INSTITUTE OF COMPUTER SCIENCE

ACADEMY OF SCIENCES OF THE CZECH REPUBLIC

Additional properties of shifted variable metric methods

J. Vlček, L. Lukšan

Technical report No. V 899

January 2004

Institute of Computer Science, Academy of Sciences of the Czech Republic
Pod vodárenskou věží 2, 182 07 Prague 8, Czech Republic
phone: (+420)266052083 fax: (+420)286585789
e-mail: luksan@cs.cas.cz, vlcek@cs.cas.cz

INSTITUTE OF COMPUTER SCIENCE

ACADEMY OF SCIENCES OF THE CZECH REPUBLIC

Additional properties of shifted variable metric methods

J. Vlček, L. Lukšan[1]

Technical report No. V 899
January 2004

Abstract

Some supplements to shifted variable metric or quasi-Newton methods for unconstrained minimization are given, including new limited-memory methods. Global convergence of these methods can be established for convex sufficiently smooth functions. Some encouraging numerical experience is reported.

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

[1]This work was supported by the Grant Agency of the Czech Academy of Sciences, project code A1030405, and with the subvention from Ministry of Education of the Czech Republic, project code MSM 242200002, L. Lukšan is also from Technical University of Liberec, Hálkova 6, 461 17 Liberec

1 Introduction

In this report we present some new properties of shifted variable metric (VM) line search methods, see [9], for unconstrained minimization. We recall that these methods are iterative. Starting with an initial point $x_1 \in \mathcal{R}^N$, they generate a sequence $x_k \in \mathcal{R}^N$, $k \geq 1$, by the process $x_{k+1} = x_k + s_k$, $s_k = t_k d_k$, where $d_k \in \mathcal{R}^N$ is a direction vector and the stepsize t_k is chosen in such a way that $t_k > 0$ and

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

$k \geq 1$, with $0 < \varepsilon_1 < 1/2$ and $\varepsilon_1 < \varepsilon_2 < 1$, where $f_k = f(x_k)$, $g_k = \nabla f(x_k)$.

We assume that the problem function $f : \mathcal{R}^N \to \mathcal{R}$ has continuous second-order derivatives on the level set $\{x \in \mathcal{R}^N : \ f(x) \leq f(x_1)\}$ and the direction vector satisfies

$$d_k = -H_k g_k, \qquad k \geq 1, \tag{1.2}$$

where H_k is a symmetric positive definite matrix. We denote $y_k = g_{k+1} - g_k$, $k \geq 1$.

We give some additional properties of shifted variable metric methods in Section 2 and numerical results in Section 3.

2 Supplements to shifted variable metric methods

In shifted VM methods, see [9], matrices H_k have the form

$$H_k = \zeta_k I + A_k, \tag{2.1}$$

$k \geq 1$, where $\zeta_k > 0$ and A_k are symmetric positive semidefinite matrices; usually $A_1 = 0$ and A_{k+1} is obtained from $\gamma_k A_k$ ($\gamma_k > 0$ is a scaling parameter) by a rank-two VM update to satisfy the shifted quasi-Newton condition (in generalized form)

$$A_{k+1} y_k = \varrho_k \tilde{s}_k, \quad \zeta_{k+1} = \varrho_k \sigma_k, \tag{2.2}$$

where $\tilde{s} = s_k - \sigma_k y_k$, $\sigma_k = \mu_k s_k^T y_k / y_k^T y_k$ is a shift parameter, $\mu_k \in (0,1)$ is a relative shift parameter and $\varrho_k > 0$ is a nonquadratic correction parameter (see [5]). Obviously, relations (2.1)-(2.2) imply that matrix H_{k+1} satisfies the quasi-Newton condition $H_{k+1} y_k = \varrho_k s_k$. In the subsequent analysis we use the following notation

$$\begin{array}{lllll} a_k = y_k^T H_k y_k, & b_k = s_k^T y_k, & c_k = s_k^T B_k s_k, & B_k = H_k^{-1}, & \tilde{b}_k = \tilde{s}_k^T y_k, \\ \bar{a}_k = y_k^T A_k y_k, & \bar{b}_k = s_k^T B_k A_k y_k, & \bar{c}_k = s_k^T B_k A_k B_k s_k, & \bar{\delta}_k = \bar{a}_k \bar{c}_k - \bar{b}_k^2, & \hat{a}_k = y_k^T y_k, \end{array}$$

$k \geq 1$. Note that the Schwarz inequality implies $\bar{\delta}_k \geq 0$, $k \geq 1$. To simplify the notation we frequently omit index k and replace index $k+1$ by symbol $+$. 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 it is usual in case of VM methods (see [5]).

Involving the scaling and the nonquadratic correction and using the same argumentation as in standard VM methods, we can write the shifted analogy of the Broyden class (see [2], [5]) for $\tilde{b} > 0$ (which implies $\tilde{s} \neq 0$, $y \neq 0$) and $\eta \geq 0$ in the form

$$\frac{1}{\gamma} A_+ = A + \frac{\varrho}{\gamma} \frac{\tilde{s}\tilde{s}^T}{\tilde{b}} - \frac{A y y^T A}{\bar{a}} + \frac{\eta}{\bar{a}} \left(\frac{\bar{a}}{\tilde{b}} \tilde{s} - A y \right) \left(\frac{\bar{a}}{\tilde{b}} \tilde{s} - A y \right)^T \tag{2.3}$$

(if $\bar{a} = 0$, i.e. $Ay = 0$, we simply omit the last two terms, because they tend to zero for $Ay = \lim_{\xi \to 0} \xi q$, $\bar{a} = \lim_{\xi \to 0} \xi q^T y$, $q^T y \neq 0$), where η is a free parameter (verification of $A_+ y = \varrho \tilde{s}$ for this update is straightforward). There are two important special cases. For $\eta = 0$ we obtain the shifted DFP update, for $\eta = 1$ the shifted BFGS update

$$\frac{1}{\gamma} A_+^{sDFP} = A + \frac{\varrho}{\gamma} \frac{\tilde{s}\tilde{s}^T}{\tilde{b}} - \frac{Ayy^TA}{\bar{a}}, \quad \frac{1}{\gamma} A_+^{sBFGS} = A + \left(\frac{\varrho}{\gamma} + \frac{\bar{a}}{\tilde{b}} \right) \frac{\tilde{s}\tilde{s}^T}{\tilde{b}} - \frac{\tilde{s}y^TA + Ay\tilde{s}^T}{\tilde{b}}. \tag{2.4}$$

In limited-memory VM methods, matrix A has the form $A = UU^T$, where $U = (u_1, \ldots, u_m)$ is a rectangular matrix with m columns, $m \geq 1$, and use the VM update

$$A_+ = \gamma VAV^T, \tag{2.5}$$

where transformation matrix V has the form $I + pq^T$ for the type 1 methods, or $I + p_1 y^T + p_2 s^T B$ for the type 2 methods. Thus we need to store only matrix U, which can be updated using relation $U_+ = \sqrt{\gamma}\, VU$.

2.1 General expression of limited-memory method for $m \leq 2$

Theorem 3.3 in [9] easily follows from the following theorem.

Theorem 2.1. *Let $A = uu^T$, $u^Ty \neq 0$. Then $A_+^{sDFP} = \varrho\, \tilde{s}\tilde{s}^T / \tilde{b}$.*

Proof. Since $Ay = (u^Ty)u$, we obtain from (2.4)

$$\frac{1}{\gamma} A_+^{sDFP} = uu^T + \frac{\varrho}{\gamma} \frac{\tilde{s}\tilde{s}^T}{\tilde{b}} - (u^Ty)^2 \frac{uu^T}{(u^Ty)^2} = \frac{\varrho}{\gamma} \frac{\tilde{s}\tilde{s}^T}{\tilde{b}}. \qquad \square$$

This result can be generalized for rank-two matrix A:

Theorem 2.2. *Let $A = u_1 u_1^T + u_2 u_2^T$, $v_2 = \bar{a}ABs - \bar{b}Ay$, $\bar{a}\bar{\delta} \neq 0$. Then*

$$\frac{1}{\gamma} A_+^{sDFP} = \frac{\varrho}{\gamma} \frac{\tilde{s}\tilde{s}^T}{\tilde{b}} + \frac{v_2 v_2^T}{\bar{a}\bar{\delta}}. \tag{2.6}$$

Proof. Denoting $\alpha_i = u_i^T y$, $i = 1, 2$, we obtain $Ay = \alpha_1 u_1 + \alpha_2 u_2$, $\bar{a} = \alpha_1^2 + \alpha_2^2$ and similar relations for ABs, $\bar{b}$ and $\bar{c}$. Thus for $\bar{\delta} = \bar{a}\bar{c} - \bar{b}^2$ and v_2 we obtain

$$\begin{aligned}
\bar{\delta} &= (\alpha_1^2 + \alpha_2^2)((u_1^TBs)^2 + (u_2^TBs)^2) - (\alpha_1 u_1^TBs + \alpha_2 u_2^TBs)^2 = (\alpha_2 u_1^TBs - \alpha_1 u_2^TBs)^2, \\
v_2 &= (\alpha_1^2 + \alpha_2^2)(u_1^TBs \cdot u_1 + u_2^TBs \cdot u_2) - (\alpha_1 u_1^TBs + \alpha_2 u_2^TBs)(\alpha_1 u_1 + \alpha_2 u_2) \\
&= (\alpha_2 u_1^TBs - \alpha_1 u_2^TBs)(\alpha_2 u_1 - \alpha_1 u_2).
\end{aligned}$$

Since $A - (1/\bar{a})Ayy^TA = [(\alpha_1^2 + \alpha_2^2)(u_1u_1^T + u_2u_2^T) - (\alpha_1 u_1 + \alpha_2 u_2)(\alpha_1 u_1 + \alpha_2 u_2)^T]/\bar{a}$, we have by (2.4)

$$\frac{1}{\gamma} A_+^{sDFP} - \frac{\varrho}{\gamma} \frac{\tilde{s}\tilde{s}^T}{\tilde{b}} = \frac{1}{\bar{a}} \left[\alpha_2^2\, u_1 u_1^T + \alpha_1^2\, u_2 u_2^T - \alpha_1\alpha_2 (u_1 u_2^T + u_2 u_1^T) \right] = \frac{v_2 v_2^T}{\bar{a}\bar{\delta}}. \qquad \square$$

Using (2.4), we can combine (2.6) with the general type 2 method expression (4.24) in [9], which gives the general form of type 1 or type 2 update for limited memory methods with $m \leq 2$

$$\frac{1}{\gamma} A_+ = \frac{\varrho}{\gamma} \frac{\tilde{s}\tilde{s}^T}{\tilde{b}} + \frac{q_2 q_2^T}{\bar{a}\bar{\delta}}. \tag{2.7}$$

Especially, the choice $q_2 = 0$ or $q_2 = v_2$ gives the shifted DFP update for $m = 1$ or $m = 2$. This interesting formula needs not store any VM matrix, similarly as conjugate gradient methods, but can be much more efficient.

Good results were also obtained with the choice $q_2 = \hat{w} \triangleq \sqrt{\eta\bar{\delta}}\left((\bar{a}/\tilde{b})\tilde{s} - Ay\right)$ (method SSBC in [9]), similar as with the shifted DFP method.

2.2 Variationally-derived limited-memory methods

Standard VM methods can be obtained by solving a certain variational problem - we find an update with the smallest correction of VM matrix in the sense of some norm (see [5]). Using the product form of the update, we can extend this approach to limited-memory methods to derive a very efficient class of methods. First we give the following general theorem, where the shifted quasi-Newton condition $U_+U_+^Ty = A_+y = \varrho\tilde{s}$ is equivalently replaced by (the first two conditions imply the third one)

$$U_+^T y = \sqrt{\gamma} z, \qquad U_+(\sqrt{\gamma} z) = \varrho \tilde{s}, \qquad z^T z = (\varrho/\gamma)\tilde{b}. \tag{2.8}$$

Theorem 2.3. *Let T be a symmetric positive definite matrix, $z \in \mathcal{R}^N$ and denote $\mathcal{U}$ the set of $N \times m$ matrices. Then the unique solution to*

$$\min\{\varphi(U_+) : U_+ \in \mathcal{U}\} \text{ s.t. } (2.8), \quad \varphi(U_+) = y^T T y \, \|T^{-1/2}(U_+ - \sqrt{\gamma} U)\|_F^2, \tag{2.9}$$

(Frobenius matrix norm) is

$$\frac{1}{\sqrt{\gamma}} U_+ = U - \frac{Ty}{y^T T y} y^T U + \left(\frac{\varrho}{\gamma}\tilde{s} - Uz + \frac{y^T U z}{y^T T y} T y\right)\frac{z^T}{z^T z} \tag{2.10}$$

and for this solution the value of $\varphi(U_+)/\gamma$ is

$$\frac{1}{\gamma}\varphi(U_+) = |U^T y - z|^2 + \frac{y^T T y}{z^T z} v^T T^{-1} v, \quad v = \frac{\varrho}{\gamma}\tilde{s} - Uz - \frac{(\varrho/\gamma)\tilde{b} - y^T U z}{y^T T y} T y. \tag{2.11}$$

Proof. Setting $U_+ = (u_1^+, \dots, u_m^+)$, define Lagrangian function $\mathcal{L} = \mathcal{L}(U_+, e_1, e_2)$ as

$$\begin{aligned} \mathcal{L} \;&=\; \frac{1}{2}\varphi(U_+) + e_1^T(U_+^T y - \sqrt{\gamma} z) + e_2^T(\sqrt{\gamma} U_+ z - \varrho \tilde{s}) \;=\; -\sqrt{\gamma} e_1^T z - \varrho e_2^T \tilde{s} \\ &+ \sum_{i=1}^m \left[\frac{y^T T y}{2}(u_i^+ - \sqrt{\gamma} u_i)^T T^{-1}(u_i^+ - \sqrt{\gamma} u_i) + e_{1i} y^T u_i^+ + \sqrt{\gamma} z_i e_2^T u_i^+\right]. \end{aligned}$$

A local minimizer U_+ satisfies the equations $\partial\mathcal{L}/\partial u_i^+ = 0$, $i = 1, \dots, m$, which gives $y^T T y \, T^{-1}(u_i^+ - \sqrt{\gamma} u_i) + e_{1i} y + \sqrt{\gamma} z_i e_2 = 0$, $i = 1, \dots, m$, yielding

$$U_+ = \sqrt{\gamma} U - \frac{Ty}{y^T T y} e_1^T - \sqrt{\gamma} \frac{T e_2}{y^T T y} z^T. \tag{2.12}$$

Using the first condition in (2.8), we have $e_1 = \sqrt{\gamma} U^T y - \sqrt{\gamma}(1 + y^T T e_2 / y^T T y) z$.

Substituing this e_1 to (2.12), we obtain $(1/\sqrt{\gamma}) U_+ = U - T y y^T U / y^T T y + \bar{e} z^T$ with some vector $\bar{e}$. The second condition in (2.8) yields

$$\bar{e} = \frac{1}{z^T z}\left(\frac{\varrho}{\gamma}\tilde{s} - Uz + \frac{y^T U z}{y^T T y} T y\right) \tag{2.13}$$

and (2.10) follows. Matrix U_+ obtained in this way minimizes φ in view of convexity of Frobenius norm. Furthermore, we get

$$\bar{e} - \frac{Ty}{y^T Ty} = \frac{1}{z^T z}\left(\frac{\varrho}{\gamma}\tilde{s} - Uz - \frac{z^T z - y^T U z}{y^T T y} Ty\right) = \frac{v}{z^T z} \tag{2.14}$$

by (2.8) and (2.11), thus by (2.10) and $v^T y = 0$

$$\begin{aligned} \frac{1}{\gamma}\frac{\varphi(U_+)}{y^T T y} &= \left\| T^{-1/2}\left(\frac{Ty}{y^T T y} y^T U - \bar{e} z^T\right)\right\|_F^2 = \left\| T^{-1/2}\left(\frac{Ty}{y^T T y}(U^T y - z)^T - \frac{v}{z^T z} z^T\right)\right\|_F^2 \\ &= \mathrm{Tr}\left(\frac{(U^T y - z)(U^T y - z)^T}{y^T T y} + \frac{v^T T^{-1} v}{(z^T z)^2} z z^T\right) = \frac{|U^T y - z|^2}{y^T T y} + \frac{v^T T^{-1} v}{z^T z}. \quad \square \end{aligned}$$

The choice of matrix T, when vectors Ty, $(\varrho/\gamma)\tilde{s} - Uz$, are linearly dependent, represents an important special case, since then $v = 0$ (thus the value of $\varphi(U_+)$ reaches its minimum on the set of symmetric positive definite matrices T), which implies $\bar{e} = Ty/y^T Ty = ((\varrho/\gamma)\tilde{s} - Uz)/((\varrho/\gamma)\tilde{b} - y^T U z)$ by (2.14) and in view of (2.13), update (2.10) can be written in the form

$$\frac{1}{\sqrt{\gamma}} U_+ = U - \frac{(\varrho/\gamma)\tilde{s} - Uz}{(\varrho/\gamma)\tilde{b} - y^T U z}\left(U^T y - z\right)^T. \tag{2.15}$$

The first term in (2.11), i.e. $|U^T y - z|^2$, can also be easily minimized subject to $z^T z = (\varrho/\gamma)\tilde{b}$. The solution is $z = \pm\sqrt{(\varrho/\gamma)\tilde{b}/\bar{a}}\, U^T y$, which gives the shifted DFP method in view of the following lemma.

Lemma 2.1. *Every update of the form $(1/\sqrt{\gamma})U_+ = U + p y^T U$, $p \in \mathcal{R}^N$, which satisfies the shifted quasi-Newton condition (2.8), is the shifted DFP method.*

Proof. From (2.8) we have $z = (1 + p^T y) U^T y = \pm\sqrt{(\varrho/\gamma)\tilde{b}/\bar{a}}\, U^T y$. Furthermore, using again (2.8), we obtain $(\varrho/\gamma)\tilde{s} = \pm\sqrt{(\varrho/\gamma)\tilde{b}/\bar{a}}\,(Ay + \bar{a})p$, thus $p = (\pm\sqrt{(\varrho/\gamma)\bar{a}/\tilde{b}}\,\tilde{s} - Ay)/\bar{a}$, which is the shifted DFP method (see [5] or [9]). $\square$

Using this lemma, we can also see that the only limited-memory type 2 method satisfying $\tau_2 = p_2^T y = 0$ (see [9]), which can be written in the form (2.10), is the shifted DFP method, since $\bar{e}^T y = 1$ by (2.13) and (2.8). It may explain the less efficiency of methods with $\tau_2 = 0$ (all type 2 methods in our report [9]) in comparison with variationally-derived methods.

The advantage of variationally-derived update formulas (2.10), (2.15) consists in possibility of parameters choice (z and in (2.10) also Ty). By comparison with the standard Broyden class (see [5]), we get meaning of these parameters. To use Theorem 2.3 for the standard Broyden class, we set $H = SS^T$ and replace U, $\tilde{s}$ and $\tilde{b}$ by S, s and b. Then update (2.10) will be replaced by

$$\frac{1}{\sqrt{\gamma}} S_+ = S - \frac{Ty}{y^T T y} y^T S + \left(\frac{\varrho}{\gamma}s - Sz + \frac{y^T S z}{y^T T y} Ty\right)\frac{z^T}{z^T z} \tag{2.16}$$

and the following assertion holds. Note that scaling of Ty has no influence on vector $Ty/y^T Ty$.

Lemma 2.2. *Every update (2.16) with $z = \alpha_1 S^T y + \alpha_2 S^T Bs$, $Ty = \beta_1 s + \beta_2 Hy$, satisfying $z^T z = (\varrho/\gamma)b$ and $b\beta_1 + a\beta_2 > 0$ (i.e. $y^T Ty > 0$), belongs to the Broyden class with*

$$\eta = b\frac{b\beta_1^2 - a(\gamma/\varrho)(\alpha_1\beta_1 - \alpha_2\beta_2)^2}{(b\beta_1 + a\beta_2)^2}, \qquad \eta' = b\frac{b(\alpha_2 - \varrho/\gamma)^2 - a(\varrho/\gamma)\alpha_1^2}{[b(\alpha_2 - \varrho/\gamma) + a\alpha_1]^2}, \tag{2.17}$$

where η' in the second formula corresponds to η in the special case, when vectors Ty, $(\varrho/\gamma)s - Sz$ are linearly dependent. Then we obtain $\eta = 1$ (the BFGS update) for $\alpha_1 = \beta_2 = 0$, $\alpha_2 = \pm\sqrt{(\varrho/\gamma)b/c}$, $\eta = 0$ (the DFP update) for $\alpha_2 = 0$, $\alpha_1 = \pm\sqrt{(\varrho/\gamma)b/a}$.

Proof. Since $Sz = \alpha_1 Hy + \alpha_2 s$ and $y^T Sz\, Ty - y^T Ty\, Sz = (\alpha_1 a + \alpha_2 b)(\beta_1 s + \beta_2 Hy) - (\beta_1 b + \beta_2 a)(\alpha_1 Hy + \alpha_2 s) = \omega(as - bHy)$, where $\omega = \alpha_1\beta_1 - \alpha_2\beta_2$, we can write (2.16) in the form

$$\frac{1}{\sqrt{\gamma}}S_+ = S - \frac{\beta_1 s + \beta_2 Hy}{\beta_1 b + \beta_2 a} y^T S + \left[\frac{s}{b} + \frac{(\gamma/\varrho)\omega}{\beta_1 b + \beta_2 a}\left(\frac{a}{b}s - Hy\right)\right]\left(\alpha_1 y^T + \alpha_2 s^T B\right) S \tag{2.18}$$

by (2.8). This gives $(1/\gamma)H_+ = H + \Delta H$, where matrix ΔH is expressed using only vectors s, Hy. Every such update, which satisfies quasi-Newton condition, belongs to the Broyden class. To determine η, it suffices to compare the terms, which contain $Hyy^T H$ (this term has coefficient $(\eta - 1)/a$ for the Broyden class). Since $(\varrho/\gamma)b = z^T z = (\alpha_1 y^T + \alpha_2 s^T B)SS^T(\alpha_1 y + \alpha_2 Bs)$ by (2.8), we obtain from (2.18)

$$\frac{\eta - 1}{a} = \frac{-2\beta_2 - 2(\gamma/\varrho)\omega\alpha_1}{\beta_1 b + \beta_2 a} + \frac{a\beta_2^2 + 2(\gamma/\varrho)\omega(\alpha_1 a + \alpha_2 b)\beta_2 + b(\gamma/\varrho)\omega^2}{(\beta_1 b + \beta_2 a)^2}.$$

Observing that

$$\frac{-\alpha_1}{\beta_1 b + \beta_2 a} + \frac{(\alpha_1 a + \alpha_2 b)\beta_2}{(\beta_1 b + \beta_2 a)^2} = \frac{-\alpha_1(\beta_1 b + \beta_2 a) + \beta_2(\alpha_1 a + \alpha_2 b)}{(\beta_1 b + \beta_2 a)^2} = \frac{-b\omega}{(\beta_1 b + \beta_2 a)^2},$$

we have $(\beta_1 b + \beta_2 a)^2(\eta - 1) = a[-2b\beta_1\beta_2 - a\beta_2^2 - b(\gamma/\varrho)\omega^2] = -(\beta_1 b + \beta_2 a)^2 + b^2\beta_1^2 - ab(\gamma/\varrho)\omega^2$, which gives the first equality in (2.17). Substituing e.g. $\beta_1 = (\varrho/\gamma) - \alpha_2$, $\beta_2 = -\alpha_1$, we get the second equality. The rest follows immediately by $z^T z = (\varrho/\gamma)b$. □

Note that Lemma 2.2 gives simple possibility, how to derive the product form of the Broyden-class update.

To choice parameters z, Ty by comparison with the standard Broyden class, we concentrate on the BFGS method, which we obtain e.g. for $z = \pm\sqrt{(\varrho/\gamma)b/c}\, S^T Bs$, $Ty = s$ by Lemma 2.2. Now we turn back to the shifted VM methods. By analogy with the BFGS method, we set

$$z = \vartheta U^T Bs, \qquad \vartheta = \pm\sqrt{(\varrho/\gamma)\tilde{b}/\bar{c}}. \tag{2.19}$$

Then (2.15) gives the type 1 method

$$\frac{1}{\sqrt{\gamma}}U_+ = U - \frac{(\varrho/\gamma)\tilde{s} - \vartheta ABs}{(\varrho/\gamma)\tilde{b} - \vartheta\bar{b}}\left(y - \vartheta Bs\right)^T U, \tag{2.20}$$

which gives the best results for the choice $\operatorname{sgn}\vartheta = -\operatorname{sgn}\bar{b}$ (compare with Theorem 2.11). Similarly, (2.10) leads to type 2 methods. With the simple choice $Ty = \tilde{s}$ we get

$$\frac{1}{\sqrt{\gamma}}U_+ = U - \frac{\tilde{s}}{\tilde{b}}y^T U + \left(\frac{\varrho}{\gamma}\frac{\tilde{s}}{\vartheta} - ABs + \frac{\bar{b}}{\tilde{b}}\tilde{s}\right)\frac{s^T BU}{\bar{c}}, \tag{2.21}$$

where ϑ is given by (2.19). The more general case, when Ty is a linear combination of vectors $\tilde{s}$, ABs and Ay, we will investigate in Section 2.6.

Note that neither update (2.20) nor (2.21) need not calculate vector Ay.

2.3 General expression of variationally-derived methods

General form of variationally-derived update (2.10) can be easily rewritten, using (2.8)

$$\frac{1}{\sqrt{\gamma}}U_+ = \frac{\tilde{s}z^T}{\tilde{b}} + \left(I - \frac{Tyy^T}{y^T Ty}\right)U\left(I - \frac{zz^T}{z^T z}\right). \tag{2.22}$$

Since $z^T(I - zz^T/z^T z) = 0$ and $(I - zz^T/z^T z)^2 = I - zz^T/z^T z$, this yields

$$\frac{1}{\gamma}A_+ = \frac{\varrho}{\gamma}\frac{\tilde{s}\tilde{s}^T}{\tilde{b}} + \left(I - \frac{Tyy^T}{y^T Ty}\right)U\left(I - \frac{zz^T}{z^T z}\right)U^T\left(I - \frac{yy^T T}{y^T Ty}\right) \tag{2.23}$$

by $A_+ = U_+ U_+^T$. This expression shows the meaning of parameters z, Ty.

In case $Ty = (\sqrt{\eta}/\tilde{b})\tilde{s} + (1/\bar{a} - \sqrt{\eta}/\bar{a})Ay$ we can easily compare update (2.23) with the shifted Broyden class update (2.3) in the following quasi-product form, see [9]

$$\frac{1}{\gamma}A_+^{sBC} = \frac{\varrho}{\gamma}\frac{\tilde{s}\tilde{s}^T}{\tilde{b}} + \left(I - \left(\frac{\sqrt{\eta}}{\tilde{b}}\tilde{s} + \frac{1-\sqrt{\eta}}{\bar{a}}Ay\right)y^T\right)A\left(I - y\left(\frac{\sqrt{\eta}}{\tilde{b}}\tilde{s} + \frac{1-\sqrt{\eta}}{\bar{a}}Ay\right)^T\right). \tag{2.24}$$

Denoting $\tilde{V} = I - Tyy^T/y^T Ty$ and observing that $U(I - zz^T/z^T z)U^T = A - Uzz^T U^T/z^T z$ and $y^T Ty = 1$, (2.23) and (2.24) give

$$\frac{1}{\gamma}A_+ = \frac{1}{\gamma}A_+^{sBC} - \frac{\tilde{V}Uz(\tilde{V}Uz)^T}{z^T z}. \tag{2.25}$$

For z chosen after (2.19) and $\eta = 1$, which represents method (2.21), update (2.23) can be easily rewritten in the form (2.38). We have $\tilde{V} = I - \tilde{s}y^T/\tilde{b}$, $Uz = \vartheta ABs$ and

$$\sqrt{\frac{\bar{c}}{z^T z}}\tilde{V}Uz = ABs - \frac{\bar{b}}{\tilde{b}}\tilde{s} = ABs - \frac{\bar{b}}{\bar{a}}Ay - \frac{\bar{b}}{\bar{a}}\left(\frac{\bar{a}}{\tilde{b}}\tilde{s} - Ay\right) = \frac{1}{\bar{a}}\left(v_2 - \frac{\bar{b}}{\sqrt{\delta}}\hat{w}\right),$$

where we denoted $v_2 = \bar{a}ABs - \bar{b}Ay$ and $\hat{w} = \sqrt{\eta\delta}\left((\bar{a}/\tilde{b})\tilde{s} - Ay\right)$ as in [9]. Thus it follows from (2.3) and (2.25) that

$$\frac{1}{\gamma}A_+ = A + \frac{\varrho}{\gamma}\frac{\tilde{s}\tilde{s}^T}{\tilde{b}} - \frac{Ayy^T A}{\bar{a}} + \frac{\hat{w}\hat{w}^T - uu^T}{\bar{a}\delta}, \quad u = \sqrt{1 - \frac{\bar{b}^2}{\bar{a}\bar{c}}}\,v_2 - \frac{\bar{b}}{\sqrt{\bar{a}\bar{c}}}\hat{w}. \tag{2.26}$$

The term $\hat{w}\hat{w}^T - uu^T$ can be rewritten

$$\hat{w}\hat{w}^T - uu^T = \frac{\bar{\delta}}{\bar{a}\bar{c}}\hat{w}\hat{w}^T + \frac{\bar{b}\sqrt{\delta}}{\bar{a}\bar{c}}\left(v_2\hat{w}^T + \hat{w}v_2^T\right) + \frac{\bar{b}^2}{\bar{a}\bar{c}}v_2v_2^T - v_2v_2^T,$$

which finally gives by (2.26)

$$\frac{1}{\gamma}A_+ = A + \frac{\varrho}{\gamma}\frac{\tilde{s}\tilde{s}^T}{\tilde{b}} - \frac{Ayy^TA}{\bar{a}} + \frac{q_2q_2^T - v_2v_2^T}{\bar{a}\bar{\delta}}, \quad q_2 = \sqrt{1 - \frac{\bar{b}^2}{\bar{a}\bar{c}}}\,\hat{w} + \frac{\bar{b}}{\sqrt{\bar{a}\bar{c}}}v_2. \tag{2.27}$$

2.4 Balanced variationally-derived shifted VM methods

Efficiency of the variationally-derived shifted VM methods can be significantly improved by using the modified quasi-Newton condition

$$A_+y = \varrho\tilde{\varrho}\tilde{s}, \quad \zeta_+ = \varrho\sigma, \tag{2.28}$$

where $\tilde{\varrho} > 0$ is a correction parameter. Very good results were obtained with the following choices of $\tilde{\varrho}$: $\nu \triangleq \mu/(1-\mu)$, $\varepsilon \triangleq \sqrt{\zeta\hat{a}/a}$ and $\sqrt{\nu\varepsilon}$, where $\mu = \sigma\hat{a}/b \in (0,1)$ is a relative shift parameter and ε is the damping factor of μ, see [9]. The first choice might be explained by the following assertion.

Lemma 2.3. *For the choice $\tilde{\varrho} = \mu/(1-\mu)$, equality $y^TA_+y/y^Ty = \zeta_+$ holds, i.e. this value of $\tilde{\varrho}$ balances the both parts of $y^TH_+y = \zeta_+\,y^Ty + y^TA_+y$.*

Proof. By $\tilde{b} = \tilde{s}^Ty = b - \sigma\hat{a} = b(1-\mu)$, we have from (2.28)

$$\frac{y^TA_+y}{y^Ty} = \frac{\varrho\tilde{\varrho}\tilde{b}}{\hat{a}} = \frac{\varrho\mu b}{\hat{a}} = \varrho\sigma = \zeta_+. \qquad \square$$

2.5 Global convergence of the shifted Broyden class methods

Global convergence is defined by the relation

$$\liminf_{k\to\infty} |g_k| = 0. \tag{2.29}$$

First we recall the basic assumptions and assertions from [9] (Assumption 2.3 is new).

Theorem 2.4. *Let the objective function $f : \mathcal{R}^N \to \mathcal{R}$ be bounded from below and have bounded second derivatives. Consider the line search method satisfying (1.1)-(1.2). If*

$$\sum_{k=1}^{\infty} cos^2\theta_k \triangleq \sum_{k=1}^{\infty} \frac{(g_k^TH_kg_k)^2}{g_k^Tg_k\,g_k^TH_k^2g_k} = \infty, \tag{2.30}$$

then (2.29) holds.

Assumption 2.1. *The objective function $f : \mathcal{R}^N \to \mathcal{R}$ is uniformly convex and has bounded second-order derivatives (i.e. $0 < \underline{G} \le \underline{\lambda}(G(x)) \le \overline{\lambda}(G(x)) \le \overline{G} < \infty$, $x \in \mathcal{R}^N$, where $\underline{\lambda}(G(x))$ and $\overline{\lambda}(G(x))$ are the lowest and the greatest eigenvalues of the Hessian matrix $G(x)$).*

Assumption 2.2. *Parameters* ϱ_k *and* μ_k *of the shifted VM method are uniformly positive and bounded, in the sense that* $0 < \underline{\varrho} \le \varrho_k \le \overline{\varrho}$ *and* $0 < \underline{\mu} \le \mu_k \le \overline{\mu} < 1$, $k \ge 1$.

Assumption 2.3. *Parameters* γ_k *and* μ_k *of the shifted VM method satisfy* $0 < \underline{\gamma} \le \gamma_k \le \overline{\gamma}$, $\mu_k^2 \le \zeta_k \hat{a}_k / a_k$, $k \ge 1$.

Lemma 2.4. *Let* $s \ne 0$, *the objective function satisfy Assumption 2.1 and parameter* μ *satisfy Assumption 2.2. Then* $y \ne 0$, $\tilde{s} \ne 0$, $b > 0$, $\tilde{b} > 0$, $\hat{a}/b \in [\underline{G}, \overline{G}]$ *and* $b/|s|^2 \ge \underline{G}$.

Theorem 2.5. *Consider any shifted variable metric method satisfying (2.1)-(2.2) and Assumption 2.2, with the line search method fulfilling (1.1)-(1.2). Let the objective function satisfy Assumption 2.1. If there is a constant* $0 < C < \infty$ *such that*

$$\mathrm{Tr} A_{k+1} \le \mathrm{Tr} A_k + C, \quad k \ge 1, \tag{2.31}$$

then (2.29) holds.

Theorem 2.6. *Consider the shifted variable metric method (2.3) satisfying Assumption 2.2 and* $\gamma_k \le 1$, $k \ge 1$, *with the line search method fulfilling (1.1)-(1.2). Let the objective function satisfy Assumption 2.1. If there is a constant* $0 \le C < \infty$ *such that*

$$\eta_k \left| \frac{\bar{a}_k}{\tilde{b}_k} \tilde{s}_k - A_k y_k \right|^2 \le C \frac{\bar{a}_k}{\tilde{b}_k} |\tilde{s}_k|^2 + |A_k y_k|^2, \quad k \ge 1, \tag{2.32}$$

then (2.29) holds.

In this section, we extend the set of methods from the shifted Broyden class (2.3), for which global convergence can be established. First result is based on the following lemma.

Lemma 2.5. *Let* $\eta < 1$. *Then*

$$\eta \left| \frac{\bar{a}}{\tilde{b}} \tilde{s} - Ay \right|^2 \le \frac{\eta}{1-\eta} \frac{\bar{a}^2}{\tilde{b}^2} |\tilde{s}|^2 + |Ay|^2. \tag{2.33}$$

Proof. The desired inequality follows from the identity

$$\eta \left| \frac{\bar{a}}{\tilde{b}} \tilde{s} - Ay \right|^2 - \frac{\eta}{1-\eta} \frac{\bar{a}^2}{\tilde{b}^2} |\tilde{s}|^2 - |Ay|^2 = (\eta - 1) \left| \frac{\eta}{1-\eta} \frac{\bar{a}}{\tilde{b}} \tilde{s} + Ay \right|^2. \qquad \square$$

Theorem 2.7. *Consider the shifted variable metric method (2.3) satisfying Assumption 2.2 and* $0 < \underline{\gamma} \le \gamma_k \le 1$, $k \ge 1$, *with the line search method fulfilling (1.1)-(1.2). Let the objective function satisfy Assumption 2.1. If* $\eta/(1-\eta)\, \bar{a}/\tilde{b} \le C$ *with any constant* $0 < C < \infty$ *(e.g.* $\eta = (\varrho/\gamma)/(\varrho/\gamma + \bar{a}/\tilde{b})$, *which corresponds to the shifted analogy of Hoshino self-dual method, see[5]), (2.29) holds.*

Proof. If $\eta/(1-\eta)\, \bar{a}/\tilde{b} \le C$, we can use Lemma 2.5 and Theorem 2.6. In case $\eta = (\varrho/\gamma)/(\varrho/\gamma + \bar{a}/\tilde{b})$ we obtain $\eta/(1-\eta)\, \bar{a}/\tilde{b} = \varrho/\gamma \le \overline{\varrho}/\underline{\gamma}$. $\square$

Now, denoting $\tilde{H}_+ = \gamma \zeta I + A_+$, we establish global convergence of methods from the shifted Broyden class for $\eta \le 1$ and $\mu^2 \le \zeta \hat{a}/a$. The following lemma plays basic role.

Lemma 2.6. *Consider the shifted variable metric method (2.3) with $\eta \le 1$. Then*

$$\frac{\det(\frac{1}{\gamma}\tilde{H}_+)}{\det H} \le \frac{\tilde{s}^T B \tilde{s}}{\tilde{b}} \left(\frac{\varrho}{\gamma} + \frac{\zeta \hat{a}}{\tilde{b}} \right). \tag{2.34}$$

Proof. It suffices to prove the desired inequality for $\eta = 1$ by (2.3) and the identity $\det(\tilde{H}_+ - uu^T) = \det \tilde{H}_+(1 - u^T \tilde{H}_+^{-1} u)$. Since the shifted BFGS update can be written in the form

$$\frac{1}{\gamma}\tilde{H}_+ = H^{1/2} \left(I + \frac{B^{1/2}(\omega\tilde{s} - Ay)(\omega\tilde{s} - Ay)^T B^{1/2} - B^{1/2} A y y^T A B^{1/2}}{\tilde{b}\,\omega} \right) H^{1/2},$$

where $\omega = \varrho/\gamma + \bar{a}/\tilde{b}$, and since

$$\begin{aligned} \det(I + (u-v)(u-v)^T - vv^T) &= (1 + |u-v|^2)(1 - |v|^2) + ((u-v)^T v)^2 \\ &= |u|^2 + (1 - u^T v)^2 - |u|^2 |v|^2, \end{aligned}$$

we obtain

$$\frac{\det(\frac{1}{\gamma}\tilde{H}_+)}{\det H} = \omega \frac{\tilde{s}^T B \tilde{s}}{\tilde{b}} + \left(1 - \frac{\tilde{s}^T B A y}{\tilde{b}} \right)^2 - \frac{\tilde{s}^T B \tilde{s} \cdot y^T A B A y}{\tilde{b}^2}.$$

Observing that $\tilde{s}^T B A y = \tilde{b} - \zeta \tilde{s}^T B y$ and $y^T A B A y = \bar{a} - \zeta \hat{a} + \zeta^2 y^T B y$, we find

$$\begin{aligned} \frac{\det(\frac{1}{\gamma}\tilde{H}_+)}{\det H} &= \omega \frac{\tilde{s}^T B \tilde{s}}{\tilde{b}} + \frac{\zeta^2 (\tilde{s}^T B y)^2}{\tilde{b}^2} - \frac{\tilde{s}^T B \tilde{s} \cdot y^T A B A y}{\tilde{b}^2} \\ &= \left(\frac{\varrho}{\gamma} + \frac{\zeta \hat{a}}{\tilde{b}} \right) \frac{\tilde{s}^T B \tilde{s}}{\tilde{b}} + \zeta^2 \frac{(\tilde{s}^T B y)^2 - \tilde{s}^T B \tilde{s} \cdot y^T B y}{\tilde{b}^2} \le \left(\frac{\varrho}{\gamma} + \frac{\zeta \hat{a}}{\tilde{b}} \right) \frac{\tilde{s}^T B \tilde{s}}{\tilde{b}} \end{aligned}$$

by the Schwarz inequality. □

Lemma 2.7. *Consider any shifted variable metric method satisfying (2.1)-(2.2). Then*

$$\frac{\det H_+}{\det \tilde{H}_+} < \left(1 + \frac{\varrho}{\gamma} \frac{b}{\zeta \hat{a}} \right)^N. \tag{2.35}$$

Proof. We have $H_+ = \tilde{H}_+ + \xi I$, where $\xi = \zeta_+ - \gamma\zeta = \varrho\,\sigma - \gamma\zeta$. Denoting $\tilde{\lambda}_1, \ldots, \tilde{\lambda}_N$ the eigenvalues of $\tilde{H}_+$, we can apply the geometric/aritmetic mean inequality to obtain

$$\begin{aligned} \frac{\det H_+}{\det \tilde{H}_+} &= \left(1 + \frac{\xi}{\tilde{\lambda}_1} \right) \cdots \left(1 + \frac{\xi}{\tilde{\lambda}_N} \right) \le \left[1 + \frac{\xi}{N} \left(\frac{1}{\tilde{\lambda}_1} + \cdots + \frac{1}{\tilde{\lambda}_N} \right) \right]^N \\ &= \left(1 + \frac{\xi}{N} \mathrm{Tr}\, \tilde{H}_+^{-1} \right)^N \le \left(1 + \xi \|\tilde{H}_+^{-1}\| \right)^N < \left(1 + \frac{\varrho}{\gamma} \frac{b}{\zeta \hat{a}} \right)^N \end{aligned}$$

by $\xi < \varrho\,\sigma = \varrho\,\mu\, b/\hat{a} < \varrho\, b/\hat{a}$ and $\|\tilde{H}_+^{-1}\| \le 1/(\gamma\zeta)$ in view of $\tilde{H}_+ = \gamma\zeta I + A_+$. □

Lemma 2.8. *Consider any shifted variable metric method satisfying (2.1)-(2.2), with $\mu^2 \le \zeta \hat{a}/a$. Then*

$$\tilde{s}^T B \tilde{s} \le 4c. \tag{2.36}$$

Proof. Assumption $\mu^2 \le \zeta\hat{a}/a$ implies $\sigma^2 = \mu^2(b/\hat{a})^2 \le \zeta b^2/(\hat{a}a) \le \zeta c/\hat{a}$ by the Schwarz inequality. Observing that $\zeta\, y^T By/y^T y \le \zeta\|B\| \le 1$ by (2.1), we obtain $\sigma^2 y^T By \le c\,\zeta\, y^T By/\hat{a} \le c$. Since $\tilde{s} = s - \sigma y$, we get

$$\tilde{s}^T B\tilde{s} = c - 2\sigma s^T By + \sigma^2 y^T By \le c + 2\sigma\sqrt{cy^T By} + \sigma^2 y^T By \le c + 2c + c = 4c. \qquad \square$$

Theorem 2.8. *Consider the shifted variable metric method (2.3) satisfying Assumption 2.2 and Assumption 2.3, with the line search method fulfilling (1.1)-(1.2). Let the objective function satisfy Assumption 2.1. Then for every $\eta \in [0,1]$ (2.29) holds.*

Proof. Since $\hat{a}/b \in [\underline{G}, \overline{G}]$ by Lemma 2.4 and $\zeta_+ = \varrho\,\sigma = \varrho\mu b/\hat{a}$ by (2.2), we deduce $\zeta_+ \in [\underline{\zeta}, \overline{\zeta}]$, where $\underline{\zeta} = \underline{\mu}\,\underline{\varrho}/\overline{G}$, $\overline{\zeta} = \overline{\mu}\,\overline{\varrho}/\underline{G}$. Combining Lemma 2.6, Lemma 2.7 and Lemma 2.8, we find

$$\frac{\det H_+}{\det H} < \frac{4c}{\tilde{b}}\left(\frac{\varrho}{\gamma} + \frac{\zeta\hat{a}}{\tilde{b}}\right)\left(\gamma + \varrho\frac{b}{\zeta\hat{a}}\right)^N \le \frac{4c\left(\overline{\varrho}/\underline{\gamma} + \overline{\zeta}\,\overline{G}\right)}{b(1-\overline{\mu})^2}\left(\overline{\gamma} + \frac{\overline{\varrho}}{\underline{\zeta}\,\underline{G}}\right) \triangleq C_2\frac{c}{b}. \qquad (2.37)$$

Observing that $\det H \ge \zeta^N \ge \underline{\zeta}^N$ by (2.1), we get by (2.37)

$$C_1 \triangleq \frac{\underline{\zeta}^N}{\det H_2} \le \frac{\det H_{k+2}}{\det H_2} = \prod_{i=2}^{k+1}\frac{\det H_{i+1}}{\det H_i} < C_2^k \prod_{i=2}^{k+1}\frac{c_i}{b_i},$$

$k \ge 1$, which yields $\sum_{i=2}^{k+1} c_i/b_i \ge k(\prod_{i=2}^{k+1} c_i/b_i)^{1/k} > kC_1^{1/k}/C_2$ with $C_1 > 0$, $C_2 > 0$. Now we show that (2.30) holds. From (1.2) and $g^T Hg \ge \zeta g^T g$ by (2.1) we obtain

$$\sum_{i=2}^{k+1} cos^2\theta_i = \sum_{i=2}^{k+1}\frac{(g_i^T H_i g_i)^2}{g_i^T g_i\, g_i^T H_i^2 g_i} = \sum_{i=2}^{k+1}\frac{g_i^T H_i g_i}{g_i^T g_i}\,\frac{b_i}{s_i^T s_i}\,\frac{c_i}{b_i} \ge \sum_{i=2}^{k+1}\zeta_i\underline{G}\,\frac{c_i}{b_i} > k\frac{\underline{\zeta}\,\underline{G}C_1^{1/k}}{C_2},$$

$k \ge 1$, by Lemma 2.4. Thus $\sum_{i=1}^{\infty} cos^2\theta_i = \infty$ and (2.29) follows from Theorem 2.4. $\square$

Note that assumption $\mu^2 \le \zeta\hat{a}/a$ gives reasons for the choice of damping coefficient $\varepsilon = \sqrt{\zeta\hat{a}/a}$ for the shift parameter μ (see [9], recall that we require $\varepsilon = 1$ for $\bar{a} = 0$).

All assertions here can also be proved, if we use the modified quasi-Newton condition (2.28) with correction parameter $\tilde{\varrho}$ satisfying $\varrho\tilde{\varrho} \le \overline{\varrho}$.

2.6 Global convergence of limited-memory methods

In this section we denote $v_1 = \bar{c}Ay - \bar{b}ABs$, $v_2 = \bar{a}ABs - \bar{b}Ay$, $q_1 = \bar{\delta}p_1 + v_1$, $q_2 = \bar{\delta}p_2 + v_2$ and $\hat{w} = \sqrt{\eta\bar{\delta}}\left((\bar{a}/\tilde{b})\tilde{s} - Ay\right)$ as in [9]. First we recall the following general forms of type 2 update formula (4.24) and (4.25) in [9] for $\bar{\delta} \ne 0$ (which implies $\bar{a}\bar{c} \ne 0$ by the Schwarz inequality)

$$\begin{aligned} \frac{1}{\gamma}A_+ &= A + \frac{\varrho}{\gamma}\frac{\tilde{s}\tilde{s}^T}{\tilde{b}} - \frac{Ayy^TA}{\bar{a}} + \frac{q_2q_2^T - v_2v_2^T}{\bar{a}\bar{\delta}}, \qquad q_2^Ty = 0, \qquad (2.38)\\ &= A + \frac{\varrho}{\gamma}\frac{\tilde{s}\tilde{s}^T}{\tilde{b}} - \frac{ABss^TBA}{\bar{c}} + \frac{q_1q_1^T - v_1v_1^T}{\bar{c}\bar{\delta}}, \qquad q_1^Ty = 0 \end{aligned}$$

(note that in case $\bar{\delta} = 0$ the choice of q_2 or q_1 is irrelevant, see [9], and we can use e.g. the shifted DFP method, which does not violate global convergence by Theorem 2.7). Using identity $\bar{a}(v_1 v_1^T + \bar{\delta} ABss^T BA) = \bar{c}(v_2 v_2^T + \bar{\delta} Ayy^T A)$, the last form can be written

$$\frac{1}{\gamma} A_+ = A + \frac{\varrho}{\gamma} \frac{\tilde{s}\tilde{s}^T}{\tilde{b}} - \frac{Ayy^T A}{\bar{a}} + \frac{q_1 q_1^T}{\bar{c}\bar{\delta}} - \frac{v_2 v_2^T}{\bar{a}\bar{\delta}}, \qquad q_1^T y = 0, \tag{2.39}$$

which is (2.38) with $q_2 q_2^T/\bar{a}$ replaced by $q_1 q_1^T/\bar{c}$. First we prove the following basic assertion.

Theorem 2.9. *Let $\bar{\delta}_k \neq 0$, $k \geq 1$. Consider the shifted variable metric method (2.38) with $q_2 = \alpha\hat{w} + \beta v_2$ (or (2.39) with $q_1\sqrt{\bar{a}/\bar{c}} = \alpha\hat{w} + \beta v_2$), satisfying Assumption 2.2 and Assumption 2.3, with the line search method fulfilling (1.1)-(1.2). Let the objective function satisfy Assumption 2.1. If $\alpha^2 + \beta^2 \leq 1$ and $\eta \in [0, 1]$, then (2.29) holds.*

Proof. Obviously, we can restrict to update (2.38). Assumption $q_2 = \alpha\hat{w} + \beta v_2$ yields

$$q_2 q_2^T - v_2 v_2^T = \alpha^2 \hat{w}\hat{w}^T + \alpha\beta\hat{w} v_2^T + \alpha\beta v_2 \hat{w}^T + (\beta^2 - 1) v_2 v_2^T. \tag{2.40}$$

If $\beta^2 = 1$, condition $\alpha^2 + \beta^2 \leq 1$ implies $\alpha = 0$ and (2.38) represents the shifted DFP method, which is globally convergent by Theorem 2.8. If $\beta^2 < 1$, we can write by (2.40)

$$\frac{q_2 q_2^T - v_2 v_2^T}{\bar{a}\bar{\delta}} = \frac{\eta'}{\bar{a}} \left(\frac{\bar{a}}{\tilde{b}} \tilde{s} - Ay \right) \left(\frac{\bar{a}}{\tilde{b}} \tilde{s} - Ay \right)^T - uu^T, \quad u = \frac{(1 - \beta^2) v_2 - \alpha\beta\hat{w}}{\sqrt{\bar{a}\bar{\delta}(1 - \beta^2)}},$$

where $\eta' = \eta\alpha^2/(1 - \beta^2) \leq \eta \leq 1$. Thus (2.38) represents update (2.3) with adding term $-uu^T$. Without this adding term, this update satisfies assumptions of Lemma 2.6. Therefore, in view of identity $\det(\tilde{H}_+ - uu^T) = \det \tilde{H}_+ (1 - u^T \tilde{H}_+^{-1} u)$, inequality (2.34) holds and the desired result follows as in the proof of Theorem 2.8. □

Corollary 2.1. *Let the objective function satisfy Assumption 2.1. For the shifted variable metric methods SSBC, NSBC and DSBC described in [9] (method DSBC is denoted DVSBC in [9]), satisfying Assumption 2.2 and Assumption 2.3, with $\eta \in [0, 1]$ and with the line search method fulfilling (1.1)-(1.2), (2.29) holds.*

Proof. It follows from Theorem 4.2 in [9] that $\alpha^2 + \beta^2 = 1$ for all three these methods and we use Theorem 2.9. □

Now we concentrate on update (2.10) with the choice (2.19), which is type 2 method with $p_1 = -Ty/y^T Ty$. Thus $p_1^T y = -1$, yielding $q_1^T y = -\bar{\delta} + v_1^T y = 0$. Therefore we can express this update in the form (2.39) (see [9]) and the following theorem enables us to derive its global convergence from update (2.3).

Theorem 2.10. *Let $\eta > 0$. Consider update (2.10) with the choice (2.19) and with*

$$Ty = \tilde{s} + \beta_1 ABs + \beta_2 Ay. \tag{2.41}$$

If

$$(\bar{a}\beta_2 + \tilde{b})^2 \geq \bar{a}\bar{c}\beta_1^2 + \tilde{b}^2/\eta \tag{2.42}$$

holds, then the assumption $\alpha^2 + \beta^2 \leq 1$ of Theorem 2.9 is satisfied.

Proof. From $p_1 = -Ty/y^TTy$ and (2.41) we obtain

$$q_1 = \bar\delta p_1 + v_1 = -\bar\delta\,\frac{\tilde s - (\tilde b/\bar a)Ay + \beta_1 ABs + (\beta_2 + \tilde b/\bar a)Ay}{\tilde b + \bar b\beta_1 + \bar a\beta_2} + \bar c Ay - \bar b ABs$$
$$= \frac{-\bar\delta\tilde b/\bar a}{\tilde b + \beta_1\bar b + \beta_2\bar a}\left(\frac{\bar a}{\tilde b}\tilde s - Ay\right) + \left(\frac{-\bar\delta\beta_1}{\tilde b + \bar b\beta_1 + \bar a\beta_2} - \bar b\right)ABs - \left(\frac{\bar\delta(\beta_2 + \tilde b/\bar a)}{\tilde b + \bar b\beta_1 + \bar a\beta_2} - \bar c\right)Ay$$
$$= \frac{-\tilde b\sqrt{\bar\delta}}{\bar a\sqrt{\eta}(\tilde b + \bar b\beta_1 + \bar a\beta_2)}\hat w - \frac{\bar b\tilde b/\bar a + \bar c\beta_1 + \bar b\beta_2}{\tilde b + \bar b\beta_1 + \bar a\beta_2}v_2 \triangleq \sqrt{\frac{\bar c}{\bar a}}(\alpha\hat w + \beta v_2),$$

using identities

$$\begin{aligned}
\bar\delta\beta_1 + \bar b(\tilde b + \bar b\beta_1 + \bar a\beta_2) &= \left(\bar b\tilde b/\bar a + \bar c\beta_1 + \bar b\beta_2\right)\bar a,\\
-\bar\delta(\beta_2 + \tilde b/\bar a) + \bar c(\tilde b + \bar b\beta_1 + \bar a\beta_2) &= \left(\bar b\tilde b/\bar a + \bar c\beta_1 + \bar b\beta_2\right)\bar b.
\end{aligned}$$

Thus we have

$$\begin{aligned}
\alpha^2 + \beta^2 &= \frac{\bar\delta\tilde b^2/\eta + \left[\bar b(\bar a\beta_2 + \tilde b) + \bar a\bar c\beta_1\right]^2}{\bar a\bar c(\bar a\beta_2 + \tilde b + \bar b\beta_1)^2}\\
&= \frac{\bar\delta\tilde b^2/\eta + \bar b^2(\bar a\beta_2 + \tilde b)^2 + 2\bar a\bar b\bar c\beta_1(\bar a\beta_2 + \tilde b) + \bar a^2\bar c^2\beta_1^2}{\bar a\bar c(\bar a\beta_2 + \tilde b)^2 + 2\bar a\bar b\bar c\beta_1(\bar a\beta_2 + \tilde b) + \bar a\bar b^2\bar c\beta_1^2}\\
&= 1 - \bar\delta\,\frac{(\bar a\beta_2 + \tilde b)^2 - \bar a\bar c\beta_1^2 - \tilde b^2/\eta}{\bar a\bar c(\bar a\beta_2 + \tilde b + \bar b\beta_1)^2} \le 1
\end{aligned}$$

by (2.42) and $\bar\delta \ge 0$. □

Corollary 2.2. *Consider the shifted variable metric method (2.21) satisfying Assumption 2.2 and Assumption 2.3, with the line search method fulfilling (1.1)-(1.2). Let the objective function satisfy Assumption 2.1. Then (2.29) holds.*

Proof. Choosing $\beta_1 = \beta_2 = 0$ in (2.41), (2.42) gives $\eta \ge 1$ and it suffices to use Theorem 2.9 with $\eta = 1$. □

This approach cannot be used for method (2.20), which uses $\beta_2 = 0$ and $\beta_1^2 = (\gamma/\varrho)\tilde b/\bar c$ by (2.19). Then condition (2.42) has the form $\bar a\gamma/\varrho \le \tilde b(1 - 1/\eta)$, which cannot be satisfied in general. Fortunately, similar assertion as Lemma 2.6 holds. Denote again $\tilde H_+ = \gamma\zeta I + A_+$.

Lemma 2.9. *Consider the shifted variable metric method (2.20) in the form*

$$(1/\sqrt{\gamma})U_+ = U - p\,q^TU, \quad p = \tilde s - \vartheta(\gamma/\varrho)ABs, \quad q = (y - \vartheta Bs)/p^Ty, \tag{2.43}$$

with $\vartheta^2 \le (\varrho/\gamma)\tilde b/\bar c$ and $\vartheta\bar b \le 0$. Then

$$\frac{\det(\frac{1}{\gamma}\tilde H_+)}{\det H} \le \frac{p^TBp}{\tilde b}\left(\frac{\varrho}{\gamma} + \frac{\zeta\hat a}{\tilde b}\right). \tag{2.44}$$

Proof. Update (2.43) can be written $(1/\gamma)A_+ = A - Aqp^T - pq^TA + q^TAq\,pp^T$, or

$$\frac{1}{\gamma}\tilde{H}_+ = H^{1/2}\left(I + \frac{B^{1/2}(q^TAq\,p - Aq)(q^TAq\,p - Aq)^TB^{1/2} - B^{1/2}Aqq^TAB^{1/2}}{q^TAq}\right)H^{1/2}.$$

Since $\det(I + (u-v)(u-v)^T - vv^T) = |u|^2 + (1 - u^Tv)^2 - |u|^2|v|^2$ (see the proof of Lemma 2.6), we obtain

$$\det(\tfrac{1}{\gamma}\tilde{H}_+)/\det H = q^TAq \cdot p^TBp + (1 - p^TBAq)^2 - p^TBp \cdot q^TABAq.$$

Observing that $q^TABAq = q^TAq - \zeta q^Tq + \zeta^2 q^TBq$ and $1 - p^TBAq = 1 - p^Tq + \zeta p^TBq = (\vartheta/p^Ty)p^TBs + \zeta p^TBq$, we find by the Schwarz inequality

$$\begin{aligned}
\frac{\det(\frac{1}{\gamma}\tilde{H}_+)}{\det H} &= p^TBp\left[\zeta q^Tq - \zeta^2 q^TBq\right] + \left[p^TB\Big((\vartheta/p^Ty)s + \zeta q\Big)\right]^2 \\
&\le p^TBp\left[\zeta q^Tq - \zeta^2 q^TBq + \Big((\vartheta/p^Ty)s + \zeta q\Big)^T B\Big((\vartheta/p^Ty)s + \zeta q\Big)\right] \\
&= \frac{p^TBp}{(p^Ty)^2}\left[\zeta|y - \vartheta Bs|^2 + \vartheta^2 c + 2\zeta\vartheta s^TB(y - \vartheta Bs)\right] \\
&= \frac{p^TBp}{(p^Ty)^2}\Big(\zeta\hat{a} + \vartheta^2 c - \zeta\vartheta^2|Bs|^2\Big) = \frac{p^TBp}{[\tilde{b} - \vartheta(\gamma/\varrho)\bar{b}]^2}\Big(\zeta\hat{a} + \vartheta^2\bar{c}\Big) \\
&\le \frac{p^TBp}{\tilde{b}^2}\Big(\zeta\hat{a} + \vartheta^2\bar{c}\Big) \le \frac{p^TBp}{\tilde{b}^2}\left(\zeta\hat{a} + \frac{\varrho}{\gamma}\tilde{b}\right)
\end{aligned}$$

and by assumptions. □

Lemma 2.10. *Consider the shifted variable metric method (2.43), satisfying Assumption 2.2 and Assumption 2.3, with $|\vartheta| \le \tilde{C}$ for some $0 < \tilde{C} < \infty$. Then*

$$p^TBp \le 2c\,[\,4 + (\tilde{C}\,\overline{\gamma}/\underline{\varrho})^2\,]. \tag{2.45}$$

Proof. Observing that $\zeta s^TB^3s/s^TB^2s \le \zeta\|B\| \le 1$, we get $s^TBABABs = c - 2\zeta s^TB^2s + \zeta^2 s^TB^3s \le c - \zeta s^TB^2s \le c$. Using Lemma 2.8, we obtain

$$p^TBp = |B^{1/2}(\tilde{s} - (\vartheta\gamma/\varrho)ABs)|^2 \le 2[\tilde{s}^TB\tilde{s} + (\vartheta\gamma/\varrho)^2c] \le 2c\,[4 + (\tilde{C}\,\overline{\gamma}/\underline{\varrho})^2\,]. \qquad \square$$

Theorem 2.11. *Consider the shifted variable metric method (2.20) satisfying Assumption 2.2 and Assumption 2.3, with the line search method fulfilling (1.1)-(1.2) and with*

$$\vartheta_k = -\mathrm{sgn}\bar{b}_k \min\left[\tilde{C}, \sqrt{(\varrho_k/\gamma_k)\tilde{b}_k/\bar{c}_k}\,\right], \quad k \ge 1, \tag{2.46}$$

for some $0 < \tilde{C} < \infty$. If the objective function satisfy Assumption 2.1, then (2.29) holds.

Proof. Using Lemma 2.9, Lemma 2.7 and Lemma 2.10, we can proceed in the same way as in the proof of Theorem 2.8. □

Note that all assertions here can also be proved, if we use the modified quasi-Newton condition (2.28) with correction parameter $\tilde{\varrho}$ satisfying $\varrho\tilde{\varrho} \le \overline{\varrho}$.

3 Computational experiments

Our new limited-memory VM methods were thoroughly tested, using the collection of relatively difficult problems with optional dimension chosen from [7] (Test 28, many of problems are dense), collection of problems for large-scale nonlinear least squares from [6] (Test 15, sparse but usually ill-conditioned problems) and collection of problems for general sparse and partially separable unconstrained optimization from [6] (Test 14, usually well-conditioned problems). We have used $m = 10$, 20 for $N = 1000$ and $m = 5$, 10 for $N = 5000$, the final precision $|g(x^\star)| \leq 10^{-6}$, $\eta = 1$ for the corresponding shifted Broyden class (methods SSBC, NSBC and DSBC, see [9]) and the choice of the shift parameter μ after [9]. For starting iterates we use the shifted BFGS method.

Results of our experiments are given in four tables, where NIT is the total number of iterations (over all problems), NFV the total number of function and also gradient evaluations, 'Fail' denotes the number of problems which were not solved successfully (usually NFV reached its limit) and 'Time' is the total computational time.

	$m = 10$				$m = 20$			
Method	NIT	NFV	Fail	Time	NIT	NFV	Fail	Time
SSBC	85211	91287	-	8:58.7	92589	95836	1	9:38.5
NSBC	100550	104347	1	9:57.3	105011	122247	3	11:05.3
DSBC	101139	103781	-	9:21.6	92285	94755	-	9:21.3
VAR1	91225	94406	-	10:26.1	95486	98252	1	10:26.1
VAR2	83047	86385	-	8:31.4	88540	91413	-	8:49.8
NS	85750	91533	-	7:45.3	84246	89349	-	8:15.2
BNS	87850	102109	-	7:11.9	89587	112245	2	9:40.9
RH	83232	101884	-	7:11.6	90466	110183	-	8:11.9
CGM	108929	222722	-	15:51.2				

Table 1 (Test 28, $N = 1000$, 80 problems)

	$m = 10$				$m = 20$			
Method	NIT	NFV	Fail	Time	NIT	NFV	Fail	Time
SSBC	40976	45574	-	50.86	39791	41254	-	60.67
NSBC	52045	53736	1	70.06	53131	68160	2	87.27
DSBC	48089	49222	-	61.45	47766	48698	-	72.99
VAR1	49120	50831	-	58.81	47591	48670	-	67.86
VAR2	48103	49746	-	55.47	45071	46106	-	61.05
NS	33765	36443	-	36.86	34994	37527	-	65.66
BNS	36485	55576	2	54.68	44381	76519	4	121.30
RH	36122	42187	-	40.80	32649	41093	1	46.05
CGM	36472	75466	-	46.45				

Table 2 (Test 15, $N = 1000$, 22 problems)

The first three rows of tables give results for various methods described in [9]: SSBC – the simple method based on the shifted Broyden class with $q_2 = \hat{w}$, NSBC – the method nearest to the shifted Broyden class and DSBC – the method with direction vector after the shifted Broyden class. Then results for the new variationally-derived limited memory methods follows: VAR1 – method (2.20) and VAR 2 – method (2.21) with $\tilde{\varrho} = \sqrt{\nu\varepsilon}$ in Table 1 and Table 2 and $\tilde{\varrho} = \nu$ in Table 3 and Table 4, see Section 2.4.

For comparison, the last four rows contain results for the following limited-memory methods: NS – the Nocedal method based on the Strang formula, see [8], BNS – the method after [1], RH – the reduced-Hessian method described in [4] and CGM – the conjugate gradient method (Hestenes and Stiefel version), see [3]. Note that methods BNS and NS store m pairs of vectors and method CGM stores no additional vectors.

	$m = 10$				$m = 20$			
Method	NIT	NFV	Fail	Time	NIT	NFV	Fail	Time
SSBC	20838	21070	-	16.64	17651	17872	-	17.74
NSBC	22585	23059	-	19.91	20125	20531	-	21.99
DSBC	21801	22041	-	18.05	18556	18765	-	19.19
VAR1	19658	19908	-	14.75	17953	18174	-	16.41
VAR2	18880	19121	-	14.34	16784	17033	-	15.72
NS	20427	21456	-	15.17	19418	20392	-	23.05
BNS	20555	26003	1	16.55	18356	24554	1	35.06
RH	22385	33181	-	24.09	22644	35167	-	35.11
CGM	20520	41049	-	17.91				

Table 3 (Test 14, $N = 1000$, 22 problems)

	$m = 5$				$m = 10$			
Method	NIT	NFV	Fail	Time	NIT	NFV	Fail	Time
SSBC	114600	115295	1	12:01.5	84671	85066	-	13:59.4
NSBC	105694	108073	1	11:34.2	86069	87878	-	14:49.1
DSBC	103796	104632	2	10:57.7	94639	95148	-	15:28.6
VAR1	96859	97941	-	8:37.9	72861	73539	-	10:17.2
VAR2	82456	83702	-	7:48.2	67372	68315	-	9:42.3
NS	108315	111456	2	9:33.8	82222	84426	-	11:02.2
BNS	102313	105828	1	10:32.6	73806	77803	-	11:38.1
RH	98046	154931	-	10:41.4	95430	150827	2	12:34.6
CGM	69805	168471	1	6:45.3				

Table 4 (Test 14, $N = 5000$, 20 problems)

Results of computational experiments imply several conclusions. First, new variationally-derived methods VAR1 and especially VAR2 are usually better than methods SSBC, NSBC and DSBC proposed in [9]. These new methods give best results for sparse well-conditioned problems, but they can be outperformed by standard limited memory methods (e.g., NS and RH) in the sparse ill-conditioned case. New methods work also well for problems contained in Test 28, measured by the total number of function evaluations. The worse computational time in Table 4 is caused by larger number of function evaluations in case of several time consuming dense problems.

Bibliography

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

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

[3] J.Ch. Gilbert, J. Nocedal: *Global convergence properties of conjugate gradient methods for optimization*, SIAM J. Optim. 2 (1992) 21-42.

[4] P.E. Gill, M.W. Leonard: *Limited-memory reduced-Hessian methods for large-scale unconstrained optimization*, SIAM J. Optim. 14 (2003), 380-401.

[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] L. Lukšan, J. Vlček: *Test problems for unconstrained optimization*, Report V-897, Prague, ICS AS CR, 2003.

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

[9] J. Vlček, L. Lukšan: *New variable metric methods for unconstrained minimization covering the large-scale case*, Report V-876, Prague, ICS AS CR, 2002.