A secant method for nonsmooth optimization

Adil M. Bagirov and Asef Nazari Ganjehlou

Centre for Informatics and Applied Optimization, School of Information Technology and Mathematical Sciences, University of Ballarat, Victoria 3353, Australia

Abstract

The notion of a secant for locally Lipschitz continuous functions is introduced and a new algorithm to locally minimize nonsmooth, nonconvex functions based on secants is developed. We demonstrate that the secants can be used to design an algorithm to find descent directions of locally Lipschitz continuous functions. This algorithm is applied to design a minimization method, called a secant method. It is proved that the secant method generates a sequence converging to Clarke stationary points. Numerical results are presented demonstrating the applicability of the secant method in wide variety of nonsmooth, nonconvex optimization problems. We also compare the proposed algorithm with the bundle method using numerical results.

Keywords: nonsmooth optimization, nonconvex optimization, subdifferential, Lipschitz functions.
Mathematical Subject Classification (2000): 65K05, 90C25.

1 Introduction

Consider the following unconstrained minimization problem:

$$\text{minimize} \quad f(x) \quad \text{subject to} \quad x \in \mathbb{R}^n \tag{1}$$

where the objective function f is assumed to be locally Lipschitz continuous. Numerical methods for solving Problem (1) have been studied extensively over more than four decades. Subgradient methods [32], the bundle method and its variations [13, 15, 16, 17, 18, 19, 20, 21, 22, 25, 27, 28, 29, 33, 35], a discrete gradient method [1, 3, 4], a gradient sampling method [8], and methods based on smoothing techniques [30] are among them.

Subgradient methods are simple, however, they are not effective to solve many nonsmooth optimization problems. Algorithms based on smoothing techniques are applicable only to the special class of nonsmooth functions such as the maximum and max-min functions. Bundle algorithms are more general algorithms and they build up information about the subdifferential of the objective function using ideas known as bundling and aggregation. These algorithms are known to be very effective for nonsmooth, convex problems. For nonconvex functions subgradient information is meaningful only locally, and must be discounted when no longer relevant. Special rules have to be designed to identify those irrelevant subgradients. As a result, bundle type algorithms are more complicated in the nonconvex case [8].

The gradient sampling method proposed in [8] does not require rules for discounting irrelevant subgradients. It is a stochastic method and relies on the fact that locally Lipschitz functions are differentiable almost everywhere. At each iteration of this method a given number of gradients are evaluated at points randomly chosen from a given neighborhood of a current point. Then a quadratic programming problem is solved to compute a descent direction by finding least distance between a polytope, given by convex hull of the gradients, and the origin. This method was successfully applied to solve some practical problems [8].

In this paper a new algorithm for solving Problem (1) is proposed. We introduce the notion of a secant for locally Lipschitz functions, which is an approximation to a subgradient. An algorithm for the computation of descent directions based on secants is given and using that we design a minimization algorithm and study its convergence.

There are some similarities between the proposed method and the bundle and the gradient sampling methods. In the new method we build up information about the approximation of the subdifferential using bundling idea which makes it similar to the bundle methods. However, the new method does not require subgradient discounting rules which makes it similar to the gradient sampling method.

The new method is applied to solve nonsmooth optimization test problems from [23] as well as some cluster analysis problems. We compare the new algorithm with two versions of bundle method using numerical results. These two versions are: PMINU - the bundle method for unconstrained minimax optimization and PBUNU - the bundle method for general unconstrained nonsmooth optimization [24].

The structure of the paper is as follows. Section 2 provides some necessary preliminaries. The notion of an r-secant is introduced in Section 3. Necessary conditions for a minimum using r-secants are derived in Section 4. In Section 5 we describe an algorithm for the computation of a descent direction. A secant method is introduced in Section 6. In Section 7 we describe an algorithm to approximate subgradients. Results of numerical experiments are given and discussed in Section 8. Section 9 concludes the paper.

2 Preliminaries

In this section we will recall some basic concepts from nonsmooth analysis needed in this paper.

Let f be a locally Lipschitz continuous function defined on $\mathbb{R}^n$. The function f is differentiable almost everywhere and one can define for it a Clarke subdifferential

[9] (see also [10]) by

$$\partial f(x) = \text{co}\left\{v \in \mathbb{R}^n : \exists (x^k \in D(f)) : \; x = \lim_{k\to\infty} x^k \;\text{ and }\; v = \lim_{k\to\infty} \nabla f(x^k)\right\},$$

here $D(f)$ denotes the set where f is differentiable, and co denotes the convex hull of a set. It is shown in [9, 10] that the mapping $x \mapsto \partial f(x)$ is upper semicontinuous and bounded on bounded sets.

The generalized directional derivative of f at x in the direction g is defined as

$$f^0(x, g) = \limsup_{y\to x, \alpha\to+0} \alpha^{-1}[f(y + \alpha g) - f(y)].$$

If the function f is locally Lipschitz continuous then the generalized directional derivative exists and

$$f^0(x, g) = \max\left\{\langle v, g\rangle : v \in \partial f(x)\right\}.$$

Here $\langle \cdot, \cdot \rangle$ stands for an inner product in $\mathbb{R}^n$. f is called a regular function at $x \in \mathbb{R}^n$, if it is directionally differentiable at x and $f'(x, g) = f^0(x, g)$ for all $g \in \mathbb{R}^n$ where $f'(x, g)$ is a derivative of the function f at the point x in the direction g:

$$f'(x, g) = \lim_{\alpha\to+0} \alpha^{-1}[f(x + \alpha g) - f(x)].$$

There exists a full calculus for regular functions, meaning that it is in the form of equalities. However, in general, such a calculus does not exist for non-regular functions and it is only in the form of inclusions [10]. Although, it cannot be applied to compute subgradients, we will consider a special scheme to approximate subgradients of such functions.

Let f be a locally Lipschitz continuous function defined on $\mathbb{R}^n$. For a point x to be a local minimizer of the function f on $\mathbb{R}^n$, it is necessary that $0_n \in \partial f(x)$. A point $x \in \mathbb{R}^n$ is called a (Clarke) stationary point if $0_n \in \partial f(x)$.

The function $f : \mathbb{R}^n \to \mathbb{R}^1$ is called semismooth at $x \in \mathbb{R}^n$, if it is locally Lipschitz at x and for each $g \in \mathbb{R}^n$ and for any sequences $\{t_k\} \subset \mathbb{R}^1$, $\{g^k\} \subset \mathbb{R}^n$, and $\{v^k\} \subset \mathbb{R}^n$ such that $t_k \downarrow 0$, $g^k \to g$, and $v^k \in \partial f(x + t_k g^k)$, the limit

$$\lim_{k\to\infty} \langle v^k, g\rangle$$

exists. The semismooth function f is directionally differentiable and

$$f'(x, g) = \lim_{k\to\infty} \langle v^k, g\rangle, \quad v^k \in \partial f(x + t_k g^k).$$

The class of semismooth functions contains convex, concave, max-type and min-type functions (see [26]).

3 An r-secant

In this section we introduce the notion of an r-secant for locally Lipschitz functions.

Let $S_1 = \{g \in \mathbb{R}^n : \|g\| = 1\}$ be the unit sphere in $\mathbb{R}^n$. For any $g \in S_1$ define

$$g_{max} = \max \{|g_i|, \quad i = 1, \dots, n\}.$$

It is clear that $g_{max} \geq n^{-1/2}$ for any $g \in S_1$. Let

$$I(g) = \{i \in \{1, \dots, n\} : \quad |g_i| = g_{max}\}.$$

Definition 1 *Let $g \in S_1,\ r > 0$ and $v \in \partial f(x + rg)$ be any subgradient. Select any $i \in I(g)$. A vector $s = s(x, g, r) \in \mathbb{R}^n$ where*

$$s = (s_1, \dots, s_n) : \ s_j = v_j, \ j = 1, \dots, n, \ j \neq i$$

and

$$s_i = \frac{f(x + rg) - f(x) - r \sum_{j=1, j \neq i}^{n} s_j g_j}{r g_i}$$

is called an r-secant of the function f at a point x in the direction g.

Remark 1 For a given $g \in S_1$, r-secants can be defined for all $i \in I(g)$ and $v \in f(x + rg)$.

Remark 2 One can see that the r-secant $s(x, g, r)$ is defined with respect to a given direction $g \in S_1$ and it is unique if and only if the set $I(g)$ and the subdifferential $\partial f(x + rg)$ are the singleton sets.

Remark 3 If $n = 1$ then an r-secant of the function f at a point $x \in \mathbb{R}^1$ is defined as follows:

$$s = \frac{f(x + rg) - f(x)}{rg},$$

where $g = 1$ or $g = -1$.

Proposition 1 *Let $x \in \mathbb{R}^n$ and $g \in S_1$. Then for a given $r > 0$*

$$f(x + rg) - f(x) = r\langle s(x, g, r), g \rangle, \tag{2}$$

where $s(x, g, r)$ is an r-secant of the function f at a point x in the direction g.

Proof: Proof follows immediately from Definition 1, more exactly from the definition of the i-th coordinate s_i of the r-secant $s(x, g, r)$. □

Remark 4 The equation (2) can be considered as a version of the mean value theorem for r-secants.

We define the following set

$$S_r f(x) = \{s \in \mathbb{R}^n : \exists g \in S_1 : \quad s = s(x, g, r)\}.$$

$S_r f(x)$ is the set of all possible r-secants of the function f at the point x.

Remark 5 In general, $S_r f(x)$ is not singleton set even for continuously differentiable functions. However, for any affine function f, $S_r f(x)$ is the singleton set for all $r > 0$ and $x \in \mathbb{R}^n$.

It is proved in [10] that for any bounded subset $X \subset \mathbb{R}^n$ there exists $M > 0$ such that

$$\sup \{\|v\| : v \in \partial f(x), \ x \in X\} \le M. \tag{3}$$

Proposition 2 *Let $f : \mathbb{R}^n \to \mathbb{R}^1$ be a locally Lipschitz function. Then for any bounded subset $X \subset \mathbb{R}^n$ there exists $M_0 \equiv M_0(n) > 0$ such that*

$$\sup \{\|s\| : s \in S_r f(x), \ x \in X\} \le M_0. \tag{4}$$

Proof: Take any $x \in X$ and $s \in S_r f(x)$. It follows from the definition of the set $S_r f(x)$ that there exists $g \in S_1$ such that $s = s(x, g, r)$ where $s(x, g, r)$ is an r-secant in the direction g. Since $s_j(x, g, r) = v_j, \ j \in \{1, \ldots, n\}, \ j \ne i, \ i \in I(g)$ for some $v \in \partial f(x + rg)$, it follows from (3) that

$$\sum_{j=1, j \ne i}^{n} s_j^2 \le M^2.$$

Since the function f is locally Lipschitz continuous there exists $L > 0$ such that

$$|f(x + rg) - f(x)| \le Lr.$$

Then we have

$$|s_i| \le Ln^{1/2} + M(n - 1).$$

Denote

$$M_0 \equiv M_0(n) = \left(M^2 + (Ln^{1/2} + M(n - 1))^2\right)^{1/2}.$$

Then we get

$$\|s\| \le M_0 \quad \forall s \in S_r f(x).$$

$\square$

Proposition 3 *The set $S_r f(x)$, $r > 0$ is closed.*

Proof: Take any sequence $\{s^k\}$, $s^k \in S_r f(x)$ and assume that $s^k \to s$ as $k \to \infty$. For each $s^k \in S_r f(x)$ there exists $g^k \in S_1$ such that $s^k = s^k(x, g^k, r)$. Without loss of generality we assume $g^k \to g$ as $k \to +\infty$. Since S_1 is a compact set $g \in S_1$. It is clear that $g_{k,max} \to g_{max}$ as $k \to \infty$. Moreover, there exists $k_0 > 0$ such that $I(g^k) \subseteq I(g)$ for all $k \geq k_0$.

For the secant s^k there exist $i \in I(g^k)$ and $v^k \in \partial f(x + rg^k)$ such that

$$s_j^k = v_j^k, \quad j = 1, \ldots, n, \ j \neq i$$

and $i \in I(g)$ for all $k \geq k_0$. Since $g^k \to g$ the upper semicontinuity of the set-valued mapping $x \mapsto \partial f(x)$ implies that if $v = \lim_{k \to \infty} v^k$ then $v \in \partial f(x + rg)$. The latter means that

$$s_j = v_j, \quad j = 1, \ldots, n, \ j \neq i.$$

The continuity of the function f implies that $s_i(x, g^k, r) \to s_i(x, g, r)$ as $k \to \infty$. Thus, $s \in S_r f(x)$. □

Corollary 1 *The set $S_r f(x)$, $r > 0$ is compact.*

Proof: The proof follows immediately from Propositions 2 and 3. □

Proposition 4 *The set-valued mapping $(x, r) \mapsto S_r f(x), r > 0$ is closed.*

Proof: Take any sequences $\{x^k\} \subset \mathbb{R}^n$, $\{r_k\} \subset \mathbb{R}^1$, $r_k > 0$, and $\{s^k\} \subset \mathbb{R}^n$ such that $x^k \to x, r_k \to r > 0$, and $s^k \in S_{r_k} f(x^k)$, $s^k \to s$ as $k \to +\infty$. We have to show that $s \in S_r f(x)$. Since $s^k \in S_{r_k} f(x^k)$ there exist $g^k \in S_1$ and $v^k \in \partial f(x^k + r_k g^k)$ such that

$$s_j^k = v_j^k, \ j = 1, \ldots, n, \ j \neq i, \ i \in I(g^k).$$

Without loss of generality we assume that $g^k \to g$ and $v^k \to v$ as $k \to \infty$. It is obvious that $g \in S_1$. Since the set-valued mapping $x \mapsto \partial f(x)$ is closed, $v \in \partial f(x + rg)$. Obviously there exists $k_0 > 0$ such that $I(g^k) \subseteq I(g)$ for all $k \geq k_0$. Then it follows from the continuity of the function f that the vector s is defined as follows:

$$s_j = v_j, \ j = 1, \ldots, n, \ j \neq i, \ i \in I(g^k) \subseteq I(g), \ k \geq k_0,$$

$$s_i = \frac{f(x + rg) - f(x) - r \sum_{j=1, j \neq i}^{n} s_j g_j}{r g_i},$$

that is $s \in S_r f(x)$. □

Proposition 5 *The set-valued mapping $(x, r) \mapsto S_r f(x), r > 0$ is upper semi-continuous.*

Proof: The upper semi-continuity of the mapping $(x,r) \mapsto S_r f(x)$ means that for any $\varepsilon > 0$ there exists $\delta > 0$ such that

$$S_t f(y) \subset S_r f(x) + B_\varepsilon$$

for all $y \in B_\delta(x)$ and $t > 0,\ |t-r| < \delta$. Here

$$B_\delta(x) = \{y \in \mathbb{R}^n :\ \|y-x\| < \delta\}, \quad B_\delta = B_\delta(0_n).$$

Assume the contrary that is the mapping $(x,r) \mapsto S_r f(x), r > 0$ is not upper semi-continuous at (x,r). Then there exists $\varepsilon_0 > 0$ such that for any $\delta > 0$, it is possible to find $y \in B_\delta(x)$ and $t > 0,\ |t-r| < \delta$ such that

$$S_t f(y) \not\subset S_r f(x) + B_{\varepsilon_0}.$$

This means that there exist sequences $\{y^k\} \subset \mathbb{R}^n$, and $\{t_k\} \subset \mathbb{R}^1$ such that $y^k \to x$, and $t_k \downarrow 0$ as $k \to \infty$ and

$$S_{t_k} f(y^k) \not\subset S_r f(x) + B_{\varepsilon_0},\ k = 1, 2, \ldots.$$

This contradicts to the fact that the mapping $(x,r) \mapsto S_r f(x)$ is closed. □

Now we introduce the following sets at a point $x \in \mathbb{R}^n$:

$$S_{0,g} f(x) = \left\{ v \in \mathbb{R}^n : \exists \{r_k\} :\ r_k > 0,\ \lim_{k\to\infty} r_k = 0 \text{ and } v = \lim_{k\to\infty} s(x, g, r_k) \right\},\ g \in S_1,$$

and

$$S_0 f(x) = \bigcup_{g \in S_1} S_{0,g} f(x). \tag{5}$$

$S_{0,g} f(x)$ is the set of limit points of r-secants in the direction $g \in S_1$. The set $S_0 f(x)$ can be also defined as

$$S_0 f(x) = \left\{ v \in \mathbb{R}^n : \exists (g \in S_1,\ \{r_k\}) :\ r_k > 0,\ \lim_{k\to\infty} r_k = 0 \text{ and } v = \lim_{k\to\infty} s(x, g, r_k) \right\},$$

which is the set of limit points of all r-secants as $r \downarrow 0$.

In the sequel we study relation between these two sets and the Clarke subdifferential.

Proposition 6 *Let f be a locally Lipschitz and directionally differentiable at $x \in \mathbb{R}^n$ and, $g \in S_1$. Then for any $e = \lambda g, \lambda > 0$*

$$f'(x, e) = \langle v, e \rangle, \quad \forall v \in S_{0,g} f(x).$$

Proof: The proof follows from (2) and the definition of the set $S_{0,g} f(x)$. □

Corollary 2 *Let f be a locally Lipschitz and directionally differentiable at $x \in \mathbb{R}^n$. Then for any $e \in \mathbb{R}^n$*

$$f'(x, e) \le \max \{\langle v, e\rangle : \ v \in S_0 f(x)\}.$$

Proof: The proof follows from (2) and the definition of the set $S_0 f(x)$. □

Proposition 7 *Let f be a semismooth function at $x \in \mathbb{R}^n$. Then*

$$S_{0,g} f(x) \subseteq \partial f(x), \quad \forall g \in S_1.$$

Proof: Given $g \in S_1$ consider the following set at the point x:

$$Q(x, g) = \{v \in \mathbb{R}^n : \exists(\{r_k\}, \ \{v^k\}) : \ r_k > 0, \ \lim_{k\to\infty} r_k = 0, \ v^k \in \partial f(x + r_k g)$$
$$\text{and} \ \ v = \lim_{k\to\infty} v^k\}.$$

Since the function f is semismooth

$$f'(x, g) = \langle v, g\rangle \ \ \forall \ v \in Q(x, g). \tag{6}$$

It follows from the definition of the set $S_{0,g} f(x)$ that for any $\bar{s} \in S_{0,g} f(x)$ there exists $\{r_k\}, \ r_k > 0, \ r_k \to 0, \ k \to \infty$ such that

$$\bar{s} = \lim_{k\to\infty} s(x, g, r_k).$$

By the definition of r-secants there exists $v^k \in \partial f(x + r_k g)$ such that

$$s_j(x, g, r_k) = v^k_j, \ j = 1, \dots, n, \ j \neq i, \ i \in I(g).$$

Without loss of generality assume that $v^k \to v$ as $k \to \infty$. It is clear that $v \in Q(x, g)$ and $\bar{s}_j = v_j, \ j = 1, \dots, n, j \neq i$. On the other hand it follows from Proposition 6 and (6) that

$$f'(x, g) = \langle \bar{s}, g\rangle = \langle v, g\rangle.$$

Then we get $\bar{s}_i = v_i$ that is $\bar{s} = v$ and $\bar{s} \in Q(x, g) \subset \partial f(x)$. Since $s \in S_{0,g} f(x)$ is arbitrary we get the proof. □

Corollary 3 *Let f be a semismooth at $x \in \mathbb{R}^n$. Then*

$$S_0 f(x) \subseteq \partial f(x).$$

Proof: The proof follows from (5) and Proposition 7. $\square$

At a point $x \in \mathrm{I\!R}^n$ consider the following two sets:

$$S_r^c f(x) = \mathrm{co}\, S_r f(x), \;\; r > 0,$$
$$S_0^c f(x) = \mathrm{co}\, S_0 f(x).$$

The following corollary follows from Proposition 2.

Corollary 4 *Let $f : \mathrm{I\!R}^n \to \mathrm{I\!R}^1$ be a locally Lipschitz function. Then for any bounded subset $X \subset \mathrm{I\!R}^n$*

$$\sup\{\|s\| : s \in S_r^c f(x), \;\; x \in X\} \leq M_0, \;\; \forall r > 0, \tag{7}$$

where M_0 is a constant from Proposition 2.

From Corollary 3 we get the following

Corollary 5 *Assume that the function f is semismooth at a point $x \in \mathrm{I\!R}^n$. Then*

$$S_0^c f(x) \subseteq \partial f(x).$$

Proposition 8 *Assume that the function f is regular and semismooth at a point $x \in \mathrm{I\!R}^n$. Then*

$$\partial f(x) = S_0^c f(x).$$

Proof: It follows from Corollary 5 and semismoothness of the function f that

$$S_0^c f(x) \subseteq \partial f(x).$$

Therefore we have to show that

$$\partial f(x) \subseteq S_0^c f(x).$$

Since the function f is regular it is directionally differentiable and

$$f'(x,e) = \max\{\langle v, e\rangle : v \in \partial f(x)\}.$$

Then it follows from Corollaries 2 and 5 that for any $e \in \mathrm{I\!R}^n$

$$\begin{aligned} f'(x,e) &\leq \max\{\langle v, e\rangle : \; v \in S_0^c f(x)\} \\ &\leq \max\{\langle v, e\rangle : \; v \in \partial f(x)\} \\ &= f'(x,e). \end{aligned}$$

Therefore for any $e \in \mathrm{I\!R}^n$

$$\max\{\langle v, e\rangle : \; v \in S_0^c f(x)\} = \max\{\langle v, e\rangle : \; v \in \partial f(x)\}.$$

Since both sets $S_0^c f(x)$ and $\partial f(x)$ are convex and compact we have that [12]

$$\partial f(x) = S_0^c f(x).$$

$\square$

Remark 6 Results of this section show that the set-valued mapping $x \mapsto S_0^c f(x)$, in general, is not upper semicontinuous.

4 Necessary conditions and descent directions

In this section we introduce r-stationary points, formulate necessary conditions for a minimum and demonstrate that the set $S_0^c f(x)$ can be used to compute descent directions. We assume that a function $f : \mathbb{R}^n \to \mathbb{R}^1$ is locally Lipschitz.

Proposition 9 *Assume that $r > 0$ is given and $f(x + rg) \geq f(x)$ for any $g \in S_1$. Then*

$$0_n \in S_r^c f(x). \tag{8}$$

Proof: It follows from (2) that for any $g \in S_1$

$$f(x + rg) - f(x) = \langle s(x, g, r), g\rangle \geq 0.$$

Since $S_r f(x) \subseteq S_r^c f(x)$ we have

$$\max\{\langle s, g\rangle : \ s \in S_r^c f(x)\} \geq 0 \tag{9}$$

for all $g \in S_1$. For any $e \in \mathbb{R}^n$, $e \neq 0$ there exist $g \in S_1$ and $\lambda > 0$ such that $e = \lambda g$, therefore the inequality (9) holds true for any $e \in \mathbb{R}^n$. Furthermore, $S_r^c f(x)$ is convex compact set and as a consequence we get the inclusion (8) (see [12]). □

Corollary 6 *Let $x \in \mathbb{R}^n$ be a local minimizer of the function f. Then there exists $r_0 > 0$ such that $0_n \in S_r^c f(x)$ for all $r \in (0, r_0]$.*

Proof: If $x \in \mathbb{R}^n$ is a local minimizer then there exists $r_0 > 0$ such that $f(x+rg) \geq f(x)$ for any $g \in S_1$ and $r \in (0, r_0]$. Then the proof follows from Proposition 9. □

Proposition 10 *Let $x \in \mathbb{R}^n$ be a local minimizer of the function f and it is directionally differentiable at x. Then*

$$0_n \in S_0^c f(x). \tag{10}$$

Proof: Since x is a local minimizer $f'(x, g) \geq 0$ for all $g \in \mathbb{R}^n$.Then it follows from Corollary 2 that

$$\max\{\langle v, g\rangle : \ v \in S_0^c f(x)\} \geq 0, \quad \forall g \in \mathbb{R}^n.$$

The set $S_0^c f(x)$ is convex compact and therefore $0_n \in S_0^c f(x)$. □

Definition 2 *Let $r > 0$. A point $x \in \mathbb{R}^n$ is said to be an r-stationary point for a function f on $\mathbb{R}^n$ if $0_n \in S_r^c f(x)$.*

Definition 3 *Let $r > 0$. A point $x \in \mathbb{R}^n$ is called an (r, δ)-stationary point for a function f on $\mathbb{R}^n$ if*

$$0_n \in S_r^c f(x) + B_\delta.$$

Assume that a point $x \in \mathbb{R}^n$ is not an r-stationary point of a function f on $\mathbb{R}^n$. This means that

$$0_n \notin S_r^c f(x).$$

In this case one can compute a descent direction using the set $S_r^c f(x)$.

Proposition 11 *Let $f : \mathbb{R}^n \to \mathbb{R}^1$ be a locally Lipschitz and $r > 0$. Assume that $x \in \mathbb{R}^n$ is not an r-stationary point, that is*

$$\min\{\|v\| : v \in S_r^c f(x)\} = \|v^0\| > 0.$$

Then for $g^0 = -\|v^0\|^{-1} v^0$

$$f(x + rg^0) - f(x) \leq -r\|v^0\|.$$

Proof: Since $S_r^c f(x)$ is compact and convex set we have

$$\max\left\{\langle v, g^0 \rangle : v \in S_r^c f(x)\right\} = -\|v^0\|.$$

Then it follows from (2) that

$$\begin{aligned} f(x + rg^0) - f(x) &= \langle s(x, g^0, r), g^0 \rangle \\ &\leq r \max\left\{\langle v, g^0 \rangle : v \in S_r^c f(x)\right\} \\ &= -r\|v^0\|. \end{aligned}$$

□

Remark 7 Proposition 11 shows that if $x \in \mathbb{R}^n$ is not an r-stationary point then the set $S_r^c f(x)$ can be used to compute descent directions at x. However, the computation of this set is not easy. In the next section we introduce an algorithm which computes descent directions using only a few elements from the set $S_r^c f(x)$.

Remark 8 If a point $x \in \mathbb{R}^n$ is not an (r, δ)-stationary point then in Proposition 11 $\|v^0\| \geq \delta$. In this case $f(x + rg^0) - f(x) \leq -r\delta$ and the set $S_r^c f(x)$ allows us to find a direction of sufficient decrease at x.

5 Computation of a descent direction

Proposition 11 implies that for the computation of the descent direction we have to solve the following problem:

$$\text{minimize } \|v\|^2 \quad \text{subject to } v \in S_r^c f(x). \tag{11}$$

Problem (11) is difficult to solve due to difficulties to compute the set $S_r^c f(x)$. In this section we propose an algorithm which uses only a few elements of $S_r^c f(x)$.

Let the numbers $r > 0,\ c \in (0,1)$ and a small enough number $\delta > 0$ be given.

Algorithm 1 The computation of descent directions.

Step 1. Select any $g^1 \in S_1$ and compute an r-secant $s^1 = s(x, g^1, r)$ in the direction g^1. Set $\overline{W}_1(x) = \{s^1\}$ and $k = 1$.

Step 2. Compute the vector $\|w^k\|^2 = \min\{\|w\|^2 : w \in \text{co}\, \overline{W}_k(x)\}$. If

$$\|w^k\| < \delta, \tag{12}$$

then stop. Otherwise go to Step 3.

Step 3. Compute the search direction by $g^{k+1} = -\|w^k\|^{-1} w^k$.

Step 4. If

$$f(x + rg^{k+1}) - f(x) \le -cr\|w^k\|, \tag{13}$$

then stop. Otherwise go to Step 5.

Step 5. Compute an r-secant $s^{k+1} = s(x, g^{k+1}, r)$ in the direction g^{k+1}, construct the set $\overline{W}_{k+1}(x) = \text{co}\left\{\overline{W}_k(x) \bigcup \{s^{k+1}\}\right\}$, set $k = k+1$ and go to Step 2.

Remark 9 In Step 1 we find an r-secant in the initial direction g^1. In Step 2 the least distance between the convex hull of r-secants (computed so far) and the origin is computed. It is a quadratic programming problem and effective algorithms exist for its solution (see, for example, [14, 34]). If the distance is less than a given tolerance $\delta > 0$ then the point x is an (r, δ)-stationary point, otherwise we compute a new search direction in Step 3. If it is descent direction the algorithm stops and the descent direction has been found (Step 4). If it is not descent direction then in Step 5 we compute a new r-secant in this direction. It improves the approximation of the set $S_r^c f(x)$.

In the next proposition we prove that Algorithm 1 is a terminating.

Proposition 12 *Assume the function f is locally Lipschitz and $c \in (0,1), \delta \in (0, M_0)$, where M_0 is a constant from Corollary 4. Then Algorithm 1 terminates after m steps, where*

$$m \le \frac{2\log_2(\delta/M_0)}{\log_2 M_1} + 2, \quad M_1 = 1 - [(1-c)(2M_0)^{-1}\delta]^2.$$

Proof: First, we show that if both stopping criteria (12) and (13) are not satisfied, then the r-secant s^{k+1}, computed in Step 5, does not belong to $\overline{W}_k(x)$ and consequently the approximation of the set $S_r^c f(x)$ is improved. Indeed, in this case $\|w^k\| > \delta$ and

$$f(x + rg^{k+1}) - f(x) > -cr\|w^k\|.$$

On the other hand it follows from (2) that

$$f(x + rg^{k+1}) - f(x) = r\langle s^{k+1}, g^{k+1}\rangle.$$

Then from the definition of the direction g^{k+1} we get

$$\langle s^{k+1}, w^k\rangle < c\|w^k\|^2. \tag{14}$$

Since $w^k = \operatorname{argmin}\{\|w\|^2 : w \in \overline{W}_k(x)\}$, the necessary condition for a minimum implies that

$$\langle w^k, w - w^k\rangle \geq 0, \quad \forall w \in \overline{W}_k(x)$$

or

$$\langle w^k, w\rangle \geq \|w^k\|^2, \ \forall w \in \overline{W}_k(x).$$

The latter along with (14) means that $s^{k+1} \notin \overline{W}_k(x)$.

Now we show that the algorithm is a terminating. It is sufficient to get an upper estimation for the number m when

$$\|w^m\| \leq \delta. \tag{15}$$

Clearly $\|w^{k+1}\|^2 \leq \|ts^{k+1} + (1-t)w^k\|^2$ for all $t \in [0, 1]$ or

$$\|w^{k+1}\|^2 \leq \|w^k\|^2 + 2t\langle w^k, s^{k+1} - w^k\rangle + t^2\|s^{k+1} - w^k\|^2.$$

It follows from Corollary 4 that there exists $M_0 > 0$ such that

$$\|s^{k+1} - w^k\| \leq 2M_0.$$

Hence taking into account the inequality (14), we have

$$\|w^{k+1}\|^2 \leq \|w^k\|^2 - 2t(1-c)\|w^k\|^2 + 4t^2M_0^2.$$

If $t = (1-c)(2M_0)^{-2}\|w^k\|^2 \in (0, 1)$ we get

$$\|w^{k+1}\|^2 \leq \left\{1 - [(1-c)(2M_0)^{-1}\|w^k\|]^2\right\} \|w^k\|^2. \tag{16}$$

Take any $\delta \in (0, M_0)$. It follows from (16) and the condition $\|w^k\| > \delta, k = 1, \ldots, m-1$ that

$$\|w^{k+1}\|^2 \leq \{1 - [(1-c)(2M_0)^{-1}\delta]^2\}\|w^k\|^2.$$

Let $M_1 = 1 - [(1-c)(2M_0)^{-1}\delta]^2$. It is clear that $M_1 \in (0, 1)$. Consequently,

$$\|w^m\|^2 \leq M_1\|w^{m-1}\|^2 \leq \ldots \leq M_1^{m-1}\|w^1\|^2 \leq M_1^{m-1}M_0^2.$$

If $M_1^{m-1}M_0^2 \leq \delta^2$, then the inequality (15) holds true and

$$m \leq 2\log_2(\delta/M_0)/\log_2 M_1 + 2.$$

□

6 A secant method and its convergence

In this section we describe the secant method for solving Problem (1). First, we describe an algorithm for finding (r, δ)-stationary points.

Let $r > 0,\ \delta > 0,\ c_1 \in (0,1), c_2 \in (0, c_1]$ be given numbers.

Algorithm 2 A secant method for computation of (r, δ)-stationary points.

Step 1. Select any starting point $x^0 \in \mathbb{R}^n$ and set $k = 0$.

Step 2. Apply Algorithm 1 for the computation of the descent direction at $x = x^k$ for given $\delta > 0, c = c_1$. This algorithm terminates after a finite number of iterations $m > 0$. As a result we get the set $\overline{W}_m(x^k)$ and an element v^k such that

$$\|v^k\|^2 = \min\left\{\|v\|^2 :\ v \in \overline{W}_m(x^k)\right\}.$$

Furthermore either $\|v^k\| \leq \delta$ or for the search direction $g^k = -\|v^k\|^{-1}v^k$

$$f(x^k + rg^k) - f(x^k) \leq -c_1 r\|v^k\|. \tag{17}$$

Step 3. If

$$\|v^k\| < \delta \tag{18}$$

then stop. Otherwise go to Step 4.

Step 4. Construct the following iteration $x^{k+1} = x^k + \alpha_k g^k$, where α_k is defined as follows

$$\alpha_k = \text{argmax}\left\{\alpha \geq 0 :\ f(x^k + \alpha g^k) - f(x^k) \leq -c_2\alpha\|v^k\|\right\}.$$

Set $k = k + 1$ and go to Step 2.

For the point $x^0 \in \mathbb{R}^n$ consider the set $\mathcal{L}(x^0) = \{x \in \mathbb{R}^n : f(x) \leq f(x^0)\}$.

Theorem 1 *Assume that f is locally Lipschitz function and the set $\mathcal{L}(x^0)$ is bounded for starting points $x^0 \in \mathbb{R}^n$. Then after finitely many iterations $K > 0$ Algorithm 2 finds (r, δ)-stationary point, where*

$$K \leq 1 + \frac{f(x^0) - f_*}{c_2 r\delta},\quad f_* = \inf\{f(x) : x \in \mathbb{R}^n\}.$$

Proof: The function f is locally Lipschitz and the set $\mathcal{L}(x^0)$ is bounded and therefore, $f_* > -\infty$. Since $c_2 \in (0, c_1]$ it follows from (17) that $\alpha_k \geq r$ and the definition of α_k implies that

$$\begin{aligned} f(x^{k+1}) - f(x^k) &\leq -c_2\alpha_k\|v^k\| \\ &\leq -c_2 r\|v^k\|. \end{aligned}$$

If x^k is not an (r, δ)-stationary point then $\|v^k\| > \delta$ and

$$f(x^{k+1}) - f(x^k) < -c_2 r\delta.$$

Thus Algorithm 2 at iteration x^k, which is not an (r, δ)-stationary point, decreases the value of the function f at least for $-c_2 r\delta$. Since the sequence $\{f(x^k)\}$ is decreasing the number of such iterations cannot be more than

$$K = 1 + \frac{f(x^0) - f_*}{c_2 r\delta}$$

and after at most K iterations the algorithm must find a point x^K where $\|v^k\| < \delta$. □

Remark 10 Since $c_2 \leq c_1$ the step-length $\alpha_k \geq r$ and therefore $r > 0$ is a lower bound for α_k. This leads to the following rule for the computation of α_k. Consider a sequence:

$$\theta_l = lr, \quad l = 1, 2, \ldots.$$

Then α_k is defined as the largest θ_l satisfying the inequality in Step 4.

Algorithm 2 can be applied to compute Clarke stationary points of the function f. Let $\{r_k\}, \ \{\delta_k\}$ be sequences such that $r_k \to +0$ and $\delta_k \to +0$ as $k \to +\infty$. Applying Algorithm 2 for each $\{r_k\}, \ \{\delta_k\}$ we get a sequence $\{x^k\}$ of $(r_k, \ \delta_k)$-stationary points. Thus, we get the following algorithm.

Algorithm 3 A secant method for computation of stationary points.

Step 1. Choose any starting point $x^0 \in \mathbb{R}^n$ and set $k = 1$.

Step 2. Apply Algorithm 2 starting from the point x^{k-1} for $r = r_k$ and $\delta = \delta_k$. This algorithm terminates after a finitely many iterations $p > 0$ and as a result it finds (r_k, δ_k)-stationary point x^k.

Step 3. Set $k = k + 1$ and go to Step 2.

Convergence of Algorithm 3 can be proved under more restrictive assumptions. We assume that the function f satisfies the following

Assumption 1 *At a given point $x \in \mathbb{R}^n$ for any $\varepsilon > 0$ there exist $\eta > 0$ and $r_0 > 0$ such that*

$$S_r^c f(x) \subset \partial f(x + B_\varepsilon) + B_\varepsilon \tag{19}$$

for all $y \in B_\eta(x)$ and $r \in (0, r_0)$. Here

$$\partial f(x + B_\varepsilon) = \bigcup_{y \in B_\varepsilon(x)} \partial f(y).$$

Theorem 2 *Assume that the function f is locally Lipschitz, it satisfies Assumption 1 at any $x \in \mathbb{R}^n$ and the set $\mathcal{L}(x^0)$ is bounded for any $x^0 \in \mathbb{R}^n$. Then every accumulation point of the sequence $\{x^k\}$ belongs to the set $X^0 = \{x \in \mathbb{R}^n : 0_n \in \partial f(x)\}$.*

Proof: All conditions of Theorem 1 are satisfied and therefore Algorithm 2 for all $k \geq 0$ generates (r_k, δ_k)-stationary point after the finite many iterations. Since for any $k > 0$ the point x^k is (r_k, δ_k)-stationary it follows from the definition of the (r_k, δ_k)-stationarity that

$$\min \left\{ \|v\| : v \in S^c_{r_k} f(x^{k+1}) \right\} \leq \delta_k. \tag{20}$$

Since $x^k \in \mathcal{L}(x^0)$ for all $k \geq 0$ and the set $\mathcal{L}(x^0)$ is bounded, it follows that the sequence $\{x^k\}$ has at least one accumulation point. Let x^* be an accumulation point and $x^{k_i} \to x^*$ as $i \to +\infty$. Then it follows from (20) that

$$\min \left\{ \|v\| : v \in S^c_{r_{k_i}} f(x^{k_i}) \right\} \leq \delta_{k_i}. \tag{21}$$

Assumption 1 implies that at the point x^*, for any $\varepsilon > 0$ there exist $\eta > 0$ and $r_0 > 0$ such that

$$S^c_r f(y) \subset \partial f(x^* + B_\varepsilon) + B_\varepsilon \tag{22}$$

for all $y \in B_\eta(x^*)$ and $r \in (0, r_0)$. Since the sequence $\{x^{k_i}\}$ converges to x^* for $\eta > 0$ there exists $i_0 > 0$ such that $x^{k_i} \in B_\eta(x^*)$ for all $i \geq i_0$. On the other hand since $\delta_k, r_k \to +0$ as $k \to +\infty$ there exists $k_0 > 0$ such that $\delta_k < \varepsilon$ and $r_k < r_0$ for all $k > k_0$. Then there exists $i_1 \geq i_0$ such that $k_i \geq k_0 + 1$ for all $i \geq i_1$. Thus it follows from (21) and (22) that

$$\min\{\|v\| : \ v \in \partial f(x^* + B_\varepsilon)\} \leq 2\varepsilon.$$

Since $\varepsilon > 0$ is arbitrary and the mapping $x \mapsto \partial f(x)$ is upper semicontinuous $0 \in \partial f(x^*)$. □

Remark 11 There are some similarities between the secant method and the bundle and the gradient sampling methods. In order to find descent directions in the secant method a bundle of r-secants is computed which makes it similar to the bundle method. Both the secant method and the gradient sampling method do not use subgradient discounting rules. In order to find descent directions the secant method computes a finite number of secants at the current iteration and the gradient sampling method computes a finite number of gradients at points randomly chosen from some neighborhood of the current iteration. However, all three methods are based on different approaches.

7 Approximation of subgradients

In the definition of the r-secants it is assumed that one can compute at least one subgradient at any point. However, there are many interesting nonconvex, nonsmooth, nonregular functions where the computation of even one subgradient is difficult task. In this section we consider one approach to approximate subgradients. This approach is introduced in [4, 5]. All necessary proofs also can be found in these papers.

We consider a function f defined on $\mathbb{R}^n$ and assume that this function is quasidifferentiable. The function f is called quasidifferentiable at a point x if it is locally Lipschitz continuous, directionally differentiable at this point and there exist convex, compact sets $\underline{\partial} f(x)$ and $\overline{\partial} f(x)$ such that:

$$f'(x, g) = \max_{u \in \underline{\partial} f(x)} \langle u, g \rangle + \min_{v \in \overline{\partial} f(x)} \langle v, g \rangle.$$

The set $\underline{\partial} f(x)$ is called a subdifferential, the set $\overline{\partial} f(x)$ is called a superdifferential and the pair $[\underline{\partial} f(x), \overline{\partial} f(x)]$ is called a quasidifferential of the function f at a point x [11].

We assume that both sets $\underline{\partial} f(x)$ and $\overline{\partial} f(x)$ are polytopes at any $x \in \mathbb{R}^n$. This assumption holds true, for example, for functions represented as a maximum, minimum or max-min of a finite number of smooth functions.

Let

$$G = \{e \in \mathbb{R}^n : e = (e_1, \ldots, e_n), |e_j| = 1, \ j = 1, \ldots, n\}$$

be the set of all vertices of the unit hypercube in $\mathbb{R}^n$. We take $e \in G$ and consider the sequence of n vectors $e^j = e^j(\alpha), \ j = 1, \ldots, n$ with $\alpha \in (0, 1]$:

$$\begin{array}{lcl} e^1 & = & (\alpha e_1, 0, \ldots, 0), \\ e^2 & = & (\alpha e_1, \alpha^2 e_2, 0, \ldots, 0), \\ \ldots & = & \ldots\ldots\ldots \\ e^n & = & (\alpha e_1, \alpha^2 e_2, \ldots, \alpha^n e_n). \end{array}$$

Given a point $x \in \mathbb{R}^n$ we define the following points

$$x^0 = x, \quad x^j = x^0 + \lambda e^j(\alpha), \quad j = 1, \ldots, n.$$

It is clear that

$$x^j = x^{j-1} + (0, \ldots, 0, \lambda \alpha^j e_j, 0, \ldots, 0), \quad j = 1, \ldots, n.$$

Define a vector $v = v(e, \alpha, \lambda) \in \mathbb{R}^n$ with the following coordinates:

$$v_j = (\lambda \alpha^j e_j)^{-1} \left[f(x^j) - f(x^{j-1}) \right], \quad j = 1, \ldots, n. \tag{23}$$

For the fixed $e \in G$ and $\alpha > 0$ we introduce the following set:

$$V(e, \alpha) = \left\{ w \in \mathbb{R}^n : \quad \exists (\lambda_k \to +0, \quad k \to +\infty), \quad w = \lim_{k \to +\infty} v(e, \alpha, \lambda_k) \right\}.$$

Proposition 13 *[4, 5] Assume that f is a quasidifferentiable function and its subdifferential and superdifferential are polytopes at x. Then there exists $\alpha_0 > 0$ such that*

$$V(e,\alpha) \subset \partial f(x)$$

for any $\alpha \in (0, \alpha_0]$.

Remark 12 It follows from Proposition 13 that in order to approximate subgradients of quasidifferentiable functions one can choose a vector $e \in G$, sufficiently small $\alpha > 0,\ \lambda > 0$ and apply (23) to compute a vector $v(e,\alpha,\lambda)$. This vector is an approximation to a subgradient.

Remark 13 Quasidifferentiable functions present a broad class of nonsmooth functions, including many interesting non-regular functions. Thus, this approach allows one to approximate subgradients of a broad class of nonsmooth functions.

8 Results of numerical experiments

The efficiency of the proposed algorithm was verified by applying it to some academic test problems with nonsmooth objective functions. We consider three types of problems:

1. Problems with nonsmooth convex objective functions;
2. Problems with nonsmooth nonconvex regular objective functions;
3. Problems with nonsmooth, nonconvex and nonregular objective functions.

Test Problems 2.1-7, 2.9-12, 2.14-16, 2.18-21 and 2.23-25 from [23] and Problems 1-3, 5 and 7 from [2] have been used in numerical experiments. We also include the following problem with nonsmooth, nonconvex and nonregular objective function.

Problem 1

$$\text{minimize} \quad f(x) = \sum_{i=1}^{p} \min_{j=1,\dots,k} \|x^j - a^i\|^2$$

Here $p = 20,\ k = 5,\ x = (x^1, \dots, x^5) \in \mathbb{R}^{15}$ and the vectors $a^i \in R^3,\ i = 1, \dots, 20$ are as follows:

This is a well known clustering function (see [7]). We apply it to solve clustering problem on two real-world data sets: TSPLIB1060 and TSPLIB3038. The description of these data sets can be found in [31]. The first data set contains 1060 2-dimensional points and the second data set contains 3038 2-dimensional points. We compute 3, 5 and 10 clusters for each data set.

The brief description of test problems are given in Table 1, where the following notation is used:

a^1	a^2	a^3	a^4	a^5	a^6	a^7	a^8	a^9	a^{10}
1.1	0.8	0.1	0.6	-1.2	0.9	0.2	-0.3	-0.8	0.0
1.0	-1.6	-1.0	0.2	1.0	1.9	0.2	-0.2	0.6	-0.4
-0.1	0.3	-0.3	0.2	1.4	-0.8	0.0	0.8	-0.2	0.6
a^{11}	a^{12}	a^{13}	a^{14}	a^{15}	a^{16}	a^{17}	a^{18}	a^{19}	a^{20}
1.0	0.0	0.0	2.1	0.2	-2.1	-1.0	0.3	1.1	3.1
0.0	1.0	0.0	-1.4	-1.0	0.0	0.5	-2.0	1.2	-1.5
0.0	0.0	1.0	1.0	1.0	-1.0	1.5	0.9	1.0	2.1

- n - number of variables;
- n_m - the total number of functions under maximum and minimum (if the function contains maximum and minimum functions);
- f_{opt} - optimum value.

The objective functions in Problems P27-P33 are the clustering function. In Problem P27 the number of clusters $k = 5$, the number of data points $p = 20$ and they are given in Table 1. In Problems P28-P30 data points are from the data set TSPLIB1060 and the number of clusters in Problem P28 is 3, in Problem P29 it is 5 and in Problem P30 10. In Problems P31-P33 data points are from the data set TSPLIB3038 and the number of clusters in Problem P31 is 3, in Problem P32 is 5 and in Problem P33 it is 10.

In our experiments we use two bundle algorithms for comparisons: PMIN - a recursive quadratic programming variable metric algorithm for minimax optimization and PBUN - a proximal bundle algorithm. The description of these algorithms can be found in [24]. PMIN is applied to minimize maximum functions and PBUN is applied to the rest of problems. In PMIN exact subgradients are used however in PBUN we approximate them using the scheme from Section 7 where $\lambda = 10^{-8},\ \alpha = 0.8$.

In Algorithm 3, $c_1 \in (0.2, 0.5),\ c_2 = 0.001$ and subgradients are approximated using the scheme from Section 7 where $\lambda = 10^{-8},\ \alpha = 0.8$. The sequences $\{r_k\}$ and $\{\delta_k\}$ were chosen as follows: $r_{k+1} = 0.6 r_k,\ r_0 = 5$ and $\delta_k \equiv 10^{-7}$.

20 starting points have been randomly generated and well posed for each of test problems. To compare the performance of the algorithms, we use two indicators: n_b - the number of successful runs considering the best known solution and n_s - the number of successful runs considering the best found solution by these two algorithms. Assume that f_{opt} and $\bar{f}$ are the values of the objective function at the best known solution and at the best found solution, respectively. Then we say that an algorithm finds the best solution with respect to a tolerance $\varepsilon > 0$ if

$$\frac{f_* - f_0}{1 + |f_*|} \le \varepsilon$$

where f_* is equal either to f_{opt} (for n_b) or to $\bar{f}$ (for n_s) and f_0 is the optimal value of the objective function found by an algorithm. In our experiments $\varepsilon = 10^{-4}$.

The results of numerical experiments are presented in Table 2. We also present f_{av} the average objective value over 20 runs of algorithms.

Results presented in Table 2 show that both the secant method and the bundle method (PMIN) are effective methods for the minimization of nonsmooth convex functions. However, the bundle method is faster and more accurate than the secant method. Results for nonconex nonsmooth, regular functions show that overall the bundle method (PMIN) produces better results than the secant method on this class of nonsmooth optimization problems. Indeed, the bundle method in 69.71 % of cases finds the best known solutions whereas for the secant method it is only 49.71 %. The bundle method in 79.80 % of cases produces the best solutions among two algorithms, but for the secant method it is 68.82 %. On the same time we should note that the secant method performs much better than the bundle on some nonsmooth, nonconvex, regular problems (Problems P7, P14, P15).

Overall the secant method performs better than the bundle method (PBUN) on nonconvex, nonsmooth nonregular functions. The secant method in 55.83 % of cases finds the best known solutions whereas the bundle method finds only in 44.58 % of all cases. The secant method produces best solutions among two algorithms in 94.58 % of all cases, but the bundle method only in 64.17 % of all cases. However, for some nonconvex, nonsmooth nonregular functions the results by both algorithms are comparable (P22, P23, P31, P32). It should be noted the objective functions in Problems P22-P33 are quasidiferentiable semismooth and their subdifferential and superdifferential are polytopes.

Table 3 presents the average number of objective function (n_f) and subgradient (n_{sub}) evaluations as well as the average CPU time (t) for both algorithms on each of test problems.

Results presented in Table 3 demonstrate that in most of cases the bundle method requires significantly less objective function, subgradient evaluations and also CPU time. However on some problems (Problems: P9, P11, P15, P16, P21) the bundle method requires significantly more function, subgradient evaluations and CPU time than the secant method. In this problems most of starting points are far away from local minimizers and the bundle method fails to quickly find the solution.

9 Conclusions

In this paper we developed the secant method for minimizing nonsmooth nonconvex functions. We introduced the notion of an r-secant and studied some of its properties. The convex hull of all limit points of r-secant is subset of the clarke subdifferential of semismooth functions. An algorithm based on r-secants for the computation of descent direction is developed and it is proved that this algorithm is terminating.

We presented results of numerical experiments. In these experiments minimization problems with nonsmooth, nonconvex and nonregular objective functions were considered. The computational results show that the secant method outperforms the bundle method when the objective is nonsmooth, nonconvex nonregular and it is semismooth qusidifferentiable function where both subdifferential and superdifferential are polytopes. However, in most cases the secant method requires significantly more objective function, subgradient evaluations as well as CPU time. We applied the secant method straitforward. However, taking into account the special structure of the objective functions such as their piecewise partially separability (see [6]) one can significantly decrease the number of the objective function and subgradient evaluations and also CPU time.

Although the secant method is not better than the bundle method for minimizing nonsmooth convex functions as well as many nonconvex, nonsmooth regular functions, however, it is better than the bundle method for minimizing many nonconvex nonsmooth nonregular functions.

Acknowledgements

Dr. Adil Bagirov is the recipient of an Australian Research Council Australian Research Fellowship (Project number: DP 0666061).

References

[1] A.M. Bagirov, Minimization methods for one class of nonsmooth functions and calculation of semi-equilibrium prices, In: A. Eberhard et al. (eds.) Progress in Optimization: Contribution from Australasia, Kluwer Academic Publishers, 1999, pp. 147-175.

[2] A.M. Bagirov, A method for minimizing of quasidifferentiable functions, Optimization Methods and Software, 17(1)(2002), pp. 31-60.

[3] A.M. Bagirov, Continuous subdifferential approximations and their applications, Journal of Mathematical Sciences, 115(5)(2003), pp. 2567-2609.

[4] A.M. Bagirov, M. Ghosh and D. Webb, A derivative-free method for linearly constrained nonsmooth optimization, *Journal of Industrial and Management Optimization,* 2(3)(2006), pp. 319-338.

[5] A.M. Bagirov, B. Karasozen and M. Sezer, Discrete gradient method: a derivative free method for nonsmooth optimization, *Journal of Optimization Theory and Applications,* 2008, accepted for publication.

[6] A.M. Bagirov, and J. Ugon, Piecewise partially separable functions and a derivative-free method for large-scale nonsmooth optimization. *Journal of Global Optimization* 35, 2006, pp. 163-195.

[7] A.M. Bagirov and J. Yearwood, A new nonsmooth optimization algorithm for minimum sum-of-squares clustering problems, European Journal of Operational Research, 170(2)(2006), pp. 578-596.

[8] J.V. Burke, A.S. Lewis and M.L. Overton, A robust gradient sampling algorithm for nonsmooth, nonconvex optimization, SIAM Journal on Optimization, 15(3)(2005), pp. 751-779.

[9] F.H. Clarke, Generalized gradients and applications, Trans. Amer. Math. Soc., 205(1975), pp. 247 - 262.

[10] F.H. Clarke, Optimization and Nonsmooth Analysis, New York, John Wiley, 1983.

[11] V.F. Demyanov, and A.M. Rubinov, Constructive Nonsmooth Analysis, Peter Lang, Frankfurt am Main, 1995.

[12] V.F. Demyanov, and L. Vasilyev, Nondifferentiable Optimization, Optimization Software Inc., New York, 1985.

[13] A. Frangioni, Generalized bundle methods, *SIAM J. Optimization,* 113(1)(2002), pp. 117-156.

[14] A. Frangioni, Solving semidefinite quadratic problems within nonsmooth optimization algorithms, *Computers & Operations Research,* Vol. 23, pp. 1099-1118, 1996.

[15] M. Fukushima and L. Qi, A globally and superlinearly convergent algorithm for nonsmooth convex minimization, *SIAM J. Optimization,* 6(1996), pp. 1106-1120.

[16] M. Gaudioso and M.F. Monaco, A bundle type approach to the unconstrained minimization of convex nonsmooth functions, Mathematical Programming, 23(1982), pp. 216-226.

[17] J.B. Hiriart-Urruty and C. Lemarechal, Convex Analysis and Minimization Algorithms, Springer Verlag, Heidelberg, Vol. 1 and 2, 1993.

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

[19] K.C. Kiwiel, Proximal control in bundle methods for convex nondifferentiable minimization, Mathematical Programming, 29(1990), pp. 105-122.

[20] C. Lemarechal, An extension of Davidon methods to nondifferentiable problems, Nondifferentiable Optimization, Balinski, M.L. and Wolfe, P. (eds.), Mathematical Programming Study, 3(1975), pp. 95-109.

[21] C. Lemarechal and C.A. Sagastizabal, Practical aspects of the Moreau-Yosida regularization I: Theoretical preliminaries, *SIAM J. Optimization,* 7(1997), pp. 367-385.

[22] L. Luksan and J. Vlcek, A bundle Newton method for nonsmooth unconstrained minimization, Mathematical Programming, 83(1998), pp. 373-391.

[23] L. Luksan and J. Vlcek, Test Problems for Nonsmooth Unconstrained and Linearly Constrained Optimization, Technical Report No. 78, Institute of Computer Science, Academy of Sciences of the Czech Republic, 2000.

[24] L. Luksan and J. Vlcek, Algorithm 811: NDA: Algorithms for nondifferentiable optimization, *ACM Trans. on Math. Software,* 27(2), 2001, pp. 193-213.

[25] M.M. Makela and P. Neittaanmaki, Nonsmooth Optimization, World Scientific, Singapore, 1992.

[26] R. Mifflin, Semismooth and semiconvex functions in constrained optimization, SIAM J. Control and Optimization, 15(6)(1977), pp. 959-972.

[27] R. Mifflin, An algorithm for constrained optimization with semismooth functions, Mathematics of Operations Research, 2(1977), pp. 191-207.

[28] R. Mifflin, A quasi-second-order proximal bundle algorithm, Mathematical Programming, 73(1996), pp. 51-72.

[29] R. Mifflin, L. Qi and D. Sun, Quasi-Newton bundle-type methods for nondifferentiable convex optimization, SIAM J. Optimization, 8(2)(1998), pp. 583-603.

[30] E. Polak and J.O. Royset, J.O., Algorithms for finite and semi-infinite min-max-min problems using adaptive smoothing techniques, Journal of Optimization Theory and Applications, 119(3)(2003), pp. 421-457.

[31] G. Reinelt, TSP-LIB-A Traveling salesman library, *ORSA J. Comput.*, 3, 1991, pp. 319-350.

[32] N.Z. Shor, Minimization Methods for Non-Differentiable Functions, Springer-Verlag, Heidelberg, 1985.

[33] P.H. Wolfe, A method of conjugate subgradients of minimizing nondifferentiable convex functions, Math. Prog. Study, 3(1975), pp. 145 - 173.

[34] P.H. Wolfe, Finding the nearest point in a polytope, Mathematical Programming, 11(2)(1976), pp. 128-149.

[35] J. Zowe, Nondifferentiable optimization: A motivation and a short introduction into the subgradient and the bundle concept, In: NATO SAI Series, 15, Computational Mathematical Programming, Schittkowski, K., (ed.), Springer-Verlag, New York, 1985, pp. 323-356.

Table 1: The description of test problems

Function type	Problems	n	n_m	f_{opt}
Nonsmooth convex	P1 (Problem 2.1 [23])	2	3	1.9522245
	P2 (Problem 2.5 [23])	4	4	-44
	P3 (Problem 2.23 [23])	11	10	261.08258
Nonsmooth nonconvex regular	P4 (Problem 2.2 [23])	2	3	0
	P5 (Problem 2.3 [23])	2	2	0
	P6 (Problem 2.4 [23])	3	6	3.5997193
	P7 (Problem 2.6 [23])	4	4	-44
	P8 (Problem 2.7 [23])	3	21	0.0042021
	P9 (Problem 2.9 [23])	4	11	0.0080844
	P10 (Problem 2.10 [23])	4	20	115.70644
	P11 (Problem 2.11 [23])	4	21	0.0026360
	P12 (Problem 2.12 [23])	4	21	0.0020161
	P13 (Problem 2.14 [23])	5	21	0.0001224
	P14 (Problem 2.15 [23])	5	30	0.0223405
	P15 (Problem 2.16 [23])	6	51	0.0349049
	P16 (Problem 2.18 [23])	9	41	0.0061853
	P17 (Problem 2.19 [23])	7	5	680.63006
	P18 (Problem 2.20 [23])	10	9	24.306209
	P19 (Problem 2.21 [23])	20	18	133.72828
	P20 (Problem 2.24 [23])	20	31	0.0000000
	P21 (Problem 2.25 [23])	11	65	0.0480274
Nonsmooth nonconvex nonregular	P22 (Problem 1 [2])	2	6	2
	P23 (Problem 2 [2])	2	-	0
	P24 (Problem 3 [2])	4	-	0
	P25 (Problem 5 [2])	5	-	0
	P26 (Problem 7 [2])	5	-	0
	P27 (Problem 1)	15	100	13.311214
	P28 (Problem 1)	6	3180	6.32621×10^6
	P29 (Problem 1)	10	5300	3.57642×10^6
	P30 (Problem 1)	20	10600	2.13505×10^6
	P31 (Problem 1)	6	9114	7.16372×10^5
	P32 (Problem 1)	10	15190	3.94402×10^5
	P33 (Problem 1)	20	30380	1.84415×10^5

Table 2: Results of numerical experiments

Prob.	Secant			Bundle		
	f_{av}	n_b	n_s	f_{av}	n_b	n_s
P1	1.95222	20	20	1.95222	20	20
P2	-44	20	20	-44	20	20
P3	3.70348	20	20	3.70348	20	20
P4	0.90750	17	17	0.90750	17	17
P5	0.57592	4	4	0.00000	20	20
P6	3.59972	20	20	3.59972	20	20
P7	-44	20	20	-28.14011	12	12
P8	0.03565	7	13	0.03051	9	17
P9	0.02886	0	7	0.01520	2	16
P10	115.70644	20	20	115.70644	20	20
P11	0.00291	0	0	0.00264	20	20
P12	0.01773	0	6	0.02752	14	16
P13	0.10916	3	9	0.30582	3	15
P14	0.24037	8	18	0.32527	5	9
P15	0.03490	20	20	0.30572	12	12
P16	0.12402	0	11	0.39131	2	9
P17	680.63006	20	20	680.63006	20	20
P18	24.30621	20	20	24.30621	20	20
P19	93.90566	20	20	93.90525	20	20
P20	0.00302	0	0	0.00000	20	20
P21	0.21456	0	9	0.22057	1	14
P22	2.00000	20	20	2.00000	20	20
P23	0.10000	18	18	0.07607	17	17
P24	1.50000	7	18	2.30008	2	10
P25	0.00000	20	20	0.00000	20	20
P26	0.00000	20	20	0.00000	20	20
P27	24.72876	0	18	35.19230	0	2
P28	8.53416×10^6	10	18	10.29861×10^6	5	12
P29	5.56355×10^6	3	20	7.10874×10^6	0	9
P30	2.93562×10^6	2	19	3.16944×10^6	0	2
P31	7.27436×10^5	11	19	7.26213×10^5	12	20
P32	3.94879×10^5	10	19	3.94922×10^5	5	13
P33	1.86467×10^5	13	18	1.88295×10^5	6	9

Table 3: The number of function and subgradient evaluations and CPU time

Prob.	Secant			Bundle		
	n_f	n_{sub}	t	n_f	n_{sub}	t
P1	221	160	0.000	10	10	0.001
P2	1113	601	0.002	12	11	0.001
P3	974	701	0.143	65	55	0.003
P4	749	593	0.002	22	9	0.000
P5	4735	461	0.002	493	251	0.001
P6	574	309	0.001	20	16	0.000
P7	1225	579	0.003	146	56	0.001
P8	1184	342	0.009	1626	171	0.009
P9	2968	601	0.003	57891	4843	0.186
P10	1192	619	0.010	29	15	0.001
P11	1829	811	0.005	26081	1904	0.122
P12	4668	1327	0.014	372	173	0.007
P13	1310	749	0.014	61	26	0.002
P14	837	686	0.020	51	23	0.002
P15	1373	1090	0.085	4985	445	0.108
P16	3463	1989	0.302	60331	4761	1.099
P17	1270	860	0.012	58	33	0.000
P18	2234	1592	0.029	18	15	0.000
P19	4752	4001	0.362	35	26	0.003
P20	3399	2733	3.393	160	52	0.030
P21	1529	849	0.249	66280	4974	2.971
P22	289	198	0.001	32	32	0.000
P23	2278	277	0.001	22	22	0.000
P24	470	450	0.003	37	37	0.000
P25	1022	983	0.010	25	25	0.001
P26	2677	920	0.005	68	68	0.001
P27	725	707	0.039	37	37	0.002
P28	305	289	0.099	16	16	0.007
P29	417	401	0.360	21	21	0.022
P30	766	736	2.734	51	51	0.205
P31	283	257	0.254	19	19	0.024
P32	406	370	0.955	28	28	0.085
P33	703	653	7.003	67	67	0.763