Constr. Approx. (1990) 6: 35-64

CONSTRUCTIVE APPROXIMATION

Constrained Best Approximation in Hilbert Space

Charles K. Chui, Frank Deutsch, and Joseph D. Ward

Abstract. In this paper we study the characterization of the solution to the extremal problem

$$\inf\{\|x\| \mid x \in C \cap M\},$$

where x is in a Hilbert space H, C is a convex cone, and M is a translate of a subspace of H determined by interpolation conditions. We introduce a simple geometric property called the "conical hull intersection property" that provides a unifying framework for most of the basic results in the subject of optimal constrained approximation. Our approach naturally lends itself to considering the data cone as opposed to the constraint cone. A nice characterization of the solution occurs, for example, if the data vector associated with M is an interior point of the data cone.

1. Introduction

An important area of research in approximation theory is constrained approximation, which can be briefly described as follows. In addition to approximating or interpolating a given set of data, the approximant is also required to preserve certain shapes such as positivity, monotonicity, and/or convexity that the data set dictates. In various specific formulations, this problem is often posed in data analysis, computer-aided geometric design, and mathematical modeling (see [6]). One common deficiency is the lack of a good criterion since the problem typically has infinitely many solutions. The usual approach is to select the solution from the ones that satisfy an optimization property. In other words, the mathematical problem is to study the existence, uniqueness, characterization, and computational aspects of the solution to the extremal problem

$$\inf\{\|x\| \mid x \in C \cap M\},$$

where x is in a normed linear space X, C is a convex cone that defines the constraint, and M is a translate of a subspace of X determined by the interpolation conditions. The available techniques in the study of this problem include the Kuhn-Tucker optimization method, the Langrange duality, and various methods

Date received: November 2, 1987. Date revised: October 15, 1988. Communicated by Charles A. Micchelli.
AMS classification: Primary 41A65; Secondary 41A29, 41A05.
Key words and phrases: Conical hull, Polar of a cone, Best constrained interpolation.

in variational calculus and optimal control theory. All these techniques are essentially built on the Hahn–Banach separation theorem.

In this paper we introduce a simple geometric property called the "conical hull intersection property" (or CHIP for short) that provides a unifying framework for the basic results in the subject of optimal constrained approximation. In addition to its simplicity, the property CHIP allows us to handle finitely many constraint cones and it encompasses much of the known material. However, we do not investigate the "dual" problem which is important for the development of algorithms. For this line of thinking, the reader is referred to [4] and [17]. Our approach naturally lends itself to considering the *data* cone as opposed to the constraint cone. Thus, for example, we show that the assumption used in Micchelli and Utreras [17], which we call the MU property for short, is equivalent to the corresponding data vector being interior to the data cone.

This paper is divided into six sections. Following the introduction, Section 2 lays the groundwork for the remainder of the paper by introducing the property CHIP. In addition, several ancilliary results about CHIP are derived. In Section 3 we restrict attention to the case of a cone intersecting a flat. Various conditions including the Slater point condition and the MU property appearing in [17] are shown to have the property CHIP. Along the way we show that the MU property is equivalent to the data vector being an interior point in the data cone, and thus recover the result that the minimum norm interpolant has the form $P_C(x)$, where P_C denotes the metric projection onto C. In Section 4 we also recover the main results of [16] and [22] by using the CHIP approach. Section 5 deals with the case of infinite interpolation constraints while Section 6 interprets the CHIP property in a geneal normed linear space setting.

Much of the notation is defined as it is used. However, certain symbols are used repeatedly. In particular, int K, $\bar{K}$, span K, and co K corresponding to a set of vectors K are used to denote the interior of K, K closure, the linear span of K, and the convex hull of K, respectively. In addition $B(x, \varepsilon)$ denotes the open ball of radius ε centered at x, and $\mathcal{R}(A)$ represents the range of a linear operator A.

In this paper we concentrate our attention to the Hilbert space setting although most of our results have Banach space formulations.

2. Dual Cones

The concept which is fundamental to our work is the notion of a dual cone of a given set. Throughout this paper, unless otherwise specified, X always denotes a (real) Hilbert space.

Definition. Let S be any nonempty set in X. The *dual cone* of S is defined by

$$S^0 := \{x \in X \mid \langle x, y\rangle \leq 0 \text{ for every } y \in S\}. \tag{2.1}$$

We remark that the dual cone of S is also called the (negative) polar of S.

The importance of the dual cone from our viewpoint stems from its role in characterizing best approximations from convex sets. More precisely, we have the following.

Proposition 2.1. *Let K be a convex subset of X, $x \in X$, and $k_0 \in K$. Then the following statements are equivalent*:

(1) *k_0 is the best approximation to x from K*;
(2) *$\langle x - k_0, k - k_0 \rangle \leq 0$ for every $k \in K$*;
(3) *$x - k_0 \in (K - k_0)^0$.*

The equivalence of the first two statements is a classical result (see Moskovitz and Dines [18]), while the equivalence of (2) and (3) is obvious. Here, recall that $k_0 \in K$ is a *best approximation* to x if $\|x - k_0\| = \inf\{\|x - k\| | k \in K\}$. Further, since K is convex, a best approximation, when it exists, is necessarily unique, and we denote it by $P_K(x)$. Thus, $k_0 = P_K(x)$ if and only if $x - k_0 \in (K - k_0)^0$. This shows that to obtain detailed information on which element $k_0 \in K$ is the best approximation to x, we must have a precise description of the dual cone $(K - k_0)^0$. In practice, K often has a representation as the intersection of a collection of "simpler" convex sets K_i, $K = \bigcap_i K_i$, so that the dual cones $(K_i - k_0)^0$ are "easier" to obtain. It would be desirable in this case to be able to express the dual cone $(K - k_0)^0$ in terms of the individual dual cones $(K_i - k_0)^0$. We give the precise condition on the collection $\{K_i | i \in I\}$ in order that this is possible (see property CHIP in the definition following Lemma 2.4). First, however, we record some useful facts about dual cones.

Recall that a set C is called a *convex cone* if it satisfies

$$C + C \subset C \quad \text{and} \quad \rho C \subset C \quad \text{for all} \quad \rho \geq 0.$$

The *conical hull* of a nonempty set S in X, denoted by con S, is the intersection of all convex cones which contains S. With the exception of part (4) of Lemma 2.2 and Lemma 2.4 whose derivations require the separation theorem, the following four lemmas are standard and their proofs are therefore omitted.

Lemma 2.1.

(1) *$\operatorname{con} S$ is a convex cone.*
(2) *$\overline{\operatorname{con}}\, S := \overline{\operatorname{con} S}$ is a closed convex cone.*
(3) *$S = \operatorname{con} S$ if and only if S is a convex cone.*
(4) $\operatorname{con} S = \{\sum_1^n \rho_i x_i | \rho_i \geq 0, x_i \in S, n \in \mathbb{N}\}$.
(5) *If S is convex and $0 \in S$, then* $\operatorname{con} S = \{\rho y | \rho > 0, y \in S\}$.

Lemma 2.2. *Let $\varnothing \neq S \subset X$. Then:*

(1) *S^0 is a closed convex cone*;
(2) $S^0 = (\bar{S})^0$;
(3) $S^0 = (\operatorname{con} S)^0 = (\overline{\operatorname{con}}\, S)^0$;
(4) $S^{00} = \overline{\operatorname{con}}\, S$.

As usual, the *annihilator* of a set S is denoted by $S^{\perp}$.

Lemma 2.3. *Let C be a convex cone and let M be a (linear) subspace of X. Then*:

(1) $C^{00} = \bar{C}$;
(2) $(C-y)^0 = C^0 \cap y^{\perp}$ *for any* $y \in C$;
(3) $M^0 = M^{\perp}$.

In addition, the *sum* of a collection of subsets S_i of X is denoted by $\sum S_i$.

Lemma 2.4. *Let $\{C_i | i \in I\}$ be a collection of closed convex cones in X. Then*

$$\left(\bigcap_i C_i\right)^0 = \overline{\sum_i C_i^0}. \tag{2.2}$$

We are now ready to introduce the notion of property CHIP.

Definition. A collection of convex sets $\{K_i | i \in I\}$ is said to have property *CHIP* (Conical Hull Intersection Property) if

$$\overline{\text{con}}\bigcap_i (K_i - k) = \bigcap_i \overline{\text{con}}(K_i - k) \tag{2.3}$$

for all $k \in \bigcap_i K_i$.

As mentioned earlier, property CHIP is the precise condition which allows a representation of the dual cone of $\bigcap_i (K_i - k)$ in terms of the dual cones of the individual sets $K_i - k$. This is the content of the equivalence of (1) and (4) in the next lemma.

Lemma 2.5. *Let $\{K_i | i \in I\}$ be a collection of convex sets in X. Then the following statements are equivalent*:

(1) $\{K_i | i \in I\}$ *has property* CHIP;
(2) $\overline{\text{con}}\bigcap_i (K_i - k) \supset \bigcap_i \overline{\text{con}}(K_i - k)$ *for every* $k \in \bigcap_i K_i$;
(3) $\text{con}\bigcap_i (K_i - k)$ *is dense in* $\bigcap_i \overline{\text{con}}(K_i - k)$ *for every* $k \in \bigcap_i K_i$;
(4) $(\bigcap_i K_i - k)^0 = \sum_i (K_i - k)^0$ *for every* $k \in \bigcap_i K_i$.

Proof. Since $\overline{\text{con}}\bigcap_i (K_i - k) \subset \bigcap_i \overline{\text{con}}(K_i - k)$ is always true, the equivalence of (1) and (2) is obvious. In addition, the equivalence of (2) and (3) is trivial.

To prove that (1) implies (4), we note that, by Lemmas 2.2 and 2.4,

$$\begin{aligned}(\textstyle\bigcap_i K_i - k)^0 &= \left[\bigcap_i (K_i - k)\right]^0 = [\overline{\text{con}}\bigcap_i (K_i - k)]^0 \\ &= \left[\bigcap_i \overline{\text{con}}(K_i - k)\right]^0 = \overline{\sum_i [\overline{\text{con}}(K_i - k)]^0} = \overline{\sum_i (K_i - k)^0}\end{aligned}$$

for any $k \in \bigcap_i K_i$.

Now suppose that (4) holds for all $k \in \bigcap_i K_i$. If (2) fails for some k, there exists an

$$x \in \bigcap_i \overline{\operatorname{con}}(K_i - k) \backslash \overline{\operatorname{con}} \textstyle\bigcap_i (K_i - k).$$

By the separation theorem, there is a $z \in X$ so that

$$\sup\{\langle z, y\rangle | y \in \overline{\operatorname{con}} \bigcap_i (K_i - k)\} = 0 < \langle z, x\rangle. \tag{2.4}$$

Then $\langle z, y\rangle \leq 0$ for all $y \in \bigcap_i (K_i - k)$ so

$$z \in \left[\bigcap_i (K_i - k)\right]^0 = \left[\bigcap_i K_i - k\right]^0 = \overline{\sum_i (K_i - k)^0} = \overline{\sum_i [\overline{\operatorname{con}}(K_i - k)]^0}.$$

Thus, there exist $z_n \in \sum_i [\overline{\operatorname{con}}(K_i - k)]^0$ converging to z. Since $x \in \overline{\operatorname{con}}(K_i - k)$ for every i, we obtain $\langle z_n, x\rangle \leq 0$ for all n and so $\langle z, x\rangle \leq 0$. But this contradicts (2.4). Hence (2) must hold. ■

As an immediate consequence of Proposition 2.1 and Lemma 2.5, we obtain the main characterization theorem of this paper.

Theorem 2.1. *Let $\{K_i | i \in I\}$ be a collection of convex subsets of X which has property CHIP and $K = \bigcap_i K_i$. Then, for each $k_0 \in K$,*

$$(K - k_0)^0 = \overline{\sum_i (K_i - k_0)^0}. \tag{2.5}$$

Moreover, for any $x \in X$ and $k_0 \in K$, $k_0 = P_K(x)$ if and only if

$$x - k_0 \in \overline{\sum_i (K_i - k_0)^0}. \tag{2.6}$$

To apply this theorem, it is essential to know when K can be represented as the intersecton of some convex sets $\{K_i | i \in I\}$ that have property CHIP. The following examples shed some light on this question. In the next section we examine in detail the situation when $K = C \cap V$, where C is a closed convex cone and V is a closed linear variety.

The next result follows easily from Lemma 2.4 and Theorem 2.1.

Proposition 2.2. *Any collection of closed linear varieties $\{V_i | i \in I\}$ has property CHIP. Moreover, for each $v \in V := \bigcap_i V_i$,*

$$(V - v)^0 = \overline{\sum_i M_i^\perp}, \tag{2.7}$$

where $M_i = V_i - v_i$ with $v_i \in V_i$. In addition, for any $x \in X$ and $v_0 \in V$, $v_0 = P_V(x)$ if and only if

$$x - v_0 \in \overline{\sum_i M_i^\perp}. \tag{2.8}$$

Before proceeding, it is perhaps worthwhile to give an example which shows that not all collections have property CHIP.

Example 2.1. Consider the closed convex cone

$$K_1 = \text{con}\{(\alpha, \beta, 1) | \alpha^2 + (\beta - 1)^2 = 1\}$$

and the closed linear variety

$$K_2 = k_0 + \{\alpha(1, 0, 0) | \alpha \in \mathbf{R}\},$$

where $k_0 = (0, 0, 1)$. It is easy to check that $k_0 \in K_1 \cap K_2$, $(K_1 - k_0) \cap (K_2 - k_0) = (0, 0, 0)$, and so $\overline{\text{con}}[(K_1 - k_0) \cap (K_2 - k_0)] = (0, 0, 0)$. But with $\overline{\text{con}}(K_1 - k_0) = \{(\alpha, \beta, \gamma) | \beta \geq 0\}$ and $M := \overline{\text{con}}(K_2 - k_0) = \{\beta(1, 0, 0) | \beta \in \mathbf{R}\}$, we have

$$\overline{\text{con}}(K_1 - k_0) \cap \overline{\text{con}}(K_2 - k_0) = M \neq (0, 0, 0),$$

so that $\{K_1, K_2\}$ fails to satisfy property CHIP.

While property CHIP fails in general even for a collection of two convex sets, it is easy to show that the conical hull (rather than closed conical hull) can *always* be interchanged with intersection, provided that the collection of convex sets is finite.

3. Best Constrained Interpolation

In this section we consider the special case when the convex set K is the intersection of a closed convex cone and a closed linear variety. This situation includes a number of problems (e.g., "shape-preserving" or constrained interpolation) studied by several authors in recent years (see, e.g., [4]-[6], [15], and the survey article [23]).

Our set-up in this section is the following: X and Y are Hilbert spaces, C is a closed convex cone in X, $d \in Y$, $A \in \mathcal{B}(X, Y)$, i.e., A is a bounded linear operator from X into Y, $V = \{x \in X | Ax = d\}$, $K = C \cap V = \{x \in C | Ax = d\}$, $x \in X$, and $k_0 \in K$. We wish to characterize $k_0 = P_K(x)$ by using Theorem 2.1.

Note that $V = \mathcal{N}(A) + k_0$, where $\mathcal{N}(A) = \{x \in X | Ax = 0\}$, and

$$K - k_0 = (C - k_0) \cap \mathcal{N}(A).$$

Theorem 3.1. *Let C, A, V, and K be as above and assume that $\{C, V\}$ has property CHIP. Then for any $x \in X$ and $k_0 \in K$ the following statements are equivalent:*

(1) $k_0 = P_K(x)$;

(2) $x - k_0 \in \overline{C^0 \cap k_0^\perp + \mathcal{R}(A^*)}$, *where $\mathcal{R}(A^*)$ denotes the range of the adjoint A^*.*

Proof. By Theorem 2.1, $k_0 = P_K(x)$ if and only if $x - k_0 \in \overline{(C - k_0)^0 + \mathcal{N}(A)^\perp}$. But since it is well known (see [9]) that $\mathcal{N}(A)^\perp = \overline{\mathcal{R}(A^*)}$ and $(C - k_0)^0 = C^0 \cap k_0^\perp$ from (2) in Lemma 2.3, we have

$$\overline{(C - k_0)^0 + \mathcal{N}(A)^\perp} = \overline{C^0 \cap k_0^\perp + \overline{\mathcal{R}(A^*)}} = \overline{C^0 \cap k_0^\perp + \mathcal{R}(A^*)}$$

which completes the proof. ■

Before giving applications of this theorem, it will be useful to discuss and compare various properties which guarantee that the pair $\{C, V\}$ has property CHIP.

Note that $K := C \cap V \neq \varnothing$ if and only if $d \in AC := \{Ax | x \in C\}$. We call any point in AC an *admissible data point*, and AC the *data cone*. Some conditions stronger than d being an admissible data point are given below, and their relationships are shown in Lemma 3.1.

Definition.

(1) d is called a *Slater point* if $d \in A(\operatorname{int} C)$ (or equivalently, $V \cap \operatorname{int} C \neq \varnothing$).
(2) d is called a *strong interior data point* if there exist $v_0 \in C \cap V$ and $r > 0$ so that $d \in \operatorname{int} A[C \cap B(v_0, r)]$.
(3) d is called an *interior data point* provided $d \in \operatorname{int} AC$.

Lemma 3.1. *Consider the following statements*:

(1) *d is a Slater point*;
(2) *d is a strong interior data point*;
(3) *d is an interior data point*;
(4) *$\{C, V\}$ has property CHIP.*

Then $(2) \Rightarrow (3) \Rightarrow (4)$.

In addition, if A is surjective, then $(1) \Rightarrow (2)$. *Furthermore, if Y is finite-dimensional, then* $(2) \Leftrightarrow (3)$. *However, in general,* $(2) \not\Rightarrow (1)$ *and* $(4) \not\Rightarrow (3)$.

Proof. (i) $(2) \Rightarrow (3)$ is obvious.

(ii) To show $(3) \Rightarrow (4)$, assume $d \in \operatorname{int} AC$. Then

$$\overline{B(d, \varepsilon)} \subset AC \tag{3.1}$$

for some $\varepsilon > 0$. We first show that there exists $r > 0$, $\delta > 0$, and $y \in Y$ so that

$$B(y, \delta) \subset A[(C - k_0) \cap B(0, r)]. \tag{3.2}$$

To see this, note that by (3.1)

$$\bigcup_{N=1}^{\infty} \{A[(C - k_0) \cap B(0, N)] \cap \overline{B(0, \varepsilon)}\} = \overline{B(0, \varepsilon)}.$$

By the Baire category theorem, there exists an integer $r = N_0$ so that

$$\overline{A[(C - k_0) \cap B(0, r)] \cap \overline{B(0, \varepsilon)}}$$

contains an interior point y; namely,

$$B(y, \delta) \subset \overline{A[(C - k_0) \cap B(0, r)] \cap \overline{B(0, \varepsilon)}} \subset \overline{A[(C - k_0) \cap B(0, r)]}. \tag{3.3}$$

Given any $y' \in B(y, \delta)$, there exists an $x_n \in (C - k_0) \cap B(0, r)$ so that $y' = \lim Ax_n$. Since $\{x_n\}$ is bounded, by passing to a subsequence, we may assume $x_n \to x$ weakly. But $(C - k_0) \cap \overline{B(0, r)}$ is closed and convex, hence weakly closed. Thus, $x \in (C - k_0) \cap \overline{B(0, r)}$. Further, since A is weakly continuous, we have

$$y' = \lim Ax_n = Ax \in A[(C - k_0) \cap \overline{B(0, r)}].$$

Replacing r with anything bigger, we deduce that

$$y' \in A[(C-k_0) \cap B(0, r)]$$

and so (3.2) holds.

Next we show that there exist $\delta' > 0$ and $r' > 0$ so that

$$B(0, \delta') \subset A[(C-k_0) \cap B(0, r')]. \tag{3.4}$$

To see this, note that by (3.3), $y \in \overline{B(0, \varepsilon)}$ so that $\|y\| \leq \varepsilon$. By (3.1), we have

$$y \in \overline{B(0, \varepsilon)} \subset AC - d = A(C-k_0).$$

Now since $-y \in \overline{B(0, \varepsilon)}$, there exists $x_0 \in (C-k_0)$ so that $Ax_0 = -y$. Let $r' = \frac{1}{2}r + \frac{1}{2}\|x_0\|$. Then for each $w \in B(0, \delta/2)$, $2w \in B(0, \delta)$ so $(2w+y) \in B(y, \delta)$ and (3.2) implies that there exists an $x \in (C-k_0) \cap B(0, r)$ such that $2w + y = Ax$. Hence

$$\tfrac{1}{2}(x+x_0) \in (C-k_0) \cap B(0, r')$$

and

$$w = \tfrac{1}{2}Ax - \tfrac{1}{2}y = A(\tfrac{1}{2}(x+x_0)) \subset A[(C-k_0) \cap B(0, r')].$$

This proves that (3.4) holds with $\delta' = \delta/2$.

To verify that $\{C, V\}$ has property CHIP, it suffices by Lemma 2.5 to show that for each $k_0 \in C \cap V$ and each $z \in \overline{\text{con}}(C-k_0) \cap M$, there exists $z_n \in \text{con}(C-k_0) \cap M$ so that $z_n \to z$, where $M := V - k_0 = \mathcal{N}(A)$. If $z = 0$, we take $z_n = 0$ for all n. If $z \neq 0$, then by scaling we may assume $\|z\| = 1$. Since $z \in \overline{\text{con}}(C-k_0)$, it follows from Lemma 2.1 that there exist $x_n \in (C-k_0)$ and $\rho_n > 0$ so that $\rho_n x_n \to z$. Then $\rho_n \|x_n\| \to \|z\| = 1$. There are two cases to consider.

Case 1. Suppose that $\{\rho_n\}$ is unbounded.

Then by passing to a subsequence, we may assume $\rho_n \to \infty$, so that $\|x_n\| \to 0$. Now

$$\begin{aligned}\|Ax_n\| &= \|x_n\| \left\| A\left(\frac{x_n}{\|x_n\|}\right) \right\| = \|x_n\| \left\| A\left(\frac{x_n}{\|x_n\|} - z\right) \right\| \\ &\leq \|x_n\| \, \|A\| \left\| \frac{x_n}{\|x_n\|} - z \right\| = \|x_n\| \, \|A\| \Delta_n,\end{aligned}$$

where

$$\Delta_n := \left\| \frac{x_n}{\|x_n\|} - z \right\| = \left\| \frac{\rho_n x_n}{\rho_n \|x_n\|} - z \right\| \to 0$$

since $\rho_n \|x_n\| \to 1$. Set $y_n = Ax_n$. Then it follows that $\|y_n\| \leq \|A\| \, \|x_n\| \Delta_n$ for all n.

Next choose $r_n \to \infty$ so that:

(a) $r_n y_n \in B(0, \delta')$ (and hence $(r_n - 1)y_n \in B(0, \delta')$) and
(b) $r_n \|x_n\| \to \infty$ (and hence $1/r_n \|x_n\| \to 0$ and $\rho_n / r_n \to 0$). For example, we may take $r_n = \delta'/2\|y_n\|$ if $y_n \neq 0$ and $r_n = n$ if $y_n = 0$.

Finally, by (3.4) we may choose $w_n \in (C - k_0) \cap B(0, r')$ so that

$$Aw_n = (1 - r_n)y_n \qquad \text{for all } n.$$

Then $\{w_n\}$ is bounded,

$$w'_n := \left(\frac{r_n - 1}{r_n}\right)x_n + \frac{1}{r_n} w_n \in (C - k_0)$$

by convexity, and

$$Aw'_n = \frac{r_n - 1}{r_n} Ax_n + \frac{1}{r_n} Aw_n = \frac{r_n - 1}{r_n} y_n + \frac{1}{r_n}(1 - r_n)y_n = 0.$$

Thus, $w'_n \in (C - k_0) \cap M$. Setting $z_n = \rho_n w'_n$, we see that $z_n \in \text{con}(C - k_0) \cap M$ and

$$\|z_n - z\| = \left\| \rho_n \left[\left(\frac{r_n - 1}{r_n}\right) x_n + \frac{1}{r_n} w_n \right] - z \right\|$$
$$\leq \left\| \frac{r_n - 1}{r_n} [\rho_n x_n - z] \right\| + \frac{1}{r_n} + \frac{\rho_n}{r_n} \|w_n\| \to 0.$$

Case 2. Suppose that $\{\rho_n\}$ is bounded.

Then by passing to a subsequence, we may assume that $\rho_n \to \rho_0 \geq 0$. Note that, for all $\alpha_n > 0$,

$$\rho_n \alpha_n \left(\frac{x_n}{\alpha_n}\right) = \rho_n x_n \to z.$$

Choose $\alpha_n > 1$ large enough so that $\rho'_n := \rho_n \alpha_n \to \infty$. Then by convexity,

$$\frac{x_n}{\alpha_n} = \frac{1}{\alpha_n} x_n + \left(1 - \frac{1}{\alpha_n}\right) 0 \in (C - k_0)$$

so $x_n/\alpha_n = x'_n \in (C - k_0)$, $\rho'_n x'_n \to z$, and $\rho'_n \to \infty$. Now Case 1 applies. This completes the proof of (3)⇒(4).

(iii) Now assume that A is surjective. We have to show (1)⇒(2). Suppose that d is a Slater point. Then there exists $v_0 \in V$ and $\varepsilon > 0$ so that $B(v_0, \varepsilon) \subset C$. By the open mapping theorem, $A[B(0, \varepsilon)] \supset B(0, \delta)$ for some $\delta > 0$. Thus,

$$A[C \cap B(v_0, \varepsilon)] = A[B(v_0, \varepsilon)] = A[B(0, \varepsilon) + v_0]$$
$$= A[B(0, \varepsilon)] + d \supset B(0, \delta) + d = B(d, \delta).$$

Thus, $d \in \text{int}\, A[C \cap B(v_0, \varepsilon)]$, so that (2) holds.

(iv) Finally, assume Y is finite-dimensional and d is an interior data point. In this case we may assume that $Y = \mathbf{R}^n$ and there exists an independent set $\{x_1, x_2, \ldots, x_n\}$ in X so that

$$Ax = (\langle x, x_1 \rangle, \langle x, x_2 \rangle, \ldots, \langle x, x_n \rangle), \qquad x \in X.$$

Since $d \in \text{int}\, AC$ and all norms are equivalent in $\mathbf{R}^n$, there is an $\varepsilon > 0$ so that $\overline{B_1(d, \varepsilon)} \subset AC$, where $B_1(d, \varepsilon)$ denotes the l_1-ball about d of radius ε. Since the set of extreme points ext $\overline{B_1(d, \varepsilon)}$ is finite, say

$$\text{ext}\, \overline{B_1(d, \varepsilon)} = \{e_1, e_2, \ldots, e_m\},$$

by choosing $c_i \in C$ so that $Ac_i = e_i$ for $i = 1, 2, \ldots, m$, we have

$$\overline{B_1(d, \varepsilon)} = \mathrm{co}\{e_1, e_2, \ldots, e_m\} = \mathrm{co}\{Ac_1, Ac_2, \ldots, Ac_m\} = AC_1,$$

where $C_1 := \mathrm{co}\{c_1, c_2, \ldots, c_m\} \subset C$. In particular, C_1 is a bounded convex subset of C. Choose any $v_0 \in C \cap V$ (so $Av_0 = d$), and let $C_2 = \mathrm{co}\{C_1, v_0\}$, we see that C_2 is a bounded convex subset of C. Finally, choose $r > 0$ large enough so that $B(v_0, r) \supset C_2$. Then $B(v_0, r) \cap C \supset C_2$ and

$$B_1(d, \varepsilon) \subset \overline{B_1(d, \varepsilon)} = AC_2 \subset A[C \cap B(v_0, r)].$$

Since for some $0 < \varepsilon' < \varepsilon$ we have $B(d, \varepsilon') \subset B_1(d, \varepsilon)$, it follows that d is a strong interior data point. Thus, (3)$\Rightarrow$(2) when Y is finite-dimensional.

The last statement of the lemma follows from examples given below. ∎

Recently, Micchelli and Utreras [17] showed that under certain additional restrictions, the conclusion of Theorem 3.1 may be strengthened; namely, the closure bar may be removed from (2) and (3). (However, as we shall see in the examples below, the closure bar is necessary in general.) It turns out that the condition they assumed is actually equivalent to d being an interior data point (Lemma 3.4). We now turn to this circle of ideas.

For the remainder of this section, we assume that Y is finite-dimensional and A is surjective. Thus, we may assume $Y = \mathbf{R}^n$ and

$$Ax := (\langle x, x_1\rangle, \langle x, x_2\rangle, \ldots, \langle x, x_n\rangle), \qquad x \in X,$$

for some linearly independent set $\{x_1, x_2, \ldots, x_n\}$ in X. Note that the adjoint mapping $A^*: \mathbf{R}^n \to X$ is then given by

$$A^*\lambda = \sum_1^n \lambda_i x_i, \qquad \lambda = (\lambda_1, \lambda_2, \ldots, \lambda_n) \in \mathbf{R}^n,$$

and the range of A^* is the subspace $\mathscr{R}(A^*) = \mathrm{span}\{x_1, x_2, \ldots, x_n\}$.

Clearly, $K = C \cap V \neq \varnothing$ if and only if $d \in AC$. We should mention that the data cone AC is not closed in general, even if C is closed and the range space is finite-dimensional.

Example 3.1. Let $X = L_2[0, 1]$ and $C = \{x \in X \mid x \geq 0\}$. We construct functions $x_1, x_2 \in X$ so that if $A: X \to \mathbf{R}^2$ is given by

$$Ax := (\langle x, x_1\rangle, \langle x, x_2\rangle), \qquad x \in X,$$

then the data cone AC is not closed.

For any set $S \subset [0, 1]$, let χ_S denote its characteristic function. Set $x_1 = \chi_{[0,1]}$ and $x_2 = \sum_1^\infty n\chi_{[2^{-n}, 2^{-n+1}]}$. Then $\|x_1\| = 1$ and

$$\|x_2\|^2 = \sum_1^\infty \int_{2^{-n+1}}^{2^{-n}} n^2\, dt = \sum_1^\infty n^2 2^{-n} < \infty.$$

Further, for any $x \in C$,

$$\langle x, x_1\rangle = \int_0^1 x(t)\, dt = \sum_1^\infty \int_{2^{-n+1}}^{2^{-n}} x(t)\, dt = \sum_1^\infty \alpha_n,$$

where

$$\alpha_n := \int_{2^{-n+1}}^{2^{-n}} x(t)\,dt \geq 0.$$

Also,

$$\langle x, x_2\rangle = \int_0^1 x(t) \sum_1^\infty n\chi_{[2^{-n},2^{-n+1}]}(t)\,dt = \sum_1^\infty n \int_{2^{-n+1}}^{2^{-n}} x(t)\,dt = \sum_1^\infty n\alpha_n.$$

Thus,

$$AC = \left\{ \left(\sum_1^\infty \alpha_n, \sum_1^\infty n\alpha_n \right) \middle| \alpha_n = \int_{2^{-n+1}}^{2^{-n}} x(t)\,dt, x \geq 0 \right\}$$
$$= \left\{ \sum_1^\infty n\alpha_n \left(\frac{1}{n}, 1 \right) \middle| \alpha_n = \int_{2^{-n+1}}^{2^{-n}} x(t)\,dt, x \geq 0 \right\}.$$

Hence, we may observe that $(1/n, 1) \in AC$ for every n, but $(0, 1) \notin AC$. This shows that $(0, 1) \in \overline{AC} \backslash AC$ and AC is not closed.

The next two lemmas characterize when a point is in the closure of the data cone and the interior of the data cone, respectively.

Lemma 3.2. *Let $d \in \mathbf{R}^n$. Then $d \in \overline{AC}$ if and only if $\sum_1^n \lambda_i d_i \leq 0$ whenever $\sum_1^n \lambda_i x_i \in C^0$. In particular, if $d \in AC$, then $\sum_1^n \lambda_i d_i \leq 0$ whenever $\sum_1^n \lambda_i x_i \in C^0$.*

Proof. Since $\overline{AC}$ is a closed convex cone in $\mathbf{R}^n$, it follows by the separation theorem that $d \notin \overline{AC}$ if and only if there exists $\lambda \in \mathbf{R}^n$ so that

$$\sup_{x \in C} \langle \lambda, Ax\rangle = 0 < \langle \lambda, d\rangle. \tag{3.5}$$

But (3.5) holds if and only if

$$\sum_1^n \lambda_i \langle x, x_i\rangle \leq 0 < \sum_1^n \lambda_i d_i$$

for every $x \in C$, or, equivalently,

$$\left\langle x, \sum_1^n \lambda_i x_i \right\rangle \leq 0 < \sum_1^n \lambda_i d_i$$

for every $x \in C$. This may be formulated as

$$\sum_1^n \lambda_i x_i \in C^0 \quad \text{and} \quad \sum_1^n \lambda_i d_i > 0. \tag{3.6}$$

Thus, $d \notin \overline{AC}$ if and only if there exists $\lambda \in \mathbf{R}^n$ so that (3.6) holds. From this the result follows. ∎

Lemma 3.3. *Let $d \in AC$. Then $d \in \operatorname{int} AC$ if and only if $\sum_1 \lambda_i x_i = 0$ whenever $\sum_1^n \lambda_i x_i \in C^0$ and $\sum_1^n \lambda_i d_i = 0$.*

Proof. (i) Suppose $\sum_1^n \lambda_i x_i \in C^0$ and $\sum_1^n \lambda_i d_i = 0$, but $\sum_1^n \lambda_i x_i \neq 0$. Since $\lambda \neq 0$, there exist $y^{(k)} = (y_1^{(k)}, \ldots, y_n^{(k)}) \in \mathbf{R}^n$ so that $\|y^{(k)} - d\| \to 0$ and $\sum_1^n \lambda_i y_i^{(k)} > 0$ for each k. By Lemma 3.2, $y^{(k)} \notin \overline{AC}$. Since $y^{(k)} \to d$, $d \notin \operatorname{int} AC$.

(ii) Suppose $d \notin \operatorname{int} AC$. It suffices to show there exists $\sum_1^n \lambda_i x_i \in C^0 \backslash \{0\}$ with $\sum_1^n \lambda_i d_i = 0$. If not, then, for all $\sum_1^n \lambda_i x_i \in C^0 \backslash \{0\}$, we must have, by Lemma 3.2, that $\sum_1^n \lambda_i d_i < 0$. By scaling, we deduce that $\sum_1^n \lambda_i d_i < 0$ for all $\lambda \in \mathbf{R}^n$ with $\sum_1^n \lambda_i x_i \in C^0$ and $\|\sum_1^n \lambda_i x_i\| = 1$. Since the set

$$\Lambda := \left\{ \lambda \in \mathbf{R}^n \,\middle|\, \sum_1^n \lambda_i x_i \in C^0,\ \left\| \sum_1^n \lambda_i x_i \right\| = 1 \right\}$$

is closed and bounded, hence compact, and the mapping $\lambda \in \Lambda \to \sum_1^n \lambda_i d_i \in \mathbf{R}$ is continuous and negative on Λ, there exists $\delta > 0$ so that $\sum_1^n \lambda_i d_i \leq -\delta$ for all $\lambda \in \Lambda$.

Hence, there is $\varepsilon > 0$ so that, for every $y \in B(d, \varepsilon)$, $\sum_1^n \lambda_i y_i \leq -\delta/2 < 0$ for all $\lambda \in \Lambda$. This proves that $\sum_1^n \lambda_i y_i < 0$ whenever $\sum_1^n \lambda_i x_i \in C^0$, $\|\sum_1^n \lambda_i x_i\| = 1$, and $y \in B(d, \varepsilon)$. Again by scaling, $\sum_1^n \lambda_i y_i < 0$ for every $y \in B(d, \varepsilon)$ whenever $\sum_1^n \lambda_i x_i \in C^0 \backslash \{0\}$. By Lemma 3.2, $y \in \overline{AC}$ for all $y \in B(d, \varepsilon)$. That is, $d \in \operatorname{int} \overline{AC}$. Since AC is a finite-dimensional convex set, it follows that $\operatorname{int} \overline{AC} = \operatorname{int} AC$ [20]. Thus, $d \in \operatorname{int} AC$ which contradicts the hypothesis. ∎

The following condition of Micchelli and Utreras [17] was an essential hypothesis to their main results:

(MU) $$\{A^* y \mid \langle y, d\rangle \geq 0\} \cap C^0 = \{0\}.$$

We now show that (among other things) the (MU)-condition is equivalent to the data point d being an *interior* data point.

Lemma 3.4. *Let $d \in \mathbf{R}^n$ and assume that $K = C \cap V \neq \varnothing$. Then the following statements are equivalent*:

(1) *the* (MU)*-condition is satisfied*;
(2) $\{A^* y \mid \langle y, d\rangle = 0\} \cap C^0 = \{0\}$;
(3) $\mathcal{R}(A^*) \cap k^\perp \cap C^0 = \{0\}$ *for each* $k \in K$;
(4) $\mathcal{R}(A^*) \cap K^\perp \cap C^0 = \{0\}$;
(5) $\{\sum_1^n \lambda_i x_i \mid \sum_1^n \lambda_i x_i \in C^0, \sum_1^n \lambda_i d_i = 0\} = \{0\}$;
(6) *d is a strong interior data point*;
(7) *d is an interior data point, i.e.,* $d \in \operatorname{int} AC$.

Moreover, if any of the above conditions holds, then

$$C^0 \cap k^\perp + \mathcal{R}(A^*)$$

is closed for each $k \in K$.

Proof. By Lemmas 3.1 and 3.3 we see that (5), (6), and (7) are equivalent. Now consider the sets

$$S_1 = \{A^* y \mid \langle y, d\rangle \geq 0\} \cap C^0,$$
$$S_2 = \{A^* y \mid \langle y, d\rangle = 0\} \cap C^0,$$
$$S_3 = \mathcal{R}(A^*) \cap K^\perp \cap C^0,$$
$$S_4 = S_4(k) = \mathcal{R}(A^*) \cap k^\perp \cap C^0, \qquad k \in K,$$

and

$$S_5=\left\{\sum_1^n \lambda_i x_i \,\middle|\, \sum_1^n \lambda_i x_i \in C^0, \sum_1^n \lambda_i d_i = 0\right\}.$$

To complete the proof of the equivalence of statements (1)-(7), it suffices to show that all the sets $S_1, \ldots, S_5$ are equal.

Let $x \in S_1$. Then $x \in C^0$ and $x = A^*y$ for some $y \in \mathbf{R}^n$ with $\langle y, d\rangle \geq 0$. For each $k \in C \cap V$,

$$0 \leq \langle y, d\rangle = \langle y, Ak\rangle = \langle A^*y, k\rangle = \langle x, k\rangle \leq 0$$

so that $\langle y, d\rangle = 0$ and $x \in S_2$.

If $x \in S_2$, then $x \in C^0$ and $x = A^*y \in \mathbf{R}(A^*)$ for some $y \in d^\perp$. Hence, for any $k \in K$,

$$\langle x, k\rangle = \langle A^*y, k\rangle = \langle y, Ak\rangle = \langle y, d\rangle = 0$$

and this yields $x \in K^\perp$ and $x \in S_3$.

Since $K^\perp \subset k^\perp$ for each $k \in K$, it follows that $S_3 \subset S_4(k)$.

If $x \in S_4(k)$ for some $k \in K$, then $x \in k^\perp \cap C^0$ and $x = A^*\lambda$ for some $\lambda \in \mathbf{R}^n$. Thus, $x = \sum_1^n \lambda_i x_i$. Also,

$$\sum_1^n \lambda_i d_i = \langle \lambda, d\rangle = \langle \lambda, Ak\rangle = \langle A^*\lambda, k\rangle = \langle x, k\rangle = 0.$$

Thus, $x \in S_5$.

Finally, let $x \in S_5$. Then $x = \sum_1^n y_i x_i$ for some $y \in \mathbf{R}^n$, $x \in C^0$, and $\sum_1^n y_i d_i = 0$. Hence $x = A^*y$ and $\langle y, d\rangle = 0$. Thus, $x \in S_1$.

We have shown that

$$S_1 \subset S_2 \subset S_3 \subset S_4(k) \subset S_5 \subset S_1$$

for any $k \in K$. Thus, all these sets are equal, or (1)-(7) are equivalent.

The proof of the last statement of the lemma follows from the fact which derives immediately from the "Dieudonne separation theorem" [10, p. 105].

Fact. *If D is a closed convex cone and M is a finite-dimensional subspace in X such that $D \cap M = \{0\}$, then $D + M$ is closed.*

To prove the last statement of the lemma, take $D = C^0 \cap k^\perp$ and $M = \mathcal{R}(A^*)$, and note that indeed $D \cap M = \{0\}$ by (3). ∎

During the course of the proof of the above lemma we have verified the following curious fact which will be needed in Section 4.

Lemma 3.5. *For each $k \in K$, $C^0 \cap k^\perp \cap \mathcal{R}(A^*) = C^0 \cap K^\perp \cap \mathcal{R}(A^*)$.*

Also, we should mention that each of the statements in Lemma 3.4 is equivalent to the statement

(8) $X = \overline{\operatorname{con}(C-k) + \mathcal{N}(A)}$ for each $k \in K$.

This follows, for example, by taking dual cones of both sides of the equality in (3) and (8). Statement 8 is closely related to (2.6b) of Proposition 2.1 of [4].

Now we can easily prove a variant of Theorem 3.1 which will be useful to us throughout the sequel.

Theorem 3.2. *Let C be a closed convex cone in the Hilbert space X, let Y be a finite-dimensional Hilbert space, $d \in Y$, $A \in \mathscr{B}(X, Y)$, $K = \{x \in C | Ax = d\}$, and assume $\{C, V\}$ has CHIP and, for each $k \in K$, $C^0 \cap k^\perp + \mathscr{R}(A^*)$ is closed. Then, for any $x \in X$ and $k_0 \in K$, the following statements are equivalent:*

(1) $k_0 = P_K(x)$;
(2) $x - k_0 \in C^0 \cap k_0^\perp + \mathscr{R}(A^*)$;
(3) $k_0 = P_C(x + A^*y)$ *for some* $y \in Y$.

Note that the hypothesis in the above theorem is satisfied if, for instance, d is an interior data point or any of the six equivalent conditions in Lemma 3.4 holds.

Proof. The equivalence of (1) and (2) follows from Theorem 3.1 using Lemmas 3.1 and 3.4. By Proposition 2.1, (3) holds if and only if $x + A^*y - k_0 \in (C - k_0)^0$ for some $y \in Y$. But, by Lemma 2.3, $(C - k_0)^0 = C^0 \cap k_0^\perp$ and the equivalence of (2) and (3) follows. ∎

Using variational methods, Micchelli and Utreras [17] proved the equivalence of (1) and (3) in Theorem 3.2. However, as we shall see in the next section, Theorem 3.1 is valid even in certain cases where Theorem 3.2 does not apply.

4. Best Positive Constrained Interpolation

In this section we apply the theory of the preceding section to characterize best approximations in $L_2(\mu)$ from the set of nonnegative functions which also lie in a finite-codimensional linear variety. The main result is Theorem 4.2.

Let μ be a measure on a σ-algebra of subsets of a set T. As usual, $L_2(\mu)$ denotes the Hilbert space of all (real) measurable functions x on T with

$$\|x\| := \left[\int_T |x|^2 \, d\mu \right]^{1/2} < \infty.$$

Then $L_2(\mu)$ is a Hilbert space with the inner product

$$\langle x, y \rangle = \int_T xy \, d\mu.$$

Here, and in the following, we do not distinguish between two functions which agree almost everywhere. Also, all sets are defined only up to a set of measure zero. For any $x \in L_2(\mu)$, the support of x is defined by

$$\operatorname{supp} x := \{t \in T | x(t) \neq 0\},$$

and the positive part of x is the function

$$x_+(t) = \max\{x(t), 0\}.$$

Let C denote the cone of nonnegative functions in $L_2(\mu)$, let $\{x_1, x_2, \ldots, x_N\}$ be a finite set in $L_2(\mu)$, $d = (d_1, d_2, \ldots, d_N) \in \mathbf{R}^N$,

$$V := \{x \in L_2(\mu) | \langle x, x_i \rangle = d_i (i = 1, 2, \ldots, N)\},$$

and

$$K := C \cap V = \{x \in L_2(\mu) | x \geq 0, \langle x, x_i \rangle = d_i (i = 1, 2, \ldots, N)\}.$$

The problem then is to determine the best approximation $P_K(x)$ of a given $x \in L_2(\mu)$ from K.

First we note that since K is closed and convex, best approximations always exist. Also, if we define A on $L_2(\mu)$ by

$$Ax := (\langle x, x_1 \rangle, \langle x, x_2 \rangle, \ldots, \langle x, x_N \rangle),$$

it follows that $A \in \mathscr{B}(L_2(\mu), \mathbf{R}^N)$ and the adjoint map $A^*: \mathbf{R}^N \to L_2(\mu)$ is given by

$$A^*\lambda = \sum_1^N \lambda_i x_i$$

so $\mathscr{R}(A^*) = \text{span}\{x_1, x_2, \ldots, x_N\}$. Thus,

$$K = \{x \in L_2(\mu) | x \geq 0 \text{ and } Ax = d\}.$$

Our problem is now in the same form as that in Section 3.

To apply Theorem 3.1, first we need to prove the following.

Theorem 4.1. *$\{C, V\}$ has property CHIP.*

Proof. By Lemma 2.5, it suffices to show that, for each $k \in K$, $\text{con}(C-k) \cap M$ is dense in $\overline{\text{con}}(C-k) \cap M$, where

$$M := V - k = \{x \in L_2(\mu) | \langle x, x_i \rangle = 0 \ (i = 1, 2, \ldots, N)\}.$$

Consider

$$E_0 = \{t \in T | k(t) = 0\}, \qquad E_1 = \{t \in T | k(t) > 1\},$$

and, for $n = 2, 3, \ldots,$

$$E_n = \{t \in T | n^{-1} < k(t) \leq (n-1)^{-1}\}.$$

Then $\{E_n | n \geq 0\}$ is a disjoint sequence of measurable sets whose union is T.

Claim 1. $\text{con}(C-k) \supset U$, where

$$U := \{x \in L_2(\mu) | x = y\chi_{E_0} + \sum_1^n z_j \chi_{E_j}, y \in L_2(\mu), y\chi_{E_0} \geq 0, z_j \in L_\infty(\mu), n < \infty\}.$$

To verify this, let $y \in L_2(\mu)$, $y\chi_{E_0} \geq 0$, and $c = y\chi_{E_0} + k$. Then $c \in C$ and

$$y\chi_{E_0} = c - k \in (C-k) \subset \text{con}(C-k).$$

Next fix any integer $j \geq 1$, and consider $z_j \in L_\infty(\mu)$ and $z = z_j \chi_{E_j}$. Then

$$|z| \leq |z_j| \leq \|z_j\|_\infty < \infty.$$

Also, for $t \in E_j$, we have that $k(t) > 1/j$. Thus, for any $\rho > 0$ with $\rho\|z_j\|_\infty < 1/j$, we see that, for $t \in E_j$,

$$k(t) + \rho z(t) \geq \frac{1}{j} - \rho\|z\|_\infty > 0$$

and, for $t \notin E_j$,

$$k(t) + \rho z(t) = k(t) \geq 0.$$

Thus, $k + \rho z \in C$ and

$$z_j \chi_{E_j} = z = \frac{1}{\rho}(k + \rho z - k) \in \operatorname{con}(C - k).$$

Since $\operatorname{con}(C-k)$ is a cone, sums of functions of the type $y\chi_{E_0}$ and $z_j\chi_{E_j}$ are also in $\operatorname{con}(C-k)$. This proves the first claim.

Claim 2. Let

$$W := \{x \in L_2(\mu) | x\chi_{E_0} \geq 0\}.$$

Then, $\overline{\operatorname{con}}(C - k) = W$.

To prove this, let $x \in \operatorname{con}(C-k)$. Then $x = \rho(c - k)$ for some $\rho > 0$ and $c \in C$. Thus,

$$x\chi_{E_0} = \rho c \chi_{E_0} \geq 0$$

and $x \in W$. Since W is closed, $\overline{\operatorname{con}}(C-k) \subset W$. If there exists an $x_0 \in W \backslash \overline{\operatorname{con}}(C-k)$, then by the separation theorem there exists $z \in L_2(\mu)$ so that

$$\sup\{\langle z, y\rangle | y \in \overline{\operatorname{con}}(C-k)\} = 0 < \langle z, x_0\rangle.$$

By (2) in Lemma 2.3 and (3) in Lemma 2.2, we have $z \in C^0 \cap k^\perp$. Thus, $z \leq 0$ and $\langle z, k\rangle = 0$. Since $k \geq 0$, it follows that $zk = 0$ so that $z = 0$ on E_0^c, the complement of E_0. Hence,

$$0 < \langle z, x\rangle = \int_{E_0} zx_0 \, d\mu + \int_{E_0^c} zx_0 \, d\mu = \int_{E_0} zx_0 \, d\mu \leq 0$$

which is absurd. Thus, $W \subset \overline{\operatorname{con}}(C-k)$ and Claim 2 is proved.

To prove the proposition, it suffices by Claims 1 and 2 to verify that $U \cap M$ is dense in $W \cap M$. Fix any $x_0 \in W \cap M$. Then $x_0\chi_{E_0} \geq 0$ and $\langle x_0, x_i\rangle = 0$ $(i = 1, 2, \ldots, N)$. Setting

$$\beta_i = \int_T x_0 \chi_{E_0} x_i \, d\mu \qquad (i = 1, 2, \ldots, N),$$

we see that $\int_T x_0 \chi_{E_0^c} x_i \, d\mu = -\beta_i$ for all i.

In the space

$$L_2(E_0^c) := \{x \in L_2(\mu) | x\chi_{E_0} = 0\},$$

we see that the convex set

$$U_0=\left\{\sum_1^n g_j\chi_{E_j} \mid g_j\in L_\infty(\mu),\ n<\infty\right\}$$
$$=\{x\in U \mid x\chi_{E_0}=0\}$$

contains all simple functions in $L_2(E_0^c)$ and thus is dense in $L_2(E_0^c)$. Further, the set

$$V_0:=\{x\in L_2(E_0^c) \mid \langle x, x_i\rangle=-\beta_i\ (i=1,2,\dots,N)\}$$

is a linear variety of finite-codimension in $L_2(E_0^c)$. It follows that $U_0\cap V_0$ is dense in V_0. Further, $x_0\chi_{E_0^c}\in V_0$. Hence, given any $\varepsilon>0$, choose $u_0\in U_0\cap V_0$ such that $\|u_0-x_0\chi_{E_0^c}\|<\varepsilon$. But since $u_0=\sum_1^n z_j\chi_{E_j}$ for some $z_j\in L_\infty(\mu)$, setting $u=x_0\chi_{E_0}+u_0$, we see that $u\in U$,

$$\langle u, x_i\rangle=\int_T x_0\chi_{E_0}x_i\,d\mu+\int_T u_0x_i\,d\mu=\beta_i-\beta_i=0$$

$(i=1,\dots,N)$, so that $u\in M$, and

$$\|u-x_0\|=\|u_0-x_0\chi_{E_0^c}\|<\varepsilon.$$

This proves that $U\cap M$ is dense in $W\cap M$. ∎

Lemma 4.1. *Let $x\in L_2(\mu)$ and $k_0\in C$. Then*:

(1) $C^0\cap k_0^\perp=\{u\in L_2(\mu) \mid u\le 0,\ u=0 \text{ on supp } k_0\}$.
(2) *The following statements are equivalent*:
  (i) $k_0=P_C(x)$;
  (ii) $x-k_0\in C^0\cap k_0^\perp$;
  (iii) $k_0=x_+$.

Proof. (1) Let $u\in C^0\cap k_0^\perp$. Then $u\le 0$ and $\langle u, k_0\rangle=0$. Since $uk_0\le 0$, it follows that $uk_0=0$ so $u=0$ on supp k_0. Conversely, if $u\le 0$ and $u=0$ on supp k_0, then $u\in C^0$ and $\langle u, k_0\rangle=0$.

(2) Using Theorem 2.1, and statements (1) and (2) in Lemma 2.3, we see that $k_0=P_C(x)$ if and only if $x-k_0\in C^0\cap k_0^\perp$, or, equivalently, $x-k_0\le 0$ and $x-k_0=0$ on supp k_0, that is, $k_0=x_+$. ∎

Before stating the main result of this section, it is convenient to isolate a special case that will simplify the proof of this result as well as some subsequent ones.

Lemma 4.2. *Let K be as defined above and define*

(4.1) $$\Omega:=\bigcup_{k\in K}\text{supp } k=\{t\in T \mid k(t)>0 \textit{ for some } k\in K\}.$$

If $x_iX_\Omega=0$ for $i=1,2,\dots,N$, then for each $x\in L_2(\mu)$,

(4.2) $$P_K(x)=x_+\chi_\Omega.$$

Proof. Since $x_i = 0$ on Ω for each i, we have $\langle k, x_i\rangle = 0$ for each $k \in K$. That is,

$$K = \{y \in L_2(\mu) | y \geq 0, \langle y, x_i\rangle = 0 \ (i = 1, 2, \ldots, N)\}.$$

It follows that K contains every nonnegative function whose support is in Ω. In particular, $k_0 := x_+\chi_\Omega$ is in K. For any $k \in K$, $k = k\chi_\Omega \geq 0$ and so

$$\begin{aligned}\|x - k_0\|^2 &= \int_\Omega |x - x_+|^2 \, d\mu + \int_{T\setminus\Omega} |x|^2 \, d\mu \\ &\leq \int_\Omega |x - k|^2 \, d\mu + \int_{T\setminus\Omega} |x|^2 \, d\mu = \|x - k\|^2.\end{aligned}$$

That is, $P_K(x) = k_0$. ■

Now we come to the main characterization of best approximations in this section. It states that the best approximation to x from K is the positive part of the sum of x and an element in $\operatorname{span}\{x_1, x_2, \ldots, x_N\}$ multiplied by the characteristic function of an explicitly prescribed subset of T.

Theorem 4.2. *Let* $\{x_1, x_2, \ldots, x_N\} \subset L_2(\mu)$, $d = (d_1, d_2, \ldots, d_N) \in \mathbf{R}^N$, *and assume that the set*

$$K := \{y \in L_2(\mu) | y \geq 0, \langle y, x_i\rangle = d_i \ (i = 1, 2, \ldots, N)\}$$

is not empty. Define

$$\Omega = \{t \in T | k(t) > 0 \text{ for some } k \in K\}$$

and set $\Gamma = T$ *if* $\{x_1, x_2, \ldots, x_N\}$ *is linearly independent over* Ω *and* $\Gamma = \Omega$ *otherwise. Then, for each* $x \in L_2(\mu)$,

$$P_K(x) = \left(x + \sum_1^N \alpha_i x_i\right)_+ \chi_\Gamma \tag{4.3}$$

for any set of scalars α_i *chosen to satisfy the equations*

$$\left\langle \left(x + \sum_1^N \alpha_i x_i\right)_+ \chi_\Gamma, x_j \right\rangle = d_j \qquad (j = 1, \ldots, N). \tag{4.4}$$

Proof. Equations (4.4) guarantee that the element in (4.3) lies in K. Also, by replacing $\{x_1, x_2, \ldots, x_N\}$ by a maximal linearly independent subset, we may assume that $\{x_1, x_2, \ldots, x_N\}$ is linearly independent. Let $M = \operatorname{span}\{x_1, x_2, \ldots, x_N\}$. We consider two cases.

Case 1. $\{x_1, x_2, \ldots, x_N\}$ is linearly independent over Ω.

We first show that $C^0 \cap k^\perp + M$ is closed for each $k \in K$. By Lemma 3.4, it suffices to show that $M \cap C^0 \cap K^\perp = \{0\}$. Let $\sum_1^N \alpha_i x_i \in M \cap C^0 \cap K^\perp$. Then $\sum_1^N \alpha_i x_i \leq 0$ and $\langle \sum_1^N \alpha_i x_i, k\rangle = 0$ for every $k \in K$. Hence $(\sum_1^N \alpha_i x_i)k = 0$ a.e. for each $k \in K$ implies that $\operatorname{supp}(\sum_1^N \alpha_i x_i) \cap \operatorname{supp} k = \varnothing$ for each $k \in K$. It follows that $\operatorname{supp}(\sum_1^N \alpha_i x_i) \cap \Omega = \varnothing$ so $\sum_1^N \alpha_i x_i = 0$ on Ω. By the hypothesis, $\alpha_1 = \alpha_2 = \cdots = \alpha_N = 0$.

This proves $C^0 \cap k^\perp + M$ is closed. By Theorem 3.2, for any $x \in L_2(\mu)$,

$$P_K(x) = P_C\left(x + \sum_1^N \alpha_i x_i\right)$$

for some scalars α_i. By Lemma 4.1,

$$P_K(x) = \left(x + \sum_1^N \alpha_i x_i\right)_+.$$

Case 2. $\{x_1, x_2, \ldots, x_N\}$ is linearly dependent over Ω.

If $x_i\chi_\Omega = 0$ for each i, then $P_K(x) = x_+\chi_\Omega$ by Lemma 4.2. Thus, we may assume $x_i\chi_\Omega \neq 0$ for some i and, by reindexing, that $\{x_1, x_2, \ldots, x_j\}$ is a maximal linearly independent subset over Ω, $1 \leq j < N$. Note that $\{x_1\chi_\Omega, x_2\chi_\Omega, \ldots, x_j\chi_\Omega\}$ is linearly independent over Ω,

$$K = \{y \in L_2(\mu) | y \geq 0, \langle y, x_i\chi_\Omega\rangle = d_i \ (i = 1, 2, \ldots, j)\},$$

and $P_K(x) = P_K(x\chi_\Omega)$ for every $x \in L_2(\mu)$. Applying Case 1 (to $x\chi_\Omega$ instead of x), we obtain

$$P_K(x) = P_K(x\chi_\Omega) = \left(x\chi_\Omega + \sum_1^j \alpha_i x_i \chi_\Omega\right)_+ = \left(x + \sum_1^j \alpha_i x_i\right)_+ \chi_\Omega$$

for some scalars α_i.

Remarks. (1) Micchelli, Smith, Swetits, and Ward [16] considered the problem of best positive constrained interpolation in any $L_p(\mu)$-space $(1 < p < \infty)$. In particular, using arguments motivated by Lagrange duality, they proved Theorem 4.2 under the additional assumptions that $L_2(\mu)$ is separable, $x = 0$, and $\{x_1, x_2, \ldots, x_N\}$ is independent.

(2) Theorem 4.2 suggests a possible method for computing best approximations. We first determine the set Ω and decide whether $\Gamma = T$ or $\Gamma = \Omega$. In order that the element

$$k_0 = \left(x + \sum_1^N \alpha_i x_i\right)_+ \chi_\Gamma$$

will be the best approximation to x from K, it is necessary and sufficient that the scalars α_i be chosen to satisfy the interpolation conditions

$$\langle k_0, x_j\rangle = d_j \qquad (j = 1, 2, \ldots, N).$$

These conditions represent N (nonlinear) equations for the N unknown scalars α_i.

An example in [16] shows that the characteristic function cannot be dropped as a factor in the best approximation. However, there is a class of examples where the characteristic function may be dispensed with as a factor in the best approximation. The next result gives a useful sufficient condition which allows this.

Theorem 4.3. *Let $\{x_1, \ldots, x_N\}$ be a finite set in $L_2(\mu)$, $(d_1, \ldots, d_N) \in R^N$, and assume*

$$K := \{y \in L_2(\mu) | y \geq 0, \langle y, x_i \rangle = d_i \ (i = 1, \ldots, N)\}$$

is not empty. If $C^0 \cap k^\perp + \operatorname{span}\{x_1, \ldots, x_N\}$ is closed for each $k \in K$, then, for any $x \in L_2(\mu)$,

$$P_K(x) = \left(x + \sum_1^N \alpha_i x_i\right)_+ \tag{4.5}$$

for any scalars α_i chosen to satisfy the equations

$$\left\langle \left(x + \sum_1^N \alpha_i x_i\right)_+, x_j \right\rangle = d_j \qquad (j = 1, \ldots, N). \tag{4.6}$$

Corollary 4.1. *Let $\{x_1, x_2, \ldots, x_N\}$ be a finite set in $L_2(\mu)$, $d = (d_1, d_2, \ldots, d_N) \in \mathbf{R}^N$, and $K = \{y \in L_2(y) | y \geq 0, \langle y, x_i \rangle = d_i \ (i = 1, 2, \ldots, N)\}$. Assume that d is an interior data point or, equivalently, any of the other six statements of Lemma 3.4 holds. Then, for each $x \in L_2(\mu)$, $P_K(x) = (x + \sum_1^N \alpha_i x_i)_+$ for any $\alpha_i \in \mathbf{R}$ chosen to satisfy equations (4.6) of Theorem 4.3.*

Easy examples show that the interior data point condition is not necessary to obtain best approximations which are the positive parts of functions (without characteristic functions as factors).

The following result is an easy consequence of Theorem 4.3.

Corollary 4.2. *If $x_1, \ldots, x_N$ are nonnegative continuous piecewise linear functions on $[a, b]$ and $d_1, \ldots, d_N$ are positive numbers, and*

$$K = \{y \in L_2[a, b] | y \geq 0, \langle y, x_i \rangle = d_i \ (i = 1, \ldots, N)\} \neq \varnothing,$$

then for any $x \in L_2[a, b]$,

$$P_K(x) = \left(x + \sum_{i=1}^N \alpha_i x_i\right)_+$$

for any $\alpha_1, \ldots, \alpha_N \in \mathbf{R}$, chosen to satisfy equations (4.6) of Theorem 4.3.

A nice application of Theorem 4.2 is that of constrained spline interpolation. The reader should see [16], along with [10], [11], [12], [19], and [1], for more details.

As a final application of Theorem 4.3, we recover and extend a result of Smith and Wolkowicz [22].

Theorem 4.4. *Let A be an $m \times n$ real matrix, $d \in l_2(m)$, and*

$$K = \{x \in l_2(n) | x \geq 0, Ax = d\} \neq \varnothing.$$

Then, for each $x \in l_2(n)$,

$$P_K(x) = (x + A^T \lambda)_+$$

for any $\lambda \in l_2(m)$ *chosen so that* $A[(x+A^{\mathrm{T}}\lambda)_+]=d$. *(Here* A^{T} *denotes the transpose of* A.)

Proof. Letting x_i denote the ith row of A, we see that $x_i \in l_2(n)$ and A is the linear operator from $l_2(n)$ to $l_2(m)$ defined by $Ax=(\langle x, x_1\rangle, \langle x, x_2\rangle, \ldots, \langle x, x_m\rangle)^{\mathrm{T}}$. Also, A^{T} may be identified with the adjoint of A:

$$A^{\mathrm{T}}\alpha = A^*\alpha = \sum_1^m \alpha_i x_i, \qquad \alpha \in l_2(m).$$

The result now follows from Theorem 4.3 if we show that

$$C^0 \cap k^{\perp} + \operatorname{span}\{x_1, x_2, \ldots, x_m\}$$

is closed for each $k \in K$, where $C=\{x \in l_2(n) | x \geq 0\}$. That this set is closed follows immediately from the next lemma. ■

Lemma 4.3. *Let* $\{y_1, y_2, \ldots, y_N\}$ *be a linearly independent set in* X, *let* E *be the convex cone*

$$E=\left\{\sum_1^N \rho_i y_i \middle| \rho_i \geq 0 \text{ for each } i\right\},$$

and let L *and* M *be finite-dimensional subspaces of* X. *Then* $(E \cap L)+M$ *is closed.*

Proof. Assume first that $L=X$. We claim that, for each $x \in E+M$, there are vectors $e \in E$ and $m \in M$ so that $x=e+m$, where

$$\|e\|+\|m\|=\inf\{\|e'\|+\|m'\| | x=e'+m', e' \in E, m' \in M\}=: r(x).$$

To see this, choose $e_n \in E$, $m_n \in M$ so that $x=e_n+m_n$ and

$$\|e_n\|+\|m_n\|<r(x)+\frac{1}{n}, \qquad n \geq 1.$$

This proves that the sequences $\{e_n\}$ and $\{m_n\}$ are bounded. Since each of these sequences lies in a finite-dimensional subspace, by passing to a subsequence, we may assume that $m_n \to m \in M$ and $e_n \to e \in E$ (since M and E are closed). Thus, $x=\lim(e_n+m_n)=e+m$ and $\|e\|+\|m\|=r(x)$.

Now assume $x_n \in E+M$ and $x_n \to x$. We need to show that $x \in E+M$. Choose $e_n \in E$ and $m_n \in M$ so that $x_n=e_n+m_n$ and $\|e_n\|+\|m_n\|=r(x_n)$. If either $\{e_n\}$ or $\{m_n\}$ has a bounded subsequence, say $\{e_n\}$, then by passing to a subsequence we may assume $e_n \to e \in E$. Thus,

$$m_n = x_n - e_n \to x-e =: m \in M$$

and $x=e+m \in E+M$.

The other possibility is that $\|e_n\| \to \infty$ and $\|m_n\| \to \infty$. But we show that this cannot happen. By passing to a subsequence, we may assume that

$$\frac{m_n}{\|m_n\|} \to m \in M, \qquad \|m\|=1.$$

Then

$$\frac{e_n}{\|m_n\|}=\frac{x_n}{\|m_n\|}-\frac{m_n}{\|m_n\|}\to 0-m=: e\in E.$$

By reindexing the vectors y_i, we may assume that

$$e=\sum_1^N \rho_i y_i,$$

where $\rho_i>0$ for $i\le k$, and $\rho_i=0$ for $i>k$.

Also, we can write

$$\frac{e_n}{\|m_n\|}=\sum_{i=1}^N \rho_{ni}y_i \qquad \text{for some} \quad \rho_{ni}\ge 0,$$

where

$$\rho_{ni}\to\rho_i \qquad (i=1,2,\ldots,N).$$

In particular, since $\rho_i>0$ for $i\le k$, there exists an integer n_0 so that

$$\rho_{ni}>\tfrac{1}{2}\rho_i \qquad (i=1,2,\ldots,k)$$

for all $n\ge n_0$. Hence

$$\frac{e_n}{\|m_n\|}-\tfrac{1}{2}e=\sum_{i=1}^N (\rho_{ni}-\tfrac{1}{2}\rho_i)y_i\in E$$

for $n\ge n_0$.

Since $e=-m\in M$, we have that

$$\frac{m_n}{\|m_n\|}+\tfrac{1}{2}e\in M.$$

It follows that

$$\frac{m_n}{\|m_n\|}+\tfrac{1}{2}e\to m+\tfrac{1}{2}e=-\tfrac{1}{2}e$$

and

$$\frac{e_n}{\|m_n\|}-\tfrac{1}{2}e\to e-\tfrac{1}{2}e=\tfrac{1}{2}e.$$

Hence we see that

$$\text{(4.7)} \qquad \begin{aligned} x_n&=\|m_n\|\left(\frac{e_n}{\|m_n\|}-\frac{e}{2}\right)+\|m_n\|\left(\frac{m_n}{\|m_n\|}+\frac{e}{2}\right)\\ &=e_n'+m_n',\end{aligned}$$

where

$$e_n'=\|m_n\|\left(\frac{e_n}{\|m_n\|}-\frac{e}{2}\right)\in E \qquad \text{for} \quad n\ge n_0$$

and

$$m_n' = \|m_n\|\left(\frac{m_n}{\|m_n\|} + \frac{e}{2}\right) \in M.$$

From (4.7) and the choice of e_n and m_n, we conclude that, for $n \geq n_0$,

$$\|e_n\| + \|m_n\| \leq \|e_n'\| + \|m_n'\| = \|m_n\| \left\| \frac{e_n}{\|m_n\|} - \frac{e}{2} \right\| + \|m_n\| \left\| \frac{m_n}{\|m_n\|} + \frac{e}{2} \right\|$$

so

$$\frac{\|e_n\|}{\|m_n\|} + 1 \leq \left\| \frac{e_n}{\|m_n\|} - \frac{e}{2} \right\| + \left\| \frac{m_n}{\|m_n\|} + \frac{e}{2} \right\|.$$

Passing to the limit in this last expression yields

$$\|e\| + 1 \leq \tfrac{1}{2}\|e\| + \tfrac{1}{2}\|e\| = \|e\|$$

which is absurd. This proves that $E + M$ is closed.

But a moment's thought reveals that the *same* proof shows that $(E \cap L) + M$ is closed for any subspace L. ■

To see that Lemma 4.3 actually completes the proof of Theorem 4.4, note that

$$C^0 = \{x \in l_2(n) | x \leq 0\} = \left\{ \sum_1^n \rho_i y_i | \rho_i \geq 0 \right\},$$

where $y_i \in l_2(n)$ satisfy $y_i(j) = -\delta_{ij}$. Then take $E = C^0$, $L = k^\perp$, and $M = \text{span}\{x_1, x_2, \ldots, x_m\}$.

5. Infinite Interpolation

Much of the theory on constrained approximation has been devoted to the case of finitely many constraints. There is good reason for this. Many of the techniques employed to handle finite contraints fail badly in the infinite constraint situation.

While making no claims on handling *general* infinite constraint problems, there are many infinite-dimensional situations where the techniques herein apply. In this section we first wish to point out how a certain class of infinite-dimensional problems is derived from the approach of this paper. Then we give an application to interpolation by piecewise linear splines.

As usual, X denotes a Hilbert space. Let $\{x_i\}_1^\infty \subset X$ be an unconditional basis (see [7]) for its closed linear span $H := \overline{\text{span}\{x_1, x_2, \ldots\}}$, and let $A: X \to l_2$ be given by

$$Ax = (\langle x, x_1 \rangle, \langle x, x_2 \rangle, \ldots), \qquad x \in X,$$

so that $A^*: l_2 \to X$ is given by

$$A^*(\alpha_1, \alpha_2, \ldots) = \sum_1^\infty \alpha_j x_j.$$

Let C be a closed convex cone in X, and let $d=(d_1, d_2, \ldots)$ be an admissible data vector in the cone AC; that is, $d \in AC$. Also, let

$$K := \{x \in C \mid Ax = d\}$$
$$= \{x \in X \mid x \in C, \langle x, x_j\rangle = d_j \; (j=1,2,\ldots)\}.$$

The *infinite constrained interpolation problem* may then be stated as follows:

Characterize $P_K(x)$ for any $x \in X$.

First we define a sequence of finite-dimensional problems based on the infinite problem. For each $n=1,2,\ldots$ define $A_n\colon X \to l_2(n)$ by

$$A_n x := (\langle x, x_1\rangle, \langle x, x_2\rangle, \ldots, \langle x, x_n\rangle), \qquad x \in X,$$

so that $A_n^*\colon l_2(n) \to X$ is given by

$$A_n^* \lambda = \sum_1^n \lambda_i x_i, \qquad \lambda \in l_2(n).$$

For the given data vector $d \in l_2$, define

$$d^{(n)} := (d_1, d_2, \ldots, d_n)$$

and let

$$K_n := \{x \in C \mid A_n x = d^{(n)}\}$$
$$= \{x \in X \mid x \in C, \langle x, x_i\rangle = d_i \; (i=1,2,\ldots,n)\}.$$

We have the following result.

Theorem 5.1. *Assume that* $d^{(n)} \in \operatorname{int} A_n C$ *for each* n, $x \in X$, *and assume that* $\lambda_n = (\lambda_{n1}, \lambda_{n2}, \ldots, \lambda_{nn}) \in l_2(n)$ *is chosen so that*

$$P_{K_n}(x) = P_C(x + A_n^* \lambda_n) \qquad (n=1,2,\ldots). \tag{5.1}$$

If

$$\rho' := \sup_n \left(\sum_{j=1}^n \lambda_{nj}^2 \right) < \infty, \tag{5.2}$$

then

$$P_K(x) = P_C(x + A^* \lambda) \qquad \text{for some} \quad \lambda \in l_2. \tag{5.3}$$

Moreover,

$$P_K(x) = \lim_n P_{K_n}(x).$$

Proof. We first verify that the range of A^*, denoted by $\mathcal{R}(A^*)$, is closed. Now A^* has a closed range if (and only if) A^* is bounded below on the norm-one elements of $(\ker A^*)^{\perp}$. Clearly, $\ker A^* = \{0\}$. Since unconditional bases for a Hilbert space are unique, there is a constant $\rho > 0$ such that

$$\|A^*\lambda\|^2 = \left\|\sum_j \lambda_j x_j\right\|^2 \geq \frac{1}{\rho^2}\sum_j \lambda_j^2 = \frac{1}{\rho^2} \tag{5.4}$$

for all $\{\lambda_j\}$ with $\sum \lambda_j^2 = 1$. Hence, $\mathcal{R}(A^*)$ is closed.

Next, let $r_n = d(x, K_n)$ for $n = 1, 2, \dots$ and $r_\infty = d(x, K)$. Since $K \subset K_{n+1} \subset K_n$, we have $r_n \leq r_{n+1} \leq r_\infty$ and

$$\lim r_n \leq r_\infty < \infty.$$

By (5.1),

$$r_n = \|x - P_C(x + A_n^*\lambda_n)\| \qquad (n = 1, 2, \dots)$$

and hence $\{P_C(x + A_n^*\lambda_n)\}$ is bounded. By the Eberlein-Smulian theorem, we may pass to a subsequence and assume that

$$P_C(x + A_n^*\lambda_n) \overset{w}{\rightharpoonup} c \tag{5.5}$$

for some $c \in X$, where $\rightharpoonup^w$ denotes weak convergence. Clearly, $c \in K_n$ for every n so $c \in K$. Thus,

$$r_\infty \leq \|x - c\| \leq \liminf_n \|x - P_C(x + A_n^*\lambda_n)\| \leq r_\infty \tag{5.6}$$

implies that $\|x - c\| = r_\infty$ and $c = P_K(x)$.

To complete the proof, we verify that $c = P_C(x + A^*\lambda)$ for some $\lambda \in l_2$.

By (5.2) the sequence $\{A_n^*\lambda_n\}$ is bounded, so (again by the Eberlein-Smulian theorem), by passing to a subsequence, we may assume that

$$A_n^*\lambda_n \overset{w}{\rightharpoonup} z$$

for some $z \in X$. Since $A_n^*\lambda_n \in \mathcal{R}(A^*)$ for each n and $\mathcal{R}(A^*)$ is closed, hence weakly closed, it follows that $z = A^*\lambda$ for some $\lambda \in l_2$ and

$$A_n^*\lambda_n \overset{w}{\rightharpoonup} A^*y.$$

By (5.5) and (5.6) we see that

$$x - P_C(x + A_n^*\lambda_n) \overset{w}{\rightharpoonup} x - c$$

and

$$\|x - P_C(x + A_n^*\lambda_n)\| \to \|x - c\|.$$

This implies that

$$\|c - P_C(x + A_n^*\lambda_n)\| \to 0. \tag{5.7}$$

Using Proposition 2.1 and (2) of Lemma 2.3, we deduce that, for an element $y_0 \in C$, $y_0 = P_C(x)$ if and only if $z - y_0 \in C^0 \cap y_0^\perp$. Using this, we see that, for each $y \in C$,

$$\langle x + A^*\lambda - c, y\rangle = \lim\langle x + A_n^*\lambda_n - P_C(x + A_n^*\lambda_n), y\rangle \le 0$$

and

$$\begin{aligned}\langle x + A^*\lambda - c, c\rangle &= \lim\langle x + A_n^*\lambda_n - P_C(x + A_n^*\lambda_n), c\rangle \\ &= \lim\langle x + A_n^*\lambda_n - P_C(x + A_n^*\lambda_n), c - P_C(x + A_n^*\lambda_n)\rangle = 0\end{aligned}$$

by (5.7). Again using the characterization of best approximations from C, we deduce that $c = P_C(x + A^*\lambda)$ which completes the proof. ■

The following example illustrates an application of Theorem 5.1. Let M_0 be the piecewise linear B-spline having the values 0, 1, and 0 at the integers $-1, 0$, and 1, respectively, which is supported on $[-1, 1]$ and set

$$M_j(t) = M_0(t - j), \qquad -\infty < j < \infty.$$

It can be deduced from [3] that the B-splines $\{M_j | j = 0, \pm 1, \pm 2, \ldots\}$ from an unconditional basis for the closed subspace they span in $L_2(\mathbf{R})$. Consider the set

$$\begin{aligned}K &= \{y \in L_2(\mathbf{R}) | y \ge 0, \langle y, M_j\rangle = d_j, -\infty < j < \infty\} \\ &= \{y \in C | Ay = d\},\end{aligned}$$

where

$$C = \{y \in L_2(\mathbf{R}) | y \ge 0\},$$

$d = (\ldots, d_{-1}, d_0, d_1, \ldots)$ is in $l_2 = l_2(\mathbf{Z})$, $d_j = \langle k, M_j\rangle$, and A: $L_2(\mathbf{R}) \to l_2$ is defined by

$$Ax := (\ldots, \langle x, M_{-1}\rangle, \langle x, M_0\rangle, \ldots).$$

That is, $(Ax)(j) = \langle x, M_j\rangle$, $-\infty < j < \infty$. In the following, we use the notation

$$\{d_j\} \approx \{f_j\}$$

if there exist positive constants c_1 and c_2 such that $c_1 d_j \le f_j \le c_2 d_j$ for all j. Hence, by setting $d_j = \langle k, M_j\rangle$ where $k = \sum_{i=-\infty}^{\infty} (5/(|i|+1))\chi_{[i, i+1/2]}$ it is clear that $k \in L_2(\mathbf{R})$ and

$$\{d_j\} \approx \left\{\frac{1}{1 + |j|}\right\}.$$

Consequently, K is nonempty. Our problem is to characterize $P_K(x)$ for each $x \in L_2(\mathbf{R})$. By a simple translation, we may simply consider the case $x = 0$. Since $\langle y, M_j\rangle \to 0$ as $|j| \to \infty$, the data cone will contain no interior points. Nevertheless, we have the following result.

Proposition 5.1. *The element of minimal norm in K is given by*

$$\left(\sum_{-\infty}^{\infty} \lambda_j M_j\right)_+ \tag{5.8}$$

for some $\lambda \in l_2$.

Proof. For each integer $n > 0$, define A_n: $L_2(\mathbf{R}) \to l_2\{-n, -n+1, \ldots, n\}$ by

$$A_n x := (\langle x, M_{-n}\rangle, \langle x, M_{-n+1}\rangle, \ldots, \langle x, M_n\rangle),$$

and set

$$d^{(n)} = (d_{-n}, d_{-n+1}, \ldots, d_n) \in l_2\{-n, \ldots, n\}$$

and

$$K_n = \{x \in C \mid A_n x = d^{(n)}\}.$$

Then by Corollary 4.2, since $d_j > 0$ for all j, $d^{(n)} \in \operatorname{int} A_n C$ for each n. By Theorem 3.2, there exists $\lambda_n = (\lambda_{n,-n}, \ldots, \lambda_{n,n}) \in l_2\{-n, \ldots, n\}$ such that

$$P_{K_n}(0) = P_C(A_n^* \lambda_n). \tag{5.9}$$

Let

$$\rho := \sup_n \|A_n^* \lambda_n\| = \sup_n \left\| \sum_{j=-n}^{n} \lambda_{nj} M_j \right\|.$$

If $\rho < \infty$, then Theorem 5.1 implies that

$$P_K(0) = P_C(A^*\lambda) \qquad \text{for some} \quad \lambda \in l_2.$$

But, by Lemma 4.1,

$$P_C(A^*\lambda) = (A^*\lambda)_+ = \left(\sum_{-\infty}^{\infty} \lambda_j M_j\right)_+.$$

Thus, the proof is complete if $\rho < \infty$.

While it can be shown that $\rho < \infty$, the proof is rather involved and lengthy. We prefer to take an alternate (and shorter) approach to verifying (5.8). By Lemma 4.1,

$$S_n^+ := P_C(A_n^* \lambda_n) = \left(\sum_{j=-n}^{n} \lambda_{n_j} M_j\right)_+.$$

We next prove that, for each j,

$$\rho_j := \sup_n |\lambda_{nj}| < \infty.$$

This is accomplished by showing that $\lim_n \sup|\lambda_{nj}| = \infty$ leads to a contradiction.

Case 1. Sone subsequence of $\{\lambda_{nj}\}$ diverges to $+\infty$.

By passing to a subsequence, we may assume that

$$\lambda_{nj} = |\lambda_{nj}| \to \infty \qquad \text{as} \quad n \to \infty.$$

Then since $\{d_j\} \approx \{1/(|j|+1)\}$ and

$$\text{(5.10)} \qquad \frac{1}{|j|+1} = \int_{j-1}^{j+1} S_n^+ M_j = \int_{j-1}^{j+1} (\lambda_{n,j-1} M_{j-1} + \lambda_{nj} M_j + \lambda_{n,j+1} M_{j+1})_+ M_j,$$

the following must hold:

$$\lambda_{n,j-1} = -|\lambda_{n,j-1}| \to -\infty$$

and

$$\lambda_{n,j+1} = -|\lambda_{n,j+1}| \to -\infty \qquad \text{as} \quad n \to \infty$$

for otherwise the integral would tend to infinity. Note that the restriction of S_n^+ to $[j-1, j+1]$, call it $S_{n,j}^+$, is a piecewise linear spline supported on $[j-\varepsilon_n, j+\delta_n]$, where $\varepsilon_n \to 0$, $\delta_n \to 0$, and $S_{n,j}^+(j) = \lambda_{nj}$.

Now since $\langle S_{n,j}^+, M_j \rangle = d_j$, d_j is approximately equal to $2^{-1}(\varepsilon_n + \delta_n)\lambda_{nj}$ which equals the area under $S_{n,j}^+$. On the other hand, the square of the L_2-norm of $S_{n,j}^+$ is equal to $(\varepsilon_n + \delta_n)\lambda_{nj}^2$ which is approximately equal to $2d_j\lambda_{nj}$. Hence $\|P_K(0)\|^2 \geq \|S_n^+\|^2 \geq \|S_{n,j}^+\|^2 \approx 2d_j\lambda_{nj} \to \infty$ as $n \to \infty$. This contradiction shows Case 1 cannot occur.

Case 2. Some subsequence of $\{\lambda_{nj}\}$ diverges to $-\infty$.

By passing to a subsequence, we may assume that

$$\lambda_{nj} = -|\lambda_{nj}| \to -\infty.$$

We consider three possible subcases.

Subcase 2(a). $\lambda_{n,j-1} < 0$, $\lambda_{nj} \to -\infty$, and $\lambda_{n,j+1} \geq 0$.

From Case 1, $\lambda_{n,j+1} \leq \rho_{j+1}$ for all n and hence the function

$$(\lambda_{n,j-1} M_{j-1} + \lambda_{nj} M_j + \lambda_{n,j+1} M_{j+1})_+$$

has the graph of a triangle having maximum height ρ_{j+1} and base shrinking to zero. Thus, the expression in (5.10) is equal to

$$0 < d_j = \int_{j-1}^{j+1} (\lambda_{n,j-1} M_{j-1} + \lambda_{nj} M_j + \lambda_{n,j+1} M_{j+1}) M_j \to 0$$

as $n \to \infty$ which is absurd.

Subcase 2(b). $\lambda_{n,j-1} > 0$, $\lambda_{nj} \to -\infty$, and $\lambda_{n,j+1} < 0$. This is similar to Subcase 2(a).

Subcase 2(c). $0 < \lambda_{n,j-1} \leq \rho_{j-1}$, $\lambda_{nj} \to -\infty$, and $0 < \lambda_{n,j+1} \leq \rho_{j+1}$. This also follows along the lines of Subcase 2(a).

Thus, we have proved that, for each j, $\{\lambda_{nj}\}$ is a bounded sequence.

Note that

$$\|P_K(0)\| \geq \|P_{K_n}(0)\| = \|S_n^+\|$$

for all n so $\{S_n^+\}$ is a bounded sequence. By the Eberlein–Smulian theorem, we may pass to a subsequence and assume that

$$S_n^+ \overset{w}{\rightharpoonup} g$$

for some $g \in L_2(\mathbf{R})$. The same proof as in Theorem 5.1 shows that $g = P_K(0)$ and $\|S_n^+ - g\| \to 0$.

It follows that on any interval $[-N, N]$, $\lim_{n\to\infty} S_n^+ = S^+$, i.e., g is the plus function of a spline on $[-N, N]$. The coefficients in $S^+ = (\sum_{-\infty}^{\infty} \lambda_j M_j)_+$ are the limits of the λ_{nj}. It follows that $S^+ = g$ in $L_2(\mathbf{R})$, so $S^+ = P_K(0)$ and the proof is complete. ■

6. Final Remarks

We point out that the definition of a dual cone may be formulated in any normed linear space. Hence, property CHIP may be defined just as in the inner product space setting. In particular, a result characterizing best approximations that generalizes Theorem 2.1 and is the basis for many applications to non-Hilbert space can be stated as follows.

Theorem 6.1. *Let $\{K_i | i \in I\}$ be a collection of convex subsets of a normed linear space X, $K = \bigcap_i K_i$, $x \in X$, and $k_0 \in K$. Then $k_0 \in P_K(x)$ if and only if*

$$D(x - k_0) \cap (K - k_0)^0 \neq \varnothing.$$

Moreover, if $\{K_i | i \in I\}$ has property CHIP, then $k_0 \in P_K(x)$ if and only if

$$D(x - k_0) \cap \text{w*-cl}\left[\sum_{i \in I} (K_i - k_0)^0\right] \neq \varnothing.$$

Here D denotes the norm-duality map $D: X \to 2^{X^*}$ defined by

$$D(x) = \{x^* \in X^* | \|x^*\| = \|x\|, x^*(x) = \|x^*\| \, \|x\|\}$$

and "w*-cl" denotes "weak* closure of."

Acknowledgments. We wish to thank C. Micchelli for providing us with a preprint of [17] and making several helpful comments on this material and Phil Smith for alerting us to [22]. In addition, we are indebted to Wu Li for supplying the proof of Lemma 4.3 which simplified our original approach. The research of C. K. Chui and J. D. Ward was supported by the National Science Foundation under Grant Numbers DMS-8602337 and DMS-8701190. The work of F. Deutsch was begun and mainly performed while he was a Visiting Professor of Mathematics at the Center for Approximation Theory at Texas A&M University during the Spring Semester of 1987.

References

1. V. E. Anderson, P. A. Ivert (1987): *Constrained interpolants with minimal $W^{k,p}$-norm.* J. Approx. Theory, 49:283–288.
2. A. Ben-Israel (1969): *Linear equations and inequalities on finite-dimensional, real or complex, vector spaces: a unified theory.* J. Math. Anal. Appl., 27:367–389.

3. C. DE BOOR (1973): *The quasi-interpolant as a tool in elementary polynomial spline theory.* In: Approximation Theory (G. G. Lorentz, ed.). New York: Academic Press, 269-276.
4. J. M. BORWEIN, H. WOLKOWICZ (to appear): *A simple constraint qualification in infinite-dimensional programming.* Math Programming.
5. J. P. BUTLER, J. A. REEDS, S. V. DAWSON (1981): *Estimating solutions of first kind integral equations with nonnegative constraints and optimal smoothing.* SIAM. J. Numer. Anal., **10**:381-397.
6. C. K. CHUI (1988): Multivariate Splines: Theory and Applications. CBMS-NSF Lecture Series. Philadelphia: SIAM.
7. J. DIESTEL (1984): Sequences and Series in Banach Spaces. New York: Springer-Verlag.
8. A. L. DONTCHEV, G. ILLIEV, M. M. KONSTANTINOV (1985): *Constrained interpolation via optimal control.* In: Proceedings of the Fourteenth Spring Conference of the Union of Bulgarian mathematicians, Sunny Beach, April 6-9, 1985, pp. 385-392.
9. P. R. HALMOS (1967): A Hilbert Space Problem Book. New York: Van-Nostrand.
10. R. B. HOLMES (1975): Geometric Functional Analysis and Its Applications. New York: Springer-Verlag.
11. G. ILLIEV, W. POLLUL (1984): *Convex interpolation with minimal L_∞-norm of the second derivative.* Math Z., **186**:49-56.
12. G. ILLIEV, W. POLLUL (1984): *Convex interpolation by functions with minimal L_p-norm* $(1 < p < \infty)$ *of the kth derivative.* In: Proceedings of the Thirteenth Spring Conference of the Union of Bulgarian Mathematicians, Sunny Beach, April 6-9, 1984.
13. L. IRVINE (1985): Constrained Interpolation. Masters Thesis, Old Dominion University.
14. L. D. IRVINE, S. P. MARIN, P. W. SMITH (1986): *Constrained interpolation and smoothing.* Constr. Approx., **2**:129-151.
15. D. G. LUENBERGER (1968): Optimization by Vector Space Methods. New York: Wiley.
16. C. A. MICCHELLI, P. W. SMITH, J. SWETITS, J. D. WARD (1985): *Constrained L_p Approximation.* Constr. Approx., **1**:93-102.
17. C. A. MICCHELLI, F. UTRERAS (1988): *Smoothing and interpolation in a convex subset of a Hilbert space.* SIAM J. Sci. Statist. Comput., **9**:728-746.
18. D. MOSKOVITZ, L. L. DINES (1939): *Convexity in a linear space with an inner product.* Duke Math. J., **5**:520-534.
19. G. OPFER, H. J. OBERLE (to appear): *The derivation of cubic splines with obstacles by methods of optimization and optimal control.* Numer. math.
20. R. T. ROCKAFELLAR (1970): Convex Analysis. Princeton, NJ: Princeton University Press.
21. P. W. SMITH, J. D. WARD (1988): *Distinguished solutions to an L_∞ minimization problem.* Approx. Theory Appl., **4**:29-40.
22. P. W. SMITH, H. WOLKOWICZ (1986): *A nonlinear equation for linear programming.* Math Programming, **34**:235-238.
23. J. WARD (1986): *Some constrained approximation problems.* in: Approximation Theory V (C. K. Chui, L. L. Schumaker, J. D. Ward eds.) New York: Academic Press, pp. 211-229.

C. K. Chui
Department of Mathematics
Texas A&M University
College Station
Texas 77843-3368
U.S.A.

F. Deutsch
Department of Mathematics
The Pennsylvania State University
University Park
Pennsylvania 16802
U.S.A.

J. D. Ward
Department of Mathematics
Texas A&M University
College Station
Texas 77843-3368
U.S.A.