JOURNAL OF APPROXIMATION THEORY 9, 316–326 (1973)

Nonlinear Optimization and Approximation

WERNER KRABS

Department of Mathematics, Michigan State University and Fachbereich Mathematik der Technischen Hochschule Darmstadt, Darmstadt, West Germany

Communicated by E. W. Cheney

Received May 18, 1971

This paper is concerned with nonlinear optimization problems in normed linear spaces. Necessary and sufficient conditions for optimal points are given and the range of applicability of these conditions is studied.

The results are applied to nonlinear approximation problems.

1. INTRODUCTION

Let E be a normed linear space over the real or complex numbers. Let X be a nonempty subset of E and $\hat{X}$ an open subset of E containing X. We consider a family $\{g_j: j \in I\}$ of functionals $g_j: \hat{X} \to R$ (= set of real numbers) and assume that the set

$$S = \{x \in X: g_j(x) \leqslant 0 \text{ for all } j \in I\} \tag{1.1}$$

is nonempty. We shall be concerned with the problem of minimizing a given functional $f: \hat{X} \to R$ on S. Throughout this paper we assume the following conditions to hold:

(a) I is a compact Hausdorff space (if I is a finite set we assume that it is provided with the discrete topology).

(b) For each $x \in \hat{X}$ the function $j \to g_j(x)$ is continuous as a function mapping I into R.

(c) Let $C(I)$ be the vector space of the real valued continuous functions on I with the maximum norm. Then the function $g: \hat{X} \to C(I)$ defined by $g(x) = (g_j(x))_{j \in I}$ is Fréchet differentiable, i.e., for each $j \in I$ there exists a linear functional $g'_{jx}: E \to R$ such that

$$\max_{j \in I} | g_j(x + h) - g_j(x) - g'_{jx}(h)| \leqslant \| h \| \cdot \alpha(\| h \|)$$

where

$$\lim_{\|h\|\to 0} \alpha(\| h \|) = 0,$$

and for each $x \in \hat{X}$ the function $(j, h) \to g'_{jx}(h)$ is continuous.

(d) $f: \hat{X} \to R$ is Fréchet differentiable, i.e., for each $x \in \hat{X}$ there exists a continuous linear functional $f_x': E \to R$ such that

$$|f(x + h) - f(x) - f_x'(h)| \leqslant \| h \| \cdot \epsilon(\| h \|)$$

where

$$\lim_{\|h\|\to 0} \epsilon(\| h \|) = 0.$$

The functionals f_x' and g'_{jx} are called the Fréchet derivatives at x. In the following we shall be mainly interested in establishing necessary and sufficient conditions for an element $\hat{x} \in S$ to be optimal, i.e., to satisfy

$$f(\hat{x}) \leqslant f(x) \qquad \text{for all} \quad x \in S \tag{1.2}$$

and to study the range of applicability of these conditions. During the following investigations the concept of tangent cones due to Hestenes [3] and Abadie [1] will play a fundamental role.

DEFINITION. Let Y be an arbitrary nonempty subset of E. Then a vector $h \in E$ will be called a tangent vector of Y at a given $y \in Y$ if there exists a sequence $\{y_k\}$ of points $y_k \in Y$ and a sequence $\{\lambda_k\}$ of positive real numbers λ_k such that

$$y = \lim_{k\to\infty} y_k \qquad \text{and} \qquad h = \lim_{k\to\infty} \lambda_k(y_k - y).$$

We denote by $T(Y, y)$ the set of all tangent vectors of Y at y. Obviously $T(Y, y)$ is nonempty since $\theta_E \in T(Y, y)$ where θ_E is the zero element of E. Furthermore the following implication holds

$$\lambda \geqslant 0, \qquad h \in T(Y, y) \Rightarrow \lambda \cdot h \in T(Y, y).$$

Therefore, $T(Y, y)$ is called the *tangent cone of Y at y*. In general $T(Y, y)$ is not convex. However, it is well known (cf. [1, 3]) that $T(Y, y)$ is closed.

EXAMPLES. (1) If Y is a nonempty open subset of E then we have $T(Y, y) = E$ for any $y \in Y$.

(2) If Y is a nonempty convex subset of E then it can be proved that

$$T(Y, y) = \overline{\bigcup_{\lambda>0} \{\lambda(x - y) : x \in Y\}}, \tag{1.3}$$

where $\bar{A}$ denotes the closure of A.

(3) If, for instance, Y is a linear submanifold of E then f we have $Y = y + V$ where V is a linear subspace of E and $T(Y, y) = \overline{V}$.

To obtain necessary conditions for optimal elements in S well known theorem [1, 3, 6] will be used.

THEOREM 1.1. *Let S be an arbitrary nonempty subset of E a subset of E which contains S. If f: $\hat{X} \to R$ is Fréchet different $\hat{x} \in S$ to satisfy* (1.2) *the following condition is necessary*:

$$f_{\hat{x}}'(h) \geqslant 0 \qquad \textit{for all} \quad h \in T(S, \hat{x}).$$

This condition contains a series of classical conditions:

(1) If $S = \hat{X}$ then (1.4) is equivalent to $f_{\hat{x}}' \equiv 0$.

(2) If S is convex then (1.4) is equivalent to $f_{\hat{x}}'(x - \hat{x}$ $x \in S$.

(3) If S is a linear submanifold of E, say $S = \hat{x} + V$, linear subspace of E then (1.4) is equivalent to $f_{\hat{x}}'(h) = 0$ for a

In Section 2 we introduce the concept of regular points x Theorem 1.1 we give a necessary condition for a regular poir optimal (Theorem 2.2). Under the assumption that all the po regular we then establish a sufficient condition for a point $\hat{x} \in S$ (Theorem 2.3). In order to insure that both conditions are nec as sufficient for optimal points we assume the so called al property (Theorem 2.4). Finally, we study the range of applic property (Theorem 2.5).

In Section 3 we apply the results to nonlinear approximati In this case it turns out that all the points of S are regular almost-convex-property is implied by a condition which is, satisfied for linear, rational, and exponential approximation.

2. NECESSARY AND SUFFICIENT CONDITIONS FOR OPTIMAL

Let the subset S of E be defined by (1.1). To each $x \in S$ \ following subset of I:

$$I(x) = \{j \in I: g_j(x) = 0\}.$$

$I(x)$ is a closed subset of I which may be empty in which case

$$g_j(x) < 0 \qquad \text{for all} \quad j \in I.$$

DEFINITION. A point $x \in S$ is called a regular point of S if either $I(x)$ is empty or the set

$$l(S, x) = \bigcap_{j \in I(x)} \{h \in T(X, x) : g'_{jx}(h) < 0\} \tag{2.3a}$$

is nonempty where $T(X, x)$ denotes the tangent cone of X at x. In the case where $I(x)$ is empty we define

$$l(S, x) = T(X, x). \tag{2.3b}$$

Then we have the following theorem.

THEOREM 2.1. *If $x \in S$ is regular then $l(S, x)$ is contained in $T(S, x)$ and there is a sequence $\{x_k\}$ of points $x_k \in S$ such that*

$$x = \lim_{k \to \infty} x_k \qquad \text{and} \qquad I(x_k) \text{ is empty for all } k. \tag{2.4}$$

Proof. (a) Let $I(x)$ be empty. Then for the second part of the assertion we can take $x_k = x$ for all k. We have to prove that $T(X, x) \subseteq T(S, x)$. Therefore, we consider $h \in T(X, x)$. The case $h = \theta_E$ is trivial. Hence, let $h \neq \theta_E$. Let $\{x_k\}$ and $\{\lambda_k\}$, $x_k \in X$ and $\lambda_k > 0$, be such that

$$x = \lim_{k \to \infty} x_k \qquad \text{and} \qquad h = \lim_{k \to \infty} \lambda_k(x_k - x). \tag{*}$$

If we put $h_k = \lambda_k(x_k - x)$ then we have $x_k = x + (1/\lambda_k)\, h_k$ which implies $\lim_{k \to \infty} (1/\lambda_k) \cdot h_k = \theta_E$ and $\lim_{k \to \infty} \lambda_k = \infty$. For any $j \in I$ we have

$$g_j(x_k) \leqslant g_j(x) + (1/\lambda_k)\{g'_{jx}(h_k) + \| h_k \| \, \alpha(1/\lambda_k \| h_k \|)\},$$

where $\lim_{k \to \infty} \alpha(1/\lambda_k \| h_k \|) = 0$.

Due to the assumptions (a)–(c) in Section 1 there exist numbers $\delta > 0$ and $m > 0$ such that

$$g_j(x) \leqslant -\delta < 0 \qquad \text{for all} \quad j \in I$$

and

$$g'_{jx}(h_k) - \| h_k \| \, \alpha(1/\lambda_k \| h_k \|) \leqslant m \qquad \text{for all } j \in I \text{ and all } k.$$

Hence, we obtain

$$g_j(x_k) \leqslant -\delta + (1/\lambda_k)m \leqslant 0 \qquad \text{for all} \quad j \in I \text{ and all } k \text{ such that } \lambda_k \geqslant m/\delta.$$

This implies $x_k \in S$ for sufficiently large values of k which in turn implies $h \in T(S, x)$.

(b) Let $I(x)$ be nonempty. Let $h \in l(S, x)$. Then $h \neq \theta_E$ and there exist sequences $\{x_k\}$ and $\{\lambda_k\}$, $x_k \in X$, $\lambda_k > 0$, such that (*) holds. Putting $h_k = \lambda_k(x_k - x)$ we again have

$$x_k = x + (1/\lambda_k)\, h_k\,, \qquad \lim_{k\to\infty} (1/\lambda_k)\, h_k = \theta_E \qquad \text{and} \qquad \lim_{k\to\infty} \lambda_k = \infty.$$

Let $g'_{jx}(h) \leqslant -\delta < 0$ for all $j \in I(x)$ and put $I(x, h) = \{j \in I\colon g'_{jx}(h) < -\delta/2\}$. Now we have as above

$$g_j(x_k) \leqslant g_j(x) + (1/\lambda_k)\{g'_{jx}(h_k) + \| h_k \|\, \alpha(1/\lambda_k \| h_k \|)\},$$

where $\lim_{k\to\infty} \alpha(1/\lambda_k \| h_k \|) = 0$. If we choose k such that

$$g'_{jx}(h_k) < -\delta/4 \qquad \text{for all} \quad j \in I(x, h) \quad \text{and} \quad \| h_k \|\, \alpha(1/\lambda_k \| h_k \|) \leqslant \delta/4$$

then we obtain $g_j(x_k) < 0$ for all $j \in I(x, h)$. If $I(x, h) = I$ we have $x_k \in S$ and $I(x_k)$ empty whenever k is sufficiently large which implies $h \in T(S, x)$ and completes the proof. If $I(x, h) \neq I$ then $I - I(x, h)$ is compact and there are numbers $n > 0$ and $m > 0$ such that

$$g_j(x) \leqslant -n \qquad \text{for all} \quad j \in I - I(x, h)$$

and

$$g'_{jx}(h_k) + \| h_k \|\, \alpha(1/\lambda_k \| h_k \|) \leqslant m \qquad \text{for all} \quad j \in I - I(x, h) \quad \text{and all } k.$$

Hence, for k sufficiently large we have $g_j(x_k) \leqslant -n + (1/\lambda_k)m < 0$ for all $j \in I - I(x, h)$. So again we obtain $x_k \in S$, $I(x_k)$ empty for k sufficiently large which implies $h \in T(S, x)$ and completes the proof.

An immediate consequence of Theorems 1.1 and 2.1 is the following theorem.

THEOREM 2.2. *If $\hat{x} \in S$ is optimal, i.e., $\hat{x}$ satisfies* (1.2) *and if $\hat{x}$ is regular then we have*

$$f_{\hat{x}}'(h) \geqslant 0 \qquad \textit{for all} \quad h \in l(S, \hat{x}), \tag{2.5}$$

where $l(S, \hat{x})$ is defined by (2.3a) *or* (2.3b).

If we define $g_0 = f$ and $I_0(\hat{x}) = I(\hat{x}) \cup \{0\}$ then (2.5) is equivalent to the following statement:

$$\max_{j\in I_0(\hat{x})} g'_{j\hat{x}}(h) \geqslant 0 \qquad \text{for all} \quad h \in T(X, \hat{x}). \tag{2.6}$$

Proof. If $I(\hat{x})$ is empty then (2.6) is the same condition as (2.5) by virtue of (2.3b). If $I(\hat{x})$ is nonempty there are two cases to be distinguished:

(a) $h \in l(S, \hat{x})$. Then (2.6) follows from (2.5) and (2.3a).

(b) $h \in T(X, \hat{x})$, but $h \notin l(S, \hat{x})$. Then there is an index $j \in I(\hat{x})$ such that $g'_{j\hat{x}}(h) \geqslant 0$ which also implies (2.6). The implication (2.6) $\Rightarrow$ (2.5) is clear.

The following theorem states a sufficient condition for optimal elements.

THEOREM 2.3. *We assume all the points x of S to be regular. Let $\hat{x} \in S$ be such that*

$$\max_{j \in I_0(\hat{x})} \{g_j(x) - g_j(\hat{x})\} \geqslant 0 \qquad \textit{for all} \quad x \in X, \tag{2.7}$$

where g_0 and $I_0(\hat{x})$ are defined as above then $\hat{x}$ is optimal, i.e., $\hat{x}$ satisfies (1.2).

Proof. If $I(\hat{x})$ is empty, then (2.7) implies $f(x) \geqslant f(\hat{x})$ for all $x \in X$ which in turn implies (1.2).

If $I(\hat{x})$ is nonempty there are two cases to be distinguished:

(a) Let $x \in S$ be such that $g_j(x) < 0$ for all $j \in I(\hat{x})$. Then (2.7) implies $f(x) \geqslant f(\hat{x})$.

(b) Let $x \in S$ be such that $g_j(x) = 0$ for at least one $j \in I(\hat{x})$. Since x is regular by Theorem 2.1 there exists a sequence $\{x_k\}$ of points $x_k \in S$ such that (2.4) holds. This implies $\max_{j \in I(\hat{x})} \{g_j(x_k) - g_j(\hat{x})\} < 0$, and, hence, $f(x_k) \geqslant f(\hat{x})$ for all k. Since f is continuous on $\hat{X}$ (as a Fréchet differentiable function) we obtain $f(x) \geqslant f(\hat{x})$ which completes the proof.

To insure that both conditions (2.6) and (2.7) are necessary as well as sufficient for $\hat{x} \in S$ to satisfy (1.2) we make the following assumption which we will call the almost-convex-property: *For any closed subset $\hat{I}$ of I and any pair of points $x, \hat{x} \in X$ such that*

$$\max_{j \in \hat{I} \cup \{0\}} \{g_j(x) - g_j(\hat{x})\} < 0 \tag{2.8}$$

there exists an $h \in T(X, \hat{x})$ such that

$$\max_{j \in \hat{I} \cup \{0\}} g'_{j\hat{x}}(h) < 0. \tag{2.9}$$

Remark. If X is convex and f and all the g_j's, $j \in I$, are convex functionals on X, then the almost-convex-property holds.

THEOREM 2.4. *If all the points $x \in S$ are regular and the almost-convex-property holds then the conditions* (2.6) *and* (2.7) *are both necessary and sufficient for $\hat{x} \in S$ to satisfy* (1.2).

Proof. Obviously we have the implication (2.6) $\Rightarrow$ (2.7) yielding the result.

In order to study the range of applicability of the almost-convex-property we consider an arbitrary function $\alpha \in C(I)$ and define

$$S_\alpha = \{x \in X: g_j(x) + \alpha_j \leqslant 0 \text{ for all } j \in I\}. \tag{2.10}$$

To each $x \in S_\alpha$ we assign the index set

$$I_\alpha(x) = \{j \in I: g_j(x) + \alpha_j = 0\}. \tag{2.11}$$

We are now concerned with the problem of minimizing f on S_α, that is, to find an $\hat{x} \in S_\alpha$ such that

$$f(\hat{x}) \leqslant f(x) \qquad \text{for all} \quad x \in S_\alpha . \tag{2.12}$$

THEOREM 2.5. *For every function $\alpha \in C(I)$ we assume the condition* (2.6) *to be sufficient for $\hat{x} \in S_\alpha$ to satisfy* (2.12) *where $I(\hat{x})$ has to be replaced by $I_\alpha(\hat{x})$ defined by* (2.11). *Then the almost-convex-property holds.*

Proof. Let $\hat{I}$ be a closed subset of I and $x^*, \hat{x} \in X$ a pair such that

$$\max_{j \in \hat{I} \cup \{0\}} \{g_j(x^*) - g_j(\hat{x})\} < 0.$$

Then we define the set $\tilde{I} = \{j \in I: g_j(x^*) - g_j(\hat{x}) \geqslant 0\}$. If $\tilde{I}$ is empty we put

$$\alpha_j = -g_j(\hat{x}) \qquad \text{for all} \quad j \in I.$$

Then we obtain

$$g_j(\hat{x}) + \alpha_j = 0 \quad \text{for all} \quad j \in I \quad \text{and} \quad g_j(x^*) + \alpha_j < 0 \quad \text{for all} \quad j \in I.$$

Since $f(x^*) < f(\hat{x})$, $\hat{x} \in S_\alpha$ cannot satisfy (2.12). If $\tilde{I}$ is nonempty, then $\hat{I}$ and $\tilde{I}$ are disjoint. By virtue of Urysohn's lemma there is a function $\epsilon \in C(I)$ such that

$$\epsilon_j \begin{cases} = 0 & \text{for all } j \in \hat{I}, \\ \in (0, 1) & \text{for all } \; j \notin \hat{I} \cup \tilde{I}, \\ = 1 & \text{for all } j \in \tilde{I}. \end{cases}$$

If we put

$$\alpha_j = -g_j(x^*) - | g_j(x^*) - g_j(\hat{x})| - \epsilon_j$$

we obtain

$$g_j(\hat{x}) + \alpha_j \begin{cases} = 0 \text{ for all } j \in \hat{I} \\ < 0 \text{ for all } j \notin \hat{I} \end{cases} \Rightarrow I_\alpha(\hat{x}) = \hat{I}$$

and $g_j(x^*) + \alpha_j < 0$ for all $j \in I$.

Hence, again $\hat{x} \in S_\alpha$ cannot satisfy (2.12). Therefore, by assumption there exists an $h \in T(X, \hat{x})$ such that

$$\max_{j \in \hat{I} \cup \{0\}} g'_{j\hat{x}}(h) < 0,$$

which completes the proof.

3. Application to Nonlinear Approximation

Let Y be a nonempty subset of R^n and $\hat{Y}$ an open subset of R^n such that $Y \subseteq \hat{Y}$. Let I as above be a compact Hausdorff space and $C(I)$ the vector space of the real valued continuous functions on I. Let $\varphi: \hat{Y} \to C(I)$ be a given map and $f \in C(I)$ be a given function. We are concerned with the problem of finding a $\hat{y} \in Y$ such that

$$\| \varphi(\hat{y}) - f \| \leqslant \| \varphi(y) - f \| \qquad \text{for all} \quad y \in Y, \tag{3.1}$$

where $\| \cdot \|$ denotes the maximum norm of $C(I)$.

This problem is equivalent to the following problem of optimization: Under the conditions

$$(\varphi_j(y) - f_j)^2 - \gamma \leqslant 0 \qquad \text{for all} \quad j \in I, \tag{3.2}$$

$$y \in Y, \qquad \gamma \in R, \tag{3.3}$$

γ is to be minimized. (Every real valued function r on I is written as above in the form $r = r_j$, $j \in I$.)

If we define $E = R^{n+1}$, $\hat{X} = \hat{Y} \times R$, $X = Y \times R$,

$$g_j(x) = g_j(y, \gamma) = (\varphi_j(y) - f_j)^2 - \gamma, \qquad j \in I,$$

and $f(x) = f(y, \gamma) = \gamma$, then we obtain an optimization problem as studied above.

We assume that for each $j \in I$ and $y \in \hat{Y}$ the gradient vector $\nabla \varphi_j(y)$ exists and that the mapping $(j, y) \to \nabla \varphi_j(y)$ is continuous with respect to both variables. Then all the conditions (a)–(d) of Section 1 hold. The key for proving this is a result in [4] which states that under these assumptions φ is Fréchet differentiable.

In particular we have

$$g'_{j(y,\gamma)}(h, \lambda) = 2\{\varphi_j(y) - f_j\}[\nabla \varphi_j(y), h] - \lambda, \qquad j \in I,$$

$$f'_{(y,\gamma)}(h, \lambda) = \lambda$$

for all $(y, \gamma) \in \hat{Y} \times R$, $(h, \lambda) \in R^{n+1}$. ($[\cdot, \cdot]$ denotes the scalar product in R^n.) Furthermore we obtain $T(X, x) = T(Y, y) \times R$, where $x = (y, \gamma)$. To each $y \in Y$ we assign the index set

$$I^*(y) = \{j \in I : |\varphi_j(y) - f_j| = \| \varphi(y) - f \|\}. \tag{3.4}$$

LEMMA 3.1. *If we define S by*

$$S = \{(y, \gamma) \in Y \times R : (\varphi_j(y) - f_j)^2 - \gamma \leqslant 0 \textit{ for all } j \in I\} \tag{3.5}$$

then every point $(y, \gamma) \in S$ is regular.

Proof. There are two cases to be distinguished:

(a) $\gamma > \| \varphi(y) - f\|^2$. Then $I(y, \gamma)$ is empty and nothing has to be shown.

(b) $\gamma = \| \varphi(y) - f\|^2$. Then $I(y, \gamma)$ is nonempty and for each $h \in T(Y, y)$ there is a number $\lambda \in R$ such that

$$g'_{j(y,\gamma)}(h, \lambda) = 2\{\varphi_j(y) - f_j\}[\nabla\varphi_j(y), h] - \lambda < 0 \qquad \text{for all} \qquad j \in I(y, \gamma).$$

If $\hat{y} \in Y$ is a solution of the approximation problem and $\hat{\gamma} = \| \varphi(\hat{y}) - f\|^2$ then $(\hat{y}, \hat{\gamma})$ is a solution of the corresponding optimization problem. Hence, $I(\hat{y}, \hat{\gamma}) = I^*(\hat{y})$ is nonempty.

By applying Theorem 2.2 we, therefore, obtain the following result: If for $\hat{y} \in Y$ the condition (3.1) is satisfied, i.e., if $\hat{y}$ is a solution of the approximation problem then we have

$$\max[\max_{j \in I^*(\hat{y})} \{\varphi_j(\hat{y}) - f_j\}[\nabla\varphi_j(\hat{y}), h] - \lambda, \lambda] \geqslant 0$$

for all $\lambda \in R$ and all $h \in T(Y, \hat{y})$ which is equivalent to

$$\max_{j \in I^*(\hat{y})} \{\varphi_j(\hat{y}) - f_j\}[\nabla\varphi_j(\hat{y}), h] \geqslant 0 \qquad \text{for all} \quad h \in T(Y, \hat{y}). \tag{3.6}$$

Now we assume $\hat{y} \in Y$ to be given and put $\hat{\gamma} = \| \varphi(\hat{y}) - f\|^2$. Then again $I(\hat{y}, \hat{\gamma}) = I^*(\hat{y})$ is nonempty and the condition (2.7) has the form

$$\max[\max_{j \in I^*(\hat{y})} \{(\varphi_j(y) - f_j)^2 - (\varphi_j(\hat{y}) - f_j)^2 - \gamma + \hat{\gamma}\}, \gamma - \hat{\gamma}] \geqslant 0$$

for all $y \in Y$ and $\gamma \in R$ which is equivalent to

$$\max_{j \in I^*(\hat{y})} \{(\varphi_j(y) - f_j)^2 - (\varphi_j(\hat{y}) - f_j)^2\} \geqslant 0 \qquad \text{for all} \quad y \in Y.$$

Making use of the identity

$$(\varphi_j(y) - f_j)^2 - (\varphi_j(\hat{y}) - f_j)^2 \\ = [2(\varphi_j(\hat{y}) - f_j) - (\varphi_j(\hat{y}) - \varphi_j(y))](\varphi_j(y) - \varphi_j(\hat{y})), \tag{3.7}$$

we obtain that the last condition is a consequence of

$$\max_{j \in I^*(\hat{y})} (\varphi_j(\hat{y}) - f_j)(\varphi_j(y) - \varphi_j(\hat{y})) \geqslant 0 \quad \text{for all} \quad y \in Y. \tag{3.8}$$

Hence, by Theorem 2.3 we have the following statement: If for some $\hat{y} \in Y$ the condition (3.8) is fulfilled then $\hat{y}$ is a solution of the approximation problem, i.e., $\hat{y}$ satisfies (3.1).

The following condition guarantees that (3.6) implies (3.8) so that both conditions are necessary as well as sufficient for $\hat{y} \in Y$ to be a solution of the approximation problem.

ASSUMPTION. *For any pair $y, \hat{y} \in Y$ and any closed nonempty subset of $\hat{I}$ of I such that*

$$\min_{j \in \hat{I}} | \varphi_j(y) - \varphi_j(\hat{y})| > 0 \tag{3.9}$$

there is an $h \in T(Y, \hat{y})$ such that

$$\max_{j \in \hat{I}} (\varphi_j(\hat{y}) - \varphi_j(y))[\nabla\varphi_j(\hat{y}), h] < 0. \tag{3.10}$$

If (3.8) is violated under this assumption then for $\hat{I} = I^*(\hat{y})$ (3.9) must hold. This implies the existence of an $h \in T(Y, \hat{y})$ such that (3.10) is true which in turn implies

$$\max_{j \in I^*(\hat{y})} (\varphi_j(\hat{y}) - f_j)[\nabla\varphi_j(\hat{y}), h] < 0.$$

Hence, (3.6) is violated too which implies that (3.6) $\Rightarrow$ (3.8) by using contraposition.

THEOREM 3.2. *Under the above assumption the almost-convex-property holds.*

Proof. By using the identity (3.7) and the above reasoning showing that the condition (2.7) is a consequence of (3.8) we obtain that (2.8) implies

$$\max_{j \in \hat{I}} (\varphi_j(\hat{y}) - f_j)(\varphi_j(y) - \varphi_j(\hat{y})) < 0$$

where $\hat{I}$ is a closed subset of I. This implies (3.9) which in turn implies the existence of an $h \in T(Y, \hat{y})$ such that (3.10) holds. However, this implies

$$\max_{j \in \hat{I}} (\varphi_j(\hat{y}) - f_j)[\nabla\varphi(\hat{y}), h] < 0$$

which is equivalent to (2.9) according to the above reasoning. This completes the proof.

In [5] we have shown that for $Y = \hat{Y}$ the above assumption holds if and only if for all $f \in C(I)$ the condition (3.6) is sufficient and the condition (3.8) is necessary for $\hat{y} \in Y$ to solve the approximation problem. Brosowski and Wegmann have shown in [2] that the above assumption holds if and only if for all $f \in C(I)$ the condition (3.6) is sufficient for $\hat{y} \in Y$ to be a solution of the approximation problem. However, they use a slightly different definition for tangent cones. By the results of [4] the above assumption is, for instance, satisfied in the case of linear, generalized rational, and exponential approximation.

References

1. J. Abadie, On the Kuhn–Tucker theorem, *in* "Nonlinear Programming," pp. 21–36, North-Holland, Amsterdam, 1967.
2. B. Brosowski und R. Wegmann, Charakterisierung bester Approximationen in normierten Vektorräumen, *J. Approximation Theory* **3** (1970), 369–397.
3. M. R. Hestenes, "Calculus of Variations and Optimal Control Theory," John Wiley and Sons, New York/London/Sydney, 1966.
4. W. Krabs, Ueber differenzierbare asymptotisch konvexe Funktionenfamilien bei der nicht-linearen gleichmäßigen Approximation, *Arch. Rat. Mech. Anal.* **27** (1967), 275–288.
5. W. Krabs, Ueber die Reichweite des lokalen Kolmogoroff-Kriteriums bei der nicht-linearen gleichmäßigen Approximation, *J. Approximation Theory* **2** (1969), 258–264.
6. W. Krabs, Nichtlineare Optimierung mit unendlich vielen Nebenbedingungen, *Computing* **7** (1971), 204–214.