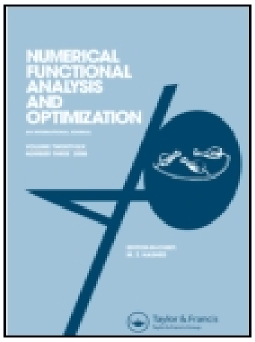

Numerical Functional Analysis and Optimization

ISSN: 0163-0563 (Print) 1532-2467 (Online) Journal homepage: https://www.tandfonline.com/loi/lnfa20

The rate of convergence of dykstra's cyclic projections algorithm: The polyhedral case

Frank Deutsch & Hein Hundal

To cite this article: Frank Deutsch & Hein Hundal (1994) The rate of convergence of dykstra's cyclic projections algorithm: The polyhedral case, Numerical Functional Analysis and Optimization, 15:5-6, 537-565, DOI: 10.1080/01630569408816580

To link to this article: https://doi.org/10.1080/01630569408816580

Published online: 15 May 2007.

Submit your article to this journal

Article views: 168

View related articles

Citing articles: 5 View citing articles

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=lnfa20

NUMER. FUNCT. ANAL. AND OPTIMIZ., 15(5&6), 537-565 (1994)

THE RATE OF CONVERGENCE OF DYKSTRA'S CYCLIC PROJECTIONS ALGORITHM: THE POLYHEDRAL CASE

FRANK DEUTSCH[1] AND HEIN HUNDAL

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

ABSTRACT. Suppose K is the intersection of a finite number of closed half-spaces in a Hilbert space X. Starting with any point $x \in X$, it is shown that the sequence of iterates $\{x_n\}$ generated by Dykstra's cyclic projections algorithm satisfies the inequality

$$||x_n - P_K(x)|| \leq \rho c^n$$

for all n, where $P_K(x)$ is the nearest point in K to x, ρ is a constant, and $0 \leq c < 1$. In the case when K is the intersection of just two closed half-spaces, a stronger result is established: the sequence of iterates is either finite or satisfies $||x_n - P_K(x)|| \leq c^{n-1}||x - P_K(x)||$ for all n, where c is the cosine of the angle between the two functionals which define the half-spaces. Moreover, the constant c is the best possible. Applications are made to isotone and convex regression, and linear and quadratic programming.

AMS(MOS) subject classification: Primary 41A65; Secondary 47N10, 49M30
Key Words: rate of convergence of Dykstra's algorithm, cyclic projections, alternating projections, iterative projections, best approximations from polyhedra, isotone regression, convex regression, linear programming, quadratic programming, Hildreth's algorithm, linear inequalities.

1 Introduction

Dykstra's cyclic projections algorithm is an attractive method for determining best approximations from a closed convex subset of a Hilbert space. Suppose $K = \bigcap_{i=1}^{r} K_i$ is the intersection of a finite number of closed convex sets $K_1, \ldots, K_r$ in a Hilbert space X. To find the best approximation from K to any $x \in X$, Dykstra's method essentially reduces the problem to an iterative scheme which involves computing the best approximations from the *individual* sets K_i which make up the intersection K. The efficacy of Dykstra's method thus depends on one's ability to compute best approximations from

[1]This author was supported in part by NSF Grant DMS-9100228.

Copyright © 1994 by Marcel Dekker, Inc.

the individual sets K_i. In the particular case when all the K_i are subspaces (or, more generally, linear varieties), Dykstra's algorithm reduces to the *method of alternating projections* due to von Neumann [25] for two subspaces and Halperin [18] for more than two subspaces. It is now well-documented that the method of alternating projections has had many important applications in at least a dozen different areas of mathematics including solving linear equations, computed tomography, linear prediction theory, and image restoration (see, e.g., the survey by the first author [11]).

For general closed convex sets which are not subspaces, Dykstra's algorithm is more involved than the method of alternating projections. In the Dykstra algorithm, certain "residual" vectors must be computed at each step and carried along in subsequent steps. In the method of alternating projections, these residuals do not appear. Some applications of Dykstra's algorithm to certain problems in statistics have been made by Dykstra [13] and Boyle and Dykstra [4].

In both the practical and theoretical applications of the method of alternating projections and Dykstra's algorithm, it is useful to have a priori error bounds on the rate of convergence of the sequence of iterates. In the case of the method of alternating projections, an error analysis was made for the case of two subspaces by Aronszajn [1] (and also by [26], [9,10], and [15] independently) and, for the case of more than two subspaces, by Smith, Solmon, and Wagner [26]. More recently, Kayalar and Weinert [23] have shown that Aronszajn's bound in the case of two subspaces is sharp; moreover, they have given the sharpest known bounds for the case of more than two subspaces. In all of these error bounds, the notion of angle between two subspaces played a dominant role. Furthermore, we can show that these error bounds remain valid if the subspaces are replaced, more generally, by linear varieties, i.e., translates of subspaces.

In this paper we develop an error analysis for Dykstra's algorithm in the important case when K is polyhedral, i.e., the intersection of a finite number of closed half-spaces. It will be seen that the error bound will depend on the angles between the linear varieties which comprise the boundaries of the half-spaces in question. Moreover, this bound is sharp in the case of two half-spaces. When K is polyhedral and X is Euclidean n-space, Dykstra's algorithm reduces to *Hildreth's algorithm* [20]. Linear convergence of Hildreth's algorithm, with less sharp bounds, was established by Iusem and De Pierro [21] using the methods of optimization theory.

In section 2, Dykstra's algorithm is described and a few general facts are recorded concerning the polyhedral case. In section 3, we give a rate of convergence theorem for the general polyhedral case. The main result here is Theorem 3.8. In section 4, we show that (with some extra effort) a substantially sharper bound can be established in the case of two half-spaces (Theorem 4.1). Indeed, this bound is shown to be the best possible. In section 5, some applications are mentioned. They include isotone and convex regression, and linear and convex programming.

We conclude this introduction by mentioning some recent work related to Dykstra's algorithm. Han [19], apparently unaware of Dykstra's algorithm, reproved the finite-dimensional version. Iusem and De Pierro [22] applied Han's (i.e., Dykstra's) algorithm to prove a (finite-dimensional) variant thereof, and also gave a convergence proof for the case when the intersection of the convex sets is empty. Gaffke and Mathar [17] gave a simpler proof of the convergence of Dykstra's algorithm. Their approach, via duality, also provided the motivation for Dykstra's algorithm. Bauschke and Borwein [2] generalized the results of Iusem and De Pierro to infinite-dimensional spaces, and also considered the case when the distance between the two convex sets is not attained.

Finally, we should mention that if one is only seeking *some* point in the intersection of convex sets, and not necessarily the best approximation to a prescribed point, then successive projection methods for this simpler problem go back at least to the mid 1960's (see, e.g., [5]). However, their applicability is naturally more restrictive.

2 General Definitions and Lemmas

Let X be a Hilbert space and let $\mathcal{H}_1, \dots, \mathcal{H}_r$ be r closed half-spaces in X with nonempty intersection K. Thus

$$\mathcal{H}_i = \{x \in X \mid \langle x, f_i \rangle \le c_i\} \quad (i = 1, \dots, r), \tag{2.0.1}$$

where $f_i \in X \backslash \{0\}$ and $c_i \in \mathbb{R}$. By scaling we may assume $||f_i|| = 1$.

Given any $x \in X$, we want to determine the best approximation $P_K(x)$ in $K = \bigcap_{i=1}^r \mathcal{H}_i$ to x. (Recall that for any nonempty closed convex set K in X, there always exists a unique best approximation $P_K(x)$ in K to any $x \in X$. That is, $||x - P_K(x)|| = \inf\{||x - y|| \mid y \in K\}$.) We will compute $P_K(x)$ via Dykstra's algorithm [13],[4]. Starting with any $x \in X$, Dykstra's algorithm generates the sequence $\{x_n\}$ as follows.

$$\begin{gathered} x_0 := x, \quad e_n := 0 \quad \text{when} \quad n \le 0, \\ [n] := \{n + mr \mid m \in \mathbb{Z}\} \cap \{1, \dots, r\}, \qquad x_n := P_{\mathcal{H}_{[n]}}(x_{n-1} + e_{n-r}), \\ e_n := x_{n-1} + e_{n-r} - x_n, \qquad (n = 1, 2, \dots). \end{gathered} \tag{2.0.2}$$

The Boyle–Dykstra Theorem [4] implies that

$$||x_n - P_K(x)|| \to 0 \quad \text{as} \quad n \to \infty.$$

For notational convenience we define

$$\begin{gathered} x_\infty := \lim_{n \to \infty} x_n = P_K(x), \qquad \text{and} \\ H_i := \{x \in X \mid \langle x, f_i \rangle = c_i\} \end{gathered} \tag{2.0.3}$$

is the hyperplane which forms the boundary of $\mathcal{H}_i$. The *distance* from x to any set K is defined by $d(x, K) := \inf\{||x - y|| \mid y \in K\}$. The *interior* of $\mathcal{H}_i$, denoted $\operatorname{int} \mathcal{H}_i$, is the set $\operatorname{int} \mathcal{H}_i = \mathcal{H}_i \backslash H_i = \{x \in X \mid \langle x, f_i \rangle < c_i\}$. Now we are ready to state and prove some basic lemmas. The following is well known (see, e.g., [12]).

Lemma 2.1. *Let $f \in X$ have norm 1, $c \in \mathbb{R}$, let H denote the hyperplane $H = \{x \in X \mid \langle x, f\rangle = c\}$ and $\mathcal{H}$ the half-space $\mathcal{H} = \{x \in X \mid \langle x, f\rangle \le c\}$. Then*

$$P_{\mathcal{H}}(x) = \begin{cases} x & \text{if } x \in \mathcal{H} \\ x - [\langle x, f\rangle - c]f & \text{if } x \notin \mathcal{H}, \end{cases} \tag{2.1.1}$$

$$P_H(x) = x - [\langle x, f\rangle - c]f, \qquad x \in X, \qquad \text{and} \tag{2.1.2}$$

$$d(x, H) = |\langle x, f\rangle - c|. \tag{2.1.3}$$

In particular,

$$P_H(y + \alpha f) = P_H(y), \qquad y \in X, \quad \alpha \in \mathbb{R}. \tag{2.1.4}$$

For the remaining lemmas, we assume that $x \in X$ is given and the sequences $\{x_n\}$ and $\{e_n\}$ are defined as in (2.0.2). Moreover, we will use the notation (2.0.3) for H_i and x_∞.

Lemma 2.2. *For each positive integer n,*

$$e_n = k_n f_{[n]}, \tag{2.2.1}$$

where $k_n := d(x_{n-1} + e_{n-r}, \mathcal{H}_{[n]}) \ge 0$.

In particular,

(1) $$x_{n-1} + e_{n-r} \in \mathcal{H}_{[n]} \iff k_n = 0;$$

(2) $$x_{n-1} + e_{n-r} \notin \mathcal{H}_{[n]} \iff k_n = \langle x_{n-1} + e_{n-r}, f_{[n]}\rangle - c_{[n]} > 0.$$

Proof. If $x_{n-1} + e_{n-r} \in \mathcal{H}_{[n]}$, then

$$x_n = P_{\mathcal{H}_{[n]}}(x_{n-1} + e_{n-r}) = x_{n-1} + e_{n-r}$$

and

$$e_n = x_{n-1} + e_{n-r} - x_n = 0.$$

Thus $k_n = 0$.

If $x_{n-1} + e_{n-r} \notin \mathcal{H}_{[n]}$, then, using Lemma 2.1,

$$\begin{aligned} x_n &= P_{\mathcal{H}_{[n]}}(x_{n-1} + e_{n-r}) = P_{H_{[n]}}(x_{n-1} + e_{n-r}) \\ &= x_{n-1} + e_{n-r} - [\langle x_{n-1} + e_{n-r}, f_{[n]}\rangle - c_{[n]}]f_{[n]} \end{aligned}$$

and

$$e_n = x_{n-1} + e_{n-r} - x_n = [\langle x_{n-1} + e_{n-r}, f_{[n]}\rangle - c_{[n]}]f_{[n]} = k_n f_{[n]},$$

where

$$k_n = \langle x_{n-1} + e_{n-r}, f_{[n]}\rangle - c_{[n]} > 0$$

since $x_{n-1} + e_{n-r} \notin \mathcal{H}_{[n]}$. But using (2.1.3),

$$\begin{aligned} d(x_{n-1} + e_{n-r}, \mathcal{H}_{[n]}) &= d(x_{n-1} + e_{n-r}, H_{[n]}) \\ &= |\langle x_{n-1} + e_{n-r}, f_{[n]}\rangle - c_{[n]}| \\ &= k_n. \quad \blacksquare \end{aligned}$$

Lemma 2.3. *For any integer $n \geq 1$, the following statements are equivalent.*

(1) $x_n = P_{H_{[n]}}(x_{n-1})$;

(2) $x_{n-1} + e_{n-r} \notin \operatorname{int} \mathcal{H}_{[n]}$;

(3) $\langle x_{n-1} + e_{n-r}, f_{[n]}\rangle \geq c_{[n]}$.

Proof. Since $\operatorname{int} \mathcal{H}_{[n]} = \{y \in X \mid \langle y, f_{[n]}\rangle < c_{[n]}\}$, the equivalence of (2) and (3) is immediate.

(1) $\Rightarrow$ (2). Assume (1) holds. Then using (2.1.4),

$$\begin{aligned} P_{\mathcal{H}_{[n]}}(x_{n-1} + e_{n-r}) &= x_n = P_{H_{[n]}}(x_{n-1}) \\ &= P_{H_{[n]}}(x_{n-1} + k_{n-r} f_{[n]}) = P_{H_{[n]}}(x_{n-1} + e_{n-r}). \end{aligned}$$

This implies (2).

(2) $\Rightarrow$ (1). If (2) holds, then

$$x_n = P_{\mathcal{H}_{[n]}}(x_{n-1} + e_{n-r}) = P_{H_{[n]}}(x_{n-1} + e_{n-r}) = P_{H_{[n]}}(x_{n-1})$$

using Lemmas 2.1 and 2.2. That is, (1) holds. ■

A key step the convergence proof for an arbitrary number of half-spaces rests on the following theorem. Here we use $\operatorname{co}(A)$ to denote the **convex hull** of A, the intersection of all convex sets which contain A.

Theorem 2.4. *For each positive integer n,*

$$x_n \in \operatorname{co}\{x_{n-1}, P_{H_{[n]}}(x_{n-1})\}. \tag{2.4.1}$$

Proof. Using Lemma 2.3, we may assume that $x_{n-1} + e_{n-r} \in \operatorname{int} \mathcal{H}_{[n]}$. Then

$$\langle x_{n-1} + e_{n-r}, f_{[n]}\rangle < c_{[n]} \tag{2.4.2}$$

and

$$x_n = P_{\mathcal{H}_{[n]}}(x_{n-1} + e_{n-r}) = x_{n-1} + e_{n-r}.$$

Thus

$$x_n - x_{n-1} = e_{n-r} = k_{n-r} f_{[n-r]} = k_{n-r} f_{[n]} \tag{2.4.3}$$

using Lemma 2.2. Also, by Lemma 2.1,

$$P_{H_{[n]}}(x_{n-1}) = x_{n-1} - [\langle x_{n-1}, f_{[n]}\rangle - c_{[n]}] f_{[n]}. \tag{2.4.4}$$

If $k_{n-r} = 0$, then $x_n = x_{n-1} \in \operatorname{co}\{x_{n-1}, P_{H_{[n]}}(x_{n-1})\}$. If $k_{n-r} > 0$, solve (2.4.3) for $f_{[n]}$ and substitute into (2.4.4) to get

$$P_{H_{[n]}}(x_{n-1}) = x_{n-1} - [\langle x_{n-1}, f_{[n]}\rangle - c_{[n]}]\frac{x_n - x_{n-1}}{k_{n-r}}$$

or

$$x_n = (1-\lambda)x_{n-1} + \lambda P_{H_{[n]}}(x_{n-1}),$$

where

$$\lambda = \frac{k_{n-r}}{c_{[n]} - \langle x_{n-1}, f_{[n]}\rangle}.$$

To complete the proof, it suffices to show that $0 < \lambda < 1$. But

$$\begin{aligned} 0 < k_{n-r} &= k_{n-r}\langle f_{[n-r]}, f_{[n-r]}\rangle = \langle e_{n-r}, f_{[n]}\rangle \\ &= \langle x_{n-1} + e_{n-r}, f_{[n]}\rangle - \langle x_{n-1}, f_{[n]}\rangle < c_{[n]} - \langle x_{n-1}, f_{[n]}\rangle \end{aligned}$$

using Lemma 2.2 and (2.4.2). It follows that $0 < \lambda < 1$. ∎

The next lemma is very useful. It is found in Boyle-Dykstra [4] and can be proved easily by induction.

Lemma 2.5. $x - x_{n-1} = e_{n-1} + e_{n-2} + \cdots + e_{n-r}$.

3 Intersection of r Closed Half-Spaces

In this section we establish that for r half-spaces Dykstra's algorithm converges geometrically (i.e., there exist constants ρ and $0 \le c < 1$ such that $\|x_n - x_\infty\| \le \rho c^n$ for all n). We will use the notation and definitions for Dykstra's algorithm established in the beginning of section 2 ((2.0.1)–(2.0.3)).

For simplicity, we shall assume throughout this section that $x_\infty = 0$. At the end of the section, we will see how the general case can by reduced to this case by a translation argument. Next we partition the set $\{1, 2, \ldots, r\}$ into two subsets. Let

$$A := \{i \in \{1, 2, \ldots, r\} \mid x_\infty \in H_i\}$$

and

$$B := \{1, 2, \ldots, r\} \backslash A = \{i \in \{1, 2, \ldots, r\} \mid x_\infty \in \operatorname{int} \mathcal{H}_i\}.$$

Here A may be regarded as the set of "active" constraints as in linear programming. These sets play a more important role than the B half-spaces as shown by the next lemma.

Lemma 3.1. *There exists an integer N_1 such that whenever $[n] \in B$ and $n \ge N_1$, then $x_n = x_{n-1}$ and $e_n = 0$.*

Proof. Let $C = \bigcap_{i \in B} \operatorname{int} \mathcal{H}_i$. By the definition of B, $x_\infty \in C$. Noting that C is open and $x_k \to x_\infty$, we conclude that there exists an integer N such that $x_k \in C$ for all $k \ge N$. Let $N_1 = N + r$. Under the conditions that $n \ge N_1$ and $[n] \in B$, it follows that $x_{n-r} \in C$ so $x_{n-r} \in \operatorname{int} \mathcal{H}_{[n]}$. But

$$x_{n-r} := P_{\mathcal{H}_{[n-r]}}(x_{n-r-1} + e_{n-2r}) = P_{\mathcal{H}_{[n]}}(x_{n-r-1} + e_{n-2r}),$$

and the fact that x_{n-r} is in the interior of $\mathcal{H}_{[n]}$ imply that $x_{n-r} = x_{n-r-1} + e_{n-2r}$. It follows that $e_{n-r} := x_{n-r-1} + e_{n-2r} - x_{n-r} = 0$. Now x_n is also in the interior of $\mathcal{H}_{[n]}$ so, by similar reasoning, $x_n = x_{n-1} + e_{n-r} = x_{n-1}$. Finally, $e_n := x_{n-1} + e_{n-r} - x_n = 0$. ■

The following lemma will be useful in the proof of Theorem 3.3. It states that the metric projection onto a linear variety is affine, and hence commutes with the convex hull operation. (Recall that a *linear variety* is a translate of a subspace.)

Lemma 3.2. *If H is a closed linear variety in a Hilbert space X, then P_H is "affine"; that is,*

$$P_H\left(\sum_1^n \alpha_i x_i\right) = \sum_1^n \alpha_i P_H(x_i) \tag{3.2.1}$$

for all $x_i \in X$ and any $\alpha_i \in \mathbb{R}$ which satisfy $\sum_1^n \alpha_i = 1$.

In particular, if A is any nonempty subset of X, then

$$P_H(\operatorname{co}(A)) = \operatorname{co}(P_H(A)). \tag{3.2.2}$$

Proof. We can write $H = M + x_0$ for some closed subspace M and any $x_0 \in H$. Note that $P_H(x) = P_M(x - x_0) + x_0$ for all $x \in X$ and P_M is linear. Thus, for any $x_i \in X$ and $\alpha_i \in \mathbb{R}$ with $\sum_1^n \alpha_i = 1$, we obtain

$$\begin{aligned} P_H\left(\sum_1^n \alpha_i x_i\right) &= P_M\left(\sum_1^n \alpha_i x_i - x_0\right) + x_0 = P_M\left(\sum_1^n \alpha_i (x_i - x_0)\right) + x_0 \\ &= \sum_1^n \alpha_i P_M(x_i - x_0) + x_0 = \sum_1^n \alpha_i [P_M(x_i - x_0) + x_0] = \sum_1^n \alpha_i P_H(x_i). \end{aligned}$$

this proves (3.2.1).

The proof of (3.2.2) follows from this since

$$\begin{aligned} P_H(\operatorname{co}(A)) &= P_H\left(\left\{\sum_i \lambda_i a_i \mid a_i \in A, \lambda_i \geq 0, \sum_i \lambda_i = 1\right\}\right) \\ &= \left\{P_H\left(\sum_i \lambda_i a_i\right) \mid a_i \in A, \lambda_i \geq 0, \sum_i \lambda_i = 1\right\} \\ &= \left\{\sum_i \lambda_i P_H(a_i) \mid a_i \in A, \lambda_i \geq 0, \sum_i \lambda_i = 1\right\} \\ &= \operatorname{co}(P_H(A)). \quad ■ \end{aligned}$$

The following definitions are required for the statement of Theorem 3.3. Define

$$A_n := A \cap \{[n+i] \mid 1 \le i \le r,\ x_{n+i} = P_{H_{[n+i]}}(x_{n+i-1})\}.$$

The set-valued function $K_n(E,D)$ is defined as follows:

$$K_n(E,D) := \text{co}\,\{P_{S,n}(x_n) \mid E \supset S \supset D\}, \quad \text{where}$$

$$P_{S,n} := Q_{S,n+r}Q_{S,n+r-1}\cdots Q_{S,n+1} \quad \text{and} \quad Q_{S,m} := \begin{cases} P_{H_{[m]}} & \text{if } [m] \in S \\ I & \text{if } [m] \notin S. \end{cases}$$

Theorem 3.3. *For each $n \ge N_1$, $x_{n+r} \in K_n(A, A_n)$.*

Proof. Fix $n \ge N_1$. To establish Theorem 3.3, we shall prove the stronger result that for $i = 1, 2, \ldots, r$,

$$x_{n+i} \in K_n(E_i, D_i), \quad \text{where} \tag{3.3.1}$$
$$E_i := A \cap \{[n+1], \ldots, [n+i]\} \quad \text{and} \quad D_i := A_n \cap \{[n+1], \ldots, [n+i]\}.$$

Note $K_n(E_r, D_r) = K_n(A, A_n)$. We establish (3.3.1) by induction on i. First note that for $i = 1, 2, \ldots, r$,

$$x_{n+i} = \begin{cases} x_{n+i-1} & \text{if } \ [n+i] \in B \\ P_{H_{[n+i]}}(x_{n+i-1}) & \text{if } \ [n+i] \in A_n \end{cases} \tag{3.3.2}$$

by Lemma 3.1 and the definition of A_n. If $[n+i] \in A \backslash A_n$, then

$$x_{n+i} \in \text{co}\,\left\{x_{n+i-1}, P_{H_{[n+i]}}(x_{n+i-1})\right\} \tag{3.3.3}$$

by Theorem 2.4. We begin our induction with $i = 1$.

Case 1.1: $[n+1] \in B$.

By (3.3.2), $x_{n+1} = x_n$, $D_i = E_i = \emptyset$ so $K_n(E_i, D_i) = K_n(\emptyset, \emptyset) = \{x_n\} = \{x_{n+1}\}$ and hence (3.3.1) holds.

Case 1.2: $[n+1] \in A_n$.

By (3.3.2), $x_{n+1} = P_{H_{[n+1]}}(x_n)$, $D_i = E_i = \{[n+1]\}$ so

$$\begin{aligned} K_n(E_i, D_i) &= K_n(\{[n+1]\}, \{[n+1]\}) \\ &= \text{co}\,\{(Q_{\{[n+1]\},n+r}\cdots Q_{\{[n+1]\},n+1})(x_n)\} \\ &= \{P_{H_{[n+1]}}(x_n)\} = \{x_{n+1}\} \end{aligned}$$

hence (3.3.1) holds.

Case 1.3: $[n+1] \in A \backslash A_n$.

$D_i = \emptyset$ and $E_i = \{[n+1]\}$ so

$$K_n(E_i, D_i) = K_n(\{[n+1]\}, \emptyset) = \text{co}\,\{x_n, P_{H_{[n+1]}}(x_n)\}.$$

By (3.3.3), $x_{n+1} \in \operatorname{co}\{x_n, P_{H_{[n+1]}}(x_n)\}$ and hence (3.3.1) holds.

This proves (3.3.1) for $i = 1$. Now we prove that if (3.3.1) holds for $i-1$, then (3.3.1) holds for i. If (3.3.1) holds for $i-1$, then $x_{n+i-1} \in K_n(E_{i-1}, D_{i-1})$.

Case 2.1: $[n+i] \in B$.

Note

$$\begin{aligned} D_i &= A_n \cap \{[n+1], \dots, [n+i]\} = A_n \cap \{[n+1], \dots, [n+i-1]\} = D_{i-1}, \\ E_i &= A \cap \{[n+1], \dots, [n+i]\} = A \cap \{[n+1], \dots, [n+i-1]\} = E_{i-1}. \end{aligned}$$

By (3.3.2), $x_{n+i} = x_{n+i-1} \in K_n(E_{i-1}, D_{i-1}) = K_n(E_i, D_i)$. Hence (3.3.1) holds.

Case 2.2: $[n+i] \in A_n$.

Note

$$\begin{aligned} D_i &= A_n \cap \{[n+1], \dots, [n+i]\} = (A_n \cap \{[n+1], \dots, [n+i-1]\}) \cup \{[n+i]\} \\ &= D_{i-1} \cup \{[n+i]\}, \\ E_i &= A \cap \{[n+1], \dots, [n+i]\} = (A \cap \{[n+1], \dots, [n+i-1]\}) \cup \{[n+i]\} \\ &= E_{i-1} \cup \{[n+i]\}. \end{aligned}$$

By (3.3.2),

$$x_{n+i} = P_{H_{[n+i]}}(x_{n+i-1}) \in P_{H_{[n+i]}}(K_n(E_{i-1}, D_{i-1})). \tag{3.3.4}$$

Next we observe the following claim.

Claim. *For any subset D of E_{i-1},*

$$P_{H_{[n+i]}}(K_n(E_{i-1}, D)) = K_n(E_{i-1} \cup \{[n+i]\}, D \cup \{[n+i]\}).$$

Proof. Using the definition of K_n, (3.2.2), and the fact that $\{[n+i], \dots, [n+r]\} \cap \{[n+1], \dots, [n+i-1]\} = \emptyset$ gives

$$\begin{aligned} P_{H_{n+i}}(K_n(E_{i-1}, D)) &= \\ &= P_{H_{[n+i]}}(\operatorname{co}\{Q_{S,n+r} Q_{S,n+r-1} \cdots Q_{S,n+1}(x_n) \mid E_{i-1} \supset S \supset D\}) \\ &= P_{H_{[n+i]}}(\operatorname{co}\{I \cdots I Q_{S,n+i-1} \cdots Q_{S,n+1}(x_n) \mid E_{i-1} \supset S \supset D\}) \\ &= \operatorname{co}\{I \cdots I P_{H_{[n+i]}} Q_{S,n+i-1} \cdots Q_{S,n+1}(x_n) \mid E_{i-1} \supset S \supset D\}. \end{aligned}$$

Let $S' = S \cup \{[n+i]\}$. Note that when $E_{i-1} \supset S$,

$$Q_{S',m} = \begin{cases} I & \text{if } \; n+r \geq m > n+i \\ P_{H_{[n+i]}} & \text{if } \; m = n+i \\ Q_{S,m} & \text{if } \; n+i > m \geq n+1. \end{cases}$$

Thus

$$\begin{aligned}P_{H_{n+i}}(K_n(E_{i-1},D)) &= \mathrm{co}\,\{Q_{S',n+r}\cdots Q_{S',n+1}(x_n) \mid E_{i-1}\supset S\supset D\}\\ &= \mathrm{co}\,\{Q_{S',n+r}\cdots Q_{S',n+1}(x_n) \mid \\ &\qquad (E_{i-1}\cup\{[n+i]\})\supset S'\supset(D\cup\{[n+i]\})\}\\ &= K_n(E_{i-1}\cup\{[n+i]\}, D\cup\{[n+i]\})\end{aligned}$$

which proves the claim.

Applying the claim to (3.3.4) yields

$$\begin{aligned}x_{n+i} &\in P_{H_{[n+i]}}(K_n(E_{i-1},D_{i-1})) = K_n(E_{i-1}\cup\{[n+i]\}, D_{i-1}\cup\{[n+i]\})\\ &= K_n(E_i,D_i)\end{aligned}$$

proving Case 2.2.

Case 2.3: $[n+i]\in A\backslash A_n$.

Note

$$D_i = D_{i-1}, \quad E_i = E_{i-1}\cup\{[n+i]\}.$$

Since $x_{n+i}\in \mathrm{co}\,\{x_{n+i-1}, P_{H_{[n+i]}}(x_{n+i-1})\}$ by (3.3.3), and the induction hypothesis implies

$$x_{n+i-1}\in K_n(E_{i-1},D_{i-1}),$$

we obtain

$$x_{n+i}\in \mathrm{co}\,(K_n(E_{i-1},D_{i-1}) \ \cup\ P_{H_{[n+i]}}(K_n(E_{i-1},D_{i-1}))).$$

Using the fact that $\mathrm{co}\,(\mathrm{co}\,(S)\cup\mathrm{co}\,(T)) = \mathrm{co}\,(S\cup T)$ for any sets S and T, we deduce

$$\begin{aligned}&\mathrm{co}\,[K_n(E_{i-1},D_{i-1}) \ \cup\ P_{H_{[n+i]}}(K_n(E_{i-1},D_{i-1}))]\\ &\quad= \mathrm{co}\,[K_n(E_{i-1},D_{i-1}) \ \cup\ K_n(E_{i-1}\cup\{[n+i]\}, D_{i-1}\cup\{[n+i]\})]\\ &\quad= \mathrm{co}\,[\mathrm{co}\,\{P_{S,n}(x_n)\mid E_{i-1}\supset S\supset D_{i-1}\}\\ &\qquad\quad \cup\,\mathrm{co}\,\{P_{S,n}(x_n)\mid E_i\supset S\supset D_{i-1}\cup\{[n+i]\}\}]\\ &\quad\subset \mathrm{co}\,\{P_{S,n}(x_n)\mid E_i\supset S\supset D_{i-1}\} = K_n(E_i,D_{i-1})\\ &\quad= K_n(E_i,D_i).\end{aligned}$$

Thus $x_{n+i}\in K_n(E_i,D_i)$ and this completes the proof. ∎

The following lemma establishes, in particular, that A_n is nonempty. Here we are using the *convention* that $\mathrm{span}\,\emptyset = \{0\} = \sum_{i\in\emptyset} e_i$.

Lemma 3.4. *There exists $N_2\geq N_1$ such that for all $n\geq N_2$,*

$$x_{n+r}\in \mathrm{span}\,\{f_i \mid i\in A_n\}.$$

Furthermore, if $x\neq x_\infty$ and $n\geq N_2$, then $A_n\neq\emptyset$.

Proof. If $x = x_\infty (= 0)$, then $x_n = 0$ for all n and the lemma is immediate. So we assume that $x \neq x_\infty$. Define

$$\epsilon = \min_{F \in \mathcal{F}} d(0, F) = \min_{F \in \mathcal{F}} d(x_\infty, F), \quad \text{where}$$
$$\mathcal{F} = \{F \mid F = \operatorname{span}\{f_i \mid i \in C\} + x, \quad C \subset A, \quad \text{and} \quad d(0, F) > 0\}.$$

Note that ϵ exists and $\epsilon > 0$ because there are only a finite number of subsets of A and $\emptyset \subset A$. Also, $x_n \longrightarrow 0$ implies there exists $N_2 \geq N_1$ such that $n \geq N_2$ implies $||x_n|| < \epsilon$. Now we claim that if $n \geq N_2$, then $x_{n+r} \in x + \operatorname{span}\{f_i \mid i \in A_n\}$. To see this, note that by the definition of A_n, if $[n+i] \notin A_n$ then either $x_{n+i} \neq P_{H_{[n+i]}}(x_{n+i-1})$ or $[n+i] \in B$. In the first case, $x_{n+i-1} + e_{n+i-r} \in \operatorname{int} \mathcal{H}_{[n+i]}$ by Lemma 2.3 which implies $e_{n+i} = 0$ by Lemma 2.2. In the second case, $e_{n+i} = 0$ by Lemma 3.1. We conclude that if $[n+i] \notin A_n$, then $e_{n+i} = 0$. It now follows from Lemmas 2.5 and 2.2 that

$$\begin{aligned} x_{n+r} &= x - \sum \{e_{n+i} \mid 1 \leq i \leq r\} \\ &= x - \sum \{e_{n+i} \mid 1 \leq i \leq r,\ [n+i] \in A_n\} \\ &= x - \sum \{k_{n+i} f_{[n+i]} \mid 1 \leq i \leq r,\ [n+i] \in A_n\} \\ &\in x + \operatorname{span}\{f_i \mid i \in A_n\}. \end{aligned} \tag{3.4.1}$$

By the definition of ϵ, either

$$d(0, x + \operatorname{span}\{f_i \mid i \in A_n\}) \geq \epsilon \quad \text{or} \quad d(0, x + \operatorname{span}\{f_i \mid i \in A_n\}) = 0.$$

But when $n \geq N_2$, $x_{n+r} \in x + \operatorname{span}\{f_i \mid i \in A_n\}$ and $||x_{n+r}|| < \epsilon$ so that we must have $d(0, x + \operatorname{span}\{f_i \mid i \in A_n\}) = 0$. That is, $x \in \operatorname{span}\{f_i \mid i \in A_n\}$. It follows that $x_{n+r} \in \operatorname{span}\{f_i \mid i \in A_n\}$. This proves the first statement of Lemma 3.4.

For the second statement we assume $x \neq x_\infty (= 0)$, $n \geq N_2$, and $A_n = \emptyset$. Now (3.4.1) implies $x = x_{n+r}$, but $x_{n+r} \in \operatorname{span}\{f_i \mid i \in A_n\} = \{0\}$. This is a contradiction. We conclude that if $x \neq x_\infty$ and $n \geq N_2$, then $A_n \neq \emptyset$. ■

The rate of convergence of Dykstra's algorithm will be seen to depend on the notion of "angle" between two subspaces.

Let M and N be closed subspaces in the Hilbert space X. The **angle** between M and N is the angle between 0 and $\pi/2$ whose cosine is given by

$$\begin{aligned} c(M, N) := \sup\{ |\langle x, y \rangle| \ \mid x \in M \cap (M \cap N)^\perp,\ ||x|| \leq 1, \\ y \in N \cap (M \cap N)^\perp,\ ||y|| \leq 1\}. \end{aligned}$$

This definition is due to Friedrichs [16].

We will need the following known facts concerning angles.

Theorem 3.5.

(1) *$c(M, N) < 1$ if and only if $M + N$ is closed.*

(2) $c(M,N) = c(M^{\perp}, N^{\perp})$.

Proof. These statements seem to be part of the folklore. Proofs can also be found, for example, in the handwritten lecture notes of D.C. Solmon [27]. Since these notes are not widely accessible, let us sketch the proof of (1) here.

(1) Deutsch [10] showed that $c(M,N) < 1$ iff $M \cap (M \cap N)^{\perp} + N \cap (M \cap N)^{\perp}$ is closed. Bauschke and Borwein [3] reported a short proof due to A. Simonič that $M + N$ is closed iff $M \cap (M \cap N)^{\perp} + N \cap (M \cap N)^{\perp}$ is closed. Hence (1) holds. ■

As an immediate consequence of Theorem 3.5, we get a result which is essential for our main theorem.

Corollary 3.6. *If M and N are closed subspaces, one of which has finite codimension, in the Hilbert space X, then $c(M,N) < 1$.*

Proof. $M^{\perp}$ and $N^{\perp}$ are closed subspaces one of which is finite-dimensional. Thus $M^{\perp} + N^{\perp}$ is closed. By Theorem 3.5, $c(M,N) = c(M^{\perp}, N^{\perp}) < 1$. ■

Lemma 3.7. *There exist r constants $\alpha_1, \alpha_2, \ldots, \alpha_r < 1$ such that for all $n \geq N_2$, $||x_{n+r}|| \leq \alpha_{[n]} ||x_n||$.*

Proof. If $x = x_\infty (= 0)$, then $x_n = x_\infty$ for all n and the result is immediate. So assume $x \neq x_\infty$. Fix any $n \geq N_2$ and let $M_n = \operatorname{span}\{f_i \mid i \in A_n\}$. By Lemma 3.4, $x_{n+r} \in M_n$. By Theorem 3.3, $x_{n+r} \in K_n(A, A_n)$. It follows that $x_{n+r} \in P_{M_n}(K_n(A, A_n))$. We can now obtain a bound on $||x_{n+r}||$. Note first that by Lemma 3.2,

$$\begin{aligned} P_{M_n}(K_n(A, A_n)) &= P_{M_n}\left(\operatorname{co}\{P_{S,n}(x_n) \mid A \supset S \supset A_n\}\right) \\ &= \operatorname{co}\{P_{M_n} P_{S,n}(x_n) \mid A \supset S \supset A_n\}. \end{aligned}$$

Since x_{n+r} is contained in this set,

$$\begin{aligned} ||x_{n+r}|| &\leq ||x_n|| \max\{||P_{M_n} P_{S,n}|| \mid A \supset S \supset A_n\} \\ &\leq ||x_n|| \max_{\substack{C \subset A \\ C \neq \emptyset}} \left(\max\{||P_{\operatorname{span}\{f_i | i \in C\}} P_{S,n}|| \mid A \supset S \supset C\}\right) \\ &\quad (\text{since } A_n \neq \emptyset) \\ &\leq ||x_n|| \max_{\substack{C \subset A \\ C \neq \emptyset}} ||P_{\operatorname{span}\{f_i | i \in C\}} P_{C,n}|| \\ &= ||x_n|| \max_{\substack{C \subset A \\ C \neq \emptyset}} ||P_{\operatorname{span}\{f_i | i \in C\}} Q_{C,n+r} \cdots Q_{C,n+1}|| \\ &= ||x_n|| \max \left\{ \begin{array}{c} ||P_{\operatorname{span}\{f_i | i \in \{[n_1], \ldots, [n_k]\}\}} P_{H_{[n_k]}} \cdots P_{H_{[n_1]}}|| \quad \Big| \\ n < n_1 < \cdots < n_k \leq n + r, \\ A \supset \{[n_1], \ldots, [n_k]\} \neq \emptyset \end{array} \right\} \\ &= \alpha_n ||x_n||, \quad \text{where} \end{aligned}$$

$$\alpha_n := \max \left\{ \begin{array}{c|} \|P_{(H_{[n_k]}\cap\cdots\cap H_{[n_1]})^\perp} P_{H_{[n_k]}} \cdots P_{H_{[n_1]}}\| \\ n < n_1 < \cdots < n_k \le n+r, \\ A \supset \{[n_1],\ldots,[n_k]\} \neq \emptyset \end{array} \right\}$$

In the last equality we used $(H_{[n_k]}^\perp + \cdots + H_{[n_1]}^\perp) = (H_{[n_k]} \cap \cdots \cap H_{[n_1]})^\perp$ (see [6; Lemma 2.4]). It can be shown that if $[n] = [m]$ then $\alpha_n = \alpha_m$. This implies that $\alpha_n = \alpha_{[n]}$ and thus $\|x_{n+r}\| \le \alpha_{[n]}\|x_n\|$.

Using the facts that for any closed subspaces $M \subset N$, $P_{M^\perp} = I - P_M$ and $P_M P_N = P_M$, we deduce for any $n_1, n_2, \ldots, n_k$,

$$\begin{aligned} \beta := & \left\|P_{(H_{[n_k]}\cap\cdots\cap H_{[n_1]})^\perp} P_{H_{[n_k]}} \cdots P_{H_{[n_1]}}\right\| \\ = & \left\|\left[I - P_{(H_{[n_k]}\cap\cdots\cap H_{[n_1]})}\right] P_{H_{[n_k]}} \cdots P_{H_{[n_1]}}\right\| \\ = & \left\|P_{H_{[n_k]}} \cdots P_{H_{[n_1]}} - P_{(H_{[n_k]}\cap\cdots\cap H_{[n_1]})}\right\|. \end{aligned}$$

By a result of Smith, Solmon, and Wagner [26], we have that

$$\beta \le \left[1 - \prod_1^{k-1} s_i^2\right]^{1/2},$$

where $s_i^2 = 1 - \left[c\left(H_{[n_i]}, \bigcap_{j=i+1}^k H_{[n_j]}\right)\right]^2$. Since the subspaces $H_{[n_i]}$ have finite codimension, it follows by Corollary 3.6 that $\beta < 1$. Since every α_n is the maximum of a finite number of expressions like β, we conclude $\alpha_n < 1$ for all n. We now have the result

$$\|x_{n+r}\| \le \alpha_n\|x_n\| = \alpha_{[n]}\|x_n\|$$

for any $n \ge N_2$. ■

Theorem 3.8. *There exist constants $0 \le c < 1$ and $\rho > 0$ such that*

$$\|x_n - x_\infty\| \le \rho c^n$$

for all n.

Proof. From Lemma 3.7, there exist $\alpha_1, \alpha_2, \ldots, \alpha_r < 1$, and N_2 such that

$$\|x_{n+r}\| \le \alpha_{[n]}\|x_n\|$$

for all $n \ge N_2$. Choose α_l to be the smallest of these. Let N_3 be a number greater than N_2 such that $[N_3] = l$. A simple induction yields $\|x_{N_3+mr}\| \le \alpha_l^m\|x_{N_3}\|$ for all $m \ge 0$. Also, $n \ge N_3 \ge N_1$ implies $x_n = x_{n-1}$ if $[n] \in B$ by Lemma 3.1 or $x_n \in \operatorname{co}\{x_{n-1}, P_{H_{[n]}}(x_{n-1})\}$ if $[n] \in A$ by Theorem 2.4. Note that $H_{[n]}$ is a subspace when $[n] \in A$ so in either case $\|x_n\| \le \|x_{n-1}\|$. We conclude that for all $m \ge 0$,

$$\|x_{N_3+mr+r}\| \le \|x_{N_3+mr+r-1}\| \le \cdots \le \|x_{N_3+mr}\| \le \alpha_l^m\|x_{N_3}\|.$$

Hence, for any $m \ge 0$ and $i = 0, 1, \ldots, r-1$,

$$\begin{aligned}||x_{N_3+mr+i}|| &\le \alpha_l^m ||x_{N_3}|| \\ &\le \alpha_l^m \alpha_l^{(i-r)/r} ||x_{N_3}|| \\ &= \alpha_l^{(mr+i-r)/r} ||x_{N_3}||.\end{aligned}$$

Let $c = \alpha_l^{1/r}$ and $\rho_1 = ||x_{N_3}|| c^{-r-N_3}$. Then

$$||x_{N_3+mr+i}|| \le c^{mr+i-r} ||x_{N_3}|| = c^{N_3+mr+i} c^{-r-N_3} ||x_{N_3}|| = \rho_1 c^{N_3+mr+i}.$$

For any integer $j \ge 0$, the division algorithm implies that $j = mr + i$ for some integer $m \ge 0$ and $0 \le i \le r-1$. Hence we conclude that $||x_{N_3+j}|| \le \rho_1 c^{N_3+j}$ for $j \ge 0$. Let $\rho_2 = \max_{n=0,\ldots,N_3}(||x_n||/c^n)$ and $\rho = \max\{\rho_1, \rho_2\}$. Then $||x_n|| \le \rho c^n$ for all $n \ge 0$. ∎

Up to this point, we have been assuming that $x_\infty = 0$. Suppose now that $x_\infty \ne 0$. Let $\{x_n\}$ be the sequence generated by Dykstra's algorithm with $x_0 = x$, and let $\{x'_n\}$ be the sequence generated by Dykstra's algorithm with $x'_0 = x - x_\infty$ and $\mathcal{H}'_i = \mathcal{H}_i - x_\infty$ ($i = 1, 2, \ldots, r$). That is, $\{x'_n\}$ is the sequence obtained from Dykstra's algorithm when the initial point and all the sets $\mathcal{H}_i$ are translated by the same vector x_∞. A simple induction argument yields

Lemma 3.9. *$x'_n = x_n - x_\infty$ and $e'_n = e_n$ for all $n \ge 0$.*

Using this lemma and applying Theorem 3.8 to $\{x'_n\}$ (where $x'_\infty = x_\infty - x_\infty = 0$), we obtain

$$||x_n - x_\infty|| = ||x'_n|| \le \rho c^n$$

for all $n \ge 0$ for some constants $\rho > 0$ and $0 \le c < 1$. This proves that Theorem 3.8 is also valid for $x_\infty \ne 0$.

4 Intersection of Two Closed Half-Spaces

In the particular case of two half-spaces, Theorem 3.8 of course applies. However, with some extra effort, a substantial strengthening of this result is available. In particular, the constant c is sharp. This result may be stated as follows.

Theorem 4.1. *Either Dykstra's algorithm is finite, or $c := |\langle f_1, f_2\rangle| < 1$ and*

$$||x_n - P_{\mathcal{H}_1\cap\mathcal{H}_2}(x)|| \le c^{n-1} ||x - P_{\mathcal{H}_1\cap\mathcal{H}_2}(x)|| \qquad (n = 1, 2, \ldots). \tag{4.1.1}$$

Moreover, the constant c^{n-1} in (4.1.1) is smallest possible independent of x.

Remark. If the vectors $\{f_1, f_2\}$ are linearly dependent, then $\mathcal{H}_1 \cap \mathcal{H}_2$ is either $\mathcal{H}_1$, $\mathcal{H}_2$, or a "slab" (when $f_1 = -f_2$). In each of these cases, it is not difficult to verify that the algorithm is finite. That is, $x_n = P_{\mathcal{H}_1\cap\mathcal{H}_2}(x)$ for all n sufficiently large. Thus, *we may assume in the sequel that $\{f_1, f_2\}$ is linearly independent and hence* (by the condition of equality in Schwarz's inequality) *that $|\langle f_1, f_2\rangle| < 1$.*

The first lemma involves the dual cone. Recall that the **dual cone** of a nonempty set Y in X is the set

$$Y^\circ := \{x \in X \mid \langle x, y\rangle \leq 0 \quad \text{for all} \quad y \in Y\}.$$

In particular, Y° is a nonempty ($0 \in Y^\circ$) closed convex cone in X. As usual, the *interior* (resp., *boundary*) of Y is denoted by $\operatorname{int} Y$ (resp., $\operatorname{bd} Y$). Finally, the **conical hull** of Y, denoted $\operatorname{con}(Y)$, is the intersection of all convex cones in X which contain Y.

Lemma 4.2. (see [6; Lemma 2.1]) *If K is a closed convex set and $0 \in K$, then*

$$\operatorname{con}(K) = \{\rho y \mid y \in K, \rho > 0\}.$$

Lemma 4.3. *Let $\mathcal{G}_1$ and $\mathcal{G}_2$ be two closed half-spaces with $0 \in \mathcal{G}_1 \cap \mathcal{G}_2$. Thus*

$$\mathcal{G}_i = \{x \in X \mid \langle x, g_i\rangle \leq c_i\} \quad (i = 1, 2),$$

where $g_i \in X\backslash\{0\}$ and $c_i \geq 0$.

(1) *If $0 \in \operatorname{int} \mathcal{G}_1 \cap \operatorname{int} \mathcal{G}_2$, then $(\mathcal{G}_1 \cap \mathcal{G}_2)^\circ = \{0\}$.*
(2) *If $0 \in \operatorname{bd} \mathcal{G}_1 \cap \operatorname{bd} \mathcal{G}_2$, then $(\mathcal{G}_1 \cap \mathcal{G}_2)^\circ = \{\lambda g_1 + \mu g_2 \mid \lambda \geq 0, \mu \geq 0\}$;*
(3) *If $0 \in \operatorname{int} \mathcal{G}_1 \cap \operatorname{bd} \mathcal{G}_2$, then $(\mathcal{G}_1 \cap \mathcal{G}_2)^\circ = \{\mu g_2 \mid \mu \geq 0\}$.*
(4) *If $0 \in \operatorname{bd} \mathcal{G}_1 \cap \operatorname{int} \mathcal{G}_2$, then $(\mathcal{G}_1 \cap \mathcal{G}_2)^\circ = \{\lambda g_1 \mid \lambda \geq 0\}$.*

In particular, in every case we have

(4.3.1) $$(\mathcal{G}_1 \cap \mathcal{G}_2)^\circ = \mathcal{G}_1^\circ + \mathcal{G}_2^\circ.$$

Proof. The proof is an easy consequence of the facts that $\mathcal{G}_i^\circ = \{0\}$ if $0 \in \operatorname{int} \mathcal{G}_i$, $\mathcal{G}_i^\circ = \operatorname{con}\{g_i\} = \{\lambda g_i \mid \lambda \geq 0\}$ if $0 \in \operatorname{bd} \mathcal{G}_i$, and $(\mathcal{G}_1 \cap \mathcal{G}_2)^\circ = \overline{\mathcal{G}_1^\circ + \mathcal{G}_2^\circ}$ (see [6; Lemma 2.4]), and we omit the details. ■

Corollary 4.4. *If $x_n = P_{H_1 \cap H_2}(x)$ for some n, then*

(4.4.1) $$x_m = P_{H_1 \cap H_2}(x)$$

for all $m \geq n$, and hence

(4.4.2) $$P_{\mathcal{H}_1 \cap \mathcal{H}_2}(x) = P_{H_1 \cap H_2}(x).$$

Proof. Let $y = P_{H_1 \cap H_2}(x) = x_n$. By Theorem 2.4,

$$x_{n+1} \in \operatorname{co}\{x_n, P_{H_{[n]}}(x_n)\} = \operatorname{co}\{y, P_{H_{[n]}}(y)\} = \operatorname{co}\{y, y\} = \{y\}.$$

That is, $x_{n+1} = y$. By induction, it follows that $x_m = y$ for all $m \geq n$. Also, since $x_n \to P_{\mathcal{H}_1 \cap \mathcal{H}_2}(x)$, it follow that $y = P_{\mathcal{H}_1 \cap \mathcal{H}_2}(x)$. ∎

The most interesting case of Theorem 4.1 is when $P_{\mathcal{H}_1 \cap \mathcal{H}_2}(x)$ is on the boundary of both $\mathcal{H}_1$ and $\mathcal{H}_2$. That is, $P_{\mathcal{H}_1 \cap \mathcal{H}_2}(x) \in H_1 \cap H_2$. It turns out that for all other cases, the Dykstra algorithm is finite. To handle this case, the following two lemmas will be useful. The first lemma is related to a general result of Fan [14; Theorem 14]. In our case, a shorter proof can be given.

Lemma 4.5. *If $P_{\mathcal{H}_1 \cap \mathcal{H}_2}(x) \in H_1 \cap H_2$, then*

$$x - P_{\mathcal{H}_1 \cap \mathcal{H}_2}(x) = \alpha f_1 + \beta f_2 \tag{4.5.1}$$

for some scalars $\alpha, \beta \geq 0$.

Proof. Let $x_\infty = P_{\mathcal{H}_1 \cap \mathcal{H}_2}(x)$. By the characterization theorem for best approximations using the dual cone approach ([8] or [12]), we see that $x - x_\infty \in (\mathcal{H}_1 \cap \mathcal{H}_2 - x_\infty)^\circ$. But $(\mathcal{H}_1 \cap \mathcal{H}_2 - x_\infty)^\circ = [(\mathcal{H}_1 - x_\infty) \cap (\mathcal{H}_2 - x_\infty)]^\circ$ and $\mathcal{H}_1 - x_\infty$ and $\mathcal{H}_2 - x_\infty$ are closed half-spaces with $0 \in (\mathcal{H}_1 - x_\infty) \cap (\mathcal{H}_2 - x_\infty)$. In fact, since $x_\infty \in \operatorname{bd}\mathcal{H}_1 \cap \operatorname{bd}\mathcal{H}_2$, it follows that $0 \in \operatorname{bd}(\mathcal{H}_1 - x_\infty) \cap \operatorname{bd}(\mathcal{H}_2 - x_\infty)$. By Lemma 4.3, (with $\mathcal{G}_i = \mathcal{H}_i - x_\infty$), we obtain that

$$x - x_\infty \in (\mathcal{H}_1 - x_\infty)^\circ + (\mathcal{H}_2 - x_\infty)^\circ = \{\lambda f_1 + \mu f_2 \mid \lambda \geq 0, \mu \geq 0\}.$$

This completes the proof. ∎

Lemma 4.6. *Suppose that $P_{\mathcal{H}_1 \cap \mathcal{H}_2}(x) \in H_1 \cap H_2$, and hence $x - P_{\mathcal{H}_1 \cap \mathcal{H}_2}(x) = \alpha f_1 + \beta f_2$ for some scalars $\alpha, \beta \geq 0$. For any integer $n \geq 1$, let*

$$\delta(n) := \alpha\langle f_1, f_{[n]}\rangle + \beta\langle f_2, f_{[n]}\rangle - k_{n-1}\langle f_1, f_2\rangle,$$

where k_{n-1} is defined as in Lemma 2.2. Then

$$x_n = P_{H_{[n]}}(x_{n-1}) \tag{4.6.1}$$

if and only if

$$\delta(n) \geq 0. \tag{4.6.2}$$

Proof. By Lemma 2.3, (4.6.1) holds if and only if

$$\langle x_{n-1} + e_{n-2}, f_{[n]}\rangle - c_{[n]} \geq 0. \tag{4.6.3}$$

But $x - x_{n-1} = e_{n-1} + e_{n-2}$ by Lemma 2.5. Lemma 2.2 implies

$$x_{n-1} + e_{n-2} = x - e_{n-1} = x_\infty + \alpha f_1 + \beta f_2 - k_{n-1} f_{[n-1]}, \tag{4.6.4}$$

where $x_\infty = P_{\mathcal{H}_1\cap\mathcal{H}_2}(x)$. Substituting (4.6.4) into (4.6.3), we see that (4.6.3) holds if and only if $\delta(n) \geq 0$. ■

Remark. The proof also shows that $\delta(n) = 0 \Leftrightarrow \langle x_{n-1} + e_{n-2}, f_{[n]}\rangle = c_{[n]} \Leftrightarrow x_{n-1} + e_{n-2} \in H_{[n]}$.

We are now in a position to prove Theorem 4.1.

Proof of Theorem 4.1. We will use the definitions and notation developed for the lemmas of sections 2 and 4.

By the well-known characterization of best approximations from closed convex subsets of a Hilbert space (see, e.g., [12; Corollary 3.1]), we deduce that

$$x - x_\infty \in (\mathcal{H}_1 \cap \mathcal{H}_2 - x_\infty)^\circ = [(\mathcal{H}_1 - x_\infty) \cap (\mathcal{H}_2 - x_\infty)]^\circ$$

Note that $\mathcal{H}_i - x_\infty \quad (i = 1, 2)$ is a closed half-space which contains 0. There are several cases to consider. They are summarized below along with the number of iterations sufficient for convergence:

TABLE I

Case		Conditions		Iterations
1	$x_\infty \in \operatorname{int}\mathcal{H}_1 \cap \operatorname{int}\mathcal{H}_2$			0
2	$x_\infty \in \operatorname{bd}\mathcal{H}_1 \cap \operatorname{int}\mathcal{H}_2$			1
3	$x_\infty \in \operatorname{int}\mathcal{H}_1 \cap \operatorname{bd}\mathcal{H}_2$			finite
4	$x_\infty \in \operatorname{bd}\mathcal{H}_1 \cap \operatorname{bd}\mathcal{H}_2$			
4a		$\langle f_1, f_2\rangle = 0$		2
4b		$\langle f_1, f_2\rangle > 0$		∞
4c		$\langle f_1, f_2\rangle < 0$		
4ci			$x \notin \mathcal{H}_1 \cup \mathcal{H}_2$	∞
4cii			$x \in \mathcal{H}_2 \backslash \mathcal{H}_1$	∞
4ciii			$x \in \mathcal{H}_1 \backslash \mathcal{H}_2$	∞
4civ			$x \in \mathcal{H}_1 \cap \mathcal{H}_2$	0

Case 1: $x_\infty \in \operatorname{int}\mathcal{H}_1 \cap \operatorname{int}\mathcal{H}_2$.

Then $0 \in \operatorname{int}(\mathcal{H}_1 - x_\infty) \cap \operatorname{int}(\mathcal{H}_2 - x_\infty)$ so that Lemma 4.3(1) implies

$$x - x_\infty \in [(\mathcal{H}_1 - x_\infty) \cap (\mathcal{H}_2 - x_\infty)]^\circ = \{0\},$$

and hence $x = x_\infty$. It follows that $x_n = x_\infty$ for all n, so Dykstra's algorithm is finite.

Case 2: $x_\infty \in \operatorname{bd}\mathcal{H}_1 \cap \operatorname{int}\mathcal{H}_2$.

Then $0 \in \operatorname{bd}(\mathcal{H}_1 - x_\infty) \cap \operatorname{int}(\mathcal{H}_2 - x_\infty)$ and Lemma 4.3(4) implies that

$$x - x_\infty \in [(\mathcal{H}_1 - x_\infty) \cap (\mathcal{H}_2 - x_\infty)]^\circ = \{\lambda f_1 \mid \lambda \geq 0\}.$$

Thus $x - x_\infty = \rho f_1$ for some $\rho \geq 0$. If $\rho = 0$, $x_n = x_\infty$ for all n. Thus we may assume that $\rho > 0$. We will verify the following

Claim. *$x_n = x_\infty$ for all n, and $e_n = 0$ if n is even and $e_n = \rho f_1$ if n is odd for all $n \geq 1$.*

Proof. To prove the claim, we proceed by induction on n. Since

$$\langle x, f_1\rangle = \langle x_\infty + \rho f_1, f_1\rangle = c_1 + \rho > c_1,$$

we have $x \notin \mathcal{H}_1$. Thus, using Lemma 2.1,

$$\begin{aligned} x_1 &= P_{\mathcal{H}_1}(x_0 + e_{-1}) = P_{\mathcal{H}_1}(x) = P_{H_1}(x) = P_{H_1}(x - \rho f_1) \\ &= P_{H_1}(x_\infty) = x_\infty, \qquad \text{and} \\ e_1 &= x_0 + e_{-1} - x_1 = x - x_\infty = \rho f_1. \end{aligned}$$

This proves the claim when $n = 1$. Now assume the claim is valid for all positive integers less that or equal to n. Consider the cases:

(a) n is odd.

Then $n + 1$ is even so that

$$\begin{aligned} x_{n+1} &= P_{\mathcal{H}_2}(x_n + e_{n-1}) = P_{\mathcal{H}_2}(x_n) = P_{\mathcal{H}_2}(x_\infty) = x_\infty \\ \text{and} \qquad e_{n+1} &= x_n + e_{n-1} - x_{n+1} = e_{n-1} = 0. \end{aligned}$$

(b) n is even.

Then $n + 1$ is odd so that

$$\begin{aligned} x_{n+1} &= P_{\mathcal{H}_1}(x_n + e_{n-1}) = P_{\mathcal{H}_1}(x_\infty + \rho f_1) = P_{\mathcal{H}_1}(x) \\ &= x_1 = x_\infty \end{aligned}$$

using the case $n = 1$. Also, $e_{n+1} = x_n + e_{n-1} - x_{n+1} = \rho f_1$.

This completes the induction and proves the claim. From the claim it follows that Dykstra's algorithm is finite.

Case 3: $x_\infty \in \operatorname{int}\mathcal{H}_1 \cap \operatorname{bd}\mathcal{H}_2$.

Then $0 \in \operatorname{int}(\mathcal{H}_1 - x_\infty) \cap \operatorname{bd}(\mathcal{H}_2 - x_\infty)$ and Lemma 4.3(3) implies that

$$x - x_\infty \in [(\mathcal{H}_1 - x_\infty) \cap (\mathcal{H}_2 - x_\infty)]^\circ = \{\lambda f_2 \mid \lambda \geq 0\}.$$

Thus

$$x - x_\infty = \rho f_2 \tag{4.1.2}$$

for some $\rho \geq 0$. If $\rho = 0$, $x = x_\infty$ so that $x_n = x_\infty$ for all n and Dykstra's algorithm is finite. Thus we may assume $\rho > 0$.

By the Boyle-Dykstra Theorem, $x_n \longrightarrow x_\infty$. Thus there is an integer N so that $x_n \in \operatorname{int}\mathcal{H}_1$ for all $n \geq N$. Since $x_n = P_{\mathcal{H}_{[n]}}(x_{n-1} + e_{n-2})$, it follows that $x_{n-1} + e_{n-2} = x_n$ and $e_n = 0$ for all $n \geq N$ with $[n] = 1$. In particular, we deduce that

(4.1.3)
$$e_{n-2} = 0 \qquad \text{and} \qquad x_{n-1} = x_n$$

for all $n \geq N+2$ with $[n] = 1$.

Next suppose there exists $n \geq N+2$ with $[n] = 2$ and $x_n \in \operatorname{int} \mathcal{H}_2$. Then

$$P_{\mathcal{H}_2}(x_{n-1} + e_{n-2}) = P_{\mathcal{H}_{[n]}}(x_{n-1} + e_{n-2}) = x_n \in \operatorname{int} \mathcal{H}_2$$

implies $x_{n-1} + e_{n-2} = x_n$ and $e_n = 0$. Now Lemma 2.5 implies that $x - x_n = e_n + e_{n-1} = 0$ using (4.1.3). Thus $x = x_n \in \operatorname{int} \mathcal{H}_1 \cap \operatorname{int} \mathcal{H}_2$ which implies that $x_\infty \in \operatorname{int} \mathcal{H}_1 \cap \operatorname{int} \mathcal{H}_2$, contradicting the assumption of case 3.

Thus we must have that

(4.1.4)
$$x_n \in H_2 := \operatorname{bd} \mathcal{H}_2$$

for all $n \geq N+2$ with $[n] = 2$. Using (4.1.3) and (4.1.4), we deduce that $x_n \in H_2$ for all $n \geq N+2$. From Lemmas 2.2 and 2.5, we obtain

(4.1.5)
$$x - x_n = e_n + e_{n-1} = k_n f_{[n]} + k_{n-1} f_{[n-1]}$$

for all $n \geq 2$. Using (4.1.3) and (4.1.5), we obtain

(4.1.6)
$$x - x_n = \rho_n f_2$$

for all $n \geq N+2$ and some $\rho_n \geq 0$. Combining (4.1.6) with (4.1.2), we obtain that

(4.1.7)
$$x_n - x_\infty = (\rho - \rho_n) f_2$$

for all $n \geq N+2$. By (4.1.4), $x_n \in H_2$ for all $n \geq N+2$ and $x_\infty \in H_2$ by hypothesis. Thus, for all $n \geq N+2$,

$$0 = \langle x_n - x_\infty, f_2 \rangle = (\rho - \rho_n)\langle f_2, f_2 \rangle = \rho - \rho_n.$$

By (4.1.7), this implies that $x_n = x_\infty$ for all $n \geq N+2$. In particular, Dykstra's algorithm is finite.

Case 4: $x_\infty \in \operatorname{bd} \mathcal{H}_1 \cap \operatorname{bd} \mathcal{H}_2$.

Hence $x - x_\infty = \alpha f_1 + \beta f_2$ for some $\alpha, \beta \geq 0$. If $\beta = 0$, $x - x_\infty = \alpha f_1$ and the same argument as in Case 2 shows that $x_n = x_\infty$ for all $n \geq 1$. Hence the algorithm is finite.

Thus we may assume $\beta > 0$.

Case 4a: $\langle f_1, f_2 \rangle = 0$.

Then

$$\langle x, f_1 \rangle = \langle x_\infty + \alpha f_1 + \beta f_2, f_1 \rangle = c_1 + \alpha \geq c_1,$$

so $x \notin \operatorname{int} \mathcal{H}_1 \Rightarrow x_1 = P_{\mathcal{H}_1}(x) = P_{H_1}(x)$ and

$$e_1 = x + e_{-1} - x_1 = x - P_{H_1}(x) = [\langle x, f_1 \rangle - c_1] f_1 = \alpha f_1.$$

Also,

$$\begin{aligned}\langle x_1, f_2\rangle &= \langle x - \alpha f_1, f_2\rangle = \langle x, f_2\rangle \\ &= \langle x_\infty + \alpha f_1 + \beta f_2, f_2\rangle = c_2 + \beta > c_2\end{aligned}$$

$\Rightarrow x_1 \notin \mathcal{H}_2 \Rightarrow$

$$\begin{aligned}x_2 &= P_{\mathcal{H}_2}(x_1 + e_0) = P_{\mathcal{H}_2}(x_1) = P_{H_2}(x_1) \\ &= P_{H_2}(x - \alpha f_1) = P_{H_2}(x_\infty + \beta f_2) = P_{H_2}(x_\infty) = x_\infty\end{aligned}$$

using (4.1.4). By Corollary 4.4, $x_n = x_\infty$ for all $n \geq 2$. Hence the algorithm is finite.

Case 4b: $\langle f_1, f_2\rangle > 0$.

Define

$$\delta(n) := \alpha\langle f_1, f_{[n]}\rangle + \beta\langle f_2, f_{[n]}\rangle - k_{n-1}\langle f_1, f_2\rangle$$

for all $n \geq 1$ as in Lemma 4.6. Then $\delta(n) \geq (\alpha + \beta - k_{n-1})\langle f_1, f_2\rangle$. To show $\delta(n) \geq 0$ for all $n \geq 1$, it suffices to show that

$$\alpha + \beta - k_{n-1} \geq 0 \tag{4.1.8}$$

for all $n \geq 1$. We proceed by induction on n. For $n = 1$, $k_0 = 0$ so $\alpha + \beta - k_0 = \alpha + \beta \geq 0$. Now suppose (4.1.8) is valid for some $n \geq 1$. If $k_n = 0$, then $\alpha + \beta - k_n = \alpha + \beta \geq 0$ so (4.1.8) holds with n replaced by $n+1$. If $k_n \neq 0$, then $k_n > 0$ and Lemma 2.2 implies that $k_n =: \langle x_{n-1} + e_{n-2}, f_{[n]}\rangle - c_{[n]}$. But $x_{n-1} + e_{n-2} = x - e_{n-1}$ by Lemma 2.5, so $x_{n-1} + e_{n-2} = x_\infty + \alpha f_1 + \beta f_2 - k_{n-1} f_{[n-1]}$. Hence

$$\begin{aligned}k_n &= \langle x_\infty + \alpha f_1 + \beta f_2 - k_{n-1} f_{[n-1]}, f_{[n]}\rangle - c_{[n]} \\ &= \alpha\langle f_1, f_{[n]}\rangle + \beta\langle f_2, f_{[n]}\rangle - k_{n-1}\langle f_1, f_2\rangle.\end{aligned}$$

It follows that

$$\alpha + \beta - k_n = \alpha[1 - \langle f_1, f_{[n]}\rangle] + \beta[1 - \langle f_2, f_{[n]}\rangle] + k_{n-1}\langle f_1, f_2\rangle \geq 0$$

since each term on the right is nonnegative. This shows that (4.1.8) holds with n replaced by $n+1$ and thus completes the induction. Thus $\delta(n) \geq 0$ for all $n \geq 1$. By Lemma 4.6, we deduce that $x_n = P_{H_{[n]}}(x_{n-1})$ for all $n \geq 1$.

Case 4c: $\langle f_1, f_2\rangle < 0$.

Suppose that $\alpha = 0$. Then $\langle x, f_1\rangle = \langle x_\infty + \beta f_2, f_1\rangle = c_1 + \beta\langle f_1, f_2\rangle < c_1$ implies that $x \in \operatorname{int}\mathcal{H}_1$. If $x \in \operatorname{int}\mathcal{H}_2$, then $x \in \operatorname{int}\mathcal{H}_1 \cap \operatorname{int}\mathcal{H}_2$ implies that $x_\infty = x \in \operatorname{int}\mathcal{H}_1 \cap \operatorname{int}\mathcal{H}_2$, which is a contradiction. Hence $x \notin \operatorname{int}\mathcal{H}_2$. Now $x \in \operatorname{int}\mathcal{H}_1$ so that $x_1 = P_{\mathcal{H}_1}(x) = x \in \operatorname{int}\mathcal{H}_1$ and $e_1 = x - x_1 = 0$. Since $x \notin \operatorname{int}\mathcal{H}_2$,

$$\begin{aligned}x_2 &= P_{\mathcal{H}_2}(x_1 + e_0) = P_{\mathcal{H}_2}(x) = P_{H_2}(x) \\ &= P_{H_2}(x_\infty + \beta f_2) = P_{H_2}(x_\infty) = x_\infty\end{aligned}$$

using Lemma 2.1. By Corollary 4.4, $x_n = x_\infty$ for all $n \geq 2$ and the algorithm is finite.

Thus we may assume that $\alpha > 0$. We will show that

$$\delta(n) := \alpha\langle f_1, f_{[n]}\rangle + \beta\langle f_2, f_{[n]}\rangle - k_{n-1}\langle f_1, f_2\rangle \geq 0 \tag{4.1.9}$$

for all $n \geq 2$.

Case 4c(i): $x \notin \mathcal{H}_1 \cup \mathcal{H}_2$.

Then $\langle x, f_i\rangle > c_i \quad (i = 1, 2)$ implies

$$\begin{aligned} c_i < \langle x_\infty + \alpha f_1 + \beta f_2, f_i\rangle = c_i + \alpha\langle f_1, f_i\rangle + \beta\langle f_2, f_i\rangle \qquad \text{or} \\ \alpha\langle f_1, f_i\rangle + \beta\langle f_2, f_i\rangle > 0 \quad (i = 1, 2). \end{aligned} \tag{4.1.10}$$

Thus $\delta(n) \geq \alpha\langle f_1, f_{[n]}\rangle + \beta\langle f_2, f_{[n]}\rangle > 0$ by (4.1.10). Thus (4.1.9) holds for all $n \geq 1$.

Case 4c(ii): $x \in \mathcal{H}_2 \backslash \mathcal{H}_1$.

Then $\langle x, f_1\rangle > c_1$ and $\langle x, f_2\rangle \leq c_2$ implies $\alpha + \beta\langle f_2, f_1\rangle > 0$ and $\alpha\langle f_1, f_2\rangle + \beta \leq 0$. If n is odd, then

$$\delta(n) = \alpha + \beta\langle f_2, f_1\rangle - k_{n-1}\langle f_1, f_2\rangle > -k_{n-1}\langle f_1, f_2\rangle \geq 0$$

since $k_{n-1} \geq 0$. If $n = 2m$ is even, then

$$\delta(n) = \delta(2m) = \alpha\langle f_1, f_2\rangle + \beta - k_{2m-1}\langle f_1, f_2\rangle. \tag{4.1.11}$$

Since $x_{2m-2} + e_{2m-3} = x - e_{2m-2}$ by Lemma 2.5, it follows that

$$\begin{aligned} \langle x_{2m-2} + e_{2m-3}, f_1\rangle &= \langle x - e_{2m-2}, f_1\rangle \\ &= \langle x_\infty + \alpha f_1 + \beta f_2 - e_{2m-2}, f_1\rangle \\ &= c_1 + \alpha + \beta\langle f_2, f_1\rangle - \langle e_{2m-2}, f_1\rangle \\ &= c_1 + \alpha + \beta\langle f_2, f_1\rangle - k_{2m-2}\langle f_2, f_1\rangle \\ &> c_1 - k_{2m-2}\langle f_2, f_1\rangle \geq c_1. \end{aligned} \tag{4.1.12}$$

Thus $x_{2m-2} + e_{2m-3} \notin \mathcal{H}_1$ implies by Lemma 2.2, that

$$\begin{aligned} 0 < k_{2m-1} = d(x_{2m-2} + e_{2m-3}, H_1) &= \langle x_{2m-2} + e_{2m-3}, f_1\rangle - c_1 \\ &= \alpha + \beta\langle f_1, f_2\rangle - k_{2m-2}\langle f_2, f_1\rangle \end{aligned}$$

using (4.1.12). Substituting this expression for k_{2m-1} into (4.1.11), we obtain

$$\begin{aligned} \delta(n) &= \alpha\langle f_1, f_2\rangle + \beta - [\alpha + \beta\langle f_2, f_1\rangle - k_{2m-2}\langle f_1, f_2\rangle]\langle f_1, f_2\rangle \\ &= \beta[1 - \langle f_1, f_2\rangle^2] + k_{2m-2}\langle f_1, f_2\rangle^2 \geq 0. \end{aligned}$$

This proves that $\delta(n) \geq 0$ for all $n \geq 1$.

Case 4c(iii): $x \in \mathcal{H}_1 \backslash \mathcal{H}_2$.

Then $\langle x, f_1 \rangle \le c_1$ and $\langle x, f_2 \rangle > c_2$, which implies that $\alpha + \beta\langle f_2, f_1 \rangle \le 0$ and $\alpha\langle f_1, f_2 \rangle + \beta > 0$. Now $x_1 = P_{\mathcal{H}_1}(x) = x$ and $e_1 = x + e_0 - x_1 = 0$. If n is even, then

$$\delta(n) = \alpha\langle f_2, f_1 \rangle + \beta - k_{n-1}\langle f_1, f_2 \rangle > -k_{n-1}\langle f_1, f_2 \rangle \ge 0.$$

If $n = 2m + 1$ is odd, $m \ge 1$, then

$$\delta(n) = \delta(2m+1) = \alpha + \beta\langle f_2, f_1 \rangle - k_{2m}\langle f_1, f_2 \rangle. \tag{4.1.13}$$

Since

$$x_{2m-1} + e_{2m-2} = x - e_{2m-1}$$

by Lemma 2.5, it follows that

$$\begin{aligned} \langle x_{2m-1} + e_{2m-2}, f_2 \rangle = \langle x - e_{2m-1}, f_2 \rangle &= \langle x_\infty + \alpha f_1 + \beta f_2 - e_{2m-1}, f_2 \rangle \\ &= c_2 + \alpha\langle f_1, f_2 \rangle + \beta - k_{2m-1}\langle f_1, f_2 \rangle \\ &> c_2 - k_{2m-1}\langle f_1, f_2 \rangle \ge c_2 \end{aligned} \tag{4.1.14}$$

implies that $x_{2m-1} + e_{2m-2} \notin \mathcal{H}_2$. Thus, by Lemma 2.2,

$$\begin{aligned} 0 < k_{2m} = d(x_{2m-1} + e_{2m-2}, H_2) &= \langle x_{2m-1} + e_{2m-2}, f_2 \rangle - c_2 \\ &= \alpha\langle f_1, f_2 \rangle + \beta - k_{2m-1}\langle f_1, f_2 \rangle \end{aligned}$$

using (4.1.14). Substituting into (4.1.13), we obtain

$$\begin{aligned} \delta(n) &= \alpha + \beta\langle f_1, f_2 \rangle - [\alpha\langle f_1, f_2 \rangle + \beta - k_{2m-1}\langle f_1, f_2 \rangle]\langle f_1, f_2 \rangle \\ &= \alpha[1 - \langle f_1, f_2 \rangle^2] + k_{2m-1}\langle f_1, f_2 \rangle^2 \ge 0. \end{aligned}$$

This proves that $\delta(n) \ge 0$ for all $n \ge 2$.

Case 4c(iv): $x \in \mathcal{H}_1 \cap \mathcal{H}_2$.

Then $x = x_\infty$ which contradicts the assumption that $x = x_\infty + \alpha f_1 + \beta f_2$ and $\beta > 0$.

Using Lemma 4.6, we may summarize all the cases as follows: in every case when $x_\infty \notin H_1 \cap H_2$, Dykstra's algorithm is finite. When $x_\infty \in H_1 \cap H_2$ and Dykstra's algorithm is not finite, then either

$$x_n = P_{H_{[n]}}(x_{n-1}) \quad \text{for all } n \ge 1 \quad \text{or,} \tag{4.1.15}$$

$$x = x_1 \text{ and } x_n = P_{H_{[n]}}(x_{n-1}) \quad \text{for all } n \ge 2. \tag{4.1.16}$$

Assume that Dykstra's algorithm is not finite. It follows that

$$x_\infty = P_{\mathcal{H}_1 \cap \mathcal{H}_2}(x) = P_{H_1 \cap H_2}(x)$$

and

$$x - x_\infty = \alpha f_1 + \beta f_2, \; \alpha \geq 0, \; \beta \geq 0.$$

Let $\gamma := -\langle f_1, f_2 \rangle$ so that $c = |\gamma|$.

We first verify that if (4.1.15) holds, then

$$x_{2n-1} - x_\infty = \gamma^{2(n-1)}(x_1 - x_\infty) \tag{4.1.17}$$

and

$$x_{2n} - x_\infty = \gamma^{2(n-1)}(x_2 - x_\infty) \tag{4.1.18}$$

for all $n \geq 1$.

We prove this by simultaneous induction on (4.1.17) and (4.1.18). For $n = 1$, both equalities are obvious. Let

$$M_i := \{x \in X \mid \langle f_i, x \rangle = 0\} = H_i - x_\infty \quad (i = 1, 2).$$

Now,

$$\begin{aligned} x_1 - x_\infty &= P_{H_1}(x) - x_\infty = P_{M_1}(x - x_\infty) = P_{M_1}(\alpha f_1 + \beta f_2) \\ &= \beta P_{M_1}(f_2) = \beta[f_2 - \langle f_2, f_1 \rangle f_1] = \beta(\gamma f_1 + f_2) \end{aligned} \tag{4.1.19}$$

and thus

$$\begin{aligned} x_2 - x_\infty &= P_{H_2}(x_1) - x_\infty = P_{M_2}(x_1 - x_\infty) = P_{M_2}[\beta(\gamma f_1 + f_2)] \\ &= \beta\gamma P_{M_2}(f_1) = \beta\gamma[f_1 - \langle f_2, f_1 \rangle f_2] = \beta\gamma(f_1 + \gamma f_2). \end{aligned} \tag{4.1.20}$$

Now suppose (4.1.17) and (4.1.18) hold for some $n \geq 1$. Then

$$\begin{aligned} x_{2n+1} - x_\infty &= P_{H_1}(x_{2n}) - x_\infty = P_{M_1}(x_{2n} - x_\infty) \\ &= P_{M_1}[\gamma^{2(n-1)}(x_2 - x_\infty)] = \gamma^{2(n-1)} P_{M_1}(x_2 - x_\infty) \\ &= \gamma^{2(n-1)} P_{M_1}[\beta\gamma(f_1 + \gamma f_2)] = \gamma^{2n}\beta P_{M_1}(f_2) \\ &= \gamma^{2n}\beta[f_2 - \langle f_2, f_1 \rangle f_1] = \gamma^{2n}\beta(\gamma f_1 + f_2) \\ &= \gamma^{2n}(x_1 - x_\infty) \end{aligned} \tag{4.1.21}$$

using (4.1.19) and (4.1.20). This proves that (4.1.17) holds with n replaced by $n + 1$. Finally, using (4.1.19)–(4.1.21), we obtain

$$\begin{aligned} x_{2n+2} - x_\infty &= P_{H_2}(x_{2n+1}) - x_\infty = P_{M_2}(x_{2n+1} - x_\infty) \\ &= P_{M_2}[\gamma^{2n}(x_1 - x_\infty)] = \gamma^{2n} P_{M_2}(x_1 - x_\infty) \\ &= \gamma^{2n} P_{M_2}[\beta(\gamma f_1 + f_2)] = \gamma^{2n+1}\beta P_{M_2}(f_1) \end{aligned}$$

$$= \gamma^{2n+1}\beta[f_1 - \langle f_1, f_2\rangle f_2] = \gamma^{2n+1}\beta[f_1 + \gamma f_2]$$
$$= \gamma^{2n}(x_2 - x_\infty)$$

which shows that (4.1.18) holds with n replaced by $n+1$.

In the same way, we can verify that if (4.1.16) holds, then

$$x_{2n} - x_\infty = \gamma^{2(n-1)}(x_2 - x_\infty) \quad \text{and} \tag{4.1.22}$$
$$x_{2n+1} - x_\infty = \gamma^{2(n-1)}(x_3 - x_\infty) \quad (n = 1, 2, \dots), \tag{4.1.23}$$

where

$$x_2 - x_\infty = \alpha(f_1 + \gamma f_2) \quad \text{and} \tag{4.1.24}$$
$$x_3 - x_\infty = \alpha\gamma(\gamma f_1 + f_2). \tag{4.1.25}$$

To complete the proof of Theorem 4.1, assume first that (4.1.15) holds. Then using (4.1.17)–(4.1.19), we get

$$||x_{2n} - x_\infty|| = \gamma^{2(n-1)}||x_2 - x_\infty||. \tag{4.1.26}$$

But

$$||x_1 - x_\infty||^2 = ||\beta(\gamma f_1 + f_2)||^2 = \beta^2(\gamma^2 - 2\gamma^2 + 1) = \beta^2(1 - \gamma^2)$$

and

$$||x_2 - x_\infty||^2 = ||\beta\gamma(f_1 + \gamma f_2)||^2 = \beta^2\gamma^2(1 - 2\gamma^2 + \gamma^2)$$
$$= \beta^2\gamma^2(1 - \gamma^2) = \gamma^2||x_1 - x_\infty||^2$$

imply that $||x_2 - x_\infty|| = |\gamma|\ ||x_1 - x_\infty||$. Thus from (4.1.26) we obtain

$$||x_{2n} - x_\infty|| = |\gamma|^{2n-1}||x_1 - x_\infty|| = c^{2n-1}||x_1 - x_\infty||. \tag{4.1.27}$$

Since $||x_1 - x_\infty|| = ||P_{H_1}(x) - x_\infty|| = ||P_{M_1}(x - x_\infty)|| \le ||x - x_\infty||$, we deduce from (4.1.27) that $||x_{2n} - x_\infty|| \le c^{2n-1}||x - x_\infty||$.

Assume next that (4.1.16) holds. Then using (4.1.22)–(4.1.25), we obtain $||x_{2n} - x_\infty|| = \gamma^{2(n-1)}||x_2 - x_\infty||$. But

$$||x - x_\infty||^2 = ||\alpha f_1 + \beta f_2||^2 = \alpha^2 - 2\alpha\beta\gamma + \beta^2 \quad \text{and}$$
$$||x_2 - x_\infty||^2 = ||\alpha(f_1 + \gamma f_2)||^2 = \alpha^2(1 - 2\gamma^2 + \gamma^2) = \alpha^2(1 - \gamma^2).$$

Now choose $\lambda \in (0,1]$ so that the element $x_1' := \lambda x + (1-\lambda)x_2$ is in H_1 (i.e., $\langle x_1', f_1\rangle = c_1$). In particular, $\lambda = 1$ if $x \in H_1$ (so $x = x_1 = x_1'$) and $\lambda = \frac{\alpha(1-\gamma^2)}{\gamma(\beta-\alpha\gamma)}$ if $x \notin H_1$. In the former case, it is easy to deduce that $\alpha = \beta\gamma$. In either case, it is straightforward, but a little tedious, to verify the second equality in the equation

$$\frac{||x_2 - x_\infty||^2}{||x_1' - x_\infty||^2} = \frac{\alpha^2(1-\gamma^2)}{\lambda^2||x - x_\infty||^2 + (1-\lambda^2)||x_2 - x_\infty||^2} = \gamma^2 = c^2.$$

Thus $||x_2 - x_\infty|| = c||x_1' - x_\infty||$ implies that $||x_{2n} - x_\infty|| = c^{2n-1}||x_1' - x_\infty||$. Since $P_{H_2}(x_1') = P_{H_2}(x_1) = x_2$, we deduce

$$\begin{aligned} ||x_1' - x_\infty||^2 &= ||x_1' - P_{H_2}(x_1')||^2 + ||P_{H_2}(x_1') - x_\infty||^2 \\ &= ||x_1' - x_2||^2 + ||x_2 - x_\infty||^2 \\ &\leq ||x_1 - x_2||^2 + ||x_2 - x_\infty||^2 = ||x_1 - x_\infty||^2 = ||x - x_\infty||^2. \end{aligned}$$

This implies $||x_1' - x_\infty|| \leq ||x - x_\infty||$ so that $||x_{2n} - x_\infty|| \leq c^{2n-1}||x - x_\infty||$. The preceding two paragraphs verify (4.1.1) when n is even.

To prove (4.1.1) for n odd, we note that if (4.1.15) holds, then

$$||x_{2n+1} - x_\infty|| = c^{2n}||x_1 - x_\infty|| \leq c^{2n}||x - x_\infty||$$

by (4.1.21). If (4.1.16) holds, then by (4.1.22)–(4.1.25),

$$\frac{||x_{2n} - x_\infty||}{||x_{2n+1} - x_\infty||} = \frac{||x_2 - x_\infty||}{||x_3 - x_\infty||} = \frac{||f_1 + \gamma f_2||}{c||\gamma f_1 + f_2||} = \frac{1}{c}.$$

Thus

$$||x_{2n+1} - x_\infty|| = c||x_{2n} - x_\infty|| \leq c\, c^{2n-1}||x - x_\infty|| = c^{2n}||x - x_\infty||.$$

In either case (4.1.1) holds. Thus (4.1.1) holds when n is odd.

It remains to observe that the constant $c^{n-1} = |\gamma|^{n-1}$ is the best possible independent of x. But this is a consequence of the fact that if $x = x_1 \in H_1$, then $\lambda = 1$ in the above and we obtain the equality $||x_n - x_\infty|| = c^{n-1}||x - x_\infty||$. This completes the proof of Theorem 4.1. ∎

5 Applications

Theorem 3.8 states that Dykstra's algorithm converges like $O(c^n)$ for some $c < 1$. The exact value for c can't be computed unless A is known, but a bound can be developed. To form a crude bound we can take a maximum over all the subsets A of $\{1, \ldots, r\}$ and limit our attention to α_r as follows. In the notation of section 3,

$$
\begin{aligned}
c^r = \alpha_l \le \alpha_r &= \alpha_0 \\
&= \max\Big\{ \left\| P_{H_{[n_k]}} \cdots P_{H_{[n_1]}} - P_{(H_{[n_k]} \cap \cdots \cap H_{[n_1]})} \right\| \ \Big| \\
&\qquad\qquad 0 < n_1 < n_2 < \cdots < n_k \le r, \quad \{[n_1], \ldots, [n_k]\} \subset A \Big\} \\
&\le \max\left\{ \left[1 - \prod_1^{k-1} s_i^2 \right]^{1/2} \ \Bigg| \ s_i^2 = 1 - \left[c\left(H_{n_i}, \bigcap_{j=i+1}^{k} H_{n_j} \right) \right]^2, \right. \\
&\qquad\qquad \left. 0 < n_1 < n_2 < \cdots < n_k \le r \right\} \\
&= \max\left\{ \left[1 - \prod_1^{k-1} s_i^2 \right]^{1/2} \ \Bigg| \right. \\
&\qquad\qquad s_i^2 = 1 - \left[c\left(\operatorname{span}\{f_{n_i}\}, \operatorname{span}\{f_{n_{i+1}}, \ldots, f_{n_k}\} \right) \right]^2, \\
&\qquad\qquad \left. 0 < n_1 < n_2 < \cdots < n_k \le r \right\}.
\end{aligned}
$$

Letting $M_i = \operatorname{span}\{f_{n_i}\}$ and $N_i = \operatorname{span}\{f_{n_{i+1}}, \ldots, f_{n_k}\}$, we see that $M_i + N_i$ is finite-dimensional, hence closed. By Theorem 3.5, $c(M_i, N_i) < 1$. We deduce that $s_i^2 > 0$ for $i = 1, 2, \ldots, k-1$, and hence $c < 1$. One way of actually computing c, or at least getting an upper bound on it, is to estimate upper bounds on the numbers $c(M_i, N_i)$. And for this, it may be helpful to observe that

$$
c(M_i, N_i) = \begin{cases} 0 & \text{if } f_{n_i} \in N_i \\ \|P_{N_i}(f_{n_i})\| & \text{if } f_{n_i} \notin N_i. \end{cases}
$$

Isotone Regression.

Isotone regression is a problem often found in statistics. A noisė function has been added to an increasing data function so the result is no longer increasing. The object of isotone regression is to reconstruct the original data using the information that the signal was originally increasing. In the language of approximation theory, the question can be restated as follows. Find the best approximation to a point x in $l_2(N)$ from the set

$$
\begin{aligned}
C &= \{y \in l_2(N) \mid y \text{ is increasing } \} \\
&= \{y \in l_2(N) \mid y(1) \le y(2) \le \cdots \le y(N)\} \\
&= \bigcap_{i=1}^{N-1} \{y \in l_2(N) \mid y(i) \le y(i+1)\} \\
&= \bigcap_{i=1}^{N-1} \{y \in l_2(N) \mid \langle y, f_i \rangle \le 0\},
\end{aligned}
$$

where $f_i \in l_2(N)$ is defined by $f_i(j) = 1$ if $j = i$, -1 if $j = i+1$, and 0 otherwise.

The bound for c was

$$\max\left\{\left[1-\prod_1^{k-1} s_i^2\right]^{1/2r} \;\middle|\; s_i^2 = 1-\left[c\left(\operatorname{span}\{f_{n_i}\}, \operatorname{span}\{f_{n_{i+1}},\dots,f_{n_k}\}\right)\right]^2,\right.$$
$$\left. 0 < n_1 < n_2 < \cdots < n_k \le r\right\}$$

with $r = N-1$ in this case. The largest that $c(\operatorname{span}\{f_{n_i}\}, \operatorname{span}\{f_{n_{i+1}},\dots,f_{n_k}\})$ can be is 1/2, consequently $s_i^2 \ge 3/4$. From this we get the bound

$$c \le \left(1-\left(\frac{3}{4}\right)^{N-2}\right)^{1/(2(N-1))} =: \hat{c}.$$

When x_∞ is a constant multiple of $(1,1,\dots,1)$ convergence is slowest. Numerical tests have shown that in this case, Dykstra's algorithm will converge slightly faster than the bound $O(\hat{c}^n)$.

Convex Regression.

Convex regression is similar to isotone regression. The original data function is convex (i.e. $f(i+1) \le [f(i)+f(i+2)]/2$). A noise function distorts it into a nonconvex function, and convex regression estimates the original data using the fact that it was convex. In the language of approximation theory, the question can be restated as follows. Find the best approximation to x in $l_2(N)$ from the set

$$\begin{aligned} C &= \{y \in l_2(N) \mid y \text{ is convex }\} \\ &= \bigcap_{i=1}^{N-2} \{y \in l_2(N) \mid -y(i)+2y(i+1)-y(i+2) \le 0\} \\ &= \bigcap_{i=1}^{N-2} \{y \in l_2(N) \mid \langle y, f_i\rangle \le 0\}, \end{aligned}$$

where $f_i \in l_2(N)$ is defined by $f_i(j) = 2$ if $j = i+1$, -1 if $|i+1-j| = 1$, and 0 otherwise. Using a procedure similar to the isotone case, the bound

$$c \le \left(1-\left(\frac{19}{36}\right)^{N-3}\right)^{1/(2(N-2))}$$

can be obtained.

Quadratic Programming.

We can also apply Dykstra's algorithm to the problem

$$\min_{x \in C} g(x)$$

where $g(x) = \frac{1}{2}x^T Qx + bx$, $C = \bigcap_{i=1}^r \mathcal{G}_i$, $\mathcal{G}_i = \{x \in l_2(N) \mid \langle x, g_i \rangle \le d_i\}$, and Q is an $N \times N$ symmetric positive definite matrix.

Defining $(x,y) := \langle Qx, y \rangle$, $||x||_Q = (x,x)^{1/2}$, and $b_0 := -Q^{-1}b$, we note that $(\ ,\ )$ defines an inner product on $\mathbb{R}^N$ and

$$\begin{aligned} g(x) &= \frac{1}{2}(x,x) + \langle b, x \rangle \\ &= \frac{1}{2}(x,x) + \langle Q(Q^{-1}b), x \rangle \\ &= \frac{1}{2}[(x,x) - 2(b_0, x)] \\ &= \frac{1}{2}[(x,x) - 2(b_0,x) + (b_0,b_0)] - \frac{1}{2}||b_0||_Q^2 \\ &= \frac{1}{2}||x - b_0||_Q^2 - \frac{1}{2}||b_0||_Q^2. \end{aligned}$$

Thus the problem of minimizing $g(x)$ over $C = \bigcap_{i=1}^r \mathcal{G}_i$ is equivalent to minimizing the distance in the Q-norm from b_0 to C. Letting X denote the space $\mathbb{R}^N$ with the inner product $(x,y) := \langle Qx, y \rangle$, we see that X is a Hilbert space and x is the minimum of $g(x)$ over C if and only if $x = P_C(b_0)$. Since

$$\mathcal{G}_i = \{x \in X \mid (x, Q^{-1}g_i) \le d_i\}$$

is a closed half-space, the solution can be obtained via Dykstra's algorithm starting with $x_0 = b_0$.

Note that in the special case when $Q = I$ is the identity and $b = 0$, the above problem reduces to finding the minimum norm solution to a system of linear inequalities.

Linear Programming.

In [24], Mangasarian showed that the standard linear programming problem is equivalent to finding the best approximation to a certain vector from the polyhedral set which defines the feasible points. Since our results apply to the latter problem, they have applications to solving linear programming problems. This idea will be more fully explored elsewhere.

REFERENCES

1. N. Aronszajn, *Theory of reproducing kernels*, Trans. Amer. Math. Soc. **68** (1950), 337–404.
2. H.H. Bauschke and J.M. Borwein, *Dykstra's alternating projection algorithm for two sets*, J. Approx. Theory (to appear).
3. ______, *On the convergence of von Neumann's alternating projection algorithm for two sets*, Set-Valued Analysis **1** (1993), 185–212.
4. 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**, Lecture Notes in Statistics, Springer-Verlag, 1985, pp. 28–47.
5. L.M. Bregman, *The method of successive projection for finding a common point of convex sets*, Doklady **162** (1965), 688–692.
6. C.K. Chui, F. Deutsch, and J.D. Ward, *Constrained best approximation in Hilbert space*, Constructive Approx. **6** (1990), 35–64.

7. ———. *Constrained best approximation in Hilbert space, II*, J. Approx. Theory **71** (1992), 213–238.
8. F. Deutsch, *Some Applications of Functional Analysis to Approximation Theory*, doctoral thesis, Brown Univ., 1965.
9. ———. *Von Neumann's alternating method: the rate of convergence*, in **Approximation Theory IV** (C.K. Chui, L.L. Schumaker, and J.D. Ward, eds.), Academic Press, New York, 1983, pp. 427–434.
10. ———. *Rate of convergence of the method of alternating projections*, in **Parametric Optimization and Approximation** (B. Brosowski and F. Deutsch, eds.), ISNM 72, Birkhäuser Verlag, Basel, 1984, pp. 96–107.
11. ———, *The method of alternating orthogonal projections*, in **Approximation Theory, Spline Functions and Applications** (S.P. Singh, ed.), Kluwer Acad. Publ., Boston, 1992, pp. 105–121.
12. F.R. Deutsch and P.H. Maserick, *Applications of the Hahn-Banach theorem in approximation theory*, SIAM Review **9** (1967), 516–530.
13. R.L. Dykstra, *An algorithm for restricted least squares regression*, J. Amer. Statist. Assoc. **78** (1983), 837–842.
14. Ky Fan, *On systems of linear inequalities*, in **Linear Inequalities and Related Systems** (H.W. Kuhn and A.W. Tucker, eds.), Annals of Math. Studies, 38, Princeton University Press, Princeton, New Jersey, 1956, pp. 99–156.
15. C. Franchetti and W. Light, *On the von Neumann alternating algorithm in Hilbert space*, J. Math. Anal. Appl. **114**, 305–314.
16. K. Friedrichs, *On certain inequalities and characteristic value problems for analytic functions and for functions of two variables*, Trans. Amer. Math. Soc. **41** (1937), 321–364.
17. N. Gaffke and R. Mathar, *A cyclic projection algorithm via duality*, Metrika **36** (1989), 29–54.
18. I. Halperin, *The product of projection operators*, Acta Sci. Math. (Szeged) **23** (1962), 96–99.
19. S.P. Han, *A successive projection method*, Math. Programming **40** (1988), 1–14.
20. C. Hildreth, *A quadratic programming procedure*, Naval Research Logistics Quarterly **4** (1957), 79–85 [Erratum, ibid, p. 361].
21. A.N. Iusem and A.R. De Pierro, *On the convergence properties of Hildreth's quadratic programming algorithm*, Math. Programming **47** (1990), 37–51.
22. ———, *On the convergence of Han's method for convex programming with quadratic objective*, Math. Programming **52** (1991), 265–284.
23. S. Kayalar and H.L. Weinert, *Error bounds for the method of alternating projections*, Math. Control Signal Systems **1** (1988), 43–59.
24. O.L. Mangasarian, *Normal solutions of linear programs*, Math. Programming Study **22** (1984), 206–216.
25. J. von Neumann, **Functional Operators—Vol. II. The Geometry of Orthogonal Spaces.** (This reprint of mimeographed lecture notes first distributed in 1933.) Annals of Math. Studies #22 (1950), Princeton University Press.
26. 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.
27. D.C. Solmon, *Mathematical Aspects of Computed Tomography* (1976–77), (handwritten lecture notes).

Received: August 24, 1993
Revised: March 2, 1994