JOURNAL OF APPROXIMATION THEORY **90**, 385–414 (1997)
ARTICLE NO. AT963082

A Dual Approach to Constrained Interpolation from a Convex Subset of Hilbert Space

Frank Deutsch*

Department of Mathematics, The Pennsylvania State University, University Park, Pennsylvania 16802

Wu Li‡

Department of Mathematics, Old Dominion University, Norfolk, Virginia 23529

and

Joseph D. Ward†

Department of Mathematics, Texas A&M University, College Station, Texas 77843

Communicated by Charles K. Chui

Received September 7, 1995; accepted in revised form August 29, 1996

Many interesting and important problems of best approximation are included in (or can be reduced to) one of the following type: in a Hilbert space X, find the best approximation $P_K(x)$ to any $x \in X$ from the set $K := C \cap A^{-1}(b)$, where C is a closed convex subset of X, A is a bounded linear operator from X into a finite-dimensional Hilbert space Y, and $b \in Y$. The main point of this paper is to show that $P_K(x)$ is *identical* to $P_C(x + A^*y)$—the best approximation to a certain perturbation $x + A^*y$ of x—from the convex set C or from a certain convex extremal subset C_b of C. The latter best approximation is generally much easier to compute than the former. Prior to this, the result had been known only in the case of a convex cone or for *special* data sets associated with a closed convex set. In fact, we give an *intrinsic characterization* of those pairs of sets C and $A^{-1}(b)$ for which this can always be done. Finally, in many cases, the best approximation $P_C(x + A^*y)$ can be obtained numerically from existing algorithms or from modifications to existing algorithms. We give such an algorithm and prove its convergence. © 1997 Academic Press

* Supported by NSF Grant DMS-9303705. E-mail: deutsch@math.psu.edu.
† Supported by NSF Grant DMS-9505460 and AFOSR Grant F49620-95-1-0194. E-mail: Joe.Ward@math.tamu.edu.
‡ E-mail: wuli@math.odu.edu.

0021-9045/97 $25.00

1. INTRODUCTION

Many interesting and important problems of best approximation are included in (or can be reduced to) one of the following type: in a Hilbert space X, find the best approximation $P_K(x)$ to any $x \in X$ from the set

$$K := C \cap A^{-1}(b),$$

where C is a closed convex subset of X, A is a bounded linear operator from X into a finite-dimensional Hilbert space Y, and $b \in Y$.

The main point of this paper is to give an *intrinsic characterization* of those pair of sets $\{C, A^{-1}(b)\}$ for which the following *perturbation property* holds: *for every $x \in X$, there exists $y \in Y$ such that*

$$P_K(x) = P_C(x + A^* y).$$

(We will later see that this equation is equivalent to the following nonlinear one: $AP_C(x + A^* y) = b$.) That is, when is it always true that the best approximation to x from K is the same as the best approximation to some perturbation $x + A^* y$ of x from C? Up to now, such characterizations had been noticed only in certain special cases (see [12, 13, 2, 3, 19, 7]). In particular, we will show that if C is polyhedral or if b is in the relative interior of $A(C)$, then the perturbation property holds. Also, if we replace C by the extremal subset C_b (see Definition 4.1 below), then the perturbation property *always* holds! As a consequence of these results, we can obtain all of the previous characterizations of best approximations from the intersection $C \cap A^{-1}(b)$ that are known to us.

The merit of this theorem is based mainly on four facts:

1. For most applications, it is *easier* to compute best approximations from C than from K.

2. In many applications, X is infinite-dimensional and hence the computation of $P_K(x)$ is intrinsically an infinite-dimensional problem. However, as will be seen, the computation of $P_C(x + A^* y)$ involves only a *finite* number of parameters.

3. This problem includes the general "shape-preserving interpolation" problem that arises in curve and surface fitting (see, e.g., [12] or [5]).

4. In many cases, $AP_C(x + A^* y) - b$ is the gradient of a differentiable convex function $\Phi(y)$. Thus, solving the nonlinear equation $AP_C(x + A^* y) = b$ is equivalent to finding a global minimizer of $\Phi(y)$ which can be resolved by various unconstrained minimization techniques (cf. [10, 9]).

The elementary facts from convex analysis necessary for the dual approach to this problem are listed in Section 2. The main theoretical results are stated in Sections 3 and 4. In Section 5, we include a worked example showing how the theory can be applied. In Section 6, we show that $AP_C(x+A^*y)-b$ is the gradient of a convex *quadratic spline* (i.e., a convex differentiable piecewise quadratic function) if C is a polyhedral set. As a consequence, we propose a steepest descent method for solving the nonlinear equation $AP_C(x+A^*y)=b$, which generates a sequence of iterates converging linearly to a solution y of this equation.

We conclude the Introduction by describing some notation used and stating a theorem which characterizes best approximations.

Recall that a subset K of the Hilbert space X is *convex* (resp., a *convex cone*) provided that

$$\lambda K+(1-\lambda)K\subset K \qquad \text{for all} \quad 0\leqslant\lambda\leqslant 1$$

(resp., $\rho K\subset K$ and $K+K\subset K$ for all $\rho\geqslant 0$). For any nonempty subset S of X, the *convex hull* (resp., *conical hull*) of S is the intersection of all convex sets (resp., convex cones) which contain S. The convex hull (resp., conical hull) of S is denoted by $co(S)$ (resp., $\operatorname{con}(S)$). If S is nonempty, the *dual cone* (resp., *orthogonal complement*) of S is the set

$$S^0:=\{x\in X \mid \langle x, y\rangle\leqslant 0 \text{ for all } y\in S\}$$

(resp., $S^\perp:=\{x\in X \mid \langle x, y\rangle=0 \text{ for all } y\in S\}$). Note that S^0 (resp., $S^\perp$) is a nonempty closed convex cone (resp., closed linear subspace). The closure of any set S is denoted by $\bar{S}$.

It is well-known that if K is a closed convex subset of the Hilbert space X, every $x\in X$ has a unique best approximation $P_K(x)$ in K to x. That is, $P_K(x)\in K$ and

$$\|x-P_K(x)\|=\inf\{\|x-y\| \mid y\in K\}.$$

The following result is well-known, but we are not certain where or when it first appeared.

THEOREM 1.1 (Characterization of Best Approximations). *Let K be a closed convex subset of the Hilbert space X, $x\in X$, and $x_0\in K$. Then $x_0=P_K(x)$ if and only if $x-x_0\in(K-x_0)^0$.*

If A is a bounded linear operator from X to Y, then A^*, $\mathscr{R}(A)$, and $\mathscr{N}(A)$ denote its adjoint map, range, and nullspace, respectively. All other terminology and notation is standard and can be found, e.g., in [8].

2. DUAL CONES

In this section we collect for reference purposes some basic facts about dual cones that will be useful to us. We also define the strong property CHIP which will prove fundamental to our work. Throughout this section X will denote an arbitrary Hilbert space.

LEMMA 2.1. *Let $S, S_1, ..., S_m$ be nonempty subsets of X. Then*

$$S^0 = (\bar{S})^0 = (\text{con } S)^0, \tag{2.1.1}$$

$$S^{00} := (S^0)^0 = \overline{\text{con}}\, S, \quad \textit{and} \tag{2.1.2}$$

$$\left(\bigcap_1^m S_i\right)^0 \supset \overline{\sum_1^m S_i^0}. \tag{2.1.3}$$

LEMMA 2.2. *Let $C_1, ..., C_n$ be closed convex cones in X. Then*

$$\left(\sum_1^n C_i\right)^0 = \bigcap_1^m C_i^0, \quad \textit{and} \tag{2.2.1}$$

$$\left(\bigcap_1^m C_i\right)^0 = \overline{\sum_1^n C_i^0}. \tag{2.2.2}$$

Both these lemmas are well known and can be found, for example, in [15, Theorem 14.1 and Corollary 16.4.2]. Also, it should be mentioned that Lemma 2.2 is false, in general, for closed convex sets which are not cones, and the closure bar in (2.2.2) cannot be removed, even if X is finite-dimensional. The important definition for our purposes is the following.

DEFINITION 2.3. A collection of convex sets $\{C_1, C_2, ..., C_m\}$ having a nonempty intersection will be said to have the **strong property CHIP**, provided that for every $x \in \bigcap_1^m C_i$,

$$\left(\bigcap_1^m C_i - x\right)^0 = \sum_1^m (C_i - x)^0. \tag{2.3.1}$$

Remark. Recall [2] that $\{C_1, C_2, ..., C_m\}$ was said to have "property CHIP" (the "conical hull intersection property") provided that for every $x \in \bigcap_1^m C_i$, (2.3.1) holds with a *closure bar* over the set on the right. Thus the strong property CHIP requires in addition that the sum of the dual cones on the right of (2.3.1) always be closed.

Alternatively, one can show that the strong property CHIP is equivalent to the statement that *the subdifferential is additive on the sum of the* *indicator functions of the* C_i (Lemma 2.4 below). To explain this, recall that the *indicator function* of a set C is defined by

$$I_C(x) := \begin{cases} 0, & \text{if } \ x \in C, \\ \infty, & \text{if } \ x \notin C. \end{cases}$$

Also, the *subdifferential* of a function $f: X \to \mathbb{R} \cup \{\infty\}$ is the set-valued mapping ∂f defined on X by

$$\partial f(x) := \{x^* \in X \mid f(x) - f(x') \leqslant \langle x^*, x - x' \rangle \text{ for every } x' \in X\}.$$

Moreover, it is well-known and easy to check that

$$I_{\bigcap_1^m C_i} = \sum_1^m I_{C_i}, \quad \text{and} \tag{2.3.2}$$

$$\partial I_C(x) = (C - x)^0 \qquad \text{for each} \quad x \in C. \tag{2.3.3}$$

We now show that the strong property CHIP can be characterized in terms of an additivity property of the subdifferential mapping as well as in terms of a distributive property of the closed conical hull mapping. This latter property will prove especially useful later when verifying that the pair of sets $\{C, A^{-1}(b)\}$ has the strong property CHIP whenever either C is polyhedral or b is in the relative interior of $A(C)$.

LEMMA 2.4. *The following statements are equivalent*:

(1) $\{C_1, C_2, ..., C_m\}$ *has the strong property CHIP*;

(2) *For each* $x \in \bigcap_1^m C_i$,

$$\left(\bigcap_1^m C_i - x\right)^0 \subset \sum_1^m (C_i - x)^0;$$

(3) *For each* $x \in \bigcap_1^m C_i$,

$$\partial\left(\sum_1^m I_{C_i}\right)(x) = \sum_1^m \partial I_{C_i}(x); \quad \textit{and} \tag{2.4.1}$$

(4) *For each* $x \in \bigcap_1^m C_i$,

$$\overline{\text{con}}\left(\bigcap_1^m C_i - x\right) = \bigcap_1^m \overline{\text{con}}(C_i - x), \quad \textit{and} \tag{2.4.2}$$

$$\sum_1^m (C_i - x)^0 \textit{ is closed.} \tag{2.4.3}$$

Proof. The equivalence of (1) and (2) follows using (2.1.3) with S_i replaced by $C_i - x$. Moreover, the equivalence of (1) and (3) follows from (2.3.2) and (2.3.3).

Now suppose (1) holds and $x \in \bigcap_1^m C_i$. Using Lemma 2.1 we get

$$\left(\bigcap_1^m C_i - x\right)^0 = \left[\overline{\text{con}}\left(\bigcap_1^m C_i - x\right)\right]^0.$$

By the strong property CHIP, we obtain

$$\left(\bigcap_1^m C_i - x\right)^0 = \sum_1^m (C_i - x)^0. \tag{2.4.4}$$

It follows that

$$\left[\overline{\text{con}}\left(\bigcap_1^m C_i - x\right)\right]^0 = \sum_1^m (C_i - x)^0.$$

Taking dual cones of both sides of this equation, and using Lemmas 2.1 and 2.2, we obtain

$$\begin{aligned}\overline{\text{con}}\left(\bigcap_1^m C_i - x\right) &= \left[\left(\overline{\text{con}}\left(\bigcap_1^m C_i - x\right)\right)^0\right]^0 \\ &= \left(\sum_1^m (C_i - x)^0\right)^0 = \bigcap_1^m ((C_i - x)^0)^0 = \bigcap_1^m \overline{\text{con}}(C_i - x).\end{aligned}$$

That is, (2.4.2) holds. Furthermore, by (2.4.4) and the fact that dual cones are closed, it follows that (2.4.3) holds. Thus (4) holds.

Conversely, if (4) holds, then taking dual cones of both sides of (2.4.2) and using Lemmas 2.1 and 2.2, we obtain

$$\begin{aligned}\left(\bigcap_1^m C_i - x\right)^0 &= \left[\overline{\text{con}}\left(\bigcap_1^m C_i - x\right)\right]^0 = \left[\bigcap_1^m \overline{\text{con}}(C_i - x)\right]^0 \\ &= \overline{\sum_1^m [\overline{\text{con}}(C_i - x)]^0} = \overline{\sum_1^m (C_i - x)^0} = \sum_1^m (C_i - x)^0.\end{aligned}$$

Thus (1) holds, and this proves the equivalence of (1) and (4). ∎

Remark. The equivalence of statements (1) and (3) of Lemma 2.4 states that $\{C_1, C_2, ..., C_m\}$ has the strong property CHIP if and only if the subdifferential is additive on the sum of the indicator functions of the C_i.

3. MAIN RESULTS

In this section we present two of our main theoretical results. Here and in the sequel, unless explicitly stated otherwise, X and Y will always denote (real) Hilbert spaces with Y finite-dimensional, A is a bounded linear operator from X to Y, C is a nonempty closed convex subset of X, $b \in Y$, and

$$K := C \cap A^{-1}(b) = \{x \in C \mid Ax = b\}.$$

Note $K \neq \emptyset$ if and only if $b \in A(C)$. (See [7] for conditions which guarantee this.) For our central results, it will be *assumed* that $K \neq \emptyset$ (i.e., $b \in A(C)$). Then each x in X has a unique best approximation $P_K(x)$ in K. Our essential purpose is to give a useful *characterization* of $P_K(x)$ which lends itself to actual computation. Before proving the main characterization theorem, it is convenient to give an alternate formulation of the strong property CHIP for the pair of sets $\{C, A^{-1}(b)\}$.

LEMMA 3.1. *The following statements are equivalent*:

(1) $\{C, A^{-1}(b)\}$ *has the strong property CHIP*;

(2) *For every* $x_0 \in C \cap A^{-1}(b)$,

$$[C \cap A^{-1}(b) - x_0]^0 = (C - x_0)^0 + \mathcal{R}(A^*); \tag{3.1.1}$$

(3) *For every* $x_0 \in C \cap A^{-1}(b)$,

$$[C \cap A^{-1}(b) - x_0]^0 \subset (C - x_0)^0 + \mathcal{R}(A^*); \quad \textit{and} \tag{3.1.2}$$

(4) *For every* $x_0 \in C \cap A^{-1}(b)$,

$$\overline{\text{con}}[(C - x_0) \cap \mathcal{N}(A)] = \overline{\text{con}}(C - x_0) \cap \mathcal{N}(A), \quad \textit{and} \tag{3.1.3}$$

$$(C - x_0)^0 + \mathcal{R}(A^*) \qquad \textit{is closed.} \tag{3.1.4}$$

Proof. First observe that for every $x_0 \in C \cap A^{-1}(b)$, we have

$$C \cap A^{-1}(b) - x_0 = (C - x_0) \cap (A^{-1}(b) - x_0) = (C - x_0) \cap \mathcal{N}(A).$$

Next note that from [8], we obtain

$$\mathcal{N}(A)^0 = \mathcal{N}(A)^{\perp} = \overline{\mathcal{R}(A^*)} = \mathcal{R}(A^*),$$

where the last equality holds since $\mathcal{R}(A^*)$ is finite-dimensional, hence closed. Using these facts, an application of Lemma 2.4 yields the result. ∎

Now we can prove our main characterization theorem. It shows that the strong property CHIP for the sets $\{C, A^{-1}(b)\}$ is the *precise* condition that allows us to always replace the approximation of any $x \in X$ from the set K by approximating a perturbation of x from the set C.

THEOREM 3.2. *The following statements are equivalent*:

(1) $\{C, A^{-1}(b)\}$ *has the strong property CHIP*;

(2) *For every* $x \in X$, *there exists* $y \in Y$ *such that*

$$A[P_C(x + A^* y)] = b; \tag{3.2.1}$$

(3) *For every* $x \in X$, *there exists* $y \in Y$ *such that*

$$P_K(x) = P_C(x + A^* y). \tag{3.2.2}$$

In fact, the y *that works in* (3.2.1) *also works in* (3.2.2), *and conversely. That is*, (3.2.1) *holds if and only if* (3.2.2) *holds.*

Proof. First note that if (3.2.2) holds, then $P_C(x + A^* y) \in K$ and, hence, (3.2.1) holds. Conversely, suppose (3.2.1) holds. Then $x_0 = P_C(x + A^* y) \in K$ and by Theorem 1.1,

$$x + A^* y - x_0 \in (C - x_0)^0. \tag{3.2.3}$$

It follows from (3.2.3) that $x - x_0 \in (C - x_0)^0 + \mathcal{R}(A^*)$. But the equivalence of (2) and (3) in Lemma 3.1 implies that

$$(C - x_0)^0 + \mathcal{R}(A^*) \subset (K - x_0)^0.$$

Thus $x - x_0 \subset (K - x_0)^0$. By Theorem 1.1 again, $x_0 = P_K(x)$. Thus $P_K(x) = P_C(x + A^* y)$; i.e., (3.2.2) holds. This proves the equivalence of (2) and (3) as well as the last statement of the theorem.

Using Lemma 3.1, (1) holds if and only if for each $x_0 \in K$,

$$(K - x_0)^0 = (C - x_0)^0 + \mathcal{R}(A^*). \tag{3.2.4}$$

If (1) holds and $x \in X$, let $x_0 = P_K(x)$. Then Theorem 1.1 implies that $x - x_0 \in (K - x_0)^0$. Using (3.2.4), there exists $y \in Y$ such that $x - x_0 \in (C - x_0)^0 - A^* y$. Thus $x + A^* y - x_0 \in (C - x_0)^0$ implies (by Theorem 1.1) that $x_0 = P_C(x + A^* y)$. That is, $P_K(x) = P_C(x + A^* y)$ and (3) holds.

Finally, suppose (3) holds and let $x_0 \in K$. Choose any $z \in (K - x_0)^0$ and set $x := z + x_0$. Note that $x - x_0 = z \in (K - x_0)^0$ so that by Theorem 1.1,

$x_0 = P_K(x)$. Since (3) holds, there exists $y \in Y$ such that $x_0 = P_K(x) = P_C(x + A^*y)$. Hence, Theorem 1.1 implies that

$$\begin{aligned} z &= x - x_0 = x - P_C(x + A^*y) \\ &= x + A^*y - P_C(x + A^*y) - A^*y \in [C - P_C(x + A^*y)]^0 + \mathscr{R}(A^*) \\ &= (C - x_0)^0 + \mathscr{R}(A^*). \end{aligned}$$

Since z was an arbitrary element of $(K - x_0)^0$, this shows that

$$(K - x_0)^0 \subset (C - x_0)^0 + \mathscr{R}(A^*).$$

We conclude from Lemma 3.1 that (1) holds. ∎

Remarks. (1) This theorem allows us to determine the best approximation in K to any x by instead determining the best approximation in C to a *perturbation* of x. The usefulness of this is that it is usually much easier to determine best approximations from C than the intersection K. The price we pay for this simplicity is that now we must determine just *which* perturbation of x works! However, this is determined by the (generally nonlinear) Eq. (3.2.1) for the unknown vector y. Moreover, since y lies in the finite-dimensional space Y, (3.2.1) is an equation involving only a *finite* number of parameters and is often amenable to standard algorithms (e.g., descent methods) for their solution. In fact, in Section 6 we describe a descent method to solve this equation, and we prove the algorithm converges linearly.

(2) Of course, to apply the theorem, one must first determine whether the pair of sets $\{C, A^{-1}(b)\}$ has the strong property CHIP. Fortunately, some of the more interesting pairs that arise in practice do have this property. In the remainder of this section, we will show that if C is a "polyhedral" set, then the pair $\{C, A^{-1}(b)\}$ has the strong property CHIP. A number of consequences, some well-known, will follow immediately from this fact. In the next section, we will show that if b is in the relative interior of $A(C)$, then the pair $\{C, A^{-1}(b)\}$ has the strong property CHIP. An important consequence of this fact is that we can define a certain convex "extremal" subset C_b of C such that $C_b \cap A^{-1}(b) = K := C \cap A^{-1}(b)$ and b is in the relative interior of $A(C_b)$. Thus we can apply the above result to see that the pair $\{C_b, A^{-1}(b)\}$ has the strong property CHIP, and then apply Theorem 3.2 to this pair. In short, if the pair $\{C, A^{-1}(b)\}$ has the strong property CHIP, then Theorem 3.2 can be applied. If not, then we can still apply Theorem 3.2, but to the pair $\{C_b, A^{-1}(b)\}$! Thus in *every* case, we have a formula for the best approximation $P_K(x)$ either as $P_C(x + A^*y)$ or $P_{C_b}(x + A^*y)$.

(3) As an alternative to the *numerical* computation of $P_K(x)$, let us mention the following. If the convex set K can be written as the intersection of a finite number of closed convex sets K_i, $K = \bigcap_1^m K_i$, and it is straightforward to compute best approximations from the K_i (e.g., if there is a formula for computing $P_{K_i}(x)$ for any $x \in X$ and any i), then the problem of computing $P_K(x)$ is amenable to *Dykstra's method of alternating projections*. This is an iterative algorithm that reduces the problem to one involving computing best approximations from only the *individual* sets K_i. (See [1] or the exposition [5].) This will be the case, for example, when C is the cone of nonnegative functions ("shape-preserving interpolation") that was studied by many authors (e.g., [9, 2, 3, 6]). Computing best approximations in this way is currently under investigation.

We are next going to show that when C is a polyhedron, then $\{C, A^{-1}(b)\}$ has the strong property CHIP. Recall the the closed convex set C is called a *polyhedron* if it is the intersection of a finite number of closed halfspaces. That is, C is a polyhedron if and only if $C = \bigcap_1^k \{z \in X \mid \langle z, x_i \rangle \leqslant d_i\}$ for some $x_i \in X \setminus \{0\}$ and real scalars d_i. Also, a convex cone D is called *finitely generated* if there exists a finite set $\{z_1, z_2, ..., z_k\}$ such that $D = \operatorname{con}\{z_1, z_2, ..., z_k\}$. That is, $D = \{\sum_1^k \rho_i z_i \mid \rho_i \geqslant 0\}$. It is well-known that if C_1 and C_2 are polyhedral sets in the Euclidean space $\mathbb{R}^n$, then $\{C_1, C_2\}$ has the strong property CHIP (cf. [15, Theorem 20.1]). In the following three lemmas, we are going to show that when C is a polyhedron in any Hilbert space X (not necessarily finite-dimensional), then $\{C, A^{-1}(b)\}$ has the strong property CHIP.

LEMMA 3.3. *If C is a polyhedron, then for each $x \in C$,*

(1) *$(C - x)^0$ is a finitely generated cone, and*

(2) *$(C - x)^0 + \mathcal{R}(A^*)$ is closed.*

Proof. Let $C = \bigcap_1^k \{y \in X \mid \langle z_i, y \rangle \leqslant d_i\}$, where $z_i \in X$ and $d_i \in \mathbb{R}$. For any $x \in C$, let $I(x) := \{i \mid \langle z_i, x \rangle = d_i\}$ be the set of indices for the "active" constraints at x. Then it is not too difficult to show that

$$(C - x)^0 = \left\{ \sum_{i \in I(x)} \lambda_i z_i \mid \lambda_i \geqslant 0 \right\}$$

(cf. [15, Corollary 9.2.2]), and hence (1) holds.

The statement (2) is a consequence of statement (1) and Lemma 4.3 of [2]. ▮

LEMMA 3.4. *If B and D are convex sets and $0 \in B \cap D$, then*

$$\operatorname{con}(B \cap D) = \operatorname{con} B \cap \operatorname{con} D.$$

Proof. Obviously, $\operatorname{con}(B \cap D) \subset \operatorname{con}(B) \cap \operatorname{con}(D)$. Now let $x \in \operatorname{con}(B) \cap \operatorname{con}(D)$. Since $0 \in B \cap D$, we have

$$\operatorname{con}(B) = \{\lambda x \mid \lambda \geqslant 0, x \in B\}, \qquad \operatorname{con}(D) = \{\lambda x \mid \lambda \geqslant 0, x \in D\}.$$

If $x = 0$, then $x \in \operatorname{con}(B \cap D)$. Otherwise, there exist positive constants λ_1, λ_2 and $x_1 \in B$, $x_2 \in D$ such that

$$x = \lambda_1 x_1 = \lambda_2 x_2.$$

Let $\varepsilon = \min\{\lambda_1^{-1}, \lambda_2^{-2}\}$. Then $\varepsilon x = \varepsilon \lambda_1 x_1 \in B$ since it is a convex combination of x_1 and 0. Similarly, $\varepsilon x = \varepsilon \lambda_2 x_2 \in D$ since it is a convex combination of x_2 and 0. Therefore, $x = \varepsilon^{-1} \varepsilon x \in \operatorname{con}(B \cap D)$. ▮

LEMMA 3.5. *If $x \in C$ and $\operatorname{con}(C - x)$ is closed, then*

$$\overline{\operatorname{con}}(C - x) \cap \mathcal{N}(A) = \overline{\operatorname{con}}[(C - x) \cap \mathcal{N}(A)]$$

Proof. By Lemma 3.4, $\operatorname{con}(C - x) \cap \mathcal{N}(A) = \operatorname{con}[(C - x) \cap \mathcal{N}(A)]$. Since $\operatorname{con}(C - x)$ is closed, it follows that $\operatorname{con}[(C - x) \cap \mathcal{N}(A)]$ is closed and, hence,

$$\begin{aligned} \overline{\operatorname{con}}(C - x) \cap \mathcal{N}(A) &= \operatorname{con}(C - x) \cap \mathcal{N}(A) = \operatorname{con}[(C - x) \cap \mathcal{N}(A)] \\ &= \overline{\operatorname{con}}[(C - x) \cap \mathcal{N}(A)]. \end{aligned}$$

THEOREM 3.6. *If C is a polyhedron, then for each $x \in X$, there exists $y \in Y$ such that*

$$A[P_C(x + A^* y)] = b. \tag{3.6.1}$$

Moreover,

$$P_K(x) = P_C(x + A^* y) \tag{3.6.2}$$

for each $y \in Y$ which satisfies (3.6.1).

Proof. Lemmas 3.3(2), 3.5, and 3.1 show that $\{C, A^{-1}(b)\}$ has the strong property CHIP. The result then follows from Theorem 3.2. ▮

When $X = \mathbb{R}^n$, special cases of Theorem 3.6 were obtained by Smith and Wolkowicz [16] and by Chui, Deutsch, and Ward [2] for $C := \{x \in \mathbb{R}^n \mid x \geqslant 0\}$ and by Li, Pardalos, and Han [10] for $C := \{x \in \mathbb{R}^n \mid l \leqslant x \leqslant u\}$, where l, u are vectors of n components and some components of l, u can be $-\infty, +\infty$, respectively. See also [10, 9] for algorithms for solving the piecewise linear equation: $AP_C(x + A^* y) = b$ when $X = \mathbb{R}^n$ and $C := \{x \in \mathbb{R}^n \mid l \leqslant x \leqslant u\}$.

Next we will establish our second main result. Namely, if b is in the relative interior of $A(C)$, then $\{C, A^{-1}(b)\}$ has the strong property CHIP. We need some preliminary results. The first fact we need is the following well-known metric regularity theorem, which is a consequence of the celebrated Robinson–Ursescu theorem (see [14, Theorem 1; 17; 4, Theorem 2.2]). Here and in the sequel, the *interior* (*relative interior*) of a set S will be denoted by int S (ri S).

THEOREM 3.7. *Let Z and W be two (real) normed linear spaces and 2^W be the collection of all subsets of W. Suppose that $\Gamma: Z \to 2^W$ is a convex multifunction with closed graph, i.e.,*

$$\lambda\Gamma(z_1) + (1-\lambda)\,\Gamma(z_2) \subset \Gamma(\lambda z_1 + (1-\lambda)\,z_2)$$
$$\textit{for} \quad z_1, z_2 \in Z, \qquad 0 \leqslant \lambda \leqslant 1,$$

and $\text{Graph}(\Gamma) := \{(z, y) \mid z \in Z,\ y \in \Gamma(z)\}$ *is a closed subset of $Z \times W$. The inverse of Γ, denoted by Γ^{-1}, is defined by*

$$\Gamma^{-1}(S) := \{z \in Z \mid S \cap \Gamma(z) \neq \varnothing\}.$$

Let $z_0 \in Z$ and $y_0 \in \Gamma(z_0)$. Then $z_0 \in \operatorname{int} \Gamma^{-1}(W)$ if and only if Γ^{-1} is "metrically regular" at (y_0, z_0), i.e., there exist a neighborhood V of (y_0, z_0) and a constant $\gamma > 0$ such that

$$d(y, \Gamma(z)) \leqslant \gamma d(z, \Gamma^{-1}(y)) \qquad \textit{for every} \quad (y, z) \in V. \tag{3.7.1}$$

LEMMA 3.8. *Suppose that $b \in \operatorname{ri} A(C)$. Then for every $x \in K$,*

$$\overline{\text{con}}(C - x) \cap \mathcal{N}(A) = \overline{\text{con}}[(C - x) \cap \mathcal{N}(A)]. \tag{3.8.1}$$

Proof. Let $Z = \text{span}[A(C) - b]$ and $W = \text{span}(C - x)$. Define

$$\Gamma(z) = \{y \in C - x \mid Ay = z\}.$$

Then

$$\Gamma^{-1}(y) = \begin{cases} \{Ay\}, & \text{if} \quad y \in C - x, \\ \varnothing, & \text{if} \quad y \notin C - x. \end{cases}$$

Since $C - x$ is convex and A is linear, we have

$$\lambda\Gamma(z_1) + (1-\lambda)\,\Gamma(z_2) \subset \Gamma(\lambda z_1 + (1-\lambda) z_2)$$
$$\text{for} \quad z_1, z_2 \in Z, \qquad 0 \leqslant \lambda \leqslant 1.$$

That is, Γ is a convex multifunction. Since $C - x$ is closed and A is continuous, $\text{Graph}(\Gamma) = \{(z, y) \mid z \in Z,\ y \in C - x,\ z = Ay\}$ is a closed subset of $Z \times W$. Let $y_0 = 0$ and $z_0 = 0$. Then $z_0 \in \operatorname{int} \Gamma^{-1}(W)$. By Theorem 3.7, there exists positive constants ε and γ such that

$$d(y, \Gamma(z)) \leqslant \gamma d(z, \Gamma^{-1}(y)) \qquad \text{for} \quad z \in Z,\ y \in W,\ \|z\| < \varepsilon,\ \|y\| < \varepsilon. \tag{3.8.2}$$

Let $y = 0$. Then

$$d(y, \Gamma(z)) = d(0, \Gamma(z)) = \min\{\|w\| \mid Aw = z,\ w \in C - x\},$$
$$d(z, \Gamma^{-1}(y)) = d(z, \Gamma^{-1}(0)) = d(z, \{0\}) = \|z\|.$$

Therefore, (3.8.2) implies that for every $z \in \operatorname{span} A(C - x)$ with $\|z\| < \varepsilon$, we have

$$\min\{\|w\| \mid Aw = z,\ w \in C - x\} \leqslant \gamma \|z\|. \tag{3.8.3}$$

It is obvious that $\overline{\operatorname{con}}(C - x) \cap \mathcal{N}(A) \supset \overline{\operatorname{con}}[(C - x) \cap \mathcal{N}(A)]$. On the other hand, let $u \in \overline{\operatorname{con}}(C - x) \cap \mathcal{N}(A)$. By the definition, there exist elements $u_k \in \operatorname{con}(C - x)$ that converge to u. Since $\lim_{k \to \infty} Au_k = Au = 0$, there is $k_0 \geqslant 0$ such that $\|Au_k\| < \varepsilon$ for $k \geqslant k_0$. Applying (3.8.3) with $z = -Au_k$ (in span $A(C - x)$), we obtain that there exists $w_k \in C - x$ such that $Aw_k = -Au_k$ and $\|w_k\| \leqslant \gamma \|Au_k\|$. Then $(u_k + w_k) \in \operatorname{con}(C - x)$, $A(u_k + w_k) = 0$, and $(u_k + w_k) \to u$ as $k \to \infty$. Thus, $u \in \overline{\operatorname{con}}[(C - x) \cap \mathcal{N}(A)]$. This proves that

$$\overline{\operatorname{con}}(C - x) \cap \mathcal{N}(A) = \overline{\operatorname{con}}[(C - x) \cap \mathcal{N}(A)]$$

and completes the proof. ∎

The following result, which seems to be new, characterizes when a point is in the relative interior of a convex set.

LEMMA 3.9. *Let D be a convex set in a finite-dimensional space Y and $x \in D$. Then $x \in \operatorname{ri} D$ if and only if $(D - x)^0 = (D - x)^\perp$.*

Proof. Since $x \in \operatorname{ri} D$ if and only if $0 \in \operatorname{ri}(D - x)$, by replacing D with $D - x$, we may assume $x = 0$. We use the notation $B(\varepsilon)$ for the open ball in Y centered at the origin with radius ε. We have $0 \in \operatorname{ri} D$ if and only if there exists $\varepsilon > 0$ such that $B(\varepsilon) \cap \operatorname{aff} D \subset D$ if and only if there exists $\varepsilon > 0$ such that $B(\varepsilon) \cap \operatorname{span} D \subset D$ which implies $[B(\varepsilon) \cap \operatorname{span} D]^0 \supset D^0$. Using Lemmas 2.1, 3.4, and 2.2(2), this implies

$$D^0 \subset [\operatorname{con}\{B(\varepsilon) \cap \operatorname{span} D\}]^0 = [\operatorname{con} B(\varepsilon) \cap \operatorname{span} D]^0$$
$$= (\operatorname{span} D)^0 = (\operatorname{span} D)^\perp = D^\perp \subset D^0.$$

Thus $D^0 = D^\perp$.

Conversely, suppose $D^0 = D^\perp$. If $0 \notin \operatorname{ri} D$, then by setting $Y_0 = \operatorname{span} D = \operatorname{aff} D$, we see that $0 \notin \operatorname{int} D$ (relative to Y_0). Thus by the separation theorem (cf., e.g., [15, Theorem 11.6]), there exists $y \in Y_0 \setminus \{0\}$ such that $0 \geqslant \langle y, d \rangle$ for every $d \in \operatorname{int} D$. By continuity, $0 \geqslant \langle y, d \rangle$ for every $d \in \overline{\operatorname{int} D} = \bar{D}$ which implies $y \in (D)^0 = D^\perp = (\operatorname{span} D)^\perp = Y_0^\perp = \{0\}$. But this contradicts $y \neq 0$. ∎

As a consequence of this lemma, we easily obtain the following fact.

LEMMA 3.10. *The following statements are equivalent*:

(1) $b \in \operatorname{ri} A(C)$;

(2) $[A(C) - b]^0 = [A(C) - b]^\perp$; *and*

(3) $\mathscr{R}(A^*) \cap (C - x)^0 = \mathscr{R}(A^*) \cap (C - x)^\perp$ *for every* $x \in C \cap A^{-1}(b)$.

Proof. The equivalence of (1) and (2) follows by Lemma 3.9. Since $Ax = b$ for every $x \in C \cap A^{-1}(b)$, we have that, for such x, (2) holds if and only if $[A(C-x)]^0 = [A(C-x)]^\perp$ if and only if $(A^*)^{-1}(C-x)^0 = (A^*)^{-1}(C-x)^\perp$ if and only if (3) holds. ∎

THEOREM 3.11. *In a Hilbert space X, suppose that D is a closed convex cone, M is a finite-dimensional subspace, and $D \cap M$ is a subspace. Then $D + M$ is closed. In particular, if $(C - x)^0 \cap \mathscr{R}(A^*)$ is a subspace, then $(C - x)^0 + \mathscr{R}(A^*)$ is closed.*

Proof. Assume first that $D \cap M = \{0\}$. Let $x_n \in D + M$ and $x_n \to x$. We must show that $x \in D + M$. Write $x_n = d_n + y_n$, where $d_n \in D$ and $y_n \in M$. If some subsequence of $\{y_n\}$ is bounded, then, by passing to a subsequence if necessary, we may assume $y_n \to y \in M$. Then $d_n = x_n - y_n \to x - y$. Since D is closed, $d := x - y \in D$ and $x = d + y \in D + M$.

If no subsequence of $\{y_n\}$ is bounded, then $\|y_n\| \to \infty$. It follows that $\{y_n / \|y_n\|\}$ is bounded in M so by passing to a subsequence if necessary, we may assume that $y_n / \|y_n\| \to y$ and $\|y\| = 1$. Then $d_n / \|y_n\| \in D$ for every n and

$$\frac{d_n}{\|y_n\|} = \frac{x_n}{\|y_n\|} - \frac{y_n}{\|y_n\|} \to 0 - y \in D$$

since D is closed. Thus $-y \in D \cap M = \{0\}$ which contradicts $\|y\| = 1$.

This proves the theorem when $D \cap M = \{0\}$. In general, $V = D \cap M$ is a closed subspace of M so we can write $M = V + V^\perp$, where $V^\perp$ is the orthogonal complement of V in M. Then $D + M = D + V + V^\perp = D + V^\perp$ and $D \cap V^\perp = \{0\}$. By the first part of the proof (with $M = V^\perp$), $D + V^\perp$ is closed. Hence $D + M = D + V^\perp$ is closed. ∎

Remark. In the special case of Theorem 3.11 when $D \cap M = \{0\}$, we recover what some writers call the "Dieudonné separation theorem."

Now we can prove the second of our main results of this section.

THEOREM 3.12. *If $b \in \operatorname{ri} A(C)$, then for each $x \in X$, there exists $y \in Y$ such that*

$$A[P_C(x + A^* y)] = b. \tag{3.12.1}$$

Moreover,

$$P_K(x) = P_C(x + A^* y) \tag{3.12.2}$$

for each $y \in Y$ which satisfies (3.12.1).

Proof. If $b \in \operatorname{ri} A(C)$, then Lemma 3.10 implies that $\mathcal{R}(A^*) \cap (C - x)^0$ is a subspace. Theorem 3.11 implies that $(C - x)^0 + \mathcal{R}(A^*)$ is closed. It follows from Lemmas 3.8 and 3.1 that $\{C, A^{-1}(b)\}$ has the strong property CHIP. Now the conclusion follows from Theorem 3.2. ▮

Various special cases of this theorem were obtained earlier by: Micchelli and Utreras [13, Theorem 2.1] when C is a closed convex cone, $x = 0$, and $b \in \operatorname{int} A(C)$; [13, Theorem 2.2] when C is the translate of a closed convex cone, $x = 0$, and $b \in \operatorname{int} A(C)$; Chui, Deutsch, and Ward [2, Theorem 3.2] when C is a closed convex cone and $b \in \operatorname{int} A(C)$; Chui, Deutsch, and Ward [3, Theorem 2.3] when C is a closed convex cone; and by [3, Theorem 4.7] when $b \in \operatorname{int} A(C)$.

4. MINIMAL EXTREMAL SUBSET OF C

In this section we will show that there exists a certain convex extremal subset C_b of C with the property that $C_b \cap A^{-1}(b) = C \cap A^{-1}(b) =: K$ and $b \in \operatorname{ri} A(C_b)$. Then we can apply Theorem 3.12, with C replaced by C_b.

Recall that a convex subset E of a convex set D is called an *extremal subset* of D if $x, y \in D$, $0 < \lambda < 1$, and $\lambda x + (1 - \lambda) y \in E$ implies $x, y \in E$. Clearly, the intersection of any collection of extremal subsets of D is either empty or extremal in D. Also, D is trivially extremal in D.

DEFINITION 4.1. Let C_b denote the smallest closed convex extremal subset of C such that $C_b \supset K$, i.e.,

$$C_b \cap A^{-1}(b) = C \cap A^{-1}(b) := K.$$

More precisely,

$$C_b = \bigcap \{E \mid E \subset C, E \text{ closed convex extremal in } C, \text{ and } E \cap A^{-1}(b) = K\}.$$

The extremal set C_b will play the essential role in our main characterization theorem (Theorem 4.5) below. Let us first note that there is an alternate way of describing C_b. This will prove useful in recovering some known results in the special case when C is a convex cone.

DEFINITION 4.2. Let F_b denote the smallest closed convex extremal subset of $A(C)$ which contains b and set

$$C_{F_b} := C \cap A^{-1}(F_b).$$

In the special case when C is a cone, this definition was given in [3].

PROPOSITION 4.3. *With C_b and C_{F_b} defined as above, we have*

(1) $C_{F_b} = C_b$.

(2) $A(C_b) = F_b$.

(3) $b \in \operatorname{ri} A(C_b)$.

Proof. Clearly, C_{F_b} is closed and convex, since C is and A is linear and continuous. Now let $x, y \in C, 0 < \lambda < 1$, and suppose $z = \lambda x + (1-\lambda) y \in C_{F_b}$. Then $z \in C$ and $Az \in F_b$. Hence $\lambda Ax + (1-\lambda) Ay = Az \in F_b$. But $Ax, Ay \in A(C)$ and F_b is extremal in $A(C)$ implies that $Ax, Ay \in F_b$. Hence $x, y \in A^{-1}(F_b) \cap C = C_{F_b}$. Thus C_{F_b} is extremal in C.

Next we observe that the mapping $A_C := A|_C : C \to A(C)$ is surjective so that

$$A(C_{F_b}) = A_C(C_{F_b}) = A_C[C \cap A^{-1}(F_b)] = A_C[A_C^{-1}(F_b)] = F_b.$$

That is,

$$A(C_{F_b}) = F_b. \tag{4.3.1}$$

Now we verify that $b \in \operatorname{ri} F_b$, or equivalently that $0 \in \operatorname{ri}(F_b - b)$. Since $0 \in F_b - b$ and $\operatorname{aff}(F_b - b) = \operatorname{span}(F_b - b)$, by working in the space $Y_0 := \operatorname{span}(F_b - b)$ rather than Y, it is equivalent to show that $0 \in \operatorname{int}(F_b - b)$. If $0 \notin \operatorname{int}(F_b - b)$, then a well-known separation theorem (e.g., see [15, Theorem 11.6]) implies that there exists $y \in Y_0 \setminus \{0\}$ such that

$$\langle y, f - b \rangle \leqslant 0 \qquad \text{for all} \quad f \in F_b.$$

If $y \in (F_b - b)^{\perp}$, then $y \in [\operatorname{span}(F_b - b)]^{\perp} = Y_0^{\perp} = \{0\}$ which contradicts $y \neq 0$. Thus there is $f_0 \in F_b$ such that $\langle y, f_0 - b \rangle < 0$. Set

$$H = \{z \in Y_0 \mid \langle y, z \rangle = 0\}.$$

Then the hyperplane H supports $F_b - b$ at 0 and $E := H \cap (F_b - b)$ is a closed convex extremal subset of $F_b - b$ such that $E \neq F_b - b$ since $f_0 - b \in (F_b - b) \backslash E$. It follows that $E + b$ is a closed convex extremal subset of F_b with $b \in E + b$ and $E + b \neq F_b$. Since F_b is extremal in $A(C)$, $E + b$ is also extremal in $A(C)$. But this contradicts the minimality of F_b. This proves that $0 \in \operatorname{int}(F_b - b)$ and hence that

$$b \in \operatorname{ri} F_b = \operatorname{ri} A(C_{F_b}). \tag{4.3.2}$$

Since C_{F_b} is a closed convex extremal subset of C and $C_{F_b} \supset K$, it follows by the minimality of C_b that $C_b \subset C_{F_b}$. For the reverse inclusion, let $c_0 \in C_{F_b}$. Then $Ac_0 = d \in F_b$. By (4.3.2), we can choose $d' \in F_b$ and $0 < \lambda < 1$ so that $b = \lambda d' + (1 - \lambda)\, d$. Next choose $c' \in C_{F_b}$ so that $d' = Ac'$. Then $b = \lambda Ac' + (1 - \lambda)\, Ac_0 = A[\lambda c' + (1 - \lambda)\, c_0]$ implies that $c := \lambda c' + (1 - \lambda)\, c_0 \in K \subset C_b$. By extremality of C_b in C, it follows that $c_0, c' \in C_b$. Thus $C_{F_b} \subset C_b$. This proves (1) which, along with (4.3.1) and (4.3.2), verifies (2) and (3). ∎

Next we characterize when b is in the relative interior of $A(C)$.

Lemma 4.4. *$b \in \operatorname{ri} A(C)$ if and only if $C = C_b$.*

Proof. The "if" part follows from Lemma 4.2(3).

Conversely, suppose $b \in \operatorname{ri} A(C)$. It suffices to prove that $C_b \supset C$. Let $x \in C$ and set $y := Ax$. Since $b \in \operatorname{ri} A(C)$, there exists $y_1 \in A(C)$ and $0 < \lambda < 1$ so that $b = \lambda y + (1 - \lambda)\, y_1$. Choose $x_1 \in C$ such that $Ax_1 = y_1$. Then

$$A[\lambda x + (1 - \lambda)\, x_1] = \lambda y + (1 - \lambda)\, y_1 = b$$

implies that $\lambda x + (1 - \lambda)\, x_1 \in K \subset C_b$. By extremality of C_b in C, x and x_1 are in C_b. Thus $C \subset C_b$. ∎

As an immediate consequence of Lemma 4.3(3) and Theorem 3.12 (applied to C_b rather than C), we obtain

Theorem 4.5. *For each $x \in X$, there exists $y \in Y$ such that*

$$A[P_{C_b}(x + A^* y)] = b. \tag{4.5.1}$$

Moreover,

$$P_K(x) = P_{C_b}(x + A^* y) \tag{4.5.2}$$

for any $y \in Y$ *which satisfies* (4.5.1).

In many practical applications of Theorem 4.5, it is useful to know how to recognize or construct the set C_b. The following proposition is often useful in this regard.

PROPOSITION 4.6. *Suppose* E *is any closed convex extremal subset of* C *with the property that* $E \supset C \cap A^{-1}(b)$ (*i.e.,* $E \cap A^{-1}(b) = C \cap A^{-1}(b)$). *Then, for any* $x \in C \cap A^{-1}(b)$, *the following statements are equivalent*:

(1) $E = C_b$;

(2) $b \in \operatorname{ri} A(E)$;

(3) $[A(E) - b]^0 = [A(E) - b]^{\perp}$.

(4) $\mathscr{R}(A^*) \cap (E - x)^0 = \mathscr{R}(A^*) \cap (E - x)^{\perp}$;

(5) $A^* y \in (E - x)^{\perp}$ *whenever* $A^* y \in (E - x)^0$.

Proof. The equivalence of (1) and (2) is just Lemma 4.4 with C replaced by E. The equivalence of (2)–(5) is just Lemma 3.10 with C replaced by E. ∎

5. AN APPLICATION OF THEOREM 4.5

In any nontrivial applications of the theories of Sections 3 and 4, one must be able to compute $P_C(z)$ or $P_{C_b}(z)$ for any $z \in X$. In particular, if $b \notin \operatorname{ri} A(C)$, one must first be able to compute C_b before applying the main characterization Theorem 4.5.

The following example is indicative of what must be done to apply the theory of Section 4.

EXAMPLE 5.1. Let $X = L_2[0, 3]$, $\phi_i(t) = [1 - |t - i|]_+$ $(i = 1, 2)$,

$$C = \{x \in X \mid 0 \leqslant x(t) \leqslant 10 \text{ for almost all } t \in [0, 3]\},$$

and define $A \colon X \to l_2(2)$ by

$$Ax = (\langle x, \phi_1 \rangle, \langle x, \phi_2 \rangle), \qquad x \in X. \tag{5.1.1}$$

Then $A^* \colon l_2(2) \to X$ is given by

$$A^*(\alpha, \beta) = \alpha\phi_1 + \beta\phi_2. \tag{5.1.2}$$

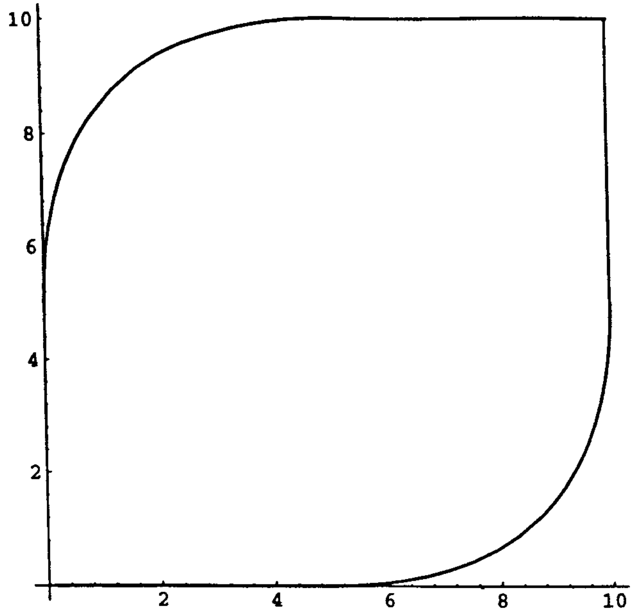

FIG. 1. The bounded convex region $A(C)$.

First observe that for each $x \in X$,

$$P_C(x) = [x]_0^{10}, \tag{5.1.3}$$

where $[x]_0^{10}$ is the "truncated" function on $[0, 3]$ defined by

$$[x]_0^{10}(t) := \begin{cases} 0, & \text{if } x(t) < 0 \\ x(t), & \text{if } 0 \leqslant x(t) \leqslant 10 \\ 10, & \text{if } x(t) > 10. \end{cases} \tag{5.1.4}$$

(This can easily be proved directly or deduced from Theorem 1.1.)

Next we compute the region $A(C)$. Note that *$A(C)$ is symmetric about the 45-degree line* $L = \{(\alpha, \alpha) \mid \alpha \in \mathbb{R}\}$. That is, if $(\alpha, \beta) \in A(C)$, then $(\beta, \alpha) \in A(C)$. To see this, choose $x \in C$ so that $Ax = (\alpha, \beta)$. Then the function $\tilde{x}$, defined by $\tilde{x}(t) = x(3 - t)$, $t \in [0, 3]$, is in C and $A\tilde{x} = (\beta, \alpha)$. This symmetry will simplify the description of $A(C)$ since now it suffices to only describe $A(C)$ above the line L (Fig. 1).

Since the functions $x = 0$, $x = 10$, $x = 10\chi_{[2, 3]}$, and $x = 10\chi_{[1, 3]}$ are all in C, it follows that $Ax \in A(C)$ for each of these functions x. Thus,

$$\{(0, 0), (10, 10), (0, 5), (5, 10)\} \subset A(C).$$

By symmetry, $(5, 0)$ and $(10, 5)$ are also in $A(C)$. Since $A(C)$ is convex, we have

$$co\{(0,0), (0,5), (5,0), (5,10), (10,5), (10,10)\} \subset A(C).$$

Moreover, it is easy to see that if $\beta > 5$, *then* $(0, \beta) \notin A(C)$. By symmetry, $(\alpha, 0) \notin A(C)$ *if* $\alpha > 5$. By a similar argument, $(\alpha, 10) \notin A(C)$ if $\alpha < 5$ *and* $(10, \beta) \notin A(C)$ *if* $\beta < 5$. Also, it is trivially true that if $(\alpha, \beta) \in A(C)$, then $\alpha, \beta \geqslant 0$ and $\alpha, \beta \leqslant 10$ (Fig. 1).

Finally, we will compute the upper boundary of $A(C)$ between the points $(0, 5)$ and $(5, 10)$. That is, if $0 < \alpha < 5$, we want to determine the *largest* number β so that $(\alpha, \beta) \in A(C)$.

PROPOSITION. *For any* $0 \leqslant \alpha \leqslant 5$, *the largest number* β *such that* $(\alpha, \beta) \in A(C)$ *is given by* $\beta = 5 - \alpha + 2\sqrt{5\alpha}$.

Proof. If $\alpha = 0$ or 5, the above argument shows that $\beta = 5$ or 10, respectively. Thus, we may assume $0 < \alpha < 5$. Let

$$\begin{aligned}\bar{\beta} = \bar{\beta}(\alpha) &:= \max\{\beta \mid \langle x, \phi_1 \rangle = \alpha, \langle x, \phi_2 \rangle = \beta, x \in C\} \\ &= \max\{\beta \mid Ax = (\alpha, \beta), x \in C\}.\end{aligned}$$

That is,

$$\bar{\beta} = \max\{\langle x, \phi_2 \rangle \mid x \in C, \langle x, \phi_1 \rangle = \alpha\}. \tag{5.1.5}$$

It suffices to show that $\bar{\beta} = 5 - \alpha + 2\sqrt{5\alpha}$.

Since $\phi_1, \phi_2 \geqslant 0$ and $\phi_1 = 0$ on $[2, 3]$, it is clear that the $x \in C$ which maximizes $\langle x, \phi_2 \rangle$ must satisfy $x = 10$ on $[2,3]$. For such x,

$$\langle x, \phi_2 \rangle = \int_0^2 x(t)\, \phi_2(t)\, dt + \int_2^3 10\phi_2(t)\, dt = \langle x\chi_{[0,2]}, \phi_2 \rangle + 5$$

and our problem can be restated as follows: determine

$$\begin{aligned}\bar{\beta} &= 5 + \max\{\langle x\chi_{[0,2]}, \phi_2 \rangle \mid x \in C, \langle x, \phi_1 \rangle = \alpha\} \\ &= 5 + \max\{\langle y, \phi_2 \rangle \mid 0 \leqslant y \leqslant 10, \operatorname{supp} y \subset [0, 2], \langle y, \phi_1 \rangle = \alpha\},\end{aligned}$$

where supp y is the *support* of y:

$$\operatorname{supp} y := \{t \in [0, 3] \mid y(t) \neq 0\}.$$

Let

$$D := \{y \in X \mid 0 \leqslant y \leqslant 10, \operatorname{supp} y \subset [0, 2], \langle y, \phi_1 \rangle = \alpha\}.$$

Then

$$\bar{\beta} = 5 + \max\{\langle y, \phi_2 \rangle \mid y \in D\}. \tag{5.1.6}$$

Since the set D is a weakly closed convex subset of X which is bounded, it is weakly compact. Since the linear mapping $y \mapsto \langle y, \phi_2 \rangle$ is weakly continuous, it attains its maximum on D at an extreme point of D. Thus we next describe the extreme points of D, ext D. It can be shown that

$$\text{ext } D = \{10\chi_\Omega \mid \Omega \subset [0, 2] \text{ is measurable}, \langle 10\chi_\Omega, \phi_1 \rangle = \alpha\}. \tag{5.1.7}$$

We omit the details of the derivation.

As noted above, the search for a maximum of $\langle y, \phi_2 \rangle$ over all $y \in D$ may be confined to $y \in \text{ext } D$. Thus from (5.1.6) we may rewrite $\bar{\beta}$ as

$$\bar{\beta} = 5 + \max\left\{\langle 10\chi_\Omega, \phi_2 \rangle \mid \Omega \subset [0, 2], \int_0^3 10\chi_\Omega \phi_1 = \alpha\right\}. \tag{5.1.8}$$

For any subset $\Omega \subset [0, 2]$, let $\Omega_1 = \Omega \cap [0, 1)$ and $\Omega_2 = \Omega \cap [1, 2]$. It follows from (5.1.8) that we may write

$$\bar{\beta} = 5 + \max\left\{\langle 10\chi_{\Omega_2}, \phi_2 \rangle \mid \Omega_1 \subset [0, 1), \Omega_2 \subset [1, 2], \int 10\chi_{\Omega_1}\phi_1 + \int 10\chi_{\Omega_2}\phi_1 = \alpha\right\}. \tag{5.1.9}$$

Since ϕ_1 is decreasing on $[1, 2]$ and ϕ_2 is increasing on $[1, 2]$, it follows from (5.1.9) that the search for a maximum may be further restricted:

$$\bar{\beta} = 5 + \max\left\{\langle 10\chi_{[\delta, 2]}, \phi_2 \rangle \mid \delta \in [1, 2), \int_\delta^2 10\phi_1(t)\, dt = \alpha\right\}. \tag{5.1.10}$$

Now $\delta \in [1, 2]$ and $\int_\delta^2 10\phi_1(t)\, dt = \alpha$ implies that

$$\alpha = \int_\delta^2 10(2-t)\, dt = 5(2-\delta)^2.$$

Solving this equation for δ, we obtain $\delta = 2 - \sqrt{\alpha/5}$. A substitution for δ into (5.1.10) yields

$$\bar{\beta} = 5 + \langle 10\chi_{[2-\sqrt{\alpha/5}, 2]}, \phi_2 \rangle = 5 + \int_{2-\sqrt{\alpha/5}}^2 10(t-1)\, dt = 5 - \alpha + 2\sqrt{5\alpha}$$

which proves the proposition. ∎

It follows that the upper boundary of the region $A(C)$ is given by the union of the curves $co\{(0,0),(0,5)\}$, $\{(\alpha, 5-\alpha+2\sqrt{5\alpha}) \mid 0<\alpha<5\}$, and $co\{(5,10),(10,10)\}$. The lower boundary is obtained using the symmetry of $A(C)$ about the 45-degree line. Furthermore, the extreme points of the upper boundary are given by $(0,0), (0,5), (\alpha, 5-\alpha+2\sqrt{5\alpha})$ for each $0<\alpha<5$, $(5,10)$, and $(10,10)$.

From this knowledge of $A(C)$, we can now easily compute C_b and $P_K(x)$ for various $b \in A(C)$ and $x \in X$. We exhibit this by a few specific examples.

Case 1. $b=(5,5)$.

Then $b \in \operatorname{int} A(C) = \operatorname{ri} A(C)$ so that by Theorem 3.12 and Eq. (5.1.3), we see that $C_b = C$ and

$$P_K(0) = P_C(\alpha_1\phi_1 + \alpha_2\phi_2) = [\alpha_1\phi_1 + \alpha_2\phi_2]_0^{10}$$

for any choice of scalars α_i satisfying

$$\langle [\alpha_1\phi_1 + \alpha_2\phi_2]_0^{10}, \phi_1\rangle = 5, \tag{5.1.11}$$

$$\langle [\alpha_1\phi_1 + \alpha_2\phi_2]_0^{10}, \phi_2\rangle = 5. \tag{5.1.12}$$

Since $0 \leqslant [\alpha_1\phi_1 + \alpha_2\phi_2]_0^{10}(t) \leqslant 10$ for all t, it follows that $0 \leqslant \alpha_i \leqslant 10$ $(i=1,2)$ and, hence,

$$[\alpha_1\phi_1 + \alpha_2\phi_2]_0^{10} = \alpha_1\phi_1 + \alpha_2\phi_2. \tag{5.1.13}$$

Substituting (5.1.13) into (5.1.11)–(5.1.12), we obtain a pair of linear equations for the α_i whose solution is $\alpha_1 = \alpha_2 = 6$. Thus,

$$P_K(0) = 6\phi_1 + 6\phi_2.$$

Case 2. $b=(0,0)$, $(5,10)$, or $(10,10)$.

In each case, $b \in \operatorname{ext} A(C)$ so that $F_b = \{b\}$ and C_b is the *singleton* set

$$C_b = C \cap A^{-1}(F_b) = \{0\}, \{10\chi_{[1,3]}\}, \qquad \text{or} \quad \{10\chi_{[0,3]}\},$$

respectively. Hence, for any $x \in X$,

$$P_K(x) = 0,\ 10\chi_{[1,3]}, \qquad \text{or} \quad 10\chi_{[0,3]},$$

respectively.

Case 3. $b = (\alpha, 5-\alpha+2\sqrt{5\alpha})$ for some $0<\alpha<5$.

Then $b \in \operatorname{ext} A(C)$ so $F_b = \{b\}$ and the argument given in the above proposition shows that the maximizer is uniquely given by $x = 10\chi_{[\delta,3]}$, where $\delta = 2 - \sqrt{\alpha/5}$. Thus

$$C_b = C \cap A^{-1}(F_b) = \{10\chi_{[2-\sqrt{\alpha/5}]}\},$$

and, hence,

$$P_K(x) = 10\chi_{[2-\sqrt{\alpha/5},\, 3]}$$

for every $x \in X$.

Case 4. $b = (7, 10)$.

Then $F_b = co\{(5, 10), (10, 10)\}$ and

$$\begin{aligned} C_b &= C \cap A^{-1}(F_b) = \{ y \in X \mid 0 \leqslant y(t) \leqslant 10,\ Ay \in F_b \} \\ &= \{ y \in X \mid 0 \leqslant y(t) \leqslant 10,\ 5 \leqslant \langle y, \phi_1 \rangle \leqslant 10,\ \langle y, \phi_2 \rangle = 10 \}. \end{aligned}$$

From this it is easy to deduce that

$$C_b = \{ y\chi_{[0,1)} + 10\chi_{[1,3]} \mid 0 \leqslant y(t) \leqslant 10 \}.$$

Moreover, for any $z \in X$,

$$P_{C_b}(z) = [z]_0^{10} \chi_{[0,1)} + 10\chi_{[1,3]}.$$

In particular,

$$\begin{aligned} P_K(0) = P_{C_b}(\alpha_1\phi_1 + \alpha_2\phi_2) &= [\alpha_1\phi_1 + \alpha_2\phi_2]_0^{10} \chi_{[0,1)} + 10\chi_{[1,3]} \\ &= [\alpha_1\phi_1]_0^{10} \chi_{[0,1)} + 10\chi_{[1,3]} \end{aligned}$$

for any α_1 which satisfies the equations

$$\begin{aligned} \langle [\alpha_1\phi_1]_0^{10} \chi_{[0,1)} + 10\chi_{[1,3]}, \phi_1 \rangle &= 7, \\ \langle [\alpha_1\phi_1]_0^{10} \chi_{[0,1)} + 10\chi_{[1,3]}, \phi_2 \rangle &= 10. \end{aligned}$$

Since the second equation is always satisfied, independent of the choice of α_1, we need only satisfy the first. Since $0 \leqslant [\alpha_1\phi_1]_0^{10}(t) \leqslant 10$ for all t, it follows that $0 \leqslant \alpha_1 \leqslant 10$ and, hence, that $[\alpha_1\phi_1]_0^{10} = \alpha_1\phi_1$. Substituting this expression into the first equation and solving for α_1, we obtain $\alpha = 1$ and thus

$$P_K(0) = \phi_1\chi_{[0,1)} + 10\chi_{[1,3]}.$$

6. A LINEARLY CONVERGENT DESCENT METHOD

Since all m-dimensional Hilbert spaces are isometric, without loss of generality *we may assume that* $Y = \mathbb{R}^m$ *in this section.* In both the main results of this paper (Theorems 3.6 and 3.12), the problem of determining the best approximation $P_K(x)$ is reduced to solving the (generally non-linear) equation

$$A[P_C(x + A^* y)] - b = 0 \tag{6.0.1}$$

for the unknown element $y \in Y$. This problem can be solved by standard optimization methods if one knows that the left side of Eq. (6.0.1) is the gradient of some convex function defined on $\mathbb{R}^m$. In this section, we will show that this is indeed the case when C is a polyhedron, and we will also describe a linearly convergent algorithm to solve this problem. For the first two results below, C may be any closed convex set in a Hilbert space X.

LEMMA 6.1. *For any* $w, z \in X$,

$$\langle P_C(w) - P_C(z), w - z \rangle \geqslant \|P_C(w) - P_C(z)\|^2.$$

Thus, P_C *is a monotone mapping.*

Proof. We have

$$\begin{aligned}\langle P_C(w) - P_C(z), w - z \rangle = {} & \langle P_C(w) - P_C(z), w - P_C(w) \rangle \\ & + \langle P_C(w) - P_C(z), P_C(w) - P_C(z) \rangle \\ & + \langle P_C(w) - P_C(z), P_C(z) - z \rangle.\end{aligned}$$

The first and third terms on the right are nonnegative by Theorem 1.1. The second term is $\|P_C(w) - P_C(z)\|^2$. ∎

COROLLARY 6.2. *Fix any* $x \in X$. *Then the function* $\varphi \colon \mathbb{R}^m \to \mathbb{R}^m$ *defined by*

$$\varphi(y) := AP_C(x + A^* y) - b \tag{6.2.1}$$

is a monotone mapping.

Proof. Let $w = x + A^* u$ and $z = x + A^* v$. Then, using Lemma 6.1,

$$\begin{aligned}\langle \varphi(u) - \varphi(v), u - v \rangle &= \langle AP_C(x + A^* u) - AP_C(x + A^* v), u - v \rangle \\ &= \langle P_C(x + A^* u) - P_C(x + A^* v), A^* u - A^* v \rangle \\ &= \langle P_C(w) - P_C(z), w - z \rangle \geqslant 0.\end{aligned}$$

Thus $\varphi(y)$ is a monotone mapping of y. ∎

THEOREM 6.3. *If C is a polyhedral set, then, for any fixed $x \in X$ and $b \in \mathbb{R}^m$, the following hold*:

(1) *The function $\varphi: \mathbb{R}^m \to \mathbb{R}^m$ defined as in* (6.2.1) *is a piecewise affine mapping.*

(2) *There exists a convex function $\Phi: \mathbb{R}^m \to \mathbb{R}$ such that*

$$\nabla\Phi(y) = AP_C(x + A^*y) - b,$$

where $\nabla\Phi$ denotes the gradient of Φ.

(3) *If $K := C \cap A^{-1}(b) \neq \emptyset$, then Φ has a global minimizer in $\mathbb{R}^m$. Moreover, $P_K(x) = P_C(x + A^*\bar{y})$ for any global minimizer $\bar{y}$ of Φ.*

Proof. Let $C := \{x \in X \mid \langle x, a_i \rangle \leqslant \beta_i, 1 \leqslant i \leqslant n\}$, where $a_i \in X$ and $\beta_i \in \mathbb{R}$. Consider the finite-dimensional subspace of X generated by $a_1, ..., a_n$:

$$X_0 := \operatorname{span}\{a_1, ..., a_n\}.$$

Then we claim that

$$(C \cap X_0) + X_0^{\perp} = C. \tag{6.3.1}$$

Obviously, $(C \cap X_0) + X_0^{\perp} \subset C$. Now let $w \in C$. Then $u := w - P_{X_0^{\perp}}(w) \in X_0$ by Theorem 1.1. Since $\langle a_i, u \rangle = \langle a_i, w \rangle$, it follows that $u \in C$. Thus, $w = u + P_{X_0^{\perp}}(w) \in (C \cap X_0) + X_0^{\perp}$. This completes the proof of (6.3.1).

In particular, $C = C - v$ for any $v \in X_0^{\perp}$. Thus for any $v \in X_0^{\perp}$ and $u \in X$, $P_C(u) = P_{C-v}(u) = P_C(u + v) - v$ or $P_C(u + v) = P_C(u) + v$. Therefore,

$$P_C(z) = P_C[P_{X_0}(z) + (z - P_{X_0}(z))] = P_C[P_{X_0}(z)] + (z - P_{X_0}(z)). \tag{6.3.2}$$

For $w \in C$, by (6.3.1), there exist $u \in C \cap X_0$ and $v \in X_0^{\perp}$ such that $w = u + v$. Hence,

$$\|P_{X_0}(z) - w\|^2 = \|P_{X_0}(z) - u - v\|^2 = \|P_{X_0}(z) - u\|^2 + \|v\|^2$$

which implies $P_C[P_{X_0}(z)] \in C \cap X_0$, so that $P_C[P_{X_0}(z)] = P_{C \cap X_0}[P_{X_0}(z)]$. Thus, from (6.3.2), we derive the following identity: for any $z \in X$,

$$P_C(z) = P_{C \cap X_0}[P_{X_0}(z)] + (z - P_{X_0}(z)). \tag{6.3.3}$$

Based on the piecewise affine property of $P_{C \cap X_0}$, we prove statements (1) and (2) simultaneously.

For any index subset J of $\{1, 2, ..., n\}$, define the following polyhedral sets in $X_0 \times \mathbb{R}^n$:

$$W_J := \Big\{ (z, \lambda) \in X_0 \times \mathbb{R}^n \mid \lambda \geqslant 0,\ \lambda_i = 0 \text{ for } i \notin J,$$

$$z - \sum_{j \in J} \lambda_j a_j \in C,\ \langle z - \sum_{j \in J} \lambda_j a_j, a_i \rangle = \beta_i \text{ for } i \in J \Big\}.$$

By Theorem 19.3 in [15], the linear projection of W_J onto X_0 is also polyhedral. That is,

$$X_J := \{ x \in X_0 \mid \text{there exists } \lambda \text{ such that } (x, \lambda) \in W_J \}$$

is a polyhedral subset of X_0. Again, by Theorem 19.3 in [15],

$$Y_J := \{ y \in Y \mid P_{X_0}(x + A^* y) \in X_J \}$$

is a polyhedral set since $P_{X_0}(x + A^* y)$ is an affine mapping of y. (Note that Y_J depends only on J.)

Next we prove that for any fixed index set J, $\varphi(y)$ is an affine mapping on Y_J and its derivative on Y_J is self-adjoint (or its Jacobian matrix is symmetric).

Let $C_J := \{ z \in X_0 \mid \langle z, a_i \rangle = \beta_i \text{ for } i \in J \}$. We claim that for every $y \in Y_J$,

$$P_C(x + A^* y) = P_{C_J}[P_{X_0}(x + A^* y)] + [(x + A^* y) - P_{X_0}(x + A^* y)]. \quad (6.3.4)$$

In fact, for $y \in Y_J$, $P_{X_0}(x + A^* y) \in X_J$ (by the definition of Y_J). Thus, there exists $\lambda \in \mathbb{R}^n$ such that $(z, \lambda) \in W_J$ with $z := P_{X_0}(x + A^* y)$. Let $u^* := z - \sum_{j \in J} \lambda_j a_j$. Then, by the definition of W_J,

$$u^* \in C \qquad \text{and} \qquad u^* \in C_J.$$

Moreover, $z - u^* = \sum_{j \in J} \lambda_j a_j \in (C_J - u^*)^0$. By Theorem 1.1,

$$u^* = P_{C_J}(z) = P_{C_J}[P_{X_0}(x + A^* y)].$$

On the other hand, $v^* := (x + A^* y) - P_{X_0}(x + A^* y) \in X_0^{\perp}$, i.e., $\langle v^*, a_i \rangle = 0$ for $1 \leqslant i \leqslant n$. Thus, $u^* + v^* \in C$. Since $\lambda_j \geqslant 0$ and $\langle u^* + v^*, a_j \rangle = \langle u^*, a_j \rangle = \beta_j$ whenever $\lambda_j > 0$, we have

$$(x + A^* y) - (u^* + v^*) = \sum_{j \in J} \lambda_j a_j \in [C - (u^* + v^*)]^0.$$

Again, by Theorem 1.1, $u^* + v^* = P_C(x + A^* y)$, which is equivalent to (6.3.4).

Since C_J is an affine set, P_{C_J} is an affine mapping. By (6.3.4), $\varphi(y) = AP_C(x + A^*y) - b$ is an affine mapping on Y_J. Let $G_J = \{z \in X_0 \mid \langle z, a_j \rangle = 0$ for $j \in J\}$. Then $C_J = G_J + z_o$ for any $z_0 \in C_J$ and, hence, for any $z \in X$, $P_{C_J}(z) = P_{G_J}(z - z_0) + z_0$. It follows from (6.3.4) that for $y_1, y_2 \in Y_J$

$$\varphi(y_1) - \varphi(y_2) = AP_{G_J} P_{X_0} A^*(y_1 - y_2) + AA^*(y_1 - y_2) - AP_{X_0} A^*(y_1 - y_2).$$

Since G_J is a subspace of X_0, $P_{G_J} P_{X_0} = P_{G_J}$. Therefore, we have

$$\varphi(y_1) - \varphi(y_2) = (AP_{G_J} A^* + AA^* - AP_{X_0} A^*)(y_1 - y_2) \qquad \text{for} \quad y_1, y_2 \in Y_J. \tag{6.3.5}$$

Since the orthogonal projection onto a subspace of a Hilbert space is a linear self-adjoint mapping, it follows from (6.3.5) that the derivative of $\varphi(y)$ for $y \in Y_J$ is the linear operator in parenthesis on the right of (6.3.5) and, hence, is self-adjoint.

To prove (1) we only have to show that $Y = \bigcup_J Y_J$. Let $y \in Y$. Define $w := P_{X_0}(x + A^*y) \in X_0$ and $J := \{j \mid \langle P_C(w), a_j \rangle = \beta_j\}$. Then, by Theorem 1.1 and the representation of the dual cone for a polyhedral set (cf. the proof of Lemma 3.3), there exist $\lambda_j \geqslant 0$ such that

$$w - P_C(w) = \sum_{j \in J} \lambda_j a_j.$$

Let $\lambda_i = 0$ for $i \notin J$. Then $(w, \lambda) \in W_J$. Thus, $w = P_{X_0}(x + A^*y) \in X_J$ and $y \in Y_J$. So φ is a piecewise affine mapping on Y.

Since the derivative of φ is self-adjoint (i.e., its Jacobian matrix is symmetric) on each polyhedral set Y_J, φ is a conservative field on Y (cf. [18, Theorem 2.6, p. 359]). As a consequence, there is a potential function Φ on Y such that $\nabla\Phi = \varphi$. Since $\nabla\Phi$ is a monotone mapping by Corollary 6.2, the function Φ is convex on Y.

Since $\Phi(y)$ is convex and differentiable, $\bar{y}$ is a global minimizer of $\Phi(y)$ if and only if $\nabla\Phi(y) = 0$. By Theorem 3.6, it follows that (3) holds. ∎

Remark. Note that (6.3.3) reduces the problem of determining $P_C(z)$ to a finite-dimensional problem, namely, determining $P_{X_0}(z)$ and $P_{C \cap X_0}[P_{X_0}(z)]$.

Algorithm 6.4 (A Steepest Descent Method). Let $0 \leqslant \alpha < 1$, $\beta > 0$, and $y_0 \in \mathbb{R}^m$. Generate a sequence of iterates y_{k+1} for $k = 0, 1, ...$ by the following steepest descent method with line search:

(1) Let $z_k = x + A^*y_k$;

(2) Let $d_k = b - AP_C(z_k)$;

(3) Find a stepsize $t_k > 0$ such that

$$\alpha\langle d_k, AP_C(z_k) - b\rangle \leqslant \langle d_k, AP_C(z_k + t_k A^* d_k) - b\rangle \leqslant 0; \quad \text{and}$$

(4) Set $y_{k+1} = y_k + \min\{t_k, \beta\}\, d_k$.

Note that the above algorithm does not require an explicit form of the potential for $AP_C(x + A^* y) - b$. Therefore, it can be applied to any closed convex subset C of X. However, our convergence analysis of iterates is based on the fact that $AP_C(x + A^* y) - b$ is the gradient of a convex quadratic spline function.

THEOREM 6.5. *Suppose that* C *is a polyhedral set in* X *and* $K := C \cap A^{-1}(b) \neq \emptyset$. *Let* $0 \leqslant \alpha < 1$ *and* $\beta > 0$ *be given. Then, for any initial point* $y_0 \in \mathbb{R}^m$ *and the sequence* $\{y_k\}$ *generated by Algorithm* 6.4, *there exists a solution* y^* *of* (6.0.1) *and positive constants* $0 \leqslant \lambda < 1$ *and* $\gamma > 0$ (*depending only on* K *and* y_0) *such that*

$$\|y_k - y^*\| \leqslant \gamma \cdot \lambda^k \qquad \textit{for} \quad k = 0, 1, 2, ... \tag{6.5.1}$$

Moreover, let $x_k = P_C(x + A^* y_k)$. *Then*

$$\|x_k - P_K(x)\| \leqslant \gamma\, \|A^*\| \cdot \lambda^k \qquad \textit{for} \quad k = 0, 1, 2, ... \tag{6.5.2}$$

Proof. A minor modification of the proof of Theorem 2.3 in [9] will yield a proof of Theorem 6.5. For easy reference, we give a complete proof here.

By Theorem 6.3, there is a convex function Φ such that $\nabla\Phi(y) = AP_C(x + A^* y) - b$ for all $y \in \mathbb{R}^m$. If $\alpha\langle d_k, \nabla\Phi(y_k)\rangle \leqslant \langle d_k, \nabla\Phi(y_{k+1})\rangle$, then, by Lemma 3.1 in [11], there exists a positive constant κ (depending only on Φ and α) such that

$$\left(\frac{\langle d_k, \nabla\Phi(y_k)\rangle}{\|d_k\|}\right)^2 \leqslant \kappa(\Phi(y_k) - \Phi(y_{k+1})).$$

Since $d_k = -\nabla\Phi(y_k)$, it follows that

$$\|\nabla\Phi(y_k)\|^2 \leqslant \kappa(\Phi(y_k) - \Phi(y_{k+1})). \tag{6.5.3}$$

If $0 > \alpha\langle d_k, \nabla\phi(y_k)\rangle > \langle d_k, \nabla\phi(y_{k+1})\rangle$, then by the definition of t_k, we obtain $y_{k+1} = y_k + \beta d_k$. Since $g(\theta) := \langle y_{k+1} - y_k, \nabla\Phi[y_k + \theta(y_{k+1} - y_k)]\rangle$ is the derivative of the convex function

$$\Phi[y_k + \theta(y_{k+1} - y_k)],$$

g is a nondecreasing function of θ. Therefore,

$$\alpha\langle d_k, \nabla\Phi(y_k)\rangle \geqslant \langle d_k, \nabla\Phi[y_k + \theta(y_{k+1} - y_k)]\rangle \quad \text{for } 0 \leqslant \theta \leqslant 1. \tag{6.5.4}$$

By the mean-value theorem and (6.5.4), there exists $0 < \theta_k < 1$ such that

$$\begin{aligned} \Phi(y_k) - \Phi(y_{k+1}) &= -\beta\langle d_k, \nabla\Phi[y_k + \theta_k(y_{k+1} - y_k)]\rangle \\ &\geqslant -\beta\alpha\langle d_k, \nabla\Phi(y_k)\rangle; \end{aligned}$$

i.e.,

$$\Phi(y_k) - \Phi(y_{k+1}) \geqslant \alpha\beta \, \|\nabla\phi(y_k)\|^2. \tag{6.5.5}$$

By the definition of y_k, we have

$$\|y_k - y_{k+1}\| = \|\min\{t_k, \beta\} \, \nabla\Phi(y_k)\| \leqslant \beta \, \|\nabla\Phi(y_k)\|. \tag{6.5.6}$$

By Theorem 2.1 in [9], it follows from (6.5.3), (6.5.5), and (6.5.6) that there exist a global minimizer y^* of $\Phi(y)$ (i.e., a solution of $AP_C(x + A^*y) = b$) and positive constants λ, γ with $0 \leqslant \lambda < 1$ such that (6.5.1) holds. The estimate (6.5.2) follows from (6.5.1), $P_K(x) = P_C(x + A^*y^*)$, and $\|P_C(u) - P_C(v)\| \leqslant \|u - v\|$ (cf. Lemma 6.1). ∎

Note that the solutions of (6.0.1) might form an unbounded subset of Y. Theorem 6.5 ensures that the iterates $\{y_k\}$ not only are bounded, but they also converge to one specific solution of (6.0.1). Different choices of α, β, and y_0 could result in sequences of iterates converging to different solutions of (6.0.1).

REFERENCES

1. J. P. Boyle and R. L. Dykstra, A method for finding projections onto the intersection of convex sets in Hilbert spaces, *in* "Advances in Order Restricted Statistical Inference," pp. 28–47, Lecture Notes in Statistics, Springer-Verlag, New York/Berlin, 1985.
2. C. K. Chui, F. Deutsch, and J. D. Ward, Constrained best approximation in Hilbert space, *Constr. Approx.* **6** (1990), 35–64.
3. C. K. Chui, F. Deutsch, and J. D. Ward, Constrained best approximation in Hilbert space II, *J. Approx. Theory* **71** (1992), 213–238, .
4. R. Cominetti, Metric regularity, tangent sets, and second-order optimality conditions, *Appl. Math. Optim.* **21** (1990), 265–287.
5. F. Deutsch, Interpolation from a convex subset of Hilbert space: A survey of some recent results, *in* "Approximation Theory, Wavelets and Applications" (S. P. Singh, Ed.), pp. 95–105, Kluwer Academic, Boston, 1995.
6. F. Deutsch, Dykstra's cyclic projections algorithm: The rate of convergence, *in* "Approximation Theory, Wavelets and Applications" (S. P. Singh, Ed.), pp. 87–94, Kluwer Academic, Boston, 1995.

7. F. Deutsch, V. Ubhaya, J. Ward, and Y. Xu, Constrained best approximation in Hilbert space III: Applications to n-convex functions, *Constr. Approx.* **12** (1996), 361–384.
8. N. Dunford and J. T. Schwartz, “Linear Operators Part I: General Theory,” Interscience, New York, 1958.
9. W. Li, Linearly convergent descent methods for unconstrained minimization of convex quadratic splines, *J. Optim. Theory Appl.* **86** (1995), 145–172.
10. W. Li, P. M. Pardalos, and C. G. Han, Gauss–Seidel method for least-distance problems, *J. Optim. Theory Appl.* **75** (1992), 487–500.
11. W. Li and J. Swetits, A new algorithm for solving strictly convex quadratic programs, *SIAM J. Optim.*, (1997), to appear.
12. C. Micchelli, P. Smith, J. Swetits, and J. Ward, Constrained L_p-approximation, *Constr. Approx.* **1** (1985), 93–102.
13. C. Micchelli and F. Utreras, Smoothing and interpolation in a convex subset of Hilbert space, *SIAM J. Sci. Statist. Comput.* **9** (1988), 728–746.
14. S. Robinson, Regularity and stability for convex multivalued functions, *Math. Oper. Res.* **1** (1976), 130–143.
15. R. T. Rockafellar, “Convex Analysis,” Princeton Univ. Press, Princeton, NJ, 1970.
16. P. W. Smith and H. Wolkowicz, A nonlinear equation for linear programming, *Math. Programming* **34** (1986), 235–238.
17. C. Ursescu, Multifunctions with closed convex graph, *Czech. Math. J.* **25** (1975), 438–441.
18. R. E. Williamson, R. H. Crowell, and H. F. Trotter, “Calculus of Vector Functions,” 2nd ed., Prentice–Hall, Englewood Cliffs, NJ, 1968.
19. K. Zhao, Best interpolation with convex constraints, *J. Approx. Theory* **73** (1993), 119–135.