Mathematical Programming 57 (1992) 103-120
North-Holland

Optimality conditions for non-finite valued convex composite functions

J.V. Burke*

Department of Mathematics, GN-50, University of Washington, Seattle, WA 98195, USA

R.A. Poliquin**

Department of Mathematics, University of Alberta, Edmonton, Alberta, Canada T6G 2G1

Received 13 February 1989
Revised manuscript received 14 May 1990

Dedicated to the memory of Robin W. Chaney

Burke (1987) has recently developed second-order necessary and sufficient conditions for convex composite optimization in the case where the convex function is finite valued. In this note we present a technique for reducing the infinite valued case to the finite valued one. We then use this technique to extend the results in Burke (1987) to the case in which the convex function may take infinite values. We conclude by comparing these results with those established by Rockafellar (1989) for the piecewise linear-quadratic case.

AMS Subject Classifications: 90C30, 90C20, 65K05, 49A52, 49B99.

Key words: Convex composite functions, second-order optimality conditions, constraint qualification.

1. Introduction

In convex composite optimization one studies the problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x \in \mathbb{R}^n \end{array} \tag{1.1}$$

where $f := h \circ F$ is the composition of a lower semi-continuous convex function $h : \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ and a function $F : \mathbb{R}^n \to \mathbb{R}^m$ which is assumed to be either locally Lipschitzian, C^1, or C^2, depending on the application. This problem model is extremely versatile, subsuming most of the problem models usually considered in the nonlinear programming literature. As an illustration of this versatility, we offer the following example.

* Research supported in part by the National Science Foundation under grants DMS-8602399 and DMS-8803206, and by the Air Force Office of Scientific Research under grant ISSA-860080.

** Research supported in part by the Natural Sciences and Engineering Research Council of Canada under grant OGP41983.

Example. The mathematical program

$$\begin{aligned} &\text{minimize} && F_0(x) \\ &\text{subject to} && F_i(x) \leqslant 0, \quad i = 1, 2, \dots, s, \\ &&& F_i(x) = 0, \quad i = s+1, \dots, m, \end{aligned}$$

where the functions F_i, for $i = 0, 1, \dots, m$, are C^1, can be reformulated as an equivalent non-finite valued convex composite optimization problem if one defines $h : \mathbb{R} \times \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ by

$$h(y_1, y_2) := \begin{cases} y_1, & \text{if } \ y_2 \in \mathbb{R}^s_- \times \{0\}_{\mathbb{R}^{m-s}}, \\ +\infty, & \text{elsewhere}, \end{cases}$$

and F is the mapping whose components are the functions F_i, for $i = 0, 1, \dots, m$.

In recent years, convex composite optimization has received considerable attention, since it offers a unified framework within which many of the traditional problem types occurring in mathematical programming can be studied (e.g., [1, 2, 3, 5, 6, 7, 8, 10, 12, 13, 14]). In [2], Burke expands upon a technique of Ioffe [7] to develop first- and second-order optimality conditions for (1.1) in the case where the convex function h is assumed to be finite valued. In this note we extend these results to the case in which h may take infinite values.

In Section 2 we develop a first order theory for (1.1) based on chain rules due to Rockafellar [15, 16]. Section 3 contains the main technical result of the paper. In this result we show that if a constraint qualification is satisfied at a local minimizer $\bar{x}$ of (1.1), then $\bar{x}$ is also a local minimizer of the finite value convex composite function

$$f_\alpha := h_\alpha \circ F$$

for all $\alpha > 0$ sufficiently small, where the convex function $h_\alpha : \mathbb{R}^m \to \mathbb{R}$ is given by

$$h_\alpha(y) := \inf\{h(z) + \alpha^{-1} \| y - z \| : z \in \mathbb{R}^m\}.$$

Example. If the functions h and F are as given in the previous example, and the norm on $\mathbb{R} \times \mathbb{R}^m$, in the definition of h_α, is chosen to be the l_1 norm, then the function f_α defined above is given by

$$f_\alpha(x) = F_0(x) + \alpha \left[\sum_{i=1}^{s} F_i(x)_+ + \sum_{i=s+1}^{m} |F_i(x)| \right],$$

where $z_+ := \max\{0, z\}$ for all $z \in \mathbb{R}$.

The strategy in Section 4 is to apply the results of [2] to f_α and then take the limit as $\alpha \downarrow 0$ to obtain second-order optimality conditions for (1.1). The function $h_\alpha(\cdot)$ is the inf-convolution of $h(\cdot)$ and $a^{-1}\| \cdot \|$ (see [17] for inf-convolution). There is considerable reference in the literature to this type of operation. In Wets [18], the set of functions $\{\lambda\|\cdot\|: \lambda > 0\}$ is an example of what is called a cast. Casts have been used by Wets to characterize epi-convergence. By far the most common operation, in the literature, is inf-convolution with $\frac{1}{2}\|\cdot\|^2$; this operation is called the Moreau–Yosida approximation. For another application of the functions f_α we refer to Poliquin [11], where the technique of extending convex functions is used to establish the "proto-derivative" of the subdifferential of the composition of a "piecewise linear-qudratic" convex function with a C^2 mapping.

In Section 5 we discuss the relationship between the necessary and sufficient second-order optimality conditions obtained in Section 4; in doing so we correct an error present in Burke [2, Section 4].

In [13], Rockafellar develops a new theory of first- and second-order variations for functions $g: \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$. He calls these variations the first- and second-order epi-derivatives of the function g when they exist. The prefix "epi" is used to emphasize the fact that these derivatives are defined as the epi-graphical limits of certain first- and second-order variations of the epigraph of g,

$$\operatorname{epi}(g) := \{(x, \alpha) \in \mathbb{R}^n \times \mathbb{R}: g(x) \leqslant \alpha\}.$$

In the context of problem (1.1), Rockafellar [14] employs these epi-derivatives to obtain first- and second-order optimality conditions for (1.1) in the case where h is assumed to be piecewise linear-quadratic. In Section 6 we compare our results with those of Rockafellar. The comparison suggests a modification of our second-order necessary conditions; this is carried out in Theorem 6.2.

The notation that we use is for the most part standard, and follows that which is given in [4, 17]. We provide a partial list for the readers' convenience. For C a non-empty subset of $\mathbb{R}^n$, we have

ri C is the interior of C relative to the smallest affine set containing C;

$\bar{C}$ is the closure of C;

$$\operatorname{cone}(C) := \bigcup_{\lambda \geqslant 0} \{\lambda C\}$$

is the cone generated by C;

$$C^0 := \{x^* \in \mathbb{R}^n: \langle x^*, x\rangle \leqslant 1 \ \forall x \in C\}$$

is the polar of C, in particular, if C is a cone, then one can show that

$$C^0 = \{x^* \in \mathbb{R}^n: \langle x^*, x\rangle \leqslant 0 \ \forall x \in C\};$$

$$\psi(x \mid C) := \begin{cases} 0, & \text{if } x \in C, \\ +\infty, & \text{else,} \end{cases}$$

is the indicator function for C; and

$$\psi^*(x^* \mid C) := \sup\{\langle x^*, x\rangle: x \in C\}$$

is the support function for C. If, moreover, C is convex with $x \in C$ then

$$N(x \mid C) := \{x^* \in \mathbb{R}^n: \langle x^*, z - x\rangle \leqslant 0 \ \forall z \in C\}$$

is the normal cone to C at x and $T(x \mid C) := N(x \mid C)^0$ is the tangent cone to C at x.

The set $\mathbb{B} \subset \mathbb{R}^s$ is the closed unit ball for the norm $\|\cdot\|$, i.e., $\mathbb{B} := \{x \in \mathbb{R}^s : \|x\| \leq 1\}$. The set $\mathbb{B}^0 \subset \mathbb{R}^s$ is the polar of $\mathbb{B}$, and is the closed unit ball of the norm that is dual to $\|\cdot\|$. Given $C \subset \mathbb{R}^s$ the distance function to C is given by

$$\operatorname{dist}(y \mid C) := \inf\{\|y - x\| : x \in C\}.$$

Of particular interest is the distance function associated with the 2-norm. It is denoted by $\operatorname{dist}_2(y \mid C)$.

Given $M \in \mathbb{R}^{s \times k}$ we have $\operatorname{Ran}(M) := \{y \in \mathbb{R}^k : \exists x \in \mathbb{R}^s \text{ with } y = Mx\}$ is the range of M, $\operatorname{Nul}(M) := \{x \in \mathbb{R}^s : Mx = 0\}$ is the nullity of M. Given $g : \mathbb{R}^s \to \mathbb{R}^k$, we denote by $g'(x)$ the Fréchet derivative of g at x. If $k = 1$, then $g'(x; d) := \lim_{t \downarrow 0}(g(x + td) - g(x))/t$ is the directional derivative of g in the direction d, $\operatorname{lev}_g(x) := \{y : g(y) \leq g(x)\}$ is the lower level set for g at x, and

$$\operatorname{argmin} g := \{\bar{x} \in \mathbb{R}^s : g(\bar{x}) = \inf\{g(x)\}\}.$$

Let the mapping $h : \mathbb{R}^m \to \mathbb{R} \cup \{\pm\infty\}$ be convex, i.e., $\operatorname{epi}(h)$ is a convex set. The function h is said to be proper if the effective domain of h,

$$\operatorname{dom}(h) := \{y \in \mathbb{R}^m : h(y) < +\infty\},$$

is nonempty, and $h(y) > -\infty$ for at least one $y \in \operatorname{dom}(h)$. The subdifferential of h at a point $y \in \operatorname{dom}(h)$ is the set

$$\partial h(y) := \{z \in \mathbb{R}^m : (z, -1) \in N((y, h(y)) \mid \operatorname{epi}(h))\}.$$

The convex conjugate of h is the function $h^* : \mathbb{R}^m \to \mathbb{R} \cup \{\pm\infty\}$ defined by the relation

$$h^*(z) := \sup\{\langle z, y\rangle - h(y) : y \in \mathbb{R}^m\}.$$

The function h^* is clearly convex, since it is the pointwise maximum of a collection of affine functions. If h is lower semicontinuous and proper, then the subdifferentials of h and h^* have the following relationship:

$$z \in \partial h(y) \quad \text{if and only if} \quad y \in \partial h^*(z).$$

For more information on convex functions and sets see Rockafellar [17].

2. First-order theory

Problem (1.1) is best viewed as a constrained optimization problem where the constraint region is the effective domain of the function f,

$$\begin{aligned}\operatorname{dom}(f) &:= \{x \in \mathbb{R}^n : f(x) < +\infty\}\\ &= \{x \in \mathbb{R}^n : F(x) \in \operatorname{dom}(h)\}.\end{aligned}$$

Thus problem (1.1) can be restated as

$$\begin{aligned}&\text{minimize} && f(x),\\ &\text{subject to} && F(x) \in \operatorname{dom}(h).\end{aligned} \tag{2.1}$$

In this context, the subdifferential regularity of f (in the sense of Clarke [4]) will in general depend upon a constraint qualification for (2.1).

Definition 2.1. Let $h:\mathbb{R}^m\to\mathbb{R}\cup\{+\infty\}$ be lower semi-continuous and convex, and let $F:\mathbb{R}^n\to\mathbb{R}^m$ be locally Lipschitzian. We say that the function $f:\mathbb{R}^n\to\mathbb{R}\cup\{+\infty\}$, defined as the composition $f:=h\circ F$, satisfies the basic constraint qualification at a point $x\in\operatorname{dom}(f)$ if the only point $z\in N(F(x)|\operatorname{dom}(h))$ for which $0\in z^{\mathrm{T}}\partial F(x)$ is $z=0$, where $\partial F(x)$ is the generalized Jacobian of F at x (see Clarke [4, p. 69]).

Remarks. (1) If $F:\mathbb{R}^n\to\mathbb{R}^m$ is C^1, then the basic constraint qualification is satisfied by f at x if

$$\operatorname{Nul}(F'(x)^{\mathrm{T}})\cap N(F(x)|\operatorname{dom}(h))=\{0\}.$$

(2) For the mathematical program discussed in the introduction, the basic constraint qualification is equivalent to the Mangasarian–Fromovitz constraint qualification [9].

(3) The basic constraint qualification for (1.1) was introduced in Rockafellar [13]. Second-order optimality conditions for (1.1) are obtained in [14] for the case in which h is assumed to be piecewise linear-quadratic. A further discussion of Rockafellar's results appears in Section 5.

The basic constraint qualification is just the tool needed to obtain first order necessary conditions for optimality in (1.1). The following result is an immediate consequence of Rockafellar [15, Corollary 5.2.3].

Theorem 2.1. *Let $f:\mathbb{R}^n\to\mathbb{R}\cup\{+\infty\}$ be as in* (1.1) *with $F:\mathbb{R}^n\to\mathbb{R}^m$ locally Lipschitzian. If $x\in\operatorname{dom}(f)$ is a local minimizer for f at which the basic constraint qualification is satisfied, then there exists $y\in\partial h(F(x))$ such that*

$$0\in y^{\mathrm{T}}\partial F(x).\qquad\square$$

In the case where F is C^1, a much stronger result can be obtained from Rockafellar [16, Theorem 3].

Theorem 2.2. *Let $f:\mathbb{R}^n\to\mathbb{R}\cup\{+\infty\}$ be as in* (1.1) *with $F:\mathbb{R}^n\to\mathbb{R}^m$ C^1, and let $x\in\operatorname{dom}(f)$ be a point at which the constraint qualification is satisfied and $\partial h(F(x))\neq\emptyset$. Then f is subdifferentially regular at x and one has*

$$\partial f(x)=F'(x)^{\mathrm{T}}\partial h(F(x)).\qquad\square$$

Corollary 2.1. *Let $f:\mathbb{R}^n\to\mathbb{R}\cup\{+\infty\}$ and $x\in\operatorname{dom}(f)$ be as in Theorem* 2.2. *If x is a local minimizer for f, then*

$$0\in F'(x)^{\mathrm{T}}\partial h(F(x)).\qquad\square$$

We close this section by recalling the Lagrangian for (1.1) introduced in Burke [2]. The Lagrangian is given by the expression

$$L(x,y):=\langle y,F(x)\rangle-h^*(y)\tag{2.2}$$

where $h^*:\mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ is the Fenchel conjugate of h. By analogy to constrained optimization, we define the set of optimal multipliers at a local solution $\bar{x}$ to (1.1) to be

$$M(\bar{x}) := \{y \in \partial h(F(\bar{x})) : 0 \in \partial_x L(\bar{x}, y)\}. \tag{2.3}$$

If the basic constraint qualification is satisfied at $\bar{x}$, then $M(\bar{x})$ is guaranteed to be non-empty by Theorem 2.1. In Lemma 4.2 we will show that the basic constraint qualification is equivalent to the compactness of $M(x)$ whenever $M(x)$ is nonempty and F is strictly differentiable at x. This result extends the analogous result in nonlinear programming for the Mangasarian–Fromowitz constraint qualification.

3. The reduction theorem

Given $\alpha > 0$ and $h:\mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ as in (1.1), define $h_\alpha:\mathbb{R}^m \to \mathbb{R}$ by

$$h_\alpha(y) := \inf\{h(z) + \alpha^{-1}\|y - z\| : z \in \mathbb{R}^m\} \tag{3.1}$$

and $f_\alpha:\mathbb{R}^n \to \mathbb{R}$ by

$$f_\alpha(x) = h_\alpha(F(x)), \tag{3.2}$$

where $\|\cdot\|$ is a given norm on $\mathbb{R}^m$.

In this section we show that if $x \in \text{dom}(f)$ is a local minimizer of f at which the basic constraint qualification is satisfied, then there exists an $\bar{\alpha} > 0$ such that for all $\alpha \in (0, \bar{\alpha})$, the point x is also a local minimizer for f_α. This result is the keystone of our development, as it allows us to extend second-order optimality results for the finite valued case to the infinite valued case. We begin with a few preparatory lemmas.

Lemma 3.1. *Let $f:\mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be as in* (1.1) *with $F:\mathbb{R}^n \to \mathbb{R}^n$ locally Lipschitzian. If $x \in \text{dom}(f)$ is such that f satisfies the basic constraint qualification at x, then there is a neighborhood U of x such that f satisfies the basic constraint qualification at every point of* $\text{dom}(f) \cap U$.

Proof. The result follows immediately from the upper semi-continuity of the normal cone operator and the generalized Jacobian $\partial F(x)$. □

Lemma 3.2. *Let $h:\mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ be as in* (1.1) *and let $\{(y_i, z_i)\} \subset \text{graph}(\partial h)$ be such that $y_i \to y \in \text{dom}(h)$ and $\|z_i\| \uparrow +\infty$. Then every cluster point of the sequence $\{z_i/\|z_i\|\}$ is an element of the normal cone to* $\text{dom}(h)$ *at y.*

Proof. The result follows immediately from the subdifferential inequality. □

Lemma 3.3. *Let $h:\mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ be as in* (1.1), *and let $h_\alpha:\mathbb{R}^m \to \mathbb{R}$ be as defined in* (3.1). *If $h_\alpha(\bar{y}) = h(\bar{z}) + \alpha^{-1}\|\bar{y} - \bar{z}\|$, where $\bar{z} \in \text{dom}(h)$, then $u \in \partial h_\alpha(\bar{y})$ if and only if $u \in \partial h(\bar{z}) \cap (\alpha^{-1}\mathbb{B}^0)$ and $(\bar{y} - \bar{z}) \in N(u|\alpha^{-1}\mathbb{B}^0)$.*

Proof. ($\Rightarrow$) Since $h^*_\alpha = h^* + \psi(\cdot \mid \alpha^{-1}\mathbb{B}^0)$, we have

$$\bar{y} \in \partial h^*_\alpha(u) = \partial h^*(u) + N(u \mid \alpha^{-1}\mathbb{B}^0)$$

and $u \in \alpha^{-1}\mathbb{B}^0$. For all $z \in \operatorname{dom}(h)$,

$$h(z) \geq h_\alpha(z) \geq h_\alpha(\bar{y}) + \langle u, z - \bar{y}\rangle = h(\bar{z}) + \alpha^{-1}\|\bar{y} - \bar{z}\| + \langle u, z - \bar{y}\rangle. \tag{3.3}$$

Hence

$$h(z) \geq h(\bar{z}) + \langle u, \bar{y} - \bar{z}\rangle + \langle u, z - \bar{y}\rangle = h(\bar{z}) + \langle u, z - \bar{z}\rangle,$$

since $\alpha^{-1}\|\bar{y} - \bar{z}\| \geq \langle u, \bar{y} - \bar{z}\rangle$. This shows that $u \in \partial h(\bar{z})$. Setting $z = \bar{z}$ in (3.3) yields

$$0 \geq \alpha^{-1}\|\bar{y} - \bar{z}\| + \langle u, \bar{z} - \bar{y}\rangle.$$

But

$$\alpha^{-1}\|\bar{y} - \bar{z}\| \geq \langle v, \bar{y} - \bar{z}\rangle$$

if $v \in \alpha^{-1}\mathbb{B}^0$. Hence $0 \geq \langle v - u, \bar{y} - \bar{z}\rangle$ for all $v \in \alpha^{-1}\mathbb{B}^0$, i.e., $(\bar{y} - \bar{z}) \in N(u|\alpha^{-1}\mathbb{B}^0)$.

($\Leftarrow$) If $u \in \partial h(\bar{z})$ and $(\bar{y} - \bar{z}) \in N(u|\alpha^{-1}\mathbb{B}^0)$, then $\bar{y} = \bar{z} + (\bar{y} - \bar{z}) \in \partial h^*_\alpha(u)$. Therefore, $u \in \partial h_\alpha(\bar{y})$. □

The reduction theorem now follows.

Theorem 3.1. *Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be as in* (1.1) *with $F : \mathbb{R}^n \to \mathbb{R}^m$ locally Lipschitzian, and let $\bar{x}$ be a local minimizer of f at which the basic constraint qualification is satisfied. If $f_\alpha : \mathbb{R}^n \to \mathbb{R}$ is as defined in* (3.2), *then there is an $\bar{\alpha} > 0$ such that $\bar{x}$ is a local minimizer for f_α with $f_\alpha(\bar{x}) = f(\bar{x})$ for all $\alpha \in (0, \bar{\alpha})$.*

Remark. If the functions h and F are as given in the examples in the Introduction, and the norm on $\mathbb{R} \times \mathbb{R}^m$ is chosen to be the l_1 norm, then as we observed, the function f_α defined above is given by

$$f_\alpha(x) = F_0(x) + \alpha\left[\sum_{i=1}^{s} F_i(x)_+ + \sum_{i=s+1}^{m} |F_i(x)|\right],$$

where $z_+ := \max\{0, z\}$ for all $z \in \mathbb{R}$. Therefore, Theorem 3.1 extends the classical results for exact penalization [4, 5, 7]. Further investigations along these lines are pursued in Burke [1].

Proof of Theorem 3.1. Let $\varepsilon > \delta > 0$ be such that $f(x) \geq f(\bar{x})$ for all $x \in \bar{x} + \varepsilon\mathbb{B}$, and f satisfies the basic constraint qualification on $\operatorname{dom}(f) \cap (\bar{x} + \delta\mathbb{B})$ (δ exists by Lemma 3.1). Set $\xi := 1 + \max\{\|F(x) - F(\bar{x})\| : x \in \bar{x} + \varepsilon\mathbb{B}\}$, define

$$\tilde{h}_\alpha(y) := \inf\{h(z) + \psi(z \mid F(\bar{x}) + \xi\mathbb{B}) + \alpha^{-1}\|y - z\| : z \in \mathbb{R}^m\}$$

i.e., $\tilde{h}_\alpha(\cdot)$ is the inf-convolution of $[h(\cdot)+\psi(\cdot\,|F(\bar{x})+\xi\mathbb{B})]$ with $\alpha^{-1}\|\cdot\|$, and consider the function

$$\hat{f}_\alpha(x) := \tilde{f}_\alpha(x)+\varphi(x)+\psi(x\,|\,\bar{x}+\varepsilon\mathbb{B})$$

where $\tilde{f}_\alpha := \tilde{h}_\alpha \circ F$ and $\varphi:\mathbb{R}^n \to \mathbb{R}$ is the C^1 function

$$\varphi(x) := \text{dist}_2^2(x\,|\,\bar{x}+\delta\mathbb{B}).$$

Observe that argmin $\hat{f}_\alpha$ is non-empty as $\hat{f}_\alpha$ is lower semi-continuous and $\bar{x}+\varepsilon\mathbb{B}$ is compact. Consequently, there is a sequence $\alpha_i \downarrow 0$ for which there is a corresponding sequence $\{x_i\} \subset \bar{x}+\varepsilon\mathbb{B}$ converging to some element $\hat{x}$ of $\bar{x}+\varepsilon\mathbb{B}$, and such that

$$x_i \in \text{argmin}\, \hat{f}_{\alpha_i}$$

for each $i = 1, 2, \ldots$. Also, from the lower semi-continuity of h and the compactness of $F(\bar{x})+\xi\mathbb{B}$, there exists for each $i = 1, 2, \ldots$, a y_i in $\text{dom}(h) \cap (F(\bar{x})+\xi\mathbb{B})$ such that

$$\tilde{f}_{\alpha_i}(x_i) = h(y_i)+\alpha_i^{-1}\|y_i - F(x_i)\|.$$

Clearly,

$$f(\bar{x}) = h(F(\bar{x})) \geq \tilde{h}_{\alpha_i}(F(\bar{x})) \geq \hat{f}_{\alpha_i}(\bar{x}) \geq \hat{f}_{\alpha_i}(x_i) \geq \tilde{f}_{\alpha_i}(x_i), \tag{3.4}$$

from the definitions of x_i, $\tilde{h}_{\alpha_i}$, and $\hat{f}_{\alpha_i}$. Therefore, as $\alpha_i \downarrow 0$ we have $\|y_i - F(x_i)\| \to 0$ so that $y_i \to F(\hat{x})$, and thus eventually $y_i \in \text{int}(F(\bar{x})+\xi\mathbb{B})$, which implies that $\tilde{f}_{\alpha_i}(x_i) = f_{\alpha_i}(x_i)$. From (3.4) we also obtain that $F(\hat{x}) \in \text{dom}(h) \cap (F(\bar{x})+\xi\mathbb{B})$ and

$$\begin{aligned} f(\bar{x}) &\geq \hat{f}_{\alpha_i}(x_i) = f_{\alpha_i}(x_i)+\varphi(x_i) \\ &= h(y_i)+\alpha_i^{-1}\|y_i - F(x_i)\|+\varphi(x_i) \geq h(y_i)+\varphi(x_i). \end{aligned}$$

Now let $i \to \infty$, and use the fact that h is lower semi-continuous to obtain

$$f(\bar{x}) \geq f(\hat{x})+\varphi(\hat{x}) \geq f(\hat{x}).$$

But, since $\hat{x} \in \bar{x}+\varepsilon\mathbb{B}$, the hypotheses imply that $f(\bar{x}) = f(\hat{x})$ and $\hat{x} \in \bar{x}+\delta\mathbb{B}$.

We now show that eventually $F(x_i) \in \text{dom}(h)$. Since $x_i \in \text{argmin}\, \hat{f}_{\alpha_i}$ and $x_i \to \hat{x} \in \bar{x}+\delta\mathbb{B}$, we know that eventually

$$0 \in \partial\tilde{f}_{\alpha_i}(x_i)+\nabla\varphi(x_i).$$

As above, $h_{\alpha_i}(F(x_i)) = f_{\alpha_i}(x_i) = \tilde{f}_{\alpha_i}(x_i) = h(y_i)+\alpha_i^{-1}\|y_i - F(x_i)\|$, so applying Lemma 3.3, $u \in \partial h_{\alpha_i}(F(x_i))$ if and only if $u \in \partial h(y_i) \cap \alpha_i^{-1}\mathbb{B}^0$ and $F(x_i) - y_i \in N(u\,|\,\alpha_i^{-1}\mathbb{B}^0)$. Now since $x_i \to \hat{x} \in \bar{x}+\varepsilon\mathbb{B}$,

$$F(x_i) \to F(\hat{x}) \in \text{int}(F(\bar{x})+\xi\mathbb{B}),$$

so eventually $\partial\tilde{h}_{\alpha_i}(F(x_i)) = \partial h_{\alpha_i}(F(x_i))$ and $\partial\tilde{f}_{\alpha_i}(x_i) = \partial f_{\alpha_i}(x_i)$. Applying Theorem 2.1, there exists $v_i \in \partial h_{\alpha_i}(F(x_i))$ with

$$0 \in v_i^{\mathrm{T}}\,\partial F(x_i)+\nabla\varphi(x_i)^{\mathrm{T}}. \tag{3.5}$$

If the sequence $\{v_i\}$ possesses a divergent subsequence $\{v_i\}_J$, then by Lemma 3.2, the sequence $\{v_i/\|v_i\|\}_J$ possesses a cluster point $v \in N(F(\hat{x})|\text{dom}(h))$ with $\|v\|=1$. But for such a cluster point v we obtain from (3.5) that $0 \in v^{\mathrm{T}}\,\partial F(\hat{x})$, which contradicts the choice of δ, therefore the sequence $\{v_i\}$ is bounded. Hence for $\bar{\alpha}$ sufficiently small $\{v_i\} \subset \bar{\alpha}^{-1}\mathbb{B}^0$ so that $N(v_i \mid \alpha_i^{-1}\mathbb{B}^0) = \{0\}$ for all i, such that $\alpha_i < \bar{\alpha}$. But then $y_i = F(x_i)$, so that $F(x_i) \in \text{dom}(h)$ whenever $\alpha_i < \bar{\alpha}$. Therefore, for all $\alpha_i < \bar{\alpha}$,

$$f(\bar{x}) \geq \hat{f}_{\alpha_i}(x_i) \geq \tilde{f}_{\alpha_i}(x_i) = f_{\alpha_i}(x_i) = h(F(x_i)) \geq f(\bar{x}),$$

so that $\bar{x} \in \operatorname{argmin} \hat{f}_{\alpha_i}$. Consequently, $\bar{x}$ is also a local minimizer of f_{α_i} for all $\alpha_i < \bar{\alpha}$. □

Remark. The method of proof also shows that if $\bar{x}$ is a strict local minimizer of f, then it is also strict local minimizer of f_α.

4. Second-order optimality conditions

We begin by recalling the results of Burke [2]. To this end we define the following two sets of non-ascent for f at a point $x \in \text{dom}(f)$: set

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

and set

$$K(x) := \{d \in \mathbb{R}^n : \exists \bar{t} > 0 \text{ such that } h(F(x) + tF'(x)d) \leq h(F(x))\ \forall t \in (0, \bar{t})\}.$$

Remark. The set $K(x)$ was introduced in Ioffe [7] where it is referred to as the critical cone.

Theorem 4.1 (Burke [2, Theorem 4.1]). *Let $f : \mathbb{R}^n \to \mathbb{R}$ be such that $f := h \circ F$ with $h : \mathbb{R}^m \to \mathbb{R}$ convex and lower semi-continuous, and $F : \mathbb{R}^n \to \mathbb{R}^m$ is such that F'' exists at $\bar{x} \in \mathbb{R}^n$.*

(1) *If f attains a local minimum at $\bar{x}$, then*

$$\max\{d^{\mathrm{T}}\nabla^2_{xx}L(\bar{x}, y)d : y \in M(\bar{x})\} \geq 0$$

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

(2) *If $\bar{x}$ is such that $M(\bar{x}) \neq \emptyset$ and*

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

for all $d \in D(\bar{x})$, then f attains an isolated local minimum at $\bar{x}$. □

We now use Theorem 3.1 to extend Theorem 4.1 to the infinite valued case. This requires an understanding of the relationship between the sets $M(x)$, $D(x)$ and $K(x)$ and the sets

$$M(x, \alpha) := \{y \in \partial h_\alpha(F(x)) : 0 \in \partial_x L(x, y)\}, \tag{4.1}$$

$$D(x, \alpha) := \{d \in \mathbb{R}^n : f'_\alpha(x;\, d) \leq 0\} \tag{4.2}$$

and

$$K(x,\alpha) := \{d \in \mathbb{R}^n : \exists \bar{t} > 0 \text{ such that } h_\alpha(F(x) + tF'(x)d) \leq h_\alpha(F(x)) \ \forall t \in (0, \bar{t})\}. \quad (4.3)$$

Remark. The set $M(x, \alpha)$ is related to the function f_α just as the set $M(x)$ is related to the function f. In order to see that this is correct the reader should verify that $(h_\alpha)^* = h^* + \psi(\cdot \mid \alpha^{-1}\mathbb{B})$.

Lemma 4.1. *Let $h : \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ be lower semi-continuous and convex and let $y \in \mathrm{dom}(\partial h)$. Then*

$$\partial h_\alpha(y) = \alpha^{-1}\mathbb{B}^0 \cap \partial h(y)$$

whenever $\alpha^{-1} \geq \mathrm{dist}(0, \partial h(y))$ where h_α is defined in (3.1).

Proof. This is an immediate consequence of Lemma 3.3. □

Lemma 4.2. *Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be as in* (1.1) *and suppose that $\bar{x} \in \mathrm{dom}(f)$ is such that $M(\bar{x}) \neq \emptyset$.*

(1) *If f satisfies the basic constraint qualification at $\bar{x}$, then $M(\bar{x})$ is compact.*

(2) *If F is strictly differentiable at $\bar{x}$ and $M(\bar{x})$ is compact, then f satisfies the basic constraint qualification at $\bar{x}$.*

Remark. Recall that $F : \mathbb{R}^n \to \mathbb{R}^m$ is said to be strictly differentiable at $\bar{x} \in \mathbb{R}^n$ if for all v in $\mathbb{R}^n$,

$$\liminf_{\substack{x \to \bar{x} \\ t \downarrow 0}} \frac{F(x+tv) - F(x)}{t} = F'(\bar{x})v,$$

where $F'(\bar{x})$ is the Fréchet derivative of F at $\bar{x}$.

Proof of Lemma 4.2. (1) The fact that $M(\bar{x})$ is closed follows from the upper semi-continuity of ∂h and ∂F. Now if $M(\bar{x})$ is unbounded, then there is a sequence $\{v_i\} \subset M(\bar{x}) \subset \partial h(F(\bar{x}))$ with $\|v_i\| \uparrow \infty$ and $v_i / \|v_i\| \to v$ with $\|v\| = 1$. By Lemma 3.2 we have $v \in N(F(\bar{x}) \mid \mathrm{dom}(h))$. Also, $0 \in v^{\mathrm{T}} \partial F(\bar{x})$ since $\partial F(\bar{x})$ is compact. This contradicts the basic constraint qualification. Hence $M(\bar{x})$ is compact.

(2) Since F is strictly differentiable at $\bar{x}$ we obtain from [16, Theorem 3] that $\partial_x L(\bar{x}, y) = F'(\bar{x})^{\mathrm{T}} y$. It then follows that the recession cone of $M(\bar{x})$ is the set

$$\{z \in \mathbb{R}^m : z \in N(F(\bar{x}) \mid \mathrm{dom}(h)) \text{ and } F'(\bar{x})^{\mathrm{T}} z = 0\}.$$

Hence the result is a consequence of [17, Theorem 8.4]. □

Proposition 4.1. *Let $f : \mathbb{R}^n \to \mathbb{R} \cap \{+\infty\}$ be as in* (1.1) *with $F : \mathbb{R}^n \to \mathbb{R}^m$ in C^1, and let $\bar{x} \in \mathrm{dom}(f)$ be a point at which f satisfies the basic constraint qualification.*

(1) *For all $\alpha > 0$ such that $\alpha^{-1} \geq \mathrm{dist}(0 \mid \partial h(F(\bar{x})))$ we have that $D(\bar{x}) \subset D(\bar{x}, \alpha)$.*

(2) *If $\bar{x}$ is a local minimizer of f, then for all $\alpha > 0$ sufficiently small,*

$$M(\bar{x}) = M(\bar{x}, \alpha) \quad \text{and} \quad K(\bar{x}) \subset K(\bar{x}, \alpha).$$

Proof. (1) From Burke [2], f_α is subdifferentially regular on $\mathbb{R}^n$ with

$$\partial f_\alpha(\bar{x}) = F'(\bar{x})^{\mathrm{T}} \partial h_\alpha(F(\bar{x})).$$

If $\alpha^{-1} \geqslant \mathrm{dist}\,(0|\partial h(F(\bar{x})))$, then $\partial h(F(\bar{x})) \neq \emptyset$, and so $\partial f(\bar{x}) = F'(\bar{x})^{\mathrm{T}} \partial h(F(\bar{x}))$ by Theorem 2.2. Consequently, Lemma 4.1 implies that

$$\begin{aligned}\partial f_\alpha(\bar{x}) &= F'(\bar{x})^{\mathrm{T}} \partial h_\alpha(F(\bar{x})) = F'(\bar{x})^{\mathrm{T}}[\alpha^{-1}\mathbb{B}^0 \cap \partial h(F(\bar{x}))]\\ &\subset F'(\bar{x})^{\mathrm{T}} \partial h(F(\bar{x})) = \partial f(\bar{x}).\end{aligned}$$

Therefore,

$$f'_\alpha(\bar{x}; d) = \sup\{\langle z, d\rangle : z \in \partial f_\alpha(\bar{x})\} \leqslant \sup\{\langle z, d\rangle : z \in \partial f(\bar{x})\} = f'(\bar{x}; d),$$

whereby the result is established.

(2) By the reduction Theorem 3.1, there is an $\bar{\alpha}_1 > 0$ such that $\bar{x}$ is a local minimizer of f_α for all $\alpha \in (0, \bar{\alpha}_1)$. Hence $M(\bar{x}, \alpha) \neq \emptyset$ for all $\alpha \in (0, \bar{\alpha}_1)$ by Theorem 2.1. From Lemma 4.2, we know that $M(\bar{x})$ is compact and so, by Lemma 4.1, there is an $\bar{\alpha}_2 > 0$ such that $M(\bar{x}) \subset \partial h_\alpha(F(\bar{x}))$ for all $\alpha \in (0, \bar{\alpha}_2)$. But then $M(\bar{x}) = M(\bar{x}, \alpha)$ for all $\alpha \in (0, \min\{\bar{\alpha}_1, \bar{\alpha}_2\})$.

Next observe that

$$h_\alpha(F(\bar{x}) + tF'(\bar{x})d) \leqslant h(F(\bar{x}) + tF'(\bar{x})d)$$

for every choice of $t \geqslant 0$ and $d \in \mathbb{R}^n$. Also, by Theorem 3.1, $h_\alpha(F(\bar{x})) = h(F(\bar{x}))$ for all $\alpha \in (0, \bar{\alpha}_1)$. Hence $K(\bar{x}) \subset K(\bar{x}, \alpha)$ for all $\alpha \in (0, \bar{\alpha}_1)$. □

The following extension of Theorem 4.1 to the infinite valued case is now easily established.

Theorem 4.2. *Let $f: \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be as in* (1.1), *and let $\bar{x} \in \mathrm{dom}(f)$ be such that F'' exists at $\bar{x}$.*

(1) *If $\bar{x}$ is a local minimizer of f at which f satisfies the basic constraint qualification, then*

$$\max\{d^{\mathrm{T}}\nabla^2_{xx}L(\bar{x}, y)d : y \in M(\bar{x})\} \geqslant 0 \tag{4.4}$$

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

(2) *If $\bar{x}$ is such that $M(\bar{x}) \neq \emptyset$ and*

$$\max\{d^{\mathrm{T}}\nabla^2_{xx}L(\bar{x}, y)d : y \in M(\bar{x})\} > 0 \tag{4.5}$$

for all $d \in D(\bar{x})$, then there is a $\gamma > 0$ such that

$$f(x) \geqslant f(\bar{x}) + \gamma\|x - \bar{x}\|^2$$

for all x near $\bar{x}$.

Proof. (1) By Lemma 4.2, $M(\bar{x})$ is non-empty and compact. Hence the map

$$d \mapsto \max\{d^{\mathrm{T}}\nabla^2_{xx}L(\bar{x}, y)d : y \in M(\bar{x})\}$$

is continuous. Consequently, we need only establish (4.4) for $d\in K(\bar{x})$. From Proposition 4.1 and Theorem 3.1, we know that there is an $\bar{\alpha}>0$ such that $M(\bar{x})=M(\bar{x},\alpha)$ and $\bar{x}$ is a local minimizer for f_α for all $\alpha\in(0,\bar{\alpha})$. Therefore, by Theorem 4.1,

$$\max\{d^{\mathrm{T}}\nabla^2_{xx}L(\bar{x},y)d: y\in M(\bar{x})\}=\max\{d^{\mathrm{T}}\nabla^2_{xx}L(\bar{x},y)d: y\in M(\bar{x},\alpha)\}\geqslant 0$$

whenever $d\in K(\bar{x},\alpha)$ for all $\alpha\in(0,\bar{\alpha})$. But by Proposition 4.1, $K(\bar{x})\subset K(\bar{x},\alpha)$ for all α sufficiently small, whereby the result follows.

(2) Suppose to the contrary that the result is not valid. Then there exist sequences $\gamma_i\downarrow 0$, and $x_i\to\bar{x}$ with $(x_i-\bar{x})/\|x_i-\bar{x}\|\to d$, such that

$$f(x_i)\leqslant f(\bar{x})+\gamma_i\|x_i-\bar{x}\|^2 \tag{4.6}$$

for all $i=1,2,\ldots$. Dividing (4.6) through by $\|x_i-\bar{x}\|$ and taking the limit as $i\to\infty$, we see that $d\in D(\bar{x})$. Moreover, since h is lower semi-continuous,

$$\begin{aligned}
f(x_i)&=\sup\{\langle y,F(x_i)\rangle-h^*(y): y\in\mathbb{R}^m\}\\
&\geqslant\sup\{\langle y,F(x_i)\rangle-h^*(y): y\in M(\bar{x})\}\\
&\geqslant\sup\{\langle y,F(\bar{x})\rangle-h^*(y)+\tfrac{1}{2}(x_i-\bar{x})^{\mathrm{T}}\nabla^2_{xx}L(\bar{x},y)(x_i-\bar{x}): y\in M(\bar{x})\}\\
&\quad+\mathrm{o}(\|x_i-\bar{x}\|^2)\\
&=f(\bar{x})+\tfrac{1}{2}\max\{(x_i-\bar{x})^{\mathrm{T}}\nabla^2_{xx}L(\bar{x},y)(x_i-\bar{x}): y\in M(\bar{x})\}+\mathrm{o}(\|x_i-\bar{x}\|^2)
\end{aligned}$$

where the last equality follows from the inclusion $M(\bar{x})\subset\partial h(F(\bar{x}))$ i.e., $\langle y,F(\bar{x})\rangle-h^*(y)=f(\bar{x})$. Combining this with (4.6), dividing through by $\|x_i-\bar{x}\|^2$, and taking the limit as $i\to\infty$ yields the contradiction

$$0\geqslant\max\{d^{\mathrm{T}}\nabla^2_{xx}L(\bar{x},y)d: y\in M(\bar{x})\}.$$

Hence part (2) is established. □

5. A comparison of the sets $D(x)$ and $\overline{K(x)}$

The second-order necessary and sufficient conditions of the previous section differ in their use of the sets $D(x)$ and $\overline{K(x)}$. Thus an understanding of the relationship between $D(x)$ and $\overline{K(x)}$ is necessary for an appreciation of these second-order results. The relationship between these two sets was examined in Burke [2]; however, the primary result of that investigation [2, Proposition 5.1] is incorrect due to improper use of the closure operation. In this section we correct the errors present in [2, Section 5].

In Burke [2, Section 5] it was observed that

$$D(x)=[F'(x)^{\mathrm{T}}\operatorname{cone}(\partial h(F(x)))]^{0}$$

and

$$K(x)=[F'(x)]^{-1}\bigcup_{t>0} t^{-1}(\mathrm{lev}_h(F(x))-F(x))$$

where $[F'(x)]^{-1}$ is the multivalued inverse of $F'(x)$, i.e.,

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

From Rockafellar [17, Corollary 16.3.2], we know that

$$\left[[F'(x)]^{-1}\bigcup_{t>0} t^{-1}(\mathrm{lev}_h(F(x))-F(x))\right]^0=F'(x)^{\mathrm{T}}N(F(x)\,|\,\mathrm{lev}_h(F(x)))$$

if

$$\mathrm{Ran}(F'(x))\cap \mathrm{ri}[T(F(x)\,|\,\mathrm{lev}_h(F(x)))]\neq\emptyset, \tag{5.1}$$

where (5.1) is satisfied at any point $x\in\mathrm{dom}(f)$ at which the basic constraint qualification is satisfied (Burke [1, Proposition 4.3]). Moreover, by Rockafellar [17, Theorem 23.7], we have

$$\mathrm{cone}(\partial h(F(x)))=N(F(x)\,|\,\mathrm{lev}_h(F(x))) \tag{5.2}$$

at every point $x\in\mathbb{R}^n$ for which $0\notin\partial h(F(x))$. By combining these facts we obtain the following result.

Proposition 5.1. *Let $f\colon\mathbb{R}^n\to\mathbb{R}\cup\{+\infty\}$ be as in* (1.1) *with $F\colon\mathbb{R}^n\to\mathbb{R}^m$ a C^1 function. If $x\in\mathrm{dom}(f)$ is such that* (5.1) *and* (5.2) *hold, then $D(x)=\overline{K(x)}$. Moreover, if f satisfies the basic constraint qualification at x, then* (5.1) *holds at x, and if $0\notin\partial h(F(x))$, then* (5.2) *holds at x.* □

If $\bar{x}$ is a local minimum for f, it may happen that $F(\bar{x})$ is also a global minimum for h, in which case the validity of (5.2) is in question. However, this is typically not the case and (5.2) is satisfied. An extremely useful class of functions for which (5.2) is always satisfied are the indicator functions for convex sets. It is of course possible that (5.1), (5.2) or both fail to hold at the solution to (1.1).

Examples. (1) Let $h\colon\mathbb{R}\to\mathbb{R}$ be defined by

$$h(y):=\max\{0,y\},$$

and let $F\colon\mathbb{R}\to\mathbb{R}$ be given by

$$F(x)=x^2.$$

Then $x=0$ solves (1.1). In this case (5.1) fails to hold although (5.2) does hold.

(2) Let $h\colon\mathbb{R}\to\mathbb{R}$ be defined by

$$h(y):=[\max\{0,y\}]^2,$$

and let $F\colon\mathbb{R}\to\mathbb{R}$ be the identity map,

$$F(x)=x.$$

Then the set $\{x : x\leqslant 0\}$ solves (1.1). In this case (5.1) holds at $x=0$ but (5.2) does not.

(3) If one takes $h\colon\mathbb{R}\to\mathbb{R}$ as given in example (2) and $F\colon\mathbb{R}\to\mathbb{R}$ as given in example (1), then neither (5.1) nor (5.2) hold at the solution $x=0$ of (1.1).

We conclude by noting that the condition that is missing from the analysis given in Burke [2, Section 5] is condition (5.1).

6. The second-order conditions of Rockafellar

We begin this section by recalling the second-order optimality conditions of Rockafellar.

Definition 6.1. A function $h:\mathbb{R}^m \to \mathbb{R}\cup\{+\infty\}$ with effective domain D is called piecewise linear-quadratic if D can be expressed as the union of finitely many sets D_j such that, for each j, the set D_j is a convex polyhedron and the restriction of h to D_j is a quadratic (or affine) function.

Theorem 6.1 (Rockafellar [14, Theorem 3.4]). *Let $f:\mathbb{R}^n \to \mathbb{R}\cup\{+\infty\}$ be as in (1.1) with $h:\mathbb{R}^m \to \mathbb{R}\cup\{+\infty\}$ piecewise linear-quadratic convex and $F:\mathbb{R}^n \to \mathbb{R}^m$ is C^2, and let $\bar{x}\in \operatorname{dom}(f)$ be such that f satisfies the basic constraint qualification at $\bar{x}$.*

(1) *If f has a local minimum at $\bar{x}$, then $0\in\partial f(\bar{x})$ and*

$$h''(F(\bar{x}); F'(\bar{x})d) + \max\{d^{\mathrm{T}}\nabla^2_{xx}L(\bar{x}, y)d: y\in M(\bar{x})\} \geq 0 \tag{6.1}$$

for all $d\in D(\bar{x})$ where

$$h''(F(\bar{x}); F'(\bar{x})d) := \lim_{t\downarrow 0} \frac{h(F(\bar{x})+tF'(\bar{x})d) - h(F(\bar{x})) - th'(F(\bar{x}); F'(\bar{x})d)}{\frac{1}{2}t^2} \tag{6.2}$$

and this limit exists and may equal $+\infty$.

(2) *If $0\in\partial f(\bar{x})$ and*

$$h''(F(\bar{x}); F'(\bar{x})d) + \max\{d^{\mathrm{T}}\nabla^2_{xx}L(\bar{x}, y)d: y\in M(\bar{x})\} > 0 \tag{6.3}$$

for all $d\in D(\bar{x})\backslash\{0\}$, then there is an $\alpha>0$ such that

$$f(y)\geq f(\bar{x}) + \alpha\|y-\bar{x}\|^2$$

for all y near $\bar{x}$. □

One can obtain some insight into the nature of the expressions (6.1) and (6.3) by considering the C^2 case. If both h and F are C^2 near $\bar{x}$, then so is f, in which case

$$\begin{aligned} d^{\mathrm{T}}\nabla^2 f(\bar{x})d &= h''(F(\bar{x}); F'(\bar{x})d) + d^{\mathrm{T}}\nabla^2 L(\bar{x}, \nabla h(F(\bar{x})))d \\ &= d^{\mathrm{T}}\left[F'(\bar{x})^{\mathrm{T}}\nabla^2 h(F(\bar{x}))F'(\bar{x}) + \sum_{i=1}^{m}\frac{\partial h}{\partial y_i}(F(\bar{x}))\nabla^2 F_i(\bar{x})\right]d. \end{aligned}$$

Thus it appears that Theorem 6.1 is able to capture the second-order behavior of h, whereas Theorems 4.1 and 4.2 do not. Of course, in the case where h and F are C^2 as above, then the appropriate second-order results can be obtained from Theorem 4.1 by redefining F as $h\circ F$ and h as the identity map. Nevertheless, there are many examples in which Theorem 6.1 provides a sharper result.

Example. Let $h:\mathbb{R}\to\mathbb{R}$ be defined as follows:

$$h(x):=\begin{cases}2x^2, & \text{if } x\leq 0,\\ x^2, & \text{if } x>0,\end{cases}$$

and let $F:\mathbb{R}\to\mathbb{R}$ be the identity map $F(x)\equiv x$. Then part (2) of Theorem 6.1 applies to yield that $x=0$ is a local minimizer of f. However, Theorem 4.2 does not apply.

One can also verify that if f satisfies the basic constraint qualification at $\bar{x}\in\operatorname{dom}(f)$ and $0\in\partial f(\bar{x})$, then

$$h''(F(\bar{x});\ F'(\bar{x})d)\geq 0$$

for all $d\in\mathbb{R}^n$, and

$$h''(F(\bar{x});\ F'(\bar{x})d)=0$$

for all $d\in K(\bar{x})$. Therefore, part (1) of Theorem 6.1 implies part (1) of Theorem 4.2, and part (2) of Theorem 4.2 implies part (2) of Theorem 6.1. Consequently, Theorem 6.1 is a stronger result for the case in which h is piecewise linear-quadratic and $D(\bar{x})\neq\overline{K(\bar{x})}$. In light of this fact we establish the following result concerning a second-order necessary condition for optimality.

Theorem 6.2. *Let $f:\mathbb{R}^n\to\mathbb{R}\cup\{+\infty\}$ be as in (1.1) with $F:\mathbb{R}^n\to\mathbb{R}^n$ a C^1 function. If $\bar{x}\in\operatorname{dom}(f)$ is a local minimizer of f at which F'' exists and f satisfies the basic constraint qualification at $\bar{x}$, then*

$$h''_*(F(\bar{x});\ F'(\bar{x})d)+\max\{d^{\mathrm{T}}\nabla^2_{xx}L(\bar{x},y)d:\ y\in M(\bar{x})\}\geq 0 \tag{6.4}$$

for all $d\in D(\bar{x})$, where

$$h''_*(F(\bar{x});\ F'(\bar{x})d):=\liminf_{\substack{u\to d\\ t\downarrow 0}}\frac{h(F(\bar{x})+tF'(\bar{x})u)-h(F(\bar{x}))-th'(F(\bar{x});\ F'(\bar{x})d)}{\frac{1}{2}t^2}.$$

Proof. We begin by establishing (6.4) in the case where h is finite valued. In order to do this we need to reintroduce some of the tools developed in Ioffe [7] and Burke [2]. Consider the function

$$\theta_{\xi\varepsilon}(x):=\max\{L(x,y):\ y\in M_{\xi\varepsilon}(\hat{x})\}$$

where $L:\mathbb{R}^n\times\mathbb{R}^m\to\mathbb{R}\cup\{-\infty\}$ is the Lagrangian defined in (2.2), $\hat{x}\in\mathbb{R}^n$ is such that $0\in\partial f(\hat{x})$, $\xi>0$, $\varepsilon>0$, and

$$M_{\xi\varepsilon}(\hat{x}):=\{y\in\partial_\varepsilon h(F(\hat{x})):\ \|\nabla_x L(\hat{x},y)\|\leq\xi\}.$$

Here $\partial_\varepsilon h(y)$ denotes the usual ε-subdifferential of convex analysis. In Burke [2, Theorem 3.1] it was shown that if $F:\mathbb{R}^n\to\mathbb{R}^m$ is C^1 and $h:\mathbb{R}^m\to\mathbb{R}$ is lower semi-continuous and convex, then $\hat{x}$ is a local minimizer for f if and only if $\hat{x}$ is a local minimizer for $\theta_{\xi\varepsilon}$ for some or all $\xi>0$ and $\varepsilon>0$. Since $\bar{x}$ is a local minimizer of f and $\theta_{\xi\varepsilon}(\bar{x})=h(F(\bar{x}))$, we have

$$\liminf_{\substack{u\to d\\ t\downarrow 0}} \frac{\theta_{\xi\varepsilon}(\bar{x}+tu)-h(F(\bar{x}))}{\frac{1}{2}t^2}\geqslant 0$$

for all $d\in\mathbb{R}^n$. Furthermore, $h'(F(\bar{x});\ F'(\bar{x})d)\leqslant 0$ for all $d\in D(\bar{x})$ and

$$\begin{aligned}\theta_{\xi\varepsilon}(\bar{x}+tu)&=\max\{L(\bar{x}+tu,y)\colon y\in M_{\xi\varepsilon}(\bar{x})\}\\ &=\max\{\langle y, F(\bar{x}+tu)\rangle-h^*(y)\colon y\in M_{\xi\varepsilon}(\bar{x})\}\\ &=\max\{\langle y, F(\bar{x})+tF'(\bar{x})u+\tfrac{1}{2}t^2F''(\bar{x})(u,u)\rangle+\mathrm{o}(t^2)-h^*(y)\colon y\in M_{\xi\varepsilon}(\bar{x})\}\\ &\leqslant h(F(\bar{x})+tF'(\bar{x})u)+\tfrac{1}{2}t^2\max\{u^{\mathrm{T}}\nabla^2_{xx}L(\bar{x},y)u\colon y\in M_{\xi\varepsilon}(\bar{x})\}+\mathrm{o}(t^2)\end{aligned}$$

for u near d and all t sufficiently small. These observations yield the inequality

$$h''_*(F(\bar{x});\ F'(\bar{x})d)+\max\{d^{\mathrm{T}}\nabla^2_{xx}L(\bar{x},y)d\colon y\in M_{\xi\varepsilon}(\bar{x})\}\geqslant 0$$

for all $d\in D(\bar{x})$. Expression (6.4) now follows since

$$M(\bar{x})=\bigcap_{\substack{\xi>0\\ \varepsilon>0}} M_{\xi\varepsilon}(\bar{x}).$$

Next let us suppose that h is not necessarily finite valued. By Theorem 3.1 and Proposition 4.1, there is an $\bar{\alpha}>0$ such that $\bar{x}$ is a local minimizer of f_α where f_α is defined in (3.2), $M(\bar{x})=M(\bar{x},\alpha)$, and $D(\bar{x})\subset D(\bar{x},\alpha)$ for all $\alpha\in(0,\bar{\alpha})$. Hence, by the first part of the proof,

$$(h_\alpha)''_*(F(\bar{x});\ F'(\bar{x})d)+\max\{d^{\mathrm{T}}\nabla^2_{xx}L(\bar{x},y)d\colon y\in M(\bar{x})\}\geqslant 0$$

for all $d\in D(\bar{x})$ and $\alpha\in(0,\bar{\alpha})$. Now since $h_\alpha(F(\bar{x}))=h(F(\bar{x}))$ for $\alpha\in(0,\bar{\alpha})$ and $h(F(\bar{x})+tF'(\bar{x})d)\geqslant h_\alpha(F(\bar{x})+tF'(\bar{x})d)$ for all $t\in\mathbb{R}$, $d\in\mathbb{R}^n$, and $\alpha>0$, we have that

$$h''_*(F(\bar{x});\ F'(\bar{x})d)\geq(h_\alpha)''_*(F(\bar{x});\ F'(\bar{x})d)$$

for all $d\in D(\bar{x})$ and $\alpha\in(0,\bar{\alpha})$, whereby (6.4) is established. $\square$

Theorem 6.2 agrees with part (1) of Rockafellar's result, Theorem 6.1, for the piecewise linear-quadratic case. However, extending part (2) of Rockafellar's Theorem 6.1 to more general functions h does not seem to be as simple task and is a topic of ongoing research.

7. Final comment

Our study of these more general types of second order conditions was originally motivated by algorithmic interests. We wished to develop a local convergence theory for nonlinear programming algorithms that did not suffer from the rather severe local hypotheses required by the available proof techniques. Our investigations indicated the need to extend the standard second order theory. We believe that the approach developed in this paper and in [1, 2, 13, 14] is the correct way to go. Regardless, from an algorithmic viewpoint, the approach does present many challenges. Even now we are unsure as to how this type of second order structure can be used to guide the construction of algorithms possessing a more robust local convergence theory. Indeed, it even remains uncertain whether or not a numerically practical theory can be developed for general problems of this type. Nonetheless, we do believe that certain intermediary results are possible and that a more general understanding of the second order structure of convex composite functions will be required to obtain them.

Correction

The first author regrets the presence of errors in Section 5 of [2]. As previously stated, the most fundamental error is the absence of condition (5.1) in [2, Proposition 5.1]. But there is yet another error in [2] that needs to be addressed. Specifically, one of the implications in [2, Lemma 5.8] is false. Consequently, the second implication in [2, Proposition 5.9] is also invalid. We thank Danny Ralph for questioning the proof of [2, Lemma 5.8]. His queries prompted us to construct a counter-example demonstrating this result to be invalid.

References

[1] J.V. Burke, "An exact penalization viewpoint of constrained optimization," *SIAM Journal on Control and Optimization* 29 (1991).

[2] J.V. Burke, "Second order necessary and sufficient conditions for convex composite NDO," *Mathematical Programming* 38 (1987) 287-302.

[3] JV. Burke, "Descent methods for composite nondifferentiable optimization problems," *Mathematical Programming* 33 (1985) 260-279.

[4] F.H. Clarke, *Optimization and Nonsmooth Analysis* (Wiley, New York, 1983).

[5] R. Fletcher, *Practical Methods of Optimization* (Wiley, New York, 1987).

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

[7] A. D. Ioffe, "Necessary and sufficient conditions for a local minimum 2: Conditions of Levitin-Miljutin-Osmolovskii type," *SIAM Journal on Control and Optimization* 17 (1979) 251-265.

[8] K. Madsen, "Minimization of non-linear approximation functions," Ph.D Thesis, Institute for Numerical Analysis, Technical University of Denmark (Lyngby, Denmark, 1985).

[9] O. L. Mangasarian and S. Fromovitz, "The Fritz John necessary optimality conditions in the presence of equality and inequality constraints," *Journal of Mathematical Analysis and Applications* 17 (1967) 37-47.

[10] M.R. Osborne and G.A. Watson, "Nonlinear approximation problems in vector norms," in: G.A. Watson, ed., *Numerical Analysis (Dundee 1977), Lecture Notes in Mathematics No. 630* (Springer, New York, 1978) pp. 115-132.

[11] R.A. Poliquin, "Proto-differentiation of subgradient set-valued mappings," *Canadian Journal of Mathematics* XLII(3) (1990) 520-532.

[12] M.J.D. Powell, "General algorithms for discrete nonlinear approximation calculations," in: C.K. Chui, L.L. Schumaker and J.D. Ward, eds., *Approximation Theory IV* (Academic Press, New York, 1983) pp. 187–218.

[13] R.T. Rockafellar, "First- and second-order epi-differentiability in nonlinear programming," *Transactions of the American Mathematical Society* 307 (1988) 75–108.

[14] R.T. Rockafellar, "Second-order optimality conditions in nonlinear programming obtained by way of epi-derivatives," *Mathematics of Operations Research* 14 (1989) 462–484.

[15] R.T. Rockafellar, "Extensions of subgradient calculus with applications to optimization," *Nonlinear Analysis* 9 (1985) 665–698.

[16] R.T. Rockafellar, "Directionally Lipschitzian functions and subdifferential calculus," *Proceedings of the London Mathematical Society* 39 (1979) 331–355.

[17] R.T. Rockafellar, *Convex Analysis* (Princeton University Press, Princeton, NJ, 1970).

[18] R.J.-B. Wets, "Convergence of convex functions, variational inequalities and convex optimization problems," in: R. Cottle, F. Giannessi and J.-L. Lions, eds., *Variational Inequalities and Complimentary Problems* (Wiley, New York, 1980) pp. 375–403.