Available online at www.sciencedirect.com
ScienceDirect
Journal of Approximation Theory 142 (2006) 36 – 55
Journal of Approximation Theory
www.elsevier.com/locate/jat

The rate of convergence for the cyclic projections algorithm I: Angles between convex sets

Frank Deutsch[a,∗,1], Hein Hundal[b]

[a] *Department of Mathematics, Penn State University, University Park, PA 16802, USA*
[b] *146 Cedar Ridge Drive, Port Matilda, PA 16870, USA*

Received 6 June 2005; accepted 7 February 2006

Communicated by Joseph Ward
Available online 15 May 2006

Abstract

The cyclic projections algorithm is an important method for determining a point in the intersection of a finite number of closed convex sets in a Hilbert space. That is, for determining a solution to the "convex feasibility" problem. We study the rate of convergence for the cyclic projections algorithm. The notion of angle between convex sets is defined, which generalizes the angle between linear subspaces. The rate of convergence results are described in terms of these angles.

MSC: 41A65; 46C05; 46N10; 47H09

Keywords: Convex feasibility problem; Cyclic projections algorithm; Rate of convergence; Projections onto convex sets; POCS; Alternating projections; Orthogonal projections; Angle between convex sets; Angle between subspaces

1. Introduction

A frequent problem that arises in various areas of mathematics and physical sciences is to determine a point in the intersection of finitely many closed convex sets in a Hilbert space. This is called the *convex feasibility* problem. (See [5] for a nice review of this problem and of the various projection algorithms for solving this problem, and [10] for an in-depth exposition on the convex feasibility problem as it pertains to image recovery.) The cyclic projections algorithm (CPA) is

∗ Corresponding author. Fax: +1 814 865 3735.
E-mail addresses: deutsch@math.psu.edu (F. Deutsch), hundalhh@yahoo.com (H. Hundal).
1 Supported in part by NSF Grant 0204569.

0021-9045/$ - see front matter
doi:10.1016/j.jat.2006.02.005

arguably the most important and useful of all the algorithms for solving the convex feasibility problem (see, e.g., [9,12,5,6]).

In this paper we investigate the *rate of convergence* for the CPA. In the special case when all the convex sets are *linear subspaces*, this has already been considered in the papers [28,12,24,15]. For that situation, it was seen that (upper bounds on) the rate of convergence could be described in terms of *angles* between the various subspaces involved as well as the norms of various products of orthogonal projections. For general convex sets, this has motivated us to define a (localized) angle between *convex* sets, which generalizes the well-known notion of angle between linear subspaces. Then we show how to describe the rate of convergence of the CPA in terms of these angles.

In Section 2 we describe the problem and record some basic facts. In Section 3, we give a localized definition of the dual cone of a set—called the *ε-dual cone* (which reduces to the usual dual cone if the set happens to be a convex cone). Then we show that the error vector at the *n*th step of the CPA lies in the closure of a sum of certain ε-dual cones. This result provides the motivation for the definition of *ε-angle* between convex sets in Section 4. The ε-angle is a localized version of angle (which reduces to the usual—global—notion of angle when the sets are linear subspaces), and allows us to prove the main rate of convergence result (Theorem 4.6). When we specialize this convergence result to the case of *two linear subspaces*, we recover the well-known (and sharp!) rate-of-convergence result of Aronszajn [2] and Kayalar and Weinert [24] (Corollary 4.8).

In a sequel to this paper [16], we show that (the cosines of) these angles can often be expressed in terms of the norms of certain *nonlinear* operators: namely, the product of certain metric projections onto convex sets. Moreover, in the applications, it is important to know exactly when such cosines are strictly less than 1, and we will also give such conditions in [16].

We conclude this Introduction by noting that while the notion of *angle* appears explicitly in some areas of mathematics (see, e.g., [13] and the references cited therein), it often appears implicitly in many other areas. For example, it can be shown to appear (implicitly) in probability and statistics, tensor analysis and Grassman algebras, and in signal processing. To see this, we will provide a very simple example for these areas mentioned. Suppose $u = (u_1, u_2, \ldots, u_n)$ and $v = (v_1, v_2, \ldots, v_n)$ are vectors in $\mathbb{R}^n$ with norm one, $\langle u, v\rangle := \sum_1^n u_i v_i \geqslant 0$, and $\sum_1^n u_i = \sum_1^n v_i = 0$. Then the cosine of the angle between u and v is defined by $c := \langle u, v\rangle$. If two random variables U and V take on the values in $\{u_1, u_2, \ldots, u_n\}$ and $\{v_1, v_2, \ldots, v_n\}$ respectively with equal frequency, then the covariance and the correlation [18] of U and V are both c. The *covariance matrix* corresponding to the vectors u and v is given by

$$A = \begin{bmatrix} \langle u, u\rangle & \langle u, v\rangle \\ \langle v, u\rangle & \langle v, v\rangle \end{bmatrix}.$$

The largest (resp., smallest) eigenvalue of A is given by $1 + c$ (resp., $1 - c$) since $c \geqslant 0$. A principal component of the data set $\{(u_i, v_i) \mid i = 1, 2, \ldots, n\}$ is the span of the eigenvector of the covariance matrix A corresponding to the largest eigenvalue $(1 + c)$. The condition number $\|A\|\,\|A^{-1}\|$ of the covariance matrix A [21, p. 25] is given by $(1+c)/(1-c)$. For the tensor analysis example [7], if the operator $h : \mathbb{R}^n \times (\mathbb{R}^n)^* \to \mathbb{R}$ is defined by $h(x, y^*) = \langle u, x\rangle\, y^*(v)$, then h is in the tensor product space $T_1^1(\mathbb{R}^n)$. The contraction of h is defined by $\sum_1^n h(e_i, e_i^*)$, where the vectors $e_1, e_2, \ldots, e_n$ and $e_1^*, e_2^*, \ldots, e_n^*$ are the canonical bases in $\mathbb{R}^n$ and $(\mathbb{R}^n)^*$, respectively. Note that

$$\sum_1^n h(e_i, e_i^*) = \sum_1^n \langle u, e_i\rangle e_i^*(v) = \sum_1^n u_i v_i = c,$$

that is, the contraction of h is c. If $n = 3$, then the vector $u \wedge v$, defined by

$$u \wedge v = (u_2 v_3 u_3 v_2)(e_2 \wedge e_3) + (u_1 v_3 u_3 v_1)(e_1 \wedge e_3) + (u_1 v_2 u_2 v_1)(e_1 \wedge e_2),$$

is in the Grassman algebra $\bigwedge^2(\mathbb{R}^3)$ (see [7, pp. 92–95]), and is the familiar cross product. The cross product has norm equal to the sine of the angle between u and v. Lastly, in signal processing, the angle of arrival of an incoming signal across a receiver array is often found by maximizing the Friedrichs angle between the steering vector and the noise subspace [27].

2. The problem

Throughout the paper H will denote a real Hilbert space with inner product $\langle x, y\rangle$ and induced norm $\|x\| = \sqrt{\langle x, x\rangle}$. If K is a nonempty closed convex subset of H, the well-known result of Riesz [26] states that each $x \in H$ has a unique best approximation (or nearest point) $P_K(x)$ in K. That is,

$$\|x - P_K(x)\| < \|x - y\| \quad \text{for every } y \in K \backslash \{P_K(x)\}.$$

The mapping $P_K : H \to K$ thus defined is called the *metric projection* onto K.

If $D \subset H$, the (*negative*) *dual cone* of D is the set

$$D^\circ := \{x \in H \mid \langle x, d\rangle \leqslant 0 \text{ for every } d \in D\}.$$

The dual cone is a closed convex cone in H. Recall that a *convex cone* is a convex set C with the property that $\rho x \in C$ whenever $x \in C$ and $\rho \geqslant 0$. The *conical hull* of a set A, denoted cone A, is the intersection of all convex cones that contain A. The closure of cone A will be denoted by $\overline{\text{cone}}\, A$. The *interior* of the set A is denoted by int A. Throughout the paper, we use the term *subspace* to mean linear subspace. All other undefined terminology and notation is standard and can be found, for example, in [14].

We will repeatedly be using some facts that we list here for convenience.

Fact 2.1 (*Characterization of best approximations*). *Let K be a closed convex set in H, $x \in H$, and $x_0 \in K$. Then $x_0 = P_K(x)$ if and only if $x - x_0 \in (K - x_0)^\circ$. That is, if and only if*

$$\langle x - x_0, y - x_0\rangle \leqslant 0 \quad \textit{for every } y \in K.$$

Fact 2.2 (*Best approximations from translates*). *If D is a closed convex set and y is any point in H, then*

$$P_{D-y}(x) = P_D(x + y) - y \quad \textit{for every } x \in H.$$

Fact 2.3. *Let K be a nonempty closed convex subset of H and let $\emptyset \neq S \subset H$.*

(1) *If $0 \in K$, then $K \cap K^\circ = \{0\}$.*
(2) *$P_{K^\circ}(S) = \{0\}$ if and only if $S \subset \overline{\text{cone}}\,(K)$.*
(3) *If $0 \in K$, then $P_K(S) = \{0\}$ if and only if $S \subset K^\circ$.*

Fact 2.4 (*Metric projections are nonexpansive*). *If K is a nonempty closed convex subset of H, then P_K is nonexpansive; that is,*

$$\|P_K(x) - P_K(y)\| \leqslant \|x - y\| \quad \textit{for all } x, y \in H.$$

In particular, if $0 \in K$, *then*

$$\|P_K(x)\| \leqslant \|x\| \quad \text{for all } x \in H.$$

Fact 2.5 (*Commuting metric projections*). *Let M and N be closed* (*linear*) *subspaces in H. Then* P_M *and* P_N *commute if and only if* $P_M P_N = P_{M\cap N}$.

In particular, since $M \cap N \subset M$, P_M *and* $P_{M\cap N}$ *must commute. It follows that* P_M *and* $P_{(M\cap N)^\perp}$ *must also commute, and hence that* $P_M P_{(M\cap N)^\perp} = P_{M\cap(M\cap N)^\perp}$ *always holds.*

Fact 2.1 goes back at least to Aronszajn [2] (see also [14, Theorem 4.1, p. 43]), while Fact 2.3 follows easily from Fact 2.1. Fact 2.2 is well-known and easy to prove (see [14, Theorem 2.7, p. 25]). Fact 2.4 is from Phelps [25] (see also, [14, Theorem 5.5, p. 72]). Finally, Fact 2.5 is classical (see, e.g., [14, Theorem 9.2, p. 194]).

Now let us describe the cyclic projections algorithm. Let $C_1, C_2, \ldots, C_r$ be closed convex subsets of the Hilbert space H with $C := \bigcap_1^r C_i \neq \emptyset$. To determine a point in C, the *cyclic projections algorithm* (CPA) is an iterative scheme that can be described as follows. Start with any point $x \in H$, and define the sequence (x_n) by

$$x_0 = x \quad \text{and} \quad x_n = P_{C_{[n]}}(x_{n-1}) \quad (n = 1, 2, \ldots), \tag{2.1}$$

where $[\cdot] : \mathbb{N} \to \{1, 2, \ldots, r\}$ is the function “mod r” with values in $\{1, 2, \ldots, r\}$. That is,

$$[n] := \{1, 2, \ldots, r\} \cap \{n - kr \mid k = 0, 1, 2, \ldots\} \quad (n = 1, 2, \ldots).$$

In particular,

$$x_{nr} = (P_{C_r} P_{C_{r-1}} \cdots P_{C_1})^n(x) \quad (n = 0, 1, 2, \ldots). \tag{2.2}$$

The following result is well-known [9]. We include a brief proof for completeness.

Proposition 2.6. *The sequence* (x_n) *defined in* (2.1) *is Fejer monotone relative to C. That is, for each* $c \in C$ *and each* $n \geqslant 1$,

$$\|x_n - c\| \leqslant \|x_{n-1} - c\|. \tag{2.3}$$

Proof. Since the P_{C_i} are nonexpansive, it follows that for each $c \in C$ and $n \in \mathbb{N}$,

$$\|x_n - c\| = \|P_{C_{[n]}}(x_{n-1}) - c\| = \|P_{C_{[n]}}(x_{n-1}) - P_{C_{[n]}}(c)\| \leqslant \|x_{n-1} - c\|. \quad \square \tag{2.4}$$

The following result shows among other things that the sequence (x_n) generated by the CPA *always converges weakly to some point in C*.

Theorem 2.7. *The sequence* (x_n) *defined in* (2.1) *converges weakly to some point* $x_\infty \in C$:

$$x_n \to x_\infty \quad \text{weakly as } n \to \infty. \tag{2.5}$$

In particular,

$$x_{nr} = (P_{C_r} P_{C_{r-1}} \cdots P_{C_1})^n(x) \to x_\infty \quad \text{weakly as } n \to \infty.$$

Moreover,

$$\lim_{n\to\infty} \|P_C(x_n) - x_\infty\| = 0. \tag{2.6}$$

Remark. The first statement of Theorem 2.7 was proved by Bregman [9]. Baillon and Brezis [3] essentially showed that $(P_C(x_n))$ converged in norm to some element of C. Bauschke [4, Theorem 6.2.2(iii)] established that the norm limit of $(P_C(x_n))$ was in fact the weak limit of (x_n), i.e., he established Eq. (2.6).

We should mention that in the special case when all the convex sets are *affine* sets (i.e., translates of subspaces), then the CPA yields a stronger result than Theorem 2.7 (see also [14, p. 217]):

Theorem 2.8 (*von Neumann–Halperin*). *Let $A_1, A_2, \ldots, A_r$ be closed affine sets in H with $A := \bigcap_1^r A_i \neq \emptyset$. Then for each $x \in H$,*

$$\lim_{n\to\infty} \|(P_{A_r} P_{A_{r-1}} \cdots P_{A_1})^n(x) - P_A(x)\| = 0. \tag{2.7}$$

In other words, in this case $x_\infty = P_A(x)$ and the sequence (x_n) converges strongly to $P_A(x)$.

Actually, both von Neumann and Halperin stated their results for *subspaces* (von Neumann for $r = 2$ and Halperin for general r), but it is not hard to show that their results also hold more generally for affine sets.

This result shows that *in the special case where all the convex sets are affine sets, the CPA produces a sequence that actually converges strongly to the best approximation in the intersection of the initial point*! Also, in a finite-dimensional space, weak and norm convergence are equivalent, so that the CPA iterates always converge in norm. Up to just recently, it was not known whether the convergence in Bregman's theorem must only be weak. However, Hundal [22] has constructed an example of two convex cones in ℓ_2 for which the convergence of the iterates generated by the CPA is *not* strong.

In contrast to the von Neumann–Halperin theorem, it is easy to construct an example of two convex sets in the plane such that the sequence generated by the CPA converges in norm to a point in the intersection that is *not* the best approximation of the initial point. (If one is specifically interested in finding the *nearest point* in the intersection to any given point in the space, then there is a beautiful method, due to Dykstra [19] in a special case and to Boyle and Dykstra [8] (see also [14, Chapter 9]) in general, called *Dykstra's algorithm* that does this. But it is generally more complicated than the cyclic projections algorithm, and we shall not consider the Dykstra algorithm in this paper.)

3. ε-Dual Cones

To develop results concerning the rate of convergence of the sequence (x_n) (defined in Theorem 2.7) to x_∞, it turns out to be convenient to introduce the following variant of the notion of the dual cone of a convex set.

Definition 3.1. For any closed convex set $A \subset H$ and any $\varepsilon \geqslant 0$, the *ε-dual cone of A* is the set

$$A^{\circ,\varepsilon} := \operatorname{cone}\{x - P_A(x) \mid x \in B(0,\varepsilon)\}, \tag{3.1}$$

where $B(y,\varepsilon) := \{x \in H \mid \|x - y\| \leqslant \varepsilon\}$ is the closed ball centered at y with radius ε.

We collect a few simple but useful consequences of this definition in the following lemma.

Lemma 3.2. *Let A be a closed convex set in H.*

(1) $A^{\circ,\varepsilon}$ *is a convex cone, and* $A^{\circ,\varepsilon} \subset A^{\circ,\beta}$ *for each* $0 \leqslant \varepsilon < \beta$.

(2) *For each* $y \in H$,

$$(A-y)^{\circ,\varepsilon} = \text{cone}\,\{x - P_A(x) \mid x \in B(y,\varepsilon)\}. \tag{3.2}$$

(3) *If* $y \in \text{int}\, A$, *then*

$$(A-y)^{\circ,\varepsilon} = \{0\} = (A-y)^{\circ}$$

for each $\varepsilon > 0$ *sufficiently small.*

(4) *In general*, $A^{\circ,\varepsilon} \neq A^{\circ}$.

Proof. (1) is obvious, while (2) follows using the easily verified fact that $P_{A-y}(x) = P_A(x+y)-y$ (see, e.g., [14, Theorem 2.7(1)(ii), p. 25]). Obviously, (3) is a simple consequence of (2).

Finally, to verify (4), let $H = \mathbb{R}^2$ and take A to be a unit disk centered at the point (0, 1). Then, for any $0 < \varepsilon < \sqrt{2}$, $A^{\circ,\varepsilon}$ is a cone in the lower half of the plane whose "thickness" depends on ε, while A° is a ray—the lower half of the vertical axis. More precisely, $A^{\circ} = \text{cone}\,\{(0,-1)\}$ and $A^{\circ,\varepsilon} = \text{cone}\left\{(-\varepsilon, \frac{\varepsilon^2-2}{\sqrt{4-\varepsilon^2}}), (\varepsilon, \frac{\varepsilon^2-2}{\sqrt{4-\varepsilon^2}})\right\}$. $\square$

In spite of statement (4) in Lemma 3.2, the next result shows that ε-dual cones may always be expressed *entirely* in terms of ordinary dual cones!

Theorem 3.3. *Let A be a closed convex set*, $y \in A$, *and* $\varepsilon > 0$. *Then*

$$(A-y)^{\circ,\varepsilon} = \text{cone}\left\{ \bigcup_{x \in A \cap B(y,\varepsilon)} (A-x)^{\circ} \right\}. \tag{3.3}$$

In particular,

$$(A-y)^{\circ,\varepsilon} \supset (A-y)^{\circ}. \tag{3.4}$$

Proof. Assume first that $y = 0 \in A$. We must show that

$$A^{\circ,\varepsilon} = \text{cone}\left\{ \bigcup_{x \in A \cap B(0,\varepsilon)} (A-x)^{\circ} \right\}. \tag{3.5}$$

Let $D = \{x - P_A(x) \mid x \in B(0,\varepsilon)\}$. It follows by definition that cone $D = A^{\circ,\varepsilon}$. Let

$$E = \bigcup_{x \in A \cap B(0,\varepsilon)} (A-x)^{\circ}.$$

Thus to verify (3.5), we must show that cone D = cone E. To do this, it suffices to show that $D \subset E$, and for each $e \in E$ there exists $\alpha > 0$ such that $\alpha e \in D$.

To prove $D \subset E$, let $y \in D$. Then $y = x - P_A(x)$ for some $x \in B(0,\varepsilon)$. Thus

$$y = x - P_A(x) \in (A-z)^{\circ}, \quad \text{where } z = P_A(x). \tag{3.6}$$

Since $0 \in A$, Fact 2.4 implies that

$$\|z\| = \|P_A(x)\| = \|P_A(x) - P_A(0)\| \leqslant \|x - 0\| = \|x\| \leqslant \varepsilon.$$

Thus $z \in A \cap B(0, \varepsilon)$ and hence

$$y \in (A - z)^\circ \subset \bigcup_{z' \in A \cap B(0,\varepsilon)} (A - z')^\circ = E.$$

It remains to show that for each $e \in E$, there exists $\alpha > 0$ such that $\alpha e \in D$. Fix any $e \in E$. If $e = 0$, then $1 \cdot e = 0 \in D$, since $0 \in A$. Thus we may assume that $e \neq 0$. Then $e \in (A - x)^\circ$ for some $x \in A \cap B(0, \varepsilon)$. Let $\varepsilon' = \frac{1}{2}(\varepsilon - \|x\|)$ and $z = x + (\varepsilon'/\|e\|)e$. Then $\varepsilon' \geqslant 0$ and

$$\|z\| \leqslant \|x\| + \varepsilon' = \tfrac{1}{2}(\varepsilon + \|x\|) \leqslant \varepsilon$$

implies that $z \in B(0, \varepsilon)$ and $z - x \in (A - x)^\circ$. Since $x \in A$, Fact 2.1 implies that $x = P_A(z)$. Thus

$$\frac{\varepsilon'}{\|e\|} e = z - x = z - P_A(z) \in D.$$

Taking $\alpha = \varepsilon'/\|e\|$, we see that $\alpha e \in D$. Thus we have verified (3.5).

Now let $y \in A$ and $\varepsilon > 0$. Applying (3.5) to the set $A - y$ (which does contain 0), we obtain

$$(A - y)^{\circ,\varepsilon} = \operatorname{cone} \left\{ \bigcup_{x \in (A-y) \cap B(0,\varepsilon)} (A - y - x)^\circ \right\}.$$

But $x \in (A - y) \cap B(0, \varepsilon)$ if and only if $z := x + y \in A \cap B(y, \varepsilon)$. Thus

$$\bigcup_{x \in (A-y) \cap B(0,\varepsilon)} (A - y - x)^\circ = \bigcup_{z \in A \cap B(y,\varepsilon)} (A - z)^\circ,$$

and we obtain (3.3).

The inclusion (3.4) follows from (3.3) by noting that $y \in A \cap B(y, \varepsilon)$. □

In the case when C is a polyhedron and $y \in C$, the ε-dual cone of $C - y$ is the same as its ordinary dual cone whenever ε is sufficiently small. This is the content of our first corollary of Theorem 3.3.

Corollary 3.4 (ε-*dual cone of a polyhedron*). *Let h_i be a nonzero element of H, let α_i be a real scalar, and let*

$$H_i := \{x \in H \mid \langle x, h_i \rangle \leqslant \alpha_i\} \quad \textit{for } i = 1, 2, \ldots, r.$$

Then, for each y in the polyhedron $\mathcal{P} := \bigcap_1^r H_i$ and each $\varepsilon = \varepsilon(y) > 0$ sufficiently small,

$$(\mathcal{P} - y)^{\circ,\varepsilon} = \operatorname{cone}\{h_i \mid i \in I(y)\} = (\mathcal{P} - y)^\circ, \tag{3.7}$$

where $I(y) := \{i \mid \langle y, h_i \rangle = \alpha_i\}$ is the set of active indices *for y relative to $\mathcal{P}$.*

In particular, for $\varepsilon > 0$ sufficiently small,

$$(\mathcal{P} - y)^{\circ,\varepsilon} = \sum_1^r (H_i - y)^{\circ,\varepsilon} = \sum_1^r (H_i - y)^\circ = (\mathcal{P} - y)^\circ. \tag{3.8}$$

Proof. Let $I = \{1, 2, \ldots, r\}$. Without loss of generality, we may assume that $\|h_i\| = 1$ for each $i \in I$. By [14, Theorem 6.40, p. 115] we see that the second equality of Eq. (3.7) holds. Also, using relation (3.4), we have that for each $\varepsilon > 0$,

$$(\mathcal{P} - y)^{\circ,\varepsilon} \supset (\mathcal{P} - y)^{\circ}. \tag{3.9}$$

Thus to complete the proof of the first statement, it suffices to show that for $\varepsilon > 0$ sufficiently small,

$$(\mathcal{P} - y)^{\circ,\varepsilon} \subset \text{cone}\,\{h_i \mid i \in I(y)\}. \tag{3.10}$$

We first show that there exists an $\varepsilon > 0$ such that for every $x \in \mathcal{P} \cap B(y, \varepsilon)$, it follows that $I(x) \subset I(y)$. If $I(y) = I$, this is trivial: any $\varepsilon > 0$ works. Otherwise, for each $i \in I \backslash I(y)$, we have that $\langle y, h_i\rangle < \alpha_i$. Choose any ε such that

$$0 < \varepsilon < \min\{\alpha_i - \langle y, h_i\rangle \mid i \in I\backslash I(y)\}.$$

Then for each $x \in \mathcal{P} \cap B(y, \varepsilon)$ and each $i \in I\backslash I(y)$, we see that

$$\langle x, h_i\rangle = \langle x - y, h_i\rangle + \langle y, h_i\rangle \leqslant \varepsilon + \langle y, h_i\rangle < \alpha_i.$$

Thus $I(x) \subset I(y)$ and using Theorem 3.3 and [14, Theorem 6.40, p. 115] again, we obtain

$$\begin{aligned}(\mathcal{P} - y)^{\circ,\varepsilon} &= \text{cone}\left\{\bigcup_{x\in\mathcal{P}\cap B(y,\varepsilon)} (\mathcal{P} - x)^{\circ}\right\}\\ &= \text{cone}\left\{\bigcup_{x\in\mathcal{P}\cap B(y,\varepsilon)} \text{cone}\,\{h_i \mid i \in I(x)\}\right\}\\ &\subset \text{cone}\left\{\bigcup_{x\in\mathcal{P}\cap B(y,\varepsilon)} \text{cone}\,\{h_i \mid i \in I(y)\}\right\}\\ &= \text{cone}\ \{\text{cone}\,\{h_i \mid i \in I(y)\}\} = \text{cone}\,\{h_i \mid i \in I(y)\}.\end{aligned}$$

This completes the proof of the first statement of the theorem.

To verify the last statement of the theorem, use the first statement (with $r = 1$) and [14, Theorem 6.40, p. 115]. $\square$

There are also other strengthenings of Theorem 3.3 in the particular cases where A is a convex cone or an affine set. We state them next.

Corollary 3.5. *Let K be a closed convex cone, A a closed affine set, $y \in A$, and $\varepsilon > 0$. Then*

$$K^{\circ,\varepsilon} = K^{\circ} \tag{3.11}$$

and

$$(A - y)^{\circ,\varepsilon} = (A - y)^{\circ}. \tag{3.12}$$

In particular, if M is a closed subspace and $y \in M$, then

$$(M - y)^{\circ,\varepsilon} = (M - y)^{\circ} = M^{\perp}. \tag{3.13}$$

Proof. By Theorem 3.3, we have that

$$\begin{aligned} K^{\circ,\varepsilon} &= \operatorname{cone}\left\{ \bigcup_{x\in K\cap B(0,\varepsilon)} (K - x)^{\circ} \right\} \\ &= \operatorname{cone}\left\{ \bigcup_{x\in K\cap B(0,\varepsilon)} (K^{\circ} \cap x^{\perp}) \right\} \quad \text{using [14, Theorem 4.5(5), p. 46]} \\ &\subset \operatorname{cone} K^{\circ} = K^{\circ} \quad \text{since } K^{\circ} \text{ is a cone.} \end{aligned}$$

But since $0 \in K$, it follows from (3.4) that $K^{\circ,\varepsilon} \supset K^{\circ}$. Thus (3.11) holds.

Since $y \in A$, we have that $M := A - y$ is a subspace, hence a convex cone, so that by the first part,

$$(A - y)^{\circ,\varepsilon} = M^{\circ,\varepsilon} = M^{\circ} = (A - y)^{\circ},$$

which proves (3.12). □

Unlike the polyhedron case in Corollary 3.4 and the affine set case in Corollary 3.5, when K is a closed convex cone it is *not* true in general that $(K - y)^{\circ,\varepsilon} = (K - y)^{\circ}$ for each $y \in K$. However, there is something positive we can say about the translated cone case.

Lemma 3.6. *Let K be a closed convex cone with $y \in K$ and $\varepsilon > 0$.*

(1) *We have*

$$(K - y)^{\circ,\varepsilon} = \operatorname{cone}\{P_{K^{\circ}}(z) \mid z \in B(y, \varepsilon)\}. \tag{3.14}$$

(2) *If $y \in \operatorname{int} K$, then*

$$(K - y)^{\circ,\varepsilon} = (K - y)^{\circ} = \{0\} \tag{3.15}$$

holds for $\varepsilon > 0$ sufficiently small.

(3) *In general, however,*

$$(K - y)^{\circ,\varepsilon} \neq (K - y)^{\circ}. \tag{3.16}$$

Proof. (1) Using Definition 3.1 and the fact that for cones K, $x - P_K(x) = P_{K^{\circ}}(x)$ for each $x \in H$ (see [14, p. 74]), we obtain (3.14).

(2) This is a special case of Lemma 3.2(3).

(3) This is a consequence of the example given next. □

Example 3.7. Let $H = \mathbb{R}^3$, $K = \{x = (\alpha, \beta, \gamma) \mid \alpha \geqslant 0,\ \beta^2 + \gamma^2 \leqslant \alpha^2\}$, and $y = (1, 1, 0)$. Then K is a closed convex cone, $y \in K$, $K^{\circ} = -K$,

$$(K - y)^{\circ} = \operatorname{cone}\{(-1, 1, 0)\},$$

and for each $\varepsilon > 0$, $(K - y)^{\circ,\varepsilon}$ properly contains $(K - y)^{\circ}$.

We outline the proof of this example. The proof that $y \in K$ is obvious. To see that K is a closed convex cone can be done by verifying that $C := \{(1, \beta, \gamma) \mid \beta^2 + \gamma^2 \leqslant 1\}$ is a convex set and $K = \operatorname{cone} C = \{\rho c \mid \rho \geqslant 0, c \in C\}$. It follows that $K^\circ = (\operatorname{cone} C)^\circ = C^\circ$. So to verify that $K^\circ = -K$, it suffices to verify that $C^\circ = -K$. To this end, let $x \in -K$. Then $x = -\rho(1, x_2, x_3)$, where $\rho \geqslant 0$ and $x_2^2 + x_3^2 \leqslant 1$. For each $c \in C$, we have that $c = (1, c_2, c_3)$, where $c_2^2 + c_3^2 \leqslant 1$. Using Schwarz's inequality, we obtain

$$\begin{aligned}\langle c, x\rangle &= \rho(-1 - c_2x_2 - c_3x_3) = \rho[-2 - \langle(1, c_2, c_3), (1, x_2, x_3)\rangle] \\ &\leqslant \rho(-2 + \|c\|\,\|(1, x_2, x_3)\|) \leqslant \rho(-2 + \sqrt{2}\sqrt{2}) = 0.\end{aligned}$$

Hence $x \in C^\circ$ and thus $-K \subset C^\circ$. Conversely, suppose $x \in C^\circ$. Then for each $c \in C$, we have

$$0 \geqslant \langle x, c\rangle = \langle(x_1, x_2, x_3), (1, c_2, c_3)\rangle = x_1 + x_2c_2 + x_3c_3, \tag{3.17}$$

In particular, choosing $c = (1, 0, 0)$ in (3.17), we deduce that $x_1 \leqslant 0$. Similarly, choosing $c = (1, x_2(x_2^2+x_3^2)^{-1/2}, x_3(x_2^2+x_3^2)^{-1/2})$, we deduce that $\sqrt{x_2^2 + x_3^2} \leqslant -x_1 = |x_1|$ and so $x_2^2+x_3^2 \leqslant x_1^2$. Thus $x \in -K$ and hence $C^\circ \subset -K$. This proves that $C^\circ = -K$; that is, $K^\circ = -K$.

Next note that $(K - y)^\circ = K^\circ \cap y^\perp$ (see [14, Theorem 4.5(5), p. 46]). Moreover, it is easy to check that $y^\perp = \operatorname{span}\{(-1, 1, 0), (0, 0, 1)\}$ and hence, since $K^\circ = -K$, that $(K - y)^\circ = \operatorname{cone}\{(-1, 1, 0)\}$.

By Theorem 3.3, we see that $(K - y)^{\circ,\varepsilon} \supset (K - y)^\circ$ for each $\varepsilon > 0$. Now let $\varepsilon > 0$ and $z = (1, 1, \varepsilon)$ to complete the proof, it suffices to show that the element $z_0 := P_{K^\circ}(z) \in (K - y)^{\circ,\varepsilon} \backslash (K - y)^\circ$. By Lemma 3.6(1), we see that $z_0 \in (K - y)^{\circ,\varepsilon}$. By way of contradiction, suppose that $z_0 \in (K - y)^\circ$. Then by the above, $z_0 = \rho(-1, 1, 0)$ for some $\rho \geqslant 0$. There are two possibilities: $\rho = 0$ or $\rho > 0$. In the former case, $z_0 = 0$ and by using Fact 2.1, we deduce that

$$(1, 1, \varepsilon) = z = z - z_0 \in (K^\circ - z_0)^\circ = K^{\circ\circ} = K.$$

Thus we must have that $1^2 + \varepsilon^2 \leqslant 1^2$, which is absurd. In the case when $\rho > 0$, we again use Fact 2.1 to deduce

$$z - z_0 \in (K^\circ - z_0)^\circ = K^{\circ\circ} \cap z_0^\perp = K \cap z_0^\perp.$$

Hence

$$0 = \langle z - z_0, z_0\rangle = \langle(1 + \rho, 1 - \rho, \varepsilon), (-\rho, \rho, 0)\rangle = -2\rho^2 \neq 0,$$

which is absurd. This verifies the example.

The basis for all the rate of convergence results for the CPA that are given in this paper can be stated as follows.

Theorem 3.8. *Let $C_1, C_2, \ldots, C_r$ be closed convex sets with a nonempty intersection, let $x \in H$, let the sequence (x_n) be defined as in* (2.1), *and let x_∞ be its weak limit* (*see* (2.5)). *Let $\varepsilon > 0$, let m be a nonnegative integer, and suppose that $\|x_m - x_\infty\| \leqslant \varepsilon$. Then for every integer $n \geqslant m$,*

$$x_n - x_\infty \in \overline{\sum_1^r (C_i - x_\infty)^{\circ,\varepsilon}}. \tag{3.18}$$

In particular, for any $\varepsilon > 0$ with $\varepsilon \geqslant \|x - x_\infty\|$, relation (3.18) *holds for all $n \geqslant 0$.*

Proof. By Proposition 2.6, we have $\|x_n - x_\infty\| \leqslant \varepsilon$ for each $n \geqslant m$. Fix any $k > n + r$. Then, using Fact 2.2, we obtain

$$\begin{aligned}
x_k &= \sum_{i=n+1}^{k} (x_i - x_{i-1}) + x_n = \sum_{i=n+1}^{k} \left[P_{[i]}(x_{i-1}) - x_{i-1}\right] + x_n \\
&= - \sum_{i=n+1}^{k} (x_{i-1} - P_{[i]}(x_{i-1})) + x_n \\
&= - \sum_{i=n+1}^{k} [(x_{i-1} - x_\infty) - (P_{[i]}(x_{i-1}) - x_\infty)] + x_n \\
&= - \sum_{i=n+1}^{k} [(x_{i-1} - x_\infty) - P_{C_{[i]}-x_\infty}(x_{i-1} - x_\infty)] + x_n \\
&\in - \sum_{i=n+1}^{k} (C_{[i]} - x_\infty)^{\circ,\varepsilon} + x_n.
\end{aligned}$$

Since $k - (n + 1) \geqslant r$ and the sets $(C_{[i]} - x_\infty)^{\circ,\varepsilon}$ are all convex cones, it follows that

$$\sum_{i=n+1}^{k} (C_{[i]} - x_\infty)^{\circ,\varepsilon} = \sum_{j=1}^{r} (C_j - x_\infty)^{\circ,\varepsilon}.$$

Thus we obtain that

$$x_k \in - \sum_{j=1}^{r} (C_j - x_\infty)^{\circ,\varepsilon} + x_n,$$

or

$$x_n - x_k \in \sum_{j=1}^{r} (C_j - x_\infty)^{\circ,\varepsilon}. \tag{3.19}$$

Since $x_k \to x_\infty$ weakly and since the weak and norm closures of a convex set are the same, we let $k \to \infty$ in (3.19) to obtain the result. $\square$

4. ε-angles

We will now define the notion of *ε-angle* between two convex sets. This concept will turn out to be very useful in further developing and refining an error analysis for the cyclic projections algorithm. First recall that the *angle* between two closed *subspaces* M and N is the angle in the interval $[0, \pi/2]$ whose cosine is given by

$$c(M, N) := \sup\{\langle x, y\rangle \mid x \in M \cap (M \cap N)^\perp \cap B_H,\ y \in N \cap (M \cap N)^\perp \cap B_H\}, \tag{4.1}$$

where $B_H := B(0, 1)$ is the closed unit ball in H. (This angle was first defined and used by Friedrichs [20]. See [13] for an exposition on the angle between subspaces.)

It is not hard to verify that (see [12])

$$c(M, N) = \|P_{N\cap(M\cap N)^\perp} P_{M\cap(M\cap N)^\perp}\|. \tag{4.2}$$

It follows that (assuming $M \neq \{0\}$)

$$c(M, N) \tag{4.3}$$
$$= \sup\left\{ \frac{\|P_{N\cap(M\cap N)^\perp} P_{M\cap(M\cap N)^\perp} x\|}{\|x\|} \,\middle|\, \|x\| = 1 \right\}$$
$$= \sup\left\{ \frac{\|P_{N\cap(M\cap N)^\perp} P_{M\cap(M\cap N)^\perp} x\|}{\|x\|} \,\middle|\, \|x\| = \varepsilon \right\}$$
$$= \sup\left\{ \frac{\|P_{N\cap(M\cap N)^\perp} P_{M\cap(M\cap N)^\perp} x\|}{\|x\|} \,\middle|\, x \in M \cap (M\cap N)^\perp,\ \|x\| = \varepsilon \right\}$$
$$= \sup\left\{ \frac{\|P_{N\cap\overline{(M^{\circ,\varepsilon}+N^{\circ,\varepsilon})}} P_{M\cap\overline{(M^{\circ,\varepsilon}+N^{\circ,\varepsilon})}} x\|}{\|x\|} \,\middle|\, x \in M \cap \overline{(M^{\circ,\varepsilon}+N^{\circ,\varepsilon})},\ \|x\| = \varepsilon \right\} \tag{4.4}$$

for any $\varepsilon > 0$, since for subspaces

$$(M\cap N)^\perp = \overline{M^\perp + N^\perp} = \overline{M^\circ + N^\circ} = \overline{M^{\circ,\varepsilon} + N^{\circ,\varepsilon}} \tag{4.5}$$

by Corollary 3.5.

It turns out that it is this latter way of expressing the cosine (i.e., Eq. (4.4)) that suggests a very useful generalization to arbitrary closed convex sets. Moreover, in contrast to the subspace case, simple examples in the Euclidean plane show that any reasonable definition of the angle between two arbitrary intersecting convex sets will likely have to depend on a particular point in the intersection, as well as a prescribed length of vectors that one restricts his attention to.

Specifically, consider the following definition of *ε-angle* between two convex sets.

Definition 4.1. Let D_1, D_2 be closed convex sets with $0 \in D_1 \cap D_2$, $\varepsilon \geqslant 0$, and $i \in \{1, 2\}$. The *i*th *ε-angle* between the ordered pair D_1 and D_2 is the angle in the interval $[0, \pi/2]$ whose cosine, $c_i(D_1, D_2; \varepsilon)$, is defined by

$$\sup\left\{ \frac{\|P_{D_2\cap\overline{(D_1^{\circ,\varepsilon}+D_2^{\circ,\varepsilon})}} P_{D_1\cap\overline{(D_1^{\circ,\varepsilon}+D_2^{\circ,\varepsilon})}}(x)\|}{\|x\|} \,\middle|\, x \in D_i \cap \overline{(D_1^{\circ,\varepsilon}+D_2^{\circ,\varepsilon})} \cap S(\varepsilon) \right\},$$

where $S(\varepsilon)$ is the ε-sphere in H:

$$S(\varepsilon) := \{x \in H \mid \|x\| = \varepsilon\}.$$

In case $\varepsilon = 0$ or $D_i \cap \overline{(D_1^{\circ,\varepsilon}+D_2^{\circ,\varepsilon})} \cap S(\varepsilon)$ is the empty set, we define $c_i(D_1, D_2; \varepsilon) = 0$.

The condition in Definition 4.1 that 0 is in the intersection of the sets is *not* a restriction for us. This is because, in the application we make to the CPA, the convex sets in question will always be translates of other convex sets obtained by subtracting a point in their intersection. Such translates obviously contain 0.

In the case of *subspaces*, we observe that there is a close connection between the *i*th ε-angle and the usual angle. Indeed, we have the following lemma.

Lemma 4.2. *Let M_1 and M_2 be closed subspaces in H, $M = M_1 \cap M_2$, and $\varepsilon > 0$. Then*

$$c_1(M_1, M_2; \varepsilon) = \|P_{M_2\cap M^\perp} P_{M_1\cap M^\perp}\| = c(M_1, M_2) \tag{4.6}$$

and

$$c_2(M_1, M_2; \varepsilon) = \|P_{M_2\cap M^\perp} P_{M_1\cap M^\perp} P_{M_2\cap M^\perp}\| = c^2(M_1, M_2). \tag{4.7}$$

In particular, the ith *ε-angles are independent of* ε *and are symmetric in* M_1 *and* M_2.

Proof. Since $M_i^{\circ,\varepsilon} = M_i^\perp$ by Corollary 3.5, and since $M^\perp = \overline{M_1^{\circ,\varepsilon} + M_2^{\circ,\varepsilon}}$ by (4.5), we obtain

$$\begin{aligned} c_1(M_1, M_2; \varepsilon) &= \sup\left\{ \frac{\|P_{M_2\cap M^\perp} P_{M_1\cap M^\perp}(x)\|}{\|x\|} \,\middle|\, x \in M_1 \cap M^\perp \cap S(\varepsilon) \right\} \\ &= \sup\left\{ \|P_{M_2\cap M^\perp} P_{M_1\cap M^\perp}(x)\| \mid x \in S(1) \right\} \\ &= \|P_{M_2\cap M^\perp} P_{M_1\cap M^\perp}\| \\ &= c(M_1, M_2) \quad \text{by (4.2).} \end{aligned}$$

This proves (4.6).

Similarly,

$$\begin{aligned} c_2(M_1, M_2; \varepsilon) &= \sup\left\{ \frac{\|P_{M_2\cap M^\perp} P_{M_1\cap M^\perp}(x)\|}{\|x\|} \mid x \in M_2 \cap M^\perp \cap S(\varepsilon) \right\} \\ &= \sup\left\{ \|P_{M_2\cap M^\perp} P_{M_1\cap M^\perp} P_{M_2\cap M^\perp}(x)\| \mid x \in S(1) \right\} \\ &= \|P_{M_2\cap M^\perp} P_{M_1\cap M^\perp} P_{M_2\cap M^\perp}\|. \end{aligned}$$

However, using the idempotency and self-adjointness of metric projections onto subspaces and the fact that $\|AA^*\| = \|A\|^2$ for any bounded linear operator A (see, e.g., [14, Theorem 8.25(3), p. 172]), we see that

$$\begin{aligned} \|P_{M_2\cap M^\perp} P_{M_1\cap M^\perp} P_{M_2\cap M^\perp}\| &= \|P_{M_2\cap M^\perp} P_{M_1\cap M^\perp} P_{M_1\cap M^\perp} P_{M_2\cap M^\perp}\| \\ &= \|(P_{M_2\cap M^\perp} P_{M_1\cap M^\perp})(P_{M_2\cap M^\perp} P_{M_1\cap M^\perp})^*\| \\ &= \|P_{M_2\cap M^\perp} P_{M_1\cap M^\perp}\|^2 \\ &= c^2(M_1, M_2) \quad \text{by (4.2).} \end{aligned}$$

This proves (4.7). □

Actually, Lemma 4.2 has a generalization that is valid for any pair of closed convex *cones*, not just subspaces. However, our proof of this fact depends on the notion of the *norm* of a nonlinear operator, and this will be taken up in the sequel [16].

It is sometimes desirable to allow more than two convex sets in the definition of ε-angle. This suggests the following definition.

Definition 4.3. Let $D_1, D_2, \ldots, D_r$ be r closed convex sets with $0 \in \bigcap_1^r D_i$, let $\varepsilon \geqslant 0$, and let $i = 1$ or $i = r$. The ith *ε-angle* of the ordered collection $\{D_1, D_2, \ldots, D_r\}$ is the angle in $[0, \pi/2]$ whose cosine is defined by

$$\begin{aligned} &c_i(D_1, D_2, \ldots, D_r; \varepsilon) \\ &\quad := \sup\left\{ \frac{\|P_{D_r\cap D^\varepsilon} P_{D_{r-1}\cap D^\varepsilon} \cdots P_{D_1\cap D^\varepsilon}(x)\|}{\|x\|} \,\middle|\, x \in D_i \cap D^\varepsilon \cap S(\varepsilon) \right\} \quad (i = 1, r), \end{aligned} \tag{4.8}$$

where $D^\varepsilon := \overline{\sum_{j=1}^r D_j^{\circ,\varepsilon}}$. In case the set $D_i \cap D^\varepsilon \cap S(\varepsilon)$ is empty or $\varepsilon = 0$, we define $c_i(D_1, D_2, \ldots, D_r; \varepsilon) = 0$.

It is important to note that although we defined *two* angles for a given ordered collection of convex sets, it is actually possible to describe both of them in terms of just a *single* angle! More precisely, it is easy to check that

$$c_r(D_1, D_2, \dots, D_r; \varepsilon) = c_1(D_r, D_1, D_2, \dots, D_r; \varepsilon).$$

In words, *the rth ε-angle of the ordered collection of r sets $\{D_1, D_2, \dots, D_r\}$ is just* 1st *ε-angle of the ordered collection of $r+1$ sets $\{D_r, D_1, D_2, \dots, D_r\}$*. While in principle it is satisfying to know that all of our results could have been described in terms of a single angle, in practice we felt that it was somewhat simpler notationally to describe these results in terms of both c_1 and c_r, and this is what we have done.

By a repeated application of Fact 2.4, one deduces that a product of nonexpansive maps is nonexpansive and, since $0 \in D^\varepsilon \cap (\bigcap_1^r D_i)$, that

$$0 \leqslant c_i(D_1, D_2, \dots, D_r; \varepsilon) \leqslant 1 \quad (i = 1, r)$$

always holds, and hence the ith ε-angles are well-defined. Moreover, if all the sets $D_i = M_i$ are *subspaces* and $M := \bigcap_1^r M_i$, by mimicking the proof of Lemma 4.2, we obtain

$$c_1(M_1, M_2, \dots, M_r; \varepsilon) = \|P_{M_r \cap M^\perp} P_{M_{r-1} \cap M^\perp} \cdots P_{M_1 \cap M^\perp}\| \tag{4.9}$$

and

$$c_r(M_1, M_2, \dots, M_r; \varepsilon) = \|P_{M_r \cap M^\perp} P_{M_{r-1} \cap M^\perp} \cdots P_{M_1 \cap M^\perp} P_{M_r \cap M^\perp}\|. \tag{4.10}$$

In particular, *the* ith *ε-angle $(i = 1, r)$ for a collection of subspaces is* independent *of ε.*

By a repeated application of Fact 2.5, we deduce that these may be rewritten in the form

$$c_1(M_1, \dots, M_r; \varepsilon) = \|P_{M_r} P_{M_{r-1}} \cdots P_{M_1} P_{M^\perp}\|$$

and

$$c_r(M_1, \dots, M_r; \varepsilon) = \|P_{M_r} P_{M_{r-1}} \cdots P_{M_1} P_{M_r} P_{M^\perp}\|.$$

In this latter form, we immediately recognize $c_1(M_1, \dots, M_r; \varepsilon)$ as the cosine of the *angle of the r-tuple* $(M_1, M_2, \dots, M_r)$ that was first defined and used by Bauschke et al. [6].

The relevance of the definition of ith ε-angle for general convex sets is that it is essential to the rate of convergence results obtained in Theorem 4.6 below. Before establishing this, it is convenient to first isolate a few simple but useful facts.

Lemma 4.4. *Let A and B be closed convex sets with $A \subset B$ and $x \in H$. If $P_B x \in A$, then $P_A x = P_B x$.*

Proof. Let $y = P_B x$. Then $y \in A$ and

$$\|x - y\| = d(x, B) \leqslant d(x, A) \leqslant \|x - y\|.$$

Thus the equality $d(x, A) = \|x - y\|$ holds, and hence $y = P_A x$. $\square$

Lemma 4.5. *Let $C_1, C_2, \dots, C_r$ be closed convex sets in H having nonempty intersection, let $x = x_0 \in H$, let (x_n) be the sequence defined in* (2.1), *and let x_∞ be its weak limit. For each $n \geqslant 0$ and each $i \in \{1, 2, \dots, r\}$,*

$$x_{nr+i} - x_\infty = P_{C_i - x_\infty}(x_{nr+i-1} - x_\infty). \tag{4.11}$$

Moreover, for each fixed $n \geqslant 0$ *and each* $\varepsilon > 0$ *that satisfies* $\varepsilon \geqslant \|x_{nr} - x_\infty\|$, *we have*

$$x_{(n+1)r} - x_\infty = P_{C^\varepsilon \cap (C_r - x_\infty)} \cdots P_{C^\varepsilon \cap (C_1 - x_\infty)}(x_{nr} - x_\infty) \tag{4.12}$$

$$= P_{C^\varepsilon \cap (C_r - x_\infty)} \cdots P_{C^\varepsilon \cap (C_1 - x_\infty)}(x_{nr+1} - x_\infty), \tag{4.13}$$

where

$$C^\varepsilon := \overline{\sum_{j=1}^{r} (C_j - x_\infty)^{\circ,\varepsilon}}.$$

In particular, for any $\varepsilon > 0$ *with* $\varepsilon \geqslant \|x - x_\infty\|$, *the relations* (4.12) *and* (4.13) *hold for all* $n \geqslant 0$.

Proof. Eq. (4.11) follows from Fact 2.2. Using (4.11), we next note that since

$$P_{C_i - x_\infty}(x_{nr+i-1} - x_\infty) = x_{nr+i} - x_\infty \in C^\varepsilon \cap (C_i - x_\infty)$$

by Theorem 3.8, it follows from Lemma 4.4 (with $A = C^\varepsilon \cap (C_i - x_\infty)$ and $B = C_i - x_\infty$) that

$$P_{C^\varepsilon \cap (C_i - x_\infty)}(x_{nr+i-1} - x_\infty) = P_{C_i - x_\infty}(x_{nr+i-1} - x_\infty) = x_{nr+i} - x_\infty. \tag{4.14}$$

By a repeated application of (4.14), we deduce that

$$\begin{aligned} x_{(n+1)r} - x_\infty &= x_{nr+r} - x_\infty = P_{C^\varepsilon \cap (C_r - x_\infty)}(x_{nr+r-1} - x_\infty) \\ &= P_{C^\varepsilon \cap (C_r - x_\infty)}\left[P_{C^\varepsilon \cap (C_{r-1} - x_\infty)}(x_{nr+r-2} - x_\infty)\right] \\ &= \cdots \\ &= P_{C^\varepsilon \cap (C_r - x_\infty)} \cdots P_{C^\varepsilon \cap (C_1 - x_\infty)}(x_{nr} - x_\infty), \end{aligned}$$

which verifies (4.12).

But by (4.14) (with $i = 1$) and Theorem 3.8, we get

$$P_{C^\varepsilon \cap (C_1 - x_\infty)}(x_{nr} - x_\infty) = x_{nr+1} - x_\infty \in C^\varepsilon \cap (C_1 - x_\infty)$$

and hence

$$P_{C^\varepsilon \cap (C_1 - x_\infty)}(x_{nr} - x_\infty) = P_{C^\varepsilon \cap (C_1 - x_\infty)}(x_{nr+1} - x_\infty). \tag{4.15}$$

Finally, substitute (4.15) into (4.12) to obtain (4.13), and this completes the proof. □

The use of Lemma 4.5 now easily yields *upper bounds* for the method of cyclic projections.

Theorem 4.6 (*Rate of convergence bounds*). *Let* $C_1, C_2, \ldots, C_r$ *be closed convex sets in* H *with a nonempty intersection, let* $x \in H$, *let the sequence* (x_n) *be defined as in Eq.* (2.1), *and let* x_∞ *denote the weak limit of* (x_n) (*see Theorem* 2.7). *Then for each* $n \geqslant 1$,

$$\begin{aligned} \|(P_{C_r} P_{C_{r-1}} \cdots P_{C_1})^n(x) - x_\infty\| &\leqslant c_{r,n-1} \|(P_{C_r} P_{C_{r-1}} \cdots P_{C_1})^{n-1}(x) - x_\infty\| \\ &\leqslant c_{1,1}\left[\prod_{k=1}^{n-1} c_{r,k}\right] \|x - x_\infty\|, \end{aligned} \tag{4.16}$$

where

$$c_{i,k} := c_i(C_1 - x_\infty, C_2 - x_\infty, \ldots, C_r - x_\infty; \|x_{ik} - x_\infty\|)$$

for $i = 1$ *or* r, *each* $k = 1, 2, \ldots, n-1$, *and* $c_{r,0} := 1$.

Proof. Since $x_{rn} = (P_{C_r} P_{C_{r-1}} \cdots P_{C_1})^n(x)$, it suffices to prove the first inequality of relation (4.16) with x_{rn} instead of $(P_{C_r} P_{C_{r-1}} \cdots P_{C_1})^n(x)$. Since $\|x_1 - x_\infty\| \leqslant \|x_0 - x_\infty\| = \|x - x_\infty\|$ by Proposition 2.6, it suffices to prove the second inequality of (4.16) when $\|x - x_\infty\|$ is replaced by $\|x_1 - x_\infty\|$. Thus it suffices to verify the two inequalities:

$$\|x_{nr} - x_\infty\| \leqslant c_{r,n-1}\|x_{(n-1)r} - x_\infty\|, \tag{4.17}$$

$$c_{r,n-1}\|x_{(n-1)r} - x_\infty\| \leqslant c_{1,1}\left[\prod_{k=1}^{n-1} c_{r,k}\right]\|x_1 - x_\infty\|. \tag{4.18}$$

If $x_{(n-1)r} - x_\infty = 0$, then $x_{nr} - x_\infty = 0$ by Proposition 2.6 and the inequality (4.17) holds trivially, so we may assume that $x_{(n-1)r} - x_\infty \neq 0$. Setting $\varepsilon := \|x_{(n-1)r} - x_\infty\|$, we see by Theorem 3.8 that

$$x_{(n-1)r} - x_\infty \in (C_r - x_\infty) \cap \overline{\sum_{i=1}^{r} (C_i - x_\infty)^{\circ,\varepsilon}}.$$

Using Eq. (4.12), we obtain that

$$x_{nr} - x_\infty = P_{C^\varepsilon \cap (C_r - x_\infty)} \cdots P_{C^\varepsilon \cap (C_1 - x_\infty)}(x_{(n-1)r} - x_\infty),$$

where C^ε is defined as in Lemma 4.5. Thus

$$\begin{aligned}
\|x_{nr} - x_\infty\| &= \|P_{C^\varepsilon \cap (C_r - x_\infty)} \cdots P_{C^\varepsilon \cap (C_1 - x_\infty)}(x_{(n-1)r} - x_\infty)\| \\
&= \|P_{C^\varepsilon \cap (C_r - x_\infty)} \cdots P_{C^\varepsilon \cap (C_1 - x_\infty)} P_{C^\varepsilon \cap (C_r - x_\infty)}(x_{(n-1)r} - x_\infty)\| \\
&\leqslant c_r(C_1 - x_\infty, \ldots, C_r - x_\infty; \|x_{(n-1)r} - x_\infty\|)\|x_{(n-1)r} - x_\infty\| \\
&= c_{r,n-1}\|x_{(n-1)r} - x_\infty\|
\end{aligned}$$

and this proves (4.17).

By a repeated application of (4.17), we obtain

$$\begin{aligned}
\|x_{nr} - x_\infty\| &\leqslant c_{r,n-1}\|x_{(n-1)r} - x_\infty\| \leqslant c_{r,n-1}c_{r,n-2}\|x_{(n-2)r} - x_\infty\| \\
&\leqslant \cdots \leqslant \left(\prod_{k=1}^{n-1} c_{r,k}\right)\|x_r - x_\infty\|.
\end{aligned} \tag{4.19}$$

Using (4.19), we see that to verify (4.18), and thus complete the proof of the theorem, it suffices to show that

$$\|x_r - x_\infty\| \leqslant c_{1,1}\|x_1 - x_\infty\|. \tag{4.20}$$

If $x_1 - x_\infty = 0$, then $x_r - x_\infty = 0$ by Proposition 2.6 and the inequality (4.20) holds trivially. If $x_1 - x_\infty \neq 0$, then with $\varepsilon := \|x_1 - x_\infty\|$, we have that

$$x_1 - x_\infty \in (C_1 - x_\infty) \cap \overline{\sum_{i=1}^{r}(C_i - x_\infty)^{\circ,\varepsilon}}$$

by Theorem 3.8. It follows by relation (4.13) (with $n = 0$) that

$$\begin{aligned}\|x_r - x_\infty\| &= \frac{\|P_{C^\varepsilon \cap (C_r - x_\infty)} \cdots P_{C^\varepsilon \cap (C_1 - x_\infty)}(x_1 - x_\infty)\|}{\|x_1 - x_\infty\|} \|x_1 - x_\infty\| \\ &\leqslant c_1(C_1 - x_\infty, \ldots, C_r - x_\infty; \|x_1 - x_\infty\|)\|x_1 - x_\infty\| \\ &= c_{1,1}\|x_1 - x_\infty\|\end{aligned}$$

and this proves (4.20). □

Our first corollary of Theorem 4.6 is the case when all the convex sets are subspaces.

Corollary 4.7. *Let $M_1, M_2, \ldots, M_r$ be r closed subspaces and $M = \bigcap_1^r M_i$. Then, for every $x \in H$,*

$$\begin{aligned}&\|(P_{M_r} P_{M_{r-1}} \cdots P_{M_1})^n(x) - P_M(x)\| \\ &\quad \leqslant \|P_{M_r \cap M^\perp} \cdots P_{M_1 \cap M^\perp}\| \, \|P_{M_r \cap M^\perp} \cdots P_{M_1 \cap M^\perp} P_{M_r \cap M^\perp}\|^{n-1} \|x\|. \end{aligned} \tag{4.21}$$

Proof. Now $x_\infty = P_M(x)$ by Theorem 2.8, and $\|x - P_M(x)\| \leqslant \|x\|$ since $0 \in M$. Letting $P_i = P_{M_i}$, Theorem 4.6 implies the bound

$$\|(P_r \cdots P_1)^n(x) - P_M(x)\| \leqslant c_{1,1} \left(\prod_{k=1}^{n-1} c_{r,k} \right) \|x - P_M(x)\|, \tag{4.22}$$

where $c_{i,k} = c_i(M_1, \ldots, M_r; \|x_{ik} - x_\infty\|)$. Also, using (4.9) and (4.10) we see that $c_1(M_1, \ldots, M_r; \|x_1 - x_\infty\|) = \|P_{M_r \cap M^\perp} \cdots P_{M_1 \cap M^\perp}\|$ and

$$c_r(M_1, \ldots, M_r; \|x_{rk} - x_\infty\|) = \|P_{M_r \cap M^\perp} \cdots P_{M_1 \cap M^\perp} P_{M_r \cap M^\perp}\|.$$

Substituting these quantities into (4.22), we obtain (4.21). □

In the case of *two* subspaces, the bound given in the inequality (4.21) is *best possible*! We state this next.

Corollary 4.8 (*Aronszajn [1], Kayalar and Weinert [24]*). *Let M_1 and M_2 be closed subspaces with $M := M_1 \cap M_2$. Then for every $x \in H$,*

$$\|(P_{M_2} P_{M_1})^n(x) - P_M(x)\| \leqslant c(M_1, M_2)^{2n-1} \|x\|. \tag{4.23}$$

Moreover, the constant $c(M_1, M_2)^{2n-1}$ in (4.23) *is smallest possible independent of x.*

Proof. Taking $r = 2$ in Corollary 4.7, we obtain

$$\begin{aligned}&\|(P_{M_2} P_{M_1})^n(x) - P_M(x)\| \\ &\quad \leqslant \|P_{M_2 \cap M^\perp} P_{M_1 \cap M^\perp}\| \, \|P_{M_2 \cap M^\perp} P_{M_1 \cap M^\perp} P_{M_2 \cap M^\perp}\|^{n-1} \|x\|. \end{aligned} \tag{4.24}$$

Using Lemma 4.2 completes the proof of (4.23).

Finally, the proof that the constant is smallest possible is due to Kayalar and Weinert [24] since they showed that $\|(P_{M_2} P_{M_1})^n - P_M\| = c(M_1, M_2)^{2n-1}$ (see also [14, Theorem 9.31, p. 219]). □

In contrast to the case when $r = 2$, the following example (adapted from [15, Example 3.8]) shows that for $r > 2$, the bound given in (4.21) is *not* best possible in general.

Example 4.9. Fix any integer $r \geqslant 3$ and choose any $0 < \theta < \pi/2$. In $H = \mathbb{R}^3$, define the r subspaces $M_1 = \operatorname{span}\{e_2, e_3\}$, $M_2 = \operatorname{span}\{(\sin\theta)e_1 + (\cos\theta)e_2, e_3\}$, and $M_3 = M_4 = \cdots = M_r = \operatorname{span}\{e_1, e_2\} = (M_1 \cap M_2)^\perp$, where e_i is the vector that is 1 in the ith coordinate and 0 elsewhere. Then $M := \bigcap_1^r M_i = \{0\}$, $\|P_{M_r} P_{M_{r-1}} \cdots P_{M_1}\| = \|P_{M_r} P_{M_{r-1}} \cdots P_{M_1} P_{M_r}\| = \cos\theta$, and $\|(P_{M_r} P_{M_{r-1}} \cdots P_{M_1})^n - P_M\| = (\cos\theta)^{2n-1}$. Thus the bound given in Corollary 4.7 is given by $(\cos\theta)^n$ which is strictly larger than the sharp upper bound $(\cos\theta)^{2n-1}$ for every $n > 1$.

Proof. It is easy to verify that $M := \bigcap_1^r M_i = \{0\}$. For simplicity, let $P_i = P_{M_i}$ for $i = 1, 2, 3$. We then observe that $P_r P_{r-1} \cdots P_1 = P_3 P_2 P_1$ and $P_r P_{r-1} \cdots P_1 P_r = P_3 P_2 P_1 P_3$. Using Fact 2.1, it is simple to verify the following facts:

$$P_1(e_1) = P_3(e_3) = 0, \quad P_1(e_2) = P_3(e_2) = e_2,$$

$$P_2(e_3) = e_3, \quad P_3(e_1) = e_1, \quad P_2(e_1) = \sin\theta[\sin\theta e_1 + \cos\theta e_2],$$

$$P_2(e_2) = \cos\theta[\sin\theta e_1 + \cos\theta e_2].$$

Using these facts and the linearity of the P_i, it is now easy to deduce that

$$\begin{aligned} &P_3 P_2 P_1 P_3(e_1) = P_3 P_2 P_1(e_1) = 0,\\ &P_3 P_2 P_1 P_3(e_3) = P_3 P_2 P_1(e_3) = 0,\\ &P_3 P_2 P_1 P_3(e_2) = P_3 P_2 P_1(e_2) = \cos\theta[(\sin\theta)e_1 + (\cos\theta)e_2]. \end{aligned}$$

It follows that $P_3 P_2 P_1 P_3 = P_3 P_2 P_1$ and hence that $\|P_3 P_2 P_1 P_3\| = \|P_3 P_2 P_1\| = \cos\theta$. Using these latter facts and induction, we deduce that for each $n \in \mathbb{N}$,

$$\begin{aligned} &(P_3 P_2 P_1)^n(e_1) = 0,\\ &(P_3 P_2 P_1)^n(e_3) = 0,\\ &(P_3 P_2 P_1)^n(e_2) = (\cos\theta)^{2n-1}[(\sin\theta)e_1 + (\cos\theta)e_2]. \end{aligned}$$

Finally, it is straightforward to deduce from the latter equations that $\|(P_3 P_2 P_1)^n\| = (\cos\theta)^{2n-1}$, and this completes the proof. $\square$

Clearly, for the applications of Corollary 4.7, it is important to know when $\|P_{M_r \cap M^\perp} P_{M_{r-1} \cap M^\perp} \cdots P_{M_1 \cap M^\perp} P_{M_r \cap M^\perp}\| < 1$. The following result characterizes this situation.

Theorem 4.10. *Let $M_1, M_2, \ldots, M_r$ be r closed subspaces in the Hilbert space H and let $M = \bigcap_1^r M_i$. Then the following statements are equivalent*:

(1) $M_1^\perp + M_2^\perp + \cdots + M_r^\perp$ *is closed*;
(2) $\|P_{M_r \cap M^\perp} P_{M_{r-1} \cap M^\perp} \cdots P_{M_1 \cap M^\perp}\| < 1$;
(3) $\|P_{M_r \cap M^\perp} P_{M_{r-1} \cap M^\perp} \cdots P_{M_1 \cap M^\perp} P_{M_r \cap M^\perp}\| < 1$;
(4) $c_1(M_1, \ldots, M_r) < 1$;
(5) $c_r(M_1, \ldots, M_r) < 1$.

Proof. The equivalence of statements (1) and (2) is due to Bauschke et al. [6]. The equivalence of statements (2) and (4) (respectively, (3) and (5)) follows from Eq. (4.9) (respectively, (4.10)). Finally, since

$$M_1^\perp + M_2^\perp + \cdots + M_r^\perp + M_r^\perp = M_1^\perp + M_2^\perp + \cdots + M_r^\perp,$$

it follows from the equivalence of statements (1) and (2) that statements (1) and (3) are equivalent. This completes the proof. □

Remarks. (a) Bauschke et al. [6] showed that the statements (1) and (2) are also each equivalent to several "regularity" properties of the subspaces. Moreover, statement (1) is readily seen to be equivalent to the collection of subspaces having the "strong conical hull intersection property" (= strong CHIP) (see [14, Lemma 10.3, p. 239].)

(b) There is a generalization of Theorem 4.10 that is valid for convex *cones*, not necessarily subspaces. Moreover, for the convex cones case, there is a connection to the strong CHIP as well as other regularity properties. This can be found in the sequel to this paper [17].

(c) We should mention that Xu and Zikatanov [29, Theorem 5.2] have given an *identity* for the expression $\|P_{M_r \cap M^\perp} \cdots P_{M_1 \cap M^\perp}\|$.

By considering several different examples, it can be shown that the bound given in Theorem 4.6 is often *best possible* in the sense that the inequality (4.16) is actually an equality for every $n \geqslant 1$. Indeed, with $r = 2$, this bound is *best possible* in each of the following cases:

1. $\{C_1, C_2\}$ is any pair of closed subspaces in a Hilbert space. (This is just Corollary 4.8.)
2. C_1 is a round ball lying on a flat table C_2. (This is a well-known example which exhibits the phenomenon of "tunneling", i.e., a slowing down of the convergence as the limit is approached; see Crombez [11].)
3. $\{C_1, C_2\}$ is any pair of closed convex cones in the Euclidean plane.

In fact (when $r = 2$), we suspect that for any pair of closed convex sets $\{C_1, C_2\}$, the rate of convergence given by Theorem 4.6 will always be best possible, at least for some starting point $x = x_0$ that is not in the intersection $C_1 \cap C_2$.

In a sequel to this paper [16], we will show that there are often useful alternate expressions for the cosines $c_1(C_1, \ldots, C_r; \varepsilon)$ and $c_r(C_1, \ldots, C_r; \varepsilon)$ in terms of the norms of certain *nonlinear* operators (viz., products of metric projections onto convex sets). We will also give conditions on the sets $C_1, \ldots, C_r$ that guarantee that the relevant cosines are strictly less than 1, which is the most important case for the applications of Theorem 4.6.

References

[1] N. Aronszajn, Theory of reproducing kernels, Trans. Amer. Math. Soc. 68 (1950) 337–404.

[2] N. Aronszajn, Introduction to the Theory of Hilbert Spaces, vol. 1, Research Foundation of Oklahoma A & M College, Stillwater, OK, 1950.

[3] J.B. Baillon, H. Brezis, Une remarque sur le comportement asymptotique des semigroupes non lineaires, Houston J. Math. 2 (1) (1976) 5–7.

[4] H.H. Bauschke, Projection algorithms and monotone operators, Doctoral Thesis, Simon Fraser University, 1996.

[5] H.H. Bauschke, J.M. Borwein, On projection algorithms for solving convex feasibility problems, SIAM Rev. 38 (1996) 367–426.

[6] H.H. Bauschke, J.M. Borwein, A.S. Lewis, The method of cyclic projections for closed convex sets in Hilbert space, Contemp. Math. 204 (1997) 1–38.

[7] R.L. Bishop, S.I. Goldberg, Tensor Analysis on Manifolds, Dover edition New York, New York, 1980, pp. 78, 86, 95.

[8] J.P. Boyle, 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, Berlin, 1985, pp. 28–47.

[9] L.M. Bregman, The method of successive projection for finding a common point of convex sets, Soviet Math. Dokl. 6 (1965) 688–692.

[10] P.L. Combettes, The convex feasibility problem in image recovery, Adv. in Imaging Electron Phys. 95 (1996) 155–270.

[11] G. Crombez, Improving the speed of convergence in the method of projections onto convex sets, Publ. Math. Debrecen 58/1–2 (2001) 29–48.
[12] F. Deutsch, Rate of Convergence of the Method of Alternating Projections, ISNM, vol. 72, Birkhäuser Verlag, Basel, 1984, pp. 96–107.
[13] F. Deutsch, The angle between subspaces of a Hilbert space, in: S.P. Singh (Ed.), Approximation Theory Wavelets and Applications, Kluwer Academic Publ., The Netherlands, 1995, pp. 107–130.
[14] F. Deutsch, Best Approximation in Inner Product Spaces, Springer, New York, 2001.
[15] F. Deutsch, H. Hundal, The rate of convergence for the method of alternating projections, II, J. Math. Anal. Appl. 205 (1997) 381–405.
[16] F. Deutsch, H. Hundal, The rate of convergence for the cyclic projections algorithm II: norms of nonlinear operators, J. Approx. Theory, in press, doi:10.1016/j.jat.2006.02.006.
[17] F. Deutsch, H. Hundal, The rate of convergence for the cyclic projections algorithm III: regularity of convex sets, preprint.
[18] E.R. Dougherty, Probability and Statistics for the Engineering, Computing, and Physical Sciences, Prentice-Hall, Englewood Cliffs, New Jersey, 1990, pp. 240, 242.
[19] R.L. Dykstra, An algorithm for restricted least squares regression, J. Amer. Statist. Assoc. 78 (1983) 837–842.
[20] K. Friedrichs, On certain inequalities and characteristic value problems for analytic functions and functions of two variables, Trans. Amer. Math. Soc. 41 (1937) 321–364.
[21] G. Golub, C.F. Van Loan, Matrix Computations, The Johns Hopkins University Press, Baltimore, 1985.
[22] H. Hundal, An alternating projection that does not converge in norm, Nonlinear Anal. 57 (2004) 35–61.
[23] I.T. Jolliffe, Principal Component Analysis, second ed., Springer, New York, 2002.
[24] S. Kayalar, H. Weinert, Error bounds for the method of alternating projections, Math. Control Signals Systems 1 (1988) 43–59.
[25] R.R. Phelps, Convex sets and nearest points I, Proc. Amer. Math. Soc. 8 (1957) 790–797.
[26] F. Riesz, Zur Theorie des Hilbertschen Raumes, Acta Sci. Math. Szeged 7 (1934) 34–38.
[27] R.O. Schmid, Multiple emitter location and signal parameter estimation, IEEE Trans. Antennas and Propagation 34 (1986) 276–280.
[28] K.T. Smith, D.C. Solmon, S.L. Wagner, Practical and mathematical aspects of the problem of reconstructing objects from radiographs, Bull. Amer. Math. Soc. 83 (1977) 1227–1270.
[29] J. Xu, L. Zikatanov, The method of alternating projections and the method of subspace corrections in Hilbert space, J. Amer. Math. Soc. 15 (2002) 573–597.