GENERALIZED DIFFERENTIABILITY, DUALITY AND OPTIMIZATION FOR PROBLEMS DEALING WITH DIFFERENCES OF CONVEX FUNCTIONS

Jean-Baptiste HIRIART-URRUTY
Université Paul Sabatier (TOULOUSE III ,
U. E. R.
Mathématiques, Informatique, Gestion
118, route de Narbonne
F 31062 TOULOUSE Cédex

ABSTRACT. *A function is called d.c. if it can be expressed as a difference of two convex functions. In the present paper we survey the main known results about such functions from the viewpoint of Analysis and Optimization.*

INTRODUCTION

The analysis and optimization of convex functions have received a great deal of attention during the last two decades. If we had to choose two key-words from these developments, we would retain the concept of *subdifferential* and the *duality theory*. As it usual in the development of mathematical theories, people had since tried to extend the known definitions and properties to new classes of functions, including the convex ones. For what concerns the generalization of the notion of subdifferential, tremendous achievements have been carried out in the past decade and any mathematician who is faced with a nondifferentiable nonconvex function has now a panoply of generalized subdifferentials or derivatives at his disposal. A lot remains to be done in this area, especially concerning *vector-valued* functions ; however we think the golden age for these researches is behind us.

Duality theory has also fascinated many mathematicians since the underlying mathematical framework has been laid down in the context of Convex Analysis. The various duality schemes which have emerged in the recent years, despite of their mathematical elegance, have not always proved as powerful as expected.

The present paper is of a limited scope since it deals with generalized differentiability, duality and optimization for problems dealing with *differences of convex functions*. A real-valued function f defined on a convex set X is called *d.c.* on X (abbreviation for difference of convex functions on X) if there are two convex functions g and h such that :

$$\forall x \in X \qquad f(x) = g(x)-h(x). \tag{0.1}$$

Why to study such functions ? At first sight it might look as a mathematician's crotchet. There are in fact many reasons for considering such functions. Firstly there are some "mathematical" reasons for doing so. The class of d.c. functions on X, denoted by DC(X), is clearly the vector space generated by the cone of convex functions on X (denoted Conv(X) throughout). It also happens that some classes $K(X)$ of functions recently considered in the literature, and including for example convex or C^2 functions on an open convex set X, give rise to the vector space DC(X). The framework is typically as follows :

$$\begin{array}{ll} K(X) \text{ is a convex cone of functions on } X, & \\ \text{Conv }(X) \subset K(X) \subset DC(X), & (0.2) \\ K(X) - K(X) = DC(X). & \end{array}$$

Another mathematical reason explaining our interest in d.c. functions is that DC(X) is dense in the set C(X) of continuous functions over a compact convex set X, endowed with the topology of uniform convergence over X. This is a mere application of the STONE-WEIERSTRASS theorem. Density results of the same vein can be derived when X is no more closed or bounded. Finally the class DC(X) enjoys a remarkable stability with respect to operations usually encountered in Optimization, like taking the maximum (or minimum) of a finite number of functions, taking the product (or the quotient, the sum, etc...) of functions. We shall harck back to it later on.

Further incentives for studying d.c. functions come from applications. When dealing with a nonconvex optimization problem, it often happens that the data are actually d.c. (sometimes after a transformation of the original problem or through some dualization scheme). Our genuine feeling is that most of the "reasonable" optimization problems actually involve d.c. functions, even if we are not always able to recognize them as such ! To begin with, let us mention very simple examples.

Example 0.1. Let A be any (n,n) symmetric matrix and let Q be the associated quadratic form on $\mathbb{R}^n$, namely :

$$Q(x) = \frac{1}{2} < Ax,x >.$$

Q is obviously d.c. on $\mathbb{R}^n$ and there are several ways of finding positive semi-definite A^+ and A^- such that :

$$Q(x) = \frac{1}{2} < A^+x,x > - \frac{1}{2} < A^-x,x>. \qquad (0.3)$$

Example 0.2. Let S be an arbitrary nonempty subset of $\mathbb{R}^n$. It turns out that the square of the distance to S, denoted by d_S^2, is d.c. :

$$\forall x \in \mathbb{R}^n \quad d_S^2(x) = \|x\|^2 - \{ \|x\|^2 - d_S^2(x)\}. \qquad (0.4)$$

Surprisingly enough, $h : x \to h(x) = \frac{1}{2}\|x\|^2 - \frac{1}{2} d_S^2(x)$ is convex whatever S be (ASPLUND, 1973).d_S^2 is thus always d.c. ; it is convex whenever S is convex.

Example 0.3. Let A be a symmetric positive definite (n,n) matrix and let λ_M be the largest eigenvalue of A. By transforming RAYLEIGH's formulation :

$$\frac{\lambda_M}{2} = \max \{\frac{1}{2} < Ax,x >; \ \|x\| \leq 1\}$$

via dualization schemes which will be mentioned later (section IV), we obtain that :

$$-\frac{\lambda_M}{2} = \min_{x^* \in \mathbb{R}^n} \{\|x^*\| - \frac{1}{2} < A^{-1} x^*, x^* >\} \qquad (0.5)$$

or

$$-\frac{\lambda_M}{2} = \min_{x^* \in \mathbb{R}^n} \{\frac{1}{2} \|x^*\|^2 - \sqrt{< Ax^*, x^* >}\} . \qquad (0.6)$$

Calculating λ_M can therefore be viewed as a problem of minimizing (or maximizing) a d.c. function.

Looking at the above-displayed examples or considering more gèneral d.c. functions give rise to the following fundamental questions.

Q1. *How to recognize d.c. functions ?*

In view of the function f itself, or the derivative of f (or some generalized version of it), or the second derivative of f (in whatever sense), how to decide that f is actually d.c. ?

Q2. *Given a d.c. function, what is the "best" decomposition of it as a difference of convex functions ?*

There are infinitely many ways of decomposing a d.c. function as differences of convex functions. What should it mean that a decomposition is "better" than another one ?

Q3. *What (more !) about optimality conditions for problems dealing with d.c. functions ?*

D.c. functions are examples of locally Lipschitz functions for which a great deal has been done concerning optimality conditions. So, one may wonder what more can be said when the involved functions are d.c.. A related but more fundamental question concerns the duality schemes : *what is the involution* $f = g - h \to \varphi(g-h)$ *corresponding to* $f \to \varphi = f^*$ *for convex* *f ?* We shall see in that respect that TOLAND's involution $g-h \to h^* - g^*$ takes root in a basic formula yielding the conjugate (in the sense of Convex Analysis) of g-h.

Q4. *How to use the richness of structure of d.c. functions for designing global minimization algorithms ?*

Since convexity is present twice in the decomposition of a d.c. function (in an antagonistic way however), one may think of using this particular structure to devise algorithms which would converge to a global minimum (or a global maximum) of the given function.

We do not pretend to answer fully all these questions here. Our intention in this paper is to take into account some recent contributions in this area of research and to pose in a clear-cut manner the problems which remain unsolved. Before going further we lay out the setting of our presentation. Throughout the underlying space will be $\mathbb{R}^n$ and, for the sake of simplicity, we suppose the convex set X on which are defined d.c. functions is the whole space, so that we are concerned with DC $(\mathbb{R}^n)$. Most of the conclusions derived throughout extend to the DC(X), where X is an open convex set. To assume a finite-dimensional setting is a stronger limitation. Solving some partial differential equations amount to solving $f'(u) = 0$ where f is a d.c. function defined on an appropriate Hilbert space. These approaches however go beyond the scope of this paper. The paper is organized as follows :

I. D.c. functions : first properties ;

II. Recognizing a d.c. function ;

III. Decomposing d.c. functions ;

IV. Optimality conditions, duality for d.c. functions ;

V. Preview on minimization procedures for d.c. functions.

I. D.C. FUNCTIONS : FIRST PROPERTIES

Let $f \in DC(\mathbb{R}^n)$, so that there exist convex functions g and h from $\mathbb{R}^n$ into $\mathbb{R}$ such that $f = g-h$. Some properties of f are directly inherited from those of convex functions ; we list the main of them.

First of all, f is *locally Lipschitz* on $\mathbb{R}^n$. The derivative of f(*) does exist almost everywhere (a.e.) and if one denotes by Ω_f the set of points where $\nabla f(x)$ fails to exist, one clearly has that $\Omega_f \subset \Omega_g \cup \Omega_h$ whatever the decomposition of f as a difference of convex g and h. Since the nature of the set of points where a convex function is not differentiable is known in a great detail, one can therefore get a better insight into the structure of Ω_f. The directional derivative $d \to f'(x;d)$ exists everywhere and :

$$f'(x;d) = g'(x;d) - h'(x;d) \text{ for all } x \text{ and } d. \qquad (1.1)$$

So, "the tangent problem at x" defined by $f'(x;\cdot)$ is itself d.c. The functions for which $f'(x;\cdot)$ can be written as a difference of two positively homogeneous (finite) convex functions are called *quasi-differentiable* by DEMYANOV and his associates (cf. [12], [13] for example). For various reasons we prefer the terminology "*tangentially d.c.*" to "quasi-differentiable" for such functions. The gap between the class of d.c. functions and that of tangentially d.c. functions is actually small. As for example, f is tangentially d.c. whenever it is differentiable, while a little more is required on ∇f for f to be d.c. (see next section).

Concerning the generalized gradient (in CLARKE's sense) of f, the following estimate holds :

$$\partial f(x) \subset \partial g(x) - \partial h(x) \text{ for all } x. \qquad (1.2)$$

This estimate however may be very coarse. In fact, by writting f as $(g+\varphi) - (h+\varphi)$ with an ad hoc convex function φ, the estimate (1.2) can be rendered as rough as desired ! An exact evaluation of $\partial f(x)$ would require us to know the generalized derivative of the $\mathbb{R}^2$-valued function $(g,-h)^T$. But, for that purpose, to know that g and h are convex do not help very

---

(*) Derivatives in the sense of GATEAUX, HADAMARD or FRECHET are equivalent for $f \in DC(\mathbb{R}^n)$.

much. There is however a consequence one can draw immediately from (1.2) and which is as follows : *$\partial f(x)$ is reduced to $\{\nabla f(x)\}$ a.e.*. Surprisingly enough and contrary to what happens for convex functions, *$\partial f(x_0)$ does not necessarily reduce to a singleton when f is differentiable at x_0.*

Example 1.1. (from [45]).

Lef $f : \mathbb{R}^2 \to \mathbb{R}$ be defined as $f(\xi_1, \xi_2) = |\xi_2 - \xi_1^3| - |\xi_2|$. For reasons which will appear later (section II) f is d.c. One verifies that f is differentiable at (0,0) with $\nabla f(0,0) = (0,0)$, while $\partial f(0,0) = \{0\} \times [-2,2]$. Calculating $\partial f(\xi_1, \xi_2)$ at points (ξ_1, ξ_2) satisfying $\xi_2 = \xi_1^3$ or $\xi_2 = 0$ gives a good idea of how $\partial f(\xi_1, \xi_2)$ behaves when (ξ_1, ξ_2) approaches (0,0).

Therefore there may be points where f is differentiable without being strictly differentiable. Let Ω_f^1 denote the set (of null measure) where f fails to be strictly differentiable. If $x_0 \in \Omega_f^1 \setminus \Omega_f$, it comes from (1.1) that :

$$\partial g(x_0) = \partial h(x_0) + \nabla f(x_0), \tag{1.3}$$

that is $\partial g(x_0)$ is a translation of $\partial h(x_0)$. For the f displayed in Example 1.1, a decomposition as a difference of convex functions is as follows :

$$\begin{aligned} f(\xi_1, \xi_2) &= 2 \max \{(\xi_1^3)^+, \xi_2 + (\xi_1^3)^-\} - (|\xi_1|^3 + 2\,\xi_2^+). \\ &= g(\xi_1, \xi_2) - h(\xi_1, \xi_2). \end{aligned}$$

At $x_0 = (0,0)$ we have that $\partial g(x_0) = \partial h(x_0) = [0,2]$. The phenomenon we have observed raises the question whether the generalized gradient is appropriate to d.c. functions ; for further discussion about the various first-order generalized derivatives, the reader may refer to [25]. Given any $\varphi \in \mathrm{Conv}(\mathbb{R}^n)$ and $a, b \in \mathbb{R}^n$, the difference $\varphi(b) - \varphi(a)$ can be represented in an integral form as follows :

$$\varphi(b) - \varphi(a) = \int_a^b \langle \partial\varphi(a + t(b-a)), b-a \rangle \, dt, \tag{1.4}$$

where the right-hand means the integral of the multifunction $\Gamma_{a,b}$: $t \rightrightarrows \langle \partial\varphi(a+t(b-a)), b-a \rangle$ over [a,b]. This is a formal writing since the integrals of all the measurable selections of $\Gamma_{a,b}$ over [a,b] yield the same value, namely $\varphi(b) - \varphi(a)$. Despite of the inclusion (1.2) and the possible discrepancy between the $\partial f(x)$ and $\{\nabla f(x)\}$, an integral representation analogous to (1.4) holds true for a d.c. function f. For any locally Lipschitz function f on $\mathbb{R}^n$ we know that :

$$f(b) - f(a) \in \int_a^b < \partial f(a+ t(b-a)), b-a > dt. \tag{1.5}$$

Now, (1.4) holds for $g(b) - g(a)$ and $h(b) - h(a)$ whenever f is expressed as a difference of convex functions g and h. Since the integral of the sum of two integrable multifunctions is the sum of their integrals, we have :

$$f(b) - f(a) = \int_a^b < \partial g(a+ t(b-a)) - \partial h(a+ t(b-a)), b - a > dt.$$

Whence

$$f(b) - f(a) = \int_a^b < \partial f(a+ t(b-a)), b-a > dt. \tag{1.6}$$

Knowing ∂f therefore allows us to recover f.

Concerning second-order derivatives, d.c. functions inherit from convex functions they are "*twice-differentiable a.e.*". Let us make this statement more precise. The multifunction ∂f is said to be differentiable at x_0 if f is differentiable at x_0 and if there exists a linear mapping denoted by $\nabla^2 f(x_0)$ such that :

$$\|\partial f(x) - \nabla f(x_0) - \nabla^2 f(x_0)(x-x_0)\| = o\ (\ \|x - x_0\|\ ), \tag{1.7}$$

or in other words :

$$\forall \eta > 0,\ \exists \delta > 0,\ \forall x \ \text{ with } \|x - x_0\| \leq \delta,\ \forall x^* \in \partial f(x),$$

$$\|x^* - \nabla f(x_0) - \nabla^2 f(x_0)(x - x_0)\| \ \leq \eta\ \|x - x_0\|\ .$$

$\nabla^2 f(x_0)$ is then a symmetric mapping which is called the derivative of ∂f at x_0. As an application of MIGNOT's differentiability theorem on maximal monotone multifunctions [34, Theorem 1.3] we have that the subdifferential multifunction of $\varphi \in \text{Conv}\ (\mathbb{R}^n)$ is differentiable a.e. on $\mathbb{R}^n$. It is therefore just a corollary to claim that "*the generalized gradient of a d.c. function is differentiable a.e.*". Another formulation, better known in the context of Convex Analysis, states that a convex function on $\mathbb{R}^n$ has a second-order TAYLOR expansion at almost all points of $\mathbb{R}^n$ (ALEXANDROFF, 1939).

That means that, at almost every $x_0 \in \mathbb{R}^n$, there is a (symmetric) linear mapping denoted by $A^2f(x_0)$ such that :

$$f(x) = f(x_0) + < \nabla f(x_0),\ x - x_0 > + \frac{1}{2} < A^2f(x_0)(x-x_0), x-x_0 >$$
$$+\ o(\|x - x_0\|^2). \tag{1.8}$$

Clearly f has a second-order expansion at x_0 whenever ∂f is differentiable at x_0. Thus "*a d.c. function has a second-order TAYLOR expansion a.e.*".

Let Ω_f^2 denote the set of null measure where ∂f fails to be differentiable. It readily comes from (1.7) that $\partial f(x_0) = \{\nabla f(x_0)\}$ whenever ∂f is differentiable at x_0. We thus summarize the differentiability properties of a d.c. function f by stating :

$$\begin{aligned} &\Omega_f \subset \Omega_f^1 \subset \Omega_f^2 \\ &\Omega_f^2 \text{ is of null measure.} \end{aligned} \tag{1.9}$$

There are limitations in extending properties of convex functions to d.c. functions. We briefly mention here some of them.

The graph of the subdifferential of a convex function on $\mathbb{R}^n$ is of a very special structure since it is a *Lipschitz manifold* of $\mathbb{R}^n \times \mathbb{R}^n$ (see [43] for a recent account on that subject). The graph of the generalized gradient of a d.c. function is no more a Lipschitz manifold.

Example 1.2. Let $f : \mathbb{R} \to \mathbb{R}$ be defined by $f(0) = 0$ and $f'(x) = |x|^{1/2}$ for all x. $f \in DC(\mathbb{R})$ but the graph of f' is not a Lipschitz curve in $\mathbb{R}^2$.

Apart from some nasty points like 0 in the previous example, the graph of the generalized gradient of a d.c. function looks pretty much alike a Lipschitz manifold. So, most of the geometrical properties of tangent cones to the graph of the subdifferential of a convex function (such as displayed in [43]) should have their counterparts for d.c. functions. Another property of convex functions which is no longer true for d.c. functions is the following : *a limit of d.c. functions is not necessarily d.c.*.

Example 1.3. (from [45])

Let $f_n : \mathbb{R} \to \mathbb{R}$ be defined as $f_n(x) = \text{Min}\{|x-\frac{1}{k}|, k=1,\dots,n\}$. f_n is d.c. on $\mathbb{R}$ but the derivative of $f = \lim_{n\to+\infty} f_n$ is not of bounded variation in a neighbourhood of 0, i.e., f is not d.c. around 0. The same example could serve to show that the infimum (or the supremum) of an infinite family of d.c. functions is not necessarily d.c..

A more severe drawback is that, contrary to convex functions, a function $f : \mathbb{R}^n \to \mathbb{R}$ may be d.c. on lines (i.e., $t \to f(a+t(b-a))$ is d.c. whatever a and b) without being d.c. on the whole space $\mathbb{R}^n$. YOMDIN has shown us examples of such functions. This state of affairs is somewhat baffling since, as recalled in the next section, d.c. functions on the real line bear an easy characterization. All the properties of d.c. functions displayed in this section can serve in their negative form : a function which does not satisfy one of the properties mentioned is not d.c..

Example 1.4. Let A be a Borel set of $\mathbb{R}$ satisfying the next property : for all nonempty open interval I of $\mathbb{R}$, $\lambda(A \cap I) > 0$ and $\lambda(A^c \cap I) > 0$. The function f defined on $\mathbb{R}$ by :

$$f(x) = \int_0^x 1_A(t)\,dt$$

is strictly increasing and locally Lipschitz on $\mathbb{R}$. Is f d.c. on $\mathbb{R}$? Although there are various reasons for answering "no", we retain the following one : $\partial f(x) = [0,1]$ for all $x \in \mathbb{R}$ so that $\partial f(x)$ does not reduce to $\{f'(x)\}$ a.e..

II. RECOGNIZING D.C. FUNCTIONS

Whether a locally Lipschitz function *on* $\mathbb{R}$ is d.c. or not depends on the variation of f' (defined a.e.). f is d.c. on $\mathbb{R}$ if and only if f' is of bounded variation on compact intervals of $\mathbb{R}$. This is easy to imagine since, according to JORDAN's decomposition, f' is of bounded variation if and only if it can be expressed as a difference of two increasing functions. A further property which is worth noticing is the following.

THEOREM 2.1. *Let f be a differentiable d.c. function on* $\mathbb{R}$. *f is then continuously differentiable and can be written as a difference of (continuously) differentiable convex functions.*

Proof. f', as a derivative, satisfies DARBOUX' property, that is : the image by f' of any interval is an interval. Therefore, f' is continuous whenever it is of bounded variation. Hence f' can be written as a difference of two continuous increasing functions.□

To recognize d.c. functions among locally Lipschitz functions on $\mathbb{R}^n$ is not an easy matter. We will give some indications about that, according to what is at our disposal : the function itself, or its first or second (generalized) derivative.

II.1. Apart from some contributions peculiar to the two-dimensional case ([58] for example), the main result on recognizing d.c. functions on $\mathbb{R}^n$ by referring only to the function goes back to HARTMAN (1959). Before stating it, we recall that $f : \mathbb{R}^n \to \mathbb{R}$ is said *locally d.c.* on $\mathbb{R}^n$ if, for every x_0, there exist a convex neighborhood V of x_0, convex functions g_V and h_V such that :

$$f(x) = g_V(x) - h_V(x) \text{ for all } x \in V. \qquad (2.1)$$

THEOREM 2.2. [HARTMAN, 1959]. Every locally d.c. function on $\mathbb{R}^n$ is globally d.c. on $\mathbb{R}^n$.

HARTMAN actually stated his theorem in a slightly more general form, by extending locally d.c. functions on a closed (or open) convex set. This type of result, although well known in the convex case, is rather surprising since it means that information on f around each point is enough to decide that f is globally d.c. on $\mathbb{R}^n$. HARTMAN's proof needs to be "freed from dust" somewhat ; the extension procedures HARTMAN uses can be adapted by relying on extension techniques from modern convex analysis (infimal convolution of a function with $k\|\cdot\|$ for example). As a mere consequence of HARTMAN's theorem, we note that *every function f which is C^2 on $\mathbb{R}^n$ is d.c. on $\mathbb{R}^n$*. It is easy to see that such a function is locally d.c.. Indeed, due to the continuity of $\nabla^2 f$, one can find a decomposition of f on $\bar{B}(\bar{x},r)$ as :

$$f(x) = (f(x) + \rho\|x\|^2) - \rho\|x\|^2,$$

where ρ is chosen such that $f + \rho\|\cdot\|^2$ be convex on $\bar{B}(\bar{x},r)$. Hence, for every C^2 function f, there are convex functions g and h such that :

$$f(x) = g(x) - h(x) \text{ for all } x \in \mathbb{R}^n. \qquad (2.3)$$

Moreover, it has been proved by BOUGEARD ([7]) that the function h in the decomposition above could be chosen to be C^∞. Hence *any C^2 function on $\mathbb{R}^n$ can be written as a difference of two convex functions, one of them being C^2 and the other one C^∞*. POMMELLET ([37]) also offered an alternative proof that a C^2 function could be written as a difference of two C^2 convex functions. The basic idea to prove such type of results is to build up global decompositions starting from local decompositions and using extension (or regularization) procedures of some kind. However, all the proofs we know are "constructive" in the sense that they indeed yield g and h satisfying (2.3) but could hardly be carried over computational aspects.

Quite a little is missing for a C^1 function to be d.c.. Actually, a $C^{1,1}$ function, that is a function f whose gradient is locally Lipschitz, is d.c.. $C^{1,1}$ functions constitute a subclass of the so-called *lower-C^2 functions*. The class of lower-C^2 functions has been studied in the literature by several authors, using different names and apparently unaware that the functions they were talking about belong to the same class (cf. [26], [29], [42], [53], [54], etc...). Among the various characterizations of lower-C^2 functions, we retain the following one : f is lower-C^2 on $\mathbb{R}^n$ if for every x_0, there exist a convex neighborhood V of x_0, a convex function g_V, a quadratic convex function h_V such that :

$$f(x) = g_V(x) - h_V(x) \text{ for all } x \in V. \tag{2.4}$$

The class of lower-C^2 functions on $\mathbb{R}^n$ is denoted by $LC^2(\mathbb{R}^n)$. This class remains unchanged if we impose that h_V in decomposition (2.4) be $C^{1,1}$ (or C^2, or C^∞) and convex. We clearly have :

$$\mathrm{Conv}(\mathbb{R}^n) \subset LC^2(\mathbb{R}^n) \subset DC(\mathbb{R}^n) \quad ; \tag{2.5}$$

$$C^{1,1}(\mathbb{R}^n) \subset LC^2(\mathbb{R}^n). \tag{2.6}$$

$LC^2(\mathbb{R}^n)$ is an example of convex cone of functions getting in the framework announced in (0.2). The vector space generated by $LC^2(\mathbb{R}^n)$ *is* $DC(\mathbb{R}^n)$ while the largest vector space $\mathcal{L}$ contained in $LC^2(\mathbb{R}^n)$ (that is : $\mathcal{L} = LC^2(\mathbb{R}^n) \cap - LC^2(\mathbb{R}^n))$ contains $C^{1,1}(\mathbb{R}^n)$. Actually the f belonging to $\mathcal{L}$ are characterized as follows : f is differentiable and for all $x_0 \in \mathbb{R}^n$ there is a neighborhood V of x_0 and a positive k such that :

$$|\langle \nabla f(x) - \nabla f(x'), x-x' \rangle| \leq k \|x-x'\|^2 \text{ whenever } x,x' \in V. \tag{2.7}$$

Hence $\mathcal{L}$ *contains* $C^{1,1}(\mathbb{R}^n)$ and is *included* in $C^1(\mathbb{R}^n)$. f is said to be *globally lower*-C^2 on $\mathbb{R}^n$ if the neighborhood V in the definition (2.4) is imposed to be the whole of $\mathbb{R}^n$. Contrary to what happens for the d.c. character, a lower-C^2 function on $\mathbb{R}^n$ is not necessarily globally lower-C^2 on $\mathbb{R}^n$.

To see there is a strong gap between $LC^2(\mathbb{R}^n)$ and $DC(\mathbb{R}^n)$, we consider again the function d_S^2 (cf. example 0.2).

PROPOSITION 2.3. *Let S be a nonempty closed set in* $\mathbb{R}^n$. *Then the function* d_S^2 *is lower*-C^2 *if and only if S in convex.*

Proof. d_S^2 is convex, hence lower-C^2, whenever S is convex. Due to an alternative characterization of lower-C^2 functions by ROCKAFELLAR ([42]), one easily shows that $-d_S^2$ is lower-C^2 on $\mathbb{R}^n$. So, assuming that d_S^2 itself is lower-C^2, we are in the presence of a differentiable function d_S^2. We know from ASPLUND (1973) that the function $h : x \to \frac{1}{2}\|x\|^2 - \frac{1}{2} d_S^2(x)$ is convex and that the subdifferential $\partial h(x_0)$ contains $P_S(x_0) = \{u \in S ; \|x_0-u\| = d_S(x_0)\}$. Thus, in our case, $P_S(x)$ is a singleton for all x. That means that S is a so-called CHEBYSHEV set, and, in $\mathbb{R}^n$, the only CHEBYSHEV sets are convex sets. □

II.2. Lower-C^2 functions can be characterized via their generalized gradients. Following ROCKAFELLAR's terminology ([42]), the generalized gradient ∂f of a locally Lipschitz function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be *strictly hypomonotone* at x_0 if there exists a neighborhood V of x_0 and a positive k such that :

$$\langle x-x', y-y' \rangle \geq -k \|x-x'\|^2 \text{ for all } x,x' \text{ in } V$$
$$\text{and } y \in \partial f(x),\ y' \in \partial f(x'). \tag{2.8}$$

ROCKAFELLAR's characterization of a lower-C^2 function via its generalized gradient comes as follows : *a locally Lipschitz* $f : \mathbb{R}^n \to \mathbb{R}$ *is lower*-C^2 *if and only if, for each* $x_0 \in \mathbb{R}^n$, ∂f *is strictly hypomonotone at* x_0. One may wonder whether there is a similar characteristic property for d.c. functions. Simple examples show that such a characterization is by no means easy to derive. The function f in Example 1.4 has a generalized gradient "varying very nicely" since $\partial f(x) = [0,1]$ for all x ; in imprecise terms, ∂f is of "bounded variation" on $\mathbb{R}$ (the HAUSFORFF distance between $\partial f(x)$ and $\partial f(x')$ is null for all x,x'). The one-dimensional case suggests us to look at a notion of "bounded variation" for functions of several variables. There are, in the literature, at least half a dozen different definitions for "$F : \Omega \subset \mathbb{R}^n \to \mathbb{R}^m$ is of bounded variation on Ω". Grafting them to our situation has not yielded very significant results ; in particular, the following desired result could not be obtained : a differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ is d.c. if and only if its gradient mapping ∇f is of bounded variation on a neighborhood of each point. This is not, after all, so surprising. Each definition of "bounded variation" has been introduced for particular purposes (e.g., the concept of bounded variation or bounded deformation in the theory of plasticity), while the notion we are looking for has to be strongly connected to that of monotonicity such as used in the context of Convex Analysis. The next definition is an attempt by ELLAIA ([18, chapter II]) to cope with that problem.

<u>*DEFINITION 2.4.*</u> *A multifunction $\Gamma : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is of bounded variation at $x_0 \in \mathbb{R}^n$ if there exists a maximal cyclically monotone multifunction $M : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ containing x_0 in the interior of its domain, such that :*

$$\begin{array}{ll}\displaystyle\limsup_{\substack{x \to x_0,\ x' \to x_0 \\ y \in \Gamma(x),\ y' \in \Gamma(x') \\ z \in M(x),\ z' \in M(x') \\ \langle x-x',\ z-z'\rangle \neq 0}} \frac{|\langle x-x',\ y-y'\rangle|}{\langle x-x',\ z-z'\rangle} < +\infty. & (2.9)\end{array}$$

In other words, for some positive k and a neighborhood V of x_0,

$$|\langle x-x',\ \Gamma(x) - \Gamma(x')\rangle| \leq k \langle x-x',\ M(x) - M(x')\rangle$$
$$\text{for all } x,x' \text{ in } V. \qquad (2.10)$$

This is actually a requirement on the "angles" $\langle x-x',\ \Gamma(x) - \Gamma(x')\rangle$ and $\langle x-x',\ M(x) - M(x')\rangle$. This concept is easier to grasp when Γ is merely a mapping ; of course, when n = 1, it amounts to the usual notion of bounded variation in a neighborhood of x_0. If Γ_1 and Γ_2 are of bounded variation at x_0, so is $\lambda\Gamma_1 + \mu\Gamma_2$ for any $\lambda, \mu \in \mathbb{R}$.

Due to what has been assumed on M, M is the subdifferential of a convex function finite in a neighborhood of x_0. So, the next result was foreseeable.

<u>THEOREM 2.5</u> ([18]). *A locally Lipschitz function $f : \mathbb{R}^n \to \mathbb{R}$ is d.c. if and only if ∂f is of bounded variation at all $x_0 \in \mathbb{R}^n$.*

The definition of bounded variation we have considered for ∂f is not quite satisfactory ; in particular it is heavy to use. More work should be done to get a clear-cut characterization of d.c. functions via their first-order (generalized) derivatives.

<u>II.3</u>. Since C^2 functions on $\mathbb{R}^n$ are always d.c., little room is left for characterizing (non C^2) d.c. functions f when having some (generalized) second derivative of f at our disposal. An approach which can be carried out to a certain extent is that of considering *second derivatives of f in the distributional sense*. When f is defined on the real line (n = 1), it is known that f is d.c. if and only if the second derivative of f is a RADON measure. When n > 1, some *necessary* conditions can be derived from DUDLEY's results on convex functions ([14]) ; for example, if $f \in DC(\mathbb{R}^n)$, the second derivative of the distribution associated with f is a (n,n) matrix-valued RADON measure. For the converse, decomposing a (n,n) matrix-valued RADON measure into the difference of two nonnegative (n,n) matrix-valued RADON measures does not help very much since one does not know whether the resulting distributions are second derivatives. The same type of difficulties will arise in decomposing d.c. functions (see Section III).

When pointwise derivatives of f are available, again some necessary conditions can be derived in terms of the total Gaussian curvature of f ([54]).

III. <u>DECOMPOSING D.C. FUNCTIONS</u>

Given $f \in DC(\mathbb{R}^n)$, it is evident there are infinitely many manners of expressing it as a difference of convex functions. Is there somehow a way of reducing the choice of such convex functions ? How to define that a decomposition is better than another one ? Finally, from the practical viewpoint, how to find a decomposition of a function f which has been constructed from other functions whose decompositions are better known ? The present section addresses to these questions.

Firstly, let us remark that if a decomposition f = g-h is available, one can always choose g and h as being *strictly (or uniformly) convex* ; it suffices to write :

$$f = (g+\varphi) - (h+\varphi) = \overline{g} - \overline{h}, \qquad (3.1)$$

with φ strictly (resp. uniformly) convex. The simplest choice consists in taking $\varphi(x) = k \|x\|^2$ with $k > 0$. Adding functions like $k \|\cdot\|^2$ may be deliberate since it adds more structure to the functions $\overline{g}$ and $\overline{h}$ involved in the decomposition of f ; in particular - and this is of importance in the context of duality (see next section) - it ensures that both $(\overline{g})^*$ and $(\overline{h})^*$ are everywhere finite and differentiable.

We say that a decomposition $f = g - h$ is *normalized* if $\inf_{\mathbb{R}^n} h(x) = 0$.

<u>PROPOSITION 3.1</u>. *Every d.c. function enjoys normalized decompositions.*

Proof. Given $h \in \mathrm{Conv}(\mathbb{R}^n)$, we show there is an affine function θ such that $\inf_{\mathbb{R}^n} (h+\theta) = 0$.

Let $x_0 \in \mathrm{dom}\, h^*$. We pose $\theta(x) = -\langle x_0^*, x \rangle + h^*(x_0^*)$. Thus

$$(h+\theta)^* (x^*) = -h^*(x_0^*) + h^*(x^* + x_0^*) \text{ for all } x^*.$$

Whence $(h+\theta)^*(0) = 0$. □

<u>III.1</u>. A normalized decomposition $f = g_{min} - h_{min}$ is called *minimal* if $g_{min} \leq g$ (and, consequently, $h_{min} \leq h$) whatever $f = g-h$ be a normalized decomposition. This definition is very stringent since it requires g and h to be comparable with g_{min} and h_{min} whatever g and h appear in a normalized decomposition of f. A weaker condition would consist in requiring $g_{min} \leq g$ whenever g_{min} and g are comparable. Nevertheless, apart from some particular cases (like d.c. functions on $\mathbb{R}$ or polyhedral functions on $\mathbb{R}^n$), there is no "minimal" decomposition as long as the pointwise ordering of functions ($\varphi \leq \psi$ if $\varphi(x) \leq \psi(x)$ for all x) is considered. There are counterexamples even for d.c. functions of two variables.

Example 3.1. (from [21])

Let $f \in DC(\mathbb{R}^2)$ be defined as $f(\xi_1, \xi_2) = 2\,\xi_1 \xi_2$.

The following are normalized decompositions of f :

$$f(\xi_1, \xi_2) = g_\varepsilon(\xi_1, \xi_2) - h_\varepsilon(\xi_1, \xi_2),$$

where

$$g_\varepsilon(\xi_1, \xi_2) = (\varepsilon\xi_1 + \frac{\xi_2}{\varepsilon})^2,$$
$$h_\varepsilon(\xi_1, \xi_2) = \varepsilon^2 \xi_1^2 + \frac{\xi_2^2}{\varepsilon^2}, \tag{3.2}$$

and $\varepsilon > 0$. Observe the antagonistic role played by ε^2 and $1/\varepsilon^2$.

If a minimal (normalized) decomposition $f = g_{min} - h_{min}$ should exist, we would have :

$$0 \le h_{min}(\xi_1, \xi_2) \le \varepsilon^2 \xi_1^2 + \frac{\xi_2^2}{\varepsilon^2} \text{ for all } (\xi_1, \xi_2) \in \mathbb{R}^2$$

and all $\varepsilon > 0$.

Since h_{min} is convex, the above implies that $h_{min} \equiv 0$. Thus, $f = g_{min}$, which is impossible because f is not convex on $\mathbb{R}^2$.

Let us go back to the example of quadratic forms on $\mathbb{R}^n$ (cf. example 0.1)

$$Q(x) = \frac{1}{2} < Ax, x >.$$

$g : x \to g(x) = \frac{1}{2}\{<Ax,x> + k \|x\|^2\}$ is convex for k large enough, so that Q can be decomposed as :

$$Q(x) = \frac{1}{2}\{<Ax,x> + k\|x\|^2\} - \frac{1}{2} k \|x\|^2. \tag{3.3}$$

Another possible decomposition of Q is via the diagonalization D of A. Let D^+ (resp. D^-) denote the diagonal matrix whose elements are positive parts (resp. negative parts) of D. The decomposition $D = D^+ - D^-$ yields a decomposition of Q,

$$Q(x) = \frac{1}{2} <A^+ x, x> - \frac{1}{2} <A^- x, x>, \tag{3.4}$$

where A^+ and A^- are positive semi-definite (and singular if Q is neither convex nor concave). Is decomposition (3.4) "better" than decomposition (3.3) ? Is decomposition (3.4) optimal in some sense ? This example just as the previous ones clearly show that *minimal decompositions have to be searched for in restricted classes of decompositions*. So, a more sensible question for the decomposition of quadratic forms would be : is decomposition (3.4) minimal in the class of quadratic decompositions of Q ? When trying to decompose a C^2 function as a difference of convex functions, one might be tempted to decompose $\nabla^2 f(x)$ as it is done in the scheme leading to (2.4) for example ;

$$<\nabla^2 f(x)d,d> = <A^+(x)d,d> - <A^-(x)d,d>. \tag{3.5}$$

The drawback, foreseen in the previous section, is that the $A^+(x)$ and $A^-(x)$ obtained in such a way are not necessarily second derivatives of functions ! The case of $f(\xi_1, \xi_2) = \xi_1 / \xi_2$ on $(\mathbb{R}_+^*)^2$ is a typical counterexample.

In sum, given a function $f \in DC(\mathbb{R}^n)$, there is no general rule for finding in an automatic way a decomposition of f as a difference of convex functions.

III.2. As mentioned earlier, $DC(\mathbb{R}^n)$ is closed under all the operations usually considered in Optimization. So, given d.c. functions f_i whose decompositions $g_i - h_i$ are known, how to find a decomposition $g - h$ of a d.c. function f constructed from the f_i ? The problem is not as difficult as it might appear at the first glance ; that depends, of course, of the operation carried out on the f_i.

If f is a linear combination of the f_i, no trouble arises since a decomposition of f comes out immediately. If now f is the maximum (or the minimum) of the f_i, one gets a decomposition of f as follows :

$$\max_{i=1,\ldots,k} f_i = \max_{i=1,\ldots,k} \{g_i + \sum_{\substack{j=1 \\ j\neq i}}^{k} h_j\} - \sum_{i=1}^{k} h_i \; ; \tag{3.6}$$

$$\min_{i=1,\ldots,k} f_i = \sum_{i=1}^{k} g_i - \max_{i=1,\ldots,k} \{h_i + \sum_{\substack{j=1 \\ j\neq i}}^{k} g_j\} . \tag{3.7}$$

As for example, if a decomposition $g - h$ of f is available, we have :

$$\begin{aligned} |f| &= 2 \max (g,h) - (g+h) \; ; \\ f^+ &= \max (g,h) - h \; ; \\ f^- &= \max (g,h) - g. \end{aligned} \tag{3.8}$$

As for the product of two d.c. functions f_1 and f_2, we firstly decompose f_i as $f_i^+ - f_i^-$ so that the question reduces to finding a decomposition of the product of two *positive* d.c. functions. Let therefore $f_1 = g_1 - h_1$

and $f_2 = g_2 - h_2$ be normalized decompositions of positive d.c. functions f_1 and f_2. The g_i and h_i are thus positive convex functions. A decomposition of $f_1 \cdot f_2$ is then at hand :

$$f_1 \cdot f_2 = \frac{1}{2}[(h_1 + h_2)^2 + (g_1 + g_2)^2] - \frac{1}{2}[(h_1 + g_1)^2 + (h_2 + g_2)^2]. \quad (3.9)$$

By reiterating or combining such decomposition rules, one often is able to find a decomposition of a function which has been constructed from other functions whose decompositions are better known. See, for instance, the decomposition of the d.c. function f arising in Example 1.1.

IV. OPTIMALITY CONDITIONS, DUALITY FOR D.C. FUNCTIONS

IV.1. *Optimality conditions*

D.c. functions are locally Lipschitz and possess directional derivatives ; optimality conditions for d.c. mathematical programs can therefore be deduced from those derived in the context of mathematical programs with directionally differentiable data ([36], [39]), or tangentially d.c. data ([12], [45], [46]), or merely locally Lipschitz data ([11]). However, knowing that the data can be written as differences of convex functions adds to the structure of the problem and may be used for optimality conditions. To begin with, consider the unconstrained case :

$$(\mathcal{P}) \text{ Minimize } f(x) = g(x) - h(x) \text{ over } \mathbb{R}^n.$$

A necessary condition for x_0 to be a local minimum of f is that :

$$0 \in \partial f(x_0) \subset \partial g(x_0) - \partial h(x_0) . \quad (4.1)$$

In other words, the subdifferentials $\partial g(x_0)$ and $\partial h(x_0)$ must overlap :

$$\partial g(x_0) \cap \partial h(x_0) \neq \phi . \quad (4.2)$$

The same condition holds true when x_0 is a local maximum of f. The major drawback of (4.2) is that it *depends on the decomposition of f* as a difference of convex functions g and h (besides the fact that it is not always informative). In particular - and this is also true for the definition of critical points of f (see paragraph IV.3) - a given x_0 may satisfy condition (4.2) for an ad hoc decomposition of f. A further approach consists in expressing necessary conditions for optimality via the *lower-subdifferen-*

tial of f in the sense of PENOT ([36]). If x_0 is a local minimum of f, then

$$f'(x_0;d) \geq 0 \text{ for all } d \in \mathbb{R}^n, \tag{4.3}$$

so that :

$$0 \in \underline{\partial} f(x_0), \tag{4.4}$$

where $\underline{\partial} f(x)$ stands for :

$$\{x^* \in \mathbb{R}^n \mid \langle x^*, d \rangle \leq f'(x;d)\}. \tag{4.5}$$

There is an alternate way of looking at the optimality condition (4.4), that is by way of the $*$-difference of $\partial g(x_0)$ and $\partial h(x_0)$. Given two nonempty compact convex sets A and B, the $*$-difference of A and B is the set

$$A \underline{*} B = \{x \in \mathbb{R}^n \mid x + B \subset A\}.$$

This operation was introduced in the context of linear differential games by PONTRYAGIN and further exploited by PSHENICHNYI ([38]) for convex optimization. If the algebraic difference A - B of A and B is known to be "too large", the $*$-difference turns out to be often "too small" ! $A \underline{*} B$ is actually a compact convex set whose support function is the biconjugate of the difference of the support functions of A and B ;

$$\psi^*_{A \underline{*} B} = (\psi^*_A - \psi^*_B)^{**}. \tag{4.6}$$

We infer from (4.3) that the biconjugate function of $d \to f'(x_0;d) = g'(x_0;d) - h'(x_0;d)$ is positive. Whence a necessary condition for x_0 to be a local minimum of $f(x) = g(x) - h(x)$ is that :

$$0 \in \partial g(x_0) \underline{*} \partial h(x_0). \tag{4.7}$$

Observe that $\partial g(x_0) \underline{*} \partial h(x_0)$ does not depend on g and h but on the difference g - h. Finally, note that :

$$\underline{\partial} f(x_0) = \partial g(x_0) \underline{*} \partial h(x_0),$$

so that (4.7) is a reformulation of the optimality condition (4.4). To summarize, a necessary condition for x_0 to be a local minimum of f is that :

$$0 \in \partial g(x_0) \underline{*} \partial h(x_0) \subset \partial f(x_0) \subset \partial g(x_0) - \partial h(x_0). \tag{4.8}$$

A *sufficient* condition for x_0 to be a (strict) local minimum of f can also be expressed in terms of $\partial g(x_0) \underset{-}{*} \partial h(x_0)$ by just exploiting the sufficient condition for minimality :

$$f'(x_0;d) \geq 0 \text{ for all non-null } d. \tag{4.9}$$

The following is easy to prove : *if* o *lies in the interior of* $\partial g(x_0) \underset{-}{*} \partial h(x_0)$, *then* x_0 *is a strict local minimum of* f.

Consider now the constrained minimization problem :

$$(\mathcal{P}) \text{ Minimize } f(x) = g(x) - h(x) \text{ over } S.$$

The intended work, in its generality, would consist in deriving optimality conditions when the constraint set is represented via inequalities and equalities involving d.c. functions. That has been done to a certain extent in [12], [18, p. 95-100], [46]. We show here how the above-displayed conditions extend to the constrained case, when the constraints are formulated as : $x \in S$. For $x_0 \in S$, we recall that the *contingent cone* to S at x_0 (or BOULIGAND's tangent cone to S at x_0) is the closed cone with apex 0, denoted by $T(S;x_0)$ and defined as :

$$T(S;x_0) = \{d \mid \exists\, (d_n) \to d,\ \exists (\alpha_n) \to 0^+ \text{ such that } x_0 + \alpha_n d_n \in S \text{ for all } n\}.$$

<u>THEOREM 4.1</u>. *Let* $x_0 \in S$ *be a local minimum of* f *on* S, *let* T *be a closed convex cone included in* $T(S;x_0)$. *Then :*

$$o \in [\partial g(x_0) + T^\circ] \underset{-}{*} \partial h(x_0) . \tag{4.10}$$

Let now N *be a closed convex cone included in* $T(S;x_0)^\circ$. *If*

$$o \in \mathrm{int}\{[\partial g(x_0) + N] \underset{-}{*} \partial h(x_0)\}, \tag{4.11}$$

then x_0 *is a strict local minimum of* f *on* S.

Proof. If x_0 is a local minimum of f on S, then $f'(x_0;d) \geq 0$ for all $d \in T(S;x_0)$. This implies that :

$$f'(x_0;d) + \psi_T(d) \geq 0 \text{ for all } d,$$

where $\psi_T = \psi^*_{T^\circ}$ is the indicator function of T (or the support function of T°). Consequently

$$g'(x_0;d) + \psi^*_{T^\circ}(d) \geq h'(x_0;d) \text{ for all } d.$$

In equivalent terms,

$$\psi^*_{\{\partial g(x_o)+T^\circ\}}(d) \geq \psi^*_{\partial h(x_o)}(d) \text{ for all } d.$$

When the announced condition (4.10), by just extending the definition of $A \underline{*} B$ to the case where A is not bounded.

If $N \subset T(S;x_o)^\circ$ and if condition (4.11) is satisfied, we have :

$$g'(x_o;d) + \psi^*_{T(S;x_o)^\circ}(d) > h'(x_o;d) \text{ for all non-null } d,$$

so that

$$f'(x_o;d) > 0 \text{ for all non-null } d \text{ in } T(S;x_o). \tag{4.12}$$

Suppose there is a sequence $(x_n) \subset S$ converging to x_o, $x_n \neq x_o$, such that $f(x_n) \leq f(x_o)$. Subsequencing if necessary, we may suppose that

$$\frac{x_n - x_o}{\|x_n - x_o\|}$$

does converge to a limit $\overline{d}$, $\overline{d} \neq 0$. $\overline{d}$ belongs to $T(S;x_o)$ since

$$x_n = x_o + \|x_n - x_o\|\,\overline{d} \quad \text{converges to } x_o \text{ in S.}$$

The inequality $f(x_n) \leq f(x_o)$ makes that $f'(x_o;\overline{d}) \leq 0$, which is in contradiction with (4.12). □

IV.2. *Duality*

Given $f = g-h \in DC(\mathbb{R}^n)$, the conjugate f^* of f can be expressed in terms of the conjugate of the (convex) functions g and h.. Duality schemes involving d.c. functions actually take root in this basic formula we state now.

THEOREM 4.2. *The conjugate of $f = g-h$ is given as :*

$$\forall\, x^* \in \mathbb{R}^n \quad f^*(x^*) = \sup_{y^* \in \operatorname{dom} h^*} \{g^*(x^*+y^*) - h^*(y^*)\}\,, \tag{4.13}$$

where $\operatorname{dom} h^* = \{x^* \mid h^*(x^*) < +\infty\}$.

Observe the symmetry with the formula giving the conjugate of (g + h) :

$$(g+h)^*(x^*) = \inf_{y^* \in \mathrm{dom}\, h^*} \{g^*(x^*-y^*) + h^*(y^*)\} . \tag{4.14}$$

The two formulae (4.13) and (4.14) overlap only when h is an affine function.

Since $-f^*(o) = \inf_{x \in \mathbb{R}^n} f(x)$, we deduce from (4.13) :

$$\inf_{x \in \mathbb{R}^n} \{g(x) - h(x)\} = \inf_{x^* \in \mathrm{dom}\, h^*} \{h^*(x^*) - g^*(x^*)\}. \tag{4.15}$$

By just using that $\sup(\cdot) = -\inf(-\cdot)$, we also derive :

$$\sup_{x \in \mathbb{R}^n} \{g(x) - h(x)\} = \sup_{x^* \in \mathrm{dom}\, g^*} \{h^*(x^*) - g^*(x^*)\} . \tag{4.16}$$

Theorem 4.2. is due to PSHENICHNYI ([38]). As an application, he proved that the conjugate of the difference of two support functions, $\psi_A^* - \psi_B^*$, is precisely the indicator function of $A \underline{*} B$. The result (4.13) remains true when g is an arbitrary function taking possibly the value $+\infty$ (cf. [19] for a short proof) ; this is of importance in order to include the constraint in the objective function when building up associated dual problems.

As for an example, consider the problem of maximizing a convex function h over a convex set S :

$$\alpha = \sup_{x \in S} h(x). \tag{4.17}$$

We clearly have that :

$$-\alpha = \inf_{x \in \mathbb{R}^n} \{\psi_S(x) - h(x)\}.$$

It comes from (4.15) that

$$-\alpha = \inf_{x^* \in \mathrm{dom}\, h^*} \{h^*(x^*) - \psi_S^*(x^*)\}. \tag{4.18}$$

To go further in this example, suppose $h(x) = \frac{1}{2} < Ax,x >$, where A is a symmetric positive definite (n,n) matrix and S the euclidean unit ball. Then 2α is the largest eigenvalue λ_M of A and we deduce from (4.18) the dual formulation announced in Example 0.3, that is :

$$- \alpha = - \lambda_M/2 = \inf_{x^* \in \mathbb{R}^n} \{ \|x^*\| - \frac{1}{2} < A^{-1} x^*, x^*>\} . \qquad (4.19)$$

Another variational formulation for λ_M is :

$$\alpha = \lambda_M/2 = \sup \{ \frac{1}{2} \|x\|^2 \ ; <A^{-1}x, x> \leq 1\} .$$

$\frac{1}{2}\|\cdot\|^2$ equals its conjugate function while the support function of the elliptic set $\{x | <A^{-1}x,x> \leq 1\}$ is $x^* \to \sqrt{< Ax^*,x^*>}$. Consequently, we infer from (4.18) the alternate formulation (0.6) given in example 0.3..

In successive papers ([49], [50], [51]), TOLAND proved and exploited thoroughly the equality (4.15). The mapping which assigns $h^* - g^*$ to $g - h$ is sometimes referred to as *TOLAND's involution*. The duality schemes TOLAND proposed in the context of Calculus of Variations have been further developed by AUCHMUTY ([6]). Relationships with LEGENDRE's transformation are laid out in EKELAND's papers ([15], [16]). A formula like (4.13) in its most general and abstract setting has recently been derived by the author ([24]).

A further interesting consequence of relation (4.15) concerns with the *regularization* of f. Given the convex function g or h, a natural way of regularizing (and, possibly, "smoothing") g is by performing the infimal convolution of g with some "kernel" θ :

$$g \nabla \theta : x \to (g \nabla \theta)(x) = \inf_{y \in \mathbb{R}^n} \{g(y) + \theta(x-y)\} . \qquad (4.20)$$

Two important examples are :

$$g_r = g \nabla \frac{1}{2r} \|\cdot\|^2, \ r > 0 \quad ;$$
$$\overline{g}_r = g \nabla \frac{1}{r} \|\cdot\| \ , \ r > 0 \qquad (4.21)$$

Each of these regularization procedures has its own advantages. The first one, widely used in Nonlinear Analysis (the so-called MOREAU-YOSIDA regularization scheme) or Convex Optimization (algorithms based upon the so-called proximal mapping), gives rise to a $C^{1,1}$ function g_r which coincides with g only at minimumpoints. The second one is also of interest since it gives rise to a Lipschitz function (for r small enough) which coincides with g at those points x where $\partial g(x)$ and the ball $\overline{B}(o,1/r)$ overlap. Since $(g \nabla \theta)^* = g^* + \theta^*$ in the examples we have considered, we deduce from (4.15) that :

$$\inf_{x \in \mathbb{R}^n} f(x) = \inf_{x \in \mathbb{R}^n} \{g(x) - h(x)\} = \inf_{x \in \mathbb{R}^n} \{g_r(x) - h_r(x)\}. \qquad (4.22)$$

$$= \inf_{x \in \mathbb{R}^n} \{\overline{g}_r(x) - \overline{h}_r(x)\}. \qquad (4.23)$$

The relationship (4.22) was noticed by GABAY ([20]) who used it for algorithmic purposes.

IV.3. *Comparing critical points of $g - h$ and $h^* - g^*$*

The first question which arises concerning critical points of $f = g-h$ is that of the *definition* itself. According to TOLAND ([49], [51]), x_0 is a critical point of f if $\partial g(x_0) \cap \partial h(x_0) \neq \phi$ (cf. (4.1) and (4.2)). This definition is not quite satisfactory since, as stated earlier, it depends on the decomposition of f as a difference of convex functions g and h. Another definition, drawing inspiration from (4.7), comes as follows ([18, ch. II]) :

DEFINITION 4.3. *x_0 is said to be a lower critical point of $f = g - h$ (resp. an upper critical point of f) if :*

$$o \in \partial g(x_0) \overset{*}{-} \partial h(x_0) \quad (\textit{resp. } o \in \partial h(x_0) \overset{*}{-} \partial g(x_0)).$$

This definition is stringent ; it however does not depend on g and h in the decomposition $g - h$ of f. Apart from the case where f is differentiable, either $\partial g(x_0) \overset{*}{-} \partial h(x_0)$ or $\partial h(x_0) \overset{*}{-} \partial g(x_0)$ is empty.

Quite interesting relationships between critical points of $g-h$ and $h^* - g^*$ have been proved by TOLAND ([49], [51]). We state here an example of such comparison results.

THEOREM 4.4. *If x_0 is a critical point of $g - h$, then every $x_0^* \in \partial g(x_0) \cap \partial h(x_0)$ is a critical point of $h^* - g^*$ and $g(x_0) - h(x_0) = h^*(x_0^*) - g^*(x_0^*)$.*

If x_0 is a global minimum of $g - h$, then every $x_0^* \in \partial h(x_0)$ is a global minimum of $h^* - g^*$.

$h^* - g^*$ is unambiguously defined by adopting the addition rule $\infty - \infty = \infty$.

Counterparts with lower (or upper) critical points can be derived to a certain extent. Usable results however do not go beyond TOLAND's ones.

Counterexample (PENOT).

Let $f \in DC(\mathbb{R}^n)$ be defined as the difference of the support functions of A and B :

$$f = \psi_A^* - \psi_B^*$$
$$= g - h ,$$

with $A = [-1,1] \times \{o\}$ and B the closed unit ball in $\mathbb{R}^n$. $x_0 = (\alpha,o), \alpha > o$, is a lower critical point of f since $\partial g(x_0) = \partial h(x_0) = \{(1,o)\}$. But for $x_0^* = (1,o)$, $\partial g^*(x_0^*) = \mathbb{R}_+ \times \mathbb{R}$ and $\partial h^*(x_0^*) = \mathbb{R}_+ \times \{0\}$. So, x_0^* is a critical point of $h^* - g^*$ in TOLAND's sense but is no more a lower critical point.

Further classifications of critical points using "second-order information" on g and h rely on MORSE theory for d.c. functions. For that, peruse [7] and [37].

V. PREVIEW ON MINIMIZATION PROCEDURES FOR D.C. FUNCTIONS

Minimizing (globally) a d.c. function is strongly related to some other difficult optimization problems like that of finding the global maximum of a convex function over a convex set. We have seen in IV.2. how the problem of *maximizing a convex function over a convex set* could be transformed into that of *minimizing a d.c. function over the whole space*. Conversely, consider the following d.c. minimization problem :

$$(\mathcal{P}) \text{ Minimize } f(x) = g(x) - h(x) \text{ over } S,$$

where $g, h \in \mathrm{Conv}(\mathbb{R}^n)$ and S is convex.

Define $\overline{g}, \overline{h} : \mathbb{R}^{n+1} \to \mathbb{R}$ by $\overline{g}(x,\xi) = g(x) - \xi$ and $\overline{h}(x,\xi) = h(x) - \xi$. Consider now :

$$(\mathcal{Q}) \text{ Maximize } \overline{h}(x,\xi) \text{ subject to } \overline{g}(x,\xi) \le 0 \text{ and } (x,\xi) \in S \times \mathbb{R}.$$

$(\mathcal{Q})$ is a maximization problem of a convex function over a convex set. $(\mathcal{P})$ and $(\mathcal{Q})$ are related in the following way : *$(x_o, \xi_o = g(x_o))$ is a solution of $(\mathcal{Q})$ if and only if x_o is a solution of $(\mathcal{P})$.*

A further example giving rise to a d.c. minimization problem comes from *fractional programming*. Consider the so-called "convex-convex fractional program" :

$$(\mathcal{F})\ \text{Maximize}\ \ q(x) = \frac{g(x)}{h(x)}\ \text{over}\ S,$$

where both g and h are convex (h strictly positive) and S is a compact convex set. An old approach in fractional programming (DINKELBACH, 1967) consists in transforming $(\mathcal{F})$ into a parametric d.c. program :

$$(\mathcal{F}_r)\ \text{Minimize}\ \{rh(x) - g(x)\}\ \ \text{over}\ S,\ r \in \mathbb{R}.$$

Denote by $\overline{r}$ the unique zero of the strictly increasing function

$$r \to \overline{q}(r) = \text{Min}\ \{rh(x) - g(x) \mid x \in S\}\ .$$

Then, *a solution of $(\mathcal{F}_r)$ with $r = \overline{r}$ is also a solution of $(\mathcal{F})$.*

Convexity is present twice (in g and h) for any decomposition of a d.c. function $f = g - h$. One may think of using alternately $\partial g(\cdot)$ and $\partial h(\cdot)$ to generate a sequence (x_n) which would converge to some local (global ?) minimum of f. Most of the attempts using this "first-order information" $\partial g(x_n)$, $\partial h(x_n)$ at the current point x_n lead to sequences which converge to stationary or critical points (of some kind) of f. Quite a little has been done to devise algorithms for a *global* minimization of d.c. functions ([35], [52]). This is a promising and important area of research, although one should not expect miracles...

CONCLUSION

Concerning the analysis and the optimization of d.c. functions, the main contributions are either old (around the thirties for Analysis) or quite recent (1979 - now for what concerns Optimization). The present survey reflects the thinking of the author while guiding ELLAIA's thesis ([18]).

It seems to us there is a growing interest in studying d.c. minimization problems, the main incentive coming from modelling in Applied

Mathematics. That induced H. TUY and the author to undertake the edition of a collection of research works in this field. That should be viewed as a natural extension of the present survey.

REFERENCES

[1] A.D. ALEXANDROFF, *Almost everywhere existence of the second differential of a convex function and some properties of convex surfaces connected with it,* Ucenye Zapiski Leningr. Gos. Univ. Ser. Mat. 37 n° 6, (1939) 3-35 (in Russian)

[2] A.D. ALEXANDROFF, *On surfaces represented as the difference of convex functions,* Izv. Akad. Nauk Kaz. SSR 60, Ser. Mat. Mekh. 3, (1949) 3-20 (in Russian).

[3] A.D. ALEXANDROFF, *Surfaces represented by the differences of convex functions,* Dokl. Akad. Nauk SSSR 72, (1959) 613-616 (in Russian).

[4] G. ARSOVE, *Functions representable as the difference of subharmonic functions,* Transactions Amer. Math. Soc. 75, (1953) 327-365.

[5] J-P. AUBIN, *Lipschitz behavior of solutions to convex minimization problems,* Math. of Operations Research 9, n° 1 (1984) 87-111.

[6] G. AUCHMUTY, *Duality for nonconvex variational principles,* J. of Differential Equations 50 (1983), 80-145.

[7] M. BOUGEARD, *Contribution à la théorie de Morse en dimension finie,* Thèse de 3ème cycle de l'Université de Paris IX, 1978.

[8] L.N. BRYSGALOVA, *Singularities of max of functions depending on parameters,* Funct. Anal. Appl. 11, 1 (1977), 59-60.

[9] L.N. BRYSGALOVA, *On max functions of families depending on parameters,* Funct. Anal. Appl. 12, 1 (1978), 66-67.

[10] H. BUSEMANN, *Convex surfaces,* Interscience Tracts in Pure and Applied Mathematics, 1958.

[11] F.H. CLARKE, *Nonsmooth analysis and optimization,* J. Wiley Interscience, 1983.

[12] V.F. DEMYANOV and L.N. POLYAKOVA, *Conditions for minimum of a quasi-differentiable function on a quasi-differentiable set,* U.S.S.R. Comput. Math. Phys. 20, (1981) 34-43.

[13] V.F. DEMYANOV and A.M. RUBINOV, *On quasi-differentiable mappings*, Math. Operationsforsch. u. Stat. ser. Optimization 14, (1983) 3-21.

[14] R.M. DUDLEY, *On second derivatives of convex functions*, Math. Scand. 41 (1977) 159-174 & 46 (1980) 61.

[15] I. EKELAND, *Legendre duality in nonconvex optimization and calculus of variations*, SIAM J. Control Optimization 15 (1977), 905-934.

[16] I. EKELAND, *Nonconvex duality*, Bull. Soc. Math. France, Mémoire 60 (1979), 45-55.

[17] I. EKELAND and J.-M. LASRY, *Problèmes variationnels non convexes en dualité*, C.R. Acad. Sc. Paris, t. 291 (1980), 493-496.

[18] R. ELLAIA, *Contribution à l'analyse et l'optimisation de différences de fonctions convexes*, Thèse de 3ème cycle de l'Université Paul Sabatier, 1984.

[19] R. ELLAIA and J.-B. HIRIART-URRUTY, *The conjugate of the difference of convex functions*, to appear in J. of Optim. Theory and Applications.

[20] D. GABAY, *Minimizing the difference of two convex functions : Part I : Algorithms based on exact regularization*, Working paper, I.N.R.I.A. (1982).

[21] P. HARTMAN, *On functions representable as a difference of convex functions*, Pacific J. Math. 9, (1959) 707-713.

[22] B. HERON and M. SERMANGE, *Non-convex methods for computing free boundary equilibria of axially symmetric plasmas*, Appl. Math. Optimization 8 (1982), 351-382.

[23] J.-B. HIRIART-URRUTY, *The approximate first-order directional derivatives for a convex function* in "Mathematical Theories of Optimization" Lecture notes in Mathematics 979, (1983) 144-177.

[24] J.-B. HIRIART-URRUTY, *A general formula on the conjugate of the difference of functions*, Séminaire d'Analyse Numérique, Université Paul Sabatier (1984).

[25] J.-B. HIRIART-URRUTY, *Miscellanies on the analysis and optimization of nonsmooth functions*, to appear.

[26] R. JANIN, *Sur la dualité et la sensibilité dans les problèmes de programmation mathematique*, Thèse Université Paris IX (1974).

[27] R. JANIN, *Sur des multiapplications qui sont des gradients généralisés*, Note aux C.R.A.S. Paris 294, (1982) 115-117.

[28] E.M. LANDIS, *On functions representable as the difference of two convex functions*, Dokl. Akad. Nauk SSSR 80, 1 (1951), 9-11.

[29] C. MALIVERT, *Méthodes de descente sur un fermé non convexe*, Bull. de la Soc. Math. de France, Mémoire n° 60, (1979) 113-124.

[30] J.N. MATHER, *Distance from a manifold in Euclidean space*, Proc. of Symp. in Pure Math., Vol. 40, Part 2 (1983), 199-216.

[31] V.I. MATOV, *Topological classification of the germs of max and minimax of the families in general position*, Russian Math. Surv. 37, 4 (1982), 167-168.

[32] V.I. MATOV, *Singularities of max functions on the manifold with boundary*, Trudy Sem. I.G. Petrovskogo, Mosk. Gos. Univ. 6 (1981), 195-222.

[33] D. MELZER, *Expressibility of piecewise linear continuous functions as a difference of two piecewise linear convex functions*, preprint Humboldt Universität Berlin (1984).

[34] F. MIGNOT, *Contrôle dans les inéquations variationnelles elliptiques*, J. of Funct. Analysis 22, (1976) 130-185.

[35] J. NOAILLES, *Une méthode d'optimisation globale en programmation non convexe*, Colloque National d'Analyse Numérique de Gouvieux, 1980.

[36] J.-P. PENOT, *Calcul sous-différentiel et optimisation*, J. of Funct. Analysis 27, (1978) 248-276.

[37] A. POMMELLET, *Analyse convexe et théorie de Morse*, Thèse de 3ème cycle de l'Université de Paris IX, 1982.

[38] B.N. PSHENICHNYI, in *Contrôle optimal et Jeux différentiels*, Cahiers de l'I.R.I.A. n° 4 (1971).

[39] B.N. PSHENICHNYI, *Necessary conditions for an extremum*, Marcel Dekker N.Y., 1971.

[40] A.W. ROBERTS and D.E. VARBERG, *Convex functions*, Academic Press, 1973.

[41] R.T. ROCKAFELLAR, *Convex analysis*, Princeton University Press, 1970.

[42] R.T. ROCKAFELLAR, *Favorable classes of Lipschitz continuous functions in subgradient optimization*, *in* "Progress in nondifferentiable optimization", E. Nurminskii, ed., Pergamon Press (1981).

[43] R.T. ROCKAFELLAR, *Maximal monotone relations and the second derivatives of nonsmooth functions*, preprint 1984.

[44] A. SHAPIRO and Y. YOMDIN, *On functions representable as a difference of two convex functions and necessary conditions in a constrained optimization*, preprint Ben-Gurion University of the Negev (1982).

[45] A. SHAPIRO, *On functions representable as a difference of two convex functions in inequality constrained optimization*, Research report University of South Africa (1983).

[46] A. SHAPIRO, *On optimality conditions in quasidifferentiable optimization*, SIAM J. Control and Optimization 22, 4 (1984) 610-617.

[47] R. SCHNEIDER, *Boundary structure and curvature of convex bodies* in "Contributions to Geometry", J. Tölke and J.M. Wills ed., Birkhäuser Verlag, 1979.

[48] J. SPINGARN, *Submonótone subdifferentials of Lipschitz functions*, Trans. Amer. Math. Soc. 264, (1981) 77-89.

[49] J.F. TOLAND, *Duality in nonconvex optimization*, J. Math. Analysis and Applications 66 (1978), 399-415.

[50] J.F. TOLAND, *A duality principle for nonconvex optimization and the calculus of variations*, Arch. Rational Mech. Anal. 71 (1979), 41-61.

[51] J.F. TOLAND, *On subdifferential calculus and duality in nonconvex optimization*, Bull. Soc. Math. France, Mémoire 60 (1979), 177-183.

[52] H. TUY, *Global minimization of a difference of two convex functions*, to appear.

[53] J.-P. VIAL, *Strong and weak convexity of sets and functions*, Math. of Operations Research 8, 2 (1983), 231-259.

[54] Y. YOMDIN, *On functions representable as a supremum of a family of smooth functions*, SIAM J. Math. Analysis 14, 2 (1983), 239-246.

[55] Y. YOMDIN, *On functions representable as a supremum of a family of smooth functions, II*, to appear.

[56] Y. YOMDIN, *Maxima of smooth families III : Morse-Sard theorem*, preprint 1984.

[57] Y. YOMDIN, *On representability of convex functions as maxima of linear families*, preprint 1984.

[58] V.A. ZALGALLER, *On the representation of a function of two variables as the difference of convex functions*, Vestn. Leningrad Univ. Ser. Mat. Mekh. 18, (1963) 44-45 (in Russian).

GUIDE FOR A BIBLIOGRAPHY

Examples of d.c. optimization problems : [20], [22], [50].

Characterization of d.c. functions, decomposition of d.c. functions : [2], [3], [18, ch. 2], [21], [28], [33], [40, ch. I], [44], [58].

Analysis and optimization of lower-C^k functions : [18, ch. I], [26], [27], [29], [42], [48], [53], [54], [55], [56].

Singularities of max and minimax functions : [8], [9], [30], [31], [32].

Second-order derivatives of convex functions ;

pointwise derivatives : [1], [5, §2], [10], [34, §1], [43], [47, § II.4] ;

distributional derivatives : [14] ;

approximate derivatives : [23].

First-order necessary conditions for optimality

for directionally differentiable functions : [25], [36], [39] ;

for tangentially d.c. functions : [12], [18, ch. 2], [45], [46].

Morse theory for d.c. functions and study of their critical points : [7], [37], [56].

Duality theory for d.c. minimization problems ; formulae and applications : [6], [15], [16], [17], [19], [24], [38], [49], [50], [51].

Algorithms for minimizing d.c. functions : [20], [22], [35], [52].