Sard theorems for Lipschitz functions and applications in Optimization.

L. Barbet, M. Dambrine, A. Daniilidis, L. Rifford

Abstract We establish a "preparatory Sard theorem" for smooth functions with a partial affine structure. By means of this result, we improve a previous result of Rifford [14, 16] concerning the generalized (Clarke) critical values of Lipschitz functions defined as minima of smooth functions. We also establish a nonsmooth Sard theorem for the class of Lipschitz functions from $\mathbb{R}^d$ to $\mathbb{R}^p$ that can be expressed as finite selections of C^k functions (more generally, continuous selections over a compact countable set). This recovers readily the classical Sard theorem and extends a previous result of Barbet-Daniilidis-Dambrine [1] to the case $p > 1$. Applications in semi-infinite and Pareto optimization are given.

Key words Sard theorem, Lipschitz function, Clarke generalized Jacobian, semi-infinite optimization.

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

Contents

1 Introduction

The classical Sard theorem asserts that the critical values of a C^k smooth function $f : \mathbb{R}^d \to \mathbb{R}^p$ are contained in a subset of null measure of $\mathbb{R}^p$, provided $k \geq \max\{1, d-p+1\}$, see [18]. The case $p = 1$, known as the Morse-Sard theorem, had been previously established in [12]. The Sard theorem can be readily extended to C^k-functions $f : \mathcal{M} \to \mathcal{N}$ where $\mathcal{M}$, $\mathcal{N}$ are C^k-manifolds of dimensions d and p respectively (see [9, Theorem 1.3] *e.g.*). Notice that if $d < p$ then every point is critical and $f(\mathcal{M})$ has null measure in $\mathcal{N}$. The above results are essentially sharp. (We refer to [13] and [2] for refinements and precise sharp statements.) Known counterexamples show that the degree of regularity cannot be lowered, unless an extra structural assumption is made over the function: see [11] for C^1 *semialgebraic* functions from $\mathbb{R}^d$ to $\mathbb{R}^p$.

In this work we focus on the case of (locally) Lipschitz continuous functions $f : \mathbb{R}^d \to \mathbb{R}^p$. The notion of criticality is now defined by means of the generalized (Clarke) Jacobian (see definition in Section 2.1). If f is Lipschitz, the notation $\mathrm{Crit} f$ (respectively, $f(\mathrm{Crit} f)$) will refer to the set of Clarke critical points (respectively, values) of f. Let us notice that if $d < p$, then $f(\mathbb{R}^d)$ is a null set in $\mathbb{R}^p$ and the conclusion of the Sard theorem holds trivially. For this reason, unless overwise specified, we assume $d \geq p$. However, endeavoring to generalize the Sard theorem for (nonsmooth) Lipschitz functions is a huge challenge even in the simplest nontrivial case $d = p$. Indeed, approximating the given Lipschitz function by functions of class C^1 outside a set of small measure (see [8, 3.1.15]), we deduce that the image of the set of the classical critical points (that is, points where the derivative of f exists and is not surjective) has measure zero. Nonetheless, this does not yield any control on the set of Clarke critical values: it can be shown that every point of a *generic* (with respect to the uniform topology) Lipschitz function $f : \mathbb{R} \to \mathbb{R}$ is Clarke critical (see [4] for the more general case of functions defined in $\mathbb{R}^d$). This wipes out any possible nonsmooth Sard-type result in full generality, indicating instead that extra assumptions are required. Should this be a structural assumption in the spirit of Grothendieck (subanalyticity, tameness of the graph of the Lipschitz function), a strong version of Morse-Sard theorem can then be established (local finiteness of the Clarke critical values), see [3]. Without such structural assumptions, a couple of ad-hoc nonsmooth Morse-Sard results can still be found in the literature for particular Lipschitz functions: the distance function to a Riemanian submanifold [14] and the viscosity solutions of Hamiltonians of certain type [16].

In this work, we improve the aforementioned results of Rifford, by establishing the following nonsmooth Morse–Sard theorem, which has interesting applications in Riemannian [14] and sub-Riemannian geometry [15] and more generally in the Hamilton-Jacobi theory [16].

Theorem 1 (Morse–Sard for min-type functions)**.** *Let $\mathcal{N}$ be a compact manifold of class C^k and of dimension ℓ. Let $\phi : \mathbb{R}^n \times \mathcal{N} \to \mathbb{R}$ be a smooth function of class C^k. If $k \geq n + \ell(n+1)$, then the set of Clarke critical values of the Lipschitz continuous function*

$$\begin{cases} f : \mathbb{R}^n \to \mathbb{R} \\ f(z) := \min_{q \in \mathcal{N}} \{\phi(z, q)\}, \qquad \textit{for all } z \in \mathbb{R}^n. \end{cases}$$

has measure zero.

A second contribution of this work is to provide a nonsmooth generalization of the Sard theorem for Lipschitz functions $f : \mathbb{R}^d \to \mathbb{R}^p$ that are selections over a finite family of C^k smooth functions, where $k \geq d - p + 1$. (This is of course the minimal regularity that one should require, it corresponds to the regularity for the classical Sard theorem to hold.) This result recovers remarkably the classical Sard theorem (consider the trivial selection over a singleton). Our forthcoming theorem will be stated in an even more general form, considering selections over a (possibly infinite) countable family. In particular it extends the recent result of Barbet, Daniilidis, Dambrine [1, Theorem 5], from the real-valued case $p = 1$ (Morse theorem) to the vector valued case $p > 1$ (Sard theorem).

Theorem 2 (Sard for Lipschitz selections)**.** *Let $T \neq \emptyset$ be a compact countable set. Assume that*
– $F : \mathbb{R}^d \times T \to \mathbb{R}^p$ is a continuous function with $d \geq p$;
– $x \mapsto F(x,t)$ is of class C^k with $k \geq d-p+1$, for all $t \in T$, and
– the function $D_xF : \mathbb{R}^d \times T \to \mathbb{R}^{pd}$ is continuous.
If $f : \mathbb{R}^d \to \mathbb{R}^p$ is continuous and

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

then f is locally Lipschitz and $f(\mathrm{Crit} f)$ is null in $\mathbb{R}^p$.

(If T is finite, the assumptions of continuity of F and D_xF become superfluous.)

Both Theorem 1 (Morse–Sard for min-type functions) and Theorem 2 (Sard for Lipschitz selections) are obtained as corollaries –though not straightforward– of the forthcoming (main) result that we call "Preparatory Sard theorem". Some notation is needed in order to state this latter result: $\mathrm{inn}\,\Delta^m$ stands for the algebraic interior of the simplex Δ^m of $\mathbb{R}^{m+1}$, $\mathcal{M}$ denotes a C^k smooth manifold and

$$\Psi(\lambda,x) := \sum_{i=0}^{m} \lambda_i \phi^i(x), \quad (\lambda,x) \in \mathrm{inn}\,\Delta^m \times \mathcal{M}$$

where $\phi^i : \mathcal{M} \to \mathbb{R}^p$ are C^k smooth functions. We denote by $\mathrm{Crit}\Psi$ the set of critical points of Ψ, that is, the set of points (λ,x) for which the derivative $D\Psi(\lambda,x)$ is not surjective:

$$(\lambda,x) \in \mathrm{Crit}\Psi \qquad \Longleftrightarrow \qquad \mathrm{rank}\,(D\Psi(\lambda,x)) < p.$$

We also define the set $\widehat{\mathrm{Crit}}\Psi$ of *strongly critical points*, as follows:

$$(\lambda,x) \in \widehat{\mathrm{Crit}}\Psi \qquad \Longleftrightarrow \qquad \begin{cases} \phi^i(x) = \phi^0(x), & i \in \{0,\ldots,m\} \\ \mathrm{rank}\,\left(\sum_{i=0}^m \lambda_i D\phi^i(x)\right) < p. \end{cases} \tag{1.1}$$

In the above definitions by rank of a linear operator we mean the dimension of its image. It follows easily that $\widehat{\mathrm{Crit}}\Psi \subset \mathrm{Crit}\Psi$ and that equality holds if either $p=1$ or $m=0$. We are now ready to state our main result.

Theorem 3 ("Preparatory Sard theorem")**.** *Let $\mathcal{M}$ be a paracompact C^k manifold of dimension d with $k \geq 1$. Given $m+1$ functions $\phi^0, \phi^1, \ldots, \phi^m : \mathcal{M} \to \mathbb{R}^p$ of class C^k we set:*

$$\begin{cases} \Psi : \mathrm{inn}\,\Delta^m \times \mathcal{M} \to \mathbb{R}^p \\ \Psi(\lambda,x) := \sum_{i=0}^m \lambda_i \phi^i(x). \end{cases} \tag{1.2}$$

for all $\lambda = (\lambda^0, \ldots, \lambda^m) \in \mathrm{inn}\,\Delta^m$ and $x \in \mathcal{M}$. Then the following properties hold:

(I) *If $k \geq d-p+1$, then $\Psi(\widehat{\mathrm{Crit}}\Psi)$ is null in $\mathbb{R}^p$.*

(II) *If $k \geq \min\{d+1, m+d-p+1\}$, then $\Psi(\mathrm{Crit}\Psi)$ is null in $\mathbb{R}^p$.*

The above statement has two parts: the first one concerns the strongly critical values, a more restrictive notion, which turns out to be exactly what is required in the sequel, in order to establish the aforementioned nonsmooth results (*c.f.* Theorem 1 and Theorem 2). Once we will have established Theorem 3 (I), its second part is obtained almost for granted, since essentially the same pattern of proof applies. This second part, although not needed for our purposes, has an independent interest and is stated for completeness. The main instrument of the proof is an adaptation of the quantitative method of Yomdin [19] (see also [6]) aiming at exploiting the affine part of Ψ (namely, the variables $\lambda_0, \ldots, \lambda_m$) in order to obtain a better regularity than the one stemming from a blind application of the classical result. We recall that this latter bound would be $k \geq (m+d) - p + 1$.

As already mentioned, Theorem 2 (Sard for Lipschitz selections) is an extension of [1, Theorem 5] (Morse for Lipschitz selections), albeit not a straightforward one: the proof of [1, Theorem 5] relies upon geometrical arguments, leading nonsmooth criticality to a tractable smooth criticality of some smooth functions in naturally arising manifolds, and applying the classical Morse-Sard theorem to them. This approach is however much compromised by the real-valued case and does not seem to admit any obvious extension if $p > 1$. In contrast to that, the proof of Theorem 2, even specificated to the case $p = 1$, is considerably different in spirit. It passes through the highly technical–analytical approach of Theorem 3 ("Preparatory Sard theorem"), which is an improved version of Sard theorem. Still a common ground is tractable: both approaches eventually recover the classical Morse-Sard (respectively, Sard) theorem.

The manuscript is organized as follows: Notation and some basic definitions are recalled in Section 2, where several useful lemmas (required for the "Preparatory Sard theorem") are established. The proofs of the main results are given in Section 3, while in Section 4 we present applications of Theorem 2 to semi-infinite and vector optimization.

2 Preliminaries

2.1 Notation

In this work we shall consider the following abbreviations:

– We denote by |J| the cardinality of any finite set J.

– $\mathbb{N}_m = \{1, \ldots, m\}$, for any integer $m \geq 0$, and

– $\Delta^m = \left\{\boldsymbol{\lambda} = (\lambda_0, \lambda_1, \ldots, \lambda_m) \in \mathbb{R}^{m+1} : \lambda_i \geq 0 : \sum_{i=0}^m \lambda_i = 1\right\}$ for the m-dimensional simplex Δ^m in $\mathbb{R}^{m+1}$.

Further, we denote by $\operatorname{rank}(A)$ the rank of a matrix A and by $\operatorname{co}(C)$ the convex envelope of any subset C of $\mathbb{R}^d$. Then $x \in \operatorname{co}(C)$ if and only if there exists $m \geq 0$, $\boldsymbol{\lambda} \in \Delta^m$ and $\{x_0, x_1, \ldots, x_m\} \subset C$ such that $x = \sum_{i=0}^m \lambda_i x_i$. Let us recall that, thanks to the Caratheodory theorem, if $x \in \operatorname{co}(C)$ then we can always obtain a representation involving at most $d+1$ points. We also denote by $\operatorname{inn} \Delta^m$ the algebraic interior of Δ^m. In words, $\boldsymbol{\lambda} \in \operatorname{inn} \Delta^m$ if and only if $\boldsymbol{\lambda} \in \Delta^m$ and $\lambda_i \neq 0$ for all $i \in \{0, \ldots, m\}$. We finally denote by $\mathcal{L}(A)$ the Lebesgue measure of any measurable subset A of $\mathbb{R}^p$.

A function $f : \mathbb{R}^d \to \mathbb{R}^p$ is called Lipschitz continuous if there exists $K > 0$ such that for all $x, y \in \mathbb{R}^d$ we have $||f(x) - f(y)|| \leq K||x - y||$. If this property holds locally (with possibly different constants) at a neighborhood of every point of the domain of f, then f is called locally

Lipschitz. By the Rademacher theorem, every (locally) Lipschitz continuous function f is almost everywhere differentiable. In particular the set D_f of points where the derivative of f exists is a dense subset of $\mathbb{R}^d$.

The Clarke (generalized) Jacobian $J^C f(x)$ of f at a point $x \in \mathbb{R}^d$ (at which f might or might not be differentiable) is defined as the convex hull of the set $J^L f(x)$ made up of all accumulation points of sequences $\{Df(x_n)\}_n$ where $\{x_n\}_n \subset D_f$ and $\{x_n\} \to x$, see [5, Chapter 2.6]. Notice that fixing the bases in $\mathbb{R}^d$ and $\mathbb{R}^p$ the sets $J^L f(x)$, $J^C f(x)$ are naturally identified as subsets of $p \times d$ matrices. Should the set $J^C f(x)$ contain an element (matrix) of rank strictly less than p, the point x will be called *(Clarke) critical* for the Lipschitz function f. In case f is continuously differentiable, it holds $J^C f(x) = J^L f(x) = \{Df(x)\}$ and Clarke criticality collapses to the standard notion of criticality (that is, the derivative fails to be surjective).

We call $y \in \mathbb{R}^p$ a *(Clarke) critical value* for a locally Lipschitz function f, if $f^{-1}(y)$ contains a Clarke critical point. We denote by $\mathrm{Crit} f$ (subset of $\mathbb{R}^d$) the set of Clarke critical points, and by $f(\mathrm{Crit} f)$ (subset of $\mathbb{R}^p$) the set of Clarke critical values. Notice that in case $p = 1$ the Clarke Jacobian is identified to the Clarke subdifferential $\partial^C f(x)$ and criticality simply means $0 \in \partial^C f(x)$.

A locally Lipschitz function $f : \mathbb{R}^d \to \mathbb{R}$ is called *Clarke regular*, if for every $x \in \mathbb{R}^d$ and every direction $u \in \mathbb{R}^d \setminus \{0\}$ the directional derivative $f'(x,u)$ and the Clarke generalized derivative $f^\circ(x,u)$ are equal, see [5, Definition 2.3.4]. In words:

$$f'(x,u) := \lim_{t \searrow 0^+} \frac{f(x+tu) - f(x)}{t} = \limsup_{y \to x\, ;\, t \searrow 0^+} \frac{f(y+tu) - f(x)}{t} := f^\circ(x,u).$$

It follows that the Clarke subdifferential $\partial^C f(x)$ (*i.e.* the convex hull of the set of all accumulation points of sequences $\{Df(x_n)\}_n$ where $\{x_n\}_n \subset D_f$ and $\{x_n\} \to x$) is equal to the Fréchet subdifferential $\hat{\partial} f(x)$, defined as follows:

$$\hat{\partial} f(x) = \left\{ p \in \mathbb{R}^d : \liminf_{||u|| \to 0} \frac{f(x+u) - f(x) - \langle p, u \rangle}{||u||} \geq 0 \right\}.$$

We finally recall that the *modulus of (uniform) continuity* of a function $g : \mathbb{R}^d \to \mathbb{R}^p$ on a compact subset $\mathcal{T} \subset \mathbb{R}^d$ is defined as follows:

$$\begin{cases} \omega : [0, +\infty) \to [0, +\infty] \\ \omega(r) = \sup \{ ||g(x) - g(y)|| : \, x, y \in \mathcal{T}, \, ||x - y|| \leq r \}. \end{cases}$$

2.2 Preliminary results

(Semialgebraic sets) A set $A \subset \mathbb{R}^d$ is called *semialgebraic* if it can be obtained by means of a finite number of Boolean operations (union, intersection, complementary) of sets of the form $\{f = 0\}, \{g > 0\}$ where f, g are polynomials on $\mathbb{R}^d$. It is known that semialgebraic sets enjoy good stability properties (Tarski-Seidenberg principle). In particular, if $A \subset \mathbb{R}^d$ is a semialgebraic set and $F : \mathbb{R}^d \to \mathbb{R}^p$ is a polynomial mapping, then the set $F(A) \subset \mathbb{R}^p$ is semialgebraic (see [19, Proposition 2.7] e.g.). Moreover the following statement holds:

Fact (see [19, Proposition 2.6, Corollary 2.10]). There exists a number $N_* > 0$ depending on the dimensions d, p and the degrees of F and the polynomials involved in the construction of A such that:

(a) The number of connected components of $F(A)$ is bounded by N_*.

(b) For any ball $\mathcal{B}$ of radius $r > 0$ in $\mathbb{R}^p$, any two points belonging to the same connected component of $F(A) \cap \mathcal{B}$ can be joined by means of an absolutely continuous path (in fact, a semialgebraic curve) whose length is bounded by $N_* r$.

(Linear operators and modulus of surjectivity) Given a linear mapping $\mathbf{L} : \mathbb{R}^d \to \mathbb{R}^p$ we denote by $\rho(\mathbf{L})$ the smallest semi-axis of the ellipsoid

$$\mathcal{E}_{\mathbf{L}} = \{\mathbf{L}u \,|\, \|u\| \leq 1\} \subset \mathbb{R}^p. \tag{2.1}$$

Notice that $\rho(\mathbf{L}) > 0$ if and only if $\mathbf{L}$ is surjective. If $\rho(\mathbf{L})$ is getting close to zero, then we approach to nonsurjectivity. In particular, if the ellipsoid $\mathcal{E}_{\mathbf{L}}$ is contained in a hyperplane, then $\rho(\mathbf{L})$ is zero, and the linear mapping $\mathbf{L}$ is not surjective.

We shall need the following lemma. In the sequel, $\{e_1, \ldots, e_p\}$ denotes the canonical basis of $\mathbb{R}^p$. For $i \in \{1, \ldots, p\}$ and $\alpha \in (0, 1)$ we define closed convex cone

$$C_{i,\alpha} := \{y \in \mathbb{R}^p : |\langle y, e_i \rangle| \leq \alpha \|y\|\}\,.$$

Lemma 4 (inclusion of hyperplanes to cones)**.** *There exists $\bar{\alpha} = \bar{\alpha}(p) \in (1/\sqrt{2}, 1)$ such that for any linear hyperplane $\mathbf{H}$ of $\mathbb{R}^p$ there exists $i \in \{1, \ldots, p\}$ such that $\mathbf{H} \subset C_{i,\bar{\alpha}}$.*

Proof. The statement is trivial for $p = 1$. If the assertion were not true for some $p > 1$, then there would exist a sequence $\{\alpha_n\}_n \nearrow 1$ together with a sequence of linear hyperplanes $\{\mathbf{H}_n\}_n$ such that each $\mathbf{H}_n$ would intersect the set $\mathbb{R}^p \setminus C_{i,\alpha_n}$ for all $i \in \{1, \ldots, p\}$. By compactness of the Grasmannian manifold $\mathbb{G}(p-1, p)$, $\{\mathbf{H}_n\}_n$ converges (up to a subsequence) to some linear hyperplane $\mathbf{H}$ which must contain all of the vector of the canonical basis, a contradiction. The assertion follows. □

We further denote by $L_1, \ldots, L_p$ the coordinates of the linear mapping $\mathbf{L}$ and we consider, for $p > 1$, the linear mapping $\hat{\mathbf{L}}$ from $\mathbb{R}^d$ to $\mathbb{R}^{p-1}$ defined by

$$\hat{\mathbf{L}}u = (L_1 u, \ldots, L_{p-1} u) \qquad \forall u \in \mathbb{R}^d. \tag{2.2}$$

The following result controls, up to a permutation, the last coordinate of the image of $\mathbf{L}$ in terms of the other coordinates and $\rho(\mathbf{L})$.

Lemma 5 (Coordinate control)**.** *Let $\mathbf{L} : \mathbb{R}^d \to \mathbb{R}^p$ $(p > 1)$ be a linear operator. Under the above notation, up to a permutation of the coordinates, there exists $C = C(p) > 0$ such that*

$$|L_p u| \leq C \left[\rho(\mathbf{L})\|u\| + \left\|\hat{\mathbf{L}}u\right\|\right] \qquad \forall u \in \mathbb{R}^d. \tag{2.3}$$

Proof. Let us first consider the case where $\rho(\mathbf{L}) = 0$, that is $\mathbf{L}$ is not surjective. Let $\mathbf{H}$ denote a hyperplane of $\mathbb{R}^p$ containing the image of $\mathbf{L}$. Applying again Lemma 4 we deduce that for some $\bar{\alpha} \in (1/\sqrt{2}, 1)$ and $i \in \{1, \ldots, p\}$ we have $\mathbf{H} \subset C_{i,\bar{\alpha}}$. Permuting the coordinates if necessary, we may assume that $i = p$. It follows that for every $\mathbf{z} = (\hat{\mathbf{z}}, z_p) \in \mathbf{H}$ we have

$$|z_p| \leq \frac{\bar{\alpha}}{\sqrt{1 - \bar{\alpha}^2}} \|\hat{\mathbf{z}}\|\,. \tag{2.4}$$

Taking $\mathbf{z} = \mathbf{L}u$ for $u \in \mathbb{R}^d$ we obtain (2.3) for $\rho(\mathbf{L}) = 0$. Notice that the obtained constant in (2.4) depends only on p.

Let us now assume $\rho(\mathbf{L}) \neq 0$ and let $\bar{u} \in \mathbb{R}^d$ with $\|\bar{u}\| = 1$ such that $\|\mathbf{L}\bar{u}\| = \rho(\mathbf{L})$. The tangent plane to the ellipsoid $\mathcal{E}_{\mathbf{L}} \subset \mathbb{R}^p$ at $\mathbf{L}\bar{u}$ is the affine hyperplane $\mathbf{L}\bar{u} + \mathbf{H}$ with

$$\mathbf{H} = \{\mathbf{z} \in \mathbb{R}^p \,|\, \langle \mathbf{L}\bar{u}, \mathbf{z}\rangle = 0\}\,.$$

Applying Lemma 4 and permuting the coordinates if necessary, we deduce as before that $\mathbf{H} \subset C_{p,\bar{\alpha}}$ and (2.4) holds. By convexity, the ellipsoid $\mathcal{E}_L$ is contained in the strip

$$\{t\,\mathbf{L}\bar{u} + \mathbf{z} \,:\; t \in [-1,1],\, \mathbf{z} \in \mathbf{H}\}\,.$$

Using any point $\mathbf{y} = (\hat{\mathbf{y}}, y_p)$ in the above strip satisfies

$$|y_p| \,\leq\, C(\rho(\mathbf{L}) + \|\hat{\mathbf{y}}\|),$$

for $C = 2\bar{\alpha}\left(1-\bar{\alpha}^2\right)^{-1/2}$. Since for every $\|u\| = 1$ the point $\mathbf{L}(u) = (\hat{\mathbf{L}}(u), L_p(u))$ belongs to the ellipsoid, we get the result for every $u \in \mathbb{R}^d$ with $\|u\| = 1$ and by homogeneity for all $u \in \mathbb{R}^d$. $\square$

3 Proofs of the results

Roughly speaking, the "preparatory Sard theorem" (Theorem 2) states that the dimension of the affine manifold $\operatorname{inn}\Delta^m$ (corresponding to linear part of Ψ) does not (fully) appear in the required regularity of Ψ in order to conclude that the set $\Psi(\widehat{\mathrm{Crit}}\Psi)$ (respectively $\Psi(\mathrm{Crit}\Psi)$) is null. Indeed, it what follows we exploit this partial affine structure of Ψ, by adapting carefully the so-called Yomdin approach (see [19]). The proof is given in Section 3.1. This result will be subsequently used to establish Theorem 1 (in Section 3.2) and Theorem 2 (in Section 3.3).

3.1 Proof of "Preparatory Sard Theorem" (Theorem 3).

There is no loss of generality to assume $\mathcal{M} = \mathbb{R}^d$. We then denote by $\mathcal{T}$ the unit cube of $\mathbb{R}^d$. Since $\mathbb{R}^d$ is covered by a countable union of translations of $\mathcal{T}$ and since a countable union of null sets is null, it is sufficient to establish the result for the restriction of Ψ to $\operatorname{inn}\Delta^m \times \mathcal{T}$.

To this end, fix a positive integer $l > m$. The cube $\mathcal{T}$ can be divided into l^d cubes $\mathcal{T}_1, \ldots, \mathcal{T}_{l^d}$ of side $r := 1/l$. For each $i \in \{0, \ldots, m\}$ (counting functions ϕ^i) and $j \in \{1, \ldots, l^d\}$ (counting cubes), let $\mathbf{P}^i_j$ be the k-Taylor polynomial of ϕ^i at the center of the cube $\mathcal{T}_j$ and let ω denote a modulus of continuity for $D^k\phi$ on the unit cube $\mathcal{T}$, where $\phi := (\phi^0, \ldots, \phi^m)$. It follows that for any $x \in \mathcal{T}_j$ and $i \in \{0, \ldots, m\}$ we have

$$\left\|\phi^i(x) - \mathbf{P}^i_j(x)\right\| \leq \omega(r) r^k \tag{3.1}$$

and

$$\left\|D\phi^i(x) - D\mathbf{P}^i_j(x)\right\| \leq \omega(r) r^{k-1}. \tag{3.2}$$

Let further $K_j > 0$ be an upper bound for all derivatives $\left\{\left\|D\phi^i(x)\right\| : x \in \mathcal{T}_j, i \in \{1, \ldots, m\}\right\}$ (that is, a common Lipschitz constant of all functions ϕ^i on $\mathcal{T}_j$). In view of (3.2), taking possibly a larger value, we may assume that K_j is also a common upper bound for the derivatives $\left\{\left\|D\mathbf{P}^i_j(x)\right\| : x \in \mathcal{T}_j, i \in \{1,\ldots,m\}\right\}$. In the sequel we set

$$K = \max_{j\in\{1,\ldots,l^d\}} K_j\,. \tag{3.3}$$

For $j \in \{1,\ldots,l^d\}$ we also denote by M_j an upper bound for the diameter of $\Psi_j\,(\mathrm{inn}\,\Delta^m \times \mathcal{T}_j)$, and we define the function

$$\widetilde{\Psi_j}(\lambda, x) := \sum_{i=0}^{m} \lambda_i \mathbf{P}^i_j(x). \tag{3.4}$$

The above function is obtained by simply replacing the functions ϕ^i by the polynomials $\mathbf{P}^i_j$ in the definition of Ψ in (1.2). Therefore, in view of (3.1), taking possibly a larger value we may assume that M_j is also a common upper bound for the diameter of the set $\widetilde{\Psi}_j\,(\mathrm{inn}\,\Delta^m \times \mathcal{T}_j)$. We set

$$M = \max_{j\in\{1,\ldots,l^d\}} M_j\,. \tag{3.5}$$

Our first aim, roughly speaking, is to show that any (strongly) critical point of Ψ lying in $\mathcal{T}_j$ is an almost (strongly) critical for the function $\widetilde{\Psi_j}$. Recalling notation from Section 2.2 we have:

Lemma 6 (Approximating critical values)**.** *Let* $(\lambda, x) \in \mathrm{inn}\,\Delta^m \times \mathcal{T}_j$*. Then*

(i) It holds

$$\left\|\Psi(\lambda,x) - \widetilde{\Psi_j}(\lambda,x)\right\| \leq \omega(r) r^k. \tag{3.6}$$

In addition:

(ii) If $(\lambda, x) \in \widehat{\mathrm{Crit}\Psi}$ *(strongly critical), then*

$$\begin{cases} \left\|\mathbf{P}^i_j(x) - \mathbf{P}^0_j(x)\right\| \leq 2\omega(r) r^k \\ \quad \textit{for all } i \in \{1,\ldots,m\} \end{cases} \quad \textit{and} \quad \rho\left(\sum_{i=0}^{m} \lambda_i D\mathbf{P}^i_j(x)\right) \leq \omega(r) r^{k-1}.$$

(iii) If $(\lambda, x) \in \mathrm{Crit}\Psi$ *(critical), then*

$$\rho\left(D\widetilde{\Psi_j}(\lambda,x)\right) \leq 2\,\omega(r) r^{k-1},$$

Proof. Assertion (i) is straightforward from (3.1). The first inequality in (ii) follows immediately from (1.1) and (3.1). To prove the second inequality of (ii), we recall that since (λ, x) is strongly critical for Ψ, the linear mapping $\sum_{i=0}^{m} \lambda_i D\phi^i(x)$ is not surjective, hence its image $\mathrm{Im}\left(\sum_{i=0}^{m} \lambda_i D\phi^i(x)\right)$ is contained in a hyperplane $\mathbf{H} \subset \mathbb{R}^p$. Thus, since by (3.2) we have

$$\left\|\sum_{i=0}^{m} \lambda_i D\phi^i(x) - \sum_{i=0}^{m} \lambda_i D\mathbf{P}^i_j(x)\right\| \leq \omega(r) r^{k-1},$$

the image by $\sum_{i=0}^{m} \lambda_i D\mathbf{P}^i_j(x)$ of the unit ball of $\mathbb{R}^d$ is contained in the strip

$$\left\{y + s\,\hat{v}\,|\, y \in \mathbf{H}, s \in [-\omega(r) r^{k-1}, \omega(r) r^{k-1}]\right\},$$

where $\hat{v}$ is a unit vector orthogonal to $\mathbf{H}$. The second inequality in (ii) follows.

To prove (iii), we note that the formulae for $D\Psi$ and $D\widetilde{\Psi_j}$ at (λ, x) are respectively,

$$\begin{cases} D\Psi(\lambda,x)(\mu,\mathbf{u}) &= \sum_{i=0}^{m}\mu_i\phi^i(x)+\sum_{i=0}^{m}\lambda_i D\phi^i(x)(\mathbf{u}) \\ D\widetilde{\Psi_j}(\lambda,x)(\mu,\mathbf{u}) &= \sum_{i=0}^{m}\mu_i\mathbf{P}_j^i(x)+\sum_{i=0}^{m}\lambda_i D\mathbf{P}_j^i(x)(\mathbf{u}) \end{cases} \qquad \forall(\mu,\mathbf{u})\in \mathrm{T}_{\mathrm{inn}\,\Delta^m}(\lambda)\times\mathbb{R}^d,$$

where

$$\mathrm{T}_{\mathrm{inn}\,\Delta^m}(\lambda)\times\mathbb{R}^d \left(\simeq \mathbb{R}^m\times\mathbb{R}^d\right)$$

denotes the tangent space of $\mathrm{inn}\,\Delta^m\times\mathbb{R}^d$ at (λ, x). We deduce easily, in view of (3.1), (3.2) and the Cauchy-Schwarz inequality (recall $r = 1/l < 1/m$), that

$$\left\| D\Psi(\lambda,x) - D\widetilde{\Psi_j}(\lambda,x)\right\| \le (r(m+1)+1)\,\omega(r)r^{k-1} \le 2\,\omega(r)r^{k-1}.$$

If (λ, x) is critical, then $D\Psi(\lambda, x)$ is not surjective, yielding as before that the image by $D\widetilde{\Psi_j}(\lambda,x)$ of the unit ball of $\mathrm{T}_{\mathrm{inn}\,\Delta^m}(\lambda)\times\mathbb{R}^d$ is contained in a symmetric strip of width $4\,\omega(r)r^{k-1}$. We conclude easily. $\square$

Let ν, r be positive real numbers with $\nu << r$. Motivated by the above Lemma 6 (parts (ii), (iii)), for every $j \in \{1,\dots,l^d\}$ we define the semialgebraic sets

$$\widehat{\Gamma(\nu,r)}_j,\ \Gamma(\nu,r)_j \ \subset\ \mathrm{inn}\,\Delta^m\times\mathcal{T}_j$$

as follows:

$$(\lambda,x)\in\widehat{\Gamma(\nu,r)}_j \quad\Longleftrightarrow\quad \begin{cases} \left\|\mathbf{P}_j^i(x)-\mathbf{P}_j^0(x)\right\|\le \nu r, \\ \text{for all } i\in\{1,\dots,m\}\end{cases} \quad \&\quad \rho\left(\sum_{i=0}^{m}\lambda_i D\mathbf{P}_j^i(x)\right)\le\nu, \tag{3.7}$$

and respectively

$$(\lambda,x)\in\Gamma(\nu,r)_j \qquad\Longleftrightarrow\qquad \rho\left(D\widetilde{\Psi_j}(\lambda,x)\right)\le\nu, \tag{3.8}$$

Roughly speaking, $\widehat{\Gamma(\nu,r)}_j$ (respectively, $\Gamma(\nu,r)_j$) can be seen as the set of "ν-almost strongly critical" (respectively, "ν -almost critical") points of the function

$$\tilde{\Psi}_j : \mathrm{inn}\,\Delta^m\times\mathcal{T}_j\left(\approx[0,r]^d\right)\to\mathbb{R}^p,$$

given in (3.4). The following result provides an upper bound for the diameter of the image of these functions $\tilde{\Psi}_j$, $j\in\{1,\dots,l^d\}$, in terms of K (common Lipschitz constant of $\mathbf{P}_j^i$, see (3.3)) and r (size of $\mathcal{T}_j$).

Lemma 7 (Size of the image of $\tilde{\Psi}_j$)**.** *Assume that $\nu > 0$ is sufficiently small. Then*

(i) If $\widehat{\Gamma(\nu,r)}_j$ is nonempty, then $\mathrm{diam}\left(\widetilde{\Psi}_j\left(\mathrm{inn}\,\Delta^m\times\mathcal{T}_j\right)\right)\le 3Kr$.

(ii) If $\Gamma(\nu,r)_j$ is nonempty, then $\widetilde{\Psi}_j\left(\mathrm{inn}\,\Delta^m\times\mathcal{T}_j\right)$ is contained in a bounded strip determined by a hyperplane $\mathbf{H}$ and a vertical width $3Kr$.

Proof. Assume $\widehat{\Gamma(\nu,r)}_j \neq \emptyset$ and pick any $(\bar\lambda,\bar x) \in \widehat{\Gamma(\nu,r)}_j$. We shall show that $\widetilde{\Psi}_j\left(\operatorname{inn}\Delta^m \times \mathcal{T}_j\right)$ is contained in a ball of center $\mathbf{P}_j^0(\bar x) \in \mathbb{R}^p$ and radius $(K+\nu)r$. Indeed, in view of (3.7), for any $(\lambda, x) \in \operatorname{inn}\Delta^m \times \mathcal{T}_j$ we have

$$\left\|\widetilde{\Psi}_j(\lambda,x) - \mathbf{P}_j^0(\bar x)\right\| \leq \left\|\sum_{i=0}^m \lambda_i\left(\mathbf{P}_j^i(x) - \mathbf{P}_j^i(\bar x)\right)\right\| + \left\|\sum_{i=0}^m \lambda_i\left(\mathbf{P}_j^i(\bar x) - \mathbf{P}_j^0(\bar x)\right)\right\| \leq Kr + \nu r.$$

Let us now assume $\Gamma(\nu,r)_j \neq \emptyset$, and let $(\bar\lambda,\bar x) \in \Gamma(\nu,r)_j$. Then, roughly speaking, the vectors $\mathbf{P}_j^0(\bar x), \ldots, \mathbf{P}_j^m(\bar x)$ are close to a hyperplane $\mathbf{H}$ of $\mathbb{R}^p$, and consequently, the bounded set

$$\tilde\Psi_j(\operatorname{inn}\Delta^m \times \{\bar x\})$$

(convex hull of the $\mathbf{P}_j^i(\bar x)$) is contained in a narrow strip around this hyperplane. Pick now any $(\lambda, x) \in \operatorname{inn}\Delta^m \times \mathcal{T}_j$ and notice that the distance of $\tilde\Psi_j(\lambda,x)$ to the above set is majorized by

$$\left\|\widetilde{\Psi}_j(\lambda,x) - \widetilde{\Psi}_j(\lambda,\bar x)\right\| \leq \left\|\sum_{i=0}^m \lambda_i\left(\mathbf{P}_j^i(x) - \mathbf{P}_j^i(\bar x)\right)\right\| \leq Kr.$$

The result follows. □

The following lemma borrows heavily from the work of Yomdin [19]. It gives a quantitative estimation of the size of the sets of ν -almost strongly critical values (respectively, ν-almost critical values) by means of standard arguments of real algebraic geometry.

Lemma 8 (Semialgebraic estimates)**.** *There exists an integer $\tilde N$ depending only on d, p (dimensions of the spaces), k (degree of Taylor approximation for the functions ϕ^i), m ($m+1$ being the number of functions ϕ^i) and the constants K and M defined in (3.3), (3.5) (depending on the functions ϕ^i) such that for each $j \in \{1,\ldots,l^d\}$ it holds:*

(i) the (semialgebraic) set

$$\widehat{\Delta(\nu,r)}_j := \left\{\sum_{i=0}^m \lambda_i \mathbf{P}_j^i(x) \mid (\lambda,x) \in \widehat{\Gamma(\nu,r)}_j\right\} = \widetilde{\Psi}_j\left(\widehat{\Gamma(\nu,r)}_j\right)$$

is contained in the union of $\tilde N\left(\frac{1}{\nu}\right)^{p-1}$ cubes of $\mathbb{R}^p$ of side νr.

(ii) the (semialgebraic) set

$$\Delta(\nu,r)_j := \left\{\sum_{i=0}^m \lambda_i \mathbf{P}_j^i(x) \mid (\lambda,x) \in \Gamma(\nu,r)_j\right\} = \widetilde{\Psi}_j\left(\Gamma(\nu,r)_j\right)$$

is contained in the union of $\tilde N\left(\frac{1}{\nu}\right)^{p-1}$ cubes of $\mathbb{R}^p$ of radius ν.

Proof. Let $j \in \{1,\ldots,l^d\}$ (fixing the cube $\mathcal{T}_j$). We can clearly assume that the sets $\widehat{\Gamma(\nu,r)}_j$ and $\Gamma(\nu,r)_j$ are nonempty. (If one of these sets is empty, the corresponding conclusion holds true trivially.) We now proceed with the proofs of (i), (ii) to obtain the required integer bounds N_1, N_2 respectively. Then $\tilde N$ will be simply the maximum of N_1 and N_2.

In the remaining of the proof, we implicitly assume $p > 1$. The case $p = 1$ is an easy adaptation and is left to the reader. (Replacing formally $p = 1$ in what follows (and using obvious conventions) would lead to a disproportionally long proof with superfluous parts.)

Let us now denote by R any rectangle (lamella) in $\mathbb{R}^p$ of the form

$$R = \hat{R} \times I_p,$$

where $\hat{R}$ is a cube in $\mathbb{R}^{p-1}$ of side $\alpha > 0$ and I_p is an interval with length $|I_p|$ greater than the diameter of $\widetilde{\Psi}_j\,(\mathrm{inn}\,\Delta^m \times \mathcal{T}_j)$. (Typically $\alpha > 0$ is much smaller than r and $|I_p|$.)

(i). Since $\widehat{\Gamma(\nu, r)}_j$ is nonempty, by Lemma 7 (i) we deduce that $\widetilde{\Psi}_j\,(\mathrm{inn}\,\Delta^m \times \mathcal{T}_j)$ is contained in a rectangular ball of side $3Kr$. Therefore, it can be covered by $(3Kr/\alpha)^{p-1}$ lamellae R. In the sequel we shall indeed consider such a covering; from now on, the notation R will assign an arbitrary element of this covering.

In order to keep notation reasonably simple, we shall drop the index j from the polynomial functions $\mathbf{P}^i_j : \mathcal{T}_j \subset \mathbb{R}^d \to \mathbb{R}^p$ and the sets $\widehat{\Gamma(\nu, r)}_j$ and $\widehat{\Delta(\nu, r)}_j$ and will identify $\mathcal{T}_j$ with its translation $[0, r]^d$. Then each polynomial $\mathbf{P}^i_j$ will be simply denoted by $\mathbf{P}^i = (P^i_1, \ldots, P^i_p)$ (keeping in mind that it is the Taylor approximation of order k of the function ϕ^i on $\mathcal{T}_j$).

Applying Lemma 5 (coordinate control) to the linear operators

$$\mathbf{L}(x) := \sum_{i=0}^{m} \lambda_i D\mathbf{P}^i(x), \qquad x \in \mathcal{T}_j \tag{3.9}$$

we deduce that for some $\sigma \in \mathfrak{S}_p$ (permutation of the coordinates $\{1, \ldots, p\}$)

$$|\mathbf{L}(x)_p(u)| \,\le\, C\,\left(\rho\,(\mathbf{L}(x))\,\|u\| + \left\|\widehat{L(x)}(u)\right\|\right) \qquad \forall u \in \mathbb{R}^d, \tag{3.10}$$

where according to the notation introduced in (2.2)

$$\widehat{L(x)} = \sum_{i=0}^{m} \lambda_i \widehat{D\mathbf{P}^i(x)}, \qquad \text{with} \qquad \widehat{D\mathbf{P}^i(x)} = (DP^i_1(x), \ldots, DP^i_{p-1}(x)).$$

This yields a natural partition of the semialgebraic set $\widehat{\Gamma(\nu, r)}$ into a finite number of sets $\{\widehat{\Gamma_\sigma(\nu, r)}\}_{\sigma \in \mathfrak{S}_p}$. Namely, $(\lambda\;, x) \in \widehat{\Gamma_\sigma(\nu, r)}$ provided that after applying the permutation $\sigma \in \mathfrak{S}_p$ to the coordinates of $\mathbf{L}(x)$ we get (3.10). It follows readily that $\widehat{\Gamma_\sigma(\nu, r)}$ is a semialgebraic subset of $\mathrm{inn}\,\Delta^m \times [0, r]^d$ for every $\sigma \in \mathfrak{S}_p$.

Let us observe that for any permutation $\sigma \in \mathfrak{S}_p$ and any lamella $R = \hat{R} \times I_p$ of the covering of $\widetilde{\Psi}_j\,(\mathrm{inn}\,\Delta^m \times \mathcal{T}_j)$ the set

$$\widehat{\Gamma^R_\sigma(\nu, r)} := \left\{(\lambda, x) \in \widehat{\Gamma_\sigma(\nu, r)} \mid \sum_{i=0}^{m} \lambda_i \mathbf{P}^i(x) \in R\right\}$$

is semialgebraic. Notice further that since

$$\widetilde{\Psi}_j\left(\widehat{\Gamma^R_\sigma(\nu, r)}\right) = \widetilde{\Psi}_j\left(\widehat{\Gamma_\sigma(\nu, r)}\right) \cap R,$$

the projection of the set $\widetilde{\Psi}_j\left(\widehat{\Gamma_\sigma^R(\nu,r)}\right)$ to the first $p-1$ coordinates is contained in a cube of side $\alpha>0$. Let us now apply the "semialgebraic fact" given in Section 2.1 for each of these sets $A=\widehat{\Gamma_\sigma^R(\nu,r)}$ and for the polynomial mapping $F=\widetilde{\Psi}_j$. We deduce that there exists an integer N_* depending only on d, p (dimensions of the spaces) and k (degree of the polynomials $\mathbf{P}^i$) such that (for any $\sigma\in\mathfrak{S}_p$ and any lamella R)

- the set $\widehat{\Gamma_\sigma^R(\nu,r)}\subset\operatorname{inn}\Delta^m\times[0,r]^d$ has at most N_* connected components ;
  (in the sequel $\widehat{\Gamma_\sigma^R(\nu,r)}_{\text{conn}}$ will denote an arbitrary connected component of $\widehat{\Gamma_\sigma^R(\nu,r)}$)
- any $(\lambda^x,x),(\lambda^y,y)\in\widehat{\Gamma_\sigma^R(\nu,r)}_{\text{conn}}$ can be joined by means of an absolutely continuous semialgebraic path $t\longmapsto(\lambda(t),\gamma(t))$, lying entirely in $\widehat{\Gamma_\sigma^R(\nu,r)}_{\text{conn}}$ such that

$$(\lambda(0),\gamma(0))=(\lambda^x,x),\quad(\lambda(1),\gamma(1))=(\lambda^y,y),$$

$$\int_0^1\|\dot\lambda(t)\|dt\le N_*\quad\text{and}\quad\int_0^1\|\dot\gamma(t)\|dt\le N_*r. \tag{3.11}$$

$$\text{and}\quad\int_0^1\left\|\frac{d}{dt}\left(\sum_{i=0}^m\lambda_i(t)\mathbf{P}^i\widehat{(\gamma(t))}\right)\right\|dt\le N_*\alpha. \tag{3.12}$$

(The first inequality is obtained by applying the aforementioned fact to the set of tuples (λ,x) in $[0,r]^{m+1}\times[0,r]^d$ such that $\lambda/r\in\operatorname{inn}\Delta^m$ and by dividing by r.)

Our strategy is now the following: our aim is to obtain an upper bound $L_*>0$ (independent of the cube $\mathcal{T}_j$, the position of the lamelle R in $\widetilde{\Psi}_j(\operatorname{inn}\Delta^m\times\mathcal{T}_j)$ and the permulation $\sigma\in\mathfrak{S}_p$) for the length of the maximum interval made up by elements of the form

$$\left\{\widetilde{\Psi}(\lambda,x)_p:=\sum_{i=0}^m\lambda_iP_p^i(x)\mid(\lambda,x)\in\widehat{\Gamma_\sigma^R(\nu,r)}_{\text{conn}}\right\}\subset I_p. \tag{3.13}$$

Since (3.13) is the projection onto the last coordinate of the set

$$\widetilde{\Psi}\left(\widehat{\Gamma_\sigma^R(\nu,r)}_{\text{conn}}\right)\subset\mathbb{R}^p,$$

we can then deduce that this latter can be covered by L_*/α cubes of size α. Multiplying then this number by N_* (upper bound for the number of connected components of $\widehat{\Gamma_\sigma^R(\nu,r)}$) and by $p!$ (number of permutations) we obtain an upper bound

$$\frac{p!N_*L_*}{\alpha}$$

for the number of cubes of size $\alpha>0$ required to cover the set

$$\widehat{\Delta(\nu,r)}\cap R=\widetilde{\Psi}\left(\widehat{\Gamma(\nu,r)}\right)\cap R\subset\mathbb{R}^p,$$

that is, the set of ν-almost strongly critical values of $\widetilde{\Psi}$ (restricted to $\operatorname{inn}\Delta^m\times\mathcal{T}_j$) in the lamella R. Multiplying the above bound by $(3Kr/\alpha)^{p-1}$ (number of lamellae R of the covering of $\widetilde{\Psi}_j(\operatorname{inn}\Delta^m\times\mathcal{T}_j)$) we deduce that the set $\widehat{\Delta(\nu,r)}$ can be covered by a maximum of

$$(3K)^{p-1}p!N_*\frac{L_*}{\alpha}\left(\frac{r}{\alpha}\right)^{p-1} \tag{3.14}$$

cubes of size $\alpha > 0$.

Let us now proceed to obtain the required upper bound $L_* > 0$. To this end, we fix a permutation $\sigma \in \mathfrak{S}_p$, a lamella $R = \hat{R} \times I_p$ and a connected component $\widehat{\Gamma^R_\sigma(\nu, r)}_{\text{conn}}$ and we consider two arbitrary elements $(\lambda^x, x), (\lambda^y, y)$ in $\widehat{\Gamma^R_\sigma(\nu, r)}_{\text{conn}}$. We readily get

$$\begin{aligned}\sum_{i=0}^m \lambda_i^y P_p^i(y) - \sum_{i=0}^m \lambda_i^x P_p^i(x) &= \int_0^1 \frac{d}{dt}\left(\sum_{i=0}^m \lambda_i(t) P_p^i(\gamma(t))\right) dt \\ &= \int_0^1 \sum_{i=0}^m \dot{\lambda}_i(t) P_p^i(\gamma(t)) dt \; + \; \int_0^1 \sum_{i=0}^m \lambda_i(t) D P_p^i(\gamma(t))\left(\dot{\gamma}(t)\right)\, dt\,.\end{aligned} \tag{3.15}$$

We shall now obtain appropriate bounds of small order for the above terms.

Step 1. Bound for the first integral of (3.15) of order νr.

Since $\lambda(t) \in \operatorname{inn} \Delta^m$, it follows that $\sum_{i=0}^m \dot{\lambda}_i(t) = 0$, whence

$$\sum_{i=0}^m \dot{\lambda}_i(t) \mathbf{P}^i(\gamma(t)) = \sum_{i=1}^m \dot{\lambda}_i(t)\left(\mathbf{P}^i(\gamma(t)) - \mathbf{P}^0(\gamma(t))\right).$$

By (3.7) and the Cauchy-Schwarz inequality, we obtain

$$\left\| \sum_{i=0}^m \dot{\lambda}_i(t) \mathbf{P}^i(\gamma(t)) \right\| \leq \sqrt{m}\, \nu r\, ||\dot{\lambda}(t)||,$$

which thanks to (3.11) yields:

$$\int_0^1 \left| \sum_{i=0}^m \dot{\lambda}_i(t) P_p^i(\gamma(t)) \right| dt \leq \int_0^1 \left\| \sum_{i=0}^m \dot{\lambda}_i(t) \mathbf{P}^i(\gamma(t)) \right\| dt \leq \sqrt{m}\, N_* \nu r. \tag{3.16}$$

Step 2. Bound for the second integral of (3.15) of order $\max\{\alpha, \nu r\}$.

According to the notation of (3.9) we have

$$\mathbf{L}(\gamma(t)) = \sum_{i=0}^m \lambda_i(t) D\mathbf{P}^i(\gamma(t)).$$

In view of Lemma 2.3, (3.7), (3.11) and (3.12) we deduce that for some $C = C(p)$

$$\begin{aligned}\left| \int_0^1 \sum_{i=0}^m \lambda_i(t) D P_p^i(\gamma(t))\left(\dot{\gamma}(t)\right) dt \right| &\leq C \int_0^1 \rho(\mathbf{L}(\gamma(t)))\, \|\dot{\gamma}(t)\|\, dt + C \int_0^1 \left\| \widehat{\mathbf{L}(\gamma(t))}\left(\dot{\gamma}(t)\right) \right\| dt \\ &\leq C \nu N_* r \; + \; C \int_0^1 \left\| \frac{d}{dt}\left(\widehat{\mathbf{L}(\gamma(t))}\right) \right\| dt \\ &\qquad + C \int_0^1 \left\| \sum_{i=0}^m \dot{\lambda}_i(t) \widehat{\mathbf{P}^i(\gamma(t))} \right\| dt \\ &\leq C N_* \nu r + C N_* \alpha + C\sqrt{m}\, N_* \nu r.\end{aligned} \tag{3.17}$$

Combining the bounds (3.16) and (3.17) obtained in the above steps, we obtain

$$\left|\sum_{i=0}^{m} \lambda_i^y P_p^i(y) - \sum_{i=0}^{m} \lambda_i^x P_p^i(x)\right| \le N_* \left(\sqrt{m}(1+C)+C\right)\nu r + N_* C\alpha.$$

Setting

$$D_* = N_* \left(\sqrt{m}(1+C)+C\right) \qquad \text{and} \qquad \alpha = \nu r$$

we deduce that

$$\left|\sum_{i=0}^{m} \lambda_i^y P_p^i(y) - \sum_{i=0}^{m} \lambda_i^x P_p^i(x)\right| \le 2D_*\nu r := L_*$$

where $D_* = D_*(d,p,k,m)$. Replacing the values $L_* = 2D_*\nu r$ and $\alpha = \nu r$ to (3.14) we deduce that the set $\widehat{\Delta(\nu,r)}_j$ can be covered by at most

$$(3K)^{p-1} p! N_*(2D_*) \left(\frac{1}{\nu}\right)^{p-1}$$

rectangles of side νr. Therefore, the first assertion follows by taking

$$N_1 = (3K)^{p-1} p! N_*(2D_*).$$

(ii). The proof of this part follows the patterns of the proof of part (i) above. Given $j \in \{1,\ldots,l^d\}$ (fixing the semialgebraic set $\Gamma(\nu,r)_j \subset \operatorname{inn}\Delta^m \times \mathcal{T}_j$) we apply Lemma 5 (coordinate control) to the linear operators

$$D\widetilde{\Psi_j}(\lambda,x) : \mathrm{T}_{\operatorname{inn}\Delta^m}(\lambda) \times \mathbb{R}^d \to \mathbb{R}^p \qquad (\lambda,x) \in \Gamma(\nu,r)_j$$

to deduce that under an appropriate permutation $\sigma \in \mathfrak{S}_p$ of the coordinates $\{1,\ldots,p\}$ relation (3.10) holds, that is,

$$\left|D\widetilde{\Psi_j}(\lambda,x)_p(u)\right| \le C\left(\rho\left(D\widetilde{\Psi_j}(\lambda,x)\right)\|u\| + \left\|\widehat{D\widetilde{\Psi_j}(\lambda,x)}(u)\right\|\right) \qquad \forall u \in \mathrm{T}_{\operatorname{inn}\Delta^m}(\lambda) \times \mathbb{R}^d. \tag{3.18}$$

We recall again the notation of (2.2)

$$\widehat{D\widetilde{\Psi_j}(\lambda,x)} = \left(D\widetilde{\Psi_j}(\lambda,x)_1,\ldots,D\widetilde{\Psi_j}(\lambda,x)_{p-1}\right).$$

For each partition $\sigma \in \mathfrak{S}_p$ we define $\Gamma_\sigma(\nu,r)_j$ to be the set of those elements $(\lambda,x) \in \Gamma(\nu,r)_j$ for which after applying the permutation σ to the coordinates of $D\widetilde{\Psi_j}(\lambda,x)$, relation (3.18) holds true. Since $\Gamma(\nu,r)_j$ is nonempty, by Lemma 7 (ii) we deduce that the bounded set $\widetilde{\Psi}_j(\operatorname{inn}\Delta^m \times \mathcal{T}_j)$ is contained in a strip of width less or equal to $3Kr$. Recalling from (3.5) that $M > 0$ is an upper bound for the diameter of $\widetilde{\Psi}_j(\operatorname{inn}\Delta^m \times \mathcal{T}_j)$ we therefore conclude that $\widetilde{\Psi}_j(\operatorname{inn}\Delta^m \times \mathcal{T}_j)$ can be covered by at most

$$\left(\frac{M}{\alpha}\right)^{p-1}$$

lamellae of the form $R = \hat{R} \times I_p$, where $\hat{R}$ is a cube in $\mathbb{R}^{p-1}$ of side $\alpha > 0$ and $|I_p| \geq M$. Let us consider such a covering and let R be an arbitrary element (lamella) of it.

Fixing now the partition $\sigma \in \mathfrak{S}_p$ and the lamella R we set

$$\Gamma_\sigma^R(\nu, r) := \left\{(\lambda, x) \in \Gamma_\sigma(\nu, r) \,|\, \widetilde{\Psi}_j(\lambda, x) \in R\right\}.$$

The above set being semialgebraic, we can apply again the "semialgebraic fact" (Section 2.1) for the polynomial mapping $F = \widetilde{\Psi}_j$. Notice again that the projection of the set $\widetilde{\Psi}_j\left(\Gamma_\sigma^R(\nu, r)\right)$ to the first $p-1$ coordinates is contained in a cube of side $\alpha > 0$. Let N_* be an integer (depending on d, p and k) such that the set $\Gamma_\sigma^R(\nu, r)$ has at most N_* connected components (each of which will be denoted as $\Gamma_\sigma^R(\nu, r)_{\mathrm{conn}}$) and that for any $(\lambda^x, x), (\lambda^y, y) \in \Gamma_\sigma^R(\nu, r)_{\mathrm{conn}}$ there exists a semialgebraic path $\mathbf{u}(t) = (\lambda(t), \gamma(t))$, $t \in [0, 1]$, lying entirely in $\Gamma_\sigma^R(\nu, r)_{\mathrm{conn}}$ with

$$\mathbf{u}(0) = (\lambda^x, x), \quad \mathbf{u}(1) = (\lambda^y, y),$$

$$\int_0^1 \|\dot{\mathbf{u}}(t)\| dt \leq N_* \tag{3.19}$$

$$\text{and} \qquad \int_0^1 \left\| \frac{d}{dt}\left(\widehat{D\widetilde{\Psi}_j(\mathbf{u(t)})}\right)\right\| dt \leq N_*\alpha. \tag{3.20}$$

Following the same strategy as in part (i), we seek for an upper bound $L'_* > 0$ for the length of the maximum interval contained in the set

$$\widetilde{\Psi}_j\left(\Gamma_\sigma^R(\nu, r)_{\mathrm{conn}}\right)_p \subset I_p.$$

Once we get this bound at hand, we can deduce, as before, that the set $\Delta(\nu, r)_j$ can be covered by a maximum of

$$p! N_* \frac{L'_*}{\alpha}\left(\frac{M}{\alpha}\right)^{p-1} \tag{3.21}$$

cubes of size $\alpha > 0$. To determine $L'_* > 0$ we pick $(\lambda^x, x), (\lambda^y, y)$ in $\Gamma_\sigma^R(\nu, r)_{\mathrm{conn}}$ and we obtain in view of (3.18) and up to a permutation of coordinates:

$$\begin{aligned}
&\left|\sum_{i=0}^m \lambda_i^y P_p^i(y) - \sum_{i=0}^m \lambda_i^x P_p^i(x)\right| = \left|\int_0^1 \frac{d}{dt}\left(D\widetilde{\Psi}_j(\mathbf{u(t)})_p\right) dt\right| \\
&\leq \int_0^1 C\rho\left(D\widetilde{\Psi}_j(\mathbf{u(t)})\right)\|\dot{\mathbf{u}}(t)\|\, dt \quad + \quad \int_0^1 \left\|\widehat{D\widetilde{\Psi}_j(\mathbf{u(t)})}(\dot{\mathbf{u}}(t))\right\| dt \\
&\leq CN_*\nu + CN_*\alpha = CN_*(\nu + \alpha) := L'_*.
\end{aligned} \tag{3.22}$$

Taking $\alpha = \nu$, replacing $L'_* = 2CN_*\nu$ in (3.21) and setting

$$N_2 = 2CN_*^2 p! M^{p-1}$$

we deduce that the set $\Delta(\nu, r)_j$ can be covered by at most

$$N_2\left(\frac{1}{\nu}\right)^{p-1}$$

cubes of side $\alpha = \nu$ as asserted. This concludes the proof of (ii) and in turn of Lemma 8. $\square$

Return now to the proof of Theorem 3. To prove **(I)**, we note that in view of Lemma 6 (ii), we may apply Lemma 8 (i) with

$$\nu = 2\omega(r)r^{k-1} \tag{3.23}$$

to deduce that for every $j \in \{1, \dots, l^d\}$ the set

$$E_j := \left\{ \sum_{i=0}^{m} \lambda_i \mathbf{P}_j^i(x) \,|\, x \in \mathcal{T}_j,\ (\lambda, x) \in \widehat{\mathrm{Crit}\Psi} \right\}$$

is contained in a union of $\tilde{N}(1/\nu)^{p-1}$ cubes of side νr, with $\tilde{N}$ depending only on k, d, p, m and K. This together with (3.6) yields that the set

$$\Psi\left(\widehat{\mathrm{Crit}\Psi}\right)$$

of strongly critical values of the function Ψ defined in (1.2) is contained in the union of $\tilde{N}\, l^d\, (1/\nu)^{p-1}$ cubes of side $3\nu r$. Therefore, its measure is bounded by

$$\tilde{N}\, l^d\, \frac{1}{\nu^{p-1}}\, (3\nu r)^p = 3^p\, \tilde{N}\, l^d\, r^p\, \nu$$

Using (3.23) and replacing $r = 1/l$ in the above, we get the following bound for the Lebesgue measure of $\Psi\left(\widehat{\mathrm{Crit}\Psi}\right)$

$$\mathcal{L}\left(\Psi\left(\widehat{\mathrm{Crit}\Psi}\right)\right) \le 2 \cdot 3^p\, \tilde{N}\, \omega(\frac{1}{l})\, l^{(d-p+1)-k}.$$

It follows that for $k \ge d-p+1$ taking the limit as l tends to infinity, we obtain the result.

To prove **(II)**, we note that in view of Lemma 6 (ii), we may apply Lemma 8 (iii) with

$$\nu = 2\,\omega(r)r^{k-1} \tag{3.24}$$

to deduce that for every $j \in \{1, \dots, l^d\}$ the set

$$E_j := \left\{ \sum_{i=0}^{m} \lambda_i P_j^i(x) \,|\, x \in \mathcal{T}_j,\ (\lambda, x) \in \mathrm{Crit}\Psi \right\}$$

is contained in a union of $\tilde{N}(1/\nu^{p-1})$ cubes of side ν, with $\tilde{N}$ depending only on k, d, p, m and the constants K and M. This together with (3.6) yields that the set $\Psi\left(\mathrm{Crit}\Psi\right)$ of critical values of Ψ is contained in the union of $\tilde{N}\, l^d\, (1/\nu^{p-1})$ cubes of side $\nu + 2\nu r \le 3\nu$. Therefore, its measure is bounded by

$$l^d\, \tilde{N}\, \frac{1}{\nu^{p-1}}(3\nu)^p = 3^p\, \tilde{N}\, l^d\, \nu.$$

Using (3.23) and replacing $r = 1/l$ in the above, we get the following bound for the Lebesgue measure of $\Psi\left(\mathrm{Crit}\Psi\right)$

$$\mathcal{L}\left(\Psi\left(\mathrm{Crit}\Psi\right)\right) \le 3^p \tilde{N}\, 2\, \omega(\frac{1}{l})\, l^{d+1-k}.$$

Therefore, if $k \ge d+1$, then taking the limit as l tends to infinity, we obtain the result. Notice that this regularity bound is interesting only when the dimension of the simplex inn Δ^m is greater or equal to the dimension of the arrival space. If $p > m$ then the classical Sard theorem will provide a better result. $\square$

3.2 Proof of Morse-Sard for min-type functions (Theorem 1).

We denote by $\mathrm{Crit}\, f$ the set of Clarke critical points of the function

$$f(z) := \min_{q\in\mathcal{N}} \{\phi(z,q)\}$$

and we recall that in this case the Clarke subdifferential is given by the formula

$$\partial f(\bar z) = \mathrm{co}\,\{D_z\phi(\bar z,q)\,|\,q\in \arg\min \phi(\bar z,\cdot)\}\,.$$

Therefore, thanks to Caratheodory's lemma, if $\bar z \in \mathrm{Crit}\, f$, there exist $m\in\{0,\dots,n\}$, $\bar\lambda = (\bar\lambda_0,\dots,\bar\lambda_m)\in \mathrm{inn}\,\Delta^m$ and $\bar{\mathbf{q}} := (\bar q^0,\dots,\bar q^m)\in\mathcal{N}^{m+1}$ such that

$$\sum_{i=0}^{m} \bar\lambda_i D_z\phi(\bar z,\bar q^i) = 0.$$

Notice that $\phi(\bar z,\bar q^i) = f(\bar z)$ and $D_q\phi(\bar z,\bar q^i) = 0$, for all $i\in\{0,\dots,m\}$. In particular we deduce

$$\sum_{i=0}^{m} \bar\lambda_i D\phi(\bar z,\bar q^i) = 0.$$

Let us now consider the function

$$\begin{cases} \Psi_m : \mathrm{inn}\,\Delta^m \times \left(\mathbb{R}^n\times\mathcal{N}^{m+1}\right) \to \mathbb{R} \\ \Psi_m\left(\lambda,z,\mathbf{q}\right) := \sum_{i=0}^m \lambda_i\phi(z,q^i), \end{cases}$$

where $\mathbf{q} := (q^0,\dots,q^m)$. Setting $x := (z,\mathbf{q})\in\mathcal{M} := \mathbb{R}^n\times\mathcal{N}^{m+1}$ and defining

$$\phi^i(x) := \phi(z,\pi_i(\mathbf{q})),\quad i\in\{0,\dots,m\}$$

(π_i denotes the i-projection of $\mathbf{q}$) we deduce easily that Ψ_m is of the form (1.2) with $p=1$ and the tuple

$$(\bar\lambda,\bar x) := (\bar\lambda,\bar z,\bar q^0,\dots,\bar q^m)$$

is a (strongly) critical point of Ψ_m. Moreover,

$$f(\bar z) = \Psi_m(\bar\lambda,\bar z,\bar{\mathbf{q}}).$$

Thus the set $f(\mathrm{Crit}\, f)$ of Clarke critical values of f is contained to the finite union (from $m=0$ to n) of the (strongly) critical values of the functions Ψ_m. Since $m\le n$ we have $k\ge \dim\mathcal{M} = n+\ell(m+1)$ and the result follows from Theorem 3 (I) for $p=1$. □

3.3 Proof of Sard for Lipschitz selections (Theorem 2)

Let $F=(F_1,\dots,F_p):\mathbb{R}^d\times T\to\mathbb{R}^p$ satisfy the assumptions of Theorem 2 and let $f=(f_1,\dots,f_p)$ be a continuous selection of F, that is, f is continuous and

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

Then the assertion that f is locally Lipschitz is a straightforward consequence of [1, Proposition 2], since for each $i \in \{1, \ldots, p\}$ the function $f_i : \mathbb{R}^d \to \mathbb{R}$ is a continuous selection of the family $\{F_i(\cdot, t) : t \in T\}$ with T countable compact, whence locally Lipschitz continuous.

The second part of the theorem asserts that $f(\operatorname{Crit} f)$ is null in $\mathbb{R}^p$. In order to establish this part we shall need the following notation: We set $T_i(x) = \{t \in T : f_i(x) = F_i(x,t)\}$, for $i \in \{1, \ldots, p\}$ and

$$T(x) = \{t \in T : f(x) = F(x,t)\} = \bigcap_{i \in \{1,\ldots,p\}} T_i(x).$$

We further set

$$A_{f_i}(x) = \operatorname{co}\{D_x F_i(x,t) : t \in T_i(x)\}, \qquad \text{for all } i \in \{1, \ldots, p\} \tag{3.25}$$

and

$$A_f(x) = A_{f_1}(x) \otimes \ldots \otimes A_{f_p}(x). \tag{3.26}$$

The following claim is important for our purposes:

Claim 1 (broadening the notion of criticality): It holds $J^C f(x) \subset A_f(x)$ for all $x \in \mathbb{R}^d$.

Proof of Claim 1. Let $\mathcal{D}_f$ (respectively, $\mathcal{D}_{f_i}$) denote the set of points of differentiability of f (respectively, of f_i). By definition of the Clarke Jacobian of f (respectively, Clarke subdifferential of f_i) we have

$$J^C f(x) = \operatorname{co}\left\{\lim_{x_n \to x} Df(x_n) : \{x_n\}_n \subset \mathcal{D}_f\right\}$$

$$\left(\text{respectively,} \quad \partial^C f_i(x) = \operatorname{co}\left\{\lim_{x_n \to x} Df_i(x_n) : \{x_n\}_n \subset \mathcal{D}_{f_i}\right\}\right).$$

Since $\mathcal{D}_f \subset \mathcal{D}_{f_i}$ for all $i \in \{1, \ldots, p\}$, it follows readily from the above definitions that

$$J^C f(x) \subset \partial^C f_1(x) \otimes \ldots \otimes \partial^C f_p(x).$$

By [1, Proposition 4] we deduce $\partial^C f_i(x) \subset A_{f_i}(x)$ for all $i \in \{1, \ldots, p\}$, and the claim follows.♦

To prove the result it is sufficient to establish that the set $f(\widetilde{\operatorname{Crit}} f)$ is null in $\mathbb{R}^p$, where

$$\widetilde{\operatorname{Crit}} f := \{x \in \mathbb{R}^d : \exists A \in A_f(x), \operatorname{rank}(A) < p\}. \tag{3.27}$$

(Indeed, in view of Claim 1, we deduce readily that $\operatorname{Crit} f \subset \widetilde{\operatorname{Crit}} f$, whence $f(\operatorname{Crit} f) \subset f(\widetilde{\operatorname{Crit}} f)$.)

Let us now consider the following (countable) set:

$$\mathcal{F} := \{J \subset T : |J| \le d+1\}. \tag{3.28}$$

For $\mathcal{J} = (J_1, \ldots, J_p) \in \mathcal{F}^p$ we set $m_i = |J_i| - 1$ and we define the function

$$\begin{cases} G_{\mathcal{J}} : \operatorname{inn}\Delta^{m_1} \times \ldots \times \operatorname{inn}\Delta^{m_p} \times \mathbb{R}^d \longrightarrow \mathbb{R}^p \\ \\ G_{\mathcal{J}}(\boldsymbol{\lambda}^1, \ldots, \boldsymbol{\lambda}^p, x) = \left(\sum_{j_1=0}^{m_1} \boldsymbol{\lambda}^1_{j_1} F_1(x, t^1_{j_1}), \ldots, \sum_{j_p=0}^{m_p} \boldsymbol{\lambda}^p_{j_p} F_p(x, t^p_{j_p})\right) \end{cases} \tag{3.29}$$

where for $i \in \{1, \ldots, p\}$ we have

$$\boldsymbol{\lambda}^i = (\lambda_0^i, \ldots, \lambda_{m_i}^i) \in \operatorname{inn} \Delta^{m_i} \quad \text{and} \quad J_i = \left\{t_0^i, \ldots, t_{m_i}^i\right\}.$$

Notice that $G_{\mathcal{J}}$ is of class C^k with $k \geq d - p + 1$ (inheriting the regularity of the functions $x \mapsto F(x,t)$, $t \in T$).

Let further $\widehat{\operatorname{Crit}}\, G_{\mathcal{J}}$ denote the set of "strongly critical points" of the function $G_{\mathcal{J}}$, that is,

$$(\boldsymbol{\lambda}^1, \ldots, \boldsymbol{\lambda}^p, x) \in \widehat{\operatorname{Crit}}\, G_{\mathcal{J}} \iff \begin{cases} F_i(x, t_{j_i}^i) = F_i(x, t_0^i), & \begin{cases} \text{for all } i \in \{1, \ldots, p\} \\ \text{for all } j_i \in \{0, \ldots, m_i\} \end{cases} \\[2ex] \operatorname{rank}\left(\sum_{j_1=0}^{m_1} \boldsymbol{\lambda}_{j_1}^1 D_x F_1(x, t_{j_1}^1), \ldots, \sum_{j_p=0}^{m_p} \boldsymbol{\lambda}_{j_p}^p D_x F_p(x, t_{j_p}^p)\right) < p. \end{cases}$$

The following claim is crucial for our considerations.

Claim 2 (transferring criticality from f to some $G_{\mathcal{J}}$): For every $x \in \operatorname{Crit} f$, there exist $\mathcal{J} = (J_1, \ldots, J_p) \in \mathcal{F}^p$ (depending on x) and $\boldsymbol{\lambda}^i \in \operatorname{inn} \Delta^{|J_i|-1}$, $i \in \{1, \ldots, p\}$ (depending on x) such that:

– the point $(\boldsymbol{\lambda}^1, \ldots, \boldsymbol{\lambda}^p, x)$ is strongly critical for the function $G_{\mathcal{J}}$;

– $f(x) = G_{\mathcal{J}}(\boldsymbol{\lambda}^1, \ldots, \boldsymbol{\lambda}^p, x)$.

Proof of Claim 2. Let $x \in \operatorname{Crit} f$. It follows from Claim 1 that $x \in \widetilde{\operatorname{Crit}} f$, that is, there exists $\mathbf{A} \in A_f(x)$ (identified with a $p \times d$ matrix) with $\operatorname{rank}(\mathbf{A}) < p$. Let $\mathbf{A}^1, \ldots, \mathbf{A}^p$ stand for the p lines of $\mathbf{A}$. By (3.26) we obtain $\mathbf{A}^i \in A_{f_i}(x)$, for $i \in \{1, \ldots, p\}$. Using (3.25) and the Caratheodory theorem we deduce that for some $0 \leq m_i \leq d$, $\boldsymbol{\lambda}^i = (\lambda_0^i, \ldots, \lambda_{m_i}^i) \in \operatorname{inn} \Delta^{m_i}$ and $J_i = \left\{t_0^i, \ldots, t_{m_i}^i\right\} \subset T_i(x)$ we have

$$\mathbf{A}^i = \sum_{j_i=0}^{m_i} \lambda_{j_i}^i \, D_x F_i(x, t_{j_i}^i).$$

Set $\mathcal{J} = (J_1, \ldots, J_p)$ and consider the function $G_{\mathcal{J}}$ defined in (3.29). Since $J_i \subset T_i(x)$ for all $i \in \{1, \ldots, p\}$, we have

$$F_i(x, t_{j_i}^i) = f_i(x), \qquad \text{for all } j_i \in \{0, \ldots, m_i\}.$$

It follows that

$$G_{\mathcal{J}}(\boldsymbol{\lambda}^1, \ldots, \boldsymbol{\lambda}^p, x) = f(x),$$

and the claim is proved. ♦

Therefore, provided we establish that the set $G_{\mathcal{J}}(\widehat{\operatorname{Crit}}\, G_{\mathcal{J}})$ is null, we shall deduce from Claim 2, since $\mathcal{F}^p$ is countable, that the set

$$f(\operatorname{Crit} f) \subset \bigcup_{\mathcal{J} \in \mathcal{F}^p} G_{\mathcal{J}}(\widehat{\operatorname{Crit}}\, G_{\mathcal{J}})$$

is null in $\mathbb{R}^p$ and we are done. Therefore, it remains to establish that for a given $\mathcal{J} = (J_1, \ldots, J_p) \in \mathcal{F}^p$ and function $G_{\mathcal{J}}$ the set $G_{\mathcal{J}}(\widehat{\operatorname{Crit}}\, G_{\mathcal{J}})$ is null.

To this end, we shall use Theorem 3 to an appropriate function Ψ of the form (1.2) that we define below. We set

$$m = \left(\prod_{i=1}^{p} |J_i|\right) - 1 \qquad \text{and} \qquad \begin{cases} \vec{i} = (i_1, i_2, \ldots, i_p) \\ i_j \in \{0, \ldots, m_j\} \end{cases}$$

and we describe the elements of $\operatorname{inn}\Delta^m$ by means of the above multi-indices, that is, $\lambda = (\lambda_{\vec{i}}) \in \operatorname{inn}\Delta^m$. We now consider the following $m+1$ functions

$$\begin{cases} \phi^{\vec{i}} : \mathbb{R}^d \to \mathbb{R}^p, \quad \vec{i} = (i_1, i_2, \ldots, i_p) \\ \phi^{\vec{i}}(x) = \Big(F_1(x, t^1_{i_1}), \ldots, F_p(x, t^p_{i_p}) \Big) \end{cases} \tag{3.30}$$

and the function

$$\Psi(\lambda, x) := \sum_{i_1=0}^{m_1} \cdots \sum_{i_p=0}^{m_p} \lambda_{\vec{i}} \phi^{\vec{i}}(x), \quad (\lambda, x) \in \operatorname{inn}\Delta^m \times \mathbb{R}^d.$$

Notice that the functions $\phi^{\vec{i}}$ and consequently the function Ψ are of class C^k with $k \geq d-p+1$. Thus Theorem 3 (I) applies yielding

$$\Psi(\widehat{\mathrm{Crit}}\Psi) \text{ is null in } \mathbb{R}^p.$$

It remains to establish the following.

Claim 3 (transferring strong criticality from $G_{\mathcal{J}}$ to Ψ): Assume $(\boldsymbol{\lambda}^1, \ldots, \boldsymbol{\lambda}^p, x) \in \widehat{\mathrm{Crit}}\, G_{\mathcal{J}}$. Then for $\lambda = (\lambda_{\vec{i}}) \in \operatorname{inn}\Delta^m$ with

$$\lambda_{(i_1,\ldots,i_p)} = \lambda^1_{i_1} \cdots \lambda^p_{i_p} \tag{3.31}$$

it holds

$$(\boldsymbol{\lambda}, x) \in \widehat{\mathrm{Crit}}\Psi \qquad \text{and} \qquad \Psi(\boldsymbol{\lambda}, x) = G_{\mathcal{J}}(\boldsymbol{\lambda}^1, \ldots, \boldsymbol{\lambda}^p, x).$$

Proof of Claim 3. The proof is straightforward in view of (3.30), (3.31) and the definition of strongly critical points for $G_{\mathcal{J}}$. ♦

This concludes the proof of the claim and in turn of the result. □

4 Applications in Optimization

In this section we discuss applications of Theorem 2 (Sard for Lipschitz selections) in Optimization.

4.1 Semi-infinite programming

Let $g_t : \mathbb{R}^d \to \mathbb{R}$, $t \in T$, be a family of C^k functions, $k \geq d-p+1$, parameterized on a nonempty countable compact set T such that both $(x,t) \longmapsto g_t(x)$ and $(x,t) \longmapsto \nabla g_t(x)$ are continuous. We consider the following semi-infinite minimization problem (depending on a scalar parameter $r \in \mathbb{R}$):

$$(\mathcal{P}_r) \qquad \min_{g_t(x) \leq r} \quad u(x)$$

where $u : \mathbb{R}^d \to \mathbb{R}$ is a Clarke regular Lipschitz continuous function. In what follows we assume that the feasibility set

$$C_r := \{x \in \mathbb{R}^d : g_t(x) \leq r\,;\ t \in T\}$$

is nonempty. A feasible point $\bar{x} \in \mathbb{R}^d$ is said to be a *solution* of $(\mathcal{P}_r)$ provided for some $\delta > 0$ it holds:

$$x \in C_r \cap B_\delta(\bar{x}) \Longrightarrow u(x) \geq u(\bar{x}). \tag{4.1}$$

Let us observe that

$$C_r := \{x \in \mathbb{R}^d : f(x) \leq r\},$$

where

$$f(x) := \max\{g_t(x) : t \in T\}. \tag{4.2}$$

We further denote by

$$T(x) := \{t \in T : g_t(x) = f(x)\} \tag{4.3}$$

the set of active indices at $x \in \mathbb{R}^d$. Thanks to our assumptions, the above set is always a nonempty compact subset of T. In the sequel we are interested in determining necessary optimality conditions for the solutions of $(\mathcal{P}_r)$.

Proposition 9 (Genericity of optimality conditions for $(\mathcal{P}_r)$)**.** *For almost all values of the parameter $r \in \mathbb{R}$, every solution $\bar{x}$ of $(\mathcal{P}_r)$ satisfies a necessary Karush-Kuhn-Tucker (in short, KKT) condition. Namely there exist $\lambda_1, \ldots, \lambda_d \geq 0$ and $\{t_1, \ldots, t_d\} \subset T(\bar{x})$ such that*

$$0 \in \partial^C u(\bar{x}) + \sum_{i=1}^{d} \lambda_i \nabla g_{t_i}(\bar{x}) \tag{4.4}$$

Proof. The function f defined in (4.2) is lower-C^1 (see [7] for example), hence locally Lipschitz continuous and Clarke regular. Moreover it satisfies the assumptions of Theorem 2 (Sard for Lipschitz selections), thus the set of its Clarke critical values has measure zero.

Let us now fix a *regular* value $r \in \mathbb{R}$ of the above function f and consider a solution (local minimum) $\bar{x}$ of $(\mathcal{P}_r)$, *i.e.* $\bar{x} \in C_r$ and (4.1) holds. If $f(\bar{x}) < r$ then $\bar{x}$ belongs to the interior of C_r and $0 \in \partial^C u(\bar{x})$. Thus (4.4) holds trivially, by picking $t_* \in T(\bar{x})$ and taking $t_1 = \ldots = t_d = t_*$ and $\lambda_1 = \ldots = \lambda_d = 0$. Consequently, in the sequel we may assume $f(\bar{x}) = r$. We now define:

$$\begin{cases} \Phi : \mathbb{R}^d \to \mathbb{R} \\ \Phi(x) = \max\{f(x), u(x) - u(\bar{x}) + r\} \end{cases}$$

It follows readily that Φ is locally Lipschitz. By [5, Prop. 2.3.12] we deduce that Φ is Clarke regular and

$$\partial^C \Phi(x) \subset \operatorname{co}\left\{\partial^C f(x), \partial^C u(x)\right\}.$$

Notice that if $x \notin C_r$ then $\Phi(x) \geq f(x) > r = \Phi(\bar{x})$. On the other hand, if $x \in C_r \cap B_\delta(\bar{x})$, then $u(x) \geq u(\bar{x})$ and $f(x) \leq r$ which yields $\Phi(x) \geq r = \Phi(\bar{x})$. It follows that $\bar{x}$ is a local minimum of Φ, thus

$$0 \in \partial^C \Phi(\bar{x}) = \operatorname{co}\left(\partial^C u(\bar{x}) \cup \partial^C f(\bar{x})\right). \tag{4.5}$$

In view of (4.2), we deduce that

$$\partial^C f(\bar{x}) = \text{co}\,\{\nabla g_t(\bar{x}) : t \in T(\bar{x})\}$$

which combined with (4.5) finally yields

$$0 \in \partial^C \Phi(\bar{x}) = \text{co}\,\left(\partial^C u(\bar{x}) \cup \{\nabla g_t(\bar{x}) : t \in T(\bar{x})\}\right).$$

By the Caratheodory theorem there exist $k \in \{1, \ldots, d\}$, $\mu = (\mu_0, \ldots, \mu_k) \in \text{inn}\,\Delta^k$ and

$$\{p_i\}_{i=0}^k \subset \partial^C u(\bar{x}) \cup \{\nabla g_t(\bar{x}) : t \in T(\bar{x})\}$$

such that

$$0 \,=\, \mu_0 p_0 \,+ \ldots +\, \mu_k p_k\,.$$

Clearly $\{p_i\}_{i=0}^k \cap \partial^C u(\bar{x}) \neq \emptyset$ (else $0 \in \partial^C f(\bar{x})$ and r would be a critical value of f, which is excluded by the choice of the value r to be a regular value). If $\{p_i\}_{i=0}^k \subset \partial^C u(\bar{x})$ then $0 \in \partial^C u(\bar{x})$ and (4.4) holds trivially true. We may thus assume that there exists at least one $p_i \notin \partial^C u(\bar{x})$. Without loss of generality we may assume that $\{p_0, \ldots, p_m\} \subset \partial^C u(\bar{x})$ and $\{p_{m+1}, \ldots, p_k\} \subset \{\nabla g_t(\bar{x}) : t \in T(\bar{x})\}$ for some $m \in \{0, \ldots, k-1\}$. We set

$$\hat{\mu} = \sum_{i=0}^{m} \mu_i \in (0, 1),$$

$\hat{\mu}_i = \mu_i / \hat{\mu}$ for $i \in \{0, \ldots, m\}$, and

$$\lambda_i = \begin{cases} \mu_{m+i}/\hat{\mu}\,, & \text{for } i \in \{1, \ldots, k-m\} \\ 0\,, & \text{for } i \in \{k-m+1, \ldots, d\} \end{cases}$$

Then $(\hat{\mu}_0, \ldots, \hat{\mu}_m) \in \Delta^m$ and by convexity of $\partial^C u(\bar{x})$,

$$\sum_{i=0}^{m} \hat{\mu}_i p_i \in \partial^C u(\bar{x}),$$

whence (4.4) follows. □

4.2 Semi-infinite programming with loose (nonrigorous) constraints

We adopt the same framework as in Section 4.1, namely, we consider again a regular Lipschitz continuous objective function $u : \mathbb{R}^d \to \mathbb{R}$ which is to be minimized under a countable set of restrictions $g_t(x) \leq r$, $t \in T$, where $r \in \mathbb{R}$ is a scalar parameter, T is a nonempty countable compact set T and the functions $(x, t) \longmapsto g_t(x)$ and $(x, t) \longmapsto \nabla g_t(x)$ are continuous. However, in contrast to $(\mathcal{P}_r)$, we now allow a certain number of constraints, say $k \in \mathbb{N}$, to be violated. The feasibility set is therefore larger and given as follows:

$$\begin{aligned} & C_r^{(-k)} := \{x \in \mathbb{R}^d : f^{(-k)}(x) \leq r\}, \\ \text{where} \quad & \\ & f^{(-k)}(x) \;=\; \inf_{F \subset T\,;\, |F| = k}\ \sup_{t \in T \setminus F} g_t(x). \end{aligned} \tag{4.6}$$

Notice that $x \in C_r^{(-k)}$ if and only if $g_t(x) \le r$ for all but k restrictions $t \in T$. Under this notation, we are interested in determining necessary optimality conditions for the solutions of the following minimization problem:

$$(\mathcal{P}_r^{(-k)}) \qquad \min_{f^{(-k)}(x) \le r} \quad u(x)$$

To this end, we shall need the following result:

Proposition 10 (Continuous Selection). *The function $f^{(-k)}$ given in (4.6) is a continuous selection of the family $\{g_t\}_{t\in T}$. In particular, $f^{(-k)}$ is Lipschitz continuous and satisfies the generalized Morse-Sard theorem for its Clarke critical values.*

Proof. Let us recall that the Cantor derivative T' of the set T is the set of all accumulation points of T, that is,

$$T' = \{t \in T : t \in \overline{T \diagdown \{t\}}\}.$$

For every subset F of T we set

$$f_{-F}(x) := \sup_{t \in T \diagdown F} g_t(x).$$

Notice that if $F \subset T \diagdown T'$ then the above supremum is attained (*i.e.* it is a maximum). For every $m \in \{1, \ldots, k\}$ we consider the function

$$f^{[-m]}(x) \; = \inf_{F \subset T \diagdown T';\, |F|=m} f_{-F}(x) = \inf_{F \subset T \diagdown T';\, |F|=m} \; \max_{t \in T \diagdown F} g_t(x). \tag{4.7}$$

Claim. The infimum in (4.7) is attained (*i.e.* it is a minimum). In particular, $f^{[-m]}$ is a (Lipschitz continuous) selection of $\{g_t\}_{t\in T}$.

*Proof of Claim***.** Let us recall from (4.2) the function

$$f(x) := \max\,\{g_t(x) : t \in T\}$$

and the corresponding set $T(x)$ of active indices given by (4.3). If $T(x) \cap T' \neq \emptyset$, then $f_{-F}(x) = f(x)$ for all $F \subset T \diagdown T'$ and consequently $f^{[-m]}(x) = f_{-F}(x) = f(x)$, for all $m \in \{1, \ldots, k\}$ and $F \subset T \diagdown T'$ with $|F| = m$. In this case the infimum in (4.7) is trivially attained. Therefore, in what follows, we may assume that $T(x) \cap T' = \emptyset$.

If $|T(x)| > m$, then again $f^{[-m]}(x) = f_{-F}(x) = f(x)$, for every $F \subset T$ with $|F| = m$ (and the infimum is trivially attained). If $|T(x)| = m$, then for $\hat{F} = T(x)$ we have

$$f^{[-m]}(x) \; = \; f_{-\hat{F}}(x) \; < \; f_{-F}(x), \quad \text{for all } F \neq \hat{F} \text{ with } |F| = m$$

and the infimum is attained at $\hat{F}$. It remains to consider the case $|T(x)| < m$ (and $T(x) \cap T' = \emptyset$). In this case we set

$$m_1 := m - |T(x)| \, \in \{1, \ldots, m-1\},$$

$T_1 = T \diagdown T(x)$ and

$$f_1(x) = \max\,\{g_t(x) : t \in T_1\} < f(x).$$

Notice that $T_1' = T'$ and

$$f^{[-m]}(x) = f_1{}^{[-m_1]}(x) = \inf_{F_1 \subset T_1 \setminus T'; \, |F_1| = m_1} (f_1)_{-F_1}(x). \tag{4.8}$$

The above analysis applying to $f_1{}^{[-m_1]}(x)$ yields that the above infimum is attained whenever either $T_1(x) \cap T' \neq \emptyset$ or $|T_1(x)| \geq m_1$. Notice that if the infimum in the definition of $f_1{}^{[-m_1]}(x)$ is attained at F_1 then the infimum in the definition of $f^{[-m]}(x)$ is attained at $F := F_1 \cup T(x)$, see (4.7) and (4.8) respectively. If now $T_1(x) \cap T' = \emptyset$ and $|T_1(x)| < m_1$, then we set $m_2 := m_1 - |T_1(x)| < m_1$, $T_2 = T_1 \setminus T_1(x)$ and continue as before. This process will eventually stop, since $m > m_1 > m_2 > \ldots$ is a strictly decreasing sequence of natural numbers, which has to be finite. We conclude that

$$f^{[-m]}(x) = \min_{F \subset T \setminus T' \, ; \, |F| = m} \; \max_{t \in T \setminus F} g_t(x).$$

This shows that $f^{[-m]}$ is a selection of $\{g_t\}_{t \in T}$. Moreover, thanks to our assumption on the family $\{g_t\}$, the functions $\{f_{-F} : F \subset T \setminus T' \, ; \, |F| = m\}$ are uniformly Lipschitz continuous. A standard argument now shows that $f^{[-m]}$ is Lipschitz continuous. This concludes the proof of the claim. ♦

We deduce easily from the above claim that

$$f^{(-k)}(x) = \min_{m \leq k} f^{[-m]}(x) = \min_{F \subset T \setminus T'; \, |F| \leq k} \; \max_{t \in T \setminus F} g_t(x)$$

is a selection of the family $\{g_t : t \in T\}$ and Lipschitz continuous. □

We now obtain the following result:

Corollary 11 (Genericity of optimality conditions for $(\mathcal{P}_r^{(-k)}$)). *For almost all values of the parameter $r \in \mathbb{R}$, every solution $\bar{x}$ of $(\mathcal{P}_r^{(-k)})$ satisfies a necessary KKT type condition. Namely there exist $\lambda_1, \ldots, \lambda_d \geq 0$ and $\{t_1, \ldots, t_d\} \subset T(\bar{x})$ such that*

$$0 \in \partial^C u(x) + \sum_{i=1}^{d} \lambda_i \nabla g_{t_i}(x)$$

Proof. By Proposition 10 and Theorem 2 almost every value $r \in \mathbb{R}$ is Clarke regular for $f^{(-k)}$. Fix any such value r and consider an optimal solution $x \in C_r^{(-k)}$ of the problem $(\mathcal{P}_r^{(-k)})$. Then $f^{(-k)}(x) \leq r$. Repeating the proof of Proposition 9 with f being replaced by $f^{(-k)}$ and C_r by C_r^{-k} we obtain the result. □

4.3 Vector Optimization: Pareto minimal values

Let P be a nontrivial closed convex cone of $\mathbb{R}^d$ and $f : \mathbb{R}^d \to \mathbb{R}^d$. A point $\bar{x} \in \mathbb{R}^d$ is called *Pareto minimum point* for the function f provided there exists $\delta > 0$ such that

$$\left. \begin{array}{c} x \in B(\bar{x}, \delta) \\ f(\bar{x}) \in f(x) + P \end{array} \right\} \Longrightarrow f(x) = f(\bar{x}).$$

In this case $\bar{y} = f(\bar{x})$ is called a *Pareto minimal value.* We now present an application of our main result concerning the size of the set of Pareto minimal values.

Proposition 12 (Pareto values)**.** *Let $F : \mathbb{R}^d \times T \to \mathbb{R}^d$ be a continuous function. Assume F is C^k, $k \geq d - p + 1$, with respect to the first variable and $D_x F$ is continuous. Then the set of Pareto minimal values of any continuous selection f of the family $\{F_t = F(\cdot, t)\}_{t \in T}$ is null.*

Proof. Let $\bar{y} = f(\bar{x})$ be a Pareto minimal value of f. Since $f(B(\bar{x}, \delta)) \cap (\bar{y} - P) = \{\bar{y}\}$ it follows directly that f is not locally surjective there. By [5, Theorem 7.1.1] $\bar{y}$ has to be a Clarke critical value. The result follows from Theorem 2. $\square$

Acknowledgement. Parts of this work have been elaborated under short research visits of the third author at the University of Pau (France) and of the fourth author at the University Autonomous of Barcelona (Spain) and the University of Santiago (Chile). These authors thank their hosts for hospitality.

References

[1] L. Barbet, A. Daniilidis, M. Dambrine, The Morse-Sard Theorem for Clarke critical values, *Advances in Mathematics* **242** (2013), 217–227.

[2] S. Bates, Toward a precise smoothness hypothesis in Sard's theorem, *Proc. Amer. Math. Soc.* **117** (1993), 279–283.

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

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

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

[6] G. Comte, Y. Yomdin, Tame geometry with application in smooth analysis, *Lecture Notes in Mathematics* **1834** (Springer, 2004).

[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] H. Federer, Curvature measures, *Trans. Amer. Math. Soc.* **93** (1959), 418–491.

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

[10] A. Jofre, *Image of critical points set for nonsmooth mappings and applications*, preprint 22p, Université de Pau et des Pays de l'Adour,

[11] K. Kurdyka, P. Orro, S. Simon, Semialgebraic Sard theorem for generalized critical values, *J. Differential Geom.* **56** (2000), 67–92.

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

[13] A. Norton, The Zygmund Morse-Sard theorem, *J. Geom. Anal.* **4** (1994), 403–424.

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

[15] L. Rifford, A propos des sphères sous-riemanniennes, *Bull. Belg. Math. Soc. Simon Stevin* **13** (2006), 521–526.

[16] 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.

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

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

[19] 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

DIM–CMM, UMI CNRS 2807
Blanco Encalada 2120, piso 5, Universidad de Chile

E-mail: `arisd@dim.uchile.cl`
`http://www.dim.uchile.cl/~arisd`

Research supported by the grants:
BASAL PFB-03 (Chile), FONDECYT 1130176 (Chile) and MTM2011-29064-C03-01 (Spain).

Ludovic Rifford

Laboratoire J.A. Dieudonné
Université Nice Sophia Antipolis
Parc Valrose, F-06108 Nice Cedex 2, France

E-mail: `ludovic.rifford@math.cnrs.fr`
`http://math.unice.fr/~rifford/`