LIMITED-MEMORY VARIABLE METRIC METHODS BASED ON INVARIANT MATRICES*

Jan Vlček, Ladislav Lukšan

We present a new family of limited-memory variationally-derived variable metric (VM) line search methods with quadratic termination property (see [4]) for unconstrained minimization. Starting with $x_0 \in \mathcal{R}^N$, VM line search methods (see [4]) 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. $g_k^T d_k < 0$, and the stepsizes $t_k > 0$ satisfy

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

$k \ge 0$, 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 $y_k = g_{k+1} - g_k$, $k \ge 0$ and by $\|.\|_F$ the Frobenius matrix norm.

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 (it is significant in case of ill-conditioned problems), where $N \times \min(k, m)$ matrices U_k, $1 \le m \ll N$, are obtained by limited-memory updates that satisfy the quasi-Newton condition

$$\bar{H}_{k+1} y_k = s_k \qquad \text{or equivalently} \qquad U_+^T y = z, \quad U_+ z = s, \quad z^T z = b. \tag{2}$$

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 $r \in \mathcal{R}^N$, $r^T y \ne 0$ (projection matrix),

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

Standard VM updates can be derived as updates with the **minimum change** of VM matrix (see [4]), which we extend to limited-memory methods (see [6], [7]).

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

$$U_+ = s z^T / b + V_p U \Big(I - z z^T / z^T z\Big), \quad \bar{H}_+ = s s^T / b + V_p U \Big(I - z z^T / z^T z\Big) U^T V_p^T. \tag{3}$$

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

Updates (3) can be **invariant** under linear transformations, i.e. can preserve the same transformation property of $\bar{H} = UU^T$ as inverse Hessian (see [7]).

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 $p \in \text{span}\{s, \bar{H}y, Uz\}$ and suppose that z 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 (3) also transforms to $\tilde{U}_+ = RU_+$.*

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

$$p = (\lambda/b)s + [(1-\lambda)/\bar{a}]\bar{H}y \ \text{ if } \ \bar{a} \neq 0, \qquad p = (1/b)s, \ \lambda = 1 \ \text{ otherwise} \tag{4}$$

we can easily compare (3) with the Broyden class update of $\bar{H}$ with parameter $\eta = \lambda^2$, to obtain $\bar{H}_+ = \bar{H}_+^{BC} - V_p Uz(V_p Uz)^T/z^T z$, where $\bar{H}_+^{BC} = ss^T/b + V_p \bar{H} V_p^T$ (see [6]). The last update is useful for **starting** iterations. Setting $U_+ = [\sqrt{1/b}\, s]$ in the first iteration, every such update modifies U and adds one column $\sqrt{1/b}\, s$ to U_+. Except for the starting iterations, we will assume that matrix U has $m \geq 1$ columns.

To choose parameter z, we utilize analogy with standard VM methods (see [7]).

Lemma 1.3. *Let $H = SS^T$ with $N \times N$ matrix S and let $z = \alpha(S^T Bs + \theta S^T y)$, α, $\theta \in \mathcal{R}$, with $z^T z = b$. Then every update (3) with S, S_+, $S_+ S_+^T$ instead of U, U_+, $\bar{H}_+$ and with p given by (4) belongs to the scaled Broyden class with*

$$\eta = \lambda^2 - b\,\alpha^2\, y^T H y \left(\theta\lambda/b - (1-\lambda)/y^T H y\right)^2. \tag{5}$$

Thus we concentrate here on the choice $z = \alpha(S^T Bs + \theta S^T y)$, $z^T z = b$. Lemma 1.4 (see [7]) gives simple conditions for this z to be invariant under linear transformations.

Lemma 1.4. *Let number θ/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 used in Lemma 1.3 is invariant under this transformation.*

We use the choice $\theta = -\bar{b}/\bar{a}$. Then θ/t is invariant and $z^T z = b$ gives (if $\bar{a}\bar{\delta} = 0$, we do not update) $z = \pm\sqrt{b/(\bar{a}\bar{\delta})}\,(\bar{a}\, U^T Bs - \bar{b}\, U^T y)$, $y^T Uz = 0$, and $V_p Uz = Uz$.

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 find the minimum correction (in the sense of Frobenius matrix norm, see [7]) of matrix $\bar{H}_+ + \zeta I$, $\zeta > 0$, in order that the resultant matrix H_+ may satisfy the quasi-Newton condition $H_+ y = s$. First we give the **projection variant** of the well-known Greenstadt's theorem, see [3].

For $M = \bar{H}_+ + \zeta I$, the resulting correction (8) together with update (3) give our family of limited-memory VM methods.

Theorem 2.1. *Let M, W be symmetric matrices, W positive definite, $q = Wy$ and denote $\mathcal{M}$ the set of $N \times N$ symmetric matrices. Then the unique solution to*

$$\min\{\|W^{-1/2}(M_+ - M)W^{-1/2}\|_F : M_+ \in \mathcal{M}\} \quad \text{s.t.} \quad M_+ y = s \tag{6}$$

is given by $V_q\,(M_+ - M)\,V_q^T = 0$ and can be written (the usual form is on the right)

$$M_+ = E + V_q\,(M - E)V_q^T \equiv M + (wq^T + qw^T)/q^T y - w^T y \cdot qq^T/(q^T y)^2\,, \tag{7}$$

where E is any symmetric matrix satisfying $Ey = s$, e.g. $E = ss^T/b$, $w = s - My$.

Theorem 2.2. *Let W be a symmetric positive definite matrix, $\zeta > 0$, $q = Wy$ 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

$$H_+ = \bar{H}_+ + \zeta V_q V_q^T. \tag{8}$$

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

$$\zeta = \varrho\, b/(y^T y + 4\,\bar{a}). \tag{9}$$

The following lemmas (see [7]) help us to obtain vector q in such a way that corrections (7), (8) represent the **Broyden class** updates of $\bar{H}_+ + \zeta I$ with parameter η.

Lemma 2.3. *Let A be a symmetric matrix and denote $a = y^T Ay$. Then every update (7) with $M = A$, $M_+ = A_+$, $q = s - \alpha Ay$, $a \neq 0$, and $\alpha a \neq b$ represents the Broyden class update with parameter $\eta = (b^2 - \alpha^2 ab)/(b - \alpha a)^2$.*

Lemma 2.4. *Let $\zeta > 0$, $\kappa = \zeta\, y^T y/b$, $\eta > -1/(1 + \kappa)$ and let matrix $\bar{H}_+$ satisfy the quasi-Newton condition (2). Then correction (8) with $q = s - \sigma y$, where $\sigma = b(1 \pm \sqrt{(1 + \kappa)/(1 + \eta\kappa)}\,)/y^T y$ represents the Broyden class update of matrix $\bar{H}_+ + \zeta I$ with parameter η.*

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 [6], [7]):

$$\eta = \min\left[1, \max\left[0\,, 1 + (1/\kappa)(1 + 1/\kappa)\,(1.2\,\zeta_-/(\zeta_- + \zeta) - 1)\right]\right]. \tag{10}$$

3. Correction formula

Corrections in Section 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 (8), using **previous vectors** s_i, y_i, $i = k-j, \ldots, k-1$, $j \le k$. Our experiments indicate that the choice $j = 1$ brings the maximum improvement. This leads to the formula $H_+ = ss^T/b + V_s\left[s_- s_-^T/b_- + V_s^-\left(\bar{H}_+ + \zeta V_q V_q^T\right)(V_s^-)^T\right]V_s^T$, where $V_s^- = I - s_- y_-^T/b_-$, which is less sensitive to the choice of ζ than (8).

4. Quadratic termination property

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 (see [7]).

Theorem 4.1. *Let $m \in \mathcal{N}$ be given and let $Q(x) = \frac{1}{2}(x - x^*)^T G(x - x^*)$, where G is an $N \times N$ symmetric positive definite matrix. Suppose that $\zeta_k > 0$, $t_k > 0$, $k \geq 0$, 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*

$$H_0 = I, \quad H_{k+1} = U_{k+1} U_{k+1}^T + \zeta_k V_{q_k} V_{q_k}^T, \ k \geq 0, \tag{11}$$

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

$$U_1 = \left[s_0 / \sqrt{b_0} \right], \quad U_{k+1} U_{k+1}^T = s_k s_k^T / b_k + V_{p_k} U_k U_k^T V_{p_k}^T, \ 0 < k < m, \tag{12}$$

$$U_{k+1} U_{k+1}^T = s_k s_k^T / b_k + V_{p_k} U_k \Big(I - z_k z_k^T / z_k^T z_k \Big) U_k^T V_{p_k}^T, \ z_k \in \mathcal{R}^m, \ k \geq m, \tag{13}$$

vectors p_k, q_k, $k > 0$, lie in $\mathrm{range}([U_k, s_k])$ and satisfy $p_k^T y_k \neq 0$, $q_k^T y_k \neq 0$, and $q_0 = s_0$. Then there exists a number $\bar{k} \leq N$ with $g_{\bar{k}} = 0$ and $x_{\bar{k}} = x^*$.

5. Computational experiments

Our VM methods were tested, using the collection [5] of sparse, usually ill-conditioned problems for large-scale nonlinear least squares (Test 15, 21 problems) with $N = 500$ and 1000, $m = 10$, ζ given by (9) and the final precision $\|g(x^\star)\|_\infty \leq 10^{-5}$.

η_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
1.5	1-51094	22808	20487	20060	2-66220	42906	36775	36323
2.0	1-50776	24318	21710	21639	2-66594	46139	40279	39199
BNS	18444				33131			

Tab. 1: *Comparison of various correction methods.*

Results of these experiments are given in two tables, where $\eta_p = \lambda^2$ is the value of parameter η of the Broyden class used to determine parameter p by (4) and η_q is the value of this parameter used in Lemma 2.4 to determine parameter $q = s - \sigma y$.

In Table 1 we compare the method described in [2] (BNS) with our new family, using various values of η_p and the following **correction methods:** Corr-0 – the adding of matrix ζI to $\bar{H}_+$, Corr-1 – correction (8), Corr-2 – correction from Section 3. We use $\eta_q = 1$ (i.e. $q = s$) in columns Corr-0, Corr-1 and Corr-2 and η_q given by (10) in columns Corr-q together with correction from Section 3. We present the total numbers of function and also gradient evaluations (over all problems), preceded by the minus-sign with the number of problems (if any occurred) which were not solved successfully (the number of evaluations reached its limit 19000 evaluations).

	η_p , $N = 500$							η_p , $N = 1000$						
η_q	0.4	0.5	0.6	0.7	0.8	0.9	1.0	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0.0	-343	-394	-967	-813	-538	32	141	1916	-912	-681	-876	-119	-744	116
0.1	211	-1154	-1028	-1100	-880	-585	-188	1052	-732	-974	-1647	-1043	-1215	320
0.2	2424	1902	1759	2088	1869	2268	2746	903	-187	-1669	-1708	-1219	-28	-567
0.3	-492	-1064	-1136	-992	-1036	-901	-939	793	-363	-975	-1731	-289	360	-484
0.4	-599	-1069	-718	-1160	-668	-934	-512	925	-1398	-1708	-1554	-1184	-498	-482
0.5	-493	-722	-727	-665	-487	-516	-399	-757	-644	-965	-1729	-1380	-926	-207
0.6	-251	-648	-798	-965	-750	-176	-371	1	-1396	-1291	-835	-1044	-767	190
0.7	-342	-764	-441	-320	-474	-749	-284	-195	-901	-356	-1019	-1482	-398	-454
0.8	-481	-706	-857	-579	-449	-497	-606	-770	-690	-1763	-886	-1009	-256	-977
0.9	-872	-759	-370	-559	-820	275	-135	8	-821	-939	-674	-696	-764	657
1.0	-346	-1004	-644	-1023	-762	-342	-335	-728	-323	-1277	-786	-839	-205	408
1.1	1939	1265	2326	791	2444	1958	1910	-773	115	183	48	-411	-619	736
1.2	1024	700	719	1452	967	1479	1982	269	155	-670	295	-649	-113	647
1.5	-596	-436	-912	-937	-770	-285	307	377	-181	-29	908	1323	441	1310
2.0	150	-396	85	259	336	222	684	2164	767	994	2035	2577	2869	3036
(10)	-771	-1263	-1280	-1423	-1368	-1020	-531	1306	-1257	-2347	-2329	-632	-1746	-675

Tab. 2: *Comparison with BNS for various* η_p, η_q.

In Table 2 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 η_p and η_q with correction from Section 3 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 η_q given by (10).

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 $\|g(x^\star)\|_\infty \le 10^{-6}$. In Table 3 we give the values of $(n_{p,q} - n_{BNS})/(n_{p,q} + n_{BNS}) * 100$ for $\eta_p = \eta_q = 0.5$ (all the others are the same as above).

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

Problem	N	%	Problem	N	%	Problem	N	%
ARWHEAD	5000	=0	BDQRTIC	5000	16	BROWNAL	500	=0
BROYDN7D	2000	-3	BRYBND	5000	-2	CHAINWOO	1000	-0
COSINE	5000	22	CRAGGLVY	5000	=0	CURLY10	1000	-5
CURLY20	1000	-9	CURLY30	1000	-3	DIXMAANA	3000	4
DIXMAANB	3000	21	DIXMAANC	3000	10	DIXMAAND	3000	12
DIXMAANE	3000	23	DIXMAANF	3000	28	DIXMAANG	3000	30
DIXMAANH	3000	22	DIXMAANI	3000	50	DIXMAANJ	3000	38
DIXMAANK	3000	27	DIXMAANL	3000	41	DQRTIC	5000	59
EDENSCH	5000	-2	EG2	1000	=0	ENGVAL1	5000	7
EXTROSNB	5000	-3	FLETCBV2	1000	3	FLETCHCR	1000	11
FMINSRF2	1024	9	FMINSURF	1024	4	FREUROTH	5000	19
GENHUMPS	1000	9	GENROSE	1000	-4	LIARWHD	1000	2
MOREBV	5000	=0	MSQRTALS	529	3	NCB20	510	20
NCB20B	1010	12	NONCVXU2	1000	-19	NONCVXUN	1000	<-35
NONDIA	5000	-22	NONDQUAR	5000	64	PENALTY1	1000	-2
PENALTY3	100	-1	POWELLSG	5000	11	POWER	1000	64
QUARTC	5000	61	SCHMVETT	5000	-6	SINQUAD	5000	7
SPARSINE	1000	-2	SPARSQUR	1000	-2	SPMSRTLS	4999	-4
SROSENBR	5000	-10	TOINTGSS	5000	-8	TQUARTIC	5000	-9
VARDIM	1000	1	VAREIGVL	1000	-8	WOODS	4000	11

Tab. 3: *Comparison with BNS for CUTE.*

References

[1] I. Bongartz, A.R. Conn, N. Gould, P.L. Toint: *CUTE: constrained and unconstrained testing environment*, ACM Trans. Math. 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] J. Greenstadt: *Variations on variable metric methods*, Math. Comput. **24** (1970), 1–22.

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

[5] L. Lukšan, J. Vlček: *Sparse and partially separable test problems for unconstrained and equality constrained optimization*, Report V-767, Inst. of Computer Science, Acad. Sci., Czech Rep., 1998.

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

[7] J. Vlček, L. Lukšan: *New class of limited-memory variationally-derived variable metric methods*, Report V-973, Inst. of Computer Science, Acad. Sci., Czech Rep., 2006.