J. Aust. Math. Soc. **101** (2016), 376–417
doi:10.1017/S1446788715000695

METRIC REGULARITY—A SURVEY PART II. APPLICATIONS

A. D. IOFFE

(Received 6 April 2015; accepted 11 August 2015; first published online 8 July 2016)

Communicated by J. M. Borwein

Abstract

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

2010 *Mathematics subject classification*: primary 49J53; secondary 54C60, 58C06, 47H04, 49K40, 46A30, 90C31.

Keywords and phrases: set-valued mapping, regularity, transversality, implicit function, Lipschitz stability, Sard theorem.

Contents

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

Part II. Applications

7. Special classes of mappings

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

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

$$d(x, [f \le \alpha]) \le K(f(x) - \alpha)^+ \tag{7.1}$$

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

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

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

https://doi.org/10.1017/S1446788715000695 Published online by Cambridge University Press

Proof. Without loss of generality, we may assume that f is nonnegative: just take f^+ instead of f. So take an $\bar{x}$ as in the statement. By Ekeland's principle, there is a $\bar{u}$ such that $d(\bar{u}, \bar{x}) \le r^{-1} f(\bar{x})$ and $f(x) + rd(x, \bar{u}) > f(\bar{u})$ if $x \ne \bar{u}$. We claim that $f(\bar{u}) \le 0$. Indeed, otherwise, by the assumption, there would be an $x \ne \bar{u}$ such that $f(\bar{u}) - f(x) \ge rd(x, \bar{u})$, which is a contradiction. □

Here, for simplicity, we shall speak mainly about *global* error bounds, corresponding to $U = X$, at the zero level. We shall denote by K_f the lower bound of K such that (7.1) holds for all x. For brevity, we also set

$$S = [f \le 0], \quad S_0 = [f = 0].$$

7.1.1. Error bounds for convex functions. We shall start with the simplest case of a convex function f (extended-real-valued, in general) on a Banach space X.

Theorem 7.2. *Let X be a Banach space and f be a proper closed convex function on X. Assume that $S = [f \le 0] \ne \emptyset$. Then*

$$K_f^{-1} = \inf_{x \notin S} \sup_{\|h\| \le 1} (-f'(x; h)) = \inf_{x \notin S} d(0, \partial f(x)) = \inf_{x \notin S} \mathrm{sur}(\mathrm{Epi}\, f)(x, f(x)).$$

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

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

Take a positive r' and δ such that $\delta < r' < r$ and let $TU(x)$ be the set of pairs (u, t) satisfying

$$\|u - x\| \le t, \quad f(u) \le f(x) - r't.$$

By Ekeland's variational principle, for any $\delta > 0$ there is a $(\bar{u}, \bar{t}) \in TU(x)$ such that $f(u) + \delta\|u - \bar{u}\|$ attains its minimum at $\bar{u}$. Clearly, $\bar{t} > 0$ (as $f(x) > 0$). We claim that $f(\bar{u}) = 0$. Indeed, if $f(\bar{u}) > 0$, then there is an h with $\|h\| = 1$ such that $-f'(\bar{u}; h) > r'$: that is, $f(\bar{u} + th) < f(\bar{u}) - r't$ for some $t > 0$. Set $u = \bar{u} + th$. Then $f(u) < f(\bar{u}) - \delta$ $\|u - \bar{u}\|$ and we get a contradiction with the definition of $\bar{u}$.

Thus $f(\bar{u}) = 0$ which means that

$$d(x, S_0) \le \|\bar{u} - x\| \le t \le \frac{1}{r'} f(x),$$

and this completes the proof as r' can be chosen arbitrarily close to r and x is an arbitrary point of $[f > 0]$. □

There is another way to characterize K_f in terms of normal cones to $[f \le 0]$.

Theorem 7.3. *For any continuous convex function f on a Banach space X*

$$K_f = \inf_{x \in [f=0]} \inf\{\tau > 0 : N([f = 0], x) \cap B_{X^*} \subset [0, \tau]\partial f(x)\}.$$

https://doi.org/10.1017/S1446788715000695 Published online by Cambridge University Press

7.1.2. Some general results on global error bounds. Let us turn now to the general case of a lower semicontinuous function on a complete metric space.

Now denote by $K_f(\alpha,\beta)$ (where $\beta > \alpha \geq 0$) the lower bound of K such that

$$d(x,[f \leq \alpha]) \leq K f(x)^+ \quad \text{if } \alpha < f(x) \leq \beta.$$

Clearly, $K_f = \lim_{\beta\to\infty} K_f(0,\beta)$.

Theorem 7.4. *Let X be a complete metric space and f a lower semicontinuous function on X. If $[f \leq 0] \neq \emptyset$, then*

$$\inf_{x\in[0<f\leq\beta]} |\nabla f|(x) = \inf_{\alpha\in[0,\beta)} K_f(\alpha,\beta)^{-1}.$$

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

As an immediate consequence, we get the following corollary.

Corollary 7.5. *Under the assumption of the theorem*

$$K_f^{-1} \geq \inf_{x\in[f>0]} |\nabla f|(x).$$

A trivial example of a function f having an isolated local minimum at a certain $\overline{x}$ and such that $\inf f < f(\overline{x})$ shows that the inequality can be strict. This may happen, of course, even if the slope is different from zero everywhere on $[f > 0]$. In this case, an estimate of another sort can be obtained. Set (for $\beta > 0$)

$$d_f(\beta) = \sup_{x\in[f\leq\beta]} d(x,[f \leq 0])$$

and define the functions

$$\kappa_{f,\varepsilon}(t) = \sup\left\{\frac{1}{|\nabla f|(x)} : |f(x) - t| < \varepsilon\right\}, \quad \kappa_f(t) = \lim_{\varepsilon\to 0} \kappa_{f,\varepsilon}(t).$$

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

$$d_f(\beta) \leq \int_0^\beta \kappa_f(t)\,dt.$$

https://doi.org/10.1017/S1446788715000695 Published online by Cambridge University Press

Following the pioneering 1952 work by Hoffmann [45] (to be proved later in this section), error bounds, both for nonconvex and, especially, convex functions have been intensively studied, in particular, during last two or three decades, both theoretically, in connection with metric regularity, and also in view of their role in numerical analysis (see, for example [27, 40, 76, 81, 98, 109]). The basic lemma was proved in [53], its earlier version corresponding to $U = X$ was proved by Azé-Corvellec-Lucchetti and appeared in [9]. Finite dimensional versions of Theorems 7.2 and 7.3 were proved in Lewis–Pang [76] and Klatte–Li [69]. The equality

$$K_f^{-1} = \inf\{d(0, \partial f(x)) : x \in [f > 0]\}$$

in Theorem 7.2 was proved by Zalinescu (see [108]). The first two equalities in the theorem can be found in [7, 8] and the third equality for polyhedral functions on $\mathbb{R}^n$ in [80]. Theorem 7.3 was proved by Zheng and Ng [109] and Theorem 7.4 by Azé and Corvellec in [7]. The papers also contain sufficiently thorough bibliographic comments. Here we follow [57], where proofs of all the stated and some other results can be found.

7.2. Mappings with convex graphs.

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

According to [59, Definition 5.1], given a convex process $\mathcal{A} : X \rightrightarrows Y$, the *adjoint process* $\mathcal{A}^* : Y^* \rightrightarrows X^*$ (always closed) is defined by

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

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

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

$$\mathcal{A}(x_1) + \mathcal{A}(x_2) \subset \mathcal{A}(x_1 + x_2).$$

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

https://doi.org/10.1017/S1446788715000695 Published online by Cambridge University Press

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

Theorem 7.9 (Regularity moduli of a convex process). *For any closed convex process $\mathcal{A} : X \rightrightarrows Y$ from one Banach space into another*

$$C(\mathcal{A}) = C^*(\mathcal{A}^*) = \operatorname{sur} \mathcal{A}(0|0) = \operatorname{contr} \mathcal{A}(0|0).$$

Note that the left-hand inequality is equivalent to $\|\mathcal{A}^{-1}\|_- = \|(\mathcal{A}^{-1})^*\|_+$ (compare with [16]).

Proof. We first observe that the right-hand equality is a consequence of the other two, in view of [59, Proposition 5.2]. The inequality $C^*(\mathcal{A}^*) \geq C(\mathcal{A})$ follows from [59, Theorem 5.4]. The same theorem, together with the definition of Banach constants, implies that

$$C^*(\mathcal{A}^{**}) \geq C^*((\mathcal{A}^*)^*) \geq C(\mathcal{A}^*) \geq C^*(\mathcal{A}^*).$$

But $\mathcal{A}^{**} = \mathcal{A}$, as $\mathcal{A}$ is closed, so that $C^*(\mathcal{A}^{**}) = C^*(\mathcal{A})) \leq C(\mathcal{A})$ (see, again, [59, Theorem 5.4]). This proves the left-hand equality.

Passing to the proof of the middle equality, we first observe that, by [59, Proposition 5.2], $C(\mathcal{A}) = \operatorname{contr} \mathcal{A}(0|0) \geq \operatorname{sur} \mathcal{A}(0|0)$ as the rate of surjection can never exceed the modulus of controllability. On the other hand, by Proposition 7.8, $D\mathcal{A}(0,0)(h) \subset D\mathcal{A}(x,y)(h)$ for all $(x,y) \in \operatorname{Graph} \mathcal{A}$ and all h. Hence, by [59, Theorem 5.13], $\operatorname{sur} \mathcal{A}(0|0) \geq C(D\mathcal{A}(0,0))$. But $D\mathcal{A}(0,0)(h) = \mathcal{A}(h)$ as the tangent cone to a closed convex cone at zero coincides with the latter. Thus $\operatorname{sur} \mathcal{A}(0|0) \geq C(\mathcal{A})$. □

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

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

We conclude this subsection by considering the effect of linear perturbations. If $\mathcal{A}$ is a convex process, then so is $\mathcal{A} + A$, where A is a linear bounded operator from X into Y. Thus, if $\mathcal{A}$ is closed, then $\mathcal{A} + A$ is perfectly regular at the origin and we get the following theorem, as an immediate consequence of [59, Theorem 5.28].

Theorem 7.11 (Radius of regularity of a convex process). *If $\mathcal{A} : X \rightrightarrows Y$ is a closed convex process, then*

$$\operatorname{rad} \mathcal{A}(0|0) = \operatorname{sur} \mathcal{A}(0|0).$$

Convex processes were introduced by Rockafellar [95, 96] as an extension of linear operators and, subsequently, thoroughly studied by Robinson [86], Borwein [14, 15] and Lewis [73, 74]. In particular, [86] contains an extension to the convex processes of the Banach–Schauder open mapping theorem. Another remarkable result (which is actually a special case of Theorem 5 in the paper) can be reformulated as follows: *let X and Y be Banach spaces, and let $\mathcal{A} : X \rightrightarrows Y$ and $\mathcal{T} : X \rightrightarrows Y$ be closed convex*

https://doi.org/10.1017/S1446788715000695 Published online by Cambridge University Press

processes. Then $C(\mathcal{A} - \mathcal{T}) \geq C(\mathcal{A}) - \|\mathcal{T}\|_-$. The result, which is equivalent to the equality $C(\mathcal{A}) = C^*(\mathcal{A}^*)$ (Theorem 7.9), was proved and further discussed in [14, 15] and in Theorem 7.11 in [73], along with the equality of the radius and distance to infeasibility for convex processes.

7.2.2. Theorem of Robinson–Ursescu.

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

$$\operatorname{sur} F(\bar{x}|\bar{y}) \geq \frac{\beta}{\alpha}.$$

PROOF. We can set $\bar{x} = 0$, $\bar{y} = 0$. It is clear that $F(t\alpha B_X)$ is dense in $t\beta B_Y$ for any $t \in (0, 1)$. Denote $r = \beta/\alpha$. We shall show that, given a $\gamma > 0$, there is an $\varepsilon > 0$ such that $F(B(x, (1 + \gamma)t))$ is dense in $B(v, rt)$ if $\|x\| < \varepsilon$, $\|v\| < \varepsilon$ and $v \in F(x)$. The theorem will then follow from [59, Corollary 3.8].

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

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

Also, $\|y_1 - y\| = \lambda \|v_0 - y\|$: that is,

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

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

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

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

Then $v \in F(x)$ as $\operatorname{Graph} F$ is convex. Also, $\|y - v\| \leq \delta/(1 + \lambda) \leq \delta/2$ and

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

If

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

this completes the proof, as γ, ε and δ can be chosen to be arbitrary small. □

As a corollary, we get the following theorem.

THEOREM 7.13 (Robinson–Ursescu [89, 100]). *Let X and Y be Banach spaces. If the graph of $F : X \rightrightarrows Y$ is convex and closed and $\bar{y} \in \operatorname{int} F(X)$, then F is regular at any $(\bar{x}, \bar{y}) \in \operatorname{Graph} F$.*

PROOF. Let $\overline{y} \in F(\overline{x})$. We have to show that there are $\alpha > 0$ and $\beta > 0$ such that $F(B(\overline{x}, \alpha))$ is dense in $B(\overline{y}, \beta)$, which is easy to do with the help of the standard argument using the Baire category theorem. □

7.2.3. Mappings with convex graphs. Regularity rates. Here we give two results containing exact formulas for the rate of surjection of set-valued mappings with convex graphs.

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

$$\operatorname{sur} F(\overline{x}|\overline{y}) = \lim_{\varepsilon \to +0} \inf_{\|y^*\|=1} \inf_{x^*} \Big(\|x^*\| + \frac{1}{\varepsilon} S_{\operatorname{Graph}(F-(\bar{x},\bar{y}))}(x^*, y^*) \Big).$$

The theorem was proved in Ioffe and Sekiguchi [63] (see, also, [57] for a short proof). It allows one to also get a 'primal' representation for the rate of surjection of a convex set-valued mapping. The key to this development is the concept of *homogenization* $\mathcal{Q}$ of a convex set $Q \subset X$, which is the closed convex cone in $X \times \mathbb{R}$ generated by the set $Q \times \{1\}$. It is an easy matter to verify (if Q is also closed) that $(x, t) \in \mathcal{Q}$ if and only if $x \in tQ$ if $t > 0$ and $x \in Q^\infty$ (the recession cone of Q) if $t = 0$ (recall that $Q^\infty = \{h \in Q : x + h \in Q, \forall x \in Q\}$).

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

$$\mathcal{F}_{(\bar{x},\bar{y})}(h, t) = \begin{cases} t\Big(F\Big(\overline{x} + \dfrac{h}{t}\Big) - \overline{y}\Big) & \text{if } t > 0, \\ F^\infty(h) & \text{if } t = 0, \\ \emptyset & \text{if } t < 0, \end{cases}$$

where F^∞ is the 'horizon' mapping of F whose graph is the recession cone of $\operatorname{Graph} F$:

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

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

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

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

$$\operatorname{sur} F(\overline{x}|\overline{y}) = \lim_{\varepsilon \to +0} C_\varepsilon(\mathcal{F}_{(\bar{x},\bar{y})}).$$

https://doi.org/10.1017/S1446788715000695 Published online by Cambridge University Press

Proof. Setting $h = t(x - \overline{x}), v = t(y - \overline{y})$, we get

$$\begin{aligned}\operatorname{Graph}\mathcal{F}^*_{(\bar{x},\bar{y})} &= \{(x^*, y^*, \lambda) : \langle x^*, h\rangle - \langle y^*, v\rangle + \lambda t \le 0 : \forall (h, v, t) \in \operatorname{Graph}\mathcal{F}_{(\bar{x},\bar{y})}\}\\ &= \{(x^*, y^*, \lambda) : t[\langle x^*, x - \overline{x}\rangle - \langle y^*, y - \overline{y}\rangle + \lambda] \le 0 : \forall (x, y) \in \operatorname{Graph} F, t > 0\}\\ &= \{(x^*, y^*, \lambda) : s_{\operatorname{Graph} F-(\bar{x},\bar{y})}(x^*, -y^*) + \lambda \le 0\}.\end{aligned}$$

As the support function of $\operatorname{Graph} F - (\bar{x}, \bar{y})$ is nonnegative, it follows that $\lambda \le 0$ whenever $(x^*, y^*, \lambda) \in \operatorname{Graph}\mathcal{F}_{(\bar{x},\bar{y})}$. The norm in $X^* \times \mathbb{R}$ dual to $\|\cdot\|_\varepsilon$ is $\|(x^*, \lambda)\|_\varepsilon = \|x^*\| + \varepsilon^{-1}|\lambda|$. Let d_ε stand for the distance in $X^* \times \mathbb{R}$ corresponding to this norm. Then

$$\begin{aligned}d_\varepsilon(0, \mathcal{F}^*_{(\bar{x},\bar{y})}(\bar{x}, \bar{y})(y^*)) &= \inf\{\|x^*\| + \varepsilon^{-1}|\lambda| : s_{\operatorname{Graph} F-(\bar{x},\bar{y})}(x^*, -y^*) + \lambda \le 0\}\\ &= \inf_{x^*}(\|x^*\| + \varepsilon^{-1} s_{\operatorname{Graph} F-(\bar{x},\bar{y})}(x^*, -y^*)).\end{aligned}$$

It remains to compare this with Theorem 7.14 to see that

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

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

The concept of homogenization was introduced by Hörmander [46]. The idea to apply homogenization for regularity estimation goes back to Robinson [88]. His main result actually says that $\operatorname{sur} F(\overline{x}|\overline{y}) \ge C_1(\mathcal{F}_{(\bar{x},\bar{y})})$. In a somewhat different context, homogenization techniques were applied by Lewis [74] for estimating the distance to infeasibility of so-called conic systems. The full statement of Theorem 7.15 was proved in [63]. Here, we will not discuss some well-developed problems relating to regularity of maps with convex graphs, for example stability under perturbations of systems of convex inequalities (see, for example, [18, 57, 87] and the references in the first two quoted papers).

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

$$\mathcal{D}^* F(x)(y^*) = \partial(y^* \circ F)(x)$$

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

The following result is straightforward.

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

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

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

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

$$F(x + th) - F(x) \subset t\mathcal{A}(h) + r(t, x)t\|h\|B_Y \quad \forall h \in Q, \tag{7.2}$$

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

There are some canonical ways for constructing prederivatives. The first to mention is the generalized Jacobian introduced by Clarke [22] for mappings in the finite-dimensional case and then extended to some classes of Banach spaces by Páles and Zeidan [83, 84]. Another construction, not associated with linear operators, was introduced in [48]. Take an $\varepsilon > 0$ and set

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

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

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

that is, $\mathcal{H}_\varepsilon(th) = t\mathcal{H}_\varepsilon(h)$. Thus $\mathcal{H}_\varepsilon$ is positively homogeneous and it is an easy matter to see that (7.2) holds with $r(t, x) = 0$.

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

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

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

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

Proposition 7.18 [52]. *If $F : X \to Y$ is Lipschitz continuous near x, then*

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

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

$$D_H^* F(x)(y^*) = \partial_H(y^* \circ F)(x) \quad \textit{and} \quad D_G^* F(x)(y^*) = \partial_G(y^* \circ F)(x), \quad \forall y^* \in Y^*.$$

Combining this proposition with [59, Theorem 5.21] we get the following theorem.

https://doi.org/10.1017/S1446788715000695 Published online by Cambridge University Press

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

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

The obvious inequality

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

(where $\mathcal{H}(x)$ is a strict prederivative at x) leads to the estimate $\operatorname{sur} F(\bar{x}) \geq \liminf_{x \to \bar{x}} C^*(\mathcal{H}(x))$, under the assumptions of the theorem. A better result can be proved with the help of the general metric regularity criteria if F has a strict Fréchet prederivative at $\bar{x}$.

THEOREM 7.20. *Assume that Y is Gâteaux smooth and $F : X \to Y$ satisfies the Lipschitz condition in a neighborhood of $\bar{x}$ and, moreover, admits at $\bar{x}$ a strict Fréchet prederivative $\mathcal{H}$ with norm compact values such that, for any y^* with $\|y^*\| = 1$,*

$$\sup_{\|h\|=1} \inf_{w \in \mathcal{H}(h)} \langle y^*, w \rangle \geq \rho > 0. \tag{7.3}$$

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

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

$$F(x') - F(x) \in \mathcal{H}(x) + \varepsilon\|x' - x\| \tag{7.4}$$

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

$$\lim_{t \to 0} t^{-1}(\|y - F(x) + tw\| - \|y - F(x)\|) = \langle y^*, w \rangle \quad \text{for every } w \in Y. \tag{7.5}$$

By (7.3), there is an $h \in S_X$ such that

$$\langle y^*, w \rangle > \rho - \varepsilon \quad \text{for all } w \in \mathcal{H}(h). \tag{7.6}$$

Since the set $-\mathcal{H}(h)$ is compact and the limit in (7.5) is uniform with respect to w from any fixed compact set, we conclude that, for sufficiently small $t > 0$,

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

This, and (7.6) imply that

$$\|y - F(x) - tw\| < \|y - F(x)\| - \langle y^*, tw \rangle + \varepsilon t \leq \|y - F(x)\| - t(\rho - 2\varepsilon) \tag{7.7}$$

for all $w \in \mathcal{H}(h)$. Let $x' := x + th$. Then $\|x' - x\| = \|th\| = t < r/2$, and hence $x' \in B(\bar{x}, r)$. Since $\mathcal{H}$ is positively homogeneous, $\mathcal{H}(x' - x) = \mathcal{H}(th) = t\mathcal{H}(h)$. Thus, by (7.4), there is a $w \in \mathcal{H}(h)$ such that

$$\|F(x') - F(x) - tw\| \leq t\varepsilon. \tag{7.8}$$

https://doi.org/10.1017/S1446788715000695 Published online by Cambridge University Press

Now, we are ready for the chain of estimates

$$\begin{aligned}\|y - F(x')\| &\le \|F(x) - F(x') + tw\| + \|y - F(x) - tw\| \\ &< \varepsilon t + \|y - F(x)\| - (\rho - 2\varepsilon)t \quad \text{(by (7.8) and (7.7))} \\ &= \|y - F(x)\| - (\rho - 3\varepsilon)t = \|y - F(x)\| - (\rho - 3\varepsilon)\|x' - x\|.\end{aligned}$$

It remains to apply the criterion of [59, Theorem 3.2]. □

A slight modification of the proof allows to get the following theorem.

THEOREM 7.21. *Assume that $F : X \to Y$ satisfies the Lipschitz condition in a neighborhood of $\bar{x}$ and, moreover, there are a homogeneous set-valued mapping $\mathcal{H} : X \rightrightarrows Y$ with norm compact values and $\beta \ge 0$ such that (7.3) holds and*

$$F(x + h) - F(x) \subset \mathcal{H}(h) + \beta\|x' - x\|B_Y.$$

Then $\operatorname{sur} F(\bar{x}) \ge \rho - \beta$.

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

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

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

$$\|F(x) - F(x') - T(x - x')\| \le \beta\|x - x'\|.$$

(b) *There are $\rho > 0$ and $\varepsilon > 0$ such that, for any $T \in T$,*

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

Then $\operatorname{sur} F(\bar{x}) \ge \rho - \beta$.

Scalarization formulas first appeared in [49] for mappings between finite-dimensional spaces and [71] for mappings between Fréchet smooth spaces, although scalarized coderivatives had already been considered in [48, 70]. The very term ‘coderivative’ was introduced in [48]. The concept of prederivative was introduced in [48] and a characterization of directional compactness in [52] (see, also, [66] for an earlier result).

Theorems 7.20 and 7.21 will appear in [20]. Theorem 7.22 was proved in [19]. An earlier result without constraints on the domain of the mapping was proved by Páles in [82]. We also refer to [20] for a shorter proof of the last theorem. Note that the convexity requirement in Theorem 7.22 is essential (consider, for instance, $F(x) = |x| : \mathbb{R} \to \mathbb{R}$ and $\mathcal{T}$ containing two operators $T_1(x) = x$ and $T_2(x) = -x$). Because of this requirement, the estimate provided by Theorem 7.22 is generally less precise than those of Theorems 7.19 and 7.20 (consider, for instance, the mapping $\mathbb{R}^2 \to \mathbb{R}$: $F(x_1, x_2) = |x_1| - |x_2|$), but it can be easier to apply in certain cases (for example, in the finite-dimensional case when we can take the generalized Jacobian as $\mathcal{T}$; see [22]).

7.4. Polyhedral sets and mappings. This subsection contains some elementary results concerning geometry of polyhedral sets in $\mathbb{R}^n$ and regularity of set-valued mappings with polyhedral graphs. Deeper problems associated with variational inequalities over convex polyhedral sets will be discussed in the next section.

Definition 7.23 (Polyhedral sets). A *convex polyhedral set* (or a *convex polyhedron*) $Q \subset \mathbb{R}^n$ is an intersection of a finite number of closed linear subspaces and hyperplanes: that is

$$Q = \{x \in \mathbb{R}^n : \langle x_i^*, x\rangle \le \alpha_i, i = 1, \ldots, k; \langle x_i^*, x\rangle = \alpha_i, i = k+1, \ldots, m\} \tag{7.9}$$

for some nonzero $x_i^* \in \mathbb{R}^n$ and $\alpha_i \in \mathbb{R}$. Following [33], we shall use the term *polyhedral set* for finite unions of convex polyhedra.

Clearly, any polyhedral set is closed. Also, as any linear equality can be replaced by two linear inequalities, we can represent any polyhedral set by means of a system of linear inequalities only. An elementary geometric argument allows us to reveal one of the most fundamental properties of polyhedral sets: an *orthogonal projection of a polyhedral set is a polyhedral set*. In fact, a linear image of a polyhedral set is polyhedral (see [96] for this and other basic properties of polyhedral sets).

A set-valued mapping $R^n \rightrightarrows \mathbb{R}^m$ is *(convex) polyhedral* if its graph is a polyhedral set. Our primary interest in this section is to study regularity properties of such mappings.

Proposition 7.24 (Local tangential representation). *Let $Q \subset \mathbb{R}^n$ be a polyhedral set and $\overline{x} \in Q$. Then there is an $\varepsilon > 0$ such that*

$$Q \cap B(\overline{x}, \varepsilon) = \overline{x} + T(Q, \overline{x}) \cap (\varepsilon B).$$

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

One more useful corollary concerns normal cones of polyhedral sets.

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

Our first result is the famous Hoffmann theorem on error bounds for a system of linear inequalities. Set $a = (\alpha_1, \ldots, \alpha_m) \in \mathbb{R}^m$ and let $Q(a)$ be defined by (7.9).

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

$$d(x, Q(a)) \le K\Big(\sum_{i=1}^{k}(\langle x_i^*, x\rangle - \alpha_i)^+ + \sum_{i=k+1}^{m} |\langle x_i^*, x\rangle - \alpha_i|\Big)$$

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

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

$$f(x) = \sum_{i=1}^{k}(\langle x_i^*, x\rangle - \alpha_i)^+ + \sum_{i=k+1}^{m} |\langle x_i^*, x\rangle - \alpha_i|.$$

Then $Q(a) = [f \le 0]$. Set

$$I_1(x) = \{i \in \{1, \dots, k\} : \langle x_i^*, x\rangle = \alpha_i\}, \quad J_+(x) = \{i \in \{1, \dots, m\} : \langle x_i^*, x\rangle > \alpha_i\},$$
$$I_0(x) = \{i \in \{k+1, \dots, m\} : \langle x_i^*, x\rangle = \alpha_i\}, \quad J_-(x) = \{i \in \{k+1, \dots, m\} : \langle x_i^*, x\rangle < \alpha_i\}.$$

Then

$$\partial f(x) = \sum_{i\in I_1(x)} [0, 1]x_i^* + \sum_{i\in I_0(x)} [-1, 1]x_i^* + \sum_{i\in J_+(x)} x_i^* - \sum_{i\in J_-(x)} x_i^*.$$

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

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

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

For any $\sigma \in \Sigma$, denote by $\gamma(\sigma)$ the distance from zero to the set on the right-hand side of the above inclusion, and let K stand for the upper bound of $\gamma(\sigma)^{-1}$ over $\sigma \in \Sigma$. Then $K < \infty$ since Σ is a finite set. Clearly, K does not depend on either a or x. On the other hand, $Kd(0, \partial f(x)) \ge 1$. It remains to refer to Theorem 7.2 to conclude the proof. □

As an immediate consequence, we get the following theorems.

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

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

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

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

$$d(x, Q_1 \cap Q_2) \le K(d(x, Q_1) + d(x, Q_2)).$$

https://doi.org/10.1017/S1446788715000695 Published online by Cambridge University Press

To prove Theorem 7.27, we need to apply the Hoffmann estimate to the graph of F. Concerning Theorem 7.28, it should be observed that global subtransversality does not imply transversality at any point. As a simple example, consider the half-spaces $S_1 = \{x : \langle x^*, x\rangle \geq 0\}$ and $S_2 = \{x : \langle x^*, x\rangle \leq 0\}$ with some $x^* \neq 0$. The intersection of the sets is $\operatorname{Ker} x^* \neq \emptyset$. But the inclusions $x_1 - x \in S_1$ and $x_2 - x \in S_2$ imply $\langle x^*, x_1\rangle \geq \langle x^*, x_2\rangle$, and hence (see [59, Definition 6.11] S_1 and S_2 are not transversal at points of $\operatorname{Ker} x^*$.

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

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

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

(b) *there is a $K > 0$ (may differ from that in (a)) such that, for any $(\bar{x}, \bar{y}) \in \operatorname{Graph} F$, there is an $\varepsilon > 0$ such that $d(x, F^{-1}(\overline{y})) \leq Kd(\overline{y}, F(x))$ if $\|x - \overline{x}\| < K\varepsilon$.*

Thus F is subregular at any point of its graph.

Proof. We have $F(x) = \bigcup_{i=1}^{k} F_i(x)$, where all F_i are convex polyhedral set-valued mappings. By Theorem 7.27, for any i there is a K_i such that $d(y, F_i(x)) \leq K_i\|x - \overline{x}\|$ for any $\overline{x} \in \operatorname{dom} F_i$ and any $(x, y) \in \operatorname{Graph} F_i$. Now fix some $\overline{x} \in \operatorname{dom} F$, and let $I = \{i : \overline{x} \in \operatorname{dom} F_i\}$. Choose an $\varepsilon > 0$ so small that $d(x, \operatorname{dom} F_i) > \varepsilon$ if $i \notin I$ and $\|x - \overline{x}\| < \varepsilon$ (clearly, such an ε can be found as all $\operatorname{dom} F_i$ are polyhedral sets, and hence closed). If now $y \in F(x)$ and $\|x - \overline{x}\| < \varepsilon$, then $I(x, y) = \{i : y \in F_i(x)\} \subset I$. On the other hand, as we have seen, there are K_i such that $y \in F_i(x)$ implies that $d(y, F_i(\overline{x})) \leq K_i\|x - \overline{x}\|$. Thus, if $y \in F(x)$ and $\|x - \overline{x}\| < \varepsilon$, then

$$d(y, F(\overline{x})) \leq \max_{i\in I(x,y)} d(y, F_i(\overline{x})) \leq (\max_i K_i)\|x - \overline{x}\|.$$

This proves the first statement.

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

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

Example 7.30. Let $X = Y = R, Y$, and let

$$F_1(x) = \begin{cases} \mathbb{R}_+ & \text{if } x > 0,\\ \mathbb{R} & \text{if } x = 0,\\ \emptyset & \text{if } x < 0,\end{cases} \qquad F_2(x) = \begin{cases} \mathbb{R}_- & \text{if } x < 0,\\ \mathbb{R} & \text{if } x = 0,\\ \emptyset & \text{if } x > 0,\end{cases}$$

and $F(x) = F_1(x) \cup F_2(x)$. Fix some $y > 0$ and $x < 0$. Then $F^{-1}(y) = \mathbb{R}_+$ and $d(x, F^{-1}(y)) = |x|$, $d(y, F(x)) = y$ so that there are no K for which the inequality $d(x, F^{-1}(y) \leq Kd(y, F(x))$ holds in a neighborhood of $(0, 0)$.

https://doi.org/10.1017/S1446788715000695 Published online by Cambridge University Press

COROLLARY 7.31 (Subtransversality of polyhedral sets). *Any two polyhedral sets Q_1 and Q_2 with nonempty intersection are subtransversal at any common point of the sets: that is, for any $\overline{x} \in Q_1 \cap Q_2$, there is a K such that the inequality*

$$d(x, Q_1 \cap Q_2) \leq K(d(x, Q_1) + d(x, Q_2))$$

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

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

7.5. Semialgebraic mappings, stratifications and the Sard theorem. Most of the results of this subsection (including the Sard theorem) can be extended to a wide class of objects, so-called *definable* sets, mappings and functions. However, we confine ourselves here to semialgebraic functions whose definition is much simpler (compare with the general definition of definability) and does not require any specific effort[1].

We shall concentrate, basically, on two topics: consequences of the general theory and studies associated with semialgebraic geometry, mainly in connection with the Sard theorem.

7.5.1. Basic properties (see [12, 26]). A semialgebraic set in $\mathbb{R}^n$ is, by definition, a union of a finite number of sets of solutions of a finite system of polynomial equalities and inequalities of n variables: that is,

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

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

A mapping (whether single or set-valued) is semialgebraic if its graph is semialgebraic. Here is a list of some basic properties of semialgebraic sets and mappings.

- The closure and interior of a semialgebraic set is semialgebraic.
- The Cartesian product of semialgebraic sets is semialgebraic.
- The composition of semialgebraic mappings is semialgebraic.
- The image and preimage of a semialgebraic set under a semialgebraic mapping is semialgebraic.

[1]It should be mentioned that, recently, Barbet *et al.* [10] proved a remarkable result containing extensions of the Sard theorem to some other important classes of nonsmooth functions.

https://doi.org/10.1017/S1446788715000695 Published online by Cambridge University Press

- The derivative of a (single-valued) semialgebraic mapping is semialgebraic.
- The upper and lower bound of a finite collection of extended-real-valued semialgebraic functions is semialgebraic.
- If we have a semialgebraic function of two (vector) variables, then its upper or lower bound with respect to one of the variables on a semialgebraic set is semialgebraic.
- If F is a semialgebraic set-valued mapping such that every $F(x)$ is a finite set, then the number of elements in each $F(x)$ does not exceed a certain finite N.

For us, in the context of variational analysis and, especially, regularity theory, the most important results are the following.

- A subdifferential mapping of a semialgebraic function or the coderivative mapping of a semialgebraic map is semialgebraic (regardless of which subdifferential on $\mathbb{R}^n$: Fréchet, Dini–Hadamard, limiting or Clarke, we are talking about).
- The slope of a semialgebraic function is a semialgebraic function of the point.
- Rates of regularity of a semialgebraic functions are also semialgebraic functions of the point of the graph.

Definition 7.32. A finite partition (M_i) of a set $Q \subset \mathbb{R}^n$ is called a C^r-Whitney stratification of Q if each M_i is a C^r-manifold and the following two properties are satisfied.

(a) If $(x_k) \subset M_i$ converges to some x belonging to another element (M_j) of the partition and the unit normal vectors $v_k \in N_{x_k}M_i$ converge to some v, then $v \in N_x M_j$.

(b) If $M_j \cap \operatorname{cl} M_i \neq \emptyset$, then $M_j \subset \operatorname{cl} M_i$.

Elements of partitions are usually called *strata*. The following remarkable fact is due to S. Łojasievicz.

Theorem 7.33 (Stratification theorem). *Given a semialgebraic set $Q \subset \mathbb{R}^n$ and an $r \in \mathbb{N}$, Q admits a Whitney stratification into semialgebraic C^r-manifolds.*

Of course, stratification is not unique. But it is easy to understand that maximal dimensions of the strata coincide for all Whitney stratifications. This observation justifies the following definition.

Definition 7.34. The dimension $\dim Q$ of a semialgebraic set Q is the maximal dimension of the strata in Whitney stratifications of Q.

The most important consequence of the stratification theorem is a Sard-type theorem for semialgebraic set-valued mappings.

Definition 7.35. Let $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ be a set-valued mapping with a semialgebraic graph, and let ∂ stand either for the limiting or for the Clarke subdifferential. A point $\overline{y} \in \mathbb{R}^m$ is a *critical value* of F if there is an $x \in \mathbb{R}^n$ such that $y \in F(x)$ and $0 \in D^*F(x|y)(y^*)$ for some $y^* \neq 0$.

https://doi.org/10.1017/S1446788715000695 Published online by Cambridge University Press

Theorem 7.36 (Semialgebraic Sard theorem). *Critical values of a semialgebraic set-valued mapping $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ form a semialgebraic set of dimension not exceeding $m - 1$.*

In particular, an extended-real-valued semialgebraic function can have at most a finite number of critical values.

For the theory of semialgebraic sets and mappings, see [12, 106]. The Sard theorem was first proved by Bolte *et al.* [13] for real-valued functions and then by Ioffe [54] for set-valued mappings (in both cases, the theorems were stated for more general classes of objects—semianalytic functions in [13] and arbitrarily stratifiable maps in [54]).

7.5.2. Transversality. We are finally ready to extend transversality theory (not just the definition) beyond the smooth domain. To begin with, we observe that a direct extension of [59, Proposition 1.12] does not hold if F is not smooth.

Example 7.37. Consider the function

$$f(x, w) = |x| - |w|$$

viewed as a mapping from $\mathbb{R}^2$ into $\mathbb{R}$. This mapping is clearly semialgebraic, even polyhedral. It is easy to verify that the mapping is regular at every point with the modulus of surjection identically equal to one (if we take the ℓ^∞ norm in $\mathbb{R}^2$). Furthermore,

$$Q = f^{-1}(0) = \{(x, w) : |x| = |w|\}$$

and the restriction to Q of the projection $(x, w) \to w$ is also a regular mapping with the modulus of surjection equal to one. However, the partial mapping $x \to f(x, 0) = |x|$ is not regular at zero.

However, the following statement is true.

Proposition 7.38 [56]. *Let $F : \mathbb{R}^m \times \mathbb{R}^k \rightrightarrows \mathbb{R}^n$ be a semialgebraic set-valued mapping with a locally closed graph, and let $\overline{y} \in F(\overline{x}, \overline{p})$. Assume that:*

(a) *F is regular at $((\overline{x}, \overline{p}), \overline{y})$;*
(b) *the set-valued mapping $\mathbb{R}^m \times \mathbb{R}^n \rightrightarrows \mathbb{R}^k$ associating the set $\{p : \ y \in F(x, p)\}$ with any $(x, y) \in \mathbb{R}^n \times \mathbb{R}^n$ is regular at $((\overline{x}, \overline{y}), \overline{p})$; and*
(c) *there is a Whitney stratification (M_i) of* Graph F *such that the restriction of the projection $(x, p) \to p$ to the set $S_i = \{(x, p) : (x, p, \overline{y}) \in M_i\}$, where M_i is the stratum containing $(\overline{x}, \overline{p}, \overline{y})$, is regular at $(\overline{x}, \overline{p})$.*

Then $F_{\bar{p}} : x \mapsto F(x, \bar{p})$ is regular at $(\bar{x}, \bar{y})$.

It is now possible to state, and prove, a semialgebraic set-valued version of Thom transversality theorem [59, Theorem 1.13].

Theorem 7.39. *Let the mapping $F : \mathbb{R}^n \times \mathbb{R}^k \rightrightarrows \mathbb{R}^m$ with closed graph and a closed set $S \subset \mathbb{R}^m$ both be semialgebraic. Denote by F_p the set-valued mapping $x \mapsto F(x, p)$. If F is transversal to S, then, for all p (with the possible exception of a semialgebraic set of dimension smaller than k), F_p is transversal to S.*

https://doi.org/10.1017/S1446788715000695 Published online by Cambridge University Press

Proof. The theorem is trivial if $F(x,p) \cap S = \emptyset$ for all (x,p), so we assume that $F(x,p)$ meets S for some values of the arguments. Then $(0,0)$ is a regular value of the mapping $\Psi : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^k \to \mathbb{R}^m$, $\Psi(x,y,p) = (F(x,p) - y) \times (S - y)$. Let $Q = \Psi^{-1}(0,0)$. This is a semialgebraic set, so, by Theorem 7.36, there is a semialgebraic set $C_0 \in \mathbb{R}^k$ such that $\dim C_0 < k$ and every $p \in \mathbb{R}^k \backslash C_0$ is a regular value of the restriction $\pi|_Q$ of the projection $(x,y,p) \mapsto p$.

Take an $r > N + m - k$, and let $(M_i)_{i=1,\ldots,r}$ be a C^1-Whitney stratification of Graph Ψ with all M_i being semialgebraic manifolds. Then, for any i, there is a semialgebraic set $C_i \subset \mathbb{R}^k$ such that any $p \in \mathbb{R}^k \backslash C_i$ is a regular value of $\pi|_{M_i}$. The union $C = \bigcup_{i=0}^r C_i$ is also a semialgebraic set of dimension smaller than k and, as we have just seen, for any $p \notin C$, all of the assumptions of Proposition 7.38 are satisfied for Ψ. Therefore $(0,0)$ is a regular value of Ψ_p. By [59, Theorem 6.15], this means that F_p is transversal to S. □

8. Some applications to analysis and optimization

In this section, we give several examples illustrating the power of regularity theory as a working instrument for treating various problems in analysis and optimization. We do not try, each time, to prove the result under the most general assumptions. The purpose is, rather, to demonstrate how regularity considerations help to understand and/or simplify the analysis of one or another phenomenon. Again, it should be said that some interesting applications of metric regularity remain outside the scope of the paper, such as the role of regularity in numerical optimization (see, for example, [33, 67, 68]) or connections with metric fixed point theory (for example, [3, 30, 31, 55, 58]) or recent developments associated with tilt stability, quadratic growth and so on (for example [1, 2, 34, 37, 67, 85]).

8.1. Subdifferential calculus. In each of the three calculus rules stated in [59, Proposition 5.9], we assume that one function is Lipschitz. One of the reasons (especially important in the proof of the exact sum rule) is that Lipschitz functions have bounded subdifferentials. But what happens when both functions are not Lipschitz? For instance, what can be said about a normal cone to an intersection of sets? As in the calculus of convex subdifferentials, we do need some qualification conditions to ensure the result.

Theorem 8.1. *Let X be a Banach space and let S_i, $i = 1,2$ be closed subsets of X. Further, let $\bar{x} \in S = S_1 \cap S_2$. If S_1 and S_2 are subtransversal at $\bar{x}$, then*

$$N_G(S,\bar{x}) \subset N_G(S_1,\bar{x}) + N_G(S_2,\bar{x}).$$

Explicitly, this theorem was first mentioned in [53] but *de facto* it had already been proved in [51] (see, also, [62, Proposition 3]). It turns out that subtransversality is the most general of all so far available conditions that would guarantee the inclusion. The most popular subdifferential transversality condition (condition (b) of [59, Theorem 6.12]) may be much stronger.

The inclusion is among the most fundamental facts of subdifferential calculus: enough to mention that, in the majority of publications on the subject, it is used as the starting point for deriving all other calculus rules. Below is a sketch of the proof of the theorem for the finite-dimensional situation.

Proof. We need the following elementary and/or well-known facts of functions on and sets in $\mathbb{R}^n$.

- $\hat{N}(Q, x) \cap B = \hat{\partial} d(\cdot, Q)(x)$ if $x \in Q$.
- If $x^* \in \hat{\partial} d(\cdot, Q)(x)$ and $u \in Q$ is the closest to x, then $x^* \in \hat{N}(Q, u)$.
- If $x \in Q$ and $f(\cdot)$ is nonnegative, equal to zero at x and $f(u) \geq d(u, Q)$ in a neighborhood of x, then $\hat{\partial} d(\cdot, Q)(x) \subset \hat{\partial} f(x)$.

Combining this with the definition of the limiting subdifferential, we conclude that, for Q, f and x, as above, $\partial d(\cdot, Q)(x) \subset \partial f(x)$ (the fact that is surprisingly missing from monographic publications).

By the assumption, there is a $K > 0$ such that $d(x, S) \leq K(d(x, S_1) + d(x, S_2))$, so, applying the above to $f(x) = K(d(x, S_1) + d(x, S_2))$ along with the exact calculus rule of Proposition, we conclude that $\partial d(\cdot, S)(\bar{x}) \subset K(\partial(\cdot, S_1)(\bar{x}) + \partial(\cdot, S_1)(\bar{x}))$ and the result follows. □

8.2. Necessary conditions in constrained optimization. We discuss, here, two ways to apply regularity theory to necessary optimality conditions and then a general approach to necessary conditions associated with one of them. Both differ substantially from classical proofs that include linearization and separation as the major steps (see, for example, [38, 42, 64, 88, 90]). Verification of relevance of linearization is usually the central and most difficult part of the proofs. It is established under certain constraint qualifications which always imply, and are often equivalent to, regularity of the constraint mapping (as in case of the popular Mangasarian–Fromovitz and Slater qualification conditions) (see, for example, [88], where the connection with regularity was made explicit).

We refer to [71, 78, 79] for extensions of the classical approach to nondifferentiable optimization in which convex separation is replaced by an ‘extremal principle’. The point is, however, that a fuller use of regularity arguments makes the way to necessary conditions much shorter. To begin with, we shall consider the problem

$$\text{minimize } f(x) \quad \text{subject to } F(x) \in Q, \ x \in C \tag{8.1}$$

(where $F : X \to Y$ is single-valued and $Q \subset Y$ and $C \subset X$ are closed sets) assuming, for simplicity, that both X and Y are finite-dimensional, although the results were originally proved in much more general situations.

8.2.1. Noncovering principle. Let $\bar{x} \in C$ be a solution of the problem. Let Ψ stand for the restriction to C of the set-valued mapping $x \mapsto \{f(x) - \mathbb{R}_-\} \times (F(x) - Q)$ from X into $Z = \mathbb{R} \times Y$. Clearly, this mapping cannot be regular near $(\bar{x}, (f(\bar{x}), 0))$. (Indeed, if U is a small neighborhood of $\bar{x}$, then $\Psi(U)$ cannot contain points $(f(\bar{x}) - \varepsilon, 0)$).

https://doi.org/10.1017/S1446788715000695 Published online by Cambridge University Press

It follows that the negation of any condition sufficient for regularity is a necessary condition for $\overline{x}$ to be a local solution of the problem. Applying [59, Theorem 6.17] and [59, Corollary 6.18], we get the following result.

THEOREM 8.2. *Assume that* $F : \mathbb{R}^n \to \mathbb{R}^m$ *is Lipschitz in a neighborhood of* $\overline{x}$. *If* $\overline{x}$ *is a local solution of (8.1), then there is a nonzero pair* (λ, y^*) *such that* $\lambda \geq 0$, $y^* \in N(Q, \overline{y})$ *and*

$$0 \in \partial(\lambda f + (y^* \circ F|_C))(\overline{x}). \tag{8.2}$$

This formulation needs some comments. We have stated the theorem in finite dimensions for simplicity, its infinite-dimensional version can be found, for example, in [50]. Note, further, that a more customary formulation would be

$$0 \in \partial(\lambda f + (y^* \circ F))(\overline{x}) + N(C, \overline{x}).$$

This condition is usually more convenient (constraints are separated) but, in general, is weaker than (8.2). It is equivalent to (8.2) if, for example, $C = X$ (obvious) or if both f and F are continuously differentiable and the constraint qualification

$$0 \in F'(\overline{x})y^* + N_C(\overline{x}), \quad y^* \in N_Q(F(\overline{x})) \Rightarrow y^* = 0$$

is satisfied (see, for example, [97], Example 10.8) which means that $F|_C$ is transversal to Q at $\overline{x}$ (Proposition 7.38).

Finally, we observe that the necessary condition is stated in the Lagrangian form. Again, such a condition can be substantially more precise than the 'separated' condition $0 \in \lambda\partial f(\overline{x}) + \partial(y^* \circ F)(\overline{x})$ (say, in the absence of the constraint $x \in C$) which, in various forms, often appears in the literature. Both conditions are equivalent if, say, f is continuously differentiable.

The 'noncovering' approach to the necessary optimality condition was first applied (probably by Warga [104]) in a fairly classical setting of the standard optimal control problem. Warga refers not to the Lyusternik–Graves theorem but to the result of Yorke [107], which is a weakened version of the theorem for integral operators associated with ordinary differential equations. But already, in the same year, the controllability–optimality dichotomy appeared as the main tool for proving necessary conditions for nonsmooth optimal control in the papers by Clarke [23] and Warga [105]. In the context of an abstract optimization problem, a noncovering criterion seems to have been first applied by Dmitruk *et al.* [28] to problems with a finite number of functional constraints and, recently, to problems with mixed structure (partly smooth and partly close to convex), by Avakov *et al.* [6]. In the next Section 8.3, we demonstrate the work of this techniques for an abstract relaxed optimal control problem. Theorem 8.2 in an infinite-dimensional setting was obtained in [50] with the same proof based on the noncovering criterion.

https://doi.org/10.1017/S1446788715000695 Published online by Cambridge University Press

8.2.2. Exact penalty. The immediate predecessor of the approach we are going to discuss here was the idea of an 'exact penalty' offered by Clarke [21, 24]: if f attains a local minimum on a closed set S at $\bar{x} \in S$ and satisfies the Lipschitz condition near $\bar{x}$, then $\bar{x}$ is a point of unconstrained minimum of $g(x) = f(x) + Kd(x, S)$ with K greater than the Lipschitz constant of f near $\bar{x}$. Clarke used a fairly sophisticated reduction technique to apply this idea to problems with functional constraints. However, the arguments are simplified dramatically by directly invoking regularity considerations.

Let us return to the problem (8.1), assuming, as above, that F is single-valued Lipschitz $X = \mathbb{R}^n$, $Y = \mathbb{R}^m$ and set as in [59, Theorem 6.17]: that is,

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

Then our problem can be reformulated as

$$\text{minimize } f(x) \quad \text{subject to } 0 \in \Phi(x).$$

Suppose that Φ is subregular at $(\bar{x}, 0)$. This means that there is some $K_0 > 0$ such that $d(x, \Phi^{-1}(0)) \le K_0 d(0, \Phi(x))$ for x of a neighborhood of $\bar{x}$. But $\Phi^{-1}(0)$ is the feasible set of our problem, so that there is some other $K_1 > 0$ such that the function $f(x) + K_1 d(0, \Phi(x))$ attains a local minimum at $\bar{x}$ or, equivalently, the function $f(x) + K_1 d(y, F(x) - Q)$ attains a local minimum at $\bar{x}$ subject to $x \in C$. The last function is Lipschitz continuous near $\bar{x}$, and hence there is a K such that

$$g(x) = f(x) + K(d(y, F(x) - Q) + d(x, C)$$

attains an unconditional minimum at $\bar{x}$.

If, on the other hand, Φ is not subregular at $\bar{x}$, [59, Theorem 6.1] and [59, Theorem 6.17] imply, together, that $0 \in \partial(y^* \circ F)(\bar{x}) + N(C, \bar{x})$ for some nonzero $y^* \in N(Q, F(\bar{x}))$. From here, we easily get a weakened version of Theorem 8.2 with the Lagrangian condition replaced by its 'separated' versions

$$0 \in \partial f(\bar{x}) + \partial(y^* \circ F)(\bar{x}) + N(C, \bar{x}), \quad y^* \in N(Q, F(\bar{x})).$$

This is a definite drawback, as we have already mentioned, which, however, is counterbalanced by some serious advantages. First, we note that g is defined in terms of the original data which makes it possible to study higher-order optimality conditions using this function. This is how such a technique was used for the first time in [47] in order to get the necessary optimality conditions that had been obtained earlier by Levitin *et al.* [72].

Another advantage is that the second approach is more universal. It can work for problems for which using scalarized coderivatives is either difficult or just impossible, for example in problems involving inclusions $0 \in \Phi(x)$ with general set-valued Φ. This is a typical case in optimal control of dynamic systems described by differential inclusions. Loewen [77] was the first to use this approach to prove a maximum principle in a free right end point problem of that sort. The analytic challenge in his proof was to find an upper estimate for the distance to the feasible set. However, the next step in the development, the 'optimality alternative' discussed below, excludes any need for such an estimate.

https://doi.org/10.1017/S1446788715000695 Published online by Cambridge University Press

8.2.3. Optimality alternative. Consider the abstract problem with (X, d) being a complete metric space: that is,

$$\text{minimize } f(x) \quad \text{subject to } x \in Q \subset X.$$

THEOREM 8.3. *Let φ be a nonnegative lower semicontinuous function on X equal to zero at $\overline{x}$. If $\overline{x} \in Q$ is a local solution to the problem, then the following alternative holds true:*

- *either there is a $\lambda > 0$ such that the function $\lambda f + \varphi$ has an unconstrained local minimum at $\overline{x}$; or*
- *there is a sequence $(x_n) \to \overline{x}$ such that $\varphi(x_n) < n^{-1} d(x_n, Q)$ and the function $x \mapsto \varphi(x) + n^{-1} d(x, x_n)$ attains a local minimum at x_n, for each n.*

We shall speak about a *regular case* if the first option takes place and a *singular* or *nonregular case* otherwise.

PROOF. Indeed, either there is an $R > 0$ such that $R\varphi(x) \geq d(x, Q)$ for all x of a neighborhood of $\overline{x}$, or there is a sequence (z_n) converging to $\overline{x}$ and such that $n^2\varphi(z_n) < d(z_n, Q)$. In the first case (as f is Lipschitz), for $x \notin Q$ and $u \in Q$ close to x (so that, for example, $d(x, u) < 2d(x, Q)$),

$$f(x) \geq f(u) - Ld(x, u) \geq f(\overline{x}) - 2LR\varphi(x),$$

if L is a Lipschitz constant of f.

As X is complete and φ is lower semicontinuous, we can apply Ekeland's principle to φ (taking into account that $\varphi(z_n) < \inf \varphi + n^{-2} d(z_n, Q)$) and find x_n such that $d(x_n, z_n) \leq n^{-1} d(z_n, Q)$, $\varphi(x_n) \leq \varphi(z_n)$ and $\varphi(x) + n^{-1} d(x, x_n) > \varphi(x_n)$ for $x \neq x_n$. Finally,

$$d(x_n, Q) \geq d(z_n, Q) - d(x_n, z_n) \geq (1 - n^{-1})\, d(z_n, Q) \geq (1 - n^{-1}) n^2 \varphi(z_n) \geq n\varphi(x_n),$$

as claimed. □

Thus a constrained problem reduces to one or a sequence of unconstrained minimization problems. Hopefully, such problems can be easier to analyse thanks to the freedom of choosing φ which we call a *test function* in the subsequent work. Even before the alternative was explicitly stated, it was *de facto* used to prove the maximum principle in various problems of optimal control (see [43, 52, 101]). Here is a brief account of how the alternative works for optimal control of systems governed by differential inclusions.

8.2.4. Optimal control of differential inclusion. As the first example of application of the alternative, we shall briefly consider the following problem of optimal control of a system governed by differential inclusion: (see, also, Section 8.3 below) minimize

$$\ell(x(0), x(T)) \tag{8.3}$$

on trajectories of the differential inclusion

$$\dot{x} \in F(t, x), \tag{8.4}$$

https://doi.org/10.1017/S1446788715000695 Published online by Cambridge University Press

satisfying the end point condition

$$(x(0), x(T)) \in S. \tag{8.5}$$

The natural space to treat the problem is $W^{1,1}$. Let $\overline{x}(\cdot)$ be a local solution. For any $x(\cdot) \in W^{1,1}$ set

$$\varphi(x(\cdot)) = \int_0^T d(\dot{x}(t), F(t, x(t)))\, dt + d((x(0), x(T)), S).$$

Clearly, φ is nonnegative and $\varphi(\overline{x}(\cdot)) = 0$. Thus, if ℓ is a Lipschitz function, we can apply the alternative to get the necessary optimality condition. According to the alternative, either there is a $\lambda > 0$ such that $\overline{x}(\cdot)$ is a local minimum of

$$\lambda\ell(x(0), x(T)) + d((x(0), x(T)), S) + \int_0^T d(\dot{x}(t), F(t, x(t)))\, dt,$$

or there is a sequence $(x_n(\cdot))$ converging to $\overline{x}(\cdot)$ such that every $x_n(\cdot)$ is not feasible in (8.3)–(8.5) and is a local minimum of the functional

$$d((x(0), x(T)), S) + \int_0^T d(\dot{x}(t), F(t, x(t)))\, dt + n^{-1}(\|x(0) - x_n(0)\| + \int_0^T \|\dot{x}(t) - \dot{x}_n(t)\|\, dt).$$

In both cases, we get an (unconstrained) Bolza problem. Analysis of such a problem needs different techniques and we refer to [52, 101], where necessary optimality conditions for the problem were obtained along these lines. A more general result was established a few years later by Clarke [25] (actually the most general for optimal control of differential inclusions so far) but a shorter proof of Clarke's theorem based on optimality alternative is now also available [60].

To conclude, I wish to note that this is not the only possible application of regularity related ideas to optimal control. We can refer to [102] for the discussion of the role of metric regularity in the Hamilton–Jacoby theory of optimal control.

8.2.5. Constraint qualification. The last question we intend to briefly discuss in this subsection concerns constraint qualifications in optimization problems. They often play an important role in proofs, but their basic function is to guarantee that the multiplier λ of the cost function in the necessary (for example, Lagrangian) optimality conditions is positive. The point is that constraint qualifications are often connected with regularity properties of the constraint mapping. We shall discuss just one example.

Let us say that the problem is *normal* at a certain feasible point if the constraint mapping is regular at the point. The *problem is normal* if either the feasible set is empty or the problem is normal at every feasible point. In the case of the problem (8.1), the constraint mapping is the restriction of F to C, so, by [59, Theorem 6.17], normality is guaranteed if F is transversal to Q, that is, if $y^* \in N(Q, F(x))$ and $0 \in D^*F|_C(\overline{x}, 0)(y^*)$ imply, together, that $y^* = 0$, which, in turn, implies that

$$0 \in \partial(y^* \circ F)(x) + N(C, x) \quad \text{and} \quad y^* \in N(Q, F(x)) \Rightarrow y^* = 0. \tag{8.6}$$

This is the now standard constrained qualification in nonsmooth optimization (see, for example, [33, 67, 79, 97]). If f and F are continuously differentiable and the sets C and Q are convex, (8.6) is dual to Robinson's constraint qualification [88].

8.3. An abstract relaxed optimal control problem. Here we apply the optimality alternative to get the necessary optimality condition in the problem

$$\text{minimize } f(x) \quad \text{subject to } F(x,u) = 0,\ x \in S, u \in U. \tag{8.7}$$

Here, $F : X \times U \to Y$, X and Y are separable Banach spaces and U is a set. The problem is similar to problems with mixed smooth and convex structures studied in [64, 99]. But, contrary to [64, 99], here we do not assume that F is continuously differentiable in x. We shall formulate the requirements on F a bit later. First, we need to introduce and discuss some necessary concepts.

We say that a continuous mapping $F : X \to Y$ is *semi-Fredholm* at $\overline{x}$ if it has at $\overline{x}$ a strict prederivative of the form $\mathcal{H}(x) = Ax + \|h\|Q$, where $A : X \to Y$ is a linear bounded operator that sends X onto a closed subspace of Y of finite codimension and $Q \subset Y$ is a compact set (that can be assumed to be convex and symmetric). We say, furthermore, that $S \subset X$ is *finite-dimensionally generated* if $S = \Lambda^{-1}(P)$, where $\Lambda : X \to R^n$ is a continuous linear operator and $P \subset \mathbb{R}^n$ is closed.

Proposition 8.4 (Noncovering principle for (8.7) [43, 50]). *Let $F : X \to Y$ be semi-Fredholm at $\overline{x}$, and let S be a finite-dimensionally generated subset of X. Further, let $F|_S$ be the restriction of F to S, that is, the set-valued mapping equal to $\{F(x)\}$ on S and $\emptyset$ outside of S. If $F|_S$ is not regular near $\overline{x}$, then there is a $y^* \neq 0$ such that $0 \in \partial_G(y^* \circ F)(\overline{x}) + N_G(S, \overline{x})$. Moreover, the weak*-closure of the set of such y^* with norm one does not contain zero*[1].

We intend to use this principle to prove the following theorem.

Theorem 8.5. *Let $(\overline{x}, \overline{u})$ be a solution of (8.7). We assume that:*

($\mathbf{A}_1$) *f satisfies the Lipschitz condition in a neighborhood of $\overline{x}$;*

($\mathbf{A}_2$) *for any $u \in U$, the mapping $F(\cdot, u)$ is Lipschitz in a neighborhood of $\overline{x}$ and $F(\cdot, \overline{u})$ is semi-Fredholm at $\overline{x}$;*

($\mathbf{A}_3$) *$F(x, U)$ is a convex set for any x of a neighborhood of $\overline{x}$; and*

($\mathbf{A}_4$) *S is finite-dimensionally generated.*

Further, let $\mathcal{L}(\lambda, y^*, x, u) = \lambda f(x) + \langle y^*, F(x,u)\rangle$ be the Lagrangian of the problem. Then there are $\lambda \geq 0$ and $y^* \in Y^*$ such that the following relations hold true:

$$\lambda + \|y^*\| > 0 \quad (\textit{nontriviality});$$

$$0 \in \partial_G \mathcal{L}(\lambda, y^*, \cdot, \overline{u})(\overline{x}) + N_G(S, \overline{x}) \quad (\textit{Euler–Lagrange inclusion}); \textit{and}$$

$$\langle y^*, F(\overline{x}, \overline{u})\rangle \geq \langle y^*, F(\overline{x}, u)\rangle \quad \forall u \in U\ (\textit{the maximum principle}).$$

Proof. Given a finite collection $\mathcal{U} = (u_1, \ldots, u_k)$ of elements of U, we define a mapping $\Phi_{\mathcal{U}} : X \times \mathbb{R}^k \to Y$ by

$$\Phi_{\mathcal{U}}(x, \alpha_1, \ldots, \alpha_k) = F(x, \overline{u}) + \sum_{i=1}^{k} \alpha_i (F(x, u_i) - F(x, \overline{u})).$$

[1]A more general version of this result can be found in many publications related to 'point estimates' and compactness properties of subdifferentials; see, for example [29, 51, 65, 66, 79].

https://doi.org/10.1017/S1446788715000695 Published online by Cambridge University Press

It is an easy matter to see that this mapping is also semi-Fredholm at $(\bar{x}, 0)$.

Consider the problem

$$\text{minimize } f(x) \quad \text{subject to } \Phi_{\mathcal{U}}(x, \alpha_1, \ldots, \alpha_k) = 0, x \in S, \alpha_i \geq 0. \tag{$\mathbf{P}_{\mathcal{U}}$}$$

Then $(\bar{x}, 0, \ldots, 0)$ solves the problem (as immediately follows from $(\mathbf{A}_3)$). Further, let $\Psi : X \times \mathbb{R}^k \to Y$ be defined by

$$\Psi(x, \alpha_0, \ldots, \alpha_k) = (f(x) + \alpha_0, \Phi_{\mathcal{U}}(x, \alpha_1, \ldots, \alpha_k)).$$

This mapping cannot be regular in a neighborhood of $(\bar{x}, 0, \ldots, 0)$ because no point $(f(\bar{x}) - \varepsilon, 0, \ldots, 0)$ can be a value of Ψ at $x \in S$ close to $\bar{x}$ and α close to zero. It is an easy matter to verify that Ψ is also semi-Fredholm at $(\bar{x}, 0, \ldots, 0)$ and we can apply Proposition 8.4.

Set $\hat{S} = S \times \mathbb{R}_-^{k+1}$, $\hat{\mathcal{L}}(\lambda, y^*, x, \alpha_0, \ldots, \alpha_k) = \lambda(f(x) + \alpha_0) + \langle y^*, \phi_n(x, \alpha_0, \ldots, \alpha_k)\rangle$. By the proposition, there are multipliers $(\lambda, y^*) \neq 0$ such that

$$0 \in \partial_G \hat{\mathcal{L}}(\lambda, y^*, \cdot)(\bar{x}, 0, \ldots, 0) + N_G(\hat{S}, (\bar{x}, 0, \ldots, 0)).$$

Using the standard rules of subdifferential calculus,

$$\begin{aligned}
&N_G(\hat{S}, (\bar{x}, 0, \ldots, 0)) = N_G(\bar{x}, S) \times \mathbb{R}_-^{k+1},\\
&\partial_G \hat{\mathcal{L}}(\lambda, y^*, \cdot)(\bar{x}, 0, \ldots, 0)\\
&\qquad \subset \partial_G \mathcal{L}(\lambda, y^*, \cdot, \bar{u})(\bar{x}) + (\lambda, \langle y^*, F(\bar{x}, u_1) - F(\bar{x}, \bar{u})\rangle, \ldots, \langle y^*, F(\bar{x}, u_i) - F(\bar{x}, \bar{u})\rangle).
\end{aligned}$$

It follows that there are $\xi_i \leq 0, i = 0, \ldots, k$ such that

$$\begin{gathered}
0 \in \partial_G \mathcal{L}(\lambda, y^*, \cdot, \bar{u})(\bar{x}) + N_G(S, \bar{x}),\\
\lambda = -\xi_0 \geq 0,\\
\langle y^*, F(\bar{x}, u_i) - F(\bar{x}, \bar{u})\rangle = \xi_i \geq 0, \quad i = 1, \ldots, k.
\end{gathered}$$

The relations remain obviously valid if we replace λ, y^* by $r\lambda, ry^*$ with some positive r. Thus, for any finite collection $(u_1, \ldots, u_k) \subset U$ we can find a pair of multipliers (λ, y^*) satisfying the three above mentioned relations and the normalization condition $\lambda + \|y^*\| = 1$. Let $\Omega(u_1, \ldots, u_k)$ be the weak*-closure of all such pairs. Then $\Omega(u_1, \ldots, u_k)$ is weak*-compact and, by Proposition 8.4, does not contain zero. It remains to notice that the increase of the set $(u_1, \ldots, u_k)$ may result only in decrease of $\Omega(u_1, \ldots, u_k)$ and, therefore, there is a nonzero pair λ, y^* common to all sets $\Omega(u_1, \ldots, u_k)$. □

8.4. Genericity in tame optimization. Here by 'tame optimization' we mean optimization problems with semialgebraic data. We consider the same class of problem as in (8.1). This time, however, we are interested in the effects of perturbations and shall work with a family of problems depending on a parameter p: that is,

$$\text{minimize } f(x, p) \quad \text{subject to } F(x, p) \in Q, x \in C.$$

https://doi.org/10.1017/S1446788715000695 Published online by Cambridge University Press

Here x is an argument in the problem and p is a parameter. So subdifferentials and derivatives that will appear below are always with respect to x alone. If p is fixed, then we denote the corresponding problem by $\mathcal{P}_p$.

Before we continue, we have to mention that, for a semialgebraic set $S \subset \mathbb{R}^n$, the following properties are equivalent:

- S is a set of first Baire category in $\mathbb{R}^n$;
- S has n-dimensional Lebesgue measure zero; and
- $\dim S < n$.

Thus, when we deal with semialgebraic objects, for example in $\mathbb{R}^k$, the word 'generic' means 'up to a semialgebraic set of dimension smaller than k'.

We shall assume that p is taken from an open set $P \subset \mathbb{R}^k$ and, as before, $x \in \mathbb{R}^n$ and F takes values in $\mathbb{R}^m$. Our main assumption is that

the restriction $F|_C(x, p)$ of F to C is transversal to Q.

This is definitely the case when $k = m$ and $F(x, p) = F(x) - p$. As to F itself, we assume that it is continuous with respect to (x, p) and locally Lipschitz in x. The sets C and Q, as usual, are assumed to be closed.

Theorem 8.6 (Generic normality). *Under the stated assumptions for a generic $p \in P$, the mapping $F|_C(\cdot, p)$ is transversal to Q. Thus, for a generic p, the problem $\mathcal{P}_p$ is normal.*

Proof. The first statement is immediate from Theorem 7.39, while the second follows from the comments following the statement of [59, Theorem 6.17]. □

Let us call a point x, which is feasible in $\mathcal{P}_p$, a *critical point* of the problem if the nondegenerate Lagrangian necessary condition of 8.2.1

$$0 \in \partial(f + (y^* \circ F|_C))(x, p), \quad y^* \in N(Q, F(x, p))$$

is satisfied. In this case, the value of f at x is called a *critical value* of $\mathcal{P}_p$.

Theorem 8.7 (Generic finiteness of critical values). *If under the stated assumptions, $\mathcal{P}_p$ is normal, then the problem may have only a finite number of critical values. Thus there is an integer N such that, for a generic p, the number of critical values in the problem does not exceed N.*

Proof. Consider the function

$$\mathcal{L}_p(x, y, y^*) = f(x, p) + \langle y^*, F|_C(x, p) - y\rangle + i_Q(y).$$

As follows from the standard calculus rules,

$$\partial\mathcal{L}_p(x, y, y^*) = \partial(f + y^* \circ F|_C)(x, p) \times (N(Q, y) - y^*) \times \{F(x, p) - y\}.$$

Thus (x, y, y^*) is a critical point of $\mathcal{L}_p$ if and only if $F(x, p) = y$, $0 \in N(Q, y) - y^*$: that is, if $y \in Q$ and $y^* \in N(Q, y)$ and $0 \in \partial(f + y^* \circ F|_C)(x, p)$. In other words, (x, y, y^*)

https://doi.org/10.1017/S1446788715000695 Published online by Cambridge University Press

is a critical point of $\mathcal{L}_p$ if and only if x is a feasible point in (**P**), $y = F(x, p)$ and the necessary optimality condition is satisfied at x with y^* being the Lagrange multiplier. We also see that, in this case, $\mathcal{L}_p(x, y, y^*) = f(x, p)$. In other words, critical values of the problem are precisely critical values of $\mathcal{L}$.

By the Sard theorem, $\mathcal{L}_p$ may have at most a finite number of critical values, which proves the theorem. □

The last result that we are going to present here has been so far proved only under some additional assumptions about elements of the problem. We shall explain it for the classical case, although the semialgebraic nature of the data remains crucial.

Theorem 8.8 (Generic finiteness of critical points). *Assume that $p = (q, y)$ with $q \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$ and $f(x, p) = f(x) - \langle q, x\rangle$, $F(x, p) = F(x) - y$, with $f(x)$ and $F(x)$ both being continuously differentiable. Assume, further, that the sets C and Q are closed and convex. Then there is an integer N such that, for a generic p, the number of pairs (x, y^*), such that x is a critical point in $\mathcal{P}_p$ and y^* is a corresponding Lagrange multiplier does not exceed N.*

The theorem follows from the two results below that contain valuable information about geometry of subdifferential mappings of semialgebraic functions.

Proposition 8.9 (Dimension of the subdifferential graph [36]). *The dimension of the graph of the subdifferential (whether Fréchet, limiting or Clarke) mapping of a semialgebraic function on R^n is n.*

Proposition 8.10 (Finiteness of preimage [36, 56]). *Let $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ be a semialgebraic set-valued mapping such that* $\dim(\operatorname{Graph} F) \leq n$. *If y is a regular value of F, then $F^{-1}(y)$ contains at most a finite number of elements. Thus there is an integer N such that, for a generic y, the number of elements in $F^{-1}(y)$ cannot exceed N.*

To see how the propositions lead to the proof of the theorem, we first note that $D^*F|_C(x)(y^*) = F'(x)y^* + N_C(x)$ if $x \in C$, F is smooth and C is convex. By [59, Theorem 6.15], $F|_C$ is transversal to Q if and only if

$$x \in C, \quad F(x) \in Q + y, \quad 0 \in F'(x)y^* + N_C(x), \quad y^* \in N(Q, F(x) - y) \;\Rightarrow y^* = 0, \tag{8.8}$$

and, by Theorem 7.39, this holds for a generic y.

Consider the function

$$g(x, y) = f(x) + i_C(x) + i_Q(F(x) - y).$$

By Proposition 8.9, the dimension of the graph of its subdifferential is $n + m$. Then so is the dimension of the graph of the mapping

$$\Gamma(x, y^*) = \{(q, y) : (q, y^*) \in \partial g(x, y)\}.$$

Now, by the Sard theorem, generic (q, y) is a regular value of Γ so, by (Proposition 8.10), for a generic (q, y) there is a finite number of (x, y^*) such that $(q, y) \in \Gamma(x, y^*)$.

Finally, if, for such (q, y), the qualification condition (8.8) is satisfied, then

$$\partial g(x, y) = \{(q, y^*) : f'(x) + (y^* \circ F(\cdot))'(x) + N(C, x), y^* \in N(Q, F(x) - y)\}$$

(even if Q is not convex; see, again, [97, Exercise 10.8]) which, in particular, means that x is a critical point of $\mathcal{P}_p$ and y^* is a Lagrange multiplier in the problem.

8.5. Method of alternating projection. This is one of the most popular methods to solve the feasibility problem due to its simplicity and efficiency. The feasibility problem, in its simplest form, consists of finding a common point of two sets, say, Q and S. The recipe offered by the method of alternating projection is the following: starting with a certain x_0, we choose, for $k = 0, 1, \ldots,$

$$x_{2k+1} \in \pi_Q(x_{2k}), \quad x_{2k+2} \in \pi_S(x_{2k+1}),$$

where $\pi_Q(x)$ is the collection of points of Q closest to x, and so on.

Von Neumann was the first to show, in the mid-1930s (see [103]), that, in the case of two subspaces, the method converges to a certain point in the intersection of two closed subspaces in a Hilbert space (depending, of course, on the starting point). Later, in the 1960s, Bregman [17] and Gubin *et al.* [44] applied it to convex subsets in $\mathbb{R}^n$. In particular, it was shown, in [44], that the convergence is linear if relative interiors of the sets meet. Later, Bauschke and Borwein [11] proved linear convergence if the sets are subtransversal at any common point.

But in computational practice, the method was successfully applied even for nonconvex sets. The first explanation was given by Lewis *et al.* [75]: if, at a certain point $\overline{x}$ in the intersection, the sets are transversal and at least one of the sets is not 'too nonconvex' in a certain sense (super-regular in the terminology of the authors), then there is linear convergence of alternating projections to a certain point, common to the sets (not necessarily $\overline{x}$), if the starting point is sufficiently close to $\overline{x}$. Recently, it was shown, by Druzviatskyj *et al.* [35], that transversality alone guarantees linear convergence. In fact, linear convergence was proved in [35] under a substantially weaker condition of 'intrinsic transversality' of the sets, but we believe that the geometric essence of the phenomenon is captured by the fact that transversality implies linear convergence.

Here is a short proof of linear convergence under the transversality assumption. Set

$$\varphi(x, y) = i_Q(x) + i_S(y) + \|x - y\|.$$

We claim that, if Q and S are transversal at $\overline{x} \in Q \cap S$, then there are $\kappa > 0$ and $\delta > 0$ such that, for any $x \in Q$, $y \in S$ close to $\overline{x}$,

$$\max\{|\nabla\varphi(\cdot, y)|(x), |\nabla\varphi(x, \cdot)|(y)\} \geq \kappa.$$

To this end, we first note that, by [59, Theorem 6.12],

$$\theta = \sup\{\langle u, v\rangle : u \in N(Q, \overline{x}), v \in -N(S, \overline{x}), \|u\| = \|v\| = 1\} < 1.$$

https://doi.org/10.1017/S1446788715000695 Published online by Cambridge University Press

Fix a certain $\kappa \in (0,1)$ and assume that there are sequences $(x_n) \subset Q$, $(y_n) \subset S$, $x_n \neq y_n$, converging to $\overline{x}$ and such that

$$|\nabla\varphi(\,\cdot\,, y_n)|(x_n) < \kappa, \quad |\nabla\varphi(x_n, \cdot)|(y_n) < \kappa,$$

that is, the functions

$$x \mapsto \varphi(x, y_n) + \kappa\|x - x_n\| \quad \text{and} \quad y \mapsto \varphi(x_n, y) + \kappa\|y - y_n\|$$

attain local minima, respectively, at x_n and y_n. This means that

$$0 \in w_n^* + \frac{x_n - y_n}{\|x_n - y_n\|} + \kappa B, \quad 0 \in z_n^* + \frac{y_n - x_n}{\|x_n - y_n\|} + \kappa B$$

for some $w_n^* \in N(Q, x_n)$ and $z_n^* \in N(S, y_n)$. Thus, for any limit point (w^*, z^*) of (w_n^*, z_n^*),

$$w^* = e + a, \quad z^* = -e + b,$$

where $\|e\| = 1, \|a\| \le \kappa, \|b\| \le \kappa$. Consequently,

$$\theta \ge \frac{\langle e + a, e + b\rangle}{\|e + a\|\,\|e + b\|} \ge \frac{(1-\kappa)^2}{(1+\kappa)^2}$$

and we get

$$\kappa \ge \frac{1 - \sqrt{\theta}}{1 + \sqrt{\theta}}.$$

This proves the claim.

Then $\pi_Q(y) = \operatorname{argmin} \varphi(\,\cdot\,, y)$ and the method of alternating projections can be written as

$$x_{n+1} \in \operatorname{argmin} \varphi(x_n, \cdot), \quad x_{n+2} \in \operatorname{argmin} \varphi(\,\cdot\,, x_{n+1}).$$

We, obviously, have $|\nabla\varphi(x_n, \cdot)|(x_{n+1})| = 0$. For a given x (not necessarily in Q), consider the function $\psi_x(y) = i_S(y) + \|x - y\|$. For any $c \in (0,1)$, condition $|\nabla\psi_x|(x_{n+1}) \le c$ obviously holds if

$$\langle x - x_{n+1}, x_n - x_{n+1}\rangle \ge \sqrt{1 - c^2}\|x - x_{n+1}\|\,\|x_n - x_{n+1}\|.$$

Take a $c < \kappa$ and let K_c be the collection of x satisfying the above inequality. This is an ice-cream cone with vertex at x_{n+1}. If $x \in Q \cap K_c$, then $\nabla\varphi(\,\cdot\,, x_n + 1)(x) \ge \kappa > c$. On the other hand, as is easy to see, the distance from x_n to the boundary of K_c is precisely cr, where $r = \|x_n - x_{n+1}\|$. Applying the basic lemma for error bounds (Lemma 7.1), we conclude that there is an $x \in Q$ with $\varphi(x, x_{n+1}) \le \varphi(x_n, x_{n+1}) - c\kappa\|x_{n+1} - x_n\|$. It follows that

$$\|x_{n+2} - x_{n+1}\| = \varphi(x_{n+2}, x_{n+1}) \le (1 - c^2)\|x_n - x_{n+1}\|,$$

which is linear convergence of (x_n).

https://doi.org/10.1017/S1446788715000695 Published online by Cambridge University Press

8.6. Generalized equations. By a generalized equation we mean the relation

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

where f is single-valued and $F : X \rightrightarrows Y$ is a set-valued mapping. Variational inequalities and necessary optimality conditions in constraint optimization with smooth cost and constraint functions are typical examples. The problem discussed in the theorem below is what happens with the set of solutions of the generalized equation if the single-valued term is slightly perturbed.

THEOREM 8.11 (Implicit function for generalized equations). *Let X, P be metric spaces and let Y be a normed space. Consider the generalized equation*

$$0 \in f(x, p) + F(x), \tag{8.9}$$

where $f : X \times P \to Y$ and $F : X \rightrightarrows Y$. Let $(\bar{x}, \bar{p})$ be a solution to the equation. Set $\bar{y} = -f(\bar{x}, \bar{p})$ and suppose that the following two properties hold:

(a) *either X or the graph of F is complete in the product metric and F is regular near $(\bar{x}, \bar{y})$ with $\operatorname{sur} F(\bar{x}|\bar{y}) > r$;*

(b) *there is a $\rho > 0$ such that f is continuous on $\mathring{B}(\bar{x}, \rho) \times \mathring{B}(\bar{p}, \rho)$ and $f(\cdot, p)$ satisfies, on $\mathring{B}(\bar{x}, \rho)$, the Lipschitz condition with constant $\ell < r$ for all $p \in \mathring{B}(\bar{p}, \rho)$.*

Let $S(p)$ stand for the solution mapping of (8.9). Then

$$d(x, S(p')) \le (r - \ell)^{-1} \|f(x, p) - f(x, p')\|$$

if $x \in S(p)$ is close to $\bar{x}$ and p, p' are sufficiently close to $\bar{p}$. Thus, if $f(x, \cdot)$ satisfies the Lipschitz condition with constant α on a neighborhood of $\bar{p}$ for all $x \in \mathring{B}(\bar{x}, \rho)$, then $S(\cdot)$ has the Aubin property near $(\bar{p}, \bar{x})$ with $\operatorname{lip} S(\bar{p}|\bar{x}) \le \alpha(r - \ell)^{-1}$.

Finally, if, in addition, F is strongly regular near $(\bar{x}, \bar{y})$, then $S(\cdot)$ has a Lipschitz localization $s(\cdot)$ at $(\bar{x}, \bar{y})$ with Lipschitz constant not greater than $\alpha(r - \ell)^{-1}$, so that

$$d(s(p), s(p')) \le (r - \ell)^{-1} \|f(s(p), p) - f(s(p), p')\| \le \alpha(r - \ell)^{-1} d(p, p').$$

PROOF. Set $G(x, p) = f(x, p) + F(x)$ and let $H(p, z) = (G(\cdot, p))^{-1}(z)$, so that $S(p) = H(p, 0)$. As the Lipschitz constants of functions $f(\cdot, p)$ are bounded by the same ℓ for all $p \in \mathring{B}(\bar{p}, \rho)$, it follows, from [59, Theorem 4.5], that there is a $\delta > 0$ such that, for every $p \in \mathring{B}(\bar{p}, \rho)$, the inequality $d(x, H(p, z)) \le (r - \ell)^{-1} d(z, G(x, p))$ holds if $d(x, \bar{x}) < \delta$ and $\|z - z(p)\| < \delta$, where $z(p) = f(\bar{x}, p) - f(\bar{x}, \bar{p}) \in G(\bar{x}, p)$. As f is continuous, we can choose $\lambda > 0$ such that $\|z(p)\| < \delta$ for $p \in \mathring{B}(\bar{p}, \lambda)$. For such p, $0 \in \mathring{B}(z(p), \delta)$ and, therefore, if $d(p, p') < \lambda$ and $0 \in f(x, p) + F(x) = G(x, p)$, we get,

$$\begin{aligned} d(x, S(p')) &\le (r - \ell)^{-1} d(0, G(x, p')) = (r - \ell)^{-1} d(0, f(x, p') + F(x)) \\ &= (r - \ell)^{-1} d(-f(x, p'), F(x)) \le (r - \ell)^{-1} \|f(x, p') - f(x, p)\|. \end{aligned}$$

This proves the first part of the theorem. The second now follows from [59, Theorem 4.13]. ◻

https://doi.org/10.1017/S1446788715000695 Published online by Cambridge University Press

The concept of a generalized equation was introduced by Robinson in [91]. The theorems proved in [91, 92] corresponded to f being continuously differentiable in x and F being either a maximal monotone operator or $F(x) = N(C, x)$, where C is a closed convex set. We refer to [33] for further results and bibliographic comments on generalized equations, which are among the central objects of interest in the monograph.

An earlier version of part (a) of the theorem, with a less precise estimate, can be found in [67, Theorem 4.9]. Part (b) of the theorem, relating to strong regularity, is the basic statement of [33, Theorem 5F.4] (generalizing the earlier results of Robinson in [92, 93]; see, also, [29] for an earlier result). Our proof, however, is different: here, the theorem appears as a direct consequence of Milyutin's perturbation theorem. Note that, in most of the related results in [33], it is assumed (following [93]) that there exists a 'strict estimator $h(x)$ for f of modulus ℓ' such that $\mathrm{sur}(F + h)(x|\overline{y} + h(\overline{x})) \geq r$. This is a fairly convenient device for practical purposes but it adds no generality to the result as the case with h reduces to the setting of the theorem if we replace $F + h$ by F and $f - h$ by f.

8.7. Variational inequalities over polyhedral sets. *Variational inequality* is a relation of the form

$$0 \in \varphi(x) + N(C, x), \tag{8.10}$$

where $\varphi : \mathbb{R}^n \to \mathbb{R}^n$ is a single-valued mapping and $C \subset \mathbb{R}^n$ is a convex set. If C is a cone, it is equivalent to

$$x \in K, \quad -\varphi(x) \in K^\circ, \quad \langle x, F(x)\rangle = 0.$$

The problem of finding such an x is known as a *complementarity problem* (see, for example, [40]). Problems of this kind typically appear in nonlinear programming in connection with necessary optimality conditions.

Consider, for instance, the problem

$$\text{minimize } f_0(x) \quad \text{subject to } f_i(x) = 0,\ i = 1, \ldots, k,\ f_i(x) \leq 0,\ i = k+1, \ldots, m,$$

with $f_0, \ldots, f_m$ twice continuously differentiable. If $\overline{x}$ is a solution of the problem, then (assuming that the problem is normal and setting $f = (f_1, \ldots, f_m)$) there is a $\overline{y} \in \mathbb{R}^m$ such that

$$\nabla f_0(\overline{x}) + \langle \overline{y}, \nabla f(\overline{x})\rangle = 0.$$

Setting

$$\varphi(x, y) = \begin{pmatrix} \nabla f_0(\overline{x}) + \langle \overline{y}, \nabla f(\overline{x})\rangle, \\ f(x) \end{pmatrix}; \quad C = \mathbb{R}^n \times \mathbb{R}^{m-k},$$

we see that $(\bar{x}, \bar{y})$ solves (8.10) (with x replaced by (x, y)).

Consider the set-valued mapping $\Psi(x) = \varphi(x) + N(C, x)$ associated with (8.10) and assume that C is a convex polyhedral set. What can be said about regularity of such a mapping near a certain $(\bar{x}, \bar{y}) \in \mathrm{Graph}\, \Phi$? Applying Milyutin's perturbation theorem [59, Theorem 4.5] and [59, Theorem 4.11] and taking into account that the Lipschitz constant of $h \to \varphi(x + h) - \phi'(x)h$ at zero is zero, we immediately get the following proposition.

https://doi.org/10.1017/S1446788715000695 Published online by Cambridge University Press

PROPOSITION 8.12. *Let* $\bar{y} \in \Psi(\bar{x})$ *for some* $\bar{x} \in C$. *Set* $A = \varphi'(\bar{x})$ *and* $\hat{\Psi}(x) = Ax + N(C - \bar{x}, x)$. *Then* Ψ *is (strongly) regular near* $(\bar{x}, \bar{y})$ *if and only if* $\hat{\Psi}$ *is (strongly) regular near* $(0, 0)$ *and* $\operatorname{sur}\Psi(\bar{x}|\bar{y}) = \operatorname{sur}\hat{\Psi}(0|0)$.

In other words, the regularity properties of Ψ are the same as of its 'linearization' $\hat{\Psi}$. Therefore, in what follows, we can deal only with the *linear variational inequality*

$$0 \in Ax + N(C, x)$$

and the associated mapping

$$\Phi(x) = Ax + N(C, x).$$

The key role in our analysis is played by the concept of a *face* of a polyhedral set C which is any closed subset F of C such that any segment $\Delta \subset C$ containing a point $x \in F$ in its interior lies in F. A face of C is *proper* if it is different from C. We refer to [96] for all necessary information about faces. The following facts are important for our discussion.

- The set $\mathcal{F}_C$ of all faces of C is finite.
- $F \in \mathcal{F}_C$ if and only if there is a $y \in \mathbb{R}^n$ such that $F = \{x \in C : \langle y, x\rangle \geq \langle y, u\rangle, \forall u \in C\}$.
- If $F, F' \in \mathcal{F}_C$ and $F \cap \operatorname{ri} F' \neq \emptyset$, then $F' \subset F$; a proper face of C lies in the relative boundary of C.
- If $F \in \mathcal{F}_C$ and $x_1,\ x_2$ belong to the relative interior of F, then $T(C, x_1) = T(C, x_2)$ and $N(C, x_1) = N(C, x_2)$.

The last property allows us to speak about the tangent and normal cones to C at F, which we shall denote by $T(C, F)$ and $N(C, F)$, respectively. It is an easy matter to see that

$$\dim F + \dim N(C, F) = n, \quad \dim(F + N(C, F)) = n. \tag{8.11}$$

For any $x \in C$, denote by $F_{\min}(x)$ the minimal element of $\mathcal{F}_C$ containing x. Clearly,

$$x \in F \in \mathcal{F}_C \quad \text{and} \quad F = F_{\min}(x) \ \Leftrightarrow\ x \in \operatorname{ri} F.$$

PROPOSITION 8.13. *If* Φ *is regular near* (x, y) *and* $F = F_{\min}(x)$, *then*

$$\dim(A(F) + N(C, F)) = n.$$

In particular, A *is one-to-one on* F.

PROOF. If $\dim F = 0$, then x is an extreme point of C, in which case $T(C, x)$ is a closed convex cone containing no lines and its polar, therefore, has nonempty interior. On the other hand, if $x \in \operatorname{int} C$, then $N(C, u) = \{0\}$ for all u of a neighborhood of x and $\Phi(u) = Au$ for such u. So, by regularity, A is an isomorphism.

Thus, in the subsequent work, we may assume that the dimensions of both F and $N(C, F)$ are positive. By changing (x, y) slightly, we can guarantee that y belong to the relative interior of $N(C, F)$. Let $\varepsilon > 0$ be so small that the distances from x and y to the relative boundaries of F and $N(C, F)$ are greater than ε. Then any (u, v)

https://doi.org/10.1017/S1446788715000695 Published online by Cambridge University Press

such that $u \in C$, $v \in N(C,u)$, $\|u - x\| < \varepsilon$, $\|v - y\| < \varepsilon$ must belong to $F \times N(C,F)$. This means that $\Phi(B(x,\varepsilon)) \cap B(y,\varepsilon) \subset A(F) + N(C,F)$ and the result follows from (8.11). Indeed, the dimension equality is immediate from the last inclusion. On the other hand, if A is not one-to one on F, then $\dim A(F) < \dim F$ and, by (8.11), $\dim A(F) + \dim N(C,F) < n$. □

Let $C \subset \mathbb{R}^n$ be a convex polyhedron, and let F be a proper face of C. Let L be the linear subspace spanned by F and let M be the linear subspace spanned by $N(C,F)$. These subspaces are complementary by (8.11) and orthogonal. By Proposition 8.13, $A(L)$ and M are also complementary subspaces if Φ is regular near any point of the graph.

Let π_M be the projection onto M parallel to $A(L)$, so that $\pi_M(A(F)) = 0$. Set $K_M = (T(C,F)) \cap M$ and let A_M be the restriction of $\pi_M \circ A$ to M. Then K_M is a convex polyhedral cone in M and its polar K_M° (in M) coincides with $N(C,F)$.

Definition 8.14. The set-valued mapping $\Phi_M(x) = A_M x + N(K_M, x)$, viewed as a mapping from M into M, will be called the *factorization of* Φ *along* F.

Observe that the graph of a factorization mapping is a union of convex polyhedral cones.

Proposition 8.15. *If* Φ *is regular near* $(\bar{x}, A\bar{x})$ *for some* $\bar{x} \in C$, *then the factorization of* Φ *along* $F = F_{\min}(\bar{x})$ *is globally regular on* $\mathbb{R}^n$.

Proof. Set $K_1 = T(C,F) = T(C,\bar{x})$ and consider the mapping $\Phi_1(x) = Ax + N(K_1, x)$. By Proposition 7.24, $\Phi_1(x) = \Phi(\bar{x} + x) - A\bar{x}$ for x close to zero. Therefore Φ_1 is regular near $(0,0)$, and hence globally regular, by Proposition 7.24. Observe that $K_1 = K_M + L$ and $K_1^\circ = N(K,F)$ and, consequently, $N(K_1,x) \subset N(K,\bar{x}) = N(K,F)$ for any $x \in K_1$.

As Φ_1 is globally regular, there is a $\rho > 0$ such that $d(x, \Phi_1^{-1}(z)) \le \rho d(z, \Phi_1(x))$ for all $x, z \in \mathbb{R}^n$. Take now $x, z \in M$. We have (taking into account that $N(K_M, x) = N(K_1, x + \xi)$ for any $\xi \in L$ and $A_M x = A(x + \xi)$ for some $\xi \in L$)

$$\begin{aligned} d(z, \Phi_M(x)) &= \inf\{\|z - A_M x - y\| : y \in N(K_M, x)\} \\ &\ge \inf\{\|z - A(x+\xi) - y\| : \xi \in L, y \in N(K_1, x+\xi)\} \\ &= \inf_{\xi \in L} d(z, \Phi_1(x+\xi)) = d(z, \Phi_1(w)) \end{aligned}$$

for some $w \in x + L$. On the other hand, there is a $w' \in \mathbb{R}^n$ such that $z \in \Phi_1(w')$ and $\|w - w'\| = d(w, \Phi_1^{-1}(z))$. Let x' be the orthogonal projection of w' to M. We have $z = Aw' + y$ for some $y \in N(K_1, w') \subset M$. Therefore $Aw' \in M$ and, moreover, $A_M x' = Aw'$. The latter is a consequence of the simple observation

$$v = Aw \in M, \quad x \in M, x \perp (w - x) \;\Rightarrow\; A_M x = v. \tag{8.12}$$

Indeed, $z = w - x \in L$, and hence $Ax = Aw + Az = v + Az$ and, as $v \in M$ and $Az \in A(L)$, $\pi_M(Ax) = v + \pi_M(Az) = v$.

It follows, as $N(K_M, x') = N(K_1, w')$), that $z \in \Phi_M(x')$ and

$$d(x, \Phi_M^{-1}(z)) \le \|x - x'\| \le \|w - w'\| = d(w, \Phi_1^{-1}(z)) \le \rho d(z, \Phi_1(w)) \le d(x, \Phi_M(x)),$$

that is, Φ_M is metrically regular on M (with the rate not greater than ρ). □

The following theorem is the key observation that paves the way for proof of the main result.

Theorem 8.16. *Let $C = K$ be a convex polyhedral cone. If Φ is regular near* $(0, 0)$ *(and hence globally regular by [59, Proposition 5.6]), then $A(K) \cap K^\circ = \{0\}$.*

Proof. The result is trivial if $n = 1$. Assume that it holds for $n = m - 1$, and let $m = n$. Note that the inclusion $A(K) \subset K^\circ$ can hold only if $K = \{0\}$. Indeed, if the inclusion is valid, then $\Phi(x) \in A(K) + K^\circ = K^\circ$ for any $x \in K$, so, by regularity, K° must coincide with the whole of $\mathbb{R}^n$ and hence $K = \{0\}$. Thus, if there is a nonzero $u \in A(K) \cap K^\circ$, we can harmlessly assume that u is a boundary point of K° and there is a nonzero $w \in N(K^\circ, u)$. Then $w \in K$ and $u \in N(K, w)$. Let $F = F_{\min}(w)$ so that $u \in N(K, F)$. As before, let L be the linear subspace spanned by F and M be the linear subspace spanned by $N(K, F)$. These subspaces are complementary by (8.11) and orthogonal. By Proposition 8.13, $A(L)$ and M are also complementary subspaces. Clearly, u does not belong either to L or to $A(L)$, the latter because, otherwise, the dimension of $A(F) + N(K, F)$ would be strictly smaller than n.

Consider the factorization Φ_M of Φ along F. Then $u \in K_M^\circ$, by definition. But, as follows from (8.12), u also belongs to $A_M(K_M)$. As Φ_M is regular, by Proposition 8.15, and $\dim M < m$, the existence of such a u contradicts the induction hypothesis. □

We are ready to state and prove the main result of the subsection.

Theorem 8.17 (Regularity implies strong regularity). *Let C be a polyhedral set and $\Phi(x) = Ax + N(C, x)$. If Φ is globally regular, then the inverse mapping Φ^{-1} is single-valued and Lipschitz on $\mathbb{R}^n$. Thus global regularity of Φ implies global strong regularity.*

In other words, the solution map of $y \in \Phi(x)$ is everywhere single-valued and Lipschitz.

Proof. We only need to show that Φ^{-1} is single-valued: the Lipschitz property will then automatically follow from regularity. The theorem is trivially valid if $n = 1$. Suppose it is true for $n \le m - 1$ and consider the case $n = m$. We have to show that, given a convex polyhedron $C \in \mathbb{R}^m$ and a linear operator A in $\mathbb{R}^m$ such that $\Phi(x) = Ax + N(C, x)$ is globally regular on $\mathbb{R}^n$, the equality $Ax + y = Au + z$ for some $x, u \in C$, $y \in N(C, x)$, $z \in N(C, u)$ can hold only if $x = u$ and $y = z$.

Step 1. To begin with, we observe that the equality $Au = Ax + y$ for some $u, x \in C$ and $y \in N(C, x)$ may hold only if $u = x$. Indeed, $u - x \in T(C, x)$. The same argument as in the proof of Proposition 8.15 shows that $\Phi_1(w) = Aw + N(T(C, x), w)$ is also globally regular and, therefore, by Theorem 8.16, $A(T(C, x)) \cap N(C, x) = \{0\}$. It follows that $A(u - x) = y = 0$. But regularity of Φ_1 implies (by Proposition 8.13) that A is one-to one on $T(C, x)$, and hence $u = x$.

Step 2. Now assume that for some $x, u \in C$, $u \neq x$, the equality $Ax + y = Au + z$, or $A(u - x) = y - z$, holds with $y \in N(C, x)$, $z \in N(C, u)$. We first show that this is impossible if $x \in F_{\min}(u)$. If under this condition $x \in \operatorname{ri} C$, then u is also in $\operatorname{ri} C$ which means that $N(C,x) = N(C,u)$ coincides with the orthogonal complement E to the subspace spanned by $C - C$. We have $y - z \in E$ and $u - x \in C - C$. By Proposition 8.13, $A(u - x) = y - z = 0$ and the second part of the proposition implies that $u = x$.

Now let $F = F_{\min}(x)$ be a proper face of C. Then $F \subset F_{\min}(u)$ and therefore $z \in N(C,F)$. As before, denote by L the subspace spanned by F and by M the subspace spanned by $N(C,F)$, and let Φ_M be the factorization of Φ along F. Set $v = A(u - x) = y - z$. Then $v \in M$ as both y and z are in $N(C,F)$. Let w be the orthogonal projection of $u - x$ onto M. Then, by (8.12), $Aw = v$ and, therefore, $A_M w = v$.

Thus (recall that $y, z \in M$)

$$A_M w + z = (\pi_M \circ A)(u - x) + z = \pi_M(A(u - x) + z) = \pi_M y = y.$$

On the other hand, it is clear that $y \in N(K_M, 0)$ and $z \in N(K_M, w)$. Indeed, $z \in N(T(C,x), u - x)$ (since $\langle z, v - x\rangle \le \langle z, u - x\rangle$ for all $v \in C$ on the one hand and, as we have seen, $z \in N(C,x)$, on the other) and, therefore, $z \in N(K_M, w)$ as $z \in M$ and $w - (u - x) \in L$. As $\dim M < m$, we conclude, by the induction hypotheses, that $w = 0$, and hence $u - x \in L$. But $A(u - x) = y - z \in M$ and a reference to proposition 8.13, again, proves that $u = x$.

Step 3. It remains to consider the case when neither x nor u belong to the minimal face of the other. Let κ be the modulus of metric regularity of Φ or any bigger number. Choose $\varepsilon > 0$ so small that the ball of radius $(1 + \kappa)\varepsilon$ around x does not meet any face $F \in \mathcal{F}_C$ not containing x. This means that $x \in F_{\min}(w)$ whenever $w \in C$ and $\|w - x\| \le (1 + \kappa)\varepsilon$. Further, let N be an integer big enough to guarantee that $\delta = N^{-1}\|y\| < \varepsilon$. Regularity of Φ allows us to construct recursively a finite sequence of pairs (u_k, z_k), $k = 0, 1, \dots, m$ such that

$$(u_0, z_0) = (u, z), \quad z_k \in F_{\max}(u_k), \quad u_k + z_k = x + (1 - m^{-1}k)y, \quad \|u_k - u_{k-1}\| \le \kappa\delta.$$

Then $u_N + z_N = x$. It follows, from the result obtained in the first step of the proof, that $u_N = x$. This, in turn, implies, as $u_0 \ne x$, that, for a certain k, we have $u_k \ne x$, $\|u_k - x\| \le \kappa\delta < \kappa\varepsilon$. By the choice of ε, this implies that $x \in F_{\min}(u_k)$. But, in this case, the result obtained at the second step excludes the possibility of the equality $u_k + z_k = x + (1 - m^{-1}k)y$ unless $u_k = x$. So, again, we get a contradiction that completes the proof. □

The material presented in this subsection is a part of my recent paper [61] which also contains a proof (based on a similar ideas) of another principal result concerning uniqueness and Lipschitz behavior of solutions to variational inequalities over polyhedral sets due to Robinson [94]. Theorem 8.17 was first stated by Dontchev and Rockafellar [32] with a comment that it follows from a comparison of the mentioned Robinson's result and another theorem (proved by Eaves and Rothblum [39]) containing an openness criterion for piecewise affine mappings. The given proof seems to give the first self-contained and reasonably short justification for the result. We refer the reader to [33, 40] for further details.

8.8. Differential inclusions—existence of solutions. Here we consider the Cauchy problem for differential inclusions

$$\dot{x} \in F(t,x), \quad x(0) = x_0, \tag{8.13}$$

where $F : \mathbb{R} \times \mathbb{R}^n \rightrightarrows \mathbb{R}^n$. We assume that;

- F is defined on some $\Delta \times U$ (that is, $F(t,x) \neq \emptyset$ for all $x \in U$ and almost all $t \in \Delta$), where $\Delta = [0,T]$ and U is an open subset of $\mathbb{R}^n$ containing x_0;
- the graph of $F(t,\cdot)$ is closed for almost every $t \in \Delta$; and
- F is measurable in t in the sense that the function $t \mapsto d((x,y), \mathrm{Graph}\, F(t,\cdot))$ is measurable for all pairs $(x,y) \in \mathbb{R}^n \times \mathbb{R}^n$.

By a solution of (8.13) on $[0,\tau] \subset [0,\Delta]$ we mean any absolutely continuous $x(t)$ defined on $[0,\tau]$ and such that $\dot{x}(t) \in F(t,x(t))$ almost everywhere on $[0,\tau]$.

THEOREM 8.18. *Assume that there is a summable $k(t)$ such that*

$$h(F(t,x), F(t,x')) \le k(t)\|x - x'\| \quad \forall x, x' \in U, \textit{almost everywhere on } [0,1]. \tag{8.14}$$

Further, let $x_0(\cdot)$ be an absolutely continuous function on $[0,T]$ with values in U such that $x_0(0) = x_0$ and $\rho(t) = d(\dot{x}_0(t), F(t,x_0(t)))$ is a summable function.

Then there is a solution of (8.13) defined on some $[0,\tau]$, $\tau > 0$. Specifically, set $r = d(x_0, \mathbb{R}^n \backslash U)$, and let $\tau \in (0,T]$ be so small that

$$1 > k_\tau = \int_0^\tau k(t)\,dt; \quad (1 - k_\tau)r > \xi_\tau = \int_0^\tau d(\dot{x}_0(t), F(t,x_0(t)))\,dt. \tag{8.15}$$

Then, for any $\varepsilon > 0$, there is a solution $x(\cdot)$ of (8.13) defined on $[0,\tau]$ and satisfying

$$\int_0^\tau \|\dot{x}(t) - \dot{x}_0(t)\| \le \frac{1+\varepsilon}{1-k_\tau}\xi_\tau. \tag{8.16}$$

Recall that $h(P,Q)$ is the Hausdorff distance between P and Q.

PROOF. We may set $x_0(t) \equiv 0$ (replacing, if necessary, $F(t,x)$ by $F(t,x_0(t)+x) - \dot{x}_0(t)$ and U by $r\overset{\circ}{B}$). Let $X = W_0^{1,1}[0,\tau]$ stand for the space of $\mathbb{R}^n$-valued absolutely continuous functions on $[0,\tau]$ equal to zero at zero with the norm

$$\|x(\cdot)\|_\tau = \int_0^\tau \|\dot{x}(t)\|\,dt,$$

and let I denote the identity map in X. Finally, let $\mathcal{F}$ be the set-valued mapping from X into itself that associates with every $x(\cdot)$ the collection of absolutely continuous functions $y(\cdot)$ such that $y(0) = 0$ and $\dot{y}(t) \in F(t,x(t))$ almost everywhere. We have to prove the existence of an $x(\cdot) \in X$ satisfying (8.16) and

$$0 \in (I - \mathcal{F})(x(\cdot)).$$

First, note that the graph of $\mathcal{F}$ is closed: that is, whenever $x_n(\cdot) \to x(\cdot)$, $y_n(\cdot) \in \mathcal{F}(x_n(\cdot))$ and $y_n(\cdot)$ norm converge to $y(\cdot)$, then $y(\cdot) \in \mathcal{F}(x(\cdot))$. Let $\mathcal{U}$ be the open ball

https://doi.org/10.1017/S1446788715000695 Published online by Cambridge University Press

of radius r around zero in X. Thus $x(t) \in U$ for any $t \in [0, \tau]$ whenever $x(\cdot) \in \mathcal{U}$ and, therefore, by (8.14), $\mathcal{F}$ is Lipschitz on $\mathcal{U}$ with $\operatorname{lip}\mathcal{F}(\mathcal{U}) \le k_\tau$. On the other hand, I is Milyutin regular on $\mathcal{U}$ with $\operatorname{sur}_m I(\mathcal{U}) = 1$. By [59, Theorem 4.2],

$$\operatorname{sur}_m(I - \mathcal{F})(\mathcal{U}) \ge 1 - k_\tau.$$

In particular, $B(y(\cdot), (1 - k_\tau)\rho) \subset (I - \mathcal{F})(\rho B)$ for any $y(\cdot) \in (I - \mathcal{F})(0)$ if $\rho < r$. Take a $y(\cdot) \in X$ such that $\dot{y}(t) \in F(t, 0)$ and $\|\dot{y}(t)\| = d(0, F(t, 0))$ almost everywhere. Then $\|y(\cdot)\|_\tau = \xi_\tau < (1 - k_\tau)r$ by (8.15). Thus $0 \in B(y(\cdot), (1 - k_\tau)\rho)$ for some $\rho < r$ and, therefore, there is an $x(\cdot)$ with $\|x(\cdot)\|_\tau < \rho$, $0 \in (I - \mathcal{F})(x(\cdot))$. □

The theorem is close to the original result of Filippov [41]. Versions of this results and its applications can be found in many subsequent publications (see, for example [4, 5]). Typical proofs of existence results for differential inclusions use either some iteration procedures or selection theorems to reduce the problem to existence of solutions of differential equations. Observe that our proof appeals to nonlocal regularity theory.

References

[1] F. J. Aragon Artacho and M. H. Geoffroy, 'Characterization of metric regularity of subdifferential', *J. Convex Anal.* **15** (2008), 365–380.

[2] F. J. Aragon Artacho and M. H. Geoffroy, 'Metric subregularity of convex subdifferential in Banach sopaces', *J. Nonlinear Convex Anal.* **15** (2014), 35–47.

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

[4] J.-P. Aubin and A. Cellina, *Differential Inclusions* (Springer, Berlin, 1984).

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

[6] E. R. Avakov, G. G. Magaril-Il'yaev and V. M. Tikhomirov, 'Lagrange's principle in extremal problems with constraints', *Russian Math. Surveys* **68**(3) (2013), 401–434.

[7] D. Azé and J.-N. Corvellec, 'On the sensitivity analysis of Hoffmann's constant for systems of linear inequalities', *SIAM J. Optim.* **12** (2002), 913–927.

[8] D. Azé and J.-N. Corvellec, 'Characterization of error bounds for lower semicontinuous functions on metric spaces', *ESAIM Control Optim. Calc. Var.* **10** (2004), 409–425.

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

[10] L. Barbet, M. Dambrine, A. Daniilidis and L. Rifford, 'Sard theorems for Lipschitz functions and applications in optimization', *Israel J. Math.*, to appear.

[11] H. Bauschke and J. M. Borwein, 'On projection algorithms for solving convex feasibility problems', *SIAM Rev.* **38** (1996), 367–426.

[12] J. Bochnak, M. Coste and M.-F. Roy, *Real Algebraic Geometry* (Springer, 1998).

[13] J. Bolte, A. Daniilidis and A. S. Lewis, 'The Morse–Sard theorem for non-differentiable subanalytic functions', *J. Math. Anal. Appl.* **321** (2006), 729–740.

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

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

[16] J. M. Borwein and A. S. Lewis, *Convex Analysis and Nonlinear Optimization*, 2nd edn. (Springer, 2006).

[17] L. M. Bregman, 'The method of successive projections for finding a common point of convex sets', *Dokl. Akad. Nauk* **6** (1965), 688–692.

https://doi.org/10.1017/S1446788715000695 Published online by Cambridge University Press

[18] M. J. Cánovas, M. A. Lopez, M. S. Mordukhovich and J. Parra, 'Variational analysis in linear semi-infinite and infinite programming 1: stability of linear inequality system of feasible solutions', *SIAM J. Optim.* **20** (2009), 1504–1526.

[19] R. Cibulka and M. Fabian, 'A note on Robinson–Ursescu and Lyusternik–Graves theorems', *Math. Program. B* **139** (2013), 89–101.

[20] R. Cibulka, M. Fabian and A. D. Ioffe, 'On primal regularity estimates for single-valued mappings', *J. Fixed Point Theory Appl.* **17** (2015), 187–208.

[21] F. H. Clarke, 'A new approach to Lagrange multipliers', *Math. Oper. Res.* **1** (1976), 165–174.

[22] F. H. Clarke, 'On the inverse function theorem', *Pacific J. Math.* **64** (1976), 97–102.

[23] F. H. Clarke, 'Necessary conditions for a general control problem', in: *Calculus of Variations and Control Theory* (ed. D. L. Russel) (Academic Press, 1976), 257–279.

[24] F. H. Clarke, *Optimization and Nonsmooth Analysis* (Wiley-Interscience, New York, 1983).

[25] F. H. Clarke, *Necessary Conditions in Dynamic Optimization*, Memoirs of the American Mathematical Society, 816 (AMS, Providence, RI, 2005).

[26] M. Coste, 'An introduction to *o*-minimal geometry', *Inst. Rech. Math.*, Univ. de Rennes, 1999 (http://name.math.univ-rennes1.fr/michel.coste/polyens/OMIN.pdf).

[27] A. Coulibali and J.-P. Crouzeix, 'Conditions numbers and error bounds in convex programming', *Math. Program. B* **116** (2009), 79–113.

[28] A. V. Dmitruk, A. A. Milyutin and N. P. Osmolovskii, 'Ljusternik's theorem and the theory of extrema', *Russian Math. Surveys* **35**(6) (1980), 11–51.

[29] A. L. Dontchev, 'Characterizations of Lipschitz stability in optimization', in: *Recent Developments in Well-Posed Variational Problems* (eds. R. Lucchetti and J. Revalski) (Kluwer, 1995), 95–115.

[30] A. L. Dontchev and H. Frankowska, 'Lyusternik–Graves theorem and fixed points', *Proc. Amer. Math. Soc.* **139** (2011), 521–534.

[31] A. L. Dontchev and H. Frankowska, 'Lyusternik–Graves theorem and fixed points 2', *J. Convex Anal.* **19** (2012), 955–974.

[32] A. L. Dontchev and R. T. Rockafellar, 'Characterizations of strong regularity for variational inequalities over polyhedral convex sets', *SIAM J. Optim.* **6** (1996), 1087–1105.

[33] A. L. Dontchev and R. T. Rockafellar, *Implicit Functions and Solution Mappings*, 2nd edn. (Springer, 2014).

[34] D. Drusvyatsky and A. D. Ioffe, 'Quadratic growth and critical point stability of semi-algebraic functions', *Math. Program. A* **153** (2015), 635–653.

[35] D. Drusvyatsky, A. D. Ioffe and A. S. Lewis, 'Alternating projections and coupling slope', *Found. Comput. Math.* **18** (2015), 1637–1651.

[36] D. Drusvyatsky and A. S. Lewis, 'Semi-algebraic functions have small subdifferentials', *Math. Program. B* **140** (2013), 5–29.

[37] D. Drusvyatsky and A. S. Lewis, 'Tilt stability, uniform quadratic growth and strong metric regularity of the subdifferential', *SIAM J. Optim.* **23** (2013), 256–267.

[38] A. Y. Dubovitslii and A. A. Milyutin, 'Zadachi na ekstremum pri nalichii ogranichenij', *Zh. Vychisl. Mat. Mat. Fiz.* **5**(3) (1965), 395–453; (in Russian; English translation: 'Problems for extremum under constraints', *USSR Comput. Math. Math. Phys.* **5** (1965)).

[39] B. C. Eaves and U. G. Rothblum, 'Relationships of properties of piecewise affine maps over ordered fields', *Linear Algebra Appl.* **132** (1990), 1–63.

[40] F. Facchinei and J. S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems* (Springer, New York, 2003).

[41] A. F. Filippov, 'Classical solutions of differential inclusions with multivalued right-hand sides', *SIAM J. Control Optim.* **5** (1967), 609–621.

[42] R. V. Gamkrelidze, *Foundations of Optimal Control* (TGU, Tbilisi, 1977), (in Russian).

[43] B. Ginsburg and A. D. Ioffe, 'Maximum principle for general semilinear optimal control problems', in: *Nonsmooth Analysis and Geometric Methods in Optimal Control* (eds. B. Mordukhovich and H. Sussmann) (Springer, 1995), 81–110.

[44] L. G. Gubin, B. T. Polyak and E. V. Raik, 'The method of projections for finding a common point of convex sets', *Comput. Math. Math. Phys.* **7** (1967), 1–24.

[45] A. J. Hoffman, 'On approximate solutions of systems of linear inequalities', *J. Res. Nat. Bureau Standards* **e49** (1952), 263–265.

[46] L. Hörmander, 'Sur la fonction d'appui des ensembles convexes dans un espace localement convex', *Ark. Mat.* **3** (1955), 181–186.

[47] A. D. Ioffe, 'Necessary and sufficient conditions for a local minimum. Reduction theorem and first order conditions', *SIAM J. Control Optim.* **17** (1979), part 1, 245–251, part 2, 261–265.

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

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

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

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

[52] A. D. Ioffe, 'Directional compactness, scalarization and nonsmooth semi-Fredholm mappings', *Nonlinear Anal.* **29** (1997), 201–219.

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

[54] A. D. Ioffe, 'Critical values of set-valued mappings with stratifiable graphs. Extensions of Sard and Smale–Sard theorems', *Proc. Amer. Math. Soc.* **136** (2008), 3111–3119.

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

[56] A. D. Ioffe, 'Variational analysis and mathematical economics 2. Nonsmooth regular economies', *Adv. Math. Econ.* **14** (2011), 17–38.

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

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

[59] A. D. Ioffe, 'Metric regularity—a survey. Part 1. Theory', *J. Aust. Math. Soc.*, in press.

[60] A. D. Ioffe, 'On necessary conditions for minimum', *Fundam. Prikl. Mat.* **19** (2014), 121–152.

[61] A. D. Ioffe, 'On variational inequalities over polyhedral sets', *Math. Program.* (submitted) arXiv:1508.06607v1.

[62] A. D. Ioffe and J.-P. Penot, 'Subdifferentials of performance functions and calculus of coderivatives of set-valued mappings', *Serdica Math. J.* **22** (1996), 359–384.

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

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

[65] A. Jourani and L. Thibault, 'Metric regularity and subdifferential calculus in Banach spaces', *Set-Valued Anal.* **3** (1995), 87–100.

[66] A. Jourani and L. Thibault, 'Verifiable conditions for openness and metric regularity in Banach spaces', *Trans. Amer. Math. Soc.* **347** (1995), 1255–1268.

[67] D. Klatte and B. Kummer, *Nonsmooth Equations in Optimization* (Kluwer, New York, 2002).

[68] D. Klatte and B. Kummer, 'Optimization methods and stability of inclusions in Banach spaces', *Math. Program. B* **117** (2009), 305–330.

[69] D. Klatte and W. Li, 'Asymptotic constraint qualifications and global error bounds for convex inequalities', *Math. Program.* **84** (1999), 137–160.

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

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

[72] E. S. Levitin, A. A. Milyutin and N. P. Osmolovskii, 'On conditions for a local minimum in problems with constraints', in: *Mathematical Economics and Functional Analysis* (ed. B. S. Mityagin) (Nauka, Moscow, 1974), (in Russian).

[73] A. S. Lewis, 'Ill-conditioned convex processes and conic linear systems', *Math. Oper. Res.* **24** (1999), 829–834.

[74] A. S. Lewis, 'Ill-conditioned inclusions', *Set-Valued Anal.* **9** (2001), 375–381.

[75] A. S. Lewis, D. R. Luke and J. Malick, 'Local linear convergence for alternating and averaged nonconvex projections', *Found. Comput. Math.* **9** (2009), 485–513.

[76] A. S. Lewis and J. S. Pang, 'Error bounds for convex inequality systems', in: *Generalized Convexity, Generalized Monotonicity: Recent Results* (eds. J. P. Crouzeix, J. E. Martinez-Legas and M. Volle) (Kluwer, 1998), 75–110.

[77] P. D. Loewen, 'Limits of Fréchet normals in nonsmooth analysis', in: *Optimization and Nonlinear Analysis* (eds. A. Ioffe, M. Markus and S. Reich) (Longman, New York, 1992).

[78] B. S. Mordukhovich, *Approximation Methods in Problems of Optimization and Control* (Nauka, Moscow, 1988), (in Russian).

[79] B. S. Mordukhovich, *Variational Analysis and Generalized Differentiation, vol. 1, 2* (Springer, 2006).

[80] B. S. Mordukhovich, J. F. Peña and V. Roschina, 'Applying metric regularity to compute a condition measure of a smoothing algorithm for matrix games', *SIAM J. Optim.* **20** (2010), 3490–3511.

[81] K. F. Ng and W. H. Yang, 'Error bounds for some convex functions and distance composite functions', *SIAM J. Optim.* **15** (2005), 1042–1056.

[82] Z. Páles, 'Inverse and implicit function theorems for nonsmooth maps in Banach spaces', *J. Math. Anal. Appl.* **209** (1997), 202–220.

[83] Z. Páles and V. Zeidan, 'Infinite dimensional Clarke generalized Jacobian', *J. Convex Anal.* **14** (2007), 433–454.

[84] Z. Páles and V. Zeidan, 'Generalized Jacobian for functions with infinite dimensional range and domain', *Set-Valued Anal.* **15** (2007), 331–375.

[85] R. Poliquin and R. T. Rockafellar, 'Tilt stability of a local minimum', *SIAM J. Optim.* **8** (1998), 287–299.

[86] S. M. Robinson, 'Normed convex processes', *Trans. Amer. Math. Soc.* **174** (1972), 127–140.

[87] S. M. Robinson, 'Stability theory for system of inequalities. Part 1: Linear systems', *SIAM J. Numer. Anal.* **12** (1975), 754–769.

[88] S. M. Robinson, 'Stability theory for system of inequalities. Part II: Differentiable nonlinear systems', *SIAM J. Numer. Anal.* **13** (1976), 497–513.

[89] S. M. Robinson, 'Regularity and stability for convex multivalued functions', *Math. Oper. Res.* **1** (1976), 130–143.

[90] S. M. Robinson, 'First order conditions for general nonlinear optimization', *SIAM J. Appl. Math.* **30** (1976), 597–608.

[91] S. M. Robinson, 'Generalized equations and their solutions. Part 1: Basic theory', *Math. Program. Study* **10** (1979), 128–141.

[92] S. M. Robinson, 'Strongly regular generalized equations', *Math. Oper. Res.* **5** (1980), 43–62.

[93] S. M. Robinson, 'An implicit function theorem for a class of nonsmooth functions', *Math. Oper. Res.* **16** (1991), 292–309.

[94] S. M. Robinson, 'Normal maps induced by linear transformations', *Math. Oper. Res.* **17** (1992), 691–714.

[95] R. T. Rockafellar, *Monotone Processes of Convex and Concave Type*, Memoirs of the American Mathematical Society, 77 (AMS, Providence, RI, 1967).

https://doi.org/10.1017/S1446788715000695 Published online by Cambridge University Press

[96] R. T. Rockafellar, *Convex Analysis* (Princeton University Press, Princeton, NJ, 1970).

[97] R. T. Rockafellar and R. J. B. Wets, *Variational Analysis* (Springer, 1998).

[98] W. Song, 'Calmness and error bounds for convex constraint systems', *SIAM J. Optim.* **17** (2006), 353–371.

[99] V. M. Tikhomirov, *Fundamental Principles of the Theory of Extremal Problems* (John Wiley & Sons, 1982).

[100] C. Ursescu, 'Multifunctions with closed convex graphs', *Czechoslovak Math. J.* **25** (1975), 438–411.

[101] R. B. Vinter, *Optimal Control* (Birkhäuser, Boston, 2000).

[102] R. B. Vinter, 'The role of metric regularity in state constrained optimal control', in: *Proceedings of the 44th IEEE Conference on Decision and Control, and the European Control Conference Seville, Spain* (2005), 262–265.

[103] J. von Neumann, 'Functional operators 2. The geometry of orthogonal spaces', *Ann. Math. Stud.* **22** (1950), Princeton.

[104] J. Warga, 'Controllability and necessary conditions in unilateral problems without differentiability assumptions', *SIAM J. Conrol Optim.* **14** (1976), 546–573.

[105] J. Warga, 'Derivate containers, inverse functions and controllability', in: *Calculus of Variations and Optimal Control* (ed. D. L. Russel) (Academic, New York, 1976), 13–46.

[106] Y. Yomdin and G. Comte, *Tame Geometry with Applications to Smooth Analysis*, Springer LNM Series, 1834 (2004).

[107] J. A. Yorke, 'The maximum principle and controllability of nonlinear equations', *SIAM J. Control Optim.* **10** (1972), 334–338.

[108] C. Zalinescu, *Convex Analysis in General Vector Spaces* (World Scientific, New Jersey, 2002).

[109] X. Y. Zheng and K. F. Ng, 'Metric regularity and constraint qualifications for convex inequalities in Banach spaces', *SIAM J. Optim.* **14** (2004), 757–772.

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

https://doi.org/10.1017/S1446788715000695 Published online by Cambridge University Press