METRIC REGULARITY – A SURVEY

A.D. IOFFE *

October 27, 2015

In science things should be made as simple as possible.
Albert Einstein

All the great things are simple.
Winston Churchill

Abstract

Metric regularity theory lies in the very heart of variational analysis, a relatively new discipline whose appearance was to a large extent determined by needs of modern optimization theory in which such phenomena as non-differentiability and set-valued mappings naturally appear. The roots of the theory go back to such fundamental results of the classical analysis as the implicit function theorem, Sard theorem and some others. The paper offers a survey of the state-of-the-art of some principal parts of the theory along with a variety of its applications in analysis and optimization.

Contents

Introduction

Part 1. Theory

1. Classical theory: five great theorems
   1.1 Banach-Shauder open mapping theorem
   1.2 Regular points of smooth maps
   1.3 Inverse and implicit function theorems
   1.4 Sard theorem. Transversality.

2. Metric theory. Definitions and equivalences
   2.1 Local regularity
   2.2 Non-local regularity

3. Metric theory. Regularity criteria
   3.1 General criteria
   3.2 An application: density theorem

*Department of Mathematics, Technion, Haifa 32000, Israel
3.3 Infinitesimal criteria
3.4 Related concepts: metric subregularity, calmness, controllability, linear recession.

4. Metric theory. Perturbations and stability
4.1 Stability under Lipschitz perturbation
4.2 Strong regularity and metric implicit function theorem

5. Banach space theory
5.1 Techniques of variational analysis in Banach spaces
   5.1.1 Homogeneous set-valued mappings
   5.1.2 Tangent cones and contingent derivatives
   5.1.3 Subdifferentials, normal cones and coderivatives
5.2 Separable reduction
5.3 Contingent derivatives and primal regularity estimates
5.4 Dual regularity estimates
   5.4.1 Neighborhood estimates
   5.4.2 Perfect regularity and linear perturbations

6. Finite dimensional theory
6.1 Regularity
6.2 Subregularity and error bounds
6.3 Transversality

Part 2. Applications

7. Special classes of mappings
7.1 Error bounds
   7.1.1 Error bounds for convex functions
   7.1.2 Some general results on global error bounds
7.2 Mappings with convex graphs
   7.2.1 Convex processes
   7.2.2 Theorem of Robinson-Ursescu
   7.2.3 Mappings with convex graphs. Regularity rates
7.3 Single-valued Lipschitz maps
7.4 Polyhedral and semi-linear sets and mapping
7.5 Semialgebraic mappings, stratifications and the Sard theorem
   7.5.1 Basic properties
   7.5.2 Transversality

8. Some applications to analysis and optimization
8.1 Subdifferential calculus
8.2 Necessary conditions in constrained optimization
   8.2.1 Noncovering principle
   8.2.2 Exact penalty
   8.2.3 Optimality alternative
   8.2.4 Optimal control of differential inclusions
   8.2.5 Constraint qualifications
8.3 An abstract relaxed optimal control problem
8.4 Genericity in tame optimization
8.5 Method of alternating projection
8.6 Generalized equations
8.7 Variational inequalities over polyhedral sets
8.8 Differential inclusions (existence of solutions)

Introduction

Metric regularity has emerged during last 2-3 decades as one of the central concepts of a young discipline now often called *variational analysis*. The roots of this concept go back to a circle of fundamental regularity ideas of classical analysis embodied in such results as the implicit function theorem, Banach open mapping theorem, theorems of Lyusternik and Graves, on the one hand, and the Sard theorem and the Thom-Smale transversality theory, on the other.

Smoothness is the key property of the objects to which the classical results are applied. Variational analysis, on the other hand, appeals to objects that may lack this property: functions and maps that are non-differentiable at points of interest, set-valued mappings etc.. Such phenomena naturally appear in optimization theory and not only there.

In the traditional nonlinear analysis, regularity of a mapping (e.g. from a normed space or a manifold to another) at a certain point means that its derivative at the point is onto (the target space or the tangent space of the target manifold). This property, translated through available analytic or topological means to corresponding local properties of the mapping, plays a crucial role in studying some basic problems of analysis such as existence and behavior of solutions of a nonlinear equation $F(x) = y$ (with $F$ and $y$ viewed as data and $x$ as unknown) under small perturbations of the data. Similar problems appear if, instead of equation, we consider inclusion

$$y \in F(x)$$

(with $F$ a set-valued mapping this time) which, in essence, is the main object to study in variational analysis. The challenge here is evident: no clear way to approximate the mapping by simple objects like linear operators in the classical case.

The key step in the answer to the challenge was connected with the understanding of the metric nature of some basic phenomena that appear in the classical theory. This eventually led to the choice of the class of metric spaces as the main playground and subsequently to abandoning approximation as the primary tool of analysis in favor of a direct study of the phenomena as such. The ”metric theory” offers a rich collection of results that, being fairly general and stated in purely metric language, are nonetheless easily adaptable to Banach and finite dimensional settings (still among the most important in applications) and to various classes of mappings with special structure. Moreover, however surprising this may sound, the techniques coming from the metric theory sometimes appear more efficient, flexible and easy to use than the available Banach space techniques (associated with subdifferentials and coderivatives, especially in infinite dimensional Banach spaces).

---

1 Grothendick mentions ”ubiquity of stratified structures in practically all domains of geometry” in his 1984 *Esquisse d’un Programme*, see [74]
We shall not once see that proper use of metric criteria may lead to dramatic simplification of proofs and clarification of the ideas behind them. This occurs at all levels of generality, from results valid in arbitrary metric spaces to specific facts about even fairly simple classes of finite dimensional mappings.

It should be added furthermore that the central role played by distance estimates has determined a quantitative character of the theory (contrary to the predominantly qualitative character of the classical theory). Altogether, this opens gates to a number of new applications, such as say metric fixed point theory, differential inclusions, all chapters of optimization theory, numerical methods.

This paper has appeared as a result of two short courses I gave in the University of Newcastle and the University of Chile in 2013-2014. The goal was to give a brief account of some major principles of the theory of metric regularity along with the impression of how they work in various areas of analysis and optimization. The three principal themes that will be in the focus of attention are:

(a) regularity criteria (containing quantitative estimates for rates of regularity) including formal comparisons of their relative power and precision;

(b) stability problems relating to the effect of perturbations of the mapping on its regularity properties, on the one hand, and to solutions of equations, inclusions etc. on the other;

(c) role of metric regularity in analysis and optimization.

The existing regularity theory of variational analysis may look very technical. Many available proofs take a lot of space and use heavy techniques. But the ideas behind most basic results, especially in the metric theory, are rather simple and in many cases proper application of the ideas leads to noticeable (occasionally even dramatic) simplification and clarification of the proofs. This is a survey paper, so many results are quoted and discussed, often without proofs. As a rule, a proof is given if (a) the result is of a primary importance and the proof is sufficiently simple, (b) the result is new, (c) the access to the original publication containing the result is not very easy and especially (d) the proof is simpler (shorter, or looking more transparent) than available in the literature known to me.

And of course there are topics (some important) not touched upon in the paper, especially those that can be found in monographic literature. I mean first of all the books by Dontchev and Rockafellar [55] and Klatte and Kummer [109] in which metric regularity, in particular its finite dimensional chapter, is prominently presented. Among more specialized topics not touched upon in the survey, I would mention nonlinear regularity models, point subdifferential regularity criteria with associated compactness properties of subdifferentials and directional regularity.

The survey consists of two parts. The first part called ‘Theory’ contains an account of the basic ideas and principles of the metric regularity theory, first in traditional settings of the classical analysis and then for arbitrary set-valued mappings between various classes of spaces. In the second part ‘Applications’ we show how the theory works for some specific classes of maps that typically appear in variational analysis and and for a variety of fundamental existence, stability and optimization problems. In preparing this part of the survey the main efforts were focused on finding a productive balance between general principles and specific results and/or methods associated with the problem. This
declaration may look as a sort of truism but the point is that publications in which over-attachment to certain particular techniques of variational analysis (e.g. associated with generalized differentiation) leads to long and poorly digestible proofs of sufficiently simple and otherwise easily provable results is not an exceptional phenomenon.

To conclude the introduction I wish to express my thanks to J. Borwein and A. Joffre for inviting me to give the lectures that were the basis for this paper and to J. Borwein especially for his suggestion to write the survey. I also wish to thank D. Drusvyatskij and A. Lewis for the years of cooperation and many fruitful discussions and to A. Kruger and D. Klatte for many helpful remarks.

Dedication. 2015 and late 2014 have witnessed remarkable jubilees of six my good old friends. I dedicate this paper, with gratitude for the past and warm wishes for the future to

Prof. Vladimir Lin            Prof. Terry Rockafellar
Prof. Louis Nirenberg         Prof. Vladimir Tikhomirov
Prof. Boris Polyak            Prof. Nikita Vvedenskaya

Notation.

\(d(x, Q)\) – distance from \(x\) to \(Q\);
\(d(Q, P) = \inf\{\|x - u\| : x \in Q, u \in P\}\) – distance between \(Q\) and \(P\);
\(\text{ex}(Q, P) = \sup\{d(x, P) : x \in Q\}\) – excess of \(Q\) over \(P\);
\(h(Q, P) = \max\{\text{ex}(Q, P), \text{ex}(P, Q)\}\) – Hausdorff distance between \(Q\) and \(P\);
\(B(x, r)\) – closed ball of radius \(r\) and center at \(x\);
\(\partial B(x, r)\) – open ball of radius \(r\) and center at \(x\);
\(F\mid_Q\) – the restriction of a mapping \(F\) to the set \(Q\);
\(F : X \rightrightarrows Y\) – set-valued mapping;
\(\text{Graph} F = \{(x, y) : y \in F(x)\}\) – graph of \(F\);
\(I\) – the identity mapping (subscript, if present, indicates the space, e.g. \(I_X\));
\(\text{epi}\ f = \{(x, \alpha) : \alpha \geq f(x)\}\) – epigraph of \(f\);
\(\text{dom}\ f = \{x : f(x) < \infty\}\) – domain of \(f\);
\(i_Q(x)\) – indicator of \(Q\) (function equal to 0 on \(Q\) and \(+\infty\) outside);
\([f \leq \alpha] = \{x : f(x) \leq \alpha\}\) etc.;
\(X \times Y\) – Cartesian product of spaces;
\(X^*\) – adjoint of \(X\);
\(\langle x^*, x \rangle\) – the value of \(x^*\) on \(x\) (canonical bilinear form on \(X^* \times X\));
\(\mathbb{R}^n\) – the \(n\)-dimensional Euclidean space;
\(B\) – the closed unit ball in a Banach space (sometimes indicated by a subscript, e.g. \(B_X\) is the unit ball in \(X\));
\(S_X\) – the unit sphere in \(X\);
\(\text{Ker}\ A\) – kernel of the (linear) operator \(A\);
\(L^\perp = \{x^* \in X^* : \langle x^*, x \rangle = 0, \forall x \in L\}\) – annihilator of a subspace \(L \subset X\);
\(K^0 = \{x^* \in X^* : \langle x^*, x \rangle \leq 0, \forall x \in K\}\) – the polar of a cone \(K \subset X\);
\(\text{Im}\ A\) – image of the operator \(A\);
\(S(X)\) – collection of closed separable subspaces of \(X\).
\( L(X,Y) \) – the space of linear bounded operators \( X \to Y \) with the \textit{operator norm}:

\[
\|A\| = \sup_{\|x\| \leq 1} \|Ax\|.
\]

\( L \oplus M \) – direct sum of subspaces;
\( T_x M, N_x M \) – tangent and normal space to a manifold \( M \) at \( x \in M \);
\( T(Q, x) \) – contingent cone to a set \( Q \) at \( x \in Q \);
\( N(Q, x) \) – normal cone to \( Q \) at \( x \in Q \), often with a subscript (e.g. \( N_F \) is a Fréchet normal cone etc.)

We use the standard conventions \( d(x, \emptyset) = \infty; \inf \emptyset = \infty; \sup \emptyset = -\infty \) with one exception: when we deal with non-negative quantities we set \( \sup \emptyset = 0 \).

Part 1. Theory

1 Classical theory: five great theorems.

In this section all spaces are Banach.

1.1 Banach-Shauder open mapping theorem

\textbf{Theorem 1.1 (17, 161).} Let \( A : X \to Y \) be a linear bounded operator onto \( Y \), that is \( A(X) = Y \). Then \( 0 \in \text{int} \ A(B) \).

The theorem means that there is a \( K > 0 \) such that for any \( y \in Y \) there is an \( x \in X \) such that \( A(x) = y \) and \( \|x\| \leq K\|y\| \) (take as \( K \) the reciprocal of the radius of a ball in \( Y \) contained in the image of the unit ball in \( X \) under \( A \)).

\textbf{Definition 1.2 (Banach constant).} Let \( A : X \to Y \) be a bounded linear operator. The quantity

\[
C(A) = \sup \{ r \geq 0 : rB_Y \subset A(B_X) \} = \inf \{ \|y\| : y \notin A(B_X) \}
\]

will be called the \textit{Banach constant} of \( A \).

The following simple proposition offers two more expressions for the Banach constant. Given a linear operator \( A : X \to Y \), we set

\[
\|A^{-1}\| = \sup_{\|y\| \leq 1} d(0, A^{-1}(y)) = \sup_{\|y\| = 1} \inf \{ \|x\| : Ax = y \}.
\]

Of course, if \( A \) is a linear homeomorphism, this coincides with the usual norm of the inverse operator.

\textbf{Proposition 1.3 (calculation of \( C(A) \)).} For a bounded linear operator \( A : X \to Y \)

\[
C(A) = \inf_{\|y^*\| = 1} \|A^* y^*\| = \|A^{-1}\|^{-1}.
\]
1.2 Regular points of smooth maps. Theorems of Lyusternik and Graves.

Let $F : X \to Y$ be Fréchet differentiable at $x \in X$. It is said that $F$ is regular at $x$ if its derivative $F'(x)$ is a linear operator onto $Y$. Let $M \subset X$ be a smooth manifold. The tangent space $T_x M$ to $M$ at $x \in M$ is the collection of $h \in X$ such that $d(x + th, S) = o(t)$ when $t \to +0$.

**Theorem 1.4** (Lyusternik [126]). Suppose that $F$ is continuously differentiable and regular at $x$. Then the tangent space to the level set $M = \{ x : F(x) = F(x) \}$ at $x$ coincides with $\text{Ker} F'(x)$.

**Theorem 1.5** (Graves [73]). Let $F$ be a continuous mapping from a neighborhood of $x \in X$ into $Y$. Suppose that there are a linear bounded operator $A : X \to Y$ and positive numbers $\delta > 0$, $\gamma > 0$, $\varepsilon > 0$ such that $C(A) > \delta + \gamma$ and

$$\| F(x') - F(x) - A(x' - x) \| < \delta \| x' - x \|,$$

whenever $x$ and $x'$ belong to the open $\varepsilon$-ball around $x$. Then

$$B(F(x), \gamma t) \subset F(B(x, t))$$

for all $t \in (0, \varepsilon)$.

Here is a slight modification (quantities explicitly added) of the original proof by Graves.

**Proof.** We may harmlessly assume that $F(x) = 0$. Take $K > 0$ such that $KC(A) > 1 > K(\delta + \gamma)$, and let $\| y \| < \gamma t$ for some $t < \varepsilon$. Set $x_0 = x$, $y_0 = y$ and define recursively $x_n$, $y_n$ as follows:

$$y_{n-1} = A(x_n - x_{n-1}), \quad \| x_n - x_{n-1} \| \leq K \| y_{n-1} \|; \quad y_n = A(x_n - x_{n-1}) - (F(x_n) - F(x_{n-1})).$$

It is an easy matter to verify that

$$\| x_n - x_{n-1} \| \leq (K\delta)^{n-1} K \| y \|, \quad \| y_n \| \leq (K\delta)^n \| y \|$$

and $y_{n-1} - y_n = F(x_n) - F(x_{n-1})$, so that $(x_n)$ converges to some $x$ such that $F(x) = y$ and

$$\| x - x \| \leq \frac{K}{1 - K\delta} \| y \| \leq \gamma^{-1} \| y \| < t$$

as claimed. \qed

The theorem of Lyusternik was proved in 1934 and the theorem of Graves in 1950. Graves was apparently unaware of Lyusternik’s result and Lyusternik, in turn, of the open mapping theorem by Banach-Shauder. Nonetheless the methods they used in their proves were very similar. For that reason the following statement which is somewhat weaker than the theorem of Graves and somewhat stronger than the theorem of Lyusternik is usually called the Lyusternik-Graves theorem.
Theorem 1.6 (Lyusternik-Graves theorem). Assume that \( F : X \to Y \) is continuously differentiable and regular at \( \overline{x} \). Then for any positive \( r < C(F'(\overline{x})) \), there is an \( \varepsilon > 0 \) such that
\[
B(F(\overline{x}), rt) \subset F(B(\overline{x}, t)),
\]
whenever \( \|x - \overline{x}\| < \varepsilon, \ 0 \leq t < \varepsilon \).

It should be also emphasized that no differentiability assumption is made in the theorem of Graves. In this respect Graves was much ahead of time. Observe that the mapping \( F \) in the theorem of Graves can be viewed as a perturbation of \( A \) by a \( \delta \)-Lipschitz mapping. With this interpretation the theorem of Graves can be also viewed as a direct predecessor of Milyutin’s perturbation theorem (Theorem 4.2 in the fourth section), which is one of the central results in the regularity theory of variational analysis.

1.3 Inverse and implicit function theorem

Theorem 1.7 (Inverse function theorem). Suppose that \( F \) is continuously differentiable at \( \overline{x} \) and the derivative \( F'(\overline{x}) \) is an invertible operator onto \( Y \). Then there is a mapping \( G \) into \( X \) defined in a neighborhood of \( \overline{y} = F(\overline{x}) \), strictly differentiable at \( \overline{y} \) and such that
\[
G'(\overline{y}) = (F'(\overline{x}))^{-1} \quad \text{and} \quad F \circ G = I_Y
\]
in the neighborhood.

The shortest among standard proofs of the theorem is based on the contraction mapping principle (see e.g. the second proof of the theorem in [55]). But equally short proof follows from the theorem of Lyusternik-Graves.

Proof. Set \( A = F'(\overline{x}) \). Then \( F(x') - F(x) - A(x' - x) = r(x', x)\|x' - x\| \), where \( \|r(x', x)\| \to 0 \) when \( x, x' \to \overline{x} \). As \( A \) is invertible, there is a \( K > 0 \) such that \( \|Ah\| \geq K\|h\| \). Hence \( \|F(x') - F(x)\| \geq (K - r(x, x'))\|x' - x\| > 0 \) if \( x, x' \) are close to \( \overline{x} \). This means that \( F \) is one-to-one in a neighborhood of \( \overline{x} \). But by the Lyusternik-Graves theorem, \( F(U) \) covers a certain open neighborhood of \( \overline{y} \). Hence \( G = F^{-1} \) is defined in a neighborhood of \( F(\overline{x}) \). So given \( y \) and \( y' \) close to \( \overline{y} = F(\overline{x}) \) and let \( x', x \) be such that \( F(x') = y', F(x) = y \). Then as we have seen \( \|y - y'\| \geq K\|x - x'\| \). We have
\[
A^{-1}(F(x') - F(x) - A(x' - x)) = A^{-1}(y' - y) - G(y') - G(y),
\]
so that
\[
\|G(y') - G(y) - A^{-1}(y' - y)\| \leq \|A^{-1}\|\|F(x') - F(x) - A(x' - x)\| \\
= \|A^{-1}\|\|r(x', x)\|\|x' - x\| \leq q(y, y')\|y' - y\|,
\]
where \( q(y, y') = Kr(G(y), G(y')) \) obviously goes to zero when \( y, y' \to \overline{y} \).

Theorem 1.8 (Implicit function theorem). Let \( X, Y, Z \) be Banach spaces, and let \( F \) be a mapping into \( Z \) which is defined in a neighborhood of \( (\overline{x}, \overline{y}) \in X \times Y \) and strictly differentiable at \( (\overline{x}, \overline{y}) \). Suppose further that the partial derivative \( F_y(\overline{x}, \overline{y}) \) is an invertible operator. Then there are neighborhoods \( U \subset X \) of \( \overline{x} \) and \( W \subset Z \) of \( \overline{y} = F(\overline{x}, \overline{y}) \) and a
mapping $S : U \times W \rightarrow Y$ such that $(x, z) \mapsto (x, S(x, z))$ is a homeomorphism of $U \times W$ onto a neighborhood of $(\bar{x}, \bar{y})$ in $X \times Y$ and

$$F(x, S(x, z)) = z, \quad \forall \ x \in U, \ \forall \ z \in W$$

The mapping $S$ is strictly differentiable at $(\bar{x}, \bar{z})$ with

$$S_x(\bar{x}, \bar{z}) = (F_y(\bar{x}, \bar{y}))^{-1}, \quad S_x(\bar{x}, \bar{z}) = (F_y(\bar{x}, \bar{y}))^{-1}F_x(\bar{x}, \bar{y}).$$

The simplest proof of the theorem is obtained by application of the inverse mapping theorem to the following map $X \times Y \rightarrow X \times Z$ (see e.g. [55]):

$$\Phi(x, y) = \left( \begin{array}{c}
x \\
F(x, y)
\end{array} \right).$$

### 1.4 Sard theorem. Transversality.

**Definition 1.9** (critical and regular value). Let $X$ and $Y$ be Banach spaces, and let $F$ be a mapping into $Y$ defined and continuously differentiable on an open set of $U \subset X$. A vector $y \in Y$ is called a critical value of $F$ if there is an $x \in U$ such that $F(x) = y$ and $x$ is a singular point of $F$. Any point in the range space which is not a critical value is called a regular value, even if it does not belong to $\text{Im} \ F$. Thus $y$ is a regular value if either $y \neq F(x)$ for any $x$ of the domain of $F$ or $\text{Im} \ F'(x) = Y$ for every $x$ such that $F(x) = y$.

**Theorem 1.10** (Sard [160]). Let $\Omega$ be an open set in $\mathbb{R}^m$ and $F$ a $C^k$-mapping from $\Omega$ into $\mathbb{R}^m$. Then the Lebesgue measure of the set of critical values of $F$ is equal to zero, provided $k \geq n - m + 1$.

For a proof of a "full" Sard theorem see [1]; a much shorter proof for $C^\infty$ functions can be found in [137].

**Definition 1.11** (transversality). Let $F : X \rightarrow Y$ be a $C^1$-mapping, and let $M \subset Y$ be a $C^1$-submanifold. Let finally $x$ be in the domain of $F$. We say that $F$ is transversal to $M$ at $x$ if either $y = F(x) \notin M$ or $y \in M$ and $\text{Im} \ F'(x) + T_y M = Y$. It is said that $F$ is transversal to $M$ : $F \pitchfork M$, if it is transversal to $M$ at every $x$ of the domain of $F$.

We can also speak about transversality of two manifolds $M_1$ in $M_2$ in $X$: $M_1 \pitchfork M_2$ at $x \in M_1 \cap M_2$ if $T_x M_1 + T_x M_2 = X$. For our future discussions, it is useful to have in mind that the latter property can be equivalently expressed in dual terms: $N_x M_1 \cap N_x M_2 = \{0\}$, where $N_x M \subset X^*$ is the normal space to $M$ at $x$, that is the annihilator of $T_x M$.

A connection with regularity is immediate from the definition: if $(L, \varphi)$ is a local parametrization for $M$ at $y$ and $y = F(x)$, then transversality of $F$ to $M$ at $x$ is equivalent to regularity at $(x, 0, 0)$ of the mapping $\Phi : X \times L \rightarrow Y$ given by $\Phi(u, v) = F(u) - \varphi(v)$.

The connection of transversality and regularity is actually much deeper. Let $P$ be also a Banach space and let $F : X \times P \rightarrow Y$. We can view $F$ as a family of mappings from $X$ into $Y$ parameterized by elements of $P$. Let us denote “individual” mappings $x \mapsto F(x, p)$ by $F(\cdot, p)$. Let further $M \subset Y$ be a submanifold, and let $\pi : X \times P \rightarrow P$ be the standard Cartesian projection $(x, p) \rightarrow p$.
Proposition 1.12. Suppose $F$ is transversal to $M$ and $Q = F^{-1}(M)$ is a manifold. Let finally $\pi|_Q$ stands for the restriction of $\pi$ to $Q$. Then $F(\cdot, p)$ is transversal to $M$, provided $p$ is a regular value of $\pi|_Q$.

Combining the proposition with the Sard theorem, we get the following (simple version of) transversality theorem of Thom

Theorem 1.13 (see e.g. [76]). Let $X$, $Y$ and $P$ be finite dimensional Banach spaces Let $M \subset Y$ be a $C^r$-manifold, and let $F : X \times P \to Y$ be a $C^k$-mapping ($k \leq r$). Assume that $F \cap M$ and $k > \dim X - \operatorname{codim} M$. Then $F(\cdot, p) \cap M$ for each $p \in P$ outside of a subset of $P$ with dim $P$-Lebesgue measure zero.

2 Metric theory. Definitions and equivalences.

Here $X$ and $Y$ are metric space. We use the same notation for the metrics in both hoping this would not lead to any difficulties.

2.1 Local regularity

We start with the simplest and the most popular case of local regularity near a certain point of the graph. So let an $F : X \Rightarrow Y$ be given as well as a $(\bar{x}, \bar{y}) \in \text{Graph } F$.

Definition 2.1 (local regularity properties). We say that $F$ is

- open or covering at a linear rate near $(\bar{x}, \bar{y})$ if there are $r > 0$, $\varepsilon > 0$ such that

$$B(y, rt) \cap B(\bar{y}, \varepsilon) \subset F(B(x, t)), \quad \forall (x, y) \in \text{Graph } F, \ d(x, \bar{x}) < \varepsilon, \ t \geq 0.$$  

The upper bound $\text{sur } F(\bar{x}|\bar{y})$ of such $r$ is the modulus or rate of surjection of $F$ near $(\bar{x}, \bar{y})$. If no such $r$, $\varepsilon$ exist, we set $\text{sur } F(\bar{x}|\bar{y}) = 0$;

- metrically regular near $(\bar{x}, \bar{y}) \in \text{Graph } F$ if there are $K > 0$, $\varepsilon > 0$ such that

$$d(x, F^{-1}(y)) \leq K d(y, F(x)), \quad \text{if } d(x, \bar{x}) < \varepsilon, \ d(y, \bar{y}) < \varepsilon.$$  

The lower bound $\text{reg } F(\bar{x}|\bar{y})$ of such $K$ is the modulus or rate of metric regularity of $F$ near $(\bar{x}, \bar{y})$. If no such $K$, $\varepsilon$ exist, we set $\text{reg } F(\bar{x}|\bar{y}) = \infty$.

- pseudo-Lipschitz or has the Aubin property near $(\bar{x}, \bar{y})$ if there are $K > 0$ and $\varepsilon > 0$ such that

$$d(y, F(x)) \leq K d(x, u), \quad \text{if } d(x, \bar{x}) < \varepsilon, \ d(y, \bar{y}) < \varepsilon, \ y \in F(u).$$  

The lower bound $\text{lip } F(\bar{x}|\bar{y})$ is the Lipschitz modulus or rate of $F$ near $(\bar{x}, \bar{y})$. If no such $K$, $\varepsilon$ exist, we set $\text{lip } F(\bar{x}|\bar{y}) = \infty$.

Note a difference between the covering property and the conclusions of theorems of Lyusternik and Graves: the theorems deal only with the given argument $\bar{x}$ while in the definition we speak about all $x \in \text{dom } F$ close to $\bar{x}$. This difference that was once a subject of heated discussions is in fact illusory as under the assumptions of the theorems of
Lyusternik and Graves the covering property in the sense of the just introduced definitions is automatically satisfied.

The key and truly remarkable fact for the theory is that the three parts of the definition actually speak about the same phenomenon. Namely the following holds true unconditionally for any set-valued mapping between two metric spaces.

**Proposition 2.2** (local equivalence). $F$ is open at a linear rate near $(\bar{x}, \bar{y}) \in \text{Graph } F$ if and only if it is metrically regular near $(\bar{x}, \bar{y})$ and if and only if $F^{-1}$ has the Aubin property near $(\bar{y}, \bar{x})$. Moreover, under the convention that $0 \cdot \infty = 1$,

$$\text{sur } F(\bar{y}|\bar{y}) \cdot \text{reg } F(\bar{y}|\bar{y}) = 1; \quad \text{reg } F(\bar{y}|\bar{y}) = \text{lip } F^{-1}(\bar{y}|\bar{x}).$$

**Remark 2.3.** In view of the proposition it makes sense to use the word *regular* to characterize the three properties. This terminology would also emphasize the ties with the classical regularity concept. We observe further that while the rates of regularity are connected with specific distances in $X$ and $Y$, the very fact that $F$ is regular near certain point is independent of the choice of specific metrics. Thus, although the definitions explicitly use metrics the regularity is a topological property.

The proof of the proposition is fairly simple (we shall get it as a consequence of a more general equivalence theorem later in this section). But the way to it was surprisingly long (see brief bibliographic comments at the end of the section).

There are other equivalent formulations of the properties. For instance, the definition of linear openness/covering can be modified by adding the constraint $0 \leq t < \varepsilon$ (see [92]); a well known modification of the definition of metric regularity includes the condition that $d(y, F(x)) < \varepsilon$. The only difference is that the $\varepsilon$’s in the original and modified definitions may be different.

**Definition 2.4** (graph regularity [164]). $F$ is said to be graph-regular at (or near) $(\bar{x}, \bar{y}) \in \text{Graph } F$ if there are $K > 0$, $\varepsilon > 0$ such that the inequality

$$d(x, F^{-1}(y)) \leq Kd((x,y), \text{Graph } F), \quad (2.1)$$

holds, provided $d(x, \bar{x}) < \varepsilon$, $d(y, \bar{y}) < \varepsilon$.

**Proposition 2.5** (metric regularity vs graph regularity [164]). Let $F : X \rightrightarrows Y$, and $(\bar{x}, \bar{y}) \in \text{Graph } F$. Then $F$ is metrically regular at $(\bar{x}, \bar{y})$ if and only if it is graph-regular at $(\bar{x}, \bar{y})$.

Note that, unlike the equivalence theorem, the last proposition is purely local: the straightforward non-local extension of this result (e.g. along the lines of the subsection below) is wrong.

### 2.2 Non-local regularity.

As we have already mentioned, most of current researches focus on local regularity. (although the first abstract definition of the covering property given in [45] was absolutely non-local). To a large extent this is because of the close connection of modern variational analysis studies with optimization theory which is basically interested in local results:
optimality conditions, stability of solutions under small perturbations, etc. Another less visible reason is that non-local regularity is a more delicate concept: in the non-local case we cannot freely change the regularity domain that is an integral part of the definition. Meanwhile non-local regularity is a powerful instrument for proving e.g. various existence theorems (see e.g. subsection 8.7).

Let \( U \subset X \) and \( V \subset Y \) (we usually assume \( U \) and \( V \) open), let \( F : X \rightrightarrows Y \), and let \( \gamma(\cdot) \) and \( \delta(\cdot) \) be extended-real-valued functions on \( X \) and \( Y \) assuming positive values (possibly infinite) respectively on \( U \) and \( V \).

**Definition 2.6** (non-local regularity properties [92]). We say that \( F \) is

- \( \gamma \)-open (or \( \gamma \)-covering) at a linear rate on \( U \times V \) if there is an \( r > 0 \) such that
  \[
  B(F(x), rt) \bigcap V \subset F(B(x, t)),
  \]
  if \( x \in U \) and \( t < \gamma(x) \). Denote by \( \text{sur}_\gamma F(U|V) \) the upper bound of such \( r \). If no such \( r \) exists, set \( \text{sur}_\gamma F(U|V) = 0 \). We shall call \( \text{sur}_\gamma F(U|V) \) the modulus (or rate) of \( \gamma \)-openness of \( F \) on \( U \times V \);

- \( \gamma \)-metrically regular on \( U \times V \) if there is a \( K > 0 \) such that
  \[
  d(x, F^{-1}(y)) \leq K d(y, F(x)),
  \]
  provided \( x \in U \), \( y \in V \) and \( K d(y, F(x)) < \gamma(x) \). Denote by \( \text{reg}_\gamma F(U|V) \) the lower bound of such \( K \). If no such \( K \) exists, set \( \text{reg}_\gamma F = \infty \). We shall call \( \text{reg}_\gamma F(U|V) \) the modulus (or rate) of \( \gamma \)-metric regularity of \( F \) on \( U \times V \);

- \( \delta \)-pseudo-Lipschitz on \( U \times V \) if there is a \( K > 0 \) such that
  \[
  d(y, F(x)) \leq K d(x, u)
  \]
  if \( x \in U \), \( y \in V \), \( K d(x, u) < \delta(y) \) and \( y \in F(u) \). Denote by \( \text{lip}_\delta F(U|V) \) the lower bound of such \( K \). If no such \( K \) exists, set \( \text{lip}_\delta F = \infty \). We shall call \( \text{lip}_\delta F(U|V) \) the \( \delta \)-Lipschitz modulus of \( F \) on \( U \times V \).

If \( U = X \) and \( V = Y \), let us agree to write \( \text{sur}_\gamma, \text{reg}_\gamma, \text{lip}_\delta \) instead of \( \text{sur}_\gamma F(X|Y), \text{reg}_\gamma F(X|Y), \text{lip}_\delta F(X|Y) \), etc. The role of the functions \( \gamma \) and \( \delta \) is clear from the definitions. They determine how far we shall reach from any given point in verification of the defined properties. It is therefore natural to call them regularity horizon functions. Such functions are inessential for local regularity (see e.g. Exercise 2.8 below). But for fixed \( U \) and \( V \) regularity horizon function is an essential element of the definition. Regularity properties corresponding to different \( \gamma \) may not be equivalent (see Example 2.2 in [97] and also Exercise 2.8 below).

**Theorem 2.7** (equivalence theorem). The following three properties are equivalent for any pair of metric spaces \( X, Y \), any \( F : X \rightrightarrows Y \), any \( U \subset X \) and \( V \subset Y \) and any (extended-real-valued) function \( \gamma(x) \) which is positive on \( U \):

(a) \( F \) is \( \gamma \)-open at a linear rate on \( U \times V \);

(b) \( F \) is \( \gamma \)-metrically regular on \( U \times V \);

(c) \( F^{-1} \) is \( \gamma \)-pseudo-Lipschitz on \( V \times U \).

Moreover (under the convention that \( 0 \cdot \infty = 1 \))

\[
\text{sur}_\gamma F(U|V) \cdot \text{reg}_\gamma F(U|V) = 1, \quad \text{reg}_\gamma F(U|V) = \text{lip}_\gamma F^{-1}(V|U).
\]
Proof. The implication (b) \( \Rightarrow \) (c) is trivial. Hence \( \text{lip}_x \gamma F^{-1}(V|U) \leq \text{reg}_x F(U|V) \). To prove that (c) \( \Rightarrow \) (a), take a \( K > \text{lip}_x \gamma F^{-1} \) and an \( r < K^{-1} \), let \( t < \gamma(x) \), and let \( x \in U \), \( y \in V \), \( v \in F(x) \) and \( y \in B(v, tr) \). Then \( d(y,v) < r\gamma(x) \) and by (c) \( d(x,F^{-1}(y)) \leq Kd(y,v) < r^{-1}d(y,v) \leq t \). It follows that there is a \( u \) such that \( y \in F(u) \) and \( d(x,u) < t \). Hence \( y \in F(B(x,t)) \). It follows that \( r \leq \text{sur}_x F \), or equivalently \( 1 \leq K\text{sur}_x F \). But \( r \) can be chosen arbitrarily close to \( K^{-1} \) and and \( K \) can be chosen arbitrarily close to \( \text{lip}_x \gamma F^{-1} \). So we conclude that \( \text{sur}_x F \cdot \text{lip}_x \gamma F^{-1} \geq 1 \).

Let finally (a) hold with some \( r > 0 \), let \( x \in U \), \( y \in V \), and let \( d(y,F(x)) < \gamma(x) \). Choose a \( v \in F(x) \) such that \( d(y,v) < r\gamma(x) \) and set \( t = d(y,v)/r \). By (a) there is a \( u \in F^{-1}(y) \) such that \( d(x,u) \leq t \). Thus \( d(x,F^{-1}(y)) \leq t = d(y,v)/r \). But \( d(y,v) \) can be chosen arbitrarily close to \( d(y,F(x)) \) and we get \( d(x,F^{-1}(y)) \leq r^{-1}d(y,F(x)) \), that is \( r \cdot \text{reg}_x F \leq 1 \). On the other hand \( r \) can be chosen arbitrarily close to \( \text{sur}_x F \) and we can conclude that \( \text{sur}_x F \cdot \text{reg}_x F \leq 1 \) so that

\[
1 \geq \text{sur}_x F(U|V) \cdot \text{reg}_x F(U|V) \geq \text{sur}_x F(U|V) \cdot \text{lip}_x \gamma F(V|U) \geq 1
\]

which completes the proof of the theorem.

The most important example of the horizon function is \( m(x) = d(x,X\setminus U) \). The meaning is that we need not look at points beyond \( U \). We shall call \( F \) Milyutin regular on \( U \times V \) if it is \( m \)-regular. (This is actually the type of regularity implicit in the definition given in [45].) In what follows we shall deal only with Milyutin regularity when speaking about non-local matters.

Exercise 2.8. Prove that \( F \) is regular near \((\bar{x}, \bar{y}) \in \text{Graph} F \) if and only if it is Milyutin regular on \( \tilde{B}(\bar{x}, \varepsilon) \times \tilde{B}(\bar{y}, \varepsilon) \) for all sufficiently small \( \varepsilon \).

We conclude the section with a useful result (a slight modification of the corresponding result in [88]) showing that, as far as metric regularity is concerned, any set-valued mapping can be equivalently and in a canonical way replaces by a single-valued mapping continuous on its domain.

Proposition 2.9 (single-valued reduction). Let \( X \times Y \) be endowed with the \( \xi \)-metric. Let \( F \) be Milyutin regular on \( U \times V \) with \( \text{sur}_m F(U|V) \geq r > 0 \). Consider the mapping \( \mathcal{P}_F : \text{Graph} F \to Y \) which is the restriction to \( \text{Graph} F \) of the Cartesian projection \((x,y) \to y \).

Then \( \mathcal{P}_F \) is Milyutin regular on \((U \times Y) \times V \) and \( \text{sur}_m \mathcal{P}_F(U \times Y|V) = \text{sur}_m F(U|V) \) if \( X \times Y \) is considered e.g. with the \( \xi \)-metric.

A few bibliographic comments. To begin with, it is worth mentioning that in the classical theory no interest to metric estimates can be traced. The covering property close to the covering part of Milyutin regularity was introduced in [45] and attributed to Milyutin. An estimate of metric regularity type first time appeared in Lyusternik’s paper [126] but for \( x \) restricted to the kernel of the derivative. In Ioffe-Tikhomirov [103] metric regularity was proved under the assumptions of the Graves theorem. Robinson was probably the first to consider set-valued mappings. In [150] he proved metric regularity of the mapping \( F(x) = f(x) + K \) (even of the restriction of this mapping to a convex closed subset of \( X \)), under the assumptions that \( f : X \to Y \) is continuously differentiable and \( K \subset Y \) is a closed
convex cone, under certain qualification condition extending Lyusternik’s $\text{Im } F'(x) = Y$. The definition of $\gamma$-regularity was given in [92].

Equivalence of covering and metric regularity was explicitly mentioned (without proof) in the paper of Dmitruk-Milyutin-Osmolovski [45] that marked the beginning of systematic study of the regularity phenomena, in particular in metric spaces, and Ioffe in [82] stated a certain equivalence result (Proposition 11.12 – see [87] for its proof) which, as was much later understood, contains even more precise information about the connection of the covering and metric regularity properties. And the pseudo-Lipschitz property was introduced by Aubin in [6].

This was the sequence of events prior to the proof of the equivalence of the three properties by Borwein-Zhuang [26] and Penot [142]. It has to be mentioned that in both papers more general "nonlinear" properties were considered. In this connection we also mention the paper by Frankowska [67] with a short proof of nonlinear openness and some pseudo-Hölder property.

3 Metric theory. Regularity criteria.

This section is central. Here we prove necessary and sufficient conditions for regularity. The key results is Theorems 3.1, 3.2 and 3.3 containing general regularity criteria. The criteria (especially the first of them) will serve us as a basis for obtaining various qualitative and quantitative characterizations of regularity in this and subsequent sections. The criteria are very simple to prove and, at the same time, provide us with an instrument of analysis which is both powerful and easy to use. We shall see this already in this section and many times in what follows. In the second subsection we consider infinitesimal criteria for local regularity based on the concept of slope, the central in the local theory.

Given a set-valued mapping $F : X \rightrightarrows Y$, we associate with it the following functions that will be systematically used in connection with the criteria and their applications:

$$
\varphi_y(x,v) = \begin{cases} 
  d(y,v), & \text{if } v \in F(x); \\
  +\infty, & \text{otherwise}
\end{cases}
\quad \psi_y(x) = d(y,F(x)); \quad \psi_y(x) = \liminf_{u \to x} \psi(u).
$$

Note that $\varphi_y$ is Lipschitz continuous on Graph $F$, hence it is lower semicontinuous whenever Graph $F$ is a closed set.

3.1 General criteria.

Given a $\xi > 0$, we define the $\xi$-metric on $X \times Y$ by

$$
d_\xi((x,y),(x',y')) = \max\{d(x,x'),\xi d(y,y')\}.
$$

**Theorem 3.1** (criterion for Milyutin regularity). Let $U \subset X$ and $V \subset Y$ be open sets, and let $F : X \rightrightarrows Y$ be a set-valued mapping whose graph is complete in the product metric. Let further $r > 0$ and there be a $\xi > 0$ such that for any $x \in U$, $y \in V$, $v \in F(x)$ with $0 < d(y,v) < r\text{m}(x)$, there is a pair $(u,w) \in \text{Graph } F$ different from $(x,v)$ and such that

$$
d(y,w) \leq d(y,v) - r d_\xi((x,v),(u,w)). \quad (3.1)
$$
Then $F$ is Milyutin regular on $U \times V$ with\n\[ \text{sur}_n F(U|V) \geq r. \]

Conversely, if $F$ is Milyutin regular on $U \times V$, then for any positive $r < \text{sur}_r F(U|V)$,\nany $\xi \in (0, r^{-1})$, any $x \in U$, $v \in F(x)$ and $y \in V$ satisfying $0 < d(y, v) < r\gamma(x)$, there is a\npair $(u, w) \in \text{Graph } F$ different from $(x, v)$ such that (3.1) holds.

The theorem offers a very simple geometric interpretation of the regularity phenomenon:\nit means that $F$ is regular if for any $(x, v) \in \text{Graph } F$ and any $y \neq v$ there is a point in\nthe graph whose $Y$-component is closer to $y$ (than $v$) and the distance from the new point\nto the original point $(x, v)$ is proportional to the gain in the distance to $y$.

**Proof.** We have to verify that, given $(\bar{x}, \bar{v}) \in \text{Graph } F$ with $\bar{x} \in U$, $y \in V$ and $0 < d(y, \bar{v}) \leq rt$, $t < m(\bar{v})$, there is a $u \in B(x, t)$ such that $y \in F(u)$. We have $\varphi_y(\bar{x}, \bar{v}) \leq rt$.

By Ekeland’s variational principle (see e.g. [27]) there is a pair $(\hat{x}, \hat{v}) \in \text{Graph } F$ such that\nde $\varphi_y((\hat{x}, \hat{v})) = 0$, that is $y = \hat{v} \in F(\hat{x})$. Indeed, $\hat{x} \in U$, so\nby the assumption if $y \neq \hat{v}$, there is a pair $(u, w) \neq (\hat{x}, \hat{v})$ and such that (3.1) holds with\nde $(\hat{x}, \hat{v})$ as $(x, v)$ which however contradicts (3.2). This proves the first statement.

Assume now that $F$ is Milyutin regular on $U \times V$ with the surjection modulus not\nsmaller than $r$. Take a positive $\xi < r^{-1}$ and $x \in U$, $y \in V$, $v \in F(x)$ with $d(y, v) < r\gamma(x)$.\nTake a small $\epsilon \in (0, r)$ and choose a $t \in (0, m(x))$ such that $(r - \epsilon)t \leq d(y, v) < rt$. By\nregularity there is a $u$ such that $d(u, x) < t$ and $y \in F(u)$. Note that $t > \xi d(y, v)$ by the\nchoice of $\xi$. So setting $w = y$ we have $t > \xi d(v, w)$ and
\[ d(y, w) = 0 \leq d(y, v) - (r - \epsilon)t \leq d(y, v) - (r - \epsilon)d((x, v), (u, w)). \]

Since $\epsilon$ can be chosen arbitrarily small, the result follows. \hfill \Box

**Theorem 3.2** (second criterion for Milyutin regularity). Let $X$ be a complete metric\nspace, $U \subset X$ and $V \subset Y$ open sets and $F : X \rightrightarrows Y$ a set-valued mapping with\nclosed graph. Then $F$ is Milyutin regular on $U \times V$ with $\text{sur}_n F(U|V) \geq r$ if and only if for any\nx $\in U$ and any $y \in V$ with $0 < \bar{\psi}_y(x) < rm(x)$ there is a $u \neq x$ such that\nde\n\[ \bar{\psi}_y(u) \leq \bar{\psi}_y(x) - rd(x, u). \]

**Proof.** The proof of sufficiency is similar to the proof of the first part of the previous\ntheorem.

To prove that (3.3) is necessary for Milyutin regularity\ntake $x \in U$, $y \in V$ such that\n$0 < d(y, F(x)) < rm(x)$. Take $\rho < r$ such that still $d(y, F(x)) < r\rho m(x)$, and let $\rho < \rho' < r$.
Let $x_n \to x$ be such that $d(y, F(x_n)) < \bar{\psi}_y(x)$. We may assume that $d(y, F(x_n)) < r\rho m(x)$\nfor all $n$. Choose positive $\delta_n \to 0$ such that $d(y, F(x_n)) \leq (1 + \delta_n)\bar{\psi}_y(x)$, and let\nt be defined by $\rho't_n = (1 + \delta_n)\bar{\psi}_y(x)$. Then $y \in B(F(x_n), \rho't_n)$, $t_n < m(x_n)$ (at least for\nlarge $n$) and due to the regularity assumption on $F$ for any $n$ we can find a $u_n$ such that\nde $d(u_n, x_n) < t_n$ and $y \in F(u_n)$. Note that $u_n$ are bounded away from $x$ for otherwise (as
Graph $F$ is closed) we would inevitably conclude that $y \in F(x)$ which cannot happen as $\overline{\psi}_y(x) > 0$. This means that $\lambda_n = d(u_n, x_n)/d(u_n, x)$ converge to one. Thus

$$\overline{\psi}_y(u_n) = 0 = \overline{\psi}_y(x) - \overline{\psi}_y(x) = \overline{\psi}_y(x) - \frac{\rho \epsilon_n}{1 + \delta_n}$$

$$\leq \overline{\psi}_y(x) - \frac{\rho'}{1 + \delta_n}d(u_n, x_n)$$

$$= \overline{\psi}_y(x) - \frac{\lambda_n \rho'}{1 + \delta_n}d(u_n, x) \leq \overline{\psi}_y(x) - \rho d(u_n, x),$$

the last inequality being eventually true as $\lambda_n \rho' > \rho(1 + \delta_n)$ for large $n$.

The theorem is especially convenient when $\psi_y$ is lower semicontinuous for every $y \in V$. Otherwise, the need for preliminary calculation of $\overline{\psi}_y$, the lower closure of $\psi_y$, may cause difficulties. It is possible however to modify the condition of the theorem and get a statement that requires verification of (3.3)-like inequality for $\psi$ rather than $\overline{\psi}$, although at the expense of some additional uniformity assumption.

**Theorem 3.3** (modified second criterion for Milyutin regularity). Let $X, Y, F, U$ and $V$ be as in Theorem 3.2. A necessary and sufficient condition for $F$ to be Milyutin regular on $U \times V$ with $\text{sur} F(U|V) \geq r$ is that there is a $\lambda \in (0, 1)$ and for any $x \in U$ and $y \in V$ with $0 < \psi_y(x) < \rho m(x)$ there is a $u \neq x$ such that

$$\psi_y(u) \leq \psi_y(x) - r \rho d(x, u), \quad \lambda \psi_y(u) \leq \lambda \psi_y(x). \tag{3.4}$$

**Proof.** The key for understanding the theorem is the following implication

$$\overline{\psi}_y(x) = 0 \Rightarrow y \in F(x) \tag{3.5}$$

of course valid, under the condition of the theorem for $x \in U, y \in V$. Indeed, $\overline{\psi}_y(x) = 0$ means that there is a sequence $(x_n)$ converging to $x$ such that $\psi_y(x_n) \rightarrow 0$. This in turn implies the existence of $x_n \in F(x_n)$ converging to $y$. As the graph of $F$ is closed, it follows that $(x, y) \in \text{Graph } F$ as claimed.

Now we can verify that under the assumptions of the theorem, the condition of Theorem 3.2 holds. So let $x \in U, y \in V$ and $0 < \alpha = \overline{\psi}_y(x)$. Take $x_n \rightarrow x$ such that $\psi_y(x_n) = \alpha_n \rightarrow \alpha$ and for each $n$ a $u_n$ such that $\psi_y(u_n) \leq \lambda \alpha_n$ and $\psi_y(u_n) \leq \psi_y(x_n) - r \rho d(x_n, u_n)$. An easy calculation shows that

$$\psi_y(u_n) \leq \overline{\psi}_y(x) - r \rho d(x, u_n) + \epsilon_n,$$

where $\epsilon_n \rightarrow 0$. As $d(x, u_n)$ are bounded away from zero by a positive constant, we have $\epsilon_n = \delta_n d(x, u_n)$, where $\delta_n \rightarrow 0$. Combining this with the above inequality, we conclude that for any $r' < r$ that $u_n \neq x$ and inequality

$$\overline{\psi}_y(u_n) \leq \overline{\psi}_y(x) - r' \rho d(x, u_n)$$

holds for sufficiently large $n$. This allows to apply Theorem 3.2 and conclude (by virtue of (3.5)) that there is a $w \in B(x, (r')^{-1})$ such that $y \in F(x)$, that is $\text{sur} \rho F(U|V) \geq r'$.
Theorem 3.4 (criterion for local regularity). Let \( F : X \to Y \) be a set-valued mapping with closed graph, and let \((x, y) \in \text{Graph } F\). Then \( F \) is regular near \((x, y)\) if and only if there are \( \varepsilon > 0 \), \( \xi > 0 \) and \( r > 0 \) such that for any \( x, v \) and \( y \) satisfying \( d(x, x) < \varepsilon \), \( d(y, y) < \varepsilon \), \( v \in F(x) \) and \( 0 < d(y, v) < \varepsilon \) either of the following two properties is valid:

(a) Graph \( F \) is locally complete and there is a pair \((u, w) \in \text{Graph } F\), \( (u, w) \neq (x, y) \) such that (3.7) holds.

(b) \( X \) is a complete metric space, the graph of \( F \) is closed and either (3.3) or (3.4) holds true.

Moreover, in either case \( \text{sur} F(\overline{\mathbf{z}}|\mathbf{y}) \geq r \).

Theorem 3.5 (density theorem [45, 92]). Let \( U \subset X \) and \( V \subset Y \) be open sets, let \( F : X \to Y \) be a set-valued mapping with complete graph. We assume that whenever \( x \in U \), \( v \in F(x) \) and \( t < m(x) \), the set \( F(B(x, t)) \) is a \( \ell \)-net in \( B(v, rt) \cap V \), where \( 0 \leq \ell < r \). Then \( F \) is Milyutin regular on \( U \times V \) and \( \text{sur}_m F \geq r - \ell \). In particular, if \( F(B(x, t)) \) is dense in \( B(F(x), rt) \cap V \) for \( x \in U \) and \( t < m(x) \), then \( \text{sur}_m F(U|V) \geq r \).

**Proof.** Take \( x \in U \) and suppose \( y \in V \) is such that \( d(y, F(x)) < rm(x) \). Take a \( v \in F(x) \) such that \( d(y, v) < rm(x) \) and set \( t = d(y, v)/r \). Then \( t < m(x) \) and by the assumption we can choose \((u, w) \in \text{Graph } F\) such that \( d(x, u) \leq t \) and \( d(y, w) \leq \ell t = (\ell/r)d(y, v) \). Then

\[
d(v, w) \leq d(y, v) + d(y, w) \leq (1 + \frac{\ell}{r})d(y, v) \leq 2d(y, v).
\]

Take a \( \xi > 0 \) such that \( \xi r \leq 1/2 \). Then \( \xi d(v, w) < 2\xi rt \leq t \) and therefore

\[
d(y, w) \leq \ell t = rt - (r - \ell)t = d(y, v) - (r - \ell)t \leq d(y, v) - (r - \ell)d\xi((x, v), d(u, w)).
\]
Exercise 3.6. Prove the theorem under the assumptions of Theorem 3.2 rather than Theorem 3.1.

Exercise 3.7. Prove Banach-Shauder open mapping theorem using the density theorem (and the Baire category theorem).

The specification of Theorem 3.5 for local regularity at \((\bar{x}, \bar{y})\) is

Corollary 3.8 (density theorem - local version). Suppose there are \(r > 0\), and \(\varepsilon > 0\) such that \(F(B(x,t))\) is an \(\ell t\)-net in \(B(v,rt)\) whenever \(d(x,\bar{x}) < \varepsilon\), \(d(v,\bar{y}) < \varepsilon\), \(v \in F(x)\) and \(t < \varepsilon\). Then \(\text{sur} F(y|x) \geq r - \ell\). Thus if \(B(v,rt) \subset \text{cl} F(B(x,t))\) for all \(x\) and \(v\) and \(t\) satisfying the specified above conditions, then \(B(v,rt) \subset F(B(x,t))\) for the same set of the variables.

The density phenomenon was extensively discussed, especially at the early stage of developments. Results in the spirit of Corollary 3.8 were first considered in Ptak [146], Tziskaridze [166] and Dolecki [46, 47] in mid-1970s. The very idea (and to a large extent the techniques used) could be traced back to Banach’s proof of the closed graph/open mapping theorem. Some of the subsequent studies (e.g. [26, 169]) were primarily concentrated on results of such type. We refer to [16] for detailed discussions and many references. Dmitruk-Milyutin-Osmolovskii in [45] made a substantial step forward when they replaced (in the global context) the density requirement by the assumption that \(F(B(x,t))\) is an \(\ell t\)-net in \(B(F(x),rt)\). This opened way to proving the Milyutin perturbation theorem (see the next section). A similar advance in the framework of the infinitesimal approach (for mappings between Banach spaces) was made by Aubin [5].

3.3 Infinitesimal criteria.

The main tool of the infinitesimal regularity theory in metric spaces is provided by the concept of (strong) slope – which is just the maximal speed of descent of the function from a given point – introduced in 1980 by DeGiorgi-Marino-Tosques [44] and since then widely used in various chapters of metric analysis.

Definition 3.9 (slope). Let \(f\) be an extended-real-valued function on \(X\) which is finite at \(x\). The quantity

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

is called the (strong) slope of \(f\) at \(x\). We also agree to set \(|\nabla f|(x) = \infty\) if \(f(x) = \infty\). The function is called calm at \(x\) if \(|\nabla f|(x) < \infty\).

We shall consider only local regularity in this subsection (although it is possible to give slope-based characterizations of Milyutin regularity as well). It is easy to observe that \(|\nabla f|(x) > r\) means that arbitrarily close to \(x\) there are \(u \neq x\) such that \(f(x) > f(u) + rd(x,u)\). This allows to reformulate the sufficient part of the regularity criteria of Theorem 3.4 in infinitesimal terms. To this end set as before

\[
\varphi_y(x,v) = d(y,v) + i_{\text{Graph} F(x,v)}, \quad \psi_y(x) = d(y,F(x)), \quad \underline{\psi}_y(x) = \liminf_{u \to x} \psi_y(u),
\]
and let $\nabla_\xi$ stand for the slope of functions on $X \times Y$ with respect to the $d_\xi$-metric: 
$$d_\xi((x, v), (x', v')) = \max\{d(x, x'), \xi d(v, v')\}.$$

Things are more complicated with the necessity part: to prove it, an additional assumption on the target space is needed. Namely, let us say that a metric space $X$ is \textit{locally coherent} if for any $x$
$$\lim_{u, w \to x, u \neq w} |\nabla d(u, \cdot)|(w) = 1.$$ 

It can be shown that a convex set and a smooth manifold in a Banach space are locally coherent in the induced metric [89] and that any length metric space (space whose metric is defined by minimal lengths of curves connecting points) is locally coherent (as follows from [14]).

\textbf{Theorem 3.10} (local regularity criterion 1 [89]). Let $X$ and $Y$ be metric spaces, let $F : X \Rightarrow Y$ be a set-valued mapping, and let $(\overline{x}, \overline{y}) \in \text{Graph} F$. We assume that $\text{Graph} F$ is locally complete at $(\overline{x}, \overline{y})$. Suppose further that there are $\varepsilon > 0$, and $r > 0$ such that for some $\xi > 0$
$$|\nabla_\xi \varphi_y|(x, v) > r$$
if
$$v \in F(x), \ d(x, \overline{x}) < \varepsilon, \ d(y, \overline{y}) < \varepsilon, \ d(v, \overline{y}) < \varepsilon, \ v \neq y. \tag{3.6}$$
Then $F$ is regular near $(\overline{x}, \overline{y})$ with $\text{sur} F(\overline{x}, \overline{y}) \geq r$.

Conversely, let $Y$ be locally coherent at $\overline{y}$. Assume that $\text{sur} F(\overline{x}, \overline{y}) > r > 0$. Take a $\xi < r^{-1}$. Then for any $\delta > 0$ there is an $\varepsilon > 0$ such that $|\nabla_\xi \varphi_y|(x, v) \geq (1 - \delta) r$ whenever $(x, y, v)$ satisfy (3.7). Thus, in this case
$$\text{sur} F(\overline{x}, \overline{y}) = \lim_{\varepsilon \to 0} \inf_{y \to \overline{y}, y \neq v} \inf_{(x, v) \in \text{Graph} F} |\nabla_\xi \varphi_y|(x, v). \tag{3.8}$$

For mappings into metrically convex spaces (for any two points there is a shortest path connecting the points) the final statement of Theorem 3.10 can be slightly improved.

\textbf{Corollary 3.11}. Suppose under the conditions of Theorem 3.10 that $Y$ is metrically convex. Then for any neighborhood $V$ of $\overline{y}$
$$\text{sur} F(\overline{x}, \overline{y}) = \lim_{\varepsilon \to 0} \inf_{(x, v) \in \text{Graph} F} \inf_{y \in V \setminus \{v\}} |\nabla_\xi \varphi_y|(x, v) \tag{3.9}$$

\textbf{Theorem 3.12} (local regularity criterion 2). Suppose that $X$ is complete and the graph of $F$ is closed. Assume further that there are neighborhood $U \subset X$ of $\overline{x}$ and $V \subset Y$ of $\overline{y}$, $r > 0$ and $\varepsilon > 0$ such that that $|\nabla_\xi \varphi_y|(x) > r$ for all $(x, y) \in U \times V$ such that $\varepsilon > \varphi_y(x) > 0$. Then $\text{sur} F(\overline{x}, \overline{y}) \geq r$.

Conversely, if in addition $Y$ is a length space and $\text{sur} F(\overline{x}, \overline{y}) > r > 0$, then there is a neighborhood of $(\overline{x}, \overline{y})$ and an $\varepsilon > 0$ such that $|\nabla_\xi \varphi_y|(x) \geq r$ for all $(x, y)$ of the neighborhood such that $y \notin F(x)$ and $0 < \varphi_y(x) < \varepsilon r$. Thus in this case
$$\text{sur} F(\overline{x}, \overline{y}) = \lim_{\varepsilon \to 0} \inf_{(x, y) \to (\overline{x}, \overline{y})} \inf_{y \neq d(y, F(x)) = 0} |\nabla_\xi \varphi_y|(x).$$
In particular, if \( \psi_y = d(y, F(\cdot)) \) is lower semicontinuous at every \( x \) of a neighborhood of \( \overline{x} \) and for every \( y \not\in F(x) \) close to \( \overline{y} \), then

\[
\text{sur} F(\overline{x}|\overline{y}) = \liminf_{(x,y) \to (\overline{x},\overline{y}) \atop 0 \not\in d(y, F(x)) \to 0} |\nabla \psi_y|(x).
\]

The starting point for developing slope-based regularity theory was the paper by Azé-Corvellec-Lucchetti [15] (its first version was circulated in 1998) who obtained a global error bound in terms of "variational pairs" that include slope on a metric space as a particular case. Theorem 3.10 and specifically the fact that the slope estimate is precise, was proved in [55] for a detailed study of the concepts mainly for mappings between finite dimensional spaces, and begin with parallel concepts relating to linear openness which are rather new in many respects. Subregularity and calmness attract much attention last years. We refer to [10] for a systematic exposition of the slope-based approach to local regularity. Theorem 3.12 is a slightly modified version of the mentioned result of Ngai-Tron-Thera [136] (proved originally for \( Y \) being a Banach space).

To explain how the additional assumption on \( Y \) is used to get necessity e.g. in Theorem 3.10 let us consider, following the original argument in [55], \( (x, y, v) \) sufficiently close to \( \overline{x} \) and \( \overline{y} \) respectively and such that \( y \neq v \in F(x) \). For any \( n \) take \( \delta_n = o(n^{-1}) \) and a \( v_n \) such that \( d(v_n, v) \leq (n^{-1} + \delta_n)d(y, v) \) and \( d(v_n, y) \leq (1 - n^{-1} + \delta_n)d(y, v) \). If \( Y \) is a length space such \( v_n \) can be found. As \( F \) covers near \( (\bar{x}, \bar{y}) \) with modulus greater than \( r \), there is a \( u_n \) such that \( v_n \in F(u_n) \) and \( d(u_n, x) \leq r^{-1}d(v_n, v) \to 0 \) when \( n \to \infty \). We have \( |d(y, v) - (d(y, v_n) + d(v, v_n))| = o(d(v_n, v)) \). Therefore (as \( r\xi < 1 \))

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

Similar argument, modified as the definition of \( \overline{\psi}_y \) includes a limit operation, can be used also for the proof of necessity in Theorem 3.12.

It should be observed that the class of locally coherent spaces is strictly bigger than the class of length spaces. For instance a smooth manifold in a Banach space with the induced metric is a locally coherent space but not a length space (unless it is a linear manifold).

### 3.4 Related concepts: metric subregularity, calmness, controllability, linear recession

In the definitions of the local versions of the three main regularity properties we scan entire neighborhoods of the reference point of the graph of the mapping. Fixing one or both components of the point leads to new weaker concepts that differ from regularity in many respects. Subregularity and calmness attract much attention last years. We refer to [55] for a detailed study of the concepts mainly for mappings between finite dimensional spaces, and begin with parallel concepts relating to linear openness which are rather new in the context of variational analysis. We skip (really elementary) proofs of almost all results in this subsection.

**Definition 3.13** (controllability). A set valued mapping \( F : X \rightrightarrows Y \) is said to be (locally) controllable at \( (\bar{x}, \bar{y}) \) if there are \( \varepsilon > 0, \gamma > 0 \) such that

\[
B(\bar{y}, rt) \subset F(B(\bar{x}, t)), \quad \text{if } 0 \leq t < \varepsilon.
\]
The upper bound of such \( r \) is the rate or modulus of controllability of \( F \) at \((\bar{x}, \bar{y})\). We shall denote it \( \text{contr} F(\bar{x}|\bar{y}) \) and \( \text{contr} F(\bar{x}) \) if \( F \) is single-valued.

**Proposition 3.14** (Regularity vs. controllability). Let \( X \) and \( Y \) be metric spaces, let \( F : X \rightrightarrows Y \) have locally complete graph, and let \((\bar{x}, \bar{y}) \in \text{Graph} \ F \). Then

\[
\text{sur} F(\bar{x}|\bar{y}) = \liminf_{\varepsilon \to 0} \{ \text{contr} F(x|y) : (x, y) \in \text{Graph} \ F, \max\{d(x, \bar{x}), d(y, \bar{y})\} < \varepsilon \}. \tag{3.11}
\]

**Definition 3.15** (linear recession). Let us say that \( F \) recedes from \( \bar{y} \) at \((\bar{x}, \bar{y})\) at a linear rate if there are \( \varepsilon > 0 \) and \( K \geq 0 \) such that

\[
d(\bar{y}, F(x)) \leq Kd(x, \bar{x}), \quad \text{if } d(x, \bar{x}) < \varepsilon. \tag{3.12}
\]

We shall call the lower bound of such \( K \) the speed of recession of \( F \) from \( \bar{y} \) at \((\bar{x}, \bar{y})\) and denote it \( \text{ress} F(\bar{x}|\bar{y}) \).

The other possible way to “pointify” the Aubin property is to fix \( \bar{x} \) and allow \((x, y)\) to change within \text{Graph} \( F \). Then, instead of (3.12) we get the inequality

\[
d(y, F(\bar{x})) \leq Kd(x, \bar{x}). \tag{3.13}
\]

**Definition 3.16** (calmness). It is said that \( F : X \rightrightarrows Y \) is calm at \((\bar{x}, \bar{y})\) if there are \( \varepsilon > 0, K \geq 0 \) such that \( d(x, \bar{x}) < \varepsilon, d(y, \bar{y}) < \varepsilon \) and \( y \in F(x) \). The lower bound of all such \( K \) will be called the modulus of calmness of \( F \) at \((\bar{x}, \bar{y})\). We shall denote it by \( \text{calm} F(\bar{x}|\bar{y}) \) (\( \text{calm} F(\bar{x}) \) if \( F \) is single-valued).

Again we can easily see that uniform calmness, that is calmness at every \((x, y)\) of the intersection of \text{Graph} \( F \) with a neighborhood of \((\bar{x}, \bar{y})\) with the same \( \varepsilon \) and \( K \) for all such \((x, y)\), is equivalent to the Aubin property of \( F \) near \((\bar{x}, \bar{y})\).

**Definition 3.17** (subregularity). Let \( F : X \rightrightarrows Y \) and \( \bar{y} \in F(\bar{x}) \). It is said that \( F \) is (metrically) subregular at \((\bar{x}, \bar{y})\) if there is a \( K > 0 \) such that

\[
d(x, F^{-1}(\bar{y})) \leq Kd(\bar{y}, F(x)) \quad \text{if } d(x, \bar{x}) < \varepsilon. \tag{3.14}
\]

for all \( x \) of a neighborhood of \( \bar{x} \). The lower bound of such \( K \) is called the rate or modulus of subregularity of \( F \) at \((\bar{x}, \bar{y})\). It will be denoted \( \text{subreg} F(\bar{x}|\bar{y}) \).

We say that \( F \) is strongly subregular at \((\bar{x}, \bar{y})\) if it is subregular at the point and \( \bar{y} \notin F(x) \) for \( x \neq \bar{x} \) of a neighborhood of \( \bar{x} \).

**Proposition 3.18.** The equalities

\[
\text{subreg} F(\bar{x}|\bar{y}) = \text{calm} F^{-1}(\bar{y}|\bar{x}), \quad \text{contr} F(\bar{x}|\bar{y}) \cdot \text{ress} F^{-1}(\bar{y}|\bar{x}) = 1
\]

always hold. If moreover, \( F \) is strongly subregular at \((\bar{x}, \bar{y})\), then

\[
\text{contr} F(\bar{x}|\bar{y}) \cdot \text{subreg} F(\bar{x}|\bar{y}) \geq 1.
\]
Theorem 3.19 (slope criterion for calmness). Let $X$ and $Y$ be arbitrary metric spaces, let $F : X \rightrightarrows Y$ be a set-valued mapping with closed graph and let $(\bar{x}, \bar{y}) \in \text{Graph } F$. Then

$$\text{calm} F(\bar{x} | \bar{y}) \geq \limsup_{y \to \bar{y}} |\nabla \psi_y(\bar{x})|,$$

where, as earlier, $\psi_y(x) = d(y, F(x))$.

Proof. Let $K > \text{calm} F(\bar{x} | \bar{y})$ then there is an $\varepsilon > 0$ such that (3.13) holds, provided $d(\bar{x}, x) < \varepsilon$ and $y \in F(x)$. To prove the theorem, it is sufficient to show that $|\nabla \psi_y(x)| \leq K$ for all $y$ sufficiently close to $\bar{y}$. To this end, it is sufficient to verify that there is a $\delta > 0$ such that the inequality

$$d(y, F(\bar{x})) - d(y, F(x)) \leq K d(x, \bar{x})$$

holds for all $x, y$ satisfying $d(\bar{x}, x) < \delta$, $d(y, \bar{y}) < \delta$.

If $y \in F(x)$, then (3.14) reduces to (3.13). Take a positive $\delta < \varepsilon/2$, and let $x$ and $y$ be such that $d(\bar{x}, x) < \delta$, $d(y, \bar{y}) < \delta$. If $d(y, F(x)) \geq \delta$, then (3.14) obviously holds. If $d(y, F(x)) < \delta$, we can choose a $v \in F(x)$ such that $d(y, v) < \delta$. Then $d(v, \bar{y}) < \varepsilon$ and therefore $d(v, F(\bar{x})) \leq d(x, \bar{x})$. Thus

$$d(y, F(\bar{x})) - d(y, F(x)) \leq d(y, v) + d(v, F(\bar{x})) - d(y, F(x))$$

$$\leq K d(x, \bar{x}) + d(y, v) - d(y, F(x))$$

and (3.14) follows as $d(y, v)$ can be arbitrarily close to $d(y, F(x))$. \qed

Theorem 3.20 (slope criterion for subregularity). Assume that $X$ is a complete metric space. Let $F : X \rightrightarrows Y$ be a closed set-valued mapping and $(\bar{x}, \bar{y}) \in \text{Graph } F$. Assume that the function $\psi(x) = d(\bar{y}, F(x))$ is lower semicontinuous and there are $\varepsilon > 0$ and $r > 0$ such that $|\nabla \psi_{\bar{y}}(x) = |\nabla d(\bar{y}, F(\cdot))(x)| \geq r$, if $d(x, \bar{x}) < \varepsilon$ and $0 < d(\bar{y}, F(\bar{x})) < \varepsilon$. Then $F$ is subregular at $(\bar{x}, \bar{y})$ with modulus of subregularity (and hence the modulus of calmness of $F^{-1}$ at $(\bar{y}, \bar{x})$) not greater than $r^{-1}$.

4 Metric theory. Perturbations and stability.

In this section we concentrate on two fundamental questions:

(a) what happens with regularity (and subregularity) properties of $F$ if the mapping is slightly perturbed?

(b) how the set of solutions of the inclusion $y \in F(x, p)$ (where $F$ depends on a parameter $p$) depends on $(y, p)$?

The answer to the second question leads us to a fairly general implicit function theorems. The key point in both cases is that we have to require a certain amount of Lipschitzness of perturbations to get desirable results.
4.1 Stability under Lipschitz perturbation

**Theorem 4.1** (stability under Lipschitz perturbation). Let $X$, $Y$ be metric spaces, let $U \subset X$ and $V \subset Y$ be open sets. Consider a set-valued mapping $\Psi : X \times X \rightrightarrows Y$ with closed graph assuming that either $X$ or the graph of $\Psi$ is complete. Let $F(x) = \Psi(x,x)$. Suppose that

(a) for any $u \in U$ the mapping $\Psi(\cdot,u)$ is Milyutin regular on $(U,V)$ with modulus of surjection greater than $r$, that is for any $x \in U$, any $v \in \Psi(x,u)$ and any $y \in B(v,rt) \cap V$ with $d(x,X\setminus U) < (r-\ell)d(y,v)$ and $y \in F(x')$;

(b) for any $x \in U$ the mapping $\Psi(x,\cdot)$ is pseudo-Lipschitz on $(U,V)$ with modulus $\ell < r$, that is for any $u,v \in U$

$$d(x,x') < \ell d(u,v).$$

Then $F(x) = \Psi(x,x)$ is Milyutin regular on $(U,V)$ with $\text{sur}_{m} F(U|V) \geq r - \ell$.

**Proof.** We shall consider only the case of complete Graph $\Psi$. According to the general regularity criterion of Theorem 3.1 all we have to show is that there is a $\xi > 0$ such that, given $(x,v) \in gr F$ and $y$ such that $x \in U$, $y \in V$ and $0 < d(y,v) < rm(x)$, there is another point $(x',v') \neq (x,v)$ in the graph of $F$ such that

$$d(y,v') \leq d(y,v) - (r-\ell) \max\{d(x,x'),\xi(v,v')\}.$$

We have by (a): $B(v,rt) \cap V \subset \Psi(B(x,t),x)$ if $t < m(x)$. As $d(y,v) < rm(x)$, it follows that there is a $x' \in B(x,t)$ such that $y \in \Psi(x',x)$ and $d(x,x') \leq r^{-1}d(y,v)$.

Clearly, $x' \in U$. Therefore by (b) $d(y,\Psi(x',x')) < \ell d(x,x')$. This means that there is a $v' \in F(x')$ such that

$$d(y,v') \leq \ell d(x,x') \leq \frac{\ell}{r}d(y,v).$$

Take $\xi < (r+\ell)^{-1}$. Then

$$\xi d(v,v') \leq (r+\ell)^{-1}(d(v,y) + d(y,v')) \leq (r+\ell)^{-1}\left(1 + \frac{\ell}{r}\right)d(y,v) = \frac{1}{r}d(y,v).$$

Thus $\max\{d(x,x'),\xi d(v,v')\} \leq r^{-1}d(y,v)$ and we have

$$d(y,v') < (\ell/r)d(y,v) = d(y,v) - \frac{r-\ell}{r}d(y,v) \leq d(y,v) - (r-\ell) \max\{d(x,x'),\xi d(v,v')\}.$$

as needed. \hfill $\square$

**Corollary 4.2** (Milyutin’s perturbation theorem [45]). Let $X$ be a metric space, let $Y$ be a normed space and $F : X \rightrightarrows Y$ and $G : X \rightrightarrows Y$. We assume that either the graphs of $F$ and $G$ are complete or $X$ is a complete space. Let further $U \subset X$ be an open set such that $F$ is Milyutin regular on $U$ with $\text{sur} F(U) \geq r$ and $G$ is (Hausdorff) Lipschitz with $\text{lip} G(U) \leq \ell < r$. If either $F$ or $G$ is single-valued continuous on $U$, then $F + G$ is Milyutin regular on $U$ and $\text{sur} (F + G)(U) \geq r - \ell$.

**Proof.** Apply the theorem to $\Psi(x,u) = F(x) + G(u)$.

To state a local version of the theorem, we need the following
Definition 4.3 (uniform regularity). Let $P$ be a topological space, let $F : P \times X \rightrightarrows Y$, let $\bar{p} \in P$, and let $(\bar{x}, \bar{y}) \in \text{Graph } F(\bar{p}, \cdot)$. We shall say that $F$ is regular near $(\bar{x}, \bar{y})$ uniformly in $p \in P$ near $\bar{p}$ if for any $r < \text{sur} F(\bar{p}, \cdot)(\bar{x}, \bar{y})$ there are $\varepsilon > 0$ and a neighborhood $W \subset P$ of $\bar{p}$ such that for any $p \in W$ and any $x$ with $d(x, \bar{x}) < \varepsilon$

$$B(F(p, x), rt) \cap B(\bar{y}, \varepsilon) \subset F(p, B(x, t)), \quad \text{if } 0 \leq t < \varepsilon.$$ 

Theorem 4.4 (stability under Lipschitz perturbations: local version). Let $X$, $Y$, $\Psi : X \times X \rightrightarrows Y$ and $F(x) = \Psi(x, x)$ be as in Theorem 4.1, and let $(\bar{x}, \bar{y}) \in \text{Graph } F$. We assume that

(a) $\Psi(\cdot, u)$ is regular near $(\bar{x}, \bar{y})$ uniformly in $u$ near $\bar{x}$;

(b) $\Psi(x, \cdot)$ is pseudo-Lipschitz near $(\bar{x}, \bar{y})$ uniformly in $x$ near $\bar{x}$.

If $\text{lip}(\Psi(\cdot, \cdot))(\bar{y}) < \ell < r < \text{sur}(\Psi(\cdot, \cdot))(\bar{y})$, then $F$ is regular near $(\bar{x}, \bar{y})$ with modulus of surjection greater than $r - \ell$.

The last theorem in turn immediately implies Milyutin’s theorem and its versions correspond to $\Psi(x, y) = F(x) + g(y)$ with $g$ being single-valued Lipschitz. The following corollary from the theorems is straightforward.

Theorem 4.5 (Milyutin’s perturbation theorem - local version). Let $X$ be a metric space, let $Y$ be a normed space, and let $F : X \rightrightarrows Y$ and $G : X \rightrightarrows Y$. Given $\bar{x} \in \text{dom } F \cap \text{dom } G$, $\bar{y} \in F(\bar{x})$, $\bar{z} \in G(\bar{x})$, we assume that $F$ is regular near $(\bar{x}, \bar{y})$ with $\text{sur} F(\bar{x}, \bar{y}) \geq r$ and $G$ has the Aubin property near $(\bar{x}, \bar{z})$ with $\text{lip}(G(\bar{x}, \bar{z})) \leq \ell$. If either $F$ or $G$ is single-valued continuous on its domain and the graph of the other is complete in the product metric, then

$$\text{sur}(F + G)(\bar{x}, \bar{y} + \bar{z}) \geq r - \ell.$$ 

Proof. Set $\Psi(x, y) = F(x) + G(y)$. It is an easy matter to check that the conditions of Theorem 4.4 are valid. □

As an immediate consequence of the last theorem we mention a stronger version of the Lyusternik-Graves theorem stating that its condition is not only sufficient but also necessary for regularity is an immediate corollary of the last theorem.

Corollary 4.6 (Lyusternik-Graves from Milyutin). Let $X$ and $Y$ be Banach spaces, and let $F : X \to Y$ be strictly differentiable at $\bar{x}$. Then $\text{sur} F(\bar{x}) = C(F'(\bar{x}))$.

Proof. Indeed, let $X, Y$ be Banach spaces, and let $F : X \to Y$ be strictly differentiable at $\bar{x}$. Set $g(x) = F(x) - F'(\bar{x})(x - \bar{x})$. As $F$ is strictly differentiable at $\bar{x}$, the Lipschitz constant of $g$ at $\bar{x}$ is zero which by Milyutin’s theorem means that the moduli of surjection of $F$ at $\bar{x}$ and $F'(\bar{x})$ coincide. □

We observe next that in Theorem 4.3 one of the mappings is assumed single-valued. This assumption is essential. With both mappings set-valued the result may be wrong as the following example shows.
Example 4.7 (cf. [55]). Let $X = Y = \mathbb{R}$, $G(x, y) = \{x^2, -1\}$, $F(x) = \{-2x, 1\}$. It is easy to see that $F$ is regular near $(0, 0)$ and $G$ is Lipschitz in the Hausdorff metric. On the other hand, 

$$\Phi(x) = \{x^2 - 2x, x^2 + 1, -2x - 1, 0\}$$

is not even regular at $(0, 0)$. Indeed $(\xi, 0) \in \text{Graph } \Phi$ for any $\xi$. However, if $\xi \neq 0$, then the $\Phi$-image of a sufficiently small neighborhood of $\xi$ does not contain points of a small neighborhood of zero other than zero itself.

Perturbation analysis of regularity properties was initiated by Dmitruk-Milyutin-Osmolovski in [45] with a proof of a global version of Theorem 4.2 (attributed in [45] to Milyutin) with both the mapping and the perturbation single valued. The first perturbation result for set-valued mappings was proved probably by Ursescu [168] (see also [88]). Observe that global theorems are valid for Lipschitz set-valued perturbations as well.

Till very recently the main attention was devoted to additive perturbations into a linear range space, especially in connection with implicit function theorems for generalized equations - see e.g. [11, 55]. Interest to non-additive Lipschitz set-valued perturbations of set-valued mappings appeared just a few years ago, partly in connection with fixed point and coincidence theorems [4, 51, 92, 97].

The Graves theorem can be viewed as a perturbation theorem for a linear regular operator. For that reason in some publications (e.g. [50, 55]) this theorem is called "extended Lyusternik-Graves theorem". I believe the name "Milyutin theorem" is adequate. It is quite obvious that Graves did not have in mind the perturbation issue and was interested only in a quality of approximation needed to get the result. (Tikhomirov and I had similar idea when proving the metric regularity counterpart of the Graves theorem for [103] without any knowledge of the Graves’ paper.) And the fact that the Lipschitz property of the perturbation as the key for the estimate was explicitly emphasized in [45]. Note also that even Corollary 4.6 cannot be obtained from the Graves theorem.

Milyutin’s theorem can also be viewed as a regularity result for a composition $\Phi(x, F(x))$, where $\Phi(x, y) = G(x) + y$. Theorems 4.1 and 4.4 can be applied to prove regularity of more general compositions, with arbitrary $\Phi$, just by taking $\Psi(x, u) = \Phi(x, F(u))$. However, a certain caution is needed to guarantee that such a $\Psi$ satisfies the required assumptions (as say in [92] where $\Phi(x, \cdot)$ is assumed to be an isometry or in [61] where a certain "composition stability" is a priori assumed). Corollary 4.6 was first stated in [49] with a direct proof, not using Milyutin’s theorem.

4.2 Strong regularity and metric implicit function theorem.

Generally speaking, the essence of the inverse function theorem is already captured by the main Equivalence Theorem 2.7. But in view of the very special role of the inverse and implicit function theorems in the classical theory, it seems appropriate to make the connection with the classical results more transparent.

So let $F(x, p) : X \times P \rightarrow Y$. We shall view $P$ as a parameter space. Let $S(y, p) = \{x \in X : y \in F(x, p)\}$ stand for the solution mapping of the inclusion $y \in F(x, p)$. In all theorems to follow we consider $Y \times P$ with an $\ell^1$-type distance

$$d_{\alpha}^1((y, p), (y', p')) = \alpha d(y, y') + d(p, p'),$$

where $\alpha$ is a positive constant. This setting is motivated by the desire to retain the classical results as special cases of our general theorems. The $\ell^1$-type distance is also convenient for applications to variational inequalities and optimization problems.

The following theorems are proved in [40] using a slight modification of the proof of [168]. The main idea is to use a version of the implicit function theorem for set-valued mappings due to Ursescu [168] and a version of the metric regularity counterpart of the Graves theorem due to Tikhomirov [103]. The proofs are based on a careful analysis of the behavior of the solution mapping $S(y, p)$ near the origin.

Theorem 4.8. Let $F(x, p) : X \times P \rightarrow Y$ be a set-valued mapping. Assume that $F(x, p)$ is weakly lower semicontinuous, weakly continuous, and weakly upper semicontinuous with respect to $x$ for each $p$ fixed. Suppose that $F(x, p)$ is weakly $\alpha$-Lipschitz in the Hausdorff metric for each $p$ fixed. Then $S(y, p)$ is a continuous mapping from $Y \times P$ to $X$.

Theorem 4.9. Let $F(x, p) : X \times P \rightarrow Y$ be a set-valued mapping. Assume that $F(x, p)$ is weakly lower semicontinuous, weakly continuous, and weakly upper semicontinuous with respect to $x$ for each $p$ fixed. Suppose that $F(x, p)$ is weakly $\alpha$-Lipschitz in the Hausdorff metric for each $p$ fixed. Then $S(y, p)$ is a continuous mapping from $Y \times P$ to $X$.

Corollary 4.10. Let $F(x, p) : X \times P \rightarrow Y$ be a set-valued mapping. Assume that $F(x, p)$ is weakly lower semicontinuous, weakly continuous, and weakly upper semicontinuous with respect to $x$ for each $p$ fixed. Suppose that $F(x, p)$ is weakly $\alpha$-Lipschitz in the Hausdorff metric for each $p$ fixed. Then $S(y, p)$ is a continuous mapping from $Y \times P$ to $X$. Moreover, $S(y, p)$ is weakly differentiable at $(y, p)$ for each $(y, p) \in S(y, p)$.

Theorem 4.11. Let $F(x, p) : X \times P \rightarrow Y$ be a set-valued mapping. Assume that $F(x, p)$ is weakly lower semicontinuous, weakly continuous, and weakly upper semicontinuous with respect to $x$ for each $p$ fixed. Suppose that $F(x, p)$ is weakly $\alpha$-Lipschitz in the Hausdorff metric for each $p$ fixed. Then $S(y, p)$ is a continuous mapping from $Y \times P$ to $X$.

4.3 Lipschitz regularity and metric implicit function theorem.

In this section we consider a more general setting for the implicit function theorem. Let $F(x, p) : X \times P \rightarrow Y$ be a set-valued mapping. Assume that $F(x, p)$ is weakly lower semicontinuous, weakly continuous, and weakly upper semicontinuous with respect to $x$ for each $p$ fixed. Suppose that $F(x, p)$ is weakly $\alpha$-Lipschitz in the Hausdorff metric for each $p$ fixed. Then $S(y, p)$ is a continuous mapping from $Y \times P$ to $X$. Moreover, $S(y, p)$ is weakly differentiable at $(y, p)$ for each $(y, p) \in S(y, p)$.

Theorem 4.12. Let $F(x, p) : X \times P \rightarrow Y$ be a set-valued mapping. Assume that $F(x, p)$ is weakly lower semicontinuous, weakly continuous, and weakly upper semicontinuous with respect to $x$ for each $p$ fixed. Suppose that $F(x, p)$ is weakly $\alpha$-Lipschitz in the Hausdorff metric for each $p$ fixed. Then $S(y, p)$ is a continuous mapping from $Y \times P$ to $X$.
where $\alpha$ will be further determined by Lipschitz moduli of mappings involved.

**Theorem 4.8** (general proposition on implicit functions). We assume that $\overline{y} \in F(\overline{x}, \overline{p})$ and $F$ satisfies the following conditions: there are constants $K > 0$, $\alpha > 0$ and a sufficiently small $\varepsilon > 0$ such that the following relations hold:

(a) $F(\cdot, p)$ is regular near $((\overline{x}, \overline{y}), \overline{p})$ uniformly in $p$ with the rate of metric regularity not grater than $K$;

(b) $F(x, \cdot)$ is pseudo-Lipschitz near $(\overline{x}, (\overline{p}, \overline{y}))$ uniformly in $x$ with the Lipschitz modulus not greater than $\alpha$.

Then $S$ has the Aubin property near $((\overline{y}, \overline{p}), \overline{x})$ with the Lipschitz modulus with respect to the metric $d_1$ in $Y \times P$ not greater than $\text{reg} F(\cdot, \overline{p})(\overline{x}, \overline{y})$.

In particular, if we are interested in solutions of the inclusion $\overline{y} \in F(x, p)$ (with fixed $\overline{y}$), then under the assumption of the theorem the solution mapping $p \mapsto S_{\overline{y}}(p)$ has the Aubin property near $(\overline{p}, \overline{x})$ with Lipschitz modulus not exceeding $K\alpha$.

**Proof.** As $F(\overline{x}, \overline{p}) \neq \emptyset$, the uniform pseudo-Lipschitz property implies that $S(y, p) \neq \emptyset$ for $(y, p)$ close to $(\overline{y}, \overline{p})$. If now $y \in F(x, p)$, then

$$d(x, S(y', p')) \leq Kd(y', F(x, p')) \leq K(d(y, y') + d(y, F(x, p'))),$$

and the proof is completed. $\square$

**Definition 4.9.** Let $F : X \rightrightarrows Y$, and let $\overline{y} \in F(\overline{x})$. We say that $F$ is strongly (metrically) regular near $(\overline{x}, \overline{y}) \in \text{Graph } F$ if for some $\varepsilon > 0$, $\delta > 0$ and $K \in [0, \infty)$

$$B(\overline{y}, \delta) \subset F(B(\overline{x}, \varepsilon)) \quad \& \quad d(x, u) \leq Kd(y, F(x)) \quad (4.1)$$

whenever $x \in B(\overline{x}, \varepsilon)$, $u \in B(\overline{x}, \varepsilon)$ and $y \in F(u) \cap B(\overline{y}, \delta)$.

We shall also say following [55] that $F$ has a single-valued localization near $(\overline{x}, \overline{y})$ if there are $\varepsilon > 0$, $\delta > 0$ such that the restriction of $F(x) \cap B(\overline{y}, \delta)$ to $B(\overline{x}, \varepsilon)$ is single-valued. If in addition, the restriction is Lipschitz continuous, we say that $F$ has Lipschitz localization near $(\overline{x}, \overline{y})$.

It is obvious from the definition that strong regularity implies regularity: the second relation in (4.1) is clearly stronger than metric regularity.

**Proposition 4.10** (characterization of strong regularity). Let $F : X \rightrightarrows Y$ and $(\overline{x}, \overline{y}) \in \text{Graph } F$. Then the following properties are equivalent

(a) $F$ is strongly regular near $(\overline{x}, \overline{y})$;

(b) there are $\varepsilon > 0$ and $\delta > 0$ such that $B(\overline{y}, \delta) \subset F(B(\overline{x}, \varepsilon))$ and

$$F(x) \cap F(u) \cap B(\overline{y}, \delta) = \emptyset, \quad (4.2)$$

whenever $u \neq x$ and both $x$ and $u$ belong to $B(\overline{x}, \varepsilon)$;

(c) $F$ is regular near $(\overline{x}, \overline{y})$ and there are $\varepsilon > 0$, $\delta > 0$ such that $F^{-1}$ has a
single-valued localization near \((\eta, \xi)\);

(d) \(F^{-1}\) has a Lipschitz localization \(G(y)\) near \((\eta, \xi)\). In particular \(y \in F(G(y))\) for all \(y\) of a neighborhood of \(\eta\).

Moreover, if \(F\) is strongly regular near \((\bar{x}, \bar{y})\), then the lower bound of \(K\) for which the second part of (4.1) holds and the Lipschitz modulus of its Lipschitz localization \(G\) at \(y\)

coincide with \(\text{reg} F(\xi, \eta)\).

**Theorem 4.11** (persistence of strong regularity under Lipschitz perturbation). We consider a set-valued mapping \(\Phi : X \rightrightarrows Y\) with complete graph, and a (single-valued) mapping \(G : X \times Y \to Z\). Let \(y \in \Phi(x)\) and \(z = G(x, y)\). We assume that

(a) \(\Phi\) is strongly regular near \((\bar{x}, \bar{y})\) with \(\text{sur} \Phi(\bar{x} | \bar{y}) > r\);

(b) \(G(x, \cdot)\) is an isometry from \(Y\) onto \(Z\) for any \(x\) of a neighborhood of \(\bar{x}\);

(c) \(G(\cdot, y)\) is Lipschitz with constant \(\ell < r\) in a neighborhood of \(\bar{x}\), the same for all \(y\) of a neighborhood of \(\bar{y}\).

Set \(F(x) = G(x, \Phi(x))\). Then \(F\) is strongly regular near \((\xi, \eta)\).

In particular, if \(Y\) is a normed space, \(\Phi\) is strongly regular near \((\bar{x}, \bar{y})\) \(\in \text{Graph} \Phi\) and \(G(x, y) = g(x) + y\) with \(\text{lip} g(\bar{x}) < \text{sur} \Phi(\bar{x} | \bar{y})\), then \(F(x) = \Phi(x) + g(x)\) is strongly regular near \((\xi, \eta + g(\bar{x}))\).

**Remark 4.12.** It is to be observed in connection with the last theorem that strong regularity is not preserved under set-valued perturbations like those in Theorem 4.1. Here is a simple example:

\[
\Psi(x, u) = x + u^2[-1, 1] \quad (x, u \in \mathbb{R}), \quad \eta = 0.
\]

Clearly \(\Psi(\cdot, 0)\) is strongly regular but \(F(x) = x + x^2[-1, 1]\) is of course regular but not strongly regular.

It follows that strong regularity is somewhat less robust compare to the standard regularity.

**Theorem 4.13** (implicit function theorem - metric version). Assume in addition to the assumptions of Theorem 4.8 that

\[
F(x, p) \cap F(x', p) \cap B(\eta, \varepsilon) = \emptyset \quad \forall x, x' \in B(\eta, \varepsilon), \quad x \neq x', \quad p \in B(\bar{p}, \varepsilon).
\] (4.3)

Then the solution map \(S\) has a Lipschitz localization \(G\) near \(((\bar{p}, \bar{y}), \xi)\) with \(\text{lip} G(\bar{p}, \bar{y}) \leq K\) (with respect to the \(d_\alpha\)-metric in \(Y \times P\). In particular \(z \in F(S(p, y), y)\) for all \((p, y)\) of a neighborhood of \((\bar{p}, \bar{y})\).

The conclusion is already very similar to the conclusion of the classical implicit function theorem. Indeed, it contains precisely the same information about the solution, namely its uniqueness in a neighborhood and its Lipschitz continuity (replacing differentiability) with the Equivalence Theorem 2.7 providing, along with the concluding part of Proposition 4.10 an estimate for the Lipschitz constant of the solution map (replacing the formulas for partial derivative in the classical theorem). Moreover, the proof below is based on the same main idea as the proof of the classical theorem, say the second proof in [55].
Proof. Consider the set-valued mapping $\Phi$ from $X \times P$ into $P \times Y$. defined by

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

Then $(\bar{p}, \bar{y}) \in \Phi(\bar{x}, \bar{p})$. We claim that $\Phi$ is strongly regular near $((\bar{x}, \bar{p}), (\bar{p}, \bar{y}))$. Indeed, we have for $x, p, y$ sufficiently close to $\bar{x}, \bar{p}, \bar{y}$

$$\Phi^{-1}(x, y) = \{p\} \times S(p, y) \quad (4.4)$$

By Theorem 4.8 $S$ has the Aubin property at $((\bar{p}, \bar{y}), \bar{x})$. This obviously implies that $\Phi^{-1}$ has the Aubin property at $((\bar{p}, \bar{y}), (\bar{x}, \bar{p}))$. The latter means that $\Phi$ is regular at $((\bar{x}, \bar{p}), (\bar{p}, \bar{y}))$.

On the other hand, $(p, y) \in \Phi(x, p) \cap \Phi(x', p')$ means that $p = p'$ and $y \in F(x, p) \cap F(x', p')$, so that (4.3) may happen only if $x = x'$. This proves the claim.

By Proposition 4.10 there is a Lipschitz localization of $\Phi^{-1}$ defined in a neighborhood of $(\bar{p}, \bar{y})$. By (4.3) this localization has the form $(p, G(p, y))$, where $G(p, y) \in S(p, y)$. Thus $G$ is a Lipschitz localization of $S$ and by Theorem 4.8 its Lipschitz constant is not greater than $K$.

Theorem 4.14 (metric infinitesimal implicit function theorem). Let $\bar{y} \in F(\bar{x}, \bar{p})$, and assume that there are $\xi > 0$, $r > 0$, $\ell > 0$, $\varepsilon > 0$ are such that for all $x, y, p, v$ satisfying

$$d(x, \bar{x}) < \varepsilon, \ d(y, \bar{y}) < \varepsilon, \ d(p, \bar{p}) < \varepsilon,$$

either Graph $F$ is complete and

(a) $|\nabla \varphi_y(x, v)| > r$ if $v \in F(x, p)$ and $d(y, v) > 0$

or $X$ is a complete space and

(a) $|\nabla \varphi_y(x, v)| > r$ if $\varphi_y(x, p) > 0$

holds along with

(b) $|\nabla \psi_y(x, v)| < \ell d(p, p')$, if $y \in F(x, p')$ for some $p' \in B(\bar{p}, \varepsilon)$.

Then $S$ has the Aubin property near $(\bar{y}, \bar{p})$ with $\text{lip} S((\bar{y}, \bar{p}), \bar{x}) \leq r^{-1}$ if $Y \times P$ is considered with the distance $d_1((y, p), (y', p')) = \ell d(p, p') + d(y, y')$.

The proof of the theorem consists in verifying the assumptions of Theorem 4.8 for all $(x, y, p)$ of a neighborhood of $(\bar{x}, \bar{p}, \bar{y})$ and $p'$ close to $\bar{p}$.

The next theorem is an infinitesimal counterpart of Theorem 4.14

Theorem 4.15. In addition to the conditions of Theorem 4.14 we assume that

(c) $|\nabla \psi_y(x, v)(x)| > 0$ if $y \in F(x', p)$ for some $x' \neq x$.

Then $S$ has a Lipschitz localization $G$ in a neighborhood of $(\bar{p}, \bar{x})$ with $G(\bar{p}, \bar{y}) = \bar{x}$ and the Lipschitz constant (with respect to the $d_1$-metric in $P \times Y$) not exceeding $r^{-1}$.

Proof. Indeed, it follows from (c) that $y \notin F(x, p)$ that is $(F(x, p) \cap F(x', p)) \cap B(\bar{y}, \varepsilon) = \emptyset$ for $x, x'$ close to $\bar{x}$ and $p$ close to $\bar{p}$ and the reference to Theorems 4.14 and 4.13 completes the proof.
There have been numerous publications extending, one way or another, the implicit function theorem to settings of variational analysis, see e.g. [11, 55, 70, 88, 124, 135, 136]. Most of them deal with Banach spaces and/or specific classes of mappings, e.g. associated with generalized equations. It should be also said that some results named “implicit function theorem” are rather parametric regularity or subregularity theorems giving uniform (w.r.t parameter) estimates for regularity rates of a mapping depending on a parameter.

The concept of strong regularity was introduced by Robinson in [154]. A number of characterizations of strong regularity can be found in [55]. It is appropriate to mention (especially because we do not discuss these questions in the paper) that there are certain important classes of mappings for which regularity and strong regularity are equivalent. Such are monotone operators, in particular subdifferentials of convex functions, or Kojima mappings associated with constrained optimization [55, 109].

5 Banach space theory.

Needless to say that the vast majority of applications of the theory of metric regularity relate to problems naturally stated in Banach spaces. Variational analysis and metric regularity theory in Banach spaces are distinguished by
(a) the existence of an approximation mechanisms, both primal and dual, using homogeneous mappings (graphical derivatives and coderivatives) in case of set-valued mappings or directional subderivatives and subdifferentials for functions;
(b) the possibility of separable reduction for metric regularity that allows to reduce much of analysis to mappings between separable spaces;
(c) the existence of a class of linear perturbations, most natural and interesting in many cases.

5.1 Techniques of variational analysis in Banach spaces.
5.1.1 Homogeneous set-valued mappings.

Definition 5.1. A set valued mapping $\mathcal{H} : X \rightrightarrows Y$ is homogeneous if its graph is a pointed cone. The latter means that $0 \in \mathcal{H}(0)$. The mapping
$$\mathcal{H}^*(y^*) = \{x^* : \langle x^*, x \rangle - \langle y^*, y \rangle \leq 0, \forall (x, y) \in \text{Graph } \mathcal{H}\}$$
is called adjoint or dual to $\mathcal{H}$ (or the dual convex process as it is often called for the reasons to be explained in the next chapter). It is an easy matter to see, that
$$\text{Graph } \mathcal{H}^* = \{(y^*, x^*) : (x^*, -y^*) \in (\text{Graph } \mathcal{H})^c\}.$$With every homogeneous mapping $\mathcal{H}$ we associate the upper norm
$$\|\mathcal{H}\|_+ = \sup\{\|y\| : y \in \mathcal{H}(x), x \in \text{dom } \mathcal{H}, \|x\| \leq 1\},$$
and the lower norm
$$\|\mathcal{H}\|_- = \sup_{x \in B \cap \text{dom } \mathcal{H}} \inf\{\|y\| : y \in \mathcal{H}(x)\} = \sup_{x \in B \cap \text{dom } \mathcal{H}} d(0, \mathcal{H}(x)).$$
For single-valued mappings with \( \text{dom } H = X \) both quantities coincide and we may speak about the norm of \( H \). The mapping \( H \) is bounded if \( \|H\|_+ < \infty \). This obviously means that there is an \( r > 0 \) such that \( H(x) \subset r\|x\|B_Y \) for all \( x \).

Very often however, in the context of regularity estimates, it is more convenient to deal with different quantities defined by way of the norms as follows:

\[
C(H) = \|H^{-1}\|_{-1} \quad \text{and} \quad C^*(H) = \|H^{-1}\|_{+1}.
\]

The quantities are respectively called the Banach constant and the dual Banach constant of \( H \). To justify the terminology, note that for linear operators they coincide with the Banach constants introduced for the latter in the first section.

The proposition below containing important geometric interpretation of the concepts shows that the Banach constants are actually very natural objects..

**Proposition 5.2** (cf. Proposition [1,3]). For any homogeneous \( H : X \rightrightarrows Y \)

\[
C(H) = \text{contr}H(0|0) = \sup\{r \geq 0 : rB_Y \subset H(B_X)\};
\]

\[
C^*(H) = (\text{subreg}H(0|0))^{-1} = \inf\{\|y\| : y \in H(x), \|x\| = 1\} = \inf_{\|x\|=1} d(0,H(x)).
\]

**Proof.** The equality \( \text{contr}H(0|0) = \sup\{r \geq 0 : rB_Y \subset H(B_X)\} \) follows from homogeneity of \( H \). On the other hand, saying that \( rB_Y \subset H(B_X) \) is the same as saying that for any \( y \) with \( \|y\| = r \) there is an \( \|x\| \) with \( \|x\| \leq 1 \) such that \( x \in H^{-1}(y) \) which means that \( \|H^{-1}\|_{-} \leq r^{-1} \) and therefore \( C(H) \geq \text{contr}H(0|0) \). Likewise, \( \|H^{-1}\|_{+} < r^{-1} \) means that for any \( y \) with \( \|y\| = 1 \) there is an \( x \) with \( \|x\| \leq r^{-1} \) such that \( y \in H(x) \) from which we get that \( rB_Y \subset H(B_X) \) and the first equality follows.

To prove the second equality, consider first the case \( C^*(H) < \infty \). Then

\[
C^*(H) = \inf_{\|y\|=1} \inf\{\|x\|^{-1} : x \in H^{-1}(y)\}
\]

\[
= \inf\{\|y\| : y \in H(x), \|x\| = 1\}.
\]

If \( C^*(H) = \infty \), and therefore \( \|H^{-1}\|_{+} = 0 \), then for any \( y \) the set \( H^{-1}(y) \) is either empty (recall our convention: \( \inf \emptyset = \infty \), \( \sup \emptyset = 0 \)) or contains only the zero vector. Hence the domain of \( H \) is a singleton containing the origin. It follows that \( \inf\{\|y\| : y \in H(x), \|x\| = 1\} = \inf \emptyset = \infty \).

This proves the left equality. Consider again the case \( C^*(H) > 0 \). Then \( \|H^{-1}\|_{+} < \infty \) and consequently, \( H^{-1}(0) = \{0\} \). It follows that \( d(x,H^{-1}(0)) = \|x\| \). Setting \( K = (C^*(H))^{1} \), we get for any \( x \) with \( \|x\| = 1 \):

\[
Kd(0,H(x)) \geq 1 = \|x\| = d(x,H^{-1}(0))
\]

and on the other hand for any \( K' < K \) we can find an \( x \) with \( \|x\| = 1 \) such that \( K'd(0,H(x)) < 1 \). It follows that \( K = \text{subreg}H(0|0) \). The case \( C^*(H) = 0 \) is treated as above.

**Corollary 5.3.** For any homogeneous mappings \( H : X \rightrightarrows Y \) and \( \mathcal{E} : Y \rightrightarrows Z \)

\[
C(\mathcal{E} \circ H) \geq C(\mathcal{E}) \cdot C(H).
\]
Proof. Take \( \rho < C(\mathcal{H}) \). Then \( \rho(B_Y) \subset \mathcal{H}(B_X) \) and therefore

\[
C(\mathcal{E} \circ \mathcal{H}) = \sup\{ r \geq 0 : rB_Z \subset (\mathcal{E} \circ \mathcal{H})(B_X) \} \geq \sup\{ r \geq 0 : rB_Z \subset \mathcal{E}(\rho B_Y) \} = \rho C(\mathcal{E})
\]

and the result follows. \( \square \)

We shall see that the tangential (primal) regularity estimates are stated in terms of Banach constants of contingent derivatives of the mapping while the subdifferential estimate need dual Banach constants of coderivatives. The following theorem is the first indicator that (surprisingly!) the dual estimates can be better.

**Theorem 5.4** (basic inequality for Banach constants). For any homogeneous set-valued mapping \( H : X \rightrightarrows Y \)

\[
C^*(\mathcal{H}^*) \geq C(\mathcal{H}) \geq C^*(\mathcal{H}).
\]

Note that for linear operators we have equality – see Proposition 1.3. In the next section we shall see that the equality also holds for convex processes and some other set-valued mappings.

**Proof.** The right inequality is immediate from the definition. If \( C(\mathcal{H}) = \infty \), that is \( \|\mathcal{H}^{-1}\|_- = 0 \), then for any \( y \in Y \) there is a sequence \( (x_n) \subset X \) norm converging to zero and such that \( y \in \mathcal{H}(x_n) \). It is easy to see that in this case

\[
\mathcal{H}(y^*) = \begin{cases} 
0, & \text{if } y^* \neq 0; \\
X^*, & \text{if } y^* = 0,
\end{cases}
\]

that is \( (\mathcal{H}^*)^{-1} \equiv \{0\} \), \( \| (\mathcal{H}^*)^{-1} \|^* = 0 \) and hence \( C^*(\mathcal{H}^*) = \infty \).

Let now \( \infty > C(\mathcal{H}) = r > 0 \). Set \( \lambda = r^{-1} \). Then \( \|\mathcal{H}^{-1}\|_- = \lambda \) so that for any \( y \) with \( \|y\| = 1 \) and any \( \varepsilon > 0 \) there is an \( x \) such that \( \|x\| \leq \lambda + \varepsilon \) and \( y \in \mathcal{H}(x) \). Let now \( x^* \in \mathcal{H}^*(y^*) \), that is \( \langle x^*, x \rangle - \langle y^*, y \rangle \leq 0 \) if \( y \in \mathcal{H}(x) \). Take \( y \in S_Y \) such that \( \langle y^*, y \rangle \leq (-1 + \varepsilon)\|y^*\| \) and choose an \( x \in \mathcal{H}^{-1}(y) \) with \( \|x\| \leq \lambda + \varepsilon \). Then

\[
-(\lambda + \varepsilon)\|x^*\| \leq \langle x^*, x \rangle \leq \langle y^*, y \rangle \leq (-1 + \varepsilon)\|y^*\|
\]

that is \( (\lambda + \varepsilon)\|x^*\| \geq (1 - \varepsilon)\|y^*\| \). As \( \varepsilon \) can be chosen arbitrarily close to zero this implies that \( \| (\mathcal{H}^*)^{-1} \|_+ \leq r^{-1} \) and therefore \( C^*(\mathcal{H}^*) \geq r = C(\mathcal{H}) \). \( \square \)

The following property plays an essential role in future discussions.

**Definition 5.5** (non-singularity). We say that \( \mathcal{H} \) is non-singular if \( C^*(\mathcal{H}) > 0 \). Otherwise we shall call \( \mathcal{H} \) singular.

We conclude the subsection with showing that regularity of a homogeneous mapping near the origins implies its global regularity.

**Proposition 5.6.** Let \( X \) and \( Y \) be two Banach spaces, and let \( F : X \rightrightarrows Y \) be a homogeneous set-valued mapping. If \( F \) is regular near \((0,0)\), then it is globally regular with the same rates.
Proof. By the assumption, there are \( K > 0 \) and \( \varepsilon > 0 \) such that \( d(x, F^{-1}(y)) \leq Kd(y, F(x)) \) if \( \max\{\|x\|,\|y\|\} < \varepsilon \). Let now \((x, y)\) be an arbitrary point of the graph. Set \( \|m\| = \max\{\|x\|,\|y\|\} \), and let \( \mu < \varepsilon/m \). Then

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

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

The norms for homogeneous multifunctions were originally introduced first by Rockafellar [157] and Robinson [148] in the context of convex processes (lower norm) and then by Ioffe [82] (upper norm for arbitrary homogeneous maps) and Borwein [23] (upper norm and duality for convex processes - see also [24, 25, 55]). The dual Banach constant \( C^* \) was also introduced in [82]. The meaning of the primal constant has undergone some evolution since it first appeared in [82]. The \( C(H) \) introduced here is reciprocal to that in [84] mainly because the connection of Banach constants with the norms of homogeneous mappings makes the present definition more natural.

5.1.2 Tangent cones and contingent derivatives

Given a set \( Q \subset X \) and an \( \overline{x} \in Q \). The tangent (or contingent) cone \( T(Q, \overline{x}) \) is the collection of \( h \in X \) with the following property: there are sequences of \( t_k \searrow 0 \) and \( h_k \to h \) such that \( \overline{x} + t_k h_k \in Q \) for all \( k \). If \( F : X \rightrightarrows Y \) then the contingent or graphical derivative of \( F \) at \((\overline{x}, \overline{y})\) is the set-valued mapping

\[
X \ni h \mapsto DF(\overline{x}, \overline{y})(h) = \{v \in Y : (h, v) \in T(\text{Graph } F, (\overline{x}, \overline{y}))\}.
\]

Let now \( f \) be a function on \( X \) finite at \( \overline{x} \). The function

\[
h \mapsto f^-(\overline{x}; h) = \liminf_{(t, h') \to (0^+, h)} t^{-1}(f(\overline{x} + th') - f(\overline{x}))
\]

is called the Dini-Hadamard lower directional derivative of \( f \) at \( \overline{x} \). This function is either lsc and equal to zero at the origin or identically equal to \(-\infty \). The latter of course cannot happen if \( f \) is Lipschitz near \( \overline{x} \).

The connection between the two concepts is very simple: \( h \in T(Q, \overline{x}) \) if and only if \( d^-(\cdot, Q)(\overline{x}; h) = 0 \) and \( \alpha = f^-(\overline{x}; h) \) if and only if \( (h, \alpha) \in T(\text{epi } f, (\overline{x}, f(\overline{x}))) \).

If \( F : X \rightrightarrows Y \) then the contingent derivative of \( F \) at \( \overline{x} \) is the set-valued mapping

\[
X \ni h \mapsto DF(\overline{x}; h) = \{v \in Y : (h, v) \in T(\text{Graph } F, (\overline{x}, F(\overline{x})))\}.
\]

The contingent tangent cone and contingent derivative were introduced by Aubin in [5] (see [8] for detailed comments concerning genesis of the concept.)

5.1.3 Subdifferentials, normal cones and coderivatives.

From now on, unless the opposite is explicitly said, all spaces are assumed separable. Thanks to the separable reduction theorem to be proved in the next subsection such a restriction is justifiable in the context of regularity theory. On the other hand, it provides
for a substantial economy of efforts, especially in the non-reflexive (or to be precise, non-Asplund) case.

Subdifferential is among the most fundamental concepts in local variational analysis. Essential for the infinite dimensional variational analysis are five types of subdifferentials: Fréchet subdifferential, Dini-Hadamard subdifferential (the two are sometimes called “elementary subdifferentials”), limiting Fréchet subdifferential, G-subdifferential and the generalized gradient. In Hilbert space there is one more convenient construction, “proximal subdifferential”. We shall introduce it in § 7.

So let $f$ be a function on $X$ which if finite at $x$. The sets

$$\partial_H f(x) = \{ x^* \in X^* : \langle x^*, h \rangle \leq f^*(x; h), \forall h \in X \}$$

and

$$\partial_F f(x) = \{ x^* \in X^* : \langle x^*, h \rangle \leq f(x + h) - f(x) + o(||h||) \}$$

are called respectively the Dini-Hadamard and Fréchet subdifferential of $f$ at $x$. The corresponding limiting subdifferential at $x$ (we denote them for a time being $\partial_{LH}$ and $\partial_{LF}$) is defined as the collection of $x^*$ such that there is a sequence $(x_n, x_n^*)$ with $x_n$ norm converging to $x$ and $x_n^*$ weak*-converging to $x^*$. The essential point in the definition of the limiting subdifferentials is that only sequential weak*-limits of elements of elementary subdifferentials are considered. The limiting Dini-Hadamard subdifferential is basically an intermediate product in the definition of the G-subdifferential. Given a set $Q \subset X$, the G-normal cone to $Q$ at $x \in Q$ is

$$N_G(S, x) = \bigcup_{\lambda \geq 0} \lambda \partial_{LH} d(\cdot, Q)(x).$$

The G-subdifferential of $f$ at $x$ is defined as follows

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

The cone $N_C(Q, x) = \text{cl(conv } N_G(Q, x))$ is Clarke’s normal cone to $Q$ at $x$ and the set

$$\partial_C f(x) = \{ x^* : (x^*, -1) \in N_C(Q, x) \}$$

is the subdifferential or generalized gradient of Clarke.

**Proposition 5.7** (some basic properties of subdifferentials). The following statements hold true:

(a) for any lsc function $\partial_H f(x) \neq \emptyset$ on a dense subset of dom $f$;

(b) the same is true for $\partial_F$ if there is a Fréchet differentiable (off the origin) norm in $X$ (that is if $X$ is an Asplund space);

(c) if $f$ is Lipschitz near $x$, then $\partial_G f(x) \neq \emptyset$ and the set-valued mapping $x \mapsto \partial_G f(x)$ is compact-valued (see (f) below) and upper semicontinuous;

(d) if $f$ is continuously (or strictly) differentiable at $x$, then $\partial f(x) = \{ f'(x) \}$ for any of the mentioned subdifferentials;

(e) if $f$ is convex, then all mentioned subdifferentials coincide with the subdifferential in the sense of convex analysis: $\partial f(x) = \{ x^* : f(x + h) - f(x) \geq \langle x^*, h \rangle, \forall h \}$;

(f) if $f$ is Lipschitz near $x$ with Lipschitz constant $K$, then $\| x^* \| \leq K$ for any $x^* \in \partial f(x)$ and any of the mentioned subdifferentials;
(g) if \( f \) is Lipschitz near \( x \), then \( \partial_{LH} f(x) = \partial_G f(x) \) and \( \partial_G f(x) = \text{cl}(\text{conv} \partial_G(x)) \);
(h) if \( f \) is lsc and \( X \) is an Asplund space, then \( \partial_{LG} f(x) = \partial_G f(x) \) for any \( x \);
(i) if \( f(x, y) = \varphi(x) + \psi(y) \), then \( \partial f(x, y) = \partial \varphi(x) + \partial \psi(y) \), where \( \partial \) any of \( \partial_F, \partial_H, \partial_G \) (but not \( \partial_C \)).

Remark 5.8. It should be observed in connection with the proposition that
- \( \partial_{LH} \) has little interest for non-Lipschitz functions: it may be too big to contain any useful information about the function.
- If \( X \) is not Asplund, \( \partial_{LG} f(x) \) may be identically empty even for a very simple Lipschitz function (e.g. \( -\|x\| \) in \( C[0,1] \)). In terminology of the subdifferential calculus this means that \( \partial_F \) cannot be trusted on non-Asplund spaces.

We do not need here a formal definition for the concept of a subdifferential trusted on a space or a class of spaces (see e.g. \([94]\)). Loosely speaking this means that a version of the fuzzy variational principle is valid for the subdifferentials of lsc functions on the space. Just note that the Fréchet subdifferential is trusted on Asplund spaces and only on them, Dini-Hadamard subdifferential is trusted on Gâteaux smooth spaces and the G-subdifferential and the generalized gradient are trusted on all Banach spaces.

There is one more important property of subdifferentials that has not been mentioned in the proposition. This property is called tightness and it characterizes a reasonable quality of lower approximation provided by the subdifferential (see \([94]\)). It turns out that the Dini-Hadamard, Fréchet and G-subdifferentials are tight but Clarke’s generalized gradient is not. This determines a relatively small role played by generalized gradient in the regularity theory. On the other hand, generalized gradient typically is much easier to compute and work with. Moreover, convexity of the generalized gradient makes it the only subdifferential that can be used in the critical point theory associated with the concept of “weak slope”, not considered here.

We do not need here the general theory of subdifferentials. Just mention in connection with the property (h) in Proposition 5.7 that in separable spaces the G-subdifferential is a unique subdifferential having a certain collection of properties (including tightness, (c), (e), (f) and "exact calculus" as defined in the proposition below). It is to be again emphasized that we assume all spaces separable.

Proposition 5.9 (basic calculus rules). Let \( f(x) = f_1(x) + f_2(x) \), where both functions are lsc and one of them is Lipschitz near \( \mathcal{T} \). Then the following statements are true

1. Fuzzy variational principle: If \( f \) attains a local minimum at \( \mathcal{T} \), then there are sequences \( (x_{in}) \) and \( (x^*_{in}) \), \( i = 1, 2 \) such that \( x_{in} \to \mathcal{T}, x^*_{in} \in \partial_H f_{in}(x_{in}) \) and \( \|x^*_{1in} + x^*_{2in}\| \to 0 \),

2. Fuzzy sum rule: if \( X \) is Asplund and \( x^* \in \partial_F f(\mathcal{T}) \), then there are sequences \( (x_{in}) \) and \( (x^*_{in}) \), \( i = 1, 2 \) such that \( x_{in} \to \mathcal{T}, x^*_{in} \in \partial_H f_{in}(x_{in}) \) and \( \|x^*_{1in} + x^*_{2in} - x^*\| \to 0 \).

3. Exact sum rule: \( \partial_G f(\mathcal{T}) \subset \partial_G f_1(\mathcal{T}) + \partial_G f_2(\mathcal{T}) \).

Let \( Q \subset X \) and \( x \in Q \). Given a subdifferential \( \partial \), the set
\[
N(Q, x) = \partial i_Q(x),
\]
always a cone, is called the normal cone to $Q$ at $x$ associated with $\partial$. It is an easy matter to see that in case of $\partial_C$ this definition coincides with the given earlier. For normal cones associated with $\partial_H$ and $\partial_F$ we use notation $N_H$ and $N_F$.

Let $F : X \rightrightarrows Y$ and $\overline{y} \in F(\overline{x})$. Given a subdifferential $\partial$ and normal cone associated with $\partial$, the set-valued mapping

$$y^* \mapsto D^* F(\overline{x}, \overline{y})(y^*) = \{ x^* : (x^*, -y^*) \in N(\text{Graph } F, (\overline{x}, \overline{y})) \}$$

is called the coderivative of $F$ at $(\overline{x}, \overline{y})$ associated with $\partial$. We use notation $D^*_H$, $D^*_F$, and $D^*_G$ for the coderivatives, associated with the mentioned subdifferentials.

There is a number of monographs and survey articles in which subdifferentials are studied at various levels of generality: [159] (finite dimensional theory), [27] 130, 144, 162 (Asplund spaces), [94, 144] (general Banach spaces), [38, 40] (generalized gradients).

Concerning the sources of the main concepts: Clarke’s subdifferential was first to appear - it was introduced in Clarke’s 1973 thesis [33] and in printed form first appeared in [34]. In $[35]$, it is not clear where the Fréchet subdifferential first appeared, probably in [20], the Dini-Hadamard subdifferential was introduced by Penot in [141], the sequential limiting Fréchet subdifferential for functions on Fréchet smooth spaces was introduced by Kruger in mimeographed paper [112] in 1981 (not in [116] as stated in e.g. [132, 130] and many other publications - the definition given in [116] is purely topological and does not involve sequential weak*-limits) and in printed form appeared in [113] (see [94] for details). The $G$-subdifferential was first defined in [81] but its definition was later modified in [85].

5.2 Separable reduction.

In this subsection $X$ and $Y$ are general Banach spaces, not necessarily separable. Recall that by $S(X)$ we denote the collection of separable subspaces of $X$.

**Proposition 5.10.** Assume that $\text{sur}(F(\overline{x}, \overline{y})) > r$. Then for any $L_0 \subset S(X)$ and $M \subset S(Y)$ there is an $L \in S(X)$ containing $L_0$ such that for sufficiently small $t \geq 0$

$$y + rt(B_Y \cap M) \subset \text{cl}(F(x + t(1 + \delta)(B_X \cap L))),$$

if $\delta > 0$ and the pair $(x, y) \in (\text{Graph } F) \cap (L \times M)$ is sufficiently close to $(\overline{x}, \overline{y})$.

**Proof.** Take an $\varepsilon > 0$ to guarantee that the inclusion below holds for $x \in B(\overline{x}, \varepsilon)$.

$$F(x) \cap B(\overline{y}, \varepsilon) + trB_Y \subset F(B(x, t)). \quad (5.2)$$

We shall prove that there is a nondecreasing sequence $(L_n)$ of separable subspaces of $X$ such that:

$$y + rt(B_Y \cap M) \subset \text{cl}(F(x + t(1 + \delta)(B_X \cap L_{n+1}))), \quad (5.3)$$

for all $\delta > 0$ and all $(x, y) \in (\text{Graph } F) \cap (L_n \times M)$ sufficiently close to $(\overline{x}, \overline{y})$. Then to complete the proof, it is sufficient to set $L = \text{cl}(\cup L_n)$.

Assume that we have already $L_n$ for some $n$. Let $(x_i, y_i)$ be a dense countable subset of the intersection of $(\text{Graph } F) \cap (L_n \times M)$ with the neighborhood of $(\overline{x}, \overline{y})$ in which (5.2) is guaranteed, let $(v_j)$ be a dense countable subset of $B_Y \cap M$, and let $(t_k)$ be a dense...
Proof. For any $i,j,k = 1,2,\ldots$ we find from (5.2) an $h_{ijk} \in B_X$ such that $y_i + nh_{ijk} \in F(x_i + t_i h_{ijk})$, and let $\hat{L}_n$ be the subspace of $X$ spanned by the union of $L_n$ and the collection of all $h_{ijk}$.

If now $(x,y) \in (\text{Graph } F) \cap (L_n \times M)$, $t \in (0,1)$, $v \in B_Y$ and $(x_{im},y_{jm})$, $t_{km}$, $v_{jm}$ converge respectively to $(x,y)$, $t$ and $v$, then as $x_{im} + t_{km}(B_X \cap M_n) \subset x + t(1+\delta)(B_X \cap M_n)$ for sufficiently large $m$, we conclude that (5.3) holds with $\hat{L}_n$ instead of $L_{n+1}$.

**Theorem 5.11** (separable reduction of regularity [96]). Let $X$ and $Y$ be Banach spaces. A set-valued mapping $F : X \rightrightarrows Y$ with closed graph is regular at $(\bar{x},\bar{y}) \in \text{Graph } F$ if and only if for any separable subspace $M \subset Y$ and any separable subspace $L_0 \subset X$ with $(\bar{x},\bar{y}) \in L_0 \times M$ there exists a bigger separable subspace $L \in S(X)$ such that the mapping $F_{L \times M} : L \rightrightarrows M$ whose graph is the intersection of $\text{Graph } F$ with $L \times M$ is regular at $(\bar{x},\bar{y})$. Moreover, if there is an $r > 0$ such that for any separable $M_0 \subset Y$ and $L_0 \subset X$ there are bigger separable subspaces $M \supset M_0$ and $L \supset L_0$ such that $\text{sur}F_{L \times M}(\bar{x},\bar{y}) \geq r$, then $F$ is regular at $(\bar{x},\bar{y})$ with $\text{sur}F(\bar{x},\bar{y}) \geq r$.

**Proof.** So assume that $F$ is regular at $(\bar{x},\bar{y})$ with $\text{sur}F(\bar{x},\bar{y}) \geq r$. Then, given $L_0$ and $M$, we can find a closed separable subspace $L \subset X$ containing $L_0$ such that (5.3) holds for any $\delta > 0$, any $(x,y) \in (\text{Graph } F) \cap (L \times M)$ sufficiently close to $(\bar{x},\bar{y})$ and any sufficiently small $t > 0$.

By the Density theorem we can drop the closure operation, so that $F_{L \times M}$ is indeed regular near $(\bar{x},\bar{y})$ with $\text{sur}F_{L \times M}(\bar{x},\bar{y}) \geq (1 + \delta)^{-1}r$. As $\delta$ can be arbitrarily small we get the desired estimate for the modulus of surjection of $F_{L \times M}$.

On the other hand, if $F$ were not regular at $(\bar{x},\bar{y})$, then we could find a sequence $(x_n,y_n) \in \text{Graph } F$ converging to $(\bar{x},\bar{y})$ such that $y_n + (t_n/n)v_n \notin F(B(x_n,t_n))$ for some $t_n < 1/n$ and $v_n \in B_Y$ (respectively $y_n + t_n(r-\delta)v_n \notin F(B(x_n,t_n))$ for some $\delta > 0$).

Clearly this carries over to any closed separable subspace $L \subset X$ and $M \subset Y$ containing respectively all $x_n$, all $y_n$ and all $v_n$, so that no such $F_{L \times M}$ cannot be regular at $(\bar{x},\bar{y})$ (with the modulus of surjection $\geq r$) contrary to the assumption.

### 5.3 Contingent derivatives and primal regularity estimates

The following simple proposition establishes connection between slope of $f$ and its lower directional derivative.

**Proposition 5.12.** For any function $f$ and any $x$ at which $f$ is finite

$$|\nabla f|(x) \geq - \inf_{\|h\| = 1} f^-(x;h).$$

**Proof.** Take an $h$ with $\|h\| = 1$. We have

$$|\nabla f|(x) = \lim_{t \searrow 0} \sup_{\|u\| = 1} \frac{(f(x) - f(x + tu)^+)}{t} \geq \lim_{(t,u) \to (0^+,0)} \frac{f(x) - f(x + tu)}{t} = -f^-(x;h)$$

as claimed.
The following result is now immediate from the proposition and Theorem 5.2.

**Theorem 5.13** (tangential regularity estimate 1). Let \((\bar{x}, \bar{y}) \in \text{Graph } F\). Assume that there are neighborhoods \(U\) of \(\bar{x}\) and \(V\) of \(\bar{y}\) such that for any \(y \in V\) the function \(\psi_y\) is lower semicontinuous \(U\) and \(\inf_{\|h\| = 1} \psi_y'(x; h) \leq -r\) for \(x \in U\) and \(y \in V\). Then

\[
\text{sur} F(x|y) \geq r. \tag{5.4}
\]

(Of course a similar estimate can be obtained from Theorem 3.10.)

**Theorem 5.14** (tangential regularity estimate 2). Suppose there are a neighborhood \(U\) of \((\bar{x}, \bar{y})\) and two numbers \(c > 0\) and \(\lambda \in [0, 1)\) such that for any \((x, y) \in U \cap \text{Graph } F\)

\[
\exp(S_Y, DF(x, y)(cB_X)) \leq \lambda, \tag{5.5}
\]

then

\[
\text{sur} F(x|y) \geq \frac{1 - \lambda}{c}. \tag{5.6}
\]

**Proof.** Take an \((x, v) \in U \cap \text{Graph } F\) with \(v \neq y\) and set \(z = \|y - v\|^{-1}(y - v)\). By the assumption for any \(\lambda' > \lambda\) there is a pair \((\bar{h}, \bar{w})\) with \(\bar{w} \in DF(x, v)(\bar{h})\) such that \(\|\bar{h}\| = c\) and \(\|z - \bar{w}\| \leq \lambda'\). As \((\bar{h}, \bar{w})\) belongs to the contingent cone to the \(\text{Graph } F\) at \((x, v)\), we can find (for sufficiently small \(t > 0\)) vectors \(h(t)\) and \(w(t)\) norm converging to \(\bar{h}\) and \(\bar{w}\) respectively and such that \(v + tw(t) \in F(x + th(t))\). We have

\[
\|y - (v + tw(t))\| = \|y - v - t\bar{w}\| + o(t) \\
\leq \|y - v - tz\| + t\|z - \bar{w}\| + o(t) \\
\leq \|y - v\|(1 - \|y - v\|) + t\lambda' + o(t), \tag{5.7}
\]

so that

\[
\varphi_y^-(x, v; (\bar{h}, \bar{w})) \leq \lim_{t \to 0} \frac{\|y - t(v + w(t))\| - \|y - v\|}{t} \leq -(1 - \lambda').
\]

Take a \(\xi > 0\) such that \(\xi(1 + \lambda) < c\) and consider the \(\xi\)-norm in \(X \times Y\), Then \(\|\bar{h}, \bar{w}\|_\xi \leq \max\{c, \xi(1 + \lambda')\} = c\) (if \(\lambda'\) is sufficiently close to \(\lambda\)) and we get from \((5.8)\)

\[
\inf\{\varphi_y^-(x, v; (h, w)) : \|(h, w)\|_\xi \leq 1\} \leq \frac{1}{c}\varphi_y^-(x, v; (\bar{h}, \bar{w})) \leq -\frac{1 - \lambda'}{c}.
\]

It remains to refer to Proposition 5.12 and Theorem 3.10. \qed

**Theorem 5.15** (tangential regularity estimate 3). Let \(X\) and \(Y\) be Banach spaces, and let \(F : X \Rightarrow Y\) be a set-valued mapping with locally closed graph. Let finally \(\bar{y} \in F(\bar{x})\). Then

\[
\text{sur} F(x|\bar{y}) \geq \lim \inf_{\varepsilon \to 0} \{C(DF(x, y)) : (x, y) \in (\text{Graph } F) \cap B((\bar{x}, \bar{y}), \varepsilon), \tag{5.8}
\]

or equivalently,

\[
\text{reg} F(x|\bar{y}) \leq \lim \sup_{\varepsilon \to 0} \{(DF(x, y))^{-1} : y \in F(x), \|x - \bar{x}\| < \varepsilon, \|y - \bar{y}\| < \varepsilon, \tag{5.9}
\]

\[
= \lim_{\varepsilon \to 0} \{ \sup_{\|v\| = 1} \inf \{\|h\| : v \in DF(x, y)(h)\} : (x, y) \in (\text{Graph } F) \cap B((\bar{x}, \bar{y}), \varepsilon)\}.
\]

37
Proof. We first note that \( DF(x, v)(B_X) \) is a star-shaped set as it contains zero and \( z \in DF(x, v)(h) \) implies that \( \lambda z \in DF(x, v)(\lambda h) \) for \( \lambda > 0 \). On the other hand, by Proposition 5.2 \( C(DF(x, v)) > r > 0 \) means that \( rB_Y \subset DF(x, v)(B_X) \). It follows that \( B_Y \subset DF(x, v)(r^{-1}B_X) \). If this is true for all \((x, v) \in \text{Graph } F \) close to \((\bar{x}, \bar{y})\), this in turn means that the condition of Theorem 5.14 is satisfied with \( c = r^{-1} \) and \( \lambda = 1 \), whence the theorem.

\[ \square \]

Remark 5.16. In fact the last two theorems are equivalent. Indeed, let the conditions of Theorem 5.14 be satisfied. Then \((1 - \lambda)B_Y \subset DF(x, v)(\epsilon B_X) \) for all \((x, v) \in \text{Graph } F \) close to \((\bar{x}, \bar{y})\) and setting \( r = c^{-1}(1 - \lambda) \) we get \( rB_Y \subset DF(x, v)(B_X) \) for the same \((x, v) \).

It follows from the proofs that the estimate provided by Theorem 5.13 is never worse than the estimates given by the other two theorems. But it can actually be strictly better (unless both spaces are finite dimensional). Informally, this is easy to understand: the quality of approximation provided by the contingent derivative for a map into an infinite dimensional spaces maybe much lower than for a real-valued function. The following example illustrates the phenomenon.

Example 5.17. Let \( X = Y \) be a separable Hilbert space, and let \((e_1, e_2, \ldots)\) an orthonormal basis in \( X \). Consider the following mapping from \([0, 1]\) into \( X \):

\[ \eta(t) = \begin{cases} 0, & \text{if } t \in \{0, 1\} \\ 2^{-(n+2)}e_n, & \text{if } t = 2^{-n}, \end{cases} \]

and \( \eta(\cdot) \) is linear on every segment \([2^{-(n+1)}, 2^{-n}], n = 0, 1, \ldots \). Define a mapping from the unit ball of \( \ell_2 \) into \( \ell_2 \) by

\[ F(x) = x - \eta(\|x\|). \]

It is an easy matter to see that \( x \mapsto \eta(\|x\|) \) is \((\sqrt{5}/4)\)-Lipschitz, hence by Milyutin’s perturbation theorem \( F \) is open near the origin with the rate of surjection at least \( 1 - (\sqrt{5}/4) \).

Let us look what we get applying both statements of the theorem for the mapping. If \( \|h\| = 1 \) and \( t \in (2^{-(n+1)}, 2^{-n}] \), then \( F(th) = th - (t/2)(e_n - e_{n+1}) - 2^{-(n+2)}(2e_{n+1} - e_n) \), and it is easy to see that for no sequence \((t_k)\) converging to zero \( t_k^{-1}F(t_k) \) converge. Hence the tangent cone to the graph of \( F \) at zero consists of a single point \((0, 0)\) and the first statement gives \( \text{sur}F(0) \geq 0 \) - a trivial conclusion.

Now take an \( x \) with \( \|x\| < 1 \) and a \( y \neq F(x) \). We have

\[ \|F(x + th) - y\| = \|x + th - \eta(\|x + th\|) - y\| \leq \|x + th - \eta(\|x\|) - y\| + \|\eta(\|x + th\|) - \eta(\|x\|)\| \leq \|F(x) + th - y\| + (3/4)t\|h\|. \]

Taking \( h = (y - F(x))/\|y - F(x)\| \), we get

\[ \varphi_y(x; h) \leq \lim_{t \downarrow 0} t^{-1}\left( \left(\frac{t}{\|F(x) - y\|}\right)\|F(x) - y\| - \|F(x) - y\| \right) + \frac{\sqrt{5}}{4} = -\frac{4 - \sqrt{5}}{4} \]

which gives \( \text{sur}F(x) \geq 1 - (\sqrt{5}/4) \) for all \( x \) with \( \|x\| < 1 \).
A tangential regularity estimate, similar to but somewhat weaker than that in Theorem 5.14 was first obtained by Aubin in [6] (see also [8]) under the same assumptions. The very estimate (5.6) was obtained in [84]. Theorem 5.15 was proved by Dontchev-Quincampoix-Zlateva in [53]. Theorem 5.13 seems to have been state for the first time in [32]. Example 5.17 has also been borrowed from that paper.

5.4 Dual regularity estimates.

This is the part of the local regularity theory that attracted main attention in the 80s and 90s. The role of coderivatives was in the center of the studies. Further developments, however, that followed the discovery of the role of slope open gates for potentially stronger (and often easier to apply) results involving subdifferentials of the functions $\varphi_y$ and $\psi_y$.

5.4.1 Neighborhood estimates

There is a simple connection between slopes and norms of elements of subdifferentials.

Proposition 5.18 (slopes and subdifferentials). Let $f$ be lsc, and let an open set $U$ have nonempty intersection with $\text{dom} f$. Then for any subdifferential $\partial f(x)$

$$\inf_{x \in U} d(0, \partial f(x)) \leq \inf_{x \in U} |\nabla f(x)|.$$  

On the other hand, $\|x^*\| \geq |\nabla f(x)|$ if $x^* \in \partial f(x)$.

Combining this with Theorems 3.10 and 3.12, we get

Theorem 5.19 (subdifferential regularity estimate 1). Let $X$ and $Y$ be Banach spaces, let $F : X \rightrightarrows Y$ have a locally closed graph, and let $\partial$ be a subdifferential trusted on a class of Banach spaces containing both $X$ and $Y$. Then for any $(\bar{x}, \bar{y}) \in \text{Graph} F$ and any $\xi > 0$

$$\text{sur} F(\bar{x}|\bar{y}) \geq \liminf_{(x,v) \rightarrow (\bar{x},\bar{y})} \inf_{y \neq v} \{\|x^*\| + \xi^{-1}\|v^*\| : (x^*, y^*) \in \partial \varphi_y(x,v)\}. \quad (5.9)$$

and

$$\text{sur} F(\bar{x}|\bar{y}) \geq \liminf_{(x,v) \rightarrow (\bar{x},\bar{y})} d(0, \partial \psi_y(x,v)). \quad (5.10)$$

Theorem 5.20 (subdifferential regularity estimate 2). Let $(\bar{x}, \bar{y}) \in \text{Graph} F$. Assume that there are neighborhoods $U$ of $\bar{x}$ and $V$ of $\bar{y}$ such that for any $y \in V$ the function $\psi_y$ is lower semicontinuous and $\|x^*\| \geq r$ if $x^* \in \partial_H \psi_y(x)$ for all $x \in U$ and $y \in V$. Then

$$\text{sur} F(\bar{x}|\bar{y}) \geq r. \quad (5.11)$$

The obvious inequality $\|x^*\| \geq -f^-(x^*, h)$ if $x^* \in \partial_H f(x)$ and $\|h\| = 1$ shows that the estimate provided by the last theorem cannot be worse that the estimate of Theorem 5.13.

Our next purpose is to derive coderivative estimates for regularity rates.
Theorem 5.21 (coderivative regularity estimate 1). Let $F : X \Rightarrow Y$ be a set-valued mapping with locally closed graph containing $(\bar{x}, \bar{y})$. Then

\[
\text{sur} F(\bar{x}|\bar{y}) \geq \liminf_{\varepsilon \to 0} \{C^*(D^*_H F(x,y)) : y \in F(x), \|x - \bar{x}\| < \varepsilon, \|y - \bar{y}\| < \varepsilon\}
\]

\[
= \liminf_{\varepsilon \to 0} \{\|x^*\| : x^* \in D^*_H F(x,y)(y^*), \|y^*\| = 1, (x,y) \in (\text{Graph } F) \cap B((\bar{x}, \bar{y}), \varepsilon)\},
\]

or equivalently,

\[
\text{reg} F(\bar{x}|\bar{y}) = \text{lip}^{-1} F^{-1}(\bar{y}|\bar{x}) \leq \limsup_{\varepsilon \to 0} \{\|D^*_H F^{-1}(x,y)\|_+ : (x,y) \in (\text{Graph } F) \cap B((\bar{x}, \bar{y}), \varepsilon)\}
\]

\[
= \limsup_{\varepsilon \to 0} \{\|y^*\| : x^* \in D^*_H F(x,y)(y^*), \|y^*\| = 1, (x,y) \in (\text{Graph } F) \cap B((\bar{x}, \bar{y}), \varepsilon)\}.
\]

To furnish the proof we can use either any of the estimates of the preceding theorem or apply directly the slope-based results of Theorems 3.10 and 3.12 via (5.18). We choose the second option as it actually leads to a shorter proof. The first approach requires to work with weak* neighborhoods to estimate subdifferential of a sum of functions (that inevitably appears in the course of calculation) which makes estimating norms of subgradients difficult (if possible at all).

Proof. We only need to show that, given $(x,w) \in \text{Graph } F$, for any neighborhoods $U \subset X$ and $V \subset Y$ of $x$ and $y$

\[
\inf \{\|x^*\| : x^* \in D^* F(u,v)(y^*), (u,v) \in \text{Graph } F \cap (U \times V), \|y^*\| = 1\} \leq m.
\]

if $|\nabla \varphi_y|(x,w) < m$ for small $\xi$. Then the theorem is immediate from Theorem 3.10 in view of Proposition 5.18.

So let $|\nabla \varphi_y|(x,w) < m$. Take an $m' < m$ but still greater than $|\nabla \varphi_y|(x,v)$ and set

\[
f(u,v) = \varphi_y (u,v) + m' \max \{\|u - x\|, \|v - w\|\}
\]

\[
= \|v - w\| + i_{\text{Graph } F}(u,v) + m' \max \{\|u - x\|, \|v - w\|\}.
\]

Then $f$ attains a local minimum at $(x,w)$.

We thus can apply Proposition 5.9 given a $\delta > 0$, there are $v_i$, $i = 0, 1, 2, u_i$, $i = 1, 2$ with $(u_1, v_1) \in \text{Graph } F$ and $v_0 \in \partial \|\cdot\|(y - v_0)$, $(u_1^*, v_1^*) \in N(\text{Graph } F, (u_1, v_1))$ and $(u_2^*, v_2^*)$ with $\|u_2^*\| + \xi^{-1} \|v_2^*\| \leq m'$ such that

\[
\|v_i - w\| < \delta, \quad \|u_i - x\| < \delta, \quad \|u_1^* + u_2^*\| < \delta, \quad \|v_0^* + v_1^* + v_2^*\| < \delta.
\]

Take $\delta < \|y - w\|$, $(1 + 2\delta)m' < m$ and $\xi$ so small that $\xi m^2 < \delta$. Then $y \neq v_0$, so that $\|v_0^*\| = 1, \|x_1^*\| \leq m'$ and $\|v_2^*\| < \delta$. We thus have $\|x^*\| \leq m' + \delta < m$ and $||v_1^* - 1| < 1 + 2\delta$. It remains to set $y^* = v_1^*/||v_1^*||, x^* = x_1^*/||v_1^*||$ to complete the proof.

Theorem 5.22 (coderivative regularity estimate 2). If in addition to the assumptions of Theorem 5.21 both $X$ and $Y$ are Asplund spaces, then

\[
\text{sur} F(\bar{x}|\bar{y}) = \liminf_{\varepsilon \to 0} \{C^*(D^*_F F(x,y)) : y \in F(x), \|x - \bar{x}\| < \varepsilon, \|y - \bar{y}\| < \varepsilon\}
\]

\[
= \liminf_{\varepsilon \to 0} \{\|x^*\| : x^* \in D^*_F F(x,y)(y^*), \|y^*\| = 1, (x,y) \in (\text{Graph } F) \cap B((\bar{x}, \bar{y}), \varepsilon)\}.
\]
or equivalently,
\[
\text{reg} F(\overline{x} | \overline{y}) = \text{lip} F^{-1}(\overline{y} | \overline{x}) = \limsup_{\varepsilon \to 0} \{ \| D^*_F F^{-1}(x, y) \|_+ : (x, y) \in (\text{Graph } F) \cap B((\overline{x}, \overline{y}), \varepsilon) \} = \limsup_{\varepsilon \to 0} \{ \| y^* \| : x^* \in D^*_F F(x, y)(y^*), \| x^* \| = 1, (x, y) \in (\text{Graph } F) \cap B((\overline{x}, \overline{y}), \varepsilon) \}.
\]

**Proof.** If the spaces are Asplund, then the same arguments as in the proof of the preceding theorem lead to the same conclusion with $D^*_H$ replaced by $D^*_F$. So we have to show that the opposite inequality holds. This however is an elementary consequence of the definition. Indeed, fix certain $(x, y) \in \text{Graph } F$ close to $(\overline{x}, \overline{y})$ and let
\[
m = \inf \{ \| x^* \| : x^* \in D^*_F F(x, y)(y^*), \| y^* \| = 1 \}.
\]
If sur$D^*_F F(\overline{x} | \overline{y})\geq 0$ or $D^*_F F(x, y)(y^*) = \emptyset$ (in which case $m = \infty$ by the general convention), the inequality is trivial. So we can take a positive $r < \text{sur} \{ \| y^* \| : x^* \in D^*_F F(x, y)(y^*), \| x^* \| = 1, (x, y) \in (\text{Graph } F) \cap B((\overline{x}, \overline{y}), \varepsilon) \}$ and suppose that $m < \infty$. Take a $x^* \in D^*_F F(x, y)(y^*)$ with $\| y^* \| = 1$ and $\| x^* \| < m + \delta$ for some $\delta > 0$. Then $(x^*, h) - \langle y^*, v \rangle \leq o(\| h \| + \| v \|)$ whenever $(x + h, y + v) \in \text{Graph } F$. Now take $v(t) \in B(y, rt)$ such that $(y^*, v(t)) \leq -(1 - t^2)\| v(t) \|$ and an $h(t)$ with $\| h(t) \| \leq t$ such that $(x + th(t), y + v(t)) \in \text{Graph } F$. Then
\[
-t\| x^* \| + (1 - t^2)rt \leq (x^*, h(t)) - \langle y^*, v(t) \rangle \leq o(\| h(t) \| + \| v(t) \|) = o(t)
\]
which implies that $r \leq m$ and the result follows.  

**Remark 5.23.** Note that the just given proof (that the inequality $\leq$ holds) works in any space, not necessarily Asplund. In other words, the part of the theorem that incorporates essential properties of the space (that is that the Fréchet subdifferential is trusted) is contained in Theorem 5.21

Comparing the last theorem with Example 5.17, we conclude that in Asplund spaces the coderivative estimate using Fréchet coderivative can be strictly better than the tangential estimate provided by Theorem 5.15. What about connection of the estimates from Theorems 5.15 and 5.21?

**Proposition 5.24** (DH-coderivative vs. tangential criterion). The regularity estimate involving Dini-Hadamard coderivative (Theorem 5.21) is never worse than tangential estimate provided by Theorem 5.15.

**Proof.** Indeed, by definition $D^*_H F(x, y) = (DF(x, y))^*$ and we only need to recall that $C^*(D^*_H F(x, y)) \geq C(DF(x, y))$ for any $(x, y) \in \text{Graph } F$ by Theorem 5.21.

Theorem 5.21 was proved in [84] for subdifferentials satisfying a bit stronger requirements than the subdifferential of Dini-Hadamard. However a minor change in the proof allows to extend it to all subdifferentials trusted on the given Banach space (see e.g. [88, 114] also for a proof), in particular to the DH-subdifferential on any Gâteaux smooth space. Likewise, Theorem 5.22 was proved in [114], in a somewhat different form and in terms of
\( \varepsilon \)-Fréchet subdifferential on Fréchet smooth spaces. And again, a minor change is needed to extend the proof to standard Fréchet subdifferentials. Theorem 5.22 as stated was proved in [133] (see also [130] for a proof, for all Asplund spaces, not necessarily separable). This extension can be viewed as a consequence of the Fréchet smooth spaces version of the theorem and the separable reduction theorem of Fabian-Zhivkov [64] (and actually was proved that way). Proposition 5.24 seems to have never been mentioned earlier. It sounds rather surprising with all its simplicity. It would be interesting to find an example with a Dini-Hadamard coderivative estimate strictly better than the tangential estimate (or to prove that the estimates are equal). It is still unclear whether strict inequality is possible. The general consideration (the dual object cannot contain more information that its original predecessor) suggests that this is rather unlikely. But no proof is available for the moment. It should be mentioned however that the tangential estimate is valid in all Banach spaces while the Dini-Hadamard coderivative makes sense basically in Gâteaux smooth spaces.

5.4.2 Perfect regularity and linear perturbations

The main inconvenience of the regularity criteria that have been just established, no matter primal or dual, comes from the necessity to scan an entire neighborhood of the point of interest. Below we define what can be viewed as an ideal situation.

**Definition 5.25.** We shall say that \( F \) is perfectly regular at \((\bar{x}, \bar{y}) \in \text{Graph } F\) if

\[
\text{sur } F(x | y) = C^* (D^*_G F(\bar{x}, \bar{y})) = \min \{ \| x^* \| : x^* \in D^*_G F(\bar{x}, \bar{y})(y^*), \| y^* \| = 1 \}. \tag{5.12}
\]

Later we shall come across some classes of perfectly regular mappings and meanwhile consider an important class of additive linear perturbations of maps.

**Definition 5.26.** Given a set-valued mapping \( F : X \rightrightarrows Y \) and an \((\bar{x}, \bar{y}) \in \text{Graph } F\). The radius of regularity of \( F \) at \((\bar{x}, \bar{y})\) is the lower bound of norms of linear continuous operators \( A : X \to Y \) such that \( \text{sur}(F + A)(\bar{x}, \bar{y} + A\bar{x}) = 0 \). We shall denote it \( \text{rad } F(\bar{x}, \bar{y}) \).

By Milyutin’s theorem \( \text{sur } F(\bar{x}, \bar{y}) \leq \text{rad } F(\bar{x}, \bar{y}) \). It turns out that for perfectly regular mappings the equality holds. To show this we need the following proposition, not very difficult to prove.

**Proposition 5.27.** Let \( X \) and \( Y \) be normed spaces, let \( F : X \rightrightarrows Y \) be set-valued mapping with closed graph, and let \( A \in \mathcal{L}(X,Y) \). Assume that \( F \) is regular at \((\bar{x}, \bar{y}) \in \text{Graph } F\) and set \( G = F + A \) (that is \( G(x) = F(x) + Ax \)). Then

\[
D^*_G ( \bar{x}, \bar{y} ) = D^*_G ( \bar{x}, \bar{y} ) + A^*
\]

Note that the equality is elementary in case of Dini-Hadamard or Fréchet subdifferentials.

**Theorem 5.28** (perfect regularity and radius formula). Assume that \( X \) and \( Y \) are Banach spaces, \( F : X \rightrightarrows Y \), \((\bar{x}, \bar{y}) \in \text{Graph } F\) and \( F + A \) is perfectly regular at \((\bar{x}, \bar{y} + A\bar{x})\) for any \( A \in \mathcal{L}(X,Y) \) of rank 1. Then

\[
\text{sur } F(\bar{x}, \bar{y}) = \text{rad } F(\bar{x}, \bar{y}). \tag{5.13}
\]
Moreover, for any \( \varepsilon > 0 \) there is a linear operator \( A_\varepsilon \) of rank one such that \( \| A_\varepsilon \| \leq \text{sur}(\overline{x} \mid y) + \varepsilon \) and \( \text{sur}(F + A)(\overline{x}, \overline{y} + A\overline{x}) = 0 \).

In the sequel we call (5.13) the radius formula.

**Proof.** Set \( r = \text{sur}(\overline{x} \mid y) \). The theorem is obviously valid if \( r = 0 \). So we assume that \( r > 0 \). Take an \( \varepsilon > 0 \) and find a \( y_\varepsilon^* \) and an \( x_\varepsilon^* \in D_{\overline{y}}(\overline{x}, y_\varepsilon^*)(y_\varepsilon^*) \) such that \( \| y_\varepsilon^* \| = 1 \), \( \| x_\varepsilon^* \| \leq (1 + \varepsilon)r \). Let further \( x_\varepsilon \in X \) and \( y_\varepsilon \in Y \) satisfy

\[
\| x_\varepsilon \| = \| y_\varepsilon \| = 1, \quad \langle x_\varepsilon^*, x \rangle \geq (1 - \varepsilon)\| x_\varepsilon^* \|. \quad \langle y_\varepsilon^*, y \rangle \geq (1 - \varepsilon).
\]

(5.14)

We use these four vectors to define an operator \( A_\varepsilon : X \rightarrow Y \) as follows:

\[
A_\varepsilon x = \frac{\langle x_\varepsilon^*, x \rangle}{\langle y_\varepsilon^*, y \rangle} y_\varepsilon.
\]

Then \( \| A_\varepsilon \| \leq \frac{1 + \varepsilon}{1 - \varepsilon}r \) and

\[
A_\varepsilon y_\varepsilon^* = \frac{\langle y_\varepsilon^*, y_\varepsilon \rangle}{\langle y_\varepsilon^*, y \rangle} x_\varepsilon^*.
\]

In particular we see that \( -x_\varepsilon^* = A_\varepsilon^* y_\varepsilon^* \). Combining this with Proposition 5.27 we get \( 0 = x_\varepsilon^* - A_\varepsilon^* y_\varepsilon^* \in D_{\overline{y}}(F + A)(\overline{x}, y_\varepsilon^*)(y_\varepsilon^*) \) and therefore by the prefect regularity assumption, \( \text{sur}(F + A)(\overline{x} \mid y) + A\overline{x}) = 0 \), that is \( \text{rad}(\overline{x}, y) \leq \| A_\varepsilon \| \rightarrow r \) as \( \varepsilon \rightarrow 0 \).

Let \( S(y, A) \) be the set of solutions of the inclusion

\[
y \in F(x) + Ax,
\]

(5.15)

where \( A \in \mathcal{L}(X, Y) \). Let \( \overline{x} \) be a nominal solution of (5.15) with \( y = \overline{y}, \ A = \overline{A} \). The question we are going to consider concerns Lipschitz stability of \( S \) with respect to small variations of both \( y \) and \( A \) around the nominal value \( (\overline{y}, \overline{A}) \) and their effect on regularity rates.

In other words, we are interested in finding \( \text{lip}|S((\overline{y}, \overline{A})|\overline{x}) \). By the equivalence theorem, this is the same as finding the modulus of surjection of the mapping \( \Phi = S^{-1} \) at \( (\overline{x}, (\overline{y}, \overline{A})) \).

Clearly

\[
\Phi(x) = \{(y, A) \in Y \times \mathcal{L}(X, Y) : y \in F(x) + Ax(x)\}.
\]

We shall consider \( Y \times \mathcal{L}(X, Y) \) with the norm \( \|(y, A)\| = \nu(||y||, ||A||) \), where \( \nu \) is a norm in \( \mathbb{R}^2 \). The dual norm is \( \nu^*(||y^*||, ||\ell||) \), where \( \ell \in (\mathcal{L}(X \times Y))^* \) and \( \nu^* \) is the norm in \( \mathbb{R}^2 \) dual to \( \nu \). \( \nu^*(u) = \sup\{\alpha \xi + \beta \eta : \nu(\alpha, \beta) \leq 1\} \). As to the space dual to \( \mathcal{L}(X, Y) \), we only need the simplest elements of the space, rank one tensors \( y^* \otimes x \) whose action on \( A \in \mathcal{L}(X, Y) \) is defined by \( \langle y^* \otimes x, A \rangle = \langle A^* y^*, x \rangle \) and whose norm is \( ||y^* \otimes x|| = ||y^*|| ||x|| \).

The following theorem gives an answer to the question.

**Theorem 5.29 (95).** Let \( X \) and \( Y \) be Banach spaces, and let \( F : X \rightrightarrows Y \) be a set-valued mapping with closed graph. Let \( (\overline{x}, \overline{y}) \in \text{Graph} \ F \) and let \( \overline{A} \in \mathcal{L}(X, Y) \) be given. Then

\[
\text{lip}|S((\overline{y}, \overline{A})|\overline{x}) \leq \nu^*(1, ||\overline{x}||)\text{reg}(F + \overline{A})(\overline{x} \mid y).
\]
To prove the theorem we only need to show that
\[
\text{sur} \Phi(\overline{y}, A) \geq \frac{1}{\nu^*(1, \|x\|)} \text{sur}(F + A)(x, y).
\] (5.16)

So the proof (involving some calculation) can be obtained either from Theorem 4.5 or directly from the general regularity criterion of Theorem 3.1.

The concepts of perfect regularity and radius of regularity were introduced respectively in [102] and [52]. Theorem 5.28 is a new result. A finite dimensional version of Theorem 5.29 for a class of \(F\) with convex graph was proved in [30]. We shall discuss the problems considered in this subsection in more details for finite dimensional mappings later in Section 8.

6 Finite dimensional theory.

In this section we concentrate on characterizations of regularity, subregularity and transversality for set-valued mappings between finite dimensional spaces. There are several basic differences that make the finite dimensional case especially rich. The first is that the subdifferential calculus is much more efficient. In addition certain properties different in the general case appear to be identical in \(\mathbb{R}^n\). In particular, for a lower semicontinuous function the Dini-Hadamard subdifferential and the Fréchet subdifferential are identical. Therefore the usual notation used in the literature for this common subdifferential is \(\partial\) rather than \(\partial_H\) or \(\partial_F\). Likewise, as the limiting Fréchet and the \(G\)-subdifferentials are also equal, it is convenient to speak simply about limiting subdifferential and denote it simply by \(\partial\).

The second circumstance to be mentioned is the abundance of some special classes of objects of practical importance and definite theoretical interest. Enough to mention polyhedral and semi-algebraic sets and mappings (to be considered in the second part of the paper), semi-smooth functions, prox-regular functions and sets etc.. We do not discuss some interesting and important subjects, e.g. Kummer’s inverse function theorem and its applications (well presented in the literature: much on the subjects can be found in [55, 109]) or semismooth mappings (see e.g. [68]).

6.1 Regularity.

Theorem 6.1. A set-valued mapping \(F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m\) with locally closed graph is perfectly regular near any point of its graph.

Proof. This is immediate from Theorem 5.22.

Theorem 6.2. The radius formula holds at any point of the graph of a set-valued mapping \(F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m\) with locally closed graph. Moreover, the lower bound in the definition of the radius of regularity is attained at a linear operator \(A : \mathbb{R}^n \rightarrow \mathbb{R}^m\) of rank one.

Proof. This is immediate from Theorem 5.28.
Theorem 6.3. Let \( F : \mathbb{R}^n \rightarrow \mathbb{R}^m \) be a set-valued mapping with locally closed graph, and let \((\bar{x}, \bar{y}) \in \text{Graph } F\). Then

\[
\text{surF}(\bar{x}, \bar{y}) = \lim_{\varepsilon \to 0} \text{inf} \{ C(DF(x, y)) : (x, y) \in (\text{Graph } F) \cap B((\bar{x}, \bar{y}), \varepsilon) \}. \quad (6.1)
\]

Proof. In view of Theorem 5.15, it is enough to verify that \( C(DF(x, y)) \geq r \) if \( B(y, tr) \subset F(B(x, t)) \) for all sufficiently small \( t \) (of course for \((x, y) \in \text{Graph } F\)). So take a \( v \in \mathbb{R}^m \) with \( \|v\| \leq r \) and let \( h(t) \) be such that \( \|h(t)\| \leq 1 \) and \( y + tv \in F(x + th(t)) \). If now \( h \) is any limiting point of \( h(t) \) as \( t \to 0 \), then \( v \in DF(x, y)(h) \). This shows that \( rB_{\mathbb{R}^m} \subset DF(x, y)(B_{\mathbb{R}^n}) \).

Similarly, inequality can be replaced by equality in the estimate of Lipschitz stability of solutions of the inclusion

\[
y \in F(x) + Ax \quad (6.2)
\]

with both \( y \) and \( A \) viewed as perturbations (cf. Theorem 5.29). But first we have to do some preliminary job. As in 5.4.2 we denote by \( S(y, A) \) the set of solutions of \((6.2)\) and by \( \Phi \) the inverse mapping

\[
\Phi(x) = \{(y, A) : y \in F(x) + Ax\}.
\]

Lemma 6.4. For any \( x \in X \), let \( E(x) : Y \times \mathcal{L}(X, Y) \rightarrow Y \) be the linear operator defined by \( E(y, \Lambda) = y - \Lambda x \). Then, under the assumptions of Theorem 5.27

\[
\nu(1, \|x\|) C(E(x) \circ D\Phi(x, (y, A))) \leq C(D(F + A)(x, y)),
\]

whenever \( y \in F(x) + Ax \).

Proof. By definition \((h, v, \Lambda) \in X \times Y \times \mathcal{L}(X, Y)\) belongs to \( T(\text{Graph } \Phi, (x, y, A)) \) if there are sequences \((h_n) \to h\), \((v_n) \to v\), \((\Lambda_n) \to \Lambda\) and \((t_n) \to +0\) such that

\[
y + t_n v_n - (A + t_n \Lambda_n)(x + t_n h_n) \in F(x + t_n h_n)
\]

or

\[
y + t_n (v_n - \Lambda_n x + t_n \Lambda_n h_n) \in (F + A)(x + t_n h_n).
\]

As \( t_n \|\Lambda_n h_n\| \to 0 \), it follows that

\[
T(\text{Graph } \Phi, (x, y, A)) = \{(h, v, \Lambda) : (h, v - \Lambda x) \in T(\text{Graph } (F + A), (x, y))\}
\]

which amounts to

\[
E(x) \circ D\Phi(x, (y, A)) = D(F + A)(x, y). \quad (6.3)
\]

We have (Corollary 5.3) \( C(E(x)) \cdot C(D\Phi(x, (y, A))) \leq C(D(F + A)(x, y)) \). On the other hand \( E(x)^*(y^*) = (y^*, -y^* \otimes x) \) and therefore (Proposition 1.3)

\[
C(E(x)) = \inf_{\|y^*\|=1} \|E(x)^* y^*\| = \nu(1, \|x\|).
\]

This completes the proof of the lemma.
**Theorem 6.5** (linear perturbations - finite dimensional case). Let \( F : \mathbb{R}^n \Rightarrow \mathbb{R}^m \) be a set-valued mapping with locally closed graph, and let \( \overline{y} \in F(\overline{x}) \). We consider \( \mathbb{R}^m \times \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m) \) with the norm \( \nu(\|y\|, \|A\|) \), where \( \nu \) is a certain norm in \( \mathbb{R}^2 \). Then, given an \( A \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m) \), we have

\[
\text{lip} \left( ((\overline{y}, A) | | \overline{x} \rangle \right) = \nu^\ast(1, | | \overline{x} \rangle) \text{reg}(F + A)(\overline{x} | \overline{y})
\]

**Proof.** Immediate from the lemma and Theorem 5.29.

Finally, we have to mention that a continuous single-valued mapping \( f : \mathbb{R}^n \to \mathbb{R}^m \) can be strongly regular only if \( m = n \). This is a simple consequence of Brouwer’s invariance of domain theorem (see e.g. [109]).

Theorem 6.1 was announced by Mordukhovich in a somewhat different form [128] (see also [129]). But the lower estimate for the modulus of surjection (which is actually the major step in the proof) is immediate from Ioffe [83]. Theorem 6.2 was proved by Dontchev-Lewis-Rockafellar in [52] and Theorem 6.3 by Dontchev-Quincampoix-Zlateva [53]. Theorem 6.5 is a slightly generalized version of already mentioned result of Cándanos, Gómez and Senent-Parra [30].

### 6.2 Subregularity and error bounds.

Let \( f \) be an extended-real-valued lsc function on \( \mathbb{R}^n \). We can associate with this function the epigraphic map

\[
\text{Epi} f(x) = \{ \alpha \in \mathbb{R} \mid \alpha \geq f(x) \}
\]

Subregularity of such a mapping at a point \((\overline{x}, \alpha)\) (if \( \alpha = f(\overline{x}) \) is finite) means that there is a \( K > 0 \) such that

\[
d(x, [f \leq \alpha]) \leq K(F(x) - \alpha)^+
\]

for all \( x \) close to \( \overline{x} \). The constant \( K \) in this case is usually called a local error bound for \( f \) at \( x \). We shall say more about error bounds in the second part of the paper.

To characterize the subregularity property of epigraphic maps we define outer limiting subdifferential of \( f \) at \( x \) as follows:

\[
\partial^> f(x) = \left\{ \lim_{k \to \infty} x_k^* : \exists x_k \to x, f(x_k) > f(x), x_k^* \in \hat{\partial} f(x_k) \right\}
\]

**Theorem 6.6** (error bounds in \( \mathbb{R}^n \)). Let \( f \) be a lower semicontinuous function on \( \mathbb{R}^n \) that is finite at \( \overline{x} \). Then \( K > 0 \) is a local error bound of \( f \) at \( \overline{x} \) if either of the following two equivalent conditions is satisfied:

(a) \( K \cdot \lim_{\varepsilon \to 0} \inf \{ |\nabla f|(x) : \|x - \overline{x}\| < \varepsilon, f(\overline{x}) < f(x) < f(\overline{x}) + K\varepsilon \} \geq 1 \);

(b) \( K \cdot d(0, \partial^> f(\overline{x})) \geq 1 \).

Thus, if \( F : \mathbb{R}^n \Rightarrow \mathbb{R}^m \) has locally closed graph and \((\overline{x}, \overline{y}) \in \text{Graph} \ F\), then

\[
\text{subreg} F(\overline{x} | \overline{y}) \leq \left[ \inf \{ \|x^*\| : x^* \in \partial^> d(\overline{x}, F(\cdot))(\overline{x}) \} \right]^{-1}.
\]
Proof. If (a) holds, then $K$ is a local error bound by Lemma 7.4 to be proved in the next section. To prove that (a)$\Rightarrow$(b), let $x^* \in \partial^\circ f(\bar{x})$. This means that there are sequences $(x_k)$ and $(x_k^*)$ such that $x_k \to f \bar{x}$, $f(x_k) > f(\bar{x})$, $x_k^* \to x^*$ and $x_k^* \in \partial f(x_k)$. Choose $\varepsilon_k \downarrow 0$ such that $\|x_k-x\| < \varepsilon_k$ and $f(x_k) - f(\bar{x}) < K\varepsilon_k$. If (a) holds, then $K \liminf |\nabla f|(x_k) \geq 1$. But $\|x_k\| \geq \nabla f(x_k)$ (Proposition 5.18) and (b) follows.

The opposite implication (b)$\Rightarrow$(a) also follows from Proposition 5.18 indeed, denote by $r$ the value of the limit in the left side of (a), take an $\varepsilon > 0$ and let $x$ satisfy the bracketed inequalities in (a) along with $\|\nabla f|(x) < r + \varepsilon$. This means that $f + (r+\varepsilon)\cdot-x$ Applying the fuzzy variational principle, we shall find $u$ and $u^* \in \partial_F(u)$ such that $\|u-x\| < \varepsilon$, $f(u) < f(\bar{x}) + \varepsilon/K$ and $\|u^*\| < r + 2\varepsilon$. This means that there is a sequence of pairs $(x_k, x_k^*)$ such that $x_k \to f \bar{x}$, $x_k^* \in \partial_F f(x_k)$ and $\limsup \|x_k\| \leq r$. As (b) holds, it follows that $Kr \geq 1$.

Conditions (a) and (b) are not necessary for $K$ to be an error bound of $f$ at $\bar{x}$.

Example 6.7. Consider

$$f(x) = \begin{cases} 0, & \text{if } x \leq 0; \\ x + x^2 \sin x^{-1}, & \text{if } x > 0. \end{cases}$$

It is an easy matter to see that any $K > 1$ is an error bound for $f$ at zero but at the same time $0 \in \partial^2 f(0)$.

Such a pathological situation, however, does not occur if the function is "not too nonconvex" near $\bar{x}$.

Proposition 6.8. Let $f$ be a lower semicontinuous function on $\mathbb{R}^n$ finite at $\bar{x}$. Suppose there are a $\theta > 0$ and a function $r(t) = o(t)$ such that

$$f(u) - f(x) \geq (x^*, u-x) - r(\|u-x\|)$$

for all $x$, $u$ of a neighborhood of $\bar{x}$, provided $f(\bar{x}) < f(x) < f(\bar{x}) + \theta$ and $x^* \in \hat{\partial}(x)$. If under these conditions, $K > 0$ is an error bound of $f$ at $\bar{x}$, then the conditions (a) and (b) of Theorem 6.6 hold.

Proof. Assume the contrary. Then there are $\varepsilon > 0$ and a sequence of pairs $(x_k, x_k^*) \in \hat{\partial}f(x_k))$ such that $x_k \to f \bar{x}$, $f(x_k) > f(\bar{x})$ and $\|x_k\| \leq K^{-1} - \varepsilon$. For any $k$ take an $\bar{x}_k \in [f \leq f(\bar{x})]$ closest to $x_k$. Then $\bar{x}_k \to f(\bar{x})$ and by the assumption

$$f(\bar{x}_k) - f(x_k) \geq (x_k^*, \bar{x}_k - x_k) - r(\|\bar{x}_k - x_k\|).$$

As $\|\bar{x}_k - x_k\| \to 0$, for large $k$ we have $r(\|\bar{x}_k - x_k\|) \leq (\varepsilon/2)\|\bar{x}_k - x_k\|$. For such $k$

$$f(x_k) \leq f(\bar{x}_k) + (\|x_k^*\| + (\varepsilon/2))\|\bar{x}_k - x_k\|.$$

It follows that

$$d(x_k, [f \leq f(\bar{x})]) = \|\bar{x}_k - x_k\| \geq \frac{1}{\|x_k^*\| + (\varepsilon/2)} f(x_k),$$

that is $(K^{-1} - (\varepsilon/2))d(x_k, [f \leq f(\bar{x})]) \geq f(x_k)$ contrary to the assumption. \qed
The last result of this subsection contains infinitesimal characterization of strong subregularity.

**Theorem 6.9** (characterization of subregularity and strong subregularity). Let again $F : \mathbb{R}^n \Rightarrow \mathbb{R}^m$ have locally closed graph and $(\bar{x}, \bar{y}) \in \text{Graph} F$. Then

- $F$ is subregular at $(\bar{x}, \bar{y})$ if $d(0, \partial^2 \psi_{\bar{y}}(\bar{y})) > 0$;
- a necessary and sufficient condition for $F$ to be strongly subregular at $(\bar{x}, \bar{y})$ is that $DF(\bar{x}, \bar{y})$ is nonsingular, that is $C^*(DF(\bar{x}, \bar{y})) > 0$.

**Proof.** The first statement is a consequence of Theorem 6.6. To prove the second, assume first that $F$ is strongly subregular at $(\bar{x}, \bar{y})$, that is there is a $K > 0$ such that $\|x - \bar{x}\| \leq Kd(\bar{y}, F(x))$ for $x$ sufficiently close to $\bar{x}$. If $DF(\bar{x}, \bar{y})$ were singular, Proposition 5.2 would guarantee the existence of sequences $(h_k) \subset \mathbb{R}^n$ and $(v_k) \subset \mathbb{R}^m$ such that $\|h_k\| = 1$, $\|v_k\| \to 0$ and $\bar{y} + t_k v_k \in F(\bar{x} + t_k h_k)$, so that for large $k$

$$\|\bar{x} + t_k h_k - \bar{y}\| = t_k > K t_k \|v_k\| = K \|\bar{y} + t_k v_k - \bar{y}\| \geq K d(\bar{y}, F(\bar{x} + t_k h_k)),$$

contrary to our assumption.

Let now $DF(\bar{x}, \bar{y})$ be nonsingular. This means that $\|v\| \geq \kappa > 0$ whenever $v \in DF(\bar{x}, \bar{y})(h)$ with $\|h\| = 1$. It immediately follows that, say, $\|y - \bar{y}\| \geq (\kappa/2)\|x - \bar{x}\|$ whenever $y \in F(x)$ and $x$ is sufficiently close to $\bar{x}$ which is strong subregularity of $F$ at $(\bar{x}, \bar{y})$. 



Literature on local error bounds in $\mathbb{R}^n$ is very rich - see e.g. the monograph by Facchinei and Pang [65] that summarizes developments prior to 2003. Theorem 6.6 and Proposition 6.8 seem to be new as stated but they are closely connected with the results of Ioffe-Outrata [100] and Meng and Yang [127] among others. The second part of Theorem 6.9 as well as other results relating to strong subregularity and applications can be found in [55] and [109]. (In [109] the authors use the term ”locally upper Lipschitz” property. The term ”strong subregularity” seem to have appeared later.) Another sufficient condition for subregularity was suggested by Gfrerer [71]. It would be interesting to understand how the two are connected. It should also be noted that no characterization for strong subregularity in terms of coderivatives is so far known.

6.3 Transversality.

We have mentioned already that the classical concepts of transversality and regularity are closely connected. To see how the concept of transversality can be interpreted in the context of variational analysis, we first consider the case of two intersecting manifolds in a Banach space.

Let $X$ be a Banach space and $M_1$ and $M_2$ smooth manifolds in $X$, both containing some $\bar{x}$. As was mentioned in Subsection 1.4, the manifolds are transversal at $\bar{x}$ if either $\bar{x} \notin M_1 \cap M_2$ or the sum of the tangent subspaces to the manifolds at $\bar{x}$ is the whole of $X$: $T_{\bar{x}} M_1 + T_{\bar{x}} M_2 = X$. The following simple lemma is the key to interpret this in regularity terms in a way suitable for extensions to the settings of variational analysis.

**Lemma 6.10.** Let $L_1$ and $L_2$ be closed subspaces of a Banach space $X$ such that $L_1 + L_2 = X$. Then for any $u, v \in X$ there is $h \in X$ such that $u + h \in L_1$ and $v + h \in L_2$. 

48
Proof. If \( u + h \in L_1 \), then \( h \in -u + L_1 \), so if the statement were wrong, we would have \( (v - u + L_1) \cap L_2 = \emptyset \). In this case there is a nonzero \( x^* \) separating \( v - u + L_1 \) and \( L_2 \), that is such that \( \langle x^*, x \rangle = 0 \) for all \( x \in L_2 \) and \( \langle x^*, v - u + x \rangle \geq 0 \) for all \( x \in L_1 \). But this means that \( x^* \) vanishes on \( L_1 \) as well. In other words, both \( L_1 \) and \( L_2 \) belong to the annihilator of \( x^* \) and so their sum cannot be the whole of \( X \).

The lemma effectively says that the linear mapping \( (u, v, h) \mapsto (u + h, v + h) \) maps \( L_1 \times L_2 \times X \) onto \( X \times X \), that is this mapping is regular. If \( \varpi \in M_1 \cap M_2 \), then applying the density theorem (Theorem \( \text{3.5} \)), we get as an immediate corollary that the set-valued mapping \( \Phi(x) = (M_1 - x) \times (M_2 - x) \) from \( X \) into \( X \times X \) is regular at zero. This justifies the following definition

**Definition 6.11.** Let \( S_i \subset X \), \( i = 1, \ldots, k \) be closed subsets of \( X \). We say that \( S_i \) are transversal at \( \varpi \in X \) if either \( \varpi \not\in \cap S_i \) or \( \varpi \in \cap S_i \) and the set-valued mapping

\[
x \mapsto F(x) = (S_1 - x) \times \cdots \times (S_k - x)
\]

is regular near \( (\varpi, 0, \ldots, 0) \). In the latter case, we also say that \( S_i \) have transversal intersection at \( \varpi \).

This definition may look strange at the first glance but the following characterization theorem shows that it is fairly natural.

**Theorem 6.12.** Let \( S_i \subset \mathbb{R}^n \), \( i = 1, \ldots, k \) and \( \varpi \in \cap S_i \). Then the following statements are equivalent

(a) \( S_i \) are transversal at \( \varpi \);

(b) \( x_i^* \in N(S_i, \varpi) \), \( x_1^* + \cdots + x_k^* = 0 \Rightarrow x_1^* = \cdots = x_k^* = 0 \);

(c) \( d(x, \bigcap_{i=1}^k (S_i - x_i)) \leq K \max_i d(x, S_i - x_i) \) if \( x_i \) are close to zero and \( x \) is close to \( \varpi \).

Proof. It is not a difficult matter to compute the limiting coderivative of \( F \): if \( (x_1, \ldots, x_k) \in F(x) \), then

\[
D^*F(x)(x_1, \ldots, x_k) = \begin{cases}
    \sum_{i=1}^k x_i^*, & \text{if } x_i^* \in N(S_i, x_i + x); \\
    \emptyset, & \text{otherwise}.
\end{cases}
\]

Combining this with Theorem \( \text{6.1} \), we prove equivalence (a) and (b).

Furthermore, \( F^{-1}(x_1, \ldots, x_k) = (S_1 - x_1) \cap \cdots \cap (S_k - x_k) \), whence equivalence of (a) and (c).

Note that implicit in (c) is the statement that the intersection of \( S_i - x_i \) is nonempty if \( x_i \) are sufficiently small. In case of two sets one more convenient characterization of transversality is available.

**Corollary 6.13.** Two sets \( S_1 \) and \( S_2 \) both containing \( \varpi \) are transversal at \( \varpi \) if and only if the set-valued mapping \( \Phi : \mathbb{R}^n \times \mathbb{R}^n \Rightarrow \mathbb{R}^n : \)

\[
\Phi(x_1, x_2) = \begin{cases}
    x_1 - x_2, & \text{if } x_i \in S_i; \\
    \emptyset, & \text{otherwise}
\end{cases}
\]
is regular near \((\overline{\tau}, \overline{\tau}, 0)\).

**Proof.** We have \(T(\text{Graph } \Phi, ((x_1, x_2), x_1-x_2)) = \{(h_1, h_2, v) : h_i \in T(S_i, x_i), v = h_1 - h_2\}\), so that
\[
D^*\Phi(\overline{\tau}, \overline{\tau}, 0)(x^*) = \{(x_1^*, x_2^*) : x_i^* \in N(S_i, \overline{\tau} + x^*)\}.
\]
If we consider the max-norm \(\|(x_1, x_2)\| = \max\{\|x_1\|, \|x_2\|\}\) in \(\mathbb{R}^n \times \mathbb{R}^m\), then it follows from Theorem 6.1 that \(\Phi\) is regular near \((\overline{\tau}, \overline{\tau}, 0)\) if and only if
\[
\inf\{|x_1^*| + \|x_2^* + x^*\| : x_i^* \in N(S_i, x_0), \|x^*\| = 1\} > 0.
\]
This amounts to \(N(S_1, \overline{\tau}) \cap (-N(S_2, \overline{\tau})) = \{0\}\), which is exactly the property in the part (b) of the theorem. \(\square\)

In view of the equivalence between (a) and (c) in Theorem 6.12, the following definition looks now very natural.

**Definition 6.14** (subtransversality). We shall say that closed sets \(S_1, \ldots, S_k\) are **subtransversal at** \(\overline{\tau} \in \cap S_i\) if there is a \(K > 0\) such that for any \(x\) close to \(\overline{\tau}\)
\[
d(x, \bigcap_{i=1}^k S_i) \leq K \sum_{i=1}^k d(x, S_i).
\]

In a similar way, it is easy to see that subtransversality is equivalent to subregularity of the same mapping \(F\) and to get a sufficient subtransversality condition from Theorem 6.6.

In the next section we shall be able to see the key role subtransversality plays in some problems of optimization and subdifferential calculus.

We conclude with a brief discussion of transversality of a mapping and a set.

**Theorem 6.15.** Let \(F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m\) have locally closed graph, and let \(S \subset \mathbb{R}^m\) be closed. Assume that \(\overline{\gamma} \in F(\overline{\tau}) \cap S\). Then the following statements are equivalent:

(a) the set-valued mapping \(\Phi : (x, y) \mapsto (F(x)-y) \times (S-y)\) is regular near \(((\bar{x}, \bar{y}), (0, 0))\);

(b) the sets Graph \(F\) and \(\mathbb{R}^n \times S\) have transversal intersection near \((\bar{x}, \bar{y})\);

(c) \(0 \in D^*F(\bar{x}, \bar{y})(y^*) \& y^* \in N(S, \bar{y}) \Rightarrow y^* = 0\).

**Proof.** Equivalence of (b) and (c) follows from Theorem 6.12. To prove that (a) and (b) are equivalent, set \(\Psi(x, y) = (\text{Graph } F - (x, y) \times (\mathbb{R}^n \times S - (x, y)))\). If \(((\xi, \mu), (\eta, \nu)) \in \Psi(x, y),\) then \((\mu, \nu) \in \Phi(u, y)\) with \(u = \xi + x\). Conversely, if \((\mu, \nu) \in \Phi(u, y),\) then \((u, \mu + y) \in \text{Graph } F\) and \((w, \nu + y) \in \mathbb{R}^n \times S\) for any \(w \in \mathbb{R}^n\). Then for any \(x\), we have, setting \(\xi = u - x, \eta = w - x\), that \(((\xi, \mu), (\eta, \nu)) \in \Psi(x, y).\)

(b) \(\Rightarrow\) (a). If (b) holds, then \(\Psi\) is regular near \(((\bar{x}, \bar{y}), (0, 0), (0, 0))\). So let \(((\xi, \mu), (\eta, \nu)) \in \Psi(x, y)\) with \((x, y)\) sufficiently close to \((\bar{x}, \bar{y})\) and \(\xi, \mu, \eta, \nu\) sufficiently close to zeros of the corresponding spaces. Take a small \(t > 0\) and let \(\|\xi' - \xi\| < t\) etc. Then by (b) there is a \(K > 0\) and \((x', y')\) such that \(\|x' - x\| \leq Kt, \|y' - y\| \leq Kt\) and \(((\xi', \mu'), (\eta', \nu')) \in \Psi(x', y').\)

We have
\[
\xi' = u' - x', \quad \mu' \in F(u') - y', \quad \eta' = w' - x', \quad \nu' \in S - y'
\]
for some \((u', v') \in \text{Graph } F\) and \(w' \in \mathbb{R}^n\). We have therefore \(\|u' - u\| \leq \|x' - x\| + \|\xi' - \xi\| \leq (K + 1)t\).
Thus, whenever \((\mu, \nu) \in \Phi(u, y)\) with \((u, y)\) close to \((\bar{x}, \bar{y})\) and \((\mu, \nu)\) close to \((0, 0)\) and \(t > 0\) is sufficiently small, for any \(\mu', \nu' \in \mathbb{R}^m\) that differ from \(\mu, \nu\) at most by \(t\), there is a pair \((u', y')\) within \((K + 1)t\) of \((u, y)\) such that \(\mu' \in F(u') - y'\) and \(\nu' \in S - y'\), that is (a).

(a) \Rightarrow (b). Here the arguments are similar, actually even a bit shorter. Let \(((\xi, \mu), (\eta, \nu)) \in \Psi(x, y)\) with \((x, y)\) close to \((\bar{x}, \bar{y})\) and \(((\xi, \mu), (\eta, \nu))\) close to \(((0, 0), (0, 0))\). Then as we have seen, \((\mu, \nu) \in \Phi(u, y)\) with \(u = \xi + x\), also close to \(\overline{x}\). Let further \(||\mu' - \mu|| < t, ||\nu' - \nu|| < t\). If \(t\) is sufficiently small, then by (a) we can find \(u', y'\) such that \(||u' - u|| \leq Kt, ||y' - y|| \leq Kt\) with some positive \(K\) such that \((\mu', \nu') \in \Phi(u, y)\). Take \(x' = x, \xi' = \overline{x} - x, \eta' = \eta\). Then as is immediate from what was explained in the first paragraph of the proof \(((\xi, \mu'), (\eta', \nu')) \in \Psi(x', y')\). Thus \(\Psi\) is regular near \(((\bar{x}, \bar{y}), ((0, 0), (0, 0)))\).

The proposition justifies the following definition.

**Definition 6.16.** Let \(F : \mathbb{R}^n \Rightarrow \mathbb{R}^m\) have locally closed graph, let \(S \subset \mathbb{R}^m\) be a closed set, and let \((\bar{x}, \bar{y}) \in \text{Graph} F\). We say that \(F\) is **transversal to** \(S\) at \((\bar{x}, \bar{y})\) if either \(\overline{y} \notin S\) or \(\overline{y} \in S\) and \(\text{Graph} F\) and \(\mathbb{R}^n \times S\) are transversal at \((\bar{x}, \bar{y})\). We say that \(F\) is **transversal to** \(S\) if it is transversal to \(S\) at any point of the graph.

Likewise, if \(\overline{y} \in F(\overline{x}) \cap S\), we shall say that \(F\) is **subtransversal to** \(S\) and \((\bar{x}, \bar{y})\), provided

\[
d((x, y), \text{Graph } F \cap (X \times S)) \leq Kd((x, y), \text{Graph } F) + d(y, S)
\]

for \((x, y)\) of a neighborhood of \((\bar{x}, \bar{y})\).

It is almost obvious from (a) that in case \(\overline{y} \in F(\overline{x}) \cap S\), transversality of \(F\) to \(S\) at \((\bar{x}, \bar{y})\) implies regularity of the mapping \(x \mapsto F(x) - S\) near \((\overline{x}, 0)\). Without going into technical details the explanation is as follows. Suppose we wish to find an \(x\) such that \(z \in F(x) - S\). By (a) there are some \((x, y)\) such that \((0, z) \in \text{Graph } F - (x, y)\) and \((0, 0) \in \mathbb{R}^n \times S - (x, y)\). This means that \(z \in F(x) - y\), on the one hand, and \(y \in S\), on the other hand, as required.

The converse however does not seem to be valid at least for a set-valued \(F\). The situation here is similar to that considered in Example 4.7. However there the converse is also true in one important case.

**Theorem 6.17.** Assume that \(F : \mathbb{R}^n \to \mathbb{R}^m\) is Lipschitz in a neighborhood of \(\overline{x}\) and \(C \subset \mathbb{R}^n, Q \subset \mathbb{R}^m\) are nonempty and closed. Assume further that \(\overline{y} = F(\overline{x}) \in Q\). Let finally

\[
\Phi(x) = \begin{cases} F(x) - Q, & \text{if } x \in C; \\ \emptyset, & \text{otherwise.} \end{cases} ; \quad F_C(x) = \begin{cases} F(x), & \text{if } x \in C; \\ \emptyset, & \text{otherwise.} \end{cases}
\]

Then \(D^* \Phi(\overline{x}, 0)(y^*) = \partial(y^* \circ F_C)(\overline{x})\), if \(y^* \in N(Q, 0)\) and \(D^* \Phi(\overline{x}, 0)(y^*) = \emptyset\) otherwise. Thus

\[
\text{sur} \Phi(\overline{x}, 0) = \min \{\|x^*\| : x^* \in \partial(y^* \circ F_C)(\overline{x}), y^* \in N(Q, \overline{y}), \|y^*\| = 1\}.
\]

(Here of course \((y^* \circ F_C)(x) = \infty\) if \(x \notin C\).) If we compare this with Theorem 6.15, we see that transversality of \(F_C\) to \(Q\) at \(\overline{x}\) is equivalent to regularity of \(F_C - Q\) near \((\overline{x}, 0)\). We note also the following simple corollary of the theorem.
Corollary 6.18. Under the assumption of the theorem

\[ D^*\Phi(x,0)(y^*) \subset \partial(y^* \circ F)(x) + N(C,F(x)), \quad \text{if } y^* \in N(Q,0). \]

The set-valued mapping in Definition 6.11 was introduced in [88] where it was shown that subtransversality of a collection of sets is equivalent to subregularity of the mapping. Theorem 6.12 was partly proved in [115] (equivalence of (a) and (c)) and partly in [122] (equivalence of (a) and (b)). We refer to [115] for more equivalent descriptions (some looking very technical) of transversality and related properties. The results relating to transversality of set-valued mappings and sets in the image space seem to be new. The exception is Theorem 6.17 that can be extracted from Theorem 5.23 of [130].

Part 2. Applications

7 Special classes of mappings

If additional information on the structure of a mapping is available, it is often possible to get stronger results and/or better estimates for regularity rates and to develop more convenient mechanisms to compute or estimate the latter. In this section we briefly discuss how this can be implemented for three important classes of mappings.

7.1 Error bounds.

By an error bound for \( f \) (at level \( \alpha \)) on a set \( U \) we mean any estimate for the distance to \( [f \leq \alpha] \) in terms of \( (f(x) - \alpha)^+ \) for \( x \in U \). We shall be mainly interested in estimates of the form

\[ d(x, [f \leq \alpha]) \leq K(f(x) - \alpha)^+ \quad (7.1) \]

(which sometimes are called linear or Lipschitz error bounds).

As follows from the definition, error bounds can be viewed as rates of metric subregularity of the set-valued mapping \( \text{Epi} f(x) = [f(x), \infty) = \{ \alpha : (x, \alpha) \in \text{epi} f \} \) from \( X \) into \( \mathbb{R} \).

Lemma 7.1 (Basic lemma on error bounds). Let \( X \) be a complete metric space, let \( U \subset X \) be an open set, and let \( f \) be a lower semi-continuous function. Suppose that \( |\nabla f|(x) > r > 0 \) for any \( u \in U \setminus [f \leq 0] \). Then for any \( \overline{x} \in U \) such that \( f(\overline{x}) < rd(\overline{x}, X \setminus U) \) there is a \( \overline{u} \) such that \( f(\overline{u}) \leq 0 \) and \( d(\overline{u}, \overline{x}) \leq r^{-1}(f(\overline{x}))^+ \).

Proof. Without loss of generality, we may assume that \( f \) is nonnegative: just take \( f^+ \) instead of \( f \). So take an \( \overline{u} \) as in the statement. By Ekeland’s principle there is a \( \overline{u} \) such that \( d(\overline{u}, \overline{x}) \leq r^{-1} f(\overline{x}) \) and \( f(x) + rd(x, \overline{u}) > f(\overline{u}) \) if \( x \neq \overline{u} \). We claim that \( f(\overline{u}) \leq 0 \). Indeed, otherwise, by the assumption there would be an \( x \neq \overline{u} \) such that \( f(\overline{u}) - f(x) \geq rd(x, \overline{u}) \) — a contradiction. \( \square \)
For simplicity we shall speak here mainly about global error bounds, corresponding to \( U = X \), at the zero level. We shall denote by \( K_f \) the lower bound of \( K \) such that (7.1) holds for all \( x \). We also set for brevity
\[
S = [f \leq 0], \quad S_0 = [f = 0].
\]

### 7.1.1 Error bounds for convex functions.

We shall start with the simplest case of a convex function \( f \) (extended-real-valued in general) on a Banach space \( X \).

**Theorem 7.2.** Let \( X \) be a Banach space and \( f \) a proper closed convex function on \( X \). Assume that \( S = [f \leq 0] \neq \emptyset \). Then
\[
K_f^{-1} = \inf_{x \in S} \sup_{\|h\| \leq 1} (-f'(x; h)) = \inf_{x \in S} \sup_{\|h\| \leq 1} (d(0, \partial f(x))) = \inf_{x \in S} \sup_{\|h\| \leq 1} (\text{sur}(\text{Epi} f)(x, f(x))). \tag{7.2}
\]

Here \( \partial f(x) = \{x^* : f(x + h) - f(x) \geq \langle x^*, h \rangle \} \) is the convex subdifferential.

**Proof.** Equality of the three quantities on the right is not connected with regularity and we omit the proof. To prove the first equality, we observe that the inequality \( K_f^{-1} \leq r = \inf_{x \in [f > 0]} \sup_{\|h\| \leq 1} (\langle -f'(x; h) \rangle) \) is immediate from Basic Lemma because for a convex function \( |\nabla f| (x) = -\inf_{\|h\| \leq 1} f'(x; h) \). So it remains to prove the opposite inequality for which we can assume that \( r > 0 \).

Take a positive \( r' \) and \( \delta \) such that \( \delta < r' < r \) and let \( TU(x) \) be the set of pairs \((u, t)\) satisfying
\[
\|u - x\| \leq t, \quad f(u) \leq f(x) - r't \tag{7.3}
\]
By Ekeland’s variational principle for any \( \delta > 0 \) there is a \((\overline{\pi}, \overline{t}) \in TU(x) \) such that \( f(u) + \delta\|u - \overline{\pi}\| \) attains its minimum at \( \overline{\pi} \). Clearly \( \overline{t} > 0 \) (as \( f(x) > 0 \)). We claim that \( f(\overline{\pi}) = 0 \). Indeed, if \( f(\overline{\pi}) > 0 \), then there is an \( h \) with \( \|h\| = 1 \) such that \(-f'(\overline{\pi}; h) > r'\), that is \( f(\overline{\pi} + th) < f(\overline{\pi}) - r't \) for some \( t > 0 \). Set \( u = \overline{\pi} + th \). Then \( f(u) < f(\overline{\pi}) - \delta\|u - \overline{\pi}\| \) and we get a contradiction with the definition of \( \overline{\pi} \).

Thus \( f(\overline{\pi}) = 0 \) which means that
\[
d(x, S_0) \leq \|\overline{\pi} - x\| \leq t \leq \frac{1}{r'}f(x)
\]
and we are done as \( r' \) can be chosen arbitrarily close to \( r \) and \( x \) is an arbitrary point of \([f > 0]\).

There is another way to characterize \( K_f \) in terms of normal cones to \([f \leq 0]\).

**Theorem 7.3.** For any continuous convex function \( f \) on a Banach space \( X \)
\[
K_f = \inf_{x \in [f = 0]} \inf\{\tau > 0 : N([f = 0], x) \cap B_{X^*} \subset [0, \tau]\text{d}f(x)\}. \tag{7.4}
\]
7.1.2 Some general results on global error bounds.

Let us turn now to the general case of a lsc function on a complete metric space. Denote now by $K_f(\alpha, \beta)$ (where $\beta > \alpha \geq 0$) the lower bound of $K$ such that
\[
d(x, [f \leq \alpha]) \leq K_f(x)^+ \text{ if } \alpha < f(x) \leq \beta.
\]
Clearly, $K_f = \lim_{\beta \to \infty} K_f(0, \beta)$.

**Theorem 7.4.** Let $X$ be a complete metric space and $f$ a lower semicontinuous function on $X$. If $[f \leq 0] \neq \emptyset$, then
\[
\inf_{x \in [0 < f \leq \beta]} |\nabla f|(x) = \inf_{\alpha \in [0, \beta)} K_f(\alpha, \beta)^{-1}.
\]

**Proof.** Set $r = \inf_{x \in [0 < f \leq \beta]} |\nabla f|(x)$. The inequality $K_f(\alpha, \beta)^{-1} \geq r$ for $0 \leq \alpha < \beta$ is immediate from Lemma 7.1. This proves that the left side of the equality cannot be greater than the quantity on the right. To prove the opposite inequality it is natural to assume that $K_f(\alpha, \beta)^{-1} \geq \xi > 0$ for all $\alpha \in [0, \beta)$. For any $x \in [f > \alpha]$ and any $\varepsilon > 0$ such that $f(x) - \varepsilon > \alpha$ choose a $u = u(\varepsilon) \in [f \leq f(x) - \varepsilon]$ such that $d(x, u) \leq (1 + \varepsilon) d(x, [f \leq f(x) - \varepsilon]) \leq (1 + \varepsilon)\xi^{-1} \varepsilon$ and therefore $u \to x$ as $\varepsilon \to 0$. On the other hand, $\xi d(x, u) \leq f(x) - f(u)$ which (as $u \neq x$) implies that $\xi \leq |\nabla f|(x)$, whence $\xi \leq |\nabla f|(x)$, and the result follows.

As an immediate consequence we get

**Corollary 7.5.** Under the assumption of the theorem
\[
K_f^{-1} \geq \inf_{x \in [f > 0]} |\nabla f|(x).
\]

A trivial example of a function $f$ having an isolated local minimum at a certain $\overline{x}$ and such that $\inf f < f(\overline{x})$ shows that the inequality can be strict. This may happen of course even if the slope is different from zero everywhere on $[f > 0]$. In this case an estimate of another sort can be obtained. Set (for $\beta > 0$)
\[
d_f(\beta) = \sup_{x \in [f \leq \beta]} d(x, [f \leq 0])
\]
and define the functions
\[
\kappa_{f, \varepsilon}(t) = \sup \left\{ \frac{1}{|\nabla f|(x)} : |f(x) - t| < \varepsilon \right\}; \quad \kappa_f(t) = \lim_{\varepsilon \to 0} \kappa_{f, \varepsilon}(t).
\]

**Proposition 7.6.** Let $\beta > 0$. Assume that $[f \leq 0] \neq \emptyset$ and $|\nabla f|(x) \geq r > 0$ if $x \in [0 < f \leq \beta]$. Then
\[
d_f(\beta) \leq \int_0^\beta \kappa_f(t) dt.
\]

Following the pioneering 1952 work by Hoffmann [78] (to be proved later in this section), error bounds, both for nonconvex and, especially, convex functions have been intensively studied, especially during last 2-3 decades, both theoretically, in connection

54
with metric regularity, and also in view of their role in numerical analysis, see e.g. [43, 65, 123, 134, 163, 178]. Basic lemma was proved in [88], its earlier version corresponding to $U = X$ was proved by Azé-Corvellec-Lucchetti and appeared in [15]. A finite dimensional versions of Theorems 7.2 and 7.3 were proved in Lewis-Pang [123]. Klatte and Li [111]. The equality $K_{f^{-1}} = \inf\{d(0, \partial f(x)) : x \in [f > 0]\}$ in Theorem 7.2 was proved by Zalinescu (see [177]). The first two equalities in the theorem can be found in [12, 13] and the third equality for polyhedral functions on $\mathbb{R}^n$ in [131]. Theorem 7.3 was proved by Zheng and Ng [178] and Theorem 7.4 by Azé and Corvellec in [12]. The papers also contain sufficiently thorough bibliographic comments. Here we follow [95] where proofs of all stated and some other results can be found.

7.2 Mappings with convex graphs.

7.2.1 Convex processes.

We start with the simplest class of convex mappings known as convex processes. By definition a convex process is a set-valued mapping $A : X \nrightarrow Y$ from one Banach space into another whose graph is a convex cone. A convex process is closed if its graph is a closed convex cone. The closure $\text{cl}A$ of a convex process $A$ is defined by $\text{Graph}(\text{cl}A) = \text{cl}(\text{Graph}A)$. We shall usually work with closed convex processes. A convex process is bounded if there is an $r > 0$ such that $\|y\| \leq r \|x\|$ whenever $y \in A(x)$. A simplest nontrivial example of an unbounded closed convex process is a densely defined closed unbounded linear operator, as say the mapping $x(\cdot) \mapsto \dot{x}(\cdot)$ from $C[0, 1]$ into itself which associates with every continuously differentiable $x(\cdot)$ its derivative and the empty set with any other element of $C[0, 1]$.

According to Definition 5.1, given a convex process $A : X \nrightarrow Y$, the adjoint process $A^* : Y^* \nrightarrow X^*$ (always closed) is defined by

$$A^*(y^*) = \{x^* \in X^* : \langle x^*, x \rangle \leq \langle y^*, y \rangle, \forall (x, y) \in \text{Graph} A\}.$$  

By $A^{**}$ we denote a convex process from $X$ into $Y$ whose graph is the intersection of $-\text{Graph}(A^\ast)$ with $X \times Y$, that is $A^{**}(x) = \{y : -y \in (A^\ast)^*(-x)\}$. Simple separation arguments show that $A^{**} = \text{cl} A$ for any convex process.

**Proposition 7.7.** Let $A : X \nrightarrow Y$ be a convex process. Then $A(Q)$ is a convex set if so is $Q$ and for any $x_1, x_2 \in X$

$$A(x_1) + A(x_2) \subset A(x_1 + x_2).$$

**Proposition 7.8.** Let $K \subset X$ be a convex closed cone. Then for any $x \in K$ the tangent cone $T(K, x)$ is the closure of the cone generated by $K - x$. In particular $K \subset T(K, x)$.

The propositions are the key element in the proof of the following fundamental property of convex processes.

**Theorem 7.9** (regularity moduli of a convex process). For any closed convex process $A : X \nrightarrow Y$ from one Banach space into another

$$C(A) = C^*(A^*) = \text{sur } A(0|0) = \text{contr } A(0|0).$$
Note that the left inequality is equivalent to \( \|A^{-1}\| = \|(A^{-1})^*\| \) (cf [25]).

**Proof.** We first observe that the right equality is a consequence of the other two in view of Proposition 5.2. The inequality \( C^*(A^*) \geq C(A) \) follows from Theorem 5.4. The same theorem together with the definition of Banach constants implies that

\[
C^*(A^{**}) \geq C^*((A^*)^*) \geq C(A^*) \geq C^*(A^*).
\]

But \( A^{**} = A \), as \( A \) is closed, so that \( C^*((A^{**})^*) = C^*(A) \) (see again Theorem 5.4). This proves the left equality.

Passing to the proof of the middle equality, we first observe that by Proposition 5.2

\[
C(A) = \text{contr}A(0|0) \geq \text{sur}A(0|0)
\]

as the rate of surjection can never exceed the modulus of controllability. On the other hand, by Proposition 7.8 \( DA(0,0)(h) \subset DA(x,y)(h) \) for all \( (x,y) \) in Graph \( A \) and all \( h \). Hence by Theorem 5.13 \( \text{sur}A(0|0) \geq C(DA(0,0)) \). But \( DA(0,0)(h) = A(h) \) as the tangent cone to a closed convex cone at zero coincides with the latter. Thus \( \text{sur}A(0|0) \geq C(A) \).

**Corollary 7.10** (perfect regularity of convex processes). *Any closed convex process is perfectly regular at the origin.*

Note that a convex process may be not perfectly regular outside of the origin. For instance, consider in the space \( C[0,1] \) the mapping into itself defined by \( A(x(\cdot)) = x(\cdot) + K \) where \( K \) is the cone of nonnegative functions.

We conclude this subsection by considering the effect of linear perturbations. If \( A \) is a convex process, then so is \( A + A \) where \( A \) is a linear bounded operator from \( X \) into \( Y \). Thus if \( A \) is closed, then \( A + A \) is perfectly regular at the origin and we get as an immediate consequence of Theorem 5.28

**Theorem 7.11** (radius of regularity of a convex process). *If \( A : X \Rightarrow Y \) is a closed convex process, then*

\[
\text{rad}A(0|0) = \text{sur}A(0|0).
\]

Convex processes were introduced by Rockafellar [157] as an extension of linear operators and subsequently thoroughly studied by Robinson [148], Borwein [23] and Lewis [120]. In particular, [148] contains an extension to convex processes of Banach-Schauder open mapping theorem. Another remarkable result (which is actually a special case of Theorem 5 in the paper) can be reformulated as follows: *let \( X \) and \( Y \) be Banach spaces, and let \( A : X \Rightarrow Y \) and \( T : X \Rightarrow Y \) be closed convex processes. Then \( C(A - T) \geq C(A) - \|T\| \). The result equivalent to the equality \( C(A) = C^*(A^*) \) (Theorem 7.9) was proved and further discussed in [23] and Theorem 7.11 in [120] along with the equality of the radius and distance to infeasibility for convex processes.*

### 7.2.2 Theorem of Robinson-Ursescu.

**Theorem 7.12** (surjection modulus of a convex map). *Let \( X \) and \( Y \) be Banach spaces, and let \( F : X \Rightarrow Y \) be a set-valued mapping with convex and locally closed graph. Suppose
there are \((\bar{x}, \bar{y}) \in \text{Graph } F\), \(\alpha > 0\) and \(\beta > 0\) such that \(F(B(\bar{x}, \alpha))\) is dense in \(B(\bar{y}, \beta)\). Then

\[
\text{sur } F(\bar{x} | \bar{y}) \geq \frac{\beta}{\alpha}.
\]

(7.5)

**Proof.** We can set \(\bar{x} = 0, \bar{y} = 0\). It is clear that \(F(t\alpha B_X)\) is dense in \(t\beta B_Y\) for any \(t \in (0, 1)\). Denote \(r = \beta/\alpha\). We shall show that, given a \(\gamma > 0\), there is an \(\varepsilon > 0\) such that \(F(B(x, (1 + \gamma)t))\) is dense in \(B(v, rt)\) if \(\|x\| < \varepsilon\), \(\|v\| < \varepsilon\) and \(v \in F(x)\). The theorem then will follow from Corollary 3.8

So take a small \(\varepsilon > 0\), and let \(\|x_0\| < \varepsilon\), \(\|v_0\| < \varepsilon\) and \(v_0 \in F(x_0)\). Let further \(y \in B(v_0, rt)\) for some \(t \in (0, \varepsilon)\). Consider the ray emanating from \(v_0\) through \(y\) and let \(y_1\) be the point of the ray with \(\|y_1\| = \beta\), that is there is a \(\lambda > 0\) such that

\[
y = \frac{1}{1 + \lambda} y_1 + \frac{\lambda}{1 + \lambda} v_0, \quad \lambda \geq \frac{\beta - \varepsilon}{rt}.
\]

We have \(\|y_1 - y\| = \lambda\|v_0 - y\|\), that is

\[
\lambda = \frac{\|y_1 - y\|}{\|v_0 - y\|} \geq \frac{\beta - \varepsilon}{rt}; \quad 1 + \lambda \geq \frac{\beta - \varepsilon}{rt}
\]

In particular, if \(\beta \geq (1 + 2r)\varepsilon\), which we may assume, then \(\lambda \geq 1\).

Take a \(\delta > 0\). By the assumption there is an \(x_1 \in \alpha B\) such that \(\|y_1 - v_1\| < \delta\) for some \(v_1 \in F(x_1)\). Set

\[
v = \frac{1}{1 + \lambda} v_1 + \frac{\lambda}{1 + \lambda} v_0, \quad x = \frac{1}{1 + \lambda} x_1 + \frac{\lambda}{1 + \lambda} x_0
\]

Then \(v \in F(x)\) as \(\text{Graph } F\) is convex. We have \(\|y - v\| \leq \delta/(1 + \lambda) \leq \delta/2\) and

\[
\|x - x_0\| \leq \frac{1}{1 + \lambda}\|x_1 - x_0\| \leq \frac{\alpha + \varepsilon}{1 + \lambda} \leq \frac{\alpha + \varepsilon}{\beta - \varepsilon} rt.
\]

If

\[
1 + \gamma \geq \frac{\alpha + \varepsilon}{\beta - \varepsilon} \cdot \frac{\beta}{\alpha},
\]

this completes the proof as \(\delta\) can be chosen arbitrary small. \(\square\)

As a corollary we get

**Theorem 7.13** (Robinson-Ursescu \[151\] [167]). Let \(X\) and \(Y\) be Banach spaces. If the graph of \(F : X \rightrightarrows Y\) is convex and closed and \(\overline{y} \in \text{int } F(X)\), then \(F\) is regular at any \((\bar{x}, \bar{y}) \in \text{Graph } F\).

**Proof.** Let \(\overline{y} \in F(\bar{x})\). We have to show that there are \(\alpha > 0\) and \(\beta > 0\) such that \(F(B(\bar{x}, \alpha))\) is dense in \(B(\overline{y}, \beta)\) which is easy to do with the help of the standard argument using Baire category. \(\square\)
7.2.3 Mappings with convex graphs. Regularity rates.

Here we give two results containing exact formulas for the rate of surjection of set-valued mappings with convex graph.

**Theorem 7.14.** Let \( F : X \rightrightarrows Y \) be a set-valued mapping with convex and locally closed graph. If \( y \in F(x) \), then

\[
\text{sur} F(x|y) = \lim_{\varepsilon \to +0} \inf_{\|y^*\|=1} \inf_{x^*} \left( \|x^*\| + \frac{1}{\varepsilon} s_{\text{Graph } (F-(x,y))}(x^*, y^*) \right).
\]

The theorem was proved in Ioffe-Sekiguchi [102], see also for [95] for a short proof. It allows to also get a "primal" representation for the rate of surjection of a convex set-valued mapping. Here we give two results containing exact formulas for the rate of surjection of set-valued mappings with convex graph. The key to this development is the concept of homogenization \( Q \) of a convex set \( Q \subset X \) which is the closed convex cone in \( X \times R \) generated by the set \( Q \times \{1\} \). It is an easy matter to verify (if \( Q \) is also closed) that \( (x,t) \in Q \) if and only if \( x \in tQ \) if \( t > 0 \) and \( x \in Q^\infty \), the recession cone of \( Q \), if \( t = 0 \). (Recall that \( Q^\infty = \{ h \in Q : x + h \in Q, \forall x \in Q \} \).

Given a set-valued mapping \( F : X \rightrightarrows Y \) with convex closed graph, we associate with \( F \) and any \((\bar{x},\bar{y}) \in X \times Y \) (not necessarily in the graph of \( F \)) a convex process \( F_{(\bar{x},\bar{y})} : X \times R \rightrightarrows Y \) whose graph is the homogenization of \( F - (\bar{x},\bar{y}) \). It is easy to see that

\[
F_{(\bar{x},\bar{y})}(h,t) = \begin{cases} t(F(x+\frac{h}{t})-\bar{y}), & \text{if } t > 0, \\ F^\infty(h), & \text{if } t = 0, \\ \emptyset, & \text{if } t < 0, \end{cases}
\]

where \( F^\infty \) is the “horizon” mapping of \( F \) whose graph is the recession cone of Graph \( F \):

\[
\text{Graph } F^\infty = \{(h,v) : (x+h,y+v) \in \text{Graph } F, \forall (x,y) \in \text{Graph } F\}.
\]

If \((\bar{x},\bar{y}) = (0,0)\), we shall simply write \( F \) (without the subscript) and call this convex process the homogenization of \( F \).

In the theorem below we use the \( \varepsilon \)-norms in \( X \times R \): \( \|(h,t)\|_\varepsilon = \max\{\|x\|, \varepsilon t\} \) and denote by \( C_\varepsilon(F_{(\bar{x},\bar{y})}) \) the Banach constant of \( F_{(\bar{x},\bar{y})} \) corresponding to this norm.

**Theorem 7.15** (primal representation of the surjection modulus). If \( F : X \rightrightarrows Y \) is a set-valued mapping with convex and locally closed graph, then

\[
\text{sur} F(x|y) = \lim_{\varepsilon \to +0} C_\varepsilon(F_{(\bar{x},\bar{y})}).
\]

**Proof.** We have (setting below \( h = t(x-\bar{x}), \ v = t(y-\bar{y}) \))

\[
\text{Graph } F^*_{{(\bar{x},\bar{y})}} = \{(x^*,y^*,\lambda) : \langle x^*,h \rangle - \langle y^*,v \rangle + \lambda t \leq 0 : \forall (h,v,t) \in \text{Graph } F_{(\bar{x},\bar{y})} \}
\]

\[
= \{(x^*,y^*,\lambda) : t(\langle x^*,x-\bar{x} \rangle - \langle y^*,y-\bar{y} \rangle + \lambda) \leq 0 : \forall (x,y) \in \text{Graph } F, t > 0 \}
\]

\[
= \{(x^*,y^*,\lambda) : s_{\text{Graph } F-(\bar{x},\bar{y})}(x^*, -y^*) + \lambda \leq 0 \}.
\]
As the support function of $\text{Graph } F - (\bar{x}, \bar{y})$ is nonnegative, it follows that $\lambda \leq 0$ whenever $(x^*, y^*, \lambda) \in \text{Graph } F_{(\bar{x}, \bar{y})}$. The norm in $X^* \times IR$ dual to $\| \cdot \|_\varepsilon$ is $\|(x^*, \lambda)\|_\varepsilon = \|x^*\| + \varepsilon^{-1}\|\lambda\|$. Let $d_\varepsilon$ stand for the distance in $X^* \times IR$ corresponding to this norm. Then

$$d_\varepsilon(0, F^\varepsilon_{(\bar{x}, \bar{y})}(\bar{x}, \bar{y}))(y^*)) = \inf\{|\|x^*\| + \varepsilon^{-1}|\lambda| : s_{\text{Graph } F_{-(\bar{x}, \bar{y})}}(x^*, -y^*) + \lambda \leq 0\}$$

$$= \inf\{|\|x^*\| + \varepsilon^{-1}s_{\text{Graph } F_{-(\bar{x}, \bar{y})}}(x^*, -y^*)\|$$

It remains to compare this with Theorem 7.14 to see that

$$\text{sur } F(\varepsilon|\varepsilon) = \lim_{\varepsilon \to +0} \inf_{\|y^*\| = 1} d_\varepsilon(0, F^\varepsilon_{(\bar{x}, \bar{y})}(y^*))$$

and then to refer to Theorem 7.9 to conclude that the quantity on the right is precisely the limit as $\varepsilon \to 0$ of $\inf_{\|y^*\| = 1} C_\varepsilon(cF_{(\bar{x}, \bar{y})}(y^*))$, where the closure operation can be dropped because as we mentioned the norms (and therefore the Banach constants) of a convex process and its closure coincide.

The concept of homogenization was introduced by Hörmander [79]. The idea to apply homogenization for regularity estimation goes back to Robinson’s [150], His main result actually says that $\text{sur } F(\varepsilon|\varepsilon) \geq C_1(F_{(\bar{x}, \bar{y})})$. In a somewhat different context homogenization techniques was applied by Lewis [121] for estimating distance to infeasibility of so called conic systems. Full statement of Theorem 7.15 was proved also in [102]. We have not discussed here some well developed problems relating to regularity of maps with convex graphs, e.g. stability under perturbations of systems of convex inequalities - see e.g. [29, 95, 149] and references in the first two quoted papers.

### 7.3 Single-valued Lipschitz maps.

The collection of analytic tools that allow to compute and estimate regularity moduli of Lipschitz single-valued mappings contains at least two devices, not available in the general situation, which are a lot more convenient to work with than coderivatives. The first is the scalarized coderivative (associated with a subdifferential):

$$D^s F(x)(y^*) = \partial(y^* \circ F)(x)$$

and the other results from suitable local approximations of the mapping either by homogeneous set-valued mappings or by sets of linear operators.

The following result is straightforward.

**Proposition 7.16.** If $F : X \to Y$ is Lipschitz continuous near $x \in X$, then for every $y^* \in Y^*$

$$\partial_F(y^* \circ F)(x) = D^s_F(x)(y^*).$$

(7.6)

Things are more complicated with the Dini-Hadamard subdifferential. From now on we assume that all spaces are Gâteaux smooth.

**Definition 7.17.** A homogeneous set-valued mapping $A : X \Rightarrow Y$ is a strict Hadamard prederivative of $F : X \to Y$ at $\varepsilon$ if $\|A\|_+ < \infty$, and for any norm compact set $Q \subset X$

$$F(x + th) - F(x) \subset tA(h) + r(t, x)h\|h\|B_Y, \ \forall h \in Q,$$

(7.7)
where \( r(t, x) = r(t, x, Q) \to 0 \) when \( x \to \bar{x}, t \to +0 \). If moreover the inclusion holds with \( Q \) replaced by \( B_X \), then \( \mathcal{A} \) is called strict Fréchet prederivative of \( F \) at \( \bar{x} \). Clearly, for a Fréchet prederivative we can write \( r(t, x) \) in the form \( \rho(t, \|x - \bar{x}\|) \).

There are some canonical ways for constructing prederivatives. The first to mention is the generalized Jacobian introduced by Clarke [36] for mappings in the finite dimensional case and then extended to some classes of Banach spaces by Páles and Zeidan [139, 140]. Another construction, not associated with linear operators was introduced in [82]. Take an \( \varepsilon > 0 \) and set

\[
\mathcal{H}_\varepsilon(h) := \{ \lambda^{-1}(F(x + \lambda h) - F(x)) : x, x + \lambda h \in \text{dom } F \cap B(\bar{x}, \varepsilon), \lambda > 0 \}, \quad h \in X.
\]

Then \( 0 \in \mathcal{H}_\varepsilon(0) \) and for \( t > 0 \) we have

\[
\mathcal{H}_\varepsilon(th) = t\{ (t\lambda)^{-1}(F(x + t\lambda h) - F(x)) : x, x + t\lambda h \in \text{dom } F \cap B(\bar{x}, \varepsilon), \lambda > 0 \},
\]

that is \( \mathcal{H}_\varepsilon(th) = t\mathcal{H}_\varepsilon(h) \). Thus \( \mathcal{H}_\varepsilon \) is positively homogeneous and it is an easy matter to see that (7.7) holds with \( r(t, x) = 0 \).

We say that \( F : X \to Y \) is directionally compact at \( \bar{x} \in \text{dom } F \) if it has a (norm) compact-valued strict Hadamard prederivative with closed graph. It is strongly directionally compact if there is a compact-valued strict Fréchet prederivative with closed graph.

The simplest, and probably the most important example of a directionally compact (actually even strongly directionally compact) mapping is an integral operator associated with a differential equation, e.g.

\[
x(\cdot) \mapsto F(x(\cdot))(t) = x(t) - \int_0^t f(s, x(s))ds
\]

with \( f(t, \cdot) \) Lipschitz with summable rate.

**Proposition 7.18** ([86]). If \( F : X \to Y \) is Lipschitz continuous near \( x \), then

\[
\partial_H(y^* \circ F)(x) \subset D_H^*F(x)(y^*), \quad \forall y^* \in Y^*.
\]

If furthermore \( F : X \to Y \) is directionally compact at \( x \), then

\[
D_H^*F(x)(y^*) = \partial_H(y^* \circ F)(x) \quad \& \quad D_G^*F(x)(y^*) = \partial_G(y^* \circ F)(x), \quad \forall y^* \in Y^*.
\]

Combining this proposition with Theorem [5.21] we get

**Theorem 7.19.** Let \( F : X \to Y \) satisfy the Lipschitz condition in a neighborhood of \( \bar{x} \). If \( F \) is directionally compact at all \( x \) of the neighborhood, then

\[
sur F(\bar{x}) \geq \lim_{\varepsilon \to 0} \inf \{ \|x^*\| : x^* \in \partial_H(y^* \circ F)(x), \|y^*\| = 1, \|x - \bar{x}\| < \varepsilon \},
\]

The obvious inequality

\[
(y^* \circ F)(x + h) - (y^* \circ F)(x) \geq \inf_{w \in \mathcal{H}(x)(h)} \langle y^*, w \rangle
\]

(where \( \mathcal{H}(x) \) is a strict prederivative at \( x \)) leads to the estimate \( \sur F(\bar{x}) \geq \lim_{x \to \bar{x}} \inf_{x \to \bar{x}} C^*(\mathcal{H}(x)) \) under the assumptions of the theorem. A better result can be proved with the help of the general metric regularity criteria if \( F \) has a strict Fréchet prederivative at \( \bar{x} \).
Theorem 7.20. Assume that $Y$ is Gâteaux smooth and $F : X \to Y$ satisfies the Lipschitz condition in a neighborhood of $\bar{x}$ and, moreover, admits at $\bar{x}$ a strict Fréchet prederivative $H$ with norm compact values such that for any $y^*$ with $\|y^*\| = 1$

$$\sup_{\|h\|=1} \inf_{w \in H(h)} \langle y^*, w \rangle \geq \rho > 0. \quad (7.8)$$

Then $\operatorname{sur} F(\bar{x}) \geq \rho$.

Proof. With no loss of generality we may assume that the norm in $Y$ is Gâteaux smooth off the origin. Take an $\varepsilon \in (0, \rho/3)$ and an $r > 0$ such that

$$F(x') - F(x) \in H(x) + \varepsilon \|x' - x\|, \quad (7.9)$$

if $x, x' \in B(\bar{x}, r)$. Take an $x \in \bar{B}(\bar{x}, r/2)$ and a $y \in Y$, different from $F(x)$. Let $y^*$ denote the derivative of $\| \cdot \|$ at $y - F(x)$. Then

$$\lim_{t \to 0} t^{-1} (\|y - F(x) + tw\| - \|y - F(x)\|) = \langle y^*, w \rangle, \quad \text{for every } w \in Y. \quad (7.10)$$

By (7.8), there is an $h \in S_X$ such that

$$\langle y^*, w \rangle > \rho - \varepsilon, \quad \text{for all } w \in H(h). \quad (7.11)$$

Since the set $-H(h)$ is compact and the limit in (7.10) is uniform with respect to $w$ from any fixed compact set, we conclude that for sufficiently small $t > 0$

$$\|y - F(x) - tw\| - \|y - F(x)\| + \langle y^*, tw \rangle < t\varepsilon \quad \text{for all } w \in H(h).$$

This and (7.11) imply that

$$\|y - F(x) - tw\| < \|y - F(x)\| - \langle y^*, tw \rangle + \varepsilon t \leq \|y - F(x)\| - t(\rho - 2\varepsilon) \quad (7.12)$$

for all $w \in H(h)$. Let $x' := x + th$. Then $\|x' - x\| = \|th\| = t < r/2$, hence $x' \in B(\bar{x}, r)$. Since $H$ is positively homogeneous, we have $H(x' - x) = H(th) = tH(h)$. Thus by (7.9) there is a $w \in H(h)$ such that

$$\|F(x') - F(x) - tw\| \leq t\varepsilon. \quad (7.13)$$

Now, we are ready for the following chain of estimates

$$\|y - F(x')\| \leq \|F(x) - F(x') + tw\| + \|y - F(x) - tw\|$$

$$< \varepsilon t + \|y - F(x)\| - (\rho - 2\varepsilon)t \quad \text{(by (7.13) and (7.12))}$$

$$= \|y - F(x)\| - (\rho - 3\varepsilon)t = \|y - F(x)\| - (\rho - 3\varepsilon)\|x' - x\|.$$ 

It remains to apply the criterion of Theorem 3.2.

A slight modification of the proof allows to get the following
Theorem 7.21. Assume that $F : X \to Y$ satisfies the Lipschitz condition in a neighborhood of $\bar{x}$ and, moreover, there are a homogeneous set-valued mapping $H : X \to Y$ with norm compact values and $\beta \geq 0$ such that (7.8) holds and

$$F(x + h) - F(x) \subset H(h) + \beta\|x' - x\|B_Y.$$  \hfill (7.14)

Then $\text{sur} F(\bar{x}) \geq \rho - \beta$.

This theorem, in turn, allows us to look at what happens when a Lipschitz mapping is approximated by a bunch of linear operators. Indeed, if $\mathcal{T}$ is a collection of linear operators from $X$ to $Y$, then the set-valued mapping $X \ni x \mapsto -\to H(x) := \{Tx : T \in \mathcal{T}\}$ is of course positively homogeneous. It is an easy matter to see that $H$ inherits some properties of $\mathcal{T}$: for us it is important to observe that when $\mathcal{T}$ is (relatively) norm compact in $L(X,Y)$ with the norm $\|T\| = \sup\{\|Tx\| : \|x\| \leq 1\}$, then so are the values of $H$, if $\mathcal{T}$ is bounded, then the values of $H$ are also bounded etc.. Thus we come to the following conclusion.

Theorem 7.22. Assume that for a given $\bar{x} \in \text{dom} F$ there is a convex subset $T \subset L(X,Y)$ which is norm compact in $L(X,Y)$ and has the following two properties:

(a) there is a $\beta > 0$ such that for any $x, x'$ in a neighborhood of $\bar{x}$ there is a $T \in \mathcal{T}$ such that

$$\|F(x) - F(x') - T(x - x')\| \leq \beta\|x - x'\|;$$  \hfill (7.15)

(b) there are $\rho > 0$ and $\varepsilon > 0$ such that for any $T \in \mathcal{T}$

$$\varepsilon \rho B_Y \subset T(\varepsilon B_X).$$  \hfill (7.16)

Then $\text{sur} F(\bar{x}) \geq \rho - \beta$.

Scalarization formulas first appeared in [83] for mappings between finite dimensional spaces and [113] for mappings between Fréchet smooth spaces, although scalarized coderivatives were considered already in [82] [112]. The very term “coderivative” was introduced in [82]. The concept of prederivative was introduced in [82] and a characterization of directional compactness in [86], see also [106] for an earlier result.

Theorems 7.20 and 7.21 will appear in [32]. Theorem 7.22 was proved in [31]. An earlier result without constraints on the domain of the mapping was proved by Páles in [138]. We also refer to [32] for a shorter proofs of the last theorem. Note that the convexity requirement in Theorem 7.22 is essential (consider, for instance, $F(x) = |x| : IR \to IR$ and $\mathcal{T}$ containing two operators $T_1(x) = x$ and $T_2(x) = -x$). Because of this requirement the estimate provided by Theorem 7.22 is generally less precise than those of Theorems 7.19 and 7.20 (consider for instance the mapping $IR^2 \to IR : F(x_1, x_2) = |x_1| - |x_2|$), but it can be easier to apply in certain cases (e.g. in the finite dimensional case when we can take the generalized Jacobian as $\mathcal{T}$ - see [36]).

7.4 Polyhedral sets and mappings

This subsection contains some elementary results concerning geometry of polyhedral sets in $IR^n$ and regularity of set-valued mappings with polyhedral graphs. Deeper problems associated with variational inequalities over convex polyhedral sets will be discussed in the next section.
Proposition 7.24 (local tangential representation). The interest in this section is to study regularity properties of such mappings.

\[ \epsilon > 0 \]

Let \( \epsilon > 0 \). Then there is an \( \epsilon > 0 \) such that

\[ Q \cap B(\mathbf{x}, \epsilon) = \mathbf{x} + T(Q, \mathbf{x}) \cap (\epsilon B). \]

As an immediate consequence, we conclude that regularity properties of a polyhedral set valued mapping with closed graph at a point of the graph are fully determined by the corresponding properties at zero of its graphical derivative at the point.

One more useful corollary concerns normal cones of a polyhedral sets.

Proposition 7.25. Let \( Q \subset \mathbb{R}^n \) be a polyhedral set. Then for any \( \mathbf{x} \in Q \) there is an \( \epsilon > 0 \) such that \( N(Q, \mathbf{x}) \subset N(Q, \mathbf{x}) \) for any \( \mathbf{x} \in Q \cap B(\mathbf{x}, \epsilon) \).

Our first result is the famous Hoffmann theorem on error bounds for a system of linear inequalities. Set \( a = (\alpha_1, \ldots, \alpha_m) \in \mathbb{R}^m \) and let \( Q(a) \) be defined by (7.17).

Theorem 7.26 (Hoffmann). Given \( x_i^* \in \mathbb{R}^n \). Then there is a \( K > 0 \) such that the inequality

\[ d(x, Q(a)) \leq K \left( \sum_{i=1}^{k} (x_i^*, x) - \alpha_i \right)^+ + \sum_{i=k+1}^{m} |(x_i^*, x) - \alpha_i| \]

holds for all \( x \in \mathbb{R}^n \) and all \( a \in \mathbb{R}^m \) such that \( Q(a) \neq \emptyset \).

Proof. We shall apply Theorem 7.2. Take an \( a \) and set

\[ f(x) = \sum_{i=1}^{k} (x_i^*, x) - \alpha_i \] \[ + \sum_{i=k+1}^{m} |(x_i^*, x) - \alpha_i| \]

Then \( Q(a) = \{ f \leq 0 \} \). Set

\[ I_1(x) = \{ i \in \{1, \ldots, k\} : (x_i^*, x) \leq \alpha_i \} \]
\[ J_+(x) = \{ i \in \{1, \ldots, m\} : (x_i^*, x) > \alpha_i \} \]
\[ I_0(x) = \{ i \in \{k+1, \ldots, m\} : (x_i^*, x) = \alpha_i \} \]
\[ J_-(x) = \{ i \in \{k+1, \ldots, m\} : (x_i^*, x) < \alpha_i \} \]
Then
\[ \partial f(x) = \sum_{i \in I_1(x)} [0, 1]x_i^* + \sum_{i \in I_0(x)} [-1, 1]x_i^* + \sum_{i \in J_+(x)} x_i^* - \sum_{i \in J_-(x)} x_i^*. \]

If \( x \notin Q(\alpha) \), then \( 0 \notin \partial f(x) \) and \( d(0, \partial f(x)) > 0 \).

We observe now that the dependence of \( \partial f(x) \) on \( x \) and \( a \) is fully determined by the decomposition of the index set \( 1, \ldots, m \). Let \( \Sigma \) be the collection of all decompositions of the index set into four subsets \( I_1, I_0, J_+, J_- \) such that \( I_1 \subset \{1, \ldots, k\}, I_0, J_- \subset \{k+1, \ldots, m\} \) and

\[ 0 \notin \sum_{i \in I_1} [0, 1]x_i^* + \sum_{i \in I_0} [-1, 1]x_i^* + \sum_{i \in J_+} x_i^* - \sum_{i \in J_-} x_i^*. \]

For any \( \sigma \in \Sigma \) denote by \( \gamma(\sigma) \) the distance from zero to the set in the right-hand side of the above inclusion, and let \( K \) stand for the upper bound of \( \gamma(\sigma)^{-1} \) over \( \sigma \in \Sigma \). Then \( K < \infty \) since \( \Sigma \) is a finite set. Clearly, \( K \) does not depend on either \( a \) or \( x \). On the other hand, \( K \partial f(x) \geq 1 \). It remains to refer to Theorem 7.22 to conclude the proof.

As an immediate consequence, we get

**Theorem 7.27** (regularity of convex polyhedral mappings). Let \( F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m \) be a polyhedral set-valued mapping. Then

(a) there is a \( K > 0 \) such that \( d(y, F(\overline{x})) \leq K\|x - \overline{x}\| \) for any \( \overline{x} \in \text{dom} \ F \) and any \( (x, y) \in \text{Graph} \ F \);

(b) there is a \( K > 0 \) (different from that in (a)) such that \( d(x, F^{-1}(y)) \leq Kd(y, F(x)) \) for any \( x \in \text{dom} \ F \) and \( y \in F(X) \).

and

**Theorem 7.28** (global subtransversality of convex polyhedral sets). Any two convex polyhedral sets \( Q_1 \) and \( Q_2 \) with nonempty intersection are globally subtransversal: there is a \( K > 0 \) such that

\[ d(x, Q_1 \cap Q_2) \leq K(d(x, Q_1) + d(x, Q_2)). \]

To prove Theorem 7.27 we have to apply the Hoffmann estimate to the graph of \( F \). Concerning Theorem 7.28 it should be observed that global transversality does not imply transversality at any point. As a simple example, consider the half spaces \( S_1 = \{ x : \langle x^*, x \rangle \geq 0 \} \) and \( S_2 = \{ x : \langle x^*, x \rangle \leq 0 \} \) with some \( x^* \neq 0 \). The intersection of the sets is \( \text{Ker} x^* \neq \emptyset \). But the inclusions \( x_1 - x \in S_1 \) and \( x_2 - x \in S_2 \) imply \( \langle x^*, x_1 \rangle \geq \langle x^*, x_2 \rangle \), hence (see Definition 6.11) \( S_1 \) and \( S_2 \) are not transversal at points of \( \text{Ker} x^* \).

The results easily extend to all (not necessarily convex) polyhedral mappings.

**Theorem 7.29** (subregularity of polyhedral mappings). Let \( F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m \) be a semi-linear set-valued mapping with closed graph. Then

(a) there is a \( K > 0 \) such that for any \( \overline{x} \in \text{dom} \ F \) there is an \( \varepsilon > 0 \) such that \( d(y, F(\overline{x})) \leq K\|x - \overline{x}\| \) for all \( (x, y) \in \text{Graph} \ F \) such that \( \|x - \overline{x}\| < \varepsilon \);

(b) there is a \( K > 0 \) (different from that in (a)) such that for any \( (\bar{x}, \bar{y}) \in \text{Graph} \ F \) there is an \( \varepsilon > 0 \) such that \( d(x, F^{-1}(y)) \leq Kd(y, F(x)) \) if \( \|x - \overline{x}\| < K\varepsilon \). Thus \( F \) is subregular at any point of its graph.
Proof. We have \( F(x) = \bigcup_{i=1}^{k} F_i(x) \), where all \( F_i \) are convex polyhedral set-valued mappings. By Theorem 7.27 for any \( i \) there is a \( K_i \) such that \( d(y, F_i(x)) \leq K_i \|x - \overline{x}\| \) for any \( \overline{x} \in \text{dom} \ F_i \) and any \( (x, y) \in \text{Graph} \ F_i \). Now fix some \( \overline{x} \in \text{dom} \ F \), and let \( I = \{ i : \overline{x} \in \text{dom} \ F_i \} \). Choose an \( \varepsilon > 0 \) so small that \( d(x, \text{dom} \ F_i) > \varepsilon \) if \( i \not\in I \) and \( \|x - \overline{x}\| < \varepsilon \). (Clearly, such an \( \varepsilon \) can be found as all \( \text{dom} \ F_i \) are polyhedral sets, hence closed.) If now \( y \in F(x) \) and \( \|x - \overline{x}\| < \varepsilon \), then \( I(x, y) = \{ i : y \in F_i(x) \} \subset I \). On the other hand, as we have seen, there are \( K_i \) such that \( y \in F_i(x) \) implies that \( d(y, F_i(\overline{x})) \leq K_i \|x - \overline{x}\| \). Thus, if \( y \in F(x) \) and \( \|x - \overline{x}\| < \varepsilon \), then
\[
d(y, F(\overline{x})) \leq \max_{i \in I(x, y)} d(y, F_i(\overline{x})) \leq (\max K_i) \|x - \overline{x}\|.
\]
This proves the first statement.

To prove the second, we apply the first to \( F^{-1} \) and find \( K \) and \( \varepsilon \) such that \( d(x, F^{-1}(\overline{y})) \leq K \|v - \overline{y}\| \) if \( v \in F(x) \) and \( \|v - \overline{y}\| < \varepsilon \). If \( d(\overline{y}, F(x)) < \varepsilon \), it follows that \( d(x, F^{-1}(\overline{y})) \leq K d(\overline{y}, F(x)) \). This inequality trivially holds if \( d(\overline{y}, F(x)) \geq \varepsilon \) and \( \|x - \overline{x}\| \leq K \varepsilon \).

The property in the second part of the theorem falls short of metric regularity because it does not guarantee that the \( \varepsilon \) will be uniformly bounded away from zero if we slightly change \( \overline{y} \). The following simple example illustrates the phenomenon.

**Example 7.30.** Let \( X = Y = R \), and let
\[
F_1(x) = \begin{cases} \mathbb{R}_+, & \text{if } x > 0, \\ \mathbb{R}, & \text{if } x = 0, \\ \emptyset, & \text{if } x < 0 \end{cases}
\]
and \( F_2(x) = \begin{cases} \mathbb{R}_-, & \text{if } x < 0, \\ \mathbb{R}, & \text{if } x = 0, \\ \emptyset, & \text{if } x > 0 \end{cases} \)
and \( F(x) = F_1(x) \cup F_2(x) \). Fix some \( y > 0 \) and \( x < 0 \). Then \( F^{-1}(y) = H_+ \) and \( d(x, F^{-1}(y)) = |x| \), \( d(y, F(x)) = y \) so that for no \( K \) the inequality \( d(x, F^{-1}(y)) \leq K d(y, F(x)) \) holds in a neighborhood of \((0, 0)\).

**Corollary 7.31** (subtransversality of polyhedral sets). Any two semi-linear sets \( Q_1 \) and \( Q_2 \) with nonempty intersection are subtransversal at any common point of the sets.
\[
d(x, Q_1 \cap Q_2) \leq K(d(x, Q_1) + d(x, Q_2)).
\]

To conclude, we mention that for any polyhedral mapping \( F : \mathbb{R}^n \Rightarrow \mathbb{R}^n \) the set of critical values (that is such \( y \in \mathbb{R}^n \) such that sur\( F(x|y) = 0 \) for some \( x \in F^{-1}(y) \)) is a polyhedral set of dimension smaller than \( n \). This will immediately follow from the semi-algebraic Sard theorem stated in the next subsection.

### 7.5 Semi-algebraic mappings, stratifications and the Sard theorem.

Most of the results of this subsection, including the Sard theorem can be extended to a wide class of objects, so called definable sets, mappings and functions. We however confine ourselves here to semi-algebraic functions whose definition is much simpler (compare with the general definition of definability) and does not require any specific effort.\(^2\)

\(^2\)It should be mentioned that recently Barbet, Dambrine, Daniilidis, Rifford \cite{[18]} proved a remarkable result containing extensions of the Sard theorem to some other important classes of non-smooth functions.
We shall concentrate basically on two topics: consequences of the general theory and studies associated with semi-algebraic geometry, mainly in connection with the Sard theorem.

### 7.5.1 Basic properties (see [21][42]).

A semi-algebraic set in $\mathbb{R}^n$ is by definition a union of finitely many sets of solutions of a finite system of polynomial equalities and inequalities of $n$ variables:

$$\{x \in \mathbb{R}^n : P_i(x) = 0, \ i = 1, \ldots, k, \ P_i(x) < 0, \ i = k + 1, \ldots, m\}.$$ 

As immediately follows from the definition, every algebraic set is semi-algebraic, every polyhedral set is semi-algebraic, unions and intersections of finite collections of semi-algebraic sets are again semi-algebraic. The main fact of the semi-algebraic geometry is the deep Tarski-Seidenberg theorem which roughly speaking says that a linear projection of a semi-algebraic set is a semi-algebraic set. This theorem determines stability of the class of semi-algebraic sets with respect to a broad variety of transformations.

A mapping (no matter single or set-valued) is semi-algebraic if its graph is semi-algebraic. Here is a list of some basic properties of semi-algebraic sets and mappings:

- the closure and interior of a semi-algebraic set is semi-algebraic;
- Cartesian product of semialgebraic sets is semi-algebraic;
- composition of semi-algebraic mappings is semi-algebraic;
- image and preimage of a semi-algebraic set under a semi-algebraic mapping is semi-algebraic;
- derivative of a (single-valued) semi-algebraic mapping is semi-algebraic;
- the upper and lower bound of a finite collection of extended-real-valued semi-algebraic functions is semi-algebraic;
- if we have a semi-algebraic function of two (vector) variables, then its upper or lower bound with respect to one of the variables on a semi-algebraic set is semi-algebraic;
- if $F$ is a semi-algebraic set-valued mapping such that every $F(x)$ is a finite set, then the number of elements in each $F(x)$ does not exceed certain finite $N$.

For us, in the context of variational analysis and, especially, regularity theory, the most important is that

- subdifferential mapping of a semi-algebraic function or the coderivative mapping of a semi-algebraic map is semi-algebraic (no matter of which subdifferential on $\mathbb{R}^n$: Fréchet, Dini-Hadamard, limiting or Clarke, we are talking about);
- slope of a semi-algebraic function is a semi-algebraic function of the point;
- rates of regularity of a semi-algebraic functions are also semi-algebraic functions of the point of the graph.

**Definition 7.32.** A finite partition $(M_i)$ of a set $Q \subset \mathbb{R}^n$ is called $C^r$-Whitney stratification of $Q$ if each $M_i$ is a $C^r$-manifold and the following two properties are satisfied:

(a) if $(x_k) \subset M_i$ converges to some $x$ belonging to another element $(M_j)$ of the partition, and the unit normal vectors $v_k \in N_{x_k} M_i$ converge to some $v$, then $v \in N_x M_j$;

(b) if $M_j \cap \text{cl} M_i \neq \emptyset$, then $M_j \subset \text{cl} M_i$. 

66
Elements of partitions are usually called *strata*. The following remarkable fact is due to S. Lojasie\v{v}cz:

**Theorem 7.33** (stratification theorem). Given a semi-algebraic set $Q \subset \mathbb{R}^n$ and an $r \in \mathbb{N}$. Then $Q$ admits a Whitney stratification into semi-algebraic $C^r$-manifolds.

Of course, stratification is not unique. But it is easy to understand that maximal dimensions of the strata coincide for all Whitney stratifications. This observation justifies the following

**Definition 7.34.** The *dimension* $\dim Q$ of a semi-algebraic set $Q$ is the maximal dimension of the strata in Whitney stratifications of $Q$.

The most important consequence of the stratification theorem is a Sard-type theorem for semi-algebraic set-valued mappings,

**Definition 7.35.** Let $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ be a set-valued mapping with semi-algebraic graph, and let $\partial$ stand either for the limiting or for the Clarke subdifferential. A point $\overline{y} \in \mathbb{R}^m$ is a *critical value* of $F$ if there is an $x \in \mathbb{R}^n$ such that $y \in F(x)$ and $0 \in D^s F(x|y)(y^*)$ for some $y^* \neq 0$.

**Theorem 7.36** (semi-algebraic Sard theorem). Critical values of a semi-algebraic set-valued mapping $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ form a semi-algebraic set of dimension not exceeding $m - 1$.

In particular an extended-real valued semi-algebraic function can have at most finitely many critical values.

For the theory of semi-algebraic sets and mappings see [21, 175]. The Sard theorem was first proved by Bolte-Daniilidis-Lewis [22] for real-valued functions and then by Ioffe [90] for set-valued mappings (in both cases the theorems were stated for more general classes of objects - semi-analytic functions in [22] and arbitrarily stratifiable maps in [90]).

### 7.5.2 Transversality.

We are finally ready to extend transversality theory (not just the definition) beyond the smooth domain. To begin with, we observe that a direct extension of Proposition 1.12 does not hold if $F$ is not smooth.

**Example 7.37.** Consider the function

$$f(x, w) = |x| - |w|$$

viewed as a mapping from $\mathbb{R}^2$ into $\mathbb{R}$. This mapping is clearly semi-algebraic, even polyhedral. It is easy to verify that the mapping is regular at every point with the modulus of surjection identically equal to one (if we take the $\ell^\infty$ norm in $\mathbb{R}^2$). Furthermore

$$Q = f^{-1}(0) = \{(x, w) : |x| = |w|\}$$

and the restriction to $Q$ of the projection $(x, w) \to w$ is also a regular mapping with the modulus of surjection equal one. However, the partial mapping $x \to f(x, 0) = |x|$ is not regular at zero.
However, the following statement is true.

**Proposition 7.38 (93).** Let \( F : \mathbb{R}^m \times \mathbb{R}^k \rightrightarrows \mathbb{R}^n \) be a semi-algebraic set-valued mapping with locally closed graph, and let \( \overline{y} \in F(\overline{x}, \overline{p}) \). Assume that

1. \( F \) is regular at \((\overline{x}, \overline{p}, \overline{y})\);
2. the set-valued mapping \( \mathbb{R}^m \times \mathbb{R}^n \rightrightarrows \mathbb{R}^k \) associating the set \( \{ p : y \in F(x, p) \} \) with any \((x, y) \in \mathbb{R}^n \times \mathbb{R}^n\) is regular at \((\overline{x}, \overline{p}, \overline{y})\);
3. there is a Whitney stratification \((M_i)\) of Graph \( F \) such that the restriction of the projection \((x, p) \rightarrow p\) to the set \( S_i = \{(x, p) : (x, p, \overline{y}) \in M_i\} \), where \( M_i \) is the stratum containing \((\overline{x}, \overline{p}, \overline{y})\), is regular at \((\overline{x}, \overline{y})\).

Then \( F_p : x \mapsto F(x, \overline{p}) \) is regular at \((\overline{x}, \overline{y})\).

It is now possible to state and prove a set-valued version of Theorem 1.13.

**Theorem 7.39.** Let the mapping \( F : \mathbb{R}^n \times \mathbb{R}^k \rightrightarrows \mathbb{R}^m \) with closed graph and a closed set \( S \subset \mathbb{R}^m \) be both semi-algebraic. Denote by \( F_p \) the set-valued mapping \( x \mapsto F(x, p) \). If \( F \) is transversal to \( S \), then for all \( p \), with possible exception of a semi-algebraic set of dimension smaller than \( k \), \( F_p \) is transversal to \( S \).

**Proof.** The theorem is trivial if \( F(x, p) \cap S = \emptyset \) for all \((x, p)\), so we assume that \( F(x, p) \) meets \( S \) for some values of the arguments. Then \((0, 0)\) is a regular value of the mapping \( \Psi : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^k \rightarrow \mathbb{R}^m \), \( \Psi(x, y, p) = (F(x, p) - y) \times (S - y) \). Let \( Q = \Psi^{-1}(0, 0) \). This is a semi-algebraic set, so by Theorem 7.36 there is a semi-algebraic set \( C_0 \subset \mathbb{R}^k \) with \( \dim C_0 < k \) and every \( p \in \mathbb{R}^k \setminus C_0 \) is a regular value of the restriction \( \pi|_Q \) of the projection \((x, y, p) \mapsto p\).

Take an \( r > N + m - k \), and let \((M_i)_{i=1, \ldots, r}\) be a \( C^1 \)-Whitney stratification of Graph \( \Psi \) with all \( M_i \) being semi-algebraic manifolds. Then for any \( i \) there is a semi-algebraic set \( C_i \subset \mathbb{R}^k \) such that any \( p \in \mathbb{R}^k \setminus C_i \) is a regular value of \( \pi|_{M_i} \). The union \( C = \bigcup_{i=0} C_i \) is also a semi-algebraic set of dimension smaller than \( k \) and, as we have just seen, for any \( p \notin C \) all of the assumptions of Proposition 7.38 are satisfied for \( \Psi \). Therefore \((0, 0)\) is a regular value of \( \Psi_p \). By Proposition 6.13 this means that \( F_p \) is transversal to \( S \).

8 Some applications to analysis and optimization

In this section we give several examples illustrating the power of regularity theory as a working instrument for treating various problems in analysis and optimization. We do not try each time to prove the result under the most general assumptions. The purpose is rather to demonstrate how regularity considerations help to understand and/or simplify the analysis of one or another phenomenon. Again, it should be said that some interesting areas of application of metric regularity remain outside the scope of the paper. Just mention the role of regularity in numerical optimization (see e.g. 55, 109, 110) or connections with metric fixed point theory (e.g. 2, 50, 51, 92, 97) or recent developments associated with tilt stability, quadratic growth etc. (e.g. 2, 8, 54, 59, 109, 143) .
8.1 Subdifferential calculus

In each of the three calculus rules stated in Proposition 5.9 we assume one function Lipschitz. One of the reasons (especially important in the proof of the exact sum rule) is that Lipschitz functions have bounded subdifferentials. But what happens when both functions are not Lipschitz? For instance, what can be said about normal cone to an intersection of sets? As in the calculus of convex subdifferentials, we do need some qualification conditions to ensure the result.

**Theorem 8.1.** Let $X$ be a Banach space and $S_i$, $i = 1, 2$ are closed subsets of $X$. Let further $\bar{x} \in S = S_1 \cap S_2$. If $S_1$ and $S_2$ are subtransversal at $\bar{x}$, then

$$NG(S, \bar{x}) \subset NG(S_1, \bar{x}) + NG(S_2, \bar{x}).$$

Explicitly, this theorem was first mentioned in [SS] but de facto it was proved already in [S5] (see also [101], Proposition 3). It turns out that subtransversality is the most general of all so far available conditions that would guarantee the inclusion. The most popular subdifferential transversality condition (condition (b) of Theorem 6.12) may be much stronger.

The inclusion is among the most fundamental facts of the subdifferential calculus: enough to mention that in the majority of publications on the subject it is used as the starting point for deriving all other calculus rules. Below is a sketch of the proof of the theorem for the finite dimensional situation.

**Proof.** We need the following elementary and/or well known facts of functions on and sets in $\mathbb{R}^n$:

- $\hat{N}(Q, x) \cap B = \hat{d}(\cdot, Q)(x)$ if $x \in Q$;
- if $x^* \in \hat{d}(\cdot, Q)(x)$ and $u \in Q$ is the closest to $x$, then $x^* \in \hat{N}(Q, u)$;
- if $x \in Q$ and $f(\cdot)$ is nonnegative, equal to zero at $x$ and $f(u) \geq d(u, Q)$ in a neighborhood of $x$, then $\hat{d}(\cdot, Q)(x) \subset \hat{\partial} f(x)$.

Combining this with the definition of the limiting subdifferential, we conclude that for $Q$, $f$ and $x$ as above, $\hat{d}(\cdot, Q)(x) \subset \partial f(x) -$ the fact that is surprisingly missing from monographic publications.

By the assumption there is a $K > 0$ such that $d(x, S) \leq K(d(x, S_1) + d(x, S_2))$, so applying the above to $f(x) = K(d(x, S_1) + d(x, S_2))$ along with the exact calculus rule of Proposition we conclude that $\partial d(\cdot, S)(\bar{x}) \subset K(\partial(\cdot, S_1)(\bar{x}) + \partial(\cdot, S_1)(\bar{x}))$ and the result follows.

8.2 Necessary conditions in constrained optimization.

We discuss here two ways to apply regularity theory to necessary optimality conditions and then a general approach to necessary conditions associated with one of them. Both substantially differ from classical proofs that include linearization and separation as the major steps (see e.g. [60, 69, 103, 150, 152]). Verification of relevance of linearization is usually the central and most difficult part of the proofs. It is established under certain constraint qualifications which always imply and often are equivalent to regularity of the constraint mapping (as in case of the popular Mangasarian-Fromovitz and Slater...
qualification conditions) (see e.g. [150] where the connection with regularity was made explicit).

We refer to [113, 128, 130] for extensions of the classical approach to nondifferentiable optimization in which convex separation is replaced by an "extremal principle". The point is however that a fuller use of regularity arguments makes the way to necessary conditions much shorter. To begin with we shall consider the problem

\[
\text{minimize } f(x), \quad \text{s.t. } F(x) \in Q, \; x \in C
\]  

(8.1)

(where \( F : X \to Y \) is single-valued and \( Q \subset Y \) and \( C \subset X \) are closed sets) assuming for simplicity that both \( X \) and \( Y \) are finite dimensional although the results have been originally proved in much more general situations.

### 8.2.1 Non-covering principle.

So let \( \bar{x} \in C \) be a solution of the problem. Let \( \Psi \) stand for the restriction to \( C \) of the set-valued mapping \( x \mapsto \{ f(x) - IR \} \times (F(x) - Q) \) from \( X \) into \( Z = IR \times Y \). Clearly, this mapping cannot be regular near \( (\bar{x}, (f(\bar{x}), 0)) \). (Indeed, if \( U \) is a small neighborhood of \( \bar{x} \), then \( \Psi(U) \) cannot contain points \( (f(\bar{x}) - \varepsilon, 0) \).)

It follows that the negation of any condition sufficient for regularity is a necessary condition for \( \bar{x} \) to be a local solution in the problem. Applying Theorem 6.17 and Corollary 6.18 we get the following result.

**Theorem 8.2.** Assume that \( F : IR^n \to IR^m \) is Lipschitz in a neighborhood of \( \bar{x} \). If \( \bar{x} \) is a local solution of (8.1), then there is a nonzero pair \((\lambda, y^*)\) such that \( \lambda \geq 0 \), \( y^* \in N(Q, \bar{x}) \) and

\[
0 \in \partial(\lambda f + (y^* \circ F|_C))(\bar{x}).
\]  

(8.2)

This formulation needs some comments. We have stated the theorem in finite dimensions for simplicity, its infinite dimensional version can be found e.g. in [84]. Note further that a more customary formulation would be

\[
0 \in \partial(\lambda f + (y^* \circ F))(\bar{x}) + N(C, \bar{x}).
\]  

(8.3)

This condition is usually more convenient (constraints are separated) but in general weaker than (8.2). It is equivalent to (8.2) if e.g. \( C = X \) (obvious) or if both \( f \) and \( F \) are continuously differentiable and the constraint qualification

\[
0 \in F'(\bar{x})y^* + N_C(\bar{x}), \quad y^* \in N_Q(F(\bar{x})) \Rightarrow y^* = 0
\]  

(8.4)

is satisfied (see e.g. [159], Example 10.8) which means that \( F|_C \) is transversal to \( Q \) at \( \bar{x} \) (Proposition 7.38).

Finally, we observe that the necessary condition is stated in the Lagrangian form. Again, such condition can be substantially more precise than the "separated" condition \( 0 \in \lambda \partial f(\bar{x}) + \partial(y^* \circ F)(\bar{x}) \) (say in the absence of the constraint \( x \in C \)) which in various forms often appears in literature. Both conditions are equivalent if, say \( f \) is continuously differentiable.

The “non-covering” approach to necessary optimality condition was first applied probably by Warga [173] in a fairly classical setting of the standard optimal control problem.
Warga refers not to the Lyusternik-Graves theorem but to the result of Yorke [176] which is a weakened version of the theorem for integral operators associated with ordinary differential equations. But already the same year the controllability - optimality dichotomy appeared as the main tool of proving necessary conditions for nonsmooth optimal control in the papers by Clarke [37] and Warga [174]. In the context of an abstract optimization problem a non-covering criterion seems to have been first applied by Dmitruk-Milyutin-Osmolowski in [45] to problems with finitely many functional constraints and recently, to problems with mixed structure (partly smooth and partly close to convex), by Avakov, Magaril-Ilyaev and Tikhomirov [9]. In the next subsection 8.3 we demonstrate the work of this techniques for an abstract relaxed optimal control problem. Theorem 8.2 in an infinite dimensional setting was obtained in [84] with the same proof based on the non-covering criterion.

8.2.2 Exact penalty.

The immediate predecessor of the approach we are going to discuss here was the idea of an “exact penalty” offered by Clarke [35, 38]: if \( f \) attains a local minimum on a closed set \( S \) at \( x \in S \) and satisfies the Lipschitz condition near \( x \), then \( x \) is a point of unconstrained minimum of \( g(x) = f(x) + Kd(x, S) \) with \( K \) greater than the Lipschitz constant of \( f \) near \( x \). Clarke used a fairly sophisticated reduction technique to apply this idea to problems with functional constraints. The arguments however are dramatically simplified by direct invoking regularity considerations.

Let us return to the problem (8.1), assuming as above that \( F \) is single-valued Lipschitz \( X = \mathbb{R}^n, Y = \mathbb{R}^m \), and set as in Theorem 6.17

\[
\Phi(x) = \begin{cases} 
F(x) - Q, & \text{if } x \in C; \\
\emptyset, & \text{otherwise}. 
\end{cases}
\]

Then our problem can be reformulated as

\[
\text{minimize } f(x), \quad \text{s.t. } 0 \in \Phi(x). \tag{8.5}
\]

Suppose that \( \Phi \) is subregular at \((\bar{x}, 0)\). This means that there is some \( K_0 > 0 \) such that \( d(x, \Phi^{-1}(0)) \leq K_0 d(0, \Phi(x)) \) for \( x \) of a neighborhood of \( \bar{x} \). But \( \Phi^{-1}(0) \) is the feasible set of our problem, so that there is some other \( K_1 > 0 \) such that the function \( f(x) + K_1 d(0, \Phi(x)) \) attains local minimum at \( \bar{x} \) or equivalently, the function \( f(x) + K_1 d(y, F(x) - Q) \) attains a local minimum at \( \bar{x} \) subject to \( x \in C \). The last function is Lipschitz continuous near \( \bar{x} \), hence there is a \( K \) such that

\[
g(x) = f(x) + K(y, F(x) - Q) + d(x, C) \tag{8.6}
\]

attains an unconditional minimum at \( \bar{x} \).

If on the other hand, \( \Phi \) is nor subregular at \( \bar{x} \), Theorems 6.11 and 6.17 imply together that \( 0 \in \partial(y^* \circ F)(\bar{x}) + N(C, \bar{x}) \) for some nonzero \( y^* \in N(Q, F(\bar{x})) \). From here we easily get a weakened version of Theorem 8.2 with the Lagrangian condition replaced by its “separated” versions

\[
0 \in \partial f(\bar{x}) + \partial(y^* \circ F)(\bar{x}) + N(C, \bar{x}), \quad y^* \in N(Q, F(\bar{x})).
\]

71
This is a definite drawback, as we have already mentioned which however is counterbalanced by some serious advantages. First we note that $g$ is defined in terms of the original data which makes it possible to study higher order optimality conditions using this function. This is how such a techniques was used for the first time in [80] in order to get necessary optimality conditions earlier obtained by Levitin-Milyutin-Osmolowski in [119].

Another advantage is that the second approach is more universal. It can work for problems for which using scalarized coderivatives is either difficult or just impossible as say, in problems involving inclusions $0 \in \Phi(x)$ with general set-valued $\Phi$. This is a typical case in optimal control of dynamic systems described by differential inclusions. Loewen [125] was the first to use this approach to prove a maximum principle in a free right end point problem of that sort. The analytic challenge in his proof was to find an upper estimate for the distance to the feasible set. However the next step in the development, the "optimality alternative" discussed below, excludes even any need in such an estimate.

8.2.3 Optimality alternative.

Consider the abstract problem with $(X,d)$ being a complete metric space:

$$\text{minimize } f(x), \text{ subject to } x \in Q \subset X.$$ 

**Theorem 8.3.** Let $\varphi$ be a nonnegative lsc function on $X$ equal to zero at $\bar{x}$. If $\bar{x} \in Q$ is a local solution to the problem, then the following alternative holds true:

- either there is a $\lambda > 0$ such that the function $\lambda f + \varphi$ has an unconstrained local minimum at $\bar{x}$;
- or there is a sequence $(x_n) \to \bar{x}$ such that $\varphi(x_n) < n^{-1}d(x_n, Q)$ and the function $x \mapsto \varphi(x) + n^{-1}d(x, x_n)$ attains a local minimum at $x_n$ for each $n$.

We shall speak about regular case if the first option takes place and singular or non-regular case otherwise.

**Proof.** Indeed, either there is an $R > 0$ such that $R\varphi(x) \geq d(x, Q)$ for all $x$ of a neighborhood of $\bar{x}$, or there is a sequence $(z_n)$ converging to $\bar{x}$ and such that $n^2\varphi(z_n) < d(z_n, Q)$.

In the first case (as $f$ is Lipschitz) we have for $x \notin Q$ and $u \in Q$ close to $x$ (so that e.g. $d(x, u) < 2d(x, Q)$):

$$f(x) \geq f(u) - Ld(x, u) \geq f(\bar{x}) - 2LR\varphi(x),$$

if $L$ is a Lipschitz constant of $f$.

As $X$ is complete and $\varphi$ is lower semicontinuous, we can apply Ekeland’s principle to $\varphi$ (taking into account that $\varphi(z_n) < \inf \varphi + n^{-2}d(z_n, Q)$) and find $x_n$ such that $d(x_n, z_n) \leq n^{-1}d(z_n, Q)$, $\varphi(x_n) \leq \varphi(z_n)$ and $\varphi(x) + n^{-1}d(x, x_n) > \varphi(x_n)$ for $x \neq x_n$. We have finally

$$d(x_n, Q) \geq d(z_n, Q) - d(x_n, z_n) \geq (1 - n^{-1})d(z_n, Q) \geq (1 - n^{-1})n^2\varphi(z_n) \geq n\varphi(x_n)$$

as claimed.

Thus, a constrained problem reduces to one or a sequence of unconstrained minimization problems. Hopefully, such problems can be easier to analyze thanks to the freedom of choosing $\varphi$ which we call test function in the sequel. Even before the alternative was
explicitly stated it was de facto used to prove the maximum principle in various problems of optimal control \[72, 86, 170\]. Here is a brief account of how the alternative works for optimal control of systems governed by differential inclusions.

### 8.2.4 Optimal control of differential inclusion.

As the first example of application of the alternative we shall briefly consider the following problem of optimal control of a system governed by differential inclusion (see also the next subsection 8.3): minimize

$$\ell(x(0), x(T))$$

on trajectories of the differential inclusion

$$\dot{x} \in F(t, x),$$

satisfying the end point condition

$$(x(0), x(T)) \in S.$$ (8.9)

The natural space to treat the problem is $W_{1}^{1}$. Let $x(\cdot)$ be a local solution. For any $x(\cdot) \in W_{1}^{1}$ set

$$\varphi(x(\cdot)) = \int_{0}^{T} d(\dot{x}(t), F(t, x(t)))dt + d((x(0), x(T)), S).$$

Clearly, $\varphi$ is nonnegative and $\varphi(\overline{x}(\cdot)) = 0$. Thus, if $\ell$ is a Lipschitz function, we can apply the alternative to get necessary optimality condition. According to the alternative, either there is a $\lambda > 0$ such that $\overline{x}(\cdot)$ is a local minimum of

$$\lambda \ell(x(0), x(T)) + d((x(0), x(T)), S) + \int_{0}^{T} d(\dot{x}(t), F(t, x(t)))dt,$$

or there is a sequence $(x_n(\cdot))$ converging to $\overline{x}(\cdot)$ such that every $x_n(\cdot)$ is not feasible in $(8.7)$-$(8.9)$ and is a local minimum of the functional

$$d((x(0), x(T)), S) + \int_{0}^{T} d(\dot{x}(t), F(t, x(t)))dt + n^{-1}\left(\|x(0) - x_n(0)\| + \int_{0}^{T} \|\dot{x}(t) - \dot{x}_n(t)\|dt\right).$$

In both cases we get an (unconstrained) Bolza problem. Analysis of such problem needs different techniques and we refer to \[86, 170\] where necessary optimality conditions for the problem were obtained along these lines. A more general result was established a few years later by Clarke \[39\] (actually the most general for optimal control of differential inclusions so far) but a shorter proof of Clarke’s theorem based on optimality alternative is now also available \[58\].

To conclude, I wish to note that this is not the only possible application of regularity related ideas to optimal control. We can refer to \[171\] for the discussion of the role of metric regularity in the Hamilton-Jacoby theory of optimal control.
8.2.5 Constraint qualification.

The last question we intend to briefly discuss in this subsection concerns constraint qualifications in optimization problems. They often play an important role in proofs but their basic function is to guarantee that the multiplier $\lambda$ of the cost function is in the necessary (e.g. Lagrangian) optimality conditions is positive. The point is that constraint qualifications are often connected with regularity properties of the constraint mapping. We shall discuss just one example.

Let us say that the problem is normal at a certain feasible point if the constraint mapping is regular at the point. The problem is normal if either the feasible set is empty or the problem is normal at every feasible point. In the case of the problem (8.1) the constraint mapping is the restriction of $F$ to $C$, so by Theorem 6.17 normality is guaranteed if $F$ is transversal to $Q$, that is if $y^* \in N(Q, F(x))$ and $0 \in D^*F|_C(\bar{r}, 0)(y^*)$ imply together that $y^* = 0$ which in turn imply that

$$0 \in \partial(y^* \circ F)(x) + N(C, x), \quad y^* \in N(Q, F(x)) \Rightarrow y^* = 0. \quad (8.10)$$

This is the now standard constrained qualification in nonsmooth optimization (see e.g. [55, 109, 130, 159]). If $f$ and $F$ are continuously differentiable and the sets $C$ and $Q$ are convex, (8.10) is dual to Robinson’s constraint qualification [150].

8.3 An abstract relaxed optimal control problem.

Here we apply the optimality alternative to get necessary optimality condition in the problem

$$\text{minimize } f(x) \quad \text{s.t. } F(x, u) = 0, \; u \in U. \quad (8.11)$$

Here $F : X \times U \to Y$, $X$ and $Y$ are separable Banach spaces and $U$ is a set. The problem is similar to problems with mixed smooth and convex structures studied in [103, 165]. But contrary to [103, 165], here we do not assume that $F$ is continuously differentiable in $x$. We shall formulate the requirements on $F$ a bit later. First we need to introduce and discuss some necessary concepts.

We say that a continuous mapping $F : X \to Y$ is semi-Fredholm at $\bar{x}$ it has at $\bar{x}$ a strict prederivative of the form $\mathcal{H}(x) = Ax + \|h\|Q$, where $A : X \to Y$ is a linear bounded operator that send $X$ onto a closed subspace of $Y$ of finite codimension and $Q \subset Y$ is a compact set (that can be assumed convex and symmetric). We say furthermore that $S \subset X$ is finite-dimensionally generated if $S = \Lambda^{-1}(P)$ where $\Lambda : X \to \mathbb{R}^n$ is a continuous linear operator and $P \subset \mathbb{R}^n$ is closed.

**Proposition 8.4** (non-covering principle for (8.11) [84, 72]. Let $F : X \to Y$ be semi-Fredholm at $\bar{x}$, and let $S$ be a finite-dimensionally generated subset of $X$. Let further $F|_S$ be the restriction of $F$ to $S$, that is the set-valued mapping equal to $\{F(x)\}$ on $S$ and $\emptyset$ outside of $S$. If $F|_S$ is not regular near $\bar{x}$, then there is a $y^* \neq 0$ such that $0 \in \partial G(y^* \circ F)(\bar{x}) + N_G(S, \bar{x})$. Moreover, the weak$^*$-closure of the set of such $y^*$ with norm 1 does not contain zero.\(^3\)

\(^3\)More general versions of this result can be found in many publications related to “point estimates” and compactness properties of subdifferentials - see e.g [85, 105, 106, 108, 130].
We intend to use this principle to prove the following theorem.

**Theorem 8.5.** Let \((\overline{x}, \overline{y})\) be a solution of (8.4). We assume that
(A1) \(f\) satisfies the Lipschitz condition in a neighborhood of \(\overline{x}\);
(A2) for any \(u \in U\) the mapping \(F(\cdot, u)\) is Lipschitz in a neighborhood of \(\overline{x}\), and \(F(\cdot, \overline{u})\) is semi-Fredholm at \(\overline{x}\);
(A3) \(F(x, U)\) is a convex set for any \(x\) of a neighborhood of \(\overline{x}\);
(A4) \(S\) is finite-dimensionally generated.

Let further \(\mathcal{L}(\lambda, y^*, x, u) = \lambda f(x) + \langle y^*, F(x, u) \rangle\) be the Lagrangian of the problem. Then there are \(\lambda \geq 0\) and \(y^* \in Y^*\) such that the following relations hold true:

\[
\begin{align*}
\lambda + \|y^*\| & > 0 \quad \text{(non-triviality)}; \\
0 & \in \partial_G \mathcal{L}(\lambda, y^*, \cdot, \overline{u})(\overline{x}) + N_G(S, \overline{x}) \quad \text{(Euler-Lagrange inclusion)}; \\
\langle y^*, F(x, \overline{u}) \rangle & \geq \langle y^*, F(x, u) \rangle, \quad \forall u \in U \quad \text{(the maximum principle)}. \\
\end{align*}
\]

**Proof.** Given a finite collection \(U = (u_1, \ldots, u_k)\) of elements of \(U\), we define a mapping \(\Phi_U : \overline{x} \times \mathbb{R}^k \rightarrow Y\) by

\[
\Phi_U(x, \alpha_1, \ldots, \alpha_k) = F(x, \overline{u}) + \sum_{i=1}^{k} \alpha_i (F(x, u_i) - F(x, \overline{u})).
\]

It is an easy matter to see that this mapping is also semi-Fredholm at \((\overline{x}, 0)\).

Consider the problem

**minimize** \(f(x)\) \(\text{s.t.} \ \Phi_U(x, \alpha_1, \ldots, \alpha_k) = 0, \ x \in S, \ \alpha_i \geq 0. \) \((P_U)\)

Then \((\overline{x}, 0, \ldots, 0)\) solves the problem (as immediately follows from (A3)). Let further \(\Psi : \overline{x} \times \mathbb{R}^k \rightarrow Y\) be defined by

\[
\Psi(x, \alpha_0, \ldots, \alpha_k) = (f(x) + \alpha_0, \Phi_U(x, \alpha_1, \ldots, \alpha_k)).
\]

This mapping cannot be regular in a neighborhood of \((\overline{x}, 0, \ldots, 0)\) because no point \((f(\overline{x}) - \varepsilon, 0, \ldots, 0)\) can be a value of \(\Psi\) at \(x \in S\) close to \(\overline{x}\) and \(\alpha\) close to zero. It is an easy matter to verify that \(\Psi\) is also semi-Fredholm at \((\overline{x}, 0, \ldots, 0)\) and we can apply Proposition 8.4.

Set \(\hat{S} = S \times \mathbb{R}^k_{-1}, \ \hat{\mathcal{L}}(\lambda, y^*, x, \alpha_0, \ldots, \alpha_k) = \lambda f(x) + \alpha_0 + \langle y^*, \Psi(x, \alpha_0, \ldots, \alpha_k) \rangle\). By the proposition there are multipliers \((\lambda, y^*) \neq 0\) such that

\[
0 \in \partial_G \hat{\mathcal{L}}(\lambda, y^*, \cdot, 0, \ldots, 0) + N_G(\hat{S}, (\overline{x}, 0, \ldots, 0)).
\]

We have (using the standard rules of subdifferential calculus)

\[
\begin{align*}
N_G(\hat{S}, (\overline{x}, 0, \ldots, 0)) &= N_G(\overline{x}, S) \times \mathbb{R}^k_{-1}, \\
\partial_G \hat{\mathcal{L}}(\lambda, y^*, \cdot, 0, \ldots, 0) &\subset \partial_G \mathcal{L}(\lambda, y^*, \cdot, \overline{u})(\overline{x}) \\
&\quad + (\lambda, \langle y^*, F(x, u_1) - F(x, \overline{u}) \rangle, \ldots, \langle y^*, F(x, u_k) - F(x, \overline{u}) \rangle)).
\end{align*}
\]

It follows that there are \(\xi_i \leq 0, \ i = 0, \ldots, k\) such that

\[
\begin{align*}
0 & \in \partial_G \mathcal{L}(\lambda, y^*, \cdot, \overline{u})(\overline{x}) + N_G(S, \overline{x}); \\
\lambda &= -\xi_0 \geq 0; \\
\langle y^*, F(x, u_i) - F(x, \overline{u}) \rangle &= \xi_i \geq 0, \quad i = 1, \ldots, k.
\end{align*}
\]
The relations remain obviously valid if we replace \( \lambda, y^* \) by \( r\lambda, ry^* \) with some positive \( r \). Thus for any finite collection \((u_1, \ldots, u_k) \subset U \) we can find a pair of multipliers \((\lambda, y^*)\) satisfying the three above mentioned relations and the normalization condition \( \lambda + \|y^*\| = 1 \). Let \( \Omega(u_1, \ldots, u_k) \) be the weak*-closure of all such pairs. Then \( \Omega(u_1, \ldots, u_k) \) is weak*-compact and by Proposition \[8.4\] does not contain zero. It remains to notice that the increase of the set \((u_1, \ldots, u_k)\) may result only in decrease of \( \Omega(u_1, \ldots, u_k) \) and therefore there is a nonzero pair \( \lambda, y^* \) common to all sets \( \Omega(u_1, \ldots, u_k) \).

\[\square\]

### 8.4 Genericity in tame optimization.

Here by “tame optimization” we mean optimization problems with semi-algebraic data. We consider the same class of problems as in \[8.1\]. This time however we are interested in the effects of perturbations and shall work with a family of problems depending on a parameter \( p \):  

\[
\text{minimize } f(x, p), \quad \text{s.t. } F(x, p) \in Q, \; x \in C.
\]

(8.12)

Here \( x \) is an argument in the problem and \( p \) is a parameter. So subdifferentials and derivatives that will appear below are always with respect to \( x \) alone. If \( p \) is fixed, then we denote the corresponding problem by \( \mathcal{P}_p \).

Before we continue, we have to mention that for a semi-algebraic set \( S \subset \mathbb{R}^n \) the properties

- \( S \) is a set of first Baire category in \( \mathbb{R}^n \);
- \( S \) has \( n \)-dimensional Lebesgue measure zero;
- \( \dim S < n \)

are equivalent. Thus, when we deal with semi-algebraic objects e.g. in \( \mathbb{R}^k \), the word “generic” means “up to a semi-algebraic set of dimension smaller than \( k \).”

We shall assume that \( p \) is taken from an open set \( P \subset \mathbb{R}^k \) and, as before, \( x \in \mathbb{R}^n \) and \( F \) takes values in \( \mathbb{R}^m \). Our main assumption is that the restriction \( F|_{C}(x, p) \) of \( F \) to \( C \) is transversal to \( Q \).

This is definitely the case when \( k = m \) and \( F(x, p) = F(x) - p \). As to \( F \) itself, we assume that it is continuous with respect to \((x, p)\) and locally Lipschitz in \( x \). The sets \( C \) and \( Q \) as usual are assumed closed.

**Theorem 8.6** (generic normality). Under the stated assumptions for a generic \( p \in P \), the mapping \( F|_{C}(\cdot, p) \) is transversal to \( Q \). Thus for a generic \( p \) the problem \( \mathcal{P}_p \) is normal.

**Proof.** The first statement is immediate from Theorem \[7.39\] while the second from the comments following the statement of Theorem \[6.17\].

Let us call a point \( x \) feasible in \( \mathcal{P}_p \) a *critical point* of the problem if the non-degenerate Lagrangian necessary condition of 8.2.1

\[
0 \in \partial(f + (y^* \circ F|_{C}))(x, p), \quad y^* \in N(Q, F(x, p))
\]

is satisfied. In this case the value of \( f \) at \( x \) is called a *critical value* of \( \mathcal{P}_p \).
Theorem 8.7 (generic finiteness of critical values). If under the stated assumptions, \( \mathcal{P}_p \) is normal, then the problem may have only finitely many critical values. Thus there is an integer \( N \) such that for a generic \( p \) the number of critical values in the problem does not exceed \( N \).

Proof. Consider the function

\[ L_p(x, y, y^*) = f(x, p) + \langle y^*, F|_C(x, p) - y \rangle + i_Q(y). \]

As follows from the standard calculus rules,

\[ \partial L_p(x, y, y^*) = \partial(f + y^* \circ F|_C)(x, p) \times (N(Q, y) - y^*) \times \{F(x, p) - y\}. \]

Thus, \((x, y, y^*)\) is a critical point of \( L_p \) if and only if \( F(x, p) = y, 0 \in N(Q, y) - y^* \), that is \( y \in Q \) and \( y^* \in N(Q, y) \), and \( 0 \in \partial(f + y^* \circ F|_C)(x, p) \). In other words, \((x, y, y^*)\) is a critical point of \( L_p \) if and only if \( x \) is a feasible point in \( (P) \), \( y = F(x, p) \) and the necessary optimality condition is satisfied at \( x \) with \( y^* \) being the Lagrange multiplier. We also see that in this case \( L_p(x, y, y^*) = f(x, p) \). In other words, critical values of the problem are precisely critical values of \( L \).

By the Sard theorem \( L_p \) may have at most finitely many critical values, whence the theorem.

The last result we are going to present here has been so far proved only under some additional assumptions on elements of the problem. We shall explain it for the classical case, although semi-algebraic nature of the data remains crucial.

Theorem 8.8 (generic finiteness of critical points). Assume that \( p = (q, y) \) with \( q \in \mathbb{R}^n \) and \( y \in \mathbb{R}^m \), \( f(x, p) = f(x) - \langle q, x \rangle \), \( F(x, p) = F(x) - y \) with \( f(x) \) and \( F(x) \) both continuously differentiable. Assume further that the sets \( C \) and \( Q \) are closed and convex. Then there is an integer \( N \) such that for a generic \( p \) the number of pairs \((x, y^*)\), such that \( x \) is a critical point in \( \mathcal{P}_p \) and \( y^* \) a corresponding Lagrange multiplier does not exceed \( N \).

The theorem follows from the two results below that contain valuable information about geometry of subdifferential mappings of semi-algebraic functions.

Proposition 8.9 (dimension of the subdifferential graph [58]). The dimension of the graph of the subdifferential (no matter which, Fréchet, limiting or Clarke) mapping of a semi-algebraic function on \( \mathbb{R}^n \) is \( n \).

Proposition 8.10 (finiteness of preimage [93, 58]). Let \( F : \mathbb{R}^n \rightarrow \mathbb{R}^n \) be a semi-algebraic set-valued mapping such that \( \dim \text{Graph } F \leq n \). If \( y \) is a regular value of \( F \), then \( F^{-1}(y) \) contains at most finitely many elements. Thus, there is an integer \( N \) such that for a generic \( y \) the number of elements in \( F^{-1}(y) \) cannot exceed \( N \).

To see how the propositions lead to the proof of the theorem, we note first that \( D^*F|_C(x)(y^*) = F'(x)y^* + N_C(x) \) if \( x \in C \), \( F \) is smooth and \( C \) convex. By Theorem 8.15 \( F|_C \) is transversal to \( Q \) if and only if

\[ x \in C, \ F(x) \in Q + y, \ 0 \in F'(x)y^* + N_C(x), \ y^* \in N(Q, F(x) - y) \Rightarrow y^* = 0, \quad (8.13) \]
and by Theorem 7.39 this holds for a generic \( y \).

Consider the function
\[
g(x, y) = f(x) + i_{C}(x) + i_{Q}(F(x) - y)
\]
By Proposition 8.9 the dimension of the graph of its subdifferential is \( n + m \). Then so is the dimension of the graph of the mapping
\[
\Gamma(x, y^{*}) = \{(q, y) : (q, y^{*}) \in \partial g(x, y)\}.
\]
Now by the Sard theorem generic \((q, y)\) is a regular value of \( \Gamma \) so (Proposition 8.10) for a generic \((q, y)\) there are finitely many \((x, y^{*})\) such that \((q, y^{*}) \in \Gamma(x, y^{*})\). Finally, if for such \((q, y)\) the qualification condition (8.13) is satisfied, then
\[
\partial g(x, y) = \{(q, y^{*}) : f'(x) + (y^{*} \circ F(\cdot))'(x) + N(C, x), y^{*} \in N(Q, F(x) - y)\}
\]
(even if \( Q \) is not convex - see again Exercise 10.8 in [159]) which in particular means that \( x \) is a critical point of \( \mathcal{P}_{p} \) and \( y^{*} \) is a Lagrange multiplier in the problem.

**8.5 Method of alternating projection.**

This is one of the most popular methods to solve feasibility problem due to its simplicity and efficiency. The feasibility problem in its simplest form consists in finding a common point of two sets, say \( Q \) and \( S \). The recipe offered by the method of alternating projection is the following: starting with a certain \( x_{0} \), we choose for \( k = 0, 1, \ldots \)
\[
x_{2k+1} \in \pi_{Q}(x_{2k}), \quad x_{2k+2} \in \pi_{S}(x_{2k+1}),
\]
where \( \pi_{Q}(x) \) is the collection of points of \( Q \) closest to \( x \) etc..

Von Neumann was the first to show in mid-30s (see [172]) that in case of two subspaces the method converges to a certain point in the intersection of two closed subspaces in a Hilbert space (depending of course on the starting point). Later in the 60s Bregman [28] and Gubin-Polyak-Raik [75] applied it to convex subsets in \( \mathbb{R}^{n} \). In particular it was shown in [75] that the convergence is linear if relative interiors of the sets meet. Later Bauschke and Borwein [19] proved linear convergence if the sets are subtransversal at any common point.

But in computational practice the method was successfully applied even for nonconvex sets. The first explanation was given by Lewis, Luke and Malik [122]: if at a certain point \( \mathbf{F} \) in the intersection the sets are transversal and at least one of the sets is not “too non-convex” in a certain sense (super-regular in the terminology of the authors) then linear convergence of alternating projections to a certain point common to the sets (not necessarily \( \mathbf{F} \)) if the starting point is sufficiently close to \( \mathbf{F} \). And very recently it was shown by Druziatskyj, Ioffe and Lewis [57] that transversality alone guarantees linear convergence. In fact linear convergence was proved in [57] under a substantially weaker condition of “intrinsic transversality” of the sets, but we believe that geometric essence of the phenomenon is captured by the transversality \( \Rightarrow \) linear convergence implication. The question whether linear convergence is guaranteed by subtransversality, as in the convex case, remains open (see [77]).
Here is a short proof of linear convergence under the transversality assumption. Set
\[ \varphi(x, y) = i_Q(x) + i_S(y) + \|x - y\|. \]

We claim that if \( Q \) and \( S \) are transversal at \( \bar{x} \in Q \cap S \), then there are \( \kappa > 0 \) and \( \delta > 0 \) such that for any \( x \in Q, y \in S \) close to \( \bar{x} \)
\[ \max\{|\nabla \varphi(\cdot, y)|(x), |\nabla \varphi(x, \cdot)|(y)\} \geq \kappa. \]

To this end, we first note that by Theorem 6.12
\[ \theta = \sup\{\langle u, v \rangle : u \in N(Q, \bar{x}), v \in -N(S, \bar{x}), \|u\| = \|v\| = 1\} < 1. \]

Fix a certain \( \kappa \in (0, 1) \) and assume that there are sequences \( (x_n) \subset Q, (y_n) \subset S, x_n \neq y_n \), converging to \( \bar{x} \) and such that
\[ |\nabla \varphi(\cdot, y_n)|(x_n) < \kappa, \quad |\nabla \varphi(x_n, \cdot)|(y_n) < \kappa, \]
that is the functions
\[ x \mapsto \varphi(x, y_n) + \kappa\|x - x_n\| \quad \text{and} \quad y \mapsto \varphi(x_n, y) + \kappa\|y - y_n\| \quad (8.14) \]
attain local minima respectively at \( x_n \) and \( y_n \). This means that
\[ 0 \in w_n^* + \frac{x_n - y_n}{\|x_n - y_n\|} + \kappa B; \quad 0 \in z_n^* + \frac{y_n - x_n}{\|x_n - y_n\|} + \kappa B \quad (8.15) \]
for some \( w_n^* \in N(Q, x_n) \) and \( z_n^* \in N(S, y_n) \). Thus, for any limit point \( (w^*, z^*) \) of \( (w_n^*, z_n^*) \), we have
\[ w^* = e + a, \quad z^* = -e + b, \]
where \( \|e\| = 1, \|a\| \leq \kappa, \|b\| \leq \kappa \). Consequently
\[ \theta \geq \frac{\langle e + a, e + b \rangle}{\|e + a\|\|e + b\|} \geq \frac{(1 - \kappa)^2}{(1 + \kappa)^2} \]
and we get
\[ \kappa \geq \frac{1 - \sqrt{\theta}}{1 + \sqrt{\theta}}. \quad (8.16) \]

This proves the claim.

Then \( \pi_Q(y) = \arg\min \varphi(\cdot, y) \) and the method of alternating projections can be written as follows:
\[ x_{n+1} \in \arg\min \varphi(x_n, \cdot); \quad x_{n+2} \in \arg\min \varphi(\cdot, x_{n+1}). \]

We obviously have \( |\nabla \varphi(x_n, \cdot)|(x_{n+1})| = 0 \). For a given \( x \) (not necessarily in \( Q \)), consider the function \( \psi_x(y) = i_S(y) + \|x - y\|. \) For any \( c \in (0, 1) \) condition \( |\nabla \psi_x|(x_{n+1}) \leq c \) obviously holds if
\[ \langle x - x_{n+1}, x_n - x_{n+1} \rangle \geq \sqrt{1 - c^2}\|x - x_{n+1}\|\|x_n - x_{n+1}\|. \]
Take a $c < \kappa$, and let $K_c$ be the collection of $c$ satisfying the above inequality. This is an ice-cream cone with vertex at $x_{n+1}$. If $x \in Q \cap K_c$, then $\nabla \varphi(\cdot, x_{n+1})(x) \geq \kappa > c$. On the other hand, as is easy to see, the distance from $x_n$ to the boundary of $K_c$ is precisely $cr$, where $r = \|x_n - x_{n+1}\|$. Applying Basic lemma for error bounds (Lemma 7.11), we conclude that there is an $x \in Q$ with $\varphi(x, x_{n+1}) \leq \varphi(x_n, x_{n+1}) - c\kappa\|x_{n+1} - x_n\|$. It follows that
\[ \|x_{n+2} - x_{n+1}\| = \varphi(x_{n+2}, x_{n+1}) \leq (1 - c^2)\|x_{n+1} - x_{n+1}\| \]
which is linear convergence of $(x_n)$.

### 8.6 Generalized equations.

By a generalized equation we mean the relation
\[ 0 \in f(x) + F(x), \]
where $f$ is a single-valued and $F : X \rightrightarrows Y$ a set-valued mapping. Variational inequalities and necessary optimality conditions in constraint optimization with smooth cost and constraint functions are typical examples. The problem discussed in the theorem below is what happens with the set of solutions of the generalized equation if the single-valued term is slightly perturbed.

**Theorem 8.11** (implicit function for generalized equations). *Let $X, Y$ be metric spaces, and let $Z$ be a normed space. Consider the generalized equation*
\[ 0 \in f(x, p) + F(x), \]  
\[ (8.17) \]
*where $f : X \times P \to Z$ and $F : X \rightrightarrows Z$. Let $(\overline{x}, \overline{p})$ be a solution to the equation. Set $\overline{z} = -f(\overline{x}, \overline{p})$ and suppose that the following two properties hold:*

(a) Either $X$ or the graph of $F$ is complete in the product metric and $F$ is regular near $(\overline{x}, \overline{z})$ with $\text{sur} F(\overline{x}, \overline{z}) > r$;

(b) there is a $\rho > 0$ such that $f$ is continuous on $\overline{B}(\overline{x}, \rho) \times \overline{B}(\overline{p}, \rho)$ and $f(\cdot, p)$ satisfies on $\overline{B}(\overline{x}, \rho)$ the Lipschitz condition with constant $\ell < r$ for all $p \in \overline{B}(\overline{p}, \rho)$.

*Let $S(p)$ stand for the solution mapping of $(8.17)$. Then*
\[ d(x, S(p')) \leq (r - \ell)^{-1}\|f(x, p) - f(x, p')\|. \]
*if $x \in S(p)$ is close to $\overline{x}$ and $p, p'$ are sufficiently close to $\overline{p}$. Thus, if $f(x, \cdot)$ satisfies the Lipschitz condition with constant $\alpha$ on a neighborhood of $\overline{p}$ for all $x \in \overline{B}(\overline{x}, \rho)$, then $S(\cdot)$ has the Aubin property near $(\overline{p}, \overline{x})$ with $\text{lip} S(\overline{p}; \overline{x}) \leq \alpha(r - \ell)^{-1}$.

*Finally, if in addition $F$ is strongly regular near $(\overline{x}, \overline{z})$, then $S(\cdot)$ has a Lipschitz localization $s(\cdot)$ at $(\overline{x}, \overline{y})$ with Lipschitz constant not greater than $\alpha(r - \ell)^{-1}$, so that*
\[ d(s(p), s(p')) \leq (r - \ell)^{-1}\|f(s(p), p) - f(s(p), p')\| \leq \alpha(r - \ell)^{-1}d(p, p'). \]

Note that in view of Theorem 8.10 condition (a) is equivalent to the assumption that there are $r > 0$ and $\xi > 0$ such that $|\nabla \xi \varphi_\xi|(x, v) > r$ (where $\varphi_\xi(x, v) = d(z, v) + \mathcal{G}_F(x, v)$) if e.g. $d(z, \overline{p}) < \rho$, $\|z\| < \rho$ and $z \neq v \in F(x)$.
Proof. Set \( G(x,p) = f(x,p) + F(x) \) and let \( H(p,z) = (G(\cdot,p))^{-1}(z) \), so that \( S(p) = H(p,0) \). As the Lipschitz constants of functions \( f(\cdot,p) \) are bounded by the same \( \ell \) for all \( p \in B(\bar{p}, \rho) \), it follows from Theorem 4.13 that there is a \( \delta > 0 \) such that for every \( p \in B(\bar{p}, \rho) \) the inequality \( d(x,H(p,z)) \leq (r - \ell)^{-1}d(z,G(x,p)) \) holds if \( d(x,\bar{x}) < \delta \) and \( \|z - z(p)\| < \delta \), where \( z(p) = f(\bar{x},p) - f(\bar{x},\bar{p}) \in G(\bar{x},p) \). As \( f \) is continuous, we can choose \( \lambda > 0 \) such that \( \|z(p)\| < \delta \) for \( p \in B(\bar{p}, \lambda) \). For such \( p \) we have \( 0 \in B(z(p), \delta) \) and therefore if \( d(\bar{p}, p') < \lambda \), we get, taking into account that \( 0 \in f(x,p) + F(x) \) by the assumption,

\[
d(x,S(p')) \leq (r - \ell)^{-1}d(0,G(x,p')) = (r - \ell)^{-1}d(0,f(x,p') + F(x)) = (r - \ell)^{-1}d(-f(x,p'), F(x)) \leq (r - \ell)^{-1}\|f(x,p') - f(x,p)\|
\]

This proves the first part of the theorem. The second now follows from Theorem 4.13.

The concept of generalized equation was introduced by Robinson in \([153]\). The theorem proved in \([153, 154]\) corresponded to \( f \) continuously differentiable in \( x \) and \( F \) being either a maximal monotone operator or \( F(x) = N(C, x) \), where \( C \) is a closed convex set. We refer to \([55]\) for further results and bibliographic comments on generalized equations which is one of the central objects of interest in the monograph.

An earlier version of part (a) of the theorem with a less precise estimate can be found in \([109]\) (Theorem 4.9). Part (b) of the theorem relating to strong regularity is the basic statement of Theorem 5F.4 of \([55]\) (generalizing the earlier results of Robinson in \([151, 155]\); see also \([48]\) for an earlier result). Our proof however is different: here the theorem appears as a direct consequence of Milyutin’s perturbation theorem. Note that in most of the related results in \([55]\) it is assumed (following \([155]\)) that there exists a “strict estimator \( h(x) \) for \( f \) of modulus \( \ell' \)” such that \( \operatorname{sur}(F + h)(x\bar{y} + h(\bar{y})) \geq r \). This is a fairly convenient device for practical purpose but it adds no generality to the result as the case with \( h \) reduces to the setting of the theorem if we replace \( F + h \) by \( F \) and \( f - h \) by \( f \).

### 8.7 Variational inequalities over polyhedral sets.

Variational inequality is a relation of the form

\[
0 \in \varphi(x) + N(C, x),
\]

where \( \varphi : \mathbb{R}^n \to \mathbb{R}^n \) is a single-valued mapping and \( C \subset \mathbb{R}^n \) is a convex set. If \( C \) is a cone, it is equivalent to

\[
x \in K, \quad F(x) \in K^\circ, \quad \langle x, F(x) \rangle = 0.
\]

The problem of finding such an \( x \) is known as a complementarity problem (see e.g. \([65]\)). Problems of this kind typically appear in nonlinear programming in connection with necessary optimality conditions.

Consider for instance the problem

\[
\text{minimize } f_0(x) \quad \text{s.t } f_i(x) \leq 0, \quad i = 1, \ldots, k, \quad f_i(x) \leq 0, \quad i = k + 1, \ldots, m.
\]
with \( f_0, \ldots, f_m \) twice continuously differentiable. If \( \varphi \) is a solution of the problem, then (assuming that the problem is normal and setting \( f = (f_1, \ldots, f_m) \)) there is a \( \bar{y} \in \mathbb{R}^m \) such that

\[
\nabla f_0(\bar{x}) + (\bar{y}, \nabla f(\bar{x})) = 0.
\]

Setting

\[
\varphi(x, y) = \left( \nabla f_0(x) + (\bar{y}, \nabla f(x)), f(x) \right); \quad C = \mathbb{R}^n \times \mathbb{R}_+^m,
\]

we see that \((\bar{x}, \bar{y})\) solves \((8.18)\) (with \( x \) replaced by \((x, y)\)).

Consider the set valued mapping \( \Psi(x) = \varphi(x) + \mathcal{N}(C, x) \) associated with \((8.18)\) assuming that \( C \) is a convex polyhedral set. What can be said about regularity of such mapping near a certain \((\bar{x}, \bar{y})\) \( \in \text{Graph } \Phi \)? Applying Milyutin’s perturbation theorem (Theorem 4.5) and Theorem 4.11 and taking into account that the Lipschitz constant of \( h \to \varphi(x + h) - \varphi'(x)h \) at zero is zero, we immediately get

Proposition 8.12. Let \( \bar{y} \in \Psi(x) \) for some \( x \in C \). Set \( \Phi(x) = \varphi(x) + \mathcal{N}(C, x) \). Then \( \Psi \) is (strongly) regular near \((\bar{x}, \bar{y})\) if and only if \( \Phi \) is (strongly) regular near \((0, 0)\) and \( \text{sur } \Psi(x | y) = \text{sur } \Phi(0 | 0) \).

In other words, the regularity properties of \( \Psi \) are the same as of its “linearization” \( \hat{\Psi} \). Therefore in what follows we can deal only with the linear variational inequality

\[
0 \in Ax + N(C, x)
\]

and the associated mapping

\[
\Phi(x) = Ax + N(C, x).
\]

The key role in our analysis is played by the concept of a face of a polyhedral set \( C \) which is any closed subset \( F \) of \( C \) such that any segment \( \Delta \subset C \) containing a point \( x \in F \) in its interior lies in \( F \). A face of \( C \) proper if it is different from \( C \). We refer to [158] for all necessary information about faces. The following facts are important for our discussion:

- the set \( \mathcal{F}_C \) of all faces of \( C \) is finite;
- \( F \in \mathcal{F}_C \) if and only if there is a \( y \in \mathbb{R}^n \) such that \( F = \{x \in C : \langle y, x \rangle \geq \langle y, u \rangle, \forall u \in C\} \);
- if \( F, F' \in \mathcal{F}_C \) and \( F \cap \text{ri } F' \neq \emptyset \), then \( F' \subset F \); a proper face of \( C \) lies in the relative boundary of \( C \);
- if \( F \in \mathcal{F}_C \) and \( x_1, x_2 \) belong to the relative interior of \( F \), then \( T(C, x_1) = T(C, x_2) \) and \( N(C, x_1) = N(C, x_2) \).

The last property allows to speak about the tangent and normal cones to \( C \) at \( F \) which we shall denote by \( T(C, F) \) and \( N(C, F) \). It is an easy matter to see that

\[
\dim F + \dim N(C, F) = n; \quad \dim(F + N(C, F)) = n.
\]

For any \( x \in C \) denote by \( F_{\min}(x) \) the minimal element of \( \mathcal{F}_C \) containing \( x \). The is straightforward

\[
x \in F \in \mathcal{F}_C, \& F = F_{\min}(x) \iff x \in \text{ri } F.
\]
Proposition 8.13. If $\Phi$ is regular near $(x, y)$ and $F = F_{\min}(x)$, then
\[ \dim(A(F) + N(C, F)) = n. \]

In particular, $A$ is one-to-one on $F$.

Proof. If $\dim F = 0$, then $x$ is an extreme point of $C$ in which case $T(C, x)$ is a convex cone containing no lines and its polar therefore has nonempty interior. On the other hand, if $x \in \text{int } C$, then $N(C, u) = \{0\}$ for all $u$ of a neighborhood of $x$ and $\Phi(u) = Au$ for such $u$. So by regularity $A$ is an isomorphism.

Thus in the sequel we may assume that the dimensions of both $F$ and $N(C, F)$ are positive. By changing $(x, y)$ slightly, we can guarantee that $y$ belong to the relative interior of $N(C, F)$. Let $\varepsilon > 0$ be so small that the distances from $x$ and $y$ to the relative boundaries of $F$ and $N(C, F)$ are greater than $\varepsilon$. Then any $(u, v)$ such that $u \in C$, $v \in N(C, u)$, $\|u - x\| < \varepsilon$, $\|v - y\| < \varepsilon$ must belong to $F \times N(C, F)$. This means that $\Phi(B(x, \varepsilon)) \cap B(y, \varepsilon) \subset A(F) + N(C, F)$ and the result follows from (8.21). Indeed, the dimension equality is immediate from the last inclusion. On the other hand, if $A$ is not one-to-one on $F$, then $\dim A(F) < \dim F$ and by (8.21) $\dim A(F) + \dim N(C, F) < n$. \qed

Let $C \subset \mathbb{R}^n$ be a convex polyhedron, and let $F$ be a proper face of $C$. Let $L$ be the linear subspace spanned by $F$ and $M$ the linear subspace spanned by $N(C, F)$. These subspaces are complementary by (8.21) and orthogonal. By Proposition 8.13 $A(L)$ and $M$ are also complementary subspaces if $\Phi$ is regular near any point of the graph.

Let $\pi_M$ be the projection onto $M$ parallel to $A(L)$, so that $\pi_M(A(F)) = 0$. Set $K_M = (T(C, F)) \cap M$, and let $A_M$ be the restriction of $\pi_M \circ A$ to $M$. Then $K_M$ is a convex polyhedral cone in $M$ and its polar $K_M^o$ (in $M$) coincides with $N(C, F)$.

Definition 8.14. The set-valued mapping $\Phi_M(x) = A_M x + N(K_M, x)$ viewed as a mapping from $M$ into $M$ will be called factorization of $\Phi$ along $F$.

Observe that the graph of a factorization mapping is a union of convex polyhedral cones.

Proposition 8.15. If $\Phi$ is regular near $(x, A\overline{x})$ for some $\overline{x} \in C$, then the factorization of $\Phi$ along $F = F_{\min}(\overline{x})$ is globally regular on $\mathbb{R}^n$.

Proof. Set $K_1 = T(C, F) = T(C, \overline{x})$ and consider the mapping $\Phi_1(x) = A x + N(K_1, x)$. By Proposition 7.24 $\Phi_1(x) = \Phi(x) + x - A\overline{x}$ for $x$ close to zero. Therefore $\Phi_1$ is regular near $(0, 0)$, hence globally regular by Proposition 7.24. Observe that $K_1 = K_M + L$ and $K_1^o = N(K, F)$ and consequently $N(K_1, x) \subset N(K, \overline{x}) = N(K, F)$ for any $x \in K_1$.

As $\Phi_1$ is globally regular, there is a $\rho > 0$ such that $d(x, \Phi_1^{-1}(z)) \leq \rho d(z, \Phi_1(x))$ for all $x, z \in \mathbb{R}^n$. Take now $x, z \in M$. We have (taking into account that $N(K_M, x) = N(K_1, x + \xi)$ for any $\xi \in L$ and $A_M x = A(x + \xi)$ for some $\xi \in L$)

\[
\begin{align*}
d(z, \Phi_M(x)) &= \inf \{\|z - A_M x - y\| : y \in N(K_M, x)\} \\
&\geq \inf \{\|z - A(x + \xi) - y\| : \xi \in L, y \in N(K_1, x + \xi)\} \\
&= \inf_{\xi \in L} d(z, \Phi_1(x + \xi)) = d(z, \Phi_1(x))
\end{align*}
\]
for some $w \in x + L$. On the other hand, there is a $w' \in \mathbb{R}^n$ such that $z \in \Phi_1(w')$ and $\|w - w'\| = d(w, \Phi_1^{-1}(z))$. Let $x'$ be the orthogonal projection of $w'$ to $M$. We have $z = Aw' + y$ for some $y \in N(K_1, w') \subset M$. Therefore $Aw' \in M$ and moreover $A_M x' = Aw'$. The latter is a consequence of the following simple observation:

\[ v = Aw \in M, \quad x \in M, \quad x \perp (w - x) \Rightarrow A_M x = v. \quad (8.23) \]

Indeed, $z = w - x \in L$, hence $Ax = Aw + Az = v + Az$ and, as $v \in M$ and $Az \in A(L)$ we have $\pi_M(Ax) = v + \pi_M(Az) = v$.

It follows, as $N(K_M, x') = N(K_1, w'))$, that $z \in \Phi_M(x')$ and

\[ d(x, \Phi_M^{-1}(z)) \leq \|x - x'\| \leq \|w - w'\| = d(w, \Phi_1^{-1}(z)) \leq \rho d(z, \Phi_1(w)) \leq d(x, \Phi_M(x)), \]

that is $\Phi_M$ is regular on $M$ (with the rate of metric regularity not greater than $\rho$).

The following theorem is the key observation that paves way for proofs of the main result.

**Theorem 8.16.** Let $C = K$ be a convex polyhedral cone. If $\Phi$ is regular near $(0, 0)$ (hence globally regular by Proposition 5.4), then $A(K) \cap K^0 = \{0\}$.

**Proof.** The result is trivial if $n = 1$. Assume that it holds for $n = m - 1$, and let $m = n$. Note that the inclusion $A(K) \subset K^0$ can hold only if $K = \{0\}$. Indeed, if the inclusion is valid, then $\Phi(x) \in A(K) + K^0 = K^0$ for any $x \in K$, so by regularity $K^0$ must coincide with the whole of $\mathbb{R}^m$ and hence $K = \{0\}$. Thus if there is a nonzero $u \in A(K) \cap K^0$, we can harmlessly assume that $u$ is a boundary point of $K^0$ and there is a nonzero $w \in N(K^0, u)$. Then $w \in K$ and $u \in N(K, w)$. Let $F = F_{\text{min}}(w)$ so that $u \in N(K, F)$. Let as before, $L$ be the linear subspace spanned by $F$ and $M$ the linear subspace spanned by $N(K, F)$. These subspaces are complementary by (8.21) and orthogonal. By Proposition 8.13 $A(L)$ and $M$ are also complementary subspaces. Clearly, $u$ does not belong either to $L$ or to $A(L)$, the latter because otherwise the dimension of $A(F) + N(K, F)$ would be strictly smaller than $n$.

Consider the factorization $\Phi_M$ of $\Phi$ along $F$. Then $u \in K_M^0$, by definition. But as follows from (8.23) $u$ also belongs to $A_M(K_M)$. As $\Phi_M$ is regular by Proposition 8.13 and $\dim M < m$, the existence of such a $u$ contradicts to the induction hypothesis.

We are ready to state and proof the main result of the subsection.

**Theorem 8.17** (regularity implies strong regularity). Let $C$ be a polyhedral set and $\Phi(x) = Ax + N(C, x)$. If $\Phi$ is globally regular then the inverse mapping $\Phi^{-1}$ is single-valued and Lipschitz on $\mathbb{R}^m$. Thus, global regularity of $\Phi$ implies global strong regularity.

In other words, the solution map of $y \in \Phi(x)$ is everywhere single-valued and Lipschitz.

**Proof.** We only need to show that $\Phi^{-1}$ is single-valued: the Lipschitz property will then automatically follow from regularity. The theorem is trivially valid if $n = 1$. Suppose it is true for $n \leq m - 1$ and consider the case $n = m$. We have to show that, given a convex polyhedron $C \subset \mathbb{R}^m$ and a linear operator $A$ in $\mathbb{R}^m$ such that $\Phi(x) = Ax + N(C, x)$ is
globally regular on $\mathbb{R}^n$, the equality $Ax + y = Au + z$ for some $x, u, v \in C$, $y \in N(C, x)$, 
$z \in N(C, u)$ can hold only if $x = u$ and $y = z$.

Step 1. To begin with we observe that the equality $Au = Ax + y$ for some $u, x \in C$ and 
y $\in N(C, x)$ may hold only if $u = x$. Indeed, $u - x \in T(C, x)$. The same argument as in the 
proof of Proposition 8.15 shows that $\Phi_1(w) = Aw + N(T(C, x), w)$ is also globally regular 
and therefore by Theorem 8.16 $A(T(C, x)) \cap N(C, x) = \{0\}$. It follows that $A(u - x) = 
y = 0$. But regularity of $\Phi_1$ implies (by Proposition 8.13) that $A$ is one-to one on $T(C, x)$, 
therefore $u = x$.

Step 2. Assume now that for some $x, u \in C$, $u \neq x$, the equality $Ax + y = Au + z$, or 
$A(u-x) = y-z$, holds with $y \in N(C, x)$, $z \in N(C, u)$. We first show that this is impossible 
if $x \in F_{\min}(u)$. If under this condition $x \in riC$, then $u$ is also in $riC$ which means that 
$N(C, x) = N(C, u)$ coincides with the orthogonal complement $E$ to the subspace spanned
by $C-C$. We have $y-z \in E$ and $u-x \in C-C$. By Proposition 8.13 $A(u-x) = y-z = 0$ 
and the second part of the proposition implies that $u = x$.

Let now $F = F_{\min}(x)$ be a proper face of $C$. Then $F \subset F_{\min}(u)$ and therefore $z 
in N(C, F)$. Denote as before by $L$ the subspace spanned by $F$ and by $M$ the subspace 
spanned by $N(C, F)$, and let $\Phi_M$ be the factorization of $\Phi$ along $F$. Set $v = A(u-x) = 
y-z$. Then $v \in M$ as both $y$ and $z$ are in $N(C, F)$. Let $w$ be the orthogonal projection 
of $u-x$ onto $M$. Then by (8.23) $A(w) = v$ and therefore $A_M w = v$.

Thus (recall that $y, z \in M$)

$$A_M w + z = (\pi_M \circ A)(u-x) + z = \pi_M(A(u-x) + z) = \pi_M y = y.$$ 

On the other hand, it is clear that $y \in N(K_M, 0)$ and $z \in N(K_M, w)$. Indeed, $z \in N(T(C, x), u-x)$ (since $\langle z, v-x \rangle \leq \langle z, u-x \rangle$) for all $v \in C$ on the one hand and, as 
we have seen, $z \in N(C, x)$, on the other and therefore $z \in N(K_M, w)$ as $z \in M$ and 
$w-(u-x) \in L$. As $\dim M < m$, we conclude by the induction hypotheses that $w = 0$, 
hence $u-x \in L$. But $A(u-x) = y-z \in M$ and a reference to proposition 8.13 again 
proves that $u = x$.

Step 3. It remains to consider the case when neither $x$ nor $u$ belongs to the minimal face 
of the other. Let $\kappa$ be the modulus of metric regularity of $\Phi$ or any bigger number. Choose 
$\varepsilon > 0$ so small that the ball of radius $(1 + \kappa)\varepsilon$ around $x$ does not meet any face $F \in F_C$ 
not containing $x$. This means that $x \in F_{\min}(w)$ whenever $w \in C$ and $\|w-x\| \leq (1+\kappa)\varepsilon$. 
Let further $N$ be an integer big enough to guarantee that $\delta = N^{-1}\|y\| < \varepsilon$. Regularity of 
$\Phi$ allows to construct recursively a finite sequence of pairs $(u_k, z_k), k = 0, 1, \ldots, m$ such that 

$$(u_0, z_0) = (u, z), \quad z_k \in F_{\max}(u_k), \quad u_k + z_k = x + (1-m^{-1}k)y, \quad \|u_k - u_{k-1}\| \leq \kappa \delta.$$

Then $u_N + z_N = x$. As follows from the result obtained at the first step of the proof, 
this means that $u_N = x$. This in turn implies, as $u_0 \neq x$, that for a certain $k$ we have 
$u_k \neq x$, $\|u_k - x\| \leq \kappa \delta < \kappa \varepsilon$. By the choice of $\varepsilon$ this implies that $x \in F_{\min}(u_k)$. But 
in this case the result obtained at the second step excludes the possibility of the equality 
$u_k + z_k = x + (1-m^{-1}k)y$ unless $u_k = x$. So we again get a contradiction that completes 
the proof.
8.8 Differential inclusions – existence of solutions.

Here we consider the Cauchy problem for differential inclusions:

\[ \dot{x} \in F(t,x), \quad x(0) = x_0, \quad (8.24) \]

where \( F : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n \). We assume that

- \( F \) is defined on some \( \Delta \times U \) (that is \( F(t,x) \neq \emptyset \) for all \( x \in U \) and almost all \( t \in \Delta \)), where \( \Delta = [0,T] \) and \( U \) is an open subset of \( \mathbb{R}^n \) containing \( x_0 \);

- the graph of \( F(t, \cdot) \) is closed for almost every \( t \in \Delta \);

- \( F \) is measurable in \( t \) in the sense that the function \( t \mapsto d((x,y), \text{Graph } F(t, \cdot)) \) is measurable for all pairs \((x,y) \in \mathbb{R}^n \times \mathbb{R}^n \).

By a solution of \( (8.24) \) on \([0,\tau] \subseteq [0,\Delta] \) we mean any absolutely continuous \( x(t) \) defined on \([0,\tau] \) and such that \( \dot{x}(t) \in F(t,x(t)) \) almost everywhere on \([0,\tau] \).

**Theorem 8.18.** Assume that there is a summable \( k(t) \) such that

\[ h(F(t,x),F(t,x')) \leq k(t)\|x-x'\|, \quad \forall x,x' \in U, \text{ a.e. on } [0,1]. \quad (8.25) \]

Let further \( x_0(\cdot) \) be an absolutely continuous function on \([0,T]\) with values in \( U \) such that \( x_0(0) = x_0 \) and \( \rho(t) = d(\dot{x}_0(t), F(t,x_0(t))) \) is a summable function.

Then there is a solution of \( (8.24) \) defined on some \([0,\tau]\), \( \tau > 0 \). Specifically, set \( r = d(x_0, \mathbb{R}^n \setminus U) \), and let \( \tau \in (0,T] \) be so small that

\[ 1 > k_\tau = \int_0^\tau k(t)dt; \quad (1 - k_\tau)r > \xi_\tau = \int_0^\tau d(\dot{x}_0(t), F(t,x_0(t)))dt. \quad (8.26) \]

Then for any \( \varepsilon > 0 \) there is a solution \( x(\cdot) \) of \( (8.24) \) defined on \([0,\tau]\) and satisfying

\[ \int_0^\tau \|\dot{x}(t) - \dot{x}_0(t)\| \leq \frac{1 + \varepsilon}{1 - k_\tau} \xi_\tau. \quad (8.27) \]

Recall that \( h(P,Q) \) is the Hausdorff distance between \( P \) and \( Q \).

**Proof.** We may set \( x_0(t) \equiv 0 \) (replacing if necessary \( F(t,x) \) by \( F(t,x_0(t) + x) - \dot{x}(t) \) and \( U \) by \( rB \)). Let \( X = W_{0}^{1,1}[0,\tau] \) stand for the space of \( \mathbb{R}^n \)-valued absolutely continuous functions on \([0,\tau]\) equal to zero at zero with the norm

\[ \|x(\cdot)\|_\tau = \int_0^\tau \|\dot{x}(t)\|dt, \]

86
and let $I$ denote the identity map in $X$. Let finally $F$ be the set-valued mapping from $X$ into itself that associates with every $x(\cdot)$ the collection of absolutely continuous functions $y(\cdot)$ such that $y(0) = 0$ and $\dot{y}(t) \in F(t, x(t))$ a.e.. We have to prove the existence of an $x(\cdot) \in X$ satisfying (8.27) and

$$0 \in (I - F)(x(\cdot)) \quad (8.28)$$

Note first that the graph of $F$ is closed, that is whenever $x_n(\cdot) \to x(\cdot)$, $y_n(\cdot) \in F(x_n(\cdot))$ and $y_n(\cdot)$ norm converge to $y(\cdot)$, then $y(\cdot) \in F(x(\cdot))$. Let $U$ be the open ball of radius $r$ around zero in $X$. Thus $x(t) \in U$ for any $t \in [0, \tau]$ whenever $x(\cdot) \in U$ and therefore by (8.25) $F$ is Lipschitz on $U$ with $\text{lip}F(U) \leq k_\tau$. On the other hand, $I$ is Milyutin regular on $U$ with $\text{sur}m I(U) = 1$. By Theorem 4.2

$$\text{sur}m(I - F)(U) \geq 1 - k_\tau. \quad (8.29)$$

In particular $B(y(\cdot), (1 - k_\tau)\rho) \subset (I - F)(\rho B)$ for any $y(\cdot) \in (I - F)(0)$ if $\rho < r$. Take a $y(\cdot) \in X$ such that $\dot{y}(t) \in F(t, 0)$ and $\|\dot{y}(t)\| = d(0, F(t, 0))$ a.e.. Then $\|y(\cdot)\|_r = \xi_r < (1 - k_\tau)\rho$ by (8.26). Thus $0 \in B(y(\cdot), (1 - k_\tau)\rho)$ for some $\rho < r$ and therefore there is an $x(\cdot)$ with $\|x(\cdot)\|_r < \rho$, $0 \in (I - F)(x(\cdot))$. \hfill $\Box$

The theorem is close to the original result of Filippov [66]. Versions of this results and its applications can be found in many subsequent publications, see e.g [7, 8]. Typical proofs of existence results for differential inclusions use either some iteration procedures or selection theorems to reduce the problem to existence of solutions of differential equations. Observe that our proof appeals to non-local regularity theory.

References

[1] P. Abraham and J. Robin, *Transversal Mappings and Flows*, Benjamin Press, N.Y. 1967.

[2] F.J. Aragon Artacho and M.H. Geoffroy, Characterization of metric regularity of subdifferential, *J. Convex Analysis*, 15 (2008), 365-380.

[3] F.J. Aragon Artacho and M.H. Geoffroy, Metric subregularity of convex subdifferential in Banach spaces, *J. Nonlinear Convex Analysis*, 15 (2014), 35-47.

[4] A. Arutyunov, E. Avakov, B. Gelnan, A. Dmitruk and V. Obukhovski, Locally covering maps in metric spaces and coincidence points, *J. Fixed Point Theory and Applications* 5 (2009), 106-127

[5] J.-P. Aubin, Contingent derivatives of set-valued maps and existence of solutions to nonlinear inclusions and differential inclusions, in “Advances in Mathematics. Supplementary Studies”, L. Nachbin, ed. 1981, pp 160-232.

[6] J.-P. Aubin, Lipshitz behavior of solutions to convex optimization problems, *Math. Oper. Res.* 9 (1984), 87-111.

[7] J.-P. Aubin and A. Cellina, *Differential Inclusions*, Springer 1984.
[8] J.P. Aubin and H. Frankowska, *Set-Valued Analysis*, Birkhäuser, 1990.

[9] E.R. Avakov, G.G. Magaril-Il’yaev and V.M. Tikhomirov, Lagrange’s principle in extremal problems with constraints, *Russian Math. Surveys* 68:3 (2013), 401-434.

[10] D. Azé, A unified theory for metric regularity of multifunctions, *J. Convex Analysis* 13 (2006), 225-252.

[11] D. Azé and S. Benahmed, On implicit multifunction theorems, *Set-Valued Analysis*, 16 (2008), 129-155.

[12] D. Azé and J.-N. Corvellec, On the sensitivity analysis of Hoffmann’s constant for systems of linear inequalities, *SIAM J. Optimization*, 12 (2002), 913-927.

[13] D. Azé and J.-N. Corvellec, Characterization of error bounds for lower semicontinuous functions on metric spaces, *ESIAM Control. Optimization, Calc. Variations* 10 (2004), 409-425.

[14] D. Azé and J.-N. Corvellec, On some variational properties of metric spaces, *J. Fixed Point Theory Applications* 5 (2009), 185-200.

[15] D. Azé, J.-N. Corvellec and R. Lucchetti, Variational pairs and application to stability in nonsmooth analysis, *Nonlinear Analysis TMA* 49 (2002), 643-670.

[16] D. Azé, C.C. Chou and J.P. Penot, Subtraction theorem and approximate openness of multifunctions, *J. Math. Anal. Appl.*, 221 (1998), 33-58.

[17] S. Banach, *Théorie des opérations linéaires*, Monografje Matematyczne, Warszawa, 1932.

[18] L. Barbet, M. Dambrine, A. Daniilidis and L. Rifford, Sard theorems for Lipschitz functions and applications in Optimization, *Israel J. Mathematics*, to appear.

[19] H. Bauschke and J.M. Borwein, On projection algorithms for solving convex feasibility problems, *SIAM Review*, 38 (1996), 367-426.

[20] M.S. Bazaraa, J.J. Goode and M.Z. Nashed, On the cones of tangents with applications to mathematical programming, *J. Optimization Theory Appl.* 13 (1974), 389-426.

[21] J. Bochnak, M. Coste and M.-F. Roy, *Real Algebraic Geometry*, Springer 1998.

[22] J. Bolte, A. Daniilidis and A.S. Lewis, The Morse-Sard theorem for non-differentiable subanalytic functions *J. Math. Anal. Appl.* 321 (2006), 729-740.

[23] J.M. Borwein. Adjoint process duality, *Math. Oper. Res.*, 8 (1983), 403-434.

[24] J.M. Borwein, Norm duality for convex processes and applications. *Journal of Optimization Theory and Applications*, 48 (1986), 53-64.
[25] J.M. Borwein and A.S. Lewis, *Convex Analysis and Nonlinear Optimization*, 2nd edition, Springer 2006.

[26] J.M. Borwein and D.M. Zhuang, Verifiable necessary and sufficient conditions for openness and regularity of set-valued and single-valued maps, *J. Math. Anal. Appl* 134 (1988), 441-459.

[27] J.M. Borwein and J.Zhu, *Techniques of Variational Analysis*, CMS Books in Mathematics, v. 20, Springer 2006.

[28] L.M. Bregman, The method of successive projections for finding a common point of convex sets, *Soviet Math. Doklady*, 6 (1965), 688-692.

[29] M.J. Cánovas, M.A. Lopez, M.S. Mordukhovich and J. Parra. Variational analysis in linear semi-infinite and infinite programming 1: Stability of linear inequality system of feasible solutions, *SIAM J. Optimization* 20 (2009), 1504-1526.

[30] M.J. Cánovas, F.J. Gómez-Senent and J. Parra, Linear regularity, equi-regularity and intersection mappings for convex semi-infinite inequality systems, *Math. Programming*, Ser. B 123 (2010), 33-60.

[31] R. Cibulka and M. Fabian, A note on Robinson-Ursescu and Lyusternik-Graves theorems, *Math. Programming*, Ser. B, 139 (2013), 89-101.

[32] R. Cibulka, M. Fabian and A.D. Ioffe, On primal regularity estimates for single-valued mappings, to appear in *J. Fixed Point Theory Appl.*, 17 (2015), 187208

[33] F.H. Clarke, Necessary Conditions for Nonsmooth Problems in Optimal Control and the Calculus of Variations, Ph.D. dissertation, Uniw. of Washington, Seattle, 1973.

[34] F.H. Clarke, Generalized gradients and applications, *Trans. Amer. Math. Soc.* 205 (1975), 247-262.

[35] F.H. Clarke, A new approach to Lagrange multipliers, *Math. Oper. Research*, 1 (1976), 165-174.

[36] F.H. Clarke, On the inverse function theorem, *Pacific J. Math.*, 64 (1976), 97-102.

[37] F.H. Clarke, Necessary conditions for a general control problem, in *Calculus of Variations and Control Theory*, D.L. Russel, editor, Acad. Press 1976, pp. 257-279.

[38] F.H. Clarke, *Optimization and Nonsmooth Analysis*, Wiley-Interscience, 1983.

[39] F.H. Clarke, *Necessary Conditions in Dynamic Optimization*, Memoirs AMS, v. 816, 2005.

[40] F.H. Clarke, Yu.S. Ledyaev, R.J. Stern and P.R. Wolenski, *Nonsmooth Analysis and Control Theory*, Springer 1998.

[41] R. Cominetti, Metric regularity, tangent sets and second order optimality conditions, *Appl. Math. Optimization*, 21 (1990), 265-287.
[42] M. Coste, *An Introduction to o-Minimal Geometry*, Inst. Rech. Math., Univ. de Rennes, 1999 (http://name.math.univ-rennes1.fr/michel.coste/polyens/OMIN.pdf).

[43] A. Coulibali and J.-P. Crouzeix, Conditions numbers and error bounds in convex programming, *Math. Programming*, ser. B **116** (2009), 79-113.

[44] E. De Giorgi, A. Marino and M. Tosques, Problemi di evoluzione in spazi metrici e curve di massima pendenza, *Atti Acad. Nat. Lincei, Rend. Cl. Sci. Fiz. Mat. Natur. 68* (1980), 180-187.

[45] A.V. Dmitruk, A.A. Milyutin and N.P. Osmolovskii, Ljusternik’s theorem and the theory of extrema, *Russian Math. Surveys 35:6* (1980), 11-51.

[46] S. Dolecki, Semicontinuity in constraint optimization 1, *Control & Cybernetics 7* no. 2 (1978), 5-15.

[47] S. Dolecki, Semicontinuity in constraint optimization 2, *Control & Cybernetics 7* no. 3 (1978), 17-26.

[48] A.L. Dontchev, Characterizations of Lipschitz stability in optimization, in *Recent Developments in Well-Posed Variational Problems*, R. Lucchetti and J. Revalski eds., Kluwer 1995, pp. 95-115.

[49] A.L. Dontchev, The Graves theorem revisited, *J. Convex Analysis, 3* (1996), 45-54.

[50] A.L. Dontchev and H. Frankowska, Lyusternik-Graves theorem and fixed points, *Proc. Amer. Math. Soc., 139* (2011), 521-534.

[51] A.L. Dontchev and H. Frankowska, Lyusternik-Graves theorem and fixed points 2, *J. Convex Analysis, 19* (2012), 955-974.

[52] A.L. Dontchev, A.S. Lewis and R.T. Rockafellar, The radius of metric regularity, *Trans. Amer. Math. Soc. 355* (2003), 493-517.

[53] A.L. Dontchev, M. Quincampoix and N. Zlateva, Aubin criterion for metric regularity, *J. Convex Analysis, 13* (2006), 281-297.

[54] A. L. Dontchev and R. T. Rockafellar, Characterizations of strong regularity for variational inequalities over polyhedral convex sets, *SIAM Journal on Optimization, 6* (1996), 1087-1105.

[55] A.L. Dontchev and R.T. Rockafellar, *Implicit Functions and Solution Mappings* (second edition), Springer 2014.

[56] D. Drusvyatsky and A.D. Ioffe, Quadratic growth and critical point stability of semi-algebraic functions, *Math. Programming, Ser. A* DOI 10.1007/s10107-014-0820-y.

[57] D. Drusvyatsky, A.D. Ioffe and A.S. Lewis, Alternating projections and coupling slope, arXiv:1401.7569v1 [math.OC] 29 Jan 2014.
[58] D. Drusvyatsky and A.S. Lewis, Semi-algebraic functions have small subdifferentials, *Mathematical Programming* Ser. B, 140 (2013), 5-29.

[59] D. Drusvyatsky and A.S. Lewis, Tilt stability, uniform quadratic growth and strong metric regularity of the subdifferential, *SIAM J. Optimization*, 23 (2013), 256-267.

[60] A.Ya. Dubovitslii and A.A. Milyutin, Zadachi na ekstremum pri nalichii ogranichenij, Zh. Vychisl. Matematiki i Mat. Fiziki, 5:3 (1965), 395-453 (in Russian; English translation: Problems for extremum under constraints, USSR Comput. Math. Math. Physics, 5 (1965)).

[61] M. Durea and R. Strugariu, Openness, stability and implicit function theorems, *Nonlinear Analysis TMA*, 75 (2012), 1246-1259.

[62] B.C. Eaves and U.G. Rothblum, Relationships of properties of piecewise affine maps over ordered fields, *Linear Algebra Appl.*, 132 (1990), 1-63.

[63] M. Fabian and D. Preiss, A generalization of the interior mapping theorem of Clarke and Pourciau, Comment. Math. Univ. Caroliniae, 28 (1987), 311–324.

[64] M. Fabian and N.V. Zhivkov, A characterization of Asplund spaces with the help of ε-supports of Ekeland and Lebourg, *C.R. Acad. Bulgare Sci.*, 38 (1985), 671-674.

[65] F. Facchinei and J.S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems*, Springer N.Y. 2003.

[66] A.F. Filippov, Classical solutions of differential inclusions with multivalued right-hand sides, *SIAM J. Control Optimization*, 5 (1967), 609-621.

[67] H. Frankowska, Some inverse mapping theorems, *Ann. Inst. Henri Poincaré. Analyse Non Linéaire*, 7 (1990), 183-234.

[68] P. Fusek, On metric regularity of weakly almost piecewise smooth functions and some applications in nonlinear semi-definite programming, *SIAM J. Optimization*, 23 (2013), 1041-1061.

[69] R.V. Gamkrelidze, *Foundations of Optimal Control*, TGU, Tbilisi 1977 (in Russian).

[70] B.D. Gelman, A generalized theorem about implicit mappings, *Funktsional’nyj Analiz Prilozhen.*, 35 (2001), 28-35.

[71] H. Gfrerer, First and second order characterizations of metric subregularity and calmness of constraint set mappings, *SIAM J. Optimization*, 21 (2011), 1439-1474.

[72] B.Ginsburg and A.D. Ioffe, Maximum principle for general semilinear optimal control problems, in *Nonsmooth Analysis and Geometric Methods in Optimal Control*, B. Mordukhovich and H. Sussmann, eds., Springer 1995, pp. 81-110.

[73] L.M. Graves, Some mapping theorems, *Duke Math. J.* 17 (1950), 111-114.
[74] A. Grothendieck, Sketch of a proposal, in Geometric Galois Actions, L. Schneps and P. Lochak, eds. Cambridge Univ. Press 1997.

[75] L.G. Gubin, B.T. Polyak and E.V. Raik, The method of projections for finding a common point of convex sets, USSR Computational Mathematics and Math. Physics, 7 (1967), 1-24.

[76] V. Guillemin and A. Pollack, Differential Topology, Prentice Hall, 1976.

[77] R. Hesse and D.R. Luke, Nonconvex notion of regularity and convergence of fundamental algorithms for feasibility problems, SIAM J. Optimization, 23 (2013), 2397-2419.

[78] A.J. Hoffman, On approximate solutions of systems of linear inequalities, J. Research Nat. Bureau Standards e49 (1952), 263-265.

[79] L. Hörmander, Sur la fonction d’appui des ensembles convexes dans un espace localement convex, Arkiv för Matematik, 3 (1955), 181-186.

[80] A.D. Ioffe, Necessary and sufficient conditions for a local minimum. Reduction theorem and first order conditions, SIAM J. Control Optimization, 17 (1979), part 1 245-251, part 2 261-265.

[81] A.D. Ioffe, Sous-differentielles approchées des fonctions numériques, C.R. Acad. Sci. Paris, 292 (1981), 675-678.

[82] A.D. Ioffe, Non-smooth analysis: differential calculus of non-differentiable mappings, Trans. Amer. Math. Soc. 255 (1981), 1-55.

[83] A.D. Ioffe, Approximate subdifferentials and applications 1. The finite dimensional theory, Trans. Amer. Math. Soc. 28 (1984), 389-416

[84] A.D. Ioffe, On the local surjection property, Nonlinear Analysis. TMA, 11 (1987), 565-592.

[85] A.D. Ioffe, Approximate subdifferentials and applications 3. The metric theory, Mathematika 36 (1989), 1-38.

[86] A.D. Ioffe, Directional compactness, scalarization and nonsmooth semi-Fredholm mappings, Nonlinear Analysis TMA, 29 (1997), 201-219.

[87] A.D. Ioffe, Variational methods in local and global nonsmooth analysis, in Nonlinear Analysis, Differential Equations and Control, F.H. Clarke and R.J. Stern, eds, NATO Science Series C: Mathematical and Physical Sciences, vol. 258, (1999), pp. 447-502.

[88] A.D. Ioffe, Metric regularity and subdifferential calculus, Uspehi Mat. Nauk 55:3 (2000), 103-162 (in Russian), English translation: Russian Math. Surveys, 55:3 (2000), 501-558.

[89] A.D. Ioffe, On regularity estimates for mappings between embedded manifolds, Control and Cybernetics, 36 (2007), 659-668.
[90] A.D. Ioffe, Critical values of set-valued mappings with stratifiable graphs. Extensions of Sard and Smale-Sard theorems, *Proc. Amer. Math. Soc.* **136** (2008), 3111-3119.

[91] A.D. Ioffe, The principle of Lagrange and critical values of optimization problems, *Pacific J. Optimization* **6** (2010), 341-352.

[92] A.D. Ioffe, Regularity on fixed sets, *SIAM J. Optimization* **21** (2011), 1345-1370.

[93] A.D. Ioffe, Variational analysis and mathematical economics 2. Nonsmooth regular economies, *Adv. Math. Economics* **14** (2011), 17-38.

[94] A.D. Ioffe, On the general theory of subdifferentials, *Adv. Nonlinear Analysis* **1** (2012), 47-120.

[95] A.D. Ioffe, Convexity and variational analysis, in *Computational and Analytical Mathematics*, D. Bailey, H. Bauschke, F. Garvan, M. Thera, J. Vanderwerff and H. Wolkovicz, eds., Springer 2013, pp. 397-428.

[96] A.D. Ioffe. Separable reduction of metric regularity properties, in *Constructive Nonsmooth Analysis and Related Topics*, V.F. Demyanov. P.M. Pardalos and M. Batsin, Eds., Springer Optimization and Applications v. 87, Springer 2013.

[97] A.D. Ioffe, Metric regularity, fixed points and some associated problems of variational analysis, *J. Fixed Point Theory Appl.*, **15** (2014), 67-99.

[98] A.D. Ioffe, On necessary conditions for and extremum, *Foundations of Pure and Applied Mathematics*, Moscow State Univ., to appear.

[99] A.D. Ioffe, Complementarity and regularity with application to variational inequality over a polyhedral set, arXiv:1508.06607v1 [math.OC] 26 Aug 2015.

[100] A.D. Ioffe and J.V. Outrata On metric and calmness qualification conditions in subdifferential calculus, *Set-Valued Analysis*, **16** (2008), 199-227.

[101] A.D. Ioffe and J.-P. Penot, Subdifferentials of performance functions and calculus of coderivatives of set-valued mappings, *Serdica Math. J.*, **22** (1996), 359-384.

[102] A.D. Ioffe and Y. Sekiguchi, Regularity estimates for convex multifunctions, *Math. Programming* ser. B **117** (2009), 255-270.

[103] A.D. Ioffe and V.M. Tihomirov, *Theory of Extremal Problems*, Nauka, Moscow 1974 (in Russian); English translation: North Holland 1979.

[104] A. Jourani and L. Thibault, Approximate subdifferential and metric regularity: the finite dimensional case, *Math. Programming*, **47** (1990), 203-218.

[105] A. Jourani and L. Thibault, Metric regularity and subdifferential calculus in Banach spaces, *Set-Valued Analysis* **3** (1995), 87-100.

[106] A. Jourani and L. Thibault, Verifiable conditions for openness and metric regularity in Banach spaces, *Trans. Amer. Math. Soc.*, **347** (1995), 1255-1268.
[107] A. Jourani and L. Thibault, Metric regularity and strongly compactly Lipschitzian mappings, *Nonlinear Analysis. TMA*, **24** (1995), 229-240.

[108] A. Jourani and L. Thibault, Coderivatives of multivalued mappings, locally compact cones and metric regularity, *Nonlinear Analysis TMA*, **35** (1999), 925-945.

[109] D. Klatte and B. Kummer, *Nonsmooth Equations in Optimization*, Kluwer 2002.

[110] D. Klatte and B. Kummer, Optimization methods and stability of inclusions in Banach spaces, *Math. Programming, Ser B*, **117** (2009), 305-330.

[111] D. Klatte and Wu Li, Asymptotic constraint qualifications and global error bounds for convex inequalities, *Math. Programming*, **84** (1999), 137160.

[112] A.Y.Kruger, Generalized differentials of nonsmooth functions, mimeographed notes, Belorussian State Univ., 1332-81, 1981, 64 pp. (in Russian).

[113] A.Y. Kruger, Generalized differentials of nonsmooth functions and necessary conditions for an extremum, *Siberian Math. J.* **26** (1985), 370-379.

[114] A.Y. Kruger, A covering theorem for set-valued mappings, *Optimization* **19** (1988), 763-780.

[115] A.Y. Kruger, About regularity of collections of sets, *Set-Valued Analysis*, **14** (2006), 187-206.

[116] A.Y. Kruger and B.S. Mordukhovich, Extremal points and the Euler equation in nonsmooth optimization, *Dokl. Acad. Nauk BSSR*, **24** (1980), 684-687 (in Russian).

[117] A. Y. Kruger, H. Van Ngai, M. Théra, Stability of error bounds for convex constraint systems in Banach spaces, *SIAM Journal on Optimization*, **20** (2010), 3280-3296

[118] Y.S. Ledyaev and Q.J. Zhu, Implicit multifunction theorem, *Set-Valued Analysis* **7** (1999), 147-162.

[119] E.S. Levitin, A.A. Milyutin and N.P. Osmolovskii, On conditions for a local minimum in problems with constraints, in *Mathematical Economics and Functional Analysis*, B.S. Mityagin, ed., “Nauka”, Moscow 1974 (in Russian).

[120] A.S. Lewis, Ill-conditioned convex processes and conic linear systems, *Math. Operation Research* **24** (1999), 829-834.

[121] A.S. Lewis, Ill-conditioned inclusions, *Set-Valued Analysis*, **9** (2001), 375-381.

[122] A.S. Lewis, D.R. Luke, and J. Malick, Local linear convergence for alternating and averaged nonconvex projections, *Foundations of Computational Mathematics* **9** (2009), 485–513.

[123] A.S. Lewis and J.S. Pang, Error bounds for convex inequality systems, in *Generalized Convexity, Generalized Monotonicity: Recent Results*, J.P. Crouzeix, J.E. Martinez-Legas and M. Volle, eds., Kluwer 1998, pp. 75-110.
[124] Y.S. Ledyaev and Q.J. Zhu, Implicit multifunction theorem, *Set-Valued Analysis* 7 (1999), 147-162.

[125] P.D. Loewen, Limits of Fréchet normals in nonsmooth analysis, in *Optimization and Nonlinear Analysis*, A. Ioffe, M. Markus and S. Reich (editors), Longman, 1992.

[126] L.A. Lyusternik, On conditional extremums of functionals, *Matem. Sbornik* (3) 41 1934, 390-401 (in Russian).

[127] K.W. Meng and X.Q. Yang, Equivalent conditions for local error bounds, *Set-Valued Var. Analysis* 20 (2012), 617636.

[128] B.S. Mordukhovich, *Approximation Methods in Problems of Optimization and Control*, Nauka, Moscow 1988 (in Russian).

[129] B.S. Mordukhovich, Complete characterization of openness, metric regularity and Lipschitzian properties of multifunctions, *Trans. Amer. math. Soc.* 340 (1993), 1-35.

[130] B.S. Mordukhovich, *Variational Analysis and Generalized Differentiation*, vol 1,2, Springer, 2006.

[131] B.S. Mordukhovich, J.F. Peña and V. Roschina, Applying metric regularity to compute a condition measure of a smoothing algorithm for matrix games, *SIAM J. Optimization*, 20 (2010), 3490-3511.

[132] B.S.Mordukhovich and Y. Shao, On nonconvex subdifferential calculus in Banach spaces, *J. Convex Analysis* 2 (1995), 211-227.

[133] B.S.Mordukhovich and Y. Shao, Nonsmooth sequential analysis in Asplund spaces, *Trans. Amer. Math. Soc.* 348 (1996), 1235-1280.

[134] K.F. Ng and W.H. Yang, Error bounds for some convex functions and distance composite functions, *SIAM J. Optimization* 15 (2005), 1042-1056.

[135] H.V. Ngai and M. Théra, Error bounds and implicit multifunctions in smooth Banach spaces and applications to optimization, *Set-Valued Analysis*, 12 (2004), 195-223.

[136] H.V. Ngai, N.H. Tron and M. Théra, Implicit multifunction theorem in complete metric spaces, *Math. Programming*, Ser. B 139 (2013), 301–326.

[137] L. Nirenberg, *Topics in Nonlinear Functional Analysis*, Lecture notes, Courant Inst. Math. Sci., N.Y. Univ. 1974.

[138] Z. Páles, Inverse and implicit function theorems for nonsmooth maps in Banach spaces, *J. Math. Anal. Appl.* 209 (1997), 202–220.

[139] Z. Páles and V. Zeidan, Infinite dimensional Clarke generalized Jacobian, *Journal of Convex Analysis*, 14 (2007), 433–454
[140] Z. Páles and V. Zeidan, Generalized Jacobian for functions with infinite dimensional range and domain, *Set-Valued Anal* **15** (2007), 331–375.

[141] J.-P. Penot, Sous-différentiels de fonctions numériques non convexes, *C.R. Acad. Sci. Paris* **278** (1974), 1553-1555.

[142] J.-P. Penot, Metric regularity, openness and Lipschitzian behaviour of multifunctions, *Nonlinear Anal.* **13** (1989), 629-643.

[143] J.-P. Penot, Error bounds, calmness and their applications to nonsmooth analysis, *Contemporary Math.* **514** (2010), 225-247.

[144] J.-P. Penot, *Calculus Without Derivatives*, Graduate Texts in Mathematics, v. **266**, Springer, 2012.

[145] R. Poliquin and R.T. Rockafellar, Tilt stability of a local minimum, *SIAM J. Optimization*, **8** (1998), 287-299.

[146] V. Ptáček, A quantitative refinement of a closed graph theorem. *Czechoslovak Math. J.* **24** (1974), 503-506.

[147] D. Ralph, A new proof of Robinson’s homeomorphism theorem for PL-normal maps, *Linear Algebra and Its Applications*, **178** (1993), 249-260.

[148] S.M. Robinson, Normed convex processes, *Trans. Amer. Math. Soc.*, **174** (1972), 127-140.

[149] S.M. Robinson, Stability theory for system of inequalities. Part I: Linear systems, *SIAM J. Numer. Anal.* **12** (1975), 754-769.

[150] S.M. Robinson, Stability theory for system of inequalities. Part II: Differentiable nonlinear systems, *SIAM J. Num. Anal.* **13** (1976), 497-513.

[151] S.M. Robinson, Regularity and stability for convex multivalued functions, *Math. Oper. Res.*, **1** (1976), 130-143.

[152] S.M. Robinson, First order conditions for general nonlinear optimization, *SIAM J. Appl. Math*, **30** (1976), 597-608.

[153] S.M. Robinson, Generalized equations and their solutions. Part I: Basic theory, *Math. Programming Study* **10** (1979), 128-141.

[154] S.M. Robinson, Strongly regular generalized equations, *Math. Oper. Res.* **5** (1980), 43-62.

[155] S.M. Robinson, An implicit function theorem for a class of nonsmooth functions, *Math. Oper. Res.*, **16** (1991), 292-309.

[156] S.M. Robinson, Normal maps induced by linear transformations, *Math. Oper. Res.*, **17** (1992), 691–714.
[157] R.T. Rockafellar. Monotone processes of convex and concave type, Memoires Amer. Math. Soc., 77 (1967).

[158] R.T. Rockafellar, Convex Analysis, Princeton Univ. Press 1970.

[159] R.T. Rockafellar and R.J.B. Wets, Variational Analysis, Springer 1998.

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

[161] J. Schauder, Uber die Umkehrung linearer, stetiger Funktionaloperationen, Studia Math. 2 (1930), 1-6.

[162] W. Schirotzek, Non-Smooth Analysis, Universitext, Springer, 2007.

[163] W. Song, Calmness and error bounds for convex constraint systems, SIAM J. Optimization 17 (2006), 353-371.

[164] L. Thibault, Various forms of metric regularity, unpublished note, Univ. de Montpellier 1999.

[165] V.M. Tikhomirov, Fundamental Principles of the Theory of Extremal Problems, John Wiley & Sons, 1982.

[166] K.Sh. Tsiskaridze, Extremal problems in Banach spaces, in “Nekotorye Voprosy Matematicheskoy Teorii Optimalnogo Upravleniya” (Some Problems of the Mathematical Theory of Optimal control), Inst. Appl. Math., Tbilisi State Univ. 1975 (in Russian).

[167] C. Ursescu, Multifunctions with closed convex graphs, Czechoslovak Math. J. 25 (1975), 438-411.

[168] C. Ursescu, Inherited openness, Revue Roumaine des Mathématiques Pures et Appliquées 41 (1996), 5–6, 401–416.

[169] C. Ursescu, Linear openness of multifunctions in metric spaces, International Journal of Mathematics and Mathematical Sciences, (2005), no.2 203–214.

[170] R.B. Vinter, Optimal Control, Birkhauser, 2000.

[171] R.B. Vinter, The role of metric regularity in state constrained optimal control, Proceedings of the 44th IEEE Conference on Decision and Control, and the European Control Conference Seville, Spain 2005, pp. 262-265.

[172] J. von Neumann, Functional operators 2. The geometry of orthogonal spaces, Annals of Math. Studies, 22, Princeton 1950.

[173] J. Warga, Controllability and necessary conditions in unilateral problems without differentiability assumptions, SIAM J. Control Optimization, 14 (1976), 546-573.

97
[174] J. Warga, Derivate containers, inverse functions and controllability, in *Calculus of Variations and Optimal Control*, D.L. Russel, editor, Acad. Press 1976, pp. 13-46.

[175] Y. Yomdin and G. Comte, *Tame Geometry with Applications to Smooth Analysis*, Springer LNM series, vol. 1834, 2004.

[176] J.A. Yorke, The maximum principle and controllability of nonlinear equations, *SIAM J. Control Optimization*, 10 (1972), 334-338.

[177] C. Zalinescu, *Convex Analysis in General Vector Spaces*, World Scientific 2002.

[178] X.Y. Zheng and K.F. Ng, Metric regularity and constraint qualifications for convex inequalities in Banach spaces, *SIAM J. Optimization* 14 (2004), 757-772.