JOURNAL OF FUNCTIONAL ANALYSIS 104, 121–148 (1992)

The Multi-dimensional Von Neumann Alternating Direction Search Algorithm in $C(B)$ and L_1

ALLAN PINKUS*

Department of Mathematics, Technion, Haifa, Israel

AND

DANIEL WULBERT

Department of Mathematics, University of California at San Diego, La Jolla, California 92093

Communicated by D. Sarason

Received July 18, 1989; revised April 2, 1991

Let $u_1, \ldots, u_m$ span a subspace U of a normed linear space X. Let $f \in X$. An algorithm to find a best approximation to f from U can be constructed by cyclically searching the one-dimensional spaces spanned by each u_i. The best approximation from the one-dimensional spaces is subtracted from f before searching in the next direction. This accumulating sum of best one-dimensional approximations is an algorithm for finding a best approximation from U. Variations of the method have been called Von Neumann alternating search, Diliberto–Strauss, and median polish algorithms. We first present the effects on the algorithm of smoothness and strict convexity of X. We then give detailed consideration to the algorithm in the two spaces $C(B)$ and L_1. The main results are characterizations for convergence to a best approximation.

1. INTRODUCTION

In this paper we investigate in detail the following algorithm for finite-dimensional subspaces. Let X be a normed linear space (over $\mathbb{R}$). Let U be an n-dimensional subspace of X, and $u_1, \ldots, u_m$ be any m nonzero elements of U which span U. The case where $m = n$ is of particular interest. For

* Supported in part by the Technion V.P.R. Loewengart Research Fund.

0022-1236/92 $3.00

$f\,(=f^{(0)}) \in X$, and $k+1 = rm+i$, $r \geqslant 0$, $i \in \{1, ..., m\}$, set $f^{(k+1)} = f^{(k)} - t_k u_i$ where the $t_k \in \mathbb{R}$ is chosen to minimize

$$\|f^{(k)} - tu_i\|$$

over $t \in \mathbb{R}$. In other words, at each stage we find a best approximant from the 1-dimensional subspace span$\{u_i\}$, and we do this cyclically. Thus

$$f^{(k)} = f - u^{(k)},$$

where $u^{(k)} \in U$.

In the analysis of such an algorithm, three main questions arise. Do the $\{u^{(k)}\}$ converge? If so, do they converge to a best approximant to f from U? At what rate do they converge? This paper is mainly concerned with the first two of these questions.

The above algorithm is not new. It is essentially a cyclic coordinate algorithm (see Zangwill [11]). For $m=2$, it is a special case of the Von Neumann alternating algorithm, or Diliberto–Straus algorithm (see Light and Cheney [4], Deutsch [2], and references therein). If X is a smooth strictly convex normed linear space (e.g., L^p, $1<p<\infty$), then it was shown by Sullivan and Atlestam [8] that the above $u^{(k)}$ necessarily converge to the unique best approximant to f from U. In Sections 2 and 3, we show exactly what role the smoothness and the strict convexity of the norm play in the above algorithm. For example, if X is not strictly convex it may be that no cluster point u^* of the $\{u^{(k)}\}$ satisfies

$$\|f-u^*\| = \min_t \|f-u^*-tu_i\|$$

for all $i=1, ..., m$; while if X is strictly convex, but not smooth, then

$$\|f-u^*\| = \min_t \|f-u^*-tu_i\|$$

for each $i=1, ..., m$, for any cluster point u^* of the $\{u^{(k)}\}$, but u^* need *not* be a best approximant to f from U.

Nothing seems to be known about convergence rates except in the Hilbert space setting, where Smith, Solomon and Wagner [7] proved geometric convergence of the $u^{(k)}$ to the unique best approximant to f from U. We conjecture that such a result should hold in any smooth uniformly convex normed linear space.

For approximating continuous functions in the L^1-norm, this algorithm was suggested by Usow [10]. However, his claim that $\|f-u^{(k)}\|_1$ converged to the error in approximating f from U was incorrect, as was pointed out by Marti [5]. Marti's suggestion for overcoming this problem involved a search in an unbounded number of directions. In fact, if the function f is such that for every $u \in U$, the zero set of $f-u$ has measure zero, then every cluster point of $\{u^{(k)}\}$ is necessarily a best approximant to f from U; see Pinkus [6].

Our investigation into this algorithm was prompted by a paper of Kemperman [3] on "median polish." Median polish, as developed by Tukey [9], is an elementary method for approximating, in the l_1-norm, an $n \times m$ matrix $A=(a_{ij})_{i=1}^{n}{}_{,}{}_{j=1}^{m}$ by a matrix of the form $D=(b_i+c_j)_{i=1}^{n}{}_{,}{}_{j=1}^{m}$. Here the error is given by

$$\|A-D\|_1=\sum_{i=1}^{n}\sum_{j=1}^{m}|a_{ij}-b_i-c_j|.$$

The idea of median polish is to start with A, and determine $\{\beta_i^{(1)}\}_{i=1}^{n}$ so as to minimize

$$\sum_{j=1}^{m}|a_{ij}-\beta_i^{(1)}|$$

for each $i=1, ..., n$. One then determines $\{\gamma_j^{(1)}\}_{j=1}^{m}$ so as to minimize

$$\sum_{i=1}^{n}|a_{ij}-\beta_i^{(1)}-\gamma_j^{(1)}|$$

for each $j=1, ..., m$. Setting $a_{ij}^{(1)}=a_{ij}-\beta_i^{(1)}-\gamma_j^{(1)}$, one then reapplies this same procedure with $a_{ij}^{(1)}$ replacing a_{ij}, to obtain $a_{ij}^{(2)}$, and then cycles through again and again. At the kth stage one obtains

$$a_{ij}^{(k)}=a_{ij}-b_i^{(k)}-c_j^{(k)},$$

where $b_i^{(k)}=\beta_i^{(1)}+\cdots+\beta_i^{(k)}$ and $c_j^{(k)}=\gamma_j^{(1)}+\cdots+\gamma_j^{(k)}$. The idea is that this process will hopefully converge, and that

$$\sum_{i=1}^{n}\sum_{j=1}^{m}|a_{ij}-b_i^{(k)}-c_j^{(k)}|$$

will approach

$$\min_{b_i,\, c_j}\sum_{i=1}^{n}\sum_{j=1}^{m}|a_{ij}-b_i-c_j|.$$

Some variants of this method are also considered. Kemperman [3] does not discuss the problem of the convergence of this algorithm. But he does ask the following: If the $A=(a_{ij})_{i=1}^{n},{}_{j=1}^{m}$ is such that one can take $\beta_i^{(1)}=0$, $i=1,...,n$, and $\gamma_j^{(1)}=0$, $j=1,...,m$, in the above, does it then follow that

$$\sum_{i=1}^{n}\sum_{j=1}^{m}|a_{ij}|=\min_{b_i,c_j}\sum_{i=1}^{n}\sum_{j=1}^{m}|a_{ij}-b_i-c_j|?$$

As Kemperman [3] noted, the answer in general is no.

The above result is not surprising since the l_1-norm is not smooth, and median polish is a special case of the algorithm we first described. To see this, consider the $n\times m$ matrix $A=(a_{ij})_{i=1}^{n},{}_{j=1}^{m}$ as a vector $\mathbf{a}\in\mathbb{R}^{nm}$, whose entries are given by $(a_{ij})_{i=1}^{n},{}_{j=1}^{m}$, arranged in lexicographic order. Then

$$\min_{b_i,c_j}\sum_{i=1}^{n}\sum_{j=1}^{m}|a_{ij}-b_i-c_j|$$

may be reformulated as a best approximation (in l_1^{nm}) to $\mathbf{a}$ from a finite-dimensional (in this case $(n+m-1)$-dimensional) subspace. In fact the above is equivalent to

$$\min_{b_i,c_j}\left\|\mathbf{a}-\sum_{i=1}^{n}b_i\mathbf{v}^i-\sum_{j=1}^{m}c_j\mathbf{w}^j\right\|_1,$$

where the $\mathbf{v}^i,\mathbf{w}^j\in\mathbb{R}^{nm}$. The $\mathbf{v}^i$ have all zero entries, except for the components with indices $\{n(i-1)+l\}_{l=1}^{m}$ which have entries one, and the $\mathbf{w}^j$ have all zero entries, except for the components with indices $\{lm+j\}_{l=0}^{n-1}$, which have entries one. Considered from this point of view, median polish is exactly the algorithm initially described.

Kemperman [3] also shows that this algorithm (median polish) converges to the correct value if we replace the l_1-norm by the l_p-norm, $1<p<\infty$. This is a special case of the result of Sullivan and Atlestam [8]. In fact Bradu [1] proposes an ε-median polish algorithm, which is perturbing the space l_1^{nm} to a smooth, strictly convex space (easily done here since $X=l_1^{nm}$ is finite-dimensional), using the above algorithm, and then perturbing back to the original norm.

2. The Algorithm and Variations

Let X be a normed linear space over $\mathbb{R}$, and U a finite-dimensional subspace of X. Let $u_1,...,u_m$ be any m elements of U. (We will generally assume that the $u_1,...,u_m$ span U, although they need not be linearly independent.)

The main algorithm we consider is the following. Given $f \in X$, we start with some $u^{(0)} \in U$. Given $u^{(k)} \in U$, write $k+1$ in the form $rm+i$, where $r \geqslant 0$ and $i \in \{1, ..., m\}$. Then

$$u^{(k+1)} = u^{(k)} + t_k u_i$$

where t_k is chosen so that

$$\|f - u^{(k)} - t_k u_i\| = \min_t \|f - u^{(k)} - t u_i\|.$$

Such a t_k exists (but is not necessarily unique) since we are approximating $f - u^k$ from the one-dimensional subspace span $\{u_i\}$.

We study this algorithm and two simple variants of it. In this section we address the question of the "convergence" of the algorithm. In the next section we ask to what it converges (if it converges). To understand what we mean by "convergence," we define the following concept.

DEFINITION. For $f \in X$ and $u^* \in U$, we say that u^* is a *stationary point* for f with respect to $\{u_1, ..., u_m\}$ if

$$\|f - u^*\| \leqslant \|f - u^* - t u_j\|$$

for all $t \in \mathbb{R}$ and each $j \in \{1, ..., m\}$. Of course, u^* is stationary for f if and only if 0 is stationary for $f - u^*$.

We consider the questions: Does the sequence $\{u^{(k)}\}$ converge to a stationary point? Is every cluster point of this sequence a stationary point? The answer to the latter question is no in general, but yes if $m = 2$ (the case for $m = 1$ is trivial) or the norm is strictly convex. We start with some simple facts.

LEMMA 2.1. *The sequence $\{u^{(k)}\}$ is a bounded sequence, and*

$$\lim_{k \to \infty} \|f - u^{(k)}\|$$

exists.

Proof. This follows from

$$0 \leqslant \|f - u^{(k+1)}\| \leqslant \|f - u^{(k)}\|. \quad \blacksquare$$

Let

$$\lim_{k \to \infty} \|f - u^{(k)}\| = \sigma.$$

If $\sigma = 0$, i.e., $\lim_{k \to \infty} u^{(k)} = f$, then the algorithm converges (to the best approximant), and there is nothing to prove. In addition, the algorithm may stop after a finite number of steps. If, for example, $u^{(k)} = u^{(k+1)} = \cdots = u^{(k+m)}$ for some $k \geqslant 0$, then it is easily checked that $u^{(k)}$ is a stationary point for f. This does not, a priori, imply that the algorithm will stop, since values of t other than zero may be selected at subsequent stages. However, if the norm is strictly convex, then the algorithm cannot possibly advance past this point.

We will prove that if $m = 2$ or if the norm is strictly convex then every cluster point u^* of $\{u^{(k)}\}$ is a stationary point for f. The idea used in the proof of both cases is much the same.

Let u^* be a cluster point of the sequence $\{u^{(k)}\}$. Thus there exists a subsequence $\{k_p\}$ of $\{k\}$ such that

$$\lim_{p \to \infty} u^{(k_p)} = u^*.$$

Since m is finite, we may and shall assume that for each p

$$k_p = r_p m + i$$

for some $r_p \geqslant 0$ and fixed $i \in \{1, ..., m\}$.

LEMMA 2.2. *Let u^*, i, and k_p be as above. We have*

$$\|f - u^*\| \leqslant \|f - u^* - tu_j\|$$

for all $t \in \mathbb{R}$ and $j \in \{i, i+1\}$ (where, if $i = m$, then we understand that $i + 1 = 1$).

Proof. By definition, $\|f - u^{(k_p)}\| \leqslant \|f - u^{(k_p)} - tu_i\|$. Hence $\|f - u^*\| \leqslant \|f - u^* - tu_i\|$.

Since the sequence $\{u^{(k_p+1)}\}$ is a bounded sequence, on some subsequence, again denoted $\{k_p + 1\}$,

$$\lim_{p \to \infty} u^{(k_p+1)} = \tilde{u}$$

for some $\tilde{u} \in U$. Now,

$$u^{(k_p+1)} = u^{(k_p)} + t_{k_p} u_{i+1}.$$

Thus

$$\lim_{p \to \infty} t_{k_p} = t^*$$

exists, and $\tilde{u} = u^* + t^* u_{i+1}$. From Lemma 2.1,

$$\|f - u^*\| = \|f - \tilde{u}\| = \sigma,$$

and from the previous analysis,

$$\|f - \tilde{u}\| \leqslant \|f - \tilde{u} - tu_{i+1}\|$$

for all $t \in \mathbb{R}$. Therefore

$$\|f - u^*\| = \|f - \tilde{u}\| \leqslant \|f - \tilde{u} - tu_{i+1}\| = \|f - u^* - t^* u_{i+1} - tu_{i+1}\|$$

for all $t \in \mathbb{R}$. That is,

$$\|f - u^*\| \leqslant \|f - u^* - tu_{i+1}\|$$

for all $t \in \mathbb{R}$. ∎

As an immediate consequence of Lemma 2.2, we have:

PROPOSITION 2.3. *For $m = 2$ every cluster point u^* of the sequence $\{u^{(k)}\}$ is a stationary point for f with respect to $\{u_1, u_2\}$.*

In addition,

PROPOSITION 2.4. *If X is a strictly convex normed linear space, and u^* is a cluster point of the sequence $\{u^{(k)}\}$, then u^* is a stationary point for f with respect to $\{u_1, ..., u_m\}$.*

Proof. The proposition is an immediate consequence of Lemma 2.2 if we can prove that for the u^* and $\tilde{u}$ as therein, $u^* = \tilde{u}$.

Recall that $\tilde{u} = u^* + t^* u_{i+1}$, and

$$\|f - u^*\| = \|f - \tilde{u}\| = \|f - u^* - t^* u_{i+1}\| \leqslant \|f - u^* - tu_{i+1}\|$$

for all $t \in \mathbb{R}$. Since X is a strictly convex normed linear space, the best approximation to $f - u^*$ from span $\{u_{i+1}\}$ is *uniquely* attained. Since both 0 and $t^* u_{i+1}$ are best approximants, it follows that $t^* = 0$, i.e., $u^* = \tilde{u}$. ∎

If $m \geqslant 3$ and the norm on X is not strictly convex, then it is *not* true that every cluster point u^* of the $\{u^{(k)}\}$ is a stationary point for f with respect to $\{u_1, ..., u_m\}$. The following is an example verifying this statement.

EXAMPLE. Let $X = l_1^3$, i.e., $\mathbf{f} = (f_1, f_2, f_3) \in \mathbb{R}^3$ with norm $\|\mathbf{f}\|_1 = |f_1| + |f_2| + |f_3|$. Set $\mathbf{u}_1 = (1, 1, 0)$, $\mathbf{u}_2 = (1, 0, 1)$, and $\mathbf{u}_3 = (0, 1, 1)$, i.e., $m = 3$. Let $\mathbf{f} = (-1, 0, 1)$ and $\mathbf{u}^{(0)} = (0, 0, 0)$. It may be checked that we can choose $t_k = -1$ if $k = 0$ or 1 (mod 4) and $t_k = 1$ otherwise. Computation shows that $\mathbf{u}^{(12)}=\mathbf{0}$ (the process is cyclical with period 12), $\|\mathbf{f}-\mathbf{u}^{(k)}\|_1=2$ for all k, and $\mathbf{f}-\mathbf{u}^{(12k+1)}=\mathbf{u}_3$ for all k. That is, $\mathbf{u}^{(12k+1)}=\mathbf{u}^*=(-1,-1,0)$ is a cluster point of the $\{\mathbf{u}^{(k)}\}$, but $\mathbf{u}^*$ is not a stationary point for $\mathbf{f}$ with respect to $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$, since

$$2=\|\mathbf{f}-\mathbf{u}^*\|_1>\|\mathbf{f}-\mathbf{u}^*-\mathbf{u}_3\|_1=0.$$

In fact $\mathbf{u}^{(12k+5)}$ and $\mathbf{u}^{(12k+9)}$ are also cluster points which are not stationary points, since $\mathbf{f}-\mathbf{u}^{(12k+5)}=\mathbf{u}_1$, and $\mathbf{f}-\mathbf{u}^{(12k+9)}=\mathbf{u}_2$.

Remark. This exact same example gives the identical result with $X=l^3_\infty$, i.e., $\mathbf{f}=(f_1, f_2, f_3)\in\mathbb{R}^3$ with norm $\|\mathbf{f}\|_\infty=\max\{|f_i|: i=1,2,3\}$. The only difference is that $\|\mathbf{f}-\mathbf{u}^{(k)}\|_\infty=1$ for all k.

A simple modification of the algorithm will allow us to prove that every cluster point is a stationary point. This modified algorithm runs as follows.

We are given $f\in X$ and $u_1, ..., u_m$ as previously. Given $u^{(k)}$, we define $u^{(k+1)}$ as follows. Let

$$\|f-u^{(k)}-t_k u_{j_k}\|=\min_{i=1,...,m}\min_t \|f-u^{(k)}-tu_i\|.$$

Set $u^{(k+1)}=u^{(k)}+t_k u_{j_k}$. That is, at each step we look along each of the m directions $\{u_i\}$, and we choose the direction which maximally minimizes the error.

PROPOSITION 2.5. *In the above modified algorithm, if u^* is a cluster point of the sequence $\{u^{(k)}\}$, then u^* is a stationary point for f with respect to $\{u_1, ..., u_m\}$.*

Proof. Let $\lim_{p\to\infty} u^{(k_p)}=u^*$. Since $\|f-u^k\|\leqslant\|f-u^l\|$ for all $l\leqslant k$ it follows that $\|f-u^{(k_p)}\|\leqslant\|f-u^{(k_{p-1}+1)}\|\leqslant\|f-u^{(k_{p-1})}-tu_i\|$ for all $t\in\mathbb{R}$ and all i. Hence $\|f-u^*\|\leqslant\|f-u^*-tu_i\|$. ∎

Remark. Another variant (generalization) of the main algorithm is the following. Let $U_1, ..., U_m$ be subspaces of U. As previously, given $u^{(k)}$ with $k+1=rm+i$, $r\geqslant 0$, $i\in\{1, ..., m\}$, we let

$$\|f-u^{(k)}-u_i\|=\min_{u\in U_i}\|f-u^{(k)}-u\|,$$

where $u_i\in U_i$. Such a u_i of course exists since each U_i is finite-dimensional. Set $u^{(k+1)}=u^{(k)}+u_i$. This is a generalization of the main algorithm. It is easily checked that the results of Propositions 2.3 and 2.4 hold in this setting. That is, if u^* is a cluster point of the $\{u^{(k)}\}$, then u^* is a "stationary point" provided that either $m=2$ or that the norm is strictly convex.

3. Stationary Points

The question we address in this section is the following. If u^* is a stationary point for f with respect to $\{u_1, ..., u_m\}$, is u^* a best approximant to f from $\text{span}\{u_1, ..., u_m\}$? We show that in order that the response be positive for *all* f and $\{u_1, ..., u_m\}$, it is both necessary and sufficient that X be a smooth normed linear space. However, it should be noted that given $\{u_1, ..., u_m\}$ in a nonsmooth X, there may exist $\{v_1, ..., v_r\}$ satisfying $\text{span}\{u_1, ..., u_m\} = \text{span}\{v_1, ..., v_r\}$ for which the answer is yes if we replace the $\{u_1, ..., u_m\}$ by $\{v_1, ..., v_r\}$.

For ease of exposition, we say that u^* is a *phantom approximation* to f from $\{u_1, ..., u_m\}$ if u^* is a stationary point for f with respect to $\{u_1, ..., u_m\}$, but not a best approximant to f from $\text{span}\{u_1, ..., u_m\}$.

Definition. The normed linear space X is *smooth* if to each $f \in X$ of norm one, there exists a unique l in the unit ball of X^* (the continuous dual of X) for which $l(f) = 1$.

Theorem 3.1. *If X is a smooth normed linear space, $f \in X$, and $u_1, ..., u_m \in X$, then there does not exist a phantom approximation to f from $\{u_1, ..., u_m\}$.*

Proof. Assume $f - u^* \neq 0$. Otherwise there is nothing to prove. Let u^* be a stationary point for f with respect to $\{u_1, ..., u_m\}$. Let l be the unique norm one linear functional in X^* satisfying $l(f - u^*) = \|f - u^*\|$. Now, for each $i \in \{1, ..., m\}$, the zero function is a best approximant to $f - u^*$ from $\text{span}\{u_i\}$. From the Hahn–Banach theorem, there exists an $l_i \in X^*$ satisfying

(1) $\|l_i\|_{X^*} = 1$.

(2) $l_i(u_i) = 0$.

(3) $l_i(f - u^*) = \|f - u^*\|$.

From (1) and (3), $l_i = l$ for each $i = 1, ..., m$. Thus we have an $l \in X^*$ satisfying

(1′) $\|l\|_{X^*} = 1$.

(2′) $l(u_i) = 0$, $i = 1, ..., m$.

(3′) $l(f - u^*) = \|f - u^*\|$.

Condition (2′) implies that $l(u) = 0$ for every $u \in \text{span}\{u_1, ..., u_m\}$, and therefore u^* is a best approximant to f from $\text{span}\{u_1, ..., u_m\}$. ∎

The condition of smoothness is necessary for the above result to hold for all f and $\{u_1, ..., u_m\}$.

PROPOSITION 3.2. *Assume X is a normed linear space which is not smooth. Then there exist $\{u_1, \ldots, u_m\}$ and f in X for which the zero function is a phantom approximation to f from $\{u_1, \ldots, u_m\}$.*

Proof. Since X is not smooth, there exist an $f \in X$ of norm one, and two distinct linear functionals l_1 and l_2 in the unit ball of X^* satisfying $l_1(f) = l_2(f) = 1$. Because $l_1 \neq l_2$ (implying $\dim X > 1$) there exists a $g \in X$ for which $l_1(g) \neq l_2(g)$. Set $W = \operatorname{span}\{f, g\}$. Since $\dim W = 2$, we can find $u_1, u_2 \in W$ such that

$$l_i(u_j) = \delta_{ij}, \qquad i, j = 1, 2.$$

Let $u_3, \ldots, u_m$ be arbitrary functions in X such that either $l_1(u_j) = 0$ or $l_2(u_j) = 0$, $j = 3, \ldots, m$.

For each $j \in \{1, \ldots, m\}$ there exists an l_i, $i \in \{1, 2\}$ satisfying

(1) $\|l_i\|_{X^*} = 1$.

(2) $l_i(u_j) = 0$.

(3) $l_i(f) = \|f\|$.

Thus the zero function is a stationary point for f with respect to $\{u_1, \ldots, u_m\}$. The zero function is *not* a best approximant to f from $\operatorname{span}\{u_1, \ldots, u_m\}$ since $f \in \operatorname{span}\{u_1, \ldots, u_m\}$. ∎

In the next two sections we investigate in more detail two nonsmooth normed linear spaces: namely, continuous functions with the uniform norm, and the space L^1.

4. THE SPACE $C(B)$

Let B be a compact Hausdorff set, and $C(B)$ the space of continuous real-valued functions defined on B with norm

$$\|f\|_\infty = \max\{|f(x)| : x \in B\}.$$

In what follows, for $f \in C(B)$ we let

$$\begin{aligned} \operatorname{crit} f &= \{x : |f(x)| = \|f\|\} \\ \operatorname{supp} f &= \{x : f(x) \neq 0\} \\ Z(f) &= \{x : f(x) = 0\} (= B \backslash \operatorname{supp} f), \end{aligned}$$

and if U is a subspace of $C(B)$, then

$$\operatorname{supp} U = \bigcup_{u \in U} \operatorname{supp} u.$$

We show that $C(B)$ is a particularly unsuitable space for the algorithm under consideration.

PROPOSITION 4.1. *Let U be an n-dimensional subspace of $C(B)$ $(n>1)$. Assume that we are given $u_1, ..., u_m$ in U with $m<2^{n-1}$. Then there exists an $f \in C(B)$ for which the zero function is a phantom approximation to f from $\{u_1, ..., u_m\}$.*

Proof. Let $\{x_1, ..., x_n\} \subseteq B$ satisfy $\dim U|_{\{x_1, ..., x_n\}} = n$. With no loss of generality assume that $u_j(x_1) \geqslant 0$ for $j=1, ..., m$. Since $m<2^{n-1}$ there exists a choice of $\varepsilon_i \in \{-1, 1\}$, $i=2, ..., n$, $\varepsilon_1 = 1$, such that for *no* $j \in \{1, ..., m\}$ do we have

$$\varepsilon_i u_j(x_i) > 0, \qquad i=1, ..., n.$$

Let $f \in C(B)$ satisfy $f(x_i) = \varepsilon_i$, $i=1, ..., n$, and $\operatorname{crit} f = \{x_1, ..., x_n\}$. Such an f exists. Since

$$\min_{i=1,...,n} \sigma u_j(x_i) f(x_i) \leqslant 0$$

for each $j=1, ..., m$, and $\sigma \in \{-1, 1\}$, we have that the zero function is a stationary point for f with respect to $\{u_1, ..., u_m\}$. But $\dim U|_{\{x_1, ..., x_n\}} = n$, and there therefore exists a $u \in U$ for which

$$u(x_i) f(x_i) > 0, \qquad i=1, ..., n.$$

Thus the zero function is not a best approximant to f from U. ∎

For $m=2^{n-1}$ we can exactly delineate those functions $u_1, ..., u_m$ for which the results of the previous proposition hold.

PROPOSITION 4.2. *Let U be an n-dimensional subspace of $C(B)$ $(n>1)$, and $m=2^{n-1}$. Let $u_1, ..., u_m \in U$. No $f \in C(B)$ has a phantom approximation from $\{u_1, ..., u_m\}$ if and only if the following hold.*

(a) *There exists a basis $v_1, ..., v_n$ for U with*

$$\operatorname{supp} v_k \cap \operatorname{supp} v_l = \varnothing$$

for all $k \neq l$.

(b) *Let $(\varepsilon_1^j, ..., \varepsilon_n^j)$, $j=1, ..., m(=2^{n-1})$ denote all possible distinct vectors with $\varepsilon_i^j \in \{-1, 1\}$, $i=2, ..., n$, and $\varepsilon_1^j = 1$. Then*

$$u_j = \delta_j \sum_{i=1}^{n} \varepsilon_i^j a_i^j v_i, \qquad j=1, ..., m,$$

for some $\delta_j \in \{-1, 1\}$, *and*

$$a_i^j > 0, \qquad i = 1, ..., n, \; j = 1, ..., m.$$

Proof. ($\Leftarrow$). Assume the $\{u_1, ..., u_m\}$ satisfy (a) and (b). It suffices to prove that the zero function is not a phantom approximation to any $f \in C(B)$ from $\{u_1, ..., u_m\}$. As such let us assume the existence of an $f \in C(B)$ for which the zero function is not a best approximant to f from U. Then there exists a $u = \sum_{i=1}^{n} b_i v_i$ satisfying

$$u(y) f(y) > 0 \qquad \text{for all} \quad y \in \operatorname{crit} f.$$

By assumption (b), there exist a $\delta \in \{-1, 1\}$ and a $j \in \{1, ..., m\}$ such that

$$\delta u_j = \sum_{i=1}^{n} a_i v_i$$

satisfies $a_i \neq 0$, $i = 1, ..., n$, and $a_i b_i \geqslant 0$, $i = 1, ..., n$. From (a), $(\delta u_j)(x) \neq 0$ for all $x \in \operatorname{supp} U$, and

$$(\delta u_j)(x)\, u(x) \geqslant 0$$

for all $x \in \operatorname{supp} U$. Thus

$$(\delta u_j)(y) f(y) > 0 \qquad \text{for all} \quad y \in \operatorname{crit} f.$$

This implies that the zero function is not a stationary point for f with respect to $\{u_1, ..., u_m\}$.

($\Rightarrow$). We now assume that every stationary point is a best approximant. Let $\{x_1, ..., x_n\} \subseteq B$ satisfy $\dim U|_{\{x_1, ..., x_n\}} = n$. It follows from the proof of the previous proposition that if $u_j(x_1) \geqslant 0$, $j = 1, ..., m$, then for each choice of $(\varepsilon_1, ..., \varepsilon_n)$ with $\varepsilon_i \in \{-1, 1\}$, $i = 2, ..., n$, $\varepsilon_1 = 1$, there must exist a $j \in \{1, ..., m\}$ such that

$$\varepsilon_i u_j(x_i) > 0, \qquad i = 1, ..., n.$$

Since $m = 2^{n-1}$, to each such choice of $(\varepsilon_1, \varepsilon_2, ..., \varepsilon_n)$ there exists a unique $j \in \{1, ..., m\}$ satisfying the above.

Assume $y \notin \{x_1, ..., x_n\}$. If for some $k, l \in \{1, ..., n\}$, $k \neq l$,

$$\dim U|_{\{x_1, ..., x_n, y\} \setminus \{x_k\}} = \dim U|_{\{x_1, ..., x_n, y\} \setminus \{x_l\}} = n,$$

then a contradiction ensures. This follows from the result of the previous paragraph and the fact that the vector $(u_1(y), ..., u_m(y))$ cannot have the same (strict) sign pattern as both $(u_1(x_k), ..., u_m(x_k))$ and $(u_1(x_l), ..., u_m(x_l))$.

Thus for each $y \in \operatorname{supp} U$ there exists one and only one $k \in \{1, ..., n\}$ for which

$$n = \dim U|_{\{x_1, ..., x_n, y\} \setminus \{x_k\}}.$$

Let $v_1, ..., v_n \in U$ satisfy $v_j(x_i) = \delta_{ij}$, $i, j = 1, ..., n$. Such v_j exist since $\dim U|_{\{x_1, ..., x_n\}} = n$. The $v_1, ..., v_n$ are a basis for U. From the above, it follows that if $y \in \operatorname{supp} U$, there exists a unique $k \in \{1, ..., n\}$ for which $v_k(y) \neq 0$. Thus for $k, l \in \{1, ..., n\}$, $k \neq l$,

$$\operatorname{supp} v_k \cap \operatorname{supp} v_l = \varnothing.$$

Since on the set $\{x_1, ..., x_n\}$, the $\{\pm u_j\}_{j=1}^m$ take on all (strict) sign patterns, it follows that the $\{u_j\}_{j=1}^m$ must satisfy (b). ∎

Remark. If $v_1, ..., v_n$ is a basis for U, and for all $k, l \in \{1, ..., n\}$, $k \neq l$,

$$\operatorname{supp} v_k \cap \operatorname{supp} v_l = \varnothing,$$

then it is easy to find a best approximant to any $f \in C(B)$. Let a_i attain the minimum in

$$\min_a \max_{x \in \operatorname{supp} v_i} |(f - av_i)(x)|.$$

Then $u = \sum_{i=1}^n a_i v_i$ is a best approximant to f from U.

If $m > 2^{n-1}$, then the situation becomes more complicated. If $C(B) = l_\infty^d$, where

$$l_\infty^d = \{\mathbf{x} : \mathbf{x} \in \mathbb{R}^d, \|\mathbf{x}\|_\infty = \max_{i=1,...,d} |x_i|\},$$

then for any finite-dimensional subspace U of l_∞^d, it is possible to find $\mathbf{u}^1, ..., \mathbf{u}^m \in U$ (m finite) such that there is no phantom approximation to any $\mathbf{b} \in l_\infty^d$ from $\{\mathbf{u}^1, ..., \mathbf{u}^m\}$. We must simply include in the set $\{\pm \mathbf{u}^j\}_{j=1}^m$ all possible (strict) sign patterns of every $\mathbf{u} \in U$ on any subset of the indices $\{1, 2, ..., d\}$. This is what was done in Proposition 4.2. It is not a recommended procedure. Let us go to the other extreme and consider one other case which further emphasizes the unsuitability of this algorithm in the uniform norm.

Let μ be any positive non-atomic Borel measure defined on a compact set B.

PROPOSITION 4.3. *Let U be a finite-dimensional subspace of $C(B)$. Assume there exists a set $A \subseteq B$ such that $\mu(A) > 0$ and $\dim U|_A \geq 2$. Further assume that if $u \in U$ and $u|_A \neq 0$, then $\mu(Z(u)|_A) = 0$. Then for any $u_1, ..., u_m$ which span U, there exists an $f \in C(B)$ for which the zero function is a phantom approximation to f from $\{u_1, ..., u_m\}$.*

Proof. For each $j \in \{1, ..., m\}$, set

$$A_j^1 = \{x : u_j(x) \geqslant 0\} \cap A$$
$$A_j^2 = \{x : u_j(x) \leqslant 0\} \cap A.$$

Then there exist $i_1, ..., i_m$, each in $\{1, 2\}$, such that

$$\mu\left(\bigcap_{j=1}^{m} A_j^{i_j}\right) \geqslant \mu(A)/2^m > 0.$$

Let $C = \bigcap_{j=1}^m A_j^{i_j}$. If $\dim U|_C < 2$, then there exists a $u \in U$ for which $u|_A \neq 0$, but $u|_C = 0$. But then $\mu(Z(u)|_A) > 0$, a contradiction. Thus $\dim U|_C \geqslant 2$, and there exist $y_1, y_2 \in C$ for which $\dim U|_{\{y_1, y_2\}} = 2$.

Thus

$$u_j(y_1)\, u_j(y_2) \geqslant 0, \qquad j = 1, ..., m,$$

and there exists a $u \in U$ satisfying $u(y_1)\, u(y_2) < 0$. Let $f \in C(B)$, $f(y_1) = 1$, $f(y_2) = -1$, and $\operatorname{crit} f = \{y_1, y_2\}$. The zero function is a phantom approximation to f from $\{u_1, ..., u_m\}$. ∎

5. The L^1-Norm

Let B be a set, Σ a σ-field of subsets of B, and ν a positive measure defined on Σ, i.e., $\nu(E) \geqslant 0$ for all $E \in \Sigma$. By $L^1(B, \nu)$ (we suppress the Σ for brevity), we mean the space of all real-valued ν-measurable functions f for which $|f|$ is ν-integrable. $L^1(B, \nu)$ is topologized by the pseudonorm

$$\|f\|_1 = \int_B |f(x)|\, d\nu(x).$$

As previously,

$$\operatorname{supp} f = \{x : f(x) \neq 0\}$$

and

$$Z(f) = \{x : f(x) = 0\}.$$

The sets $\operatorname{supp} f$ and $Z(f)$ are ν-measurable. For $f \in L^1(B, \nu)$, we set

$$\operatorname{sgn} f(x) = \begin{cases} 1, & f(x) > 0 \\ 0, & f(x) = 0 \\ -1, & f(x) < 0. \end{cases}$$

We recall that if $f \in L^1(B, \nu)$ and U is a linear subspace of $L^1(B, \nu)$, then the zero function is a best approximant to f from U if and only if

$$\left| \int_B u \operatorname{sgn}(f)\, d\nu \right| \leqslant \int_{Z(f)} |u| \, d\nu$$

for all $u \in U$.

Remark. When working in L_p spaces we interpret all functions, including f and u_i to be equivalence classes of functions. In this section we will only consider the case $m = n$.

In contrast to the previous section, phantom approximations are not nearly as prevalent in this setting (especially for $L^1(B, \nu) = l_1^d$). We start with a simple result.

PROPOSITION 5.1. *Let $U = \operatorname{span}\{u_1, ..., u_n\}$ be an n-dimensional subspace of $L^1(B, \nu)$. Assume that*

$$\operatorname{supp} u_k \cap \operatorname{supp} u_l = \varnothing$$

for all $k \neq l$. Then no $f \in L^1(B, \nu)$ has a phantom approximation from $\{u_1, ..., u_n\}$.

Proof. Let u^* be a stationary point for f with respect to $\{u_1, ..., u_n\}$. Thus the zero function is a best approximinant to $f - u^*$ from each $\operatorname{span}\{u_i\}$, $i = 1, ..., n$. Therefore

$$\left| \int_B u_i \operatorname{sgn}(f - u^*)\, d\nu \right| \leqslant \int_{Z(f - u^*)} |u_i| \, d\nu,$$

$i = 1, ..., n$. Since the $\{u_i\}_{i=1}^n$ have distinct support, for any $a_i \in \mathbb{R}$, $i = 1, ..., n$,

$$\begin{aligned} \left| \int_B \sum_{i=1}^n a_i u_i \operatorname{sgn}(f - u^*)\, d\nu \right| &\leqslant \sum_{i=1}^n |a_i| \left| \int_B u_i \operatorname{sgn}(f - u^*)\, d\nu \right| \\ &\leqslant \sum_{i=1}^n |a_i| \int_{Z(f - u^*)} |u_i| \, d\nu \\ &= \int_{Z(f - u^*)} \sum_{i=1}^n |a_i u_i| \, d\nu \\ &= \int_{Z(f - u^*)} \left| \sum_{i=1}^n a_i u_i \right| d\nu. \end{aligned}$$

Thus u^* is a best approximant to f from U. ∎

Remark. In the above case the algorithm converges to a best approximant after one cycle through the $u_1, \ldots, u_n$.

There is an essential difference in the problem of best approximation in $L^1(B, \nu)$ depending upon whether or not the measure ν has atoms.

THEOREM 5.2. *Assume that ν is a nonatomic, σ-finite, positive measure. Let $U = \operatorname{span}\{u_1, \ldots, u_n\}$ be an n-dimensional subspace of $L^1(B, \nu)$. Assume that*

$$\nu(\operatorname{supp} u_k \cap \operatorname{supp} u_l) > 0$$

for some $k \neq l$. Then there exists an $f \in L^1(B, \nu)$ for which the zero function is a phantom approximation to f from $\{u_1, \ldots, u_n\}$.

Proof. For ease of notation, assume that $\operatorname{supp} u_1 \cap \operatorname{supp} u_2 = R$, and $\nu(R) > 0$.

From Liapunov's Theorem (since ν is non-atomic and σ-finite), if $G \subseteq B$, then the set

$$\mathscr{A}_G = \left\{ \left(\int_G hu_1 \, d\nu, \ldots, \int_G hu_n \, d\nu \right) : h \in L^\infty(G, \nu), |h(x)| = 1, \text{ all } x \in G \right\}$$

is a compact, convex subset of $\mathbb{R}^n$, symmetric about the origin, and equals

$$\left\{ \left(\int_G hu_1 \, d\nu, \ldots, \int_G hu_n \, d\nu \right) : h \in L^\infty(G, \nu), \|h\|_\infty \leqslant 1 \right\}.$$

Furthermore, if the $u_1, \ldots, u_n$ are linearly independent over G, then $\mathbf{0} \in \operatorname{int} \mathscr{A}_G$. For if $\mathbf{0} \in \partial \mathscr{A}_G$ (obviously $\mathbf{0} \in \mathscr{A}_G$), there exists a $\mathbf{b} = (b_1, \ldots, b_n) \neq \mathbf{0}$ for which

$$\sum_{i=1}^n b_i \int_G hu_i \, d\nu \geqslant 0$$

for all $h \in L^\infty(G, \nu)$ satisfying $|h(x)| = 1$ for all $x \in G$. Setting $u = \sum_{i=1}^n b_i u_i$ and $h = -\operatorname{sgn} u$ on $\operatorname{supp} u \cap G$, a contradiction ensues since $u \not\equiv 0$.

Since $\mathbf{0} \in \operatorname{int} \mathscr{A}_B$, there exists an $\varepsilon > 0$ such that

$$D_\varepsilon = \{ \mathbf{d} : \mathbf{d} \in \mathbb{R}^n, \|\mathbf{d}\|_\infty \leqslant \varepsilon \} \subseteq \mathscr{A}_B.$$

Set

$$P_i^j = \{ x : (-1)^j u_i(x) > 0 \} \cap R, \qquad i, j = 1, 2.$$

Since $P_i^1 \cup P_i^2 = R$ for $i = 1, 2$, at least one of the four sets $P_1^{j_1} \cap P_2^{j_2}$ has positive ν-measure for $j_1, j_2 \in \{1, 2\}$. Let $\tilde{C}$ denote such a set.

From the absolute continuity of the integral, given $\varepsilon > 0$ there exists a $\delta > 0$ such that if $Q \subseteq B$ and $\nu(Q) < \delta$, then

$$\int_Q |u_i| \, d\nu < \varepsilon/3, \qquad i = 1, ..., n.$$

Choose $C \subset \tilde{C}$ such that $0 < \nu(C) < \delta$. Thus $\mathscr{A}_C \subset D_{\varepsilon/3}$. Since

$$\mathscr{A}_{B \setminus C} + \mathscr{A}_C = \mathscr{A}_B$$

and $D_\varepsilon \subseteq \mathscr{A}_B$, we have that $D_{\varepsilon/3} \subset \mathscr{A}_{B\setminus C}$. Thus $\mathscr{A}_C \subset \mathscr{A}_{B \setminus C}$. There therefore exists an $h \in L^\infty(B \setminus C, \nu)$ with $|h(x)| = 1$ for all $x \in B \setminus C$ such that

$$\int_{B \setminus C} hu_1 \, d\nu = \int_C -u_1 \, d\nu$$

and

$$\int_{B \setminus C} hu_i \, d\nu = \int_C u_i \, d\nu \qquad i = 2, ..., n.$$

From the above, we obtain

$$\left| \int_{B \setminus C} hu_i \, d\nu \right| \leqslant \int_C |u_i| \, d\nu, \qquad i = 1, ..., n. \tag{1}$$

Choose $\alpha, \beta \in \mathbb{R}$ satisfying $\alpha u_1(x) < 0$ and $\beta u_2(x) > 0$ for all $x \in C$. Since $C \subseteq \tilde{C}$, such a choice is possible. Thus

$$\begin{aligned} \int_{B \setminus C} h(\alpha u_1 + \beta u_2) \, d\nu &= \int_C -\alpha u_1 + \beta u_2 \, d\nu \\ &= \int_C |\alpha u_1| + |\beta u_2| \, d\nu \\ &\geqslant \int_C |\alpha u_1 + \beta u_2| \, d\nu. \end{aligned}$$

For equality to hold in the above inequality, it is necessary that

$$|\alpha u_1 + \beta u_2| = |\alpha u_1| + |\beta u_2| \qquad \nu \text{ a.e. on } C.$$

But $(\alpha u_1)(\beta u_2) < 0$ on C. Thus, setting $\tilde{u} = \alpha u_1 + \beta u_2$, we have

$$\int_{B \setminus C} h\tilde{u} \, d\nu > \int_C |\tilde{u}| \, d\nu. \tag{2}$$

Finally, let $f \in L^1(B, \nu)$ satisfy

$$\operatorname{sgn} f(x) = \begin{cases} h(x), & x \in B \setminus C \\ 0, & x \in C. \end{cases}$$

Then from (1) and (2)

$$\left| \int_B u_i \operatorname{sgn} f \, d\nu \right| \leqslant \int_{Z(f)} |u_i| \, d\nu, \qquad i = 1, ..., n$$

$$\int_B \tilde{u} \operatorname{sgn} f \, d\nu > \int_{Z(f)} |\tilde{u}| \, d\nu.$$

The first set of inequalities imply that the zero function is a stationary point for f with respect to $\{u_1, ..., u_n\}$. The second inequality says that the zero function is not a best approximant to f from U. ∎

Remark. For the above-constructed f, $\nu(Z(f)) > 0$. If u^* is a stationary point for f with respect to $\{u_1, ..., u_n\}$, and $\nu(Z(f - u^*)) = 0$, then a simple consequence of the characterization theorem of best approximants is that u^* is a best approximant to f from $\operatorname{span}\{u_1, ..., u_n\}$. If we further assume that f and the $\{u_i\}_{i=1}^n$ are continuous (see the next paragraph) and $\nu(Z(f-u)) = 0$ for all $u \in \operatorname{span}\{u_1, ..., u_n\}$, then it was proved, in Pinkus [6], that every cluster point of the algorithm is a stationary point and hence a best approximant.

We wish to extend the above Theorem 5.2 to continuous functions. To avoid complications, we let $B = K$, where K is a compact subset of $\mathbb{R}^d$ with $K = \overline{\operatorname{int} K}$. We also set $\nu = \mu$, where μ is a nonatomic, positive, finite, regular, Borel measure on K whose support is K. Thus $\|\cdot\|_1$ is a true norm on $C(K)$ and we do not need to consider equivalence classes of functions. By $C_1(K, \mu)$ we mean the space $C(K)$ endowed with the norm $\|\cdot\|_1$. $C_1(K, \mu)$ is not a Banach space. It is not complete.

THEOREM 5.3. *Let the above assumptions hold. Assume that* $U = \operatorname{span}\{u_1, ..., u_n\}$ *is an n-dimensional subspace of* $C(K)$ *and* $u_k(x)\, u_l(x) \neq 0$ *for some* $x \in K$ *and* $k \neq l$. *There then exists an* $f \in C(K)$ *such that in* $C_1(K, \mu)$ *the zero function is a phantom approximation to f from* $\{u_1, ..., u_n\}$.

Proof. By assumption $\mu(\operatorname{supp} u_k \cap \operatorname{supp} u_l) > 0$ (since $U \subset C(K)$). Thus from the previous theorem there exist a $g \in L^1(K, \mu)$ and $\tilde{u} \in U$ for which

$$\left| \int_K u_i \operatorname{sgn} g \, d\mu \right| \leqslant \int_{Z(g)} |u_i| \, d\mu, \qquad i = 1, ..., n \tag{3}$$

$$\int_K \tilde{u} \operatorname{sgn} g \, d\mu > \int_{Z(g)} |\tilde{u}| \, d\mu. \tag{4}$$

The proof of this theorem is divided into two steps. The first step proves that we can perturb g to $\tilde{g} \in L^1(K, \mu)$ so that strict inequality holds in (3) for all $i \in \{1, ..., n\}$, and in (4). The second step shows that we can replace $\tilde{g}$ by $f \in C(K)$. This implies the statement of the theorem. Since the proof of each step is lengthy and technical, we separate them.

LEMMA 5.4. *Assume* (3) *and* (4) *hold. Then there exists a* $\tilde{g} \in L^1(K, \mu)$ *satisfying* (3) *and* (4) *with strict inequality in* (3) *for each* $i = 1, ..., n$.

Proof. Assume (multiplying the u_i by -1 if necessary) that

$$\left| \int_K u_i \operatorname{sgn} g \, d\mu \right| < \int_{Z(g)} |u_i| \, d\mu, \qquad i = 1, ..., r-1 \tag{5}$$

$$\int_K u_i \operatorname{sgn} g \, d\mu = \int_{Z(g)} |u_i| \, d\mu, \qquad i = r, ..., n, \tag{6}$$

where $1 \leqslant r \leqslant n$.

Let $\varepsilon = \int_K \tilde{u} \operatorname{sgn} g \, d\mu - \int_{Z(g)} |\tilde{u}| \, d\mu > 0$. From the absolute continuity of the integral, there exists a $\delta > 0$ such that if $A \subset K$ and $\mu(A) < \delta$, then

$$2 \int_A |\tilde{u}| \, d\mu < \varepsilon.$$

Now, if $\int_K u_r \operatorname{sgn} g \, d\mu = \int_{Z(g)} |u_r| \, d\mu = 0$ then $\operatorname{supp} u_r \subseteq \operatorname{supp} g$, and there exists a set $A \subseteq \operatorname{supp} u_r$ such that $\mu(A) < \delta, 0 < \int_A |u_r| \, d\mu$, and $\int_A u_r \operatorname{sgn} g \, d\mu = 0$.

If $\int_K u_r \operatorname{sgn} g \, d\mu = \int_{Z(g)} |u_r| \, d\mu > 0$, there exists a set $A \subseteq \operatorname{supp} g$ such that $\mu(A) < \delta$, $\int_A u_r \operatorname{sgn} g \, d\mu = \int_A |u_r| \, d\mu > 0$, and $\int_A |u_r| \, d\mu < \int_K u_r \operatorname{sgn} g \, d\mu$.

In either case we set

$$\tilde{g}(x) = \begin{cases} g(x), & x \notin A \\ 0, & x \in A. \end{cases}$$

In the first case we have

$$\int_K u_r \operatorname{sgn} \tilde{g} \, d\mu = \int_K u_r \operatorname{sgn} g \, d\mu - \int_A u_r \operatorname{sgn} g \, d\mu = 0,$$

while

$$\int_{Z(\tilde{g})} |u_r| \, d\mu = \int_A |u_r| \, d\mu > 0.$$

In the second case we have

$$\int_K u_r \operatorname{sgn} \tilde{g}\, d\mu = \int_K u_r \operatorname{sgn} g\, d\mu - \int_A u_r \operatorname{sgn} g\, d\mu$$

$$= \int_K u_r \operatorname{sgn} g\, d\mu - \int_A |u_r|\, d\mu > 0,$$

while

$$0 < \int_K u_r \operatorname{sgn} \tilde{g}\, d\mu = \int_K u_r \operatorname{sgn} g\, d\mu - \int_A |u_r|\, d\mu$$

$$= \int_{Z(g)} |u_r|\, d\mu - \int_A |u_r|\, d\mu$$

$$< \int_{Z(g) \cup A} |u_r|\, d\mu$$

$$= \int_{Z(\tilde{g})} |u_r|\, d\mu.$$

Thus in all cases,

$$\left| \int_K u_r \operatorname{sgn} \tilde{g}\, d\mu \right| < \int_{Z(\tilde{g})} |u_r|\, d\mu.$$

For $i \in \{1, ..., n\}$, $i \neq r$,

$$\left| \int_K u_i \operatorname{sgn} \tilde{g}\, d\mu \right| = \left| \int_K u_i \operatorname{sgn} g\, d\mu - \int_A u_i \operatorname{sgn} g\, d\mu \right|$$

$$\leqslant \left| \int_K u_i \operatorname{sgn} g\, d\mu \right| + \int_A |u_i|\, d\mu$$

$$\leqslant \int_{Z(g)} |u_i|\, d\mu + \int_A |u_i|\, d\mu$$

$$= \int_{Z(\tilde{g})} |u_i|\, d\mu.$$

Thus the inequalities (3) are maintained. Since (5) holds for $i = 1, ..., r-1$, there must be strict inequality in the above set of inequalities. Thus (5) now holds for $i = 1, ..., r$.

Finally, we prove that (4) is maintained. Then applying the above process a finite number of times, we will have proven the lemma.

Recall that since $\mu(A) < \delta$,

$$2\int_A |\tilde{u}|\, d\mu < \int_K \tilde{u} \operatorname{sgn} g\, d\mu - \int_{Z(g)} |\tilde{u}|\, d\mu.$$

Now, from the above inequality,

$$\begin{aligned}\int_K \tilde{u} \operatorname{sgn} \tilde{g}\, d\mu &= \int_K \tilde{u} \operatorname{sgn} g\, d\mu - \int_A \tilde{u} \operatorname{sgn} g\, d\mu \\ &\geqslant \int_K \tilde{u} \operatorname{sgn} g\, d\mu - \int_A |\tilde{u}|\, d\mu \\ &> \int_{Z(g)} |\tilde{u}|\, d\mu + \int_A |\tilde{u}|\, d\mu \\ &= \int_{Z(\tilde{g})} |\tilde{u}|\, d\mu.\end{aligned}$$

Thus (4) continues to hold. ∎

LEMMA 5.5. *Let* $U = \operatorname{span}\{u_1, ..., u_n\}$ *be an* n*-dimensional subspace of* $C(K)$. *Assume that* $\tilde{u} \in U$ *and* $\tilde{g} \in L^1(K, \mu)$ *satisfy*

$$\left| \int_K u_i \operatorname{sgn} \tilde{g}\, d\mu \right| < \int_{Z(\tilde{g})} |u_i|\, d\mu, \qquad i = 1, ..., n, \tag{7}$$

and

$$\int_K \tilde{u} \operatorname{sgn} \tilde{g}\, d\mu > \int_{Z(\tilde{g})} |\tilde{u}|\, d\mu. \tag{8}$$

Then there exists an $f \in C(K)$ *such that* (7) *and* (8) *hold with* f *replacing* $\tilde{g}$.

Proof. Let

$$\eta = \min \begin{cases} \int_{Z(\tilde{g})} |u_i|\, d\mu - |\int_K u_i \operatorname{sgn} \tilde{g}\, d\mu|, & i = 1, ..., n \\ \int_K \tilde{u} \operatorname{sgn} \tilde{g}\, d\mu - \int_{Z(\tilde{g})} |\tilde{u}|\, d\mu. \end{cases}$$

From (7) and (8), $\eta > 0$. For ease of notation, set $\tilde{u} = u_{n+1}$ in what follows. Let

$$v_i(C) = \int_C |u_i|\, d\mu, \qquad i = 1, ..., n+1,$$

and

$$P = \{x : \tilde{g}(x) > 0\}$$
$$N = \{x : \tilde{g}(x) < 0\}.$$

From the regularity of μ, and the absolute continuity of the integral, for each $\varepsilon > 0$ there exist compact sets A, B and disjoint open sets V, W satisfying $A \subset P$, $B \subset N$, $A \subset V$, $B \subset W$, where each of the $v_i(P \backslash A)$, $v_i(N \backslash B)$, $v_i(V \backslash A)$, $v_i(W \backslash B)$ is strictly less than ε for $i = 1, ..., n+1$.

From Urysohn's Lemma, there exists an $f \in C(K)$ satisfying

$$f(x) = \begin{cases} 1, & x \in A \\ -1, & x \in B \\ 0, & x \notin (V \cup W). \end{cases}$$

Now, for $i = 1, ..., n+1$

$$\begin{aligned}
&\left| \int_K u_i \operatorname{sgn} \tilde{g} \, d\mu - \int_K u_i \operatorname{sgn} f \, d\mu \right| \\
&\quad = \left| \left(\int_P u_i \, d\mu - \int_N u_i \, d\mu \right) \right. \\
&\qquad \left. - \left(\int_A u_i \, d\mu - \int_B u_i \, d\mu + \int_{\operatorname{supp} f \backslash (A \cup B)} u_i \operatorname{sgn} f \, d\mu \right) \right| \\
&\quad \leqslant \int_{P \backslash A} |u_i| \, d\mu + \int_{N \backslash B} |u_i| \, d\mu + \int_{\operatorname{supp} f \backslash (A \cup B)} |u_i| \, d\mu \\
&\quad \leqslant \int_{P \backslash A} |u_i| \, d\mu + \int_{N \backslash B} |u_i| \, d\mu + \int_{(V \cup W) \backslash (A \cup B)} |u_i| \, d\mu \\
&\quad = v_i(P \backslash A) + v_i(N \backslash B) + v_i((V \cup W) \backslash (A \cup B)) \\
&\quad = v_i(P \backslash A) + v_i(N \backslash B) + v_i(V \backslash A) + v_i(W \backslash B) < 4\varepsilon.
\end{aligned}$$

For $i = 1, ..., n+1$,

$$\begin{aligned}
\int_{Z(f)} |u_i| \, d\mu - \int_{Z(\tilde{g})} |u_i| \, d\mu &\leqslant \int_{K \backslash (A \cup B)} |u_i| \, d\mu - \int_{K \backslash (P \cup N)} |u_i| \, d\mu \\
&= v_i((K \backslash (A \cup B)) \backslash (K \backslash (P \cup N))) \\
&= v_i((P \cup N) \backslash (A \cup B)) \\
&= v_i(P \backslash A) + v_i(N \backslash B) < 2\varepsilon,
\end{aligned}$$

while

$$\int_{Z(\bar{g})} |u_i|\, d\mu - \int_{Z(f)} |u_i|\, d\mu \leqslant \int_{K\backslash(P\cup N)} |u_i|\, d\mu - \int_{K\backslash(V\cup W)} |u_i|\, d\mu$$

$$\leqslant \int_{K\backslash(A\cup B)} |u_i|\, d\mu - \int_{K\backslash(V\cup W)} |u_i|\, d\mu$$

$$= v_i((K\backslash(A\cup B))\backslash(K\backslash(V\cup W))$$

$$= v_i(V\backslash A) + v_i(W\backslash B) < 2\varepsilon.$$

Thus

$$\left| \int_{Z(\bar{g})} |u_i|\, d\mu - \int_{Z(f)} |u_i|\, d\mu \right| < 2\varepsilon$$

for $i = 1, ..., n+1$. Choosing $6\varepsilon < \eta$, the lemma follows. ∎

Lemmas 5.4 and 5.5 prove Theorem 5.3.

In Theorems 5.2 and 5.3 we considered nonatomic measures. We now go to the other extreme and consider

$$l_1^d = \left\{ \mathbf{x} : \mathbf{x} \in \mathbb{R}^d,\ \|\mathbf{x}\|_1 = \sum_{i=1}^{d} |x_i| \right\}.$$

We present two positive results.

PROPOSITION 5.6. *Let $U = \operatorname{span}\{\mathbf{u}^1, ..., \mathbf{u}^n\}$ be an n-dimensional subspace of $\mathbb{R}^d$. Assume there exists a set of $n-1$ distinct indices $K = \{k_1, ..., k_{n-1}\} \subset \{1, ..., d\}$ for which*

$$u_{k_i}^j = \delta_{ij}, \qquad i, j = 1, ..., n-1,$$

$$u_{k_i}^n = 0, \qquad i = 1, ..., n-1,$$

and

$$\sum_{k \notin K} |u_k^j| < 1, \qquad j = 1, ..., n-1.$$

Then there does not exist a phantom approximation to any $\mathbf{b} \in l_1^d$ from $\operatorname{span}\{\mathbf{u}^1, ..., \mathbf{u}^n\}$.

Proof. Let $\mathbf{u}^*$ be a stationary point for $\mathbf{b}$ with respect to $\{\mathbf{u}^1, ..., \mathbf{u}^n\}$. Then

$$\left| \sum_{i=1}^{d} u_i^j \operatorname{sgn}(b_i - u_i^*) \right| \leqslant \sum_{i \in Z(\mathbf{b} - \mathbf{u}^*)} |u_i^j|, \qquad j = 1, ..., n. \tag{9}$$

The crux of our proof is the fact that $K \subseteq Z(\mathbf{b}-\mathbf{u}^*)$. Assume to the contrary that $k_1 \notin Z(\mathbf{b}-\mathbf{u}^*)$. Consider (9) for $j=1$. We have

$$1 - \sum_{\substack{i \notin Z(\mathbf{b}-\mathbf{u}^*) \\ i \neq k_1}} |u_i^1| \leqslant \left| \sum_{i=1}^{d} u_i^1 \operatorname{sgn}(b_i - u_i^*) \right| \leqslant \sum_{i \in Z(\mathbf{b}-\mathbf{u}^*)} |u_i^1|.$$

Thus

$$1 \leqslant \sum_{\substack{i=1 \\ i \neq k_1}}^{d} |u_i^1| = \sum_{k \notin K} |u_k^1|,$$

contradicting our hypothesis. Therefore $K \subseteq Z(\mathbf{b}-\mathbf{u}^*)$.

Let $\mathbf{u}^* = \sum_{i=1}^n a_i^* \mathbf{u}^i$. Since $K \subseteq Z(\mathbf{b}-\mathbf{u}^*)$, we have $b_{k_i} = u_{k_i}^*$, $i=1, ..., n-1$, implying by hypothesis that $a_i^* = b_{k_i}$, $i=1, ..., n-1$. Thus

$$\mathbf{b}-\mathbf{u}^* = \mathbf{b} - \sum_{i=1}^{n-1} b_{k_i} \mathbf{u}^i - a_n^* \mathbf{u}^n.$$

Since $\mathbf{u}^*$ is a stationary point for $\mathbf{b}$ with respect to $\{\mathbf{u}^n\}$, we have

$$\|\mathbf{b}-\mathbf{u}^*\|_1 = \min_{a_n} \left\| \mathbf{b} - \sum_{i=1}^{n-1} b_{k_i} \mathbf{u}^i - a_n \mathbf{u}^n \right\|_1.$$

Now, if $\tilde{\mathbf{u}} \in U$ is any best approximant to $\mathbf{b}$ from U, then (9) holds with $\tilde{\mathbf{u}}$ replacing $\tilde{\mathbf{u}}$. Thus as above, $K \subseteq Z(\mathbf{b}-\tilde{\mathbf{u}})$ and consequently $\tilde{\mathbf{u}} = \sum_{i=1}^{n-1} b_{k_i} \mathbf{u}^i - \tilde{a}_n \mathbf{u}^n$, where

$$\begin{aligned}\|\mathbf{b}-\tilde{\mathbf{u}}\|_1 &= \min_{a_1, ..., a_n} \left\| \mathbf{b} - \sum_{i=1}^{n} a_i \mathbf{u}^i \right\|_1 \\ &= \min_{a_n} \left\| \mathbf{b} - \sum_{i=1}^{n-1} b_{k_i} \mathbf{u}^i - a_n \mathbf{u}^n \right\|_1.\end{aligned}$$

Therefore $\|\mathbf{b}-\mathbf{u}^*\|_1 = \|\mathbf{b}-\tilde{\mathbf{u}}\|_1$, and $\mathbf{u}^*$ is a best approximant to $\mathbf{b}$ from U. ∎

PROPOSITION 5.7. *Let U be an n-dimensional subspace of $\mathbb{R}^{n+1}$. Then there exist $\mathbf{u}^1, ..., \mathbf{u}^n$ in U such that no $\mathbf{b} \in \mathbb{R}^{n+1}$ has a phantom approximation from $\{\mathbf{u}^1, ..., \mathbf{u}^n\}$.*

Proof. Since $\dim U = n$, $U \subset \mathbb{R}^{n+1}$, there exists a unique (up to multiplication by -1) vector $\mathbf{y} \in \mathbb{R}^{n+1}$ satisfying

(1) $\|\mathbf{y}\|_\infty = 1$.

(2) $\sum_{i=1}^{n+1} y_i u_i = 0$, all $\mathbf{u} \in U$.

Let $|y_j| = 1$. Then

$$|u_j| = |y_j u_j| = \left| -\sum_{\substack{i=1 \\ i \neq j}}^{n+1} y_i u_i \right| \leqslant \sum_{\substack{i=1 \\ i \neq j}}^{n+1} |u_i|$$

for all $\mathbf{u} \in U$. Without loss of generality, assume that $j = n+1$. Thus

$$|u_{n+1}| \leqslant \sum_{i=1}^{n} |u_i|$$

for all $\mathbf{u} \in U$. This implies that U restricted to the indices $\{1, ..., n\}$ is of full rank. For otherwise there exists a $\mathbf{u} \in U$, $\mathbf{u} \neq \mathbf{0}$, satisfying $u_i = 0$, $i = 1, ..., n$, in contradiction to the above inequality.

Define $\mathbf{v}^1, ..., \mathbf{v}^n$ in U by

$$v_i^j = \delta_{ij}, \qquad i, j = 1, ..., n.$$

The $\mathbf{v}^1, ..., \mathbf{v}^n$ span U, and

$$|v_{n+1}^j| \leqslant \sum_{i=1}^{n} |v_i^j| = |v_j^j| = 1, \qquad j = 1, ..., n.$$

If $|v_{n+1}^j| < 1$ for any $n-1$ of the $\mathbf{v}^1, ..., \mathbf{v}^n$, then we can set $\mathbf{u}^j = \mathbf{v}^j$, $j = 1, ..., n$, and apply the previous proposition.

Assume

$$|v_{n+1}^j| < 1, \qquad j = 1, ..., r-1$$

$$|v_{n+1}^j| = 1, \qquad j = r, ..., n$$

where $1 \leqslant r \leqslant n-1$. Note that the vector

$$\mathbf{y} = (v_{n+1}^1, ..., v_{n+1}^n, -1)$$

satisfies (1) and (2), i.e., $y_j = v_{n+1}^j$, $j = 1, ..., n$, $y_{n+1} = -1$. Thus $|y_j| = 1$, $j = r, ..., n$.

We define $\mathbf{u}^1, ..., \mathbf{u}^n$ as follows. For $j = 1, ..., r-1$, set $\mathbf{u}^j = \mathbf{v}^j$. For $j = r, ..., n$, set

$$\mathbf{u}^j = (0, ..., 0, (n-j+1)\, y_j, -y_{j+1}, ..., -y_n, 1).$$

Note that $\mathbf{u}^r, ..., \mathbf{u}^n \in U$ since $|y_j| = 1$, $j = r, ..., n$, and

$$\mathbf{u}^n = y_n \mathbf{v}^n$$

$$\mathbf{u}^j = (n-j+1)\, y_j \mathbf{v}^j - y_{j+1} \mathbf{v}^{j+1} - \cdots - y_n \mathbf{v}^n$$

for $j = r, ..., n-1$. In addition, $U = \operatorname{span}\{\mathbf{u}^1, ..., \mathbf{u}^n\}$.

Now assume that $\mathbf{u}^*$ is a stationary point for $\mathbf{b}$ with respect to $\{\mathbf{u}^1, ..., \mathbf{u}^n\}$. Then

$$\left|\sum_{i=1}^{n+1} u_i^j \operatorname{sgn}(b_i - u_i^*)\right| \leqslant \sum_{i \in Z(\mathbf{b}-\mathbf{u}^*)} |u_i^j|, \qquad j=1, ..., n.$$

As in the proof of the previous proposition, it follows that $\{1, ..., r-1\} \subseteq Z(\mathbf{b}-\mathbf{u}^*)$.

We next prove that there exists a $\delta \in \{-1, 1\}$ for which

$$\delta y_i = \operatorname{sgn}(b_i - u_i^*)$$

for all $i \notin Z(\mathbf{b}-\mathbf{u}^*)$. Let $k = \min\{i : i \notin Z(\mathbf{b}-\mathbf{u}^*)\}$. If $k=n+1$ there is nothing to prove. Assume $k \leqslant n$. Since $\{1, ..., r-1\} \subseteq Z(\mathbf{b}-\mathbf{u}^*)$, we have $r \leqslant k$. Now

$$\left|\sum_{i=1}^{n+1} u_i^k \operatorname{sgn}(b_i - u_i^*)\right| \leqslant \sum_{i \in Z(\mathbf{b}-\mathbf{u}^*)} |u_i^k|.$$

Thus

$$\begin{aligned} &\Bigg|(n-k+1)\, y_k \operatorname{sgn}(b_k - u_k^*) \\ &\qquad - \sum_{i=k+1}^{n} y_i \operatorname{sgn}(b_i - u_i^*) + \operatorname{sgn}(b_{n+1} - u_{n+1}^*)\Bigg| \\ &\quad \leqslant \#\{i : i \in Z(\mathbf{b}-\mathbf{u}^*), \qquad i \geqslant k+1\}, \end{aligned}$$

where $\#\{A\}$ indicates the number of elements in A. Since $|y_i| = 1$ for $i=k, ..., n$,

$$\begin{aligned} &\Bigg|(n-k+1)\, y_k \operatorname{sgn}(b_k - u_k^*) \\ &\qquad - \sum_{i=k+1}^{n} y_i \operatorname{sgn}(b_i - u_i^*) + \operatorname{sgn}(b_{n+1} - u_{n+1}^*)\Bigg| \\ &\quad \geqslant (n-k+1) - \#\{i : i \notin Z(\mathbf{b}-\mathbf{u}^*), i \geqslant k+1\}. \end{aligned} \tag{10}$$

Because

$$\begin{aligned} &\#\{i : i \in Z(\mathbf{b}-\mathbf{u}^*), i \geqslant k+1)\} \\ &\qquad + \#\{i : i \notin Z(\mathbf{b}-\mathbf{u}^*), i \geqslant k+1\} = n-k+1, \end{aligned}$$

we must have equality in (10), implying that if we set $\delta = y_k \operatorname{sgn}(b_k - u_k^*)$, then

$$\delta = y_i \operatorname{sgn}(b_i - u_i^*),$$

for $i \notin Z(\mathbf{b} - \mathbf{u}^*)$, $k+1 \leqslant i \leqslant n$, and

$$\delta y_{n+1} = \operatorname{sgn}(b_{n+1} - u_{n+1}^*)$$

if $n+1 \notin Z(\mathbf{b} - \mathbf{u}^*)$ since $y_{n+1} = -1$. Thus

$$\delta y_i = \operatorname{sgn}(b_i - u_i^*)$$

for all $i \notin Z(\mathbf{b} - \mathbf{u}^*)$.

Now, for any $\mathbf{u} \in U$, we have from properties (1) and (2) of $\mathbf{y}$

$$\begin{aligned}\left|\sum_{i=1}^{n+1} u_i \operatorname{sgn}(b_i - u_i^*)\right| &= \left|\sum_{i \notin Z(\mathbf{b}-\mathbf{u}^*)} u_i\, \delta y_i\right| \\ &= \left|\sum_{i \notin Z(\mathbf{b}-\mathbf{u}^*)} u_i y_i\right| \\ &= \left|-\sum_{i \in Z(\mathbf{b}-\mathbf{u}^*)} u_i y_i\right| \\ &\leqslant \sum_{i \in Z(\mathbf{b}-\mathbf{u}^*)} |u_i|.\end{aligned}$$

Then $\mathbf{u}^*$ is a best approximant to $\mathbf{b}$ from U. ∎

Remark. There is a general result which can provide a better method of approximating in l_1^d. If U is an n-dimensional subspace of $\mathbb{R}^d$, and there exists a set of n distinct indices $L = \{l_1, ..., l_n\}$ in $\{1, ..., d\}$ for which

$$\sum_{i \notin L} |u_i| \leqslant \sum_{i \in L} |u_i|$$

for all $\mathbf{u} \in U$, then given any $\mathbf{b} \in \mathbb{R}^d$ there exists a $\mathbf{u}^* \in U$ satisfying $b_{l_i} - u_{l_i}^* = 0$, $i = 1, ..., n$. In addition, $\mathbf{u}^*$ is necessarily a best approximant to U from $\mathbf{b}$ (Pinkus [6, p. 137]). If $d = n+1$, then such a set L necessarily exists (see the beginning of the proof of Proposition 5.7).

Remark. In contrast to our results for $C(B)$, in this section we had only considered linearly independent $u_1, ..., u_n$. That is, we always set $m = n$. It is known that the problem of best approximation in l_1^d (for any d) is equivalent to a linear programming problem. If one could find a reasonable bound m dependent on d and n, such that for any n-dimensional subspace

U of l_1^d there exist $\mathbf{u}^1, \ldots, \mathbf{u}^m$ in U such that no $\mathbf{b}$ has a phantom approximation from $\{\mathbf{u}^1, \ldots, \mathbf{u}^m\}$, then one could develop an alternative algorithm to the simplex method for solving the standard linear programming problem.

ACKNOWLEDGMENT

Thanks are due to Professor R. Olshen for introducing us to the topic of median polish.

REFERENCES

1. D. BRADU, An ε-median polish algorithm, *Comput. Statist. Data Anal.* **5** (1987), 327–336.
2. F. DEUTSCH, The alternating method of von Neumann, *in* "Multivariate Approximation Theory" (W. Schempp and K. Zeller, Eds.), pp. 83–96, International Series of Numerical Mathematics, Vol. 51, Birkhäuser, Basel, 1979.
3. J. H. B. KEMPERMAN, Least absolute value and median polish, *in* "Inequalities in Statistics and Probability" (Y. L. Tong, Ed.), pp. 84–103, Lecture Notes–Monograph Series, Vol. 5, Inst. Math. Statist., Hayward, CA, 1984.
4. W. A. LIGHT AND E. W. CHENEY, "Approximation Theory in Tensor Product Spaces," Lecture Notes in Mathematics, Vol. 1169, Springer-Verlag, Berlin, 1985.
5. J. T. MARTI, A method for the numerical computation of best L_1-approximation of continuous functions, *in* "Numerische Methoden der Approximationstheorie, Band 2" (L. Collatz and G. Meinardus, Eds.), pp. 79–92, International Series of Numerical Mathematics, Vol. 26, Birkhäuser, Basel, 1975.
6. A. PINKUS, "On L^1-Approximation," Cambridge Tracts in Mathematics, Vol. 93, Cambridge University Press, Cambridge, 1989.
7. K. T. SMITH, D. C. SOLOMON, AND S. L. WAGNER, Practical and mathematical aspects of the problem of reconstructing objects from radiographs, *Bull. Amer. Math. Soc.* **83** (1977), 1227–1270.
8. F. E. SULLIVAN AND B. ATLESTAM, Descent methods in smooth, rotund spaces with applications to approximation in L^p, *J. Math. Anal. Appl.* **48** (1974), 155–164.
9. J. W. TUKEY, "Exploratory Data Analysis," Addison-Wesley, Reading, MA, 1977.
10. K. H. USOW, On L_1 approximation. I. Computation for continuous functions and continuous dependence, *SIAM J. Numer. Anal.* **4** (1967), 70–88.
11. W. I. ZANGWILL, "Nonlinear Programming: A Unified Approach," Prentice-Hall, Englewood Cliffs, NJ, 1969.