APPROXIMATE HESSIAN MATRICES AND SECOND-ORDER OPTIMALITY CONDITIONS FOR NONLINEAR PROGRAMMING PROBLEMS WITH C^1-DATA

V. JEYAKUMAR[1] and X. WANG[1]

Abstract

In this paper, we present generalizations of the Jacobian matrix and the Hessian matrix to continuous maps and continuously differentiable functions respectively. We then establish second-order optimality conditions for mathematical programming problems with continuously differentiable functions. The results also sharpen the corresponding results for problems involving $C^{1,1}$-functions.

1. Introduction

This work was motivated both by the need of sharper second-order optimality conditions for non-smooth mathematical programming problems and by the recent work ([8, 9]) on second-order non-smooth analysis of approximate Jacobian matrices which extends basic calculus of locally Lipschitz vector-valued maps to continuous maps and unifies and strengthens various results of second-order calculus. In this paper we examine mathematical programming problems involving continuously differentiable functions. Mathematical programming problems involving continuously differentiable functions which are not necessarily $C^{1,1}$-functions, that is, continuously differentiable functions with locally Lipschitz gradients, arise in applications, particularly, in reformulating certain variational inequality problems with continuously differentiable maps as equivalent optimization problems. We present necessary and sufficient optimality conditions for such mathematical programming problems. The optimality conditions, which are expressed in terms of approximate Hessian matrices, extend and sharpen the corresponding results for mathematical programming problems with $C^{1,1}$-functions [5].

---

[1]Department of Applied Mathematics, University of New South Wales, Sydney 2052, Australia

A new notion of an *approximate Hessian* is defined as the closed and bounded set of matrices which is an approximate Jacobian of the vector-valued gradient mapping. For a scalar-valued mapping, the approximate Jacobian means there exists a compact set which generates both an upper convex and a lower concave approximation to the mapping at a point (see [15, 8]). It is a key property that is shared by most of the generalized sub-differentials in non-smooth analysis. Our approach also extends the idea of generalized Hessian matrix introduced and studied for $C^{1,1}$-functions in [5].

The outline of the paper is as follows. In Section 2 we introduce the notions of approximate Jacobian and Hessian matrices and develop calculus rules such as the sum formula, the mean-value condition and generalized Taylor's expansions which will then be used to establish optimality conditions. We also provide examples to show the significance of such matrices in particular in the locally Lipschitz case. In Section 3 we present necessary optimality conditions for nonlinear programming problems with equality and inequality constraints, where the functions are assumed to be C^1. The optimality conditions are compared with the corresponding results for problems involving $C^{1,1}$ functions. A numerical example is given to show that the results provide sharper conditions for such $C^{1,1}$-problems. In Section 5 we establish second-order sufficient optimality conditions.

This paper is dedicated to Professors Bruce Craven and Bert Mond on the occasion of their retirement.

2. Approximate Jacobian & Hessian matrices

We assume throughout the paper that F is a function mapping $\mathbb{R}^n$ into $\mathbb{R}^m$, that is, $F : \mathbb{R}^n \to \mathbb{R}^m$ and that F has components, $(f_1, \dots, f_m)$. For each $v \in \mathbb{R}^m$ the composite function, $(vF) : \mathbb{R}^n \to \mathbb{R}$, is defined by

$$(vF)(x) = \langle v, \ F(x) \rangle = \sum_{i=1}^{m} v_i f_i(x).$$

The lower Dini directional derivative and the upper Dini directional derivative of vF at x in the direction $u \in \mathbb{R}^n$ are defined by

$$(vF)^-(x, u) := \liminf_{t \downarrow 0} \frac{(vF)(x + tu) - (vF)(x)}{t},$$

$$(vF)^+(x, u) := \limsup_{t \downarrow 0} \frac{(vF)(x + tu) - (vF)(x)}{t}.$$

We denote by $L(\mathbb{R}^n, \mathbb{R}^m)$ the space of all $n \times m$ matrices. The *convex hull* and the *closed convex hull* of a set A in a topological vector space are denoted by $co(A)$ and $\overline{co}(A)$ respectively.

DEFINITION 2.1. The map $F : \mathbb{R}^n \to \mathbb{R}^m$ admits an **approximate Jacobian** $\partial^* F(x)$ at $x \in \mathbb{R}^n$ if $\partial^* F(x) \subseteq L(\mathbb{R}^n, \mathbb{R}^m)$ is closed and bounded and for each $v \in \mathbb{R}^m$,

$$(vF)^+(x, u) \leq \max_{M \in \partial^* F(x)} \langle Mv, u \rangle \quad \forall u \in \mathbb{R}^n. \tag{2.1}$$

A matrix M of $\partial^* F(x)$ is called an **approximate Jacobian matrix** of F at x. Note that condition (2.1) is equivalent to the condition

$$(vF)^-(x, u) \geq \min_{M \in \partial^* F(x)} \langle Mv, u \rangle \quad \forall u \in \mathbb{R}^n \tag{2.2}$$

and is a strengthened version of the initial definition of an approximate Jacobian, given in Jeyakumar and Luc [8], where the approximate Jacobian was assumed to be merely closed, not necessarily bounded, and the inequality (2.1) was given in terms of the lower Dini directional derivative $(vF)^-(x, .)$ of vF at x. The connections between the approximate Jacobian and the co-derivatives ([6, 7, 11, 12, 13, 14]) are discussed in [8].

Clearly, if $F : \mathbb{R}^n \to \mathbb{R}^m$ is continuously differentiable at x, then any compact subset Ω of $L(\mathbb{R}^n, \mathbb{R}^m)$ containing the Jacobian $\nabla F(x)^T$ is an approximate Jacobian of F at x. Here we write $\nabla F(x)$ for the usual $m \times n$ Jacobian matrix of partial derivatives.

Suppose that $F : \mathbb{R}^n \to \mathbb{R}^m$ is locally Lipschitz at x. Then the Clarke generalized Jacobian $\partial_C F(x)$ is an approximate Jacobian of F at x. Indeed, for each $v \in \mathbb{R}^m$,

$$\partial^\circ(vF)(x) = \partial_C F(x)v. \tag{2.3}$$

Consequently, for each $u \in \mathbb{R}^n$,

$$(vF)^+(x, u) \leq (vF)^\circ(x, u) = \max_{\xi \in \partial^\circ(vF)(x)} \langle \xi, u \rangle = \max_{M \in \partial_C F(x)} \langle Mv, u \rangle,$$

where

$$\partial_C F(x) = co\{ \lim_{n \to \infty} \nabla F(x_n)^T : x_n \in \Omega, \ x_n \to x \},$$

Ω is the set of points in $\mathbb{R}^n$ where F is differentiable and the Clarke directional derivative of vF is given by

$$(vF)^\circ(x, u) = \limsup_{\substack{x' \to x \\ t \downarrow 0}} \frac{\langle v, F(x' + tu) - F(x') \rangle}{t}.$$

For the locally Lipschitz map $F : \mathbb{R}^n \to \mathbb{R}^m$, the set

$$\partial_B F(x) := \{ \lim_{n \to \infty} \nabla F(x_n)^T : x_n \in \Omega, \ x_n \to x \}$$

is also an approximate Jacobian of F at x. The set $\partial_B F(x)$ is known as the B-subdifferential of F at x, which plays a significant role in the development of non-smooth Newton methods. Let us look at a numerical example of a locally Lipschitz map where the Clarke generalized Jacobian strictly contains the convex hull of an approximate Jacobian.

EXAMPLE 2.1. Consider the function $F : \mathbb{R}^2 \to \mathbb{R}^2$

$$F(x, y) = (|x| - |y|, |y| - |x|).$$

Then it can be verified that

$$\partial^* F(0) = \left\{ \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}, \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} \right\}$$

is an approximate Jacobian of F at 0. On the other hand, the Clarke generalized Jacobian is $\partial_C F(0) = co(\partial_B F(0))$, where

$$\partial_B F(0) = \left\{ \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}, \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}, \begin{pmatrix} -1 & 1 \\ -1 & 1 \end{pmatrix}, \begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix} \right\}.$$

Observe that

$$\begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix} \notin co(\partial^* F(0)).$$

Let us present some basic calculus for approximate Jacobians that will be useful later in applications to optimization problems.

PROPOSITION 2.1. Let $F : \mathbb{R}^n \to \mathbb{R}^m$ and let $0 \neq \alpha \in \mathbb{R}$. If $\partial^* F(x)$ is an approximate Jacobian of F at $x \in \mathbb{R}^n$ then $\alpha \partial^* F(x)$ is an approximate Jacobian of αF at x.

PROOF. The proof follows by standard arguments and so is omitted.

The following proposition shows that a simple sum formula holds for our approximate Jacobians. However, the general notion of an approximate Jacobian, given in [8], requires a regularity hypothesis.

PROPOSITION 2.2. Let $F_i : \mathbb{R}^n \to \mathbb{R}^m$ for $i = 1, 2$. Suppose that for each $i = 1, 2$, $\partial^* F_i(x)$ is an approximate Jacobian of F_i at x. Then the set $\partial^* F_1(x) + \partial^* F_2(x)$ is an approximate Jacobian of $F := F_1 + F_2$ at x.

PROOF. Let $v \in \mathbb{R}^m$, $u \in \mathbb{R}^n$ be arbitrary. The sub-additivity of upper limits shows that

$$(vF)^+(x, u) \leq \limsup_{t\downarrow 0} \frac{\langle v, F_1(x + tu) - F_1(x) \rangle}{t} + \limsup_{t\downarrow 0} \frac{\langle v, F_2(x + tu) - F_2(x) \rangle}{t}.$$

Since $\partial^* F_1(x)$ and $\partial^* F_2(x)$ are approximate Jacobians of F_1 and F_2 at x, it follows that

$$(vF)^+(x, u) \leq \max_{M \in \partial^* F_1(x)} \langle Mv, u \rangle + \max_{N \in \partial^* F_2(x)} \langle Nv, u \rangle = \langle (M_0 + N_0)v, u \rangle,$$

for some $M_0 \in \partial^* F_1(x)$ and $N_0 \in \partial^* F_2(x)$. Hence,

$$(vF)^+(x, u) \leq \max_{P \in \partial^* F_1(x) + \partial^* F_2(x)} \langle Pv, u \rangle$$

and so the conclusion follows.

In the following proposition we show how an approximate Jacobian of F can be constructed from the approximate Jacobians of its components. Recall that when $m = 1$, $\partial^* F(x)$ consists of $n \times 1$ matrices (that is, column vectors).

PROPOSITION 2.3. *Let* $F(x) = (f_1(x), f_2(x), \dots, f_m(x))$. *Let* $\partial^* f_i(x)$ *be the approximate Jacobian of* f_i *at* x. *Then the set*

$$\partial^* F(x) := \partial^* f_1(x) \times \partial^* f_2(x) \times \dots \times \partial^* f_m(x)$$

is an approximate Jacobian of F *at* x. *Here the latter set denotes the set of all matrices whose* i^{th} *column belongs to* $\partial^* f_i(x)$ *for each* i.

PROOF. Since for each i, $\partial^* f_i(x)$ is compact, $\partial^* F(x)$ is clearly a closed and bounded subset of $L(\mathbb{R}^n, \mathbb{R}^m)$. Let $u \in \mathbb{R}^n$ and let $v \in \mathbb{R}^m$. Then

$$(vF)^+(x, u) = \limsup_{t\downarrow 0} \frac{\sum_{i=1}^m v_i(f_i(x + tu) - f_i(x))}{t}$$

$$\leq \sum_{i=1}^m \limsup_{t\downarrow 0} \frac{v_i(f_i(x + tu) - f_i(x))}{t}.$$

So,

$$(vF)^+(x, u) \leq \sum_{i=1}^m \max_{M_i \in \partial^* f_i(x)} \langle M_i v_i, u \rangle = \sum_{i=1}^m \langle M_i^o v_i, u \rangle = \langle \sum_{i=1}^m M_i^o v_i, u \rangle = \langle M^o v, u \rangle,$$

where $M_i^o \in \partial^* f_i(x)$ and $M^o = (M_1^o, M_2^o, \dots, M_m^o)$. Hence,

$$(vF)^+(x, u) \leq \langle M^o v, u \rangle \leq \max_{M \in \partial^* F(x)} \langle Mv, u \rangle$$

and so $\partial^* F(x)$ is an approximate Jacobian of F at x.

The following mean-value theorem holds for continuous maps as a special case of the corresponding result in [8] since the approximate Jacobians here are closed and bounded sets.

THEOREM 2.1. ([8]) *Let* $a, b \in \mathbb{R}^n$ *and let* $F : \mathbb{R}^n \to \mathbb{R}^m$ *be continuous. Assume that for each* $x \in [a, b]$, $\partial^* F(x)$ *is an approximate Jacobian of* F *at* x. *Then*

$$F(b) - F(a) \in co(\partial^* F([a, b])(b - a)). \qquad (2.4)$$

Note that the right-hand side in (2.4) is the convex hull of all points of the form $M(b - a)$ where $M \in \partial^* F(\zeta)$ for some $\zeta \in [a, b]$. The following example illustrates that the previous theorem provides a sharper mean-value condition even for a locally Lipschitz map than the corresponding result of Clarke [2].

EXAMPLE 2.2. Let $F : \mathbb{R}^2 \to \mathbb{R}^2$ be defined by

$$F(x, y) = (|x| - |y|, |y| - |x|).$$

Let $a = (-1, -1)$ and $b = (1, 1)$. Then the mean value condition (2.4) is verified by

$$\partial^* F(0) = \left\{ \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}, \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} \right\}.$$

Moreover, condition (2.4) yields the mean value condition in the sense of Clarke (see [4]), since $co(\partial^* F(0)) \subset \partial_C F(0)$.

It is also worth noting that it easily follows from Mean-Value Theorem 2.1 that if an approximate Jacobian of a continuous map F is bounded in a neighbourhood of a then F is locally Lipschitz at a (see [8] for further details).

Let us now introduce the notion of approximate Hessian for continuously differentiable (C^1-)functions. Note that the derivative of f which is denoted by ∇f is a continuous map from $\mathbb{R}^n$ to $\mathbb{R}^n$; however, it is not necessarily locally Lipschitz.

DEFINITION 2.2. The function f admits an **approximate Hessian** $\partial_*^2 f(x)$ at x if this set is an approximate Jacobian to ∇f at x.

Note that $\partial_*^2 f(x) = \partial^* \nabla f(x)$ and the matrix $M \in \partial_*^2 f(x)$ is an approximate Hessian matrix of f at x. Clearly, if f is twice differentiable at x, then $\nabla^2 f(x)$ is a symmetric approximate Hessian matrix of f at x.

Recall that if $f : \mathbb{R}^n \to \mathbb{R}$ is $C^{1,1}$, then the generalized Hessian in the sense of Hiriart-Urruty *et al.* [5] is given by

$$\partial_H^2 f(x) = co\{M : M = \lim_{n \to \infty} \nabla^2 f(x_n), \ x_n \in \Delta, \ x_n \to x\},$$

where Δ is the set of points in $\mathbb{R}^n$ where f is twice differentiable. Clearly, $\partial^2_H f(x)$ is a nonempty convex compact set of symmetric matrices. The second-order directional derivative of f at x in the directions $(u, v) \in \mathbb{R}^n \times \mathbb{R}^n$ is defined by

$$f^{\circ\circ}(x; u, v) = \limsup_{\substack{y \to x \\ s \to 0}} \frac{\langle \nabla f(y + su), v \rangle - \langle \nabla f(y), v \rangle}{s}.$$

Since for each $(u, v) \in \mathbb{R}^n \times \mathbb{R}^n$,

$$(v \nabla f)^+(x, u) \le f^{\circ\circ}(x; u, v) = \max_{M \in \partial^2_H f(x)} \langle M u, v \rangle = \max_{M \in \partial^2_H f(x)} \langle M v, u \rangle,$$

$\partial^2_H f(x)$ is an approximate Hessian of f at x. For a related notion of approximate Hessian, see [10].

The following generalized Taylor's expansion holds for C^1-functions as a special case of the corresponding result in [8]. For completeness we here provide a strengthened form of the proof given in [8] (see also [1]).

THEOREM 2.2. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be continuously Gâteaux differentiable on $\mathbb{R}^n$; let $x, y \in \mathbb{R}^n$. Suppose that for each $z \in [x, y]$, $\partial^2_* f(z)$ is an approximate Hessian of f at z. Then there exists $\zeta \in (x, y)$ such that*

$$f(y) \in f(x) + \langle \nabla f(x), y - x \rangle + \frac{1}{2} co \langle \partial^2_* f(\zeta)(y - x), (y - x) \rangle.$$

PROOF. Let $h(t) = f(y + t(x - y)) + t \langle \nabla f(y + t(x - y)), y - x \rangle + \frac{1}{2} a t^2 - f(y)$, where $a = -2(f(x) - f(y) + \langle \nabla f(x), y - x \rangle)$. Then $h(0) = 0, h(1) = f(x) - f(y) + \langle \nabla f(x), y - x \rangle + \frac{1}{2} a = 0$ and h is continuous. So h attains its extremum at some $\gamma \in (0, 1)$. Suppose that γ is a minimum point of h. Now by necessary conditions, we have for each $v \in \mathbb{R}$, $h^+(\gamma; v) \ge 0$. Let

$$g(t) = \langle \nabla f(y + t(x - y)), y - x \rangle.$$

Then by direct calculations

$$h^+(\gamma; 1) = -g(\gamma) + g(\gamma) + \gamma g^+(\gamma; 1) + a\gamma$$
$$= a\gamma + \gamma \limsup_{\lambda \to 0^+} \langle \frac{\nabla f(\zeta + \lambda(x - y)) - \nabla f(\zeta)}{\lambda}, y - x \rangle,$$

where $\zeta = y + \gamma(x - y) \in (x, y)$. Since $h^+(\gamma; 1) \ge 0$,

$$a \ge \liminf_{\lambda \to 0^+} \langle \frac{\nabla f(\zeta + \lambda(x - y)) - \nabla f(\zeta)}{\lambda}, x - y \rangle.$$

Now, from (2.1) we see that

$$a \geq \min_{M \in \partial_*^2 f(\zeta)} \langle M(x - y), x - y \rangle.$$

Similarly, by calculating $h^+(\gamma; -1)$ we obtain that

$$a \leq \max_{M \in \partial_*^2 f(\zeta)} \langle M(x - y), x - y \rangle.$$

Hence, it follows that

$$\min_{M \in \partial_*^2 f(\zeta)} \langle M(y - x), y - x \rangle \leq a \leq \max_{M \in \partial_*^2 f(\zeta)} \langle M(y - x), y - x \rangle,$$

and so,

$$a \in co \langle \partial_*^2 f(\zeta)(y - x), (y - x) \rangle.$$

Thus

$$f(y) - f(x) - \langle \nabla f(x), y - x \rangle = \frac{a}{2} \in \frac{1}{2} co \langle \partial_*^2 f(\zeta)(y - x), (y - x) \rangle. \qquad (2.5)$$

The case where γ is a maximum point of h also yields the same condition (2.5). The details are left to the reader.

3. Second-order necessary optimality conditions

Consider the nonlinear programming problem

(P) Minimize $f(x)$

subject to $x \in \mathbb{R}^n$, $g_i(x) \leq 0$, $i = 1, 2, \ldots, p$; $h_j(x) = 0$, $j = 1, 2, \ldots, q$,

where $f, g_i, i = 1, 2, \ldots, p$ and $h_j, j = 1, 2, \ldots, q$ are C^1-functions on $\mathbb{R}^n$. The Lagrangian function is given by

$$L(x, \lambda, \mu) = f(x) + \sum_{i=1}^{p} \lambda_i g_i(x) + \sum_{j=1}^{q} \mu_j h_j(x).$$

If a is a local minimum of (P) then there exist $\lambda_0^* \geq 0$, $\lambda_i^* \geq 0$ and $\lambda_i^* g_i(a) = 0$, for $i = 1, 2, \ldots, p$, $\mu^* \in \mathbb{R}^q$ such that

$$\lambda_0^* \nabla f(a) + \sum_{i=1}^{p} \lambda_i^* \nabla g_i(a) + \sum_{j=1}^{q} \mu_j^* \nabla h_j(a) = 0.$$

This Fritz-John condition yields the following Kuhn-Tucker conditions under a first-order constraint qualification (see [3]):

$$\lambda_i^* \geq 0, \quad \lambda_i^* g_i(a) = 0, \text{ for } i = 1, 2, \ldots, p, \quad \mu \in \mathbb{R}^q, \quad \nabla L(a, \lambda^*, \mu^*) = 0.$$

Here $\lambda_0 = 1$. Let $C = \{x \in \mathbb{R}^n \mid g_i(x) \leq 0, i = 1, 2, \ldots, p, h_j(x) = 0, j = 1, 2, \ldots, q\}$ be the feasible set and let

$$C(\lambda) = \{x \in C \mid \sum_{i=1}^{p} \lambda_i g_i(x) = 0\}.$$

The *cone of feasible directions* to a subset S of $\mathbb{R}^n$ at $x \in S$ is given by

$$F(S, x) = \{u \in \mathbb{R}^n \mid \exists \delta > 0, \quad \forall 0 \leq \alpha \leq \delta, \quad x + \alpha u \in S\}.$$

THEOREM 3.1. *Assume that the problem* (P) *attains a local minimum at* a. *Suppose that for each* $\lambda \in \mathbb{R}_+^p$ *and* $\mu \in \mathbb{R}^q$, $L(., \lambda, \mu)$ *admits an approximate Hessian* $\partial_*^2 L(a, \lambda, \mu)$ *at* a. *If a first- order constraint qualification holds at* a *then there exist* $\lambda_i^* \geq 0, \lambda_i^* g_i(a) = 0$, *for* $i = 1, 2, \ldots, p$, $\mu^* \in \mathbb{R}^q$, $\nabla L(a, \lambda^*, \mu^*) = 0$ *and*

$$(\forall u \in F(C(\lambda^*), a)) \ (\exists M \in \partial_*^2 L(a, \lambda^*, \mu^*)) \ \langle Mu, u \rangle \geq 0. \tag{3.1}$$

PROOF. From the Karush-Kuhn-Tucker conditions we get

$$
\begin{aligned}
(u \nabla L)^+((a, \lambda^*, \mu^*), u) &= \limsup_{t \downarrow 0} \frac{(u \nabla L)(a + tu, \lambda^*, \mu^*) - (u \nabla L)(a, x^*, \mu^*)}{t} \\
&= \limsup_{t \downarrow 0} \frac{\langle u, \nabla L(a + tu, \lambda^*, \mu^*) \rangle - \langle u, \nabla L(a, \lambda^*, \mu^*) \rangle}{t} \\
&= \limsup_{t \downarrow 0} \frac{\langle u, \nabla L(a + tu, \lambda^*, \mu^*) \rangle}{t}.
\end{aligned}
$$

Let $u \in F(C(\lambda^*), a)$. Then there exists $\delta > 0$ such that for all $0 \leq \alpha \leq \delta$, $a + \alpha u \in C(\lambda^*)$. Then for $0 < \frac{1}{k} < \delta$,

$$L(a + \frac{1}{k}u, \lambda^*, \mu^*) = f(a + \frac{1}{k}\mu) \geq f(a) = L(a, \lambda^*, \mu^*).$$

It then follows from the standard Mean-Value Theorem that there exists an integer $N > 0$ and $0 < t_k < \frac{1}{k}$ for each $k > N$, such that

$$\langle u, \nabla L(a + t_k u, \lambda^*, \mu^*) \rangle \geq 0, \ \forall k \geq N.$$

So,

$$\limsup_{t \downarrow 0} \frac{\langle u, \nabla L(a + tu, \lambda^*, \mu^*) \rangle}{t} \geq 0.$$

Since $\partial_*^2 L(a, \lambda^*, \mu^*)$ is an approximate Hessian of $L(., \lambda^*, \mu^*)$ at a, we have for any $u \in F(C(\lambda^*), a)$

$$\max_{N \in \partial_*^2 L(a, \lambda^*, \mu^*)} \langle Nu, u \rangle \geq (u\nabla L)^+((a, \lambda^*, \mu^*), u)$$

$$= \limsup_{t \downarrow 0} \frac{\langle u, \nabla L(a + tu, \lambda^*, \mu^*) \rangle}{t}$$

$$\geq 0.$$

It then follows from the fact that $\partial_*^2 L(a, \lambda^*, \mu^*)$ is closed and bounded that there exists $M \in \partial_*^2 L(a, \lambda^*, \mu^*)$ such that

$$\langle Mu, u \rangle = \max_{N \in \partial_*^2 L(a, \lambda^*, \mu^*)} \langle Nu, u \rangle \geq 0.$$

COROLLARY 3.1. *Assume that the problem* (P) *attains a local minimum at* a. *Suppose that* $\partial_*^2 f(a)$, $\partial_*^2 g_i(a)$, *for* $i = 1, 2, \ldots, p$ *and* $\partial_*^2 h_j(a)$, *for* $j = 1, 2, \ldots, q$ *are approximate Hessians of* f, g_i *and* h_j *at* a *respectively. If a first- order constraint qualification holds at* a *then there exist* $\lambda_i^* \geq 0$ *such that* $\lambda_i^* g_i(a) = 0$, *for* $i = 1, 2, \ldots, p$, $\mu^* \in \mathbb{R}^q$ *such that* $\nabla L(a, \lambda^*, \mu^*) = 0$ *and for each* $u \in F(C(\lambda^*), a))$ *there exist* $M_0 \in \partial_*^2 f(a)$, $M_i \in \partial_*^2 g_i(a)$ *and* $M_j \in \partial_*^2 h_j(a)$ *such that*

$$\langle (M_0 + \sum_{i=1}^{p} \lambda_i M_i + \sum_{j=1}^{q} \mu_j M_j)u, u \rangle \geq 0.$$

PROOF. From Propositions 2.1 and 2.2, we know that the set

$$\partial_*^2 L(a, \lambda^*, \mu^*) = \partial_*^2 f(a) + \sum_{i=1}^{p} \lambda_i^* \partial_*^2 g_i(a) + \sum_{j=1}^{q} \mu_j^* \partial_*^2 h_j(a)$$

is an approximate Hessian of $L(x, \lambda^*, \mu^*)$ at a. Thus by Theorem 3.1 we can find a matrix $M \in \partial_*^2 L(a, \lambda^*, \mu^*)$ such that for any $u \in F(C(\lambda^*), a)$

$$(Mu, u) \geq 0.$$

Since $M = M_0 + \sum_{i=1}^{p} \lambda_i^* M_i + \sum_{j=1}^{q} \mu_j^* M_j$ for some $M_0 \in \partial_*^2 f(a)$, $M_i \in \partial_*^2 q_i(a)$, $i = 1, 2, \ldots, p$ and $M_j \in \partial_*^2 q_j(a)$, $j = 1, 2, \ldots, q$, we then have

$$\langle (M_0 + \sum_{i=1}^{p} \lambda_i^* M_i + \sum_{j=1}^{q} \mu_j^* M_j)u, u \rangle \geq 0, \forall u \in F(C(\lambda^*), a).$$

A stronger necessary optimality condition than (3.1) can be obtained in terms of the Bouligand tangent cone [5] under an additional hypothesis on the approximate Hessians. Recall that the *Bouligand tangent cone* to the set S at x is given by

$$B(S, x) = \{u \in \mathbb{R}^n \mid \exists s_k \downarrow 0, u_k \to u, \ x + s_k u_k \in S\}.$$

DEFINITION 3.1. Let $S \subset \mathbb{R}^n$, let $f : \mathbb{R}^n \to \mathbb{R}$ be C^1 and let $a \in S$. We say that the approximate Hessian set-valued map $\partial_*^2 f : x \Longrightarrow \partial_*^2 f(x)$ is *regular at* a with respect to S if for each $u \in S$

$$\limsup_{\substack{A' \in \partial_*^2 f(a + tu') \\ u' \to u, \, t \downarrow 0}} \langle A'u', u' \rangle \leq \max_{A \in \partial_*^2 f(a)} \langle Au, u \rangle. \tag{3.2}$$

Condition (3.2) means that for each $u \in S$ and for each sequence $u_n \to u, \quad t_n \downarrow 0$, $A_n \in \partial_*^2 f(a + t_n u_n)$ and

$$\limsup_{n \to \infty} \langle A_n u_n, u_n \rangle \leq \max_{A \in \partial_*^2 f(a)} \langle Au, u \rangle.$$

It is easy to see from the definition that if the map $\partial_*^2 f$ is locally bounded at a then

$$\limsup_{\substack{A' \in \partial_*^2 f(a + tu') \\ u' \to u, \, t \downarrow 0}} \langle A'u', u' \rangle$$

is finite. We now see that upper semi-continuity of the map $\partial_*^2 f$ at a guarantees regularity at a. Note first that a set-valued mapping $G : \mathbb{R}^n \to L(\mathbb{R}^n, \mathbb{R}^m)$ is *locally bounded* at x if there exist a neighbourhood U of x and a positive α such that $\|A\| \leq \alpha$, for each $A \in G(U)$. The map G is said to be *upper semi-continuous* at x if for each open set V containing $G(x)$, there is a neighbourhood U of x such that $G(U) \subset V$. It then follows from the definitions that if the map $\partial_*^2 f$ is upper semi-continuous at x then it is locally bounded at x since $\partial_*^2 f(x)$ is bounded.

LEMMA 3.1. *Let f be a C^1-function, let $\partial_*^2 f(x)$ be an approximate Hessian of f for each $x \in \mathbb{R}^n$ and let $a \in S \subset \mathbb{R}^n$. If the set-valued map $\partial_*^2 f$ is upper semi-continuous at a, then $\partial_*^2 f$ is regular at a with respect to S.*

PROOF. Let $u \in S$ and let the sequences $u_n \to u, \quad t_n \downarrow 0$ and $A_n \in \partial_*^2 f(a + t_n u_n)$. Since $\partial_*^2 f$ is locally bounded

$$l := \limsup_{\substack{A' \in \partial_*^2 f(a + tu') \\ u' \to u, \, t \downarrow 0}} \langle A'u', u' \rangle$$

is finite. Suppose that

$$l > \max_{A \in \partial_*^2 f(a)} \langle Au, u \rangle = \langle A_o u, u \rangle,$$

where $A_o \in \partial_*^2 f(a)$. Define $\epsilon = l - \langle A_o u, u \rangle > 0$. Then there exists a subsequence, again denoted by $\langle A_n u_n, u_n \rangle$ such that

$$\langle A_o u, u \rangle = l - \epsilon < \lim_{n \to \infty} \langle A_n u_n, u_n \rangle.$$

Since $\partial_*^2 f$ is upper semi-continuous at a, we can find a subsequence $A_{k_n} \in \partial_*^2 f(a + t_{k_n} u_{k_n})$ such that $A_{k_n} \to \bar{A} \in \partial_*^2 f(a)$ as $n \to \infty$. Hence

$$\langle A_o u, u \rangle < \lim_{n \to \infty} \langle A_n u_n, u_n \rangle$$
$$= \langle \bar{A} u, u \rangle \le \langle A_o u, u \rangle,$$

which is a contradiction and so

$$l \le \max_{A \in \partial_*^2 f(a)} \langle A u, u \rangle.$$

Clearly if f is twice continuously differentiable then $\partial_*^2 f(.) = \{\nabla^2 f(.)\}$ is regular at x with respect to each subset S of $\mathbb{R}^n$. If f is $C^{1,1}$ then $\partial_*^2 f := \partial_H^2 f$ is regular at each point. In other words, condition (3.2) is satisfied for a $C^{1,1}$-function by $\partial_*^2 f = \partial_H^2 f$. The following example shows that an approximate Hessian set-valued map of a $C^{1,1}$-function, which is not upper semi-continuous, satisfies the regularity condition (3.2).

EXAMPLE 3.1. Let $h : \mathbb{R} \to \mathbb{R}$ be an odd function which is defined for $x \ge 0$ by

$$h(x) = \begin{cases} 2x - 1, & x \ge \frac{1}{2}; \\ -x + \frac{1}{2^{2n-1}}, & x \in \left[\frac{1}{2^{2n}}, \frac{1}{2^{2n-1}}\right], n = 1, 2, \ldots; \\ 2x - \frac{1}{2^{2n}}, & x \in \left[\frac{1}{2^{2n+1}}, \frac{1}{2^{2n}}\right], n = 1, 2, \ldots; \\ 0, & x = 0. \end{cases}$$

Define $f : \mathbb{R}^2 \to \mathbb{R}$ by

$$f(x_1, x_2) = \int_0^{|x_1|} h(t) dt + \frac{x_2^2}{2}.$$

Then f is a $C^{1,1}$-function since $\nabla f(x_1, x_2) = (h(x_1), x_2)$ is a locally Lipschitz function. An approximate Hessian set-valued map $\partial_*^2 f$ is given by

$$\partial_*^2 f(x_1, x_2) = \begin{cases} \left\{ \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} \right\}, & x_1 = \pm\frac{1}{2^n}, n = 1, 2, \ldots; \\ \left\{ \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} \right\}, & x_1 = 0; \\ \left\{ \begin{pmatrix} h'(x_1) & 0 \\ 0 & 1 \end{pmatrix} \right\}, & \text{otherwise.} \end{cases}$$

It is easy to verify that $\partial_*^2 f$ is regular at $(0, 0)$ and locally bounded at $(0, 0)$. However, it is not upper semi-continuous at $(0, 0)$ since

$$\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \in \partial_*^2 f((\frac{1}{2^n}, 0)); \quad \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \notin \partial_*^2 f((0, 0)).$$

It is also worth noting that

$$\partial_H^2 f((0, 0)) = \left\{ \begin{pmatrix} \alpha & 0 \\ 0 & 1 \end{pmatrix} \mid \alpha \in [-1, 2] \right\}$$

and that $co(\partial_*^2 f((0, 0)) \subset \partial_H^2 f((0, 0))$.

THEOREM 3.2. *Assume that the problem (P) attains a local minimum at a. Let the Kuhn-Tucker conditions be satisfied at a by λ_i^* and μ_j^*. Suppose that for each $x \in \mathbb{R}^n$, $\partial_*^2 L(x, \lambda, \mu)$ is an approximate Hessian of $L(., \lambda^*, \mu^*)$ at x. If the set-valued map $\partial_*^2 L(\cdot, \lambda^*, \mu^*)$ is locally bounded at a and regular at a with respect to $B(C(\lambda^*), a)$ then*

$$(\forall u \in B(C(\lambda^*), a)) \, (\exists M \in \partial_*^2 L(a, \lambda^*, \mu^*)) \, \langle Mu, u \rangle \geq 0. \tag{3.3}$$

PROOF. Let $u \in B(C(\lambda^*), a)$. Then there exist sequences $t_k \downarrow 0$ and $u_k \rightarrow u$ as $k \rightarrow \infty$ such that, for every k, $a + t_k u_k \in C(\lambda^*)$. So

$$L(a + t_k u_k, \lambda^*, \mu^*) = f(a + t_k u_k).$$

Now it follows from the generalized Taylor's expansion (Theorem 2.2) that

$$L(a + t_k u_k, \lambda^*, \mu^*) \leq L(a, \lambda^*, \mu^*) + t_k \langle \nabla L(a, \lambda^*, \mu^*), u_k \rangle + \frac{t_k^2}{2} \langle N_k u_k, u_k \rangle,$$

where $N_k \in \partial_*^2 L(a + \bar{t}_k u_k, \lambda^*, \mu^*)$ and $0 < \bar{t}_k < t_k$. Noting that a is a local minimum of (P), we get

$$L(a, \lambda^*, \mu^*) = f(a),$$
$$\nabla L(a, \lambda^*, \mu^*) = 0,$$
$$f(a + t_k u_k) \geq f(a),$$

for sufficiently large k. Thus for sufficiently large k, $\langle N_k u_k, u_k \rangle \geq 0$. Since the set-valued map $\partial_*^2 f$ is locally bounded at a, the sequence N_k is bounded. Hence this sequence has a subsequence, again denoted by N_k which converges to a matrix N. As $k \rightarrow \infty$, the sequence $a + \bar{t}_k u_k \rightarrow a$. Then it follows that

$$\langle Nu, u \rangle = \lim_{k \rightarrow \infty} \langle N_k u_k, u_k \rangle \geq 0.$$

Hence

$$\limsup_{\substack{A' \in \partial_*^2 f(a+tu') \\ u' \to u, \, t \downarrow 0}} \langle A'u', u' \rangle \geq \langle Nu, u \rangle \geq 0,$$

and so by the regularity assumption we get that $\max_{A \in \partial_*^2 f(a)} \langle Au, u \rangle \geq 0$.

COROLLARY 3.2. *Assume that the problem* (P) *attains a local minimum at* a. *Suppose that the Kuhn-Tucker conditions are satisfied at* a *by* λ_i^* *and* μ_j^*. *If* $L(., \lambda^*, \mu^*)$ *admits an approximate Hessian* $\partial_*^2 L(a, \lambda, \mu)$ *at* a *and if the set-valued map* $\partial_*^2 L(\cdot, \lambda^*, \mu^*)$ *is upper semi-continuous at* a *then*

$$(\forall u \in B(C(\lambda^*), a)) \ (\exists M \in \partial_*^2 L(a, \lambda^*, \mu^*)) \ (Mu, u) \geq 0. \tag{3.4}$$

PROOF. The conclusion follows from Theorem 3.2 and Lemma 3.1.

COROLLARY 3.3. *Assume that the functions* f, g_i *and* h_j, *for each* i, j *in problem* (P) *are* $C^{1,1}$ *and that the problem* (P) *attains a local minimum at* a. *If a first-order constraint qualification holds at* a *then there exist* $\lambda_i^* \geq 0$, $\lambda_i^* g_i(a) = 0$, *for* $i = 1, 2, .., p$, $\mu^* \in \mathbb{R}^q$, $\nabla L(a, \lambda^*, \mu^*) = 0$ *and*

$$(\forall u \in B(C(\lambda^*), a)) \ (\exists M \in \partial_H^2 L(a, \lambda^*, \mu^*)) \ (Mu, u) \geq 0. \tag{3.5}$$

PROOF. Choose $\partial_H^2 L(a, \lambda^*, \mu^*)$ as the approximate Hessian of $L(\cdot, \lambda^*, \mu^*)$ at a. The result then follows from Theorem 3.2 since the map $\partial_H^2 L(\cdot, \lambda^*, \mu^*)$ is upper semi-continuous at a.

The following example shows that Theorem 3.2 provides sharper optimality conditions than the conditions of Corollary 3.3 (see [5]).

EXAMPLE 3.2. Consider the problem

$$\text{Minimize} \quad \int_0^{|x_1|} h(t)dt + \frac{x_2^2}{2}$$

$$\text{subject to } x_1 \geq 0, \quad x_2 \geq 0,$$

where $f(x_1, x_2) = \int_0^{|x_1|} h(t)dt + \frac{x_2^2}{2}$, $g_1(x_1, x_2) = x_1$, $g_2(x_1, x_2) = x_2$ and h is given as in Example 3.1. Then f is a $C^{1,1}$ function. The point $(0, 0)$ is a minimum point of the problem. The Kuhn-Tucker conditions are satisfied at $(0, 0)$ by $\lambda^* = (\lambda_1^*, \lambda_2^*) = (0, 0)$ and condition (3.3) is verified by the matrix

$$\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \in \partial_*^2 L((0, 0), \lambda^*) = \partial_*^2 f((0, 0)) \subset \partial_H^2 f((0, 0)),$$

for each $(u_1, u_2) \in B(C(\lambda^*), (0, 0)) = \{(x_1, x_2) \in \mathbb{R}^2 \mid x_1 \geq 0, \ x_2 \geq 0\}$.

It is known that under a second-order constraint qualification (see [3]) the necessary optimality conditions in Theorem 3.2 and Corollary 3.2 can be expressed in terms of the gradients of the functions g_i and h_j. A second-order constraint qualification such as the linear independence condition that the vectors

$$\nabla g_i(a), \quad i \in I(a), \quad \nabla h_j(a), \quad j = 1, 2, \dots, q$$

are linear independent guarantees that $u \in B(C(\lambda^*), a)$ if and only if $\langle \nabla g_i(a), u \rangle = 0$ for i such that $\lambda_i^* > 0$, $\langle \nabla g_i(a), u \rangle \leq 0$ for i such that $\lambda_i^* = 0$ and $g_i(a) = 0$ and $\langle \nabla h_j(a), u \rangle = 0$ for $j = 1, 2, \dots, q$. Here $I(a) = \{i : g_i(a) = 0\}$.

4. Second-order sufficient conditions

In this section, we present sufficient optimality conditions in terms of approximate Hessian for a feasible point to be a local minimum of (P).

Let $J = \{i \in I(a) : \lambda_i > 0\}$ and let $B_n = \{y \in \mathbb{R}^n : ||y|| = 1\}$. Define

$$Y = \{y \in B_n : y^T \nabla g_i(a) = 0, \ i \in J, \ y^T \nabla h_j(a) = 0, \ j = 1, 2, \dots, q\}$$

and for $\epsilon > 0$ and $\delta > 0$ define

$$Z(\epsilon, \delta) = \{u \in B_n : ||u - y|| < \epsilon, \text{ for some } y \in Y,$$
$$\text{and } a + \bar{\delta}(u)u \in C, \text{ for some } 0 < \bar{\delta}(u) < \delta\}.$$

THEOREM 4.1. *Let a be a feasible point for (P). Suppose that the Kuhn-Tucker conditions are satisfied at a by λ_i^* and μ_j^*. Assume that for each x in a neighbourhood of a, $\partial_*^2 L(x, \lambda^*, \mu^*)$ is an approximate Hessian of $L(., \lambda^*, \mu^*)$ at a. If there exist $\epsilon > 0$ and $\delta > 0$ such that for each $u \in Z(\epsilon, \delta)$ and for each $0 < \alpha < 1$,*

$$\forall M \in \partial_*^2 L(a + \alpha u, \lambda^*, \mu^*) \quad \langle Mu, u \rangle \geq 0$$

then a is a local minimum of (P).

PROOF. If a is not a local minimum, then there exists $\{x_k\}$ such that x_k is feasible for (P), $x_k \to a$ as $k \to +\infty$, and $f(x_k) < f(a)$ for each k.

Let $x_k = a + \delta_k u_k$, where $||u_k|| = 1$, $\delta_k > 0$, $\delta_k \to 0$ as $k \to +\infty$. Since $||u_k|| = 1$, the sequence $\{u_k\}$ has a convergence subsequence. Without loss of generality, we assume that $u_k \to y$ as $k \to +\infty$, with $||y|| = 1$.

By the Mean-Value Theorem, we have

$$0 > f(x_k) - f(a) = \delta_k u_k^T \nabla f(a + \eta_{0k}\delta_k u_k), \ 0 < \eta_{0k} < 1,$$
$$0 \geq g_i(x_k) - g_i(a) = \delta_k u_k^T \nabla g_i(a + \eta_{ik}\delta_k u_k), \ 0 < \eta_{ik} < 1, \ \forall i \in I(a),$$
$$0 = h_j(x_k) - h_j(a) = \delta_k u_k^T \nabla h_j(a + \xi_{jk}\delta_k u_k), \ 0 < \xi_{jk} < 1, \ \forall j = 1, \dots, q.$$

Dividing the above inequalities and the equality by δ_k and taking limits as $k \to +\infty$, we obtain

$$y^T \nabla f(a) \leq 0, \ y^T \nabla g_i(a) \leq 0, \ \forall i \in I(a), \ y^T \nabla h_j(a) = 0, \ \forall j.$$

Suppose that $y^T \nabla g_i(a) < 0$, for at least one $i \in J$. Then we get

$$0 \geq y^T \nabla f(a) = -\sum_{i \in J} \lambda_i^* y^T \nabla g_i(a) - \sum_{j=1}^{q} \mu_j^* y^T \nabla h_j(a) > 0.$$

This is a contradiction. Thus $y^T \nabla g_i(a) = 0$, for all $i \in J$, or $J = \phi$. Then $y \in Y$. Since the Kuhn-Tucker conditions are satisfied at a by λ_i^*, μ_j^*, we have

$$\lambda_i^* \geq 0, \ \lambda_i^* g_i(a) = 0, \ i = 1, \ldots, p, \tag{4.1}$$

$$\nabla L(a, \lambda^*, \mu^*) = \nabla f(a) + \sum_{i \in I(a)} \lambda_i^* \nabla g_i(a) + \sum_{j=1}^{q} \mu_j^* \nabla h_j(a) = 0. \tag{4.2}$$

It follows from the inequality $f(a) > f(x_k)$, (4.1), (4.2) and the Taylor's expansion for $L(x, \lambda^*, \mu^*)$ at a (Theorem 2.2) that

$$f(a) > f(x_k)$$

$$\geq f(x_k) + \sum_{i \in I(a)} \lambda_i^* g_i(x_k) + \sum_{j=1}^{q} \mu_j^* h_j(x_k)$$

$$\geq f(a) + \sum_{i \in I(a)} \lambda_i^* g_i(a) + \sum_{j=1}^{q} \mu_j^* h_j(a)$$

$$+ \delta_k u_k^T (\nabla f(a) + \sum_{i \in I(a)} \lambda_i^* \nabla g_i(a) + \sum_{j=1}^{q} \mu_j^* \nabla h_j(a))$$

$$+ \frac{1}{2} \min_{M_k \in co\partial_*^2 L(a+\theta_k \delta_k u_k, \ \lambda^*, \mu^*)} \langle M_k \delta_k u_k, \ \delta_k u_k \rangle$$

$$= f(a) + \frac{1}{2} \delta_k^2 \min_{M_k \in \partial_*^2 L(a+\theta_k \delta_k u_k, \lambda^*, \mu^*)} \langle M_k u_k, \ u_k \rangle$$

$$= f(a) + \frac{1}{2} \delta_k^2 \langle M_k^0 u_k, \ u_k \rangle$$

for some $M_k^0 \in \partial_*^2 L(a + \theta_k \delta_k u_k, \lambda^*, \mu^*)$ and $0 < \theta_k < 1$.
Hence for any k

$$0 > \langle M_k^0 u_k, \ u_k \rangle. \tag{*}$$

By construction, $\|u_k\| = 1$, $u_k \to y \in Y$, $\delta_k \to 0$ as $k \to +\infty$, $0 < \theta_k \delta_k < 1$ when k is large and $a + \delta_k u_k$ is feasible for every k. Hence for k large, $u_k \in Z(\epsilon, \delta)$ and by assumption

$$\langle M_k^0 u_k, u_k \rangle \geq 0.$$

This is a contradiction to $(*)$. Thus a is a local minimum of (P).

THEOREM 4.2. *Let a be a feasible point for (P). Suppose that the Kuhn-Tucker conditions are satisfied at a by λ_i^* and μ_j^*. Assume that for each x in a neighbourhood of a, $\partial_*^2 L(x, \lambda^*, \mu^*)$ is an approximate Hessian of $L(., \lambda^*, \mu^*)$ at a. If there exist $\epsilon > 0$ and $\delta > 0$ such that for each $u \in Z(\epsilon, \delta)$ and for each $0 < \alpha < 1$,*

$$\forall M \in \partial_*^2 L(a + \alpha u, \lambda^*, \mu^*) \quad \langle Mu, u \rangle > 0$$

then a is a strict local minimum of (P).

PROOF. The proof is only a slight modification of that of Theorem 4.1 and so is omitted.

In this section we have shown how the descriptions of second-order sufficient conditions in terms of generalized Hessian matrices can be improved and sharpened using approximate Hessian matrices. In [16], second-order sufficient conditions for $C^{1,1}$ problems were given in terms of a generalized second-order directional derivatives. However, this directional derivative does not appear to have a matrix representation of the form (2.1) which allowed us to directly compare the corresponding results in [5].

Acknowledgment

References

[1] W. L. Chan, L. R. Huang and K. F. Ng, 'On generalized second-order derivatives and Taylor expansions in nonsmooth optimization', *SIAM J. Control Optim.* **32** (1994) 591–611.
[2] F. H. Clarke, *Optimization and nonsmooth analysis* (Wiley-Interscience, New York, 1983).
[3] A. V. Fiacco and G. P. McCormick, *Nonlinear Programming: Sequential unconstrained minimization techniques* (SIAM Publication, Philadelphia, USA, 1990).

[4] J. B. Hiriart-Urruty, 'Mean value theorems for vector valued mappings in nonsmooth optimization', *Numer. Funct. Anal. Optim.* **2** (1980) 1–30.

[5] J. B. Hiriart-Urruty, J. J. Strodiot and V. Hien Nguyen, 'Generalized Hessian matrix and second-order optimality conditions for problems with $C^{1,1}$ data', *Applied Math. Optimiz.* **11** (1984) 43–56.

[6] A. D. Ioffe, 'Nonsmooth analysis: differential calculus of nondifferential mappings', *Trans. Amer. Math. Soc.* **266** (1981) 1–56.

[7] A. D. Ioffe, 'Approximate subdifferentials and applications I: The finite dimensional theory', *Trans. Amer. Math. Soc.* **281** (1984) 389–416.

[8] V. Jeyakumar and D. T. Luc, 'Approximate Jacobian matrices for nonsmooth continuous maps and C^1-optimization', *SIAM J. Control and Optim.* **36**(5) (1998) 1815–1832.

[9] V. Jeyakumar, D. T. Luc and S. Schaible, 'Characterizations of generalized monotone nonsmooth continuous maps using approximate jacobians', Applied Mathematics Research Report AMR 97/2, (University of New South Wales, Australia, 1997), J. Convex Analysis, to appear.

[10] V. Jeyakumar and X. Q. Yang, 'Approximate generalized Hessians and Taylor's expansions for continuously Gâteaux differentiable functions', Applied Mathematics Research Report AMR96/20, (University of New South Wales, Australia, 1996), Nonlinear Analysis T, M & A, to appear.

[11] B. S. Mordukhovich, 'Metric approximations and necessary optimality conditions for general classes of nonsmooth extremal problems', *Soviet. Math. Dokl.* **22** (1980) 526–530.

[12] B. S. Mordukhovich, 'Generalized differential calculus for nonsmooth and set-valued mappings', *J. Math. Anal. Appl.* **183** (1994) 250–288.

[13] B. S. Mordukhovich and Y. Shao, 'On nonconvex subdifferential calculus in banach spaces', *J. Convex. Anal.* **2** (1995) 211–228.

[14] R. T. Rockafellar and J. B. Wets, 'Variational analysis', (Springer Verlag, 1997).

[15] M. Studniarski and V. Jeyakumar, 'A generalized mean-value theorem and optimality conditions in composite nonsmooth minimization', *Nonlinear Anal. T.M.& A* **24**(6) (1995) 883–894.

[16] X. Q. Yang, 'Second-order conditions for $C^{1,1}$ optimization with applications', *Numer. Funct. Anal. Optim.* **14** (1993) 621–632.

[17] X. Q. Yang and V. Jeyakumar, 'Generalized second-order directional derivatives and optimization with $C^{1,1}$ functions', *Optimization* **26** (1992) 165–185.