Yokohama Publishers
ISSN 1349-8169 ONLINE JOURNAL

Pacific Journal of Optimization
Vol. 14, No. 4, pp. 551–565, 2018

A NEW NONSMOOTH TRUST-REGION METHOD EQUIPPED WITH A LINE SEARCH FOR MINIMIZING LOCALLY LIPSCHITZ FUNCTIONS

Z. Akbari*, M. Reza Peyghami† and R. Yousefpour

Abstract: In this paper, we employ a line search technique in the framework of trust region methods and propose a new nonsmooth trust region method for minimization of locally Lipschitz functions. Despite existing nonsmooth trust region methods, our new proposed approach performs a line search along a descent direction rather than re-solving the subproblem in the lack of sufficient reduction. This causes a significant decrease in the number of subproblem that need to be solved and, consequently, a decrease in the number of function evaluations. Under some standard assumptions, the global convergence property of the new proposed method is established for minimization of locally Lipschitz functions. The proposed algorithm is implemented in MATLAB environment and applied on some test problems. Numerical results confirm the efficiency of the new approach in comparison with some existing trust region methods for nonsmooth functions.

Key words: *nonsmooth trust region method, line search method, Lipschitz functions, CG-Steihaug method*

Mathematics Subject Classification: *49J52, 90C26*

1 Introduction

In this paper, we deal with the following unconstrained optimization problem

$$\min_{x\in\mathbb{R}^n} f(x), \tag{1.1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is a locally Lipschitz and therefore almost everywhere differentiable function.

Several methods, including the bundle method [12], the gradient sampling method [15], the descent direction algorithm [14], discrete gradient method [4] and trust region methods [2, 7, 9, 18, 20], have been developed in the literature for solving problems of the form (1.1). Among them, trust region (TR) methods received much more attention due to their strong convergence property and robustness; see, e.g., [16, 21, 6] and the references therein.

Although TR methods have been well developed for solving smooth optimization problems, they have been modified for the nonsmooth optimization context. The earliest version

*The first and last authors are thankful from University of Mazandaran for partially supporting this research.

†Corresponding author. The second author would like to thank the Research Council of K.N. Toosi University of Technology for supporting this work.

was suggested by Fletcher in [9] for solving a special nonsmooth composite function. Then, Powell in [18] proposed a nonsmooth trust region method for solving another type of composite functions in which a differentiable function is composed with a nonsmooth convex function. Global convergence property of Powell's method was investigated by Yuan in [20]. A crucial aspect in TR methods for nonsmooth optimization is the way of constructing an appropriate quadratic model for the TR subproblem. Under some conditions on the quadratic model, Dennis et al. in [7] presented a unified approach for global convergence property of the TR methods for a class of nonsmooth functions, named as regular functions. They showed that these conditions are fulfilled by several nonsmooth TR methods in the literature. By relaxing the regularity condition in [7], Qi and Sun in [19] proposed a quadratic model for locally Lipschitz functions in which the first order information is replaced by the so called *iteration function* in its structure. Based on the proposed model, they constructed a TR algorithm for locally Lipschitz functions and established its global convergence property. Despite several nonsmooth TR methods, a local model that is practically and efficiently implementable on locally Lipschitz functions has been rarely presented in the literature. Recently, Akbari et al [2] introduced a practical nonsmooth TR method for locally Lipschitz functions. In their suggested method, the quadratic model was constructed based on the approximation of the steepest descent direction, as proposed in [14]. They established the global convergence property of their algorithm under some standard assumptions and compared it with some other nonsmooth optimization algorithms in terms of number of iterations and function evaluations. A survey on the reported numerical results shows that the proposed algorithm has acceptable performance on medium-scale problems and performs well on large-scale problems.

Solving the TR subproblem is the most computationally expensive part in numerical performance of TR methods as it may require solving one or more linear systems or an iterative process with high computational cost. Therefore, the concept of line search techniques along the rejected trial steps have been exploited in the structure of TR methods for smooth optimization to reduce the number of subproblems that need to be solved; see e.g. [1, 3, 17]. From this point of view, if the sufficient reduction in the objective function is not achieved by the trial step, a line search along the trial step, which is usually a decreasing direction for smooth problems, is performed instead of re-solving the subproblem. In this regard, none of trial steps are rejected and therefore the number of subproblem solving and function evaluations are reduced significantly.

In this paper, we exploit line search techniques in the structure of TR methods for minimizing locally Lipschitz functions. In fact, we equip the nonsmooth TR method as proposed in [2] with a backtracking line search technique in order to improve its practical performance. Unlike TR methods for smooth problems, the trial step generated by the TR subproblem might not be a descent direction for the objective function of nonsmooth problems. Therefore, a search along this direction may not cause any reduction in the objective function. In this case, at the current point, we introduce a descent direction and perform a line search in this direction whenever a sufficient reduction is not achieved by the trial step. Under some standard assumptions, we establish the global convergence property of the new algorithm for locally Lipschitz functions. Preliminary numerical results on some test problems show the efficiency and effectiveness of the new algorithm in comparison with that of proposed in [2], which in turn has better performance than some other existing methods in the literature, especially in the large-scale settings.

The rest of the paper is organized as follows: Section 2 is devoted to describe some preliminaries in the context of nonsmooth problems and to recall the proposed TR method in [2]. The structure of the new proposed nonsmooth TR algorithm for locally Lipschitz functions is constructed in Section 3. The global convergence property of this algorithm is established in Section 4 under some standard assumptions. An implementation of the proposed algorithm along with its numerical results are provided in Section 5. The paper ends up with some concluding remarks in Section 6.

2 Preliminaries

In this section, after a short review on the concepts of nonsmooth analysis, we recall the nonsmooth TR method as proposed in [2] for minimizing locally Lipschitz functions.

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a locally Lipschitz function. The Clarke generalized directional derivative of the function f at point x in the direction p, denoted by $f^\circ(x,p)$, is defined by [5]:

$$f^\circ(x,p) := \limsup_{y\to x, t\downarrow 0} \frac{f(y+tp)-f(x)}{t}.$$

Based on this definition, the Clarke generalized subdifferential is defined as follows [5]:

$$\partial f(x) := \left\{ v \in \mathbb{R}^n \ : \ f^\circ(x,p) \geq \ v^T p, \ \forall p \in \mathbb{R}^n \right\}.$$

It can be seen that $f^\circ(x,p) = \sup_{v\in\partial f(x)} v^T p$. Let L be the Lipschitz constant in the neighborhood of x. Then, we have:

$$\|v\|_2 \leq L, \quad \forall v \in \partial f(x).$$

If f is differentiable at x, then $\nabla f(x) \in \partial f(x)$. Furthermore, if f is continuously differentiable at x, then

$$\partial f(x) = \{\nabla f(x)\}.$$

If x is a minimizer (or stationary point) of f, then $0 \in \partial f(x)$.

For given $\varepsilon > 0$, the ε-subdifferential of the function f at point x is defined by [10]:

$$\partial_\varepsilon f(x) := \mathrm{conv}\{v \ : \ v \in \partial f(y), \ \|x-y\|_2 \leq \varepsilon\}.$$

If $0 \in \partial_\varepsilon f(x)$, then x is called as an ε-stationary point.

Now, let us briefly describe the nonsmooth TR algorithm, denoted by "Ntrust", for minimizing locally Lipschitz functions as proposed in [2]. In this algorithm, at first, an approximation of the steepest descent direction for locally Lipschitz functions at point x_k is computed by solving the following problem:

$$\min_{v\in\partial_{\Delta_k} f(x_k)} \|v\|_2, \tag{2.1}$$

where Δ_k is the TR radius. Let v_k be a solution of (2.1). Then, a sufficient reduction along v_k is achieved whenever the following inequality holds:

$$f(x_k - \Delta_k v_k) - f(x_k) \leq -c_1 \Delta_k \|v_k\|_2^2, \tag{2.2}$$

where $c_1 > 0$ is a constant. An approach for efficiently computing the solution v_k for the problem (2.1) has been proposed in [14], called as MY algorithm, which guarantees a sufficient reduction in the objective function, i.e. (2.2) holds. In the MY algorithm, $\partial_\varepsilon f(x)$ is approximated iteratively until a search direction which staisfies in the Armijo condition is found. Let W be a finite subset of $\partial_\varepsilon f(x)$ and its convex hull is considered as an aproximation of $\partial_\varepsilon f(x)$. In each iteration, the following minimization problem is solved:

$$v_0 = \arg \min_{v \in \mathsf{conv} W} \|v_2\|_2.$$

If $-v_0$ is satisfied the Armijo condition then $\mathsf{conv}W$ is considered as an acceptable approxiamtion for $\partial_\varepsilon f(x)$. Else a new element of $\partial_\varepsilon f(x)$ is added to W such that this element does not belong to $\mathsf{conv}W$ [14] (Algorithm 3.1). If $\|v_0\|_2$ is less than a predefined threshold or $-v_0$ is satisfied in the Armijo condition then My algorithm is terminated. This algorithm is terminated after finitely many iterations.

Using MY algorithm, the Ntrust method constructs a local quadratic model as below:

$$m(x_k, p) = f(x_k) + v_k^T p + \frac{1}{2} p^T B_k p,$$

where B_k is a positive definite matrix. Therefore, at point x_k, the trial step is computed by solving the following subproblem:

$$\begin{array}{ll} \min & m(x_k, p) = f(x_k) + v_k^T p + \frac{1}{2} p^T B_k p \\ s.t. & \|p\|_2 \leq \Delta_k. \end{array} \tag{2.3}$$

Using the so-called CG-Steihaug method [16], an approximate solution p_k is computed for the problem (2.3). Note that a sufficient reduction in the objective function along p_k is achieved whenever the following inequality holds:

$$f(x_k + p_k) - f(x_k) \leq c_1 v_k^T p_k.$$

As long as the sufficient reduction is not achieved, Ntrust reduces the TR radius and recomputes an approximation of steepest descent by MY algorithm. Soon after achieving a sufficient reduction, the TR ratio is computed by:

$$\rho_k = \frac{f(x_k + p_k) - f(x_k)}{m(x_k, p_k) - m(x_k, 0)}. \tag{2.4}$$

Based on the magnitude of ρ_k, the new point and the radius are updated appropriately. The structure of the Ntrust algorithm for minimizing locally Lipschitz functions is outlined in Algorithm 2.1.

Algorithm 2.1: A nonsmooth trust region algorithm (Ntrust) [2]

Step 0: Given $\Delta_0, \Delta_1 > 0$, $\eta > 0$, $\theta_\Delta, \delta_1, \theta_\delta \in (0, 1)$, $x_1 \in \mathbb{R}^n$, $\xi_1 \in \partial f(x_1)$, $c_4 > 1$, $0 < c_2 < c_3 < 1$ and $c_1 \in (0, 1)$, let $B_1 = I$ and $k = 1$.

Step 1: Apply the MY algorithm [14] at point x_k with the parameters $\varepsilon = \Delta_k$, $\delta = \delta_k$ and $c = c_1$, and obtain an approximate steepest descent direction v_k.

Step 2: If $\|v_k\|_2 \leq \eta$, then Stop. If $\|v_k\|_2 \leq \delta_k$, then set $\Delta_{k+1} = \theta_\Delta \Delta_k$, $x_{k+1} = x_k$, $\delta_{k+1} = \delta_k \theta_\delta$, $k := k + 1$ and go to Step 1. Else, set $\delta_{k+1} = \delta_k$ and go to Step 3.

Step 3: Solve the subproblem (2.3) to obtain the trial step p_k.

Step 4: If $f(x_k + p_k) - f(x_k) \leq c_1 v_k^T p_k$, then go to Step 5. Else, set $\Delta_{k+1} = \theta_\Delta \Delta_k$, $x_{k+1} = x_k$, $k := k+1$ and go to Step 1.

Step 5: Compute ρ_k using (2.4).
If $\rho_k > c_2$, then set $x_{k+1} = x_k + p_k$. Else set $x_{k+1} = x_k$.
If $\rho_k > c_3$ and $\|p_k\|_2 = \Delta_k$, then set $\Delta_{k+1} = \min\{\Delta_0, c_4\Delta_k\}$. Else if $\rho_k < c_2$, then set $\Delta_{k+1} = \theta_\Delta \Delta_k$. Else, set $\Delta_{k+1} = \Delta_k$.

Step 6: Select a subgradient $\xi_{k+1} \in \partial f(x_{k+1})$, and update the matrix B_k by the BFGS formula. Set $k := k+1$ and go to Step 1.

Remark 2.1. We select $\xi_{k+1} \in \partial f(x_{k+1})$, if the secant equation is satisfied and

$$y_k^T p_k > 0,$$

where $y_k = \xi_{k+1} - \xi_k$ and $\xi_k \in \partial f(x_k)$, then we update B_{k+1} by the BFGS formula as following:

$$B_{k+1} = B_k - \frac{B_k p_k p_k^T B_k}{p_k^T B_k p_k} + \frac{y_k y_k^T}{y_k^T p_k}.$$

Remark 2.2. Let B_k be a positive definite matrix. Then, the Cauchy point algorithm returns the following solution for the subproblem (2.3); see e.g. [16];

$$p_k = -\Delta_k \frac{v_k}{\|v_k\|_2}.$$

This solution implies a sufficient reduction in the objective function. Besides, if v_k satisfies

$$\frac{v_k^T v_k}{v_k^T B_k v_k} \|v_k\|_2 \geq \Delta_k,$$

then the CG-Steihaug method returns exactly the Cauchy point.

In the next section, by appropriately exploiting the concept of line search techniques in the framework of TR algorithms, we try to improve the efficiency of the Ntrust algorithm.

$\boxed{3}$ A New Nonsmooth TR Algorithm

In Ntrust algorithm, as long as a sufficient reduction is not achieved, i.e., $\rho_k < c_2$, the trial step p_k is rejected by the algorithm, the TR radius is shrunk, a new approximation of the steepest descent direction is computed by the MY algorithm and the subproblem is re-solved. This procedure causes an increase in the number of function evaluations, which in turn may harm the efficiency of the algorithm.
To prevent further subproblem re-solving, we develop a variant of the Ntrust algorithm which is equipped with a line search technique. It has to be noted that the computed trial direction in Step 3 of Algorithm 2.1 is not necessarily a descent direction for locally Lipschitz functions. Therefore, instead of performing a line search along this direction, whenever it is rejected, we perform a line search along the direction $-v_k$, which is a descent direction. More precisely, in Step 4 of Algorithm 2.1, if the following condition holds:

$$f(x_k + p_k) - f(x_k) > c_1 v_k^T p_k,$$

i.e., the sufficient decrease is not achieved, then the backtracking procedure, as outlined in Algorithm 3.1, is performed along the direction $-v_k$.

Algorithm 3.1: Backtracking procedure.

Given $c_1 > 0$, $\Delta_k > 0$, $\varrho \in (0,1)$, x_k and v_k, set $\alpha := 1$;

While $f\left(x_k - \alpha \frac{v_k}{\|v_k\|_2}\right) - f(x_k) > -c_1\alpha\|v_k\|_2$ and $\alpha_k \geq \Delta_k$
$\alpha := \varrho\alpha$

End(While)

Return $\alpha_k = \max\{\alpha, \Delta_k\}$.

Now, our new proposed nonsmooth TR algorithm for minimizing locally Lipschitz functions is constructed by modifying Step 4 of Algorithm 2.1 as below:

Step 4: If $f(x_k + p_k) - f(x_k) \leq c_1 v_k^T p_k$, then go to Step 5. Else, Apply Algorithm 3.1 in the direction $-v_k$ to obtain the step length $\alpha_k \in (0,1]$. Set $x_{k+1} = x_k - \alpha_k \frac{v_k}{\|v_k\|_2}$, $\Delta_{k+1} = \theta_\Delta \Delta_k$, $k := k+1$ and go to Step 6.

From now on, we call the new proposed nonsmooth TR algorithm as "Ltrust". In fact, the structure of Ltrust is exactly the same as Ntrust in which Step 4 is modified as above.

4 Convergence Analysis

In this section, our aim is to analyze the global convergence property of the Ltrust algorithm for minimizing locally Lipschitz functions. Note that, as Ltrust is a modification of Ntrust, the proofs in this section are very similar to those presented in [2].
Let $\mathcal{K}_1$ be the set of all successful iterations, i.e.,

$$\mathcal{K}_1 = \{k \ : \ \rho_k \geq c_2\},$$

and $\mathcal{K}_2$ be the set of all iterations that the line search is performed, i.e.,

$$\mathcal{K}_2 = \{k \ : \ f(x_k + p_k) - f(x_k) > c_1 v_k^T p_k\}.$$

Moreover, assume that $\mathcal{K}$ is the set of all iterations and $\bar{\mathcal{K}} = \mathcal{K}_1 \bigcup \mathcal{K}_2$.

Lemma 4.1. *Assume that the level set $\mathcal{L} := \{x : \ f(x) \leq f(x_1)\}$ is bounded and there exists a constant $M > 0$ so that $\|B_k\| \leq M$, for all k. Furthermore, let v_k be an approximation of the steepest descent direction, which is generated by the MY algorithm. If Ltrust does not terminate after finitely many iterations, then*

$$\liminf_{k\to\infty, k\in\bar{\mathcal{K}}} \|v_k\|_2 = 0. \tag{4.1}$$

Proof. Suppose that, on the contrary, there exists $\epsilon > 0$ so that, for all $k \in \bar{\mathcal{K}}$,

$$\|v_k\|_2 \geq \epsilon. \tag{4.2}$$

Using Theorem 4.4 in [16], at each iteration $k \in \mathcal{K}$, there exists $\alpha \in (0, 1)$ so that

$$m(x_k, 0) - m(x_k, p_k) \geq \alpha \|v_k\|_2 \min \left\{ \Delta_k, \frac{\|v_k\|_2}{\|B_k\|} \right\}. \tag{4.3}$$

Using (4.3) and the stopping criteria of the MY algorithm [14], for all $k \in \mathcal{K}_1$, we have:

$$\begin{aligned} 1 - \rho_k &= \frac{f(x_k + p_k) - m(x_k, p_k)}{m(x_k, 0) - m(x_k, p_k)} \\ &= \frac{f(x_k + p_k) - f(x_k) - v_k^T p_k - \frac{1}{2} p_k^T B_k p_k}{m(x_k, 0) - m(x_k, p_k)} \\ &\leq \frac{(c_1 - 1) v_k^T p_k - \frac{1}{2} p_k^T B_k p_k}{\alpha \|v_k\|_2 \min \left\{ \Delta_k, \frac{\|v_k\|_2}{\|B_k\|} \right\}} \\ &\leq \frac{(1 - c_1) \|p_k\|_2}{\alpha \min \left\{ \Delta_k, \frac{\|v_k\|_2}{\|B_k\|} \right\}}, \end{aligned} \tag{4.4}$$

where the last inequality is obtained from the Cauchy-Schwartz inequality and the fact that B_k is a positive definite matrix. Thus, for all $k \in \mathcal{K}_1$, (4.2) and (4.4) imply that

$$1 - \rho_k \leq \frac{(1 - c_1)\Delta_k}{\alpha \min \left\{ \Delta_k, \frac{\epsilon}{M} \right\}}. \tag{4.5}$$

On the other, since $\|B_k\| \leq M$ and by (4.2), we have

$$\frac{v_k^T v_k}{v_k^T B_k v_k} \|v_k\|_2 \geq \frac{v_k^T v_k}{M \|v_k\|_2^2} \|v_k\|_2 = \frac{\|v_k\|_2}{M} \geq \frac{\epsilon}{M}.$$

Now, let $\tilde{\Delta}$ be defined as follows:

$$\tilde{\Delta} = \frac{\epsilon}{M}.$$

Then, for all $\Delta_k \leq \tilde{\Delta}$, we have $\min \left\{ \Delta_k, \frac{\epsilon}{M} \right\} = \Delta_k$ and

$$\frac{v_k^T v_k}{v_k^T B_k v_k} \|v_k\|_2 \geq \frac{\epsilon}{M} \geq \Delta_k,$$

which in turn implies that the CG-Steihaug method returns exactly the Cauchy point and the sufficient reduction is achieved. Now, using (4.5), we obtain

$$1 - \rho_k \leq \frac{(1 - c_1)}{\alpha}.$$

If $c_1 \geq 1 - \alpha(1 - c_2)$, then, for all $\Delta_k \leq \tilde{\Delta}$, we have $\rho_k \geq c_2$ and the CG-Stiehaug method returns the Cauchy point and the sufficient reduction is achieved. In this situation, for sufficiently large k, let say $k \geq k'$, in the Ltrust algorithm we have $\Delta_{k+1} \geq \Delta_k$, which in turn implies that

$$\Delta_k \geq \min \left\{ \Delta_{k'}, \frac{\tilde{\Delta}}{\theta_\Delta} \right\}, \qquad \forall k \in \mathcal{K}_1, \; k \geq k'. \tag{4.6}$$

To complete the proof of lemma, we show that $\bar{\mathcal{K}}$ is a finite set. For this purpose, on the contrary, assume that $\bar{\mathcal{K}}$ is an infinite set. Then, either $\mathcal{K}_1$ or $\mathcal{K}_2$ is infinite.
If $\mathcal{K}_1$ is an infinite set, then for $k \in \mathcal{K}_1$, we have:

$$\begin{aligned} f(x_k) - f(x_{k+1}) &\geq c_2 \left(m(x_k, 0) - m(x_k, p_k)\right) \\ &\geq c_2 \alpha \|v_k\|_2 \min \left\{ \Delta_k, \frac{\|v_k\|_2}{\|B_k\|} \right\} \\ &\geq c_2 \alpha \epsilon \min \left\{ \Delta_k, \frac{\epsilon}{M} \right\}. \end{aligned} \tag{4.7}$$

Now, since the sequence $\{f(x_k)\}_{k \in \mathcal{K}_1}$ is a decreasing and bounded below, (4.7) implies that

$$\lim_{k \in \mathcal{K}_1, k \to \infty} \Delta_k = 0. \tag{4.8}$$

On the other hand, if $\mathcal{K}_2$ is an infinite set, then for all $k \in \mathcal{K}_2$, we have $\Delta_k \leq \alpha_k$. This implies that

$$f(x_k) - f(x_{k+1}) \geq c_1 \alpha_k \|v_k\|_2 \geq c_1 \Delta_k \epsilon, \tag{4.9}$$

which in turn concludes that

$$\lim_{k \in \mathcal{K}_2, k \to \infty} \Delta_k = 0. \tag{4.10}$$

Therefore, considering (4.8) and (4.10), we have:

$$\lim_{k \in \bar{\mathcal{K}}, k \to \infty} \Delta_k = 0, \tag{4.11}$$

whenever $\bar{\mathcal{K}}$ is an infinite set. This means that there exists k'' so that for all $k > k''$ and $k \in \bar{\mathcal{K}}$, we have $\Delta_k \leq \tilde{\Delta}$, which is a contradiction with (4.6). This shows that $\bar{\mathcal{K}}$ is a finite set. Therefore, $\rho_k < c_2$ or a sufficient reduction is not achieved, for all sufficiently large k. In this situation, Δ_k is multiplied by θ_Δ at every iteration, and we have $\Delta_k \to 0$, which is again a contradiction with (4.6). Hence, the assumption (4.2) is false and (4.1) holds. $\square$

Now, we are ready to provide the global convergence property of the Ltrust algorithm. The proof is very similar to the proof of Theorem 3.1 in [2]. In [2], $\mathcal{K}_1$ is just the set of successful iterations. But here the successful iterations are $\mathcal{K}_1 \bigcup \mathcal{K}_2$.

Theorem 4.2. *Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is a locally Lipschitz function and the level set $\mathcal{L} = \{x : f(x) \leq f(x_1)\}$ is bounded. Assume that there exist positive constants M and $\bar{m}$ so that $\bar{m}\|p\|_2^2 \leq p^T B_k p \leq M\|p\|_2^2$, for all k and $p \in \mathbb{R}^n$. Then, the Ltrust algorithm either stops at a certain k_0 with $\|v_{k_0}\|_2 = 0$, or generates an infinite sequence $\{x_k\}_{k \in \bar{\mathcal{K}}}$ with $0 \in \partial f(x^*)$, where x^* is its limit point.*

Proof. If the Ltrust algorithm terminates at a certain iteration k_0 with $\|v_{k_0}\|_2 = 0$, then $0 \in \partial_{\Delta_{k_0}} f(x_{k_0})$ which shows that x_{k_0} is a Δ_{k_0}-stationary point of f.
Now, assume that $\{x_k\}_{k \in \bar{\mathcal{K}}}$ is an infinite sequence. Since $\mathcal{L}$ is a bounded set, then the sequence $\{x_k\}_{k \in \bar{\mathcal{K}}}$ has at least one limit point. Let x^* be a limit point of $\{x_k\}_{k \in \bar{\mathcal{K}}}$ and $\{x_{k_n}\}$ be its subsequence so that $x_{k_n} \to x^*$. On the other hand, the sequence $\{f(x_k)\}_{k \in \bar{\mathcal{K}}}$ is a decreasing sequence and bounded below, therefore it converges to a point, let say $f^* > -\infty$. Assume that, on the contrary, $0 \notin \partial f(x^*)$. Since $\partial f(\cdot)$ is an upper semicontinuous function, then there exists $r > 0$ so that $0 \notin \partial_r f(x^*)$. Let

$$\gamma = \min_{v \in \partial_r f(x)} \|v\|_2. \tag{4.12}$$

Since $0 \notin \partial_r f(x^*)$, then $\gamma > 0$. We first prove that $\|v_{k_n}\|_2$ converges to 0. To do so, on the contrary, assume that there exist $\epsilon > 0$ and a subsequence $\{\|v_t\|_2\}$ of $\{\|v_{k_n}\|_2\}$, so that

$$\|v_t\|_2 \geq \epsilon, \qquad \forall t. \tag{4.13}$$

Due to Lemma 4.1, we have $\liminf_{k\to\infty, k\in\bar{\mathcal{K}}} \|v_k\|_2 = 0$. Therefore, there exist some $k \in \bar{\mathcal{K}}$ so that $\|v_k\|_2 < \bar{\kappa} := \min\left\{\frac{\epsilon}{2}, \frac{\gamma}{2}\right\}$. For each t, let $l_t > t$ be the first iteration in $\bar{\mathcal{K}}$ so that $\|v_{l_t}\|_2 < \bar{\kappa}$. Thus, we have

$$\|v_k\|_2 \geq \bar{\kappa}, \text{ for all } k \in \bar{\mathcal{K}} \text{ and } \ t \leq k < l_t.$$

Now, for every $k \in \mathcal{K}_1 \subseteq \bar{\mathcal{K}}$ and $t \leq k < l_t$, we have

$$\begin{aligned} f(x_k) - f(x_{k+1}) &\geq c_2\left(m(x_k, 0) - m(x_k, p_k)\right) \\ &\geq c_2\alpha\|v_k\|_2 \min\left\{\Delta_k, \frac{\|v_k\|_2}{\|B_k\|}\right\} \\ &\geq c_2\alpha\bar{\kappa} \min\left\{\Delta_k, \frac{\bar{\kappa}}{M}\right\}. \end{aligned} \tag{4.14}$$

On the other hand, for each $k \in \mathcal{K}_2 \subseteq \bar{\mathcal{K}}$ and $t \leq k < l_t$, we have

$$f(x_k) - f(x_{k+1}) \geq c_1\alpha_k\|v_k\|_2 \geq c_1\Delta_k\bar{\kappa}. \tag{4.15}$$

Now, using (4.14) and (4.15) and the fact that $\{f(x_k)\}$ is a decreasing and convergent sequence, we have:

$$\Delta_k \to 0, \quad \text{as } k \in \bar{\mathcal{K}}, \ t \leq k < l_t \text{ and } t \to \infty.$$

Thus, for sufficiently large $k \in \bar{\mathcal{K}}$ with $t \leq k < l_t$, we conclude that:

$$\Delta_k \leq \min\left\{\frac{1}{c_2\alpha\bar{\kappa}}\left[f(x_k) - f(x_{k+1})\right], \frac{1}{c_1\bar{\kappa}}\left[f(x_k) - f(x_{k+1})\right]\right\}.$$

Letting $\gamma = \min\left\{\frac{1}{c_2\alpha\bar{\kappa}}, \frac{1}{c_1\bar{\kappa}}\right\}$, for sufficiently large t, we have

$$\|x_t - x_{l_t}\|_2 \leq \sum_{k=t, k\in\bar{\mathcal{K}}}^{l_t - 1} \|x_{k+1} - x_k\|_2 \leq \sum_{k=t, k\in\bar{\mathcal{K}}}^{l_t - 1} \Delta_k \leq \gamma\left[f(x_t) - f(x_{l_t})\right]. \tag{4.16}$$

Now, from the fact that the right-hand side of (4.16) converges to zero, we obtain that

$$\|x_t - x_{l_t}\|_2 \to 0, \text{ as } t \to \infty.$$

Therefore, $x_{l_t} \to x^*$. On the other hand, since $\Delta_k \to 0$, then for sufficiently large t, we obtain that $\partial_{\Delta_{l_t}} f(x_{l_t}) \subset \partial_r f(x^*)$, which is a contradiction with (4.12) considering the fact that $\|v_{l_t}\|_2 \leq \bar{\kappa} \leq \frac{\gamma}{2}$. Thus, $\|v_{k_n}\|_2 \to 0$, as $k_n \to \infty$.

Now, we show that $\Delta_{k_n} \to 0$. Assume that, on the contrary, there exists $\sigma > 0$ so that $\Delta_{k_n} \geq \sigma$. For each k_n, we have

$$\begin{aligned} m(x_{k_n}, p_{k_n}) - m(x_{k_n}, 0) &= v_{k_n}^T p_{k_n} + p_{k_n}^T B_{k_n} p_{k_n} \\ &\geq v_{k_n}^T p_{k_n} + \Delta_{k_n}^2 \bar{m} \\ &\geq v_{k_n}^T p_{k_n} + \bar{m}\sigma^2. \end{aligned}$$

Now, since $\|v_{k_n}\|_2 \to 0$, we have

$$\lim_{k_n \to \infty} (m(x_{k_n}, p_{k_n}) - m(x_{k_n}, 0)) \geq \bar{m}\sigma^2 > 0,$$

which is a contradiction with the fact that $m(x_{k_n}, p_{k_n}) < m(x_{k_n}, 0)$. Thus, $\Delta_{k_n} \to 0$ as $k_n \to \infty$.
Finally, we show that $0 \in \partial f(x^*)$. Since, for each k_n, we have $v_{k_n} \in \partial_{\Delta_{k_n}} f(x_{k_n})$, then there exists $y_{k_n} \in N_{\Delta_{k_n}}(x_{k_n})$ so that $v_{k_n} \in \partial f(y_{k_n})$, where $N_{\Delta_{k_n}}(x_{k_n})$ stands for the neighborhood with radius Δ_{k_n} centered at x_{k_n}. Besides, $\Delta_{k_n} \to 0$ implies that $y_{k_n} \to x^*$. Now, using the fact that the function $\partial f(\cdot)$ is an upper semicontinuous function and $v_{k_n} \to 0$, we obtain that $0 \in \partial f(x^*)$, which completes the proof of the theorem. □

5 Numerical Results

The promising practical behavior of the Ntrust algorithm [2] against some other existing nonsmooth methods has been well studied in [2]. In this section, our aim is to compare the numerical performance of the Ltrust and Ntrust algorithms. More precisely, we investigate the effect of utilizing the line search technique in the structure of Ntrust algorithm.
The Ltrust and Ntrust algorithms are implemented in MATLAB environment on a PC with CPU 2.0 GHz, 2GB RAM memory and double precision format. The specifications of the considered test problems, taken from [11, 13], are provided in Table 1. These problems are exactly those considered in [2] and are divided in two classes; the first class consists of convex functions and the second class considers the nonconvex test functions.

Table 1: Test problems and their optimum values

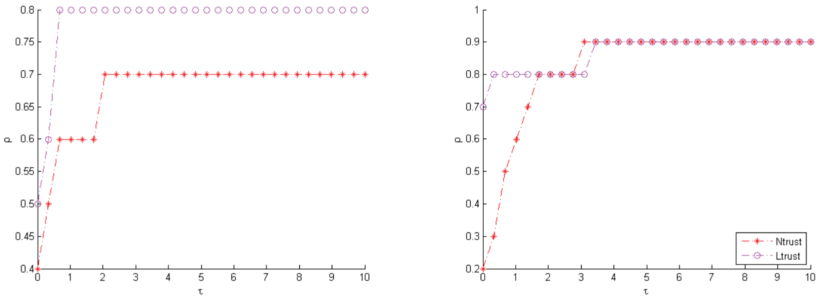
First class of problems			Second class of problems		
No.	problem	optimal value	No.	problem	optimal value
1	MAXQ	0	11	problem 2 from TEST29	0
2	MAXHILB	0	12	problem 5 from TEST29	0
3	LQ	-1.412799e+003	13	problem 6 from TEST29	0
4	CB3I	1998	14	problem 11 from TEST29	1.203128e+004
5	CB3II	1998	15	problem 13 from TEST29	5.661313e+002
6	NACTFACES	0	16	problem 17from TEST29	0
7	Brown 2	0	17	problem 19 from TEST29	0
8	Mifflin 2	-7.065034e+002	18	problem 20 from TEST29	0
9	Crescent I	0	19	problem 22 from TEST29	0
10	Crescent II	0	20	problem 24 from TEST29	0

We have also utilized the advantages of the performance profile of Dolan and Moré [8] (in $\log_2$ scale) in order to have a better comparison.
The parameters and stopping criteria in both Ntrust and Ltrust algorithms are initiated same as the Ntrust algorithm in [2]. Numerical results are given in Tables 2 and 3. Since the number of subgradient and function evaluations are equal, we just report the number of function evaluations. In these tables, "nfeval" stands for the number of function evaluations and f^* denotes the computed optimal value. Moreover, "NLS" in Ltrust shows the number of line search usages for solving a problem and n is the dimention of the problem.
The performance profile of the considered algorithms in terms of number of function evaluations for the first and second classes of test problems are respectively drawn in Figures 1

and 2. These figures reveal that the number of function evaluations are reduced significantly in Ltrust. Moreover, Ltrust fails less often than Ntrust in the case $n = 100$. Besides, Ltrust solves more problems than Ntrust, especially in large-scale settings, in the minimum number of function evaluations. For $n = 1000$, Ltrust is more efficient in 80% of test problems in Figure 1 and 70% in Figure 2. In summary, Figures 1 and 2 show that Ltrust is more efficient than the Ntrust as the number of fuction evaluations and iterations are reduced significantly in Ltrust.

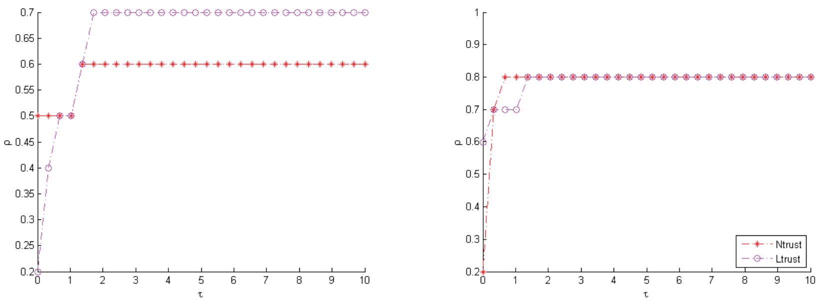
On the other hand, the number of subproblem resolvings is reduced in Ltrust. This fact is shown in the performance profile drawn in terms of number of subproblem solvings in Figures 3 and 4. These figures show that the number of subproblem resolvings is decreased significanly in the new proposed algorithm.

It is worth mentioning that the step length $\alpha_k = \Delta_k$ satisfies the Armijo condition. In our experience, this step length did not give us any better result than the usual line search method and therefore we ignored this case in our results. We believe that the usual line search technique mostly finds a larger step length than Δ_k, and therefore the objective function can be decreased more.



n=100 n=1000

Figure 1: Performance profile of the algorithms for the first class of problems in terms of nfeval



n=100 n=1000

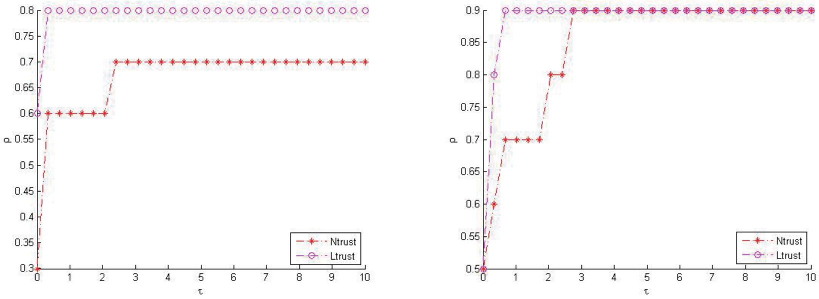
Figure 2: Performance profile of the algorithms for the second class of problems in terms of nfeval

Table 2: Numerical results of test problems for $n = 100$

	Ntrust		Ltrust		
No.	f^*	nfeval	f^*	nfeval	NLS
1	1.236e+02	10002	1.221e+03	10002	257
2	1.232e-04	10006	8.926e-06	46927	10
3	-1.400e+02	10015	-1.400e+02	12383	1
4	1.980e+02	7803	1.980e+02	1999	1
5	1.980e+02	1863	1.980e+02	1199	3
6	6.883e-15	101	3.961e-11	144	1
7	5.507e-10	521	2.006e-10	673	1
8	-7.015e+01	10008	-7.014e+01	11921	3
9	3.071e-09	332	5.055e-09	293	2
10	2.000e+00	2500	2.226e-09	882	3
11	5.684e-11	5112	1.905e-10	8206	338
12	4.278e-05	10014	1.053e-06	22898	11
13	1.158e-11	197	1.786e-11	203	4
14	1.190e+03	10022	1.191e+03	9776	3
15	5.559e+01	5527	5.559e+01	2301	4
16	6.120e-11	4409	2.448e-10	5001	74
17	1.424e-01	7150	8.479e-09	12595	587
18	1.106e-07	10005	5.543e-10	26701	477
19	1.725e-04	10023	1.544e-04	7979	2
20	1.889e-02	10001	6.721e-03	23772	250

Table 3: Numerical results of test problems for $n = 1000$

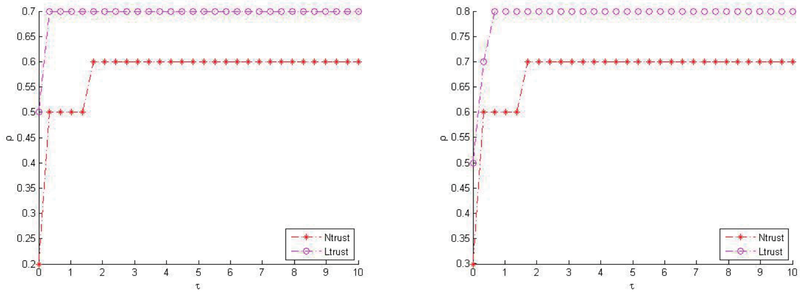
	Ntrust		Ltrust		
No.	f^*	nfeval	f^*	nfeval	NLS
1	5.262e+05	100001	9.162e+05	100000	1047
2	8.404e-05	100007	1.058e-05	57868	21
3	-1.413e+03	100180	-1.413e+03	900030	1
4	1.998e+03	23398	1.998e+03	3073	2
5	1.998e+03	2723	1.998e+03	1226	2
6	3.996e-10	550	1.125e-09	210	3
7	5.829e-10	817	5.725e-10	660	4
8	-7.065e+02	1138	-7.065e+02	807	4
9	3.209e-09	403	6.694e-10	411	4
10	4.146e-09	1468	2.000e-09	1135	4
11	5.814e-11	48222	9.498e-11	30415	2653
12	3.336e-05	42349	5.414e-06	36967	13
13	2.824e-11	203	1.786e-11	160	4
14	1.203e+04	100054	1.203e+04	36065	15
15	5.661e+02	3108	5.661e+02	3830	5
16	5.808e-11	39251	9.776e-11	35227	513
17	4.059e-08	45890	5.066e-10	39006	5502
18	1.015e-05	100007	1.285e-05	100006	4197
19	4.085e-05	42096	1.023e-05	100163	20
20	4.016e-01	100033	1.387e-02	100010	2868



n=100 n=1000

Figure 3: Performance profile of the algorithms for the first class of problems in terms of number of subproblem solvings

s



n=100 n=1000

Figure 4: Performance profile of the algorithms for the second class of problems in terms of number of subproblem solvings

6 Conclusions

In this paper, a backtracking line search technique is exploited in the structure of nonsmooth trust region (TR) methods for minimizing locally Lipschitz functions. In this method, a line search performs along an approximation of the steepest descent direction for locally Lipschitz functions whenever the computed trial step of the TR subproblem is rejected by the nonsmooth TR algorithm. This may significantly reduce the number of function evaluations and therefore increases the efficiency of the algorithm. Under some standard assumptions, the global convergence property of the new algorithm is established. Preliminary numerical results show the efficiency and robustness of the new proposed algorithm compared with the Ntrust algorithm in [2]. It is worth mentioning that the priority and well promising behavior of Ntrust algorithm to some other existing nonsmooth methods for minimizing locally Lipschitz functions has been investigated in [2].

References

[1] M. Ahookhosh, K. Amini and M. Reza Peyghami, A nonmonotone trust-region line search method for large-scale unconstrained optimization, *Appl. Math. Model.* 36 (2012) 478–487.

[2] Z. Akbari, R. Yousefpour and M. Reza Peyghami, A new nonsmooth trust region algorithm for locally Lipschitz unconstrained optimization problems, *J. Optim. Theory Appl.* 164 (2015) 733–754.

[3] D. Ataee Tarzanagh, M. Reza Peyghami and H. Mesgarani, A new nonmonotone trust region method for unconstrained optimization equipped by an efficient adaptive radius, *Optim. Methods Softw.* 29 (2014) 819–836.

[4] A.M. Bagirov, B. Karasözen and M. Sezeri, Discrete gradient method: derivative-free method for nonsmooth optimization, *J. Optim. Theory Appl.* 137 (2008) 317–334.

[5] F.H. Clarke, *Optimization and Nonsmooth Analysis*, Society for Industrial and Applied Mathematics, 1987.

[6] A.R. Conn, N.I.M. Gould and Ph.L. Toint, *Trust-Region Methods*, SIAM, 2000.

[7] J. Dennis, S. Li and R. Tapia, A unified approach to global convergence of trust region methods for nonsmooth optimization, *Math. Program.* 68 (1995) 319–346.

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

[9] R. Fletcher, A model algorithm for composite nondifferentiable optimization problem, *Math. Program. Studies* 17 (1982) 67–76.

[10] A.A. Goldstein, Optimization of Lipschitz continuous functions, *Math. Program.* 13 (1977) 14–22.

[11] N. Haarala, K. Miettinen and M.M. Mäkelä, Globally convergent limited memory bundle method for large-scale nonsmooth optimization, *Math. Program.* 109 (2007) 181-205.

[12] K.C. Kiwiel, *Methods of Descent for Nondifferentiable Optimization*, Lecture Notes in Mathematics 1133, Springer-Verlag, 1985.

[13] L. Lukšan, M. Tuma, M. Siska, J. Vlček and N. Ramešová, *UFO 2002: interactive System for Universal Functional Optimization*, Academy of Sciences of the Czech Republic, 2002. Also available as http://www.cs.cas.cz/luksan/ufo.pdf

[14] N. Mahdavi-Amiri and R. Yousefpour, An effective nonsmooth optimization algorithm for locally Lipschitz functions, *J. Optim. Theory Appl.* 155 (2012) 180–195.

[15] M.M. Mäkelä, Survey of bundle methods for nonsmooth optimization, *Optim. Methods Softw.* 17 (2002) 1–29.

[16] J. Nocedal and S.J. Wright, *Numerical Optimization*, Springer, 1999.

[17] J. Nocedal and Y. Yuan, Combining Trust Region and Line Search Techniques, in *Advances in Nonlinear Programming*, Volume 14 of the series Applied Optimization, Sringer, 1998, pp. 153–175.

[18] M.J.D. Powell, General Algorithms for Discrete Nonlinear Approximation Calculations, in *Approximation Theory IV*, C.K. Chui, L.L. Schumaker and J.D. Wards (Eds.), Academic Press, London, 1983, pp. 187–218.

[19] L. Qi and J. Sun, A trust region algorithm for minimization of locally Lipschitzian functions, *Math. Program.* 66 (1994) 25–43.

[20] Y. Yuan, *Global Convergence of Trust Region Algorithms for Nonsmooth Optimization*, Department of Applied Mathematics and Theoretical Physics, Cambridge University, DAMTP 1983/NA13, 1983.

[21] Y. Yuan, Recent advances in trust region algorithms. *Math. Program.* 151 (2015) 249–281.

Manuscript received 11 December 2015
revised 18 July 2016, 13 February 2017
accepted for publication 7 April 2017

Z. Akbari
Department of Mathematical Sciences
University of Mazandaran, Babolsar, Iran
E-mail address: z.akbari@umz.ac.ir

M. Reza Peyghami
Faculty of Mathematics, K.N. Toosi Univ. of Tech.
P.O. Box 16315-1618, Tehran, Iran
Scientific Computations in OPtimization and
Systems Engineering (SCOPE), K.N. Toosi Univ. of Tech., Tehran, Iran
E-mail address: peyghami@kntu.ac.ir

R. Yousefpour
Department of Mathematical Sciences
University of Mazandaran, Babolsar, Iran
E-mail address: yousefpour@umz.ac.ir