Institute of Computer Science
Academy of Sciences of the Czech Republic

Limited-memory projective variable metric methods for unconstrained minimization

J. Vlček, L. Lukšan

Pod Vodárenskou věží 2, 182 07 Prague 8 phone: +420 2 688 42 44, fax: +420 2 858 57 89, e-mail:luksan@cs.cas.cz, vlcek@cs.cas.cz

Institute of Computer Science
Academy of Sciences of the Czech Republic

Limited-memory projective variable metric methods for unconstrained minimization

J. Vlček, L. Lukšan[1]

Technical report No. V 1036

December 2008

Abstract:

A new family of limited-memory variable metric or quasi-Newton methods for unconstrained minimization is given. The methods are based on a positive definite inverse Hessian approximation in the form of the sum of identity matrix and two low rank matrices, obtained by the standard scaled Broyden class update. To reduce the rank of matrices, various projections are used. Numerical experience is encouraging.

Keywords:
Unconstrained minimization, variable metric methods, limited-memory methods, Broyden class updates, projection matrix, numerical results.

1   Introduction

In this report we present a new family of limited-memory variable metric (VM) line search methods for unconstrained minimization, which are projective, ie. they reduce the rank of matrices in the inverse Hessian approximation by suitable projection.

VM line search methods, see [5], [3], are iterative. Starting with an initial point $x_0 \in \mathcal{R}^N$, they 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 d_k is the direction vector and $t_k > 0$ is a stepsize.

We assume that the problem function $f : \mathcal{R}^N \to \mathcal{R}$ is differentiable, $d_k = -H_k g_k$ and stepsize t_k is chosen in such a way that

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

$k \geq 0$, where $0 < \varepsilon_1 < 1/2$, $\varepsilon_1 < \varepsilon_2 < 1$, $f_k = f(x_k)$, $g_k = \nabla f(x_k)$ and H_k is a symmetric positive definite matrix. We denote $B_k = H_k^{-1}$, $y_k = g_{k+1} - g_k$, $k \geq 0$ and by $\|.\|_F$ the Frobenius matrix norm.

Various possibilities how to construct positive definite inverse Hessian approximations based on the low rank matrices are discussed in Section 2. In Section 3 we show how properties of the standard scaled BFGS update can be utilized and generalized to construct new limited-memory methods. Numerical results are presented in Section 4.

2   Positive definite inverse Hessian approximation

A simple form of positive definite matrices H_k based on low rank matrices is $\zeta_k I + U_k U_k^T$, $k \geq 0$, where $\zeta_k > 0$ and U_k are $N \times m_k$ rectangular matrices, $1 \leq m_k \ll N$, but updating matrices of this form appears to be difficult. To avoid this drawback, shifted VM methods, see [7], were developed. They appeared to be very efficient, except for ill-conditioned problems, which is probably caused by the fact that shifted VM updates are not invariant under linear transformations (significance of the invariance property of methods for solution of ill-conditioned problems is discussed in [3]).

Another possibility, how to update matrices of this form, is described in [8]. Here invariant variationally-derived updates are applied, but VM matrices cannot be directly used to calculate direction vectors d_k, and thus complicated corrections are necessary.

Our new methods are based on matrices H_k of the form

$$H_k = \zeta_k I + U_k U_k^T - R_k R_k^T, \tag{2.1}$$

$k \geq 0$, where $\zeta_k > 0$ and U_k, R_k are $N \times \min(k,m)$ rectangular matrices, $1 \leq m \ll N$. Most of VM matrices in standard VM methods, see [3], [5], can be written in this way. These methods set usually $H_0 = I$ and H_{k+1} is obtained from $\gamma_k H_k$ (γ_k is a scaling parameter) by a rank-two VM update to satisfy the quasi-Newton condition; its generalized form is

$$H_{k+1} y_k = \varrho_k s_k, \tag{2.2}$$

where $\varrho_k > 0$ is a nonquadratic correction parameter (see [5]).

Therefore matrices H_k can be easily updated in starting iterations, i.e. for $k < m$. When $k \geq m$, the rank of matrices U, R needs to be reduced before each update and

updating formulas modified, which is described in Section 2.1. Some conditions to have the result positive definite are discussed in Section 2.2.

We use in this report matrices U_k, R_k with the same number of columns only for simplicity. Numerical experiments indicate that the number of columns of R_k can be substantially reduced without significant increasing of the number of function and gradient evaluations, which can increase efficiency. All methods here can be easily adapted in this way.

For given $q_k \in \mathcal{R}^N$, $q_k^T y_k \neq 0$, we denote by V_{q_k} the projection matrix $I - q_k y_k^T / q_k^T y_k$. To simplify the notation we frequently omit index k and replace index $k+1$ by symbol $+$. In the subsequent analysis we use the following notation

$$ C = \zeta I - RR^T, \quad b = s^T y, \quad V = V_s = I - (1/b)sy^T. $$

Note that $b > 0$ by (1.1).

2.1   Limited-memory VM matrices updating

To be able to add next columns, when matrices U, R have rank m, we first need to make their columns dependent, i.e. we need matrices $\underline{U}$, $\underline{R}$, near to U, R, and suitable vectors $z_1, z_2 \in \mathcal{R}^m$, satisfying $\underline{U} z_1 = 0$, $\underline{R} z_2 = 0$. Such matrix $\underline{U}$ can be in a general way written as the product of any $N \times m$ matrix and orthogonal projection matrix $P_1 = I - z_1 z_1^T / z_1^T z_1 = I - \bar{z}_1 \bar{z}_1^T$, $\bar{z}_1 = z_1 / |z_1|$, similarly for matrix $\underline{R}$ with $P_2 = I - \bar{z}_2 \bar{z}_2^T$, $\bar{z}_2 = z_2 / |z_2|$, instead of P_1.

The following lemma can be used to find the nearest (in the sense of Frobenius matrix norm) matrices to U, R.

Lemma 2.1. *Let T be symmetric positive definite, $q \in \mathcal{R}^N$, $z \in \mathcal{R}^m$, $z \neq 0$, $1 \leq m \leq N$, and denote $\mathcal{W}$ the set of $N \times m$ matrices, $W \in \mathcal{W}$. Then the unique solution to*

$$ \min\{\|T^{-1/2}(\underline{W} - W)\|_F : \underline{W} \in \mathcal{W}\} \quad \text{s.t.} \quad \underline{W} z = 0 \tag{2.3} $$

is

$$ \underline{W} = WP, \qquad P = I - zz^T / z^T z, \tag{2.4} $$

independently on the choice of T. Moreover, we can write

$$ \underline{W}\,\underline{W}^T + qq^T = \left(\underline{W} + qz^T/|z|\right)\left(\underline{W} + qz^T/|z|\right)^T. \tag{2.5} $$

Proof. Setting $W = (w_1, \ldots, w_m)$, $\underline{W} = (\underline{w}_1, \ldots, \underline{w}_m)$, $z = (\xi_1, \ldots, \xi_m)^T$, we define Lagrangian function $\mathcal{L} = \mathcal{L}(\underline{W}, e)$, $e \in \mathcal{R}^N$, as

$$ \mathcal{L} = \frac{1}{2}\|T^{-1/2}(\underline{W} - W)\|_F^2 + e^T \underline{W} z = \frac{1}{2}\sum_{i=1}^m (\underline{w}_i - w_i)^T T^{-1}(\underline{w}_i - w_i) + \sum_{i=1}^m \xi_i e^T \underline{w}_i. $$

A local minimizer $\underline{W}$ satisfies the equations $\partial \mathcal{L}/\partial \underline{w}_i = 0$, $i = 1, \ldots, m$, which gives $T^{-1}(\underline{w}_i - w_i) + \xi_i e = 0$, $i = 1, \ldots, m$, yielding $\underline{W} = W - Tez^T$. Using the condition $\underline{W} z = 0$, we have $Te = Wz/z^T z$, i.e. $\underline{W} = WP$. Matrix $\underline{W}$ satisfies (2.3) in view of convexity of Frobenius norm and the rest follows immediately from (2.4).   □

In view of Lemma 2.1 we set

$$\underline{U} = UP_1, \quad \underline{R} = RP_2, \quad \underline{C} = \zeta I - \underline{R}\,\underline{R}^T, \quad \underline{H} = \underline{C} + \underline{U}\,\underline{U}^T. \qquad (2.6)$$

If matrix $\underline{H}$ is positive definite and we choose a suitable standard VM update of $\underline{H}$ in the form $(1/\gamma)H_+ = \underline{H} + uu^T - rr^T$ which maintain positive definiteness of VM matrices, limited-memory updating of H can be completely covered by Lemma 2.1. Writing $H_+ = \zeta_+ I - R_+ R_+^T + U_+ U_+^T$ and using (2.5), we can set $\zeta_+ = \gamma\zeta$, $(1/\sqrt{\gamma})U_+ = \underline{U}P_1 + u\bar{z}_1^T$, $(1/\sqrt{\gamma})R_+ = \underline{R}P_2 + r\bar{z}_2^T$.

Unfortunately, in general case we cannot find vectors z_1, z_2 which guarantee positive definiteness of $\underline{H}$. Some conditions that ensure it will be given in the next section.

2.2 Conditions for positive definiteness of reduced VM matrices

In view of $P_i^2 = P_i$, $i = 1, 2$, we can write

$$\underline{U}\,\underline{U}^T = UP_1 U^T = UU^T - U\bar{z}_1\bar{z}_1 U^T, \quad \underline{R}\,\underline{R}^T = RP_2 R^T = RR^T - R\bar{z}_2\bar{z}_2 R^T. \qquad (2.7)$$

In view of $\underline{C} = C + R\bar{z}_2\bar{z}_2 R^T$ by (2.6), we see that situation is quite simple, when matrix C is positive definite, since then also matrix $\underline{C}$ and $\underline{H} = \underline{C} + \underline{U}\,\underline{U}^T$ are positive definite for any z_1, z_2. The next section will be devoted to such standard VM updates that can be adapted to maintain this property.

Even if matrix C is not positive definite, the choice of z_2 has no influence to preserving of positive definiteness of H after reduction of matrix R, since $\zeta I - \underline{R}\,\underline{R}^T + UU^T = H + R\bar{z}_2\bar{z}_2 R^T$ by (2.7). The following lemma shows one possibility how to attain positive definiteness of the VM matrix after reduction of matrix U.

Lemma 2.2. *Let $A = \underline{H} + vv^T$, $v \in \mathcal{R}^N$, and denote $\underline{\lambda}(A)$, $\underline{\lambda}(\underline{H})$ the minimum eigenvalues of A, $\underline{H}$. If $[\underline{U}, \underline{R}]^T v = 0$ then $\underline{\lambda}(\underline{H}) = \underline{\lambda}(A)$. Conditions $\underline{U}^T v = 0$, $\underline{R}^T v = 0$ can be satisfied e.g. by the choice $z_1 = U^T v$, $z_2 = R^T v$.*

Proof. We first show that $\underline{\lambda}(A) \leq \zeta$. Denoting $K_1 = \{q \in \mathcal{R}^N \colon [\underline{U}, v]^T q = 0, |q| = 1\}$, we have

$$\underline{\lambda}(A) = \min_{|q|=1} q^T A q \leq \min_{q \in K_1} q^T A q = \min_{q \in K_1} q^T(\zeta I - \underline{R}\,\underline{R}^T)q \leq \zeta.$$

We will consider the following two cases:

(a) If $\underline{\lambda}(\underline{H}) = \zeta$, then $\underline{\lambda}(\underline{H}) = \underline{\lambda}(A)$ holds, since $\zeta = \underline{\lambda}(\underline{H}) \leq \underline{\lambda}(A) \leq \zeta$.

(b) If $\underline{\lambda}(\underline{H}) \neq \zeta$, then $\underline{H}\,w = \underline{\lambda}(\underline{H})\,w$, $w \neq 0$, implies $\underline{U}\,\underline{U}^T w - \underline{R}\,\underline{R}^T w = (\underline{\lambda}(\underline{H}) - \zeta)w$ and thus $w \in K_2 \overset{\Delta}{=} \{q \in \text{range}([\underline{U}, \underline{R}]) \colon q \neq 0\}$. By $[\underline{U}, \underline{R}]^T v = 0$ we get

$$\underline{\lambda}(\underline{H}) \leq \underline{\lambda}(A) = \min_{q \neq 0} \frac{q^T A q}{q^T q} \leq \min_{q \in K_2} \frac{q^T A q}{q^T q} = \min_{q \in K_2} \frac{q^T \underline{H}\,q}{q^T q} \leq \frac{w^T \underline{H}\,w}{w^T w} = \underline{\lambda}(\underline{H}),$$

therefore we again obtain $\underline{\lambda}(\underline{H}) = \underline{\lambda}(A)$.

If $z_1 = U^T v$, we have $\underline{U}^T v = P_1 U^T v = P_1 z_1 = 0$ and similarly $\underline{R}^T v = 0$ for $z_2 = R^T v$. $\qquad\square$

3

Note that in our case $v = U\bar{z}_1$ by (2.7), thus the choice $z_1 = U^T v$ leads to the eigenvalue problem $U^T U z_1 = |z_1|z_1$; in view of (2.7) the solution of this problem minimizes matrix UU^T damage, measured by $\|U\bar{z}_1\bar{z}_1 U^T\| = |U\bar{z}_1|^2 = z_1 U^T U z_1/z_1^T z_1$, if we choose as z_1 the eigenvector corresponding to the smallest eigenvalue of $U^T U$. This is valuable also in situations when we attain positive definiteness of the VM matrix after reduction of matrix U in another way.

3 New limited-memory methods

In this section we focus on adaptation of standard VM updates to maintain positive definiteness of matrix C. Having reduced matrix $\underline{C}$ positive definite, we want to find updated matrix H_+ satisfying the quasi-Newton condition (2.2), in the form $H_+ = C_+ + U_+ U_+^T$, where C_+ is positive definite. An easy way to do it is shown in Section 3.1 in case of the scaled BFGS method, see [5]. The variant of this approach based on the partly inverse representation of matrix C is described in Section 3.2. These results are generalized in Section 3.3, using variational approach, and in Section 3.4, using a special transformation. The choice of parameters, including z_1 and z_2, is discussed in Section 3.5.

3.1 Adaptation of the scaled BFGS method

The standard scaled BFGS update of $\underline{H}$ can be written in the form (see [5])

$$\frac{1}{\gamma}\underline{H}_+^{BFGS} = \frac{\varrho}{\gamma b}ss^T + V\underline{H}V^T, \tag{3.1}$$

which in view of $\underline{H} = \underline{C} + \underline{U}\,\underline{U}^T$ after easy rearrangements leads to

$$\frac{1}{\gamma}\underline{H}_+^{BFGS} = \frac{1}{\gamma}\underline{C}_+^{BFGS} + V\underline{U}\,\underline{U}^T V^T = \underline{C} - rr^T + uu^T + V\underline{U}\,\underline{U}^T V^T, \tag{3.2}$$

where

$$u = (1/\sqrt{\omega b})\,(\omega s - \underline{C}\,y), \quad r = (1/\sqrt{\omega b})\,\underline{C}\,y, \quad \omega = \varrho/\gamma + \tilde{a}/b, \quad \tilde{a} = y^T\underline{C}\,y. \tag{3.3}$$

If we define $(1/\gamma)C_+ = \underline{C} - rr^T$, we can easily show that matrix C_+ is positive definite.

Lemma 3.1. *Let matrix $\underline{C}$ be positive definite and vector r be given by (3.3). Then matrix $\underline{C} - rr^T$ is also positive definite.*

Proof. For any $q \in \mathcal{R}^N$ we obtain by the Schwarz inequality, (3.3) and $b > 0$

$$\frac{q^T(\underline{C} - rr^T)q}{q^T\underline{C}\,q^T} = 1 - \frac{(q^T\underline{C}\,y)^2}{(\omega b)\,q^T\underline{C}\,q^T} \geq 1 - \frac{\tilde{a}}{\omega b} = 1 - \frac{\tilde{a}}{b\varrho/\gamma + \tilde{a}} = \frac{b\varrho/\gamma}{b\varrho/\gamma + \tilde{a}} > 0. \qquad \square$$

Replacing $\underline{C} - rr^T$ in (3.2) by $(1/\gamma)C_+$, we see that to get

$$H_+ = C_+ + U_+ U_+^T, \tag{3.4}$$

4

we need matrix U_+ satisfying $(1/\gamma)U_+U_+^T = V\underline{U}\,\underline{U}^T V^T + uu^T$. Similarly, considering that $\underline{C} = \zeta I - \underline{R}\,\underline{R}^T$, we see that to get

$$C_+ = \zeta_+ I - R_+ R_+^T, \tag{3.5}$$

we need $\zeta_+ > 0$ and matrix R_+ satisfying $(1/\gamma)R_+R_+^T = \underline{R}\,\underline{R}^T + rr^T$. By (2.6) we can fulfil all these requirements if we define

$$\zeta_+ = \gamma\zeta, \quad (1/\sqrt{\gamma})U_+ = V\underline{U} + u\bar{z}_1^T, \quad (1/\sqrt{\gamma})R_+ = \underline{R} + r\bar{z}_2^T. \tag{3.6}$$

For better clearness, we summarize the updating process. First we choose vectors z_1, z_2, see recommendations in Section 3.5, and then we determine matrices $\underline{U}$, $\underline{R}$, $\underline{C}$, $\underline{H}$ by (2.6) and compute vectors u, r by (3.3). Finally we compute number ζ_+ and matrices U_+, R_+ by (3.6) and determine matrices C_+ by (3.5) and H_+ by (3.4).

In starting iterations we omit reductions of matrices U, R, i.e. $\underline{C} = C$, $\underline{U} = U$, $\underline{R} = R$ and simply add column u to VU and r to R. Note that then H_+ is the BFGS update of H, unlike C_+, which is <u>not</u> any standard VM update of matrix C.

3.2   Partly inverse representation of the scaled BFGS method

In the previous method we can change ζ only by scaling (3.6) to guarantee that C_+ will be positive definite. This drawback can be overcome when we use partly inverse representation of the scaled BFGS update.

In starting iterations we can in (3.2) rewrite the BFGS update of matrix $\underline{C} = C$ in the form

$$\frac{1}{\gamma}\underline{C}_+^{BFGS} = \left(\underline{C}^{-1} + \frac{\gamma}{\varrho b}yy^T\right)^{-1} + uu^T \tag{3.7}$$

with u given again by (3.2), which can be readily verified. Thus instead of $(1/\gamma)C_+ = \underline{C} - rr^T$, we can use equivalent formula

$$C_+^{-1} = \frac{1}{\gamma}\underline{C}^{-1} + \frac{1}{\varrho b}yy^T. \tag{3.8}$$

Starting with $C_0 = I$ and updating $C_0^{-1}, \ldots, C_{k-1}^{-1}$ according to (3.8), we obtain

$$\underline{C}^{-1} = C^{-1} = (1/\bar{\gamma})\,I + YDY^T, \tag{3.9}$$

where $\bar{\gamma} = \gamma_1 \cdots \gamma_k$, $Y = [y_1, \ldots, y_k]$ and diagonal positive definite matrix D is given by $D_1 = [1/(\varrho_1 b_1)]$ and by update formula

$$D_+ = \mathrm{diag}\left(\frac{1}{\gamma}D, \frac{1}{\varrho b}\right). \tag{3.10}$$

In view of (3.9), we store matrices Y, D instead of R.

When $k \geq m$, we reduce matrix U in the same way as before, i.e. $\underline{U} = UP_1$, while in case of matrices Y, D, we can simply delete their first columns, i.e. supposing that $Y = [y_{k-m}, \ldots, y_{k-1}]$, $D = \mathrm{diag}(d_1, \ldots, d_m)$ and $C^{-1} = (1/\zeta)\,I + YDY^T$, $\zeta > 0$, we set

$$\underline{Y} = [y_{k-m+1}, \ldots, y_{k-1}], \quad \underline{D} = \mathrm{diag}(d_2, \ldots, d_m), \quad \underline{C}^{-1} = (1/\zeta)\,I + \underline{Y}\,\underline{D}\,\underline{Y}^T. \tag{3.11}$$

To get $H_+ = C_+ + U_+ U_+^T$, we define ζ_+, U_+ again by (3.6) and C_+ in view of (3.8) by

$$C_+^{-1} = \frac{1}{\zeta_+} I + Y_+ D_+ Y_+^T, \quad Y_+ = [\underline{Y}, y], \quad D_+ = \mathrm{diag}\left(\frac{1}{\gamma} \underline{D}, \frac{1}{\varrho b}\right). \tag{3.12}$$

To be able to compute $\underline{C} y$, we store $m \times m$ matrix $\underline{Y}^T \underline{Y}$ and use the Woodbury formula, which yields

$$\underline{C} = \left(\frac{1}{\zeta} I + \underline{Y} \, \underline{D} \, \underline{Y}^T\right)^{-1} = \zeta I - \zeta \underline{Y} \left(\frac{1}{\zeta} \underline{D}^{-1} + \underline{Y}^T \underline{Y}\right)^{-1} \underline{Y}^T, \tag{3.13}$$

i.e. to compute $\underline{C} y$, we need to solve a linear system with symmetric positive definite matrix $(1/\zeta) \underline{D}^{-1} + \underline{Y}^T \underline{Y}$. Note that during computation of $\underline{C} y$ we can save vector $\underline{Y}^T y$ and use it in updating of matrix $Y^T Y$, since by (3.12) we obtain

$$Y_+^T Y_+ = \begin{pmatrix} \underline{Y}^T \underline{Y} & \underline{Y}^T y \\ y^T \underline{Y} & y^T y \end{pmatrix}. \tag{3.14}$$

Now we summarize the updating process. First we choose vector z_1, see recommendations in Section 3.5, and then we determine matrices $\underline{U}$ by (2.6), $\underline{Y}$, $\underline{D}$, $\underline{C}$ by (3.11) and $\underline{H}$ by (2.6) and compute vector u by (3.3), using (3.13). Finally we compute number ζ_+ and matrix U_+ by (3.6) and determine matrices Y_+, D_+, C_+ by (3.12), using (3.14), and H_+ by (3.4).

If we wish to use a different projection instead of deleting the first column of Y and D (although we are not able to say in this time, how to choose a suitable projection), we can use another representation of $\underline{C}$. In starting iterations we rewrite (3.9) in the form

$$\underline{C}^{-1} = C^{-1} = (1/\bar{\gamma}) I + \tilde{Y} \tilde{Y}^T, \quad \tilde{Y} = Y D^{1/2}. \tag{3.15}$$

Then for $k \geq m$ we reduce matrix U as before, i.e. $\underline{U} = U P_1$, and matrix $\tilde{Y}$ in a similar way, i.e. $\underline{\tilde{Y}} = \tilde{Y} P_3$, $P_3 = I - \bar{z}_3 \bar{z}_3^T$, $\bar{z}_3 = z_3/|z_3|$, with some $z_3 \in \mathcal{R}^m$, and set

$$\underline{C}^{-1} = (1/\zeta) I + \underline{\tilde{Y}} \, \underline{\tilde{Y}}^T, \quad \zeta > 0. \tag{3.16}$$

To get $H_+ = C_+ + U_+ U_+^T$, we define ζ_+, U_+ again by (3.6) and C_+ in view of (3.8) by

$$C_+^{-1} = \frac{1}{\zeta_+} I + \tilde{Y}_+ \tilde{Y}_+^T, \quad \tilde{Y}_+ = \frac{1}{\sqrt{\gamma}} \underline{\tilde{Y}} + \frac{1}{\sqrt{\varrho b}} y \bar{z}_3^T. \tag{3.17}$$

To be able to compute $\underline{C} y$, we store $m \times m$ matrix $\underline{\tilde{Y}}^T \underline{\tilde{Y}}$ and use the Woodbury formula, which yields

$$\underline{C} = \left(\frac{1}{\zeta} I + \underline{\tilde{Y}} \, \underline{\tilde{Y}}^T\right)^{-1} = \zeta I - \zeta \underline{\tilde{Y}} \left(\frac{1}{\zeta} I + \underline{\tilde{Y}}^T \underline{\tilde{Y}}\right)^{-1} \underline{\tilde{Y}}^T.$$

Note that during computation of $\underline{C} y$ we can again save vector $\underline{\tilde{Y}}^T y$ and use it in updating of matrix $\tilde{Y}^T \tilde{Y}$, since by (3.17) we obtain

$$\tilde{Y}_+^T \tilde{Y}_+ = \frac{1}{\gamma} \underline{\tilde{Y}}^T \underline{\tilde{Y}} + \frac{1}{\sqrt{\varrho \gamma b}} \left(\underline{\tilde{Y}}^T y \bar{z}_3^T + \bar{z}_3 y^T \underline{\tilde{Y}}\right) + \frac{y^T y}{\varrho \, b} \bar{z}_3 \bar{z}_3^T.$$

6

3.3 Variationally-derived generalization

Although the scaled Broyden class updates of $\underline{H}$ with positive value of parameter η can be written in the quasi-product form similar to (3.1), see [7], we cannot construct limited-memory methods in a similar way as in the previous sections, since then projection matrix contains vector $\underline{H}\,y \neq \underline{C}\,y$ and therefore we do not obtain the relation similar to (3.2). In spite of that, it is possible to generalize this process for the standard scaled Broyden class updates, as we show in the next section.

In this section we briefly describe two other possibilities. The first approach is based on the projection variant, see [8], of the well-known Greenstadt's theorem, see [4]:

Theorem 3.1. *Let M, T be symmetric matrices, T positive definite, $\varrho > 0$, $p = Ty$ and denote $\mathcal{M}$ the set of $N \times N$ symmetric matrices. Then the unique solution to*

$$\min\{\|T^{-1/2}(M_+ - M)T^{-1/2}\|_F : M_+ \in \mathcal{M}\} \quad \text{s.t.} \quad M_+ y = \varrho s$$

is determined by the relation $V_p\,(M_+ - M)\,V_p^T = 0$ and can be written in the form

$$M_+ = E + V_p\,(M - E)\,V_p^T,$$

where E is any symmetric matrix satisfying $Ey = \varrho s$, e.g. $E = (\varrho/b)ss^T$.

Using this theorem with $M = \gamma\underline{H} = \gamma(\underline{C} + \underline{U}\,\underline{U}^T)$, $M_+ = H_+$, $p = s - \alpha\,\underline{C}\,y$, $\alpha \in \mathcal{R}$, and $E = (\varrho/b)ss^T$, we obtain

$$\frac{1}{\gamma}H_+ = V_p\,\underline{U}\,\underline{U}^T V_p^T + \frac{1}{\gamma}\underline{C}\,_+^{BC}, \quad \frac{1}{\gamma}\underline{C}\,_+^{BC} = \frac{\varrho}{\gamma}\frac{ss^T}{b} + V_p\left(\underline{C} - \frac{\varrho}{\gamma}\frac{ss^T}{b}\right)V_p^T, \qquad (3.18)$$

where by Lemma 2.3 in [8], this $\underline{C}\,_+^{BC}$ is the Broyden class update, see [5], of $\underline{C}$ with parameter

$$\eta_C = \frac{b^2}{(b - \alpha\tilde{a})^2}\left(1 - \alpha^2\frac{\varrho}{\gamma}\frac{\tilde{a}}{b}\right). \qquad (3.19)$$

For given η_C satisfying

$$0 < \eta_C < \tilde{\omega}\frac{\gamma}{\varrho}, \quad \tilde{\omega} \triangleq \frac{\varrho}{\gamma} + \frac{\tilde{a}}{b}\eta_C, \qquad (3.20)$$

we can readily verify that the corresponding α can be obtained from

$$\alpha_1 = \frac{\eta_C + \sqrt{\eta_C + (1-\eta_C)\varrho b/(\gamma\tilde{a})}}{\tilde{\omega}} \quad \text{or} \quad \alpha_2 = \frac{(b/\tilde{a})(\eta_C - 1)}{\eta_C + \sqrt{\eta_C + (1-\eta_C)\varrho b/(\gamma\tilde{a})}}. \qquad (3.21)$$

Using the usual form of the Broyden class update of $\underline{C}$ in (3.18), after straightforward arrangements we obtain the relation, similar to (3.2)

$$\frac{1}{\gamma}H_+ = \frac{1}{\gamma}\underline{C}\,_+^{BC} + V_p\,\underline{U}\,\underline{U}^T V_p^T = \underline{C} - \tilde{r}\tilde{r}^T + \tilde{u}\tilde{u}^T + V_p\,\underline{U}\,\underline{U}^T V_p^T, \qquad (3.22)$$

where

$$\tilde{u} = \sqrt{\frac{\tilde{\omega}}{b}} \left(s - \frac{\eta_C}{\tilde{\omega}} \underline{C}\, y \right), \quad \tilde{r} = \sqrt{1 - \frac{\varrho}{\gamma}\frac{\eta_C}{\tilde{\omega}}} \frac{\underline{C}\, y}{\sqrt{\tilde{a}}}. \tag{3.23}$$

Note that H^+ is not here the Broyden class update of $\underline{H}$. Similarly as in Section 3.1 we define $(1/\gamma)C_+ = \underline{C} - \tilde{r}\tilde{r}^T = \zeta I - \underline{R}\,\underline{R}^T - \tilde{r}\tilde{r}^T$. The following lemma gives simple conditions for matrix C_+ to be positive definite.

Lemma 3.2. *Let matrix $\underline{C}$ be positive definite, η_C satisfy (3.20) and vector $\tilde{r}$ be given by (3.23). Then matrix $\underline{C} - \tilde{r}\tilde{r}^T$ is also positive definite.*

Proof. For any $q \in \mathcal{R}^N$ we obtain by the Schwarz inequality and $b > 0$

$$\frac{q^T(\underline{C} - rr^T)\, q}{q^T \underline{C}\, q^T} = 1 - \left(1 - \frac{\varrho}{\gamma}\frac{\eta_C}{\tilde{\omega}}\right)\frac{(q^T \underline{C}\, y)^2}{\tilde{a}\, q^T \underline{C}\, q^T} \geq 1 - \left(1 - \frac{\varrho}{\gamma}\frac{\eta_C}{\tilde{\omega}}\right) = \frac{\varrho}{\gamma}\frac{\eta_C}{\tilde{\omega}} > 0. \qquad \square$$

To get $H_+ = C_+ + U_+ U_+^T$, we can similarly as in (3.6) for $k \geq m$ set

$$\zeta_+ = \gamma\zeta, \quad (1/\sqrt{\gamma})U_+ = V_p \underline{U} + \tilde{u}\bar{z}_1^T, \quad (1/\sqrt{\gamma})R_+ = \underline{R} + \tilde{r}\bar{z}_2^T, \tag{3.24}$$

for $k < m$ simply add column $\tilde{u}$ to $V_p U$ and $\tilde{r}$ to R.

The second approach utilizes idea of shifting, similarly as shifted VM methods, see [7]. We again update $\underline{C}$, using the scaled standard Broyden class update (3.18) in the usual form $(1/\gamma)\underline{C}_+^{BC} = \underline{C} - \tilde{r}\tilde{r}^T + \tilde{u}\tilde{u}^T$ with $\tilde{r}$, $\tilde{u}$ given by (3.23), set $(1/\gamma)C_+ = \underline{C} - \tilde{r}\tilde{r}^T$ and update ζ and $\underline{R}$ according to (3.24).

To update matrix U, we use shifted quasi-Newton condition

$$U_+ U_+^T y = \varrho\tilde{s}, \tag{3.25}$$

where in view of $H_+ = C_+ + U_+ U_+^T$, (2.2) and (3.23) we define

$$\tilde{s} = \frac{1}{\varrho}\left(H_+ y - C_+ y\right) = s - \frac{\gamma}{\varrho}\left(\underline{C}\, y - \tilde{r}^T y\,\tilde{r}\right) = s - \frac{\gamma}{\varrho}\underline{C}\, y\left[1 - \left(1 - \frac{\varrho}{\gamma}\frac{\eta_C}{\tilde{\omega}}\right)\right] = s - \frac{\eta_C}{\tilde{\omega}}\underline{C}\, y. \tag{3.26}$$

Important property of shifting defined in this way is that number $\tilde{b} \triangleq \tilde{s}^T y$ is always positive, since by (3.26)

$$\tilde{b} = \left(s - \frac{\eta_C}{\tilde{\omega}}\underline{C}\, y\right)^T y = b - \frac{\eta_C}{\tilde{\omega}}\tilde{a} = \frac{b\tilde{\omega} - \tilde{a}\,\eta_C}{\tilde{\omega}} = \frac{(\varrho/\gamma)\, b}{\tilde{\omega}} > 0. \tag{3.27}$$

Now we define matrix U_+ by the following theorem, see [7].

Theorem 3.2. *Let $\tilde{T}$ be a symmetric positive definite matrix, $\varrho > 0$, $\gamma > 0$, $\tilde{p} = \tilde{T}y$ and denote $\mathcal{U}$ the set of $N \times m$, $1 \leq m \leq N$, matrices. Then the unique solution to*

$$\min\{\varphi(U_+) : U_+ \in \mathcal{U}\} \text{ s.t. } (3.25), \quad \varphi(U_+) = y^T \tilde{T} y\, \|\tilde{T}^{-1/2}(U_+ - \sqrt{\gamma}\, U)\|_F^2,$$

is

$$(1/\sqrt{\gamma})U_+ = V_{\tilde{p}} U P + (1/\tilde{b})\tilde{s}z^T, \quad z = (1/\sqrt{\gamma})U_+^T y, \quad P = I - zz^T/z^T z.$$

8

It follows from (3.25) and (3.27) that vector z defined by this theorem satisfy

$$z^T z = (1/\gamma)y^T U_+ U_+^T y = (\varrho/\gamma)\tilde{s}^T y = \tilde{b}^2(\varrho/\gamma)/\tilde{b} = \tilde{b}^2\tilde{\omega}/b,$$

which yields by (3.26) and (3.23) $(\sqrt{z^T z}/\tilde{b})\tilde{s} = \sqrt{\tilde{\omega}/b}\,\tilde{s} = \tilde{u}$. Setting $z_1 = z$, Theorem 3.2 gives for U_+ the updating formula

$$(1/\sqrt{\gamma})U_+ = V_{\tilde{p}}\underline{U} + (\sqrt{z_1^T z_1}/\tilde{b})\tilde{s}\bar{z}_1^T = V_{\tilde{p}}\underline{U} + \tilde{u}\bar{z}_1^T, \tag{3.28}$$

which is the corresponding formula in (3.24) with $V_{\tilde{p}}$ instead of V_p, i.e. we see that the projection matrix in this update formula can be derived not only from update of matrix $\underline{C}$, but also in another way, e.g. from update of $\underline{H}$.

Summarizing the updating process, we see that it is very similar to the process in Section 3.1. First we choose vectors z_1, z_2, see recommendations in Section 3.5, and then we determine matrices $\underline{U}$, $\underline{R}$, $\underline{C}$, $\underline{H}$ by (2.6) and compute vectors $\tilde{u}$, $\tilde{r}$ by (3.23). Finally we compute number ζ_+ and matrices U_+ by (3.24) or (3.28) and R_+ by (3.24) and determine matrices C_+ by (3.5) and H_+ by (3.4).

Previous methods can also be modified to use the partly inverse representation of updates. We can readily verify the following generalization of (3.7) for the scaled standard Broyden class update with parameter $\eta_C \neq 0$

$$\frac{1}{\gamma}\underline{C}_+^{BC} = \left[\underline{C}^{-1} + \left(\frac{1-\eta_C}{\eta_C} + \frac{\tilde{a}}{b}\frac{\gamma}{\varrho}\right)\frac{yy^T}{\tilde{a}}\right]^{-1} + \tilde{u}\tilde{u}^T, \tag{3.29}$$

which can be used instead of $(1/\gamma)\underline{C}_+^{BC} = \underline{C} - \tilde{r}\tilde{r}^T + \tilde{u}\tilde{u}^T$ in (3.22).

3.4   Generalization to the scaled Broyden class using transformation

Methods based on adaptation of the scaled BFGS method can be generalized to the standard Broyden class updates with parameter η if we use the following special representation of these updates. We denote

$$\underline{a} = y^T \underline{H} y, \quad \underline{\omega} = \frac{\varrho}{\gamma} + \frac{a}{b}\eta, \quad \mu = \eta + (1-\eta)\frac{\varrho}{\gamma}\frac{b}{\underline{a}}.$$

Theorem 3.3. *Let $\varrho > 0$, $\gamma > 0$, $\underline{a}\,\underline{\omega} \neq 0$, $\mu \geq 0$ and $\beta = (\eta \pm \sqrt{\mu})/\underline{\omega}$. Then the scaled standard Broyden class update of $\underline{H}$ with parameter η, scaling parameter γ and nonquadratic correction parameter ϱ can be expressed in the form*

$$\frac{1}{\gamma}\underline{H}_+^{BC} = \frac{\varrho}{\gamma}\eta\frac{\hat{s}\hat{s}^T}{b} + \hat{V}\underline{H}\hat{V}^T, \quad \hat{s} = s - \beta\underline{H}y, \quad \hat{V} = I \pm \frac{\sqrt{\mu}}{b}\hat{s}y^T. \tag{3.30}$$

Moreover, if $\hat{\omega} \overset{\Delta}{=} (\varrho/\gamma)\eta + (\tilde{a}/b)\mu > 0$, we can write

$$\frac{1}{\gamma}\underline{H}_+^{BC} = \underline{C} - \hat{r}\hat{r}^T + \hat{u}\hat{u}^T + \hat{V}\underline{U}\,\underline{U}^T\hat{V}^T, \quad \hat{r} = \sqrt{\frac{\mu}{\hat{\omega}b}}\underline{C}y, \quad \hat{u} = \sqrt{\frac{\hat{\omega}}{b}}\hat{s} \pm \hat{r}. \tag{3.31}$$

9

Proof. Consider the scaled Broyden class update with parameters η, γ and ϱ in the form, see [5],

$$\frac{1}{\gamma}\underline{H}_+^{BC} = \underline{H} + \frac{\omega}{b}ss^T - \frac{\eta}{b}\left(\underline{H}\,ys^T + sy^T\underline{H}\right) + \frac{\eta-1}{\underline{a}}\underline{H}\,yy^T\underline{H}\,.$$

Setting $s = \hat{s} + \xi\underline{H}\,y$, $\xi \in \mathcal{R}$, we obtain

$$\frac{1}{\gamma}\underline{H}_+^{BC} = \underline{H} + \frac{\omega}{b}\hat{s}\hat{s}^T + \frac{\xi\omega-\eta}{b}\left(\underline{H}\,y\hat{s}^T + \hat{s}y^T\underline{H}\right) + \left(\frac{\eta-1}{\underline{a}} + \frac{\xi^2\omega-2\xi\eta}{b}\right)\underline{H}\,yy^T\underline{H}.$$

The last term vanishes for $\xi^2\omega - 2\xi\eta + (b/\underline{a})(\eta-1) = 0$, i.e. for $\xi = (\eta \pm \sqrt{\mu})/\omega = \beta$; then $\xi\omega - \eta = \pm\sqrt{\mu}$ and

$$\frac{1}{\gamma}\underline{H}_+^{BC} = \underline{H} + \frac{\omega}{b}\hat{s}\hat{s}^T + \frac{\pm\sqrt{\mu}}{b}\left(\underline{H}\,y\hat{s}^T + \hat{s}y^T\underline{H}\right) = \hat{V}\underline{H}\hat{V}^T + \left(\omega - \frac{a}{b}\mu\right)\frac{\hat{s}\hat{s}^T}{b}\,,$$

which is (3.30) in view of $b\omega - \underline{a}\,\mu = (\varrho/\gamma)b[1 - (1-\eta)] + \underline{a}\eta - \underline{a}\eta = (\varrho/\gamma)\eta b$.

Setting $\underline{H} = \underline{C} + \underline{U}\,\underline{U}^T$ into (3.30), we get

$$
\begin{aligned}
\frac{1}{\gamma}\underline{H}_+^{BC} &= \frac{\varrho}{\gamma}\eta\frac{\hat{s}\hat{s}^T}{b} + \hat{V}\underline{U}\,\underline{U}^T\hat{V}^T + \left(I \pm \frac{\sqrt{\mu}}{b}\hat{s}y^T\right)\underline{C}\left(I \pm \frac{\sqrt{\mu}}{b}y\hat{s}^T\right) \\
&= \hat{V}\underline{U}\,\underline{U}^T\hat{V}^T + \underline{C} + \frac{\pm\sqrt{\mu}}{b}\left(\underline{C}\,y\hat{s}^T + \hat{s}y^T\underline{C}\right) + \left(\frac{\varrho}{\gamma}\eta + \frac{\tilde{a}}{b}\mu\right)\frac{\hat{s}\hat{s}^T}{b} \\
&= \hat{V}\underline{U}\,\underline{U}^T\hat{V}^T + \underline{C} + \frac{\hat{\omega}}{b}\left(\hat{s} \pm \frac{\sqrt{\mu}}{\hat{\omega}}\underline{C}\,y\right)\left(\hat{s} \pm \frac{\sqrt{\mu}}{\hat{\omega}}\underline{C}\,y\right)^T - \frac{\mu}{\hat{\omega}b}\underline{C}\,yy^T\underline{C}\,,
\end{aligned}
$$

which is (3.31). $\qquad\qquad\square$

Note that $\hat{V}$ is not projection matrix in general and that we prefer the minus sign in β, $\hat{V}$ and $\hat{u}$, since then for $\eta = 1$ (BFGS) we get $\beta = 0$, $\hat{s} = s$ and $\hat{V} = V$. For $\eta \approx 1$ it is also $\sqrt{\mu} \approx 1$, therefore the formula for β above should be rewritten in another form. In view of

$$\eta^2 - \mu = \eta^2 - \eta - (1-\eta)\varrho b/(\gamma\underline{a}) = (\eta-1)(\eta + \varrho b/(\gamma\underline{a})) = (\eta-1)\omega b/\underline{a}$$

we obtain
$$\beta = \frac{\eta - \sqrt{\mu}}{\omega} = \frac{\eta^2 - \mu}{\omega(\eta + \sqrt{\mu})} = \frac{(\eta-1)\omega b/\underline{a}}{\omega(\eta + \sqrt{\mu})} = \frac{(\eta-1)b/\underline{a}}{\eta + \sqrt{\mu}}\,.$$

For better understanding, condition $\mu \geq 0$ can be rewritten as $\eta(\varrho b - \gamma\underline{a}) \leq \varrho b$, i.e. $\eta \leq \eta_{SR1}$ for $\eta_{SR1} > 0$, or $\eta \geq \eta_{SR1}$ for $\eta_{SR1} < 0$, where η_{SR1} is the value of parameter η for the SR1 method, $\eta_{SR1} = \varrho b/(\varrho b - \gamma\underline{a})$, see [5].

Now we define $(1/\gamma)C_+ = \underline{C} - \hat{r}\hat{r}^T$ and give conditions for matrix C_+ to be positive definite.

Lemma 3.3. *Let the assumptions of Theorem 3.3 be satisfied, matrix $\underline{C}$ be positive definite, vector $\hat{r}$ be given by (3.31) and $\eta > 0$. Then $\underline{C} - \hat{r}\hat{r}^T$ is also positive definite.*

Proof. For any $q \in \mathcal{R}^N$ we obtain by the Schwarz inequality and $b > 0$

$$\frac{q^T(\underline{C} - \hat{r}\hat{r}^T)q}{q^T\underline{C}\,q^T} = 1 - \frac{\mu(q^T\underline{C}\,y)^2}{(\hat{\omega}b)\,q^T\underline{C}\,q^T} \geq 1 - \frac{\mu\tilde{a}}{\hat{\omega}b} = 1 - \frac{\mu\tilde{a}}{\eta b\varrho/\gamma + \mu\tilde{a}} = \frac{\eta b\varrho/\gamma}{\eta b\varrho/\gamma + \mu\tilde{a}} > 0.$$

$\square$

To get $H_+ = C_+ + U_+U_+^T$, we can set for $k \geq m$ similarly as in (3.6)

$$\zeta_+ = \gamma\zeta, \quad (1/\sqrt{\gamma})U_+ = \hat{V}\underline{U} + \hat{u}\bar{z}_1^T, \quad (1/\sqrt{\gamma})R_+ = \underline{R} + \hat{r}\bar{z}_2^T, \tag{3.32}$$

for $k < m$ simply add column $\hat{u}$ to $\hat{V}U$ and $\hat{r}$ to R.

Summarizing the updating process, we see that it is again very similar to the process in Section 3.1. First we choose vectors z_1, z_2, see recommendations in Section 3.5, and then we determine matrices $\underline{U}$, $\underline{R}$, $\underline{C}$, $\underline{H}$ by (2.6) and compute vectors $\hat{s}$ by (3.30) and $\hat{u}$, $\hat{r}$ by (3.31). Finally we compute number ζ_+ and matrices U_+, R_+ by (3.32) and determine matrices C_+ by (3.5) and H_+ by (3.4).

It can be readily verified that the partly inverse representation of (3.31) has the form

$$\frac{1}{\gamma}\underline{H}_+^{BC} = \left[\underline{C}^{-1} + \left(\frac{1-\eta}{\eta} + \frac{\underline{a}}{b}\frac{\gamma}{\varrho}\right)\frac{yy^T}{\underline{a}}\right]^{-1} + \hat{u}\hat{u}^T + \hat{V}\underline{U}\,\underline{U}^T\hat{V}^T. \tag{3.33}$$

3.5 Choice of parameters

Efficiency of all methods described in this report depends very much on the suitable choice of parameters. Since analysis of VM matrix damage caused by projections is very complicated, recommendations given here have empirical character.

The choice of scaling parameter γ plays the main role. The value $\gamma = 1$, i.e. scaling is omitted in some iterations, is not good here in general. In starting iterations ($\underline{H} = H$, $\underline{C} = C$), the value $\gamma = b/\underline{a}$ appears to be the best choice; the choices $\gamma = b/\tilde{a}$, $\gamma = b/\sqrt{\underline{a}\tilde{a}}$ give approximately the same results. In other iterations the choice $\gamma = b/\sqrt{\underline{a}\tilde{a}}$ gives good results, the value $\gamma = b/\sqrt{\tilde{a}\max(\underline{a}, \tilde{a} + |U^TBs|^2)}$ seems to be the best choice. In all iterations, if γ computed in this way is less than 10^{-3}, we set $\gamma = b/\tilde{a}$.

Also the choice of the projection vector z_1 has great influence. The good choices are $z_1^{(1)} = U^TBs - (y^TUU^TBs/|U^Ty|^2)U^Ty$, $z_1^{(2)} = U^TBs - (|U^TBs|/\sqrt{y^THy})\,U^Ty$, we use combination $z_1 = (1-\theta)z_1^{(1)} + \theta z_1^{(2)}$ with $\theta = |U^Ty|^2/(|U^Ty|^2 + |U^TBs|^2)$. Note that $y^TUz_1^{(1)} = 0$.

On the other hand, the choice of the vector z_2 has only little influence. We use the vector $z_2 = (y^TRR^TBs)\,R^TBs - (|R^TBs|^2)R^Ty$, which satisfies $s^TBRz_2 = 0$.

As regards nonquadratic correction parameter ϱ, no value seems to be the best, therefore we choose $\varrho = 1$.

4 Computational experiments

Our new limited-memory VM methods were tested, using the collection of sparse and partially separable test problems from [6] (Test 14 and Test 15, 22 problems each) with $N = 1000$, $m = 10$, $\varrho = 1$, $\eta = 0.8$ in starting iterations and $\eta = 1$ otherwise, the final precision $\|g(x^\star)\|_\infty \leq 10^{-6}$ for Test 14 and $\|g(x^\star)\|_\infty \leq 10^{-5}$ for Test 15.

In Table 1 we compare our method described in Section 3.4 (PLM) with the following methods: method given in [2] (BNS), shifted limited-memory method (SLM), see [7], and variationally-derived limited-memory method (VLM), see[8]. We present the total numbers of function and also gradient evaluations over all problems (NFE) and the total computational time (Time) in seconds.

Collection	BNS		SLM		VLM		PLM	
of problems	NFE	Time	NFE	Time	NFE	Time	NFE	Time
Test 14	22101	12.3	18693	10.3	19410	13.8	19684	18.2
Test 15	33241	21.3	38323	26.2	30862	24.5	31126	32.3

Table 1. Comparison with other methods for Test 14 and Test 15.

We see that the numerical results of the new methods are comparable with other methods as regards the number of evaluations. The longer total computational time is caused by the fact that the number of columns of matrices U, R are the same in this test version of method. We plan to reduce the number of columns of matrix R in future, see Section 2.

For a better comparison with method BNS, we performed additional tests with problems from the widely used CUTE collection [1] with various dimensions N and the final precision $\|g(x^\star)\|_\infty \leq 10^{-6}$. The results are given in Table 2, where NFE is the number of function and also gradient evaluations and Time the computational time in seconds.

CUTE		BNS		SLM		VLM		PLM	
Problem	N	NFV	Time	NFV	Time	NFV	Time	NFV	Time
BROYDN7D	2000	3021	10.14	3597	11.63	2858	10.61	2815	11.64
CURLY30	1000	7010	6.78	8246	7.50	6544	7.34	6761	8.97
DIXMAANI	3000	877	1.95	1368	2.84	2653	7.42	1020	3.52
DQRTIC	5000	236	0.43	242	0.51	967	3.02	288	1.24
FLETCBV2	1000	1182	1.09	1132	1.05	1248	1.45	919	1.34
GENHUMPS	1000	2271	3.63	4159	6.61	2698	5.06	2884	5.75
GENROSE	1000	2374	1.52	2347	1.51	2199	1.80	2102	2.17
MSQRTALS	529	2947	6.16	3501	7.13	3142	6.99	2832	6.91
NCB20B	1010	1715	8.86	2492	12.41	2204	11.47	2103	11.59
NONCVXU2	1000	3685	3.09	2917	2.59	2493	2.42	4027	5.17
NONDQUAR	5000	3588	8.47	7330	14.56	16090	54.03	5082	22.23
POWER	1000	110	0.01	135	0.03	496	0.13	99	0.04
QUARTC	5000	236	0.53	242	0.50	967	3.20	288	1.28

Table 2: Comparison with other methods for CUTE.

Our limited numerical experiments indicate that the additional improvements and testing of the new methods could be useful.

References

[1] 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.

[2] 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.

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

[4] J. Greenstadt, *Variations on variable metric methods*, Math. Comput. 24 (1970) 1-22.

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

[6] L. Lukšan, J. Vlček: *Sparse and partially separable test problems for unconstrained and equality constrained optimization*, Report V-767, Prague, ICS AS CR, 1998.

[7] J. Vlček, L. Lukšan: *Shifted variable metric methods for large-scale unconstrained optimization*, J. Comput. Appl. Math. 186 (2006) 365-390.

[8] J. Vlček, L. Lukšan: *New class of limited-memory variationally-derived variable metric methods*, Report V-973, Prague, ICS AS CR, 2006.