Metric regularity and subdifferential calculus

A. D. Ioffe

Abstract. The theory of metric regularity is an extension of two classical results: Lyusternik tangent space theorem and Graves surjection theorem. The developments of non-smooth analysis in 80-ies and 90-ies opened a number of possibilities of far reaching extensions of these results. It was also well understood that the phenomena behind the results have a metric origin, not connected with any linear structures. At the same time it became clear that some basic hypotheses of the subdifferential calculus are closely connected with metric regularity of certain set-valued maps. The survey is devoted to the metric theory of metric regularity and its connection with subdifferential calculus in Banach spaces.

Bibliography 97 items

Contents

Introduction 000
Chapter 1. The metric theory 000
§1. Terminology, notation, definitions. Equivalence theorems 000
§2. General and local regularity criteria 000
§3. Applications and extensions 000
§4. Comments 000
Chapter 2. Subdifferentials 000
§1. Definitions and examples of subdifferentials 000
§2. Basic theorems of subdifferential calculus 000
Chapter 3. Subdifferential calculus and local regularity 000
§1. Subdifferential criteria of local metric regularity 000
§2. Regularity at a point and subdifferential calculus of non-Lipschitz functions 000
§3. Comments 000
Bibliography 000

To the memory of Lazar Aronovich Lyusternik on his centennial

Introduction

The concept of metric regularity of (set-valued) maps appeared in the end of 70-ies at the very initial stage of development of the branch of analysis that later became known as “non-smooth” analysis. The development of non-smooth analysis was basically stimulated by needs of optimization theory. But the sources of the
concept of metric regularity should be sought in classical theorems of the differenti- 
al calculus and linear analysis. Among them, the first to be mentioned is the Banach–Schauder open mapping theorem. Similar proofs of the theorem were given in the famous Banach’s treatise [9], which appeared in 1931 and a year later was translated into French, and in a slightly earlier Schauder’s note [92]. The follow- 
ing two equivalent formulations of the Banach–Schauder theorem are essential for understanding of the subsequent material.

Let $X$ and $Y$ be Banach spaces, and let $A: X \mapsto Y$ be a linear bounded operator sending $X$ onto $Y$. Then

(a) there is an $N > 0$ such that

$$B_Y \subset A(N B_X),$$  \hspace{1cm} (0.1)

which means that the unit ball in $Y$ is covered by the image of the ball of radius $N$ in $X$;

(b) there is a $K > 0$ such that for any $(x, y)$

$$d(x, A^{-1}(y)) \leq K \|y - Ax\|,$$  \hspace{1cm} (0.2)

where $A^{-1}(y)$ is the inverse image of $y$ under $A$ and $d( \cdot, \cdot )$ stands for distance.

It is not a difficult matter to verify that the lower bound of those $N$ for which (0.1) holds and the lower bound of those $K$ for which (0.2) is valid coincide and are equal to $\|(A^*)^{-1}\|$, the norm of the inverse to the adjoint of $A$.

By the guiding principle of the smooth non-linear analysis a map must locally satisfy properties of its derivative. As far as the Banach–Schauder theorem is concerned, this principle was embodied in two now already classical theorems: the tangent space theorem of Lyusternik (1934) [72] and the surjection theorem of Graves (1950) [43], each connected with one of the two given interpretations of the Banach–Schauder theorem. The standard statement of the Lyusternik theorem says that the tangent manifold to the $\overline{\mathcal{F}}$-level set of a continuously differentiable map $F: X \mapsto Y$ at a $\overline{\mathcal{F}}$ coincides with $\text{Ker} F'(\overline{\mathcal{F}})$, provided $F(\overline{\mathcal{F}}) = \overline{\mathcal{F}}$ and the regularity (or Lyusternik’s) condition $\text{Im} F'(\overline{\mathcal{F}}) = Y$ is satisfied. Meanwhile, Lyusternik’s proof contains a more precise statement, namely that there is a $K > 0$ such that

$$d(x, F^{-1}(\overline{\mathcal{F}})) \leq K d(\overline{\mathcal{F}}, F(x))$$  \hspace{1cm} (0.3)

for all $x \in \text{Ker} F'(\overline{\mathcal{F}}) + \overline{\mathcal{F}}$ sufficiently close to $\overline{\mathcal{F}}$.

In fact, Lyusternik’s proof can be easily modified to guarantee that (0.3) is valid for all $x$ close to $\overline{\mathcal{F}}$.

Lyusternik’s immediate goal in [72] was to prove a Lagrange multiplier rule in an abstract minimization problem:

$$\text{minimize } f(x) \text{ subject to } F(x) = 0.$$  

In general, the implicit function theorem, which plays the key role in proofs of the finite-dimensional Lagrange multiplier rule or the Euler–Lagrange equation in

---

1As I remember, A. A. Milyutin mentioned such a possibility in personal conversations already at the end of 60-ies.
the isoperimetric problem and the problem of Lagrange in the classical calculus of variations, can no longer be applied (because of the possible absence of a topological complement for $\text{Ker} \; F'(\overline{x})$). Lyusternik's regularity condition $\text{Im} \; F'(\overline{x}) = Y$ was the first general regularity condition on the constraint which was not connected with the specific structure of $F$. (Later, after the famous 1948 work of Kuhn and Tucker, such conditions got the name of constraint qualifications.) In this sense, Lyusternik's paper should be considered a precursor of the general theory of extremal problems whose development was triggered by Pontryagin’s maximum principle and subsequent studies of Dubovitskii–Milyutin and Neustadt.

The commonly accepted statement of the Graves theorem is the following: if $F: X \to Y$ is continuously differentiable at $\overline{x}$ and the regularity condition $\text{Im} \; F'(\overline{x}) = Y$ holds, then there is an $m > 0$ such that for any sufficiently small $t > 0$ the $F$-image of the ball of radius $t$ around $\overline{x}$ contains the ball of radius $mt$ around $\overline{y}$. Meanwhile, the main result of Graves which remained basically unnoticed\(^2\) contains a stronger statement: if $F$ is continuous and there are a linear bounded operator $A: X \to Y$, $m > \delta > 0$ such that $A$ satisfies (0.1) with $N = m^{-1}$ and

$$\|F(x) - F(x') - A(x - x')\| \leq \delta \|x - x'\| \quad (0.4)$$

for all $x$, $x'$ sufficiently close to $\overline{x}$, then the equation $F(x) = y$ has a solution $x$ with $\|x - \overline{x}\| \leq t$ if $\|y - F(\overline{x})\| < t(m - \delta)$ for any sufficiently small $t$. Much later, in the beginning of 70-ies, Ioffe and Tihomirov showed in [56] that, under the assumption of the Graves theorem, (0.3) holds for all $x$ sufficiently close to $\overline{x}$ with $K = (m - \delta)^{-1}$.

It is clear that the condition of the Graves theorem is satisfied at every point of a neighborhood of $\overline{x}$ if it is satisfied at $\overline{x}$. Therefore under the condition, the following estimate holds for all $(x, y)$ close to $(\overline{x}, \overline{y})$:\^6

$$d(x, F^{-1}(y)) \leq Kd(y, F(x)).\quad (0.5)$$

This is the property which is now called metric regularity of $F$ in a neighborhood of $\overline{x}$.

The Graves theorem is especially remarkable because of the absence of any differentiability requirement on $F$. In this respect it was much ahead of its time. We observe that (0.4) contains also the case of strict differentiability of $F$, explicitly considered only in 1974 [82]. This is the case when $\delta$ in the right-hand side of (0.4) is not a constant but a function of $(x, x')$ going to zero as both $x$ and $x'$ go to zero. (A formal definition of strict differentiability was given ten years after the paper of Graves—see [68].) We finally note that the proofs by Lyusternik and Graves, as well as Banach’s proof of the open mapping theorem, use similar iterative procedures.

Thus, to sum up the achievements of the “classical” stage, we can state a general result which can be naturally called the **Lyusternik–Graves theorem**: let $X, Y$ be Banach spaces, and let $F: X \to Y$ be strictly differentiable at $\overline{x}$ with $\text{Im} \; F'(\overline{x}) = Y$. Then there is a $K > 0$ such that

(a) $B(F(x), t) \subset F(B(x, Kt))$ for $x$ close to $\overline{x}$ and small $t > 0$;
(b) $d(x, F^{-1}(y)) \leq Kd(y, F(x))$ for $(x, y)$ close to $(\overline{x}, F(\overline{x}))$.

\(^2\)Up to a recent paper by Dontchev [28].
Moreover, the lower bound of those $K$ for which (a) and (b) hold is equal to $\|[(F'(x))^{-1}]^{-1}\|$. The Banach–Schauder open mapping theorem is an immediate consequence of the Lyusternik–Graves theorem. On the other hand, neither the Lyusternik theorem, nor the Graves theorem follow from the Banach–Schauder theorem and need independent proofs (although both Lyusternik and Graves do use the latter in their proofs). Both Lyusternik and Graves also remark that the implicit function theorem can be obtained from their results, hence also from the just stated Lyusternik–Graves theorem. Taking this into account, along with important applications which go beyond the framework of the implicit function theorem, first of all in optimization theory, we can qualify the Lyusternik–Graves theorem as the principal result of (smooth) non-linear analysis.

The next stage is characterized by a growing interest to more complicated objects than just everywhere defined maps and systems of equations. First to be mentioned are systems of equalities and inequalities defined either on the entire space or on more specialized sets, as say cones etc. The beginning of this stage should be probably dated back to publication in 1952 by Hoffmann [44] a paper that gave estimates for distances to sets of solutions of linear systems in finite-dimensional spaces. The central among the subsequent results are undoubtedly the theorem of Robinson-Ursescu [88], [95] extending the Banach–Schauder theorem to set-valued maps with convex graph (see Chapter 1, Theorem 4) and the following extension of the Lyusternik–Graves theorem obtained by Robinson [89]: Let $C \subset X$ be a convex set, let $G: X \to Y$ be strictly differentiable at $x \in C$, and let $K \subset Y$ be a closed convex cone with $0 \in G(x) + K$. Set

$$F(x) = \begin{cases} G(x) + K, & \text{if } x \in C; \\ \emptyset, & \text{otherwise.} \end{cases}$$

If

$$0 \in \text{int}(G(x) + G'(x)(C - x) + K),$$

then there is a neighborhood $U$ of $x$ and a neighborhood $V$ of zero in $Y$ such that

(a) for any $y \in V$ there is a solution to $y \in F(x)$ belonging to $U$;

(b) there is a $K > 0$ such that (0.5) holds\(^3\) for all $x \in U$, $y \in V$.

Theorems of Robinson–Ursescu and Robinson were the first application-oriented generalizations of fundamental results of analysis to set-valued maps. It was a breakthrough which had substantial psychological effect and which to a great extent determined the direction and the style of future research. With all that, both theorems belong to the realm of classical analysis since their statements and the machinery used in the proofs do not need anything beyond the classical differential calculus and linear analysis. For non-smooth analysis, being newly born around that time, these frameworks soon became very narrow.

---

\(^3\)Robinson proved a result which seems to be more general at a first glance. It can be interpreted as a variant of the implicit function theorem. But the result we have stated above and the theorem proved by Robinson [89] are equivalent in the sense that each of them can be obtained from the other.
In the survey we discuss certain outcomes of the “post-classical” studies connected with metric regularity and its applications. It seems that, in the most general terms, we have to mention as the main achievements the understanding of

A) metric nature of the phenomenon of metric regularity, not connected with any linear structures;

B) the equivalence of metric regularity to certain surjection (or openness) and Lipschitz properties of set-valued maps;

C) the central role of metric regularity in non-smooth analysis, similar to that of the Lyusternik–Graves theorem in the “smooth” analysis and in a certain sense even more significant. A fact of a matter is that, unlike in the classical differential calculus, the main theorems of the calculus of subdifferentials (such as theorems on subdifferentials of sums or compositions) require additional qualification assumptions which are always connected with metric regularity of some suitable set-valued maps (see Chapter 3).

The passage (from normed) to metric spaces resulted, in turn, in the understanding of an essential fact that from the point of view of regularity theory there is no difference between set-valued and single-valued maps, in the sense that the regularity or surjection problems for any set-valued map can be equivalently formulated in terms of a simple single-valued continuous map (see Chapter 1, Proposition 3). This is known already for some time but the prejudice about greater generality of set-valued theory is still widely spread. On the other hand, methodologically set-valued maps are often more convenient, basically because of the additional flexibility offered by (formally) more general formulations. Therefore in the survey we mainly consider set-valued maps.

It is interesting to observe that a similar evolution from differential manifolds and smooth functions to metric spaces and continuous functions occurred, already in the 90-ies, in the calculus of variations “in the large”, mainly in connection with the theory of Lyusternik-Schnirelman and its more recent variants and generalizations (see, for example, [24], [55], [62]).

The paper consists of three chapters. The first is devoted to metric regularity, the second to the theory of subdifferentials and the third to the connection between subdifferential calculus and metric regularity. The first and the third chapters are the main parts of the survey. The second chapter containing a summary of necessary results from the calculus of subdifferentials plays the role of a bridge between the other two.

The paper contains a number new results and, probably, even more new formulations and proofs. Basically, the entire theory presented in the first chapter is new. Its main conclusion is that the theory of metric regularity and a parallel Lipschitz theory of set-valued maps do not need anything like subdifferentials, directional derivatives, coderivatives and tangent cones. At the local level the adequate technical apparatus is based on the concept of strong slope introduced in 1980 by DeGiorgi, Marino and Tosques [26]. Strong slope is more universal than the above mentioned objects of local non-smooth analysis as it makes sense in every metric space. It is also, at least in principle, easier to calculate and, what is the main, it allows to obtain “uniformly” better estimates from which known estimates and
criteria can be obtained without much effort. However, in *applications* (for example, to necessary conditions in constrained optimization) sufficient subdifferential criteria for metric regularity and openness are extremely important.

In the first and the third chapters results, as a rule, are accompanied with complete proofs. These two chapters are concluded with sections of comments in which I have tried to describe, as completely as possible, interconnections between basic results and ideas. This has proved to be not an easy task as mutual ties, influence, authorship are often difficult to determine in the flow of publications on the subjects to be discussed. I apologize in advance for unavoidable omissions and imprecisions, hoping that they are not too many. Each chapter has autonomous numbering of sections, propositions and formulae. In cross references the first mentioned is the number of a chapter so that a reference in Chapter 3 to Theorem 1.3 (or formula (1.3)) sends to Theorem 3 of Chapter 1.

I was not personally acquainted with L.A. Lyusternik. But among my brightest guiding stars, I have to mention, without any doubt, the theory of Lyusternik–Schnirel’man and the Lyusternik–Graves theorem. With this paper I wish to express my gratitude to Lazar Aronovich Lyusternik.

**Chapter 1. The metric theory**

§1. Terminology, notation, definitions. Equivalence theorems

The main classes of objects to be studied in the paper are set-valued maps and extended real valued functions. Let $X$ and $Y$ be metric spaces and $F$ a set-valued map from $X$ into $Y$. (In the sequel we often use the symbol $F : X \rightrightarrows Y$.) This means that with every $x \in X$ we associate a set $F(x) \subseteq Y$ which can well be empty. If nothing else is said, we always assume set-valued maps closed-valued, that is to say, we assume that all sets $F(x)$ are closed. The sets

$$\text{dom} F = \{ x : F(x) \neq \emptyset \} \quad \text{and} \quad \text{Im} F = \bigcup_x F(x)$$

are respectively called the domain and image of $F$. We say that $F$ is *not proper* if either $\text{dom} F = \emptyset$ or $F(x) = Y$ on $\text{dom} F$. Otherwise $F$ is a proper map.

A set-valued map is often identified with its graph

$$\text{Gr} F = \{(x,y) \in X \times Y : y \in F(x)\}.$$

Therefore the set-valued map whose graph is the closure of $\text{Gr} F$ is called the closure of $F$. To denote the latter we use the symbol $\text{cl} F$. It is said that $F$ is closed if $F = \text{cl} F$.

The *inverse* to $F$ is naturally defined by

$$F^{-1}(y) = \{ x \in X : y \in F(x) \}.$$

Clearly, $\text{dom} F^{-1} = \text{Im} F$.

Suppose now that $f(x)$ is a function on $X$ with values in the extended real line $\mathbb{R} = [-\infty, \infty]$. Every such function defines a set-valued map

$$x \mapsto \{ \alpha \in \mathbb{R} : \alpha \geq f(x) \}.$$
The graph of this set-valued map is called the epigraph of $f$:

$$\text{epi } f = \{(x, \alpha) \in X \times \mathbb{R} : \alpha \geq f(x)\}.$$ 

Accordingly, we define

$$\text{dom } f = \{x \in X : |f(x)| < \infty\}$$

and call a function proper if $\text{dom } f \neq \emptyset$.

We use the same symbol $d(\cdot, \cdot)$ for the distance in any space. Hopefully, this will cause no confusion. In the product $X \times Y$ we usually consider the additive distance $d((x, y), (u, v)) = d(x, u) + d(y, v)$ or a more general $\alpha$-metric ($\alpha > 0$)

$$d_{\alpha}((x, y), (u, v)) = d(x, u) + \alpha d(y, v).$$

We shall also adopt the following conventions:

$$\inf \emptyset = \infty; \quad d(x, \emptyset) = \infty.$$

For the closed ball of radius $r$ around $x$ we use the symbol $B(x, r)$ and for the corresponding open ball the symbol $B^o(x, r)$. The unit ball in a normed space is denoted by $B_X$, or just $B$.

**Definition 1.** Let $V$ be a subset of $X \times Y$. We say that $F$ is metrically regular on $V$ if there is a $K > 0$ such that

$$(x, y) \in V \implies d(x, F^{-1}(y)) \leq K d(y, F(x)). \quad (1)$$

The smallest $K$ with which (1) holds will be called the norm of metric regularity of $F$ on $V$ which we shall denote by $\text{Reg}_V F$.

The following local version of metric regularity plays an important part in many applications.

**Definition 1 (loc).** $F$ is metrically regular near $(\overline{x}, \overline{y}) \in \text{Gr } F$ if for some $\varepsilon > 0$ it is metrically regular on the set $V = B(\overline{x}, \varepsilon) \times B(\overline{y}, \varepsilon)$. In other words, $F$ is metrically regular near $(\overline{x}, \overline{y})$ if (1) holds for all $(x, y)$ such that $d(x, \overline{x}) \leq \varepsilon$, $d(y, \overline{y}) \leq \varepsilon$.

The lower bound of such $K$ in this case will be called the norm of (metric) regularity of $F$ near $(\overline{x}, \overline{y})$. We shall denote it by $\text{Reg } F(\overline{x}, \overline{y})$ (or $\text{Reg } F(\overline{x})$ if $F$ is single-valued).

Local metric regularity can be equivalently defined with the help of smaller sets.

**Proposition 1.** A set-valued map is metrically regular near $(\overline{x}, \overline{y}) \in \text{Gr } F$ if and only if it is metrically regular on the set $V = \{(x, y) \in B(\overline{x}, \varepsilon) \times B(\overline{y}, \varepsilon) : d(y, F(x)) \leq \varepsilon\}$ for some $\varepsilon > 0$.

**Proof.** Clearly, $F$ is metrically regular on $V$ if it is regular near $(\overline{x}, \overline{y})$. Suppose that $F$ is metrically regular on $V$. Take $\delta > 0$ so small that $(K + 1)\delta < \varepsilon$, and let $d(x, \overline{x}) \leq \delta$, $d(y, \overline{y}) \leq \delta$. If $d(y, F(x)) \leq \varepsilon$ as well, then $d(x, F^{-1}(y)) \leq K d(y, F(x))$.


by the assumption. Otherwise we get the desired estimate, taking into account that
\(d(y, F(\overline{x})) \leq d(y, \overline{y}) < \varepsilon:\)

\[d(x, F^{-1}(y)) \leq d(\overline{x}, F^{-1}(y)) + d(x, \overline{x})\]
\[\leq Kd(y, F(\overline{x})) + d(x, \overline{x}) \leq Kd(y, \overline{y}) + d(x, \overline{x}) \leq \varepsilon \leq d(y, F(x)).\]

**Remark 1.** The latter property is often taken as the definition of local regularity. As we see, there is no need to assume a priori that \(d(y, F(x)) \leq \varepsilon\) but in many cases such an assumption is rather convenient.

The definitions have a very clear meaning: in case of a single-valued map they just mean that the distance from \(x\) to the \(y\)-level set of \(F\) is controlled by the norm of deviation of \(F(x)\) from \(y\).

In what follows, speaking about regularity, we shall usually omit the adjective “metric” as no other type of regularity is considered in the survey. If \(V\) has the form \(P \times Y\) or \(X \times Q\) where \(P \subset X,\ Q \subset Y\), we say that \(F\) is regular on \(P\) or on \(Q\) and do not mention \(V\) at all if it is the whole \(X \times Y\). Finally, we say that \(F\) is regular in a neighborhood of \(\overline{x}\) if \(F\) is single-valued on \(\text{dom}\ F\).

With every set-valued map \(F\) we associate two families of set-valued maps:

\[
F_t(x) = \{ y : d(y, F(x)) < t \};
\]
\[
F^t(x) = F(B(x, t)) = \bigcup_{d(x, u) \leq t} F(u)
\]

**Definition 2.** We say that \(F\) covers on \(V\) at a linear rate if there is a \(K > 0\) such that

\[(x, y) \in V, v \in F(x) & d(v, y) < t \implies \exists u : d(u, x) \leq Kt & y \in F(u).\]  

(2)

The lower bound of those \(K\) for which (2) holds will be called the norm of covering (or surjection) of \(F\) on \(V\), and the inverse of the norm the constant of covering. To denote the latter we shall use the symbol \(\text{Sur}_V F\).

If \(V = P \times Q\), where \(P \subset X\) and \(Q \subset Y\), then (2) simply means that for any \(x \in P\) the \(t\)-neighborhood of \(F(x) \cap Q\) is contained in the \(F\)-image of the \(Kt\)-neighborhood of \(x\).

The local version of the definition is obvious: \(F\) covers at a linear rate near \((\overline{x}, \overline{y})\) if for some \(\varepsilon > 0\), (2) is satisfied for all \((x, y)\) with \(d(x, \overline{x}) \leq \varepsilon, d(y, \overline{y}) \leq \varepsilon\). We can verify, in a way similar to Proposition 1, that in the local definition we can restrict ourselves to \(v\) satisfying \(d(v, \overline{y}) \leq \varepsilon\).

The lower bound of those \(K\) with which the local condition holds is called the norm of covering near \((\overline{x}, \overline{y})\). By \(\text{Sur}(\overline{x}, \overline{y})\) we denote the inverse to the norm of covering near \((\overline{x}, \overline{y})\) which is called the constant of covering (or constant of surjection) of \(F\) near \((\overline{x}, \overline{y})\).

**Definition 3.** Let \(W \subset X \times Y\). We say that \(F\) is pseudo-Lipschitz on \(W\) if there is a \(K > 0\) such that

\[(x, y) \in W, & y \in F(u) \implies d(y, F(x)) \leq Kd(x, u).\]  

(3)
The smallest $K$ for which (3) holds is called the pseudo-Lipschitz norm of $F$ on $W$. We can equivalently rewrite (3) in the form

$$Kd(x,u) < t \implies F(u) \cap W_x \subset F_t(x),$$

where $W_x = \{y : (x,y) \in W\}$. If $W = S \times Y$, $S \subset X$ we get the standard Lipschitz condition for set-valued maps.

We say that $F$ is pseudo-Lipschitz near $(\bar{x}, \bar{y})$ if there are $K > 0$ and $\varepsilon > 0$ such that (3) holds for all $(x,y) \in W = B(\bar{x}, \varepsilon) \times B(\bar{y}, \varepsilon)$. If $Y$ is a normed linear space this means that

$$F(u) \cap B(\bar{y}, \varepsilon) \subset F(x) + Kd(x,u)B$$

if $d(x, \bar{x}) \leq \varepsilon$, $d(y, \bar{y}) \leq \varepsilon$. The lower bound of all such $K$ is called the pseudo-Lipschitz norm of $F$ near $(\bar{x}, \bar{y})$.

The following proposition shows that in all three definitions we basically speak about the same phenomenon. (That is why we do not need any special notation for the pseudo-Lipschitz norm).

**Proposition 2.** The following statements are equivalent:

(a) $F$ is regular on $V$;  
(b) $F$ covers on $V$ at a linear rate;  
(c) $F^{-1}$ is pseudo-Lipschitz on $W = V^{-1}\{(y, x) : (x,y) \in V\}$.

Moreover, the norms of regularity and covering of $F$ on $V$ and the pseudo-Lipschitz norm of $F^{-1}$ on $W$ are equal. In particular, $\text{Reg}_V F \cdot \text{Sur}_V F = 1$.

**Proof.** (a) and (c) are equivalent because (3) applied to $F^{-1}$ is precisely (1) (since (3) does not contain any restrictions on $u$).

(a)⇒(b). Suppose (1) holds with some $K$. Let $(x,y) \in V$, $v \subset nF(x)$ and $d(y,v) < t$. By (1) $d(x, F^{-1}(y)) \leq Kd(y, F(x)) \leq d(y,v) < Kt$. This means that there is an $u$ such that $d(x,u) < Kt$ and $y \in F(u)$.

(b)⇒(a). Suppose (2) holds with some $K > 0$. Let $(x,y) \in V$ and $d(y, F(x)) = \xi$. Take an arbitrary $t > \xi$. By (2) there is an $u \in X$ such that $d(x,u) < Kt$ and $y \in F(u)$. This means that $d(x, F^{-1}(y)) \leq Kt$. The latter is valid for every $t > \xi = d(y, F(x))$. Therefore $d(x, F^{-1}(y)) \leq Kd(\bar{y}, F(\bar{x}))$.

**Remark 2.** As follows from the definition of covering at a linear rate, the open $t$-neighborhood of $v$ is covered by the image of the closed $Kn$-neighborhood of $x$. It follows from the proof of Proposition 2 that metric regularity is equivalent to the property that can be defined as “open covering”, namely that the open $t$-neighborhood of $v$ is covered by the image of the open ball of radius $Kt$ around $x$. Of course this property is equivalent to covering at a linear rate of Definition 2.

By analogy, we can consider also “closed covering” when the closed $t$-neighborhood of $v$ is covered by the image of the closed ball of radius $Kt$ around $x$. It is clear that closed covering implies covering at a linear rate in the sense of Definition 2 and hence open covering.

Returning back to the proof of Proposition 2, we can notice that, with a slight modification, it implies closed covering as well when $F(x)$ contains an element nearest to $y$ for any $(x,y) \in V$, in particular when the values of $F$ are compact sets. In this case the proof of Proposition 2 allows to conclude that regularity is
equivalent to closed covering and, as a consequence, that all three types of covering are equivalent.

The next question we intend to discuss in this section concerns connection between metric regularity properties for single-valued and set-valued maps. Let $F$ and $V$ be as above, and let $\Pi_Y: \text{Gr} F \mapsto Y$ stand for the Cartesian projection (that is $\Pi_Y ((x,y)) = y$). We emphasize that we consider it only on the graph of $F$. We further consider a set $W = \{(x,y,v) : (x,v) \in \text{Gr} F, (x,y) \in V \}$.

**Proposition 3.** The following conditions are equivalent:
(a) $F$ is regular (covers) on $V$ with norm not greater than $K$;
(b) $\Pi_Y$ is regular (covers) on $W$ with norm not greater than $K + \alpha$ if $\text{Gr} F$ is considered with the $\alpha$-metric.

In other words, the study of the regularity or covering properties for a set-valued maps can always be reduced to the study of the corresponding property for a simple single-valued map, namely, the projection of $\text{Gr} F$ onto $Y$.

**Proof.** We shall check the implication (a) $\Rightarrow$ (b) for regularity and the converse for covering.
(a) $\Rightarrow$ (b). Let $(x,v) \in \text{Gr} F, (x,y) \in V$, that is $(x,v,y) \in W$, and let $d(y,v) < t$. Since $F$ is regular on $V$, there is a $u \in X$ such that $y \in F(u)$ and $d(u,x) < Kt$. This means that $\Pi_Y (u,y) = y$ and $d_n ((x,v),(u,y)) = d(x,u) + ad(y,v) \leq (K + \alpha) t$, whence (b).

(b) $\Rightarrow$ (a) Let $(x,y) \in V$ and $d(y,F(x)) = \tau$. Take an arbitrary $t > \tau$ and find a $v \in F(x)$ with $d(y,v) < t$. By definition $(x,v,y) \in W$. As $\Pi_Y$ covers on $W$ at a linear rate, for any $N$ greater than the norm of covering there is a pair $(u,w) \in \text{Gr} F$ such that $y = \Pi_Y (u,w)$ (that is $y = w$) and $Nt \geq d_n ((x,v),(u,y)) = d(x,u) + ad(y,v)$. It follows that $d(x,u) \leq Nt - \alpha \tau = (N - \alpha(\tau/t)) t$. As the latter holds for any $t > \tau$, we conclude that $F$ covers on $V$ with the norm not greater than $N - \alpha$.

Local regularity admits still another convenient equivalent description. Let us say that $F$ is graph-regular on $V$ with norm not greater than $K$ if
\[
d(x,F^{-1}(y)) \leq d_K ((x,y), \text{Gr} F) \quad \text{for all} \quad (x,y) \in V.
\]

The definition is specified for graph-regularity near $(\overline{x}, \overline{y}) \in \text{Gr} F$ in an obvious way.

**Proposition 4.** (a) If $F$ is graph-regular on $V$ with norm not greater than $K$, then it is regular on $V$ with norm not greater than $K$;
(b) If $F$ is regular near $(\overline{x}, \overline{y}) \in \text{Gr} F$ then it is graph-regular near $(\overline{x}, \overline{y})$ with the same norm. Thus, local regularity and graph regularity are equivalent.

**Proof.** (a) Let $(x,y) \in V$. Then for any $(u,v) \in \text{Gr} F$
\[
d(x,F^{-1}(y)) \leq d(x,u) + Kd(y,v),
\]
and it is sufficient to set $u = x$ and and take the infimum over $v \in F(u)$ in the right-hand side.
(b) Suppose that (1) holds for all $x$ and $y$ of the $\varepsilon$-neighborhoods of $\mathcal{F}$ and $\mathcal{G}$ respectively. Let $\delta > 0$ be so small that
\[
d_K((x, y), \text{Gr} F) = \inf \{d_K((x, y), (u, v)) : d(u, \mathcal{F}) \leq \varepsilon, d(v, \mathcal{G}) \leq \varepsilon, v \in F(u) \}
\]
for all $(x, y)$ with $d(x, \mathcal{F}) < \delta$, $d(y, \mathcal{G}) < \delta$. Such a $\delta$ clearly exists as $\mathcal{G} \in F(\mathcal{F})$. For any $(u, v) \in \text{Gr} F$ with $d(u, \mathcal{F}) \leq \varepsilon$, $d(v, \mathcal{G}) \leq \varepsilon$ and any $(x, y)$ sufficiently close to $(\mathcal{F}, \mathcal{G})$ we have:
\[
d(x, F^{-1}(y)) \leq d(u, x) + d(u, F^{-1}(y)) \leq d(u, x) + K d(y, F(u))
\]
\[
\leq d(u, x) + K d(y, v) = d_K((x, y), (u, v)).
\]
We get the desired result by comparing this with the preceding equality assuming that $d(x, \mathcal{F}) < \delta$ and $d(y, \mathcal{G}) < \delta$.

Finally, prior to studying criteria for metric regularity, we shall discuss the question about natural assumptions on the very spaces $X$ and $Y$. As a rule, one or another completeness assumption is necessary for the proof of any criterion. Typical requirements include either completeness of $X$ alone or completeness of $\text{Gr} F$. (We observe that both in the Banach–Schauder open mapping theorem and in the Lyusternik–Graves theorem both spaces were initially assumed complete.) The following proposition clarifies the situation.

**Proposition 5.** Let $X$ and $Y$ be metric spaces, let $F : X \rightrightarrows Y$ be a closed set-valued map, and let $V \subset X \times Y$. Let further $\hat{X}, \hat{Y}, \hat{F} : \hat{X} \rightrightarrows \hat{Y}$ and $\hat{V} \subset \hat{X} \times \hat{Y}$ be respectively the completions of $X$ and $Y$, the set-valued map whose graph is the closure of $\text{Gr} F$ in $\hat{X} \times \hat{Y}$ and the closure of $V$ in $\hat{X} \times \hat{Y}$. Assume finally that $\text{Gr} F \subset V$ and $F$ and $V$ satisfy the following compatibility condition:

If $(x_n, v_n) \in \text{Gr} F$, $(x_n, v_n) \to (\hat{x}, \hat{v}) \in \text{Gr} \hat{F}$ and $(\hat{x}, \hat{y}) \in \hat{V},$

then there are $y_n \in Y$ such that $(x_n, y_n) \in V, y_n \to \hat{y}$.

Under this assumptions in each of the following two cases, $\hat{F}$ covers on $\hat{V}$ if and only if $F$ covers on $V$ and the constants of covering of $\hat{F}$ and $F$ coincide:

(a) Gr $F$ is complete;

(b) $X$ is a complete space.

**Proof.** In both cases it is clear that $F$ covers on $V$ if $\hat{F}$ covers on $\hat{V}$. We further observe that (a) reduces to (b) by means of Proposition 3. Thus, only the case (b) should be considered.

So assume that $F$ covers on $V$ with norm not greater than $K$. Let $(\hat{x}, \hat{v}) \in \text{Gr} \hat{F}$ and $(\hat{x}, \hat{y}) \in \hat{V}$. By the compatibility condition there is a sequence of triples $(x_n, y_n, v_n) \to (\hat{x}, \hat{v}, \hat{v})$ such that $(x_n, v_n) \in \text{Gr} F$ and $(x_n, y_n) \in V$. Take a $t > d(\hat{y}, \hat{v})$. Then $t > d(y_n, v_n) + 2^{-n}$ if $n$ is sufficiently large. Since $F$ covers on $V$ with norm not greater than $K$, there are $u_n \in X$ such that

\[y_n \in F(u_n) \quad \text{and} \quad d(x_n, u_n) \leq K(d(y_n, v_n) + 2^{-2n}).\]
Set $\beta_n = d(x_n, \hat{x}) \to 0$. Taking, if necessary a subsequence, we may assume without loss of generality that $\sum d(y_{n+1}, y_n) < \infty$. Then

$$
\alpha_n = \sum_{m=n}^{\infty} d(y_{m+1}, y_m) \to 0.
$$

By the assumption $(u_n, y_n) \in V$. Therefore for any $n$ and any $i = 1, 2, \ldots$ we can choose an $u_{ni} \in X$ such that

$$
y_{n+i} \in F(u_{ni}) \quad \text{and} \quad d(u_{n(i+1)}, u_{ni}) \leq K(x(n+i+1), y_{n+i}) + 2^{-2(n+i+1)}.
$$

It is clear that $\{u_{ni}\}_{i=1}^{\infty}$ is a Cauchy sequence. Therefore it converges in $X$ to a certain $\hat{w}_n$ for which we have $d(u_n, \hat{w}_n) \leq K(\alpha_n + 2^{-2n})$ and $\hat{y} \in F(\hat{w}_n)$. If $n$ is sufficiently large, then

$$
d(\hat{w}_n, \hat{x}) \leq d(\hat{w}_n, x_n) + \beta_n \leq d(\hat{w}_n, u_n) + (u_n, x_n) + \beta_n
$$

$$
\leq K(d(y_n, v_n) + \alpha_n + 2^{-2(n-1)}) + \beta_n < Kt
$$

which completes the proof.

§2. **General and local regularity criteria**

To avoid discussions of inessential technical details, we assume in all statements and proofs in this section that all metric spaces are complete. Although certain results remain valid without the assumption, Proposition 5 shows that it is not very restrictive.

Let $f$ be a proper function on $X$. We set

$$
M_0 = M_0(f) = \{x : f(x) \leq 0\}, \quad M_+ = M_+(f) = \{x : f(x) > 0\}.
$$

**Basic Lemma.** Let $X$ be a complete metric space and $f$ a lower semicontinuous proper function on $X$. Assume that there are $\overline{x} \in \text{dom } f$, $K > 0$ and $r \in (0, \infty]$ such that $K f(\overline{x}) < r$ and for any $t > K$, $x \in M_+ \cap B(\overline{x}, r)$ there is an $u \in X$, $u \neq x$ such that

$$
f^+(u) \leq f(x) - t^{-1}d(x, u).
$$

Then $M_0 \neq \emptyset$ and

$$
d(x, M_0) \leq K f^+(\overline{x})
$$

(where $\alpha^+ = \max\{0, \alpha\}$).

**Proof.** There is nothing to prove if $\overline{x} \in M_0$. So we assume that $f(\overline{x}) > 0$ and set $\varepsilon = f(\overline{x})$. Let $t > K$ satisfy $t \varepsilon < r$. Applying Ekeland’s variational principle [33][4] to $f^+(x) = \max\{0, f(x)\}$ with $\lambda = \varepsilon t$ we can find a $w \in X$ such that

$$
d(\overline{x}, w) \leq \lambda; \quad f^+(z) + (\varepsilon / \lambda) d(z, w) > f^+(w), \quad \forall z \neq w, \quad \forall z \neq w.
$$

---

[4] If $f$ is a lower semicontinuous and bounded below proper function on a complete metric space, then for any $\overline{x}$ with $f(\overline{x}) \leq \inf f + \varepsilon$ and any $\lambda > 0$ there is a $w \in X$ with the following three properties: $f(w) \leq f(\overline{x}) - (\varepsilon / \lambda) d(\overline{x}, w)$; $d(w, \overline{x}) \leq \lambda$ and $f(x) - (\varepsilon / \lambda) d(x, w) > f(w)$ for any $x \neq w$. 

If \( f(w) > 0 \), then by the assumption (since \( w \in B(\overline{\mathcal{X}}, r) \)) there is a \( u \neq w \) such that \( f(u) + t^{-1}d(u, w) \leq f(w) \) which in impossible according to (4). Therefore \( u \in M_0 \) and consequently \( d(\overline{\mathcal{X}}, M_0) \leq d(\overline{\mathcal{X}}, w) \leq \lambda \) by the first inequality in (4). The last inequality is valid for all \( \lambda > \varepsilon K \), hence \( d(x, M_0) \leq K \varepsilon = K f(\overline{\mathcal{X}}) \). The lemma has been proved.

Let us return back to \( F \) and consider the function \( f_y(x) = d(y, F(x)) \). This is a non-negative function and its zero level set is precisely \( F^{-1}(y) \). Therefore for this function the inequality in the Basic Lemma reduces to the regularity inequality (1) for the given \( y \). This observation can easily be turned into an exact statement about regularity of \( F \).

**Theorem 1.** Let \( F: X \rightrightarrows Y \) be a closed set-valued map, and let \( V \) be a closed subset of \( X \times Y \) containing \( \text{Gr} F \). Then \( F \) is regular on \( V \) if and only if there are \( m > 0 \) and \( \alpha > 0 \) such that for any triple \( (x, y, v) \) satisfying

\[
(x, y) \in V, \quad y \neq v, \quad v \in F(x),
\]

and any \( m' < m \) there is a pair \( (u, w) \neq (x, v) \) such that \( w \in F(u) \) and

\[
d(y, w) \leq d(y, v) - m' (d(u, x) + \alpha d(v, w)).
\]

Moreover, the norm of regularity of \( F \) on \( V \) does not exceed \( (m^{-1} - \alpha) \).

**Proof.** We shall consider \( X \times Y \) along with the \( \alpha \)-metric. Let us consider the function

\[
\varphi_y(x, v) = \begin{cases} 
    d(y, v), & \text{if } (x, y) \in V, \ v \in F(x); \\
    \infty, & \text{otherwise}. 
\end{cases}
\]

It is lower semicontinuous since the graph of \( F \) and \( V \) are closed sets. The condition of the theorem means that for any \( (x, v) \) and \( m' < m \) we can find a pair \( (u, w) \) such that

\[
\varphi_y^+(u, w) = \varphi_y(u, w) \leq \varphi_y(x, v) - m' d_\alpha((x, v), (u, w)).
\]

We have furthermore (since \( \text{Gr} F \subset V \))

\[
M_0(\varphi_y) = \{(x, v) : (x, y) \in V, y = v \in F(x)\} = F^{-1}(y) \times \{y\}.
\]

Therefore \( d_\alpha((x, v), M_0) = d(x, F^{-1}(y)) + \alpha d(y, v) \) and applying Basic Lemma with \( r = \infty \), we conclude that

\[
d(x, F^{-1}(y)) \leq (m^{-1} - \alpha) d(y, v)
\]

for all \( (x, y, v) \) satisfying (5). But (5) does not impose any restrictions on \( v \) in addition to \( v \in F(x) \). Therefore the last inequality implies that

\[
d(x, F^{-1}(y)) \leq (m^{-1} - \alpha) d(y, F(x)),
\]

which means that \( F \) is regular on \( V \) with norm not exceeding \( (m^{-1} - \alpha) \).

The proof of the converse, namely that regularity implies the condition of the theorem, is elementary: if the last inequality is satisfied and \( m' < m \), then we take an \( u \in F^{-1}(y) \) with \( d(x, u) \leq ((m')^{-1} - \alpha) d(y, v) \) and setting \( w = y \), get \( 0 = d(y, w) \leq d(y, v) - m' (d(x, u) + \alpha d(v, w)) \).

We state below several variants of Theorem 1 which are proved in a similar way. Let us start with a slightly more flexible sufficient regularity condition (which in general may not be necessary). Let us call a function \( \omega(x, v) \) a *test function* for \( f \) if it is lower semicontinuous, non-negative and equals zero if and only if \( v \in F(x) \).
Theorem 1a.  Let $F$ and $V$ be the same as in Theorem 1, and let $\omega(x,v)$ be a test function for $F$.  Suppose further that there are $m > 0$, $\alpha > 0$ such that $m^{-1} > \alpha$ and for any $(x,y) \in V$, $y \neq v \in Y$ there is a pair $(u,w) \neq (x,v)$ such that $(u,y) \in V$ and

$$d(u,w) + \omega(u,w) \leq d(y,v) + \omega(x,v) - m d_\alpha((x,v),(u,w)).$$

Then $F$ is regular on $V$ with norm not greater than $(m^{-1} - \alpha)$.

In the proof of this theorem one has to replace $\varphi_y$ by the function $d(y,v) + \omega(x,v) + \chi_V(x,y) + \chi_{\text{Gr} F}(x,v)$.

Theorem 1b. Assume in addition to the condition of Theorem 1 that for any $y \in Y$ the function $y \mapsto d(y,F(x))$ is lower semicontinuous. Then $F$ is regular on $V$ with constant $K$ if and only if for any $(x,y) \in V$ with $y \not\in F(x)$ and any $t > K$ there is an $u \neq x$ such that

$$(u,y) \in V \quad \& \quad d(y,F(u)) \leq d(y,F(x)) - t^{-1}d(x,u).$$

In this case the proof is also similar to the proof of Theorem 1. The only difference is that instead of $\varphi_y$ we can use a simpler function

$$\psi_y(x) = \begin{cases} d(y,F(x)), & \text{if } (x,y) \in V; \\ \infty, & \text{otherwise.} \end{cases}$$

Of course, Theorem 1b is equivalent to Theorem 1 in view of Proposition 3.

Finally, the third theorem contains a necessary and sufficient condition for graph regularity of $F$ on $V$.

Theorem 1c. Under the assumptions of Theorem 1, $F$ is graph-regular on $V$ with norm not exceeding $K$ if and only if for any $(x,y) \in V$, $y \not\in F(x)$ and $t < 1$ there is an $u \neq x$ such that

$$(u,y) \in V \quad \& \quad d_K((u,y),\text{Gr} F) \leq d_K((x,y),\text{Gr} F) - td(x,u).$$

In this case we have to consider (instead of $\varphi_y$) the function

$$\psi_y(x) = \begin{cases} d_K((x,y),\text{Gr} F), & \text{if } (x,y) \in V; \\ \infty, & \text{otherwise.} \end{cases}$$

Then $M_0(\psi_y) = F^{-1}(y)$ and graph-regularity of $F$ coincides with what is obtained after the application of Basic Lemma to the condition of the theorem.

Remark 3. Ekeland’s principle still works if only a certain part of the space containing $\text{Gr} F$ is complete. Therefore only completeness of $\text{Gr} F$ is needed for Theorem 1. Theorems 1a and 1c remain valid if $V$ is complete and Theorem 1b is valid when $X$ is complete and $Y$ is an arbitrary metric space.

To state local results we need one more concept which plays an important role in the sequel.
Definition 4. Let $x \in \text{dom } f$. The quantity

$$|\nabla f|(x) = \limsup_{u \to x} \frac{(f(x) - f(u))^+}{d(x, u)}$$

is called the (strong) slope of $f$ at $x$.

For example, for a Fréchet differentiable function on a normed linear space, $|\nabla f|(x)$ is equal to the norm of $f'(x)$.

Theorem 2. Let $(\bar{x}, \bar{y}) \in X \times Y$, and let positive numbers $m > 0$ and $r > 0$ satisfy $mr > d(\bar{y}, F(\bar{x}))$. Assume further that for any $x \in B(\bar{x}, r)$, $y \in B(\bar{y}, r)$, $v \in B(\bar{y}, r)$ with $y \neq v \in F(x)$ and any sufficiently small $\alpha > 0$

$$|\nabla \varphi_y|(x, v) \geq m,$$

where $\varphi_y$ is the function defined by (6) for $V = X \times Y$ and its strong slope is defined with respect to the $\alpha$-metric in $X \times Y$.

Then the inequality

$$d(x, F^{-1}(y)) \leq m^{-1}d(y, F(x))$$

holds for any $(x, y) \in B(\bar{x}, r) \times B(\bar{y}, r)$ such that $d(y, F(x)) < m(r - d(x, \bar{x}))$. Thus, if $(\bar{x}, \bar{y}) \in \text{Gr } F$, then $F$ is regular near $(\bar{x}, \bar{y})$ with norm not greater than $m^{-1}$.

Moreover, if $Y$ is a Banach space, then the stated condition is also necessary for regularity of $F$ near $(\bar{x}, \bar{y})$. Namely, if $F$ is regular near $(\bar{x}, \bar{y})$ with norm not exceeding $K$, then for any $x$ of a small neighborhood of $\bar{x}$, all $y, \neq v \in F(x)$ of a small neighborhood of $\bar{y}$ and all $\alpha > 0$ the strong slope of $\varphi_y$ at $(x, v)$ with respect to $d_\alpha$ satisfies

$$|\nabla \varphi_y|(x, v) \geq (K + \alpha)^{-1}.$$ 

Proof. Under the condition of the first part of the theorem, for any $(x, y, v)$ satisfying the condition and any sufficiently small $\alpha > 0$ and $\gamma > 0$ there is a pair $(u, w) \neq (x, v)$ such that

$$\frac{d(y, v) - d(y, w)}{d(u, x) + \alpha d(v, w)} \geq m - \gamma = m',$$

that is

$$d(y, w) \leq d(y, v) - m'(d(u, x) + \alpha d(v, w)).$$

If, in addition, $d(y, F(x)) < (m' - \alpha)(r - d(x, \bar{x}))$, then the desired inequality follows from Basic Lemma as in the proof of Theorem 1, this time however with a finite $r$, if we take into account that $\alpha$ and $\gamma$ can be arbitrarily small. For $(\bar{x}, \bar{y}) \in \text{Gr } F$ this means, in view of Proposition 1, that $F$ is regular near $(\bar{x}, \bar{y})$. This proves the first part of the theorem.

To prove the second, take $x, y, v$ sufficiently close to $\bar{x}$ and $\bar{y}$ respectively and such that $y \neq v \in F(x)$. Let $v_n = (1 - (1/n))v + (1/n)y$. By Proposition 2, $F$
covers near \((\bar{x}, \bar{y})\) with norm not greater than \(K\). Therefore for any \(n\) there is an \(u_n\) such that \(v_n \in F(u_n)\) and \(d(u_n, x) \leq K\|v_n - v\| \to 0\) when \(n \to \infty\).

We have \(\|y - v\| = \|y - v_n\| + \|v - v_n\|\). Therefore

\[
\nabla \varphi_y(x, v) \geq \lim_{n \to \infty} \frac{\varphi_y(x, v) - \varphi_y(u_n, v_n)}{d(u_n, x) + \alpha \|v_n - v\|} \geq \lim_{n \to \infty} \frac{\|v_n - v\|}{(K + \alpha)\|v_n - v\|} = (K + \alpha)^{-1}.
\]

The theorem has been proved.

As in the general case, we can get local regularity criteria, sometimes more convenient for calculation, using other functions. In particular, local versions of Theorems 1a and 1b need only minor changes in the proof of Theorem 2. The following is the statement of a local version of Theorem 1b.

**Theorem 2b.** Let \((\bar{x}, \bar{y}) \in X \times Y\), and let positive \(K\) and \(r\) be such that \(r > Kd(\bar{y}, F(\bar{x}))\). Assume that the functions \(x \mapsto \psi_y(x) = d(y, F(x))\) are lower semi-continuous near \(\bar{x}\) and for all \(x \in B(\bar{x}, r)\), \(y \in B(\bar{y}, r)\) such that \(y \not\in F(x)\)

\[
|\nabla \psi_y|(x) > K^{-1}\]

Then the inequality

\[
d(x, F^{-1}(y)) \leq Kd(y, F(x))
\]

holds for any \((x, y) \in B(\bar{x}, r) \times B(\bar{y}, r)\) satisfying \(Kd(y, F(x)) < r - d(x, \bar{x})\).

In particular, if \((\bar{x}, \bar{y}) \in \text{Gr } F\) then a sufficient condition for \(F\) to be regular near \((\bar{x}, \bar{y})\) is that for any \(t > K\) there is a neighborhood of \((\bar{x}, \bar{y})\) such that \(|\nabla \psi_y|(x) > t^{-1}\) for all \((x, y)\) of the neighborhood not belonging to the graph of \(F\). This condition is also necessary if \(Y\) is a Banach space.

We conclude the section by proving a local criterion connected with the distance to the graph of \(F\), that is to say, a local version of Theorem 1c. This version, however, differs from Theorem 2 both by its statement and proof.

**Definition 5.** Let \(F: X \rightrightarrows Y\) and \((\bar{x}, \bar{y}) \in \text{Gr } F\). For any \(K > 0\) and \(y \in Y\) we define the function \(g^K_y(x) = d_K((x, y), \text{Gr } F)\). We further set for any \(\varepsilon > 0\)

\[
U_\varepsilon = \{(x, y) : d((x, y), (\bar{x}, \bar{y})) < \varepsilon, y \not\in F(x)\}
\]

and define the quantity

\[
|\nabla F|_K(\bar{x}, \bar{y}) = \lim_{\varepsilon \to 0} \inf_{(x, y) \in U_\varepsilon} |\nabla g^K_y|(x).
\]

which will be called the (limiting) \(K\)-slope of \(F\) at \((\bar{x}, \bar{y})\).

**Theorem 3.** Let \(F: X \rightrightarrows Y\) be a set-valued map with closed graph, and let \((\bar{x}, \bar{y}) \in \text{Gr } F\). A sufficient condition for \(F\) to be regular near \((\bar{x}, \bar{y})\) is that

\[
|\nabla F|_K(\bar{x}, \bar{y}) \geq 1.
\]

This condition is also necessary if both spaces \(X\) and \(Y\) are Banach.

**Proof.** If \(U_\varepsilon = \emptyset\), the \(K\)-slope of \(F\) at \((\bar{x}, \bar{y})\) is \(\infty\) by definition. On the other hand, as follows from the definition of local covering, \(F\) in this case covers with constant
greater than any fixed number. So the theorem is trivially true if \( U_e = \emptyset \) and in the rest of the proof we assume that \( U_e \neq \emptyset \) for any positive \( \varepsilon \).

**Sufficiency.** Suppose (8) holds. Then for any \( t > 1 \) we can choose positive \( \gamma \) and \( \delta < \gamma / 2 \) to make sure that

\[
|\nabla g^K_y(x)| > t^{-1}, \quad tg^K_y(x) < \gamma / 2
\]

for all \( x \) and \( y \) of the \( \delta \)-neighborhoods of (\( \overline{\mathcal{F}}, \mathcal{J} \)). As \( U_e \neq \emptyset \), we can choose \( x \) and \( y \) for which \( g^K_y(x) > 0 \). Applying to \( g^K_y \) the variational principle of Ekeland with \( \varepsilon = g^K_y(x) \) and \( \lambda = tg^K_y(x) \), we find an \( x \in X \) such that

\[
d(w, x) \leq tg^K_y(x) \quad \text{and} \quad g^K_y(u) + t^{-1}d(w, u) > g^K_y(w), \quad \forall u \neq w. \tag{9}
\]

It is clear that \( d(w, x) < \gamma \). Therefore, having assumed that \( g^K_y(w) > 0 \) we would get \( |\nabla g^K_y(x)| > t^{-1} \). This means that in every neighborhood of \( w \) there is an \( u \neq w \) such that \( g^K_y(u) \leq g^K_y(w) - t^{-1}d(w, u) \) in contradiction with (9). Thus \( g^K_y(w) = 0 \).

It is clear furthermore that

\[
M_0(g^K_y) = \{ u : y \in F(u) \} = F^{-1}(y),
\]

so that \( d(x, F^{-1}(y)) \leq d(x, w) \leq td_K((x, y), \text{Gr} f) \). This inequality is valid for every \( t > 1 \), hence for \( t = 1 \). This completes the proof of sufficiency.

**Necessity.** Suppose now that \( X \) and \( Y \) are Banach spaces and \( F \) is regular (hence covers) near \( (\overline{\mathcal{F}}, \mathcal{J}) \) with norm not greater than \( K \). Let us take again a \( t > 1 \) and choose a neighborhood of \( (\overline{\mathcal{F}}, \mathcal{J}) \) to make sure that (0.5) is valid for all points of the neighborhood with \( tK \) instead of \( K \). Let \( (x, y) \) belong to the chosen neighborhood and \( y \not\in F(x) \), that is, \( d_K((x, y), \text{Gr} F) > 0 \). Let finally \( (x_n, y_n) \in \text{Gr} F \) be such that

\[
\|x - x_n\| + K\|y - y_n\| = d_K((x, y), (x_n, y_n)) \leq d_K((x, y), \text{Gr} F) + n^{-2}.
\]

We can consider two possible situations.

(a) \( \|y - y_n\| \to 0 \). Then \( \|x - x_n\| \geq \alpha > 0 \) for all \( n \). Set \( u_n = (1 - n^{-1})x + n^{-1}x_n \). Then \( \|u_n - x_n\| = (1 - n^{-1})\|x - x_n\| \) and

\[
\frac{1}{n^2\|x - u_n\|} = \frac{1}{n\|x - x_n\|} \leq \frac{1}{n\alpha} \to 0.
\]

It follows that

\[
\lim_{n \to \infty} \frac{g^K_y(x) - g^K_y(u_n)}{\|x - u_n\|} = \lim_{n \to \infty} \frac{g^K_y(x) + n^{-2} - g^K_y(u_n)}{\|x - u_n\|} \geq \lim_{n \to \infty} \frac{\|x - x_n\| - \|u_n - x_n\|}{\|x - u_n\|} = 1.
\]

(b) \( \|y - y_n\| \geq \alpha > 0 \) for sufficiently large \( n \). Set \( v_n = (1 - n^{-1})y_n + n^{-1}y \).
Since $F$ covers near $(\overline{x}, \overline{y})$ with norm not greater than $K$, we can find a $w_n \in X$ such that $v_n \in F_w(w_n)$ and $\|w_n - x_n\| \leqslant tK\|v_n - y_n\|$. Set $u_n = x + w_n - x_n$. Then $u_n - w_n = x - x_n$ and therefore

$$d_K((x, y), (x_n, y_n)) - d_K((u_n, y), (w_n, v_n)) = K\|y - y_n\| - K\|y - v_n\|
\leqslant n^{-1}K\|y - y_n\| \leqslant n^{-1}K\alpha. \quad (10)$$

It follows that for large $n$

$$g^K_y(x) - g^K_y(u_n) \geqslant d_K((x, y), (x_n, y_n)) - n^{-2} - d_K((u_n, y), (w_n, v_n))
\geqslant \alpha K/n - 1/n^2 > 0.$$ 

Taking into account that $\|y_n - v_n\| = n^{-1}\|y - y_n\| \sim n^{-1}$, we get from the last inequality and (10)

$$\limsup_{n \to \infty} \frac{g^K_y(x) - g^K_y(u_n)}{\|x - u_n\|} = \limsup_{n \to \infty} \frac{g^K_y(x) + n^{-2} - g^K_y(u_n)}{\|x - u_n\|}
\geqslant \limsup_{n \to \infty} \frac{d_K((x, y), (x_n, y_n)) - d_K((u_n, y), (w_n, v_n))}{tK\|y_n - v_n\|} = \frac{1}{t}.$$

Thus, for any $(x, y)$ sufficiently close to $(\overline{x}, \overline{y})$ the strong slope of $f_y(\cdot)$ at $x$ is bounded below by a number arbitrarily close to 1. This completes the proof of the theorem.

**Remark 4.** As in the non-local case, we notice that Theorem 2 remains valid if only $\operatorname{Gr} F$ is complete, while Theorem 2b needs completeness of $X$ alone. We also observe that, as immediately follows from the proofs, the second (necessity) parts of Theorems 2 and 3 without change extend to the case when $Y$ has the following geodesic property:

$$\forall x, y \quad \forall \varepsilon > 0 \quad \exists z : \quad \max \{d(x, z), d(y, z)\} \leqslant \frac{1}{2}(d(x, y) + \varepsilon).$$

**§3. Applications and extensions**

In all cases we are aware of, no extreme effort is needed to get more special criteria or sufficient conditions for regularity, openness or pseudo-Lipschitz property from the general theorems of the previous section. Below we give several examples connected with a number of widely known results.

**3.1. Robinson–Ursesku theorem.** We already mentioned that this theorem is a generalization of the Banach–Schauder open mapping theorem to convex set-valued maps which by definition are set-valued maps with convex graphs. As follows from Proposition 2, the regularity problem for convex set-valued maps can be reduced to the regularity problem for restrictions of linear bounded operators to convex closed sets. Therefore the statement given below is equivalent to the theorem of Robinson-Ursesku.
Theorem 4. Let $X$ and $Y$ be Banach spaces, let $A: X \mapsto Y$ be a linear bounded operator, let $S \subset X$ be a convex closed set, let $\overline{x} \in S$, and let $C_{\overline{x}}$ be the cone generated by the set $S - \overline{x}$. If $A(C_{\overline{x}}) = Y$, then the restriction of $A$ to $S$ is regular near $\overline{x}$.

Proof. As we are interested in regularity on a neighborhood, we can assume without loss of generality that $S$ is a bounded set, say $\text{diam}S \leq 1$. Set $Q = A(S - \overline{x})$, where the bar stands for the closure. Obviously, the cone generated by $Q$ contains $A(C_{\overline{x}})$ and therefore coincides with $Y$. Applying the Baire category theorem, we conclude that the interior of $Q$ is nonempty and contains zero. Therefore there is a $\delta > 0$ such that $Q$ contains the ball of radius $2\delta$.

In particular, $B(y, \delta) \subset Q$ whenever $\|y\| \leq \delta$. Since $S$ is convex and $A$ is linear, it follows that for any $x \in U = \{x : \|A(x - \overline{x})\| < \delta\}$ and any $\xi$ of a dense subset of $\delta B$ there is an $u \in S$ such that $Au = Ax + \xi$ and $\|u - x\| \leq \|\xi\|/\delta$. Take an arbitrary $x \in U$, set $v = Ax$, and let $y \neq v$, $\xi = \delta\|y - v\|^{-1}(y - v)$. Set further $w(\lambda) = v + \lambda\xi$, where $\lambda \in (0, 1)$. Then $\|w(\lambda) - Au(\lambda)\| = o(\lambda)$ for some $u(\lambda)$ such that $\|u(\lambda) - x\| \leq \lambda$. On the other hand,

$$\|y - w(\lambda)\| = \|y - v\| - \lambda\delta \leq \|y - v\| - \delta\|u(\lambda) - x\|,$$

so that

$$\limsup_{u \to x} \frac{\|y - Ax\| - \|y - Au\|}{\|u - x\|} \geq \delta \geq \lim_{\lambda \to 0} \frac{\|y - v\| - \|y - w(\lambda)\| - \|w(\lambda) - Au(\lambda)\|}{\|u(\lambda) - x\|}$$

(as $\|u(\lambda) - x\| = O(\lambda)$). Applying Theorem 2, we get the result.

3.2. Covering on a system of balls. Let $\rho(x)$ be a nonnegative function on $X$ satisfying the Lipschitz condition with constant one:

$$|\rho(x) - \rho(u)| \leq d(x, u),$$

and let $\Sigma$ be the collection of balls $B(x, \rho)$ with $0 \leq \rho \leq \rho(x)$. Following [27], we write $(x, \rho) \in \Sigma$ if the ball $B(x, \rho)$ belongs to $\Sigma$.

We say that $F$ covers on $\Sigma$ with constant not smaller than $m$ if

$$F_{mt}(x) \subset F^t(x)$$

for all $x$ and $t \leq \rho(x)$.

Theorem 5. Let $F: X \mapsto Y$ be a closed set-valued map. Assume that there are numbers $a > b > 0$ such that $F^r(x)$ is an $rb$-net for $F_{ar}(x)$ for any $(x, \rho) \in \Sigma$, where $r$ is defined by $ar = (a - b)\rho$. Then $F$ covers on $\Sigma$ with constant not smaller than $m = a - b$.

Proof. Assume first that $F$ is single-valued and continuous. Set

$$V = \bigcup_{(x, \rho) \in \Sigma} \{x\} \times B(F(x), m\rho) = \bigcup_x \{x\} \times B(F(x), m\rho(x)).$$
It is obvious that $V$ is closed (as $F$ is continuous). Let further $(x, y) \in V, y \neq F(x)$. Then $y \in B(F(x), m\rho)$ for some $\rho \leq \rho(x)$. Define $r$ by the equality $d(y, F(x)) = ar$. By the assumption (as $ar \leq m\rho$) there is an $u \in X$ such that $d(u, x) \leq r$ and $d(y, F(u)) \leq br = ar - mr \leq d(y, F(x)) - md(x, u)$. The desired result now follows from Theorem 1b. Thus, the theorem is true for single-valued maps.

The general case is now easily proved with the help of Proposition 3. Indeed, consider the projection $\Pi_Y$ of the graph of $F$ onto $Y$. By the assumption, for any $(x, y, v)$ with $v \in F(x), d(y, v) \leq ar$ we can find a pair $(u, w)$ such that $d(u, x) \leq r$ and $d(y, w) \leq br$. Take an arbitrary $\delta > 0$ and consider in $\text{Gr} F$ the $\alpha$-distance $d_{\alpha}((x, v), (u, w)) = d(x, u) + \alpha d(v, w)$, where $\alpha = \delta/(a + b)$. Then $d_{\alpha}((x, v), (u, w)) \leq (1 + \delta)r = cr$. This means that the projection of the ball of radius $cr$ in $\text{Gr} F$ onto $Y$ is a $br$-net in the ball $B(y, ar)$. We therefore arrive at the situation considered in the first part of the proof but with $a/c$ and $b/c$ instead of $a$ and $b$. It follows that $\Pi_Y$ covers with constant not smaller than $(1 + \delta)^{-1}(a - b)$ on the family $\Sigma'$ of balls $B((x, v), \rho')$ in $\text{Gr} F$ such that $(x, (1 + \delta)^{-1}\rho') \in \Sigma$. It remains to apply Proposition 3 and take into account that $\delta$ can be arbitrarily small.

Remark 5. If $d(y, F(x))$ is lower semicontinuous for all $y$, we can drop the assumption that $Y$ is complete and use Theorem 1b.

3.3. Perturbation theorems. It is well known that a small linear perturbation of a linear bounded operator onto remains an operator onto. The theorems below extend this to set-valued maps.

Theorem 6. Let $F$ and $\Phi$ be closed set-valued maps from $X$ into $Y$. Set $F_1(x) = F(x) + \Phi(x)$. Suppose that $F$ covers with constant $a$ and $\Phi$ satisfies the Lipschitz condition with constant $b$ (in the sense that $h(\Phi(x), \Phi(x')) \leq bd(x, x'), h(C, D)$ being the Hausdorff distance between $C$ and $D$), where $a > b > 0$. Then $F_1$ covers with constant not smaller than $a - b$.

It may happen that the sets $F_1(x)$ are not closed. Therefore Theorem 6 formally does not follow from Theorem 1. However, with a simple trick similar to that used in the proof of Proposition 1, we can get Theorem 6 from Theorem 5 (and hence from Theorem 1). We shall also give another proof using a suitable modification of the “Lyusternik” iteration process.

Proof 1. We first consider the case when $F$ is single-valued and continuous on $\text{dom} F$. In this case $\text{Gr} F_1$ is closed, hence complete, and a simple calculation shows that $F_1$ satisfies the condition of Theorem 5, that is $(F_1)^r$ is a $br$-net for $(F_1)_{ar}(x)$ for all $x$ and $r$. Thus, in this case Theorem 6 follows from Theorem 5.

In the general case when both maps are set-valued, we consider the set $W = \text{Gr} F$ with the $\alpha$-metric $d_{\alpha}((x, y), (x', y')) = d(x, x') + \alpha d(y, y')$ and a set-valued map $H(x, y) = y + \Phi(x)$ from $W$ into $Y$. By Proposition 3, the first component of the map covers with constant $a_1 = (a^{-1} + \alpha)^{-1}$, whereas the second obviously satisfies the Lipschitz condition with constant $b$. If $\alpha$ is sufficiently small, then $a_1 > b$ and applying the just proved result for a single-valued first component, we conclude that $H$ covers with constant not smaller than $a_1 - b$. Using the arguments similar to those in the second part of the proof of Proposition 3, we come to a conclusion that $F_1$ covers with constant not smaller than $[(a_1 - b)^{-1} - \alpha]^{-1}$. The latter is true
for any $\alpha$, hence $F_1$ covers with constant not smaller than $\lim_{n \to 0} [(a_1-b)^{-1} -\alpha]^{-1} = a - b$.

**Proof 2.** Take $b' \in (b, a)$ and set $\theta = b'/a < 1$. We choose next a $(\bar{x}, \bar{y}) \in \text{Gr} F$, a

\[ \bar{x} \in \Phi(\bar{x}) \text{ and } \bar{y} \in Y \text{ and set } x_0 = \bar{x}, \ y_0 = \bar{y}, \ w_0 = \bar{x}, \ v_0 = \bar{y} + \bar{x} - y, \ t_0 = \|v_0\| . \]

Suppose we have already found $x_n, \ y_n, \ w_n, \ v_n$. Then $x_{n+1}, \ y_{n+1}, \ w_{n+1}, \ v_{n+1}$ are

chosen as follows:

\[
\begin{align*}
y_{n+1} &= y_n - v_n; \quad t_n = \|v_n\|; \\
x_{n+1} &= F^{-1}(y_{n+1}); \quad d(x_{n+1}, x_n) \leq t_n / a; \\
w_{n+1} &= \Phi(x_{n+1}); \quad \|w_{n+1} - w_n\| \leq b' d(x_{n+1}, x_n); \\
v_{n+1} &= w_{n+1} - w_n.
\end{align*}
\]

Such a choice is possible (not unique) thanks to the assumptions: we can chose an $x_{n+1}$ since $F$ covers with constant $a$, and we can choose $w_{n+1}$ since $\Phi$ is Lipschitz with constant $b < b'$. We have

\[
d(x_{n+1}, x_n) \leq t_n / a \leq \theta^{n+1} t_0 / a; \\
\|v_{n+1}\| \leq b' d(x_{n+1}, x_n) \to 0.
\]

The first inequality shows that $\{x_n\}$ is a fundamental sequence. Therefore it converges to some $x$ such that

\[
d(x, \bar{x}) \leq a^{-1} t_0 \sum_{n=0}^{\infty} \theta^n = (a - b')^{-1} t_0 = (a - b')^{-1} d(y, \bar{y} + \bar{x}). \tag{11}
\]

Furthermore

\[
w_{n+1} = \sum_{i=1}^{n+1} v_i; \quad y_{n+1} = y - \sum_{i=1}^{n} v_i,
\]

so that

\[
F(x_{n+1}) + \Phi(x_{n+1}) \ni y_{n+1} + w_{n+1} = y + v_{n+1} \to y.
\]

On the other hand,

\[
\|w_{n+1} - w_n\| = \|v_{n+1}\| \leq \theta^n t_0,
\]

which shows that $\{w_n\}$ is also a fundamental sequence converging to a certain $w$. But this implies that $\{y_n\}$ also converges to some $z$ and $z$ and $w$ are clearly such that $z + w = y$, and $z \in F(x)$ and $w \in \Phi(x)$ since the graph of each map is closed. Thus $y \in F(x) + \Phi(x)$ and a comparison with (11) concludes the proof (as $b'$ can be arbitrarily close to $b$).

It is not a difficult matter to modify the proof of the theorem to get the following perturbation result for local covering.
Theorem 6a. Suppose under the condition of Theorem 6 that $\Phi$ is a single-valued map such that the following holds in a neighborhood of $\overline{\mathcal{X}}$:

$$d(\Phi(x), \Phi(u)) \leq r(x, u)d(x, u), \quad \text{where} \quad r(x, u) \to 0 \quad \text{as} \quad x, u \to \overline{\mathcal{X}}.$$ 

Then the norms of regularity and covering of $\mathcal{F}$ and $\mathcal{F}_1 = F + \Phi$ near $(\overline{\mathcal{X}}, \overline{\mathcal{Y}})$ and $(\overline{\mathcal{X}}, \overline{\mathcal{Y}} + \overline{\Phi(\mathcal{X})})$ respectively are equal (as are the pseudo-Lipschitz norms of the inverse maps).

3.4. A general implicit function theorem. Let $F : X \times Y \rightarrow Z$. Given an $x$, we denote by $F_x$ the set-valued map $y \mapsto F(x, y)$. Accordingly, $F_x^{-1}$ is the inverse of $F_x$. In what follows we assume that $\overline{\mathcal{X}} \in F(\overline{\mathcal{X}}, \overline{\mathcal{Y}})$ and denote the function $y \mapsto d(z, F(x, y))$ by $\varphi_x^z(y)$.

Lemma 1. Suppose there are $K > 0$ and $L > 0$ such that

(a) $d(z, F(x, y)) - d(z, F(u, y)) \leq Ld(x, u)$;
(b) $d(y, F_x^{-1}(x, z)) \leq Kd(z, F(x, y))$

for all $x$ and $u$ of a neighborhood of $\overline{\mathcal{X}}$, all $y$ of a neighborhood of $\overline{\mathcal{Y}}$, and all $z$ of a neighborhood of $\overline{\mathcal{Z}}$.

Then the mapping $(x, z) \mapsto F_x^{-1}(z)$ is pseudo-Lipschitz in a neighborhood of $((\overline{\mathcal{X}}, \overline{\mathcal{Z}}), \overline{\mathcal{Y}})$.

The first condition actually means that for any $y$ sufficiently close to $\overline{\mathcal{Y}}$ the map $x \mapsto F(x, y)$ is pseudo-Lipschitz with norm not exceeding $L$ in a neighborhood of every point of $\text{Gr} F$ close to $(\overline{\mathcal{X}}, \overline{\mathcal{Z}})$. The second condition means that the map $y \mapsto F(x, y)$ is regular with norm not exceeding $K$ near $(\overline{\mathcal{Y}}, \overline{\mathcal{Z}})$ if $x$ is sufficiently close to $\overline{\mathcal{X}}$.

Proof. We have to prove that there are $\delta > 0$ and $N > 0$ such that

$$F_x^{-1}(z) \cap B(\overline{\mathcal{Y}}, \delta) \subset (F_x^{-1})_N((x, z), (x', z'))(z') \quad (12)$$

for all $x, x', z$ satisfying $d(x, \overline{\mathcal{X}}) < \delta$, $d(x', \overline{\mathcal{X}}) < \delta$, $d(z, \overline{\mathcal{Z}}) < \delta$.

Consider a set-valued mapping from $X \times Y$ into $X \times Z$ defined by

$$\Phi(x, y) = \{x\} \times F(x, y).$$

We claim that it is regular near $((\overline{\mathcal{X}}, \overline{\mathcal{Y}}), (\overline{\mathcal{X}}, \overline{\mathcal{Z}}))$. Indeed,

$$\Phi^{-1}(x, z) = (x, F_x^{-1}(z)). \quad (13)$$

If we introduce the additive metrics with $\alpha = 1$ in $X \times Y$ and $X \times Z$, then

$$d((x, y), \Phi^{-1}(u, z)) = d(x, u) + d(y, F_u^{-1}(z));$$
$$d((u, z), \Phi(x, y)) = d(x, u) + d(z, F(x, y)).$$
So using (a) and (b), we get
\[
d((x, y), \Phi^{-1}(u, z)) = d(x, u) + d(y, F_u^{-1}(z))
\leq d(x, u) + Kd(z, F(u, y))
\leq d(x, u) + Kd(z, F(x, y)) + KLd(x, u)
\leq (1 + K + KL)d((u, z), \Phi(x, y)),
\]
which proves that $\Phi$ is indeed regular near $((x, \gamma), (x, \overline{\gamma}))$ with norm not exceeding $N = 1 + K + KL$.

By Proposition 2, $\Phi^{-1}$ is pseudo-Lipschitz in a neighborhood of $(x, \overline{\gamma})$, that is there is a $\gamma > 0$ such that
\[
\Phi^{-1}(x, z) \cap B((x, \overline{\gamma}), \gamma) \subset \Phi^{-1}_{d((x, z), (x', z'))}(x', z').
\]
Together with (13), this immediately implies (12) if we take, for instance, $\delta = \gamma/2$.

**Lemma 2.** Suppose $F$ is a closed set-valued map. Suppose further that there are $\delta > 0$ and $K > 0$ such that for any $z \in B(\overline{z}, \delta)$

(i) the function $x \mapsto d(z, F(x, y))$ is upper semicontinuous at $\overline{x}$ for any $y$ of the $\delta$-neighborhood of $\overline{x}$;

(ii) the function $y \mapsto d(z, F(x, y))$ is lower semicontinuous in the $\delta$-neighborhood of $\overline{y}$ for any $x$ of a neighborhood of $\overline{x}$;

(iii) $|\nabla \phi^x(y)| \geq K^{-1}$ for any $y$ of the $\delta$-neighborhood of $\overline{y}$, provided $z \not\in F_x(y)$. Then there is a neighborhood $\mathcal{O} \subset X$ of $\overline{x}$ and a $\gamma > 0$ such that
\[
F_x^{-1}(z) \cap B(\overline{y}, \gamma/2) \neq \emptyset
\]
and
\[
d(y, F_x^{-1}(z)) \leq Kd(z, F_x(y))
\]
for any $x \in \mathcal{O}, \ y \in B(\overline{y}, \gamma), \ z \in B(\overline{z}, \gamma)$.

**Proof.** By (i) for any $\varepsilon > 0$ there is a neighborhood $\mathcal{O} = \mathcal{O}(\varepsilon)$ of $\overline{x}$ such that $d(z, F(x, \overline{y})) < \varepsilon$ if $d(z, \overline{z}) < \varepsilon/2, x \in \mathcal{O}$. Choose a positive $\varepsilon < \delta/2K$. Applying Theorem 2b to $F_x$ and taking (iii) into account we get
\[
d(\overline{y}, F_x^{-1}(z)) \leq Kd(z, F(x, \overline{y})) \leq K\varepsilon < \delta/2
\]
which means that $F_x^{-1}(z) \cap B(\overline{y}, \delta/2) \neq \emptyset$.

Assume now that $d(y, \overline{y}) \leq \delta/2, \ d(z, F(x, y)) < \varepsilon$. In this case, as above, we can be sure, by virtue of Theorem 2b, that $d(y, F_x^{-1}(z)) \leq Kd(z, F(x, y))$. It remains to take $\gamma \leq (1 + K)^{-1} \min\{\varepsilon/2, \delta/2\}$ and refer to Proposition 1 (or rather to its proof) keeping in mind that our argument is equally applicable to any $x \in \mathcal{O}$.

Combining Lemmas 1 and 2, we get the following metric version of the implicit function theorem.
Theorem 7. Let $F : X \times Y \rightrightarrows Z$ be a closed set-valued map and $\bar{z} \in F(\bar{x}, \bar{y})$. Assume that there are $\delta > 0$, $K > 0$ such that 

(i) the function $x \mapsto F(x, y)$ is Lipschitz in the $\delta$-neighborhood of $\bar{x}$ for all $y, z$ of the $\delta$-neighborhoods of $\bar{y}$ and $\bar{z}$ respectively;

(ii) the function $y \mapsto F(x, y)$ is lower semicontinuous in the $\delta$-neighborhood of $\bar{y}$ for any $x, z$ of the $\delta$-neighborhoods of $\bar{x}$ and $\bar{z}$ respectively;

(iii) for any $x, y, z$ of the $\delta$-neighborhoods of $\bar{x}, \bar{y}, \bar{z}$ respectively such that $z \not\in F(x, y)$ the strong slope of the function $v \mapsto d(z, F(x, v))$ at $y$ is not smaller than $K^{-1}$.

Then there is a $\gamma > 0$ such that $(F^{-1}_{x}(z) \neq \emptyset$ and

$$d(y, F^{-1}_{x}(z)) \leq K d(z, F(x, y))$$

for all $x, y, z$ of the $\gamma$-neighborhoods of $\bar{x}, \bar{y}, \bar{z}$ respectively and the set-valued map $(x, z) \mapsto F^{-1}_{x}(z)$ is pseudo-Lipschitz near $(\bar{x}, \bar{z})$.

Corollary 7.1. Suppose, in addition to the assumptions of the theorem, that $F(x, y) \cap F(x', y) = \emptyset$ for all $y$ and $x \neq x'$. Then there is a single-valued map $\Phi : (x, z) \mapsto \Phi(x, z)$ from $X \times Z$ into $Y$ satisfying the Lipschitz condition and such that the relations $z \in F(x, y)$ and $\Phi(x, z) = y$ are equivalent.

Remark 6. As easily follows from the proof of Lemma 2, condition (ii) can be omitted if we use Theorem 2 instead of Theorem 2b and accordingly change the statements of Lemma 2 and Theorem 7.

3.5. **Regularity of order $\rho(\cdot)$**. **Moduli of regularity and covering.** Each of the properties we have considered (metric regularity, covering at a linear rate, pseudo-Lipschitz property) is characterized by a proportional dependence of certain quantities: metric regularity suggests a distance estimate proportional to the deviation from the target set, covering at a linear rate means that the radius of a covered ball in the range space is proportional to the radius of the covering ball in the domain space, and the pseudo-Lipschitz property (as well as the standard Lipschitz property) offers an estimate for the deviation proportional to the distance between arguments. It seems natural to try to consider estimates of a more general kind, as say power, exponential etc. As a result, we get a theory which is very similar to its “linear” predecessor and practically does not need new proofs.

All functions denoted below in this subsection by small Greek letters are defined on the half-line of non-negative reals, non-negative, non-decreasing, equal to zero at zero and strictly positive for positive value of their arguments. For any such function $\rho(t)$ we denote by $\underline{\rho}(t)$ the greatest lower semicontinuous function majorized by $\rho(t)$ and by $\overline{\rho}(t)$ the smallest upper semicontinuous function majorizing $\rho(t)$:

$$\underline{\rho}(t) = \lim_{\tau \to t^-} \rho(\tau); \quad \overline{\rho}(t) = \lim_{\tau \to t^+} \rho(\tau).$$

We shall also consider the upper and lower inverse functions

$$\underline{\rho}^{-1}(t) = \sup\{\xi : \rho(\xi) < t\}; \quad \overline{\rho}^{-1}(t) = \inf\{\xi : \rho(\xi) > t\}.$$ 

In both cases the upper and lower functions obviously coincide everywhere except at possibly countably many points.
Definition 6. We say that
(a) $F$ is $\rho$-regular on $V \subset X \times Y$ if
\[(x, y) \in V, \ d(y, F(x)) < t \implies d(x, F^{-1}(y)) \leq \rho(t);\]
(b) $F$ $\mu$-covers on $V$ if
\[(x, y) \in V, \ v \in F(x), \ d(y, v) < \mu(t) \implies \exists u \in X : d(u, x) \leq t, \ y \in F(u);\]
(c) $F$ is $\delta$-Hölder on $W \subset X \times Y$ if
\[(x, y) \in W, \ y \in F(u), \ d(x, u) < t \implies d(y, F(x)) \leq \delta(t).\]

The functions $\rho$, $\mu$ and $\delta$ in the definitions are not assumed continuous and even semicontinuous. If $\rho$ or $\delta$ are continuous, the strict inequalities can be replaced by non-strict in the corresponding definitions. On the other hand, the definition permits us to assume $\rho$ and $\delta$ lower semicontinuous without any loss of generality.

Proposition 6.
(a) If $F$ is $\rho$-regular on $V$ then $F$ $\mu$-covers on $V$ with $\mu(t) = \overline{(\rho^{-1})(t)}$;
(b) If $F$ $\mu$-covers on $V$, then $F$ is $\rho$-regular on $V$ with $\rho(t) = \overline{(\mu^{-1})(t)}$, and hence with $\rho(t) = \overline{(\mu^{-1})(t)}$;
(c) $F$ is $\rho$-regular on $V$ if and only if $F^{-1}$ is $\rho$-Hölder on $W = \{(y, x) \in Y \times X : (x, y) \in V\}$.

The proof of the proposition does not differ much from the proof of Proposition 2. Let us prove, for instance, the second statement. Let $(x, y) \in V$, $v \in F(x)$, and $d(y, v) < t$. Then $d(y, v) < \mu(\xi)$ if $\mu(\xi) > t$. Since $F$ $\mu$-covers, it follows that there is an $u \in X$ such that $y \in F(u)$ and $d(x, u) \leq \xi$. This means that $d(x, F^{-1}(y)) \leq \xi$ for any $\xi$ for which $\mu(\xi) > t$. Therefore $d(x, F^{-1}(y)) \leq \overline{(\mu^{-1})(t)}$.

Theorem 8. Let $F : X \Rightarrow Y$ be a closed set-valued map, and let $V \subset X \times Y$ be a closed set containing $\text{Gr} F$. Then $F$ is $\rho$-regular on $V$, provided there is an $\alpha > 0$ such that for any triple $(x, y, v)$ satisfying (5) there is a $(u, w) \neq (x, v)$ such that
\[\rho(d(y, w)) \leq \rho(d(y, v)) - (d(x, u) + \alpha d(v, w)).\]

The proof of the theorem repeats almost word for word the proof of sufficiency in Theorem 1. The difference is that instead of $\varphi_y$ we have to apply Basic Lemma to
\[\rho(d(y, v)) = \begin{cases} \rho(d(y, v)), & \text{if } (x, y) \in V, \ v \in F(x); \\ \infty, & \text{otherwise}, \end{cases} \tag{14}\]
and afterwards pass from the inequality
\[d(x, F^{-1}(y)) + \alpha d(y, v) \leq \rho \cdot \varphi_y(x, v),\]
to
\[d(x, F^{-1}(y)) \leq (\rho \cdot \varphi_y)(x, v).\]
(This explains, by the way, why the condition of the theorem is only sufficient but not necessary.)
**Definition 7.** The functions

\[ \lambda(t) = \inf \{ \eta > 0 : d(x, F^{-1}(y)) \leq \eta, \forall (x, y) \in V, d(y, F(x)) < t \} \]

and

\[ \omega(t) = \sup \{ r > 0 : y \in F^t(x), \forall (x, y) \in V, v \in F(x), d(y, v) < r \} \]

are called the *modulus of regularity* and the *modulus of surjection* of \( F \) on \( V \). We shall denote them by \( \text{reg}_V F \) and \( \text{sur}_V F \).

It is clear that \( \lambda(t) \leq \rho(t) \) for all \( t \) if \( F \) is \( \rho \)-regular on \( V \). It follows that \( \lambda(t) \) is lower semicontinuous. Likewise, \( \mu(t) \leq \omega(t) \) if \( F \) \( \mu \)-covers on \( V \). Therefore we get from Proposition 6.

**Proposition 7.**

\[ \lambda(t) = \sup \{ \xi : \omega(\xi) \leq t \} \]

In particular, if one of the moduli is strictly monotone and continuous, then so is the other one and the functions are mutually inverse.

A local version of the “nonlinear theory” is constructed in an obvious way: we have to require in the definitions of \( \rho \)-regularity and \( \mu \)-covering that the corresponding property be satisfied for \( (x, y) \) of a small neighborhood of \( (\bar{x}, \bar{y}) \). We can use Theorem 8 to get a local sufficient regularity criterion just like Theorem 2 was deduced from Theorem 1 in the linear case.

**Theorem 9.** Suppose as usual that \( F : X \rightrightarrows Y \) is a closed set-valued map and \( \bar{y} \in F(\bar{x}) \). Assume further that \( \rho(t) \) is lower semicontinuous

\[ |\nabla (\rho \circ \varphi_y)| (x, v) \geq 1 \]

(where the slope is calculated with respect to an additive metric in \( X \times Y \)) for all \( x, y, v \) of \( \varepsilon \)-neighborhoods of \( \bar{x}, \bar{y}, \bar{y} \) respectively and such that \( v \in F(x), \ y \neq v \). Then \( F \) is \( \rho \)-regular near \( (\bar{x}, \bar{y}) \).

**3.6. Regularity at a point and optimization problems.** In certain cases, in particular for the theory of necessary optimality conditions, weaker properties, which are usually defined as regularity and covering at a point, appear to be very useful. Namely, it is said that \( F \) is regular at \( (\bar{x}, \bar{y}) \in \text{Gr} F \) if there is a \( K > 0 \) such that (0.3) holds for all \( x \) of a neighborhood of \( \bar{x} \) and that \( F \) covers at \( (\bar{x}, \bar{y}) \) if there is a \( r > 0 \) such that \( B(\bar{y}, rt) \subset F(B(\bar{x}, t)) \) for all sufficiently small \( t \). It is clear that these two properties are no longer equivalent, unlike the earlier considered metric regularity and covering at a linear rate in a neighborhood. (It is sufficient to notice that in the first of them \( F \) does not necessarily covers a neighborhood of \( \bar{x} \).) We also observe that regularity and covering at a point do not reduce to regularity and covering on a one-point set in the sense of Definitions 1 and 2.

Let us consider the problem

\[ \text{minimize } f(x) \text{ subject to } 0 \in F(x), \]

where \( f \) is a function on \( X \), and \( F \) is a set-valued mapping from \( X \) into \( Y \).
Theorem 10. Assume that \( \mathfrak{T} \) solves the problem. If \( f \) satisfies the Lipschitz condition in a neighborhood of \( \mathfrak{T} \) and \( F' \) is regular at \( (\mathfrak{T}, 0) \), then there is an \( N > 0 \) such that \( \mathfrak{T} \) is an unconditional local minimum of \( f(x) + Nd(0, F(x)) \).

Proof. By the assumptions there are \( K > 0, L > 0 \) such that

\[
f(x) - f(x') \leq Ld(x, x'), \quad d(x, F^{-1}(0)) \leq Kd(0, F(x))
\]

for all \( x, x' \) sufficiently close to \( \mathfrak{T} \). Take such an \( x \), and let \( u \in F^{-1}(\overline{y}) \) is such that \( d(x, u) \leq (K + 1)d(0, F(x)) \). Then

\[
f(x) \geq f(u) - Ld(x, u) \geq f(\mathfrak{T}) - Ld(x, F^{-1}) \geq f(x) - L(K + 1)d(0, F(x)),
\]

that is \( f(x) + L(K + 1)d(0, F(x)) \geq f(\mathfrak{T}) \).

§4. Comments

4.1. For the first time the expression “metric regularity” (which seems to be very appropriate) probably appeared in Penot’s paper [83] in 1989. But the very word “regular”, or rather its Russian equivalent “pravil’nyi” is already present in the paper by Lyusternik: a map is regular at a point in Im \( F'(x) = Y \). In the same sense the word “regularity” was used in [56]. It became widely accepted after Robinson’s papers [88], [89] in which regularity was defined as a covering-type condition for local approximations. Robinson’s condition is equivalent to the Mangasarian–Fromovitz constraint qualification condition if the set-valued map is the solution set for a system of equalities and inequalities (as a function, say of their right-hand parts).

Finally, a definition in terms of a distance estimate similar to (0.5) was given in [45]. The definition of regularity with respect to \( V \) is introduced in this paper as an attempt to get a unified scheme that would include both the case of “complete system of balls” considered in [27] and sets like \( \{(x, y) : d(y, F(x)) \leq \delta\} \).

4.2. Banach already knew that the open mapping and closed graph theorems are valid for linear maps of Fréchet spaces. Ptak [86], [87], using an elegant modification of the Banach iteration scheme, extended the closed graph theorem to set-valued maps from a complete metric to a metric space. Ptak’s theorem is connected with the results discussed in this paper approximately like Banach–Schauder open mapping theorem with the theorem of Lyusternik–Graves: it does not imply Theorem 5 and the subsequent results (including the Robinson–Uršesku theorem) which, however, can be proved with the help of modifications of Ptak’s method. On the other hand, Ptak’s theorem can be easily derived from Theorem 8 (but not Theorem 5).

An important observation by Dmitruk, Milyutin, and Osmolovskii, later independently made also by Aubin [1], is that there is no need to a priori require that the image of a ball be dense in a suitable ball in \( Y \) (as say in Ptak’s theorem): the covering property actually follows from a weaker assumption that image points can be found sufficiently close to any point of the latter.

We mention in this connection a nice paper by Khan [63] which offers a synthetic iteration procedure containing both the iteration scheme of Ptak and the scheme developed in [27].
The papers [27], [86], [87] were the first publications in which the metric approach to the covering and regularity problems was subsequently implemented. This is certainly the most natural approach and problems not connected with analysis of local criteria were later considered by many authors in the general context of metric spaces (see, for example, [17], [23], [37], [83]).

4.3. The equivalence of the local properties of metric regularity and covering at a linear rate, including the equality Reg \( F \cdot \text{Sur} \ F = 1 \), was well understood by the beginning of the 80-ies (see the introduction to [27] and Proposition 11.12 in [46]), even before the concept of the pseudo-Lipschitz property was first formulated. Proposition 2, as it stated here, appeared in [17], [83].

The pseudo-Lipschitz property was defined by Aubin [2] in 1984. From the formal point of view, this is of course a reformulation of the regularity property for the inverse map. But it was this definition that led to the development of a new chapter of non-smooth analysis devoted to the study of Lipschitz properties of set-valued maps (see [91]). In non-smooth analysis, this is not a unique case when a properly formulated statement offers a new insight into a seemingly well studied phenomenon.

Thanks to the equivalence between regularity and the pseudo-Lipschitz property of the inverse map, we can interpret any statement containing a sufficient regularity condition as a multi-valued “inverse function theorem”. This interpretation, well understood before the equivalence itself, is a key element in the circle of problems connected with “Lipschitz stability” of solutions to equations, variational inequalities, optimization problems etc.. The studies of Lipschitz stability initiated by works of Robinson [88], [89] and Aubin [2] proved to be very important for many applications and were continued by many researches (see for instance [5], [11], [29], [31], [32], [77], [80] and many references therein).

A proof of Proposition 3 can be found in [51] but the fact itself, as we have already mentioned, was known much earlier (see for instance the introduction to [50]). The concept of (local) graph-regularity was introduced by Jourani and Thibault [58] and its equivalence to local regularity was recently discovered by Thibault [93], [94]. The second of the quoted Thibault’s notes contains many other equivalent characterizations for metric regularity of set-valued maps, including a result equivalent to Proposition 3. Proposition 5 probably appears for the first time.

4.4. The results of the first chapter have a long history. In the vast majority of publications attention is devoted to sufficient regularity (openness) conditions. The few exceptions I am aware of include a recent paper by Kummer [66] who noticed that in certain cases the arguments based on Ekeland’s principle can be reversed, and earlier results of [50], [64], [78] of which we speak in greater details in the third chapter as they relate to the subdifferential theory. It has to be observed, however, that these are sufficient conditions that are typically needed in applications.

The first version of Basic Lemma with a regularity criterion stated in terms of Clarke’s generalized gradients was proved by Ioffe [45]. It was considerably strengthened by Cominetti [23] who was the first to give a purely metric condition of that sort. A version the result of [45] using proximal subgradients can be found

---

5 In fact, even more general class of so-called quasi-metric spaces is considered in [27]. It seems, however, that the results at this level of generality did not find much application.
in [22] under the name “decrease lemma” and a somewhat more subtle statement involving “weak slope”—in [55]. We also mention a recent paper by Ledyaev and Zhu [69] with a parametric analog of the first version of Basic Lemma involving Fréchet subdifferentials. Finally, very recently Azé, Corvellec, and Lucchetti [8] proved a local version of Basic Lemma corresponding to \( r = \infty \).

Theorems 1–3 are new. But the lower semicontinuity condition on functions \( x \mapsto d(y, F(x)) \) (Theorem 1b) appeared earlier in the above mentioned paper by Kummer [66], and the first (sufficiency) part of Theorem 2 has much in common with the quoted paper by Azé, Corvellec, and Lucchetti [8]. The latter was probably the first publication which drew attention to the possibility to use the strong slope of [26] to get characterizations of metric regularity. A very recent proof of necessity of the condition of Theorem 2 in a slightly more general situation can be found in [54]. We can state with a high level of confidence that Theorems 2 and 3 offer a key to various local regularity criteria.

4.5. There are two basic methods used in proofs of general regularity and covering theorems. The first, going back to Banach, Schauder, Lyusternik and Graves, employs special iteration procedures. They are closely connected with the classical Newton method but differ from it in one essential point. In Newton-type algorithms the next iteration is completely defined by the previous iteration whereas in the scheme of of Banach, Schauder, Lyusternik and Graves we have to choose one of many possibilities offered by the previous iteration at every step of the procedure. This determines a tremendous universality and flexibility of the scheme.\(^6\) The above mentioned works of Ptak and Khan [63], [86], [87] follow the basic lines of the scheme as well as many proofs of the “multi-valued contraction mapping principle” first proposed by Nadler [81] (see for instance [30], [56]). Among other publications in which this technique is developed we mention [7], [23], [27], [89], [90] (the chronological order of these publication is precisely the opposite!).

Another approach based on the variational principle of Ekeland was offered by Ioffe in [45] and later was applied and developed in many publications, mainly to prove local regularity criteria (see for instance [1], [8], [17], [46], [50], [59], [60], [66], [69], [80], [85]). The proof of Basic Lemma in this paper as well as the proof of the corresponding result in [8] are modifications of the proof given in [45].\(^7\)

The approach based on Ekeland’s principle is more universal: it does not need uniformity conditions necessary to guarantee convergence of iterations (usually with a linear rate) independent of initial approximation. This approach is certainly adequate to the problem of metric regularity which is emphasized by the fact that using it, we can also get necessary regularity conditions.\(^8\) On the other hand, it was shown in [57] that any existence theorem based on iteration schemes of the first kind also follows from Ekeland’s principle.

---

\(^6\)It is appropriate to quote here a remark of the authors of [27] that the history of achievements relating to problems being discussed reduces to a search of suitable new formulations for already existing proof!

\(^7\)So that, the observation of the previous footnote applies, to a certain extent, also for the second approach.

\(^8\)This circumstance was not immediately recognized as can be seen from a rather sceptical estimate of the potential of Ekeland’s principle (as regards the problem of regularity) in [27] and the interpretation by Ekeland in [33] of the local version of Basic Lemma proved in [45] as a “generalized mean value theorem”.
4.6. For single-valued mappings Theorems 5 and 6 were proved in [27]. There is also a thorough survey of earlier results there. In this case we can certainly guarantee closed covering. The proof of the single-valued version of Theorem 5 given in [27] can easily be adjusted to the conditions of the theorem. It also follows from Proposition 3 that Theorems 5 and 6 are equivalent to the corresponding results for single-valued maps. For instance, to get Theorem 6 from Theorem 1.3 of [27] we have to consider the maps \((x, y) \mapsto y\) and \((x, y) \mapsto y + \Phi(x)\) on the graph of \(F\).

An independent proof of Theorem 6a with the help of the multi-valued contraction mapping principle in Banach spaces can be found in [30] (see also [28]). In this specific case the assumption of the corollary means that \(\Phi\) is strictly differentiable at \(x\) with the zero derivative.

4.7. In some recent publications the name “implicit function theorem” is used for results for which the name “parametric regularity” theorem would be more suitable (see, for example, [8], [69], [90]). Meanwhile, as the authors of [27] justifiably remark, the main content of the implicit function theorem is that the implicitly defined function inherits analytic properties of the defining map (for example, the degree of smoothness). In the metric theory the most natural candidates for such properties are, of course, condition of Lipschitz or Hölder type.

Theorem 7 appears for the first time. It is closely connected with parametric theorems of [8], [69]. Among earlier “real” non-smooth implicit function theorems we mention the result obtained in [11] which follows from Theorem 7. In Corollary 7.1 we explicitly require that the inverse mapping be single-valued mainly to simplify the arguments. It has to be noted nonetheless that local non-smooth analysis does offer some infinitesimal means to provide this property [3, 20, 46, 65].

4.8. Proposition 7 was stated in [46] (Proposition 11.12) without proof. Its detailed proof can be found in [52]. But a thorough study of “higher order” covering property for set-valued maps was first undertaken in a series of papers by Frankowska [37], [38]. (In smooth case corresponding extensions of the Ljusternik-Graves theorem were studied by Grasse [41].) The given definitions of regularity, covering and Hölder properties of order \(\delta\) are slight modifications of the definitions introduced in [17]. A conceivable local version of Proposition 6 is a part of the main theorem of [17] and, follow, in turn from Proposition 6. It was also proved in [17] that for \(\delta(t) = t^r\) the list of equivalent properties of the local version of Proposition 6 can be augmented by the condition that the \(r\)-th derivative of the mapping at \((\overline{x}, \overline{y})\) (in the sense defined in [38]) is finite.

4.9. We have already mentioned that Lyusternik’s immediate motivation that led him to prove his famous theorem was the creation of an adequate machinery for proving Lagrange multiplier rule in situations when the traditional techniques based on the implicit function theorem cannot be applied (for the reason that the kernel of the derivative of the constraint map does not have a topological complement).

Regularity estimates proved to be a much more powerful tool. In addition to being effectively used in proofs of the first order necessary conditions in optimization problems, they play a central role in the theory of higher order optimality conditions developed by Levitin, Milyutin, and Osmolovskii [70], [74] (see also [56]) and, on the other hand, offer a convenient instrument for stability analysis of solutions to optimization problems (see, for example, [14]).
Chapter 2. Subdifferentials

§1. Definitions and examples of subdifferentials

1.1. Subdifferentials associated with derivatives. This is the class of subdifferentials defined in the simplest and most natural way. Recall that a function \( f \) is Fréchet differentiable at \( x \) if there is an \( x^* \in X^* \) (called the Fréchet derivative of \( f \) at \( x \)) such that

\[
f(x + h) - f(x) - \langle x^*, h \rangle = r(h)\|h\|,
\]

where \( r(h) \to 0 \) if \( \|h\| \to 0 \). Clearly, the derivative, if it exists, is uniquely defined. To get the definition of a Fréchet subgradient we simply change the equality in the definition of Fréchet derivative to \( \geq \).

**Definition 1.** A functional \( x^* \in X^* \) is a Fréchet subgradient of \( f \) at \( x \) if

\[
f(x + h) - f(x) - \langle x^*, h \rangle \geq r(h)\|h\|, \quad \text{where} \quad r(h) \to 0 \quad \text{when} \quad \|h\| \to 0.
\]

The set of all Fréchet subgradients of \( f \) at \( x \) is called the Fréchet subdifferential of \( f \) at \( x \). It is usually denoted by \( \partial_{F} f(x) \).

Subdifferentials associated with other types of derivatives are defined in a similar way. We recall (see, for example, [6]) that a collection \( \beta \) of centrally symmetric convex bounded subsets of \( X \) is called a bornology if (a) the union of elements of \( \beta \) coincides with \( X \) and (b) for any two elements of \( \beta \) there is a third element containing the first two. Given a bornology \( \beta \) in \( X \), the \( \beta \)-derivative of \( f \) at \( x \) is a linear bounded functional \( x^* \) such that for any \( S \in \beta \)

\[
\lim_{t \to 0} t^{-1}(f(x + th) - f(x) - t\langle x^*, h \rangle) = 0
\]

uniformly for \( h \in S \). As in the case of Fréchet derivatives, we get the definition of a \( \beta \)-subgradient of \( f \) at \( x \) by changing equality to \( \geq \) in the definition of the \( \beta \)-derivative. The set \( \partial_{\beta} f(x) \) of all \( \beta \)-subgradients of \( f \) at \( x \) is the \( \beta \)-subdifferential of \( f \) at \( x \).

The Fréchet derivative corresponds to the bornology formed by all balls around the origin. Another important bornology is formed by all compact centrally symmetric sets. The corresponding subdifferential is called the Dini–Hadamard subdifferential. We shall denote it by \( \partial^{-} f(x) \). The following two propositions contain equivalent descriptions of Fréchet and Dini–Hadamard subdifferentials.

**Proposition 1.** The following conditions are equivalent:

(a) \( x^* \in \partial_{F} f(x) \);

(b) the function \( g(h) = f(x + h) - f(x) - \langle x^*, h \rangle + \varepsilon\|h\| \) attains a local minimum at zero for any \( \varepsilon > 0 \);

(c) there is a function \( \varphi(x) \) defined and satisfying the Lipschitz condition in a neighborhood of \( x \), Fréchet differentiable at \( x \) and such that \( \varphi'(x) = x^* \) and \( f - \varphi \) attains a local minimum at \( x \).

The equivalence of (a) and (b), as well as the implication (c)\( \Rightarrow \) (a), is an immediate consequence of the definition. A proof of the inverse implication (a)\( \Rightarrow \) (c) can be distilled from a proof of a more subtle version of this result in [25].
Proposition 2. Assume that $f$ is lower semicontinuous in a neighborhood of $x$. Then the following conditions are equivalent:
(a) $x^* \in \partial^- f(x)$;
(b) $d^- f(x; h) \geq \langle x^*, h \rangle$ \(\forall h \in X\), where
\[
d^- f(x; h) = \liminf_{t\to 0} t^{-1} (f(x + tu) - f(x))
\]
is the Dini–Hadamard directional derivative of $f$ at $x$ along $h$;
(c) for any $\varepsilon > 0$ and every finite-dimensional subspace $L \subset X$, the restriction of $g(h) = f(x + h) - f(x) - \langle x^*, h \rangle + \varepsilon\|h\|$ to $L$ has a local minimum at zero.

The $\beta$-subdifferentials that have been defined are called canonical. We observe that unlike in the case of the Fréchet subdifferential, $g(h)$ may fail to have an unconditional local minimum at zero if $x^* \in \partial^- f(x)$. (This is the most distinctive property of the Fréchet subdifferential among all subdifferentials associated with derivatives.) This may cause some technical complications, and to cope with them another class of subdifferentials associated with derivatives is often considered. It is said that $x^*$ is a viscosity $\beta$-subgradient of $f$ at $x$ if there is a function $\varphi(x)$, defined $\beta$-differentiable and Lipschitz in a neighborhood of $x$, such that $\varphi'(x) = x^*$ and $f - \varphi$ have a local minimum at $x$. Subdifferentials of this type proved to be very useful in the Hamilton-Jacoby theory. But in what follows we mainly speak about canonical subdifferentials. (Note that Fréchet viscosity and canonical subdifferentials coincide if there is a Fréchet differentiable Lipschitz bump function on $X$—see below.)

The definitions we have introduced make sense in every Banach (even normed) space. But the possibility to effectively apply them in specific situations very much depends on geometric properties of the corresponding space. For instance, if $X$ is not an Asplund space, then there is a continuous concave (hence Lipschitz on every ball) function whose Fréchet subdifferential is identically empty. Therefore, generally speaking, the Fréchet subdifferential makes sense only on Asplund spaces, although for special classes of functions it can be considered on more general classes of spaces. For instance, integral functionals like $\int f(t, x(t)) \, dt$ have good Fréchet subdifferentiability properties on spaces of continuous functions.

We shall call $X$ a $\beta$-subdifferentiability space if every lower semicontinuous function has a nonempty $\beta$-subdifferential on a dense subset of its domain. Clearly, $\beta$-subdifferentiability is a minimal requirement on the space if we wish to work with lower semicontinuous function using $\beta$-subdifferentials. The following concept of bump function is essential in available descriptions of $\beta$-subdifferentiability spaces. We say that $\varphi(x)$ is a bump function if $0 \leq \varphi(x) \leq 1 = \varphi(0)$, $\varphi(x) = 0$ outside of the unit ball and $\varphi$ has a strong maximum at zero (that is, every maximizing sequence converges to zero).

Proposition 3. (a) Suppose that there is a $\beta$-differentiable bump function on $X$ satisfying the Lipschitz condition. Then $X$ is a $\beta$-subdifferentiability space.

(b) $X$ is a Fréchet subdifferentiability space if and only if it is Asplund.
that the viscosity $\beta$-subdifferential is nonempty under the assumption. The second statement was proved by Fabian [35, 36] for canonical Fréchet subdifferentials. It needs a more complicated proof based on separable reduction of Asplund spaces.

It follows from the proposition that any Banach space with an equivalent $\beta$-differentiable norm is a $\beta$-subdifferentiability space. Indeed,

$$\varphi(x) = \begin{cases} (1 - \|x\|^2)^2, & \text{if } \|x\| \leq 1; \\ 0, & \text{otherwise} \end{cases}$$

is a bump function which is $\beta$-differentiable and satisfies the Lipschitz condition.

To conclude we note that Fréchet and Dini–Hadamard subdifferentials coincide in a finite-dimensional space.

1.2. Limiting subdifferentials. In one respect subdifferentials associated with derivatives are apparently inconvenient: even for a Lipschitz function they can be empty on a rich subset of the domain. This does not allow to use such subdifferentials in many important cases. For instance, necessary conditions in constraint optimization problems involving such subdifferentials can be stated only in terms of certain minimizing sequences whose elements can not be in principle identified.

**Definition 2.** The limiting $\beta$-subdifferential of $f$ at $x$ is the collection of all weak* limits of sequences $\{x_n^*\}$ such that $x_n^* \in \partial_\beta(x_n)$, where $x_n \to x$ and $f(x_n) \to f(x)$. We shall occasionally denote them by $\partial_\beta f(x)$.

A limiting subdifferential may not be a convex set. It is clear that Proposition 3 extends to subdifferentials of this type. But we can state for them even more.

**Proposition 4.** Let $X$ be a $\beta$-subdifferentiability space such that the unit ball in $X^*$ is sequentially weak* compact. Then $\partial_\beta f(x) \neq \emptyset$ if $f$ satisfies the Lipschitz condition in a neighborhood of $x$.

The proof follows from the obvious inequality $\|x^*\| \leq K$ if $x^* \in \partial_\beta f(x)$ and $K$ is a Lipschitz constant of $f$ near $x$.

We recall that a Banach space with a $\beta$-differentiable renorm is a space of $\beta$-subdifferentiability by Proposition 3. On the other hand in such space every bounded sequence of linear functionals contains a weak*-converging subsequence. It follows that Proposition 4 is valid for such spaces.

In case of the (canonical) Fréchet subdifferential the conclusion of Proposition 4 holds true in every Asplund spaces.

1.3. Approximate subdifferential. If $X$ is not a subdifferentiability space for any reasonable bornology, we need to find another way to construct subdifferentials. The simplest idea consists in taking restrictions of the function to “good” subspaces with subsequent looking at what happens when they infinitely expand.

We shall first try to implement this idea for functions satisfying the Lipschitz condition. In this case

$$d^-(x; h) = \liminf_{t \to 0} t^{-1}(f(x + th) - f(x)).$$

Let $L$ be a closed subspace of $X$. We set

$$\partial_L^- f(x) = \{x^* : d^-(x, h) \geq \langle x^*, h \rangle \ \forall h \in L\}.$$
This is the Dini–Hadamard subdifferential of the restriction of the function \( \psi(h) = f(x + h) \) to \( L \) (which is the function equal to \( f(x + h) \) on \( L \) and infinity outside of \( L \)). Let finally \( \mathcal{F} \) denote the collection of all finite-dimensional subspaces of \( X \).

**Approximate subdifferential** of \( f \) at \( x \) is defined by the equality

\[
\partial_G f(x) = \bigcap_{L \in \mathcal{F}} \limsup_{u \to x} \partial_L^* f(u).
\]

Sometimes more convenient is a lengthier (but equivalent) formula

\[
\partial_G f(x) = \bigcap_{L \in \mathcal{F}} \limsup_{u \to x} [\partial_L^* f(u) \cap kB],
\]

where \( k \) is any number not smaller than the Lipschitz constant of \( f \).

Suppose now that \( f \) is an arbitrary lower semicontinuous function. Then the approximate subdifferential of \( f \) at \( x \) is defined as follows

\[
\partial_G f(x) = \left\{ x^* : (x^*, -1) \in \bigcup_{\lambda > 0} \lambda \partial d((x, f(x)), \text{epi} f) \right\}. \tag{1}
\]

Here as above, \( d((x, \alpha), \text{epi} f) \) is the distance from \((x, \alpha)\) to \( \text{epi} f \). The described construction is universal in the sense that it works in all Banach spaces and in “good” spaces produces a subdifferential coinciding with the above introduced limiting subdifferentials.

**1.4. Clarke’s generalized gradient.** This is chronologically the first well studied class of subdifferentials introduced in mid-seventieth. Its original definition also uses a two step construction. For a function satisfying the Lipschitz condition near \( x \) we set

\[
f^\circ(x; h) = \limsup_{t \to 0, u \to x} t^{-1} (f(u + th) - f(u))
\]

and then define the generalized gradient by

\[
\partial_C f(x) = \{ x^* : f^\circ(x; h) \geq \langle x^*, h \rangle \ \forall h \in X \}.
\]

This is always a convex set. If now \( f \) is an arbitrary lower semicontinuous function, then the generalized gradient of \( f \) is defined as in (1) but with the addition of the weak* closure operation:

\[
\partial_C f(x) = \text{cl}^* \text{conv} \left\{ x^* : (x^*, -1) \in \bigcup_{\lambda > 0} \lambda \partial_C d((x, f(x)), \text{epi} f) \right\}.
\]

Rockafellar [94] gave an equivalent description of the generalized gradient as a support set of a certain “directional derivative” of \( f \) at \( x \).

---

\(^9\)Index “\( G \)” appeared because originally two types of approximate subdifferential were considered, “analytic” and “geometric”. But it became clear later that the first is less interesting, at least in the context of Banach spaces.
A complete description of the connection between the approximate subdifferential and the generalized gradient will be given later in Proposition 6.

1.5. Axiomatics of the theory of subdifferentials. All subdifferential we have mentioned have the following properties:

(SD$_1$) $\partial f(x) = \emptyset$ if $x \not\in \text{dom } f$;
(SD$_2$) $\partial f(x) = \partial g(x)$ if $f$ and $g$ coincide in a neighborhood of $x$;\footnote{The Russian version of the paper was already in print when R. T. Rockafellar brought my attention to the fact that a better (and more general) way to state this property would be the following: $\partial f(x) = \partial g(x)$ if there is a neighborhood $U$ of $(x, f(x))$ in $X \times R$ such that $U \cap \text{epi } f = U \cap \text{epi } g.$}
(SD$_3$) if $f$ convex then $\partial f(x)$ is the subdifferential of $f$ in the sense of convex analysis;
(SD$_4$) if $f$ satisfies the Lipschitz condition with constant $K$ in a neighborhood of $x$ then $\|x^*\| \leq K$ for any $x^* \in \partial f(x)$.
(SD$_5$) $0 \in \partial f(x)$ if $f$ attains a local minimum at $x$;
(SD$_6$) if $X = X_1 \times X_2$ and $f(x) = f_1(x_1) + f_2(x_2)$ then $\partial f(x) \subset \partial f_1(x_1) \times \partial f_2(x_2)$;
(SD$_7$) if $f(x) = \lambda g(Ax + y)$, where $\lambda > 0$, $A$ is a linear bounded operator from $X$ onto $Y$, then $\partial f(x) = \lambda A^* \partial g(Ax + y)$.
(SD$_8$) $\partial f(x) = \{x^* : (x^*,-1) \in \partial \chi_{\text{epi } f}(x, f(x))\}$, where

$$\chi_{S}(x) = \begin{cases} 0, & \text{if } x \in S; \\ \infty, & \text{otherwise} \end{cases}$$

is the so called indicator of $S$.

In many cases only these properties are needed in proofs and further constructions rather than a specific structure of one or another subdifferential. In such situations an axiomatic approach, when we interpret these properties as axioms of subdifferential calculus, appears to be very convenient.

1.6. Normal cones and coderivatives. Let $S \subset X$, and let $\partial$ be a certain subdifferential. The obvious equality $\lambda \chi_{S}(x) = \chi_{S}(x), (\lambda > 0)$ along with (SD$_7$) implies that $\lambda \partial \chi_{S}(x) = \partial \chi_{S}(x)$ for any positive $\lambda$. This means that the subdifferential of an indicator function at any point of the set is a cone containing zero by (SD$_4$). It is called the normal cone (to $S$ at $x$) associated with with $\partial$, and it is usually denoted by the symbol $N(S, x)$ supplied, if necessary, with the same descriptive subscript as the subdifferential with which the cone is associated. With this notation, (SD$_8$) assumes the form

$$\partial f(x) = \{x^* : (x^*,-1) \in N(\text{epi } f, (x, f(x)))\}.$$  

The following proposition contains several important facts specific for the approximate subdifferential and the canonical Fréchet subdifferential of the distance function. Among them there is an equivalent description of the corresponding normal cone, often very convenient.

**Proposition 5.** Suppose that either $X$ is an arbitrary Banach space and $\partial$ is the approximate subdifferential or $X$ is an Asplund space and $\partial$ is the (limiting)
canonical Fréchet subdifferential. Let $S \subset X$. Then for any $x \in S$
\[ N(S, x) = \bigcup_{\lambda > 0} \lambda \partial d(x, S), \quad \partial d(x, X) = \bigcup_{\lambda \in [0, 1]} \lambda \partial d(x, S), \]
and for any function $f(\cdot)$ which is equal to zero on $S$ and nowhere smaller than $d(\cdot, S)$
\[ \partial d(x, S) \subset \partial f(x). \]

Now we can give the complete description of the connection between the generalized gradient and approximate subdifferential. We set (for any subdifferential)
\[ \partial^\infty f(x) = \{ x^* : (x^*, 0) \in N(\text{epi } f, (x, f(x))) \}. \]

This is the so called singular subdifferential of $f$ at $x$. Geometrically it corresponds to “horizontal” normals to the epigraph of $f$ at $(x, f(x))$. It is easy to verify that the singular subdifferential of a Lipschitz function is trivial.

**Proposition 6.**
\[ N_G(S, x) = \text{cl}^* \text{conv } N_G(S, x); \]
\[ \partial_G f(x) = \text{cl}^* \text{conv } (\partial_G f(x) + \partial_G^\infty f(x)). \]

In particular the approximate subdifferential is always contained in the corresponding generalized gradient. Moreover, if $f$ satisfies the Lipschitz condition in a neighborhood of $x$, then
\[ \partial_G f(x) = \text{cl}^* \text{conv } \partial_G f(x). \]

Suppose now that we are given a set-valued map $F$ from $X$ into $Y$ and $(x, y) \in \text{Gr } F$. The coderivative of $F$ at $(x, y)$ associated with $\partial$ is the set-valued map $D^* f(x, y)$ from $Y^*$ into $X^*$ defined by
\[ D^* F(x, y)(y^*) = \{ x^* : (x^*, -y^*) \in N(\text{Gr } F, (x, y)) \}. \]

The coderivative of a bounded linear operator at any point coincides with the adjoint operator, no matter with which subdifferential it is associated. Likewise, any coderivative of a map strictly Fréchet differentiable at a point is the adjoint to the derivative at the point.

**§2. Basic theorems of subdifferential calculus**

**2.1. Fuzzy calculi and trustworthy spaces.** Speaking about calculus of derivatives and subdifferentials, we usually mean rules that enable us to calculate or estimate derivatives or subdifferentials of composite functions (sums, compositions etc.), mean value theorems etc. As far as sums are concerned, the inclusion
\[ \partial(f_1 + \cdots + f_n)(x) \supset \partial f_1(x) + \cdots + \partial f_n(x) \quad (2) \]
for $\beta$-subdifferentials is obvious, whereas the opposite inclusion is obviously not valid. On the other hand, the experience of convex analysis suggests that this is the opposite inclusion that is usually needed in applications. Fortunately, in many important cases the necessary inclusion does hold for limiting subdifferentials (in suitable spaces) and for the approximate subdifferential and generalized gradient (in all Banach spaces). The decisive for non-smooth analysis fact is that also for subdifferentials associated with derivatives and certain others we have a calculus of a new type in which the desired inclusion is “almost” satisfied in the sense precisely defined below.
Definition 4. Suppose that we are given a subdifferential \( \partial \) and a Banach space \( X \). We say that \( \partial \) satisfies the basic fuzzy principle on \( X \) if for any \( \varepsilon > 0 \) and any finite collection \( (f_1, \ldots, f_k) \) of lower semicontinuous functions on \( X \) satisfying, but for at most one of them, the Lipschitz condition in a neighborhood of a certain \( x \), there are \( x_1, \ldots, x_k, x^*_1, \ldots, x^*_k \) such that

\[
\|x_i - x\| < \varepsilon, \quad |f_i(x_i) - f_i(x)| < \varepsilon, \quad x^*_i \in \partial f_i(x_i) \quad \forall i = 1, \ldots, k, \quad \left\| \sum x^*_i \right\| < \varepsilon. \tag{3}
\]

If the basic fuzzy principle is satisfied for \( \partial \) on \( X \), we shall also say that \( X \) is a \( \partial \)-trustworthy space.

Theorem 1. (a) \( X \) is a trustworthy space for any \( \beta \)-subdifferential, provided there is on \( X \) a \( \beta \)-differentiable bump function satisfying Lipschitz condition.

(b) A Banach space is a trustworthy space for the canonical Fréchet subdifferential if and only if it is Asplund.

(c) Any Banach space is a trustworthy space for the approximate subdifferential.

It is obvious from the definition that any \( \partial \)-trustworthy space is also a \( \partial' \)-trustworthy space if \( \partial' \) is bigger than \( \partial \) in the sense that the inclusion \( \partial f(x) \subset \partial f'(x) \) always holds. Therefore a space satisfying (a) is a trustworthy space for the corresponding limiting \( \beta \)-subdifferential, any Asplund space is a trustworthy space for the limiting (canonical) Fréchet subdifferential and every Banach space is a trustworthy space for the generalized gradient.

The following minimality property of the approximate subdifferential should be mentioned in this connection: if \( X \) is a \( \partial \)-trustworthy space and \( f \) satisfies the Lipschitz condition near \( x \), then

\[
\partial_G(x) \subset \limsup_{x \to x} \partial f(x).
\]

(Here \( \limsup \) is the topological (not sequential) upper limit with respect to the weak*-topology in \( X^* \).) It can also be observed that in nearly all important for application cases, namely when \( X \) is a weakly compactly generated space, any limiting \( \beta \)-subdifferential of a Lipschitz function coincides with its approximate subdifferential, provided \( X \) is also a \( \partial \)-trustworthy space. In case of an Asplund space and the limiting Fréchet (canonical) subdifferential, this applies to any lower semicontinuous function, not only to Lipschitz functions. Therefore, say in a reflexive space, all limiting subdifferentials coincide with the approximate subdifferential if the function is Lipschitz, and in a finite-dimensional space this coincidence extends to all lower semicontinuous functions. In this sense we can speak about the universality of the approximate subdifferential (see also footnote 14 in Chapter 3).

Proof of the theorem. We shall prove only the first statement. It is also the main part of the proof of the second statement as every separable subspace of an Asplund space has a Fréchet differentiable norm, and to complete the proof we have to apply the so-called separable reduction (lifting from separable spaces to the entire space). Finally, the proof of the third statement is based on the first and Proposition 2 and 5.
1) Let \( \varphi(x) \) be a \( \beta \)-differentiable bump function satisfying the Lipschitz condition. Set \( U = \{ x : \varphi(x) > 0 \} \), and for any \( r > 0 \) define

\[
\gamma_r(x) = \begin{cases} 
[\gamma_0(x)]^{-1}, & \text{if } x \in r^{-1}U, \\
\infty, & \text{otherwise.}
\end{cases}
\]

By the assumption, \( U \) lies in the unit ball, that is \( \gamma_r(x) < \infty \) only if \( ||x|| < r^{-1} \). It is clear that \( \gamma_r \) is \( \beta \)-differentiable at every point of its domain and satisfies the Lipschitz condition in a neighborhood of any such point.

2) So let \( \varepsilon > 0 \), and let \( f_1, \ldots, f_k \) be lower semicontinuous and all of them, but for at most one, satisfy the Lipschitz condition in a neighborhood of \( \overline{X} \). Suppose that \( f = f_1 + \cdots + f_k \) attains a local minimum at \( \overline{X} \). We may assume without loss of generality that \( \overline{X} = 0 \) and, on the unit ball, \( f(x) \geq f(0) \), \( f_i(x) \geq f_i(0) - 1 \) and \( f_1, \ldots, f_{k-1} \) satisfy the Lipschitz condition with constant \( C \). Consider the function

\[
p_r(x_1, \ldots, x_k) = \sum_{i=1}^{k} f_i(x_i) + \sum_{i=1}^{k-1} \gamma_r(x_i - x_k) + \delta(1 - \varphi(x_k)),
\]

where \( \delta \) is so small that the Lipschitz constant of \( \delta \varphi(x) \) is not greater than \( \varepsilon/2 \). This function is bounded below and lower semicontinuous on the product of \( k \) copies of the unit ball of \( X \). Set

\[
a_r = \inf\{ p_r(x_1, \ldots, x_k) : ||x_i|| \leq 1 \},
\]

and choose \( u_{ir} \) satisfying \( ||u_{ir}|| < 1 \) and

\[
a_r + 1/r \geq p_r(u_{1r}, \ldots, u_{kr}).
\]

3) We have

\[
a_r + 1/r \geq p_r(u_{1r}, \ldots, u_{kr})
\]

\[
= \sum_{i=1}^{k} f_i(u_{ir}) + \sum_{i=1}^{k-1} \gamma_r(u_{ir} - u_{kr}) + \delta(1 - \varphi(u_{kr}))
\]

\[
\geq \sum_{i=1}^{k} f_i(u_{kr}) + \sum_{i=1}^{k-1} \gamma_r(u_{ir} - u_{kr}) - C \sum_{i=1}^{k-1} ||u_{ir} - u_{kr}||
\]

\[
\geq f(0) + \sum_{i=1}^{k-1} \gamma_r(u_{ir} - u_{kr}) - 2kC.
\]

It follows that \( \gamma_r(u_{ir} - u_{kr}) < \infty \), hence \( ||u_{ir} - u_{kr}|| \leq r^{-1} \to 0 \) for all \( i \).

On the other hand,

\[
f(0) + k\gamma_r(0) \geq p_r(0, \ldots, 0) \geq a_r
\]

\[
\geq \sum_{i=1}^{k} f_i(u_{ir}) + \delta(1 - \varphi(u_{kr})) - 1/r
\]

\[
\geq f(0) - C \sum_{i=1}^{k-1} ||u_{ir} - u_{kr}|| + \delta(1 - \varphi(u_{kr})) - 1/r.
\]
It follows that \( \varphi(u_{kr}) \to 1 \), hence \( u_{kr} \to 0 \). Thus, all \( u_{ir} \) go to zero as \( r \to \infty \).

4) Consider the space \( \mathcal{F}_0 \) of bounded everywhere \( \beta \)-differentiable functions on \( X^k \) satisfying the Lipschitz condition. We shall consider this space along with the \( C^1 \)-norm. It is not a difficult matter to verify that this space is complete. Denote by \( \mathcal{F} \) the subspace of \( \mathcal{F}_0 \) formed by separable functions \( \psi(x_1, \ldots, x_k) = \psi_1(x_1) + \cdots + \psi_k(x_k) \). Applying the variational principle of Deville–Godefroy–Zizler\(^{11}\) \cite{25}, to \( p_r \) and \( \mathcal{F} \), we find a \( \psi_r \in \mathcal{F} \) such that its \( \mathcal{F} \)-norm is not greater than \( 1/r \) and \( p_r - \psi_r \) attains its minimum at a certain \( (u_{1r}, \ldots, u_{kr}) \). This means that for any \( i < k \) the function

\[
x \mapsto f_i(x) + \gamma_r(x - u_{kr}) - \psi_{r,i}(x)
\]

attains minimum at \( u_{ir} \). Denote by \( -u_{ir}^* \) the \( \beta \)-derivative of \( \gamma_r \) at \( u_{ir} - u_{kr} \) and by \( v_{ir}^* \) the \( \beta \)-derivative of \( \psi_{ir} \) at \( u_{ir} \). It follows that

\[
x_{ir}^* = u_{ir}^* + v_{ir}^* \in \partial \beta f_i(u_{ir}).
\]

Likewise, the function

\[
x \mapsto f_k(x) + \sum_{i=1}^{k-1} \gamma_r(u_{ir} - x) + \delta(1 - \varphi(x)) - \psi_{r,k}(x)
\]

attains minimum at \( u_{kr} \) and, if \( w_r^* \) is the \( \beta \)-derivative of \( \delta \varphi \) at \( u_{kr} \), then

\[
x_{kr}^* = w_r^* - \sum_{i=1}^{k-1} u_{ir}^* + v_{kr}^* \in \partial \beta f_k(u_{kr}).
\]

Thus

\[
\sum x_{ir}^* = w_r^* + \sum v_{ir}^* \in \sum \partial \beta f_i(u_{ir}).
\]

It remains to note that the value of \( p_r \) at \( (u_{1r}, \ldots, u_{kr}) \) differs from \( a_r \) by less than \( 1/r \) (by the choice of \( \psi_r \)), \( \|w_r^*\| < \varepsilon/2 \) (by the choice of \( \delta \)) and \( \|v_{ir}^*\| \to 0 \) as \( r \to 0 \).

It turns out that fuzzy analogues of many fundamental analytic results, including the mean value theorem, theorem of Lyusternik–Graves (see the next chapter), the Lagrange multiplier rule can be established for functions on trustworthy spaces. On the other hand, property (3) alone is not sufficient to get analogues of the basic calculus rules for derivatives and convex subdifferentials in the abstract setting. However for the main specific subdifferentials such analogues do exist.

\(^{11}\) Let \( f \) be a bounded below lower semicontinuous function on a Banach space \( X \), and let \( \mathcal{F} \) be a Banach space of continuous functions on \( X \) with the following properties: (a) \( \mathcal{F} \) is stable with respect to translations and homotheties in \( X \) (that is, along with a given \( \varphi \) it contains all functions \( x \mapsto \varphi(\lambda x + u) \)), the corresponding operations in \( \mathcal{F} \) are continuous and the norm is invariant with respect to translations; (b) the norm in \( \mathcal{F} \) is not weaker than the the uniform norm on \( X \); (c) \( \mathcal{F} \) contains a bump function. Then the collection of \( \varphi \in \mathcal{F} \) for which \( f - \varphi \) attains its minimum at a unique point is a set of the second Baire category in \( X \).
Theorem 2. Let $X$ be a Banach space, and let $f_1, \ldots, f_k$ satisfy the Lipschitz condition in a neighborhood of an $x \in X$. Set $f = f_1 + \cdots + f_k$ and suppose that $x^* \in \partial f(x)$ for a certain subdifferential $\partial$. Then

1. if $\partial$ is a $\beta$-subdifferential and there is a $\beta$-differentiable Lipschitz bump function on $X$, then for any $\varepsilon > 0$ and any weak* neighborhood $V \subset X^*$ of zero there are $x_1, \ldots, x_k \in X$ and $x^*_1, \ldots, x^*_k \in X^*$ such that

$$\|x_i - x\| < \varepsilon, \quad x^*_i \in \partial f_i(x), i = 1, \ldots, k, \quad x^* \in x^*_1 + \cdots + x^*_k + V;$$

2. if $X$ is an Asplund space and $\partial = \partial_F$, then for any $\varepsilon > 0$ there are $x_1, \ldots, x_k, x^*_1, \ldots, x^*_k$ such that

$$\|x_i - x\| < \varepsilon, x^*_i \in \partial f_i(x), i = 1, \ldots, k, \quad \|x^*_1 + \cdots + x^*_k - x^*\| < \varepsilon;$$

3. The inclusion

$$\partial(f_1 + \cdots + f_k)(x) \subset \partial f_1(x) + \cdots + \partial f_k(x)$$

holds in the following four cases

3a) there is a Lipschitz $\beta$-differentiable bump function on $X$, the unit ball in $X^*$ is sequentially weak* compact and $\partial$ is a limiting $\beta$-subdifferential;

3b) $X$ is an Asplund space and $\partial$ is the limiting Fréchet (canonical) subdifferential;

3c) $X$ is an arbitrary Banach space and $\partial$ is the approximate subdifferential;

3d) $X$ is a Banach space and $\partial$ is Clarke’s generalized gradient.

The first three statement follow from Theorem 1 and Propositions 1 and 2. The last statement is an immediate consequence of the corresponding fact for convex functions as $(f_1 + \cdots + f_k)^*(x) \leq f_1^*(x) + \cdots + f_k^*(x)$ (which is obvious from the definition).

For more information about subdifferentials and their properties we refer to [13], [16], [15], [18], [21], [22], [36], [47], [49], [51], [53], [67], [75], [79], [91], [97].

Chapter 3. Subdifferential calculus and local regularity

Mutual cooperation between metric regularity and subdifferential calculus has been very fruitful since the very beginning of non-smooth analysis. On the one hand, the theory of subdifferentials has for many years been the main source of means for characterization of local regularity (and sometimes even non-local), and subdifferential regularity criteria have found many applications\(^\text{12}\). On the other hand, all qualification condition s in the most developed theorems of subdifferential calculus turn out to be regularity conditions for certain set-valued maps. In this chapter we everywhere speak only about metric regularity. The reader can easily reformulate the results for the covering and pseudo-Lipschitz properties.

---

\(^\text{12}\)See, however, the introduction.
§1. Subdifferential criteria of local metric regularity

1.1. The main sufficient condition. The following proposition explains why subdifferentials is an appropriate language for local regularity criteria.

Proposition 1. Let \( \partial \) be a subdifferential, and let \( X \) be a \( \partial \)-trustworthy space. Let further \( f \) be lower semicontinuous, and let \( U \) be an open subset of \( X \). Then

\[
\inf_{x \in U} |\nabla f|(x) \geq \inf \{ ||x^*|| : x^* \in \partial f(x), \ x \in U \}.
\]

Proof. Set \( m = \inf_{x \in U} |\nabla f|(x) \). Take an \( \varepsilon > 0 \) and choose an \( \overline{x} \) such that \( |\nabla f(\overline{x})| < m + \varepsilon \). Then for all \( u \) of a neighborhood of \( \overline{x} \)

\[
g(u) = f(u) + (m + \varepsilon)||\overline{x} - u|| \geq f(\overline{x}) = g(\overline{x}).
\]

By (SD3) and (SD7) the subdifferential of \((m + \varepsilon)||\overline{x} - u||\) at any point belongs to \((m + \varepsilon)B \). Since \( X \) is a \( \partial \)-trustworthy space, it follows that there is an \( x \in U \) such that \( \partial f(x) + (m + \varepsilon)B \) meets the ball of radius \( \varepsilon \) around the origin. This means that there an \( x^* \in \partial f(x) \) such that \( ||x^*|| \leq m + 2 \varepsilon \) and the result follows.

A direct application of the proposition to derivation of sufficient regularity conditions from Theorem 1.3 is possible only if we have good estimates, say, similar to those in (2.4) for approximate or limiting subdifferentials of composite functions. Such estimates cannot be obtained in the abstract situation, so Proposition 1 should rather be considered an illustration explaining the essence of the matter. Below we prove a theorem containing the main subdifferential sufficient condition for local regularity which in general requires an independent proof.

Theorem 1. Let \( F : X \rightrightarrows Y \) be a closed set-valued map, and let \((\overline{x}, \overline{y}) \in \text{Gr} F\). Suppose that \( X \times X \times Y \) is a \( \partial \)-trustworthy space and there are \( r > 1 \) and \( K > 0 \) such that for any \((x, y)\) of some neighborhood \( U \) of \((\overline{x}, \overline{y})\) the inequality

\[
||x^*|| \geq 1
\]

holds whenever \((x^*, y^*) \in r\partial d_K ((x, y), \text{Gr} F) \) and \( ||y^*|| = K \).

Then \( F \) is regular near \((\overline{x}, \overline{y})\) with norm not exceeding \( K \).

Proof. Assume that, on the contrary, \( F \) is not regular near \((\overline{x}, \overline{y})\) with norm not exceeding \( K \). By Theorem 1.3 this means that in any neighborhood of \((\overline{x}, \overline{y})\) there is a point \((\overline{x}, \overline{y})\) with \( \overline{y} \notin F(\overline{x}) \) such that

\[
|\nabla f|(\overline{x}) < 1,
\]

where \( f(u) = d_K ((u, \overline{y}), \text{Gr} F) \). It follows (by the definition of the strong slope) that there are \( \mu \in (0, 1) \), \( \alpha \in (0, 1) \) such that

\[
d_K ((x, \overline{y}), \text{Gr} F) + \mu||x - \overline{x}|| \geq d_K ((\overline{x}, \overline{y}), \text{Gr} F)
\]

if \( ||x - \overline{x}|| < \alpha \). We have \( d_K ((\overline{x}, \overline{y}), \text{Gr} F) = \gamma > 0 \) by the choice of \((\overline{x}, \overline{y})\). Let us check that there are \( \varepsilon > 0, \xi > 0 \) such that

\[
(u, v) \in \text{Gr} F \quad \& \quad ||u - \overline{x}|| + K||v - \overline{y}|| \leq \gamma + \varepsilon \quad \Longrightarrow \quad ||v - \overline{y}|| \geq \xi.
\]
If this were not true, then we could find a sequence of pairs \((u_n, v_n) \in \text{Gr} F\) such that \(\|u_n - \tilde{x}\| \to \gamma, \|v_n - \tilde{y}\| \to 0\). Chose a \(\beta \in (0, 1)\) to guarantee that \(\beta \gamma < \alpha\) and set \(x_n = \beta u_n + (1 - \beta) \tilde{x}\). Then \(\|x_n - \tilde{x}\| = \beta \|u_n - \tilde{x}\| \to \beta \gamma\) which means that (2) is satisfied for \(x = x_n\) if \(n\) is large. On the other hand,

\[
d_K((x_n, \tilde{y}), \text{Gr} F) \leq \|x_n - u_n\| + K\|v_n - \tilde{y}\|
\]

\[
= (1 - \beta)\|u_n - \tilde{x}\| + K\|v_n - \tilde{y}\|
\]

\[
\to (1 - \beta)\gamma = d_K((\tilde{x}, \tilde{y}), \text{Gr} F) - \lim \|x_n - \tilde{x}\|,
\]

in contradiction with (2).

Assume that \(\delta\) has been chosen so small that

\[
\delta^2 < \varepsilon, \quad 2\delta < \xi; \quad \mu + 4\delta < 1; \quad \delta < (r - 1) \min\{1, K\}. \tag{4}
\]

Take a \((\tilde{u}, \tilde{v}) \in \text{Gr} F\) such that

\[
\|\tilde{u} - \tilde{x}\| + K\|\tilde{v} - \tilde{y}\| < d_K((\tilde{x}, \tilde{y}), \text{Gr} F) + \delta^2.
\]

By (3) we have

\[
\|\tilde{v} - \tilde{y}\| \geq \xi. \tag{5}
\]

Applying Ekeland’s principle to the function

\[
\|x - u\| + K\|v - \tilde{y}\| + \mu\|x - \tilde{x}\|
\]

c onsidered as a function on \(X \times \text{Gr} F\), we find a \((x_1, u_1, v_1)\) such that \(v_1 \in F(u_1),\)

\[
\|x_1 - \tilde{x}\| + \|u_1 - \tilde{u}\| + \|v_1 - \tilde{v}\| < \delta
\]

and the function

\[
g(x, u, v) = \|x - u\| + K\|v - \tilde{y}\| + \mu\|x - \tilde{x}\| + \delta(\|x - x_1\| + \|u - u_1\| + \|v - v_1\|)
\]

attains its minimum on \(X \times \text{Gr} F\) at \((x_1, u_1, v_1)\).

This function satisfies the Lipschitz condition with respect to \(u\) with constant \(1 + \delta\) with respect to \(u\) and with constant \(K + \delta < rK\) with respect to \(v\). Therefore

\[
g(x, u, v) + rd_K((u, v), \text{Gr} F) \tag{6}
\]

attains at \((x_1, u_1, v_1)\) its absolute minimum on the entire \(X \times X \times Y\). As the latter is a \(\delta\)-trustworthy space, there are triples \((x_2, u_2, v_2)\) and \((x_3, u_3, v_3)\) with components belonging to the \(\delta\)-neighborhoods of the corresponding components of \((x_1, u_1, v_1)\) and functionals:

\((x^*, y^*)\) belonging to the subdifferential of \(rd_K(\cdot, \text{Gr} F)\) at \((u_3, v_3)\);

\(u^*\) belonging to the subdifferential of \(\|\cdot\|_X\) at \(x_2 - u_2\);

\(w^*\) belonging to the subdifferential of \(\|\cdot\|_X\) at \(x_3 - \tilde{x}\);

\(u^*\) belonging to the subdifferential of \(\|\cdot\|_X\) at \(v_2 - \tilde{y}\);
such that
\[ \|u^* + w^\| < 2\delta, \quad \| - u^* + x^* \| < 2\delta, \quad \|K v^* + y^\| < 2\delta. \]

We have \( \|v_2 - \bar{y}\| \geq \|\bar{v} - \bar{y}\| - \|v_1 - v_1\| - \|v_1 - v_2\| \geq \xi - 2\delta > 0 \) which implies that \( \|v^*\| = 1 \) and therefore \( \|y^*\| - K \leq 2\delta. \) On the other hand, \( \|x^*\| \leq \mu + 4\delta \) and we come to a contradiction with the assumptions as \( \|y^*\| \to K \) as \( \delta \to 0 \) while \( \lim_{\delta \to 0} \|x^*\| \leq \mu < 1. \)

**Remark 1.** If \( \partial \) is an approximate or a limiting subdifferential for which the inclusion \( \partial(f + g)(\cdot) \subset \partial f(\cdot) + \partial g(\cdot) \) is valid for Lipschitz functions on \( X \times X \times Y, \) then in the concluding part of the proof we can take \( (x_i, u_i, v_i) \) equal to \( (x_1, u_1, v_1). \) Therefore in this case it is sufficient to assume in Theorem 1 that \( (1) \) is valid only on \( \text{Gr } F. \) A similar weakening of the condition is possible if both \( X \) and \( Y \) have \( \beta \)-differentiable renorms and \( \partial \) is a \( \beta \)-subdifferential.

**Remark 2.** If our space is a trustworthy space for two subdifferentials \( \partial \) and \( \partial' \) such that one of them is smaller in the sense that, say, the inclusion \( \partial f(x) \subset \partial' f(x) \) is always valid, then the estimates involving the smaller subdifferential are in general more precise. Therefore smaller subdifferentials are more preferable (as far as the regularity, covering and Lipschitz properties are considered).

We conclude the discussion of general subdifferential regularity criteria by stating a version of Theorem 1 whose proof is easily obtained from the proof of Theorem 1 if instead of \( (6) \) we take the function \( g(x, u, v) + \chi_{\text{Gr } F}(u, v). \) Theorem 1 gives a slightly better estimate but we shall see that in certain interesting cases even the “weakened” estimate of the theorem below is exact.

**Theorem 1a.** Under the assumptions of Theorem 1, the following estimate is valid for the constant of covering near \((\bar{x}, \bar{y})\):

\[
\text{Sur } F(\bar{x}, \bar{y}) \geq \liminf_{\varepsilon \to 0} \{ \|x^*\| : x^* \in D^* F(x, y)(y^*), \|y^*\| \geq 1, (x, y) \in B(\bar{x}, \varepsilon) \times B(\bar{y}, \varepsilon) \}.
\]

**Proof:** The conclusion of Theorem 1 can be reformulated as follows: the norm of regularity of \( F \) near \( \bar{x} \) satisfies

\[
\text{Reg } F(\bar{x}, \bar{y}) \leq \limsup_{\varepsilon \to 0} \{ \|y^*\| : x^* \in D^* F(x, y)(y^*), \|x^*\| \leq 1, (x, y) \in B(\bar{x}, \varepsilon) \times B(\bar{y}, \varepsilon) \}.
\]

So it remains to remember that the norm of regularity and the constant of covering are mutually inverse (Proposition 1.2).

**1.2. Approximate subdifferential and point criteria.** The Lyusternik–Graves theorem is a simple consequence of Theorem 1. But there is one noticeable difference in the way they are stated. In the first case we use only the derivative at a given point; in the second we have to take into account subdifferentials at all
points of a neighborhood. In applications this may not be convenient, as say when we speak about necessary optimality conditions\textsuperscript{13}.

The possibility to use only the derivative at a given point in the Lyusternik–Graves theorem follows from the condition that the derivative of \( F \) is continuous. The furthest available generalization of the condition that leads to a “point regularity criteria” is stated as follows.

**Definition 1.** Let \( F: X \rightharpoonup Y \), and let \( \partial \) be a certain subdifferential. We say that \( F \) is codirectionally compact with respect to \( \partial \) (or \( \partial \)-codirectionally compact) at \((\overline{x}, \overline{y})\) if for any sequence of quadruples \((x_n, y_n, x_n^*, y_n^*)\) such that

1. \((x_n, y_n) \rightarrow (\overline{x}, \overline{y})\), \(y_n \in F(x_n)\);
2. \(x_n^* \in D^*F(x_n, y_n)(y_n^*)\), \(\|x_n^*\| \rightarrow 0\);
3. \(\sup\|y_n^*\| < \infty\),


\( y_n \) norm converge to zero, provided they converge in the weak* topology.

Here are three important examples in which the codirectional compactness property is satisfied for the approximate subdifferential:

(a) \( \dim Y < \infty \);
(b) \( F(x) = Ax + G(x) \), where \( A \) is a linear bounded operator whose image is a subspace of finite codimension and \( G: X \rightharpoonup Y \) has the property that \( G(x + h) \subset G(x) + \|h\|Q + r(x, \|h\|)B \), where \( Q \) is a norm compact subset of \( Y \) and \( t^{-1}r(x, t) \rightarrow 0 \) as \( t \rightarrow 0 \) and \( x \rightarrow \overline{x} \). In particular, if \( F \) is continuously (or strictly) differentiable at \( \overline{x} \), then it is codirectionally compact if and only if \( \text{codim}(\text{Im}F'(\overline{x})) < \infty \);
(c) \( X = Y = C^n[0, 1] \) and \( F \) is defined by

\[
x(t) \mapsto x(t) - \int_0^t \Phi(x(s)) \, ds,
\]

where \( \Phi: \mathbb{R}^n \rightrightarrows \mathbb{R}^n \) is Lipschitz and bounded-valued. (Such mappings naturally appear in connection with differential inclusions.)

It follows from the definition that the codirectional compactness property is inherited by smaller subdifferentials, that is, for any two subdifferentials \( \partial, \partial' \) such that \( \partial f(x) \subset \partial' f(x) \), codirectional compactness with respect to \( \partial \) follows from codirectional compactness with respect to \( \partial' \). On the other hand, it is clear that a point criteria can be formulated only in terms of subdifferentials that have a certain amount of stability with respect to variations of the point. Therefore, in view of the minimality property, the approximate subdifferential seems to be the most suitable for universal point criteria\textsuperscript{14}.

\textsuperscript{13}In principle it is possible to state a “fuzzy” Lagrange multiplier rule as was done, for instance, in [16], [47]. But such necessary conditions by no means can be considered a final result. This is rather an unfinished product than something ready for consumption, for all necessary conditions, from the Lagrange multiplier rule to Pontryagin’s maximum principle, are devised with the purpose to replace an extremum problem by an equation, inequality etc.

\textsuperscript{14}Borwein and Fitzpatrick [12] constructed an example of a Lipschitz function on an Asplund space whose limiting Fréchet subdifferential at a certain point is smaller than the approximate subdifferential at the same point. So far this is the only example in which the first behaves better than the second. It is noteworthy that the space in the example is highly “pathological”: the space of continuous functions on the ordinal segment from zero to the first uncountable ordinal.
Under the assumptions of the Lyusternik–Graves theorem the codirectional compactness property is satisfied than ks to the Lyusternik condition \( \text{Im} F'(\pi) = Y \) and continuity of \( F'(\cdot) \) in the uniform operator topology which jointly guarantee that there is a \( k > 0 \) such that \( \|[F']^*(x)y^*\| \geq k\|y^*\| \) for any \( y^* \) and any \( x \) of a sufficiently small neighborhood of \( \pi \).

In the general case we set

\[
\text{Ker } D^*F(x, y) = \{y^* : 0 \in D^*F(x, y)(y^*)\}.
\]

**Theorem 2.** Suppose that a closed set-valued map \( F : X \to Y \) is codirectionally compact with respect to the approximate subdifferential at \((\pi, \overline{y}) \in \text{Gr } F\). Suppose also that \( \text{Ker } D^*F(\pi, \overline{y}) = \{0\} \). Then \( F \) is metrically regular near \((\pi, \overline{y})\).

**Proof.** We need to verify that the conditions of Theorem 1 are satisfied under the assumptions of Theorem 2. Assuming that this is not true, we can find a sequence of quadruples \((x_n, y_n, x_n^*, y_n^*)\) such that \((x_n, y_n)\) belongs to the \((1/n)\)-neighborhood of \((\pi, \overline{y})\), \( (x_n^*, y_n^*) \in 2\partial d((x_n, y_n), \text{Gr } F) \), \( \|y_n^*\| = n \) and \( \|x^*\| \leq 1 \) for any \( n \). According to Remark 1, we may assume in addition that \( y_n \in F(x_n) \). In this case \( \lambda(x_n^*, y_n^*) \in 2\partial d((x_n, y_n), \text{Gr } F) \) for any \( \lambda \in [0, 1] \) by Proposition 5.

Let \( v^* \) be a weak*-limit point of \( \{y_n^*/n\} \). Then \((0, v^*) \in 2\partial d((\pi, \overline{y}), \text{Gr } F)\), that is \( 0 \in D^*F(\pi, \overline{y})(v^*) \). By the assumption this means that \( v^* = 0 \). In other words, zero is the only weak*-limit point of the sequence, hence that \( v^* \) is the weak* limit of \( y_n^*/n \) as \( n \to \infty \). As \( F \) is codirectionally compact at \((\pi, \overline{y})\), it follows that the norms of \( y_n^*/n \) converge to zero. But the norms are all equal to one by definition. The contradiction completes the proof.

### 1.3. Fréchet subdifferential and a necessary regularity condition.

**Proposition 2.** Let \( \partial \) stands for the Fréchet (canonical) subdifferential. Then for any \( f \) and \( x \)

\[
|\nabla f|(x) \leq \inf \left\{ \|x^*\| : x^* \in \partial f(x) \right\}.
\]

**Proof.** If \( \partial f(x) = \emptyset \), then the inequality holds by the general convention. If \( x^* \in \partial f(x) \), then by definition

\[
f(x + h) - f(x) - \langle x^*, h \rangle \geq r(h)\|h\|,
\]

where \( r(h) \to 0 \) if \( \|h\| \to 0 \). It follows that for any \( h \)

\[
\frac{f(x) - f(x + h)}{\|h\|} \leq \|x^*\| + |r(h)|,
\]

whence the desired inequality.

Combining this with Proposition 1 and the second part of Proposition 2.3, we get

---

All other attempts to find a meaningful example (say, connected with non-regularity) which could demonstrate that the limiting Fréchet subdifferential in an Asplund space (which necessarily must not be weakly compactly generated) allows to get stronger results have not been successful and the question remains open.
Proposition 3. Let $X$ be an Asplund space and $\partial$ the (canonical) Fréchet subdifferential. Then for any lower semicontinuous $f$ and any open $U \subset X$

$$\inf_{x \in U} \nabla f(x) = \inf \{ \|x^*\| : x^* \in \partial f(x), \ x \in U \}.$$ 

Thus, taking Theorem 1.3 into account, we should expect that a necessary condition for metric regularity of maps between arbitrary Banach space can be stated in terms of Fréchet subdifferentials which, by Theorem 1, must be also sufficient in case of Asplund spaces.

Theorem 3. Let $X$ and $Y$ be Banach spaces, and let $D^*$ stand for the coderivative associated with the (canonical) Fréchet subdifferential. Let further $F$ be a closed set-valued map from $X$ into $Y$ which is metrically regular near $(\overline{x}, \overline{y}) \in \text{Gr} F$ with norm not greater than $K > 0$. Then there is a neighborhood $U$ of $(\overline{x}, \overline{y})$ such that the estimate $\|x^*\| \geq K^{-1}$ holds whenever $(x, v) \in U \cap \text{Gr} F$ and $(x^*, y^*)$ is $s$ such that $\|y^*\| = 1, x^* \in D^* F(x, v)(y^*)$.

Thus, by Theorem 1 the condition is necessary and sufficient for regularity of $F$ near $(\overline{x}, \overline{y})$ with norm not exceeding $K$ if both $X$ and $Y$ are Asplund spaces and $\partial$ is the canonical Fréchet subdifferential.

Proof. By Theorem 1.2 there is a neighborhood $U$ of $(\overline{x}, \overline{y})$ and an $\alpha > 0$ such that $|\nabla \varphi_y|(x, v) \geq (K + \alpha)^{-1}$ for all $(x, v) \in U \cap \text{Gr} F$ and $y \in Y$ (with the slope calculated with respect to the $\alpha$-metric in $X \times Y$). Consider the subset $Q$ of the unit sphere of $Y^*$ such that for any $y^* \in Q$ there is a nonzero $y \in Y$ normal to the unit ball at $y^*$.

Suppose now that $(x, v) \in U \cap \text{Gr} F$, $x^* \in D^* F(x, v)(y^*)$ (that is $(x^*, -y^*) \in \partial_{\chi_{\text{Gr} F}}(x, v)$) and $\|y^*\| = 1$. By the Bishop–Phelps theorem [10] for any $\varepsilon > 0$ there is a pair $(\overline{w}, w^*)$ such that $\overline{w} \in Y$, $w^* \in Y^*$, $\|w\| = \|w^*\| = \langle w, w^* \rangle = 1$ and $\|y^* - w^*\| < \varepsilon$. Set $y = v + \overline{w}$ and consider the function

$$\varphi_y(u, w) = \|y - w\| + \chi_{\text{Gr} F}(u, w).$$

Clearly $w^*$ belongs to the subdifferential of the norm in $Y$ at $\overline{w}$. Therefore $(0, w^*)$ belongs to the subdifferential of $\|y - \cdot\|_Y$, considered as a function on $X \times Y$, at $(x, v)$. It follows (by virtue of (2.2)) that

$$(0, w^*) + (x^*, -y^*) \in \partial \varphi_y(x, v).$$

By Proposition 2 this implies that

$$\|x^*\| + \varepsilon \alpha \geq |\nabla \varphi_y|(x, v) \geq K^{-1}.$$ 

The latter is valid for any $\varepsilon > 0$, whence $\|x^*\| \geq K^{-1}$ and the proof is completed.

As an immediate consequence we get the following nice criterion for local regularity of maps between finite-dimensional spaces.
Theorem 4. If $X$ and $Y$ are finite-dimensional Banach spaces, then $F$ is regular near $(\bar{x}, \bar{y}) \in \text{Gr} F$ if and only if

$$\text{Ker} D^* F(\bar{x}, \bar{y}) = \{0\},$$

where $D^*$ is the coderivative associated with the limiting subdifferential.

The sufficiency part follows from Theorem 2 in view of automatic codirectional compactness of finite-dimensional maps. To prove necessity, we have to notice that (by Theorem 3) $\|x^*\| \geq k^{-1}\|y^*\|$ if $x^*$ belongs to the value at $y^*$ of the Fréchet coderivative of $F$ at $(x, y)$ for $(x, y)$ sufficiently close to $(\bar{x}, \bar{y})$. Therefore the same relation must hold for the limiting coderivative at such points, in particular at $(\bar{x}, \bar{y})$ itself.

§2. Regularity at a point and subdifferential calculus of non-Lipschitz functions

We have seen in Chapter 2 (Theorem 2.2) that the approximate subdifferential of a sum of Lipschitz functions belongs to the sum of the approximate subdifferentials of the summands at the same point. Though one-sided, this is already a real calculus and it is no surprise that the result plays an important part in many theorems and applications of non-smooth analysis. It is easy to see that the result in general is not correct for functions not satisfying the Lipschitz condition. Let us consider, for example, the following two functions on the real line:

$$f_1(x) = \begin{cases} \sqrt{|x|}, & \text{if } x < 0, \\ 0, & \text{if } x \geq 0; \end{cases} \quad f_2(x) = \sqrt{|x|} \cdot \text{sign } x.$$

Then

$$(f_1 + f_2)(x) = \begin{cases} 0, & \text{if } x < 0, \\ \sqrt{x}, & \text{if } x \geq 0. \end{cases}$$

and $\partial f_1(0) = (-\infty, 0], \partial f_2(0) = \emptyset$, $\partial (f_1 + f_2)(0) = [0, \infty)$ (where $\partial$ is the limiting subdifferential).

Nonetheless, “exact” inclusions for subdifferentials of composite functions with non-Lipschitz components can be obtained under additional assumptions. Usually these assumptions contain conditions of two types: compactness conditions similar to those considered in §1 and so-called qualification conditions similar to the Mangasarian–Fromowitz condition in nonlinear programming. It turns out that the latter are always connected with regularity conditions of some special set-valued maps at a point.

In what follows we talk only about approximate subdifferentials, so all explanatory indices and words will be omitted. With obvious changes, the results carry over to the limiting Fréchet subdifferentials in Asplund spaces. It is not known whether similar results are valid for limiting $\beta$-subdifferentials in $\partial_\beta$-trustworthy spaces.

2.1. Normal cone to an intersection. Let $S_1, \ldots, S_k$ be subsets of a Banach space $X$ and $\bar{x} \in S = \bigcap S_i$. 

Proposition 4. Suppose that there is a $K > 0$ such that

$$d(x, S) \leq K \sum_{i=1}^{k} d(x, S_i)$$

(7)

for all $x$ of a neighborhood of $\overline{x}$. Then

$$N(\overline{x}, S) \subset N(S_1, \overline{x}) + \cdots + N(S_k, \overline{x}).$$

(8)

The proof follows immediately from Proposition 2.5 and Theorem 2.2.

Proposition 5. The inequality (7) is equivalent to regularity at $(\overline{x}, (0, \ldots, 0))$ of the following set-valued map $F: X \rightrightarrows X^k$:

$$F(x) = (S_1 - x) \times \cdots \times (S_k - x).$$

To prove the proposition it is sufficient to notice (assuming that $X^k$ is considered together with the additive metric) that

$$d((x_1, \ldots, x_k), F(x)) = d(x + x_1, S_1) + \cdots + d(x + x_k, S_k),$$

so that

$$F^{-1}((x_1, \ldots, x_k)) = \bigcap \{x : x_i \in S_i - x\} = \bigcap (S_i - x_i),$$

and $F^{-1}((0, \ldots, 0)) = \bigcap S_i = S$.

Definition 2. We say that the sets $S_1, \ldots, S_k$ are in a general position near $\overline{x} \in S_1 \cap \cdots \cap S_k$ if for any sequence of $2k$-tuples $(x_{1n}, \ldots, x_{kn}, x_{in}^*, \ldots, x_{kn}^*)$ such that $x_{in} \in S_i$, $x_{in} \to \overline{x}$, $x_{in}^* \in N(S_i, x_{in})$, $i = 1, \ldots, k$,

$$\left\| \sum_{i=1}^{k} x_{in}^* \right\| \to 0 \quad \& \quad \sup_n \sum_{i=1}^{k} \|x_{in}^*\| < \infty \implies \|x_{in}^*\| \to 0, \quad i = 1, \ldots, k.$$

Proposition 6. If sets $S_i$, $i = 1, \ldots, k$ are in the general position near $\overline{x} \in \bigcap S_i$, then $F$ is codirectionally compact at $(\overline{x}, (0, \ldots, 0))$.

Proof. We shall begin with calculation of the coderivative. We have

$$d((x, (x_1, \ldots, x_k)), \text{Gr} F) = \sum_{i=1}^{k} d(x, S_i - x) = \sum_{i=1}^{k} d(x + x_i, S_i).$$

Set

$$g_i(x, x_1, \ldots, x_k) = d(x + x_i, S_i), \quad h_i(x, u) = d(x + u, S_i)$$

By Theorem 2.2

$$\partial d((x, (x_1, \ldots, x_k)), \text{Gr} F) \subset \partial g_1(x, x_1, \ldots, x_k) + \cdots + \partial g_k(x, x_1, \ldots, x_k)$$
According to (SD)$_6$, $\partial g_i(x, x_1, \ldots, x_k)$ is the Cartesian product of $\partial h_i(x, x_i)$ and zeros in all components corresponding to $x_j, j \neq i$. On the other hand by (SD)$_7$, $\partial h_i(x, u) \subset \{(x^*, x^*): x^* \in \partial d(x + u, S_i)\}$. Therefore

$$(x^*, (x_1^*, \ldots, x_k^*)) \in \partial d(x, (x_1, \ldots, x_k), Gr F) \implies x_i^* \in \partial d(x + x_i, S_i) \quad \& \quad x_1^* + \cdots + x_k^* = x^*.$$ 

Thus, setting $y^* = -(x_1^*, \ldots, x_k^*)$ and taking Proposition 2.5 into account, we get

$$D^*F(\overline{F}, (0, \ldots, 0))(y^*) = \begin{cases} x_1^* + \cdots + x_k^*, & \text{if } x_i^* \in N(S_i, \overline{F}); \\ \varnothing, & \text{otherwise}. \end{cases}$$

It remains to refer to the definition of codirectional compactness to conclude the proof.

We observe further that the assumption that $S_1, \ldots, S_k$ are in general position near $\overline{F}$ implies the following qualification condition

$$x_i^* \in N(S_i, \overline{F}), \quad i = 1, \ldots, n \quad \& \quad x_1^* + \cdots + x_k^* = 0 \implies x_1^* = \cdots = x_k^* = 0.$$ (9)

The equality $\text{Ker } D^*F(\overline{F}, (0, \ldots, 0)) = \{0\}$ reduces to this condition in view of the above found expression for the coderivative. Thus, Propositions 4–6 and Theorem 2 imply the following result.

**Theorem 5.** Suppose that $S_i$ are closed sets which are in the general position near $\overline{F} \in \bigcap S_i$. Then (8) holds.

The general position condition is in particular satisfied if the sets have a sufficiently simple structure near $\overline{F}$ and normals to the sets at the point are independent.

**Definition 3.** A set $Q \subset X$ is normally compact at $\overline{F}$ if any sequence of $x_n^*$ such that $x_n^* \in N(Q, x_n)$ for some $x_n \rightarrow \overline{F}$ and $x_n^*$ weak$^*$ converge to zero, converges to zero in the norm topology.

It is clear that every set in a finite-dimensional space is normally compact at every its point and any set in a Banach space is normally compact at interior points. The simplest non-trivial example of a normally compact set is given by so called epi-Lipschitz set which has a structure of the epigraph of a Lipschitz function near the point in question.

**Proposition 7.** The sets $S_i$ are in the general position if all of them, but for at most one, are normally compact at $\overline{F}$ and the qualification condition (9) holds.

Let us sum up the results of our discussion. A natural sufficient condition for (8) to hold is (7) which amounts to regularity of $F$ at the corresponding point. We do not have in our disposal any criterion of regularity at a point which would be finer than criteria of regularity in a neighborhood. So we have to use the latter. In our specific case we get as a result a sufficient condition for (8) based on the point regularity criterion applied to $F$ at a corresponding point. The situation with functions is similar.
2.2. Sum of functions. Suppose now that we are given lower semicontinuous functions \( f_1, \ldots, f_k \) which are finite at \( \bar{x} \). Set \( f(x) = f_1(x) + \cdots + f_k(x) \). Here we shall discuss the main rule of the calculus of approximate subdifferentials which establishes a connection between the subdifferential of a sum and the subdifferentials of its terms at the same point. We begin by quoting the following result.

**Proposition 8** ([49], Proposition 6.1). Suppose that there is a \( K > 0 \) such that for any \( (x, \alpha) \) of a neighborhood of \( (\bar{x}, f(\bar{x})) \)

\[
d((x, \alpha), \text{epi} f) \leq K \sum_{i=1}^{k} d((x, \alpha_i), \text{epi} f_i)
\]

(10)

for all \( \alpha_i \) sufficiently close to \( f_i(\bar{x}) \), \( i = 1, \ldots, k \) and such that \( \alpha_1 + \cdots + \alpha_k = \alpha \). Then

\[
\partial f(\bar{x}) \subset \partial f_1(\bar{x}) + \cdots + \partial f_k(\bar{x})
\]

(11)

and

\[
\partial^\infty f(\bar{x}) \subset \partial^\infty f_1(\bar{x}) + \cdots + \partial^\infty f_k(\bar{x}).
\]

(12)

The proposition reduces the problem of estimating the subdifferential of a sum to the situation considered in the preceding subsection. Set

\[
\mathcal{X} = X \times \mathbb{R}^k; \quad \mathcal{G}_i = \{\bar{y} = (x, \alpha_1, \ldots, \alpha_k) \in \mathcal{X} : \alpha_i \geq f_i(x)\}, \quad \mathcal{G} = \bigcap_i \mathcal{G}_i.
\]

Then obviously,

\[
d((x, \alpha_1, \ldots, \alpha_k), \text{epi} f) \leq d((x, \alpha_1, \ldots, \alpha_k), \mathcal{G}).
\]

In other words, if the set-valued map

\[
\mathcal{F}(\bar{y}) = (\mathcal{G}_1 - \bar{y}) \times \cdots \times (\mathcal{G}_k - \bar{y})
\]

from \( \mathcal{X} \) into \( \mathbb{R}^k \) (cf. Proposition 5) is regular at \( (\bar{x}, (0, \ldots, 0)) \), then the inclusions (11), (12) hold.

Analysis of the general position condition and the normal compactness condition for \( \mathcal{G}_i \) leads to natural functional analogues of the conditions.

**Definition 4.** We say that functions \( f_1, \ldots, f_k \) are in general position near \( \bar{x} \in \text{dom} f_1 \cap \cdots \cap \text{dom} f_k \) if for any sequence of tuples

\[
((x_{in}, \alpha_{in}), \ldots, (x_{kn}, \alpha_{kn}), (x^*_{1n}, \beta_{1n}), \ldots, (x^*_{kn}, \beta_{kn}))
\]

such that

\[
\alpha_{in} \geq f_i(x_{in}); \quad (x^*_{in}, \beta_{in}) \in N(\text{epi} f_i, (x_{in}, \alpha_{in})); \quad (x_{in}, \alpha_{in}) \to (\bar{x}, f(\bar{x})),
\]

\[
\|x^*_{1n} + \cdots + x^*_{kn}\| \to 0, \quad \|x^*_{in}\| + \cdots + \|x^*_{kn}\| \leq C < \infty, \quad \beta_{in} \to 0, \quad i = 1, \ldots, n,
\]

each sequence \( \{x^*_{in}\}, i = 1, \ldots, k \) norm converges to zero.
(An exact reformulation of Definition 2 in our case gives: $\sum_i \beta_i \to 0$ and $\beta_i$ are uniformly bounded. But as far as $\beta \leq 0$ for $(x^*, \beta) \in N(\text{epi} f, (x, \alpha))$, it follows that all $\beta_i$ converge to zero.)

Taking the latter into account, we find that the qualification condition for functions analogous to (9) assumes the form

$$x_i^* \in \partial f(x), \ i = 1, \ldots, k, \quad x_1^* + \cdots + x_k^* = 0$$

$$\implies \ x_1^* = \cdots = x_k^* = 0. \quad (13)$$

**Definition 5.** A lower semicontinuous function $f(x)$ is normally compact at $\overline{x} \in \text{dom} f$ if for any sequence $(x_n, \alpha_n, x_n^*, \beta_n)$ such that $\alpha_n \geq f(x_n)$, $x_n \to \overline{x}$, $\alpha_n \to f(\overline{x})$, $(x_n^*, \beta_n) \in N(\text{epi} f, (x_n, \alpha_n))$, $\beta_n \to 0$, weak* convergence of $x_n^*$ to zero implies convergence in the norm topology.

The simplest non-trivial example of a normally compact function is given by the restriction of a Lipschitz function $f(x)$ to a set $S$ which is the closure of the intersection of an open set and a manifold of finite codimension, that is to say, the function $f_S$ equal to $f(x)$ on $S$ and zero outside of $S$.

The following theorem summarizes the discussion.

**Theorem 6.** Suppose that $f_1, \ldots, f_k$ are in general position near $\overline{x}$. Then the inclusions (11) and (12) hold for the approximate subdifferential. This is in particular true if all functions, but for at most one of them, are normally compact at $\overline{x}$ and the qualification condition (13) holds.

### 2.3. Composite function.

We follow quite a similar scheme when estimating the subdifferential of composition $f(x) = g(\Phi(x))$ of a lower semicontinuous function $g$ on $Y$ and a continuous map $\Phi: X \mapsto Y$. Let us fix as usual an $\overline{x} \in X$ and set $\overline{y} = \Phi(\overline{x})$, $\overline{\alpha} = f(\overline{x}) = g(\overline{y})$.

**Proposition 9.** Suppose that there is a $K > 0$ such that for all $(x, y, \alpha)$ of a neighborhood of $(\overline{x}, \overline{y}, \overline{\alpha})$

$$d((x, \alpha), \text{epi} f) \leq K(d((y, \alpha), \text{epi} g) + d((x, y), \text{Gr} \Phi)). \quad (14)$$

Then

$$\partial f(\overline{x}) \subseteq \bigcup_{y^* \in \partial g(\overline{y})} D^* \Phi(x)(y^*) \quad (15)$$

and

$$\partial^\infty f(\overline{x}) \subseteq \bigcup_{y^* \in \partial^\infty g(\overline{y})} D^* \Phi(x)(y^*). \quad (16)$$

We do not give a proof as it is not closely connected with the main topic of the paper and only remark that a possible proof would follow the same scheme as the proof of Proposition 6.1 in [49] (Proposition 8 above).

As in case of a sum of functions, the proposition allows us to reduce the situation to intersection of sets considered in 2.1. Indeed, let us consider the following sets in $X \times Y \times \mathbb{R}$:

$$S_1 = \{(x, y, \alpha) : (y, \alpha) \in \text{epi} g\} = X \times \text{epi} g;$$

$$S_2 = \{(x, y, \alpha) : (x, y) \in \text{Gr} \Phi\} = \text{Gr} \Phi \times \mathbb{R}.$$
The epigraph of $f$ is the projection of $S_1 \cap S_2$ onto $X \times \mathbb{R}$. Therefore
\[ d((x, \alpha), \text{epi } f) \leq d((x, y, \alpha), S_1 \cap S_2). \]
It follows that the inequality
\[ d((x, y, \alpha), S_1 \cap S_2) \leq K(d((x, y, \alpha), S_1) + d((x, y, \alpha), S_2)) \]
is sufficient to get (15) and (16).

The general position condition of Definition 2 reduces in our case to the following.

**Definition 6.** We say that $g$ and $\Phi$ are in general position near $(\overline{\mathbf{x}}, \overline{\mathbf{y}})$ if for any sequence of 8-tuples $(x_n, y_{1n}, y_{2n}, \alpha_n, x_n^*, y_{1n}^*, y_{2n}^*, \beta_n)$ such that

\[ x_n \to \overline{\mathbf{x}}, \quad y_{1n}, y_{2n} \to \overline{\mathbf{y}}, \quad \alpha_n \to g(\overline{\mathbf{y}}), \quad \|x_n^*\| \to 0, \quad \beta_n \to 0; \]
\[ (y_{1n}^*, \beta_n) \in N(\text{epi } g, (y_{1n}, \alpha_n)), \quad x_n^* \in D^*F(x_n)(-y_{2n}^*), \quad \|y_{1n}^*\| + \|y_{2n}^*\| \leq C < \infty \]
both $y_{1n}^*$ and $y_{2n}^*$ norm converge to zero, provided $\|y_{1n}^* + y_{2n}^*\| \to 0$.

The qualification condition (9) applied to $S_1$ and $S_2$ assumes the form

\[ y^* \in \partial^\infty g(\overline{\mathbf{y}}) \quad \& \quad 0 \in D^* \Phi(\overline{\mathbf{x}})(y^*) \quad \implies \quad y^* = 0 \quad (17) \]
(or just $\partial^\infty(\overline{\mathbf{y}}) \cap \text{Ker } D^* \Phi(\overline{\mathbf{x}}) = \{0\}$).

Finally, the normal compactness condition for $\Phi$ is formulated as follows: a sequence $(x_n^*, y_n^*)$ norm converges to zero if it converges in the weak*-topology and there are $(x_n, y_n) \to (\overline{\mathbf{x}}, \overline{\mathbf{y}})$ such that $(x_n, y_n) \in \text{Gr } \Phi$, $(x_n, y_n) \to (\overline{\mathbf{x}}, \overline{\mathbf{y}})$ and $x_n^* \in D^* \Phi(x_n)(y_n^*)$. This condition is excessively strong as the general position condition already contains the requirement that $\|x_n^*\| \to 0$. Therefore in this case it is sufficient to require that $y_{1n}^*$ norm converge to zero if they converge weakly* and $\|x_n^*\| \to 0$. The latter is precisely the condition that $\Phi$ is codirectionally compact near $(\overline{\mathbf{x}}, \overline{\mathbf{y}})$. Thus, we get the following result.

**Theorem 7.** If $g$ and $\Phi$ are in the general position near $(\overline{\mathbf{x}}, \overline{\mathbf{y}})$, then (15) and (16) hold. This is in particular true if the qualification condition (17) is satisfied and either $g$ is normally compact at $\overline{\mathbf{x}}$ or $\Phi$ is codirectionally compact at $(\overline{\mathbf{x}}, \overline{\mathbf{y}})$.

§ 3. Comments

3.1. It seems very strange that interrelations between the strong slope and subdifferentials has not been thoroughly studied, especially in view of the fact that the connection between the weak slope and Clarke’s generalized gradient has been in the center of attention of creators of the “non-smooth critical point theory” since the very beginning (see, for instance, [19] where the connection is discussed in details). However, Propositions 1–3 for seem to appear for the first time in this paper.

3.2. Theorem 1 (or to be more a accurate, Theorem 1a) was basically proved in [50]. The only difference is that a slightly stronger form of trustworthiness is used in [50], which covers the case of the approximate subdifferential, generalized gradient, all Fréchet subdifferentials and their limiting versions but not $\beta$-subdifferentials
of other types. At the same time, the proof given in [50] carries over to the situation considered here without any change. There are many subsequent publications in which sufficient subdifferential regularity conditions are proved for special classes of spaces, set-valued maps and/or subdifferentials [3], [8], [16], [22], [59], [64], [69], [78], [80]. All these results follow from [50], hence from Theorems 1 and 1a.

3.3. “Point” sufficient regularity conditions in terms of coderivatives (associated with the approximate, or limiting subdifferentials) were first obtained in [47] for set-valued maps of finite-dimensional spaces. They were later extended in to maps of arbitrary Banach spaces into finite-dimensional in [3] (for generalized gradient)\(^{15}\) and [50] (for approximate subdifferentials).

The search for additional assumptions that would lead to point regularity criteria for maps between infinite-dimensional spaces continued rather for a long time. Probably the first such condition was proved in [48] for integral operators associated with control systems. It was based on the so called “finite codimension property” which has received its final form much later in [50], [40]. In particular, a first version of Theorem 2 was proved in [50]. An extensive study of compactness conditions started a bit later, to a great extent stimulated by works of Borwein and Strojwas [14] and Loewen [71]. A number of new general concept were proposed in the course of the study: partial epi-Lipschitz property [59], partial normal compactness [80] and some others. The last paper, along with [61] and [85] led to crystallization of the concept defined here as the (sequential) codirectional compactness. At the same time it was shown in [51] that the topological equivalent of the codirectional compactness property, as well as the property of partial normal compactness and certain others are equivalent to the finite codimension property.

3.4. Theorem 3 belongs to Mordukhovich and Shao [78]. Its second part can be interpreted as the statement that the lower estimate of Theorem 1a gives an exact value of the constant of covering when the space is Asplund and \(\partial\) is the canonical Fréchet subdifferential. Theorem 4, as stated here, appeared in [76]. The necessity part of the theorem was proved by Mordukhovich [75] and, probably independently, by Krüger [64]. The sufficiency part was proved by Ioffe [47] somewhat earlier. Theorem 4 can be naturally viewed as a partial extension of the Fredholm alternative to set-valued maps. It is possible to state a more precise result, namely that under the assumptions of Theorem 4, the constant of covering of \(F\) near \((\bar{x}, \bar{y})\) is equal to

\[
\min \left\{ \|x^*\| : x^* \in D^* F(\bar{x}, \bar{y}), \|y^*\| \leq 1 \right\}.
\]

This estimate first appeared in [50] as a lower bound for the constant and then the opposite inequality was established in [76].

3.5. The equivalence of the Mangasarian–Fromowitz condition and stability of the map “right-hand part of the constraint map ⇒ feasible set” in smooth problems of mathematical programming was mentioned already by Robinson [89]. Auslender [5] extended this result to non-smooth problems with Lipschitz functions using the techniques of generalized gradients. The final result in terms of the limiting Fréchet subdifferential can be found in [91].

---

\(^{15}\)It should be said that [3] contains a “primal” theorem in which the conclusion is given in terms of Clarke’s tangent cones. But as was already mentioned in the comments section of Chapter 1, such results can always be reformulated in terms of Clarke’s normal cones and generalized gradients.
Proposition 4 was proved in [49]. But a systematic approach to subdifferential calculus with regularity estimates for certain set-valued maps used as qualification conditions was probably pioneered by Jourani and Thibault in [60]. Here we follow the scheme developed in [51] with slight modifications.

3.6. To conclude, we observe that in the local theory, in addition to subdifferential criteria, there are results of a different kind, so called “primal” space criteria stated in terms of the domain space objects such that tangent cones, directional derivatives etc. [1], [2], [4], [37], [38], [39], [42] and also intermediate criteria using approximations by sets of linear operators [20], [46], [65], [96]. Primal space criteria are often more transparent while subdifferential criteria are more convenient to formulate optimality conditions such as multiplier rules. Along with the fact that the first necessary regularity conditions were obtained in subdifferential terms, the latter was probably one of the reasons why subdifferential criteria have got more attention, especially last years. But the results of the first chapter seem to suggest that Aubin’s contention that primal theorems can potentially produce better results could be justified.
Bibliography


L. Thibault, Equivalence between metric regularity and graphical metric regularity for set-valued mappings, Unpublished note, Univ. de Montpellier, Montpellier 1999.
L. Thibault, Various forms of metric regularity, Unpublished note, Univ. de Montpellier, Montpellier 1999.

Department of Mathematics, Technion, Haifa 32000, Israel
E-mail address: ioffe@math.technion.ac.il

Received 26/JUN/2000

Typeset by \LaTeX