Set-Valued and Variational Analysis
https://doi.org/10.1007/s11228-021-00610-3

SI: OPTIMIZATION, CONVEX AND VARIATIONAL ANALYSIS

Convergence of the Gradient Sampling Algorithm on Directionally Lipschitz Functions

J. V. Burke[1] **· Q. Lin**[2]

Received: 2 July 2021 / Accepted: 25 September 2021 / Published online: 14 October 2021

Abstract
The convergence theory for the gradient sampling algorithm is extended to directionally Lipschitz functions. Although directionally Lipschitz functions are not necessarily locally Lipschitz, they are almost everywhere differentiable and well approximated by gradients and so are a natural candidate for the application of the gradient sampling algorithm. The main obstacle to this extension is the potential unboundedness or emptiness of the Clarke subdifferential at points of interest. The convergence analysis we present provides one path to addressing these issues. In particular, we recover the usual convergence theory when the function is locally Lipschitz. Moreover, if the algorithm does not drive a certain measure of criticality to zero, then the iterates must converge to a point at which either the Clarke subdifferential is empty or the direction of steepest descent is degenerate in the sense that it does lie in the interior of the domain of the regular subderivative.

Keywords Gradient sampling algorithm · Non-Lipschitzian · Directionally Lipschitz · Nonsmooth optimization

Mathematics Subject Classification (2010) 49J22 · 65K05 · 65K10 · 90C26

1 Introduction

The gradient sampling (GS) algorithm is designed to solve non-smooth optimization problems by using locally sampled gradients to approximate the Clarke subdifferential and the associated direction of steepest descent. The objective is assumed to be continuously

This paper is dedicated to Terry Rockafellar on the occasion of his 85th birthday.

Supported in part by the U.S. National Science Foundation grant DMS-1908890.

✉ J. V. Burke
jvburke01@gmail.com

Q. Lin
qiuyinglin5499@gmail.com

[1] University of Washington, Seattle, WA, USA

[2] Amazon Corp., 410 Terry Ave N., Seattle, WA, USA

differentiable on an open set $\mathcal{D}$ of full Lebesgue measure. Although the method was originally applied to minimize non-Lipschitzian nonsymmetric spectral functions [4–6], the existing convergence theory only applies to locally Lipschitz functions. The purpose of this note is to extend the convergence theory to directionally Lipschitz functions (see Definition 1). Directionally Lipschitz functions were introduced by Rockafellar in [16] and further developed in [17]. Loosely speaking, a function is directionally Lipschitz at a point $\bar{x}$ if it is possible to "tilt" its epigraph in such a way that the tilted set is the epigraph of a function that is locally Lipschitz at $\bar{x}$. A function can be directionally Lipschitz at a point but not locally Lipschitz or even continuous at that point. Some of the ideas for our approach appear in [12] and are motivated by the results from [3, 6, 7, 11]. In particular, our choice of directionally Lipschitz functions is inspired by [3, Corollary 6.1] (see Theorem 3) where it is shown that nearby gradients can be used to approximate their subdifferential. The primary difficulty in the non-Lipschitzian case is the potential unboundedness or emptiness of the subdifferential. Indeed, in this setting, it is not entirely clear what kind of convergence result can reasonably be expected.

Both our choice of how the algorithm is stated and the consequent convergence theory closely parallels those proposed by Kiwiel in [11] since his approach provides the most complete picture in the Lipschitzian case. A nice discussion of this approach as well as other recent advances and ongoing work is given in [7]. The paper proceeds as follows. Section 2 is broken into 4 parts: (1) notation and a review of the subdifferential calculus especially the Clarke subdifferential and its relationship to the generalized (Mordukhovich or limiting) subdifferential, (2) pointedness of cones and its use in approximating the distance to a convex set, (3) the direction of steepest descent for nonsmooth functions, and (4) an introduction to directionally Lipschitz functions. In Section 3 we state the version of the gradient sampling algorithm to be examined and present our convergence results. We conclude in Section 4 with a few comments on the algorithm and its convergence.

2 Preliminaries

2.1 Notation

Our notation is based on that used in [14]. We work in the n-dimensional Euclidean space $\mathbb{R}^n$ with the standard inner product $\langle x\,,\,y\rangle$, with $\|\cdot\|$ denoting the associated 2-norm whose closed unit ball is $\mathbb{B} := \{x \in \mathbb{X} \mid \|x\| \leq 1\}$. Given $x \in \mathbb{R}^n$, define the open $\epsilon > 0$ ball about x as the set $B_\epsilon(x) := \{y \mid \|x - y\| < \epsilon\}$. Let C be a subset of a Euclidean space $\mathbb{X}$ whose norm is denoted by $\|\cdot\|$. We say C is convex if every line segment connecting two points in C is contained in C, and C is affine if it is the translate of a subspace. The affine hull of C, denoted $\operatorname{aff} C$, and the convex hull of C, denoted $\operatorname{conv} C$, are the intersection of all affine, respectively, convex sets that contain it. If C is convex, its relative interior is the set $\operatorname{ri} C := \{x \in C \mid \exists\, \epsilon > 0 \text{ s.t. } \operatorname{aff} C \cap B_\epsilon(x) \subset C\}$. Denote the closure and interior of C by $\operatorname{cl} C$ and $\operatorname{int} C$, respectively. The distance to C is defined by $\operatorname{dist}(x \mid C) := \inf_{z \in C} \|x - z\|$. The Projection Theorem for convex sets tells us that for a nonempty closed convex set $C \subset \mathbb{X}$ and any $x \in \mathbb{X}$ there is a unique vector $\hat{x} \in C$ such that $\operatorname{dist}(x \mid C) = \|x - \hat{x}\|$. The vector $\hat{x}$ is called the projection of x onto C and is denoted by $\operatorname{proj}_C(x)$.

The set of natural numbers is denoted by $\mathbb{N} := \{1, 2, \ldots\}$. Let $\Delta_n := \left\{\lambda \in \mathbb{R}^{n+1}_+ \mid \lambda_1 + \cdots + \lambda_{n+1} = 1\right\}$ be the unit simplex in $\mathbb{R}^{n+1}$, $\mathbb{R}_+$ the set of non-negative reals, and $\mathbb{R}_{++}$ the set of positive reals.

A subset K of the Euclidean space $\mathbb{X}$ is said to be a cone if $0 \in K$ and $\lambda x \in K$ for all $x \in K$ and $\lambda \geq 0$. It is said to be a convex cone if it is both a cone and a convex set. The cone $K \subset \mathbb{R}^n$ is said to be pointed if for all $k \geq 2$ and $x^1, x^2, \ldots, x^k \in K$ one has $x^1 + x^2 + \cdots + x^k = 0$ if and only if $x^i = 0,\ i = 1, \ldots, k$.

The horizon cone and polar of $C \subset \mathbb{X}$ are given by

$$C^\infty := \left\{ w \;\middle|\; \exists \{x^k\} \subset C,\ t_k \downarrow 0 \text{ s.t. } t_k x^k \to w \right\} \text{ and}$$
$$C^* := \left\{ v \in \mathbb{X}^* \mid \langle v\,, x \rangle \leq 1\ \forall, x \in C \right\},$$

respectively. The polar of a nonempty set is aways a closed convex set. In addition, if C is a cone, then one can show that $C^* = \{ v \in \mathbb{X} \mid \langle v\,, x \rangle \leq 0\ \forall, x \in C \}$. The convex indicator and support function for C are give by

$$\delta_C(x) := \begin{cases} 0, & x \in C, \\ +\infty, & x \notin C \end{cases} \qquad \text{and} \qquad \delta^*_C(v) := \sup_{x \in C} \langle v\,, x \rangle\ ,$$

respectively.

Given Euclidean spaces $\mathbb{X}$ and $\mathbb{Y}$, a mapping S from $\mathbb{X}$ to $\mathbb{Y}$ for which $S(x)$ is a subset of $\mathbb{Y}$ for every $x \in \mathbb{X}$ (possibly empty) is called a multivalued mapping and is denoted by $S : \mathbb{X} \rightrightarrows \mathbb{Y}$. The domain of S is the set $\operatorname{dom}(S) := \{x \mid S(x) \neq \emptyset\}$. Such a mapping S is said to be outer semicontinuous (osc) if

$$\left\{ v \;\middle|\; \exists\, (x^k, v^k) \to (x, v) \text{ with } v^k \in S(x^k)\, \forall\, k \right\} \subset S(x) \quad \forall\, x \in \operatorname{dom}(S)\,.$$

The graph of S is the set $\operatorname{graph}(S) := \{(x, y) \mid y \in S(x)\}$ and the osc hull of S is the multivalued mapping $\operatorname{cl} S : \mathbb{X} \rightrightarrows \mathbb{Y}$ such that $\operatorname{graph}(\operatorname{cl} S) = \operatorname{cl} \operatorname{graph}(S)$.

Let $f : \mathbb{R}^n \to \overline{\mathbb{R}} := \mathbb{R} \cup \{+\infty\}$ and set

$$\operatorname{dom}(f) := \{x \mid f(x) < \infty\}$$
$$\operatorname{epi} f := \{(x, \mu) \mid f(x) \leq \mu\} \subset \mathbb{R}^n \times \mathbb{R}.$$

Let $\bar{x} \in \operatorname{dom}(f)$. The regular subdifferential of f at $\bar{x}$ is given by $\hat{\partial} f(x) := \{v \mid f(z) \geq f(x) + \langle v\,, z - x \rangle + o(\|z - x\|)\}$. This set is always closed and convex, but may be empty. The subdifferential of f at $\bar{x}$ is given by $\partial f(\bar{x}) = \left\{ v \;\middle|\; \exists\, x^k \to \bar{x},\ v^k \to v \text{ s.t. } v^k \in \hat{\partial} f(x^k)\ \forall\, k \in \mathbb{N} \right\}$, and the horizon subdifferential of f at $\bar{x}$ is given by

$$\partial^\infty f(\bar{x}) := \left\{ v \;\middle|\; \begin{array}{c} \exists\, x^k \to \bar{x},\ t_k \downarrow 0,\ t_k v^k \to v,\ \text{s.t.} \\ v^k \in \hat{\partial} f(x^k)\ \forall\, k \in \mathbb{N} \end{array} \right\}. \tag{1}$$

These sets are always closed, and if f is lower semi-continuous (lsc) at $\bar{x}$ then either $\partial f(\bar{x}) \neq \emptyset$ or $\partial^\infty f(\bar{x})$ contains at least one nonzero element [14, Corollary 8.10]. These subdifferentials are all mutivalued mappings with ∂f and $\partial^\infty f$ osc along f-attentive sequences by construction (an f-attentive sequence is any sequence $\{x^k\} \subset \operatorname{dom}(f)$ such that if $x^k \to \bar{x}$ then $f(x^k) \to f(\bar{x})$).

Given a closed nonempty set $C \subset \mathbb{R}^n$ and a point $\bar{x} \in C$, the regular normal cone to C at $\bar{x}$ is the set

$$\widehat{N}_C(\bar{x}) := \{v \mid \langle v\,, x - \bar{x} \rangle \leq o(\|x - \bar{x}\|) \text{ for } x \in C\}\,.$$

The osc hull of this multivalued mapping is called the normal cone mapping and is denoted by $N_C(\bar{x})$. The Clarke normal cone to C at x is given by $\overline{N}_C(x) := \operatorname{cl} \operatorname{conv} N_C(\bar{x})$. The

cone of regular tangents to C at a point $x \in C$ where C is locally closed is defined by $\widehat{T}_C(x) := N_C(x)^*$ [14, Theorem 6.28].

Given $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ and $\bar{x} \in \operatorname{dom}(f)$ at which f is lsc, the Clarke subdifferential of f at $\bar{x}$ is $\bar{\partial} f(x) := \left\{ v \;\middle|\; (v, -1) \in \overline{N}_{\operatorname{epi} f}(x, f(x)) \right\}$, and $\bar{\partial}^\infty f(x) := \left\{ v \;\middle|\; (v, 0) \in \overline{N}_{\operatorname{epi} f}(x, f(x)) \right\}$ is the Clarke horizon subdifferential of f at $\bar{x}$ [14, Theorem 8.49]. The subdifferential and the Clarke subdifferential reduce to the usual subdifferential in convex analysis when f is convex. Finally, the regular subderivative of f at $x \in \operatorname{dom}(f)$, denoted $\hat{d} f(x) : \mathbb{R}^n \to \mathbb{R} \cup \{\pm\infty\}$, at points where f is lsc is defined by the relation $\operatorname{epi}(\hat{d} f(x)) = \widehat{T}_{\operatorname{epi} f}(x, f(x))$ [14, Theorem 8.17]. The regular subderivative coincides with Clarke's directional derivative when f is locally Lipschitz [10]. The following theorem establishes the relationships between the subdifferential and the Clarke subdifferential.

Theorem 1 (Subdifferential Relationships) [14, Theorem 8.49 and Exercise 8.23] *Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be locally lsc and finite-valued at $\bar{x} \in \mathbb{R}^n$. Then the following hold:*

1. *$\partial f(x)$ and $\partial^\infty f(x)$ are osc at $\bar{x}$ with respect to $f-$attentive convergence, that is, with respect to sequences $\{x^k\} \subset \operatorname{dom}(f)$ such that $(x^k, f(x^k)) \to (x, f(x))$.*
2. *$\bar{\partial} f(\bar{x})$ is a closed convex set and $\bar{\partial}^\infty f(\bar{x})$ is a closed convex cone.*
3. *$\bar{\partial}^\infty f(\bar{x}) = \bar{\partial} f(\bar{x})^\infty$ when $\bar{\partial} f(\bar{x}) \neq \emptyset$, or equivalently, $\partial f(\bar{x}) \neq \emptyset$.*
4. *If the cone $\partial^\infty f(\bar{x})$ is pointed (or equivalently, $\bar{\partial}^\infty f(\bar{x})$ is pointed), then*

$$\bar{\partial} f(\bar{x}) = \operatorname{conv} \partial f(\bar{x}) + \operatorname{conv} \partial^\infty f(\bar{x}) \;\text{ and }\; \bar{\partial}^\infty f(\bar{x}) = \operatorname{conv} \partial^\infty f(\bar{x}).$$

Moreover, if $\bar{\partial} f(\bar{x}) \neq \emptyset$ (equivalently, $\partial f(\bar{x}) \neq \emptyset$), then $\hat{d} f(x) = \delta^*_{\bar{\partial} f(x)}$.

Finally, at various points in the paper we say that a set $\mathcal{Q}$ is a full measure subset of another set $\mathcal{V}$. By this we mean a full Lebesgue measure subset, i.e., both $\mathcal{Q}$ and $\mathcal{V}$ are Lebesgue measurable, $\mathcal{Q} \subset \mathcal{V}$, and for every Lesbesgue measurable set $\mathcal{M}$, $\ell(\mathcal{M} \cap \mathcal{Q}) = \ell(\mathcal{M} \cap \mathcal{V})$, where ℓ is Lebesgue measure on $\mathbb{R}^n$ with the value of n determined by context. In addition, we say that an event occurs with probability 1 if it occurs with probability 1 relative to any probability measure that is absolutely continuous with respect to Lebesgue measure.

2.2 Pointedness

We review pointedness and a few of its properties.

Lemma 1 *Let K be a non-empty closed cone in $\mathbb{R}^n$ and consider the following statements:*

(i) *K is pointed.*
(ii) $K \cap (-K) = \{0\}$.
(iii) $\operatorname{int} K^* \neq \emptyset$.

Statements (i) and (ii) are equivalent, and if K is convex, both are equivalent to (iii). Moreover, in the convex case, $z \in \operatorname{int} K^*$ if and only if there exist $\epsilon > 0$ such that $\langle z, w \rangle \leq -\epsilon \|w\|$ for all $w \in K$.

Proof The statements concerning (i)-(iii) follow from [14, Proposition 3.14, Exercise 6.22]. Therefore, we need only establish the the final statement of the lemma. Let $z \in \operatorname{int} K^*$ and

$\epsilon > 0$ be such that $z + \epsilon\mathbb{B} \subset K^*$. Then, for all $w \in K$ and $u \in \mathbb{B}$, $0 \geq \langle z + \epsilon u\,,\, w\rangle = \langle z\,,\, w\rangle + \epsilon\,\langle u\,,\, w\rangle$. Hence, $0 \geq \langle z\,,\, w\rangle + \epsilon \sup_{u\in\mathbb{B}} \langle u\,,\, w\rangle = \langle z\,,\, w\rangle + \epsilon\,\|w\|$.

On the other hand, if there is a $z \in \mathbb{R}^n$ and $\epsilon > 0$ is such that $\langle z\,,\, w\rangle \leq -\epsilon\,\|w\|$ for all $w \in K$, then, for all $u \in \mathbb{B}$ and $w \in K$, $\left\langle z + \frac{\epsilon}{2}u\,,\, w\right\rangle \leq -\epsilon\,\|w\| + \frac{\epsilon}{2}\,\|w\| = -\frac{\epsilon}{2}\,\|w\|$ so that $z \in \operatorname{int} K^*$. □

We now connect the pointedness of C^∞ to projections and the distance function for a non-empty closed convex set. This result extends lemma [11, Lemma 3.1] and introduces a condition that is key to our analysis of the non-Lipschitzian setting.

Lemma 2 (Pointedness, and Projections) *Let C be a non-empty closed convex subset of $\mathbb{R}^n$ such that C^∞ is pointed. Let $z \notin C$ be such that*

$$z - \operatorname{proj}_C(z) \in \operatorname{int}(C^\infty)^*\,. \tag{2}$$

Then, for all $\beta \in (0, 1)$, there is a $\delta > 0$ such that if $u, v \in C$ with $\|z - u\| \leq \operatorname{dist}(z \mid C) + \delta$, then $\langle z - v\,,\, z - u\rangle > \beta\,\|z - u\|^2$. In particular, if $z = 0$, then $\langle v\,,\, u\rangle > \beta\,\|u\|^2$ whenever $u, v \in C$ and u satisfies $\|u\| \leq \operatorname{dist}(0 \mid C) + \delta$.

Proof Let $\beta \in (0, 1)$. If the result is false, there exists a sequence $\{(u^k, v^k)\} \subset C \times C$ with $\left\|z - u^k\right\| \leq \operatorname{dist}(z \mid C) + 1/k$ such that

$$\left\langle z - v^k\,,\, z - u^k\right\rangle \leq \beta\left\|z - u^k\right\|^2 \quad \forall\, k. \tag{3}$$

Since $\{u^k\}$ is bounded we can assume with no loss of generality that $u^k \to \operatorname{proj}_C(z)$. The projection theorem tells us that

$$\left\langle v - \operatorname{proj}_C(z)\,,\, z - \operatorname{proj}_C(z)\right\rangle \leq 0 \quad \forall\, v \in C,$$

or equivalently,

$$\operatorname{dist}(z \mid C)^2 \leq \left\langle z - v\,,\, z - \operatorname{proj}_C(z)\right\rangle \quad \forall\, v \in C. \tag{4}$$

If $\{v^i\}$ has a bounded subsequence, we can again assume with no loss in generality that $v^i \to \bar{v} \in C$. Then, by (3), $\left\langle z - \bar{v}\,,\, z - \operatorname{proj}_C(z)\right\rangle \leq \beta\operatorname{dist}(z \mid C)^2$ which contradicts (4) since $\beta \in (0, 1)$ and $\operatorname{dist}(z \mid C) > 0$. Hence, the sequence $\{v^i\}$ is divergent. Consequently, we can assume, with no loss in generality, that $v^i / \left\|v^i\right\| \to \bar{v} \in C^\infty$ with $\|\bar{v}\| = 1$. Dividing (3) by $\left\|v^i\right\|$ and taking the limit yields $\left\langle \bar{v}\,,\, z - \operatorname{proj}_C(z)\right\rangle \geq 0$. But $\bar{v} \in C^\infty$ and $z - \operatorname{proj}_C(z) \in \operatorname{int}(C^\infty)^*$, so, by Lemma 1, there is an $\epsilon > 0$ such that $\left\langle \bar{v}\,,\, z - \operatorname{proj}_C(z)\right\rangle \leq -\epsilon\,\|\bar{v}\|$. This contradiction establishes the result. □

Condition (2) plays a central role in our analysis of the GS algorithm. The following lemma gives insight into this condition by describing properties of the horizon cone C^∞ and its polar.

Lemma 3 (Normal, Barrier, and Horizon Cones) [8, Lemma 5] *Let C be a non-empty closed convex set and define $K := \bigcup_{x\in C} N_C(x)$. Then $\operatorname{ri}\operatorname{bar} C \subset K \subset \operatorname{bar} C$, and*

$$\operatorname{cl} K = \operatorname{cl}\operatorname{bar} C = (C^\infty)^* \;\text{ and }\; C^\infty = (\operatorname{bar} C)^*,$$

where $\operatorname{bar} C := \operatorname{dom}\left(\delta_C^*\right)$ is called the barrier cone of C.

Recall that it is always the case that $z - \text{proj}_C(z) \in N_C\left(\text{proj}_C(z)\right)$, and so, by Lemma 3, we have

$$z - \text{proj}_C(z) \in N_C\left(\text{proj}_C(z)\right) \subset \text{cl} \bigcup_{x \in C} N_C(x) = (C^\infty)^*.$$

In particular, if C is bounded, then $(C^\infty)^* = \mathbb{R}^n$ so condition (2) is trivially satisfied. Intuitively, the "smaller" the horizon cone of C the more "likely" condition (2) is satisfied.

2.3 Steepest Descent Directions

In the smooth setting the direction of steepest decent is given by the direction of unit length that minimizes the directional derivative. By contrast, in the nonsmooth setting there are several notions of directional derivative to choose from. From a numerical perspective, the most useful permit a dual representation as the support function of an associated subdifferential which in turn yields a dual representation of the direction of steepest descent via the Minimum Norm Duality Theorem, e.g. see [9, Theorem 2.8].

Since our analysis uses the Clarke subdifferential, our direction of steepest descent is based on the regular subderivative (see Theorem 1). That is, the direction of steepest descent for f at x is given by

$$\bar{d}_x := \arg \min_{\|x\| \leq 1} \hat{d} f(x)(d) \,. \tag{5}$$

The dual to this optimization problem is given by the Minimum Norm Duality Theorem.

Theorem 2 (Minimum Norm Duality Theorem) [13] *Let $\mathbb{X}$ be a normed linear space with norm $\|\cdot\|$ and dual norm $\|\cdot\|_*$, and let $\mathbb{B}$ denote the closed unit ball in $\mathbb{X}$. Given a nonempty closed convex set $C \subset \mathbb{X}^*$ and $\bar{z} \in \mathbb{X}^*$ with $\bar{z} \notin C$, we have*

$$dist_*(\bar{z} \mid C) = \sup_{\|v\| \leq 1} [\langle v \,, \bar{z} \rangle - \delta_C^*(v)], \tag{6}$$

where δ_C is the convex indicator of C and f^* denotes the convex conjugate of a function f. In particular, if $\bar{z} = 0$, then

$$\inf_{\|v\| \leq 1} \delta_C^*(v) = -dist_*(0 \mid C).$$

The Projection Theorem for convex sets tells us that for a nonempty closed convex set C and any $\bar{z} \in \mathbb{X}$ one has $\hat{z} = \text{proj}_C(\bar{z})$ if and only if $\bar{z} - \hat{z} \in N_C(\hat{z})$. This implies that $\bar{v} := \frac{(\bar{z} - \text{proj}_C(\bar{z}))}{\|\bar{z} - \text{proj}_C(\bar{z})\|}$ is the unique solution to the supremum problem in (6). By taking $C = \bar{\partial} f(x)$, we obtain a dual interpretation for the direction of steepest descent.

Corollary 1 (Steepest Descent Duality) [9, Theorem 2.8] *Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ and $x \in \text{dom}\,(\partial f)$ be such that f is lsc at x. Then*

$$\inf_{\|d\| \leq 1} \hat{d} f(x)(d) = -dist\left(0 \;\middle|\; \bar{\partial} f(x)\right),$$

and the vector $\bar{d}_x$ in (5) is given by $\bar{d}_x = -\text{proj}_{\bar{\partial} f(x)}(0) / \left\|\text{proj}_{\bar{\partial} f(x)}(0)\right\|$.

Proof By Theorem 1, $\bar{\partial} f(x)$ is a nonempty closed convex set with $\hat{d} f(x) = \delta^*_{\bar{\partial} f(x)}$. The corollary follows by taking $C = \bar{\partial} f(x)$ and $\bar{z} = 0$ in Theorem 2. □

2.4 Directionally Lipschitz Functions and Subdifferential Approximation

Rockafellar builds the notion of a directionally Lipschitzian function on that of epi-Lipschitzian sets [15]. He then establishes a useful characterization of directionally Lipschitzian functions through horizon subgradients [16]. We circumvent the epi-Lipschitzian construction and instead use the characterization given in [14, Exercise 9.42] as our definition.

Definition 1 (Directionally Lipschitzian Functions) Suppose $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is finite at $\bar{x} \in \mathbb{R}^n$. We say f is directionally Lipschitz at $\bar{x}$ if f is locally lsc at $\bar{x}$ and there is a unit vector u such that

$$\limsup_{\substack{x \xrightarrow{f} \bar{x} \\ v \to u \\ t \downarrow 0}} \frac{f(x+tv) - f(x)}{t} < \infty,$$

where the notation $x \xrightarrow[f]{} \bar{x}$ means that we consider only f-attentive convergence to $\bar{x}$, i.e., $x \to \bar{x}$ with $f(x) \to f(\bar{x})$. We say that f is directionally Lipschitz if it is directionally Lipschitz at every point of $\mathbb{R}^n$.

A simple characterization of directionally Lipschitz functions is obtained through the pointedness of the horizon cone of the subdifferential.

Lemma 4 [14, Exercise 9.42(b)] *A function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ finite at $x \in \mathbb{R}^n$ is directionally Lipschitz at x if and only if f is locally lsc at x and the horizon subdifferential $\partial^\infty f(x)$ is pointed.*

In particular, locally Lipschitz functions are directionally Lipschitz. In [14, Exercise 9.42(c)], Rockafellar and Wets show that a function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ that is finite and locally lsc at $\bar{x}$ is directionally Lipschitz at $\bar{x}$ if there is a convex cone $K \subset \mathbb{R}^n$ having nonempty interior such that f is K-nonincreasing, i.e. $f(x+w) \leq f(x)$ for all $x \in \mathbb{R}^n$ and $w \in K$. In [2, Theorem 6], it is shown that if $\operatorname{int} K \neq \emptyset$, then K-monotone functions (K-nonincreasing or K-nondecreasing) are continuous and almost everywhere differentiable. These authors also establish the following characterization of continuous directionally Lipschitz functions in terms of monotonicity.

Proposition 1 [2, Proposition 8] *A continuous function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is directionally Lipschitz at x if and only if it is locally representable near x as $f = g + l$ where g is monotone with respect to a convex cone with interior and l is linear.*

The pointedness of $\partial^\infty f(x)$, or equivalently, $\bar{\partial}^\infty f(x)$, is also related to the continuity of the regular subderivative $\hat{d} f(x)$.

Lemma 5 (Continuity of $\hat{d} f(x)$) *Suppose $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is finite at $x \in \mathbb{R}^n$ with $\partial f(\bar{x}) \neq \emptyset$, then $\hat{d} f(x)$ is continuous on*

$$\operatorname{int}\left[(\bar{\partial}^\infty f(\bar{x}))^*\right] = \operatorname{int}\left[\operatorname{dom}\left(\hat{d} f(\bar{x})(\cdot)\right)\right].$$

Proof By Theorem 1, Lemma 3 and the closure properties of convex sets, we have

$$\begin{aligned}\text{int}\,[(\bar{\partial}^\infty f(\bar{x}))^*] &= \text{int}\,[(\bar{\partial} f(\bar{x})^\infty)^*] = \text{int}\,[\text{cl bar}\,\bar{\partial} f(\bar{x})]\\ &= \text{int}\,[\text{bar}\,\bar{\partial} f(\bar{x})] = \text{int}\,[\text{dom}\left(\delta^*_{\bar{\partial} f(\bar{x})}\right)]\\ &= \text{int}\left[\text{dom}\left(\hat{d} f(\bar{x})(\cdot)\right)\right].\end{aligned}$$

Since $\hat{d} f(\bar{x})$ is convex, it is continuous on the interior of its domain. □

In general, directionally Lipschitzian functions need not be locally Lipschitz or even continuous at $\bar{x}$. For example, for every $\eta \geq 0$, the function

$$f(x) := \begin{cases} x^{1/3} - \eta, & x \leq 0,\\ x^{1/3} + \eta, & x > 0,\end{cases}$$

is directionally Lipschitz at $\bar{x} = 0$ and continuous at $\bar{x} = 0$ if and only if $\eta = 0$. Nonetheless, in [3, Corollary 6.1] it is shown that when $\partial^\infty f(x)$ is pointed, then $\bar{\partial} f(x)$ can be locally approximated by nearby gradients. We offer a slight improvement of this result that is useful to our discussion. We begin with the following technical lemma.

Lemma 6 (Limits of Gradients) *Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ and $\bar{x} \in \text{dom}\,(f)$ be such that, $\partial^\infty f(\bar{x})$ is pointed, f is continuous on an open neighborhood $\mathcal{V}$ of $\bar{x}$ and differentiable on an open set $\mathcal{Q} \subset \mathcal{V}$ of full measure in $\mathcal{V}$. For each $x \in \mathcal{V}$ and $\delta > 0$ such that $\bar{x} + \delta\mathbb{B} \subset \mathcal{V}$ set*

$$G_\delta(x) := clconv\,\nabla f((x + \delta I\!B) \cap \mathcal{Q}). \tag{7}$$

Let $x^k \to \bar{x}$ and $\delta_k \downarrow 0$ with $x^k + \delta_k\mathbb{B} \subset \mathcal{V}$ for all $k \in \mathbb{N}$.

(a) *If $w^k \to \bar{w}$ with $w^k \in G_{\delta_k}(x^k)$ for all $k \in \mathbb{N}$, then $\bar{w} \in \bar{\partial} f(\bar{x})$.*
(b) *If $v^k \to \bar{v}$ with $v^k \in G_{\delta_k}(x^k)^\infty$ for all $k \in \mathbb{N}$, then $\bar{v} \in \bar{\partial}^\infty f(\bar{x})$.*

Proof We only show (b) since the proof of (a) follows the same pattern but is significantly simpler. By Carathéodory's Theorem, for each $k \in \mathbb{N}$, there exist sequences $\{(x^{kj1}, \ldots, x^{kj(n+1)}) \mid j \in \mathbb{N}\} \subset \mathsf{X}_{i=1}^{n+1}\mathbb{R}^n$, $\{\alpha^{kj} \in \Delta_n \mid j \in \mathbb{N}\}$, and $\{t_{kj} \mid j \in \mathbb{N}\} \subset \mathbb{R}_+$ such that $t_{kj} \downarrow_j 0$, $x^{kji} \in (\bar{x} + \delta_k I\!B) \cap \mathcal{Q}$ $((j,i) \in \mathbb{N} \times \{1, 2, ..., n+1\})$, and $t_{kj}\sum_{i=1}^{n+1}\alpha_i^{kj}\nabla f(x^{kji}) \overset{j}{\to} v^k$. Choose $\epsilon_k \downarrow 0$. For each $k \in \mathbb{N}$, let $j_k \in \mathbb{N}$ be such that $t_{kj_k} < \epsilon_k$ and $\left\| v^k - t_{kj_k}\sum_{i=1}^{n+1}\alpha_i^{kj_k}\nabla f(x^{kj_ki})\right\| \leq \epsilon_k$. For each $k \in \mathbb{N}$, set $(\bar{x}^{k1}, \ldots, \bar{x}^{k(n+1)}) := (x^{kj_k1}, \ldots, x^{kj_k(n+1)})$, $\bar{\alpha}^k := \alpha^{kj_k}$ and $\bar{t}_k = t_{kj_k}$ so that $(\bar{x}^{k1}, \ldots, \bar{x}^{k(n+1)}) \to (\bar{x}, \ldots, \bar{x})$, $\bar{t}_k \downarrow 0$, and $\bar{t}_k\sum_{i=1}^{n+1}\bar{\alpha}_i^k\nabla f(\bar{x}^{ki}) \to \bar{v}$. By compactness, we can assume that $\bar{\alpha}^k \to \bar{\alpha} \in \Delta_n$.

Suppose the sequence $\{\bar{t}_k(\nabla f(\bar{x}^{k1}), \ldots, \nabla f(\bar{x}^{k(n+1)}))\}$ is unbounded. Let ν_k denote the norm of the kth member of this sequence. We can assume that $\nu_k \uparrow \infty$ since the sequence is unbounded. Then, with no loss in generality, there exists $(w^1, \ldots, w^{(n+1)})$ such that

$$(\tilde{t}_{k1}\nabla f(\bar{x}^{k1}), \ldots, \tilde{t}_{k(n+1)}\nabla f(\bar{x}^{k(n+1)})) \to (w^1, \ldots, w^{(n+1)}) \neq (0, \ldots, 0),$$

where $\tilde{t}_{ki} := (\bar{t}_k\bar{\alpha}_i^k)/\nu_k$ for $i = 1, \ldots, (n+1)$ and $k \in \mathbb{N}$. Since $\tilde{t}_{ki} \downarrow 0$ and $\nabla f(\bar{x}^{ki}) \in \hat{\partial} f(\bar{x}^{ki})$ for $i = 1, \ldots, n+1$ and $k \in \mathbb{N}$, we have $w^i \in \partial^\infty f(\bar{x})$ for $i = 1, \ldots, (n+1)$ (see (1)). We have

$$\sum_{i=1}^{n+1}\tilde{t}_k\nabla f(\bar{x}^{ki}) = \nu_k^{-1}\left(\bar{t}_k\sum_{i=1}^{n+1}\bar{\alpha}_i^k\nabla f(\bar{x}^{ki})\right) \to 0,$$

since $\bar{t}_k \sum_{i=1}^{n+1} \bar{\alpha}_i^k \nabla f(\bar{x}^{ki}) \to \bar{v}$ and $\nu_k \uparrow \infty$. But $\sum_{i=1}^{n+1} \tilde{t}_k \nabla f(\bar{x}^{ki}) \to \sum_{i=1}^{n+1} w^i$ by construction. Hence $0 = \sum_{i=1} w^i$ with $(w^1, \ldots, w^{(n+1)}) \neq (0, \ldots, 0)$ which contradicts the pointedness of $\partial^\infty f(\bar{x})$. Therefore, $\{\bar{t}_k(\nabla f(\bar{x}^{k1}), \ldots, \nabla f(\bar{x}^{k(n+1)}))\}$ is bounded, so we may assume that there exist $w^i \in \partial^\infty f(\bar{x})$ such that $\bar{t}_k \nabla f(\bar{x}^{ki}) \to w^i$ for $i = 1, \ldots, (n+1)$. Consequently, by Theorem 1, $v = \sum_{i=1}^{n+1} \bar{\alpha}_i w^i \in \operatorname{conv} \partial^\infty f(\bar{x}) = \bar{\partial}^\infty f(\bar{x})$ proving the result. □

The outer semi-continuity of ∂f and $\partial^\infty f$ along f-attentive sequences [14, Proposition 8.7] implies that pointedness is a local property and that the pointedness of $\partial^\infty f$ and G_δ are related.

Lemma 7 (Pointedness of $\partial^\infty f(x)$ is a local property) *Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ and $\bar{x} \in \operatorname{dom}(f)$ be such that $\partial^\infty f(\bar{x})$ is pointed, f is continuous on an open neighborhood $\mathcal{V}$ of $\bar{x}$ and differentiable on an open set $\mathcal{Q} \subset \mathcal{V}$ of full measure in $\mathcal{V}$. Let $G_\delta(x)$ be as in (7). Then the following statements hold.*

(i) *There is an $\epsilon > 0$ with $(\bar{x} + \epsilon\mathbb{B}) \subset \mathcal{V}$ such that $\partial^\infty f(x)$ is pointed on $(\bar{x} + \epsilon\mathbb{B})$.*
(ii) *There is a $\bar{\delta} > 0$ with $(\bar{x} + \bar{\delta}\mathbb{B}) \subset \mathcal{V}$ such that $G_\delta(\bar{x})^\infty$ is pointed for all $\delta \in (0, \bar{\delta}]$.*
(iii) *There exist $\epsilon,\ \bar{\delta} > 0$ with $(\bar{x} + \epsilon\mathbb{B}) \subset \mathcal{V}$ and $(x + \bar{\delta}\mathbb{B}) \subset \mathcal{V}$ for all $x \in (\bar{x} + \epsilon\mathbb{B})$ such that both $\partial^\infty f(x)$ and $G_\delta(x)^\infty$ are pointed for all $x \in \bar{x} + \epsilon\mathbb{B}$ and $0 < \delta < \bar{\delta}$.*

Proof The statements (i)-(iii) are proved in essentially the same manner. Therefore we only prove (iii). If the result is false, then there exist $\bar{\epsilon} > 0$ with $(\bar{x} + 2\bar{\epsilon}\mathbb{B}) \subset \mathcal{V}$ and sequences $\{x^k\} \subset (\bar{x} + \bar{\epsilon}\mathbb{B})$ and $\{\delta_k\} \subset (0, \bar{\epsilon})$ with $x^k \to \bar{x}$ and $\delta_k \downarrow 0$ such that either $\partial^\infty f(x^k)$ is not pointed for all $k = 1, 2, \ldots$ or $G_{\delta_k}(x^k)^\infty$ is not pointed for all $k = 1, 2, \ldots$. Let us first suppose that the cone $\partial^\infty f(x^k)$ is not pointed for all $k = 1, 2, \ldots$. Then there exist $v^{k1},\ v^{k2} \in \partial^\infty f(x^k)$ such that $v^{k1} + v^{k2} = 0$ and $\|v^{k1}\| + \|v^{k2}\| = 1$ for all k. Compactness and the osc of $\partial^\infty f$ at $\bar{x}$ (Theorem 1) tells us that we can also assume there exist $\bar{v}^1, \bar{v}^2 \in \partial^\infty f(\bar{x})$ with $(v^{k1}, v^{k2}) \to (\bar{v}^1, \bar{v}^2)$, $\bar{v}^1 + \bar{v}^2 = 0$ and $\|v^1\| + \|v^2\| = 1$. This contradicts the pointedness of $\partial^\infty f(\bar{x})$. Next suppose that the cone $G_{\delta_k}(x^k)^\infty$ is not pointed for all $k = 1, 2, \ldots$. Again, there exist $v^{k1},\ v^{k2} \in G_{\delta_k}(x^k)^\infty$ such that $v^{k1} + v^{k2} = 0$ and $\|v^{k1}\| + \|v^{k2}\| = 1$ for all $k \in \mathbb{N}$. Compactness tells us that we can assume there exist $\bar{v}^1, \bar{v}^2$ with $(v^{k1}, v^{k2}) \to (\bar{v}^1, \bar{v}^2)$, $\bar{v}^1 + \bar{v}^2 = 0$ and $\|v^1\| + \|v^2\| = 1$. But Lemma 6(b) tells us that $v^1, v^2 \in \bar{\partial}^\infty f(\bar{x})$ contradicting the pointedness of $\bar{\partial}^\infty f(\bar{x})$. □

In the next lemma we establish a relationship between regular subgradients and gradients at nearby points. The lemma extracts a portion of the proof of [3, Theorem 5.2] which we will use to extend [3, Corollary 6.1].

Lemma 8 (Gradients and Regular Subgradients) *Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be continuous on $B_\delta(\bar{x})$ for $\bar{x} \in \mathbb{R}^n$ and $\delta > 0$, and assume that $\mathcal{Q}$ is a full measure subset of $B_\delta(\bar{x})$ consisting of points where f is differentiable. If either f is absolutely continuous on line segments in $B_\delta(\bar{x})$ or $\mathcal{Q}$ is open and f is continuously differentiable on $\mathcal{Q}$, then $\hat{\partial} f(\bar{x}) \subset G_\delta(\bar{x})$.*

Proof If f is absolutely continuous on line segments in $B_\delta(\bar{x})$, the result is the first statement established in the proof of [3, Theorem 5.2]. If $\mathcal{Q}$ is open and f is continuously differentiable on $\mathcal{Q}$, the result requires only a very small change to this proof.

If $y \notin G_\delta(\bar{x})$, the separation theorem tells us that there exists a non-zero vector z and $k \in \mathbb{R}$ such that

$$\langle y\,,\,z\rangle > k \text{ but } \langle \nabla f(x)\,,\,z\rangle \leq k \;\forall\, x \in \mathcal{Q} \cap (\bar{x} + \delta\mathbb{B}).$$

If $y \in \hat{\partial} f(\bar{x})$, then $f(x+tz) \geq f(\bar{x}) + t\,\langle y\,,\,z\rangle + o(t)$. Let $\bar{t} > 0$ be such that $t\,\langle y\,,\,z\rangle + o(t) > kt$ for all $t \in (0, \bar{t}]$ so that $f(\bar{x} + tz) > f(\bar{x}) + k\bar{t}$ for all $t \in (0, \bar{t})$. By continuity, given $t \in (0, \bar{t})$, for all points w sufficiently close to $\bar{x}$, $f(w + \bar{t}z) > f(w) + k\bar{t}$. Hence, we can choose $\bar{w} \in \mathcal{Q}$ and $\hat{t} \in (0, \bar{t})$ so that

$$w + sz \in \mathcal{Q} \cap (\bar{x} + (\delta/2)\mathbb{B})\;\forall\, s \in (0, \hat{t}] \;\text{ with }\; f(w + \hat{t}z) > f(w) + k\hat{t}. \tag{8}$$

Now consider the function $g : [0, \hat{t}] \to \mathbb{R}$ defined by $g(s) := f(w+sz)$. By construction g is continuously differentiable on $(0, \hat{t})$ with $g'(s) = \langle \nabla f(w + sz)\,,\,z\rangle \leq k$. Therefore, by the Fundamental Theorem of Calculus, $f(w + \hat{t}z) = g(\hat{t}) \leq g(0) + k\hat{t} = f(w) + k\hat{t}$ which contradicts (8). □

Theorem 3 (Subdifferential Approximation) *Suppose that, close to $\bar{x} \in \mathbb{R}^n$, the function $f : \mathbb{R}^n \to \mathbb{R}$ is continuous and absolutely continuous on line segments, with $\partial^\infty f(\bar{x})$ pointed. If $\mathcal{Q}$ is a full measure subset of a neighborhood of $\bar{x}$ consisting of points where f is differentiable, then*

$$\bar{\partial} f(\bar{x}) = \bigcap_{\delta>0} G_\delta(\bar{x}) \quad \textit{and} \quad \bar{\partial}^\infty f(\bar{x}) = \bigcap_{\delta>0} G_\delta(\bar{x})^\infty.$$

Moreover, if $\mathcal{Q}$ is open with f continuously differentiable on $\mathcal{Q}$, then the requirement that f be absolutely continuous on line segments can be dropped.

Proof The statement of the theorem differs in two respects from the result given in [3, Corollary 6.1]. First, the result in [3] makes no mention of the case when $\mathcal{Q}$ is open, and, second, there is no formula for the horizon cone equivalence. The case when $\mathcal{Q}$ is open follows from Lemma 8 since the lemma tells us that the implication in [3, Theorem 5.2] follows from this hypothesis. Consequently, [3, Corollary 6.1] follows from this hypothesis as well.

We now prove the horizon cone equivalence. By Lemma 8, for all small $\delta > 0$, $\hat{\partial} f(x) \subset G_{\delta/2}(x) \subset G_\delta(\bar{x})$ for all $x \in B_{\delta/2}(\bar{x})$. Hence $\partial^\infty f(\bar{x}) \subset G_\delta(\bar{x})^\infty$, and so, by Theorem 1(4), $\bar{\partial}^\infty f(\bar{x}) \subset G_\delta(\bar{x})^\infty$ for all small $\delta > 0$. Consequently, $\bar{\partial}^\infty f(\bar{x}) \subset \bigcap_{\delta>0} G_\delta(\bar{x})^\infty$. For the reverse inclusion let $v \in \bigcap_{\delta>0} G_\delta(\bar{x})^\infty$. Then there exist sequences $\delta_k \downarrow 0$ and $v^k \to v$ such that $v^k \in G_{\delta_k}(\bar{x})^\infty$ for all $k \in \mathbb{N}$. By Lemma 6(b), $v \in \bar{\partial}^\infty f(\bar{x})$ which proves the result. □

The next lemma establishes a key property of approximate directions of steepest descent for directionally Lipschitz functions and extends the content of [11, Lemma 3.1] to these functions.

Lemma 9 (Approximate Directions of Steepest Descent) *Let $\bar{x} \in \mathbb{R}^n$ be such that $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is differentiable on a full measure subset $\mathcal{Q}$ of an open convex neighborhood $\mathcal{N}$ of $\bar{x}$. Further suppose that f is either continuous and absolutely continuous along line segments in $\mathcal{N}$ or that $\mathcal{Q}$ is open. If*

$$0 \notin \bar{\partial} f(\bar{x}),\; \emptyset \neq \bar{\partial} f(\bar{x}),\; \textit{and}\; -\operatorname{proj}_{\bar{\partial} f(\bar{x})}(0) \in \operatorname{int}(\bar{\partial}^\infty f(\bar{x}))^*, \tag{9}$$

then, for all $\beta \in (0,1)$*, there exists* $\delta > 0$ *and* $\eta > 0$ *such that* $0 \notin G_\eta(\bar{x})$ *and, for every* $u, v \in G_\eta(\bar{x})$ *with* $\|u\| \leq dist\left(0 \,\middle|\, G_\eta(\bar{x})\right) + \delta$*, we have* $\langle v, u\rangle > \beta \|u\|^2$.

Proof Since $0 \notin \bar{\partial} f(\bar{x})$, Theorem 3 tells us that there is an $\bar{\eta} > 0$ such that $0 \notin G_\eta(\bar{x})$ for all $\eta \in (0, \bar{\eta}]$. Theorem 3 also tells us that $\bar{\partial} f(\bar{x}) \subset G_\eta(\bar{x})$ for all $\eta \geq 0$. Therefore, $\text{dist}\left(0 \,\middle|\, G_\eta(\bar{x})\right) \leq \text{dist}\left(0 \,\middle|\, \bar{\partial} f(\bar{x})\right) < \infty$ for all $\eta \geq 0$.

We suppose the result is false and establish a contradiction. Since the result is false, there exist $\hat{\beta} \in (0,1)$ and sequences $\{(u^i, v^i)\}$ in $\mathbb{R}^{2n}$ and $\{(\eta_i, \delta_i)\}$ in $\mathbb{R}^2_+$ with $\eta_i \downarrow 0$ and $\delta_i \downarrow 0$ such that, for all $i \in \mathbb{N}$,

$$u^i, v^i \in G_{\eta_i}(\bar{x}),\ \left\|u^i\right\| \leq \text{dist}\left(0 \,\middle|\, G_{\eta_i}(\bar{x})\right) + \delta_i \text{ and } \left\langle v^i, u^i \right\rangle \leq \hat{\beta} \left\|u^i\right\|^2. \tag{10}$$

Let $\delta_0 \geq \delta_1$. Since $\{\|u^i\|\}$ is bounded by $\text{dist}\left(0 \,\middle|\, \bar{\partial} f(\bar{x})\right) + \delta_0$, we may assume that $u^i \to \bar{u}$, where $\bar{u} \in \bar{\partial} f(\bar{x})$ by Lemma 6. Theorem 3 tells us that $\bar{\partial} f(\bar{x}) \subset G_{\eta_i}(\bar{x})$ for all $i \in \mathbb{N}$. Hence, for all i large, $\|u^i\| \leq \text{dist}\left(0 \,\middle|\, \bar{\partial} f(\bar{x})\right) + \delta_i$. Therefore $\bar{u} = \text{proj}_{\bar{\partial} f(\bar{x})}(0)$ and so $-\bar{u} \in \text{int}\,(\bar{\partial}^\infty f(\bar{x}))^*$ by (9).

Next consider the sequence $\{v^i\}$. If this sequence is bounded, then, again, Lemma 6 tells us that, with no loss in generality, there is a $\bar{v} \in \bar{\partial} f(\bar{x})$ such that $v^i \to \bar{v}$. The projection theorem for convex sets tells us that $\langle \bar{v}, \bar{u}\rangle \geq \|\bar{u}\|^2$, but, by construction, $\langle \bar{v}, \bar{u}\rangle \leq \hat{\beta}\|\bar{u}\|^2 < \|\bar{u}\|^2$. This contradiction implies that the sequence $\{v^i\}$ is unbounded. Therefore, with no loss in generality, $\{v^i\}$ is divergent. By Carathéodory's Theorem, there exists $\lambda^i \in \Delta_n$ and $x^{ij} \in G_{\eta_i}(\bar{x})$ such that $v^i = \sum_{j=1}^{n+1} \lambda_{ij} \nabla f(x^{ij})$ for all i. Since the sequence $\{v^i\}$ is divergent, the sequence defined by $\hat{g}_i := (\lambda_{i1}\nabla f(x^{i1}), \ldots, \lambda_{i(n+1)}\nabla f(x^{i(n+1)}))$ must also be divergent, and so, again with no loss in generality, there is a $(\bar{g}^1, \ldots, \bar{g}^{n+1})$ such that $\hat{g}_i / \|\hat{g}_i\| \to (\bar{g}^1, \ldots, \bar{g}^{n+1}) \neq 0$, where we have taken $\|\hat{g}_i\| := \max_{j=1,\ldots,n+1} \lambda_{ij} \|\nabla f(x^{ij})\|$. Theorem 1 tells us that $\bar{g}^j \in \partial^\infty f(\bar{x})$, $j = 1, \ldots, n+1$. Clearly, $\|v^i\| \leq \|\hat{g}_i\|$ for all $i = 1, 2, \ldots$. If $\{v^i / \|\hat{g}_i\|\}$ has a subsequence convergent to zero, then taking the limit along this subsequence yields $\sum_{j=1}^{n+1} \bar{g}^j = 0$ which contradicts the fact that $\partial^\infty f(\bar{x})$ is pointed. So we can assume that $v^i / \|\hat{g}_i\| = \sum_{j=1}^{n+1} \lambda_j^i \nabla f(x^{ij}) / \|\hat{g}_i\| \to \tilde{v} \in \partial^\infty f(\bar{x}) \setminus \{0\}$. Theorem 1 tells us that $\tilde{v} \in \bar{\partial}^\infty f(\bar{x}) = \bar{\partial} f(\bar{x})^\infty$. But $-\bar{u} \in \text{int}\,(\bar{\partial} f(\bar{x})^\infty)^*$, so, by Lemma 1, $\langle \tilde{v}, \bar{u}\rangle > 0$ while $\langle \tilde{v}, \bar{u}\rangle \leq 0$ by (10). This final contradiction establishes the result. □

The condition $-\text{proj}_{\bar{\partial} f(\bar{x})}(0) \in \text{int}\,(\bar{\partial}^\infty f(\bar{x}))^*$ in (9) plays an important role in our analysis. Although examples where it fails to hold are easily generated, such points are *degenerate* in the sense that the direction of steepest descent for the regular subderivative does not lie in the interior of its domain (see Lemma 5). Further discussion of this issue is given in our concluding remarks.

Example 1 Let $h : \mathbb{R}^2 \to \mathbb{R}$ be given by $h(x) := \langle y, x\rangle + [\text{dist}\left(x \,\middle|\, \mathbb{R}^2_+\right)]^{1/2}$, where $y := (-1, \beta)^T$. Then $\bar{\partial} h(0) = y + \mathbb{R}^2_-$, $(\bar{\partial}^\infty h(0))^* = \mathbb{R}^2_+$ and

$$-\text{proj}_{\bar{\partial} f(\bar{x})}(0) = \begin{cases} (1,0)^T \notin \text{int}\,(\bar{\partial}^\infty f(0))^* & , \ \beta \geq 0, \\ (1, -\beta) \in \text{int}\,(\bar{\partial}^\infty f(0))^* & , \ \beta < 0. \end{cases}$$

3 The Gradient Sampling Algorithm

Assume that $f : \mathbb{R}^n \to \mathbb{R}$ satisfies the following hypothesis:

$\mathcal{H}$: f is continuous on $\mathbb{R}^n$ and continuously differentiable on an open full measure set $\mathcal{D} \subset \mathbb{R}^n$.

We use the form of the gradient sampling algorithm given in [7] which is based on the version proposed by Kiwiel in [11].

The GS Algorithm (Gradient sampling algorithm).

Initialization: Let $x^0 \in \widehat{\mathcal{D}}$ the set of points where f is differentiable. Choose termination tolerances $(\epsilon_{opt}, \nu_{opt}) \in [0, \infty) \times [0, \infty)$ and the initial sampling radius $\epsilon_0 \in (\epsilon_{opt}, \infty)$, initial stationarity target $\nu_0 \in [\nu_{opt}, \infty)$, sample size $m \geq n + 1$, line search parameters $(\beta, \gamma) \in (0, 1) \times (0, 1)$, and reduction factors $(\theta_\epsilon, \theta_\nu) \in (0, 1] \times (0, 1]$.

For $k \in \mathbb{N}$ **do**
(i) Independently sample $\{x^{k,1}, \ldots, x^{k,m}\}$ uniformly from $x^k + \epsilon_k \mathbb{B}$.
(ii) Terminate the algorithm if $\{x^{k,1}, \ldots, x^{k,m}\} \not\subset \widehat{\mathcal{D}}$.
(iii) Compute g^k as the solution of $\min_{g \in \mathcal{G}^k} \frac{1}{2}\|g\|^2$, where
$\mathcal{G}^k := \operatorname{conv}\{\nabla f(x^k), \nabla f(x^{k,1}), \ldots, \nabla f(x^{k,m})\}$.
(iv) **If** $\nabla f(x^k) = 0$ or ($\|g^k\|_2 \leq \nu_{opt}$ and $\epsilon_k \leq \epsilon_{opt}$), **then** terminate.
(v) **If** $\|g^k\|_2 \leq \nu_k$
(vi) **then** set $\nu_{k+1} \leftarrow \theta_\nu \nu_k$, $\epsilon_{k+1} \leftarrow \theta_\epsilon \epsilon_k$, and $t_k \leftarrow 0$
(vii) **else** set $\nu_{k+1} \leftarrow \nu_k$, $\epsilon_{k+1} \leftarrow \epsilon_k$, $d^k \leftarrow -g^k / \|g^k\|$, and
$t_k \leftarrow \max\left\{t \in \{1, \gamma, \gamma^2, \ldots\} : f(x^k + td^k) < f(x^k) - \beta t \|g^k\|\right\}$. (11)
(viii) **If** f is differentiable at $x^k + t_k d^k$
(ix) **then** set $x^{k+1} \leftarrow x^k + t_k d^k$
(ix) **else** set x^{k+1} randomly as any point where f is differentiable and such that
$$f(x^{k+1}) < f(x^k) - \beta t_k \|g^k\| \quad \text{and}$$
$$\|x^k + t_k d^k - x^{k+1}\|_2 \leq \min\{t_k, \epsilon_k\}$$
End for

Remark 1 As shown in [6, Page 756], the line search (11) in the algorithm is finitely terminating when $\nabla f(x^k) \neq 0$.

Remark 2 In [11, Section 4.1] it is observed that one can also take d^k to be the unnormalized direction $-g^k$ when f is Lipschitz. However, the argument in [11] explicitly depends on f being Lipschitz continuous. In the non-Lipschitzian case, our proof of convergence requires the normalized direction in the statement of the GS algorithm given above.

Remark 3 The algorithm terminates in Step (ii) if $\{x^{k,1}, \ldots, x^{k,m}\} \not\subset \widehat{\mathcal{D}}$. But the reader should note that the hypotheses imply that $\{x^{k,1}, \ldots, x^{k,m}\} \subset \mathcal{D}$ for all k with probability 1 since the countable intersection of probability one events has probability one.

3.1 Convergence

We now introduce the key tools in analyzing the GS algorithm introduced in [6]: for $\epsilon, \delta > 0$ and $\bar{x}, x \in \mathbb{R}^n$, let

$$\rho_\epsilon(x) := \operatorname{dist}(0 \mid G_\epsilon(x))$$

and set

$$\mathcal{D}_\epsilon^m(x) := \prod_1^m ((x + \epsilon\mathbb{B}) \cap \mathcal{D}) \subset \mathbb{R}^n \qquad \text{and}$$

$$V_\epsilon(\bar{x}, x, \delta) := \left\{ (y^1, y^2, \ldots, y^m) \in \mathcal{D}_\epsilon^m(x) \,\middle|\, \operatorname{dist}(0 \mid \operatorname{conv}\{\nabla f(y^i)\}_{i=1}^m) \le \rho_\epsilon(\bar{x}) + \delta \right\},$$

where $m \ge n + 1$ is as given in the statement of Algorithm I.

The next lemma shows that the convex hull of a collection of gradients can be used to obtain directions of approximate steepest descent.

Lemma 10 [6, Lemma 3.2(i)] [11, Lemma 3.2(i)] *Let $\epsilon > 0$ and $\bar{x} \in \mathbb{R}^n$. For all $\delta > 0$ there is a $\tau > 0$ and a non-empty open set $\overline{V}$ such that $\overline{V} \subset V_\epsilon(\bar{x}, x, \delta)$ for all $x \in B_\tau(\bar{x})$ with $\operatorname{dist}\left(0 \,\middle|\, \operatorname{conv}\{\nabla f(y^i)\}_{i=1}^m\right) \le \rho_\epsilon(\bar{x}) + \delta$ for all $(y^1, \ldots, y^m) \in \overline{V}$.*

Remark 4 The statement of this lemma parallels the form given in [11, Lemma 3.2(i)] rather than the form given in [6, Lemma 3.2(i)]. Essentially the same proof is given in both papers and follows from the continuity of ∇f on $\mathcal{D}$.

We make use of the following mean value theorem to provide a lower bound on the step sizes t_k in step (vi) of the GS algorithm when $0 \notin \partial f(x)$.

Theorem 4 (Approximate Mean Value Theorem) [1, Theorem 3.4.7]
Let $\varphi : \mathbb{R}^n \to \mathbb{R}$ be lsc and assume that $r \in \mathbb{R}$ and $x, y \in \mathbb{R}^n$ are such that $x \ne y$, $\varphi(x) < +\infty$, and $r < \varphi(y) - \varphi(x)$. Then there is a $\hat{x} \in [x, y)$ such that for all $\epsilon > 0$ there exists $(\tilde{x}, \varphi(\tilde{x})) \in B_\epsilon((\hat{x}, \varphi(\hat{x})))$ and $\tilde{v} \in \hat{\partial}\varphi(\tilde{x})$ for which

$$\left\langle \tilde{v}, \hat{x} - \tilde{x} \right\rangle > -\epsilon, \quad \langle \tilde{v}, y - x \rangle > r, \quad \text{and} \quad \varphi(\tilde{x}) \le \varphi(x) + |r| + \epsilon.$$

Lemma 11 (Stepsize Bound) *Let $f : \mathbb{R}^n \to \mathbb{R}$ be such that $\mathcal{H}$ holds. Let $\beta, \gamma \in (0, 1)$ be given, and let $\bar{x} \in \mathbb{R}^n$ be such that all three conditions in (9) hold. Then there exist $\eta > 0$ and $\delta > 0$ so that the consequences of Lemma 9 hold. Moreover, given $\epsilon > 0$, we can choose $\tau \in (0, \epsilon/3)$ so that the consequences of Lemma 10 hold for this δ. That is, there exists a non-empty open set $\overline{V}$ such that $\overline{V} \subset V_\epsilon(\bar{x}, x, \delta)$ for all $x \in B_\tau(\bar{x})$ with $\operatorname{dist}\left(0 \,\middle|\, \operatorname{conv}\{\nabla f(y^i)\}_{i=1}^m\right) \le \rho_\epsilon(\bar{x}) + \delta$ for all $(y^1, \ldots, y^m) \in \overline{V}$. Then, for all $x \in B_\tau(\bar{x})$ and $(x^1, \ldots, x^m) \in \overline{V}$,*

$$\bar{t} := \min\{1, \gamma\epsilon/3\} \le \hat{t} := \max\left\{ t \,\middle|\, \begin{array}{l} f(x + td) < f(x) - \beta t \|g\| \\ t \in \{1, \gamma, \gamma^2, \ldots\} \end{array} \right\}, \tag{12}$$

where $g := \arg\min\left\{ \|v\| \,\middle|\, v \in \operatorname{conv}\{\nabla f(x^i)\}_{i=1}^m \right\}$ and $d := -g/\|g\|$.

Proof Since the hypotheses of Lemmas 9 and 10 are satisfied, the parameters η, δ and τ can be chosen as required. Set $\widehat{G} := \operatorname{conv}\{\nabla f(x^i)\}_{i=1}^m$. Since $(x^1, \ldots, x^m) \in \overline{V} \subset V_\epsilon(\bar{x}, \bar{x}, \delta)$,

Lemma 10 tells us that $\operatorname{dist}\left(0 \,\middle|\, \widehat{G}\right) \leq \rho_\epsilon(\bar{x}) + \delta$ and $\widehat{G} \subset G_\epsilon(\bar{x})$. Hence, $g \in G_\epsilon(\bar{x})$ and $\|g\| \leq \rho_\epsilon(\bar{x}) + \delta$. Consequently, Lemma 9 tells us that

$$\langle v\,,\, g\rangle > \beta \|g\|^2 \quad \forall\, v \in G_\epsilon(\bar{x}). \tag{13}$$

Assume to the contrary that the inequality (12) is false. Then $\hat{t} < 1$ and so

$$-\beta\gamma^{-1}\hat{t}\,\|g\| \leq f(x + \gamma^{-1}\hat{t}d) - f(x).$$

By taking $f = \varphi$, $x = x$, $y = x + \gamma^{-1}\hat{t}d$ and $r = -\beta\gamma^{-1}\hat{t}\,\|g\|$ in Theorem 4, there exists $\hat{x} \in [x,\ x + \gamma^{-1}\hat{t}d)$ such that for all $\tilde{\epsilon} > 0$ there exists $(\tilde{x},\, f(\tilde{x})) \in B_\epsilon(\hat{x},\, f(\hat{x}))$ and $\tilde{v} \in \hat{\partial} f(\tilde{x})$ such that

$$-\beta\gamma^{-1}\hat{t}\,\|g\| < \gamma^{-1}\hat{t}\,\langle \tilde{v}\,,\, d\rangle\,,$$

or equivalently,

$$\langle \tilde{v}\,,\, g\rangle < \beta\,\|g\|^2\,.$$

So $\tilde{v} \notin G_\epsilon(\bar{x})$ by (13). Assume that we have chosen $\tilde{\epsilon} \in (0, \epsilon/3)$. Since the inequality (12) is false, $\hat{t} < \gamma\epsilon/3$ or equivalently, $\gamma^{-1}\hat{t}\,\|d\| < \epsilon/3$. Consequently, $\tilde{x} \in B_\epsilon(\bar{x})$ and $\partial f(\tilde{x}) \subset G_\epsilon(\bar{x})$. Therefore, $\tilde{v} \in \hat{\partial} f(\tilde{x}) \subset \partial f(\tilde{x}) \subset G_\epsilon(\bar{x})$. This contradiction establishes the result. □

The main convergence result for the GS Algorithm now follows. Our proof is inspired by Kiwiel's proof of [11, Theorem 3.3].

Theorem 5 (Convergence: $0 = v_{opt} = \epsilon_{opt}$) *Suppose $f : \mathbb{R}^n \to \mathbb{R}$ satisfies $\mathcal{H}$. Let $\{x^k\}$ be a sequence generated by the GS Algorithm with $v_0,\ \epsilon_0 \in \mathbb{R}_{++}$, $\theta_\epsilon,\ \theta_v \in (0, 1)$, and $\epsilon_{opt} = v_{opt} = 0$. With probability 1 the algorithm does not terminate in line (ii) and one of the following must occur:*

(a) *There is a $k_0 \in \mathbb{N}$ such that $\nabla f(x^{k_0}) = 0$ and the algorithm terminates.*
(b) $f(x^k) \downarrow -\infty$.
(c) $0 < \bar{v} := \inf_k v_k$ *and the sequence converges to some $\bar{x} \in \mathbb{R}^n$ for which at least one of the three conditions in (9) must be violated, that is, either*

$$\emptyset = \bar{\partial} f(\bar{x}), \quad 0 \in \bar{\partial} f(\bar{x}), \quad \text{or } -\operatorname{proj}_{\bar{\partial} f(\bar{x})}(0) \notin \operatorname{int}(\bar{\partial}^\infty f(\bar{x}))^*. \tag{14}$$

(d) $v_k \downarrow 0$ *and every cluster point $\bar{x}$ of $\{x^k\}$ (if one exists) satisfies $0 \in \bar{\partial} f(\bar{x})$.*

Moreover, if f is locally Lipschitz, then outcome (c) cannot occur.

Proof If f is locally Lipschitz, then the result follows from [11, Theorem 3.3]. The assumptions on the function f imply that, with probability 1, the algorithm does not terminate in line (ii) of the GS Algorithm and we can assume $\{x^{k,1}, \ldots, x^{k,m}\} \subset \mathcal{D}$ for all k (see Remark 3). We also assume that neither (a) nor (b) occurs and show that either (c) or (d) must occur. Let $J \subset \mathbb{N}$ be those iterations for which $x^k \neq x^{k+1}$. Observe that if J is finite with maximum value k_0, then $\nabla f(x^{k_0}) = 0$, hence, J is infinite. Since

$$f(x^{k+1}) \leq f(x^k) - \beta t_k \left\|g^k\right\| \quad \forall\, k \in \mathbb{N},$$

the sequence $\{f(x^k)\}$ is non-increasing and bounded below, and so has a limit $\tilde{f}$. Summing this inequality over k and taking the limit tells us that

$$\beta \sum_{k=1}^{\infty} \left\|x^{k+1} - x^k\right\| \left\|g^k\right\| \leq \beta \sum_{k=1}^{\infty} t_k \left\|g^k\right\| \leq f(x^0) - \tilde{f} < \infty, \tag{15}$$

where the first inequality follows from lines (viii)-(x) of the GS algorithm. In particular, $\|x^{k+1} - x^k\| \|g^k\| \to 0$. We decompose this fact into two mutually exclusive possibilities: either $0 < \bar{\nu} := \inf_k \nu_k$ or $\nu_k \downarrow 0$.

Let us first suppose that $0 < \bar{\nu} := \inf_k \nu_k$. By lines (v) and (vi) of the algorithm, $0 < \bar{\epsilon} := \inf_k \epsilon_k$ and $\bar{\nu} \leq \inf_{k \in J} \|g^k\|$. Therefore, (15) tells us that $\sum_{k=1}^{\infty} \|x^{k+1} - x^k\| < \infty$ and $t_k \downarrow 0$. In particular, this implies that the sequence $\{x^k\}$ is Cauchy, and so there exists $\bar{x}$ such that $x^k \to \bar{x}$. Assume to the contrary that none of the conditions in (14) holds, or equivalently, the hypotheses of Lemma 11 (9) hold at $\bar{x}$. Let $\epsilon,\ \eta,\ \delta,\ \tau \in \mathbb{R}_{++}$ and $\overline{V} \subset \mathbb{R}^n$ be an open set satisfying the conditions of Lemma 11. We may assume that $\tau < \inf_k \epsilon_k$. Since for all k sufficiently large $t_k < \gamma\epsilon/3$, we must have $(x^{k1}, \ldots, x^{km}) \notin \overline{V}$ for all large k. But since $\overline{V}$ is open, the probability of this event is zero. Hence, with probability 1, Lemma 11 tells us that at least one of the three conditions in (14) must hold, that is, (c) is satisfied.

Finally, suppose that $\nu_k \downarrow 0$. By line (vi) of the algorithm, $\epsilon_k \downarrow 0$. Let $\bar{x}$ be a cluster point of the sequence $\{x^k\}$. If there is any subsequence $\widehat{J} \subset \mathbb{N}$ such that

$$x^k \overset{\widehat{J}}{\to} \bar{x} \quad \text{and} \quad \left\|g^k\right\| \overset{\widehat{J}}{\to} 0, \tag{16}$$

then $0 \in \bar{\partial} f(\bar{x})$ by Lemma 6. Therefore, we assume that no such subsequence exists and establish a contradiction. In particular, this implies that $x^k \not\to \bar{x}$. Since no subsequence satisfies (16), there exist $\bar{\nu} > 0$ such that if $\|x^k - \bar{x}\| \leq \bar{\nu}$, then $\|g^k\| > \bar{\nu}$; otherwise, there exists $\widehat{J} \subset \mathbb{N}$ and $\bar{\nu}_k \downarrow_{\widehat{J}} 0$ and such that $\|x^k - \bar{x}\| \leq \bar{\nu}_k$ and $\|g^k\| \leq \bar{\nu}_k$ for all $k \in \widehat{J}$ which implies that $\widehat{J}$ satisfies (16), a contradiction. Since $\bar{x}$ is a cluster point, the set $K := \left\{k \mid \|x^k - \bar{x}\| \leq \bar{\nu}\right\}$ is infinite with $\|g^k\| > \bar{\nu}$ for all $k \in K$. Since $x^k \not\to \bar{x}$, we can reduce $\bar{\nu}$ if necessary so that the set $\mathbb{N} \setminus K$ is infinite. Observe that inequality (15) tell us that $\sum_{k \in K} \|x^k - \bar{x}\| < \infty$. Let $\widehat{K} := \left\{k \mid \|x^k - \bar{x}\| \leq \bar{\nu}/3\right\}$. Again $\widehat{K}$ is infinite since $\bar{x}$ is a cluster point of $\{x^k\}$. Both K and $\widehat{K} \subset K$ define subsequences of $\{x^k\}$. Since $\mathbb{N} \setminus K$ is infinite, for each $k \in \widehat{K}$ there is a $\hat{k} > k$ such that $\hat{k} \notin K$ but $x^i \in K$ for $k \leq i < \hat{k}$. By construction, $\left\|x^{\hat{k}} - x^k\right\| \geq \bar{\nu}/3$ for all $k \in \widehat{K}$; otherwise, $x^{\hat{k}} \in K$, a contradiction. By the triangle inequality, we have $\bar{\nu}/3 \leq \left\|x^{\hat{k}} - x^k\right\| \leq \sum_{i=k}^{\hat{k}-1} \|x^{i+1} - x^i\|$ for all $k \in \widehat{K}$. But $\sum_{k \in K} \|x^k - \bar{x}\| < \infty$ and $\widehat{K} \subset K$ so that $\sum_{i=k}^{\hat{k}-1} \|x^{i+1} - x^i\| \overset{\widehat{K}}{\to} 0$. This contradiction implies that our assumption that there is no subsequence satisfying (16) is false. That is, $0 \in \bar{\partial} f(\bar{x})$. □

In the Lipschitzian case, Theorem 5 differs from [11, Theorem 3.3] with the introduction of possible outcome (c). Kiwiel's proof of [11, Theorem 3.3] shows that the case $0 < \bar{\nu} := \inf_k \nu_k$ does not occur if f is locally Lipschitz continuous. The absence of the case (c) requires that $\bar{\partial} f$ is an osc, compact, convex valued operator whose domain is all of $\mathbb{R}^n$, in particular, it requires that f be locally Lipschitz. On the other hand, if f is not locally Lipschitz, then $\bar{\partial} f$ is not locally bounded and possibly empty at some points. These possibilities are reflected in the outcome (c), and only in (c). This does not imply that $\bar{\partial} f(\bar{x})$ is bounded in outcome (d), but outcome (d) does require that $\bar{\partial} f(\bar{x})$ be nonempty. Note that outcome (c) signals why ν_k is not reduced to zero. These observations are reviewed in our final comments. We conclude this section by stating two corollaries that describe the behavior of the algorithm under standard variations in the choice of of initial parameters.

Corollary 2 (Convergence: $0 < \epsilon_{opt},\ 0 < \nu_{opt}$) *Suppose $f : \mathbb{R}^n \to \mathbb{R}$ satisfies $\mathcal{H}$. Let $\{x^k\}$ be a sequence generated by the GS Algorithm with $\nu_0,\ \epsilon_0 \in \mathbb{R}_{++}$, $\theta_\epsilon,\ \theta_\nu \in (0, 1)$ and*

$0 < \epsilon_{opt},\ 0 < \nu_{opt}$. *With probability 1 the algorithm does not terminate in line (ii) and one of the following must occur:*

(a) *There is a* $k_0 \in \mathbb{N}$ *such that dist* $\left(0 \mid \bar{\partial}_{\epsilon_{opt}} f(x^{k_0})\right) \leq \nu_{opt}$ *and the algorithm terminates.*
(b) $f(x^k) \downarrow -\infty$.
(c) $\nu_{opt} < \bar{\nu} := \inf_k \nu_k$ *and the sequence converges to some* $\bar{x} \in \mathbb{R}^n$ *at which at least one of the three statements in* (14) *is true.*

Proof By assumption the algorithm does not terminate in line (ii) of the GS Algorithm with probability 1 and we can assume $\{x^{k,1}, \ldots, x^{k,m}\} \subset \mathcal{D}$ for all k (see Remark 3). Next we assume that neither (a) nor (b) occur and show that (c) must occur. Since (a) does not occur and $\nabla f(x^k) \in \bar{\partial}_{\epsilon_k} f(x^k)$, the algorithm does not terminate in step (iv) and step (v) of the algorithm occurs at most finitely many times. Therefore, $\nu_{opt} < \bar{\nu} := \inf_k \nu_k$, the algorithm does not terminate and the sequence $\{x^k\}$ is infinite. Consequently, Theorem 5 tells us that (c) must occur and the final statement of the corollary follows. □

Corollary 3 (Convergence: $0 < \epsilon_{opt} = \epsilon_0,\ 0 = \nu_{opt} = \nu_0$) *Suppose* $f : \mathbb{R}^n \to \mathbb{R}$ *satisfies* $\mathcal{H}$. *Let* $\{x^k\}$ *be a sequence generated by the GS Algorithm with* $\nu_{opt} = \nu_0 = 0$, $\epsilon_{opt} = \epsilon_0 > 0$ *and* $0 = \theta_\nu,\ 1 = \theta_\epsilon$. *Let* $J \subset \mathbb{N}$ *be those iterations for which* $x^k \neq x^{k+1}$. *With probability 1 the algorithm does not terminate in line (ii) and one of the following must occur:*

(a) *The algorithm terminates at some iteration* $k_0 \in \mathbb{N}$ *with either* $\nabla f(x^{k_0}) = 0$ *or* $g^{k_0} = 0$, *and consequently* $0 \in \bar{\partial}_{\epsilon_{opt}} f(x^{k_0})$.
(b) $f(x^k) \downarrow -\infty$.
(c) *The sequence* $\{x^k\}$ *is infinite with* $\inf_{k \in J} \left\|g^k\right\| > 0$ *in which case there exists* $\bar{x} \in \mathbb{R}^n$ *such that* $x^k \to \bar{x}$ *and at least one of the conditions in* (14) *is satisfied.*
(d) *The sequence* $\{x^k\}$ *is infinite with* $\inf_{k \in J} \left\|g^k\right\| = 0$ *in which case every cluster point* $\bar{x}$ *of* $\{x^k\}$ *(if one exists) satisfies* $0 \in \bar{\partial} f(\bar{x})$.

Proof The proof strategy follows that of the Theorem 5. By assumption the algorithm does not terminate in line (ii) of the GS Algorithm with probability 1 and we can assume $\{x^{k,1}, \ldots, x^{k,m}\} \subset \mathcal{D}$ for all k (see Remark 3). We also assume that neither (a) nor (b) occurs and show that either (c) or (d) must occur. Observe that if J is finite with maximum value k_0, then, by step (iv) of the algorithm, (a) occurs, hence, J is infinite. Following the proof of Theorem 5, we have that (15) holds. We analyze the two mutually exclusive possible outcomes $\inf_{k \in J} \left\|g^k\right\| > 0$ and $\inf_{k \in J} \left\|g^k\right\| = 0$ separately.

First suppose that $\bar{\nu} := \inf_{k \in J} \left\|g^k\right\| > 0$. By (15), the sequence $\{x^k\}$ is Cauchy so that $x^k \to \bar{x}$ for some $\bar{x} \in \mathbb{R}^n$. The argument used in Theorem 5 applies to show that one of the conditions in (14) is satisfied.

Next suppose that $\inf_{k \in J} \left\|g^k\right\| = 0$ and $\bar{x}$ is a cluster point of the sequence $\{x^k\}$. As in the proof of Theorem 5, assume that there is no subsequence $\widehat{J} \subset \mathbb{N}$ satisfying (16). Following the proof of Theorem 5, we again find that $0 \in \bar{\partial} f(\bar{x})$. □

4 Concluding Remarks

The extension of the gradient sampling algorithm to non-Lischitzian, continuous, directionally Lipschitz functions addresses the possibility of unbounded and potentially empty

Clarke subdifferentials. These possibilities affect both the construction of the algorithm and the convergence results. Specifically, in line (vii) of the algorithm, we require that the direction of steepest descent be normalized to have unit magnitude since it may happen that the sequence $\{g^k\}$ is unbounded. Although other normalization strategies are possible, we chose a unit normalization for simplicity. As for the convergence results, the results differ from the Lipschitzian case only by the inclusion of outcome (c) in Theorem 5 as well as Corollaries 2 and 3. This outcome occurs only if the sequence $\{g^k\}$ does not converge to zero in which case it is shown that the sequence $\{x^k\}$ converges to a limit $\bar{x}$. Lemma 11 indicates that this can be manifested in excessively short stepsizes. Nonetheless, in this case failure to converge to a Clarke stationary point only occurs when either $\bar{\partial} f(\bar{x}) = \emptyset$ or $\bar{\partial} f(\bar{x})$ is unbounded and

$$-\text{proj}_{\bar{\partial} f(\bar{x})}(0) \notin \text{int}\,[(\bar{\partial}^{\infty} f(\bar{x}))^*] = \text{int}\left[\text{dom}\left(\hat{d} f(\bar{x})(\cdot)\right)\right],$$

or equivalently, the regular subderivative $\hat{d} f(\bar{x})(\cdot)$ is not continuous at the direction of steepest descent (see Lemma 5). This observation yields two open questions in the directionally Lipschitz case. First, is it possible for $\bar{\partial} f(\bar{x}) = \emptyset$, and if so, when does this occur? Second, is there a way to modify the search direction so that the iterates are not attracted to non-stationary points at which $-\text{proj}_{\bar{\partial} f(\bar{x})}(0) \notin \text{int}\,(\bar{\partial}^{\infty} f(\bar{x}))^*$, or is this a fundamental limitation of the method?

Finally, we note that the class of directionally Lipschitz functions is still not sufficiently broad to capture the non-symmetric spectral functions even though the method has successfully been applied in this case [4–6]. For these functions, there is still much more work to do and it is likely that a very different approach to the convergence analysis is required.

Acknowledgements The authors sincerely thank two referees for their care careful reading and thoughtful comments. Their efforts have significantly improved the clarity of the presentation.

References

1. Borwein, J.M., Zhu, Q.J.: Techniques of Variational Analysis. Canadian Math Society (2005)
2. Borwein, J.M., Burke, J.V., Lewis, A.S.: Differentiability of cone-monotone functions on separable Banach space. Pro. Amer. Math Soc. **132**, 1067–1076 (2003)
3. Burke, J.V., Lewis, A.S., Overton, M.L.: Approximating subdifferentials by random sampling of gradients. Math. Oper Res. **27**(3), 567–584 (2002)
4. Burke, J.V., Lewis, A.S., Overton, M.L.: Two numerical methods for optimizing matrix stability. Linear Algebra Appl. **351/352**, 117–145 (2002)
5. Burke, J.V., Lewis, A.S., Overton, M.L.: A Nonsmooth, Nonconvex optimization approach to robust stabilization by static output feedback and Low-Order controllers. IFAC Proceedings Volumes **36**, 175–181 (2003)
6. Burke, J.V., Lewis, A.S., Overton, M.L.: A robust gradient sampling algorithm for nonsmooth nonconvex optimization. SIAM J. Optim. **15**(3), 751–779 (2005)
7. Burke, J.V., Curtis, F.E., Lewis, A.S., Overton, M.L., Simoes, L.E.A.: Gradient sampling methods for nonsmooth optimization. In: Bagirov, A.M., Gaudioso, M., Karmitsa, N., Mäkella, M.M., Taheri, S. (eds.) Numerical Nonsmooth Optimization: State of the Art Algorithms, chapter 6, pp. 201–225. Springer (2020)
8. Burke, J.V., Deng, S.: Weak sharp minima revisited, part ii: Applications to linear regularity and error bounds. Math. Prog. **104**, 236–261 (2005)
9. Burke, J.V.: Methods for Solving Generalized Systems of Inequalities with Application to Nonlinear Programming. PhD thesis University of Illinois at Urbana-Champaign (1983)
10. Clarke, F.H.: Optimization and Nonsmooth Analysis. Wiley, New York (1983). Reprinted by SIAM Philadelphia 1990

11. Kiwiel, K.C.: Convergence of the gradient sampling algorithm for nonsmooth nonconvex optimization. SIAM J. Optim. **18**(2), 379–388 (2007)
12. Lin, Q.: Sparsity and Non-Convex, Non-Smooth Optimization. Phd Thesis, University of Washington, Seattle (2009)
13. Nirenberg, L.: Lecture notes in functional analysis (1961)
14. Rockafellar, R.T., Wets, R.J.-B.: Variational Analysis. Springer (1998)
15. Rockafellar, R.T.: Clarke's tangent cones and the boundaries of closed sets in $\mathbb{R}^n$. Nonlin. Anal. Th. Meth. and Appl. **3**, 145–154 (1979)
16. Rockafellar, R.T.: Directionally lischitzian functions and subdifferential calculus. Proc. London Math. Soc. **39**, 331–355 (1979)
17. Rockafellar, R.T.: Generalized directional derivatives and subdifferentials of nonconvex functions. Canadian J. Math. **32**, 157–180 (1980)