J. Nonlinear Var. Anal. 1 (2017), No. 2, pp. 265-279
Available online at http://jnva.biemdas.com

CONCEPTS ON GENERALIZED ε-SUBDIFFERENTIALS FOR MINIMIZING LOCALLY LIPSCHITZ CONTINUOUS FUNCTIONS

MARTIN KNOSSALLA

Department of Mathematics, Friedrich-Alexander University Erlangen-Nürnberg, Cauerstr. 11, 91058 Erlangen, Germany

Abstract. In this paper, a general concept around converging descent methods for unconstrained nonsmooth optimization problems is introduced. This concept is of constructive nature. Based on the subdifferential according to Clarke or Mordukhovich, respectively, a general approach to ε-subdifferentials, a kind of continuous outer approximations, is given in an axiomatic way. This leads to a constructive way to create a converging descent method for a given locally Lipschitz continuous objective function. Hence, from a theoretical point of view, convergence is established through the construction of ε-subdifferentials. This is in contrast to other approaches which are usually based on the assumption of semismoothness of the objective function. The proposed framework generalizes some other approximation concepts for various types of subdifferentials. Furthermore, based on these continuous outer approximations an algorithm is presented and its global convergence to stationary points is proved.

Keywords. Nonsmooth optimization; Variational analysis; Set-valued analysis; Generalized subdifferential; Convergence analysis.

2010 Mathematics Subject Classification. 49J52, 49J53, 90C26.

1. INTRODUCTION

In this paper, we are focussing on optimization problems of the type

$$\min f(x) \quad \text{subject to} \quad x \in \mathbb{R}^n, \tag{1.1}$$

where the objective function $f : \mathbb{R}^n \to \mathbb{R}$ is supposed to be locally Lipschitz continuous. Note that no semismoothness or convexity assumptions for f are made.

The purpose of this paper is to give a constructive concept to achieve converging descent methods for problems of the type (1.1). In general, information about the behavior of the objective function f is needed to compute descent directions, such information can be obtained by so-called subdifferentials ∂f. The reader can find more information on the theory of subdifferentials in [5, 6, 16, 25, 27]. Descent methods, which depend on a subdifferential for a convex or locally Lipschitz continuous function, are described in detail in e.g. [4, 13, 14, 15, 18, 21]. However, to achieve converging methods information about the local behavior of the graph of f is needed. This was observed in e.g. [7, 14, 28], where examples were considered showing the non-convergence working with the subdifferential of convex

E-mail address: knossalla@math.fau.de.

Received July 3, 2017; Accepted July 14, 2017.

analysis. The background of this phenomenon is the lack of continuity of the set-valued mapping ∂f. For this reason, a kind of outer approximation is introduced to remedy this lack, by the so-called ε-subdifferential for the subdifferential of convex analysis. Concepts around outer approximations to the Clarke subdifferential are described in e.g. [8, 10, 11]. Particulary, A.M. Bagirov took and elaborated this idea by Hausdorff continuous approximations to the Clarke subdifferential, see [2, 3]. But to guarantee convergence of the corresponding algorithm the objective function f has to be semismooth, that is more restrictive than locally Lipschitz continuity, see [22] for detailed information about semismoothness. Methods based on semismooth functions are studied in e.g. [12, 17, 20, 23]. But in applications there is often no exact information about the whole Clarke subdifferential, like e.g. for marginal functions, see e.g. [9]. In this case, the semismoothness of the cost functions cannot be proved or is violated.

Here, the concept on outer approximations to the Clarke and Mordukhovich subdifferential is introduced in an axiomatic way, see Definition 3.1 below. By this means, it is possible to construct a suitable continuous subdifferential. Due to the simple fact that in practice the optimization problems are explicitly given, these theoretical constructions are realizable in practice. Moreover, this is also realizable in the case if, for instance, the Clarke subdifferential is not known completely, see Remark 3.1 below. On the whole, this leads to a constructive approach around converging descent methods. Hence, from a theoretical point of view, convergence is established through the construction of generalized ε-subdifferentials. This is in contrast to other approaches which are usually based on the assumption of semismoothness.

This paper is organized as follows. Section 2 provides preliminary material from set-valued analysis and nonsmooth analysis needed in the subsequences. In section 3, generalized ε-subdifferentials are introduced in an axiomatic way. Based on these subdifferentials directional derivatives are defined and relevant properties are studied. Section 4 describes a descent method corresponding to a generalized ε-subdifferential. The global convergence to stationary points of this method is proved for locally Lipschitz continuous functions. Finally, section 5 concludes this paper.

2. Preliminaries

In this section we recall some terms and facts for set-valued mappings between finite-dimensional normed vector spaces, in which the Euclidean norm $\|x\| := \sqrt{x^T x}$ is always used in this paper. Also, we recall the definitions of the subdifferentials according to Clarke and Mordukhovich, and give some properties on these subdifferentials needed in the subsequences.

The notation $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is used for set-valued mappings. Throughout the paper we use standard notations of set-valued analysis, e.g. $\operatorname{dom}(S) := \{x \in \mathbb{R}^n : S(x) \neq \emptyset\}$ for the domain of S or $\operatorname{gph}(S) := \{(x,y) \in \mathbb{R}^n \times \mathbb{R}^m : y \in S(x)\}$ for the graph of S. Moreover, recall that the closed ball with center $x \in \mathbb{R}^n$ and radius $\delta > 0$ is denoted by $B_\delta(x)$. For a given subset A of $\mathbb{R}^n$, we denote the boundary of A by $\operatorname{bd} A$, the convex hull of A by $\operatorname{conv} A$ and the closure of A by $\operatorname{cl} A$. The Minkowski sum of subsets A, B of $\mathbb{R}^n$ is defined as always by $A + B := \{a + b : a \in A,\, b \in B\}$.

The following definition is extracted from [27]. It describes the continuity concept of a set-valued mapping in the sense of Painlevé-Kuratowski, cf. also Fig. **1** below.

Definition 2.1. Let $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ and $\bar{x} \in \mathrm{dom}(S)$.

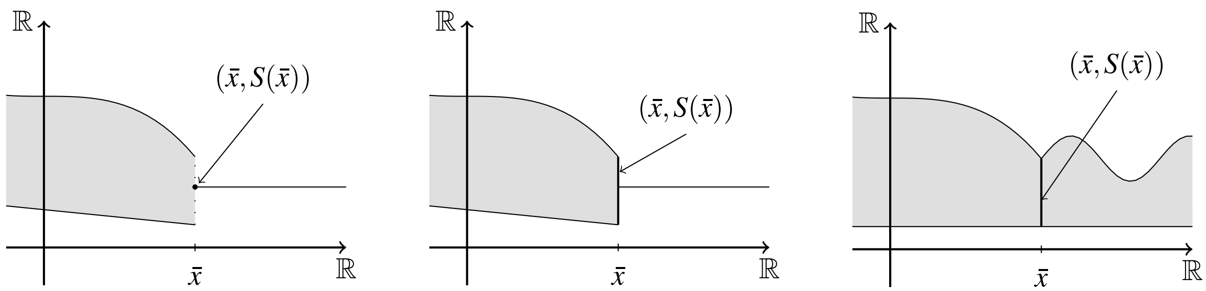
(i) The mapping S is called *outer semicontinuous (osc) at* $\bar{x}$, if

$$\{g \in \mathbb{R}^m : \exists x_k \to \bar{x}\, \exists g_k \to g \text{ with } g_k \in S(x_k)\, \forall k\} \subseteq S(\bar{x}).$$

(ii) The mapping S is called *inner semicontinuous (isc) at* $\bar{x}$, if

$$\{g \in \mathbb{R}^m : \forall x_k \to \bar{x}\, \exists g_k \to g \text{ with } g_k \in S(x_k)\, \forall k\} \supseteq S(\bar{x}).$$

(iii) The mapping S is called *Painlevé-Kuratowski continuous (PK-continuous) at* $\bar{x}$, if S is osc and isc at $\bar{x}$.



S is isc at $\bar{x}$, fails to be osc — S is osc at $\bar{x}$, fails to be isc — S is PK-continuous at $\bar{x}$

FIG. 1 Illustrations of inner and outer semicontinuity of $S : \mathbb{R} \rightrightarrows \mathbb{R}$, respectively

Definition 2.2. A set-valued mapping $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is called *locally bounded at* $\bar{x} \in \mathbb{R}^n$, if there exists a neighborhood $N_{\bar{x}}$ of $\bar{x}$ such that the set $\{y \in \mathbb{R}^m : \exists x \in N_{\bar{x}} \text{ s.t. } y \in S(x)\}$ is bounded.

In this paper we are interested in the subdifferential concepts of Clarke and Mordukhovich, see Remark 3.4 below (the reader finds more information on these subdifferentials in [6, 25]).

Definition 2.3. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function.

(i) The *Clarke subdifferential* $\partial_C f : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is defined by

$$\partial_C f(x) := \mathrm{conv}\left\{\lim_{k\to\infty} \nabla f(x_k) : \exists x_k \to x \text{ s.t. } \nabla f(x_k) \text{ exists } \forall k\right\}.$$

(ii) The *limiting subdifferential* $\partial_L f : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is defined by

$$\partial_L f(x) := \left\{g \in \mathbb{R}^n : \exists x_k \xrightarrow{f} x\, \exists g_k \to g \text{ s.t. } \liminf_{\substack{y\to x_k\\ y\neq x_k}} \frac{f(y) - f(x_k) - g_k^T(y - x_k)}{\|y - x_k\|} \geq 0\, \forall k\right\},$$

where $x_k \xrightarrow{f} x$ means $x_k \to x$ with $f(x_k) \to f(x)$.

(iii) The *symmetric subdifferential* $\partial_S f : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is defined by

$$\partial_S f(x) := -\partial_L(-f)(x) \cup \partial_L f(x).$$

Throughout this paper, we will represent each of the above defined subdifferentials by $\partial_\star f$ with $\star \in \{C, L, S\}$. The following results collect several well-known properties of $\partial_\star f$ with purpose to design a descent method.

Theorem 2.1. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be locally Lipschitz continuous and $\star \in \{C, L, S\}$.*

(a) *If $\bar{x} \in \mathbb{R}^n$ is a local minimizer of f, then $\bar{x}$ is a $\star$-stationary point of f, i.e. $0 \in \partial_\star f(\bar{x})$.*

(b) *If $\max\{g^T \bar{d} : g \in \partial_\star f(\bar{x})\} < 0$ at $(\bar{x}, \bar{d}) \in \mathbb{R}^n \times \mathbb{R}^n$, then $\bar{d}$ is a descent direction of f at $\bar{x}$, i.e. there exists $T > 0$ such that $f(\bar{x} + t\bar{d}) < f(\bar{x})$ for all $t \in (0, T)$.*

Proof. (a). In the case of $\star = C$ see [21, I. Theorem 5.1.1] and of $\star \in \{L, S\}$ see [24, Theorem 5.6].
(b). In the case of $\star = C$ see [21, I. Theorem 5.2.5]. Now, let $\star \in \{L, S\}$. Then, by [25, Theorem 3.57] we have immediately

$$\operatorname{conv} \partial_\star f(x) = \partial_C f(x) \qquad \forall x \in \mathbb{R}^n.$$

Thus, for $(\bar{x}, \bar{d})$ with $\max\{g^T \bar{d} : g \in \partial_\star f(\bar{x})\} < 0$, we obtain

$$\max\{g^T \bar{d} : g \in \partial_C f(\bar{x})\} = \max\{g^T \bar{d} : g \in \operatorname{conv} \partial_\star f(\bar{x})\} = \max\{g^T \bar{d} : g \in \partial_\star f(\bar{x})\} < 0,$$

in which the last equality follows by [26, Theorem 32.2]. Hence, the proof is similar to that with the Clarke subdifferential. □

Also, the mapping $\partial_\star f : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is locally bounded and outer semicontinuous with $\operatorname{dom}(\partial_\star f) = \mathbb{R}^n$. But as described in Section 1, in general, the mapping $\partial_\star f$ is not inner semicontinuous, cf. also Remark 3.1 below, which is the main reason for non-convergence of descent methods for nonsmooth optimization problems. However, we use for our algorithmic concept, as described in Section 4 below, a property of $\partial_\star f$ which allows to approximate the function values of f:

Theorem 2.2 (Mean-Value Theorem)**.** *Let $x, y \in \mathbb{R}^n$ with $x \neq y$, $f : \mathbb{R}^n \to \mathbb{R}$ Lipschitz continuous on an open set containing the line segment $[x, y]$ and $\star \in \{C, S\}$. Then there exists $\theta \in (0, 1)$ such that*

$$f(y) - f(x) \in \left\{ g^T(y - x) : g \in \partial_\star f(x + \theta(y - x)) \right\}.$$

Proof. In the case of $\star = C$ see [21, I. Theorem 3.2.7] and of $\star = S$ see [25, Theorem 3.47]. □

In general, the Mean-Value Theorem is not satisfied for the limiting subdifferential. A counterexample is already given by the simple function $\mathbb{R} \ni x \mapsto -|x|$. For this reason, we are not interested in the limiting subdifferential in the following sections.

3. Generalized ε-Subdifferentials and Corresponding Directional-Derivatives

The construction of generalized ε-subdifferentials is based on the Clarke subdifferential or the symmetric subdifferential of a real-valued and locally Lipschitz continuous function. Also, this concept is meaningful for other subdifferentials satisfying certain properties, see Remark 3.4 below.

We start with the definition of a generalized ε-subdifferential, which is given in an axiomatic way:

Definition 3.1. Let $\star \in \{C, S\}, \varepsilon > 0$ and let $\partial_\star f : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ be a $\star$-subdifferential of a locally Lipschitz continuous function $f : \mathbb{R}^n \to \mathbb{R}$. A mapping $G^f_{\star,\varepsilon} : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is called a $\star(\varepsilon)$*-subdifferential of f (or a generalized ε-subdifferential of f)*, if the following conditions hold:

(OS$_1$) $\partial_\star f(x) \subseteq G^f_{\star,\varepsilon}(x)$ for all $x \in \mathbb{R}^n$,

(OS$_2$) $G^f_{\star,\varepsilon}$ is convex-valued and locally bounded on $\mathbb{R}^n$,

(OS$_3$) $G^f_{\star,\varepsilon}$ is PK-continuous on $\mathbb{R}^n$,

(OS$_4$) $G^f_{\star,\eta}(x) \subsetneqq G^f_{\star,\varepsilon}(x)$ for all $x \in \mathbb{R}^n$, if $0 < \eta < \varepsilon$,

(OS$_5$) $\forall x \in \mathbb{R}^n, \eta > 0 \, \exists \tilde{\varepsilon} \in (0,\varepsilon), \delta > 0 : \bigcup_{y \in B_\delta(x)} G^f_{\star,\tilde{\varepsilon}}(y) \subseteq \partial_\star f(x) + B_\eta(0)$.

Condition (OS$_4$) and (OS$_5$) are needed to generate $\star$-stationary points via the algorithm introduced in Section 4. Whereas in this section, the results depend only on the conditions (OS$_1$), (OS$_2$) and (OS$_3$).

Remark 3.1. The idea behind the axiomatic approach of generalized ε-subdifferentials can be construed as follows: Usually the challenge to achieve convergence in nonsmooth optimization is to have a semismooth objective function, respectively to investigate if the considered problem has this property, which can be impossible, as explained in the introduction. Now, the challenge to achieve convergence is to construct a suitable outer approximation of a given $\star$-subdifferential. Hence, the difficulty is transformed from investigation to construction. This approach is also applicable, when just a superset of a $\star$-subdifferential is given. The lack of information of a subdifferential, as described in the introduction, is to be closed by (OS$_1$) and (OS$_3$). Note that an outer semicontinuous mapping is also closed-valued. Therefore, condition (OS$_2$) and (OS$_3$) imply that the set $\operatorname{argmin}\{\|g\| : g \in G^f_{\star,\varepsilon}(x)\}$ is singleton for every $x \in \mathbb{R}^n$, cf. also Theorem 3.1 below.

Typically, a descent direction of a locally Lipschitz continuous function f is computed via the projection of the origin onto $\partial_\star f(x)$. We keep this approach for computing descent directions in this paper, but now by the projection onto the set $G^f_{\star,\varepsilon}(x)$. In this case the projection is continuous, where x is the subject to perturbation:

Theorem 3.1. *Let $\star \in \{C,S\}, \varepsilon > 0$, and let $G^f_{\star,\varepsilon} : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ be a $\star(\varepsilon)$-subdifferential of f. Then the function $\pi_\varepsilon : \mathbb{R}^n \to \mathbb{R}^n$ defined by*

$$\pi_\varepsilon(x) := \pi^f_{\star,\varepsilon}(x) := \operatorname{argmin}\{\|g\| : g \in G^f_{\star,\varepsilon}(x)\} \tag{3.1}$$

is continuous on $\mathbb{R}^n$.

Proof. Since the mapping $G^f_{\star,\varepsilon}$ is closed-valued, convex-valued and PK-continuous, the assertion follows immediately by [27, Proposition 4.9]. □

To become more familiar with the concept of Definition 3.1, we like to indicate how to derive calculus rules for generalized ε-subdifferentials at the beginning and on the general case. The basic idea is to transfer existing calculus rules from a subdifferential. We demonstrate this in the following remark at hand of the sum rule with the Clarke subdifferential and for *Clarke regular functions*, that is when the Clarke directional-derivative coincides with the usual directional-derivative.

Remark 3.2. For given locally Lipschitz continuous Clarke regular functions $f,g:\mathbb{R}^n\to\mathbb{R}$ and scalars $\lambda,\mu>0$ we obtain, see [21, I. Theorem 3.2.6],

$$\partial_C(\lambda f+\mu g)(x)=\lambda\partial_C f(x)+\mu\partial_C g(x)\qquad \forall x\in\mathbb{R}^n. \tag{3.2}$$

For a $C(\varepsilon)$-subdifferential of f and g, respectively, the mapping

$$\mathbb{R}^n\rightrightarrows\mathbb{R}^n,\qquad x\longmapsto \lambda G^f_{C,\varepsilon}(x)+\mu G^g_{C,\varepsilon}(x)$$

is a $C(\varepsilon)$-subdifferential of the function $\lambda f+\mu g$ or more precise of the pair $(\lambda f+\mu g,\partial_C(\lambda f+\mu g))$, see the sketch of proof below. Thereby, we can use the following calculus rule for generalized ε-subdifferentials:

$$G^{\lambda f+\mu g}_{C,\varepsilon}(x)=\lambda G^f_{C,\varepsilon}(x)+\mu G^g_{C,\varepsilon}(x)\qquad \forall x\in\mathbb{R}^n \tag{3.3}$$

Note that generalized ε-subdifferentials are not necessarily uniquely determinated. But there is always the possibility to design new generalized ε-subdifferentials on the basis of already designed ones, as introduced in (3.3).

Sketch of proof of formula (3.3): The proof is straightforward for the conditions (OS_1) till (OS_4). To prove condition (OS_5) choose arbitrary $x\in\mathbb{R}^n$ and $\eta>0$. Then there exist $\varepsilon_f,\varepsilon_g\in(0,\varepsilon)$ and $\delta_f,\delta_g>0$ such that

$$\bigcup_{y\in B_{\delta_f}(x)} G^f_{C,\varepsilon_f}(y)\subseteq\partial_C f(x)+B_{\eta/2\lambda}(0)\quad\text{and}\quad \bigcup_{y\in B_{\delta_g}(x)} G^g_{C,\varepsilon_g}(y)\subseteq\partial_C g(x)+B_{\eta/2\mu}(0). \tag{3.4}$$

Define $\tilde\varepsilon:=\min\{\varepsilon_f,\varepsilon_g\}$ and $\delta:=\min\{\delta_f,\delta_g\}$. It follows

$$\begin{aligned}
\bigcup_{y\in B_\delta(x)}\left(\lambda G^f_{C,\tilde\varepsilon}(x)+\mu G^g_{C,\tilde\varepsilon}(x)\right) &\overset{(\mathrm{OS}_4)}{\subseteq} \bigcup_{y\in B_\delta(x)}\left(\lambda G^f_{C,\varepsilon_f}(y)+\mu G^g_{C,\varepsilon_g}(y)\right)\\
&\subseteq \bigcup_{y\in B_\delta(x)}\lambda G^f_{C,\varepsilon_f}(y)+\bigcup_{y\in B_\delta(x)}\mu G^g_{C,\varepsilon_g}(y)\\
&\subseteq \bigcup_{y\in B_{\delta_f}(x)}\lambda G^f_{C,\varepsilon_f}(y)+\bigcup_{y\in B_{\delta_g}(x)}\mu G^g_{C,\varepsilon_g}(y)\\
&\overset{(3.4)}{\subseteq} \lambda\partial_C f(x)+\mu\partial_C g(x)+\lambda B_{\eta/2\lambda}(0)+\mu B_{\eta/2\mu}(0)\\
&\overset{(3.2)}{=} \partial_C(\lambda f+\mu g)(x)+B_\eta(0),
\end{aligned}$$

and $\tilde\varepsilon\in(0,\varepsilon)$. □

We discuss some candidates for $\star(\varepsilon)$-subdifferential of a locally Lipschitz continuous function f at the end of this section. In advance, we study the properties of the support function of the set $G^f_{\star,\varepsilon}(x)$, hereby, the perturbation behavior of x is of special interest.

Every non-empty closed convex set in $\mathbb{R}^n$ is uniquely determined by the support function of this set, and vice versa. This well-known concept in convex analysis can be used to introduce a directional-derivative. Next, we adapt and generalize familiar techniques for establishing the calculus of a directional-derivative. It is necessary to check, however, that these techniques work also in our general framework of generalized ε-subdifferentials.

Definition 3.2. Let $\star \in \{C, S\}, \varepsilon > 0$, and let $G^f_{\star,\varepsilon} : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ be a $\star(\varepsilon)$-subdifferential of f. The $\star(\varepsilon)$-*directional-derivative of f at $x \in \mathbb{R}^n$ in direction $d \in \mathbb{R}^n$* is defined by

$$f^G_{\star,\varepsilon}(x,d) := \max\{g^T d : g \in G^f_{\star,\varepsilon}(x)\}. \tag{3.5}$$

Due to the concept of $G^f_{\star,\varepsilon}$ the $\star(\varepsilon)$-directional-derivative is only interesting for locally Lipschitz continuous functions $f : \mathbb{R}^n \to \mathbb{R}$. For this reason we will generally assume that f is locally Lipschitz continuous.

Theorem 3.2. *Let $\star \in \{C, S\}, \varepsilon > 0$, and $\bar{x} \in \mathbb{R}^n$.*

(a) *The value $f^G_{\star,\varepsilon}(\bar{x}, d)$ is finite for each direction $d \in \mathbb{R}^n$.*

(b) *The function $f^G_{\star,\varepsilon}(\bar{x}, \cdot) : \mathbb{R}^n \to \mathbb{R}$ is lower semicontinuous and sublinear.*

Proof. The support function $\sigma_{G^f_{\star,\varepsilon}(\bar{x})}(\cdot) : \mathbb{R}^n \to \mathbb{R}$ of the non-empty compact convex set $G^f_{\star,\varepsilon}(\bar{x})$ is given by

$$\sigma_{G^f_{\star,\varepsilon}(\bar{x})}(d) = \sup\{g^T d : g \in G^f_{\star,\varepsilon}(\bar{x})\} = f^G_{\star,\varepsilon}(\bar{x}, d).$$

Thereby, the assertions follow from [27, Theorem 8.24]. □

It follows that the $\star(\varepsilon)$-directional-derivative of f can be considered as a function from $\mathbb{R}^n \times \mathbb{R}^n$ onto $\mathbb{R}$. The next result gives some properties on this function:

Theorem 3.3. *Let $f^G_{\star,\varepsilon} : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ be a $\star(\varepsilon)$-directional-derivative of $f : \mathbb{R}^n \to \mathbb{R}$.*

(a) *The function $f^G_{\star,\varepsilon}$ is continuous on $\mathbb{R}^n \times \mathbb{R}^n$.*

(b) *Suppose that $0 \notin G^f_{\star,\varepsilon}(\bar{x})$ for $\bar{x} \in \mathbb{R}^n$. Then there exists a neighborhood $N_{\bar{x}}$ of $\bar{x}$ such that the function $d_\varepsilon : N_{\bar{x}} \to \mathbb{R}^n$ given by*

$$d_\varepsilon(x) := d^f_{\star,\varepsilon}(x) := -\frac{\pi_\varepsilon(x)}{\|\pi_\varepsilon(x)\|} \tag{3.6}$$

is well-defined and continuous on $N_{\bar{x}}$, where π_ε is defined as in (3.1). Moreover, we have

$$f^G_{\star,\varepsilon}(x, d_\varepsilon(x)) = \pi_\varepsilon(x)^T d_\varepsilon(x) = -\|\pi_\varepsilon(x)\| \qquad \forall x \in N_{\bar{x}}. \tag{3.7}$$

Proof. (a). The main idea of the proof is based on the results of [1, Theorem 1.4.16]. Therefore, we consider the set-valued mapping $S : \mathbb{R}^n \times \mathbb{R}^n \rightrightarrows \mathbb{R}^n \times \mathbb{R}^n$ with

$$S(x,d) := G^f_{\star,\varepsilon}(x) \times \{0\}, \tag{3.8}$$

and the functions $h : \mathrm{gph}(S) \to \mathbb{R}$, $\vartheta : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ with

$$h(x,d,g,y) := (d,x)^T \begin{pmatrix} g \\ y \end{pmatrix} \quad \text{and} \quad \vartheta(x,d) := \max\{h(x,d,g,y) : (g,y) \in S(x,d)\}.$$

Due to the fact that

$$\mathrm{gph}(S) = \{(x,d,g,0) : g \in G^f_{\star,\varepsilon}(x)\},$$

we obtain

$$h(x,d,g,y) = g^T d \qquad \forall (x,d,g,y) \in \mathrm{gph}(S).$$

Hence, the function h is continuous on $\mathrm{gph}(S)$. Moreover, the mapping S is PK-continuous and locally bounded, since $G^f_{\star,\varepsilon}$ has these properties. Consequently, the identity

$$\vartheta(x,d) = f^G_{\star,\varepsilon}(x,d)$$

and [1, Theorem 1.4.16] prove the assertions, cf. also Remark 3.3 below.
(b). For $\bar{x} \in \mathbb{R}^n$ with $0 \notin G^f_{\star,\varepsilon}(\bar{x})$ we obtain $\pi_\varepsilon(\bar{x}) \neq 0$. Thus, by Theorem 3.1 there is a neighborhood $N_{\bar{x}}$ of $\bar{x}$ such that

$$\pi_\varepsilon(x) \neq 0 \qquad \forall x \in N_{\bar{x}}.$$

In the following let $x \in N_{\bar{x}}$. By the projection theorem we have

$$\pi_\varepsilon(x)^T(g - \pi_\varepsilon(x)) \geq 0 \qquad \forall g \in G^f_{\star,\varepsilon}(x).$$

Due to $\|\pi_\varepsilon(x)\| > 0$ it follows

$$\pi_\varepsilon(x)^T\left(-\frac{\pi_\varepsilon(x)}{\|\pi_\varepsilon(x)\|}\right) \leq g^T\left(-\frac{\pi_\varepsilon(x)}{\|\pi_\varepsilon(x)\|}\right) \qquad \forall g \in G^f_{\star,\varepsilon}(x).$$

Finally, we obtain

$$f^G_{\star,\varepsilon}(x,d_\varepsilon(x)) = \max\{g^T d_\varepsilon(x) : g \in G^f_{\star,\varepsilon}(x)\} = \pi_\varepsilon(x)^T\left(-\frac{\pi_\varepsilon(x)}{\|\pi_\varepsilon(x)\|}\right) = -\|\pi_\varepsilon(x)\|.$$

□

Thus, the continuity property of $G^f_{\star,\varepsilon}$ transfers to the continuity of the directional-derivative $f^G_{\star,\varepsilon}$ with respect to both variables (x,d).

Remark 3.3. The referred results of [1] depend on the continuity term according to Bouligand, Kuratowski and Wilson. In general, this definition is not equivalent to the PK-continuity which we introduced in this paper and used to prove the theorem above. However, the set-valued mapping in (3.8) is locally bounded. Therefore, both continuity terms are equivalent in this situation, see [27, Theorem 5.19].

Next, our interest is focussing on the computation of descent directions of f. The first theorem describes a sufficient condition for a descent direction at the point x, which follows immediately by the separation theorem on the disjoint non-empty compact sets $\{0\}$ and $G^f_{\star,\varepsilon}(x)$.

Theorem 3.4. *Let $f^G_{\star,\varepsilon} : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ be a $\star(\varepsilon)$-directional-derivative of $f : \mathbb{R}^n \to \mathbb{R}$. If $(x,d) \in \mathbb{R}^n \times \mathbb{R}^n$ is a point with $f^G_{\star,\varepsilon}(x,d) < 0$, then d is a descent direction of f at x.* □

The following result is an immediate consequence of Theorem 3.3.

Corollary 3.1. *Let $f^G_{\star,\varepsilon} : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ be a $\star(\varepsilon)$-directional-derivative of $f : \mathbb{R}^n \to \mathbb{R}$, and let $x \in \mathbb{R}^n$ be a point with $0 \notin G^f_{\star,\varepsilon}(x)$. Then $f^G_{\star,\varepsilon}(x,d_\varepsilon(x)) < 0$, where d_ε is defined as in (3.6), i.e. $d_\varepsilon(x)$ is a descent direction of f at x.* □

Due to the properties and assumptions on $\partial_\star f$ and $G^f_{\star,\varepsilon}$, respectively, the local behavior of the graph of f is completely known, see also Remark 3.4 below. The following key result describes this in detail:

Lemma 3.1. *Let $f^G_{\star,\varepsilon}:\mathbb{R}^n\times\mathbb{R}^n\to\mathbb{R}$ be a $\star(\varepsilon)$-directional-derivative of $f:\mathbb{R}^n\to\mathbb{R}$, $c\in(0,1)$, and let $(\bar{x},\bar{d})\in\mathbb{R}^n\times\mathbb{R}^n$ with $\|\bar{d}\|=1$ be a point satisfied*

$$f^G_{\star,\varepsilon}(\bar{x},\bar{d})<0. \tag{3.9}$$

Then there exists $\bar{\delta}>0$ such that for all $x\in B_{\bar{\delta}}(\bar{x})$, $d\in B_{\bar{\delta}}(\bar{d})\cap B_1(0)$, $t\in[0,\bar{\delta}]$ the inequality

$$f(x+td)\leq f(x)+t(1-c)f^G_{\star,\varepsilon}(\bar{x},\bar{d}) \tag{3.10}$$

holds.

Proof. Our proof starts with the observation that for $-cf^G_{\star,\varepsilon}(\bar{x},\bar{d})>0$ there exists $\delta_1>0$ such that for all $(x,d)\in B_{\delta_1}(\bar{x})\times B_{\delta_1}(\bar{d})$ the inequality

$$f^G_{\star,\varepsilon}(x,d)\leq -cf^G_{\star,\varepsilon}(\bar{x},\bar{d})+f^G_{\star,\varepsilon}(\bar{x},\bar{d})=(1-c)f^G_{\star,\varepsilon}(\bar{x},\bar{d}) \tag{3.11}$$

holds, cf. Theorem 3.3 (a). Moreover, since the subdifferential $\partial_\star f$ satisfies the Mean-Value Theorem, see Theorem 2.2, there exists $\delta_2>0$ such that for all $y_1,y_2\in B_{\delta_2}(\bar{x})$ there exists $\theta\in(0,1)$ such that

$$f(y_1)-f(y_2)\in\left\{g^T(y_1-y_2):g\in\partial_\star f(y_2+\theta(y_1-y_2))\right\}. \tag{3.12}$$

Define

$$\bar{\delta}:=\tfrac{1}{2}\min\{\delta_1,\delta_2\}.$$

For all $x\in B_{\bar{\delta}}(\bar{x})$, $d\in B_{\bar{\delta}}(\bar{d})\cap B_1(0)$ and $t\in[0,\bar{\delta})$, we obtain

$$\|(x+t\lambda d)-\bar{x}\|\leq\|x-\bar{x}\|+t\lambda\|d\|\leq\min\{\delta_1,\delta_2\}\qquad\forall\lambda\in(0,1).$$

Thus, inequality (3.11) holds for $(x+t\lambda d,d)$, and property (3.12) holds for $x+td$ and x. Hence, for every $(x,d)\in B_{\bar{\delta}}(\bar{x})\times\left(B_{\bar{\delta}}(\bar{d})\cap B_1(0)\right)$ and $t\in[0,\bar{\delta}]$ there exist $\theta\in(0,1)$ and $g_\theta\in\partial_\star f(x+t\theta d)$ such that

$$f(x+td)-f(x)\overset{(3.12)}{=}tg_\theta^Td\overset{(\mathrm{OS}_1)}{\leq}tf^G_{\star,\varepsilon}(x+t\theta d,d)\overset{(3.11)}{\leq}t(1-c)f^G_{\star,\varepsilon}(\bar{x},\bar{d}).$$

□

Remark 3.4. Basically, the main tools to prove the result of Lemma 3.1 are the Mean-Value Theorem, valid for the Clarke and the symmetric subdifferential, and the Painlevé-Kuratowski continuity of the $\star(\varepsilon)$-subdifferential, requested in condition (OS_3). Thereby, the construction of generalized ε-subdifferential is also meaningful for other subdifferentials, which satisfies the Mean-Value Theorem. A general concept of subdifferentials, which can possibly be used to build the basis, in the sense of (OS_1), of the construction of a generalized ε-subdifferential, is discussed in detail in [19].

Corollary 3.2. *Let $f^G_{\star,\varepsilon}:\mathbb{R}^n\times\mathbb{R}^n\to\mathbb{R}$ be a $\star(\varepsilon)$-directional-derivative of $f:\mathbb{R}^n\to\mathbb{R}$, $c\in(0,1)$, and let $0\notin G^f_{\star,\varepsilon}(\bar{x})$ for $\bar{x}\in\mathbb{R}^n$. Then there exists $\bar{\delta}>0$ such that*

$$f(x+td)\leq f(x)-t(1-c)\|\pi_\varepsilon(\bar{x})\|\qquad\forall x\in B_{\bar{\delta}}(\bar{x})\,\forall d\in B_{\bar{\delta}}(d_\varepsilon(\bar{x}))\cap B_1(0)\,\forall t\in[0,\bar{\delta}], \tag{3.13}$$

where d_ε is defined as in (3.6). In particular,

$$f(x+td_\varepsilon(x))\leq f(x)-t(1-c)\|\pi_\varepsilon(\bar{x})\|\qquad\forall x\in B_{\bar{\delta}}(\bar{x})\,\forall t\in[0,\bar{\delta}]. \tag{3.14}$$

Proof. The proof of assertion (3.13) is straightforward by Lemma 3.1 and representation (3.7). It follows from (3.13) and Theorem 3.3 (b) that assertion (3.14) holds. □

We conclude this section by discussing the concepts of the ε-subdifferential of convex analysis and the Goldstein-ε-subdifferential in relation to the concept of generalized ε-subdifferentials. Note that every convex function is also locally Lipschitz continuous and that the subdifferential of convex analysis is identical with the Clarke subdifferential.

Theorem 3.5. *Let $f:\mathbb{R}^n \to \mathbb{R}$ be convex, and $\varepsilon > 0$. Then the ε-subdifferential of f, $\partial_{C,\varepsilon} f:\mathbb{R}^n \rightrightarrows \mathbb{R}^n$ defined by*

$$\partial_{C,\varepsilon} f(x) := \{g \in \mathbb{R}^n : g^T(y-x) \le f(y) - f(x) + \varepsilon \;\forall y \in \mathbb{R}^n\},$$

is a $C(\varepsilon)$-subdifferential.

Proof. The conditions (OS_1) and (OS_2) follow from [21, Theorem 3.3.1.4 and Theorem 3.3.1.5]. Immediately, by definition we have (OS_4). The inner semicontinuity of $\partial_{C,\varepsilon} f$ follows from [13, Theorem 4.1.3 and the subsequent comments]. To prove the outer semicontinuity of $\partial_{C,\varepsilon} f$ on $\mathbb{R}^n$, and thus to complete the proof of the (OS_3) condition, choose an arbitrary $x \in \mathbb{R}^n$ and sequences $x_k \to x$, $g_k \to g$ with $g_k \in \partial_{C,\varepsilon} f(x_k)$ for every $k \in \mathbb{N}$. Hence, for all $k \in \mathbb{N}$ we have

$$g_k^T(y-x_k) \le f(y) - f(x_k) + \varepsilon.$$

By letting $k \to \infty$ on both sides, we obtain $g \in \partial_{C,\varepsilon} f(x)$. Finally, the condition (OS_5) follows essentially by [13, XI. Theorem 4.1.2] and [21, I. Theorem 2.1.5]. □

Thus, the ε-subdifferential fits in our concept. Thereby, we obtain a well-known formula for a $C(\varepsilon)$-directional-derivative of a convex function:

Example 3.1. For a convex function $f:\mathbb{R}^n \to \mathbb{R}$ and $\varepsilon > 0$ we have

$$f^G_{C,\varepsilon}(x,d) = \inf_{t>0} \frac{f(x+td) - f(x) + \varepsilon}{t} \qquad \forall (x,d) \in \mathbb{R}^n \times \mathbb{R}^n,$$

see Theorem 3.5 and [13, XI. Theorem 2.1.1].

Remark 3.5. The ε-subdifferential of convex analysis shows that there does not have to exist a smallest $\star(\varepsilon)$-subdifferential with respect to the partial order $\subseteq$: For a convex function $f:\mathbb{R}^n \to \mathbb{R}$ and for every $x \in \mathbb{R}^n$ we have $\bigcap_{\varepsilon>0} \partial_{C,\varepsilon} f(x) = \partial_C f(x)$.

However, the Goldstein-ε-subdifferential does not satisfy the (OS_3) condition, which is needed, for instance, to guarantee the result of Lemma 3.1. We discuss the Goldstein approach in the following remark:

Remark 3.6. For the simple function $f:\mathbb{R} \to \mathbb{R}$ with $f(x) := -|x|$ and $\varepsilon > 0$ the Goldstein-ε-subdifferential of f is given by

$$\partial^G_{C,\varepsilon} f(x) := \operatorname{conv} \bigcup_{y \in B_\varepsilon(x)} \partial_C f(y) = \begin{cases} \{-1\} & \text{for } x > \varepsilon, \\ [-1,1] & \text{for } -\varepsilon \le x \le \varepsilon, \\ \{1\} & \text{for } x < -\varepsilon, \end{cases}$$

see Figure 2 below. In general, the (OS_3) condition is not satisfied for this mapping.

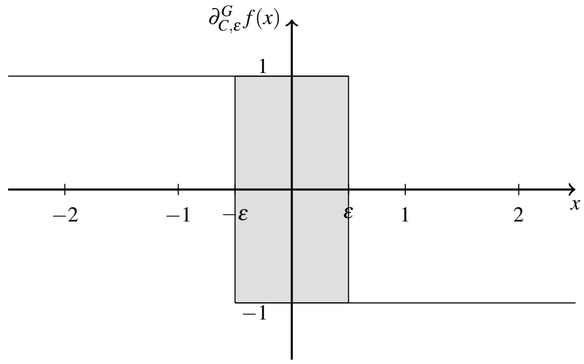


FIG. 2 Goldstein-ε-subdifferential of the function $f(x) := -|x|$

Obviously, this mapping is not inner semicontinuous at the points $\pm\varepsilon$, since

$$\left\{g \in \mathbb{R} : \forall x_k \to \pm\varepsilon\ \exists g_k \to g \text{ with } g_k \in \partial^G_{C,\varepsilon} f(x_k)\ \forall k\right\} = \{\mp 1\} \not\supseteq [-1,1] = \partial^G_{C,\varepsilon} f(\pm\varepsilon).$$

4. Conceptual Algorithm for the Calculation of $\star$-Stationary Points

Let $G^f_{\star,\varepsilon} : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ be a generalized ε-subdifferential of a locally Lipschitz continuous function $f : \mathbb{R}^n \to \mathbb{R}$, and $\star \in \{C,S\}$. In this section, we are focussed on a descent method based on $\star(\varepsilon)$-subdifferentials to achieve convergence to $\star$-stationary points, i.e. points $\bar{x}$ with $0 \in \partial_\star f(\bar{x})$. In contrast to the previous section, here the conditions (OS_4) and (OS_5) are essential. However, it is possible to achieve convergence to ε-stationary points, i.e. points $\bar{x}$ with $0 \in G^f_{\star,\varepsilon}(\bar{x})$, by descent methods based on $\star(\varepsilon)$-subdifferentials just satisfying the conditions (OS_1), (OS_2), and, (OS_3). Such algorithms are discussed in detail in [19].

Algorithm 4.1 (Conceptual algorithm)**.**

Input: $x_0 \in \mathbb{R}^n$ and $\varepsilon_0 > 0$

1: Set $k := 0$
2: **while** $0 \notin G^f_{\star,\varepsilon_k}(x_k)$ **do**
3: Compute the unique solution $\pi_{\varepsilon_k}(x_k)$ of the problem $\min\left\{\|g\| : g \in G^f_{\star,\varepsilon_k}(x_k)\right\}$, cf. Theorem 3.1
4: Set $d_k := -\dfrac{\pi_{\varepsilon_k}(x_k)}{\|\pi_{\varepsilon_k}(x_k)\|}$
5: Compute a step size $t_k > 0$ such that $f(x_k + t_k d_k) < f(x_k)$, e.g. with (4.1) or (4.9)
6: Update $x_{k+1} := x_k + t_k d_k$
7: Choose ε_{k+1} with $0 < \varepsilon_{k+1} < \varepsilon_k$ and set $k+1 \mapsto k$
8: **end while**
9: **return** x_k

Due to Corollary 3.2 the vector d_k is a descent direction of f at x_k. Thus Step 5 is well-defined. Similar to descent methods for smooth objective functions we need a requirement on the step size t_k to guarantee convergence. On this account, we precise Step 5 in the following theorems.

Theorem 4.1. *Let $(x_k)_{k\in\mathbb{N}_0}$ be generated by Algorithm 4.1 and let the sublevel set $\{x\in\mathbb{R}^n : f(x)\le f(x_0)\}$ be bounded for every $x_0\in\mathbb{R}^n$. Furthermore, specify Step 5 by the following condition:*

$$\exists\, \tau>0\;\forall k\in\mathbb{N}_0 \text{ such that } f(x_k+t_kd_k)\le f(x_k)-t_k\|\pi_{\varepsilon_k}(x_k)\| \;\wedge\; \tau\le t_k \tag{4.1}$$

Then Algorithm 4.1 either terminates after finitely many iterations at some $\bar{k}\in\mathbb{N}_0$ with $0\in G^f_{\star,\varepsilon_{\bar{k}}}(x_{\bar{k}})$ or every accumulation point of the sequence $(x_k)_{k\in\mathbb{N}_0}$ is a $\star$-stationary point of f.

Proof. If the algorithm terminates, then there exists an index $\bar{k}$ such that the assertion holds, by construction. To prove the more interesting part of the theorem, let $\hat{x}$ be an accumulation point of $(x_k)_{k\in\mathbb{N}_0}$, i.e there exists a subsequence $(x_{k_l})_{l\in\mathbb{N}_0}$ such that

$$\lim_{l\to\infty} x_{k_l}=\hat{x}. \tag{4.2}$$

By condition (4.1) there exists $\tau>0$ such that

$$f(x_{k_{l+1}})\le f(x_{k_l+1})=f(x_{k_l}+t_{k_l}d_{k_l})\le f(x_{k_l})-t_{k_l}\|\pi_{\varepsilon_{k_l}}(x_{k_l})\| \quad\text{and}\quad \tau\le t_{k_l} \qquad \forall l\in\mathbb{N}_0. \tag{4.3}$$

Since the sublevel set is bounded and f is continuous, we have

$$\inf\{f(y):y\in\mathbb{R}^n\}=\min\{f(y):y\in\{x:f(x)\le f(x_0)\}\}>-\infty,$$

i.e. the sequence $(f(x_{k_l}))_{l\in\mathbb{N}_0}$ is bounded from below. By construction this sequence is also monotonically decreasing, and therefore, it is a Cauchy sequence. It follows that for each $\alpha>0$ there exists $L_1\in\mathbb{N}_0$ such that

$$f(x_{k_l})-f(x_{k_{l+1}})=|f(x_{k_l})-f(x_{k_{l+1}})|\le\alpha \qquad \forall l\ge L_1. \tag{4.4}$$

Hence, we obtain

$$\|\pi_{\varepsilon_{k_l}}(x_{k_l})\| \overset{(4.3)}{\le} \tfrac{1}{t_{k_l}}\left(f(x_{k_l})-f(x_{k_{l+1}})\right) \overset{(4.4)}{\le} \tfrac{1}{t_{k_l}}\alpha\le\frac{1}{\tau}\alpha \qquad \forall l\ge L_1, \tag{4.5}$$

i.e.

$$\min\left\{\|g\|:g\in G^f_{\star,\varepsilon_{k_l}}(x_{k_l})\right\}\le\frac{1}{\tau}\alpha \qquad \forall l\ge L_1. \tag{4.6}$$

Moreover, by condition (OS_5) for $\hat{x}$ and for each $\eta>0$, we can find $\varepsilon:=\varepsilon(\hat{x},\eta)>0$ and $\delta:=\delta(\hat{x},\eta)>0$ such that

$$\bigcup_{x\in B_\delta(\hat{x})} G^f_{\star,\varepsilon}(x)\subseteq\partial_\star f(\hat{x})+B_\eta(0). \tag{4.7}$$

Due to (4.2) and Step 7 of the algorithm there exists an index $L_2\in\mathbb{N}_0$ such that

$$x_{k_l}\in B_\delta(\hat{x}) \quad\text{and}\quad \varepsilon_{k_l}\le\varepsilon \qquad \forall l\ge L_2. \tag{4.8}$$

Finally, by (4.6), (4.7), (4.8), and the (OS_4) condition, for all $l\ge\max\{L_1,L_2\}$ we obtain

$$\min\{\|g\|:g\in\partial_\star f(\hat{x})+B_\eta(0)\}\le\frac{1}{\tau}\alpha$$

and thus

$$\min\{\|g\|:g\in\partial_\star f(\hat{x})\}\le\frac{1}{\tau}\alpha+\eta.$$

By letting $\alpha\searrow 0$ and $\eta\searrow 0$ we have

$$\min\{\|g\|:g\in\partial_\star f(\hat{x})\}=0,$$

i.e. $\hat{x}$ is a $\star$-stationary point of f. □

The boundedness assumption on the sublevel set implies that problem (1.1) has at least one minimal solution. Note, that this assumption is also common for problems of type (1.1) with a smooth objective function. Therefore, we do not require more than usually demanded.

Theorem 4.2. *Let $(x_k)_{k\in\mathbb{N}_0}$ be generated by Algorithm 4.1 and let the sublevel set $\{x \in \mathbb{R}^n : f(x) \leq f(x_0)\}$ be bounded for every $x_0 \in \mathbb{R}^n$. Furthermore, specify Step 5 by the following condition:*

$$\exists\, \vartheta > 0\ \forall k \in \mathbb{N}_0 \text{ such that } f(x_k + t_k d_k) \leq f(x_k) - \vartheta \|\pi_{\varepsilon_k}(x_k)\|. \tag{4.9}$$

Then Algorithm 4.1 either terminates after finitely many iterations at some $\bar{k} \in \mathbb{N}_0$ with $0 \in G^f_{\star,\varepsilon_{\bar{k}}}(x_{\bar{k}})$ or every accumulation point of the sequence $(x_k)_{k\in\mathbb{N}_0}$ is a $\star$-stationary point of f.

Proof. If the algorithm terminates, then there exists an index $\bar{k}$ such that the assertion holds, by construction. To prove the more interesting part of the theorem, let $\hat{x}$ be an accumulation point of $(x_k)_{k\in\mathbb{N}_0}$. For simplicity denote the associated subsequence by $(x_k)_{k\in\mathbb{N}_0}$ i.e.

$$\lim_{k\to\infty} x_k = \hat{x}. \tag{4.10}$$

Analogously, to the prove of Theorem 4.1 we have that $(f(x_k))_{k\in\mathbb{N}_0}$ is a Cauchy sequence. Therefore, for each $\alpha > 0$ there exists $K_1 \in \mathbb{N}_0$ such that

$$f(x_k) - f(x_{k+1}) = |f(x_k) - f(x_{k+1})| \leq \alpha \qquad \forall k \geq K_1. \tag{4.11}$$

Hence, we obtain

$$\|\pi_{\varepsilon_k}(x_k)\| \overset{(4.9)}{\leq} \tfrac{1}{\vartheta}(f(x_k) - f(x_{k+1})) \overset{(4.11)}{\leq} \frac{\alpha}{\vartheta} \qquad \forall k \geq K_1, \tag{4.12}$$

i.e.

$$\min\left\{\|g\| : g \in G^f_{\star,\varepsilon_k}(x_k)\right\} \leq \frac{\alpha}{\vartheta} \qquad \forall k \geq K_1. \tag{4.13}$$

Moreover, by condition (OS_5) for $\hat{x}$ and for each $\eta > 0$, we can find $\varepsilon := \varepsilon(\hat{x},\eta) > 0$ and $\delta := \delta(\hat{x},\eta) > 0$ such that

$$\bigcup_{x\in B_\delta(\hat{x})} G^f_{\star,\varepsilon}(x) \subseteq \partial_\star f(\hat{x}) + B_\eta(0). \tag{4.14}$$

Due to (4.10) and Step 7 of the algorithm there exists an index $K_2 \in \mathbb{N}_0$ such that

$$x_k \in B_\delta(\hat{x}) \quad \text{and} \quad \varepsilon_k \leq \varepsilon \qquad \forall k \geq K_2. \tag{4.15}$$

Finally, by (4.13), (4.14), (4.15), and the (OS_4) condition, for all $k \geq \max\{K_1, K_2\}$ we obtain

$$\min\{\|g\| : g \in \partial_\star f(\hat{x}) + B_\eta(0)\} \leq \frac{\alpha}{\vartheta}$$

and thus

$$\min\{\|g\| : g \in \partial_\star f(\hat{x})\} \leq \frac{\alpha}{\vartheta} + \eta.$$

By letting $\alpha \searrow 0$ and $\eta \searrow 0$ we have

$$\min\{\|g\| : g \in \partial_\star f(\hat{x})\} = 0,$$

i.e. $\hat{x}$ is a $\star$-stationary point of f. □

Note that Theorem 4.1 is a corollary of Theorem 4.2, since condition (4.1) implies condition (4.9). But from a theoretical point of view, the condition (4.1) is a generalization of the well-known efficient step size rule. On this account, the result of Theorem 4.1 is more familiar to the results of unconstrained smooth optimization problems. In particular, for a deeper understanding of this result we like to present the corresponding proof separately.

5. Conclusion

In this paper a general theory for the realization of converging descent methods for nonsmooth optimization problems is described with an approach of generalized ε-subdifferentials. Since the algorithm framework requires neither semismoothness nor calculation of exact subgradients, this theory is applicable for a broader class of locally Lipschitz continuous functions than other methods usually used in nonsmooth optimization, as described in the introduction. In particular, the proposed approach is also useable for semismooth functions, and thus also for convex ones. From a theoretical point of view convergence is established through the construction of generalized ε-subdifferentials. Thus, the challenge lies in the design of it. In general, we have shown that for a convex function the ε-subdifferential fulfills all the required properties. This fact emphasizes the approach of our theory. Nevertheless, we have also discussed that for a locally Lipschitz continuous function the Goldstein-ε-subdifferential is not inner semicontinuous. For this reason, we propose for future applications to design a generalized ε-subdifferential particularly for the problem of current interest. The construction can be done by means of the calculus rules, given supersets or exact formulas of the Clarke subdifferential or the limiting subdifferential, respectively.

References

[1] J.-P. Aubin, H. Frankowska, Set-Valued Analysis, Birkhäuser, 1st edition, 1990.
[2] A. M. Bagirov, Continuous subdifferential approximations and their applications, J. Math. Sci. 115 (2003), 2567-2609.
[3] A. M. Bagirov, A. N. Ganjehlou, An approximate subgradient algorithm for unconstrained nonsmooth, nonconvex optimization, Math. Methods Oper. Res. 67 (2008), 187-206.
[4] A. M. Bagirov, N. Karmitsa, M. M. Mäkelä, Introduction to Nonsmooth Optimization, Springer, 1st edition, 2014.
[5] F. Clarke, Generalized gradients of lipschitz functionals, Adv. Math. 40 (1981), 52-67.
[6] F. H. Clarke, Optimization and Nonsmooth Analysis, Wiley, 1st edition, 1983.
[7] V. F. Demyanov, V. N. Malozemov, Einführung in Minimax-Probleme, Akademische Verlagsgesellschaft Geest & Portig K.-G, 1st edition, 1975.
[8] V. F. Demyanov, A. M. Rubinov, Constructive Nonsmooth Analysis, Lang, 1st edition, 1995.
[9] F. Dubeau, J. Gauvin, Differential properties of the marginal function in mathematical programming, Math. Program. Stud. 19 (1982), 101-119.
[10] B. M. Glover, A. M. Rubinov, H. Xu, Continuous approximations to generalized jacobians, Optimization 46 (1999), 221-246.
[11] A. A. Goldstein, Optimization of lipschitz continuous functions, Math. Program. 13 (1977), 14-22.
[12] N. Haarala, K. Miettinen, M. M. Mäkelä, Globally convergent limited memory bundle method for large-scale nonsmooth optimization, Math. Program. 109 (2007), 181-205.
[13] J.-B. Hiriart-Urruty, C. Lemaréchal, Convex Analysis and Minimization Algorithms II, Springer, 1st edition, 1993.
[14] J.-B. Hiriart-Urruty, C. Lemaréchal, Convex Analysis and Minimization Algorithms I, Springer, 2nd edition, 1996.

[15] J.-B. Hiriart-Urruty, C. Lemaréchal, Fundaments of Convex Analysis, Springer, 2nd edition, 2004.

[16] A. Ioffe, Calculus of dini subdifferentials of functions and contingent coderivatives of set-valued maps, Nonlinear Anal. 8 (1984), 517-539.

[17] N. Karmitsa, Diagonal bundle method for nonsmooth sparse optimization, J. Optim. Theory Appl. 166 (2015), 889-905.

[18] K. C. Kiwiel, Methods of Descent for Nondifferentiable Optimization, Springer, 1st edition, 1985.

[19] M. Knossalla, Continuous outer subdifferentias in nonsmooth optimization, submitted, 2016.

[20] L. Lukan, J. Vlek, Globally convergent variable metric method for nonconvex nondifferentiable unconstrained minimization, J. Optim. Theory Appl. 111 (2001), 407-430.

[21] M. M. Mäkelä, P. Neittaanmäki, Nonsmooth Optimization, World Scientific, 1st edition, 1992.

[22] R. Mifflin, Semismooth and semiconvex functions in constrained optimization, SIAM J. Control Optim. 15 (1976), 959-972.

[23] R. Mifflin, An algorithm for constrained optimization with semismooth functions, Math. Oper. Res. 2 (1977), 191-207.

[24] B. Mordukhovich, Y. Shao, On nonconvex subdifferentials calculus in banach spaces, J. Convex Anal. 2 (1955), 211-227.

[25] B. S. Mordukhovich, Variational Analysis and Generalized Differentiation I, Springer, 1st edition, 2006.

[26] R. T. Rockafellar, Convex Analysis, Princeton University Press, 1st edition, 1970.

[27] R. T. Rockafellar, R. J.-B. Wets, Variational Analysis, Springer, 3rd edition, 2009.

[28] P. Wolfe, A method of conjugate subgradients for minimizing nondifferentiable functions, Math. Program. Stud. 3 (1975), 145-173.