International Journal of Pure and Applied Mathematics
Volume 23 No. 3 2005, 299-310

# ON FRITZ JOHN AND KKT NECESSARY CONDITIONS OF CONSTRAINED QUASIDIFFERENTIABLE OPTIMIZATION

Zun-Quan Xia[1 §], Chun-Ling Song[2], Li-Wei Zhang[3]
[1,2,3]Centre for Optimization Research and Applications (CORA)
Department of Applied Mathematics
Dalian University of Technology
Dalian 116024, Liaoning, P.R. CHINA
[1]e-mail: zqxiazhh@dlut.edu.cn
[2]e-mail: mary_sohu@163.com
[3]e-mail: lwzhang@online.ln.cn

**Abstract:** The Fritz John and KKT necessary conditions of constrained quasidifferentiable optimization in the sense of Demyanov and Rubinov (1980), in the form of Demyanov sum of quasidifferential (1981), are presented and derived via Ekeland principle under the assumption of upper semicontinuity of the mapping of Demyanov sum of quasidifferential, that is weaker than the one given by Gao (2000). As an application, the optimality conditions of multi-objective optimization are also presented.

**AMS Subject Classification:** 26A27, 90C26, 90C30, 49K99
**Key Words:** nonsmooth optimization, quasidifferentiable functions, Ekeland principle, Demyanov difference (sum), multi-objective optimization

## 1. Introduction

Necessary optimality conditions, in the multiplier form, of constrained quasidifferentiable optimization have being studied since the early of 1980's, see for instance, Shapiro [13], Uderzo [14], Eppler et al [7], Gao [8] and Xia [15]. But, in these necessary conditions, mutipliers usually depend on super-gradients, in other words, different super-gradients would deduce different multipliers.

Received: May 28, 2005 
§Correspondence author

Gao [9] presented a set of Fritz John necessary conditions via Demyanov difference, due to Demyanov [2], for constrained quasidifferentiable minimization with inequality and equality constraints under the assumption that the objective function and the constraints are required to be uniformly directionally differentiable, and some constrained qualification is required. In this paper, the result is improved by a weaker assumption.

This paper is organized as follows: In Section 2, Fritz John necessary conditions in Gao's form, Gao [9], are given under the assumption that Demyanov sum of quasidifferential is upper semi-continuous. In Section 3, we prove that the assumption of the upper semi-continuity is rational. In Section 4, the necessary conditions of KKT type are presented. In Section 5, the necessary optimality conditions of multi-objective programming are given.

## 2. Fritz John Necessary Conditions

We first recall some related concepts that will be used later on. A function $f$, defined on an open set $\mathcal{O} \subseteq R^n$, is said to be quasidifferentiable (q.d.) at $x \in \mathcal{O}$ in the sense of Demyanov and Rubinov, if $f$ is directionally differentiable at $x$ and $f'(x;d)$ can be expressed as the difference of two sublinear functions, i.e., there exists a pair of nonempty compact convex subsets $[\underline{\partial} f(x), \overline{\partial} f(x)]$ of $R^n$ such that $f'(x;d) = \max_{v \in \underline{\partial} f(x)} \langle v, d\rangle + \min_{w \in \overline{\partial} f(x)} \langle w, d\rangle, \forall d \in R^n$. $\underline{\partial} f(x)$ is called the subdifferential of $f$ at $x$, and $\overline{\partial} f(x)$ the superdifferential of $f$ at $x$. Let $U$ and $V$ be nonempty compact convex subsets of $R^n$, and $T$ be a subset of $R^n$ with full measure such that $\nabla\delta^*(x\,|\,U)$ and $\nabla\delta^*(x\,|\,V)$ are differentiable at $x \in T$, $G_x(U) = \{u \in U \,|\, \langle u, x\rangle = \max_{u\in U}\langle x, u\rangle\}$. The set $U \dot{-} V = \operatorname{clco}\{\nabla\delta^*(x\,|\,U) - \nabla\delta^*(x\,|\,V) \,|\, x \in T\}$ is called the Demyanov difference of $U$ and $V$, the set $U \ddot{-} V = \operatorname{clco}\{G_x(U) - G_x(V) \,|\, x \neq 0\}$ is called quasi-difference of $U$ and V, see for instance, Demyanov et al [3]. For convenience, the notations will be used: $U \dot{+} V := U \dot{-} (-V)$, $U \ddot{+} V := U \ddot{-} (-V)$, $\partial^{\dot{+}} f(x) := \underline{\partial} f(x) \dot{+} \overline{\partial} f(x)$ and $\partial^{\ddot{+}} f(x) := \underline{\partial} f(x) \ddot{+} \overline{\partial} f(x)$.

A set-valued mapping $\alpha : R^n \to 2^{R^m}$ is said to be upper semi-continuous at $x$ if for any $\varepsilon > 0$, there exists $\delta > 0$ such that $\alpha(x') \subseteq \alpha(x) + B_\varepsilon, \forall x' \in B(x,\delta)$, where $B_\varepsilon := B(0,\varepsilon)$ and $B(x,\delta) := \{y \,|\, \|x-y\| \le \delta\}$. For simplicity, "u.s.c." stands for "upper semi-continuous" or "upper semi-continuity".

**Lemma 2.1.** (Diamond et al [5]) *Let $U$ and $V$ be nonempty convex compact subsets of $R^n$, and $U \supseteq V$. Then $0 \in U \dot{-} V$.*

**Theorem 2.1.** *Suppose that $f$ is a quasidifferentiable function defined on $R^n$, and $\overline{x} \in R^n$ is the minimizer of $f$. Then $0 \in \partial^{\dot{+}} f(\overline{x})$.*

*Proof.* It follows from Theorem 3.1 in Demyanov et al [3] and Lemma 2.1. □

**Lemma 2.2.** (Diamond et al [5]) *Suppose $U, V, U_1, V_1, U_2, V_2$ and $W$ are nonempty compact convex subsets of $R^n$. Then one has,*

*(1)* $-(U \dot{-} V) = V \dot{-} U$;

*(2)* $\lambda(U \dot{-} V) = \lambda U \dot{-} \lambda V, \lambda \geq 0$;

*(3)* $(U + W) \dot{-} (V + W) = U \dot{-} V$;

*(4)* $(U_1 + U_2) \dot{-} (V_1 + V_2) \subseteq (U_1 \dot{-} V_1) + (U_2 \dot{-} V_2)$.

Consider the optimization problem:

$$\text{(P)} \qquad \begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \cdots, m, \\ & h_j(x) = 0, \quad j = 1, \cdots, p, \end{array}$$

where $f_i, h_j : R^n \to R^1$, $i = 0, \cdots, m$, $j = 1, \cdots, p$, are continuous and quasidifferentiable.

**Theorem 2.2.** (Theorem 1.1, Ekeland [6]; Proposition 1.43, Rockfellar et al [12]) *Suppose that $f : R^n \to R^1$ is lower semi-continuous and $\hat{x}$ is an $\varepsilon-$ minimizer of $f$. Then for any $\delta > 0$, there exists a point $\overline{x}$ possessing the following properties:* (1) $f(\overline{x}) < f(\hat{x})$; (2) $\overline{x} \in B(\hat{x}, \varepsilon/\delta)$; (3) $\{\overline{x}\} = \arg\min_x \{f(x) + \delta \|x - \overline{x}\|\}$.

**Lemma 2.3.** *Let $\alpha : R^n \to 2^{R^m}$ be u.s.c. at $x$ and $T$ be a nonempty subset of $R^1$ not containing 0. Define a set-valued mapping $(\alpha, I) : R^n \times T \to 2^{R^m}$ as follows,*

$$(x, t) \mapsto t \cdot \alpha(x). \tag{2.1}$$

*Then, $(\alpha, I)$ is u.s.c. at $(x, t)$ for every $t \in T$.*

*Proof.* First, by the definition of u.s.c. it is easy to check that $(\alpha, I)$ is locally bounded at $x$, i.e., there exists a neighborhood $V$ of $x$ such that $(\alpha, I)(V)$ is bounded. Therefore, it is sufficient to check that $(\alpha, I)$ is closed at $(x, t)$ (see Hogan [10]), i.e., for any $x_k \to x, t_k \to t, y_k \in (\alpha, I)(x_k, t_k)$ and $y_k \to y$, we shall prove that $y \in (\alpha, I)(x, t)$. From $y_k \in (\alpha, I)(x_k, t_k)$ and $t_k \neq 0$, one has that $y_k/t_k \in \alpha(x_k)$. Since $\alpha$ is u.s.c., we have that $y/t \in \alpha(x)$, i.e., $y \in t \cdot \alpha(x)$, in other words, $y \in (\alpha, I)(x, t)$. In consequence, $(\alpha, I)$ is u.s.c. at $(x, t)$. □

**Theorem 2.3.** *Let $\overline{x} \in R^n$ be a minimizer to problem* (P). *Suppose the set-valued mapping $x \mapsto \partial^{\dot{+}} f_i(x), i = 0, \cdots, m$ and $x \mapsto \partial^{\dot{+}} h_j(x), j = 1, \cdots, p$,* *are u.s.c. at $\overline{x}$. Then, there exist scalars $\overline{\lambda}_i \geq 0, i = 0, 1, \cdots, m$, $\overline{\mu}_j, j = 1, \cdots, p$, not all zero, such that*

$$0 \in \sum_{i=0}^{m} \overline{\lambda}_i \partial^{\dot{+}} f_i(\overline{x}) + \sum_{j=1}^{p} \overline{\mu}_j \partial^{\dot{+}} h_j(\overline{x}), \tag{2.2}$$

$$\overline{\lambda}_i f_i(\overline{x}) = 0, \quad i = 1, \cdots, m. \tag{2.3}$$

*Proof.* Let

$$T = \{t = (\lambda_0, \lambda, \mu) \,|\, \lambda_0 \in R^1, \lambda \in R^m, \mu \in R^p, \lambda_0 \geq 0, \lambda \geq 0_m, \|t\| = 1\}$$

and for any $\varepsilon > 0$, define

$$\begin{aligned} F_\varepsilon(x) &:= \max\nolimits_{t \in T} L_\varepsilon(t, x) \\ &:= \max\nolimits_{t \in T} \{\lambda_0[f_0(x) - f_0(\overline{x}) + \varepsilon] + \textstyle\sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{j=1}^{p} \mu_j h_j(x)\}. \end{aligned}$$

Then it can be checked that $\overline{x}$ is an $\varepsilon-$ minimizer of minimizing $F_\varepsilon(x)$ over $R^n$. In fact, $F_\varepsilon(\overline{x}) = \varepsilon$, since the maximizer in $F_\varepsilon(\overline{x})$ must be attained at $(1, 0_m, 0_p)$. In addition, for all $x \in R^n$, it can be checked that $F_\varepsilon(x) > 0$:

*Case 1.* If $x$ is feasible for (P), then $F_\varepsilon(x) = f_0(x) - f_0(\overline{x}) + \varepsilon \geq \varepsilon > 0$;

*Case 2.* If there exists an $i_0 \in \{1, \cdots, m\}$, such that $f_{i_0}(x) > 0$, then one has $L_\varepsilon(\widetilde{t}, x) > 0$, where $\widetilde{t} \in T$ and the $i_0$-th component of $\widetilde{t}$ is 1 and other components are 0, so $F_\varepsilon(x) > 0$;

*Case 3.* If there exists a $j_0 \in \{1, \cdots, p\}$, such that $h_{j_0}(x) \neq 0$, then one has $L_\varepsilon(\widetilde{t}, x) > 0$, where $\widetilde{t} \in T$ and the $j_0$-th component of $\widetilde{t}$ is $\operatorname{sign}(h_{j_0}(x))$ and others are 0, so $F_\varepsilon(x) > 0$.

So, $F_\varepsilon(\overline{x}) \leq \inf_x F_\varepsilon(x) + \varepsilon$ and $\overline{x}$ is an $\varepsilon-$ minimizer of $\min F_\varepsilon(x)$ over $R^n$. By Theorem 2.2, for $\delta = \sqrt{\varepsilon}$, there exists an $x_\varepsilon \in R^n$ shares the following properties:

(a) $\|x_\varepsilon - \overline{x}\| \leq \sqrt{\varepsilon}$;

(b) $F_\varepsilon(x_\varepsilon) < F_\varepsilon(\overline{x})$;

(c) $\{x_\varepsilon\} = \arg\min\{F_\varepsilon(x) + \sqrt{\varepsilon}\|x - x_\varepsilon\|\}$.

Moreover, since $F_\varepsilon(\cdot) + \sqrt{\varepsilon}\|\cdot - x_\varepsilon\|$ is quasidifferentiable, it follows from (c) and Theorem 2.1 that

$$\begin{aligned} 0 &\in \underline{\partial}(F_\varepsilon(x) + \sqrt{\varepsilon}\|x - x_\varepsilon\|)|_{x=x_\varepsilon} \dot{+} \overline{\partial}(F_\varepsilon(x) + \sqrt{\varepsilon}\|x - x_\varepsilon\|)|_{x=x_\varepsilon} \\ &= [\underline{\partial} F_\varepsilon(x_\varepsilon) + \sqrt{\varepsilon} B_n] \dot{+} \overline{\partial} F_\varepsilon(x_\varepsilon), \end{aligned} \tag{2.4}$$

where $B_n$ is the unit ball in $R^n$.

Since $F_\varepsilon(x_\varepsilon) > 0$, there is a unique $t_\varepsilon = (\lambda_{\varepsilon_0}, (\lambda_{\varepsilon_i})_{i=1,\cdots,m}, (\mu_{\varepsilon_j})_{j=1,\cdots,p})$ lying in $T$ at which the maximum in the definition of $F_\varepsilon(x_\varepsilon)$ is attained. Therefore,

$$\underline{\partial} F_\varepsilon(x_\varepsilon) = \sum_{i=0}^{m} \lambda_{\varepsilon_i} \underline{\partial} f_i(x_\varepsilon) + \sum_{j=1}^{p} \underline{\partial}(\mu_{\varepsilon_j} h_j)(x_\varepsilon), \tag{2.5}$$

and

$$\overline{\partial} F_\varepsilon(x_\varepsilon) = \sum_{i=0}^{m} \lambda_{\varepsilon_i} \overline{\partial} f_i(x_\varepsilon) + \sum_{j=1}^{p} \overline{\partial}(\mu_{\varepsilon_j} h_j)(x_\varepsilon). \tag{2.6}$$

Hence, it follows from (2.4), (2.5) and (2.6) that

$$0 \in [\sum_{i=0}^{m} \lambda_{\varepsilon_i} \underline{\partial} f_i(x_\varepsilon) + \sum_{j=1}^{p} \underline{\partial}(\mu_{\varepsilon_j} h_j)(x_\varepsilon) + \sqrt{\varepsilon} B_n] \dot{+} [\sum_{i=0}^{m} \lambda_{\varepsilon_i} \overline{\partial} f_i(x_\varepsilon) + \sum_{j=1}^{p} \overline{\partial}(\mu_{\varepsilon_j} h_j)(x_\varepsilon)]. \tag{2.7}$$

By Lemma 2.2(4), we have

$$0 \in \sum_{i=0}^{m} \lambda_{\varepsilon_i} \partial^{\dot{+}} f_i(x_\varepsilon) + \sum_{j=1}^{p} \partial^{\dot{+}}(\mu_{\varepsilon_j} h_j)(x_\varepsilon) + \sqrt{\varepsilon} B_n. \tag{2.8}$$

In what follows, we prove $\partial^{\dot{+}}(\mu_{\varepsilon_j} h_j)(x_\varepsilon) = \mu_{\varepsilon_j} \partial^{\dot{+}} h_j(x_\varepsilon)$. For the case where $\mu_{\varepsilon_j} \geq 0$, it is clear. It is enough to prove the case where $\mu_{\varepsilon_j} < 0$:

$$\begin{aligned}
& \partial^{\dot{+}}(\mu_{\varepsilon_j} h_j)(x_\varepsilon) = \underline{\partial}(\mu_{\varepsilon_j} h_j)(x_\varepsilon) \dot{-} [-\overline{\partial}(\mu_{\varepsilon_j} h_j)(x_\varepsilon)] \\
= \; & -\mu_{\varepsilon_j}(-\overline{\partial} h_j(x_\varepsilon) \dot{-} \underline{\partial} h_j(x_\varepsilon)) && \text{(Lemma 2.2(2))} \\
= \; & \mu_{\varepsilon_j}[\underline{\partial} h_j(x_\varepsilon) \dot{-} (-\overline{\partial} h_j(x_\varepsilon))] && \text{(Lemma 2.2(1))} \\
= \; & \mu_{\varepsilon_j} \partial^{\dot{+}} h_j(x_\varepsilon).
\end{aligned}$$

Hence, we have from (2.8) that

$$0 \in \sum_{i=0}^{m} \lambda_{\varepsilon_i} \partial^{\dot{+}} f_i(x_\varepsilon) + \sum_{j=1}^{p} \mu_{\varepsilon_j} \partial^{\dot{+}} h_j(x_\varepsilon) + \sqrt{\varepsilon} B_n. \tag{2.9}$$

Since $\varepsilon$ is arbitrary, taking a sequence $\{\varepsilon_k\}_{k=1}^{\infty}$ such that $\varepsilon_k \downarrow 0$ as $k \to \infty$, we have that $\|x_k - \overline{x}\| \leq \sqrt{\varepsilon_k}$ and

$$0 \in \sum_{i=0}^{m} \lambda_{k_i} \partial^{\dot{+}} f_i(x_k) + \sum_{j=1}^{p} \mu_{k_j} \partial^{\dot{+}} h_j(x_k) + \sqrt{\varepsilon_k} B_n, \tag{2.10}$$

where $t_k := t_{\varepsilon_k}$ and $x_k := x_{\varepsilon_k}$.

Since $T$ is bounded, without loss of generality, we can assume that $t_k \to \overline{t} = (\overline{\lambda}_0, \overline{\lambda}, \overline{\mu})$. By the definition of $T$ and the upper semi-continuity of $\partial^{\dot{+}} f_i(x)$ and $\partial^{\dot{+}} h_j(x)$, one has from Lemma 2.3 that the set-valued mapping $(x,t) \mapsto t\partial^{\dot{+}} f_i(x)$ and $(x,t) \mapsto t\partial^{\dot{+}} h_j(x)$ are u.s.c. Thus, the set-valued mapping

$$(x,t,\varepsilon) \mapsto \sum_{i=0}^{m} \lambda_i \partial^{\dot{+}} f_i(x) + \sum_{j=1}^{p} \mu_j \partial^{\dot{+}} h_j(x) + \sqrt{\varepsilon} B_n \tag{2.11}$$

is u.s.c., where $t = (\lambda_0, \lambda, \mu)$, $\lambda = (\lambda_1, \cdots, \lambda_m)$ and $\mu = (\mu_1, \cdots, \mu_p)$. Therefore, taking the limit to (2.10) as $k \to \infty$, we have that

$$0 \in \sum_{i=0}^{m} \overline{\lambda}_i \partial^{\dot{+}} f_i(\overline{x}) + \sum_{j=1}^{p} \overline{\mu}_j \partial^{\dot{+}} h_j(\overline{x}). \tag{2.12}$$

Since $\overline{t} = (\overline{\lambda}_0, \overline{\lambda}, \overline{\mu}) \in T$, it is obvious that $\overline{t} \neq 0_{1+m+p}$. Furthermore, if there exists an $i_0 \in \{1, \cdots, m\}$ such that $f_{i_0}(\overline{x}) < 0$, then $f_{i_0}(x_k) < 0$ for $k$ large enough by the continuity of $f_{i_0}$. So the maximum in the definition of $F_{\varepsilon_k}(x_k)$ must be attained at $\lambda_{k_{i_0}} = 0$ and hence $\overline{\lambda}_{i_0} = 0$, i.e., (2.3) holds. □

In the next section, some weaker conditions would be presented under which Demyanov sum of quasidifferential is u.s.c. Therefore, the assumption in Theorem 2.3 is rational.

## 3. On the Upper Semi-Continuity of $\partial^{\dot{+}} f$

In this section, by $\mathcal{K}(X)$ we denote the set of nonempty compact convex subsets of $X$, and $G_d(A)$ is the max-face of the convex compact set $A$ determined by $d$.

**Definition 3.1.** A quasidifferential (mapping), $[\underline{\partial}, \overline{\partial}]f$, of a q.d. function $f$ is said to be u.s.c. at $x$ if its sub- and super-differential (mapping) are u.s.c. at $x$.

**Lemma 3.1.** *Suppose the set-valued mapping $\alpha : R^n \to \mathcal{K}(R^m)$ is u.s.c. and let $f : R^m \times R^n \to R^1, f(x,y) = \max\{\langle x, u\rangle \mid u \in \alpha(y)\}$. Then the function $f$ is u.s.c.*

*Proof.* Omitted. □

**Corollary 3.1.** (Demyanov et al [4]) *Suppose the set-valued mapping $\alpha : R^n \to \mathcal{K}(R^m)$ is u.s.c. Let $f : R^n \to R^1, f(y) = \max\{\langle x, u\rangle \mid u \in \alpha(y)\}$, $x \in R^m$. Then the function $f$ is u.s.c.*

**Lemma 3.2.** *Let $\alpha : R^n \to \mathcal{K}(R^m)$ be an u.s.c. set-valued mapping, and $G(\alpha) : R^m \times R^n \to \mathcal{K}(R^m), (d, x) \mapsto G_d(\alpha(x))$. Then, $G(\alpha)$ is u.s.c.*

**Corollary 3.2.** *Suppose $\alpha : R^n \to \mathcal{K}(R^m)$ is u.s.c. Define $f : R^m \times R^n \to R^1$ as follows: $f(x, y) = \max\{\langle d, u\rangle \,|\, u \in G(\alpha)(x, y)\}$, $d \in R^m$. Then the function $f$ is u.s.c.*

**Corollary 3.3.** *Suppose $A$ is a compact convex subset of $R^n$. For any $d \in R^n$, define $f : R^n \to R^1$, $f(x) = \max\{\langle d, u\rangle \,|\, u \in G_x(A)\}$. Then $f$ is u.s.c.*

**Theorem 3.1.** *Suppose that $f$ is quasidifferentiable on $R^n$, and there exists a quasidifferential mapping $Df \in \mathcal{D}f$ such that $Df$ is u.s.c. at $x$ and $\partial^{\dot{+}} f(x) = \partial^{\ddot{+}} f(x)$. Then, Demyanov sum $\partial^{\dot{+}} f$ is u.s.c. at $x$.*

*Proof.* By contradiction, assume that $\partial^{\dot{+}} f$ is not u.s.c. at $x$. Then there exists $\varepsilon > 0$ such that for every $1/k, k = 1, 2, \cdots$, there exists $x_k$ satisfying $\|x_k - x\| \le 1/k$ and $\partial^{\dot{+}} f(x_k) \not\subseteq \partial^{\dot{+}} f(x) + B_\varepsilon$. But, we will prove the fact that for any $x_k \to x, k \to \infty$ and $\varepsilon > 0$, there exists $K > 0$, such that

$$\partial^{\dot{+}} f(x_k) \subseteq \partial^{\dot{+}} f(x) + B_\varepsilon, \quad \forall k \ge K. \tag{3.13}$$

This contradiction leads to completing the proof.

Since Demyanov sum of two convex compact sets is convex compact, (3.13) is equivalent to the following inequality

$$\delta^*(d \,|\, \partial^{\dot{+}} f(x_k)) \le \delta^*(d \,|\, \partial^{\dot{+}} f(x)) + \varepsilon, \quad \forall d \in B_n, \forall k \ge K, \tag{3.14}$$

where $B_n$ is the unit ball in $R^n$, see for instance, Demyanov et al [3] and Rockfellar [11]. Therefore, we prove (3.14) instead of doing (3.13). For any $d \in B_n$, one has from the definition of Demyanov sum that

$$\begin{aligned} &\delta^*(d \,|\, \partial^{\dot{+}} f(x_k)) \\ &\quad = \sup\{\langle d,\, v\rangle \,|\, v \in \{\nabla\delta^*(h \,|\, \underline{\partial} f(x_k)) - \nabla\delta^*(h \,|\, -\overline{\partial} f(x_k)) \,|\, h \in T_k\}\}, \end{aligned} \tag{3.15}$$

where $T_k = \{h \in R^n \,|\, \delta^*(\cdot \,|\, \underline{\partial} f(x_k))$ and $\delta^*(\cdot \,|\, -\overline{\partial} f(x_k))$ are differentiable at $h\}$.

For "sup" in (3.15) there are two cases:

*Case 1.* For every $k$ there exists a $h_k := h_k(d) \in T_k$ such that the supremum in (3.15) can be attained;

*Case 2.* Otherwise, for some $k$ there is no $h_k := h_k(d) \in T_k$ such that the supremum in (3.15) can be attained.

We now consider the Case 1, in other words,

$$\delta^*(d \,|\, \partial^{\dot{+}} f(x_k)) = \langle d, \nabla\delta^*(h_k \,|\, \underline{\partial} f(x_k)) - \nabla\delta^*(h_k \,|\, -\overline{\partial} f(x_k))\rangle. \tag{3.16}$$

Without loss of generality, it can be assumed that $\{h_k\}_{k=1}^{\infty} \longrightarrow h$ and $\|h_k\| = 1$. If $\delta^*(\cdot\,|\,\underline{\partial} f(x))$ and $\delta^*(\cdot\,|\,-\overline{\partial} f(x))$ are both differentiable at $h$, then there exists $K > 0$ such that

$$\delta^*(d\,|\,\partial^{\dot{+}} f(x_k)) \le \langle d, \nabla\delta^*(h\,|\,\underline{\partial} f(x)) - \nabla\delta^*(h\,|\,-\overline{\partial} f(x))\rangle + \varepsilon, \quad \forall k \ge K, \tag{3.17}$$

according to Lemma 3.2, Lemma 3.1 and the u.s.c of $\underline{\partial} f$ and $-\overline{\partial} f$. For

$$\nabla\delta^*(h\,|\,\underline{\partial} f(x)) - \nabla\delta^*(h\,|\,-\overline{\partial} f(x)) \in \partial^{\dot{+}} f(x),$$

one has that for any $k \ge K$, $\delta^*(d\,|\,\partial^{\dot{+}} f(x_k)) \le \delta^*(d\,|\,\partial^{\dot{+}} f(x)) + \varepsilon$. Therefore, the conclusion holds. Otherwise, $\delta^*(\cdot\,|\,\underline{\partial} f(x))$ and $\delta^*(\cdot\,|\,-\overline{\partial} f(x))$ are not both differentiable at $h$. By Lemma 3.2, Lemma 3.1 and the u.s.c of $\underline{\partial} f$ one has

$$\delta^*(d\,|\,\partial^{\dot{+}} f(x_k)) \le \delta^*(d\,|\,\widetilde{G}_h(x)) + \varepsilon, \quad \forall k \ge K, \tag{3.18}$$

where $\widetilde{G}_h(x) = G_h(\underline{\partial} f(x)) - G_h(-\overline{\partial} f(x))$. Moreover, by the definition of quasi-difference and $\|h\| = 1$, one has $\widetilde{G}_h(x) \subseteq \partial^{\ddot{+}} f(x) = \partial^{\dot{+}} f(x)$. In consequence, one has

$$\delta^*(d\,|\,\partial^{\dot{+}} f(x_k)) \le \delta^*(d\,|\,\partial^{\dot{+}} f(x)) + \varepsilon, \quad \forall k \ge K.$$

We now turn to Case 2 for some $k$ there is no $h_k := h_k(d) \in T_k$ such that the supremum of $\delta^*(d\,|\,\partial^{\dot{+}} f(x_k))$ can be attained at $\nabla\delta^*(h_k\,|\,\underline{\partial} f(x_k)) - \nabla\delta^*(h_k\,|\,-\overline{\partial} f(x_k))$. In this case, it follows from the definition of Demyanov sum that for any $\varepsilon > 0$, there must exist $h_k = h_k(d)$ such that $\delta^*(\cdot\,|\,\underline{\partial} f(x_k))$ and $\delta^*(\cdot\,|\,-\overline{\partial} f(x_k))$ are differentiable at $h_k$ and

$$\delta^*(d\,|\,\partial^{\dot{+}} f(x_k)) \le \langle d, \nabla\delta^*(h_k\,|\,\underline{\partial} f(x_k)) - \nabla\delta^*(h_k\,|\,-\overline{\partial} f(x_k))\rangle + \varepsilon \tag{3.19}$$

holds. The rest of the proof can be done by a way similar to the one used in Case 1. The demonstration is completed. □

## 4. KKT Necessary Conditions

To obtain KKT necessary conditions for (P), a constraint qualification will be presented in this section.

**Definition 4.1.** Let $A_1, A_2, \cdots, A_m$ be a family of the subsets of $R^n$. If for any $\lambda_1, \lambda_2, \cdots, \lambda_m \in R^1$, $0_n \in \sum_{i=1}^{m} \lambda_i A_i$ implies that $\lambda_1 = \lambda_2 = \cdots = \lambda_m = 0$, then $\{A_i\,|\,i = 1, \cdots, m\}$ is said to be of maximal rank.

**Remark 4.1.** If $R^n = R^{n_1} \oplus \cdots \oplus R^{n_m}, n_1 + \cdots + n_m = n$, and $A_i \subseteq \pi_{n_i}(R^n), i = 1, \cdots, m$, where $\pi_{n_i}$ denotes the projection mapping from $R^n$ to $R^{n_i}$. Then the condition of maximal rank is satisfied naturally.

**Theorem 4.1.** *Let $\overline{x}$ be the minimizer of* (P). *Suppose there exists a vector $v$ such that*

$$\delta^*(v \,|\, \partial^{\dot{+}} f_i(\overline{x})) < 0, \quad i = 1, \cdots, m, \tag{4.20}$$

$$\delta^*(v \,|\, \partial^{\dot{+}} h_j(\overline{x})) = 0, \quad j = 1, \cdots, p, \tag{4.21}$$

*and $\{\partial^{\dot{+}} h_j(\overline{x}) \,|\, j = 1, \cdots, p\}$ is of maximal rank. Then there exist scalars $\overline{\lambda}_i \geq 0, i = 1, \cdots, m$, and $\overline{\mu}_j, j = 1, \cdots, p$, not all zero, such that*

$$0 \in \partial^{\dot{+}} f_0(\overline{x}) + \sum_{i=1}^{m} \overline{\lambda}_i \partial^{\dot{+}} f_i(\overline{x}) + \sum_{j=1}^{p} \overline{\mu}_j \partial^{\dot{+}} h_j(\overline{x}), \tag{4.22}$$

$$\overline{\lambda}_i f_i(\overline{x}) = 0, \quad i = 1, \cdots, m. \tag{4.23}$$

*Proof.* By Theorem 2.3, there exist $\overline{\lambda}_i \geq 0, i = 0, \cdots, m$, $\overline{\mu}_j, j = 1, \cdots, p$, not all zero, such that

$$0 \in \sum_{i=0}^{m} \overline{\lambda}_i \partial^{\dot{+}} f_i(\overline{x}) + \sum_{j=1}^{p} \overline{\mu}_j \partial^{\dot{+}} h_j(\overline{x}), \tag{4.24}$$

$$\overline{\lambda}_i f_i(\overline{x}) = 0, \quad i = 1, \cdots, m. \tag{4.25}$$

It suffices to prove $\overline{\lambda}_0 \neq 0$. Assume to the contrary that $\overline{\lambda}_0 = 0$, then we have

$$0 \in \sum_{i=1}^{m} \overline{\lambda}_i \partial^{\dot{+}} f_i(\overline{x}) + \sum_{j=1}^{p} \overline{\mu}_j \partial^{\dot{+}} h_j(\overline{x}). \tag{4.26}$$

Since $\{\partial^{\dot{+}} h_j(\overline{x}) \,|\, j = 1, \cdots, p\}$ is of maximal rank, there must exist an $i \in \{1, \cdots, m\}$, such that $\overline{\lambda}_i \neq 0$. Otherwise, one has from (4.26) that

$$0 \in \sum_{j=1}^{p} \overline{\mu}_j \partial^{\dot{+}} h_j(\overline{x}), \tag{4.27}$$

and so $\overline{\mu}_j = 0, j = 1, \cdots, p$, which contradict to $((\overline{\lambda}_i)_{i=0}^m, (\overline{\mu}_j)_{j=1}^p) \neq 0_{1+m+p}$. Taking support function of the set $\Omega := \sum_{i=1}^m \overline{\lambda}_i \partial^{\dot{+}} f_i(\overline{x}) + \sum_{j=1}^p \overline{\mu}_j \partial^{\dot{+}} h_j(\overline{x})$, then one has from (4.20) and (4.21) that $\delta^*(v \,|\, \Omega) < 0$, which contradicts to (4.26). In consequence, the conclusion holds. $\square$

## 5. Fritz John Necessary Conditions of Multi-Objective Optimization

Let $f$ be a vector function $[f_0, f_1, \cdots, f_m]$ and consider the following problem:

$$\text{(PM)} \qquad \begin{array}{ll} \text{minimize} & f(x)\,, \\ \text{subject to} & g_j(x) \le 0, \quad j = 1, \cdots, p, \\ & h_k(x) = 0, \quad k = 1, \cdots, q, \end{array}$$

where $f_i, g_j, h_k : R^n \to R^1$, $i = 0, \cdots, m$, $j = 1, \cdots, p$ and $k = 1, \cdots, q$, are quasidifferentiable.

The feasible point $x$ is said to be Pareto optimal for (PM), if there is no feasible point $y$ such that $f_i(y) < f_i(x), i = 0, \cdots, m$.

**Theorem 5.1.** *Suppose that the set-valued mapping $x \longmapsto \partial^{\dot{+}} f_i(x), i = 0, \cdots, m$, $x \longmapsto \partial^{\dot{+}} g_j(x), j = 1, \cdots, p$ and $x \longmapsto \partial^{\dot{+}} h_k(x), k = 1, \cdots, q$, are u.s.c. at $\overline{x}$. If $\overline{x}$ is Pareto optimal for* (PM), *then there exist scalars $\overline{\lambda}_i \ge 0, i = 0, 1, \cdots, m$, $\overline{\mu}_j, j = 1, \cdots, p$ and $\overline{\gamma}_k, k = 1, \cdots, q$, not all zero, such that*

$$0 \in \sum_{i=0}^{m} \overline{\lambda}_i \partial^{\dot{+}} f_i(\overline{x}) + \sum_{j=1}^{p} \overline{\mu}_j \partial^{\dot{+}} g_j(\overline{x}) + \sum_{k=1}^{q} \overline{\gamma}_k \partial^{\dot{+}} h_k(\overline{x}), \tag{5.28}$$

$$\overline{\mu}_j g_j(\overline{x}) = 0, \quad j = 1, \cdots, p. \tag{5.29}$$

*Proof.* Define $L_\varepsilon(t, x)$ and $F_\varepsilon(x)$ as follows:

$$L_\varepsilon(t, x) = \sum_{i=0}^{m} \lambda_i (f_i(x) - f_i(\overline{x}) + \varepsilon) + \sum_{j=1}^{p} \mu_j g_j(x) + \sum_{k=1}^{q} \gamma_k h_k(x), \tag{5.30}$$

$$F_\varepsilon(x) = \max_{t \in \Lambda} L_\varepsilon(t, x), \tag{5.31}$$

where $\Lambda := \{t = (\lambda_0, \cdots, \lambda_m, \mu_1, \cdots, \mu_p, \gamma_1, \cdots, \gamma_q) \in R^{1+m+p+q} \,|\, \lambda_i \ge 0, i = 0, \cdots, m, \mu_j \ge 0, j = 1, \cdots, p, \|t\| = 1\}$.

It can be checked that $\overline{x}$ is $\varepsilon-$ optimal for minimizing $F_\varepsilon(x)$ over $R^n$. In fact, it is obvious that $F_\varepsilon(\overline{x}) = \varepsilon$. Moreover, if $x$ is feasible for (PM), then there must exist an $i_0 \in \{0, \cdots, m\}$ such that $f_{i_0}(x) \ge f_{i_0}(\overline{x})$, since $\overline{x}$ is Pareto optimal for (PM). So $L_\varepsilon(\widetilde{t}, x) \ge \varepsilon > 0$, where $\widetilde{t} \in \Lambda$, the $i_0$-th component of $\widetilde{t}$ is 1 and other components are 0. Hence, $F_\varepsilon(x) > 0$. If $x$ is not a feasible solution of (PM), one has $F_\varepsilon(x) > 0$ by the similar way to the part in the proof of Theorem 2.3.

The rest of the proof can be done by the technique used in proving Theorem 2.3 without any change. ☐

**Acknowledgements**

The authors are grateful to Professor Y. Gao for his useful comments and suggestions on Corollary 3.3.

The work was supported by the Foundations of Ph.D. Units of China (20020141013), State Foundation of Natural Science of China (10471015).

## References

[1] V.F. Demyanov, A.M. Rubinov, On quasidifferentiable functionals, *Doklady Akademi Nauk SSSR*, **250** (1980), 21-25. (Translated in *Soviet Mathematics Doklady*, **21** (1980), 14-17)

[2] V.F. Demyanov, On a relation between the Clarke subdifferential and quasidifferential, *Vestnik Leningrad University*, **13** (1981), 183-189.

[3] V.F. Demyanov, A.M. Rubinov, Constructive Nonsmooth Analysis, *Verlag Peter Lang*, Berlin (1995).

[4] V.F. Demyanov, A.M. Rubinov, Quasidifferential Calculus, *Optimization Software Inc.*, New York (1986).

[5] P. Diamond, P. Kloeden, A. Rubinov, A. Vladimirov, Comparative properties of three metrics in the space of compact convex sets, *Set-Valued Analysis*, **5** (1997), 267-289.

[6] I. Ekeland, On the variational principle, *Journal of Mathematical Analysis and Applications*, **47** (1974), 324-353.

[7] E. Eppler, B. Luderer, The Lagrange principle and quasidifferentil calculus, *University of Techonology of Karl-Marx-Stadt*, **29** (1987), 187-192.

[8] Y. Gao, A weak Fritz John condition for quasidifferentiable functions, *Operations Research Transactions*, **9** (1990), 43-44, In Chinese.

[9] Y. Gao, Demyanov difference of two sets and optimality conditions of Lagrange multiplier type for constrained quasidifferentiable optimization, *Journal of Optimization Theory and Applications*, **104** (2000), 377-394.

[10] W.W. Hogan, Point-To-Set maps in mathematical programming, *SIAM Review*, **15**, No. 3 (1973), 591-603.

[11] R.T. Rockfellar, *Convex Analysis*, Princeton University Press, Princeton, New Jersey (1970).

[12] R.T. Rockfellar, R.J.-B. Wets, *Variational Analysis*, Springer-Verlag, Berlin Heidelberg (1998).

[13] A. Shapiro, On optimality conditions in quasidifferentiable optimization, *SIAM Journal of Control and Optimization*, **24** (1984), 610-617.

[14] A. Uderzo, Quasi-Multipliers rules for quasidifferentiable extremum problems, *Optimization*, **51**, No. 6 (2002), 761-795.

[15] Z.-Q. Xia, Generalized Fritz John necessary conditions for quasidifferentiable minimization, *Approximation, Optimization and Computing: Theory and Applications*, Elsevier Science Publishers B.V. (1990), 325-327.