Nonsmooth Steepest Descent Method by Proximal Subdifferentials in Hilbert Spaces

Zhou Wei · Qing Hai He

Abstract In this paper, we first study nonsmooth steepest descent method for nonsmooth functions defined on Hilbert space and establish the corresponding algorithm by proximal subgradients. Then, we use this algorithm to find stationary points for those functions satisfying prox-regularity and Lipschitzian continuity. As one application, the established algorithm is used to search the minimizer of lower semicontinuous and convex functions on finite-dimensional space. The convergent theorem, as one extension and improvement of the existing converging result for twice continuously differentiable convex functions, is also presented therein.

1. Introduction

Given a smooth function on an Euclidean space, one kind of optimization problem is to find the global/local minimizers of the given function. This kind of problems has been receiving a great deal of attention and it is of significance to solve it by using the iteration of line search methods. For solving this problem, it is used generally to find the steepest descent unit direction, and then the given function can be reduced along this direction and finally attain the global minimizer hopefully. This so-called steepest descent method nowadays has been applied to various application areas such as numerical optimization, computing mathematics, probabilistic analysis, algorithm, differential equations, vector optimization and so on (cf. [1,2,3,4,5,6,7,8] and references therein). This method could be traced back

Zhou Wei(✉)
Department of Mathematics, Yunnan University, Kunming 650091, P. R. China
E-mail: wzhou@ynu.edu.cn

Qing Hai He
Department of Mathematics, Yunnan University, Kunming 650091, P. R. China
E-mail: heqh@ynu.edu.cn

to Cauchy and subsequently studied and improved by many authors. One of most useful improvement is the conjugate gradient method. It is known that Hestenes and Stiefel [9] studied the linear conjugate gradient method for solving linear systems with positive definite coefficient matrices in 1950s, and afterwards in 1960s, the first nonlinear conjugate gradient method, as one of the earliest known techniques for solving large-scale nonlinear optimization problems, was introduced by Fletcher and Reeves in [10]. The classic algorithm for smooth functions by steepest descent method is presented in Section 2.

Note that the main idea of the steepest descent method is, by using the steepest descent unit direction, to produce a gradient sequence that converges to zero with mild assumptions such as Zoutendijk, Wolfe or Goldstein conditions. It cannot be guaranteed that this method converges to a minimizer, but only that it is attracted by stationary points. Naturally, it is interesting and important to establish the corresponding algorithm for nonsmooth functions on Hilbert space along the line by this method. Motivated by it, in this paper, we mainly study the nonsmooth version of steepest descent method by proximal subdifferential of nonsmooth functions defined on Hilbert space, and aim to construct one appropriate algorithm being used to find the stationary point by the nonsmooth steepest descent method.

The paper is organized as follows. In Section 2, we will give some definitions and preliminary through this paper. Our notation is basically standard and conventional in the area of variational analysis. Section 3 is devoted to establishing the algorithm for nonsmooth functions on Hilbert space by nonsmooth steepest descent method. This algorithm is used to find stationary points of those functions with prox-regularity and Lipschitzian continuity assumptions in Hilbert space. The newly-established algorithm is also applied to the search of the minimizer for lower semicontinuous Gâteaux differentiable convex functions on the finite-dimensional space (see Theorem 3.4). The conclusion of this paper is presented in Section 4.

2. Preliminaries

Let H be a *Hilbert space* equipped with the *inner product* $\langle \cdot, \cdot \rangle$ and the corresponding *norm* $\| \cdot \|$. Given a Hilbert space H and a multifunction $F : H \rightrightarrows H$, the symbol

$$\operatorname*{Limsup}_{y \to x} F(x) := \Big\{ \zeta \in H : \exists \text{ sequences } x_n \to x \text{ and } \zeta_n \xrightarrow{w} \zeta \text{ with}$$
$$\zeta_n \in F(x_n) \text{ for all } n \in \mathbb{N} \Big\}$$

signifies the *sequential Painlevé-Kuratowski outer/upper limit* of $F(x)$ as $y \to x$.

For $x \in H$ and $\delta > 0$, let $B(x, \delta)$ denote the *open ball* with center x and radius δ. Given a set $A \subset H$, denote $\overline{A}$ and $\operatorname{int}(A)$ the *norm closure* and the *interior* of A, respectively. If a sequence $\{v_n\}$ converges weakly to v, it is denoted by $v_n \xrightarrow{w} v$.

For any point $z \in H$, the distance between z and S is given by

$$d(z, S) := \inf\{\|z - s\| : s \in S\}.$$

The set of *closest points* to z in S is denoted by

$$P_S(z) := \{s \in S : \|z - s\| = d(z, S)\}.$$

Let $x \in S$. Recall that the *proximal normal cone* of S at x, denoted by $N^p(S, x)$, is defined as:

$$N^p(S, x) := \left\{ \zeta \in H : \zeta = t(z - x) \text{ for some } (z, t) \in H \times (0, +\infty) \text{ with } x \in P_S(z) \right\}.$$

It is known from [11] that $\zeta \in N^p(S, x)$ if and only if there exist two constants $\sigma, \delta \in]0, +\infty[$ such that

$$\langle \zeta, s - x \rangle \leq \sigma \|s - x\|^2, \ \ \forall s \in B(x, \delta) \cap S.$$

The *limiting(Mordukhovich) normal cone* of S at x, denoted by $N(S, x)$, is defined as

$$N(S, x) := \operatorname*{Limsup}_{y \xrightarrow{S} x} N^p(S, y),$$

where $y \xrightarrow{S} x$ means $y \to x$ and $y \in S$. Thus, $\zeta \in N(S, x)$ if and only if there exists a sequence $\{(x_n, \zeta_n)\}$ in $S \times H$ such that $x_n \to x$, $\zeta_n \xrightarrow{w} \zeta$ and $\zeta_n \in N^p(S, x_n)$ for each $n \in \mathbb{N}$.

Let $f : H \to \mathbb{R} \cup \{+\infty\}$ be a lower semicontinuous function. We denote

$$\operatorname{dom}(f) := \{y \in H : f(y) < +\infty\} \ \text{ and } \ \operatorname{epi} f := \{(x, \alpha) \in X \times \mathbb{R} : f(x) \leq \alpha\}$$

the *domain* and the *epigraph* of f, respectively. Let $x \in \operatorname{dom}(f)$. Recall that f is said to be *Gâteaux differentiable* at x, provided there exists $\xi \in H$ such that

$$\lim_{t \downarrow 0} \frac{f(x + th) - f(x)}{t} = \langle \xi, h \rangle, \ \ \forall h \in H.$$

Recall from [11] that a vector $\zeta \in H$ is said to be a *proximal subgradient* of f at x provided that

$$(\zeta, -1) \in N^p(\operatorname{epi} f, (x, f(x))).$$

The set of all such ζ, denoted $\partial_p f(x)$, is said to be the *proximal subdifferential* of f at x. It is known from [11] that one useful and important characterization for proximal subgradient is as follows: $\zeta \in \partial_p f(x)$ if and only if there exist two constants $r, \delta \in]0, +\infty[$ such that

$$f(y) \geq f(x) + \langle \zeta, y - x \rangle - \frac{r}{2} \|y - x\|^2, \ \ \forall \ y \in B(x, \delta).$$

The *limiting(Mordukhovich) subdifferential* of f at x is defined as:

$$\partial f(x) := \{\zeta \in H : (\zeta, -1) \in N(\operatorname{epi} f, (x, f(x)))\}.$$

It is proved in [12,13] that

$$\partial f(x) = \operatorname*{Limsup}_{y \xrightarrow{f} x} \partial_p f(y),$$

where $y \xrightarrow{f} x$ means $y \to x$ and $f(y) \to f(x)$; that is, $\zeta \in \partial f(x)$ if and only if there exist $x_n \xrightarrow{f} x$ and $\zeta_n \xrightarrow{w} \zeta$ such that $\zeta_n \in \partial_p f(x_n)$ for all $n \in \mathbb{N}$. When f is convex, the proximal subdifferential and the limiting subdifferential of f at x coincide and both reduce the subdifferential in the sense of convex analysis, that is

$$\partial_p f(x) = \partial f(x) = \{\zeta \in H : \langle \zeta, y - x \rangle \leq f(y) - f(x), \ \ \forall y \in H\}.$$

Readers are invited to consult books [11,12,14] for more details on these various normal cones and subdifferentials.

In the analysis, we need the notion of *upper semicontinuity*. Let X, Y be Hausdorff topological spaces and let $F : X \rightrightarrows Y$ be a set-valued mapping from X into Y. Recall from [15] that F is *upper semicontinuous* at $x \in X$ iff, for each open set V in Y containing $F(x)$, there is an open neighborhood U of x such that $F(x') \subset V$ for all $x' \in U$.

Let $f : \mathbb{R}^m \to \mathbb{R}$ be a smooth function. We consider the following unconstraint optimization problem:

$$\min_{x \in \mathbb{R}^m} f(x). \tag{1}$$

Note that it is of significance to solve it by using the iteration of line search methods. Each iteration of a line search method computes a search direction u_{n-1} and then decides how far to move along this direction. The iteration is given by

$$x_n := x_{n-1} + \varepsilon_{n-1} u_{n-1},$$

where the positive scalar ε_{n-1} is called the step length. When it comes to problem (1), it is used generally to find the steepest descent unit direction, and then f can be reduced along this direction and finally attain the global minimizer hopefully. The classical algorithm for smooth functions by the steepest descent method is as follows:

Algorithm SDM *Let $f : \mathbb{R}^m \to \mathbb{R}$ be smooth and let x_0 be given.*
Step 1: Let $n := 1$. Go to Step 2.
Step 2: Check vector $\triangledown f(x_{n-1})$. If $\triangledown f(x_{n-1}) = 0$, then terminate; otherwise, set

$$u_{n-1} := -\frac{\triangledown f(x_{n-1})}{\|\triangledown f(x_{n-1})\|} \quad and \quad x_n := x_{n-1} + \varepsilon_{n-1} u_{n-1},$$

where the step length ε_{n-1} is chosen by

$$f(x_{n-1} + \varepsilon_{n-1} u_{n-1}) = \min_{t \geq 0} f(x_{n-1} + t u_{n-1}).$$

Go to Step 3.
Step 3: Set $n := n + 1$ and return to Step 2.

For the smooth function case, the sequence u_n, produced by Algorithm SDM, can converge to a local minimizer only by making additional requirements on the search direction u_n. However, the following known convergent result shows that, for twice differentiable convex functions, the sequence generated by Algorithm SDM would converge to the global minimizer with no assumption on search direction (cf. [7, Theorem 2.3.8]):

Theorem 2.1. *Let $f : \mathbb{R}^m \to \mathbb{R}$ be a strictly convex, twice continuously differentiable function and $x_0 \in \mathbb{R}^m$. Denote $\Omega := \{x \in \mathbb{R}^m : f(x) \leq f(x_0)\}$ the level set. Suppose that there exist constants $m, M \in]0, +\infty[$ such that*

$$m\|y\|^2 \leq \langle y, \triangledown^2 f(z) y \rangle \leq M\|y\|^2 \quad \forall z \in \Omega \text{ and } \forall y \in \mathbb{R}^m. \tag{2}$$

Let $\{x_n\}$ be generated by Algorithm SDM. Then there exists a subsequence $\{x_{n_k}\}$ of $\{x_n\}$ converging to a point $\bar{x}$ that minimizes $f(z)$ over $z \in \mathbb{R}^m$, where $\triangledown^2 f(z)$ denotes the Hessian matrix of f at z.

Note that qualification condition (2) could guarantee the boundedness of the level set Ω (by [7, Theorem B2.8]) and the bounds m, M may be the smallest and largest eigenvalue of the Hessian matrix $\triangledown^2 f(z)$. Our main work in this paper is to establish the corresponding algorithm for nonsmooth functions defined on Hilbert space and use it to find stationary points of those functions satisfying prox-regularity and Lipschitzian continuity (see Theorem 3.1). Then, we apply the newly-established algorithm to the search of minimizer for lower semicontinuous Gâteaux differentiable convex functions defined on the finite-dimensional space so as to extend and improve Theorem 2.1 above.

3. Nonsmooth Steepest Descent Method

In this section, we study nonsmooth steepest descent method for nonsmooth functions on the Hilbert space and establish the corresponding algorithm by proximal subgradients so as to extend the original smooth algorithm by steepest descent method and steepest descent gradient. Then, we use the newly-established algorithm to find stationary points for those nonsmooth functions satisfying prox-regularity and Lipschitzian continuity. We begin with recalling the definition of *prox-regularity* taken from [14, 16].

Definition 3.1. Let $f : H \to \mathbb{R} \cup \{+\infty\}$, and let $\bar{y} \in \text{dom}(f)$. We say that f is *prox-regular* at $\bar{y}$ for $\bar{w} \in \partial_p f(\bar{y})$ iff there exist $\delta > 0$ and $L \geq 0$ such that

$$f(y') \geq f(y) + \langle w, y' - y \rangle - \frac{L}{2}\|y' - y\|^2 \quad \forall y' \in B(\bar{y}, \delta), \tag{3}$$

whenever $y \in B(\bar{y}, \delta)$ with $|f(y) - f(\bar{y})| < \delta$ and $w \in \partial_p f(y) \cap B(\bar{w}, \delta)$. We say that f is prox-regular at $\bar{y}$ iff this holds for every $\bar{w} \in \partial_p f(\bar{y})$.

The following concept introduced in [17] is also relevant here.

Definition 3.2. Let $f : H \to \mathbb{R} \cup \{+\infty\}$, and let $E \subset H$. We say that f is *uniformly prox-regular* on E iff there exist $\delta > 0$ and $L > 0$ such that, for any $\bar{y} \in E$ and $\bar{w} \in \partial_p f(\bar{y})$, one has

$$f(y') \geq f(y) + \langle w, y' - y \rangle - \frac{L}{2}\|y' - y\|^2 \quad \forall y' \in B(\bar{y}, \delta), \tag{4}$$

whenever $y \in B(\bar{y}, \delta)$ with $|f(y) - f(\bar{y})| < \delta$ and $w \in \partial_p f(y) \cap B(\bar{w}, \delta)$. We say that f is *locally uniformly prox-regular* around $y_0 \in H$ iff E can be taken as a neighborhood of y_0, i.e., $E = B(y_0, r)$ for some $r > 0$.

The following proposition, obtained from [17, Proposition 3.3], is one characterization for the local uniform prox-regularity.

Proposition 3.1 *Let $f : H \to \mathbb{R} \cup \{+\infty\}$ and $y_0 \in H$. Then f is locally uniformly prox-regular around y_0 if and only if there are $\delta > 0$ and $L > 0$ such that, for any $y, y' \in B(y_0, \delta)$ and $w \in \partial_p f(y)$, we have the estimate*

$$f(y') \geq f(y) + \langle w, y' - y \rangle - \frac{L}{2}\|y' - y\|^2. \tag{5}$$

Now, we provide one algorithm of nonsmooth steepest descent method for non-smooth functions by proximal subgradients and use it to find stationary points for nonsmooth functions with prox-regularity and Lipschitzian continuity assumptions in the Hilbert space.

Theorem 3.1 *Let $f : H \to \mathbb{R} \cup \{+\infty\}$ be lower semicontinuous and bounded below. Let $x_0 \in \text{dom}(f)$ with $\partial_p f(x_0) \neq \emptyset$. We define the level set as*

$$\Omega := \{x \in H : f(x) \leq f(x_0)\}. \tag{6}$$

Suppose that there exist $\delta, L > 0$ such that f is uniformly prox-regular on the set $B(\Omega, \delta) := \{x \in H : d(x, \Omega) < \delta\}$ and

$$\partial_p f(y) \subset \partial_p f(z) + L\|y - z\|B_H \quad \forall y \in \Omega \text{ with } \|y - z\| < \delta. \tag{7}$$

We use the following Algorithm NSDM to construct a sequence $\{x_n : n = 1, 2, \cdots\}$:

Algorithm NSDM *Let x_0 be given with $\partial_p f(x_0) \neq \emptyset$.*

Step 1: Let $n := 1$. Go to Step 2*.*

Step 2: Choose $v_{n-1} \in \partial_p f(x_{n-1})$ and check vector v_{n-1}. If $v_{n-1} = 0$, then terminate; otherwise, set

$$u_{n-1} := -\frac{v_{n-1}}{\|v_{n-1}\|} \text{ and } x_n := x_{n-1} + \varepsilon_{n-1}u_{n-1},$$

where ε_{n-1} is chosen as follows:

$$f(x_{n-1} + \varepsilon_{n-1}u_{n-1}) = \min_{t \geq 0} f(x_{n-1} + tu_{n-1}). \tag{8}$$

Go to Step 3*.*

Step 3: Set $n := n + 1$ and return to Step 2*.*

Then Algorithm NSDM is valid and every cluster point of $\{x_n : n = 0, 1, \cdots\}$ is the stationary point.

Proof First, we have to show that $\Omega \subset \text{dom}(\partial_p f)$ so that the validness of Algorithm NSDM could be guaranteed. Let $y \in \Omega$. By the Density Theorem (cf. [11, Theorem 3.1]), one has $\text{dom}(\partial_p f) \cap \Omega$ is dense in Ω. Then, there exist $y_k \in \text{dom}(\partial_p f) \cap \Omega$ such that $y_k \to y$. By virtue of (7), when k is sufficiently large, one has

$$\partial_p f(y_k) \subset \partial_p f(y) + L\|y_k - y\|B_H.$$

This implies that $\partial_p f(y) \neq \emptyset$ as $\partial_p f(y_k) \neq \emptyset$ for all k sufficiently large. Hence $\Omega \subset \text{dom}(\partial_p f)$.

Since f is uniformly prox-regular, by Proposition 3.1, without any loss of generality, we can assume that, for any $\bar{y} \in B(\Omega, \delta)$ and $\bar{w} \in \partial_p f(\bar{y})$, one has

$$f(y') \geq f(y) + \langle w, y' - y \rangle - \frac{L}{2}\|y' - y\|^2 \quad \forall y' \in B(\bar{y}, \delta) \tag{9}$$

whenever $w \in \partial_p f(y) \cap B(\bar{w}, \delta)$ and $y \in B(\bar{y}, \delta)$ with $|f(y) - f(\bar{y})| < \delta$, perhaps after reducing the value of $\delta > 0$ and increasing the value of $L > 0$. If there exists n_0 such that $v_{n_0} = 0$, then the conclusion holds. Next, we suppose that $v_n \neq 0$ for all $n \in \mathbb{N}$. Let $\varepsilon \in]0, \delta[$. Noting that $v_{n-1} \in \partial_p f(x_{n-1})$, it follows from (7) that there exist $w_{n-1} \in \partial_p f(x_{n-1} + \varepsilon u_{n-1})$ and $b_{n-1} \in B_H$ such that

$$v_{n-1} = w_{n-1} + L\varepsilon b_{n-1} \tag{10}$$

(thanks to $\varepsilon \in]0, \delta[$). By (9) and (10), one has

$$f(x_{n-1}) \geq f(x_{n-1} + \varepsilon u_{n-1}) + \langle w_{n-1}, -\varepsilon u_{n-1} \rangle - \frac{L}{2}\varepsilon^2$$

$$\geq f(x_{n-1} + \varepsilon u_{n-1}) - \varepsilon \langle v_{n-1}, u_{n-1} \rangle - L\varepsilon^2 - \frac{L}{2}\varepsilon^2.$$

This and (8) imply that

$$\|v_{n-1}\| \leq \frac{f(x_{n-1}) - f(x_n)}{\varepsilon} + \frac{3}{2}L\varepsilon. \tag{11}$$

Since $\{f(x_n)\}$ is monotonic decreasing by (8) and f is bounded below, it follows that $\lim_{n \to \infty} f(x_n)$ exists. Hence, there exists $N \in \mathbb{N}$ such that when $n > N$, one has

$$\frac{f(x_{n-1}) - f(x_n)}{\varepsilon} < \varepsilon.$$

From this and (11), we have

$$\|v_{n-1}\| \leq (1 + \frac{3}{2}L)\varepsilon, \quad n > N,$$

and thus $v_n \to 0$. Let $\bar{x}$ be any cluster point of $\{x_n : n = 0, 1, \cdots\}$. Then there exists a subsequence $\{x_{n_k}\}$ of $\{x_n\}$ such that $x_{n_k} \to \bar{x} \in H$. It follows from the lower semicontinuity of f that

$$f(\bar{x}) \leq \liminf_{k \to \infty} f(x_{n_k}) \leq f(x_0),$$

and hence $\bar{x} \in \Omega$. By applying (9) again, for k large enough, one has

$$f(\bar{x}) \geq f(x_{n_k}) + \langle v_{n_k}, \bar{x} - x_{n_k} \rangle - \frac{L}{2}\|\bar{x} - x_{n_k}\|^2,$$

and consequently $f(\bar{x}) \geq \limsup_{k \to \infty} f(x_{n_k})$. This and lower semicontinuity of f imply that $f(x_{n_k}) \to f(\bar{x})$ as $k \to \infty$. Noting that $v_{n_k} \in \partial_p f(x_{n_k})$ and $v_{n_k} \to 0$, it follows that $0 \in \partial f(\bar{x})$. The proof is complete. $\qquad \square$

Remark 3.1. (i) Condition (7) in Theorem 3.1 refers to the Lipschitzian property of subdifferential $\partial_p f$. Recently Bacak, Borwein, Eberhard and Mordukhovich [18] mainly studied the Lipschitzian property of subdifferentials for prox-regular functions in Hilbert space, and obtained results on this Lipschitzian property of subdifferentials (cf. [18, Lemma 5.2 and Theorem 5.3]). For the smooth functions, subdifferential $\partial_p f$ reduces to Fréchet derivative operator ∇f, and condition (7) means that ∇f is Lipschitzian continuous with modulus $L > 0$. This condition

holds for $C^{1,1}$ functions, para-convex and para-concave functions and twice differentiable functions on finite-dimensional space with bounded Hessian matrices. In particular, the convex function studied in Theorem 2.1 satisfies condition (7).

(ii) Algorithm NSDM provides a search direction u_{n-1} by proximal subgradient for nonsmooth functions defined on the Hilbert space and generates a sequence $\{x_n\}$ having some subsequence converging the stationary point with mild assumptions. When restricted to the smooth case, Algorithm NSDM reduces to Algorithm SDM mentioned in Section 2.

The following example shows that Algorithm NSDM is practicable to find stationary points (or even minimizers) of functions satisfying prox-regularity and Lipschitzian continuity assumptions in Hilbert space.

Example 3.1. Let $H := l^2$ and $y = (\eta(1), \cdots, \eta(k), \cdots) \in H$. Define a function f on H by

$$f(x) := \frac{1}{2} \sum_{i=1}^{\infty} \xi(i)^2 - \sum_{i=1}^{\infty} \xi(i)\eta(i) + 3, \quad \text{for all } x = (\xi(1), \cdots, \xi(k), \cdots) \in H.$$

Clearly f is convex, and thus satisfies the uniform prox-regularity on H. Since $\partial_p f(x) = x - y$ (by computing), then $\partial_p f$ is Lipschitzian continuous on H. Take $x_0 := (0, \cdots, 0, \cdots) \in H$ and let Ω be defined as (6). One can verify that (7) holds with $L := 1$. Using Algorithm NSDM, the sequence $\{x_n : n = 1, 2, \cdots\}$ can be generated and $v_n = x_n - y$. By virtue of the proof of Theorem 3.1, sequence $\{v_n\}$ converges to vector $0 \in H$, and thus y is a cluster point of $\{x_n\}$. It follows from Theorem 3.1 that y is a stationary point, and consequently y minimizes $f(x)$ over $x \in H$ due to the convexity of f.

Let $F : H \rightrightarrows H$ be a set-valued mapping and $x \in H$ with $F(x) \neq \emptyset$. Recall from [15,12] that F is said to be *locally Lipschitzian* at x iff there exist $\delta, \kappa > 0$ such that

$$F(x') \subset F(x'') + \kappa\|x' - x''\|B_H \quad \forall x', x'' \in B(x, \delta).$$

Theorem 3.2 *Let $f : \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ be lower semicontinuous and bounded below. Let $x_0 \in \text{dom}(f)$ with $\partial_p f(x_0) \neq \emptyset$ and the level set Ω be defined as (6). Suppose that Ω is bounded, $\partial_p f$ is locally Lipschitzian on Ω and that f is locally uniformly prox-regular on Ω. Let the sequence $\{x_n : n = 0, 1, \cdots\}$ be generated by Algorithm NSDM. Then there exists a subsequence $\{x_{n_k}\}$ of $\{x_n\}$ such that $\{x_{n_k}\}$ converges to a point $\bar{x} \in \Omega$ with $0 \in \partial f(\bar{x})$.*

Proof First, we need to show that $\Omega \subset \text{dom}(\partial_p f)$ so that the Algorithm NSDM is valid. Let $y \in \Omega$. By the Density Theorem, one has $\text{dom}(\partial_p f) \cap \Omega$ is dense in Ω. Then, there exists $y_k \in \text{dom}(\partial_p f) \cap \Omega$ such that $y_k \to y$. Noting that $\partial_p f$ is locally Lipschitzian at y, it follows that there exist $r > 0$ and $\kappa_y > 0$ such that

$$\partial_p f(z_1) \subset \partial_p f(z_2) + \kappa_y\|z_1 - z_2\|B_H \quad \forall z_1, z_2 \in B(y, r).$$

This means that when k is sufficiently large, one has

$$\partial_p f(y_k) \subset \partial_p f(y) + \kappa_y\|y_k - y\|B_H,$$

and consequently $\partial_p f(y) \neq \emptyset$ as $\partial_p f(y_k) \neq \emptyset$ for all k sufficiently large. Hence $\Omega \subset \mathrm{dom}(\partial_p f)$.

Let $y \in \Omega$. Since $\partial_p f$ is locally Lipschitzian and locally uniformly prox-regular around y, by Proposition 3.1, there exist $\delta_y > 0$ and $L_y > 0$ such that, for any $y', y'' \in B(y, \delta_y)$ and $w \in \partial_p f(y')$, one has

$$f(y'') \geq f(y') + \langle w, y'' - y' \rangle - \frac{L_y}{2} \|y'' - y'\|^2, \tag{12}$$

and

$$\partial_p f(z_1) \subset \partial_p f(z_2) + L_y \|z_1 - z_2\| B_H \quad \forall z_1, z_2 \in B(y, \delta_y), \tag{13}$$

perhaps after reducing the value of $\delta_y > 0$ and increasing the value of $L_y > 0$ if necessary. Then

$$\Omega \subset \bigcup_{y \in \Omega} B(y, \frac{\delta_y}{2}). \tag{14}$$

Note that Ω is closed by the semicontinuity of f and bounded, so Ω is compact. It follows from (14) that there exist $y_1, \cdots, y_l \in \Omega$ such that

$$\Omega \subset \bigcup_{i=1}^{l} B(y_i, \frac{\delta_{y_i}}{2}). \tag{15}$$

Let

$$\delta := \min\left\{\frac{\delta_{y_1}}{2}, \cdots, \frac{\delta_{y_l}}{2}\right\} \quad \text{and} \quad L := \max\{L_{y_1}, \cdots, L_{y_l}\},$$

and let $\varepsilon \in]0, \delta[$. By the Algorithm NSDM, we have $x_{n-1} \in \Omega$, and thus there exists $i \in \{1, \cdots, l\}$ such that $x_{n-1} \in B(y_i, \frac{\delta_{y_i}}{2})$. Noting that $v_{n-1} \in \partial_p f(x_{n-1})$ and $x_{n-1} + \varepsilon u_{n-1} \in B(y_i, \delta_{y_i})$, it follows from (13) that

$$v_{n-1} \in \partial_p f(x_{n-1}) \subset \partial_p f(x_{n-1} + \varepsilon u_{n-1}) + L\varepsilon B_{\mathbb{R}^m}.$$

(thanks to $L \geq L_{y_i}$). Then, there are $w_{n-1} \in \partial_p f(x_{n-1} + \varepsilon u_{n-1})$ and $b_{n-1} \in B_{\mathbb{R}^m}$ such that

$$v_{n-1} = w_{n-1} + L\varepsilon b_{n-1} \tag{16}$$

Since $x_{n-1}, x_{n-1} + \varepsilon u_{n-1} \in B(y_i, \delta_{y_i})$ and $w_{n-1} \in \partial_p f(x_{n-1} + \varepsilon u_{n-1})$, by (12) and (16), one has

$$f(x_{n-1}) \geq f(x_{n-1} + \varepsilon u_{n-1}) + \langle w_{n-1}, -\varepsilon u_{n-1} \rangle - \frac{L_{y_i}}{2} \varepsilon^2$$

$$\geq f(x_{n-1} + \varepsilon u_{n-1}) - \varepsilon \langle v_{n-1}, u_{n-1} \rangle - L\varepsilon^2 - \frac{L}{2}\varepsilon^2.$$

This and (8) imply that

$$\|v_{n-1}\| \leq \frac{f(x_{n-1}) - f(x_n)}{\varepsilon} + \frac{3}{2} L\varepsilon.$$

Since $\{f(x_n)\}$ is monotonic decreasing and bounded below, one has $\lim_{n \to \infty} f(x_n)$ exists, and consequently there exists $N \in \mathbb{N}$ such that, whenever $n > N$, one has

$$\frac{f(x_{n-1}) - f(x_n)}{\varepsilon} < \varepsilon.$$

Hence

$$\|v_{n-1}\| < (1 + \frac{3}{2}L)\varepsilon, \quad \forall n > N,$$

and consequently $v_n \to 0$. Noting that $\{x_n\} \subset \Omega$ and Ω is compact, there exists a subsequence $\{x_{n_k}\}$ of $\{x_n\}$ converging to a point $\bar{x} \in \Omega$. By using (12), when k is sufficiently large, we have

$$f(\bar{x}) \geq f(x_{n_k}) + \langle v_{n_k}, \bar{x} - x_{n_k} \rangle - \frac{L}{2}\|\bar{x} - x_{n_k}\|^2.$$

Thus $f(\bar{x}) \geq \limsup_{k \to \infty} f(x_{n_k})$, and consequently $f(x_{n_k}) \to f(\bar{x})$ as f is lower semicontinuous at $\bar{x}$. Hence $0 \in \partial f(\bar{x})$. The proof is complete.   □

Next, we use the established Algorithm NSDM to find the minimizer of lower semicontinuous and convex functions defined on finite-dimensional space. Since convexity implies local uniform prox-regularity automatically, the following theorem is immediate from Theorem 3.2.

Theorem 3.3 *Let $f : \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ be a lower semicontinuous, convex and bounded below function, $x_0 \in \text{dom}(f)$ with $\partial f(x_0) \neq \emptyset$ and the level set Ω be defined as (6). Suppose that Ω is bounded and ∂f is locally Lipschitzian on Ω. Let the sequence $\{x_n : n = 0, 1, \cdots\}$ be generated by Algorithm NSDM. Then there exists a subsequence $\{x_{n_k}\}$ of $\{x_n\}$ such that $\{x_{n_k}\}$ converges to a point $\bar{x} \in \Omega$ with $0 \in \partial f(\bar{x})$, and consequently $f(\bar{x}) = \min_{z \in \mathbb{R}^m} f(z)$.*

Now, we provide one convergent theorem for lower semicontinuous, convex and Gâteaux differentiable functions defined on the finite-dimensional space. This theorem shows that the sequence generated by Algorithm NSDM converges to the minimizer even without Lipschitzian assumption of subdifferential.

Theorem 3.4 *Let $f : \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ be lower semicontinuous, convex, Gâteaux differentiable and bounded below function, $x_0 \in \text{dom}(f)$ and the level set Ω be defined as (6). Suppose that Ω is bounded and $\Omega \subset \text{int}(\text{dom}(f))$. Let $\{x_n\}$ be generated by the Algorithm NSDM. Then there exists a subsequence $\{x_{n_k}\}$ of $\{x_n\}$ such that $\{x_{n_k}\}$ converges to a point $\bar{x} \in \Omega$ that minimizes $f(z)$ over $z \in \mathbb{R}^m$.*

Proof First, the validness of the Algorithm NSDM follows from the Gâteaux differentiability of f.

Using [19, Proposition 3.3], we have that f is continuous on $\text{int}(\text{dom}(f))$, and consequently the subdifferential mapping ∂f is norm-to-norm upper semicontinuous on $\text{int}(\text{dom}(f))$ by [19, Proposition 2.5]. Let $\varepsilon \in]0, +\infty[$. Noting that $\mathbb{R}^m$ is finite-dimensional and $\Omega \subset \text{int}(\text{dom}(f))$, by the upper semicontinuity, for each $y \in \Omega$ there exists $\delta_y > 0$ such that

$$\partial f(y') \subset \partial f(y) + \frac{\varepsilon}{3}B_{\mathbb{R}^m} \quad \forall y' \in B(y, \delta_y). \tag{17}$$

Note that Ω is bounded, and consequently Ω is compact. This implies that there exist $y_1, \cdots, y_l \in \Omega$ such that

$$\Omega \subset \bigcup_{i=1}^{l} B(y_i, \frac{\delta_{y_i}}{2}). \tag{18}$$

Let $\delta := \min\left\{\frac{\delta_{y_1}}{2}, \cdots, \frac{\delta_{y_l}}{2}\right\}$ and $\alpha \in]0, \delta[$. By the Algorithm NSDM, we have $x_{n-1} \in \Omega$, and there exists $i \in \{1, \cdots, l\}$ such that $x_{n-1} \in B(y_i, \frac{\delta_{y_i}}{2})$. Since f is Gâteaux differentiable at y_i, then the subdifferential $\partial f(y_i)$ is the singleton. Noting that $v_{n-1} \in \partial f(x_{n-1})$ and $x_{n-1} + \alpha u_{n-1} \in B(y_i, \delta_{y_i})$, it follows from (18) that there exist $w_{n-1} \in \partial f(x_{n-1} + \alpha u_{n-1})$ and $a_{n-1}, b_{n-1} \in B_{\mathbb{R}^m}$ such that

$$v_{n-1} = w_{n-1} + \frac{\varepsilon}{3}(b_{n-1} + a_{n-1}).$$

This and convex inequality in convex analysis imply that

$$f(x_{n-1}) - f(x_{n-1} + \alpha u_{n-1}) \geq \langle w_{n-1}, -\alpha u_{n-1}\rangle \geq \langle v_{n-1}, -\alpha u_{n-1}\rangle - \frac{2}{3}\alpha\varepsilon$$

and, by (8), one has

$$\|v_{n-1}\| \leq \frac{f(x_{n-1}) - f(x_n)}{\alpha} + \frac{2}{3}\varepsilon. \qquad (19)$$

Since $\{f(x_n)\}$ is monotonic decreasing and bounded below, then $\lim_{n \to \infty} f(x_n)$ exists and there exists $N \in \mathbb{N}$ such that, whenever $n > N$, we have

$$\frac{f(x_{n-1}) - f(x_n)}{\alpha} < \frac{\varepsilon}{3}.$$

From this and (19), one has

$$\|v_{n-1}\| < \varepsilon, \quad \forall n > N.$$

This implies that $v_n \to 0$. Since $\{x_n\} \subset \Omega$ and Ω is compact, there exists a subsequence $\{x_{n_k}\}$ of $\{x_n\}$ such that $x_{n_k} \to \bar{x} \in \Omega$. Note that $f(x_{n_k}) \to f(\bar{x})$ follows from the continuity of f at $\bar{x}$, so $0 \in \partial f(\bar{x})$ and this is equivalent to that $f(\bar{x}) = \min_{z \in \mathbb{R}^m} f(z)$ (thanks to the convexity of f). The proof is complete. $\qquad \square$

Remark 3.2. In Theorem 3.4, Algorithm NSDM established has been used to construct a sequence having some subsequence converging to the minimizer of lower semicontinuous convex and Gâteaux differentiable functions that are more general than strictly convex and twice continuously differentiable functions. Hence this convergent result improves Theorem 2.1 in the sense of relaxing the twice continuously differentiable assumption.

4. Concluding Remarks

This paper is devoted to presenting an algorithm for finding stationary points of nonsmooth functions based on nonsmooth steepest descent method and proximal subgradients in Hilbert space. The algorithm established is used to produce one sequence converging to zero vector so as to obtain the stationary point of lower semicontinuous functions satisfying prox-regularity and Lipschitzian continuity on Hilbert space. This algorithm is also applied to the search of minimizers for convex functions on finite-dimension space. The obtained convergent theorem extends the existing result on twice continuously differentiable convex functions on finite-dimensional space.

References

1. Attouch, H., Cominetti, R.: A dynamical approach to convex minimization coupling approximation with the steepest descent method, J. Differ. Equations., 128, 519-540 (1996).
2. Bazaraa, M. S., Shetty, C. M.: Nonlinear programming: theory and algorithm, New York: Wiley (1979).
3. Chuong, T. D., Yao, J.-C.: Steepest descent methods for critical points in vector optimization problems, Optimization, 91, 1811-1829 (2012).
4. Grana Drummond, L. M., Svaiter, B. F.: A steepest descent method for vector optimization, J. Comput. Appl. Math., 175, 395-414 (2005).
5. Fliege, J., Svaiter, B. F.: Steepest descent methods for multicriteria optimization, Math. Methods Oper. Res., 51, 479-494 (2000).
6. Nocedal, J., Wright, S. J.: Numerical optimization, Springer, New York (2006).
7. Polak, E.: Computational methods in optimization, Academic Press, New York (1971).
8. Smyrlis, G., Zisis, V.: Local convergence of the steepest descent method in Hilbert spaces, J. Math. Anal. Appl., 300, 436-453 (2004).
9. Hestenes, M. R., Stiefel, E.: Methods of conjugate gradients for solving linear systems, Journal of Research of the National Bureau of standards, 49, 409-436 (1952).
10. Fletcher, R., Reeves, C. M.: Functions minimization by conjugate gradients, Comput. J., 7, 149-154 (1964).
11. Clarke, F. H., Ledyaev, Yu. S., Stern, R. J., Wolenski, P. R.: Nonsmooth analysis and control theory, Springer (1998).
12. Mordukhovich, B. S.: Variational Analysis and Generalized differentiation I/II, Springer-verlag, Berlin (2006).
13. Mordukhovich, B. S., Shao, Y.: Nonsmooth sequential analysis in Asplund spaces, Trans. Amer. Math. Soc., 348, 1235-1280 (1996).
14. Rockafellar, R. T., Wets, R. J-B.: Variational Analysis, Springer-Verlag, New York (1998).
15. Aubin, J. P., Frankowska, H.: Set-valued Analysis, Birkhäuser, Boston (1990).
16. Poliquin, R., Rockafellar, R.T.: Prox-regular functions in variational analysis, Trans. Amer. Math. Soc., 348, no. 5, 1805-1838 (1996).
17. Bernard, F., Thilbaut, L.: Uniform prox-regularity of functions and epigraphs in Hilbert spaces, Nonlinear Anal., 60, 187-207 (2005).
18. Bacak, M., Borwein, J. M., Eberhard, A., Mordukhovich, B. S.: Infimal convolutions and Lipschitzian properties of subdifferentials for prox-regular functions in Hilbert spaces, J. Convex Anal., 17, 737-763 (2010).
19. Phelps, R. R.: Convex functions, Monotone operators and Differentiability, Lecture Notes in Math. 1364, Springer, New York (1989).