JOURNAL OF APPROXIMATION THEORY 79, 418–443 (1994)

Dykstra's Alternating Projection Algorithm for Two Sets

H. H. BAUSCHKE

Department of Mathematics, Statistics and Computing Science, Dalhousie University, Halifax, Nova Scotia, Canada, B3H 3J5

AND

J. M. BORWEIN

Department of Combinatorics and Optimization, University of Waterloo, Waterloo, Ontario, Canada, N2L 3G1

Communicated by Frank Deutsch

Received June 1, 1992; accepted in revised form April 28, 1993

We analyze Dykstra's algorithm for two arbitrary closed convex sets in a Hilbert space. Our technique also applies to von Neumann's algorithm. Various convergence results follow. An example allows one to compare qualitative and quantitative behaviour of the two algorithms. We discuss the case of finitely many sets.

1. INTRODUCTION

The problem of finding the projection of a given point in a Hilbert space onto the nonempty intersection of finitely many closed convex sets arises in many areas. Perhaps the earliest, but clearly one of the most successful solutions dates back to von Neumann [18]. He treated the case of two closed subspaces. Since then, many results have been found for this case. For a recent overview, see the paper of Deutsch [8].

Dykstra [9] suggested an algorithm which solves the problem for closed convex cones in a Euclidean space. Boyle and Dykstra [4] showed that Dykstra's algorithm, which coincides with von Neumann's algorithm for closed subspaces, solves the problem for general closed convex sets in a Hilbert space. Han [14] discovered Dykstra's algorithm in a Euclidean space via duality. The same approach led to a beautiful proof by Gaffke and Mathar [12]. All these authors discuss numerous applications.

0021-9045/94 $6.00

In this paper, we study Dykstra's algorithm for two arbitrary closed convex sets in a Hilbert space. Our main theorem generalizes recent results by Iusem and De Pierro [16] who assumed that the space is finite dimensional and the distance between the sets is attained. The paper is organized as follows. In Section 2, the geometry of two sets is briefly discussed. Quantities are introduced that are crucial for the understanding of Dykstra's algorithm for two sets.

A careful analysis of Dykstra's algorithm is given in Section 3. Our main result also applies to von Neumann's algorithm, as we demonstrate in Section 4. Specifically, we obtain an apparently new strong convergence result for two closed affine sets. In Section 5, we provide sufficient conditions for the main sequences of the two algorithms to be bounded. Examples give additional insight into the asymptotic behaviour and rates of convergence. Through reduction to two sets in a suitable product space, we can handle finitely many sets and also obtain a result for a weighted projections method. This is done in Section 6, where we conclude with some comments on the two algorithms for finitely many sets.

A more detailed analysis of some of these topics can be found in the associated technical report [1]. Throughout the paper, we frequently make use of the following.

If X is a (always real) Hilbert space, M a closed convex subset, and x a point in X, then the point in M nearest to x is called the *projection* of x onto M and denoted by $P_M x$ or $P_M(x)$. It is well-known that this point is characterized by *Kolmogorov's criterion*:

$$P_M x \in M \qquad \text{and} \qquad \langle M - P_M x, x - P_M x \rangle \leqslant 0.$$

An immediate consequence is the formula for the *projection onto a translate of* M: $P_{M+a}(x) = a + P_M(x-a)$ for any vectors a, x in X. Given two subsets A and B, their *distance* is defined by

$$d(A, B) := \inf \|A - B\| := \inf\{\|a-b\| : a \in A, b \in B\}.$$

The *fixed point set* of a mapping Q from X to X is $\operatorname{Fix} Q := \{x \in X : Qx = x\}$. With $\to$ (rsp. $\rightharpoonup$) we abbreviate *norm* (rsp. *weak*) *convergence* of sequences in X and $\overline{\lim}$ (rsp. $\underline{\lim}$) stands for *limit superior* (rsp. *limit inferior*) in $\mathbb{R}$. The closure of a set S in X is denoted by $\bar{S}$. Recall finally that $x_n \rightharpoonup x$ implies $\|x\| \leqslant \underline{\lim} \|x_n\|$ and that X has the *Kadec–Klee property*:

$$\left.\begin{array}{l} x_n \rightharpoonup x \\ \|x_n\| \to \|x\| \end{array}\right\} \text{implies } x_n \to x.$$

These two properties are well-known and easy to verify using the Cauchy–Schwarz inequality and the parellelogram law.

2. Geometrical Preliminaries

Let X be a Hilbert space, A, B, closed convex subsets, and $x \in X$. For abbreviation, let us define

$$\delta := d(A, B) \qquad \text{and} \qquad v := P_{\overline{B-A}}(0).$$

LEMMA 2.1. (i) *The vector v is the unique vector in $\overline{B-A}$ with minimal norm*: $\|v\| = \delta$.

(ii) *The constant δ is attained* (*i.e., the infimum* $\inf \|A - B\|$ *is attained*) *if and only if $v \in B - A$. In particularly, δ is attained whenever $B - A$ is closed.*

(iii) *Also,*

$$\|v\|^2 \leqslant \delta^2 \leqslant \langle v, B - A \rangle. \qquad (*)$$

In fact, $(*)$ *characterizes* $P_{\overline{B-A}}(0)$.

Proof. Only the last part needs a proof. If $v = P_{\overline{B-A}}(0)$, then $\|v\| = \delta$ and by Kolmogorov's criterion $\langle \overline{B-A} - v, 0 - v \rangle = \langle \overline{B-A}, -v \rangle + \delta^2 \leqslant 0$ and this verifies $(*)$. Conversely, let $v \in X$ be a vector satisfying $(*)$. We would like to show that $v = P_{\overline{B-A}}(0)$ and may assume $\delta > 0$. Pick a sequence $b_n - a_n \in B - A$ with $\|b_n - a_n\| \to \delta$. Then

$$\delta^2 \leqslant \langle v, b_n - a_n \rangle \leqslant \|v\| \cdot \|b_n - a_n\| \to \|v\| \cdot \delta,$$

thus $\|v\| = \delta$. Suppose v' is another vector satisfying $(*)$. So does $(v + v')/2$ and by what we just proved $\|v\| = \|v'\| = \|(v + v')/2\| = \delta$. The strict convexity of X implies $v = v'$; therefore, $P_{\overline{B-A}}(0) = v$ is the only vector which satisfies $(*)$. ∎

An interpretation of the vector v in the framework of convex programming can be found in [2].

Consider now the points in A (rsp. B) *nearest* to B (rsp. A):

$$E := \{a \in A : d(a, B) = \delta\}, \qquad F := \{b \in B : d(b, A) = \delta\}.$$

Three simple examples illustrate the definition:

(1) If $X := \mathbb{R}^2$, $A := \{(x, y) : x > 0,\ y \geqslant 1/x\}$, and $B := \{(x, 0) : x \in \mathbb{R}\}$, then $v = 0$, $\delta = 0$, $E = F = \varnothing$.

(2) If $X := \mathbb{R}^2$, $A := \{(x, 1) : x \in \mathbb{R}\}$, $B := \{(x, y) : x^2 + y^2 \leqslant 1\}$, then $v = 0$, $\delta = 0$, $E = F = A \cap B = \{(0, 1)\}$.

(3) If $X := \mathbb{R}^2$, $A := \{(x, -2) : x \in \mathbb{R}\}$, $B := \{(x, y) : |x|, |y| \leqslant 1\}$, then $v = (0, 1)$, $\delta = 1$, $E = \{(x, -2) : |x| \leqslant 1\}$, $F = E + v$.

We collect some basic facts about the *nearest point sets*, E and F. Part (i) in the next lemma was proved by Cheney and Goldstein [6].

LEMMA 2.2. (i) $E = \mathrm{Fix}(P_A P_B)$, $F = \mathrm{Fix}(P_B P_A)$.

(ii) *E, F are closed convex sets.*

(iii) *If δ is attained, then E, F are nonempty.*

(iv) *If E or F is nonempty, then δ is attained. Moreover,*

$$P_B e = e + v = P_F e \; (e \in E), \qquad P_A f = f - v = P_E f \; (f \in F), \qquad \text{so}$$

$$E + v = F, \qquad E = A \cap (B - v), \qquad F = (A + v) \cap B.$$

For the rest, suppose that δ is attained. Then

(v) $\langle A - E, v \rangle \leqslant 0$, $\langle B - F, -v \rangle \leqslant 0$, *and* $\langle E - E, v \rangle = \langle F - F, v \rangle = 0$.

(vi) $P_F x = P_E x + v$.

(vii) *If $A = \tilde{a} + M$, $B = \tilde{b} + N$ are closed affine subspaces (i.e., M, N are closed subspaces and $\tilde{a}, \tilde{b} \in X$), then E, F are also affine and*

$$v \in M^{\perp} \cap N^{\perp}.$$

Proof. (ii) follows from continuity and convexity of the distance functions $d(\cdot, A)$, $d(\cdot, B)$.

(iii) If δ is attained, then by Lemma 2.1(ii) $v \in B - A$, say $v = b - a$. It is easy to check that $P_B a = b$, $P_A b = a$. Thus $a \in E$ and $b \in F$.

(iv) Fix $e \in E$. Then $\|P_B e - e\| = \delta$ and $P_B e - e \in B - A$. By Lemma 2.1, $P_B e - e = v$. The rest of (iv) follows.

(v) Fix $e \in E$. Then $P_B e = e + v$ (by (iv)) and Kolmogorov's criterion yields $\langle B - P_B e, e - P_B e \rangle = \langle B - (e + v), -v \rangle \leqslant 0$. Since $E + v = F$ (by (iv)), we conclude $\langle B - F, -v \rangle \leqslant 0$. The inequality $\langle A - E, v \rangle \leqslant 0$ is proved analogously. Now $E \subseteq A$, $F \subseteq B$, so in particular $\langle E - E, v \rangle$, $\langle F - F, -v \rangle \leqslant 0$ implying $\langle E - E, v \rangle = \langle F - F, -v \rangle = 0$.

(vi) Let $e_0 := P_E x$. Then for any $e \in E$,

$$
\begin{aligned}
0 &\geqslant \langle e - e_0, x - e_0 \rangle && \text{(by Kolmogorov's criterion)} \\
&= \langle e - e_0, x - e_0 \rangle + \langle e - e_0, -v \rangle && \text{(by (v))} \\
&= \langle (e + v) - (e_0 + v), x - (e_0 + v) \rangle.
\end{aligned}
$$

Since $E + v = F$, we have $\langle F - (e_0 + v), \; x - (e_0 + v) \rangle \leqslant 0$, thus by Kolmogorov's criterion $P_F x = e_0 + v = P_E x + v$.

(vii) As observed in Section 1,

$$P_A(y) = \tilde{a} + P_M(y - \tilde{a}) \qquad \text{and} \qquad P_B(y) = \tilde{b} + P_N(y - \tilde{b})$$

for every vector y in X. Now if $e_1, e_2 \in E$ and $\lambda_1, \lambda_2 \in \mathbb{R}$ with $\lambda_1 + \lambda_2 = 1$, then by linearity of P_N and (iv),

$$\begin{aligned} P_B(\lambda_1 e_1 + \lambda_2 e_2) &- (\lambda_1 e_1 + \lambda_2 e_2) \\ &= \tilde{b} + P_N(\lambda_1 e_1 + \lambda_2 e_2 - \tilde{b}) - (\lambda_1 e_1 + \lambda_2 e_2) \\ &= \lambda_1(\tilde{b} + P_N(e_1 - \tilde{b}) - e_1) + \lambda_2(\tilde{b} + P_N(e_2 - \tilde{b}) - e_2) \\ &= \lambda_1(P_B e_1 - e_1) + \lambda_2(P_B e_2 - e_2) = v; \end{aligned}$$

thus, $\lambda_1 e_1 + \lambda_2 e_2$ is not only in A but also in E. Therefore E is affine, and so is $F = v + E$. Choose $\tilde{m} \in M$ s.t. $\tilde{a} + \tilde{m} \in E$. By (v),

$$\langle A - (\tilde{a} + \tilde{m}), v \rangle = \langle M, v \rangle \leqslant 0,$$

so $v \in M^{\perp}$ and similarly $v \in N^{\perp}$. ∎

LEMMA 2.3. *Suppose* $b_n - a_n \in B - A$ *is a sequence such that* $\|b_n - a_n\| \to \delta$. *Then*

$$b_n - a_n \to v$$

and every weak cluster point of (a_n) *(rsp.* (b_n)*) lies in* E *(rsp.* F*). Consequently, if* δ *is not attained, then* $\|a_n\|, \|b_n\| \to +\infty$.

Proof. Any weak cluster point of $(b_n - a_n)$ lies in $\overline{B - A}$ and has norm less or equal to δ. But the only point with this property is v (by Lemma 2.1(i)). Hence $b_n - a_n \rightharpoonup v$. Since X has the Kadec-Klee property, we obtain $b_n - a_n \to v$. Assume now that a^* is a weak cluster point of (a_n), say $a_{n'} \rightharpoonup a^*$ for some subsequence (n') of (n). Then $b_{n'} \rightharpoonup a^* + v \in B$ and hence $a^* \in E$. Similarly, every weak cluster point of (b_n) lies in F. Finally, if δ is not attained, then (by Lemma 2.2(iv)) $E = F = \varnothing$ and $(a_n), (b_n)$ cannot have weak cluster points. ∎

3. THE MAIN RESULT

For every integer $n \geqslant 1$, the terms of the sequences of Dykstra's algorithm (or more shortly *Dykstra sequences*) are defined as follows:

$$\begin{aligned} &p_0 := q_0 := 0, && b_0 := x, \\ &a_n := P_A(b_{n-1} + p_{n-1}), && p_n := b_{n-1} + p_{n-1} - a_n, \\ &b_n := P_B(a_n + q_{n-1}), && q_n := a_n + q_{n-1} - b_n. \end{aligned} \tag{DA}$$

We sometimes refer to the sequences (a_n), (b_n) (rsp. (p_n), (q_n)) as the *main sequences* (rsp. *auxiliary sequences*) *of Dykstra's algorithm.*

We should note that some authors prefer other auxiliary sequences and thus their formulae for the algorithm look different. Of course, the main sequences are the same and it is easy to make a "conversion table" for different formulations.

Let us collect the basic properties of the Dykstra sequences.

LEMMA 3.1. *The sequence $(a_n)_{n\geq 1}$ (rsp. $(b_n)_{n\geq 1}$)) lies in A (rsp. B). Also, for every integer $n \geq 1$:*

(i) $\langle p_n, a_n - A\rangle, \langle q_n, b_n - B\rangle \geq 0$,

(ii) $p_n = \sum_{k=1}^{n} (b_{k-1} - a_k)$, $q_n = \sum_{k=1}^{n} (a_k - b_k)$,

(iii) $x = a_n + p_n + q_{n-1} = b_n + p_n + q_n$,

(iv) $\|a_n - b_n\| \geq \|b_n - a_{n+1}\| \geq \|a_{n+1} - b_{n+1}\|$.

Proof. Part (i) follows directly from the definition (DA) of the Dykstra sequences and Kolmogorov's criterion. Part (ii) is an easy induction. Part (iii) is also proved inductively using (ii). We prove (iv):

$$\begin{aligned}
\|a_n - b_n\|^2 &= \|b_n - a_{n+1}\|^2 + \|a_{n+1} - a_n\|^2 + 2\langle b_n - a_{n+1}, a_{n+1} - a_n\rangle \\
&\geq \|b_n - a_{n+1}\|^2 + 2\langle p_{n+1} - p_n, a_{n+1} - a_n\rangle \\
&\geq \|b_n - a_{n+1}\|^2 \qquad \text{(by (i))} \\
&= \|a_{n+1} - b_{n+1}\|^2 + \|b_{n+1} - b_n\|^2 + 2\langle a_{n+1} - b_{n+1}, b_{n+1} - b_n\rangle \\
&\geq \|a_{n+1} - b_{n+1}\|^2 + 2\langle q_{n+1} - q_n, b_{n+1} - b_n\rangle \\
&\geq \|a_{n+1} - b_n\|^2. \qquad \text{(by (i))} \quad \blacksquare
\end{aligned}$$

The following formulae are crucial for our analysis. Similar formulae appear also in the articles by Boyle and Dykstra [4] and Iusem and De Pierro [16] and play an equally important role there.

LEMMA 3.2. *Suppose $a \in A$, $b \in B$, and $u := b - a$. Then*

(i) *for $1 \leq m \leq n$:*

$$\begin{aligned}
\|a_m - a\|^2 = \|a_n - a\|^2 &+ \sum_{k=m}^{n-1} \{\|a_k - b_k\|^2 + \|b_k - a_{k+1}\|^2 + 2\langle a_k - b_k, u\rangle\} \\
&+ 2\sum_{k=m}^{n-1} [\langle q_{k-1}, b_{k-1} - b_k\rangle + \langle p_k, a_k - a_{k+1}\rangle] \\
&+ 2[\langle q_{n-1}, b_{n-1} - b\rangle + \langle p_n, a_n - a\rangle \\
&- \langle q_{m-1}, b_{m-1} - b\rangle - \langle p_m, a_m - a\rangle].
\end{aligned}$$

(ii) *for* $0 \leqslant m \leqslant n$:

$$\begin{aligned}\|b_m - b\|^2 = \|b_n - b\|^2 + \sum_{k=m+1}^{n} \{ &\|b_{k-1} - a_k\|^2 \\ &+ \|a_k - b_k\|^2 + 2\langle b_{k-1} - a_k, -u\rangle \} \\ &+ 2 \sum_{k=m+1}^{n} [\langle p_{k-1}, a_{k-1} - a_k\rangle + \langle q_{k-1}, b_{k-1} - b_k\rangle] \\ &+ 2[\langle p_n, a_n - a\rangle + \langle q_n, b_n - b\rangle \\ &- \langle p_m, a_m - a\rangle - \langle q_m, b_m - b\rangle].\end{aligned}$$

Proof. (i)

$$\begin{aligned}\|a_m - a\|^2 &= \|a_m - b_m\|^2 + 2\langle a_m - b_m, b_m - a\rangle + \|b_m - a\|^2 \\ &= \|a_{m+1} - a\|^2 + \|a_m - b_m\|^2 + \|b_m - a_{m+1}\|^2 \\ &\quad + 2\langle a_m - b_m, b_m - a\rangle + 2\langle b_m - a_{m+1}, a_{m+1} - a\rangle \\ &= \|a_{m+1} - a\|^2 + \|a_m - b_m\|^2 + \|b_m - a_{m+1}\|^2 \\ &\quad + 2\langle a_m - b_m, b_m - b\rangle + 2\langle b_m - a_{m+1}, a_{m+1} - a\rangle \\ &\quad + 2\langle a_m - b_m, u\rangle \\ &= \|a_{m+1} - a\|^2 + \|a_m - b_m\|^2 + \|b_m - a_{m+1}\|^2 + 2\langle a_m - b_m, u\rangle \\ &\quad + 2\langle q_m - q_{m-1}, b_m - b\rangle + 2\langle p_{m+1} - p_m, a_{m+1} - a\rangle \\ &= \|a_{m+1} - a\|^2 + \{ \|a_m - b_m\|^2 + \|b_m - a_{m+1}\|^2 + 2\langle a_m - b_m, u\rangle \}_1 \\ &\quad + 2\{ \langle q_{m-1}, b_{m-1} - b_m\rangle + \langle p_m, a_m - a_{m+1}\rangle \}_2 \\ &\quad + 2\{ \langle q_m, b_m - b\rangle + \langle p_{m+1}, a_{m+1} - a\rangle - \langle q_{m-1}, b_{m-1} - b\rangle \\ &\quad - \langle p_m, a_m - a\rangle \}_3.\end{aligned}$$

Now substitute similarly for $\|a_{m+1} - a\|^2$ to get $\|a_m - a\|^2 = \|a_{m+2} - a\|^2 + \cdots$ and repeat until $\|a_m - a\|^2 = \|a_n - a\|^2 + \cdots$. One gets the desired formula by observing that the $\{\ \}_1$, $\{\ \}_2$-brackets add up to the two $\sum$-sums in the formula and that the $\{\ \}_3$-bracket telescopes. Part (ii) is proved in the same manner. ∎

The power of these formulae lies in our knowledge about the signs of the terms. Indeed, because of Lemma 3.1(i), we recognize the structure

$$\begin{aligned}\|a_m - a\|^2 = \|a_n - a\|^2 + \sum_{k=m+1}^{n-1} \{ &\|b_{k-1} - a_k\|^2 + \|a_k - b_k\|^2 \\ &+ 2\langle a_k - b_k, u\rangle \} + \alpha(m, n)\end{aligned} \tag{1}$$

and

$$\|b_m - b\|^2 = \|b_n - b\|^2 + \sum_{k=m+1}^{n-1} \{\|a_k - b_k\|^2 + \|b_k - a_{k+1}\|^2 + 2\langle b_k - a_{k+1}, -u\rangle\} + \beta(m, n), \tag{2}$$

where $\underline{\lim}_n \alpha(m, n)$, $\underline{\lim}_n \beta(m, n)$ are strictly bigger than $-\infty$. Therefore,

$$\overline{\lim_n} \sum_{k=1}^{n} \{\|b_{k-1} - a_k\|^2 + \|a_k - b_k\|^2 + 2\langle a_k - b_k, u\rangle\},$$

$$\overline{\lim_n} \sum_{k=1}^{n} \{\|a_k - b_k\|^2 + \|b_k - a_{k+1}\|^2 + 2\langle b_k - a_{k+1}, -u\rangle\}$$

are strictly less than $+\infty$. Under additional hypothesis, interesting consequences can be drawn. We extract the rather simple but very useful idea in the following.

OBSERVATION 3.3. *Suppose (x_n), (y_n) are sequences in X such that*

$$\inf\{\|x_n - y_n\| : n \in \mathbb{N}\}, \quad \inf\{\|y_n - x_{n+1}\| : n \in \mathbb{N}\} \geqslant \|u\|$$

for some vector u in X. Suppose further that the partial sums

$$\sum_{k=1}^{n} \{\|y_{k-1} - x_k\|^2 + \|x_k - y_k\|^2 + 2\langle x_k - y_k, u\rangle\},$$

$$\sum_{k=1}^{n} \{\|x_k - y_k\|^2 + \|y_k - x_{k+1}\|^2 + 2\langle y_k - x_{k+1}, -u\rangle\},$$

are bounded above. Because of

$$\begin{aligned} &\|y_{k-1} - x_k\|^2 + \|x_k - y_k\|^2 + 2\langle x_k - y_k, u\rangle \\ &\quad = (\|y_{k-1} - x_k\|^2 - \|u\|^2) + \|x_k - y_k + u\|^2 \end{aligned}$$

and

$$\begin{aligned} &\|x_k - y_k\|^2 + \|y_k - x_{k+1}\|^2 + 2\langle y_k - x_{k+1}, -u\rangle \\ &\quad = (\|x_k - y_k\|^2 - \|u\|^2) + \|x_{k+1} - y_k + u\|^2, \end{aligned}$$

the above partial sums consist entirely of nonnegative terms. Consequently, the corresponding series together with

$$\sum_k (\|y_{k-1} - x_k\|^2 - \|u\|^2), \qquad \sum_k \|y_k - x_k - u\|^2,$$

and

$$\sum_k (\|x_k - y_k\|^2 - \|u\|^2), \qquad \sum_k \|y_k - x_{k+1} - u\|^2$$

are convergent. In particular, $y_k - x_k$, $y_k - x_{k+1} \to u$.

We apply this observation to Dykstra's algorithm:

COROLLARY 3.4. *The Dykstra sequences* (a_n), (b_n) *satisfy*

$$b_n - a_n, \quad b_n - a_{n+1} \to v.$$

Moreover,

If $d(A, B)$ *is not attained, then* $\|a_n\|$, $\|b_n\| \to +\infty$.

If $d(A, B)$ *is attained, then* (a_n), (b_n) *are bounded and the series*

$$\sum_n (\|b_{n-1} - a_n\|^2 - \|v\|^2), \qquad \sum_n \|b_n - a_n - v\|^2,$$

$$\sum_n (\|a_n - b_n\|^2 - \|v\|^2), \qquad \sum_n \|b_n - a_{n+1} - v\|^2$$

are convergent.

Proof. *Case* 1. $d(A, B)$ is not attained.

Claim. $\inf_n \|a_n - b_n\| = \inf_n \|b_n - a_{n+1}\| = d(A, B)$.

By Lemma 3.1(iv), the infima coincide and are limits of the sequences $(\|a_n - b_n\|)$, $(\|b_n - a_{n+1}\|)$. Assume to the contrary that the common limit, say L, is strictly bigger than $d(A, B)$. Pick $u \in B - A$ s.t. $d(A, B) < \|u\| < L$. Observation 3.3 applies for this vector u and the sequences (a_n), (b_n) and results in $b_n - a_n \to u$. This contradicts the definition of L and verifies the claim. The claim, together with Lemma 2.3, yields all we have announced for this case.

Case 2. $d(A, B)$ is attained.

Again Observation 3.3 applies, this time for v and (a_n), (b_n). So once more $b_n - a_n$, $b_n - a_{n+1} \to v$ and the four series are convergent. Moreover, choosing $u = v$, we see that every term of the $\sum$-sums in (1) and (2) is nonnegative. Therefore the sequences (a_n), (b_n) are bounded. ∎

The following Lemma was also used and a stronger version of it proved by Gaffke and Mathar [12]. An elementary proof based solely on the Cauchy–Schwarz inequality can be found in [1].

LEMMA 3.5. *Suppose (ξ_n), (η_n) are sequences s.t. $\sum \xi_n^2$, $\sum \eta_n^2$ are finite. Then*

$$\underline{\lim}_n |\xi_n| \cdot \sum_{k=1}^{n} |\eta_k| = 0.$$

LEMMA 3.6. *Suppose $d(A, B)$ is attained. Let f be a vector in F, n be a positive integer, and $e := f - v$. Then*

$$\langle f - b_n, x - b_n \rangle = \langle f - b_n, q_n \rangle + \langle e - a_n, p_n \rangle + n \cdot \langle v + a_n - b_n, v \rangle + \langle v + a_n - b_n, p_n - nv \rangle.$$

On the right side, the first three terms are nonpositive and the last term satisfies

$$\underline{\lim}_n |\langle v + a_n - b_n, p_n - nv \rangle| = 0.$$

Proof. By Lemma 3.1(iii),

$$\begin{aligned}\langle f - b_n, x - b_n \rangle &= \langle f - b_n, q_n + p_n \rangle \\ &= \langle f - b_n, q_n \rangle + \langle e - a_n, p_n \rangle + \langle v + a_n - b_n, p_n \rangle.\end{aligned}$$

Now $\langle v + a_n - b_n, p_n \rangle = n \cdot \langle v + a_n - b_n, v \rangle + \langle v + a_n - b_n, p_n - nv \rangle$ and this gives the formula. The first term, $\langle f - b_n, q_n \rangle$, is clearly nonpositive by Lemma 3.1(i). The same is true for the second term, because $e \in A$ by Lemma 2.2(iv). Using Lemma 2.1(*), we obtain (with $\delta = d(A, B)$)

$$\langle v + a_n - b_n, v \rangle = \|v\|^2 + \langle a_n - b_n, v \rangle = \delta^2 + \langle a_n - b_n, v \rangle \leqslant \delta^2 + (-\delta^2) = 0$$

and the nonpositivity of the third term. It remains to investigate the last term. By Lemma 3.1(ii), $p_n - nv = \sum_{k=1}^{n} (b_{k-1} - a_k - v)$; thus, involving the Cauchy–Schwarz inequality,

$$|\langle v + a_n - b_n, p_n - nv \rangle| \leqslant \|v + a_n - b_n\| \cdot \sum_{k=1}^{n} \|b_{k-1} - a_k - v\|.$$

Viewing the latter term as a term of a sequence, we conclude that the limit inferior of this sequence is 0 (by Corollary 3.4 and Lemma 3.5). Therefore, the proof is complete. ▮

THEOREM 3.7. *Suppose $d(A, B)$ is attained. Then $a_n \to P_E x$, $b_n \to P_F x$.*

Proof. Let (n') be a subsequence of (n) s.t. $\lim_{n'} \langle v + a_{n'} - b_{n'}, p_{n'} - n'v \rangle = 0$, $\lim_{n'} \|b_{n'}\|$ exists, and $f^* :=$ weak-$\lim_{n'} b_{n'}$ exists. This is possible by Lemma 3.6, Corollary 3.4, and the Eberlein-Smulian Theorem (see, e.g., the book by Holmes [15]). The notation for f^* is justified: f^* lies in F by Corollary 3.4 and Lemma 2.3. Now fix an arbitrary vector f in F. Then

$$\begin{aligned}\langle f-f^*, x-f^*\rangle &= \|f^*\|^2 - \langle f, f^*\rangle - \langle f^*, x\rangle + \langle f, x\rangle \\ &\leqslant \lim_{n'} [\|b_{n'}\|^2 - \langle f, b_{n'}\rangle - \langle b_{n'}, x\rangle + \langle f, x\rangle] \\ &= \lim_{n'} \langle f - b_{n'}, x - b_{n'}\rangle \leqslant 0.\end{aligned}$$

We used the weak lower semicontinuity of $\|\cdot\|^2$ and the weak convergence of the sequence $(b_{n'})$ to establish the first inequality; the second inequality is implied by Lemma 3.6 and the choice of (n'). Since f was arbitrary, Kolmogorov's criterion yields

$$f^* = P_F x.$$

Now if we choose $f = f^*$, then the above chain of inequalities becomes one of equalities implying $\lim_{n'} \|b_{n'}\| = \|f^*\|$. By the Kadec–Klee property, $\lim_{n'} b_{n'} = f^*$. Let $e^* := f^* - v$. Having established the norm convergence of the sequence $(b_{n'})$, we conlude from Lemma 3.6 that

$$\lim_{n'} \langle f^* - b_{n'}, q_{n'}\rangle, \qquad \lim_{n'} \langle e^* - a_{n'}, p_{n'}\rangle = 0.$$

By Lemma 3.2(ii) (with $m = n'$, $a = e^*$, $b = f^*$, $u = v$), we get for $n \geqslant n'$

$$\begin{aligned}\|b_{n'} - f^*\|^2 = \|b_n - f^*\|^2 + \sum_{k=n'+1}^{n} &\{\|b_{k-1} - a_k\|^2 \\ &+ \|a_k - b_k\|^2 + 2\langle b_{k-1} - a_k, -v\rangle\} \\ &+ 2 \cdot \sum_{k=n'+1}^{n} [\langle p_{k-1}, a_{k-1} - a_k\rangle + \langle q_{k-1}, b_{k-1} - b_k\rangle] \\ &+ 2[\langle p_n, a_n - e^*\rangle + \langle q_n, b_n - f^*\rangle \\ &- \langle p_{n'}, a_{n'} - e^*\rangle - \langle q_{n'}, b_{n'} - f^*\rangle].\end{aligned}$$

Now

$$\begin{aligned}&\|b_{k-1} - a_k\|^2 + \|a_k - b_k\|^2 + 2\langle b_{k-1} - a_k, -v\rangle \\ &\qquad \geqslant \|a_k - b_k\|^2 - \|b_{k-1} - a_k\|^2 \geqslant \|v\|^2 - \|b_{k-1} - a_k\|^2,\end{aligned}$$

by the Cauchy–Schwarz inequality and definition of v. Combining with Lemma 3.1(i) gives the estimate

$$\begin{aligned}\|b_{n'} - f^*\|^2 \geqslant \|b_n - f^*\|^2 - &\sum_{k=n'+1}^{n} [\|b_{k-1} - a_k\|^2 - \|v\|^2] \\ &- 2[\langle p_{n'}, a_{n'} - e^*\rangle + \langle q_{n'}, b_{n'} - f^*\rangle],\end{aligned}$$

and hence

$$\|b_n - f^*\|^2 \leqslant \|b_{n'} - f^*\|^2 + \sum_{k=n'+1}^{\infty} [\|b_{k-1} - a_k\|^2 - \|v\|^2]$$
$$+ 2[\langle p_{n'}, a_{n'} - e^* \rangle + \langle q_{n'}, b_{n'} - f^* \rangle].$$

But we have seen that all terms on the right hand side become small when n' gets large (for the series, review Corollary 3.4) and therefore $b_n \to f^*$. Finally, again by Corollary 3.4, $a_n = b_n - (b_n - a_n) \to f^* - v = e^*$, and this completes the proof. ∎

We summarize in the following theorem.

THEOREM 3.8. *Let X be a Hilbert space, A, B closed convex subsets, and x a point in X. Define the terms of the Dykstra sequences for every integer $n \geqslant 1$ by*

$$\begin{aligned} &p_0 := q_0 := 0, && b_0 := x, \\ &a_n := P_A(b_{n-1} + p_{n-1}), && p_n := b_{n-1} + p_{n-1} - a_n, \\ &b_n := P_B(a_n + q_{n-1}), && q_n := a_n + q_{n-1} - b_n. \end{aligned} \tag{DA}$$

Then

$$b_n - a_n,\ b_n - a_{n+1} \to v,$$

where $v := P_{\overline{B-A}}(0)$ and $\|v\| = d(A, B)$. In particular, $\|b_n - a_n\|$, $\|b_n - a_{n+1}\| \to d(A, B)$ and a_n/n, $b_n/n \to 0$, $p_n/n \to v$, $q_n/n \to -v$. Furthermore,

If $d(A, B)$ is not attained, then $\|a_n\|$, $\|b_n\| \to +\infty$.

If $d(A, B)$ is attained, then $a_n \to P_E x$, $b_n \to P_F x$, where $E := \{a \in A: d(a, B) = d(A, B)\}$, $F := \{b \in B: d(b, A) = d(A, B)\}$ are nonempty closed convex sets with $E + v = F$.

Proof. We have only to verify the part dealing with the Dykstra sequences divided by n. But by Lemma 3.1(ii)

$$\frac{p_n}{n} = \frac{1}{n} \sum_{k=1}^{n} (b_{k-1} - a_k), \qquad \frac{q_n}{n} = \frac{1}{n} \sum_{k=1}^{n} (a_k - b_k);$$

so we recognize the sequences (p_n/n), (q_n/n) as arithmetic means of convergent sequences. By a well-known result of Cesaro, $p_n/n \to v$, $q_n/n \to -v$. The statement for (a_n/n), (b_n/n) follows now from Lemma 3.1(iii). ∎

Remarks 3.9. Theorem 3.8 implies the following:

- If $A \cap B \neq \varnothing$, then $d(A, B)$ is attained, $E = F = A \cap B$ and hence $a_n, b_n \to P_{A \cap B} x$. This is Theorem 2 for two sets in Boyle and Dykstra's article [4].
- If X is finite dimensional and $d(A, B)$ is attained, then $a_n \to P_E x$, $b_n \to P_F x$. This is Theorem 1 in the article by Iusem and De Pierro [16].
- If $d(A, B) > 0$, then the auxiliary sequences $(p_n), (q_n)$ are unbounded. This may happen even when $d(A, B) = 0$ as Han [14] demonstrated.

4. Some Remarks on von Neumann's Algorithm

von Neumann's algorithm (or method of alternating projections) is another, much less complicated attempt to find a point in the intersection of two closed convex sets, A, B in a Hilbert space X. For a starting point x in X, the terms of the sequences of von Neumann algorithm's are defined by

$$b_0 := x,$$

$$a_n := P_A(b_{n-1}), \qquad b_n := P_B(a_n),$$

for every integer $n \geqslant 1$.

As we mentioned in Section 1, the method is well-understood in the case where A, B are linear subspaces. Dykstra [9] pointed out that in this case his and von Neumann's method coincide; the same is true when A, B are affine (with possible empty intersection), as Gaffke and Mathar [12] remarked. In particular, Theorem 3.8 applies and yields (together with Lemma 2.2(vii)) the following theorem.

Theorem 4.1. *Let X be a Hilbert space, $A = \tilde{a} + M$, $B = \tilde{b} + N$ closed affine subspaces (here M, N are closed subspaces and $\tilde{a}, \tilde{b} \in X$), and x a point in X. Define for any integer $n \geqslant 0$ the terms of the von Neumann sequences by*

$$b_0 := x, \qquad a_{n+1} := P_A(b_n), \qquad \textit{and} \qquad b_{n+1} := P_B(a_{n+1}).$$

Then

$$b_n - a_n,\ b_n - a_{n+1} \to v,$$

where $v := P_{\overline{B-A}}(0)$ *and* $v \in M^{\perp} \cap N^{\perp}$. *In particular,* $\|b_n - a_n\|$, $\|b_n - a_{n+1}\| \to d(A, B)$ *and* a_n/n, $b_n/n \to 0$. *Moreover,*

If $d(A, B)$ *is not attained, then* $\|a_n\|, \|b_n\| \to \infty$.

If $d(A, B)$ *is attained, then*

$$a_n \to P_E x, \qquad b_n \to P_F x,$$

where $E := \{a \in A: d(a, B) = d(A, B)\}$, $F := \{b \in B: d(b, A) = d(A, B)\}$ *are nonempty closed affine subspaces with* $E + v = F$.

Remark 4.2. von Neumann's result is included as the case where A, B are closed subspaces (and hence $d(A, B) = 0$ and attained). Deutsch [7] observed that von Neumann's algorithm works for intersecting affine closed subspaces. His results can be used to obtain a different proof of the case in the last theorem, where $d(A, B)$ is attained (see [1]).

The question arises if the case "$d(A, B)$ is not attained" occurs. The (somewhat counter-intuitive) answer is yes.

EXAMPLE 4.3. Suppose X is a Hilbert space containing two closed subspaces M, N with non-closed sum, e.g., as constructed by Franchetti and Light [11]. Assume that $\overline{M + N^{\perp}}$ contains a unit vector u (otherwise we embed X, M, N into $X \times \mathbb{R}$ in the obvious way). Fix a vector $z \in \overline{M + N} \setminus (M + N)$ and a nonnegative real number δ. Define

$$A := M, \qquad B := N + \delta(z + u).$$

Then $d(A, B) = \delta$ and $d(A, B)$ is attained if and only if $\delta = 0$.

Conclusion. If $\delta > 0$, then the von Neumann sequences tend to infinity. If $\delta = 0$, then the von Neumann sequences converge "arbitrarily slowly," as demonstrated by Franchetti and Light [11].

This indicates that the closedness of $B - A$ plays an important role. We discuss this in more detail in the next section. We return to the discussion of the general von Neumann algorithm, i.e., A, B are arbitrary closed convex sets. The following properties of the von Neumann sequence are immediate by Kolmogorov's criterion.

LEMMA 4.4. *The sequence* $(a_n)_{n \geq 1}$ (*rsp.* $(b_n)_{n \geq 1}$) *lies in* A (*rsp.* B). *For any integer* $n \geq 1$:

(i) $\langle b_{n-1} - a_n, a_n - A \rangle, \langle a_n - b_n, b_n - B \rangle \geq 0$;

(ii) $\|a_n - b_n\| \geq \|b_n - a_{n+1}\| \geq \|a_{n+1} - b_{n+1}\|$.

The following Lemma is the counterpart to Lemma 3.2.

LEMMA 4.5. *Suppose $a \in A$, $b \in B$, and $u := b - a$. Then for $1 \leqslant m \leqslant n$:*

(i)
$$\|a_m - a\|^2 = \|a_n - a\|^2 + \sum_{k=m}^{n-1} \{\|a_k - b_k\|^2 + \|b_k - a_{k+1}\|^2 + 2\langle a_k - b_k, u\rangle\} + \sum_{k=m}^{n-1} [\langle a_k - b_k, b_k - b\rangle + \langle b_k - a_{k+1}, a_{k+1} - a\rangle];$$

(ii)
$$\|b_m - b\|^2 = \|b_n - b\|^2 + \sum_{k=m+1}^{n} \{\|b_{k-1} - a_k\|^2 + \|a_k - b_k\|^2 + 2\langle b_{k-1} - a_k, -u\rangle\} + 2\sum_{k=m+1}^{n} [\langle b_{k-1} - a_k, a_k - a\rangle + \langle a_k - b_k, b_k - b\rangle].$$

Proof. (i)

$$\begin{aligned}\|a_m - a\|^2 &= \|a_m - b_m\|^2 + \|b_m - a\|^2 + 2\langle a_m - b_m, b_m - a\rangle \\ &= \|a_m - b_m\|^2 + \|b_m - a_{m+1}\|^2 + \|a_{m+1} - a\|^2 \\ &\quad + 2\langle a_m - b_m, b_m - b\rangle + 2\langle a_m - b_m, u\rangle \\ &\quad + 2\langle b_m - a_{m+1}, a_{m+1} - a\rangle\end{aligned}$$

Repeat this for $\|a_{m+1} - a\|^2$ until one reaches $\|a_n - a\|^2$. Part (ii) is proved similarly. ∎

We observe that the formulae for the von Neumann sequences are of the same structure as the formulae (1), (2) in the last section for the Dykstra sequences. So Observation 3.3 applies and we can imitate the proof of Corollary 3.4 to obtain

COROLLARY 4.6. *The von Neumann sequences (a_n), (b_n) satisfy*

$$b_n - a_n, \quad b_n - a_{n+1} \to v.$$

Moreover: If $d(A, B)$ is not attained, then $\|a_n\|, \|b_n\| \to +\infty$. If $d(A, B)$ is attained, then (a_n), (b_n) are bounded and the series

$$\sum (\|b_{n-1} - a_n\|^2 - \|v\|^2), \qquad \sum \|b_n - a_n - v\|^2,$$
$$\sum (\|a_n - b_n\|^2 - \|v\|^2), \qquad \sum \|b_n - a_{n+1} - v\|^2$$

are convergent.

We will show that (a_n), (b_n) are weakly convergent when $d(A, B)$ is attained. To do so, we need a result on nonexpansive maps proved by Reich [20]. Since we need only a very special case of it, we give the proof. Recall that a map Q is called *nonexpansive* if it is Lipschitz continuous with Lipschitz constant less or equal to 1 on its domain. It is easy to check that the composition of nonexpansive maps is nonexpansive. Also, using Kolmogorov's criterion, one verifies that projections are nonexpansive.

LEMMA 4.7. *Suppose $Q: X \to X$ is nonexpansive,* $\operatorname{Fix} Q \neq \varnothing$, *and the sequence*

$$x_0 \in X, \qquad x_{n+1} = Qx_n \qquad (n \geqslant 0)$$

has no weak cluster points outside $\operatorname{Fix} Q$. *Then (x_n) is weakly convergent.*

Proof. Since $\operatorname{Fix} Q \neq \varnothing$, the sequence (x_n) is bounded. In fact, for any $c \in \operatorname{Fix} Q$:

$$\|x_{n+1} - c\| = \|Qx_n - Qc\| \leqslant \|x_n - c\|,$$

so $(\|x_n - c\|^2)_{n \in \mathbb{N}}$ as well as

$$(\|x_n\|^2 - 2\langle x_n, c\rangle)_{n \in \mathbb{N}} \tag{†}$$

are convergent. Choosing c_1, c_2 as weak cluster points of (x_n) and subtracting the corresponding sequences (†) shows that $(\langle x_n, c_1 - c_2\rangle)_{n \in \mathbb{N}}$ is convergent. But since c_1, c_2 are weak cluster points we conclude $\langle c_1, c_1 - c_2\rangle = \langle c_2, c_1 - c_2\rangle$ or $c_1 = c_2$. ∎

We now obtain the main result of this section.

THEOREM 4.8. *Let X be a Hilbert space, A, B closed convex subsets, and x a point in X. Define the terms of the von Neumann sequences by*

$$b_0 := x, \qquad a_{n+1} := P_A b_n, \qquad b_{n+1} := P_B a_{n+1}, \tag{vN}$$

for $n \geqslant 0$. Then $b_n - a_n$, $b_n - a_{n+1} \to v$, where $v = P_{\overline{B-A}}(0)$ and $\|v\| = d(A, B)$. In particular, $\|b_n - a_n\|$, $\|b_n - a_{n+1}\| \to d(A, B)$ and a_n/n, $b_n/n \to 0$. If $d(A, B)$ is not attained, then $\|a_n\|$, $\|b_n\| \to +\infty$. If $d(A, B)$ is attained, then

$$a_n \rightharpoonup e^* \in E, \qquad b_n \rightharpoonup f^* = e^* + v \in F,$$

for some $e^* \in E := \{a \in A: d(a, B) = d(A, B)\}$ and $F := \{b \in B: d(b, A) = d(A, B)\}$, where E, F are closed convex sets with $E + v = F$.

Proof. Only the last part needs a proof, the rest is easy. By Lemmata 2.3, 2.2(i), every weak cluster point of $(b_n) = ((P_B P_A)^n x)$ lies in $F = \mathrm{Fix}(P_B P_A)$. Hence Lemma 4.7 yields the weak convergence of (b_n) to some $f^* \in F$ and $a_n = b_n - (b_n - a_n) \rightharpoonup f^* - v \in E$. ∎

COROLLARY 4.9 (Theorem 4 in the Paper of Cheney and Goldstein [6]). *If $d(A, B)$ is attained and A or B is locally compact, then the von Neumann sequences (a_n), (b_n) are convergent.*

Remarks 4.10.

- It is not surprising that parts of Theorem 4.8 may be proved by a nonexpansive mapping approach. Bruck and Reich [5] introduced the class of strongly nonexpansive mappings, to which P_A, P_B belong and their theory applies.
- It is a long standing open question as to whether or not the convergence of the von Neumann sequences can actually be only weak. For some positive results on norm convergence, see [2] and the references therein.

5. FURTHER REMARKS ON DYKSTRA'S ALGORITHM

Let X be a Hilbert space and A, B closed convex subsets. By Lemma 2.1, the distance $d(A, B)$ between A and B is attained if and only if $P_{\overline{B-A}}(0) \in B - A$. It is important to know when $d(A, B)$ is attained, because if so then the sequences (a_n), (b_n) in Dykstra's (rsp. von Neumann's) algorithm are norm (rsp. weakly) convergent!

Clearly,

$$d(A, B) \text{ is attained whenever}$$

$$(1)\ A \cap B \neq \varnothing \qquad \text{or} \qquad (2)\ B - A \text{ is closed.}$$

For applications, it is a priori not always known whether a problem is feasible or not, i.e., whether, or not (1) holds. The sufficient condition (2) has the advantage that it does not require knowledge of $A \cap B$. We now give a list of sufficient conditions for (2).

FACTS 5.1. *Each of the following conditions is sufficient for $B - A$ to be closed and therefore for $d(A, B)$ to be attained:*

(i) *One set (i.e., A or B) is bounded.*

(ii) *Both sets are polyhedral.*

(iii) *Both sets are closed affine subspaces and one set has finite dimension or codimension.*

(iv) *One set is locally compact and the intersection of the recession cones of both sets is linear.*

Sketch of the Proof. (i) Since one set is bounded, it is weakly compact. Hence weak closedness of $B-A$ is proved readily, since for convex sets being weakly closed is the same as being closed.

(ii) First we must say what we mean by a polyhedral set. We call a set *polyhedral* if it is the intersection of a finite number of *halfspaces.* In turn a halfspace is a lower level set of some non-zero continuous linear functional. Clearly, polyhedra are closed. If X is finite-dimensional, then the difference of two polyhedra is again a polyhedron. For a proof, the reader is referred to Sect. 19 in the monograph by Rockafellar [21]. We claim that this remains true for infinite dimensional X. The difficulty in proving this stems from the fact that, unlike in finite dimension, a polyhedron is not a finitely generated set. A complete proof for the general case will appear in the thesis of the first author.

(iii) Let $A=\tilde{a}+M$, $B=\tilde{b}+N$, where M, N are closed subspaces and $\tilde{a}, \tilde{b} \in X$. Clearly, $B-A$ is closed $\Leftrightarrow M+N$ is closed. The result now follows by Sect. 9, Coro. 1, in the book by Holmes [15] and Corollary 35.6 in the book by Jameson [17].

(iv) This is a lemma in Sect. 15 of Holmes' book [15]. Recall the definition of the *recession cone* $O^+(S)$ for a set S in X: $O^+(S) := \{x \in X: x+S \subseteq S\}$. If S is closed and convex, then $O^+(S) := \{x \in S: x = (\text{weak-}) \lim_n t_n s_n,\ t_n \geqslant 0,\ t_n \to 0,\ s_n \in S\}$. ∎

EXAMPLE 5.2. Let $H := \ell_2(\mathbb{N})$ and L be the linear and continuous operator

$$L\colon H \to H\colon (x_n) \mapsto \left(\frac{x_n}{2^n}\right).$$

Define further $I = \{x=(x_n) \in H: |x_n| \leqslant 1 \text{ for all } n\}$, $y := (1/2^n)$ and note that $y \in \overline{L(I)} \setminus L(I)$, since

$$L(\underbrace{1, 1, \ldots, 1}_{n\text{-times}}, 0, 0, \ldots) \to y \qquad \text{and} \qquad (1) \notin H.$$

The set $L(I)$ is unbounded but *linearly bounded* (i.e., it contains no rays), and contained in I. Now let $v \in \mathbb{R}$ and define

$$X := H \times H \times \mathbb{R}, \qquad B_v := H \times \{0\} \times \{v\},$$
$$A := [\text{graph}\, ({}^{L}|_{I}) \times \{0\}] + (0, y, 0),$$

where graph $({}^{L}|_{I}) = \{(x, Lx) \in H \times H: x \in I\}$. But

$$\begin{aligned} B_v - A &= (0, 0, v) + \{[H \times \{0\} \times \{0\}] - [\text{graph}({}^{L}|_{I}) \times \{0\}] - (0, y, 0)\} \\ &= (0, 0, v) + \{[H \times L(I) \times \{0\}] - (0, y, 0)\}, \end{aligned}$$

and therefore

$$P_{\overline{B_v - A}}(0, 0, 0) = (0, 0, v), \qquad d(A, B_v) = |v|,$$

where $d(A, B_v)$ is not attained. Consequently, the Dykstra sequence (a_n) in A will tend to infinity (for every starting point). However, A is unbounded but linearly bounded, so $O^+(A) = \{0\}$. Thus if weak-lim $t_n a_n = d$, where $t_n \to 0$, $t_n \geqslant 0$, then $d \in O^+(A)$, so $d = 0$. In particular, $a_n/\|a_n\| \rightharpoonup 0$ and there is no chance of proving in general that $(a_n/\|a_n\|)$ converges even weakly to a non-zero direction.

Note that the sets involved in this example are not artificial: One is a closed affine subspace, the other the restriction of a graph. Also, the distance between both sets can be chosen arbitrarily.

EXAMPLE 5.3. Let X be $\mathbb{R}^2$, $R > 0$, and define

$$A := \{(x, y) \in X: y = R\}, \qquad B := \{(x, y) \in X: x^2 + y^2 \leqslant 1\}.$$

Let $\mathbf{x} := (a_0, R)$ be the starting point of either algorithm where $a_0 > 0$ and if $R < 1$, then $a_0 > (1 - R^2)^{1/2}$. Both algorithms produce sequences $(\mathbf{a}_n)$, $(\mathbf{b}_n)$, where the sequence $(\mathbf{a}_n)$ lies in A and thus its terms are of the form

$$\mathbf{a}_n := (a_n, R)$$

for $n \geqslant 0$. It is easy to see that the limit of the sequence $(\mathbf{a}_n)$ is the same for both algorithms. With some work one can be quite precise about the convergence rates; see [3] and [1] for more details. We summarize the behaviour of the sequence (a_n) in the following table:

	$\lim_n a_n$	von Neumann	Dykstra
$0 < R < 1$	$(1 - R^2)^{1/2}$	$\dfrac{a_{n+1} - (1 - R^2)^{1/2}}{a_n - (1 - R^2)^{1/2}} \to R^2$	$\dfrac{a_{n+1} - (1 - R^2)^{1/2}}{a_n - (1 - R^2)^{1/2}} \to 1 - \dfrac{(1 - R^2)^{3/2}}{a_0}$
$R = 1$	0	$\dfrac{a_n}{n^{-1/2}} \to 1$	$\dfrac{a_n}{\left(\dfrac{3n}{2a_0}\right)^{-1/3}} \to 1$
$1 < R$	0	$\dfrac{a_{n+1}}{a_n} \to R^{-1}$	$\dfrac{a_n}{\dfrac{a_0}{R - 1} n^{-1}} \to 1$

The example lays bare remarkable facts:

- The rate of convergence for the Dykstra sequence $(\mathbf{a}_n)$ depends, unlike the rate of convergence for the von Neumann sequence $(\mathbf{a}_n)$, on the starting point.
- In this example, the rate of convergence for the Dykstra sequence $(\mathbf{a}_n)$ is always worse than the rate of convergence for the von Neumann sequence $(\mathbf{a}_n)$. If $R < 1$, then although both sequences yield geometric convergence, the rate for the Dykstra sequence $(\mathbf{a}_n)$ can be arbitrarily close to 1 (by choosing a_0 large enough). If $R > 1$, then we even observe a qualitative difference: geometric convergence for the von Neumann sequence $(\mathbf{a}_n)$ versus the much slower "arithmetic" convergence for Dykstra's sequence $(\mathbf{a}_n)$.

Both algorithms have their characteristic pros and cons:

Dykstra's algororithm can be "slow" and probably does not allow an easy error-analysis (dependence on starting point in the above example). However, Dykstra's algorithm yields "nearest points via norm convergence," as Theorem 3.8 shows.

von Neumann algorithm's is very easy to compute and probably "faster" than Dykstra's algorithm. Unfortunately, only weak convergence results are established for the general case.

6. Finitely Many Sets

We apply our main results on the two algorithms in a product setting.

Throughout this section, let X be a Hilbert space, $C_1, C_2, ..., C_N (N \geqslant 2)$ closed convex subsets with corresponding projections $P_1, P_2, ..., P_N$ and x a point in X.

A clever product approach, due to Pierra [19] and developed by Flåm and Zowe [10] and Iusem and De Pierro [16], is as follows: Given strictly positive weights $\lambda_1, \lambda_2, ..., \lambda_N$, i.e., $\sum_{i=1}^{N} \lambda_i = 1$, define

$$\mathbf{X} := \prod_{i=1}^{N} (X, \lambda_i \langle \cdot, \cdot \rangle), \qquad \mathbf{A} := \{(x_1, ..., x_N) \in \mathbf{X} : x_1 = x_2 = \cdots = x_N \in X\},$$

$$\mathbf{B} := \{(x_1, ..., x_N) \in \mathbf{X} : x_i \in C_i \text{ for } i = 1 \cdots N\}.$$

For the distance of two points $\mathbf{y} = (y_1, ..., y_N)$, $\mathbf{z} = (z_1, ..., z_N) \in \mathbf{X}$ we have $\|\mathbf{y} - \mathbf{z}\|^2 = \sum_{i=1}^{N} \lambda_i \|y_i - z_i\|^2$, which implies

$$I := \inf_{y \in X} \sum_{i=1}^{N} \lambda_i d^2(y, C_i) = d^2(\mathbf{A}, \mathbf{B}).$$

Using Kolmogorov's criterion, we readily verify

$$P_A(y_1, ..., y_N) = \left(\sum_{i=1}^{N} \lambda_i y_i, \sum_{i=1}^{N} \lambda_i y_i, ..., \sum_{i=1}^{N} \lambda_i y_i \right),$$

$$P_B(y_1, y_2, ..., y_N) = (P_1 y_1, P_2 y_2, ..., P_N y_N).$$

Consequently,

$$\begin{aligned} \mathbf{E} &:= \{\mathbf{a} \in \mathbf{A}: d(\mathbf{a}, \mathbf{B}) = d(\mathbf{A}, \mathbf{B})\} = \operatorname{Fix}(P_{\mathbf{A}} P_{\mathbf{B}}) \qquad \text{(by Lemma 2.2(i))} \\ &= \left\{ (y, y, ..., y) \in \mathbf{X}: y = \sum_{i=1}^{N} \lambda_i P_i y \right\} \\ &= \left\{ (y, y, ..., y) \in \mathbf{X}: \sum_{i=1}^{N} \lambda_i d^2(y, C_i) = I \right\}. \end{aligned}$$

Define

$$C := \bigcap_{i=1}^{N} C_i,$$

$$\begin{aligned} E &:= \operatorname{Fix}\left(\sum_{i=1}^{N} \lambda_i P_i \right) = \left\{ y \in X: \sum_{i=1}^{N} \lambda_i d^2(y, C_i) = I \right\} \\ &= \operatorname{arginf} \sum_{i=1}^{N} \lambda_i d^2(\cdot, C_i), \end{aligned}$$

and observe that (i) E is closed and convex, (ii) E is nonempty if and only if I is attained, and (iii) if C is nonempty, then $E = C$.

THEOREM 6.1. *Define* $2N+1$ *sequences* (a_n), (b_n^i), (q_n^i), *by*

$$b_0^i := x, \qquad q_0^i := 0,$$

$$a_{n+1} := \sum_{j=1}^{N} \lambda_j b_n^j, \;\; b_{n+1}^i := P_i(a_{n+1} + q_n^i), \;\; q_{n+1}^i := a_{n+1} + q_n^i - b_{n+1}^i,$$

for $n \geq 0$ *and* $i = 1, 2, ..., N$. *Then* $\sum_{i=1}^{N} \lambda_i \|b_n^i - a_n\|^2$, $\sum_{i=1}^{N} \lambda_i \|b_n^i - a_{n+1}\|^2 \to I$ *and* a_n/n, b_n^1/n, b_n^2/n, ..., $b_n^N/n \to 0$. *Moreover,*

If I *is not attained, then* $\|a_n\|$, $\max\{\|b_n^1\|, \|b_n^2\|, ..., \|b_n^N\|\} \to \infty$.

If I *is attained, then*

$$a_n \to P_E x, \qquad b_n^i \to P_i P_E x,$$

for $i = 1, 2, ..., N$.

Proof. Apply Theorem 3.8 to $\mathbf{X}$, $\mathbf{A}$, $\mathbf{B}$, and $(x, x, ..., x)$.

Since $\mathbf{A}$ is an affine closed set, the auxiliary sequence $(\mathbf{p}_n)$ need not be taken into account, as Gaffke and Mathar [12] noted. ∎

Remarks 6.2.

- If C is nonempty, then the sequences (a_n), (b_n^i) converge to $P_C x$.
- The theorem generalizes Theorems 3 and 4 in the article by Iusem and De Pierro [16] who assumed that X is finite-dimensional and that I is attained. We should note that Iusem and De Pierro [16] used a different auxiliary sequence, so their algorithm looks different on the first sight.
- A geometric interpretation of the set

$$\mathbf{F} := \{\mathbf{b} \in \mathbf{B} \colon d(\mathbf{b}, \mathbf{A}) = d(\mathbf{B}, \mathbf{A})\} = \mathrm{Fix}(P_{\mathbf{B}} P_{\mathbf{A}})$$

can be found in [2].

The next theorem follows easily from Theorem 4.1.

THEOREM 6.3. *Suppose that $C_1, ..., C_N$ are closed affine subspaces and define the terms of the sequence (a_n) by*

$$a_1 := x, \qquad a_{n+1} := \sum_{i=1}^{N} \lambda_i P_i a_n,$$

for $n \geqslant 1$. Then $\sum_{i=1}^{N} \lambda_i \|a_n - P_i a_n\|^2$, $\sum_{i=1}^{N} \lambda_i \|P_i a_n - a_{n+1}\|^2 \to I$ and a_n/n, $P_1 a_n/n$, $P_2 a_n/n$, ..., $P_N a_n/n \to 0$. Moreover,

If I is not attained, then $\|a_n\|$, $\max\{\|P_1 a_n\|, \|P_2 a_n\|, ..., \|P_N a_n\|\} \to \infty$.

If I is attained, then

$$a_n \to P_E x, P_1 a_n \to P_1 P_E x, ..., P_N a_n \to P_N P_E x$$

and E is a nonempty closed affine subspace.

In our general setting we obtain at least the following theorem on weighted projections (by Theorem 4.8).

THEOREM 6.4. *Define the terms of the sequence (a_n) by*

$$a_1 := x, \qquad a_{n+1} := \sum_{i=1}^{N} \lambda_i P_i a_n,$$

for $n \geqslant 1$. Then $\sum_{i=1}^{N} \lambda_i \|a_n - P_i a_n\|^2$, $\sum_{i=1}^{N} \lambda_i \|P_i a_n - a_{n+1}\|^2 \to I$ for $n \geqslant 1$. Moreover,

If I is not attained, then $\|a_n\|$, $\max\{\|P_1 a_n\|, \|P_2 a_n\|, ..., \|P_N a_n\|\} \to \infty$.
If I is attained, then

$$a_n \rightharpoonup e^*, \qquad P_1 a_n \rightharpoonup P_1 e^*, ..., P_N a_n \rightharpoonup P_N e^*.$$

for some $e^* \in E$.

Remarks 6.5.

- *If X is finite-dimensional, then (a_n) converges in norm.*
- *If C is nonempty, then (a_n) converges weakly to some point in C.*
- *If X is finite-dimensional and C is nonempty, then (a_n) converges in norm to some point in C. This also follows by Theorem* 1 *in the paper by Flåm and Zowe* [10] *which studies a very general iteration scheme in Euclidean spaces.*

In view of the last theorems, it is important to know when E is nonempty. The next lemma generalizes Theorem 5 in the paper of Iusem and De Pierro [16].

LEMMA 6.6. *E is nonempty or equivalently I is attained whenever (at least) one of the following conditions hold:*

(i) *C is nonempty.*

(ii) *X is finite-dimensional and $C_1, C_2, ..., C_N$ are polyhedral.*

(iii) *One of the sets $C_1, ..., C_N$ is bounded.*

(iv) *There exist points $c_i^* \in C_i$ such that for $i, j \in \{1, 2, ..., N\}$:*

$$P_i c_j^* = c_i^*.$$

Proof. (i), (ii) E is nonempty if and only if $\mathbf{E}$ is. In turn $\mathbf{E}$ is nonempty when $\mathbf{B} - \mathbf{A}$ is closed; see Lemma 2.1(ii). The conclusion now follows from Facts 5.1.

(iii) In this case, the weakly lower semi-continuous function $\sum_{i=1}^N \lambda_i d^2(\cdot, C_i)$ has weakly compact lower level sets and therefore I is attained.

(iv) Fix $i \in \{1, 2, ..., N\}$. Then $c_j^* \in P_i^{-1}(c_i^*)$ for $j = 1, 2, ..., N$ but the last set is convex and thus contains $\sum_{j=1}^N \lambda_j c_j^* =: c^*$. Hence $P_i c^* = c^*$ and, since i was chosen arbitrarily, c^* is a fixed point of $\sum \lambda_j P_j$. Thus E is nonempty. ∎

Remark 6.7. Define the N sets

$$\begin{aligned} F_1 &:= \mathrm{Fix}(P_1 P_N P_{N-1} \cdots P_3 P_2), \\ F_2 &:= \mathrm{Fix}(P_2 P_1 P_N P_{N-1} \cdots P_4 P_3), \\ &\vdots \\ F_N &:= \mathrm{Fix}(P_N P_{N-1} \cdots P_2 P_1). \end{aligned}$$

As fixed point sets of nonexpansive maps in a strictly convex space, the sets $F_1, ..., F_N$ are closed and convex. Results of Bruck and Reich [5] allow one to prove that the sets $F_1, ..., F_N$ are actually translates of each other!

- If C is nonempty, then $F_1 = F_2 = \cdots = F_N = C$.
- If $N = 2$, then F_1 is the set of points in C_1 nearest to C_2 and vice versa for F_2 (by Lemma 2.2(i)).
- If $N \geqslant 3$, then unfortunately this intuitive geometric interpretation of the sets $F_1, ..., F_N$ is lost: take C_1 to be a singleton, say $C_1 = \{c_1\}$, and $C_2, C_3, ..., C_N$ as you like. Then

$$F_1 = \{c_1\}, \quad F_2 = \{P_2 c_1\}, \quad F_3 = \{P_3 P_2 c_1\}, ..., F_N = \{P_N P_{N-1} \cdots P_3 P_2 c_1\},$$

and therefore aspects of the "relative geometry" between the sets, e.g., whether or not $d(C_2, C_3)$ is attained, are irrelevant for the sets $F_1, ..., F_N$.

Concerning von Neumann's algorithm, we can prove the following theorem.

THEOREM 6.8. *Define the terms of the von Neumann sequence (x_n) by*

$$x_0 := x, \quad x_1 := P_1 x_0, \quad x_2 := P_2 x_1, ..., x_N := P_N x_{N-1}, \quad x_{N+1} := P_1 x_N, ...$$

Then either the sets $F_1, F_2, ..., F_N$ are empty and the sequence (x_n) tends to infinity or the sets $F_1, ..., F_N$ are nonempty and the N subsequences

$$(x_{kN+1})_{k \in \mathbb{N}}, \qquad (x_{kN+2})_{k \in \mathbb{N}}, ..., (x_{kN})_{k \in \mathbb{N}}$$

tend weakly to limits $f_1, f_2, ..., f_N$ where $f_i \in F_i$ for $i = 1, 2, ..., N$, and the weak limits are connected by

$$f_2 = P_2 f_1, \qquad f_3 = P_3 f_2, ..., f_N = P_N f_{N-1}, \qquad f_1 = P_1 f_N;$$

also

$$(x_{kN+2} - x_{kN+1})_{k \in \mathbb{N}}, \qquad (x_{kN+3} - x_{kN+2})_{k \in \mathbb{N}}, ..., (x_{kN+1} - x_{kN})_{k \in \mathbb{N}}$$

are norm convergent to vectors $v_1, ..., v_N$, where

$$F_2 = F_1 + v_1, \qquad F_3 = F_2 + v_2, ..., F_N = F_{N-1} + v_{N-1}, \qquad F_1 = F_N + v_N$$

and

$$v_1 + v_2 + \cdots + v_N = 0.$$

We omit the proof since the theorem is not in its final form. Related results were established by Gubin *et al.* [13].

In the case where $F_1, ..., F_N$ are nonempty, we also have that

$$P_{F_2}x = P_2 P_{F_1}x, ..., P_{F_N}x = P_N P_{F_{N-1}}x, \qquad P_{F_N}x = P_1 P_{F_N}x.$$

This supports the following conjecture on Dykstra's algorithm.

Conjecture. If (x_n) denotes the main sequence of Dykstra's algorithm with starting point x, then either $F_1, F_2, ..., F_N$ are empty and (x_n) tends to infinity or $F_1, ..., F_N$ are nonempty and the sequences

$$(x_{kN+1})_{k \in \mathbb{N}}, \qquad (x_{kN+2})_{k \in \mathbb{N}}, ..., (x_{kN})_{k \in \mathbb{N}}$$

tend *in norm* to

$$P_{F_1}x, P_{F_2}x, ..., P_{F_N}x.$$

Acknowledgments

The authors wish to thank David Borwein for discussion of Example 5.3, Judith Borwein for preparing the manuscript, and two anonymous referees for helpful suggestions.

References

1. H. H. Bauschke and J. M. Borwein, "Dykstra's Alternating Projection Algorithm for Two Sets," Technical Report, CORR 92-15, Faculty of Mathematics, University of Waterloo, 1992.
2. H. H. Bauschke and J. M. Borwein, "On the Convergence of von Neumann's Alternating Projection Algorithm for Two Sets," Set-Valued Analysis, Vol. 1, No. 2, pp. 185–212, 1993.
3. D. Borwein and J. M. Borwein, Problem 10355, in *American Mathematical Monthly*, Vol. 100, No. 8, Oct. 1993, p. 797.
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. Proceedings, Iowa City, 1985," Lecture Notes in Statistics, pp. 28–47, Springer-Verlag, Berlin/New York, 1985.
5. R. E. Bruck and S. Reich, Nonexpansive projections and resolvents of accretive operators in Banach spaces, *Houston J. Math.* **3** (1977), 459–470.
6. W. Cheney and A. A. Goldstein, Proximity maps for convex sets, *Proc. Amer. Math. Soc.* **10** (1959), 448–450.
7. F. Deutsch, Applications of von Neumann's alternating projections algorithm, *in* "Mathematical Methods in Operations Research" (P. Kenderov, Ed.), pp. 44–51, Sophia, Bulgaria, 1983.
8. F. Deutsch, The method of alternating orthogonal projections, *in* "Approximation Theory, Spline Functions and Applications" (S. P. Singh, Ed.), pp. 105–121, Kluwer, Dordrecht, 1991.
9. R. L. Dykstra, An algorithm for restricted least squares regression, *J. Amer. Statist. Assoc.* **78**, No. 384 (Dec. 1983), 837–842.
10. S. D. Flåm and J. Zowe, Relaxed outer projections, weighted averages and convex feasibility, *BIT* **30** (1990), 289–300.

11. C. Franchetti and W. Light, On the von Neumann alternating algorithm in Hilbert space, *J. Math. Anal. Appl.* **114** (1986), 305–314.
12. N. Gaffke and R. Mathar, A cyclic projection algorithm via duality, *Metrika* **36** (1989), 29–54.
13. L. G. Gubin, B. T. Polyak, and E. V. Raik, The method of projections for finding the common point of convex sets, *U.S.S.R. Comput. Math. and Math. Phys.* **7**, No. 6 (1967), 1–24.
14. S. P. Han, A successive projection method, *Math. Programming* **40** (1988), 1–14.
15. R. B. Holmes, "Geometric Functional Analysis and its Applications," Springer-Verlag, New York, 1975.
16. A. N. Iusem and A. R. De Pierro, On the convergence of Han's method for convex programming with quadratic objective, *Math. Programming* **52** (1991), 265–284.
17. G. J. O. Jameson, "Topology and Normed Spaces," Chapman and Hall, London, 1974.
18. J. von Neumann, "Functional Operators Vol. II," No. 22, Princeton Univ. Press, Princeton, 1950. [Reprint of mimeographed lecture notes first distributed in 1933.]
19. G. Pierra, Decomposition through formalization in a product space, *Math. Programming* **28** (1984), 96–115.
20. S. Reich, Weak convergence theorems for nonexpansive mappings in Banach spaces, *J. Math. Anal. Appl.* **67** (1979), 274–276.
21. R. T. Rockafellar, "Convex Analysis," Princeton Univ. Press, Princeton, 1970.