Journal of Convex Analysis
Volume 28 (2021), No. 1, 157–178

Versions of the Sard Theorem for Essentially Smooth Lipschitz Maps and Applications in Optimization and Nonsmooth Equations*

Truong Xuan Duc Ha
Institute for Mathematics and Applied Sciences, Thang Long University, Hanoi, Vietnam
txdha@math.ac.vn

Received: January 21, 2020
Accepted: April 27, 2020

The classical Sard theorem (in a special case) states that the set of critical values of a C^1-map from an open set of $\mathbb{R}^n$ to $\mathbb{R}^n$ has Lebesgue measure zero. Motivated by a recent work of Barbet, Dambrine, Daniilidis and Rifford [*Sard theorems for Lipschitz functions and applications in optimization*, Israel J. Math. 212(2) (2016) 757-–790], we obtain in this paper versions of this theorem for a finite family of essentially smooth Lipschitz maps and for a locally Lipschitz continuous selection of this family. Here, a locally Lipschitz map is essentially smooth if its Clarke's subdifferential reduces to a singleton almost everywhere. As applications, we establish the genericity of Karush-Kuhn-Tucker type necessary condition for scalar/vector parametrized constrained optimization problems, Lebesgue zero measure of the set of Pareto optimal values of a map and the genericity of the finiteness of the solution set for a nonsmooth equation.

Keywords: Sard theorem, essentially smooth Lipschitz map, critical points, optimization, nonsmooth equation.

2010 Mathematics Subject Classification: 49J40.

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 [28]. Recall that a point x is critical point of f if $\operatorname{rank} f'(x) < p$ and y is a critical value of f if it is an image of some critical point. The case $p = 1$, known as the Morse-Sard theorem, has been considered earlier in [20]. The Sard theorem has been extended to d.c. functions (d.c. means "difference of convex functions") defined on $\mathbb{R}^2$ in [18]. Mentioning that the hypothesis in [28] is based solely on the differentiability of f, Sard relaxed it and imposed other conditions on the set of critical points [29]. Later, it was shown that the C^k smoothness assumption can be weakened to $C^{k-1,1}$, see [4] and its refinement [22] for the case $n > p$ and [15, 16, 26] 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 in the theories of differentiable manifolds and of degrees of maps, in optimization, ect..., and extensions of this result to other classes of maps. We cannot mention all

*Dedicated to the memory of my mother.

ISSN 0944-6532 / $ 2.50 © Heldermann Verlag

of the works in this area and therefore restrict ourselves to the survey [1] and the references therein.

We would like to look longer at the case when f is a map between spaces of the same dimension. In the classical versions of the Sard theorem, the differentiability condition is that f is a C^1-map (i.e., f is continuously differentiable) [28] or at least f is differentiable at considered points [29]. However, the map f is dictated by a problem at hand, and certainly may fail to satisfy the above conditions. Since nonsmooth phenomena in mathematics occur naturally and frequently, several approaches have been used for defining the notion of criticality in such situations and we would like to recall three of them:

The approach taken by Rader [26] is to work with the classical derivative and he adds points at which the derivative does not exist to the set of (classical) critical points. The second approach, stemming from the work of Clarke [11] for locally Lipschitz maps, 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 map. In Jofre's approach [17], the critical point notion is associated to a "Fréchet style" approximation of a so called Fréchet-stable map.

Locally Lipschitz maps appear nearly everywhere in mathematics see for instance [12] and, as it was remarked in [3], endeavoring to generalize the Sard theorem for these 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 [19] and 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) map f when $n = 1$ [3] (for the general case $n > 1$, see [10]). To obtain the Sard theorem for the Clarke critical values, an extra assumption is needed. In [15], such an assumption is that f is *essentially smooth Lipschitz*, i.e., f is locally Lipschitz and the Clarke subdifferential of f reduces to a singleton almost everywhere (a.e.). Here, "almost everywhere" means "except for a set of Lebesgue measure zero".

Recently, Barbet, Dambrine, Daniilidis and Rifford established in [3] the so called "preparatory Sard theorem" for a compact countable set I of C^k maps from an open subset of $\mathbb{R}^n$ to $\mathbb{R}^p$ with $k \geq n - p + 1$ and a Sard theorem for a locally Lipschitz continuous selection of this family. Since essentially smooth Lipschitz maps form a significant subclass of the class of locally Lipschitz maps (see Section 2), further research on the Sard theorem for such maps is of interest.

In this paper, we prove that the C^1 assumption in the mentioned above results of [3] can be weakened to "essential smooth Lipschitz" one in the special case $n = p$ and the indexed family I is finite. As applications, we show that the following assertions hold true under an essential smooth Lipschitz hypothesis: (i) for almost all parameters, any optimal solution/weak Pareto optimal solution to scalar/vector parametrized constrained optimization problems satisfies Karush-Kuhn-Tucker necessary condition; (ii) the set of Pareto optimal values of a map has Lebesgues measure zero; (iii) for almost all y, a nonsmooth equation $f(x) = y$ has at most a finite number of solutions. Our techniques are based on results about the Clark subdifferential from [11, 12] and some arguments of [2, 3, 15].

The paper 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 is devoted to applications in optimization and nonsmooth equations.

2. Preliminaries

Notations: Let be given an Euclidean space X. In what follows $\mathbb{B}$ and $\mathbb{B}(x, r)$ are the closed unit ball and the closed ball centered at $x \in X$ with radius r, respectively, and for a set $U \subset X$, $\bar{U}$, ∂U and clconv$U$ stand for the closure, the boundary and the closed convex hull of U. From now on unless otherwise stated, let

$$f = (f_1, ..., f_p) : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}^p$$

be a map and Ω be a nonempty open set.

The map under our consideration is assumed to be locally Lipschitz and it is known that such a map enjoys remarkable properties. For instance, it satisfies the so called *Condition (N)* about the image of a null set, which was introduced by Lusin, see [21] (a null set means a set with Lebesgue measure zero). We state this property in the following.

Proposition 2.1. *If f is locally Lipschitz on Ω, then it satisfies Condition (N), i.e., the image of any null subset of Ω by f is a null set.*

Proposition 2.1 can be proved, for instance, by using [23, Lemma 4.1], which states a similar property for the case f is globally Lipschitz on Ω, see also [14].

Another remarkable property of a locally Lipschitz map (formulated in the Rademacher theorem) is that f is Fréchet differentiable (each f_i is Fréchet differentiable) a.e. on any neighborhood of x in Ω. Let us recall the Clarke subdifferential based on this property of locally Lipschitz maps.

Definition 2.2. Assume that f is locally Lipschitz on Ω. The *Clarke subdifferential* of f at $x \in \Omega$, denoted by $\partial f(x)$, is the closed convex hull of all $n \times p$-matrices obtained as the limit of a sequence of derivatives $f'(x_i)$, where $x_i \to x$ and $f'(x_i)$ is defined:

$$\partial f(x) := \text{clconv } \{v \in \mathbb{R}^{n\times p} : v = \lim_{x_i \to x, f'(x_i) \text{ exists}} f'(x_i)\}. \qquad \square$$

Note that if f is Fréchet differentiable at x, its derivative $f'(x)$ coincides with the $n \times p$- matrix of partial derivatives of f at x. Basic properties of the Clarke subdifferential are collected in the following proposition, see [12, Propositions 2.2.1, 2.2.2, 2.2.4. and 2.6.2].

Proposition 2.3. *Assume that f is locally Lipschitz on Ω. Let $x \in \Omega$. Then the following assertions hold:*

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

(b) *∂f is upper semicontinuous at x: for any $\epsilon > 0$ there is $\delta > 0$ such that, for all u in $x + \delta\mathbb{B}$: $\partial f(u) \subset \partial f(x) + \epsilon\mathbb{B}_{n\times p}$.*

(c) *If f admits Fréchet derivative $f'(x)$, then $f'(x) \in \partial f(x)$.*

(d) *If f is continuously differentiable at x, then $\partial f(x) = \{f'(x)\}$.*

(e) *If $\partial f(x)$ reduces to a singleton $\{\xi\}$, then $\xi = f'(x)$ and for each $v \in \mathbb{R}^n$ one has*

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

Remark 2.4. Borwein used the expression f is "Fréchet-type strictly differentiable" (in short, *strictly differentiable*) at x in [8, p.323], which means, in the finite dimensional setting, that the convergence in (1) is uniform with respect to v in compact sets [12, p. 30-31] and which is equivalent to "$\partial f(x)$ reduces to a singleton", see [12, Proposition 2.2.4]. □

Next, we state a result relating the equality (1) and the strong differentiability, which is one of key tools in the proof of our version of the Sard theorem. Recall [23, p.324]) that f is *strongly differentiable* at x if for any $\epsilon > 0$ there is exists $\delta_x > 0$ 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

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

Lemma 2.5. *If f satisfies the equality (1) at x, then f is strongly differentiable at x.*

Proof. Let $\epsilon > 0$. Then the equality (1) implies the existence of $\delta_x > 0$ 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

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

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 $0 < t \le 2\delta_x$, $v \in \mathbb{B}$ and $x_2 = x_1 + tv$, the inequality (2) implies

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

or $\|f(x_1) - f(x_2) - f'(x)(x_1 - x_2)\| \le \epsilon\|x_1 - x_2\|$. □

Now, we explain the name of the main object of our study, the so called essentially smooth Lipschitz maps. Note that Lebourg's concept of smoothness for a real-valued function [19, 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 Ω and $\partial f(x)$ reduces to a singleton. On the other hand, Borwein's concept of *essential smoothness* [6, p.68] (as it was recalled in [8, p.323]) for the locally Lipschitz map f restricted to a finite dimensional space setting and in terms of the Clarke subdifferential leads to the following definition.

Definition 2.6. Assume that f is locally Lipschitz on Ω. We say that f is *essentially smooth* on Ω if $\partial f(x)$ reduces to a singleton a.e. on Ω. □

As noted in Remark 2.4, a locally Lipschitz map is essentially smooth if and only if it is strictly differentiable a.e. Essentially smooth Lipschit maps have been considered in a finite dimensional space setting in [15, 24, 25] and in a separable Banach space setting in [6, 7], see also Borwein-Woors papers in late 1990s [8, 9]. In [24, 25], 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 a.e. (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 Lipschitz maps on an open subset of $\mathbb{R}^n$ form a broad linear space containing, for instance, d.c. functions [25, Theorem 7, p. 1008]. It was shown that demand functions of some economies are essentially smooth Lipschitz [5]. Below are some illustrating examples.

Example 2.7. (i) See [19, 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 $0 < \mathrm{mes}(M \cap I)$, 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 a.e. 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 [23, 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 the 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 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 $f_i : \mathbb{R}^n \to \mathbb{R}$ by $f_i(x) = \mathrm{d}(x, \mathcal{C}_i)$. Let $f \colon \mathbb{R}^n \to \mathbb{R}^n$ be the map defined by $f(x) = (f_1(x), ..., f_n(x))$. This map is essentially smooth Lipschitz on $\mathbb{R}^n$.

Next, we recall some concepts of critical points/values in the case $n = p$.

Definition 2.8. Let $f \colon \Omega \subseteq \mathbb{R}^n \to \mathbb{R}^n$.

(i) The set $\mathrm{Cri}f$ of critical points in the classical sense [29] is given by

$$\mathrm{Cri}f := \{x \in \Omega \ : \ f'(x) \text{ exists and } \det f'(x) = 0\}.$$

(ii) The set $\mathrm{Cri}_R f$ of critical points in the sense of Rader [26] is given by

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

(iii) (f is assumed to be locally Lipschitz) The set $\mathrm{Cri}_C f$ of critical points in the sense of Clarke [12] is given by

$$\mathrm{Cri}_C f := \{x \in \Omega \ : \ \exists A \in \partial f(x),\ \det A = 0\}.$$

(iv) $x \in \Omega$ is a regular point in some sense if it is not a critical point in this sense.

(v) $y \in f(\Omega)$ is a critical value in some sense if it is an image of a critical point in this sense and y is a regular value in some sense if all points in its preimage $f^{-1}(y)$ are regular in this sense. □

Note that Clarke used an equivalent definition for regularity in [12]: *x is non-singular if any $A \in \partial f(x)$ has maximal rank n.* In what follows we omit the term "classical" and will use also the terms "the Rader/Clarke critical/regular point/value" for the concepts in Definition 2.8.

The classical Sard theorem [28] states that the set of critical values of a C^k map from $\Omega \subseteq R^n$ to R^p has the Lebesgues measure zero provided $k \geq n - p + 1$. Some extensions of the Sard theorem for the particular case $n = p$ are collected in the following.

Theorem 2.9. *Let f be a map from an open set Ω of $\mathbb{R}^n$ into $\mathbb{R}^n$.*

(i) ([26, Lemma 2]) *If f is Fréchet differentiable a.e. and satisfies Condition* (N)*, then* $\mathrm{mes} f(\mathrm{Cri}_R f) = 0$.

(ii) ([15, Theorem 3.2]) *If f is essentially smooth Lipschitz, then* $\mathrm{mes} f(\mathrm{Cri}_C f) = 0$.

Remark 2.10. (i) Let f be the function in Example 2.7 (i). Theorem 2.9 (i) yields $\mathrm{mes} f(\mathrm{Cri}_R f) = 0$. Note that for this map one has $\mathrm{mes}(\mathrm{Cri}_C f \setminus \mathrm{Cri}_R f) > 0$, i.e., $\mathrm{Cri}_R f$ is a proper subset of $\mathrm{Cri}_C f$. This example also shows that one can not drop the assumption on the essential smoothness in Theorem 2.9 (ii). Indeed, since $\mathrm{Cri}_C f = \mathbb{R}$ and f is strictly increasing continuous, we have $\mathrm{mes} f(\mathrm{Cri}_C f) > 0$.

(ii) Theorem 2.9 (ii) can be applied for instance to the map in Example 2.7 (iii).

(iii) It has been established that if $A \subseteq \mathrm{Cri} f$ is a countable union of sets each of which has a finite outer n-dimensional Hausdorff measure, then $\mathrm{mes} f(A) = 0$ [29, Theorem 1]. The assertions of Theorem 2.9 can be proved with the help of this result and Proposition 2.1 (see also [16, Theorem 3.1] for a detailed proof of the assertion (i) in the case of locally Lipschitz maps). Note that by the Rademacher theorem and Proposition 2.1, a locally Lipschitz map satisfies all conditions of the assertion (i).

3. A version of the preparatory Sard theorem for essentially smooth Lipschitz maps

In this section, we establish a version of the so called preparatory Sard theorem [2] in the case $n = p$ for essentially smooth Lipschitz maps. As before, Ω is an open subset of $\mathbb{R}^n$.

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

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

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

Clearly, Ψ 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 Ψ given in [2].

Definition 3.1. The set $\widehat{\mathrm{Cri}}\Psi$ of *strongly critical points* of Ψ is defined by

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

A scalar $t \in \mathbb{R}$ is called *strongly critical value of* Ψ if we have $\Psi(\lambda, x) = t$ for some $(\lambda, x) \in \widehat{\mathrm{Cri}}\Psi$. □

Our main result reads as follows.

Theorem 3.2. *Given $m + 1$ maps $\phi_i : \Omega \to \mathbb{R}^n$, $i = 0, 1, ..., m$, we set*

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

for all $\lambda = (\lambda_0, \lambda_1, ..., \lambda_m) \in$ innΔ^m and $x \in \Omega$. Assume that the maps ϕ_i, where $i = 0, 1, ..., m$, are essentially smooth Lipschitz on Ω. Then mes$\Psi(\widehat{\mathrm{Cri}}\Psi) = 0$.

Note that Theorem 3.2 reduces to the assertion (ii) in Theorem 2.9 when $m = 0$. Let us establish an auxiliary result.

Lemma 3.3. *The set* $\widehat{\mathrm{Cri}}\Psi$ *is closed in* inn$\Delta^m \times \Omega$.

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

Define $\lambda = (\lambda_0, \lambda_1, ..., \lambda_m)$ and $\lambda^j = (\lambda_0^j, \lambda_1^j, ..., \lambda_m^j)$. Observe that $\lambda \in$ innΔ^m. Since $(\lambda^j, x^j) \in \widehat{\mathrm{Cri}}\Psi$, for all $i \in \{1, ..., m\}$ we have $\phi_i(x^j) = \phi_0(x^j)$ and the continuity of the functions ϕ_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 2.3 (a) implies $M < +\infty$. If $M = 0$, then $(\lambda, x) \in \widehat{\mathrm{Cri}}\Psi$ 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 2.3 (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

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

Since $(\lambda^j, x^j) \in \widehat{\mathrm{Cri}}\Psi$, 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 $\mathrm{rank}A_j < n$. By (3), 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 2.3 (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

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

Thus, $(\lambda, x) \in \widehat{\mathrm{Cri}}\Psi$. □

We are ready now for the proof of Theorem 3.2.

Proof. Without loss of generality, we may assume that Ω is bounded. Denote

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

Since ϕ_i is smooth a.e. on Ω for all $i = 0, 1, ..., m$, the set $\Omega \backslash D$ has Lebesgue measure zero and so is the set $\mathrm{inn}\Delta^m \times (\Omega \setminus D)$. Therefore, the set $\widehat{\mathrm{Cri}}\Psi \cap (\mathrm{inn}\Delta^m \times (\Omega \setminus D))$ also has Lebesgue measure zero. Observe that the map Ψ is locally Lipschitz on $\mathrm{inn}\Delta^m \times \Omega$. Therefore, Proposition 2.1 implies that

$$\mathrm{mes}\ \Psi(\widehat{\mathrm{Cri}}\Psi \cap (\mathrm{inn}\Delta^m \times (\Omega \setminus D))) = 0.$$

Hence, it remains to prove that

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

For simplicity, we write $K := \widehat{\mathrm{Cri}}\Psi \cap (\mathrm{inn}\Delta^m \times D)$.

By Lemma 3.3, the set $\widehat{\mathrm{Cri}}\Psi$ 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 $\mathrm{mes}(K \setminus \cup_{i=1}^{\infty}\Gamma_i) = 0$. Again, Proposition 2.1 implies that $\mathrm{mes}\ \Psi((K \setminus \cup_{i=1}^{\infty}\Gamma_i)) = 0$. Hence, our proof will be achieved if we prove that for any nonempty compact subset Γ of K we have $\mathrm{mes}\ \Psi(\Gamma) = 0$.

Let $\epsilon > 0$ be an arbitrary scalar. To any $\lambda := (\lambda_0, ..., \lambda_m) \in \mathrm{inn}\Delta^m$ we associate a map Φ_λ 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 Φ_λ is essentially smoth Lipschitz on Ω.

Now, let $(\lambda, x) \in \Gamma \subset K$. Then $x \in D$ and ϕ_i is smooth at x for all $i = 0, 1, ..., m$ and so is the map Φ_λ. 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 ϕ_i at x. Since the map Φ_λ is strictly differentiable at x, Lemma 2.5 implies that 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 $\lambda \in \mathrm{inn}\Delta^m$, we can choose $\delta_{x,\lambda}$ such that $\mathbb{B}(\lambda, \delta_{x,\lambda}) \subset \mathrm{inn}\Delta^m$.

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

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

and (5) holds. Since Γ is a compact set, there exist a finite number l of pairs denoted by (λ^i, x_i), $i = 1, ..., l$ in Γ 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 $\Gamma \subset \mathrm{inn}\ \Delta^m \times \mathcal{C}$ 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

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

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

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

where α is positive scalar depending on Γ. Indeed, suppose that there is a pair $(\bar{\lambda}, \bar{x}) \in \Gamma$ such that $(\bar{\lambda}, \bar{x}) \in \mathrm{inn}\ \Delta^m \times C$.

By (6), one can find an index $i \in \{1, ..., l\}$ and the corresponding pair (λ^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 (β, 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,λ^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 $\lambda^i = (\lambda_0^i, ..., \lambda_m^i)$, where $\lambda_k^i > 0$ for $k = 0, ..., m$ and with $\sum_{k=0}^m \lambda_k^i = 1$. Therefore,

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

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 $(\lambda^i, x_i) \in \Gamma \subset K \subset \widehat{\text{Cri}}\Psi$, 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 $\quad |\rho_k| = |\langle A_{x_i,\lambda^i}(y - b), e_k\rangle| \le \|A_{x_i,\lambda^i}(y - b)\| \le \alpha\|y - b\| \le \alpha\sqrt{n}\frac{a}{N},$

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

Since $\partial\phi_j$ $(j = 0, ..., m)$ is upper semicontinuous from Ω with nonempty convex compact values in $\mathbb{R}^{n\times n}$, one can prove that α is finite and depends only on Γ. Let η_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 we have for all $k = 1, 2, ...n$

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

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

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

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

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

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 Γ, is less than or equal to N^n, the inequality (7) implies

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

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

Example 3.4. Theorem 4.1 can be applied 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 2.7(iii), and $\phi_1(x, y) = (|x|, |y|)$. Note that these maps are essentially smooth Lipschitz but are not C^1 on $\mathbb{R}^2$ and hence, Theorem 3 of [3] cannot be applied to this case. □

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 Lipschitz maps. Throughout this section, I is a finite set of indexes and Ω is a nonempty open subset of $\mathbb{R}^n$. We say that a map $h : \Omega \to \mathbb{R}^n$ is a selection of maps $f_i : \Omega \to \mathbb{R}^n$ $(i \in I)$ on Ω 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 $f_i\colon \Omega \to \mathbb{R}^n$ $(i \in I)$ be essentially smooth Lipschitz maps. Assume that $h\colon \Omega \to \mathbb{R}^n$ is a continuous selection of the maps f_i $(i \in I)$ on Ω. Then h is locally Lipschitz on Ω and* mes $h(\text{Cri}_C\ h) = 0$.

To prove Theorem 4.1, we need some auxiliary results.

Proposition 4.2. *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 Ω. If h is continuous on Ω, 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 $j \notin I(x)$. Therefore

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

Since $S \subset \cup_{x \in S} \mathbb{B}_S(x, \delta_x)$, we can choose $x_1, ..., x_k \in S$ with the properties:

(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 4.2 is a special case of [2, 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 following proposition.

Proposition 4.3. *Let $f_i : \Omega \to \mathbb{R}$ $(i \in I)$ be essentially smooth Lipschitz maps and $h : \Omega \to \mathbb{R}$ be a continuous selections of these maps. Then (h is a locally Lipschitz map and) for every point $x \in \mathcal{D}$ one has*

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

Here, $\mathcal{D} := \{x \in \Omega : h$ is Fréchet differentiable and f_i is smooth, $\forall i \in I\}$ and $I(x) := \{i \in I : h(x) = f_i(x)\}$.

Proof. Proposition 4.3 is an extension of [2, Proposition 4] (the C^1-differentiability is relaxed to essential smoothness) in the special case when the family of parameter I is finite and the proof of the former is similar to the one of the latter.

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

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

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

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

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

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

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

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

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

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

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

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

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

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 Ω to $\mathbb{R}$. The maps $f_i^j : \Omega \to \mathbb{R}$ $(j = 1, ..., n,\ i \in I)$ are essentially smooth Lipschitz. For each j, one has

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

By Proposition 4.3, h is locally Lipschitz on Ω. Let $\bar{x} \in \mathrm{Cri}_C h \cap \mathcal{D}$ be an arbitrary point, where $\mathcal{D}$ is the set defined in Proposition 4.3. 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 4.3 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 Clarke 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

$$\phi_{\tilde{i}}(x) = (f_{i_1}^1(x), ..., f_{i_n}^n(x)) \ \text{ and let } \ \bar{\lambda}_{\tilde{i}} := \bar{\lambda}_{i_1}^1 ... \bar{\lambda}_{i_n}^n.$$

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

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

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 $\Psi : \text{inn}\Delta^m \times \mathbb{R}^n \to \mathbb{R}^n$ of the form

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

(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 3.2 applied to Ψ implies that the set of strongly critical values of Ψ has Lebesgue measure zero

$$\text{mes}\Psi(\widehat{\text{Cri}}\Psi) = 0. \tag{20}$$

Let $\bar{\lambda} := (\bar{\lambda}_{\tilde{i}})_{\tilde{i} \in \tilde{I}(\bar{x})} \in \mathbb{R}^{m+1}$. Then (17) implies $\bar{\lambda} \in \text{inn}\Delta^m$. Hence, $(\bar{\lambda}, \bar{x}) \in \text{inn}\Delta^m \times \mathbb{R}^n$. We will show that $(\bar{\lambda}, \bar{x}) \in \widehat{\text{Cri}}\Psi$. 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 $\text{rank}A < n$, 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

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

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

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

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

Observe that the number of maps of the form (19) is finite. Therefore, (20) and (21) imply mes $h(\text{Cri}_C h) = 0$. □

We illustrate Theorem 4.1 by an example.

Example 4.4. Let $f_1, f_2, f_3 : \mathbb{R}^2 \to \mathbb{R}^2$ be maps given by

$$f_1(x, y) = (|x| + y, 2x + |y|), \quad \text{and} \quad f_2(x, y) = (|x| + 2y, x + |y|)$$

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

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

where $C \subset \mathbb{R}$ is the Cantor set. These maps are essentially smooth Lipschitz but are not C^1. 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. Note that Theorem 2 of [3] cannot be applied to them. □

5. Some applications in optimization and nonsmooth equations

5.1. Applications in optimization

It is known that the Sard theorem plays an important role in the study of generic optimality conditions in nonlinear programming. We would like to mention the initiating work in this domain of research by Saigal and Simon [27] and generalizations in the works by Spingarn and Rockafellar [33], Spingarn [31, 32], Jofre [17] and Barbet, Dambrine, Daniilidis and Rifford [3]. In this subsection, we apply Theorem 4.1 to study the set of parameters for which 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. Note that in contrast to the work [3], Theorem 6.1.3 in [12] allows us to consider here also the vector parametrized constrained optimization problems. We provide examples illustrating obtained results.

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

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

Here, $f, g_i, h_j : \mathbb{R}^n \to \mathbb{R}$ ($i=1,...,m$, $j=1,...,n-1$) are functions, $(r,s) \in \mathbb{R} \times \mathbb{R}^{n-1}$, $s = (s_1,...,s_{n-1}) \in \mathbb{R}^{n-1}$. 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 $f(x) \ge f(\bar{x})$ 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 $\lambda \ge 0$, $\alpha_i \ge 0$ ($i = 1,...,m$) and β_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 $\lambda = 1$ is of interest. We will need the following hypotheses:

(H1) The functions $g_i : \mathbb{R}^n \to \mathbb{R}$ $(i = 1, ..., m)$ are essentially smooth 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 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 Karush-Kuhn-Tucker type condition, i.e., there exist $\alpha_i \geq 0$ $(i = 1, ..., m)$ and β_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.

To prove this proposition, we need the following fact.

Proposition 5.3. [12, Propositions 2.3.6, 2.3.12] *Let $\Omega \subset \mathbb{R}^n$ a nonempty open set, $g_i : \Omega \to \mathbb{R}$, $i = 1, ..., m$ be locally Lipschitz functions and $g : \mathbb{R}^n \to \mathbb{R}$ a function defined by $g(x) = \max\{g_i(x), i = 1, ..., m\}$. Let $x \in \Omega$ 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*

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

Let us return to the proof of Proposition 5.2.

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

$$\Phi(x) = (g(x), h_1(x), ..., g_{n-1}(x)), \quad \text{and} \quad \Phi_i(x) = (g_i(x), h_1(x), ..., g_{n-1}(x)).$$

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

Let (r, s) be a Clarke regular value of Φ. 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 5.3 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 β_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 β_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 β_j are zero yields that u_j $(j = 1, ..., n-1)$ are not linearly independent, a contradiction to (25). Hence, at least one scalar α_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 λ, we get (24). □

Similarly, we consider the following parametrized vector optimization problem

$(\mathcal{VP}_{r,s})$ $\quad \min f(x)$ s.t. $x! \in \mathbb{R}^n$, $g_i(x) \leq r, i = 1, ..., m$; $h_j(x) = s_j$, $j = 1, ..., n-1$.

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 5.4. *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 Karush-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 β_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.

Next, we show that the Pareto optimal 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 $\bar{y} = f(\bar{x})$ 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 5.5. *Let $f_i \colon \mathbb{R}^n \to \mathbb{R}^n$ $(i \in I)$ be essentially smooth 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 optimal 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 5.5.

Example 5.6. Let I and $f_1, f_2, f_3 : \mathbb{R}^2 \to \mathbb{R}^2$ be as in Example 4.4. 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 5.5 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 $\{(u, v) \in \mathbb{R}^2 \ : \ u + v = 0\}$. Note that Proposition 13 of [3] cannot be applied to this case because the maps under consideration are not C^1. □

5.2. An application in the study of Nonsmooth Equations

Many problems in economic equilibrium sensitivity analysis are based on solving and analyzing solutions to an equation $f(x) = y$. A mathematical model which attempts to explain economic equilibrium must have a nonempty set of solutions. One would also wish the solution to be unique or the equation to have at least a finite set of solutions. In his seminal paper [13], Debreu established the genericity of economies with a finite set of equilibria in the C^1 case. Continuing the research carried out in the works [13, 26], Shannon studied economies, which are regular in the sense of Rader. She established in [30, Corollary 2] the finiteness for the solution set when y is a Rader regular value of f and applied the Rader version of the Sard theorem to obtain the genericity of such property of this equation. Afterward, she established the finiteness of equilibrium of a regular economy (a "regular economy" is defined through the Rader concept of regular value).

Motivated by Shannon's work, we will show that Theorem 4.1 gives information about how "big" is the set of y for which the equation $f(x) = y$ has at most a finite number of solutions on a given compact set. Firstly, we show that if f is locally Lipschitz and y is a Clarke regular value of f, the above equation has at most a finite number of solutions on any compact set (we make a convention that this number is zero when the equation has no solution). We will need the following inverse function theorem due to Clarke.

Proposition 5.7. [12, Theorem 7.1.1] *Let Ω be an open subset of $\mathbb{R}^n$, $f : \Omega \to \mathbb{R}^p$ be a locally Lipschit and $x \in \Omega$. If $\partial f(x)$ is of maximal rank (this means that every $A \in \partial f(x)$ is of maximal rank), then there exist neighborhoods U in Ω and V in $\mathbb{R}^p$ of x and $f(x)$, respectively, and a Lipschitz map $g : V \to \mathbb{R}^n$ such that*

(a) $g(f(u)) = u$ *for all* $u \in U$.

(b) $f(g(v)) = v$ *for all* $v \in V$.

Our results read as follows.

Proposition 5.8. *Let $\Omega \subseteq \mathbb{R}^n$ be an open set and $f : \Omega \to \mathbb{R}^n$ be a locally Lipschitz map. Assume that U is a nonempty compact subset of Ω. Then for any Clarke regular value y of f, the equation $f(x) = y$ has at most a finite number of solutions on U.*

Proof. Let $S_U := f^{-1}(y) \cap U$. The assertion is automatically true when $S_U = \emptyset$, so we assume that $S_U \neq \emptyset$. The compactness of U and the continuity of f imply that S_U is compact. For any solution $x \in S_U$, the Clarke subdifferential $\partial f(x)$ is of maximal rank because y is a Clarke regular value of f. Proposition 5.7 then implies that f is an one-to-one map between some two neighborhoods of x and y, which means that x is an isolated point of S_U. Thus, S_U is a discrete set and being compact, it has a finite number of points. □

Proposition 5.9. *Let $f_i : \Omega \to \mathbb{R}^n$ $(i \in I)$ be essentially smooth Lipschitz maps, $f : \Omega \to \mathbb{R}^n$ a continuous selection of these maps and $U \subset \Omega$ a nonempty compact set. Then the equation $f(x) = y$ has at most a finite number of solutions on U for all y outside a set of Lebesgue measure zero.*

Proof. Observe first that f is locally Lipschitz by Theorem 4.1. If the equation $f(x) = y$ has infinite number of solutions on U, Proposition 5.8 implies that y cannot be a Clarke regular value. The assertion follows from Theorem 4.1. □

Example 5.10. Let $f : \mathbb{R}^2 \to \mathbb{R}^2$ be the map $f(x_1, x_2) = (|x_1| + x_2, 2x_1 + |x_2|)$. We have

$$\partial f(0,0) = \left\{ \begin{bmatrix} s & 1 \\ 2 & t \end{bmatrix} \ : \ |s| \leq 1, |t| \leq 1 \right\},$$

see [12, Remarks 7.1.2.(iii)]). Since $f^{-1}(0,0) = \{(0,0)\}$ and for any $A \in \partial f(0,0)$, we have $\det A \leq -1$, it follows that $(0,0)$ is a Clarke regular value of f. Proposition 5.8 implies that the equation $f(x_1, x_2) = (0,0)$ has at most a finite number of solutions in any compact set (in fact, it has an unique solution $(x_1, x_2) = (0,0)$). It is clear that Proposition 5.9 is applicable to this map. Note that $(0,0)$ is not a Rader regular value because f is not differentiable at $(0,0)$ and some results of [26, 30] are not applicable to this map. □

In conclusion, we would like to mention that demand functions of some economies are proved to be continuously differentiable on an open set of full Lebesgue measure, in other words, they are essentially smooth Lipschitz, see the work [5] by Bonnisseau and Rivera-Cayupi. Therefore, similar to [30], results of this subsection could be used to study these economies the regularity of which will be defined by means of the Clarke regularity of a locally Lipschitz map.

Acknowledgments. The author would like to express her sincere thanks to the referee for useful remarks which help to improve the paper. This work was partially carried out during the author's stay at the Vietnam Institute for Advanced Study in Mathematics.

References

[1] D. Azagra, J. Ferrera, J. Gomez-Gil: *The Morse-Sard theorem revisited*, Quart. J. Math. 69(3) (2018) 887–913.

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

[3] L. Barbet, M. Dambrine, A. Daniilidis, L. Rifford: *Sard theorems for Lipschitz functions and applications in optimization*, Israel J. Math. 212(2) (2016) 757–790.

[4] S. M. Bates: *Toward a precise smoothness hypothesis in Sard's theorem*, Proc. Amer. Math. Soc. 117(1) (1993) 279–283.

[5] J.-M. Bonnisseau, J. Rivera-Cayupi: *Constrained consumptions, Lipschitzian demands, and regular economies*, J. Optimization Theory Appl. 131(2) (2006) 179–193.

[6] J. M. Borwein: *Minimal cuscos and subgradients of Lipschitz functions*, in: *Fixed Point Theory and its Applications*, J.-B. Baillion, M. Théra (eds.), Pitman Lecture Notes in Mathematics 252, Longman, Harlow (1991) 57–81.

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

[8] J. M. Borwein, W. B. Moors: *Essentially smooth Lipschitz functions*, J. Funct. Analysis 149(2) (1997) 305–351.

[9] J. M. Borwein, W. B. Moors: *Null sets and essentially smooth Lipschitz functions*, SIAM J. Optimization 8(2) (1998) 309–323.

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

[11] F. Clarke: *Generalized gradient and applications*, Trans. Amer. Math. Soc. 205 (1975) 247–265.

[12] F. Clarke: *Optimization and Nonsmooth Analysis*, Wiley, New York (1983).

[13] G. Debreu: *Economies with a finite set of equilibria*, Econometrica 38 (1970) 387–392.

[14] H. Federer: *Geometric Measure Theory*, Springer, New York (1969).

[15] T. X. D. Ha: *The Sard's theorem for a class of locally Lipschitz mappings*, Sém. d'Analyse Convexe 17, Exp. No. 9, Univ. du Languedoc, Montpellier (1989) 13 pp.

[16] R. Howard: *Alexandrov's theorem on the second derivatives of convex functions via Rademacher's theorem on the first derivatives of Lipschitz functions*, Notes from Lectures, available at http://people.math.sc.edu/howard/Notes/alex.pdf (1998).

[17] A. Jofre: *Image of Critical Points Set for Nonsmooth Mappings and Applications*, in: *Contributions à l'Analyse Non-Différentiable avec Applications à l'Optimisation et à l'Analyse Non-Linéaire*, Thèse, Université de Pau et des Pays de l'Adour, Pau, France (1989).

[18] E. Landis: *On the functions representable as difference of two convex functions (Russian)*, Dokl. Akad. Nauk SSSR 80(1) (1951) 9–11.

[19] G. Lebourg: *Generic differentiability of Lipschitz functions*, Trans. Amer. Math. Soc. 256 (1979) 125–144.

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

[21] I. P. Natanson: *Theory of Functions of a Real Variable. Volume II*, Frederick Ungar Publishing Co., New York (1957).

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

[23] B. H. Pourciau: *Analysis and optimization of Lipschitz continuous mappings*, J. Optimization Theory Appl. 22 (1977) 311–351.

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

[25] L. Qi: *The maximal normal operator space and integration of subdifferentials of nonconvex functions*, Nonlinear Analysis 13 (1989) 1003–1012.

[26] T. Rader: *Nice demand functions*, Econometrica 41 (1973) 913–935.

[27] R. Saigal, C. Simon: *Generic properties of the complementarity problem*, Math. Programming 4 (1973) 324–335.

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

[29] A. Sard: *Images of critical sets*, Ann. of Math. 68 (1958) 247–259.

[30] C. Shannon: *Regular nonsmooth equations*, J. Math. Economics 23 (1994) 147–165.

[31] J. Spingarn: *Fixed and variable constraints in sensitivity analysis*, SIAM J. Control Optimization 18(3) (1980) 297–310.

[32] J. Spingarn: *On optimality conditions for structured families of nonlinear programming problems*, Math. Programming 22(1) (1982) 82–92.

[33] J. Spingarn, R. T. Rockafellar: *The generic nature of optimality conditions in nonlinear programming*, Math. Oper. Res. 4(4) (1979) 425–430.