Semi-algebraic functions have small subdifferentials

D. Drusvyatskiy* A.S. Lewis†

March 31, 2010

Abstract

We prove that the subdifferential of any semi-algebraic extended-real-valued function on $\mathbf{R}^n$ has n-dimensional graph. We discuss consequences for generic semi-algebraic optimization problems.

1 Introduction

A principal goal of variational analysis is the search for generalized critical points of nonsmooth functions $f\colon \mathbf{R}^n \to \mathbf{R}$. For example, given a locally Lipschitz function f, we might be interested in points $x \in \mathbf{R}^n$ having zero in the "Clarke generalized gradient" (or "subdifferential") $\partial_c f(x)$, a set consisting of all limits of convex combinations of gradients of f at points near x [12].

Adding a linear perturbation, we might seek critical points of the function $x \mapsto f(x) - v^T x$ for a given vector $v \in \mathbf{R}^m$, or, phrased in terms of the graph of the subdifferential mapping $\partial_c f$, solutions to the inclusion

$$(x, v) \in \operatorname{gph} \partial_c f.$$

*Department of Operations Research and Information Engineering, Cornell University, Ithaca, New York, USA; dd379@cornell.edu. Work of Dmitriy Drusvyatskiy on this paper has been partially supported by the NDSEG grant from the Department of Defense.

†School of Operations Research and Information Engineering, Cornell University, Ithaca, New York, USA; http://people.orie.cornell.edu/ aslewis/. Research supported in part by National Science Foundation Grant DMS-0806057.

More generally, given a smooth function $G\colon \mathbf{R}^m \to \mathbf{R}^n$, we might be interested in solutions $(x, y) \in \mathbf{R}^m \times \mathbf{R}^n$ to the system

$$(G(x)), y) \in \operatorname{gph} \partial_c f \quad \text{and} \quad \nabla G(x)^* y = v \tag{1}$$

(where $*$ denotes the adjoint). Such systems arise naturally when we seek critical points of the composite function $x \mapsto f(G(x)) - v^T x$.

Generalized critical points of *smooth* functions f are, of course, simply the critical points in the classical sense. However, the more general theory is particularly interesting to optimization specialists, because critical points of continuous convex functions are just minimizers [27, Proposition 8.12], and more generally, for a broader class of functions (for instance, those that are Clarke regular [12]), a point is critical exactly when the directional derivative is nonnegative in every direction.

The system (1) could, in principal, be uninformative if the graph $\operatorname{gph} \partial_c f$ is large. In particular, if the dimension (appropriately defined) of the graph is larger than n, then we could not typically expect the system to be a very definitive tool, since it involves $m + n$ variables constrained by only m linear equations and the inclusion. Such examples are not hard to construct: indeed, there exists a function $f\colon \mathbf{R} \to \mathbf{R}$ with Lipschitz constant one and with the property that its Clarke subdifferential is the interval $[-1, 1]$ at every point [12]. Alarmingly, in a precise mathematical sense, this property is actually typical for such functions [9].

Optimization theorists often consider subdifferentials that are smaller than Clarke's, the "limiting" subdifferential ∂f being a popular choice [27, 13, 23, 10]. However, the potential difficulty posed by functions with large subdifferential graphs persists [8].

Notwithstanding this pathology, concrete functions $f\colon \mathbf{R}^n \to \mathbf{R}$ encountered in practice have subdifferentials $\partial_c f$ whose graphs are, in some sense, small. As a consequence, for instance, Robinson [26] considers algorithmic aspects of functions whose subdifferentials are *thin*: their subdifferential graphs are everywhere locally Lipschitz homeomorphic to an open subset of $\mathbf{R}^n$. As above, dimensional considerations suggest reassuringly that this property should help the definitive power of critical point systems like (1), and Robinson furthermore argues that an assumption of thinness carries powerful computational promise.

When can we be confident that a function has a subdifferential graph that is, in some sense, small? The study of classes of functions that are favorable for subdifferential analysis, in particular excluding the pathological examples above, is well-developed. The usual starting point is a unification of smooth and convex analysis, arriving at such properties as amenability [27, Chapter 10.F.], prox-regularity [25], and cone-reducibility [6, Section 3.4.4]. Poliquin and Rockafellar [25] showed that prox-regular functions, in particular, have thin subdifferentials. Aiming precisely at a class of functions with small subdifferentials (in fact minimal in the class of outer semicontinuous mappings with nonempty compact convex images), [7] considers "essential strict differentiability".

In this work we take a different, very concrete approach. We focus on the dimension of the subdifferential graph, unlike the abstract minimality results of [7], but we consider the class of *semi-algebraic* functions—those functions whose graphs are semi-algebraic, meaning composed of finitely-many sets, each defined by finitely-many polynomial inequalities—and prove that such functions have small subdifferentials in the sense of dimension: the Clarke subdifferential has n-dimensional graph. This result subsumes neither the simple case of a smooth function, nor the case of a convex function, neither of which is necessarily semi-algebraic. Nonetheless, it has a certain appeal: semi-algebraic functions are common and easy to recognize (as a consequence of the Tarski-Seidenberg theorem on preservation of semi-algebraicity under projection), they serve as an excellent model for "concrete" functions in variational analysis [19], and in marked contrast with prox-regular functions, such functions may not even be Clarke regular.

To illustrate, consider the critical points of the function $x \mapsto f(x) - v^T x$ for a semi-algebraic function $f\colon \mathbf{R}^n \to [-\infty, +\infty]$. As a consequence of the subdifferential graph being small, we show that for a *generic* choice of the vector v, the number of critical points is finite. More precisely, there exists a number N, and a semi-algebraic set $S \subset \mathbf{R}^n$ of dimension strictly less than n, such that for all vectors v outside S, there exist at most N critical points. A result of a similar flavor can be found in [20], where criticality of so called "constraint systems" is considered. Specifically, [20] shows that if a semi-algebraic constrained minimization problem is "normal", then it has only finitely many critical points. Furthermore, it is shown that normality is a generic property. To constrast their approach to ours, we should note that [20] focuses on perturbations to the constraint structure, whereas we address linear perturbations to the function itself.

To be concrete, we state our results for semi-algebraic sets and functions. Analogous results, with essentially identical proofs, hold for the case of “tame” sets and functions definable in an “o-minimal structure”. (In the case of tame function, the word “finite” should be replaced with “countable” in Proposition 4.3 and Corollary 4.4.) For a quick introduction to these concepts in an optimization context, see [19].

2 Preliminaries

2.1 Variational Analysis

In this section, we summarize some of the fundamental tools used in variational analysis and nonsmooth optimization. We refer the reader to the monographs Borwein-Zhu [10], Mordukhovich [23, 24], Clarke-Ledyaev-Stern-Wolenski [13], and Rockafellar-Wets [27], for more details. Unless otherwise stated, we follow the terminology and notation of [27].

Consider the extended real line $\overline{\mathbf{R}} := \mathbf{R} \cup \{-\infty\} \cup \{+\infty\}$. We say that an extended-real-valued function is proper if it is never $\{-\infty\}$ and is not always $\{+\infty\}$.

For a function $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$, we define the *domain* of f to be

$$\operatorname{dom} f := \{x \in \mathbf{R}^n : f(x) < +\infty\},$$

and we define the *graph* of f to be

$$\operatorname{gph} f := \{(x, y) \in \mathbf{R}^n \times \mathbf{R} : x \in \operatorname{dom} f, y = f(x)\}.$$

The *epigraph* of f is the set

$$\operatorname{epi} f := \{(x, r) \in \mathbf{R}^n \times \mathbf{R} : r \geq f(x)\}.$$

A *set-valued mapping* F from $\mathbf{R}^n$ to $\mathbf{R}^m$, denoted by $F\colon \mathbf{R}^n \rightrightarrows \mathbf{R}^m$, is a mapping from $\mathbf{R}^n$ to the power set of $\mathbf{R}^m$. Thus for each point $x \in \mathbf{R}^n$, $F(x)$ is a subset of $\mathbf{R}^m$. For a set-valued mapping $F\colon \mathbf{R}^n \rightrightarrows \mathbf{R}^m$, we define the *domain* of F to be

$$\operatorname{dom} F := \{x \in \mathbf{R}^n : F(x) \neq \emptyset\},$$

and we define the *graph* of F to be

$$\operatorname{gph} F := \{(x, y) \in \mathbf{R}^n \times \mathbf{R}^m : y \in F(x)\}.$$

Definition 2.1. Consider a set-valued mapping $F\colon \mathbf{R}^n \rightrightarrows \mathbf{R}^m$.

1. F is *outer semicontinuous* at a point $\bar{x} \in \mathbf{R}^n$ if for any sequence of points $x_r \in \mathbf{R}^n$ converging to $\bar{x}$ and any sequence of points $y_r \in F(x_r)$ converging to $\bar{y}$, we must have $\bar{y} \in F(\bar{x})$.

2. F is *inner semicontinuous* at $\bar{x}$ if for any sequence of points $x_r \in \mathbf{R}^n$ converging to $\bar{x}$ and any point $\bar{y} \in F(\bar{x})$, there exists a sequence $y_r \in \mathbf{R}^m$ converging to $\bar{y}$ such that $y_r \in F(x_r)$ for all r.

If both properties hold, then we say that F is *continuous* at $\bar{x}$.

Definition 2.2. Consider a set $S \subset \mathbf{R}^n$ and a point $\bar{x} \in S$. The *regular normal cone* to S at $\bar{x}$, denoted $\hat{N}_S(\bar{x})$, consists of all vectors $v \in \mathbf{R}^n$ such that

$$\langle v, x - \bar{x} \rangle \leq o(|x - \bar{x}|) \text{ for } x \in S,$$

where we denote by $o(|x - \bar{x}|)$ for $x \in S$ a term with the property that

$$\frac{o(|x - \bar{x}|)}{|x - \bar{x}|} \to 0$$

when $x \xrightarrow{S} \bar{x}$ with $x \neq \bar{x}$.

The mapping $x \mapsto \hat{N}_S(x)$ does not necessarily have a closed graph. To correct for that, the following definition is introduced.

Definition 2.3. Consider a set $S \subset \mathbf{R}^n$ and a point $\bar{x} \in S$. The *limiting normal cone* to S at $\bar{x}$, denoted $N_S(\bar{x})$, consists of all $v \in \mathbf{R}^n$ such that there are sequences $x_r \xrightarrow{S} \bar{x}$ and $v_r \to v$ with $v_r \in \hat{N}_S(x_r)$.

For a set $S \subset \mathbf{R}^n$, we denote its topological closure by $\operatorname{cl} S$ and its convex hull by $\operatorname{conv} S$.

Definition 2.4. Consider a set $S \subset \mathbf{R}^n$ and a point $\bar{x} \in S$. The *Clarke normal cone* to S at $\bar{x}$, denoted $N_S^c(\bar{x})$ is defined by

$$N_S^c(\bar{x}) = \text{cl conv } N_S(\bar{x}).$$

We summarize some simple facts about normal cones that we will need.

Theorem 2.5. *Consider a set $S \subset \mathbf{R}^n$ and a point $\bar{x} \in S$.*

1. $\hat{N}_S(\bar{x}) \subset N_S(\bar{x}) \subset N^c_S(\bar{x})$.

2. $N_S(\bar{x})$, $\hat{N}_S(\bar{x})$, *and* $N^c_S(\bar{x})$ *are closed cones.* $\hat{N}_S(\bar{x})$ *and* $N^c_S(\bar{x})$ *are, in addition, convex.*

3. *Consider a set* $F \subset \mathbf{R}^n$ *containing* $\bar{x}$ *such that* $S \subset F$*. Then we have* $\hat{N}_F(\bar{x}) \subset \hat{N}_S(\bar{x})$.

Definition 2.6 (Clarke regularity of sets)**.** A set $C \subset \mathbf{R}^n$ is said to be *Clarke regular* at a point $\bar{x} \in C$ if it is locally closed at $\bar{x}$ and every limiting normal vector to C at $\bar{x}$ is a regular normal vector, that is $N_C(\bar{x}) = \hat{N}_C(\bar{x})$.

Given any set $S \subset \mathbf{R}^n$ and a mapping $f\colon S \to \widetilde{S}$, where $\widetilde{S} \subset \mathbf{R}^m$, we say that f is *smooth* if for each point $\bar{x} \in S$, there is a neighborhood U of $\bar{x}$ and a $\mathbf{C^1}$ mapping $\hat{f}\colon \mathbf{R}^n \to \mathbf{R}^m$ that agrees with f on $S \cap U$. If a smooth function f is bijective and its inverse is also smooth, then we say that f is a *diffeomorphism*.

Definition 2.7 ([22, Proposition 8.12])**.** Consider a set $M \subset \mathbf{R}^n$. We say that M is a *submanifold* of dimension r if for each point $\bar{x} \in M$, there is an open neighborhood U around $\bar{x}$ such that $M \cap U = F^{-1}(0)$, where $F\colon U \to \mathbf{R}^{n-r}$ is a smooth map with $\nabla F(\bar{x})$ of full rank. In this case we call F a *local defining function* for M around $\bar{x}$.

Theorem 2.8 ([27, Example 6.8])**.** *If* S *is a manifold, then for every point* $\bar{x} \in S$*, the manifold* S *is Clarke regular at* $\bar{x}$ *and* $N_S(\bar{x})$ *is equal to the normal space to* S *at* $\bar{x}$*, in the sense of differential geometry.*

Normal cones allow us to study geometric objects. We now define subdifferentials, which allow us to analyze behavior of functions.

Definition 2.9. Consider a function $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$ and a point $\bar{x} \in \mathbf{R}^n$ where f is finite. The *regular*, *limiting*, and *Clarke subdifferentials* of f at $\bar{x}$, respectively, are defined by

$$\hat{\partial} f(\bar{x}) = \{v \in \mathbf{R}^n : (v, -1) \in \hat{N}_{\operatorname{epi} f}(\bar{x}, f(\bar{x})\},$$

$$\partial f(\bar{x}) = \{v \in \mathbf{R}^n : (v, -1) \in N_{\operatorname{epi} f}(\bar{x}, f(\bar{x})\},$$

$$\partial_c f(\bar{x}) = \{v \in \mathbf{R}^n : (v, -1) \in N^c_{\operatorname{epi} f}(\bar{x}, f(\bar{x})\}.$$

For $x \notin \operatorname{dom} f$, we follow the convention that $\hat{\partial} f(x) = \partial f(x) = \partial_c f(x) = \emptyset$.

Definition 2.10 (Subdifferential regularity)**.** A function $f\colon \mathbf{R}^n \to \bar{\mathbf{R}}$ is called *subdifferentially regular* at $\bar{x}$ if $f(\bar{x})$ is finite and epi f is Clarke regular at $(\bar{x}, f(\bar{x}))$ as a subset of $\mathbf{R}^n \times \mathbf{R}$.

Theorem 2.11 ([27, Exercise 8.8, Corollary 10.9])**.** *Consider the function $h = f + g$, where $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$ is finite at $\bar{x}$ and $g\colon \mathbf{R}^n \to \overline{\mathbf{R}}$ is smooth on a neighborhood of $\bar{x}$. Then we have*

$$\hat{\partial} h(\bar{x}) = \hat{\partial} f(\bar{x}) + \nabla g(\bar{x}), \quad \partial h(\bar{x}) = \partial f(\bar{x}) + \nabla g(\bar{x}).$$

Furthermore, h is regular at $\bar{x}$ if and only if f is regular at $\bar{x}$.

For a set $C \subset \mathbf{R}^n$, we define $\delta_C\colon \mathbf{R}^n \to \overline{R}$ to be a function that is 0 on C and $+\infty$ elsewhere. We refer to δ_C as the *indicator function* of the set C.

Theorem 2.12 ([27, Exercise 8.14])**.** *Consider the indicator function δ_C of a set $C \subset \mathbf{R}^n$. Then we have*

$$\partial \delta_C(\bar{x}) = N_C(\bar{x}), \quad \hat{\partial} \delta_C(\bar{x}) = \hat{N}_C(\bar{x}).$$

Furthermore, δ_C is subdifferentially regular at $\bar{x}$ if and only if C is Clarke regular at $\bar{x}$.

2.2 Semi-algebraic Geometry

A *semi-algebraic* set $S \subset \mathbf{R}^n$ is a finite union of sets of the form

$$\{x \in \mathbf{R}^n : P(x) = 0, Q_1(x) < 0, \ldots, Q_l(x) < 0\}$$

where $P, Q_1, \ldots, Q_l$ are polynomials in n variables. In other words, S is a union of finitely many sets, each defined by finitely many polynomial equalities and inequalities. A map $F\colon \mathbf{R}^n \rightrightarrows \mathbf{R}^m$ is *semi-algebraic* if gph $F \subset \mathbf{R}^{n+m}$ is a semi-algebraic set. Semi-algebraic sets enjoy many nice structural properties. We discuss some of these properties in this section. See the monographs of Basu-Pollack-Roy [1], Lou van den Dries [30], and Shiota [29]. For a quick survey, see the article of van den Dries-Miller [31] and the surveys of Coste [15, 14]. Unless otherwise stated, we follow the notation of [31] and [15].

The most striking and useful fact about semi-algebraic sets is that they can be partitioned into finitely many semi-algebraic manifolds that fit together in a regular pattern. The particular stratification that we are interested in is defined below.

Definition 2.13. Consider a semi-algebraic set Q in $\mathbf{R}^n$. A *Whitney stratification* of Q is a finite partition of Q into semi-algebraic manifolds M_i (called strata) with the following properties:

1. For distinct i and j, if $M_i \cap \operatorname{cl} M_j \neq \emptyset$, then $M_i \subset \operatorname{cl} M_j \setminus M_j$.
2. For any sequence of points (x_k) in a stratum M_j converging to a point x in a stratum M_i, if the corresponding normal vectors $y_k \in N_{M_j}(x_k)$ converge to a vector y, then $y \in N_{M_i}(x)$.

Observe that property 1 of Definition 2.13 gives us topological information on how the strata fit together, while property 2 gives us control over how sharply the strata fit together. Property 1 is called the *frontier condition* and property 2 is called *Whitney condition (a)*. We should note that Whitney stratification, as defined above, is normally referred to as $\mathbf{C^1}$-Whitney stratification. Furthermore, Whitney condition (a) is usually stated somewhat differently. The equivalence is noted in [18]. One simple example of this type of a stratification to keep in mind throughout the discussion is the partition of a polytope into its open faces.

Definition 2.14. Given finite collections $\{B_i\}$ and $\{C_j\}$ of subsets of $\mathbf{R}^n$, we say that $\{B_i\}$ is *compatible* with $\{C_j\}$ if for all B_i and C_j, either $B_i \cap C_j = \emptyset$ or $B_i \subset C_j$.

As discussed above, the following theorem is true.

Theorem 2.15 ([31, Theorem 4.8])**.** *Let $Q, C_1, \ldots, C_l$ be semi-algebraic sets in $\mathbf{R}^n$. Then Q admits a Whitney stratification that is compatible with $C_1, \ldots, C_l$.*

The notion of a stratification being compatible with some predefined sets might not look natural; in fact, it is crucial. This property gives us the ability to construct refinements of stratifications and will be used extensively. We will have occasion to use the following simple result.

Theorem 2.16 ([31, Theorem 4.2])**.** *Consider a semi-algebraic set $S \subset \mathbf{R}^n$ and a semi-algebraic function $f\colon S \to \mathbf{R}$. Let $\mathcal{A}$ be a finite collection of semi-algebraic subsets of S. Then there exists a finite partition of S into semi-algebraic submanifolds $\{B_i\}$ such that $f|_{B_i}$ is smooth and $\{B_i\}$ is compatible with $\mathcal{A}$.*

In particular, it follows that semi-algebraic functions are "generically" (in a sense about to be made clear) smooth. In fact, much more can be said about how well-behaved semi-algebraic functions are. Given manifolds M and N, and a smooth map $F\colon M \to N$, we say that F has *constant rank* if its derivative has constant rank throughout the domain. Such maps are particularly simple, since locally they look like projections. See [22, Theorem 7.13] for more details.

Theorem 2.17 ([31, Theorem 4.8])**.** *Consider a semi-algebraic set $S \subset \mathbf{R}^n$ and a semi-algebraic map $f\colon S \to \mathbf{R}^m$. Let $\mathcal{A}$ be a finite collection of semi-algebraic subsets of S and $\mathcal{B}$ a finite collection of semi-algebraic subsets of $\mathbf{R}^m$. Then there exists a Whitney stratification $\mathcal{A}'$ of S that is compatible with $\mathcal{A}$ and a Whitney stratification $\mathcal{B}'$ of $\mathbf{R}^m$ compatible with $\mathcal{B}$ such that for every stratum $Q \in \mathcal{A}'$, we have that the restriction $f|_Q$ is smooth with constant rank, and $f(Q) \in \mathcal{B}'$.*

Definition 2.18. Let $A \subset \mathbf{R}^n$ be a nonempty semi-algebraic set. Then we define the *dimension* of A, $\dim A$, to be the maximal dimension of a stratum in any Whitney stratification of A. We adopt the convention that $\dim \emptyset = -\infty$.

It can be easily shown that the dimension does not depend on the particular stratification. See [30, Chapter 4] for more details.

Theorem 2.19. *Let A and B be nonempty semi-algebraic sets in $\mathbf{R}^n$. Then the following hold.*

1. *If $A \subset B$, then* $\dim A \leq \dim B$.
2. $\dim A = \dim \operatorname{cl} A$.
3. $\dim(\operatorname{cl} A \setminus A) < \dim A$.
4. *If $f\colon A \to \mathbf{R}^n$ is a semi-algebraic function, then $f(A)$ is a semi-algebraic set and* $\dim f(A) \leq \dim A$. *If f is one-to-one, then we have* $\dim f(A) = \dim A$. *In particular, semi-algebraic homeomorphisms preserve dimension.*
5. $\dim A \cup B = \max\{\dim A, \dim B\}$.
6. *$A \times B$ is a semi-algebraic set and* $\dim A \times B = \dim A + \dim B$.

A set $U \subset \mathbf{R}^n$ is said to be "generic", if it is large in some precise mathematical sense, depending on context. Two popular choices are that of U being a *full-measure* set, meaning its complement has Lebesgue measure zero, and that of U being *topologically generic*, meaning it contains a countable intersection of dense open sets. In general, these notions are very different. However for semi-algebraic sets, the situation simplifies drastically. Indeed, if $U \subset \mathbf{R}^n$ is a semi-algebraic set, then the following are equivalent.

- U is full-measure.
- U is topologically generic.
- The dimension of U^c is strictly smaller than n.

We will say that a certain property holds for a generic vector $v \in \mathbf{R}^n$ if the set of vectors for which this property holds is generic in the sense just described. Generic properties of semi-algebraic optimization problems will be discussed in Section 4.

Definition 2.20. A continuous semi-algebraic mapping $p\colon A \to \mathbf{R}^n$ is *semi-algebraically trivial* over a semi-algebraic set $C \subset \mathbf{R}^n$ if there is a semi-algebraic set F and a semi-algebraic homeomorphism $h\colon p^{-1}(C) \to C \times F$ such that $p|_{p^{-1}(C)} = \mathrm{proj} \circ h$, or in other words the following diagram commutes:

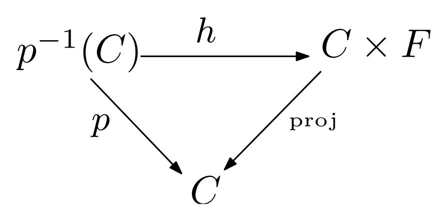

Figure 1: Semi-algebraic Triviality.

We call h a *semi-algebraic trivialization* of p over C.

Henceforth, we use the symbol $\cong$ to indicate that two semi-algebraic sets are semi-algebraically homeomorphic.

Remark 2.21. If p is trivial over some semi-algebraic set C, then we can decompose $p|_{p^{-1}(C)}$ into a homeomorphism followed by a simple projection. Also, since the homeomorphism h in the definition is surjective and $p =$

proj $\circ\, h$, it follows that $h(p^{-1}(c)) = \{c\} \times F$ for any $c \in C$. Thus for any point $c \in C$, we have $p^{-1}(c) \cong F$ and $p^{-1}(C) \cong C \times p^{-1}(c)$.

The following is a simple example of semi-algebraic triviality.

Example 2.22. We follow the notation of Definition 2.20. Consider the semi-algebraic function $p\colon \mathbf{R} \to \mathbf{R}$ defined by $p(x) = x^2$. Now consider the semi-algebraic mapping

$$h\colon \mathbf{R} \setminus \{0\} \to \mathbf{R}_{++} \times \{\pm 1\}, \qquad x \mapsto (x^2, \operatorname{sgn} x).$$

It is easy to check that h is a semi-algebraic homeomorphism and furthermore we have $p = \operatorname{proj} \circ h$. Thus h is a semi-algebraic trivialization of p over $\mathbf{R}_{++}$.

Definition 2.23. In the notation of Definition 2.20, a trivialization h is *compatible* with a semi-algebraic set $B \subset A$ if there is a semi-algebraic set $H \subset F$ such that $h(B \cap p^{-1}(C)) = C \times H$.

If h is a trivialization over C, then certainly for any set $B \subset A$, we know h restricts to a homeomorphism from $B \cap p^{-1}(C)$ to $h(B \cap p^{-1}(C))$. The content of the definition above is that if p is compatible with B, than h restricts to a homeomorphism between $B \cap p^{-1}(C)$ and $C \times H$ for some semi-algebraic set $H \subset F$. Here is a simple example.

Example 2.24. Let the semi-algebraic functions p and h be as defined in Example 2.22. Now notice that $h(\mathbf{R}_{++} \cap p^{-1}(\mathbf{R}_{++})) = \mathbf{R}_{++} \times \{+1\}$. Thus h is compatible with $\mathbf{R}_{++}$.

The following result will be used extensively in the rest of this work. See [30, Chapter 9, Theorem 1.2] for more details.

Theorem 2.25 (Hardt triviality)**.** *Let $A \subset \mathbf{R}^n$ be a semi-algebraic set and $p\colon A \to \mathbf{R}^m$, a continuous semi-algebraic mapping. There is a finite partition of the image $p(A)$ into semi-algebraic sets $C_1, \ldots, C_k$ such that p is semi-algebraically trivial over each C_i. Furthermore, if $Q_1, \ldots, Q_l$ are semi-algebraic subsets of A, we can require each trivialization $h_i\colon p^{-1}(C_i) \to C_i \times F_i$ to be compatible with all Q_j.*

Example 2.26. Consider the following elaboration on Example 2.22. Let the semi-algebraic functions p and h be defined as in Example 2.22. We saw that h is a semi-algebraic trivialization of p over $\mathbf{R}_{++}$. Let $f\colon \{0\} \to \{0\} \times \{0\}$ be the zero map. Observe f is a semi-algebraic trivialization of p over $\{0\}$. Thus $\{\mathbf{R}_{++}, \{0\}\}$ is a partition of $p(\mathbf{R})$ guaranteed to exist by Theorem 2.25.

Given a continuous semi-algebraic function p, Theorem 2.25 states that we can partition the image of p into semi-algebraic sets $C_1, \ldots, C_k$, so that for each index $i = 1, \ldots, k$, the restricted mapping $p|_{C_i}$ has a very simple form. By applying Theorem 2.25 to various naturally occurring mappings, many interesting results can be obtained. See [30, Chapter 9] for more details. In particular, by applying this theorem to the projection map we can break up semi-algebraic sets into simple building blocks that have product structure and analyze each one separately. This type of reasoning leads to the following corollary.

Corollary 2.27. *Let $F \colon \mathbf{R}^n \rightrightarrows \mathbf{R}^m$ be a semi-algebraic set-valued mapping. Then there exists a partition of the domain of F into semi-algebraic sets $X_1, X_2, \ldots, X_k$ with the following properties:*

1. *For each index $i = 1, 2, \ldots k$, there exists a semi-algebraic set $Y_i \subset \mathbf{R}^m$ and a semi-algebraic homeomorphism $\theta_i \colon \operatorname{gph} F|_{X_i} \to X_i \times Y_i$ satisfying*
$$\theta_i(\{x\} \times F(x)) = \{x\} \times Y_i \quad \textit{for all } x \in X_i.$$
Consequently, for all $x \in X_i$, we have $F(x) \cong Y_i$ and
$$\operatorname{gph} F|_{X_i} \cong X_i \times F(x).$$

2. *If in addition, $\widetilde{F} \colon \mathbf{R}^n \rightrightarrows \mathbf{R}^m$ is another semi-algebraic set-valued mapping with $\widetilde{F}(x) \subset F(x)$, then we may also require that for each index $i = 1, 2, \ldots, k$, there exists a semi-algebraic set $\widetilde{Y}_i \subset Y_i$, such that $\theta_i(\operatorname{gph} \widetilde{F}|_{X_i}) = X_i \times \widetilde{Y}_i$. Consequently, for all $x \in X_i$, we have $\widetilde{F}(x) \cong \widetilde{Y}_i$ and*
$$\operatorname{gph} \widetilde{F}|_{X_i} \cong X_i \times \widetilde{F}(x).$$

Proof Assume that we are given semi-algebraic set-valued maps F and $\widetilde{F}$ such that $\widetilde{F}(x) \subset F(x)$ for all $x \in \mathbf{R}^n$. If $\widetilde{F}$ was not given, proceed with the proof with $\widetilde{F}(x) = \emptyset$ for all $x \in \mathbf{R}^n$. Consider $\operatorname{gph} F \subset \mathbf{R}^n \times \mathbf{R}^m$. Let $p \colon \operatorname{gph} F \to \mathbf{R}^n$ be the projection onto the first n coordinates. By applying Theorem 2.25 to p, we get a partition of the domain of F into semi-algebraic sets $X_1, X_2, \ldots, X_k$ such that p is semi-algebraically trivial over each X_i and each trivialization is compatible with $\operatorname{gph} \widetilde{F}$. Thus there exist semi-algebraic sets $Y_1, Y_2, \ldots, Y_k \subset \mathbf{R}^m$ and $\widetilde{Y}_1, \widetilde{Y}_2, \ldots, \widetilde{Y}_k \subset \mathbf{R}^m$ with $\widetilde{Y}_i \subset Y_i$, such that for each i, there is a semi-algebraic homeomorphism $\theta_i \colon p^{-1}(X_i) \to X_i \times Y_i$, where we have

$$
\begin{aligned}
\theta_i(\operatorname{gph} \widetilde{F} \cap p^{-1}(X_i)) &= X_i \times \widetilde{Y}_i, \\
\operatorname{proj}_{X_i} \circ \, \theta_i &= p|_{p^{-1}(X_i)}.
\end{aligned} \tag{2}
$$

Observe that $p^{-1}(X_i) = \operatorname{gph} F|_{X_i}$ and since $\operatorname{gph} \widetilde{F}$ is contained in $\operatorname{gph} F$, it follows that $\operatorname{gph} \widetilde{F} \cap p^{-1}(X_i) = \operatorname{gph} \widetilde{F}|_{X_i}$. Thus to summarize, we have

$$
\begin{aligned}
\operatorname{gph} F|_{X_i} &\cong X_i \times Y_i, \\
\operatorname{gph} \widetilde{F}|_{X_i} &\cong X_i \times \widetilde{Y}_i.
\end{aligned} \tag{3}
$$

Finally from (2) and (3), it follows that for all points $x \in X_i$, we have

$$\theta_i(\{x\} \times F(x)) = \{x\} \times Y_i,$$

completing the proof. □

The following proposition appears in [2, 3]; as observed there, this result is an easy and important consequence of Theorem 2.25, and even though we will not have occasion to use it in this work, we include it and its proof below as an elegant illustration.

Proposition 2.28. *Let $F \colon \mathbf{R}^n \rightrightarrows \mathbf{R}^m$ be a semi-algebraic set-valued mapping. Then there exists a finite partition of the domain of F into semi-algebraic sets $X_1, \ldots, X_k$, such that for each index $i = 1, \ldots, k$, the restricted mapping $F|_{X_i}$ is inner semicontinuous. If in addition, the mapping F is compact-valued, then we can also require the restricted mapping $F|_{X_i}$ to be outer semicontinuous for each index $i = 1, \ldots, k$. (In fact, the partition guaranteed to exist by Corollary 2.27 is one such partition.)*

Proof Applying Corollary 2.27 to the mapping F, we get a finite partition of the domain of F into semi-algebraic sets $X_1, \ldots, X_k$, so that in particular property 1 of the corollary holds. To see the inner semicontinuity of the restricted map $F|_{X_i}$, consider any point $(\bar{x}, \bar{y}) \in \operatorname{gph} F|_{X_i}$, and any sequence of points $x_r \to \bar{x}$ in the set X_i. We want to construct a sequence of points $y_r \in F(x_r)$ converging to $\bar{y}$. Notice that $\theta_i(\bar{x}, \bar{y}) = (\bar{x}, \hat{y})$ for some point $\hat{y} \in Y_i$. Since $(x_r, \hat{y}) \to (\bar{x}, \hat{y})$, we deduce $\theta_i^{-1}(x_r, \hat{y}) \to \theta_i^{-1}(\bar{x}, \hat{y}) = (\bar{x}, \bar{y})$. But for each index $r = 1, 2, \ldots$, we know $\theta_i^{-1}(x_r, \hat{y}) = (x_r, y_r)$ for some point $y_r \in F(x_r)$, so the result follows.

Assume now that F is compact valued. Consider any point $\bar{x} \in X_i$ and any sequence of points $(x_r, y_r) \to (\bar{x}, \bar{y})$, where $\bar{y}$ is some point in $\mathbf{R}^m$ and $y_r \in F(x_r)$ for each r. We want to argue that $\bar{y}$ is in $F(\bar{x})$. Consider the sequence $(\bar{x}, \mathrm{proj}_{Y_i}(\theta_i(x_r, y_r)))$. Observe that this sequence is contained in $\{\bar{x}\} \times Y_i$, which is a compact set since it is homeomorphic to $F(\bar{x})$. Thus without loss of generality, we can assume that $(\bar{x}, \mathrm{proj}_{Y_i}(\theta_i(x_r, y_r)))$ converges to $(\bar{x}, \hat{y})$ for some point $\hat{y} \in Y_i$. So we have

$$(x_r, y_r) = \theta_i^{-1}(x_r, \mathrm{proj}_{Y_i}(\theta_i(x_r, y_r))) \to \theta_i^{-1}(\bar{x}, \hat{y}) \in \{\bar{x}\} \times F(\bar{x}).$$

By the uniqueness of the limit, we must have $\bar{y} \in F(\bar{x})$. □

As a consequence of Proposition 2.28, it follows that any semi-algebraic set-valued mapping $F\colon \mathbf{R}^n \to \mathbf{R}^m$ is generically inner semicontinuous. If in addition, F is compact valued, then F is generically continuous. In fact, we can do better. If we require the mapping F just to be closed valued, than we can still partition its domain into semi-algebraic sets $X_1, \ldots, X_k$, such that for each index $i = 1, \ldots, k$, the restricted mapping $F|_{X_i}$ is continuous. To see this, we need the following theorem that appears in [27, Theorem 5.55], and is attributed to [11, 21, 28].

Theorem 2.29 (Kuratowski)**.** *Consider a set $X \subset \mathbf{R}^n$ and a closed-valued set-valued mapping $F\colon X \rightrightarrows \mathbf{R}^m$. Assume that F is either outer semicontinuous or inner semicontinuous relative to X. Then the set of points where S fails to be continuous relative to X is meager in X.*

It is easy to see that if a semi-algebraic set S is meager in another semi-algebraic set X, then the dimension of S is strictly less than the dimension of X (see [4] for more details).

Proposition 2.30. *Let $F\colon \mathbf{R}^n \rightrightarrows \mathbf{R}^m$ be a semi-algebraic closed-valued set-valued mapping. Then there exists a finite partition of the domain of F into semi-algebraic sets $X_1, \ldots, X_k$, such that for each index $i = 1, \ldots, k$, the restricted mapping $F|_{X_i}$ is continuous.*

Proof Applying Proposition 2.28 to the mapping F, we get a partition of the domain of F into semi-algebraic sets $X_1, \ldots, X_k$, so that the restricted map $F|_{X_i}$ is inner semicontinuous. Fix some set X_i. Let $S^0 := X_i$ and let $S^1 \subset X_i$ be the set of points at which $F|_{S^0}$ fails to be continuous. By Theorem 2.29, it follows that $\dim S^1 < \dim S^0$. Now by applying this argument inductively, we can create a sequence of semi-algebraic sets $S^0 \supset \ldots \supset S^k$, for some integer k, such that the collection $\{S_j \setminus S_{j+1}\}_{j=0}^{k-1}$ is a partition of X_i and F is continuous when restricted to each $S_j \setminus S_{j+1}$. By applying this argument to all the sets X_i, for $i = 1, \ldots, k$, we get the result. $\square$

Remark 2.31. In fact, it is shown in Daniilidis-Pang [16] that closed-valued semi-algebraic maps are generically strictly continuous (see [27] for the definition). Their proof of this rather stronger result requires more sophisticated tools.

Finally, we have the following result:

Theorem 2.32 ([31, Theorem 4.4])**.** *Let A be a semi-algebraic subset of $\mathbf{R}^n \times \mathbf{R}^m$. There is an integer β such that for every point $x \in \mathbf{R}^n$, the number of connected components of the set $A_x = \{y \in \mathbf{R}^m : (x, y) \in A\}$ is no greater than β.*

The following is a simple special case of Theorem 2.32. We record it here for convenience.

Corollary 2.33. *Let $F \colon \mathbf{R}^n \rightrightarrows \mathbf{R}^m$ be a semi-algebraic mapping. There is $\beta \in \mathbb{N}$ such that for every $x \in \mathbf{R}^n$, the number of connected components of $F(x)$ is no greater than β.*

Proof Apply Theorem 2.32 to $\operatorname{gph} F \subset \mathbf{R}^n \times \mathbf{R}^m$ and notice that $(\operatorname{gph} F)_x = F(x)$. $\square$

3 Main Results

Consider a convex semi-algebraic set $K \subset \mathbf{R}^n$. We denote by $\operatorname{aff} K$, the affine span of K. It is not hard to see that the dimension of K as a convex set, which is the dimension of $\operatorname{aff} K$, is equal to the dimension of K as a semi-algebraic set. We will have occasion to use the following simple result.

Proposition 3.1. *Consider a function $f \colon \mathbf{R}^n \to \overline{\mathbf{R}}$ and a point $\bar{x} \in \mathbf{R}^n$. If the regular subdifferential $\hat{\partial} f(\bar{x})$ is nonempty, then $\hat{\partial} f(\bar{x})$ has dimension $\dim \hat{N}_{\operatorname{epi} f}(\bar{x}, f(\bar{x})) - 1$. The same relationship holds in the Clarke case.*

Proof Consider a point $\bar{x} \in \mathbf{R}^n$ with $\hat{\partial} f(\bar{x})$ nonempty. Let K denote the regular normal cone, $\hat{N}_{\operatorname{epi} f}(\bar{x}, f(\bar{x}))$. Observe

$$\operatorname{ri} K \not\subset \{y \in \mathbf{R}^{n+1} : y_{n+1} \geq 0\},$$

since otherwise taking closures gives $y_{n+1} \geq 0$ for all $y \in K$ and hence we have $\hat{\partial} f(\bar{x}) = \emptyset$, which is a contradiction. Thus there exists a point $\hat{y} \in \operatorname{ri} K$ with $\hat{y}_{n+1} < 0$. Since K is a cone, we can rescale to get $\hat{y} \in \operatorname{ri} K$ with $\hat{y}_{n+1} = -1$. So we have

$$\operatorname{ri} K \cap \operatorname{ri} \{y \in \mathbf{R}^{n+1} : y_{k+1} = -1\} \neq \emptyset.$$

It follows, using [27, Proposition 2.42], that

$$\begin{aligned} \operatorname{aff} \{y \in K : y_{k+1} = -1\} &= \operatorname{aff} K \cap \operatorname{aff} \{y \in \mathbf{R}^{n+1} : y_{k+1} = -1\} \\ &= (K - K) \cap \{y \in \mathbf{R}^{n+1} : y_{k+1} = -1\}. \end{aligned}$$

Observe that $K - K$ is a subspace and $\{y \in \mathbf{R}^{n+1} : y_{k+1} = -1\}$ is a hyperplane not containing the origin. Elementary linear algebra shows that the dimension of the right hand side is $\dim(K - K) - 1 = \dim K - 1$ and thus the dimension of $\hat{\partial} f(\bar{x})$ is $\dim \hat{N}_{\operatorname{epi} f}(\bar{x}, f(\bar{x})) - 1$. The same argument works for $\partial_c f(\bar{x})$. □

Theorem 3.2. *Let $f \colon \mathbf{R}^n \to \overline{\mathbf{R}}$ be a semi-algebraic function. Then the graph of the limiting subdifferential,* $\operatorname{gph} \partial f$*, has dimension no greater than n.*

Proof Let $D = \operatorname{dom} \hat{\partial} f$. Applying Corollary 2.27 to the semi-algebraic set-valued mapping $\hat{\partial} f$, we get a finite partition of D into semi-algebraic sets $\{C_i\}$, such that

$$\operatorname{gph} \hat{\partial} f|_{C_i} \cong C_i \times \hat{\partial} f(c) \tag{4}$$

for any $c \in C_i$. Now by Theorem 2.16, we can partition the domain of f into finitely many semi-algebraic submanifolds $\{B_j\}$ such that $f|_{B_j}$ is smooth and $\{B_j\}$ is compatible with $\{C_k\}$. Now consider some C_i. Without loss of generality, assume that $C_i = B_1 \cup \ldots \cup B_l$ for some l. Let B be a stratum of C_i of maximal dimension and let $c \in B$ be arbitrary. Observe that since f is smooth on B, it follows that $\operatorname{gph} f|_B$ is a manifold of the same dimension as B. Using this observation and Proposition 3.1, we have

$$\begin{aligned} \dim \hat{\partial} f(c) &\leq \dim \hat{N}_{\operatorname{epi} f}(c, f(c)) - 1 \leq \dim \hat{N}_{\operatorname{gph} f}(c, f(c)) - 1 \\ &\leq \dim \hat{N}_{\operatorname{gph} f|_B}(c, f(c)) - 1 = n + 1 - \dim B - 1 \\ &= n - \dim C_i, \end{aligned} \tag{5}$$

Combining (4) and (5), it follows that

$$\dim \operatorname{gph} \hat{\partial} f|_{C_i} = \dim C_i + \dim \hat{\partial} f(c) \leq \dim C_i + (n - \dim C_i) = n.$$

Finally, we have

$$\dim \operatorname{gph} \hat{\partial} f = \dim \Big(\bigcup_i \operatorname{gph} \hat{\partial} f|_{C_i} \Big) = \max_i (\dim \operatorname{gph} \hat{\partial} f|_{C_i}) \leq n.$$

Since $\operatorname{gph} \partial f \subset \operatorname{cl} \operatorname{gph} \hat{\partial} f$, we deduce that $\dim \operatorname{gph} \partial f \leq n$ as well. □

With some effort, we can improve the result of Theorem 3.2. Namely we will show that the same result holds with ∂f replaced by $\partial_c f$. One indication of the difficulty is that the monotonicity property of the regular normal cone (Theorem 2.5, Property 3), crucial in the proof above, can fail for the Clarke normal cone. Furthermore to pass from $\hat{\partial} f$ to ∂f, we used the fact that taking closure of a semi-algebraic set does not increase dimension. This type of reasoning will not suffice for dealing with $\partial_c f$, since for some point x, the convex hull operation can make $\partial_c f(x)$ much larger than $\hat{\partial} f(x)$. In particular, the dimension of $\partial_c f(x)$ can be strictly larger than that of $\hat{\partial} f(x)$. Fortunately, this potential increase in dimension turns out to be irrelevant. The following simple geometric result is key. This result is essentially equivalent to a projection lemma due to Bolte-Daniilidis-Lewis-Shiota [5, Proposition 4]. Since its proof is short and illuminating, we include it here.

Proposition 3.3. *Suppose the set $Q \subset \mathbf{R}^n$ admits a Whitney stratification $\{M_i\}$. Then at any point in a stratum M_i, the Clarke normal cone to Q is contained in the normal space to M_i.*

Proof Consider an arbitrary point x in some stratum M_i. We claim that the limiting normal cone $N_Q(x)$ is contained in $\hat{N}_{M_i}(x)$. To see this, consider a vector $v \in N_Q(x)$. By definition of the limiting normal cone, there exist sequences (x_r) and (v_r) such that $x_r \xrightarrow{Q} x$ and $v_r \to v$ with $v_r \in \hat{N}_Q(x_r)$. Since there are finitely many strata, we can assume that there is some stratum M_j such that the entire sequence (x_r) is contained in M_j. Since $M_j \subset Q$, it follows that $\hat{N}_Q(x_r) \subset \hat{N}_{M_j}(x_r)$. Thus $v_r \in \hat{N}_{M_j}(x_r)$. Therefore by Whitney condition (a), we have $v \in \hat{N}_{M_i}(x)$. Since v was arbitrarily chosen from $N_Q(x)$, it follows that $N_Q(x) \subset \hat{N}_{M_i}(x)$ and thus $N^c_Q(x) = \operatorname{cl} \operatorname{conv} N_Q(x) \subset \hat{N}_{M_i}(x)$. Finally observe that since M_i is a manifold, the regular normal cone $\hat{N}_{M_i}(x)$ is the normal space to M_i at x. □

We are now ready for the main result. Our proof uses the triviality tools we have developed, an appealingly systematic framework. An alternative approach uses a stratification argument directly, along with the projection lemma of [5, Proposition 4].

Theorem 3.4. *Let $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$ be a semi-algebraic function. Then the graph of the Clarke subdifferential,* $\operatorname{gph}\partial_c f$*, has dimension no greater than* n.

Proof Let $D = \operatorname{dom}\partial_c f$. Applying Corollary 2.27 to the semi-algebraic set-valued mapping $\partial_c f$, we get a finite partition of D into semi-algebraic sets $\{C_i\}$, such that

$$\operatorname{gph}\partial_c f|_{C_i} \cong C_i \times \partial_c f(x)$$

for any $x \in C_i$.

Now we will use Theorem 2.17 applied to the projection map $p\colon \operatorname{epi} f \to \mathbf{R}^n$ obtained by restricting the canonical projection $\mathbf{R}^n \times \mathbf{R} \to \mathbf{R}^n$. Namely, in the notation of the theorem, we let $\mathcal{A} = \{\operatorname{gph} f\}$ and $\mathcal{B} = \{C_i\}$. Thus we get a Whitney stratification $\mathcal{A}'$ of $\operatorname{epi} f$ compatible with $\mathcal{A}$ and a Whitney stratification $\mathcal{B}'$ of $\mathbf{R}^n$ compatible with $\mathcal{B}$ such that for every set $Q \in \mathcal{A}'$, the restriction $p|_Q$ is smooth with constant rank, and $p(Q) \in \mathcal{B}'$. Let the strata of $\mathcal{A}'$ be $\{M_i\}$ and let I be the index set for the strata that are contained in $\operatorname{gph} f$. Thus we have

$$\operatorname{gph} f = \bigcup_{i \in I} M_i.$$

Observe that for $i \in I$, the restricted map $p|_{M_i}$ is smooth with constant rank, and is a bijection onto its image $p(M_i)$. By a standard fact from differential geometry (see [22, Theorem 7.15]), it follows that $p|_{M_i}$ is a diffeomorphism. Thus $p(M_i) \in \mathcal{B}'$ is a manifold of the same dimension as M_i, that is we know for each $i \in I$, $\dim p(M_i) = \dim M_i$.

Consider some set $C_i \in \mathcal{B}$. It has been partitioned into strata:

$$C_i = \bigcup_{j \in J} p(M_j)$$

for some index set $J \subset I$. Let $p(M_j)$ be the stratum of C_i of maximal dimension and let $x \in p(M_j)$ be arbitrary. We have

$$\operatorname{gph}\partial_c f|_{C_i} \cong C_i \times \partial_c f(x). \tag{6}$$

By Proposition 3.1, we have that

$$\dim \partial_c f(x) \leq \dim N^c_{\operatorname{epi} f}(x, f(x)) - 1.$$

Now observe that $(x, f(x))$ is contained in M_j and therefore by Proposition 3.3, we have

$$N^c_{\operatorname{epi} f}(x, f(x)) \subset N_{M_j}(x, f(x)).$$

Therefore

$$(7)\qquad \begin{aligned} \dim \partial_c f(x) &\leq \dim N_{M_j}(x, f(x)) - 1 = n + 1 - \dim M_j - 1 \\ &= n - \dim p(M_j) = n - \dim C_i, \end{aligned}$$

where the last equality follows from maximality of $p(M_j)$. Combining (6) and (7), we deduce

$$\dim \operatorname{gph} \partial_c f|_{C_i} = \dim C_i + \dim \partial_c f(x) \leq \dim C_i + (n - \dim C_i) = n.$$

Finally, observe that

$$\dim \operatorname{gph} \partial_c f = \dim \Big(\bigcup_i \operatorname{gph} \partial_c f|_{C_i} \Big) = \max_i (\dim \operatorname{gph} \partial_c f|_{C_i}) \leq n,$$

completing the proof. □

Shortly we will show that for a proper semi-algebraic function $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$, both $\dim \operatorname{gph} \partial_c f$ and $\dim \operatorname{gph} \partial f$ are actually equal to n. If the domain of f is full-dimensional, then this fact is easy to show. The argument is as follows. By Theorem 2.16, the domain of f can be partitioned into semi-algebraic submanifolds $\{X_i\}$ such that $f|_{X_i}$ is smooth. Let X_i be the manifold of maximal dimension. Observe that for $x \in X_i$, we have that $\partial f(x) = \{\nabla f(x)\}$ and it easily follows that $\dim \operatorname{gph} \partial f|_{X_i} = n$. Thus we have

$$n \leq \dim \operatorname{gph} \partial f \leq \dim \operatorname{gph} \partial_c f \leq n,$$

where the last inequality follows from Theorem 3.4, and hence there is equality throughout. The argument just presented no longer works when the domain of f is not full-dimensional. A slightly more involved argument is required. We record the following simple observation for reference.

Proposition 3.5. *Consider a smooth manifold $M \subset \mathbf{R}^n$ and a smooth real-valued function $f\colon M \to \mathbf{R}$. Define a function $h : \mathbf{R}^n \to \overline{\mathbf{R}}$ agreeing with f on M and equaling plus infinity elsewhere. Then h is subdifferentially regular throughout M. Furthermore, at any point $\bar{x} \in M$, we have*

$$\partial h(\bar{x}) = N_M(\bar{x}) + \nabla g(\bar{x}),$$

where $g\colon \mathbf{R}^n \to \mathbf{R}$ is any smooth function agreeing with f on M on a neighborhood of $\bar{x}$. Consequently, $\partial h(\bar{x})$ is nonempty with dimension $n - \dim M$.

Proof Observe that near the point $\bar{x}$, we have $h = \delta_M + g$. Combining Theorem 2.8 and Theorem 2.12, we have that the function δ_M is regular at $\bar{x}$. By Theorem 2.11, it follows that h is regular at $\bar{x}$ and

$$\partial h(\bar{x}) = \partial \delta_M(\bar{x}) + \nabla g(\bar{x}) = N_M(\bar{x}) + \nabla g(\bar{x}),$$

as we needed to show. □

Theorem 3.6. *Let $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$ be a proper semi-algebraic function. Then the graphs of the regular, limiting, and Clarke subdifferentials have dimension exactly n.*

Proof We know that

$$\dim \operatorname{gph} \hat{\partial} f \leq \dim \operatorname{gph} \partial f \leq \dim \operatorname{gph} \partial_c f \leq n,$$

where the last inequality follows from Theorem 3.4. Thus if we show that $\dim \operatorname{gph} \hat{\partial} f \geq n$, we will be done. Let $D = \operatorname{dom} \hat{\partial} f$. First, we apply Corollary 2.27 to the semi-algebraic set-valued mapping $\hat{\partial} f$. Thus we get a finite partition of D into semi-algebraic sets $\{C_i\}$, such that

$$\operatorname{gph} \hat{\partial} f|_{C_i} \cong C_i \times \hat{\partial} f(x), \tag{8}$$

for any point $x \in C_i$. Next we use Theorem 2.17 applied to the mapping f. Specifically, we get a Whitney stratification $\mathcal{A}'$ of the domain of f compatible with $\{C_i\}$, such that for every stratum $Q \in \mathcal{A}'$, the restriction $f|_Q$ is smooth. Let the strata of $\mathcal{A}'$ be $\{M_i\}$ and let M_j be a stratum of $\operatorname{dom} f$ of maximal dimension.

Lemma 3.7. *There exists a neighborhood $B \subset \mathbf{R}^n$ around $\bar{x}$ so that*

$$B \cap \operatorname{dom} f = B \cap M_j.$$

Proof Assume otherwise. Then there is a sequence $x_r \in \operatorname{dom} f$ converging to $\bar{x}$ with $x_r \notin M_j$. Since there are finitely many strata, we can assume that the whole sequence is contained in some stratum M. It follows that $\bar{x}$ is a limit point of M. By the first condition of the Whitney stratification, it must be that $\dim M_j < \dim M$, which is a contradiction since the stratum M_j was chosen to have maximal dimension. □

Now consider the function $h\colon \mathbf{R}^n \to \overline{\mathbf{R}}$, which agrees with f on M_j and is plus infinity elsewhere. By Lemma 3.7, the functions h and f coincide on a neighborhood of $\bar{x}$. Applying Proposition 3.5, we deduce that f is subdifferentially regular at $\bar{x}$ and $\partial f(\bar{x})$ is nonempty with dimension $n - \dim M_j$. So in particular, since $\partial f(\bar{x})$ is nonempty, it follows that $\bar{x} \in M_j \subset C_i$ for some i. So by (8), we have

$$\dim \operatorname{gph} \hat{\partial} f|_{C_i} = \dim C_i + \dim \hat{\partial} f(\bar{x}) = \dim C_i + n - \dim M_j = n,$$

where the last equality follows since M_j has maximal dimension in C_i. It follows that $\dim \operatorname{gph} \hat{\partial} f \geq n$. □

We end this section by relating our results to the notion of minimal cusco, introduced in [7]. Consider a set $A \subset \mathbf{R}^n$ and a set-valued mapping $F \colon A \rightrightarrows \mathbf{R}^m$. The mapping F is a *cusco* if it is outer-semicontinuous on A and $F(x)$ is a nonempty compact convex set for each point $x \in A$. A *minimal cusco* is a cusco, whose graph does not strictly contain the graph of any other cusco. It is tempting to think that in the semi-algebraic setting, the graph of a minimal cusco should have small dimension. However, it is not hard to see that this is not the case. For instance, we will now exhibit a semi-algebraic minimal cusco $F \colon \mathbf{R}^3 \rightrightarrows \mathbf{R}^3$, whose graph is 4-dimensional. Thus semi-algebraic minimal cuscos with low dimensional graphs are somewhat special.

To simplify notation, we let $[y < 0, z < 0]$ be an alias for the set $\{(x,y,z) \in \mathbf{R}^3 : y < 0, z < 0\}$ and we reserve analogous notation for relaters '>' and '='. Consider the semi-algebraic set-valued mapping $F \colon \mathbf{R}^3 \rightrightarrows \mathbf{R}^3$, defined as follows

$$F|_{[y>0,z>0]} = \{(0,0,0)\}, \ F|_{[y<0,z>0]} = \{(0,0,1)\},$$

$$F|_{[y<0,z<0]} = \{(0,1,0)\}, \ F|_{[y>0,z<0]} = \{(1,0,0)\},$$

$$F|_{[y>0,z=0]} = \operatorname{conv}\{(0,0,0),(0,0,1)\}, \ F|_{[y=0,z>0]} = \operatorname{conv}\{(0,0,1),(0,1,0)\},$$

$$F|_{[y<0,z=0]} = \operatorname{conv}\{(0,1,0),(1,0,0), \ F|_{[y=0,z<0]} = \operatorname{conv}\{(0,0,0),(1,0,0)\},$$

$$F|_{[y=0,z=0]} = \operatorname{conv}\{(0,0,0),(0,0,1),(0,1,0),(1,0,0)\}$$

It is easy to verify that F is indeed a minimal cusco with a 4-dimensional graph. In particular, Theorem 3.6 implies that F is not the Clarke subdifferential mapping $\partial_c f$ for any semi-algebraic function $f \colon \mathbf{R}^3 \rightrightarrows \overline{\mathbf{R}}$.

Let $U \subset \mathbf{R}^n$ be an open set and consider a semi-algebraic locally Lipschitz function $f \colon U \to \mathbf{R}$. It follows by a direct application of [7, Corollary 2.2] and generic smoothness of f that the the set-valued mapping $\partial_c f$ is a minimal cusco. Furthermore, combining this with Theorem 3.6, we deduce that the set-valued mapping $\partial_c f$ is a minimal cusco with n-dimensional graph.

4 Consequences

Definition 4.1. Consider a function $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$. We say that a point $x \in \mathbf{R}^n$ is *Clarke-critical* for the function f if $0 \in \partial_c f(x)$, and we call such a critical point x *nondegenerate* if the stronger property $0 \in \operatorname{ri} \partial_c f(x)$ holds.

Recall that for a proper convex function $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$ and a point $\bar{x} \in \operatorname{dom} f$, the subdifferentials $\hat{\partial} f(\bar{x})$, $\partial f(\bar{x})$, and $\partial_c f(\bar{x})$ all coincide and are equal to the convex subdifferential of f at $\bar{x}$. So in this case, the notions of Clarke-criticality and Clarke-nondegeneracy reduce to more familiar notions from Convex Analysis. Consider the following theorem (part 1 follows from [4, Proposition 1] and for part 2 see [17]).

Theorem 4.2. *Let $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$ be a proper convex function. Consider the collection of perturbed functions $h_v(x) = f(x) - \langle v, x\rangle$, indexed by vectors $v \in \mathbf{R}^n$. Then for a topologically generic and full measure set of vectors $v \in \mathbf{R}^n$, the function h_v has at most one minimizer, which furthermore is nondegenerate.*

Shortly, we will prove that a natural analogue of Theorem 4.2 holds for arbitrary semi-algebraic functions, with no assumption of convexity. We will then reference an example of a locally Lipschitz function that is not semi-algebraic, and for which the conclusion of our analogous result fails, thus showing that the assumption of semi-algebraicity is not superfluous. In what follows, for a set S, the number of elements in S will be denoted by $S^{\#}$. We begin with the following simple proposition.

Proposition 4.3. *Let $F\colon \mathbf{R}^n \rightrightarrows \mathbf{R}^m$ be a semi-algebraic set-valued mapping such that $\dim \operatorname{gph} F \le n$. Then there exists $\beta \in \mathbb{N}$ such that for a generic set of points $c \in \mathbf{R}^n$, we have $F(c)^{\#} \le \beta$.*

Proof Let $D = \operatorname{dom} F$. If the dimension of D is strictly less than n, then we are done since then the complement of D is a set satisfying the claimed property with $\beta = 0$. Thus assume that D has dimension n. Applying

Corollary 2.27 to the mapping F, we get a finite partition of D into semi-algebraic sets $\{C_i\}$, such that

$$\operatorname{gph} F|_{C_i} \cong C_i \times F(c)$$

for any $c \in C_i$. Let C_i be a stratum of maximal dimension. So $\dim C_i = n$ and we have

$$n \geq \dim \operatorname{gph} F|_{C_i} = n + \dim F(c)$$

for any $c \in C_i$. Thus $\dim F(c) = 0$ and since it is a semi-algebraic set, we have that it must be finite. Since this argument holds for any C_i of maximal dimension, we have that for a generic vector c, the set $F(c)$ is finite. Observe that if $F(c)$ is a finite non-empty set, then $F(c)^{\#}$ is equal to the number of connected components of $F(c)$. By Corollary 2.33, there exists $\beta \in \mathbb{N}$ such that for all $c \in \mathbf{R}^n$, the number of connected components of $F(c)$ is no greater than β. So in particular, for generic c, we have $F(c)^{\#} \leq \beta$ □

Corollary 4.4. *Let $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$ be a semi-algebraic function and consider the collection of perturbed functions $h_v(x) = f(x) - \langle v, x\rangle$, indexed by vectors $v \in \mathbf{R}^n$. Then there exists a positive integer β, such that for generic $v \in \mathbf{R}^n$, the number of Clarke-critical points of the perturbed function h_v is no greater than β.*

Proof Observe

$$0 \in \partial_c h_v(x) \Leftrightarrow v \in \partial_c f(x) \Leftrightarrow x \in (\partial_c f)^{-1}(v).$$

Thus the set $(\partial_c f)^{-1}(v)$ is equal to the set of Clarke-critical points of the function h_v. By Theorem 3.2, we have $\dim \operatorname{gph} \partial_c f \leq n$, hence $\dim \operatorname{gph} (\partial_c f)^{-1} \leq n$. Applying Theorem 4.3 to $(\partial_c f)^{-1}$, we deduce that there exists a positive integer β, such that for generic v, we have $((\partial_c f)^{-1}(v))^{\#} \leq \beta$. The result follows. □

Corollary 4.5. *Let $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$ be a semi-algebraic function and consider the collection of perturbed functions $h_v(x) = f(x) - v^T x$, indexed by points $v \in \mathbf{R}^n$. Then for generic $v \in \mathbf{R}^n$, every Clarke-critical point of the function h_v is nondegenerate.*

Corollary 4.5 follows immediately from the observation

$$0 \in \operatorname{ri} \partial_c h_v(x) \Leftrightarrow v \in \operatorname{ri} \partial_c f(x),$$

and the following result.

Corollary 4.6. *Let* $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$ *be a semi-algebraic function. Then for generic* $v \in \mathbf{R}^n$*, we have that*

$$x \in (\partial_c f)^{-1}(v) \Longrightarrow v \in \operatorname{ri} \partial_c f(x).$$

Proof Let $D = \operatorname{dom} \partial_c f$. Consider the semi-algebraic set-valued mapping

$$\widetilde{F}\colon \mathbf{R}^n \rightrightarrows \mathbf{R}^n$$

$$x \longmapsto \text{rb } \partial_c f(x).$$

Our immediate goal is to show that the dimension of $\operatorname{gph} \widetilde{F}$ is no greater than $n-1$. Observe that for each $x \in \mathbf{R}^n$, we have $\widetilde{F}(x) \subset \partial_c f(x)$. Applying Corollary 2.27 to the mapping $\partial_c f$, we get a finite partition of D into semi-algebraic sets $\{X_i\}$, such that

$$\operatorname{gph} \partial_c f|_{X_i} \cong X_i \times \partial_c f(x)$$

and

$$\operatorname{gph} \widetilde{F}|_{X_i} \cong X_i \times \widetilde{F}(x)$$

for any $x \in X_i$ (for each i). By Theorem 3.4, we have that

$$n \geq \dim \operatorname{gph} \partial_c f|_{X_i} = \dim X_i + \dim \partial_c f(x).$$

Since $\widetilde{F}(x) = \text{rb } \partial_c f(x)$, it follows that

$$\dim \widetilde{F}(x) \leq \dim \partial_c f(x) - 1.$$

Therefore

$$\dim \operatorname{gph} \widetilde{F}|_{X_i} = \dim X_i + \dim \widetilde{F}(x) \leq \dim X_i + \dim \partial_c f(x) - 1 \leq n - 1.$$

Thus

$$\dim \operatorname{gph} \widetilde{F} = \dim \Big(\bigcup_i \operatorname{gph} \widetilde{F}|_{X_i} \Big) \leq n - 1.$$

And so if we let

$$\pi\colon \operatorname{gph} \widetilde{F} \to \mathbf{R}^n$$

be the projection onto the last n coordinates, we deduce that $\dim \pi(\operatorname{gph} \widetilde{F}) \leq n - 1$. Finally, observe

$$\pi(\operatorname{gph} \tilde{F}) \; = \; \Big\{ v \in \mathbf{R}^n : v \in \operatorname{rb} \partial_c f(x) \text{ for some } x \in \mathbf{R}^n \Big\}.$$

□

Remark 4.7. Observe that if a convex function has finitely many minimizers then, in fact, it has a unique minimizer. Thus, for a proper convex semi-algebraic function, Corollaries 4.4 and 4.5 reduce to Theorem 4.2.

Remark 4.8. In Corollaries 4.4 and 4.5, if the function f is not semi-algebraic, then the results of these corollaries can fail. In fact, these results can fail even if the function f is locally Lipschitz continuous. For instance, there is a locally Lipschitz function $f\colon \mathbf{R} \to \mathbf{R}$ such that $\partial_c f(x) = [-x, x]$ for every $x \in \mathbf{R}$. See the article of Borwein-Moors-Wang [8]. For all $v \in \mathbf{R}$, the perturbed function h_v has infinitely many critical points, and for all $v \in \mathbf{R} \setminus \{0\}$, the function h_v has critical points that are degenerate.

References

[1] S. Basu, R. Pollack, and M. Roy. *Algorithms in Real Algebraic Geometry (Algorithms and Computation in Mathematics)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.

[2] L.E. Blume and W.R. Zame. *The Algebraic Geometry of Competitive Equilibrium*. Sep 1992. In: T.J. Rader and W. Neuefeind and R.G. Riezman (Eds.), Economic Theory and International Trade. Essays in Memoriam of J. Trout Rader. Springer-Verlag, Berlin.

[3] L.E. Blume and W.R. Zame. The algebraic geometry of perfect and sequential equilibrium. Game Theory and Information 9309001, EconWPA, Sep 1993.

[4] J. Bolte, A. Daniilidis, and A.S. Lewis. Generic optimality conditions for semi-algebraic convex programs. *submitted to Mathematics of Operations Research, 2009.*

[5] J. Bolte, A. Daniilidis, A.S. Lewis, and M. Shiota. Clarke subgradients of stratifiable functions. *SIAM J. on Optimization*, 18(2):556–572, 2007.

[6] J.F. Bonnans and A. Shapiro. *Perturbation Analysis of Optimization Problems*. Springer, New York, 2000.

[7] J.M. Borwein and W.B. Moors. Essentially smooth lipschitz functions. *Journal of Functional Analysis*, 149(2):305 – 351, 1997.

[8] J.M. Borwein, W.B. Moors, and X. Wang. Lipschitz functions with prescribed derivatives and subderivatives. *Nonlinear Analysis*, 29(1):53 – 63, 1997.

[9] J.M. Borwein and X. Wang. Lipschitz functions with maximal clarke subdifferentials are generic. *Proceedings of the American Mathematical Society*, 128(11):3221–3229, 2000.

[10] J.M. Borwein and Q.J. Zhu. *Techniques of Variational Analysis.* Springer Verlag, New York, 2005.

[11] G. Choquet. Convergences. *Ann. Univ. Grenoble*, 23:57–112, 1948.

[12] F.H. Clarke. *Optimization and Nonsmooth Analysis.* Wiley Interscience, New York, 1983.

[13] F.H. Clarke, Yu. Ledyaev, R.I. Stern, and P.R. Wolenski. *Nonsmooth Analysis and Control Theory.* Texts in Math. 178, Springer, New York, 1998.

[14] M. Coste. *An Introduction to o-minimal Geometry.* RAAG Notes, 81 pages, Institut de Recherche Mathématiques de Rennes, November 1999.

[15] M. Coste. *An Introduction to Semialgebraic Geometry.* RAAG Notes, 78 pages, Institut de Recherche Mathématiques de Rennes, October 2002.

[16] A. Daniilidis and C.H.J. Pang. Continuity of set-valued maps revisited in the light of tame geometry. 2009.

[17] D. Drusvyatskiy and A.S. Lewis. Generic properties of semi-algebraic optimization problems. *in preparation.*

[18] A.D. Ioffe. Critical values of set-valued maps with stratifiable graphs. extensions of sard and smale-sard theorems. *Proc. Amer. Math. Soc.*, 136:3111 – 3119, 2008.

[19] A.D. Ioffe. An invitation to tame optimization. *SIAM Journal on Optimization*, 19(4):1894–1917, 2009.

[20] A.D. Ioffe. The principle of lagrange and critical values of optimization problems. *To appear in Pacific J. Optimization*, 2010.

[21] K. Kuratowski. Les fonctions semi-continues dans lespace des ensembles fermés. *Fundamenta Mathematicae*, 18:148–159, 1932.

[22] J.M. Lee. *Introduction to Smooth Manifolds.* Springer, New York, 2003.

[23] B.S. Mordukhovich. *Variational Analysis and Generalized Differentiation I: Basic Theory.* Grundlehren der mathematischen Wissenschaften, Vol 330, Springer, Berlin, 2005.

[24] B.S. Mordukhovich. *Variational Analysis and Generalized Differentiation II: Applications.* Grundlehren der mathematischen Wissenschaften, Vol 331, Springer, Berlin, 2005.

[25] R.A. Poliquin and R.T. Rockafellar. Prox-regular functions in variational analysis. *Trans. Amer. Math. Soc.*, 348:1805–1838, 1996.

[26] S.M. Robinson. Equations on locally thin sets. *preprint*, 2010.

[27] R.T. Rockafellar and R. J-B. Wets. *Variational Analysis.* Springer, Berlin, 1998.

[28] Shu-Chung Shi. Semi-continuités génériques de multi-applications. *C.R. Acad. Sci. Paris (Série A)*, 293:27–29, 1981.

[29] M. Shiota. *Geometry of subanalytic and semialgebraic sets.* Birkhauser Boston Inc., Cambridge, MA, USA, 1997.

[30] L. van den Dries. *Tame topology and o-minimal structures*, volume 248 of *LMS Lecture Note Series.* Cambridge University Press, Cambridge, 1998.

[31] L. van den Dries and C. Miller. Geometric categories and o-minimal structures. *Duke Mathematics Journal*, 84:497–540, 1996.