Morse-Sard theorem for Clarke critical values

Luc Barbet, Marc Dambrine, Aris Daniilidis

Abstract The Morse-Sard theorem states that the set of critical values of a C^k smooth function defined on a Euclidean space $\mathbb{R}^d$ has Lebesgue measure zero, provided $k \geq d$. This result is hereby extended for (generalized) critical values of continuous selections over a compactly indexed countable family of C^k functions: it is shown that these functions are Lipschitz continuous and the set of their Clarke critical values is null.

Key words Morse-Sard theorem, Lipschitz function, Cantor–Bendixson derivative, Clarke critical value.

AMS Subject Classification *Primary* 49J52 ; *Secondary* 58E05, 26B40.

1 Introduction

The classical Morse-Sard theorem states that the set of critical values of a C^k smooth function defined on a Euclidean space $\mathbb{R}^d$ has Lebesgue measure zero, provided $k \geq d$ (see [11], [16]). The result is sharp as far as the order of smoothness of the function is concerned [17], [12] and is naturally extended to C^k-functions defined on a C^k-manifold of dimension d (see [8, Theorem 1.3] *e.g.*). Quantitative Sard-type theorems are obtained in [5].

Generalized Morse-Sard results are known in variational analysis, under a generalized notion of criticality, usually defined in terms of the Clarke subdifferential [4] (the definition is recalled in Section 2). Positive results are known in some particular cases: we quote for instance [14] for the distance function to the Riemanian submanifold and [13] for viscosity solutions of Hamiltonians of certain type. The Morse-Sard theorem obviously fails for general Lipschitz functions —it already fails for C^1 functions in $\mathbb{R}^d$ with $d \geq 2$— but the failure is of a different type: in the classical case, the failure of Morse-Sard theorem is due to the (bad) structure of the set of critical points (smooth functions are constant on rectifiable arcs made up of critical points, as a consequence of the chain rule). On the contrary, in [1] it is shown that for a generic (for the uniform topology) set of 1-Lipschitz functions, the Clarke subdifferential at any point equals the ball $B(0,1)$, that is, all points of a generic 1-Lipschitz function are critical (chain rule fails generically for the Clarke subdifferential).

In recent years, tame variational analysis is an alternative way to circumvent smoothness and to deal with generalized critical values. This leads to considering particular subclasses of Lipschitz functions enjoying *a prior* structural assumption: Lipschitz (nonsmooth) functions whose graphs are definable in some o-minimal structure (see [6] for the relevant definition). Indeed, under this tame assumption, Morse-Sard result follows as a consequence of the existence of a sufficiently smooth Whitney (normally regular) stratification, and the so-called projection formula for the Clarke subdifferential to the tangent space of the corresponding stratum ([2, Proposition 4]). This leads to two fundamental results in tame variational analysis:

– Clarke critical values for tame Lipschitz functions are locally finite [2] ;

– Tame closed graph set-valued maps, or more generally, set-valued maps admitting a sufficiently smooth normally regular stratification, satisfy Sard theorem for the set of their critical values (values where metric regularity fails) [9].

Let us point out that o-minimality (tameness) of a function does not automatically guarantee a Morse-Sard type result for any variational notion of criticality. For instance, let us call a point $\bar{x} \in \mathbb{R}^d$ broadly critical for the function f, if for every $\varepsilon > 0$, the closed convex hull of all derivatives $\nabla f(x)$ at points $x \in B(\bar{x}, \varepsilon)$ contains 0 (definability of the function guarantees differentiability in an open dense set). In [3], an example of a continuous (globally) subanalytic function $f : \mathbb{R}^3 \to \mathbb{R}$ which is strictly increasing in a segment of broadly critical points is presented, showing that the Morse-Sard theorem fails for this notion of criticality.

In contrast to the aforementioned tame-geometrical results, in this work we do not make any prior structural assumption on the nature of the functions. Instead, we consider the class of continuous functions $f : \mathbb{R}^d \to \mathbb{R}$ of the form

$$f(x) \in \{F(x,t) : t \in T\}, \quad \text{for all } x \in \mathbb{R}^d \tag{1}$$

where

($\mathcal{H}_1$) T is a nonempty compact countable set ;

($\mathcal{H}_2$) for each $t \in T$, the function $x \mapsto F(x,t)$ is C^k-smooth, with $k \geq d$;

($\mathcal{H}_3$) the functions F and $\nabla_x F$ (defined on $\mathbb{R}^d \times T$) are continuous.

We denote by $T(x)$ the set of active indices of F at x, that is,

$$T(x) = \{t \in T : f(x) = F(x,t)\}. \tag{2}$$

The main result of the manuscript is the following:

- Under ($\mathcal{H}_1$)-($\mathcal{H}_3$), every continuous function of the form (1) is locally Lipschitz and satisfies a generalized Morse-Sard theorem.

As a consequence, the max function $f(x) = \max\,\{F(x,t) : t \in T\}$ has a null set of generalized critical values. This result appears to be new even in the case where T is finite. The result is sharp, in the sense that it fails for $d \geq 3$ if the compact set T is uncountable (Remark 7).

The proof of our main result is based on geometrical arguments leading nonsmoothness to a kind of tractable smooth assumption in naturally arising manifolds, in which the classical Morse-Sard theorem applies.

The paper is organized as follows: in the next section we fix our notation, we recall the definition of the Clarke subdifferential and the Cantor-Bendixson rank of a set. The main result is proved in Section 3.

2 Preliminaries, Notation

For any integer $m \geq 0$, we set $\mathbb{N}_m = \{1, \ldots, m\}$ and we denote by

$$\Delta^m = \left\{ \alpha \in \mathbb{R}^{m+1} : \alpha_i \geq 0 : \sum_{i=1}^{m+1} \alpha_i = 1 \right\}$$

the m-dimensional simplex Δ^m. Given any nonempty subset C of $\mathbb{R}^d$, we denote by $\text{co}\,(C)$ the convex hull of C. Then $x \in \text{co}\,(C)$ if and only if there exists $m \geq 0$, $\alpha \in \Delta^m$ and $\{x_1, \ldots, x_{m+1}\} \subset C$ such that $x = \sum_{i=1}^{m+1} \alpha_i x_i$. By the Caratheodory theorem, we can always assume $m \leq d+1$. We finally denote by $|\mathrm{J}|$ the cardinality of any finite set J.

A function $f : \mathbb{R}^d \to \mathbb{R}$ is called locally Lipschitz continuous, if for every point $\bar{x}$ there exist a neighbordhood $\mathcal{U}$ of $\bar{x}$ and a constant $M > 0$ such that

$$|f(x) - f(y)| \leq M||x - y||, \quad \text{for all } x, y \in \mathcal{U}. \tag{3}$$

We recall that locally Lipschitz functions are differentiable almost everywhere (Rademacher theorem). Denoting by $\mathcal{D}_f$ the set of points where the derivative exists, the Clarke subdifferential of f at any point $\bar{x}$ is defined as follows:

$$\partial f(\bar{x}) = \text{co} \left\{ \lim_{x_n \to \bar{x}} \nabla f(x_n) : \{x_n\}_n \subset \mathcal{D}_f \right\}. \tag{4}$$

It follows that $\partial f(\bar{x})$ is a nonempty convex compact subset of $\mathbb{R}^d$ containing $\nabla f(\bar{x})$ whenever the latter exists, and being equal to $\{\nabla f(\bar{x})\}$ if and only if f is strictly differentiable at $\bar{x}$ (see [4] for details). A point $\bar{x} \in \mathbb{R}^d$ is called *Clarke critical point* if $0 \in \partial f(\bar{x})$. A value $\bar{r} \in f(\mathbb{R}^d)$ is called *Clarke critical value* if $\bar{r} = f(\bar{x})$ for some Clarke critical point $\bar{x}$.

Given two metric spaces X and Y, a set-valued mapping $A : X \rightrightarrows Y$ is said to be upper semicontinuous, if for every open set $\mathcal{U}$ containing $A(x_0)$, there exists an open set $\mathcal{V}$ containing x_0 such that $A(x) \subset \mathcal{U}$ for all $x \in \mathcal{V}$. We say that A is locally bounded (respectively, locally compact), if every $x_0 \in X$ has a neighborhood $\mathcal{V}$ whose image $A(\mathcal{V}) := \bigcup_{x \in \mathcal{V}} A(x)$ is bounded (respectively, relatively compact). If f is locally Lipschitz, the Clarke subdifferential $\partial f : \mathbb{R}^d \rightrightarrows \mathbb{R}^d$ is locally compact and upper semicontinuous (see [4]) with nonempty convex compact calues. We further say that A has a closed graph, if $\text{Graph}\,(A) := \{(x, y) : y \in A(x)\}$ is a closed subset of $X \times Y$. The following statement is well-known.

- Let $A : X \rightrightarrows Y$ be locally compact with nonempty closed values. Then A is upper semicontinuous if, and only if, it has a closed graph.

Given a nonempty closed set T we denote by T' the *(Cantor) derivative* of T, that is, the set of all accumulation points of T (see [10] *e.g.*). Obviously T' is a closed subset of T (if $T = T'$ the set T is called perfect). In general, using transfinite induction, we define a decreasing chain of closed subsets T^β as follows: $T^0 := T$, $T^{\beta+1} = (T^\beta)'$ for a successive ordinal $\beta + 1$, and $T^\xi = \bigcap_{\beta < \xi} T^\beta$ for a limiting ordinal ξ. In the sequel we shall use the following fact:

- If T is a compact countable space, then there exists a countable ordinal λ such that $T^\lambda = \emptyset$.

The above statement is a consequence of the fact that every countable compact space is metrizable, thus by Baire theorem it has at least one isolated point.

3 Main result

In this section we establish our main result, namely, the set of Clarke critical values of any continuous selection (1) is null, provided $(\mathcal{H}_1)$-$(\mathcal{H}_3)$ hold. To this end, we shall need several intermediate results, some of them of independent interest. Let us state the following result, whose proof is standard and will be omitted.

Lemma 1 (Upper semicontinuity of the map of active indices)**.** *The set-valued mapping*

$$x \rightrightarrows T(x) = \{t \in T : f(x) = F(x,t)\}$$

(defined in (2)) is upper semicontinuous with nonempty compact values.

Our next aim is to show that continuous selections of the form (1) inherit from the family $F(\cdot,t)$ the property of being (locally) Lipschitz continuous, as a consequence of assumption $(\mathcal{H}_1)$. For this result assumptions $(\mathcal{H}_2)$-$(\mathcal{H}_3)$ are relaxed. On the other hand, obvious counterexamples can be found if T is uncountable or if it is not compact.

Proposition 2 (Lipschitz continuity of f)**.** *Let $T \neq \emptyset$ be compact countable and $F : \mathbb{R}^d \times T \to \mathbb{R}$ be continuous. Let $f : \mathbb{R}^d \to \mathbb{R}$ be a selection of F, that is,*

$$f(x) \in \{F(x,t) : t \in T\}, \quad \textit{for all } x \in \mathbb{R}^d.$$

(i) If f has closed graph, then it is continuous.

(ii) If in addition, $x \mapsto F(x,t)$ is locally Lipschitz, uniformly on $t \in T$, then f is locally Lipschitz.

Proof. (i) Let $\{x_n\} \to \bar{x}$ and let $t_n \in T(x_n)$ so that $f(x_n) = F(x_n, t_n)$. Since F is continuous and T is compact, the sequence $\{f(x_n)\}_n$ is bounded. Further, any accumulation point of the sequence $\{f(x_n)\}_n$ has to be equal to $f(\bar{x})$, since f has closed graph. This proves that $\{f(x_n)\}_{n\geq 1}$ converges to $f(\bar{x})$ and the assertion follows.

(ii) Let $x_0 \in \mathbb{R}^d$. According to our assumption, there exists $M > 0$ and a convex neighborhood $\mathcal{U}$ of x_0 such that for all $x, y \in \mathcal{U}$ and $t \in T$

$$|F(x,t) - F(y,t)| \leq M\,||x - y||.$$

We shall show that f is Lipschitz continuous on $\mathcal{U}$ with constant M, that is, for any $\bar{x}, \bar{y} \in \mathcal{U}$ we have

$$|f(\bar{y}) - f(\bar{x})| \leq M\,||\bar{y} - \bar{x}||. \tag{5}$$

To prove (5), fix $\bar{x}, \bar{y} \in \mathcal{U}$ with $\bar{x} \neq \bar{y}$ and define the function $\rho : [\bar{x}, \bar{y}] \to [\bar{x}, \bar{y}]$ as follows:

$$\rho(x) = \sup\,\{\, z \in [x, \bar{y}] : |f(z) - f(x)| \leq M\,||z - x||\}\,. \tag{6}$$

Since f is continuous, the above supremum is in fact a maximum. We further define

$$\mathcal{F} = \{x \in [\bar{x}, \bar{y}] : \rho(x) = x\}. \tag{7}$$

Notice that $\rho(x) \geq x$ and that

$$\rho(\rho(x)) = \rho(x). \tag{8}$$

Thus, if we prove that $\mathcal{F} = \{\bar{y}\}$, then (5) follows. The last part of the proof is devoted to show that $\mathcal{F}$ reduces to the singleton $\{\bar{y}\}$. We shall need the following assertion.

Claim 1. If $(\hat{x}, \hat{x}+\delta) \subset [\bar{x}, \bar{y}] \diagdown \mathcal{F}$ for some $\delta > 0$ and $\hat{x} \in [\bar{x}, \bar{y})$, then $\hat{x} \notin \mathcal{F}$.

Proof of Claim 1. Let us first show that for every $x \in (\hat{x}, \hat{x}+\delta)$ we have $\rho(x) \geq \hat{x}+\delta$. Indeed, if this were not the case, then for some $x_1 \in (\hat{x}, \hat{x}+\delta)$ we would have $\hat{x} < x_1 < \rho(x_1) < \hat{x}+\delta$, which in view of (8) yields $\rho(x_1) \in (\hat{x}, \hat{x}+\delta) \cap \mathcal{F}$, a contradiction. Now let us consider a strictly decreasing sequence $\{\varepsilon_n\} \searrow 0$. By continuity of f, for every $n \geq 1$ there exists $\delta_n > 0$ such that for all $x \in (\hat{x}, \hat{x}+\delta_n)$ we have $|f(\hat{x}) - f(x)| < \varepsilon_n$. We can clearly assume that the sequence $\{\delta_n\}_n$ is decreasing and $\delta_1 < \delta$. Then for $x_n \in (\hat{x}, \hat{x}+\delta_n)$, we have $\rho(x_n) \geq \hat{x}+\delta$ and $|f(\hat{x}) - f(x_n)| < \varepsilon_n$. It follows

$$|f(\hat{x}) - f(\rho(x_n))| \leq \varepsilon_n + |f(x_n) - f(\rho(x_n))| \leq \varepsilon_n + M||x_n - \rho(x_n)|| < \varepsilon_n + M||\hat{x} - \rho(x_n)||. \quad (9)$$

Since $\rho(x_n) \in [\hat{x}+\delta, \bar{y}]$, passing to a subsequence if necessary, we may assume $\rho(x_n) \to z^* \in [\hat{x}+\delta, \bar{y}]$ as $n \to \infty$. We deduce from (9) that

$$|f(\hat{x}) - f(z^*)| \leq M||\hat{x} - z^*||,$$

yielding $\rho(\hat{x}) \geq z^* \geq \hat{x}+\delta$, that is, $\hat{x} \notin \mathcal{F}$. $\diamond$

Before we proceed to the rest of the proof, let us introduce the following notation:

For any ordinal β, we define

$$X_\beta := \{x \in [\bar{x}, \bar{y}] : T(x) \cap T^\beta \neq \emptyset\},$$

where $T(x)$ is defined in (2) and T^β denotes the Cantor-Bendinxon derivative of order β of the compact set T. We shall now prove the following assertion.

Claim 2. For every ordinal β it holds:

$$\hat{x} \in \mathcal{F} \diagdown \{\bar{y}\} \Longrightarrow \hat{x} \in X_\beta. \quad (10)$$

Proof of Claim 2. Let $\hat{x} \in \mathcal{F} \diagdown \{\bar{y}\}$. We deduce by Claim 1 that there exists $\{x_n\}_n \subset \mathcal{F} \cap (\hat{x}, \bar{y})$ with $x_n \to \hat{x}$. Pick $t_n \in T(x_n)$, so that $f(x_n) = F(x_n, t_n)$. We may assume, passing possibly to a subsequence, that $\{t_n\}_n$ converges to some $\hat{t} \in T$. Then taking the limit as $n \to +\infty$ we deduce $f(\hat{x}) = F(\hat{x}, \hat{t})$, thus $\hat{t} \in T(\hat{x})$. Note further that $\hat{t}$ cannot be equal to t_n for some n, since in such case we would have

$$|f(\hat{x}) - f(x_n)| = |F(\hat{x}, \hat{t}) - F(x_n, \hat{t})| \leq M||\hat{x} - x_n||,$$

yielding $\rho(\hat{x}) \geq x_n > \hat{x}$ and contradicting the fact that $\hat{x} \in \mathcal{F}$. It follows that $\hat{t} \in T(\hat{x}) \cap T'$ and $\hat{x} \in X'$. Thus the assertion holds for $\beta = 1$.

We shall use transfinite induction. Let us first assume that $\mathcal{F} \diagdown \{\bar{y}\} \subset X_\beta$ for some ordinal β and let us assume, towards a contradiction, that there exists $\hat{x} \in \mathcal{F} \diagdown \{\bar{y}\}$ with $\hat{x} \notin X_{\beta+1}$. It follows easily by the latter and the compactness of T^β that for some $\delta > 0$ we have $(\hat{x}, \hat{x}+\delta) \cap X_\beta = \emptyset$. In view of our inductive assumption, $(\hat{x}, \hat{x}+\delta) \cap \mathcal{F} = \emptyset$, hence by Claim 1, $\hat{x} \notin \mathcal{F}$ a contradiction.

It remains to consider the case of a limiting ordinal ξ. To this end, let us assume that (10) holds for all ordinals $\beta < \xi$ and that $\hat{x} \in \mathcal{F} \diagdown \{\bar{y}\}$. It follows that $\hat{x} \in X_\beta$ for every $\beta < \xi$, hence the set $\mathcal{T}_\beta := T(\hat{x}) \cap T^\beta$ is nonempty and compact. Moreover, for $\beta_1 < \beta_2$ we have $\mathcal{T}_{\beta_1} \supset \mathcal{T}_{\beta_2}$. It follows readily that the family $\{\mathcal{T}_\beta\}_{\beta<\xi}$ has the finite intersection property, thus $\bigcap_{\beta<\xi} \mathcal{T}_\beta = T(\hat{x}) \cap T^\xi \neq \emptyset$. Therefore, $\hat{x} \in X_\xi$. This completes the transfinite induction and the claim is proved. $\diamond$

Since T is a countable compact set, there exists a (countable) ordinal λ such that $T^\lambda = \emptyset$. It follows that $X_\lambda = \emptyset$ and $\mathcal{F} = \{\bar{y}\}$. The proof is complete. $\square$

As a consequence of the previous proposition, the Clarke subdifferential enters naturally into consideration for the study of continuous functions of the form (1). Our next objective is to associate the subdifferential $\partial f(x)$ of the selection f with the active derivatives $\nabla_x F(x,t)$, $t \in T(x)$, of the corresponding family. This is the aim of the forthcoming Proposition 4. We shall need the following one-dimensional result.

Lemma 3 (one-dimensional case)**.** *Let $h : \mathbb{R} \to \mathbb{R}$ be a continuous function satisfying*

$$h(x) \in \{g_t(x) : t \in T\}, \quad \textit{for all } x \in \mathbb{R},$$

where the functions $(x,t) \longmapsto g_t(x)$, and $(x,t) \longmapsto g_t'(x)$ are continuous and T is a countable compact set. Then (h is locally Lipschitz and) for every point of differentiability $x \in \mathcal{D}_h$ it holds

$$h'(x) \in \mathrm{co}\left\{g_t'(x) : t \in T(x)\right\}. \tag{11}$$

Proof. Let us assume, towards a contradiction, that (11) fails for some $\bar{x} \in \mathcal{D}_h$. We may assume $\bar{x} = 0$, $h(0) = 0$, $h'(0) = 0$ and $g_t'(0) \geq 2$ for all $t \in T(0)$ (there is no loss of generality in doing this). Since the mapping $(x,t) \longmapsto g_t'(x)$ is continuous around $(0,\bar{t})$, for all $\bar{t} \in T(0)$ and since $T(0)$ is compact, we infer that for some $\delta_0 > 0$ and some open neighborhood T_0 of $T(0)$ we have

$$g_t'(x) \geq 1, \quad \text{for all } x \in [0,\delta_0] \text{ and all } t \in T_0. \tag{12}$$

Since $h(0) = h'(0) = 0$, shrinking $\delta_0 > 0$ if needed, we guarantee that

$$|h(x)| \leq \frac{1}{2}x, \quad \text{for all } x \in [0,\delta_0]. \tag{13}$$

Let us define the function $\hat{\rho} : [0,\delta_0] \to [0,\delta_0]$ as follows:

$$\hat{\rho}(x) = \sup\left\{\, z \in [x,\delta_0] : h(z) - h(x) \geq z - x \right\}. \tag{14}$$

By continuity of h we deduce that the supremum in (14) is in fact a maximum. Notice also that $\hat{\rho}(x) \geq x$ and $\hat{\rho}(\hat{\rho}(x)) = \hat{\rho}(x)$. Defining

$$\hat{\mathcal{F}} := \{x \in [0,\delta_0] : \hat{\rho}(x) = x\},$$

using (12) and proceeding in a similar way as in the proof of Proposition 2 (*Claims* 1 and 2) we can show that $\mathcal{F} = \{\delta_0\}$. This yields $\hat{\rho}(0) > 0$, thus combining with (14) and (13) we get

$$\frac{1}{2}\hat{\rho}(0) \geq h(\hat{\rho}(0)) = h(\hat{\rho}(0)) - h(0) \geq \hat{\rho}(0), \tag{15}$$

a clear contradiction. $\square$

We are now ready to relate the Clarke subdifferential of the selection f with the active derivatives of the family $F(\cdot,t)$, $t \in T$.

Proposition 4 (Clarke subdifferential versus active derivatives)**.** *Let $f : \mathbb{R}^d \to \mathbb{R}$ be any continuous function satisfying*

$$f(x) \in \{F(x,t) : t \in T\}, \quad \textit{for all } x \in \mathbb{R}^d,$$

where F, $\nabla_x F$ are continuous and T is a countable compact set. Then f is locally Lipschitz and

$$\partial f(x) \subset \operatorname{co} \{\nabla_x F(x,t) : t \in T(x)\}, \qquad \textit{for all } x \in \mathbb{R}^d. \tag{16}$$

Before we proceed, let us introduce the set-valued map $A_f : \mathbb{R}^d \rightrightarrows \mathbb{R}^d$ defined by

$$A_f(x) = \operatorname{co} \{\nabla_x F(x,t) : t \in T(x)\}. \tag{17}$$

The above proposition actually claims that the Clarke subdifferential $\partial f(x)$ is contained in $A_f(x)$. This set-valued map will be also in use in our main result.

Proof of Proposition 4. The set-valued mapping $x \rightrightarrows C(x) := \{\nabla_x F(x,t) : t \in T(x)\}$ has a closed graph, nonempty compact values and is locally compact, thus it is upper semicontinuous. It follows readily from the definition that the set-valued mapping $A_f(x) := \operatorname{co} C(x)$ is upper semicontinuous with nonempty convex compact values, therefore, it has a closed graph. In view of the definition of Clarke subdifferential, see (4), inclusion (16) will follow from the relation

$$\nabla f(x) \in A_f(x), \qquad \text{for all } x \in \mathcal{D}_f. \tag{18}$$

Let us prove that (18) holds. We proceed by contradiction: assume that for some $\bar{x} \in \mathcal{D}_f$ the functional $\ell := \nabla f(\bar{x})$ does not belong to $A_f(\bar{x})$. We may assume for simplicity that $\ell = 0$, since we can always replace $f(x)$ by $f(x) - \langle \ell, x \rangle$ and $F(x,t)$ by $F(x,t) - \langle \ell, x \rangle$. Then by the Hahn-Banach theorem, there exists a direction $e \in \mathbb{R}^d$, such that for all $t \in T(\bar{x})$ we have $\langle \nabla_x F(\bar{x},t), e \rangle > 1 > 0 = \langle \nabla f(\bar{x}), e \rangle$. Let us consider the restrictions of f and $F(\cdot,t)$, for every $t \in T$, on the line $\bar{x} + \mathbb{R}e$, that we call h and $g_t(\cdot)$ respectively. In other words, for every $s \in \mathbb{R}$ we set

$$h(s) := f(\bar{x} + se) \quad \text{and} \quad g_t(s) := F(\bar{x} + se, t).$$

It follows that $h'(0) = \langle \nabla f(\bar{x}), e \rangle = 0$ and $g_t'(0) = \langle \nabla_x F(\bar{x},t), e \rangle > 1$, $t \in T(\bar{x})$. Applying Lemma 3 we get a contradiction. The proof is complete. $\square$

We are now ready to prove our main result for continuous selections of the form (1). The notion of criticality is based on the set-valued mapping A_f, so it applies *a fortiori* for the Clarke critical values.

Theorem 5 (Generalized Morse-Sard theorem)**.** *Let $f : \mathbb{R}^d \to \mathbb{R}$ be a continuous function of the form*

$$f(x) \in \{F(x,t) : t \in T\}, \quad \textit{for all } x \in \mathbb{R}^d$$

where $x \mapsto F(x,t)$ is a family of C^k-smooth functions, with $k \geq d$, and T a countable compact set. We assume that F and $\nabla_x F$ are continuous on $\mathbb{R}^d \times T$. Then the set of Clarke critical values of f has measure zero.

Proof. Recalling by (17) the definition of the set-valued mapping A_f, we define the set of *broadly critical* points of f as follows:

$$S = \{x \in \mathbb{R}^d : 0 \in A_f(x)\}.$$

In view of Proposition 4, we have $x \in S$ whenever $0 \in \partial f(x)$. Therefore the set of broadly critical values $f(S)$ contains the set of Clarke critical values. We shall establish the following (generalized) Morse-Sard theorem: the set $f(S)$ of broadly critical values of f is null.

To this end, let $x \in S$. Applying the Caratheodory theorem we deduce that there exist a finite subset J of $T(x)$, with $|\mathrm{J}| \leq d+1$, and $\alpha \in \Delta^{|\mathrm{J}|}$ such that

$$\sum\nolimits_{t\in \mathrm{J}} \alpha_t \nabla_x F(x,t) = 0. \tag{19}$$

We say that (19) defines a minimal representation of 0 for the critical point $x \in S$, if it involves a minimum number of active gradients $\nabla_x F(x,t)$, $t \in T(x)$. We further denote by $S(\ell)$, for $1 \leq \ell \leq d+1$, the set of broadly critical points of f, admitting a minimal representation of 0 made of exactly ℓ active gradients. Clearly,

$$S = \bigcup\nolimits_{\ell=1}^{d+1} S(\ell). \tag{20}$$

In order to establish that $f(S)$ has measure zero, it suffices to prove that for every $\ell \in \mathbb{N}_{d+1}$ the set $f(S(\ell))$ is contained in a null-set.

- Let us first consider the case $\ell = 1$. Then

  $$S(1) = \{x \in \mathbb{R}^d : \exists t \in T(x),\ \nabla_x F(x,t) = 0\}.$$

  Denoting by S_t the set of critical points of the C^k function $x \mapsto F(x,t)$, we have $S(1) \subset \cup_{t\in T} S_t$ and

  $$f(S(1)) \subset \bigcup\nolimits_{t\in T} F(S_t, t).$$

  Applying the classical Morse-Sard theorem to each function $x \mapsto F(x,t)$, we deduce that $f(S(1))$ is contained in a countable union of null-sets, thus it is also a null-set.

- Fix $1 < \ell \leq d+1$. For every finite subset J of T of cardinality $|\mathrm{J}| = \ell$ we set

  $$S(\ell,\mathrm{J}) := \left\{x \in S(\ell) : \mathrm{J} \subset T(x),\ \exists \alpha \in \Delta^{|\mathrm{J}|} : \sum\nolimits_{t\in\mathrm{J}} \alpha_t \nabla_x F(x,t) = 0\right\}. \tag{21}$$

  Let us prove that for every $x \in S(\ell, \mathrm{J})$, the vector space

  $$W_x := \mathrm{span}\ \{\nabla_x F(x,t) : t \in \mathrm{J}\}$$

  has dimension $\ell - 1$. Indeed, since $0 \in \mathrm{co}\{\nabla_x F(x,t) : t \in \mathrm{J}\} \subset W_x$, the Caratheodory theorem asserts that for some $\mathrm{J}^* \subset \mathrm{J}$ of cardinality $|\mathrm{J}^*| = \dim W_x + 1$ we have $0 \in \mathrm{co}\{\nabla_x F(x,t) : t \in \mathrm{J}^*\}$. Since the representation (19) in (21) is minimal, we deduce $\mathrm{J}^* = \mathrm{J}$ and $\ell = \dim W_x + 1$, that is, $\dim W_x = \ell - 1$.

  Assume $\mathrm{J} := \{t_1, t_2, \ldots, t_\ell\}$, set $\Phi_i(x) = F(x,t_i) - F(x,t_\ell)$ for $i \in \mathbb{N}_{\ell-1}$ and consider the function

  $$\begin{cases} \Phi : \mathbb{R}^d \to \mathbb{R}^{\ell-1} \\ \Phi = (\Phi_1, \ldots, \Phi_{\ell-1}). \end{cases}$$

We further define $E_{\rm J} = \Phi^{-1}(0)$ and

$$\mathcal{M}_{\rm J} = \{x \in E_{\rm J} : \operatorname{rg} D\Phi(x) = \ell - 1\},$$

where $\operatorname{rg} D\Phi(x)$ stands for the dimension of the image of $\mathbb{R}^d$ under $D\Phi(x)$. It follows by the implicit function theorem that $\mathcal{M}_{\rm J}$ is a C^k-differentiable manifold of co-dimension $\ell-1$ (for $\ell = d+1$ it reduces to a locally finite union of points).

Notice that if $x \in S(\ell, {\rm J})$, then $x \in E_{\rm J}$ and the gradients $\{\nabla\Phi_i(x) : i \in \mathbb{N}_{\ell-1}\}$ are linearly independent. It follows that

$$S(\ell, {\rm J}) \subset \mathcal{M}_{\rm J}.$$

Let $\varphi_{\rm J}$ denote the restriction of any of the functions $x \mapsto F(x, t_i)$ to $\mathcal{M}_{\rm J}$ (they are all equal there). Then $\varphi_{\rm J}$ is C^k differentiable on $\mathcal{M}_{\rm J}$ and its Riemann gradient at $x \in \mathcal{M}_{\rm J}$ satisfies

$$d\varphi_{\rm J}(x) = \operatorname{proj}_{T_{\mathcal{M}_{\rm J}}(x)} \left[\nabla_x F(x, t_i)\right], \quad \text{for all } i \in \mathbb{N}_\ell,$$

where $\operatorname{proj}_{T_{\mathcal{M}_{\rm J}}(x)}$ denotes the projection operator onto the tangent space $T_{\mathcal{M}_{\rm J}}(x)$ at x of the manifold $\mathcal{M}_{\rm J}$. The above formula together with (19) show that any critical point $x \in S(\ell, {\rm J})$ satisfies $d\varphi_{\rm J}(x) = 0$, that is, $S(\ell, {\rm J})$ is contained in the set $S_{\rm J} = \{x \in \mathcal{M}_{\rm J} : d\varphi_{\rm J}(x) = 0\}$ of critical points of the function $\varphi_{\rm J} : \mathcal{M}_{\rm J} \to \mathbb{R}$. Notice further that $\varphi_{\rm J}(x) = f(x) = F(x, t_1)$ for all $x \in S(\ell, {\rm J})$, yielding

$$f(S(\ell, {\rm J})) \subset \varphi_{\rm J}(S_{\rm J}). \tag{22}$$

Applying the Morse-Sard theorem to the C^k function $\varphi_{\rm J}$ over the C^k differentiable manifold $\mathcal{M}_{\rm J}$ ([8, Theorem 1.3] *e.g.*) we conclude that $f(S(\ell, {\rm J}))$ has measure zero. Let us finally notice that

$$S(\ell) \subset \bigcup_{{\rm J} \subset T,\, |{\rm J}| = \ell} S(\ell, {\rm J}). \tag{23}$$

It follows that $f(S(\ell))$ is contained in the countable union of the null-sets $f(S(\ell, {\rm J}))$, thus it has measure zero.

It now follows from (20) that $f(S)$ has measure zero, which completes the proof. $\square$

The following corollary is an immediate consequence of Theorem 5.

Corollary 6 (Morse-Sard theorem for functions of max type)**.** *Consider the locally Lipschitz function*

$$f(x) = \max_{t \in T} F(x, t), \quad x \in \mathbb{R}^d \tag{24}$$

where T is a countable compact set and for every $t \in T$ the function $x \mapsto F(\cdot, t)$ is C^k-differentiable with $k \geq d$. If F and $\nabla_x F$ are continuous, the set of Clarke critical values of f has measure zero.

We recall that a function f is called lower-C^k if it can be represented as a maximum of a family of C^k-functions $x \mapsto F(x, t)$, indexed on a compact set T, and such that $F, \nabla_x F, \ldots, \nabla_k F$ are continuous on $\mathbb{R}^d \times T$ (see [15], [7] *e.g.*). Lower-C^k functions are Lipschitz continuous and Clarke regular (see [4], [15] for relevant definitions and proofs). Corollary 6 reveals in particular that the Morse-Sard theorem holds true for lower-C^k functions, provided the compact set T is countable. The following remark shows that this assumption cannot be relaxed, revealing that both Corollary 6 and Theorem 5 are sharp.

Remark 7 (Sharpness of the main result)**.** (i). It is well-known ([15]) that every lower-C^2 function, thus a fortiori, every C^2 function, can be parameterized to become lower-C^∞. Therefore, the Morse-Sard theorem fails for functions of the form (24) —and consequently for Lipschitz continuous functions of the form (1)— if T is not countable and $d \geq 3$.

(ii) Compactness of T is not essential for the proof of our main result, it had been previously required to relate ∂f to A_f (*c.f.* Proposition 4). In particular, the proof of Theorem 5 shows that the set of broadly critical values of any function satisfying (1) is null, provided T is countable and $(\mathcal{H}_2)$-$(\mathcal{H}_3)$ hold.

Another consequence of Theorem 5 is the following result.

Corollary 8 (Max/min operations over a collection of smooth functions)**.** *Let $\mathcal{F}_0 = \{f_i\}_{i \in I}$ be a (possibly uncountable) collection of C^k functions on $\mathbb{R}^d$ with $k \geq d$. Set $\mathcal{F}_1$ be the collection of all max- or min-type functions of the form $\max\{f_i : i \in \mathrm{J}\}$ or $\min\{f_i : i \in \mathrm{J}\}$ where J is a finite subset of I. We construct inductively $\mathcal{F}_{m+1}$ as the collection of all finite maxima or minima of functions in $\mathcal{F}_m$. Then every function $f \in \bigcup_m \mathcal{F}_m$ is locally Lipschitz and has a null set of Clarke critical values.*

Proof. It is easily seen that if $f \in \mathcal{F}_m$ for some $m \in \mathbb{N}$, then there exists a finite set $\mathrm{J} \subset I$ such that f is locally Lipschitz and satisfies (1). The result follows from Theorem 5. $\square$

Acknowledgement. This work has been realized during a research stay of the third author at the University of Pau. This author thanks the host institution for hospitality.

References

[1] J. Borwein, X. Wang, Lipschitz functions with maximal subdifferentials are generic, *Proc. Amer. Math. Soc.* **128** (2000), 3221–3229.

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

[3] J. Bolte, A. Daniilidis, A. Lewis, M. Shiota, Clarke critical values of subanalytic Lipschitz continuous functions. *Ann. Polon. Math.* **87** (2005), 13–25.

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

[5] G. Compte, Y. Yomdin, *Tame Geometry with Applications in Smooth Analysis*, LNM 1834 (Springer 2004).

[6] M. Coste, *An introduction to o-minimal geometry*, RAAG Notes, 81 p., Institut de Recherche Mathématiques de Rennes, November 1999.

[7] A. Daniilidis, J. Malick, Filling the gap between lower-C^1 and lower-C^2 functions, *J. Convex Anal.* **12** (2005), 315–329.

[8] W. Hirsch, *Differential Geometry*, Graduate Texts in Mathematics **33** (Springer, 1976).

[9] A. Ioffe, Critical values of set-valued maps with stratifiable graphs. Extensions of Sard and Smale-Sard theorems, *Proc. Amer. Math. Soc.* **136** (2008), 3111–3119.

[10] A.S. Kechris, *Classical Descriptive Set Theory* (Springer 1995).

[11] A. Morse, The behavior of a function on its critical set, *Ann. of Math.* **40** (1939), 62–70.

[12] A. Norton, A critical set with nonnull image has large Hausdorff dimension, *Trans. Amer. Math. Soc.* **296** (1986), 367–376.

[13] L. Rifford, On viscosity solutions of certain Hamilton-Jacobi equations: regularity results and generalized Sard's theorems. *Comm. Partial Differential Equations* **33** (2008), 517–559.

[14] L. Rifford, A Morse-Sard theorem for the distance function on Riemannian manifolds, *Manusc. Math.* **113** (2004), 251–265.

[15] R. T. Rockafellar, R. Wets, *Variational Analysis* (Springer, 1998).

[16] A. Sard, The measure of the critical values of differentiable maps, *Bull. Amer. Math. Soc.* **48** (1942), 883–890.

[17] H. Whitney, A function not constant on a connected set of critical points, *Duke Math. J.* **1** (1935), 514–517.

[18] Y. Yomdin, The geometry of critical and near-critical values of differentiable mappings. *Math. Ann.* **264** (1983), 495–515

Luc Barbet, Marc Dambrine

Laboratoire de Mathématiques et de leurs Applications, UMR CNRS 5142
Université de Pau et des Pays de l'Adour
Avenue de l'Université - BP 1155, 64013 Pau, France.

E-mail: `luc.barbet@univ-pau.fr ; marc.dambrine@univ-pau.fr`

Aris Daniilidis

Departament de Matemàtiques, C1/308
Universitat Autònoma de Barcelona
E-08193 Bellaterra, Spain

E-mail: `arisd@mat.uab.es`
`http://mat.uab.es/~arisd`

Research supported by the MEC Grant No. MTM2011-29064-C01 (Spain).