Conservative and Semismooth Derivatives are Equivalent for Semialgebraic Maps

Damek Davis[1] · Dmitriy Drusvyatskiy[2]

Abstract
Subgradient and Newton algorithms for nonsmooth optimization require generalized derivatives to satisfy subtle approximation properties: conservativity for the former and semismoothness for the latter. Though these two properties originate in entirely different contexts, we show that in the semi-algebraic setting they are equivalent. Both properties for a generalized derivative simply require it to coincide with the standard directional derivative on the tangent spaces of some partition of the domain into smooth manifolds. An appealing byproduct is a new short proof that semi-algebraic maps are semismooth relative to the Clarke Jacobian.

Keywords Semismoothness · Conservativity · Semialgebraic · Generalized differentiation

Mathematics Subject Classification (2010) Primary: 49J53 · 49J52 · Secondary: 32B20 · 14P15

1 Introduction

Algorithms for nonsmooth optimization, such as subgradient and semi-smooth Newton methods, crucially rely on first-order approximations of Lipschitz maps. Two seemingly disparate first-order approximation properties underlie existing results.

Conservativity Infull generality, Lipschitz continuous functions can be highly pathological, resulting in failure of subgradient methods to find any critical points [8]. Consequently, it is essential to limit the class of functions under consideration. With this in mind, recent analysis of subgradient methods [3, 9, 15] requires *conservativity* of a generalized gradient

---

✉ Damek Davis
dsd95@cornell.edu

[1] School of ORIE, Cornell University, Ithaca, NY 14850, USA

[2] Department of Mathematics, University of Washington, Seattle, WA 98195, USA

mapping $G(x)$ for f. This property, introduced by Bolte and Pauwels [3], stipulates the validity of a formal chain rule

$$\frac{d}{dt}(f \circ \gamma)(t) = \langle G(\gamma(t)), \dot{\gamma}(t) \rangle \qquad \text{for a.a. } t \in (0, 1),$$

along any absolutely continuous curve γ. Most importantly, conservativity holds automatically when f is semialgebraic and G is the Clarke subdifferential $\partial_C f$ [9]. More generally, Bolte and Pauwels [3, 4] showed that conservativity holds when G is an output of an automatic differentiation scheme on a composition of semialgebraic functions.

Semismoothness Newton methods for solving nonsmooth equations $F(x) = 0$, for a Lipschitz map F require a *semismoothness* property of F with respect to a generalized Jacobian mapping $G(x)$. This property, introduced by Mifflin [16] and explored for Newton methods by Qi and Sun [21], stipulates the estimate:

$$F(y) - F(x) - G(y)(y - x) \subset o(\|x - y\|)\mathbb{B} \qquad \text{as } y \to x,$$

at any point $x \in \mathbf{R}^n$. Bolte, Daniilidis and Lewis [1] famously showed that semismoothness holds when F is semialgebraic and $G = \partial_C F$ is the Clarke Jacobian.

Seemingly disparate, there is reason to believe conservativity and semismoothness are closely related. For example, Norkin's seminal work [18, 19] analyzed subgradient methods on functions f that are semismooth with respect to certain generalized subdifferential operators. On the other hand, Ruszczyński [23] recently showed that semismoothness implies a weaker notion of conservativity along semismooth curves. In this paper, we prove that the notions of conservativity and semismoothness are equivalent for semialgebraic maps. Our proof strategy is to relate the two properties to an intermediate "stratified derivative" condition, already known from [14] to be equivalent to conservativity. The end result applies to a natural family of (set-valued) directional derivatives $D(x, u)$, akin to $G(x)u$. In particular, $D(x, u)$ could be a directional derivative, the map induced by the coordinate-wise Clarke Jacobian, or the output of an automatic differentiation procedure.

Theorem 1.1 (Informal) *Suppose $F(\cdot)$ and $D(\cdot, \cdot)$ are semialgebraic. Then conservativity and semismoothness are both equivalent to the following condition: there exists a partition of $\mathbf{R}^d$ into finitely many semialgebraic C^1 manifolds such that for any point x lying in a manifold M equality holds:*

$$D(x, u) = \{F'(x, u)\} \qquad \forall u \in T_M(x),$$

Thus both conservativity and semismoothness hold if and only if $D(x, u)$ coincides with the standard directional derivative $F'(\cdot, \cdot)$ on the tangent spaces of some partition of $\mathbf{R}^n$ into finitely many smooth manifolds. In dual terms, this means that semismooth generalized derivatives are simply selections of the map $(x, u) \to J(x)u$ where the rows of $J(x)$ are Clarke subgradients of F_i shifted by the normal space. Building on [3], the authors of [14] recently showed that the latter property is equivalent to $J(x)$ being conservative. Summarizing, the results of our paper together with those already available in the literature [3, 14] imply that semismooth directional derivatives, stratified directional derivatives/subgradients, and conservative set-valued vector fields are equivalent in the semi-algebraic setting. We emphasize that the paper [2] plays a foundational role in our work, both introducing stratifications of graphs and establishing the projection formula, which we heavily use here.

2 Notation and Preliminaries

In this section, we record basic notation and preliminaries from variational analysis and semi-algebraic geometry that will be used in the paper.

2.1 Variational Analysis

We follow standard notation of variational analysis, as set out for example in the monographs [5, 12, 13, 17, 20, 22]. Throughout, fix a Euclidean n-dimension space, denoted by $\mathbf{R}^n$, equipped with an inner product $\langle \cdot, \cdot \rangle$ and the induced norm $\|x\| = \sqrt{\langle x, x \rangle}$. A set-valued map $G \colon \mathbf{R}^n \to \mathbf{R}^m$, maps points to subsets $G(x) \subset \mathbf{R}^m$. The domain and graph of G, respectively, are defined by

$$\mathrm{dom}\ G = \{x \in \mathbf{R}^n : G(x) \neq \emptyset\} \quad \text{and} \quad \mathrm{gph}\ G = \{(x, y) \in \mathbf{R}^n \times \mathbf{R}^m : y \in G(x)\}.$$

The set $\mathrm{Limsup}_{x \to \bar{x}} G(x)$ consists of all limits of sequences $y_i \in G(x_i)$ where x_i is any sequence converging to $\bar{x}$. The map G is called outer-semicontinuous if $\mathrm{gph}\ G$ is a closed set. We say that G is inner-semicontinuous on a set $Q \subset \mathbf{R}^n$ if for any points $x \in Q$ and $y \in G(x)$ and for any sequence $x_i \xrightarrow{Q} x$, there exist points $y_i \in G(x_i)$ converging to y.

The distance function and the projection onto a set $X \subset \mathbf{R}^n$ are defined by:

$$\mathrm{dist}(y, X) := \inf_{x \in X} \|y - x\|, \qquad \mathrm{proj}(y, X) := \underset{x \in X}{\mathrm{argmin}} \|y - x\|.$$

The deviation between two nonempty compact sets $X, Y \subset \mathbf{R}^n$ will be measured by the Hausdorff distance

$$\mathrm{dist}(X, Y) := \max \left\{ \sup_{x \in X} \mathrm{dist}(x, Y), \sup_{y \in Y} \mathrm{dist}(y, X) \right\}.$$

The directional derivative of any map $F \colon \mathbf{R}^n \to \mathbf{R}^m$ is defined by

$$F'(x, u) := \lim_{t \searrow 0} \frac{F(x + tu) - F(x)}{t}. \tag{2.1}$$

The map F is called directionally differentiable if $F'(x, u)$ is well-defined, meaning the limit exists in (2.1) for every $x, u \in \mathbf{R}^n$. It is straightforward to verify that when F is locally Lipschitz continuous and directionally differentiable, equality holds:

$$F'(x, u) = \lim_{t \searrow 0,\ v \to u} \frac{F(x + tv) - F(x)}{t}$$

Abusing notation, for any curve $\gamma \colon [0, 1) \to \mathbf{R}^n$, we define the one-sided velocity $\gamma'(0) := \lim_{t \searrow 0} \frac{\gamma(t) - \gamma(0)}{t}$ and we let $\dot{\gamma}(t)$ denote the derivative of γ at any point $t \in (0, 1)$ where γ is differentiable.

Given any locally Lipschitz continuous map $F \colon \mathbf{R}^n \rightrightarrows \mathbf{R}^m$, the Clarke Jacobian of F at x is the set

$$\partial_C F(x) = \mathrm{conv} \left\{ \lim_{i \to \infty} \nabla F(x_i) : x_i \xrightarrow{\Omega} x \right\},$$

where Ω is the set of points at which F is differentiable and conv $(\cdot)$ denotes the convex hull. Finally, for any C^1 manifold $M \subset \mathbf{R}^n$ and a point $x \in M$, the symbols $T_M(x)$ and $N_M(x)$ will denote the tangent and normal spaces to M at x embedded in $\mathbf{R}^n$.

2.2 Semialgebraic Geometry

We next collect a few elementary facts from semialgebraic geometry. For details we refer the reader to [6, 7, 25]. All results in the paper hold more generally, and with identical proofs, for sets and functions definable in an o-minimal structure. We focus on the semialgebraic setting only for simplicity.

A set $Q \subset \mathbf{R}^n$ is called semialgebraic if it can be written as a union of finitely many sets defined by finitely many polynomial inequalities. A set-valued map $G \colon \mathbf{R}^n \rightrightarrows \mathbf{R}^m$ is called semi-algebraic if its graph is a semialgebraic set. Univariate semialgebraic functions are particularly simple.

Lemma 2.1 (Curves) *For any semialgebraic map* $\gamma \colon [0, 1] \to \mathbf{R}^n$, *there exists* $\epsilon > 0$ *such that* γ *is* C^1-*smooth on the open interval* $(0, \epsilon)$. *Moreover, as long as the quotients* $t^{-1}(\gamma(t) - \gamma(0))$ *are bounded for all small* $t > 0$, *the derivative* $\gamma'(0)$ *exists and equality* $\gamma'(0) = \lim_{t \searrow 0} \dot{\gamma}(t)$ *holds.*

Notice that the equality $\gamma'(0) = \lim_{t \searrow 0} \dot{\gamma}(t)$ can be interpreted as semi-smoothness of univariate semi-algebraic functions—an observation we will revisit. An immediate consequence is that locally Lipschitz semi-algebraic maps are directionally differentiable. The following theorem moreover shows that semi-algebraic maps are "generically smooth." To simplify notation, we use the term C^1 semialgebraic partition of $\mathbf{R}^n$ to mean a partition of $\mathbf{R}^n$ into finitely many semialgebraic C^1 manifolds.

Theorem 2.1 (Generic smoothness) *Let* $F \colon \mathbf{R}^n \to \mathbf{R}^m$ *be a semialgebraic map. Then there exists a* C^1-*semialgebraic partition* $\mathcal{A}$ *of* $\mathbf{R}^n$ *such that the restriction of* F *to each manifold* $M \in \mathcal{A}$ *is* C^1-*smooth.*

A useful property that blends variational analysis and semialgebraic constructions is the projection formula, proved in [2, Proposition 4].

Theorem 2.2 (Projection formula) *Let* $f \colon \mathbf{R}^n \to \mathbf{R}$ *be a semialgebraic locally Lipschitz continuous function. Then there exists a* C^1-*semialgebraic partition* $\mathcal{A}$ *of* $\mathbf{R}^n$ *such that for any point* x *lying in a manifold* $M \in \mathcal{A}$ *equality holds:*

$$\{f'(x, u)\} = \langle \partial_C f(x), u \rangle \qquad \forall u \in T_M(x).$$

The next result shows that any semialgebraic map admits a semialgebraic single-valued selection.

Lemma 2.2 (Semialgebraic selection) *Any semi-algebraic map* $G \colon \mathbf{R}^n \rightrightarrows \mathbf{R}^m$ *admits a semialgebraic selection* $g \colon \mathrm{dom}\, G \to \mathbf{R}$ *satisfying* $g(x) \in G(x)$ *for all* $x \in \mathrm{dom}\, G$.

The following theorem, from [11, Proposition 2.28], shows that semialgebraic set-valued maps are generically inner-semicontinuous.

Theorem 2.3 (Generic inner semicontinuity) *Let $G\colon \mathbf{R}^n \rightrightarrows \mathbf{R}^m$ be a semialgebraic map. Then there exists a C^1-semialgebraic partition $\mathcal{A}$ of $\mathbf{R}^n$ such that the restriction $G\mid_{\mathcal{M}}$ is inner-semicontinuous for every $\mathcal{M} \in \mathcal{A}$.*

Many of the aforementioned results guarantee existence of certain partitions of $\mathbf{R}^n$. Consequently, it will be useful to refine partitions. A partition $\mathcal{A}$ of $\mathbf{R}^n$ is called *compatible* with a collection $\mathcal{B}$ of sets in $\mathbf{R}^n$ if each set in $\mathcal{B}$ is a union of some sets in $\mathcal{A}$.

Theorem 2.4 (Compatible partitions) *Let $\mathcal{B}$ be a collection of finitely many semialgebraic sets. Then there exists a C^1-semialgebraic partition $\mathcal{A}$ of $\mathbf{R}^n$ that is compatible with $\mathcal{B}$.*

3 Main Results

Throughout, we let $F\colon \mathbf{R}^n \rightarrow \mathbf{R}^m$ be a locally Lipschitz continuous and directionally differentiable map and let $D\colon \mathbf{R}^n \times \mathbf{R}^n \rightrightarrows \mathbf{R}^m$ be a set-valued map. Later, we will impose that F is semialgebraic; in this setting directional differentiability will follow from local Lipschitz continuity [1]. The reader may view $D(x, u)$ as a "generalized directional derivative" of F at x in direction u.

Assumption 1 *We introduce the following assumptions on D.*

1. **(Full domain and closed-valued)** *The image $D(x, u)$ is nonempty and compact for each $u \in \mathbf{R}^d$ for all $x, u \in \mathbf{R}^n$.*
2. **(Homogeneity)** *The map D is positively homogeneous in the second argument:*

$$D(x, 0) = \{0\} \qquad and \qquad D(x, tu) = t\,D(x, u),$$

for all $x, u \in \mathbf{R}^n$ and $t > 0$.

3. **(Lipschitz continuity)** *The assignment $D(x, \cdot)$ is Lipschitz continuous locally uniformly in x. That is, for every point $\bar{x} \in \mathbf{R}^n$, there exists $L > 0$ such that*

$$dist(D(x, u_1), D(x, u_2)) \leq L\|u_1 - u_2\|,$$

for all $u_1, u_2 \in \mathbf{R}^n$ and all x sufficiently close to $\bar{x}$.

The first condition is self-explanatory. The second condition, which asserts that gph $D(x, \cdot)$ is a cone, is natural for directional derivatives. The third condition is more nuanced but is again mild. In particular, all three conditions hold for the directional derivative $D(x, u) = F'(x, u)$ and for any map of the form $D(x, u) = G(x)u$ where $G\colon \mathbf{R}^n \rightrightarrows \mathbf{R}^{m \times n}$ is locally bounded.

Clearly, $D(x, u)$ can be regarded as a generalized directional derivative of F only if it accurately predicts the variations of F at x in direction u. There are a number of seemingly distinct conditions in the literature that model "goodness of approximation," depending on context. We record the most relevant ones for us below and comment on each.

Conditions 1 *We introduce the following conditions.*

1. **(Semismooth I)** *For any point x it holds:*

$$\operatorname*{Limsup}_{y \rightarrow x} \frac{F(y) - F(x) - D(y; y - x)}{\|y - x\|} = \{0\},$$

2. **(Semismooth II)** *For any point x it holds:*

$$\underset{y \to x}{\text{Limsup}} \frac{F(y) - F(x) + D(y; x - y)}{\|y - x\|} = \{0\}.$$

3. **(Conservative derivative)** *For any absolutely continuous curve $\gamma : [0, 1] \to \mathbf{R}^n$, equality holds:*

$$\left\{\frac{d}{dt}(F \circ \gamma)(t)\right\} = D(\gamma(t), \dot{\gamma}(t)) \qquad \text{for a.a. } t \in (0, 1). \tag{3.1}$$

4. **(Stratified derivative)** *There exists a semialgebraic C^1 partition $\mathcal{A}$ of $\mathbf{R}^n$ such that equality*

$$D(x, u) = \{F'(x, u)\}$$

holds for any manifold $M \in \mathcal{A}$, $x \in M$, and any tangent vector $u \in T_M(x)$.

5. **(Stratified subdifferential)** *There exists a semialgebraic C^1 partition $\mathcal{A}$ of $\mathbf{R}^n$ such that for any manifold $M \in \mathcal{A}$ and $x \in M$, it holds:*

$$D(x, u) \subset J(x)u,$$

where we set

$$J(x) = \{A \in \mathbf{R}^{m \times n} : A_i \in \partial_C F_i(x) + N_M(x)\},$$

and A_i denote the rows of A.

The theorem identifies the largest possible example of semismooth, conservative, and stratified derivatives. These are simply Clarke subdifferentials plus normals to the manifolds that form a partition of $\mathbf{R}^n$. Let us discuss these conditions in turn.

Semismoothness The first two conditions, 1 and 2, play a central role for ensuring super-linear convergence of Newton-type algorithms for nonsmooth equations. We refer the reader the monograph [13, Chapter 10] for details. To place these conditions in context, recall that any locally Lipschitz and directionally differentiable map F satisfies the first-order approximation property [24]:[1]

$$F(y) = F(x) + F'(x, y - x) + o(\|y - x\|). \tag{3.2}$$

Condition 1 instead asserts

$$F(y) - F(x) - D(y, y - x) \subset o(\|y - x\|)\mathbf{B}.$$

Notice that contrary to (3.2), the value $D(y, y - x)$ is computed at the basepoint y. Condition 2 asserts instead

$$F(x) - F(y) - D(y, x - y) \subset o(\|y - x\|)\mathbf{B}.$$

Clearly, when D has the form $D(x, u) = G(x)u$ for some set-valued map $G : \mathbf{R}^n \to \mathbf{R}^{m \times n}$, conditions 1 and 2 are equivalent. In particular, if F is directionally differentiable and $G(x) = \partial_C F(x)$ is the Clarke generalized Jacobian, both conditions reduce to the semismoothness property in the sense of [16, 21]. In the context of optimization, sim-ilar conditions have been used by Norkin [18, 19] to establish asymptotic convergence guarantees of subgradient methods.

---

[1] Here $o = o_x$ refers to any function satisfying $t^{-1}o(t) \to 0$ as $t \to 0$.

Conservative Derivative Condition 3 asserts that the generalized directional derivative D satisfies a formal chain rule at almost every point along any absolutely continuous path γ. Equivalently, we may write condition (3.1) as:

$$\{F'(\gamma(t), \dot{\gamma}(t))\} = D(\gamma(t), \dot{\gamma}(t)) \qquad \text{for a.a. } t \in (0, 1).$$

This property is equivalent to the conservative derivatives $G(x)$ introduced in [3], in the setting $D(x, u) = G(x)u$. Such generalized derivatives play a key role in justifying automatic differentiation techniques for nonsmooth functions [4], as well as for analyzing the asymptotic behavior of subgradient methods [9].

Stratified Derivative/Subdifferential Condition 4 is geometrically intuitive. It simply asserts that there exists a partition of $\mathbf{R}^n$ into finitely many smooth manifolds, so that $D(x, \cdot)$ coincides with the directional derivative in directions tangent to the manifold containing the point x. Condition 5 can be interpreted as a "dual" counterpart of 4. Namely it stipulates that $D(x, u)$ is a selection of the map $(x, u) \mapsto J(x)u$, where the rows of $J(x)$ consist of the Clarke subdifferentials of the component functions F_i shifted by a normal space $N_M(x)$. The "duality" between conditions 4 and 5 is explored in [2].

3.1 Equivalence of 1–5

The goal of our paper is to prove that if Assumption 1 holds and F and D are semialgebraic, then conditions 1–5 are equivalent. The authors of [14], building on [3], recently proved the equivalence $3 \Leftrightarrow 5$ for maps of the form $D(x, u) = G(x)u$ where $G(\cdot)$ locally bounded and outer-semicontinuous. Therefore we claim no originality with respect to the equivalence $3 \Leftrightarrow 5$.

We begin by proving that 3 implies 1 and 2. The following simple lemma will be useful.

Lemma 3.1 *Condition 3 implies that for any absolutely continuous curve $\gamma : [0, 1] \to \mathbf{R}^n$, equality holds:*

$$D(\gamma(t), \dot{\gamma}(t)) = -D(\gamma(t), -\dot{\gamma}(t)) \qquad \text{for a.a. } t \in (0, 1).$$

Proof Fix an absolutely continuous curve $\gamma : [0, 1] \to \mathbf{R}^n$ and define $\alpha(t) = \gamma(1 - t)$. Then $\dot{\alpha}(t) = -\dot{\gamma}(1 - t)$ and $(F \circ \alpha)'(t) = -(F \circ \gamma)'(1 - t)$ for a.a. $t \in (0, 1)$. Plugging in α in place of γ in (3.1) completes the proof. $\square$

Theorem 3.2 (Conservative derivatives and semismoothness) *Suppose that F and D are semialgebraic and that Assumption 1 holds. Then condition 3 implies both 1 and 2. Moreover, both implications are true if 3 only holds with respect to semialgebraic curves γ.*

Proof Suppose condition 3 holds; we aim to verify 1. To this end, fix a point x and assume without loss of generality $x = 0$ and $F(x) = 0$. We aim to prove

$$\operatorname*{Limsup}_{d \to 0} \frac{F(d) - D(d, d)}{\|d\|} = \{0\}.$$

For any $t \in [0, 1]$ define the semialgebraic function

$$\varphi(t) := \sup_{(d,v): \|d\|=t, \, v \in D(d,d)} \|F(d) - v\|.$$

Condition 1 will follow immediately once we establish $\varphi'(0) = 0$. Clearly, we may assume $\varphi(t) > 0$ for all small $t > 0$, since otherwise by Lemma 2.1 the equality $\varphi'(0) = 0$ holds trivially. Lemma 2.2 yields semialgebraic curves $d(\cdot)$ and $v(\cdot)$ satisfying $\|d(t)\| = t$, $v(t) \in D(d(t), d(t))$ and $\|F(d(t)) - v(t)\| \geq \frac{1}{2}\varphi(t)$ for all small $t > 0$. Observe $d(0) = 0$ and $\|d(t)\|/t = 1$. Therefore Lemma 2.1 shows that $d'(0)$ exists and $\lim_{t \to 0} \dot{d}(t) = d'(0)$.

Note the inclusion $v(t)/t \in D(d(t), \frac{d(t)}{t})$. Local Lipschitz continuity of $D(x, \cdot)$ at $x = d(0) = 0$, implies that there exists $L > 0$ and a semialgebraic curve $w(t)$ satisfying $w(t) \in D(d(t), \dot{d}(t))$ with

$$\left\| w(t) - \frac{v(t)}{t} \right\| \leq L \left\| \frac{d(t)}{t} - \dot{d}(t) \right\|$$

for all small $t > 0$. Since the right side tends to zero as $t \searrow 0$, we compute

$$\varphi'(0) = \lim_{t \searrow 0} \frac{\varphi(t)}{t} \leq \limsup_{t \searrow 0} \frac{2\|F(d(t)) - v(t)\|}{t} = 2 \cdot \limsup_{t \searrow 0} \left\| \frac{F(d(t))}{t} - w(t) \right\|.$$

Lemma 2.1 and condition 3 guarantee

$$\lim_{t \searrow 0} w(t) = \lim_{t \searrow 0} (F \circ d)'(t) = \lim_{t \searrow 0} \frac{F(d(t))}{t},$$

Note that all the limits in the displayed equation exist due to the local monotonicity of semialgebraic functions. We conclude $\varphi'(0) = 0$ and therefore condition 1 holds. The proof of condition 2 follows by identical reasoning, while taking into account Lemma 3.1. $\square$

Combining Theorem 3.2 with Theorem 2.2 immediately shows that semi-algebraic locally Lipschitz maps are semismooth with respect to the Clarke generalized Jacobian. This result original appeared in [1] with a different proof.

Corollary 3.3 (Semi-algebraic maps are semismooth) *Any locally Lipschitz semi-algebraic map $F: \mathbf{R}^n \rightrightarrows \mathbf{R}^m$ satisfies condition 1 for the map $D(x, u) = \partial_C F(x)u$.*

Proof Note the inclusion $\partial_C F(x) \subset \{A \in \mathbf{R}^{m \times n} : A_i \in \partial_C F_i(x)\}$. Let $\mathcal{A}_i$ be the partition ensured by Theorem 2.2 for each coordinate function F_i. Using Theorem 2.4, let $\mathcal{A}$ be a semialgebraic partition that is compatible with each $\mathcal{A}_i$. Fix a semialgebraic curve γ. Semialgebraicity implies that for any $M \in \mathcal{A}$ and for a.a. $t \in (0, 1)$ with $\gamma(t) \in M$, the inclusion $\dot{\gamma}(t) \in T_M(\gamma(t))$ holds. Consequently, by Theorem 2.2, condition 3 holds with respect to all semialgebraic curves. An application of Theorem 3.2 completes the proof. $\square$

The proof of the full equivalence between conditions 1–5 will make use of the following simple linear algebraic fact.

Lemma 3.4 *For any sets $A, B \subset \mathbf{R}^n$ and a subspace $V \subset \mathbf{R}^n$, the equivalence holds:*

$$A \subset B + V \qquad \Longleftrightarrow \qquad \text{proj}(A, V^{\perp}) \subset \text{proj}(B, V^{\perp}).$$

We now have all the ingredients to prove the main the result of the paper.

Theorem 3.5 (Equivalence) *Suppose that F and D are semialgebraic and that Assumption 1 holds. Then conditions 1-5 are equivalent.*

Proof We first establish the equivalences $2 \Leftrightarrow 3 \Leftrightarrow 4 \Leftrightarrow 5$ by verifying the implications $3 \Rightarrow 2 \Rightarrow 4 \Rightarrow 5 \Rightarrow 3$.

Implication $3 \Rightarrow 2$: This was proved in Theorem 3.2.

Implication $2 \Rightarrow 4$: Theorems 2.1 and 2.3 yield a partition $\mathcal{A}$ of $\mathbf{R}^n$ into finitely many semialgebraic C^1 manifolds such that when restricted to each manifold $M \in \mathcal{A}$, the function F is C^1-smooth and the map

$$x \mapsto \mathrm{gph}\ D(x, \cdot)$$

is inner semicontinuous. Fix a manifold $M \in \mathcal{A}$, a point $x \in M$, unit tangent vector $u \in T_M(x)$, and $v \in D(x, u)$. Since $-u$ also lies in $T_M(x)$, we may find sequences $x_i \xrightarrow{M} x$ and $\tau_i \searrow 0$ with $\tau_i^{-1}(x_i - x) \to -u$. By inner-semicontinuity, there exist sequences (u_i, v_i) converging to (u, v) and satisfying $v_i \in D(x_i, u_i)$. The assumed condition 2 therefore guarantees

$$\{0\} = \underset{i \to 0}{\mathrm{Limsup}}\ \frac{F(x_i) - F(x) + D(x_i, x - x_i)}{\|x_i - x\|}.$$

Rearranging and using linearity of $F'(x, \cdot)$ on $T_M(x)$ we deduce[2]

$$\{F'(x, u)\} = \{-F'(x, -u)\} = \underset{i \to \infty}{\mathrm{Limsup}}\ D\left(x_i, \frac{x - x_i}{\|x - x_i\|}\right). \tag{3.3}$$

Taking into account local Lipschitz continuity of $D(y, \cdot)$ uniformly for all y near x, we deduce that there exists $L > 0$ and a sequence $z_i \in D\left(x_i, \frac{x - x_i}{\|x - x_i\|}\right)$ satisfying

$$\left\| z_i - \frac{v_i}{\|u_i\|}\right\| \leq L \left\| \frac{x - x_i}{\|x - x_i\|} - \frac{u_i}{\|u_i\|}\right\|,$$

for all large indices i. Observe that the right side tends to zero. Continuing (3.3), we deduce

$$\{F'(x, u)\} = \lim_{i \to \infty} z_i = \lim_{i \to \infty} \frac{v_i}{\|u_i\|} = \{v\}.$$

We have thus proved $D(x, u) = \{F'(x, u)\}$ for all $u \in T_M(x)$, and therefore 4 holds.

Implication $4 \Rightarrow 5$: Let $\mathcal{A}$ be the partition of $\mathbf{R}^n$ into C^1 semialgebraic manifolds stipulated by condition 4. Using Theorem 2.2 for each coordinate function F_i and refining the partitions using Theorem 2.4, we arrive at a finite partition $\mathcal{A}'$ of $\mathbf{R}^n$ into C^1 semialgebraic manifolds that is compatible with $\mathcal{A}$ and satisfies

$$\{F_i'(x, u)\} = \langle \partial_C F_i(x), u \rangle,$$

---

[2]Note that $F'(x; \cdot)$ is linear on $T_M(x)$. Indeed, since F is semialgebraic, it is directionally differentiable. Therefore, for any $u \in T_M(x)$, the directional derivative $F'(x; u)$ coincides with the directional derivative (in the sense of manifolds) of the restriction of F to M at x in the direction u.

whenever x lies in $M \in \mathcal{A}'$ and $u \in T_M(x)$ is a tangent vector. Since for any such vector u, equality $\{F'(x, u)\} = D(x, u)$ holds by 4, we conclude

$$D(x, u) = J(x)u \qquad \forall u \in T_M(x).$$

On the other hand, for any $u \notin T_M(x)$, we trivially have

$$J_i(x)u = \langle \partial_C F_i(x), u \rangle + \langle N_M(x), u \rangle = \langle \partial_C F_i(x), u \rangle + \mathbf{R} = \mathbf{R},$$

where $J_i(x)$ denotes the set of all i'th rows of matrices in $J(x)$. Therefore, in this case, the inclusion $D(x, u) \subset J(x)u$ holds trivially.

Implication $5 \Rightarrow 3$: Let $\mathcal{A}$ be the partition of $\mathbf{R}^n$ into C^1 semialgebraic manifolds stipulated by condition 5. Refining the partition using Theorems 2.2 and 2.4, we may ensure that the equality holds:

$$\{F'(x, u)\} = J(x)u \qquad \forall u \in T_M(x),$$

whenever x lies in some manifold $M \in \mathcal{A}$.

Consider now any absolutely continuous curve $\gamma : [0, 1] \to \mathbf{R}^n$. It is elementary to verify that for a.a. $t \in (0, 1)$ the implication holds (e.g. [10, Lemma 4.13]):

$$\gamma(t) \in M \implies \dot{\gamma}(t) \in T_M(\gamma(t)). \tag{3.4}$$

We conclude that for a.a. $t \in (0, 1)$ and $M \in \mathcal{A}$ satisfying $\gamma(t) \in M$, we have

$$\left\{ \frac{d}{dt}(F \circ \gamma)(t) \right\} = \{F'(\gamma(t), \dot{\gamma}(t))\} = J(\gamma(t))\dot{\gamma}(t) \supset D(\gamma(t), \dot{\gamma}(t)),$$

as claimed.

Summarizing, we have proved the equivalence $2 \Leftrightarrow 3 \Leftrightarrow 4 \Leftrightarrow 5$. Next, observe that 5 holds for a map $D(\cdot, \cdot)$ if and only if it holds for the map $\hat{D}(x, u) := -D(x, -u)$. Noting that condition 2 for $\hat{D}$ coincides with condition 1 for D completes the proof of the theorem. $\square$

References

1. Bolte, J., Daniilidis, A., Lewis, A.: Tame functions are semismooth. Math. Program. **117**(1), 5–19 (2009)
2. Bolte, J., Daniilidis, A., Lewis, A., Shiota, M.: Clarke subgradients of stratifiable functions. SIAM J. Optim. **18**(2), 556–572 (2007)
3. Bolte, J., Pauwels, E.: Conservative set valued fields, automatic differentiation, stochastic gradient methods and deep learning. Mathematical Programming (2020)
4. Bolte, J., Pauwels, E.: A mathematical model for automatic differentiation in machine learning. arXiv:2006.02080 (2020)
5. Clarke, F.H., Ledyaev, Yu., Stern, R.I., Wolenski, P.R.: Nonsmooth Analysis and Control Theory Texts in Math, vol. 178. Springer, New York (1998)
6. Coste, M.: An introduction to o-minimal geometry. RAAG Notes, 81 pages. Institut de Recherche Mathématiques de Rennes (1999)

7. Coste, M.: An introduction to semialgebraic geometry. RAAG notes, 78 pages, Institut de Recherche Mathématiques de Rennes (2002)
8. Daniilidis, A., Drusvyatskiy, D.: Pathological subgradient dynamics. SIAM J. Optim. **30**(2), 1327–1338 (2020)
9. Davis, D., Drusvyatskiy, D., Kakade, S., Lee, J.D.: Stochastic subgradient method converges on tame functions. Foundations of Computational Mathematics **20**(1), 119–154 (2020)
10. Drusvyatskiy, D., Ioffe, A., Lewis, A.: Curves of descent. SIAM J. Control. Optim. **53**(1), 114–138 (2015)
11. Drusvyatskiy, D., Lewis, A.: Semi-algebraic functions have small subdifferentials. Math. Program. **140**(1), 5–29 (2013)
12. Ioffe, A.D.: Variational Analysis of Regular Mappings. Springer Monographs in Mathematics. Springer, Cham (2017). Theory and applications
13. Klatte, D., Kummer, B.: Nonsmooth Equations in Optimization, volume 60 of Nonconvex Optimization and its Applications. Kluwer Academic Publishers, Dordrecht (2002). Regularity, calculus, methods and applications
14. Lewis, A., Tian, T.: The structure of conservative gradient fields. arXiv:2101.00699 (2021)
15. Majewski, S., Miasojedow, B., Moulines, E.: Analysis of nonsmooth stochastic approximation: the differential inclusion approach. arXiv:1805.01916 (2018)
16. Mifflin, R.: Semismooth and semiconvex functions in constrained optimization. SIAM J. Control. Optim. **15**(6), 959–972 (1977)
17. Mordukhovich, B.S.: Variational Analysis and Generalized Differentiation I: Basic Theory. Grundlehren der mathematischen Wissenschaften, vol. 330. Springer, Berlin (2006)
18. Norkin, V.: Generalized-differentiable functions. Cybernetics **16**(1), 10–12 (1980)
19. Norkin, V.: Stochastic generalized-differentiable functions in the problem of nonconvex nonsmooth stochastic optimization. Cybernetics **22**(6), 804–809 (1986)
20. Penot, J.-P.: Calculus without derivatives, volume 266 of Graduate Texts in Mathematics. Springer, New York (2013)
21. Qi, L., Sun, J.: A nonsmooth version of Newton's method. Mathematical Programming **58**(1-3), 353–367 (1993)
22. Rockafellar, R.T., Wets, R.J.-B.: Variational Analysis Grundlehren der mathematischen Wissenschaften, vol. 317. Springer, Berlin (1998)
23. Ruszczyński, A.: Convergence of a stochastic subgradient method with averaging for nonsmooth nonconvex constrained optimization. Optim. Lett. 1–11 (2020)
24. Shapiro, A.: On concepts of directional differentiability. J. Optim. Theory Appl. **66**(3), 477–487 (1990)
25. van den Dries, L., Miller, C.: Geometric categories and o-minimal structures. Duke Math. J. **84**, 497–540 (1996)