Yokohama Publishers
ISSN 1349-8169 ONLINE JOURNAL

Pacific Journal of Optimization
Vol. 6, No. 2, pp. 341–352, May 2010

THE PRINCIPLE OF LAGRANGE AND CRITICAL VALUES OF OPTIMIZATION PROBLEMS

A.D. Ioffe

To the memory of A.M. Rubinov

Abstract: We prove that in a normal mathematical programming problem with Lipschitz data having some additional structural property (e.g. semi-algebraic or, more generally, definable in some o-minimal structure) there may be only finitely many critical values. That is to say, there are at most finite number of values the cost function may assume at points satisfying necessary optimality conditions. The Lipschitz condition can be dropped if it is a priori known that the principle of Lagrange holds for the problem. It is further shown that normality is a typical property in the sense that in a family of problems with good (in the same structural sense) dependence on the parameter the set of parameters for which the problem can be abnormal must have a smaller dimension in the parameter space.

Key words: *critical point and value, necessary optimality condition, subdifferential, normal problem, o-minimal structure, definable sets, functions and mappings*

Mathematics Subject Classification: *49J53, 58K05, 32B20, 65K10, 90C30, 90C31*

1 Introduction

One of the standard ways to treat an optimization problem, e.g. minimization of a certain function f_0 over a set of constraints suggested by the theory invokes the so called *principle of Lagrange* (see e.g.[10]). According to the principle we have to write the Lagrangian of the problem and then look for necessary optimality conditions for an unconstrained minimum of the Lagrangian. This approach can be effective only if the set of feasible points at which the Lagrangian necessary optimality condition is satisfied (*critical points* for short) is sufficiently small. One of the deepest results of the classical analysis and differential geometry, Sard's theorem, gives an estimate of the size of the set in terms of *critical values* of a sufficiently smooth function, that is to say, the values of the function at critical points.

In this paper we deal with critical values of optimization problems, which are defined as the values of the cost function at points at which the Lagrangian necessary optimality condition is satisfied. Loosely speaking, the main result of the paper (Theorem 5.1) says that an optimization problem (with a finite number of variables and constraints) has only a finite number of critical values if the cost function and constraint functions and sets have some good structural properties, the problem is normal and the principle of Lagrange holds. Moreover, it turns out that if the problem belongs to a family of problems depending on a finite dimensional parameter in a certain well structured manner, then there is a universal upper bound for the number of critical values in every problem of the family.

It has to be emphasized that in the paper we deal with, generally, non-smooth optimization problems. The simplest and easily understandable example of nonsmooth problems for which the theorem holds is provided by e.g. problems with the cost and constrained functions being differences of polyhedral functions and constraint sets being unions of polyhedral sets (so-called semi-linear functions and sets). The theorem however is valid for much broader classes of functions and sets, namely definable in a certain o-minimal structure. The theory of o-minimal structures whose development was triggered by the classical studies of semi-algebraic functions and sets by Łoiasiewizc, Hironaka and others, is now an extremely active area of model theory and algebraic geometry. It offers an enormous variety of functions and sets having remarkable analytic and geometric properties. An important point about these developments is that statements of many their results are easily understandable and are almost ready for application in analysis and optimization. Section 4 contains a brief discussion of the subject.

In the main theorem we assume that all data in the problem are locally Lipschitz. This guarantees that the principle of Lagrange does hold. The theorem remains valid if (instead of the Lipschitz assumption) we assume that the principle holds for the problem. It is not known however whether this is the case even if, say, only the cost function is lower semicontinuous and not locally Lipschitz.

The last mentioned property that makes the result possible is normality. The problem is normal if an equivalent of the Mangasarian-Fromowitz qualification condition is satisfied at every feasible point. For an abnormal problem the theorem is not valid even in the simplest case of linear programming. (Take for instance a problem with linearly dependent equality constraint functions in which every value is critical.) The subsequent theorems offer conditions which guarantee that in a family of problems (depending on a finite dimensional parameter) abnormality is a very rare phenomenon, even not in the usual sense of Baire category or measure negligibility but in the sense of lower dimensionality of the set of parameters for which the corresponding problem can be abnormal.

The next section contains necessary information from variational analysis including the definitions of critical and regular points and values and subdifferential quantitative characterization for regularity. Theorem 2.2 is an extension of well known coderivative scalarization theorem to a certain class of set-valued mappings (so called "constraint systems" which will be in the focus of our attention). In the third section we briefly discuss necessary optimality conditions in optimization problems, § 4 contains necessary information about o-minimal structures and definable sets and mappings, and the last § 5 - main results.

Throughout the papers all spaces are finite dimensional Euclidean.

$\boxed{2}$ Critical Points of Set-valued Mappings

Recall that a set-valued mapping $F : X \rightrightarrows Y$ is *metrically regular* near a point $(\bar{x}, \bar{y})$ of its graph if there are $K < \infty$ and $\varepsilon > 0$ such that the inequality

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

holds for all x, and y within ε of $\overline{x}$ and $\overline{y}$ respectively. The greatest lower bound of such K is called the *modulus of metric regularity* of F at $(\bar{x}, \bar{y})$ and is denoted $\mathrm{reg}F(\overline{x}|\overline{y})$.

An equivalent property: there are $r > 0$, $\varepsilon > 0$ (not necessarily the same) such that

$$B(y, rt) \subset F(B(x, t))$$

is called the *covering* or *openness at a linear rate* near $(\bar{x}, \bar{y})$. The least upper bound of all such r is called the *modulus of surjection* of F at $(\bar{x}, \bar{y})$ and is denoted $\mathrm{sur}F(\overline{x}|\overline{y})$.

Moreover, it turns out that the equality

$$\mathrm{sur}F(\overline{x}|\overline{y}) \cdot \mathrm{reg}F(\overline{x}|\overline{y}) = 1$$

always holds if we agree to set $\mathrm{reg}F(\overline{x}|\overline{y}) = \infty$ and $\mathrm{sur}F(\overline{x}|\overline{y}) = 0$ if the mapping is not metrically regular (open at a linear rate) near $(\bar{x}, \bar{y})$ and use the convention that $0 \times \infty = 1$.

Thus the following expression for the modulus of surjection is also available:

$$\mathrm{sur}F(\overline{x}|\overline{y}) = \sup_{\varepsilon>0} \inf_{\|(x,y)-(\bar{x},\bar{y}))\|<\varepsilon} \sup\{\alpha \geq 0 : \ \alpha d(x, F^{-1}(y)) \leq d(y, F(x))\}. \tag{2.1}$$

Definition 2.1. It is said that $(x, y) \in$ Graph F is a *critical point* of F if $\mathrm{sur}F(x|y) = 0$. Otherwise (x, y) is called a *regular* point of F and F is said to be *regular near* (x, y). A vector $y \in Y$ is a *critical value* of F if there is an x such that $y \in F(x)$ and $\mathrm{sur}F(x|y) = 0$. Otherwise, y is called a *regular* value.

Of course, if F is single-valued, then we refer only to the arguments when speaking about critical and regular points.

If f is an extended-real-valued function on X, we can associate with it the epigraphical set valued mapping

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

(with $(\mathrm{Epi}f)(x) = \emptyset$ if $f(x) = \infty$). Critical points of this mapping are called *(lower) critical* points of f. (Note that Graph [Epi f] = epi f, the epigraph of f.)

We further observe that (in case of a lower semicontinuous function) $(x^*, \beta) \in N(\mathrm{epi}\ f, (x, a))$ only if $\beta \leq 0$ when $\alpha = f(x)$ and $\beta = 0$ when $\alpha > f(x)$. Moreover $x^* \in \partial f(x)$ if and only if $(x^*, -1) \in N(\mathrm{epi}\ f, (x, f(x)))$.

Modern variational analysis provides for characterization of regular and critical points in terms of the *limiting coderivative* introduced (under a different name) in 1976 by Mordukhovich [12]. This is a brief summary of available results. We refer to [14, 17] for details.

The *Fréchet subdifferential* of f at x is the collection $\partial_F f(x)$ of vectors x^* satisfying the inequality

$$\langle x^*, h\rangle \leq \liminf_{\|h\|\to 0} \|h\|^{-1}(f(x + h) - f(x)).$$

This is always a convex closed set (possibly empty).

The *limiting subdifferential* of f at x. is

$$\partial f(x) = \limsup_{u\to x} \partial_F f(u),$$

with the“limsup” understood in the usual Painlevé-Kuratowski sense. This is a closed set which, in case of a Lipschitz function, is nonempty and bounded. If f is a convex function, its limiting subdifferential coincides with the subdifferential in the sense of convex analysis at every point.

Given a closed set $Q \subset X$, the *limiting normal cone* $N(Q, x)$ to Q at x is defined as the limiting subdifferential of the *indicator* of Q which is the function equal to zero on Q and infinity outside of Q. An equivalent and somewhat looser definition is that $x^* \in N(Q, x)$ if and there are sequences of $(x_n) \subset Q$ converging to x, $(x_n^*) \in X$ converging to x^* and (ε_n), (r_n) of positive numbers converging to zero such that

$$\langle x_n^*, u - x_n\rangle \leq \varepsilon_n\|u - x_n\|, \quad \text{if} \quad u \in Q, \ \|x_n - u\| < r_n.$$

If Q is a closed convex set and $x \in Q$, then $N(Q,x)$ is the normal cone to Q at x in the sense of convex analysis, that is $x^* \in N(Q,x)$ if and only if $\langle x, u-x\rangle \le 0$ for all $u \in Q$.

Now, if we have a (set-valued) mapping F from X into Y, the *(limiting) coderivative* of F at a point $(\bar{x},\bar{y}) \in \text{Graph } F$ is the set-valued mapping $D^*F(\bar{x},\bar{y})$ from Y into X defined as follows

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

If, for instance, F is a linear operator, the coderivative reduces to the adjoint operator.

It turns out that for a set-valued mapping with closed graph

$$\begin{aligned} \text{sur}F(\overline{x}|\overline{y}) &= \inf\{d(0, D^*F(\bar{x},\bar{y})(y^*)) : \ \|y^*\| = 1\} \\ &= \inf\{\|x^*\| : x^* \in D^*(\bar{x},\bar{y})(y^*), \ \|y^*\| = 1\} \end{aligned} \tag{2.2}$$

and the infimum is attained. Thus $(x,y) \in \text{Graph } F$ is a critical point of F if and only if $0 \in D^*F(x,y)(y^*)$ for some $y^* \ne 0$. In particular, if f is a lower semicontinuous function, then x is a lower critical point of f if and only if $0 \in \partial f(x)$.

We shall be specifically interested in set-valued mappings of a special kind, so called *constraint systems* which have the form

$$\Phi(x) = F(x) - Q,$$

where $Q \subset Y$ is closed and F is single-valued and locally Lipschitz. Coderivatives of such mappings admit a simpler representation in terms of subdifferentials of compositions $y^* \circ F$. The following theorem is the basic fact from the calculus of subdifferentials and coderivatives we need in this paper.

Theorem 2.2. *Let Φ be defined as above with F single-valued and Lipschitz near $x \in P$ and Q being a closed set. Then*

$$D^*\Phi(x,y)(y^*) = \begin{cases} \partial(y^* \circ F)(x), & \text{if} \quad y^* \in N(Q, F(x)-y), \\ \emptyset, & \text{otherwise.} \end{cases}$$

This result can be extracted from [11] and [14] (Lemma 5.23). A proof (in Russian) can also be found in [13] (Theorem 1.3.3). Below we give an alternative and sufficiently elementary proof.

Proof. The equality $D^*F(x)(y^*) = \partial(y^* \circ F)(x)$ was established in [6]. Also the inclusion

$$D^*\Phi(x,y)(y^*) \subset \begin{cases} D^*F(x)(y^*), & \text{if} \quad y^* \in N(Q, F(x)-y), \\ \emptyset, & \text{otherwise.} \end{cases} \tag{2.3}$$

follows from Theorem 4.32 of [14]. Thus we have only to prove equality in (2.3), that is that the set on the right must be a part of the set on the left.

So set $z = F(x) - y$, and let $x^* \in D^*F(x)(y^*)$, $y^* \in N(Q,z)$. We have to prove that $x^* \in D^*\Phi(x,y)(y^*)$. To this end we need to show, e.g. that there are sequences of vectors $(x_n) \subset X$, $(y_n) \subset Y$, $(x_n^*) \subset X$, $(y_n^*) \subset Y$ converging to x, y, x^* and y^* respectively and two sequences of positive numbers (δ_n), (r_n) converging to zero such that

$$\langle x_n^*, h\rangle - \langle y_n^*, w\rangle \le \delta_n(\|h\| + \|w\|), \quad \text{if } y_n + w \in \Phi(x_n + h);\ \|h\| + \|w\| < r_n. \tag{2.4}$$

As $x^* \in D^*F(x)(y^*)$, we do have sequences (x_n), (x_n^*), (y_n^*) converging to x, x^* and y^* respectively and two sequences of positive numbers (ε_n) and (ρ_n) converging to zero such that

$$\langle x_n^*, h\rangle - \langle y_n^*, v\rangle \le \varepsilon_n(\|h\| + \|v\|), \text{ if } F(x_n) + v = F(x_n + h);\ \|h\| + \|v\| < \rho_n. \tag{2.5}$$

Likewise, as $y^* \in N(Q, z)$, there are sequences $(z_n) \subset Q$ and (v_n^*) converging to z and y^* respectively and sequences of positive numbers $\varepsilon_n' \to 0$ and $\rho_n' \to 0$ such that

$$\langle v_n^*, \xi \rangle \le \varepsilon_n' \|\xi\|, \quad \text{if } z_n + \xi \in Q, \ \|\xi\| \le \rho_n'. \tag{2.6}$$

Clearly, $\|v_n^* - y_n^*\| \to 0$, so setting $\gamma_n = \varepsilon_n' + \|v_n^* - y_n^*\|$, we get from (2.6)

$$\langle y_n^*, \xi \rangle \le \gamma_n \|\xi\|, \quad \text{if } z_n + \xi \in Q, \ \|\xi\| \le \rho_n'. \tag{2.7}$$

Now take $r_n = (1 + L)^{-1} \min\{\rho_n, \rho_n'\}$, where L is Lipschitz constant of F, and let $y_n = F(x_n) - z_n \in \Phi(x_n)$. Let further $y_n + w \in \Phi(x_n + h) = F(x_n + h) - Q$ for some h, w satisfying $\|h\| + \|w\| < r_n$. Then

$$z_n' = F(x_n + h) - (y_n + w) = F(x_n + h) - F(x_n) + z_n - w \in Q.$$

Setting $\xi = z_n' - z_n$, we see that

$$\|\xi\| \le \|F(x_n + h) - F(x_n)\| + \|w\| \le L\|h\| + \|w\| \le (1 + L)r_n \le \rho_n'.$$

Thus,taking (2.4) and (2.7) into account, we get (with $v = F(x_n + h) - F(x_n)$)

$$\begin{array}{rcl} \langle x_n^*, h \rangle - \langle y_n^*, w \rangle & = & \langle x_n^*, h \rangle - \langle y_n^*, v \rangle + \langle y_n^*, \xi \rangle \\ & \le & \varepsilon_n(\|h\| + \|v\|) + \gamma_n \|\xi\| \\ & \le & \varepsilon_n(1 + L)\|h\| + \gamma_n(\|w\| + L\|h\|) \\ & \le & \delta_n(\|h\| + \|w\|) \end{array}$$

where $\delta_n = (1 + L)(\varepsilon_n + \gamma_n) \to 0$. □

Returning back to the definition of a critical point, we immediately get in view of (2.2)

Theorem 2.3. *Under the assumptions of Theorem 2.2,* $(x, y) \in \operatorname{Graph} \Phi$ *is a critical point of* Φ *if and only if there is a nonzero* $y^* \in N(Q, F(x) - y)$ *such that* $0 \in \partial(y^* \circ F)(x)$.

3 The Problem and the Necessary Optimality Condition

The problem to be considered can be stated as follows:

$$(\mathbf{P}) \qquad \text{minimize} \quad f(x) \quad \text{s.t.} \quad F(x) \in Q.$$

Here, as before, $x \in X$, $Q \subset Y$ and f is a function on X. Throughout the paper we assume (unless otherwise is explicitly stated) that the following assumptions hold:

(**A**) f and F are locally Lipschitz and Q is closed.

This is a fairly general scheme covering, of course, the standard problems of nonlinear programming with finitely many equality and inequality constraints and many other formulations (including basic problems of semi-infinite programming).

Set

$$\Psi(x) = \begin{pmatrix} f(x) - I\!R_- \\ F(x) - Q \end{pmatrix},$$

where $I\!R_-$ is the set of non-positive reals. Then $(\alpha, y) \in \Psi(x)$ means that $F(x) - y \in Q$ and $\alpha \ge f(x)$. Thus, if $\overline{x}$ is a local solution of the problem, then $(\overline{x}, (f(\overline{x}), 0))$ must be a critical point of Ψ. As $\Psi(x) = G(x) - P$, where $G(x) = (f(x), F(x))$ is locally Lipschitz and $P = I\!R_- \times Q$ is a closed set in $I\!R \times Y$, critical points of Ψ are characterized by Theorem 2.3. As a consequence of this observation we get the standard **necessary optimality condition** for a local minimum in the problem.

Theorem 3.1. *Let $\overline{x}$ is a local solution in the problem. Then there is a pair (λ, y^*) such that $\lambda \geq 0,\ \lambda + \|y^*\| > 0$, and*

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

The theorem basically says that the principle of Lagrange holds for the problem. To emphasize the subtleness of this conclusion, we note that it is still not known whether this remains true if f is only lower semicontinuous, not Lipschitz. The available necessary condition in this case is that $0 \in \partial(\lambda f)(\overline{x}) + \partial(y^* \circ F)(\overline{x})$ which may be a much weaker statement (and not implying that $\overline{x}$ is a critical point of a certain Lagrangian function).

Definition 3.2. The problem (**P**) is *normal* if for any feasible x the constraint mapping $\Phi(x) = F(x) - Q$ is regular at $(x, 0)$.

Thus in the normal problem the necessary optimality condition can be reformulated as follows: *if $\overline{x}$ is a local solution of (**P**), then there is a $y^* \in N(Q, 0)$ such that* $0 \in \partial(f + y^* \circ F)(\overline{x})$.

4 O-minimality

We start with the definition of the so-called *semi-linear sets*. An *open polyhedron* in $I\!R^n$ is by definition the intersection of a finite number of affine sets and open half-spaces:

$$P = \{x \in I\!R^n :\ \langle a_i, x\rangle = \alpha_i,\ i = 1, ..., k;\ \langle a_i, x\rangle < \alpha_i,\ i = k+1, ..., m\}.$$

A *semi-linear* set in $I\!R^n$ is defined as a finite union of open polyhedra.

Denote for a moment the collection of all semi-linear subsets of $I\!R^n$ by $\mathcal{S}_n$. It is an easy matter to verify that

(i) $\mathcal{S}_n$ is a Boolean algebra, that is it contains complements, finite unions and finite intersections of its elements;

(ii) every polyhedral set in $I\!R^n$ belongs to $\mathcal{S}_n$;

(iii) if $Q \in \mathcal{S}_n,\ P \in \mathcal{S}_m$, then $Q \times P \in S_{n+m}$;

(iv) semi-linear subsets of $I\!R$ are finite unions of points and open intervals;

(v) the projection mapping $(x_1, ..., x_n) \to (x_1, ..., x_{n-1})$ from $I\!R^n$ into $I\!R^{n-1}$ carries $\mathcal{S}_n$ into S_{n-1}.

It turns out that there are much broader systems of sets satisfying (i)-(v).

Definition 4.1. Let for any $n = 1, 2, ...,\ \mathcal{S}_n$ be a collection of subsets of $I\!R^n$. The sequence $(\mathcal{S}_n)$ is called *o-minimal structure* if the properties (i)-(v) are satisfied for all n. The elements of $\mathcal{S}_n$ are called in this case *definable sets* (in $\mathcal{S}$). A (set-valued) mapping from $I\!R^n$ into $I\!R^m$ is called *definable* if its graph is a definable subset of $I\!R^{n+m}$.

The most famous example of an o-minimal structure is provided by *semi-algebraic* sets which are finite unions of sets of solution of finite systems of algebraic equations and strict inequalities, e.g.

$$\{x \in I\!R^n :\ f_i(x) = 0,\ i = 1, ..., k,\ f_i(x) < 0,\ i = k+1, ..., r\},$$

where f_i are polynomials of n variables. Verification of the properties (i)-(iv) is equally elementary for semi-algebraic sets but (v) is a deep fact known as Tarski-Seidenberg theorem (see e.g. [2] for details). For more information about o-minimal structures see also [3].

Linear operations (addition and multiplication by scalar) are definable in any o-minimal structure; in o-minimal structures containing semi-algebraic sets also multiplication of functions is a definable mapping. There are also o-minimal structures containing semi-algebraic sets and exponential functions. In such structures taking powers is a definable operations. Moreover, given an o-minimal structure, we can be sure that

(a) *if $f(x, y)$ is a definable function, then so is $\varphi(x) = \inf_y f(x, y)$; in particular, the distance to a defnable set is a definable function;*

(b) *a composition of definable mappings is a definable mapping;*

(c) *the image and preimage of a definable set under a definable mapping are definable sets.*

We shall also mention several deeper results relating to o-minimal structures.

(d) (uniform boundedness theorem) if $F : I\!R^n \rightrightarrows I\!R^m$ is definable and $F(x)$ is either empty or finite for any x, then there is an integer N such that $\text{card}F(x) \le N$ for all x ([4], 4.4).

(e) (definable Sard theorem) the collection of critical values of a definable set-valued mapping $F : I\!R^n \rightrightarrows I\!R^m$ is a definable set of dimension strictly smaller than m ([1, 8, 9]).

In what follows we consider only o-minimal structures containing semi-algebraic sets (and hence containing multiplication among definable operations).

$\boxed{5}$ Main Theorems

We return back to our problem (**P**). Let $\mathcal{F} = \{x : \; 0 \in F(x) - Q\}$ stand for the feasible set in the problem. Following the terminology introduced in § 2, we say that $x \in \mathcal{F}$ is a *critical point of the problem* if the necessary optimality condition of § 3 holds at x. We shall say furthermore that an $\alpha \in I\!R$ is a *critical value* of the problem if there is a critical point $x \in \mathcal{F}$ such that $f(x) = \alpha$.

The first question we intend to study is how rich may be the set of critical values. In general, it could be very rich: take for instance a linear programming problem with linearly dependent equality constraints. In such a problem any admissible value is critical. Or we may identify a problem of unconstraint minimization with any function, and so the famous example of Whitney [18] shows that the set of critical values may contain an interval also in a nonlinear problem.

On the other hand, as follows from the definable Sard theorem of the preceding section, only finitely many critical values may appear in a problem of unconstraint minimization of a definable function. The theorem below shows that a similar property is also valid for constrained minimization problems with definable data.

Theorem 5.1. *Suppose that f, F and Q satisfy (**A**) and are definable in a certain o-minimal structure. Suppose further that (**P**) is a normal problem. Then (**P**) may have only finitely many critical values.*

Proof. The proof of the theorem is not difficult. Consider the function

$$\mathcal{L}(x, y, y^*) = f(x) + \langle y^*, F(x) - y\rangle + \delta_Q(y).$$

Here δ_Q is the indicator of Q. As follows from the standard calculus rules,

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

Thus, (x, y, y^*) is a lower critical point of $\mathcal{L}$ if and only if $F(x) = y$, $0 \in N(Q, y) - y^*$, that is $y \in Q$ and $y^* \in N(Q, y)$, and $0 \in \partial(f + y^* \circ F)(x)$. In other words, (x, y, y^*) is a lower critical point of $\mathcal{L}$ if and only if x is a feasible point in $(\mathbf{P})$, $y = F(x)$ and necessary optimality condition is satisfied at x with y^* being the Lagrange multiplier. We also see that in this case $\mathcal{L}(x, y, y^*) = f(x)$. In other words, critical values of the problem are precisely lower critical values of $\mathcal{L}$.

On the other hand, $\mathcal{L}$ is a definable function if so are f, F and Q. Therefore by the definable Sard theorem (e) $\mathcal{L}$ may have at most finitely many low critical values, whence the theorem. □

We next consider the case when the data depend on a certain parameter. In other words, we shall consider a family of problems

$$(\mathbf{P}(p)), \qquad \text{minimize} \quad f(x,p) \quad \text{s.t.} \quad F(x,p) \in Q(p),$$

where $Q : P \rightrightarrows Y$ is a set-valued mapping with closed and convex values and $P \subset I\!R^k$ is the set of parameters. We also denote by $\mathcal{F}(p)$ the feasible set in $(\mathbf{P}(p))$.

Theorem 5.2. *Suppose f and F are definable (as functions of both variables) as well as the set-valued mapping $Q(p)$. Assume further that for every $p \in P$ with $\mathcal{F}(p) \neq \emptyset$ the assumption $(\mathbf{A})$ is satisfied and $(\mathbf{P}(p))$ is a normal problem. Then there is an integer N such that the number of critical values in each problem does not exceed N for all $p \in P$.*

Proof. This is an immediate consequence of Theorem 5.2 and the uniform boundedness theorem (property (d) in the previous section). The only point that should be clarified is that the set-valued mapping $p \to C(p)$ which associates with every parameter p the collection of all critical values of $(\mathbf{P}(p))$ is definable. To prove the latter it is sufficient to show that the set

$$\{(x, y, p) : \ x \in \mathcal{F}(p) \text{ is a critical point of } (\mathbf{P}(p))\}$$

is definable (as the projection of this set onto the (p, y)-space is precisely the graph of $C(\cdot)$). As follows from the proof of Theorem 5.1, this set, in turn, is the projection to the (x, y, p)-space of the set

$$\{(x, y, y*, p) : \ (x, y, y^*) \text{ is a lower critical point of } \mathcal{L}(., p)\},$$

where $\mathcal{L}(x, y, y^*, p) = f(x, p) + \langle y^*, F(x, p) - y\rangle + \delta_{Q(p)}(y)$.

It is clear that $\mathcal{L}(x, y, y^*, p)$ is a definable function. Thus the problem reduces to the following: given a definable function $\varphi(x, p)$ which is lower semi-continuos in the first variable. Set

$$\mathcal{C} = \{(x, p) : \ x \text{ is a lower critical point of } \varphi(\cdot, p)\}.$$

We have to verify that $\mathcal{C}$ is a definable set.

As follows from the explanation given in §2, it is sifficient for that to show that the function $\psi(x, p) = \text{sur}\ \varphi(\cdot, p)(x, f(x))$ is definable (as $\mathcal{C}$ is the zero level set of ψ) which is an immediate consequence of (2.1).

Indeed, set $G(x, p) = \text{Epi}\ \varphi$, $H(\alpha, p) = \{x : \ \alpha \in G(x, p)\} = G(\cdot, p)^{-1}(\alpha)$. Both set-valued mappings are definable as so is φ. Therefore the distance functions $d(x, H(\lambda, p))$ and $d(\lambda, G(x, p))$ are definable as well as the function

$$\xi((x,\alpha),(u,\beta),p,r,\varepsilon) = \begin{cases} rd(x, H(\alpha,p)) - d(\alpha, G(x,p)), & \text{if } \|(x,\alpha) - (u,\beta\| \le \varepsilon, \\ & \qquad x \in \text{dom } f; \\ \infty, & \text{otherwise.} \end{cases}$$

By (2.1)

$$\psi(x,p) = \sup\{r \geq 0 : \inf_{\varepsilon>0} \xi((x,f(x)),(u,\beta),p,r,\varepsilon) \leq 0\}$$

which is a definable function due to the properties (a), (c) of the previous section. □

The last theorem immediately leads us to ask how typical is the normality property in a family of problems considered in Theorem 5.2. This question can be easily answered in a special (although probably the most important) case when $P = Y$ and

$$F(x,y) = F(x) - y. \tag{5.1}$$

Theorem 5.3. *Assume that $F(x,y)$ is defined by (5.1) and $F(x)$ and Q satisfy* (**A**). *Assume further that both $F(x)$ and Q are definable in some o-minimal structure. Then the collection of those y for which the problem* ($\mathbf{P}(y)$) *is not normal form a definable subset of Y of dimension strictly smaller than* $\dim Y$.

Proof. This also results from the definable Sard theorem applied to the same mapping $\Phi(x) = F(x) - Q$. Indeed, by Theorem 2.3 y is a critical value of the mapping if and only if there are an x with $y \in F(x) - Q$ and a nonzero $y^* \in N(Q, F(x) - y)$ such that $0 \in \partial(y^* \circ F)(x)$. But this means precisely that the problem ($\mathbf{P}(y)$) is not normal. □

The case of a general dependence on a parameter is more complicated. Below we give a positive answer when F is continuously differentiable. The following extension of the famous Lusternik's theorem is needed to get the answer. Recall that, given a closed set Q in a Banach space and an $\overline{x} \in Q$, the *tangent cone* (or *Dubovitzkii-Milyutin* tangent cone) to Q at $\overline{x}$ is defined as the collection of $h \in X$ such that $d(x + th, Q) = o(t)$ as $t \to +0$. We denote the cone $T(Q, \overline{x})$.

Proposition 5.4. *Let X and Y be Banach spaces, $F : X \to Y$ a mapping defined and continuously differentiable in a neighborhood of an $\overline{x} \in X$ and $Q \subset Y$ a closed set. We assume that $F(\overline{x}) \in Q$ and set as above $\Phi(x) = F(x) - Q$. Set finally*

$$\Gamma = \Phi^{-1}(0) = F^{-1}(Q) = \{x \in X : F(x) \in Q\}.$$

Then

$$T(\Gamma, \overline{x}) = (F'(\overline{x}))^{-1}(T(Q, F(\overline{x}))) = \{h \in X : F'(\overline{x})h \in T(Q, F(\overline{x}))\},$$

provided Φ is regular near $(\overline{x}, 0)$, that is $\mathrm{sur}\Phi(\overline{x}|0) > 0$.

Proof. The inclusion $\subset$ is trivial: if $h \in T(\Gamma, \overline{x})$, then for any $t > 0$ there is an $h(t) \in X$ such that $h(t) \to h$ as $t \to 0$ and $F(\overline{x} + th(t)) \in Q$. We have setting $v(t) = F'(\overline{x})h(t) \to F'(\overline{x})h$:

$$F(\overline{x}) + tv(t) + r(t) \in Q,$$

where $\|r(t)\| = o(t)$. Hence $F'(\overline{x})h \in T(Q, F(\overline{x}))$.

Let us prove the opposite inclusion. Let $h \in (F'(\overline{x}))^{-1}(T(Q, F(\overline{x})))$, that is $v = F'(\overline{x})h$ satisfies $d(F(\overline{x}) + tv, Q) = o(t)$. It follows that

$$d(0, \Psi(\overline{x} + th)) = d(F(\overline{x} + th), Q) = d(F(\overline{x}) + tF'(\overline{x})h, Q) + o(t) = o(t)$$

As Φ is regular near $(\overline{x}, 0)$, we can find a $K > 0$ and an $\varepsilon > 0$ such that

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

whenever $\|x - \overline{x}\| < \varepsilon,\ \|y\| < \varepsilon$. Taking $x = \overline{x} + th,\ y = 0$, we get (as $F(\overline{x}) \in Q$):

$$d(\overline{x} + th, \Phi^{-1}(0)) = d(\overline{x} + th, F^{-1}(Q)) \le o(t)$$

In other words, there is a $u(t)$ such that $F(u(t)) \in Q$ and $\|u(t) - (\overline{x} + th)\| = o(t)$, or equivalently $u(t) = \overline{x} + th(t) \in \Gamma$, and $\|h(t) - h\| \to 0$. □

Proposition 5.5. *Assume, in addition to the conditions of Proposition 5.4, that Q is convex. Then Φ is regular near $(\overline{x}, 0)$ if and only if*

$$\mathrm{Im}\ F'(\overline{x}) - T(Q, 0) = Y.$$

Proof. The proof that the equality implies regularity is due to Robinson [16]. The opposite implication in the finite dimensional context is elementary. (But it is also true in infinite dimensional case as a consequence of Robinson-Ursescu theorem - see e.g. [7], Thms 1.4 and 1.6a.) □

The following proposition establishes an inherited regularity property for constraint systems which plays a crucial role in studying parametric dependences.

Proposition 5.6. *Let again X, P and Y be finite dimensional Banach spaces. Asssume that $F:\ X \times P \to Y$ is defined and continuously differentiable on an open subset of $X \times P$ containing $(\overline{x}, \overline{p})$. Assume further that $Q \subset Y$ is closed and convex and $F(\overline{x}, \overline{p}) \in Q$. Set as before $\Phi(x, p) = F(x, p) - Q$, and $\Gamma = F^{-1}(Q) = \{(x, p) :\ F(x, p) \in Q\}$. Let finally $F^p : X \to Y$ be defined by $F^p(x) = F(x, p)$ and $\Phi^p = F^p - Q$. If zero is a regular value of Φ and $\overline{p}$ is a regular value of the projection $\pi : (x, p) \to p$ from Γ to P, then zero is a regular value of $\Phi^{\overline{p}}$.*

Proof. (cf. the proof of the transversality theorem in [5]). By Proposition 5.5 regularity of Φ at $((x, p), 0)$ is equivalent to Im $F'(x, p) - T(Q, 0) = Y$. So let $0 \in \Phi(x, \overline{p})$. By the assumptions, Φ is regular near $((x, \overline{p}), 0)$ whenever $F(x, \overline{p}) \in Q$. This means in particular that, given a $y \in Y$,we can find an $h \in X$ and a $q \in P$ and $v \in T(Q, 0)$ such that $F'_x(x, \overline{p}))h + F'_p(x, \overline{p})q - v = y$.

So take a $y \in Y$ and choose corresponding h, q and v. As Φ is regular near $(x, \overline{p})$, the tangent cone to Γ at $(x, \overline{p})$ consists of those $(\xi, \eta) \in X \times P$ which satisfy $F'_x(x, \overline{p}))\xi + F'_p(x, \overline{p})\eta \in T(Q, 0)$. (Here F'_x and F'_p stand for the partial derivatives of F with respect to x and p respectively.) As $\overline{p}$ is regular value of π, the projection of the tangent cone to P must be the whole of P. This means that we can find $h' \in X$ and $v' \in T(Q, 0)$ such that $F'_x(x, \overline{p}))h' - F'_p(x, \overline{p})q = v'$. This allows us to exclude the term containing q from the equality at the end of the previous paragraph and get

$$F'_x(x, \overline{p}))(h + h') - (v + v') = y$$

We notice now that $v + v' \in T(Q, 0)$ as the latter is a convex cone, so that $y \in \mathrm{Im}\ F'(x, \overline{p}) - T(Q, 0)$. This is true for any $y \in Y$, so by Proposition 5.5 $F^{\overline{p}}$ is regular near x. As x is, in turn an arbitrary point of $(F^{\overline{p}})^{-1}(Q)$, we conclude that zero is a regular value of $\Phi^{\overline{p}}$. □

We are able now to state and to easily prove the concluding result.

Theorem 5.7. *Assume that F and Q are definable in a certain o-minimal structure, $F(x, p)$ is continuosly differentiable on its domain and Q is closed and convex. Assume further that zero is a regular value of F. Then the collection of $p \in I\!R^k$ for which ($\mathbf{P}(p)$) is not normal is a definable subset of $I\!R^k$ of dimension strictly smaller than k.*

Proof. By definition $(\mathbf{P}(p))$ is not normal if zero is not a regular value of Φ^p. This in turn would mean (in view of Proposition 5.6) that p is not a regular value of the restriction of π to Γ, provided this restriction is a definable mapping. The latter however is immediate from the facts that Γ is a definable set as an inverse image of a point under a definable mapping (which Φ obviously is) and pi is a linear, hence definable mapping. □

References

[1] J. Bolte, A. Daniilidis, A.S. Lewis and M. Shiota, Clarke subgradients of stratifiable functions, *SIAM J. Optimization* 18 (2007) 556–572.

[2] J. Bonchak, M. Coste and M.F. Roy. *Real Algebraic Geometry*, E.M.G. vol. **36**, Springer, Berlin, 1998.

[3] L. van den Dries, *Tame Topology and O-minimal Structures*, Cambridge Univ. Press, Cambridge, 1998.

[4] L. van den Dries and C. Miller, Geometric categories and o-minimal structures, *Duke Math. J.* 84 (1996) 497–540.

[5] V. Guillemin and A. Poillack, *Differential Topology*, Prentice Hall, Englewood Cliffs N.J., 1974.

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

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

[8] A.D. Ioffe, A Sard theorem for tame set-valued mappings, *J. Math. Anal. Appl* 335 (2007) 882–901.

[9] A.D. Ioffe, Critical values of set-valued maps with stratifiable graphs: extensions of Sard and Smale-Sard theorems, *Proc. AMS* 136 (2008) 3111–3119.

[10] A.D. Ioffe and V.M. Tikhomirov, *Theory of Extremal Problems*, "Nauka", Moscow, 1974; English translation: North Holland - Elsevier, 1979.

[11] A.Ya. Kruger, Generalized differentials of nonsmooth functions and necessary optimality conditions for an extremum, *Siberian Math. J.* 26 (1985) 370–379.

[12] B.S. Mordukhovich, Maximum principle in problems of time optimal control with nonsmooth constraints, *J. Appl. Math. Mech.* 40 (1976) 960–969.

[13] B.S. Mordukhovich, *Approximation Methoids in Optimization and Control*, "Nauka", Moscow, 1988 (in Russian).

[14] B.S. Mordukhovich, *Variational Analysis and Generalized Differentiation*, vol 1. Springer, Berlin, 2006.

[15] S.M. Robinson, Regularity and stability for convex multifunctions, *Math. Oper. Res.* 1 (1976) 130–143.

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

[17] R.T. Rockafellar and R.J.B. Wets, *Variational Analysis*, Springer, Berlin, 1998.

[18] H.Whitney, A function not constant on a connected set of critical points, *Duke Math. J.* 1 (1935) 514–517.

[19] H. Whitney, Local properties of analytic varieties, in *Differential and Conbinatorial Topology*, Princeton Univ. Press, Princeton, 1965, pp. 205–244.

Manuscript received 1 August 2008
revised 28 September 2008
accepted for publication 30 September 2008

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