APPROXIMATION IN THE METRIC OF $L^1(X, \mu)$

BY

B. R. KRIPKE AND T. J. RIVLIN

0. Introduction.

0.1. Let W be a semi-normed linear space with semi-norm $\|\cdot\|$, V a closed linear subspace, and $x \in W$. If $\inf[\|x-y\| : y \in V]$ is attained, for $y = y^* \in V$, say, then we call y^* a best approximation to x out of V. A central problem of approximation theory is the study of the characterization and uniqueness of such best approximations.

This problem can be given an air of greater generality by allowing V to be an arbitrary closed affine subspace of W. However, by translating V and x by some vector in V, we can arrange that V pass through 0, so that we are reduced to the case that V is actually a linear subspace (cf. Singer [16]). In this way we can encompass the case that y varies in V subject to constraints of the form $f(y) = c$, where f is a continuous linear functional on V.

The uniqueness question will receive much attention in what follows. When V is finite dimensional, a case of most frequent interest, the existence of best approximations is proved by a routine compactness argument (Achieser [1, p. 10]). If $\|\cdot\|$ is a norm, and is strictly convex (if $\|x\| \leqq \rho$, $\|y\| \leqq \rho$ and $x \neq y$, then $\|x+y\| < 2\rho$), then best approximations, when they exist, are unique (Achieser [1]). This settles the uniqueness question for the L^p spaces with $1 < p < \infty$, since their norms are strictly convex. However, in L^1 and L^∞ the norms are not strictly convex, and the uniqueness problem requires further study.

A natural approach to the study of best approximations is to consider the derivative $d/dt\,\|x-y-tz\|\,|_{t=0}$. If y is a best approximation out of V to x, we should expect this derivative to be 0 for all $z \in V$. This approach is successful in studying approximation in L^2 and can be used to advantage in the L^p spaces with $1 < p < \infty$. However, in L^1 and L^∞ the derivative may fail to exist for the same reason that the norm is not strictly convex: spheres may have facets and corners. Therefore, since our concern is with approximation in L^1, §1 will be devoted to a careful study of the differentiability of the norm in L^1, and the proof of a basic variational lemma.

0.2. The setting for our paper is as follows: we are given a positive measure space, (X, μ), and we consider the linear space, W, of μ-summable functions on X under the semi-norm

Presented to the Society, September 1, 1960 under the title *A unified approach to approximation in the metric of L*. I, II; received by the editors September 28, 1961 and, in revised form May 8, 1962 and January 15, 1964.

$$\|f\| = \int_X |f(x)| \; d\mu(x). \tag{0.1}$$

Frequently, we shall confine our attention to certain subspaces of W on which $\|\cdot\|$ is actually a norm. For example, we consider the continuous and the meromorphic functions in the complex plane. Otherwise, it will be convenient to pass to the quotient space $L^1(X,\mu)$ of W by the functions which vanish almost everywhere (a.e.), so that $\|\cdot\|$ again becomes a norm. Here, however, we shall make the usual harmless abuse of language and speak of the equivalence classes as being functions and having values (determined almost everywhere), etc.

Thus far, nothing has been said about whether the functions in W are to be real- or complex-valued. When a distinction is to be made, we indicate it by a subscript: $L^1_C(X,\mu)$, $L^1_R(X,\mu)$. When the subscript is missing, our remarks apply to both cases. $L^1_C(X,\mu)$ can, of course, be viewed as a real vector space simply by restricting multiplication to real scalars.

We define sgn x to be $x/|x|$ if $x \neq 0$, and sgn $0 = 0$. $\overline{\text{sgn}}\, x = \bar{x}/|x|$, is the complex conjugate of sgn x.

In §2, we shall apply our results on differentiating the norm and our variational lemma to proving uniqueness and nonuniqueness in various special situations. Our principal result is the extension of a classical theorem of Jackson to the complex case, with various examples to show that our theorem cannot be much improved.

§3 deals with the case when μ is concentrated in finitely many points, and concludes with a brief survey, in the context of our approach, of mostly known results for the case of continuous functions on an interval.

0.3. Some of these results were presented before the meeting of the American Mathematical Society in Worcester in October 1960. One of the authors (Rivlin) wishes to acknowledge many enlightening conversations with Professor H. S. Shapiro of New York University on the subject of L^1 approximation. In particular, it was Professor Shapiro who first brought to his attention the variational lemma from which we have derived most of our results. Part of the work of B. R. Kripke was done during the tenure of a National Science Foundation Graduate Fellowship, and part was supported by the United States Air Force under contract AF-AFOSR-467-63. The original work on this paper was done during the summer of 1960 at the International Business Machines Corp. Mathematical Research Center.

1. **Differentiability of the norm.**

1.1. We begin with some definitions. If f is μ-summable and g is bounded a.e. and μ-measurable, we abbreviate the inner product $\int_X f(x)\overline{g(x)}d\mu$ by "(f,g)." $Z(f) = \{x \in X : f(x) = 0\}$ is the set of zeros of f. For $f \in L^1$, $Z(f)$ may be found by choosing a representative of the equivalence class f. $Z(f)$ is then an equivalence class of measurable sets, with two sets, A and B, being identified if their symmetric difference, $(A \setminus B) \cup (B \setminus A)$ has measure zero. For a set, V, of functions, put

$$Z(V) = \bigcap \{Z(f) : f \in V\}.$$

$Z(V)$ is the set of common zeros of functions in V. The remainder, $X \setminus Z(f)$ or $X \setminus Z(V)$, will be denoted by "$R(f)$" or "$R(V)$" respectively.

1.2. THEOREM 1.1. *For $f, p \in L^1(X,\mu)$ and real t,*

$$\lim_{t \to 0} 1/t\Big[\|f+tp\| - \|f\| - |t| \int_{Z(f)} |p| \, d\mu\Big] = \mathrm{Re}(p, \mathrm{sgn} f). \tag{1.1}$$

Proof.

$$1/t\Big[\|f+tp\| - \|f\| - |t| \int_{Z(f)} |p| \, d\mu\Big]$$
$$= 1/t \int_{R(f)} [|f+tp| - |f|] \, d\mu = \int_{R(f)} \frac{2\mathrm{Re}(p\bar{f}) + t|p|^2}{|f+tp| + |f|} d\mu.$$

Now as $t \to 0$, the integrand approaches $2\mathrm{Re}(p\bar{f})/2|f| = \mathrm{Re}(p\,\overline{\mathrm{sgn} f})$ on $R(f)$; while $|(|f+tp| - |f|)/t| \leqq |p|$, and p is summable. The theorem is now proved by applying the Lebesgue dominated convergence theorem, taking into account that $\mathrm{sgn} f$ vanishes on $Z(f)$.

COROLLARY 1.2. *If $\mu(Z(f) \cap R(V)) = 0$ then $\|f+tp\|$ is differentiable at 0 as a function of t for each $p \in V$, and*

$$\frac{d}{dt}\|f+tp\| \Big|_{t=0} = \mathrm{Re}(p, \mathrm{sgn}\ f). \tag{1.2}$$

(1.1) shows that the one-sided derivatives of the norm exist in any case. This, in fact, is true in any normed linear space (see Dunford and Schwartz [4, pp. 445–453, pp. 471–473] or James [7]).

The condition $\mu(Z(f) \cap R(V)) = 0$ of the corollary is actually necessary for the derivative $d/dt\|f+tp\| \,|_{t=0}$ to exist when V is a finite dimensional subspace of L^1. Say $p_1, \cdots, p_n$ is a basis for V. Then $Z(V) = Z(p_1) \cap \cdots \cap Z(p_n)$, so that $R(V) = R(p_1) \cup \cdots \cup R(p_n)$. If $\mu(Z(f) \cap R(V)) > 0$, then $\mu(Z(f) \cap R(p_k)) > 0$ for some k, and it is then clear from (1.1) that the derivative along p_k does not exist at 0.

1.3. We now present the variational lemma from which most of our results will be deduced.

THEOREM 1.3. *A necessary and sufficient condition that*

$$\|f - p_0\| \leqq \|f - (p_0 + cp)\|$$

for all scalars c (real c in the real case, complex c in the complex case) is that

$$|(p, \operatorname{sgn}(f - p_0))| \leqq \int_{Z(f-p_0)} |p| \, d\mu. \tag{1.3}$$

If

$$|(p, \operatorname{sgn}(f - p_0))| < \int_{Z(f-p_0)} |p| \, d\mu, \tag{1.4}$$

then when $c \neq 0$ *we have* $\|f - p_0\| < \|f - (p_0 + cp)\|$.

Proof. If $0 < t < 1$, the triangle inequality yields $\|f - (p_0 + tcp)\| \leqq (1-t) \|f - p_0\| + t\|f - (p_0 + cp)\|$, from which we deduce $\|f - (p_0 + tcp\| - \|f - p_0\| \leqq t[\|f - (p_0 + cp)\| - \|f - p_0\|]$. Thus (1.1) implies

$$\begin{aligned} &\int_{Z(f-p_0)} |cp| \, d\mu - \operatorname{Re}(cp, \operatorname{sgn}(f - p_0)) \\ &= \lim_{t \to 0^+} 1/t[\|f - p_0 + t(-cp)\| - \|f - p_0\|] \leqq \|f - (p_0 + cp)\| - \|f - p_0\|. \end{aligned} \tag{1.5}$$

Now if (1.3) holds, it remains valid with p replaced by cp. From this we conclude immediately that the left-hand side of (1.5) is non-negative, proving the sufficiency of (1.3). If $c \neq 0$ and (1.4) holds, this left-hand side is positive, proving the final statement of the theorem.

Suppose (1.3) fails. Then putting $c = \overline{\operatorname{sgn}}(p, \operatorname{sgn}(f - p_0))$, we have $\operatorname{Re}(cp, \operatorname{sgn}(f - p_0)) = |(cp, \operatorname{sgn}(f - p_0))| > \int_{Z(f-p_0)} |cp| \, d\mu$. Thus (1.5) gives $\lim_{t \to 0^+} 1/t[\|f - (p_0 + tcp)\| - \|f - p_0\|] < 0$. When t is positive and small enough, it follows that $\|f - (p_0 + tcp)\| < \|f - p_0\|$, proving the necessity of (1.3).

The actual form in which this variational lemma is most useful is the following:

Corollary 1.4. *Let V be a linear subspace of* L^1, $p_0 \in V$. *Then*

$$\|f - p_0\| \leqq \|f - p\|$$

for all $p \in V$ *if, and only if,* (1.3) *holds for all* $p \in V$. *If* (1.4) *holds for all nonzero* $p \in V$, *then* p_0 *is the unique best approximation to f out of V.*

Another corollary to Theorem 1.3 which is frequently used in the literature on L^1 approximation is the following orthogonality relation.

Corollary 1.5. *If* p_0 *is a best approximation to f out of V and* $\mu(Z(f - p_0)) = 0$, *then for all* $p \in V$

$$(p, \operatorname{sgn}(f - p_0)) = 0. \tag{1.6}$$

Conversely, if (1.6) *holds for some* $p_0 \in V$ *and all* $p \in V$, *then* p_0 *is a best approximation to* f *out of* V.

Theorem 1.3 has not, to the best of our knowledge, previously appeared in print in its full generality. James [7] obtains the result for the real case in a somewhat different context. The orthogonality relation, Corollary 1.5, seems to be widely known, in the real case, at least. Achieser [1, pp. 82–85] and Laasonen [11] prove it in the real case, under strong restrictions, by direct differentiation. Havinson [5] proves a similar result. He shows that when $\mu(Z(f-p_0)) > 0$, (1.6) continues to hold if $\operatorname{sgn}(f-p_0)$ is altered appropriately on $Z(f-p_0)$. Singer [15] gives a proof of the corollary in the real case, and attributes the result to M.G. Kreĭn. However, in quoting Kreĭn (as Dr. Singer has informed us) the author omitted a hypothesis equivalent to assuming that $\mu(Z(f-p_0)) = 0$. Therefore, his proof and conclusion are convincing only subject to this hypothesis. (There is a typographical error in Singer's reference to Kreĭn. Instead of "[1]," there should appear "[2]," which is Kreĭn and Achieser [10].) Indeed, we shall see several examples below in which (1.6) fails when $\mu(Z(f-p_0)) > 0$. This error, the omission of the hypothesis $\mu(Z(f-p_0)) = 0$, occurs also in an exercise in Dunford and Schwartz [4, p. 371, Ex. 84].

2. **Uniqueness.** This section is devoted to a discussion of conditions on X, μ, f, and V which insure that f have a *unique* best approximation out of V.

2.1. One phenomenon which has marked influence on uniqueness is the presence of atoms in X. An atom is a measurable subset, A, of X (for example, a point having positive mass) satisfying

(2.1) $\mu(A) > 0$,

(2.2) if E is a measurable subset of A, either $\mu(E) = 0$ or $\mu(A \setminus E) = 0$.

It is convenient when $F \subseteq E$ and $\mu(E \setminus F) = 0$ to say that E consists essentially of F.

When X has no atoms, the situation is as follows:

THEOREM 2.1. *If* (X,μ) *is atom-free, and* V *is a finite-dimensional subspace of* L^1, *then there exists an* $f \in L^1$ *which has infinitely many best approximations out of* V.

The real case of this theorem was proved by Kreĭn [10] for Lebesgue measure on an interval of the real line; and Phelps [13] proved the real case of the general theorem. Therefore, in place of a proof, we shall confine ourselves to the remark that the proof for the complex case follows the lines of Phelps' argument, with some extra work being required because of the larger number of values which $\operatorname{sgn} x$ can assume when x is a complex variable.

Some additional structure must be introduced if we want best approximations to be unique. As in the case of uniform approximation, it is fruitful to superimpose on X a topology compatible with μ. When X has a topology, we shall henceforth require that μ have the following properties:

(2.3) Each point in X and each open subset of X is μ-measurable.

(2.4) There is a closed set $C \subseteq X$ such that

(a) $\mu(X \setminus C) = 0$,

(b) if U is open and $U \cap C$ is not empty, then $\mu(U) > 0$.

It is easy to see that there can be at most one set, C, satisfying (2.4). When there is such a set, we call it the *carrier* of μ and denote it by "$C(\mu)$." $C(\mu)$ will always exist if the topology for X has a countable base; or if it is paracompact (Bourbaki [3, pp. 67–78]). In either case, $C(\mu)$ is the complement of the largest open set of measure zero.

In this context, it is natural to require that f and the functions in V be continuous. A fundamental rôle in determining the uniqueness of approximations out of V in such a situation will be played by the connectivity of the set $C(\mu) \setminus Z(V)$. As an illustration, we have the following *nonuniqueness* theorem:

THEOREM 2.2. *Let X be a normal topological space, and let $p_1, \cdots, p_n$ be continuous functions which are a basis for the subspace V of L^1.*

(a) *$C(\mu) \setminus Z(V)$ must contain at least n points.*

(b) *Suppose there are disjoint closed sets, E and F, such that*

$$\mu(X \setminus (Z(V) \cup E \cup F)) = 0$$

and

(2.5) *$\int_E |p|\, d\mu$ is a norm on V,*

(2.6) *there is at least one $p \in V$ such that $\int_F |p|\, d\mu > 0$. Then there are continuous functions $f, q_1, \cdots, q_n$ such that*

(2.7) *for any scalars $s_1, \cdots, s_n, s_1 q_1 + \cdots + s_n q_n$ and $s_1 p_1 + \cdots + s_n p_n$ have the same zeros in $E \cup F$;*

(2.8) *f has infinitely many best approximations out of the subspace, V^0, of L^1 spanned by $q_1, \cdots, q_n$.*

Proof. (a) is evident, because $p_1, \cdots, p_n$ are linearly independent as elements of L^1.

(b) We use the following facts: The closed unit sphere in a normed linear space of finite dimension is compact. Any two norms on a space of finite dimension are equivalent. (Dunford and Schwartz [4, pp. 244–245].)

Since the semi-norm $\int_F |p|\, d\mu$ is dominated by the norm $\int_X |p|\, d\mu$ on V, it is continuous, and so attains its maximum on the set $\{p \in V : \int_E |p|\, d\mu \leqq 1\}$ at a point p_0. According to (2.6), $0 < M = \int_F |p_0|\, d\mu < \infty$.

Since X is normal, there are continuous functions, g and h, on X such that $g \equiv 0$ and $h \equiv 1$ on E, while $g \equiv 1$ and $h \equiv 0$ on F. Put $q_i = (h + g/M)\, p_i$, $i = 0, \cdots, n$, and $f = (g/M) p_0$. The maximum of $\int_F |q|\, d\mu$ on the set $\{q \in V^0 : \int_E |q|\, d\mu \leqq 1\}$ is 1, and it is attained when $q = q_0$.

For any $q \in V^0$,

$$\int_{Z(f)} |q|\, d\mu \geqq \int_E |q|\, d\mu \geqq \int_F |q|\, d\mu$$

$$\geqq \int_{R(f)} |q|\, d\mu \geqq \left| \int_{X - Z(f)} q\, \overline{\mathrm{sgn}}\, f\, d\mu \right|,$$

because the set of points outside F where f does not vanish is contained in $X \setminus (Z(V) \cup E \cup F)$, whose measure is 0. We conclude by Corollary 1.4 that $q^* \equiv 0$ is a best approximation to f out of V^0.

But if $0 < t < 1$, $|q_0 - tq_0| = (1-t)|q_0|$. On F, then, $|f - tq_0| = |f| - t|q_0|$. By our choice of q_0,

$$\int_X |f - tq_0|\, d\mu = t\int_E |q_0|\, d\mu + \int_F |f|\, d\mu - t\int_F |q_0|\, d\mu$$

$$= \int_F |f|\, d\mu = \int_X |f - q^*|\, d\mu.$$

In our proofs of uniqueness, we shall need the following result:

THEOREM 2.3. *If $f' \in L^1(X,\mu)$ has two distinct best approximations out of V, then there are an $f \in L^1$, and a nonzero $q \in V$ such that $f - f' \in V$ and λq is a best approximation to f in V whenever $-1 \leqq \lambda \leqq 1$.*

If μ is σ-finite, we can choose a $p^* = \lambda q \neq 0$ so that $\overline{\mathrm{sgn}}(f - \lambda p^*) = \overline{\mathrm{sgn}}\, f$ for all $\lambda \in [-1,1]$ a.e. There is a real-valued function, g, such that $|g(x)| < 1$ for all x and $p^* = gf$ a.e. In particular, for almost all x, $f(x) = 0 \Rightarrow p(x) = 0$.

Proof(1). For the first assertion, let $p_1, p_2 \in V$ be distinct best approximations to f'. It then suffices to put $f = f' - (1/2)(p_1 + p_2)$, $q = (1/2)(p_1 - p_2)$.

Let $F(\lambda) = R(f) \cap Z(f - \lambda q)$. It is easy to see that when $\lambda_1 \neq \lambda_2$, $F(\lambda_1) \cap F(\lambda_2)$ is empty. If μ is σ-finite, at most countably many of the sets $F(\lambda)$, $-1 \leqq \lambda \leqq 1$, can have positive measure. Thus, we can choose a nonzero $\lambda \in [-1,1]$ such that $\mu(F(\lambda)) = \mu(F(-\lambda)) = 0$. Put $p^* = \lambda q$. We have

$$\int_X |f \pm (1/2)p^*|\, d\mu = (1/2)\int_X |f|\, d\mu + (1/2)\int_X |f \pm p^*|\, d\mu$$

$$= \int_X [(1/2)|f| + (1/2)|f \pm p^*|]\, d\mu,$$

which can be true (equality in the triangle inequality) only if $\mathrm{sgn}\, f = \mathrm{sgn}(f \pm p^*)$ a.e. on $R = R(f) \cap (R(f + p^*) \cup R(f - p^*))$. By our choice of p^*, it follows that

(1) Our argument is adapted from Ptak [14]. The theorem is similar to parts of Havinson's [5] basic theorems (numbers 2 and 3). However, Havinson omits the hypothesis, necessary for his proof, of σ-finiteness.

$\operatorname{sgn} f = \operatorname{sgn}(f \pm p^*)$ a.e. on $R(f)$. It is then clear that almost everywhere on $R(f)$, $p^*(x)/f(x)$ must be real with absolute value less than 1.

Now $\|f\| = \|f \pm p^*\|$ and

$$\|f \pm p^*\| = \int_{R(f)} (f \pm p^*)\overline{\operatorname{sgn}} f \, d\mu + \int_{Z(f)} |p^*| \, d\mu$$

$$= \|f\| \pm \int_{R(f)} p^* \overline{\operatorname{sgn}} f d\mu + \int_{Z(f)} |p^*| \, d\mu.$$

Hence $\pm \int_{R(f)} p^* \overline{\operatorname{sgn}} f \, d\mu = \int_{Z(f)} |p^*| \, d\mu = 0$. This shows that $p^* = 0$ a.e. on $Z(f)$ and completes the proof.

2.2. There are two kinds of discontinuities which we can allow without destroying uniqueness. In some cases, the functions in V can have discontinuities on sets of measure zero. f also can have discontinuities of a type which we now proceed to define.

Throughout this subsection, we shall suppose X to be a separable metric space. Let f be defined on a subset, E, of X. For each $x \in X$, we define $A(f, E, x)$ to be the set of points, c, in the extended complex plane (the plane plus a point at infinity) for which

(2.9) there is a sequence $\{x_n\}$ of (possibly identical) points in E such that $\lim x_n = x$ and $\lim f(x_n) = c$.

We say that x is of type I (with respect to f and E) if $A(f, E, x)$ contains three noncollinear complex numbers. We say that x is of type II if $A(f, E, x)$ is a (finite, infinite, or degenerate) line segment in the extended plane and

(2.10) if $\alpha \in A(\operatorname{sgn} f, E, x)$, there is a $z \in A(f, E, x)$ such that $\alpha = \operatorname{sgn} z$ $(z \neq \infty)$. If, in addition, $0 \in A(f, E, x)$, we shall say that x is of type II_0.

These definitions, as well as the lemmas which follow are essentially those of Havinson [5]. The rather awkward condition (2.10), which we shall need for Lemma 2.6 below, is trivial if $\infty \notin A(f, E, x)$. It is, in fact, omitted by Havinson, although without some such hypothesis, his main theorem, number 3, is false. We shall give a counterexample below. Since, as far as we know, Havinsons' paper is available only in Russian, we feel justified in sketching proofs of the following lemmas, which are variations of his arguments:

LEMMA 2.4. *Let f be μ-measurable on X. There is a subset, E_f, of X such that*

(2.11) $\mu(X \setminus E_f) = 0$,

(2.12) *if $E \subseteq E_f$ and $\mu(E_f \setminus E) = 0$, then $A(f, E_f, x) = A(f, E, x)$ for all $x \in X$.*

Proof. Let $\{U_n\}$ be a countable base for the topology of the extended plane. Put $V_n = f^{-1}(U_n)$. Let E_n be the carrier of the restriction of μ to the separable metric space V_n. Then $E_f = X \setminus \bigcup_{n=1}^{\infty}(V_n \setminus E_n)$ is the required set. In fact, if $c \in A(f, E_f, x)$, then for each U_n containing c, and each open neighborhood, N, of x, $N \cap V$

must have positive measure. Since $\mu(E_f \setminus E)=0$, $V_n \cap N \cap E$ also must have positive measure. It follows that $c \in A(f,E,x)$.

The discontinuous functions we shall allow are those in the class $T(\mu,S)$ of $f \in L^1$ such that every point in S is of type I or II with respect to f and E_f. Since every point in $C(\mu)$ is of type II with respect to a continuous function $T(\mu,S)$ contains all the continuous functions if $S \subseteq C(\mu)$.

Theorem 2.3 has some special consequences when the functions in V are continuous. Note that any point of continuity of $f'-f$ is of the same type with respect to f and f'.

Lemma 2.5. *Suppose p^* is continuous at x, and $p^* = gf$ a.e., where g is real and $|g| < 1$. Then if x is of type* I *or type* II_0 *with respect to f and E_f, and has no neighborhood of measure zero, then $p^*(x)=0$.*

Proof. Let E be the subset of E_f on which $p^*=gf$. Since $\mu(E_f \setminus E)=0$, $A(f,E,x)=A(f,E_f,x)$. If $p^*(x)\neq 0$, then for any sequence, $\{x_n\}$, in E converging to x such that $\lim f(x_n)=c$, $c\neq 0$ and $\operatorname{sgn} c = \lim \operatorname{sgn} f(x_n) = \lim \pm \operatorname{sgn} p^*(x_n)$ $= \pm \operatorname{sgn} p^*(x)$. Therefore, x cannot be of type I.

Suppose x is of type II_0. Then we can find a sequence, $\{x_n\}$, in E such that $\lim x_n = x$ and $\lim f(x_n)=0$. But since $|p^*(x_n)| \leqq |f(x_n)|$, $p^*(x)=\lim p^*(x_n)=0$.

Lemma 2.6. *Suppose $p^*=gf$ a.e., where g is real and $|g|<1$. Let U be an open connected set on which p^* is continuous and such that*

(2.13) *U contains only points of type* II *with respect to f and E_f,*

(2.14) *p^* does not vanish on U,*

(2.15) *no nonempty open subset of U has measure zero.*

Then either $\operatorname{sgn} f = \operatorname{sgn} p^*$ *or* $\operatorname{sgn} f = -\operatorname{sgn} p^*$ *a.e. on U.*

Proof. It is enough to show that $\operatorname{sgn} f$ is equal almost everywhere on U to a continuous function, α, for then $\alpha/\operatorname{sgn} p^*$ must be 1 or -1 everywhere on U.

Let E be the subset of E_f where $p^*(x)=g(x)f(x)$. For each $x\in U$, choose a $z(x)\in A(f,E,x)$ distinct from ∞. The existence of such a $z(x)$ is implied by (2.10). Put $\alpha(x)=\operatorname{sgn} z(x)$.

If $x\in U$, $c\in A(f,E,x)$, $c\neq\infty$, then there is a sequence, $\{x_n\}$, in $E\cap U$ such that $\lim x_n = x$ and $\lim f(x_n)=c$, which cannot be 0 by (2.13), (2.14), (2.15) and Lemma 2.5. Thus $\operatorname{sgn} c = \lim \operatorname{sgn} f(x_n) = \lim \pm \operatorname{sgn} p^*(x_n) = \pm \lim \operatorname{sgn} p^*(x_n)$ $= \pm \operatorname{sgn} p^*(x)$.

Moreover, if $x\in E\cap U$, $f(x)\in A(f,E,x)$. Thus $A(f,E,x)$ is an interval, all of whose points have sign $\pm\, p^*(x)$, which does not contain 0, but which does contain $f(x)$. That is, on $E\cap U$, $\alpha(x)$ must equal $\operatorname{sgn} f(x)$.

Since α is bounded, discontinuity of α at x would imply the existence of two sequences, $\{x_n\}$ and $\{y_n\}$ in U, converging to x, such that $\lim \alpha(x_n)$ and $\lim \alpha(y_n)$ exist, but are *unequal*. Because $\mu(U\setminus E)=0$, there must be, by the very definition of α, v_n and u_n in $E\cap U$ such that $|u_n - x_n|$, $|v_n - y_n|$, $|\operatorname{sgn} f(u_n) - \alpha(x_n)|$, and

$|\operatorname{sgn} f(v_n) - \alpha(y_n)|$ are all less than $1/n$. Sgn $f(u_n)$ and sgn $f(v_n)$ must also have unequal limits, which must be $\pm$ sgn $p^*(x)$. By (2.10), there are s and t in $A(f,E,x) = A(f,E_f,x)$ such that $\operatorname{sgn} s = \operatorname{sgn} p^*(x)$ and $\operatorname{sgn} t = -\operatorname{sgn} p^*(x)$. But then the interval $A(f,E,x)$ must contain 0, which contradicts either (2.14) or Lemma 2.5. Q.E.D.

REMARK. Lemmas 2.5 and 2.6 do not depend upon the particular form of E_f, but only on the two properties (a) $\mu(X \setminus E_f) = 0$ and (b) $E \subseteq E_f$, $\mu(E_f \setminus E) = 0$ $\Rightarrow A(f,E,x) = A(f,E_f,x)$.

Finally, let us remark that in the present context, an atom is essentially a point of positive mass. As a topological subspace of X, an atom, A, is itself a separable metric space, *ipso facto* having a countable base for its topology. The restriction of μ to A consequently has a carrier, C. If C contained two distinct points, x and y, then there would be disjoint neighborhoods, U and V, of x and y *in* A. Both would have positive measure. This would contradict the fact that A is an atom. Thus C is a single point, and A is essentially C.

2.3. An interval of the real line is disconnected by the removal of a single interior point. In view of this fact, Lemma 2.6 suggests that when X is an interval, an important rôle in the uniquenesss problem will be played by a condition limiting the numbers of zeros of functions in V. If V has dimension n, $E \subseteq X$, and no nonzero function in V vanishes more than $n-1$ times in E, we say that V is a *Čebyšev space over E*. For example, the polynomials of degree $\geqq n-1$ form a Čebyšev space over any subset of the complex plane containing n or more points.

We shall need the following fact (cf. Kreĭn [9]) about Čebyšev spaces.

LEMMA 2.7. *Let V be a Čebyšev space of real-valued continuous functions on $[a,b]$, of dimension n. If $x_1,\cdots,x_k$, $k < n$, are distinct points interior to $[a,b]$, there exists a nonzero $p \in V$ which changes sign at $x_1,\cdots,x_k$, but at no other point in $[a,b]$.*

Jackson [6] proved that if X is a compact interval of the real line, μ is Lebesgue measure, $1,x,\cdots,x^{n-1}$span V, and f is real and continuous, then f has a unique best approximation out of V. Historically this result has stimulated much of the research on L^1 approximation.

Jackson's theorem was extended by Kreĭn [10] to the case when V is any n-dimensional Čebyšev subspace of $L^1_R(X,\mu)$(μ is still Lebesgue measure) over the compact real interval, X, and all functions involved are continuous. The theorem which follows is, to our knowledge, the first extension of these results to the case of complex-valued functions. Results similiar to the real case of our theorem were given by Havinson [5], but with some errors.

Observe that the behavior of μ and f on any set where all the functions in V vanish cannot affect the uniqueness of best approximations to f out of V. If $V$is a finite-dimensional subspace of $L^1(X,\mu)$, $Z(V)$ is measurable. Let μ_V be the measure defined by $\mu_V(G) = \mu(G \cap R(V))$. If G is a subset of X, we denote the closure of G relative to X by "$\bar{G}$."

Condition $J(G)$: V is a subspace of $L^1(X,\mu)$ such that:

(2.16) V is an n-dimensional Čebyšev space over G ($n \geqq 1$),

(2.17) $G \subseteq C(\mu_V) \subseteq \bar{G}$,

(2.18) the functions in V are continuous on G.

If, in addition,

(2.19) $p \in V \Rightarrow \operatorname{Re} p \in V$,

(2.20) $\bar{G} \setminus Z(V)$ is atom-free,

we say that V satisfies condition $J'(G)$.

We say that $J(G)$ (respectively $J'(G)$) guarantees uniqueness for a class, $\mathscr{C}$, of functions in L^1 if each $f \in \mathscr{C}$ has a unique best approximation out of each V satisfying $J(G)$ (respectively $J'(G)$).

THEOREM 2.8. *Let X be a (finite or infinite) real interval, μ a σ-finite measure on X.*

(a) *Let $G = \bar{G}$ be a disconnected subset of X, and let there be at least one V satisfying $J(G)$. Then $J(G)$ guarantees uniqueness for the functions in L^1 which are continuous on G if, and only if, G consists of n points of positive mass.*

If $\mu(X) < \infty$ or there is a V consisting of functions continuous on all of X which satisfies $J(G)$, the same condition is necessary and sufficient for $J(G)$ to guarantee uniqueness for the functions in L^1 which are continuous on all of X.

(b) *If G is a nonempty open subinterval of X, then $J'(G)$ guarantees uniqueness for $T(\mu, G)$.*

Proof. (a) If G consists of n points of positive mass, then since V is n-dimensional, the unique best approximation to f out of V is the unique $p \in V$ which is equal to f on G.

Conversely, let G be a relatively closed disconnected subset of X, and let V satisfy $J(G)$.

Case 1. G is a finite set. Then since $G = C(\mu_V)$, each point in G must be a point of positive mass for μ_V. By Theorem 2.2(a), G must contain at least n points. Suppose it contains more than n.

Let x be one of them. Put $E = G \setminus \{x\}$, $F = \{x\}$. Then the hypotheses of Theorem 2.2(b) are satisfied if we take the space X of that theorem to be G, and the measure μ of that theorem to be the restriction of μ_V to G. In fact, (2.5) is verified because no nonzero $p \in V$ can vanish on n points, and (2.6) is verified because if x were in $Z(V)$, it could not be an isolated point of $C(\mu_V)$.

Case 2. G is infinite. Then G is a union of two disjoint sets, closed relative to X, at least one of which, say "E," must be infinite. Call the other "F." We are once more in the situation of the preceding paragraph, and again can apply Theorem 2.2(b).

In either case, we find a function, f, continuous *on* G, which has more than one best approximation out of a space V^0, satisfying $J(G)$.

If the functions in V are continuous on all of X, there is no need to restrict the application of Theorem 2.2(b) to G—we can apply it to all of X, and find that f is continuous on X.

If $\mu(X) < \infty$ and g is the characteristic function of G ($g \equiv 1$ on G, $g \equiv 0$ on $X \setminus G$), then $ge^{-x^2}, xge^{-x^2}, \cdots, x^{n-1}ge^{-x^2}$ is a basis for a space, V, of *bounded* functions satisfying $J(G)$. The function, f, which we construct on G can then be extended, by the Tietze extension theorem (Kelley [8, p. 242]) to a bounded continuous function on all of X. The extension of f lies in L^1 because $\mu(X) < \infty$.

In either case, we have a function continuous on all of X with more than one best approximation out of a space, V^0, satisfying $J(G)$.

(b) Now suppose that G is a nonempty open subinterval of X, and that V satisfies $J'(G)$. Suppose also that there were an $f' \in T(\mu, G)$ possessing more than one best approximation out of V. Theorem 2.3 then asserts that there are an $f \in T(\mu, G)$ and a nonzero $p^* \in V$ such that $\operatorname{sgn} f = \operatorname{sgn}(f - \lambda p^*)$ when $-1 \leqq \lambda \leqq 1$.

The sets $F(\lambda) = G \cap Z(\operatorname{Re}(f - \lambda p^*))$ intersect, for distinct $\lambda \in [-1, 1]$, only in points where $f = p^* = 0$. Thus, since μ is σ-finite, uncountably many of them can have positive mass only if $\operatorname{Re} f = \operatorname{Re} p^* = 0$ on a subset of G of measure greater than zero. By (2.19), (2.16), and (2.20), that would imply that $\operatorname{Re} p^*$ is identically 0.

A similar argument holds for $\operatorname{Im} p^*$. Since $p^* \neq 0$, we can choose a nonzero $\lambda \in [-1, 1]$ such that one or the other of $G \cap Z(\operatorname{Re}(f - \lambda p^*))$ and $G \cap Z(\operatorname{Im}(f - \lambda p^*))$ has measure zero. $C(\mu_V) \cap G \cap Z(f - \lambda p^*)$ is contained in both of these sets. (2.17), (2.20) and the definition of μ_V thus imply that $\mu_V(Z(f - \lambda p^*)) = 0$.

λp^* is a best approximation to f with respect to μ_V as well as with respect to μ. We can use Corollary 1.5 and Theorem 2.3 to conclude that for every $p \in V$, $0 = \int_X p \overline{\operatorname{sgn}}(f - \lambda p^*)\, d\mu_V = \int_X p \overline{\operatorname{sgn}}(f - \lambda p^*)\, d\mu = (p, \operatorname{sgn} f)$. In particular, if p is real, $(p, \operatorname{Re}\operatorname{sgn} f) = \operatorname{Re}(p, \operatorname{sgn} f) = 0 = \operatorname{Im}(p, \operatorname{sgn} f) = (p, \operatorname{Im}\operatorname{sgn} f)$.

Lemma 2.5 shows that there can be at most $n - 1$ points in G of type I with respect to f and E_f, say $x_1, \cdots, x_k$. The remaining points are all of type II with repect to f and E_f.

One or the other of $\operatorname{Re} p^*$ and $\operatorname{Im} p^*$ must not vanish identically on G—the former, let us say. Since $\operatorname{Re} p^*$ must vanish along with p^* at $x_1, \cdots, x_k$, (2.16) implies that there are at most $n - k - 1$ other points, say $x_{k+1}, \cdots, x_s$, at which $\operatorname{Re} p^* = 0$. Let $U_1, \cdots, U_{s+1}$ be the intervals which make up $G \setminus \{x_1, \cdots, x_s\}$. Then on each U_i, f and p^* satisfy the hypotheses of Lemma 2.6. We conclude that either $\operatorname{sgn} f = \operatorname{sgn} p^*$ or $\operatorname{sgn} f = -\operatorname{sgn} p^*$ a.e. on each U_i. Since $\operatorname{Re} p^*$ does not vanish on any U_i, on each of these intervals $\operatorname{sgn}\operatorname{Re} f = \pm \operatorname{sgn}\operatorname{Re} p^*$ is a *constant*.

Lemma 2.7 allows us to choose a $p_0 \in V$ which has the *same sign* as $\operatorname{Re} f$ on each U_i. Therefore, $(p_0, \operatorname{Re}\operatorname{sgn} f) > 0$, which is a contradiction. Q. E. D.

REMARK. If $\operatorname{Im} f \equiv 0$ on G, then $\operatorname{Im} p^* \equiv 0$ on X for any best approximation, p^*, in V to f. If this were not so, it is easy to see that $\operatorname{Re} p^*$ would be a better approximation out of V to f than p^*. We have *not* proved, however, that $\operatorname{Re} p^*$ is always a best approximation in V to $\operatorname{Re} f$.

Our proof of Theorem 2.8 also establishes the following characterization of the extremals, which we shall need in §3.

THEOREM 2.9. *Let X be a real interval, G an open subinterval of X, μ a σ-finite measure on X. Let f be a real-valued function in L^1 which is continuous on G. Let V be a subspace of L^1 satisfying $J'(G)$, and p^* a best approximation to f out of V. Then $f - p^*$ is real on G, and changes sign at least n times in G unless $\mu(G \cap Z(f - p^*)) > 0$.*

Walsh and Motzkin [17] have proved Theorem 2.8(b) in the real case under far more restrictive conditions. Not only does their method lead to an extremely simple proof of Jackson's Theorem, but it also gives a more detailed characterization of the extremals than that above in the case that $\mu(G \cap Z(f - p^*)) > 0$.

2.4. Our next result generalizes a theorem due to Havinson [5].

THEOREM 2.10. *Let X be a metric space, and μ a σ-finite measure on X which assigns positive mass to each nonempty open set. Let V be a subspace of $L^1(X,\mu)$ such that if $p, p' \in V$, $p' \neq 0$,*

(2.21) *p is continuous except on a set $S(p)$ of measure zero.*

(2.22) *$X \setminus (S(p) \cup S(p') \cup Z(p'))$ is nonempty and connected.*

Then each $f \in T(\mu, X)$ has at most one best approximation out of V.

Proof. Suppose there were an $f' \in T(\mu, X)$ having two best approximations in V. Then by Theorem 2.3, there are an $f \in L^1$, a real g with $|g| < 1$, a $p \in V$, and a nonzero $p^* \in V$ such that:

(a) $f \in T(\mu, X \setminus S(p))$,

(b) $\pm p^*$ and 0 are best approximations in V to f,

(c) $\operatorname{sgn} f = \operatorname{sgn}(f \pm p^*)$ a.e.,

(d) $p^* = gf$ a.e.

$D = X \setminus (S(p) \cup S(p^*) \cup Z(p^*))$ is nonempty and connected, and by Lemma 2.5, all the points of D are of type II with respect to f. D is an open subspace of $Y = X \setminus (S(p) \cup S(p^*))$. Since $\mu(X \setminus Y) = 0$, $A(f, E \cap E_f, x) = A(f, E_f, x)$ whenever $\mu(Y \setminus E) = 0$. Applying Lemma 2.6 to Y and D, we conclude that $\operatorname{sgn} f = \operatorname{sgn} p^*$ or $\operatorname{sgn} f = -\operatorname{sgn} p^*$ throughout D; the former, let us say.

Then a.e. on D, $|f - p^*| = (f - p^*)\overline{\operatorname{sgn}}(f - p^*) = |f| - |p^*|$, by (c). In view of (b),

$$\|f\| = \|f - p^*\| = \int_{D \cup Z(p^*)} |f - p^*| \, d\mu = \int_D |f - p^*| \, d\mu + \int_{Z(p^*)} |f| \, d\mu$$

$$= \int_D (|f| - |p^*|) \, d\mu + \int_{Z(p^*)} |f| \, d\mu = \|f\| - \int_D |p^*| \, d\mu.$$

This shows that $p^* \equiv 0$ a.e., a contradiction which proves the theorem.

We now state several corollaries of the preceding theorem. The first is a strengthening of Theorem 2.8 in the case that $n = 1$.

COROLLARY 2.11. *When* $n = 1$, *Theorem* 2.8 (b) *remains valid if we require only that* V *satisfy condition* $J(G)$. (2.19) *and* (2.20) *can be omitted.*

COROLLARY 2.12. *Let* X *be a simple closed curve, and* V *a two dimensional Čebyšev subspace of* L^1 *over* X, *consisting of continuous functions. If* μ *is a* σ*-finite measure on* X *which assigns positive mass to each nonempty open set, then each* $f \in T(\mu, X)$ *has a unique best approximation out of* V.

The next corollary extends a result of Havinson [5]. There is an error, however, in Havinson's proof: he applies to meromorphic functions a theorem (number 5) in which it is assumed that the approximators have no discontinuities. The error can be corrected by some slight alterations. We call a function on a complex manifold "meromorphic" if it is, at least locally, a quotient of holomorphic functions (with the denominator not identically zero).

COROLLARY 2.13. *Let* X *be a connected complex manifold of dimension* $n \geqq 1$, *and* μ *a* σ*-finite measure on* X *which assigns positive mass to each nonempty open set. If* V *is a subspace of* $L^1(X, \mu)$ *consisting of meromorphic functions, each* $f \in T(\mu, X)$ *has at most one best approximation out of* V.

Proof. It is clear that the set of singularities, $S(p)$, of a meromorphic $p \in L^1$ must have measure 0. If $p, p' \in V$, and p' is not identically 0, then $M = S(p) \cup S(p') \cup Z(p')$ is an analytic variety in X of complex dimension $n - 1$. We have to show that $X \setminus M$ is connected.

Each point $x \in X$ has a neighborhood, U_x such that $U_x \setminus M$ is connected (Bochner and Martin [2, p. 196, Lemma 5]). If x_0, $x_1 \in X \setminus M$, we can join them by an arc in $X: x = x(t), x(0) = x_0, x(1) = x_1$. Choose points $0 = t_0 < t_1 < \cdots < t_k = 1$ in $[0,1]$ such that $x([t_i, t_{i+1}])$ is contained in one of these neighborhoods, say U_i, for $i = 0, \cdots, k-1$. Clearly, $(U_i \cap U_{i+1}) \setminus M$ is nonempty for each i, since $x(t_{i+1}) \in U_i \cap U_{i+1}$, and M has complex dimension $n-1$. It follows that $(U_0 \setminus M) \cup \cdots \cup (U_{k-1} \setminus M)$ is a connected subset of $X \setminus M$ containing x_0 and x_1.

2.5. We conclude this section with some counterexamples whose purpose is to show that certain of the hypotheses of Theorem 2.8 cannot be weakened.

Although condition (2.19) had an obvious utility in our proof, one might hope that it were not actually needed to extend Jackson's theorem to the complex case, especially in view of what we have shown in Corollary 2.11. However, even when $n = 2$, uniqueness may fail if V is not spanned by its real-valued functions.

Nevertheless, it is natural to ask whether uniqueness might not at least obtain in the following setting: X is an arc in the complex plane, μ is a well-behaved measure on X, V is the Čebyšev space of polynomials of degree less than n, and f

is analytic on a neighborhood of X. Again, in general, the answer is no, and this is all the more remarkable in view of the fact that certain special geometries do guarantee uniqueness. For example, if X is a line segment, uniqueness is a consequence of Theorem 2.8(b) and the fact that on X the real part of a polynomial is again a polynomial. Approximation is also unique in this context when X is an arc of a circle (Havinson [5, Theorem 11]).

A single example will establish the possibility of nonuniqueness in both of the preceding situations.

EXAMPLE 2.14. There are an *analytic* arc, X, in the complex plane, and a function, f, analytic in a neighborhood of X which has infinitely many best approximations by *linear* polynomials on X. The measure, μ, is a *continuous* multiple of the arc length, $ds: \mu(E) = \int_E g(s)\, ds$, where g is continuous on X.

We take X to be the arc described by $z(t) = te^{it^2}$, $-\sqrt{(2\pi)} \leqq t \leqq \sqrt{(2\pi)}$. It is clear that $z(t)$ is analytic and one-to-one on $[-\sqrt{(2\pi)}, \sqrt{(2\pi)}]$. Moreover, $z'(t) = e^{it^2}(1 + 2t^2 i)$ never vanishes on this interval.

LEMMA. *$z(t)$ is one-to-one on a neighborhood, U, of $[-\sqrt{(2\pi)}, \sqrt{(2\pi)}]$ in the complex plane.*

Proof. (This argument can be formulated for compact metric spaces in general.) Let M be a compact neighborhood of $[-\sqrt{(2\pi)}, \sqrt{(2\pi)}] = S$. Because $z' \neq 0$ *on* S, there is an $\varepsilon > 0$ such that $s \in S$, $|x - s| < \varepsilon$, $|y - s| < \varepsilon$, and $z(x) = z(y)$ imply that $x = y$. Moreover, $z(t)$ is a homeomorphism on S, so there is a $\delta > 0$ such that if $s, t \in S$ and $|z(s) - z(t)| < \delta$, then $|s - t| < \varepsilon/2$.

Choose $\eta > 0$ so that $x, y \in M$, $|x - y| < \eta \Rightarrow |z(x) - z(y)| < \delta/2$. Let U be the set of points in M whose distance from S is less than $\alpha = \min(\eta, \varepsilon/2)$. If $x, y \in U$ and $z(x) = z(y)$, then there are s and t in S such that $|x - s|$ and $|y - t|$ are less than α. Therefore, $|z(s) - z(t)| \leqq |z(s) - z(x)| + |z(y) - z(t)| \leqq \delta/2 + \delta/2 = \delta$. It follows that $|s-t| < \varepsilon/2$, whence $|x-s| < \varepsilon/2 < \varepsilon$ and $|y-s| < |y-t| + |t-s| < \varepsilon$. By our choice of ε, $x = y$. Q. E. D.

If $W = z(U)$, there is a one-to-one analytic function $t(z)$ on W, such that $z = t(z)e^{i[t(z)]^2}$. Put $f(z) = e^{i[t(z)]^2}$ on W, so that on $[-\sqrt{(2\pi)}, \sqrt{(2\pi)}]$, $f(z(t)) = e^{it^2}$ and $\overline{\operatorname{sgn} f(z(t))} = e^{-it^2}$. We set $\mu(E) = \int_{z^{-1}(E)} |t|\, dt = \int_E |t(z)| (1 + 4[(t(z)]^4)^{-1/2}\, ds$. Thus $\int_X z\, \overline{\operatorname{sgn} f(z)}\, d\mu = \int_{-\sqrt{(2\pi)}}^{\sqrt{(2\pi)}} te^{it^2} e^{-it^2} |t|\, dt = 0$, because $t|t|$ is an odd function, and similarly $\int_X \overline{\operatorname{sgn} f(z)}\, d\mu = 0$.

This shows that $\overline{\operatorname{sgn} f}$ is orthogonal to the space V of linear polynomials, so that 0 is a best approximation to f out of V. But a routine calculation shows that $\|f\| = \|f - az\|$ whenever $-1/\sqrt{(2\pi)} \leqq a \leqq 1/\sqrt{(2\pi)}$.

We next investigate the condition, (2.20), that $\bar{G} \setminus Z(V)$ contain no atoms, which also is unnecessary when $n = 1$. Havinson [5] actually omits this stipulation—erroneously, as we shall see.

EXAMPLE 2.15. Let X and μ be as in Theorem 2.8, with $\mu(X) < \infty$, and suppose there is a point, c, interior to $C(\mu)$ whose mass is positive. Let $G \subseteq C(\mu)$ be an open interval containing c. Then there are p_1 and $p_2 \in L^1(X,\mu)$ which span a subspace, V, satisfying conditions $J(G)$ and (2.19), and a real-valued continuous $f \in L^1$ which has infinitely many best approximations out of V.

Say that $G =]\,r,s\,[$. We define a function $p(a,b;x)$ by setting $p(a,b;x) = e^{a(x-c)}$ on $[r,c]$, $p(a,b;x) = e^{-b(x-c)}$ on $[c,s]$, and $p(a,b;x) = 0$ on $X \setminus \bar{G}$. By hypothesis $\mu([r,c[)$, $\mu(]\,c,s]) > 0$. The Lebesgue dominated convergence theorem guarantees that we can choose a so large, and then b so large that $\int_{[r,c[} p(a,b;x)\,d\mu < 1/2\mu(c)$, $\int_{]c,s]} p(a,b;x)\,d\mu < 1/2\mu(c)$, and

$$\int_{]c,s]} p(a,b;x)(x-c)\ d\mu \leqq \left| \int_{[r,c[} (x-c)\,p(a,b;x)\ d\mu \right|. \tag{2.23}$$

The right-hand expression in (2.23) is a continuous monotonically decreasing function of a, tending to zero as a increases. Therefore, by choosing a still larger value for a if need be, we can arrange that (2.23) actually be an equality.

Having chosen a and b in this way, we now put $p_1(x) = p(a,b;x), p_2(x) = (x-c)\,p_1(x)$. It is easy to see that the subspace, V, which they span satisfies the conditions (2.19) and $J(G)$.

Let $f(x) = |\,x - c\,|$. Then

$$\begin{aligned} \left| \int_{X\setminus Z(f)} (up_1 + vp_2)\overline{\operatorname{sgn}} f\ d\mu \right| &= \left| \int_{X\setminus Z(f)} up_1\ d\mu \right| \\ &\leqq |\,u\,|\,\mu(c) = \int_{Z(f)} |\,up_1 + vp_2\,|\,d\mu. \end{aligned}$$

An application of Corollary 1.4 shows that 0 is a best approximation out of V, and another routine calculation shows that $\|f\| = \|f - \lambda p_2\|$ for all $\lambda \in [-1,1]$.

As a counterweight to this example, we show that even when there is an atom interior to $C(\mu)$, uniqueness may obtain for a *particular* V (of dimension $\geqq 2$).

EXAMPLE 2.16. Let $X = [-1,2]$, and let μ be Lebesgue measure, except that $\mu(0) = 7$. Then each $f \in T(X,\mu)$ has a unique best approximation by linear polynomials.

In fact, if $p^*(x) = c + dx$ is a best linear approximation to an $f \in T(X,\mu)$, we can easily see that c must equal $f(0)$ in order to avoid violating the condition (1.3).

Suppose (Theorem 2.3) that 0 and $\pm\, p^* \neq 0$ were all best linear approximations to f. Then we should have $c = 0$, $d \neq 0$. By Lemmas 2.5 and 2.6, $\operatorname{sgn} f$ must be constant on $[-1,0[$ and on $]0,2]$, where it must be $\pm \operatorname{sgn} d \neq 0$. By Theorem 1.3, we must have $|(x,\operatorname{sgn} f)| \leqq \int_{Z(f)} |x|\ d\mu = \int_{\{0\}} |x|\ d\mu = 0$, which is impossible under the conditions we have just derived.

It is also easy to give examples of *particular* measures, μ, and spaces V satisfying $J(C(\mu))$ such that each $f \in T(X,\mu)$ has a unique best approximation out of V, despite the fact that $C(\mu)$ consists of $\dim V + k$ atoms ($k > 0$). Moreover, the argument of Example 2.16 shows that if $X = [-1,2]$ and μ is Lebesgue measure, each $f \in T(X, \mu)$ has a unique best approximation by polynomials of the form cx, even though such polynomials do not form a Čebyšev space over the interior of X. The condition (2.16) cannot, however, be dropped in general, as we see by the following example:

EXAMPLE 2.17. Suppose $\mu(X) < \infty$ and $\int_X p \, d\mu = 0$ for every $p \in V$. There are many non-Čebyšev spaces satisfying this criterion. E.g., V might be the subspace of $L^1([0, 2\pi], dx)$ spanned by $\cos x, \cdots, \cos nx$.

Under these conditions, it is immediate that each real-valued $p \in V$ such that $|p(x)| \leqq 1$ a.e. is a best approximation out of V to the constant function, 1.

If, in the setting of Theorem 2.8(b), f is bounded and real-valued, but is *not* in $T(\mu, G)$, then $A(f, E_f, x)$ must be a disconnected subset of an interval for some x. It is not hard to show, then, that f will have more than one best approximation out of some Čebyšev space, V, of dimension 1 (Havinson [5, Theorem 10]). An example of a similar nature shows that the condition (2.10), which Havinson omitted, is actually necessary for Theorem 2.8(b) to be true.

EXAMPLE 2.18. Let $X = [-1, 1]$, and let $\mu(E) = \int_E |x| \, dx$. Let V be the space of constant functions, and $f(x) = 1/x$ for $x \neq 0$. Then any $p \in V$ such that $-1 \leqq p \leqq 1$ is a best approximation in V to f. Every point in X is of type II with respect to f and $E_f = [-1, 1] \setminus \{0\}$, except that at 0 the condition (2.10) fails.

EXAMPLE 2.19. Let $X = [0, 3]$, and let μ be the measure which assigns mass 1 to each rational in $[0, 1] \cup [2, 3]$ and is zero elsewhere. If $p \in L^1(X, \mu)$ is continuous on $C(\mu) = [0, 1] \cup [2, 3]$, then p must vanish on $C(\mu)$. This shows why we bothered to worry in stating Theorem 2.8(a) whether or not there is any V satisfying $J(G)$. The condition $\mu(X) < \infty$ serves to eliminate the unpleasant possibility that any $f \in L^1$ continuous on all of X would have to vanish a.e.

3. **Two special cases: Approximation on a finite point set and approximation to a real continuous function on an interval.** Having seen in §2 how strongly the presence of atoms may affect the uniqueness of best approximations, we devote the first part of this section to a study of the extreme case in which X is a finite point set. We could give our discussion an appearance of greater generality by allowing X to consist of finitely many atoms, but this would change no essential feature of the argument. In the second part of this section, we apply our general theory to the classical problem of approximating real continuous functions on an interval.

3.1. Throughout this section, we confine ourselves to the real case. We shall suppose that X contains N points, each of which is measurable and has positive mass. Let us fix a subspace, V, of L^1, and an $f \in L^1$. We propose to study the set B of best approximations to f out of V. It is immediate that B is closed, bounded (hence compact), convex, and since L^1 has finite dimension (N), B is nonempty. Our approach is to examine the consequences of the variational lemma (Theorem 1.3) in the present situation.

Suppose that $p_0 \in B$, so that for each $p \in V$, (1.3) holds. Suppose also that (1.4) does not hold for every $p \in V$. That is,

$$|(p, \operatorname{sgn}(f - p_0)| = \int_{Z(f-p_0)} |p| \, d\mu \tag{3.1}$$

for some $p \in V$. For otherwise $B = \{p_0\}$ according to Corollary 1.4.

As a compact convex set in L^1, B is the closed convex hull of its extreme points (the Kreĭn-Milman Theorem; cf. Dunford and Schwartz [4, p. 440]). Our principal results are summarized as follows:

THEOREM 3.1. *Let E be the set of extreme points of B. Each $q \in E$ takes the same values as f on a nonempty subset of X, but no two distinct functions in E agree with f on the same subset of X.*

If $q_1, \cdots, q_m$ are distinct functions in E, and T is defined by

$$T = \{r_1 q_1 + \cdots + r_m q_m : r_k > 0,\ k = 1, \cdots, m;\ r_1 + \cdots + r_m = 1\} \tag{3.2}$$

then any two functions in T agree with f on the same subset of X, but for each $k = 1, \cdots, m$ and $q \in T$, there is an $x \in X$ such that $q(x) \neq f(x) = q_k(x)$.

When (3.1) *holds for a nonzero $p \in V$, $B \cap \{p_0 + tp : -\infty < t < \infty\}$ is a nontrivial interval. If*

$$\int_{Z(f-p_0)} |p| \, d\mu = 0 \tag{3.3}$$

then p_0 is in the interior of this interval; othervise p_0 is an endpoint of this interval.

REMARK 1. Since, by the theorem, E is a finite set, B is actually the convex hull of E, because the latter is closed.

REMARK 2. In the present case, (1.4) is also a necessary condition for $\|f - p_0\| < \|f - (p_0 + tp)\|$ for all real $t \neq 0$.

Proof. We begin with the statements of the final paragraph. If $|t|$ is small enough, then $f - p_0 - tp$ has the same sign as $f - p_0$ on the set $R(f - p_0)$ where $f - p_0 \neq 0$. This is a consequence of our assumption that X, and hence $R(f - p_0)$, is finite. Thus for $|t|$ sufficiently small,

$$\begin{aligned} \|f - p_0 - tp\| - \|f - p_0\| &= |t| \int_{Z(f-p_0)} |p| \, d\mu \\ &\quad + \int_{R(f-p_0)} [(f - p_0 - tp) - (f - p_0)] \operatorname{sgn}(f - p_0) d\mu \\ &= |t| \int_{Z(f-p_0)} |p| \, d\mu - t \int_X p \operatorname{sgn}(f - p_0) \, d\mu. \end{aligned} \tag{3.4}$$

The case when (3.3) holds is now immediate.

Suppose

$$\int_{Z(f-p_0)} |p| \, d\mu > 0. \tag{3.5}$$

Then if $t \neq 0$, the last expression in (3.4) is either 0, or is $2|t| \int_{Z(f-p_0)} |p| \, d\mu > 0$ depending on the signs of $(p, \operatorname{sgn}(f - p_0))$ and t. This proves our final statements. They are used to prove the rest of the Theorem.

The set where p_0 agrees with f is $Z(f - p_0)$. If $p_0 \in E$, then p_0 is, by definition, contained in the interior of no segment in B. Therefore, for each $p \in V$, either (1.4) or (3.5) must hold. In neither case can $Z(f - p_0)$ be empty.

However, if p_0 is interior to a line segment $B \cap \{p_0 + tp\}$, (3.3) must hold. It follows that p vanishes on $Z(f - p_0)$ and hence $Z(f - p_0) \subseteq Z(f - p_0 - tp)$, for all t.

Let $q_1, \cdots, q_m$ be distinct functions in E, and let T be defined by (3.2). If u, $v \in T$, then both are interior to $B \cap \{u + t(v - u)\}$. Thus $Z(f - u) \subseteq Z(f - v)$; and similarly $Z(f - v) \subseteq Z(f - u)$. In other words, u and v agree with f on the same subset of X. Since q_k is an extreme point, it must be an endpoint of the segment $B \cap \{q_k + t(q - q_k)\}$ for a given $q \in T$. Thus (3.5) applies, so that there is an $x \in R(q - q_k) \cap Z(f - q_k)$. We then have $f(x) = q_k(x) \neq q(x)$.

If we apply this last result in the case $m = 2$, we find that the assumption that $Z(f - q_1) = Z(f - q_2)$ leads to a contradiction: $Z(f - q) \supseteq Z(f - q_1) = Z(f - q_2)$ for each $q \in T$. Thus q_1 and q_2 do not agree with f on the same set.

The following result of Motzkin and Walsh [12] is a corollary of what we have just proved:

THEOREM 3.2. *If V has dimension n, then each extreme function of B agrees with f on at least n points of X.*

Proof. We first note that for any $n - 1$ points $x_1, \cdots, x_{n-1} \in X$, there is a nonzero $p \in V$ which vanishes on $x_1, \cdots, x_{n-1}$. Suppose this were false. Then $p(x_1) = \cdots = p(x_{n-1}) = 0$ implies that $p \equiv 0$. Therefore, the restriction map of V into $L^1(\{x_1, \cdots, x_{n-1}\}, \mu)$ would be a linear isomorphism, which is impossible, because the dimension of $L^1(\{x_1, \cdots, x_{n-1}\}, \mu)$ is only $n - 1$.

Now if the theorem were false, there would be a set $S \subseteq X$ of, at most, $n - 1$ points, and an extreme function $p_0 \in B$ agreeing with f exactly on S. Then if we choose a nonzero $p \in V$ which vanishes on S, we have

$$\int_{Z(f-p_0)} |p| \, d\mu = \int_S |p| \, d\mu = 0.$$

But then (3.3) is satisfied, and p_0 is not an extreme point. The theorem is proved.

When $n = N - 1$, this last result can be refined by a combinatorial argument.

THEOREM 3.3. *If V has dimension $n = N - 1$, $q_1, \cdots, q_m$ are distinct extreme* *points of B, and T is defined by* (3.2), *then if $f \notin V$, each $q \in T$ agrees with f just on the $n-m+1$ points at which $f(x) = q_1(x) = \cdots = q_m(x)$.*

Proof. By Theorem 3.2, each q_k agrees with f on at least n points. Since $f \notin V$, there is then exactly one point, $x_k \in X$, at which $q_k(x_k) \neq f(x_k)$. By Theorem 3.1, the points $x_1, \cdots, x_m$ are distinct. On the remaining $n+1-m$ points of X, each of $q_1, \cdots, q_m$, and hence each $q \in T$, agrees with f. For $q \in T$ we have $q(x_k) = r_1 q_1(x_k) + \cdots + r_m q_m(x_k) = (1-r_k)f(x_k) + r_k q_k(x_k) \neq f(x_k)$.

REMARK 1. Every $q \in B$ which agrees with f on a set of n points is an extreme oint of B. B can thus be determined by examining the elements of V which agree with f at n points, since every extreme point of B is to be found among them.

REMARK 2. Suppose $f \notin V$. Let p_0 be an extreme best approximation to f out of V, agreeing with f at $x_1, \cdots, x_n$. Then for any $g \notin V$, a best approximation to g out of V can be found by choosing the unique $p \in V$ which agrees with g on $x_1, \cdots, x_n$.

In fact, g can be represented uniquely as $g = q + af$, $q \in V$. Then $q + ap_0$ is a best approximation to g out of V which interpolates g on $x_1, \cdots, x_n$. No other $p \in V$ can agree with g on these points, for if there were another, there would be a nonzero $p' \in V$ which vanished on $x_1, \cdots, x_n = Z(f - p_0)$, so that p_0 would not be extreme.

3.2. We conclude with a brief survey of some of the classical properties of L^1-approximation in the following setting: $C(\mu)$ is a nontrivial atom-free real interval, f is real and continuous on $C(\mu)$, and V is an n-dimensional Čebyšev space over the interior of $C(\mu)$ consisting of real continuous functions. It is a remarkable feature of this situation that in some cases we can determine a canonical set of n points interior to $C(\mu)$ so that the function of best approximation out of V is the one which interpolates f at these points. We shall give special attention to the case in which μ is Lebesgue measure and V consists of ordinary polynomials. In this case, the canonical nodes can be determined explicitly.

Following Laasonen [11], we shall say that the function $f \notin V$ is *adjoined* to the Čebyšev space V over the set E if the linear span of V and f is also a Čebyšev space over E. For example, if E is an interval and f has a continuous nonvanishing nth derivative on E, then n applications of Rolle's Theorem show that f is adjoined to the space of polynomials of degree $< n$. In particular, x^n is adjoined to this space.

THEOREM 3.4. *Let X be a nontrivial real interval, μ an atom-free, σ-finite measure on X which assigns positive mass to each nonempty open set. Let V be an n-dimensional Čebyšev subspace of $L^1_R(X, \mu)$ over the interior of X consisting of continuous functions, and let $g \in L^1_R(X, \mu)$ be continuous and adjoined to V. Let p_0 be the* (*unique*) *best approximation to g out of V. Then $g - p_0$ changes sign at exactly n distinct points, $x_1, \cdots, x_n$, interior to X. Suppose $f \in L^1_R(X, \mu)$*

is continuous and p^* is its (unique) best approximation out of V. Then if $f - p^*$ changes sign at most n times in X and vanishes only on a set of measure zero, $f - p^*$ changes sign exactly at the points $x_1, \cdots, x_n$. (This is true in particular if f is adjoined to V.)

Proof. $g - p_0$ vanishes at most n tines in the interior of X because g is adjoined to V, and changes sign at least n times by Theorem 2.9. The uniqueness is guaranteed by Theorem 2.8.

By the same reasoning, $f - p^*$ must change sign at exactly n points, $y_1, \cdots, y_n$. Lemma 2.7 assures us of the existence of a function $q + Bg$, $q \in V$, which changes sign just at $y_1, \cdots, y_n$. B cannot be zero, because V is a Čebyšev space of dimension n. Thus, there is a $p' \in V$ such that $g - p'$ changes sign just at $y_1, \cdots, y_n$. As a result, $\operatorname{sgn}(g - p') = \operatorname{sgn}(f - p^*)$ or $\operatorname{sgn}(g - p') = -\operatorname{sgn}(f - p^*)$ a.e. on X.

By Corollary 1.5, $(p, \operatorname{sgn}(f - p^*)) = \pm(p, \operatorname{sgn}(g - p')) = 0$ for all $p \in V$; which implies, by the same Corollary, that p' is a best approximation to g out of V. By uniqueness, $p' = p_0$ and $y_k = x_k$, $k = 1, \cdots, n$.

Theorem 3.4 is essentially proved in Achieser [1], although it is not stated explicitly. Laasonen [11] states and proves essentially the same theorem we have proved. Kreĭn [9] also gives Theorem 3.4 and some generalizations of it to situations we are not considering here.

We see from Theorem 3.4 that if f satisfies the conditions of that theorem, its best approximation with respect to Lebesgue measure by polynomials of degree $< n$ is the polynomial which interpolates it at certain canonical points. By computing the best approximation to x^n, these points can be determined: they are the points $-\cos(k\pi)/(n+1)$, $k = 1, \cdots, n$ (cf. Achieser [1], Kreĭn [9]). There are similar results for trigonometric polynomials. On the interval $[0, 2\pi[$, any $f(x)$ satisfying the conditions of Theorem 3.4 with respect to the space of trigonometric polynomials of degree $\leqq n$ has as its best approximation the polynomial $a_0 + a_1 \cos x + b_2 \sin x + \cdots + a_n \cos nx + b_n \sin nx$ which interpolates it at the points $j\pi/(n+1)$, $j = 1, \cdots, (2n+1)$.

References

1. N. I. Achieser, *Theory of approximation*, Ungar, New York, 1956.
2. S. Bochner and W. T. Martin, *Several complex variables*, Princeton, 1961.
3. N. Bourbaki, *Intégration*, Chapters I-IV, Hermann, Paris, 1952.
4. N. Dunford and J. T. Schwartz, *Linear operators*, Vol. I, Interscience, New York, 1958.
5. S. Ja. Havinson, *On uniqueness of functions of best approximation in the metric of the space* L^1, Izv. Akad. Nauk SSSR Ser. Mat. **22** (1958), 243–270. (Russian)
6. D. Jackson, *Note on a class of polynomials of approximation*, Trans. Amer. Math. Soc. **22** (1921), 320–326.
7. R. C. James, *Orthogonality and linear functionals in normed linear spaces*, Trans. Amer. Math. Soc. **61** (1947), 265–292.
8. J. L. Kelley, *General topology*, Van Nostrand, New York, 1955.
9. M. G. Kreĭn, *The ideas of P. L. Čebyšev and A. A. Markov in the theory of limiting values*

of integrals and their further development, Uspehi Mat. Nauk (N.S.) **6** (1951), 3–120; Amer. Math. Soc. Transl. (2) **12** (1959), 1–121.

10. M. G. Kreĭn and N. I. Achieser, *On some problems in the theory of moments*, Kharkov, 1938; Transl. Math. Monographs, Vol. 2, Amer. Math. Soc., Providence, R.I., 1962.

11. P. Laasonen, *Einige Sätze über Tschebyscheffsche Funktionensysteme*, Ann. Acad. Sci. Fenn. Ser. A. I. **52** (1949), 3–24.

12. T. S. Motzkin and J. L. Walsh, *Least pth power polynomials on a real finite point set*, Trans. Amer. Math. Soc. **78** (1955), 67–81.

13. R. R. Phelps, *Uniqueness of Hahn-Banach extensions and unique best approximation*, Trans. Amer. Math. Soc. **95** (1960), 238–255.

14. V. Ptak, *On approximation of continuous functions in the metric* $\int_a^b |x(t)|\, dt$, Czechoslovak Math. J. **8** (83) (1958), 267–273.

15. I. Singer, *Caractérisation des elements de meilleure approximation dans un espace de Banach quelconque*, Acta. Sci. Math. Szeged **17** (1956), 181–189.

16. ———, *La meilleure approximation interpolative dans les espace de Banach*, Rev. Math. Pures Appl. **4** (1959), 95–113.

17. J. L. Walsh and T. S. Motzkin, *Polynomials of best approximation on an interval*, Proc. Nat. Acad. Sci. U.S.A. **45** (1959), 1523–1528.

University of Texas,
Austin, Texas
International Business Machines Corporation,
Yorktown Heights, New York