Mathematical Programming 38 (1987) 287-302
North-Holland

SECOND ORDER NECESSARY AND SUFFICIENT CONDITIONS FOR CONVEX COMPOSITE NDO

James V. BURKE
Department of Mathematics GN-50, University of Washington, Seattle, WA 98195, USA

Received 20 June 1986
Revised manuscript received 13 November 1986

In convex composite NDO one studies the problem of minimizing functions of the form $F := h \circ f$ where $h: \mathbb{R}^m \to \mathbb{R}$ is a finite valued convex function and $f: \mathbb{R}^n \to \mathbb{R}^m$ is continuously differentiable. This problem model has a wide range of application in mathematical programming since many important problem classes can be cast within its framework, e.g. convex inclusions, minimax problems, and penalty methods for constrained optimization. In the present work we extend the second order theory developed by A.D. Ioffe in [11, 12, 13] for the case in which h is sublinear, to arbitrary finite valued convex functions h. Moreover, a discussion of the second order regularity conditions is given that illuminates their essentially geometric nature.

Key words: Composite functions, second order conditions, LMO approximations.

1. Introduction

The problem model that one studies in convex composite nondifferentiable optimization can be succinctly stated as

$$\text{P:}\quad \text{minimize}\quad F(x)$$
$$\text{subject to}\quad x \in \mathbb{R}^n$$

where $F := h \circ f$ with $h: \mathbb{R}^m \to \mathbb{R}$ convex and $f: \mathbb{R}^n \to \mathbb{R}^m$ continuously differentiable. This problem has recently received a great deal of attention in the literature [1, 2, 5, 9-17, 20-23], and justifiably so since many classes of problems in optimization theory can be cast within this framework e.g. convex inclusions, minimax problems, and penalty methods for nonlinear programming. Moreover, convex composite NDO provides a unified framework in which to study the convergence behavior of many of the algorithmic approaches to constrained and unconstrained optimization [2, 5, 9, 10, 14-17, 20-23]. There have been many contributors to the study of P and its implications for numerical optimization, the most notable of whom are Osborne [1, 15], Fletcher [9, 10] and Powell [16, 17]. There is however one author who has made several truly profound and inspiring contributions to the subject whose work seems to have been overlooked by many of the researchers in the area. We are speaking of the work of A.D. Ioffe in [11, 12, 13], wherein much of the theoretical foundation for convex composite NDO can be found, especially for the case in which the function h is further assumed to be sublinear. In fact, it is Ioffe's

2. A review of the first order theory for convex composite NDO

Most of the first order theory for convex composite functions is easily derived from the observation that

$$F(y)=h(f(y))=h(f(x)+f'(x)(y-x))+o(\|y-x\|). \tag{2.1}$$

Indeed, this local representation for F is only a consequence of h being locally Lipschitz on $\mathbb{R}^n$ so that

$$\begin{aligned}&|h(f(y))-h(f(x)+f'(x)(y-x))|\\&\quad\leq K_x\|y-x\|\int_0^1\|f'(x+t(y-x))-f'(x)\|\,\mathrm{d}t\end{aligned}$$

for some $K_x\geq 0$ when y is sufficiently close to x. Equation (2.1) can be written equivalently as

$$F(x+d)=h(f(x))+\Delta F(x;d)+o(\|d\|) \tag{2.2}$$

where $\Delta F(x;d)=h(f(x)+f'(x)d)-h(f(x))$ (see [14]). From (2.2) one immediately observes that F is everywhere directionally differentiable, in fact

$$\begin{aligned}F'(x;d)&=h'(f(x);f'(x)d)=\psi^*(f'(x)d\,|\,\partial h(\cdot)(f(x)))\\&=\psi^*(d\,|\,f'(x)^{\mathrm{T}}\partial h(\cdot)(f(x))).\end{aligned} \tag{2.3}$$

The last expression (2.3) indicates the following natural extension of the subdifferential calculus of convex functions to convex composite functions; the subdifferential of the convex composite function $F=h\circ f$ is the set

$$\partial F(x):=f'(x)^{\mathrm{T}}\,\partial h(f(x)) \tag{2.4}$$

for each $x\in\mathbb{R}^n$. It is well-known that this subdifferential coincides with the Clark subdifferential [8] and so possesses all of the properties of the Clark subdifferential. Equation (2.3) moreover shows that convex composite functions are subdifferentially regular [8], since $F'(x;d)=\psi^*(d\,|\,\partial F(x))$.

Representation (2.2) has been strongly exploited by several authors in the development of algorithms for P [2, 5, 9, 10, 14, 16, 17, 20–23]. One should observe that (2.2) closely resembles the usual first order expansion of F but with $\Delta F(x;d)$ replacing $F'(x;d)$. The relationship between $\Delta F(x;d)$ and $F'(x;d)$ is quite subtle and is of great significance in the algorithms for solving P. Moreover, as we shall see, this relationship is also of fundamental importance in our treatment of second order optimality conditions for P. Note in particular that since h is convex, the quotient $\lambda^{-1}\Delta F(x;\lambda d)$ is an increasing function of λ for $\lambda\geq 0$. Consequently,

$$F'(x;d)=h'(f(x);f'(x)d)=\inf_{\lambda>0}\lambda^{-1}\Delta F(x;\lambda d). \tag{2.5}$$

Employing this relationship one can easily show that $F'(x;d)$ and $\Delta F(x;d)$ are in a sense interchangeable with respect to the first order necessary conditions for P.

Theorem 2.6. *Let the functions F, h, and f be as given in* P. *If $z \in \mathbb{R}^n$ is a local solution to* P *then $0 \in \partial F(z)$. Moreover, the following conditions are equivalent*:

(1) $0 \in \partial F(z)$;

(2) $0 \leq F'(z; d)$ *for all* $d \in \mathbb{R}^n$;

(3) $0 \leq \Delta F(z; d)$ *for all* $d \in \mathbb{R}^n$.

Proof. The equivalence of (2) and (3) is obvious from (2.5) and the equivalence of (1) and (2) is a direct consequence of the geometric form of the Hahn–Banach theorem since $F'(z; d) = \psi^*(d \mid \partial F(z))$ where $\partial F(z) = f'(z)^{\mathrm{T}} \partial h(f(z))$ is a non-empty compact convex set. Finally, since F is everywhere directionally differentiable, statement (2) is a well-known consequence of the local optimality of z for P. □

In Sections 4 and 5 we will continue the discussion of the relationship between $F'(x; d)$ and $\Delta F(x; d)$.

3. A Lagrangian for P and a local dualization result

In this section we develop a local duality theory for P that parallels that which is given by Ioffe [13]. However our approach is based upon the somewhat different Lagrangian

$$L(x, y^*) := \langle y^*, f(x) \rangle - h^*(y^*)$$

where h^* is the Fenchel conjugate of h. The only difference between this Lagrangian and the one studied by both Ioffe [13] and Fletcher [9] is the use of the Fenchel conjugate h^*. It is this simple innovation though that allows us to establish the fundamental local dualization theorem for P. One can heuristically justify this definition of L by noting that the minimization of F is equivalent to a mini-max problem involving $L(x, y^*)$ as follows:

$$\inf_{x \in \mathbb{R}^n} F(x) = \inf_{x \in \mathbb{R}^n} h(f(x)) = \inf_{x \in \mathbb{R}^n} \{ \sup_{y^* \in \mathbb{R}^m} \{ \langle y^*, f(x) \rangle - h^*(y^*) \} \}$$

$$= \inf_{x \in \mathbb{R}^n} \sup_{y^* \in \mathbb{R}^m} L(x, y^*).$$

By analogy with constrained optimization theory we define the set of optimal Lagrange multipliers for P at a point $z \in \mathbb{R}^n$ to be

$$M(z) = \{ y^* \in \partial h(\cdot)(f(z)) \mid \nabla_x L(z, y^*) = 0 \},$$

and observe that $M(z) \neq \emptyset$ if and only if $0 \in \partial F(z)$, that is z satisfies the first order optimality conditions for P. As in Ioffe [13] we also define the following local approximations to $M(z)$; let η and ε be positive scalars and set

$$M_{\eta\varepsilon}(z) := \{ y^* \in \partial_\varepsilon h(\cdot)(f(z)) \colon \| \nabla_x L(z, y^*) \| \leq \eta \},$$

where ∂_ε is the usual ε-subdifferential operator of convex analysis. Observe that $M_{\eta\varepsilon}(z) = M(z)$ if $\eta = \varepsilon = 0$, and moreover

$$\bigcap_{\substack{\eta>0\\ \varepsilon>0}} M_{\eta\varepsilon}(z) = M(z)$$

with $M_{\eta\varepsilon}(z)$ being a closed compact convex set for every $\eta > 0$ and $\varepsilon > 0$. To each of the sets $M_{\eta\varepsilon}(z)$ we associate the function

$$\theta_{\eta\varepsilon}(x) := \max\{L(x, y^*) \colon y^* \in M_{\eta\varepsilon}(z)\}$$

where the maximum is always attained due to the compactness of the sets $M_{\eta\varepsilon}(z)$ whenever $M_{\eta\varepsilon}(z) \neq \emptyset$. The functions $\theta_{\eta\varepsilon}$ should be considered as local approximations to F at the point z when $M(z) \neq \emptyset$. In particular, if $M(z) \neq \emptyset$ then $\theta_{\eta\varepsilon}(z) = F(z)$ for all $\eta > 0$ and $\varepsilon > 0$. In order to see this simply note that

$$\theta_{\eta\varepsilon}(z) \leq \sup_{y^* \in \mathbb{R}^m} L(z, y^*) = h^{**}(f(z)) = h(f(z))$$

whereas if $y^* \in M(z)$ then $y^* \in M_{\eta\varepsilon}(z)$ and $y^* \in \partial h(\cdot)(f(z))$, so that

$$h(f(z)) = L(z, y^*) \leq \max\{L(z, y^*) \colon y^* \in M_{\eta\varepsilon}(z)\} = \theta_{\eta\varepsilon}(z).$$

We now establish the local dualization theorem for P. This result is a finite dimensional extension of Proposition 1 in section 2 of Ioffe [13].

Theorem 3.1. *Let F, h, and f be as given in* P. *Then the following conditions are equivalent*:

(a) *F attains a local minimum at z*;

(b) *$M(z) \neq \emptyset$ and $\theta_{\eta\varepsilon}$ attains a local minimum at z for any $\eta > 0$ and $\varepsilon > 0$*;

(c) *$M(z) \neq \emptyset$ and $\theta_{\eta\varepsilon}$ attains a local minimum at z for some $\eta > 0$ and $\varepsilon > 0$.*

Proof. By Proposition 1 in [12] we have that the mapping

$$\rho_\varepsilon(x, d) := h_\varepsilon(f(x) + f'(z)d)$$

where

$$h_\varepsilon(y) := \max\{\langle y^*, y\rangle - h^*(y^*) \colon y^* \in \partial_\varepsilon h(\cdot)(f(z))\}$$

is a so called LMO-approximation for f at z. Thus, by Proposition 5 in [12], the result will be established if we can show that

$$\theta_{\eta\varepsilon}(x) = -\min\{\rho_\varepsilon^*(x, d^*) \mid \|d^*\| \leq \eta\}$$

where

$$\rho_\varepsilon^*(x, d^*) = \sup\{\langle d^*, d\rangle - \rho_\varepsilon(x, d) \colon d \in \mathbb{R}^n\}.$$

To this end recall that if $g : \mathbb{R}^n \to \mathbb{R}$ is convex, $A \in \mathbb{R}^{m\times n}$, $a \in \mathbb{R}^m$, and we define

$$\tilde{g}(y) := g(Ay + a)$$

for all $y \in \mathbb{R}^n$, then

$$\tilde{g}^*(y^*) = \inf\{g^*(x^*) - \langle a, x^*\rangle : A^{\mathrm{T}}x^* = y^*\},$$

(e.g., see Rockafellar [19, Theorem 12.3 and Theorem 16.3]).

Consequently,

$$\rho_\varepsilon^*(x, d^*) = \inf\{h^*(y^*) + \psi(y^* \mid \partial_\varepsilon h(\cdot)(f(z))) - \langle y^*, f(x)\rangle \mid d^* = f'(z)^{\mathrm{T}}y^*\}$$

since

$$h_\varepsilon^*(y^*) = h^*(y^*) + \psi(y^* \mid \partial_\varepsilon h(\cdot)(f(z))).$$

Therefore, by straightforward computation, we have that

$$\begin{aligned}
&-\min\{\rho_\varepsilon^*(x, d^*) \mid \|d^*\| \leq \eta\} \\
&\quad = \max_{d^* \in \mathbb{R}^n}\{-\inf\{h^*(y^*) + \psi(y^* \mid \partial_\varepsilon h(\cdot)(f(z))) - \langle y^*, f(x)\rangle \mid d^* = f'(z)^{\mathrm{T}}y^*\} \\
&\qquad - \psi(d^* \mid \eta B)\} \\
&\quad = \max_{y^* \in \mathbb{R}^m}\{\langle y^*, f(x)\rangle - h^*(y^*) - \psi(y^* \mid \partial_\varepsilon h(\cdot)(f(z))) - \psi(f'(z)^{\mathrm{T}}y^* \mid \eta B)\} \\
&\quad = \max\{L(x, y^*) \mid y^* \in \partial_\varepsilon h(\cdot)(f(z)), \|\nabla_x L(z, y^*)\| \leq \eta\} = \theta_{\eta\varepsilon}(x). \quad \square
\end{aligned}$$

4. Second order necessary and sufficient conditions for P

Before stating the theorem we again return to a brief discussion of the terms $\Delta F(x; d)$ and $F'(x; d)$. Due to the inequality

$$F'(x; d) \leq \Delta F(x; d)$$

for all x and d in $\mathbb{R}^n$, we know that if $\Delta F(x; d) \leq 0$ then d is a direction of non-ascent for F. Employing this observation we define two sets of directions of non-ascent for F as follows;

$$K(x) := \{d \in \mathbb{R}^n : \Delta F(x; td) \leq 0 \text{ for some } t > 0\}$$

and

$$D(x) = \{d \in \mathbb{R}^n : F'(x; d) \leq 0\}.$$

Clearly $K(x) \subset D(x)$ and they are both convex cones, but, whereas $D(x)$ is always closed, it is possible that $K(x)$ is not. For example if

$$h((y_1, y_2)^{\mathrm{T}}) = y_1^2 + y_2^2$$

and

$$f((x_1, x_2)^T) = (x_1, x_2 + 1)^{\mathrm{T}}$$

then

$$K(0, 0) = \{(d_1, d_2)^{\mathrm{T}} : d_2 < 0\} \cup \{(0, 0)^{\mathrm{T}}\}.$$

However, note that in this example $\overline{K(0,0)} = D(0,0)$. In [13], Ioffe refers to the cone $K(x)$ as the critical cone for F at x. These two cones of directions play a central role in our development. An understanding of the exact nature of their relationship to one another is essential for an appreciation of what may be called the second order regularity of P. Specifically, we will say that F is regular at $z \in \mathbb{R}^n$ if

$$\overline{K(z)} = D(z).$$

In the next section our efforts will be devoted to obtaining an understanding of this condition. We now state our main result.

Theorem 4.1. *Let $z \in \mathbb{R}^n$ be such that f is twice continuously differentiable in a neighborhood of z.*

(1) *If F attains a local minimum at z, then*

$$\max\{d^T \nabla^2_{xx} L(z, y^*) d : y^* \in M(z)\} \geq 0 \tag{4.2}$$

whenever $d \in \overline{K(z)}$.

(2) *If $z \in \mathbb{R}^n$ is such that $M(z) \neq \emptyset$ and*

$$\max\{d^{\mathrm{T}} \nabla^2_{xx} L(z, y^*) d : y^* \in M(z)\} > 0$$

for all $d \in D(z)$, then z is an isolated local minimum for F.

Proof. (1) Note that it is sufficient to establish the result for $d \in K(z)$ since the mapping

$$d \mapsto \max\{d^{\mathrm{T}} \nabla^2_{xx} L(z, y^*) d : y^* \in M(z)\} \tag{4.3}$$

is continuous due to the compactness of $M(z)$. Also observe that by Theorem 3.1 and its preceeding discussion we know that $M(z) \neq \emptyset$ and $\theta_{\eta\varepsilon}(x) \geq \theta_{\eta\varepsilon}(z) = F(z)$ in a neighborhood of z for any $\eta > 0$ and $\varepsilon > 0$. Let $d \in K(z)$ be given. If $d = 0$, the result holds trivially, thus suppose that $d \neq 0$. Since $M_{\eta\varepsilon}(z) \subset \partial_\varepsilon h(\cdot)(f(z))$ is compact, we have

$$\begin{aligned} F(z) \leq \theta_{\eta\varepsilon}(z + td) &= \max\{\langle y^*, f(z+td)\rangle - h^*(y^*) \mid y^* \in M_{\eta\varepsilon}(z)\} \\ &\leq \max\left\{\langle y^*, f(z) + tf'(z)d\rangle + \frac{t^2}{2} d^{\mathrm{T}} \nabla^2_{xx} L(z, y^*) d \right. \\ &\qquad \left. - h^*(y^*) \mid y^* \in M_{\eta\varepsilon}(z)\right\} + o(t^2) \end{aligned} \tag{4.4}$$

Now since $d \in K(z)$ there is a $t_0 > 0$ such that

$$\Delta F(z; t_0 d) \leq 0$$

and so for each $t \in [0, t_0]$ we have from (2.5) that

$$\Delta F(z; td) \leq 0.$$

Hence, for all $t \in [0, t_0]$,

$$\langle y^*, f(z)+f'(z)\,td\rangle - h^*(y^*) \leq h(f(z)+tf'(z)d) \leq h(f(z)). \tag{4.5}$$

Combining (4.5) with (4.4) we obtain

$$0 \leq \max\{d^T\nabla^2_{xx}L(z, y^*)d \mid y^* \in M_{\eta\varepsilon}(z)\}$$

for every $\varepsilon > 0$ and $\eta > 0$. The result is now verified by again observing that

$$M(z) = \bigcap_{\substack{\eta>0\\ \varepsilon>0}} M_{\eta\varepsilon}(z) \neq \emptyset$$

where all of these sets are compact.

(2) If $D(z) = \{0\}$ we are done since then

$$0 < F'(z; d) \quad \text{for all } d \neq 0,$$

so that z is an isolated local minimum of F. Thus we shall assume that $D(z) \neq \{0\}$. Define

$$S_1 := D(z) \cap \text{bdry}(B)$$

and set

$$\gamma := \inf\{\max\{d^T\nabla^2_{xx}L(z, y^*)d \colon y^* \in M(z)\} \mid d \in S_1\}.$$

Clearly $\gamma > 0$, since S_1 is compact and the mapping (4.3) is continuous. Moreover, again by the continuity of the mapping (4.3), there is an $\varepsilon > 0$ such that

$$\max\{d^T\nabla^2_{xx}L(z, y^*)d \mid y^* \in M(z)\} \geq \gamma/2$$

for all $d \in S_1 + \varepsilon B$ and $0 \notin S_1 + \varepsilon B$. Set

$$\tilde{D}(z) := \text{cone}\{\text{bdry } B \cap (S_1 + \varepsilon B)\}.$$

Clearly $\tilde{D}(z)$ is a closed convex cone (since $0 \notin S_1 + \varepsilon B$) with

$$D(z) \subset \text{int}(\tilde{D}(z)).$$

We now claim that there is a $\nu > 0$ such that

$$\nu < F'(z; d) \tag{4.6}$$

for all $d \in S_2 := \overline{(\text{bdry}(B) \cap (\mathbb{R}^n \backslash \tilde{D}(z)))}$. If $S_2 = \emptyset$ our claim holds vacuously. On the other hand if $S_1 \neq \emptyset$ and the claim were not true, there would be a $d \in S_2$ for which $F'(x; d) = 0$ and so d is also in $\text{bdry}(B) \cap D(z)$ which is contained in the interior of $\tilde{D}(z)$. But by construction $\text{int}(\tilde{D}(z)) \cap S_2 = \emptyset$, a contradiction. Thus a ν satisfying (4.6) exists. Consequently, there is a neighborhood U_1 of z such that

$$h(f(y)) > h(f(z))$$

for all $y \in U_1 \cap (z + \text{cone}(S_2))$ with $y \neq z$.

Since $\mathbb{R}^n = \tilde{D}(z) \cup \operatorname{cone}(S_2)$, the proof would be complete if we could establish the existence of another neighborhood of z, U_2, such that

$$h(f(y)) > h(f(z))$$

for all $y \in \bar{U}_2 \cap (z + \tilde{D}(z))$ with $y \neq z$. Let us suppose to the contrary that such a neighborhood does not exist. Then there are sequences $\{t_i\} \subset \mathbb{R}_+$ and $\{d_i\} \subset \operatorname{bdry} B \cap \tilde{D}(z)$ with $t_i \downarrow 0$ such that

$$h(f(z + t_i d_i)) \leqslant h(f(z)). \tag{4.7}$$

However, for $d \in \operatorname{bdry} B \cap \tilde{D}(z)$,

$$\begin{aligned} h(f(z+td)) &= \sup\{\langle y^*, f(z+td)\rangle - h^*(y^*) \mid y^* \in \mathbb{R}^n\} \\ &\geqslant \max\{\langle y^*, f(z+td)\rangle - h^*(y^*) \mid y^* \in M(z)\} \\ &= \max\left\{\langle y^*, f(z) + tf'(z)d\rangle + \frac{t^2}{2} d^{\mathrm{T}} \nabla^2_{xx} L(z, y^*) d \right. \\ &\qquad \left. - h^*(y^*) \,\middle|\, d \in M(z)\right\} + \mathrm{o}(t^2) \\ &= h(f(z)) + \frac{t^2}{2} \max\{d^{\mathrm{T}} \nabla^2_{xx} L(z, y^*)\, d \mid y^* \in M(z)\} + \mathrm{o}(t^2) \\ &\geqslant h(f(z)) + \frac{t^2}{4} \gamma + \mathrm{o}(t^2), \end{aligned}$$

where the second equality follows from the fact that

$$h(f(z)) = \langle y^*, f(z)\rangle - h^*(y^*) \quad \text{and} \quad 0 = f'(z)^{\mathrm{T}} y^*$$

for every $y^* \in M(z)$. Hence, by (4.7),

$$0 \geqslant \gamma + 4t_i^{-2} \mathrm{o}(t_i^2)$$

for all $i = 1, 2, 3, \ldots$. Letting $i \to \infty$ we obtain the contradiction $0 \geqslant \gamma > 0$. Thus the neighborhood U_2 exists and the proof is complete. □

For the sake of comparison, we conclude this section by briefly discussing Fletcher's results on second order optimality conditions for convex composite NDO [9, Theorems 14.2.2 and 14.2.3]. Fletcher's result is as follows:

Theorem 4.8. *Let $z \in \mathbb{R}^n$ be such that f is twice continuously differentiable in a neighborhood of z.*

(1) *If F attains a local minimum at z and there is a $y^* \in M(z)$ such that*

$$D(z) = T_X(z) \tag{4.9}$$

where

$$X := \{x \colon F(x) = F(z) + \langle y^*, f(x) - f(z)\rangle\}$$

and $T_X(z)$ is the contingent cone X at z,

$$T_X(z) := \bigcap_{\substack{\varepsilon>0\\ \lambda_0>0}} \bigcup_{0<\lambda\leq\lambda_0} [\lambda^{-1}(X-z)+\varepsilon B],$$

then

$$d^{\mathrm{T}}\nabla^2_{xx}L(z, y^*)d \geq 0$$

for all $d \in D(z)$.

(2) *If there exists $y^* \in M(z)$ such that*

$$d^{\mathrm{T}}\nabla^2_{xx}L(z, y^*)d > 0$$

for all $d \in D(z)$, then z is an isolated local minimum for F.

Observe that part 2 of Theorem 4.8 is implied by part 2 of Theorem 4.1. The relationship between part 1 of these two theorems is not so obvious. They key to understanding how Fletcher's theorem allows one to employ only a single optimal dual variable in the statement of the second order necessary conditions lies in the regularity condition (4.9). However, an understanding of this condition is obscured by the complex nature of the set X and its contingent cone at z. Fletcher [9, Lemma 14.2.6] has shown that it is always the case that $T_X(z) \subset D(z)$ whenever $y^* \in M(z)$. In order to shed further light on this issue, observe that

$$\begin{aligned} X &:= \{x\colon h(f(x)) = h(f(z)) + \langle y^*, f(x) - f(z)\rangle\} \\ &= \{x\colon h(f(x)) = h(f(z)) - [\langle y^*, f(z)\rangle - h^*(y^*)] + [\langle y^*, f(x)\rangle - h^*(y^*)]\} \\ &= \{x\colon h(f(x)) = \langle y^*, f(x)\rangle - h^*(y^*)\} \\ &= \{x\colon y^* \in \partial h(\cdot)(f(x))\}, \end{aligned}$$

where the third equality follows from the fact that $y^* \in M(z) \subset \partial h(\cdot)(f(z))$.

Note that if h is affine then it is always the case that

$$D(z) = T_X(z) = \mathbb{R}^n.$$

Using this fact, one can proceed to establish certain facts about equality (4.9) when h is the pointwise maximum of a finite collection of affine functions. In fact, this exactly the course taken by Fletcher [9]. However, for more general functions h the situation becomes much less clear. For example, if h is C^1, then

$$X := \{x\colon h'(f(x)) = h'(f(z))\}.$$

Even in this case, conditions ensuring equality (4.9), when h is not affine, are nontrivial. Employing Fletcher's analysis, it is easy to show that one can relax the definition of the set X a little and replace it with the set

$$\tilde{X} := \{x\colon F(x) = F(z) + \langle y^*, f(x) - f(z)\rangle + \mathrm{o}(\|x-z\|^2)\}$$

in Theorem 4.8. Even so, equality (4.9), with either X or $\tilde{X}$, seems to impose a severe restriction on the local structure of h at x_0. One would hope that a less severe hypothesis can be found. In this connection, it is possible that the techniques employed in Rockafellar [18] would be helpful.

5. The regularity condition $\overline{K(z)} = D(z)$

In order to "close the gap" between the necessary and sufficient second order conditions in Theorem 4.1 one needs to assume that $\overline{K(z)} = D(z)$. In this section we intend to show that this condition is in fact quite mild and can be seen to arise naturally from the structure of the sets $K(x)$ and $D(x)$. To begin with observe that

$$\begin{aligned} K(x) &= \{d \colon h(f(x) + tf'(x)d) \leq h(f(x)) \text{ for some } t > 0\} \\ &= \{d \colon f(x) + tf'(x)d \in \operatorname{lev}_h(f(x))\} \\ &= \{d \colon d \in (f'(x))^{-1}[t^{-1}(\operatorname{lev}_h(f(x)) - f(x))] \text{ for some } t > 0\} \\ &= (f'(x))^{-1} \bigcup_{t>0} t^{-1}(\operatorname{lev}_h(f(x)) - f(x)) \end{aligned}$$

where $(f'(x))^{-1}$ denotes the multivalued inverse of $f'(x)$, i.e.

$$(f'(x))^{-1}(y) = \{d \in \mathbb{R}^n \colon y = f'(x)d\}.$$

Thus, by Rockafellar [19, Corollary 16.3.2],

$$K(x)^0 = f'(x)^{\mathrm{T}} N(f(x) \mid \operatorname{lev}_h(f(x))).$$

On the other hand,

$$\begin{aligned} D(x) &= \{d \colon F'(x; d) \leq 0\} = \{d \colon \psi^*(d \mid f'(x)^{\mathrm{T}}\, \partial h(\cdot)(f(x))) \leq 0\} \\ &= \{d \colon \langle d, z^* \rangle \leq 0 \text{ for all } z^* \in f'(x)^{\mathrm{T}}\, \partial h(\cdot)(f(x))\} \\ &= \{d \colon \langle d, z^* \rangle \leq 0 \text{ for all } z^* \in f'(x)^{\mathrm{T}} \operatorname{cone}(\partial h(\cdot)(f(x)))\} \\ &= [f'(x)^{\mathrm{T}} \operatorname{cone}(\partial h(\cdot)(f(x)))]^0 \end{aligned}$$

so that

$$D(x)^0 = f'(x)^{\mathrm{T}}\, \overline{\operatorname{cone}}(\partial h(\cdot)(f(x))).$$

Therefore an equivalent way of stating the regularity condition $\overline{K(x)} = D(x)$ is to specify that

$$f'(x)^{\mathrm{T}}\, \overline{\operatorname{cone}}(\partial h(\cdot)(f(x))) = f'(x)^{\mathrm{T}} N(f(x) \mid \operatorname{lev}_h(f(x))).$$

This condition is quite natural from a geometric point of view. In particular, note that by Rockafellar [19, Theorem 23.7], one in fact has that

$$\operatorname{cone}(\partial h(\cdot)(f(x))) = N(f(x) \mid \operatorname{lev}_h(f(x)))$$

at every point $x \in \mathbb{R}^n$ for which $0 \notin \partial h(\cdot)(f(x))$. We now catalogue these observations in the following proposition.

Proposition 5.1. *Let F, h, and f be as given in* P, *and let $x \in \mathbb{R}^n$, then the following conditions are equivalent*:

(1) $\overline{K(x)} = D(x)$;

(2) $f'(x)^{\mathrm{T}} N(f(x) \mid \operatorname{lev}_h(f(x))) = f'(x)^{\mathrm{T}}\, \overline{\operatorname{cone}}\{\partial h(\cdot)(f(x))\}$.

Moreover, if $0 \notin \partial h(\cdot)(f(x))$ *then it is necessarily the case that* $\bar{K}(x) = D(x)$.

In [13] Ioffe introduces the following regularity condition in order to obtain his second order results:

5.2. There exists an $\alpha > 0$ such that

$$\operatorname{dist}(d \mid K(x)) \leq \alpha \Delta F(x; d) \quad \text{for all } d \in \mathbb{R}^n.$$

We now proceed to show that if $M(x) \neq \emptyset$ then the condition 5.2 is equivalent to the statement

$$f'(x)^{\mathrm{T}} N(f(x)) \mid \operatorname{lev}_h(f(x)) = f'(x)^{\mathrm{T}} \operatorname{cone}(\partial h(\cdot)(f(x))). \tag{5.3}$$

To this end let us first observe that condition 5.2 is equivalent to saying that

$$\operatorname{dist}(d \mid K(x)) \leq \alpha F'(x; d) \tag{5.4}$$

for all $d \in \mathbb{R}^n$. Indeed, if (5.4) holds then (5.2) holds trivially since by (2.5) it is always the case that

$$F'(x; d) \leq \Delta F(x; d)$$

On the other hand, we know from [4, Theorem 3.1] that

$$\operatorname{dist}(d \mid K(x)) = \psi^*(d \mid B^0 \cap K(x)^0) \tag{5.5}$$

since $K(x)$ is a convex cone. Thus, in particular, $\operatorname{dist}(\cdot \mid K(x))$ is a positively homogeneous function and so if 5.2 holds then

$$\operatorname{dist}(d \mid K(x)) \leq \alpha t^{-1} \Delta F(x; td)$$

for all $t > 0$. Therefore, again by (2.5), we find that

$$\operatorname{dist}(d \mid K(x)) \leq \alpha F'(x; d).$$

Now combining (2.3), (5.4) and (5.5) we have that

$$\psi^*(d \mid B^0 \cap K(x)^0) \leq \psi^*(d \mid \alpha f'(x)^{\mathrm{T}} \partial h(\cdot)(f(x))) \tag{5.6}$$

for all $d \in \mathbb{R}^n$. Therefore

$$B^0 \cap K(x)^0 \subset \alpha f'(x)^{\mathrm{T}} \partial h(\cdot)(f(x)) \tag{5.7}$$

and so

$$K(x)^0 \subset f'(x)^{\mathrm{T}} \operatorname{cone}(\partial h(\cdot)(f(x)))$$

where

$$K(x)^0 = f'(x)^{\mathrm{T}} N(f(x) \mid \operatorname{lev}_h(f(x))).$$

Consequently, (5.3) holds since it is always the case that

$$\operatorname{cone}(\partial h(\cdot)(f(x))) \subset N(f(x) \mid \operatorname{lev}_h(f(x))).$$

In order to show that (5.3) implies (5.2) we employ the assumption that $M(x) \neq \emptyset$ along with the following technical lemma.

Lemma 5.8. *Let C be a closed convex subset of $\mathbb{R}^m$ with $0 \in C$. Then there is an $\alpha > 0$ such that*

$$B \cap \operatorname{cone}(C) \subset \alpha C$$

if and only if cone (C) is closed.

Proof. ($\Rightarrow$) Since C is closed we know that αC is also closed. Hence

$$B \cap \overline{\operatorname{cone}}(C) \subset \alpha C.$$

Let $\{z_i\} \subset \operatorname{cone}(C)$ be such that $z_i \to z$. Without loss of generality, we assume that there is a $y \in B$ such that $z_i/\|z_i\| \to y$. Observe that

$$y \in B \cap \overline{\operatorname{cone}}(C),$$

so that $y \in \alpha C$. Hence

$$\alpha^{-1} z_i = \alpha^{-1}\|z_i\|(z_i/\|z_i\|) \to \alpha^{-1}\|z\| y$$

where $\alpha^{-1}\|z\| y \in \operatorname{cone}(C)$. But

$$\alpha^{-1} z_i \to \alpha^{-1} z,$$

so that $z = \|z\| y \in \operatorname{cone}(C)$.

($\Leftarrow$) Since $\operatorname{cone}(C)$ is closed, the domain of the gauge function for C, $\gamma(\cdot \mid C)$, is cone (C). Moreover, $\gamma(\cdot \mid C)$ is a closed proper convex function since C is a non-empty closed convex set. Hence $\gamma(\cdot \mid C)$ is continuous on its domain. Therefore $\gamma(\cdot \mid C)$ attains a finite maximum value on the compact set $B \cap \operatorname{cone}(C)$, call this value α. Then

$$B \cap \operatorname{cone}(C) \subset \{y \colon \gamma(y \mid C) \leq \alpha\} = \alpha C$$

since $0 \in C$. $\square$

Now if we assume that (5.3) holds then $f'(x)^{\mathrm{T}} \operatorname{cone}(\partial h(\cdot)(f(x)))$ is closed. Moreover, if $M(x) \neq \emptyset$ then $0 \in f'(x)^{\mathrm{T}} \partial h(\cdot)(f(x))$. Therefore Lemma 5.8 can be applied to yield the existence of an $\alpha > 0$ such that (5.7) holds. But then, via (5.6), we have that (5.4) is also valid. Thus we have established the equivalence of (5.2) and (5.3) since we have already demonstrated the equivalence of (5.2) with (5.4). We record this result in the following proposition.

Proposition 5.9. *Let F, h, and f be as in* P, *and let $x \in \mathbb{R}^n$. Then condition* (5.2) *always implies condition* (5.3), *and if $M(x) \neq \emptyset$, then condition* (5.3) *implies condition* (5.2).

We conclude by observing that at least in the finite dimensional setting, the regularity condition that we propose is slightly weaker than the one employed by Ioffe since his condition (5.2) not only implies that $\overline{K(x)} = D(x)$ but moreover that the cone $f'(x)^{\mathrm{T}} \operatorname{cone}(\partial h(\cdot)(f(x)))$ is closed. This discrepancy, although minor, can be of significance in some cases. For example if we take $h:\mathbb{R}^n\to\mathbb{R}$ to be $h(x):=\|x\|_2+\langle u,x\rangle$ where u is any unit vector, and set $f(x)=x$, then

$$\partial h(0)=B+u,$$

so that

$$f'(0)^{\mathrm{T}}\operatorname{cone}(\partial h(\cdot)(f(0))=\{x^*:\langle u,x^*\rangle>0\}\cup\{0\},$$

which is not closed, whereas

$$K(0)=D(0)=\{-\lambda u:\lambda\geq 0\}.$$

An example illustrating the possibility that $\overline{K(x)}=D(x)$ but with $K(x)$ not closed is given in Section 4. To find an example in which neither of the regularity conditions hold one need only take $h(x)=x^2$ and $f(x)=x$.

Acknowledgment

The proof provided for part 2 of Theorem 4.1 differs markedly from the original which was modeled on the technique employed by Ioffe [13, Theorem 2]. The version of the proof contained herein was suggested by Professor R.T. Rockafellar. We gratefully acknowledge this contribution. Moreover, the careful reading and detailed comments provided by two referees are appreciated.

References

[1] D.H. Anderson and M.R. Osborne, "Discrete, nonlinear approximation problems in polyhedral norms," *Numerische Mathematik* 28 (1977) 143–156.

[2] B.M. Bell, *Nonsmooth optimization by successive quadratic programming* (Dissertation, Department of Mathematics, University of Washington, USA, 1984).

[3] A. Ben-Tal and J. Zowe, "Necessary and sufficient optimality conditions for a class of nonsmooth minimization problems," *Mathematical Programming* 24 (1982) 70–91.

[4] J.V. Burke and S.-P. Han, "A Gauss–Newton approach to solving generalized inequalities," *Mathematics of Operations Research* 11 (1986) 632–643.

[5] J.V. Burke, "Descent methods for composite nondifferentiable optimization problems," *Mathematical Programming* 33 (1985) 260–279.

[6] R.W. Chaney, "Second order necessary conditions in semi-smooth optimization," Preprint, Dept. of Math., Western Washington University, Bellingham, WA 98225, USA (1986).

[7] R.W. Chaney, "Second order sufficient conditions in nonsmooth optimization," Preprint, Western Washington University, Bellingham, WA 98225, USA (1986).

[8] F.H. Clarke, *Optimization and Nonsmooth Analysis* (Canadian Mathematical Society Series of Monographs and Advanced Texts, John Wiley and Sons, 1983).

[9] R. Fletcher, *Practical Methods for Optimization, Vol. II, Constrained Optimization* (Wiley, New York, 1981).

[10] R. Fletcher, "A model algorithm for composite nondifferentiable optimization problems," *Mathematical Programming Study* 17 (1982) 67–76.

[11] A.D. Ioffe, "Necessary and sufficient conditions for a local minimum. 1: A reduction theorem and first order conditions," *SIAM Journal Control and Optimization* 17 (1979) 245–250.

[12] A.D. Ioffe, "Necessary and sufficient conditions for a local minimum, 2: Conditions of Levitin-Miljutin-Osmolovskii Type," *SIAM Journal Control and Optimization* 17 (1979) 251-265.
[13] A.D. Ioffe, "Necessary and sufficient conditions for a local minimum. 3: Second order conditions and augmented duality," *SIAM Journal Control and Optimization* 17 (1979) 266-288.
[14] K. Madsen, *Minimization of Nonlinear Approximation Functions* (Dissertation, Institute for Numerical Analysis, Technical University of Denmark, Lyngby, Denmark, 1985).
[15] M.R. Osborne, "Algorithms for nonlinear approximation," in: C.T.H. Baker and C. Phillips, eds., *The Numerical Solution of Nonlinear Problems* (Clarendon Press, Oxford, 1981) pp. 270-286.
[16] M.J.D. Powell, "General algorithms for discrete nonlinear approximation calculation," in: C.K. Chui, L.L. Schumaker, and J.D. Ward, eds., *Approximation Theory IV*, Academic Press, NY (1983) 187-218.
[17] M.J.D. Powell, "On the global convergence of trust-region algorithms for unconstrained minimization," *Mathematical Programming* 29 (1984) 297-303.
[18] R.T. Rockafellar, "Marginal values and second-order necessary conditions for optimality," *Mathematical Programming* 26 (1983) 245-286.
[19] R.T. Rockafellar, *Convex Analysis* (Princeton University Press, 1970).
[20] R.S. Womersley, "Optimality conditions for piecewise smooth functions," Mathematical Programming Study 17 (1982) 13-27.
[21] R.S. Womersley, "Local properties of algorithms for minimizing nonsmooth composite functions," *Mathematical Programming* 32 (1985) 69-89.
[22] Y. Yuan, "Global convergence of trust region algorithms for nonsmooth optimization," Report DAMTP (1983), Cambridge University, Cambridge, England.
[23] Y. Yuan, "Some properties of trust-region algorithms for non-smooth optimization," Report DAMTP (1983), Cambridge University, Cambridge, England.