Chapter 2
Metric Theory: Phenomenology

In this chapter we, roughly speaking, translate some principal results of the classical theory presented in the preceding chapter into purely metric language in which the key word is "distance" and words like "derivative" or "tangent space" make little sense. The point is that the phenomena that appear in the conclusions of Theorems 1.12, 1.15 and 1.38 should most naturally be studied in metric spaces with no reference to any linear or differentiable structures. Moreover, the most natural objects to consider in the context of regularity theory are set-valued mappings.

Our attention will be focused on two types of behavior. The first is the proper "regular" behavior which manifests itself either through linear openness (with proportional dependence of the radii of balls in the domain space and balls in the range space covered by the images of the first), as in the Theorems of Graves and Lyusternik–Graves, or through metric estimates for distances to solution sets of inclusions or equations (error bounds), as in Theorem 1.15. The second type is the Lipschitz-like behavior which, for set-valued mappings, was first described by Aubin and for single-valued mappings reduces to the standard locally Lipschitz dependence.

A remarkable fact is that the regular behavior of a (set-valued) mapping and a certain type of Lipschitz behavior of its inverse are equivalent phenomena, as well as the two manifestations of the regular behavior described in the previous paragraph. These equivalences are valid unconditionally for all set-valued mappings between metric spaces and, moreover, they can be expressed not just qualitatively but also in precise quantitative terms involving certain regularity rates (or moduli) which provide us with quantitative measures of regularity associated with each of the three types of phenomena. This underscores the metric nature of the phenomena: the equivalences remained somewhat unnoticed in the classical theory, although the understanding that regularity is a key to stability was always among the leading principles. The Equivalence Theorem proved in §2 (also containing all main definitions) is the first fundamental result of the metric regularity theory. Moreover, in concrete situations, it allows us to choose which of the three equivalent properties is most convenient to work with, and this is a valuable practical asset.

A.D. Ioffe, *Variational Analysis of Regular Mappings*, Springer Monographs in Mathematics, DOI 10.1007/978-3-319-64277-2_2

The second principal result of the theory (or rather a group of results) are quantitative regularity criteria established in the third section. They are the most general and all other criteria to be discussed later in the book are their consequences. The criteria offer verifiable necessary and sufficient regularity conditions containing mechanisms to determine regularity rates (the rate of surjection first of all). We shall often see that the criteria offer a convenient and powerful instrument of analysis, and application of the criteria often needs very little calculation (compare even to more specified infinitesimal criteria to be discussed in the next and the 5th chapters). In §3 we provide a simple proof of a very basic density theorem which says that if the images of the balls of the domain space are sufficiently dense in some balls in the range space then the former actually cover the latter.

In the fourth section we consider four weaker regularity related concepts: subregularity, calmness, controllability, and linear recession. Controllability and subregularity relate to linear openness and metric regularity, while linear recession and calmness can be viewed as weakened versions of the Aubin property, the first equivalent to controllability of the inverse mapping and the second to its subregularity. The pair subregularity–calmness is rather well studied (especially in the finite-dimensional setting) and plays an important role in a number of applications. The term "controllability" is of course borrowed from control theory and its relation to the standard controllability concept is obvious from the definition. The distinguished role of controllability is that it provides still another, and in some cases the most convenient, characterization for local regularity.

In the short fifth section we discuss another central problem of regularity theory: the effect of small perturbations of the mapping on regularity and regularity rates in particular. The general principle for mappings into Banach spaces (and more generally to linear metric spaces with shift invariant metric) is that an additive single-valued Lipschitz perturbation does not kill regularity as long as the Lipschitz constant of the perturbing mapping does not exceed the rate of regularity of the given map. Here we consider a more general perturbation scheme that works for arbitrary metric domain and range spaces and allows us to extend the principle to purely metric settings.

The sixth section is devoted to implicit function theorems for set-valued mappings between metric spaces. This is essentially a consequence of the equivalence theorem, a part of which can be easily interpreted as a sort of inverse mapping theorem for set-valued mappings. However surprising it may look at first glance, it is possible to establish meaningful results bearing some principal features of the classical implicit function theorem even in this extremely general situation. The key setting here is a mapping of two variables which displays regular behavior as a function of one of the variables and the Lipschitz-type behavior as a function of the other. Subsequently we shall watch the evolution of the implicit function theorems following accumulation of structural requirements on the mapping.

In the seventh section we consider nonlinear regularity models which correspond, say in the case of openness, to nonlinear dependence of the radii of balls in the domain space and the radii of balls in the range space covered by the former. We prove nonlinear analogues of the equivalence theorem, of the general regularity criterion and of the density theorem.

Our principal technical instrument is the variational principle of Ekeland, although we show that the perturbation theorem can also be obtained with the help of a Newton-like iterative procedure similar to that used by Lyusternik and Graves.[1] We give a complete proof of the principle in the introductory section and then apply it to give a short proof of the famous Bishop–Phelps–Bollobás theorem (a more precise version of the famous Bishop–Phelps theorem) which will be needed in what follows. Note that Ekeland's principle is also the key element in developing in the next chapter an infinitesimal mechanism of slopes which allows us to give a precise quantitative characterization of local regularity properties in metric spaces and opens a way to obtain subdifferential regularity characterizations for mappings in Banach spaces.

We conclude the chapter with a supplement containing a brief discussion of the regularity problem for compositions of set-valued mappings.

2.1 Introduction

Notation and terminology. In this chapter all spaces are metric, with a completeness requirement occasionally added since Ekeland's principle and Newton-type techniques need completeness. We shall keep all notation of the previous chapter: as in the case of a Banach space, $B(x,r)$ and $\overset{\circ}{B}(x,r)$ are respectively the closed and open balls of radius r around x. If $Q \subset X$, then by $B(Q,r)$ (resp $\overset{\circ}{B}(Q,r)$) we denote the union of the closed (resp. open) r-balls around elements of Q:

$$B(Q,r) = \{x :\ \exists u \in Q \text{ such that } d(x,u) \le r\} = \bigcup_{u \in Q} B(u,r).$$

We shall usually denote the distance by the same letter d, no matter which space is considered – this should not create any confusion. The distance to the empty set is $+\infty$ by the standard convention.

Given two metric spaces X and Y, we shall mainly deal with the following three metrics in $X \times Y$, each associated with some numerical parameter: the ℓ^∞-type ξ-metric

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

and two ℓ^1-type metrics

$$\boxed{d_{K,1}((x,y),(x',y')) = Kd(x,x') + d(y,y')}$$

[1] It seems to be appropriate to quote here [82]: "in essence the whole history of the generalizations of the Lyusternik theorem reduces to finding new formulations from the standard process of proof". But it soon became clear that certain results proved with the help of Ekeland's principle (e.g. the general regularity criterion of Theorem 2.46) can hardly be obtained using Lyusternik–Graves iterations.

and

$$\boxed{d_{1,K}((x, y), (x', y')) = d(x, x') + Kd(y, y').}$$

If $K = 1$, we usually omit the subscript and write simply $d((x, y), (x', y'))$.

For two sets $P,\ Q \subset X$ we define the *distance* (or *gap*) between P and Q by

$$d(P, Q) = \inf\{d(x, u) :\ x \in P,\ u \in Q\},$$

the *excess* of P from Q by

$$\mathrm{ex}(P, Q) = \inf\{r > 0 :\ P \subset B(Q, r)\} = \sup\{d(x, Q) :\ x \in P\}$$

and the *Hausdorff distance* between P and Q by

$$\mathbf{H}(P, Q) = \max\{\mathrm{ex}(P, Q), \mathrm{ex}(Q, P)\}.$$

The general convention is that for a nonempty Q

$$\mathrm{ex}(\emptyset, Q) = 0, \quad \mathrm{ex}(Q, \emptyset) = \infty.$$

Given a set $Q \subset X$, by cl Q, int Q and bd Q = (cl Q)\(int Q) we denote the closure, interior and boundary of Q.

Starting with this chapter, we shall typically consider *extended-real-valued* functions which can assume values $\pm\infty$ along with real values. The function is *proper* if it is everywhere greater than $-\infty$ and not everywhere equal to ∞. We write $I\!R_+$ for $[0, \infty)$ and $I\!R_-$ for $(-\infty, 0]$, and sometimes use the notation $\overline{I\!R}$ for $[-\infty, \infty]$. So let f be a function on X. We associate with f two sets

$$\mathrm{epi}\ f = \{(x, \alpha) \in X \times I\!R :\ \alpha \geq f(x)\} \quad \text{and} \quad \mathrm{dom}\ f = \{X \in X :\ |f(x)| < \infty\},$$

the first called the *epigraph* of f and the second the *domain* of f. The *indicator function* (or just *indicator*) of a set $Q \subset X$ is

$$i_Q(x) = \begin{cases} 0, & \text{if} \quad x \in Q; \\ \infty, & \text{otherwise.} \end{cases}$$

Clearly dom $i_Q = Q$, epi $i_Q = Q \times I\!R_+$.

If f is a function on X and $Q \subset X$, then the *restriction of f to Q* is the function $f|_Q$ equal to $f(x)$ if $x \in Q$ and $+\infty$ if $x \notin Q$. In particular, if f is everywhere greater than $-\infty$, then $f|_Q(x) = f(x) + i_Q(x)$.

To denote level sets, sublevel sets etc. of (extended real-valued) functions we use the symbols

$$[f = \alpha] = \{x :\ f(x) = \alpha\}, \quad [f \leq \alpha] = \{x :\ f(x) \leq \alpha\} \quad \text{etc.}$$

The symbol $x \to_Q \bar{x}$ means $x \to \bar{x}$ and $x \in Q$. If f is a function, then we write $u \to_f x$ as an abbreviation of $u \to x$ and $f(u) \to f(x)$. By $\alpha^+ = \max\{\alpha, 0\}$ we denote the positive part of an $\alpha \in I\!R$.

We shall use the expression "lsc" as an abbreviation for *lower semicontinuous*. Recall that f is lsc if all sublevel sets $\{x \in X : f(x) \le \alpha\}$ are closed. This amounts to the epigraph of f being a closed set. We say that f is *upper semicontinuous* if $-f$ is lower semicontinuous. Given a function f we denote by $\overline{f}$ its *lower envelope*:

$$\overline{f}(x) = \liminf_{u \to x} f(u).$$

It is said that $F : X \to Y$ *satisfies the Lipschitz condition* (or *is Lipschitz*, or *is Lipschitz continuous*) on a set $S \subset X$ if there is a $K \ge 0$ such that $d(F(x), F(x')) \le Kd(x, x')$ for all $x, x' \in S$. The lower bound of all such K (*the Lipschitz constant* (or *rank*)) will be denoted $\partial_S F$. If F is Lipschitz continuous in a neighborhood of x, then we also say that F satisfies the Lipschitz condition *at* x and define *the Lipschitz constant of* F *at* x by

$$\text{lip } F(x) = \lim_{\varepsilon \to 0} \text{lip}_{B(x,\varepsilon)} F.$$

If X is a metric space and $Q \subset X$, then the *induced metric* on Q is the restriction to Q of the metric of the ambient space.

Set-valued mappings. The symbol $F : X \rightrightarrows Y$ means "F is a set-valued mapping from X into Y", that is, a correspondence which to every x associates a set $F(x)$, possibly empty. As with functions, with every set-valued mapping F, we associate two sets, the *graph* and the *domain*:

$$\text{Graph } F = \{(x, y) \in X \times Y : y \in F(x)\}, \quad \text{dom } F = \{x \in X : F(x) \ne \emptyset\}.$$

A set-valued mapping $F : X \rightrightarrows Y$ is *Lipschitz*, or *satisfies the Lipschitz condition* near $\bar{x} \in \text{dom } F$, if there is a $K > 0$ such that

$$\mathbf{H}(F(x), F(x')) \le Kd(x, x')$$

for all x, x' in a neighborhood of $\bar{x}$. The *restriction* of F to $Q \subset X$ is the set-valued mapping $F|_Q : X \rightrightarrows Y$ coinciding with F on Q and assuming the empty value outside of Q:

$$F|_Q(x) = \begin{cases} F(x), & \text{if } x \in Q \\ \emptyset, & \text{otherwise.} \end{cases}$$

To every set-valued mapping $F : X \rightrightarrows Y$ and any $y \in Y$ we associate three functions:

$$\boxed{\varphi_y(x, v) = \varphi_{F,y}(x, v) = d(y, v) + i_{\text{Graph } F}(x, v)}$$

(that is, the restriction of $d(y, v)$ (viewed as a function of (x, v)) to the graph of F;

$$\boxed{\psi_y(x) = \psi_{F,y}(x) = d(y, F(x));}$$

and

$$\boxed{\omega_y^K(x) = \omega_{F,y}^K(x) = d_{1,K}((x, y), \text{Graph } F).}$$

As a rule, we omit the subscript F and simply write $\varphi_y,\ \psi_y,\ \omega_y^K$ since it is typically clear which mapping we are talking about. The definitions of the functions have been put into boxes because the functions will be a key element in all regularity criteria and the notation will be used throughout the text. Occasionally we shall recall the definitions. As a rule we shall omit the subscript F and simply write φ_y, ψ_y and ω_y^K.

Note that ψ_y may fail to be a lower semicontinuous function. Therefore we shall often be compelled to consider instead the lower closure of the function:

$$\overline{\psi}_y(x) = \liminf_{u \to x} \psi_y(u).$$

The *inverse* of F is the mapping $F^{-1} : Y \rightrightarrows X$ defined by $F^{-1}(y) = \{x \in X : y \in F(x)\}$. Clearly Graph $F^{-1} = \{(y, x) \in Y \times X :\ (x, y) \in \text{Graph } F\}$.

A set-valued mapping F is *closed-valued* (resp. *compact-valued* etc.) if for any x the set $F(x)$ is closed (resp. compact); F is a *closed mapping* if its graph is a closed set; F is *locally closed* at a point of the graph if there is a closed neighborhood of the point whose intersection with the graph is closed. The map is *locally closed* if it is locally closed at any point of its graph.

We shall say that a closed set-valued mapping F is *upper (lower) semicontinuous at* $\overline{x}$ if the function $x \mapsto d(y, F(x))$ is lower (upper) semicontinuous at $\overline{x}$ for every $y \in Y$. This definition is weaker than the standard neighborhood definition (for any $\varepsilon > 0$ there is a $\delta > 0$ such that $\text{ex}(F(x), F(\overline{x})) < \varepsilon$ if $d(x, \overline{x}) < \delta$) but convenient when working with regularity properties. Useful observation: if F is single-valued and upper (lower) semicontinuous, then it is continuous.

If we have a set-valued mapping $F : X \rightrightarrows Y$ and $x \in \text{cl}\,(\text{dom } F)$, then the *outer limit* of F at x is

$$\limsup_{u \to x} F(u) := \{y \in Y :\ \exists\, x_n \to x,\ y_n \in F(x_n),\ y_n \to y\},$$

and the *lower (or inner) limit* of F at x is

$$\liminf_{u \to x} F(u) := \{y :\ \forall\, \{x_n\} \to x,\ \exists\, y_n \in F(x_n),\ y_n \to y\}.$$

(In both definitions $\exists\, y_n$ should be understood as "there is an n_0 such that $\exists\, y_n$ for $n \geq n_0$".)

Exercise 2.1. Given a set-valued mapping $F : X \rightrightarrows Y$, then
(a) the set-valued mapping $\tilde{F}(x) = \liminf\limits_{u \to x} F(u)$ is lower semicontinuous;
(b) a closed compact-valued mapping is upper semicontinuous;
(c) a closed-valued upper semicontinuous mapping is closed.

This example explains why we use the term "outer limit" rather than "upper limit": it is an easy matter to find an example of a closed mapping which is not upper semicontinuous. It is also worth mentioning at this point that sometimes closed set-valued mappings are called "outer semicontinuous" (see e.g. [287]).

Here are a few examples of set-valued mappings that often appear in analysis and in applications.

Example 2.2 (epigraphical mapping). Let f be a function on X. The mapping

$$\mathrm{Epi}_f(x) = \{\alpha \in I\!R : \ \alpha \geq f(x)\}$$

is the *epigraphical mapping* associated with f. Its graph is of course Epi f and the domain is $\{x \in X : \ f(x) < \infty\}$. Epi f is a closed-graph mapping if and only if f is lower semicontinuous.

Example 2.3 (solution mapping). Let $F : X \times P \rightrightarrows Y$ and $\overline{y} \in Y$. We view x as an argument of the mapping and p as a parameter. The set-valued mapping

$$S(p) = \{x \in X : \overline{y} \in F(x, p)\}$$

is called the *solution mapping* of the inclusion $\overline{y} \in F(x, p)$. Observe that often, when F does not depend on p and the left-hand side $\overline{y}$ itself is viewed as a parameter, the inverse mapping $F^{-1}(y)$ is the solution mapping of the inclusion $y \in F(x)$.

Example 2.4 (closed operator). Let X and Y be Banach spaces, and let $A : X \to Y$ be a closed and possibly unbounded operator with the domain dom A. Then

$$F(x) = \begin{cases} Ax, & \text{if } x \in \operatorname{dom} A, \\ \emptyset, & \text{otherwise} \end{cases}$$

is a closed set-valued mapping.

Example 2.5 (generalized equation). These are relations of the form

$$0 \in f(x, y) + F(x),$$

where $f : X \times Y \to Z$ is a single-valued mapping and $F : X \rightrightarrows Z$ is a set-valued mapping into a normed space Z.

There are some standard ways to produce new set-valued mappings. Given two set-valued mappings $F : \ X \rightrightarrows Y$ and $G : \ Y \rightrightarrows Z$, then

$$(G \circ F)(x) = \bigcup_{y \in F(x)} G(y)$$

is the *composition* of F and G. An important particular case is the *restriction* of F to $Q \subset X$:

$$F|_Q(x) = \begin{cases} F(x), & \text{if } x \in Q, \\ \emptyset, & \text{otherwise.} \end{cases}$$

We can view it as the composition of the embedding of Q into X and F.

Variational principle of Ekeland. Although this result is widely available in the monographic literature (e.g. [13, 14, 49, 149, 246]), we give it with a full proof in view of its fundamental importance for the theory and variational analysis in general.

Theorem 2.6 (Ekeland). *Let f be a lower semicontinuous and bounded from below extended-real-valued function on a complete metric space (X, d), not identically equal to $+\infty$. Assume that $f(\overline{x}) \le \inf f + \varepsilon$ for some $\overline{x} \in X$ and $\varepsilon > 0$. Then for any $\lambda > 0$ there is a $z = z(\lambda)$ such that*

(a) $f(z) \le f(\overline{x}) - \dfrac{\varepsilon}{\lambda} d(z, \overline{x})$;

(b) $d(z, \overline{x}) \le \lambda$;

(c) $f(x) + \dfrac{\varepsilon}{\lambda} d(x, z) > f(z), \quad \forall x \ne z.$

Proof. First we observe that it is sufficient to prove the theorem for $\varepsilon = \lambda = 1$. Indeed, if the theorem has been proved in this specific case, then, given arbitrary $\varepsilon > 0,\ \lambda > 0$, we apply the theorem to the function $\varphi(x) = \varepsilon^{-1} f(x)$ on the space (X, d'), where $d'(x, y) = \lambda^{-1} d(x, y)$.

1. For any $v \in \operatorname{dom} f$ set $S(v) = \{u \in X : \ f(u) \le f(v) - d(u, v)\}$. Then $S(v)$ is a nonempty ($v \in S(v)$) closed set and

$$u \in S(v) \quad \text{implies} \quad S(u) \subset S(v) \tag{2.1.1}$$

(indeed, if $u \in S(v)$ and $x \in S(u)$, then by the triangle inequality

$$f(x) \le f(u) - d(x, u) \le f(v) - d(u, v) - d(x, u) \le f(v) - d(x, v).)$$

2. Consider a sequence $x_0, x_1, \ldots$ with $x_0 = \overline{x}$ and other elements chosen according to the following rule:

$$x_{i+1} \in S(x_i), \quad f(x_{i+1}) \le (1/2)(f(x_i) + \inf_{x \in S_i} f(x)). \tag{2.1.2}$$

(The second condition simply means that $f(x_{i+1})$ must not exceed the middle value of f on S_i.) Then $\{x_i\}$ is a Cauchy sequence. Indeed, $f(x_{i+1}) \le f(x_i)$, which means that the sequence $\{f(x_i)\}$ does not increase and hence, as it is bounded below by $\inf f$, it converges to a certain α. Furthermore (using again the triangle inequality) we can write

$$d(x_i, x_{i+m}) \leq \sum_{j=i}^{j=i+m-1} d(x_j, x_{j+1})$$
$$\leq \sum_{j=i}^{j=i+m-1} (f(x_j) - f(x_{j+1})) = f(x_i) - f(x_{i+m}) \leq f(x_i) - \alpha,$$

which proves the claim.

3. As X is complete, the sequence $\{x_i\}$ converges to a certain z. Then (a) follows from the fact that by (2.1.2) $f(x_{i+1}) \leq f(x_i) - d(x_i, x_{i+1})$, so that $f(x_n) \leq f(x_0) - (d(x_n, x_{n-1} + \ldots + d(x_1, x_0)) \leq f(\overline{x}) - d(x_n, \overline{x}))$, and (b) follows from the last inequality applied for $i = 1$ (as $\alpha \geq \inf f$):

$$d(\overline{x}, z) = \lim_{m\to\infty} d(\overline{x}, x_m) \leq f(\overline{x}) - \alpha \leq f(\overline{x}) - \inf f \leq 1.$$

To verify (c) we first observe that it is equivalent to

$$S(z) = \{z\}.$$

By (2.1.1) $x_{i+m} \in S(x_i)$ for all i and m. Therefore $z \in \bigcap S(x_i)$, hence again by (2.1.1), $S(z) \subset S(x_i)$ and for any $u \in S(z)$ we have

$$f(z) - d(u, z) \geq f(u) \geq \lim_{i\to\infty} \inf_{S_i} f(x) \geq \lim_{i\to\infty} [f(x_{i+1}) - (f(x_i) - f(x_{i+1}))] = f(z),$$

whence $d(u, z) = 0$. This completes the proof of Ekeland's principle. □

As an application of Ekeland's principle we prove below the celebrated Bishop–Phelps–Bollobás theorem, which is actually an equivalent of Ekeland's principle for Banach spaces.

Theorem 2.7 (Bishop–Phelps–Bollobás). *Let X be a Banach space, and let S be a closed convex bounded subset of X. Let an $x^* \in X^*$ be given with $\|x^*\| = 1$, and let $w \in S$ satisfy $\langle x^*, w\rangle \geq \sup_{x\in S}\langle x^*, x\rangle - \varepsilon^2/4$. Then there are an $\overline{x} \in S$ and $y^* \in Y^*$ with $\|y^*\| = 1$ such that $\|\overline{x} - w\| < \varepsilon,\ \|y^* - x^*\| < \varepsilon$ and $\langle y^*, \overline{x}\rangle = \max_{x\in S}\langle y^*, x\rangle$. In particular, the collection of the elements of the dual sphere of X^* which attain their maximum on S is norm dense in the sphere.*

Proof. Denote by B^* the unit ball in X^*. As S is closed, it is a complete metric space with respect to the distance induced by the norm in X. Consider the function $f(x) = -\langle x^*, x\rangle$ on S. Then $f(w) \leq \inf f + \varepsilon^2/4$. Take an arbitrary ε, say $\varepsilon < 1/2$, and apply Ekeland's principle to find an $\overline{x} \in S$ such that $\|\overline{x} - w\| \leq \varepsilon/2$ and $f(x) + (\varepsilon/2)\|x - \overline{x}\| > f(\overline{x})$ for all $x \in S$ other than $\overline{x}$, that is to say, such that $-\langle x^*, x - \overline{x}\rangle + (\varepsilon/2)\|x - \overline{x}\| > 0$ if $x \in S, x \neq \overline{x}$.

Set $g(x) = -\langle x^*, x\rangle + (\varepsilon/2)\|x\|$ and $Q = \{x \in X :\ g(x - \overline{x}) \leq 0\}$, which is a translation of the zero sublevel set of g. The latter is a convex cone with a nonempty interior. (Indeed take a $u \in X$ with $\langle x^*, u\rangle > (\varepsilon/2)$. Then $g(u) \leq -\langle x^*, u\rangle + (\varepsilon/2) <$

0. On the other hand, g is homogeneous of degree 1: $g(\lambda x) = \lambda g(x)$ whenever $\lambda > 0$.) It follows that $\overline{x}$ belongs to the boundary of Q, and moreover, this is the only point of S that belongs to Q. Therefore we can separate Q and S by a nonzero linear functional $y^* \in X^*$:

$$\sup_{x\in S}\langle y^*, x\rangle = \langle y^*, \overline{x}\rangle = \inf_{x\in Q}\langle y^*, x\rangle. \tag{2.1.3}$$

Multiplying y^* by a positive scalar if necessary, we can assume that $\|y^*\| = 1$. We shall show that $\|x^* - y^*\| \le \varepsilon$ and this will complete the proof.

The right equality in (2.1.3) says that $z^* = -y^*$ is a normal vector to Q at $\overline{x}$ or, equivalently, a normal vector at zero to $\{x : \ g(x) \le 0\}$. As we have seen, g is a continuous convex function which also assumes negative values. In this case it follows from standard rules of convex analysis that z^* must be positively proportional to an element of the subdifferential of g at zero, that is, there is a $\lambda > 0$ such that $\lambda(-y^*) \in -x^* + (\varepsilon/2)B^*$, which means that $\|\lambda y^* - x^*\| \le \varepsilon/2$. But $\|y^*\| = \|x^*\| = 1$, so $|\lambda - 1| = |\lambda\|y^*\| - \|x^*\|| \le \|\lambda y^* - x^*\| \le \varepsilon/2$ and therefore $\|y^* - x^*\| \le \|\lambda y^* - x^*\| + |1-\lambda|\|y^*\| \le \varepsilon$, as claimed. □

2.2 Regularity: Definitions and Equivalences

In this section we introduce the most fundamental concepts of the theory. Unlike the notion of regularity in the classical setting stated in terms of linear approximations of mappings, in the definitions we instead use descriptions of phenomena as such. A clear reason for this is the absence of a good local approximation for a set-valued mapping near a point of its graph comparable in precision with the linear approximation of a smooth mapping by its derivative. However, we shall see in the next chapter that a sort of infinitesimal characterization can also be given in the general setting of set-valued maps between metric spaces.

2.2.1 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 $(\bar{x}, \bar{y}) \in \text{Graph}\, F$.

Definition 2.8 (*local regularity properties*). We say that F is

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

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

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

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

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

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

• has the *Aubin property* (or is *pseudo-Lipschitz*)[2] near $(\bar{x}, \bar{y})$ if there are $K > 0$ and $\varepsilon > 0$ such that

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

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

Exercise 2.9. Prove that in the definition of linear openness we can equivalently replace the closed ball $B(y, rt)$ by the open ball $\overset{\circ}{B}(y, rt)$. In other words, *$F$ is open at a linear rate near $(\bar{x}, \bar{y})$ if and only if*

$$\overset{\circ}{B}(y, rt) \cap B(\overline{y}, \varepsilon) \subset F(B(x, t)), \quad \text{if } (x, y) \in \text{Graph } F,\ d(x, \overline{x}) < \varepsilon,\ t \ge 0,$$

and the upper bound of such r coincides with $\text{sur } F(\overline{x}|\overline{y})$. Therefore we can use either definition, whichever convenient.

The key 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 2.10 (local equivalence). *F is open at a linear rate near $(\bar{x}, \bar{y}) \in$ 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 $(\overline{y}, \overline{x})$. Moreover, under the convention that* $0 \cdot \infty = 1$,

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

We shall obtain the proposition as a corollary of the general Equivalence Theorem 2.25 proved later in this section. In view of the proposition it is natural to call F *regular near* $(\bar{x}, \bar{y})$ if the three equivalent properties hold near the point. It should be mentioned that in the literature the expression ‘regularity *at*’ (rather than *near*) is used, since it sounds better in certain contexts. (I would not even exclude the possibility that this expression has been occasionally used in the present text as well!) However, the word *near* gives a more precise description of the phenomenon because, according to the definitions, if the properties are satisfied for a certain point of the graph, then they are automatically valid for all points in the intersection of the graph with a neighborhood of the point.

---

[2]We shall use the term “Aubin property” only in the local context, leaving the term “pseudo-Lipschitz” for non-local situations.

There are some other equivalent descriptions which are specific to local regularity and do not have analogues in the general case. The first characterizes a certain level of robustness of local regularity. For instance, it enough to verify linear openness only for e.g. $0 < t < \varepsilon$, not for all $t \in [0, \infty)$.

Proposition 2.11 . *F is open at a linear rate near $(\bar{x}, \bar{y})$ if and only if there are $r > 0$ and $\varepsilon > 0$ such that*

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

and the upper bound of such r is precisely $\operatorname{sur} F(\overline{x}|\overline{y})$.

Proof. The only if part of the proposition is of course trivial as well as the fact that the upper bound of r here is not smaller than $\operatorname{sur} F(\overline{x}|\overline{y})$. So let the inclusion above hold for some $r > 0$ and $\varepsilon > 0$. Take $\delta < \varepsilon/2$ so small that

$$B(\overline{y}, \delta) \subset B(F(x), rt) \text{ if } t \ge \varepsilon \quad \text{and} \quad B(\overline{y}, \delta) \subset F(B(\overline{x}, \varepsilon/2)),$$

which is of course possible. Then for any x with $d(x, \overline{x}) < \delta$

$$\overset{\circ}{B}(F(x), rt) \cap B(\overline{y}, \delta) \subset \overset{\circ}{B}(F(x), rt) \cap B(\overline{y}, \varepsilon) \subset F(B(x, t))$$

if $t < \varepsilon$, and for $t \ge \varepsilon$ we have

$$\overset{\circ}{B}(F(x), rt) \cap B(\overline{y}, \delta) \subset B(\overline{y}, \delta) \subset F(B(\overline{x}, \varepsilon/2)) \subset F(B(x, t)),$$

as claimed. □

The following exercise offers analogues of Proposition 2.11 for metric regularity and Aubin properties. (To see that the exercise is indeed valid one can use Theorem 2.25 with $U = \overset{\circ}{B}(\overline{x}, \varepsilon)$ and $\gamma(x) \equiv \varepsilon$, to be proved in the next subsection, along with Proposition 2.10. But a direct proof of the statements also seems to be a worthy enterprise.)

Exercise 2.12. Prove the following statement.
Let again $F : X \rightrightarrows Y$ and $(\bar{x}, \bar{y}) \in \text{Graph } F$. Then

(a) F is metrically regular near $(\bar{x}, \bar{y})$ with $\operatorname{reg} F(\overline{x}|\overline{y}) \le K$ if and only if for any $K' > K$ there is an $\varepsilon > 0$ such that
$d(x, F^{-1}(y)) \le K' d(y, F(x)), \quad \text{if } d(x, \overline{x}) < \varepsilon,\ d(y, \overline{y}) < \varepsilon,\ d(y, F(x)) < \varepsilon;$

(b) F has the Aubin property near $(\bar{x}, \bar{y})$ with $\operatorname{lip} F(\overline{x}|\overline{y}) \le K$ if and only if for any $K' > K$ there is an $\varepsilon > 0$ such that
$d(y, F(x)) \le K' d(x, u), \quad \text{if } d(x, \overline{x}) < \varepsilon,\ d(y, \overline{y}) < \varepsilon,\ d(x, u) < \varepsilon.$

Note that an indirect proof of (a) follows from Proposition 2.29 in the next subsection.

Exercise 2.13. Assume that $F : X \rightrightarrows Y$ is regular near $(\bar{x}, \bar{y})$ and $g : Y \to X$ is defined and Lipschitz continuous in a neighborhood of $\overline{y}$. Prove that $(F^{-1} + g)^{-1}$ is regular at $(\overline{x} + g(\overline{y}), \overline{y})$ and

$$\text{reg } (F^{-1} + g)^{-1}(\overline{x} + g(\overline{y})|\overline{y}) \le \text{lip } F^{-1}(\overline{y}|\overline{x}) + \text{lip } g(\overline{y}).$$

Local regularity will remain the main object of interest throughout the book (as it is in the literature in general). Looking back to the classical theory, we see that a C^1-mapping F, regular at a certain $\overline{x}$ (in the sense of the previous chapter), is open at a linear rate at $\overline{x}$ with the rate of surjection not smaller that $C(F'(\overline{x}))$ (by the Lyusternik–Graves theorem) and metrically regular at $\overline{x}$ with the modulus of metric regularity not greater than $C(F'(\overline{x}))^{-1}$ (by Theorem 1.15). We can also deduce from the implicit function theorem that the inverse (set-valued!) mapping F^{-1} has the Aubin property near $(\overline{y}, \overline{x})$ (where as usual $\overline{y} = F(\overline{x})$) if Ker $F'(\overline{x})$ splits X.

Proposition 2.14 (characterization of Aubin's property). *Let X and Y be metric spaces, let $F : X \rightrightarrows Y$, and let $\overline{y} \in F(\overline{x})$. Then the following properties are equivalent for any $K > 0$:*

(a) F has the Aubin property near $(\bar{x}, \bar{y})$ with $\text{lip } F(\overline{x}|\overline{y}) \le K$*;*

(b) for any $\delta > 0$ there are neighborhoods $U \subset X$ of $\overline{x}$ and $V \subset Y$ of $\overline{y}$ such that the inclusion

$$F(u) \bigcap V \subset B(F(x), (K + \delta)d(x, u))$$

holds for any $x, u \in U$ or equivalently,

$$\text{ex}(F(u) \cap V, F(x)) \le (K + \delta)d(x, u).$$

In particular, $F(u) \cap B(\overline{y}, (K + \delta)\varepsilon) \neq \emptyset$ if u is sufficiently close to $\overline{x}$;

(c) for any $\delta > 0$ there are neighborhoods $U \subset X$ of $\overline{x}$ and $V \subset Y$ of $\overline{y}$ such that for any y the function $x \to d(y, F(x))$ satisfies on U the Lipschitz condition with constant $\le K + \delta$.

Proof. The implications (b) $\Rightarrow$ (a) and (c) $\Rightarrow$ (b) are immediate. Indeed, taking $x = \overline{x}$ in either (b) or (c), we see that $F(x) \neq \emptyset$ for $x \in U$. If now (b) holds, then taking $x, u \in U$ and $y \in F(u) \bigcap V$, we have $y \in B(F(x), (K + \delta)d(x, u))$, which is the same as $d(y, F(x)) \le (K + \delta)d(x, u)$, whence (a). Likewise, if (c) holds, then taking $x, u \in U$ and $y \in F(u) \bigcap V$, we have $d(y, F(x)) \le (K + \delta)d(x, u)$, which means that $y \in B(F(x), (K + \delta)d(x, u))$, whence (b).

So we need to prove that (a) implies (c). If (a) holds, then for any $\delta > 0$ there is an $\varepsilon > 0$ such that $d(y, F(x)) \le (K + \delta)d(x, u)$ if x, u are within ε of $\overline{x}$ and $y \in F(u)$ is within ε of $\overline{y}$. Applying this to $u = \overline{x}$ and $y = \overline{y}$, we conclude that for every x with $d(x, \overline{x}) < \xi = \varepsilon / \max\{1, K + \delta\}$ the set $F(x)$ contains a v with $d(v, \overline{y}) < \varepsilon$.

Take now x and x' within ξ of $\overline{x}$, a $v \in F(x)$ such that $d(v, \overline{y}) < \varepsilon$ and a certain y. We have

$$\begin{aligned} d(y, F(x)) - d(y, F(x')) &\le d(y, v) - (d(y, v) - d(v, F(x')) \\ &= d(v, F(x')) \le (K + \delta) d(x, x'), \end{aligned}$$

and changing the roles of x and x', we get $|d(y, F(x)) - d(y, F(x'))| \le (K + \delta)d(x, x')$. □

Since any function of two variables which is Lipschitz with respect to each of them in a neighborhood of a certain point is jointly Lipschitz in the neighborhood, we get as an immediate consequence

Corollary 2.15. *A set-valued mapping $F : X \rightrightarrows Y$ has the Aubin property near $(\bar{x}, \bar{y}) \in \text{Graph } F$ if and only if there are neighborhoods $U \subset X$ of $\overline{x}$ and $V \subset Y$ of $\overline{y}$ such that the function $(x, y) \to d(y, F(x))$ is Lipschitz on $U \times V$.*

A set-valued mapping $F : X \rightrightarrows Y$ which has the Aubin property near $(\overline{x}, y)$ for any $y \in F(\overline{x})$ need not be Lipschitz near $\overline{x}$. Consider for instance the mapping $F(x) = \{x^* : \langle x^*, x\rangle \ge 0\}$ from a Banach space into its dual (considered with the norm topology). However, the implication does hold if Y is a compact space.

Proposition 2.16 (a pseudo-Lipschitz map into a compact space is Lipschitz). *Let $F : X \rightrightarrows Y$, where Y is a compact metric space, be a set-valued mapping with closed graph. Assume that for some $\overline{x} \in \text{dom } F$ and all $y \in F(\overline{x})$ it has the Aubin property near $(\overline{x}, y)$. Then F is Lipschitz in a neighborhood of $\overline{x}$.*

Proof. By the assumption, all sets $F(x)$ are closed. Let us check that F is upper semicontinuous, that is, for every $x \in \text{dom } F$ and every $\delta > 0$ there is an $\varepsilon > 0$ such that $F(u) \subset B(F(x), \delta)$ for any $u \in B(x, \varepsilon)$. Assuming the contrary, we shall find a $\delta > 0$ and a sequence $(x_n, y_n) \in \text{Graph } F$ such that $x_n \to x$ and $d(y_n, F(x)) \ge \delta$. As Y is a compact set, we may assume that (y_n) converges to some y. Clearly $y \notin F(x)$. But the graph of F is closed, so y must belong to $F(x)$. The contradiction proves the claim.

By Proposition 2.14(b) for any $y \in F(\overline{x})$ there are neighborhoods $U(y)$ of $\overline{x}$ and $V(y)$ of y and a positive number $K(y)$ such that for any $x,\ x' \in U(\overline{x})$

$$F(x) \cap V(y) \subset B(F(x'), K(y)d(x, x')).$$

As $F(\overline{x})$ is a compact set, we can find a finite collection $\{y_1, \dots, y_k\}$ of elements of $F(\overline{x})$ such that the union of $V_i = V(y_i)$ covers $F(\overline{x})$. Then V is a neighborhood of $F(\overline{x})$ and we can find a $\delta > 0$ such that $B(F(\overline{x}), \delta) \subset V$. Choose an $\varepsilon > 0$ such that $F(B(\overline{x}, \varepsilon)) \subset B(F(\overline{x}), \delta)$. Set $U = (\cap U(y_i)) \cap \overset{\circ}{B}(\overline{x}, \varepsilon)$. Then for any $x, x' \in U$

$$F(x) = F(x) \cap V = \bigcup_i F(x) \cap V_i \subset B(F(x'), K(y)d(x, x')),$$

as claimed. □

Exercise 2.17. Prove that for any metric spaces X, Y, any set-valued mapping $F : X \rightrightarrows Y$ and any $(\bar{x}, \bar{y}) \in \text{Graph } F$

$$\text{lip } F(\overline{x}|\overline{y}) \geq \text{sur } F(\overline{x}|\overline{y}).$$

This is an extension of the obvious inequality $\|A\| \geq C(A)$ for a linear bounded operator from one Banach space to another.

We conclude the subsection with two more equivalent characterizations of local metric regularity.

Recall that $\overline{\psi}_y$ stands for the lower closure of ψ_y:

$$\overline{\psi}_y(x) = \liminf_{u \to x} \psi_y(u).$$

Proposition 2.18. *Let* $F : X \rightrightarrows Y$ *be a mapping with closed graph, and let* $(\bar{x}, \bar{y}) \in \text{Graph } F$. *Then* F *is metrically regular at* $(\bar{x}, \bar{y})$ *if and only if there is a* $K \geq 0$ *such that the inequality*

$$d(x, F^{-1}(y)) \leq K\overline{\psi}_y(x)$$

holds for all (x, y) *in a neighborhood of* $(\bar{x}, \bar{y})$. *Moreover, the lower bound of such* K *is precisely* reg $F(\overline{x}|\overline{y}))$.

Proof. As $\overline{\psi}_y(x) \leq d(y, F(x))$, we only need to verify the "only if" part of the proposition. So let F be metrically regular at $(\bar{x}, \bar{y})$, and let $d(x, F^{-1}(y)) \leq Kd(y, F(x))$ for all (x, y) in a neighborhood of $(\bar{x}, \bar{y})$.

First we note that $y \in F(x)$ if $\overline{\psi}_y(x)) = 0$ (see Lemma 2.48 in the next section). Assume now that (x, y) is sufficiently close to $(\bar{x}, \bar{y})$ and $\overline{\psi}_y(x) > 0$. Let as above $x_n \to x$ and $\psi_y(x_n) \to \overline{\psi}_y(x)$. As F is metrically regular, $d(x_n, F^{-1}(y)) \leq Kd(y, F(x_n)) \to \overline{\psi}_y(x)$ and the result follows. $\square$

To state the second result we need the following definition.

Definition 2.19 (*graph regularity*). F is said to be *graph-regular near* $(\bar{x}, \bar{y}) \in \text{Graph } F$ if there are $K > 0,\ \varepsilon > 0$ such that the inequality

$$d(x, F^{-1}(y)) \leq d_{1,K}((x, y), \text{Graph } F) = \omega_y^K(x), \tag{2.2.1}$$

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

It turns out that local regularity is equivalent to graph regularity near the same point. The advantage of the latter is that in certain cases it is easier to work with graph regularity because the function $(x, y) \mapsto d((x, y), \text{Graph } F)$ is Lipschitz continuous whereas $x \mapsto d(y, F(x))$ may not be.

Proposition 2.20 (metric regularity vs. graph regularity). *Let* $F : X \rightrightarrows Y$, *and* $(\bar{x}, \bar{y}) \in (\text{Graph } F)$. *Then* F *is regular at* $(\bar{x}, \bar{y})$ *if and only if it is graph-regular at*

$(\bar{x}, \bar{y})$. *Moreover,* reg $F(\overline{x}|\overline{y})$ *is the lower bound of* $K > 0$ *for which (2.2.1) holds under a suitable choice of* $\varepsilon > 0$.

Proof. Suppose F is regular near $(\bar{x}, \bar{y})$ and reg $F(\overline{x}|\overline{y}) < K$, that is to say, there is an $\varepsilon > 0$ such that $d(x, F^{-1}(y)) \le Kd(y, F(x))$ if $d(x, \overline{x}) < \varepsilon$ and $d(y, \overline{y}) < \varepsilon$. Let $\delta > 0$ be so small that

$$d_{1,K}((x, y), \text{Graph } F) = \inf\{d(x, u) + Kd(y, v) : \\ d(u, \overline{x}) < \varepsilon,\ d(v, \overline{y}) < \varepsilon, v \in F(u)\} \tag{2.2.2}$$

if $d(x, \overline{x}) < \delta,\ d(y, \overline{y}) < \delta$. As $\overline{y} \in F(\overline{x})$ such a δ exists. Indeed, let $2(1+K)^2\delta < \varepsilon$. If $d(x, \overline{x}) < \delta$ and $d(y, \overline{y}) < \delta$, then $d_{1,K}((x, y), (\bar{x}, \bar{y})) < (1 + K)\delta$ and any $(u, v) \in$ Graph F with $d_{1,K}((x, y), (u, v)) \le d_{1,K}((x, y), (\bar{x}, \bar{y})) < (1 + K)\delta$ satisfy $d_{1,K}((u, v), (\bar{x}, \bar{y})) < 2(1 + K)\delta \le (1 + K)^{-1}\varepsilon$, hence $d(u, \overline{x}) < \varepsilon$ and $d(v, \overline{y}) < \varepsilon$.

Thus for any such (x, y) and any $(u, v) \in$ Graph F satisfying (2.2.2) we have

$$\begin{aligned} d(x, F^{-1}(y)) &\le d(u, x) + d(u, F^{-1}(y)) \\ &\le d(u, x) + Kd(y, F(u)) \le d(u, x) + Kd(y, v). \end{aligned}$$

This inequality holds for any $(u, v) \in$ Graph $F \times B((\bar{x}, \bar{y}), \varepsilon)$, so applying (2.2.2), we conclude the proof of graph regularity of F.

Conversely, if (2.2.1) holds then the last inequality is valid and we prove the metric regularity of F by setting $u = x$ and taking the infimum over $v \in F(x)$. □

2.2.2 General (Non-local) Case: Definitions and Discussion

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

Definition 2.21 (*linear openness on* (U, V)). F is said to be γ*-open (or* γ*-covering) at a linear rate* on (U, V) if there is an $r > 0$ such that

$$B(F(x), rt) \bigcap V \subset F(B(x, t)),$$

if $x \in U$ and $t < \gamma(x)$.

In other words, F is γ-open at a linear rate on (U, V) if the inclusion

$$B(v, rt) \bigcap V \subset F(B(x, t))$$

holds whenever $(x, v) \in$ Graph F, $x \in U$, and $t < \gamma(x)$. Denote by $\text{sur}_\gamma F(U|V)$ the upper bound of such r. If no such r exists, we set $\text{sur}_\gamma F(U|V) = 0$. We shall call $\text{sur}_\gamma F(U|V)$ the *rate* (or *modulus*) *of* γ*-surjection* of F on (U, V).

Note that as in the local case, here the "closed" balls $B(F(x), rt)$ can be replaced by their open counterparts $\overset{\circ}{B}(F(x), rt)$ and we can use any of the options, whichever convenient.

Definition 2.22 (*metric regularity on* (U, V)). F is said to be *γ-metrically regular on* (U, V) if there is a $K > 0$ such that

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

provided $x \in U$, $y \in V$ and $Kd(y, F(x)) < \gamma(x)$. Denote by $\mathrm{reg}_\gamma F(U|V)$ the lower bound of such K. If no such K exists, set $\mathrm{reg}_\gamma F = \infty$. We shall call $\mathrm{reg}_\gamma F(U|V)$ the *modulus* (or *rate*) *of γ-metric regularity* of F on (U, V).

Definition 2.23 (*the pseudo-Lipschitz property on* (U, V)). F is said to have the *δ-pseudo-Lipschitz property* on (U, V) if there is a $K > 0$ such that

$$d(y, F(x)) \leq Kd(x, u)$$

if $x \in U$, $y \in V$, $Kd(x, u) < \delta(y)$ and $y \in F(u)$. Denote by $\mathrm{lip}_\delta F(U|V)$ the lower bound of such K. If no such K exists, set $\mathrm{lip}_\delta F = \infty$. We shall call $\mathrm{lip}_\delta F$ the *δ-Lipschitz modulus* of F on (U, V).

We should say at this point that we will not work with this general definition often in this book. In the non-local context, the most attention will be paid to what will be defined as "Milyutin regularity" in Definition 2.28 later in this subsection, which corresponds to a special choice of γ.

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. So it is natural to call them *regularity horizon functions*. If we look back to Proposition 2.11 and Exercise 2.12 we see that such functions are not needed for local properties. This is because (see the proof of Proposition 2.11) we can freely change a neighborhood of $(\bar{x}, \bar{y})$ in the course of verification.

In the case of non-local regularity on a fixed set we do not have such flexibility, so that for fixed U and V a regularity horizon function is an essential element of the definition. Indeed, we shall see later in Example 2.32 that regularity properties corresponding to different γ may fail to be equivalent.

The following observations add valuable information about the concepts by emphasizing a subtle but significant difference between the last definitions and the local Definition 2.8.

Remark 2.24. (a) Observe that in the definition of linear openness we do not require the sets appearing in the left terms of the inclusions to be nonempty for some or even for all $x \in U$ and $t < \gamma(x)$. Likewise, the definition of metric regularity does not exclude the possibility that the inequality $Kd(y, F(x)) < \gamma(x)$ is not satisfied for all $(x, y) \in U \times V$ and the definition of the pseudo-Lipschitz property allows the possibility that for no $(x, y) \in U \times V$ is there a u such that $y \in F(u)$ and

$Kd(x,u) < \delta(y)$. This is similar to the inclusion of points not belonging to the image of the mapping in the classical definition of a regular value.

(b) According to a general convention, the distance to the empty set is equal to infinity. Therefore the definition of metric regularity implies that $F^{-1}(y) \neq \emptyset$ if $d(y, F(x)) < \gamma(x)$ for some $x \in U$. Likewise the definition of the pseudo-Lipschitz property contains an implicit statement that $F(x) \neq \emptyset$ for any $x \in U$ satisfying $d(x,u) < \rho = \sup_{y\in V} \delta(y)$ whenever $u \in U$ is such that $F(u) \bigcap V \neq \emptyset$. In particular, if for some $\overline{x} \in U$ the set $F(\overline{x})$ meets V and $U \subset \overset{\circ}{B}(\overline{x}, \rho)$, then $F(x) \neq \emptyset$ for all $x \in U$.

Concerning the Lipschitz modulus, we note that in the case when F is single-valued and $\delta(y) \equiv \infty$, the Lipschitz modulus coincides with the Lipschitz constant of F on U, so it is legitimate to use the same notation for these two quantities.

The equivalence theorem we are going to prove next is one of the principal results of regularity theory. With all its simplicity, it is a remarkable and fundamental fact that underscores the metric nature of the three regularity phenomena defined above.

Theorem 2.25 (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, V);

(b) F is γ-metrically regular on (U, V);

(c) F^{-1} has γ-pseudo-Lipschitz property on (V, U).

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

$$\operatorname{sur}_\gamma F(U|V) \cdot \operatorname{reg}_\gamma F(U|V) = 1, \quad \operatorname{reg}_\gamma F(U|V) = \operatorname{lip}_\gamma F^{-1}(V|U). \tag{2.2.3}$$

Proof. The implication (b) $\Rightarrow$ (c) is trivial. Hence $\operatorname{lip}_\gamma F^{-1} \leq \operatorname{reg}_\gamma F$. To prove that (c) $\Rightarrow$ (a), assume first that $\operatorname{lip}_\gamma F^{-1} < \infty$ and take a $K > \operatorname{lip}_\gamma F^{-1}$ and an $r < K^{-1}$. Let further $t < \gamma(x)$, $x \in U$, $y \in V$, $v \in F(x)$ and $y \in B(v, tr)$. Then $d(y,v) < r\gamma(x)$ and by (c) $d(x, F^{-1}(y)) \leq Kd(y,v) < r^{-1}d(y,v) \leq t$. It follows that there is a u such that $y \in F(u)$ and $d(x,u) < t$. Hence $y \in F(B(x,t))$. Then $r \leq \operatorname{sur}_\gamma F(U|V)$, or equivalently $1 \leq K \operatorname{sur}_\gamma F(U|V)$. But r can be chosen arbitrarily close to K^{-1} and K can be chosen arbitrarily close to $\operatorname{lip}_\gamma F^{-1}(V|U)$. So we conclude that $\operatorname{sur}_\gamma F(U|V) \cdot \operatorname{lip}_\gamma F^{-1}(V|U) \geq 1$. In view of our convention, the inequality is all the more valid if $\operatorname{lip}_\gamma F^{-1}(V|U) = \infty$.

It follows that the first equality in (2.2.3) automatically holds if $\operatorname{sur} F(U|V) = 0$. Assume now that (a) holds with some $r > 0$, let $x \in U$, $y \in V$, and let $d(y, F(x)) < \gamma(x)$. Choose a $v \in F(x)$ such that $d(y,v) < r\gamma(x)$ and set $t = d(y,v)/r$. By (a) there is a $u \in F^{-1}(y)$ such that $d(x,u) \leq t$. Thus $d(x, F^{-1}(y)) \leq t = d(y,v)/r$. But $d(y,v)$ can be chosen arbitrarily close to $d(y, F(x))$ and we get $d(x, F^{-1}(y)) \leq r^{-1}d(y, F(x))$, that is, $r \cdot \operatorname{reg}_\gamma F \leq 1$. On the other hand r can be chosen arbitrarily close to $\operatorname{sur}_\gamma F(U|V)$ and we can conclude that $\operatorname{sur}_\gamma F(U|V) \cdot \operatorname{reg}_\gamma F(U|V) \leq 1$ so that

$$1 \geq \operatorname{sur}_\gamma F(U|V) \cdot \operatorname{reg}_\gamma F(U|V) \geq \operatorname{sur}_\gamma F(U|V) \cdot \operatorname{lip}_\gamma F(V|U) \geq 1,$$

which completes the proof of the theorem. □

Local equivalence (Proposition 2.10) is now immediate: enough to apply the theorem to $U = \overset{\circ}{B}(\bar{x}, \varepsilon)$, $V = \overset{\circ}{B}(\bar{y}, \varepsilon)$, $\gamma(x) \equiv \infty$. The theorem justifies the following definition.

Definition 2.26 (*regularity*). We say that the set-valued mapping $F : X \rightrightarrows Y$ is γ*-regular* on (U, V) if the three equivalent properties of Theorem 2.25 are satisfied. If $V = Y$ we shall speak about γ*-regularity on* U and usually write $\operatorname{sur}_\gamma F(U)$ rather than $\operatorname{sur}_\gamma F(U|Y)$. If γ is inessential (which as we shall see is the case for local regularity) or is clear from the context, we shall omit the prefix γ and just speak about regularity of F. Finally, we say that F is *globally regular* if it is regular on $\operatorname{dom} F \times Y$ with $\gamma \equiv \infty$.

Furthermore, we shall say that $y \in Y$ is a *regular value* of F if either $y \notin F(x)$ for all x or F is regular near (x, y) whenever $y \in F(x)$ (Fig. 2.1). If y is not a regular value of F, we say that it is a *critical value* of the mapping. Likewise, we say that x is a *regular point* of F if F is regular near (x, y), $y \in F(x)$. Otherwise we say that x is a *critical point* of F.

Now we can have a closer look at the details of the definitions. As a rule, we choose open U and V but neither the definition nor the equivalence theorem need such an assumption. In certain cases it is convenient to drop openness and take closed or arbitrary sets as U and V. Next we mention the following elementary fact.

Proposition 2.27 (monotonicity of regularity). *The function* $(\gamma, U, V) \mapsto \operatorname{sur}_\gamma F(U|V)$ *is non-increasing. Specifically, if* $U' \subset U$, $V' \subset V$ *and* $\gamma'(x) \leq \gamma(x)$ *for all* x*, then*

$$\operatorname{sur}_{\gamma'} F(U'|V') \geq \operatorname{sur}_\gamma F(U|V).$$

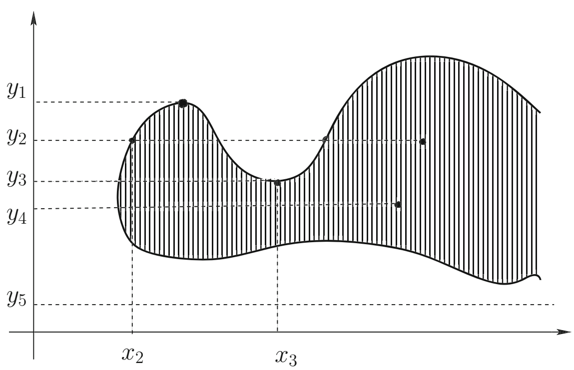

Fig. 2.1 Regular (y_2, y_4, y_5) and critical (y_1, y_3) values; regular (x_2) and critical (x_3) points

Thus ∞-regularity (that is, γ-regularity with $\gamma(x) \equiv \infty$) on (U, V) implies regularity for any γ and we again may simplify notation by omitting the subscript and write

$$\text{sur } F(U|V), \qquad \text{reg } F(U|V), \qquad \text{lip } F(U|V)$$

instead of $\text{sur}_{\infty} F(U|V)$ etc. (Of course, this type of "universal" regularity corresponds to the absence of any constraint on t in the definition of linear openness or on $d(y, F(x))$ and $d(x, u)$ in the definitions of metric regularity and the pseudo-Lipschitz property.) We shall also say in this case that F is *regular on* (U, V), without mention of the horizon function.

The type of regularity defined below plays a central role in the sequel. It is associated with a regularity horizon function that can reasonably be viewed as the smallest.

Definition 2.28 (*Milyutin regularity*). Set

$$m(x) = d(x, X \backslash U).$$

We shall say that F is *Milyutin regular* on (U, V) if it is γ-regular on (U, V) with $\gamma(x) = m(x)$. We shall denote the corresponding regularity rates (moduli) by

$$\text{sur}_m F(U|V), \quad \text{reg}_m F(U|V), \quad \text{lip } F_m^{-1}(V|U).$$

If $V = Y$, we say that F is *Milyutin regular on* U and denote the corresponding rate of regularity by $\text{sur}_m F(U)$.

Milyutin regularity on (U, V) with $\text{sur}_m F(U|V) = r$ implies that for any $x \in U$ and $y \in V$ with $d(y, F(x)) < rm(x)$ there is a $u \in U$ with $y \in F(u)$. This in turn means that when dealing with Milyutin regularity we do not need to look at points outside U (e.g. for computing $d(x, F^{-1}(y))$), in particular in the assumptions concerning the behavior of F. For that reason we essentially work with Milyutin regularity in the non-local context. The proposition below shows that Milyutin regularity is sufficient to adequately express local regularity as well.

Proposition 2.29. *F is regular near $(\bar{x}, \bar{y}) \in \text{Graph } F$ with $\text{sur } F(\overline{x}|\overline{y}) > r$ if and only if there is an $\varepsilon > 0$ such that F is Milyutin regular on $(\mathring{B}(\overline{x}, \varepsilon), \mathring{B}(\overline{y}, \varepsilon))$ with $\text{sur}_m(\mathring{B}(\overline{x}, \varepsilon)|\mathring{B}(\overline{y}, \varepsilon)) > r$.*

Proof. This is an immediate consequence of the monotonicity property. Indeed, let F be regular near $(\bar{x}, \bar{y}) \in \text{Graph } F$, that is, for some $\varepsilon > 0$

$$B(F(x), rt) \cap \mathring{B}(\overline{y}, \varepsilon) \subset F(B(x, t))$$

if $d(x, \overline{x}) < \varepsilon,\ t > 0$. But then the inclusion is all the more valid if $t < \varepsilon - d(x, \overline{x})$, which means Milyutin regularity on $(\mathring{B}(\overline{x}, \varepsilon), \mathring{B}(\overline{y}, \varepsilon))$.

Conversely, if F is Milyutin regular on $(\overset{\circ}{B}(\overline{x}, \varepsilon), \overset{\circ}{B}(\overline{y}, \varepsilon))$ for some $\varepsilon > 0$, taking $\delta < \varepsilon/2$, we see that

$$B(y, rt) \cap \overset{\circ}{B}(\overline{y}, \delta) \subset B(F(x), rt) \cap \overset{\circ}{B}(\overline{y}, \varepsilon) \subset F(B(x, t))$$

for $(x, y) \in (\text{Graph } F) \cap (\overset{\circ}{B}(\overline{x}, \delta) \times \overset{\circ}{B}(\overline{y}, \delta))$ and $t < \delta$, and we have $m(x) > \delta$, which by Proposition 2.11 means that F is regular near $(\bar{x}, \bar{y})$. □

As immediate consequences we get the following two corollaries:

Corollary 2.30 (Milyutin regularity implies local regularity). *If F is Milyutin regular on (U, V), then it is regular near any $(u, v) \in (\text{Graph } F) \cap (U \times V)$ with* $\text{sur}\, F(x|v) \geq \text{sur}_m F(U|V)$,

and (see Exercise 2.12)

Corollary 2.31. *F is regular near $(\bar{x}, \bar{y}) \in \text{Graph } F$ if and only if there is an $\varepsilon > 0$ such that the inequality in the definition of metric regularity (e.g. Definition 2.8) holds whenever x and y are ε-close respectively to $\overline{x}$ and $\overline{y}$ and in addition $d(y, F(x)) < \varepsilon$.*

A comparison between ∞-regularity and Milyutin regularity shows that the regularity property corresponding to different horizon functions are not equivalent. Indeed, ∞-regularity of F on (U, V) implies that $V \subset F(X)$, but for Milyutin regularity the inclusion is not necessary. Next we present a slightly more sophisticated example.

Example 2.32. Let $X = Y = R^2$. We consider both spaces with the standard Euclidean metric but represent all the vectors in polar coordinates, which we write with square brackets, e.g. $[\rho, \varphi]$. Consider the mapping $F : [r, \varphi] \mapsto [r, k\varphi]$ from $I\!R^2$ into itself (with $F([r, \pm\pi]) = \{[r, k\pi], [r, -k\pi]\}$) where $k \in (1, 2)$. Let further $U = \{x = [r, \varphi] \in I\!R^2 : \ r < \overline{r}, \ |\varphi| < \pi/2\}$ be the open right semi-disc of radius $\overline{r}$, and set $V = F(U) = \{[r, \varphi] : \ r < \overline{r}, \ |\varphi| < k(\pi/2)\}$. It is an easy matter to see that F is Milyutin regular on (U, V) with all the rates of surjection equal to 1.

But with say $\gamma(x) \equiv (3/2)\overline{r}$ and k sufficiently close to 2 (e.g. $k > 11/6$) the rate of surjection is substantially smaller. Indeed, let $\varepsilon = (\pi/2)(2-k)$, and let $2\varepsilon < \pi/6$. Take an $x = [\rho, \beta] \in U$ so close to $(\overline{r}, \pi/2)$ that for $u = (\rho, -\beta)$, $y = F(x)$ and $v = F(u)$ we have $\|u-x\| > (3/2)\overline{r}$, $\|y-v\| < (1/2)\overline{r}$. Then the open ball of radius $(1/2)\overline{r}$ around y contains v but on the other hand, v is not in $F(x, t)$ if $t \leq (3/2)\overline{r}$. This means that $\text{sur}_\gamma(U|V)$ cannot be greater than one third.

We conclude the subsection with a proposition that offers still another equivalence property, maybe not as fundamental as that of Theorem 2.25 but often very useful. It says that, as far as the regularity properties are concerned, every set-valued mapping can be equivalently replaced by a single-valued mapping canonically associated with F, namely the mapping $\mathcal{P}_F : \text{Graph } F \to Y$, which is the restriction to Graph F of the Cartesian projection $(x, y) \to y$.

Proposition 2.33 (single-valued reduction). *Let $X \times Y$ be endowed with the ξ-metric. Let F be γ-regular on (U, V) with* $\mathrm{sur}_\gamma F(U|V) \geq r > 0$. *Set* $\gamma'(x, v) = \min\{1, (\xi r)^{-1}\}\gamma(x)$. *Then $\mathcal{P}_F$ is γ'-regular on $(U \times Y) \times V$ and*

$$\mathrm{sur}_{\gamma'} \mathcal{P}_F(U \times V|V) = \min\{\mathrm{sur}_\gamma F(U|V), \frac{1}{\xi}\}.$$

Thus, $\mathrm{sur}_\gamma \mathcal{P}_F(U \times V|V) = \mathrm{sur}_\gamma F(U|V)$ *if* $\xi \cdot \mathrm{sur}_\gamma F(U|V) \leq 1$.

Proof. Indeed, $y \in \mathcal{P}_F(B((x, v), t))$ means that there is a $u \in X$ such that $y \in F(u)$ and $d_\xi((x, v), (u, y)) \leq t$, which means that $d(x, u) \leq t$ and $d(y, v) \leq t/\xi$. On the other hand, by the assumption $y \in F(B(x, t))$ if $y \in V$ and $d(y, v) \leq rt$ for some $v \in F(x)$. Finally, for any t the relations $t < \gamma(x)$ and $\min\{1, (\xi r)^{-1}\}t < \gamma'(x, v)$ are equivalent. Therefore if $y \in V$, $(x, v) \in (U \times Y) \bigcap \mathrm{Graph}\ F$ and $d(y, v) \leq \min\{1, (\xi r)^{-1}\}t < \gamma'(x, v)$, there is a $u \in X$ with $y \in F(u)$ and $d_\xi((x, v), (u, v)) = d(x, u) \leq t$ and the result follows. □

The local version of the result, most needed in the sequel, reads as follows.

Proposition 2.34 (single-valued reduction – local version). *Let $F : X \rightrightarrows Y$ be regular at $(\bar{x}, \bar{y}) \in$ Graph F with* $\mathrm{sur}\ F(\bar{x}|\bar{y}) \geq r > 0$. *Then $\mathcal{P}_F$ is regular near $((\bar{x}, \bar{y})|\bar{y})$ and, moreover,* $\mathrm{sur}\,\mathcal{P}_F((\bar{x}, \bar{y})|\bar{y}) \geq r$, *if $X \times Y$ is endowed with the ξ-metric with $\xi r < 1$.*

2.2.3 Restricted Regularity

In this subsection we briefly discuss an alternative way to define regularity properties on fixed sets. It seems that a mention of such a possibility may be useful, but we shall not use this approach in what follows, so the subsection is optional and can be harmlessly omitted by the reader.

In the definitions of the three regularity properties we do not require that either $F(x) \subset V$ in Definition 2.21 or that $u \in U$ in Definition 2.23. Such requirements, however, can be included in the definitions as follows.

Definition 2.35 (*restricted regularity*). Set $F^V(x) = F(x) \bigcap V$. We define *restricted γ-openness at a linear rate* and *restricted γ-metric regularity* on (U, V) by replacing F by F^V in Definitions 2.21 and 2.22. Likewise, we define *restricted γ-pseudo Lipschitz property* by adding the requirement $u \in U$ in Definition 2.23. We can also define *restricted rates* in an obvious way.

Exercise 2.36. Prove that the equivalence theorem also holds for the restricted versions of the three properties.

The following example shows that the restricted versions of the regularity properties may be strictly weaker, up to the extent that F^V can be ∞-regular while F itself is not regular for any choice of the regularity horizon function.

Example 2.37. Let $X = Y = I\!R$, $U = (0, 1)$, $V = (0, 2)$, $F(x) = \{x, 2\}$. Then $F^V(x) = x$ if $0 \le x \le 2$ and $F^V(x) = \emptyset$ otherwise, and we see that F^V is γ-regular on (U, V) for any γ with the rate of surjection is equal to 1. On the other hand F itself is not regular on (U, V) for any γ. Indeed, if $x \in U$, $t > 0$ and $x + t < 1$, then $(2-rt, 2) \subset B(F(x), rt) \cap V$ for any $r > 0$ while $F(B(x, t)) = (x-t, x+t) \cup \{2\}$.

Thus, in principle, it is reasonable to deal with F^V rather than F itself, but for technical convenience we shall not do this. Nevertheless, we will keep in mind that the regularity assumptions can always be harmlessly weakened to restricted regularity with the same regularity horizon function.

Another idea that may occur is that perhaps it is also reasonable to deal with the restriction of F to (the closure of) U. However, this does not work: restricting the mapping to U can just kill regularity.

Example 2.38. Let $X = I\!R^2$ and

$$F(x) = \begin{cases} (1 - \frac{1}{2}\|x\|)\frac{x}{\|x\|}, & \text{if } x \neq 0 \\ S, & \text{if } x = 0. \end{cases}$$

Here S is the unit sphere in $I\!R^2$. Let further $U = V = \overset{\circ}{B}$, the open unit ball. Then F is γ-regular on (U, V) with e.g. $\gamma = \text{const} < 1/4$. But $F|_U$ is no longer γ-regular on (U, V) as the F-image of U does not contain points of $(1/2)\overset{\circ}{B}$.

Exercise 2.39. Show that the δ-pseudo-Lipschitz property on (U, V) can be equivalently expressed by the relation

$$\text{ex}(F(u)\cap V, F(x)) \le Kd(x, u), \quad \forall x \in U,\ u \in F^{-1}(V),\ d(x, u) < \delta(y),\ \forall y \in F(u)\cap V,$$

while the restricted δ-pseudo-Lipschitz property by the relation

$$\text{ex}(F(u) \cap V, F(x)) \le Kd(x, u), \quad \forall x \in U,\ u \in U,\ d(x, u) < \delta(y). \forall y \in F(u) \cap V.$$

Note finally that while the values of the rates in Definition 2.21 are determined by the specific metrics in the domain spaces, the very property of F being regular does not depend on the specific choice of equivalent metrics.

2.2.4 Regularity and Completeness

So far we have not imposed any additional constraints on the spaces and on the mapping. However, as we shall see in the next section, all efficient regularity criteria can be proved only under certain completeness requirements. Therefore it is reasonable to look in advance at the interplay between the regularity and completeness properties.

Proposition 2.40 (completeness requirements). *Let X and Y be metric spaces, let $F : X \rightrightarrows Y$ be a closed set-valued map, let $U \subset X$ and $V \subset Y$ be open sets, and let γ be positive and continuous on U. Denote by i_X and i_Y the natural imbedding $X \to \hat{X}$ and $Y \to \hat{Y}$ and set $i = i_X \times i_Y$. Let further:*

- *$\hat{X}$, $\hat{Y}$ be the completions of X and Y;*
- *the graph of $\hat{F} : \hat{X} \rightrightarrows \hat{Y}$ be the closure of $i(\text{Graph } F)$ in $\hat{X} \times \hat{Y}$;*
- *$\hat{U} = \text{int }(\text{cl } i_X(U))$, $\hat{V} = \text{int }(\text{cl } i_Y(V))$.*

If now F is Milyutin regular on (U, V), then $\hat{F}$ is Milyutin-regular on $\hat{U} \times \hat{V}$ with $\text{sur}_m \hat{F}(\hat{U}|\hat{V}) = \text{sur}_m F(U|V)$. The opposite implication is also valid if either

(a) Graph *F is complete*

or

(b) X is a complete space.

Proof. Let $\text{sur}_m(U|V) > r > 0$. Set as before $m(x) = d(x, X \backslash U)$ and also $\hat{m}(x) = d(\hat{x}, \hat{X} \backslash \hat{U})$. Clearly $\hat{m}(i_X(x)) = m(x)$ for $x \in U$. Let now $\hat{u} \in \hat{U}$, $\hat{v} \in \hat{F}(\hat{u})$, $t < \hat{m}(\hat{u})$ be such that $\hat{y} \in \overset{\circ}{B}(\hat{v}, rt) \cap \hat{V}$. Fix an $\varepsilon > 0$ to make sure that

$$\varepsilon < t, \quad (1+\varepsilon)t < \hat{m}(\hat{u}), \quad d(\hat{y}, \hat{v}) + 2\varepsilon rt < rt$$

and choose a sequence $(y_n) \subset V$ such that $d(i_Y(y_n), \hat{y}) < 2^{-(n+1)}\varepsilon rt,\ n = 1, 2, \ldots$. We may assume without loss of generality that $d(i_Y(y_n), \hat{v}) < (r - 2\varepsilon)t$ for all n.

Let further $(u, v) \in$ Graph F satisfy $d(i_X(u), \hat{u}) < \varepsilon t,\ d(i_Y(v), \hat{v}) < \varepsilon rt$. Then $t < m(u)$ and

$$d(y_n, v) = d(i_Y(y_n), i_Y(v)) \le d(i_Y(y_n), \hat{v}) + d(\hat{v}, i_Y(v)) < (1-\varepsilon)rt.$$

By regularity of F there is an x_1 such that $y_1 \in F(x_1)$ and $d(x_1, u) \le r^{-1}d(y_1, v) < (1-\varepsilon)t$. It follows that $m(x_1) \ge m(u) - d(x_1, u) > \varepsilon t$.

As

$$d(y_{n+1}, y_n) \le d(i_Y(y_{n+1}), \hat{y}) + d(\hat{y}, i_Y(y_n)) \le \frac{3}{4}2^{-n}\varepsilon t < 2^{-n}\varepsilon rt,$$

we can, again by regularity of F, find a sequence (x_n) such that $y_n \in F(x_n)$ and

$$d(x_{n+1}, x_n) \le 2^{-n}\varepsilon t.$$

Indeed, suppose we have already found $x_k,\ k = 1, \ldots n$. Then

$$m(x_n) \ge m(x_1) - d(x_1, x_2) - \ldots - d(x_{n-1}, x_n) \ge 2^{n-1}\varepsilon t > r^{-1}d(y_{n+1}, y_n)$$

and consequently there is an x_{n+1} such that $y_{n+1} \in F(x_{n+1})$ and

$$d(x_{n+1}, x_n) \le r^{-1}d(y_{n+1}, y_n).$$

This means that (x_n) is also a Cauchy sequence, so $i_Y(x_n)$ converges to some $\hat{x}$ such that $d(\hat{x}, \hat{u}) \le (1+\varepsilon)t$ and as the graph of $\hat{F}$ is closed, $\hat{y} \in F(\hat{x})$. As t could have been chosen (greater but) arbitrarily close to $r^{-1}d(\hat{y}, \hat{v})$, we can be sure that e.g. $d(\hat{x}, \hat{u}) < r^{-1}d(\hat{y}, \hat{v})$. This completes the proof of the first statement as ε can be chosen arbitrarily small.

To prove the second statement, we first note that it is obvious in the case of (b) because if X is complete and $i_Y(y) \in \hat{F}(x)$ for some $y \in Y$, then necessarily $y \in F(x)$ as the graph of F is closed. On the other hand, (a) reduces to (b) in view of Proposition 2.33. □

Reformulation of the proposition for local regularity is immediate (in view of Exercise 2.12).

Corollary 2.41 (completeness requirements: local regularity). *Let X and Y be metric spaces, let $F : X \rightrightarrows Y$ be a closed set-valued map, and let $(\bar{x}, \bar{y}) \in$* Graph *F. If F is regular at $(\bar{x}, \bar{y})$ with* sur $(\overline{x}|\overline{y}) > r$*, then $\hat{F}$ is also regular near $(i_X(\overline{x}), i_Y(\overline{y}))$ with the same rate of surjection. The opposite implication is also valid if either*

(a) Graph *F is complete or*

(b) X is a complete space.

In view of the proposition, we will essentially work with the two types of completeness assumptions specified in the statements. The proposition in principle allows us to prove regularity under the assumption that both spaces and the graph of the mapping are complete, and then obtain regularity under only (a) or (b) as a consequence. However, this requires some caution for it may happen that, unlike the regularity properties, some of the assumptions do not translate to the completions of the spaces or/and the graph or vice versa.

2.3 General Regularity Criteria

This section is central. Here we prove necessary and sufficient conditions for regularity. The key results of the section are several theorems containing general criteria (necessary and sufficient conditions for regularity) which, unlike the equivalence theorems of the previous section, already have an algorithmic flavor. The criteria will serve as a basis for obtaining various qualitative and quantitative characterizations of regularity in this and subsequent chapters. The proofs of the criteria based on Ekeland's principle are rather simple. At the same time the criteria provide us with an instrument of analysis which is both powerful and easy to use. We shall demonstrate how they work already in this section by proving the density theorem, which reveals a rather surprising circumstance that the covering property automatically holds if we assume that images of balls in the domain space are only sufficiently dense in the corresponding balls in the target space. At the end of the section we introduce and briefly discuss the concepts of error bound and metric subregularity, to be considered in more detail in the subsequent chapters.

2.3.1 The Basic Lemma and Error Bounds

We begin the discussion of regularity criteria with the following fundamental fact following from Ekeland's variational principle. Given an open $U \subset X$, and a function γ positive on U, we set

$$U_\gamma = \bigcup_{x \in U} \overset{\circ}{B}(x, \gamma(x)).$$

The following lemma plays a key role in proofs of all regularity criteria that will be proved in this section and later on.

Lemma 2.42 (Basic Lemma). *Let X be a complete metric space, let $U \subset X$ be an open set, let f be a lower semicontinuous function, let $r > 0$, and let $\gamma(\cdot)$ be a nonnegative function on X with Lipschitz constant not greater than 1 which is positive on U. Suppose that for any $x \in U_\gamma$ satisfying $0 < f(x) < r\gamma(x)$ there is a $u \neq x$ such that*

$$f(u) \le f(x) - rd(u, x). \tag{2.3.1}$$

Then for any $\overline{x} \in U$ such that $f(\overline{x}) < r\gamma(\overline{x})$ there is a $\overline{u}$ such that $f(\overline{u}) \le 0$ and $d(\overline{u}, \overline{x}) \le f(\overline{x})/r$.

Proof. With no loss of generality we may assume that $f(x) \ge 0$ for all x (replacing, if necessary, f by $\max\{f(x), 0\}$). Take an $\overline{x}$ satisfying the conditions. Set $\varepsilon = f(\overline{x})$. Applying Ekeland's variational principle to f we shall find a $\overline{u}$ satisfying

$$d(\overline{u}, \overline{x}) \le \frac{\varepsilon}{r} = \frac{f(\overline{x})}{r} < \gamma(\overline{x}), \quad f(\overline{u}) \le f(\overline{x}) - rd(\overline{u}, \overline{x})$$

and such that for any $x \neq \overline{u}$

$$f(x) + rd(x, \overline{u}) > f(\overline{u}). \tag{2.3.2}$$

Then $\overline{u} \in U_\gamma$ and

$$f(\overline{u}) \le f(\overline{x}) - rd(\overline{u}, \overline{x}) < r\gamma(\overline{x}) - rd(\overline{u}, \overline{x}) \le r\gamma(\overline{u}).$$

So if $f(\overline{u}) > 0$, we would be able to find an x satisfying (2.3.1) (with u replaced by $\overline{u}$). This, however, would contradict (2.3.2). Thus $f(\overline{u}) = 0$. ☐

Remark 2.43. Note that the lemma and its proof remain valid if we assume that (2.3.1) holds not for all $x \in U_\gamma$ satisfying $0 < f(x) < r\gamma(x)$ but only for x which in addition satisfy $f(x) < f(\overline{x})$.

If f is not lower semicontinuous, the conclusion of the lemma does not hold even if we assume that the set $\{x : f(x) \le 0\}$ is nonempty and closed and any sequence (x_n) with $f(x_n) \to 0$ converges to some $x \in [f \le 0]$. As a simple example, consider a function $f(t)$ on $[-1/2, 1/2]$ with the following properties

- $t^2 < f(t) \le 2t^2$;
- $f(t) = f(-t)$;
- there is a sequence of $t_n \searrow 0$ such that $2t_n^2 - t_{n+1}^2 = t_n - t_{n+1}$, $f(t_n) = 2t_n^2$ and $f(t) = f(t_n) - (t_n - t)$ for $t \in (t_{n+1}, t_n]$.

Take $\gamma(t) \equiv 1$. It is an easy matter to see that all conditions of the lemma, except lower semicontinuity of f, are satisfied but the distance estimate in the conclusion fails to be valid for any r.

The first immediate consequence of the Basic Lemma is a characterization theorem for error bounds. Recall that given a function f on X and an $\alpha \in I\!R$, we denote by $[f \le \alpha] = \{x \in X : f(x) \le \alpha\}$ the corresponding sublevel set of f. The meaning of the notation $[f < \alpha]$ and $[f = \alpha]$ is also obvious. By an *error bound* for f (at level α) we mean any estimate for the distance to $[f \le \alpha]$ in terms of $(f(x) - \alpha)^+$. We shall be mainly interested in estimates of the form

$$d(x, [f \le \alpha]) \le K(f(x) - \alpha)^+ \tag{2.3.3}$$

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

We speak about a *local error bound* at some $\overline{x}$ with $f(\overline{x}) = \alpha$ if (2.3.3) holds for all x in a neighborhood of $\overline{x}$ or, more generally, about an error bound on some set $U \subset X$ if (2.3.3) holds for all $x \in U$ or finally about a *global error bound* when (2.3.3) is satisfied for all $x \in X$. In applications we usually deal with $\alpha = 0$. Occasionally we somewhat abuse the language and call the number K itself an error bound of f.

Theorem 2.44 (characterization theorem for error bounds). *Let f be a lower semicontinuous function on X which is finite at $\overline{x}$. Then $K > 0$ is a local error bound for f at $\overline{x}$ if and only if there is a $\delta > 0$ such that for any x with $d(\overline{x}, x) < \delta$ and $f(x) > f(\overline{x})$ and any $r < K^{-1}$ the inequality*

$$f(u) \le f(x) - rd(u, x) \tag{2.3.4}$$

holds for some $u \ne x$.

Proof. If K is an error bound for f at $\overline{x}$, that is, (2.3.3) holds with $\alpha = f(\overline{x})$ for all x in a neighborhood V of $\overline{x}$, then for any $x \in V$ with $f(x) > f(\overline{x})$ we can find an $u \in [f \le \alpha]$ such that $d(x, u) \le r^{-1}(f(x) - \alpha)$, whence (2.3.4).

Conversely, assume that for any $x \in \overset{\circ}{B}(\overline{x}, \delta)$ and $f(x) > f(\overline{x})$ we can find a $u \ne x$ such that (2.3.4) holds. Set $\varepsilon = \delta/2$, and let $U = \overset{\circ}{B}(\overline{x}, \varepsilon)$. If for some $x \in U$ we have $f(x) - f(\overline{x}) \ge K\varepsilon$, then (2.3.3) is satisfied for x. If, on the other hand, $f(x) - f(\overline{x}) < K\varepsilon$, then we can apply Lemma 2.42 with $\gamma(\cdot) \equiv \varepsilon$ (in which case $U_\gamma = \overset{\circ}{B}(\overline{x}, \delta)$) and again get (2.3.3). $\square$

2.3.2 Main Regularity Criteria

The Basic Lemma offers a simple approach to verification of regularity. If, for instance, X is a complete space and for any y a lower semicontinuous function f_y is given whose zero sublevel set coincides with $F^{-1}(y)$, then the Basic Lemma gives an estimate of the distance to the set in terms of f_y, provided, of course, that f_y satisfies the conditions of the lemma. Alternatively, we may assume the graph of F is complete in the product metric and consider functions f_y on Graph F with $f_y(x, v) \le 0$ if and only if $v = y$. In this way we obtain a series of powerful regularity criteria, both for local and non-local regularity.

We shall work not with general functions f_y but with the three functions from the introduction section:

$$\begin{aligned}
&\varphi_y(x, v) = d(y, v) + i_{\text{Graph } F}(x, v),\\
&\psi_y(x) = d(y, F(x)),\\
&\omega_y^K(x) = d_{1,K}((x, y), \text{Graph } F).
\end{aligned}$$

This is more than sufficient for practical purposes and allows us to avoid unnecessary abstractions. To warm up, we shall prove a criterion for the Milyutin regularity of a single-valued continuous mapping whose formulation and proof are especially simple. Note that for a single-valued continuous mapping it is natural to consider Milyutin regularity only on U, that is, taking $V = Y$.

Theorem 2.45 (Milyutin regularity of a single-valued mapping). *Let X be a complete metric space, let Y be a metric space, and let $U \subset X$ be an open set. Let further $F : X \to Y$ be a single-valued mapping defined and continuous on U. Then F is Milyutin regular on U with $\text{sur}_m F(U) \ge r > 0$ if and only if for any $r' < r$, any $x \in U$ and any $y \in Y$ for which $0 < \psi_y(x) < r'm(x)$ there is a $u \ne x$ such that*

$$\psi_y(u) \le \psi_y(x) - r'd(x, u). \tag{2.3.5}$$

(Recall that $m(x) = d(x, X \setminus U)$.)

Proof. If F is Milyutin regular on U with $\text{sur}_m F(U) \ge r$ and $0 < r' < r$, then by definition, the relation $\psi_y(x) < rm(x)$ implies that there is a u such that $d(x, u) < r'^{-1}\psi_y(x)$ and $y \in F(u)$, so that (2.3.5) holds. This proves the necessity of the criterion. To prove sufficiency, we apply the Basic Lemma to $f(x) = \psi_y(x)$, which is possible if x and y satisfy the assumption of the theorem. By the lemma, there is a u satisfying $d(u, x) \le r'^{-1}\psi_y(u)$ and $\psi_y(u) = 0$, that is, $y \in F(u)$. This means that sur $F(U|V) \ge r'$ and the result follows as r' can be arbitrarily close to r. ◻

We are ready now to prove a series of general criteria for regularity of a set-valued mapping. The function ψ_y may no longer be lower semicontinuous in this case, so we have to work with other functions.

Theorem 2.46 (general regularity criterion). *Let $U \subset X$ and $V \subset Y$ be open sets, and let $F : X \rightrightarrows Y$ be a set-valued mapping whose graph is complete in the product metric. Let further $\xi > 0$, $r > 0$ and a nonnegative function $\gamma(\cdot)$ on X, positive on U, satisfy the Lipschitz condition with constant ≤ 1. Given a $y \in V$, we assume that for any pair $x \in U_\gamma$, $v \in F(x)$ with $0 < d(y, v) < r\gamma(x)$ we can find another pair $(u, w) \in \text{Graph } F$ different from (x, v) and such that*

$$\varphi_y(u, w) \leq \varphi_y(x, v) - rd_\xi((x, v), (u, w)). \tag{2.3.6}$$

Then for any $(x, v) \in \text{Graph } F$, $x \in U$ with $d(y, v) \leq rt < r\gamma(x)$, there is a $u \in B(x, t)$ such that $y \in F(u)$. In particular, F is γ-regular on (U, V) with $\text{sur}_\gamma F(U|V) \geq r$, provided the assumption of the theorem is satisfied for any $y \in V$.

Conversely, if F is γ-regular on (U, V), then for any positive $r < \text{sur}_\gamma F(U|V)$, any $\xi > 0$ such that $r\xi < 1$, any $x \in U$, $v \in F(x)$ and any $y \in V$ satisfying $0 < d(y, v) < r\gamma(x)$, there is a pair $(u, w) \in \text{Graph } F$ different from (x, v) such that (2.3.6) holds.

The theorem offers a very simple geometric interpretation of the regularity phenomenon: it means that F is regular if for any $(x, v) \in \text{Graph } F$ 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. We emphasize once again that v is not required to be in V.

Of course, (2.3.6) can be written in the form

$$d(y, w) \leq d(y, v) - rd_\xi((x, v), (u, w)) \tag{2.3.7}$$

without any mention of φ_y. We prefer, however, to use (2.3.6) in order to unify the statement of the general criterion with the the statements of uther regularity criteria, some of which use functions other than φ_y.

Proof. Consider on Graph F the function $\varphi_y(x, v) = d(y, v)$. This is a Lipschitz continuous function, hence lower semicontinuous. We apply the Basic Lemma for this function with (Graph F, d_ξ) being the domain spaces and U replaced by $(U \times Y) \bigcap \text{Graph } F$. Fix a $x \in U$ and $v \in F(x)$. Then all conditions of the lemma are satisfied for φ_y if $d(y, v) < \gamma(x)$. In this case by the lemma there is a pair $(u, w) \in \text{Graph } F$ such that $\varphi_y(u, w) = 0$, that is, $y = w \in F(u)$, and $rd(u, x) \leq d_\xi((x, v), (u, w)) \leq d(y, v) < rt$. This proves the first statement.

Conversely, let F be γ-regular on (U, V), and let $0 < r < \text{sur}_\gamma F(U|V)$. If $x \in U$ and $v \in F(x) \cap V$, then by definition there is a $u \in X$ such that $y \in F(u)$ and $rd(x, u) \leq d(y, v)$. Taking $\xi > 0$ satisfying $r\xi < 1$, we get

$$\begin{aligned} d(y, y) = 0 &\leq \max\{d(y, v) - rd(x, u), d(y, v) - r\xi d(y, v)\} \\ &= d(y, v) - r \max\{d(x, u), \xi d(y, v)\}, \end{aligned}$$

which is (2.3.6). $\square$

Note that according to the definition, F is γ-regular on (U, V) even if there is no x, y and v satisfying the conditions of the theorem. As in the classical theory when a value is regular also if it does not belong to the range of the mapping, this does not create any additional problems.

A certain inconvenience of the criterion comes from the fact that there is a gap between its necessary and sufficient conditions. This gap, however, disappears in case of Milyutin regularity (and also for local regularity – see Theorem 2.54 below), and for that reason we shall essentially work with Milyutin regularity in the non-local context.

Theorem 2.47 (first criterion for Milyutin regularity). *If the graph of F is complete, then a necessary and sufficient condition for F to be Milyutin regular on (U, V) with $\mathrm{sur}_m F \geq r$ is that there is an $\xi > 0$ such that for any $r' < r$, any $x \in U$ and $v \in F(x)$ and $y \in V$ satisfying $0 < d(y, v) < r'm(x)$ there are $(u, w) \in \mathrm{Graph}\, F$ different from (x, v) such that*

$$d(y, w) \leq d(y, v) - r' d_\xi((x, v), (u, w)). \tag{2.3.8}$$

Proof. Apply Theorem 2.46 taking into account that in this case $U_m = U$. □

To prove the second criterion for Milyutin regularity we need the following simple fact.

Lemma 2.48. *If the graph of F is closed, then $y \in F(x)$ whenever $\overline{\psi}_y(x) = 0$.*

Proof. Indeed, take a sequence $(x_n) \to x$ such that $\psi_y(x_n) \to \overline{\psi}_y(x) = 0$. This means that there are $y_n \in F(x_n)$ such that $d(y_n, y) \to 0$. As the graph of F is closed it follows that $y \in F(x)$. □

Theorem 2.49 (second criterion for Milyutin regularity). *Let X be a complete metric space, $U \subset X$ and $V \subset Y$ open sets and $F : X \rightrightarrows Y$ a set-valued mapping with closed graph. Then F is Milyutin regular on (U, V) with $\mathrm{sur}_m F(U|V) \geq r$ if and only if for any $r' < r$, $x \in U$ and any $y \in V$ with $0 < \overline{\psi}_y(x) < r'm(x)$ there is a $u \neq x$ such that*

$$\overline{\psi}_y(u) \leq \overline{\psi}_y(x) - r' d(x, u). \tag{2.3.9}$$

Proof. We start with the sufficiency part: (2.3.9) implies regularity. Let $x \in U$, $y \in V$ and $d(y, F(x)) < r'm(x)$. Then $\overline{\psi}_y(x) < r'm(x)$. As $\overline{\psi}_y$ is lsc, all conditions of the Basic Lemma are satisfied and there is a u such that $d(u, x) \leq \overline{\psi}_y(x)$ and $\overline{\psi}_y(u) = 0$. By Lemma 2.48 $y \in F(x)$, that is,

$$d(x, F^{-1}(y)) \leq d(x, u) \leq \frac{\overline{\psi}_y(x)}{r'} \leq \frac{1}{r'} d(y, F(x)).$$

To prove that (2.3.9) is necessary for Milyutin regularity take $x \in U$, $y \in V$ such that $0 < \overline{\psi}_y(x) < r'm(x)$. Take a $\rho \in (r', r)$.

Let now $x_n \to x$ be such that $d(y, F(x_n)) \to \overline{\psi}_y(x)$. We may assume that $d(y, F(x_n)) < r'm(x)$ for all n. Choose positive $\delta_n \to 0$ such that $d(y, F(x_n)) < (1+\delta_n)\overline{\psi}_y(x)$, and let t_n be defined by $\rho t_n = (1+\delta_n)\overline{\psi}_y(x)$. Then $y \in \overset{\circ}{B}(F(x_n), \rho t_n)$, $t_n < m(x_n)$ (at least for large n) since $m(\cdot)$ is continuous 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 the u_n are bounded away from x for otherwise (as Graph F is closed) we would inevitably conclude that $y \in F(x)$, which cannot happen as $\overline{\psi}_y(x) > 0$. This means that $\lambda_n = d(u_n, x_n)/d(u_n, x)$ converges to one. Thus

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

(the last inequality being eventually true as $\lambda_n \rho > r'(1+\delta_n)$ for large n) and (2.3.9) follows. ◻

The last theorem is especially convenient when ψ_y is lower semicontinuous for every $y \in V$. In particular, when F is single-valued it reduces to Theorem 2.45. Otherwise, the need for the preliminary calculation of $\overline{\psi}_y$, the lower closure of ψ_y, may cause difficulties. Surprisingly, it is possible to modify the condition of the theorem and get a statement that requires the verification of a (1.2)-like inequality for ψ rather than $\overline{\psi}$, although at the expense of some additional uniformity assumption which, however, automatically holds for regular mappings.

Theorem 2.50 (modified second criterion for Milyutin regularity). *Let X, Y, F, U and V be as in Theorem 2.49. A necessary and sufficient condition for F to be Milyutin regular on (U, V) with* sur $F(\overline{x}|\overline{y}) \ge r$ *is that there is a $\lambda \in (0, 1)$ such that for any $r' < r$, any $x \in U$ and $y \in V$ with $0 < \psi_y(x) < rm(x)$ there is a $u \ne x$ such that*

$$\psi_y(u) \le \psi_y(x) - r'd(x, u), \quad \psi_y(u) \le \lambda \psi_y(x). \tag{2.3.10}$$

Proof. Let F be Milyutin regular on (U, V) with sur $F(U|V) > r$, then given $x \in U$ and $y \in V$ with $0 < \psi_y(x) = d(y, F(x)) < r'm(x)$, we can find a u such that $y \in F(u)$ and $d(x, u) \le r'^{-1} d(y, F(x))$, so that (2.3.10) is satisfied with any $\lambda \ge 0$.

So we only need to verify that under the assumptions of the theorem, the condition of Theorem 2.49 holds. So let the conditions of the theorem be satisfied with some λ. Take $x \in U$, $y \in V$ and $0 < \alpha = \overline{\psi}_y(x)$. Let $x_n \to x$ be such that $\psi_y(x_n) = \alpha_n \to \alpha$ and for each n find a u_n such that $\psi_y(u_n) \le \lambda \alpha_n$ and $\psi_y(u_n) \le \psi_y(x_n) - r'd(x_n, u_n)$. An easy calculation shows that

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

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

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

holds for sufficiently large n. This allows us to apply Theorem 2.49 and conclude that there is a $w \in B(x, (r')^{-1})$ such that $y \in F(w)$, that is, $\mathrm{sur}_m F(U|V) \geq r'$. $\square$

Remark 2.51. As we have already mentioned (right after Definition 2.28), we do not need to look at x outside of U when dealing with Milyutin regularity. In particular, Theorems 2.47 and 2.49 remain valid if we assume that the graph of F is closed only relative to $U \times Y$ rather than on the entire $X \times Y$.

Remark 2.52. The first two criteria for Milyutin regularity may not, in principle, be equivalent. The first criterion can be obtained from the second with the help of Proposition 2.33 but whether the converse is true is not clear.

Nonetheless the very existence of various criteria adds much flexibility in applications by allowing us to choose the most convenient criterion (among those applicable) in each specific situation.

We can also use the function ω_y^K to get a criterion for Milyutin-type non-local graph regularity.

Theorem 2.53 *Let X be a complete metric space, $U \subset X$ and $V \subset Y$ open sets and $F : X \rightrightarrows Y$ a set-valued mapping with closed graph. Then the inequality*

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

holds for all $x \in U$ and $y \in V$ satisfying $0 < \omega_y^K(x) < m(x)$, provided that for any such (x, y) there is a $u \neq x$ such that

$$\omega_y^K(u) \leq \omega_y^K(x) - d(x, u). \tag{2.3.11}$$

We omit the proof: all we need is to apply the Basic Lemma to the function ω_y^K, which is possible, of course, as this function is continuous, even Lipschitz. Reformulation of the theorem for local graph regularity does not present any difficulty. Moreover, equivalence of local regularity and graph regularity together with the Basic Lemma 2.42 allows us to state one more criterion for local regularity that does not have a non-local analogue. Indeed, assume that X is a complete space, U and V are certain neighborhoods of $\bar{x}$ and $\bar{y}$ and for any $x \in U$ and $y \in V$ with $y \notin F(x)$ there is a $u \neq x$ such that (2.3.11) holds.

For any y the function ω_y^K is Lipschitz (with constant one), so we can apply to it the Basic Lemma. As obviously $[\omega_y^K \leq 0] = F^{-1}(y)$, it follows that F is regular near $(\bar{x}, \bar{y})$. A slight modification of the argument allows to use in the proof a bit better sufficient condition for local regularity with (2.3.11) replaced by

$$\omega_y^K(u) \le \omega_y^K(x) - \lambda d(x,u). \tag{2.3.12}$$

for $\lambda \in (0,1)$ arbitrarily close to 1. It is an easy matter to see, precisely as in the proofs of the regularity criteria above, that (2.3.12) is also necessary for regularity. Summarising, we get the following theorem.

Theorem 2.54 (criteria for local regularity). *Let $F : X \rightrightarrows Y$ and $(\bar{x}, \bar{y}) \in \text{Graph } F$. Then F is regular near $(\bar{x}, \bar{y})$ with* sur $F(\overline{x}|\overline{y}) \ge r > 0$ *if and only if any of the following three properties holds for $U = \overset{\circ}{B}(\overline{x}, \varepsilon)$, $V = \overset{\circ}{B}(\overline{y}, \varepsilon)$ with some $\varepsilon > 0$.*

(a) Graph F *is locally complete and there is a $\xi > 0$ such that for any $r' < r$, any $x \in U$, any $v \in F(x)$ and any $y \in V$, $y \ne v$ there is a pair $(u, w) \in$* Graph F *such that (2.3.8) holds.*

(b) X is complete and for any $r' < r$, any $x \in (\text{dom } F) \cap U$ and any $y \in V$, $y \notin F(x)$ there is a $u \ne x$ such that either (2.3.9) or (2.3.10) holds.

(c) X is a complete space and for any $\lambda \in (0,1)$, any $x \in U$ and any $y \in V$ with $y \notin F(x)$ there is a $u \ne x$ such that (2.3.12) holds.

2.3.3 A Fundamental Application – The Density Theorem

The theorem below is a highly useful and easy consequence of any of the criteria we have established.

Theorem 2.55 (density theorem). *Let $U \subset X$ and $V \subset Y$ be open sets and $F : X \rightrightarrows Y$ 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 an ℓt-net in $B(v, rt) \bigcap V$, where $0 \le \ell < r$. Then F is Milyutin regular on (U, V) and* $\text{sur}_m F \ge r - \ell$.

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

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

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

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

Since $t < m(x)$ and $v \ne w$, reference to Theorem 2.47 completes the proof. $\square$

The following result is an immediate consequence of the theorem.

Corollary 2.56. *Let X, Y, U, V and F be as in the theorem. If $F(B(x,t))$ is dense in $B(F(x), rt) \bigcap V$ for $x \in U$ and $t < m(x)$, then* $\mathrm{sur}_m F(U|V) \geq r$.

It is not clear whether we can replace in the theorem the assumption that Graph F is complete by the assumption that X is complete and Graph F is closed. But a slightly weaker fact, implying in particular that the above corollary can be extended to the latter case, can be proved.

Exercise 2.57. Prove the following modification of the theorem: *let X be a complete space, let the graph of F be closed, and let U and V be as in the theorem. Assume that for any $x \in U$ and any $t < t' < m(x)$ the set $F(B(x,t'))$ is an ℓt-net in $B(F(x), t) \cap V$. Then* $\mathrm{sur}_m F(U|V) \geq r - \ell$.
Hint: check that the result is valid if Y is also a complete space; prove that the assumption extends to the completion of Y and $\hat{F}$ defined as in the proof of Proposition 2.40.

The specification of Theorem 2.55 for local regularity at $(\bar{x}, \bar{y})$ is

Corollary 2.58 (density theorem – local version). *Suppose there are $r > 0$ and $\varepsilon > 0$ such that $F(B(x,t))$ is an ℓt-net in $B(v, rt) \cap B(\overline{y}, \varepsilon)$ whenever $d(x, \overline{x}) < \varepsilon$, $d(v, \overline{y}) < \varepsilon$, $v \in F(x)$ and $t < \varepsilon$. Then* $\mathrm{sur}\, F(\overline{x}|\overline{y}) \geq r - \ell$. *Thus if $B(v, rt) \cap B(\overline{y}, \varepsilon) \subset \mathrm{cl}\, F(B(x,t))$ for all x, v and t satisfying the above specified conditions, then $B(v, rt) \subset F(B(x,t)$ for the same set of variables.*

2.4 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 component points leads to new, weaker concepts that differ from regularity in many respects. The three most interesting, which often appear in applications, are subregularity, controllability and calmness, the first connected with metric regularity, the second with linear openness and the third with the Aubin (pseudo-Lipschitz) property. The fourth, and rather new, concept of linear recession is also connected with the Aubin property.

Definition 2.59 (*subregularity*). Let $F : X \rightrightarrows Y$ and $\overline{y} \in F(\overline{x})$. It is said that F is *(metrically) subregular* at $(\bar{x}, \bar{y})$ if there is a $K > 0$ such that

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

for all x in a neighborhood of $\overline{x}$. The lower bound of such K is called the *modulus of subregularity* of F at $(\bar{x}, \bar{y})$. We shall denote it by $\mathrm{subreg}\, F(\overline{x}|\overline{y})$.

Subregularity can be viewed as the existence of a local error bound at the zero level for any of the three functions in the local regularity criteria of Theorem 2.54, namely $\varphi_{\overline{y}}$ near $(\bar{x}, \bar{y})$ and $\psi_{\overline{y}}$ and $\omega_{\overline{y}}^K$ near $\overline{x}$. On the other hand, for any function f the existence of a local error bound for f at $\overline{x}$ at the level $\alpha = f(\overline{x})$ is precisely subregularity at $(\overline{x}, \alpha)$ of the epigraphical mapping Epi $f : X \rightrightarrows I\!R$ defined by Epi $f(x) = \{\alpha \in I\!R : \alpha \geq f(x)\}$. Note also that regularity of F near $(\bar{x}, \bar{y})$ implies subregularity of F at $(x, y) \in$ Graph F sufficiently close to $(\bar{x}, \bar{y})$. An important point is that subregularity of a mapping at every point of its graph close to $(\bar{x}, \bar{y})$, even if there is a common lower bound for the moduli of subregularity at all such points, does not imply regularity near $(\bar{x}, \bar{y})$, unless there is a uniform lower estimate for the sizes of neighborhoods of the points for which the inequalities in the definition of subregularity hold (see e.g. Example 8.39 in Chap. 8).

The good news, however, is that, being a local property, subregularity admits a graph equivalent similar to what we have defined earlier for local regularity (Definition 2.19).

Definition 2.60 (*graph subregularity*). F is said to be *graph-subregular at* $(\bar{x}, \bar{y}) \in$ Graph F if there are $K > 0,\ \varepsilon > 0$ such that the inequality

$$d(x, F^{-1}(\overline{y})) \leq d_{1,K}((x, \overline{y}), \text{Graph } F) = \omega_{\overline{y}}^K(x) \tag{2.4.1}$$

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

Repeating word for word the proof of Proposition 2.20 with y replaced by $\overline{y}$, we prove

Proposition 2.61 (metric subregularity vs graph subregularity). *Let $F : X \rightrightarrows Y$, and $(\bar{x}, \bar{y}) \in$ (Graph F). Then F is subregular at $(\bar{x}, \bar{y})$ if and only if it is graph-subregular at $(\bar{x}, \bar{y})$. Moreover,* subreg$F(\overline{x}|\overline{y})$ *is the lower bound of $K > 0$ for which the inequality holds under a suitable choice of $\varepsilon > 0$.*

Taking this into account, we can state a sufficient condition for subregularity that easily follows from the Basic Lemma 2.42 (cf. Theorem 2.54).

Theorem 2.62 (subregularity criteria). *Let $F : X \rightrightarrows Y$ and $(\bar{x}, \bar{y}) \in$ Graph F. Given $\varepsilon > 0$, $r' > 0$, we assume that for some $\varepsilon > 0$ one of the following three properties holds true:*

(a) Graph F *is locally complete near $(\bar{x}, \bar{y})$ and there is a $\xi > 0$ such that for any x with $d(x, \overline{x}) < \varepsilon$ and any $v \in F(x)$, $v \neq \overline{y}$ there is a pair $(u, w) \in$* Graph F *such that (2.3.6) holds for $y = \overline{y}$;*

(b) X is complete and for any $x \in$ (dom $F) \cap \overset{\circ}{B}((\overline{x}, \varepsilon))$ *such that $\overline{y} \notin F(x)$ there is a $u \in$* dom F *such that either (2.3.9) holds for $y = \overline{y}$;*

(c) X is a complete space and for any x with $d(x, \overline{x}) < \varepsilon$ with $\overline{y} \notin F(x)$ there is a $u \neq x$ such that (2.3.12) holds for $y = \overline{y}$.

Then F is subregular at $(\bar{x}, \bar{y})$ with subreg $F(\overline{x}|\overline{y}) \leq K'$.

Proof. The proof is standard and again, practically the same for each of the three cases. So let us consider only (c). Applying the Basic Lemma 2.42 to the function $f(x) = \omega_{\overline{y}}^K(x)$ and x satisfying $d(x, \overline{x}) < \varepsilon/2$, we get that $d(x, [f \le 0]) \le f(x)$, which is precisely (2.4.1). $\square$

Definition 2.63 (*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

$$d(y, F(\overline{x})) \le Kd(x, \overline{x}) \tag{2.4.2}$$

whenever $d(x, \overline{x}) < \varepsilon$, $d(y, \overline{y}) < \varepsilon$ 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 calm $F(\overline{x}|\overline{y})$ (calm $F(\overline{x})$ if F is single-valued).

As in case of the Aubin property, it is possible to give some equivalent characterization of the calmness property (cf. Proposition 2.14 in the previous chapter).

Proposition 2.64 (characterization of calmness). *F is calm at $(\bar{x}, \bar{y})$ if and only if there are $\varepsilon > 0$ and $K \ge 0$ such that*

$$F(x) \cap B(\overline{y}, \varepsilon) \subset B(F(\overline{x}), Kd(x, \overline{x})), \quad \text{if } d(x, \overline{x}) < \varepsilon, \tag{2.4.3}$$

or equivalently

$$\text{ex}(F(x) \cap B(\overline{y}, \varepsilon)), F(\overline{x})) \le Kd(x, \overline{x}), \quad \text{if } d(x, \overline{x}) < \varepsilon.$$

Proof. Elementary. $\square$

Calmness of the inverse mapping F^{-1} is expressed by the inequality $d(x, F^{-1}(\overline{y})) \le Kd(y, \overline{y})$ valid for all $(x, y) \in$ Graph F sufficiently close to $(\bar{x}, \bar{y})$. The latter, as we shall see shortly, is equivalent to

$$d(x, F^{-1}(\overline{y})) \le Kd(\overline{y}, F(x)) \quad \text{if } d(x, \overline{x}) < \varepsilon, \tag{2.4.4}$$

which is precisely *metric subregularity* of F at $(\bar{x}, \bar{y})$: there are $\varepsilon > 0$ and $K \ge 0$ such that (2.4.4) holds if $d(x, \overline{x}) < \varepsilon$.

Here, as in the case of metric regularity, we can see that calmness of a set-valued mapping at every point of the graph in a neighborhood of $(\bar{x}, \bar{y})$ does not imply the Aubin property unless there is a common positive lower bound for ε and K^{-1} for all such points.

Proposition 2.65 (equivalence of metric subregularity and calmness of the inverse). *F is subregular at $(\bar{x}, \bar{y}) \in$ Graph F if and only if F^{-1} is calm at $(\overline{y}, \overline{x})$. Moreover,*

$$\text{subreg } F(\overline{x}|\overline{y}) = \text{calm } F^{-1}(\overline{y}|\overline{x}).$$

Proof. The implication metric subregularity of $F \Rightarrow$ calmness of F^{-1} is straightforward. To prove the converse, let us assume the contrary: for any $\delta \le \varepsilon$ there is an x with $d(x, \overline{x}) < \delta$ and such that $d(x, F^{-1}(\overline{y})) > Kd(\overline{y}, F(x))$. Take $\delta \le \min\{1, K\}$. We have $d(x, F^{-1}(\overline{y})) \le d(x, \overline{x}) < \delta \le K\varepsilon$. Thus there is a $y \in F(x)$ such that $d(x, F^{-1}(\overline{y})) > Kd(y, \overline{y})$ and $d(y, \overline{y}) \le K^{-1}d(x, \overline{x}) < \varepsilon$. But for every such y we have by calmness $d(x, F^{-1}(\overline{y})) \le Kd(y, \overline{y})$. □

As we shall see later, this pair of equivalent properties (calmness and, especially, metric subregularity) plays an important part in subdifferential calculus and optimization theory, in particular by being closely associated with the subtransversality property in Banach spaces, which in turn is a key element of so-called "metric qualification conditions", which are very natural and in a sense the weakest possible.

We shall proceed with a "point" counterpart of local openness.

Definition 2.66 (*controllability*). A set-valued mapping $F : X \rightrightarrows Y$ is said to be *(locally) controllable at* $(\bar{x}, \bar{y})$ if there are $r > 0$ and $\varepsilon > 0$ such that

$$B(\overline{y}, rt) \subset F(B(\overline{x}, t)), \quad \text{if } 0 \le t < \varepsilon. \tag{2.4.5}$$

The upper bound of such r is the *controllability rate* of F *at* $(\bar{x}, \bar{y})$. We shall denote it by $\text{contr}\ F(\overline{x}|\overline{y})$ (and $\text{contr}\ F(\overline{x})$ if F is single-valued).[3]

It is clear that F is open at a linear rate near $(\bar{x}, \bar{y})$ if and only if it is *uniformly controllable* in a neighborhood of $(\bar{x}, \bar{y})$, that is to say, if it is controllable at every (x, y) in the intersection of Graph F with the neighborhood with the same ε and r for all such (x, y). In fact, a much stronger statement, which is an immediate consequence of Corollary 3.27 to be proved in the next chapter, is true if Y is length space, namely the following statement holds true:

Theorem 2.67 (Regularity vs. controllability). *Let X be a metric space, let Y be a length space, let $F : X \rightrightarrows Y$ have a locally complete graph, and let* $(\bar{x}, \bar{y}) \in \text{Graph}\ F$. *Then*

$$\begin{aligned} \text{sur}\ F(\overline{x}|\overline{y}) = \lim_{\delta\to 0} \inf\{&\text{contr} F(x|y) : \\ &(x, y) \in \text{Graph}\ F,\ \max\{d(x, \overline{x}), d(y, \overline{y})\} < \delta\}. \end{aligned} \tag{2.4.6}$$

Exercise 2.68. Prove the theorem using Corollary 3.27 and a combination of Theorems 2.54 and 3.4.

---

[3]To explain the terminology, consider a "control system" governed by the differential equation $\dot{x} = f(x, u)$, where the control function $u(t)$ is taken from a pool of admissible controls $\mathcal{U}$. Once a control $u(t)$ and the initial state x_0 at, say $t = 0$, of the system are given, the equation defines a trajectory $x(t)$ of the system. Let $x_1 = x(1)$. The system is called locally controllable if small variations of $u(t)$ allow us to transfer x_0 to any point of a neighborhood of x_1.

The theorem essentially says that F is regular near $(\bar{x}, \bar{y}) \in \text{Graph } F$ if it is controllable at any point of the graph sufficiently close to $(\bar{x}, \bar{y})$ with r bounded away from zero (and no uniformity requirements concerning ε). It is not clear whether such a property is valid for non-local regularity on fixed sets. But for global regularity a similar property does hold, as we shall see in the last section of the next chapter.

We next note that there is another possible way to "pointify" the Aubin property, namely by fixing $\overline{y}$ and allowing x to change in a neighborhood of $\overline{x}$.

Definition 2.69 (*linear recession*). Let us say that F *recedes from* $\overline{y}$ *(at a linear rate)* near $\overline{x}$ if there are $\varepsilon > 0$ and $K \geq 0$ such that

$$d(\overline{y}, F(x')) \leq Kd(x, x') \tag{2.4.7}$$

if $\overline{y} \in F(x)$ and $x,\ x'$ are ε-close to $\overline{x}$.

We can also consider a weaker version of the property:

$$d(\overline{y}, F(x)) \leq Kd(x, \overline{x}), \quad \text{if } d(x, \overline{x}) < \varepsilon. \tag{2.4.8}$$

We shall say that F *recedes from* $\overline{y}$ **at** $(\bar{x}, \bar{y})$ if (2.4.8) holds. In the latter case we shall call the lower bound of such K the *speed of recession* of F from $\overline{y}$ at $(\bar{x}, \bar{y})$ and denote it by $\text{ress } F(\overline{x}|\overline{y})$. As usual, we set $\text{ress } F(\overline{x}|\overline{y}) = \infty$ if no such K exists.

Note that (2.4.7) can be equivalently written as

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

so it can be viewed as a sort of "anti-subregularity"!

We shall see that the first stronger property (2.4.7) plays an essential role in the metric implicit function theorem to be proved in §6. It is also an easy matter to verify that F has the Aubin property near $(\bar{x}, \bar{y})$ if and only if it recedes with finite speed at every (x, y) in an intersection of Graph F with a neighborhood of $(\bar{x}, \bar{y})$ and this property is uniform in the sense that the same ε and K can be chosen for all such (x, y).

Proposition 2.70 (controllability vs. linear recession of the inverse). *$F : X \rightrightarrows Y$ is controllable at $(\bar{x}, \bar{y})$ if and only if F^{-1} recedes from $\overline{x}$ at a linear rate at $(\overline{y}, \overline{x})$. Moreover,*

$$\text{contr } F(\overline{x}|\overline{y}) \cdot \text{ress } F^{-1}(\overline{y}|\overline{x}) = 1$$

(under the standard convention that $0 \cdot \infty = 1$).

Proof. To avoid confusion, we mention that the linear recession property for F^{-1} at $(\overline{y}, \overline{x})$ means that there is an $\varepsilon > 0$ such that

$$d(\overline{x}, F^{-1}(y)) \leq Kd(y, \overline{y}), \qquad \text{if} \quad d(y, \overline{y}) < \varepsilon. \tag{2.4.9}$$

Let F be controllable at $(\bar{x}, \bar{y})$, and let ε and r be as in Definition 2.66. If $d(y, \overline{y}) < r\varepsilon$, then taking $t = d(y, \overline{y})/r < \varepsilon$ we conclude that $y \in F(B(\overline{x}, t))$, so that there is a u with $d(u, \overline{x}) \le r^{-1}d(y, \overline{y})$ such that $y \in F(u)$. This means that $d(\overline{x}, F^{-1}(y)) \le r^{-1}d(y, \overline{y})$, which is the same as (2.4.9) with $K = r^{-1}$.

Conversely, if (2.4.9) holds, take a small δ and assume that $d(y, \overline{y}) \le t/K$ for some $t < \min\{1, K\}\varepsilon$. Then $d(\overline{x}, F^{-1}(y)) \le t$ and therefore there is a u such that $y \in F(u)$ and $d(u, x) < (1+\delta)t$. Setting $\tau = t(1+\delta)$ and taking into account that y is an arbitrary point of $B(\overline{y}, K^{-1}t$, we conclude that $B(\overline{y}, [(1+\delta)K]^{-1}\tau) \subset F(B(\overline{x}, \tau)$, which gives us (2.4.5) with ε replaced by $\min\{1, K^{-1}\}\varepsilon$ and $r = [(1+\delta)K]^{-1}$. As δ can be chosen arbitrarily small, the result follows □

Our final remark concerns the relationship between controllability and subregularity. Unlike linear openness and metric regularity, these two properties are no longer equivalent. A simple example of a subregular mapping which is not controllable is a set-valued mapping $I\!R \rightrightarrows I\!R$ which assumes the empty values at negative x and is equal to $[0, x]$ at $x \in I\!R_+$. On the other hand, the set-valued mapping (also from $I\!R$ into itself) equal to $[x, -x]$ on R_- and $\{x^2\}$ on $I\!R_+$ is controllable but not subregular at (0, 0) (see Fig. 2.2 below).

However, the following holds true.

Exercise 2.71. Prove that a mapping $F : X \rightrightarrows Y$ is controllable at $(\bar{x}, \bar{y}) \in \text{Graph } F$ if it is subregular at the point and open in the sense that for any $\varepsilon > 0$ there is a $\delta > 0$ such that

$$B(\overline{y}, \delta) \subset F(B(\overline{x}, \varepsilon)).$$

2.5 Perturbations and Stability

In this section we begin to discuss another question of primary importance: what happens to the regularity rates if the mapping is perturbed one way or another.

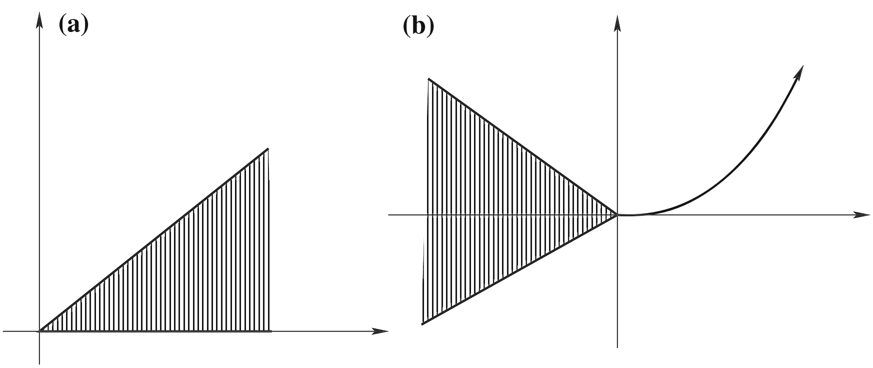

Fig. 2.2 Subregular but not controllable mapping in **a** and controllable but not subregular in **b**

Questions of this sort are often encountered in typical applications and we shall periodically return to them. Here we consider the problem in the most general setting of set-valued mappings between metric spaces but also with special attention to the most important class of additive perturbations that require linear structure in the range space. The general quantitative answer is that the regularity property cannot be destroyed by a perturbation which is Lipschitz in a certain sense with Lipschitz constant smaller than the rate of surjection of the unperturbed mapping. The following theorem is an umbrella result from which all other results of this kind (at least present in this book) follow.

Theorem 2.72 (stability under Lipschitz perturbation). *Let X be a complete metric space, let Y be a metric space, 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, and let $F(x) = \Psi(x, x)$. We assume that*

(a) for any $u \in U$ the mapping $\Psi(\cdot, u)$ is Milyutin regular on (U, V) with rate of surjection greater than r, that is,

$$d(x, (\Psi(\cdot, u))^{-1}(y)) < r^{-1} d(y, \Psi(x, u))$$

for any $x \in U$ and any $y \in V$ such that $d(y, \Psi(x, u)) \leq rm(x)$;

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

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

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

Proof. The proof of the theorem is amazingly simple. Take $x \in U$, $y \in V$ such that $r^{-1} d(y, F(x)) < m(x)$. By (a) there is a u such that $d(x, u) < r^{-1} d(y, F(x))$ and $y \in \Psi(u, x)$. Then $d(u, x) < m(x)$ and therefore $u \in U$. We can now apply (b) to estimate the distance from y to $\Psi(u, u)$ and conclude that $d(y, F(u)) \leq \ell d(x, u)$. It follows that

$$\begin{aligned} d(y, F(u)) \leq \ell d(x, u) &< \frac{\ell}{r} d(y, F(x)) \\ &= d(y, F(x)) - \frac{r - \ell}{r} d(y, F(x)) \leq d(y, F(x)) - (r - \ell) d(x, u). \end{aligned}$$

The middle and the last inequalities above show that the conditions of Theorem 2.50 are satisfied with $\lambda = \ell / r$ and the proof is completed by application of the latter. □

Remark 2.73. Taking Remark 2.51 into account, we conclude that *the theorem remains valid if we assume that the graph of Ψ is closed only relative to $U \times Y$.*

To state a local version of the theorem we need a concept of uniform regularity of a family of mappings near a certain point.

Definition 2.74 (*uniform versions of the properties*). Let P be a topological space, let $F : P \times X \rightrightarrows Y$, let $\overline{p} \in P$, and let $\overline{y} \in F(\overline{p}, \overline{x})$. We shall say that $F(p, \cdot)$ is *K-regular near* $(\bar{x}, \bar{y})$ *uniformly* in $p \in P$ near $\overline{p}$ if there are $\varepsilon > 0$ and a neighborhood $W \subset P$ of $\overline{p}$ such that

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

for any $p \in W$, any x with $d(x, \overline{x}) < \varepsilon$ and any y with $d(y, \overline{y}) < \varepsilon$. If the inequality is satisfied only for $y = \overline{y}$ and all x and p as above, we say that $F(p, \cdot)$ is *K-subregular at* $(\bar{x}, \bar{y})$ *uniformly* in p near $\overline{p}$.

Likewise, we say that $F(p, \cdot)$ is *ℓ-pseudo-Lipschitz near* $(\overline{x}, \overline{y})$ *uniformly in x* near $\overline{x}$ if there are an $\varepsilon > 0$ and a neighborhood $W \subset P$ of $\overline{p}$ such that

$$d(y, F(p, x')) \leq \ell d(x, x')$$

for all $x, x' \in B(\overline{x}, \varepsilon)$, $p \in W$ and $y \in F(p, x) \cap B(\overline{y}, \varepsilon)$. If the inequality is satisfied only for $y = \overline{y}$ and all x and p as above, we say that $F(p, \cdot)$ *recedes from $\overline{y}$ near $\overline{x}$ with speed at most ℓ uniformly* in p near $\overline{p}$.

Theorem 2.75 (stability under Lipschitz perturbations: local version). *Let, as in Theorem 2.72, X be a complete metric space, let $\Psi : X \times X \rightrightarrows Y$ have locally closed graph, let $F(x) = \Psi(x, x)$, and let $(\bar{x}, \bar{y}) \in$* Graph F. *Assume that there are $r > \ell \geq 0$ such that*

(a) $\Psi(\cdot, u)$ is r^{-1}-regular near $(\bar{x}, \bar{y})$ uniformly in u near $\overline{x}$;

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

Then F is regular near $(\bar{x}, \bar{y})$ with rate of surjection not smaller than $r - \ell$.

We leave the proof to the reader: it is an easy consequence of Theorem 2.72 and Proposition 2.29.

In both theorems we consider $\Psi(\cdot, w)$ a perturbation of F. A more traditional and probably the most important case of an additive perturbation can be easily treated in the framework of the theorem. The two theorems to follow are basically reformulations of Theorems 2.72 and 2.75 for mappings Ψ having special additive structure.

Theorem 2.76 (Additive perturbation – Milyutin's theorem). *Let X be a complete metric space, let Y be a normed space, and let $F : X \rightrightarrows Y$ and $G : X \rightrightarrows Y$ be set-valued mappings with closed graphs. Let further $U \subset X$ be an open set such that F is Milyutin regular on U with* $\mathrm{sur}_m F(U) \geq r$ *and G is (Hausdorff) Lipschitz on U with* $\mathrm{lip}\, G(U) \leq \ell < r$. *If either F or G is single-valued continuous on U, then $F + G$ is Milyutin regular on U and* $\mathrm{sur}_m(F + G)(U|Y) \geq r - \ell$.

Proof. As one of the mappings is single-valued continuous on U and the graph of the other is closed, the graph of Ψ is also closed relative to $U \times Y$. Apply Theorem 2.72 to $\Psi(x, u) = F(x) + G(u)$ taking Remark 2.73 into account. □

Remark 2.77. We have to assume $V = Y$ in this theorem. Otherwise we would be compelled to assume some coordination between V and G, for instance that $F(x) + G(u) \subset V$ for all $(x, u) \in U \times U$ belonging to a neighborhood of the diagonal of $X \times X$. Such an assumption, which looks natural in Theorem 2.72, would be rather awkward when dealing with a sum of mappings.

Nonetheless it is possible to extend Theorem 2.72 to additive perturbations with smaller V.

Theorem 2.78 (additive perturbation – the case $V \neq Y$). *Let X be a complete metric space, let Y be a normed space, and let $F : X \rightrightarrows Y$ and $G : X \rightrightarrows Y$ be set-valued mappings with closed graphs. Let further $U \subset X$ and $V \subset Y$ be open sets such that F is Milyutin regular on (U, V) with* $\operatorname{sur} F(U|V) \geq r$ *and G is (Hausdorff) Lipschitz on U with* $\operatorname{lip} G(U|Y) \leq \ell < r$. *Assume furthermore that there is an open set $W \subset Y$ such that $W - G(U) \subset V$. If F or G is single-valued continuous on U, then $F + G$ is Milyutin regular on (U, W) and* $\operatorname{sur} (F + G)(U|W) \geq r - \ell$.

Proof. Consider again the set-valued mapping $\Psi(x, u) = F(x) + G(u)$ from $U \times U$ into Y. All we need is to check that $\Psi(\cdot, u)$ is Milyutin regular on (U, W). Indeed, the graph of Ψ is closed relative to $(U \times U) \times Y$, since one of the mappings is single-valued continuous on U, $\Psi(x, \cdot)$ is pseudo-Lipschitz on $U \times Y$ and, all the more, on $U \times W$. In other words, if $\Psi(\cdot, u)$ is Milyutin regular on (U, W), the result again follows from Theorem 2.72 and Remark 2.73.

To prove that $\Psi(\cdot, u)$ is Milyutin regular on (U, W) we have to elaborate a bit on the proof of Theorem 2.72.

So let $v \in \Psi(x, u)$ for some $x, u \in U$ and let $y \in W$ satisfy $d(y, v) \leq rt$ for some $t < m(x)$. We have $v = v' + w'$, where $v' \in F(x)$ and $w' \in G(u)$. Since either F or G is single-valued, one of the inclusions is actually an equality, which means that v' and w' are uniquely defined by v.

Set $y' = y - w'$. By the assumption, $y' \in V$. On the other hand $d(y', v') = d(y, v) \leq rt$. As F is Milyutin regular on (U, V), there is an $x' \in U$ such that $d(x, x') \leq r^{-1} d(y', v')$ and $y' \in F(x')$. This means that $d(x, x') \leq r^{-1} d(y, v) \leq t$ and $y \in F(x') + G(u) = \Psi(x', u)$.

As y can be arbitrarily chosen in $B(v, rt) \cap W$, it follows that $B(v, rt) \cap W \subset \Psi(B(x, t), u)$. This is true for any $(x, u) \in U \times U$ and any $v \in \Psi(x, u)$, hence $\Psi(\cdot, u)$ is Milyutin regular on (U, W). $\square$

A local version of the Milyutin perturbation theorem is now straightforward.

Theorem 2.79 (Milyutin's perturbation theorem – local version). *Let X be a complete metric space, let Y be a normed space, let $F : X \rightrightarrows Y$ have locally closed graph and be regular near $(\bar{x}, \bar{y}) \in$* Graph F *with* $\operatorname{sur} F(\overline{x}|\overline{y}) \geq r$, *and let $g : X \to Y$ be defined and Lipschitz in a neighborhood of $\overline{x}$ with* $\operatorname{lip} g(\overline{x}) \leq \ell < r$. *Specifically, suppose that there is an $\varepsilon > 0$ such that*

$$B(F(x), rt) \cap B(\overline{y}, \varepsilon) \subset F(B(x(t))),\ \ d(g(x), g(x')) \leq \ell d(x, x')$$

if $x' \in B(\overline{x}, \varepsilon)$ *and* $t < \varepsilon$. *Then given a* $\delta \le \varepsilon/2$,

$$B((F+g)(x), (r-\ell)t) \cap B(\overline{y} + g(\overline{x}), \delta) \subset (F+g)(B(x,t))$$

for any $x \in \overset{\circ}{B}(\overline{x}, \delta)$. *In particular,*

$$\text{sur}\,(F+g)(\overline{x}, \overline{y}+\overline{z}) \ge r - \ell.$$

Proof. By Proposition 2.29 F is Milyutin regular on (U, V) (with the rate of surjection not smaller than r), where $U = \overset{\circ}{B}(\overline{x}, \varepsilon)$, $V = \overset{\circ}{B}(\overline{y}, r\varepsilon)$. Then $g(U) \subset g(\overline{x}) + \overset{\circ}{B}(\overline{x}, \ell\varepsilon)$. Set $W = g(\overline{x}) + \overset{\circ}{B}(\overline{y}, (r-\ell)\varepsilon)$. Then $W - g(U) \subset V$ and, by Theorem 2.76, $F+g$ is Milyutin regular on (U, W) with $\text{sur}_m(F+g)(U|W) \ge r-\ell$. The result follows. ☐

The last theorem, in turn, allows us to get a stronger version of the Lyusternik–Graves theorem stating that its condition is not only sufficient but also necessary for regularity.

Corollary 2.80 (Lyusternik–Graves from Milyutin). *Let X and Y be Banach spaces, and let* $F : X \to Y$ *be strictly differentiable at* $\overline{x}$. *Then the rates of surjection of F near* $\overline{x}$ *and of* $F'(\overline{x})$ *coincide.*

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 becomes arbitrarily small on neighborhoods of $\overline{x}$ as the diameter of the neighborhoods goes to zero. Applying Milyutin's theorem, we conclude that the rates of surjection of F near $\overline{x}$ and of $F'(\overline{x})$ coincide. ☐

It is possible to look at Theorem 2.79 from a slightly different angle. Assume that G is single-valued Lipschitz. Then the mapping $F(x) + G(x)$ can be viewed as a composition $\Phi(x, G(x))$ with $\Phi(x, y) = F(x) + y$. The following proposition is a certain elaboration on this.

Proposition 2.81 (single-valued perturbations). *Let* X, Y *and* Z *be metric spaces with* X *being complete, and let* $\Phi : X \rightrightarrows Y$ *and* $G : X \times Y \to Z$. *We assume that*

(a) the graph of Φ *is closed and* Φ *is Milyutin regular on* $U \times Y$, *where* U *is an open subset of* X, *with* $\text{sur}_M \Phi(U|Y) \ge r$;

(b) $G(x, \cdot)$ *is an isometry from* Y *onto* Z *for any* $x \in X$;

(c) $G(\cdot, y)$ *satisfies the Lipschitz condition with constant* $\ell < r$ *for any* $y \in Y$.

Set $F(x) = G(x, \Phi(x))$. *Then* F *is Milyutin regular on* $U \times Z$ *and* $\text{sur}_m F(U|Z) \ge r - \ell$.

Proof. We first note that the graph of F is closed in the product metric of $X \times Z$. This follows from the simple observation (due to (b) and (c)): if $z = G(x, y)$ and $z' = G(x', y')$, then $|d(z, z') - d(y, y')| \le \ell d(x, x')$. Thus, if (x_n, z_n) is a sequence of elements of Graph F converging to (x, z), then $z_n = G(x_n, y_n)$ for some uniquely defined $y_n \in \Phi(x_n)$ and $G(x, y_n) \to z$. As $G(x, \cdot)$ is an isomorphism, it follows that there is a $y \in Y$ such that $G(x, y) = z$ and y_n converge to y. Finally, $y \in \Phi(x)$ as the graph of Φ is closed.

It remains to set $\Psi(x, u) = G(u, F(x))$ and apply Theorem 2.72. □

Another circumstance to be mentioned is that in Theorems 2.76 and 2.78 one of the mappings is assumed single-valued. This assumption is essential. With both mappings set-valued the result may be wrong, as the following example shows, unless we are dealing with global regularity, as we shall see in the next chapter – see Theorem 3.45.

Example 2.82 (cf. [96]). Let $X = Y = I\!R$, $G(x) = \{x^2, -1\}$, $F(x) = \{-2x, 1\}$. Set $U = (-1/4, 1/4)$, $V = (-1/2, 1/2)$. Then it is easy to see that F is γ-regular on (U, V) with $\mathrm{sur}_\gamma F(U|V) = 2$ if $\gamma(x) \le 1/2$, and G is Hausdorff Lipschitz (hence pseudo-Lipschitz) on U with $\mathrm{lip}\, G = 1/2$. However,

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

is not even regular at (0, 0). Indeed, $(\xi, 0) \in \mathrm{Graph}\, \Phi$ for any ξ. However, if $\xi \ne 0$, 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.

We conclude the section with the observation that, unlike linear openness and metric regularity, the subregularity property is not stable and can disappear under arbitrarily small perturbation (see Example 3.23 in the next chapter). However, in the next section we shall see that a certain strong version of the subregularity property is free from this flaw.

2.6 Metric Implicit Function Theorems: Strong Regularity

In this section we make the first step to approach the inverse and implicit function theorems in the context of variational analysis. In the course of future discussions we shall return to these theorems several times to study the effects the properties of spaces and mappings may have on their conclusions. Here we shall be interested in purely metric aspects of the problem.

Generally speaking, the essence of the inverse function theorem is already captured by the main Equivalence Theorem 2.25. 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. To this end we shall consider a set-valued mapping $F : P \times X \rightrightarrows Z$ defined on a product of two metric spaces P and X and associated inclusion

$$y \in F(p, x) \tag{2.6.1}$$

in which we interpret p as a parameter and x as an argument.

Let $S(p, y) = \{x \in X : \ y \in F(p, x)\}$ stand for the solution mapping. In the theorems to follow we consider $P \times Y$ with an ℓ^1-type distance $d_{\alpha,1}((p, y), (p', y')) = \alpha d(p, p') + d(y, y')$, where α is determined by Lipschitz moduli of mappings involved. We shall begin with the simple general statement below. In a nutshell, this theorem says that the Lipschitz behavior of the solution map (as a function of both p and y) is guaranteed by the combination of the regular behavior of F as a function of x and the pseudo-Lipschitz behavior as a function of p (cf. Theorem 2.72).

Theorem 2.83 (implicit function theorem – metric version). *Let $\overline{y} \in F(\overline{p}, \overline{x})$. Suppose that*

(a) $F(p, \cdot)$ is K-regular near $(\overline{x}, \overline{y})$ uniformly in p in a neighborhood of $\overline{p}$;

(b) $F(\cdot, x)$ is α-pseudo-Lipschitz near $(\overline{p}, \overline{y})$ uniformly in x in a neighborhood of $\overline{x}$.

Then S has the Aubin property near $((\overline{p}, \overline{y}), \overline{x})$ with

$$\operatorname{lip} S((\overline{p}, \overline{y})|\overline{x}) \le K$$

if $P \times Y$ is considered with the $d_{\alpha,1}$-metric.

Proof. Formally the assumptions of the theorem mean that we can find an $\varepsilon > 0$ such that

(a') $d(x, S(p, y)) \le Kd(y, F(p, x))$ whenever $d(x, \overline{x}) < \varepsilon$, $d(y, \overline{y}) < \varepsilon$, $d(p, \overline{p}) < \varepsilon$;

(b') $d(y, F(p', x)) \le \alpha d(p, p')$ for any $x,\ y,\ p$ as in (a') such that $y \in F(p, x)$.

Take a $\delta \in (0, \varepsilon)$ small enough to make sure that the p, x and y that appear in the proof do not leave the open ε-neighborhoods of $\overline{p}$, $\overline{x}$ and $\overline{y}$. We first observe that $S(p, y) \ne \emptyset$ for all p and y in the open δ-balls around $\overline{p}$ and $\overline{y}$, respectively. Indeed, as $F(\overline{p}, \overline{x}) \ne \emptyset$, it follows from the definition of the pseudo-Lipschitz property (see Remark 2.24) that $F(p, \overline{x}) \ne \emptyset$ for all p with $d(p, \overline{p}) < \delta$. Now (a) implies that $S(p, y) \ne \emptyset$.

Let now (p, x, y) be within ε of $(\overline{p}, \overline{x}, \overline{y})$ and $y \in F(p, x)$. If $d(p', p) < \varepsilon$, $d(y', y) < \varepsilon$, then

$$\begin{aligned} d(x, S(p', y')) &\le Kd(y', F(p', x)) \\ &\le K\big(d(y, y') + d(y, F(p', x)\big) \le K\big(d(y, y') + \alpha d(p, p')\big). \end{aligned}$$

Here the first inequality follows from (a') and the third from (b'). The proof has been completed. □

It is often convenient to consider y as an additional perturbation. But the standard statements of implicit function theorems deal with fixed y, say $y = \overline{y}$. For that case, as follows from the proof of the above theorem, we need to verify the conditions only for $y' = y = \overline{y}$. In other words, specifying the theorem and its proof for the inclusion $\overline{y} \in F(p, x)$ (with fixed $\overline{y}$) we get the following result.

Theorem 2.84 (implicit function theorem with fixed y). *Suppose that*

(a) $F(p,\cdot)$ is K-subregular at $(\overline{x},\overline{y})$ uniformly in p in a neighborhood of $\overline{p}$;

(b) $F(\cdot,x)$ recedes from $\overline{y}$ near $\overline{p}$ with speed at most α uniformly in x in a neighborhood of $\overline{x}$.

Then $S(\cdot,\overline{y})$ has the Aubin property near $(\overline{p},\overline{x})$ with

$$\operatorname{lip} S(\cdot,\overline{y})(\overline{p}|\overline{x}) \le K\alpha.$$

The two theorems can be viewed as very general versions of the implicit function theorem. They are actually equivalent if Y is a normed space (or more generally, linear metric space with invariant metric). To see this it is enough to write the inclusion $y \in F(x)$ as $0 \in F(x)-y$. Note also that, like the equivalence theorem, the theorems hold unconditionally, for all set-valued mappings satisfying the assumptions without any completeness or even closedness requirements.

In the subsequent discussions we shall follow step by step the evolution of the results accompanying specializations of the structural requirements on the mapping and modes of the behavior we wish the solution map to have. As the first step towards the classical implicit function theorem, we shall look here at set-valued mappings with a single-valued inverse. The latter property can be best formalized in the framework of the important concept of strong regularity defined below.

Let us say that a set-valued mapping $F : X \rightrightarrows Y$ is *linearly disjoint* on (U,V) (with, as usual, $U \subset X$ and $V \subset Y$) if there is a $K > 0$ such that

$$d(x,u) \le Kd(y,F(u)),$$

whenever $x,u \in U$ and $y \in F(x) \cap V$. Clearly, if $F(x) \cap V \ne \emptyset$ for some $x \in U$, then F is linearly disjoint on (U,V) if and only if $G(y) = F^{-1}(y) \cap U$ is at most single-valued on the intersection of its domain with V. We shall be interested mainly in the case when F is linearly disjoint near some $(\bar{x},\bar{y}) \in \text{Graph } F$, that is, with U and V being neighborhoods of $\overline{x}$ and $\overline{y}$, respectively.

Definition 2.85. Let $F : X \rightrightarrows Y$, and let $\overline{y} \in F(\overline{x})$. We say that F is *strongly (metrically) regular* near $(\bar{x},\bar{y}) \in \text{Graph } F$ if F is linearly disjoint near $(\bar{x},\bar{y})$ and the F-image of some neighborhood of $\overline{x}$ contains a neighborhood of $\overline{y}$, in other words, if there are $\varepsilon > 0$, $\delta > 0$ and $K \in [0,\infty)$ such that

$$B(\overline{y},\delta) \subset F(B(\overline{x},\varepsilon)) \quad \& \quad d(x,u) \le Kd(y,F(x)) \tag{2.6.2}$$

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

We shall also say, following [96], that *F has a single-valued localization near* $(\bar{x},\bar{y})$ if there are $\varepsilon > 0$, $\delta > 0$ such that the restriction of $F(x) \cap B(\overline{y},\delta)$ to $B(\overline{x},\varepsilon)$ is single-valued. If, in addition, the restriction is Lipschitz continuous, we say that F has *Lipschitz localization* near $(\bar{x},\bar{y})$.

The obvious difference between metric regularity and strong metric regularity (justifying the use of the word "strong") is that in the latter we require that the distance from u to every $x \in F^{-1}(y)$ does not exceed $Kd(y, F(u))$ while the first requires that the distance from u to $F^{-1}(y)$ is not greater than $Kd(y, F(u))$. It is interesting to observe a certain symmetry between the given definition of strong regularity and the Aubin property. Each of them can be viewed as obtained from the definition of metric regularity by replacing the distance to a set by the distance to an arbitrary element of the set, in the left-hand side of the inequality for the strong metric regularity and in the right-hand side for the Aubin property. But if in the second case we get just an equivalent of metric regularity, strong metric regularity is a much stronger property. The following proposition summarizes some basic properties of strong metric regularity which easily follow from the definition.

Proposition 2.86 (equivalence for strong regularity). *Let $F : X \rightrightarrows Y$ and $(\bar{x}, \bar{y}) \in$* Graph F. *Then the following five statements are equivalent*

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

(b) F is regular near $(\bar{x}, \bar{y})$ and there are $\varepsilon > 0$ and $\delta > 0$ such that

$$F(x) \bigcap F(u) \bigcap B(\overline{y}, \delta) = \emptyset \tag{2.6.3}$$

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

(c) F is regular near $(\bar{x}, \bar{y})$ and F^{-1} has a single-valued localization near $(\overline{y}, \overline{x})$.

(d) F^{-1} has the Aubin property near $(\overline{y}, \overline{x})$ and a single-valued localization in a neighborhood of the point.

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

Moreover, if F is strongly regular at $(\bar{x}, \bar{y})$, then the lower bound of K for which the second part of (2.6.2) holds and the Lipschitz modulus of its Lipschitz localization G at $\overline{y}$ coincide with reg $F(\overline{x}|\overline{y})$.

Proof. (a) $\Rightarrow$ (b). If under the conditions specified in Definition 2.85 $y \in F(x)$ with $d(x, \overline{x}) < \varepsilon$, then $x = u$ by (2.6.2). This means that $F^{-1}(y) \cap B(\overline{y}, \delta) = \{u\}$ and the second relation in (1) reduces to $d(x, F^{-1}(y)) \le Kd(y, F(x))$. By the first part of (2.6.2) this applies to every $y \in B(\overline{y}, \delta)$ and all $x \in B(\overline{x}, \varepsilon)$, whence regularity and (2.6.3).

The equivalence of (b) and (c) is obvious while the equivalence of (c) and (d) is an immediate consequence of the Equivalence Theorem 2.25. The implications (d) $\Rightarrow$ (e) follows directly from the definition of the Aubin property. The equivalence theorem is also instrumental in the proof that (e) $\Rightarrow$ (a).

Indeed, if (e) holds, then by Theorem 2.25 F is open at a linear rate near $(\bar{x}, \bar{y})$. Let now $G(y)$ be a Lipschitz localization of F^{-1} near $(\overline{y}, \overline{x})$. Then there are $\varepsilon > 0,\ \delta > 0$ such that G is defined on $B(\overline{y}, \delta)$, $G(\overline{y}) = \overline{x}$ and $\{G(y)\} = F^{-1}(y) \cap B(\overline{x}, \varepsilon)$ if $d(y, \overline{y}) \le \delta$ and $d(G(y), \overline{x}) < \varepsilon$. Take a $\delta' < \delta/3$ such that $K\delta' < \varepsilon$. Then $G(B(\overline{y}, \delta')) \subset B(\overline{x}, \varepsilon)$ and therefore $B(\overline{y}, \delta') \subset F(B(\overline{x}, \varepsilon))$. This is precisely the first relation in (2.6.2).

To prove the last statement, we first note that $d(x, u) \le Kd(y, v)$ if $x = F^{-1}(y)$ and $u = F^{-1}(v)$ if y and v are within δ of $\overline{y}$ and K is the Lipschitz constant of G

in $B(\overline{y}, \delta)$. If $d(y, \overline{y}) < \delta',\ d(v, \overline{y}) < \delta'$, $x = F^{-1}(v)$ and $u = F^{-1}(y)$, then the distances of x and u from $\overline{x}$ do not exceed $K\delta'$. If now $d(y, v') \le d(y, v)$ for some $v' \in F(x)$, then $d(\overline{y}, v') \le \delta$, and therefore $d(x, u) \le Kd(y, v')$. It follows that $d(x, u) \le Kd(y, F(x))$.

Finally, let $x \in B(\overline{x}, K\delta')$ be such that $F(x) \cap B(\overline{y}, \delta) = \emptyset$. Then $d(y, F(x)) \ge 2\delta'$ and, on the other hand, $d(x, u) \le d(x, \overline{x}) + d(u, \overline{x}) \le 2K\delta' \le Kd(y, F(x))$. Thus, the second inequality in (2.6.2) also holds whenever $x,\ u$ are within $K\delta'$ of $\overline{x}$, and $y \in F(u)$ is within δ' of $\overline{y}$. □

The equivalence of (a) and (c) can be interpreted as an inverse mapping theorem. We proceed to look at the effect of strong regularity on the properties discussed in the previous sections. Although it will not be immediately used in this section, the theorem below contains important information about strong regularity.

Theorem 2.87 (persistence of strong regularity under Lipschitz perturbation). *Let X be a complete metric space. We consider a set-valued mapping $\Phi : X \rightrightarrows Y$ with closed graph, and a (single-valued) mapping $G : X \times Y \to Z$. Let $\overline{y} \in \Phi(\overline{x})$ and $\overline{z} = G(\overline{x}, \overline{y})$. We assume that*

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

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

(c) $G(\cdot, y)$ is Lipschitz with constant $\ell < r$ in a neighborhood of $\overline{x}$, and likewise for all y in a neighborhood of $\overline{y}$.

Set $F(x) = G(x, \Phi(x))$. Then F is strongly regular near $(\overline{x}, \overline{z})$.

Proof. In view of (local version of) Proposition 2.81 and Proposition 2.86 we only need to check that F is disjoint near $(\overline{x}, \overline{z})$, that is, that there are $\varepsilon > 0$ and $\delta > 0$ such that the intersection $F(x) \cap F(u) \cap B(\overline{z}, \delta)$ is empty if x and u are within ε of $\overline{x}$.

Set $K = r^{-1}$ and choose $\varepsilon > 0$ and $\delta > 0$ to make sure that $B(\overline{x}, \varepsilon)$ and $B(\overline{y}, \delta)$ lie within the neighborhoods in (b) and (c), $B(\overline{y}, \delta) \subset \Phi(B(\overline{x}, \varepsilon))$ and $\Phi(x) \cap \Phi(u) \cap B(\overline{y}, \delta) = \emptyset$ for all $x, u \in B(\overline{x}, \varepsilon)$.

Assume that there are $x \ne u$ belonging to $B(\overline{x}, \varepsilon)$ and a $z \in B(\overline{z}, \delta)$ which belongs both to $F(x)$ and to $F(u)$. Then there are $y_x \in \Phi(x)$ and $y_u \in \Phi(u)$ such that $z = G(x, y_x) = G(u, y_u)$. We have $d(y_x, \overline{y}) = d(z, \overline{z}) \le \delta$ and likewise $d(y_u, \overline{y}) \le \delta)$ and

$$\begin{aligned} d(z, \overline{z}) &= d(G(x, y_x), G(\overline{x}, \overline{y})) \\ &\ge d(G(\overline{x}, y_x), G(\overline{x}, \overline{y}) - d(G(x, y_x), G(\overline{x}, \overline{y}) \\ &\ge d(G(\overline{x}, y_x), G(\overline{x}, \overline{y}) - \ell d(x, \overline{x}) = d(y_x, \overline{y}) - \ell d(x, \overline{x}), \end{aligned}$$

that is, $d(y_x, \overline{y}) \le \ell d(x, \overline{x}) + d(z, \overline{z})$ and a similar inequality with u and y_u replacing x and y_x also holds. Likewise, as $G(x, \cdot)$ is an isometry and z belongs to both $F(x)$ and $F(u)$,

$$d(y_x, y_u) = d(G(x, y_x), G(x, y_u)) \le d(G(x, y_x), G(u, y_u)) + \ell d(x, u) = d(x, u).$$

On the other hand, as Φ is strongly regular, we have taking into account the choice of ε and δ,

$$d(x,u) \le r^{-1}d(y_x, \Phi(u)) \le r^{-1}d(y_x, y_u) \le (\ell/r)d(x,u) < d(x,u),$$

and we arrive at a contradiction. □

Remark 2.88. It is to be observed, in connection with the last proposition, that strong regularity is not preserved under set-valued perturbations like those in Theorem 2.72, even if the mapping itself is single-valued. Here is a simple example:

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

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 standard regularity.

Finally, we have to introduce the strong analogue of the subregularity property.

Definition 2.89 (*strong subregularity*). $F : X \rightrightarrows Y$ is *strongly subregular* at $(\bar{x}, \bar{y}) \in \text{Graph } F$ if there are $K > 0$ and $\varepsilon > 0$ such that

$$d(x,\overline{x}) \le Kd(\overline{y}, F(x)), \quad \text{if } d(x,\overline{x}) < \varepsilon. \tag{2.6.4}$$

As immediately follows from the definition, strong subregularity simply means that F is subregular at $(\bar{x}, \bar{y})$ and $\overline{y} \notin F(x)$ for all x in a neighborhood of $\overline{x}$, on the one hand, and the lower bound of K for which (2.6.4) holds coincides with subreg $F(\overline{x}|\overline{y})$ on the other.

A remarkable fact is that strong subregularity, unlike its general counterpart, remains stable under Lipschitz perturbations of the mapping.

Proposition 2.90 (stability of strong subregularity). *Let X and Y be metric spaces, and let $F : X \rightrightarrows Y$ be strongly subregular at $(\bar{x}, \bar{y}) \in$ Graph F with* subreg $F(\overline{x}|\overline{y}) = K > 0$. *Let further Z be another metric space and $G : X \times Y \to Z$ a continuous mapping such that $G(x,\cdot)$ is an isometry for all x in a neighborhood of $\overline{x}$ and for any y the mapping $G(\cdot, y)$ is Lipschitz continuous in a neighborhood of $\overline{x}$ with Lipschitz constant ℓ such that $K\ell < 1$. Set $\overline{z} = G(\overline{x}, \overline{y})$. Then $\Phi(x) = (G \circ F)(x) = G(x, F(x))$ is strongly subregular at $(\overline{x}, \overline{z})$ with*

$$\text{subreg } \Phi(\overline{x}|\overline{z}) \le \frac{K}{1 - K\ell}.$$

In particular, if Y is a Banach space and $g : X \to Y$ is Lipschitz with constant $\ell < K^{-1}$, then the above inequality holds for $\Phi(x) = F(x) + g(x)$ and $\overline{z} = \overline{y} + g(\overline{x})$.

Proof. We have

$$\begin{aligned} d(\overline{z}, \Phi(x)) &= d(\overline{z}, G(x, F(x)) \geq d(\overline{z}, G(\overline{x}, F(x)) - \ell d(x, \overline{x}) \\ &= d(G(\overline{x}, \overline{y}), G(\overline{x}, F(x)) - \ell d(x, \overline{x}) \\ &= d(\overline{y}, F(\overline{x})) - \ell d(x, \overline{x}) \geq (K^{-1} - \ell) d(x, \overline{x}) = \frac{1 - K\ell}{K} d(x, \overline{x}), \end{aligned}$$

as claimed. □

Observe that, unlike in case of "full" regularity (cf. Theorem 2.72), here we do not need any completeness assumptions. Note also that we do not use in the proof the full power of the Lipschitz property of $G(\cdot, x)$. All we need is that $d(\overline{z}, G(\overline{x}, F(x))) - d(\overline{z}, G(x, F(x)) \leq \ell d(x, \overline{x})$ for all x in a neighborhood of $\overline{x}$. If G has an additive structure: $G(x, y) = g(x) + y$, this observation shows that it is sufficient to assume that g is calm at $\overline{x}$ with modulus not greater than ℓ.

We further need uniform versions of strong regularity and subregularity to use the properties in implicit function theorems. Let us say that $F(p, \cdot)$ is *strongly K-regular near* $(\bar{x}, \bar{y})$ *uniformly in p* near $\overline{p}$ if $F(p, \cdot)$ is K-regular uniformly in p near $(\bar{x}, \bar{y})$ and for some $\varepsilon > 0$ and all $p \in \overset{\circ}{B}(\overline{\psi}, \varepsilon),\ x, x' \in \overset{\circ}{B}(\overline{x}, \varepsilon),\ x \neq x',\ y \in \overset{\circ}{B}(\overline{y}, \varepsilon)$

$$F(p, x) \cap F(p, x') \cap \overset{\circ}{B}(\overline{y}, \varepsilon) = \emptyset \tag{2.6.5}$$

In the spirit of the definition of strong subregularity in which we deal with a single value of y, it is natural to call $F(p, \cdot)$ *strongly K-subregular at* $(\bar{x}, \bar{y})$ *uniformly* in p near $\overline{p}$ (where, of course, we assume that $\overline{y} \in F(\overline{p}, \overline{x})$) if $F(p, \cdot)$ is K-subregular at $(\bar{x}, \bar{y})$ uniformly in p and $\max\{d(\overline{y}, F(p, x)), d(\overline{y}, F(p, x'))\} > 0$ for all p, x, $x \neq x'$, close respectively to $\overline{p}$ and $\overline{x}$. Clearly, uniform strong regularity of F (in p) implies uniform strong subregularity.

We can now return to the main subject of this section.

Theorem 2.91 (implicit function theorem with strong regularity/subregularity). *Suppose that*

(a) $F(p, \cdot)$ is strongly K-regular near $(\overline{x}, \overline{y})$ uniformly in p in a neighborhood of $\overline{p}$;

(b) $F(\cdot, x)$ is α-pseudo-Lipschitz near $(\overline{p}, \overline{y})$ uniformly in x in a neighborhood of $\overline{x}$.

Then the solution map $S(p, y)$ has a Lipschitz localization G near $((\overline{p}, \overline{y}), \overline{x})$. In particular, $y \in F(p, S(p, y))$ for all (p, y) in a neighborhood of $(\overline{p}, \overline{y})$.

Likewise, assume instead of (a), that

(a_s) $F(p, \cdot)$ is strongly K-subregular at $(\bar{x}, \bar{y})$ uniformly in p near $\overline{p}$,
and instead of (b) that

(b_s) $F(\cdot, x)$ recedes from $\overline{y}$ at $(\overline{p}, \overline{y})$ with speed not greater than α uniformly in x in a neighborhood of $\overline{x}$.

Then $S(\cdot, \overline{y})$ has Lipschitz localization near $(\overline{p}, \overline{x})$. The estimates for the Lipschitz moduli found in Theorems 2.83 and 2.84 remain valid in both cases.

Proof. The proofs are identical in both cases (with the only difference that in the first we refer to Theorem 2.83 and in the second to Theorem 2.84), so we shall consider only the second.

Clearly, F satisfies the assumptions of Theorem 2.84 and therefore $S(\cdot, \overline{y})$ has the Aubin property near $(\overline{p}, \overline{x})$. Suppose $x, x' \in S(p, \overline{y})$ for some p, x, x' sufficiently close respectively to $\overline{p}$ and $\overline{x}$ to allow us to apply the definition of uniform strong subregularity. We have $\overline{y} \in F(p, x) \cap F(p, x')$, which by the definition may happen only if $x = x'$. This means that $S(\cdot, \overline{y})$ has a single-valued localization in a neighborhood of $(\overline{p}, \overline{x})$. By Proposition 2.86 this localization is necessarily Lipschitz. □

The proofs of all implicit function theorems in this section are strikingly simple. Still, the role of the theorems should not be underestimated as they clearly identify the (so far minimal) properties that should be verified in specific situations to get a desired implicit function theorem.

It seems to be reasonable to observe furthermore that the implicit function theorem just proved already contains the classical implicit function theorem as an easy corollary (modulo the Lyusternik–Graves theorem). Indeed, if the spaces are Banach and F is a C^1-mapping along with $F'(\overline{p}, \cdot)(\overline{x})$, the derivative of F with respect to the second argument at $(\overline{p}, \overline{x})$, an invertible operator from X onto Y, then all the assumptions of Theorem 2.91 are satisfied. This is obvious, as far as (2.6.5) is concerned, that $F'(p, \cdot)(x)$ remains a linear homeomorphism of X and Y since F is continuously differentiable by the assumption. The latter also immediately implies condition (b) of Theorem 2.83 and, together with the Lyusternik–Graves theorem, condition (a) of that theorem. Differentiability of the solution mapping and the formula for the derivative remain, of course, beyond the scope of Theorem 2.91, but this is rather a technical and simple part of the proof of the classical theorem.

Of course the Lyusternik–Graves theorem alone implies the classical implicit function theorem. The important point is, however, that a meaningful result containing the principal statements of the classical theorem can be established even in the very general setting of arbitrary metric spaces. The simplicity of the proofs is surprising. But (in addition to the fact that metric theory offers a natural language to treat regularity problems) it is also due to the uniformity assumption, whose verification in principle may not be easy. But for some special classes of mappings (e.g. associated with generalized equations) verification is an easy matter. In the next chapter we shall also consider some verifiable infinitesimal conditions that guarantee the necessary uniformity properties.

2.7 Nonlinear Regularity Models

In this short section we consider regularity models which appear when, say in the case of openness, the radius of a ball around x in X is not proportional to the radius of the neighborhood of $F(x)$ covered by the F-image of the ball. In spite of this, results and arguments involved in the analysis of such models display a lot of similarities with the linear case we have studied so far. Roughly speaking, up to certain technical nuances, the key change reduces to replacement of rt by a certain nonlinear function in the definition of openness and corresponding adjustment of the other two properties. Specifically, we shall consider *gauge functions*, by which we mean any non-negative strictly increasing function on $[0, \infty)$ equal to zero at 0 and continuous on its domain.

Definition 2.92. Given an $F : X \rightrightarrows Y$, where as usual X and Y are metric spaces, let, as before, $U \subset X$ and $V \subset Y$ be open sets, let $\gamma(\cdot)$ be a function on X which is positive on U and let $\delta(\cdot)$ be a function on Y which is positive on V. Assume finally that we are given three gauge functions $\mu(\cdot)$, $\nu(\cdot)$ and $\eta(\cdot)$.

(a) F is *γ-open on (U, V) with functional modulus not smaller than* $\mu(\cdot)$ if the inclusion

$$B(F(x), \mu(t)) \bigcap V \subset F(B(x), t))$$

holds whenever $x \in U$ and $t < \gamma(x)$.

(b) F is *γ-metrically regular on (U, V) with functional modulus not greater than* $\nu(\cdot)$ if the inequality

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

holds whenever $x \in U,\ y \in V,\ \nu(d(y, F(x))) < \gamma(x)$.

(c) F is *δ-Hölder with functional modulus not greater than* $\eta(\cdot)$ if the inequality

$$d(y, F(x)) \leq \eta(d(x, u))$$

holds, provided $x \in U,\ y \in V \bigcap F(u)$ and $\eta(d(x, u)) < \delta(y)$.

If we compare this definition with Definition 2.21 we notice that the only difference is that in the definition of openness rt has been replaced by $\mu(t)$ and in the definitions of metric regularity and pseudo Lipschitz-pseudo Hölder properties Kt has been replaced by $\nu(t)$ and $\eta(t)$, respectively. We also observe that the definition of e.g. nonlinear openness simply means (as $\mu(\cdot)$ is a continuous function) that whenever $x \in U,\ v \in F(x)$, $y \in V$ and $d(y, v) < \mu(\gamma(x))$, there is a u such that $d(x, u) \leq \mu^{-1}(d(y, v))$ if $y \in F(u)$. Finally, it should be mentioned that, as in the linear case, $B(F(x), \mu(t))$ can be equivalently replaced by $\overset{\circ}{B}(F(x), \mu(t))$.

Remark 2.93. Local versions of the definitions follow the same pattern as in the standard linear case. Namely F is *open at* $(\bar{x}, \bar{y}) \in \text{Graph } F$ *with functional modulus not smaller than* $\mu(\cdot)$ if there is an $\varepsilon > 0$ such that the property (a) of Definition 2.92 holds with $\gamma(x) \equiv \varepsilon$ and $V = \overset{\circ}{B}(\overline{y}, \varepsilon)$ whenever $d(x, u) < \varepsilon$ etc.

Exercise 2.94. Prove that the just defined local surjection property at $(\bar{x}, \bar{y})$ with functional modulus not smaller than $\mu(\cdot)$ does not change if we take $\gamma(t) \equiv \infty$.

The equivalence theorem and the general regularity criterion also extend to this set of definitions without much trouble.

Theorem 2.95. *The following properties are equivalent, given a gauge function μ:*

(a) F is γ-open on (U, V) with functional modulus not smaller than μ;

(b) F is γ-metrically regular on (U, V) with functional modulus not greater than μ^{-1};

(c) F^{-1} is γ-Hölder on (V, U) with functional modulus not greater than μ^{-1}.

Proof. The proof essentially repeats the proof of Theorem 2.25. The implication (b) $\Rightarrow$ (c) is trivial. To prove that (c) $\Rightarrow$ (a), let $t < \gamma(x)$, and let $y \in B(F(x), \mu(t))$, that is, $y \in B(v, \mu(t))$ for some $v \in F(x)$. Then $\mu^{-1}(d(y, v)) < \gamma(x)$ and by (c) $d(x, F^{-1}(y)) \le \mu^{-1}\big(d(y, v)\big)$. In other words, there is a $u \in X$ such that $y \in F(u)$ and $d(x, u) \le \mu^{-1}(d(y, v)) \le \mu^{-1}(\mu(t))) = t$. As y is an arbitrary element of $B(F(x), \mu(t)) \bigcap V$, (a) follows.

(a) $\Rightarrow$ (b). Take an $x \in U$ and $y \in V$. Let $\mu^{-1}(d(y, F(x))) < \gamma(x)$. Take a t and $\varepsilon > 0$ to make sure that $\mu^{-1}(d(y, F(x)) + \varepsilon) = \mu(t) < \gamma(x)$. Then $y \in B(F(x), \mu(t))$ and by (a) there is a u such that $y \in F(u)$, $d(x, u) \le t = \mu^{-1}(d(y, F(x)) + \varepsilon)$. As μ is continuous and ε can be arbitrarily small, it follows that $d(x, F^{-1}(y)) \le \mu^{-1}\big(d(y, F(x))\big)$. ◻

We shall next prove a "nonlinear analogue" of the general regularity criterion. Let again $U \subset X$ and $V \subset Y$ be open sets and let $\gamma(x)$ be a Lipschitz function on X with Lipschitz constant 1. We set as before

$$U_\gamma = \bigcup_{u \in U} \overset{\circ}{B}(x, \gamma(x)).$$

Let further a gauge function $\mu(t)$ be given.

Theorem 2.96 (nonlinear regularity criterions). *Let $F : X \rightrightarrows Y$ be a set-valued mapping whose graph is complete in the product metric. Then F is γ-open on (U, V) with functional modulus not smaller than μ if the following holds: there is a $\xi > 0$ such that for any $x \in U_\gamma$, $y \in V$, $v \in F(x)$ with $0 < d(y, v) < \gamma(x)$ there is a pair $(u, w) \in \operatorname{Graph} F$, $(u, w) \ne (x, v)$ such that*

$$\mu^{-1}\big(d(y, w)\big) \le \mu^{-1}\big(d(y, v)\big) - d_\xi((x, v), (u, w)). \tag{2.7.1}$$

This condition applied only to $x \in U$ is also necessary for γ-openness of F on (U, V) with the functional modulus not smaller than μ, provided there is a $\xi > 0$ such that $\xi\tau \le \mu^{-1}(\tau)$ for all $\tau \in [0, \mu(\gamma(x))]$ and all $x \in U$.

Proof. As in the case of the equivalence theorem the proof follows the lines of the proof of Theorem 2.46. Given a $y \in Y$, set $\psi_y(x, v) = \mu^{-1}(d(y, v)) + i_{\text{Graph } F}(x, v)$. Let $x \in U$, $y \in V$, $v \in F(x)$ and $d(v, y) < \mu(\gamma(x))$. Set $\varepsilon = \psi_y(x, v)$ and find, using Ekeland's principle, a pair $(\hat{u}, \hat{w})$ such that

$$\begin{aligned}
&d_\xi((x, v), (\hat{u}, \hat{w})) \le \varepsilon;\\
&\psi_y(\hat{u}, \hat{w}) \le \psi_y(x, v) - d_\xi((x, v), (\hat{u}, \hat{w}));\\
&\psi_y(u, w) + d_\xi((u, w), (\hat{u}, \hat{w})) > \psi_y(\hat{u}, \hat{w}), \quad \text{if } (u, w) \ne (\hat{u}, \hat{w}).
\end{aligned} \tag{2.7.2}$$

Then $\psi_y(\hat{u}, \hat{w})) = 0$, that is, $\hat{w} = y$. Assuming the contrary we shall get a contradiction as in the proof of Theorem 2.46. To this end, we first note that $d(x, \hat{u}) < \gamma(x)$. This is a consequence of the second relation in (2.7.2), which implies that $d(x, \hat{u}) \leq \psi_y(x, v)$. The inequality means that $\hat{u} \in U_\gamma$ and therefore by (2.7.1) there is a $(u, w) \neq (\hat{u}, \hat{w})$ in Graph F such that

$$\mu^{-1}\big(d(y, w)\big) \leq \mu^{-1}\big(d(y, \hat{w})\big) - d_\xi((\hat{u}, \hat{w}), (u, w)).$$

The latter contradicts the last relation in (2.7.2).

The starting point of the proof of the second statement is also similar to that in the proof of Theorem 2.46. So suppose F is γ-open with functional modulus not smaller than $\mu(t)$. Given $x,\ y,\ v$ as above, we find a u such that $y \in F(u)$ and $d(x, u) \leq \mu^{-1}(d(y, v))$. Let further $\xi > 0$ satisfy the conditions specified in the statement. Then,

$$d_\xi((x, v), (u, y)) = \max\{d(x, u), \xi d(v, y)\} \leq \mu^{-1}(d(y, v)),$$

so that setting $w = y$, we get

$$0 = \mu^{-1}(d(y, w)) \leq \mu^{-1}(d(y, v)) - d_\xi((x, v), (u, w)),$$

as claimed. $\square$

Remark 2.97. It is worth noting that the theorem contains the general regularity criterion of Theorem 2.46. To see this, we only need to take $\mu(t) = rt$.

Theorem 2.98 (nonlinear density theorem). *Let $U \subset X$ and $V \subset Y$ be open sets, let $F : X \rightrightarrows Y$ be a set-valued mapping with complete graph, and let γ be a Lipschitz function on X with Lipschitz constant one, positive on U and equal to zero on the boundary of U. We assume that there are a gauge function $\mu(t)$, a $\lambda \in (0, 1)$ and an $\eta > 0$ such that $\mu(\eta t) \leq t$ for all $t < \sup_U \gamma(x)$ and for any $x \in U$ and $t < \gamma(x)$ the set $F(B(x, t))$ is a $\lambda\mu(t)$-net in $B(F(x), \mu(t)) \bigcap V$.*

Let $\nu(t)$ be another gauge function satisfying

$$\nu^{-1}(\tau) - \nu^{-1}(\lambda\tau) \geq \mu^{-1}(\tau), \quad \forall\, t \in (0, \gamma(x)),\ \forall\, x \in U. \tag{2.7.3}$$

Then F is γ-open on (U, V) with functional modulus not smaller than $\nu(t)$.

Proof. Note that $\nu^{-1}(t) \geq \mu^{-1}(t)$ and therefore $\nu(t) \leq \mu(t)$. Let $x \in U,\ y \in V,\ v \in F(x),\ d(y, v) < \mu(\gamma(x))$. Set $t = \mu^{-1}(d(y, v))$. Then $t < \gamma(x)$. By the assumption there is a $(u, w) \in$ Graph F such that $d(x, u) \leq t$ and

$$d(y, w) \leq \lambda t = \lambda\mu^{-1}(d(y, v)). \tag{2.7.4}$$

We have by (2.7.3)

$$\begin{aligned}\nu^{-1}(d(y,w)) &\leq \nu^{-1}(\lambda d(y,v))\\ &= \nu^{-1}(d(y,v)) - (\nu^{-1}(d(y,v)) + \nu^{-1}(\lambda d(y,v)))\\ &\leq \nu^{-1}(d(y,v)) - \mu^{-1}(d(y,v)).\end{aligned} \tag{2.7.5}$$

On the other hand, $\mu^{-1}(d(y,v)) = t \geq d(u,x)$ and by (2.7.5) $d(v,w) \leq (1+\lambda)d(y,v)$, so that $\xi d(v,w) \leq (1+\lambda)\xi d(y,v) \leq \mu^{-1}(d(y,v))$ if $\xi(1+\lambda) < \eta$, which, together with (2.7.5), gives $\nu^{-1}(d(y,w)) \leq \nu^{-1}(d(y,v)) - d_\xi((x,v),(u,w))$. A reference to Theorem 2.96 completes the proof. □

In particular, we can look for $\nu(t)$ of the form $\nu(t) = \rho\mu(t)$ for some $\rho < 1$. Then $\nu^{-1}(t) = \mu^{-1}(t/\rho)$ and (2.7.3) reduces to

$$\mu^{-1}(\rho^{-1}t) - \mu^{-1}(\rho^{-1}\lambda t) \geq \mu^{-1}(t).$$

In the most interesting case of $\mu(t) = rt^k$ we get

Corollary 2.99. *If under the assumptions of Theorem 2.98, $\mu(t) = rt^k$, then F is γ-open on (U,V) with the functional modulus not smaller than $r(1-\lambda^{1/k})^k t^k$.*

Observe that this corollary implies Theorem 2.55 if we set $k = 1$ and $\lambda = \ell/r$. In the last corollary we deal with local regularity.

Corollary 2.100. *Let $F : X \rightrightarrows Y$ be a set-valued mapping whose graph is locally complete at $(\bar{x},\bar{y}) \in \operatorname{Graph} F$. Assume that there are an $\varepsilon > 0$ and a functional modulus $\mu(\cdot)$ such that for any $x \in \overset{\circ}{B}(\overline{x},\varepsilon)$ and $t \in (0,\varepsilon)$ the set $F(B(x,t))$ is dense in $B(F(x)\mu(t)) \bigcap \overset{\circ}{B}(\overline{y},\varepsilon)$. Then F is open at $(\bar{x},\bar{y})$ with functional modulus not smaller than $\mu(\cdot)$.*

2.8 Supplement: Regularity of a Composition

We have seen in Example 2.82 that a sum of a regular and a pseudo-Lipschitz set-valued mapping can fail to be regular even if the Lipschitz modulus of the second mapping is much smaller than the rate of surjection of the first. The example can be easily modified to show that a composition of two regular set-valued mappings can fail to be regular.

Example 2.101. Let $X = Y = I\!R$ and

$$F(x) = \begin{cases} \{0\}, & \text{if } x = 0;\\ \{x, x^{-2}\}, & \text{if } x \neq 0; \end{cases} \qquad G(y) = \begin{cases} \{0\}, & \text{if } y = 0;\\ \{y, y^{-1}\}, & \text{if } y \neq 0. \end{cases}$$

Clearly, both F and G are γ-regular on $(0,1)\times(0,1)$ with $\gamma(t) = 1 - |t|$ and the rate of surjection equal to one. On the other hand

$$(G \circ F)(x) = \begin{cases} \{0\}, & \text{if } x = 0, \\ \{x, x^2, x^{-1}, x^{-2}\}, & \text{if } x \neq 0 \end{cases}$$

is not regular even near $(0, 0)$ because of the x^2 component.

The cause of the phenomenon is that, staying within the (U, W)-boundaries, we cannot control the values of y that contribute to $(G \circ F)(x)$ for $x \in U$. This puts forward the problem of regularity of a composition as such. Clearly, we have to impose some additional assumptions to get regularity of compositions. Below we prove three propositions of that sort. As we shall see, the additional assumptions are sufficiently strong but they can hardly be substantially weakened.

Proposition 2.102. *Let $F : X \rightrightarrows Y$ be γ-regular on (U, V) for some $\gamma(\cdot)$ with $\mathrm{sur}_\gamma(U|V) > r$. Let further $G : Y \rightrightarrows Z$, and let the restriction $G|_V$ of G to V be δ-regular on (V, W) with $\delta \equiv const > 0$ and $\mathrm{sur}\,_\delta G|_V(V|W) > s$. Set $\eta(x) = \min\{\gamma(x), \delta/r\}$. Then $\Phi = G|_V \circ F$ is η-regular on (U, W) and $\mathrm{sur}\,_\eta \Phi(U|W) \geq rs$.*

Proof. Let $(x, z) \in \mathrm{Graph}\,\Phi$. This means that $z \in G(y)$ for some $y \in F(x) \cap V$. Take a $t < \eta(x)$ and set $\tau = rt$. As $\tau < \delta$, we have $B(z, s\tau) \cap W \subset G(B(y, rt) \cap V)$. But $t < \gamma(x)$, so by regularity of F we have $B(y, rt) \cap V \subset F(B(x, t)$. Thus $B(z, s\tau) \cap W \subset F(B(x, t)$. □

It should be emphasized that the condition that the restriction of G to V (rather than G itself) is δ-regular on (V, W) is fairly strong. A typical situation when restriction kills regularity occurs when (e.g. for F regular on (U, V)) the image of a point close to the boundary of U contains points deeply inside V (see e.g. Example 2.38).

Proposition 2.103. *Assume that $F : X \rightrightarrows Y$ is γ-regular on (U, V) with $\mathrm{sur}_\gamma(U|V) > r > 0$ and $G : Y \rightrightarrows Z$ is δ-regular on $V \times W$ with $\mathrm{sur}\,_\delta G(V|W) > s$ and $\delta(y) \leq d(y, Y \backslash V)$. Set for $\xi > 0$*

$$\theta(\xi) = \inf\{r^{-1}\delta(y) : \ y \in F(x) \cap V, \ x \in U, \ \gamma(x) \geq \xi\}, \quad \eta(\xi) = \min\{\xi, \theta(\xi)\}$$

and assume that $\eta(\xi) > 0$ for any $\xi > 0$. Finally, set $F^V(x) = F(x) \cap V$. Then $\Phi = G \circ F^V$ is $(\eta \circ \gamma)$-regular on $U \times W$ with $\mathrm{sur}\,_{\eta\circ\gamma} \Phi(U|W) > rs$.

Proof. Let $x \in U, \ z \in W$ and $z \in \Phi(x)$. This means that there is a $y \in F(x) \cap V$ such that $z \in G(y)$. Let z' satisfy $d(z', z) < rs(\eta \circ \gamma)(x)$, that is to say, there is a $t < (\eta \circ \gamma)(x)$ such that $d(z', z) < rst$. Set $\tau = rt$. Then $\tau < \delta(y)$ by definition of η and $d(z', z) < s\tau$. As G is δ-regular on (V, W), there is a y' such that

$$z' \in G(y') \quad \& \quad d(y', y) < s^{-1} d(z', z) < \tau < \delta(y) \leq d(y, Y \backslash V).$$

This means that $y' \in V$. On the other hand, as $d(y', y) < \tau = rt$ and F is γ-regular on (U, V), there is an x' with $d(x', x) < t$ such that $y' \in F(x')$. As y' is also in V, we have $y' \in F(x') \cap V$ and therefore $z' \in \Phi(x')$. □

Here is a typical situation when the composition may fail to be regular if F maps some points inside U to an arbitrary small vicinity of the boundary of V.

Example 2.104. Let again $X = Y = I\!R^2$, $U = V = \overset{\circ}{B}$. Let further F be itself a composition $F_2 \circ F_1$, where F_1 is a folding of the plane: $(\xi, \eta) \to (|\xi|, \eta)$ and F_2 is Lipschitz and maps the boundary of the semicircle $B \cap \{(\xi, \eta) : \ \xi \geq 0\}$ to the boundary of the unit circle with the interior of the first going into the interior of the second. It is clear that, no matter which G we take, we shall have $\eta(x) \equiv 0$.

The simplest case when the proposition works corresponds to $V = Y$ and $\delta(y)$ identically equal to a sufficiently big constant, possible even to ∞. This immediately implies

Corollary 2.105. *Assume that $F : X \rightrightarrows Y$ is γ-regular on $U \times Y$ with* $\operatorname{sur}_{\gamma}(U|V) > r > 0$ *and $G : Y \rightrightarrows Z$ is δ-regular on $Y \times W$ with* $\operatorname{sur}_{\delta}G(V|W) > s$ *and $\delta(y) \equiv const \geq r$. Then $G \circ F$ is γ-regular on $U \times W$ with* $\operatorname{sur}_{\gamma}(U|W) \geq rs$.

If V is distinct from the whole of Y, $\eta(\xi)$ can be identical zero if the values of F are sufficiently big. This is the weak point of Proposition 2.103.

Proposition 2.106. *Suppose that there are open $U' \subset U$, $V' \subset V$ and $W' \subset W$ and a $\delta > 0$ such that*

*(**a**) $B(V', \delta) \subset V$ and*

*(**b**) for any $(x, z) \in U' \times W'$ such that $z \in (G \circ F)(x)$ there is a $y \in F(x) \cap V'$ such that $z \in G(y)$.*

If under these conditions F is γ-regular on (U, V) with rate of surjection $r > 0$, where γ is a positive number, and G is δ-regular on $V' \times W'$ with rate of surjection $s > 0$, then $G \circ F$ is ξ-regular on $U \times W'$ with rate of surjection rs for any $\xi < \min\{r^{-1}\delta, \gamma\}$.

Proof. The proof is similar to the proof of Proposition 2.103. Let $z \in (G \circ F)(x) \cap W'$ for some $x \in U'$. By (**b**) there is a $y \in F(x) \cap V'$ such that $z \in G(y)$. Let further $t \leq \xi$ and $d(z', z) < rst = s\tau$ where $\tau = rt$. As $\tau < \delta$ and G is δ-regular on $V' \times W'$, there is a y' such that $z' \in G(y')$ and $d(y', y) \leq s^{-1}d(z'z) < \tau < \delta$. It follows from (**a**) that $y' \in V$ and by regularity of F there is an x' such that $y' \in F(x')$ and $d(x', x) \leq r^{-1}d(y',) \leq (rs)^{-1}d(z'z) < t$. Thus $z' \in (G \circ F)(x')$ and $d(x', x) < \gamma$. □

A local version of property (b) of the proposition for given $\bar{y} \in F(\bar{x})$, $\bar{z} \in G(\bar{y})$ reads as follows:

(**b**) loc For any $\varepsilon > 0$ there is a $\delta > 0$ such that for any $(x, z) \in$ Graph $(G \circ F)$ satisfying $d(x, \bar{x}) < \delta$, $d(z, \bar{z}) < \delta$ there is a $y \in Y$ such that $d(y, \bar{y}) < \varepsilon$, $y \in F(x)$ and $z \in G(y)$.

It is said that F and G are *composition stable* at $(\bar{x}, \bar{y}, \bar{z})$ if property (**b**) loc is satisfied.

The following corollary is immediate from Proposition 2.106.

Corollary 2.107. *Assume that F and G are composition stable at $(\bar{x}, \bar{y}, \bar{z})$. If F is regular at $(\bar{x}, \bar{y})$ with* sur $F(\overline{x}|\overline{y}) \geq r$ *and G is regular at $(\overline{y}, \overline{z})$ with* sur $G(\overline{y}, \overline{z}) \geq s$, *then $G \circ F$ is regular at $(\overline{x}, \overline{z})$ with* sur $(G \circ F)(\overline{x}|\overline{z}) \geq rs$.

Proposition 2.106 and the corollary do not look very practical because they establish some coordination between the two mappings involved in the composition. However, they do have corollaries with independent conditions on the mappings.

Corollary 2.108. *Let $F : X \rightrightarrows Y$, $G : Y \rightrightarrows Z$, $\overline{z} \in G(\overline{y})$, $\overline{y} \in F(\overline{x})$ and* sur $F(\overline{x}|\overline{y}) \geq r$, sur $G(\overline{y}|\overline{z}) \geq s$. *Then* sur $(G \circ F)(\overline{z}|\overline{x}) \geq rs$ *in the following cases:*

(a) if F is single-valued continuous or, more generally, if $F(\overline{x}) = \{\overline{y}\}$ and F is upper semicontinuous at $\overline{x}$;

(b) if $G(\overline{y}) = \{\overline{z}\}$ and there is a gauge function μ such that $d(z, \overline{z}) \geq \mu(d(y, \overline{y}))$ if $z \in G(y)$;

(c) if G is strongly regular at $(\overline{y}|\overline{z})$.

2.9 Comments

Introduction. The classical reference for set-valued mappings is the 1959 monograph by Berge [31]. More recent accounts of the basic properties of set-valued mappings fully covering our needs can be found in the monographs by Aubin and Frankowska [14] and Rockafellar and Wets [287].

For the original publications of Ekeland's principle see [112, 113]. In the proof we follow [164]. An important fact to keep in mind is that the variational principle of Ekeland is a characteristic property of complete metric spaces, namely, a metric space is necessarily complete if the variational principle is valid for lsc functions on the space [301] (see also [149]. For the most general formulation of Ekeland's principle, see Borwein and Zhu [49]. Much additional information can be found in Hyer, Isac and Rassias [149].

The Bishop–Phelps theorem was proved in [32] and its extension by Bollobás in [35] (see [267]).

Section 2. The development of regularity theory after Lyusternik and Graves actually started much later, in the very late 60 s and early 70s, when the power of the Lyusternik theorem and its potential role in optimization theory were first emphasized by Milyutin. The choice of the Lyusternik theorem as the main instrument of analysis in optimization theory in our 1974 book with Tikhomirov [189] was substantially influenced by conversations with him. The appearance of Ekeland's variational principle [112] was definitely a turning point in the developments (see e.g. [10, 150]). Still, the idea remained for some time rather foreign, even in spite of the series of Robinson's seminal papers of 1975–76.[4]

[4]See, for instance, Ekeland's comments in [113], where the main result of [150] was interpreted as a nonsmooth mean value theorem.

The systematic study of regularity in the context of metric spaces was initiated in 1980 by Dmitruk–Milyutin–Osmolowskii [82] in a paper dedicated to Lyusternik's 80th anniversary. The main concept studied in the paper is close to what is called here openness in the sense of Milyutin.[5] But we should mention first the papers by Ptak [271], Tziskaridze [305] from the mid-1970s, where open mapping theorems in metric spaces in the spirit of Corollary 2.56 were first considered (see also [198], p. 202 for a still earlier result of this sort).

All mentioned works dealt exclusively with the openness property. The distance estimates in general, but close to classical contexts, appeared in the book by Ioffe–Tikhomirov [189] published in 1974 and Robinson's 1976 paper [276]. Robinson was also the first to consider set-valued mappings in the context of regularity theory. The very concept of "metric regularity" was being gradually worked out and took its final form in the late 1980s: see [41, 50, 150, 261]. The Aubin (local pseudo-Lipschitz) property was introduced by Aubin in 1984 in [11] for set-valued mappings between Banach spaces. The term "Aubin's property" was suggested by Dontchev and Rockafellar in [94]. We prefer to use the original pseudo-Lipschitz terminology in the non-local settings to avoid possible confusion. The original definition had the form (4) (in Banach terms). The equivalence with the Lipschitz-type estimate we use here in Definition 2.21(c) was noted by Rockafellar [286].

It also has to be mentioned that definitions of local linear openness vary considerably in the literature. For instance, in [50, 166, 246] it was defined by means of the inclusion

$$B(F(x) \cap V, rt) \subset F(B(x,t)), \quad \text{if } x \in U,\ 0 \le t < \varepsilon$$

(rather than $B(F(x), rt) \cap V$, as we have done here). Fortunately, for local openness (when the neighborhoods for which the relations hold are not fixed) this is an equivalent definition. But they are no longer equivalent if the neighborhoods are fixed.

The vast majority of studies of metric regularity concentrate on local regularity. However, Milyutin's original definition of openness in [82] was rather global and applied to single-valued mappings. The three properties for the case of fixed sets were introduced in [175] (with a slightly different terminology).

The understanding of the equivalence of linear openness and metric regularity, and later the Aubin property, was being developed at the same time as the understanding of the properties themselves. Explicit mention of the equivalence of the first two can already be found in [82]. The final results were obtained by Borwein–Zhuang [50] and Penot [261] for the local setting. We also mention the paper of Frankowska [130], which has a short proof of equivalence of local openness and the inverse pseudo-Hölder property. It has to be mentioned that nonlinear regularity properties

[5] In [82] the authors considered systems of balls in X which are 'full' in the sense that all balls contained in an element of the system also belong to the system, and called a (single-valued continuous) mapping $F : X \to Y$ an *a-covering* on the system if $B(F(x), at) \subset F(B(x,t))$ whenever $B(x,t)$ belongs to the system.

were considered in the three quoted papers. The equivalence for the case of fixed sets (Theorem 2.25) was proved in [175].

Various proofs of equivalences of local and certain non-local properties can now be found in numerous publications, including surveys and monographs, e.g. [96, 166, 200, 246, 287]. We mention in this connection an earlier result stated in [154] without proof (for a simple proof see [164]), which is in a sense the most precise because it gives a sort of stratification over points of the graph and not just estimates averaged over a neighborhood. It gives yet another approach to the question, both for linear and nonlinear regularity. For an $F : X \rightrightarrows Y$ and $(x, v) \in \text{Graph } F$ consider two functions on $[0, \infty)$:

$$S_F(x, v, t) = \sup\{r \geq 0 : \ B(v, r) \subset F(B(x, t))\};$$

$$R_F(x, v, t) = \inf\{\eta > 0 : \ d(x, F^{-1}(y)) \leq \eta, \ \text{if } d(y, v) \leq t\}.$$

Exercise 2.109 ([154], Proposition 11.12). Prove that

$$R_F(x, v, t) = \inf\{\xi > 0 : \ S_F(x, v, \xi) > t\}.$$

The equality basically says that the maximal monotone mappings $I\!R \rightrightarrows I\!R$ generated by $S_F(x, y, \cdot)$ and $R_F(x, y, \cdot)$ are mutually inverse for each point of the graph of any set-valued map.[6] To see that this implies the equivalence of local linear openness and metric regularity, it is enough to note that whenever μ is a functional modulus satisfying $\mu(t) < S_F(x, v, t)$ for all $(x, v) \in \text{Graph } F$ in a neighborhood of $(\bar{x}, \bar{y})$ and all small positive t, then for some $\varepsilon > 0$ F is ε-open at $(\bar{x}, \bar{y})$ with functional modulus not smaller than μ (cf. Definition 2.92(a)). Likewise if $\eta(t) > R_F(x, y, \cdot)$ for all (x, y) in a neighborhood of $(\bar{x}, \bar{y})$, then there is an $\varepsilon > 0$ such that F is ε-metrically regular at $(\bar{x}, \bar{y})$ with functional modulus not smaller than ν. Finally, the proposition in the exercise says that $\mu(t) < S_F(x, v, t)$ for all $(x, v) \in \text{Graph } F$ in a neighborhood of $(\bar{x}, \bar{y})$ and e.g. $t \in (0, \varepsilon)$ is the same as $\mu^{-1}(t) > R_F(x, v, t)$ for the same (x, v) and $t < \mu(\varepsilon)$.

Exercise 2.110. Prove that (cf. Exercise 2.109)

$$\text{contr } F(x|y) = \liminf_{t \to +0} t^{-1} S_F(x, y, t); \quad \text{recess} F(x|y) = \limsup_{t \to +0} t^{-1} R_F(x, y, t).$$

The equivalence of local metric regularity and graph regularity was discovered by Thibault [303]. Note that this equivalence cannot be extended to the case of fixed sets. The equivalence property of Proposition 2.18 was established by Ngai, Tron and Théra in [258] (Theorem 3(ii)). The other two equivalent results, Propositions 2.33 and 2.40, first appeared in [166].

[6] In [154, 164] single-valued mappings between Banach spaces were considered. But the proof in [164] carries over to set-valued maps between metric spaces with minor changes.

Section 3. The central role of the general regularity criterion of Theorem 2.46 in regularity theory was emphasized in the 2000 survey of Ioffe [166]. A surprising and recently discovered fact (see [59]) is that a version of the criterion was first mentioned in a 1987 paper by Fabian and Preiss [122] (Remark 2(c)) who, however, never used it (to such an extent that even one of the authors, at least, forgot about it!), and the result remained unknown until it was rediscovered in [166]. A close version of the criterion for the case when the function $d(y, F(\cdot))$ is lower semicontinuous (in the form of a criterion for non-regularity) can also be found in a 1999 paper by Kummer [214].

The Basic Lemma 2.42 has considerable independent interest, especially in connection with error bounds. The lemma was originally proved in [166] but its sources can be traced back to [150] and to a paper by Cominetti [69] who made a substantial step forward and proved a result close to the Basic Lemma, with two notable differences. The function considered in [69] was not assumed to be lower semicontinuous, but instead satisfying an additional stronger requirement that $\psi(x) \leq (1-r)\psi(u)$ (in the notation of the Basic Lemma). Thanks to this assumption, the proof of the distance estimate was obtained in [69] using simple Picard-type iterations that can hardly be used under the assumptions of the Basic Lemma. It is not clear whether, for example, the density theorem can be proved using Cominetti's result (although it can be proved with the help of simple iterations as in [82]).

In fact, the proof of the Basic Lemma in [166] was an immediate reaction to the introduction of slope by Azé, Corvellec and Lucchetti in [23]. (Although [166] appeared in print before [23], I was acquainted with the first version of [23] while preparing [166].) The criteria of Theorems 2.49 and 2.50 are new, although the first was essentially inspired by the mentioned paper by Ngai, Tron and Théra [258] and the second by the same Cominetti's paper [69]. Theorem 2.54 appears here for the first time. But parts (a) and (b) of the theorem are immediate consequences of Proposition 2.29 and the corresponding non-local criteria for Milyutin regularity. A proof of part (c) will appear in a paper by Fabian, Ioffe and Revalski with the tentative title "Separable reduction of local metric regularity".

According to my experience, the three criteria are an extremely convenient tool when verifying regularity (even the most convenient in many cases), in particular, because application of the criteria does not require any preliminary calculations (e.g. of slopes or subdifferentials).

The density phenomenon (F covers if $F(B(x), \varepsilon)$ is dense in $B(F(x), \delta)$ for some ε and δ) has been extensively discussed, especially at the early stage of development. 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, whose key component was to show that a linear continuous image of the unit ball contains an open ball if it (the image) is dense in the latter. Ng [255] extended the result to homogeneous mappings between certain classes of vector spaces. For more general classes of mappings, results in the spirit of Corollary 2.58 were first considered in Ptak [271], Tziskaridze [305] and Dolecki [83] in the mid-1970s (see also [308] for a somewhat more detailed study of the phenomenon in the global setting). We refer to [19] for detailed discussions, many references and further results. In the proof of the density theorem we follow [175].

Dmitruk–Milyutin–Osmolovski in [82] made a substantial step forward when they replaced (in the non-local 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, which (along with its extensions) plays a central role in the theory. A similar advance in the framework of the infinitesimal approach (for mappings between Banach spaces) was made by Aubin [10] (see also the comments to Chap. 5).[7]

A further step was made by Khanh [199], who found a unified approach more or less containing all mentioned density results as particular cases. Here is his theorem:

Theorem 2.111 (Khanh). *Let X and Y be metric spaces, let $F : X \rightrightarrows Y$ be a set-valued mapping with complete graph, and let $\rho(x)$ be a positive Lipschitz function on X with Lipschitz constant one. We assume that for any $t > 0$ there are two sequences $(a_n(t))$ and $(b_n(t))$ of positive numbers such that $\sum_n a(n(t)) < t$ and $b_n(t) \to 0$ and the following holds: if $0 < t < \rho(x)$*

$$F(B(x, a_n(t)) \quad \text{is a} \ \ b_{n+1}(t)\text{-net for } B(F(x), b_n(t)).$$

Then the inclusion

$$B(F(x), b_1(t)) \subset F(B(x, t))$$

holds whenever $t < \rho(x)$.

We observe that in the last theorem, as in many of the mentioned results, nonlinear openness in the spirit of Sect. 6 is considered. Here we just mention that if in Khanh's theorem $b_1(t) = rt$, then the theorem states that F is regular with $\gamma(x) = r\rho(t)$ and $\mathrm{sur}\,_{\gamma} F \geq r$. For some further results in this direction, see [19].

Section 4. For mappings between Banach spaces, calmness and metric subregularity (originally under different names) have been thoroughly studied in the variational analysis literature. We refer to [96] for a most detailed finite-dimensional account. The term "calmness" was introduced in [62] for what in today's terminology would be calmness of the epigraphic mapping of an extended-real-valued function. For general set-valued mappings the concept of calmness was probably introduced in [281] under the name "upper Lipschitz property" and the concept of subregularity in [166] under the name "regularity at a point". The very term "subregularity" seems to have appeared in [95]. Strong subregularity was first explicitly defined in [96]. The concept of controllability seems to have been mentioned here for the first time in connection with subregularity.

Section 5. The first theorem concerning the effect of perturbations on regularity rates was published without proof in [81] and proved four years later in [82]. The authors

[7] In [11] Aubin mentioned that a result close to his had been established in the thesis of G. Lebourg which, as I understand, was never published.

of both papers attribute the result to Milyutin. The theorem states the following: *if $F : X \to Y$ (where X is a complete metric and Y is a Banach space) is a continuous mapping a-covering on a full system Σ of balls and $G : X \to Y$ is b-Lipschitz with $b < a$, then $F+G$ is $(a-b)$-covering on Σ.*[8] It is an easy matter to see that Theorem 2.79 is basically a set-valued extension of this result.

A local version of the theorem was proved in [158] and the set-valued extension of Milyutin's theorem was first obtained by Ursescu in [307] in the fully global setting – see Theorem 3.45 in the next chapter. A shorter proof of the theorem was given in [166] with an explicit mention that the proof cannot work for local regularity. A counterexample (of which Example 2.82 is a slight modification) was given in the first 2008 edition of [96]. Certain results showing that a local version of Milyutin's theorem for set-valued mappings with a set-valued perturbation may hold under additional assumptions that require a certain level of coordination between the given mapping and its perturbation can be found in [99] and [258]. Theorem 2.72 is a new result. It seems to be a natural extension of Milyutin's theorem to a purely metric setting.

The role of Milyutin's theorem and its variants and extensions in the theory can hardly be overestimated. We refer to the recent monograph of Dontchev and Rockafellar [96], where the theorem and its extensions, variants and consequences are central. It should be mentioned that in [96] a result equivalent to Milyutin's theorem (Theorem 5E.1 in [96]) is called the "Extended Lyusternik–Graves Theorem". There is a certain justification for this. Indeed, Graves' theorem can be viewed as a perturbation theorem for a linear operator. But it is quite clear from Graves' paper that he considered the linear mapping to be only a certain approximation of the primal object of interest, not an independent object subject to perturbations.[9] To emphasize the difference between the theorems of Graves and Milyutin, we mention that the necessity of the Lyusternik–Graves condition in Corollary 2.80, first observed by Dontchev in [87], follows from Milyutin's theorem but not from Graves' theorem. All of this leads me to believe that "Milyutin's Theorem" is the most adequate name for the result.

Section 6. Extensions of inverse and implicit function theorems to nonsmooth and set-valued settings occupy a distinguished place in variational analysis. See e.g. [14, 18, 86, 96, 109, 130, 131, 166, 201, 203, 213, 218, 236, 246, 257, 258, 264, 282]; the earliest was probably [270]. Many results in this direction are collected in a recent monograph by Dontchev and Rockafellar [96]. Often, however, rather a loose interpretation applies to the very concepts of an inverse or an implicit function. We

---

[8] See the footnote on p. 99.

[9] It is certain that Graves' paper was not known to Milyutin in 1976. To the best of my knowledge the first reference to it in the Russian literature appeared in [82] in 1980. Moreover, I doubt that Graves' theorem was known and/or used by the optimization community before 1980. It seems (curiously enough) that at the earlier stage of development of regularity theory all the main ideas appeared independently. Lyusternik in 1934 was likely unaware of the Banach open mapping theorem. Graves in 1950 knew Banach's result but apparently not Lyusternik's theorem. Tikhomirov and myself knew Lyusternik's theorem while writing the 1974 book, but not the result of Graves, and Robinson in 1976 seemed to have been unaware of the papers by Lyusternik, Graves and our book.

mentioned already that the equivalence theorem has a flavor of an inverse function result. For that reason any result which offers a sufficient condition for regularity of a mapping under such an approach may be considered an inverse function theorem (see e.g. [96], p. 184). Moreover, in some publications statements containing the inequality

$$d(x, S(p, \overline{y})) \le Kd(\overline{y}, F(x, p)),$$

proved under certain conditions for a certain set of parameters, are called implicit function theorems. I believe, however, that the real content of the implicit function paradigm is that the solution mapping inherits some good properties from F. This is rather a common idea (see e.g. [203], p. 8).

Theorems 2.83 and 2.84 express this idea in a very clear way. Although the proofs of the theorems are similar, we emphasize that the theorems are independent. The fact that they deal with different sets of parameters is only a part of the reason. The subtle point is that even if $F(p, \cdot)$ is uniformly subregular at $(\overline{x}, y)$ for every $y \in F(\overline{p}, \overline{x})$, the epsilons produced by Theorem 2.84 may fail to be bounded away from the origin, which makes it difficult to deduce Theorem 2.83 from Theorem 2.84.

Theorem 2.84 should be attributed to Ledyaev and Zhu [218], who de facto established the result for set-valued mappings between Fréchet smooth Banach spaces. Their paper does not contain the statement but the result can be easily extracted from the proof of Corollary 3.9 in the paper. The first explicit mention of subregularity of the x-behavior of the mapping in an implicit function theorem can probably only be found in a recent paper by Gfrerer and Outrata [135] in the finite-dimensional context. Theorem 2.83 was stated and proved by Ioffe in [166] (as Lemma 1 in Sect. 1.3) already for set-valued mappings between metric spaces.

We shall return to discussions of further developments in the comments to Chaps. 3 and 7. Here we just mention that the uniformity of regularity and the pseudo-Lipschitz property (especially the first), mainly responsible for the elementary character of the proof of the theorem, turns out to be a rather non-trivial concept and further developments are essentially connected with looking for ways to characterize and verify it.

The concept of strong regularity was introduced by Robinson in 1980 [280], in connection with generalized equations. For general set-valued mappings, especially in finite-dimensional spaces, it was extensively studied by Dontchev and Rockafellar – (see [96] and references therein). But even earlier (see [166], Corollary 7.1 in Chap. 1) strong regularity in the form of Property (b) of Proposition 2.86 was used to get an implicit function theorem for general metric spaces. In Robinson's original definition strong regularity was defined through the existence of a Lipschitz localization of the inverse mapping. The specific "quantitative" form of the definition (Definition 2.85) seems to appear here for the first time. Persistence of strong regularity under additive perturbations of a mapping into a linear metric space with shift-invariant metric was established in the first edition of [96] (Theorem 5F.1) for additive perturbations of a mapping into a linear metric space with shift invariant metric (see also [85] for any earlier results in this vein).

Proposition 2.90 is a new result. But its additive counterpart, namely that $F + g$ is strongly subregular at $(\overline{x}, \overline{y} + g(\overline{x}))$ and

$$\text{subreg}(F + g)(\overline{x}|\overline{y} + g(\overline{x})) \le \frac{K}{1 - K\ell},$$

if F is strongly subregular at $(\bar{x}, \bar{y})$ with subreg$F(\overline{x}|\overline{y}) \ge K$ and g is single-valued Lipschitz with lip $g(\overline{x}) \le \ell < K^{-1}$, can also be found in the first edition of [96].

Section 7. The study of nonlinear regularity models was initiated in the late 80s by Borwein–Zuang [50], Frankowska [128, 131] and Penot [261]. The main focus of [50, 261] was on equivalent characterizations of nonlinear regularity with arbitrary nonlinearities. A short proof of the equivalence of openness and metric regularity of "order k" (with $\mu(t) = t^k$) was given in [130]. In Theorem 2.96 we follow [166]. We have mentioned that this theorem is a nonlinear extension of the general regularity criterion of Theorem 2.46. Perhaps other regularity criteria of Sect. 3 may have non-linear analogues as well but I am not aware of such results. Theorem 2.98 does not seem to have appeared in the literature, but Corollary 2.99 is an easy consequence of the following result proved in [131].

Theorem 2.112. *Suppose that X is a metric space, Y is a metric space, $F : X \rightrightarrows Y$ is a set-valued mapping with a complete graph and $(\bar{x}, \bar{y}) \in$ Graph F. Finally, let $Q \subset Y$ be bounded. Suppose that for some $k > 0$, $\lambda \in [0, 1)$ and $\varepsilon > 0$ the inclusion*

$$F(x) \cap B(\overline{y}, \delta) + t^k Q \subset B(B(x, t)) + \lambda t^k Q$$

holds. Then for all $(x, y) \in$ Graph F in a small neighborhood of $(\bar{x}, \bar{y})$ and all sufficiently small $t > 0$, we have

$$y + (1 - \lambda^{1/k})^k t^k Q \subset G(B(x, t)).$$

Exercise 2.113. Prove the theorem.

Section 8. The problem of regularity of a composition was discussed in [109, 110] partly in connection with the two maps paradigm in fixed point theory. Most of the results of the section are taken from [182]. Corollarys 2.107 and 2.108a are from [110].

http://www.springer.com/978-3-319-64276-5

Variational Analysis of Regular Mappings
Theory and Applications
Ioffe, A.
2017, XXI, 495 p. 11 illus., 2 illus. in color., Hardcover
ISBN: 978-3-319-64276-5