# New class of limited-memory variationally-derived variable metric methods [1]

Jan Vlček, Ladislav Lukšan
Institute of Computer Science, Academy of Sciences of the Czech Republic,
L. Lukšan is also from Technical University of Liberec

We present a new family of limited-memory variationally-derived variable metric (VM) line search methods with quadratic termination property for unconstrained minimization. Starting with $x_0 \in \mathcal{R}^N$, VM line search methods (see [6], [3]) generate iterations $x_{k+1} \in \mathcal{R}^N$ by the process $x_{k+1} = x_k + s_k$, $s_k = t_k d_k$, where the direction vectors $d_k \in \mathcal{R}^N$ are descent, i.e. $g_k^T d_k < 0$, $k \geq 0$, and the stepsizes $t_k > 0$ satisfy

$$f(x_{k+1}) - f(x_k) \leq \varepsilon_1 t_k g_k^T d_k, \qquad g_{k+1}^T d_k \geq \varepsilon_2 g_k^T d_k, \tag{1}$$

$k \geq 0$, with $0 < \varepsilon_1 < 1/2$ and $\varepsilon_1 < \varepsilon_2 < 1$, where $f$ is an objective function, $g_k = \nabla f(x_k)$. We denote $y_k = g_{k+1} - g_k$, $k \geq 0$ and by $\|.\|_F$ the Frobenius matrix norm.

We describe a new family in Section 1 and in Section 2 a correction formula, which uses the previous vectors $s_{k-1}$, $y_{k-1}$. Numerical results are presented in Section 3.

## 1 A new family of limited-memory methods

Our methods are based on approximations $\bar{H}_k = U_k U_k^T$, $k > 0$, $\bar{H}_0 = 0$, of the inverse Hessian matrix, which are **invariant** under linear transformations (see [3] for significance of the invariance property in case of ill-conditioned problems), where $N \times \min(k, m)$ matrices $U_k$, $1 \leq m \ll N$, are obtained by limited-memory updates with scaling parameters $\gamma_k > 0$ (see [6]) that satisfy the quasi-Newton condition

$$\bar{H}_{k+1} y_k = \varrho_k s_k, \tag{2}$$

where $\varrho_k > 0$ is a nonquadratic correction parameter (see [6]).

We frequently omit index $k$, replace index $k+1$ by symbol $+$, index $k-1$ by symbol $-$ and denote $V_r = I - r y^T / r^T y$ for $r \in \mathcal{R}^N$, $r^T y \neq 0$ (projection matrix),

$$B = H^{-1}, \quad b = s^T y > 0, \quad \bar{a} = y^T \bar{H} y, \quad \bar{b} = s^T B \bar{H} y, \quad \bar{c} = s^T B \bar{H} B s, \quad \bar{\delta} = \bar{a}\bar{c} - \bar{b}^2 \geq 0.$$

### 1.1 Variationally-derived invariant limited-memory method

Standard VM updates can be derived as updates with the **minimum change** of VM matrix in the sense of some norm (see [6]). We extend this approach to limited-memory methods (see also [10], [12]), using the product form of the update and

[1] This work was supported by the Grant Agency of the Czech Academy of Sciences, project No. IAA1030405, the Grant Agency of the Czech Republic, and the Institutional research plan No. AV0Z10300504

replacing the quasi-Newton condition $U_+U_+^Ty = \bar{H}_+y = \varrho s$ equivalently by

$$U_+^Ty = \sqrt{\gamma}z, \qquad U_+(\sqrt{\gamma}z) = \varrho s, \qquad z^Tz = (\varrho/\gamma)b. \tag{3}$$

**Theorem 1.1.** *Let $T$ be a symmetric positive definite matrix, $\varrho > 0$, $\gamma > 0$, $z \in \mathcal{R}^m$, $1 \le m \le N$, $p = Ty$ and $\mathcal{U}$ the set of $N \times m$ matrices. Then the unique solution to* $\min\{\varphi(U_+) : U_+ \in \mathcal{U}\}$ s.t. (3), $\varphi(U_+) = y^TTy\,\|T^{-1/2}(U_+ - \sqrt{\gamma}U)\|_F^2$, *is*

$$\frac{1}{\sqrt{\gamma}}U_+ = \frac{sz^T}{b} + V_pU\left(I - \frac{zz^T}{z^Tz}\right) = U - \frac{p}{p^Ty}y^TU + \left[s - \frac{\gamma}{\varrho}\left(Uz - \frac{y^TUz}{p^Ty}p\right)\right]\frac{z^T}{b}, \tag{4}$$

*which yields the following projection form of limited-memory update of $\bar{H}$*

$$\frac{1}{\gamma}\bar{H}_+ = \frac{\varrho}{\gamma}\frac{ss^T}{b} + V_pU\left(I - \frac{zz^T}{z^Tz}\right)U^TV_p^T. \tag{5}$$

We can show that updates (4), (5) can be **invariant** under linear transformations, i.e. can preserve the same transformation property of $\bar{H} = UU^T$ as inverse Hessian.

**Theorem 1.2 .** *Consider a change of variables $\tilde{x} = Rx + r$, where $R$ is $N \times N$ nonsingular matrix, $r \in \mathcal{R}^N$. Let vector $p$ lie in the subspace generated by vectors $s$, $\bar{H}y$ and $Uz$ and suppose that $z$, $\gamma$ and coefficients in the linear combination of vectors $s$, $\bar{H}y$ and $Uz$ forming $p$ are invariant under the transformation $x \to \tilde{x}$, i.e. they are not influenced by this transformation. Then for $\tilde{U} = RU$ matrix $U_+$ given by (4) also transforms to $\tilde{U}_+ = RU_+$.*

In the special case (this choice satisfies the assumptions of Theorem 1.2)

$$p = (\lambda/b)s + [(1-\lambda)/\bar{a}]\bar{H}y \ \text{ if } \ \bar{a} \ne 0, \qquad p = (1/b)s, \ \lambda = 1 \ \text{ otherwise} \tag{6}$$

we can easily compare (5) with the scaled Broyden class update of $\bar{H}$ with parameter $\eta = \lambda^2$, to obtain $(1/\gamma)\bar{H}_+ = (1/\gamma)\bar{H}_+^{BC} - (1/z^Tz)\,V_pUz(V_pUz)^T$, where (see [11])

$$\bar{H}_+^{BC} = (\varrho/b)ss^T + \gamma V_p\bar{H}V_p^T. \tag{7}$$

Update (7) is useful for **starting** iterations. Setting $U_+ = [\sqrt{\varrho/b}\,s]$ in the first iteration, every update (7) modifies $U$ and adds one column $\sqrt{\varrho/b}\,s$ to $U_+$. Except for the starting iterations we will assume that matrix $U$ has $m \ge 1$ columns.

To choose parameter $z$, we utilize **analogy with standard VM methods**, setting $H = SS^T$, replacing $U$ by $N \times N$ matrix $S$ and using Theorem 1.1 for the standard scaled Broyden class update (see [6]) of matrix $H = B^{-1}$ and the assertion

**Lemma 1.1.** *Every update (4) with $S$, $S_+$ instead of $U$, $U_+$, $z = \alpha_1S^Ty + \alpha_2S^TBs$ satisfying $z^Tz = (\varrho/\gamma)b$ and $p$ given by (6) belongs to the scaled Broyden class with*

$$\eta = \lambda^2 - b\frac{\gamma}{\varrho}\left(\frac{\alpha_1}{b}\lambda - \frac{\alpha_2}{y^THy}(1-\lambda)\right)^2 y^THy. \tag{8}$$

Thus we concentrate here on the choice $z = \alpha_1 U^T y + \alpha_2 U^T Bs$, $\alpha_2 \neq 0$, which yields

$$z = \pm\sqrt{\frac{\varrho}{\gamma}\frac{b}{\bar{a}\theta^2 + 2\bar{b}\theta + \bar{c}}}(U^T Bs + \theta\, U^T y) \tag{9}$$

by $z^T z = (\varrho/\gamma)b$, where $\theta = \alpha_1/\alpha_2$. The following lemma gives simple conditions for $z$ to be invariant under linear transformations. Note that the standard unit values of $\varrho$, $\gamma$, used in our numerical experiments, satisfy this conditions.

**Lemma 1.2.** *Let numbers $\varrho, \gamma$ and $\theta/t$ be invariant under transformation $\tilde{x} = Rx + r$, where $t$ is the stepsize, $R$ is $N \times N$ nonsingular matrix and $r \in \mathcal{R}^N$, and suppose that $\tilde{U} = RU$. Then vector $z$ given by (9) is invariant under this transformation.*

In our numerical experiments we use the choice $\theta = -\bar{b}/\bar{a}$ for $\bar{a} \neq 0$ (if $\bar{a} = 0$, we do not update), which gives good results. Then $\theta/t$ is invariant and (9) gives $z = \pm\sqrt{(\varrho/\gamma)\; b\; /(\bar{a}\bar{\delta})}\,(\bar{a}\, U^T Bs - \bar{b}\, U^T y)$. In this case we have $y^T U z = 0$ and $V_p U z = U z$.

## 1.2 Variationally-derived simple correction

To have matrices $\bar{H}_k$ invariant, we use such updates that $-\bar{H}_k g_k$ cannot be used as the direction vectors $d_k$. Thus we replace $\bar{H}_k$ by $H_k$ to calculate $d_k = -H_k g_k$.

We will find the minimum correction (in the sense of Frobenius matrix norm) of matrix $\bar{H}_+ + \zeta I$, $\zeta > 0$, in order that the resultant matrix $H_+$ may satisfy the quasi-Newton condition $H_+ y = \varrho s$. First we give the **projection variant** of the well-known Greenstadt's theorem, see [4]. For $M = \bar{H}_+ + \zeta I$, the resulting correction (12) together with update (4) give the new family of limited-memory VM methods.

**Theorem 1.3.** *Let $M, W$ be symmetric matrices, $W$ positive definite, $\varrho > 0$, $q = Wy$ and denote $\mathcal{M}$ the set of $N \times N$ symmetric matrices. Then the unique solution to*

$$\min\{\|W^{-1/2}(M_+ - M)W^{-1/2}\|_F : M_+ \in \mathcal{M}\} \quad \text{s.t.} \quad M_+ y = \varrho s \tag{10}$$

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

$$M_+ = E + V_q\,(M - E)\,V_q^T, \tag{11}$$

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

**Theorem 1.4 .** *Let $W$ be a symmetric positive definite matrix, $\zeta > 0$, $\varrho > 0$, $q = Wy$ and denote $\mathcal{M}$ the set of $N \times N$ symmetric matrices. Suppose that matrix $\bar{H}_+$ satisfies the quasi-Newton condition (2). Then the unique solution to*

$$\min\{\|W^{-1/2}(H_+ - \bar{H}_+ - \zeta I)W^{-1/2}\|_F : H_+ \in \mathcal{M}\} \quad \text{s.t.} \quad H_+ y = \varrho s$$

*is*

$$H_+ = \bar{H}_+ + \zeta V_q V_q^T. \tag{12}$$

To choose parameter $\zeta$, the widely used choice is $\zeta = \varrho b/y^T y$, which minimizes $|(\bar{H}_+ - \zeta I)y|$. We can obtain slightly better results e.g. by the choice

$$\zeta = \varrho\, b/(y^T y + 4\,\bar{a}). \tag{13}$$

As regards parameter $q$, we use comparison with the scaled Broyden class (see [6]).

**Lemma 1.3 .** *Let $A$ be a symmetric matrix, $\gamma > 0$, $\varrho > 0$ and denote $a = y^T Ay$. Then every update (11) with $M = \gamma A$, $M_+ = A_+$, $q = s - \alpha Ay$, $a \neq 0$ and $\alpha\, a \neq b$ represents the scaled Broyden class update with $\eta = [b^2 - \alpha^2(\varrho/\gamma)ab\,]/(b - \alpha a)^2$.*

The following lemma enables us to determine vector $q$ in such a way that correction (12) represents the **Broyden class** update of $\bar{H}_+ + \zeta I$ with parameter $\eta$.

**Lemma 1.4 .** *Let $\varrho > 0$, $\zeta > 0$, $\kappa = \zeta\, y^T y/b$, $\eta > -\varrho/(\varrho + \kappa)$ and let matrix $\bar{H}_+$ satisfy the quasi-Newton condition (2). Then correction (12) with $q = s - \sigma y$, where $\sigma = b(1 \pm \sqrt{(\varrho + \kappa)/(\varrho + \eta\kappa)}\,)/y^T y$ represents the non-scaled Broyden class update of matrix $\bar{H}_+ + \zeta I$ with parameter $\eta$ and nonquadratic correction $\varrho$.*

For $q = s$, i.e. $\eta = 1$, we get the BFGS update. Better results were obtained with the formula, based on analogy with the shifted VM methods (see [9]-[12]):

$$\eta = \min\left[1, \max\left[0\,, 1 + (\varrho/\kappa)(1 + \varrho/\kappa)\,(1.2\,\zeta_-/(\zeta_- + \zeta) - 1)\right]\right]. \tag{14}$$

## 1.3 Quadratic termination property

In this section we give conditions for our family of limited-memory VM methods with exact line searches to terminate on a quadratic function in at most $N$ iterations.

**Theorem 1.5 .** *Let $m \in \mathcal{N}$ be given and let $Q : \mathcal{R}^N \to \mathcal{R}$ be a strictly convex quadratic function $Q(x) = \frac{1}{2}(x - x^*)^T G(x - x^*)$, where $G$ is an $N \times N$ symmetric positive definite matrix. Suppose that $\zeta_k > 0$, $\varrho_k > 0$, $\gamma_k > 0$, $t_k > 0$, $k \geq 0$, and that for $x_0 \in \mathcal{R}^N$ iterations $x_{k+1} = x_k + s_k$ are generated by the method $s_k = -t_k H_k g_k$, $g_k = \nabla Q(x_k)$, $k \geq 0$, with exact line searches, i.e. $g_{k+1}^T s_k = 0$, where*

$$H_0 = I, \quad H_{k+1} = U_{k+1}U_{k+1}^T + \zeta_k V_{q_k} V_{q_k}^T,\ k \geq 0, \tag{15}$$

*$N \times \min(k, m)$ matrices $U_k$, $k > 0$, satisfy*

$$U_1 = \left(\sqrt{\frac{\varrho_0}{b_0}} s_0\right), \quad \frac{1}{\gamma_k} U_{k+1}U_{k+1}^T = \frac{\varrho_k}{\gamma_k}\frac{s_k s_k^T}{b_k} + V_{p_k} U_k U_k^T V_{p_k}^T,\ 0 < k < m, \tag{16}$$

$$\frac{1}{\gamma_k} U_{k+1}U_{k+1}^T = \frac{\varrho_k}{\gamma_k}\frac{s_k s_k^T}{b_k} + V_{p_k} U_k \left(I - \frac{z_k z_k^T}{z_k^T z_k}\right) U_k^T V_{p_k}^T,\ k \geq m, \tag{17}$$

*vectors $z_k \in \mathcal{R}^m$, $k \geq m$, satisfy $z_k^T z_k = (\varrho_k/\gamma_k) b_k$, vectors $p_k$, $k > 0$, lie in $\mathrm{range}([U_k, s_k])$ and satisfy $p_k^T y_k \neq 0$, vectors $q_k$ for $k > 0$ lie in $\mathrm{span}\{s_k, U_k U_k^T y_k\}$ and satisfy $q_k^T y_k \neq 0$ and vector $q_0 = s_0$. Then there exists a number $\bar{k} \leq N$ with $g_{\bar{k}} = 0$ and $x_{\bar{k}} = x^*$.*

## 2 Correction formula

Corrections in Section 1.2 respect only the latest vectors $s_k$, $y_k$. Thus we can again correct (without scaling) matrices $\check{H}_{k+1} = \bar{H}_{k+1} + \zeta_k V_{q_k} V_{q_k}^T$, $k > 0$, obtained from (12), using **previous vectors** $s_i$, $y_i$, $i = k-j, \dots, k-1$, $j \le k$. Our experiments indicate that the choice $j = 1$ brings the maximum improvement. This leads to the formula $H_+ = (\varrho/b) s s^T + V_s \left[ (\varrho_-/b_-) s_- s_-^T + V_s^- \left( \bar{H}_+ + \zeta V_q V_q^T \right) (V_s^-)^T \right] V_s^T$, where $V_s^- = I - s_- y_-^T / b_-$, which is less sensitive to the choice of $\zeta$ than (12). To calculate the direction vector $d_+ = -H_+ g_+$, we can utilize the Strang formula, see [8].

## 3 Computational experiments

Our new limited-memory VM methods were tested, using the collection of sparse, usually **ill-conditioned** problems for large-scale nonlinear least squares from [7] (Test 15 without problem 18, which was very sensitive to the choice of the maximum stepsize in line search, i.e. 21 problems) with $N = 500$ and 1000, $m = 10$, $\varrho = \gamma = 1$, the final precision $\|g(x^\star)\|_\infty \le 10^{-5}$ and $\zeta$ given by (13).

| $\eta_p$ | $N = 500$ | | | | $N = 1000$ | | | |
|---|---|---|---|---|---|---|---|---|
| | Corr-0 | Corr-1 | Corr-2 | Corr-$q$ | Corr-0 | Corr-1 | Corr-2 | Corr-$q$ |
| 0.0 | (2)76916 | 32504 | 22626 | 24016 | (3)99957 | (1)58904 | 44608 | (1)47204 |
| 0.1 | (3)99032 | 36058 | 21839 | 35756 | (3)98270 | (1)54494 | 42649 | (1)47483 |
| 0.2 | (2)97170 | 29488 | 23732 | 29310 | (3)89898 | (1)52368 | 36178 | (1)44115 |
| 0.3 | (1)79978 | 28232 | 18388 | 18913 | (3)80087 | 47524 | 33076 | 38030 |
| 0.4 | (1)70460 | 24686 | 18098 | 17673 | (3)78498 | 44069 | 32403 | 34437 |
| 0.5 | 60947 | 22532 | 17440 | 17181 | (3)88918 | 41558 | 32808 | 31874 |
| 0.6 | 56612 | 21240 | 17800 | 17164 | (2)76264 | 38805 | 31854 | 30784 |
| 0.7 | 52465 | 20289 | 17421 | 17021 | (2)72626 | 39860 | 32345 | 30802 |
| 0.8 | 51613 | 20623 | 17682 | 17076 | (1)69807 | 37501 | 32292 | 32499 |
| 0.9 | 50877 | 20548 | 18102 | 17424 | (2)69802 | 38641 | 32926 | 31385 |
| 1.0 | 49672 | 20500 | 18109 | 17913 | (1)68603 | 38510 | 33539 | 32456 |
| 1.1 | 52395 | 20994 | 18694 | 18470 | (1)65676 | 41284 | 35103 | 33053 |
| 1.2 | 51270 | 21444 | 19230 | 18372 | (1)68711 | 41332 | 35649 | 34028 |
| 1.3 | (1)50064 | 21899 | 19289 | 19890 | (2)67976 | 41491 | 36155 | 34776 |
| 1.4 | (1)52255 | 21900 | 19737 | 19695 | (2)67340 | 43758 | 35793 | 35998 |
| 1.5 | (1)51094 | 22808 | 20487 | 20060 | (2)66220 | 42906 | 36775 | 36323 |
| 2.0 | (1)50776 | 24318 | 21710 | 21639 | (2)66594 | 46139 | 40279 | 39199 |
| 3.0 | (1)54714 | 28641 | 24634 | 24675 | (2)68680 | (1)54531 | 45366 | 44785 |
| BNS | 18444 | | | | 33131 | | | |

Table 1. Comparison of various correction methods.

Results of these experiments are given in three tables, where $\eta_p = \lambda^2$ is the value of parameter $\eta$ of the Broyden class used to determine parameter $p$ by (6) and $\eta_q$ is the value of this parameter used in Lemma 1.4 to determine parameter $q = s - \sigma y$.

| | $\eta_p$ | | | | | | |
|---|---|---|---|---|---|---|---|
| $\eta_q$ | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | 1.0 |
| 0.0 | -343 | -394 | -967 | -813 | -538 | 32 | 141 |
| 0.1 | 211 | -1154 | -1028 | -1100 | -880 | -585 | -188 |
| 0.2 | 2424 | 1902 | 1759 | 2088 | 1869 | 2268 | 2746 |
| 0.3 | -492 | -1064 | -1136 | -992 | -1036 | -901 | -939 |
| 0.4 | -599 | -1069 | -718 | -1160 | -668 | -934 | -512 |
| 0.5 | -493 | -722 | -727 | -665 | -487 | -516 | -399 |
| 0.6 | -251 | -648 | -798 | -965 | -750 | -176 | -371 |
| 0.7 | -342 | -764 | -441 | -320 | -474 | -749 | -284 |
| 0.8 | -481 | -706 | -857 | -579 | -449 | -497 | -606 |
| 0.9 | -872 | -759 | -370 | -559 | -820 | 275 | -135 |
| 1.0 | -346 | -1004 | -644 | -1023 | -762 | -342 | -335 |
| 1.1 | 1939 | 1265 | 2326 | 791 | 2444 | 1958 | 1910 |
| 1.2 | 1024 | 700 | 719 | 1452 | 967 | 1479 | 1982 |
| 1.3 | -322 | -410 | -785 | -872 | -332 | 333 | 174 |
| 1.4 | -600 | -718 | -839 | -1324 | -959 | -811 | 222 |
| 1.5 | -596 | -436 | -912 | -937 | -770 | -285 | 307 |
| 1.6 | -256 | -474 | -365 | -370 | -517 | -86 | 203 |
| 1.7 | -61 | -430 | -526 | -158 | -356 | -211 | 85 |
| 1.8 | -206 | -102 | -240 | -618 | -412 | 71 | 359 |
| 1.9 | -293 | -235 | -169 | -332 | 32 | 23 | 607 |
| 2.0 | 150 | -396 | 85 | 259 | 336 | 222 | 684 |
| 2.5 | 467 | 357 | 863 | 701 | 890 | 1274 | 356 |
| 3.0 | 7698 | 5036 | 4903 | 4337 | 4218 | 3577 | 3541 |
| (14) | -771 | -1263 | -1280 | -1423 | -1368 | -1020 | -531 |

Table 2. Comparison with BNS for N=500.

In Table 1 we compare the method after [2] (BNS) with our new family, using various values of $\eta_p$ and the following **correction methods:** Corr-0 – the adding of matrix $\zeta I$ to $\bar{H}_+$, Corr-1 – correction (12), Corr-2 – correction after Section 2. We use $\eta_q = 1$ (i.e. $q = s$) in columns Corr-0, Corr-1 and Corr-2 and $\eta_q$ given by (14) in columns Corr-$q$ together with correction after Section 2. We present the total numbers of function and also gradient evaluations (over all problems), preceded by the number of problems (in parentheses, if any occurred) which were not solved successfully (the number of evaluations reached its limit 19000 evaluations).

In Table 2 and Table 3 we give the **differences** $n_{p,q} - n_{BNS}$, where $n_{p,q}$ is the total number of function and also gradient evaluations (over all problems) for selected values of $\eta_p$ and $\eta_q$ with correction after Section 2 and $n_{BNS}$ is the number of evaluations for method BNS (negative values indicate that our method is better than BNS). In the last row we present this difference for $\eta_q$ given by (14).

| | $\eta_p$ | | | | | | |
|---|---|---|---|---|---|---|---|
| $\eta_q$ | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | 1.0 |
| 0.0 | 1916 | -912 | -681 | -876 | -119 | -744 | 116 |
| 0.1 | 1052 | -732 | -974 | -1647 | -1043 | -1215 | 320 |
| 0.2 | 903 | -187 | -1669 | -1708 | -1219 | -28 | -567 |
| 0.3 | 793 | -363 | -975 | -1731 | -289 | 360 | -484 |
| 0.4 | 925 | -1398 | -1708 | -1554 | -1184 | -498 | -482 |
| 0.5 | -757 | -644 | -965 | -1729 | -1380 | -926 | -207 |
| 0.6 | 1 | -1396 | -1291 | -835 | -1044 | -767 | 190 |
| 0.7 | -195 | -901 | -356 | -1019 | -1482 | -398 | -454 |
| 0.8 | -770 | -690 | -1763 | -886 | -1009 | -256 | -977 |
| 0.9 | 8 | -821 | -939 | -674 | -696 | -764 | 657 |
| 1.0 | -728 | -323 | -1277 | -786 | -839 | -205 | 408 |
| 1.1 | -773 | 115 | 183 | 48 | -411 | -619 | 736 |
| 1.2 | 269 | 155 | -670 | 295 | -649 | -113 | 647 |
| 1.3 | 51 | 150 | -234 | -527 | -158 | -323 | 1381 |
| 1.4 | 498 | 298 | -522 | 246 | -383 | 696 | 2533 |
| 1.5 | 377 | -181 | -29 | 908 | 1323 | 441 | 1310 |
| 1.6 | 1072 | 1135 | 766 | -39 | 853 | 1307 | 2065 |
| 1.7 | 825 | 874 | -199 | 79 | 607 | 1108 | 3370 |
| 1.8 | 1334 | 1147 | 667 | 1064 | 821 | 3854 | 2908 |
| 1.9 | 1470 | 486 | 1863 | 1047 | 1973 | 2609 | 3156 |
| 2.0 | 2164 | 767 | 994 | 2035 | 2577 | 2869 | 3036 |
| 2.5 | 2284 | 3821 | 3325 | 3337 | 3838 | 4929 | 5167 |
| 3.0 | 4570 | 4457 | 3423 | 4106 | 5172 | 4430 | 4818 |
| (14) | 1306 | -1257 | -2347 | -2329 | -632 | -1746 | -675 |

Table 3. Comparison with BNS for N=1000.

In these numerical experiments, limited-memory VM methods from our new family with suitable values of parameters $\eta_p$ (e.g. $\eta_p = 0.7$) and $\eta_q$ (e.g. $\eta_q$ given by (14)) give **better results** than method BNS.

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 $\|g(x^\star)\|_\infty \le 10^{-6}$. The results are given in Table 4, where Corr-LMM is the limited-memory VM method from our new family with $\eta_p = \eta_q = 0.5$ and correction after Section 2 (the other parameters are the same as above), NIT is the number of iterations, NFV the number of function and also gradient evaluations and Time the computer time in seconds.

| CUTE | | Corr-LMM | | | BNS | | |
|---|---|---|---|---|---|---|---|
| Problem | $N$ | NIT | NFV | Time | NIT | NFV | Time |
| ARWHEAD | 5000 | 8 | 18 | 0.19 | 8 | 18 | 0.18 |
| BDQRTIC | 5000 | 216 | 301 | 1.49 | 145 | 220 | 1.04 |
| BROWNAL | 500 | 7 | 16 | 0.30 | 6 | 16 | 0.29 |
| BROYDN7D | 2000 | 2830 | 2858 | 10.28 | 2953 | 3021 | 10.03 |
| BRYBND | 5000 | 31 | 40 | 0.34 | 31 | 42 | 0.30 |
| CHAINWOO | 1000 | 414 | 467 | 0.36 | 429 | 469 | 0.36 |
| COSINE | 5000 | 21 | 30 | 0.19 | 14 | 19 | 0.14 |
| CRAGGLVY | 5000 | 88 | 101 | 0.77 | 84 | 101 | 0.69 |
| CURLY10 | 1000 | 5428 | 5436 | 3.97 | 5827 | 5975 | 3.37 |
| CURLY20 | 1000 | 5813 | 5818 | 5.05 | 6720 | 6907 | 5.06 |
| CURLY30 | 1000 | 6537 | 6544 | 6.84 | 6831 | 7010 | 6.08 |
| DIXMAANA | 3000 | 10 | 14 | 0.06 | 9 | 13 | 0.06 |
| DIXMAANB | 3000 | 13 | 17 | 0.06 | 7 | 11 | 0.03 |
| DIXMAANC | 3000 | 12 | 16 | 0.06 | 9 | 13 | 0.06 |
| DIXMAAND | 3000 | 15 | 19 | 0.06 | 11 | 15 | 0.05 |
| DIXMAANE | 3000 | 392 | 396 | 1.08 | 237 | 249 | 0.55 |
| DIXMAANF | 3000 | 328 | 332 | 0.89 | 180 | 188 | 0.43 |
| DIXMAANG | 3000 | 345 | 349 | 0.80 | 178 | 187 | 0.44 |
| DIXMAANH | 3000 | 299 | 303 | 0.80 | 183 | 192 | 0.47 |
| DIXMAANI | 3000 | 2649 | 2653 | 6.88 | 855 | 877 | 1.97 |
| DIXMAANJ | 3000 | 776 | 780 | 1.97 | 340 | 351 | 0.84 |
| DIXMAANK | 3000 | 596 | 573 | 1.41 | 314 | 326 | 0.70 |
| DIXMAANL | 3000 | 541 | 545 | 1.42 | 221 | 230 | 0.52 |
| DQRTIC | 5000 | 966 | 907 | 2.86 | 235 | 236 | 0.52 |
| EDENSCH | 5000 | 26 | 28 | 0.25 | 25 | 29 | 0.23 |
| EG2 | 1000 | 4 | 9 | 0.01 | 4 | 9 | 0.02 |
| ENGVAL1 | 5000 | 23 | 40 | 0.24 | 26 | 35 | 0.20 |
| EXTROSNB | 5000 | 39 | 43 | 0.27 | 40 | 46 | 0.32 |
| FLETCBV2 | 1000 | 1246 | 1248 | 1.33 | 1162 | 1182 | 1.14 |
| FLETCHCR | 1000 | 68 | 73 | 0.08 | 50 | 58 | 0.08 |
| FMINSRF2 | 1024 | 405 | 408 | 2.33 | 332 | 340 | 1.73 |

Table 4a: Comparison with BNS for CUTE

| CUTE | | Corr-LMM | | | BNS | | |
|---|---|---|---|---|---|---|---|
| Problem | $N$ | NIT | NFV | Time | NIT | NFV | Time |
| FMINSURF | 1024 | 513 | 517 | 2.97 | 462 | 477 | 2.50 |
| FREUROTH | 5000 | 22 | 47 | 0.31 | 24 | 32 | 0.27 |
| GENHUMPS | 1000 | 2424 | 2698 | 4.58 | 1802 | 2271 | 3.70 |
| GENROSE | 1000 | 2088 | 2199 | 1.63 | 2106 | 2374 | 1.58 |
| LIARWHD | 1000 | 21 | 29 | 0.16 | 23 | 28 | 0.19 |
| MOREBV | 5000 | 114 | 116 | 0.45 | 112 | 116 | 0.39 |
| MSQRTALS | 529 | 3136 | 3142 | 6.81 | 2880 | 2947 | 6.08 |
| NCB20 | 510 | 783 | 845 | 4.38 | 505 | 561 | 2.81 |
| NCB20B | 1010 | 2087 | 2204 | 11.27 | 1584 | 1715 | 8.61 |
| NONCVXU2 | 1000 | 2492 | 2493 | 2.45 | 3603 | 3685 | 3.06 |
| NONCVXUN | 1000 | 23993 | 23994 | 23.42 | - | $>50000$ | - |
| NONDIA | 5000 | 14 | 19 | 0.19 | 25 | 30 | 0.27 |
| NONDQUAR | 5000 | 16080 | 16090 | 49.25 | 3210 | 3588 | 8.42 |
| PENALTY1 | 1000 | 61 | 69 | 0.00 | 64 | 72 | 0.05 |
| PENALTY3 | 100 | 61 | 91 | 0.63 | 56 | 92 | 0.66 |
| POWELLSG | 5000 | 45 | 57 | 0.09 | 37 | 46 | 0.14 |
| POWER | 1000 | 489 | 496 | 0.13 | 104 | 110 | 0.02 |
| QUARTC | 5000 | 966 | 967 | 2.70 | 235 | 236 | 0.52 |
| SBRYBND | 5000 | - | - | - | - | - | - |
| SCHMVETT | 5000 | 35 | 37 | 0.39 | 36 | 42 | 0.38 |
| SCOSINE | 5000 | - | - | - | - | - | - |
| SINQUAD | 5000 | 288 | 386 | 2.25 | 250 | 338 | 1.83 |
| SPARSINE | 1000 | 9396 | 9400 | 11.20 | 9347 | 9726 | 9.66 |
| SPARSQUR | 1000 | 35 | 41 | 0.06 | 37 | 43 | 0.05 |
| SPMSRTLS | 4999 | 201 | 204 | 1.24 | 213 | 223 | 1.14 |
| SROSENBR | 5000 | 12 | 19 | 0.08 | 18 | 23 | 0.11 |
| TOINTGSS | 5000 | 4 | 6 | 0.11 | 4 | 7 | 0.08 |
| TQUARTIC | 5000 | 19 | 25 | 0.17 | 21 | 30 | 0.20 |
| VARDIM | 1000 | 24 | 41 | 0.02 | 33 | 40 | 0.03 |
| VAREIGVL | 1000 | 143 | 146 | 0.14 | 164 | 171 | 0.16 |
| WOODS | 4000 | 34 | 41 | 0.14 | 28 | 33 | 0.11 |

Table 4b: Comparison with BNS for CUTE

Our limited numerical experiments indicate that methods from our new family can compete with the well-known BNS method.

## 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] J. Huschens, *On the use of product structure in secant methods for nonlinear least squares problems*, SIAM J. Optim. 4 (1994), 108-129.

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

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

[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.

[10] J. Vlček, L. Lukšan: *Additional properties of shifted variable metric methods*, Report V-899, Prague, ICS AS CR, 2004.

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

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