Jan Vlček; Ladislav Lukšan

A modified limited-memory BNS method for unconstrained minimization derived from the conjugate directions idea

In: Jan Chleboun and Petr Přikryl and Karel Segeth and Jakub Šístek and Tomáš Vejchodský (eds.): Programs and Algorithms of Numerical Mathematics, Proceedings of Seminar. Dolní Maxov, June 8-13, 2014. Institute of Mathematics AS CR, Prague, 2015. pp. 237–243.

Persistent URL: http://dml.cz/dmlcz/702689

Programs and Algorithms of Numerical Matematics 17
J. Chleboun, P. Přikryl, K. Segeth, J. Šístek, T. Vejchodský (Eds.)
Institute of Mathematics AS CR, Prague 2015

A MODIFIED LIMITED-MEMORY BNS METHOD FOR UNCONSTRAINED MINIMIZATION DERIVED FROM THE CONJUGATE DIRECTIONS IDEA

Jan Vlček[1], Ladislav Lukšan[1,2]
[1]Institute of Computer Science, Academy of Sciences of the Czech Republic,
Pod Vodárenskou věží 2, 182 07 Prague 8, Czech Republic
[2]Technical University of Liberec, Hálkova 6, 461 17 Liberec, Czech Republic

Abstract

A modification of the limited-memory variable metric BNS method for large scale unconstrained optimization of the differentiable function $f : \mathcal{R}^N \to \mathcal{R}$ is considered, which consists in corrections (based on the idea of conjugate directions) of difference vectors for better satisfaction of the previous quasi-Newton conditions. In comparison with [11], more previous iterations can be utilized here. For quadratic objective functions, the improvement of convergence is the best one in some sense, all stored corrected difference vectors are conjugate and the quasi-Newton conditions with these vectors are satisfied. The algorithm is globally convergent for convex sufficiently smooth functions and our numerical experiments indicate its efficiency.

1. Introduction

The BNS method (see [3]) belongs to the variable metric (VM) or quasi-Newton (QN) line search iterative methods, see [9], [10]. They start with an initial point $x_0 \in \mathcal{R}^N$ and generate iterations $x_{k+1} \in \mathcal{R}^N$ by the process $x_{k+1} = x_k + s_k$, $s_k = t_k d_k$, $k \geq 0$, where usually the direction vector $d_k \in \mathcal{R}^N$ is $d_k = -H_k g_k$, matrix H_k is symmetric positive definite and a stepsize $t_k > 0$ is chosen in such a way that

$$f_{k+1} - f_k \leq \varepsilon_1 t_k g_k^T d_k, \qquad g_{k+1}^T d_k \geq \varepsilon_2 g_k^T d_k, \qquad k \geq 0 \tag{1}$$

(the Wolfe line search conditions, see [10]), where $0 < \varepsilon_1 < 1/2$, $\varepsilon_1 < \varepsilon_2 < 1$, $f_k = f(x_k)$, $g_k = \nabla f(x_k)$; typically H_0 is a multiple of I and H_{k+1} is obtained from H_k by a VM update to satisfy the QN condition (see [9]) $H_{k+1} y_k = s_k$, $y_k = g_{k+1} - g_k$, $k \geq 0$.

Among VM methods, the BFGS method, see [9], [10], belongs to the most efficient; it preserves positive definite VM matrices and can be easily modified for large-scale optimization; the BNS and L-BFGS (see [5], [6] - subroutine PLIS) methods represent its well-known limited-memory adaptations. In every iteration, we repeatedly update an initial approximation of the inverse Hessian matrix $\zeta_k I$, $\zeta_k > 0$, by the BFGS method, using $\tilde{m} + 1$ couples of vectors $(s_{k-\tilde{m}}, y_{k-\tilde{m}}), \ldots, (s_k, y_k)$ successively (without forming approximations of the inverse Hessian matrix explicitly),

where $\tilde{m}=\min(k, m-1)$ and $m>1$ is a given parameter. In the case of the BNS method, the direction vector can be calculated without computing matrix H_+, see [3], by

$$-H_+g_+ = -\zeta g_+ - S\Big[U^{-T}\Big((D+\zeta Y^TY)U^{-1}S^Tg_+ - \zeta Y^Tg_+\Big)\Big] + Y\Big[\zeta U^{-1}S^Tg_+\Big], \quad (2)$$

(we often omit index k and replace indices $k+1, k-1$ by the symbols $+, -$ for simplification), where for $k \geq 0$ we denote $b_k = s_k^Ty_k$ and $S_k = [s_{k-\tilde{m}}, \ldots, s_k]$, $Y_k = [y_{k-\tilde{m}}, \ldots, y_k]$, $D_k = \mathrm{diag}[b_{k-\tilde{m}}, \ldots, b_k]$, $(U_k)_{i,j} = (S_k^TY_k)_{i,j}$ for $i \leq j$, $(U_k)_{i,j} = 0$ otherwise (an upper triangular matrix).

The concept of conjugacy plays an important role in optimization methods based on quadratic models, see e.g. [10]. We generalize the approach presented in [11], using vectors from more previous iterations to correct vectors s, y. Unlike [11], we use the BNS concept to calculate the direction vector, since then the increase in the number of required arithmetic operations can be relatively small. We use corrected quantities $\tilde{s}_k, \tilde{y}_k, \tilde{b}_k, \tilde{H}_k$, $k \geq 0$, defined by $\tilde{s}_0 = s_0$, $\tilde{y}_0 = y_0$, $\tilde{b}_0 = b_0$, $\tilde{H}_0 = I$ and

$$\tilde{s}_k = s_k + \hat{\underline{S}}_k\sigma_k, \quad \tilde{y}_k = y_k + \hat{\underline{Y}}_k\eta_k, \quad \tilde{b}_k = \tilde{s}_k^T\tilde{y}_k, \quad k > 0, \quad (3)$$

where matrices $\hat{\underline{S}}_k, \hat{\underline{Y}}_k$ contain some columns of $\tilde{\underline{S}}_k = [\tilde{s}_{k-\tilde{m}}, \ldots, \tilde{s}_{k-1}]$, $\tilde{\underline{Y}}_k = [\tilde{y}_{k-\tilde{m}}, \ldots, \tilde{y}_{k-1}]$ (we denote a set of indices i of these selected vectors $\tilde{s}_i, \tilde{y}_i$ by $\underline{\mathcal{I}}_k$ and $\mathcal{I}_k = \underline{\mathcal{I}}_k \cup \{k\}$; it can be $\underline{\mathcal{I}}_k = \emptyset$, in which case we set $\tilde{s}_k = s_k$, $\tilde{y}_k = y_k$, $\tilde{b}_k = b_k$) and σ_k, η_k are chosen in such a way that $\tilde{b}_k > 0$. Positive definite matrix $\tilde{H}_+$ is obtained by analogy to H_+, using corrected difference vectors. Note that matrix $\tilde{H}_+$ satisfies the QN condition $\tilde{H}_+\tilde{y} = \tilde{s}$ and that the direction vector $\tilde{d}_+ = -\tilde{H}_+g_+$ (and consequently, also an auxiliary vector $\tilde{Y}^T\tilde{H}_+g_+$) can be calculated by analogy to (2).

In Section 2 we investigate the BFGS update with corrected difference vectors

$$\ddot{H}_+ = (1/\tilde{b})\tilde{s}\tilde{s}^T + \tilde{V}\ddot{H}\tilde{V}^T, \quad \tilde{V} = I - (1/\tilde{b})\tilde{s}\tilde{y}^T, \quad (4)$$

where $\ddot{H}$ is any symmetric positive definite matrix, and discuss the choice of parameters σ, η. In Section 3 we show properties of $\ddot{H}_+$ and a role of unit stepsizes for quadratic functions. Application to the corrected BNS method and the corresponding algorithm are described in Section 4. Global convergence of the algorithm is established in Section 5 and numerical results are reported in Section 6. We will denote the Frobenius matrix norm by $\|\cdot\|_F$, the spectral matrix norm by $\|\cdot\|$ and the Euclidean vector norm by $|\cdot|$. Details and proofs of assertions can be found in [13].

2. Derivation of the method

Assuming that set $\underline{\mathcal{I}}$ is non-empty, we will investigate the influence of the correction parameters σ, η on properties of matrix $\ddot{H}_+$, given by (4). For our purpose, the satisfaction of the QN conditions $\ddot{H}_+\hat{Y} = \hat{S}$, $\hat{S} = [\hat{\underline{S}}, \tilde{s}]$, $\hat{Y} = [\hat{\underline{Y}}, \tilde{y}]$, plays a crucial role. We will suppose that the auxiliary QN conditions $\ddot{H}\hat{\underline{Y}} = \hat{\underline{S}}$ are satisfied (thus matrix $\hat{\underline{S}}^T\hat{\underline{Y}} = \hat{\underline{Y}}^T\ddot{H}\hat{\underline{Y}}$ is symmetric) and give a technique which guarantees the satisfaction of these conditions for a suitable matrix $\ddot{H}$. We denote $\ddot{B} = \ddot{H}^{-1}$, $\ddot{B}_+ = \ddot{H}_+^{-1}$, $\ddot{a} = \tilde{y}^T\ddot{H}\tilde{y}$.

The following lemma shows that, under some assumptions, conditions $\ddot{H}_+\tilde{y}_i=\tilde{s}_i$ are equivalent to the conjugacy of vector $\tilde{s}$ with vectors $\tilde{s}_i$ with respect to $\ddot{B}, \ddot{B}_+$, i.e. $\tilde{s}^T\ddot{B}\tilde{s}_i = \tilde{s}^T\ddot{B}_+\tilde{s}_i = 0$, $i \in \underline{\mathcal{I}}$, or $\hat{\underline{S}}^T\tilde{y} = \hat{\underline{Y}}^T\tilde{s} = 0$; these equations can be easily solved.

Lemma 1. *Let $\ddot{H}$ be any symmetric positive definite matrix satisfying $\ddot{H}\underline{\hat{Y}} = \underline{\hat{S}}$, matrix $\ddot{H}_+$ be given by (4) and let $\tilde{b} > 0$. Then $\ddot{H}_+$ is symmetric positive definite. If vectors $\tilde{s}$, $\ddot{H}\tilde{y}$ are linearly independent then $\ddot{H}_+\hat{Y} = \hat{S}$ if and only if $\hat{\underline{S}}^T\tilde{y} = \hat{\underline{Y}}^T\tilde{s} = 0$.*

Lemma 2. *Let matrix $\hat{\underline{S}}^T\hat{\underline{Y}}$ be nonsingular. Then the unique solution (σ, η) to $\hat{\underline{S}}^T\tilde{y} = \hat{\underline{Y}}^T\tilde{s} = 0$ is (σ^*, η^*), where $\sigma^* = -\left(\hat{\underline{Y}}^T\hat{\underline{S}}\right)^{-1}\hat{\underline{Y}}^T s$, $\eta^* = -\left(\hat{\underline{S}}^T\hat{\underline{Y}}\right)^{-1}\hat{\underline{S}}^T y$.*

Theorem 1 shows the variational characterizations of the choice $\sigma = \sigma^*$, $\eta = \eta^*$ also for non-quadratic functions, see also Theorem 3. Assumptions of Theorem 2 give our simple strategy for choosing matrices $\hat{S}, \hat{Y}$, which guarantees the satisfaction of the QN conditions $\tilde{H}_{k+1}\hat{Y}_k = \hat{S}_k$ and the corresponding auxiliary QN conditions.

Theorem 1. *Let $\tilde{b} > 0$ for $(\sigma, \eta) = (\sigma^*, \eta^*)$, matrix $\hat{\underline{S}}^T\hat{\underline{Y}}$ be nonsingular, matrices $\ddot{H}, \ddot{H}_+$ satisfy the same assumptions as in Lemma 1 and define $\mathcal{S}(\hat{\underline{S}}, \hat{\underline{Y}}) = \{(\sigma, \eta) : \hat{\underline{S}}^T\tilde{y} = \hat{\underline{Y}}^T\tilde{s}\}$. If we have any symmetric positive definite matrix $\ddot{G}$ such that $\ddot{G}\hat{\underline{S}} = \hat{\underline{Y}}$ and $\ddot{G}(s + \hat{\underline{S}}\ddot{\sigma}) = y + \hat{\underline{Y}}\ddot{\eta}$ for some $(\ddot{\sigma}, \ddot{\eta}) \in \mathcal{S}(\hat{\underline{S}}, \hat{\underline{Y}})$, then within $(\sigma, \eta) \in \mathcal{S}(\hat{\underline{S}}, \hat{\underline{Y}})$, values $\|\ddot{G}^{1/2}\ddot{H}_+\ddot{G}^{1/2} - I\|_F^2$ and $\tilde{b}$ are minimized by the choice $\sigma = \sigma^*$, $\eta = \eta^*$.*

Theorem 2. *Suppose that each set $\underline{\mathcal{I}}_k$, $k > 0$, is chosen in such a way that $\underline{\mathcal{I}}_k \subset \mathcal{I}_{k-1}$, $\tilde{b}_k > 0$ and $\hat{\underline{S}}_k^T\tilde{y}_k = \hat{\underline{Y}}_k^T\tilde{s}_k = 0$ in case that $\underline{\mathcal{I}}_k \neq \emptyset$. Then for $k > 0$: $\tilde{s}_i^T\tilde{y}_j = \tilde{y}_i^T\tilde{s}_j = 0$, $i \in \underline{\mathcal{I}}_k$, $i < j \leq k$, the QN conditions $\tilde{H}_{k+1}\hat{Y}_k = \hat{S}_k$ are satisfied and the auxiliary QN conditions $\ddot{H}_k\hat{\underline{Y}}_k = \hat{\underline{S}}_k$ are satisfied for $\underline{\mathcal{I}}_k \neq \emptyset$ with those matrices $\ddot{H}_k$ by the BFGS updating (4) of which we get matrices $\tilde{H}_{k+1} = \ddot{H}_{k+1}$.*

The first assertion of the theorem implies that all matrices $\hat{S}^T\hat{Y}$ are diagonal and thus many results can be simplified. E.g. vectors σ^*, η^* have components $-s^T\tilde{y}_i/\tilde{b}_i, -\tilde{s}_i^T y/\tilde{b}_i$, $i \in \underline{\mathcal{I}}$, and a damage of the QN condition with non-corrected vectors caused by our corrections and value $\tilde{b}$ for $(\sigma, \eta) = (\sigma^*, \eta^*)$ can be written:

$$(\ddot{H}_+ y - s)^T\ddot{B}_+(\ddot{H}_+ y - s) = b\sum\nolimits_{i\in\underline{\mathcal{I}}}(\tilde{s}_i^T y - s^T\tilde{y}_i)^2/(b\tilde{b}_i),\ \tilde{b} = b - \sum\nolimits_{i\in\underline{\mathcal{I}}} s^T\tilde{y}_i\,\tilde{s}_i^T y/\tilde{b}_i. \quad (5)$$

3. Results for quadratic functions

Here we suppose that f is a quadratic function with a symmetric positive definite Hessian G and $\eta_k = \sigma_k$, $k > 0$, which yields $\tilde{y}_k = G\tilde{s}_k$, as for non-corrected vectors. The following lemma and theorem show that for the choice $\sigma = \sigma^*$, the improvement of convergence is the best in some sense for linearly independent direction vectors.

Lemma 3. *Let f be a quadratic function $f(x) = \frac{1}{2}(x - \bar{x})^T G(x - \bar{x})$, $\bar{x} \in \mathcal{R}^N$, with a symmetric positive definite matrix G and all columns of $[\tilde{\underline{S}}, s]$ be linearly independent. Then for any selection of $\hat{\underline{S}}$, $\hat{\underline{Y}}$ from $\tilde{\underline{S}}$, $\tilde{\underline{Y}}$, matrix $\hat{\underline{S}}^T\hat{\underline{Y}}$ is symmetric positive definite, value σ^* is well defined by Lemma 2 and $\tilde{b} > 0$ for any $\sigma = \eta$.*

Theorem 3. *Let $\ddot{H}$ be any symmetric positive definite matrix satisfying $\ddot{H}\underline{\hat{Y}} = \underline{\hat{S}}$ and suppose that $\sigma = \eta$ and that the assumptions of Lemma 3 are satisfied. Then $\tilde{b} > 0$ and the choice $\sigma = \sigma^*$ implies $\ddot{H}_+ y = s$ and minimizes values $\tilde{b}$ and $\|G^{1/2}\ddot{H}_+ G^{1/2} - I\|_F$ as a function of σ, where matrix $\ddot{H}_+$ is defined by update (4) of $\ddot{H}$.*

Theorem 4 describes a situation when the case $\sigma = \sigma^*$ occurs in all iterations. Comparing these results with those given in [11] (Theorem 3.2) for the unit stepsizes, we see that they are similar. Theorem 5 gives an interesting explanation.

Theorem 4. *Let the assumptions of Lemma 3 be satisfied with the columns of every matrix $[\underline{\tilde{S}}_k, s_k]$ linearly independent and let always $\underline{\hat{S}}_k = \underline{\tilde{S}}_k$, $\underline{\hat{Y}}_k = \underline{\tilde{Y}}_k$, $\sigma_k = \sigma_k^*$, $k > 0$. Then all columns of $\tilde{S}_k$ are G-conjugate, i.e. matrices $\tilde{S}_k^T \tilde{Y}_k$ are diagonal and all QN conditions $\tilde{H}_{k+1}\tilde{Y}_k = \tilde{S}_k$, $\ddot{H}_k \underline{\tilde{Y}}_k = \underline{\tilde{S}}_k$, are satisfied, with those matrices $\ddot{H}_k$ by the BFGS updating (4) of which we get matrices $\tilde{H}_{k+1} = \ddot{H}_{k+1}$, $k > 0$.*

Theorem 5. *Let $\tilde{H}, \tilde{H}_+$ be symmetric positive definite matrices satisfying $\tilde{H}\underline{\hat{Y}} = \underline{\hat{S}}$, $\tilde{H}_+ \underline{\hat{Y}} = \underline{\hat{S}}$, $d = -\tilde{H}g$, $d_+ = -\tilde{H}_+ g_+$, $\sigma = \eta$, $t = 1$ and the assumptions of Lemma 3 be satisfied. Then $\underline{\hat{S}}^T y_+ = \underline{\hat{Y}}^T s_+ = 0$, i.e. all columns of $\underline{\hat{S}}$ are G-conjugate with s_+.*

4. Implementation

It is important to say that not all vectors $\tilde{s}_i, \tilde{y}_i$, $i \in \underline{\mathcal{I}}$, are suitable as correction vectors. Principally, we do not use vectors $\tilde{s}_i, \tilde{y}_i$, $k - \tilde{m} \le i < k$, $k > 0$, for the correction process (i.e. we decide that $i \notin \underline{\mathcal{I}}_k$) if $\tilde{b}_k \le 0$, if resultant values $b_k/\tilde{b}_k$, $b_k/\ddot{a}_k$, $b_k/\tilde{s}_k^T \ddot{B}_k \tilde{s}_k$ or $(\tilde{s}_i^T y_k - s_k^T \tilde{y}_i)^2/(b_k \tilde{b}_i)$ (see (5)) are too great or if $i \notin \mathcal{I}_{k-1}$, see Theorem 2.

In order to prove global convergence, we also exclude index i from $\underline{\mathcal{I}}$ if values $|\tilde{s}_i|/|s_i|$, $|\tilde{y}_i|/|y_i|$ are too great. Note that these values were rarely greater than 50 in our numerical experiments with $N = 5000$. Further, Theorem 5 indicates that an influence of the second and further correction vectors can be small. Thus for $i < k-1$, $k > 0$, we should not correct if a benefit of corrections is negligible, see [13] for details.

Algorithm 1 (without indices elimination details and stopping criteria)

Data: A number $m > 1$ of VM updates per iteration, line search and correction parameters and a maximum number of correction vectors $n \in [0, m-1]$.

Step 0: Initiation. Choose starting point $x_0 \in \mathcal{R}^N$, define starting matrix $\tilde{H}_0 = I$ and direction vector $d_0 = -g_0$ and initiate iteration counter k to zero.

Step 1: Line search. Set $\tilde{m} = \min(k, m-1)$. Compute $x_{k+1} = x_k + t_k d_k$, where t_k satisfies (1), $g_{k+1} = \nabla f(x_{k+1})$, $s_k = t_k d_k$, $y_k = g_{k+1} - g_k$, $b_k = s_k^T y_k$, $\zeta_k = b_k / y_k^T y_k$. If $k = 0$ set $\tilde{s}_k = s_k$, $\tilde{y}_k = y_k$, $\tilde{b}_k = \tilde{s}_k^T \tilde{y}_k$, $\mathcal{I}_k = \{0\}$, $\tilde{S}_k = [\tilde{s}_k]$, $\tilde{Y}_k = [\tilde{y}_k]$, $\tilde{S}_k^T \tilde{Y}_k = [\tilde{s}_k^T \tilde{y}_k]$, $\tilde{Y}_k^T \tilde{Y}_k = [\tilde{y}_k^T \tilde{y}_k]$, compute $\tilde{S}_k^T g_{k+1}, \tilde{Y}_k^T g_{k+1}$ and go to Step 5. Compute $\underline{\tilde{S}}_k^T g_{k+1}$, $\underline{\tilde{Y}}_k^T g_{k+1}$, $\underline{\tilde{Y}}_k^T s_k = -t_k \underline{\tilde{Y}}_k^T \tilde{H}_k g_k$, $\underline{\tilde{S}}_k^T y_k = \underline{\tilde{S}}_k^T g_{k+1} - \underline{\tilde{S}}_k^T g_k$ and $\underline{\tilde{Y}}_k^T y_k = \underline{\tilde{Y}}_k^T g_{k+1} - \underline{\tilde{Y}}_k^T g_k$.

Step 2: Elimination of indices. Set $\underline{\mathcal{I}}_k = \{i \in \mathcal{I}_{k-1} : i \ge k - n\}$. Eliminate non-suitable indices from $\underline{\mathcal{I}}_k$. If $\underline{\mathcal{I}}_k = \emptyset$ go to Step 4, otherwise form matrices $\underline{\hat{S}}_k, \underline{\hat{Y}}_k$.

Step 3: Correction. Compute $(\sigma_k)_i = -s_k^T \tilde{y}_i / \tilde{b}_i$, $(\eta_k)_i = -\tilde{s}_i^T y_k / \tilde{b}_i$ for $i \in \underline{\mathcal{I}}_k$ and $\tilde{s}_k, \tilde{y}_k, \tilde{b}_k$ by (3). Set $\mathcal{I}_k = \underline{\mathcal{I}}_k \cup \{k\}$.

Step 4: Matrix updating. Similarly as in [3] form matrices $\tilde{S}_k$, $\tilde{Y}_k$, $\tilde{S}_k^T\tilde{Y}_k$, $\tilde{Y}_k^T\tilde{Y}_k$.

Step 5: Direction vector. Compute $d_{k+1} = -\tilde{H}_{k+1}g_{k+1}$ by the BNS method with vectors $(\tilde{s}_{k-\tilde{m}}, \tilde{y}_{k-\tilde{m}}), \ldots, (\tilde{s}_k, \tilde{y}_k)$ and an auxiliary vector $\tilde{Y}_k\tilde{H}_{k+1}g_{k+1}$, see Section 1. Set $k:=k+1$. If $k \geq m$ delete the first column of $\tilde{S}_{k-1}$, $\tilde{Y}_{k-1}$ and the first row and column of $\tilde{S}_{k-1}^T\tilde{Y}_{k-1}$, $\tilde{Y}_{k-1}^T\tilde{Y}_{k-1}$ to form matrices $\underline{\tilde{S}}_k, \underline{\tilde{Y}}_k, \underline{\tilde{S}}_k^T\underline{\tilde{Y}}_k, \underline{\tilde{Y}}_k^T\underline{\tilde{Y}}_k$. Go to Step 1.

5. Global convergence

Assumption 1. *The objective function $f : \mathcal{R}^N \to \mathcal{R}$ is bounded from below and uniformly convex with bounded second-order derivatives (i.e. $0 < \underline{G} \leq \underline{\lambda}(G(x)) \leq \overline{\lambda}(G(x)) \leq \overline{G} < \infty$, $x \in \mathcal{R}^N$, where $\underline{\lambda}(G(x))$ and $\overline{\lambda}(G(x))$ are the lowest and the greatest eigenvalues of the Hessian matrix $G(x)$).*

Theorem 6. *If the objective function f satisfies Assumption 1, Algorithm 4.1 generates a sequence $\{g_k\}$ that satisfies $\lim_{k\to\infty}|g_k|=0$ or terminates with $g_k=0$ for some k.*

6. Numerical experiments

We demonstrate the influence of vector corrections on the number of evaluations and computational time, using the following collections of test problems: **Test 11** [8] (55 modified problems from CUTE collection [2] with $N = 1000 - 5000$, computed repeatedly ten times), test from [1], termed **Test 12** here, 73 problems, $N = 10000$, **Test 25** [7] (68 problems), $N = 10000$. The source texts and the corresponding reports can be downloaded from `camo.ici.ro/neculai/ansoft.htm` (Test 12) and `www.cs.cas.cz/luksan/test.html` (Tests 11 and 25).

Table 1 contains the total number of function evaluations (NFV) and the total computational time in seconds (Time) for the following limited-memory methods: L-BFGS [5], method from [11] and new Algorithm 1 for $n = 2, 4$, all implemented in the system UFO [12]. We have used $m=5$ and the final precision $\|g(x^\star)\|_\infty \leq 10^{-6}$.

Method	Test 11		Test 12		Test 25	
	NFV	Time	NFV	Time	NFV	Time
L-BFGS	80539	10.361	119338	50.88	502966	429.01
Alg. 4.1 in [11]	64395	9.614	67619	32.61	325441	318.71
Alg. 1, $n=2$	62770	8.795	67372	31.06	302908	302.62
Alg. 1, $n=4$	64127	8.977	66403	30.77	308847	298.05

Table 1: Comparison of the selected methods

For Test 25, we also compare these methods by using performance profiles [4]. Value $\rho_M(0)$ is the percentage of the test problems for which method M is the best and value $\rho_M(\tau)$ for τ large enough is the percentage of the problems that method M can solve. Performance profiles show the relative efficiency and reliability of the methods: the higher is the particular curve, the better is the corresponding method.

Acknowledgements

This work was supported by the Grant Agency of the Czech Republic, project No. 13-06684S, and the Institute of Computer Science of the AS CR (RVO: 67985807).

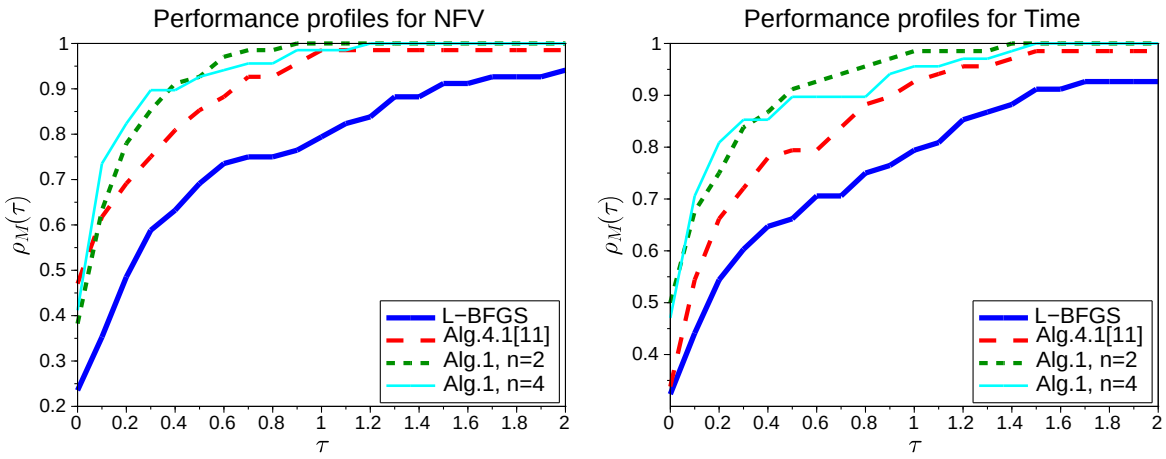

Figure 1: Comparison of $\rho_M(\tau)$ for Test 25 (68 problems) and various methods.

References

[1] Andrei, N.: An unconstrained optimization test functions collection. Advanced Modeling and Optimization **10** (2008), 147–161.

[2] Bongartz, I., Conn, A. R., Gould, N., and Toint, P. L.: CUTE: constrained and unconstrained testing environment. ACM Trans. Math. Software **21** (1995), 123–160.

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

[4] Dolan, E. D. and Moré, J. J.: Benchmarking optimization software with performance profiles. Math. Program. **91** (2002), 201–213.

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

[6] Lukšan, L., Matonoha, C., and Vlček, J.: Algorithm 896: LSA – Algorithms for Large-Scale Optimization. ACM Trans. Math. Software **36** (2009), 16:1–16:29.

[7] Lukšan, L., Matonoha, C., and Vlček, J.: Sparse test problems for unconstrained optimization. Report V-1064, ICS AS CR, Prague, 2010.

[8] Lukšan, L., Matonoha, C., and Vlček, J.: Modified CUTE problems for sparse unconstrained optimization. Report V-1081, ICS AS CR, Prague, 2010.

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

[10] Nocedal, J. and Wright S. J.: *Numerical optimization*. Springer-Verlag, New York, 1999.

[11] Vlček, J. and Lukšan, L.: A conjugate directions approach to improve the limited-memory BFGS method. Appl. Math. Comput. **219** (2012), 800–809.

[12] Lukšan, L., Tůma, M., Vlček, J., Ramešová, N., Šiška, M., Hartman, J., and Matonoha, C.: UFO 2013. Interactive System for Universal Functional Optimization. Report V-1191, ICS AS CR, Prague, 2013, `www.cs.cas.cz/luksan/ufo.html`.

[13] Vlček, J. and Lukšan, L.: A modified limited-memory BNS method for unconstrained minimization based on the conjugate directions idea. Report V-1203, ICS ASCR, Prague, 2014, `www.cs.cas.cz/vlcek/reports.html`.