Journal of Approximation Theory **93**, 415–433 (1998)
Article No. AT983174

Successive Approximate Algorithm for Best Approximation from a Polyhedron

Shusheng Xu

Department of Mathematics, Jiangnan University, Wuxi, Jiangsu, 214063, People's Republic of China

Communicated by Frank Deutsch

Received May 7, 1996; accepted in revised form May 1, 1997

Suppose K is the intersection of a finite number of closed half-spaces $\{K_i\}$ in a Hilbert space X, and $x \in X \backslash K$. Dykstra's cyclic projections algorithm is a known method to determine an approximate solution of the best approximation of x from K, which is denoted by $P_K(x)$. Dykstra's algorithm reduces the problem to an iterative scheme which involves computing the best approximation from the individual K_i. It is known that the sequence $\{x_j\}$ generated by Dykstra's method converges to the best approximation $P_K(x)$. But since it is difficult to find the definite value of an upper bound of the error $\|x_j - P_K(x)\|$, the applicability of the algorithm is restrictive. This paper introduces a new method, called the *successive approximate algorithm*, by which one can generate a finite sequence $x_0, x_1, ..., x_k$ with $x_k = P_K(x)$. In addition, the error $\|x_j - P_K(x)\|$ is monotone decreasing and has a definite upper bound easily to be determined. So the new algorithm is very applicable in practice.

1. INTRODUCTION

Suppose $K = \bigcap_{i=1}^{r} C_i$ is the nonempty intersection of a finite number of closed convex sets $C_1, ..., C_r$ in a Hilbert space X, and $x \in X \backslash K$. *Dykstra's cyclic projections algorithm* is a known method to determine an approximate solution of the best approximation of x from K, $P_k(x)$. Dykstra's algorithm essentially reduces the problem to an iterative scheme which involves computing the best approximation from the individual $C_1, ..., C_r$. According to Dykstra [1] and Boyle and Dykstra [2], the sequence $\{x_j\}$ generated by Dykstra's method, which is generally an infinite sequence except in some special cases, converges to $P_K(x)$. Then the efficacy of the method depends on the rate of convergence and one's ability of estimating the upper bound of $\|x_j - P_K(x)\|$.

In some simple cases, e.g., when all the C_i are subspaces, linear varieties (i.e., translates of subspaces), or half-spaces, one can determine by Dykstra's algorithm a sequence $\{x_j\}$ converging to $P_K(x)$ since it is easy to find the

0021-9045/98 $25.00

best approximation from each C_i. In fact, when all the C_i are subspaces, Dykstra's algorithm reduces to the *method of alternating projections* due to Halperin [3]. Error analyses were made by Smith, Solmon, and Wagner [4] and Kayalar and Weinert [5]. Furthermore, it can be shown that those error bounds remain valid if the subspaces are replaced by linear varieties. When all the C_i are half-spaces, i.e., K is a polyhedron in the Hilbert space X, certain "residual" vectors must be computed at each step of projection (but no such "residual" vectors appeared in the subspace case). Due to Deutsch and Hundal [6], the sequence $\{x_j\}$ generated by Dykstra's algorithm has an error bound of exponential type as

$$\|x_j - P_K(x)\| \leqslant \rho c^j,$$

where $\rho > 0$, $0 \leqslant c < 1$. Though [6] gave an upper bound less than 1 for the constant c, no estimation for ρ was given. So the applicability of Dykstra's algorithm for polyhedron approximation is restrictive unless we can find an active estimation for ρ.

Motivated by the fact that polyhedron approximation has many important applications (see, e.g., [6, Sect. 5]) this paper introduces a new method which we call the *successive approximate algorithm*. According to this algorithm, starting from an arbitrary point $x_0 \in K$ one can generate a finite sequence $x_0, x_1, ..., x_k$ with $x_k = P_K(x)$. Moreover, x_j $(j < k)$ can be considered to be an approximate solution of $P_K(x)$ because the error $\|x_j - P_K(x)\|$ is monotone decreasing and has a definite upper bound easily to be determined. So the new method is very applicable in practice.

We conclude this introduction by mentioning that usually it is not difficult to find a point x_0 in K for a given practical problem. Otherwise, one can get an $x_0 \in K$ by known *successive projection methods* (see, e.g., [7]).

2. MAIN RESULTS

Let X be a Hilbert space. For $i = 1, ..., r$ ($r \geqslant 2$ is a given integer), let $c_i \in \mathbb{R}$ and $f_i \in X$ with $\|f_i\| = 1$. Write

$$H_i := \{x \in X \mid \langle x, f_i \rangle = c_i\},$$

$$K_i := \{x \in X \mid \langle x, f_i \rangle \leqslant c_i\},$$

$$K := \bigcap_{i=1}^{r} K_i.$$

Assume $H_i \neq H_j$, if $i \neq j$, and K is nonempty. Since K is a closed convex set, for any given $x \in X$ there always exists a unique best approximation $P_K(x)$

of x from K. By a translation if necessary we may assume that x equals the origin O. We assume $O \notin K$ unless otherwise stated.

Suppose the dimension of the subspace $X_n := \text{span}\{f_i\}_{i=1}^r$ is n. Clearly, if $x \in K$ and the projection of x on X_n is x', then $x' \in K$ because $\langle x', f_i \rangle = \langle x' - x, f_i \rangle + \langle x, f_i \rangle \leqslant 0 + c_i$, $i \in \{1, ..., r\}$. For $x, y \in X$, by $[xy]$ we denote the set $\{(1-\lambda)\, x + \lambda y \mid \lambda \in [0, 1]\}$.

For any subset $I \subset \{1, ..., r\}$, denote the number of the elements of I by $|I|$ and write

$$H(I) := \bigcap_{i \in I} H_i,$$

$$K(I) := \bigcap_{i \in I} K_i,$$

$$P_I := P_{H(I)}(O).$$

For $x \in X$, let

$$I(x) := \{i \in \{1, ..., r\} \mid \langle x, f_i \rangle = c_i\}.$$

Based on Lemma 2 in the next section, P_I can be written as a linear combination of $\{f_i\}_{i \in I}$ if the $\{f_i\}_{i \in I}$ are linearly independent. So we can define

$$T(x) = \Big\{ I \subset I(x) \mid O < |I| < n,\ \{f_i\}_{i \in I} \text{ are linearly independent,}$$
$$\text{and } P_I \neq O \text{ can be written as } \sum_{i \in I} \alpha_i f_i \text{ with } \alpha_i \leqslant 0,\ i \in I \Big\}.$$

DEFINITION. Assume $k \geqslant 1$, $x_0, ..., x_k \in K$. If

$$x_1 = P_{K \cap [x_0' O]}(O),$$

where x_0' is the projection of x_0 on X_n, and for $j = 1, ..., k-1$ there exists $I_j \in T(x_j)$ such that

$$x_{j+1} = P_{K \cap [x_j P_{I_j}]}(P_{I_j}) \neq x_j, \tag{2.1}$$

then we call $x_0, ..., x_k$ a *successive approximate sequence* to O in K,

Obviously, if $x_0, ..., x_k$ is a successive approximate sequence then $x_0, ..., x_j$ is also, $1 \leqslant j < k$.

THEOREM. *For any point $x_0 \in K$, there exists a successive approximate sequence $x_0, x_1, ..., x_k$ to O in K with x_0 being its starting point, for which*

(i) $x_k = P_K(O)$;

(ii) *any successive approximate sequence* $x_0, x'_1, ..., x'_{k'}$ *to* O *in* K *satisfies* $k' \leqslant k$ *and* $x'_j = x_j$, $j = 1, ..., k'$;

$$\text{(iii)} \quad \begin{cases} \|x_1\| \leqslant \|x_0\|, \\ \|x_{j+1}\| < \|x_j\|, \qquad 1 \leqslant j < k, \end{cases} \tag{2.2}$$

$$0 < \|x_{j+1}\| - \|P_K(O)\| < \|x_{j+1}\| - \|P_{I_j}\|, \qquad 1 \leqslant j < k-1, \tag{2.3}$$

$$\|x_{j+1} - P_K(O)\| < \|x_j - P_K(O)\|, \qquad 1 \leqslant j < k, \tag{2.4}$$

and

$$\|x_{j+1} - P_K(O)\| < (\|x_{j+1}\|^2 - \|P_{I_j}\|^2)^{1/2}, \qquad 1 \leqslant j < k-1; \tag{2.5}$$

$$\text{(iv)} \quad k \leqslant 1 + \left[\binom{r}{1} + \cdots + \binom{r}{n-1}\right] - \left[\binom{r_+}{1} + \cdots + \binom{r_+}{n-1}\right], \tag{2.6}$$

where r_+ *denotes the number of the elements of the set*

$$I_+ := \{i \in \{1, ..., r\} \mid c_i \geqslant 0\},$$

and $\binom{r_+}{j}$ *is defined to be zero if* $j > r_+$.

The successive approximate sequence $x_0, ..., x_k$ in the Theorem can be generated by the following algorithm:

ALGORITHM FOR A SUCCESSIVE APPROXIMATE SEQUENCE. **Step** 0. For the given point $x_0 \in K$, compute its projection x'_0 on X_n and the best approximation of O from $K \cap [x'_0 O]$.

Write

$$x_1 = P_{K \cap [x'_0 O]}(O), \qquad T_1 = T(x_1), \qquad d_1 = 0.$$

Let $j \leftarrow 1$ and go to Step j.

Step j. For each $I \in T_j$ compute P_I, if $\|P_I\| > d_j$ then determine $P_{K \cap [x_j P_I]}(P_I)$.

Case j.1. If there exists an $I_j \in T_j$ for which

$$\|P_{I_j}\| > d_j \tag{j.1-1}$$

and

$$P_{K \cap [x_j P_{I_j}]} = (P_{I_j}) \qquad \text{or} \qquad \neq x_j, \tag{j.1-2}$$

then

Case j.1.1. if $P_{K\cap[x_jP_{I_j}]}(P_{I_j})=P_{I_j}=x_j$, let $k=j$ and end;

Case j.1.2. if $P_{K\cap[x_jP_{I_j}]}(P_{I_j})=P_{I_j}\neq x_j$, let $x_{j+1}=P_{K\cap[x_jP_{I_j}]}(P_{I_j})$, $k=j+1$ and end;

Case j.1.3. if $P_{K\cap[x_jP_{I_j}]}(P_{I_j})\neq P_{I_j}$), let $x_{j+1}=P_{k\cap[x_jP_{I_j}]}(P_{I_j})$ and

$$T_{j+1}=\{I\notin T_1\cup\cdots\cup T_j\mid I\in T(x_{j+1})\},$$

$$d_{j+1}=\|P_{I_j}\|$$

and $j\leftarrow j+1$. Go to Step j;

Case j.2. If there is no $I_j\in T_j$ satisfying (j.1-1) and (j.1-2), let $k=j$ and end.

3. PROOF OF MAIN RESULTS

Firstly, we point out that by Theorem 2.4.2 of [8] one can compute the projection of x_0 on the finite dimensional subspace X_n, and by following Lemma 1 and Lemma 2 one can complete the calculation of the above Algorithm for a Successive Approximate Sequence.

LEMMA 1. *For any* $x\in K$ *and* $y\in X$,

$$P_{K\cap[xy]}(y)=(1-\lambda)\,x+\lambda y, \tag{3.1}$$

where $\lambda=1$ *if*

$$\hat{I}:=\{i\in\{1,...,r\}\mid\langle y,f_i\rangle>c_i\}$$

is empty, *otherwise*

$$\lambda=\min\left\{\lambda_i\mid\lambda_j=\frac{c_i-\langle x,f_i\rangle}{\langle y,f_i\rangle-\langle x,f_i\rangle},\ i\in\hat{I}\right\}$$

and $\lambda\in[0,1)$.

Proof. Obviously, (3.1) holds if $\hat{I}=\varnothing$ and $\lambda=1$. If $\hat{I}\neq\varnothing$, it is easy to check

$$\langle(1-\lambda_i)\,x+\lambda_i y,f_i\rangle=c_i,\qquad i\in\hat{I},$$

and (3.1) holds clearly. ∎

LEMMA 2. *If* $I=\{i_1, ..., i_s\} \subset \{1, ..., r\}$, *and* $\{f_{i_j}\}_{j=1}^{s}$ *are linearly independent, then*

$$P_I = \sum_{j=1}^{s} \alpha_{i_j} f_{i_j}$$

and

$$\|P_I\|^2 = \sum_{j=1}^{s} \sum_{m=1}^{s} \alpha_{i_j}\alpha_{i_m} \langle f_{i_j}, f_{i_m}\rangle, \tag{3.2}$$

where

$$\alpha_{i_j} = \frac{G_{i_j}(i_1, ..., i_s, (c_{i_1}, ..., c_{i_s}))}{G(i_1, ..., i_s)},$$

$$G(i_1, ..., i_s) = \begin{vmatrix} \langle f_{i_1}, f_{i_1}\rangle & \langle f_{i_1}, f_{i_2}\rangle & \cdots & \langle f_{i_1}, f_{i_s}\rangle \\ \langle f_{i_2}, f_{i_1}\rangle & \langle f_{i_2}, f_{i_2}\rangle & \cdots & \langle f_{i_2}, f_{i_s}\rangle \\ \cdots & & & \\ \langle f_{i_s}, f_{i_1}\rangle & \langle f_{i_s}, f_{i_2}\rangle & \cdots & \langle f_{i_s}, f_{i_s}\rangle \end{vmatrix},$$

$$G_{i_j}(i_1, ..., i_s, (c_{i_1}, ..., c_{i_s})) = \begin{vmatrix} \langle f_{i_1}, f_{i_1}\rangle & \cdots & \langle f_{i_1}, f_{i_{j-1}}\rangle & c_{i_1} & \langle f_{i_1}, f_{i_{j+1}}\rangle & \cdots & \langle f_{i_1}, f_{i_s}\rangle \\ \langle f_{i_2}, f_{i_1}\rangle & \cdots & \langle f_{i_2}, f_{i_{j-1}}\rangle & c_{i_2} & \langle f_{i_2}, f_{i_{j+1}}\rangle & \cdots & \langle f_{i_2}, f_{i_s}\rangle \\ & \cdots & & & & & \\ \langle f_{i_s}, f_{i_1}\rangle & \cdots & \langle f_{i_s}, f_{i_{j-1}}\rangle & c_{i_s} & \langle f_{i_s}, f_{i_{j+1}}\rangle & \cdots & \langle f_{i_s}, f_{i_s}\rangle \end{vmatrix}.$$

Proof. Let $P=\sum_{j=1}^{s} \alpha_{i_j} f_{i_j}$. By Cramer's rule we have

$$\langle P, f_{i_j}\rangle = c_{i_j}, \qquad j=1, ..., s.$$

So $P \in H(I)$.

For any $x \in H(I)$, by $\langle x, f_{i_j}\rangle = c_{i_j}$, $j=1, ..., s$, we have

$$\langle x-P, P\rangle = \sum_{j=1}^{s} \alpha_{i_j} \langle x-P, f_{i_j}\rangle = \sum_{j=1}^{s} \alpha_{i_j}(\langle x, f_{i_j}\rangle - \langle P, f_{i_j}\rangle) = 0,$$

and hence

$$\|x\|^2 = \|x-P+P, x-P+P\|^2 = \|x-P\|^2 + 2\langle x-P, P\rangle + \|P\|^2 \geqslant \|P\|^2,$$

which implies $P=P_I$. Equation (3.2) holds obviously. ∎

To prove our Theorem, we need the following lemmas. Especially, Lemma 3 does not need the hypothesis of $O \notin K$.

LEMMA 3. *$x^* \in K$ is the best approximation $P_K(O)$ if and only if there exists $\alpha_i \leqslant 0$, $i \in I(x^*)$ for which*

$$x^* = \sum_{i \in I(x^*)} \alpha_i f_i.$$

Proof. For any set $S \subset X$, we write the closure of S as $\bar{S}$, and

$$S^\circ := \{x \in X \mid \langle x, y \rangle \leqslant 0, \forall y \in S\},$$

$$\mathrm{cc}(S) := \left\{ x \mid x = \sum_{i=1}^{m} \lambda_i y_i,\ y_i \in S,\ \lambda_i \geqslant 0,\ m \in \mathbb{N} \right\}.$$

Then Proposition (6.9.2) in [8] shows

$$s^{\circ\circ} = \overline{\mathrm{cc}(S)}. \tag{3.3}$$

It is well known that $x^* = P_K(O)$ iff

$$-x^* \in (K - x^*)^\circ.$$

So it is sufficient to prove

$$(K - x^*)^\circ = \mathrm{cc}\{f_i \mid i \in I(x^*)\}. \tag{3.4}$$

In fact, (a) if $f = \sum_{i \in I(x^*)} \lambda_i f_i$, $\lambda_i \geqslant 0$, then for any $x \in K - x^*$ from $\langle x^* + x, f_i \rangle \leqslant c_i$ and $\langle x^*, f_i \rangle = c_i$, $i \in I(x^*)$ we have

$$\langle f, x \rangle = \sum_{i \in I(x^*)} \lambda_i \langle f_i, x \rangle = 0.$$

So $f \in (K - x^*)^\circ$ and

$$\mathrm{cc}\{f_i \mid i \in I(x^*)\} \subset (K - x^*)^\circ. \tag{3.5}$$

(b) On the contrary, assume $f \in (K - x^*)^\circ$. Suppose x is an arbitrary point of $\bigcap_{i \in I(x^*)} (K_i - x^*)$. Then for $i \in I(x^*)$, by $x \in K_i - x^*$ and the definition of $I(x^*)$ it follows that

$$\langle x, f_i \rangle = \langle x + x^*, f_i \rangle - \langle x^*, f_i \rangle \leqslant c_i - c_i = 0.$$

Thus

$$\langle x^* + \varepsilon x, f_i \rangle \leqslant c_i, \qquad i \in I(x^*), \varepsilon > 0. \tag{3.6}$$

Since $\langle x^*, f_i \rangle < c_i$ for any $i \notin I(x^*)$, there exists an $\varepsilon > 0$ such that

$$\langle x^* + \varepsilon x, f_i \rangle \leqslant c_i, \qquad i \notin I(x^*).$$

Combined with (3.6) we see that

$$\varepsilon x \in K - x^*.$$

So we have

$$\langle x, f \rangle = \frac{1}{\varepsilon} \langle \varepsilon x, f \rangle \leqslant 0.$$

That is,

$$f \in \Big[\bigcap_{j \in I(x^*)} (K_i - x^*) \Big]^\circ. \tag{3.7}$$

Note that $y \in [\mathrm{cc}\{ f_i \mid i \in I(x^*)\}]^\circ$ implies $\langle y, f_i \rangle \leqslant 0$, $i \in I(x^*)$ and hence $y \in \bigcap_{i \in I(x^*)} (K_i - x^*)$. We have

$$[\mathrm{cc}\{ f_i \mid i \in I(x^*)\}]^\circ \subset \bigcap_{i \in I(x^*)} (K_i - x^*).$$

Noting (3.7), (3.3), and the fact that $I(x^*)$ is a finite set we conclude that

$$\begin{aligned} f \in \Big[\bigcap_{i \in I(x^*)} (K_i - x^*) \Big]^\circ &\subset [\mathrm{cc}\{ f_i \mid i \in I(x^*)\}]^{\circ\circ} \\ &= \overline{\mathrm{cc}\{ f_i \mid i \in I(x^*)\}} = \mathrm{cc}\{ f_i \mid i \in I(x^*)\}. \end{aligned}$$

Combined with (3.5) we get (3.4). ∎

LEMMA 4. *If $x \in K$ and $I \in T(x)$, then*

$$K \subset K_I, \qquad H(I) \subset H_I, \tag{3.8}$$

and

$$P_{K_I}(O) = P_I, \tag{3.9}$$

where

$$K_I = \left\{ y \in X \,\middle|\, \left\langle y, -\frac{P_I}{\|P_I\|} \right\rangle \leqslant -\|P_I\| \right\},$$

$$H_I = \left\{ y \in X \,\middle|\, \left\langle y, -\frac{P_I}{\|P_I\|} \right\rangle = -\|P_I\| \right\}.$$

Proof. Based on the definition of $T(x)$ we have

$$P_I = \sum_{i\in I} \alpha_i f_i \neq 0, \qquad \alpha_i \leqslant 0.$$

So for any $y \in K$, the fact that $P_I \in H(I)$ gives

$$\left\langle y, -\frac{P_I}{\|P_I\|}\right\rangle = \sum_{i\in I} \frac{-\alpha_i}{\|P_I\|} \langle y, f_i\rangle \leqslant -\frac{1}{\|P_I\|} \sum_{i\in I} \alpha_i c_i$$
$$= -\frac{1}{\|P_I\|} \sum_{i\in I} \alpha_i \langle f_i, P_I\rangle = -\|P_I\|,$$

which implies $K \subset K_I$. Proof of $H(I) \subset H_I$ is similar.

For any $y \in K_I$, the definition of K_I implies $\langle y, -P_I\rangle \leqslant -\langle P_I, P_I\rangle$ which is $\langle y - P_I, P_I\rangle \geqslant 0$. Thus

$$\|y\|^2 = \|y - P_I\|^2 + 2\langle y - P_I, P_I\rangle + \|P_I\|^2 \geqslant \|P_I\|^2.$$

So from $P_I \in H_I$ we have (3.9). ∎

LEMMA 5. *If $x^* = P_K(O)$, then there exists an nonempty $I^* \subset I(x^*)$ such that $\{f_i\}_{i\in I(x^*)}$ are linearly independent and*

$$x^* = \sum_{i\in I^*} \alpha_i f_i \qquad \text{with} \quad \alpha_i \leqslant 0, \quad i \in I^*, \tag{3.10}$$

moreover

$$x^* = P_{I^*}. \tag{3.11}$$

Proof. Based on Lemma 3, there exists at least one nonempty subset $I \subset I(x^*)$ such that x^* can be written as a linear combination of $\{f_i\}_{i\in I}$ with negative coefficients. If there exist more than one such subsets, take one that has least elements and denote it as I'. Then

$$x^* = \sum_{i\in I'} \alpha_i f_i, \qquad \text{with} \quad \alpha_i < 0, \quad i \in I'.$$

It is not difficult to show that the $\{f_i\}_{i\in I'}$ are linearly independent. In fact, if there exists a set $\{a_i\}_{i\in I'} \subset \mathbb{R}$ that at least one of the elements does not equal zero and

$$\sum_{i\in I'} a_i f_i = 0,$$

then

$$x^* = \sum_{i \in I'} (\alpha_i - \alpha a_i)\, f_i \tag{3.12}$$

for any $\alpha \in \mathbb{R}$. Let $\alpha = \alpha_{i_0}/a_{i_0}$ with $i_0 \in I'$ satisfy

$$\left|\frac{\alpha_{i_0}}{a_{i_0}}\right| = \min_{i \in I'} \left|\frac{\alpha_i}{a_i}\right|.$$

Then

$$|\alpha a_i| \leqslant \left|\frac{\alpha_i}{a_i}\right| \cdot |a_i| = |\alpha_i|, \qquad i \in I'.$$

So on the right hand side of (3.12) the coefficient of f_i is zero if $i = i_0$ and not larger than zero if $i \neq i_0$, which contradicts the definition of I'.

Now take a subset $I^* \subset I(x^*)$ such that $I' \subset I^*$ and $\{f_i\}_{i \in I^*}$ is a maximal linearly independent subset of $\{f_i\}_{i \in I(x^*)}$. Then (3.10) holds.

Suppose on the contrary that $x^* \neq P_{I^*}$. Then by $x^* \in H(I^*)$ we have

$$\|P_{I^*}\| < \|x^*\|. \tag{3.13}$$

It is easy to show that for $i \in I(x^*)$

$$\langle P_{I^*}, f_i \rangle = c_i = \langle x^*, f_i \rangle. \tag{3.14}$$

Actually, (3.14) holds for $i \in I^*$ obviously. For $i \in I(x^*) \backslash I^*$, writing

$$f_i = \sum_{j \in I^*} a_j f_j$$

we have

$$\begin{aligned} \langle P_{I^*}, f_i \rangle &= \sum_{j \in I^*} a_j \langle P_{I^*}, f_j \rangle = \sum_{j \in I^*} a_j c_j \\ &= \sum_{j \in I^*} a_j \langle x^*, f_j \rangle = \langle x^*, f_i \rangle = c_i. \end{aligned}$$

So (3.14) holds for any $i \in I(x^*)$. Since

$$\langle x^*, f_i \rangle < c_i, \qquad i \notin I(x^*),$$

there exists a $\lambda > 0$ for which

$$x_\lambda := (1-\lambda)\, x^* + \lambda P_{I^*} \in K_i, \qquad i = 1, ..., r.$$

But by (3.13)

$$\|x_\lambda\| < \|x^*\|$$

which contradicts the hypothesis of $x^* = P_K(O)$. ∎

LEMMA 6. *If* $x \in K \cap X_n$, $O \notin K(I(x))$, *and*

$$P_{K \cap [xP_I]}(P_I) = x$$

for any $I \in T(x)$, *then*

$$x = P_K(O).$$

Proof. Assume the best approximation of O from $K(I(x))$ is x^*, then $x^* \neq 0$. Using Lemma 5 to the polyhedron $K(I(x))$ we can get a nonempty subset

$$I^* \subset I(x^*) \cap I(x) \tag{3.15}$$

for which the $\{f_i\}_{i \in I^*}$ are linearly independent and (3.10) and (3.11) hold.

If $|I^*| < n$, then $I^* \in T(x)$ and by the hypothesis

$$P_{K \cap [xP_{I^*}]}(P_{I^*}) = P_{K \cap [xx^*]}(x^*) = x.$$

Hence from Lemma 1 we can find a $\lambda \in [0, 1]$ for which

$$(1-\lambda)\, x + \lambda x^* = x.$$

Provided $\lambda = 0$, then there exists an

$$i \in \hat{I} := \{i \mid \langle x^*, f_i \rangle \geqslant c_i\} \tag{3.16}$$

such that

$$0 = \lambda = \lambda_i = \frac{c_i - \langle x, f_i \rangle}{\langle x^*, f_i \rangle - \langle x, f_i \rangle}.$$

So $\langle x, f_i \rangle = c_i$ which implies $i \in I(x)$. So by (3.16) we have $x^* \notin K(I(x))$ which contradicts the definition of x^*. Now we obtain $\lambda \neq 0$ and hence

$$x = x^*. \tag{3.17}$$

If $|I^*| = n$, then by $x \in X_n$, (3.15), (3.10), and the linear independence of $\{f_i\}_{i \in I^*}$ we have

$$x \in \left(\bigcap_{i \in I(x)} H_i \right) \cap X_n \subset \bigcap_{i \in I^*} (H_i \cap X_n) = \{x^*\}$$

which implies (3.17) too. So

$$x = P_{K(I(x))}(O),$$

and hence

$$x = P_K(O)$$

because $K \subset K(I(x))$. ∎

LEMMA 7. *Assume* $x \in K$, $I' \in T(x)$, *and*

$$P_{K \cap [xP_{I'}]}(P_{I'}) \neq x.$$

Then for any $I \in T(x)$

$$\|P_{I'}\| \geqslant \|P_I\|,$$

and if in addition $P_I \neq P_{I'}$ *then*

$$\|P_{I'}\| > \|P_I\|. \tag{3.18}$$

Proof. Firstly, we consider the case that both $I' = \{i'\}$ and $I = \{i\}$ are subsets having only one element.

Suppose

$$\langle P_{I'}, f_i \rangle > c_i.$$

Since $I = \{i\} \in T(x)$ implies $i \in I(x)$, so $\langle x, f_i \rangle = c_i$. Based on the hypothesis and Lemma 1 there exists a $\lambda \in (0.1]$ such that

$$\langle P_{K \cap [xP_{I'}]}(P_{I'}), f_i \rangle = \langle (1 - \lambda)\, x + \lambda P_{I'}, f_i \rangle > c_i,$$

which contradicts the fact that $P_{K \cap [xP_{I'}]}(P_{I'}) \in K$. So

$$P_{I'} \in K_i. \tag{3.19}$$

Since Lemma 2 and the hypothesis of $\|f_i\| = 1$ imply

$$P_I = c_i f_i,$$

from the definition of $T(x)$ we have $c_i < 0$. Using Lemma 3 to K_i and P_I we can find that P_I is the best approximation to O from K_i. Thus by (3.19),

$$\|P_{I'}\| \geqslant \|P_I\|.$$

If in addition $P_I \neq P_{I'}$, then from the uniqueness of the best approximation to O from K_i we get (3.18).

Secondly, in the general case, using Lemma 4,

$$K \subset K_{I'}, \qquad K \subset K_I, \qquad x \in H(I') \subset H_{I'}, \qquad x \in H(I) \subset H_I.$$

So applying the above approach to $K = K_1 \cap \cdots \cap K_r \cap K_{I'} \cap K_I$, i.e., with f_i, c_i, K_i substituted by $-P_I/\|P_I\|$, $-\|P_I\|$, K_I and $f_{i'}$, $c_{i'}$, $K_{i'}$ substituted by $-P_{I'}/\|P_{I'}\|$, $-\|P_{I'}\|$, $K_{I'}$ respectively, we get the conclusion required. ∎

Proof of the Theorem. It is not difficult to check that for the sequence $x_0, ..., x_k$ generated by the Algorithm for a Successive Approximate Sequence, (2.1) holds if $1 \leqslant j < k$. In fact, since x_{j+1} is generated either in Case *j*.1.2 or in Case *j*.1.3, (2.1) is immediate in the first case, and in the latter case from (*j*.1-2) we have $x_{j+1} \neq x_j$, which is (2.1). So $x_0, ..., x_k$ is a successive approximate sequence to O in K.

(i) If the algorithm ends in Case *k*.1.1, then $x_k = P_{I_k}$ with $I_k \in T_k$. Thus by Lemma 3 we have $x_k = P_K(O)$. A similar consideration leads to $x_k = P_K(O)$ if the algorithm ends in Case $(k-1)$.1.2.

Provided the algorithm ends in Case *k*.2, if $O \notin K(I(x_k))$ and

$$P_{K \cap [x_k P_I]}(P_I) = x_k, \qquad I \in T(x_k), \tag{3.20}$$

then from $\{x_j\}_{j=1}^{k} \subset X_n$ and Lemma 6 we have $x_k = P_K(O)$. In fact, when $k = 1$, by the fact that $x_1 = P_{K \cap [x_0' O]}(O)$, $x_0' \in K$, and $O \notin K$ there must be an $i \in \{1, ..., r\}$ such that $x_1 \in H_i$ but $O \notin K_i$. So $O \notin K(I(x_1))$. When $k > 1$, obviously x_k must be generated in Case $(k-1)$.1.3 where $I_{k-1} \in T_{k-1} \subset T(x_{k-1})$. Noting the definition of $T(x)$ and the fact that $P_{I_{k-1}} \in K(I_{k-1})$, by Lemma 3 (used to $K(I_{k-1})$) we have $P_{I_{k-1}} = P_{K(I_{k-1})}(O)$. But $P_{I_{k-1}} \neq O$, so $O \notin K(I_{k-1})$. Since x_{k-1}, $P_{I_{k-1}} \in H(I_{k-1})$ and $x_k = P_{K \cap [x_{k-1} P_{I_{k-1}}]}(P_{I_{k-1}})$ implies

$$I_{k-1} \subset I(x_k). \tag{3.21}$$

we have $K(I_{k-1}) \supset K(I(x_k))$ and $O \notin K(I(x_k))$.

Now it remains to prove (3.20). Suppose on the contrary that there exists an

$$I_k \in T(x_k) \tag{3.22}$$

for which

$$P_{K \cap [x_k P_{I_k}]}(P_{I_k}) \neq x_k. \tag{3.23}$$

If $k = 1$, since the definition of $T(x_k)$ implies $P_{I_k} \neq O$, by $d_1 = 0$ we have

$$\|P_{I_k}\| > d_k. \tag{3.24}$$

If $k>1$, then similar to the approach above there exists an $I_{k-1}\in T_{k-1}\subset T(x_{k-1})$ such that (3.21) holds. From the definition of $T(x)$ we obtain $I_{k-1}\in T(x_k)$. Thus from (3.22), (3.23), and Lemma 7 we will get $\|P_{I_k}\|>\|P_{I_{k-1}}\|=d_k$, which is (3.24), if we can show $P_{I_{k-1}}\neq P_{I_k}$. In fact, if $P_{I_{k-1}}=P_{I_k}$, then by the fact that

$$x_k=P_{K\cap[x_{k-1}P_{I_{k-1}}]}(P_{I_{k-1}})\neq P_{I_{k-1}}$$

there exists a $\lambda'\in[0,1)$ such that

$$x_k=(1-\lambda')\,x_{k-1}+\lambda' P_{I_{k-1}}, \tag{3.25}$$

and

$$(1-\lambda)\,x_{k-1}+\lambda P_{I_{k-1}}\notin K, \qquad \forall\lambda\in(\lambda',1]. \tag{3.26}$$

But by (3.23) it follows that

$$P_{K\cap[x_kP_{I_k}]}(P_{I_k})=(1-\lambda'')\,x_k+\lambda'' P_{I_{k-1}}\neq x_k,$$

where $\lambda''>0$. Substituting x_k in the above expression by (3.25) we obtain

$$P_{K\cap[x_kP_{I_k}]}(P_{I_k})=(1-\lambda'')(1-\lambda')\,x_{k-1}+[(1-\lambda'')\,\lambda'+\lambda'']\;P(I_{k-1})\in K,$$

which contradicts (3.26).

Now, if $I_k\in T_k$, then (3.23) and (3.24) contradicts the condition of Case k.2, which implies (3.20).

If $I_k\notin T_k$, then by (3.22) and the definition of T_k there exists a $j\in\{1,\dots,k-1\}$ for which $I_k\in T_j$. Since in Case j.1.3 of Step j there exists a $I_j\in T_j$ such that

$$x_{j+1}=P_{K\cap[x_jP_{I_j}]}(P_{I_j})\neq x_j,$$

from the fact of $I_k,\ I_j\in T_j\subset T(x_j)$ and Lemma 7 we have

$$\|P_{I_j}\|\geqslant\|P_{I_k}\|.$$

Combined with (3.24) and $d_1<d_2<\cdots<d_k$ we obtain

$$\|P_{I_k}\|>d_k\geqslant d_{j+1}=\|P_{I_j}\|\geqslant\|P_{I_k}\|.$$

This contradiction implies

$$I_k\in T_k,$$

which completes the proof of (i).

(ii) For $x_0, x'_1, ..., x'_{k'}$, by the definition of the successive approximate sequence, $x'_1 = x_1$ obviously. Suppose inductively

$$x'_1 = x_1, ..., x'_j = x_j,$$

where $1 \leqslant j < \min\{k, k'\}$. Since

$$x_{j+1} = P_{K \cap [x_j P_{I_j}]}(P_{I_j}) \neq x_j$$

and

$$x'_{j+1} = P_{K \cap [x'_j P_{I'_j}]}(P_{I'_j}) \neq x'_j,$$

by Lemma 7 we have

$$\|P_{I'_j}\| > \|P_{I_j}\|$$

and

$$\|P_{I_j}\| > \|P_{I'_j}\|$$

provided $P_{I_j} \neq P_{I'_j}$. So $P_{I_j} = P_{I'_j}$ and hence $x'_{j+1} = x_{j+1}$. Thus

$$x'_j = x_j, \qquad j = 1, 2, ..., \min\{k, k'\}.$$

Provided $k' > k$, since (2.1) implies $\|x'_{j+1}\| < \|x'_j\|$, by (i) we have

$$\|x'_{k'}\| < \|x'_k\| = \|x_k\| = \|P_K(O)\|$$

which is a contradiction.

(iii) (a) Relation (2.2) holds obviously.

(b) For $j = 1, ..., k-2$, by the algorithm

$$x_{j+1} = P_{K \cap [x_j P_{I_j}]}(P_{I_j}) \neq P_{I_j}.$$

So using Lemma 4 we have $P_{K_{I_j}}(O) = P_{I_j} \notin K$ and $P_K(O) \in K \subset K_{I_j}$. Thus by the uniqueness of the best approximation of O in K_{I_j} we have

$$\|P_K(O)\| > \|P_{I_j}\|, \tag{3.27}$$

i.e., (2.3) holds.

(c) If $j = k-1$, (2.4) holds clearly. Now let $1 \leqslant j < k-1$. Since there exists an $I_j \in T_j$ for which

$$x_{j+1} = P_{K \cap [x_j P_{I_j}]}(P_{I_j}) \neq P_{I_j},$$

by Lemma 1 there exists a $\lambda \in [0, 1)$ such that

$$x_{j+1} = (1-\lambda)\, x_j + \lambda P_{I_j}. \tag{3.28}$$

Based on (3.8) of Lemma 4, we conclude that

$$\left\langle x_k, -\frac{P_{I_j}}{\|P_{I_j}\|} \right\rangle \leqslant -\|P_{I_j}\|.$$

That is, $\langle x_k, P_{I_j}\rangle \geqslant \langle P_{I_j}, P_{I_j}\rangle$. Write the projection of x_k on H_{I_j} as P_1, the projection of P_1 on the straight line $\{x \mid x = \alpha x_j + (1-\alpha)\, x_{j+1}, \alpha \in \mathbb{R}\}$ as P_2. Then

$$\begin{aligned}
\|x_k\|^2 &= \|x_k - P_{I_j}\|^2 + 2\langle x_k - P_{I_j}, P_{I_j}\rangle + \|P_{I_j}\|^2 \\
&= \|x_k - P_1\|^2 + \|P_1 - P_{I_j}\|^2 + 2(\langle x_k, P_{I_j}\rangle - \langle P_{I_j}, P_{I_j}\rangle) + \|P_{I_j}\|^2 \\
&\geqslant \|P_1 - P_2\|^2 + \|P_2 - P_{I_j}\|^2 + \|P_{I_j}\|^2 \\
&= \|P_1 - P_2\|^2 + \|P_2\|^2 \geqslant \|P_2\|^2.
\end{aligned} \tag{3.29}$$

Suppose

$$P_2 = \alpha_0 x_j + (1-\alpha_0)\, x_{j+1}. \tag{3.30}$$

If $\alpha_0 \geqslant 0$, then from (3.30) and (3.28) we deduce

$$\begin{aligned}
\|P_2\|^2 &= \|P_{I_j}\|^2 + \|P_{I_j} - P_2\|^2 \\
&= \|P_{I_j}\|^2 + \left\| P_{I_j} - \left[\alpha_0 \frac{x_{j+1} - \lambda P_{I_j}}{1-\lambda} + (1-\alpha_0)\, x_{j+1} \right] \right\|^2 \\
&= \|P_{I_j}\|^2 + \left(1 + \frac{\lambda\alpha_0}{1-\lambda}\right)^2 \|P_{I_j} - x_{j+1}\|^2 \\
&\geqslant \|P_{I_j}\|^2 + \|P_{I_j} - x_{j+1}\|^2 = \|x_{j+1}\|^2.
\end{aligned}$$

From (3.29)

$$\|x_k\| \geqslant \|P_2\| \geqslant \|x_{j+1}\|$$

which contradicts (2.2).

Now we conclude that $\alpha_0 < 0$. So by (3.30)

$$\begin{aligned}\|x_j - x_k\|^2 &= \|x_j - P_2\|^2 + \|P_2 - x_k\|^2 \\ &= \left\| \frac{1}{\alpha_0} P_2 - \frac{1-\alpha_0}{\alpha_0} x_{j+1} - P_2 \right\|^2 + \|P_2 - x_k\|^2 \\ &= \left(\frac{1-\alpha_0}{\alpha_0}\right)^2 \|P_2 - x_{j+1}\|^2 + \|P_2 - x_k\|^2 \\ &> \|P_2 - x_{j+1}\|^2 + \|P_2 - x_k\|^2 = \|x_{j+1} - x_k\|^2,\end{aligned}$$

which completes the proof of (2.4).

(d) From (i) and Lemma 5 there exists an $I^* \in I(x_k)$ for which $\{f_i\}_{i \in I^*}$ are linearly independent and (3.10) and (3.11) hold. If $|I^*| < n$, then $I^* \in T(x_k)$ and by Lemma 4 we have

$$K \subset K_{I^*}.$$

However, as a matter of fact the hypothesis of $|I| < n$ is not needed for the proof of (3.8), so the above expression still holds if $|I^*| = n$. Thus

$$\left\langle x_{j+1}, \frac{-x_k}{\|x_k\|} \right\rangle \leqslant -\|x_k\|,$$

and

$$\begin{aligned}\|x_{j+1} - x_k\|^2 &= \|x_{j+1}\|^2 + 2\langle x_{j+1}, -x_k\rangle + \|x_k\|^2 \\ &\leqslant \|x_{j+1}\|^2 - 2\|x_k\|^2 + \|x_k\|^2 \\ &= \|x_{j+1}\|^2 - \|x_k\|^2.\end{aligned}$$

Combined with (3.27) we get (2.5).

(iv) It is not difficult to show that for any nonempty subset $I \subset I_+$,

$$I \notin T(x_j), \qquad j = 1, 2, ..., k. \tag{3.31}$$

In fact, (3.31) holds obviously if $\{f_i\}_{i \in I}$ are linearly dependent. Otherwise, by Lemma 2 we can write

$$P_I = \sum_{i \in I} \alpha_i f_i.$$

Suppose

$$\alpha_i \leqslant 0, \qquad i \in I. \tag{3.32}$$

Then using Lemma 3 we have $P_I = P_{K(I)}(O)$. But by the definition of I_+ we have $O \in K(I)$ which implies $P_{K(I)}(O) = O$. So $P_I = O$ and hence (3.31) holds. Moreover, if (3.32) is false, then (3.31) still holds.

Let

$$T_+ = \{I \subset I_+ \mid 0 < |I| < n\},$$
$$T = \{I \subset \{1, ..., r\} \mid 0 < |I| < n\}.$$

Then the numbers of the elements of T_+ and T are $\binom{r_+}{1} + \cdots + \binom{r_+}{n-1}$ and $\binom{r}{1} + \cdots + \binom{r}{n-1}$, respectively. Note for each x_j, $1 \leqslant j < k$, there exists an

$$I_j \in T_j \subset T(x_j).$$

So

$$T_j \subset T \backslash T_+, \qquad j = 1, ..., k-1.$$

because the intersection set of any two sets of $\{T_j\}_{j=1}^{k-1}$ is empty, we obtain

$$k - 1 \leqslant \left[\binom{r}{1} + \cdots + \binom{r}{n-1}\right] - \left[\binom{r_+}{1} + \cdots + \binom{r_+}{n-1}\right]. \quad \blacksquare$$

At last, we point out that in practice, the value of k depends on the nature of the given problem and the choice of the starting point x_0, and it may be that k is far less than the upper bound given by (2.6).

ACKNOWLEDGMENTS

I thank Professor F. Deutsch, Dr. Jun Zhong, and H. Hundal for valuable discussions on approximation from a polyhedron.

REFERENCES

1. R. L. Dykstra, An algorithm for restricted least squares regression, *J. Amer. Statist. Assoc.* **78** (1983), 837–842.
2. 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.
3. I. Halperin, The product of projection operators, *Acta Sci. Math.* (*Szeged*) **23** (1962), 96–99.
4. K. T. Smith, D. C. Solmon, and S. L. Wagner, Practical and mathematical aspects of the problem of reconstructing objects from radiographs, *Bull. Amer. Math. Soc.* **83** (1977), 1227–1270.

5. S. Kayalar and H. L. Weinert, Error bounds for the method of alternating projections, *Math. Control Signals Systems* **1** (1988), 43–59.
6. F. Deutsch and H. Hundal, The rate of convergence of Dykstra's cyclic projections algorithm: The polyhedral case, *Numer. Funct. Anal. Optim.* **15**, Nos. 5–6 (1994), 537–565.
7. L. M. Bregman, The method of successive projection for finding a common point of convex sets, *Doklady* **162** (1965), 688–692.
8. P. J. Laurent, "Approximation et Optimisation," Collect. Enseignement Sci., Vol. 13, Hermann, Paris, 1972.