

CONSTRUCTING A SUBGRADIENT FROM DIRECTIONAL DERIVATIVES FOR FUNCTIONS OF TWO VARIABLES

Kamil A. Khan*

Yingwei Yuan

Abstract For any scalar-valued bivariate function that is locally Lipschitz continuous and directionally differentiable, it is shown that a subgradient may always be constructed from the function's directional derivatives in the four compass directions, arranged in a so-called "compass difference". When the original function is nonconvex, the obtained subgradient is an element of Clarke's generalized gradient, but the result appears to be novel even for convex functions. The function is not required to be represented in any particular form, and no further assumptions are required, though the result is strengthened when the function is additionally L-smooth in the sense of Nesterov. For certain optimal-value functions and certain parametric solutions of differential equation systems, these new results appear to provide the only known way to compute a subgradient. These results also imply that centered finite differences will converge to a subgradient for bivariate nonsmooth functions. As a dual result, we find that any compact convex set in two dimensions contains the midpoint of its interval hull. Examples are included for illustration, and it is demonstrated that these results do not extend directly to functions of more than two variables or sets in higher dimensions.

Keywords: subgradients, directional derivatives, support functions, Clarke's generalized gradient, interval hull

1 INTRODUCTION

Subgradient methods [2, 40] and bundle methods [2, 26, 28, 29] for nonsmooth optimization typically use a subgradient at each iteration to provide local sensitivity information that is ultimately useful enough to infer descent. For convex problems, these subgradients are elements of the convex subdifferential; for nonconvex problems, the subgradients must typically be elements of either Clarke's generalized gradient [6] or other established generalized subdifferentials [27, 31, 32]. Evaluating a subgradient directly, however, may be a challenging task; this difficulty has motivated the development of numerous subdifferential approximations [2, 11, 34].

Nevertheless, there are several settings in which evaluating directional derivatives is much simpler than evaluating a subgradient using established methods. For finite compositions of simple smooth and nonsmooth functions, directional derivatives may be evaluated efficiently [12] by extending the standard forward/tangent mode of algorithmic differentiation [13], while extensions to efficient subgradient evaluation methods require more care [21, 24]. Directional derivatives of implicit functions and inverse functions may be obtained by solving auxiliary equation systems [39], whereas subgradient results in this setting assume either special structure [25, 39] or a series of recursive equation-solves [25].

This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) under Grant RGPIN-2017-05944.

*Corresponding author.

K.A. Khan and Y. Yuan are with the Department of Chemical Engineering, McMaster University, Hamilton, ON, Canada.
Email: kamilkhan@mcmaster.ca.

For solutions of parametric ordinary differential equations (ODEs) with nonsmooth right-hand sides, directional derivatives may be evaluated by solving an auxiliary ODE system [34, Theorem 7] using a standard ODE solver, whereas the only general method for subgradient evaluation involves solving a series of ODEs punctuated by discrete jumps that must be handled carefully [19, 23]. In parametric optimization, Danskin's classical result [8, 17] describes directional derivatives for optimal-value functions as the solutions of related optimization problems in a general setting, while subgradient results such as [43, Theorem 5.1] tend to additionally require unique solutions for the embedded optimization problem.

Moreover, directional derivatives and subdifferentials of convex functions are essentially duals [15]. Hence, this article examines the question of whether, given a directional-derivative evaluation oracle for a function and little else, this oracle may be used to compute a subgradient at each iteration of a typical nonsmooth optimization method. This is clearly true for univariate functions, for example; in this case, the entire subdifferential may be constructed from directional derivatives in the positive and negative directions.

To address this question, this article defines a function's *compass differences* to be vectors obtained by arranging directional derivatives in the coordinate directions and negative coordinate directions in a certain way. Thus, for a bivariate function, a compass difference involves directional derivatives in the four compass directions. For a bivariate function that is locally Lipschitz continuous and directionally differentiable, it is shown that the compass difference at any domain point is a subgradient, with this subgradient understood to be an element of Clarke's generalized gradient in the nonconvex case. Surprisingly, while this result is simple to state, it appears to be previously unknown even for convex functions, and does not require any additional assumptions. It is also shown that this result does not extend directly to functions of more than two variables. As a related result, this article shows that a compact convex set in \mathbb{R}^2 must always contain the midpoint of its interval hull, though this does not extend directly to sets in \mathbb{R}^n for $n > 2$. Hence, four calls to a directional-derivative evaluation oracle are sufficient to compute a subgradient for a nonsmooth bivariate function, and centered finite differences for these functions are useful approximations of a subgradient. In several cases, the approach of this article appears to be the only way known thus far to evaluate a subgradient correctly.

Audet and Hare [1] studied a similar problem involving the similar setup, in the field of geometric probing [41]. Unlike our work, Audet and Hare additionally assume that: (a) their oracle D is convex (as a set's support function), (b) the bivariate function's regular subdifferential is polyhedral, and (c) the oracle D evaluates the function's directional derivative¹. These assumptions are evidently satisfied, for example, by any function that is both convex and piecewise-differentiable in the sense of Scholtes [39]. Under these assumptions, Audet and Hare present a method to use finitely many directional derivative evaluations to construct the whole regular subdifferential at a given domain point. This method proceeds by deducing each vertex of the subdifferential, and depends heavily on the assumption of subdifferential polyhedrality; its complexity scales linearly with the number of subdifferential vertices. It is readily verified, for example, that their Algorithm 1 will run forever without locating any subgradients when applied to the convex Euclidean norm function:

$$f : \mathbb{R}^2 \rightarrow \mathbb{R} : x \mapsto \sqrt{x_1^2 + x_2^2}$$

at $x = 0$. Indeed, their algorithm is not intended to work in this case. Unlike the work of [1], we do not assume that subdifferentials are polyhedral, do not require the subdifferential's support function

¹We note briefly that the piecewise-linear function $f : x \in \mathbb{R}^2 \mapsto \max(0, \min(x_1, x_2))$ has a polyhedral regular subdifferential at $\bar{x} := (0, 0)$, but here the oracle D of [1] is distinct from both the directional derivative $d \mapsto f'(\bar{x}; d)$ (c.f. Definition 2.1 below) and Clarke's generalized directional derivative $d \mapsto f^\circ(\bar{x}; d)$ (from [6]). In this case $d \mapsto f'(\bar{x}; d)$ is nonconvex, so (a) and (c) cannot be satisfied simultaneously even if D is redefined.

to be available in the nonconvex case, and do not assume directional derivatives to be convex with respect to direction. Our goal is only to identify one subgradient rather than a whole subdifferential; characterizing a whole subdifferential in closed form may be difficult or impossible when we do not know *a priori* that it is polyhedral. As mentioned above, in the nonconvex case, we evaluate an element of Clarke’s generalized gradient [6] instead of a subgradient.

The remainder of this article is structured as follows. Section 2 summarizes relevant established constructions in nonsmooth analysis, Section 3 defines compass differences in terms of directional derivatives and shows that they are valid subgradients, and Section 4 presents several examples for illustration.

2 MATHEMATICAL BACKGROUND

The Euclidean norm $\|\cdot\|$ and inner product $\langle \cdot, \cdot \rangle$ are used throughout this article. The i^{th} unit coordinate vector in \mathbb{R}^n is denoted as $e_{(i)}$, and components of vectors are indicated using subscripts, e.g. $x_i := \langle e_{(i)}, x \rangle$. The convex hull and the closure of a set $S \subset \mathbb{R}^n$ are denoted as $\text{conv } S$ and $\text{cl } S$, respectively.

2.1 DIRECTIONAL DERIVATIVES AND CONVEX SUBGRADIENTS

Definition 2.1. Consider an open set $X \subset \mathbb{R}^n$ and a function $f : X \rightarrow \mathbb{R}$. The following limit, if it exists, is the *(one-sided) directional derivative* of f at $x \in X$ in the direction $d \in \mathbb{R}^n$:

$$f'(x; d) := \lim_{t \rightarrow 0^+} \frac{f(x + td) - f(x)}{t}.$$

If $f'(x; d)$ exists in \mathbb{R} for each $d \in \mathbb{R}^n$, then f is *directionally differentiable* at x .

This article primarily considers situations where directional derivatives are available via a black-box oracle. For example, this oracle could represent symbolic calculation, the situation-specific directional derivatives described in Section 1, algorithmic differentiation [13], or even finite difference approximation if some error is tolerable.

The primary goal of this article is to use directional derivatives to evaluate a subgradient, defined for convex functions as follows, and generalized to nonconvex functions as in Section 2.2 below. Individual subgradients are used at each iteration of subgradient methods for convex minimization [40] and bundle methods for nonconvex minimization [16]. They are also used to build useful affine outer approximations for nonconvex sets [22, 38]. In each of these applications, only a single subgradient is needed at each visited domain point.

Definition 2.2. Given a convex set $X \subset \mathbb{R}^n$ and a convex function $f : X \rightarrow \mathbb{R}$, $s \in \mathbb{R}^n$ is a *subgradient* of f at $x \in X$ if

$$(2.1) \quad f(y) \geq f(x) + \langle s, y - x \rangle, \quad \forall y \in X.$$

The set of all subgradients of f at x is the *(convex) subdifferential* $\partial f(x)$.

In this definition, if X is open, then $\partial f(x)$ is convex, compact, and nonempty [15]. The directional derivative and subdifferentials of f at x are related as follows [15]. For each $d \in \mathbb{R}^n$,

$$(2.2) \quad f'(x; d) = \max\{\langle s, d \rangle : s \in \partial f(x)\} \quad \text{and} \quad \partial f(x) = \{s \in \mathbb{R}^n : f'(x; d) \geq \langle s, d \rangle, \forall d \in \mathbb{R}^n\}.$$

Thus, the subdifferential characterizes the local behavior of convex functions via (2.2), and characterizes the global behavior of convex functions via (2.1). Moreover, (2.2) shows that directional derivatives and subgradients of convex functions are essentially duals of each other.

2.2 NONSMOOTH ANALYSIS

The following constructions by Clarke [6] extend certain subgradient properties to nonconvex functions, and are used in methods for equation-solving [10, 36] and optimization [26, 30, 40].

Definition 2.3. Consider an open set $X \subset \mathbb{R}^n$ and a locally Lipschitz continuous function $f : X \rightarrow \mathbb{R}$. The (Clarke-)generalized directional derivative of f at $x \in X$ in the direction $d \in \mathbb{R}^n$ is:

$$f^\circ(x; d) := \limsup_{\substack{y \rightarrow x \\ t \rightarrow 0^+}} \frac{f(y + td) - f(y)}{t}.$$

Clarke’s generalized gradient of f at x is then:

$$\partial f(x) := \{s \in \mathbb{R}^n : f^\circ(x; d) \geq \langle s, d \rangle, \quad \forall d \in \mathbb{R}^n\}.$$

Elements of Clarke’s generalized gradient will be called *Clarke subgradients*.

With f as in the above definition, and for any $x \in X$, $\partial f(x)$ is guaranteed to be nonempty, convex, and compact in \mathbb{R}^n . As suggested by its notation, Clarke’s generalized gradient does indeed coincide with the convex subdifferential when f is convex [6]. When f is nonconvex, (2.2) is no longer guaranteed to hold with Clarke’s generalized gradient in place of the convex subdifferential. The following result for univariate functions is easily demonstrated, and is summarized in [1].

Proposition 2.4. Consider an open set $X \subset \mathbb{R}$ and a univariate function $f : X \rightarrow \mathbb{R}$ that is locally Lipschitz continuous and directionally differentiable. For each $x \in X$,

$$(2.3) \quad \partial f(x) = \text{conv} \{f'(x; 1), -f'(x; -1)\}.$$

Hence, one call to an oracle that evaluates directional derivatives is sufficient to obtain a single Clarke subgradient for such a univariate function f . It will be shown in this article that, for bivariate functions $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ that are locally Lipschitz continuous and directionally differentiable, four directional derivative evaluations are sufficient to evaluate a single Clarke subgradient.

The following definition by Nesterov [32] will be used to specialize this result in a useful way. Nesterov’s definition is based on repeated directional differentiation, and permits certain extensions of calculus rules for smooth functions to nonsmooth functions.

Definition 2.5. Consider an open set $X \subset \mathbb{R}^n$ and a locally Lipschitz continuous function $f : X \rightarrow \mathbb{R}$. The function f is *lexicographically (L-)smooth* at $x \in X$ if the following conditions are satisfied:

- f is directionally differentiable at x ,
- with $f^{(0)} := f'(x; \cdot)$, for any collection of vectors $m_{(1)}, \dots, m_{(n)} \in \mathbb{R}^n$, the following inductive sequence of higher-order directional derivatives is well-defined:

$$f^{(k)} := [f^{(k-1)}]'(m_{(k)}; \cdot), \quad \text{for each } k \in \{1, 2, \dots, n\}.$$

If these vectors $m_{(i)}$ are linearly independent, then $f^{(n)}$ is linear, and its constant gradient is called a *lexicographic subgradient* of f at x . The *lexicographic subdifferential* $\partial_L f(x)$ is the set of all lexicographic subgradients of f at x .

All convex functions on open domains in \mathbb{R}^n are L-smooth [32], as are differentiable functions, functions that are piecewise differentiable in the sense of Scholtes [24, 39], and functions that are well-defined finite compositions of other L-smooth functions [32]. Further characterizations of L-smoothness have been developed for certain optimal-value functions [43], and for parametric systems of ordinary differential equations or differential-algebraic equations [23, 42].

3 CONSTRUCTING A SUBGRADIENT FROM DIRECTIONAL DERIVATIVES

This section defines *compass differences* for functions in terms of directional derivatives, and shows that a compass difference of a bivariate function is a subgradient. As a corollary, it is also shown that any compact convex set in \mathbb{R}^2 contains the midpoint of its interval hull. As there is nothing particularly special about the compass directions in this context, other choices of directions are also considered.

3.1 COMPASS DIFFERENCES

Definition 3.1. Consider an open set $X \subset \mathbb{R}^n$ and a function $f : X \rightarrow \mathbb{R}$ that is directionally differentiable at $x \in X$. The *compass difference* of f at x is a vector $\Delta^\oplus f(x) := (\Delta_1^\oplus f(x), \dots, \Delta_n^\oplus f(x)) \in \mathbb{R}^n$ for which, for each $i \in \{1, \dots, n\}$,

$$\Delta_i^\oplus f(x) := \frac{1}{2}(f'(x; e_{(i)}) - f'(x; -e_{(i)})).$$

The compass difference is so named because it considers how f behaves when its argument is varied in each of the compass directions. This metaphor works best when $n = 2$; this case is also the focus of this article.

Evaluating $\Delta^\oplus f(x)$ ostensibly requires $2n$ directional derivative evaluations. However, if directional derivative values are not available, compass differences may instead be approximated using finite differences. Observe that the compass difference of a function is a centered finite difference of the directional derivative mapping $f'(x; \cdot)$ at 0. From the definition of the directional derivative, we have, for each $i \in \{1, \dots, n\}$,

$$\Delta_i^\oplus f(x) = \lim_{\delta \rightarrow 0^+} \frac{f(x + \delta e_{(i)}) - f(x - \delta e_{(i)})}{2\delta}.$$

So, if numerical evaluations of $f : \mathbb{R}^n \rightarrow \mathbb{R}$ are viable but evaluations of $f'(x; \cdot)$ are not, then $2n$ evaluations of f may be used to approximate $\Delta^\oplus f(x)$ using the argument of the above limit. That is, for sufficiently small $\delta > 0$,

$$(3.1) \quad \Delta^\oplus f(x) \approx \frac{1}{2\delta} \begin{bmatrix} f(x + \delta e_{(1)}) - f(x - \delta e_{(1)}) \\ f(x + \delta e_{(2)}) - f(x - \delta e_{(2)}) \\ \vdots \\ f(x + \delta e_{(n)}) - f(x - \delta e_{(n)}) \end{bmatrix},$$

which is incidentally the centered simplex gradient of f at x with a sampling set comprising the coordinate vectors (c.f. [7]). However, if f is evaluated here using a numerical method, and if δ is too small, then the subtraction operations in this approximation may introduce unacceptable numerical error. This drawback is typical of finite difference approximations.

3.2 NONCONVEX FUNCTIONS OF TWO VARIABLES

As in [11], let us say that a function $\mathbb{R}^n \rightarrow \mathbb{R}$ is *B-differentiable* if it is both directionally differentiable and locally Lipschitz continuous. This section presents the main result of this article: that any compass difference of a B-differentiable function of two variables is a Clarke subgradient. This result is strengthened somewhat when the considered function is L-smooth, and is also specialized to convex functions and convex sets in the subsequent sections.

To our knowledge, the main result in this section is the first general closed-form description of a Clarke subgradient for a nonconvex bivariate function in terms of that function's directional derivatives (in the sense of Definition 2.1). Moreover, the result shows that four calls to a directional derivative oracle are sufficient to evaluate a Clarke subgradient for a bivariate B-differentiable function, without any further structural knowledge of the function at all. Unlike established characterizations of generalized

subgradients such as [39, Proposition 4.3.1] and [21, Theorem 3.5], this result does not require f to be represented in any particular format.

The following mean-value theorem will be useful in this development.

Lemma 3.2. *Consider a function $\psi : \mathbb{R}^n \rightarrow \mathbb{R}$ that is positively homogeneous and locally Lipschitz continuous. For any $x, y \in \mathbb{R}^n$, there exists $s \in \partial\psi(0)$ for which*

$$(3.2) \quad \psi(y) - \psi(x) = \langle s, y - x \rangle.$$

If ψ is also L -smooth, then there exists $s \in \text{conv } \partial_L \psi(0) \subset \partial\psi(0)$ satisfying (3.2).

Proof. We first proceed without the L -smoothness assumption. According to Lebourg’s mean-value theorem [6, Theorem 2.3.7], the equation (3.2) holds for some $z \in \text{conv } \{x, y\}$ and some $s \in \partial\psi(z)$. Since ψ is positively homogeneous and locally Lipschitz continuous, [23, Lemma 3.1] implies that $\partial\psi(z) \subset \partial\psi(0)$, and so $s \in \partial\psi(0)$, as required.

Next, if ψ is additionally assumed to be L -smooth, then the final claimed result is obtained by a similar argument, applying Nesterov’s mean-value theorem [32, Theorem 12] instead of Lebourg’s, applying [20, Lemma 4] instead of [23, Lemma 3.1], and applying [32, Theorem 11] to establish the inclusion $\text{conv } \partial_L \psi(0) \subset \partial\psi(0)$. \square

The following theorem is the main result of this article, and rests heavily on Lemma 3.2. It shows that any compass difference of a B -differentiable function is a Clarke subgradient, and specializes this result to L -smooth functions.

Theorem 3.3. *Consider an open set $X \subset \mathbb{R}^2$ and a locally Lipschitz continuous function $f : X \rightarrow \mathbb{R}$. If f is directionally differentiable at some $x \in X$, then $\Delta^\oplus f(x) \in \partial f(x)$. Moreover, if f is L -smooth at $x \in X$, then $\Delta^\oplus f(x) \in \text{cl conv } \partial_L f(x) \subset \partial f(x)$.*

Proof. Suppose that f is directionally differentiable at $x \in X$. Consider the auxiliary mapping:

$$\psi : y \mapsto f'(x; y) - \langle \Delta^\oplus f(x), y \rangle,$$

and observe that ψ is Lipschitz continuous [39], and that $f'(x; y) = \psi(y) + \langle \Delta^\oplus f(x), y \rangle$ for each $y \in \mathbb{R}^2$. Thus, Clarke’s calculus rule for addition [6, Corollary 1 to Proposition 2.3.3] implies:

$$\partial[f'(x; \cdot)](0) = \{a + \Delta^\oplus f(x) : a \in \partial\psi(0)\}.$$

Moreover, [23, Corollary 3.1] and [18] imply $\partial[f'(x; \cdot)](0) \subset \partial f(x)$, and so

$$\{a + \Delta^\oplus f(x) : a \in \partial\psi(0)\} \subset \partial f(x).$$

It therefore suffices to show that $0 \in \partial\psi(0)$.

Now, observe that ψ is positively homogeneous, and so ψ is equivalent to $\psi'(0; \cdot)$. Thus, for each $i \in \{1, 2\}$,

$$\begin{aligned} \Delta_i^\oplus \psi(0) &= \frac{1}{2}(\psi(e_{(i)}) - \psi(-e_{(i)})) \\ &= \frac{1}{2} [(f'(x; e_{(i)}) - \langle \Delta^\oplus f(x), e_{(i)} \rangle) - (f'(x; -e_{(i)}) - \langle \Delta^\oplus f(x), -e_{(i)} \rangle)] \\ &= \Delta_i^\oplus f(x) - \langle \Delta^\oplus f(x), e_{(i)} \rangle \\ &= 0. \end{aligned}$$

Hence $\Delta^\oplus \psi(0) = 0$.

To obtain a contradiction, suppose that $0 \notin \partial\psi(0)$. Then, since $\partial\psi(0)$ is convex and closed, there must exist a strictly separating hyperplane between 0 and $\partial\psi(0)$. That is, there exist a nonzero vector $p := (p_1, p_2) \in \mathbb{R}^2$ and a scalar $a > 0$ for which $\langle p, s \rangle \geq a$ for each $s \in \partial\psi(0)$.

Since $\Delta^\oplus\psi(0) = 0$, we have $\psi(1, 0) = \psi(-1, 0)$ and $\psi(0, 1) = \psi(0, -1)$. Since ψ is positively homogeneous, we then have $\psi(p_1, 0) = \psi(-p_1, 0)$ and $\psi(0, p_2) = \psi(0, -p_2)$ (regardless of the signs of p_1 and p_2). Subtraction then yields:

$$(3.3) \quad \psi(p_1, 0) - \psi(0, -p_2) = \psi(-p_1, 0) - \psi(0, p_2).$$

Now, according to Lemma 3.2, there exist vectors $\eta, \sigma \in \partial\psi(0)$ for which

$$\psi(p_1, 0) - \psi(0, -p_2) = \langle p, \eta \rangle \quad \text{and} \quad \psi(-p_1, 0) - \psi(0, p_2) = -\langle p, \sigma \rangle.$$

Hence, since $\langle p, s \rangle \geq a$ for each $s \in \partial\psi(0)$, we have

$$\psi(p_1, 0) - \psi(0, -p_2) \geq a > 0 > -a \geq \psi(-p_1, 0) - \psi(0, p_2),$$

which contradicts (3.3). Thus, $0 \in \partial\psi(0)$ as required.

Next, suppose that f is L-smooth at $x \in X$. The inclusion $\partial_L f(x) \subset \partial f(x)$ was shown by Nesterov [32, Theorem 11]; since $\partial f(x)$ is closed and convex, it follows that $\text{cl conv } \partial_L f(x) \subset \partial f(x)$. Consider the auxiliary mapping ψ as above, and note that (3.3) still holds. The calculus rules of the lexicographic subdifferential [32, Theorem 5 and Definitions 1 and 5] imply that both $f'(x; \cdot)$ and ψ are L-smooth at 0, and that

$$\partial_L f(x) = \partial_L[f'(x; \cdot)](0) = \{a + \Delta^\oplus f(x) : a \in \partial_L \psi(0)\}.$$

From here, a similar argument to the previous case shows that $\Delta^\oplus f(x) \in \text{cl conv } \partial_L f(x)$. □

Intuitively, there is nothing special about the coordinate directions used to construct a compass difference, and a change of basis in Theorem 3.3 may be carried out as follows.

Corollary 3.4. *Consider an open set $X \subset \mathbb{R}^2$, a locally Lipschitz continuous function $f : X \rightarrow \mathbb{R}$, and a nonsingular matrix $V \in \mathbb{R}^{2 \times 2}$. If f is directionally differentiable at some $x \in X$, and if $v_{(i)}$ denotes the i^{th} column of V , then*

$$\frac{1}{2}(V^T)^{-1} \begin{bmatrix} f'(x; v_{(1)}) - f'(x; -v_{(1)}) \\ f'(x; v_{(2)}) - f'(x; -v_{(2)}) \end{bmatrix} \in \partial f(x).$$

Proof. Consider auxiliary mappings:

$$g : \mathbb{R}^2 \rightarrow \mathbb{R}^2 : y \mapsto x + V(y - x)$$

and $h := f \circ g$. Let

$$z := \frac{1}{2} \begin{bmatrix} f'(x; v_{(1)}) - f'(x; -v_{(1)}) \\ f'(x; v_{(2)}) - f'(x; -v_{(2)}) \end{bmatrix}.$$

The chain rule for directional derivatives [39, Theorem 3.1.1] implies that $\Delta^\oplus h(x) = z$, and so Theorem 3.3 shows that $z \in \partial h(x)$. Since V is nonsingular, g is surjective, in which case [6, Theorem 2.3.10] implies that:

$$\partial h(x) = \{V^T s : s \in \partial f(x)\}.$$

Thus, $z = V^T s$ for some $s \in \partial f(x)$, and so $(V^T)^{-1}z \in \partial f(x)$ as claimed. □

The particular Clarke subgradients identified by Theorem 3.3 and Corollary 3.4 do not necessarily coincide.

We may remove the directional differentiability requirement of Theorem 3.3 as follows, by employing Clarke's generalized directional derivative f° from Definition 2.3. We note, however, that the generalized directional derivative is typically inaccessible in practice.

Corollary 3.5. *Given an open set $X \subset \mathbb{R}^2$, a locally Lipschitz continuous function $f : X \rightarrow \mathbb{R}$, and some $x \in X$,*

$$\Delta^\oplus[f^\circ(x; \cdot)](0) = \frac{1}{2} \left[\begin{matrix} f^\circ(x; (1, 0)) - f^\circ(x; (-1, 0)) \\ f^\circ(x; (0, 1)) - f^\circ(x; (0, -1)) \end{matrix} \right] \in \partial[f^\circ(x; \cdot)](0) = \partial f(x).$$

Proof. Established results [15, Section V, Proposition 2.1.2] and [6, Proposition 2.1.2] imply that $f^\circ(x; \cdot)$ is convex and positively homogeneous (as the support function of $\partial f(x)$), and has the subdifferential $\partial f(x)$ at 0. Moreover, as a convex function on an open domain, $f^\circ(x; \cdot)$ is locally Lipschitz continuous and directionally differentiable [15]. Hence, [Theorem 3.3](#) implies the claimed result. \square

3.3 CONVEX FUNCTIONS OF TWO VARIABLES

This section specializes [Theorem 3.3](#) to convex functions; this specialization appears to be a novel result in convex analysis and is simpler to state. Namely, any compass difference of a bivariate convex function is in fact a subgradient in the traditional sense. Hence, four directional derivative evaluations are sufficient to construct a subgradient of a bivariate convex function.

Corollary 3.6. *Consider an open convex set $X \subset \mathbb{R}^2$ and a convex function $f : X \rightarrow \mathbb{R}$. For each $x \in X$, $\Delta^\oplus f(x) \in \partial f(x)$.*

Proof. Since f is convex and X is open, f is locally Lipschitz continuous and directionally differentiable [15]. The claimed result then follows immediately from [Theorem 3.3](#). \square

Corollary 3.7. *Consider an open convex set $X \subset \mathbb{R}^2$, a convex function $f : X \rightarrow \mathbb{R}$, and a nonsingular matrix $V \in \mathbb{R}^{2 \times 2}$. For any $x \in X$, with $v_{(i)}$ denoting the i^{th} column of V ,*

$$\frac{1}{2}(V^T)^{-1} \left[\begin{matrix} f'(x; v_{(1)}) - f'(x; -v_{(1)}) \\ f'(x; v_{(2)}) - f'(x; -v_{(2)}) \end{matrix} \right] \in \partial f(x).$$

Proof. Again, since f is locally Lipschitz continuous and directionally differentiable, the claimed corollary is a special case of [Corollary 3.4](#). \square

3.4 COMPACT CONVEX SETS IN \mathbb{R}^2

This section applies [Corollary 3.6](#) to show that any nonempty compact convex set in \mathbb{R}^2 contains the center of its smallest enclosing box (or *interval*). These notions are formalized in the following classical definitions (summarized in [33]), followed by the claimed result.

Definition 3.8. An *interval* in \mathbb{R}^n is a nonempty set of the form $\{x \in \mathbb{R}^n : a \leq x \leq b\}$, where $a, b \in \mathbb{R}^n$, and where each inequality is to be interpreted componentwise. The *midpoint* of an interval $\{x \in \mathbb{R}^n : a \leq x \leq b\}$ is $\frac{1}{2}(a + b) \in \mathbb{R}^n$.

Given a bounded set $B \subset \mathbb{R}^n$, the *interval hull* of B is the intersection in \mathbb{R}^n of all interval supersets of B .

The interval hull of a bounded set $B \subset \mathbb{R}^n$ is itself an interval, and is, intuitively, the smallest interval superset of B . Support functions of convex sets, defined as follows and discussed at length in [15], are useful when relating convex sets to properties of subdifferentials of convex functions.

Definition 3.9. Given a set $C \subset \mathbb{R}^n$, the *support function* of C is the mapping:

$$\sigma_C : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\pm\infty\} : d \mapsto \sup\{\langle d, x \rangle : x \in C\}.$$

The following corollary uses support functions to extend [Corollary 3.6](#) to the problem of locating an element of a closed convex set in \mathbb{R}^2 .

Corollary 3.10. *Any nonempty compact convex set $C \subset \mathbb{R}^2$ contains the midpoint of its interval hull.*

Proof. The interval hull of C may be expressed in terms of the support function σ_C as:

$$\{x \in \mathbb{R}^2 : -\sigma_C(-1, 0) \leq x_1 \leq \sigma_C(1, 0), \quad -\sigma_C(0, -1) \leq x_2 \leq \sigma_C(0, 1)\};$$

the midpoint of this interval hull is then

$$z := \frac{1}{2} \begin{bmatrix} \sigma_C(1, 0) - \sigma_C(-1, 0) \\ \sigma_C(0, 1) - \sigma_C(0, -1) \end{bmatrix}.$$

As shown in [15, Section VI, Example 3.1], σ_C is directionally differentiable at 0, with $(\sigma_C)'(0; d) = \sigma_C(d)$ for each $d \in \mathbb{R}^2$. Thus, $\Delta^\oplus \sigma_C(0) = z$.

Next, [15, Section VI, Example 3.1] also shows that σ_C is convex, with $\partial \sigma_C(0) = C$. Combining these observations with Corollary 3.6 yields $z = \Delta^\oplus \sigma_C(0) \in \partial \sigma_C(0) = C$, as claimed. \square

4 EXAMPLES

This section illustrates the main results of this article. Section 4.1 motivates the assumptions of Corollary 3.6 and Corollary 3.10 by showing how these results could fail if their assumptions were weakened. Section 4.2 uses compass differences to compute individual subgradients in cases where this was previously difficult or impossible.

4.1 COUNTEREXAMPLES FOR RELATED CLAIMS

The following example shows that, for functions mapping \mathbb{R}^2 into \mathbb{R} , compass differences are not necessarily elements of either the regular subdifferential [37], the lexicographic subdifferential [32], the B-subdifferential [35, 39], or the Mordukhovich upper subdifferential [31].

Example 4.1. Consider the concave piecewise-linear function:

$$f : \mathbb{R}^2 \rightarrow \mathbb{R} : x \mapsto -|x_1|.$$

Direct computation yields $\Delta^\oplus f(0, 0) = (0, 0)$, which is indeed an element of $\partial f(0, 0) = \{(\lambda, 0) : -1 \leq \lambda \leq 1\}$. However, the lexicographic subdifferential, the B-subdifferential, and the Mordukhovich upper subdifferential of f at $(0, 0)$ are each equal to $\{(-1, 0), (1, 0)\}$, which does not contain $(0, 0)$. The regular subdifferential of f at $(0, 0)$ is empty.

The following example shows that Theorem 3.3, Corollary 3.6, and Corollary 3.10 are minimal in the sense that, under the respective assumptions of these results, three support function evaluations are generally not sufficient to infer a set element, and three directional derivative evaluations are generally not sufficient to infer a function's subgradient.

Example 4.2. Suppose that $C \subset \mathbb{R}^2$ is the unit ball $\{x \in \mathbb{R}^2 : \|x\| \leq 1\}$, which has the constant support function $\sigma_C : d \mapsto 1$. Consider three nonzero points $u, v, w \in \mathbb{R}^2$ in general position. From the support function's definition, if we did not know the set C but did know that $\sigma_C(u) = \sigma_C(v) = \sigma_C(w) = 1$, then we could infer that C is a subset of the triangle:

$$T := \{x \in \mathbb{R}^2 : \langle u, x \rangle \leq 1, \quad \langle v, x \rangle \leq 1, \quad \langle w, x \rangle \leq 1\}.$$

Denote the three vertices of T as $a, b, c \in \mathbb{R}^2$, and denote the three edges of T as

$$T_1 := \text{conv}\{a, b\}, \quad T_2 := \text{conv}\{b, c\}, \quad \text{and} \quad T_3 := \text{conv}\{a, c\}.$$

Since $\{a, b\} \subset T_1 \subset T$, observe that

$$\sigma_C(u) = 1 = \sigma_T(u) \geq \sigma_{T_1}(u) \geq \langle u, x \rangle \quad \forall x \in \{a, b\}.$$

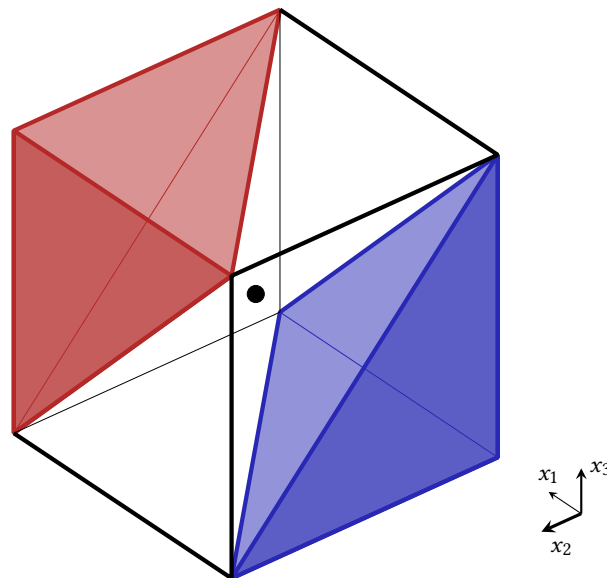


Figure 1: The disjoint convex compact sets C_1 (red) and C_2 (blue) in \mathbb{R}^3 described in Example 4.3, and the common midpoint (black dot) of their interval hulls.

But, since a, b, c are the vertices of the triangle T , and since one edge of T lies on the line $\langle u, x \rangle = 1$, it cannot be that $\langle u, a \rangle$ and $\langle u, b \rangle$ are both less than 1. Hence $\sigma_{T_1}(u) \geq 1$, and so $\sigma_C(u) = \sigma_{T_1}(u)$.

Similar logic shows that $\sigma_{T_i}(u) = \sigma_C(u)$, $\sigma_{T_i}(v) = \sigma_C(v)$, and $\sigma_{T_i}(w) = \sigma_C(w)$ for each $i \in \{1, 2, 3\}$. Each T_i is compact and convex, and the intersection $T_1 \cap T_2 \cap T_3$ is empty. Hence, there is no way to infer an element of C from the support function evaluations $\sigma_C(u)$, $\sigma_C(v)$, and $\sigma_C(w)$ and the knowledge that C is compact and convex; these support function evaluations are consistent with the incorrect hypotheses $C = T_1$, $C = T_2$, and $C = T_3$, yet these guesses have no point in common.

Similarly, considering the convex Euclidean norm function

$$f : \mathbb{R}^2 \rightarrow \mathbb{R} : x \mapsto \|x\|,$$

it is readily verified that $\partial f(0) = C$. Suppose we know nothing about f other than its convexity and the fact that $f'(0; u) = f'(0; v) = f'(0; w) = 1$. In this case, there is no way to infer an element of $\partial f(0)$ from these three directional derivatives alone, since for each $i \in \{1, 2, 3\}$, the functions

$$\phi_i : \mathbb{R}^2 \rightarrow \mathbb{R} : d \mapsto \max\{\langle s, d \rangle : s \in T_i\}$$

all have the same directional derivatives as f at 0 in the directions u, v , and w . However, their subdifferentials at 0 are the sets T_i , which have no point in common.

The following example shows that the results of this article do not extend directly to functions of more than two variables or sets in more than two dimensions.

Example 4.3. Consider the following convex compact sets in \mathbb{R}^3 :

$$C_1 := \text{conv} \{(1, 1, -1), (-1, 1, 1), (1, -1, 1), (1, 1, 1)\},$$

$$\text{and } C_2 := \text{conv} \{(1, -1, -1), (-1, 1, -1), (-1, -1, 1), (-1, -1, -1)\}.$$

These sets are illustrated in Figure 1. They are disjoint; for any $x \in C_1$ and $y \in C_2$, and with $e := (1, 1, 1) \in \mathbb{R}^3$, observe that

$$\langle e, x \rangle \geq 1 > -1 \geq \langle e, y \rangle.$$

However, it is readily verified that both C_1 and C_2 have the interval hull $[-1, 1]^3$, whose midpoint is $(0, 0, 0)$, which is in neither C_1 nor C_2 . Thus, [Corollary 3.10](#) does not extend immediately to \mathbb{R}^3 .

Similarly, consider the following two convex piecewise-linear functions:

$$\begin{aligned} f : \mathbb{R}^3 &\rightarrow \mathbb{R} : x \mapsto \max\{x_1 + x_2 - x_3, \quad x_2 + x_3 - x_1, \quad x_3 + x_1 - x_2\}, \\ \phi : \mathbb{R}^3 &\rightarrow \mathbb{R} : x \mapsto \max\{x_1 - x_2 - x_3, \quad x_2 - x_3 - x_1, \quad x_3 - x_1 - x_2\}. \end{aligned}$$

According to [[39](#), Proposition 4.3.1], $\partial f(0) \subset \text{conv}\{(1, 1, -1), (-1, 1, 1), (1, -1, 1)\} \subset C_1$, and $\partial\phi(0) \subset \text{conv}\{(1, -1, -1), (-1, 1, -1), (-1, -1, 1)\} \subset C_2$. Thus, the subdifferentials $\partial f(0)$ and $\partial\phi(0)$ are disjoint. Moreover, it is readily verified that:

$$1 = f'(0; se_{(i)}) = \phi'(0; se_{(i)}), \quad \forall s \in \{-1, +1\}, \quad \forall i \in \{1, 2, 3\}.$$

Thus, the functions f and ϕ cannot be distinguished based on their directional derivatives at 0 in any coordinate direction or negative coordinate direction, and $\Delta^\oplus f(0) = \Delta^\oplus\phi(0) = 0$, but the two functions' subdifferentials at 0 are disjoint. This shows that [Theorem 3.3](#) and [Corollary 3.6](#) do not extend immediately to functions of three variables.

The following example illustrates that the assumption in [Corollary 3.10](#) that C is closed is crucial.

Example 4.4. Consider the convex set:

$$C := \{x \in \mathbb{R}^2 : -1 < x_1, \quad x_2 < 1, \quad x_1 < x_2\}.$$

Observe that C is not closed, and that the interval hull of C is $[-1, 1]^2$. The midpoint of this hull is $(0, 0)$, which is not an element of C .

4.2 APPLICATIONS

4.2.1 SOLUTIONS OF PARAMETRIC DIFFERENTIAL EQUATIONS

This section applies [Theorem 3.3](#) to describe correct single subgradients for solutions of parametric ordinary differential equations (ODEs) with parameters in \mathbb{R}^2 . This approach reduces to the classical ODE sensitivity approach of [[14](#), Section V, Theorem 3.1] when the original ODE is defined in terms of smooth functions. Unlike existing methods [[23](#)] for generalized derivative evaluation for these systems, the approach of this article describes a subgradient in terms of auxiliary ODE systems that can be integrated numerically using off-the-shelf ODE solvers, but is of course restricted to systems with two parameters.

We consider the following setup, which is readily adapted to other ODE representations.

Assumption 4.5. Consider functions $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$, $x_0 : \mathbb{R}^2 \rightarrow \mathbb{R}^n$, and $g : \mathbb{R}^2 \times \mathbb{R}^n \rightarrow \mathbb{R}$ that are locally Lipschitz continuous and directionally differentiable. For some scalar $t_f > 0$, let $x : [0, t_f] \times \mathbb{R}^2$ be defined so that, for each $p \in \mathbb{R}^2$, $x(\cdot, p)$ solves the following ODE system uniquely:

$$\frac{dx}{dt}(t, p) = f(x(t, p)), \quad x(0, p) = x_0(p).$$

Define $\phi : \mathbb{R}^2 \rightarrow \mathbb{R}$ to be the cost function:

$$\phi : p \mapsto g(p, x(t_f, p)).$$

Under this assumption, a subgradient for ϕ may be computed by combining the results of this article with directional derivatives described by [[34](#), Theorem 7] as follows. If it is desired for the ODE right-hand-side to depend explicitly on t , then an alternative directional derivative result [[23](#), Theorem 4.1] may be used instead.

Proposition 4.6. *Suppose that Assumption 4.5 holds, and consider some particular $p \in \mathbb{R}^2$. For each $d \in \mathbb{R}^2$, let $y(\cdot, d)$ denote a solution on $[0, t_f]$ of the following ODE:*

$$(4.1) \quad \frac{dy}{dt}(t, d) = f'(x(t, p); y(t, d)) \quad y(0, d) = x'_0(p; d).$$

Then $y(\cdot, d)$ is in fact the unique solution of this ODE for each $d \in \mathbb{R}^2$. Moreover, if we define

$$\psi(d) := g'((p, x(t_f, p)); (d, y(t_f, d)))$$

for each $d \in \mathbb{R}^2$, then

$$\frac{1}{2} \begin{bmatrix} \psi(1, 0) - \psi(-1, 0) \\ \psi(0, 1) - \psi(0, -1) \end{bmatrix}$$

is an element of $\partial\phi(p)$.

Proof. According to [34, Theorem 7], $y(t, d)$ is the directional derivative $x'((t, p); (0, d))$ for each $t \in [0, t_f]$ and $d \in \mathbb{R}^2$. The result then follows immediately from Corollary 3.4 and the chain rule [39, Theorem 3.1.1]. \square

If lexicographic derivatives are unavailable for the functions in Assumption 4.5 or do not exist, then Proposition 4.6 is, to our knowledge, the first method for describing a subgradient of ϕ . The following numerical example illustrates this proposition.

Example 4.7. Consider a function $x_0 : \mathbb{R}^2 \rightarrow \mathbb{R}^3 : p \mapsto (p_1, p_2, p_1)$. For each $p \in \mathbb{R}^2$, let $x(\cdot, p)$ denote the unique solution on $[0, 1]$ of the following parametric ODE system. Here dotted variables denote derivatives with respect to t .

$$\begin{aligned} \dot{x}_1 &= |x_1| + |x_2| + x_3, \\ \dot{x}_2 &= |x_2|, \\ \dot{x}_3 &= x_3, \\ x(0, p) &= x_0(p). \end{aligned}$$

Consider a cost function $\phi : p \mapsto x_1(1, p)$. In this case, for each $d \in \mathbb{R}^2$ the ODE (4.1) becomes:

$$\begin{aligned} \dot{y}_1 &= \begin{cases} -y_1 - y_2 + y_3, & \text{if } x_1 < 0, x_2 < 0, \\ -y_1 + y_2 + y_3, & \text{if } x_1 < 0, x_2 > 0, \\ -y_1 + |y_2| + y_3, & \text{if } x_1 < 0, x_2 = 0, \\ y_1 - y_2 + y_3, & \text{if } x_1 > 0, x_2 < 0, \\ y_1 + y_2 + y_3, & \text{if } x_1 > 0, x_2 > 0, \\ y_1 + |y_2| + y_3, & \text{if } x_1 > 0, x_2 = 0, \\ |y_1| - y_2 + y_3, & \text{if } x_1 = 0, x_2 < 0, \\ |y_1| + y_2 + y_3, & \text{if } x_1 = 0, x_2 > 0, \\ |y_1| + |y_2| + y_3, & \text{if } x_1 = 0, x_2 = 0, \end{cases} \\ \dot{y}_2 &= \begin{cases} y_2, & \text{if } x_2 > 0, \\ |y_2|, & \text{if } x_2 = 0, \\ -y_2, & \text{if } x_2 < 0, \end{cases} \\ \dot{y}_3 &= y_3, \end{aligned}$$

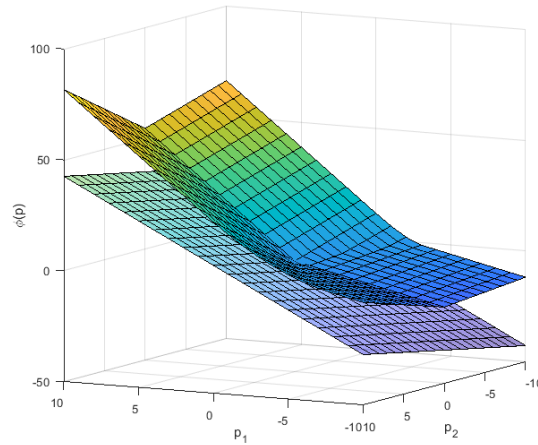


Figure 2: Plot of the function ϕ (top) described in Example 4.7, which appears to dominate the approximation $p \mapsto \phi(0) + \langle s, p \rangle$ (bottom) based on the computed compass difference s of ϕ at 0.

with $y(0, d) \equiv (d_1, d_2, d_1)$. In this case ϕ is convex. To evaluate a compass difference of ϕ , the numerical variable-step variable-order ODE solver `ode15s` was used in MATLAB to evaluate y numerically, using MATLAB’s default precision (on the order of 16 significant digits) for arithmetic, and using respective local absolute and relative tolerances of 10^{-6} and 10^{-3} for each integration step. Thus, to within the corresponding computational error, we obtained $\Delta^\oplus \phi(0) \approx (3.490, 0.772) =: s$, and Proposition 4.6 yields $\Delta^\oplus \phi(0) \in \partial\phi(0)$. Figure 2 shows that s does indeed appear to satisfy (2.1), and does thereby appear to be a subgradient of ϕ at 0 to within numerical precision.

4.2.2 OPTIMAL-VALUE FUNCTIONS

A well-known result by Danskin [8, Theorem 1] describes directional derivatives for certain optimal-value functions, and has been extended to a variety of settings (e.g. [3, 17]). The following proposition and its proof are intended to show how any of these results may be combined with Theorem 3.3 or Corollary 3.6 to describe a subgradient in each case.

Proposition 4.8. Consider a compact set $C \subset \mathbb{R}^n$, some open superset Z of C , and a continuously differentiable function $f : \mathbb{R}^2 \times Z \rightarrow \mathbb{R}$. Define an optimal-value function $\phi : \mathbb{R}^2 \rightarrow \mathbb{R}$ for which

$$\phi : x \mapsto \min\{f(x, y) : y \in C\}.$$

For some particular $\hat{x} \in \mathbb{R}^2$, define the following:

- a set $Y := \{\hat{y} \in C : f(\hat{x}, \hat{y}) \leq f(\hat{x}, y), \quad \forall y \in C\}$,
- for each $d \in \mathbb{R}^2$, a point $\psi(d) := \min\{\langle d, \nabla_x f(\hat{x}, y) \rangle : y \in Y\}$.

Then ϕ is locally Lipschitz continuous and directionally differentiable, and

$$\frac{1}{2} \begin{bmatrix} \psi(1, 0) - \psi(-1, 0) \\ \psi(0, 1) - \psi(0, -1) \end{bmatrix}$$

is an element of $\partial\phi(\hat{x})$.

Proof. The optimal-value function ϕ has already been established to be locally Lipschitz continuous [9, Theorem 2.1] and directionally differentiable [8], with directional derivatives given by $\phi'(\hat{x}; d) = \psi(d)$ for each $d \in \mathbb{R}^2$. The claimed result then follows immediately from Theorem 3.3. \square

Observe that, unlike several established sensitivity results for optimal-value functions [4, 5], the above result does not require second-order sufficient optimality conditions to hold, and does not require unique solutions of the optimization problems defining ϕ . An analogous approach describes subgradients of the Tsoukalas-Mitsos convex relaxations [44] of composite functions of two variables; the Tsoukalas-Mitsos approach is based entirely on analogous optimal-value functions.

5 CONCLUSION

For a bivariate nonsmooth function under minimal assumptions, the compass difference introduced in this article is guaranteed to be a subgradient and may be computed using four calls to a directional-derivative evaluation oracle. This remains true for nonconvex functions, with the “subgradient” understood in this case to be an element of Clarke’s generalized gradient. Thus, for such functions, centered finite differences will necessarily converge to a subgradient as the perturbation width tends to zero. The presented examples show that this new relationship between directional derivatives and subgradients may be useful for functions of two variables, and may in some cases provide the only known way to evaluate a subgradient, but does not extend directly to functions of three or more variables. Such a nontrivial extension represents a possible avenue for future work.

REFERENCES

- [1] C. Audet and W. Hare, Algorithmic construction of the subdifferential from directional derivatives, *Set-Valued Var. Anal.* 26 (2018), 431–447.
- [2] A. Bagirov, N. Karimtsa, and M. M. Mäkelä, *Introduction to Nonsmooth Optimization: Theory, Practice, and Software*, Springer, Cham, 2014.
- [3] D. P. Bertsekas, *Nonlinear Programming*, Athena Scientific, Nashua, 2nd edition, 1999.
- [4] J. F. Bonnans and A. Shapiro, Optimization problems with perturbations: A guided tour, *SIAM Rev.* 40 (1998), 228–264.
- [5] J. F. Bonnans and A. Shapiro, *Perturbation Analysis of Optimization Problems*, Springer Series in Operations Research, Springer, New York, 2000.
- [6] F. H. Clarke, *Optimization and Nonsmooth Analysis*, SIAM, Philadelphia, PA, 1990.
- [7] A. R. Conn, K. Scheinberg, and L. N. Vicente, *Introduction to Derivative-Free Optimization*, SIAM, Philadelphia, 2009.
- [8] J. M. Danskin, The theory of Max–Min, with applications, *J. SIAM Appl. Math.* 14 (1966), 641–664.
- [9] S. Dempe, B. S. Mordukhovich, and A. B. Zemkoho, Sensitivity analysis for two-level value functions with applications to bilevel programming, *SIAM J. Optim.* 22 (2012), 1309–1343.
- [10] F. Facchinei, A. Fischer, and M. Herrich, An LP-Newton method: nonsmooth equations, KKT systems, and nonisolated solutions, *Math. Program., Ser. A* 146 (2014), 1–36.
- [11] F. Facchinei and J. S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems*, Springer-Verlag New York, Inc., New York, NY, 2003.

- [12] A. Griewank, Automatic directional differentiation of nonsmooth composite functions, in *Recent Developments in Optimization, French-German Conference on Optimization*, Springer, Dijon, 1994.
- [13] A. Griewank and A. Walther, *Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation*, Other Titles in Applied Mathematics, SIAM, Philadelphia, PA, 2nd edition, 2008.
- [14] P. Hartman, *Ordinary Differential Equations (2nd ed.)*, Classics in Applied Mathematics, SIAM, 2002.
- [15] J. B. Hiriart-Urruty and C. Lemaréchal, *Convex Analysis and Minimization Algorithms I: Fundamentals*, A Series of Comprehensive Studies in Mathematics, Springer-Verlag, Berlin, 1993.
- [16] J. B. Hiriart-Urruty and C. Lemaréchal, *Convex Analysis and Minimization Algorithms II: Advanced Theory and Bundle Methods*, A Series of Comprehensive Studies in Mathematics, Springer-Verlag, Berlin, 1993.
- [17] W. Hogan, Directional derivatives for extremal-value functions with applications to the completely convex case, *Oper. Res.* 21 (1973), 188–209.
- [18] C. Imbert, Support functions of the Clarke generalized Jacobian and of its plenary hull, *Nonlinear Anal.-Theor.* 49 (2002), 1111–1125.
- [19] K. A. Khan, *Sensitivity analysis for nonsmooth dynamic systems*, PhD thesis, Massachusetts Institute of Technology, 2015.
- [20] K. A. Khan, Relating lexicographic smoothness and directed subdifferentiability, *Set-Valued Var. Anal.* 25 (2017), 233–244.
- [21] K. A. Khan, Branch-locking AD techniques for nonsmooth composite functions and nonsmooth implicit functions, *Optim. Method Softw.* 33 (2018), 1127–1155.
- [22] K. A. Khan, Subtangent-based approaches for dynamic set propagation, in *2018 IEEE Conference on Decision and Control*, IEEE, Miami Beach, 2018.
- [23] K. A. Khan and P. I. Barton, Generalized derivatives for solutions of parametric ordinary differential equations with non-differentiable right-hand sides, *J. Optimiz. Theory App.* 163 (2014), 355–386.
- [24] K. A. Khan and P. I. Barton, A vector forward mode of automatic differentiation for generalized derivative evaluation, *Optim. Method Softw.* 30 (2015), 1185–1212.
- [25] K. A. Khan and P. I. Barton, Generalized derivatives for hybrid systems, *IEEE T. Automat. Contr.* 62 (2017), 3193–3208.
- [26] K. C. Kiwiel, *Methods of Descent for Nondifferentiable Optimization*, Lecture Notes in Mathematics, Springer-Verlag, Berlin, 1985.
- [27] A. Y. Kruger, On Fréchet subdifferentials, *J. Math. Sci.* 116 (2003), 3325–3358.
- [28] C. Lemaréchal, J. J. Strodiot, and A. Bihain, On a bundle algorithm for nonsmooth optimization, in *Nonlinear Programming 4*, O. L. Mangasarian, R. R. Meyer, and S. M. Robinson (eds.), Academic Press, New York, NY, 1981.
- [29] L. Lukšan and J. Vlček, A bundle-Newton method for nonsmooth unconstrained minimization, *Math. Program.* 83 (1998), 373–391.

- [30] M. M. Mäkelä, Survey of bundle methods for nonsmooth optimization, *Optim. Method. Softw.* 17 (2002), 1–29.
- [31] B. S. Mordukhovich, *Variational analysis and generalized differentiation I: Basic theory*, Springer, Berlin, 2006.
- [32] Y. Nesterov, Lexicographic differentiation of nonsmooth functions, *Math. Program. B* 104 (2005), 669–700.
- [33] A. Neumaier, *Interval Methods for Systems of Equations*, Cambridge University Press, Cambridge, 1990.
- [34] J. S. Pang and D. E. Stewart, Solution dependence on initial conditions in differential variational inequalities, *Math. Program. B* 116 (2009), 429–460.
- [35] L. Qi, Convergence analysis of some algorithms for solving nonsmooth equations, *Math. Oper. Res.* 18 (1993), 227–244.
- [36] L. Qi and J. Sun, A nonsmooth version of Newton’s method, *Math. Program.* 58 (1993), 353–367.
- [37] R. T. Rockafellar and R. J. B. Wets, *Variational Analysis*, A Series of Comprehensive Studies in Mathematics, Springer, Berlin, 1998.
- [38] G. Rote, The convergence rate of the sandwich algorithm for approximating convex functions, *Computing* 48 (1992), 337–361.
- [39] S. Scholtes, *Introduction to Piecewise Differentiable Equations*, SpringerBriefs in Optimization, Springer, New York, NY, 2012.
- [40] N. Z. Shor, *Minimization Methods for Non-Differentiable Functions*, Springer Series in Computational Mathematics, Springer-Verlag, Berlin, 1985.
- [41] S. S. Skiena, Problems in geometric probing, *Algorithmica* 4 (1989), 599–605.
- [42] P. G. Stechlinski and P. I. Barton, Generalized derivatives of differential-algebraic equations, *J. Optim. Theory App.* 171 (2016), 1–26.
- [43] P. G. Stechlinski, K. A. Khan, and P. I. Barton, Generalized sensitivity analysis of nonlinear programs, *SIAM J. Optim.* 28 (2018), 272–301.
- [44] A. Tsoukalas and A. Mitsos, Multivariate McCormick relaxations, *J. Glob. Optim.* 59 (2014), 633–662.