Necessary conditions for local optimality in d.c. programming

Immanuel M. Bomze*

Claude Lemaréchal°

* University of Vienna, Austria

° INRIA Grenoble, France

Abstract

Using ε-subdifferential calculus for difference-of-convex (d.c.) programming, 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.

Key words: approximate subdifferential; non-smooth optimization; optimality condition; strong convexity

1 Introduction

The difference-of-convex (d.c.) paradigm in continuous global optimization goes back to [5] and has proven useful in a number of areas in optimization [3]. So the class of functions f which admit a d.c. decomposition $f(x) = g(x) - h(x)$, where $g : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ and $h : \mathbb{R}^n \to \mathbb{R}$ are convex functions, has been studied quite extensively. While this class is large (it contains the twice continuously differentiable functions), the relevance of theoretical results for algorithmic applications heavily 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 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 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 [4, 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 [2], 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 $\delta > 0$*

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

Recently [1, Theorem 8 in Section 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 $g(x) = \frac{1}{2}x^\top P_+ x + \Psi_M(x)$ where Ψ_M denotes the indicator function

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

Since P_+ is positive-definite, there is a (possibly small) $\rho > 0$ such that $g(x) - \rho\|x - \bar{x}\|^2$ is still convex. It will turn out below, that the latter property also guarantees necessity of Dür's condition 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 [1, 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 [4, 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,

$$g_\varepsilon(\bar{x}; d) = \inf_{t>0} \frac{\phi_d(t) + \varepsilon}{t} \quad \text{with} \quad \phi_d(t) = g(\bar{x} + td) - g(\bar{x})\,, \tag{4}$$

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

$$h_\varepsilon(\bar{x}; d) = \inf_{t>0} \frac{\psi_d(t) + \varepsilon}{t} \quad \text{with} \quad \psi_d(t) = h(\bar{x} + td) - h(\bar{x})\,. \tag{5}$$

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 [2]. 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

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

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, choose an appropriate $t > 0$ such that

$$\frac{\phi_d(t) + \varepsilon}{t} \approx g_\varepsilon(\bar{x}; d) < h_\varepsilon(\bar{x}; d) \leq \frac{\psi_d(t) + \varepsilon}{t}\,.$$

From (4) and (5) it follows $g(x) - g(\bar{x}) < h(x) - h(\bar{x})$, thus the claim follows.□

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 *Suppose that $\bar{x}$ is a local minimizer of $f(x) = g(x) - h(x)$ and further suppose that g is strongly convex, i.e., $g(x) - \rho\|x - \bar{x}\|^2$ is still convex for some small $\rho > 0$. Then there is some $\delta > 0$ such that* (3) *holds. In other words, $\delta(\bar{x}) > 0$. To be more precise, if*

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

then

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

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, 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 $t_o = \frac{\varepsilon}{\rho\eta^2}$. By convexity, the function $\frac{\hat{\phi}_d(t)}{t}$ 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

$$h_\varepsilon(\bar{x}; d) \leq \inf_{t\in\,]0,1]} \frac{\phi_d(t) + \varepsilon}{t} = g_\varepsilon(\bar{x}; d)\,;$$

claim (8) is proved. 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 that $\partial_\varepsilon h(\bar{x}) \subseteq \partial_\varepsilon g(\bar{x})$, which is what we wanted to prove. $\square$

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; we give two results along these lines.

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

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

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 *For any closed convex function r,*

$$\partial_\varepsilon(r + q_\rho)(\bar{x}) = \bigcup_{\delta\in[0,\varepsilon]} [\partial_\delta q_\rho(\bar{x}) + \partial_{\varepsilon-\delta} r(\bar{x})] \quad \textit{for all} \quad \varepsilon \geq 0\,. \tag{9}$$

Proof. Because q_ρ is a finite-valued function, [4, Thm.XI.3.1.1] applies. □

In what follows, $B_\tau = \{y \in \mathbb{R}^n : \|y\| \leq \tau\}$ will denote the ball of radius τ centered at the origin.

Theorem 4.1 *Suppose* (6) *holds. Then for all $\varepsilon > 0$ and all $\delta \in [0, \varepsilon]$,*

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

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

$$\partial_\varepsilon h_\rho(\bar{x}) \subseteq \partial_\varepsilon g_\rho(\bar{x}) \quad \text{for all } \varepsilon \leq \rho\eta^2\,. \tag{11}$$

Now apply Lemma 4.1 to h_ρ and to g_ρ. We have for all $\varepsilon \in [0, \rho\eta^2]$ and all $\delta \in [0, \varepsilon]$

$$\partial_\delta q_\rho(\bar{x}) + \partial_{\varepsilon-\delta} h(\bar{x}) \subseteq \bigcup_{\sigma\in[0,\varepsilon]} [\partial_\sigma q_\rho(\bar{x}) + \partial_{\varepsilon-\sigma} g(\bar{x})]\,. \tag{12}$$

Next we use Example [4, XI.1.2.2]: for all $\delta \geq 0$

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

Hence (12) reads

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

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 $\rho = \varepsilon/\eta^2$; 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 4.2 *The local optimality condition* (6) *holds if and only if*

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

Proof. Let Ψ_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 $f + \Psi_\eta = (g + \Psi_\eta) - h$ and Theorem 1.1 applies. Using Lemma 4.1 for the sum $g + \Psi_\eta$, we see that (6) holds if and only if

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

Subdifferentiating the indicator function of a ball is straightforward from the definition (1) and we obtain $\partial_\sigma \Psi_\eta(\bar{x}) = B_{\sigma/\eta}$; (13) is established. □

We conclude with some observations. Note again that (10) and (13) hold for all positive ε. As it happens, small values of δ in (10) do not bring much information, as compared to (13). On the other hand, setting $\delta = \varepsilon$ gives a definitely different inclusion, namely

$$\partial_0 h(\bar{x}) + B_{2\varepsilon/\eta} \subseteq \bigcup_{\sigma\in[0,\varepsilon]} \left[\partial_{\varepsilon-\sigma} g(\bar{x}) + B_{2\sqrt{\sigma\varepsilon}/\eta}\right] \quad \text{for all } \varepsilon > 0\,. \tag{14}$$

Finally, observe that $\partial_{\varepsilon-\sigma} g(\bar{x}) \subseteq \partial_\varepsilon g(\bar{x})$ and $B_{\sigma/\eta} \subseteq B_{\varepsilon/\eta}$ for all $\sigma \in [0, \varepsilon]$, so that the local optimality condition (6) implies via (13)

$$\partial_\varepsilon h(\bar{x}) \subseteq \partial_\varepsilon g(\bar{x}) + B_{\varepsilon/\eta} \quad \text{for all } \varepsilon > 0\,. \tag{15}$$

The limiting case $\varepsilon \searrow 0$ in all of these conditions (15), (14), (13) and (10) yields the familiar condition

$$\partial_0 h(\bar{x}) \subseteq \partial_0 g(\bar{x}) \tag{16}$$

necessary for $\bar{x}$ being a local minimizer of $f = g - h$. Hence these conditions may be viewed as a quantitative sharpening of (16), involving explicitly the size η of the optimality neighborhood in (6).

References

[1] Bomze, I.M. (2009), "Global Optimization – a Quadratic Programming Perspective," in: G. di Pillo, F. Schoen, and P. Zecca (eds.), CIME School Nonlinear Optimization, Cetraro, 2-7 July 2007, Springer Lecture Notes in Mathematics, to appear.

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

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

[4] Hiriart-Urruty, J.-B., and C. Lemaréchal (1996), *Convex Analysis and Minimization Algorithms II.* Grundlehren der Mathematik **306**, Springer, Berlin.

[5] Tuy, H. (1985), "A general deterministic approach to global optimization via d.c. programming," in: *Fermat days 85: Mathematics for optimization,* Toulouse/France 1985, North-Holland Math. Stud. **129**, 273–303.