

HIGHER-ORDER VARIATIONAL SETS, VARIATIONAL DERIVATIVES AND HIGHER-ORDER NECESSARY CONDITIONS IN ABSTRACT MATHEMATICAL PROGRAMMING

Furukawa, Nagata
Department of Mathematics, Faculty of Science, Kyushu University

Yoshinaga, Yuji
Kaho High School

<https://doi.org/10.5109/13392>

出版情報 : Bulletin of informatics and cybernetics. 23 (1/2), pp.9-40, 1988-03. Research
Association of Statistical Sciences

バージョン :

権利関係 :



HIGHER-ORDER VARIATIONAL SETS, VARIATIONAL DERIVATIVES AND HIGHER-ORDER NECESSARY CONDITIONS IN ABSTRACT MATHEMATICAL PROGRAMMING

By

Nagata FURUKAWA* and Yuji YOSHINAGA**

Abstract

We consider the problem of minimizing the value of a functional $f(x)$ under the constraints $g(x) \in B$, $x \in Q$, where g is a mapping from a Banach space to a Banach space. We assume neither the differentiability in usual sense nor the convexity on f and g . The set B is an arbitrary subset of a Banach space which has a nonempty interior. A higher-order variational theory for the optimization problem is developed. Several kinds of concepts of higher-order variational sets are introduced in a Banach space, and concepts of higher-order variational derivatives of mappings are introduced too. We study fundamental properties of those variational sets and variational derivatives. We also give higher-order necessary optimality conditions for the problem in terms of the variational sets and derivatives. We apply the results of optimality condition to derive a second-order criterion of an optimal solution in a nonlinear Tchebycheff approximation problem.

1. Introduction

In a standard theory of non-linear programming, key theorem to derive the first-order necessary conditions for a local optimal solution is the separation of infinitesimal cones describing some geometrical features of the solution. A separation theorem of such infinitesimal cones has been initially given by Dubovitskii and Milyutin [3].

In order that the same discussion as above may possibly derive higher-order necessary conditions, we need to introduce the concepts of higher-order variational sets and higher-order variational derivatives. In this paper three kinds of concepts of higher-order variational sets are defined. Two kinds of those concepts are the same that Hoffmann and Kornstaedt originally introduced in [6]. We first study systematically the variational sets and prove fundamental properties. And then we introduce three kinds of concepts of variational derivatives of functionals. In [10] Neustadt introduced a certain concept of first-order variational derivative. We extend his derivative to the higher-order derivatives, which we call higher-order Neustadt derivatives.

* Department of Mathematics, Faculty of Science, Kyushu University, Fukuoka, Japan

** Kaho High School, Iizuka, Japan

Let X and Y be real Banach spaces, Q an arbitrary subset of X , and B a subset of Y with nonempty interior $\text{int } B$. Let $f: X \rightarrow R$, the set of real numbers, and $g: X \rightarrow Y$. Then the mathematical programming problem with which we are concerned is expressed by the following form:

$$\left. \begin{array}{l} \text{minimize } f(x) \\ \text{subject to} \\ g(x) \in B, \ x \in Q. \end{array} \right\} \quad (1.1)$$

In a usual mathematical programming problem, the set B assumes a convex cone. On the contrary, in our paper B is arbitrary as far as $\text{int } B \neq \emptyset$. Moreover we assume neither differentiability nor convexity on the functionals involved in the optimization problem. Assuming the existence of higher-order Neustadt derivatives of the functionals, we derive higher-order necessary optimality conditions in a general abstract form using variational sets and Neustadt derivatives.

An important problem to which our theorem of higher-order necessary conditions can be applied is the problem of the best approximation, especially the nonlinear Tchebycheff approximation problem. In that problem the set B in (1.1) is usually taken as the unit sphere in the space of continuous functions. In this paper nonlinear Tchebycheff approximation problem is discussed as an application of the necessary condition and then the second-order criteria for the problem are given.

2. Higher-order Variational Sets

The concepts which play key roles in developing a higher-order theory of optimization are variational sets and variational derivatives of higher-order. In this section we give the definitions of three kinds of variational sets of order n and investigate some fundamental properties of them.

Let X be a real Banach space, and let Q an arbitrary nonempty subset of X .

DEFINITION 2.1. For $x_0 \in \text{bd } Q$ and for $x_1, x_2, \dots, x_{n-1} \in X$, three kinds of variational sets of Q at x_0 (with respect to the points x_1, x_2, \dots, x_{n-1}) are defined by the following:

$$F(Q; x_0, x_1, \dots, x_{n-1}) = \left\{ h \in X \left| \begin{array}{l} \exists \text{bd } V \text{ of } h, \exists t_0 > 0 \\ \text{such that} \\ \sum_{i=0}^{n-1} t^i x_i + t^n V \subset Q \text{ for } \forall t \in (0, t_0] \end{array} \right. \right\}, \quad (2.1)$$

$$V(Q; x_0, x_1, \dots, x_{n-1}) = \left\{ h \in X \left| \begin{array}{l} \exists t_0 > 0, \exists r: (0, t_0] \rightarrow X \\ \text{such that} \\ \sum_{i=0}^{n-1} t_i x_i + t^n h + r(t) \in Q \text{ for } \forall t \in (0, t_0] \\ \|r(t)\|/t^n \rightarrow 0 \text{ as } t \downarrow 0 \end{array} \right. \right\}, \quad (2.2)$$

$$T(Q; x_0, x_1, \dots, x_{n-1}) = \left\{ h \in X \left| \begin{array}{l} \exists \{y_k\} \subset Q, \exists \{\lambda_k\} \downarrow 0 \\ \text{such that} \\ (y_k - \sum_{i=0}^{n-1} \lambda_k^i x_i) / \lambda_k^n \rightarrow h \text{ as } k \rightarrow \infty \end{array} \right. \right\}, \quad (2.3)$$

For $x_0 \in \text{int } Q$ we define

$$F(Q; x_0, x_1, \dots, x_{n-1}) = V(Q; x_0, x_1, \dots, x_{n-1}) = T(Q; x_0, x_1, \dots, x_{n-1}) = X,$$

and for $x_0 \in \text{cl } Q$ we define

$$F(Q; x_0, x_1, \dots, x_{n-1}) = V(Q; x_0, x_1, \dots, x_{n-1}) = T(Q; x_0, x_1, \dots, x_{n-1}) = \emptyset.$$

In the case $n=1$, the variational sets defined above coincide with the well-known infinitesimal cones: $F(Q; x_0)$ is the cone of feasible directions ([5]), $V(Q; x_0)$ the cone of tangent directions ([5]), and $T(Q; x_0)$ is the cone of adherent displacements ([18]). Hence the variational sets in the case $n \geq 2$ can be considered as the higher-order extension of the infinitesimal cones we have used in the first-order theory. Hoffman and Kornstaedt [6] has defined two kinds of variational sets of higher-order in a topological vector space. Two of our variational sets are essentially same as theirs, that is, in a normed vector space $F(Q; x_0, x_1, \dots, x_{n-1})$ is same as the variational set $K(Q; x_0, x_1, \dots, x_{n-1})$ and $T(Q; x_0, x_1, \dots, x_{n-1})$ same as the set $K[Q; x_0, x_1, \dots, x_{n-1}]$ in their notation. Ben-Tal and Zowe [2] has defined a second-order tangent direction of a continuous mapping ξ . The set of the second-order tangent directions of ξ in their sense coincides with $V(N[\xi]; x_0, x_1)$ in our notation, where $N[\xi]$ is the kernel of the mapping ξ .

PROPOSITION 2.1. *Let $x_0 \in \text{bd } Q$. Then*

- (i) $F(Q; x_0)$, $V(Q; x_0)$ and $T(Q; x_0)$ are all cones with vertices at the origin.
- (ii) $F(Q; x_0, 0, 0, \dots, 0) = F(Q; x_0)$, $V(Q; x_0, 0, 0, \dots, 0) = V(Q; x_0)$ and $T(Q; x_0, 0, 0, \dots, 0) = T(Q; x_0)$.
- (iii) $F(Q; x_0, x_1, \dots, x_{n-1})$ is open, and $T(Q; x_0, x_1, \dots, x_{n-1})$ is closed.
- (iv) $V(Q; x_0, x_1, \dots, x_{n-1})$ is closed.

PROOF. (i) and (ii) are obvious. The former half of (iii) is also obvious. The latter of (iii) is easily proved: in this connection we can refer to the proof of Krabs [7] in the case $n=1$. Concerning (iv), a direct proof by the definition of the variational set may be possible. But we will prove (iv) by another approach using a representation theorem of the variational set in Section 4. \square

PROPOSITION 2.2. *Let $x_0 \in \text{bd } Q$. Then*

- (i) *If $h \in F(Q; x_0, x_1, \dots, x_{n-1})$, then there exists a number $t_0 > 0$ such that $\sum_{i=0}^{n-1} t^i x_i + t^n h \in \text{int } Q$ for all $t \in (0, t_0]$.*
- (ii) $F(Q; x_0, x_1, \dots, x_{n-1}) \subset V(Q; x_0, x_1, \dots, x_{n-1})$
 $\subset T(Q; x_0, x_1, \dots, x_{n-1})$.
- (iii) *If $x_n \in F(Q; x_0, x_1, \dots, x_{n-1})$, then $F(Q; x_0, x_1, \dots, x_{n-1}, x_n) = X$.*
- (iv) *If $x_n \in V(Q; x_0, x_1, \dots, x_{n-1})$, then $F(Q; x_0, x_1, \dots, x_{n-1}, x_n) = V(Q; x_0, x_1, \dots, x_{n-1}, x_n) = \emptyset$.*
- (v) *If $x_n \in T(Q; x_0, x_1, \dots, x_{n-1})$, then $T(Q; x_0, x_1, \dots, x_{n-1}, x_n) = \emptyset$.*

PROOF. (i) is obvious.

(ii) Let $h \in F(Q; x_0, x_1, \dots, x_{n-1})$. By (i) then we have a positive number t_0 such that $\sum_{i=0}^{n-1} t^i x_i + t^n h \in \text{int } Q$ for all $t \in (0, t_0]$. Hence it can be shown that

$h \in V(Q; x_0, x_1, \dots, x_{n-1})$ by putting $r \equiv 0$ in the definition of $V(Q; x_0, x_1, \dots, x_{n-1})$.

Next let $h \in V(Q; x_0, x_1, \dots, x_{n-1})$. Then from the definition there exist $t_0 > 0$ and $r: (0, t_0] \rightarrow X$ such that

$$\sum_{i=0}^{n-1} t^i x_i + t^n h + r(t) \in Q \quad \forall t \in (0, t_0], \quad (2.4)$$

$$\|r(t)\|/t^n \rightarrow 0 \quad \text{as } t \downarrow 0. \quad (2.5)$$

Choose a positive integer $N > 1/t_0$, and put

$$y_k = \sum_{i=0}^{n-1} \frac{1}{k^i} x_i + \frac{1}{k^n} h + r\left(\frac{1}{k}\right) \quad \text{for } k \geq N.$$

Then (2.4) implies that $\{y_k\} \subset Q$. By taking $\lambda_k = 1/k$, we get from (2.5) that

$$\left(y_k - \sum_{i=0}^{n-1} \lambda_k^i x_i\right) / \lambda_k^n \rightarrow h \quad \text{as } k \rightarrow \infty.$$

Thus we have $h \in T(Q; x_0, x_1, \dots, x_{n-1})$.

(iii) The proof is given in Section 4.

(iv) Let $x_n \notin V(Q; x_0, x_1, \dots, x_{n-1})$. Suppose that there exists a point $h \in F(Q; x_0, x_1, \dots, x_{n-1}, x_n)$. By (i) then we have a positive number t_0 such that $\sum_{i=0}^n t^i x_i + t^{n+1} h \in \text{int } Q$ for all $t \in (0, t_0]$. Hence, putting $r(t) = t^{n+1} h$, we have

$$\sum_{i=0}^n t^i x_i + r(t) \in Q \quad \forall t \in (0, t_0],$$

$$\|r(t)\|/t^n \rightarrow 0 \quad \text{as } t \downarrow 0,$$

which imply that $x_n \in V(Q; x_0, x_1, \dots, x_{n-1})$. This is a contradiction.

Next suppose that there exists a point $h \in V(Q; x_0, x_1, \dots, x_{n-1}, x_n)$. Then there exist $t_0 > 0$ and $r: (0, t_0] \rightarrow X$ such that

$$\sum_{i=0}^n t^i x_i + t^{n+1} h + r(t) \in Q \quad \forall t \in (0, t_0], \quad (2.6)$$

$$\|r(t)\|/t^{n+1} \rightarrow 0 \quad \text{as } t \downarrow 0. \quad (2.7)$$

Putting $\hat{r}(t) = t^{n+1} h + r(t)$, we get from (2.6) and (2.7) that

$$\sum_{i=0}^n t^i x_i + \hat{r}(t) \in Q \quad \forall t \in (0, t_0],$$

$$\|\hat{r}(t)\|/t^n \rightarrow 0 \quad \text{as } t \downarrow 0.$$

Thus we have $x_n \in V(Q; x_0, x_1, \dots, x_{n-1})$, which is a contradiction.

(v) Let $x_n \notin T(Q; x_0, x_1, \dots, x_{n-1})$. Suppose that there exists $h \in T(Q; x_0, x_1, \dots, x_{n-1}, x_n)$. Then from the definition we have a sequence $\{y_k\}$ of points of Q and a sequence $\{\lambda_k\}$ of positive numbers such that

$$\lambda_k \downarrow 0 \quad \text{as } k \rightarrow \infty, \quad (2.8)$$

$$\left(y_k - \sum_{i=0}^n \lambda_k^i x_i\right) / \lambda_k^{n+1} \rightarrow h \quad \text{as } k \rightarrow \infty. \quad (2.9)$$

(2.9) together with (2.8) implies that

$$\left(y_k - \sum_{i=0}^n \lambda_k^i x_i\right) / \lambda_k^n \rightarrow 0 \quad \text{as } k \rightarrow \infty,$$

which is equivalent to that

$$\left(y_k - \sum_{i=0}^{n-1} \lambda_k^i x_i\right) / \lambda_k^n \rightarrow x_n \quad \text{as } k \rightarrow \infty.$$

Thus $x_n \in T(Q; x_0, x_1, \dots, x_{n-1})$, which is a contradiction. \square

In the case when Q is convex, especially we have the following

PROPOSITION 2.3. *Let Q be a convex subset of X . Let $x_0 \in Q$ and $y \in Q$. Then*

$$Q - y \subset T(Q; x_0, y - x_0).$$

PROOF. Let h be an arbitrary point of Q . Let

$$x_n = \frac{1}{n^2} h + \frac{1}{n} \left(1 - \frac{1}{n}\right) y + \left(1 - \frac{1}{n}\right) x_0 \quad \text{for } n \geq 1.$$

Since x_n is a convex combination of points of Q , by the convexity of Q we have $x_n \in Q$ for all n . A simple calculation leads to

$$n^2 \left[x_n - x_0 - \frac{1}{n} (y - x_0) \right] = h - y \quad \text{for } n \geq 1,$$

which shows trivially that $h - y \in T(Q; x_0, y - x_0)$. Since the point h of Q is arbitrary, the desired inclusion relation follows. \square

3. Higher-order Variational Derivatives

In the preceding section we have introduced the concept of higher-order variational sets. In this section we shall give several kinds of higher-order variational derivatives of mappings. Among these derivatives higher-order directional derivatives and Dini derivatives are included and some useful properties of later one are shown.

Let X, Y be real Banach spaces. Let U be an open subset of X . Then we begin with the following definition.

DEFINITION 3.1. Let $f: U \rightarrow Y$ be a mapping, let $x_0 \in U$ and let x_1, x_2, \dots, x_n be points of X . If there exists a point $f'(x_0; x_1)$ of Y such that

$$f'(x_0; x_1) = \lim_{\lambda \downarrow 0} \frac{1}{\lambda} [f(x_0 + \lambda x_1) - f(x_0)], \quad (3.1)$$

then we call the point $f'(x_0; x_1)$ the first-order directional derivative of f in the direction x_1 at x_0 . If the mapping f has the derivative in every direction at x_0 , then we call the mapping $f'(x_0; \cdot)$ the first-order directional derivative of f at x_0 . If there

exists the directional derivative $f'(x_0; x_1)$ and, if there exists a point $f''(x_0, x_1; x_2)$ of Y such that

$$f''(x_0, x_1; x_2) = \lim_{\lambda \downarrow 0} \frac{1}{\lambda^2} [f(x_0 + \lambda x_1 + \lambda^2 x_2) - f(x_0) - \lambda f'(x_0; x_1)], \quad (3.2)$$

then we call the point the second-order directional derivative of f in the direction x_2 at x_0 (with respect to x_1). If f has the derivative $f''(x_0, x_1; x_2)$ for every $x_2 \in X$, then we call the mapping $f''(x_0, x_1; \cdot)$ the second-order directional derivative of f at x_0 (with respect to x_1). Inductively, the n -th order directional derivative of f at x_0 is defined by the following formula:

$$\begin{aligned} & f''^{\overbrace{\dots}^n}'(x_0, x_1, \dots, x_{n-1}; x_n) \\ &= \lim_{\lambda \downarrow 0} \frac{1}{\lambda^n} \left[f\left(\sum_{i=0}^n \lambda^i x_i\right) - f(x_0) - \sum_{i=1}^{n-1} \lambda^i f''^{\overbrace{\dots}^{i-1}}(x_0, x_1, \dots, x_{i-1}; x_i) \right]. \end{aligned}$$

The directional derivatives defined above do not operate directly on the derivation of higher-order optimality conditions in general theory of this paper, but they are necessary for calculating variational sets in concrete problems as will be seen in Section 7. For the purpose of the derivation of our optimality conditions the following is inevitable.

DEFINITION 3.2. Let $f: U \rightarrow Y$, and let $x_0 \in U$. If for every point x_1 of X there exists a point $f^{(1)}(x_0; x_1)$ of Y such that

$$f^{(1)}(x_0; x_1) = \lim_{\substack{y \rightarrow x_0 \\ \lambda \downarrow 0}} \frac{1}{\lambda} [f(x_0 + \lambda y) - f(x_0)], \quad (3.3)$$

then we call the mapping $f^{(1)}(x_0; \cdot)$ the first-order Neustadt derivative of f at x_0 . Let x_1 be a point of X . If f has the first-order Neustadt derivative $f^{(1)}(x_0; \cdot)$ at x_0 , and if for every point x_2 of X there exists a point $f^{(2)}(x_0, x_1; x_2)$ of Y such that

$$f^{(2)}(x_0, x_1; x_2) = \lim_{\substack{y \rightarrow x_0 \\ \lambda \downarrow 0}} \frac{1}{\lambda^2} [f(x_0 + \lambda x_1 + \lambda^2 y) - f(x_0) - \lambda f^{(1)}(x_0; x_1)], \quad (3.4)$$

then we call the mapping $f^{(2)}(x_0, x_1; \cdot)$ the second-order Neustadt derivative of f at x_0 (with respect to x_1). Inductively, if f has the derivatives $f^{(1)}(x_0; \cdot)$, $f^{(2)}(x_0, x_1; \cdot)$, \dots , $f^{(n-1)}(x_0, x_1, \dots, x_{n-2}; \cdot)$ up to the $(n-1)$ -th order, and if for every point x_n of X there exists a point $f^{(n)}(x_0, x_1, \dots, x_{n-1}; x_n)$ of Y such that

$$\begin{aligned} & f^{(n)}(x_0, x_1, \dots, x_{n-1}; x_n) \\ &= \lim_{\substack{y \rightarrow x_0 \\ \lambda \downarrow 0}} \frac{1}{\lambda^n} \left[f\left(\sum_{i=0}^n \lambda^i x_i + \lambda^n y\right) - f(x_0) - \sum_{i=1}^{n-1} \lambda^i f^{(i)}(x_0, x_1, \dots, x_{i-1}; x_i) \right], \quad (3.5) \end{aligned}$$

then we call the mapping $f^{(n)}(x_0, x_1, \dots, x_{n-1}; \cdot)$ the n -th order Neustadt derivative of f at x_0 (with respect to x_1, x_2, \dots, x_{n-1}). When f has the Neustadt derivative of order n at x_0 with respect to $(x_1, x_2, \dots, x_{n-1})$, f is said to be n -times Neustadt differentiable at x_0 with respect to $(x_1, x_2, \dots, x_{n-1})$. When f is n -times Neustadt differentiable at x_0 w.r.t. $(x_1, x_2, \dots, x_{n-1})$ for every points x_1, x_2, \dots, x_{n-1} of X , f is

said to be n -times Neustadt differentiable at x_0 .

The following concepts of derivatives are necessary for the representation of the variational sets.

DEFINITION 3.3. Let $f:U \rightarrow R$, $x_0 \in U$ and let x_1, x_2, \dots, x_n be points of X . We define for each $n \geq 1$

$$\bar{f}^{(n)}(x_0, x_1, \dots, x_{n-1}; x_n) = \limsup_{\lambda \downarrow 0} \frac{1}{\lambda^n} \left[f\left(\sum_{i=0}^n \lambda^i x_i\right) - f(x_0) \right], \quad (3.6)$$

$$\underline{f}^{(n)}(x_0, x_1, \dots, x_{n-1}; x_n) = \liminf_{\lambda \downarrow 0} \frac{1}{\lambda^n} \left[f\left(\sum_{i=0}^n \lambda^i x_i\right) - f(x_0) \right], \quad (3.7)$$

where the limit supremum and the limit infimum are allowed to take the values $+\infty$ and $-\infty$. We call $\bar{f}^{(n)}(x_0, x_1, \dots, x_{n-1}; \cdot)$ the n -th order upper Dini derivative of f at x_0 (with respect to x_1, x_2, \dots, x_{n-1}), and call $\underline{f}^{(n)}(x_0, x_1, \dots, x_{n-1}; \cdot)$ the n -th order lower Dini derivative of f at x_0 (with respect to x_1, x_2, \dots, x_{n-1}).

PROPOSITION 3.1. *It holds that*

$$\underline{f}^{(n)}(x_0, x_1, \dots, x_{n-1}; \cdot) = -[(\overline{-f})^{(n)}(x_0, x_1, \dots, x_{n-1}; \cdot)], \quad (3.8)$$

where $(\overline{-f})^{(n)}(x_0, x_1, \dots, x_{n-1}; \cdot)$ means the n -th order upper Dini derivative of the function $(-f)$ at x_0 .

PROPOSITION 3.2. *Let $f:U \rightarrow R$ be locally Lipschitz. Then*

- (i) *Both $\bar{f}^{(1)}(x_0; \cdot)$ and $\underline{f}^{(1)}(x_0; \cdot)$ are real-valued Lipschitz functions.*
- (ii) *For each $n \geq 2$, either one and only one of the following necessarily occurs:*

- (a) $\bar{f}^{(n)}(x_0, x_1, \dots, x_{n-1}; \cdot) \equiv +\infty$.
- (b) $\bar{f}^{(n)}(x_0, x_1, \dots, x_{n-1}; \cdot) \equiv -\infty$.
- (c) $\bar{f}^{(n)}(x_0, x_1, \dots, x_{n-1}; \cdot)$ is a real-valued Lipschitz function.

- (iii) *For $\underline{f}^{(n)}(x_0, x_1, \dots, x_{n-1}; \cdot)$, the same thing that for $\bar{f}^{(n)}(x_0, x_1, \dots, x_{n-1}; \cdot)$ in (ii) holds.*

PROOF. (i) By the assumption that f is locally Lipschitz, there exist a neighborhood V of x_0 and a constant $K > 0$ such that $|f(x) - f(y)| \leq K\|x - y\|$ for all $x, y \in V$. Hence

$$|\bar{f}^{(1)}(x_0; h)| = \left| \limsup_{\lambda \downarrow 0} \frac{1}{\lambda} [f(x_0 + \lambda h) - f(x_0)] \right| \leq K\|h\|$$

for $\forall h \in X$,

which implies that $\bar{f}^{(1)}(x_0; \cdot)$ is real-valued. The relation,

$$\begin{aligned} |\bar{f}^{(1)}(x_0; h_1) - \bar{f}^{(1)}(x_0; h_2)| &\leq \limsup_{\lambda \downarrow 0} \frac{1}{\lambda} |f(x_0 + \lambda h_1) - f(x_0 + \lambda h_2)| \\ &\leq K\|h_1 - h_2\| \quad \forall h_1, \forall h_2 \in X, \end{aligned}$$

implies that $\bar{f}^{(1)}(x_0; \cdot)$ is of Lipschitz.

The assertion for $\underline{f}^{(1)}(x_0; \cdot)$ follows from Proposition 3.1 and the results for $\bar{f}^{(1)}(x_0; \cdot)$.

(ii) First prepare the trivial relation

$$\begin{aligned} & \bar{f}^{(n)}(x_0, x_1, \dots, x_{n-1}; x_n) \\ &= \limsup_{\lambda \downarrow 0} \left\{ \frac{1}{\lambda^n} \left[f\left(\sum_{i=0}^n \lambda^i x_i\right) - f\left(\sum_{i=0}^{n-1} \lambda^i x_i\right) + \frac{1}{\lambda^n} \left[f\left(\sum_{i=0}^{n-1} \lambda^i x_i\right) - f(x_0) \right] \right] \right\}. \end{aligned} \quad (3.9)$$

The local Lipschitz property of f yields that for sufficiently small $\lambda > 0$

$$\frac{1}{\lambda^n} \left| f\left(\sum_{i=0}^n \lambda^i x_i\right) - f\left(\sum_{i=0}^{n-1} \lambda^i x_i\right) \right| \leq K \|x_n\|, \quad (3.10)$$

where K is a local Lipschitz coefficient of f at x_0 .

Let

$$\alpha = \limsup_{\lambda \downarrow 0} \frac{1}{\lambda^n} \left[f\left(\sum_{i=0}^{n-1} \lambda^i x_i\right) - f(x_0) \right].$$

In the case when $\alpha = +\infty$, from (3.9) and (3.10), for any $x_n \in X$ we have

$$\begin{aligned} & \bar{f}^{(n)}(x_0, x_1, \dots, x_{n-1}; x_n) \\ & \geq \limsup_{\lambda \downarrow 0} \left\{ -K \|x_n\| + \frac{1}{\lambda^n} \left[f\left(\sum_{i=0}^{n-1} \lambda^i x_i\right) - f(x_0) \right] \right\} \\ & = +\infty, \end{aligned}$$

which implies that (a) occurs. Similarly, if $\alpha = -\infty$ then (b) occurs. In the case when α is finite, by (3.9) and (3.10) the function $\bar{f}^{(n)}(x_0, x_1, \dots, x_{n-1}; \cdot)$ is real-valued and the following relation holds:

$$\begin{aligned} & |\bar{f}^{(n)}(x_0, x_1, \dots, x_{n-1}; h_1) - \bar{f}^{(n)}(x_0, x_1, \dots, x_{n-1}; h_2)| \\ & \leq \limsup_{\lambda \downarrow 0} \frac{1}{\lambda^n} \left| f\left(\sum_{i=0}^{n-1} \lambda^i x_i + \lambda^n h_1\right) - f\left(\sum_{i=0}^{n-1} \lambda^i x_i + \lambda^n h_2\right) \right| \\ & \leq K \|h_1 - h_2\|. \end{aligned}$$

Hence $\bar{f}^{(n)}(x_0, x_1, \dots, x_{n-1}; \cdot)$ is of Lipschitz. Thus (c) occurs.

(iii) Obvious from Proposition 3.1 and (ii). □

PROPOSITION 3.3. *Let $f: U \rightarrow R$ be locally Lipschitz. Then*

- (i) *If $0 < \bar{f}^{(n-1)}(x_0, x_1, \dots, x_{n-2}; x_{n-1}) \leq +\infty$, then $\bar{f}^{(n)}(x_0, x_1, \dots, x_{n-1}; \cdot) \equiv +\infty$.*
- (ii) *If $0 > \bar{f}^{(n-1)}(x_0, x_1, \dots, x_{n-2}; x_{n-1}) \geq -\infty$, then $\bar{f}^{(n)}(x_0, x_1, \dots, x_{n-1}; \cdot) \equiv -\infty$.*
- (iii) *If $0 < \underline{f}^{(n-1)}(x_0, x_1, \dots, x_{n-2}; x_{n-1}) \leq +\infty$, then $\underline{f}^{(n)}(x_0, x_1, \dots, x_{n-1}; \cdot) \equiv +\infty$.*
- (iv) *If $0 > \underline{f}^{(n-1)}(x_0, x_1, \dots, x_{n-2}; x_{n-1}) \geq -\infty$, then $\underline{f}^{(n)}(x_0, x_1, \dots, x_{n-1}; \cdot) \equiv -\infty$.*

PROOF. (i) Suppose that $0 < \bar{f}^{(n-1)}(x_0, x_1, \dots, x_{n-2}; x_{n-1}) \leq +\infty$. From the definition there exists a sequence $\{t_k\}$ of positive numbers such that $\{t_k\} \downarrow 0$ and

$$\lim_{k \rightarrow \infty} \frac{1}{t_k^{n-1}} \left[f\left(\sum_{i=0}^{n-1} t_k^i x_i\right) - f(x_0) \right] = \bar{f}^{(n)}(x_0, x_1, \dots, x_{n-2}; x_{n-1}). \quad (3.11)$$

Since the right-hand side of (3.11) is positive or possibly positive infinity, we may assume

$$\frac{1}{t_k^{n-1}} \left[f \left(\sum_{i=0}^{n-1} t_k^i x_i \right) - f(x_0) \right] > 0 \quad k \geq 1. \quad (3.12)$$

Hence, by (3.11) and (3.12) we have

$$\begin{aligned} & \limsup_{\lambda \downarrow 0} \frac{1}{\lambda^n} \left[f \left(\sum_{i=0}^{n-1} \lambda^i x_i \right) - f(x_0) \right] \\ & \geq \limsup_{k \rightarrow \infty} \frac{1}{t_k^n} \left[f \left(\sum_{i=0}^{n-1} t_k^i x_i \right) - f(x_0) \right] \\ & \geq \limsup_{k \rightarrow \infty} \frac{1}{t_k^{n-1}} \left[f \left(\sum_{i=0}^{n-1} t_k^i x_i \right) - f(x_0) \right] \liminf_{k \rightarrow \infty} \frac{1}{t_k} \\ & = +\infty. \end{aligned}$$

Namely,

$$\limsup_{\lambda \downarrow 0} \frac{1}{\lambda^n} \left[f \left(\sum_{i=0}^{n-1} \lambda^i x_i \right) - f(x_0) \right] = +\infty.$$

From the proof of Proposition 3.2 (ii) then we have $\bar{f}^{(n)}(x_0, x_1, \dots, x_{n-1}; \cdot) \equiv +\infty$.

(ii) Suppose that $0 > \bar{f}^{(n-1)}(x_0, x_1, \dots, x_{n-2}; x_{n-1}) \geq -\infty$. Then we may assume that for sufficiently small $\lambda > 0$,

$$f \left(\sum_{i=0}^{n-1} \lambda^i x_i \right) - f(x_0) < 0.$$

Therefore,

$$\begin{aligned} & \liminf_{\lambda \downarrow 0} \frac{1}{\lambda^n} \left[f(x_0) - f \left(\sum_{i=0}^{n-1} \lambda^i x_i \right) \right] \\ & \geq \liminf_{\lambda \downarrow 0} \frac{1}{\lambda^{n-1}} \left[f(x_0) - f \left(\sum_{i=0}^{n-1} \lambda^i x_i \right) \right] \liminf_{\lambda \downarrow 0} \frac{1}{\lambda} \\ & = -\limsup_{\lambda \downarrow 0} \frac{1}{\lambda^{n-1}} \left[f \left(\sum_{i=0}^{n-1} \lambda^i x_i \right) - f(x_0) \right] \lim_{\lambda \downarrow 0} \frac{1}{\lambda} \\ & = +\infty, \end{aligned}$$

and so

$$\limsup_{\lambda \downarrow 0} \frac{1}{\lambda^n} \left[f \left(\sum_{i=0}^{n-1} \lambda^i x_i \right) - f(x_0) \right] = -\infty.$$

The proof of Proposition 3.2 (ii) indicates that $\bar{f}^{(n)}(x_0, x_1, \dots, x_{n-1}; \cdot) \equiv -\infty$.

The assertion (iii) follows from Proposition 3.1 and (ii), and (iv) follows from Proposition 3.1 and (i). \square

PROPOSITION 3.4. *Let $f: U \rightarrow Y$ and let $x_0 \in U$. Then*

(i) *If f is Fréchet differentiable at x_0 , then it holds that*

$$f'(x_0; x_1) = f^{(1)}(x_0; x_1) = Df(x_0)x_1 \quad \text{for } x_1 \in X,$$

where $Df(x_0)$ denotes the Fréchet derivative of f at x_0 .

(ii) *If f is twice Fréchet differentiable at x_0 , then it holds that*

$$f''(x_0, x_1; x_2) = f^{(2)}(x_0, x_1; x_2) = Df(x_0)x_2 + \frac{1}{2}D^2f(x_0)(x_1, x_1) \\ \text{for } x_1, x_2 \in X,$$

where $D^2f(x_0)$ denotes the second-order Fréchet derivative of f at x_0 .

The results of this proposition can be derived by simple calculations, so the proofs are omitted.

4. Representation Formulae of Higher-order Variational Sets

In Section 2 we have defined the three kinds of higher-order variational sets. In applications to concrete optimization problems, e.g., optimal control, best approximation problems and so on, it is required to calculate the variational sets. But it is not easy to calculate directly from their definitions the variational sets of a given set.

In this section we shall give certain representation formulae of the variational sets, which are equivalent to their definitions. Those formulae will be effectively used to calculate the variational sets in Section 7 dealing with the problem of best approximation.

For a nonempty arbitrary subset S of a real Banach space X , let d_S be the distance function of S defined by

$$d_S(x) = \inf\{\|x - y\| \mid y \in S\} \quad \text{for } x \in X.$$

If S is empty, we define $d_S(\cdot) \equiv +\infty$. It is easy to prove that d_S is a Lipschitz function with the Lipschitz constant 1, if S is nonempty. For a given set Q in X , let Δ_Q be the function defined by

$$\Delta_Q(x) = d_Q(x) - d_{Q^c}(x) \quad \text{for } x \in X.$$

If Q is a nonempty proper subset, Δ_Q is a Lipschitz function.

Throughout the remainder of this section we assume that Q is a nonempty proper subset of X .

DEFINITION 4.1. Let x_0, x_1, \dots, x_{n-1} be points of X . Define three kinds of sets by the following:

$$\begin{aligned} \mathcal{F}(Q; x_0, x_1, \dots, x_{n-1}) &= \{h \in X \mid \bar{J}_Q^{(n)}(x_0, x_1, \dots, x_{n-1}; h) < 0\}, \\ \mathcal{CV}(Q; x_0, x_1, \dots, x_{n-1}) &= \{h \in X \mid \bar{J}_Q^{(n)}(x_0, x_1, \dots, x_{n-1}; h) \leq 0\}, \\ \mathcal{I}(Q; x_0, x_1, \dots, x_{n-1}) &= \{h \in X \mid \underline{J}_Q^{(n)}(x_0, x_1, \dots, x_{n-1}; h) \leq 0\}. \end{aligned}$$

PROPOSITION 4.1. The sets $\mathcal{CV}(Q; x_0, x_1, \dots, x_{n-1})$ and $\mathcal{I}(Q; x_0, x_1, \dots, x_{n-1})$ are closed, and the set $\mathcal{F}(Q; x_0, x_1, \dots, x_{n-1})$ is open.

PROOF. We state the proof for the set $\mathcal{CV}(Q; x_0, x_1, \dots, x_{n-1})$ only, because the other sets are proved similarly.

(i) The case $n=1$.

Since Δ_Q is a Lipschitz function, $\bar{J}_Q^{(1)}$ is real-valued Lipschitz continuous. Therefore $\mathcal{CV}(Q; x_0)$ is closed by its definition.

(ii) The case $n \geq 2$.

By Proposition 3.2 (ii), either one and only one of the following occurs:

- (a) $\bar{J}Q^{(n)}(x_0, x_1, \dots, x_{n-1}; \cdot) \equiv +\infty$.
- (b) $\bar{J}Q^{(n)}(x_0, x_1, \dots, x_{n-1}; \cdot) \equiv -\infty$.
- (c) $\bar{J}Q^{(n)}(x_0, x_1, \dots, x_{n-1}; \cdot)$ is real-valued Lipschitzian.

(a) implies $\mathcal{CV}(Q; x_0, x_1, \dots, x_{n-1}) = \emptyset$, (b) implies $\mathcal{CV}(Q; x_0, x_1, \dots, x_{n-1}) = X$, and (c) implies $\mathcal{CV}(Q; x_0, x_1, \dots, x_{n-1})$ is closed. Thus in any case of (a), (b), (c), the set $\mathcal{CV}(Q; x_0, x_1, \dots, x_{n-1})$ is closed. \square

THEOREM 4.1. *Let $x_0 \in \text{bd } Q$. Then*

- (i) $\mathcal{F}(Q; x_0, x_1, \dots, x_{n-1}) = F(Q; x_0, x_1, \dots, x_{n-1})$.
- (ii) $\mathcal{CV}(Q; x_0, x_1, \dots, x_{n-1}) = V(Q; x_0, x_1, \dots, x_{n-1})$.
- (iii) $\mathcal{T}(Q; x_0, x_1, \dots, x_{n-1}) = T(Q; x_0, x_1, \dots, x_{n-1})$.

PROOF. (i) First note that $F(Q; x_0, x_1, \dots, x_{n-1})$ can be expressed by the following, from its definition,

$$F(Q; x_0, x_1, \dots, x_{n-1}) = \left\{ h \in X \left| \begin{array}{l} \exists t_0 > 0, \exists \alpha > 0 \\ \text{such that} \\ d_{Q^c} \left(\sum_{i=0}^{n-1} t^i x_i + t^n h \right) \geq \alpha t^n \quad \text{for } \forall t \in (0, t_0] \end{array} \right. \right\} \quad (4.1)$$

Since $x_0 \in \text{bd } Q$, on the other hand, it holds that $h \in \mathcal{F}(Q; x_0, x_1, \dots, x_{n-1})$ if and only if

$$\limsup_{\lambda \downarrow 0} \frac{1}{\lambda^n} \left[d_Q \left(\sum_{i=0}^{n-1} \lambda^i x_i + \lambda^n h \right) - d_{Q^c} \left(\sum_{i=0}^{n-1} \lambda^i x_i + \lambda^n h \right) \right] < 0. \quad (4.2)$$

Now let $h \in \mathcal{F}(Q; x_0, x_1, \dots, x_{n-1})$. Then, by (4.2),

$$\begin{aligned} & \limsup_{\lambda \downarrow 0} \left[-\frac{1}{\lambda^n} d_{Q^c} \left(\sum_{i=0}^{n-1} \lambda^i x_i + \lambda^n h \right) \right] \\ & \leq \limsup_{\lambda \downarrow 0} \frac{1}{\lambda^n} \left[d_Q \left(\sum_{i=0}^{n-1} \lambda^i x_i + \lambda^n h \right) - d_{Q^c} \left(\sum_{i=0}^{n-1} \lambda^i x_i + \lambda^n h \right) \right] < 0, \end{aligned}$$

which implies that

$$\liminf_{\lambda \downarrow 0} \frac{1}{\lambda^n} d_{Q^c} \left(\sum_{i=0}^{n-1} \lambda^i x_i + \lambda^n h \right) > 0.$$

Hence there exists a number $\lambda_0 > 0$ such that $\alpha > 0$ if we let

$$\alpha = \inf_{0 < \lambda \leq \lambda_0} \frac{1}{\lambda^n} d_{Q^c} \left(\sum_{i=0}^{n-1} \lambda^i x_i + \lambda^n h \right),$$

and so

$$d_{Q^c} \left(\sum_{i=0}^{n-1} \lambda^i x_i + \lambda^n h \right) \geq \alpha \lambda^n \quad \text{for } \forall \lambda \in (0, \lambda_0].$$

Thus, from (4.1), we get $h \in F(Q; x_0, x_1, \dots, x_{n-1})$.

Conversely, let $h \in F(Q; x_0, x_1, \dots, x_{n-1})$. By (4.1) then we have

$$\liminf_{\lambda \downarrow 0} \frac{1}{\lambda^n} d_{Q^c} \left(\sum_{i=0}^{n-1} \lambda^i x_i + \lambda^n h \right) > 0. \quad (4.3)$$

Since

$$\sum_{i=0}^{n-1} \lambda^i x_i + \lambda^n h \in \text{int } Q \quad \text{for sufficiently small } \lambda > 0$$

by Proposition 2.2 (i), we have

$$\begin{aligned} & \limsup_{\lambda \downarrow 0} \frac{1}{\lambda^n} \left[d_Q \left(\sum_{i=0}^{n-1} \lambda^i x_i + \lambda^n h \right) - d_{Q^c} \left(\sum_{i=0}^{n-1} \lambda^i x_i + \lambda^n h \right) \right] \\ &= \limsup_{\lambda \downarrow 0} \frac{1}{\lambda^n} \left[-d_{Q^c} \left(\sum_{i=0}^{n-1} \lambda^i x_i + \lambda^n h \right) \right] < 0, \end{aligned}$$

where the last inequality follows from (4.3). By appealing to the relation (4.2) we have $h \in \mathcal{F}(Q; x_0, x_1, \dots, x_{n-1})$.

(ii) First we prove the following relation :

$$\mathcal{CV}(Q; x_0, x_1, \dots, x_{n-1}) = \{h \in X \mid \bar{d}_Q^{(n)}(x_0, x_1, \dots, x_{n-1}; h) = 0\}. \quad (4.4)$$

Let $h \in \mathcal{CV}(Q; x_0, x_1, \dots, x_{n-1})$. Suppose that $\bar{d}_Q^{(n)}(x_0, x_1, \dots, x_{n-1}; h) > 0$. From the definition, then, there exists a sequence $\{\lambda_k\}$ of positive numbers such that $\{\lambda_k\} \downarrow 0$ and

$$\begin{aligned} \frac{1}{\lambda_k^n} \left[d_Q \left(\sum_{i=0}^{n-1} \lambda_k^i x_i + \lambda_k^n h \right) \right] &\longrightarrow \bar{d}_Q^{(n)}(x_0, x_1, \dots, x_{n-1}; h) \quad (> 0) \\ &\text{as } k \rightarrow \infty. \end{aligned} \quad (4.5)$$

Therefore we can conclude that

$$\sum_{i=0}^{n-1} \lambda_k^i x_i + \lambda_k^n h \in Q \quad \text{for sufficiently large } k. \quad (4.6)$$

Noting that $\mathcal{A}_Q(x_0) = 0$, from (4.5) and (4.6) we have

$$\begin{aligned} & 0 < \bar{d}_Q^{(n)}(x_0, x_1, \dots, x_{n-1}; h) \\ &= \lim_{k \rightarrow \infty} \frac{1}{\lambda_k^n} \left[d_Q \left(\sum_{i=0}^{n-1} \lambda_k^i x_i + \lambda_k^n h \right) - d_{Q^c} \left(\sum_{i=0}^{n-1} \lambda_k^i x_i + \lambda_k^n h \right) \right] \\ &= \lim_{k \rightarrow \infty} \frac{1}{\lambda_k^n} \left[\mathcal{A}_Q \left(\sum_{i=0}^{n-1} \lambda_k^i x_i + \lambda_k^n h \right) - \mathcal{A}_Q(x_0) \right] \\ &\leq \limsup_{\lambda \downarrow 0} \frac{1}{\lambda^n} \left[\mathcal{A}_Q \left(\sum_{i=0}^{n-1} \lambda^i x_i + \lambda^n h \right) - \mathcal{A}_Q(x_0) \right] \\ &= \bar{\mathcal{A}}_Q^{(n)}(x_0, x_1, \dots, x_{n-1}; h), \end{aligned}$$

which contradicts that $h \in \mathcal{CV}(Q; x_0, x_1, \dots, x_{n-1})$. Thus we have

$$\bar{d}_Q^{(n)}(x_0, x_1, \dots, x_{n-1}; h) \leq 0. \quad (4.7)$$

But, since it holds that

$$\begin{aligned}
& \bar{d}_Q^{(n)}(x_0, x_1, \dots, x_{n-1}; h) \\
&= \limsup_{\lambda \downarrow 0} \frac{1}{\lambda^n} \left[d_Q \left(\sum_{i=0}^{n-1} \lambda^i x_i + \lambda^n h \right) - d_Q(x_0) \right] \\
&= \limsup_{\lambda \downarrow 0} \frac{1}{\lambda^n} \left[d_Q \left(\sum_{i=0}^{n-1} \lambda^i x_i + \lambda^n h \right) \right] \geq 0,
\end{aligned}$$

(4.7) implies that $\bar{d}_Q^{(n)}(x_0, x_1, \dots, x_{n-1}; h) = 0$.

Conversely, let h be such that $\bar{d}_Q^{(n)}(x_0, x_1, \dots, x_{n-1}; h) = 0$. Then it holds that

$$\begin{aligned}
& \bar{d}_Q^{(n)}(x_0, x_1, \dots, x_{n-1}; h) \\
&= \limsup_{\lambda \downarrow 0} \frac{1}{\lambda^n} \left[d_Q \left(\sum_{i=0}^{n-1} \lambda^i x_i + \lambda^n h \right) - d_Q \left(\sum_{i=0}^{n-1} \lambda^i x_i + \lambda^n h \right) \right] \\
&\leq \limsup_{\lambda \downarrow 0} \frac{1}{\lambda^n} \left[d_Q \left(\sum_{i=0}^{n-1} \lambda^i x_i + \lambda^n h \right) - d_Q(x_0) \right] \\
&= \bar{d}_Q^{(n)}(x_0, x_1, \dots, x_{n-1}; h) = 0,
\end{aligned}$$

which leads us to $h \in \mathcal{CV}(Q; x_0, x_1, \dots, x_{n-1})$. Thus (4.4) has been proved.

Next we prove the relation:

$$\{h \in X \mid \bar{d}_Q^{(n)}(x_0, x_1, \dots, x_{n-1}; h) = 0\} = V(Q; x_0, x_1, \dots, x_{n-1}). \quad (4.8)$$

Let h be such that $\bar{d}_Q^{(n)}(x_0, x_1, \dots, x_{n-1}; h) = 0$. Then it follows that

$$\limsup_{\lambda \downarrow 0} \frac{1}{\lambda^n} d_Q \left(\sum_{i=0}^{n-1} \lambda^i x_i + \lambda^n h \right) = 0. \quad (4.9)$$

On the other hand, from the definition of $d_Q(\cdot)$, for every $\lambda > 0$ there exists a point $y(\lambda) \in Q$ such that

$$\left\| y(\lambda) - \left(\sum_{i=0}^{n-1} \lambda^i x_i + \lambda^n h \right) \right\| \leq d_Q \left(\sum_{i=0}^{n-1} \lambda^i x_i + \lambda^n h \right) + \lambda^{n+1}.$$

Hence, if we let

$$r(\lambda) = y(\lambda) - \left(\sum_{i=0}^{n-1} \lambda^i x_i + \lambda^n h \right),$$

we have

$$\frac{1}{\lambda^n} \|r(\lambda)\| \leq \frac{1}{\lambda^n} d_Q \left(\sum_{i=0}^{n-1} \lambda^i x_i + \lambda^n h \right) + \lambda. \quad (4.10)$$

From (4.9) and (4.10), it follows that

$$\lim_{\lambda \downarrow 0} \frac{1}{\lambda^n} \|r(\lambda)\| = 0.$$

Rewriting the relation that $y(\lambda) \in Q$ yields

$$\sum_{i=0}^{n-1} \lambda^i x_i + \lambda^n h + r(\lambda) \in Q \quad \forall \lambda > 0.$$

The last two results together imply that $h \in V(Q; x_0, x_1, \dots, x_{n-1})$.

Conversely, let $h \in V(Q; x_0, x_1, \dots, x_{n-1})$. Then there exist a number $\lambda_0 > 0$ and a mapping $r: (0, \lambda_0] \rightarrow X$ such that

$$\sum_{i=0}^{n-1} \lambda^i x_i + \lambda^n h + r(\lambda) \in Q \quad \text{for } \forall \lambda \in (0, \lambda_0],$$

$$\frac{1}{\lambda^n} \|r(\lambda)\| \rightarrow 0 \quad \text{as } \lambda \downarrow 0.$$

Hence

$$\begin{aligned} & \bar{d}_Q^{(n)}(x_0, x_1, \dots, x_{n-1}; h) \\ &= \limsup_{\lambda \downarrow 0} \frac{1}{\lambda^n} d_Q \left(\sum_{i=0}^{n-1} \lambda^i x_i + \lambda^n h \right) \\ &\leq \limsup_{\lambda \downarrow 0} \frac{1}{\lambda^n} \left\| \left(\sum_{i=0}^{n-1} \lambda^i x_i + \lambda^n h \right) - \left(\sum_{i=0}^{n-1} \lambda^i x_i + \lambda^n h + r(\lambda) \right) \right\| \\ &= \limsup_{\lambda \downarrow 0} \frac{1}{\lambda^n} \|r(\lambda)\| = 0, \end{aligned}$$

which implies $\bar{d}_Q^{(n)}(x_0, x_1, \dots, x_{n-1}; h) \leq 0$. But, this result is equivalent to $\bar{d}_Q^{(n)}(x_0, x_1, \dots, x_{n-1}; h) = 0$, as seen in the proof of (4.4). Thus the relation (4.8) has been proved.

Finally, combining (4.4) and (4.8) yields the assertion of (ii).

(iii) First we show the relation:

$$\mathcal{T}(Q; x_0, x_0, x_1, \dots, x_{n-1}) = \{h \in X \mid \underline{d}_Q^{(n)}(x_0, x_1, \dots, x_{n-1}; h) = 0\}. \quad (4.11)$$

Let $h \in \mathcal{T}(Q; x_0, x_1, \dots, x_{n-1})$. Then $\underline{d}_Q^{(n)}(x_0, x_1, \dots, x_{n-1}; h) \leq 0$ from the definition. Suppose that $\underline{d}_Q^{(n)}(x_0, x_1, \dots, x_{n-1}; h) > 0$. Then there exists $\lambda_0 > 0$ such that

$$\inf_{0 < \lambda < \lambda_0} \frac{1}{\lambda^n} \left[d_Q \left(\sum_{i=0}^{n-1} \lambda^i x_i + \lambda^n h \right) \right] > 0,$$

and so

$$\sum_{i=0}^{n-1} \lambda^i x_i + \lambda^n h \notin Q \quad \text{for } \forall \lambda \in (0, \lambda_0].$$

Hence,

$$\begin{aligned} 0 &< \inf_{0 < \lambda < \lambda_0} \frac{1}{\lambda^n} \left[d_Q \left(\sum_{i=0}^{n-1} \lambda^i x_i + \lambda^n h \right) \right] \\ &= \inf_{0 < \lambda < \lambda_0} \frac{1}{\lambda^n} \left[d_Q \left(\sum_{i=0}^{n-1} \lambda^i x_i + \lambda^n h \right) - d_Q \left(\sum_{i=0}^{n-1} \lambda^i x_i + \lambda^n h \right) \right] \\ &= \inf_{0 < \lambda < \lambda_0} \frac{1}{\lambda^n} \left[\mathcal{A}_Q \left(\sum_{i=0}^{n-1} \lambda^i x_i + \lambda^n h \right) \right] \\ &= \inf_{0 < \lambda < \lambda_0} \frac{1}{\lambda^n} \left[\mathcal{A}_Q \left(\sum_{i=0}^{n-1} \lambda^i x_i + \lambda^n h \right) - \mathcal{A}_Q(x_0) \right] \\ &\leq \liminf_{\lambda \downarrow 0} \frac{1}{\lambda^n} \left[\mathcal{A}_Q \left(\sum_{i=0}^{n-1} \lambda^i x_i + \lambda^n h \right) - \mathcal{A}_Q(x_0) \right] \\ &= \underline{\mathcal{A}}_Q^{(n)}(x_0, x_1, \dots, x_{n-1}; h), \end{aligned}$$

which is a contradiction. Therefore we have $\underline{d}_Q^{(n)}(x_0, x_1, \dots, x_{n-1}; h) \leq 0$, which is equivalent to $\underline{d}_Q^{(n)}(x_0, x_1, \dots, x_{n-1}; h) = 0$.

Conversely let $\underline{d}_Q^{(n)}(x_0, x_1, \dots, x_{n-1}; h) = 0$. Then we have

$$\begin{aligned} & \underline{d}_Q^{(n)}(x_0, x_1, \dots, x_{n-1}; h) \\ & \leq \liminf_{\lambda \downarrow 0} \frac{1}{\lambda^n} \left[d_Q \left(\sum_{i=0}^{n-1} \lambda^i x_i + \lambda^n h \right) \right] \\ & = \liminf_{\lambda \downarrow 0} \frac{1}{\lambda^n} \left[d_Q \left(\sum_{i=0}^{n-1} \lambda^i x_i + \lambda^n h \right) - d_Q(x_0) \right] \\ & = \underline{d}_Q^{(n)}(x_0, x_1, \dots, x_{n-1}; h) = 0, \end{aligned}$$

which implies $h \in \mathcal{T}(Q; x_0, x_1, \dots, x_{n-1})$. This proves (4.11).

The next step is to prove

$$\{h \in X \mid \underline{d}_Q^{(n)}(x_0, x_1, \dots, x_{n-1}; h) = 0\} = T(Q; x_0, x_1, \dots, x_{n-1}). \quad (4.12)$$

Let h be such that $\underline{d}_Q^{(n)}(x_0, x_1, \dots, x_{n-1}; h) = 0$. Then there exists a sequence $\{\lambda_k\}$ of positive numbers such that $\{\lambda_k\} \downarrow 0$ and

$$\frac{1}{\lambda_k^n} \left[d_Q \left(\sum_{i=0}^{n-1} \lambda_k^i x_i + \lambda_k^n h \right) \right] \rightarrow 0 \quad \text{as } k \rightarrow \infty. \quad (4.13)$$

For each positive integer k , on the other hand, there exists a point $y_k \in Q$ such that

$$\left\| y_k - \left(\sum_{i=0}^{n-1} \lambda_k^i x_i + \lambda_k^n h \right) \right\| \leq d_Q \left(\sum_{i=0}^{n-1} \lambda_k^i x_i + \lambda_k^n h \right) + \lambda_k^{n+1}. \quad (4.14)$$

From (4.13) and (4.14) it follows that

$$h = \lim_{k \rightarrow \infty} \frac{1}{\lambda_k^n} \left(y_k - \sum_{i=0}^{n-1} \lambda_k^i x_i \right). \quad (4.15)$$

Hence $h \in T(Q; x_0, x_1, \dots, x_{n-1})$.

Conversely, let $h \in T(Q; x_0, x_1, \dots, x_{n-1})$. Then there exist a sequence $\{y_k\} \subset Q$ and a sequence $\{\lambda_k\}$ of positive numbers such that $\{\lambda_k\} \downarrow 0$ and the relation (4.15) holds. Let us put

$$z_k = \frac{1}{\lambda_k^n} \left(y_k - \sum_{i=0}^{n-1} \lambda_k^i x_i \right),$$

then (4.15) implies $z_k \rightarrow h$ as $k \rightarrow \infty$. By using the fact that the distance function is a Lipschitz function of the Lipschitz constant 1, hence, we have

$$\begin{aligned} \frac{1}{\lambda_k^n} \left| d_Q \left(\sum_{i=0}^{n-1} \lambda_k^i x_i + \lambda_k^n h \right) - d_Q \left(\sum_{i=0}^{n-1} \lambda_k^i x_i + \lambda_k^n z_k \right) \right| & \leq \frac{1}{\lambda_k^n} \|\lambda_k^n h - \lambda_k^n z_k\| \\ & = \|h - z_k\| \rightarrow 0 \quad \text{as } k \rightarrow \infty. \end{aligned}$$

From this and $\{y_k\} \subset Q$ it follows that

$$\begin{aligned}
d_Q^{(n)}(x_0, x_1, \dots, x_{n-1}; h) &= \liminf_{\lambda \downarrow 0} \frac{1}{\lambda^n} \left[d_Q \left(\sum_{i=0}^{n-1} \lambda^i x_i + \lambda^n h \right) \right] \\
&\leq \liminf_{k \rightarrow \infty} \frac{1}{\lambda_k^n} \left[d_Q \left(\sum_{i=0}^{n-1} \lambda_k^i x_i + \lambda_k^n h \right) \right] \\
&= \liminf_{k \rightarrow \infty} \frac{1}{\lambda_k^n} \left[d_Q \left(\sum_{i=0}^{n-1} \lambda_k^i x_i + \lambda_k^n z_k \right) \right] \\
&= \liminf_{k \rightarrow \infty} \frac{1}{\lambda_k^n} d_Q(y_k) = 0.
\end{aligned}$$

Again from the equivalency we get $d_Q^{(n)}(x_0, x_1, \dots, x_{n-1}; h) = 0$. This proves (4.12).

Combining (4.11) and (4.12) yields the assertion of (iii). This completes the proof of the theorem. \square

We close this section by proving some parts of propositions in Section 2, which still remain to be proved.

[Proof of Proposition 2.1 (iv)]

If $x_0 \in \text{int } Q$ or $x_0 \in \text{cl } Q$, the closedness of the variational set is obvious from the definition. Let $x_0 \in \text{bd } Q$. Then we have $V(Q; x_0, x_1, \dots, x_{n-1}) = \mathcal{C}V(Q; x_0, x_1, \dots, x_{n-1})$ by Theorem 4.1 (ii). But, since $\mathcal{C}V(Q; x_0, x_1, \dots, x_{n-1})$ is closed by Proposition 4.1, $V(Q; x_0, x_1, \dots, x_{n-1})$ is closed. \square

[Proof of Proposition 2.2 (iii)]

If $x_0 \in \text{int } Q$, then obvious. In the case when $x_0 \in \text{cl } Q$, $F(Q; x_0, x_1, \dots, x_{n-1})$ is empty by the definition. Hence this case can be deleted in the proposition now. Let $x_0 \in \text{bd } Q$, and suppose that $x_n \in F(Q; x_0, x_1, \dots, x_{n-1})$. Then, by Theorem 4.1 (i), $x_n \in \mathcal{F}(Q; x_0, x_1, \dots, x_{n-1})$, namely $\tilde{J}_Q^{(n)}(x_0, x_1, \dots, x_{n-1}; x_n) < 0$. Hence, from Proposition 3.3 (ii), $\tilde{J}_Q^{(n+1)}(x_0, x_1, \dots, x_{n-1}, x_n; \cdot) \equiv -\infty$, which implies $\mathcal{F}(Q; x_0, x_1, \dots, x_n) = X$. Again from Theorem 4.1 (i) we get $F(Q; x_0, x_1, \dots, x_n) = X$. \square

5. Higher-order Necessary Optimality Conditions

In this section we study higher-order necessary conditions for local minimum solutions of a general nonlinear programming problem. The necessary conditions will be given in an abstract form by the use of higher-order variational sets and variational derivatives.

Let both X and Y be real Banach spaces, Q an arbitrary nonempty subset of X , and \hat{Q} an open subset of X such that $Q \subset \hat{Q}$. Let B be an arbitrary subset of Y such that $\text{int } B \neq \emptyset$, where $\text{int } B$ denotes the interior of B . Let f a real-valued functional defined on \hat{Q} , and g a map from \hat{Q} into Y .

Then the optimization problem to be considered is

$$\left. \begin{aligned} &\text{minimize } f(x) \\ &\text{subject to} \\ &g(x) \in B, \ x \in Q \end{aligned} \right\}. \quad (\mathcal{P})$$

In many cases in the literature the set B is assumed to be a convex cone. We should note that, in this paper, nothing about B is assumed except that $\text{int } B \neq \emptyset$.

The following theorem is, of course, included as a special case in the theorem of higher-order necessary conditions which we shall give later. However we consciously present the proof of it, because the theorem is new as far as the first-order criterion concerns.

THEOREM 5.1. (*First-order necessary condition*). *Let x_0 be a local minimum solution of (\mathcal{P}) . Suppose that both f and g be Neustadt differentiable at x_0 . Then the system of h :*

$$f^{(1)}(x_0; h) < 0 \quad (5.1)$$

$$g^{(1)}(x_0; h) \in F(B; g(x_0)) \quad (5.2)$$

$$h \in T(Q; x_0) \quad (5.3)$$

is inconsistent.

PROOF. Suppose that there exist a point h satisfying (5.1), (5.2), (5.3). From (5.3) then there exist a sequence $\{x_n\} \subset Q$ and a sequence $\{\lambda_n\}$ of positive numbers such that $\lambda_n \downarrow 0$ and

$$\frac{1}{\lambda_n}(x_n - x_0) \rightarrow h \quad \text{as } n \rightarrow \infty. \quad (5.4)$$

Since we have

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{\lambda_n} [f(x_n) - f(x_0)] &= \lim_{n \rightarrow \infty} \frac{1}{\lambda_n} \left[f\left(x_0 + \lambda_n \cdot \frac{1}{\lambda_n}(x_n - x_0)\right) - f(x_0) \right] \\ &= f^{(1)}(x_0; h) < 0 \end{aligned}$$

by (5.1), there exists a positive integer n_0 such that

$$f(x_n) < f(x_0) \quad \forall n > n_0.$$

By virtue of (5.2) there exist a neighborhood V of $g^{(1)}(x; h)$ and a positive number $t_0 > 0$ such that

$$g(x_0) + tV \subset B \quad \text{for } \forall t \in (0, t_0]. \quad (5.5)$$

Let

$$l_n = \frac{1}{\lambda_n} [g(x_n) - g(x_0)],$$

then we have

$$l_n \rightarrow g^{(1)}(x_0; h) \quad \text{as } n \rightarrow \infty \quad (5.6)$$

considering (5.4) and the formula

$$l_n = \frac{1}{\lambda_n} \left[g\left(x_0 + \lambda_n \cdot \frac{1}{\lambda_n}(x_n - x_0)\right) - g(x_0) \right].$$

Since V is a neighborhood of $g^{(1)}(x_0; h)$, it follows from (5.6) that there exists a positive integer n_1 such that $l_n \in V$ for all $n > n_1$. Choose a positive integer n_2 such that $n_2 \geq n_1$ and $\lambda_{n_2} \leq t_0$. Then, (5.5) implies that if $n > n_2$ then

$$g(x_0) + tl_n \in B \quad \text{for } \forall t \in (0, \lambda_n].$$

Hence, in particular we have

$$g(x_0) + \lambda_n l_n = g(x_n) \in B \quad \text{for } n > n_2.$$

Letting $n_3 = \max\{n_0, n_2\}$, we have $f(x_n) < f(x_0)$ and $g(x_n) \in B$ for all $n > n_3$. Since (5.4) implies that $x_n \rightarrow x_0$ as $n \rightarrow \infty$, this is a contradiction. \square

Let $R_- = \{r \in R \mid r < 0\}$.

THEOREM 5.2 (*Higher-order necessary condition*). *Let x_0 be a local minimum solution of (\mathcal{P}) . Suppose that both f and g be n -times Neustadt differentiable at x_0 . Then the system of h :*

$$\begin{aligned} & f^{(n)}(x_0, x_1, \dots, x_{n-1}; h) \\ & \in F(f(x_0) + R_-; f(x_0), f^{(1)}(x_0; x_1), \dots, f^{(n-1)}(x_0, x_1, \dots, x_{n-2}; x_{n-1})) \end{aligned} \quad (5.7)$$

$$\begin{aligned} & g^{(n)}(x_0, x_1, \dots, x_{n-1}; h) \\ & \in F(B; g(x_0), g^{(1)}(x_0; x_1), \dots, g^{(n-1)}(x_0, x_1, \dots, x_{n-2}; x_{n-1})) \end{aligned} \quad (5.8)$$

$$h \in T(Q; x_0, x_1, \dots, x_{n-1}) \quad (5.9)$$

is inconsistent for all $x_1, x_2, \dots, x_{n-1} \in X$.

PROOF. Let x_1, x_2, \dots, x_{n-1} be arbitrary but fixed points of X . Suppose that there exist a point h satisfying (5.7), (5.8), (5.9). From (5.9) we have a sequence $\{y_k\}$ of points of Q and a sequence $\{\lambda_k\}$ of positive numbers such that $\lambda_k \downarrow 0$ and

$$\frac{1}{\lambda_k^n} \left(y_k - \sum_{i=0}^{n-1} \lambda_k^i x_i \right) \rightarrow h \quad \text{as } k \rightarrow \infty. \quad (5.10)$$

Note that (5.10) with $\lambda_k \downarrow 0$ implies $y_k \rightarrow x_0$ as $k \rightarrow \infty$.

Next from (5.7) we have a neighborhood V of $f^{(n)}(x_0, x_1, \dots, x_{n-1}; h)$ and a number $t_0 > 0$ such that

$$\begin{aligned} & \sum_{i=0}^{n-1} t^i f^{(i)}(x_0, x_1, \dots, x_{i-1}; x_i) + t^n V \subset f(x_0) + R_- \\ & \text{for } \forall t \in (0, t_0], \end{aligned} \quad (5.11)$$

where $f^{(0)}(x_0) = f(x_0)$. Let

$$l_k = \frac{1}{\lambda_k^n} \left[f(y_k) - \sum_{i=0}^{n-1} \lambda_k^i f^{(i)}(x_0, x_1, \dots, x_{i-1}; x_i) \right].$$

Since l_k can be written as

$$l_k = \frac{1}{\lambda_k^n} \left[f \left(\sum_{i=0}^{n-1} \lambda_k^i x_i + \lambda_k^n \cdot \frac{1}{\lambda_k^n} \left(y_k - \sum_{i=0}^{n-1} \lambda_k^i x_i \right) \right) - \sum_{i=0}^{n-1} \lambda_k^i f^{(i)}(x_0, x_1, \dots, x_{i-1}; x_i) \right],$$

by (5.10) and the assumption of the n -times Neustadt differentiability of f we have

$$l_k \rightarrow f^{(n)}(x_0, x_1, \dots, x_{n-1}; h) \quad \text{as } k \rightarrow \infty.$$

Hence there exists a positive integer k_0 such that $l_k \in V$ for all $k > k_0$, because V is a neighborhood of $f^{(n)}(x_0, x_1, \dots, x_{n-1}; h)$. Choose an integer k_1 so that $k_1 \geq k_0$ and $\lambda_{k_1} \leq t_0$. Then it follows that $l_k \in V$ and $\lambda_k \leq t_0$ for all $k > k_1$. Therefore we have from

(5.11) that

$$\sum_{i=0}^{n-1} \lambda_k^i f^{(i)}(x_0, x_1, \dots, x_{i-1}; x_i) + \lambda_k^n l_k < f(x_0) \quad \forall k > k_1.$$

Substituting the definition formula of l_k into the above inequality, we have $f(y_k) < f(x_0)$ for all $k > k_1$.

Moreover from (5.8) we have a neighborhood U of $g^{(n)}(x_0, x_1, \dots, x_{n-1}; h)$ and a number $s_0 > 0$ such that

$$\sum_{i=0}^{n-1} t^i g^{(i)}(x_0, x_1, \dots, x_{i-1}; x_i) + t^n U \subset B \quad \forall t \in (0, s_0].$$

If we let

$$\eta_k = \frac{1}{\lambda_k^n} \left[g(y_k) - \sum_{i=0}^{n-1} \lambda_k^i g^{(i)}(x_0, x_1, \dots, x_{i-1}; x_i) \right],$$

we can get

$$\eta_k \rightarrow g^{(n)}(x_0, x_1, \dots, x_{n-1}; h) \quad \text{as } k \rightarrow \infty$$

by the same manner as in the case of f . Hence there exists a positive integer k_2 satisfying $\eta_k \in U$ for all $k > k_2$. Choose an integer k_3 so that $k_3 \geq k_2$ and $\lambda_{k_3} \leq s_0$. Then we have $g(y_k) \in B$ for all $k > k_3$.

Putting $k_4 = \max\{k_1, k_2\}$, finally, above results can be summarized as follows:

$$\begin{aligned} \{y_k\} &\subset Q, \\ y_k &\rightarrow x_0 \quad \text{as } k \rightarrow \infty, \\ f(y_k) &< f(x_0) \quad \forall k > k_1, \\ g(y_k) &\in B \quad \forall k > k_4. \end{aligned}$$

This contradicts that x_0 is a local minimum solution to (\mathcal{P}) , which proves the theorem. \square

As assured by the above theorem, the inconsistency of the system (5.7)~(5.9) is true for all points x_1, x_2, \dots, x_{n-1} of X . But for some particular points x_1, x_2, \dots, x_{n-1} the inconsistency of the system is trivially true. The following proposition clarifies this fact.

PROPOSITION 5.1. *Theorem 5.2 is meaningful, only if the points x_1, x_2, \dots, x_{n-1} satisfy simultaneously the following three conditions.*

- (R1) $f^{(n-1)}(x_0, x_1, \dots, x_{n-2}; x_{n-1})$
 $\in V(f(x_0) + R_-; f(x_0), f^{(1)}(x_0; x_1), \dots, f^{(n-2)}(x_0, x_1, \dots, x_{n-3}; x_{n-2})),$
- (R2) $g^{(n-1)}(x_0, x_1, \dots, x_{n-2}; x_{n-1})$
 $\in V(B; g(x_0), g^{(1)}(x_0; x_1), \dots, g^{(n-2)}(x_0, x_1, \dots, x_{n-3}; x_{n-2})),$
- (R3) $x_{n-1} \in T(Q; x_0, x_1, \dots, x_{n-2}).$

PROOF. We prove the contraposition of the assertion of this proposition. Suppose that the condition (R1) is violated. Then it follows from Proposition 2.2 (iv) that

$$F(f(x_0)+R_-; f(x_0), f^{(1)}(x_0; x_1), \dots, f^{(n-1)}(x_0, x_1, \dots, x_{n-2}; x_{n-1}))=\emptyset.$$

So the relation (5.7) is meaningless.

The case when (R2) is violated follows the same line the case of (R1). Finally suppose the condition (R3) be violated. Then by Proposition 2.2 (v) we have

$$T(Q; x_0, x_1, \dots, x_{n-1})=\emptyset.$$

Thus the relation (5.9) is meaningless. \square

REMARK 5.1. In the case $n=2$, the condition (R1) is equivalent to that $f^{(1)}(x_0; x_1) \leq 0$. This is obvious from the relation that for any $a \in R$,

$$V(a+R_-; a)=\text{cl } R_-=\{r \in R \mid r \leq 0\}. \quad (5.12)$$

REMARK 5.2. For every points x_1, x_2, \dots, x_{n-2} satisfying that

$$f^{(1)}(x_0; x_1)=f^{(2)}(x_0, x_1; x_2)=\dots=f^{(n-2)}(x_0, x_1, \dots, x_{n-3}; x_{n-2})=0, \quad (5.13)$$

the condition (R1) is equivalent to that $f^{(n-1)}(x_0, x_1, \dots, x_{n-2}; x_{n-1}) \leq 0$.

PROOF. Suppose (5.12) be satisfied. Then by virtue of Proposition 2.1 (ii) and (5.12) it follows that the condition (R1) holds if and only if

$$f^{(n-1)}(x_0, x_1, \dots, x_{n-2}; x_{n-1}) \in V(f(x_0)+R_-; f(x_0), 0, 0, \dots, 0)$$

if and only if

$$f^{(n-1)}(x_0, x_1, \dots, x_{n-2}; x_{n-1}) \in V(f(x_0)+R_-; f(x_0))$$

if and only if $f^{(n-1)}(x_0, x_1, \dots, x_{n-2}; x_{n-1}) \leq 0$. \square

6. Second-order Directional Derivatives of Maximum Type Functionals

As we have proved in Section 4, higher-order variational sets of a given set Q are represented by using the higher-order Dini derivatives of the functional Δ_Q . On the basis of this result, we need to calculate the Dini derivatives of the functional Δ_Q in order to obtain the explicit form of the variational sets. In the case when the underlying space X is the space of continuous functions, we may expect that the functional Δ_Q is of maximum type. In reality, we find that Δ_Q is expressed by a form of maximum type when Q is, for example, the unit sphere in the space of continuous functions. In general, on the other hand, the Dini derivative of a maximum type functional coincides with the directional derivative of it under certain mild conditions.

For these reasons, in this section we shall study the directional derivative of a maximum type functional, especially the second-order directional derivative.

Let X be a normed (real) vector space, \hat{E} an open set in X . Let T be a compact set in a metric space (M, d) . Let $g(t, x)$ be a real-valued continuous function defined on $T \times \hat{E}$. We use the notation $g_t(\cdot)$ as the function of x defined by $g_t(\cdot)=g(t, \cdot)$ for each $t \in T$. For a nonempty subset $\Omega \subset M$ and a positive number δ , $B(\Omega; \delta)$ is defined by

$$B(\Omega; \delta)=\text{cl} \bigcup_{a \in \Omega} \{b \in M \mid d(a, b) \leq \delta\}.$$

Theorem 6.1 has been proved by Furukawa [4] in the case when $X=R^n$. In the case of a normed vector space the proof is identical with that of the theorem in [4] except only a slight modification.

THEOREM 6.1. Define $G: \hat{E} \rightarrow R$ by

$$G(x) = \max_{t \in T} g(t, x) \quad \text{for } x \in \hat{E}. \quad (6.1)$$

Let $z \in \hat{E}$ be arbitrary but fixed, and let $I(z)$ be the subset of T defined by

$$I(z) = \{t \in T \mid g(t, z) = G(z)\}. \quad (6.2)$$

Suppose that

- (i) for each $t \in T$, the directional derivative $g'_t(z; \cdot)$ of $g_t(\cdot)$ at z does exist,
- (ii) for each $x \in X$, there exists a positive number δ such that

$$\frac{g_t(z + \lambda x) - g_t(z)}{\lambda} \rightarrow g'_t(z; x) \text{ uniformly in } t \in B(I(z); \delta) \cap T \quad \text{as } \lambda \downarrow 0.$$

Then we have

$$G'(z; x) = \max_{t \in I(z)} g'_t(z; x) \quad \text{for } x \in X. \quad (6.3)$$

THEOREM 6.2. Let G be defined by (6.1). Let z be an arbitrary point of \hat{E} . For each $x \in X$ define $I(x)$ by

$$I(x) = \{t \in T \mid g(t, x) = G(x)\}.$$

Suppose the hypotheses (i), (ii) in Theorem 6.1 be satisfied. Let $d \in X$ be such that $G'(z; d) = 0$, and let $J(z; d)$ be given by

$$J(z; d) = \{t \in I(z) \mid g'_t(z; d) = 0\}. \quad (6.4)$$

Suppose that

- (iii) for each $t \in T$, the second-order directional derivative $g''_t(z, d; \cdot)$ of $g_t(\cdot)$ at z (with respect to d) does exist,
- (iv) for each $h \in X$, there exists a positive number δ_1 such that

$$\frac{1}{\lambda^2} [g_t(z + \lambda d + \lambda^2 h) - g_t(z) - \lambda g'_t(z; d)] \rightarrow g''_t(z, d; h) \text{ uniformly in } t \in B(I(z); \delta_1) \cap T \quad \text{as } \lambda \downarrow 0. \quad (6.5)$$

Moreover we suppose the following regularity condition:

(Q_d) there exist a positive number λ_0 and a neighborhood N of d such that

$$g'_t(z; d) \leq 0 \quad \forall t \in \bigcup_{0 \leq \lambda < \lambda_0} \bigcup_{\bar{d} \in N} \{I(z + \lambda \bar{d})\}.$$

Then we have

$$G''(z, d; h) = \max_{t \in J(z; d)} g''_t(z, d; h) \quad \text{for } h \in X. \quad (6.6)$$

[Remark] Under the assumptions of the theorem, by Theorem 6.1 the first-order directional derivative of G at z is given by (6.3). Hence for $d \in X$ such that $G'(z; d) = 0$, it holds that $g'_t(z; d) \leq 0$ for all $t \in I(z)$ and $J(z; d) \neq \emptyset$.

PROOF. We may assume that $I(z) \neq T$, because the assertion of this theorem is trivial if otherwise.

Case 1: $J(z; d) \neq I(z)$.

Let $h \in X$ be arbitrary but fixed. By virtue of (iv) there exists $\delta_1 > 0$ satisfying (6.5). Let $\varepsilon > 0$ be arbitrary. Letting $W_1 = B(I(z); \delta_1) \cap T$, from (6.5) there exists a positive number λ_1 such that

$$|g_t(z + \lambda d + \lambda^2 h) - g_t(z) - \lambda g'_t(z; d) - \lambda^2 g''_t(z, d; h)| \leq \frac{\lambda^2 \varepsilon}{2} \quad \text{for } \forall \lambda \in (0, \lambda_1), \forall t \in W_1. \quad (6.7)$$

Since $I(z) \subseteq W_1 \subset T$, by the continuity of $g_t(z)$ in t we have a positive number α satisfying

$$\sup_{t \in T \setminus W_1} g_t(z) = G(z) - \alpha. \quad (6.8)$$

We now prove that there exist a neighborhood U of z such that

$$|g(t, y) - g(t, z)| < \frac{\alpha}{2} \quad \text{for } \forall t \in T, \forall y \in U. \quad (6.9)$$

By the uniform continuity of $g(t, z)$ in t on T , there exist a $\gamma > 0$ such that if $t, t' \in T$ and $d(t, t') < \gamma$ then $|g(t, z) - g(t', z)| < \frac{\alpha}{4}$. And also by the continuity of $g(t, x)$ in (t, x) , for each $t \in T$ there exist a neighborhood $V(t)$ of t and a neighborhood $U^t(z)$ of z depending on t such that

$$|g(s, y) - g(t, z)| < \frac{\alpha}{4} \quad \text{for } \forall (s, y) \in V(t) \times U^t(z).$$

Without loss of generality we may assume that $\text{diam} V(t) < \gamma$ for all $t \in T$. Since $\{V(t); t \in T\}$ is an open covering of the compact set T , we can select a finite set $\{t_i; i=1, 2, \dots, m\}$ so that $\{V(t_i); i=1, 2, \dots, m\}$ becomes a covering of T . Let $U = \bigcap_{i=1}^m U^{t_i}(z)$. Then U is trivially a neighborhood of z . It is easily verified that the neighborhood U defined above satisfies (6.9).

From (6.9) it follows that

$$G(z) \leq G(y) + \frac{\alpha}{2} \quad \text{for } \forall y \in U. \quad (6.10)$$

We obtain for any $t \in T \setminus W_1$ and $y \in U$

$$g(t, y) < g(t, z) + \frac{\alpha}{2} \leq \sup_{s \in T \setminus W_1} g(s, z) + \frac{\alpha}{2} \quad (\text{by (6.9)})$$

$$= G(z) - \frac{\alpha}{2} \quad (\text{by (6.8)})$$

$$\leq G(y), \quad (\text{by (6.10)})$$

therefore

$$g(t, y) < G(y) \quad \forall t \in T \setminus W_1, \forall y \in U. \quad (6.11)$$

Choose a positive number λ_2 so that

$$d + \lambda h \in N, \quad z + \lambda N \subset U \quad \forall \lambda \in (0, \lambda_2),$$

where N is given in (Q_d) . This is possible because N is a neighborhood of d and U a neighborhood of z . Then we have that $z + \lambda d + \lambda^2 h \in U$ for all $\lambda \in (0, \lambda_2)$. Therefore from (6.11) it follows that for any $\lambda \in (0, \lambda_2)$

$$G(z + \lambda d + \lambda^2 h) = \sup_{t \in W_1} g_t(z + \lambda d + \lambda^2 h). \quad (6.12)$$

Now we put $\lambda_3 = \min\{\lambda_0, \lambda_1, \lambda_2\}$, and put

$$\mathcal{A}(z, d) = \bigcup_{0 \leq \lambda \leq \lambda_3} \bigcup_{d \in N} \{I(z + \lambda \bar{d})\}.$$

Then one can easily prove that for any $\lambda \in (0, \lambda_3)$

$$G(z + \lambda d + \lambda^2 h) > g_t(z + \lambda d + \lambda^2 h) \quad \forall t \in (T \setminus \mathcal{A}(z, d)). \quad (6.13)$$

Define

$$Z = W_1 \cap \mathcal{A}(z, d).$$

So the set Z is not empty, because both of W_1 and $\mathcal{A}(z, d)$ include the set $I(z)$.

If $\lambda \in (0, \lambda_3)$, then it holds that

$$\begin{aligned} G(z + \lambda d + \lambda^2 h) &= \sup_{t \in Z} g_t(z + \lambda d + \lambda^2 h) && \text{(by (6.12), (6.13))} \\ &\leq \sup_{t \in Z} [g_t(z) + \lambda g'_t(z; d) + \lambda^2 g''_t(z, d; h)] + \frac{\lambda^2 \varepsilon}{2} && \text{(by (6.7))} \\ &\leq G(z) + \lambda \sup_{t \in Z} [g'_t(z; d) + \lambda g''_t(z, d; h)] + \frac{\lambda^2 \varepsilon}{2}, \end{aligned}$$

which implies

$$G(z + \lambda d + \lambda^2 h) - G(z) \leq \lambda \sup_{t \in Z} [g'_t(z; d) + \lambda g''_t(z, d; h)] + \frac{\lambda^2 \varepsilon}{2}. \quad (6.14)$$

From the continuity of $g''_t(z, d; h)$ in $t \in W_1$, which is an immediate consequence of (iv), and from that $J(z; d) \cap I(z) \subset Z$, one can choose $\delta_2 > 0$ so small that $Z \setminus [B(J(z; d); \delta_2) \cap T] \neq \emptyset$ and

$$\sup_{t \in B(J(z; d); \delta_2) \cap T} g''_t(z, d; h) \leq \sup_{t \in J(z; d)} g''_t(z, d; h) + \frac{\varepsilon}{2}. \quad (6.15)$$

Let us put

$$W_2 = B(J(z; d); \delta_2) \cap T, \quad \hat{W} = Z \setminus W_2,$$

and let

$$c(z, d; h) = \sup_{t \in J(z; d)} g''_t(z, d; h) = \max_{t \in J(z; d)} g''_t(z, d; h).$$

Here we should note that $J(z; d)$ is compact. By choice of δ_2 we have $\hat{W} \neq \emptyset$. By the notation defined above, (6.15) is as follows.

$$\sup_{t \in W_2} g''_t(z, d; h) \leq c(z, d; h) + \frac{\varepsilon}{2}. \quad (6.16)$$

We let

$$\beta = \sup_{t \in \tilde{W}} g'_t(z; d).$$

Remembering the condition of Case 1, we may assume that $\beta < 0$ (take in advance δ_1 so small, if necessary). Then one has

$$\sup_{t \in W_2} g'_t(z; d) \geq \sup_{t \in J(z; d)} g'_t(z; d) = 0 > \beta = \sup_{t \in \tilde{W}} g'_t(z; d).$$

By this strict inequality one can choose $\lambda_4 > 0$ such that

$$\sup_{t \in W_2} [g'_t(z; d) + \lambda g''_t(z, d; h)] \geq \sup_{t \in \tilde{W}} [g'_t(z; d) + \lambda g''_t(z, d; h)]$$

$$\forall \lambda \in (0, \lambda_4).$$

Let $\lambda_5 = \min\{\lambda_3, \lambda_4\}$. Then for any $\lambda \in (0, \lambda_5)$

$$\begin{aligned} \sup_{t \in Z} [g'_t(z; d) + \lambda g''_t(z, d; h)] &= \sup_{t \in Z \cap W_2} [g'_t(z; d) + \lambda g''_t(z, d; h)] \\ &\leq \sup_{t \in Z} g'_t(z; d) + \lambda \sup_{t \in W_2} g''_t(z, d; h). \end{aligned} \quad (6.17)$$

From the condition (Q_d) it follows that

$$\sup_{t \in Z} g'_t(z; d) \leq 0. \quad (6.18)$$

Substituting (6.16), (6.18) into (6.17) we obtain

$$\sup_{t \in Z} [g'_t(z; d) + \lambda g''_t(z, d; h)] \leq \lambda c(z, d; h) + \frac{\lambda \varepsilon}{2} \quad \forall \lambda \in (0, \lambda_5),$$

which is devoted to (6.14) to yield

$$\frac{1}{\lambda^2} [G(z + \lambda d + \lambda^2 h) - G(z)] \leq c(z, d; h) + \varepsilon \quad \forall \lambda \in (0, \lambda_5). \quad (6.19)$$

On the other hand, for sufficiently small $\lambda > 0$

$$\begin{aligned} G(z + \lambda d + \lambda^2 h) &= \sup_{t \in W_1} g_t(z + \lambda d + \lambda^2 h) && \text{(by (6.12))} \\ &\geq \sup_{t \in W_1} \left[g_t(z) + \lambda g'_t(z; d) + \lambda^2 g''_t(z, d; h) - \frac{\lambda^2 \varepsilon}{2} \right] && \text{(by (6.7))} \\ &\geq G(z) + \sup_{t \in I(z)} [\lambda g'_t(z; d) + \lambda^2 g''_t(z, d; h)] - \frac{\lambda^2 \varepsilon}{2} \\ &\geq G(z) + \lambda \sup_{t \in J(z; d)} [g'_t(z; d) + \lambda g''_t(z, d; h)] - \frac{\lambda^2 \varepsilon}{2} \\ &= G(z) + \lambda^2 \sup_{t \in J(z; d)} g''_t(z, d; h) - \frac{\lambda^2 \varepsilon}{2}. \end{aligned}$$

The last equality follows from the definition of $J(z; d)$. Hence

$$\frac{1}{\lambda^2} [G(z + \lambda d + \lambda^2 h) - G(z)] \geq c(z, d; h) - \varepsilon \quad (6.20)$$

for sufficiently small $\lambda > 0$. As $\varepsilon > 0$ is arbitrary, together (6.19), (6.20) imply

$$\lim_{\lambda \rightarrow 0} \frac{1}{\lambda^2} [G(z + \lambda d + \lambda^2 h) - G(z)] = c(z, d; h),$$

which is identical with the relation

$$G''(z, d; h) = c(z, d; h),$$

since d satisfies $G'(z; d) = 0$.

Case 2: $J(z; d) = I(z)$.

It is obvious from the proof of Case 1 that (6.14) is valid also in Case 2. Hence from (6.14) and (Q_d) we have that for sufficiently small $\lambda > 0$

$$\begin{aligned} G(z + \lambda d + \lambda^2 h) - G(z) &\leq \lambda \sup_{t \in Z} [g'_t(z; d) + \lambda g''_t(z, d; h)] + \frac{\lambda^2 \varepsilon}{2} \\ &\leq \lambda \sup_{t \in Z} g'_t(z; d) + \lambda^2 \sup_{t \in W_1} g''_t(z, d; h) + \frac{\lambda^2 \varepsilon}{2} \\ &= \lambda^2 \sup_{t \in W_1} g''_t(z, d; h) + \frac{\lambda^2 \varepsilon}{2}. \end{aligned}$$

Since we may assume that δ_1 is so small that

$$\sup_{t \in W_1} g''_t(z, d; h) \leq \sup_{t \in I(z)} g''_t(z, d; h) + \frac{\varepsilon}{2}$$

(the continuity of $g''_t(z, d; h)$ in t is a consequence of (iv)), we get

$$\begin{aligned} G(z + \lambda d + \lambda^2 h) - G(z) &\leq \lambda^2 \sup_{t \in I(z)} g''_t(z, d; h) + \lambda^2 \varepsilon \\ &= \lambda^2 c(z, d; h) + \lambda^2 \varepsilon \end{aligned}$$

for sufficiently small $\lambda > 0$. Here we should not that $I(z) = J(z; d)$. Thus we have proved that (6.19) holds also in the case under consideration. It is also obvious that (6.20) is valid in this case. This completes the proof. \square

REMARK 6.1. In [1] Ben-Tal, Teboulle and Zowe have studied the second order directional derivative of a maximum type functional, and given a formula similar to our result (6.6) under some conditions. They assumed, in our notation, that $g(t, x)$ is defined on $T \times R^n$, where T is a compact set, and that for each t , $g_t(x)$ is twice differentiable in usual sense and both $\nabla g_t(x)$ and $\nabla^2 g_t(x)$ are continuous in (t, x) . On the contrary, in our paper $g(t, x)$ is only assumed to have the first and second-order directional derivatives w.r.t. x for each t , and satisfy the uniformity conditions (ii), (iv). The regularity condition (Q_d) is essentially same as the condition $P(d)$ in their paper [1].

REMARK 6.2. If T is a finite set, the conditions (ii), (iv) and (Q_d) in Theorem 6.2 are necessarily satisfied without any other restriction. Therefore, in that case, the conditions (ii), (iv) and (Q_d) are all deleted from the assumption in Theorem 6.2.

7. Nonlinear Tchebycheff Approximation

The theorem of higher-order necessary conditions we have given in Section 5 can be applied to optimization problems of various types. In this section we study an

application of the theorem to some problem of nonlinear Tchebycheff approximation. Although we can show general higher-order necessary conditions in the Tchebycheff approximation problem, we shall restrict ourselves to the case of second-order condition for the sake of simplicity of descriptions.

Let X be a real Banach space, Q an arbitrary nonempty subset of X , and \hat{Q} an open set in X such that $Q \subset \hat{Q}$. Let T be a compact set in a metric space (M, d) . Let $C(T)$ be the normed vector space of real-valued continuous functions defined on T equipped with the norm

$$\|f\| = \max_{t \in T} |f(t)| \quad \text{for } f \in C(T).$$

Let ϕ be a given map from \hat{Q} into $C(T)$. We assume that ϕ is continuous and twice Neustadt differentiable on \hat{Q} . A fixed element $b \in C(T)$ being given, consider the following problem

$$\left. \begin{array}{l} \text{minimize } \|\phi(x) - b\| \\ \text{subject to} \\ x \in Q. \end{array} \right\} \quad (7.1)$$

Let B_0 be the unit sphere in $C(T)$, i. e.,

$$\begin{aligned} B_0 &= \{y \in C(T) \mid \|y\| \leq 1\} \\ &= \{y \in C(T) \mid |y(t)| \leq 1 \quad \forall t \in T\}, \end{aligned}$$

and let $R_+ = \{r \in \mathbb{R} \mid r > 0\}$. Consider another problem

$$\left. \begin{array}{l} \text{minimize } r \\ \text{subject to} \\ \frac{1}{r}(\phi(x) - b) \in B_0 \\ (x, r) \in Q \times R_+. \end{array} \right\} \quad (7.2)$$

If x_0 is a feasible solution to (7.1), then we may assume that $\|\phi(x_0) - b\| > 0$. For, if otherwise, the problem (7.1) is trivial. Keeping this convention in mind, we see that the problem (7.1) is equivalent to the problem (7.2) in the sense that x_0 is a local minimum solution to (7.1) if and only if (x_0, r_0) with $r_0 = \|\phi(x_0) - b\|$ is a local minimum solution to (7.2). For this reason we shall be concerned with the problem (7.2) hereafter. Define $f: \hat{Q} \times R_+ \rightarrow \mathbb{R}$ and $g: \hat{Q} \times R_+ \rightarrow C(T)$ by

$$\begin{aligned} f(x, r) &= r, \\ g(x, r) &= \frac{1}{r}(\phi(x) - b) \end{aligned} \quad \text{for } (x, r) \in \hat{Q} \times R_+, \quad (7.3)$$

respectively. Then the problem (7.2) is expressed in the following canonical form.

$$\left. \begin{array}{l} \text{minimize } f(x, r) \\ \text{subject to} \\ g(x, r) \in B_0 \\ (x, r) \in Q \times R_+. \end{array} \right\} \quad (7.4)$$

First we shall calculate Neustadt derivatives of the objective functional and constraint functional in (7.2). Let (x_0, r_0) be arbitrary but fixed feasible solution to (7.2). Let (x_1, r_1) be any point of $X \times R$. Then it is easily verified that

$$\begin{aligned} f^{(1)}((x_0, r_0); (x_1, r_1)) &= r_1, \\ g^{(1)}((x_0, r_0); (x_1, r_1)) &= \frac{1}{r_0^2} [r_0 \phi^{(1)}(x_0; x_1) - (\phi(x_0) - b)r_1]. \end{aligned} \quad (7.5)$$

As we shall see later, it suffices to obtain the second-order Neustadt derivatives of the functionals with respect to (x_1, r_1) with $r_1=0$. Let (x_2, r_2) be any point of $X \times R$. It is clear that

$$f^{(2)}((x_0, r_0), (x_1, 0); (x_2, r_2)) = r_2.$$

An easy calculation shows that for any point $(x, r) \in X \times R$ and for any $\lambda > 0$,

$$\begin{aligned} & \frac{1}{\lambda^2} [g(x_0 + \lambda x_1 + \lambda^2 x, r_0 + \lambda^2 r) - g(x_0, r_0) - \lambda g^{(1)}((x_0, r_0); (x_1, 0))] \\ &= \frac{r}{r_0(r_0 + \lambda^2 r)} b + \frac{1}{r_0 \lambda^2} [\phi(x_0 + \lambda x_1 + \lambda^2 x) - \phi(x_0) - \lambda \phi^{(1)}(x_0; x_1)] \\ & \quad - \frac{r}{r_0(r_0 + \lambda^2 r)} \phi(x_0 + \lambda x_1 + \lambda^2 x). \end{aligned} \quad (7.6)$$

Taking limit as $x \rightarrow x_2$, $r \rightarrow r_2$, $\lambda \downarrow 0$ on both sides of (7.6), we have

$$g^{(2)}((x_0, r_0), (x_1, 0); (x_2, r_2)) = \frac{1}{r_0^2} [r_0 \phi^{(2)}(x_0, x_1; x_2) - r_2 (\phi(x_0) - b)] \quad (7.7)$$

since ϕ is continuous and twice Neustadt differentiable by assumption.

Next we shall calculate the variational sets to the unit sphere B_0 by the help of Theorem 4.1. Let y be a point of $C(T)$. It is easily seen that if $y \in B_0^\circ$ then

$$d_{B_0}(y) = \max \left\{ \max_{t \in T} y(t) - 1, -1 - \min_{t \in T} y(t) \right\},$$

and if $y \in B_0$ then

$$d_{B_0^c}(y) = -\max \left\{ \max_{t \in T} y(t) - 1, -1 - \min_{t \in T} y(t) \right\}.$$

From this results we get

$$A_{B_0}(y) = \max_{t \in T} |y(t)| - 1 = \|y\| - 1 \quad \text{for } y \in C(T). \quad (7.8)$$

Let y_0 be any fixed point of $C(T)$. For any $y \in C(T)$ we define

$$I(y) = \{t \in T \mid |y(t)| = \|y\|\} \quad (7.9)$$

$$J(y_0; y) = \{t \in T \mid y(t) = 0\} \cap I(y_0). \quad (7.10)$$

Define "sgn" by

$$\text{sgn } a = \begin{cases} +1 & \text{if } a > 0 \\ -1 & \text{if } a < 0 \\ 0 & \text{if } a = 0. \end{cases}$$

PROPOSITION 7.1. *Let y_0 be a point of $C(T)$ such that $\|y_0\| > 0$. Then*

$$(i) \quad \begin{aligned} \bar{J}_{B_0}^{(1)}(y_0; y) &= \underline{J}_{B_0}^{(1)}(y_0; y) = \mathcal{A}'_{B_0}(y_0; y) \\ &= \max_{t \in I(y_0)} y(t) \operatorname{sgn} y_0(t) \quad \forall y \in C(T). \end{aligned} \quad (7.11)$$

(ii) *Let y_1 be a point of $C(T)$ such that*

$$\max_{t \in I(y_0)} y_1(t) \operatorname{sgn} y_0(t) = 0. \quad (7.12)$$

Assume that there exist a positive number λ_0 and a neighborhood N of y_1 in $C(T)$ such that

$$y_1(t) \operatorname{sgn} y_0(t) \leq 0 \quad \forall t \in \bigcup_{0 \leq \lambda \leq \lambda_0} \bigcup_{y \in N} \{I(y_0 + \lambda y)\}. \quad (7.13)$$

Then it holds that

$$\begin{aligned} \bar{J}_{B_0}^{(2)}(y_0, y_1; y) &= \underline{J}_{B_0}^{(2)}(y_0, y_1; y) = \mathcal{A}''_{B_0}(y_0, y_1; y) \\ &= \max_{t \in J(y_0; y_1)} y(t) \operatorname{sgn} y_0(t) \quad \forall y \in C(T). \end{aligned} \quad (7.14)$$

PROOF. First we note that the constant term (-1) in (7.8) can be omitted in order to calculate the derivatives of \mathcal{A}_{B_0} .

(i) The last equality in (7.11) is easily derived from Theorem 6.1 with letting $g(t, y) = |y(t)|$ for $(t, y) \in T \times C(T)$. The first two equalities in (7.11) are obvious, since the directional derivative $\mathcal{A}'_{B_0}(y_0; y)$ does exist.

(ii) The last equality in (7.14) is derived from Theorem 6.2 with $g(t, y)$ being the same as above. The first two equalities in (7.14) are obvious from the existence of $\mathcal{A}''_{B_0}(y_0, y_1; y)$ with y_1 satisfying (7.12). In this connection, note that (7.12) is identical with the relation

$$\mathcal{A}'_{B_0}(y_0; y_1) = 0. \quad \square$$

PROPOSITION 7.2. *Let y_0 be a point of $C(T)$ such that $\|y_0\| = 1$. Then*

$$(i) \quad F(B_0; y_0) = \{y \in C(T) \mid y(t)y_0(t) < 0 \quad \forall t \in I(y_0)\}.$$

$$(ii) \quad \begin{aligned} V(B_0; y_0) &= T(B_0; y_0) \\ &= \{y \in C(T) \mid y(t)y_0(t) \leq 0 \quad \forall t \in I(y_0)\}. \end{aligned}$$

PROOF. (i) is immediate from the combination of Proposition 7.1 (i) and Theorem 4.1 (i) with $n=1$, and (ii) from the combination of Proposition 7.1 (i) and Theorem 4.1 (ii), (iii) with $n=1$. \square

PROPOSITION 7.3. *Let y_0 be a point of $C(T)$ such that $\|y_0\| = 1$, and let y_1 be a point of $C(T)$. Then we have the following.*

(i) *if $\max_{t \in I(y_0)} y_1(t)y_0(t) > 0$, then $F(B_0; y_0, y_1) = V(B_0; y_0, y_1) = T(B_0; y_0, y_1) = \emptyset$.*

(ii) *if $\max_{t \in I(y_0)} y_1(t)y_0(t) < 0$, then $F(B_0; y_0, y_1) = V(B_0; y_0, y_1) = T(B_0; y_0, y_1) = C(T)$.*

(iii) *if $\max_{t \in I(y_0)} y_1(t)y_0(t) = 0$, and if there exists a positive number λ_0 and a neighborhood N of y_1 with respect to the maximum norm in $C(T)$ for which (7.13) holds, then*

$$F(B_0; y_0, y_1) = \{y \in C(T) \mid y(t)y_0(t) < 0 \quad \forall t \in J(y_0; y_1)\}, \quad (7.15)$$

$$\begin{aligned} V(B_0; y_0, y_1) &= T(B_0; y_0, y_1) \\ &= \{y \in C(T) \mid y(t)y_0(t) \leq 0 \quad \forall t \in J(y_0; y_1)\}. \end{aligned} \quad (7.16)$$

PROOF. (i) From the assumption and Proposition 7.1 (i), it follows that $\bar{J}_{B_0}^{(1)}(y_0; y_1) = \underline{J}_{B_0}^{(1)}(y_0; y_1) > 0$. Hence, by Proposition 3.3, $\bar{J}_{B_0}^{(2)}(y_0, y_1; \cdot) = \underline{J}_{B_0}^{(2)}(y_0, y_1; \cdot) \equiv +\infty$. From Theorem 4.1 we have $F(B_0; y_0, y_1) = V(B_0; y_0, y_1) = T(B_0; y_0, y_1) = \emptyset$. The proof of (ii) is similar to that of (i).

(iii) Obvious from Proposition 7.1 and Theorem 4.1. \square

We now return to the approximation problem (7.1). For any $x_0 \in Q$, we let

$$E(x_0) = \{t \in T; \|\phi(x_0)(t) - b(t)\| = \|\phi(x_0) - b\|\}, \quad (7.17)$$

and

$$\theta_t(x_0) = \text{sgn}(\phi(x_0)(t) - b(t)), \quad t \in T. \quad (7.18)$$

The following gives a first-order necessary condition for a solution of (7.1).

THEOREM 7.1. *Let x_0 be a local minimum solution of the problem (7.1). Then it holds that*

$$\max_{t \in E(x_0)} \theta_t(x_0) \phi^{(1)}(x_0; x)(t) \geq 0 \quad \forall x \in T(Q; x_0). \quad (7.19)$$

PROOF. Put $r_0 = \|\phi(x_0) - b\|$. For the mapping g defined by (7.3), from Proposition 7.2 (i) and (7.5) we have

$$\begin{aligned} g^{(1)}((x_0, r_0); (x, r)) &\in F(B_0; g(x_0, r_0)) \\ &\Leftrightarrow g^{(1)}((x_0, r_0); (x, r))(t) \cdot g(x_0, r_0)(t) < 0 \quad \forall t \in E(x_0) \\ &\Leftrightarrow \{r_0 \phi^{(1)}(x_0; x)(t) - (\phi(x_0)(t) - b(t))r\} \cdot (\phi(x_0)(t) - b(t)) < 0 \quad \forall t \in E(x_0) \\ &\Leftrightarrow \max_{t \in E(x_0)} \theta_t(x_0) \phi^{(1)}(x_0; x)(t) < r. \end{aligned}$$

Hence, by applying Theorem 5.1 to the problem (7.2) which is equivalent to (7.1), we have that the system of (x, r) :

$$\begin{aligned} \max_{t \in E(x_0)} \theta_t(x_0) \phi^{(1)}(x_0; x)(t) &< r < 0, \\ x &\in T(Q; x_0), \end{aligned}$$

is inconsistent. The inconsistency of this system is equivalent to (7.19). \square

As a preparation for the second-order theorem, we shall examine the conditions (R1)~(R3) in Section 5 more closely for the problem (7.2). Naturally we are concerned with those conditions in the case that $n=2$. By Remark 5.1 we have

$$(R1) \Leftrightarrow r_1 \leq 0.$$

From (7.5) and Proposition 7.2 (ii),

$$(R2) \Leftrightarrow \max_{t \in E(x_0)} \theta_t(x_0) \phi^{(1)}(x_0; x_1)(t) \leq r_1.$$

Since $r_0 > 0$, $T(Q \times R_+; (x_0, r_0)) = T(Q; x_0) \times R$. Hence

$$(R3) \Leftrightarrow x_1 \in T(Q; x_0).$$

Thus we have

$$(R1) \sim (R3) \Leftrightarrow \left\{ \begin{array}{l} \max_{t \in E(x_0)} \theta_t(x_0) \phi^{(1)}(x_0; x_1)(t) \leq r_1 \leq 0 \\ x_1 \in T(Q; x_0) \end{array} \right\}.$$

By the help of Theorem 7.1, finally we get

$$(R1) \sim (R3) \Leftrightarrow \left\{ \begin{array}{l} r_1 = 0 \\ \max_{t \in E(x_0)} \theta_t(x_0) \phi^{(1)}(x_0; x_1)(t) = 0 \\ x_1 \in T(Q; x_0) \end{array} \right\}. \quad (7.20)$$

Now we define

$$\mathcal{E}[y] = \{t \in T; |y(t)| = \|y\|\} \quad \text{for } y \in C(T).$$

Obviously $\mathcal{E}[\phi(x_0) - b] = E(x_0)$. We define moreover

$$K(x_0; x_1) = \{t \in T; \phi^{(1)}(x_0; x_1)(t) = 0\} \quad \text{for } x_1 \in X.$$

Then we have the following

THEOREM 7.2. *Let x_0 be a local minimum solution of the problem (7.1). Then for every $x_1 \in T(Q; x_0)$ satisfying the condition*

$$\max_{t \in E(x_0)} \theta_t(x_0) \phi^{(1)}(x_0; x_1)(t) = 0 \quad (7.21)$$

and the condition

$$(D): \left\{ \begin{array}{l} \text{there exist a number } \lambda_0 > 0 \text{ and a neighborhood } V \\ \text{of } \phi^{(1)}(x_0; x_1) \text{ in } C(T) \text{ such that} \\ \theta_t(x_0) \phi^{(1)}(x_0; x_1)(t) \leq 0 \quad \forall t \in \bigcup_{0 \leq \lambda < \lambda_0} \bigcup_{y \in V} \mathcal{E}[\phi(x_0) - b + \lambda y], \end{array} \right. \quad (7.22)$$

it holds that

$$\max_{t \in E(x_0) \cap K(x_0; x_1)} \theta_t(x_0) \phi^{(2)}(x_0, x_1; x)(t) \geq 0 \quad \forall x \in T(Q; x_0, x_1). \quad (7.23)$$

PROOF. The assertions can be derived from Theorem 5.2 with $n=2$. It can be seen from Proposition 5.1 that we may restrict the points (x_1, r_1) to those satisfying (7.20). So we may put $r_1=0$.

Put $r_0 = \|\phi(x_0) - b\|$. For the function f defined by (7.3), an easy calculation shows that

$$\begin{aligned} & F(f(x_0, r_0) + R_-; f(x_0, r_0), f^{(1)}((x_0, r_0); (x_1, 0))) \\ &= F(r_0 + R_-; r_0, 0) \\ &= F(r_0 + R_-; r_0) = R_-. \end{aligned} \quad (\text{by Proposition 2.1 (ii)})$$

Hence, for the problem (7.2) the formula (5.7) with $n=2$ can be written as $r < 0$. Next, for the mapping g defined by (7.3) we shall calculate the variational set $F(B_0; g(x_0, r_0), g^{(1)}((x_0, r_0); (x_1, 0)))$. From (7.5) we have

$$g^{(1)}((x_0, r_0); (x_1, 0)) = \frac{1}{r_0} \phi^{(1)}(x_0; x_1).$$

Therefore, if we put $y_0 = g(x_0, r_0)$ and $y_1 = g^{(1)}((x_0, r_0); (x_1, 0))$ in Proposition 7.3, it is easily seen that all assumptions in (iii) of the proposition are satisfied by virtue of (7.21) and (D). Thus the relation (7.15) immediately leads us to

$$\begin{aligned} & F(B_0; g(x_0, r_0), g^{(1)}((x_0, r_0); (x_1, 0))) \\ &= \{y \in C(T) \mid y(t) \cdot (\phi(x_0)(t) - b(t)) < 0 \quad \forall t \in E(x_0) \cap K(x_0; x_1)\}. \end{aligned} \quad (7.24)$$

From (7.7) and (7.24) we obtain the following relation:

$$\begin{aligned} & g^{(2)}((x_0, r_0), (x_1; 0); (x, r)) \in F(B_0; g(x_0, r_0), g^{(1)}((x_0, r_0); (x_1, 0))) \\ & \Leftrightarrow \max_{t \in E(x_0) \cap K(x_0; x_1)} \theta_t(x_0) \phi^{(2)}(x_0, x_1; x)(t) < r. \end{aligned}$$

Viewing the above results, we apply Theorem 5.2 together with Proposition 5.1 to the problem (7.2) to yield that for every $x_1 \in T(Q; x_0)$ satisfying (7.21) and (D) the system of (x, r)

$$\begin{cases} r < 0 \\ \max_{t \in E(x_0) \cap K(x_0; x_1)} \theta_t(x_0) \phi^{(2)}(x_0, x_1; x)(t) < r \\ x \in T(Q; x_0, x_1) \end{cases}$$

is inconsistent. From this inconsistency we get the desired result (7.23). This completes the proof. \square

COROLLARY 7.1. *Let x_0 be a local minimum solution of the problem (7.1). Suppose that ϕ is twice Fréchet differentiable at x_0 . Then for every $x_1 \in T(Q; x_0)$ satisfying that*

$$\max_{t \in E(x_0)} \theta_t(x_0) (D\phi(x_0)x_1)(t) = 0$$

and the condition

$$(D_1): \begin{cases} \text{there exist a number } \lambda_0 > 0 \text{ and a neighborhood } V \\ \text{of } D\phi(x_0)x_1 \text{ such that} \\ \theta_t(x_0) (D\phi(x_0)x_1)(t) \leq 0 \quad \forall t \in \bigcup_{0 \leq \lambda < \lambda_0} \bigcup_{y \in V} \mathcal{E}[\phi(x_0) - b + \lambda y], \end{cases}$$

it holds that

$$\begin{aligned} & \max_{t \in E(x_0) \cap K_1(x_0; x_1)} \theta_t(x_0) \left\{ (D\phi(x_0)x)(t) + \frac{1}{2} (D^2\phi(x_0)(x_1, x_1))(t) \right\} \geq 0 \\ & \quad \forall x \in T(Q; x_0, x_1), \end{aligned} \quad (7.25)$$

where $K_1(x_0; x_1) = \{t \in T; (D\phi(x_0)x_1)(t) = 0\}$.

REMARK 7.1. In the case when ϕ is Fréchet differentiable, the inequality (7.19) coincides with the so-called local Kolmogoroff condition. Therefore (7.19) in the original form is a nondifferentiable extension of the local Kolmogoroff condition. On the other hand, the inequalities (7.23) and (7.25) have not appeared in the literatures of approximation theory treating first-order criteria. We can consider them the second-order versions of the local Kolmogoroff condition.

COROLLARY 7.2. *Suppose that in the problem (7.1) T is a finite set. Let x_0 be a local minimum solution of the problem. Then*

- (i) *for every $x_1 \in T(Q; x_0)$ satisfying (7.21), the relation (7.23) holds,*
- (ii) *if Q is a convex set, then for every $y \in Q$ satisfying that*

$$\max_{t \in E(x_0)} \theta_t(x_0) \phi^{(1)}(x_0; y - x_0) = 0,$$

it holds that

$$\max_{t \in E(x_0) \cap K(x_0; y - x_0)} \theta_t(x_0) \phi^{(2)}(x_0, y - x_0; x - y) \geq 0 \quad \forall x \in Q.$$

PROOF. The assertion (i) follows from the proof of Theorem 7.2 considering Remark 6.2, and (ii) follows from (i) and Proposition 2.3. \square

References

- [1] BEN-TAL, A., TEBoulLE, M. and ZOWE, J.: *Second order necessary optimality conditions for semi-infinite programming problems*. In: Hettich R(ed) *Semi-infinite programming*, Lecture notes in control and information science 15 (1979), 17-30, Springer-Verlag Berlin-Heidelberg-New York.
- [2] BEN-TAL, A. and ZOWE, J.: *A unified theory of first and second order conditions for extremum problems in topological vector spaces*. *Mathematical Programming Study* 19 (1982), 39-76.
- [3] DUBOVITSKII, A.Y. and MILYUTIN, A.A.: *The extremum problem in the presence of constraints*. *Dokl Akad Nauk SSSR* 149 (1963), 759-762.
- [4] FURUKAWA, N.: *Optimality conditions in nondifferentiable programming and their applications to best approximations*. *Appl Math Optim* 9 (1983), 337-371.
- [5] GIRSANOV, I.V.: *Lectures on mathematical theory of extremum problems*. Lecture notes in economic and mathematical systems 67 (1972).
- [6] HOFFMANN, K.H. and KORNSTAEDT, H.J.: *Higher-order necessary conditions in abstract mathematical programming*. *J. Optim Theory Appl.*, 26 (1978), 533-568.
- [7] KRABS, W.: *Optimization and approximation*. (English translation), John Wiley & Sons, New York, (1979).
- [8] LAURENT, J.P.: *Optimisation et approximation*. Herman, Paris, (1972).
- [9] LINNEMAN, A.: *Higher-order necessary conditions for infinite optimization*. *J. Optim Theory Appl.*, 38 (1982), 483-512.
- [10] NEUSTADT, L.W.: *An abstract variational theory with applications to a broad class of optimization problems I, general theory*. *SIAM J. Contr.*, 4 (1966), 505-527.

Received July 1, 1987