Weak Stationarity: Eliminating the Gap between Necessary and Sufficient Conditions

Alexander Kruger †

December 12, 2003

Abstract. Starting from known necessary extremality conditions in terms of strict subdifferentials and normals the notion of weak stationarity is introduced. It is defined in terms of initial space elements. The necessary conditions become necessary and sufficient (for stationarity).

Keywords: nonsmooth analysis, subdifferentials, normal cones, optimality conditions, set-valued mappings, Asplund spaces

1 Introduction

It is typical of nonconvex optimization problems that there is a gap between necessary and sufficient optimality conditions. If to proceed from the necessary conditions, then in general there exist points satisfying these conditions but not being optimal. Such points are usually referred to as *stationary*. Meanwhile if the necessary conditions are strong enough then the stationary points do possess some extremal properties and can be of practical interest.

To eliminate the gap between necessary and sufficient conditions, contrary to the traditional approach when stronger and stronger necessary conditions are deduced, it is suggested in the current paper to extend (weaken) the initial definition of optimality taking stationary points into consideration.

Formally, it is possible to speak of interrelations between the two conditions (groups of conditions): in terms of the initial (optimality or stationarity definitions) and dual (optimality conditions) spaces. The aim is to describe in terms of the initial space the set of points satisfying necessary optimality conditions expressed in terms of the dual space elements. As a result necessary conditions become also sufficient (for stationarity) and one can speak of (a kind of) duality between corresponding conditions.

Of course, the notion of stationarity depends on the type of the necessary conditions being considered. We consider below a group of necessary conditions (for different settings of extremal problems), obtained in the works of B. S. Mordukhovich and the author, and formulated in terms of so called *strict subdifferentials* and related to them constructions of *strict normal cones* and *strict coderivatives*.

The strict subdifferential is defined as a union of *Fréchet subdifferentials* calculated near the given point and thus accumulates information about "differential" properties of the function in the neighborhood of the point. This is also true for the defined on its basis *limiting subdifferential*.

Actually conditions, expressed in terms of strict (and limiting) subdifferentials are examples of so called "*fuzzy* conditions", when extremality at a point is characterized by "elementary" subdifferentials calculated at some points from its neighborhood. Corresponding stationarity conditions are also "fuzzy" in a sense: they estimate difference quotients for points from a neighborhood of a given point. They are weaker than traditional ones in which one of the points is assumed to be fixed. We will call the corresponding notion *weak stationarity* (the term *extended extremality* was used earlier [13], see also preceding definitions in [9, 10, 11, 12]).

There exist many different abstract definitions of optimality (extremality). In the current paper we proceed from the two definitions: of the *extremal point of the sets system* [15] (in terms of sets) and of the *(φ, Ω, M)-extremal point* [8] (in terms of mappings). The corresponding definitions of weak stationarity are introduced in Sections 3, 4.

†School of Information Technology and Mathematical Sciences, Centre of Information and Applied Optimization, University of Ballarat, POB 663, Ballarat, Vic, 3350, E–Mail: a.kruger@ballarat.edu.au

In Section 5 the definition of weak stationarity is being made specific for the case of a scalar function. This produces the notion of weak inf-stationarity which turns out equivalent to the corresponding stationarity notion introduced by B. Kummer [16]. In this case the definition of weak stationarity can be reformulated with the help of a *slope* [2] (see also [7]). Some illustrated examples of weakly inf-stationary functions are presented in Section 6.

The main subdifferential constructions which are used in statements of extremality conditions are recalled in Section 2. Comparison of different constructions and description of the potential of their use in analysis and optimization can be found in the survey paper [14].

Mainly standard notations are used throughout the paper. Symbols X^* and Y^* denote spaces topologically dual to X and Y respectively, and $\langle\cdot,\cdot\rangle$ denote the bilinear form defining duality between a space and its dual. The ball of radios ρ centered at x is denoted $B_\rho(x)$. We write B_ρ if $x=0$, and simply B if $x=0$, $\rho=1$. A unit ball in the dual space is denoted B^*.

2 Strict subdifferentials, normals, coderivatives

Let X be a real normed space. Consider a function $\varphi : X \to \bar{\mathbb{R}} = \mathbb{R}\cup\{+\infty\}$, finite at x°. The set

$$\partial\varphi(x^\circ) = \left\{x^* \in X^* :\ \liminf_{x\to x^\circ} \frac{\varphi(x)-\varphi(x^\circ)-\langle x^*, x-x^\circ\rangle}{\|x-x^\circ\|} \geq 0\right\} \tag{1}$$

is called the *Fréchet subdifferential* of φ at x°.

The convex set (1) is a natural generalization of the Fréchet derivative and the subdifferential of convex analysis. Fréchet subdifferentials possess comparatively poor calculus and are not widely used in nonsmooth analysis and optimization. But it is possible to define on their basis more powerful analysis tools: the *strict δ-subdifferential* and the *limiting subdifferential*.

The *strict δ-subdifferential* ($\delta>0$) of φ at x° is defined by the formula

$$\hat\partial_\delta\varphi(x^\circ) \ = \ \bigcup\left\{\partial(\mathrm{cl}^{\downarrow}\varphi)(x) :\ x\in B_\delta(x^\circ),\ \left|\mathrm{cl}^{\downarrow}\varphi(x)-\varphi(x^\circ)\right|\leq\delta\right\}. \tag{2}$$

The symbol $\mathrm{cl}^{\downarrow}\varphi$ denotes here a lower semicontinuous envelope of φ: $\mathrm{cl}^{\downarrow}\varphi(x) = \liminf_{u\to x}\varphi(u)$. Of course, definition (2) becomes simpler if φ is lower semicontinuous near x°.

The set (2) is nonconvex in general. It accumulates the information about "differential" properties of φ near x° (it would be more precise to speak of points of the graph of $\mathrm{cl}^{\downarrow}\varphi$ near $(x^\circ,\varphi(x^\circ))$) and generalizes the notion of strict derivative. Let us recall that φ is called *strictly differentiable* [1] at x° (with the derivative $\nabla\varphi(x^\circ)$) if

$$\lim_{x\to x^\circ,\ x'\to x^\circ} \frac{\varphi(x')-\varphi(x)-\langle\nabla\varphi(x^\circ), x'-x\rangle}{\|x'-x\|} = 0.$$

Proposition 1. *If φ is strictly differentiable at x° with the derivative $\nabla\varphi(x^\circ)$ then for any $\varepsilon>0$ there exist $\delta>0$ such that $\nabla\varphi(x^\circ)\subset\hat\partial_\delta\varphi(x^\circ)\subset\nabla\varphi(x^\circ)+\varepsilon B^*$.*

Proposition 2. *If φ is convex and $\partial\varphi(x^\circ)\neq\emptyset$ then for any $\varepsilon>0$ there exist $\delta>0$ such that $\partial\varphi(x^\circ)\subset\hat\partial_\delta\varphi(x^\circ)\subset\partial\varphi(x^\circ)+\varepsilon B^*$.*

The *limiting subdifferential* of φ at x° can be defined by the formula

$$\bar\partial\varphi(x^\circ) \ = \ \bigcap_{\delta>0}\mathrm{cl}^*\hat\partial_\delta\varphi(x^\circ), \tag{3}$$

where the symbol cl^* denotes weak* sequential closure of a set (a collection of the limits of all weakly* convergent sequences of elements of this set) in the dual space. In other words, $x^*\in\bar\partial\varphi(x^0)$ $\Leftrightarrow$ there exist sequences $\{x_k\}\subset X$, $\{x_k^*\}\subset X^*$ such that $x_k^*\in\partial\varphi(x_k)$, $k=1,2,\ldots$, and $x_k\to x^\circ$, $\varphi(x_k)\to\varphi(x^\circ)$, $x_k^*\stackrel{w^*}{\to}x^*$ when $k\to\infty$. In case of a strictly differentiable function φ (3) reduces to the strict derivative.

Remark 1. Simple examples show that the Fréchet subdifferential (1) can be empty. On the other hand, it follows from [6] that in the case of an *Asplund* space $\hat\partial_\delta\varphi(x^\circ)\neq\emptyset$ for any lower semicontinuous function φ and any $\delta>0$. If X is not Asplund the latter statement is not true. The limiting subdifferential (3) can also be empty even in case of a continuous function on a finite dimensional space. However, in finite dimensions it is possible to substitute completely the strict

δ-subdifferentials by the limiting constructions if to consider besides the limiting subdifferential (3) the so called *singular limiting subdifferential.* Unfortunately, in the infinite-dimensional case to use the limiting subdifferentials effectively one needs additional assumptions guaranteing nontriviality of the limits in the weak* topology (see. [18]).

Remark 2. Changing in (1) the inequality sign for the opposite one it is possible to define the *Fréchet superdifferential*, and on its basis also the *strict* δ- and the *limiting superdifferential.* Of course, the Fréchet sub- and superdifferential at some point can be nonempty simultaneously if and only if the function is Fréchet differentiable at the point. Strict and limiting sub- and superdifferentials can be nonempty simultaneously even for nonsmooth functions, and they can also be significantly different.

Remark 3. As it follows from Definition 1, the Fréchet subdifferentials and defined on their basis strict δ-subdifferentials are convenient for investigating differential properties of functions on an Asplund space. In an arbitrary Banach space instead of (1) one must use the ε*-subdifferential* ($\varepsilon > 0$)

$$\partial_\varepsilon\varphi(x^\circ) = \Big\{x^* \in X^* : \liminf_{x\to x^\circ} \frac{\varphi(x) - \varphi(x^\circ) - \langle x^*, x - x^\circ\rangle}{\|x - x^\circ\|} \ge -\varepsilon\Big\},$$

and instead of (2) the strict (ε, δ)-subdifferential

$$\hat{\partial}_{\varepsilon,\delta}\varphi(x^\circ) \;=\; \bigcup \left\{\partial_\varepsilon(\mathrm{cl}^{\downarrow}\varphi)(x) :\; x \in B_\delta(x^\circ),\; \left|\mathrm{cl}^{\downarrow}\varphi(x) - \varphi(x^\circ)\right| \le \delta\right\}$$

must be used.

By analogy with (1)–(3) some geometrical objects, namely the generalized normals to a set can be defined.

Let Ω be a set in a normed space X, $x^\circ \in \Omega$. The sets

$$N(x^\circ|\,\Omega) = \left\{x^* \in X^* : \limsup_{x\xrightarrow{\Omega}x^\circ} \frac{\langle x^*, x - x^\circ\rangle}{\|x - x^\circ\|} \le 0\right\}, \tag{4}$$

$$\hat{N}_\delta(x^\circ|\,\Omega) = \bigcup\left\{N(x|\,\mathrm{cl}\,\Omega) :\; x \in \mathrm{cl}\,\Omega \cap B_\delta(x^\circ)\right\}, \tag{5}$$

$$\bar{N}(x^\circ|\,\Omega) \;=\; \bigcap_{\delta>0} \mathrm{cl}^{\,*}\hat{N}_\delta(x^\circ|\,\Omega) \tag{6}$$

are called correspondingly the *Fréchet normal cone*, the *strict* δ*-normal cone* and the *limiting normal cone* to Ω at x°. The denotation $x \xrightarrow{\Omega} x^\circ$ in (4) means that $x \to x^\circ$ with $x \in \Omega$.

Of course, all the three sets (4)–(6) are cones. The cones (5), (6) can be nonconvex. The sets (4)–(6) coincide with the corresponding subdifferentials of the indicator function δ_Ω of Ω, which is defined as follows: $\delta_\Omega(x) = 0$ if $x \in \Omega$, and $\delta_\Omega(x) = +\infty$ otherwise. On the other hand, the subdifferentials (1)–(3) can be defined through the normal cones (4)–(6) to the epigraph $\mathrm{epi}\,\varphi = \{(x, \mu) \in X \times \mathbb{R} :\; \varphi(x) \le \mu\}$ of φ at $(x, \varphi(x))$.

In case of a convex set Ω the cones (4), (6) coincide with the normal cone in the sense of convex analysis, and for the cone (5) a statement similar to Proposition 2 holds true.

With the help of the normal cones (4)–(6) it is easy to define coderivatives for set-valued mappings.

Let $F :\; X \rightrightarrows Y$ be a set-valued mapping (multifunction) between normed spaces, $(x^\circ, y^\circ) \in \mathrm{gph}\,F$, where $\mathrm{gph}\,F = \{(x, y) \in X \times Y :\; y \in F(x)\}$ is a graph of F. We shall assume that the norm in $X \times Y$ agrees with the norms in X and Y, for example, $\|x, y\| = \max(\|x\|, \|y\|)$, $x \in X$, $y \in Y$.

The role of the derivative for F can be played by the set-valued mappings which for any $y^* \in Y^*$ are defined by the following relations:

$$\partial F(x^\circ, y^\circ; y^*) = \{x^* \in X^* : (x^*, -y^*) \in N(x^\circ, y^\circ|\,\mathrm{gph}\,F)\}, \tag{7}$$

$$\hat{\partial}_\delta F(x^\circ, y^\circ; y^*) = \{x^* \in X^* : (x^*, -y^*) \in \hat{N}_\delta(x^\circ, y^\circ|\,\mathrm{gph}\,F)\}, \tag{8}$$

$$\bar{\partial} F(x^\circ, y^\circ; y^*) = \{x^* \in X^* : (x^*, -y^*) \in \bar{N}(x^\circ, y^\circ|\,\mathrm{gph}\,F)\}. \tag{9}$$

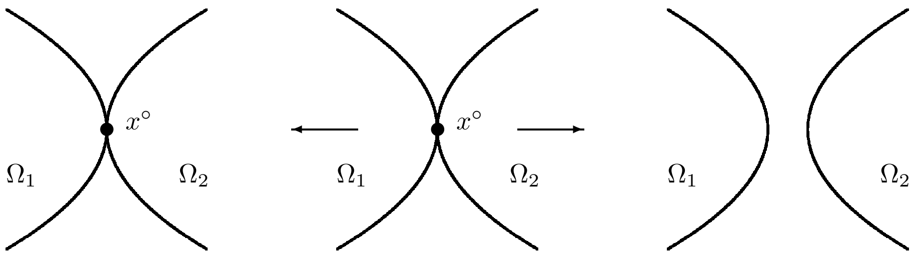
The equalities (7)–(9) define correspondingly the *Fréchet coderivative*, the *strict* δ*-coderivative* and the *limiting coderivative* of F at (x°, y°).

In the case $Y = \mathbb{R}$, $F(x) = \varphi(x) + \mathbb{R}_+$, $y^\circ = \varphi(x^\circ)$, $y^* = 1$ the sets (7)–(9) reduce to the subdifferentials (1)–(3).

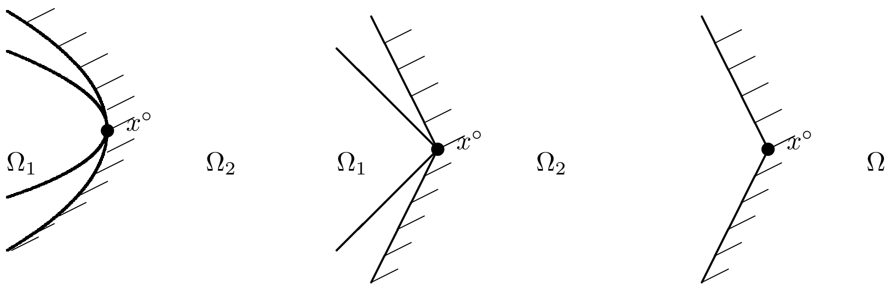
3 Weak stationarity of sets systems

In this section we consider a system of sets Ω_1, Ω_2, …, Ω_n $(n > 1)$ in a normed space X, having a common point $x^\circ \in \cap_{i=1}^n \Omega_i$ and we are interested in investigating "extremal" properties of the system.

The initial definition of the *extremal system* [15] was geometrical: all the sets have a common point and an arbitrarily small shift of the sets makes them unintersecting.



The sets are not assumed convex. Below are three more illustrated examples of the extremal systems of two sets.



In the last example the extremal system consists of the set Ω and its boundary point $\{x^\circ\}$.

The following constants can be used for characterizing the mutual arrangement of sets Ω_1, Ω_2, …, Ω_n near x°:

$$\theta_\rho[\Omega_1, \ldots, \Omega_n](x^\circ) = \sup \left\{ r \geq 0 : \left(\bigcap_{i=1}^n (\Omega_i - a_i) \right) \bigcap B_\rho(x^\circ) \neq \emptyset, \forall a_i \in B_r \right\}, \tag{10}$$

$$\tilde{\theta}[\Omega_1, \ldots, \Omega_n](x^\circ) = \liminf_{\rho \to +0} \; \theta_\rho[\Omega_1, \ldots, \Omega_n](x^\circ)/\rho, \tag{11}$$

$$\hat{\theta}[\Omega_1, \ldots, \Omega_n](x^\circ) = \liminf_{\omega_i \xrightarrow{\Omega_i} x^\circ} \; \tilde{\theta}[\Omega_1 - \omega_1, \ldots, \Omega_n - \omega_n](0). \tag{12}$$

Evidently all the constants (10)–(12) are nonnegative, the function $\rho \to \theta_\rho[\Omega_1, \ldots, \Omega_n](x^\circ)$ is nondecreasing, $\tilde{\theta}[\Omega_1, \ldots, \Omega_n](x^\circ) \geq \hat{\theta}[\Omega_1, \ldots, \Omega_n](x^\circ)$, and $\theta_\rho[\Omega_1, \ldots, \Omega_n](x^\circ) > 0$ for any $\rho > 0$ if $\tilde{\theta}[\Omega_1, \ldots, \Omega_n](x^\circ) > 0$.

It follows from definition (10) that for any $i = 1, 2, \ldots, n$ and any nonnegative $r < \theta_\rho[\Omega_1, \ldots, \Omega_n]$ the following inclusion holds:

$$B_r(x^\circ) \subset \Omega_i + B_\rho,$$

which, in its turn, yields the following statement.

Proposition 3. *The following assertions are equivalent:*
(a) $\theta_\rho[\Omega_1, \ldots, \Omega_n](x^\circ) \to 0$ *as* $\rho \to 0$*;*
(b) $x^\circ \notin \cap_{i=1}^n \mathrm{int}\, \mathrm{cl}\, \Omega_i$.

Definition 1. The system of sets Ω_1, Ω_2, …, Ω_n is *locally extremal* at x° if $\theta_\rho[\Omega_1, \ldots, \Omega_n](x^\circ) = 0$ for some $\rho > 0$.

The condition formulated above means that an arbitrarily small shift makes the sets unintersecting in a neighborhood of x°: there exist a number $\rho > 0$ and sequences $\{a_{ik}\} \subset X$ tending to zero, such that $\cap_{i=1}^n (\Omega_i - a_{ik}) \cap B_\rho(x^\circ) = \emptyset$, $k = 1, 2, \ldots$. Thus Definition 1 is equivalent to the initial one introduced in [15]. It defines a general notion of extremality embedding different solution notions in optimization problems.

Remark 4. If to exclude from (10) mentioning the neighborhood $B_\rho(x^\circ)$ (this corresponds to the case $\rho = +\infty$), then global extremality of the sets system can be defined.

The following two definitions weaken step by step the requirements to the system of sets.

Definition 2. The system of sets $\Omega_1, \Omega_2, \ldots, \Omega_n$ is *stationary* at x° if $\tilde{\theta}[\Omega_1, \ldots, \Omega_n](x^\circ) = 0$.

Definition 3. The system of sets $\Omega_1, \Omega_2, \ldots, \Omega_n$ is *weakly stationary* at x° if $\hat{\theta}[\Omega_1, \ldots, \Omega_n](x^\circ) = 0$.

Definition 2 corresponds to the traditional notion of stationarity, and the condition formulated in Definition 3 means that arbitrarily close to x° there exist points whose properties are arbitrarily close to the stationarity property.

Combining (10)–(12), one can get the following representation for (12):

$$\hat{\theta}[\Omega_1, \ldots, \Omega_n](x^\circ) = \liminf_{\substack{\omega_i \xrightarrow{\Omega_i} x^\circ \\ \rho \to +0}} \sup \left\{ r \geq 0 : \left(\bigcap_{i=1}^{n} (\Omega_i - \omega_i - a_i) \right) \bigcap B_\rho \neq \emptyset, \forall a_i \in B_r \right\} / \rho. \tag{13}$$

The next assertion is an immediate consequence of the definitions.

Proposition 4. *$\hat{\theta}[\Omega_1, \ldots, \Omega_n](x^\circ) > 0$ if and only if there exist $\alpha > 0$ and $\delta > 0$ such that*

$$\left(\bigcap_{i=1}^{n} (\Omega_i - \omega_i - a_i) \right) \bigcap B_\rho \neq \emptyset \tag{14}$$

for any $\rho \in (0, \delta]$, $\omega_i \in \Omega_i \cap B_\delta(x^\circ)$, $a_i \in B_{\alpha\rho}$, $i = 1, 2, \ldots, n$.

In the case $n = 2$ the following constant can be used along with (10):

$$\theta'_\rho[\Omega_1, \Omega_2](x^\circ) = \sup \{ r \geq 0 : \; B_r \subset \Omega_1 \cap B_\rho(x^\circ) - \Omega_2 \cap B_\rho(x^\circ) \}. \tag{15}$$

The difference of sets in (15) is understood in the algebraic sense: $\Omega_1 - \Omega_2 = \{\omega_1 - \omega_2 : \omega_1 \in \Omega_1, \; \omega_2 \in \Omega_2\}$. Contrary to (10) $\theta'_\rho[\Omega_1, \Omega_2](x^\circ)$ always tends to zero when $\rho \to 0$.

Proposition 5. *$\theta'_\rho[\Omega_1, \Omega_2](x^\circ) \leq 2\rho$.*

The next statement establishes relations between (10) and (15).

Proposition 6. *(a) $\theta'_\rho[\Omega_1, \Omega_2](x^\circ) \leq 2\theta_{\rho + \theta'_\rho[\Omega_1, \Omega_2](x^\circ)/2}[\Omega_1, \Omega_2](x^\circ)$;*
(b) $\theta'_\rho[\Omega_1, \Omega_2](x^\circ) \leq 2\theta_{2\rho}[\Omega_1, \Omega_2](x^\circ)$;
(c) $2\theta_\rho[\Omega_1, \Omega_2](x^\circ) \leq \theta'_{\rho + \theta_\rho[\Omega_1, \Omega_2](x^\circ)}[\Omega_1, \Omega_2](x^\circ)$;
(d) $2\min(\theta_\rho[\Omega_1, \Omega_2](x^\circ), \rho) \leq \theta'_{2\rho}[\Omega_1, \Omega_2](x^\circ)$.

Proof. Let $0 \leq r < \theta'_\rho[\Omega_1, \Omega_2](x^\circ)$ and $a_1, a_2 \in B_{r/2}$. Then $a_1 - a_2 \in B_r$ and in view of (15) the following inclusion holds true:

$$a_1 - a_2 \in \Omega_1 \cap B_\rho(x^\circ) - \Omega_2 \cap B_\rho(x^\circ),$$

which is equivalent to the condition

$$(\Omega_1 \cap B_\rho(x^\circ) - a_1) \cap (\Omega_2 \cap B_\rho(x^\circ) - a_2) \neq \emptyset,$$

which, in its turn, implies the condition

$$(\Omega_1 - a_1) \cap (\Omega_2 - a_2) \cap B_{\rho + r/2}(x^\circ) \neq \emptyset.$$

Due to the arbitrariness of $a_1, a_2 \in B_{r/2}$ it yields the inequality $\theta_{\rho + r/2}[\Omega_1, \Omega_2](x^\circ) \geq r/2$, which due to arbitrariness of $r < \theta'_\rho[\Omega_1, \Omega_2](x^\circ)$ and monotonicity of $\rho \to \theta_\rho[\Omega_1, \Omega_2](x^\circ)$ yields (a).

(b) follows from (a) in view of Proposition 5.

Let $0 \leq r < \theta_\rho[\Omega_1, \Omega_2](x^\circ)$ and $a \in B_{2r}$. Then owing to (10) the condition

$$(\Omega_1 - a/2) \cap (\Omega_2 + a/2) \cap B_\rho(x^\circ) \neq \emptyset,$$

holds true, which implies the inclusion

$$a \in \Omega_1 \cap B_{\rho + r}(x^\circ) - \Omega_2 \cap B_{\rho + r}(x^\circ).$$

Due to arbitrariness of $a \in B_{2r}$ it implies the estimate $\theta'_{\rho + r}[\Omega_1, \Omega_2](x^\circ) \geq 2r$, which due to arbitrariness of $r < \theta_\rho[\Omega_1, \Omega_2](x^\circ)$ and monotonicity of $\rho \to \theta'_\rho[\Omega_1, \Omega_2](x^\circ)$ yields (c) and (d). □

It follows from Proposition 6 (conditions (b) (d)) that in the case $n = 2$ replacement in (11) of the constant (10) by the constant (15) does not imply changes in Definitions 2, 3. Thus when $n = 2$ Definition 3 is equivalent to the corresponding definition from [10] (see also [13], [14]). Instead of "weak stationarity" the term "*extended extremality*" was used earlier.

On the other hand, the general case $n \geq 2$ is easily reduced to the case of two sets.

Proposition 7. *The system of sets $\Omega_1, \Omega_2, \ldots, \Omega_n$ is weakly stationary at x° if and only if the system of sets $\tilde{\Omega}_1 = \Omega_1 \times \Omega_2 \times \ldots \times \Omega_n$, $\tilde{\Omega}_2 = \{(x, x, \ldots, x) : x \in X\}$ in X^n is weakly stationary at $(x^\circ, x^\circ, \ldots, x^\circ)$.*

Let us define one more constant for $\Omega_1, \Omega_2, \ldots, \Omega_n$, this time with the help of dual space elements:

$$\eta[\Omega_1, \ldots, \Omega_n](x^\circ) = \liminf_{\delta \to 0} \left\{ \left\| \sum_{i=1}^n x_i^* \right\| : \ x_i^* \in \hat{N}_\delta(x^\circ | \Omega_i), i = 1, \ldots, n, \ \sum_{i=1}^n \|x_i^*\| = 1 \right\}. \quad (16)$$

(16) is convenient for formulating dual criteria of weak stationarity of sets systems. It can be used to reformulate the *generalized Euler equation* [8, 10, 15, 13].

Definition 4. The system of sets Ω_1, Ω_2,$\ldots$,Ω_n satisfies the generalized Euler equation at x° if $\eta[\Omega_1, \ldots, \Omega_n](x^\circ) = 0$.

Proposition 8. *$\eta[\Omega_1, \ldots, \Omega_n](x^\circ) < \varepsilon$ if and only if for any $\delta > 0$ there exist points $\omega_i \in \Omega_i \cap B_\delta(x^\circ)$, $x_i^* \in N(\omega_i | \Omega_i)$, $i = 1, 2, \ldots, n$, such that $\sum_{i=1}^n \|x_i^*\| = 1$, $\|\sum_{i=1}^n x_i^*\| < \varepsilon$.*

As it can be seen from the next theorem the dual criterion of weak stationarity becomes exact in the case of an Asplund space. Let us recall that a Banach space is called Asplund (see [19, 20]) if any continuous convex function on it is Fréchet differentiable on a dense set of points. Asplund spaces form a rather broad subclass of Banach spaces. It contains e.g. all reflexive spaces and all spaces that admit equivalent norms, Fréchet differentiable at all nonnull points.

Theorem 1. *(a) The following inequality is true:*

$$\hat{\theta}[\Omega_1, \ldots, \Omega_n](x^\circ) \leq \eta[\Omega_1, \ldots, \Omega_n](x^\circ). \quad (17)$$

(b) If X is Asplund, the sets $\Omega_1, \Omega_2, \ldots, \Omega_n$ are closed and $\hat{\theta}[\Omega_1, \ldots, \Omega_n](x^\circ) < 1$ then

$$\eta[\Omega_1, \ldots, \Omega_n](x^\circ) \leq \frac{\hat{\theta}[\Omega_1, \ldots, \Omega_n](x^\circ)}{1 - \hat{\theta}[\Omega_1, \ldots, \Omega_n](x^\circ)}. \quad (18)$$

Proof. (a) If $\hat{\theta}[\Omega_1, \ldots, \Omega_n](x^\circ) = 0$ then (17) holds trivially.

Let $\hat{\theta}[\Omega_1, \ldots, \Omega_n](x^\circ) > \alpha > 0$, $\eta[\Omega_1, \ldots, \Omega_n](x^\circ) < \varepsilon$. We shall show that $\alpha < \varepsilon$. In view of Propositions 4, 8 there exist a number $\delta > 0$ and points $\omega_i \in \Omega_i \cap B_\delta(x^\circ)$, $x_i^* \in N(\omega_i | \Omega_i)$, $i = 1, 2, \ldots, n$, such that $\sum_{i=1}^n \|x_i^*\| = 1$, $\|\sum_{i=1}^n x_i^*\| < \varepsilon$ and for any $\rho \in (0, \delta]$, $a_i \in B_{\alpha\rho}$, $i = 1, 2, \ldots, n$, condition (14) holds true.

Let us denote $\varepsilon' = (\varepsilon - \|\sum_{i=1}^n x_i^*\|)/(2n)$. It follows from definition (4) of the normal cone that for sufficiently small ρ for all $\omega \in \Omega_i \cap B_{(\alpha+1)\rho}(\omega_i)$ the inequalities $\langle x_i^*, \omega - \omega_i \rangle \leq \frac{\varepsilon'}{\alpha+1} \|\omega - \omega_i\| \leq \varepsilon' \rho$ holds true. Besides, for any $i = 1, 2, \ldots, n$ it is possible to find a point $a_i \in B_{\alpha\rho}$ such that $\langle x_i^*, a_i \rangle > \alpha\rho \|x_i^*\| - \varepsilon' \rho$. Let us make use of condition (14) now. Let $x \in (\cap_{i=1}^n (\Omega_i - \omega_i - a_i)) \cap B_\rho$. For any $i = 1, 2, \ldots, n$ we have $x = \omega_i' - \omega_i - a_i$ for some $\omega_i' \in \Omega_i$ and $\|\omega' - \omega_i\| = \|x + a_i\| \leq (\alpha+1)\rho$. Thus $\langle x_i^*, x \rangle < -\alpha\rho \|x_i^*\| + 2\varepsilon' \rho$ and consequently $\sum_{i=1}^n \langle x_i^*, x \rangle < -\alpha\rho + 2n\varepsilon' \rho$. On the other hand, $\langle \sum_{i=1}^n x_i^*, x \rangle > -\varepsilon\rho$. Comparing the last two inequalities, one can conclude that $\alpha < \varepsilon$. Due to the arbitrariness of α and ε inequality (17) is proved.

(b) When proving inequality (18) we use essentially two fundamental results of variational analysis: *Ekeland variational principle* [4] and established by M. Fabian representation for elements of the subdifferential of the sum of functions [5]. To apply the first result it is necessary to assume that X is a Banach space and $\Omega_1, \Omega_2, \ldots, \Omega_n$ are closed. To make use of the second result we must assume additionally that X is Asplund.

Let $\hat{\theta}[\Omega_1, \ldots, \Omega_n](x^\circ) < \alpha < 1$. We shall show that $\eta[\Omega_1, \ldots, \Omega_n](x^\circ) < \alpha/(1 - \alpha)$. Let us take an arbitrary $\beta \in (\hat{\theta}[\Omega_1, \ldots, \Omega_n](x^\circ), \alpha)$. For any $\delta > 0$ there exist a number $\rho \in (0, \delta/2)$ and points $\omega_i \in \Omega_i \cap B_{\delta/2}(x^\circ)$, $a_i \in B_{\beta\rho}$, $i = 1, 2, \ldots, n$, such that $(\cap_{i=1}^n (\Omega_i - \omega_i - a_i)) \cap B_\rho = \emptyset$. Let us consider in the Asplund space X^n (with the norm $\|x_1, x_2, \ldots, x_n\| = \max(\|x_1\|, \|x_2\|, \ldots, \|x_n\|)$) two closed sets $\bar{\Omega}_1 = \Omega_1 \times \Omega_2 \times \ldots \times \Omega_n$ and $\bar{\Omega}_2 = \{(x, x, \ldots, x) : x \in B_\rho\}$ and a continuous function φ_1 from $X^n \times X^n$ into $\mathbb{R}$: $\varphi_1(\bar{x}_1, \bar{x}_2) = \|\bar{x}_1 - \bar{\omega} - \bar{a} - \bar{x}_2\|$, where $\bar{\omega} = (\omega_1, \omega_2, \ldots, \omega_n)$, $\bar{a} = (a_1, a_2, \ldots, a_n)$. Evidently $\varphi_1(\bar{x}_1, \bar{x}_2) > 0$ for all $(\bar{x}_1, \bar{x}_2)$ from $\bar{\Omega}_1 \times \bar{\Omega}_2$ and $\varphi_1(\bar{\omega}, 0) = \|\bar{a}\| \leq \beta\rho$.

The space $X^n \times X^n$ with the norm $\|\bar{x}_1, \bar{x}_2\| = \max(\|\bar{x}_1\|, \|\bar{x}_2\|)$ is also Asplund. Let us take now an arbitrary $\gamma \in (\beta, \alpha)$. The Ekeland theorem [4] guarantees existence of a point $(\bar{\omega}', \bar{x}') \in (\bar{\Omega}_1 \times \bar{\Omega}_2) \cap B_{\beta\rho/\gamma}(\bar{\omega}, 0)$ minimizing the function $\varphi_1 + \varphi_2 + \varphi_3$, where $\varphi_2(\bar{x}_1, \bar{x}_2) = \gamma \|(\bar{x}_1, \bar{x}_2) - (\bar{\omega}', \bar{x}')\|$, $\varphi_3(\bar{x}_1, \bar{x}_2) = 0$ if $(\bar{x}_1, \bar{x}_2) \in \bar{\Omega}_1 \times \bar{\Omega}_2$ and $\varphi_3(\bar{x}_1, \bar{x}_2) = \infty$ otherwise (the indicator function of the set $\bar{\Omega}_1 \times \bar{\Omega}_2$). Evidently $\|(\bar{\omega}', \bar{x}') - (\bar{\omega}, 0)\| \leq \beta\rho/\gamma < \rho$. Let us denote $\varrho = \rho - \|(\bar{\omega}', \bar{x}') - (\bar{\omega}, 0)\|$.

The functions φ_1 and φ_2 are Lipschitz continuous, and one can apply Lemma 4 from [5] (*"zero" fuzzy sum rule*). There exist points $(\bar{x}_{1j}, \bar{x}_{2j})$, $(\bar{x}^*_{1j}, \bar{x}^*_{2j}) \in \partial\varphi_j(\bar{x}_{1j}, \bar{x}_{2j})$, $j = 1, 2, 3$, such that $\|(\bar{x}_{1j}, \bar{x}_{2j}) - (\bar{\omega}', \bar{x}')\| < \varrho$, $\varphi_1(\bar{x}_{11}, \bar{x}_{21}) > 0$, $(\bar{x}_{13}, \bar{x}_{23}) \in \bar{\Omega}_1 \times \bar{\Omega}_2$ and $\|(x^*_1, y^*_1) + (x^*_2, y^*_2) + (x^*_3, y^*_3)\| < \alpha - \gamma$. Evidently, $\|(\bar{x}_{13}, \bar{x}_{23}) - (\bar{\omega}, 0)\| < \rho$. Let $\bar{x}_{13} = (w_1, w_2, \ldots, w_n)$, $w_i \in \Omega_i$, $i = 1, 2, \ldots, n$, $\bar{x}_{23} = (y, y, \ldots, y)$. The last estimate guarantees that $\|w_i - x^\circ\| < \delta$ and y is an interior point of B_ρ.

The functions φ_1 and φ_2 are convex. Their Fréchet subdifferentials coincide with subdifferentials in the sense of convex analysis and can be easily calculated. The subdifferential of φ_3 reduces to the normal cone to $\bar{\Omega}_1 \times \bar{\Omega}_2$. When $j = 1$ one has $\bar{x}^*_{11} = \bar{x}^*$, $\bar{x}^*_{21} = -\bar{x}^*$, where $\bar{x}^*$ is a subgradient of the norm of the space X^n, calculated at a nonnull point. Thus, $\|\bar{x}^*\| = 1$. When $j = 2$ the estimate $\|\bar{x}^*_{12}, \bar{x}^*_{22}\| \leq \gamma$ is true. Finally, when $j = 3$ one has $\bar{x}^*_{13} = (x^*_{13}, x^*_{23}, \ldots, x^*_{n3})$, $x^*_{i3} \in N(w_i|\Omega_i)$, $i = 1, 2, \ldots, n$, $\bar{x}^*_{23} = (y^*_{13}, y^*_{23}, \ldots, y^*_{n3})$, $\sum_{i=1}^n y^*_{i3} = 0$. On the other hand, $\|\bar{x}^*_{13} + \bar{x}^*_{23}\| = \|(\bar{x}^*_{11} + \bar{x}^*_{12} + \bar{x}^*_{13}) + (\bar{x}^*_{21} + \bar{x}^*_{22} + \bar{x}^*_{23}) - \bar{x}^*_{12} - \bar{x}^*_{22}\| < \alpha$ and consequently $\|\sum_{i=1}^n x^*_{i3}\| = \|\sum_{i=1}^n (x^*_{i3} + y^*_{i3})\| \leq \|\bar{x}^*_{13} + \bar{x}^*_{23}\| < \alpha$. At the same time $\sum_{i=1}^n \|x^*_{i3}\| = \|\bar{x}^*_{13}\| \geq \|\bar{x}^*_{11}\| - \|\bar{x}^*_{11} + \bar{x}^*_{12} + \bar{x}^*_{13}\| - \|\bar{x}^*_{12}\| > 1 - \alpha > 0$. Let us denote $x^*_i = x^*_{i3}/\|\bar{x}^*_{13}\|$. Evidently, $x^*_i \in N(w_i|\Omega_i)$, $i = 1, 2, \ldots, n$, $\|\sum_{i=1}^n x^*_i\| < \alpha/(1-\alpha)$, $\sum_{i=1}^n \|x^*_i\| = 1$. Consequently, $\eta[\Omega_1, \ldots, \Omega_n](x^\circ) < \alpha/(1-\alpha)$. In view of arbitrariness of α inequality (18) is proved. □

Corollary 1.1. *In an Asplund space the Extended extremal principle holds true: a system of closed sets is weakly stationary at some point if and only if this system satisfies the generalized Euler equation at this point.*

Remark 5. Since a locally extremal sets system is weakly stationary, it follows from [17] that asplundity of the space is not only sufficient for validity of the Extended extremal principle but also necessary.

4 Weak stationarity of set-valued mappings

Let us start with considering first a set-valued mapping $F: X \rightrightarrows Y$ between normed spaces X and Y with a graph $\text{gph}\, F = \{(x, y) \in X \times Y : y \in F(x)\}$ and fix a point $(x^\circ, y^\circ) \in \text{gph}\, F$. We will assume that the product space $X \times Y$ is equipped with the maximum-type norm: $\|x, y\| = \max(\|x\|, \|y\|)$.

Similarly to (10)–(12) the following three constants can be defined for characterizing the local behavior of F near (x°, y°):

$$\theta_\rho[F](x^\circ, y^\circ) = \sup\{r \geq 0 : B_r(y^\circ) \subset F(B_\rho(x^\circ))\}, \tag{19}$$

$$\tilde{\theta}[F](x^\circ, y^\circ) = \liminf_{\rho \to +0} \frac{\theta_\rho[F](x^\circ, y^\circ)}{\rho}, \tag{20}$$

$$\hat{\theta}[F](x^\circ, y^\circ) = \liminf_{(x,y) \overset{\text{gph}\, F}{\to} (x^\circ, y^\circ)} \tilde{\theta}[F](x, y). \tag{21}$$

All the constants (19)–(21) are nonnegative, the function $\rho \to \theta_\rho[F](x^\circ, y^\circ)$ is nondecreasing, $\tilde{\theta}[F](x^\circ, y^\circ) \geq \hat{\theta}[F](x^\circ, y^\circ)$, and $\theta_\rho[F](x^\circ, y^\circ) > 0$ for any $\rho > 0$ if $\tilde{\theta}[F](x^\circ, y^\circ) > 0$.

Proposition 9. *The following assertions are equivalent:*
(a) $\theta_\rho[F](x^\circ, y^\circ) \to 0$ *as* $\rho \to 0$;
(b) $y^\circ \notin \text{int} \cap_{\rho > 0} F(B_\rho(x^\circ))$.

The three constants (19)–(21) give rise to the three definitions.

Definition 5. F is *locally extremal* at (x°, y°) if $\theta_\rho[F](x^\circ, y^\circ) = 0$ for some $\rho > 0$.

The condition formulated above means that the image of a ball centered at x° does not contain any ball centered at y°.

Definition 6. F is *stationary* at (x°, y°) if $\tilde{\theta}[F](x^\circ, y^\circ) = 0$.

Definition 7. F is *weakly stationary* at (x°, y°) if $\hat{\theta}[F](x^\circ, y^\circ) = 0$.

Combining (19)–(21), one can get the following representation for (21):

$$\hat{\theta}[F](x^\circ, y^\circ) = \liminf_{\substack{(x,y) \overset{\mathrm{gph}\, F}{\to} (x^\circ, y^\circ) \\ \rho \to +0}} \sup \{r \geq 0 : B_r(y) \subset F(B_\rho(x))\} / \rho. \tag{22}$$

The next proposition shows that weak stationarity is equivalent to the absence of the *covering* [3] property.

Proposition 10. *$\hat{\theta}[F](x^\circ, y^\circ) > 0$ if and only if there exist $\alpha > 0$ and $\delta > 0$ such that*

$$B_{\alpha\rho}(y) \subset F(B_\rho(x)) \tag{23}$$

for any $\rho \in (0, \delta]$, $(x, y) \in \mathrm{gph}\, F \cap B_\delta(x^\circ, y^\circ)$.

The dual counterpart of (22) can be defined as follows:

$$\eta[F](x^\circ, y^\circ) = \lim_{\delta \to 0} \inf \big\{ \|x^*\| : x^* \in \hat{\partial}_\delta F(x^\circ, y^\circ; y^*),\ \|y^*\| = 1 \big\}. \tag{24}$$

The element $y^* \in Y^*$ in (24) can be seen as some analog of the Lagrange multipliers vector.

Definition 8. The *generalized Lagrange multipliers rule* holds for F at (x°, y°) if $\eta[F](x^\circ, y^\circ) = 0$.

Proposition 11. *$\eta[F](x^\circ, y^\circ) < \varepsilon$ if and only if for any $\delta > 0$ there exist points $(x, y) \in \mathrm{gph}\, F \cap B_\delta(x^\circ, y^\circ)$, $(x^*, y^*) \in N(x, y|\mathrm{gph}\, F)$, such that $\|y^*\| = 1$, $\|x^*\| < \varepsilon$.*

Theorem 2. *(a) The following inequality is true:*

$$\hat{\theta}[F](x^\circ, y^\circ) \leq \eta[F](x^\circ, y^\circ). \tag{25}$$

(b) If X is Asplund, $\mathrm{gph}\, F$ is closed and $\hat{\theta}[F](x^\circ, y^\circ) < 1$ then

$$\eta[F](x^\circ, y^\circ) \leq \frac{\hat{\theta}[F](x^\circ, y^\circ)}{1 - \hat{\theta}[F](x^\circ, y^\circ)}. \tag{26}$$

The proof of the theorem is very similar to the one of Theorem 1.

Proof. (a) If $\hat{\theta}[F](x^\circ, y^\circ) = 0$ then (25) holds trivially.

Let $\hat{\theta}[F](x^\circ, y^\circ) > \alpha > 0$, $\eta[F](x^\circ, y^\circ) < \varepsilon$. We shall show that $\alpha < \varepsilon$. In view of Propositions 10, 11 there exist a number $\delta > 0$ and points $(x, y) \in \mathrm{gph}\, F \cap B_\delta(x^\circ, y^\circ)$, $(x^*, y^*) \in N(x, y|\mathrm{gph}\, F)$, such that $\|y^*\| = 1$, $\|x^*\| < \varepsilon$ and for any $\rho \in (0, \delta]$ condition (23) holds true.

Let us denote $\alpha' = \max(\alpha, 1)$, $\varepsilon' = (\varepsilon - \|x^*\|)/2$. It follows from definition (4) of the normal cone that for sufficiently small ρ for all $(u, v) \in \mathrm{gph}\, F \cap B_{\alpha'\rho}(x, y)$ the inequalities $\langle (x^*, y^*), (u, v) - (x, y) \rangle \leq (\varepsilon'/\alpha') \|(u, v) - (x, y)\| \leq \varepsilon'\rho$ hold true. Besides, it is possible to find a point $a \in B_{\alpha\rho}$ such that $\langle y^*, a \rangle > \alpha\rho - \varepsilon'\rho$. Let us denote $y' = y + a$ and make use of condition (23): there exists $x' \in B_\rho(x)$ such that $y' \in F(x')$. Thus $(x', y') \in \mathrm{gph}\, F \cap B_{\alpha'\rho}(x, y)$ and consequently $\langle (x^*, y^*), (x', y') - (x, y) \rangle \leq \varepsilon'\rho$. On the other hand, $\langle (x^*, y^*), (x', y') - (x, y) \rangle > -\|x^*\|\rho + \alpha\rho - \varepsilon'\rho$. Comparing the last two inequalities, one can conclude that $\alpha < \|x^*\| + 2\varepsilon' = \varepsilon$. Due to the arbitrariness of α and ε inequality (25) is proved.

(b) Let $\hat{\theta}[F](x^\circ, y^\circ) < \alpha < 1$. We shall show that $\eta[F](x^\circ, y^\circ) < \alpha/(1 - \alpha)$. Let us take an arbitrary $\beta \in (\hat{\theta}[F](x^\circ, y^\circ), \alpha)$. For any $\delta > 0$ there exist a number $\rho \in (0, \delta/2)$ and points $(x, y) \in \mathrm{gph}\, F \cap B_{\delta/2}(x^\circ, y^\circ)$, $\tilde{y} \in B_{\beta\rho}(y)$, such that $F^{-1}(\tilde{y}) \cap B_\rho(x) = \emptyset$. Let us consider a continuous function $\varphi_1 : X \times Y \to \mathbb{R}$ given by $\varphi_1(u, v) = \|v - \tilde{y}\|$. Evidently $\varphi_1(u, v) > 0$ for all $(u, v) \in \mathrm{gph}\, F$ such that $u \in B_\rho(x)$, and $\varphi_1(x, y) \leq \beta\rho$.

Let us take now an arbitrary $\gamma \in (\beta, \alpha)$. The Ekeland variational principle guarantees existence of a point $(x', y') \in \mathrm{gph}\, F \cap B_{\beta\rho/\gamma}(x, y)$ minimizing the function $\varphi_1 + \varphi_2 + \varphi_3$, where $\varphi_2(u, v) =$

$\gamma\,\|(u,v)-(x',y')\|$, $\varphi_3(u,v)=0$ if $(u,v)\in \operatorname{gph} F$ and $\varphi_3(u,v)=\infty$ otherwise (the indicator function of the set $\operatorname{gph} F$). Evidently $\|(x',y')-(x,y)\|\le \beta\rho/\gamma<\rho$. Let us denote $\varrho=\rho-\|(x',y')-(x,y)\|$.

The functions φ_1 and φ_2 are Lipschitz continuous, and one can apply Lemma 4 from [5] (*"zero" fuzzy sum rule*). There exist points (x_j,y_j), $(x_j^*,y_j^*)\in\partial\varphi_j(x_j,y_j)$, $j=1,2,3$, such that $\|(x_j,y_j)-(x',y')\|<\varrho$, $\varphi_1(x_1,y_1)>0$, $(x_3,y_3)\in\operatorname{gph} F$ and $\|(x_1^*,y_1^*)+(x_2^*,y_2^*)+(x_3^*,y_3^*)\|<\alpha-\gamma$. Evidently, $\|(x_3,y_3)-(x,y)\|<\rho$, in particular, x_3 is an interior point of B_ρ, and $(x_3,y_3)\in B_\delta(x^\circ,y^\circ)$.

The functions φ_1 and φ_2 are convex. Their Fréchet subdifferentials coincide with subdifferentials in the sense of convex analysis and can be easily calculated. The subdifferential of φ_3 reduces to the normal cone to $\operatorname{gph} F$. One has $x_1^*=0$, $\|y_1^*\|=1$, $\|x_2^*,y_2^*\|\le\gamma$, $(x_3^*,y_3^*)\in N(x_3,y_3|\operatorname{gph} F)$. On the other hand, $\|x_3^*\|=\|(x_1^*+x_2^*+x_3^*)-x_1^*-x_2^*\|<\alpha$, $\|y_3^*\|=\|y_1^*+y_2^*-(y_1^*+y_2^*+y_3^*)\|>1-\alpha$. Let us denote $x'^*=x_3^*/\|y_3^*\|$, $y'^*=y_3^*/\|y_3^*\|$. Evidently, $(x'^*,y'^*)\in N(x_3,y_3|\operatorname{gph} F)$, $\|x'^*\|<\alpha/(1-\alpha)$, $\|y'^*\|=1$. Consequently, $\eta[F](x^\circ,y^\circ)<\alpha/(1-\alpha)$. In view of arbitrariness of α inequality (26) is proved. □

Corollary 2.1. *In an Asplund space a closed-graph set-valued mapping is weakly stationary at some point if and only if the generalized Lagrange multipliers rule holds at this point.*

Let us consider a special case now, motivated by applications in optimization. Let the set-valued mapping F be defined by the triple $\{\Omega,M,f\}$ in the following way: $F(x)=f(x)-M$ if $x\in\Omega$ and $F(x)=\emptyset$ otherwise. Here Ω and M are subsets of X and Y correspondingly and f is a (single-valued) function from Ω to Y. We will assume that $x^\circ\in\Omega$, $f(x^\circ)\in M$. We will say that f is M-closed if $\operatorname{gph} F$ is closed in $X\times Y$.

The stationarity and other properties introduced above induce corresponding ones for the triple $\{\Omega,M,f\}$.

Definition 9. (a) $\{\Omega,M,f\}$ is *weakly stationary* at x° if F weakly stationary at $(x^\circ,0)$.

(b) The *generalized Lagrange multipliers rule* holds for $\{\Omega,M,f\}$ at x° if it holds for F at $(x^\circ,0)$.

In view of Definition 9 the next statement follows immediately from Corollary 2.1.

Corollary 2.2. *Let X, Y be Asplund, Ω, M be closed, and f be M-closed. $\{\Omega,M,f\}$ is weakly stationary at x° if and only if the generalized Lagrange multipliers rule holds for $\{\Omega,M,f\}$ at x°.*

When f is Lipschitz the generalized multipliers rule takes a more natural form: the following constant can be used instead of (24):

$$\eta[\Omega,M,f](x^\circ)=\lim_{\delta\to 0}\inf\bigl\{\|x^*\|:\ x^*\in\hat\partial_\delta\langle y^*,f\rangle(x^\circ),\ y^*\in\hat N_\delta(f(x^\circ)|M),\ \|y^*\|=1\bigr\}\tag{27}$$

Proposition 12. *Let f be Lipschitz continuous near x°. The generalized Lagrange multipliers rule holds for $\{\Omega,M,f\}$ at x° if and only if $\eta[\Omega,M,f](x^\circ)=0$.*

The term $\langle y^*,f\rangle$ in (27) can be treated as the Lagrangian function and the first inclusion in (27), combined with the condition $\eta[\Omega,M,f](x^\circ)=0$, can be considered as an analog of the traditional multipliers rule. The second inclusion in (27) generalizes conditions on the signs of multipliers and the complementarity slackness conditions.

Let us mention that the classical nonlinear programming problem

$$\begin{array}{ll}\text{minimize} & f_0(x)\\ \text{subject to} & f_i(x)\le 0,\ i=1,2,\dots,m,\\ & f_i(x)=0,\ i=m+1,\dots,n,\\ & x\in\Omega.\end{array}$$

can be plunged into the scheme described above if one assigns $f=(f_0-f_0(x),f_1,f_2,\dots,f_n):X\to\mathbb{R}^{n+1}$, $M=\mathbb{R}_-^{m+1}\times 0$.

5 Weak inf-stationarity

The case of a scalar function $\varphi: X\to\bar{\mathbb{R}}=\mathbb{R}\cup\{+\infty\}$ will be considered in this section. It is assumed to be finite at some point $x^\circ\in X$. We will relate to it the set-valued mapping $F:\ X\rightrightarrows\mathbb{R}$ defined by $F(x)=f(x)+\mathbb{R}_+=\{\mu\in\mathbb{R}:\ \mu\ge f(x)\}$ (the *epigraphical* mapping).

Of course, $y^\circ = \varphi(x^\circ) \in F(x^\circ)$ and constants (19)–(21) reduce (up to a sign) in this case to the following ones:

$$\theta_\rho[\varphi](x^\circ) = \inf_{x \in B_\rho(x^\circ)} \varphi(x) - \varphi(x^\circ), \tag{28}$$

$$\tilde{\theta}[\varphi](x^\circ) = \limsup_{\rho \to +0} \frac{\theta_\rho[\varphi](x^\circ)}{\rho}, \tag{29}$$

$$\hat{\theta}[\varphi](x^\circ) = \limsup_{x \xrightarrow{\varphi} x^\circ} \tilde{\theta}[\varphi](x). \tag{30}$$

The denotation $x \xrightarrow{\varphi} x^\circ$ in (30) means that $x \to x^\circ$ with $\varphi(x) \to \varphi(x^\circ)$.

The constants (28)–(30) are nonpositive, the function $\rho \to \theta_\rho[\varphi](x^\circ)$ is nonincreasing, $\tilde{\theta}[\varphi](x^\circ) \le \hat{\theta}[\varphi](x^\circ)$, and $\theta_\rho[\varphi](x^\circ) < 0$ for any $\rho > 0$ if $\tilde{\theta}[\varphi](x^\circ) < 0$.

Proposition 13. *The following assertions are equivalent:*
(a) $\theta_\rho[\varphi](x^\circ) \to 0$ as $\rho \to 0$;
(b) φ is lower semicontinuous at (x°).

From the point of view of optimization theory the "zero cases" of (28)–(30) are of interest. (28) corresponds to the traditional notion of local minimality, and (29), (30) lead to the corresponding stationarity notions.

Definition 10. φ is *inf-stationary* at x° if $\tilde{\theta}[\varphi](x^\circ) = 0$.

Definition 11. φ is *weakly inf-stationary* at x° if $\hat{\theta}[\varphi](x^\circ) = 0$.

The role of "inf" prefix in Definitions 10, 11 is to stress that minimization problems are addressed here[1]. *Sup-stationarity* can be defined in a similar way with obvious changes in (28)–(30).

Combining (28)–(30), one can get the following representation for (30):

$$\hat{\theta}[\varphi](x^\circ) = \limsup_{\substack{x \xrightarrow{\varphi} x^\circ \\ \rho \to +0}} \inf_{u \in B_\rho(x)} \frac{\varphi(u) - \varphi(x)}{\rho}. \tag{31}$$

The next proposition shows that the absence of weak stationarity means that all points in a neighborhood of a given point have descent directions with a uniform rate.

Proposition 14. *$\hat{\theta}[\varphi](x^\circ) < 0$ if and only if there exist $\alpha > 0$ and $\delta > 0$ such that for any $x \in B_\delta(x^\circ)$ with $|\varphi(x) - \varphi(x^\circ)| \le \delta$ and any $\rho \in (0, \delta]$ one can find $u \in B_\rho(x)$ such that $\varphi(u) - \varphi(x) < -\alpha\rho$.*

The dual constant (24) takes in the current setting a very simple form:

$$\eta[\varphi](x^\circ) = \lim_{\delta \to 0} \inf \left\{ \|x^*\| : x^* \in \hat{\partial}_\delta \varphi(x^\circ) \right\}. \tag{32}$$

It is not difficult to see that substituting $\Omega = X$, $M = \mathbb{R}_-$, $f = \varphi$ into (27) leads to the same constant (32) though φ is not assumed to be Lipschitz.

Definition 12. The *generalized Fermat condition* holds for φ at x° if $\eta[\varphi](x^\circ) = 0$.

Proposition 15. *$\eta[\varphi](x^\circ) < \varepsilon$ if and only if for any $\delta > 0$ there exist $x \in B_\delta(x^\circ)$ such that $|\varphi(x) - \varphi(x^\circ)| \le \delta$, and an element $x^* \in \partial\varphi(x)$ such that $\|x^*\| < \varepsilon$.*

Application of Corollary 2.1 brings us to the following result.

Proposition 16. *In an Asplund space a lower semicontinuous real valued function is weakly stationary at some point if and only if the generalized Fermat condition holds at this point.*

[1]The terminology was proposed by V. F. Demianov (personal communication)

Besides (28)–(30) the following pair of constants can be used for characterizing stationarity properties of φ at x°:

$$\tau[\varphi](x^\circ) = \liminf_{x \to x^\circ} \min\left(\frac{\varphi(x) - \varphi(x^\circ)}{\|x - x^\circ\|}, 0\right), \tag{33}$$

$$\hat{\tau}[\varphi](x^\circ) = \limsup_{x \xrightarrow{\varphi} x^\circ} \ \tau[\varphi](x). \tag{34}$$

Comparing (28), (29) and (33) one can easily see that $\tau[\varphi](x^\circ) \leq \tilde{\theta}[\varphi](x^\circ)$ which implies that $\hat{\tau}[\varphi](x^\circ) \leq \hat{\theta}[\varphi](x^\circ)$, and the last inequality can be strict. Fortunately, as it follows from the next proposition, constants (30) and (34) do coincide in the most important "zero case" in the Banach space setting.

Proposition 17. *Let X be a Banach space and φ be lower semicontinuous near x°. φ is weakly inf-stationary at x° if and only if $\hat{\tau}[\varphi](x^\circ) = 0$.*

The proof of Proposition 17 can be found in [14]. The necessary part was actually proved by B. Kummer[2] using Ekeland variational principle.

Remark 6. In view of (33), (34) the condition $\hat{\tau}[\varphi](x^\circ) = 0$ means that for any $\varepsilon > 0$ there exists a point $x_\varepsilon \in B_\varepsilon(x^\circ)$ such that $|\varphi(x_\varepsilon) - \varphi(x^\circ)| \leq \varepsilon$ and $\varphi(u) + \varepsilon \|u - x_\varepsilon\| \geq \varphi(x_\varepsilon)$ for all u near x_ε. Using the terminology adopted in [16], x_ε is a *local Ekeland point* of φ and x° is a *stationary point* of φ *with respect to minimization.*

Remark 7. It is easy to see that $\tau[\varphi](x^\circ) = -|\nabla\varphi|(x^\circ)$ where $|\nabla\varphi|(x^\circ)$ is a *slope* of φ at x° [2] (see also [7]).

In the smooth case inf-stationarity reduces to the classical one.

Proposition 18. *Let φ be strictly differentiable at x°. φ is weakly inf-stationary at x° if and only if $\nabla\varphi(x^\circ) = 0$.*

In general the notion of inf-stationarity appears to be stable relative to small deformations of the data.

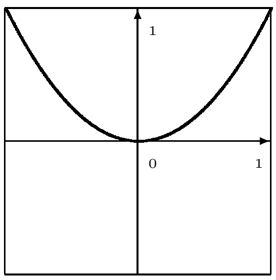
Proposition 19. *Let ψ be strictly differentiable at x° with $\nabla\psi(x^\circ) = 0$. If φ is weakly inf-stationary at x° then $\varphi + \psi$ is weakly inf-stationary at x°.*

6 Examples

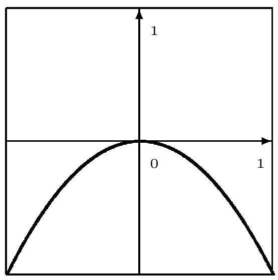
The illustrated versions of the examples of inf-stationary functions from [13] are presented in this section. All the functions below are from $\mathbb{R}$ to $\mathbb{R}$ and are weakly inf-stationary at zero, though only for the first one zero is really a point of minimum.

First, three classical examples of stationary points.

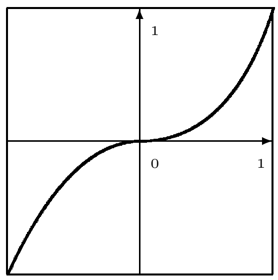
1. $y = x^2$



2. $y = -x^2$



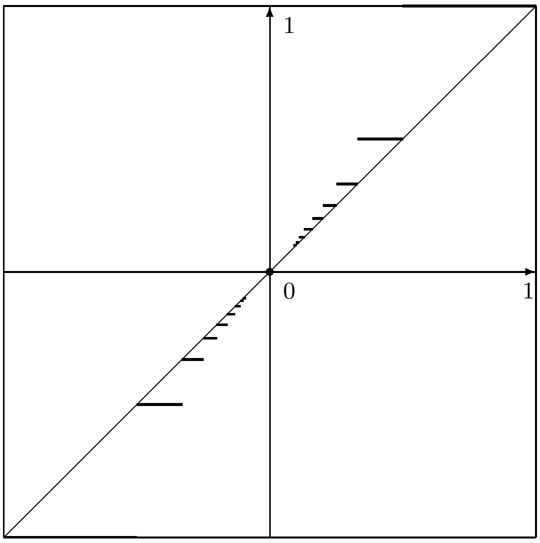
3. $y = x^3$



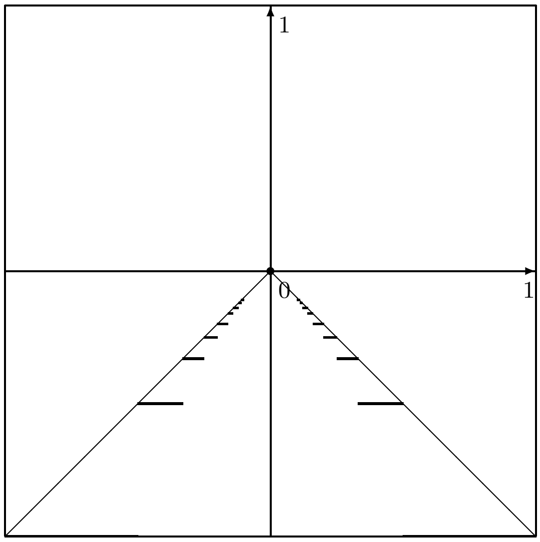
[2] personal communication

Two piecewise constant functions.

4. $$y = \begin{cases} -\dfrac{1}{n}, & -\dfrac{1}{n} < x \le -\dfrac{1}{n+1},\ n = 1, 2, \ldots, \\ 0, & x = 0, \\ \dfrac{1}{n}, & \dfrac{1}{n+1} < x \le \dfrac{1}{n},\ n = 1, 2, \ldots. \end{cases}$$

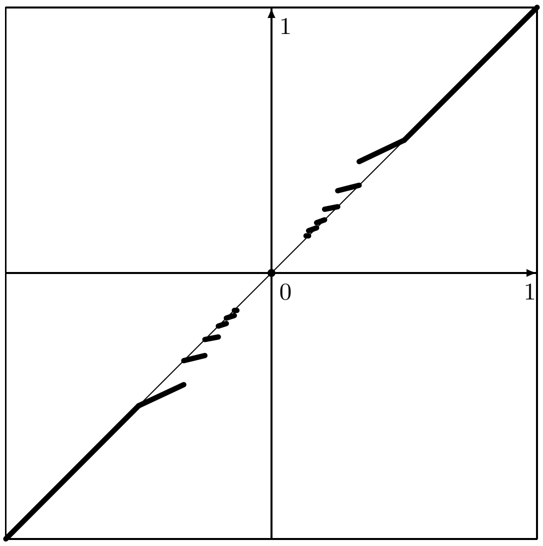


5. $$y = \begin{cases} -\dfrac{1}{n}, & -\dfrac{1}{n} < x \le -\dfrac{1}{n+1},\ n = 1, 2, \ldots, \\ 0, & x = 0, \\ -\dfrac{1}{n}, & \dfrac{1}{n+1} \le x < \dfrac{1}{n},\ n = 1, 2, \ldots. \end{cases}$$



Finally, a piecewise linear function.

6. $$y = \begin{cases} -\dfrac{1}{n} + \dfrac{1}{n}(x + \dfrac{1}{n}), & -\dfrac{1}{n} < x \le -\dfrac{1}{n+1}, \\ 0, & x = 0, \\ \dfrac{1}{n} + \dfrac{1}{n}(x - \dfrac{1}{n}), & \dfrac{1}{n+1} < x \le \dfrac{1}{n}. \end{cases}$$



References

1. N. Bourbaki, *Variétés Différentielles et Analytiques*, Herman, Paris, 1967.
2. E. De Giorgi, A. Marino, and M. Tosques, *Problemi di evoluzione in spazi metrici e curve di massima pendenza*, Atti Accad. Nat. Lincei, Cl. Sci Fiz. Mat. Natur. **68** (1980), 180–187.
3. A. V. Dmitruk, A. A. Milyutin, and N. P. Osmolovsky, *Lyusternik's theorem and the theory of extrema*, Russ. Math. Surv. **35** (1980), 11–51.

4. I. Ekeland, *On the variational principle*, J. Math. Anal. Appl. **47** (1974), 324–353.

5. M. Fabian, *Subdifferentiability and trustworthiness in the light of a new variational principle of Borwein and Preiss*, Acta Univ. Carolinae **30** (1989), 51–56.

6. A. D. Ioffe, *On subdifferentiability spaces*, Ann. New York Acad. Sci. **410** (1983), 107–119.

7. ———, *Metric regularity and subdifferential calculus*, Rus. Math. Surveys **55** (2000), 501–558.

8. A. Y. Kruger, *Generalized differentials of nonsmooth functions and necessary conditions for an extremum*, Siberian Math. J. **26** (1985), 370–379.

9. ———, *Strict ε-semidifferentials and extremality conditions*, Dokl. Akad. Nauk Belarus **41** (1997), no. 3, 21–26, in Russian.

10. ———, *On extremality of sets systems*, Dokl. Nat. Akad. Nauk Belarus **42** (1998), no. 1, 24–28, in Russian.

11. ———, *On necessary and sufficient extremality conditions*, Dynamical Systems: Stability, Control, Optimization, Abstracts, Vol. 2, Minsk, 1998, pp. 165–168.

12. ———, *Strict (ε, δ)-semidifferentials and extremality of sets and functions*, Dokl. Nat. Akad. Nauk Belarus **44** (2000), no. 4, 21–24, in Russian.

13. ———, *Strict (ε, δ)-semidifferentials and extremality conditions*, Optimization **51** (2002), 539–554.

14. ———, *On Fréchet subdifferentials*, J. Math. Sci. **116** (2003), no. 3, 3325–3358.

15. A. Y. Kruger and B. S. Mordukhovich, *Extremal points and the Euler equation in nonsmooth optimization*, Dokl. Akad. Nauk BSSR **24** (1980), no. 8, 684–687, in Russian.

16. B. Kummer, *Inverse functions of pseudo regular mappings and regularity conditions*, Math. Program., Ser. B **88** (2000), 313–339.

17. B. S. Mordukhovich and Y. Shao, *Extremal characterizations of Asplund spaces*, Proc. Amer. Math. Soc. **124** (1996), 197–205.

18. ———, *Nonsmooth sequential analysis in Asplund spaces*, Trans. Amer. Math. Soc. **348** (1996), 1235–1280.

19. R. R. Phelps, *Convex Functions, Monotone Operators and Differentiability, 2nd edition*, Lecture Notes in Mathematics, Vol. 1364, Springer, New York, 1993.

20. D. Yost, *Asplund spaces for beginners*, Acta Univ. Carol., Math. Phys. **34** (1993), no. 2, 159–177.