Nihonkai Math. J.
Vol.2(1991), 99-115

OPTIMALITY CONDITIONS FOR LIPSCHITZ FUNCTIONS ON BANACH SPACES

Hang–Chin Lai

Abstract. In convex analysis, if a convex function f defined on a Banach space X attains its minimum at x_0, then $0 \in \partial f(x_0)$, the subdifferential of f at x_0. Thus we study in this paper for optimality conditions of a minimization problem of locally Lipschitz objective subject to an inequality and equality constraints with values in Banach spaces. We replace f by the Lagrangian L for a given programming problem, and prove that the Kuhn–Tucker/Fritz John multiplier rule holds. That is,

$$\theta \in \partial_x^{\circ} L(x_0, \lambda, \mu, \nu, K),$$

the generalized gradient of L with respect to $x \in X$.

1980 (AMS) Mathematics subject classification: 49D37, 49D29, 90C48

Keywords and phrases: Lipschitz function, generalized gradient, generalized derivative, Kuhn–Tucker type conditions, Fritz John type condition, multiobjective, vector objective, Lagrangian.

1. INTRODUCTION AND PRELIMINARIES

Many authors investigated the necessary and sufficient conditions for optimal solutions of convex/concave programming problems with inequality and equality constraints. These optimal conditions are essentially established in Kuhn–Tucker/ Fritz John multiplier rules for convex programming problems. For examples, one can consult Lai and Ho [8, Theorem 3.1], Kanniapan [6, Theorems 3.2 and 3.4]; Kanniapan and Sastry [7, Theorem 2.2]. See also Lai and Yang [9], and Minami [12] etc. While the objective function and constraint functions are locally Lipschitz, Hiriart–Urruty [4 – 5] and Clarke [1–2] have established some optimal conditions for real valued functions. In this paper, we shall investigate Fritz John multiplier rule for Banach space–valued optimization problem of locally Lipschitz functions.

It is known that if a convex function f on a Banach space X attains its minimum at x_0, then $0 \in \partial f(x_0)$, the subdifferential of f at x_0. For some related optimization problems, one can refere to Lai and Lin [10–11], Yu [13] and Zowe [14–15]. To study the necessary conditions of cone optimality in a programming problem for locally Lipschitz functions with values in an ordered Banach space, we will replace such f by the Lagrangian L for a given optimization problem and prove that the Fritz John multiplier rule holds. That is,

$$\theta \in \partial^{o}_{x} L(x_0, \lambda, \mu, \nu, K),$$

the generalized gradient of L with respect to $x \in X$. Our main result is established in Theorem 2.3. It is a new expression of optimality condition for locally Lipschitz functions on Banach spaces. The proof given here is interesting by using the Ekeland variational principle.

For convenience, we start from definitions and basic properties about the generalized gradient (cf. Clarke [2]). Let X be a real Banach space. A function f: $X \longrightarrow \mathbb{R}$, the real number field, is called *locally Lipschitz* of rank K if, for any

$x \in X$, there is a neighborhood U and a constant $K > 0$ such that

$$|f(z) - f(y)| \leq K\|z - y\| \qquad \text{for all } z, y \in U. \tag{1.1}$$

The *generalized directional derivative* $f^{\circ}(x; v)$ of f at $x \in X$ in the direction v is defined by

$$f^{\circ}(x; v) = \limsup_{\substack{y \to x \\ t \downarrow 0}} \frac{f(y+tv) - f(y)}{t}. \tag{1.2}$$

If f is a locally Lipschitz function, then the lim sup in (1.2) exists. The function $v \longrightarrow f^{\circ}(x; v)$ is positively homogeneous and convex such that

$$|f^{\circ}(x; v)| \leq K\|v\|. \tag{1.3}$$

The *generalized gradient* of f at $x \in X$, denoted by $\partial^{\circ} f(x)$ to distinguish from the subdifferential $\partial f(x)$ in convex analysis, is defined to be the set of elements $\zeta \in X^*$ such that

$$< v, \zeta > \leq f^{\circ}(x; v) \qquad \text{for all } v \in X. \tag{1.4}$$

That is,

$$\partial^{\circ} f(x) = \{\zeta \in X^* \,;\, < v, \zeta > \leq f^{\circ}(x; v) \text{ for all } v \in X\}. \tag{1.5}$$

From (1.3) and (1.4), it is immediate that

$$\partial^{\circ} f(x) \subset K\overline{B}_* \tag{1.6}$$

where $\overline{B}_*$ is the closed unit ball of X^*.

Let f and g be locally Lipschitz functions. From $\limsup (f+g) \leq \limsup f + \limsup g$, one can get

$$(f + g)^{\circ}(x; v) \leq f^{\circ}(x; v) + g^{\circ}(x; v) \tag{1.7}$$

for all $x, v \in X$. It follows that

$$\partial^{\circ}(f + g)(x) \subset \partial^{\circ} f(x) + \partial^{\circ} g(x). \tag{1.8}$$

Note that if f is a locally Lipschitz function of rank K, it is easy to see that the generalized gradient $\partial^{o}f(x)$ possesses the following properties.

(1) $\partial^{o}f(x) \neq \emptyset$ is a convex, weak*–compact subset of X^* and $\| \zeta \|_* \leq K$ for every $\zeta \in \partial^{o}f(x)$.

(2) If f is also convex, then

$$\partial^{o}f(x) = \partial f(x) \qquad \text{for } x \in X, \tag{1.9}$$

where

$$\partial f(x) = \{\zeta \in X^* \mid < y-x, \zeta > \leq f(y) - f(x), \text{ for all } y \in X\} \tag{1.10}$$

stands for the subdifferential of f.

(3) If f admits a Gateaux derivative $Df(y)$ at $y \in U(x)$, a neighborhood of x, and $Df: X \longrightarrow X^*$ is continuous, then $\partial^{o}f(x) = \{Df(x)\}$.

(4) The generalized gradient $\partial^{o}f$ is an upper semi–continuous mapping from point to set, that is, for any sequence $\zeta_i \in \partial^{o}f(x_i)$, $i = 1, 2, \cdots$,

$$x_i \longrightarrow x \text{ and } \zeta_i \longrightarrow \zeta \Longrightarrow \zeta \in \partial^{o}f(x). \tag{1.11}$$

(5) If x_0 minimizes $f(x)$ over X, then

$$0 \in \partial^{o}f(x_0). \tag{1.12}$$

In this paper, the main task is to show that (1.12) still holds if f is replaced by a Lagrangian defined on a programming problem (P) (see later context) for locally Lipschitz functions with values in ordered Banach spaces. This result extends the Theorems 6.1.1 and 6.1.3 in Clarke [2], and Hiriart–Urruty [5, Theorem 3.1]. In order to get this result, the technique of Ekeland variational principle (see Ekeland [3]) is useful in our proof.

2. KUHN–TUCKER MULTIPLIER RULE FOR LIPSCHITZ FUNCTIONS ON BANACH SPACES

Let X be a real Banach space, and let Y, Z and W be reflexive real

Banach spaces. Let C_Y and C_Z be closed convex cones of Y and Z respectively, and int $C_Y \neq \emptyset$. Suppose that

$$f: X \longrightarrow Y, \quad g: X \longrightarrow Z \quad \text{and} \quad h: X \longrightarrow W$$

are locally Lipschitz mappings. We consider the following programming problem

$$(P) \quad \begin{cases} \text{minimize} \;\; f(x) \\ \text{subject to} \;\; g(x) \leq \theta,\ h(x) = \theta,\ x \in E, \text{ a closed subset of } X \end{cases}$$

where θ stands the zero vector in linear spaces. We will present a version of Lagrangian multiplier rule for vector objective problem (P). We say that $x_0 \in E$ is a *minimal/weakly minimal* solution for problem (P) if there does not exist other feasible $x \in E$ such that

$$f(x_0) \in f(x) + C_Y \;\; / \;\; f(x_0) \in f(x) + \text{int } C_Y.$$

If $C_Y \neq Y$, then the set of minimal solutions is contained in the set of weakly minimal solutions. Thus we have only to derive the theorem for weakly minimal solution only.

Let Y^* be the dual space of Y. The dual cone C_{Y^*} of C_Y is defined by

$$C_{Y^*} = \{y^* \in Y^* \mid \langle y, y^* \rangle \geq 0 \text{ for all } y \in C_Y\}.$$

Let T be a compact metric space, and for $x_0 \in X$, let the mapping: $t \in T \longrightarrow r_t(x)$ be upper semicontinuous for x near x_0.

Now let $\{r_t(x):\ t \in T\}$ be a family of bounded locally Lipschitz functions on X. Define a real valued function r on X by

$$r(x) = \max_{t \in T} r_t(x) \qquad \text{for } x \in X. \tag{2.1}$$

Since, for each t, r_t is Lipschitz, we see that r is also a Lipschitz function, thus the image of the multimapping $M:\ x \in X \longrightarrow 2^T$, defined by,

$$M(x) = \{t \in T;\ r_t(x) = r(x)\} \quad \text{for } x \text{ near } x_0 \tag{2.2}$$

is a nonempty, closed subset of T. For any $t \in T$ and $x \in X$, we define

$$\partial^o_{[T]} r_t(x) = \overline{co}^*\{\zeta \in X^* |\ \zeta_i \in \partial^o r_{t_i}(x_i) \text{ with } t_i \longrightarrow t,\ x_i \longrightarrow x \text{ and } \zeta_i \xrightarrow{w^*} \zeta\} \tag{2.3}$$

where $\overline{co}^*$ is the weak* closed convex hull of the continuous linear functional in X^*.

A multifunction: $(\tau, x) \in T \times X \longrightarrow \partial^o r_\tau(x)$ is called (weak*) *closed* at (t, x_0) if

$$\partial^o_{[T]} r_t(x_0) = \partial^o r_t(x_0). \tag{2.4}$$

It is easy to characterize that

$$\partial^o_{[T]} r_t(x_0) = \overline{co}^* \bigcap_{\delta>0} \bigcup_i \{\partial^o r_{t_i}(x_i);\ \|x_i - x_0\| < \delta,\ \|t_i - t\| < \delta\}. \tag{2.5}$$

For any $x \in X$, the Theorem 2.8.2 and Theorem 2.7.5 of Clark [2] showed that:

Proposition 2.1. Under the above assumptions, it follows that

$$\partial^o r(x) \subset \int_T \partial^o_{[T]} r_t(x) \mu(dt). \tag{2.6}$$

That is, for every $\zeta \in \partial^o r(x)$, there exist a measurable selection: $t \in T \longrightarrow \zeta_t \in \partial^o_{[T]} r_t(x) \subset X^*$ and a probability measure μ on $M(x)$ such that $t \longrightarrow \zeta_t$ is weak* μ–integrable and for any $u \in X$,

$$< u, \zeta > = \int_T < u, \zeta_t > \mu(dt). \tag{2.7}$$

We need Ekeland's variational principle (Ekeland [3]) as follows:

Lemma 2.2. Let (V, d) be a complete metric space, and let $F: V \longrightarrow \mathbb{R}$ be lower semi–continuous function which is bounded below. Given $\epsilon > 0$ and

$x_0 \in V$, if $F(x_0) \leq \inf F(v) + \epsilon$, then there exists an $x \in V$ such that

$$d(x_0, x) \leq \sqrt{\epsilon}, \tag{2.8$_1$}$$

$$F(v) + \sqrt{\epsilon}\, d(x, v) \geq F(x) \qquad \text{for all } v \in V. \tag{2.8$_2$}$$

We will apply this Lemma to the case of Banach space $V = X$ and assume F been continuous. Thus we define a Lagrangian function associated with the optimization problem (P) by

$$L(x; \lambda, \mu, \nu, K) = < f(x), \lambda > + < g(x), \mu > + < h(x), \nu > + K\|(\lambda, \mu, \nu)\| d_E(x) \tag{2.9}$$

where $\|(\lambda, \mu, \nu)\| = \| \lambda \|_{Y*} + \| \mu \|_{Z*} + \| \nu \|_{W*}$ is the norm of $Y^* \times Z^* \times W^*$, $K > 0$ is a constant no less than the Lipschitz constant of the Lipschitz function (f, g, h) and $d_E(x) = d(x, E) = \inf\{\| x-y \|; y \in E\}$.

The following theorem is main which is a new expression for optimality in infinite dimensional case. The proof is new and is interesting by applying the Ekeland variational principle.

Theorem 2.3. Let x_0 be a weakly minimal solution of problem (P). Then, for a suitable constant $K > 0$, there exist $\lambda \in C_{Y*}$, $\mu \in C_{Z*}$, $\nu \in W^*$, not all zero, such that

$$\theta \in \partial_x^o L(x_0; \lambda, \mu, \nu, K) \tag{2.10}$$

and

$$< g(x_0), \mu > = 0 \tag{2.11}$$

where ∂_x^o stands the generalized gradient with respect to the variable $x \in X$.

Proof. For convenience, we let

$$T = \{t = (\lambda, \mu, \nu) \in C_{Y*} \times C_{Z*} \times W^*; \| t \| = 1\}. \tag{2.12}$$

Since Y, Z and W are reflexive Banach spaces, the unit sphere of $Y^* \times Z^* \times W^*$ is weak (= weak*) compact. It follows that for any $\epsilon > 0$, the function $F: X \longrightarrow \mathbb{R}$ defined by

$$F(x) = \max_{t \in T} \{< (f(x) - f(x_0) + \epsilon u, g(x), h(x)), t > \quad ; \; u \in \operatorname{int}(C_Y) \text{ and } \| u \| = 1\} \tag{2.13}$$

is well defined.

Evidently, F is Lipschitz near x_0 and $F(x_0) = \epsilon$. Indeed, since x_0 is feasible and so $g(x_0) \leq \theta$, $h(x_0) = \theta$, thus the maximum in (2.13) is attained if $t = (\lambda, \mu, \nu)$ is taken to be $\mu = \nu = 0$, $\|\lambda\| = 1$ and $< u, \lambda > = \| u \| = 1$ (Hahn–Banach Theorem) whenever $u \in \operatorname{int}(C_Y)$ with $\| u \| = 1$.

Claim that F is a positive function on E.

For if $F(x) \leq 0$, it would imply that

$$< f(x) - f(x_0) + \epsilon u, \lambda > \; \leq 0, \qquad \lambda \in C_{Y^*} \tag{2.14}$$

$$< g(x), \mu > \; \leq 0, \qquad \mu \in C_{Z^*} \tag{2.15}$$

and $$< h(x), \nu > \; \leq 0, \qquad \nu \in W^*. \tag{2.16}$$

Since $\nu \in W^*$ is arbitrary, $< h(x), \nu > = 0$, and since W^* separates the points on W, we have $h(x) = 0$. It follows that

$$f(x) - f(x_0) + \epsilon u \in (-C_Y) \text{ and } g(x) \leq \theta.$$

This shows that x is still a better solution than x_0 for problem (P) which contradicts the fact that x_0 is a weakly minimal solution of (P).

Hence x_0 satisfies

$$F(x_0) \leq \inf_{x \in E} F(x) + \epsilon. \tag{2.17}$$

Applying Ekeland variational principle (Lemma 2), there is a point $v = v_\epsilon \in X$ such that

$$\|v - x_0\| < \sqrt{\epsilon} \tag*{(2.18)$_1$}$$

$$F(x) + \sqrt{\epsilon}\|x - v\| \geq F(v) \quad \text{for all } x \in E. \qquad (2.18)_2$$

Owing f is Lipschitz, we can let K be a constant larger than the Lipschitz constant K_F of F. Actually, $K \geq K_F + \sqrt{\epsilon}$, the Lipschitz constant of the function

$$x \longrightarrow F(x) + \sqrt{\epsilon}\| x - v \| \quad \text{for } x \text{ near } v. \qquad (2.19)$$

Since the above function attains to its minimum at $x = v$, it follows from that v also minimizes the function

$$x \longrightarrow F(x) + \sqrt{\epsilon}\| x - v \| + Kd_E(x). \qquad (2.20)$$

Let $G(x) = F(x) + Kd_E(x)$. Then, by (2.9), (2.13) and (2.12), we have that

$$G(x) = \max_{(\lambda, \mu, \nu) = t \in T} \{L(x; \lambda, \mu, \nu, K) - < f(x_0), \lambda > + < \epsilon u, \lambda >\}. \quad (2.21)$$

It follows that the function (2.20) can be written by

$$x \longrightarrow G(x) + \sqrt{\epsilon}\, \| x - v \|. \qquad (2.22)$$

This function is also Lipschitz and attains to its minimum at $x = v$. Hence, for a sufficient small $\epsilon > 0$, by (1.6) we have

$$\theta \in \partial^{o} G(v) + \partial^{o}(\sqrt{\epsilon}\|\cdot\|)(\theta) \subset \partial^{o} G(v) + \sqrt{\epsilon}\, \overline{B}_{*} \qquad (2.23)$$

since $\partial^{o}(f + g)(x) \subset \partial^{o} f(x) + \partial^{o} g(x)$ for Lipschitz functions f and g and $\partial^{o}\| \ \|(\theta) \subset \overline{B}_{*}$, (see (1.8) and (1.9)) where $\overline{B}_{*}$ is the closed unit ball of X^{*}.

Claim: $\partial^{o} G(v) \subset \partial^{o}_{x} L(v, t_v, K)$. At first we show that the mapping

$$(t, z) \longrightarrow \partial^{o}_{x} L(z, t, K) \qquad (2.24)$$

is closed, that is, (2.4) holds:

$$\partial^{o}_{[T]} L(x, t, K) = \partial^{o}_{x} L(x, t, K).$$

To see this fact, for any $t_1, t_2 \in T$, the mapping

$$x \longrightarrow L(x, t_1, K) - L(x, t_2, K) = (t_1 - t_2)\cdot(f, g, h)(x)$$

is Lipschitz of rank $K'\| t_1 - t_2 \|$ near v where K' stands Lipschitz constant of the function (f, g, h). From the elementary property (1.6) of generalized gradient we have that

$$\partial_x^o[(t_1-t_2)\cdot(f, g, h)](x) \subset K'\| t_1-t_2 \|\bar{B}_* . \tag{2.25}$$

Hence as the same reason of (2.23), it follows that

$$\partial_x^o L(x, t_1, K) \subset \partial_x^o L(x, t_2, K) + K'\| t_1-t_2 \|\bar{B}_* . \tag{2.26}$$

Since the generalized gradient of a Lipschitz function is upper semicontinuous in x and by (2.3) and (2.5), we see that (2.24) is closed. Therefore the maximal function $F(v)$ (> 0) defined in (2.13) is attained at some point $t_v \in T$ so does for G. Then $G(v) = L(v, t_v, K) - < f(x_0), \lambda > + < \epsilon u, \lambda >$ has generalized gradient:

$$\partial^o G(v) \subset \partial_x^o L(v, t_v, K).$$

From (2.23),

$$\theta \in \partial_x^o L(v, t_v, K) + \sqrt{\epsilon}\,\bar{B}_* . \tag{2.27}$$

Letting $\epsilon \downarrow 0$, there corresponds a subsequence t_{v_α} convergent to an element $t = (\lambda, \mu, \nu)$ in T. In this case $v = v_\epsilon$ converges to x_0. Consequently, from the closedness of $(t, v) \longrightarrow \partial_x^o L(v, t, K)$, the expression (2.27) should imply that

$$\theta \in \partial_x^o L(x_0, t, K) \qquad \text{whenever } \epsilon \downarrow 0.$$

Note that in general, $< g(v_\epsilon), \mu(\epsilon) > \lesseqgtr 0$, but in the above construction $< g(v_\epsilon), \mu(\epsilon) >$ is nonnegative, and as $\epsilon \downarrow 0$, the limit is also 0, this proves $< g(x_0), \mu > = 0$. The proof is completed. □

3. REMARKS ON SOME ALTERNATIVE CONCLUSIONS FOR FINITE DIMENSIONAL CASES

Throughout this section, we consider that the range spaces of functions in problem (P) are finite dimensional spaces. That is, in (P), $Y = \mathbb{R}^n$, $Z = \mathbb{R}^m$ and $W = \mathbb{R}^k$. Some known results are reduced from Theorem 2.3. Hiriart–Urruty [5, Theorem 3.1] established a Fritz John type optimal condition as follows.

Theorem 3.1. ([5]). Let x_0 be a local minimum of $f: X \longrightarrow \mathbb{R}$ in problem (P) with $X = \mathbb{R}^N = E$, $Y = \mathbb{R}$. Then there exist $\lambda \in \mathbb{R}_+$, $\mu \in \mathbb{R}^m_+$ and $\nu \in \mathbb{R}^k$, not all zero, such that

$$\begin{cases} \theta \in \partial^\circ(\lambda f + \sum_{i=1}^{m} \mu_i g_i + \sum_{j=1}^{k} \nu_j h_j)(x_0) \\ \text{and } < g(x_0), \ \mu > = 0. \end{cases} \tag{3.1}$$

If $E \subsetneq X = \mathbb{R}^N$ is an additional closed subset in problem (P), then the distance function $d_E(x) = \inf_{x \in E} \| x-y \|$ is Lipschitz of rank 1 and $\theta \in \partial^\circ d_E(x)$ for every feasible point x of problem (P). Since the set $N_E(x)$ of normal cone to E at x is a closed convex cone generated by $\partial^\circ d_E(x)$, the Fritz John type condition of Theorem 3.1 turns to the following corollary.

Corollary 3.2. ([5, Corollary 3.2]). Like the assumptions in Theorem 3.1 if x_0 is a local solution of problem (P) with $E \subsetneq X$, then we have

$$\begin{cases} \theta \in \lambda \partial^\circ f(x_0) + \sum_{i=1}^{m} \mu_i \partial^\circ g_i(x_0) + \sum_{j=1}^{k} \nu_j \partial^\circ h_j(x_0) + N_E(x_0) \\ \text{and } < g(x_0), \ \mu > = 0. \end{cases} \tag{3.2}$$

If the constrained inequality $g(x) \leq \theta$ in problem (P) has Slater condition, that is, there exists a feasible point $\tilde{x}$ such that $g(\tilde{x}) < \theta$, then the Kuhn–Tucker optimality condition holds, that is, $\lambda = 1$. This result is established in Theorem 4.1 of Hiriart–Urruty [5].

The above results hold for X being a real Banach space(cf. Clarke [2]). In the following part, we always assume that X is a real Banach space.

In problem (P) a feasible point x_0 is a *Pareto* (resp. *weakly Pareto*) optimal solution if there is no other feasible point x such that

$$\begin{cases} f_j(x) \leq f_j(x_0) \quad (\text{resp. } f_j(x) < f_j(x_0)), \ j = 1, 2, \cdots, n \\ \text{with } f_i(x) \neq f_i(x_0) \quad \text{for at least one } i. \end{cases} \tag{3.3}$$

Note that this definition is different from Clarke [2, P. 230]. In Clarke [2], the Pareto optimum is just the weakly Pareto optimum in our sense. Evidently, any weakly Pareto optimal solution is also a Pareto optimal solution for multiobjective programming problem, and they coincide with the usual optimal solution in the case of single valued objective function.

Lemma 3.1 of Kanniapan [6] is established in the case of convex functions. It holds also in the case of local Lipschitz functions.

Lemma 3.3. A feasible solution x_0 of problem (P) is a Pareto optimal solution if and only if x_0 solves the subproblem (P_i), $i = 1, 2, \cdots, n$ defined by

$$(P_i) \qquad \begin{cases} \text{minimize} \ f_i(x) \\ \text{subject to } x \in F_i, \end{cases}$$

where $$F_i = \{x \in F;\ f_j(x) \leq f_j(x_0) \text{ for } j = 1, \cdots, n,\ j \neq i\} \tag{3.4}$$

and $$F = \{x \in X;\ g(x) \leq \theta, h(x) = \theta\} \cap E \tag{3.5}$$

is the feasible set of (P).

Proof. Observe that if for some $i \in \{1, \cdots, n\}$, x_0 does not minimize f_i for problem (P_i), then there is a feasible solution x for problem (P) such that

$$f_j(x) \leq f_j(x_0) \quad \text{for } j \neq i, \quad h(x) = \theta, \quad g(x) \leq \theta \quad \text{and} \quad f_i(x) < f_i(x_0).$$

This shows that x_0 is not a Pareto optimal solution of (P). Conversely, if x_0 does not solve problem (P) , then we can find a feasible solution $x \neq x_0$ such that $f(x) \leq f(x_0)$ with at least one component, say f_i, such that $f_i(x) < f_i(x_0)$. This shows that x_0 does not minimize f_i over F_i. □

Employing Lemma 3.3, we can easily extended the Theorem 6.1.3 of Clarke [2] from weakly Pareto optimum to Pareto optimum as follows.

Theorem 3.4. Let x_0 be a Pareto optimal solution of problem (P) with $X = E$, a Banach space. Then there exist $\lambda \in \mathbb{R}^n_+$ and $\nu \in \mathbb{R}^k$, not all zero, such that

$$\begin{cases} \theta \in \partial^\circ(< f,\lambda > + < g, \mu > + < h, \nu >)(x_0) \\ < g(x_0) , \mu > = 0. \end{cases} \tag{3.6}$$

Proof. By Lemma 3.3, x_0 is also an optimal solution of problem (P_i) for some i. Applying Theorem 3.1 to the single–valued objective function f_i with inequality constraints:

$$g(x) \leq \theta \quad \text{and} \quad f_j(x) - f_j(x_0) \leq 0 \qquad \text{for } j = 1, 2, \cdots, n, \quad j \neq i,$$

we have that there exist $\lambda_j \geq 0$, $j = 1, 2, \cdots, n$, $j \neq i$ and $\lambda_i > 0$, $\mu \in \mathbb{R}^m_+$, $\nu \in \mathbb{R}^k$ not all zero such that $< g(x_0), \mu > = 0$ and

$$\theta \in \partial^{\circ}(\lambda_i f_i + \sum_{j\neq i} \lambda_j(f_j - f_j(x_0)) + \langle g, \mu \rangle + \langle h, \nu \rangle)(x_0)$$
$$\subset \partial^{\circ}(\langle \lambda, f \rangle + \langle g, \mu \rangle + \langle h, \nu \rangle)(x_0)$$

since the generalized gradient of a constant function is zero, that is, $\partial^{\circ}(\lambda_j f_j(x_0))(x_0) = \{0\}$ for all j. This proves the theorem. □

Like (3.2), the expression (3.6) can be written in "separated" form

$$\begin{cases} \theta \in \sum_{i=1}^{m} \lambda_i \partial^{\circ} f_i(x_0) + \sum_{j=1}^{m} \mu_j \partial^{\circ} g_j(x_0) + \sum_{\ell=1}^{k} \nu_\ell \partial^{\circ} h_\ell(x_0) \\ \text{and} \quad \sum_{j=1}^{m} \mu_j g_j(x_0) = 0. \end{cases} \tag{3.6)'$$

If $E \subsetneq X$, x_0 is a weakly Pareto optimal solution of problem (P), then the stationary condition (3.6) of Theorem 3.4 should be (a corollary of Theorem 2.3)

$$\theta \in \partial^{\circ}_{x}(\langle f, \lambda \rangle + \langle g, \mu \rangle + \langle h, \nu \rangle + \tau d_E(\cdot))(x_0)$$
$$\subset \sum_{i=1}^{n} \lambda_i \partial^{\circ} f_i(x_0) + \sum_{j=1}^{m} \mu_j \partial^{\circ} g_j(x_0) + \sum_{\ell=1}^{k} \nu_\ell \partial^{\circ} h_\ell(x_0) + N_E(x_0). \tag{3.7}$$

Here $\tau > 0$ is some constant and $|\lambda| + |\mu| + |\nu| = 1$. This result was established by Minami [10, Theorem 3.1] (cf. also Clark [2, theorem 6.1.1]).

Recently, Lai and Ho [8, theorem 3.1] established Fritz John type optimal condition for convex programming problem in matrix forms. We now can apply the stationary condition (3.7) to establish a similar type for generalized gradient in matrix forms which we state as follows.

Theorem 3.5. Let x_0 be a weakly Pareto optimal solution (equivalently that x_0 is a Pareto optimun and there is a feasible point $\tilde{x}$ such that $g(\tilde{x}) < \theta$) for problem (P) with $E \subsetneq X$. Then there exist matrices

$$\begin{array}{llll} \lambda = (\lambda_{ij})_{n\times n}, & \lambda_{ij} \geq 0, & \lambda_{ij} = 0 & \text{for } i \neq j \\ \mu = (\mu_{ij})_{n\times m}, & \mu_{ij} \geq 0 & & \\ \nu = (\nu_{ij})_{n\times k} & \nu_{ij} \in \mathbb{R} & & \end{array}$$

not all zero matrices such that

$$\begin{cases} \theta \in \lambda\partial^{o} f(x_0) + \mu\partial^{o} g(x_0) + \nu\partial^{o} h(x_0) + \tilde{N}_E(x_0) \\ \text{and} \quad \mu\cdot g(x_0) = \theta, \end{cases}$$

here, $\partial^{o} f(x_0) = (\partial^{o} f_1(x_0), \cdots, \partial^{o} f_n(x_0))^{T}$ is an n column vector,

$\partial^{o} g(x_0)$ and $\partial^{o} h(x_0)$ are, respectively, m and k column vectors;

$\tilde{N}_E(x_0)$ is n–tuple of $N_E(x_0)$ in $(X^*)^n$;

and $N_E(x_0)$ is the normal cone difined in (3.7).

Remark. The stationary condition $\theta \in \partial^{o} L$ of (2.10) in Theorem 2.3 can not imply the "separated" condition like (3.7) did.

References

[1] Clarke, F.H., A new approach to Lagrange multipliers, *Math. Oper. Res.* 1(1976), 165–174.

[2] Clarke, F.H., *Optimization and Nonsmooth Analysis*, A Wiley–Interscience Publication, 1983.

[3] Ekeland, I., On the variational principles, *J. Math. Anal. Appl.* 47(1974), 324–353.

[4] Hiriart–Urruty, J.B., Optimality conditions in nondifferentiable programming, *Math. Program.* 14(1978), 73–96.

[5] Hiriart–Urruty, J.B. Refinements of necessary optimality conditions in nondifferentiable programming I, *Appl. Math. Optim.* 5(1979), 63–82.

[6] Kanniappan, P., necessary conditions for optimality of nondifferentiable convex multiobjective programming *J. Optim. Theory Appl.* 40(1983), 167–174.

[7] Kanniappan, P. and Sastry, S.M.A., Duality theorems and an optimality condition for nondifferentiable convex programming *J. Aust. Math. Soc. (Series A)*, 32(1982), 369–379.

[8] Lai, H.C. and Ho, C.P., Duality theorem of nondifferentiable convex multiobjective programming *J. Optim. Theory Appl.* 50 (1986), 407–420.

[9] Lai, H.C. and Yang, L.S., Strong duality for infinite–dimensional vector–valued programming problems, *J. Optim. Theory Appl.* 62(1989), 447–464.

[10] Lai, H.C. and Lin, L.J., Optimality for set functions with values in ordered vector spaces, *J. Optim. Theory Appl.*, 63(1989), 405–423.

[11] Lai, H.C., Optimization theory for infinite dimensional set functions, *Approximation, Optimization and Computing: Theory and Applications,* A.C. Law and C.L. Wang (eds.) Elsevier Science Publ. B.V. (North Holland), IMACS (1990), 13–21.

[12] Minami, M., Weak pareto–optimal necessary conditions in nondifferential multiobjective program on a Banach space, *J. Optim. Theory Appl.* 41(1983), 451–461.

[13] Yu, P.L., Cone convexity, cone extreme points and nondominated solution in decision problem with multiobjective, *J. Optim. Theory Appl.*, 14(1974), 319–377.

[14] Zowe, J., A duality theorem for a convex programming problem in order complete vector lattices, *J. Math. Anal. Appl.*, 50(1975), 273–287.

[15] Zowe, J., The saddle point theorem of Kuhn and Tucker in order vector spaces, *J. Math. Anal. Appl.*, 57(1977), 47–55.

Department of Mathematics
University of Cape Town
South Aftica

Received February 22, 1991