JOURNAL OF APPROXIMATION THEORY 36, 32–42 (1982)

Some Characterizations of Best Mixed-Norm Approximations

G. A. WATSON

Department of Mathematics, University of Dundee, Dundee, Scotland

Communicated by E. W. Cheney

Received April 23, 1980

1. INTRODUCTION

Let X, Y be compact topological spaces, and let S be a linear space of bivariate functions $f(x, y)$ from $X \times Y$ to R. For a given $M \subset S$, let M_X and M_Y be sets of univariate functions defined on X and Y, respectively, by

$$M_X = \{f(\cdot, y): f \in M, y \in Y\},$$
$$M_Y = \{f(x, \cdot): f \in M, x \in X\},$$

and let $\{S_X, \|\cdot\|_X\}$, $\{S_Y, \|\cdot\|_Y\}$ be normed linear spaces such that $M_X \subset S_X$, $M_Y \subset S_Y$. Then for each $y \in Y$, $f(\cdot, y) \in M_X$, we can write

$$\|f(\cdot, y)\|_X = \sup_{v \in U(S_X^*)} \langle f(\cdot, y), v(\cdot)\rangle_X, \tag{1.1}$$

where S_X^* is the topological dual space of S_X, $U(S_X^*)$ denotes the unit ball

$$\{v: v \in S_X^*, \|v\|_X^* \leqslant 1\},$$

and $\|\cdot\|_X^*$ is the dual norm. For all $f \in M$, $y \in Y$ let us define the sets

$$V_y(f) = \{v \in U(S_X^*), \|f\|_X = \langle f, v\rangle_X\},$$

and for each $y \in Y$, let $v_y \in V_y(f)$. Then we may consider v to be the bivariate function $v(x, y)$ defined by

$$v(x, y) = v_y(x), \qquad x \in X, y \in Y;$$

we denote by $V(f)$ the (convex) set of all bivariate functions which may be defined in this way, so that

$$V(f) = \{v: v(\cdot, y) \in V_y(f), \text{ for all } y \in Y\}.$$

0021-9045/82/090032–11\$02.00/0

We will restrict consideration in what follows to sets M and corresponding normed linear spaces for which

$$\langle f, v \rangle_X \in S_Y, \qquad \text{all} \quad f \in M,\ v \in V(f).$$

Under this assumption, it is possible to embed M in a linear space S which is equipped with the norm

$$\|f(x, y)\| = \| \|f(\cdot, \cdot)\|_X \|_Y \tag{1.2}$$

and therefore to define the approximation problem

$$\text{find } f \in M \text{ to minimise } \|f\|. \tag{1.3}$$

This is a mixed-norm bivariate approximation problem. Examples of such problems have been given in [2, 6]: in particular in [2] a characterization is given of a mixed-norm problem involving an L_1 and an L_∞ norm, which has applications in the study of integral transforms. The purpose of this paper is to consider in some generality the characterization of solutions to mixed-norm approximation problems in terms of properties of the individual norms. The analysis is carried out for bivariate approximation, but the results readily generalise to functions of an arbitrary number of variables in an obvious way.

The most convenient way of defining a particular subset M of S is through an appropriate parameterization. We will assume therefore that M is the family of functions

$$M = \{f(x, y, a), a \in R^n\},$$

where $f(x, y, a)$ is a given mapping from $X \times Y \times R^n$ into R. (The parameter space could in fact be any real Banach space; however little is lost, and some simplification is gained, by retaining finite dimensionality.) Then the problem (1.3) may be restated as

$$\text{find } a \in R^n \text{ to minimise } \|f(a)\| \tag{1.4}$$

where $f(a)\colon R^n \to M$. For all $a \in R^n$, we will write $V_y(a)$ for $V_y(f(a))$, $V(a)$ for $V(f(a))$, and will denote by $W(a)$ the set

$$W(a) = \{u \in U(S_Y^*),\ \|h\| = \langle h, u \rangle_Y\}, \tag{1.5}$$

where $h(y) = \|f(a)\|_X$, $y \in Y$, and S_Y^* is the dual space of S_Y.

In the next section, we establish necessary conditions in terms of V and W for solutions to some general subclasses of the problem (1.4). When M is convex, sufficiency results are also obtained, and as a consequence of this, precise characterizations of best approximations can be given. The results are presented in terms of the partial derivatives of f with respect to the components of a, which are assumed to exist and to be continuous everywhere; this global assumption may in fact be weakened, and replaced by a local requirement, but we will not draw any distinction. The method of attack gives a form for these results which seems the most useful from a practical point of view, as a check may be made on the conditions in a straightforward manner. As final pieces of general notation, we will use g_i to denote $\partial f/\partial a_i$, $i = 1, 2,..., n$, where a_i is the ith component of a, and $l(c)$ to denote $\sum_{i=1}^{n} c_i g_i$, for any $c \in R^n$. The explicit dependence of f, g_i (and other quantities) on a will often be suppressed, when no confusion can arise.

2. Conditions for a Best Approximation

In order to derive conditions in a convenient form, it is clearly necessary to exploit the special properties which are possessed by the functions involved in (1.4). In particular, we will make use (both implicit and explicit) of the fact that the assumption that $f(a)$ is continuously differentiable leads to $\|f(a)\|_X$ being locally Lipschitz for each $y \in Y$, and thus possessing a generalized gradient. For convenience, these properties are now defined.

Definition 1. A function $\phi(a)$: $R^n \to R$ is said to be *locally Lipschitz* if every point $a \in R^n$ admits a neighbourhood N such that, for some constant K,

$$|\phi(a_1) - \phi(a_2)| \leqslant K \|a_1 - a_2\|$$

for all a_1, $a_2 \in N$.

Let ϕ be locally Lipschitz, and let a be any point in R^n.

Definition 2. The *generalized directional derivative* of ϕ at a in the direction c is given by

$$\phi^0(a; c) = \limsup_{\substack{a_1 \to a \\ \gamma \to 0+}} [\phi(a_1 + \gamma c) - \phi(a_1)]/\gamma.$$

Definition 3. The *generalized gradient* of ϕ at a, denoted by $\partial\phi(a)$, is the set of all $z \in R^n$ satisfying

$$\phi^0(a; c) \geqslant \sum_{i=1}^{n} c_i z_i,$$

for all $c \in R^n$.

If $\phi(a)$ is convex, or continuously differentiable at a, then $\partial\phi(a)$ is, respectively, the subdifferential at a in the sense of convex analysis, or the vector of partial derivatives of $\phi(a)$ with respect to the components of a (see Clarke [3, 4]).

LEMMA 1. *Let $y \in Y$ be arbitrary. Then $\|f(a)\|_X$ is a locally Lipschitz function, with the generalized gradient at a given by*

$$\partial \|f(a)\|_X = \{z \in R^n \colon z_i = \langle g_i, v\rangle_X,\ i = 1, 2,\ldots, n,\ v \in V_y\}. \tag{2.1}$$

Proof. Let $c \in R^n$, $\|c\| = 1$, be arbitrary, and let $\gamma > 0$. Then if $v \in V_y$,

$$\begin{aligned}\|f(a+\gamma c)\|_X - \|f(a)\|_X &\geqslant \langle f(a+\gamma c), v\rangle_X - \langle f(a), v\rangle_X \\ &= \gamma\langle l(c), v\rangle_X + o(\gamma).\end{aligned}$$

Also, if $v(\gamma) \in V_y(a + \gamma c)$,

$$\begin{aligned}\|f(a+\gamma c)\|_X &= \langle f(a+\gamma c), v(\gamma)\rangle_X \\ &\leqslant \|f(a)\|_X + \gamma\langle l(c), v(\gamma)\rangle_X + o(\gamma).\end{aligned}$$

Thus

$$\begin{aligned}\langle l(c), v\rangle_X + o(1) &\leqslant \frac{\|f(a+\gamma c)\|_X - \|f(a)\|_X}{\gamma} \\ &\leqslant \langle l(c), v(\gamma)\rangle_X + o(1).\end{aligned} \tag{2.2}$$

By the weak* compactness of the unit ball in S_X^* (Alaoglu–Bourbaki theorem, for example Holmes [5]) there exists a sequence $\{\gamma_j\} \to 0$ and $\bar{v} \in S_X^*$ such that

$$\langle m, v(\gamma_j)\rangle_X \to \langle m, \bar{v}\rangle_X \qquad \text{as} \quad j \to \infty,$$

for all $m \in S_X$. Also

$$\begin{aligned}0 &\leqslant \|f(a)\|_X - \langle f(a), v(\gamma)\rangle_X \\ &\leqslant \|f(a+\gamma c)\|_X - \langle f(a+\gamma c), v(\gamma)\rangle_X + O(\gamma) \\ &\leqslant O(\gamma)\end{aligned}$$

and so $\bar{v} \in V_y$. Letting $\gamma \to 0$ in the inequalities to (2.2) along the sequence $\{\gamma_j\}$, we have the limiting value of the right-hand side as

$$\langle l(c), \bar{v}\rangle_X = \max_{v \in V_y} \langle l(c), v\rangle_X$$

It follows that we must have

$$\lim_{\gamma \to 0+} \frac{\|f(a+\gamma c)\|_X - \|f(a)\|_X}{\gamma} = \max_{v \in V_y} \langle l(c), v \rangle_X. \tag{2.3}$$

Thus, for any $y \in Y$, $\|f(a)\|_X$ is a locally Lipschitz function, with generalized gradient given by the set of $z \in R^n$ satisfying

$$\max_{v \in V_y} \sum_{i=1}^{n} c_i \langle g_i, v \rangle_X \geqslant \sum_{i=1}^{n} c_i z_i$$

for all $c \in R^n$. The result (2.1) follows.

Although the above lemma is proved for $\|f\|_X$, it is evident that an equivalent result holds for any norm on the elements of M. A necessary condition for a to solve (1.4) (the existence of a zero generalized gradient) is therefore readily available in terms of the elements of S^*, the dual space of S. However, to obtain such a result in terms of properties of the individually occurring normed spaces forces some restrictions to be placed on (1.4), and it is convenient to consider separately some special cases. Perhaps the most straightforward of these arises when the norm on X may be assumed to be smooth at the points of interest.

THEOREM 1. *Let a solve* (1.4), and let $\|\cdot\|_X$ *be smooth at $f(a)$ for all $y \in Y$. Then there exists $w \in W$ such that*

$$\langle \langle g_i, v \rangle_X, w \rangle_Y = 0, \qquad i = 1, 2, \ldots, n, \tag{2.4}$$

for all $v \in V$.

Proof. By the smoothness assumption, $\|f(a)\|_X$ is differentiable at a for all $y \in Y$, and the generalized gradient (2.1) is just the unique vector in R^n with the ith component

$$\langle g_i, v_y \rangle_X, \qquad i = 1, 2, \ldots, n$$

for any $v_y \in V_y$, all $y \in Y$. If $\|f(a)\|$ is a minimum, then $\| \|f(a)\|_X \|_Y$ is a minimum over all $\|f(a)\|_X$. The result (2.4) is therefore an immediate consequence of applying the appropriate minimum norm necessary condition (or equivalently applying Lemma 1) to $\|\cdot\|_Y$ (see, for example, [7]).

THEOREM 2. *Let M be convex, let* (2.4) *hold at a, and let $\|\cdot\|_X$ be smooth at a for all $y \in Y$. Then a solves* (1.4).

Proof. Let the conditions be satisfied, and let $c \in R^n$ be arbitrary. Then for all μ, $0 < \mu \leqslant 1$, the convexity assumption gives

$$\|f(a+c)\| - \|f(a)\| \geqslant \frac{1}{\mu} (\|f(a+\mu c)\| - \|f(a)\|)$$
$$\geqslant \frac{1}{\mu} \langle \|f(a+\mu c)\|_X - \|f(a)\|_X, w\rangle_Y \qquad (2.5)$$

for all $w \in W$. Thus by Taylor expansion

$$\|f(a+c)\| - \|f(a)\| \geqslant \langle\langle l(c), v\rangle_X, w\rangle_Y + o(1) \qquad (2.6)$$

for all $v \in V$. Since this inequality holds on letting $\mu \to 0$, the result follows.

When $\|\cdot\|_X$ cannot be assumed smooth, it is necessary to restrict S_Y or $\|\cdot\|_Y$ in some way. An important requirement, which has the effect of allowing the generalized gradient of $\|f(a)\|$ to be given in an appropriate form, is the condition that $\|\cdot\|_Y$ be *monotonic* in the sense that if $g_1(y)$, $g_2(y) \in S_Y$ with $|g_1(y)| \geqslant |g_2(y)|$ for all $y \in Y$, then $\|g_1(y)\|_Y \geqslant \|g_2(y)\|_Y$. The important consequence of monotonicity for our purposes is the following. Let $s \in S$, let v be such that $v(x, y) \in U(S_X^*)$ for each $y \in Y$, and let $w \in U(S_Y^*)$. Then

$$|\langle\langle s, v\rangle_X, w\rangle_Y| \leqslant \|\langle s, v\rangle_X\|_Y$$
$$\leqslant \|s\|.$$

The next theorem is proved using approximation theoretic techniques, but the result could also have been obtained by construction of the generalized gradient of $\|f(a)\|$.

THEOREM 3. *Let $\|\cdot\|_Y$ be monotonic, and let a solve* (1.4). *Then*

$$0 \in \operatorname{conv}\{z \in R^n: z_i = \langle\langle g_i, v\rangle_X, w\rangle_Y,\ i = 1, 2,\dots, n,\ v \in V, w \in W\}. \qquad (2.7)$$

Proof. Let a solve (1.4) but (2.7) not be satisfied. Then by the theorem on linear inequalities [1], there exists $c \in R^n$, $\delta > 0$ such that

$$\langle\langle l(c), v\rangle_X, w\rangle_Y \leqslant -\delta \qquad \text{for all} \quad v \in V,\ w \in W.$$

Let $v \in V$, $w \in W$, and for any $\gamma > 0$, $v(\gamma) \in V(a+\gamma c)$, $w(\gamma) \in W(a+\gamma c)$ all be arbitrary. Then

$$\|f(a+\gamma c)\| = \langle\langle f(a+\gamma c), v(\gamma)\rangle_X, w(\gamma)\rangle_Y$$
$$= \langle\langle f(a), v(\gamma)\rangle_X, w(\gamma)\rangle_Y$$
$$+ \gamma\langle\langle l(c), v(\gamma)\rangle_X, w(\gamma)\rangle_Y + o(\gamma)$$

$$\begin{aligned}&\leqslant \|f(a)\| + \gamma\langle\langle l(c), v(\gamma)\rangle_X, w(\gamma)\rangle_Y + o(\gamma)\\ &\leqslant \|f(a)\| - \gamma\delta + \gamma\langle\langle l(c), v(\gamma)\rangle_X, w(\gamma)\rangle_Y\\ &\quad - \gamma\langle\langle l(c), v\rangle_X, w\rangle_Y + o(\gamma).\end{aligned} \tag{2.8}$$

Now if $s \in U(S)$,

$$|\langle\langle s, v(\gamma)\rangle_X, w(\gamma)\rangle_Y| \leqslant 1$$

by the monotonicity assumption, so that $\langle\langle s, v(\gamma)\rangle_X, w(\gamma)\rangle_Y \in U(S^*)$, where S^* is the dual space of S. By the weak* compactness of $U(S^*)$, there exists a positive sequence $\{\gamma_j\} \to 0$ such that $\{\langle\langle s, v(\gamma_j)\rangle_X, w(\gamma_j)\rangle_Y\}$ is convergent, for all $s \in S$. In addition

$$\begin{aligned}0 &\leqslant \|f(a)\| - \langle\langle f(a), v(\gamma)\rangle_X, w(\gamma)\rangle_Y\\ &= \langle\langle f, v\rangle_X, w\rangle_Y - \langle\langle f(a+\gamma c), v(\gamma)\rangle_X, w(\gamma)\rangle_Y + O(\gamma)\\ &= \langle\langle f(a+\gamma c), v\rangle_X, w\rangle_Y - \|f(a+\gamma c)\| + O(\gamma)\\ &\leqslant O(\gamma).\end{aligned}$$

Thus the limiting functional in the above net is one for which $\|f(a)\|$ is attained, and so has the form $\langle\langle s, \bar{v}\rangle_X, \bar{w}\rangle_Y$ for some $\bar{v} \in V$, $\bar{w} \in W$. Put $v = \bar{v}$, $w = \bar{w}$ in (2.8) and let $\gamma \to 0$ along the sequence $\{\gamma_j\}$. For γ sufficiently small, we contradict the fact that a solves (1.4) and the theorem is proved.

THEOREM 4. *Let M be convex, let $\|\cdot\|_Y$ be monotonic, and let* (2.7) *hold at a. Then a solves* (1.4).

Proof. This is virtually identical with the proof of Theorem 2, with (2.6) following in this case by the monotonicity assumption.

It is possible to give appropriate necessary conditions in the absence of both smoothness and monotonicity, but subject to the condition that the space S_Y be *finite dimensional.* This is the substance of the following theorem, which uses results from the theory of locally Lipschitz programming.

THEOREM 5. *Let $S_Y = R^m$, and let a solve* (1.4). *Then there exists $v \in V$, $w \in W$ such that*

$$\langle\langle g_i, v\rangle_X, w\rangle_Y = 0, \qquad i = 1, 2, \ldots, n. \tag{2.9}$$

Proof. We can take $Y = \{y_1, y_2, \ldots, y_m\}$. Then the problem (1.4) may be posed as follows.

$$\begin{aligned}&\text{find } a \in R^n,\ h \in R^m \text{ to minimise } \|h\|_Y\\ &\text{subject to } h_j = \|f(a)\|_{X,j},\ j = 1, 2, \ldots, m,\end{aligned} \tag{2.10}$$

where h_j is the jth component of h and $\|f(a)\|_{X,j}$ denotes $\|f(a)\|_X$ evaluated at $y = y_j$. This is a nondifferentiable optimization problem in the $n+m$ variables a, h with a convex objective function, and m equality constraints, the constraint functions being locally Lipschitz, by Lemma 1. If all the functions involved are considered as functions of $\binom{a}{h} \in R^{n+m}$, we have the generalized gradients

$$\partial \|h\|_Y = \{z \in R^{n+m}: z_i = 0,\ i = 1, 2,..., n,\ z_{i+n} = w_i,\ i = 1, 2,..., m,\ w \in W\},$$

$$\partial(\|f(a)\|_{X,j} - h_j) = \{z \in R^{n+m}: z_i = \langle g_i^j, v^j \rangle_X, i = 1, 2,..., n,$$
$$v^j \in V^j, z_{i+n} = -\delta_{ij}, i = 1, 2,..., m\},$$

$j = 1, 2,..., m$, where g_i^j denotes g_i evaluated at y_j, and $V^j \equiv V_{y_j}$. Then by Theorem 1 of Clarke [4], if $\binom{a}{h}$ solves (2.10), there exist numbers $\lambda_0 \geqslant 0$, λ_j, $j = 1, 2,..., m$ (Lagrange multipliers) not all zero such that

$$0 \in \lambda_0 \partial \|h\|_Y + \sum_{j=1}^{m} \lambda_j \partial(\|f(a)\|_{X,j} - h_j),$$

so that

$$\sum_{j=1}^{m} \lambda_j \langle g_i^j, v^j \rangle_X = 0, \qquad i = 1, 2,..., n, \tag{2.11}$$

$$\lambda_0 w_i = \lambda_i, \qquad i = 1, 2,..., m \tag{2.12}$$

for some $v^j \in V^j$, $j = 1, 2,..., m$, and some $w \in W$. It follows from (2.12) that $\lambda_0 \neq 0$, and so substituting for λ into (2.11) gives the required result.

THEOREM 6. *Let $S_Y = R^m$, let M be convex, and let* (2.9) *hold with $w_j \geqslant 0$, $j = 1, 2,..., m$. Then a solves* (1.4).

Proof. Exactly as in the proof of Theorem 2, we obtain the inequality (2.5). Now if v, $w \geqslant 0$ satisfy (2.9),

$$\langle \|f(a+\mu c)\|_X, w \rangle_Y \geqslant \langle \langle f(a+\mu c), v \rangle_X, w \rangle_Y$$

so that (2.6) also holds and the result follows.

In fact, monotonicity of $\|\cdot\|_Y$ is the natural condition which allows (2.9) to be both necessary and sufficient (without further qualification) for a to solve (1.4) when M is convex and $S_Y = R^m$. Under this assumption, a solves (2.10) if and only if a solves (2.10) with the equality constraints replaced by the *inequality* constraints

$$\|f(a)\|_{X,j} \leqslant h_j, \qquad j = 1, 2,..., m,$$

for we can reduce any of the h_j values to force equality to hold without raising the value of $\|h\|_Y$. It follows (for example, [4]) that the Lagrange multipliers λ_j, $j=1, 2,..., m$ must be non-negative, and so, by (2.12) we must have $w_j \geqslant 0$, $j=1, 2,..., m$.

3. Examples

In this final section, we illustrate the application of the theorems by taking some specific examples of pairs of normed linear spaces. First, let M_X, $\|\cdot\|_X \subset L_p(X)$, M_Y, $\|\cdot\|_Y \subset L_q(Y)$, $1 < p, q < \infty$, with X and Y (say) intervals of the real line. Then (1.4) is

$$\text{find } a \in R^n \text{ to minimize } \left(\int_Y \left(\int_X |f(x, y, a)|^p \, dx\right)^{q/p} dy\right)^{1/q}.$$

If $\|f\| > 0$, then the set W contains the unique element

$$w(y) = \|f\|_X^{q-1} \, \|f\|^{1-q}.$$

Also if $\|f\|_X > 0$ for all $y \in Y$, then V contains the unique element

$$v(x, y) = \operatorname{sign}(f) \, |f|^{p-1} \, \|f\|_X^{1-p}.$$

Theorem 1 applies, and the condition (2.4) is easily seen to be

$$\int_Y \|f\|_X^{q-p} \int_X g_i \operatorname{sign}(f) \, |f|^{p-1} \, dx \, dy = 0, \qquad i = 1, 2,..., n.$$

Of course if $p = q$, we recover the result which would be obtained directly by treating the problem as one in $L_p(X \times Y)$.

Now let $S_X = R^l$, $X = \{x_1, x_2,..., x_l\}$, normed by the L_1 norm, and let $S_Y = R^m$, $Y = \{y_1, y_2,..., y_m\}$ normed by the elliptic norm $\|h\|_Y^2 = \langle h, Gh\rangle_Y$, where G is an $m \times m$ symmetric positive definite matrix (this norm need not be monotonic). Let f_{jk} denote $f(x_j, y_k)$. Then if $\|f\| \neq 0$, W is just the singleton

$$w = \frac{Gh}{\|f\|},$$

where $h_k = \sum_{j=1}^{l} |f_{jk}|$, $k = 1, 2,..., m$. Also

$$V = \{v \in R^l \times R^m \colon v_{jk} = \operatorname{sign}((f_{jk}), f_{jk} \neq 0, |v_{jk}| \leqslant 1\}.$$

Then the condition of Theorem 5 is: there exists $v \in V$ such that

$$\sum_{k=1}^{m} \sum_{j=1}^{l} v_{jk}\, g_i(x_j, y_k)\, w_k = 0, \qquad i = 1, 2,..., n.$$

Finally, let M_X, $\|\cdot\|_X \subset L_1(X, \Sigma, \mu)$, where (X, Σ, μ) is a finite measure space, and let M_Y, $\|\cdot\|_Y \subset C(Y)$, where Y is a compact Hausdorff space. Then (1.4) is

$$\text{find } a \in R^n \text{ to minimise } \max_{y \in Y} \int_X |f(x, y, a)|\, d\mu(x). \tag{3.1}$$

At $a \in R^n$, define the sets

$$Z_y = \{x \in X : f = 0\} \qquad \text{for all} \quad y \in Y,$$
$$K = \{y \in Y : \|f\|_X = \|f\|\}.$$

THEOREM 7. *Let a solve* (3.1) *with* $\mu(Z_y) = 0$, *all* $y \in Y$. *Then there exist* $t \leqslant n + 1$ *points* $\{y_1, y_2,..., y_t\} \in K$ *and a nontrivial vector* $\lambda \in R^t$, $\lambda_j \geqslant 0$, $j = 1, 2,..., t$, *such that*

$$\sum_{j=1}^{t} \lambda_j \int_X g_i(x, y_j)\, \operatorname{sign}(f(x, y_j))\, d\mu(x) = 0, \qquad i = 1, 2,..., n.$$

Proof. We have for all $y \in Y$

$$V_y = \{v \in L_\infty(X, \Sigma, \mu) : |v| \leqslant 1, v = \operatorname{sign}(f), x \in X - Z_y\}$$

so that

$$V = \{v(x, y) : v(\cdot, y) \in L_\infty(X, \Sigma, \mu), v(\cdot, y) = \operatorname{sign}(f), x \in X - Z_y,$$
$$|v| \leqslant 1, x \in X, y \in Y\}.$$

Also

$$W = \operatorname{conv}\{\delta(y), y \in K\}$$

where $\langle g(y), \delta(y_0)\rangle_Y = g(y_0)$, $y_0 \in Y$. If $\mu(Z_y) = 0$, all $y \in Y$, then for all $v \in V$,

$$\langle g_i, v\rangle_X = \int_X g_i \operatorname{sign}(f)\, d\mu(x), \qquad i = 1, 2,..., n.$$

Thus we have Theorem 1 holding, and the result follows from (2.4), the definition of W, and Carathéodory's theorem.

REFERENCES

1. E. W. CHENEY, "Introduction to Approximation Theory," McGraw-Hill, New York, 1966.
2. E. W. CHENEY, J. H. MCCABE AND G. M. PHILLIPS, A mixed-norm bivariate approximation problem with applications to Lewanowicz operators, *in* "Multivariate Approximation Theory" (D. C. Handscomb, Ed.), Academic Press, London, 1978.
3. F. H. CLARKE, Generalized gradients and applications, *Trans. Amer. Math. Soc.* **205** (1975), 247–262.
4. F. H. CLARKE, A new approach to Lagrange multipliers, *Math. Oper. Res.* **1** (1976), 165–174.
5. R. B. HOLMES, "Geometric Functional Analysis and its Applications," Springer-Verlag, New York, 1975.
6. W. KRABS, "Optimierung und Approximation," Teubner, Stuttgart, 1975.
7. G. A. WATSON, On a class of methods for nonlinear approximation problems, *in* "Multivariate Approximation Theory" (D. C. Handscomb, Ed.), Academic Press, London, 1978.