On Functions, Representable as a Difference of Two Convex Functions, and Necessary Conditions in a Constrained Optimization

A. Shapiro, Y. Yomdin

Department of Mathematics
Ben-Gurion University of the Negev
Beer-Sheva, Israel

The class of functions, representable as a difference $f_1 - f_2$ of two convex functions, is studied. We show, that this class is closed under superposition and taking max (min) of a finite number of functions. We prove, that a supremum (infimum) of any bounded in C^2-norm family of C^2-smooth functions belongs to this class. Actually, we prove, that f can be represented as a supremum (infimum) of a bounded in C^2-norm family of C^2-smooth functions if and only if $f = f_1 - f_2$, where f_1 and f_2 are convex and f_2 (resp. f_1) - C^2-smooth. We give a simple description of such functions in terms of the generalized gradient. (An important example of a function of this type gives a square of the distance to a closed subset $T \subset R^n$, or the distance itself on $R^n \setminus T$).

We consider positively homogeneous functions, representable as a difference of two convex positively homogeneous functions, and study the isometric correspondence between this class and the normed space of pairs of convex sets. Finally, we apply these results to obtain necessary conditions for an extremum in a constrained optimization.

Introduction

Real functions representable as the difference of two convex functions were studied in a number of publications ([1], [2], [6], [12]). The importance of this class, denoted below by DC, can be explained by the following facts: functions of DC preserve many local properties of convex ones (say the differential properties) and, on the other hand, this class forms a linear space that is much wider than the cone of convex functions.

The function of one variable belongs to DC if and only if it is lipschitzian and its first derivative (defined almost everywhere) is of bounded variation on every closed bounded interval. However, relatively little is known in the case of several variables. Any piecewise-linear function can be represented as the difference of convex piecewise-linear functions ([1], [2]). The class DC is closed under superposition and then under any algebraic operation, whenever defined ([6]).

The aim of this paper is to show that the class DC is closely related to the operation of taking supremum (infimum) of a family of functions. We show that this class is closed under taking max (min) of a finite number of functions. We prove that sup (inf) of any bounded in C^2-norm family of C^2-smooth functions can be represented as the difference $g_1 - g_2$ of two convex functions with g_2 (g_1) C^2-smooth and we give a simple description of such functions in terms of the generalized gradient.

Then we consider the class DCH of positively-homogeneous functions, representable as the difference of two convex positively homogeneous functions and transfer the results above to this class.

The linear space DCH (with supremum norm on the unit shpere) is naturally isometric to the normed space of pairs of convex sets introduced in [10]. We apply this result to obtain necessary conditions for an extremum in problems of minimax optimization. These conditions can be viewed as a natural generalization of the necessary conditions for the quasidifferentiable functions of Pshenichnyi [9].

I. Let D be an open domain in R^n. We shall be dealing with functions $f:D \to R$ which will be assumed locally Lipschitz. That is, for each $x \in D$ there exists a neighborhood U of x and a constant K such that

(1.1) $$|f(x_2) - f(x_1)| \leq K\, |x_2 - x_1|$$

for all points x_1 and x_2 of U ($|.|$ denotes the Euclidean norm). It is known that such functions are differentiable almost everywhere.

We denote $\nabla f(x)$ the gradient of f at x (whenever defined) and $\partial_c f(x)$ the Clarke generalized gradient of f at x, that is the convex hull of the set of limits of the form $\lim \nabla f(x + h_i)$, where $h_i \to 0$ as $i \to \infty$. It follows that $\partial_c f(x)$ is a nonempty convex compact set ([4]), and the mean value theorem holds, i.e., $f(x + y) - f(x) = (a,y)$, where $a \in \partial_c f(x + \Theta y)$, $0 \leq \Theta \leq 1$, and interval $[x, x + y]$ belongs to D, ([8]).

Definition 1.1

A function f will be called d.c. at $x \in D$ if there exists a convex neighborhood U of x and a pair of convex functions $f^+(x)$, $f^-(x)$ on U such that

(1.2) $$f(x) = f^+(x) - f^-(x)$$

for all $x \in U$. When f is d.c. at every point x of D, it will be said to be d.c. on D. The linear space of such functions will be denoted by DC(D). A d.c. function f(x) will be called a subconvex (subconcave) if $f^-(x)$ ($f^+(x)$) is twice continuously differentiable on U in at least one of representations (1.

Theorem 1.1 (Hartman, [6]).

Let $f \in DC(D)$ and E be an open or closed convex subset of D. Then there exists a pair of convex functions $f^+(x)$, $f^-(x)$ defined on E such that equality (1.2) holds for all points x of E.

Let us now consider subconvex functions. The following definition will be useful.

Definition 1.2.

Let G(x) be a multimapping from $B \subset R^n$ to R^n. That is, for each $x \in B$, G(x) is a subset of R^n. We say that G(x) is locally lower semilipschitz (l.l.s.l.) if for every point $x \in B$ there exists a neighborhood U of x and a constant $K \geq 0$ such that

$$(1.3) \qquad (a_2 - a_1, x_2 - x_1) \geq -K|x_2 - x_1|^2$$

for all $x_1, x_2 \in B \cap U$ and all $a_1 \in G(x_1)$, $a_2 \in G(x_2)$. A family of multimappings $F(x,t) : B \times T \to R^n$ is said to be l.l.s.l. on T if the neighborhood U and constant K can be chosen independently of a parameter t. For mappings the definition is analogical.

Theorem 1.2.

The following conditions are equivalent:

(a) The function f(x) is subconvex

(b) The mapping $\nabla f(x)$ (whenever defined) is l.l.s.l.

(c) The multimapping $\partial_c f(x)$ is l.l.s.l.

Proof

Assume that (a) holds. Then $\nabla f(x) = \nabla f^+(x) - \nabla f^-(x)$ and by the convexity the mapping $\nabla f^+(x)$ is monotone, that is $(\nabla f^+(x_2) - \nabla f^+(x_1), x_2 - x_1) \geq 0$, ([11]). By the definition $f^-(x)$ is twice continuously differentiable and thus $\nabla f^-(x)$ is locally Lipschitz. Hence we obtain

$$(1.4) \qquad (\nabla f(x_2) - \nabla f(x_1), x_2 - x_1) \geq -(\nabla f^-(x_2) - \nabla f^-(x_1), x_2 - x_1) \geq$$

$$-|\nabla f^-(x_2) - \nabla f^-(x_1)|.\ |x_2 - x_1| \geq -K|x_2 - x_1|^2.$$

And this proves (b).

That (b) implies (c) is evident from the definitions.

Now suppose that $\partial_c f(x)$ is 1.1.s.1. We want to show that the function $h(x) = f(x) + K(x,x)$ is convex on U. Let $x \in U$ and $a \in \partial_c f(x)$. Then we have by the mean value theorem

$$(1.5) \qquad h(x + y) - h(x) - (a + 2Kx, y) =$$

$$f(x + y) - f(x) - (a, y) + K(y, y) = (b - a, y) + K(y, y),$$

where $\quad b \in \partial_c f(x + \Theta y)\ ,\ 0 \leq \Theta \leq 1.$

It follows from the definition of 1.1.s.1. that

$$(1.6) \qquad (b - a, y) \geq -K\Theta\,|y|^2$$

and thus the inequality

$$(1.7) \qquad h(x + y) - h(x) \geq (c, y)$$

(where $c = a + 2Kx$) holds for all y such that $x + y \in U$. Then h is a supremum of linear functions and hence convex. □

Remark 1.1

Actually we proved that function $f^{-}(x) = K(x,x)$ depends only on the constant K.

Remark 1.2.

An extension of the generalized gradient to infinite dimensional spaces can be found in [5]. The mean value theorem is still valid and the equivalence between (a) and (c) can be easily established for any real Hilbert space.

Corollary 1.1

The function $f(x)$ is subconvex and subconcave iff $f(x)$ is differentiable and the mapping $\nabla f(x)$ is locally Lipschitz. In particular if $f(x)$ is twice continuously differentiable on D then $f(x)$ is subconvex and subconcave.

The subconvex functions preserve many important properties of convex ones. In this connection the concept of quasidifferentiability of Pshenichnyi will be relevant, ([9]).

Definition 1.2.

A function $f(x)$ is said to be quasidifferentiable at x if the directional derivative

$$f'(x, y) = \lim_{t \to 0^+} \frac{f(x + ty) - f(x)}{t}$$

exists for every direction $y \in R^n$ and $f'(x, y) = s(y|A)$. Here $s(y|A) = \max\{(y, z) \,|\, z \in A\}$ is the support function of the convex compact set A, that will be denoted $\partial f(x)$ and called the subdifferential of f at x. We say that f is regularly quasidifferentiable at x if it is quasidifferentiable and uniformly differentiable in all directions at this point.

Proposition 1.1.

A subconvex function is regularly quasidifferentiable and $\partial f(x) = \partial_c f(x)$ at every point $x \in D$. The proof follows by convex analysis, [11].

Another important characteristic of convexity is that a supremum of convex functions is convex.

Let $f(x, t) : D \times T \to R$ be a family of locally Lipschitz functions. We consider the following function:

$$f(x) = \sup \{f(x, t) \mid t \in T\}. \tag{1.8}$$

We suppose f to be finite for all $x \in D$.

Theorem 1.3

Let the family of mappings $\nabla_x f(x, t)$ be l.l.s.l. on T. Then

(1) $f(x)$ is subconvex on D

(2) $f(x)$ is locally Lipschitz

(3) $f(x)$ is regularly quasidifferentiable and

(4) $\partial f(x) = \partial_c f(x)$ at every $x \in D$

If, in addition, T is a compact topological space and $f(x, t)$ is continuous on $D \times T$ then:

(5) $\partial f(x)$ is the topological closure of the convex hull of $\{\partial_x f(x, t) \mid t \in S(x)\}$, where $S(x) = \{t \mid f(x) = f(x, t)\}$.

Proof.

By remark 1.1 we have the following representation of $f(\cdot,t)$ on U

$$f(x, t) = f^+(x, t) - f^-(x) \tag{1.9}$$

with the same C^2-smooth function $f^-(x)$ for all $t \in T$. Then

$$f(x) = f^+(x) - f^-(x) \tag{1.10}$$

where $f^+(x) = \sup\{f^+(x, t) \mid t \in T\}$ is convex on U as a supremum of convex ones. Thus we have proved (1). Then (2), (3) and (4) follow from convex analysis (see [11] and proposition 1.1). Property (5) is well known for convex functions (e.g. [7]) and we obtain

$$\partial f(x) = \partial f^+(x) - \{\nabla f^-(x)\} = \overline{\text{conv}}\{\partial_x f^+(x, t) \mid t \in S(x)\} - \{\nabla f^-(x)\} \tag{1.11}$$

$$= \overline{\text{conv}}\{\partial_x f^+(x, t) - \{\nabla f^-(x)\} \mid t \in S(x)\} = \overline{\text{conv}}\{\partial_x f(x, t) \mid t \in S(x)\} \quad \square$$

Remark 1.3.

Theorem 1.3 can be easily extended to any real Hilbert space if we replace the l.l.s.l. of $\nabla_x f(x, t)$ by the l.l.s.l. of the family of multimappings $\partial_c f(x, t)$. (For finite dimensional spaces these two conditions are equivalent, see remark 1.2).

The function $f(x)$ has been the subject of much investigation (see [4], [9]). A theorem that is similar to the corollary below is given in [3], (Lemma 2.1).

Corollary 1.2.

Let $f(\cdot, t)$ be Gateaux differentiable and $\nabla_x f(x, t)$ be locally Lipschitz on T. Then properties (1)-(4) and (for a suitable hypothesis on T) (5) follow.

For an infinum of subconcave functions the analogical results can be obtained.

Corollary 1.3.

The function $f(x)$ is subconvex (subconcave) if and only if it can be locally represented as a supremum (an infinum) of a family of twice continuously differentiable functions bounded in the C^2-norm.

Corollary 1.4.

Let T be some subset of R^n (of a real Hilbert space H, see remark 1.3). Consider $d_T(x) = \inf \{|x - t| : t \in T\}$ the distance function to T. Then $d_T^2(x)$ is subconcave at any $x \in R^n$ and $d_T(x)$ is subconcave at any $x \in R^n \setminus \bar{T}$, ($\bar{T}$ is the topological closure of T in R^n). Furthermore if T is compact then

$$\partial d_T(x) = \text{conv}\, \{(x - t) / |x - t| \,:\, t \in S(x)\} \qquad \text{where}$$

$$S(x) = \{t \mid t \in T,\ d_T(x) = |\, x - t \,|\}.$$

The proof follows from corollary 1.2 for $f(x, t) = |x - t|^2$ and $f(x, t) = |x - t|$ respectively. □

Hartman [6] showed that the superposition of d.c. functions is a d.c. function. We give a simple proof of this fact and show how representation (1.2) can be constructec

Definition 1.3

We say that the mapping $F = (f_1, \dots f_m) : D \to R^m$ is d.c. (on D) and write $F \in DC(D)$ if each coordinate function f_i, $i = 1, \dots, m$, belongs to $DC(D)$.

Theorem 1.4.

Let $D_1 \subseteq R^n$, $D_2 \subseteq R^m$ be open sets and let $F_1 : D_1 \to D_2$, $F_2 : D_2 \to R^k$ be d.c. mappings. Then $F_2 \circ F_1 : D_1 \to R^k$ is d.c. mapping.

$F_2 \circ F_1$ is d.c. mapping.

In particular $DC(D)$ is closed under any algebraic operation (whenever defined).

Proposition 1.2.

Let $f_i \in DC(D)$ and $f_i = f_i^+ - f_i^-$ (on U), $i = 1, \dots, m$. Then $f = \max\{f_1 \dots, f_{\text{I}}$

Proof.

By definition, we need to prove that each coordinate function of $F_2 \circ F_1$ belongs to $DC(D_1)$. Since F_2 is d.c. mapping, its coordinate function can be represented as difference of convex functions. Then it is sufficient to prove the following lemma:

Lemma 1.1:

Let $F = (f_1,\dots,f_m) : D_1 \to D_2$ be a d.c. mapping and let $g : D_2 \to R$ be a convex function. Then $g(f_1,\dots,f_m)$ belongs to $DC(D_1)$.

Proof.

Let $x \in D_1$ and $y = F(x) \in D_2$. The convex function g can be represented in some neighborhood U_2 of y as the supremum of a bounded family of linear functions: $g = \sup \ell_t$, with $\ell_t = a_{0t} + a_{1t} y_1 + \dots + a_{mt} y_m$ ($y_1,\dots y_m$ are the standard coordinates in R^m). Let $M = \sup_{i,t} |a_{it}|$ and let $f_i(x) = f_i^+(x) - f_i^-(x)$ in some neighborhood U_1 of x such that $F(U_1) \subseteq U_2$. Then $\ell_t(f_1,\dots,f_m) = a_{0t} + \sum_{i=1}^m a_{it} f_i^+ - \sum_{i=1}^m a_{it} f_i^- = [a_{0t} + \sum_{i=1}^m (M + a_{it}) f_i^+ + \sum_{i=1}^m (M - a_{it}) f_i^-] - M(\sum_{i=1}^m f_i^+ + f_i^-) = g_t^+ - g^-$, with g_t^+ and g^- convex and g^- independent of t. Then $g(f_1,\dots,f_m) = \sup_t \ell_t(f_1,\dots,f_m) = \sup_t (g_t^+ - g^-) = \sup_t g_t^+ - g^- = g^+ - g^-$, i.e., $g(f_1,\dots f_m) \in DC(D_1)$. □

By theorem 1.4. and corollary 1.1. we have:

Corollary 1.5.

Let $F_1 : D_1 \to D_2$ be d.c. and $F_2 : D_2 \to R^k$ be C^2-smooth mappings. Then $F_2 \circ F_1$ is d.c. mapping.

In particular $DC(D)$ is closed under any algebraic operation (whenever defined).

Proposition 1.2.

Let $f_i \in DC(D)$ and $f_i = f_i^+ - f_i^-$ (on U), $i = 1,\dots,m$. Then $f = \max\{f_1 \dots, f_I$

belongs to DC(D) and $f = f^+ - f^-$ (on U), where $f^- = \sum_{i=1}^{m} f_i^-$ and $f^+ = \max\{f_i^+ + \sum_{j\neq i} f_j^- \mid i = 1,\dots,m\}$.

Remark 1.4.

Proposition 1.2 can be extended for an infinite sequence $f_1, f_2, \dots$ if the series $\sum_{i=1}^{\infty} f_i^-(x)$ converges for any $x \in U$.

Remark 1.5.

The following example shows that DC(D) is not closed under taking sup (inf) of an infinite family of functions:
Let $f_n = -\left|x + \frac{1}{n}\right|$. Then clearly, the variation of the first derivative of $f = \sup\{f_n \mid n = 1, 2, \dots\}$ is unbounded in any neighborhood of zero, i.e., f is not d.c. at zero.

Corollary 1.6.

Let functions $f_1(x),\dots,f_m(x)$ be subconvex at $\bar{x} \in R^n$, $g(y)$ be convex in some neighborhood of $\bar{y} = (f_1(\bar{x}),\dots,f_m(\bar{x}))$ and $\partial g(y) \subseteq R_+^m = \{y \mid y > 0\}$. Then $f(f_1(x),\dots,f_m(x))$ is subconvex at $\bar{x}$.

Proof.

From the upper semicontinuity of $\partial g(y)$ follows that $\partial g(y) \subset R_+^m$ for all y in some neighborhood of $\bar{y}$. Then the linear functions ℓ_t in lemma 1.1. can be taken with positive coefficients $a_{1t},\dots,a_{mt}$ and $\ell_t(f_1,\dots,f_m) = [a_{0t} + \sum_{i=1}^{m} a_{it} f_i^+ + \sum_{i=1}^{m} (M - a_{it}) f_i^-] - M(\sum_{i-1}^{m} f_i^-) = g_t^+ - g^-$, where $M = \sup_{i,t} a_{it}$ and $g^- = M(\sum_{i=1}^{m} f_i^-)$. Thus g^- is C^2-smooth as a sum of C^2-smooth functions f_i^-. □

II

Here we consider positively homogeneous functions $f: R^n \to R$, $f(tx) = tf(x)$, $t \geq 0$. Such function is defined by its values on the unit sphere $S^{n-1} = \{x: |x| = 1\}$. The restriction $f/_{S^{n-1}}$ will be denoted by $\hat{f}$. By P_x we denote a tangent space of S^{n-1} at x.

We say that f is d.c.h., if it can be represented as the difference of two convex positively homogeneous functions. The linear space of d.c.h. functions we denote $DCH(R^n)$ or simply DCH.

Theorem 2.1.

A positively homogeneous function f belongs to DCH if and only if the restriction f/P_x is a d.c. function at zero for every $x \in S^{n-1}$.

Proof.

The restriction of a convex function at an affine subspace is convex and the necessity follows from the definitions.

Now suppose that $f/_{P_x}$ is d.c. at zero, i.e. there exists a neighborhood $U \subseteq P_x$ of zero and a convex function $g(y)$ on U such that $f/_{P_x} + g$ is convex on U.

Furthermore the neighborhood can be chosen in such a way that $g(y) = \sup_t \{a_t + \ell_t(y)\}$ on U, where $\{\ell_t\}$ is the bounded family of linear functions on P_x. Consider the convex positively homogeneous function (on R^n)

$$G(\alpha x + y) = \sup_t \{\alpha a_t + \ell_t(y)\}.$$

It is clear that $G_{/P_x} = g$ on U and we obtain that $f + G$ is convex in some neighborhood of x. Then by the compactness of S^{n-1} there exist points $x_1, \ldots, x_m \in S^{n-1}$ and convex positively homogeneous functions $G_1, \ldots G_m$ such that $f + G_i$ is convex in the neighborhood U_i of x_i, $i = 1, \ldots m$, and

$S^{n-1} \subset \bigcup_{i=1}^{m} U_i$. Thus $f + \sum_{i=1}^{m} G_i$ is convex and the proof is complete. □

This theorem enables us to apply the results of the previous section to the homogeneous functions.

Corollary 2.1.

A positively homogeneous function $f: R^2 \to R$ belongs to $DCH(R^2)$ if and only if the restriction $\hat{f}$ on the circle S^1 is Lipschitz and its first derivative (with respect to the standard coordinate φ on S^1) is of bounded variation.

Corollary 2.2.

Let f be positively homogeneous function such that $\hat{f}$ is twice continuously differentiable with respect to the local coordinate systems of S^{n-1}. Then f belongs to DCH.

Corollary 2.3.

Define the supremum norm in DCH:

$$||f|| = \sup \{|f(x)| \; |x \in S^{n-1}\}.$$

Then DCH is isometric to a dense subspace of $C(S^{n-1})$, that is the space of continuous functions on S^{n-1} (with sup norm).

Corollary 2.4.

The superposition of d.c.h. functions is d.c.h. In particular max or min of finite number of d.c.h. functions is d.c.h.

Now we translate the results above to the language of convex sets. We denote Π the class of compact convex sets in R^n. The following results can be found in [11].

Proposition 2.1.

(a) The support function

$$s(x|A) = \sup \{(x, y)|y \in A\}$$

of a set A is a convex, continuous and positively homogeneous function from R^n into R.

(b) Conversely, if $h(x)$ is a convex function from R^n into R which is positively homogeneous, then it is the support function of a certain set $A \in \Pi$. This set is unique and given by $A = \{y|(x, y) \leq h(x)$ for all $x \in R^n\}$.

(c) For convex sets A_1 and A_2 and $\lambda, \mu \geq 0$ we have

$$s(\cdot|\lambda A_1 + \mu A_2) = \lambda s(\cdot|A_1) + \mu s(\cdot|A_2).$$

(d) For sets $A_1, A_2 \in \Pi$ the inequality $s(\cdot|A_1) \geq s(\cdot|A_2)$ holds if and only if $A_1 \supseteq A_2$

We have from proposition 2.1 and the definition that to each $f \in DCH(R^n)$ correspond a pair of sets $A_1, A_2 \in \Pi$ such that

(2.1) $$f = s(\cdot|A_1) - s(\cdot|A_2)$$

Consider an equivalence relation $\sim$ that is defined on $\Pi \times \Pi$ by stating that $(A_1, A_2) \sim (C_1, C_2)$ if $A_1 + C_2 = A_2 + C_1$. In other words the pairs (A_1, A_2) and (C_1, C_2) represent the same function f in (2.1). The equivalence class containing (A_1, A_2) will be denoted $\langle A_1, A_2\rangle$. The linear space $\tilde{\Pi}$ is taken to be the quotient space $\Pi \times \Pi/\sim$, where algebraic operations in $\tilde{\Pi}$ are imposed by the linear space DCH, i.e.,

(2.2) $$\langle A_1, A_2\rangle + \langle C_1, C_2\rangle \equiv \langle A_1 + C_1, A_2 + C_2\rangle$$

and if $\lambda \geq 0$, then

(2.3) $\lambda \langle A_1, A_2\rangle \equiv \langle \lambda A_1, \lambda A_2\rangle$

while if $\lambda < 0$, then

(2.4) $\lambda \langle A_1, A_2\rangle \equiv \langle -\lambda A_2, -\lambda A_1\rangle$.

The number

$$\delta(A_1, A_2) = \sup\{d(x, A_2) \mid x \in A_1\}$$

is called the deviation of the set A_1 from the set A_2. The maximum of the deviations $\delta(A_1, A_2)$ and $\delta(A_2, A_1)$:

$\rho(A_1, A_2) = \max\{\delta(A_1, A_2), \delta(A_2, A_1)\}$ is called the Hausdorff distance between the sets A_1 and A_2.

We define a norm $||\cdot||$ on $\tilde{\Pi}$ by

(2.5) $||\langle A_1, A_2\rangle|| = \rho(A_1, A_2)$

The normed space $\tilde{\Pi}$ was introduced by Radstrom [10]. It easily can be shown that $\pi(A) \equiv \langle A, 0\rangle$ is an isometric embedding of Π (with Hausdorff metric) into $\tilde{\Pi}$, [10].

Theorem 2.2

Let $A_1, A_2 \in \Pi$ and $A_1 \not\subseteq A_2$. Then

(2.6) $\delta(A_1, A_2) = \sup\{s(x|A_1) - s(x|A_2) \mid x \in S^{n-1}\}$

Proof.

Let $B = \{x \mid |x| \leq 1\}$ be a unit ball in R^n. We observe that

$$s(x|A_2 + \lambda B) = s(x|A_2) + \lambda s(x|B) = s(x|A_2) + \lambda$$

for $\lambda > 0$ and $|x| = 1$. From proposition 2.1 (d) we have that

$A_1 \subseteq A_2 + \lambda B$, $\lambda \geq 0$, if and only if $s(x|A_1) - s(x|A_2 + \lambda B) \leq 0$ or

$s(x|A_1) - s(x|A_2) \leq \lambda$ for $|x| = 1$.

Noting that $\delta(A_1, A_2) = \inf\{\lambda | A_1 \subseteq A_2 + \lambda B, \lambda \geq 0\}$ we complete the proof. □

By theorem 2.2 we obtain:

Corollary 2.5.

(2.7) $\rho(A_1, A_2) = \sup\{|s(x|A_1) - s(x|A_2)|, x \in S^{n-1}\}$

i.e. the correspondence

(2.8) $\langle A_1, A_2\rangle \longleftrightarrow s(x|A_1) - s(x|A_2)$

between $\tilde{\Pi}$ and DCH (with sup norm on S^{n-1}) is isometric.

Corollary 2.6

$\tilde{\Pi}$ is isometric to a dense subspace of $C(S^{n-1})$. (See corollary 2.3).

We denote $c(x|A)$ the contact set of a point x:

$c(x|A) = \{y \in A \mid (x, y) = s(x|A)\}$.

The contact set is known to be the subdifferential of $s(\cdot|A)$ at x, i.e.

$\partial s(x|A) = c(x|A)$, ([7]). Then using theorem 1.3 and theorem 2.1 we obtain

Corollary 2.7

Let $f_t = s(\cdot|A_t) - s(\cdot|B_t) \in DCH$, $t \in T$.

Suppose that

(a) $c(x|B_t) = \{F_t(x)\}$ is singleton for all $x \in R^n$ and $t \in T$.

(b) The mapping $F_t(x)$ is Lipschitz in some neighborhood of zero of the tangent space P_x at every $x \in S^{n-1}$ and $t \in T$, such that the neighborhood and constant K are independent of t.

Then if function $f(x) = \sup\{f_t(x) \mid t \in T\}$ is finite at every $x \in R^n$ it belongs to DCH.

III.

In this section we apply the results above to obtain necessary conditions for an extremum in constraint optimization.

Definition 3.1.

A function $f: D \to R$ is said to be pseudosmooth at $x \in D$ if the directional derivative $f'(x, y)$ at x exists for every direction $y \in R^n$ and

$$(3.1) \qquad f'(x, \cdot) = s(\cdot \mid A_1) - s(\cdot \mid A_2) ,$$

i.e. $f'(x, \cdot) \in$ DCH. The sets $A_1, A_2 \in \Pi$ will be denoted $\partial^+ f(x)$, $\partial^- f(x)$ respectively and the class $\langle \partial^+ f(x), \partial^- f(x) \rangle \in \tilde{\Pi}$ will be called the pseudogradient of f at x. If, in addition, f is uniformly differentiable in all directions we say that f is regularly pseudosmooth at x. The mapping $F = (f_1, \ldots, f_m): D \to R^m$ will be said (regularly) pseudosmooth at x if all coordinate functions f_i, $i = 1, \ldots, m$, are (regularly) pseudosmooth at x.

It is clear that quasidifferentiable and d.c. functions are pseudosmooth and for a quasidifferentiable function f we have

$$\langle \partial^+ f(x), \partial^- f(x) \rangle \equiv \langle \partial f(x), \{0\} \rangle .$$

Theorem 3.1

Let $F: D_1 \to D_2 \subseteq R^m$ be (regularly) pseudosmooth mapping at $x \in D_1$ and $g: D_2 \to R$ be regularly pseudosmooth function at $y = F(x) \in D_2$. Then function $h = g \circ F: D_1 \to R$ is (regularly) pseudosmooth at x. Furthermore if $0 \leq \alpha_i \leq M$, $0 \leq \beta_i \leq M$, $i = 1,\ldots,m$, for all $\alpha = (\alpha_1,\ldots,\alpha_m) \in \partial^+ g(y)$ and all $\beta = (\beta_1,\ldots,\beta_m) \in \partial^- g(y)$ then

(3.2) $$<\partial^+ h(x), \partial^- h(x)> \equiv$$

$$<\operatorname{conv} \bigcup_{\alpha\in\partial^+ g(y)} \{ \sum_{i=1}^{m} [\alpha_i \partial^+ f_i(x) + (M - \alpha_i)\partial^- f_i(x)]\} + M \sum_{i=1}^{m} \partial^- f_i(x),$$

$$\operatorname{conv} \bigcup_{\beta\in\partial^- g(y)} \{ \sum_{i=1} [\beta_i \partial^+ f_i(x) + (M - \beta_i) \partial^- f_i(x)]\} + M \sum_{i=1}^{m} \partial^- f_i(x)>$$

The proof is similar to the proof of lemma 1.1 and utilizes the chain rule for directional derivatives.

Remark 3.1.

If g is Frechet differentiable at y then

(3.3) $$<\partial^+ h(x), \partial^- h(x)> \equiv \sum_{i=1}^{m} \alpha_i <\partial^+ f_i(x), \partial^- f_i(x)> ,$$

where $(\alpha_1,\ldots,\alpha_m) = \nabla g(y)$.

Remark 3.2.

Let $h(x) = \max\{f_i(x) | i = 1,\ldots,m\}$ and $S(x) = \{i | h(x) = f_i(x)\}$.

Then

(3.4) $$<\partial^+ h(x), \partial^- h(x)> \equiv <\operatorname{conv} \bigcup_i \{\partial^+ f_i(x) + \sum_{j\neq i} \partial^- f_j(x)\}, \sum_j \partial^- f_j(x)>$$

where $i, j \in S(x)$ (see proposition 1.2).

For minimum the formula is similar.

Now we consider the following problem:

$$(\Omega)\qquad \begin{cases} \text{Minimize } f_0(x) \text{ subject to} \\ f_i(x) \leq 0\ ,\ i = 1,\dots,m \end{cases}$$

We suppose that $f_0, f_1, \dots, f_m$ are locally Lipschitz and pseudosmooth at every $x \in R^n$.

Let $h(x) = \max\{f_i(x) \mid i = 1,\dots,m\}$ and $X = \{x \mid h(x) \leq 0\}$. At each point $x \in X$ we consider a d.c.h. function $f'(x,\cdot)$ that is defined:

$$(3.5)\qquad f'(x,\cdot) = \begin{cases} \max\{f_0'(x,\cdot)\ ,\ h'(x,\cdot)\}\ , & h(x) = 0 \\ f_0'(x,\cdot)\ , & h(x) < 0. \end{cases}$$

(cf. [8], p. 969). We note that $h'(x,\cdot) = \max\{f_i'(x,\cdot) \mid i \in S(x)\}$.

The pseudogradient of $f'(x,\cdot)$ at zero will be denoted $<\partial^+ f'(x), \partial^- f'(x)>$. (i.e. $\partial^+ f'(x) = \operatorname{conv}\{\partial^+ f_0(x) + \partial^- h(x), \partial^- f_0(x) + \partial^+ h(x)\}$ and

$\partial^- f'(x) = \partial^- f_0(x) + \partial^- h(x)$ if $h(x) = 0$, as follows from (3,4) and the definition).

Definition 3.2.

We say that point x is stationary for problem (Ω) if $x \in X$ and

$$(3.6)\qquad \partial^- f'(x) \subseteq \partial^+ f'(x).$$

It follows from proposition 2.1 (d) that:

Proposition 3.1.

The point $x \in X$ is stationary if and only if $f'(x, y) \geq 0$ for all $y \in R^n$.

Theorem 3.2. (necessary conditions).

If $\bar{x}$ solves (Ω) locally, then $\bar{x}$ is stationary.

Proof.

Suppose that $h(\bar{x}) = 0$ and $\bar{x}$ solves (Ω) locally. By standard argument of the feasible directions method we obtain that if $h'(\bar{x}, y) < 0$ then $f_0'(\bar{x}, y) \geq 0$. In other words $\max\{f_0'(\bar{x}, y), h'(\bar{x}, y)\} \geq 0$ for all $y \in R^n$ and using proposition 3.1 we complete the proof. If $h(\bar{x}) < 0$ the proof is similar. □

Theorem 3.3. (sufficient conditions).

Suppose that $f_0(x)$ is <u>regularly</u> pseudosmooth at $\bar{x} \in X$ and there exists an open neighborhood W of $\partial^- f(\bar{x})$ such that $W \subseteq \partial^+ f(\bar{x})$. Then $\bar{x}$ solves (Ω) locally.

Proof.

It follows that $f'(\bar{x}, y) > 0$ for all $y \in S^{n-1}$. By compactness of S^{n-1} and continuity of $f'(\bar{x}, \cdot)$ we actually have that $f'(\bar{x},y) \geq \varepsilon > 0$ for all $y \in S^{n-1}$. Thus $f_0'(\bar{x}, y) \geq \varepsilon$, $y \in S^{n-1}$, in each direction y such that $h(\bar{x}, y) \leq 0$. Using the regularity of $f_0(x)$ we now can complete the proof. □

Remark 3.3

It follows from theorem 2.2 that if $\bar{x}$ is not stationary (i.e., $\partial^- f'(\bar{x}) \not\subseteq \partial^+ f'(\bar{x})$) then (the steepest descent value)

(3.7) $$\min\{f'(\bar{x}, y) \mid y \in S^{n-1}\} = -\delta(\partial^- f'(\bar{x}), \partial^+ f'(\bar{x})).$$

Similarly, if $\bar{x}$ is stationary, then

(3.8) $$\min\{f'(\bar{x}, y) \mid y \in S^{n-1}\} = \sup\{\lambda \mid \partial^- f'(\bar{x}) + \lambda B \subseteq \partial^+ f'(\bar{x})\}.$$

(3.8) implies that $\bar{x}$ in theorem 3.3 is a strong local minimizer, provided $\partial^- f'(\bar{x}) \subseteq W \subseteq \partial^+ f'(\bar{x})$.

Definition 3.2

Let $f: D \to R$ be pseudosmooth at $x \in D$. The generalized gradient of $f'(x,\cdot)$ at zero will be denoted $\partial f(x)$ and called a subdifferential of f at x.

(If $f'(x,y)$ is convex by y, i.e. f is quasidifferentiable at x, then this definition of subdifferential coincides with one given in definition 1.2).

Remark 3.4

From convex analysis and by the definition of generalized gradient we have

$$(3.9) \qquad \partial f(x) = \overline{\operatorname{conv}}\{ \bigcup_{y \in Y} [c(y|\partial^+ f(x)) - c(y|\partial^- f(x))]\}$$

where $Y \subseteq S^{n-1}$ is the set of such directions y that the contact sets $c(y|\partial^+ f(x))$ and $c(y|\partial^- f(x))$ are singletons.

Theorem 3.4

Let f be pseudosmooth at x then

$$(3.10) \qquad \partial f(x) \subseteq \partial_c f(x)$$

Proof.

Following Clark we define the generalized directional derivative of f at x in the direction y by

$$(3.11) \qquad f^0(x,y) = \limsup_{h \to 0; t \downarrow 0} [f(x+h+ty) - f(x+h)]/t.$$

It was proved in [4] that $f^0(x,\cdot)$ is the support function of $\partial_c f(x)$. Let us denote $\varphi(y) = f'(x,y)$. By proposition 2.1 (d) it will be enough to show that $f^0(x,y) \geqslant \varphi^0(0,y)$ at every $y \in R^n$. We have for $t, \alpha > 0$ and $y \neq 0$

$$[\varphi(h+ty) - \varphi(h)]/t = [\varphi(\alpha h+\alpha ty) - \varphi(\alpha h)]/\alpha t =$$

$$= [f(x+\alpha h+\alpha ty) - f(x+\alpha h) + o(\alpha)]/\alpha t$$

and thus for $h_1 = \alpha h$ and $t_1 = \alpha t$

$$\lim_{\alpha\to 0^+} [f(x+h_1+t_1 y) - f(x+h_1)]/t_1 = [\varphi(h+ty)-\varphi(h)]/t$$

Then we have

$$f^0(x,y) = \limsup_{h_1\to 0;\, t_1\downarrow 0} [f(x+h_1+t_1 y) - f(x,h_1)]/t_1 \geqslant \limsup_{h\to 0\ t\downarrow 0} [\varphi(h+ty)-\varphi(h)]/t. \tag{3.12}$$

and this proves that $f^0(x,y) \geqslant \varphi^0(0,y)$. □

Now return to problem (Ω). Let us consider a point $x \in X$ and the cone $Q(x)$ of strong feasible directions:

$$\begin{cases} Q(x) = R^n \setminus \{0\} & , \quad h(x) < 0 \\ Q(x) = \{y \mid f_i'(x,y) < 0,\ i \in S(x)\}, & \quad h(x) = 0 \end{cases} \tag{3.13}$$

. (where $S(x) = \{ i \mid f_i(x) = 0, 1 \leq i \leq$

The stationarity of x means that if $y \in Q(x)$ then $f_0'(x,y) \geqslant 0$. By continuity of $f'(x,\cdot)$ this property can be extended to the topological closure $\bar{Q}(x)$ of $Q(x)$. Note that

$$\bar{Q}(x) \subseteq \{y \mid h'(x,y) \leqslant 0\}, \text{ for } x \text{ satisfying } h(x) = 0.$$

We have that the point $\bar{x}$ is stationary if and only if $y = 0$ is the minimizer of the following problem:

$$(\Psi) \quad \begin{cases} \text{Minimize } f_0'(\bar{x},y) \text{ by } y, \\ \text{subject to } y \in \bar{Q}(\bar{x}) \end{cases}$$

Now we recall the necessary conditions of Clark ([5]) applied to problems (Ψ) and (Ω) respectively.

Theorem 3.5

If $y=0$ solves (Ψ) then there exist nonnegative numbers α_0, α_i, $i \in S(\bar{x})$, not all zero such that

$$(3.14) \qquad 0 \in \alpha_0 \partial f_0(\bar{x}) + \sum_i \alpha_i \partial f_i(\bar{x}).$$

If $\bar{x}$ solves (Ω) locally, then there exist nonnegative numbers β_0, β_i, $i \in S(\bar{x})$, not all zero such that

$$(3.15) \qquad 0 \in \beta_0 \partial_c f_0(\bar{x}) + \sum_i \beta_i \partial_c f_i(\bar{x}) .$$

Remark 3.5

We introduced three types of necessary conditions for problem (Ω): (3.6), (3.14) and (3.15). Condition (3.6) is the most powerful. It follows from theorem 3.4 that (3.14) implies (3.15).

Example.

We define $f: R \to R$ by:

$$f(x) = \begin{cases} ax-|x| + x^2\sin 1/x , & x \neq 0 \\ 0 , & x = 0 . \end{cases}$$

It can be easily verified that $f(x)$ is pseudosmooth at zero and

$$< \partial^+ f(0), \partial^- f(0) > \equiv < \{a\}, [-1,1] >$$

$$\partial f(0) = [a-1, a+1]$$

$$\partial_c f(0) = [a-2, a+2]$$

We see that condition (3.6) is not satisfied for any a, although (3.14) is satisfied for $-1 \leq a \leq 1$ and (3.15) for $-2 \leq a \leq 2$.

Acknowledgement

We wish to thank Dr. A. Galperin for his helpful comments.

References

[1] Alexandroff, A.D., "On surfaces represented as the difference of convex functions", Izv. Akad. Nauk Kas. SSR 60, 1949, Ser. Mat. Mekh. 3, 3-20 (Russian).

[2] Alexandroff, A.D., "Surfaces represented by the differences of convex functions", Dokl. Adad. Nauk SSSr, 1950, 72, 613-616 (Russian).

[3] Cheney, E.W. and Goldstein, A.A., "An extension of the method of subgradients" in Nonsmooth Optimization, Proceedings of a II ASA Workshop, March 28-April 8, 1977, 51-70.

[4] Clarke, F.H., "Generalized gradients and applications", Transations of the American Mathematical Society 205 (1975), 247-262.

[5] Clarke, F.H., "A new approach to Lagrange multipliers", Mathematics of Operations Research, 1; 2 (1976), 165-174.

[6] Hartman, P., "On functions representable as a difference of convex functions" Pac. J. Math. 9 (1959), 707-713.

[7] Ioffe, A.D. and Tihomirov, V.M., "Theory of extremal problems", North-Holland Publ. Comp., 1979.

[8] Mifflin, R., "Semismooth and semiconvex functions in constrained optimization", SIAM Journal on Control and Optimization, 15, 6 (1977), 959-972.

[9] Pshenichnyi, B.N. (1971), "Necessary conditions for an extremum", Marcel Dekker, New York.

[10] Radstrom, H., "An embedding theorem for spaces of convex sets", Proc. American Math. Soc. 3 (1952), 165-169.

[11] Rockafellar, R.T., "Convex analysis", Princeton University Press, Princeton, N.J., 1970.

[12] Roberts, A.W. and Varberg, D.E., "Convex functions", Academic Press, New York and London, 1973.