Metody s proměnnou metrikou s omezenou pamětí, založené na invariantních maticích *

Jan Vlček, Ladislav Lukšan
ÚI AVČR Praha 8, L. Lukšan také TU Liberec

J. Vlček, L. Lukšan: *New class of limited-memory variationally-derived variable metric methods*, Zpráva V-973, ÚI AV ČR, Praha, 2006.

*Tato práce byla podpořena AV ČR grantem č. IAA1030405 a výzkumným záměrem AV0Z10300504

1. NOVÁ TŘÍDA METOD

Popis: nová třída VM (kvazinewton.) LS metod pro nepodmíněnou minimalizaci s vlastností kvadratického ukončení.

VM LS metody: $x_0 \in \mathcal{R}^N$, $x_{k+1} = x_k + s_k$, $s_k = t_k d_k$, $d_k = -H_k g_k$ jsou směrové vektory, délky kroku $t_k > 0$ splňují

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

$k \ge 0$, $0 < \varepsilon_1 < \frac{1}{2}$, $\varepsilon_1 < \varepsilon_2 < 1$, f je účelová funkce, $g_k = \nabla f(x_k)$.

Matice H_k: aprox. inv. Hessovy matice, obvykle spd, $H_0 = I$, $H_{k+1} = H_k +$ aktualiz. hodn. ≤ 2, splňující QN podmínku

$$H_{k+1} y_k = s_k, \quad y_k = g_{k+1} - g_k, \quad k \ge 0.$$

Standardní VMM - pracují s plnými maticemi H_k (H_k^{-1}).

Mezi nejlepšími jsou metody patřící do **Broydenovy třídy**.
Z nich nejznámější jsou:

- BFGS (1970)
- DFP (1959-1963)
- SR1 (1967-1969) - aktualizace hodnosti 1

LMM - pracují s řádově menšími maticemi nebo bez matic.

PRINCIP: Místo H_k bereme $\bar{H}_k = U_k U_k^T$, $k > 0$, $\bar{H}_0 = 0$, (kde U_k jsou $N \times \min(k, m)$ matice, $1 \leq m \ll N$), které **aktualizujeme** tak, aby splňovaly kvazinewtonovskou podmínku

$$\bar{H}_{k+1} y_k = s_k$$

a byly **invariantní** vůči lineární transformaci, což má význam pro špatně podmíněné úlohy. Pro d_k pak $\bar{H}_k$ korigujeme.

Značení: Často vynecháme index k, nahradíme index $k+1$ resp. $k-1$ symbolem $+$ resp. $-$ a označíme:

$$V_r = I - r y^T / r^T y \text{ (matice projekce)}, \; r \in \mathcal{R}^N, \; r^T y \neq 0,$$

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

Standardní VM aktualizace lze odvodit z principu **minimální změny** VM matic (ve smyslu nějaké normy). To lze i pro LMM náhradou QN podmínky $U_+ U_+^T y = \bar{H}_+ y = s$ ekvivalentně:

$$U_+^T y = z, \quad U_+ z = s, \quad z^T z = b. \tag{1}$$

Věta 1. *Buď* T *spd,* $z \in \mathcal{R}^m$, $1 \le m \le N$, $p = Ty$ *a* $\mathcal{U}$ *množina* $N \times m$ *matic. Pak jediné řešení úlohy* $\min\{\varphi(U_+) : U_+ \in \mathcal{U}\}$ s.t. (1), $\varphi(U_+) = y^T T y \, \|T^{-1/2}(U_+ - U)\|_F^2$, *je*

$$U_+ = \frac{s z^T}{b} + V_p U \left(I - \frac{z z^T}{z^T z} \right), \quad \bar{H}_+ = \frac{s s^T}{b} + V_p U \left(I - \frac{z z^T}{z^T z} \right) U^T V_p^T. \tag{2}$$

Lze ukázat, že tyto aktualizace mohou být **invariantní** vůči lineární transformaci, tj. mohou zachovávat transformační vlastnost $\bar{H} = UU^T$, tutéž jako má inverzní Hessova matice.

Věta 2. *Uvažujme transformaci $\tilde{x} = Rx + r$, kde R je $N \times N$ regulární matice, $r \in \mathcal{R}^N$. Buď $p \in \text{span}\{s, \bar{H}y, Uz\}$ a nechť z a koeficienty lineární kombinace vektorů s, $\bar{H}y$ a Uz tvořící p jsou invariantní vůči transformaci $x \to \tilde{x}$, tj. nemění se touto transformací. Pak jestliže $\tilde{U} = RU$, transformuje se matice U_+ daná (2) rovněž na $\tilde{U}_+ = RU_+$.*

Ve speciálním případě (splňujícím předpoklady věty 2)

$$p = \frac{\lambda}{b}s + \frac{1-\lambda}{\bar{a}}\bar{H}y \quad \text{pro } \bar{a} \neq 0, \qquad p = \frac{1}{b}s, \ \lambda = 1 \quad \text{jinak}, \quad (3)$$

lze snadno **porovnat** aktualizaci (2) matice $\bar{H}$ s aktualizací ze známé Broydenovy třídy s parametrem $\eta = \lambda^2$:

$$\bar{H}_+ = \bar{H}_+^{BC} - \frac{V_pUz(V_pUz)^T}{z^Tz}, \qquad \bar{H}_+^{BC} = \frac{ss^T}{b} + V_p\bar{H}V_p^T.$$

Aktualizace $\bar{H}_+^{BC}$ je užitečná pro **startovací** iterace: položíme $U_1 = [\sqrt{1/b_1}\, s_1]$ v prvé iteraci a každá aktualizace $\bar{H}_+^{BC}$ modifikuje U a přidá sloupec $\sqrt{1/b}\, s$. Budeme předpokládat, že kromě startovacích iterací má matice U $m \geq 1$ sloupců.

Volba parametru z: Z analogie se standardními VMM. Položíme $H = SS^T$, nahradíme U čtvercovou maticí S řádu N, užijeme větu 1 pro standardní Broydenovu třídu a lemma:

Lemma 1. *Každá aktualizace (2) s S, S_+ místo U, U_+, $z = \alpha(S^T Bs + \theta S^T y)$ splňující $z^T z = b$ s vektorem p podle (3) patří do Broydenovy třídy s parametrem*

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

Následující lemma dává jednoduchou **podmínku** pro to, aby toto z bylo invariantní vůči lineární transformaci.

Lemma 2. *Buď θ/t invariantní vůči transformaci $\tilde{x} = Rx + r$, kde t je délka kroku, R je $N \times N$ regulární matice a $r \in \mathcal{R}^N$, a předpokládejme, že $\tilde{U} = RU$. Pak vektor z uvedený v lemmatu 1 je invariantní vůči této transformaci.*

V numerických experimentech užíváme volbu $\theta = -\overline{b}/\overline{a}$ (lze -li počítat z - jinak neaktualizujeme), která minimalizuje $|U^T Bs + \theta U^T y|$. Pak je θ/t invariantní, platí $y^T Uz = 0$ a $V_p Uz = Uz$ a podmínka $z^T z = b$ dává (je vždy $\overline{a}\overline{c} \geq \overline{b}^2$)

$$z = \pm\sqrt{b/[\overline{a}(\overline{a}\overline{c} - \overline{b}^2)]}\,(\overline{a}\, U^T Bs - \overline{b}\, U^T y).$$

2. JEDNODUCHÁ KOREKCE

Užívané aktualizace sice zajišťují invariantnost matic $\bar{H}_k$, ale $-\bar{H}_k g_k$ nelze použít jako směrový vektor $d_k \Rightarrow$ přidáme k $\bar{H}_k$ ζI, $\zeta > 0$, korekce této matice dá H_k a počítáme $d_k = -H_k g_k$.

Výsledná korekce matice $\bar{H}_+ + \zeta I$ spolu s aktualizací (2) bude tvořit **novou třídu** LM VMM.

Budeme hledat **minimální korekci** (ve smyslu Frobeniovy maticové normy) matice $\bar{H}_+ + \zeta I$, $\zeta > 0$, aby výsledná matice H_+ splňovala QN podmínku $H_+ y = s$. Nejprve uvedeme **novou projektivní variantu** Greenstadtovy věty (Math. Comp. 24 (1970) 1-22).

Věta 3. *Buďte matice M, W symetrické, W spd, $q = Wy$ a označme $\mathcal{M}$ množinu $N \times N$ symetrických matic. Pak jediné řešení úlohy*

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

je určeno vztahem $V_q \left(M_+ - M\right) V_q^T = 0$ a lze ho psát jako

$$M_+ = E + V_q\,(M - E)V_q^T \equiv M + \frac{wq^T + qw^T}{q^T y} - \frac{w^T y}{(q^T y)^2}\,qq^T, \quad (4)$$

kde $w = (M_+ - M)y = s - My$, E je libovolná symetrická matice splňující podmínku $Ey = s$, např. $E = \frac{ss^T}{b}$.

Případ $M = \bar{H}_+ + \zeta I \;\; \Rightarrow$ **zjednodušení:**

Věta 4. *Buď* W *spd,* $\zeta > 0$, $q = Wy$ *a označme* $\mathcal{M}$ *množinu* $N \times N$ *symetrických matic. Nechť matice* $\bar{H}_+$ *splňuje QN podmínku* $\bar{H}_+ y = s$. *Pak jediné řešení úlohy*

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

je $\quad H_+ = \bar{H}_+ + \zeta V_q V_q^T.$

Volba ζ: Pro 1. iteraci VMM se velmi často užívá $\zeta = b/y^T y$, minimalizující $|(H_+ - \zeta I)y|$. O málo lepší výsledky dává

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

Volba parametru q - porovnáním s Broydenovou třídou:

Lemma 3. *Buď* A *symetrická matice,* $a = y^T A y$. *Pak každá aktualizace (4) s* $M = A$, $M_+ = A_+$, $q = s - \alpha A y$, $a \neq 0$, $\alpha\, a \neq b$, *patří do Broydenovy třídy s* $\eta = [b^2 - \alpha^2 ab]/(b - \alpha a)^2$.

$\alpha = 0$ $(q = s)$ dá aktualizaci BFGS $(\eta = 1)$, $\alpha = 1$ dá SR1.

Lemma 4. *Buď* $\eta > -1/(1+\kappa)$, $\kappa = \zeta\, y^T y / b$ *a nechť matice* $\bar{H}_+$ *splňuje QN podmínku* $\bar{H}_+ y = s$. *Pak korekce* $\bar{H}_+ + \zeta V_q V_q^T$, $q = s - \sigma y$, $\sigma = b(1 \pm \sqrt{(1+\kappa)/(1+\eta\kappa)}\,)/y^T y$, *je aktualizací (matice* $\bar{H}_+ + \zeta I$*) z Broydenovy třídy s parametrem* η.

3. KVADRATICKÉ UKONČENÍ

Věta 5. *Buď* $m \in \mathcal{N}$, $f : \mathcal{R}^N \to \mathcal{R}$ *ryze konvexní kvadratická,* $\zeta_k > 0$, $t_k > 0$, $k \geq 0$, $x_0 \in \mathcal{R}^N$. *Mějme metodu* $x_{k+1} = x_k + s_k$, $s_k = -t_k H_k g_k$, $k \geq 0$, *s přesnými LS, tj.* $g_{k+1}^T s_k = 0$, *kde*

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

$N \times \min(k, m)$ *matice* U_k, $k > 0$, *splňují*

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

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

vektory p_k, q_k, $p_k^T y_k \neq 0$, $q_k^T y_k \neq 0$, *leží v* range$([U_k, s_k])$, $k > 0$, $q_0 = s_0$. *Pak existuje* $\bar{k} \leq N$, *že* $g_{\bar{k}} = 0$ *a* $x_{\bar{k}} = x^*$.

4. KOREKČNÍ FORMULE

Jednoduché korekce respektují jen poslední vektory s_k, y_k. Proto **znovu** korigujeme matice $\bar{H}_{k+1} + \zeta_k V_{q_k} V_{q_k}^T$, $k > 0$, (jednoduché korekce matic $\bar{H}_{k+1} + \zeta_k I$) pomocí **předchozích vektorů** s_i, y_i, $i = k - j, \ldots, k - 1$, $j \le k$. Nevětší zlepšení dávala volba $j = 1 \Rightarrow$ **formule** ($V_s^- = I - s_- y_-^T / b_-$)

$$H_+ = \frac{ss^T}{b} + V_s \left[\frac{s_- s_-^T}{b_-} + V_s^- \left(\bar{H}_+ + \zeta V_q V_q^T \right) (V_s^-)^T \right] V_s^T,$$

která je méně citlivá na volbu ζ.

K výpočtu směrového vektoru $d_+ = -H_+ g_+$ můžeme výhodně použít Strangovy rekurence (Nocedal: *Updating...*, Math. Comp. 35 (1980) 773-782).

5. NUMERICKÉ VÝSLEDKY

Nejprve testy pomocí sbírky řídkých, obvykle **špatně podmíněných** velkých úloh aproximace metodou nejmenších čtverců (zpráva ÚI V-767, test 15, 21 úloh) pro $N = 500$ a 1000, $m = 10$, přesnost $\|g(x^\star)\|_\infty \leq 10^{-5}$ a $\zeta = b/(y^T y + 4\,\bar{a})$.

V tabulkách je $\eta_p = \lambda^2$ **parametr** η Broydenovy třídy pro základní aktualizaci (užitím vektoru p), η_q tento parametr pro jednoduchou korekci (užitím vektoru $q = s - \sigma y$).

V **tabulce 1** porovnáme naši třídu s metodou BNS (Byrd, Nocedal, Schnabel, MP 63(1994),129-156) pro různá η_p a **korekce:**

Corr-0 – pouhé přičtení matice ζI k $\bar{H}_+$,

Corr-1 – jednoduchá korekce, $\eta_q = 1$ (i.e. $q = s$),

Corr-2 – korekční formule, $\eta_q = 1$.

Corr-q – korekční formule, η_q z analogie s posunutými VMM:

$$\eta_q = \min\left[1, \max\left[0, 1 + \frac{1}{\kappa}\left(1 + \frac{1}{\kappa}\right)\left(\frac{1.2\,\zeta_-}{\zeta_- + \zeta} - 1\right)\right]\right], \ \kappa = \zeta\frac{y^T y}{b}.$$

Uvádíme celkový **počet vyčíslení** funkčních hodnot (a také gradientů), s případným počtem nevyřešených úloh - kdy počet vyčíslení přesáhl 19000 (před -):

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	170460	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
3.0	1-54714	28641	24634	24675	2-68680	1-54531	45366	44785
BNS	18444				33131			

Tabulka 1. Porovnání různých typů korekce.

V tabulkách $2, 3$ uvádíme **rozdíly** $n_{p,q} - n_{BNS}$, kde $n_{p,q}$ je celkový počet vyčíslení funkčních hodnot (a také gradientů) pro vybrané hodnoty η_p a η_q užitím korekční formule a n_{BNS} je počet vyčíslení pro metodu BNS (záporné hodnoty značí, že naše metoda dává lepší výsledek než BNS).

V posledním řádku je uveden tento rozdíl pro η_q dané týmž **vzorcem** jako u tabulky 1.

	η_p						
η_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
1.5	-596	-436	-912	-937	-770	-285	307
2.0	150	-396	85	259	336	222	684
3.0	7698	5036	4903	4337	4218	3577	3541
vzorec	-771	-1263	-1280	-1423	-1368	-1020	-531

Tabulka 2. Srovnání s BNS pro N=500.

η_q	η_p						
	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
1.5	377	-181	-29	908	1323	441	1310
2.0	2164	767	994	2035	2577	2869	3036
3.0	4570	4457	3423	4106	5172	4430	4818
vzorec	1306	-1257	-2347	-2329	-632	-1746	-675

Tabulka 3. Srovnání s BNS pro N=1000.

Vidíme, že zde LM VMM z naší třídy pro vhodnou volbu parametru η_p (např. $\eta_p = 0.7$) a η_q (např. η_q dané uvedeným vzorcem) dávají **lepší výsledky** než metoda BNS.

Pro lepší srovnání s metodou BNS - další testy pro úlohy z často užívané kolekce **CUTE** (Bongartz, Conn, Gould, Toint, ACM TOMS 21 (1995) 123-160) s proměnnou dimenzí N a přesností $\|g(x^\star)\|_\infty \leq 10^{-6}$. V tabulce 4 jsou **podíly** $(n_{p,q} - n_{BNS})/(n_{p,q}+n_{BNS})*100$, kde $n_{p,q}$ je celkový počet vyčíslení funkčních hodnot (= gradientů) pro $\eta_p = \eta_q = 0.5$ a korekční formuli a n_{BNS} je počet vyčíslení pro metodu BNS (záporné hodnoty značí, že naše metoda dává lepší výsledek než BNS), ostatní parametry jako dříve.

Úloha	N	%	Úloha	N	%	Úloha	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

Tabulka 4: Srovnání počtu vyčíslení s BNS pro CUTE.

DĚKUJI ZA POZORNOST !