Control and Cybernetics

VOL. 14 (1985) No. 4

Hölder subgradients and applications in optimization

by

PHAN QUOC KHANH
Institute of Mathematics,
Hanoi, Vietnam.

NGUYEN THIEN LUAN

Institute of Mathematics,
Hanoi, Vietnam

We introduce Hölder subgradients as generalized ϕ-subgradients with a class ϕ of functionals exhibiting properties related to Hölder analysis and consider its properties needed for investigation of optimization problems. Necessary conditions for optimality and conditions for normality are derived for problems with locally Hölder data.

The newness of the results lies in the use of class ϕ of rather simple nonlinear functionals instead of X^* to approximate functionals in a relaxed sense and on the absence of Lipschitz conditions on the functionals involved in the problems.

1. Introduction

Nonsmooth analysis has been extensively developed in the last decade with successful applications in optimization and stabilization, in considering controllability and in investigations of existence of solutions to equations or inclusions. Many concepts of differentiability have been introduced in order to relax classical conditions on smoothness of data of problems in the mentioned above mathematical branches.

It was Clarke who began this process by defining *generalized gradients* [2] and then provided many uses in the calculus of variations [4], [6], [7], optimal control [5], [6], [7] and mathematical programming [3]. Soon after Clarke's beginning many experts in optimization and control theory made important contributions introducing various notions of generalized differentiability. Halkin [10] used his concept of *screen* for operators in finite dimensional spaces to obtain interior mapping theorems and necessary conditions for optimality in terms of Lagrange multipliers. Warga [19] introduced *derivate containers* as a tool for studies of inverse functions, controllability and ncessary conditions for extrema. In [17] Pourciau with the aid of his *generalized derivatives* managed to refine concepts and results of Clarke proving three fundamental theorems in analysis for generalized differentiability (interior mapping theorem, inverse mapping theorem and implicit mapping theorem) in finite dimensional spaces and a necessary condition for optimality in a multiplier rule form. A couple of years later Aubin [1] proposed a new approach to nonsmooth analysis based on *contingent derivatives.* Using the concept of *shield*, a generalization of the classical strong derivative, Pham Huy Dien [13] obtained three mentioned above theorems in infinite dimensional spaces.

The common feature in encountered concepts of generalized differentiability and generally in other formerly known notions is that a functional (or an operator) may be approximated by a set of linear functionals (or linear operators) when it is regular enough in a certain relaxed sense of smoothness. Besides, these concepts are effective mainly in problems with locally Lipschitz data. For instance, take very simple function $f:\mathbf{R}\to\mathbf{R}$, $f(x)=|x|^\alpha$, $0<\alpha<1$, we see that all concepts of differentiability based on linear approximations are not appropriate for considering the behaviour of this function at $x=0$. For this kind of situation in optimization see Example 3–4. Two following facts should be added here. First, the concept of *ϕ-subgradient* in [8] of Dolecki and Kurcyusz is quite different from those ones mentioned above. It is a set of nonlinear functionals in a general class ϕ but it is used to approximate other functionals in the classical (not relaxed) sense of subgradient. The idea is merely replacing X^* by ϕ. Second, for multifunctions many concepts of differentiability based on nonlinear approximations have appeared (see e.g. Pham Huu Sach [14], [20]).

In the present paper we introduce *α-Hölder subgradients* for functionals on a normed space X based on a Hölder approximation in a relaxed sense. Namely we define a directional derivative of a functional f by means of

$$f_\alpha(x; v)=\limsup_{\lambda\downarrow 0}\frac{f(x+\lambda v)-f(x)}{\lambda^\alpha}$$

and an α-Hölder subgradient of f at x as the set of all $\varphi\in\phi$ such that $\varphi(v)\leqslant f_\alpha(x;v)$ for all $v\in X$. Here ϕ is a class of rather simple nonlinear functionals we choose properly so that many locally Hölder functionals are Hölder subdifferentiable. After a study of calculus of Hölder subgradients we are in a position to derive necessary conditions for optimality for mathematical programming problems with locally Hölder data and conditions for the normality of such problems.

2. Hölder functionals and Hölder subgradients

Throughout the paper (unless otherwise specified) let X be a normed space, X^* be its dual space and $0 < \alpha \leqslant 1$. We use the notations $B(x_0, \delta) \overset{\mathrm{df}}{=} \{x \in X / \|x - x_0\| < \delta\}$ and $B(U, \delta) \overset{\mathrm{df}}{=} \bigcup_{x \in U} B(x, \delta)$ for $U \subset X$.

A functional $f: X \to \mathbf{R}$ is said to be *(globally) Hölder of degree* α if there exists $K > 0$ such that

$$|f(x) - f(y)| \leqslant K \|x - y\|^\alpha \quad \text{for } x, y \in X.$$

f is called *locally Hölder of degree* α at x_0 if there are a neighbourhood $U(x_0)$ of x_0 and $K > 0$ such that

$$|f(x) - f(y)| \leqslant K \|x - y\|^\alpha \quad \text{for } x, y \in U(x_0).$$

We say that f is *locally Hölder of degree* α (in a given subset of X) if for each x (in this subset) f is locally Hölder of degree α at x.

A *recession functional of degree* α of a functional f is a functional denoted by $f0^\alpha$ on X such that its epigraph is [15]

$$\operatorname{epi}(f0^\alpha) = \{(x, v) \in x \times \mathbf{R} / \operatorname{epi} f + (\lambda x, \lambda^\alpha v) \subset \operatorname{epi} f, \forall \lambda > 0\}.$$

LEMMA 2.1. *If f is subadditive and positively homogeneous of degree α, then $f0^\alpha = f$.*

Proof. Let (y, μ) be an arbitrary point in $\operatorname{epi} f$. Then for each $(x, v) \in \operatorname{epi} f$ and $\lambda > 0$ one has

$$f(x + \lambda y) \leqslant f(x) + f(\lambda y) = f(x) + \lambda^\alpha f(y) \leqslant v + \lambda^\alpha \mu,$$

so $(x, v) + (\lambda y, \lambda^\alpha \mu) \in \operatorname{epi} f$, i.e. $(y, \mu) \in \operatorname{epi} f0^\alpha$.

Conversely, if $(\zeta, \gamma) \in \operatorname{epi} f0^\alpha$ then since $(0, 0) \in \operatorname{epi} f$ one has for $\lambda > 0$ $(\lambda\zeta, \lambda^\alpha \gamma) \in \operatorname{epi} f$, i.e. $f(\lambda\zeta) \leqslant \lambda^\alpha \gamma$. Taking $\lambda = 1$ we see that $(\zeta, \gamma) \in \operatorname{epi} f$. ■

LEMMA 2.2.

a) *$f0^\alpha$ is positively homogeneous of degree α and*

$$f0^\alpha(\zeta) = \sup_{\substack{x \operatorname{dom} f \\ \lambda > 0}} \frac{f(x + \lambda\zeta) - f(x)}{\lambda^\alpha} \tag{1}$$

b) *f satisfies the Hölder condition*

$$|f(x) - f(y)| \leqslant K \|x - y|^\alpha, \tag{2}$$

K being the minimal Hölder constant if and only if

$$\sup_{\|\zeta\| \leqslant 1} f0^\alpha(\zeta) = K$$

c) *In particular, if f is subadditive and positively homogeneous of degree α, then f is (globally) Hölder with exponent α if and only if $|f(x)| \leqslant K\ \|x\|^\alpha$. In this case*

$$\sup\{|f(x)|/\|x\| \leqslant 1\} = \sup\{f(x)/\|x\| \leqslant 1\} = K.$$

Proof.

a) By definition $(\zeta, v) \in \operatorname{epi} f0^\alpha$ means that $(x+\lambda\zeta, \mu+\lambda^\alpha v) \in \operatorname{epi} f$ for every $(x, y) \in \operatorname{epi} f$ and $\lambda > 0$, i.e. $f(x+\lambda\zeta) \leqslant f(x)+\lambda^\alpha v$. Hence $(\zeta, v_1) \in \operatorname{epi} f0^\alpha$ for all $v_1 \geqslant v$, i.e. $\operatorname{epi} f0^\alpha$ is really an epigraph and we obtain (1). From this formula follows immediately the required homogeneity of $f0^\alpha$.

b) By a) we see that (2) holds if and only if

$$+\infty > K = \sup_{x \neq y} \frac{f(x)-f(y)}{\|x-y\|^\alpha} = \sup_{\substack{x-y=\lambda\zeta \\ \|\zeta\|=1, \lambda>0}} \frac{f(y+\lambda\zeta)-f(y)}{\lambda^\alpha} =$$

$$= \sup_{\|\zeta\|=1} (f0^\alpha)(\zeta) = \sup_{\|\zeta\|\leqslant 1} (f0^\alpha)(\zeta).$$

c) obvious from b) and Lemma 2.1.

LEMMA 2.3. *If f is subadditive, positively homogeneous of degree α, $0<\alpha<1$, and $|f(x)| \leqslant K\ \|x\|^\alpha$ for all $x \in X$, then f is Hölder and nonnegative.*

Proof. Assume the contrary that there is $\bar{x} \in X$, $\|\bar{x}\| = 1, f(\bar{x}) < 0$. We have $0 = f(0) = f(\bar{x}-\bar{x}) \leqslant f(\bar{x})+f(-\bar{x})$. So $f(-\bar{x}) > 0$. Using Lemmas 2.1 and 2.2 it is not difficult to see that the function $g: \mathbf{R} \to \mathbf{R}$ defined by $g(t) \stackrel{\mathrm{df}}{=} f(t\bar{x})$ is Hölder with Hölder constant $K = f(-\bar{x})$. Setting $H = f(\bar{x})$ and taking $p>0, q>0$ arbitrarily we have $q^\alpha K - p^\alpha H \leqslant K(p+q)^\alpha$ or

$$(q^\alpha + Mp^\alpha)^{1/\alpha} \leqslant (p+q), \tag{3}$$

where $M = -HK^{-1}$. Put $p = 1$ and $q = ([1/\alpha]\, M)^{1/\alpha-1}$, where $[1/\alpha]$ is the integer part of $1/\alpha$, in the left-hand side of (3). Then Taylor expansion yields

$$(q+M)^{1/\alpha} > q + M^{1/\alpha} + 1 > q+1,$$

contradicting (3).

For $f: X \to \mathbf{R}$ we call the following functional on X

$$f_\alpha(x_0; v) \stackrel{\mathrm{df}}{=} \limsup_{\lambda \downarrow 0} \frac{f(x_0+\lambda v)-f(x_0)}{\lambda^\alpha}$$

a *directional α-Hölder derivative* of f at x_0. Then $f_\alpha(x_0; \cdot)$ is positively homogeneous of degree α. If f is locally Hölder of degree α at x_0 then a straighforward computation shows that $f_\alpha(x_0; \cdot)$ is (globally) Hölder of degree α.

Let ϕ^α be the set of all continuous functionals φ on X which are positively homogeneous of degree α, $\varphi(-x) = -\varphi(x)$ and bounded in the sense that $\varphi(x) \leqslant K\,\|x\|^\alpha$.

DEFINITION 2.4. *α-Hölder subgradient* (briefly *α-subgradient*) of functional f at x, denoted by $\partial_\alpha f(x)$, is the set of all $\varphi \in \phi^\alpha$ such that $\varphi(v) \leqslant f_\alpha(x; v)$ for all $v \in X$. If $\partial_\alpha f(x) \neq \phi$ then f is called *α-subdifferentiable* at x.

REMARK 2.5. If $0 < \beta < \alpha < \gamma \leqslant 1$ and $f_\alpha(x; v)$ is finite, then $f_\beta(x; v) = 0$ and $f_\gamma(x; v)$ is infinite. If f is α-subdifferentiable at x then is also β-subdifferentiable at x and $\partial_\beta f(x) = \{0\}$. In particular, a locally Lipschitz functional (at x) f is locally Hölder of degree α and α-subdifferentiable (at x) for all $\alpha \in (0, 1)$ and $\partial_\alpha f(x) = \{0\}$.

For $f: X \to \mathbf{R}$ and a fixed $\delta > 0$ we define a transformation G^δ by $G^\delta(f)(x) \overset{\mathrm{df}}{=} |f(x)|^\delta \operatorname{sgn} f(x)$. We omit the easy proofs of the following properties

a) if $f(x) \leqslant g(x)$ then $G^\delta(f)(x) \leqslant G^\delta(g)(x)$;

b) for $\lambda > 0$ $G^\delta(\lambda f) = \lambda^\delta G^\delta(f)$;

c) $G^\delta(-f) = -G^\delta(f)$;

d) $G^{1/\delta}(G^\delta(f) = f$.

If a is a real number we write $G^\delta a = |a|^\delta \operatorname{sgn} a$.

Now we prove the following properties

e) for $\Delta \subset X$, $\sup\limits_{x \in \Delta} G^\delta(f)(x) = G^\delta\big(\sup\limits_{x \in \Delta} f(x)\big)$;

f) for $\Delta \subset X$, $\inf\limits_{x \in \Delta} G^\delta(f)(x) = G^\delta\big(\inf\limits_{x \in \Delta} f(x)\big)$;

g) for a set U of functionals on X,

$$\sup_{\psi \in U} G^\delta(\psi)(x) = G^\delta\big(\sup_{\psi \in U} \psi(x)\big);$$

h) for a set U of functionals on X,

$$\inf_{\psi \in U} G^\delta(\psi)(x) = G^\delta\big(\inf_{\psi \in U} \psi(x)\big).$$

Proof.

e) By a) we have $G^\delta(\sup f(x)) \geqslant G^\delta(f)(x)$ for all x. So $G^\delta(\sup f(x)) \geqslant \geqslant \sup G^\delta(f)(x)$. Conversely, we see that $G^\delta(f)(x) \leqslant \sup G^\delta(f)(x)$ and then $f(x) \leqslant G^{1/\delta}(\sup G^\delta(f)(x))$. Consequently, $\sup f(x) \leqslant G^{1/\delta}(\sup G^\delta(f)(x))$ or $G^\delta(\sup f(x)) \leqslant \sup G^\delta(f)(x)$.

The proofs of f), g) and h) are analogous.

For $\varphi \in \phi^\alpha$ we define $\|\varphi\|_\alpha = \sup\limits_{\|x\| \leqslant 1} \varphi(x)$. Of course $\|\cdot\|_\alpha$ is a norm. Denote $X^\alpha \overset{\mathrm{df}}{=} \operatorname{Lin}\{\varphi / \exists \xi \in X^*, G^\alpha \xi = \varphi\}$ we see that, for $0 < \alpha \leqslant 1$, X^α is a normed space with the norm $\|\cdot\|_\alpha$ (if $G^\alpha \xi = \varphi$ then $\|\varphi\|_\alpha = \|\xi\|^\alpha$) and X^α is a subspace of ϕ^α.

To compare with the Lipschitz case and ϕ-convexity let us recall some notions.

If a functional g is locally Lipschitz, then following Clarke a *directional derivative* of g at x is

$$g^0(x; v) \overset{\mathrm{df}}{=} \limsup_{\substack{\lambda \downarrow 0 \\ x' \to x}} \frac{g(x'+\lambda v)-g(x')}{\lambda}$$

and a *generalized gradient* of g at x is

$$\partial g(x) \overset{\mathrm{df}}{=} \{\xi \in X^*/\xi(v) \leqslant g^0(x; v)\ \forall v \in X\}.$$

A functional f is said to be *semiregular* at x if for all $v \in X$ we have

$$\limsup_{\substack{y \to 0 \\ \lambda \downarrow 0}} \frac{f(x+y+\lambda v)-f(x+y)}{\lambda} = \limsup_{\lambda \downarrow 0} \frac{f(x+\lambda v)-f(x)}{\lambda}.$$

Let ϕ be an arbitrary class of functionals on X. A set $\Delta \subset X$ is said to be *ϕ-convex* if Δ has the form $\{x \in X/\varphi_i(x) \leqslant \gamma_i, \varphi_i \in \phi, \gamma_i \in \mathbf{R}\}$. Δ is called *ϕ-closed* if $x_n \in \Delta$, $\varphi(x_n) \to \varphi(x)$ $\forall \varphi \in \phi$ implies $x \in \Delta$. In particular if $\phi = \phi^\alpha$ let us define for each $x \in X$ a functional $x(\varphi) = \varphi(x)$. Then x is a linear functional. We call the weakest topology in ϕ^α such that all $x \in X$ are continuous X-topology. If $U \subset \phi^\alpha$ is compact in X-topology we say that U is *X-compact*. Since X is also a class of functionals on ϕ^α we can speak about X-convexity and X-closedness.

Proposition 2.6. Suppose that a locally Lipschitz functional g is semiregular at 0 and $g(0) = 0$. Then $G^\alpha(g)$, $0 < \alpha \leqslant 1$, is α-subdifferentiable at 0 and

(i) $(G^\alpha(g))_\alpha(0; v) = G^\alpha(g^0(0; \cdot))(v)$;
(ii) $G^\alpha(\partial g(0)) \subset \partial_\alpha(G^\alpha(g))(0)$.

Proof. Set in this proof for brevity $G^\alpha(g) = f$.

(i) In view of the semiregularity of g we have

$$f_\alpha(0; v) = \limsup_{\lambda \downarrow 0} \frac{f(\lambda v)}{\lambda^\alpha} = \inf_{\beta > 0} \sup_{\lambda \leqslant \beta} \frac{f(\lambda v)}{\lambda^\alpha} =$$

$$= \inf_{\beta > 0} \sup_{\lambda \leqslant \beta} G^\alpha(g/\lambda)(\lambda v) = G^\alpha\left(\inf_{\beta > 0} \sup_{\lambda \leqslant \beta} \frac{g(\lambda v)}{\lambda}\right) =$$

$$= G^\alpha\left(\limsup_{\lambda \downarrow 0} \frac{g(\lambda v)}{\lambda}\right) = G^\alpha(g^0(0; \cdot))(v).$$

(ii) For $\xi \in \partial g(0)$, $v \in X$ we have $\xi(v) \leqslant g^0(0; v)$ and $G^\alpha(\xi)(v) \leqslant$ $\leqslant G^\alpha(g^0(0; v))(v) = f_\alpha(0; v)$, i.e. $G^\alpha(\partial g(0)) \subset \partial_\alpha f(0)$. ■

The following example shows a case in which (ii) becomes an equality.

Example 2–7. Let $g: \mathbf{R} \to \mathbf{R}$, $g(x) = \begin{cases} x & \text{if } x \geqslant 0, \\ \sin x & \text{if } x < 0. \end{cases}$
Let $\alpha = \dfrac{1}{2}$. Then

$$f(x) \overset{df}{=} G^{1/2}(g)(x) = \begin{cases} \sqrt{x} & \text{if } x \geqslant 0, \\ -|\sin x|^{1/2} & \text{if } x < 0. \end{cases}$$

g is locally Lipschitz at 0 and $\partial g(0) = \{1\}$. $f_{1/2}(0; v) = |v|^{1/2} \operatorname{sgn} v$ and $\partial_{1/2} f(0) = G^{1/2}(\partial g(0)) = \{\varphi\}$, where $\varphi(v) = |v|^{1/2} \operatorname{sgn} v$.

The semiregularity is really essential. Without it (ii) may not hold as shown by

EXAMPLE 2–8. Let $g: \mathbf{R} \to \mathbf{R}$, $g(x) = \begin{cases} x^2 \sin \dfrac{1}{x} & \text{if } x \neq 0, \\ 0 & \text{if } x = 0. \end{cases}$

Obviously g is not semiregular at 0. Let $\alpha = 1/2$. Then

$$f(x) \overset{df}{=} G^{1/2}(g)(x) = \begin{cases} |x|\, |\sin(1/x)| \operatorname{sgn}(\sin 1/x) & \text{if } x \neq 0, \\ 0 & \text{if } x = 0. \end{cases}$$

Then $f_{1/2}(0; v) \equiv 0$, so $\partial_{1/2} f(0) = \{0\}$. On the other hand $\partial g(0) = [-1, 1]$ and $G^{1/2}(\partial g(0)) = \{\varphi / \varphi(x) = |Kx|^{1/2} \operatorname{sgn} Kx, K \in [-1, 1]\}$, i.e. $G^{1/2}(\partial g(0)) \supset \supset \partial_{1/2} f(0)$.

If f is locally Lipschitz at x and semiregular at x, then f is 1-subdifferentiable at x and $\partial f(x) \subset \partial_1(f)$. If in addition $x = \mathbf{R}$, then $\partial f(x) = \partial_1 f(x)$. If f is only locally Lipschitz at x, then $f_1(x; v) \leqslant f^0(x; v)$ and $\phi^1 \supset X^*$, so we cannot say anything about the relation between $\partial f(x)$ and $\partial_1 f(x)$.

PROPOSITION 2–9.

a) $\partial_\alpha f(x)$ is X-convex and closed.

b) If f is locally Hölder of degree α at x with constant K, then $\|\varphi\|_\alpha \leqslant K$ for all $\varphi \in \partial_\alpha f(x)$ and $\partial_\alpha f(x)$ is X-compact.

Proof.

a) The X-convexity is clear by definition. Now suppose $\varphi_n \in \partial_\alpha f(x)$ and $\|\varphi_n - \varphi\|_\alpha \to 0$. Then for $v \in X$ one has $\varphi(v) = \varphi(v) - \varphi_n(v) + \varphi_n(v) \leqslant \|\varphi_n - \varphi\|_\alpha \|V\|^\alpha + f_\alpha(x; v) \leqslant \varepsilon \|v\|^\alpha + f_\alpha(x; v)$. Since ε is arbitrarily small we get $\varphi(v) \leqslant f_\alpha(x; v)$, i.e. $\partial_\alpha f(x)$ is closed in ϕ^α.

b) For $\varphi \in \partial_\alpha f(x)$ we have $\varphi(v) \leqslant f_\alpha(x; v) \leqslant K \|v\|^\alpha$ and also $-\varphi(v) = \varphi(-v) \leqslant f_\alpha(x, -v) \leqslant K \|v\|^\alpha$. Hence $-K \|v\|^\alpha \leqslant \varphi(v) \leqslant K \|v\|^\alpha$ and $\|\varphi\|_\alpha \leqslant K$.

One easily sees the X-closedness of $\partial_\alpha f(x)$. Then according to Alaoglu's theorem $\partial_\alpha f(x)$ is X-compact. ■

If f is not locally Hölder at x but $f_\alpha(x; v) \leqslant K \|v\|^\alpha$ then Proposition 2.9 is still true.

For $\Delta \subset \phi^\alpha$, $v \in X$ we call

$$C(v; \Delta) \overset{df}{=} \sup \{\varphi(v) / \varphi \in \Delta\}$$

the *support functional* of Δ. If f is locally Hölder and α-subdifferentiable at x then Proposition 2–9 yields

$$C\left(v; \partial_\alpha f(x)\right) = \max\left\{\varphi(v)/\varphi \in \partial_\alpha f(x)\right\}.$$

The following theorem will be important for our arguments later.

THEOREM 2.10. *Let f be locally Hölder of degree α and α-subdifferentiable at x. Then for $v \in X$*

$$C\left(v; \partial_\alpha f(x)\right) = f_\alpha(x; v).$$

Proof. By property g) of transformation G^δ one has

$$C\left(v; G^{1/\alpha}\left(\partial_\alpha f(x)\right)\right) = G^{1/\alpha}\left(C\left(\cdot, \partial_\alpha f(x)\right)\right)(v).$$

To finish the proof it suffices to verify

$$C\left(v; G^{1/\alpha}\left(\partial_\alpha f(x)\right)\right) = G^{1/\alpha}\left(f_\alpha(x; \cdot)\right)(v).$$

For the sake of simplicity the following notations will be used in the proof: $\Delta = G^{1/\alpha}\left(\partial_\alpha f(x)\right)$, $p(v) = G^{1/\alpha}\left(f_\alpha(x; \cdot)(v)\right)$ and $S = \{x \in X / \|x\| = 1\}$. Assume the contrary: that there exists $v_0 \in X$, $\|v_0\| = 1$ such that $\max\limits_{\psi \in \Delta} \psi(v_0) < p(v_0)$, we choose β satisfying $\max\limits_{\psi \in \Delta} \psi(v_0) < \beta < p(v_0)$ and an arbitrary $\eta \in \Delta$. Then $\eta(v_0) < \beta < p(v_0)$. Since η and p are continuous there is δ, $0 < \delta < 1$ such that $\eta(v) < \beta < p(v)$ for all $v \in B(v_0, \delta)$. We now construct $g: X \to \mathbf{R}$ as follows. We define $g(v) = \eta(v)$ if $v \notin \overline{B(v_0, \delta)} \cup \left(-\overline{B(v_0, \delta)}\right)$ and $g(v_0) = \beta$. For $v \in \overline{B(v_0, \beta)} \setminus \{v_0\}$ there is t_v, $0 \leqslant t_v \leqslant 1$ and $\bar{v} = v_0 + \delta \|v - v_0\|^{-1}(v - v_0)$ such that $v = \bar{v} + t_v(v_0 - \bar{v})$. Then we put $g(v) = t_v \beta + (1 - t_v)\eta(v)$. For $v \in -\overline{B(v_0, \delta)}$ there is $u \in \overline{B(v_0, \delta)}$, $v = -u$ and we take $g(v) = -g(u)$.

To prove the continuity on X of g we have to show only the continuity in $\overline{B(v_0, \delta)} \cup \left(-\overline{B(v_0, \delta)}\right)$. If $v_1 \in \overline{B(v_0, \delta)}$ and $\varepsilon > 0$, there is r_1 small enough such that $|\eta(v) - \eta(v_1)| < \dfrac{\varepsilon}{2}$ for $v \in B(v_1, r_1)$. Then

$$\begin{aligned}
|g(v) - g(v_1)| &= \left|(1 - \delta^{-1}\|v - v_0\|)\beta + \delta^{-1}\|v - v_0\|\,\eta(v) - \right.\\
&\quad \left. -(1 - \delta^{-1}\|v_1 - v_0\|)\beta + \delta^{-1}\|v_1 - v_0\|\,\eta(v_1)\right| \leqslant \\
&\leqslant \left|\|v - v_0\| - \|v_1 - v_0\|\right| |\beta|\,\delta^{-1} + \delta^{-1}\|v - v_0\|\,|\eta(v) - \eta(v_1)| + \\
&\quad + \left|\|v - v_0\| - \|v_1 - v_0\|\right| |\eta(v_1)|\,\delta^{-1} \leqslant \\
&\leqslant \delta^{-1}\|v_1 - v\|\left(|\beta| + |\eta(v_1)|\right) + |\eta(v) - \eta(v_1)|.
\end{aligned}$$

Put

$$D = \delta^{-1}\left(|\beta| + |\eta(v_1)|\right),\quad r = \begin{cases} r_1 & \text{if } D = 0,\\ \min\{r_1, \varepsilon/2D\} & \text{if } D \neq 0.\end{cases}$$

We have $|g(v) - g(v_1)| < \varepsilon$ for all $v \in B(v_1, r)$. For the case $v_1 \in -\overline{B(v_0, \delta)}$ the argument is similar. Thus g is continuous.

We define $\Theta : X \to \mathbf{R}$ by means of

$$\Theta(v) = \begin{cases} \|v\|\, g(\|v\|^{-1} v) & \text{if } v \neq 0, \\ 0 & \text{if } v = 0. \end{cases}$$

The conclusion of the theorem will follow if we can show $\Theta \in \Delta$ (due to the contradiction $\Theta(v_0) = \beta$).

Since $\Delta = \{\Theta \in \phi^1 / \Theta(v) \leqslant p(v), \forall v \in X\}$ the following will have to be checked

(i) $\Theta(r) \leqslant p(r)$ for $v \in S$,

(ii) $\Theta(-v) = -\Theta(v)$ for $v \in S$,

(iii) $\Theta(\lambda v) = \lambda \Theta(v)$ for $\lambda > 0$,

(iv) Θ is continuous on X,

(v) $|\Theta(v)| \leqslant c\,\|v\|$, c being a positive number.

(ii) through (v) are evident. To prove (i) we have $\Theta(v) = g(v) = t_v \beta + (1 - t_v)\,\eta(v) \leqslant \beta < p(v)$ for $v \in \overline{B(v_0, \delta)} \cap S$. For $v \in -\overline{B(v_0, \delta)} \cap S$ there is $u \in \overline{B(v_0, \delta)}$ such that $v = -u$, $\Theta(v) = g(v) = -g(u) = -(t_u \beta + (1 - t_u)\,\eta(u)) < -\eta(u) = \eta(-u) = \eta(v) \leqslant p(v)$. For the other elements v of S one sees $\Theta(v) = g(v) = \eta(v) \leqslant p(v)$.

Corollary 2.11 [12]. *Let f be locally Hölder of degree α and α-subdifferentiable at x. If*

$$-f_\alpha(x; -\bar{v}) \leqslant \varrho \leqslant f_\alpha(x, \bar{v}),$$

Then there exists $\varphi \in \partial_\alpha f(x)$ such that $\varphi(\bar{v}) = \varrho$.

Proof. Using notations Δ and $p(v)$ as in the proof of Theorem 2.10 and putting $\beta = G^{1/\alpha}(\varrho)$ we see that $-p(-\bar{v}) \leqslant \beta \leqslant p(\bar{v})$. Without loss of generality we may assume that $\|\bar{v}\| = 1$. If $\beta = p(\bar{v})$, then by Theorem 2.10 $\beta = \max\{\psi(\bar{v}) / \psi \in \Delta\}$, i.e. there is $\Theta \in \Delta$ such that $\beta = \Theta(\bar{v})$. If $\beta = -p(-\bar{v})$, then $-\beta = p(-\bar{v}) = \max\{\psi(-\bar{v}) / \psi \in \Delta\}$, then there exists $\Theta \in \Delta$ such that $-\beta = \Theta(-\bar{v})$, so $\beta = \Theta(\bar{v})$. Now suppose $-p(-\bar{v}) < \beta < p(\bar{v})$. Then there is $\eta \in \Delta$ such that $\eta(\bar{v}) < \beta < p(\bar{v})$. By an argument analogous to that in the proof of Theorem 2.10 we can find $\Theta \in \Delta$ such that $\Theta(\bar{v}) = \beta$. Hence $\varphi(v) = G^\alpha(\Theta(\cdot))(v)$ is evidently a required functional. ■

Proposition 2.13. Let f be locally Hölder and α-subdifferentiable at x. Let Ω be a X-convex subset of ϕ^α. Then $\partial_\alpha f(x)$ is contained in Ω if and only if

$$f_\alpha(x; v) \leqslant \sup\{\varphi(v) / \varphi \in \Omega\}.$$

Proof. The "necessary" is tautologous by Theorem 2.10. We show the "sufficient". Being X-convex Ω has the form

$$\Omega = \{\varphi \in \phi^\alpha / \varphi(v_t) \leqslant \beta_t,\ v_t \in X,\ \beta_t \in \mathbf{R}\}.$$

If some φ_0 in $\partial_\alpha f(x)$ were not in Ω, there would be v_{t_0} such that $\varphi_0(v_{t_0}) > \beta_{t_0} \geqslant \sup\{\varphi(v_{t_0})/\varphi\in\Omega\} \geqslant f_\alpha(x; v_{t_0})$. Hence $\varphi_0 \notin \partial_\alpha f(x)$, a contradiction. ■

Now we give some criterions for α-subdifferentiability.

PROPOSITION 2.13. If there exists a locally Lipschitz functional g such that $G^{1/\alpha}(f_\alpha(x;\cdot))(v) = g^0(x; v)$, then f is α-subdifferentiable at x and $\partial_\alpha f(x) \supset \supset G^\alpha(\partial g(x))$. ■

The proof is evident and omitted.

PROPOSITION 2.14. Let f be locally Hölder of degree α at x. If $f_\alpha(x; v)$ satisfies one of the following conditions

(i) $f_\alpha(x; v)$ is subadditive;

(ii) $G^{1/\alpha}(f_\alpha(x;\cdot))(v) \overset{\mathrm{df}}{=} p(v)$ is convex,

then f is α-subdifferentiable at x and $\partial_\alpha f(x) \supset G^\alpha(\partial p(0))$ for the case (ii), where ∂p is subdifferential of convex functional p.

Proof.

(i) immediate from an appeal to Lemma 2.3.

(ii) taking $\xi\in\partial p(0)$ one sees that

$$\varphi(v) \overset{\mathrm{df}}{=} G^\alpha(\xi)(v) \leqslant G^\alpha(p)(v) = f_\alpha(x; v),$$

so that $\varphi\in\partial_\alpha f(x)$.

PROPOSITION 2.15.

(i) If f is α-subdifferentiable at x, then for all $v\in X$, we have

$$f_\alpha(x; v) + f_\alpha(x; -v) \geqslant 0. \tag{4}$$

(ii) If f is locally Hölder at x and satisfies (4), then f is α-subdifferentiable at x.

Proof.

(i) $\varphi\in\partial_\alpha f(x)$ implies $\varphi(v) \leqslant f_\alpha(x; v)$ and $-\varphi(v) = \varphi(-v) \leqslant f_\alpha(x; -v)$. Adding two inequalities entails (4).

(ii) From (4) follows

$$\varphi(v) \overset{\mathrm{df}}{=} \frac{1}{2}\left(f_\alpha(x; v) - f_\alpha(x; -v)\right) \leqslant f_\alpha(x; v)$$

for all $v\in X$. Then it is plain that $\varphi\in\phi^\alpha$ and $\varphi\in\partial_\alpha f(x)$. ■

Now suppose $f_i: X\to\mathbf{R}$, $i = 1, 2, \ldots, n$. Let

$$m(x) \overset{\mathrm{df}}{=} \max\{f_i(x)/i = 1, 2, \ldots, n\} \tag{5}$$

and let $I(x)$ stand for the set of the indices in $\{1, 2, \ldots, n\}$ at which the maximum in (5) is attained.

PROPOSITION 2.16. If $f_i, i = 1, 2, \ldots, n$, are locally Hölder and α-subdifferentiable at x, then so is m and

$$\partial_x m(x) = co_x \{\partial_\alpha f_i(x)/i \in I(x)\},$$

where co_x denotes X-convex hulls in ϕ^α.

Proof. m is clearly locally Hölder at x. For $x \in X, v \in X$ and λ small enough we have $I(x+\lambda v) \subset I(x)$. Indeed, if $j \notin I(x)$, then $f_j(x) < m(x)$. By virtue of the continuity of m and f_j we can take a neighbourhood of x so that $f_j(y) < m(y)$ for all y in it and for all $j \notin I(x)$. If λ is small enough then $x+\lambda v$ belongs to the mentioned neighbourhood and $j \notin I(x+\lambda v)$, so $I(x+\lambda v) \subset I(x)$. Therefore

$$\begin{aligned} m_\alpha(x; v) &= \limsup_{\lambda \downarrow 0} \frac{m(x+\lambda v)-m(x)}{\lambda^\alpha} = \limsup_{\lambda \downarrow 0} \max_{i \in I(x)} \frac{f_i(x+\lambda v)-f_i(x)}{\lambda^\alpha} = \\ &= \max_{i \in I(x)} \limsup_{\lambda \downarrow 0} \frac{f_i(x+\lambda v)-f_i(x)}{\lambda^\alpha} = \max_{i \in I(x)} f_{i\alpha}(x; v) = \\ &= \max_{i \in I(x)} C(v; \partial_\alpha f_i(x)) = \max_{i \in I(x)} \max \{\varphi(v)/\varphi \in \partial_\alpha f_i(x)\} \leqslant \\ &\leqslant \max \{\varphi(v)/\varphi \in co_x(\partial_\alpha f_i(x)), i \in I(x)\}. \end{aligned}$$

Now Proposition 2.12 yields $\partial_x m(x) \subset co_x \{\partial_\alpha f_i(x), i \in I(x)\}$.

Conversely, for all $i \in I(x)$ and $v \in X$ we have

$$\frac{m(x+\lambda v)-m(x)}{\lambda^\alpha} \geqslant \frac{f_i(x+\lambda v)-f_i(x)}{\lambda^\alpha},$$

and so $m_\alpha(x; v) \geqslant f_{i\alpha}(x; v)$. Then $\partial_\alpha f_i(x) \subset \delta_\alpha m(x)$ for all $i \in I(x)$ and then

$$co_x \{\partial_\alpha f_i(x), i \in I(x)\} \subset \partial_\alpha m(x).$$

PROPOSITION 2–17. If f and g are locally Hölder and α-subdifferentiable at x, then

$$\partial_\alpha (f+g)(x) \subset co_x (\partial_\alpha f(x) + \partial_\alpha g(x)).$$

(However, the left-hand side may be empty.)

Proof.

$$\begin{aligned} (f+g)_\alpha(x; v) &\leqslant f_\alpha(x; v) + g_\alpha(x; v) = C(v; \partial_\alpha f(x)) + \\ &+ C(v; \partial_\alpha g(x)) = \max \{\varphi(v)+\psi(v)/\varphi+\psi \in \partial_\alpha f(x)+\partial_\alpha g(x)\} \leqslant \\ &\leqslant \max \{\Theta(v)/\Theta \in co_x(\partial_\alpha f(x)+\partial_\alpha g(x))\}. \end{aligned}$$

Now applying Proposition 2.12 yields the required inclusion. ■

PROPOSITION 2.18. Let $h: \mathbf{R} \to \mathbf{R}$, $h \in C^1$ and $h'(f(x_0)) \geqslant 0$. Let f be locally Hölder and α-subdifferentiable at x_0. Then the superposition hf is α-subdifferentiable at x_0 and

$$\partial_\alpha (hf)(x_0) \subset h'(f(x_0))\, \partial_\alpha f(x_0).$$

. Proof. Applying the mean value theorem we see that

$$\limsup_{\lambda \downarrow 0} \frac{hf(x_0+\lambda v)-hf(x_0)}{\lambda^\alpha} = h'(f(x_0)) f_\alpha(x_0; v).$$

So hf is α-subdifferentiable at x_0. Moreover

$$\begin{aligned}(hf)_\alpha(x_0; v) &= h'(f(x_0)) f_\alpha(x_0; v) = \\ &= h'(f(x_0)) \max\{\varphi(v)/\varphi \in \partial_\alpha f(x_0)\} = \\ &= \max\{\psi(v)/\psi \in h'(f(x_0))\, \partial_\alpha f(x_0)\}.\end{aligned}$$

■

PROPOSITION 2.19. If f has a local minimum at x_0, then f is α-subdifferentiable at x_0 and $0 \in \partial_\alpha f(x_0)$.

The proof is evident and omitted.

For $a, b \in x$ we denote $[a, b] = \{x/x = a+t(b-a), t \in [0, 1]\}$ and $(a, b) = = \{x/x = a+t(b-a), t \in (0, 1)\}$.

PROPOSITION 2.20. Let $\Omega \subset \mathbf{R}$ be an open subset and $f: \Omega \to \mathbf{R}$ be locally Hölder and α-subdifferentiable at every point in $[a, b] \subset \Omega$. Then there exist $c \in (a, b)$ and $\varphi \in \partial_\alpha f(c)$ such that $\varphi(b-a) = 0$.

Proof. 1. If $f(a) = f(b)$ putting $g(t) = f(a)+t(b-a)$, $t \in [a, b]$, one sees that $g(0) = g(1) = f(a) = f(b)$ and g is continuous on $[0, 1]$. We have three following possibilities.

(i) If $g(t) = \text{const}$, $t \in [0, 1]$, we take $\bar{t} \in (0, 1)$ arbitrarily. Then setting $c = a+\bar{t}(b-a)$ we have $f_\alpha(c; b-a) = 0$. Analogously $f_\alpha(c; a-b) = 0$. By Corollary 2.11 there is $\varphi \in \partial_\alpha f(c)$ such that $\varphi(b-a) = 0$.

(ii) If g attains the maximum at some $\bar{t} \in (0, 1)$, then

$$0 \geqslant \frac{g(\bar{t}+\Delta t)-g(\bar{t})}{\Delta t^\alpha} = \frac{f(a+(\bar{t}+\Delta t)(b-a))-f(a+\bar{t}(b-a))}{\Delta t^\alpha}.$$

Therefore $f_\alpha(c; b-a) \leqslant 0$, where $c = a+\bar{t}(b-a)$. In a similar way we get $f_\alpha(c; a-b) \leqslant 0$. Both of two last inequalities cannot be strict thanks to Proposition 2.15. Hence $\varphi(b-a) = 0$ for all $\varphi \in \partial_\alpha f(c)$.

(iii) If g attains the minimum at some $\bar{t} \in (0, 1)$, then very much like above we have $f_\alpha(c; b-a) \geqslant 0$ and $f_\alpha(c; a-b) \geqslant 0$. So $-f_\alpha(c; a-b) \leqslant 0 \leqslant \leqslant f_\alpha(\acute{s}; b-a)$. By Corollary 2.11 there is $\varphi \in \partial_\alpha f(c)$ such that $\varphi(a-b) = 0$.

2. If $f(a) \neq f(b)$, we set $h(x) = f(x)-f(b)-\dfrac{f(a)-f(b)}{a-b}(x-a)$. Then $h(a) = h(b) = f(a)-f(b)$. According to above there exist $c \in (a, b)$ and $\varphi \in \partial_\alpha h(c) = = \partial_\alpha f(c)$ such that $\varphi(b-a) = 0$. ■

COROLLARY 2.21. *Let f be locally Hölder, α-subdifferentiable at x_0 and attain a local maximum at x_0. Then*

$$\partial_\alpha f(x_0) = \{0\},$$
$$f_\alpha(x_0, v) = \lim_{\lambda \downarrow 0} \frac{f(x_0+\lambda v)-f(x_0)}{\lambda^\alpha}.$$

Let Y be another normed space and $\Gamma: X \to 2^Y$ be a multifunction. Γ is said to be *closed at* x_0 if for each pair of sequences $\{x_n\} \subset X$ and $\{y_n\} \subset Y$ with the properties $x_n \to x_0$, $y_n \in \Gamma(x_n)$, $y_n \to y_0$ it follows that $y_0 \in \Gamma(x_0)$. Γ is called *upper Hausdorff semicontinuous* (u.H.s.c.) at x_0 if for each $\delta > 0$ there is a neighbourhood $V(x_0)$ of x_0 such that $x \in V(x_0)$ implies $\Gamma(x) \subset B(\Gamma(x_0), \delta)$. It is known that if Γ is u.H.s.c. at x_0 and $\Gamma(x_0)$ is closed then Γ is closed at x_0. Γ is said to be *locally Lipschitz* at x_0 if there is a neighbourhood $U(x_0)$ of x_0 and $K > 0$ such that, for $x_1, x_2 \in U(x_0)$, $H(\Gamma(x_1), \Gamma(x_2)) \leqslant K \|x_1 - x_2\|$, $H(\cdot, \cdot)$ being the Hausdorff distance.

Let $f: X \times Y \to \mathbf{R}$ be a functional.

We are now interested in Hölder properties of the following functional

$$f\Gamma(x) \stackrel{\mathrm{df}}{=} \inf_{y \in \Gamma(x)} f(x, y).$$

This functional is an important object of investigation in parameter optimization and has been studied by many authors (and is called in varying ways: marginal function, extreme value function, primal function...). Especially, its continuity (or semicontinuity) has been extensively considered.

Concerning this functional we shall prove here merely the following proposition.

Proposition 2.22. Assume that

(i) Γ is locally Lipschitz at x_0 (with K and $U(x_0)$ as above) and $\Gamma(x_0)$ is compact;

(ii) the mapping $y \to f(x, y)$ satisfies the uniform Lipschitz condition: there are $P > 0$ and $\delta > 0$ such that for $x \in U(x_0)$ and $y_1, y_2 \in B(\Gamma(x_0), \delta)$

$$f(x, y_1) - f(x, y_2) \leqslant P \|y_1 - y_2\|;$$

(iii) the mapping $x \to f(x, y)$ satisfies the uniform Hölder condition: there is $L > 0$ such that for $x_1, x_2 \in U(x_0)$ and $y \in B(\Gamma(x_0), \delta)$

$$f(x_1, y) - f(x_2, y) \leqslant L \|x_1 - x_2\|^\alpha.$$

Then f is locally Hölder of degree α at x_0.

Proof. Since Γ is locally Lipschitz at x_0 there is a neighbourhood $W \subset U(x_0)$ of x_0 with diameter smaller than 1 such that $\Gamma(x) \subset B(\Gamma(x_0), \delta)$ for $x \in W$. Let $x_1, x_2 \in W$. For each $y_1 \in \Gamma(x_1)$ there is $y_2 \in \Gamma(x_2)$ such that $\|y_2 - y_1\| \leqslant$ $\leqslant K \|x_2 - x_1\|$. One has

$$f(x_2, y_2) - f(x_1, y_1) = f(x_2, y_2) - f(x_1, y_2) + f(x_1, y_2) - f(x_1, y_1) \leqslant$$

$$\leqslant L\|x_2-x_1\|^\alpha+P\ \|y_2-y_1\| \leqslant L\|x_2-x_1\|^\alpha+PK\ \|x_2-x_1\| \leqslant$$
$$\leqslant (L+PK)\ \|x_2-x_1\|^\alpha.$$

So,

$$f\Gamma\ (x_2) \leqslant f(x_1, y_1)+(L+PK)\ \|x_2-x_1\|^\alpha$$

for every $y_1 \in \Gamma\ (x_1)$. Consequently

$$f\Gamma\ (x_2) \leqslant f\Gamma\ (x_1)+(L+PK)\ \|x_2-x_1\|^\alpha. \qquad ■$$

3. Applications in optimization

The optimization problem to be now considered is

$$\begin{cases} \text{minimize } g_0(x), \\ g_i(x) \leqslant 0,\ i \in I \overset{\mathrm{df}}{=} \{1, 2, ..., n\}, \\ h_j(x) = 0,\ j \in J \overset{\mathrm{df}}{=} \{1, 2, ..., m\}, \end{cases} \tag{A}$$

where x varies in a Banach space X, g_0, g_i and h_j are locally Hölder of degree α, $0 < \alpha \leqslant 1$.

Since $\|x-y\|^\alpha, 0 < \alpha \leqslant 1$, is a metric and X equipped with it becomes a complete space, we get the following version of Ekeland's variational principle [9].

LEMMA 3.1 [9]. *Let $F: X \to \mathbf{R}$ be lower semicontinuous and bounded from below. Let C be a closed subset of X. Let $\bar{x}$ in C satisfy $F(\bar{x}) \leqslant \inf_{x \in C} F(x)+\varepsilon$, where $\varepsilon > 0$ is arbitrary. Then there exists ζ in C such that*

$$\|\bar{x}-\zeta\|^\alpha \leqslant \sqrt{\varepsilon}, \tag{6}$$

$$F(x)+\sqrt{\varepsilon}\ \|x-\zeta\|^\alpha \geqslant F(\zeta) \quad \text{for all } x \text{ in } C. \tag{7}$$

For a subset $\Delta \subset \phi^\alpha$ we have (see the lemma in [16])

$$co_x\, \Delta = \{\varphi \in \phi^\alpha / \varphi(x) \leqslant \sup_{\psi \in \Delta} \psi(x)\ \forall x \in X\}. \tag{8}$$

THEOREM 3.2. *Suppose that g_0, g_i and h_j are α-subdifferentiable in a neighbourhood of $\bar{x}$. If $\bar{x}$ solves Problem (A) locally and the mappings $x \to \partial_\alpha g_0(x)$, $x \to \partial_\alpha g_i(x)$ and $x \to \partial_\alpha |h_j(x)|$ are u.H.s.c. at $\bar{x}$, then three following equivalent statements hold:*

a) $0 \in co_x \bigcup_{\substack{i \in I \cup \{0\} \\ j \in J}} (\partial_\alpha g_i(\bar{x}) \cup \partial_\alpha |h_j(\bar{x})|) \overset{\mathrm{df}}{=} co_x\, P(\bar{x})$;

b) $\max\{\varphi(x)/\varphi \in P(\bar{x})\} \geqslant 0$ for all $x \in X$;

c) $C(x; P(\bar{x})) \geqslant 0$ for all $x \in X$.

Proof. The equivalence between a) and b) is trivial by (8) and between b) and c) is clear by the definition of support functionals. Now starting by an idea of Clarke we prove a). Let $\varepsilon > 0$ be given. Define $F: X \to \mathbf{R}$ by

$$F(x) \stackrel{\mathrm{df}}{=} \max \{g_0(x) - g_0(\bar{x}) + \varepsilon,\ g_i(x),\ |h_j(x)|,\ 0 / i \in I, j \in J\}.$$

Obviously F is locally Hölder of degree α, bounded from below by 0 and $F(\bar{x}) = \varepsilon$. On account of Lemma 3.1 there is $\zeta \in X$ satisfying (6) and (7). Of course $F(\zeta) > 0$. For otherwise ζ would solve (A) and $\bar{x}$ would not, since $g_0(\zeta) < g_0(\bar{x})$.

An easy computation shows that $\partial_\alpha \|\cdot\|^\alpha(0)$ is contained in the unit ball B in ϕ^α. By Proposition 2.19 and (7) we obtain

$$0 \in co_x\left(\partial_\alpha F(\zeta) + \sqrt{\varepsilon}\, B\right).$$

Then an appeal to Proposition 2.16 yields

$$0 \in co_x\left(P(\zeta) + \sqrt{\varepsilon}\, B\right). \tag{9}$$

By Lemma 3.1 ζ depends on ε and $\zeta \to \bar{x}$ as $\varepsilon \to 0$. So $P(\zeta) + \sqrt{\varepsilon}\, B \stackrel{\mathrm{df}}{=} Q(\varepsilon)$ is a multifunction. $Q(\cdot)$ is u.H.s.c. at $\varepsilon = 0$ and so is $co_x\, Q(\cdot)$. Being a X-convex hull $co_x\, Q(0)$ is closed. Hence $co_x\, Q(\cdot)$ is closed at $\varepsilon = 0$. Thus (9) gives $0 \in co_x\, Q(0) = co_x\, P(\bar{x})$. ∎

The theorem extends Clarke's Lagrange multiplier rule [3] to problems with locally Hölder data. However, in the case $\alpha = 1$ and the functionals involved in the problem are semiregular at $\bar{x}$, our rule is weaker than Clarke's one because

$$co\left\{\bigcup\left(\partial g_i(\bar{x}) \cup \partial |h_j(\bar{x})|\right)\right\} \subset co_x\left\{\bigcup\left(\partial_1 g_i(\bar{x}) \cup \partial_1 |h_j(\bar{x})|\right)\right\}.$$

In particular, if an addition $X = \mathbf{R}$, then the two rules coincide.

The hypothesis on the u.H.s.c. in Theorem 3.2 plays a crucial role. But for the following problem without equality constraints

$$\begin{cases} \text{minimize } g_0(x), \\ g_i(x) \leqslant 0, \ i \in I, \\ x \in \Omega \subset X, \end{cases} \tag{B}$$

where g_0 and g_i are locally Hölder, Theorem 3.3 below allows to omit this hypothesis.

Note that if f is locally Hölder at x, then

$$f_\alpha(x; v) = \limsup_{\substack{\lambda \downarrow 0 \\ u \to v}} \frac{f(x + \lambda u) - f(x)}{\lambda^\alpha}.$$

Let $\Omega \subset X$ and $x_0 \in \Omega$. We recall the definition of the *tangent cone* of Ω at x_0:

$T(x_0/\Omega) \stackrel{\mathrm{df}}{=} \{v \in X /$ there are $x_k \in \Omega$ and $\alpha_k > 0$ such that $x_k \to x_0,\ \alpha_k(x_k - x_0) \to v$ as $k \to \infty\}$.

Evidently $T(x_0/\Omega)$ is a closed cone.

THEOREM 3.3. *If $\bar{x}$ is a local solution of Problem (B), then*

(i) $g_{0\alpha}(\bar{x}; v) \geq 0$ *for all* $v \in T(\bar{x}/\Omega)$ *satisfying the condition* $g_{i\alpha}(\bar{x}; v) \leq 0$ $\forall i \in J(\bar{x}) \overset{\mathrm{df}}{=} \{i \in I / g_i(\bar{x}) = 0\}$;

(ii) *if* g_i, $i \in I$, *are* α*-subdifferentiable at* $\bar{x}$ *then*

$$C(v, R_0(\bar{x})) \geq 0$$

for all $v \in T(\bar{x}/\Omega)$, *where* $R_0(\bar{x}) \overset{\mathrm{df}}{=} \bigcup_{i \in J(\bar{x}) \cup \{0\}} \delta_\alpha g_i(\bar{x})$.

Proof.

(i) Let $v \in T(\bar{x}/\Omega)$ satisfy $g_{i\alpha}(\bar{x}; v) \leq 0$ $\forall i \in J(\bar{x})$. If $i \in J(\bar{x})$ one has

$$0 \geq \limsup_{\lambda \downarrow 0} \frac{g_i(\bar{x} + \lambda v) - g_i(\bar{x})}{\lambda^\alpha} = \limsup_{\substack{\lambda \downarrow 0 \\ v_k \to v}} \frac{g_i(\bar{x} + \lambda v_k)}{\lambda^\alpha}.$$

So for all sequences $\gamma_k \to 0$ and $v_k \to v$ we may assume $g_i(\bar{x} + \gamma_k v_k) \leq 0$ for k large enough. If $i \in I \setminus J(\bar{x})$ then $g_i(\bar{x}) < 0$ and by continuity we have for mentioned above γ_k, v_k, k $g_i(\bar{x} + \gamma_k v_k) \leq 0$ as well. On the other hand, if $v \in T(\bar{x}/\Omega)$ there are $\alpha_k > 0$ and $x_k \in \Omega$, $x_k \to \bar{x}$ such that $\alpha_k(x_k - \bar{x}) \to v$. Set $v_k \overset{\mathrm{df}}{=} \alpha_k(x_k - \bar{x})$, i.e. $\alpha_k^{-1} v_k + \bar{x} = x_k \in \Omega$. Letting $\alpha_k \to +\infty$ we get for k large enough and $i \in I$ $g_i(\bar{x} + \alpha_k^{-1} v_k) \leq 0$, i.e. $\bar{x} + \alpha_k^{-1} v_k$ is a feasible point. Hence

$$g_{0\alpha}(\bar{x}; v) \geq \limsup_{\substack{v_k \to v \\ \alpha_k \to +\infty}} [g_0(\bar{x} + \alpha_k^{-1} v_k) - g_0(\bar{x})] \alpha_K^\alpha \geq 0.$$

(ii) By (i) we have for $v \in T(\bar{x}/\Omega)$

$$0 \leq \max_{i \in J(\bar{x}) \cup \{0\}} g_{i\alpha}(\bar{x}; v) = \max_{i \in J(\bar{x}) \cup \{0\}} C(v; \partial_\alpha g_i(\bar{x})) = C(v; R_0(\bar{x})). \quad \blacksquare$$

EXAMPLE 3.4. Consider the problem

$$\begin{cases} g_0(x) = |x-1|^{1/2} + |x-2|^{1/2} \operatorname{sgn}(x-2) \to \min, \\ g_1(x) = |x-1|^{1/2} - 1 \leq 0. \end{cases}$$

The solution of this problem is evidently $\bar{x} = 1$, $g_0(\bar{x}) = -1$. Direct calculations supply $g_{0\,1/2}(1; v) = |v|^{1/2}$, $\partial_{1/2} g_0(1) = \{|ax|^{1/2} \operatorname{sgn}(ax)/|a| \leq 1\}$ and $\partial_{1/2} g_0(1) = \partial_{1/2} g_1(1)$. Thus 0 belongs to $\partial_{1/2} g_0(1) \cup \partial_{1/2} g_1(1)$.

On the other hand, functionals g_0, g_1 are not Lipschitz at $\bar{x} = 1$. So Clarke's necessary condition cannot be applied. In general, the formerly known necessary conditions based on linear approximations are not effectively used because of the behaviour of g_0 and g_1 near the point $\bar{x} = 1$.

COROLLARY 3–5. *If $\Omega \subset X$ is convex and compact and $\bar{x}$ is a local solution of (B), then*

(i) $\min_{x \in \Omega} \max_{i \in J(\bar{x}) \cup \{0\}} g_{i\alpha}(\bar{x}; x - \bar{x}) \geq 0$;

(ii) *if* $g_i, i\in I$, *are* α*-subdifferentiable at* $\bar{x}$, *we have*

$$\min_{x\in\Omega} \max_{\varphi\in R_0(\bar{x})} \varphi(x-\bar{x}) \geqslant 0.$$

In "bad" cases the objective functional g_0 may not play any role in the necessary conditions given by Theorems 3.2 and 3.3. So these conditions speak only about the constraints of the problems. To avoid this situation we introduce and consider a class of so-called normal problems.

Problem (B) is said to be *normal* if whenever $\bar{x}$ is a local solution, there exists $u\in T(\bar{x}/\Omega)$ such that $C(u; R(\bar{x})) < 0$, where $R(\bar{x}) = \bigcup_{i\in J(\bar{x})} \partial_\alpha g_i(\bar{x})$.

REMARK 3.6. Suppose that all g_i satisfy the *useful condition* ((U_{10}) in [11], [18]) at $\bar{x}$: there exists $v_0\in T(\bar{x}/\Omega)$ such that $g_i^0(\bar{x}; v_0) < 0$ for all $i\in J(\bar{x})$. Then Problem (B) is normal ($\alpha = 1$). For denoting $\eta(\bar{x}) = \{v\in T(\bar{x}/\Omega)/ g_{i\alpha}(\bar{x}, v) < 0\ \forall i\in J(\bar{x})\}$ we see by definition that Problem (B) is normal if and only if $\eta(\bar{x}) \neq \phi$ for each local solution $\bar{x}$.

PROPOSITION 3.7. If Problem (B) is normal and $\bar{x}$ is a local solution, then

$$\Omega\cap Q \neq \phi,$$

where $Q = \{x/g_i(x) < 0\ \forall i\in J(\bar{x})\}$.

Proof. Suppose $\Omega\cap Q = \phi$. For each $v\in T(\bar{x}/\Omega)$ there is $\{x_k\}\subset\Omega$, $x_k\to\bar{x}$, and $t_k > 0$ such that $t_k(x_k-\bar{x})\to v$. Setting $v_k = t_k(x_k-\bar{x})$, $x_k = t_k^{-1} v_k + \bar{x}\in\Omega$, one sees the existence of $i\in J(\bar{x})$ such that $g_i(x_k)\geqslant 0$. Hence $g_{i\alpha}(\bar{x}; v)\geqslant 0$ and $\eta(\bar{x}) = \phi$. Thus (B) is not normal. ■

THEOREM 3–8. *If in Problem (B)* $Q\subset\Omega$ *and*

$$W(\bar{x}) \overset{\mathrm{df}}{=} \{v\in X/g_{i\alpha}(\bar{x}; v) < 0\ \ \forall i\in J(\bar{x})\} \neq \phi$$

for each local solution $\bar{x}$, *then (B) is normal.*

Proof. Since $\eta(\bar{x}) = W(\bar{x})\cap T(\bar{x}/\Omega)$, if we can show that $W(\bar{x})\subset T(\bar{x}/\Omega)$, i.e. $\eta(\bar{x}) = W(\bar{x})$, then the proof is complete. Now let $v\in W(\bar{x})$. Because g_i, $i\in I$ are locally Hölder, we have for $i\in J(\bar{x})$

$$0 > g_{i\alpha}(\bar{x}; v) = \limsup_{\lambda\downarrow 0} \frac{g_i(\bar{x}+\lambda v)}{\lambda^\alpha} = \limsup_{\substack{\lambda_k\downarrow 0\\ v_k\to v}} \frac{g_i(\bar{x}+\lambda_k v_k)}{\lambda_k^\alpha}.$$

Hence we can choose sequence $\lambda_k\downarrow 0$ and $v_k\to v$ satisfying $g_i(\bar{x}+\lambda_k v_k) < 0$. Then $\bar{x}+\lambda_k v_k\in Q\subset\Omega$. Moreover $\bar{x}+\lambda_k v_k\to\bar{x}$. So, by definition $v\in T(\bar{x}/\Omega)$. ■

References

[1] AUBIN J. P. Contingent derivatives of set-valued maps and existence of solutions to nonlinear inclusions and differential inclusions. *Math. Anal. Appl.* Part A. Advances in Math. Supplementary study, **7A** (1981), 159–229.

[2] CLARKE F. H. Generalized gradients and applications. *Trans. Amer. Math. Soc.* **205** (1975), 247–262.

[3] CLARKE F. H. A new approach to Lagrange multipliers. *Math. Oper. Research* **1** (1976), 165–174.

[4] CLARKE F. H. The generalized problem of Bolza. *SIAM J. Control Opt.* **14** (1976), 682–699.

[5] CLARKE F. H. The maximum principle under minimal hypotheses. *SIAM J. Control Opt.* **14** (1976), 1078–1091.

[6] CLARKE F. H. The Erdmann condition and Hamiltonian inclusions in optimal control and the calculus of variations. *Can. J. Math.* **32** (1980) 2, 494–509.

[7] CLARKE F. H. Generalized gradients of Lipschitz functionals. *Advances in Math.* **40** (1981), 52–67.

[8] DOLECKI S., KURCYUSZ S. On ϕ-convexity in extremal problems. *SIAM J. Control Opt.* **16** (1978), 277–300.

[9] EKELAND I. On the variational principle. *J. Math. Anal. Appl.* **47** (1974), 324–353.

[10] HALKIN H. Mathematical programming without differentiability. "Calculus of Variations and Control Theory", D. L. Russell ed., New-York, 1976, 279–287.

[11] HIRIART-URRUTY J. B. On optimality conditions in nondifferentiable programming. *Math. Programming* **4** (1978), 73–86.

[12] LEBOURG G. Generic differentiability of Lipschitzian functions. *Trans. Amer. Math. Soc.* **256** (1979), 125–144.

[13] PHAM HUY DIEN Some results on locally Lipschitzian mappings. *Acta Math. Vietnamica*, **6** (1981) 2.

[14] PHAM HUU SACH Les points reguliers des applications multivoques et la commandabilité dans les systèmes discrets. *Analyse Appliqué dans te Informatique*, Uni. Bordeaux I, U.E.R. de Math. Infor.. 1980, n. 8022, 1–47.

[15] PHAN QUOC KHANH A concept of optimality of degree p in the observation theory and optimal observability for convex functionals. *Control and Cybernetics* **8** (1979) 1, 43–53.

[16] PHAN QUOC KHANH A formula for extremum involving ϕ-convexity. *Control and Cybernetics* **13** (1984) 4, 305–311.

[17] POURCIAU B. H. Analysis and optimization of Lipschitz continuous mappings. *J. Opt. Theory Appl.* **22** (1977), 311–351.

[18] STRODIOT J. J., HIEN NGUYEN V. Kuhn-Tucker multipliers and nonsmooth programs. *Math. Prog. Study* **19** (1982), 222–240.

[19] WARGA J. Derivate containers, inverse functions and controlability. "Calculus of Variations and Control Theory", D. L. Russell ed., New-York, 1976, 13–46.

Received, May 1984

Subgradienty Höldera i ich zastosowania w optymalizacji

Wprowadzono subgradienty Höldera jako uogólnione Φ-subgradienty z klasą funkcjonałów Φ, dla których można stosować analizę hölderowską, i rozważono ich właściwości potrzebne przy zastosowaniu w zadaniach optymalizacji. Wyprowadzono warunki konieczne optymalności i warunki normalności dla zadań z funkcjonałami lokalnie hölderowskimi.

Praca zawiera wyniki związane z przyjęciem do aproksymacji w słabym sensie klasy funkcjonałów złożonej z prostych funkcjonałów nieliniowych zamiast funkcjonałów z przestrzeni dualnej X^*. Wyniki uzyskano bez założenia, że występujące w sformułowaniu zadania funkcjonały spełniają warunki Lipschitza.

Субградиенты Гольдера и приложения в оптимизации

Мы вводим понятие субградиент Гольдера как обобщенный φ-субградиент с классом φ функционалов, обладающих свойствами, связанными с анализом Гольдера и рассмотрим его нужные свойства для исследования задач оптимизации. Мы доказали необходимые условия оптимизации и условия нормальности для задач с функционалами Гольдера.

Новинка результатов состоит в том, что мы взяли довольно простых нелинейных функционалов вместо сопряженного пространства x^* для аппроксимации функционалов и что функционалы, определяющие задачи не удовлетворяют условиям Липшица.