International Journal of Research in Advent Technology, Vol.6, No.5, May2018
E-ISSN: 2321-9637
Available online at www.ijrat.org

Nonlinear Optimization with Generalized derivatives and Fenchel Duality

Vinod Kumar Bhardwaj
Department of Mathematics, G.L.A. University, Mathura-281406(UP), India
Email: dr.vinodbhardwaj@gmail.com

Abstract- Various generalized sub gradients are developed in non-smooth analysis to describe the regularity of convex functions. Optimality conditions of the problems are discussed using continuity of sub differentials. The global solutions to the problems having local solutions are provided using non-smooth necessary conditions. Application of the non-smooth theory in this paper is to combine distinct versions of necessary optimal conditions in constraints optimization.

Index Terms- Nonlinear optimization, Sub differentials, Fenchel Duality, etc.

1. INTRODUCTION

The sub differential $\partial f(\cdot)$ of a convex function f shows important and necessary properties of the derivative in optimization. When a point is (local) minimizer it follow the necessary optimality condition $0 \in \partial f(x)$ and reduces to $\{\nabla f(x)\}$ if f is differentiable at x. This frequently satisfies certain calculus rules $\partial(f+g)(x) = \partial f(x) + \partial g(x)$. For various reasons, the sub differential $\partial f(\cdot)$ is not mainly useful if the function f is not convex. This forced to search other definitions of the sub differential for a concave function. In present study we try to outline few of such sub differentials like Dini directional derivative.

Like Karush Kuhn-Tucker conditions, Fritz John conditions and Mangasarian-Fromovitz conditions are also equivalent. A Michel-Penot sub differential is generally smaller and provides stronger necessary conditions, therefore, preferred over Clarke sub differential. In contradiction to our assumption, local Lipschitzness are not assumed around the optimal point $\bar{x}$. The aim of the non smooth theory in the present study is to combine two different versions of the necessary optimality conditions are considered in constrained optimization. The first concluding in the Karush-Kuhn-Tucker conditions depending on Gateaux differentiability, while the second used the Lagrangian necessary conditions in convexity. A main character of the Michel-Penot sub differential is to coincides with the Gateaux derivative whenever exists. The generalizations of the convex sub differential indicates that a convex function is locally Lipschitz and also regular around any point in the interior of its domain. The Clarke sub differential has a extraordinary alternate explanation frequently more convenient for calculation. Rademacher's theorem states that locally Lipschitz functions are almost everywhere differentiable. It is a logically simple measure-theoretic outcome.

2. PRELIMINARIES AND DEFINITIONS

For a convex function $f : E \to (-\infty, +\infty]$ with x in dom f, we can characterize the sub differential via the directional derivative: $\phi \in \partial f(x)$ if and only if $\langle \phi, \cdot \rangle \le f'(x;\cdot)$. A natural approach is therefore to generalize the directional derivative. Henceforth in this paper we make the simplifying assumption that the real function f (a real-valued function defined on some subset of **E**) is locally Lipschitz around the point x in **E** and partly motivated by the development of optimality conditions, a simple first type is the Dini directional derivative:

$$f^{-}(x;h) = \liminf_{t \downarrow 0} \frac{f(x+th) - f(x)}{t}$$

A disadvantage of this idea is that $f^{-}(x;\cdot)$ is not usually sub linear (consider for example $f = -|\cdot|$ on **R**), so we could not expect an analogue of the Max formula. With this in mind, we introduce the Clarke directional derivative,

$$f^{\circ}(x;h) = \liminf_{y \to x,\ t \downarrow 0} \frac{f(y+th) - f(y)}{t}$$

$$= \inf_{\delta>0} \sup_{\|y-x\| \le \delta, 0<t<\delta} \frac{f(y+th) - f(y)}{t},$$

and the Michel-Penot directional derivative,

$$f^{\bullet}(x;h) = \sup_{u \in E,\ t \downarrow 0} \limsup \frac{f(y+th) - f(y)}{t}$$

Proposition 2.1 If the real function f has Lipschitz constant K around the point x in **E** then the Clarke and Michel-Penot directional derivatives $f^{\circ}(x;\cdot)$ and $f^{\bullet}(x;\cdot)$ are sub linear, and satisfy

$$f^{-}(x;\cdot) \le f^{\bullet}(x;\cdot) \le f^{\circ}(x;\cdot) \le K\|\cdot\|$$

Proof. The positive homogeneity and upper bound are straightforward, so let us prove subadditivity in the Clarke case. For any sequences $x^r \to x$ in **E** and $t \downarrow 0$ in **R**, and any real $\varepsilon > 0$, we have

$$\frac{f(x^r + t_r(u+v)) - f(x^r + t_r u))}{t_r} \le f^{\circ}(x;v) + \varepsilon$$

, and

$$\frac{f(x^r + t_r u) - f(x^r)}{t_r} \le f^{\circ}(x;u) + \varepsilon,$$

for all large r. Adding and letting r approach 1 shows

$f^{\circ}(x;u+v) \le f^{\circ}(x;u) + f^{\circ}(x;v) + 2\varepsilon$, and the result follows. We leave the Michel-Penot case as an exercise. The inequalities are straightforward.

Using our knowledge of support functions, we can now define the Clarke subdifferential

$$\partial_{\circ} f(x) = \left\{ \phi \in E \,\middle|\, \langle \phi, h \rangle \le f^{\circ}(x;h),\ for\, all\, h \in H \right\}$$

,

and the Dini and Michel-Penot sub differentials $\partial_{-} f(x)$ and $\partial_{\circ} f(x)$ analogously.

Elements of the respective sub differentials are called sub gradients.

Corollary 2.2 (Nonsmooth max formulae) If the real function f has Lipschitz constant K around the point x in **E** then the Clarke and Michel-Penot sub differentials $\partial_{\circ} f(x)$ and $\partial_{\bullet} f(x)$ are nonempty, compact and convex, and satisfy

$$\partial_{-} f(x) \subset \partial_{\bullet} f(x) \subset \partial_{\circ} f(x) \subset KB$$

Furthermore, the Clarke and Michel-Penot directional derivatives are the support functions of the corresponding sub differentials:

(2.3)

$f^{\circ}(x;h) = \max\left\{ \langle \phi, h \rangle \,\middle|\, \phi \in \partial_{\circ} f(x) \right\}$, and

(2.4)

$f^{\bullet}(x;h) = \max\left\{ \langle \phi, h \rangle \,\middle|\, \phi \in \partial_{\bullet} f(x) \right\}$

for any direction h in **E**.

Notice the Dini subdifferential is also compact and convex, but may be empty.

Clearly if the point x is a local minimizer of f then any direction h in **E** satisfies $f^{-}(x;h) \ge 0$, and hence the necessary optimality conditions $0 \in \partial_{-} f(x) \subset \partial_{\bullet} f(x) \subset \partial_{\circ} f(x)$ hold.

If g is another real function which is locally Lipschitz around x then we would not typically expect $\partial_{\circ}(f+g)(x) = \partial_{\circ} f(x) + \partial_{\circ} g(x)$ (Consider $f = -g = |\cdot|$ on **R** at x = 0 for example).

On the other hand, if we are interested in an optimality condition like $0 \in \partial_{\circ}(f+g)(x)$, it is the inclusion $\partial_{\circ}(f+g)(x) \subset \partial_{\circ} f(x) + \partial_{\circ} g(x)$ which really matters. We address this in the next result, along with an analogue of the formula for the convex sub differential of a max-function. We write $f \vee g$ for the function $x \mapsto \max\{f(x) + g(x)\}$.

Theorem 2.5 (Nonsmooth calculus) If the real functions f and g are locally Lipschitz around the point x in **E**, then the Clarke sub differential satisfies

(2.6) $\partial_{\circ}(f+g)(x) \subset \partial_{\circ} f(x) + \partial_{\circ} g(x)$, and

(2.7) $\partial_{\circ}(f \vee g)(x) \subset conv(\partial_{\circ} f(x) + \partial_{\circ} g(x))$.

Analogous results hold for the Michel-Penot sub differential.

Proof. The Clarke directional derivative satisfies

$(f+g)^{\circ}(x;\cdot) \le f^{\circ}(x;\cdot) + g^{\circ}(x;\cdot)$, Since limsup is a sub linear function. Using the Max formula (1.1.3) we deduce

$$\delta^*_{\partial_0(f+g)(x)} \le \delta^*_{\partial_\circ f(x)+\partial_\circ g(x)},$$

and taking conjugates now gives the result using the Fenchel biconjugacy theorem and the fact that both sides of inclusion (2.6) are compact and convex.

To see inclusion (2.7), fix a direction h in **E** and choose sequences $x^r \to x$ in **E** and $t \downarrow 0$ in **R** satisfying

$$\frac{(f\vee g)(x^r+t_r h)-(f\vee g)(x^r)}{t_r} \to (f\vee g)^\circ(x;h)$$

Without loss of generality, suppose $(f\vee g)(x^r+t_r h) = f(x^r+t_r h)$ for all r in some subsequence R of **N**, and now note

$$f^\circ(x;h) \ge \limsup_{r\to\infty,\, r\in R} \frac{f(x^r+t_r h)-f(x^r)}{t_r}$$
$$\ge \limsup_{r\to\infty,\, r\in R} \frac{(f\vee g)(x^r+t_r h)-(f\vee g)(x^r)}{t_r}$$
$$= (f\vee g)^\circ(x;h)$$

We deduce $(f\vee g)^\circ(x;\cdot) \le f^\circ(x;\cdot)\vee g^\circ(x;\cdot)$, which, using the Max formula (1.1.3), we can rewrite as

$$\delta^*_{\partial_0(f\vee g)(x)} \le \delta^*_{\partial_\circ f(x)} \vee \delta^*_{\partial_\circ g(x)}$$
$$= \delta^*_{conv(\partial_\circ f(x)\vee\partial_\circ g(x))}$$

using Support functions. Now the Fenchel biconjugacy theorem again completes the proof.

Theorem 2.8 (Nonsmooth necessary condition) Suppose the point $\bar{x}$ is a local minimizer for the problem

(2.9) $\inf\left\{ f(x) \middle| g_i(x) \le 0\, (i\in I)\right\}$

where the real functions f and g_i (for i in finite index set I) are locally Lipschitz around $\bar{x}$. Let $I(\bar{x}) = \left\{ i \middle| g_i(\bar{x}) = 0\right\}$ be the active set. Then there exist real $\lambda_0 \lambda_i \ge 0$, for i in $I(\bar{x})$, not all zero, satisfying

(2.10) $0 \in \lambda_0 \partial_\circ f(\bar{x}) + \sum_{i\in I(\bar{x})} \lambda_i \partial_\circ g_i(\bar{x})$

If furthermore some direction d in **E** satisfies

(2.11) $g_i^\circ(\bar{x};d) < 0$ for all i in $I(\bar{x})$

then we can assume $\lambda_0 = 1$.

Proof. Imitating the approach of section-2.3, we note that $\bar{x}$ is a local minimizer of the function

$$x \mapsto \max\left\{ f(x)\text{-}f(\bar{x}),\ g_i(x)(i\in I(\bar{x}))\right\}.$$

We deduce

$$0 \in \partial_\circ(\max\{ f\text{-}f(\bar{x}),\ g_i\,(i\in I(\bar{x}))\})(\bar{x})$$
$$\subset conv\left(\partial_\circ f(\bar{x}) \cup \bigcup_{i\in I(\bar{x})} \partial_\circ g_i(\bar{x})\right)$$

by inclusion (2.7).

If condition (2.11) holds and λ_0 is 0 in condition (1.1.10), we obtain the contradiction

$$0 \le \max\left\{ \langle \phi, d\rangle \middle| \phi \in \sum_{i\in I(\bar{x})} \lambda_i \partial_\circ g_i(\bar{x})\right\}.$$
$$= \sum_{i\in I(\bar{x})} \lambda_i g_i^\circ(\bar{x};d) < 0$$

Thus λ_0 is strictly positive, and hence without loss of generality equals to 1. Assumption (2.11) is a Mangasarian-Fromovitz type condition and the conclusion is a Karush-Kuhn-Tucker condition. Michel-Penot sub differentials are preferred over Clarke sub differential because generally it is smaller and hence provides stronger necessary conditions. In contradiction to our assumption, local Lipschitzness are not assumed around the optimal point $\bar{x}$.

3. GENERALIZED DERIVATIVES AND NON SMOOTH OPTIMIZATION

Proposition 3.1 (Unique Michel-Penot sub gradient) A real function f which is locally Lipschitz around the point x in **E** has a unique Michel- Penot sub gradient ϕ at x if and only if ϕ is the Gateaux derivative $\nabla f(x)$.

Proof. If f has a unique Michel-Penot sub gradient ϕ at x, then all directions h in **E** satisfy

$$f^{\circ}(x;h)=\sup_{u\in E}\limsup_{t\downarrow 0}\frac{f(x+th+tu)-f(x+tu)}{t}$$
$$=\langle\phi,h\rangle$$
.

The cases $h=w$ with $u=0$, and $h=-w$ with $u=w$ show

$$\limsup_{t\downarrow 0}\frac{f(x+tw)-f(x)}{t}$$
$$\leq\langle\phi,w\rangle\leq\liminf_{t\downarrow 0}\frac{f(x+tw)-f(x)}{t}$$

so we deduce $f'(x,w)=\langle\phi,w\rangle$ as required.

Conversely, if f has Gateaux derivative ϕ at x then any directions h and u in **E** satisfy

$$\limsup_{t\downarrow 0}\frac{f(x+th+tu)-f(x+tu)}{t}$$
$$\leq\limsup_{t\downarrow 0}\frac{f(x+t(h+u))-f(x)}{t}$$
$$-\liminf_{t\downarrow 0}\frac{f(x+tu)-f(x)}{t}$$
$$=f'(x;h+u)-f'(x;u)=\langle\phi,h+u\rangle-\langle\phi,u\rangle$$
$$=\langle\phi,h\rangle=f'(x;h)\leq f^{\circ}(x;h).$$

Now taking the supremum over u shows $f^{\circ}(x;h)=\langle\phi,h\rangle$ for all h, as we claimed.

Thus for example the Fritz John condition (2.10) reduces to Fritz John Theorem of weaker conditions in the differentiable case (under the extra assumption of local Lipschitzness). The above result shows that when f is Gateaux differentiable at the point x, the Dini and Michel-Penot directional derivatives coincide. If they also equal to the Clarke directional derivative then we say f is regular at x. Thus a real function f , locally Lipschitz around x, is regular at x exactly when the ordinary directional derivative $f'(x;\cdot)$ exists and equals to the Clarke directional derivative $f^{\circ}(x;\cdot)$.

One of the reasons we are interested in regularity is that when the two functions f and g are regular at x, the nonsmooth calculus rules (2.6) and (2.7) hold with equality (assuming f (x) = g(x) in the latter).

The generalizations of the convex sub differential indicates that a convex function is locally Lipschitz and also regular around any point in the interior of its domain.

Theorem 3.2 (Regularity of convex functions) Suppose the function $f:E\to(-\infty,+\infty]$ is convex. If the point x lies in int(dom f) then f is regular at x, and hence the convex, Dini, Michel-Penot and Clarke subdifferentials all coincide:

$$\partial_{\circ}f(x)=\partial_{\bullet}f(x)=\partial_{-}f(x)=\partial f(x).$$

Proof. Fix a direction h in **E,** and choose a real $\delta>0$. Denoting the local Lipschitz constant by K, we know

$$f^{\circ}(x;h)=\lim_{\varepsilon\downarrow 0}\sup_{\|y-x\|\leq\varepsilon\delta}\sup_{0<t<\varepsilon}\frac{f(y+th)-f(y)}{t}$$
$$=\lim_{\varepsilon\downarrow 0}\sup_{\|y-x\|\leq\varepsilon\delta}\frac{f(y+\varepsilon h)-f(y)}{\varepsilon}$$
$$\leq\lim_{\varepsilon\downarrow 0}\frac{f(x+\varepsilon h)-f(x)}{\varepsilon}+2K\delta$$
$$=f'(x;h)+2K\delta$$

using the convexity of f . We deduce

$$f^{\circ}(x;h)\leq f'(x;h)=f^{-}(x;h)\leq f^{\bullet}(x;h)\leq f^{\circ}(x;h)$$

and the result follows.

Thus for example, the Karush-Kuhn-Tucker type condition that we obtained at the end reduces exactly to the Lagrangian necessary conditions, written in the form $0\in\partial f(\bar{x})+\sum_{i\in I(\bar{x})}\lambda_i\partial g_i(\bar{x})$, assuming the convex functions f and g_i (for indices i in $I(\bar{x})$) are continuous at the optimal solution $\bar{x}$.

By analogy with Proposition 3.1 (Unique Michel-Penot subgradient), we might ask when the Clarke sub differential of a function f at a point x is a singleton $\{\phi\}$? Clearly in this case f must be regular, with Gateaux derivative $\nabla f(x)=\phi$, although Gateaux differentiability is not enough, as the example $x^2\sin(1/x)$. To answer the question we need a stronger notion of differentiability.

For future reference we introduce three gradually stronger conditions for an arbitrary real function f. We say an element ϕ of **E** is the Frechet derivative of f at x if it satisfies

$$\lim_{y \to x,\, y \neq x} \frac{f(y) - f(x) - \langle \phi, y - x \rangle}{\|y - x\|} = 0,$$

and we say ϕ is the strict derivative of f at x if it satisfies

$$\lim_{y,z \to x,\, y \neq z} \frac{f(y) - f(z) - \langle \phi, y - z \rangle}{\|y - z\|} = 0$$

In either case, it is easy to see $\nabla f(x)$ is ϕ. For locally Lipschitz functions on **E**, a straightforward exercise shows Gateaux and Frechet differentiability coincide, but notice that the function $x^2 \sin(1/x)$ is not strictly differentiable at 0. Finally, if f is Gateaux differentiable close to x with gradient map $\nabla f(\cdot)$ continuous, then we say f is continuously differentiable around x. In the case $E = R^n$ we see in elementary calculus that this is equivalent to the partial derivatives of f being continuous around x. We make analogous definitions of Gateaux, Frechet, strict and continuous differentiability for a function $F : E \to Y$ (where **Y** is another Euclidean space). The derivative $\nabla f(x)$ is in this case a linear map from **E** to **Y**.

Theorem 3.3 (Strict differentiability) A real function f has strict derivative ϕ at a point x in **E** if and only if it is locally Lipschitz around x with

$$\lim_{y \to x,\, t \downarrow 0} \frac{f(y + th) - f(y)}{t} = \langle \phi, h \rangle$$

for all directions h in **E**. In particular this holds if f is continuously differentiable around x, with $\nabla f(x) = \phi$.

Theorem 3.4 (Unique Clarke subgradient) A real function f which is locally Lipschitz around the point x in **E** has a unique Clarke subgradient ϕ at x if and only if ϕ is the strict derivative of f at x. In this case f is regular at x.

Proof. One direction is clear, so let us assume $\partial_\circ f(x) = \{\phi\}$. Then we deduce

$$\begin{aligned} &\lim_{y \to x,\, t \downarrow 0} \frac{f(y + th) - f(y)}{t} \\ &= -\limsup_{y \to x,\, t \downarrow 0} \frac{f((y + th) - th) - f(y + th)}{t} \\ &= -f^\circ(x; -h) = \langle \phi, h \rangle = f^\circ(x; h) \\ &= \limsup_{y \to x,\, t \downarrow 0} \frac{f(y + th) - f(y)}{t}, \end{aligned}$$

and the result now follows, using Theorem 3.3 (Strict differentiability).

Theorem 3.5 (Intrinsic Clarke sub differential) suppose that the real function f is locally Lipschitz around the point x in **E** and that the set $S \subset E$ has measure zero. Then the Clarke sub differential of f at x is

$$\partial_\circ f(x) = conv\left\{ \lim_r \nabla f(x^r) \,\middle|\, x^r \to x,\ x^r \notin S \right\}.$$

4. Conclusion:

- The Michel-Penot sub differential is analogous to Clarke sub differential.
- Locally Lipschitz functions are almost everywhere differentiable.
- The Clarke sub differential has a extraordinary alternate explanation frequently more convenient for calculation.
- The generalizations of the convex sub differential indicates that a convex function is locally Lipschitz and also regular around any point in the interior of its domain.

REFERENCES

[1] Bessis, D.N.and Clarke, F.H., Partial sub differentials, derivates and Rademacher's Theorem. Trasactions of the American Mathematical Society, 1999.

[2] Borwein, J.M., and Zhu, Q. A survey of smooth sub differential calculus with applications. Nonlinear Analysis: Theory, Methods and Applications, 1998.

[3] Clarke, F.H. Generalized gradients and applications. Transactions of the American Mathematical Society (1975), 205:247-262.
[4] Clarke, F.H. Optimization and Nonsmooth Analysis. Wiley, New York, 1983.
[5] Dini, U. Fondamenti per la teoria delle funzioni di variabili reali. Pisa, 1878.
[6] Ioffe, A.D. Approximate sub differentials and applications. I: The finite dimensional theory. Transactions of the American Mathematical Society (1984), 281:389-416.
[7] Jahn, J. An Introduction to the Theory of Nonlinear Optimization. Springer-Verlag, Berlin, 1996.
[8] Lewis, A.S. Derivatives of spectral functions. Mathematics of Operations Research (1996), 6:576-588.
[9] Lewis, A.S. Lidskii's theorem via nonsmooth analysis. SIAM Journal on Matrix Analysis, 1999.
[10] Michel, P. and Penot, J.-P. Calcul sous-differentiel pour les functions lipschitziennes et non lipschitziennes. C. R. Acad. Sci. Paris (1984), 298:269-272.
[11] Michel, P. and Penot, J.-P. A generalized derivative for calm and stable functions. Differential and Integral Equations (1992), 5:433-454.
[12] Mordukhovich, B.S. Nonsmooth analysis with nonconvex generalized differentials and adjoint mappings. Doklady Akademia Nauk BSSR (1984), 28:976-979.
[13] Rockafellar, R.T. and Wets, R.J.-B. Variational Analysis. Springer, Berlin, 1998.