JOURNAL OF MATHEMATICAL ANALYSIS AND APPLICATIONS **187**, 118–133 (1994)

An Exterior Point Method for Computing Points That Satisfy Second-Order Necessary Conditions for a $C^{1,1}$ Optimization Problem

X. Q. YANG

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

Submitted by E. Stanley Lee

Received May 17, 1993

We obtain a second-order generalized chain rule for composite $C^{1,1}$ functions using a generalized Hessian matrix. We introduce a unified exterior point penalty method for a $C^{1,1}$ constrained minimization problem and derive second-order necessary conditions for exterior point penalty problem using the established generalized chain rule. We then show that any limiting point of the sequences obtained by the exterior point method satisfies the second-order necessary conditions.

1. INTRODUCTION

It is well known that second-order optimality conditions have many applications in sensitivity analysis and convergence theory. Second-order conditions with twice differentiability conditions have been well established. Recently the class of differentiable functions with locally Lipschitz gradients, called $C^{1,1}$ in the literature, has received a great deal of attention, see [6, 9–11, 13–15] and the references therein. This kind of functions appears, e.g., in penalty function, Augmented Lagrangian, and proximal point methods. Various generalized second-order directional derivatives and corresponding Hessians for $C^{1,1}$ functions were introduced and studied, and calculus rules were obtained. Calculus rules, such as chain rules and Taylor expansions, play a crucial role in obtaining second-order conditions, see [6, 10, 13, 14].

The main purpose of this paper is to study a unified exterior point penalty method for a $C^{1,1}$ constrained minimization problem and to derive

0022-247X/94 $6.00

second-order necessary optimality conditions for exterior penalty problem. These conditions are obtained by establishing a generalized chain rule for composite $C^{1,1}$ functions. We further show that any limiting point of the sequences obtained by the exterior point penalty method satisfies the second-order necessary conditions. Our methods use reduced differentiability assumptions and employ recently established calculus rules.

The outline of the paper is as follows. In Section 2, we present some basic property of the generalized Hessian of Hiriart-Urruty *et al.* [10]. In Section 3, we obtain a generalized chain rule for composite $C^{1,1}$ functions. As an application, we show how twice differentiability of max functions can be characterized. In Section 4, we first present a general convergence result which shows that any limiting point of certain sequences satisfies second-order necessary conditions for a $C^{1,1}$ nonlinear programming problem. We then compute such sequences by unified exterior point methods. Full details of the proof for the general convergence theorem will be given in Section 5.

2. A Generalized Hessian Matrix

We recall some basic properties of generalized Hessian and generalized second-order directional derivative of $C^{1,1}$ functions presented in Hiriart-Urruty [9] and Hiriart-Urruty *et al.* [10]. We denote the inner product of two vectors u, v in an n-dimensional space by $\mathbb{R}^n$ by $\langle u, v\rangle$. For a real valued function f: $\mathbb{R}^n \to \mathbb{R}$, ∇f, $\nabla^2 f$ denote the gradient and Hessian of f. For a vector function g: $\mathbb{R}^n \to \mathbb{R}^p$, $Jg(x)$ denotes the Jacobian of g at x.

Hiriart-Urruty *et al.* [10] introduced a generalized Hessian matrix for a $C^{1,1}$ function f at x, denoted by $\partial^2 f(x)$, as the convex hull of the set

$$\{M : \exists\ x_i \to x,\ \nabla^2 f(x_i) \text{ exists and } \nabla^2 f(x_i) \to M\}.$$

Since f is Fréchet differentiable, $\nabla^2 f(x_i)$ is a symmetric matrix whenever $\nabla^2 f(x_i)$ exists, thus $\partial^2 f(x)$ is the set of symmetric matrices which is a nonempty convex compact set of the space of $n \times n$ matrices. The support bi-function of the set $\partial^2 f(x)$ is denoted by $f^{\infty}(x; u, v)$, i.e.,

$$f^{\infty}(x: u, v) = \max\{\langle Mu, v\rangle: \langle M \in \partial^2 f(x)\}. \tag{2.1}$$

A characterization of $f^{\infty}(x; u, v)$ using gradients was given in Hiriart-Urruty [9]:

$$f^{\infty}(x; u, v) = \limsup_{\substack{y \to x \\ t \downarrow 0}} \frac{\langle \nabla f(y + tu), v\rangle - \langle \nabla f(y), v\rangle}{t}. \tag{2.2}$$

It is easy to see that $f^{\infty}(x; u, v) = (\langle\nabla f(\cdot), v\rangle)^{\circ}(x; u)$, where $g^{\circ}(x; u)$ is the Clarke generalized directional derivative of g at x in the direction u (see Clarke [5]). Furthermore it has been shown in Cominetti and Correa [6] that the following relation holds:

$$f^{\infty}(x; u, v) = \limsup_{\substack{y \to x \\ s,t \downarrow 0}} \frac{f(y + su + tv) - f(y + su) - f(y + tv) + f(y)}{st}.$$

Recall that f: $\mathbb{R}^n \to \mathbb{R}$ is twice *strictly* differentiable at x (see, e.g., Cominetti and Correa [6]) if there exists a linear operator $D^2f(x)$: $\mathbb{R}^n \to \mathbb{R}^n$ such that

$$\limsup_{\substack{y \to x \\ s,t \downarrow 0}} \frac{f(y + su + tv) - f(y + su) - f(y + tv) + f(y)}{st} = \langle D^2f(x)u, v\rangle.$$

Then it is easy to see that f is twice strictly differentiable if and only if ∇f is strictly differentiable. This is equivalent to the condition that $\partial^2 f(x) = \{\nabla^2 f(x)\}$.

Now we state some basic properties of generalized Hessian $\partial^2 f(x)$ which will be used in the sequel. For details, see Hiriart-Urruty [9] and Hiriart-Urruty *et al.* [10].

1. For $C^{1,1}$ functions f, g: $\mathbb{R}^n \to \mathbb{R}$ and r, $q \neq 0$, we have

$$\partial^2(rf + qg)(x) \subseteq r\,\partial^2 f(x) + q\,\partial^2 g(x), \tag{2.3}$$

and if either f or g twice strictly differentiable, then (2.3) holds with the equality;

2. The set-valued mapping $x \to \partial^2 f(x)$ is locally bounded at x; i.e., there exist a neighborhood V of x and a constant K such that

$$\sup\{\|M\| : M \in \partial^2 f(x), x \in V\} \leq K;$$

3. The set-valued mapping $x \to \partial^2 f(x)$ is upper semi-continuous; i.e., if $x_n \to x$, $M_n \to M$, with $M_n \in \partial^2 f(x_n)$, then $M \in \partial^2 f(x)$.

3. Chain Rules with Generalized Hessians

In this section, we obtain a generalized chain rule in terms of the generalized Hessian matrix of Hiriart-Urruty *et al.* [10]. We then apply the established chain rule to derive the characterizations of the generalized Hessian of max functions. Chain rules for compositions of a $C^{1,1}$ function and a twice differentiable function were obtained by Cominetti and Correa

[6] and Yang and Jeyakumar [13]. Generalized chain rules for composite $C^{1,1}$ functions were given by Hiriart-Urruty *et al.* [10] and Yang [14]. However, the one given in [10] is not valid as we will show by an example, Here, we obtain a generalized chain rule for composite $C^{1,1}$ functions using the generalized Hessian of Hiriart-Urruty *et al.* [10].

The following generalized chain rule for composite $C^{1,1}$ functions,

$$(f \circ g)^{\infty}(x; u, v) \leq \sum_{i=1}^{m} \frac{\partial f(g(x))}{\partial y_i} g_i^{\infty}(x; u, v) + f^{\infty}(g(x); \nabla g(x)u, \nabla g(x)v), \tag{3.1}$$

was given in [10], where f: $\mathbb{R}^m \to \mathbb{R}$, $g = (g_1, \ldots, g_m)$: $\mathbb{R}^n \to \mathbb{R}^m$ are $C^{1,1}$ (see Theorem 2.2 in [10]). The following numerical example shows that (3.1) is not valid.

EXAMPLE 3.1. Let

$$f(y_1, y_2) = y_1 - y_2,$$

$$g(x) = \left(x^2, \int_0^x t^2 \sin \frac{1}{t}\, dt\right).$$

Then $(f \circ g)(x) = x^2 - \int_0^x t^2 \sin 1/t \, dt$. Simple calculations give us that

$$(f \circ g)^{\infty}(0; u, v) = 2uv + |uv|,$$

$$f^{\infty}(g(0); \nabla g(0)u, \nabla g(0)v) + \sum_{i=1}^{2} \frac{\partial f(g(0))}{\partial y_i} g_i^{\infty}(0; u, v) = 2uv - |uv|.$$

It follows from (3.1) that

$$2uv + |uv| \leq 2uv - |uv|,$$

but this is not true, e.g., for $u = v = 1$.

Now we derive a correct formulation of the generalized chain rule for composite $C^{1,1}$ functions and provide full details of the proof.

THEOREM 3.1. *Let $F = f \circ g$, where f: $\mathbb{R}^m \to \mathbb{R}$, $g = (g_1, \ldots, g_m) : \mathbb{R}^n \to \mathbb{R}^m$. If f, $g_1, \ldots, g_m$ are $C^{1,1}$ functions, then for each $x, u, v \in \mathbb{R}^n$, we have*

$$F^{\infty}(x; u, v) \leq f^{\infty}(g(x); \nabla g(x)u, \nabla g(x)v) + h_x(u, v), \tag{3.2}$$

where $h_x(u, v) = \max\{\sum_{i=1}^m \partial f(g(x))/\partial y_i \langle M_i u, v\rangle : M_i \in \partial^2 g_i(x)\}$; equivalently, we have

$$\partial^2 F(x)u \subseteq Jg(x)^T\,\partial^2 f(g(x))Jg(x)u + \sum_{i=1}^{m} \frac{\partial f(g(x))}{\partial y_i}\,\partial^2 g_i(x)u. \tag{3.3}$$

Furthermore if g is open at x and twice differentiable, then (3.2) *and* (3.3) *hold with equalities.*

Proof. We compute $F^{\infty}(x; u, v)$. Since f, g are Fréchet differentiable, we get

$$\nabla F(x) = Jg(x)^T \nabla f(g(x)) = \sum_{i=1}^{m} \frac{\partial f(g(x))}{\partial y_i} \nabla g_i(x).$$

Then

$$\begin{aligned} &F^{\infty}(x; u, v)\\ &\quad= \limsup_{\substack{z\to x\\ t\downarrow 0}} \frac{\langle Jg(z+tu)^T\nabla f(g(z+tu)), v\rangle - \langle Jg(z)^T\nabla f(g(z)), v\rangle}{t}\\ &\quad= \limsup_{\substack{z\to x\\ t\downarrow 0}} \frac{\langle Jg(z+tu)^T\nabla f(g(z+tu)), v\rangle - \langle Jg(z+tu)^T\nabla f(g(z)), v\rangle}{t}\\ &\qquad+ \limsup_{\substack{z\to x\\ t\downarrow 0}} \frac{\langle Jg(z+tu)^T\nabla f(g(z)), v\rangle - \langle Jg(z)^T\nabla f(g(z)), v\rangle}{t}\\ &:= \alpha + \beta. \end{aligned} \tag{3.4}$$

Since

$$\lim_{\substack{z\to x\\ t\downarrow 0}} \frac{g(z+tu) - g(z) - tJg(z)u}{t} = 0,$$

we get

$$\begin{aligned} \alpha &= \limsup_{\substack{z\to x\\ t\downarrow 0}} \frac{\langle \nabla f(g(z+tu)) - \nabla f(g(z)), Jg(z+tu)v\rangle}{t}\\ &= \limsup_{\substack{z\to x\\ t\downarrow 0}} \frac{\langle \nabla f(g(z)+tJg(z)u) - \nabla f(g(z)), Jg(z+tu)v\rangle}{t}\\ &\le \limsup_{\substack{z_1\to g(x)\\ t\downarrow 0}} \frac{\langle \nabla f(z_1+tJg(x)u) - \nabla f(z_1), Jg(x)v\rangle}{t}\\ &= f^{\infty}(g(x), Jg(x)u, Jg(x)v) \end{aligned} \tag{3.5}$$

(equality holds if g is open at x). Now we prove that

$$\beta \leq \max\left\{\sum_{i=1}^{m} \frac{\partial f(g(x))}{\partial y_i} \langle M_i u, v\rangle : M_i \in \partial^2 g_i(x)\right\} = h_x(u, v). \quad (3.6)$$

Since f, g are $C^{1,1}$, we see that

$$\beta = \limsup_{\substack{z\to x\\ t\downarrow 0}} \frac{\langle (Jg(z+tu) - Jg(z))^T \nabla f(g(z)), v\rangle}{t}$$

$$= \limsup_{\substack{z\to x\\ t\downarrow 0}} \frac{\langle (Jg(z+tu) - Jg(z))^T \nabla f(g(x)), v\rangle}{t}$$

$$= \limsup_{\substack{z\to x\\ t\downarrow 0}} \frac{1}{t} \sum_{i=1}^{m} \frac{\partial f(g(x))}{\partial y_i} \langle \nabla g_i(z+tu) - \nabla g_i(z), v\rangle.$$

We claim that for any $\varepsilon > 0$, there exist $M_i \in \partial^2 g_i(x)$ and $\delta > 0$ such that for any $t \in (0,\delta]$, $\|z - x\| < \delta$ we have

$$\left|\frac{1}{t}\langle \nabla g_i(z+tu) - \nabla g_i(z), v\rangle - \langle M_i u, v\rangle\right| < \varepsilon. \quad (3.7)$$

To see this, we have

$$\max\{\langle M_i u, v\rangle : M_i \in \partial^2 g_i(x)\} = g_i^{\infty}(x; u, v)$$
$$= \limsup_{\substack{z\to x\\ t\downarrow 0}} \frac{\langle \nabla g_i(z+tu) - \nabla g_i(z), v\rangle}{t}.$$

On the other hand, it follows from (2.2) that

$$\min\{\langle M_i u, v\rangle : M_i \in \partial^2 g_i(x)\}$$
$$= -\max\{\langle M_i(-u), v\rangle : M_i \in \partial^2 g_i(x)\}$$
$$= -g_i^{\infty}(x; -u, v)$$
$$= -\limsup_{\substack{z\to x\\ t\downarrow 0}} \frac{\langle \nabla g_i(z + t(-u)) - \nabla g_i(z), v\rangle}{t}$$
$$= \liminf_{\substack{z\to x\\ t\downarrow 0}} \frac{\langle \nabla g_i(z - tu + tu) - \nabla g_i(z - tu), v\rangle}{t}$$
$$= \liminf_{\substack{z'\to x\\ t\downarrow 0}} \frac{\langle \nabla g_i(z' + tu) - \nabla g_i(z'), v\rangle}{t}.$$

Note that $\partial^2 g_i(x)$ is a compact convex set. Hence (3.7) holds. So we obtain

$$\left|\frac{1}{t}\sum_{i=1}^{m}\frac{\partial f(g(x))}{\partial y_i}\langle \nabla g_i(z+tu)-\nabla g_i(z), v\rangle - \sum_{i=1}^{m}\frac{\partial f(g(x))}{\partial y_i}\langle M_i u, v\rangle\right| \leq m\varepsilon L_f,$$

where $|\partial f(g(x))/\partial y_i| \leq L_f$. Then (3.6) holds. Hence

$$F^{\infty}(x; u, v) \leq f^{\infty}(g(x); \nabla g(x)u, \nabla g(x)v) + h_x(u, v).$$

By the Hahn–Banach Theorem, we get (3.3) from (3.2). Furthermore if g is open and twice differentiable, then (3.5)–(3.7) hold with equalities. Hence (3.2) holds with equality. ∎

EXAMPLE 3.2. Consider the same example as in Example 3.1. Let $f(y_1, y_2) = y_1 - y_2$, $g(x) = (x^2, \int_0^x t^2 \sin(1/t)dt)$. Then

$$\begin{aligned} &f^{\infty}(g(x); \nabla g(x)u, \nabla g(x)v) + h_x(u, v) \\ &\quad = 0 + \max\{M_1uv - M_2uv : M_1 \in \partial^2 g_1(0), M_2 \in \partial^2 g_2(0)\} \\ &\quad = \max\{2uv - M_2uv : M_2 \in [-1, 1]\} \\ &\quad = 2uv + |uv|. \end{aligned}$$

Hence Theorem 3.1 is verified.

We note that a generalized chain rule for composite $C^{1,1}$ functions using $\partial^{\infty}f(x)(u)$ in infinite dimensional spaces was obtained by Yang [14]. However, Theorem 3.1 cannot be derived from the chain rule given in [14] even if we assume the upper semi-continuity of the mapping $x \to f^{\infty}(x; u, v)$ since some $C^{1,1}$ functions may not have this property. As applications of the generalized chain rule, we obtain the characterizations of the generalized Hessian $\partial^2 F(x)$ and twice strictly differentiability for the function of the form

$$F(x) = [\max\{g(x), 0\}]^p,$$

where $p \geq 2$ and g is $C^{1,1}$. We point out that a special case when $p = 2$ using generalized Hessian $\partial^{\infty}f(x)(u)$ was given by Jeyakumar and Yang [11], where a direct proof was provided. Under C^2 assumptions the characterizations of the generalized Hessians for the function of this type with $p = 2$ were given by Hiriart-Urruty *et al.* [10] and Yang and Jeyakumar [13] (see also [3]). It should be noted that this kind of function appears in augmented Lagrangian methods and penalty function methods.

PROPOSITION 3.1. *Let g: $\mathbb{R}^n \to \mathbb{R}$ be $C^{1,1}$. Then for every $u \in \mathbb{R}^n$*

$$\partial^2 F(x)u \subseteq \begin{cases} pg^{p-1}(x)\partial^2 g(x)u + p(p-1)g^{p-2}(x)\langle \nabla g(x), u\rangle \nabla g(x), & \textit{if } g(x) > 0; \\ 0 & \textit{if } g(x) < 0; \\ p(p-1)g_1^{p-2}(x)\{\beta\langle \nabla g_1(x), u\rangle \nabla g_1(x) : \beta \in [0,1]\} & \textit{if } g(x) = 0. \end{cases}$$

Proof. Let $f(y) = [\max\{y, 0\}]^p$. Then the conclusion follows from Theorem 2.1. ∎

This result will be used to derive extended exterior point methods in Section 4. We now obtain the characterizations of twice strictly differentiability of the max function.

PROPOSITION 3.2. *Let x be a point satisfying $g(x) = 0$ and let $g(x)$ be twice strictly differentiable at x. Then $F(x)$ is twice strictly differentiable at x if and only if $\nabla g(x) = 0$. Moreover, when $g(x) = 0$, $\nabla g(x) = 0$, we have*

$$D^2F(x)u = 0.$$

Proof. It follows from Proposition 3.1 that $\partial^2 F(x)(u)$ is single-valued for all $u \in X$ if and only if $\nabla g(x) = 0$. Then the function F is twice strictly differentiable at x if and only if $\nabla g(x) = 0$.

4. CONVERGENCE ANALYSIS AND EXTERIOR POINT METHODS

Consider the following constrained minimization problem

$$\begin{aligned} \text{(P)}\quad & \min f(x) \\ & \text{s.t. } g_i(x) \le 0, \qquad i = 1, \dots, m, \\ & \qquad h_j(x) = 0, \qquad j = 1, \dots, l, \end{aligned}$$

where f, g_i, h_j: $\mathbb{R}^n \to \mathbb{R}$, $i = 1, \dots, m, j = 1, \dots, l$ are $C^{1,1}$ functions. In this section, we present exterior point methods in a unified way that allows one to compute points which satisfy second-order necessary conditions of (P). Recall that arc methods and penalty methods were given by McCormick [12] and Auslender [3] for inequality constrained minimization problems under C^2 requirements. Recently a method that combines curvilinear paths and trust regions was given by Bulteau and Vial [4] for the unconstrained optimization problem.

Let $C = \{x \in \mathbb{R}^n : g_i(x) \le 0, i = 1, \dots, m, h_j(x) = 1, \dots, l\}$ be the feasible set of (P) and $I(x) = \{i : g_i(x) = 0\}$. Assume that $a \in C$ is a local minimum of (P). A first-order constraint qualification is a condition that guarantees the existence of $\lambda \in \mathbb{R}^m_+$, $\mu \in \mathbb{R}^l$ satisfying

$$\nabla L(a, \lambda, \mu) = \nabla f(a) + \sum_{i=1}^{m} \lambda_i \nabla g_i(a) + \sum_{j=1}^{l} \mu_j \nabla h_j(a) = 0, \tag{4.1}$$

$$\lambda_i g_i(a) = 0, \qquad i = 1, \ldots, m, \tag{4.2}$$

where L is the Lagrangian function of (P) defined by

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

Let

$$\Phi = \{u \in \mathbb{R}^n : \langle \nabla g_i(a), u\rangle = 0, \forall i \in I(a), \langle \nabla h_j(x), u\rangle = 0, \forall j\}.$$

We say that a second-order necessary condition of (P) holds at a if for every $u \in \Phi$, there exist $A \in \partial^2 f(a)$, $B_i \in \partial^2 g_i(a)$, $C_j \in \partial^2 h_j(a)$, $i \in I(a)$, $j = 1, \ldots, l$ such that

$$\left\langle \left(A + \sum_{i \in I(a)} \lambda_i B_i + \sum_{j=1}^{l} \mu_j C_j\right) u, u \right\rangle \geq 0. \tag{4.3}$$

The following nonsmooth version of McCormick's second-order constraint qualification is sufficient for (4.3) to hold: for any $u \in \Phi$, there exists a C^2 arc $\alpha(\theta)$ (where $\theta \geq 0$) along which $g_i(\alpha(\theta)) = 0$ for all $i \in I(a)$, $h_j(\alpha(\theta)) = 0, j = 1, \ldots, l$ for small θ, and $\alpha(0) = a$ and $\alpha'(0) = u$, see [15].

To obtain convergence analysis, we will use the *linearly independent condition;* i.e., the set $\{\nabla g_i(a), i \in I(a), \nabla h_j(a), j = 1, \ldots, l\}$ is linearly independent. If the functions are C^2, then the linearly independent condition implies the above second-order constraint qualification, see [8].

We now present the following general convergence theorem which is a (second-order) nonsmooth extension of a result of Auslender [3] for (P) in which the functions are $C^{1,1}$. For the details of the proof, see Appendix.

THEOREM 4.1. *Consider the problem* (P). *Assume that the linearly independent condition holds for any* $x \in C$. *Let* $\{x_n\}$ *be a sequence which converges to* $a \in C$ *and* $\{\lambda_i^n\}$, $\{\mu_j^n\}$ *be real sequences such that*

$$\lambda_i^n \geq 0, \; i = 1, \ldots, m; \; \lim_{n \to \infty} \lambda_i^n = 0, \qquad \forall i \notin I(a), \tag{4.4}$$

Let $\{\varphi_i^n\}_{n\in N}(i \in I(a))$, $\{\psi_j^n\}_{n\in N}$ *be real-valued functions defined on* $\mathbb{R}$ *such that*

$$\varphi_i^n(0) = 0, \qquad \forall i \in I(a),\ n \in N,\ \psi_j^n(0) = 0,\ j = 1, \ldots, l,\ n \in N. \tag{4.5}$$

If for every n there exists $A_n \in \partial^2 f(x_n)$, $B_n^i \in \partial^2 g_i(x_n)$, $C_n^j \in \partial^2 h_j(x_n)$, $i \in I(a), j = 1, \ldots, l$ *such that* x_n *satisfies*

$$\nabla f(x_n) + \sum_{i=1}^{m} \lambda_i^n \nabla g_i(x_n) + \sum_{j=1}^{l} \mu_j^n \nabla h_j(x_n) = 0, \tag{4.6}$$

$$\left\langle \left(A_n + \sum_{i=1}^{m} \lambda_i^n B_n^i + \sum_{j=1}^{l} \mu_j^n C_n^j \right) u, u \right\rangle + \sum_{i\in I(a)} \varphi_i^n(\langle \nabla g_i(x_n), u\rangle)$$
$$+ \sum_{j=1}^{l} \psi_j^n(\langle \nabla h_j(x_n), u\rangle) \geq 0, \qquad \forall u \in \mathbb{R}^n, \tag{4.7}$$

then a satisfies the first-order and the second-order necessary conditions.

We now construct methods to calculate the sequence $\{x_n\}$, which satisfies the conditions in Theorem 4.1. For a given $C^{1,1}$ function f, we will assume that if f is inf-compact, i.e., for every λ, the set $\{x : f(x) \leq \lambda\}$ is compact, then we can use a descent method to find a (local) minimum of f (see Auslender [2]).

Let $z \in C$ and let

$$\rho_n \leq \rho_{n+1},\ n \in N,\ \lim_{n\to\infty} \rho_n = +\infty,$$

$$\gamma_n \leq \gamma_{n+1},\ n \in N,\ \lim_{n\to\infty} \gamma_n = +\infty.$$

The *extended exterior point method* is to minimize the penalty function

$$\phi_n(x) = f(x) + \rho_n \sum_{i=1}^{m} [\max\{g_i(x), 0\}]^p + \gamma_n \sum_{j=1}^{l} |h_j(x)|^p, \qquad (p \geq 2).$$

It is worth noting that for the case $p = 2$, even if g_i is twice differentiable, the function $[\max\{g_i, 0\}]^2$ may not be twice differentiable, but it is $C^{1,1}$. Therefore we will use the generalized calculus rules given in Section 2.

It should also be noted that if $\rho_n = \gamma_n$, then we have $\phi_n(x) = (\phi_n^0 \circ F)(x)$, where $F(x) = (f(x), g_1(x), \ldots, g_m(x), h_1(x), \ldots, h_l(x))$ and ϕ_n^0: $\mathbb{R}^{m+l+1} \to \mathbb{R}$ is the inf-convolution of the functions ϕ_n^1 and $\|.\|_p^p$ with index ρ_n, i.e.,

$$\phi_n^0(z) = \inf\left\{\phi_n^1(y) + \frac{1}{\rho_n}\|y - z\|_p^p : y \in \mathbb{R}^{m+l+1}\right\}, \tag{4.8}$$

where

$$\phi_n^1(y_1, y_2) = \begin{cases} y_1, & \text{if } y_2 \in \mathbb{R}^m_- \times \{0\}_{\mathbb{R}^l}, \\ +\infty, & \text{otherwise}, \end{cases}$$

and $\|x\|_p = (\sum_{i=1}^m |x_i|^p)^{1/p}$. We note that (4.8) is a special case of the extended inf-convolution introduced by Attouch and Wets [1]

$$G_\xi(z) = \inf\{f(z) + \xi(z - x) : x \in \mathbb{R}^n\}, \tag{4.9}$$

where $\xi : \mathbb{R}^n \to \mathbb{R}$ is a strictly convex coercive function satisfying $\xi(0) = 0$. In fact, we can let $\xi_n(x) = (1/\rho_n)\|x\|_p^p$. For the case $p = 2$, (4.8) is nothing other than the Moreau–Yosida approximation of the ordinary penalty function ϕ_n^1 and $(1/\rho_n)\|.\|_p^p$, which is again a special case of (4.9).

Let x_n be a local minimum of the function $\phi_n(x)$. Then it follows from Hiriart-Urruty *et al.* [10, Theorem 3.1] that $\nabla\phi_n(x_n) = 0$ and for each $u \in \mathbb{R}^n$, $\exists M \in \partial^2\phi_n(x_n)$ such that $\langle Mu, u\rangle \geq 0$; that is

$$\nabla f(x_n) + p\rho_n \sum_{i=1}^m [\max\{g_i(x), 0\}]^{p-1}\nabla g_i(x) + p\gamma_n \sum_{j=1}^l |h_j(x_n)|^{p-1}\nabla h_j(x_n) = 0, \tag{4.10}$$

and for each $u \in \mathbb{R}^n$, there exist $A_n \in \partial^2 f(x_n)$, $B_n^i \in \partial^2 g_i(x_n)$, $C_n^j \in \partial^2 h_j(x_n)$ such that

$$\begin{aligned} &\left\langle\left(A_n + p\rho_n \sum_{i\in I^+(x_n)} g_i^{p-1}(x_n)B_n^i + p\gamma_n \sum_{j=1}^l |h_j(x_n)|^{p-1}C_n^j\right) u, u\right\rangle \\ &\quad + p(p-1)\rho_n \sum_{i\in I(x_n)} g_i^{p-2}(x_n)\langle\nabla g_i(x_n), u\rangle^2 \\ &\quad + p(p-1)\rho_n \sum_{i\in I^+(x_n)} g_i^{p-2}(x_n)\langle\nabla g_i(x_n), u\rangle^2 \\ &\quad + p(p-1)\gamma_n \sum_{j=1}^l |h_j(x_n)|^{p-2}\langle\nabla h_j(x_n), u\rangle^2 \geq 0, \end{aligned} \tag{4.11}$$

where $I^+(x) = \{i : g_i(x) > 0\}$. We see that $0^0 = 1$ and (4.11) is obtained by using Proposition 3.1, which is derived from the generalized chan rule (3.2).

Note that when $p = 2$ a second-order necessary condition of a local minimum for the function $\phi_n(x)$ was given by Auslender [3], where the term $2\rho_n \sum_{i\in I(x_n)} \langle \nabla g_i(x_n), u\rangle^2$ in (4.11) is replaced by $2\rho_n \sum_{i\in I(x_n)} [\max\{\langle \nabla g_i(x_n), u\rangle, 0\}]^2$, thus Auslender's condition is tighter than (4.11), and a direct proof without using the chain rule was also given.

Now we obtain a result which allows us to compute the points that satisfy second-order necessary conditions. It should be noted that when $p = 2$, and without equality constraints, our results of the twice differentiable case imply the ones in [3] and that when $p = 3$, and without equality constraints, our results of the twice differentiable case coincide with the ones in [3].

THEOREM 4.2. *Assume that the linearly independent condition holds for any $x \in C$ and ϕ_0 is* inf-*compact. Let $\{x_n\}$ be a sequence satisfying the above necessary conditions* (4.10) *and* (4.11). *Then $\{x_n\}$ is bounded and every limiting point of the sequence $\{x_n\}$ is a point of C, which satisfies the first-order and the second-order necessary conditions.*

Proof. Since $\phi_0(x) \leq \phi_n(x)$, we get

$$\{x : \phi_n(x) \leq \lambda\} \subseteq \{x : \phi_0(x) \leq \lambda\}.$$

Now by the inf-compactness hypothesis, we see that the set $\{x : \phi_n(x) \leq \phi_n(z)\}$ is compact. For n fixed, we find x_n by a descent method such that

$$\phi_n(x_n) \leq \phi_n(z).$$

Thus

$$\phi_0(x_n) \leq \phi_n(x_n) \leq \phi_n(z) = f(z), \qquad \forall n.$$

Since $\{x : \phi_0(x) \leq f(z)\}$ is compact, the sequence $\{x_n\}$ is bounded. Let x^* be a limiting point of $\{x_n\}$. Without loss of generality, we assume that $x_n \to x^*$. Next we show that $x^* \in C$. Since $\phi_n(x_n) \leq f(z)$,

$$f(x_n) + \rho_n \sum_{i=1}^{m} [\max\{g_i(x_n), 0\}]^p + \gamma_n \sum_{j=1}^{l} |h_j(x_n)|^p \leq f(z).$$

Taking limit and noting that $x_n \to x^*$, we get

$$\sum_{i=1}^{m} [\max\{g_i(x^*), 0\}]^p \leq \lim_{n\to\infty} \frac{f(z) - f(x_n)}{\rho_n} = 0,$$

i.e.,

$$g_i(x^*) \le 0, \qquad i = 1, \ldots, m.$$

By using the same argument, we get $|h_j(x^*)|^p \le 0, j = 1, \ldots, l$, i.e., $h_j(x^*) = 0, j = 1, \ldots, l$. Thus $x^* \in C$. Since $x_n \to x^* \in C$, there exists $n_0 > 0$ such that

$$I(x_n) \subset I(x^*),\ I^+(x_n) \subset I(x^*), \qquad \forall n > n_0.$$

By (4.10) and (4.11), let

$$\lambda_i^n = p\rho_n[\max\{g_i(x_n), 0\}]^{p-1},$$

$$\varphi_i^n(t) = \begin{cases} p(p-1)\rho_n g_i^{p-2}(x_n)t^2, & \text{if } i \in I(x_n) \cup I^+(x_n) \\ 0, & \text{if } i \notin I(x_n) \cup I^+(x_n) \end{cases}$$

$$\mu_j^n = p\gamma_n|h_j(x_n)|^{p-1}, \qquad j = 1, \ldots, l,$$

$$\psi_j^n = p(p-1)\gamma_n|h_j(x_n)|^{p-2}t^2, \qquad j = 1, \ldots, l.$$

So the conditions in Theorem 4.1 are easily verified. Then x^* satisfies the first-order and the second-order necessary conditions by Theorem 4.1. ∎

5. Appendix: Proof of the Convergence Theorem

In this Appendix, we give the proof of Theorem 4.1. In our proof, we use the locally boundedness and upper semi-continuity of the generalized Hessian matrix $\partial^2 f(x)$ and the linearly independent condition.

Let $\{A_n\}_{n\in N}$ be a sequence of subsets of $\mathbb{R}^n$. The super limit, $\lim_{n\to\infty}\sup A_n$, and lower limit, $\lim_{n\to\infty}\inf A_n$ are defined by

$$\limsup_{n\to\infty} A_n = \{x \in \mathbb{R}^n : x = \lim_{i\to\infty} x_{n_i} \text{ with } x_{n_i} \in A_{n_i},\ n_i \in N_i \subset N\}$$

$$\liminf_{n\to\infty} A_n = \{x \in \mathbb{R}^n : x = \lim_{n\to\infty} x_n \text{ with } x_n \in A_n,\ n > n_0\}.$$

If $\lim_{n\to\infty}\sup A_n = \lim_{n\to\infty}\inf A_n$, we say that the limit $\lim_{n\to\infty} A_n$ exists and $\lim_{n\to\infty} A_n = \lim_{n\to\infty}\sup A_n = \lim_{n\to\infty}\inf A_n$. It is easy to see that $\lim_{n\to\infty} A_n = A$ is equivalent to that A is closed and $d(.,A^n) \xrightarrow{pw} d(.,A)$, where $\xrightarrow{pw}$ means pointwise convergence.

Let f be an affine function from $\mathbb{R}^n \to \mathbb{R}^p$. The rank of f, denoted by rank(f), is the rank of the (unique) $p \times n$ matrix A such that $f(x) = Ax +$

b. The following results were established in Dantzig *et al.* [7] about the stability of the feasible sets and the closedness of the optimal solution set. These will be used in the proof of Theorem 4.1.

LEMMA 5.1 [Corollary II.3.4 [7]). *Let* χ^n, χ *be affine functions from* $\mathbb{R}^n$ *to* $\mathbb{R}^q$ *with* $\chi^n \xrightarrow{pw} \chi$. *Let* $H(\chi) = \{x \in \mathbb{R}^n : \chi(x) = 0\}$. *Suppose that*

$$\limsup_{n\to\infty} \operatorname{rank}(\chi^n) \leq \operatorname{rank}(\chi). \tag{5.1}$$

Then either $H(\chi_n)$ *is empty or* $\lim_{n\to\infty} H(\chi_n) = H(\chi)$.

Let

$$v(x) = \inf\{\eta(x, y) : y \in \Omega(x)\},$$

where η is a real-valued function defined on $\mathbb{R}^n \times \mathbb{R}^p$ and Ω is a point-to-map from $\mathbb{R}^n$ to $\mathbb{R}^p$, and let

$$G(x) = \{y \in \Omega(x) : v(x) = \eta(x, y)\}.$$

Then we have the following.

LEMMA 5.2 [Theorem I.2.2. [7]). *Assume that* Ω *is a point-to-set mapping from* $\mathbb{R}^n$ *to* $\mathbb{R}^p$ *and that* η *is continuous. If for every sequence* $\{x_n\}$ *with* $x_n \to x$, $\liminf_{n\to\infty} \Omega(x_n)$ *is either empty or equal to* $\Omega(x)$, *then G is upper semi-continuous at x.*

Proof of Theorem 4.1. (i) To prove the first-order necessary condition, let us first show that $\{\lambda_i^n\}$, $\{\mu_j^n\}$, $i \in I(a)$, $j = 1, \ldots, l$ are bounded sequences by contradiction. Assume that, for example, $\{\mu_j^n\}$ is an unbounded sequence for at least one j, but $\{\lambda_i^n\}$ are bounded sequences for all $i \in I(a)$. Let $\mu_n = \sum_{j=1}^{l} \mu_j^n$. Dividing (4.6) by μ_n

$$\frac{1}{\mu_n} \nabla f(x_n) + \sum_{i=1}^{m} \frac{\lambda_i^n}{\mu_n} \nabla g_i(x_n) + \sum_{j=1}^{l} \frac{\mu_j^n}{\mu_n} \nabla h_j(x_n) = 0. \tag{5.2}$$

If we let $n \to \infty$ in (5.2), then there exist w_j, $j = 1, \ldots, l$ such that

$$\sum_{j=1}^{l} w_j = 1, \qquad \text{and} \qquad \sum_{j=1}^{l} w_j \nabla h_j(a) = 0,$$

thus $\nabla h_j(a), j = 1, \ldots, l$ are linearly dependent. This is a contradiction with the linearly independent condition. Therefore the sequences $\{\lambda_i^n\}$, $\{\mu_j^n\}$, $i \in I(a)$, $j = 1, \ldots, l$ are bounded.

Hence these sequences have converging subsequences. Without loss of generality, we assume that there exists λ_i, μ_j, $i \in I(a)$, $j = 1, \ldots, l$ such that $\lambda_i^n \to \lambda_i$, $\mu_j^n \to \mu_j$. By taking the limit in (4.6) as $n \to \infty$ and noting (4.4), we see that

$$\nabla f(a) + \sum_{i \in I(a)} \lambda_i \nabla g_i(a) + \sum_{j=1}^{l} \mu_j \nabla h_j(a) = 0.$$

(ii) We now establish the second-order necessary condition (4.3). For convenience, assume that $I(a) = \{1, \ldots, r\}$, where $0 \le r \le m$. Let

$$\Phi_n = \{u \in \mathbb{R}^n : \chi^n(u) = 0\},$$

where

$$\chi_k^n(u) = \begin{cases} \langle \nabla g_k(x_n), u \rangle, & k = 1, \ldots, r \\ \langle \nabla h_{k-r}(x_n), u \rangle, & k = r + 1, \ldots, r + l. \end{cases}$$

Then $\Phi_n \neq \emptyset$, $\forall n \in N$ since $0 \in \Phi_n$. It follows from the linearly independent condition that

$$\limsup_{n \to \infty} \operatorname{rank}(\chi^n) \le \operatorname{rank}(\chi).$$

By Lemma 5.1, we obtain $\lim_{n \to \infty} \Phi_n = \Phi$.

Let $u \in \Phi$ and u_n be the projection of u on Φ_n. Since $0 \in \Phi_n$ and the projection is a nonexpansive operator, we have $\|u_n\| \le \|u\|$. Thus $\{u_n\}$ is bounded. Let u^* be an arbitrary limiting point of this sequence. Without loss of generality, suppose that $u_n \to u^*$. Let

$$H_n = \left\{ y \in \Phi_n : \|y - u\| = \inf_{z \in \Phi_n} \|z - u\| \right\},$$

$$H = \left\{ y \in \Phi : \|y - u\| = \inf_{z \in \Phi} \|z - u\| \right\}.$$

Noting that $\lim_{n \to \infty} \Phi_n = \Phi$, it follows from Theorem 4.2 that H is upper semi-continuous at a, i.e., $u^* \in H = \{u\}$. We proved that $\lim_{n \to \infty} u_n = u$.

Then it follows from $u_n \in \Phi_n$ and (4.7) that

$$\left\langle \left(A_n + \sum_{i \in I(a)} \lambda_i^n B_n^i + \sum_{j=1}^{l} \mu_j^n C_n^j \right) u_n, u_n \right\rangle + \sum_{i \notin I(a)} \lambda_i^n \langle B_n^i u_n, u_n \rangle \ge 0. \quad (5.3)$$

Using the locally boundedness of the generalized Hessians $\partial^2 f(a)$ we have that $\{A_n\}$, $\{B^i_n\}$, $\{C^j_n\}$, $i = 1, \ldots, m$, $j = 1, \ldots, l$ are bounded sequences. Without loss of generality, we assume that $A_n \to A$, $B^i_n \to B_i$, $C^j_n \to C_j$, $i \in I(a)$, $j = 1, \ldots, l$. Noting that $x_n \to a$ and the upper semi-continuity of the mapping $x \to \partial^2 f(x)$, we have $A \in \partial^2 f(a)$, $B_i \in \partial^2 g_i(a)$, $C_j \in \partial^2 h_j(a)$, $i \in I(a)$, $j = 1, \ldots, l$. Taking the limit in (5.3) and noting (4.4), we have

$$\left\langle \left(A + \sum_{i \in I(a)} \lambda_i B_i + \sum_{j=1}^{l} \mu_j C_j \right) u, u \right\rangle \geq 0. \quad \blacksquare$$

Acknowledgment

This research was done under the supervision of Dr. V. Jeyakumar whose helpful guidance, valuable suggestions and careful reading of the paper are much appreciated.

References

1. H. Attouch and R. Wets, Epigraphical analysis. *Ann. Inst. H. Poincaré Anal. Non Linéaire* **6** (1989), 73–100.
2. A. Auslender, "Optimization: Méthods Numérques," Masson, Paris, 1976.
3. A. Auslender, Penalty methods for computing points that satisfying second-order necessary conditions, *Math. Programming* **17** (1979), 229–238.
4. J. P. Bulteau and J. P. Vial, Curvilinear path and trust region in unconstrained optimization: a convergence analysis, *Math. Programming Study* **30** (1987), 82–101.
5. F. H. Clarke, "Optimization and Nonsmooth Analysis," Wiley, New York, 1983.
6. R. Cominetti and R. Correa, A generalized second-order derivative in nonsmooth optimization, *SIAM J. Control Optim.* **28** (1990), 789–809.
7. G. B. Dantzig, J. Folkman, and N. Shapiro, On the continuity of the minimum set of a continuous function, *J. Math. Anal. Appl.* **17** (1967), 519–548.
8. A. V. Fiacco and G. P. McCormick, "Nonlinear Programming: Sequential Unconstrained Minimization Techniques," Wiley, New York, 1968.
9. J. B. Hiriart-Urruty, Characterizations of the plenary hull of the generalized Jacobian matrix, *Math. Programming Study* **17** (1982), 1–12.
10. 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, *Appl. Math. Optim* **11** (1984), 43–56.
11. V. Jeyakumar and X. Q. Yang, Convex composite minimization with $C^{1,1}$ functions *J. Optimiz. Theory Appl.*, to appear.
12. G. P. McCormick, An arc method for nonlinear programming, *SIAM J. Control* **13** (1975), 1194–1216.
13. X. Q. Yang and V. Jeyakumar, Generalized second-order directional derivatives and optimization with $C^{1,1}$ functions, *Optimization* **26** (1992), 165–185.
14. X. Q. Yang, Generalized second-order directional derivatives and optimality conditions, *Nonlinear Analysis–TMA*, to appear.
15. X. Q. Yang, Second-order conditions in $C^{1,1}$ optimization with applications, *Numer. Funct. Anal. Optimiz.* **14** (1993), 621–632.