J. Aust. Math. Soc. **101** (2016), 188–243
doi:10.1017/S1446788715000701

METRIC REGULARITY—A SURVEY PART 1. THEORY

A. D. IOFFE

(Received 6 April 2015; accepted 11 August 2015; first published online 14 July 2016)

Communicated by J. M. Borwein

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 at the very heart of variational analysis, a relatively new discipline whose appearance was, to a large extent, determined by the needs of modern optimization theory in which such phenomena as nondifferentiability 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. This 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.

2010 *Mathematics subject classification*: primary 50C06; secondary 49N99, 90C26.

Keywords and phrases: metric regularity, inverse function theorems, error bounds, variational analysis.

Contents

The author was supported in part by the US–Israel Science Foundation under the grant 2008261.

Introduction

Metric regularity has emerged during the last two or three 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: that is, functions and maps that are nondifferentiable at points of interest,

https://doi.org/10.1017/S1446788715000701 Published online by Cambridge University Press

set-valued mappings and so on. Such phenomena naturally appear in optimization theory, but not only there[1].

In the traditional nonlinear analysis, regularity of a mapping (for example, from a normed space or one 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 the inclusion

$$y \in F(x)$$

(with F a set-valued mapping this time) which, in essence, is the main object of study in variational analysis. The challenge here is evident: there is no clear way to approximate the mapping by simple objects such as 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). We shall 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 up a number of new applications, such as, say, metric fixed point theory, differential inclusions, all chapters of optimization theory and 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

[1]Grothendick mentions 'ubiquity of stratified structures in practically all domains of geometry' in his 1984 *Esquisse d'un Programme* (see [46]).

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 and so on, on the other; and
(c) the 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 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, in particular, (d) the proof is simpler (shorter, or looking more transparent) than that available in the literature known to me.

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 [35] and Klatte and Kummer [65] 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 and 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 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 like a sort of truism but the point is that publications in which over-attachment to certain particular techniques of variational analysis (for example, 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;
$\mathrm{ex}(Q, P) = \sup\{d(x, P) : x \in Q)\}$ — excess of Q over P;
$h(Q, P) = \max\{\mathrm{ex}(Q, P), \mathrm{ex}(P, Q)\}$ — Hausdorff distance between Q and P;
$B(x, r)$ — closed ball of radius r and center at x;
$\overset{\circ}{B}(x, r)$ — open ball of radius r and center at x;
$F|_Q$ — the restriction of a mapping F to the set Q;
$F : X \rightrightarrows Y$ — set-valued mapping;
$\mathrm{Graph}\, F = \{(x, y) :\ y \in F(x)\}$ — graph of F;
I — the identity mapping (subscript, if present, indicates the space, for example, I_X);
$\mathrm{epi}\, f = \{(x, \alpha) : \alpha \geq f(x)\}$ — epigraph of f;
$\mathrm{dom}\, f = \{x : f(x) < \infty\}$ — 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, for example, B_X is the unit ball in X);
S_X — the unit sphere in X;
$\mathrm{Ker}\, A$ — kernel of the (linear) operator A;
$L^\perp = \{x^* \in X^* : \langle x^*, x\rangle = 0,\ \forall\, x \in L\}$ — annihilator of a subspace $L \subset X$;
$K^\circ = \{x^* \in X^* : \langle x^*, x\rangle \leq 0,\ \forall\, x \in K\}$ — the polar of a cone $K \subset X$;
$\mathrm{Im}\, A$ — image of the operator A;
$\mathcal{S}(X)$ — collection of closed separable subspaces of X;
$\mathcal{L}(X, Y)$ — the space of linear bounded operators $X \to Y$ with the *operator norm*:

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

$L \oplus M$ — direct sum of subspaces;
T_xM, N_xM — 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 (for example, N_F is a Fréchet normal cone and so on)
We use the standard conventions $d(x, \emptyset) = \infty$; $\inf \emptyset = \infty$; $\sup \emptyset = -\infty$ with one exception: when we deal with nonnegative 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.

THEOREM 1.1 [11, 93]. *Let $A : X \to Y$ be a linear bounded operator onto Y: that is $A(X) = Y$. Then $0 \in \operatorname{int} A(B)$.*

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 $\|x\| \le K\|y\|$ (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).

DEFINITION 1.2 (Banach constant). Let $A : X \to Y$ be a bounded linear operator. The quantity

$$C(A) = \sup\{r \ge 0 : rB_Y \subset A(B_X)\} = \inf\{\|y\| : y \notin A(B_X)\}$$

will be called the *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

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

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

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 $\bar{x} \in X$. It is said that F is *regular* at $\bar{x}$ if its derivative $F'(\bar{x})$ 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 [73]). *Suppose that F is continuously differentiable and regular at $\bar{x}$. Then the tangent space to the level set $M = \{x : F(x) = F(\bar{x})\}$ at $\bar{x}$ coincides with* $\operatorname{Ker} F'(\bar{x})$.

https://doi.org/10.1017/S1446788715000701 Published online by Cambridge University Press

THEOREM 1.5 (Graves [45]). *Let F be a continuous mapping from a neighborhood of $\overline{x} \in X$ 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 ε-ball around $\overline{x}$. Then

$$B(F(\overline{x}), \gamma t) \subset F(B(\overline{x}, t))$$

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 $F(\overline{x}) = 0$. Take $K > 0$ such that $KC(A) > 1 > K(\delta + \gamma)$, and let $\|y\| < \gamma t$ for some $t < \varepsilon$. Set $x_0 = \overline{x}$, $y_0 = y$ and define, recursively, x_n, y_n as

$$Ax_n = y_{n-1} + Ax_{n-1}; \quad \|x_n - x_{n-1}\| \le K\|y_{n-1}\|; \quad y_n = y_{n-1} - (F(x_n) - F(x_{n-1})).$$

It is an easy matter to verify that

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

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

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

as claimed. □

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 proofs were 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 $\|x - \overline{x}\| < \varepsilon$, $0 \le t < \varepsilon$.

It should be also emphasized that no differentiability assumption is made in the theorem of Graves. In this respect, Graves was way ahead of his time. Observe that the mapping F in the theorem of Graves can be viewed as a perturbation of A by a δ-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 Section 4), 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 $\bar{x}$ and the derivative $F'(\bar{x})$ is an invertible operator onto Y. Then there is a mapping G into X defined in a neighborhood of $\bar{y} = F(\bar{x})$, strictly differentiable at $\bar{y}$ and such that*

$$G'(\bar{y}) = (F'(\bar{x}))^{-1} \quad \text{and} \quad F \circ G = I_Y$$

in the neighborhood.

The shortest among standard proofs of the theorem is based on the contraction mapping principle (see, for example, the second proof of the Theorem in [35]). An equally short proof follows from the theorem of Lyusternik–Graves.

PROOF. Set $A = F'(\bar{x})$. Then $F(x') - F(x) - A(x' - x) = r(x', x)\|x' - x\|$, where $\|r(x', x)\| \to 0$ when $x, x' \to \bar{x}$. 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 $\bar{x}$. This means that F is one-to-one in a neighborhood of $\bar{x}$. But, by the Lyusternik–Graves theorem, $F(U)$ covers a certain open neighborhood of $\bar{y}$. Hence $G = F^{-1}$ is defined in a neighborhood of $F(\bar{x})$. Given y and y' close to $\bar{y} = F(\bar{x})$, let x', x be such that $F(x') = y'$, $F(x) = y$. Then, as we have seen, $\|y - y'\| \geq K\|x - x'\|$.

$$A^{-1}(F(x') - F(x) - A(x' - x)) = A^{-1}(y' - y) - G(y') - G(y),$$

so that

$$\begin{aligned}\|G(y') - G(y) - A^{-1}(y' - y)\| &\leq \|A\|^{-1}\|F(x') - F(x) - A(x' - x)\| \\ &= \|A^{-1}\| \, \|r(x', x)\| \, \|x' - x\| \, \| \leq q(y, y')\|y' - y\|,\end{aligned}$$

where $q(y, y') = Kr(G(y), G(y'))$ obviously goes to zero when $y, y' \to \bar{y}$. □

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 $(\bar{x}, \bar{y}) \in X \times Y$ and strictly differentiable at $(\bar{x}, \bar{y})$. Suppose, further, that the partial derivative $F_y(\bar{x}, \bar{y})$ is an invertible operator. Then there are neighborhoods $U \subset X$ of $\bar{x}$ and $W \subset Z$ of $\bar{z} = F(\bar{x}, \bar{y})$ and a mapping $S : U \times W \to Y$ 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

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

The simplest proof of the theorem is obtained by application of the inverse mapping theorem to the map $X \times Y \to X \times Z$ (see, for example, [35]):

$$\Phi(x, y) = \begin{pmatrix} x \\ F(x, y) \end{pmatrix}.$$

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 $\operatorname{Im} F$. Thus y is a regular value if either $y \neq F(x)$ for any x of the domain of F or $\operatorname{Im} F'(x) = Y$ for every x such that $F(x) = y$.

Theorem 1.10 (Sard [92]). *Let Ω be an open set in $\mathbb{R}^n$ and F a C^k-mapping from Ω 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^∞ functions can be found in [82].

Definition 1.11 (Transversality). Let $F : X \to Y$ be a C^1-mapping, and let $M \subset Y$ be a C^1-submanifold. Finally let 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 $\operatorname{Im} F'(x) + T_yM = Y$. 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 $T_xM_1 + T_xM_2 = X$. For our future discussions, it is useful to have in mind that the latter property can be equivalently expressed in dual terms: $N_xM_1 \cap N_xM_2 = \{0\}$, where $N_xM \subset X^*$ is the *normal space* to M at x, that is, the annihilator of T_xM.

A connection with regularity is immediate from the definition: if (L, φ) 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 $\Phi : X \times L \to Y$ 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 $F : X \times P \to Y$. We can view F as a family of mappings from X into Y parameterized by elements of P. Let us denote 'individual' mappings $x \to F(x, p)$ by $F(\cdot, p)$. Further, let $M \subset Y$ be a submanifold and let $\pi : X \times P \to P$ be the standard Cartesian projection $(x, p) \to p$.

Proposition 1.12. *Suppose F is transversal to M and $Q = F^{-1}(M)$ is a manifold. Finally let $\pi|_Q$ stand for the restriction of π 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) the transversality theorem of Thom.

Theorem 1.13 (See, for example, [47]). *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 $F \pitchfork M$ and $k > \dim X - \operatorname{codim} M$. Then $F(\cdot, p) \pitchfork M$ for each $p \in P$ outside a subset of P with* dim *P-Lebesgue measure zero.*

https://doi.org/10.1017/S1446788715000701 Published online by Cambridge University Press

2. Metric theory. Definitions and equivalences

Here X and Y are metric space. We use the same notation for the metrics in both and hope that this will 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 $F : X \rightrightarrows Y$ be given as well as a $(\bar{x}, \bar{y}) \in \operatorname{Graph} F$.

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

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

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

- *metrically regular* near $(\bar{x}, \bar{y}) \in \operatorname{Graph} F$ 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 $\operatorname{reg} F(\bar{x} \,|\, \bar{y})$ of such K is the *modulus* or *rate of metric regularity* of F near $(\bar{x}, \bar{y})$. If no such K, ε exist, we set $\operatorname{reg} F(\bar{x} \,|\, \bar{y}) = \infty$;

- *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 $\operatorname{lip} F(\bar{x} \,|\, \bar{y})$ is the *Lipschitz modulus* or *rate* of F near $(\bar{x}, \bar{y})$. If no such K, ε exist, we set $\operatorname{lip} F(\bar{x} \,|\, \bar{y}) = \infty$.

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 $x \in \operatorname{dom} F$ close to $\bar{x}$. This difference, which 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 $(\bar{x}, \bar{y}) \in \operatorname{Graph} F$ 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 $0 \cdot \infty = 1$,*

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

https://doi.org/10.1017/S1446788715000701 Published online by Cambridge University Press

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 points 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 $0 \le t < \varepsilon$ (see [56]); 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 ε in the original and modified definitions may be different.

Definition 2.4 (Graph regularity [95]). F is said to be *graph regular at (or near)* $(\bar{x}, \bar{y}) \in \operatorname{Graph} F$ if there are $K > 0$, $\varepsilon > 0$ such that the inequality

$$d(x, F^{-1}(y)) \le Kd((x, y), \operatorname{Graph} F),$$

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

Proposition 2.5 (Metric regularity versus graph regularity [95]). *Let* $F : X \rightrightarrows Y$ *and* $(\bar{x}, \bar{y}) \in \operatorname{Graph} F$. *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 nonlocal extension of this result (for example, along the lines of the subsection below) is wrong.

2.2. Nonlocal regularity. As we have already mentioned, most current research focuses on local regularity (although the first abstract definition of the covering property given in [26] was absolutely nonlocal). 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 (that is, optimality conditions, stability of solutions under small perturbations and so on). Another less visible reason is that nonlocal regularity is a more delicate concept: in the nonlocal case we cannot freely change the regularity domain that is an integral part of the definition. Meanwhile, nonlocal regularity is a powerful instrument for proving, for example, various existence theorems (see, for example, [61, Section 8.7]).

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

https://doi.org/10.1017/S1446788715000701 Published online by Cambridge University Press

DEFINITION 2.6 (Nonlocal regularity properties [56]). We say that F is

- *γ-open (or γ-covering) at a linear rate* on $U \times V$ if there is an $r > 0$ such that
$$B(F(x), rt) \cap V \subset F(B(x, t))$$
if $x \in U$ and $t < \gamma(x)$. Denote by $\mathrm{sur}_{\gamma} F(U \mid V)$ the upper bound of such r. If no such r exists, set $\mathrm{sur}_{\gamma} F(U \mid V) = 0$. We shall call $\mathrm{sur}_{\gamma} F(U \mid V)$ the *modulus* (or *rate*) *of γ-openness* of F on $U \times V$;
- *γ-metrically regular on* $U \times V$ if there is a $K > 0$ such that
$$d(x, F^{-1}(y)) \le K d(y, F(x)),$$
provided $x \in U$, $y \in V$ and $K d(y, F(x)) < \gamma(x)$. Denote by $\mathrm{reg}_{\gamma} F(U \mid V)$ the lower bound of such K. If no such K exists, set $\mathrm{reg}_{\gamma} F = \infty$. We shall call $\mathrm{reg}_{\gamma} F(U \mid V)$ the *modulus* (or *rate*) *of γ-metric regularity* of F on $U \times V$; and
- *δ-pseudo-Lipschitz* on $U \times V$ if there is a $K > 0$ such that
$$d(y, F(x)) \le K d(x, u)$$
if $x \in U$, $y \in V$, $K d(x, u) < \delta(y)$ and $y \in F(u)$. Denote by $\mathrm{lip}_{\delta} F(U \mid V)$ the lower bound of such K. If no such K exists, set $\mathrm{lip}_{\delta} F = \infty$. We shall call $\mathrm{lip}_{\delta} F(U \mid V)$ the *δ-Lipschitz modulus* of F on $U \times V$.

If $U = X$ and $V = Y$, let us agree to write $\mathrm{sur}_{\gamma} F, \mathrm{reg}_{\gamma} F, \mathrm{lip}_{\delta} F$ instead of $\mathrm{sur}_{\gamma} F(X|Y)$ and so on. The role of the functions γ and δ 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, for example, Exercise 2.8 below). But, for fixed U and V, a regularity horizon function is an essential element of the definition. Regularity properties corresponding to different γ may not be equivalent (see [60, Example 2.2] 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 γ-open at a linear rate on $U \times V$;*
(b) *F is γ-metrically regular on $U \times V$; and*
(c) *F^{-1} is γ-pseudo-Lipschitz on $V \times U$.*

Moreover (under the convention that $0 \cdot \infty = 1$),
$$\mathrm{sur}_{\gamma} F(U \mid V) \cdot \mathrm{reg}_{\gamma} F(U \mid V) = 1, \quad \mathrm{reg}_{\gamma} F(U \mid V) = \mathrm{lip}_{\gamma} F^{-1}(V \mid U).$$

PROOF. The implication (b) $\Rightarrow$ (c) is trivial. Hence $\mathrm{lip}_{\gamma} F^{-1}(V \mid U) \le \mathrm{reg}_{\gamma} F(U \mid V)$. To prove that (c) $\Rightarrow$ (a), take a $K > \mathrm{lip}_{\gamma} F^{-1}$ and an $r < K^{-1}$, let $t < \gamma(x)$, and let

https://doi.org/10.1017/S1446788715000701 Published online by Cambridge University Press

$x \in U$, $y \in V$, $v \in F(x)$ and $y \in (Bv, rt)$. Then $d(y,v) < r\gamma(x)$ and, by (c), $d(x, F^{-1}(y)) \le Kd(y,v) < r^{-1}d(y,v) \le t$. 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 $r \le \mathrm{sur}_{\gamma} F$, or, equivalently, $1 \le K \mathrm{sur}_{\gamma} F$. But r can be chosen arbitrarily close to K^{-1} and K can be chosen arbitrarily close to $\mathrm{lip}_{\gamma} F^{-1}$. So we conclude that $\mathrm{sur}_{\gamma} F \cdot \mathrm{lip}_{\gamma} F^{-1} \ge 1$.

Finally, let (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 $d(x,u) \le t$. Thus $d(x, F^{-1}(y)) \le t = d(y,v)/r$. But $d(y,v)$ can be chosen arbitrarily close to $d(y, F(x))$ and we get $d(x, F^{-1}(y)) \le r^{-1} d(y, F(x))$: that is, $r \cdot \mathrm{reg}_{\gamma} F \le 1$. On the other hand, r can be chosen arbitrarily close to $\mathrm{sur}_{\gamma} F$ and we can conclude that $\mathrm{sur}_{\gamma} F \cdot \mathrm{reg}_{\gamma} F \le 1$ so that

$$1 \ge \mathrm{sur}_{\gamma} F(U \mid V) \cdot \mathrm{reg}_{\gamma} F(U \mid V) \ge \mathrm{sur}_{\gamma} F(U \mid V) \cdot \mathrm{lip}_{\gamma} F(V \mid U) \ge 1,$$

which completes the proof of the theorem. □

The most important example of the horizon function is $m(x) = d(x, X \setminus U)$. This means 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 [26].) In what follows, we shall deal only with Milyutin regularity when speaking about nonlocal matters.

Exercise 2.8. A set-valued mapping F is regular near $(\bar{x}, \bar{y}) \in \mathrm{Graph}\, F$ if and only if it is Milyutin regular on $\overset{\circ}{B}(\bar{x}, \varepsilon) \times \overset{\circ}{B}(\bar{y}, \delta)$ for all sufficiently small $\varepsilon > 0$ and $\delta > 0$.

We conclude the section with a useful result (a slight modification of the corresponding result in [54]) showing that, as far as metric regularity is concerned, any set-valued mapping can be equivalently, and in a canonical way, replaced by a single-valued mapping continuous on its domain.

Proposition 2.9 (Single-valued reduction). *Let F be Milyutin regular on $U \times V$ with $\mathrm{sur}_m F(U \mid V) \ge r > 0$. Consider the mapping $\mathcal{P}_F : \mathrm{Graph}\, F \to Y$ which is the restriction to* Graph F *of the Cartesian projection $(x,y) \to y$.*

Then $\mathcal{P}_F$ is Milyutin regular on $(U \times Y) \times V$ and $\mathrm{sur}_m \mathcal{P}_F(U \times Y | V) = \mathrm{sur}_m F(U \mid V)$ if $X \times Y$ is considered, for example, with the ξ-metric to be defined in the next subsection.

A few bibliographic comments. To begin with, it is worth mentioning that, in the classical theory, no interest in metric estimates can be traced. The covering property close to the covering part of Milyutin regularity was introduced in [26] and attributed to Milyutin. An estimate of a type of metric regularity first appeared in Lyusternik's paper [73] but for x restricted to the kernel of the derivative. In Ioffe and Tikhomirov [64], metric regularity was proved under the assumptions of the Graves theorem. Robinson was probably the first to consider set-valued mappings. In [88], 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

https://doi.org/10.1017/S1446788715000701 Published online by Cambridge University Press

that $f : X \to Y$ is continuously differentiable and $K \subset Y$ is a closed convex cone, under a certain qualification condition extending Lyusternik's $\mathrm{Im}\, F'(x) = Y$. The definition of γ-regularity was given in [56].

Equivalence of covering and metric regularity were explicitly mentioned (without proof) in the paper of Dmitruk *et al.* [26] which marked the beginning of systematic study of the regularity phenomena (in particular in metric spaces) and Ioffe, in [49], stated a certain equivalence result (Proposition 11.12—see [53] for its proof) which, as was much later understood, contains even more precise information about the connection of the covering and metric regularity properties. The pseudo-Lipschitz property was introduced by Aubin in [4].

This was the sequence of events prior to the proof of the equivalence of the three properties by Borwein and Zhuang [16] and Penot [84]. 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 [41] 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 are Theorems 3.1–3.3 containing general regularity criteria. The criteria (especially the first of them) will serve 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*, which is the central concept 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.

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

Note that φ_y is Lipschitz continuous on Graph F, and hence it is lower semicontinuous whenever Graph F is a closed set.

3.1. General criteria. Given a $\xi > 0$, we define the ξ-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. Further, let $r > 0$ and let there be a $\xi > 0$ such that, for any $x \in U$, $y \in V$, $v \in F(x)$ with $0 < d(y, v) < rm(x)$, there is a pair $(u, w) \in$ Graph F different from (x, v) and such that*

$$d(y, w) \le d(y, v) - r d_\xi((x, v), (u, w)). \tag{3.1}$$

Then F is Milyutin regular on $U \times V$ with $\mathrm{sur}_m F(U \mid V) \ge r$.

https://doi.org/10.1017/S1446788715000701 Published online by Cambridge University Press

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

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

Proof. We have to verify that, given $(\overline{x}, \overline{v}) \in \mathrm{Graph}\, F$ with $\overline{x} \in U$, $y \in V$ and $0 < d(y, \overline{v}) \leq rt$, $t < m(\overline{x})$, there is a $u \in B(x, t)$ such that $y \in F(u)$. We have $\varphi_y(\overline{x}, \overline{v}) \leq rt$. By Ekeland's variational principle (see, for example, [17]) there is a pair $(\hat{x}, \hat{v}) \in \mathrm{Graph}\, F$ such that $d_{\xi}((\hat{x}, \hat{v}), (\overline{x}, \overline{v})) \leq t$ and

$$\varphi_y(x, v) + r d_{\xi}((x, v), (\hat{x}, \hat{v})) > \varphi_y(\hat{x}, \hat{v}) \tag{3.2}$$

if $(x, v) \neq (\hat{x}, \hat{v})$. We claim that $\varphi_y(\hat{x}, \hat{v}) = 0$: that is, $y = \hat{v} \in F(\hat{x})$. Indeed, $\hat{x} \in U$, so, by 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 $(\hat{x}, \hat{v})$ as (x, v). However, this contradicts (3.2). This proves the first statement.

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

$$d(y, w) = 0 \leq d(y, v) - (r - \varepsilon)t \leq d(y, v) - (r - \varepsilon)\, d_{\xi}((x, v), (u, w)).$$

Since ε can be chosen arbitrarily small, the result follows. □

Theorem 3.2 (Second criterion for Milyutin regularity). *Let X be a complete metric space, $U \subset X$ and $V \subset Y$ be open sets and $F : X \rightrightarrows Y$ be a set-valued mapping with closed graph. Then F is Milyutin regular on $U \times V$ with $\mathrm{sur}_m F(U \mid V) \geq r$ if and only if, for any $x \in U$ and any $y \in V$ with $0 < \overline{\psi}_y(x) < rm(x)$, there is a $u \neq x$ such that*

$$\overline{\psi}_y(u) \leq \overline{\psi}_y(x) - r d(x, u). \tag{3.3}$$

Proof. The proof of sufficiency is similar to the proof of the first part of the previous theorem.

To prove that (3.3) is necessary for Milyutin regularity, take $x \in U, y \in V$ such that $0 < d(y, F(x)) < rm(x)$. Take $\rho < r$ such that still $d(y, F(x)) < \rho m(x)$, and let $\rho < \rho' < r$. Let $x_n \to x$ be such that $d(y, F(x_n)) \to \overline{\psi}_y(x)$. We may assume that $d(y, F(x_n)) < rm(x)$ for all n. Choose positive $\delta_n \to 0$ such that $d(y, F(x_n)) \leq (1 + \delta_n)\overline{\psi}_y(x)$, and let t_n be defined by $\rho' t_n = (1 + \delta_n)\overline{\psi}_y(x)$. Then $y \in \mathring{B}(F(x_n), \rho' t_n)$, $t_n < m(x_n)$ (at least for large n) and, due to the regularity assumption on F, for any n we can find a u_n such that $d(u_n, x_n) < t_n$ and $y \in F(u_n)$. Note that u_n are bounded away from x, otherwise

https://doi.org/10.1017/S1446788715000701 Published online by Cambridge University Press

(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

$$\begin{aligned}\overline{\psi}_y(u_n) = 0 = \overline{\psi}_y(x) - \overline{\psi}_y(x) &= \overline{\psi}_y(x) - \frac{\rho' t_n}{1+\delta_n}\\ &\le \overline{\psi}_y(x) - \frac{\rho'}{1+\delta_n} d(u_n, x_n)\\ &= \overline{\psi}_y(x) - \frac{\lambda_n \rho'}{1+\delta_n} d(u_n, x) \le \overline{\psi}_y(x) - \rho d(u_n, x),\end{aligned}$$

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

The theorem is especially convenient when ψ_y is lower semicontinuous for every $y \in V$. Otherwise, the need for preliminary calculation of $\overline{\psi}_y$, the lower closure of ψ_y, may cause difficulties. It is possible, however, to modify the condition of the theorem and get a statement that requires verification of a (3.3)-like inequality for ψ 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* $\operatorname{sur} F(\overline{x} \mid \overline{y}) \ge r$ *is that there is a $\lambda \in (0, 1)$ and, for any $x \in U$ and $y \in V$ with $0 < \psi_y(x) < rm(x)$, there is a $u \ne x$ such that*

$$\psi_y(u) \le \psi_y(x) - rd(x, u), \quad \psi_y(u) \le \lambda \psi_y(x). \tag{3.4}$$

Proof. The key to understanding the theorem is the implication

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

which is, 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 $\psi_y(x_n) \to 0$. This, in turn, implies the existence of $v_n \in F(x_n)$ converging to y. As the graph of F is closed, it follows that $(x, y) \in \operatorname{Graph} F$, 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 $x_n \to x$ such that $\psi_y(x_n) = \alpha_n \to \alpha$ and, for each n, a u_n such that $\psi_y(u_n) \le \lambda\alpha_n$ and $\psi_y(u_n) \le \psi_y(x_n) - rd(x_n, u_n)$. An easy calculation shows that

$$\psi_y(u_n) \le \overline{\psi}_y(x) - rd(x, u_n) + \varepsilon_n,$$

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

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

holds for sufficiently large n. This allows us to apply Theorem 3.2 and conclude (by virtue of (3.5)) that there is a $w \in B(x, (r')^{-1})$ such that $y \in F(x)$: $\operatorname{sur}_m F(U \mid V) \ge r'$. □

https://doi.org/10.1017/S1446788715000701 Published online by Cambridge University Press

Note that the proof of necessity in the two last theorems does not much differ from the proof of Theorem 3.1. Corresponding criteria for local regularity are immediate.

Theorem 3.4 (Criterion for local regularity). *Let $F : X \rightrightarrows Y$ be a set-valued mapping with closed graph, and let $(\bar{x}, \bar{y}) \in \operatorname{Graph} F$. Then F is regular near $(\bar{x}, \bar{y})$ if and only if there are $\varepsilon > 0$, $\xi > 0$ and $r > 0$ such that, for any x, v and y satisfying $d(x, \overline{x}) < \varepsilon, d(y, \overline{y}) < \varepsilon, v \in F(x)$ and $0 < d(y, v) < \varepsilon$, either of the following two properties is valid.*

(a) $\operatorname{Graph} F$ *is locally complete and there is a pair $(u, w) \in \operatorname{Graph} F$, $(u, w) \neq (x, v)$ such that (3.1) 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, $\operatorname{sur} F(\overline{x} \mid \overline{y}) \geq r$.

Theorem 3.1 is a particular case of the criterion for γ-regularity proved in [56]. Theorem 3.4 is a modification of the result established in [54]. Theorems 3.2 and 3.3 are new but the first was largely stimulated by a recent result of Ngai *et al.* [81] (see Theorem 3.12 later in this section) and by a much earlier observation by Cominetti [24] that $\overline{\psi}_y(x) = 0$ implies that $y \in F(x)$. Surprisingly, it has been recently discovered that sufficiency in the statement of part (a) of the local criterion (Theorem 3.4) is present as a remark in a much earlier paper by Fabian and Preiss [38].

The completeness assumption in the first theorem differs from the corresponding assumption of the other two theorems. So it is natural to ask if and how they are connected. It is an easy matter to see, in view of Proposition 2.9, that Theorem 3.1 follows from Theorem 3.2. On the other hand, Theorem 3.1 is easier to use as it does not need *a priori* calculation of any limit or verification of the existence of λ, as in the third theorem. However, if the functions $d(y, F(\cdot))$ are lower semicontinuous, the second criterion may be more convenient. It should also be observed that the theorems can be equivalent in some cases (as follows from [54, Proposition 1.5]).

3.2. An application: density theorem. Here is the first example demonstrating how handy and powerful the criteria are.

Theorem 3.5 (Density theorem [26, 56]). *Let $U \subset X$ and $V \subset Y$ be open sets and let $F : X \rightrightarrows Y$ 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 ℓt-net in $B(v, rt) \bigcap V$, where $0 \leq \ell < r$. Then F is Milyutin regular on $U \times V$ and* $\operatorname{sur}_m F \geq r - \ell$. *In particular, if $F(B(x, t))$ is dense in $B(F(x), rt) \bigcap V$ for $x \in U$ and $t < m(x)$, then* $\operatorname{sur}_m F(U \mid V) \geq r$.

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 $(u, w) \in \operatorname{Graph} F$ such that $d(x, u) \leq t$ and $d(y, w) \leq \ell t = (\ell/r)\, d(y, v)$. Then

$$d(v, w) \leq d(y, v) + d(y, w) \leq \left(1 + \frac{\ell}{r}\right) d(y, v) \leq 2\, d(y, v).$$

https://doi.org/10.1017/S1446788715000701 Published online by Cambridge University Press

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

$$d(y,w) \le \ell t = rt - (r-\ell)t = d(y,v) - (r-l)t \le d(y,v) - (r-\ell)\, d_{\xi}((x,v), d(u,w)).$$

Apply Theorem 3.1. □

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 the following corollary.

Corollary 3.8 (Density theorem—local version). *Suppose there are $r > 0$, and $\varepsilon > 0$ such that $F(B(x,t))$ is an ℓt-net in $B(v,rt)$ whenever $d(x,\overline{x}) < \varepsilon$, $d(v,\overline{y}) < \varepsilon$, $v \in F(x)$ and $t < \varepsilon$. Then $\operatorname{sur} F(\overline{x}\,|\,\overline{y}) \ge r - \ell$. Thus, if $B(v,rt) \subset \operatorname{cl} F(B(x,t))$ for all x, 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 development. Results in the spirit of Corollary 3.8 were first considered in Ptak [86], Tziskaridze [96] and Dolecki [27, 28] in the 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 (for example, [16, 98]) were primarily concentrated on results of such type. We refer to [10] for detailed discussions and many references. Dmitruk *et al.* in [26] 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 ℓt-net in $B(F(x),rt)$. This opened the 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 [3].

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 *et al.* [25] 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

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

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 $|\nabla f|(x) < \infty$.

https://doi.org/10.1017/S1446788715000701 Published online by Cambridge University Press

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 us to reformulate the sufficient part of the regularity criteria of Theorem 3.4 in infinitesimal terms. To this end, as before, set

$$\varphi_y(x, v) = d(y, v) + i_{\operatorname{Graph} F}(x, v), \quad \psi_y(x) = d(y, F(x)), \quad \overline{\psi}_y(x) = \liminf_{u \to x} \psi_y(u),$$

and let ∇_ξ stand for the slope of functions on $X \times Y$ with respect to the d_ξ-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 *locally coherent* if for any x

$$\lim_{\substack{u,w \to x\\ u \neq w}} |\nabla d(u, \cdot)|(w) = 1.$$

It can be shown that a convex set and a smooth manifold in a Banach space are locally coherent in the induced metric [55] and that any length metric space (space whose metric is defined by minimal lengths of curves connecting points) is locally coherent (as follows from [8]).

THEOREM 3.10 (Local regularity criterion 1 [55]). *Let X and Y be metric spaces, let $F : X \rightrightarrows Y$ be a set-valued mapping, and let $(\bar{x}, \bar{y}) \in \operatorname{Graph} F$. We assume that $\operatorname{Graph} F$ is locally complete at $(\bar{x}, \bar{y})$. Suppose, further, that there are $\varepsilon > 0$ and $r > 0$ such that, for some $\xi > 0$,*

$$|\nabla_\xi \varphi_y|(x, v) > r$$

if

$$v \in F(x), \quad d(x, \overline{x}) < \varepsilon, \quad d(y, \overline{y}) < \varepsilon, \quad d(v, \overline{y}) < \varepsilon, \quad v \neq y. \tag{3.6}$$

Then F is regular near $(\bar{x}, \bar{y})$ with $\operatorname{sur} F(\bar{x}, \bar{y}) \geq r$.

Conversely, let Y be locally coherent at $\overline{y}$. Assume that $\operatorname{sur} F(\overline{x} \mid \overline{y}) > r > 0$. 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 (3.6). Thus, in this case,

$$\operatorname{sur} F(\bar{x}, \bar{y}) = \liminf_{\substack{(x,v) \underset{\operatorname{Graph} F}{\to} (\bar{x},\bar{y})\\ y \to \bar{y},\ y \neq v}} |\nabla_\xi \varphi_y|(x, v).$$

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.

COROLLARY 3.11. *Suppose, under the conditions of Theorem 3.10, that Y is metrically convex. Then, for any neighborhood V of $\overline{y}$,*

$$\operatorname{sur} F(\bar{x}, \bar{y}) = \liminf_{(x,v) \underset{\operatorname{Graph} F}{\to} (\bar{x},\bar{y})} \inf_{y \in V \setminus \{v\}} |\nabla_\xi \varphi_y|(x, v).$$

https://doi.org/10.1017/S1446788715000701 Published online by Cambridge University Press

THEOREM 3.12 (Local regularity criterion 2). *Suppose that X is complete and the graph of F is closed. Assume, further, that there are neighborhoods $U \subset X$ of $\bar{x}$ and $V \subset Y$ of $\bar{y}$, $r > 0$ and $\varepsilon > 0$ such that that $|\nabla\overline{\psi}_y|(x) > r$ for all $(x, y) \in U \times V$ such that $\varepsilon > \overline{\psi}_y(x) > 0$. Then* $\operatorname{sur} F(\bar{x} \mid \bar{y}) \geq r$.

Conversely, if, in addition, Y is a length space and $\operatorname{sur} F(\bar{x} \mid \bar{y}) > r > 0$, *then there is a neighborhood of $(\bar{x}, \bar{y})$ and an $\varepsilon > 0$ such that $|\nabla\overline{\psi}_y|(x) \geq r$ for all (x, y) of the neighborhood such that $y \notin F(x)$ and $0 < \overline{\psi}_y(x) < \varepsilon r$. Thus, in this case,*

$$\operatorname{sur} F(\bar{x} \mid \bar{y}) = \liminf_{\substack{(x,y)\to(\bar{x},\bar{y})\\ 0\neq d(y,F(x))\to 0,}} |\nabla\overline{\psi}_y|(x).$$

In particular, if $\psi_y = d(y, F(\cdot))$ is lower semicontinuous at every x of a neighborhood of $\bar{x}$ and for every $y \notin F(x)$ close to $\bar{y}$, then

$$\operatorname{sur} F(\bar{x} \mid \bar{y}) = \liminf_{\substack{(x,y)\to(\bar{x},\bar{y})\\ 0\neq d(y,F(x))\to 0}} |\nabla\psi_y|(x).$$

The starting point for developing slope-based regularity theory was the paper by Azé *et al.* [9] (its first version was circulated in 1998) who obtained a global error bound in terms of 'variational pairs' that included 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 [54] under a somewhat stronger condition (equivalent to Y being a length space). We refer to [6] 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 *et al.* [81] (proved originally for Y being a Banach space).

To explain how the additional assumption on Y is used to get necessity, for example in Theorem 3.10, let us consider, following the original argument in [54], (x, y, v) sufficiently close to $\bar{x}$ and $\bar{y}$, 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 $r\xi < 1$)

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

A 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 have attracted much attention over the years. We refer to [35] 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

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

The upper bound of such r is the *rate* or *modulus of controllability* of F *at* $(\bar{x}, \bar{y})$. We shall denote it $\operatorname{contr} F(\bar{x} \mid \bar{y})$ and $\operatorname{contr} F(\bar{x})$ if F is single valued.

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

$$\operatorname{sur} F(\bar{x} \mid \bar{y}) = \lim_{\varepsilon \to 0} \inf\{\operatorname{contr} F(x \mid y) : (x, y) \in \operatorname{Graph} F,\ \max\{d(x, \bar{x}), d(y, \bar{y})\} < \varepsilon\}.$$

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

$$d(\bar{y}, F(x)) \le K d(x, \bar{x}) \quad \text{if } d(x, \bar{x}) < \varepsilon. \tag{3.7}$$

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 by $\operatorname{ress} F(\bar{x} \mid \bar{y})$.

The other possible way to 'pointify' the Aubin property is to fix $\bar{x}$ and allow (x, y) to change within $\operatorname{Graph} F$. Then, instead of (3.7), we get the inequality

$$d(y, F(\bar{x})) \le K d(x, \bar{x}). \tag{3.8}$$

Definition 3.16 (Calmness). It is said that $F : X \rightrightarrows Y$ is *calm* at $(\bar{x}, \bar{y})$ if there are $\varepsilon > 0$, $K \ge 0$ such that (3.8) holds if $d(x, \bar{x}) < \varepsilon,\ d(y, \bar{y}) < \varepsilon$ 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 $\operatorname{calm} F(\bar{x} \mid \bar{y})$ ($\operatorname{calm} F(\bar{x})$ if F is single valued).

Again, we can easily see that *uniform calmness* (that is, calmness at every (x, y) of the intersection of $\operatorname{Graph} F$ with a neighborhood of $(\bar{x}, \bar{y})$ with the same ε 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

$$d(x, F^{-1}(\bar{y})) \le K d(\bar{y}, F(x)) \quad \text{if } d(x, \bar{x}) < \varepsilon \tag{3.9}$$

https://doi.org/10.1017/S1446788715000701 Published online by Cambridge University Press

for all x of a neighborhood of $\overline{x}$. 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 by $\operatorname{subreg} F(\overline{x}\,|\,\overline{y})$.

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

Proposition 3.18. *The equalities*

$$\operatorname{subreg} F(\overline{x}\,|\,\overline{y}) = \operatorname{calm} F^{-1}(\overline{y}\,|\,\overline{x}), \quad \operatorname{contr} F(\overline{x}\,|\,\overline{y}) \cdot \operatorname{ress} F^{-1}(\overline{y}\,|\,\overline{x}) = 1$$

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

$$\operatorname{contr} F(\overline{x}\,|\,\overline{y}) \cdot \operatorname{subreg} F(\overline{x}\,|\,\overline{y}) \geq 1.$$

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 $(\bar{x}, \bar{y}) \in \operatorname{Graph} F$. Then*

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

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

Proof. Let $K > \operatorname{calm} F(\overline{x}\,|\,\overline{y})$. Then there is an $\varepsilon > 0$ such that (3.8) holds, provided $d(x, \overline{x}) < \varepsilon$ and $y \in F(x)$. To prove the theorem, it is sufficient to show that $|\nabla \varphi_y|(\overline{x}) \leq K$ for all y sufficiently close to $\overline{y}$. To this end, it is sufficient to verify that there is a $\delta > 0$ such that the inequality

$$d(y, F(\overline{x})) - d(y, F(x)) \leq K d(x, \overline{x}) \tag{3.10}$$

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

If $y \in F(x)$, then (3.10) reduces to (3.8). Take a positive $\delta < \varepsilon/2$, and let x and y be such that $d(x, \overline{x}) < \delta,\ d(y, \overline{y}) < \delta$. If $d(y, F(x)) \geq \delta$, then (3.10) obviously holds. If $d(y, F(x)) < \delta$, we can choose a $v \in F(x)$ such that $d(y, v) < \delta$. Then $d(v, \overline{y}) < \varepsilon$ and therefore $d(v, F(\overline{x})) \leq d(x, \overline{x})$. Thus

$$\begin{aligned} d(y, F(\overline{x})) - d(y, F(x)) &\leq d(y, v) + d(v, F(\overline{x})) - d(y, F(x)) \\ &\leq K d(x, \overline{x}) + d(y, v) - d(y, F(x)) \end{aligned}$$

and (3.10) follows as $d(y, v)$ can be arbitrarily close to $d(y, F(x))$. □

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 $(\bar{x}, \bar{y}) \in \operatorname{Graph} F$. Assume that the function $\psi_{\overline{y}} = d(\overline{y}, F(x))$ is lower semicontinuous and there are $\varepsilon > 0$ and $r > 0$ such that*

$$|\nabla \psi_{\overline{y}}|(x) = |\nabla d(\overline{y}, F(\cdot))|(x) \geq r,$$

if $d(x, \overline{x}) < \varepsilon$ and $0 < d(\overline{y}, F(x)) < \varepsilon$. Then F is subregular at $(\bar{x}, \bar{y})$ with modulus of subregularity (and hence the modulus of calmness of F^{-1} at $(\overline{y}, \overline{x})$) not greater than r^{-1}.

https://doi.org/10.1017/S1446788715000701 Published online by Cambridge University Press

4. Metric theory. Perturbations and stability

In this section, we concentrate on the following two fundamental questions.

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

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

The answer to the second question leads us to a fairly general implicit function theorem. The key point in both cases is that we have to require a certain amount of Lipschitzness of perturbations to get the desired results.

4.1. Stability under Lipschitz perturbation.

THEOREM 4.1 (Stability under Lipschitz perturbation). *Let X, Y be metric spaces and 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 Ψ 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 $y \in \overset{\circ}{B}(v, rt) \cap V$ with $t < d(x, X \backslash U)$ there is an x' such that $d(x, x') \le r^{-1} d(y, v)$ and $y \in F(x')$; and*

(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 $u, w \in U$*

$$\operatorname{ex}(\Psi(x, u) \cap V, \Psi(x, w)) < \ell d(u, w).$$

Then $F(x) = \Psi(x, x)$ is Milyutin regular on (U, V) with $\operatorname{sur}_m F(U \mid V) \ge r - \ell$.

PROOF. We shall consider only the case of complete graph of Ψ. 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 $(x, v) \in grF$ and y such that $x \in U$, $y \in V$ and $0 < d(y, v) < rm(x)$, there is another point $(x', v') \ne (x, v)$ in the graph of F such that

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

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 $d(x, x') \le r^{-1} d(y, v)$.

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

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

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

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

https://doi.org/10.1017/S1446788715000701 Published online by Cambridge University Press

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

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

as needed. □

COROLLARY 4.2 (Milyutin's perturbation theorem [26]). *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. Further, let $U \subset X$ be an open set such that F is Milyutin regular on U with $\operatorname{sur} F(U) \ge r$ and G is (Hausdorff) Lipschitz with $\operatorname{lip} G(U) \le \ell < r$. If either F or G is single valued continuous on U, then $F + G$ is Milyutin regular on $U \times Y$ and $\operatorname{sur}(F + G)(U) \ge r - \ell$.*

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.

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 $(\bar x,\bar y) \in \operatorname{Graph} F(\bar p,\cdot)$. We shall say that F is regular near $(\bar x,\bar y)$ *uniformly* in $p \in P$ near $\bar p$ if, for any $r < \operatorname{sur} F(\bar p,\cdot)(\bar x \mid \bar y)$, there are $\varepsilon > 0$ and a neighborhood $W \subset P$ of $\bar p$ such that, for any $p \in W$ and any x with $d(x,\bar x) < \varepsilon$,

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

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 $(\bar x,\bar y) \in \operatorname{Graph} F$. We assume that*

(a) $\Psi(\cdot,u)$ *is regular near $(\bar x,\bar y)$ uniformly in u near $\bar x$; and*
(b) $\Psi(x,\cdot)$ *is pseudo-Lipschitz near $(\bar x,\bar y)$ uniformly in x near $\bar x$.*

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

The last theorem, in turn, immediately implies a local version of 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, let $F : X \rightrightarrows Y$ have closed graph, let $(\bar x,\bar y) \in$ Graph F, and let $g : X \to Y$ be ℓ-Lipschitz in a neighborhood of $\bar x$. Assume finally that either X or Graph F is complete and F is regular near $(\bar x,\bar y)$ with $\operatorname{sur} F(\bar x \mid \bar y) > r > \ell$. Then*

$$\operatorname{sur}(F + g)(\bar x,\bar y + g(\bar x)) \ge r - \ell.$$

Specifically, if F is Milyutin regular on $\mathring{B}(\bar x,\varepsilon) \times \mathring{B}(\bar y,\delta)$ for some $\varepsilon > 0, \delta > 0$ with $\operatorname{sur} F(\mathring{B}(\bar x,\varepsilon)\mathring{B}(\bar y,\delta)) > r$, then $\operatorname{sur}(F + g)(\mathring{B}(\bar x,\varepsilon')\mathring{B}(\bar y + g(\bar x),\delta')) > r - \ell$, where $\varepsilon' < \varepsilon$ and $\delta' = \ell\varepsilon' < \delta$.

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. □

https://doi.org/10.1017/S1446788715000701 Published online by Cambridge University Press

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

Corollary 4.6 (Lyusternik–Graves from Mulyutin). *Let X and Y be Banach spaces, and let $F : X \to Y$ be strictly differentiable at $\overline{x}$. Then* $\operatorname{sur} F(\overline{x}) = C(F'(\overline{x}))$.

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

We observe next that, in Theorem 4.5, one of the mappings is assumed to be 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. [35]). Let $X = Y = \mathbb{R}$, $G(x, y) = \{x^2, -1\}$ and $F(x) = \{-2x, 1\}$. 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 $(\xi, 0) \in \operatorname{Graph} \Phi$ for any ξ. However, if $\xi \neq 0$, then the Φ-image of a sufficiently small neighborhood of ξ does not contain points of a small neighborhood of zero other than zero itself.

Perturbation analysis of regularity properties was initiated by Dmitruk *et al.* [26] with a proof of a global version of Theorem 4.2 (attributed in [26] to Milyutin) with both the mapping and the perturbation single valued. The first perturbation result for set-valued mappings was proved probably by Ursescu [97] (see also [54]). Observe that global theorems are valid for Lipschitz set-valued perturbations as well.

Until very recently, most attention was devoted to additive perturbations into a linear range space, especially in connection with implicit function theorems for generalized equations—see, for example, [7, 35]. Interest in nonadditive Lipschitz set-valued perturbations of set-valued mappings appeared just a few years ago, partly in connection with fixed point and coincidence theorems [2, 32, 56, 60].

The Graves theorem can be viewed as a perturbation theorem for a *linear* regular operator. For that reason, in some publications (for example, [31, 35]), Theorem 4.5 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 a similar idea when proving the metric regularity counterpart of the Graves theorem for [64] without any knowledge of Graves' paper.) There is also the fact that the Lipschitz property of the perturbation, as the key to the estimate, was explicitly emphasized in [26]. Note also that even Corollary 4.6 cannot be obtained from the Graves theorem.

https://doi.org/10.1017/S1446788715000701 Published online by Cambridge University Press

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 Φ, just by taking $\Psi(x, u) = \Phi(x, F(u))$. However, a certain caution is needed to guarantee that such a Ψ satisfies the required assumptions (as, say, in [56], where $\Phi(x, \cdot)$ is assumed to be an isometry, or in [37], where a certain 'composition stability' is *a priori* assumed). Corollary 4.6 was first stated in [30] 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 $F(x, p) : X \times P \rightrightarrows Y$. 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 ℓ^1-type distance

$$d^1_\alpha((y, p), (y', p')) = \alpha d(y, y') + d(p, p'),$$

where α will be further determined by Lipschitz moduli of the mappings involved.

THEOREM 4.8 (General proposition on implicit functions). *We assume that $\bar{y} \in F(\bar{x}, \bar{p})$ 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 $((\bar{x}, \bar{y}), \bar{p})$ uniformly in p with the rate of metric regularity not greater than K.*

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

Then S has the Aubin property near $((\bar{y}, \bar{p}), \bar{x})$ with the Lipschitz modulus with respect to the metric d^1_α in $Y \times P$ not greater than $\operatorname{reg} F(\cdot, \bar{p})(\bar{x}, \bar{y})$.

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

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

$$\begin{aligned} d(x, S(y', p')) &\le Kd(y', F(x, p')) \le K(d(y, y') + d(y, F(x, p'))) \\ &\le K(d(y, y') + \alpha d(p, p')) = Kd^1_\alpha((y, p), (y', p')) \\ &= K\alpha(d(p, p') + \alpha^{-1} d(y, y')), \end{aligned}$$

and the proof is complete. □

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

$$B(\bar{y}, \delta) \subset F(B(\bar{x}, \varepsilon)) \quad \text{and} \quad d(x, u) \le Kd(y, F(x)) \tag{4.1}$$

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

https://doi.org/10.1017/S1446788715000701 Published online by Cambridge University Press

We shall also say, following [35], that F *has a single-valued localization near* $(\bar{x}, \bar{y})$ if there are $\varepsilon > 0, \delta > 0$ such that the restriction of $F(x) \cap B(\bar{y}, \delta)$ to $B(\bar{x}, \varepsilon)$ is single valued. If, in addition, the restriction is Lipschitz continuous, we say that F has *Lipschitz localization* near $(\bar{x}, \bar{y})$.

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* $(\bar{x}, \bar{y}) \in \operatorname{Graph} F$. *Then the following properties are equivalent.*

(a) *F is strongly regular near $(\bar{x}, \bar{y})$.*

(b) *F is regular and there are $\varepsilon > 0$, $\delta > 0$ such that*

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

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

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

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

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 $\bar{y}$ coincide with $\operatorname{reg} F(\bar{x} \mid \bar{y})$.

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 $\bar{y} \in \Phi(\bar{x})$ and $\bar{z} = G(\bar{x}, \bar{y})$. We assume that:*

(a) *Φ is strongly regular near $(\bar{x}, \bar{y})$ with* $\operatorname{sur} \Phi(\bar{x} \mid \bar{y}) > r$;

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

(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 $(\bar{x}, \bar{z})$.

In particular, if Y is a normed space, Φ is strongly regular near $(\bar{x}, \bar{y}) \in \operatorname{Graph} \Phi$ *and $G(x, y) = g(x) + y$ with* $\operatorname{lip} g(\bar{x}) < \operatorname{sur} \Phi(\bar{x} \mid \bar{y})$, *then $F(x) = \Phi(x) + g(x)$ is strongly regular near $(\bar{x}, \bar{y} + g(\bar{x}))$.*

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. A simple example is

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

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 compared to the standard regularity.

https://doi.org/10.1017/S1446788715000701 Published online by Cambridge University Press

Theorem 4.13 (Implicit function theorem—metric version). *Assume, in addition to the assumptions of Theorem 4.8, that*

$$F(x,p)\cap F(x',p)\cap \overset{\circ}{B}(\bar y,\varepsilon)=\emptyset\quad \forall x,x'\in \overset{\circ}{B}(\bar x,\varepsilon), x\neq x',\ p\in \overset{\circ}{B}(\bar p,\varepsilon). \tag{4.2}$$

Then the solution map S has a Lipschitz localization G near $((\bar p,\bar y),\bar x)$ with $\operatorname{lip} G(\bar p,\bar y)\le K$ *(with respect to the d^1_α-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, (see e.g. the second proof in [35]).

Proof. Consider the set-valued mapping Φ 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 Φ is strongly regular near $((\bar x,\bar p),(\bar p,\bar y))$. Indeed, for x,p,y sufficiently close to $\bar x,\bar p,\bar y$,

$$\Phi^{-1}(x,y)=\{p\}\times S(p,y). \tag{4.3}$$

By Theorem 4.8, S has the Aubin property at $((\bar p,\bar y),\bar x)$. This obviously implies that Φ^{-1} has the Aubin property at $((\bar p,\bar y),(\bar x,\bar p)$. The latter means that Φ 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 (4.2) may happen only if $x=x'$. This proves the claim.

By Proposition 4.10, there is a Lipschitz localization of Φ^{-1} defined in a neighborhood of $(\bar p,\bar y)$. By (4.2), 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 $\xi>0, r>0, \ell>0, \varepsilon>0$ 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$_1$) $|\nabla_\xi\varphi_y(\cdot,p)|(x,v)>r$ *if $v\in F(x,p)$ and $d(y,v)>0$,*

or X is a complete space and

(a$_2$) $|\nabla\overline{\psi}_y(\cdot,p)|(x)>r$ *if* $\overline{\psi}_y(x,p)>0$

holds along with

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

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

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 4.13.

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

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

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

Proof. Indeed, it follows from (c) that $y \notin F(x,p)$: that is $(F(x,p) \cap F(x',p)) \cap \overset{\circ}{B}(\bar{y},\varepsilon) = \emptyset$ for x, x' close to $\bar{x}$ and p close to $\bar{p}$, and the references to Theorems 4.14 and 4.13 complete the proof. □

There have been numerous publications extending, one way or another, the implicit function theorem to settings of variational analysis (see, for example, [7, 35, 43, 54, 71, 80, 81]). Most of them deal with Banach spaces and/or specific classes of mappings, for example, associated with generalized equations. It should also be said that some results named 'implicit function theorem' are actually parametric regularity or subregularity theorems giving uniform (with respect to parameter) estimates for regularity rates of a mapping depending on a parameter.

The concept of strong regularity was introduced by Robinson in [89]. A number of characterizations of strong regularity can be found in [35]. 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. These are monotone operators, in particular subdifferentials of convex functions, or Kojima mappings associated with constrained optimization [35, 65].

5. Banach space theory

Needless to say, 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 approximation mechanisms, both primal and dual, using homogeneous mappings (graphical derivatives and coderivatives) in the case of set-valued mappings or directional subderivatives and subdifferentials for functions;

(b) the possibility of separable reduction for metric regularity that allows us to reduce much of analysis to mappings between separable spaces; and

https://doi.org/10.1017/S1446788715000701 Published online by Cambridge University Press

(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 $0 \in H(0)$. The mapping

$$\mathcal{H}^*(y^*) = \{x^* : \langle x^*, x\rangle - \langle y^*, y\rangle \le 0,\ \forall\,(x, y) \in \operatorname{Graph}\mathcal{H}\}$$

is called *adjoint* or *dual* to $\mathcal{H}$ (or the *dual convex process* as it is often called for reasons to be explained later). It is an easy matter to see that

$$\operatorname{Graph}\mathcal{H}^* = \{(y^*, x^*) : (x^*, -y^*) \in (\operatorname{Graph}\mathcal{H})^\circ\}.$$

With every homogeneous mapping $\mathcal{H}$, we associate the *upper norm*

$$\|\mathcal{H}\|_+ = \sup\{\|y\| : y \in \mathcal{H}(x),\ x \in \operatorname{dom}\mathcal{H},\ \|x\| \le 1\}$$

and the *lower norm*

$$\|\mathcal{H}\|_- = \sup_{x\in B\cap\operatorname{dom}\mathcal{H}} \inf\{\|y\| : y \in \mathcal{H}(x)\} = \sup_{x\in B\cap\operatorname{dom}\mathcal{H}} d(0, \mathcal{H}(x)).$$

For single-valued mappings with $\operatorname{dom}\mathcal{H} = X$, both quantities coincide and we may speak about the *norm* of $\mathcal{H}$. The mapping $\mathcal{H}$ is *bounded* if $\|\mathcal{H}\|_+ < \infty$. This obviously means that there is an $r > 0$ such that $\mathcal{H}(x) \subset r\|x\|B_Y$ 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

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

The quantities are, respectively, called the *Banach constant* and the *dual Banach constant* of $\mathcal{H}$. 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 $\mathcal{H} : X \rightrightarrows Y$*

$$C(\mathcal{H}) = \operatorname{contr}\mathcal{H}(0 \mid 0) = \sup\{r \ge 0 : rB_Y \subset \mathcal{H}(B_X)\},$$
$$C^*(\mathcal{H}) = (\operatorname{subreg}\mathcal{H}(0 \mid 0))^{-1} = \inf\{\|y\| : y \in \mathcal{H}(x),\ \|x\| = 1\} = \inf_{\|x\|=1} d(0, \mathcal{H}(x)).$$

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

https://doi.org/10.1017/S1446788715000701 Published online by Cambridge University Press

To prove that $C^*(\mathcal{H}) = \inf\{\|y\| : y \in \mathcal{H}(x), \|x\| = 1\}$ we first consider the case $C^*(H) < \infty$. Then

$$\begin{aligned} C^*(\mathcal{H}) &= \inf_{\|y\|=1} \inf\{\|x\|^{-1} : \ x \in \mathcal{H}^{-1}(y)\} \\ &= \inf\{\|y\| : y \in \mathcal{H}(x), \|x\| = 1\}. \end{aligned}$$

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

To prove the left inequality for $C^*(\mathcal{H})$ consider first the case $C^*(\mathcal{H}) > 0$. Then $\|\mathcal{H}^{-1}\|_+ < \infty$ and, consequently, $\mathcal{H}^{-1}(0) = \{0\}$. It follows that $d(x, \mathcal{H}^{-1}(0)) = \|x\|$. Setting $K = (C^*(\mathcal{H}))^{-1}$, we get, for any x with $\|x\| = 1$,

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

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

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

$$C(\mathcal{E} \circ \mathcal{H}) \geq C(\mathcal{E}) \cdot C(\mathcal{H}).$$

Proof. Take $\rho < C(\mathcal{H})$. Then $\rho(B_Y) \subset \mathcal{H}(B_X)$ and therefore

$$\begin{aligned} C(\mathcal{E} \circ \mathcal{H}) &= \sup\{r \geq 0 : \ rB_Z \subset (\mathcal{E} \circ \mathcal{H})(B_X)\} \\ &\geq \sup\{r \geq 0 : \ rB_Z \subset \mathcal{E}(\rho B_Y)\} = \rho C(\mathcal{E}) \end{aligned}$$

and the result follows. □

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 needs 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,

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

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

https://doi.org/10.1017/S1446788715000701 Published online by Cambridge University Press

Now let $\infty > C(\mathcal{H}) = r > 0$. 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 $\|x\| \le \lambda + \varepsilon$ and $y \in \mathcal{H}(x)$. Now let $x^* \in \mathcal{H}^*(y^*)$: that is $\langle x^*, x\rangle - \langle y^*, y\rangle \le 0$ if $y \in \mathcal{H}(x)$. Take $y \in S_Y$ such that $\langle y^*, y\rangle \le (-1 + \varepsilon)\|y^*\|$ and choose an $x \in \mathcal{H}^{-1}(y)$ with $\|x\| \le \lambda + \varepsilon$. Then

$$-(\lambda + \varepsilon)\|x^*\| \le \langle x^*, x\rangle \le \langle y^*, y\rangle \le (-1 + \varepsilon)\|y^*\|,$$

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

The following property plays an essential role in future discussions.

Definition 5.5 (Nonsingularity). We say that $\mathcal{H}$ is *nonsingular* if $C^*(\mathcal{H}) > 0$. Otherwise, we shall call $\mathcal{H}$ *singular*.

We conclude the subsection by 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 rate.*

Proof. By the assumption, there are $K > 0$ and $\varepsilon > 0$ such that $d(x, F^{-1}(y)) \le Kd(y, F(x))$ if $\max\{\|x\|, \|y\|\} < \varepsilon$. Now let (x, y) be an arbitrary point of the graph. Set $\|m\| = \max\{\|x\|, \|y\|\}$ and let $\mu < \varepsilon/m$. Then

$$\mu d(x, F^{-1}(y)) = d(\mu x, F^{-1}\mu y) \le d(\mu y, F(\mu x)) = \mu d(\mu y, F(\mu x))$$

and hence $d(x, F^{-1}(y)) \le Kd(y, F(x))$. □

The norms for homogeneous multifunctions were originally introduced by Rockafellar [90] and Robinson [87] in the context of convex processes (lower norm) and then by Ioffe [49] (upper norm for arbitrary homogenous maps) and Borwein [13] (upper norm and duality for convex processes; see also [14, 15, 35]). The dual Banach constant C^* was also introduced in [49]. The meaning of the primal constant has undergone some evolution since it first appeared in [49]. The $C(\mathcal{H})$ introduced here is reciprocal to that in [51], 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 $(\bar{x}, \bar{y})$ is the set-valued mapping

$$X \ni h \mapsto DF(\bar{x}, \bar{y})(h) = \{v \in Y : (h, v) \in T(\operatorname{Graph} F, (\bar{x}, \bar{y}))\}.$$

Now let f be a function on X which is finite at $\overline{x}$. The function

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

https://doi.org/10.1017/S1446788715000701 Published online by Cambridge University Press

is called the *Dini–Hadamard lower directional derivative* of f at $\bar{x}$. This function is either lower semicontinuous and equal to zero at the origin or identically equal to $-\infty$. The latter, of course, cannot happen if f is Lipschitz near $\bar{x}$.

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

If $F : X \rightrightarrows Y$, then the *contingent derivative* of F at $\bar{x}$ is the set-valued mapping

$$X \ni h \mapsto DF(\bar{x}; h) = \{v \in Y : (h, v) \in T(\operatorname{Graph} F, (\bar{x}, F(\bar{x})))\}.$$

The contingent tangent cone and contingent derivative were introduced by Aubin in [3] (see [5] 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 nonreflexive (or to be precise, non-Asplund) case.

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

Let f be a function on X which is finite at x. The sets

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

and

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

are called, respectively, the *Dini–Hadamard* and *Fréchet subdifferential* of f at x. The corresponding *limiting* subdifferentials at x (we denote them for the time being by ∂_{LH} and ∂_{LF}) are 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

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

The *G-subdifferential* of f at x is defined as

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

The cone $N_C(Q, x) = \operatorname{cl}(\operatorname{conv} N_G(Q, x))$ 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*.

https://doi.org/10.1017/S1446788715000701 Published online by Cambridge University Press

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

(a) *for any lower semicontinuous function* $\partial_H f(x) \neq \emptyset$ *on a dense subset of* dom f*;*
(b) *the same is true for* ∂_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* ∂ *being any of the mentioned subdifferentials;*
(g) *if* f *is Lipschitz near* x*, then* $\partial_{LH} f(x) = \partial_G f(x)$ *and* $\partial_C f(x) = \mathrm{cl}(\mathrm{conv}\, \partial_G(x))$*;*
(h) *if* f *is lower semicontinuous and* X *is an Asplund space, then* $\partial_{LF} f(x) = \partial_G f(x)$ *for any* x*; and*
(i) *if* $f(x, y) = \varphi(x) + \psi(y)$*, then* $\partial f(x, y) = \partial\varphi(x) + \partial\psi(y)$*, where* ∂ *is any of* $\partial_F,\ \partial_H, \partial_G$ *(but not* ∂_C*).*

Remark 5.8. The following should be observed, in connection with the proposition.

- ∂_{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, $\partial_{LF} f(x)$ may be identically empty even for a very simple Lipschitz function (for example, $-\|x\|$ in $C[0, 1]$). In the terminology of subdifferential calculus, this means that ∂_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, for example, [57]). Loosely speaking, this means that a version of the fuzzy variational principle is valid for the subdifferentials of lower semicontinuous functions on the space. Just note that the Fréchet subdifferential is trusted on Asplund spaces and only on them, the 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. This property is called *tightness* and it characterizes a reasonable quality of lower approximation provided by the subdifferential (see [57]). 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 the generalized gradient in the regularity theory. On the other hand, the generalized gradient, typically, is much easier to compute and work with. Moreover, convexity of the generalized

https://doi.org/10.1017/S1446788715000701 Published online by Cambridge University Press

gradient makes it the only subdifferential that can be used in the critical point theory associated with the concept of 'weak slope', which is not considered here.

We do not need, here, the general theory of subdifferentials. We 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 emphasized, again, that we assume that all spaces are separable.

Proposition 5.9 (Basic calculus rules). *Let $f(x) = f_1(x) + f_2(x)$, where both functions are lower semicontinuous and one of them is Lipschitz near $\bar{x}$. Then the following statements are true.*

(1) *Fuzzy variational principle: if f attains a local minimum at $\bar{x}$, then there are sequences (x_{in}) and (x^*_{in}), $i = 1, 2$ such that $x_{in} \to \bar{x}$, $x^*_{in} \in \partial_H f_{in}(x_{in})$ and $\|x^*_{1n} + x^*_{2n}\| \to 0$.*
(2) *Fuzzy sum rule: if X is Asplund and $x^* \in \partial_F f(\bar{x})$, then there are sequences (x_{in}) and (x^*_{in}), $i = 1, 2$ such that $x_{in} \to \bar{x}$, $x^*_{in} \in \partial_H f_{in}(x_{in})$ and $\|x^*_{1n} + x^*_{2n} - x^*\| \to 0$.*
(3) *Exact sum rule: $\partial_G f(\bar{x}) \subset \partial_G f_1(\bar{x}) + \partial_G f_2(\bar{x})$.*

Let $Q \subset X$ and $x \in Q$. Given a subdifferential ∂, the set

$$N(Q, x) = \partial i_Q(x),$$

which is always a cone, is called the *normal cone* to Q at x *associated with* ∂. It is an easy matter to see that, in the case of ∂_G, this definition coincides with that given earlier. For normal cones associated with ∂_H and ∂_F, we use the notation N_H and N_F.

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

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

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

There are a number of monographs and survey articles in which subdifferentials are studied at various levels of generality: [91] (finite-dimensional theory), [17, 77, 85, 94] (Asplund spaces), [57, 85] (general Banach spaces), [22, 23] (generalized gradients). Concerning the sources of the main concepts, Clarke's subdifferential was first to appear—it was introduced in Clarke's 1973 thesis [20] and in printed form first appeared in [21]. It is not clear where the Fréchet subdifferential first appeared; probably in [12]. The Dini–Hadamard subdifferential was introduced by Penot in [83], the sequential limiting Fréchet subdifferential for functions on Fréchet smooth spaces was introduced by Kruger in mimeographed paper [66] in 1981 (not in [70], as stated in, for example, [77, 78] and many other publications—the definition given in [70] is purely topological and does not involve sequential weak*-limits) and in printed form appeared in [67] (see [57] for details). The G-subdifferential was first defined in [48] but its definition was later modified in [52].

https://doi.org/10.1017/S1446788715000701 Published online by Cambridge University Press

5.2. Separable reduction. In this subsection, X and Y are general Banach spaces, which are not necessarily separable. Recall that we denote the collection of separable subspaces of X by $\mathcal{S}(X)$.

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

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

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

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

$$F(x) \cap B(\bar{y}, \varepsilon) + tr\, B_Y \subset F(B(x, t)). \tag{5.1}$$

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

$$y + tr(B_Y \cap M) \subset \operatorname{cl}(F(x + t(1+\delta)(B_X \cap L_{n+1}))) \tag{5.2}$$

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

Assume that we have already L_n for some n. Let (x_i, y_i) be a dense countable subset of the intersection of $(\operatorname{Graph} F) \cap (L_n \times M)$ with the neighborhood of $(\bar{x}, \bar{y})$ in which (5.1) is guaranteed, let (v_j) be a dense countable subset of $B_Y \cap M$ and let (t_k) be a dense countable subset of $(0, \varepsilon)$. For any $i, j, k = 1, 2, \ldots$ we find, from (5.1), an $h_{ijk} \in B_X$ such that $y_i + rt_k v_j \in F(x_i + t_k h_{ijk})$, 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 $(x, y) \in (\operatorname{Graph} F) \cap (L_n \times M)$, $t \in (0, 1)$, $v \in B_Y$ and (x_{i_m}, y_{i_m}), t_{k_m}, v_{j_m} converge, respectively, to (x, y), t and v, then, as $x_{i_m} + t_{k_m}(B_X \cap M_n) \subset x + t(1+\delta)(B_X \cap M_n)$ for sufficiently large m, we conclude that (5.2) holds with $\hat{L}_n$ instead of L_{n+1}. □

Theorem 5.11 (Separable reduction of regularity [59]). *Let X and Y be Banach spaces. A set-valued mapping* $F : X \rightrightarrows Y$ *with closed graph is regular at* $(\bar{x}, \bar{y}) \in \operatorname{Graph} F$ *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* $L \in \mathcal{S}(X)$ *such that the mapping* $F_{L\times M} : L \rightrightarrows M$ *whose graph is the intersection of* $\operatorname{Graph} F$ *with* $L \times M$ *is regular at* $(\bar{x}, \bar{y})$. *Moreover, if* $\operatorname{sur} F(\bar{x}\,|\,\bar{y}) > r$, *we can choose* $L \in \mathcal{S}(X)$ *and* $M \in \mathcal{S}(Y)$ *containing, respectively,* $\bar{x}$ *and* $\bar{y}$ *to make sure that also* $\operatorname{sur} F_{L\times M}(\bar{x}\,|\,\bar{y}) \geq r$. *Conversely, 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* $\operatorname{sur} F_{L\times M}(\bar{x}\,|\,\bar{y}) \geq r$, *then F is regular at* $(\bar{x}, \bar{y})$ *with* $\operatorname{sur} F(\bar{x}\,|\,\bar{y}) \geq r$.

Proof. So assume that F is regular at $(\bar{x}, \bar{y})$ with $\operatorname{sur} F(\bar{x}\,|\,\bar{y}) > r$. Then, given L_0 and M, we can find a closed separable subspace $L \subset X$ containing L_0 such that (5.2) holds for any $\delta > 0$, any $(x, y) \in (\operatorname{Graph} F) \cap (L \times M)$ sufficiently close to $(\bar{x}, \bar{y})$ and any sufficiently small $t > 0$.

https://doi.org/10.1017/S1446788715000701 Published online by Cambridge University Press

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 $\mathrm{sur}\, F_{L\times M}(\overline{x} \mid \overline{y}) \geq (1+\delta)^{-1} r$. As δ 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 $(x_n, y_n) \in \mathrm{Graph}\, F$ 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 $F_{L\times M}$ can be regular at $(\bar{x}, \bar{y})$ (with the modulus of surjection $\geq r$), which is contrary to the assumption. □

5.3. Contingent derivatives and primal regularity estimates. The following simple proposition establishes a connection between the 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 $\|h\| = 1$.

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

as claimed. □

The following result is now immediate from the proposition and Theorem 3.2.

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

$$\mathrm{sur}\, F(\overline{x} \mid \overline{y}) \geq r.$$

(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 $(x, y) \in U \cap \mathrm{Graph}\, F$,*

$$\mathrm{ex}(S_Y, DF(x, y)(cB_X)) \leq \lambda, \tag{5.3}$$

then

$$\mathrm{sur}\, F(\overline{x} \mid \overline{y}) \geq \frac{1-\lambda}{c}. \tag{5.4}$$

Proof. Take an $(x, v) \in U \cap \mathrm{Graph}\, F$ with $v \neq y$ and set $z = \|y - v\|^{-1}(y - v)$. By the assumption, for any $\lambda' > \lambda$, there is a pair $(\tilde{h}, \tilde{w})$ with $\tilde{w} \in DF(x, v)(\tilde{h})$ such that $\|\tilde{h}\| = c$ and $\|z - \tilde{w}\| \leq \lambda'$. As $(\tilde{h}, \tilde{w})$ belongs to the contingent cone to the Graph F at (x, v), we

https://doi.org/10.1017/S1446788715000701 Published online by Cambridge University Press

can find (for sufficiently small $t > 0$) vectors $h(t)$ and $w(t)$ norm converging to $\tilde{h}$ and $\tilde{w}$, respectively, and such that $v + tw(t) \in F(x + th(t))$. We have

$$\begin{aligned}\|y - (v + tw(t))\| &= \|y - v - t\tilde{w}\| + o(t)\\ &\le \|y - v - tz\| + t\|z - \tilde{w}\| + o(t)\\ &\le \|y - v\|\Big(1 - \frac{t}{\|y - v\|}\Big) + t\lambda' + o(t),\end{aligned}$$

so that using the same φ that was defined at the beginning of Section 3, we can write

$$\varphi_y^-((x, v); (\tilde{h}, \tilde{w})) \le \lim_{t\to+0} \frac{\|y - t(v + w(t))\| - \|y - v\|}{t} \le -(1 - \lambda').$$

Take a $\xi > 0$ such that $\xi(1 + \lambda) < c$ and consider the ξ-norm in $X \times Y$, Then $\|(\tilde{h}, \tilde{w})\|_\xi \le \max\{c, \xi(1 + \lambda')\} = c$ (if λ' is sufficiently close to λ) and, from (5.5),

$$\inf\{\varphi_y^-((x, v); (h, w)) : \ \|(h, w)\|_\xi \le 1\} \le \frac{1}{c}\varphi_y^-((x, v); (\tilde{h}, \tilde{w})) \le -\frac{1 - \lambda'}{c}.$$

It remains to refer to Proposition 5.12 and Theorem 3.10. □

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

$$\operatorname{sur} F(\bar{x}\,|\,\bar{y}) \ge \lim_{\varepsilon\to 0} \inf\{C(DF(x, y)) : (x, y) \in (\operatorname{Graph} F) \cap B((\bar{x}, \bar{y}), \varepsilon)\}, \tag{5.5}$$

or, equivalently,

$$\begin{aligned}&\operatorname{reg} F(\bar{x}\,|\,\bar{y})\\ &\quad\le \lim_{\varepsilon\to 0} \sup\{\|(DF(x, y))^{-1}\|_- : y \in F(x), \|x - \bar{x}\| < \varepsilon, \|y - \bar{y}\| < \varepsilon\}\\ &\quad= \lim_{\varepsilon\to 0}\{\sup_{\|v\|=1} \inf\{\|h\| : v \in DF(x, y)(h)\} : (x, y) \in (\operatorname{Graph} F) \cap B((\bar{x}, \bar{y}), \varepsilon)\}.\end{aligned}$$

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 $(x, v) \in \operatorname{Graph} F$ close to $(\bar{x}, \bar{y})$, this, in turn, means that the condition of Theorem 5.14 is satisfied with $c = r^{-1}$ and $\lambda = 1$, and hence this proves the theorem. □

Remark 5.16. In fact the last two theorems are equivalent. Indeed, let the conditions of Theorem 5.14 be satisfied. Then $(1 - \lambda)B_Y \subset DF(x, v)(cB_X)$ for all $(x, v) \in \operatorname{Graph} F$ 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 space may be much lower than for a real-valued function. The following example illustrates the phenomenon.

https://doi.org/10.1017/S1446788715000701 Published online by Cambridge University Press

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

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

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 ℓ_2 into ℓ_2 by

$$F(x) = x - \eta(\|x\|).$$

It is an easy matter to see that $x \mapsto \eta(\|x\|)$ is $(\sqrt{5}/4)$-Lipschitz, and 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 at what we get by 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 there is no sequence (t_k) converging to zero for which $t_k^{-1} F(t_k h)$ converges. Hence the tangent cone to the graph of F at zero consists of a single point $(0, 0)$ and the first statement gives $\text{sur}\, F(0) \geq 0$, which is a trivial conclusion.

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

$$\begin{aligned} \|F(x + th) - y\| &= \|x + th - \eta(\|x + th\|) - y\| \\ &\leq \|x + th - \eta(\|x\|) - y\| + \|\eta(\|x + th\|) - \eta(\|x\|)\| \\ &\leq \|F(x) + th - y\| + (3/4)t\|h\|. \end{aligned}$$

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

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

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 [4] (see also [5]) under the same assumptions. The very estimate (5.4) was obtained in [51]. Theorem 5.15 was proved by Dontchev *et al.* in [34]. Theorem 5.13 seems to have been stated for the first time in [19]. 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 much attention in the 1980s and 1990s. The role of coderivatives was in the center of the studies. Further developments, however, that followed the discovery of the role of slope, open up the potential for stronger (and often easier to apply) results involving subdifferentials of the functions φ_y and ψ_y.

https://doi.org/10.1017/S1446788715000701 Published online by Cambridge University Press

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 lower semicontinuous, and let an open set U have nonempty intersection with* dom f. *Then, for any subdifferential ∂,*

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

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

Combining this with Theorems 3.10 and 3.12, we get the following theorem.

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 ∂ be a subdifferential trusted on a class of Banach spaces containing both X and Y. Then, for any $(\bar{x},\bar{y}) \in$ Graph F and any $\xi > 0$,*

$$\operatorname{sur} F(\overline{x}\,|\,\overline{y}) \ge \liminf_{\substack{(x,v)\underset{\mathrm{Graph} F}{\to}(\bar{x},\bar{y})\\ y\to\bar{y},\,y\ne v}} \inf\{\|x^*\| + \xi^{-1}\|v^*\| : (x^*,y^*) \in \partial\varphi_y(x,v)\} \tag{5.6}$$

and

$$\operatorname{sur} F(\overline{x}\,|\,\overline{y}) \ge \liminf_{\substack{(x,y)\to(\bar{x},\bar{y})\\ y\notin F(x)}} d(0,\partial\overline{\psi}_y(x)). \tag{5.7}$$

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

$$\operatorname{sur} F(\overline{x}\,|\,\overline{y}) \ge r. \tag{5.8}$$

The obvious inequality $\|x^*\| \ge -f^-(x;h)$ 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 $F : X \rightrightarrows Y$ be a set-valued mapping with locally closed graph containing $(\bar{x},\bar{y})$. Then*

$$\begin{aligned}
&\operatorname{sur} F(\overline{x}\,|\,\overline{y})\\
&\quad\ge \lim_{\varepsilon\to 0}\inf\{C^*(D^*_H F(x,y)) : y \in F(x), \|x-\overline{x}\| < \varepsilon, \|y-\overline{y}\| < \varepsilon\}\\
&\quad= \lim_{\varepsilon\to 0}\inf\{\|x^*\| : x^* \in D^*_H F(x,y)(y^*),\ \|y^*\| = 1, (x,y) \in (\mathrm{Graph}\, F)\cap B((\bar{x},\bar{y}),\varepsilon)\},
\end{aligned}$$

or, equivalently,

$$\begin{aligned}
&\operatorname{reg} F(\overline{x}\,|\,\overline{y})\\
&\quad= \operatorname{lip} F^{-1}(\overline{y}\,|\,\overline{x}) \le \lim_{\varepsilon\to 0}\sup\{\|D^*_H F^{-1}(x,y)\|_+ : (x,y) \in (\mathrm{Graph}\, F)\cap B((\bar{x},\bar{y}),\varepsilon)\}\\
&\quad= \lim_{\varepsilon\to 0}\sup\{\|y^*\| : x^* \in D^*_H F(x,y)(y^*),\ \|x^*\| = 1, (x,y) \in (\mathrm{Graph}\, F)\cap B((\bar{x},\bar{y}),\varepsilon)\}.
\end{aligned}$$

https://doi.org/10.1017/S1446788715000701 Published online by Cambridge University Press

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 us to work with weak* neighborhoods to estimate the subdifferential of a sum of functions (that inevitably appears in the course of the calculation) which makes estimating norms of subgradients difficult (if possible at all).

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

$$\inf\{\|x^*\| : x^* \in D^*F(u, v)(y^*),\ (u, v) \in \operatorname{Graph} F \cap (U \times V),\ \|y^*\| = 1\} \le m$$

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

Let $|\nabla_\xi \varphi_y|(x, w) < m$. Take an $m' < m$ but still greater than $|\nabla_\xi \varphi_y|(x, v)$ and set

$$\begin{aligned} f(u, v) &= \varphi_y(u, v) + m' \max\{\|u - x\|, \xi\|v - w\|\} \\ &= \|v - y\| + i_{\operatorname{Graph} F}(u, v) + m' \max\{\|u - x\|, \xi\|v - w\|\}. \end{aligned}$$

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 $(u_1, v_1) \in \operatorname{Graph} F$ and $v_0^* \in \partial\|\cdot\|(y - v_0)$, $(u_1^*, v_1^*) \in N(\operatorname{Graph} F, (u_1, v_1))$ and (u_2^*, v_2^*) with $\|u_2^*\| + \xi^{-1}\|v_2^*\| \le m'$ 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 ξ so small that $\xi m' < \delta$. Then $y \ne v_0$, so that $\|v_0^*\| = 1$, $\|x_2^*\| \le m'$ and $\|v_2^*\| < \delta$. We thus have $\|x_1^*\| \le m' + \delta < m$ 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*

$$\begin{aligned} &\operatorname{sur} F(\overline{x} \mid \overline{y}) \\ &\quad = \lim_{\varepsilon \to 0} \inf\{C^*(D_F^* F(x, y)) : y \in F(x), \|x - \overline{x}\| < \varepsilon, \|y - \overline{y}\| < \varepsilon\} \\ &\quad = \lim_{\varepsilon \to 0} \inf\{\|x^*\| : x^* \in D_F^* F(x, y)(y^*),\ \|y^*\| = 1, (x, y) \in (\operatorname{Graph} F) \cap B((\bar{x}, \bar{y}), \varepsilon)\}, \end{aligned}$$

or, equivalently,

$$\begin{aligned} &\operatorname{reg} F(\overline{x} \mid \overline{y}) = \operatorname{lip} F^{-1}(\overline{y} \mid \overline{x}) \\ &\quad = \lim_{\varepsilon \to 0} \sup\{\|D_F^* F^{-1}(x, y)\|_+ : (x, y) \in (\operatorname{Graph} F) \cap B((\bar{x}, \bar{y}), \varepsilon)\} \\ &\quad = \lim_{\varepsilon \to 0} \sup\{\|y^*\| : x^* \in D_F^* F(x, y)(y^*),\ \|x^*\| = 1, (x, y) \in (\operatorname{Graph} F) \cap B((\bar{x}, \bar{y}), \varepsilon)\}. \end{aligned}$$

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 $(x, y) \in \operatorname{Graph} F$ close to $(\bar{x}, \bar{y})$ and let

$$m = \inf\{\|x^*\| : x^* \in D_F^* F(x, y)(y^*),\ \|y^*\| = 1\}.$$

https://doi.org/10.1017/S1446788715000701 Published online by Cambridge University Press

If $\operatorname{sur} D_F^*F(\bar{x}\,|\,\bar{y}) = 0$ 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 $r < \operatorname{sur} F(\bar{x}\,|\,\bar{y})$, in which case we may assume that $B(y, rt) \subset F(B(x,t))$ for small t and y close to $\bar{y}$, and suppose that $m < \infty$. Take a $x^* \in D_F^*(x,y)(y^*)$ with $\|y^*\| = 1$ and $\|x^*\| < m + \delta$ for some $\delta > 0$. Then $\langle x^*, h\rangle - \langle y^*, v\rangle \le o(\|h\| + \|v\|)$ whenever $(x + h, y + v) \in \operatorname{Graph} F$. Now take $v(t) \in B(y, rt)$ such that $\langle y^*, v(t)\rangle \le -(1 - t^2)\|v(t)\|$ and an $h(t)$ with $\|h(t)\| \le t$ such that $(x + th(t), y + v(t)) \in \operatorname{Graph} F$. Then

$$-t\|x^*\| + (1 - t^2)rt \le \langle x^*, h(t)\rangle - \langle y^*, v(t)\rangle \le o(\|h(t)\| + \|v(t)\|) = o(t),$$

which implies that $r \le m$ and the result follows. □

Remark 5.23. Note that the just given proof (that the inequality $\le$ 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 versus tangential criterion). *If X and Y are Gâteaux smooth spaces, then the regularity estimate involving Dini–Hadamard coderivative (Theorem 5.21) is never worse than the 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 $C^*(D_H^*F(x,y)) \ge C(DF(x,y))$ for any $(x,y) \in \operatorname{Graph} F$, by Theorem 5.4. □

Theorem 5.21 was proved in [51] for subdifferentials satisfying slightly stronger requirements than the subdifferential of Dini–Hadamard. However, a minor change in the proof allows us to extend it to all subdifferentials trusted on the given Banach space (see, for example, [54, 57] for a proof), and, in particular, to the DH-subdifferential on any Gâteaux smooth space. Likewise, Theorem 5.22 was proved in [68], in a somewhat different form and in terms of ε-Fréchet subdifferentials on Fréchet smooth spaces. Again, a minor change is needed to extend the proof to standard Fréchet subdifferentials. Theorem 5.22, as stated, was proved in [79] (see also [77]). 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 [39] (and actually was proved that way). Proposition 5.24 does not seem to have ever 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 at 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, whether 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 say that F is *perfectly regular* at $(\bar{x}, \bar{y}) \in \mathrm{Graph}\, F$ if

$$\mathrm{sur}\, F(\overline{x} \mid \overline{y}) = C^*(D^*_G F(\bar{x}, \bar{y})) = \min\{\|x^*\| : x^* \in D^*_G F(\bar{x}, \bar{y})(y^*), \|y^*\| = 1\}. \tag{5.9}$$

Later, we shall come across some classes of perfectly regular mappings but, 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 $(\bar{x}, \bar{y}) \in \mathrm{Graph}\, F$. 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 $\mathrm{sur}\,(F + A)(\overline{x}, \overline{y} + A\overline{x}) = 0$. We shall denote it by $\mathrm{rad}\, F(\overline{x} \mid \overline{y})$.

By Milyutin's theorem, $\mathrm{sur}\, F(\overline{x} \mid \overline{y}) \le \mathrm{rad}\, F(\overline{x} \mid \overline{y})$. It turns out that, for perfectly regular mappings, the equality holds. To show this we need the following proposition,which is not very difficult to prove.

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

$$D^*_G G(\overline{x} \mid \overline{y} + A\overline{x}) = D^*_G F(\bar{x}, \bar{y}) + A^*.$$

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

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

$$\mathrm{sur}\, F(\overline{x} \mid \overline{y}) = \mathrm{rad}\, F(\overline{x} \mid \overline{y}). \tag{5.10}$$

Moreover, for any $\varepsilon > 0$, there is a linear operator A_ε of rank one such that $\|A_\varepsilon\| \le$ $\mathrm{sur}\, F(\overline{x} \mid \overline{y}) + \varepsilon$ *and* $\mathrm{sur}\,(F + A)(\overline{x}, \overline{y} + A\overline{x})) = 0$.

In the sequel we call (5.10) the *radius formula*.

Proof. Set $r = \mathrm{sur}\, F(\overline{x} \mid \overline{y})$. The theorem is obviously valid if $r = 0$. So we assume that $r > 0$. Take an $\varepsilon > 0$ and find a y^*_ε and an $x^*_\varepsilon \in D^*_G F(\bar{x}, \bar{y})(y^*_\varepsilon)$ such that $\|y^*_\varepsilon\| = 1$, $\|x^*_\varepsilon\| \le (1 + \varepsilon)r$. Further, let $x_\varepsilon \in X$ and $y_\varepsilon \in Y$ satisfy

$$\|x_\varepsilon\| = \|y_\varepsilon\| = 1, \quad \langle x^*_\varepsilon, x_\varepsilon\rangle \ge (1 - \varepsilon)\|x^*_\varepsilon\|, \quad \langle y^*_\varepsilon, y_\varepsilon\rangle \ge (1 - \varepsilon).$$

https://doi.org/10.1017/S1446788715000701 Published online by Cambridge University Press

We use these four vectors to define an operator $A_\varepsilon : X \to Y$ as

$$A_\varepsilon x = -\frac{\langle x_\varepsilon^*, x\rangle}{\langle y_\varepsilon^*, y_\varepsilon\rangle} y_\varepsilon.$$

Then $\|A_\varepsilon\| \le \dfrac{1+\varepsilon}{1-\varepsilon} r$ and

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

In particular, we see that $-x_\varepsilon^* = A_\varepsilon^* y_\varepsilon^*$. Combining this with Proposition 5.27 we get $0 = x_\varepsilon^* - A_\varepsilon^* y_\varepsilon^* \in D_G^*(F+A)(\bar{x}, \bar{y} + A\bar{x})(y_\varepsilon^*)$ and therefore, by the perfect regularity assumption, $\operatorname{sur}(F+A)(\bar{x} \mid \bar{y} + A\bar{x}) = 0$: that is, $\operatorname{rad} F(\bar{x}, \bar{y}) \le \|A_\varepsilon\| \to r$ as $\varepsilon \to 0$. □

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

$$y \in F(x) + Ax, \tag{5.11}$$

where $A \in \mathcal{L}(X, Y)$. Let $\bar{x}$ be a nominal solution of (5.11) with $y = \bar{y}$, $A = \bar{A}$. 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 $(\bar{y}, \bar{A})$ and their effect on regularity rates.

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

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

We shall consider $Y \times \mathcal{L}(X, Y)$ with the norm $\|(y, A)\| = \nu(\|y\|, \|A\|)$, where ν is a norm in $\mathbb{R}^2$. The dual norm is $\nu^*(\|y^*\|, \|\ell\|)$, where $\ell \in (\mathcal{L}(X \times Y))^*$ and ν^* is the norm in $\mathbb{R}^2$ dual to ν: $\nu^*(u) = \sup\{\alpha\xi + \beta\eta : \nu(\alpha, \beta) \le 1\}$. As to the space dual to $\mathcal{L}(X, Y)$, we only need the simplest elements of the space, that is, 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 [58]. *Let X and Y be Banach spaces, and let $F : X \rightrightarrows Y$ be a set-valued mapping with closed graph. Let $(\bar{x}, \bar{y}) \in \operatorname{Graph} F$ and let $\bar{A} \in \mathcal{L}(X, Y)$ be given. Then*

$$\operatorname{lip} S((\bar{y}, \bar{A}) \mid \bar{x}) \le \nu^*(1, \|\bar{x}\|) \operatorname{reg}(F + \bar{A})(\bar{x} \mid \bar{y}).$$

To prove the theorem, we only need to show that

$$\operatorname{sur} \Phi(\bar{x} | (\bar{y}, \bar{A})) \ge \frac{1}{\nu^*(1, \|\bar{x}\|)} \operatorname{sur}(F + \bar{A})(\bar{x} \mid \bar{y}). \tag{5.12}$$

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 [63] and [33]. 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 [18]. We shall discuss the problems considered in this subsection in more details for finite-dimensional mappings later in [61, Section 8].

https://doi.org/10.1017/S1446788715000701 Published online by Cambridge University Press

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 that are 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 $\hat{\partial}$ rather than ∂_H or ∂_F. Likewise, as the limiting Fréchet and the G-subdifferentials are also equal, it is convenient to speak simply about a *limiting subdifferential* and denote it simply by ∂.

The second circumstance to be mentioned is the abundance of some special classes of objects of practical importance and definite theoretical interest. It is enough to mention polyhedral and semialgebraic sets and mappings (to be considered in the second part of the paper), semismooth functions, prox-regular functions and sets and so on. We do not discuss some interesting and important subjects, for example Kummer's inverse function theorem and its applications (well presented in the literature: much on the subject can be found in [35, 65]) or semismooth mappings (see, for example, [42]).

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 $A : \mathbb{R}^n \to \mathbb{R}^m$ of rank one.*

PROOF. This is immediate from Theorem 5.28. □

THEOREM 6.3. *Let $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ be a set-valued mapping with locally closed graph, and let $(\bar{x}, \bar{y}) \in \operatorname{Graph} F$. Then*

$$\operatorname{sur} F(\overline{x} \mid \overline{y}) = \lim_{\varepsilon \to 0} \inf\{C(DF(x, y)) : (x, y) \in (\operatorname{Graph} F) \cap B((\bar{x}, \bar{y}), \varepsilon)\}. \tag{6.1}$$

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 $(x, y) \in \operatorname{Graph} F$). 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 $rB_{\mathbb{R}^m} \subset DF(x, y)(B_{\mathbb{R}^n})$. □

https://doi.org/10.1017/S1446788715000701 Published online by Cambridge University Press

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

$$y \in F(x) + Ax, \tag{6.2}$$

with both y and A viewed as perturbations (compare with Theorem 5.29). But first we have to do some preliminary work. As in Section 5.4.2, we denote by $S(y, A)$ the set of solutions of (6.2) and by Φ the inverse mapping

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

LEMMA 6.4. *For any $x \in X$, let $E(x) : Y \times \mathcal{L}(X, Y) \to Y$ be the linear operator defined by $E(y, \Lambda) = y - \Lambda x$. Then, under the assumptions of Theorem 5.29,*

$$\nu(1, \|x\|)C(E(x) \circ D\Phi(x, (y, A)) \le C(D(F + A)(x, y))$$

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

PROOF. By definition, $(h, v, \Lambda) \in X \times Y \times \mathcal{L}(X, Y)$ belongs to $T(\operatorname{Graph}\Phi, (x, y, A))$ 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

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

which amounts to

$$E(x) \circ D\Phi(x, (y, A)) = D(F + A)(x, y).$$

We have (Corollary 5.3) $C(E(x)) \cdot C(D\Phi(x, (y, A))) \le C(D(F + A)(x, y))$. 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 $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ 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 ν is a certain norm in $\mathbb{R}^2$. Then, given an $A \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$,*

$$\operatorname{lip} S((\overline{y}, \overline{A}) \mid \overline{x}) = \nu^*(1, \|\overline{x}\|)\operatorname{reg}(F + \overline{A})(\overline{x} \mid \overline{y}).$$

PROOF. This is 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, for example, [65]).

Theorem 6.1 was announced by Mordukhovich in a somewhat different form [75] (see also [76]). But the lower estimate for the modulus of surjection (which is actually the major step in the proof) is immediate from Ioffe [50]. Theorem 6.2 was proved by Dontchev *et al.* in [33] and Theorem 6.3 by Dontchev *et al.* [34]. Theorem 6.5 is a slightly generalized version of the already mentioned result of Cánovas *et al.* [18].

6.2. Subregularity and error bounds. Let f be an extended-real-valued lower semicontinuous function on $\mathbb{R}^n$. We can associate with this function the epigraphic map

$$\operatorname{Epi} f(x) = \{\alpha \in R : \alpha \geq f(x)\}.$$

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

$$\partial^{>} f(x) = \{\lim_{k\to\infty} x_k^* : \exists x_k \underset{f}{\to} x,\ f(x_k) > f(x),\ x_k^* \in \hat{\partial} f(x_k)\}.$$

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

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

Thus, if $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ has locally closed graph and $(\bar{x}, \bar{y}) \in \operatorname{Graph} F$, then

$$\operatorname{subreg} F(\overline{x} \,|\, \overline{y}) \leq [\inf\{\|x^*\| : x^* \in \partial^{>} d(\overline{y}, F(\cdot))(\overline{x})]^{-1}\}.$$

Proof. If (a) holds, then K is a local error bound by Lemma 7.1 to be proved in the next section in [61]. To prove that (a) $\Rightarrow$ (b), let $x^* \in \partial^{>} f(\overline{x})$. This means that there are sequences (x_k) and (x_k^*) such that $x_k \to_f \overline{x}$, $f(x_k) > f(\overline{x})$, $x_k^* \to x^*$ and $x_k^* \in \partial f(x_k)$. Choose $\varepsilon_k \downarrow 0$ such that $\|x_k - \overline{x}\| < \varepsilon_k$ and $f(x_k) - f(\overline{x}) < K\varepsilon_k$. If (a) holds, then $K \cdot \liminf |\nabla f|(x_k) \geq 1$. 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 on the left-hand 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 $f + (r + \varepsilon)\| \cdot - x\|$ attains a local minimum at x. Applying the fuzzy variational principle, we shall find u and $u^* \in \partial_F(u)$ such that $\|u - x\| < \varepsilon$, $f(u) < f(\overline{x}) + \varepsilon/K$ and $\|u^*\| < r + 2\varepsilon$. This means that there is a sequence of pairs (x_k, x_k^*) such that $x_k \to_f \overline{x}$, $x_k^* \in \partial_F f(x_k)$ 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 $\overline{x}$.

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, $0 \in \partial^{>} f(0)$.

https://doi.org/10.1017/S1446788715000701 Published online by Cambridge University Press

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

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

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

for all x, u of a neighborhood of $\overline{x}$, provided $f(\overline{x}) < f(x) < f(\overline{x}) + \theta$ and $x^* \in \hat{\partial}(x)$. If, under these conditions, $K > 0$ is an error bound of f at $\overline{x}$, 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 $x_k \to_f \overline{x}$, $f(x_k) > f(\overline{x})$ and $\|x_k^*\| \leq K^{-1} - \varepsilon$. For any k, take an $\overline{x}_k \in [f \leq f(\overline{x})]$ closest to x_k. Then $\overline{x}_k \to f(\overline{x})$ and, by the assumption,

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

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

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

It follows that

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

that is, $(K^{-1} - (\varepsilon/2))\, d(x_k, [f \leq f(\overline{x})]) \geq f(x_k)$, which is contrary to the assumption. □

The last result of this subsection contains infinitesimal characterization of strong subregularity.

THEOREM 6.9 (Characterization of subregularity and strong subregularity). *Again, let $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ have a locally closed graph and $(\bar{x}, \bar{y}) \in \operatorname{Graph} F$. Then:*

- *F is subregular at $(\bar{x}, \bar{y}) \in \operatorname{Graph} F$ if $d(0, \partial^{>}\psi_{\overline{y}}(\overline{x})) > 0$; and*
- *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, firs,t that F is strongly subregular at $(\bar{x}, \bar{y})$: that is, that 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 $\overline{y} + t_k v_k \in F(\overline{x} + t_k h_k)$, so that, for large k,

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

which is 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, for example, the monograph by Facchinei and Pang [40] 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 [62] and Meng–Yang [74], among others. The second part of Theorem 6.9, as well as other results relating to strong subregularity and applications, can be found in [35] and [65]. (In [65], the authors use the term 'locally upper Lipschitz' property. The term 'strong subregularity' seems to have appeared later.) Another sufficient condition for subregularity was suggested by Gfrerer [44]. 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 $\overline{x}$. As was mentioned in Section 1.4, the manifolds are transversal at $\overline{x}$ if either $\overline{x} \notin M_1 \cap M_2$ or the sum of the tangent subspaces to the manifolds at $\overline{x}$ is the whole of X: $T_{\overline{x}}M_1 + T_{\overline{x}}M_2 = X$. The following simple lemma is the key to interpreting this, in regularity terms, in a way suitable for extension to the setting 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 $v + h \in L_2$.*

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. As an immediate corollary, that the set-valued mapping $\Phi(x) = (L_{1-x}) \times (L_{2-x})$ from X into $X \times X$ is regular at zero. This justifies the following definition.

Definition 6.11. Let $S_i \subset X$, $i = 1, \dots, k$ be closed subsets of X. We say that S_i are *transversal* at $\overline{x} \in X$ if either $\overline{x} \notin \cap S_i$ or $\overline{x} \in \cap S_i$ and the set-valued mapping

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

from X into X^k is regular near $(\overline{x}, 0, \dots, 0)$. In the latter case, we also say that the S_i *have transversal intersection* at $\overline{x}$.

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

https://doi.org/10.1017/S1446788715000701 Published online by Cambridge University Press

Theorem 6.12. *Let $S_i \subset \mathbb{R}^n$, $i = 1, \dots, k$ and $\overline{x} \in \cap S_i$. Then the following statements are equivalent:*

(a) *the S_i are transversal at $\overline{x}$;*
(b) $x_i^* \in N(S_i, \overline{x}), x_1^* + \cdots + x_k^* = 0 \Rightarrow x_1^* = \cdots = x_k^* = 0$*; and*
(c) $d(x, \bigcap_{i=1}^k (S_i - x_i) \le K \max_i d(x, S_i - x_i)$ *if x_i are close to zero and x is close to $\overline{x}$.*

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

$$D^* F(x \mid (x_1, \dots, x_k)) = \begin{cases} \sum_{i=1}^k x_i^* \text{ if } x_i^* \in N(S_i, x_i + x), \\ \emptyset \qquad \text{otherwise.} \end{cases}$$

Combining this with Theorem 6.1, we prove equivalence (a) and (b).

Furthermore, $F^{-1}(x_1, \dots, x_k) = (S_1 - x_1) \cap \cdots \cap (S_k - x_k)$, from which comes 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 the case of two sets, one more convenient characterization of transversality is available.

Corollary 6.13. *Two sets S_1 and S_2, which both contain $\overline{x}$, are transversal at $\overline{x}$ if and only if the set-valued mapping $\Phi : \mathbb{R}^n \times \mathbb{R}^n \rightrightarrows \mathbb{R}^n$, given by*

$$\Phi(x_1, x_2) = \begin{cases} x_1 - x_2 \text{ if } x_i \in S_i, \\ \emptyset \qquad \text{otherwise} \end{cases}$$

is regular near $(\overline{x}, \overline{x}, 0)$.

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

$$D^*\Phi((\overline{x}, \overline{x}), 0)(x^*) = \{(x_1^*, x_2^*) : x_i^* \in N(S_i, \overline{x}) + x^*\}.$$

If we consider the max-norm $\|(x_1, x_2)\| = \max\{\|x_1\|, \|x_2\|\}$ in $\mathbb{R}^n \times \mathbb{R}^n$, then it follows, from Theorem 6.1, that Φ is regular near $(\overline{x}, \overline{x}, 0)$ if and only if

$$\inf\{\|x_1^* - x^*\| + \|x_2^* + x^*\| : x_i^* \in N(S_i, x_n),\ \|x^*\| = 1\} > 0.$$

This amounts to $N(S_1, \overline{x}) \cap (-N(S_2, \overline{x})) = \{0\}$, which is exactly the property in part (b) of the theorem. □

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 $S_1, \dots, S_k$ are *subtransversal at* $\overline{x} \in \bigcap S_i$ if there is a $K > 0$ such that, for any x close to $\overline{x}$,

$$d\Big(x, \bigcap_{i=1}^k S_i\Big) \le K \sum_{i=1}^k d(x, S_i).$$

https://doi.org/10.1017/S1446788715000701 Published online by Cambridge University Press

In a similar way, it is easy to see that subtransversality is equivalent to subregularity of the same mapping Φ which allows 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 $\overline{y} \in F(\overline{x}) \cap S$. 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 at* $(\bar{x}, \bar{y})$*; and*
(c) $0 \in D^*F(\bar{x}, \bar{y})(y^*)$ & $y^* \in N(S, \bar{y}) \Rightarrow y^* = 0.$

The theorem justifies the following definition.

DEFINITION 6.16. Let $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ 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 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

$$d((x, y), \text{Graph}\, F \cap (X \times S)) \le K d((x, y), \text{Graph}\, F) + d(y, S))$$

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

It is almost obvious from (a) that, in the 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 $\overline{y} = F(\overline{x}) \in Q$. Finally, let*

$$\Phi(x) = \begin{cases} F(x) - Q \text{ if } x \in C, \\ \emptyset \qquad\qquad \text{otherwise,} \end{cases} \qquad F_C(x) = \begin{cases} F(x) \text{ if } x \in C, \\ \emptyset \qquad \text{otherwise.} \end{cases}$$

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

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

https://doi.org/10.1017/S1446788715000701 Published online by Cambridge University Press

(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 $\bar{x}$ is equivalent to regularity of $F_C - Q$ near $(\bar{x}, 0)$. We note, also, the following simple corollary of the theorem.

COROLLARY 6.18. *Under the assumption of the theorem*

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

The set-valued mapping in Definition 6.11 was introduced in [54], where it was shown that subtransversality of a collection of sets is equivalent to subregularity of the mapping. Equivalence of (a) and (c) in Theorem 6.12 is immediate from this result. Explicitly the theorem was partly proved in [69] (equivalence of (a) and (c)) and partly in [72] (equivalence of (a) and (b)). We refer to [69] for further equivalent descriptions (some looking very technical) of transversality and related properties. The inequality in the definition of subtransversality first appeared in a paper by Dolecki [29] as a sufficient condition for the equality of the contingent cone to an intersection and the intersection of the contingent cones. For discussions concerning the role of transversality in analysis and optimization, see Section 8. The very term 'subtransversality' was introduced in [36]. 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, which can be extracted from [77, Theorem 5.23].

References

[1] P. Abraham and J. Robin, *Transversal Mappings and Flows* (Benjamin Press, New York, 1967).

[2] A. Arutyunov, E. Avakov, B. Gelman, A. Dmitruk and V. Obukhovski, 'Locally covering maps in metric spaces and coincidence points', *J. Fixed Point Theory Appl.* **5** (2009), 106–127.

[3] J.-P. Aubin, 'Contingent derivatives of set-valued maps and existence of solutions to nonlinear inclusions and differential inclusions', Advances in Mathematics. Supplementary Studies (ed. L. Nachbin) (1981), 160–232.

[4] J.-P. Aubin, 'Lipshitz behavior of solutions to convex optimization problems', *Math. Oper. Res.* **9** (1984), 87–111.

[5] J. P. Aubin and H. Frankowska, *Set-Valued Analysis* (Birkhäuser, Boston, 1990).

[6] D. Azé, 'A unified theory for metric regularity of multifunctions', *J. Convex Anal.* **13** (2006), 225–252.

[7] D. Azé and S. Benahmed, 'On implicit multifunction theorems', *Set-Valued Anal.* **16** (2008), 129–155.

[8] D. Azé and J.-N. Corvellec, 'On some variational properties of metric spaces', *J. Fixed Point Theory Appl.* **5** (2009), 185–200.

[9] D. Azé, J.-N. Corvellec and R. Lucchetti, 'Variational pairs and application to stability in nonsmooth analysis', *Nonlinear Anal.* **49** (2002), 643–670.

[10] D. Azé, C. C. Chou and J. P. Penot, 'Subtraction theorem and approximate openness of multifunctions', *J. Math. Anal. Appl.* **221** (1998), 33–58.

[11] S. Banach, *Théorie des opérations linéaires*, Monografje Matematyczne (Warszawa, 1932).

[12] M. S. Bazaraa, J. J. Goode and M. Z. Nashed, 'On the cones of tangents with applications to mathematical programming', *J. Optim. Theory Appl.* **13** (1974), 389–426.

[13] J. M. Borwein, 'Adjoint process duality', *Math. Oper. Res* **8** (1983), 403–434.

[14] J. M. Borwein, 'Norm duality for convex processes and applications', *J. Optim. Theory Appl.* **48** (1986), 53–64.

https://doi.org/10.1017/S1446788715000701 Published online by Cambridge University Press

[15] J. M. Borwein and A. S. Lewis, *Convex Analysis and Nonlinear Optimization*, 2nd edn (Springer, 2006).
[16] 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.
[17] J. M. Borwein and J. Zhu, *Techniques of Variational Analysis*, CMS Books in Mathematics, 20 (Springer, 2006).
[18] M. J. Cánovas, F. J. Gómez-Senent and J. Parra, 'Linear regularity, equi-regularity and intersection mappings for convex semi-infinite inequality systems', *Math. Program. B* **123** (2010), 33–60.
[19] R. Cibulka, M. Fabian and A. D. Ioffe, 'On primal regularity estimates for single-valued mappings', *J. Fixed Point Theory Appl.* **17** (2015), 187–208.
[20] F. H. Clarke, 'Necessary conditions for nonsmooth problems in optimal control and the calculus of variations', PhD Dissertation, University of Washington, Seattle, 1973.
[21] F. H. Clarke, 'Generalized gradients and applications', *Trans. Amer. Math. Soc.* **205** (1975), 247–262.
[22] F. H. Clarke, *Optimization and Nonsmooth Analysis* (Wiley-Interscience, New York, 1983).
[23] F. H. Clarke, Y. S. Ledyaev, R. J. Stern and P. R. Wolenski, *Nonsmooth Analysis and Control Theory* (Springer, 1998).
[24] R. Cominetti, 'Metric regularity, tangent sets and second order optimality conditions', *Appl. Math. Optim.* **21** (1990), 265–287.
[25] 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.
[26] 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.
[27] S. Dolecki, 'Semicontinuity in constraint optimization 1', *Control Cybernet.* **7**(2) (1978), 5–15.
[28] S. Dolecki, 'Semicontinuity in constraint optimization 2', *Control Cybernet.* **7**(3) (1978), 17–26.
[29] S. Dolecki, 'Tangency and dierentiation, some applications of convergence theory', *Ann. Math. Pura Appl.* **130** (1982), 223–255.
[30] A. L. Dontchev, 'The Graves theorem revisited', *J. Convex Anal.* **3** (1996), 45–54.
[31] A. L. Dontchev and H. Frankowska, 'Lyusternik–Graves theorem and fixed points', *Proc. Amer. Math. Soc.* **139** (2011), 521–534.
[32] A. L. Dontchev and H. Frankowska, 'Lyusternik–Graves theorem and fixed points 2', *J. Convex Anal.* **19** (2012), 955–974.
[33] A. L. Dontchev, A. S. Lewis and R. T. Rockafellar, 'The radius of metric regularity', *Trans. Amer. Math. Soc.* **355** (2003), 493–517.
[34] A. L. Dontchev, M. Quincampoix and N. Zlateva, 'Aubin criterion for metric regularity', *J. Convex Anal.* **13** (2006), 281–297.
[35] A. L. Dontchev and R. T. Rockafellar, *Implicit Functions and Solution Mappings*, 2nd edn (Springer, 2014).
[36] D. Drusvyatsky, A. D. Ioffe and A. S. Lewis, 'Alternating projections and coupling slope', *Found. Comput. Math.* (2015), to appear, arXiv:1401.7569v1.
[37] M. Durea and R. Strugariu, 'Openness, stability and implicit function theorems', *Nonlinear Anal.* **75** (2012), 1246–1259.
[38] M. Fabian and D. Preiss, 'A generalization of the interior mapping theorem of Clarke and Pourciau', *Comment. Math. Univ. Carolin.* **28** (1987), 311–324.
[39] 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.
[40] F. Facchinei and J. S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems* (Springer, New York, 2003).
[41] H. Frankowska, 'Some inverse mapping theorems', *Ann. Inst. H. Poincaré Analyse Non Linéaire* **7** (1990), 183–234.

[42] P. Fusek, 'On metric regularity of weakly almost piecewise smooth functions and some applications in nonlinear semi-definite programming', *SIAM J. Optim.* **23** (2013), 1041–1061.

[43] B. D. Gelman, 'A generalized theorem about implicit mappings', *Funktsional. Anal. Prilozhen* **35** (2001), 28–35.

[44] H. Gfrerer, 'First and second order characterizations of metric subregularity and calmness of constraint set mappings', *SIAM J. Optim.* **21** (2011), 1439–1474.

[45] L. M. Graves, 'Some mapping theorems', *Duke Math. J.* **17** (1950), 111–114.

[46] A. Grothendieck, 'Sketch of a proposal', in: *Geometric Galois Actions* (eds. L. Schneps and P. Lochak) (Cambridge University Press, Cambridge, 1997).

[47] V. Guillemin and A. Pollack, *Differential Topology* (Prentice-Hall, Englewood Cliffs, New Jersey, 1976).

[48] A. D. Ioffe, 'Sous-differentielles approchées des fonctions numériques', *C.R. Acad. Sci. Paris* **292** (1981), 675–678.

[49] A. D. Ioffe, 'Non-smooth analysis: differential calculus of non-differentiable mappings', *Trans. Amer. Math. Soc.* **255** (1981), 1–55.

[50] A. D. Ioffe, 'Approximate subdifferentials and applications 1. The finite dimensional theory', *Trans. Amer. Math. Soc.* **28** (1984), 389–416.

[51] A. D. Ioffe, 'On the local surjection property', *Nonlinear Anal.* **11** (1987), 565–592.

[52] A. D. Ioffe, 'Approximate subdifferentials and applications 3. The metric theory', *Mathematika* **36** (1989), 1–38.

[53] A. D. Ioffe, 'Variational methods in local and global nonsmooth analysis', in: *Nonlinear Analysis, Differential Equations and Control*, NATO Science Series C: Mathematical and Physical Sciences, 258 (eds. F. H. Clarke and R. J. Stern) (1999), 447–502.

[54] A. D. Ioffe, 'Metric regularity and subdifferential calculus', *Uspekhi Mat. Nauk* **55**(3) (2000), 103–162; (in Russian), English translation: *Russian Math. Surveys* **55** (2000), 501–558.

[55] A. D. Ioffe, 'On regularity estimates for mappings between embedded manifolds', *Control Cybernet.* **36** (2007), 659–668.

[56] A. D. Ioffe, 'Regularity on fixed sets', *SIAM J. Optim.* **21** (2011), 1345–1370.

[57] A. D. Ioffe, 'On the general theory of subdifferentials', *Adv. Nonlinear Anal.* **1** (2012), 47–120.

[58] A. D. Ioffe, 'Convexity and variational analysis', in: *Computational and Analytical Mathematics* (eds. D. Bailey, H. Bauschke, F. Garvan, M. Thera, J. Vanderwerff and H. Wolkovicz) (Springer, 2013), 397–428.

[59] A. D. Ioffe, 'Separable reduction of metric regularity properties', in: *Constructive Nonsmooth Analysis and Related Topics*, Springer Optimization and Applications, 87 (eds. V. F. Demyanov, P. M. Pardalos and M. Batsin) (Springer, 2013).

[60] A. D. Ioffe, 'Metric regularity, fixed points and some associated problems of variational analysis', *J. Fixed Point Theory Appl.* **15** (2014), 67–99.

[61] A.D. Ioffe, 'Metric regularity—a survey, Part 2. Applications', *J. Aust. Math. Soc.*, in press.

[62] A. D. Ioffe and J. V. Outrata, 'On metric and calmness qualification conditions in subdifferential calculus', *Set-Valued Anal.* **16** (2008), 199–227.

[63] A. D. Ioffe and Y. Sekiguchi, 'Regularity estimates for convex multifunctions', *Math. Program. B* **117** (2009), 255–270.

[64] A. D. Ioffe and V. M. Tihomirov, *Theory of Extremal Problems* (Nauka, Moscow, 1974), (in Russian); English translation: North-Holland 1979.

[65] D. Klatte and B. Kummer, *Nonsmooth Equations in Optimization* (Kluwer, Dordrecht, 2002).

[66] A. Y. Kruger, 'Generalized differentials of nonsmooth functions, mimeographed notes', 1332–81 (Belorussian State University, 1981), 64 pages (in Russian).

[67] A. Y. Kruger, 'Generalized differentials of nonsmooth functions and necessary conditions for an extremum', *Sib. Math. J.* **26** (1985), 370–379.

[68] A.Y. Kruger, 'A covering theorem for set-valued mappings', *Optimization* **19** (1988), 763–780.

[69] A. Y. Kruger, 'About regularity of collections of sets', *Set-Valued Anal.* **14** (2006), 187–206.
[70] 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).
[71] Y. S. Ledyaev and Q. J. Zhu, 'Implicit multifunction theorem', *Set-Valued Anal.* **7** (1999), 147–162.
[72] A. S. Lewis, D. R. Luke and J. Malick, 'Local linear convergence for alternating and averaged nonconvex projections', *Found. Comput. Math.* **9** (2009), 485–513.
[73] L. A. Lyusternik, 'On conditional extremums of functionals', *Sb. Math. (3)* **41** (1934), 390–401; (in Russian).
[74] K. W. Meng and X. Q. Yang, 'Equivalent conditions for local error bounds', *Set-Valued Var. Anal.* **20** (2012), 617–636.
[75] B. S. Mordukhovich, *Approximation Methods in Problems of Optimization and Control* (Nauka, Moscow, 1988), (in Russian).
[76] B. S. Mordukhovich, 'Complete characterization of openness, metric regularity and Lipschitzian properties of multifunctions', *Trans. Amer. Math. Soc.* **340** (1993), 1–35.
[77] B. S. Mordukhovich, *Variational Analysis and Generalized Differentiation, vol. 1, 2* (Springer, 2006).
[78] B. S. Mordukhovich and Y. Shao, 'On nonconvex subdifferential calculus in Banach spaces', *J. Convex Anal.* **2** (1995), 211–227.
[79] B. S. Mordukhovich and Y. Shao, 'Nonsmooth sequential analysis in Asplund spaces', *Trans. Amer. Math. Soc.* **348** (1996), 1235–1280.
[80] H. V. Ngai and M. Théra, 'Error bounds and implicit multifunctions in smooth Banach spaces and applications to optimization', *Set-Valued Anal.* **12** (2004), 195–223.
[81] H. V. Ngai, N. H. Tron and M. Théra, 'Implicit multifunction theorem in complete metric spaces', *Math. Program. B* **139** (2013), 301–326.
[82] L. Nirenberg, *Topics in Nonlinear Functional Analysis*, Lecture Notes (Courant Institute of Mathematical Sciences, New York University, 1974).
[83] J.-P. Penot, 'Sous-différentiels de fonctions numériques non convexes', *C. R. Acad. Sci. Paris* **278** (1974), 1553–1555.
[84] J.-P. Penot, 'Metric regularity, openness and Lipschitzean behaviour of multifunctions', *Nonlinear Anal.* **13** (1989), 629–643.
[85] J.-P. Penot, *Calculus Without Derivatives*, Graduate Texts in Mathematics, 266 (Springer, 2012).
[86] V. Ptàk, 'A quantitative refinement of a closed graph theorem', *Czechoslovak Math. J.* **24** (1974), 503–506.
[87] S. M. Robinson, 'Normed convex processes', *Trans. Amer. Math. Soc.* **174** (1972), 127–140.
[88] S. M. Robinson, 'Stability theory for system of inequalities. Part II: Differentiable nonlinear systems', *SIAM J. Numer. Anal.* **13** (1976), 497–513.
[89] S. M. Robinson, 'Strongly regular generalized equations', *Math. Oper. Res.* **5** (1980), 43–62.
[90] R. T. Rockafellar, 'Monotone processes of convex and concave type', *Mem. Amer. Math. Soc.* **77** (1967).
[91] R. T. Rockafellar and R. J. B. Wets, *Variational Analysis* (Springer, 1998).
[92] A. Sard, 'The measure of critical values of differentiable maps', *Bull. Amer. Math. Soc.* **48** (1942), 883–890.
[93] J. Schauder, 'Uber die Umkerhrung linearer, stetiger Funktionaloperationen', *Studia Math.* **2** (1930), 1–6.
[94] W. Schirotzek, *Non-Smooth Analysis*, Universitext (Springer, 2007).
[95] L. Thibault, Various forms of metric regularity, unpublished note, Univ. de Montpellier (1999).
[96] K. Sh. Tsiskaridze, 'Extremal problems in Banach spaces', in: *Nekotorye Voprosy Matematicheskoy Theorii Optimalnogo Upravleniya (Some Problems of the Mathematical Theory of Optimal control)*, Inst. Appl. Math. (Tbilisi State University, 1975), (in Russian).

https://doi.org/10.1017/S1446788715000701 Published online by Cambridge University Press

[97] C. Ursescu, 'Inherited openness', *Revue Roumaine Math. Pures Appl.* **41** (1996), 5–6, 401–416.
[98] C. Ursescu, 'Linear openness of multifunctions in metric spaces', *Intl J. Math. Math. Sci.* (2) (2005), 203–214.

A. D. IOFFE, Department of Mathematics,
Technion, Haifa 32000, Israel
e-mail: alexander.ioffe38@gmail.com, ioffe@tx.technion.ac.il