Journal of Convex Analysis
Volume 17 (2010), No. 2, 673–680

Necessary Conditions for Local Optimality in Difference-of-Convex Programming

Immanuel M. Bomze
Dept. of Statistics, University of Vienna,
Brünner Str. 72, 1210 Vienna, Austria
immanuel.bomze@univie.ac.at

Claude Lemaréchal
INRIA Grenoble-Rhône-Alpes, 655 Avenue de l'Europe,
38334 Saint Ismier - Montbonnot, France
Claude.Lemarechal@inria.fr

Received: November 11, 2008
Revised manuscript received: June 4, 2009

Using ε-subdifferential calculus for difference-of-convex (d.c.) optimization, Dür proposed a condition sufficient for local optimality, and showed that this condition is not necessary in general. Here it is proved that whenever the convex part is strongly convex, this condition is also necessary. Strong convexity can always be ensured by changing the given d.c. decomposition slightly. This approach also allows for a formulation with perturbed ε-subdifferentials which involves only the original d.c. decomposition, even without imposing strong convexity. We relate this result with another inclusion condition on perturbed ε-subdifferentials, which even can serve as a quantitative version of a criterion both necessary and sufficient for local optimality.

Keywords: Approximate subdifferential, non-smooth optimization, optimality condition, strong convexity

1. Introduction

The difference-of-convex (d.c.) paradigm in continuous optimization can be traced back to [6, 9, 10, 11] and has been proven useful in a number of areas in optimization [5]. The underlying mathematical concept is much older and goes at least back to Alexandroff's paper from 1950, see [1] and also [4], where E. G. Straus is credited for coining the abbreviation "d.c.".

The class of functions f which admit a d.c. decomposition $f = g - h$, where $g : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ and $h : \mathbb{R}^n \to \mathbb{R}$ are convex functions, has therefore been studied quite extensively. While this class is large (it contains the twice continuously differentiable functions) the relevance of theoretical results for algorithmic applications depends on the choice of the functions g and h in the d.c. decomposition. Still, issues in the context of optimality conditions remained open, and this note tries to answer some of these questions, by using ε-subdifferential calculus.

Given an extended-valued closed (i.e., lower semicontinuous) convex function $g : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ and $\varepsilon \geq 0$, the ε-subdifferential of g at $\bar{x} \in \mathbb{R}^n$ with $g(\bar{x}) < +\infty$ is defined

ISSN 0944-6532 / \$ 2.50

as

$$\partial_\varepsilon g(\bar{x}) = \{y \in \mathbb{R}^n : g(x) - g(\bar{x}) \geq y^\top (x - \bar{x}) - \varepsilon \text{ for all } x \in \mathbb{R}^n\}. \tag{1}$$

In terms of ε-subdifferentials, we have the following characterization of global optimality, see [7] or [8, p. 101]:

Theorem 1.1. *Suppose that $\bar{x}$ satisfies $g(\bar{x}) < +\infty$. Then $\bar{x}$ is a global minimizer of $f = g - h$ if and only if*

$$\partial_\varepsilon h(\bar{x}) \subseteq \partial_\varepsilon g(\bar{x}) \quad \textit{for all } \varepsilon \geq 0. \tag{2}$$

The situation for local optimality is different: while Dür has established the following sufficient condition in [3], she also has specified a counterexample there which shows that this condition is not necessary.

Theorem 1.2. *Suppose that $\bar{x}$ satisfies $g(\bar{x}) < +\infty$. Then $\bar{x}$ is a local minimizer of $f = g - h$ if for some $\delta > 0$*

$$\partial_\varepsilon h(\bar{x}) \subseteq \partial_\varepsilon g(\bar{x}) \quad \textit{for all } \varepsilon \textit{ with } 0 \leq \varepsilon \leq \delta. \tag{3}$$

Recently [2, Thm. 8 in §1.1.5], it has been shown that (3) is necessary and sufficient in case of (possibly indefinite) quadratic programming over a polyhedron M, given a natural quadratic d.c. decomposition of $f(x) = \frac{1}{2}x^\top Qx + c^\top x$ as follows: decompose $Q = P_+ - P_-$ with $P_\pm$ positive-definite, and consider $h(x) = \frac{1}{2}x^\top P_- x - c^\top x$ as well as $g(x) = \frac{1}{2}x^\top P_+ x + \mathrm{i}_M(x)$ where i_M denotes the indicator function

$$\mathrm{i}_M(x) = \begin{cases} 0, & \text{if } x \in M, \\ +\infty, & \text{else.} \end{cases}$$

Since P_+ is positive-definite, the function $g(x) - \rho\|x - \bar{x}\|^2$ is still convex for $\rho > 0$ small enough. It will turn out below that the latter property also guarantees necessity of Dür's condition (3) in the general case.

2. The role of the violating parameter region

In light of the preceding discussion, let us denote by

$$G(\bar{x}) = \{\varepsilon > 0 : \partial_\varepsilon h(\bar{x}) \not\subseteq \partial_\varepsilon g(\bar{x})\}$$

the *violating parameter region* where the required inclusion of (3) does not hold. It was unclear for a while whether the violating parameter region $G(\bar{x})$ is convex, i.e., an interval. This has been answered recently in the negative by an example [2, Section 1.1.4] where $G(\bar{x})$ can have up to n connected components if $\bar{x} \in \mathbb{R}^n$ is a local minimizer of $g - h$.

Next we proceed to relate $G(\bar{x})$ to the question of finding an improving point. To this end, recall [8, Theorem XI.2.1.1] that the value of the support functional to the convex set $\partial_\varepsilon g(\bar{x})$ at any direction d is given by the ε-directional derivatives w.r.t. d,

$$g'_\varepsilon(\bar{x}; d) = \inf_{t>0} \frac{\phi_d(t) + \varepsilon}{t} \quad \text{with } \phi_d(t) = g(\bar{x} + td) - g(\bar{x}), \tag{4}$$

and similarly for $\partial_\varepsilon h(\bar{x})$,

$$h'_\varepsilon(\bar{x};d) = \inf_{t>0} \frac{\psi_d(t)+\varepsilon}{t} \quad \text{with } \psi_d(t) = h(\bar{x}+td) - h(\bar{x})\,. \tag{5}$$

So, whenever $\varepsilon \in G(\bar{x})$, there must be a direction d such that $h'_\varepsilon(\bar{x};d) > g'_\varepsilon(\bar{x};d)$ and vice versa; indeed, d can be chosen as a normal direction of a hyperplane separating $\partial_\varepsilon g(\bar{x})$ from a point in $\partial_\varepsilon h(\bar{x}) \setminus \partial_\varepsilon g(\bar{x})$.

The following result can be found in [3]. For the sake of being self-contained, we provide a short proof.

Theorem 2.1. *If $x \in \mathbb{R}^n$ satisfies $g(x) - h(x) < g(\bar{x}) - h(\bar{x}) < +\infty$ and $y \in \partial_0 h(x)$, then*

$$\varepsilon = h(\bar{x}) - h(x) - y^\top(\bar{x} - x) \in G(\bar{x})\,.$$

Conversely, if $\varepsilon \in G(\bar{x})$ and $h'_\varepsilon(\bar{x};d) > g'_\varepsilon(\bar{x};d)$, then d is an improving direction: there is a $t > 0$ such that $x = \bar{x} + td$ satisfies $g(x) - h(x) < g(\bar{x}) - h(\bar{x})$.

Proof. First we note that the specified y and ε satisfy $y \in \partial_\varepsilon h(\bar{x})$. Indeed, for any z, we get

$$\begin{aligned} h(z) - h(\bar{x}) &= h(z) - h(x) + h(x) - h(\bar{x}) \\ &\geq y^\top(z-x) + h(x) - h(\bar{x}) = y^\top(z-\bar{x}) - \varepsilon\,. \end{aligned}$$

On the other hand, we know

$$g(x) - g(\bar{x}) < h(x) - h(\bar{x}) = y^\top(x - \bar{x}) - \varepsilon\,,$$

so that $y \notin \partial_\varepsilon g(\bar{x})$, and hence $\varepsilon \in G(\bar{x})$.

Conversely, $g'_\varepsilon(\bar{x};d) < h'_\varepsilon(\bar{x};d)$ implies the existence of $t > 0$ such that

$$\frac{\phi_d(t)+\varepsilon}{t} < \frac{\psi_d(t)+\varepsilon}{t}\,.$$

Then (4) and (5) imply $g(x) - g(\bar{x}) < h(x) - h(\bar{x})$, which completes the proof. □

3. Dür's delta and the main result

Now define *Dür's delta* as $\delta(\bar{x}) = \inf G(\bar{x})$. Suppose $\bar{x}$ is a local, nonglobal solution. Then Theorem 2.1 gives an upper bound for $\delta(\bar{x})$ if we know an improving feasible solution. Conversely, this result suggests that searching for ε beyond $\delta(\bar{x}) < +\infty$ may be a good strategy for escaping from nonglobal optima. Further, Theorems 1.1 and 1.2 state

$$\begin{aligned} \delta(\bar{x}) = \infty \quad &\Longleftrightarrow \quad \bar{x} \text{ is a global minimizer of } g - h\,; \\ \delta(\bar{x}) > 0 \quad &\Longrightarrow \quad \bar{x} \text{ is a local minimizer of } g - h\,, \end{aligned}$$

while the following result will specify a condition on g such that the last implication arrow can be reverted, too.

Theorem 3.1. *Let $\bar{x}$ be a local minimizer of $f = g - h$:*

$$g(x) - g(\bar{x}) \geq h(x) - h(\bar{x}) \quad \text{for all } x \text{ with } \|x - \bar{x}\| \leq \eta \tag{6}$$

and suppose that g is strongly convex, i.e., $g(x) - \rho\|x - \bar{x}\|^2$ is still convex for some small $\rho > 0$. Then

$$\partial_\varepsilon h(\bar{x}) \subseteq \partial_\varepsilon g(\bar{x}) \quad \text{for all non-negative } \varepsilon \leq \rho\eta^2 .$$

In other words, $\delta(\bar{x}) \geq \rho\eta^2$.

Proof. For a direction d with $\|d\| = \eta$, we again consider the ε-directional derivatives w.r.t. d as in (4) and (5). Obviously,

$$h'_\varepsilon(\bar{x}; d) = \inf_{t>0} \frac{\psi_d(t) + \varepsilon}{t} \leq \inf_{t\in\,]0,1]} \frac{\psi_d(t) + \varepsilon}{t} \leq \inf_{t\in\,]0,1]} \frac{\phi_d(t) + \varepsilon}{t} \tag{7}$$

and we claim that

$$h'_\varepsilon(\bar{x}; d) \leq g'_\varepsilon(\bar{x}; d) \quad \text{for all } \varepsilon \leq \rho\eta^2 . \tag{8}$$

For this purpose introduce the convex function $\hat{g}(x) = g(x) - \rho\|x - \bar{x}\|^2$. Then

$$\frac{\phi_d(t) + \varepsilon}{t} = \frac{\hat{g}(\bar{x} + td) - \hat{g}(\bar{x})}{t} + \frac{\rho\eta^2 t^2 + \varepsilon}{t} = \frac{\hat{\phi}_d(t)}{t} + t\rho\eta^2 + \frac{\varepsilon}{t} ,$$

with the obvious notation $\hat{\phi}_d(t) = \hat{g}(\bar{x}+td)-\hat{g}(\bar{x})$. The minimum of the (strictly convex) function $t\rho\eta^2 + \frac{\varepsilon}{t}$ is attained at $t_o = \sqrt{\frac{\varepsilon}{\rho\eta^2}}$. By convexity of $\hat{g}$, the difference quotient $\frac{\hat{\phi}_d(t)}{t}$ is increasing in t. Altogether, $\frac{\phi_d(t)+\varepsilon}{t}$ attains its minimum at some $t_\varepsilon \leq t_o \leq 1$ if $\varepsilon \leq \rho\eta^2$. We therefore see from (7) that, when $\varepsilon \leq \rho\eta^2$, then

$$h'_\varepsilon(\bar{x}; d) \leq \inf_{t\in\,]0,1]} \frac{\phi_d(t) + \varepsilon}{t} = g'_\varepsilon(\bar{x}; d)$$

and our claim (8) is established. Hence at d, the support functional of $\partial_\varepsilon h(\bar{x})$ does not exceed that of $\partial_\varepsilon g(\bar{x})$. Because the (normalized) direction d was arbitrary, this just means $\partial_\varepsilon h(\bar{x}) \subseteq \partial_\varepsilon g(\bar{x})$ which is what we wanted to prove. □

4. Perturbations and local optimality conditions

As already mentioned, local optimality of $\bar{x}$ does not imply (3) with $\delta > 0$. On the other hand, it does imply inclusions resembling it, with perturbed forms of approximate subdifferentials.

For a motivation, recall the familiar condition

$$\partial_0 h(\bar{x}) \subseteq \partial_0 g(\bar{x}) \tag{9}$$

necessary for $\bar{x}$ to be a local minimizer of $f = g - h$. Below we will specify conditions – (10), (12) and their consequences (13), (14) – involving explicitly the size η of the optimality neighborhood in (6), as well as an arbitrary parameter ε. All of these formulae yield the above condition (9) as a limiting case for $\varepsilon \searrow 0$; hence they may be viewed as a quantitative sharpening of (9). In what follows, $B_\tau = \{y \in \mathbb{R}^n : \|y\| \leq \tau\}$ will denote the ball of radius τ centered at the origin.

Our first optimality condition can be obtained by enforcing strong convexity via an alternative d.c. decomposition: setting $q_\rho(x) = \rho\|x - \bar{x}\|^2$, we have

$$f = g_\rho - h_\rho \quad \text{with } g_\rho = g + q_\rho \text{ and } h_\rho = h + q_\rho\,.$$

In order to apply Theorem 3.1 again, we need to express the ε-subdifferential of a sum such as g_ρ or h_ρ:

Lemma 4.1. *Consider two closed convex functions r and q such that $r(\bar{x}) < +\infty$ and $q(x) < +\infty$ for all $x \in \bar{x} + B_\tau$ for some $\tau > 0$. Then*

$$\partial_\varepsilon(r + q)(\bar{x}) = \bigcup_{\sigma \in [0,\varepsilon]} [\partial_\sigma r(\bar{x}) + \partial_{\varepsilon-\sigma} q(\bar{x})] \quad \textit{for all } \varepsilon \geq 0\,.$$

Proof. By assumption, $\{\bar{x}\} \subseteq \operatorname{dom} r$ and $\{\bar{x}\} + B_\tau \subseteq \operatorname{dom} q$. It follows that $-B_\tau \subseteq \operatorname{dom} r - \operatorname{dom} q$: the qualification condition $0 \in \operatorname{int}(\operatorname{dom} r - \operatorname{dom} q)$ holds and [8, Thms. X.2.3.2, XI.3.1.1] apply. □

Note that strong convexity of g is not needed in the next result.

Theorem 4.2. *Suppose* (6) *holds. Then for all $\varepsilon > 0$ and all $\delta \in [0, \varepsilon]$,*

$$\partial_\delta h(\bar{x}) + B_{\frac{2}{\eta}\sqrt{(\varepsilon-\delta)\varepsilon}} \subseteq \bigcup_{\sigma \in [0,\varepsilon]} \left[\partial_\sigma g(\bar{x}) + B_{\frac{2}{\eta}\sqrt{(\varepsilon-\sigma)\varepsilon}}\right]. \tag{10}$$

Proof. Because g_ρ is strongly convex, we know from Theorem 3.1 that

$$\partial_\varepsilon h_\rho(\bar{x}) \subseteq \partial_\varepsilon g_\rho(\bar{x}) \quad \text{for all } \varepsilon \leq \rho\eta^2\,.$$

Now apply Lemma 4.1 twice with $q = q_\rho$, once combined with $r = h_\rho$ and once combined with $r = g_\rho$. We have for all $\varepsilon \in [0, \rho\eta^2]$ and all $\delta \in [0, \varepsilon]$

$$\partial_\delta h(\bar{x}) + \partial_{\varepsilon-\delta} q_\rho(\bar{x}) \subseteq \bigcup_{\sigma \in [0,\varepsilon]} [\partial_\sigma g(\bar{x}) + \partial_{\varepsilon-\sigma} q_\rho(\bar{x})]\,. \tag{11}$$

Next we use [8, Example XI.1.2.2]: for all $\tau \geq 0$

$$\partial_\tau q_\rho(\bar{x}) = \partial_\tau(\rho q_1)(\bar{x}) = \rho\,\partial_{\tau/\rho} q_1(\bar{x}) = \rho\, B_{2\sqrt{\tau/\rho}} = B_{2\sqrt{\tau\rho}}\,.$$

Hence (11) reads

$$\partial_\delta h(\bar{x}) + B_{2\sqrt{(\varepsilon-\delta)\rho}} \subseteq \bigcup_{\sigma \in [0,\varepsilon]} \left[\partial_\sigma g(\bar{x}) + B_{2\sqrt{(\varepsilon-\sigma)\rho}}\right].$$

This holds whenever $0 \leq \delta \leq \varepsilon \leq \rho\eta^2$. Now, given $\varepsilon > 0$, we rather define $\rho > 0$ such that these inequalities hold, namely $\rho = \varepsilon/\eta^2$ and we arrive at (10). □

Another inclusion can be obtained, in which the left-hand side just involves the approximate subdifferential itself. This inclusion turns out to be necessary *and* sufficient for local optimality.

Theorem 4.3. *The local optimality property* (6) *holds if and only if*

$$\partial_\varepsilon h(\bar{x}) \subseteq \bigcup_{\sigma\in[0,\varepsilon]} \left[\partial_\sigma g(\bar{x}) + B_{\frac{\varepsilon-\sigma}{\eta}}\right] \quad \textit{for all } \varepsilon > 0\,. \tag{12}$$

Proof. Let i_η be the indicator function of the ball $\bar{x} + B_\eta$ centered at $\bar{x}$ with radius η. Clearly, the local optimality property (6) holds if and only if $\bar{x}$ is a global minimizer of the d.c. function $f + \mathrm{i}_\eta = (g + \mathrm{i}_\eta) - h$ and Theorem 1.1 applies. Using Lemma 4.1 for the sum $g + \mathrm{i}_\eta$, we see that (6) holds if and only if

$$\partial_\varepsilon h(\bar{x}) \subseteq \bigcup_{\sigma\in[0,\varepsilon]} \left[\partial_\sigma g(\bar{x}) + \partial_{\varepsilon-\sigma}\mathrm{i}_\eta(\bar{x})\right] \quad \text{for all } \varepsilon > 0\,.$$

Subdifferentiating the indicator function of a ball is straightforward from the definition (1): we obtain $\partial_\tau \mathrm{i}_\eta(\bar{x}) = B_{\frac{\tau}{\eta}}$ and (12) is established. □

We stress that (10) and (12) hold for all positive ε. Fixing δ or σ in these formulae yields some simplifications which are worth mentioning. As it happens, large values of $\delta \le \varepsilon$ in (10) do not provide much information, as compared to (12); but setting $\delta = 0$ gives a definitely different inclusion, namely

$$\partial_0 h(\bar{x}) + B_{\frac{2\varepsilon}{\eta}} \subseteq \bigcup_{\sigma\in[0,\varepsilon]} \left[\partial_\sigma g(\bar{x}) + B_{\frac{2}{\eta}\sqrt{(\varepsilon-\sigma)\varepsilon}}\right] \quad \text{for all } \varepsilon > 0\,. \tag{13}$$

The left-hand side (lhs) of (13) is a perturbation of the lhs of (9). On the other hand, monotonicity arguments in (12) yield a really closed formula:

Corollary 4.4. *If $\bar{x}$ satisfies* (6), *then*

$$\partial_\varepsilon h(\bar{x}) \subseteq \partial_\varepsilon g(\bar{x}) + B_{\varepsilon/\eta} \quad \textit{for all } \varepsilon > 0\,. \tag{14}$$

Proof. Just observe in (12) that $\partial_\sigma g(\bar{x}) \subseteq \partial_\varepsilon g(\bar{x})$ and $B_{(\varepsilon-\sigma)/\eta} \subseteq B_{\varepsilon/\eta}$ for all $\sigma \in [0, \varepsilon]$. □

Finally, we illustrate the above four formulae on Dür's counterexample [3, Example 2.1]. Here $g(x) = [(x-1)_+]^2$ and $h(x) = [(-1-x)_+]^2$ where $t_+ = \max\{0, t\}$, so that $\bar{x} = 0$ is a local minimizer of $f = g - h$ with maximal $\eta = 1$ in (6). The approximate subdifferentials are

$$\partial_\sigma g(0) = [0, r(\sigma)] \quad \text{and} \quad \partial_\sigma h(0) = [-r(\sigma), 0]\,,$$

with $r(\sigma) := 2\sqrt{1+\sigma} - 2$. Note that r is concave with $r'(0) = 1$; this implies $r(\sigma) \le \sigma$ for all $\sigma \ge 0$. Hence lhs (14) $\subseteq [-\varepsilon, 0]$ while the right-hand side (rhs) of (14) for $\eta = 1$ is $[-\varepsilon, \varepsilon + r(\varepsilon)]$.

Setting $s(\delta) := 2\sqrt{(\varepsilon-\delta)\varepsilon}$, the sets involved in (10) are (with $\eta = 1$)

$$\begin{cases} \text{lhs}\,(10) = [-r(\delta) - s(\delta), s(\delta)]\,, \\ \text{rhs}\,(10) = \bigcup_{\sigma\in[0,\varepsilon]} [-s(\sigma), r(\sigma) + s(\sigma)]\,. \end{cases} \tag{15}$$

Note that s is concave with $s'(0) = -1$, hence decreasing; also $r + s$ is concave and decreasing, because $(r + s)'(0) = 0$. As a result,

$$\text{rhs}\,(10) = [-s(0), r(0) + s(0)] = [-2\varepsilon, 2\varepsilon]\,.$$

The same monotonicity arguments show that

$$\forall\delta \in [0, \varepsilon],\ \text{lhs}\,(10) \subseteq [-r(0) - s(0), s(0)] = [-2\varepsilon, 2\varepsilon]\,,$$

so that (10) does hold. Its consequence (13), obtained by fixing $\delta = 0$ in (15), is then an identity: $\text{lhs}\,(13) = [-2\varepsilon, 2\varepsilon] = \text{rhs}\,(10) = \text{rhs}\,(13)$.

Now we illustrate the necessary and sufficient character of (12), which must be true for $\eta \le 1$ and false for $\eta > 1$. Its left-hand side is $\partial_\varepsilon h(0) = [-r(\varepsilon), 0]$, while its right-hand side is

$$\bigcup_{\sigma\in[0,\varepsilon]} \left[\frac{\sigma-\varepsilon}{\eta}, r(\sigma) + \frac{\varepsilon-\sigma}{\eta}\right].$$

The right end-points of the segments in the above union are clearly non-negative, so (12) holds if and only if

$$r(\varepsilon) \le \frac{\varepsilon-\sigma}{\eta}\,,\quad \text{for all } \varepsilon > 0 \text{ and some } \sigma \in [0, \varepsilon],$$

which means

$$r(\varepsilon) \le \frac{\varepsilon}{\eta}\,,\quad \text{for all } \varepsilon > 0.$$

Remembering that $r(\varepsilon) \le \varepsilon$ and $\frac{r(\varepsilon)}{\varepsilon} \to r'(0) = 1$ as $\varepsilon \searrow 0$, this holds if and only if $\eta \le 1$.

Acknowledgements. The authors want to thank two anonymous referees and an associate editor for their careful reading and valuable suggestions.

References

[1] A. D. Alexandroff: Surfaces represented by the difference of convex functions, Dokl. Akad. Nauk SSSR, n. Ser. 72 (1950) 613–616.

[2] I.M. Bomze: Global optimization – a quadratic programming perspective, in: Nonlinear Optimization (Cetraro, 2007), I.M. Bomze et al. (eds.), Lect. Notes Math. 1989, Springer (2010), to appear.

[3] M. Dür: A parametric characterization of local optimality, Math. Methods Oper. Res. 57 (2003) 101–109.

[4] P. Hartman: On functions representable as a difference of convex functions, Pacific J. Math. 9 (1959) 707–713.

[5] R. Horst, N. V. Thoai: DC programming: overview, J. Optimization Theory Appl. 103 (1999) 1–43.

[6] J.-B. Hiriart-Urruty: Generalized differentiability, duality and optimization for problems dealing with differences of convex functions, in: Convexity and Duality in Optimization (Groningen, 1984), J. Ponstein (ed.), Lect. Notes Econ. Math. Syst. 256, Springer, Berlin (1985) 37–70.

[7] J.-B. Hiriart-Urruty: From convex optimization to nonconvex optimization. Part I: necessary and sufficient conditions for global optimality, in: Nonsmooth Optimization and Related Topics (Erice, 1988), F. C. Clarke, V. Demyanov, F. Gianessi (eds.), Plenum Press, New York (1989) 219–239.

[8] J.-B. Hiriart-Urruty, C. Lemaréchal: Convex Analysis and Minimization Algorithms II: Advanced Theory and Bundle Methods, Grundlehren der Mathematischen Wissenschaften 306, Springer, Berlin (1996).

[9] I. Singer: A Fenchel-Rockafellar type duality theorem for maximization, Bull. Aust. Math. Soc. 20 (1979) 193–198.

[10] J. F. Toland: Duality in nonconvex optimization, J. Math. Anal. Appl. 66 (1978) 399–415.

[11] J. F. Toland: A duality principle for nonconvex optimization and the calculus of variations, Arch. Rat. Mech. Anal. 71 (1979) 41–61.