NECESSARY AND SUFFICIENT CONDITION ON GLOBAL OPTIMALITY WITHOUT CONVEXITY AND SECOND ORDER DIFFERENTIABILITY

PÁL BURAI

ABSTRACT. The main goal of this paper is to give a necessary and sufficient condition of global optimality for unconstrained optimization problems, when the objective function is not necessarily convex. We use Gâteaux differentiability of the objective function and its bidual (the latter is known from convex analysis).

1. INTRODUCTION

One of the main topics of optimization theory, and probably one of the most important one at the same time, is to give first and second order sufficient conditions for optimality (see e.g.. [IofTih79], [BonSha00] and the references therein). Usually this is carried out via some kind of first- and second order derivatives.

In general a typical second order sufficient condition uses second order differentiability, and first order sufficient condition is only available beside some kind of additional convexity assumption on the objective function. Here we would like to give a different approach which is strongly influenced by the following theorem of Hiriart-Urruty (see [Hir86]).

Theorem 1 (Hiriart-Urruty). *Let $f \colon \mathbb{R}^n \to \mathbb{R}$ be a differentiable function. Then $\bar{x} \in \mathbb{R}^n$ is a global minimizer of f on $\mathbb{R}^n$ if and only if*

(i) $\nabla f(\bar{x}) = 0$ and
(ii) $f^{**}(\bar{x}) = f(\bar{x})$.
In such case f^{**} is differentiable at $\bar{x}$ and $\nabla f^{**}(\bar{x}) = 0$.

At first we extend this result to infinite dimensional spaces, where a more careful analysis is needed in the proof. After that we characterize the points where the objective function coincides with its bidual and show through an example how we can avoid the so called two-norm discrepancy. At last we characterize the points where the bidual is strictly less than the original function. A possible application of this is to filter out local minimizers.

2. Needed concepts

We work in a real Banach space X, X^* denotes its dual namely, the Banach space of all continuous linear functionals defined on X, and $\langle x^*, x \rangle$ is the canonical pairing between X and X^*.

Let $f \colon X \to \mathbb{R}$ be an arbitrary function (not necessarily convex), then we define its *polar (conjugate, Fenchel conjugate)* in the following way:

$$f^*(x^*) := \sup_{x \in X} \{ \langle x^*, x \rangle - f(x) \}, \qquad x^* \in X^*.$$

If we repeat this process we get the so-called *bipolar (second conjugate, bidual)* of f:

$$f^{**}(x) := \sup_{x^* \in X^*} \{ \langle x^*, x \rangle - f^*(x^*) \}, \qquad x \in X.$$

Both f^* and f^{**} are convex and extended real valued (see e.g. [Roc70] ,[EkeTem76]).

As in the convex case we define the *subdifferential at $\bar{x}$* of a not necessarily convex function in the following way.

$$\partial f(\bar{x}) := \{ x^* \in X^* | \langle x^*, x - \bar{x} \rangle \leq f(x) - f(\bar{x}), \forall x \in X \}$$

This set is possibly empty, however if f is convex then the subdifferential is nonempty at $\bar{x}$ if f is continuous and finite at $\bar{x}$ (see e.g. [Roc70] or [Eke79]).

An easy consequence of the previous definitions is the inequality

$$(1) \qquad\qquad\qquad\qquad f^{**} \leq f.$$

Let $f \colon X \to \mathbb{R}$ be a given functional. If the following limit exists then we call it the *directional derivative of f at the point $\bar{x}$ in the direction h*

$$\lim_{t \downarrow 0} \frac{f(\bar{x} + th) - f(\bar{x})}{t} =: f'(\bar{x}, h).$$

As usual f is said to be *Gâteaux differentiable at $\bar{x}$* if the previous limit exists for all h and there is an element of the dual space $f'_G(\bar{x})$ such that

$$\lim_{t \downarrow 0} \frac{f(\bar{x} + th) - f(\bar{x})}{t} = \langle f'_G(\bar{x}), h \rangle, \qquad \text{for all} \quad h \in X.$$

A point $\bar{x}$ is a *stationary point of f* if f is Gâteaux differentiable at $\bar{x}$ and $\langle f_G(\bar{x}), h \rangle = 0$ for all h.

At last f is *Fréchet differentiable at $\bar{x}$* if there is an element of the dual space $f'(\bar{x})$ such that

$$\lim_{\|h\| \to 0} \frac{f(\bar{x} + h) - f(\bar{x}) - \langle f'(\bar{x}), h \rangle}{\|h\|} = 0.$$

It is well-known that Fréchet differentiability implies Gâteaux differentiability, but the reverse is not true. Also there is a connection between the subdifferential and the Gâteaux derivative. If f is convex and

Gâteaux differentiable at $\bar{x}$ then $\partial f(\bar{x})$ contains only a single element $f'_G(\bar{x})$. Clearly it is true with Fréchet differentiability too. Conversely, if f is continuous and finite at $\bar{x}$ and $\partial f(\bar{x})$ has only one element then it must be the Gâteaux derivative of f at $\bar{x}$ and of course f is Gâteaux differentiable at $\bar{x}$ in this case (see e.g. [EkeTem76]).

3. Main results

The main theorem of this paper runs as follows.

Theorem 2. *Let X be a Banach space and $f \colon X \to \mathbb{R}$ be a function. Assume that f is Gateaux differentiable and continuous at $\bar{x} \in X$. Then $\bar{x}$ is a global minimum of f if and only if $\bar{x}$ is a stationary point of f and f coincides with its bipolar at $\bar{x}$.*

Proof. The necessity part is easy. Indeed, if $\bar{x}$ is a global minimizer then $f(\bar{x}) \le f(x)$ for all x so, $0 = \langle 0, x - \bar{x} \rangle \le f(x) - f(\bar{x})$ for all x that is, $0 \in \partial f(\bar{x})$. From this we have that $0 + f(\bar{x})$ is an exact affine minorant of f at $\bar{x}$ and this means that f is equal to its bipolar at $\bar{x}$. On the other hand, Gâteaux differentiability of f at $\bar{x}$ entails that $f'_G(\bar{x}) = 0$. Consequently $\bar{x}$ is a stationary point.

Now assume that $\bar{x}$ is such a stationary point where the bipolar is equal to the original function. Using (1) and the second assumption the following estimate is valid for all positive real t

$$(2) \qquad \frac{f^{**}(\bar{x} + tv) - f^{**}(\bar{x})}{t} \le \frac{f(\bar{x} + tv) - f(\bar{x})}{t}.$$

We intend to take the limits of both sides of the previous inequality as t tends to zero monotone decreasingly.

The limit of the right hand side exists by the assumed Gateaux differentiability at $\bar{x}$.

In regard to the left hand side we should prove that f^{**} is a "nice" convex function. The convexity comes from the definition of the bipolar. On the other hand the continuity of f at $\bar{x}$ entails its local boundedness from above at the same point. Together with (1) this means that f^{**} is also locally bounded from above at $\bar{x}$. On the other hand f^{**} is also lower semicontinuous. So, it is nowhere equal to $-\infty$ otherwise it would be equal to $-\infty$ everywhere (see e.g. [EkeTem76]) which would contradict to the fact that f^{**} is finite at $\bar{x}$. From this we get that f^{**} is locally Lipschitz at every point, so directionally differentiable at every point in all directions (see e.g. [Cla83]).

Now we can take the limits of both sides of (2).

$$\pm (f^{**})'(\bar{x}, v) \le (f^{**})'(\bar{x}, \pm v) \le \langle f'_G(\bar{x}), \pm v \rangle = 0.$$

Here we utilized that $\bar{x}$ is a stationary point of f and the elementary properties of the directional derivative. All in all we get that $(f^{**})'(\bar{x}, v) = 0$ for every direction v, that is

$$(f^{**})'(\bar{x}, \pm v) = \langle 0, v \rangle.$$

This means that $0 \in \partial(f^{**})(\bar{x})$, so $0 \leq f^{**}(x) - f^{**}(\bar{x}) = f^{**}(x) - f(\bar{x})$ for all x. Using (1) again we get $f(\bar{x}) \leq f(x)$ for every x. $\qquad \square$

The tune of the proof is similar to the original proof of Hiriart-Urruty, however in the sufficiency part we need more care and more tools because the space X is not necessarily finite dimensional.

From the second part of the proof we get the next corollary immediately.

Corollary 1. *Assume that $\bar{x}$ is a stationary point of f and f coincides with its bipolar at $\bar{x}$ then f^{**} is also Gâteaux differentiable at $\bar{x}$ and $(f^{**})'_G(\bar{x}) = 0$.*

If the objective function is Fréchet differentiable we can skip the continuity assumption.

Corollary 2. *Let X be a Banach space and $f \colon X \to \mathbb{R}$ be a function. Assume that f is Frechét differentiable at $\bar{x} \in X$. Then $\bar{x}$ is a global minimum of f if and only if $\bar{x}$ is a stationary point of f and f coincides with its bipolar at $\bar{x}$.*

Proof. Frechét differentiability at a point implies continuity and Gateaux differentiability at the same point. So, the corollary is an immediate consequence of the previous theorem. $\qquad \square$

According to the theorem sufficiency can be ensured in such cases when the function is continuous at the point in question. The continuity assumption is essential because Gâteaux differentiability does not imply continuity even in the finite dimensional case, a standard counterexample can be found e.g. in [IofTih79]. The next example (see [Hir86]) shows that a weaker differentiability, say Clarke- or other type directional differentiability, is not enough instead of Gateaux differentiability even in the simplest case when $X = \mathbb{R}$.

Example 1. *The within defined function f is locally (actually globally) Lipschitz, so it possesses Clarke directional derivative at every point.*

$$f(x) = \begin{cases} |x| & \text{if} \quad -1 \leq x \leq 1; \\ |x + \frac{3}{2}| + \frac{1}{2} & \text{if} \qquad x < -1; \\ |x - \frac{3}{2}| + \frac{1}{2} & \text{if} \qquad x > 1. \end{cases} \qquad 0 \in \partial^\circ f\left(-\frac{3}{2}\right) = [-1,1]$$

$$f^{**}(x) = \begin{cases} f(x) & \text{if} \quad |x| > \frac{3}{2}; \\ |\frac{x}{3}| & \text{if} \quad |x| \leq \frac{3}{2}. \end{cases} , \qquad 0 \notin \partial f^{**}\left(-\frac{3}{2}\right) = [-1, -\frac{1}{3}],$$

$$f\left(-\frac{3}{2}\right) = f^{**}\left(-\frac{3}{2}\right).$$

Here

$$\partial^\circ f(\bar{x}) := \{ \ x^* \in X^* \mid \langle x^*, v \rangle \leq f^\circ(\bar{x}, v) \text{ for all } v \in X \ \}$$

denotes the subgradient of f at the point $\bar{x}$ in the sense of Clarke, where

$$f^\circ(\bar{x}, v) := \limsup_{x \to \bar{x}, \ t \downarrow 0} \frac{f(x + tv) - f(x)}{t}$$

is the Clarke directional derivative of f at the point $\bar{x}$ in the direction v (for more information about these concept see [Cla83]).

It is important to notice that the condition $f^{**}(\bar{x}) = f(\bar{x})$ holds necessarily when $\bar{x}$ is a global minimizer as the next theorem shows. This condition is clearly not sufficient because lower semi-continuous convex functions are equal to their bidual everywhere.

Proposition 1. *Let X be a Banach space and $f \colon X \to \mathbb{R}$ be a function. A necessary condition of a point being a global minimizer of f is that the function itself is equal to its bidual at this point.*

Proof. Assume that $\bar{x} \in X$ is a global minimizer of f, then

$$-\sup_{x \in X}\{\langle x^*, x \rangle - f(x)\} \geq -\sup_{x \in X}\{\langle x^*, x \rangle - f(\bar{x})\} \quad \text{for all } x^* \in X^*,$$

so

$$f^{**}(\bar{x}) = \sup_{x^* \in X^*} \left\{ \langle x^*, \bar{x} \rangle - \sup_{x \in X} \{\langle x^*, x \rangle - f(x)\} \right\} \geq$$

$$\sup_{x^* \in X^*} \left\{ \langle x^*, \bar{x} \rangle - \sup_{x \in X} \{\langle x^*, x \rangle - f(\bar{x})\} \right\} =$$

$$f(\bar{x}) + \sup_{x^* \in X^*} \left\{ \langle x^*, \bar{x} \rangle - \sup_{x \in X} \{\langle x^*, x \rangle\} \right\} \geq$$

$$f(\bar{x}) + \sup_{x^* \in X^*} \left\{ \langle 0, \bar{x} \rangle - \sup_{x \in X} \{\langle 0, x \rangle\} \right\} = f(\bar{x}).$$

$\square$

4. Characterization of points where a function is equal to its bipolar

In the light of Theorem 2 it is rewarding to give a sort of description of such points where the objective function coincides with its bipolar.

Assume that $f(\bar{x}) = f^{**}(\bar{x})$, that is

$$f(\bar{x}) = f^{**}(\bar{x}) = \sup_{x^* \in X^*} \left\{ \langle x^*, \bar{x} \rangle - \sup_{x \in X} \{\langle x^*, x \rangle - f(x)\} \right\},$$

i.e. for all positive ε there exists a nonempty subset X_ε^* of the dual space, such that

$$f(\bar{x}) \leq \langle x^*, \bar{x} - x \rangle + f(x) + \varepsilon,$$

for every $x \in X$ and for every $x^* \in X_\varepsilon^*$. Here we utilized that f is real valued.

For an arbitrary positive ε and for an arbitrary (not necessarily convex) function $f\colon X \to \mathbb{R}$ define the *ε-subdifferential of f at $\bar{x}$ (or approximate subdifferential of f at $\bar{x}$)* in the following way:

$$\partial_\varepsilon f(\bar{x}) := \{x^* \in X^* | \langle x^*, x - \bar{x} \rangle \le f(x) - f(\bar{x}) + \varepsilon, \text{ for every } x \in X\}.$$

The geometric meaning of the approximate subdifferential of f at $\bar{x}$ is the collection of all the almost exact affine minorants of f at $\bar{x}$, where "almost" is measured by ε. If $\varepsilon = 0$ we get the subdifferential $\partial_0 f(\bar{x}) = \partial f(\bar{x})$. This concept was introduced in [BroRoc65] by Brønsted and Rockafellar.

It is clear that the set (possibly empty) $\partial_\varepsilon f(\bar{x})$ is convex and weak*-closed. Moreover the map $\varepsilon \mapsto \partial_\varepsilon f(\bar{x})$ is monotone increasing in the following sense, if $\varepsilon \le \varepsilon'$, then $\partial_\varepsilon f(\bar{x}) \subset \partial_{\varepsilon'} f(\bar{x})$.

Theorem 3. *Assume that a function f (not necessarily convex) is locally Lipschitz at a point. Then f is equal to its bidual at the same point, if and only if for all positive ε the corresponding approximate subdifferential of f at the point in question is not empty for every positive ε.*

Proof. Taking into account the previous the necessity part of the statement comes immediately, furthermore this part is also true without the locally Lipschitz assumption on f at $\bar{x}$.

Assume that f is locally Lipschitz at $\bar{x}$ with Lipschitz modulus L, and $\partial_\varepsilon f(\bar{x})$ is nonempty for all positive ε. Then for an arbitrary element $x^* \in \partial_\varepsilon f(\bar{x})$ we get the following estimate:

$$\langle x^*, x - \bar{x} \rangle \le f(x) - f(\bar{x}) + \varepsilon \le |f(x) - f(\bar{x}) + \varepsilon| \le L\|x - \bar{x}\| + \varepsilon$$

and interchanging x with $\bar{x}$ we get

$$|\langle x^*, x - \bar{x} \rangle| \le L\|x - \bar{x}\| + \varepsilon$$

if x is sufficiently close to $\bar{x}$. From this we have that the set $\partial_\varepsilon f(\bar{x})$ is bounded in X^*.

So, we get that the previous intersection of the nested family of these weak*-closed, bounded, convex sets is also nonempty, that is $\bigcap_{\varepsilon > 0} \partial_\varepsilon f(\bar{x}) = \partial_0 f(\bar{x}) = \partial f(\bar{x})$ has at least one element. This means that f has at least one exact affine minorant at $\bar{x}$ and this entails that f coincides its bidual at $\bar{x}$. $\square$

It is worthy to note that good analytic properties do not imply the nonemptiness of the approximate subdifferential, good examples for this are $f(x) = x^3$ and $g(x) = (x^2 - 1)^2$, $\bar{x} = \frac{1}{2}$, where ε is small.

At first sight the previous theorem can seem to be useless. The point is that as ε is increasing the set $\partial_\varepsilon f(\bar{x})$ can get large at once. A typical situation is that $\partial_0 f(\bar{x})$ has only one element and $\partial_\varepsilon f(\bar{x})$ for positive ε has infinitely many elements, so in this case it is easier to check the nonemptiness of the set $\partial_\varepsilon f(\bar{x})$.

Proposition 2. *The set $\partial_\varepsilon f(\bar{x})$ is either empty or it has exactly one element or it has infinitely many elements.*

Proof. It is enough to prove that if $\partial_\varepsilon f(\bar{x})$ has two different elements, then it has infinitely many elements, but this is a direct consequence of the convexity of $\partial_\varepsilon f(\bar{x})$. $\qquad\square$

4.1. **Avoiding two-norm discrepancy.** It is a known problem in optimization theory that in some particular cases we can not derive the desired optimality condition, because the objective functional is not (continuously) Fréchet differentiable in the required order in that space where the problem is posed (for further information on two-norm discrepancy see e.g. [Iof79] and [Trö11]). A possible way out was discovered by Ioffe (see [Iof79]).

This inconvenience can be avoided by using Theorem 2. We show this through an example which comes from [Trö11].

Example 2. *Let us consider the following optimization problem.*

$$\min_{u\in L^2[0,1]} f(u) := -\int_0^1 \cos(u(x))dx$$

Here the objective functional is not twice continuously Fréchet differentiable in $L^2[0,1]$, so the second order optimality condition is not available. On the other hand everything is nice in $L^\infty[0,1]$. It is clear that Gâteaux differentiability and continuity of f stand on the whole space. A straightforward calculation shows that $\langle (D_G f)(u), v\rangle = \langle \sin(u), v\rangle$. So, we can apply Theorem 2. From $\langle \sin(u), v\rangle = 0$ ($v \in L^2[0,1]$) we get that $u \equiv 0$ is a good candidate (of course not the only one) for being a minimizer because:

$$(D_G f)(0) = 0 \qquad and \qquad u^* \equiv 0 \in \partial_\varepsilon f(0).$$

Indeed

$$0 = \int_0^1 0u(x)dx \leq -\int_0^1 \cos(u(x))dx + 1 + \varepsilon \text{ for all } u \in L^2[0,1].$$

According to Theorem 2. $u \equiv 0$ is really a global minimizer of the previous optimization problem.

4.2. **Characterization of points where a function is not equal to its bipolar.** It seems to be useful to investigate a possible characterization of such points where a function is distinct from its bipolar for the following reason. This kind of characterization with Theorem 2 allows us to filter out local minimizers.

If $f^{**}(\bar{x}) \neq f(\bar{x})$ then according to (1) $f^{**}(\bar{x}) < f(\bar{x})$, which means

$$\sup_{x^*\in X^*} \{\langle x^*, \bar{x}\rangle - f^*(x^*)\} < f(\bar{x}) \quad \Rightarrow \quad \langle x^*, \bar{x}\rangle - \sup_{x\in X}\{\langle x^*, x\rangle - f(x)\} < f(\bar{x})$$

for all $x^* \in X^*$. That is, for every $x^* \in X^*$ there exists X_{x^*} a nonempty subset of X, such that

$$f(x) < \langle x^*, x - \bar{x} \rangle + f(\bar{x}) \qquad \text{for all } x \in X_{x^*}.$$

On the other hand if there is a positive ε such that

$$\langle x^*, \bar{x} \rangle - \sup_{x \in X} \{ \langle x^*, x \rangle - f(x) \} < f(\bar{x}) - \varepsilon$$

for all $x^* \in X^*$, then

$$\sup_{x^* \in X^*} \{ \langle x^*, \bar{x} \rangle - f^*(x^*) \} < f(\bar{x}).$$

That is, for every $x^* \in X^*$ there exists X_{x^*} a nonempty subset of X, such that

$$f(x) < \langle x^*, x - \bar{x} \rangle + f(\bar{x}) - \varepsilon \qquad \text{for all } x \in X_{x^*}.$$

Let us define the following object.for all nonnegative ε and for all $x^* \in X^*$

$$\eth_{x^*,\varepsilon} f(\bar{x}) := \{ \, x \in X \mid \langle x^*, \bar{x} - x \rangle < f(\bar{x}) - f(x) - \varepsilon \, \}.$$

Taking into account the previous we have the following proposition.

Proposition 3. *If a function f is not equal to its bipolar at the point $\bar{x}$, then $\eth_{x^*,0} f(\bar{x})$ is nonempty for every $x^* \in X^*$. Conversely, if there exists a positive ε such that $\eth_{x^*,\varepsilon} f(\bar{x})$ is nonempty for every $x^* \in X^*$, then the function is not equal to its bipolar at this point.*

References

[BonSha00] J.F. Bonnans and A. Shapiro, *Perturbation analysis of optimization problems*, Springer Series in Operations Research. Springer-Verlag, New York, 2000.

[BroRoc65] A. Brøndsted and R.T. Rockafellar, *On the subdifferentiability of convex functions*, Proc. Amer. Math. Soc. 16 (1965), 605-611.

[Cla83] F.H. Clarke, *Optimization and nonsmooth analysis*, Canadian Mathematical Society Series of Monographs and Advanced Texts. A Wiley-Interscience Publication. John Wiley and Sons, Inc., New York, 1983.

[Eke79] I. Ekeland, *Nonconvex minimization problems*, Bull. Amer. Math. Soc. (N.S.) 1 (1979), no. 3, 443-474.

[EkeTem76] I. Ekeland and R. Temam, *Convex analysis and variational problems*, Translated from the French. Studies in Mathematics and its Applications, Vol. 1. North-Holland Publishing Co., Amsterdam-Oxford; American Elsevier Publishing Co., Inc., New York, 1976.

[GeoChiPar11] P. G. Georgiev, A. Chinchuluun and P. M. Pardalos, *Optimality conditions of first order for global minima of locally Lipschitz functions*, Optimization 60 (2011), no. 1–2, 277–282.

[Hir86] J.B. Hiriart-Urruty, *When is a point x satisfying $\nabla f(x) = 0$ a global minimum of f?*, Amer. Math. Monthly 93 (1986), no. 7, 556-558.

[Iof79] A.D. Ioffe, *Necessary and sufficient conditions for a local minimum. III. Second order conditions and augmented duality*, SIAM J. Control Optim. 17 (1979), no. 2, 266-288.

[IofTih79] A.D. Ioffe and V.M. Tihomirov, *Theory of extremal problems*, Nort-Holland, Amsterdam, 1979.

[Roc70] R.T. Rockafellar, *Convex analysis*, Princeton Mathematical Series, No. 28 Princeton University Press, Princeton, N.J. 1970.

[Trö11] F. Tröltzsch, *Optimal control of partial differential equations, Theory, methods and applications*, Translated from the 2005 German original by Jürgen Sprekels. Graduate Studies in Mathematics, 112. American Mathematical Society, Providence, RI, 2010.

UNIVERSITY OF DEBRECEN, FACULTY OF INFORMATICS, H-4010 DEBRECEN, PF. 12

E-mail address: `burai.pal@inf.unideb.hu`

TU BERLIN, DEPARTMENT OF MATHEMATICS, STRAE DES 17. JUNI 136, 10623 BERLIN

E-mail address: `burai@math.tu-berlin.de`