# THE SARD THEOREM FOR ESSENTIALLY SMOOTH LOCALLY LIPSCHITZ MAPS AND APPLICATIONS IN OPTIMIZATION

TRUONG XUAN DUC HA

Abstract. The classical Sard theorem states that the set of critical values of a $C^k$-map from an open set of $\mathbb{R}^n$ to $\mathbb{R}^p$ $(n \geq p)$ has Lebesgue measure zero provided $k \geq n - p + 1$.

In the recent paper by Barbet, Dambrine, Daniilidis and Rifford, the so called "preparatory Sard theorem" for a compact countable set $I$ of $C^k$ maps from $\mathbb{R}^n$ to $\mathbb{R}^p$ and a Sard theorem for a locally Lipschitz continuous selection of this family have been established under the assumption that $k \geq n-p+1$. Here, we show that, in the special case $n = p$ and $I$ is finite, the $C^1$ smoothness assumption in these results can be relaxed to "essentially smooth locally Lipschitz". Then we apply the obtained results to study Karush-Kuhn-Tucker type necessary condition for scalar/vector parametrized constrained optimization problems and the set of Pareto optimal values of a continuous selection of a finite family of essentially smooth locally Lipschitz maps.

## 1. Introduction

The classical Sard theorem states that the set of critical values of a $C^k$-map from an open set of $\mathbb{R}^n$ to $\mathbb{R}^p$ (we restrict ourselves to the case $n \geq p$) has Lebesgue measure zero provided $k \geq n - p + 1$, see [23]. Here, $y$ is a critical value if $\text{rank} f'(x) < p$ for all $x \in f^{-1}(y)$. The case $p = 1$, known as the Morse-Sard theorem, has been considered earlier in [17]. The Sard theorem has been extended to d.c. functions (d.c. means "difference of convex functions") defined on $\mathbb{R}^2$ in [15]. Later, it was shown that the $C^k$ smoothness hypothesis can be weakened to $C^{k-1,1}$, see [5] and its refinement [18] for the case $n > p$ and [13, 14, 22] for the case $n = p$. Here, $C^{k-1,1}$ $(k > 1)$ means that $f$ is $C^{k-1}$ and its $(k-1)$th derivative is locally Lipschitz and $C^{0,1}$ means that $f$ is locally Lipschitz. There is a rich literature on important applications of the Sard theorem and extensions of this result to other classes of maps. We cannot mention all of works in this area and we restrict ourselves to the survey [2] and the references therein.

---

*Date*: December 4, 2018.

1991 *Mathematics Subject Classification.* 49J40.

*Key words and phrases.* Sard theorem, locally Lipschitz, essentially smooth map, critical points, optimization.

$^\sharp$ This work was partially carried out during the author's stay at the Vietnam Institute for Advanced Study in Mathematics.

We would like to stay more in the case $f$ is a map between spaces of the same dimension. The differentiability condition required in the Sard theorem is that $f$ is a $C^1$-map, i.e. it is continuously differentiable. However, the map $f$ is dictated by a problem at hand, and certainly may fail to be continuously differentiable, see [22, 24] for instance for problems in economy. While searching for the notions of criticality for nonsmooth maps, we found two approaches. The approach taken by Rader (1973) [22] is to work with the classical derivative. Namely, critical values of $f$ are those $y$, for which at some preimage point $x \in f^{-1}(y)$, either $f$ is not differentiable or $f$ is differentiable but $\det f'(x) = 0$. In [22, Lemma 2], an analogue of the Sard theorem with this concept of critical value has been established for a map, which is almost everywhere Fréchet differentiable and has the property that it maps any set of Lebesgue measure zero into a set with Lebesgue measure zero ("almost everywhere" means "except a set of Lebesgue measure zero"). We refer an interested reader to [24] for applications of the mentioned result in economy and to [14] for a detailed proof in the case of locally Lipschitz maps.

The second approach, stemming from the work of Clarke (1975) [11] for locally Lipschitz functions, seeks to replace the classical derivative with the Clarke subdifferential, which is a set of generalized derivative and is nonempty at every point in the domain of a considered function. As it was remarked in [4], endeavoring to generalize the Sard theorem for locally Lipschitz maps is a huge challenge even in the simplest nontrivial case $n = p$. In fact, the set of the Clarke critical values may have the positive Lebesgue measure, see Section 2. Moreover, as mentioned in [4], the set of the Clarke critical points may be equal to the whole space $\mathbb{R}$ for a generic (with respect to the uniform topology) function $f$ when $n = 1$ (see [10] for the general case $n > 1$). To obtain the Sard theorem for the Clarke critical values, an extra assumption is needed. In [13], we showed that the set of the Clarke critical values has Lebesgue measure zero provided that $f$ is *essential smooth*. This means, roughly speaking, that $f$ is almost everywhere strict differentiable. Essentially smooth maps form a significant subclass of the class of locally Lipschitz maps and contain the class of *continuous d.c.* functions. We refer the interested reader to the works by Borwein and Woors [6, 7, 8, 9] for the basic properties and applications of essential smooths maps.

Recently, Barbet, Dambrine, Daniilidis and Rifford established in [4] the so called "preparatory Sard theorem" for a compact countable set $I$ of $C^k$ maps from $\mathbb{R}^n$ to $\mathbb{R}^p$ with $k \geq n-p+1$ and a Sard theorem for a locally Lipschitz continuous selection of this family. In this paper, we prove that the $C^1$ assumption in the above results can be weakened to "essentially smooth locally Lipschitz" in the special case $n = p$ and the indexed family $I$ is finite. As applications, we show that for almost all parameters, any optimal solution/weak Pareto optimal solution to scalar/vector parametrized constrained optimization problems satisfies

Karush-Kuhn-Tucker necessary condition and the set of Pareto optimal values of a continuous selection of a finite family of essential smooth locally Lipschitz maps has Lebesgues measure zero. Our arguments are essentially based on results about the Clark subdifferential from [11] and some techniques of [3, 4, 13]. Note that in contrast to [4], applying [12, Theorem 6.1.1] allows us to consider also the case of a *vector* parametrized constrained optimization problem. We provide examples illustrating obtained results.

The parer is organized as follows. Section 2 contains preliminaries. Sections 3 and 4 concern with the "preparatory Sard theorem" and the Sard theorem for Lipschitz selections, respectively. Section 5 contains applications to optimization.

## 2. Preliminaries

In this section, we recall some concepts of differentiability, the Clarke subdifferential and some known versions of the Sard theorem for locally Lipschitz maps. The map under our consideration is assumed to be "essentially smooth", a term introduced by Borwein in [6, p.68] and "locally Lipschitz". Note that there are several related concepts of differentiability in existing literature, however they are expressed in different forms, see for instance, the works [1, 7, 12, 16, 19]. To clarify the term "essentially smooth" and for the reader's convenience, we will recall these concepts and related to them results in an unified scheme.

For a normed space $X$, $d(x, U)$ is the distance from a point $x \in X$ to a nonempty subset $U$ of $X$, $\mathbb{B}(x, \rho)$ is the closed ball centered at $x \in X$ with the radius $\rho$, $\mathbb{B}_X$ (or simply $\mathbb{B}$) is the closed unite ball of $X$ and $\rho\mathbb{B} := \mathbb{B}(0, \rho)$. Let $U$ be a nonempty subset of the Euclidean $n$-dimensional space $\mathbb{R}^n$. We shall use $\text{mes}U$ and $\text{mes}^*U$ to denote the Lebesgue measure and the outer Lebesgue measure of $U$, respectively. The term "almost everywhere"(briefly, a.e.) on $U$ means everywhere on $U$ except a set with Lebesgue measure zero.

Let $\mathbb{R}^{p\times n}$ be the space of $p \times n$-matrices equipped with the norm

$$\|(a_{ij})\|_{\mathbb{R}^{p\times n}} := (\sum_{i=1}^{p}\sum_{j=1}^{n} a_{ij}^2)^{1/2}.$$

For $A \in \mathbb{R}^{p\times n}$, $\text{rank}A$ is the dimension of the image set $A(\mathbb{R}^n)$. We can also write $A$ in the form $A = (A_1, ..., A_n)$, $A_i \in \mathbb{R}^p$ for $i = 1, 2, ..., n$, and then $\text{rank}A$ can be defined equivalently as the largest number of vectors $A_i$, which are linearly independent.

Let $X$, $Y$ be Banach spaces, $\Omega \subset X$ a nonempty open set and $f : \Omega \to Y$ a map. Let $x \in \Omega$.

**Definition 2.1.** (a) We say that $f$ is *globally Lipschitz* (or *Lipschitz continuous* as in some papers) if there exists a constant $L > 0$ such that

$$\|f(x_1) - f(x_2)\| \le L\|x_1 - x_2\|$$

for all $x_1, x_2 \in \Omega$ and that $f$ is *locally Lipschitz* at $x \in \Omega$ or near $x$ if the above inequality holds for all $x_1, x_2 \in U$, where $U \subset \Omega$ is some neighborhood of $x$.

(b) We say that a continuous linear functional from $X$ into $Y$, denoted by $f'(x)$ (for the sake of convenience, also by $Df(x)$ sometimes), is

(i) the *Gâteaux derivative* of $f$ at $x$ if

$$\lim_{t\to 0} \|\frac{1}{t}[f(x+tv) - f(x)] - f'(x)(v)\| = 0$$

for any $v \in X$ [1, p.16].

(ii) the *Fréchet derivative* of $f$ at $x$ if

$$\lim_{t\to 0^+} \|\frac{1}{t}[f(x+tv) - f(x)] - f'(x)(v)\| = 0$$

and the convergence is uniform with respect to $v$ in bounded sets [12, p. 30] or, equivalently, if

$$\lim_{v\to 0} \|\frac{1}{\|v\|}[f(x+v) - f(x)] - f'(x)(v)\| = 0,$$

see [1, p.16].

(iii) the *Fréchet-type strict derivative* of $f$ at $x$ if

$$\lim_{u\to x, t\to 0^+} \frac{1}{t}[f(u+tv) - f(u)] - f'(x)(v)\| = 0$$

and the convergence is uniform with respect to $v$ in bounded sets (see [8, p. 308], under the name *strict Fréchet derivative*) or, equivalently,

$$\lim_{u\to x, v\to 0} \|\frac{1}{\|v\|}[f(u+v) - f(u)] - f'(x)(v)\| = 0$$

(see [1, p.16], under the name *strict derivative*).

(iv) the *Hadamard-type strict derivative* of $f$ at $x$ if

$$\lim_{u\to x, t\to 0^+} \frac{1}{t}[f(u+tv) - f(u)] - f'(x)(v)\| = 0$$

and the convergence is uniform with respect to $v$ in compact sets [12, p. 30-31].

(v) the *strong derivative* of $f$ at $x$ if for all $\epsilon > 0$ there is exists $\delta > 0$ such that $\mathbb{B}(x, \delta_x) \subset \Omega$ and for any pair $x_1, x_2 \in \mathbb{B}(x, \delta_x)$, one has

$$\|f(x_1) - f(x_2) - f'(x)(x_1 - x_2)\| \leq \epsilon \|x_1 - x_2\|$$

whenever $x_1, x_2 \in \mathbb{B}(x, \delta_x)$ [19, p. 324].

In these cases, we say that $f$ is *Gâteaux differentiable*, *Fréchet differentiable*, *Fréchet-type strictly differentiable*, *Hadamard-type strictly differentiable*, *strongly differentiable* at $x$, respectively.

(b) We say that $f$ is $C^1$ or *continuously Gâteaux differentiable* at $x$ if there exists a neighborhood $U$ of $x$ such that $f$ is Gâteaux differentiable on $U$ and $f'(u) \to f'(x)$ in $\mathbb{R}^{n\times p}$ when $u \to x$, $u \in U$ [1, p.17].

**Definition 2.2.** Assume that $f$ is locally Lipschitz at $x$. The *genaralized directional derivative* of $f$ at $x$ in the direction $v \in X$, denoted by $f^0(x, v)$, is defined as follows

$$f^0(x,v) = \limsup_{u\to x, t\to 0^+} \frac{f(u+tv)-f(u)}{t}$$

and the Clarke subdifferential of $f$ at $x$, denoted by $\partial f(x)$, is defined as follows

$$\partial f(x) := \{x^* \in X^* \mid \langle x^*, v\rangle \le f^0(x,v) \text{ for all } v \in X\},$$

see [12, p. 25, 27].

We recall some known facts about relations among the mentioned derivatives.

**Remark 2.1.** (i) It has been established in [1, Lemma 1, p. 17] that

$$\begin{aligned} C^1 &\Rightarrow \text{Fréchet-type strict differentiability} \Rightarrow \text{Fréchet differentiability} \\ &\Rightarrow \text{Gâteaux differentiability.} \end{aligned}$$

(ii) If $f$ is locally Lipschitz at $x$ and admits Fréchet derivatives $f'(x)$, then $f'(x) \in \partial f(x)$ [12, Proposition 2.2.2].

(iii) If $f$ is $C^1$ at $x$, then $f$ is Hadamard-type strictly differentiable and locally Lipschitz at this point [12, Corolllary, p.32].

(iv) If $f$ is Hadamard-type strictly differentiable at $x$, then it is Lipschitz near $x$ and $\partial f(x) = \{f'(x)\}$. Conversely, if $f$ is Lipschitz near $x$ and $\partial f(x)$ is a singleton $\{\xi\}$, then $f$ is Hadamard-type strictly differentiable at $x$ and $f'(x) = \xi$ [12, Proposition 2.2.4].

(v) Assume that $\Omega \subset \mathbb{R}^n$ is an open convex set and $f : \Omega \to \mathbb{R}$ is a convex function. Then the Clarke subdifferential of $f$ at $x \in \Omega$ coincides with the subdifferential of convex analysis (hereafter denoted by the same notation $\partial f(x)$ in this convex case) given by

$$\partial f(x) := \{x^* \in \mathbb{R}^n \ : \ \langle x^*, x' - x\rangle \le f(x') - f(x) \ \ \forall x' \in \Omega\}.$$

Let us establish relations among the Fréchet-type strict differentiability, the Hadamard-type strict differentiability and the strong differentiability, which plays an important role in our proof of Sard theorems.

**Proposition 2.1.** (i) *Strong differentiability $\Rightarrow$ Hadamard-type strict differentiability.*

(ii) *If $X$ has a finite dimension, then the strong differentiability, the Fréchet-type strict differentiability and the Hadamard-type strict differentiability coincide.*

*Proof.* (i) As $f$ is strongly differentiable at $x$, for all $\epsilon > 0$ there is exists $\delta_x > 0$ such that $\mathbb{B}(x, \delta_x) \subset \Omega$ and for all $x_1, x_2 \in \mathbb{B}(x, 2\delta_x)$, one has

$$\|f(x_1) - f(x_2) - f'(x)(x_1 - x_2)\| \leq \epsilon \|x_1 - x_2\|. \tag{1}$$

Assume that $K$ is some compact set of $Y$ and $v \in K$ be an arbitrary vector. Let $\rho > 0$ be a scalar such that $K \subset \rho\mathbb{B}$. For any $x' \in B(x, \delta_x)$ and $t \in ]0, \delta_x/\rho]$, we have $\|x' + tv - x\| \leq \|x' - x\| + t\|v\| \leq \delta_x + (\delta_x/\rho)\rho = 2\delta_x$ or $x' + tv \in \mathbb{B}(x, 2\delta_x)$. The inequality (1) implies

$$\|f(x' + tv) - f(x') - f'(x)(tv)\| \leq \epsilon t \|v\|$$

or

$$\|\frac{1}{t}[f(x' + tv) - f(x')] - f'(x)(v)\| \leq \rho\epsilon.$$

This means that the convergence of the limit in the definition of the strict differentiability is uniform with respect to all $v \in \rho\mathbb{B}$ and hence, uniform with respect to all $v \in K$. Thus, $f$ is Hadamard-type strictly differentiable at $x$.

(ii) Since $X$ has a finite dimension, we can replace "compact sets" in the definition of the Hadamard-type strict differentiability by "bounded sets". Hence, the Fréchet-type strict differentiability and the Hadamard-type strict differentiability coincide. Next, assume that $f$ is Hadamard-type strictly differentiable at $x$. Then for any $\epsilon > 0$ there is exists $\delta_x > 0$ such that $\mathbb{B}(x, \delta_x) \subset \Omega$ and for all $x' \in \mathbb{B}(x, \delta_x)$ and $t \in ]0, 2\delta_x]$, the inequality

$$\|\frac{1}{t}[f(x' + tv) - f(x')] - f'(x)(v))\| \leq \epsilon \tag{2}$$

holds for all $v \in \mathbb{B}$. Let $x_1, x_2 \in \mathbb{B}(x, \delta_x)$ with $x_1 \neq x_2$. Observe that

$$\frac{1}{\|x_1 - x_2\|}[f(x_2) - f(x_1)] - f'(x)(\frac{x_2 - x_1}{\|x_1 - x_2\|}) = \frac{1}{t}[f(x_1 + tv) - f(x_1)] - f'(x)(v),$$

with $t = \|x_1 - x_2\|$, $v = (x_2 - x_1)/\|x_1 - x_2\|$. Since $0 < t \leq 2\delta_x$, $v \in \mathbb{B}$ and $x_2 = x_1 + tv$, the inequality (2) implies

$$\|\frac{1}{\|x_1 - x_2\|}[f(x_2) - f(x_1)] - f'(x)(\frac{x_2 - x_1}{\|x_1 - x_2\|})\| \leq \epsilon$$

or $\|f(x_1) - f(x_2) - f'(x)(x_1 - x_2)\| \leq \epsilon \|x_1 - x_2\|$. $\square$

Now, we explain the name of the main object of our study, the so called essentially smooth maps. Note that Lebourg's concept of smoothness for a real-valued function [16, Definition (2.9)] extended to the vector-valued map $f$ states that $f$ is *smooth* at $x \in \Omega$ if it is locally Lipschitz on some neighborhood of $x$ in $\Omega$ and $\partial f(x)$ reduces to a singleton, i.e., if $f$ is Hadamard-type strictly differentiable at $x$. On the other hand, Borwein's concept of *essential smoothness* [6, p.68] (as it was recalled in [8, p.323]) restricted to a finite dimensional space setting leads to the following definition.

**Definition 2.3.** Assume that $f$ is locally Lipschitz on $\Omega$. We say that $f$ is *essentially smooth* on $\Omega$ if it is Fréchet-type strictly differentiable a.e. on $\Omega$.

Assume that $\Omega$ is an open subset of $\mathbb{R}^n$ and $f : \Omega \to \mathbb{R}^p$. In view of Remark 2.1 and Proposition 2.1, the expression "$f$ is essentially smooth locally Lipschitz on $\Omega$" implies that $f$ is strongly differentiable a.e. on $\Omega$. This fact plays an important role in the proof of our version of the "preparatory Sard theorem".

The term "essentially smooth locally Lipschitz map" has been used in [8, p.323]). Note that essentially smooth maps have been considered in a finite dimensional space setting in [13, 20, 21] and in a separable Banach space setting in [6, 7], see also Borwein-Woors papers in late 1990s [8, 9]. In [20, 21], the author used the name *a primal function* for a locally Lipschitz map $f : \mathbb{R}^n \to \mathbb{R}$ the Clarke subdifferential of which is single-valued almost everywhere (this name is due to the fact that a primal map defined on a connected open set can be recovered by its subdifferential).

Essentially smooth locally Lipschitz maps on an open subset of $\mathbb{R}^n$ form a broad linear space. In fact, as a consequence of [21, Theorem 7, p. 1008], functions which is representable as a difference of two closed convex functions belong to this family.

Let us provide some examples.

**Example 2.1.** (i) See [16, Proposition (1.9)]. Let $M$ be a measurable subset of $\mathbb{R}$, which intersects every nonempty open interval $I \subset \mathbb{R}$ in a set of positive measure $0 < \text{mes}(M \cap I)$, and let $g$ be the indicator function of $M$. Let $f : \mathbb{R} \to \mathbb{R}$ be the continuous function defined on $\mathbb{R}$ by

$$f(x) := \int_0^x g(t)dt.$$

The function $f$ is well-defined, strictly increasing, and locally Lipschitz on $\mathbb{R}$. The derivative of $f$ is almost everywhere 0 or 1 and each value is achieved on a dense subset of $\mathbb{R}$. Thus, $\partial f(x) = [0, 1]$ for all $x \in \mathbb{R}$ and $f$ is nowhere smooth.

(ii) See [19, p.324] Let $f : \mathbb{R} \to \mathbb{R}$ be the function defined by

$$f(x) = \begin{cases} x^2 \sin(1/x) & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$$

This function is locally Lipschitz on $\mathbb{R}$ and

$$\partial f(x) = \begin{cases} \{2x \sin(1/x) - \cos(1/x)\} & \text{if } x \neq 0 \\ [-1, 1] & \text{if } x = 0 \end{cases}$$

Thus, $f$ is smooth everywhere except at $x = 0$ (it has Fréchet derivative $f'(0) = 0$) and hence, it is essentially smooth on $\mathbb{R}$.

(iii) Let $C \subset \mathbb{R}$ be the Cantor (uncountable) set with positive Lebesgue measure. The function $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = \mathrm{d}(x, C)$ is essentially smooth locally Lipschitz on $\mathbb{R}$ [8, Example 4.1]. For each $i = 1, ..., n$, we define the set $\mathcal{C}_i$

$$\mathcal{C}_i := \{(x_1, ..., x_i, ..., x_n) \in \mathbb{R}^n \ : \ x_i \in C, x_j = 0 \text{ for } j \neq i\}$$

and a function $f_i : \mathbb{R}^n \to \mathbb{R}$

$$f_i(x) = \mathrm{d}(x, \mathcal{C}_i).$$

Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be a map defined by $f(x) = (f_1(x), ..., f_n(x))$. This map is essentially smooth locally Lipschitz on $\mathbb{R}^n$.

From now on, unless otherwise stated, let $f = (f_1, ..., f_p)$ be a map from an open set $\Omega$ of $\mathbb{R}^n$ into $\mathbb{R}^p$ with $p$ being an integer satisfying $1 \leq p \leq n$. Assume that $f$ is Fréchet differentiable at $x$ with the derivative $f'(x)$. Then $f'(x) = Jf(x)$, where $Jf(x)$ denotes the Jacobian of $f$ at $x$, i.e., the $p \times n$-matrix of partial derivatives of $f$ at $x$. We assume that each $f_i$ (and hence $f$) is Lipschitz near a given point $x$ of interest. It follows from the Rademacher theorem that $f$ is Fréchet differentiable (i.e., each $f_i$ is Fréchet differentiable) a.e. on any neighborhood of $x$ on which $f$ is Lipschitz. In the finite dimensional space setting, the Clarke subdifferential can also be defined as follows.

**Definition 2.4.** [12, Definition 2.6.1] Assume that $f$ is locally Lipschitz on $\Omega \subseteq \mathbb{R}^n$. The *Clarke subdifferential* of $f$ at $x \in \Omega$ denoted by $\partial f(x)$ is the convex hull of all $p \times n$-matrices obtained as the limit of a sequence of the form $Jf(x_i)$ where $x_i \to x$ and $Jf(x_i)$ is defined.

Some properties of the Clarke subdifferential are collected in the following.

**Proposition 2.2.** [12, Proposition 2.6.2] *Let $f$ be a locally Lipschitz map from an open set $\Omega$ of $\mathbb{R}^n$ into $\mathbb{R}^p$. Let $x \in \Omega$. Then the following assertions hold:*

(a) *$\partial f(x)$ is a nonempty compact convex subset of $\mathbb{R}^{p \times n}$.*

(b) *$\partial f$ is upper semicontinuous at $x$: for any $\epsilon > 0$ there is $\delta > 0$ such that, for all $y$ in $x + \delta\mathbb{B}$,*

$$\partial f(y) \subset \partial f(x) + \epsilon \mathbb{B}_{p \times n}.$$

It is immediate from Propositions 2.3.6 and 2.3.12 in [12] the following.

**Proposition 2.3.** *Let $g_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, ..., m$ be locally Lipschitz functions and $g : \mathbb{R}^n \to \mathbb{R}$ be a function defined by*

$$g(x) = \max\{g_i(x), i = 1, ..., m\}.$$

*Let $x \in \mathbb{R}^n$ and denote by $I(x)$ the set of active indexes, i.e., $I(x) := \{i \in \{1, ..., m\}, g_i(x) = g(x)\}$. Assume that each function $g_i$ ($i = 1, ..., m$) is either strictly differentiable at $x$ or convex. Then*

$$\partial g(x) = \text{co}\{\partial g_i(x), i \in I(x)\}.$$

We recall a property of a locally Lipschitz map to be used later. Lemma 4.1 [19] states that a globally Lipschitz map from $\Omega \subseteq \mathbb{R}^n$ to $\mathbb{R}^n$ maps a subset of $\Omega$ with Lebesgue measure zero into a set with Lebesgue measure zero. Using this lemma, one can prove a similar result for the locally Lipschitz case.

**Proposition 2.4.** *Let $f$ be a locally Lipschitz map from an open set $\Omega$ of $\mathbb{R}^n$ into $\mathbb{R}^n$. Suppose that $U \subseteq \Omega$ is a set with Lebesgue measure zero. Then*

$$\text{mes} f(U) = 0.$$

*Proof.* For $x \in U$, let $r(x)$ be a positive scalar such that $f$ is locally Lipschitz on the ball $\mathbb{B}(x, r(x))$. Without loss of generality, we may assume that $U$ is contained in the unit cube $[0, 1]^n$. For $N = 1, 2, ...$, we divide this unit cube into $N^n$ closed cubes of length $1/N$ and we represent the unit cube as a countable union of these cubes $C_j$

$$[0, 1]^n = \cup_j C_j.$$

For each $x \in U$, there exists $j$ such that $C_j \subset \mathbb{B}(x, r(x))$ and hence, $f$ is (globally) Lipschitz on $C_j$. Thus, one can find $I \subseteq \{1, 2, ...\}$ such that

$$U = \cup_{j \in I}(U \cap \mathring{C}_j),$$

where $\mathring{C}_j$ is the interior of $C_j$. Note that $\text{mes}(U \cap \mathring{C}_j) = 0$. Lemma 4.1 [19] implies that $\text{mes}(f(U \cap \mathring{C}_j)) = 0$. Therefore, $\text{mes} f(U) = 0$. □

Next, we recall some concepts of critical points. We will restrict ourselves to the case when $f$ is a map from an open set $\Omega$ of $\mathbb{R}^n$ into $\mathbb{R}^n$. In [22], Rader defined the set of critical points of $f$ (here denoted by $\text{Cri}_1 f$) in terms of Jacobian as follows

$$\text{Cri}_1 f := \{x \in \Omega \ : \ Jf(x) \text{ does not exist or } Jf(x) \text{ exists and } \det Jf(x) = 0\}.$$

In the case when $f$ is locally Lipschitz on $\Omega$ of $\mathbb{R}^n$, Clarke called a point $x \in \Omega$ *non-singular* if any $A \in \partial f(x)$ has maximal rank $n$ [12]. Hence, the set $\text{Cri} f$ of the (Clarke) *critical points* of $f$ can be defined as

$$\begin{aligned} \text{Cri} f := & \ \{x \in \Omega \ : \ \exists A \in \partial f(x), \ \text{rank} A < n\} \\ = & \ \{x \in \Omega \ : \ \exists A \in \partial f(x), \ \det A = 0\}. \end{aligned}$$

The set Cri$f$ has been considered in [4, 13]. A value $r \in f(\Omega)$ is a (Clarke) *critical value* of $f$ if there exists an element $x \in \text{Cri}f$ such that $f(x) = r$. Note that for the map $f$ in Example 2.1(i), $\text{Cri}_1 f$ is a proper subset of Cri$f$ and, moreover, $\text{mes}(\text{Cri}f \setminus \text{Cri}_1 f) > 0$.

We recall two known versions of the Sard theorem for these types of critical points.

**Theorem 2.1.** [13] *Let $f$ be an essential smooth locally Lipschitz map from an open set $\Omega$ of $\mathbb{R}^n$ into $\mathbb{R}^n$. Then*

$$\text{mes} f(\text{Cri}f) = 0.$$

**Theorem 2.2.** [22, Lemma 2] *Let $f$ be an almost everywhere differentiable map from an open set $\Omega$ of $\mathbb{R}^n$ into $\mathbb{R}^n$ and has the property that for any subset $U \subset \Omega$ with the Lebesgue measure zero, $f(B)$ has the Lebesgue measure zero. Then*

$$\text{mes} f(\text{Cri}_1 f) = 0.$$

See also [14, Theorem 3.1] for a detailed proof of Theorem 2.2 in the case of locally Lipschitz maps.

**Remark 2.2.** (i) Let $f$ be the function in Example 2.1 (i). Theorem 2.2 implies that $\text{mes} f(\text{Cri}_1 f) = 0$. Meanwhile, since $\text{Cri}f = \mathbb{R}$ and $f$ is strictly increasing continuous, we have $\text{mes} f(\text{Cri}f) > 0$. This fact shows that the assumption on the essential smoothness in Theorem 2.1 cannot be dropped. Theorem 2.1 can be applied for instance to the map in Example 2.1 (iii).

(ii) Applying Proposition 2.4, one can derive Theorem 2.1 from Theorem 2.2.

## 3. A version of "preparatory Sard theorem" for essentially smooth maps

In this section, we establish a version of the "preparatory Sard theorem" of [3] in the case $n = p$ for essentially smooth Lipschitz maps. From now on, $\Omega$ is an open subset of $\mathbb{R}^n$.

Let $\Delta^m$ and inn$\Delta^m$ be the $m$-dimensional simplex in $\mathbb{R}^{m+1}$ and its algebraic interior

$$\Delta^m := \{\lambda = (\lambda_0, \lambda_1, ..., \lambda_m) \in \mathbb{R}^{m+1} \ : \ \lambda_i \geq 0, \sum_{i=0}^{m} \lambda_i = 1\}$$

and

$$\text{inn}\Delta^m := \{\lambda = (\lambda_0, \lambda_1, ..., \lambda_m) \in \mathbb{R}^{m+1} \ : \ \lambda_i > 0, \sum_{i=0}^{m} \lambda_i = 1\}.$$

Let $\phi_i : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}^n$, $i = 0, 1, ..., m$ be locally Lipschitz maps and

$$\Psi(\lambda, x) := \sum_{i=0}^{m} \lambda_i \phi_i(x), \quad (\lambda, x) \in \text{inn}\Delta^m \times \Omega.$$

Clearly, $\Psi$ is a locally Lipschitz map on inn$\Delta^m \times \Omega$. The following concept is motivated by the one of strongly critical points for a $C^1$-map $\Psi$ given in [3].

**Definition 3.1.** The set $\hat{\text{Cri}}\Psi$ of *strongly critical points* of $\Psi$ is defined by

$$(\lambda, x) \in \hat{\text{Cri}}\Psi \Longleftrightarrow \begin{cases} \phi_i(x) = \phi_0(x), \; i \in \{1, ..., m\} \\ \exists A \in \sum_{i=0}^{m} \lambda_i \partial\phi_i(x), \quad \text{rank}A < n. \end{cases}$$

A scalar $t \in \mathbb{R}$ is called a strongly critical value of $\Psi$ if there exists $(\lambda, x) \in \hat{\text{Cri}}\Psi$ such that $\Psi(\lambda, x) = t$.

Our main result reads as follows.

**Theorem 3.1.** *Given $m+1$ maps $\phi_i : \Omega \to \mathbb{R}^n$, $i = 0, 1, ..., m$, we set*

$$\begin{cases} \Psi : \text{inn}\Delta^m \times \Omega \to \mathbb{R}^n \\ \Psi(\lambda, x) := \sum_{i=0}^{m} \lambda_i \phi_i(x) \end{cases}$$

*for all $\lambda = (\lambda_0, \lambda_1, ..., \lambda_m) \in \text{inn}\Delta^m$ and $x \in \Omega$. Assume that the maps $\phi_i$ $(i = 0, 1, ..., m)$ are essentially smooth locally Lipschitz on $\Omega$. Then*

$$\text{mes}\Psi(\hat{\text{Cri}}\Psi) = 0.$$

Note that Theorem 3.1 reduces to Theorem 2.1 when $m = 0$. Let us establish an auxiliary result.

**Lemma 3.1.** *The set $\hat{\text{Cri}}\Psi$ is closed in $\text{inn}\Delta^m \times \Omega$.*

*Proof.* Let $\{(\lambda^j, x^j)\}_{j=1}^{\infty}$ be a sequence and $(\lambda, x) \in \text{inn}\Delta^m \times \Omega$ be a vector such that $(\lambda^j, x^j) \in \hat{\text{Cri}}\Psi$ for $j = 1, 2, ...$ and $\lim_{j\to\infty}(\lambda^j, x^j) = (\lambda, x)$. We have to show that $(\lambda, x) \in \hat{\text{Cri}}\Psi$.

Denote $\lambda = (\lambda_0, \lambda_1, ..., \lambda_m)$ and $\lambda^j = (\lambda_0^j, \lambda_1^j, ..., \lambda_m^j)$. Observe that $\lambda \in \text{inn}\Delta^m$. Since $(\lambda^j, x^j) \in \hat{\text{Cri}}\Psi$, for all $i \in \{1, ..., m\}$ we have $\phi_i(x^j) = \phi_0(x^j)$ and the continuity of the functions $\phi_i$ implies $\phi_i(x) = \phi_0(x)$ for all $i$.

Let $M := \sup\{\|v\| \;:\; v \in \partial\phi_i(x), \; i = 0, ..., m\}$. Proposition 2.2 (a) implies that $M < +\infty$. If $M = 0$, then $(\lambda, x) \in \hat{\text{Cri}}\Psi$ and the assertion follows. Assume that $M > 0$. For $j = 1, 2, ...$, set $\epsilon_j := 1/2^j$. Taking a subsequence if necessary, we may assume that

$$|\lambda_i^j - \lambda_i| \leq \frac{\epsilon_j}{3M(m+1)}$$

and since the subdifferential map is upper semicontinuous (Proposition 2.2 (b)), we may assume that

$$\partial\phi_i(x^j) \subset \partial\phi_i(x) + \frac{\epsilon_j}{3}\mathbb{B}$$

for $j = 1, 2, ...$ and $i = 0, ..., m$. Therefore, we have

$$\begin{aligned} \textstyle\sum_{i=0}^{m} \lambda_i^j \partial\phi_i(x^j) \; & \subset \textstyle\sum_{i=0}^{m} \lambda_i^j \partial\phi_i(x) + \frac{\epsilon_j}{3} \sum_{i=0}^{m} \lambda_i^j \mathbb{B} \\ & \subset \textstyle\sum_{i=0}^{m} \lambda_i \partial\phi_i(x) + \sum_{i=0}^{m} (\lambda_i^j - \lambda_i)\partial\phi_i(x) + \frac{\epsilon_j}{3}\mathbb{B} \\ & \subset \textstyle\sum_{i=0}^{m} \lambda_i \partial\phi_i(x) + \sum_{i=0}^{m} \frac{\epsilon_j}{3M(m+1)} M\mathbb{B} + \frac{\epsilon_j}{3}\mathbb{B} \\ & \subset \textstyle\sum_{i=0}^{m} \lambda_i \partial\phi_i(x) + \frac{2\epsilon_j}{3}\mathbb{B} \end{aligned} \quad . \tag{3}$$

Since $(\lambda^j, x^j) \in \hat{\text{Cri}}\Psi$, for each $j = 1, 2, ...$ one can find $A_j \in \sum_{i=0}^m \lambda_i^j \partial\phi_i(x^j)$ such that $\text{rank}A_j < n$. By (17), for each $j = 1, 2, ...$ one can find $\tilde{A}_j \in \sum_{i=0}^m \lambda_i \partial\phi_i(x)$ such that $\|A_j - \tilde{A}_j\| \leq \frac{2\epsilon_j}{3}$. Proposition 2.2 (a) implies that the set $\sum_{i=0}^m \lambda_i \partial\phi_i(x)$ is compact. Without lost of generality we may assume that there exists $A \in \sum_{i=0}^m \lambda_i \partial\phi_i(x)$ such that $\|A - \tilde{A}_j\| \leq \frac{\epsilon_j}{3}$. Then $\|A_j - A\| \leq \epsilon_j$ and the lower semicontinuity property of the rank function implies

$$\text{rank}A \leq \liminf_{j\to\infty} \text{rank}A_j < n.$$

Thus, $(\lambda, x) \in \hat{\text{Cri}}\Psi$. □

We are ready now for the proof of Theorem 3.1.

*Proof.* Without loss of generality, we may assume that $\Omega$ is bounded. Denote

$$D := \{x \in \Omega \ : \ \phi_i \text{ is smooth at } x \text{ for all } i = 0, 1, ..., m\}.$$

Since $\phi_i$ is smooth a.e. on $\Omega$ for all $i = 0, 1, ..., m$, the set $\Omega \setminus D$ has Lebesgue measure zero and so is the set $\text{inn}\Delta^m \times (\Omega \setminus D)$. Therefore, the set $\hat{\text{Cri}}\Psi \cap (\text{inn}\Delta^m \times (\Omega \setminus D))$ also has Lebesgue measure zero. Observe that the map $\Psi$ is locally Lipschitz on $\text{inn}\Delta^m \times \Omega$. Therefore, Proposition 2.4 implies that mes $\Psi(\hat{\text{Cri}}\Psi \cap (\text{inn}\Delta^m \times (\Omega \setminus D))) = 0$. Hence, it remains to prove that

$$\text{mes } \Psi(\hat{\text{Cri}}\Psi \cap (\text{inn}\Delta^m \times D)) = 0.$$

For simplicity, we denote

$$K := \hat{\text{Cri}}\Psi \cap (\text{inn}\Delta^m \times D).$$

By Lemma 3.1, the set $\hat{\text{Cri}}\Psi$ is measurable, and hence, so is the set $K$. Moreover, since $K$ is bounded, there exists an increasing sequence of compacts subsets of $K$, say $\{\Gamma_i\}$ $(i = 1, 2, ..)$ such that $\text{mes}(K \setminus \cup_{i=1}^\infty \Gamma_i) = 0$. Again, Proposition 2.4 implies that mes $\Psi((K \setminus \cup_{i=1}^\infty \Gamma_i)) = 0$. Hence, our proof will be achieved if we prove that for any nonempty compact subset $\Gamma$ of $K$ it holds

$$\text{mes } \Psi(\Gamma) = 0.$$

Let $\epsilon > 0$ be an arbitrary scalar. To any $\lambda := (\lambda_0, ..., \lambda_m) \in \text{inn}\Delta^m$ we associate a map $\Phi_\lambda$ defined by

$$\Phi_\lambda : u \in \Omega \to \Phi_\lambda(u) = \sum_{j=0}^m \lambda_i \phi_i(u).$$

It is clear that the map $\Phi_\lambda$ is essentially smoth locally Lipschitz on $\Omega$.

Now, let $(\lambda, x) \in \Gamma \subset K$. Then $x \in D$ and $\phi_i$ is smooth at $x$ for all $i = 0, 1, ..., m$ and so is the map $\Phi_\lambda$. Thus, $\partial\Phi_\lambda(x) = \{A_{x,\lambda}\}$, where

$$A_{x,\lambda} = \sum_{i=0}^m \lambda_i \phi_i'(x) \tag{4}$$

and $\phi_i'(x)$ $(i = 0, 1, ..., m)$ is the strict derivative of $\phi_i$ at $x$. Since the map $\Phi_\lambda$ is strictly differentiable at $x$, it is strongly differentiable at that point by Proposition 2.1. Hence, for the given $\epsilon > 0$ there exists $\delta_{x,\lambda} > 0$ such that $\mathbb{B}(x, \delta_{x,\lambda}) \subset \Omega$ and

$$\|\Phi_\lambda(u_1) - \Phi_\lambda(u_2) - A_{x,\lambda}(u_1 - u_2)\| \le \epsilon\|u_1 - u_2\|, \quad \forall\ u_1, u_2 \in \mathbb{B}(x, \delta_{x,\lambda}). \tag{5}$$

Note that $\lambda \in \text{inn}\Delta^m$, we can choose $\delta_{x,\lambda}$ such that $\mathbb{B}(\lambda, \delta_{x,\lambda}) \subset \text{inn}\Delta^m$.

Thus, for any $(\lambda, x) \in \Gamma \subset K$ and any $\epsilon > 0$ there exists $\delta_{x,\lambda} > 0$ such that

$$(\lambda, x) \in \mathbb{B}(\lambda, \delta_{x,\lambda}) \times \mathbb{B}(x, \delta_{x,\lambda}) \subset \text{inn}\Delta^m \times \Omega$$

and (5) holds. Since $\Gamma$ is a compact set, there exist a finite number $l$ of pairs denoted by $(\lambda^i, x_i)$, $i = 1, ..., l$ in $\Gamma$ such that

$$\Gamma \subset \cup_{i=1}^l \mathbb{B}(\lambda^i, \delta_{x_i,\lambda^i}) \times \mathbb{B}(x_i, \frac{1}{2}\delta_{x_i,\lambda^i}). \tag{6}$$

Let $\mathcal{C} \subset \mathbb{R}^n$ be a cube with side of length $a$ such that

$$\Gamma \subset \text{inn }\Delta^m \times \mathcal{C}$$

and $N$ be a positive integer satisfying

$$N > \frac{3a\sqrt{n}}{\min\{\delta_{x_1,\lambda^1}, ..., \delta_{x_l,\lambda^l}\}}.$$

Divide $\mathcal{C}$ into $N^n$ small cubes with side of length $\frac{a}{N}$. Assume that $C$ is one of these small cubes such that

$$(\text{inn }\Delta^m \times C) \cap \Gamma \neq \emptyset$$

and $b$ is the center of $C$. We claim that there exists $\lambda^i \in \{\lambda^1, ..., \lambda^l\})$ such that for any $(\beta, y) \in (\text{inn }\Delta^m \times C) \cap \Gamma$, $\Psi(\beta, y)$ is contained in the closed parallelepiped $\mathcal{P}$ with the center $\Phi_{\lambda^i}(b)$ and

$$\text{mes}\mathcal{P} \le 2^n(\alpha + \epsilon)^{n-1}\epsilon\sqrt{n^n}\frac{a^n}{N^n}, \tag{7}$$

where $\alpha$ is positive scalar depending on $\Gamma$. Indeed, suppose that there is a pair $(\bar{\lambda}, \bar{x}) \in \Gamma$ such that $(\bar{\lambda}, \bar{x}) \in \text{inn }\Delta^m \times C$. By (6), one can find an index $i \in \{1, ..., l\}$ and the corresponding pair $(\lambda^i, x_i)$ such that $\bar{x} \in \mathbb{B}(x_i, \frac{1}{2}\delta_{x_i,\lambda^i})$. Observe that for any pair $(z_1, z_2) \in C \times C$ we have

$$\|z_1 - z_2\| \le \sqrt{n}\frac{a}{N} \le \frac{1}{3}\min\{\delta_{x_1,\lambda^1}, ..., \delta_{x_l,\lambda^l}\}.$$

Let $(\beta, y)$ be an arbitrary pair satisfying $(\beta, y) \in (\text{inn }\Delta^m \times C) \cap \Gamma$. Then we have

$$\|y - x_i\| \le \|y - \bar{x}\| + \|\bar{x} - x_i\| \le \frac{1}{3}\min\{\delta_{x_1,\lambda^1}, ..., \delta_{x_l,\lambda^l}\} + \frac{1}{2}\delta_{x_i,\lambda^i} < \delta_{x_i,\lambda^i}$$

which means that $C \subset \mathbb{B}(x_i, \delta_{x_i,\lambda^i})$. Since $y, b \in C$, we have $\|y - b\| \le \sqrt{n}\frac{a}{N}$ and $y, b \in \mathbb{B}(x_i, \delta_{x_i,\lambda^i})$. Then we deduce from (5) that

$$\|\Phi_{\lambda^i}(y) - \Phi_{\lambda^i}(b) - A_{x_i,\lambda^i}(y - b)\| \le \epsilon\|y - b\| \le \epsilon\sqrt{n}\frac{a}{N},$$

where $A_{x_i,\lambda^i}$ is defined by (4). Since $(\beta, y) \in (\text{inn } \Delta^m \times C) \cap \Gamma$, we have $\phi_j(y) = \phi_0(y)$ for $j = 0, ..., m$ and $\sum_{k=0}^m \beta_k = 1$. Further, since $\lambda^i \in \text{inn } \Delta^m$, we have $\lambda^i = (\lambda^i_0, ..., \lambda^i_m)$, where $\lambda^i_k > 0$ for $k = 0, ..., m$ and with $\sum_{k=0}^m \lambda^i_k = 1$. Therefore,

$$\Psi(\beta, y) = \sum_{k=0}^m \beta_k \phi_k(y) = \sum_{k=0}^m \beta_k \phi_0(y) = \sum_{k=0}^m \lambda^i_k \phi_0(y) = \sum_{k=0}^m \lambda^i_k \phi_k(y) = \Phi_{\lambda^i}(y).$$

Hence,

$$\|\Psi(\beta, y) - \Phi_{\lambda^i}(b) - A_{x_i,\lambda^i}(y - b)\| \le \epsilon \|y - b\| \le \epsilon \sqrt{n} \frac{a}{N}. \tag{8}$$

To simplify the notations, put

$$r(x_i, y, b) = \Psi(\beta, y) - \Phi_{\lambda^i}(b) - A_{x_i,\lambda^i}(y - b).$$

Since $(\lambda^i, x_i) \in \Gamma \subset K \subset \hat{\text{Cri}}\Psi$, we have $\text{rank} A_{x_i,\lambda^i} < n$. Let $e_1, ..., e_n$ be an orthonormal basis of $\mathbb{R}^n$ such that $A_{x_i,\lambda^i}(\mathbb{R}^n)$ is contained in the vector space generated by $e_1, ..., e_{n-1}$. We have

$$A_{x_i,\lambda^i}(y - b) = \sum_{k=1}^{n-1} \rho_k e_k$$

with

$$|\rho_k| = |\langle A_{x_i,\lambda^i}(y - b), e_k\rangle| \le \|A_{x_i,\lambda^i}(y - b)\| \le \alpha \|y - b\| \le \alpha \sqrt{n} \frac{a}{N},$$

where

$$\alpha := \max\{\|v\|_{\mathbb{R}^{n\times n}} \ : \ v \in \sum_{j=0}^m \xi_j \partial\phi_j(u), \ ((\xi_0, ..., \xi_m), u) \in \Gamma)\}.$$

Since $\partial\phi_j$ $(j = 0, ..., m)$ is upper semicontinuous from $\Omega$ with nonempty convex compact values in $\mathbb{R}^{n\times n}$, one can prove that $\alpha$ is finite and depends only on $\Gamma$. Let $\eta_k$ $(k = 1, ..., n)$ be the coordinate of $r(x_i, y, b)$ in the orthonormal basis $e_1, ..., e_n$, i.e.,

$$r(x_i, y, b) := \sum_{k=1}^n \eta_k e_k.$$

It follows from (8) that

$$|\eta_k| = |\langle r(x_i, y, b), e_k\rangle| \le \|r(x_i, y, b)\| \le \epsilon \|y - b\| \le \epsilon \sqrt{n} \frac{a}{N}$$

for all $k = 1, 2, ...n$. Further, let $\Psi(\beta, y) - \Phi_{\lambda^i}(b) := \sum_{k=1}^n \nu_k e_k$. Since $\Psi(\beta, y) - \Phi_{\lambda^i}(b) = A_{x_i,\lambda^i}(y - b) + r(x_i, y, b)$, we get

$$\Psi(\beta, y) - \Phi_{\lambda^i}(b) = \sum_{k=1}^n \nu_k e_k = \sum_{k=1}^{n-1} \rho_k e_k + \sum_{k=1}^n \eta_k e_k.$$

Then the following estimates hold

$$\begin{aligned}\nu_k &= \rho_k + \eta_k,\ k = 1, 2, ..., n-1\\ \nu_n &= \eta_n\\ |\nu_k| &\le |\rho_k| + |\eta_k| \le (\alpha+\epsilon)\sqrt{n}\tfrac{a}{N},\ k = 1, 2, ..., n-1\\ |\nu_n| &= |\eta_n| \le \epsilon\sqrt{n}\tfrac{a}{N}.\end{aligned}$$

The above inequalities show that $\Psi(\beta, y)$ is contained in the closed parallelepiped $\mathcal{P}$ with the center $\Phi_{\lambda^i}(b)$ and with sides of length $c_k$, where

$$\begin{aligned}c_k &\le 2(\alpha+\epsilon)\sqrt{n}\tfrac{a}{N},\ k = 1, 2, ..., n-1\\ c_n &\le 2\epsilon\sqrt{n}\tfrac{a}{N}.\end{aligned}$$

Observe that the measure in $\mathbb{R}^n$ does not depend on the choice of the orthonormal bases. The inequality (7) follows.

Since the number of sets of the form $\text{inn}\Delta^m \times C$, which has a nonempty intersection with the set $\Gamma$, is less than or equal to $N^n$, the inequality (7) implies

$$\text{mes}^*\Psi(\Gamma)) \le N^n 2^n (\alpha+\epsilon)^{n-1}(\sqrt{n}\frac{a}{N})^n\epsilon = [2^n(\alpha+\epsilon)^{n-1}a^n\sqrt{n^n}]\epsilon.$$

As $\epsilon > 0$ is arbitrary, it follows that $\text{mes}\Psi(\Gamma) = 0$. □

**Example 3.1.** Theorem 4.1 can be applied for instance to the maps $\phi_0, \phi_1 : \mathbb{R}^2 \to \mathbb{R}^2$ given by $\phi_0 = f$, where $f$ is the map in Example 2.1(iii), and $\phi_1(x, y) = (|x|, |y|)$. Note that these maps are essentially smooth locally Lipschitz but are not $C^1$ on $\mathbb{R}^2$.

## 4. A version of the Sard theorem for a continuous selection

In this section, we establish a version of the Sard theorem for a continuous selection of a finite number of essentially smooth locally Lipschitz maps. Throughout this section, $I$ is a finite set of indexes and $\Omega$ is a nonempty open subset of $\mathbb{R}^n$. We say that a map $h : \Omega \to \mathbb{R}$ is a selection of maps $f_i : \Omega \to \mathbb{R}$ $(i \in I)$ on $\Omega$ if

$$h(x) \in \{f_i(x) \ : \ i \in I\} \quad x \in \Omega.$$

The main result of this section reads as follows.

**Theorem 4.1.** *Let $f_i : \mathbb{R}^n \to \mathbb{R}^n$ $(i \in I)$ be essentially smooth locally Lipschitz maps. Assume that $h : \mathbb{R}^n \to \mathbb{R}^n$ is a continuous selection of the maps $f_i$ $(i \in I)$ on $\mathbb{R}^n$. Then $h$ is locally Lipschitz and*

$$\text{mes } h(\text{Cri } h) = 0.$$

To prove Theorem 4.1, we need some auxiliary results.

**Proposition 4.1.** *Let $f_i : \Omega \to \mathbb{R}$ $(i \in I)$ be locally Lipschitz map and $h : \Omega \to \mathbb{R}$ be a selection of these maps on $\Omega$. If $h$ is continuous on $\Omega$, then it is locally Lipschitz on this set.*

*Proof.* For any $x \in \Omega$, let $I(x)$ be the set of active index at $x$, i.e.,

$$I(x) := \{i \in I \ : \ f_i(x) = h(x)\}.$$

Note that $I(x) \neq \emptyset$ for all $x \in \Omega$. Let $\bar{x} \in \Omega$. Let $L$ be a positive scalar and $U := B(\bar{x}, r) \subset \Omega$ be a ball centered of $\bar{x}$ such that

$$|f_i(z) - f_i(y)| \leq L\|z - y\|, \quad \forall y, z \in U, \ \forall i \in I.$$

Our aim is to show that for any $\bar{y}, \bar{z} \in U$ it holds

$$|h(\bar{z}) - h(\bar{y})| \leq L\|\bar{z} - \bar{y}\|.$$

Let $S := [\bar{y}, \bar{z}] := \{\bar{y} + t(\bar{z} - \bar{y}) \ : \ t \in [0, 1]\}$. Note that $S \subset U$. Let $x \in S$ be an arbitrary point. For any $j \notin I(x)$, we have $|f_j(x) - h(x)| > 0$. Since the functions $h$ and $f_i$ $(i \in I)$ are continuous and $I$ is finite, we can find $\delta_x > 0$ such that for any $u \in \mathbb{B}(x, \delta_x) \cap S =: \mathbb{B}_S(x, \delta_x)$ one has $|f_j(u) - h(u)| > 0$ for all $j \notin I(x)$. It follows that

$$I(u) \subseteq I(x), \qquad \forall u \in \mathbb{B}_S(x, \delta_x). \tag{9}$$

Since $S \subset \cup_{x\in S}\mathbb{B}_S(x, \delta_x)$, we can choose $x_1, ..., x_k \in S$ such that: (i) $\bar{y} = x_1 < x_2 < ... < x_k = \bar{z}$ (ii) $S \subset \cup_{i=1}^k \mathbb{B}_S(x_i, \delta_{x_i})$ (iii) $\mathbb{B}_S(x_i, \delta_{x_i}) \cap \mathbb{B}_S(x_{i+1}, \delta_{x_{i+1}}) \neq \emptyset$ for $k = 1, ..., k-1$, where $u_1 < u_2$ for $u_1, u_1 \in S$ mean $u_2 - u_1 = t(\bar{z} - \bar{y})$ with $t > 0$. By (9) and (iii), one can find $i_j \in I(x_j) \cap I(x_{j+1})$ for $j = 1, ..., k-1$. By the definition of the active set $I(x)$, we get $h(x_j) = f_{i_j}(x_j)$ and $h(x_{j+1}) = f_{i_j}(x_{j+1})$ for $j = 1, ..., k-1$. Hence,

$$\begin{aligned}
|h(\bar{y}) - h(\bar{z})| \ &\leq |h(x_1) - h(x_2)| + |h(x_2) - h(x_3)| + ... + |h(x_{k-1}) - h(x_k)| \\
&= |f_{i_1}(x_1) - f_{i_1}(x_2)| + |f_{i_2}(x_2) - f_{i_2}(x_3)| + ... + |f_{i_{k-1}}(x_{k-1}) - f_{i_{k-1}}(x_k)| \\
&\leq L\|x_1 - x_2\| + L\|x_2 - x_3\| + ... + L\|x_{k-1} - x_k\| \\
&= L\|x_1 - x_k\| = L\|\bar{y} - \bar{z}\|
\end{aligned}$$

(recall that $x_1, ..., x_k \in S = [\bar{y}, \bar{z}]$). □

Remark that Proposition 4.1 is a special case of [3, Proposition 2], which has been established for the general case with $I$ being a countable compact set. We presented here another simple proof based on the finite character of $I$. Moreover, the inclusion (9) in this proof will be used to prove the next proposition.

**Proposition 4.2.** *Let $f_i : \mathbb{R}^n \to \mathbb{R}$ $(i \in I)$ be essentially smooth locally Lipschitz maps and $h : \mathbb{R}^n \to \mathbb{R}$ be a continuous selections of these maps. Then (h is locally Lipschitz map and) for every point $x \in \mathcal{D}$ one has*

$$h'(x) \in \text{co}\{f_i'(x) \ : \ i \in I(x)\}.$$

*Here,*

$$\mathcal{D} := \{x \in \mathbb{R}^n : h \ \text{ is Fréchet differentiable and } f_i \text{ is smooth, } \forall i \in I\}$$

*and* $I(x) := \{i \in I : h(x) = f_i(x)\}$.

This proposition is an extension of [3, Proposition 4] (the $C^1$-differentiability is relaxed to essential smoothness) in the special case when the family of parameter $I$ is finite.

*Proof.* Our proof is similar to the one of [3, Proposition 4].

Let us begin with the case $n = 1$. Suppose to the contrary that for some $\bar{x} \in \mathcal{D}$ one has $h'(\bar{x}) \notin \text{co}\{f_i'(\bar{x}) \ : \ i \in I(\bar{x})\}$. Without loss of generality, we may assume that $\bar{x} = 0$, $h(0) = h'(0) = 0$ and $f_i'(0) > 1$ for all $i \in I(0)$. Applying the argument used to prove (9), we find $\delta > 0$ such that

$$I(x) \subseteq I(0) \quad \forall x \in [0, \delta]. \tag{10}$$

Let $\epsilon \in ]0, 0.3[$ be an arbitrary scalar. Since the functions $f_i$ $(i \in I(0))$ are smooth at $\bar{x} = 0$, we may assume that

$$|\frac{f_i(z + ty) - f_i(z)}{t} - f_i'(0)(y)| \leq \epsilon$$

for all $i \in I(0)$, $t \in ]0, \delta]$, $z \in [0, \delta]$ and $y \in \{1, -1\}$. Taking $y = 1$ and $z = 0$, we get

$$|\frac{f_i(t) - f_i(0)}{t} - f_i'(0)| \leq \epsilon, \quad \forall t \in ]0, \delta], \ \forall i \in I(0).$$

Since $f_i(0) = h(0) = 0$ and $f_i'(0) > 1$ for any $i \in I(0)$ by assumptions, we get

$$1 - \epsilon < -\epsilon + f_i'(0) \leq \frac{f_i(t)}{t}, \quad \forall t \in ]0, \delta], \ \forall i \in I(0),$$

which together with (10) implies

$$1 - \epsilon < \frac{f_i(t)}{t}, \quad \forall t \in ]0, \delta], \ \forall i \in I(t). \tag{11}$$

On the other hand, since $h$ is Fréchet differentiable at $\bar{x} = 0$, we may assume that

$$|\frac{h(t) - h(0)}{t} - h'(0)| \leq \epsilon, \quad \forall t \in ]0, \delta].$$

Let $\bar{t} \in ]0, \delta]$. Since $h(0) = 0$ and $h'(0) = 0$, it follows that $\frac{h(\bar{t})}{\bar{t}} \leq \epsilon$. Take $i \in I(\bar{t})$. Then, we have $h(\bar{t}) = f_{\bar{i}}(\bar{t})$ and by (11), we get

$$1 - \epsilon < \frac{f_{\bar{i}}(\bar{t})}{\bar{t}} = \frac{h(\bar{t})}{\bar{t}} \leq \epsilon,$$

which is a contradiction because $\epsilon < 0.3$.

Next, we consider the case $n > 1$. Denote

$$A(x) := \text{co}\{f_i'(x) \ : \ i \in I(x)\}.$$

Suppose to contrary that for some $\bar{x} \in \mathcal{D}$, we have $h'(\bar{x}) \notin A(\bar{x})$. We may assume that $h'(\bar{x}) = 0$ because otherwise we could replace $h(x)$ by $h(x) - \langle h'(\bar{x}), x\rangle$ and $f_i(x)$ by $f_i(x) - \langle h'(\bar{x}), x\rangle$. The Hahn-Banach theorem implies the existence of $v \in \mathbb{R}^n$ such that

$$0 = \langle h'(\bar{x}), v\rangle < 1 < \min_{i \in I(\bar{x})} \langle f_i'(\bar{x}), v\rangle. \tag{12}$$

Let $g$ and $g_i$ $(i \in I(\bar{x}))$ be functions from $\mathbb{R}$ to $\mathbb{R}$ given by $g(t) := h(\bar{x} + tv)$ and $g_i(t) := f_i(\bar{x} + tv)$. Observe that $g'(0) = \langle h'(\bar{x}), v\rangle$ and $g_i'(0) = \langle f_i'(\bar{x}), v\rangle$. Being in the case $n = 1$ with the maps $g$ and $g_i$, we get $g'(0) \in \mathrm{co}\{g_i'(0) \ : \ i \in I(\bar{x})\}$, contradicting (12). □

We are ready to prove Theorem 4.1.

*Proof.* We will use the following notations. Let $h = (h^1, ..., h^n)$ and $f_i = (f_i^1, ..., f_i^n)$ for $i \in I$, where $h^1, ..., h^n$, $f_i^1, ..., f_i^n$ $(i \in I)$ are maps from $\mathbb{R}^n$ to $\mathbb{R}$. The maps $f_i^j : \mathbb{R}^n \to \mathbb{R}$ $(j = 1, ..., n,\ i \in I)$ are essentially smooth locally Lipschitz and for each $j$, one has

$$h^j(x) \in \{f_i^j(x) \ : \ i \in I\}, \ \forall x \in \mathbb{R}^n.$$

By Proposition 4.2, $h$ is locally Lipschitz on $\mathbb{R}^n$. Let $\bar{x} \in \mathrm{Cri}h \cap \mathcal{D}$ be an arbitrary point, where $\mathcal{D}$ is the set defined in Proposition 4.2. Then the Fréchet derivative $Dh(\bar{x})$ exists and

$$\mathrm{rank} Dh(\bar{x}) < n. \tag{13}$$

Clearly, $Dh(\bar{x}) = (Dh^1(\bar{x}), ..., Dh^n(\bar{x}))$. Let $j \in \{1, .., n\}$ be an arbitrary index. Proposition 4.2 implies that

$$Dh^j(\bar{x}) = \mathrm{co}\{Df_i^j(\bar{x}), \ i \in I^j(\bar{x})\},$$

where $I^j(\bar{x}) := \{i \in I \ : \ h^j(\bar{x}) = f_i^j(\bar{x})\}$. Then we can find a subset $\tilde{I}^j(\bar{x})$ of $I^j(\bar{x})$ and scalars $\bar{\lambda}_i^j \in ]0, 1[$, $i \in \tilde{I}^j(\bar{x})$ such that

$$Dh^j(\bar{x}) = \sum_{i \in \tilde{I}^j(\bar{x})} \bar{\lambda}_i^j Df_i^j(\bar{x}) \tag{14}$$

and

$$\sum_{i \in \tilde{I}^j(\bar{x})} \bar{\lambda}_i^j = 1. \tag{15}$$

Since $\bar{x}$ is a critical point of $h$, the relations (13) and (14) imply

$$\mathrm{rank}\{\sum_{i \in \tilde{I}^1(\bar{x})} \lambda_i^1 Df_i^1(\bar{x}), ..., \sum_{i \in \tilde{I}^n(\bar{x})} \lambda_i^n Df_i^n(\bar{x})\} < n. \tag{16}$$

Set $\tilde{I}(\bar{x}) := \tilde{I}^1(\bar{x}) \times ... \times \tilde{I}^n(\bar{x})$. Let $\tilde{i} := (i_1, ..., i_n) \in \tilde{I}(\bar{x})$ be an arbitrary $n$-tuple. Let $\phi_{\tilde{i}} : \mathbb{R}^n \to \mathbb{R}^n$ be a map given by

$$\phi_{\tilde{i}}(x) = (f_{i_1}^1(x), ..., f_{i_n}^n(x))$$

and let

$$\bar\lambda_{\tilde i} := \bar\lambda^1_{i_1}...\bar\lambda^n_{i_n}.$$

Note that the map $\phi_{\tilde i}$ is essentially smooth locally Lipschitz. Moreover, since the equality (15) holds for all $j = 1, ..., n$, we have

$$\begin{aligned}\textstyle\sum_{\tilde i\in\tilde I(\bar x)}\bar\lambda_{\tilde i} &= \textstyle\sum_{i_1\in\tilde I^1(\bar x),...,i_n\in\tilde I^n(\bar x)}\bar\lambda^1_{i_1}...\bar\lambda^n_{i_n} = \sum_{i_2\in\tilde I^2(\bar x),...,i_n\in\tilde I^n(\bar x)}(\sum_{i\in\tilde I^1(\bar x)}\bar\lambda^1_i)\bar\lambda^2_{i_2}...\bar\lambda^n_{i_n}\\ &= \textstyle\sum_{i_2\in\tilde I^2(\bar x),...,i_n\in\tilde I^n(\bar x)}\bar\lambda^2_{i_2}...\bar\lambda^n_{i_n} = \sum_{i_3\in\tilde I^3(\bar x),...,i_n\in\tilde I^n(\bar x)}(\sum_{i\in\tilde I^2(\bar x)}\bar\lambda^2_i)\bar\lambda^3_{i_3}...\bar\lambda^n_{i_n}\\ &= \textstyle\sum_{i_3\in\tilde I^3(\bar x),...,i_n\in\tilde I^n(\bar x)}\bar\lambda^3_{i_3}...\bar\lambda^n_{i_n} = ... = \sum_{i\in\tilde I^n(\bar x)}\bar\lambda^n_i = 1.\end{aligned} \tag{17}$$

Further, since $\tilde i = (i_1, ..., i_n) \in \tilde I(\bar x) = \tilde I^1(\bar x) \times ... \times \tilde I^n(\bar x)$, we have $f^j_{i_j}(\bar x) = h^j(\bar x)$ for all $j = 1, ..., n$. Therefore,

$$\phi_{\tilde i}(\bar x) = (f^1_{i_1}(\bar x), ..., f^n_{i_n}(\bar x)) = (h^1(\bar x), ..., h^n(\bar x)) = h(\bar x). \tag{18}$$

Set $m := |\tilde I^1(\bar x)| \times |\tilde I^2(\bar x)| \times ... \times |\tilde I^n(\bar x)| - 1$, where $|\tilde I^j(\bar x)|$ is the cardinality of the set $\tilde I^j(\bar x)$ for any $j = 1, 2, ..., n$. Then $|\tilde I(\bar x)| = m + 1$ and there are $m + 1$ such maps $\phi_{\tilde i}$.

Consider now a map $\Psi : \mathrm{inn}\Delta^m \times \mathbb{R}^n \to \mathbb{R}^n$ of the form

$$\Psi(\lambda, x) := \sum_{\tilde i\in\tilde I(\bar x)}\lambda_{\tilde i}\phi_{\tilde i}(x), \quad (\lambda, x) \in \mathrm{inn}\Delta^m \times \mathbb{R}^n, \lambda := (\lambda_{\tilde i})_{\tilde i\in\tilde I(\bar x)} \in \mathbb{R}^{m+1} \tag{19}$$

(without lost of generality, we may assume that the maps $\phi_{\tilde i}$ are ordered by an one-to-one map from $\tilde I(\bar x)$ to $\{0, 1, ...m\}$). Theorem 3.1 applied to $\Psi$ implies that the set of strongly critical values of $\Psi$ has Lebesgue measure zero

$$\mathrm{mes}\Psi(\hat{\mathrm{Cri}}\Psi) = 0. \tag{20}$$

Let $\bar\lambda := (\bar\lambda_{\tilde i})_{\tilde i\in\tilde I(\bar x)} \in \mathbb{R}^{m+1}$. Then (17) implies $\bar\lambda \in \mathrm{inn}\Delta^m$. Hence, $(\bar\lambda, \bar x) \in \mathrm{inn}\Delta^m \times \mathbb{R}^n$. We will show that $(\bar\lambda, \bar x) \in \hat{\mathrm{Cri}}\Psi$. Recall that (18) is already established. Further, since $\bar x \in \mathcal{D}$, $D\phi_{\tilde i}(\bar x)$ exists and $D\phi_{\tilde i}(\bar x) = (Df^1_{i_1}(\bar x), ..., Df^n_{i_n}(\bar x))$ for all $\tilde i \in \tilde I(\bar x)$. It remains to show that $\mathrm{rank}A < n$, where

$$A := \sum_{\tilde i\in\tilde I(\bar x)}\bar\lambda_{\tilde i}D\phi_{\tilde i}(\bar x) = \sum_{\tilde i\in\tilde I(\bar x)}(\bar\lambda_{\tilde i}Df^1_{i_1}(\bar x), ..., \bar\lambda_{\tilde i}Df^n_{i_n}(\bar x)).$$

Recall that $\bar\lambda_{\tilde i} = \bar\lambda^1_{i_1}...\bar\lambda^n_{i_n}$. Then one can use the argument of the proof of (17) to check that

$$A = \{\sum_{i\in\tilde I^1(\bar x)}\lambda^1_i Df^1_i(\bar x), ..., \sum_{i\in\tilde I^n(\bar x)}\lambda^n_i Df^n_i(\bar x)\}.$$

The desired inequality $\mathrm{rank}A < n$ now follows from (16). Thus, $(\bar\lambda, \bar x) \in \hat{\mathrm{Cri}}\Psi$.

On the other hand, taking (17) and (18) into account, we obtain

$$\Psi(\bar\lambda, \bar x) = \sum_{\tilde i\in\tilde I(\bar x)}\bar\lambda_{\tilde i}\phi_{\tilde i}(\bar x) = \sum_{\tilde i\in\tilde I(\bar x)}\bar\lambda_{\tilde i}h(\bar x) = (\sum_{\tilde i\in\tilde I(\bar x)}\bar\lambda_{\tilde i})h(\bar x) = h(\bar x),$$

which yields

$$h(\bar{x}) \in \Psi(\hat{\text{Cri}}\Psi). \tag{21}$$

Observe that the number of maps of the form (19) is finite. Therefore, (20) and (21) imply

$$\text{mes} h(\text{Cri} h) = 0.$$

□

Let us illustrate Theorem 4.1 by an example.

**Example 4.1.** Let $f_1, f_2, f_3 : \mathbb{R}^2 \to \mathbb{R}^2$ be maps given by

$$f_1(x, y) = (|x| + y, 2x + |y|),$$

$$f_2(x, y) = (|x| + 2y, x + |y|)$$

(the first map is given from [12, Remarks 7.1.2. (iii)]) and

$$f_3(x, y) = (x + y\text{d}(x, C) + y + x\text{d}(y, C), x + 2y\text{d}(x, C) + y + 2x\text{d}(y, C)),$$

where $C \subset \mathbb{R}$ is the Cantor set. These maps are essentially smooth locally Lipschitz. Theorem 4.1 yields that the set of the Clarke critical values of any continuous selection of the family $\{f_1, f_2, f_3\}$ has Lebesgue measure zero.

## 5. Applications in optimization

In this section, we study the set of parameters for which almost all or all optimal solutions of the corresponding scalar/vector parametrized constrained optimization problems satisfy Karush-Kuhn-Tucker type necessary conditions and the set of Pareto optimal values of a continuous (vector-valued) selection.

Firstly, we consider the following (scalar) parametrized optimization problem with constraints given by equalities and inequalities

$$
\begin{aligned}
(\mathcal{P}_{r,s}) \qquad\qquad & \min f(x) \\
& x \in \mathbb{R}^n, g_i(x) \le r, i = 1, ..., m \\
& h_j(x) = s_j, j = 1, ..., n-1
\end{aligned}
$$

Here, $f, g_i, h_j : \mathbb{R}^n \to \mathbb{R}$ $(i = 1, ..., m,\ j = 1, ..., n-1)$ are functions and $(r, s) \in \mathbb{R} \times \mathbb{R}^{n-1}$, $s = (s_1, ..., s_{n-1}) \in \mathbb{R}^{n-1}$. Denote by $\Omega_{r,s}$ the feasible set of $(\mathcal{P}_{r,s})$

$$\Omega_{r,s} := \{x \in \mathbb{R}^n \ : \ g_i(x) \le r, i = 1, ..., m, h_j(x) = s_j, j = 1, ..., n-1\}.$$

We will assume that $\Omega_{r,s}$ is nonempty. We say that a feasible point $\bar{x} \in \Omega_{r,s}$ is a local optimal solution of $(\mathcal{P}_{r,s})$ (or simply, $\bar{x}$ solves $(\mathcal{P}_{r,s})$) if there exists $\delta > 0$ such that

$$f(x) \ge f(\bar{x}) \qquad \text{for all } x \in \Omega_{r,s} \cap \mathbb{B}(\bar{x}, \delta).$$

Theorem 6.1.1 in [12] applied to this problem gives the following necessary conditions.

**Proposition 5.1.** *Assume that* $f, g_i, h_j : \mathbb{R}^n \to \mathbb{R}$ $(i = 1, ..., m,\ j = 1, ..., n-1)$ *are locally Lipschitz functions. Let* $\bar{x}$ *solve* $(\mathcal{P}_{r,s})$. *Then there exist* $\lambda \geq 0$, $\alpha_i \geq 0$ $(i = 1, ..., m)$ *and* $\beta_j$ $(j = 1, ..., n-1)$, *not all zero, such that*

$$\sum_{i=1}^{m} \alpha_i (g_i(\bar{x}) - r) = 0 \tag{22}$$

*and*

$$0 \in \lambda \partial f(\bar{x}) + \sum_{i=1}^{m} \alpha_i \partial g_i(\bar{x}) + \sum_{j=1}^{n-1} \beta_j \partial h_j(\bar{x}). \tag{23}$$

The case when (23) holds with $\lambda = 1$It is of interest. We will need the following hypothesis:

- (H1) The functions $g_i : \mathbb{R}^n \to \mathbb{R}$ $(i = 1, ..., m)$ are essentially smooth locally Lipschitz. Let $\mathcal{D} := \{x\ :\ g_i \text{ is smooth at } x,\ \forall i = 1, ..., m\}$.
- (H2) The functions $g_i : \mathbb{R}^n \to \mathbb{R}$ $(i = 1, ..., m)$ are either smooth on $\mathbb{R}^n$ or convex.
- (H3) The functions $h_j : \mathbb{R}^n \to \mathbb{R}$ $(j = 1, ..., n-1)$ are essentially smooth locally Lipschitz.

**Proposition 5.2.** *Assume that* $f$ *is a locally Lipschitz function. Assume further that (H1) (respectively, (H2)) and (H3) are satisfied.*

*Then for almost all parameter* $(r, s) \in \mathbb{R} \times \mathbb{R}^{n-1}$, *every optimal solution* $\bar{x} \in \mathcal{D}$ *(respectively, every optimal solution* $\bar{x} \in \mathbb{R}^n$*) of* $(\mathcal{P}_{r,s})$ *satisfies a necessary Karuh-Kuhn-Tucker type condition, i.e., there exist* $\alpha_i \geq 0$ $(i = 1, ..., m)$ *and* $\beta_j$ $(j = 1, ..., n-1)$ *such that*

$$0 \in \partial f(\bar{x}) + \sum_{i=1}^{m} \alpha_i \partial g_i(\bar{x}) + \sum_{j=1}^{n-1} \beta_j \partial h_j(\bar{x}). \tag{24}$$

*Here,* $\partial g_i(\bar{x}) = \{g_i'(\bar{x})\}$ *when* $g_i$ *is smooth at* $\bar{x}$ *and* $\partial g_i(\bar{x})$ *is the subdifferential of convex analysis when* $g_i$ *is convex.*

*Proof.* Let $g : \mathbb{R}^n \to \mathbb{R}$ be a function defined by $g(x) = \max\{g_i(x), i = 1, ..., m\}$ and denote by $I(x)$ the set of active indexes at $x$, i.e., $I(x) := \{i \in \{1, ..., m\}, g_i(x) = g(x)\}$. Clearly, $g$ is locally Lipschitz. Further, let $\Phi : \mathbb{R}^n \to \mathbb{R}^n$ and $\Phi_i : \mathbb{R}^n \to \mathbb{R}^n$, $i = 1, ..., m$ be maps defined by

$$\Phi(x) = (g(x), h_1(x), ..., g_{n-1}(x)),$$

and

$$\Phi_i(x) = (g_i(x), h_1(x), ..., g_{n-1}(x)).$$

Observe that the map $\Phi$ is locally Lipschitz and hence, it is continuous. Clearly, the maps $\Phi_i$ are essentially smooth locally Lipschitz. Therefore, $\Phi$ is a continuous selection of a finite family of essentially smooth locally Lipschitz maps. Theorem 4.1 implies that the set of critical values $(r, s) \in \mathbb{R} \times \mathbb{R}^{n-1}$ of $\Phi$ has Lebesgues measure zero.

Let $(r, s)$ be a regular value of $\Phi$. Then for any $x \in \mathbb{R}^n$ with $\Phi(x) = (r, s)$ and any $v \in \partial g(x)$, $u_j \in \partial h_j(x)$ $(j = 1, ..., n-1)$, one has

$$\operatorname{rank}\{v, u_1, ..., u_{n-1}\} = n, \tag{25}$$

or, in other words, the vectors $v, u_1, ..., u_{n-1}$ are linearly independent. Assume that $\bar{x}$ solve $(\mathcal{P}_{r,s})$. Proposition 2.3 gives

$$\partial g(\bar{x}) = \operatorname{co}\{\partial g_i(\bar{x}), i \in I(\bar{x})\}. \tag{26}$$

By Proposition 5.1, there exists $\lambda \geq 0$, $\alpha_i \geq 0$ $(i = 1, ..., m)$ and $\beta_j$ $(j = 1, ..., n-1)$, not all zero, such that (22) and (23) are satisfied. Our aim is to show that $\lambda \neq 0$. Suppose to the contrary that $\lambda = 0$. Then the scalars $\alpha_i \geq 0$ $(i = 1, ..., m)$ and $\beta_j$ $(j = 1, ..., n-1)$ are not all zero. Let $v_i \in \partial g_i(\bar{x})$ $(i = 1, ..., m)$ and $u_j \in \partial h_j(\bar{x})$ $(j = 1, ..., n-1)$ such that

$$0 = \sum_{i=1}^{m} \alpha_i v_i + \sum_{j=1}^{n-1} \beta_j u_j. \tag{27}$$

It follows from (22) that $\alpha_i = 0$ if $g_i(\bar{x}) < r$. Then (27) becomes

$$0 = \sum_{i \in I(\bar{x})} \alpha_i v_i + \sum_{j=1}^{n-1} \beta_j u_j. \tag{28}$$

If $\alpha_i = 0$ for all $i \in I(\bar{x})$, then (28) implies that $0 = \sum_{j=1}^{n-1} \beta_j u_j$, which together with the fact that not all scalars $\beta_j$ are zero yields that $u_j$ $(j = 1, ..., n-1)$ are not linearly independent, a contradiction to (25). Hence, at least one scalar $\alpha_i$, $i \in I(\bar{x})$ must be nonzero. Without lost of generality, we may assume that $\alpha_i \in [0, 1]$ for $i \in I(\bar{x})$ and $\sum_{i \in I(\bar{x})} \alpha_i = 1$. It follows from (26) that

$$v := \sum_{i \in I(\bar{x})} \alpha_i v_i \in \sum_{i \in I(\bar{x})} \alpha_i \partial g_i(\bar{x}) = \partial g(\bar{x}).$$

Then (28) becomes

$$0 = v + \sum_{j=1}^{n-1} \beta_j u_j,$$

where $v \in \partial g(\bar{x})$, $u_j \in \partial h_j(\bar{x})$ $(j = 1, ..., n-1)$, which is a contradiction to (25). Thus, (23) holds with $\lambda \neq 0$. Dividing both sides of (23) by $\lambda$, we get (24). $\square$

Similarly, we consider the following parametrized vector optimization problem

$$(\mathcal{VP}_{r,s}) \qquad \text{Minimize} f(x)$$
$$x \in \mathbb{R}^n, g_i(x) \leq r, i = 1, ..., m$$
$$h_j(x) = s_j, j = 1, ..., n-1$$

Here, $f = (f_1, ..., f_q)$, $f_k : \mathbb{R}^n \to \mathbb{R}$ $(k = 1, ..., q)$ are functions. We say that a feasible point $\bar{x} \in \Omega_{r,s}$ is a local weak Pareto optimal solution of $(\mathcal{VP}_{r,s})$ if there exists $\delta > 0$ such that there is no $x \in \Omega_{r,s} \cap \mathbb{B}(\bar{x}, \delta)$ such that $f_k(x) < f_k(\bar{x})$ for all $k = 1, ..., q$. Applying Theorem 4.1 and Theorem 6.1.3 in [12], and the arguments used in the proof of Proposition 5.2, one obtains the following result for the problem $(\mathcal{VP}_{r,s})$.

**Proposition 5.3.** *Assume that $f_k$, $k = 1, ..., q$, are locally Lipschitz functions. Assume further that (H1) (respectively, (H2)) and (H3) are satisfied. Then for almost all parameter $(r, s) \in \mathbb{R} \times \mathbb{R}^{n-1}$, every weak Pareto optimal solution $\bar{x} \in \mathcal{D}$ (respectively, every weak Pareto optimal solution $\bar{x} \in \mathbb{R}^n$) of $(\mathcal{VP}_{r,s})$ satisfies a necessary Karuh-Kuhn-Tucker type condition, i.e., there exist $\lambda_k \geq 0$ ($k = 1, ..., q$), not all zero, $\alpha_i \geq 0$ ($i = 1, ..., m$) and $\beta_j$ ($j = 1, ..., n-1$) such that*

$$0 \in \sum_{i=1}^{q} \lambda_k \partial f_k(\bar{x}) + \sum_{i=1}^{m} \alpha_i \partial g_i(\bar{x}) + \sum_{j=1}^{n-1} \beta_j \partial h_j(\bar{x}).$$

*Here, $\partial g_i(\bar{x}) = \{g_i'(\bar{x})\}$ when $g_i$ is smooth at $\bar{x}$ and $\partial g_i(\bar{x})$ is the subdifferential of convex analysis when $g_i$ is convex.*

Finally, we show that the Pareto minimal values of a continuous selection has measure zero. Let $K \subset \mathbb{R}^p$ be a nontrivial pointed closed convex cone (pointedness means $K \cap (-K) = \{0\}$) and $f : \mathbb{R}^n \to \mathbb{R}^p$. Recall that the value $y = f(\bar{x})$ of $f$ at some $\bar{x} \in \mathbb{R}^n$ is said to be a *Pareto optimal value* for the map $f$ if there exists $\delta > 0$ such that

$$f(\mathbb{B}(\bar{x}, \delta)) \cap (\bar{y} - K) = \{\bar{y}\}. \tag{29}$$

**Proposition 5.4.** *Let $f_i : \mathbb{R}^n \to \mathbb{R}^n$ ($i \in I$) be essentially smooth locally Lipschitz functions. Then for any continuous selection $f$ of the family $\{f_i \; : \; i \in I\}$, the set of its Pareto optimal values has Lebesgue measure zero.*

*Proof.* Let $\bar{y} = f(\bar{x})$ be a Pareto minimal value of $f$. Then (29) holds for some $\delta > 0$, which means that $f$ is not surjective on $\mathbb{B}(\bar{x}, \delta)$. Theorem 7.1.1 from [12] yields that $\bar{y}$ is a Clarke critical point. The assertion follows from Theorem 4.1. □

We give an illustrating example for Proposition 5.4.

**Example 5.1.** Let $I$ and $f_1, f_2, f_3 : \mathbb{R}^2 \to \mathbb{R}^2$ be as in Example 4.1. Let $K = \mathbb{R}^2_+$ and $f : \mathbb{R}^2 \to \mathbb{R}^2$ be a map given by

$$f(x, y) = \begin{cases} (|x| + y, x + |y|) & \text{if}(x, y) \notin \mathbb{R}^2_+ \\ (x + y\mathrm{d}(x, C) + y + x\mathrm{d}(y, C), x + 2y\mathrm{d}(x, C) + y + 2x\mathrm{d}(y, C) & \text{otherwise} \end{cases}$$

This map is a continuous selection of the family $\{f_1, f_2, f_3\}$ and Proposition 5.4 yields that the set of Pareto optimal values of $f$ has Lebesgue measure zero. In fact, one can check that $f(\mathbb{R}^2_+) \subset \mathbb{R}^2_+$, $f(\mathbb{R}^2 \setminus \mathbb{R}^2_+) \subset \{(u, v) \in \mathbb{R}^2 \; : \; u + v \geq 0\}$ (because $|x| + y + x + |y| \geq 0$) and the set of Pareto optimal values of $f$ is $\{(u, v) \in \mathbb{R}^2 \; : \; u + v = 0\}$.

## References


[1] Aubin, J.-P., Ekeland, I., Applied Nonlinear Analysis, John Wiley and Sons, 1984.

[2] Azagra, D., Ferrera, J., Gomez-Gil J., The Morse-Sard theorem revisited, The Quarterly Journal of Mathematics, Oxford University Press https://doi.org/10.1093/qmath/hay004

[3] Barbet, L., Dambrine, M., Daniilidis, A., The Morse-Sard theorem for Clarke critical values. Adv. Math. 242 (2013), 217–227

[4] Barbet, L., Dambrine, M., Daniilidis, A., Rifford, L., Sard theorems for Lipschitz functions and applications in optimization. Israel J. Math. 212 (2016), no. 2, 757–790.

[5] S. M. Bates, Toward a precise smoothness hypothesis in Sard's theorem, Proc. Amer. Math. Soc. 117 (1993), no. 1, 279-283.

[6] Borwein, J. M., Minimal cuscos and subgradients of Lipschitz functions, in *Fixed Point Theory and Its Applications* (J.-B. Baillion and M. Thera, Eds.), Pitman Lecture Notes in Math., pp. 57-82, Longman, Essex, UK, 1991.

[7] Borwein, J. M.; Moors, W. B., Essentially smooth Lipschitz functions: compositions and chain rules. Foundations of computational mathematics (Rio de Janeiro, 1997), 16–22, Springer, Berlin, 1997.

[8] Borwein, J. M.; Moors, W. B., Essentially smooth Lipschitz functions. J. Funct. Anal. 149 (1997), no. 2, 305–351.

[9] Borwein, J. M.; Moors, W. B., Null sets and essentially smooth Lipschitz functions. SIAM J. Optim. 8 (1998), no. 2, 309–323.

[10] Borwein, J. M., Wang, X., Lipschitz functions with maximal subdifferentias are generic, Proc. Amer. Math. Soc. 128 (2000), 3221-3229.

[11] Clarke, F., Generalized gradient and applications, Trans. AMS, 205 (1975), 247-265.

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

[13] Ha, T.X.D., The Sard's theorem for a class of locally Lipschitz mappings, Séminaire d'analyse convexe, Vol. 17, Exp. No. 9, 13 pp., Univ. Sci. Tech. Languedoc, Montpellier, 1989.

[14] Howard, R., Alexandrov's theorem on the second derivatives of convex functions via Rademacher's theorem on the first derivatives of Lipschitz functions, notes from lectures, dated May 1998, available at http://people.math.sc.edu/howard/Notes/alex.pdf

[15] Landis, E., On the functions representable as difference of two convex functions, Dokl. Akad. Nauk., SSSR, Vol. 80, 1 (1951), 9-11 (in Russian).

[16] Lebourg, G., Generic differentiability of Lipschitz functions, Trans. AMS., 256 (1979), 125-144.

[17] Morse, A. P., The behavior of a function on its critical set, Ann. of Math. 40 (1939), 62-70.

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

[19] Pourciau, B.H., Analysis and Optimization of Lipschitz continuous mappings, J. Optim. Theory Appl. 22 (1977), 311-351.

[20] Qi L., Maximal normal operators and the second-order derivatives of nonsmooth functions, Industrial Engineering Technical Report 88-3, University of Wisconsin-Madison (1988).

[21] Qi, L., The maximal normal operator space and integration of subdifferentials of nonconvex functions, Nonlinear Anal. 13 (1989), 1003-1012.

[22] Rader, T., 1973, Nice demand functions, Econometrica 41, 913-935.

[23] Sard, A., The measure of critical values of differentiable maps, Bull. AMS. 48 (1942), 883-890.

[24] Shannon, C., Regular nonsmooth equations, Journal of Mathematical Economics 23 (1994) 147-165.



TIMAS, Thanglong University

*E-mail address*: `txdha@math.ac.vn`