compared with the second new update or the standard BNS update is very small for N large (it is independent of N).

For our proof of global convergence we need $\det \tilde A \ne 0$. This is guaranteed by Theorem 1.4.2 in [6] for the first new update, since $\tilde A_{22}$ and $\tilde C$ are nonsingular, or by Procedure 5.1 for the second update, since we require $L_{ii}^2 \ge \varepsilon_F \max[\operatorname{Tr} A, \operatorname{Tr}(L^T L)]$, $1 \le i \le m$, $\varepsilon_F \in (0,1)$, see below; note that $L^T L = A$ for A symmetric.

We first present two auxiliary procedures. Procedure 5.1, based on Lemma 5 in [21], is used for the factorization $A = UL$, Procedure 5.2 serves for updating of the basic matrices $S, Y, S^T Y = A, Y^T Y$; the submatrices of S, Y with columns from previous iterations are denoted by S_P, Y_P (i.e. $S = [S_P, \hat s]$, $Y = [Y_P, \hat y]$). In comparison with the corresponding algorithm in [3], which uses only the main diagonal of A and the part of A above the diagonal, we need all entries of A here. Thus we use an additional vector $Y_P^T s = -t\, Y_P^T H g$ (see Algorithm 5.1) to have the number of arithmetic operations approximately the same.

Procedure 5.1 *(UL factorization of A)*

Given: A global convergence parameter $\varepsilon_F \in (0,1)$ and the $m \times m$ matrix A.

(i): Set $Q := A$, $\nu := m$ and $\kappa := Q_{mm}$.

(ii): If $Q_{\nu\nu} < \varepsilon_F \operatorname{Tr} A$ then the factorization fails and return.

(iii): Set $Q_{ij} := Q_{ij} - Q_{i\nu} Q_{\nu j}/Q_{\nu\nu}$, $i = 1, \dots, \nu-1$, $j = 1, \dots, \nu-1$. Set $\nu := \nu - 1$ and

then $\kappa := \min[\kappa, Q_{\nu\nu}]$. If $\nu > 1$ go to *(ii)*.

(iv): Set $L_{ij} := Q_{ij}/\sqrt{Q_{ii}}$, $U_{ji} := Q_{ji}/\sqrt{Q_{ii}}$ for $1 \le j \le i \le m$ and $L_{ij} := U_{ji} := 0$ for $1 \le i < j \le m$. If $\kappa < \varepsilon_F \|L\|_F^2$ then the factorization fails. Return.

Procedure 5.2 *(Updating of basic matrices)*

Given: t, $\hat{b}/\bar{b}$, $\alpha \in \mathcal{R}$, matrices $S_P, Y_P, S_P^T Y_P, Y_P^T Y_P$ and vectors $\hat{s}, \hat{y}, g_+, S_P^T g, Y_P^T g, Y_P^T Hg$.

(i): Compute $S_P^T g_+$, $Y_P^T g_+$, $\hat{s}^T g_+$, $\hat{y}^T g_+$, $\hat{s}^T \hat{y}$, $\hat{y}^T \hat{y}$.

(ii): Compute $S_P^T y := S_P^T g_+ - S_P^T g$, $Y_P^T y := Y_P^T g_+ - Y_P^T g$, $Y_P^T s := -t\, Y_P^T Hg$.

(iii): Compute $S_P^T \hat{y} := S_P^T y - \alpha S_P^T y_-$, $Y_P^T \hat{s} := (Y_P^T s - \alpha Y_P^T s_-)\, \hat{b}/\bar{b}$, $Y_P^T \hat{y} := Y_P^T y - \alpha Y_P^T y_-$.

(iv): Set $S := [\, S_P, \hat{s}]$, $Y := [\, Y_P, \hat{y}]$, $S^T g_+ := [S_P^T g_+, \hat{s}^T g_+]$, $Y^T g_+ := [Y_P^T g_+, \hat{y}^T g_+]$.

(v): Set $A = S^T Y := \begin{bmatrix} S_P^T Y_P & S_P^T \hat{y} \\ \hat{s}^T Y_P & \hat{s}^T \hat{y} \end{bmatrix}$, $Y^T Y := \begin{bmatrix} Y_P^T Y_P & Y_P^T \hat{y} \\ \hat{y}^T Y_P & \hat{y}^T \hat{y} \end{bmatrix}$ and return.

We now state the method in details. For simplicity, we do not describe stopping criteria and contingent restarts when some computed direction vector is not a sufficiently descent direction. Note that the order μ of A_{22} is also used as a correction indicator and that the contingent restarts have occurred very rarely in our numerical experiments.

Algorithm 5.1

Data: A maximum number $\hat{m}$ of columns S, Y, line search parameters $\varepsilon_1, \varepsilon_2$, $0 < \varepsilon_1 < 1/2$, $\varepsilon_1 < \varepsilon_2 < 1$, tolerance parameters $\delta_i \in (0, 1)$, $i \in \{1, \ldots, 7\}$, a global convergence parameters $\theta > 1$, $\varepsilon_D, \varepsilon_E, \varepsilon_F \in (0, 1)$ and a chosen method number $n_M \in \{1, 2\}$.

Step 1: Initiation. Choose starting point $x_0 \in \mathcal{R}^N$, define the starting matrix $H_0 := I$ and the direction vector $d_0 := -g_0$ and initiate the iteration counter k to zero.

Step 2: Line search. Set the update indicator i_U to zero. Compute $x_+ := x + td$, where t satisfies (1.1), $g_+ := \nabla f(x_+)$, $s := td$, $y := g_+ - g$, $b := s^T y$ and $\zeta := b/y^T y$. Set $\tilde{m} := \min[k, \hat{m} - 1]$, $m := \tilde{m} + 1$ and define $H^I := \zeta I$. If $k = 0$ set $S := [s]$, $Y := [y]$, $S^T Y := [s^T y]$, $Y^T Y := [y^T y]$, $X = [1/b]$, compute $S^T g_+$, $Y^T g_+$ and go to Step 9.

Step 3: Correction preparation. Set $\mu := 1$. If $m > 1$ compute $\hat{b}, \bar{b}, \alpha, \gamma$ by Theorem 4.1 and $\bar{\delta}, \bar{\Delta}$ by (5.6). If $\bar{\delta} < \delta_1$, $\hat{b} > 0$, $\bar{b} > \delta_2 b$ and $\max[|\hat{s}_-|/|\mathring{s}_-|, |\hat{y}_-|/|\mathring{y}_-|] \le \theta$ set $\mu := 2$. If $n_M = 1$ and $\big((\alpha\gamma/\hat{b})^2 > \delta_3$ or $(\gamma/\hat{b})^2 > \delta_4\big)$ or if $n_M = 2$ and $(\alpha\gamma/\hat{b})^2 > \delta_5$ set $\mu := 1$.

Step 4: Correction. If $\mu = 1$ set $\hat{s} := s$, $\hat{y} := y$, otherwise compute $\hat{s}, \hat{y}$ by Theorem 4.1.

Step 5: Basic matrices updating. Using Procedure 5.2, form the matrices $S, Y, A, Y^T Y$.

Step 6: Update selection. If $\bar{\Delta} > \delta_6$ go to Step 9. Set $T := I$. If $n_M = 2$ go to Step 8. If $\mu = 2$, form the matrix T according to the second paragraph of Section 5 and define $\tilde{S} := ST$ and $\tilde{A} := T^T A$, otherwise define $\tilde{S} = S$ and $\tilde{A} = A$.

Step 7: VM update 1. Set $i_U := 1$. Use Theorems 3.2 and 3.1 with $\tilde{S}, \tilde{A}, \tilde{C}, \tilde{X}$ instead of S, A, C, X to compute $\tilde{C}$, $\tilde{X}$, if all diagonal entries of $\tilde{A}_{22}$ are greater then $\varepsilon_D \mathrm{Tr}\, A$ and if all diagonal entries of $\tilde{C}$ are greater then $\delta_7 \mathrm{Tr}\, A$ and if all eigenvalues of $(\tilde{C}\tilde{C}^T)^{1/2}$ divided by $1 + \|\tilde{A}_{22}^{-1} \tilde{A}_{21}\|_F^2$ are greater than or equal to $\varepsilon_E \mathrm{Tr}\, A$, otherwise set $i_U := 0$. Go to Step 9.

Step 8: VM update 2. Use Procedure 5.1 to factorize $A := UL$. If the factorization fails, go to Step 9. Set $i_U := 2$ and compute $X := U^{-T} U^{-1}$.

Step 9: Direction vector. Define H_+ by (2.3) for $i_U = 0$ or by (5.1) otherwise. Compute $d_+ = -H_+ g_+$ and an auxiliary vector $Y^T H_+ g_+$ by (2.4)–(2.5) for $i_U = 0$ or by (5.2)–(5.3) with $\bar{X} = T\tilde{X}T^T$ for $i_U = 1$ or by (5.4)–(5.5) for $i_U = 2$. Set $k := k+1$. If $k \geq \hat{m}$ delete the first column of S_-, Y_- and the first row and column of $S_-^T Y_-$, $Y_-^T Y_-$ to form S_P, Y_P, $S_P^T Y_P$, $Y_P^T Y_P$. Go to Step 2.

6 Global convergence

In this section we establish global convergence of Algorithm 5.1 in convex case and without restarts. Assumption 6.1 and Lemma 6.1 are presented in [19], Lemma 6.2 in [21].

Note that a suitable restarts technique can guarantee global convergence also for non-convex f, see e.g. Algorithm 6.1 and comments in the beginning of Section 7 in [22]. Besides, there are some other possibilities how to establish global convergence of VM methods for non-convex f, see e.g. [9, 23].

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. $0 < \underline{G} \leq \underline{\lambda}(G(x)) \leq \overline{\lambda}(G(x)) \leq \overline{G} < \infty$, $x \in \mathcal{R}^N$, where $\underline{\lambda}(G(x))$ and $\overline{\lambda}(G(x))$ are the lowest and the greatest eigenvalues of the Hessian matrix $G(x)$).*

Lemma 6.1. *Let the objective function f satisfy Assumption 6.1. Then $\underline{G} \leq |\mathring{y}|^2/\mathring{b} \leq \overline{G}$ and $\mathring{b}/|\mathring{s}|^2 \geq \underline{G}$ ($\mathring{s}, \mathring{y}$ are original, uncorrected difference vectors, see Section 5).*

Lemma 6.2. *Let $K_1, K_2 \in \mathcal{R}^{\nu\times\nu}$, $\nu > 0$, be symmetric positive semidefinite matrices. Then $0 \leq \mathrm{Tr}(K_1K_2) \leq \mathrm{Tr}\, K_1 \,\mathrm{Tr}\, K_2$. Moreover, if K_2 is symmetric positive definite, then $\mathrm{Tr}(K_1K_2^{-1}) \leq \mathrm{Tr}\, K_1\, (\mathrm{Tr}\, K_2)^{\nu-1}/\det K_2$.*

Lemma 6.3. *Let the assumptions of Theorem 3.1 be satisfied with $X_{11} = (CC^T)^{-1/2}$ and $A_{22} \in \mathcal{S}_\mu$ (symmetric positive definite of order μ, see Section 3). Then $A^TXA \in \mathcal{S}_m$ and*

$$\begin{aligned} \mathrm{Tr}\,(A^TXA)^{-1} \;&\leq\; \Big(1 + \|A_{22}^{-1}\, A_{21}\|_F^2\Big)\,\mathrm{Tr}\, X_{11} + \mathrm{Tr}\, A_{22}^{-1}, \\ \det X^{-1} \;&=\; \det X_{11}^{-1}\, \det A_{22}\,. \end{aligned}$$

Proof. We have $X_{11}^{-1} \sim C^T X_{11} C \triangleq M$ by Theorem 3.2. Further, (3.11) and (3.12) imply

$$A^TXA = \begin{bmatrix} M + A_{21}^T A_{22}^{-1} A_{21} & A_{21}^T \\ A_{21} & A_{22} \end{bmatrix} \in \mathcal{S}_m,$$

since the Schur complement of A_{22} in A^TXA is $M \in \mathcal{S}_{m-\mu}$. It can easily be verified that

$$(A^TXA)^{-1} = \begin{bmatrix} M^{-1} & -M^{-1}A_{21}^T A_{22}^{-1} \\ -A_{22}^{-1}A_{21}M^{-1} & A_{22}^{-1} + A_{22}^{-1}A_{21}M^{-1}A_{21}^T A_{22}^{-1} \end{bmatrix}.$$

The trace of a product of two matrices is independent of the order of multiplication (if the both products are defined), which gives $\mathrm{Tr}\,(A^TXA)^{-1} = \mathrm{Tr}\, M^{-1} + \mathrm{Tr}\,(A_{21}^T A_{22}^{-2} A_{21} M^{-1}) + \mathrm{Tr}\, A_{22}^{-1}$. Using Lemma 6.2, we obtain

$$\begin{aligned} \mathrm{Tr}\,(A^TXA)^{-1} \;&\leq\; \Big(1 + \mathrm{Tr}\,(A_{21}^T A_{22}^{-2} A_{21})\Big)\,\mathrm{Tr}\, M^{-1} + \mathrm{Tr}\, A_{22}^{-1} \\ &=\; \Big(1 + \|A_{22}^{-1}\, A_{21}\|_F^2\Big)\,\mathrm{Tr}\, X_{11} + \mathrm{Tr}\, A_{22}^{-1} \end{aligned}$$

and the rest follows from (3.9) by Theorem 1.4.2 in [6]. □

Lemma 6.4. *Let $\hat{b}, \bar{b}, \alpha, \gamma$ are given by Theorem 4.1, $\delta, \tilde{\delta} \in (0,1)$ and suppose that $\bar{b} > \delta b$ and $|\alpha\gamma| \le \tilde{\delta}\,\hat{b}$. Then $\hat{b}/\bar{b} \le 1/(1-\tilde{\delta})$ and $\hat{b} > (\delta/2)b$.*

Proof. In the same way as in (4.9) we get $\bar{b} = \hat{b} - \alpha\gamma$, therefore

$$\hat{b}(1-\tilde{\delta}) \le |\hat{b}| - \hat{b}\tilde{\delta} \le |\hat{b}| - |\alpha\gamma| \le |\hat{b} - \alpha\gamma| = \bar{b}, \quad \bar{b} \le |\hat{b}| + |\alpha\gamma| < 2\,\hat{b},$$

which yields $\hat{b}/\bar{b} \le 1/(1-\tilde{\delta})$ and $\hat{b} > \bar{b}/2 > (\delta/2)\,b$. □

Lemma 6.5. *Let objective function f satisfy Assumption 6.1. Then Algorithm 5.1 guarantees that the sequences $\{|\hat{s}_k|^2/\hat{b}_k\}$, $\{|\hat{y}_k|^2/\hat{b}_k\}$ (corrected or uncorrected) are always bounded.*

Proof. Since $|\mathring{s}|^2/\mathring{b} \le 1/\underline{G}$ and $|\mathring{y}|^2/\mathring{b} \le \overline{G}$ by Lemma 6.1, the assertion holds for $\hat{s}_k, \hat{y}_k, \hat{b}_k$ without corrections ($\mu_k = 1$). Let $\mu_k = 2$. The safeguarding technique in Step 3 of Algorithm 5.1 guarantees

$$\hat{b} > 0, \quad \bar{b} > \delta_2 b, \quad |\hat{s}_-| \le \theta|\mathring{s}_-|, \quad |\hat{y}_-| \le \theta|\mathring{y}_-|, \quad |\alpha\gamma| \le \tilde{\delta}\,\hat{b}, \;\; \tilde{\delta} = \sqrt{\max[\delta_3, \delta_5]}. \tag{6.1}$$

Setting $\delta = \delta_2$ and using Lemma 6.4, we have $\hat{b}/\bar{b} \le 1/(1-\tilde{\delta})$ and $\hat{b}/b > \delta/2$. Similarly we obtain $\hat{b}_-/\mathring{b}_- > \delta/2$ for $\mu_- = 2$, which is also true for $\mu_- = 1$ by $\hat{b}_-/\mathring{b}_- = 1 > \delta/2$. Thus

$$\hat{b}/\bar{b} \le 1/(1-\tilde{\delta}), \quad \hat{b}/b > \delta/2, \quad \hat{b}_-/\mathring{b}_- > \delta/2 \tag{6.2}$$

holds for any μ_-. Further we get $\alpha^2 < b/\hat{b}_-$ by $\hat{b} = b - \alpha^2\hat{b}_- > 0$, which yields

$$\begin{aligned} |s - \alpha\hat{s}_-| &\le |s| + \sqrt{b/\hat{b}_-}\,|\hat{s}_-| \le |s| + \sqrt{2b/\delta}\,\theta|\mathring{s}_-|/\sqrt{\mathring{b}_-} \le \sqrt{b/\underline{G}}\Big(1+\sqrt{2/\delta}\,\theta\Big), \\ |y - \alpha\hat{y}_-| &\le |y| + \sqrt{b/\hat{b}_-}\,|\hat{y}_-| \le |y| + \sqrt{2b/\delta}\,\theta|\mathring{y}_-|/\sqrt{\mathring{b}_-} \le \sqrt{b\,\overline{G}}\Big(1+\sqrt{2/\delta}\,\theta\Big), \end{aligned}$$

by Lemma 6.1, $\hat{b}_- > (\delta/2)\mathring{b}_-$, $|\hat{s}_-| \le \theta|\mathring{s}_-|$ and $|\hat{y}_-| \le \theta|\mathring{y}_-|$, see (6.1)–(6.2). This implies

$$\begin{aligned} \frac{|\hat{s}_k|^2}{\hat{b}_k} &= \frac{|s_k - \alpha_k\hat{s}_{k-1}|^2(\hat{b}_k/\bar{b}_k)^2}{\hat{b}_k} \le \frac{b_k\Big(1+\sqrt{2/\delta}\,\theta\Big)^2}{\underline{G}\,\hat{b}_k(1-\tilde{\delta})^2} \le \frac{2\Big(1+\sqrt{2/\delta}\,\theta\Big)^2}{\delta\,\underline{G}\,(1-\tilde{\delta})^2}, \\ \frac{|\hat{y}_k|^2}{\hat{b}_k} &= \frac{|y_k - \alpha_k\hat{y}_{k-1}|^2}{\hat{b}_k} \le \frac{b_k\,\overline{G}}{\hat{b}_k}\Big(1+\sqrt{2/\delta}\,\theta\Big)^2 \le \frac{2\,\overline{G}}{\delta}\Big(1+\sqrt{2/\delta}\,\theta\Big)^2 \end{aligned}$$

by Theorem 4.1 and (6.2). □

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. Using Lemma 6.5, we can find $\theta_1, \theta_2 \in \mathcal{R}$ satisfying

$$|\hat{s}_k|^2/\hat{b}_k < \theta_1, \tag{6.3}$$
$$|\hat{y}_k|^2/\hat{b}_k < \theta_2 \tag{6.4}$$

for all $k \ge 0$, where $\hat{s}_k, \hat{y}_k$ mean corrected or uncorrected vectors in the whole proof.

As we mentioned in Section 5, in all iterations we choose $H_k^I = \zeta_k I$, $\zeta_k = \mathring{b}_k/|\mathring{y}_k|^2$, see Step 2. Denoting $B_k^I = (H_k^I)^{-1}$, Lemma 6.1 gives

$$\mathrm{Tr}\,B_k^I = (|\mathring{y}_k|^2/\mathring{b}_k)\,\mathrm{Tr}\,I \le N\overline{G}, \quad \det B_k^I = (|\mathring{y}_k|^2/\mathring{b}_k)^N \ge \underline{G}^N, \quad k \ge 0. \tag{6.5}$$

(i) Suppose that $i_U = 0$, i.e. the BNS update (2.3) of H_k^I is used and columns of S, Y are $\hat{s}_i, \hat{y}_i$. This is equivalent to the recurrent application of the BFGS update (2.1) to H_k^I

$$H_{i+1}^{k+1} = \frac{1}{\hat b_i}\hat s_i \hat s_i^T + \Big(I - \frac{1}{\hat b_i}\hat s_i \hat y_i^T\Big) H_i^{k+1}\Big(I - \frac{1}{\hat b_i}\hat y_i \hat s_i^T\Big), \quad i \in \{k-\tilde m, \dots, k\} \triangleq \mathcal{I}_k, \tag{6.6}$$

where $H_{k-\tilde m}^{k+1} = H_k^I = \zeta_k I$, $H_{k+1} = H_{k+1}^{k+1}$. These updates satisfy (see [17])

$$\begin{aligned} \operatorname{Tr} B_{i+1}^{k+1} &= \operatorname{Tr} B_i^{k+1} + |\hat y_i|^2/\hat b_i - |B_i^{k+1}\hat s_i|^2/\hat s_i^T B_i^{k+1}\hat s_i, & (6.7)\\ \det B_{i+1}^{k+1} &= \Big(\hat b_i/\hat s_i^T B_i^{k+1}\hat s_i\Big)\det B_i^{k+1}, & (6.8)\end{aligned}$$

$i \in \mathcal{I}_k$, denoting $B_i^{k+1} = (H_i^{k+1})^{-1}$, $i = k-\tilde m, \dots, k+1$. This yields

$$\operatorname{Tr} B_i^{k+1} \le N\overline{G} + m\,\theta_2 \triangleq \theta_3 \tag{6.9}$$

by (6.4)–(6.5). Since $\hat b_i/\hat s_i^T B_i^{k+1}\hat s_i = (\hat b_i/|\hat s_i|^2)(|\hat s_i|^2/\hat s_i^T B_i^{k+1}\hat s_i) \ge 1/(\theta_1\theta_3)$ by (6.3) and (6.9), for all $k > 0$ in view of (6.9), (6.8) and (6.5) we get

$$\begin{aligned} \operatorname{Tr} B_{k+1} &= \operatorname{Tr} B_{k+1}^{k+1} \le \theta_3, & (6.10)\\ \det B_{k+1} &= \det B_{k+1}^{k+1} \ge \underline{G}^N/(\theta_1\theta_3)^m \triangleq \theta_4. & (6.11)\end{aligned}$$

(ii) Let $i_U > 0$, i.e. the update (5.1) is used (and columns of S, Y are again $\hat s_i, \hat y_i$), where for $n_M = 1$ we use matrices $\tilde S = ST, \tilde A, \tilde C, \tilde X$ instead of S, A, C, X with $\det T = 1$, see the beginning of Section 5. In view of $B_k^I = (1/\zeta_k)I$, (6.5), $\tilde S^T\tilde S = T^T(S^TS)\,T$ and $\det \tilde S^T\tilde S \ne 0$ due to $\det \tilde A \ne 0$ (see Section 5), for $k > 0$ Theorem 3 in [21] gives

$$\begin{aligned} B_{k+1} &= (1/\zeta_k)I - (1/\zeta_k)\tilde S_k(\tilde S_k^T\tilde S_k)^{-1}\tilde S_k^T + Y_k\tilde A_k^{-1}\tilde X_k^{-1}\tilde A_k^{-T}Y_k^T, & (6.12)\\ \operatorname{Tr} B_{k+1} &\le N\overline{G} + \operatorname{Tr}(Y_k^TY_k\,\tilde A_k^{-1}\tilde X_k^{-1}\tilde A_k^{-T}) = N\overline{G} + \operatorname{Tr}\Big(Y_k^TY_k(\tilde A_k^T\tilde X_k\tilde A_k)^{-1}\Big), & (6.13)\\ \det B_{k+1} &= \zeta_k^{m-N}\det \tilde X_k^{-1}/\det \tilde S_k^T\tilde S_k \ge \underline{G}^{N-m}\det \tilde X_k^{-1}/\det S_k^TS_k\,. & (6.14)\end{aligned}$$

(ii-a) For $n_M = 1$, Step 7 of Algorithm 5.1 guarantees $\operatorname{Tr}\tilde A_{22}^{-1} \le \mu/(\varepsilon_D \operatorname{Tr} A)$ (recall that $\tilde A_{22}$ is diagonal, see Section 5) and $\omega \operatorname{Tr}\tilde X_{11} \le (m-\mu)/(\varepsilon_E \operatorname{Tr} A)$, where $\omega = 1 + \|\tilde A_{22}^{-1}\tilde A_{21}\|_F^2$. Using Lemma 6.3 with $\tilde A, \tilde X$ instead of A, X, we have $\operatorname{Tr}(\tilde A^T\tilde X\tilde A)^{-1} \le m/(\varepsilon_M \operatorname{Tr} A)$, where $\varepsilon_M = \min[\varepsilon_D, \varepsilon_E]$. From (6.13) we get

$$\begin{aligned} \operatorname{Tr} B_{k+1} &\le N\overline{G} + \operatorname{Tr}(Y_k^TY_k)\operatorname{Tr}(\tilde A_k^T\tilde X_k\tilde A_k)^{-1} \le N\overline{G} + m\frac{\operatorname{Tr}(Y_k^TY_k)}{\varepsilon_M \operatorname{Tr} A_k}\\ &= N\overline{G} + \frac{m\sum_{i\in\mathcal{I}_k}|\hat y_i|^2}{\varepsilon_M\sum_{i\in\mathcal{I}_k}\hat b_i} \le N\overline{G} + \frac{m}{\varepsilon_M}\sum_{i\in\mathcal{I}_k}\frac{|\hat y_i|^2}{\hat b_i} \le \theta_5 & (6.15)\end{aligned}$$

with $\theta_5 = N\overline{G} + m^2\theta_2/\varepsilon_M > \theta_3$ by Lemma 6.2, (6.4), $m > 1$ and $\varepsilon_M < 1$.

We proceed to estimate $\det B_{k+1}$. Similarly as above, Step 7 of Algorithm 5.1 guarantees $\det \tilde A_{22} \ge (\varepsilon_D \operatorname{Tr} A)^\mu$ and $\det \tilde X_{11}^{-1} > (\varepsilon_E \operatorname{Tr} A)^{m-\mu}$ by $\omega \ge 1$. Using Lemma 6.3 with $\tilde A, \tilde X$ instead of A, X, we have $\det \tilde X^{-1} > (\min[\varepsilon_D, \varepsilon_E]\operatorname{Tr} A)^m = (\varepsilon_M \operatorname{Tr} A)^m$. Using Lemma 6.2, (6.3) and the geometric mean - arithmetic mean inequality, from (6.14) we get

$$(\underline{G}^{m-N}\det B_{k+1})^{1/m} \ge \frac{(\det \tilde X_k^{-1})^{1/m}}{\operatorname{Tr}(S_k^TS_k)/m} > m\frac{\varepsilon_M\sum_{i\in\mathcal{I}_k}\hat b_i}{\sum_{i\in\mathcal{I}_k}|\hat s_i|^2} \ge \frac{m\,\varepsilon_M}{\sum_{i\in\mathcal{I}_k}|\hat s_i|^2/\hat b_i} \ge \frac{\varepsilon_M}{\theta_1},$$

i.e.

$$\det B_{k+1} > (\varepsilon_M/\theta_1)^m\,\underline{G}^{N-m} \triangleq \theta_6. \tag{6.16}$$

(ii-b) For $n_M = 2$ we have $A_k = U_k L_k$ and $X_k = U_k^{-T} U_k^{-1}$, which yields $A_k^T X_k A_k = L_k^T L_k$. Using Lemma 6.2, from (6.13) we get

$$\operatorname{Tr} B_{k+1} \le N\overline{G} + \operatorname{Tr}(Y_k^T Y_k\,(L_k^T L_k)^{-1}) \le N\overline{G} + \frac{\operatorname{Tr}(Y_k^T Y_k)\,\operatorname{Tr}(L_k^T L_k)^{m-1}}{\det(L_k^T L_k)}. \tag{6.17}$$

The conditions in Procedure 5.1 imply $(L_k)_{ii}^2 \ge \varepsilon_F \max[\operatorname{Tr} A_k, \operatorname{Tr}(L_k^T L_k)]$, $1 \le i \le m$, thus $\det(L_k^T L_k) \ge \varepsilon_F^m \operatorname{Tr} A_k\,(\operatorname{Tr}(L_k^T L_k))^{m-1}$. Using (6.17), Lemma 6.1 and (6.4) we get

$$\operatorname{Tr} B_{k+1} \le N\overline{G} + \frac{\operatorname{Tr}(Y_k^T Y_k)}{\varepsilon_F^m \operatorname{Tr} A_k} = N\overline{G} + \frac{\sum_{i\in\mathcal{I}_k} |\hat{y}_i|^2}{\varepsilon_F^m \sum_{i\in\mathcal{I}_k} \hat{b}_i} \le N\overline{G} + \varepsilon_F^{-m} \sum_{i\in\mathcal{I}_k} \frac{|\hat{y}_i|^2}{\hat{b}_i} \le \theta_7, \tag{6.18}$$

$\theta_7 = N\overline{G} + \varepsilon_F^{-m} m\theta_2 > \theta_3$ by (6.9) and $\varepsilon_F < 1$. Since $X^{-1} = UU^T$ and the main diagonals of L, U are identical with $U_{ii}^2 = L_{ii}^2 \ge \varepsilon_F \operatorname{Tr} A = \varepsilon_F \sum_{i\in\mathcal{I}_k} \hat{b}_i$, we obtain similarly as above

$$(\underline{G}^{m-N} \det B_{k+1})^{1/m} \ge \frac{(\det(U_k U_k^T))^{1/m}}{\operatorname{Tr}(S_k^T S_k)/m} \ge m \frac{\varepsilon_F \sum_{i\in\mathcal{I}_k} \hat{b}_i}{\sum_{i\in\mathcal{I}_k} |\hat{s}_i|^2} \ge \frac{m\,\varepsilon_F}{\sum_{i\in\mathcal{I}_k} |\hat{s}_i|^2/\hat{b}_i} \ge \frac{\varepsilon_F}{\theta_1}$$

by (6.14), (6.3) and the geometric mean - arithmetic mean inequality. Thus we can write

$$\det B_{k+1} \ge \min[(\varepsilon_F/\theta_1)^m \underline{G}^{N-m}, \theta_4, \theta_6] \triangleq \theta_8\,. \tag{6.19}$$

(iii) Setting $\theta_9 = \max[\theta_5, \theta_7] > \theta_3$, we always have $\operatorname{Tr} B_i \le \theta_9$ and $\det B_i \ge \theta_8$, $i > 1$, by (6.10) – (6.19). The lowest eigenvalue $\underline{\lambda}(B_i)$ of B_i satisfies $\underline{\lambda}(B_i) \ge \det B_i/(\operatorname{Tr} B_i)^{N-1}$. Setting $q_i = H_i^{1/2} g_i$, we get

$$\frac{(\mathring{s}_i^T g_i)^2}{|\mathring{s}_i|^2 |g_i|^2} = \frac{\mathring{s}_i^T B_i \mathring{s}_i}{\mathring{s}_i^T \mathring{s}_i} \frac{g_i^T H_i g_i}{g_i^T g_i} = \frac{\mathring{s}_i^T B_i \mathring{s}_i}{\mathring{s}_i^T \mathring{s}_i} \frac{q_i^T q_i}{q_i^T B_i q_i} \ge \frac{\det B_i}{(\operatorname{Tr} B_i)^{N-1}} \frac{1}{\operatorname{Tr} B_i} \ge \frac{\theta_8}{\theta_9^N}, \quad i > 1, \tag{6.20}$$

which implies $\lim_{i\to\infty} |g_i| = 0$, see Theorem 3.2 in [17] and relations (3.17) – (3.18) ibid. □

One can show in the same way as in [10] that (6.20) with line search conditions (1.1) and Assumption 6.1 imply that the sequence $\{x_i\}$ is at least R-linearly convergent.

7 Numerical experiments

In this section, we compare our results with the results obtained by the L-BFGS method [10, 16] and by our latest limited-memory method [22].

All methods are implemented in the optimization software system UFO [15], which can be downloaded from `www.cs.cas.cz/luksan/ufo.html`. We use the following collections of test problems:

- **Test 11** – 55 chosen problems from [13] (computed repeatedly five times for a better comparison), which are problems from the CUTE collection [2], some of them modified; used N are given in Table 1, where the modified problems are marked with '*',
- **Test 12** – 73 problems from [1], $N = 10\,000$,
- **Test 25** – 67 chosen problems from [12], which are sparse test problems for unconstrained optimization, contained in the system UFO, $N = 10\,000$.

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`.

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 – the modified CUTE collection.

We have chosen $\hat{m}=5$, which is an often used value in comparisons of limited-memory methods. In [17] the results for the L-BFGS method with $\hat{m} = 3, 5, 17, 29$ are compared and it is stated that the best CPU time is often obtained for small values of $\hat{m}$, but the algorithm tends to be less robust when $\hat{m}$ is small; this is also confirmed by our numerical experiments. Note that the required amount of storage is $2(\hat{m} + 1)N$.

Furthermore, we have used $\delta_1=10^{-2}$, $\delta_2=\varepsilon_E=10^{-5}$, $\delta_3=\delta_5=0.025$, $\delta_4=0.05$, $\delta_6=0.5$, $\delta_7=\varepsilon_D=\varepsilon_F=10^{-7}$, $\theta=10^3$, $\varepsilon_1=10^{-4}$, $\varepsilon_2=0.9$ 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 12		Test 25	
	NFV	Time	NFV	Time	NFV	Time
L-BFGS	79575	20.546	114083	132.00	501657	951.55
Alg. 6.1 in [22]	59735	14.566	64242	61.39	407421	714.81
Alg. 5.1, met 1	60198	14.416	64815	62.56	372704	699.08
Alg. 5.1, met 2	60083	14.422	63482	63.11	366824	917.49

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 [5]. The performance profile $\rho_M(\tau)$, $\tau \ge 0$, is defined by the formula

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

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

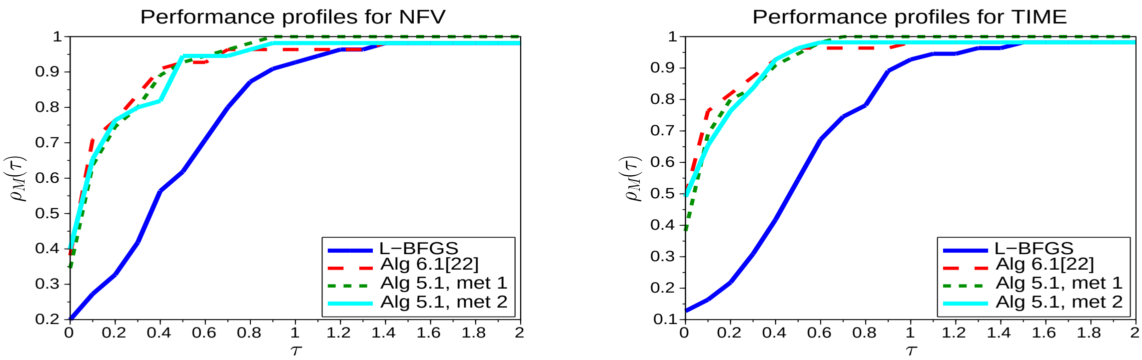

Figure 7.1: Comparison of $\rho_M(\tau)$ for Test 11 and various methods for NFV and TIME.

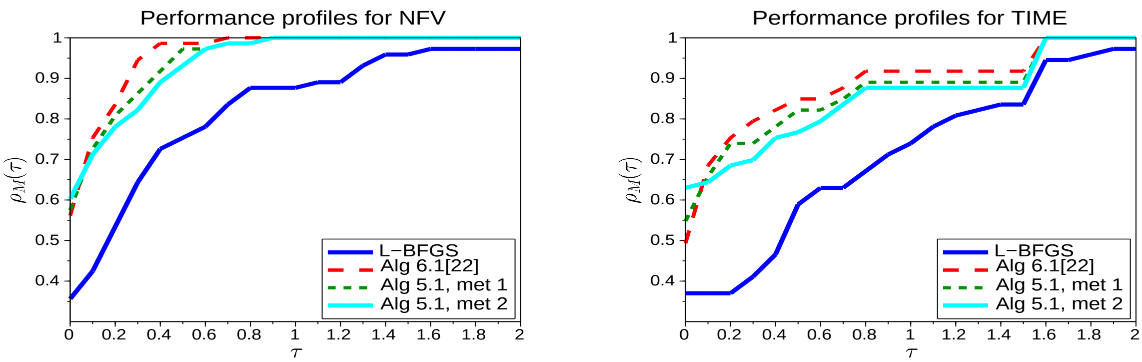

Figure 7.2: Comparison of $\rho_M(\tau)$ for Test 12 and various methods for NFV and TIME.

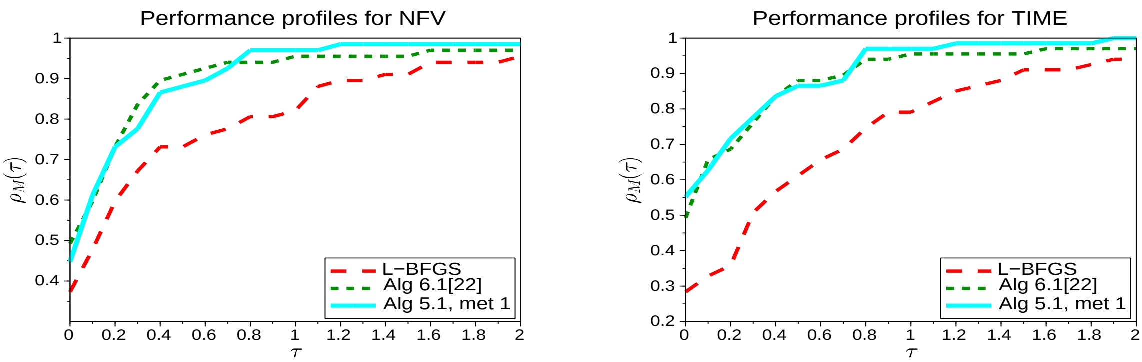

Figure 7.3: Comparison of $\rho_M(\tau)$ for Test 25 and various methods for NFV and TIME.

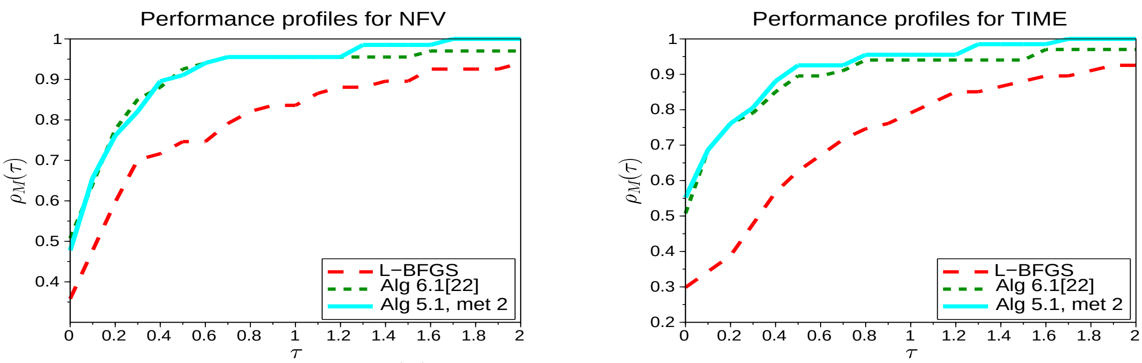

Figure 7.4: Comparison of $\rho_M(\tau)$ for Test 25 and various methods for NFV and TIME.

that method M can solve. The relative efficiency and reliability of each method can be directly seen from the performance profiles: the higher the particular curve, the better is the corresponding method. Figures 1–4, based on the results in Table 2, show graphical performance profiles for the tested methods (for Test 25, we compare our new methods separately to make the differences more visible). They demonstrate the efficiency of our methods in comparison with the L-BFGS method. We can also see that the numerical results for the new methods and for our method [22] are comparable.

8 Conclusions

In this contribution, we derive two new updates for general functions with minimum violation of the previous quasi-Newton equations in some sense, describe its properties and show how the corresponding methods can be advantageously combined with vector corrections for conjugacy.

Our experiments indicate that this approach can improve unconstrained large-scale minimization results significantly compared with the frequently used L-BFGS method.

References

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

[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] 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.

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

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

[6] M. Fiedler: Special matrices and their applications in numerical mathematics, 2nd ed. Mineola, N.Y., Dover Publications, 2008.

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

[8] 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.

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

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

[11] 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.

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

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

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

[15] L. Lukšan, M. Tůma, C. Matonoha, J. Vlček, N. Ramešová, M. Šiška, J. Hartman: *UFO 2017. Interactive System for Universal Functional Optimization*, Report V-1252, Prague, ICS AS CR, 2017.

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

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

[18] 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.

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

[20] 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.

[21] J. Vlček, L. Lukšan, *Properties of the block BFGS update and its application to the limited-memory block BNS method for unconstrained minimization*, Numerical Algorithms 80 (2019) 957-987.

[22] J. Vlček, L. Lukšan: *A limited-memory optimization method using the infinitely many times repeated BNS update and conjugate directions*, J. Comput. Appl. Math. 351 (2019) 14-28.

[23] G. Yuan, Z. Sheng, B. Wang, W. Hu, Ch. Li, *The global convergence of a modified BFGS method for nonconvex functions*, J. Comput. Appl. Math. 327 (2018) 274-294.