Transformations enabling to construct limited-memory Broyden class methods

Vlček, Jan
2008

Institute of Computer Science
Academy of Sciences of the Czech Republic

Transformations enabling to construct limited-memory Broyden class methods

J. Vlček, L. Lukšan

Technical report No. V 1037

December 2008

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

Transformations enabling to construct limited-memory Broyden class methods

J. Vlček, L. Lukšan[1]

Technical report No. V 1037

December 2008

Abstract:

The Broyden class of quasi-Newton updates for inverse Hessian approximation are transformed to the formal BFGS update, which makes possible to generalize the well-known Nocedal method based on the Strang recurrences to the scaled limited-memory Broyden family, using the same number of stored vectors as for the limited-memory BFGS method. Two variants are given, the simpler of them does not require any additional matrix by vector multiplications. Numerical results indicate that this approach can save computational time.

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

[1]This work was supported by the Grant Agency of the Czech Academy of Sciences, project No. IAA1030405, and the Institutional research plan No. AV0Z10300504, L. Lukšan is also from Technical University of Liberec, Hálkova 6, 461 17 Liberec.

1 Introduction

In this report we present a new family of limited-memory variable metric (VM) line search methods for unconstrained minimization, which generalizes the well-known limited-memory BFGS method, see [5], [2].

VM or quasi-Newton line search methods, see [3], 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 and $t_k > 0$ is a stepsize.

It is assumed 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

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

$k \ge 0$, 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; usually $H_0 = I$ and H_{k+1} is obtained from $\gamma_k H_k$ ($\gamma_k > 0$ is a scaling parameter) by a rank-two VM update to satisfy the quasi-Newton condition $H_{k+1} y_k = \varrho_k s_k$ (in generalized form), where $\varrho_k > 0$ is a nonquadratic correction parameter (see [3]), $y_k = g_{k+1} - g_k$, $k \ge 0$.

To simplify the notation we frequently omit index k and replace index $k+1$ by symbol $+$. We denote (note that $b > 0$ by (1.1))

$$B = H^{-1}, \quad a = y^T H y, \quad b = s^T y, \quad c = s^T B s, \quad V = I - (1/b) s y^T.$$

Limited-memory BFGS method is based on the following quasi-product form of the BFGS update

$$\frac{1}{\gamma} H_+ = V H V^T + \frac{\varrho}{\gamma b} s s^T. \tag{1.2}$$

If we restrict to initial scaling, recommended in [2], where also only unit nonquadratic correction parameters are used, it makes possible to define matrices H_k, $k \ge 0$, by $H_0 = I$, $H_k = H_j^k$, $j = \min(k, m)$, $k > 0$, $1 \le m \ll N$, where

$$H_0^k = \frac{b_{k-1}}{|y_{k-1}|^2} I, \tag{1.3}$$

$$H_{i+1}^k = V_{k-j+i} H_i^k V_{k-j+i}^T + \frac{\varrho_{k-j+i}}{b_{k-j+i}} s_{k-j+i} s_{k-j+i}^T, \quad 0 \le i < j\,. \tag{1.4}$$

Instead of matrices H_i^k, $2j$ vectors s_{k-j+i}, y_{k-j+i}, $0 \le i < j$, are stored and direction vectors $d_k = -H_k g_k$, $k > 0$, are computed using the Strang recurrences, see [5].

In Section 2 we mention some problems with generalization of this approach and describe two variants of transformation of the scaled standard Broyden class of VM updates to the quasi-product form resembling the BFGS update formula. In Section 3 we show how these transformations can be utilized in constructing of limited-memory methods, which we call the limited-memory Broyden class methods. Numerical results are presented in Section 4.

2 Transformations to the formal BFGS update

Although the scaled Broyden class update of H with positive value of parameter η can be written in the quasi-product form (see e.g. [6])

$$\frac{1}{\gamma}H_+ = \left(I - \left(\frac{\sqrt{\eta}}{b}s + \frac{1-\sqrt{\eta}}{a}Hy\right)y^T\right)H\left(I - y\left(\frac{\sqrt{\eta}}{b}s + \frac{1-\sqrt{\eta}}{a}Hy\right)^T\right) + \frac{\varrho}{\gamma}\frac{ss^T}{b}, \quad (2.1)$$

similar to (1.2), application of this form of update to generalizing of the limited-memory BFGS method is problematic; on the one hand we need to compute and store vectors Hy (additional matrix by vector multiplication is sometimes not necessary, see Section 3), on the other hand, VM matrices used in these vectors in previous iterations differ from the current VM matrices and thus we have not the Broyden class update.

Therefore we process in another way. We transform the Broyden class update to the formal BFGS update in transformed variables, which makes possible to construct limited-memory methods in a similar way as for the BFGS update, with the same number of stored vectors. Although we use the unit values of γ_k and ϱ_k in almost all cases, we will consider also non-unit values in the subsequent analysis as is usual in case of VM methods (see [3]). First we give the simple variant of the transformation. We denote

$$\omega = \frac{\varrho}{\gamma} + \frac{a}{b}\eta, \quad \mu = \eta + (1-\eta)\frac{\varrho}{\gamma}\frac{b}{a}. \quad (2.2)$$

Theorem 2.1. *Let $\varrho > 0$, $\gamma > 0$, $\omega \neq 0$, $\mu \geq 0$ and denote $\alpha = (\eta \pm \sqrt{\mu})/\omega$. Then $\sqrt{\mu} \neq -\eta$ and the scaled standard Broyden class update of matrix H with parameter η, scaling parameter γ and nonquadratic correction parameter ϱ can be expressed in the form*

$$\frac{1}{\gamma}H_+^{BC} = \frac{\varrho\,\eta}{\gamma b}\hat{s}\hat{s}^T + \check{V}H\check{V}^T, \quad \hat{s} = s - \alpha H\,y, \quad \check{V} = I \pm \frac{\sqrt{\mu}}{b}\hat{s}y^T. \quad (2.3)$$

Proof. (i) First we show that $\sqrt{\mu} \neq -\eta$. From (2.2) we have

$$\eta^2 - \mu = \eta^2 - \eta - (1-\eta)\varrho b/(\gamma a) = (\eta - 1)(\eta + \varrho b/(\gamma a)) = (\eta - 1)\,\omega b/a, \quad (2.4)$$

thus $\sqrt{\mu} \neq -\eta$ for $\eta \neq 1$ by $b > 0$ and obviously also for $\eta = 1$.

(ii) Consider the scaled Broyden class update with parameters η, γ and ϱ in the form, see [3],

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

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

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

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

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

In view of

$$\omega - \frac{a}{b}\,\mu = \frac{\varrho}{\gamma} + \frac{a}{b}\,\eta - \frac{a}{b}\,\eta - \frac{\varrho}{\gamma}(1-\eta) = \frac{\varrho}{\gamma}\eta \tag{2.5}$$

we have (2.3). □

Note that we prefer the minus sign in α and $\check{V}$, since then for $\eta = 1$ (BFGS) we get $\alpha = 0$, $\hat{s} = s$ and $\check{V} = V$. For $\eta \approx 1$ it is also $\sqrt{\mu} \approx 1$, therefore the formula for α above should be rewritten in another form. In view of (2.4) we obtain for $\omega \neq 0$ (thus also $\sqrt{\mu} \neq -\eta$ by Theorem 2.1)

$$\alpha = \frac{\eta - \sqrt{\mu}}{\omega} = \frac{\eta^2 - \mu}{\omega(\eta + \sqrt{\mu})} = \frac{(\eta-1)\,\omega b/a}{\omega(\eta + \sqrt{\mu})} = \frac{(\eta-1)b/a}{\eta + \sqrt{\mu}}\,. \tag{2.6}$$

For better understanding, condition $\mu \geq 0$ can be rewritten as $\eta(\varrho b - \gamma a) \leq \varrho b$, 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, $\eta_{SR1} = \varrho b/(\varrho b - \gamma a)$, see [3]. From (2.2) we obtain $\omega = (\varrho/\gamma)\eta \neq 0$ for $\eta = \eta_{SR1}$ and therefore Theorem 2.1 can also be used for the SR1 update.

For the transformation above, the similarity to the BFGS update is relatively free. Firstly $\check{V}$ is not a projection matrix in general, secondly matrix H_+^{BC} does not satisfy the quasi-Newton condition in transformed variables, since vector $H_+^{BC} y$ is not parallel to $\hat{s}$. Both these properties can be obtained if we introduce an additional transformation.

Theorem 2.2. *Let $\varrho > 0$, $\gamma > 0$, $\omega \neq 0$, $\mu > 0$, $\alpha = (\eta - \sqrt{\mu})/\omega$, $\hat{b} = b/\sqrt{\mu}$, $\hat{c} = c - 2\alpha b + \alpha^2 a$, $\hat{c} \neq 0$, $\beta = -(\varrho/\gamma)\alpha\hat{b}/\hat{c}$, $\hat{s} = s - \alpha H y$ and $\hat{y} = y - \beta B \hat{s}$. Then $\hat{b} = \hat{s}^T \hat{y} > 0$, $\hat{c} = \hat{s}^T B \hat{s} > 0$ and the scaled standard Broyden class update of matrix H with parameter η, scaling parameter γ and nonquadratic correction parameter ϱ can be expressed in the form*

$$\frac{1}{\gamma} H_+^{BC} = \frac{\hat{\varrho}}{\gamma \hat{b}} \hat{s}\hat{s}^T + \hat{V} H \hat{V}^T, \quad \hat{V} = I - \frac{1}{\hat{b}} \hat{s}\hat{y}^T, \quad \hat{\varrho} = \varrho\left(\frac{\eta}{\sqrt{\mu}} - \alpha\beta\right). \tag{2.7}$$

Moreover, if $\eta \geq 0$, then $\hat{\varrho} > 0$.

Proof. (i) First we establish $\hat{c} = \hat{s}^T B \hat{s} > 0$ and $\hat{b} = \hat{s}^T \hat{y} > 0$. From $\hat{s} = s - \alpha H y$ we get $\hat{s}^T B \hat{s} = (s - \alpha H y)^T (Bs - \alpha y) = c - 2\alpha b + \alpha^2 a = \hat{c}$. This yields

$$\hat{s}^T \hat{y} = \hat{s}^T (y - \beta B \hat{s}) = (s - \alpha H y)^T y - \beta \hat{s}^T B \hat{s} = b - \alpha a - \beta \hat{c},$$

which gives by (2.6), (2.2) and $\sqrt{\mu} \neq -\eta$, see Theorem 2.1,

$$\hat{s}^T\hat{y} = b - b\frac{\eta - 1}{\eta + \sqrt{\mu}} + \frac{\varrho}{\gamma}\hat{b}\alpha = b\frac{\sqrt{\mu} + 1}{\eta + \sqrt{\mu}} + \frac{b}{\sqrt{\mu}}\,\frac{(\eta - 1)\varrho b/(\gamma a)}{\eta + \sqrt{\mu}} = \frac{b(\mu + \sqrt{\mu} + \eta - \mu)}{\sqrt{\mu}(\eta + \sqrt{\mu})} = \hat{b}.$$

From $\hat{b} = b/\sqrt{\mu}$ and $b > 0$ we deduce $\hat{b} > 0$, thus also $\hat{s} \neq 0$, which implies $\hat{c} > 0$ by positive definiteness of B.

(ii) Next we show that $\hat{a}/\hat{b} + 2\beta = \sqrt{\mu}(a - \beta^2 \hat{c})/b$, where $\hat{a} = \hat{y}^T H \hat{y}$. From $a = y^T H y = (\hat{y} + \beta B \hat{s})^T (H\hat{y} + \beta \hat{s}) = \hat{a} + 2\beta\hat{b} + \beta^2 \hat{c}$ we obtain

$$\frac{\hat{a}}{\hat{b}} + 2\beta = \frac{\hat{a} + 2\beta\hat{b}}{\hat{b}} = \frac{a - \beta^2 \hat{c}}{\hat{b}} = \sqrt{\mu}\,\frac{a - \beta^2 \hat{c}}{b}. \tag{2.8}$$

(iii) As in the proof of Theorem 2.1 we get

$$\frac{1}{\gamma}H_+^{BC} = H + \frac{\omega}{b}\hat{s}\hat{s}^T - \frac{\sqrt{\mu}}{b}\left(H\, y\hat{s}^T + \hat{s}y^T H\right).$$

Setting $y = \hat{y} + \beta B\hat{s}$, we obtain by $b = \hat{b}\sqrt{\mu}$

$$\frac{1}{\gamma}H_+^{BC} = H - \frac{1}{\hat{b}}\left(H\hat{y}\hat{s}^T + \hat{s}\hat{y}^T H\right) + \frac{\omega - 2\beta\sqrt{\mu}}{\hat{b}\sqrt{\mu}}\hat{s}\hat{s}^T = \hat{V}H\hat{V}^T + \frac{1}{\hat{b}}\left[\frac{\omega}{\sqrt{\mu}} - 2\beta - \frac{\hat{a}}{\hat{b}}\right]\hat{s}\hat{s}^T.$$

To complete the proof, we rewrite the expression in brackets, using (2.8) and (2.5):

$$\frac{\omega}{\sqrt{\mu}} - 2\beta - \frac{\hat{a}}{\hat{b}} = \frac{\omega}{\sqrt{\mu}} - \frac{a}{b}\sqrt{\mu} + \beta^2\frac{\hat{c}}{b}\sqrt{\mu} = \frac{1}{\sqrt{\mu}}\left(\frac{\varrho}{\gamma}\eta\right) + \frac{\beta\hat{c}}{\hat{b}}\beta = \frac{\varrho}{\gamma}\left(\frac{\eta}{\sqrt{\mu}} - \alpha\beta\right) = \frac{\hat{\varrho}}{\gamma}.$$

Since $\eta = 1$ for $\alpha = 0$ due to (2.6), we see that $\hat{\varrho}/\varrho \equiv \eta/\sqrt{\mu} - \alpha\beta = \eta/\sqrt{\mu} + (\varrho/\gamma)\alpha^2\hat{b}/\hat{c} > 0$ holds for $\eta \geq 0$ by $\hat{b} > 0$ and $\hat{c} > 0$. □

Obviously, the quasi-Newton condition $H_+^{BC}\hat{y} = \hat{\varrho}\hat{s}$ in transformed variables is satisfied by (2.7). Note that $\hat{c} = c - 2\alpha b + \alpha^2 a$ can be near to zero and therefore it is better to compute it e.g. by

$$\hat{c} = \frac{ac - b^2 + (\alpha a - b)^2}{a} = c - \frac{b^2}{a} + \frac{b^2}{a}\left(\frac{\sqrt{\mu} + 1}{\sqrt{\mu} + \eta}\right)^2 \tag{2.9}$$

in view to (2.6) and $\sqrt{\mu} \neq -\eta$, see Theorem 2.1, where $ac \geq b^2$ by the Schwarz inequality.

3 Application to limited-memory methods

Theory in the previous section enables us to view the Broyden class updates with $\mu > 0$ formally as the BFGS update in transformed variables.

Theorem 3.1. *Let the assumptions of Theorem 2.1 be satisfied, $\mu > 0$, $\check{b} = \mp b/\sqrt{\mu}$ and $\check{\varrho} = \mp\varrho\eta/\sqrt{\mu}$. Then the scaled standard Broyden class update of matrix H with parameter η, scaling parameter γ and nonquadratic correction parameter ϱ can be formally expressed as the scaled BFGS update in the form (1.2) with s, b, ϱ and V replaced by $\hat{s}$, $\check{b}$, $\check{\varrho}$ and $\check{V}$.*

Proof. See Theorem 2.1. □

Theorem 3.2. *Let the assumptions of Theorem 2.2 be satisfied. Then the scaled standard Broyden class update of matrix H with parameter η, scaling parameter γ and nonquadratic correction parameter ϱ can be formally expressed as the scaled BFGS update in the form (1.2) with s, y, b, ϱ and V replaced by $\hat{s}$, $\hat{y}$, $\hat{b}$, $\hat{\varrho}$ and $\hat{V}$.*

Proof. See Theorem 2.2. □

Therefore we can again define matrices H_0^k by (1.3) and $H_k = H_j^k$, $j = \min(k, m)$, $k > 0$, by relations similar to (1.4) according to Theorem 3.1 or Theorem 3.2. Instead of matrices H_i^k, $2j \le 2m$ vectors $\hat{s}_{k-j+i}$ and y_{k-j+i} or $\hat{y}_{k-j+i}$ are stored here, together with numbers $\check{b}_{k-j+i}$, $\check{\varrho}_{k-j+i}$ or $\hat{b}_{k-j+i}$, $\hat{\varrho}_{k-j+i}$, $0 \le i < j$. In addition to vector $H_k g_k$, another vector $H_k y_k$, $k > 0$, should be computed here (it is not necessary in case of the simpler transformation according to Theorem 2.1, see below). Note that transformed nonquadratic correction parameters are not unit here, which requires a little modification of the Strang recurrences, see [5], used for computing of vectors $H_k g_k$, $H_k y_k$.

In the next algorithms which correspond to the transformation given in Theorem 2.1 we also consider the case $\mu_k = 0$, i.e. the SR1 update. Then variables $\check{b}$, $\check{\varrho}$ are not defined and thus we use in relation (3.1) term η_{k-j+i}/b_{k-j+i} instead of $\check{\varrho}_{k-j+i}/\check{b}_{k-j+i}$, which again requires a little modification of the Strang recurrences. Moreover, if $\mu_k = 0$ we have $\check{V}_k = I$ and thus the Strang recurrences can be simplified.

We shall now state the limited-memory Broyden class methods in details. In the first two algorithms we use one additional matrix by vector multiplication per iteration in comparison with the limited-memory BFGS method. Algorithm 3.1 corresponds to the transformation given in Theorem 2.1, Algorithm 3.2 to the transformation given in Theorem 2.2. Everywhere we suppose $\varrho = \gamma = 1$, except for the initial scaling (1.3).

Algorithm 3.1

Data: The number m of VM updates per iteration, line search parameters ε_1 and ε_2, $0 < \varepsilon_1 < 1/2$, $\varepsilon_1 < \varepsilon_2 < 1$.

Step 0: Initiation. Choose the starting point $x_1 \in \mathcal{R}^N$, define starting matrix $H_0^0 = I$ and direction vector $d_0 = -g_0$ and set the iteration counter $k = 0$.

Step 1: Line search. Compute $x_{k+1} = x_k + t_k d_k$, where t_k satisfies (1.1), $g_{k+1} = \nabla f(x_{k+1})$, $y_k = g_{k+1} - g_k$, b_k, $H_k y_k$ (by the modified Strang recurrences, see discussion above, using matrices $\{H_i^k\}_{i=0}^{\min(k,m)}$) and a_k.

Step 2: Transformation. Choose parameter η_k of the Broyden class update satisfying $\mu_k \ge 0$. Using Theorem 2.1, compute α_k and $\hat{s}_k$ and define $\check{V}_k$.

Step 3: Updates definition. Set $k := k + 1$, $j = \min(k, m)$ and define H_0^k by (1.3) and $H_k = H_j^k$ by

$$H_{i+1}^k = \check{V}_{k-j+i} H_i^k \check{V}_{k-j+i}^T + \frac{\eta_{k-j+i}}{b_{k-j+i}} \hat{s}_{k-j+i} \hat{s}_{k-j+i}^T \,, \quad 0 \le i < j \,. \qquad (3.1)$$

Step 4: Direction vector. Compute $d_k = -H_k g_k$ by the modified Strang recurrences, using matrices $\{H_i^k\}_{i=0}^{\min(k,m)}$, and goto Step 1.

Algorithm 3.2

Data: The number m of VM updates per iteration, line search parameters ε_1 and ε_2, $0 < \varepsilon_1 < 1/2$, $\varepsilon_1 < \varepsilon_2 < 1$.

Step 0: Initiation. Choose the starting point $x_1 \in \mathcal{R}^N$, define starting matrix $H_0^0 = I$ and direction vector $d_0 = -g_0$ and set the iteration counter $k = 0$.

Step 1: Line search. Compute $x_{k+1}=x_k+t_kd_k$, where t_k satisfies (1.1), $g_{k+1}=\nabla f(x_{k+1})$, $y_k=g_{k+1}-g_k$, b_k, H_ky_k (by the modified Strang recurrences, see discussion above, using matrices $\{H_i^k\}_{i=0}^{\min(k,m)}$) and a_k.

Step 2: Transformation. Choose parameter η_k of the Broyden class update satisfying $\mu_k > 0$ and $\omega_k \neq 0$. Using Theorem 2.2, compute α_k, $\hat{b}_k$, $\hat{c}_k$, β_k, $\hat{s}_k$, $\hat{y}_k$ and $\hat{\varrho}_k$ and define $\hat{V}_k$.

Step 3: Updates definition. Set $k := k+1$, $j = \min(k,m)$ and define H_0^k by (1.3) and $H_k = H_j^k$ by

$$H_{i+1}^k = \hat{V}_{k-j+i} H_i^k \hat{V}_{k-j+i}^T + \frac{\hat{\varrho}_{k-j+i}}{\hat{b}_{k-j+i}} \hat{s}_{k-j+i} \hat{s}_{k-j+i}^T\,, \quad 0 \le i < j\,. \tag{3.2}$$

Step 4: Compute $d_k = -H_kg_k$ by the modified Strang recurrences, using matrices $\{H_i^k\}_{i=0}^{\min(k,m)}$, and goto Step 1.

Classical Broyden class updates of matrix H, see [1], [3], need vector Hy besides the direction vector $d = -Hg$ in every iteration, but it does not mean that we must calculate two matrix by vector multiplications per iteration. If we have computed vector Hy, then the next direction vector can be expressed as a linear combination of vectors s, Hy.

Lemma 3.1. *Consider the scaled Broyden update of matrix H with parameter η, scaling parameter γ and nonquadratic correction parameter ϱ and denote by t the stepsize. Then the direction vector $d_+ = -H_+g_+$ can be written in the form*

$$td_+ = \left[\gamma\eta\left(\frac{ac}{b^2}-1\right)+\gamma+\varrho\left(\frac{c}{b}-t\right)\right]s - \frac{b}{a}\left[\gamma\eta\left(\frac{ac}{b^2}-1\right)+\gamma\right]Hy. \tag{3.3}$$

Proof. Writing VM update in the form $(1/\gamma)H_+ = H + \Delta$, we get by $s = -tHg$ and the quasi-Newton condition $H_+y = \varrho s$

$$tH_+g_+ = tH_+y + t\gamma(H+\Delta)g = (\varrho t - \gamma)s - \gamma\Delta Bs. \tag{3.4}$$

For the scaled Broyden class update we have, see [3],

$$\Delta = \frac{1}{b}\left(\frac{\varrho}{\gamma}+\frac{a}{b}\eta\right)ss^T - \frac{\eta}{b}\left(H\,ys^T + sy^TH\right) + \frac{\eta-1}{a}H\,yy^TH\,. \tag{3.5}$$

Therefore

$$\Delta Bs = \left[\frac{\varrho}{\gamma}\frac{c}{b}+\eta\left(\frac{ac}{b^2}-1\right)\right]s - \frac{b}{a}\left[\eta\left(\frac{ac}{b^2}-1\right)+1\right]Hy,$$

which together with (3.4) gives (3.3). □

Note that in the frequent case $\varrho = \gamma = t = 1$ we have $d_+ = \Delta Bs$ by (3.4).

This approach cannot be used directly for our limited-memory methods, since matrix H_k used for calculation of vector H_ky_k is created by updating of matrix $H_0^k = (b_{k-1}/|y_{k-1}|^2)I$, which is different from matrix $H_0^{k+1} = (b_k/|y_k|^2)I$, and thus matrix

$H_k = H^k_{\min(k,m)}$ satisfying $s_k = -t_k H_k g_k$ is different from the matrix which we update for the last time to obtain matrix H_{k+1}. The result is that vector s is here only approximation of $-tHg$ and the right side of (3.3) is only a poor approximation of td_+.

The situation is even worse when we start creating of matrix H_k with matrix $(b_k/|y_k|^2)I$ instead of $(b_{k-1}/|y_{k-1}|^2)I$. Nevertheless, Lemma 3.1 can be used to save one matrix by vector multiplication per iteration in case of transformation given by Theorem 2.1. The idea consist in using an approximation of vector $H_k g_{k+1}$, calculated by updating of matrix H_0^{k+1}, to approximate vector $H_k y_k$. Unfortunately, this idea is not suitable for transformation given by Theorem 2.2, where condition $s_k = -t_k H_k g_k$ must be satisfied more accurately.

Algorithm 3.3

Data: The number m of VM updates per iteration, line search parameters ε_1 and ε_2, $0 < \varepsilon_1 < 1/2$, $\varepsilon_1 < \varepsilon_2 < 1$.

Step 0: Initiation. Choose the starting point $x_1 \in \mathcal{R}^N$, define starting matrix $H_0^0 = I$ and direction vector $d_0 = -g_0$ and set the iteration counter $k = 0$.

Step 1: Line search. Compute $x_{k+1} = x_k + t_k d_k$, where t_k satisfies (1.1), $g_{k+1} = \nabla f(x_{k+1})$, $y_k = g_{k+1} - g_k$, b_k, c_k and $H_k g_{k+1}$ by the modified Strang recurrences, see discussion before Algorithm 3.1, using matrices $\{H_i^k\}_{i=0}^{\min(k,m)}$.

Step 2: Approximation. Define $H_k y_k = H_k g_{k+1} + (1/t_k) s_k$ and $a_k = y_k^T H_k y_k$.

Step 3: Transformation. Choose parameter η_k of the Broyden class update satisfying $\mu_k \geq 0$. Using Theorem 2.1, compute α_k and $\hat{s}_k$ and define $\check{V}_k$.

Step 4: Updates definition. Let $j = \min(k, m)$. Define $H_0^k = (b_k/|y_k|^2)I$ and $H_k = H_j^k$ by (3.1).

Step 5: Direction vector. Compute d_{k+1} by (3.3).

Step 6: Loop. Set $k := k + 1$ and goto Step 1.

4 Computational experiments

In this section we demonstrate the influence of parameter η on the number of evaluations and computational time, using the collection of sparse and partially separable test problems from [4] (Test 14, 22 problems each) with 1000, $m = 10$, $\varrho = \gamma = 1$ and the final precision $\|g(x^\star)\|_\infty \leq 10^{-6}$.

Results for Algorithm 3.2 and Algorithm 3.3 are given in Table 1, where 'NFE' is the total numbers of function and also gradient evaluations over all problems and 'Time' the total computational time (Time) in seconds.

Our limited numerical experiments indicate that

- the efficiency of Algorithm 3.1 and Algorithm 3.2 is practically the same,
- it is possible to generalize limited-memory BFGS method with the same number of matrix by vector multiplication and number of stored vectors,

- the suitable choice of parameter η can improve efficiency of limited-memory methods.

η	Alg. 3.1		Alg. 3.2		Alg. 3.3	
	NFE	Time	NFE	Time	NFE	Time
0.5	24443	11.45	24025	11.17	22148	8.68
0.6	23379	10.94	24041	11.25	22903	8.69
0.7	23463	11.05	22440	10.35	23176	8.81
0.8	22687	10.50	23020	10.66	22649	8.64
0.9	21513	10.04	22005	10.00	21058	8.03
1.0	22419	10.19	22389	10.08	22139	8.44
1.1	21410	9.64	22478	10.21	21179	8.03
1.2	21813	9.96	21642	10.05	22008	8.27
1.3	21181	9.49	21696	9.61	20848	7.94
1.4	21688	9.67	21589	9.61	21164	8.09
1.5	21525	9.69	22112	9.83	22285	8.23
1.6	22044	9.70	21948	9.66	22311	8.37
1.7	22248	9.87	22220	9.77	21911	7.98
1.8	22006	9.89	21628	9.63	23416	8.66
1.9	23030	10.12	22065	9.73	23259	8.30
2.0	23017	10.06	22418	9.76	24640	8.97

Table 1. Influence of parameter η.

References

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

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

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

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

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

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