On the method of cyclic projections for convex sets in Hilbert space

H.H. Bauschke*, J.M. Borwein† and A.S. Lewis‡

February 10, 1994

Abstract

The method of cyclic projections is a powerful tool for solving convex feasibility problems in Hilbert space. Although in many applications, in particular in the field of image reconstruction (electron microscopy, computed tomography), the convex constraint sets do not necessarily intersect, the method of cyclic projections is still employed. Results on the behaviour of the algorithm for this general case are improved, unified, and reviewed. The analysis relies on key concepts from convex analysis and the theory of nonexpansive mappings. The notion of the angle of a tuple of subspaces is introduced. New linear convergence results follow for the case when the constraint sets are closed subspaces whose orthogonal complements have a closed sum; this holds, in particular, for hyperplanes or in Euclidean space.

1991 M.R. Subject Classification.

Primary 47H09, 49M45, 65-02, 65J05, 90C25;

Secondary 26B25, 41A65, 46C99, 46N10, 47N10, 52A05, 52A41, 65F10, 65K05, 90C90, 92C55.

Key words and phrases. Angle, computerized tomography, convex feasibility problem, convex polyhedron, convex set, Fejér monotone sequence, Hilbert space, image reconstruction, Kaczmarz’s method, method of cyclic projections, nearest point mapping, nonexpansive mapping, projection, sum of subspaces, von Neumann’s alternating projection algorithm.

*Centre for Experimental and Constructive Mathematics, Simon Fraser University, Burnaby, B.C., Canada V5A 1S6. Electronic mail: bauschke@ccm.sfu.ca.
†Centre for Experimental and Constructive Mathematics, Simon Fraser University, Burnaby, B.C., Canada V5A 1S6. Electronic mail: jborwein@ccm.sfu.ca. Research supported by NSERC and by the Shrum Endowment.
‡Department of Combinatorics and Optimization, University of Waterloo, Waterloo, Ont., Canada N2L 3G1. Electronic mail: mlewis@orion.uwaterloo.ca. Research partially supported by NSERC.
Contents

1 The problem 1
  1.1 Introduction and formulation .............................. 1
  1.2 Interpretations and notation ............................. 4

2 Projections 4
  2.1 Basic properties ........................................... 4
  2.2 Examples .................................................. 5
  2.3 Baillon and Bruck’s quasi-projection ..................... 6

3 A toolbox 7
  3.1 Polar cones ............................................... 7
  3.2 Recession cones ......................................... 7
  3.3 Convex polyhedra ........................................ 8
  3.4 Polyhedral multifunctions .............................. 9
  3.5 Kadec/Klee sets ........................................ 10
  3.6 Regularities ............................................. 11
  3.7 The angle of a tuple of subspaces ...................... 14
  3.8 Fejér monotone sequences ................................ 17

4 Nonexpansive mappings 18
  4.1 Some bits of general theory ............................ 18
  4.2 Firmly nonexpansive mappings .......................... 19
  4.3 Bruck and Reich’s strongly nonexpansive mappings ... 20
  4.4 Nonexpansive mappings à la De Pierro and Iusem ... 22
  4.5 Strongly attracting mappings ........................... 23

5 Main results 23
  5.1 Geometry of the problem ................................ 23
  5.2 Dichotomy ................................................ 25
  5.3 Minimal displacement ................................... 26
  5.4 Convergence results I ................................. 29
  5.5 Convergence results II .................................. 30
  5.6 The convex polyhedral case ............................ 32
  5.7 The affine subspace case ............................... 33
  5.8 Related algorithms .................................... 41
1 The problem

1.1 Introduction and formulation

A frequent problem in various areas of mathematics and physical sciences consists of determining a “solution” in form of a point satisfying some convex “constraints”, i.e. lying in the intersection of certain convex sets. This problem is usually referred to as the Convex Feasibility Problem.

Often these constraint sets are closed subspaces in some Hilbert space; an immensely successful algorithm for solving this problem is the method of cyclic projections, where a sequence is generated by projecting cyclically onto the constraint subspaces. The fundamental result is due to von Neumann and Halperin:

**Fact 1.1.1** Suppose $C_1, \ldots, C_N$ are intersecting closed affine subspaces with corresponding projections $P_1, \ldots, P_N$ in some Hilbert space $H$. If $x_0 \in H$, then the sequence

$$ (x_0, x_1 = P_1 x_0, x_2 = P_2 x_1, \ldots, x_N = P_N x_{N-1}, x_{N+1} = P_1 x_N, \ldots) $$

converges in norm to the projection of $x_0$ onto $\bigcap_{i=1}^N C_i$.

In 1933, von Neumann [72, Theorem 13.7] proved this result for two closed subspaces; it was later extended to finitely many closed subspaces by Halperin [45]. (The generalization to intersecting closed affine subspaces is easy.)

An impressive survey on applications of the method of cyclic projections for intersecting closed affine subspaces can be found in Deutsch’s [25]. In particular, if each $C_i$ is a hyperplane, then one obtains the well-known method of Kaczmarz [49] for solving systems of linear equations.

In some important applications, however, Fact 1.1.1 cannot be used since either the constraints are not affine (i.e. the $C_i$ are not all closed affine subspaces) or the constraints are inconsistent (i.e. the intersection $\bigcap_{i=1}^N C_i$ is empty) or even both.

Non-affine constraints occur, for instance, when the desired solution is supposed to be nonnegative, or bounded by some prescribed value, or a solution of a linear inequality.

In theory the mathematical models employed almost always possess consistent constraints; in practice, however, due (unavoidably) to inaccurate measurements, one is often confronted with inconsistent constraints.

Inconsistent (non-affine) constraints constitute a typical example of an ill-posed problem. They occur frequently in the field of image reconstruction: for instance in signal reconstruction (Goldburg and Marks [41]), in electron microscopy (Sezan [66]), in limited-angle tomography (Boyd and Little [13]), and in medical imaging (Viergever [71]); in particular in computed tomography (Censor [20]).
Our objective in this paper is to improve, unify, and review results on the method of cyclic projections for possibly non-intersecting closed convex nonempty sets in Hilbert space. We will also speculate a little on some open questions and hope very much to transmit some of our interest and enthusiasm to the reader.

The mathematical model we analyze is described as follows:

**SETTING.** Throughout the paper, we assume that $H$ is a real Hilbert space and $C_1, \ldots, C_N$ are closed convex nonempty subsets of $H$ with corresponding projections $P_1, \ldots, P_N$. (We discuss projections in Section 2.) For an arbitrary starting point $x^0 \in H$, the method of cyclic projections generates $N$ sequences $(x^n_i)$ by

\[
x_0^1 := P_1 x^0, \quad x_1^1 := P_2 x_0^1, \quad \ldots, \quad x_{N-1}^1 := P_N x_{N-1}^1,
\]

\[
x_0^2 := P_1 x_0^1, \quad x_1^2 := P_2 x_1^2, \quad \ldots, \quad x_{N-1}^2 := P_N x_{N-1}^2,
\]

\[
x_0^3 := P_1 x_0^2, \quad \ldots.
\]

We collect these sequences cyclically in one sequence $(x^0, x_1^1, x_2^1, \ldots, x_N^1, x_1^2, x_2^2, \ldots)$ to which we refer as the orbit generated by $x^0$ or the orbit with starting point $x^0$. We further define

\[
Q_1 := P_1 P_N P_{N-1} \cdots P_2, \quad Q_2 := P_2 P_1 P_N P_{N-1} \cdots P_3, \quad \ldots, \quad Q_N := P_N P_{N-1} \cdots P_1,
\]

which allows us to write more concisely

\[
x_{i+1}^n := Q_i^{n-1} x_i^1, \quad \text{for all } n \geq 1\text{ and every } i;
\]

for convenience, we set $P_0 := P_N$, $P_{N+1} := P_1$, $x_0^n := x_{N-1}^{n-1}$, and $x_{N+1}^n := x_1^{n+1}$, so that

\[
x_{i+1}^n = P_{i+1} x_i^n, \quad \text{for all } n \geq 1\text{ and every } i.
\]

When appropriate, we will similarly identify $i = 0$ with $i = N$ and $i = N + 1$ with $i = 1$.

**OVERVIEW.**

Basic facts on projections and some examples are contained in Section 2. We touch on quasi-projections, a useful concept related to projections and recently introduced by Baillon and Bruck [5].

In Section 3, we provide a “toolbox” which makes the subsequent analysis clearer: some of our results rely indirectly on the notion of a polar cone. Recession cones are intimately connected with convex sets and their projections. One important example of a convex set is a convex polyhedron: here the study of the mappings $Q_i$ naturally involves polyhedral multifunctions. We give some
basic facts on convex Kadec/Klee sets, a notion which generalize the better-
known concept of (locally) uniformly convex sets. A discussion on (bounded) 
(linear) regularity of $N$-tuples of closed convex sets follows; this key condition 
is sufficient for norm convergence of any orbit in the consistent case. We then 
introduce the notion of the angle of an $N$-tuple $(L_1, \ldots, L_N)$ of closed subspaces 
and show that the angle is positive if and only if $L_1^+ + \cdots + L_N^+$ is closed. This 
strengthens results on the well-known two set case (Deutsch [24], Bauschke and 
Borwein [9]) and is used to yield new linear convergence results. The section 
concludes with a short review on Féjér monotone sequences.

Each projection $P_i$ is nonexpansive and so are the compositions $Q_i$. Moreover, 
they actually satisfy various stronger conditions introduced by Bruck and Reich 
[18], by De Pierro and Iusem [63], and by Bauschke and Borwein [8]. These 
conditions, which yield significant information on the behaviour of orbits, are 
discussed in Section 4.

Section 5 contains our main results. After considering the geometry of the 
problem, we give a dichotomy result on orbits which roughly says that if each $Q_i$ 
is fixed point free, then the orbit has no bounded subsequence; otherwise, each 
subsequence $(x^n_i)$ converges weakly to some fixed point of $Q_i$. (Throughout this 
section, we specialize our general results to the two set case ($N = 2$) and to the 
consistent case ($\bigcap_{i=1}^N C_i \neq \emptyset$) to allow comparison with known results for these 
important special cases.) Two central questions arise:

1. When does each $Q_i$ have a fixed point?

2. If each $Q_i$ has a fixed point, when do the subsequences $(x^n_i)$ converge in 
norm (or even linearly)?

Concerning Question 1, we provide sufficient conditions for the existence of fixed 
points or of approximate fixed points (that is, $\inf_{x \in H} ||x - Q_ix|| = 0$, for each $i$). 
It follows that while fixed points of $Q_i$ need not exist for non-intersecting closed 
affine subspaces, approximate fixed points must.

In respect to Question 2, a variety of conditions guaranteeing norm convergence 
in the presence of fixed points for each $Q_i$ is offered: one of the sets $C_i$ has 
to be (boundedly) compact or all sets are Kadec/Klee, or convex polyhedra, or 
closed affine subspaces. In the affine subspace case, each sequence $(x^n_i)$ converges 
to the fixed point of $Q_i$ nearest to $x^0$. Moreover, the convergence will be 
linear, whenever the angle of the $N$-tuple of the associated closed subspaces is 
positive. In contrast, if $N = 2$ and the angle is zero, then the convergence will 
be “arbitrarily slow”. We draw connections to the theory of generalized inverses 
by specializing to hyperplanes with linearly independent normal vectors.

The analysis in Section 5 extends results by Bauschke and Borwein [9], by Ch-
enev and Goldstein [22], by Gubin et al. [44], by De Pierro and Iusem [63], by 
Kosmol [52], and by Kosmol and Zhou [53, 54]. We end the paper by referring 
the reader to related algorithms.
1.2 Interpretations and notation

Because the orbit \( (x^n_i) \) is generated by the rule \( x^n_i := P_ix^n_{i-1} \), the problem described in the previous subsection allows interpretations in the following (related) mathematical areas:

- **Difference Equations**: the orbit \( (x^n_i) \) is given by a *nonautonomous difference equation* ([55]).
- **Discrete Dynamical Systems**: the orbit \( (x^n_i) \) is the orbit of a *discrete dynamical system* ([58]).
- **Discrete Processes**: the orbit \( (x^n_i) \) is the *positive orbit of a difference equation* ([56]).

Less seriously, one can also think about our problem as a *deterministic magnetic billiard*: the starting point \( x^0 \) is a *metal billiard-ball* and the sets \( C_i \) are *electric magnets*. Now switch the magnets on and off cyclically: the trajectory of the billiard ball corresponds exactly to the orbit \( (x^n_i) \).

**NOTATION.** The norm of a point \( x \in H \) is given by \( \|x\| := \sqrt{\langle x, x \rangle} \), where \( \langle \cdot, \cdot \rangle \) is the inner product of \( H \). We write \( I \) for the identity mapping on \( H \), \( B_H \) for the unit ball \( \{ x \in H : \|x\| \leq 1 \} \), and \( B(x_0, r) \) for \( x_0 + rB_H \), where \( x_0 \in H \) and \( r \geq 0 \). Suppose \( C \) is a nonempty subset of \( H \). The *interior* (resp. *intrinsic core*, *boundary*, *closure*, *span*) of \( C \) is denoted \( \text{int}(C) \) (resp. \( \text{icr}(C) \), \( \text{bd}(C) \), \( C \), \( \text{span}(C) \)). The *indicator function* of a \( C \), \( \iota_C \), is defined by \( \iota_C(x) = 0 \), if \( x \in C \); \(+\infty \), otherwise: its Fenchel conjugate, the *support function* of \( C \), is \( \iota_C^* = \sup(C, \cdot) \). The *normal cone* \( N_C(x) \) at \( x \in C \) is given by \( N_C(x) = \{ x^* \in H : \langle x^*, C - x \rangle \leq 0 \} \). If \( C \) is a subspace, then we write \( C^\perp \) for its *orthogonal complement*. The subdifferential of a convex function \( f \) from \( H \) to \( \mathbb{R} \) at \( x \in H \) is abbreviated \( \partial f(x) = \{ x^* \in H : \langle x^*, h \rangle \leq f(x + h) - f(x) \ \forall h \in H \} \).

If \( D \) is another nonempty subset of \( H \), then the *gap between \( C \) and \( D \)* is defined by \( d(C,D) := \inf_{c \in C, d \in D} |c - d| \). For sequences, we use the symbol \( \to \) (resp. \( \rightharpoonup \)) to indicate *norm* (resp. *weak*) *convergence*. Finally, we write \([ \cdot ]\) for the *mod function* with values in \( \{1, \ldots, N\} \).

2 Projections

2.1 Basic properties

**Definition 2.1.1** Suppose \( C \) is a closed convex nonempty subset of \( H \). Then for every \( x \in H \), there exists a unique nearest point, say \( P_Cx \), to \( x \) in \( C \):

\[
\|x - P_Cx\| \leq \|x - c\|, \quad \text{for all } c \in C.
\]
The associated mapping $P_C$ from $H$ to $C$ which sends every point to its nearest point in $C$ is called the projection onto $C$. The function
\[ d(\cdot, C) : H \rightarrow \mathbb{R} : x \mapsto \|x - P_Cx\| \]
is called the distance function (to $C$).

We list basic facts on projections. (i) and (ii) are folklore (see for instance [75, Section 1]) whereas (iii) is a simple but useful expansion.

**Proposition 2.1.2** Suppose $C$ is a closed convex nonempty subset of $H$. Then:

(i) The projection onto $C$ is well-defined and the distance function to $C$ is continuous and convex, hence weakly lower semicontinuous.

(ii) (Kolmogorov’s Criterion) For any two points $x, c^* \in H$, we have $c^* = P_Cx$ if and only if
\[ c^* \in C \quad \text{and} \quad \langle C - c^*, x - c^* \rangle \leq 0. \]

(iii) For any $x, y \in H$, the formula below contains exclusively nonnegative terms:
\[
\|x - y\|^2 = \|(I - P_C)x - (I - P_C)y\|^2 + \|P_Cx - P_Cy\|^2 + 2\langle (I - P_C)x - (I - P_C)y, P_Cx - P_Cy \rangle
\]

in particular, $\langle x - y, P_Cx - P_Cy \rangle \geq \|P_Cx - P_Cy\|^2$.

**Corollary 2.1.3** Suppose $(x^n_i)$ (resp. $(y^n_i)$) is the orbit generated by some $x^0$ (resp. $y^0$) in $H$. Then the sequences

\[ (\|x^n_i - y^n_i\|, (x^n_i, y^n_i), (x^n_i - y^n_i, x^n_i - y^n_i)) \]

are nonnegative and tend to zero for each $i$. In fact, the series generated by the terms of these sequences is convergent.

### 2.2 Examples

**Examples 2.2.1** Suppose $C$ is a closed convex nonempty subset of $H$ and $x$ is an arbitrary point in $H$.

(i) If $C = B_H$, then $P_Cx = \begin{cases} x, & \text{if } \|x\| \leq 1; \\ x/\|x\|, & \text{otherwise.} \end{cases}$

(ii) If $C$ is a hyperplane, say $C = \{x \in H : \langle a, x \rangle = b\}$ with $a \in H \setminus \{0\}$, then
\[ P_Cx = x - ((\langle a, x \rangle - b)/\|a\|^2)a. \]
(iii) If $C$ is a halfspace, say $C = \{x \in H : \langle a, x \rangle \leq b \}$ with $a \in H \setminus \{0\}$, then $P_C x = x - \frac{(\langle a, x \rangle - b)^+}{\|a\|^2}a$.

(iv) If $H$ is a Hilbert lattice with lattice cone $H^+$, then $P_C x = x^+$.

(v) Suppose $C$ is a closed convex cone. If the negative polar cone is denoted $C^\ominus$ (see Subsection 3.1), then $P_{C^\ominus}(x) = x - P_C x$. In particular, if $C$ is a closed subspace, then $P_{C^\perp}(x) = x - P_C x$.

(vi) Suppose $Y$ is another Hilbert space and $A : H \to Y$ is a continuous linear operator with closed range. Let $C = \text{range}(A^*)$. If the generalized inverse of $A$ is denoted $A^+$ (see Definition 5.7.10), then $P_C x = A^+Ax$.

(vii) Suppose $Y = \mathbb{R}^N$ and a continuous linear operator $A$ is given by $Ah = (\langle a_1, h \rangle, \ldots, \langle a_N, h \rangle)$ for some vectors $a_1, \ldots, a_N \in H$. Let $C = \text{span}\{a_1, \ldots, a_N\}$. Then $P_C x = A^+Ax$. Furthermore, if the vectors $a_1, \ldots, a_N$ are linearly independent, then $P_C x = A^*(A^A)^{-1}Ax$.

**Proof.** (i), (ii), (iii), and (v) are easy to check. (iv) is proved by Borwein and Yost’s [12, Theorem 8]. (vi) is due to Desoer and Whalen: see [23]. (vii) is a special case of (vi) and was used in [9, Section 5].

The following proposition is useful for building more examples of projections. We omit the simple proof.

**Proposition 2.2.2** Suppose $C$ is a closed convex nonempty subset of $H$ and $x$ is an arbitrary point in $H$. Then:

(i) (translation formula) If $z \in H$, then $P_{z+C}(x) = z + P_C(x - z)$.

(ii) (dilation formula) If $\rho \neq 0$, then $P_{\rho C}(x) = \rho P_C(x/\rho)$.

### 2.3 Baillon and Bruck’s quasi-projection

**Definition 2.3.1** (Baillon and Bruck’s quasi-projection [5]) Suppose $C$ is a closed convex nonempty subset of $H$. For every $x \in H$, define the quasi-projection of $x$ onto $C$, $Q_C x$, by

$$Q_C x = \{c^* \in C : \|c^* - c\| \leq \|x - c\|, \text{ for all } c \in C\}.$$ 

It is easy to verify the following proposition (see also [6]):

**Proposition 2.3.2** Suppose $C$ is a closed convex nonempty subset of $H$ and $x$ is a point in $H$. Then:

(i) $Q_C x$ is a bounded closed convex nonempty subset of $C$; moreover:

$$P_C x \in Q_C x \subseteq C \cap B(P_C x, d(x, C)).$$

Consequently, if $x \in C$, then $Q_C x = \{P_C x\}$. 

6
(ii) (translation formula) \( Q_{z+Cx} = z + Q_C(x-z) \), for all \( z \in H \).

(iii) If \( C \) is a closed affine subspace, then \( Q_C x = \{ P_C x \} \).

3 A toolbox

3.1 Polar cones

**Definition 3.1.1** The positive (resp. negative) polar cone \( S^\oplus \) (resp. \( S^\ominus \)) of a nonempty subset \( S \) of \( H \) is given by

\[ S^\oplus := \{ x \in H : \langle x, S \rangle \geq 0 \} \quad (\text{resp. } S^\ominus := \{ x \in H : \langle x, S \rangle \leq 0 \} = -S^\ominus). \]

The next proposition is readily verified:

**Proposition 3.1.2**

(i) Every positive (resp. negative) polar cone is a closed convex nonempty cone.

(ii) If \( S_1, S_2, \ldots, S_M \) are nonempty subsets of \( H \), then \( S_1^\oplus + S_2^\oplus + \cdots + S_M^\oplus \subseteq (S_1 \cap S_2 \cap \cdots \cap S_M)^\oplus \).

The class of closed subspaces can be characterized in the class of closed convex nonempty cones:

**Proposition 3.1.3** Suppose \( S \) is a closed convex nonempty cone in \( H \). Then \( S \) is a subspace if and only if \( S \cap S^\ominus = \{0\} \).

**Proof.** “\( \Rightarrow \)”: If \( S \) is a subspace, then \( S^\oplus = S^\perp \) and hence \( S \cap S^\ominus = \{0\} \).

“\( \Leftarrow \)”: By Examples 2.2.1(v) and Proposition 2.2.2, \( x = P_S x - P_{S^\ominus}(-x) \), for all \( x \in H \). Let us assume that \( S \) is not a subspace. Then there exists some vector \( x \in S \) such that \( -x \not\in S \), that is \( P_S(-x) \neq -x \). Thus \( -x = P_S(-x) - P_{S^\ominus} x \), where \( P_{S^\ominus} x \neq 0 \). Since \( P_{S^\ominus} x = P_S(-x) + x \in S \), we have found \( P_{S^\ominus} x \in (S \cap S^\ominus) \setminus \{0\} \). The proof is complete. \( \square \)

3.2 Recession cones

**Definition 3.2.1** The recession cone \( \text{rec}(C) \) of a closed convex nonempty subset \( C \) of \( H \) is given by

\[ \text{rec}(C) := \{ x \in H : x + C \subseteq C \}. \]

**Proposition 3.2.2** Suppose \( C \) is a closed convex nonempty subset of \( H \). Then:

(i) (Zarantonello's [75, Section 3]) \( \text{rec}(C) \) is a closed convex nonempty cone.
(ii) A point \( x \in H \) belongs to \( \text{rec}(C) \) if and only if there exists a sequence \( (c_n) \) in \( C \) and a sequence of reals \( (r_n) \) tending to \( +\infty \) such that \( (c_n/r_n) \) converges weakly to \( x \).

(iii) If \( D \) is a closed convex superset of \( C \), then \( \text{rec}(C) \subseteq \text{rec}(D) \).

**Proof.** (ii): \( \Longrightarrow \): Since \( \text{rec}(C) \) is cone, points of the form \( c + nx \) belong to \( C \), for all \( c \in C \) and \( n \geq 1 \). Let us choose \( c_n := c + nx \) and \( r_n := n \). Then the sequence \( (c_n/r_n) = (c/n + x) \) converges to \( x \), as desired.

\( \Longleftarrow \): Fix any \( c \in C \); then \( (1 - 1/r_n)c + c_n/r_n \in C \), for all large \( n \). Since \( C \) is weakly closed, the weak limit \( c + x \) lies in \( C \). Finally, (iii) is obvious from (ii). \( \square \)

There is an intimate connection between the projection, the recession cone, and the indicator function of a given closed convex nonempty set:

**Fact 3.2.3** Suppose \( C \) is a closed convex nonempty subset of \( H \). Then:

(i) (Castaing and Valadier’s [19, Proposition I-7]) \( \text{rec}(C) = (\text{domain}(i_C))^{\circ} \).

(ii) (Zarantanello’s [75, Theorem 3.1]) \( \text{range}(P_C - I) = (\text{rec}(C))^{\circ} \).

**Remark 3.2.4** The set \( \text{domain}(i_C) = \{ x \in H : \sup \langle x, C \rangle < \infty \} \) is also known as the **barrier cone** of \( C \); see for instance [4].

### 3.3 Convex polyhedra

**Definition 3.3.1** The intersection of finitely many halfspaces is called a **convex polyhedron**.

**Proposition 3.3.2** Suppose each \( C_i \) is a convex polyhedron given as \( \{ x \in H : \langle a_{i,j}, x \rangle \leq b_{i,j} \} \) for finitely many vectors \( a_{i,j} \) and real numbers \( b_{i,j} \). Suppose further \( K \) is a closed subspace of \( \bigcap_{i,j} \ker(a_{i,j}) \). Define \( D_i := C_i \cap K^\perp \). Then:

(i) Each \( D_i \) is a convex polyhedron in \( K^\perp \).

(ii) \( P_i = P_K + P_{D_i}P_{K^\perp} \), for each \( i \).

(iii) \( I - Q_N = P_{K^\perp} - P_{D_N} \cdots P_{D_1}P_{K^\perp} \).

(iv) \( \inf_{x \in H} \| x - Q_N x \| \) is attained if and only if \( \inf_{y \in K^\perp} \| y - P_{D_N} \cdots P_{D_1} y \| \) is.

**Proof.** (i) is obvious. (ii): Fix \( x \in H \) and an index \( i \). We check Kolmogorov’s criterion: since \( C_i = K \oplus D_i \), the point \( P_{K^\perp} x + P_{D_i} P_{K^\perp} x \) belongs to \( C_i \). Pick any \( k \in K \) and \( d_i \in D_i \). Then

\[
\langle k + d_i - (P_K x + P_{D_i} P_{K^\perp} x), x - (P_K x + P_{D_i} P_{K^\perp} x) \rangle \\
= \langle k - P_K x, (x - P_K x) - P_{D_i} P_{K^\perp} x \rangle + \langle d_i - P_{D_i} P_{K^\perp} x, P_{K^\perp} x - P_{D_i} P_{K^\perp} x \rangle \\
\leq 0,
\]

8
as desired. Hence (ii) holds. (iii) follows from (ii) inductively whereas (iv) is obvious from (iii).

\[ \Box \]

**Remark 3.3.3** It follows that \( \text{range}(I - P_i) \) is a subset of \( K^\perp \). Ky Fan ([34, Theorem 14]) showed that \( \text{range}(I - P_i) \) is actually contained in the cone generated by the \( a_{ij} \); this will, however, not be needed in the sequel.

### 3.4 Polyhedral multifunctions

**Definition 3.4.1** Suppose \( X, Y \) are Euclidean spaces and \( \Omega : X \rightrightarrows Y \) is a multifunction. If the graph of \( \Omega \) can be written as the union of finitely many convex polyhedra, then \( \Omega \) is called a **polyhedral multifunction**.

**Fact 3.4.2** (Robinson [64, Section 1]) The class of polyhedral multifunctions is closed under addition, scalar multiplication, composition, and inversion.

**Fact 3.4.3** (Robinson [64, Section 1]) If \( A \) is a linear mapping between the Euclidean spaces \( X \) and \( Y \), then \( A \) is a polyhedral multifunction. If \( D \) is a convex polyhedron, then \( \partial_{t_D}(\cdot) = N_D(\cdot) \) is a polyhedral multifunction.

**Corollary 3.4.4** If \( D_1, \ldots, D_N \) are convex polyhedra in the Euclidean space \( X \), then \( I - P_{D_N} \cdots P_{D_1} \) is a single-valued polyhedral multifunction.

**Proof.** \( y = P_{D_1} x \iff x \in (I + \partial_{t_{D_1}})(y) \iff y \in (I + \partial_{t_{D_1}})^{-1}(x) \). Now \( I \) and \( \partial_{t_{D_1}} \) are polyhedral multifunctions by Fact 3.4.3; hence so is \( I - P_{D_N} \cdots P_{D_1} \) by repeated application of Fact 3.4.2.

**Fact 3.4.5** (Robinson’s [64, Proposition 2]) If \( \Omega \) is a polyhedral multifunction between the Euclidean spaces \( X \) and \( Y \) and if \( W \) is a bounded subset of \( \text{range}(\Omega) \), then there exists some bounded subset \( V \) of \( X \) such that \( \Omega(V) \subseteq W \).

**Corollary 3.4.6** If \( D_1, \ldots, D_N \) are convex polyhedra in the Euclidean space \( X \), then the infimum \( \inf_{x \in X} \|x - P_{D_N} \cdots P_{D_1}\| \) is attained.

**Proof.** \( \Omega := I - P_{D_N} \cdots P_{D_1} \) is polyhedral and single-valued by Corollary 3.4.4. Let \( \beta > \inf_{x \in X} \|\Omega x\| \) and set \( W := \text{range}(\Omega) \cap \beta B_X \). By the previous fact, we get \( \alpha > 0 \) such that \( \Omega(\alpha B_X) \subseteq W \). Thus \( \inf_{x \in X} \|\Omega x\| = \inf_{x \in \alpha B_X} \|\Omega x\| \); however, this latter infimum is attained, since \( \Omega \) is continuous and \( \alpha B_X \) is compact.

**Remark 3.4.7** In particular, if each \( D_i \) is a halfspace in the Euclidean space \( X \), then the infimum \( \inf_{x \in X} \|x - P_{D_N} \cdots P_{D_1}\| \) is attained — this also follows from De Pierro and Iusem’s [63, Proposition 13].

Combining Proposition 3.3.2.(iv) and Corollary 3.4.6 thus yields:

**Corollary 3.4.8** If each \( C_i \) is a convex polyhedron, then the infimum \( \inf_{x \in H} \|x - Q_N x\| \) is attained.
3.5 Kadec/Klee sets

The following definition generalizes well-known notions of rotundity of unit balls in Banach spaces; see Deville et al.’s recent monograph [27].

**Definition 3.5.1** Suppose $C$ is a closed convex nonempty subset of $H$. We say that $C$ is:

(i) *Kadec*Klee, if \[ (x_n) \text{ in } C, \quad x_n \to x \in \text{bd}(C) \] implies $x_n \to x$.

(ii) *uniformly convex*, if \[ (x_n), (y_n) \text{ in } C, \quad d((x_n + y_n)/2, \text{bd}(C)) \to 0 \] implies $x_n - y_n \to 0$.

(iii) *locally uniformly convex*, if \[ (x_n) \text{ in } C, \quad x \in C, \quad d((x_n + x)/2, \text{bd}(C)) \to 0 \] implies $x_n \to x$.

(iv) *strictly convex*, if \[ x, y \in C, \quad (x + y)/2 \in \text{bd}(C) \] implies $x = y$.

**Proposition 3.5.2** Suppose $C$ is a closed convex nonempty subset of $H$. Then:

(i) $C$ uniformly convex $\Rightarrow$ $C$ locally uniformly convex $\Rightarrow$ $C$ strictly convex.

(ii) $C$ strictly convex and compact $\Rightarrow$ $C$ uniformly convex.

(iii) If $\text{int}(C) = \emptyset$, then: $C$ is strictly convex $\Rightarrow$ $C$ is singleton.

(iv) If $\text{int}(C) = \emptyset$, then: $C$ is Kadec/Klee $\iff$ $C$ is boundedly compact.

(v) $C$ boundedly compact $\Rightarrow$ $C$ Kadec/Klee $\Rightarrow$ $C$ boundedly compact or $\text{int}(C) \neq \emptyset$.

(vi) $C$ locally uniformly convex $\Rightarrow$ $C$ Kadec/Klee.

(vii) The function

\[
\delta_C(t) := \sup \{ r \geq 0 : B((x+y)/2, r) \subseteq C, \text{ for all } x, y \in C \text{ with } \|x-y\| = t \}
\]

is increasing. Moreover: $C$ is uniformly convex $\iff \delta_C(t) > 0$, for all $t > 0$.

**Proof.** We prove here only (vi); the rest is straightforward. Suppose $C$ is locally uniformly convex. In view of (iii) and (iv), we can assume that $\text{int}(C) \neq \emptyset$. Pick a sequence $(x_n)$ in $C$ that converges weakly to some $x \in \text{bd}(C)$. Then $x$ is a support point of $C$ and there is a support functional $f$ with $f(x) \geq f(C)$ and $\|f\| = 1$. Let $t_n := 1/n + (f(x) - f(x_n))/2$, for all $n$; then $t_n \to 0^+$. Also, $f(t_n f + (x_n + x)/2) = 1/n + f(x) > f(x)$ so that $t_n f + (x_n + x)/2 \notin C$, for all $n$. Hence $d((x_n + x)/2, \text{bd}(C)) \leq \|[x_n + x]/2 - (x_n + x)/2 + t_n f\| = t_n \to 0$. Therefore, since $C$ is locally uniformly convex, $x_n \to x$ and (vi) holds. \[ \square \]
Remarks 3.5.3

- Let \( H = \mathbb{R}^2 \) and suppose \( C = \{ (x, y) \in H : y \geq 1/x > 0 \} \). Then \( C \) is locally uniformly convex but not uniformly convex. Hence the assumption of compactness in (ii) is important.

- Every non-singleton boundedly compact closed convex nonempty set is Kadec/Klee; thus there are many Kadec/Klee sets which are not strictly convex.

- (Borwein and Fitzpatrick's [10, Example following Corollary 3.5]) Let \( H = \ell_2 \times \mathbb{R} \) and define a new norm \( \| \cdot \| \) on \( H \) by \( \| (x, r) \|^2 := \sum_n (x_n/n)^2 + r^2 + \max \{ \sum_n x_n^2, r^2 \} \). Suppose \( C := \{ (x, r) \in H : \| (x, r) \| \leq 1 \} \). Then \( C \) is strictly convex. Denote the \( n^{th} \) unit vector of \( \ell_2 \) by \( e_n \). Then the sequence \( (e_n, 1) / \sqrt{2 + 1/n^2} \) lies in the boundary of \( C \) and converges weakly to \( (0, 1 / \sqrt{2}) \in \text{bd}(C) \), but not in norm. Therefore, \( C \) is an strictly convex set with nonempty interior but \( C \) is neither Kadec/Klee nor locally uniformly convex.

- Borwein and Lewis [11] showed that the notions of locally uniform convexity and strict convexity coincide in Euclidean spaces.

**Definition 3.5.4** Suppose \( C \) is a uniformly convex subset of \( H \). If there is some \( \gamma > 0 \) such that \( \delta_C(t) \geq \gamma t^2 \), then \( C \) is called strongly convex.

**Remarks 3.5.5**

- If \( C = \rho B_H \), where \( \rho > 0 \), then \( \delta_C(t) \geq t^2/(8 \rho) \) (use the Parallelogram Law) and hence \( C \) is strongly convex.

- Let \( H = \mathbb{R}^2 \) and suppose \( C = \{ (x, y) \in H : x^4 + y^4 \leq 1 \} \). Fix an arbitrary \( 0 < s < 1 \) and set \( x_s = (s, (1-s)^{1/4}), \ y_s = (s, -(1-s)^{1/4}) \). Then \( x_s, y_s \in C \), \( (x_s + y_s)/2 = (s, 0) \), and \( d((x_s + y_s)/2, \text{bd}(C)) \leq \| (x_s + y_s)/2 - (1, 0) \| = 1 - s \). Also, \( \| x_s - y_s \| = 2(1 - s)^{1/4} \), and thus \( \delta_C(2(1 - s)^{1/4}) \leq 1 - s \). Changing variables, we obtain \( \delta_C(t) \leq 1 - (1 - (t/2)^{1/4})^{1/4} = t^4/64 + \mathcal{O}(t^8) \). Therefore, \( \delta_C(t)/t^2 \leq t^2/64 + \mathcal{O}(t^6) \) and \( C \) is not strongly convex.

### 3.6 Regularities

**The two set case**

**Definition 3.6.1** Suppose \( N = 2 \) and \( d_1 := P_{C_2-C_1} \in C_2 - C_1 \). We say that \( (C_1, C_2) \) is:

1. **boundedly regular**, if
   
   \[
   (x_n) \text{ is a bounded sequence in } H, \quad \max \{ d(x_n, C_1), d(x_n, C_2 - d_1) \} \to 0 \implies d(x_n, C_1 \cap (C_2 - d_1)) \to 0.
   \]
(ii) \textit{regular}, if
\[
(x_n) \text{ is a sequence in } H, \quad \max\{d(x_n, C_1), d(x_n, C_2 - d_1)\} \to 0 \quad \text{implies } d(x_n, C_1 \cap (C_2 - d_1)) \to 0.
\]

Obviously, regularity implies bounded regularity. These conditions have turned out to be sufficient for norm convergence of orbits; see Subsection 5.5.

\textbf{Fact 3.6.2} Suppose \( N = 2 \).

(i) ([9, Theorem 3.9]) If \( C_1 \) or \( C_2 \) is boundedly compact and \( d_1 \in C_2 - C_1 \), then \((C_1, C_2)\) is boundedly regular.

(ii) ([9, Corollary 3.14]) If \( C_1 \) or \( C_2 \) is bounded and \((C_1, C_2)\) is boundedly regular, then \((C_1, C_2)\) is regular.

\textbf{Correction 3.6.3} The first two authors wish to point out the following lacunae in their paper [9]:

- The definition of bounded regularity [9, Definition 3.6] needs in addition the assumption \( R_{C_2 - C_1} \in C_2 - C_1 \) and is then equivalent to the present, more handy Definition 3.6.1.

- Similarly, [9, Theorems 3.7, 3.8, and Corollary 3.10] need in addition the assumption \( R_{C_2 - C_1} \in C_2 - C_1 \).

\textit{The consistent case}

\textbf{Definition 3.6.4} Suppose \( \bigcap_{i=1}^N C_i \neq \emptyset \). We say that \((C_1, \ldots, C_N)\) is:

(i) \textit{boundedly regular}, if \( \max_i d(x_n, C_i) \to 0 \) implies \( d(x_n, \bigcap_i C_i) \to 0 \), for all bounded sequences \((x_n)\) in \( H \).

(ii) \textit{regular}, if \( \max_i d(x_n, C_i) \to 0 \) implies \( d(x_n, \bigcap_i C_i) \to 0 \), for all sequences \((x_n)\) in \( H \).

(iii) \textit{boundedly linearly regular}, if for every bounded subset \( S \) of \( H \), there exists \( \kappa_S > 0 \) such that \( d(x, \bigcap_i C_i) \leq \kappa_S \max_i d(x, C_i) \), for all \( x \in S \).

(iv) \textit{linearly regular}, if there exists \( \kappa > 0 \) such that \( d(x, \bigcap_i C_i) \leq \kappa \max_i d(x, C_i) \), for all \( x \in H \).

Clearly,

\[
\begin{align*}
\text{linearly regular} & \implies \text{boundedly linearly regular} \\
& \implies \text{regular} \\
\text{regular} & \implies \text{boundedly regular},
\end{align*}
\]

and again we prefer the more intuitive definition of (bounded) regularity to the equivalent [8, Definition 5.1].

12
**Theorem 3.6.5**

(i) ([8, Proposition 5.4.(i)]) If some $C_i$ is boundedly compact, then $(C_1, \ldots, C_N)$ is boundedly regular.

(ii) ([8, Proposition 5.4.(iii)]) If $H$ is finite-dimensional, then $(C_1, \ldots, C_N)$ is always boundedly regular.

(iii) If every $C_i$, except possibly one, is Kadec/Klee, then $(C_1, \ldots, C_N)$ is boundedly regular.

(iv) ([8, Proposition 5.4.(ii)]) If $(C_1, \ldots, C_N)$ is boundedly regular and some $C_i$ is bounded, then $(C_1, \ldots, C_N)$ is regular.

(v) (Gubin et al.’s [44, Proof of Lemma 5]) If every $C_i$, except possibly one, is uniformly convex, then $(C_1, \ldots, C_N)$ is regular.

**Proof.** (iii): After relabeling if necessary, we assume that $C_1, \ldots, C_{N-1}$ are Kadec/Klee. Now pick a bounded sequence with $\max_{i=1}^{N} d(x_n, C_i) \to 0$. After passing to a subsequence if necessary, we assume that $(d(x_n, \bigcap_{i=1}^{N} C_i))$ converges to some limit, say $L$. It suffices to show that $L = 0$. After passing once more to a subsequence, we also assume that $(x_n)$ converges weakly to some $x$, where $x$ lies necessarily in $\bigcap_{i=1}^{N} C_i$ (Proposition 2.1.2.(i)). If $x \in \text{bd}(C_i)$, for some $i \in \{1, \ldots, N-1\}$, then $(P_i x_n)$ converges in norm to $x$ (because $x_n - P_i x_n \to 0$ and $C_i$ is Kadec/Klee) and so does $(x_n)$ implying $L = 0$, as desired. Otherwise, $x \in C_N \cap \text{int}(\bigcap_{i=1}^{N-1} C_i)$. Then, by Fact 3.6.9, $(C_1, \ldots, C_N)$ is even boundedly linearly regular. The proof is complete. 

**Remark 3.6.6**

- Gubin et al. ([44, Lemma 5]) claimed only bounded regularity in (iv); their proof, however, works for regularity.

- For an example of two intersecting sets $C_1$ and $C_2$ where $(C_1, C_2)$ is not boundedly regular, see Bauschke and Borwein’s [9, Example 5.5]. Hence the assumption of bounded compactness in (i) is important.

We now give some useful results on (bounded) linear regularity; see also [8].

**Fact 3.6.7** ([8, Theorem 5.19]) Suppose $L_1, \ldots, L_N$ are closed subspaces of $H$. Then $L_1^+ + \cdots + L_N^+$ is closed if and only if $(L_1, \ldots, L_N)$ is regular in any of the four senses. Consequently, if each $C_i$ is a closed affine subspace, where $C_i = c + L_i$ and $L_i$ is linear, then $(C_1, \ldots, C_N)$ is regular in any of the four senses if and only if $L_1^+ + \cdots + L_N^+$ is closed.

**Fact 3.6.8** ([8, Corollary 5.26]) If each $C_i$ is a convex polyhedron, then $(C_1, \ldots, C_N)$ is linearly regular.
Fact 3.6.9 ([8, Corollaries 5.13 and 5.14]) If \( 0 \in \bigcap_{i=1}^{N-1} \text{int}((\bigcap_{j=1}^{i} C_j) - C_{i+1}) \), then \((C_1, \ldots, C_N)\) is boundedly linearly regular. In particular, this happens whenever \( C_N \cap \text{int}(\bigcap_{i=1}^{N-1} C_i) \neq \emptyset \).

Remark 3.6.10 Since the definition of (bounded) (linear) regularity does not depend on the order of the sets, the last fact has corresponding formulations for all other orders of the sets.

The general case

In general, without understanding the geometry of the problem (see Subsection 5.1), it seems to be hard even to suggest useful notions of regularity. We feel that a positive answer to Conjecture 3.1.6 would yield the appropriate general notion of regularity.

3.7 The angle of a tuple of subspaces

In this subsection, we generalize the notion of the angle between two subspaces to finitely many subspaces. The results are also interesting from an operator-theoretic point of view and will lead to some new linear convergence results in Subsection 5.7.

Throughout this subsection, we assume that

\[
L_1, L_2, \ldots, L_N \text{ are closed subspaces and } L := \bigcap_{i=1}^{N} L_i.
\]

Fact 3.7.1 (Kayalar and Weinert’s [50, Theorem 1])

\[
P_{L_N} \cdots P_{L_2} P_{L_1} = P_{L_N \cap L^\perp} \cdots P_{L_1 \cap L^\perp}.
\]

The next proposition is easily verified.

Proposition 3.7.2 Suppose \( B \) is a closed subspace and \( A \) is a subset of \( B^\perp \). Then \( A \) is closed if and only if \( A \oplus B \) is.

Proposition 3.7.3 The following are equivalent:

(i) \((L_1 \cap L^\perp, \ldots, L_N \cap L^\perp)\) is regular in any of the four senses.

(ii) \(L_1^\perp + \cdots + L_N^\perp + L\) is closed.

(iii) \((L_1, \ldots, L_N)\) is regular in any of the four senses.

(iv) \(L_1^\perp + \cdots + L_N^\perp\) is closed.
Proof. In view of Fact 3.6.7, it suffices to establish the equivalence of (ii) and (iv). By Proposition 3.7.2, $(L_i \cap L^\perp)^\perp = L_i^\perp + L = L_i^\perp + L$ for each $i$. Using Proposition 3.7.2 once more, we get: $L_i^\perp + \cdots + L_N^\perp$ is closed $\iff (L_i^\perp + \cdots + L_N^\perp) + L$ is closed $\iff (L_i + L) + \cdots + (L_N + L)$ is closed $\iff (L_i \cap L^\perp)^\perp + \cdots + (L_N \cap L^\perp)^\perp$ is closed. \hfill \Box \\

**Theorem 3.7.4** The Hilbert space operator norm $\|P_{L_N} \cdots P_{L_1} P_{L^\perp}\|$ is less than 1 if and only if the sum $L_i^\perp + \cdots + L_N^\perp$ is closed.

**Proof.** $\implies$ : We prove the contrapositive and assume that $L_i^\perp + \cdots + L_N^\perp$ is not closed. By Proposition 3.7.3, $(L_i \cap L^\perp, \ldots, L_N \cap L^\perp)$ is not boundedly regular. Hence there is some bounded sequence $(x_n)$ with $\max_i d(x_n, L_i \cap L^\perp) > 0$, but $d(x_n, \bigcap_i L_i \cap L^\perp) = \|x_n\| \neq 0$. After passing to a subsequence and normalizing if necessary, we may assume $\|x_n\| = 1$ for all $n$. Now $x_n - P_{L_i \cap L^\perp} x_n \to 0$, which implies, by nonexpansivity of $P_{L_i \cap L^\perp}$, that $P_{L_2 \cap L^\perp} x_n = P_{L_2 \cap L^\perp} P_{L_i \cap L^\perp} x_n \to 0$. Since $x_n - P_{L_2 \cap L^\perp} x_n \to 0$, we obtain $x_n - P_{L_2 \cap L^\perp} P_{L_1 \cap L^\perp} x_n \to 0$. Repeating this line of thought yields eventually

$$x_n - P_{L_N \cap L^\perp} \cdots P_{L_i \cap L^\perp} x_n \to 0.$$ 

The triangle inequality implies $\|x_n\| - \|P_{L_N \cap L^\perp} \cdots P_{L_i \cap L^\perp} x_n\| \to 0$; thus $\|P_{L_N \cap L^\perp} \cdots P_{L_i \cap L^\perp} x_n\| \to 1$. Therefore, by Fact 3.7.1, $\|P_{L_N \cap L^\perp} \cdots P_{L_i \cap L^\perp}\| = \|P_{L_N} \cdots P_{L_i} P_{L^\perp}\| = 1$.

$\iff$ : By Proposition 3.7.3, $(L_i \cap L^\perp, \ldots, L_N \cap L^\perp)$ is linearly regular. Hence there is some $\kappa > 0$ such that

$$\|x\| = d(x, \bigcap_i L_i \cap L^\perp) \leq \kappa \max_i d(x, L_i \cap L^\perp), \quad \text{for all } x \in H.$$ 

On the other hand, by Fact 4.5.2 below,

$$d^2(x, L_i \cap L^\perp) \leq \|x - P_{L_i \cap L^\perp} \cdots P_{L_i \cap L^\perp} x\|^2 \leq 2^{i-1} (\|x\|^2 - \|P_{L_i \cap L^\perp} \cdots P_{L_i \cap L^\perp} x\|^2) \leq 2^{N-1} (\|x\|^2 - \|P_{L_N \cap L^\perp} \cdots P_{L_i \cap L^\perp} x\|^2),$$

for all $x \in H$ and each $i$. Altogether,

$$\|x\|^2 \leq \kappa^2 2^{N-1} (\|x\|^2 - \|P_{L_N \cap L^\perp} \cdots P_{L_i \cap L^\perp} x\|^2), \quad \text{for all } x \in H,$$

which implies $\|P_{L_N \cap L^\perp} \cdots P_{L_i \cap L^\perp}\| \leq (1 - \kappa^{-2} 2^{-(N-1)})^{1/2} < 1$. The theorem now follows from Fact 3.7.1. \hfill \Box \\

**Definition 3.7.5** We define the angle $\gamma := \gamma(L_1, \ldots, L_N) \in [0, \pi/2]$ of the $N$-tuple $(L_1, \ldots, L_N)$ by

$$\cos \gamma = \|P_{L_N} \cdots P_{L_1} P_{L^\perp}\|.$$ 

15
Remarks 3.7.6

(i) The classical definition of the angle \( \gamma := \gamma(L_1, L_2) \) between two subspaces was given by Friedrich [37] in 1937:

\[
\cos \gamma = \sup \{ \langle l_1, l_2 \rangle : l_1 \in L_1 \cap L_1^\perp; \ l_2 \in L_2 \cap L_2^\perp; \ |l_1|, |l_2| \leq 1 \}. 
\]

Deutsch ([24, Proof of Lemma 2.5.(iii)]) showed that \( \cos \gamma = \| P_{L_1} P_{L_1} P_{(L_2 \cap L_1)^\perp} \| \); therefore, \( \gamma = \gamma \) and our Definition 3.7.5 of the angle is consistent with Friedrich’s original one.

(ii) The importance of the angle between two subspaces stems from the fact that a positive angle guarantees linear convergence of the sequence of alternating projections; see, for instance, [3, pages 378ff.], [9, Theorem 4.11], [24, Theorem 2.3], and [50, Theorem 2]. In Subsection 5.7, we will show among other things that a positive angle of an \( N \)-tuple of subspaces yields linear convergence of the sequence of cyclic projections as well.

Recently, Deutsch and Hundal [26] carefully analyzed Dykstra’s algorithm for intersecting halfspaces and obtained finite and linear convergence results. It is interesting to note that they provide an upper bound for the rate of convergence which depends on the angle of tuples of subspaces as in Definition 3.7.5.

(iii) By (i), the angle between two subspaces is independent of the order, that is \( \gamma(L_1, L_2) = \gamma(L_2, L_1) \). For \( N \geq 3 \), the angle of the \( N \)-tuple \( (L_1, \ldots, L_N) \) may well depend on the order of the subspaces; to see this, we pick three unit vectors \( a, b, c \) with \( \langle b, c \rangle \neq 0 \) and \( \langle a, b \rangle \neq \langle a, c \rangle \). (This requires \( H \) to be of dimension at least 2.) Let \( L_1 := L_2 := \cdots := L_{N-2} := \text{span}(a), \ L_{N-1} := \text{span}(b), \ L_N := \text{span}(c) \); then \( L_{N-1} \cap L_N = \{0\} \). Hence \( L = \{0\} \) and one easily calculates:

\[
\begin{align*}
\| P_{L_1} P_{L_{N-1}} P_{L_{N-2}} \cdots P_{L_4} P_{L_3} P_{L_2} \| &= \| P_{L_1} P_{L_{N-1}} P_{L_{N-2}} \cdots P_{L_4} P_{L_3} P_{L_2} P_{L_1} \| = |\langle a, b \rangle| |\langle b, c \rangle| \\
\| P_{L_{N-2}} P_{L_1} P_{L_{N-1}} P_{L_{N-2}} \cdots P_{L_4} P_{L_3} P_{L_2} \| &= \| P_{L_{N-1}} P_{L_1} P_{L_{N-1}} P_{L_{N-2}} \cdots P_{L_4} P_{L_3} P_{L_2} \| = |\langle a, c \rangle| |\langle c, b \rangle|.
\end{align*}
\]

Thus \( \gamma(L_1, \ldots, L_{N-2}, L_{N-1}, L_N) \neq \gamma(L_1, \ldots, L_{N-2}, L_N, L_{N-1}) \).

In general, it follows from Theorem 3.7.4 that \( \gamma(L_1, \ldots, L_N) \) is positive if and only if \( \gamma(L_{\tau(1)}, \ldots, L_{\tau(N)}) \) is, for every permutation \( \tau \) of \( \{1, \ldots, N\} \).

(iv) Although the angle of the \( N \)-tuple \( (L_1, \ldots, L_N) \) depends on the order, it is true that \( \gamma(L_1, \ldots, L_N) = \gamma(L_N, \ldots, L_1) \), because (see Fact 3.7.1)

\[
\| P_{L_N} \cdots P_{L_1} P_{L_1} \| = \| P_{L_N \cap L_1} \cdots P_{L_1 \cap L_1} \| = \| (P_{L_N \cap L_1} \cdots P_{L_1 \cap L_1})' \| = \| P_{L_1 \cap L_1} \cdots P_{L_N \cap L_1} \| = \| P_{L_1} \cdots P_{L_N} P_{L_1} \|. 
\]

This was observed by Kayalar and Weinert ([50, Corollary 1]) and it explains once more why \( \gamma(L_1, L_2) \) equals \( \gamma(L_2, L_1) \).
**Proposition 3.7.7** The angle of the $N$-tuple $(L_1, \ldots, L_N)$ is positive if and only if the sum $L_1^+ + \cdots + L_N^+$ is closed. In particular, this holds whenever one of the following conditions is satisfied:

(i) Some $L_i \cap L_i^+$ is finite-dimensional.

(ii) Some $L_i$ is finite-dimensional.

(iii) $H$ is finite-dimensional.

(iv) All $L_i$, except possibly one, are finite-codimensional.

(v) Each $L_i$ is a hyperplane.

(vi) Each angle $\gamma(L_i \cap L_j \cap \cdots \cap L_N)$ is positive.

**Proof.** The first statement is just a reformulation of Theorem 3.7.4. (i). By assumption, $L_i \cap L_i^+$ is boundedly compact. Hence Theorem 3.6.5.(i) implies bounded regularity of the $N$-tuple $(L_1 \cap L_1^+, \ldots, L_N \cap L_N^+)$ which is equivalent to the closedness of $L_1^+ + \cdots + L_N^+$ (Proposition 3.7.3). (ii) and (iii) clearly imply (i). Condition (iv) yields linear regularity of $(L_1, \ldots, L_N)$ and hence closedness of $L_1^+ + \cdots + L_N^+$ by [8, Corollary 5.21.(ii)] and Proposition 3.7.3. Obviously, (v) implies (iv). Finally, (vi) follows from [8, Example 6.18].

**Remarks 3.7.8**

- Gaffke and Mathar [38] suggested condition (i) in their study of Dykstra’s algorithm. Condition (vi) is popular in computed tomography; see Smith et al.’s [67, Theorem 2.2].

- For $N = 2$, the angle between $L_1$ and $L_2$ is positive if and only if $L_1^+ + L_2^+$ is closed which in turn is equivalent to the closedness of $L_1 + L_2$ (see, for instance, [48, Corollary 35.6]). For $N \geq 3$, however, the closedness of the sum $L_1 + \cdots + L_N$ is independent of the closedness of the sum $L_1^+ + \cdots + L_N^+$ ([8, Remarks 5.20]).

### 3.8 Fejér monotone sequences

**Definition 3.8.1** Suppose $C$ is a closed convex nonempty subset of $H$. A sequence $(x_n)_{n \geq 0}$ is called Fejér monotone w.r.t. $C$, if

$$
\|x_{n+1} - c\| \leq \|x_n - c\|, \quad \text{for every } c \in C \text{ and all } n \geq 0.
$$

Fejér monotone sequences are a useful concept when dealing with projection algorithms or sequences of iterates of nonexpansive mappings. In our context, the most important property of Fejér monotone sequences is the following (others can be found in [8]):

17
**Proposition 3.8.2** Suppose \((x_n)_{n \geq 0}\) is Fejér monotone w.r.t. \(C\). Then:

(i) \((x_n)\) converges weakly to some point in \(C\) if and only if all weak cluster points of \((x_n)\) lie in \(C\). In this case, the weak limit of \((x_n)\) belongs to \(Q_{C;x_0}\).

(ii) \((x_n)\) converges in norm to some point in \(C\) if and only if \((x_n)\) has at least one norm cluster point in \(C\).

**Proof.** (i): Combine [8, Theorem 2.16.(ii)] and [6, Proposition 2.6.(vi)]. (ii) is [8, Theorem 2.16.(v)]. □

# 4 Nonexpansive mappings

## 4.1 Some bits of general theory

**Definition 4.1.1** A selfmapping \(T\) of \(H\) is called **nonexpansive**, if

\[
\|Tx - Ty\| \leq \|x - y\|, \quad \text{for all } x, y \in H.
\]

A point \(x \in H\) is a **fixed point** of \(T\), if \(x = Tx\); the set of all fixed points of \(T\) is denoted \(\text{Fix } T\).

Obviously, the class of nonexpansive mappings is closed under composition. In view of Proposition 2.1.2, we thus have:

**Corollary 4.1.2** Each \(P_i\) and each \(Q_i\) is nonexpansive.

**Proposition 4.1.3** (minimal displacement vector) Suppose \(T\) is a nonexpansive selfmapping of \(H\). Then:

(i) (Pazy’s [60, Lemma 4]) The set \(\text{range}(T - I)\) is convex and consequently has a unique element of minimal norm (namely the projection of 0 onto \(\text{range}(T - I)\)) which we call the **minimal displacement vector** of \(T\) and denote \(v(T)\).

(ii) The set \(\{x \in H : Tx - x = v(T)\}\) is closed and convex. In particular, \(\text{Fix } T\) is a closed convex (possibly empty) subset of \(H\).

(iii) (Pazy’s [60, Corollary 2]) For every \(x \in H\), the sequence \(((1/n)T^nx)\) converges in norm to \(v(T)\).

(iv) The minimal displacement vector \(v(T)\) is contained in \(\text{rec}(\text{conv}(\text{range}(T)))\).
Proof. (ii): Let $F := (I + T)/2$. Then $F$ is firmly nonexpansive (see the following subsection). Since $\|Fx - x\| = \|Tx - x\|/2$ for all $x$, we have $\{x \in H : Tx - x = v(T)\} = \{x \in H : Fx - x = v(F)\}$. Now the latter set is convex by Corollary 4.2.4.

(iv): $T^nx \in \text{conv}(\text{range}(T))$ for all $n$, thus $v(T) \in \text{rec(\text{conv}(\text{range}(T)))}$ by (iii) and Proposition 3.2.2.(ii). 

Proposition 4.1.4 Suppose $T$ is a nonexpansive selfmapping of $H$ with Fix $T \neq \emptyset$. Then the sequence of iterates $(T^nx)_{n \geq 0}$ is Fejér monotone w.r.t. Fix $T$.

Proof. By Proposition 4.1.3.(ii), Fix $T$ is closed and convex. If $f \in \text{Fix} T$ and $n \geq 0$, then $\|T^{n+1}x - f\| = \|T^{n+1}x - Tf\| \leq \|T^nx - f\|$, the result follows. 

Corollary 4.1.5 If Fix $Q_i$ is nonempty, then the sequence $(x^i_n)$ is Fejér monotone w.r.t. Fix $Q_i$.

Fact 4.1.6 (Browder [16]/Göhde [40]/Kirk [51]) Every nonexpansive selfmapping of a bounded closed convex nonempty subset of the Hilbert space $H$ has at least one fixed point.

Remark 4.1.7 Numerous generalizations of this result have been established (see, for instance, Goebel and Kirk’s [39]); however, here we need only the Hilbert space case for which Pazy [60, Corollary 5] gave a very instructive proof.

4.2 Firmly nonexpansive mappings

Definition 4.2.1 A nonexpansive selfmapping $T$ of $H$ is called firmly nonexpansive, if

$$\langle Tx - Ty, x - y \rangle \geq \|Tx - Ty\|^2, \quad \text{for all } x, y \in H.$$ 

Fact 4.2.2 (Goebel and Kirk’s [39, Theorem 12.1], Rockafellar’s [65, Proposition 1], or Zarantonello’s [75, Lemma 1.3]) Suppose $T$ is a selfmapping of $H$. Then the following are equivalent:

(i) $T$ is firmly nonexpansive.

(ii) $I - T$ is firmly nonexpansive.

(iii) $\langle (I - T)x - (I - T)y, Tx - Ty \rangle \geq 0$, for all $x, y \in H$.

(iv) $2T - I$ is nonexpansive.

Corollary 4.2.3 Projections are strongly nonexpansive. Consequently, each $P_i$ is firmly nonexpansive.
\textbf{Proof.} Immediate from Proposition 2.1.2.(iii) and Fact 4.2.2. \hfill \Box

\textbf{Corollary 4.2.4} Suppose $T$ is a firmly nonexpansive selfmapping of $H$ and $x$ is a point in $H$. Then $Tx - x = v(T)$ if and only if $(y - Ty + v(T), Ty - v(T) - x) \geq 0$, for all $y \in H$. Consequently, the set $\{x \in H : Tx - x = v(T)\}$ is closed and convex.

\textbf{Proof.} Follows easily from Fact 4.2.2.(iii). \hfill \Box

Among all nonexpansive mappings, the class of firmly nonexpansive mappings is probably the closest to the class of projections. It possesses, however, a serious drawback:

\textbf{Example 4.2.5} The class of firmly nonexpansive mapping is not closed under composition: consider in $H := \mathbb{R}^2$ the mappings $T_1(x, y) := P_{\text{span}\{1, 0\}}(x, y) = (x, 0)$ and $T_2(x, y) := P_{\text{span}\{1, 1\}}(x, y) = (x + y, x + y)/2$. Then $T_2T_1(x, y) = (x, x)/2$ and $T_2T_1(1, -2) - T_2T_1(0, 0), (1, -2) - (0, 0) = -1/2 \not\geq 0$; thus $T_2T_1$ is not firmly nonexpansive.

\textbf{Conclusion:} The composition of two projections can fail even to be firmly nonexpansive.

Useful substitutes to “repair this defect” were provided by Bruck and Reich, by De Pierro and Iusem, and by Bauschke and Borwein: they introduced classes of nonexpansive mappings which are closed under composition and preserve characteristic features of projections.

4.3 \textbf{Bruck and Reich’s strongly nonexpansive mappings}

\textbf{Definition 4.3.1} (Bruck and Reich [18]) A nonexpansive selfmapping $T$ of $H$ is called \textit{strongly nonexpansive}, if for every pair of sequences $(x_n), (y_n)$,

\[
\begin{cases}
(x_n - y_n) \text{ is bounded,} \\
\|x_n - y_n\| - \|Tx_n - Ty_n\| \to 0
\end{cases}
\]

implies $(x_n - y_n) - (Tx_n - Ty_n) \to 0$.

\textbf{Proposition 4.3.2} Projections are strongly nonexpansive. Consequently, each $P_e$ is strongly nonexpansive.

\textbf{Proof.} Let $P$ be a projection. If $(x_n - y_n)$ is bounded, then so is $(\|Px_n - Py_n\|)$ and hence $(\|x_n - y_n\| + \|Px_n - Py_n\|)$. Thus if $\|x_n - y_n\| - \|Px_n - Py_n\| \to 0$, then $\|x_n - y_n\|^2 - \|Px_n - Py_n\|^2 \to 0$ and Proposition 2.1.2 yields $(x_n - y_n) - (Px_n - Py_n) \to 0$. \hfill \Box

\textbf{Remark 4.3.3} In fact, Bruck and Reich [18, Proposition 2.1] proved that every firmly nonexpansive mapping is strongly nonexpansive whenever the underlying space is \textit{uniformly convex} (which is the case in our Hilbert space setting).
**Fact 4.3.4** (Bruck and Reich’s [18, Lemma 2.1]) Suppose \( T_1, \ldots, T_N \) are strongly nonexpansive selfmappings of \( H \). Then \( T_N T_{N-1} \cdots T_1 \) is strongly nonexpansive. Moreover, Fix \( T_N T_{N-1} \cdots T_1 = \bigcap_{j=1}^N \text{Fix } T_j \) whenever the latter intersection is nonempty.

**Corollary 4.3.5** Each \( Q_i \) is strongly nonexpansive. Moreover, Fix \( Q_i = \bigcap_{j=1}^N C_j \) provided the latter intersection is nonempty.

**Remark 4.3.6** In view of the last remark and Example 4.2.5, we see that the class of firmly nonexpansive mappings is a proper subclass of the class of strongly nonexpansive mappings.

**Fact 4.3.7** (Bruck and Reich’s [18, Corollaries 1.3, 1.4, and 1.5]) Suppose \( T \) is a strongly nonexpansive selfmapping of \( H \). Then:

(i) Fix \( T \neq \emptyset \) if and only if \( (T^n x) \) converges weakly to some fixed point of \( T \), for all \( x \in H \).

(ii) Fix \( T = \emptyset \) if and only if \( \lim_n \|T^n x\| = +\infty \), for all \( x \in H \).

(iii) \( \lim_n T^{n+1} x - T^n x = v(T) \).

**Theorem 4.3.8** The minimal displacement vectors \( v(Q_1) = \lim_n x_i^{n+1} - x_i^n, \ldots, v(Q_N) = \lim_n x_N^{n+1} - x_N^n \) all coincide and are contained in

\[
\bigcap_{i=1}^N \text{rec}(C_i) \cap \left( \bigcap_{i=1}^N \text{rec}(C_i) \right)^{\oplus}.
\]

**Proof.** By Fact 4.3.7.(iii), \( x_i^{n+1} - x_i^n = Q_i^n x_i^1 - Q_i^{n-1} x_i^1 \rightarrow v_i \), for each \( i \). Corollary 2.1.3 implies (after choosing \( y^0 := x_N^n \))

\[
0 \leftarrow (x_i^{n+1} - x_i^n) - (x_{i+1}^{n+1} - x_{i+1}^n), \quad \text{for each } i;
\]

consequently, \( v(Q_1) = \cdots = v(Q_N) \). On the one hand, by Proposition 4.1.3.(iv) and Proposition 3.2.2.(iii), \( v(Q_i) \in \text{rec} \left( \text{conv}(\text{range}(Q_i)) \right) \subseteq \text{rec} \left( \text{conv}(\text{range}(P_i)) \right) = \text{rec}(C_i) \), for each \( i \). On the other hand,

\[
x_N^{n+1} - x_N^n = (x_N^{n+1} - x_{N-1}^{n+1}) + (x_{N-1}^{n+1} - x_{N-2}^{n+1}) + \cdots + (x_1^{n+1} - x_1^n)
\in \text{range}(P_N - I) + \text{range}(P_{N-1} - I) + \cdots + \text{range}(P_1 - I),
\]

hence \( v(Q_N) \in \text{range}(P_N - I) + \text{range}(P_{N-1} - I) + \cdots + \text{range}(P_1 - I) \). Therefore, by Fact 3.2.3.(ii) and Proposition 3.1.2.(ii),

\[
v(Q_N) \in \bigoplus \text{rec}(C_N) + \cdots + \bigoplus \text{rec}(C_i) \subseteq \left( \bigcap_{i=1}^N \text{rec}(C_i) \right)^{\oplus}.
\]

The proof is complete. \( \square \)

21
**Definition 4.3.9** We abbreviate \( v(Q_1) = v(Q_2) = \cdots = v(Q_N) \) by \( v \) and refer to \( v \) as the **minimal displacement vector**. The corresponding nonnegative real number \( \mu := \|v\| = \inf_x \|x - Q_i x\| = \inf_x \|x - Q_2 x\| = \cdots = \inf_x \|x - Q_N x\| \) is called the **minimal displacement**. We say that the **minimal displacement \( \mu \) is attained** (resp. **unattained**), if the infimum \( \inf_x \|x - Q_i x\| \) is (resp. is not) attained. Finally, we write shorter \( F_i \) instead of \( \text{Fix } Q_i \), for each \( i \).

**Proposition 4.3.10**

(i) Each \( F_i \) is a closed convex subset of \( C \). If \( \bigcap_{j=1}^N C_j \neq \emptyset \), then each \( F_i \) is equal to this intersection.

(ii) \( \mu \) is attained if and only if each infimum \( \inf_x \|x - Q_i x\| \) is.

(iii) Either each \( F_i \) is nonempty or each \( F_i \) is empty.

**Proof.** (i) follows from Proposition 4.1.3.(ii) and Corollary 4.3.5.

(ii), (iii): Suppose \( \mu \) is attained; that is \( \mu = v(Q_1) = \|\bar{x} - Q_i \bar{x}\| \), for some \( \bar{x} \in H \). Then \( \mu = v(Q_2) \leq \|Q_2 P_2 \bar{x} - P_2 \bar{x}\| = \|P_2 Q_2 \bar{x} - P_2 \bar{x}\| \leq \|Q_i \bar{x} - \bar{x}\| = \mu \); hence \( \inf_x \|x - Q_2 x\| \) is attained (at \( P_2 \bar{x} \)) and so are the remaining infima. \( \square \)

**Remark 4.3.11** As we will see in the next subsection, the following is actually true: if \( \mu \) is attained, then \( \mu = 0 \).

### 4.4 Nonexpansive mappings à la De Pierro and Iusem

**Definition 4.4.1** (De Pierro and Iusem [63]) We say a nonexpansive selfmapping \( T \) of \( H \) is **DPI-nonexpansive**, if for all \( x, y \in H \):

\[
\|Tx - Ty\| = \|x - y\| \quad \text{implies} \quad \begin{cases} 
Tx - Ty = x - y, \\
\langle x - y, Ty - y \rangle = 0.
\end{cases}
\]

**Fact 4.4.2** (De Pierro and Iusem’s [63, Proposition 1]) The class of DPI-nonexpansive mappings is closed under composition.

**Corollary 4.4.3** (De Pierro and Iusem’s [63, Propositions 8 and 9]) Projections are DPI-nonexpansive. Consequently, each \( P_i \) and each \( Q_i \) is DPI-nonexpansive.

**Proof.** Combine Proposition 2.1.2.(iii) and the preceding fact. \( \square \)

**Fact 4.4.4** (De Pierro and Iusem’s [63, Proof of Lemma 1]) Suppose \( T \) is a DPI-nonexpansive selfmapping of \( H \). If \( \inf_x \|x - Tx\| \) is attained, then it is \( 0 \).

**Corollary 4.4.5** If the minimal displacement \( \mu \) is positive, it is unattained.

**Remark 4.4.6** De Pierro and Iusem formulated their results in Euclidean space — their proofs, nevertheless, work in general Hilbert space.
4.5 Strongly attracting mappings

Strongly attracting mappings are a useful tool for studying projection algorithms; see [8, Section 2].

**Definition 4.5.1** ([8]) A nonexpansive self-mapping \( T \) of \( H \) with \( \text{Fix } T \neq \emptyset \) is called strongly attracting or \( \kappa \)-attracting, if there exists \( \kappa > 0 \) such that

\[
\kappa \| x - T x \|^2 \leq \| x - y \|^2 - \| T x - y \|^2 , \quad \text{for all } x \in H, \ y \in \text{Fix } T.
\]

**Fact 4.5.2**

(i) ([8, Corollary 2.5]) Projections are 1-attracting. In particular, each \( P_i \) is 1-attracting.

(ii) ([8, Proposition 2.10.(ii)]) If \( T_1, \ldots, T_N \) are strongly attracting with constants \( \kappa_1, \ldots, \kappa_N \) and if \( \bigcap_{i=1}^N \text{Fix } T_i \neq \emptyset \), then the composition \( T_N \cdots T_1 \) is \( \min\{\kappa_1, \ldots, \kappa_N\}/2^{N-1} \)-attracting. Consequently, if \( \bigcap_{i=1}^N C_i \neq \emptyset \), then each \( Q_i \) is \( 2^{-1}(N-1) \)-attracting.

5 Main results

5.1 Geometry of the problem

**Definition 5.1.1** Suppose each \( F_i \) is nonempty. Fix any \( f_i \in F_i \) and recursively define \( f_{i+1} := P_{i+1} f_i \) so that \( f_i \in F_i \), for each \( i \). We refer to the \( N \)-tuple \( (f_1, \ldots, f_N) \) as a cycle. The \( i \)-th difference vector \( d_i \) is defined by \( d_i := f_{i+1} - f_i \), for each \( i \).

**Theorem 5.1.2** Suppose each \( F_i \) is nonempty. Then the difference vectors are well-defined, i.e. they do not depend on the choice of \( f_i \). They satisfy \( \langle C_{i+1} - F_{i+1}, -d_i \rangle \leq 0, \langle F_{i+1} - F_{i+1}, d_i \rangle = 0 \), for each \( i \) and \( d_1 + d_2 + \cdots + d_N = 0 \). The restriction \( F_{i+1} \mid F_i \) is a bijection between \( F_i \) and \( F_{i+1} \): it is given by \( d_i + I \) and the sets \( F_i \) are all translates of each other. Finally, \( P_{F_{i+1}} = d_i + P_{F_i} \).

**Proof.** Suppose the difference vectors \( d_i \) arise from a cycle \( (f_i) \). Let \( (g_i) \) be another cycle. Then Corollary 2.1.3 implies \( d_i = g_{i+1} - g_i \) (the difference vectors are thus well-defined). Kolmogorov’s criterion yields \( \langle C_{i+1} - F_{i+1}, -d_i \rangle \leq 0 \), hence \( \langle F_{i+1} - F_{i+1}, d_i \rangle = 0 \). The rest follows readily. \( \square \)

**Remark 5.1.3** Youla and Velasco already showed that the difference vectors are well-defined; see [74, Theorem 4].
The two set case

In this situation, the geometry is well-understood:

**Fact 5.1.4** Suppose \( N = 2 \). Then

(i) (Cheney and Goldstein’s [22, Theorem 2]) \( F_1 = \{ c_1 \in C_1 : d(c_1, C_2) = d(C_1, C_2) \} \), \( F_2 = \{ c_2 \in C_2 : d(c_2, C_1) = d(C_2, C_1) \} \).

(ii) ([7, Lemmata 2.1 and 2.3]) \( d_1 = \frac{d_C - c_1}{d_C} \), \( d_2 = \frac{d_C - c_2}{d_C} = -d_1 \), \( F_1 = C_1 \cap (C_2 + d_1) \), \( F_2 = C_2 \cap (C_1 + d_1) \).

**Remarks 5.1.5**

- Goldburg and Marks rediscovered (i); see [41, Theorem 3]. Gubin et al. stated (i) without proof; see [44, Note 1 to Theorem 2].
- It is important to note that (ii) gives a description of the difference vectors which is independent of \( F_1 \) and \( F_2 \). This alternative representation is the key to the understanding of the cyclic projection algorithm (and also of Dykstra’s algorithm) for two sets.

The consistent case

Here Theorem 5.1.2 reduces essentially to Proposition 4.3.10: if \( \bigcap_{i=1}^{N} C_i \neq \emptyset \), then \( F_1 = \cdots = F_N = \bigcap_{i=1}^{N} C_i \) and \( d_1 = \cdots = d_N = 0 \).

The general case again

The general case is likely to be much more complicated: consider, for instance, in \( \mathbb{R}^2 \) the case when \( N = 3 \) and the sets \( C_i \) are the sides of an equilateral triangle. Then each \( F_i \) is a singleton. If \( (f_1, f_2, f_3) \) denotes the unique cycle, then \( \| f_2 - f_1 \| + \| f_3 - f_2 \| + \| f_1 - f_3 \| \) is not the infimal length of a closed polygonal path with a single vertex in each set (and other similar measures) — in stark distinction from the two set case (Fact 5.1.4(i))! A similar example was given by Kosmol [52].

We conclude this subsection with a conjecture which we saw holds true for the two set and for the consistent case:

**Conjecture 5.1.6** (geometry conjecture) The difference vectors can be described without referring to the fixed point sets \( F_i \) and thus they exist even when each \( F_i \) is empty. Moreover: \( d_1 + \cdots + d_N = 0 \) and

\[
F_N = C_N \cap (C_{N-1} + d_{N-1}) \cap \cdots \cap (C_1 + d_1 + \cdots + d_{N-1});
\]

corresponding formulae hold for \( F_1, \ldots, F_{N-1} \).

**Remark 5.1.7** By Theorem 5.1.2, we always have \( F_N \subseteq C_N \cap (C_{N-1} + d_{N-1}) \cap \cdots \cap (C_1 + d_1 + \cdots + d_{N-1}) \).
5.2 Dichotomy

**Theorem 5.2.1** Exactly one of the following alternatives holds:

(i) Each $F_i$ is empty. Then $\lim_{n} \|x_i^n\| = +\infty$, for all $i$.

(ii) Each $F_i$ is nonempty. Then the $N$-tuple of sequence $(x_1^n, x_2^n, \ldots, x_N^n)$ converges weakly to some cycle $(f_1, f_2, \ldots, f_N)$ and each $f_i$ is an element of $Q_{F_i}(x_i^m)$, for all $m$. The sequence of differences $(x_i^{n+1} - x_i^n)$ converges in norm to the $i$th difference vector $d_i$, for each $i$.

**Proof.** (i) follows from Corollary 4.3.5, Fact 4.3.7(ii), and Proposition 4.3.10(iii).

(ii): If (i) doesn’t hold, then each $F_i$ is nonempty, by Proposition 4.3.10(iii).

Comparison of the orbit $(x_i^n)$ with the orbit of any cycle yields $x_i^{n+1} - x_i^n - d_i \to 0$, for all $i$ (Corollary 2.1.3). Also, by Fact 4.3.7(i), $x_i^n \to f_i$ for some $f_i \in F_i$, for all $i$. Hence $f_i^{n+1} = f_i + d_i$, for each $i$ and thus $(f_i)^N_{i=1}$ is identified as a cycle. Finally, $(x_i^n)_{n \geq m}$ is Fejér monotone w.r.t. $F_i$ (Corollary 4.1.5) and converges weakly to $f_i \in F_i$; thus $f_i \in Q_{F_i}(x_i^m)$ by Proposition 3.8.2(i).

The two set case

**Fact 5.2.2** ([7, Theorem 4.8]) Suppose $N = 2$. Then

$$x_2^n - x_1^n \to d_1 = P_{C_2 - C_1}(0), \quad x_1^{n+1} - x_2^n \to d_2 = P_{C_1 - C_2}(0) = -d_1.$$ 

Moreover, exactly one of the following alternatives holds:

(i) $F_1$ and $F_2$ are empty. Then $\|x_1^n\|, \|x_2^n\| \to +\infty$.

(ii) $F_1$ and $F_2$ are nonempty. Then $x_1^n \to f_1$ and $x_2^n \to f_2$, where $f_1 \in F_1$ and $f_2 = f_1 + d_1 \in F_2$.

**The consistent case**

Here Theorem 5.2.1.(ii) results in the following:

**Theorem 5.2.3** Suppose $\bigcap_{i=1}^N C_i \neq \emptyset$. Then the entire orbit $(x_i^n)$ converges weakly to some point in $\bigcap_{i=1}^N C_i$. Also, $x_{i+1} - x_i^n \to 0$ which implies $d(x_i, C_j) \to 0$, for all $i, j$.

**Remark 5.2.4** The weak convergence of $(x_i^n)$ was already known to Bregman ([15, Theorem 1]).
The general case again

Nurtured by the results for the two set case and the consistent case, we formulate two conjectures:

**Conjecture 5.2.5** (convergent differences conjecture) The sequence of differences $(x_{i+1}^n - x_i^n)$ is norm convergent for all starting points and every $i$.

**Remark 5.2.6** If the convergent differences conjecture is true, then it follows that $\lim_n x_{i+1}^n - x_i^n$ is independent of the starting point (Corollary 2.1.3).

The second conjecture is formally weaker than the first one:

**Conjecture 5.2.7** (bounded differences conjecture) The sequences of differences $(x_{i+1}^n - x_i^n)$ are bounded for all starting points and every $i$.

We would like to mention the following long-standing open problem:

**Open Problem 5.2.8** Even when $N = 2$ and $C_1 \cap C_2 \neq \emptyset$, it is still not known whether the convergence of the entire orbit $(x_i^n)$ to some point in $C_1 \cap C_2$ can actually be only weak.

### 5.3 Minimal displacement

**Theorem 5.3.1**

(i) $\mu = 0$ whenever $\bigcap_{i=1}^N \mathrm{rec}(C_i)$ is a subspace.

(ii) $\mu = 0$ and is attained if and only if there exists a bounded closed convex nonempty subset of $H$ that is left invariant by some $Q_j$. In particular, this holds whenever some $C_j$ is bounded.

(iii) $\mu = 0$ and is attained whenever $\bigcap_{i=1}^N \mathrm{rec}(C_i)$ is a subspace and there exists some $j$ such that the “normalization” $C_j := \left\{ c_j/\|c_j\| : c_j \in C_j \setminus \{0\} \right\}$ of $C_j$ is relatively compact. The latter “normalization” condition holds whenever $H$ is finite-dimensional.

**Proof.** (i): By Proposition 3.1.3, $\bigcap_{i=1}^N \mathrm{rec}(C_i)$ is a subspace if and only if $\bigcap_{i=1}^N \mathrm{rec}(C_i) \cap \left( \bigcap_{i=1}^N \mathrm{rec}(C_i) \right) = \{0\}$. (ii) is now obvious from Theorem 4.3.8. (ii): For the “only if” part, simply choose $f_j \in F_j$ so that $\{f_j\}$ is left invariant by $Q_j$. The “if” and “in particular” parts follow from Corollary 4.1.2 and Fact 4.1.6.

(iii): Suppose to the contrary $\mu$ is not simultaneously 0 and attained. In view of Corollary 4.4.5 and Theorem 5.2.1, each $Q_j$ is fixed point free and all orbits $(x_i^n)$ tend in norm to infinity. We may also assume (after relabeling, if necessary) that $j = N$. Fix an orbit $(x_i^n)$ and extract a subsequence $(x_{n_k}^n)$ of $(n)$ such that

$$
\|x_{n_k+1}^n\| = \|Q_N x_{n_k}^n\| > \|x_{n_k}^n\| =: r_k, \quad \text{for all } k.
$$

26
Then \((r_k)\) is strictly increasing and tends to infinity. Let us abbreviate
\[
y_k^0 := x_N^n \quad \text{and} \quad y_k^i := P_i y_{k-1}^i, \quad \text{for all} \ k \ \text{and} \ i.
\]
Now choose a positive integer \(m\) so large that \(y_k^n \neq y_k^m\), for all \(n \geq m\) and every \(i\) (If this weren't possible, then we would conclude \(y_N^n = y_N^m\) infinitely often, which would be absurd because \(\|y_N^n\| > r_n \to +\infty\)).

**Claim:** \(\lim_n \|y_i^n - y_i^m\|/\|y_i^n - y_i^m\| = 1\), for all \(i\).

Since all of these quotients are at most 1, it suffices to show that \(\lim_n \|y_N^n - y_N^m\|/\|y_0^n - y_0^m\| = 1\). This, however, follows from
\[
1 \geq \frac{\|y_N^n - y_N^m\|}{\|y_0^n - y_0^m\|} \geq \frac{\|y_N^n\| - \|y_N^m\|}{\|y_0^n\| + \|y_0^m\|} > \frac{r_n - \|y_N^n\|}{r_n + \|y_0^n\|} \to 1;
\]
the claim is thus verified.

Since \(\|y_0^n\| = r_n \to +\infty\), we reach the following:

**Conclusion:** \(\lim_n \|y_i^n\| = +\infty\), for all \(i\).

After passing to another subsequence if necessary, we can also assume that the sequence \((\|y_i^n - y_i^m\|/\|y_i^n - y_i^m\|)\) converges weakly to some point \(q_i\), for each \(i\).

The Claim and Proposition 2.1.2.(iii) imply
\[
\frac{y_i^n - y_i^m}{\|y_i^n - y_i^m\|} \to 0, \quad \text{for all} \ i.
\]
Hence the weak limits \(q_i\) all coincide, say \(q := q_1 = \cdots = q_N\). Next, by the Conclusion and Proposition 3.2.2,
\[
q_i = \text{weak lim}_n \frac{y_i^n - y_i^m}{\|y_i^n - y_i^m\|} = \text{weak lim}_n \frac{y_i^n}{\|y_i^n\|} \in \text{rec}(C_i), \quad \text{for all} \ i.
\]
Thus
\[
q \in \bigcap_{i=1}^N \text{rec}(C_i).
\]
Now \(C_N\) is relatively compact (by hypothesis), hence we can assume that \((y_N^n/\|y_N^n\|)\) converges in norm to \(q\) (after passing to yet another subsequence, if necessary) which implies
\[
\|q\| = 1.
\]
On the other hand, for the orbit \((x_i^n)\), we certainly have (Fact 3.2.3.(ii) and Proposition 3.1.2.(ii)) \(x_i^{n+1} - x_i^n \in \sum_{i=1}^N \text{range}(P_i - I) \subseteq (\bigcap_{i=1}^N \text{rec}(C_i))^{\perp}\), for all \(n\). Hence \(x_N^{n+1} - x_N^n \in (\bigcap_{i=1}^N \text{rec}(C_i))^{\perp}\), for all positive integers \(n, l\).

Remembering the special structure of \((y_i^n)\), we conclude \((y_N^n - y_N^m)/\|y_N^n - y_N^m\| \in (\bigcap_{i=1}^N \text{rec}(C_i))^{\perp}\), for all \(n\), which yields
\[
q \in (\bigcap_{i=1}^N \text{rec}(C_i))^{\perp}.
\]
Altogether, we obtain
\[ 0 \neq q \in \left( \bigcap_{i=1}^{N} \text{rec}(C_i) \right) \cap \left( \bigcap_{i=1}^{N} \text{rec}(C_i) \right)^{\oplus}. \]

Finally, \( \bigcap_{i=1}^{N} \text{rec}(C_i) \) is a subspace so that \( \left( \bigcap_{i=1}^{N} \text{rec}(C_i) \right) \cap \left( \bigcap_{i=1}^{N} \text{rec}(C_i) \right)^{\oplus} = \{ 0 \} \) (Proposition 3.1.3). Therefore, \( q \) has to be 0 which is the desired contradiction. \( \square \)

**Corollary 5.3.2** If each \( C_i \) is a closed affine subspace, then \( \mu = 0 \).

**Remark 5.3.3** Theorem 5.3.1.(iii) is sharp in the following sense:

- Kosmol and Zhou [54, Theorem] and Bauschke and Borwein [7, Example 4.3] constructed examples where \( N = 2 \) and each \( C_i \) is a closed affine subspace but \( \mu = 0 \) and is unattained. Hence the preceding corollary cannot be improved and the assumption of relative compactness in Theorem 5.3.1.(iii) is important.

- The following simple example shows that the assumption on the recession cones in Theorem 5.3.1.(iii) is important: suppose \( H = \mathbb{R}^2 \), \( N = 2 \), and let \( C_1 = \{(r,0) : r \geq 0\} \), \( C_2 = \{(x,y) : y \geq 1/x > 0\} \). Then \( \text{rec}(C_1) \cap \text{rec}(C_2) = C_1 \) is not linear, \( \mu = 0 \) (see the subsection on the two set case below) and unattained.

**Remarks 5.3.4** Suppose \( C \) is a closed convex nonempty subset of \( H \) and let \( \hat{C} := \{c/\|c\| : c \in C \setminus \{0\}\} \).

- \( \hat{C} \) need not be closed: consider in \( \mathbb{R}^2 \) the epigraph of the function \( 1/x \) restricted to the positive reals, i.e. \( C = \{(x,y) \in \mathbb{R}^2 : y \geq 1/x > 0\} \). Then \( \hat{C} = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 = 1; x > 0; y > 0\} \) is not closed.

- It is easy to check that if \( \hat{C} \) is relatively compact, then \( C \) is boundedly compact.

- In general, if \( C \) is boundedly compact, then \( \hat{C} \) need not be relatively compact: consider in \( H = \ell_2 \) the compact convex nonempty set \( \hat{C} := \text{conv}\{e^n/n : n \geq 1\} \), where \( e^n \) denotes the \( n^{\text{th}} \) unit vector. Then \( \hat{C} \) contains \( \{e^1,e^2,\ldots\} \) and thus cannot be relatively compact.

**The two set case**

**Theorem 5.3.5** If \( N = 2 \), then \( \mu = 0 \).

**Proof.** By Fact 5.2.2, \( x^n_2 - x^n_1 \to d_1 \) and \( x^{n+1}_1 - x^n_2 \to d_2 = -d_1 \). Adding both sequences yields \( x^{n+1}_1 - x^n_1 = Q_1 x^n_2 - x^n_2 \to 0 \) and therefore \( \mu = \inf_x \|Q_1 x - x\| = 0 \). \( \square \)
The consistent case

Theorem 5.3.1 does not provide us with new insight: if \( \bigcap_{i=1}^{N} C_i \neq \emptyset \), then \( \mu = 0 \) and is attained.

The general case again

In view of the results of this subsection, the following conjecture is very plausible (and true for the two set and the consistent cases):

**Conjecture 5.3.6** (zero displacement conjecture) The minimal displacement \( \mu \) always equals 0.

**Theorem 5.3.7** The bounded differences conjecture is stronger than the zero displacement conjecture

**Proof.** Suppose the bounded differences conjecture is true. Comparison of the orbits generated by \( x^n \) and \( x^n_N \) via Corollary 2.1.3 yields \( \lim_n \langle x^n_i - x^n_{i+1}, x^n_{i+1} - x^n_{i+1} \rangle = 0 \), for all \( i \). Hence, by assumption and Theorem 4.3.8, \( \lim_n \langle x^n_i - x^n_{i+1}, -v \rangle = 0 \), for all \( i \). Addition yields \( 0 = \lim_n \langle x^n_i - x^n_{i+1}, -v \rangle = \langle -v, -v \rangle = \|v\|^2 \); therefore, \( \mu = \|v\| = 0 \) and the proof is complete. \( \square \)

5.4 Convergence results I

**Theorem 5.4.1**

(weak convergence) If some set \( C_j \) is bounded, then \( (x^n, \ldots, x^n_N) \) converges weakly to some cycle.

(norm convergence) \( (x^n, \ldots, x^n_N) \) converges in norm to some cycle whenever (at least) one of the following conditions holds:

(i) Each \( F_i \) is nonempty and some \( C_j \) is boundedly compact.

(ii) Some \( C_i \) is bounded and some \( C_j \) is boundedly compact.

(iii) Some \( C_j \) is compact.

(iv) Some \( C_j \) is bounded and \( H \) is finite-dimensional.

(v) Some normalization \( \hat{C}_j = \{c_j/\|c_j\| : c_j \in C_j \setminus \{0\} \} \) is relatively compact and \( \bigcap_{i=1}^N \text{rec}(C_i) \) is a subspace.

(vi) \( H \) is finite-dimensional and \( \bigcap_{i=1}^N \text{rec}(C_i) \) is a subspace.

In all situations, \( \lim_n x^n_{i+1} - x^n_i = d_i \), for each \( i \).
Proof. (weak convergence): Combine Theorem 5.3.1.(ii) with Theorem 5.2.1.(ii). (norm convergence). (i): On the one hand, by Theorem 5.2.1.(ii), \((x_j^n)\) is weakly convergent to some \(f_j \in F_j\). Since \((x_j^n)\) is bounded and \(C_j\) is boundedly compact, \(f_j\) is actually a norm cluster point of \((x_j^n)\). On the other hand, \((x_j^n)\) is Fejér monotone w.r.t. \(F_j\) (Corollary 4.1.5). Thus, by Proposition 3.8.2.(ii), \((x_j^n)\) converges in norm to \(f_j\). Therefore, by continuity of each \(P_i\), each \((x_i^n)\) is norm convergent. The rest follows since: \((iv) \Rightarrow (iii) \Rightarrow (ii) \Rightarrow (i) \Leftarrow (v) \Leftarrow (vi)\). (For “(v) \Rightarrow (i)”, use Theorem 5.3.1.(iii) and Remarks 5.3.4.) “In all situations” is nothing but Theorem 5.2.1.(ii).

Remarks 5.4.2

- A slightly less general form of the weak convergence part of the last theorem is Gubin et al.’s [44, weak convergence part of Theorem 2]; they also observed (iv).
- The two set version of Theorem 5.4.1.(i) and (iii) is due to Cheney and Goldstein; see [22, Theorem 4]
- Consider the second example in the remark following Theorem 5.3.1. Since \(F_1\) and \(F_2\) are empty, the sequences \((x_1^n)\) and \((x_2^n)\) tend to infinity. Therefore, the assumption on the recession cones in Theorem 5.4.1.(v) is important.

The two set case

In this case, there is another sufficient condition for weak convergence:

Theorem 5.4.3 If \(C_2 - C_1\) is closed, then \((x_1^n, x_2^n)\) converges weakly to some cycle.

Proof. By Fact 5.1.4.(ii), \(F_1\) and \(F_2\) are nonempty. The result now follows from Theorem 5.2.1.(ii).

5.5 Convergence results II

Theorem 5.5.1 Suppose each \(C_i\), except possibly one, is Kadec/Klee and each \(F_i\) is nonempty. Then \((x_1^n, \ldots, x_N^n)\) converges in norm to some cycle.

Proof. By Theorem 5.2.1.(ii), \((x_1^n, \ldots, x_N^n)\) converges weakly to some cycle \((f_1, \ldots, f_N)\). Case 1: \(\bigcap_{i=1}^N C_i = \emptyset\). Then there is some \(i\) such that \(f_{i+1} = P_{i+1} f_i \neq f_i\) and \(C_{i+1}\) is Kadec/Klee. Hence \(f_{i+1} \in \text{bd}(C_{i+1})\). Since \(x_i^n \rightarrow f_{i+1}\), we conclude \(x_i^n \rightarrow f_{i+1}\) and thus \((x_1^n, \ldots, x_N^n) \rightarrow (f_1, \ldots, f_N)\). Case 2: \(\bigcap_{i=1}^N C_i \neq \emptyset\). Then \((C_1, \ldots, C_N)\) is boundedly regular (Theorem 3.6.5.(iii)) and the result follows from Fact 5.5.5.(iii) below.
Corollary 5.5.2 (Gubin et al.'s [44, part of Theorem 2]) Suppose each $C_i$, except possibly one, is uniformly convex and some $C_j$ is bounded. Then $(x_n^i, \ldots, x_N^i)$ converges in norm to some cycle.

Proof. Clear from Proposition 3.5.2, Theorem 5.4.1, and Theorem 5.5.1. □

The two set case

Fact 5.5.3 ([9, Theorem 3.7 and Corollary 3.10]) Suppose $N = 2$. If $(C_1, C_2)$ is boundedly regular, then $(x_n^1, x_n^2)$ converges in norm to some cycle. In particular, this happens whenever $C_1$ or $C_2$ is boundedly compact and each $F_i$ is nonempty.

Correction 5.5.4 The first two authors' [9, Theorem 3.7 and Corollary 3.10] lack the assumption of nonemptiness of each $F_i$; see Correction 3.6.3.

The consistent case

Our observations on (linear) regularity in Subsection 3.6 combined with convergence results in [8, Section 5] yield the following two facts:

Fact 5.5.5 ([8, Theorem 5.2]) Suppose $\bigcap_i C_i \neq \emptyset$ and $(C_1, \ldots, C_N)$ is boundedly regular. Then the entire orbit $(x_n^i)$ converges in norm to some point in $\bigcap_i C_i$. In particular, this happens if (i) some $C_i$ is boundedly compact; (ii) $H$ is finite-dimensional; (iii) every $C_i$, except possibly one, is Kadec/Klee; or (iv) every $C_i$, except possibly one, is uniformly convex.

Remark 5.5.6 Fact 5.5.5.(i),(ii) give a different explanation for the consistent case of Theorem 5.4.1.(i). Fact 5.5.5.(iv) is due to Gubin et al.; see [44, Theorem 1.(b)].

Fact 5.5.7 ([8, Theorem 5.7]) Suppose $\bigcap_i C_i \neq \emptyset$ and $(C_1, \ldots, C_N)$ is boundedly linearly regular. Then: (a) the entire orbit $(x_n^i)$ converges linearly to some point in $\bigcap_i C_i$; moreover, (b) the rate of convergence is independent of the starting point whenever $(C_1, \ldots, C_N)$ is linearly regular. In particular, (a) happens whenever (i) $0 \in \bigcap_{i=1}^N \text{int}(\bigcap_{j=1}^i C_j) - C_{i+1}$ or (ii) $C_N \cap \text{int}(\bigcap_{i=1}^{N-1} C_i) \neq \emptyset$.

Remarks 5.5.8 Fact 5.5.7.(a),(ii) is also due to Gubin et al.; see [44, Theorem 1.(a)]. Corresponding formulations of Fact 5.5.7.(a),(i),(ii) with different orders hold, too; see Remark 3.6.10. We discuss some more special cases of Fact 5.5.7 in Subsections 5.6 and 5.7.

Fact 5.5.9 ([9, Theorem 5.3.(iv)]) Suppose $N = 2$ and $H$ is a Hilbert lattice with lattice cone $H^+$. If $C_1$ is a hyperplane, $C_2 = H^+$, and $C_1 \cap C_2 \neq \emptyset$, then the entire orbit $(x_n^i)$ converges in norm to some point in $C_1 \cap C_2$.

Remarks 5.5.10
• Classical examples are $H = t_2$ with $H^+ = \{ (s_n) \text{ in } H : s_n \geq 0, \text{ for all } n \}$ and $H = L_2[0,1]$ with $H^+ = \{ f \in H : f \geq 0 \text{ almost everywhere} \}$.

• Fact 5.5.9 relies heavily on the order structure of $H$. It would be interesting to know whether this result holds when $C_i$ is a closed affine finite-codimensional subspace instead of a hyperplane.

**The inconsistent case**

**Fact 5.5.11** (Gubin et al.’s [44, part of Theorem 2]) Suppose each $C_i$, except possibly one, is strongly convex, some $C_j$ is bounded, and $\bigcap_{i=1}^{N} C_i = \emptyset$. Then $(x^n_1, \ldots, x^n_N)$ converges linearly to some cycle.

**Remark 5.5.12** The simple example of two discs in $\mathbb{R}^2$ with precisely one common point shows that the assumption of emptiness of the intersection is important.

### 5.6 The convex polyhedral case

**Theorem 5.6.1** If each $C_i$ is a convex polyhedron, then each $F_i$ is nonempty and the $N$-tuple of sequences $(x_1^n, \ldots, x_N^n)$ converges in norm to some cycle.

**Proof.** Corollary 3.4.8 in conjunction with Theorem 5.2.1.(ii) implies the weak convergence of $(x_1^n, \ldots, x_N^n)$ to some cycle. Now suppose that each $C_i$ is given as $\{ x \in H : \langle a_{i,j}, x \rangle \leq b_{i,j} \}$. Let $K := \bigcap_{i,j} \text{kernel}(a_{i,j})$ and let $D_i := C_i \cap K^\perp$, for each $i$. Then, by Proposition 3.3.2,

$$x_{i+1}^n - x_i^n = P_K x_i^n + P_{D_{i+1}} K^\perp x_i^n - x_i^n = -P_K x_i^n + P_{D_{i+1}} K^\perp x_i^n \in K^\perp,$$

for all $n$ and each $i$. Telescoping yields

$$x_i^n \in x_0^0 + K^\perp,$$

hence the entire orbit $(x_i^n)$ lies in an affine finite-dimensional subspace. Here, however, weak and norm convergence coincide and the result follows. \hfill $\square$

**Remark 5.6.2** Theorem 5.6.1 is related to a result by De Pierro and Iusem [63, Proposition 13 and Theorem 1] which states that $(x_N^n)$ converges in norm to some point in $F_N$ when $H$ is finite-dimensional and each $C_i$ is a halfspace: their algorithm, however, is more general.

**The consistent case**

Combining Fact 3.6.8 with Fact 5.5.7.(b) yields:
Theorem 5.6.3 Suppose each $C_i$ is a convex polyhedron and $\bigcap_{i=1}^{N} C_i \neq \emptyset$. Then the entire orbit $(x^k)$ converges linearly to some point in $\bigcap_{i=1}^{N} C_i$ with a rate independent of the starting point.

Remark 5.6.4 In the setting of the problem, special cases of this theorem were obtained by Gubin et al. ([44, Theorem 1.(d)]) and by Herman et al. ([46, Theorem 1]); see also Mandel’s [57, Theorem 3.1] for an upper bound for the rates of convergence and [8, Section 6] for more general algorithms.

The general case again

Encouraged by the last fact, we believe the following is true:

Conjecture 5.6.5 If each $C_i$ is a convex polyhedron, then $(x^k, \ldots, x^N)$ converges linearly to some cycle with a rate independent of the starting point.

5.7 The affine subspace case

Throughout this subsection, we assume that

\[ C_i = c_i + L_i, \ldots, C_N = c_N + L_N \]

are closed affine subspaces,

where, without loss of generality,

\[ L_i \text{ is linear, } L := \bigcap_{i=1}^{N} L_i, \text{ and } c_i \in L_i \] for each $i$.

We investigate the behaviour of the method of cyclic projections in this context and restrict ourselves to the subsequence $(y^k)$ of an arbitrary but fixed orbit. Proposition 2.2.2.(i) yields $P_i x = c_i + P_i x$, for each $i$ and all $x$; thus if we let $c := (c_1, \ldots, c_N)$ and define the operator

\[ T : L_1^\perp \times \cdots \times L_N^\perp \rightarrow H : y = (y_1, \ldots, y_N) \mapsto \sum_{i=1}^{N} P_{L_i} \cdots P_{L_{i+1}} y_i; \]

it then follows that $Q_N x = P_{L_N} \cdots P_{L_i} x + T c$, and hence inductively

\[ x_N^n = Q_N^n x^0 = (P_{L_N} \cdots P_{L_i})^n x^0 + \sum_{k=0}^{n-1} (P_{L_N} \cdots P_{L_i})^k (T c), \text{ for all } n \geq 1. \]

In passing, we note that the first term on the right hand side of the last equation, $(P_{L_N} \cdots P_{L_i})^n x^0$, tends to $P_L x^0$ by the Fact 1.1.1.

Proposition 5.7.1

(i) $L^\perp$ is an invariant subspace of $P_L$ and $P_{L_N} \cdots P_{L_i}$, for each $i$. 

33
(ii) $\mathbf{T}c \in L^\perp$.

(iii) $(P_{L_N} \cdots P_{L_1})^k = P_L \oplus (P_{L_N} \cdots P_{L_1} P_{L_L})^k$, for all $k \geq 1$.

(iv) $P_L = P_L P_{L_i}$, for each $i$.

**Proof.** (i): Obviously, $L^\perp$ is an invariant subspace of $P_L$. Now fix $l^\perp \in L^\perp$ and $l \in L$. Then $\langle l, P_{L,N} \cdots P_{L_1} l^\perp \rangle = \langle (P_{L_N} \cdots P_{L_1})^n l^\perp, l^\perp \rangle = \langle P_L, l^\perp \rangle = 0$, which implies that $L^\perp$ is an invariant subspace of $P_{L,N} \cdots P_{L_1}$.

(ii): Since $c_i \in L_i^\perp \subseteq L^\perp$, for all $i$, (i) and the definition of $\mathbf{T}$ imply $\mathbf{T}c \in L^\perp$.

(iii) is proved by induction on $k$: $P_{L,N} \cdots P_{L_1} = P_{L,N} \cdots P_{L_1} (P_L \oplus P_{L_L}) = P_L \oplus P_{L,N} \cdots P_{L_1} P_{L_L}$; hence (iii) holds for $k = 1$. If it holds for some $k \geq 1$, then $(P_{L,N} \cdots P_{L_1})^{k+1} = (P_{L,N} \cdots P_{L_1}) (P_L \oplus (P_{L,N} \cdots P_{L_1} P_{L_L})^k) = P_L \oplus (P_{L,N} \cdots P_{L_1} P_{L_L})^{k+1}$, because $L^\perp$ is an invariant subspace of $P_{L,N} \cdots P_{L_1}$ by (i).

(iv): Fix $x \in H$ and $l \in L$. Then $\langle l - P_L P_{L_1} x, x - P_L P_{L_1} x \rangle = \langle l - P_L P_{L_1} x, P_L x - P_L P_{L_1} x \rangle + \langle l - P_L P_{L_1} x, P_{L_1}^\perp x \rangle = 0$; since this is true for all $l \in L$, we conclude (Proposition 2.1.2(ii)) that $P_L x = P_L P_{L_1} x$ and the result follows. \qed

**Corollary 5.7.2** The sequence $(x_n)$ satisfies

$$x_n = P_L x^0 + (P_{L,N} \cdots P_{L_1} P_{L_L})^n x^0 + \sum_{k=0}^{n-1} (P_{L,N} \cdots P_{L_1} P_{L_L})^k (\mathbf{T}c), \text{ for all } n \geq 1.$$ 

This observation is useful, since the operator norm of $P_{L,N} \cdots P_{L_1} P_{L_L}$ is less than or equal to the operator norm of $P_{L_N} \cdots P_{L_1}$. Moreover, $\|P_{L,N} \cdots P_{L_1} P_{L_L}\|$ is nothing but the angle of the $N$-tuple $(L_1, \ldots, L_N)$; see Definition 3.7.3.

The next proposition provides information on the set of fixed points $F_N$:

**Proposition 5.7.3**

(i) $F_N = (I - P_{L,N} \cdots P_{L_1})^{-1}(\mathbf{T}c)$.

(ii) (Kosmol and Zhou’s [54, Proposition 4]) If $F_N$ is nonempty, then $F_N = f_N^* + L$, for all $f_N^* \in F_N$.

**Proof.** (i): $x \in F_N \iff x - Q_N x = 0 \iff x - (P_{L_N} \cdots P_{L_1}) x = \mathbf{T}c$.

(ii): Pick $f_N, f_N^* \in F_N$. Then, using (i), the linearity of $P_{L,N} \cdots P_{L_1}$, and Corollary 4.3.5, we get: $f_N - P_{L_N} \cdots P_{L_1} f_N = \mathbf{T}c = f_N^* - P_{L_N} \cdots P_{L_1} f_N^* \iff f_N - f_N^* = (P_{L_N} \cdots P_{L_1}) (f_N - f_N^*) \iff f_N - f_N^* \in \text{Fix} (P_{L_N} \cdots P_{L_1}) \iff f_N \in f_N^* + L$. \qed

The following useful identity is due to Kosmol and Zhou; it is proved inductively using Proposition 5.7.3.(i).

**Fact 5.7.4** ([53, Proof of Theorem 3]) If $F_N$ is nonempty and $f_N^* \in F_N$, then

$$x_n = (P_{L,N} \cdots P_{L_1})^n (x^0 - f_N^*) + f_N^*, \text{ for all } n \geq 1.$$  

34
Theorem 5.7.5 (norm dichotomy) Exactly one of the following alternatives holds:

(i) \( F_N \) is empty. Then \( \|x^n_N\| \) tends to \(+\infty\).

(ii) \( F_N \) is nonempty. Then the sequence \( (x^n_N) \) converges in norm:

\[
\lim_n x^n_N = P_{F_N}x^0 = P_Lx^0 + \sum_{k=0}^{\infty}(P_{L_N} \cdots P_{L_1}P_{L})^k(Tc).
\]

Consequently, \( \sum_{k=0}^{\infty}(P_{L_N} \cdots P_{L_1}P_{L})^k(Tc) \) is equal to \( P_{F_N}0 \), the minimum norm fixed point of \( Q_N \).

Proof. In view of Theorem 5.2.1, we assume that \( F_N \) is nonempty and fix \( f_N \in F_N \). By Fact 5.7.4 and Fact 1.1.1, \( x^n_N \to P_L(x^0 - f_N) + f_N = P_Lx^0 + P_Lf_N \).

Also, by Proposition 2.2.2.(i) and Proposition 5.7.3.(ii), \( P_L(x^0 - f_N) + f_N = P_{f_N+L}(x^0) = P_{F_N}x^0 \) and \( P_Lf_N = f_N + P_L(0 - f_N) = P_{f_N+L}(0) = P_{F_N}0 \).

Hence, by Corollary 5.7.2, \( P_Lf_N = P_{F_N}0 = \sum_{k=0}^{\infty}(P_{L_N} \cdots P_{L_1}P_{L})^k(Tc) \).

\( \Box \)

Corollary 5.7.6 If all \( F_i \) are empty, then \( \lim_n \|x^n_i\| = +\infty \), for each \( i \).

Otherwise, all \( F_i \) are nonempty, each sequence \( (x^n_i) \) converges in norm to \( P_{F_i}x^0 \), and \( d_i = P_{L_i}(F_{i+1} - F_i) \).

Proof. In view of Theorem 5.2.1, we only have to verify the “otherwise” part. Theorem 5.7.5.(ii) implies \( x^n_N \to P_{F_N}x^0 \); if we apply the theorem once more (this time to \( (x^n_i) \)), we get \( x^n_i \to P_{F_i}x^i_1 \). Fix \( f_i \in F_i \). Then, by Proposition 5.7.3.(ii), Proposition 2.2.2.(i), and Proposition 5.7.1, \( P_{f_i}x^i_1 = P_{f_i+L}(x^0) + P_L(x^0 - f_i) = f_i + P_L(x^0 - f_i) = P_{f_i+L}(x^0) = P_{F_i}x^0 \).

Also, \( x^n_i \to P_{F_i}x^i_1 \), and, similarly to what we just did, \( P_{F_i}x^i_1 = P_{f_i+L}(x^0) = P_{F_i}x^0 \).

Analogously, \( P_{F_2}x^2_1 = P_{F_2}x^0, \ldots \), up to \( P_{F_{N-1}}x^0 \). Finally, by Theorem 5.1.2, Proposition 5.7.3.(ii), and Proposition 2.2.2.(i), \( d_i = P_{F_{i+1}}(x^0) - P_{F_i}(x^0) = P_{f_{i+1}+L}(x^0) - P_{f_i+L}(x^0) = f_{i+1} + P_{L}(x^0 - f_{i+1}) - (f_i + P_L(x^0 - f_i)) = P_{L_i}(f_{i+1} - f_i) \), for all \( f_i \in F_i \) and \( f_{i+1} \in F_{i+1} \).

\( \Box \)

Remarks 5.7.7

- Kosmol and Zhou essentially obtained the “otherwise” part of Corollary 5.7.6; see [53, Theorem 3].

- The fact that the limit of the entire orbit \( (x^n_i) \) has to be \( P_{F_i}x^0 \) (provided it exists) can be explained differently by using Fejér monotone sequences and quasi-projections. Indeed, the sequence \( (x^n_i) \) is Fejér monotone w.r.t. \( F_i \) (Corollary 4.1.5), hence its limit lies in \( Q_{F_i}x^i_1 \) (Proposition 3.8.2) which is — in this setting — the singleton \( \{P_{F_i}x^i_1\} \) (Proposition 2.3.2.(iii)). However, as we saw in the proof of Corollary 5.7.6, \( P_{F_i}x^i_1 = P_{F_i}x^0 \).
**Theorem 5.7.8** (angle and linear convergence) If \( L_1^\perp + \cdots + L_N^\perp \) is closed, then \( F_N \) is nonempty and \((x^*_N)\) converges linearly with a rate no worse that \( \gamma(L_1, \ldots, L_N) \), the angle of the \( N \)-tuple \((L_1, \ldots, L_N)\):

\[
\lim_n x^*_N = P_{F_N} x^0 = P_L x^0 + P_{F_N} 0 = P_L x^0 + (I - P_{L_N} \cdots P_{L_1} P_{L_1}^\perp)^{-1}(Tc).
\]

In particular, this holds whenever:

(i) Some \( L_i \cap L^\perp \) is finite-dimensional.

(ii) Some \( L_i \) is finite-dimensional.

(iii) \( H \) is finite-dimensional.

(iv) All \( L_i \), except possibly one, are finite-codimensional.

(v) Each \( C_i \) is a hyperplane.

(vi) Each angle \( \gamma(L_i, L_{i+1} \cap \cdots \cap L_N) \) is positive.

Consequently, each sequence \((x^*_n)\) converges linearly to \( P_L x^0 \) with a rate no worse than the angle \( \gamma(L_{i+1}, \ldots, L_N, L_i) \).

**Proof.** Proposition 3.7.7 implies that the angle \( \gamma(L_1, \ldots, L_N) \) of the \( N \)-tuple \((L_1, \ldots, L_N)\) is positive, in other words: \( \|P_{L_N} \cdots P_{L_1} P_{L_1}^\perp\| < 1 \). Hence, by Corollary 5.7.2, \((x^*_N)\) converges linearly with a rate no worse than \( \gamma(L_1, \ldots, L_N) \); its limit is \( P_L x^0 + (I - P_{L_N} \cdots P_{L_1} P_{L_1}^\perp)^{-1}(Tc) \). The main statement of the theorem now follows with Theorem 5.7.5. Conditions (i) through (vi) ensure, again by Proposition 3.7.7, \( \gamma(L_1, \ldots, L_N) > 0 \). Finally, the “consequently” part follows from Corollary 5.7.6 and the first statement of this theorem. \( \square \)

**Remarks 5.7.9**

- If \( N = 2 \) and \( L_1^\perp + L_2^\perp \) is not closed, then convergence can be “arbitrarily slow”; see Theorem 5.7.16.

- Kosmol proved a less general version of (iii); see [52, Satz 3].

We now specialize to the case when each \( C_i \) is a hyperplane:

\[
C_i = \{x \in H : \langle a_i, x \rangle = b_i \}, \quad \text{for some } a_i \in H \setminus \{0\} \text{ and } b_i \in \mathbb{R}.
\]

In the terminology of this subsection,

\[
c_i = \frac{b_i}{\|a_i\|^2} a_i, \quad L_i = \ker(a_i) = \{a_i\}^\perp, \quad L_i^\perp = \span\{a_i\}, \quad \text{for each } i.
\]
Theorem 5.7.8(v) implies that the angle \( \gamma(L_1, \ldots, L_N) \) is positive; hence \( \|P_{L_N} \cdots P_{L_1} P_{L_L} \| < 1 \) and each subsequence \( (x^n_k) \) converges linearly.

The mapping

\[
\Psi : \mathbb{R}^N \rightarrow \mathbb{R}^N : y = (y_1, \ldots, y_N) \mapsto \left( \frac{y_1}{\|a_1\|^2} a_1, \ldots, \frac{y_N}{\|a_N\|^2} a_N \right)
\]

not only defines a linear isomorphism between \( \mathbb{R}^N \) and \( L_1^+ \times \cdots \times L_N^+ \), but also allows us to think of \( T \) as being defined on \( \mathbb{R}^N \); this will yield a nice connection to generalized inverses.

**Definition 5.7.10** Suppose \( Y \) is another Hilbert space and \( A : H \rightarrow Y \) is a continuous linear operator with closed range. The **generalized inverse** of \( A \), denoted \( A^\dagger \), is the unique continuous linear operator from \( Y \) to \( H \) satisfying (i) \( A^\dagger A = I \) on \( (\text{kernel}(A))^\perp \) and (ii) \( A^\dagger = 0 \) on \( (\text{range}(A))^\perp \).

**Remark 5.7.11** Here, we have chosen the definition of the generalized inverse due to Desoer and Whalen [23] which is equivalent to better known definitions by Moore [59] and by Penrose [61]. A proof of these equivalences and more can be found in Groetsch’s monograph [43].

Now consider

\[
A : H \rightarrow \mathbb{R}^N : x \mapsto (\langle a_1, x \rangle, \ldots, \langle a_N, x \rangle).
\]

Then \( A^* : \mathbb{R}^N \rightarrow H : y = (y_1, \ldots, y_N) \mapsto \sum_{i=1}^N y_i a_i \). Because of \( (\text{kernel}(A))^\perp = \text{range}(A^*) \) and \( (\text{range}(A))^\perp = \text{kernel}(A^*) \), we have

\[
\text{span}\{a_1, \ldots, a_N\} = \text{range}(A^*) = (\text{kernel}(A))^\perp = (\bigcap_{i=1}^N \text{kernel}(a_i))^\perp = L^+_1,
\]

and

\[
(\text{range}(A))^\perp = \text{kernel}(A^*) = \{y = (y_1, \ldots, y_N) \in \mathbb{R}^N : \sum_{i=1}^N y_i a_i = 0\}.
\]

**Proposition 5.7.12** If the vectors \( a_1, \ldots, a_N \) are linearly independent, then \((I - P_{L_N} \cdots P_{L_1} P_{L_L})^{-1} T \Psi \) is the generalized inverse \( A^\dagger \) of \( A \).

**Proof.** We have to check conditions (i) and (ii) of Definition 5.7.10.

(i): Fix \( x \in L^+_1 = (\text{kernel}(A))^\perp \). Then

\[
T \Psi A x = T \Psi (\langle a_1, x \rangle, \ldots, \langle a_N, x \rangle)
= T \left( \frac{\langle a_1, x \rangle}{\|a_1\|^2} a_1, \ldots, \frac{\langle a_N, x \rangle}{\|a_N\|^2} a_N \right)
= \left( \frac{\langle a_1, x \rangle}{\|a_1\|^2}, \ldots, \frac{\langle a_N, x \rangle}{\|a_N\|^2} \right).
\]
\[ T(P_{L_1^x}, \ldots, P_{L_k^x}) \]
\[ = \sum_{i=1}^{N} (P_{L_N} \cdots P_{L_{i+1}})P_{L_i^x} \]
\[ = \sum_{i=1}^{N} (P_{L_N} \cdots P_{L_{i+1}})(x - P_Lx) \]
\[ = x - P_{L_N} \cdots P_{L_k}x \]
\[ = (I - P_{L_N} \cdots P_{L_k})A \cdot x \]

so that \((I - P_{L_N} \cdots P_{L_k})A = x\), as promised.

(ii): Since the vectors \(a_1, \ldots, a_N\) are linearly independent, \((\text{range}(A))^\perp = \{0\}\)
and hence (ii) clearly holds. \(\square\)

The following example shows that the assumption of linear independence in the last theorem is important:

**Example 5.7.13** Let \(H = \mathbb{R}, N = 2, a_1 = a_2 = 1.\) A simple calculation yields
\[
A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \quad (I - P_{L_1}P_{L_2})^{-1}T \Psi = (0, 1), \quad \text{but} \quad A^T = \left( \frac{1}{2}, \frac{1}{2} \right).
\]

**Theorem 5.7.14** Suppose \(a_1, \ldots, a_N\) are nonzero vectors in \(H\) and let \(A : H \rightarrow \mathbb{R}^N : x \mapsto (\langle a_i, x \rangle)_{i=1}^{N}\). Given any vector \(b = (b_i)_{i=1}^{N} \in \mathbb{R}^N\), let \(C_i = \{x \in H : \langle a_i, x \rangle = b_i\}\), for all \(i\). Then:

(i) The subsequence \((x_n^n)\) of the orbit with starting point \(x_0\) converges linearly to \((I - P_{L_N} \cdots P_{L_1}P_{L_2})^{-1}T \Psi b + (I - A^T A)x_0\).

(ii) (generalized inverses via cyclic projections) If furthermore the vectors \(a_1, \ldots, a_N\) are linearly independent, then the subsequence \((x_n^n)\) of the orbit with starting point \(x_0 = 0\) converges linearly to \(A^T b\).

**Proof.** (i): On the one hand, \((x_n^n)\) converges linearly to
\[ P_{L_N}x_0 + (I - P_{L_N} \cdots P_{L_1}P_{L_2})^{-1}T c = P_{L_N}x_0 + (I - P_{L_N} \cdots P_{L_1})^{-1}T \Psi b. \]

On the other hand, by Examples 2.2.1(v) and (vi), \(P_L = I - P_{L^+} = I - P_{\text{span}\{a_1, \ldots, a_N\}} = I - A^T A\). Hence (i) holds. (ii) follows from the last proposition and (i). \(\square\)

**Remarks 5.7.15**

- Related finite-dimensional versions of (i) were obtained by Censor et al. [21, Lemma 2], De Pierro and Iusem [63, Theorem 1 and Proposition 10], Eggermont et al. [32, Theorem 1.1], Tanabe [68, Corollary 9], and Trummer [69, Theorem 14].

- In view of Example 5.7.13, the assumption of linear independence in (ii) is important.

38
The two set case

We proved the two set version of Theorem 5.7.8 in [9, Theorem 4.11]. In addition, it should be noted that
\[ d_1 = P_{L_1^+ \cap L_2^+} (c_1 - c_1), \quad d_2 = P_{L_1^+ \cap L_2^+} (c_1 - c_2) = -d_1; \]
see [9, Example 2.2].
We conclude with a striking trichotomy result:

**Theorem 5.7.16 (trichotomy)** Suppose \( N = 2 \). Exactly one of the following alternatives holds:

(i) \( F_1, F_2 \) are empty. Then \( \gamma(L_1, L_2) = 0 \) is zero and \( \lim_{n \to \infty} \|x_n^m\| = +\infty \).

(ii) \( F_1, F_2 \) are nonempty and \( \gamma(L_1, L_2) = 0 \). Then for every starting point \( x^0 \in H \), the sequence \( (x_n^m) \) converges in norm to \( P_{F_2} x^0 \). Convergence is “arbitrarily slow” in the following sense: for each sequence \( (\lambda_k) \) of real numbers with \( 1 > \lambda_1 > \lambda_2 > \cdots > \lambda_{k-1} > \lambda_k \to 0 \), there exists some starting point \( x^0 \in H \) such that \( \|x_n^m - P_{F_2} x^0\| \geq \lambda_k \), for all \( k \).

(iii) \( F_1, F_2 \) are nonempty and \( \gamma(L_1, L_2) > 0 \). Then for every starting point \( x^0 \), the sequence \( (x_n^m) \) converges linearly to \( P_{F_2} x^0 \) with a rate no worse than \( \gamma(L_1, L_2) \).

**Proof.** By Theorems 5.7.5, 5.7.8, we have only to prove (ii). In view of Fact 5.7.4, we assume without loss of generality that \( C_1 = L_1, C_2 = L_2 \), so that \( F_2 = F_1 = C_1 \cap C_2 \). Set \( A := C_1 \cap (C_1 \cap C_2)^\perp, B := C_2 \cap (C_1 \cap C_2)^\perp \).

**Step 1:** Select a sequence of integers \( (s_k) \) such that \( s_k \lambda_k < 1 \leq (s_k + 1) \lambda_k \), for all \( k \).

Then \( (s_k) \) is increasing and every term \( s_k \) occurs only finitely often. Let \( (t_n) \) be the strictly increasing sequence of integers with \( \{t_1, t_2, \ldots\} = \{s_1, s_2, \ldots\} \). Note that \( \sum_n 1/t_n^2 \) is convergent.

**Step 2:** Define \( k_0(n) := \min\{k : s_k = t_n\}, k_1(n) := \max\{k : s_k = t_n\} \) so that
\[ s_{k_0(n)} - 1 = t_{n-1} < t_n = s_{k_0(n)} = s_{k_0(n)+1} = \cdots = s_{k_1(n)} < t_{n+1} = s_{k_1(n)+1}; \]
and further set
\[ \alpha_n := (\lambda_{k_0(n)} t_n) \frac{1}{k_1(n)} , \quad \text{for all } n. \]

Since \( 1 > \lambda_{k_0(n)} s_{k_0(n)} = \lambda_{k_0(n)} t_n \geq 1 - \lambda_{k_0(n)} \) and \( \lambda_k \to 0 \), we get \( \alpha_n \to 1^- \).

**Claim:** There exist sequences \( (e^*_n) \) in \( A \), \( (f^*_n) \) in \( B \) such that for all \( n, m \) with \( n \neq m \):
\[ \langle e^*_n, e^*_m \rangle = \langle e^*_n, f^*_m \rangle = \langle f^*_n, f^*_m \rangle = 0, \quad \|e^*_n\|, \|f^*_n\| \leq 1, \quad \langle e^*_n, f^*_m \rangle \geq \alpha_n. \]
By Remarks 3.7.6.(i), sup\{\langle a, b \rangle : a \in A, b \in B, \|a\|, \|b\| \leq 1 \} = 1. We construct the sequences inductively. Clearly, \( \epsilon'_1, f'_1 \) exist. Now suppose we have already found \( \epsilon_1, \ldots, \epsilon_m, f'_1, \ldots, f'_m \). Set \( E := \text{span}\{\epsilon_1, \ldots, \epsilon_m\}, F := \text{span}\{f'_1, \ldots, f'_m\} \). Pick sequences \( (a_n) \) in \( A \), \( (b_n) \) in \( B \) with \( \|a_n\| = \|b_n\| = 1, \langle a_n, b_n \rangle \to 1 \) and \( a_n - b_n \to 0 \) and hence \( a = b = 0 \). Because \( P_E, P_F \) are compact operators, we conclude \( P_Ea_n, P_Fb_n \to 0 \) and further

\[
1 = \lim_n \langle a_n, b_n \rangle = \lim_n \langle P_Ea_n + P_Fa_n, P_Fb_n + P_Fb_n \rangle = \lim_n \langle P_Ea_n, P_Fa_n \rangle.
\]

Consequently, for \( n \) sufficiently large, \( \epsilon'_m \) exist. Set \( e'_{m+1} := P_Ea_n, f'_{m+1} := P_Fb_n \). The job is done and the claim is verified.

**Step 3:** Define \( e_n := e'_n/\|e'_n\|, f_n := f'_n/\|f'_n\| \), for all \( n \) and \( E := \text{span}\{e_1, e_2, \ldots\} \), \( F := \text{span}\{f_1, f_2, \ldots\} \). Then the sequences \( (e_n), (f_n) \) have the same properties as the sequences \( (e'_n), (f'_n) \); moreover, their terms are norm one. Since \( E = A \cap (E + F), F = B \cap (E + F), (E + F)^\perp = E^\perp \cap F^\perp \), we obtain

\[
C_1 = (C_1 \cap C_2) \oplus E \oplus (A \cap E^\perp \cap F^\perp), C_2 = (C_1 \cap C_2) \oplus F \oplus (B \cap E^\perp \cap F^\perp).
\]

If \( z \in F \), then \( P_Ez = P_{C_1 \cap C_2}z + P_{E \cap F}z + P_{A \cap E^\perp \cap F^\perp}z = P_Ez \), because \( F \) is orthogonal to \( C_1 \cap C_2 \) and to \( A \cap E^\perp \cap F^\perp \). Similarly, if \( z \in F \), then \( P_Fz = P_Ez \). Thus \( Q_{2F} := P_{2P_1}_F \equiv P_{2F}P_E \).

**Step 4:** \( P_{C_1 \cap C_2} F = P_{E \cap F} \equiv 0 \). This follows from Fact 1.1.1 and Step 3: if \( z \in F \), then \( P_{C_1 \cap C_2}z = \lim_n (P_{2P_1})^nz = \lim_n (P_{2P_1})^nz = P_{E \cap F}z = P_Fz = 0 \).

**Step 5:** Set \( x^0 := \sum_n (1/t_n)f_n \). Then \( P_E x^0 = P_{C_1 \cap C_2} x^0 = 0 \) and one readily verifies that \( (P_{2P_1})^k x^0 = \sum_n (1/t_n)(e_n, f_n)^{2k} f_n \) which yields

\[
\|x^k - P_{F^k}x^0\| \geq \frac{1}{t_n} \alpha_n^{2k}, \quad \text{for all } n, k.
\]

**Last Step:** Fix an arbitrary \( k \) and get \( \bar{n} \) such that \( k_0(\bar{n}) \leq k \leq k_1(\bar{n}) \). Then

\[
\|x^k - P_{F^k}x^0\| \geq \frac{1}{t_n} \alpha_n^{2k} \geq \frac{\alpha_n^{2k_1(\bar{n})}}{t_n} \geq \lambda_{k_0(\bar{n})} \geq \lambda_k;
\]

the proof is complete. \( \square \)

**Remarks 5.7.17** For examples of case (i), see Kosmol and Zhou’s [54, Theorem] and Bauschke and Borwein’s [7, Example 4.3]. Franchetti and Light have an example of case (ii) in [36, Section 4]; we give another example below. Theorem 5.7.8 provides many examples of case (iii).
Example 5.7.18 Let $H := \ell_2$ and denote the unit vectors by $(u_n)$. Suppose $(\alpha_n)$ is a sequence of positive real numbers with $1 > \alpha_n > -1$ and $\lim_n \alpha_n = 1$. Set $e_n := u_{2n-1}$, $f_n := \alpha_n u_{2n-1} + \sqrt{1 - \alpha_n^2} u_{2n}$, $C_1 := \text{span}\{e_1, e_2, \ldots\}$, and $C_2 := \text{span}\{f_1, f_2, \ldots\}$. Then $C_1 \cap C_2 = \{0\}$ and the sum $C_1^\perp + C_2^\perp$ is dense in $H$ but not closed. By choosing $(\alpha_n)$ and $(x^0)$ appropriately, we can arrange arbitrarily slow convergence of $(x_n^n)$: see the proof of the last theorem.

The consistent case

Our journey is over: the consistent case of Theorem 5.7.5 is the von Neumann/Halperin result (Fact 1.1.1) we started with. A vast number of papers on this result have been appearing; we feel the reader is in good hands when we refer him or her once more to Deutsch's survey article [25]. Two special cases have received much attention:

- two subspaces with positive angle. Sharper upper bounds on the rate of convergence can be found in [3, 24, 50].
- intersecting hyperplanes. Here one obtains the well-known method of Kaczmarz [49] (see also [42]) for solving systems of linear equations.

For more general algorithmic schemes, see [8] and the references therein.

5.8 Related algorithms

Random projections

Suppose $\sigma$ is a random mapping for $\{1, \ldots, N\}$, i.e. a surjective mapping from $\mathbb{N}$ to $\{1, \ldots, N\}$ that assumes every value infinitely often. Then generate a random sequence $(x_n)$ by

$$x_0 \in H \text{ arbitrary, } \quad x_{n+1} := P_{\sigma(n+1)} x_n, \quad \text{for all } n \geq 0.$$  

This method of random projections is more general than the method of cyclic projections: indeed, if $\sigma(\cdot) \equiv [\cdot]$, then the random sequence $(x_n)$ and the orbit generated by $x_0$ (in the sense of Section 1) coincide.

It is not surprising that little is known on random projections for the inconsistent case with three or more sets, because the preceding material for the more restrictive case of cyclic projections has left some important questions open.

The two set case

Each projection $P_i$ is idempotent, i.e. $P_i^2 = P_i$; thus the random sequence $(x_n)$ is essentially the orbit (for cyclic projections) with starting point $x_0$.  

41
The consistent case

The random sequence \((x_n)\) converges weakly to some point in \(\bigcap_{i=1}^{N} C_i\), when each \(C_i\) is a closed subspace (Amemiya and Ando [2]), or \(N = 3\) (Dye and Reich [31]), or the sets \(C_i\) have a common "inner point" (Dye and Reich [31], Youla [73]; see also Bruck [17], Dye [28], Dye et al. [29], and Dye and Reich [30]). It is not known whether or not every random sequence converges weakly to some point in \(\bigcap_{i=1}^{N} C_i\).

Norm convergence of the random sequence \(x_n\) to some point in \(\bigcap_{i=1}^{N} C_i\) follows when some \(C_i\) is boundedly compact (Bauschke and Borwein [8, Example 4.6]; see also Aharoni and Censor [1], Bruck [17], Elsner et al. [33], Flâm and Zowe [35], and Tseng [70]) or when the \(N\)-tuple \((C_1, \ldots, C_N)\) is innately boundedly regular, i.e. \((C_j)_{j \in J}\) is boundedly regular, for every nonempty subset \(J\) of \(\{1, \ldots, N\}\) (Bauschke [6]).

All of the above authors have established more general results.

Weighted projections

The method of (equally) weighted projections is given by

\[
x_0 \in H \text{ arbitrary, } \quad x_{n+1} := \sum_{i=1}^{N} \frac{1}{P_i} P_i x_n, \quad \text{for all } n \geq 0.
\]

At first glance, this looks quite different from the method of cyclic projections; however Pierra’s product space formalization ([62]; see also [35, 47]) allows an interpretation as the method of cyclic projections in a suitable product space:

Let \(H := \prod_{i=1}^{N} (H_i, \|\cdot\|),\ C_1 := \{x = (x_i)_{i=1}^{N} \in H : x_1 = \cdots = x_N\}, \) and \(C_2 := \prod_{i=1}^{N} C_i\).

The projection of \(x = (x_i)_{i=1}^{N}\) onto \(C_1, C_2\) is given by

\[
(P_{C_1} x, P_{C_2} x) = \left(\sum_{i=1}^{N} \frac{1}{P_i} x_i, \cdots, \sum_{i=1}^{N} \frac{1}{P_i} x_i\right), \quad P_{C_1} x = (P_1 x_1, \ldots, P_N x_N).
\]

Moreover, by [7, 9],

\[
F_1 := \text{Fix} (P_{C_1} P_{C_2}) = \{ (x_1, \ldots, x_N) \in H : x \in \text{Fix} \left( \frac{P_1 + \cdots + P_N}{N} \right) \} = \{ (x_1, \ldots, x_N) \in H : x \in \text{arginf}_{x \in H} \sum_{i=1}^{N} d^2(x, C_i) \},
\]

and

\[
F_2 := \text{Fix} (P_{C_2} P_{C_1}) = \text{arginf}_{(x_1, \ldots, x_N) \in C_2} \sum_{i,j} \|x_i - x_j\|^2.
\]

If we now consider the method of cyclic projections for \(C_1, C_2\) in \(H\) with starting point \(x^0 = (x_0, \ldots, x_0)\), then

\[
x_n = (x_{n+1}, \ldots, x_{n+1}), \quad \text{for all } n \geq 0.
\]

Hence all result on the two set case can be "translated" to this particular setting; see, for instance, [9, Section 6].

42
Flexible weighted relaxed projections

The method of \textit{flexible weighted relaxed projections} is a generalization of the two previously discussed methods: Given \( x_0 \in H \), generate a sequence \((x_n)\) by

\[
x_{n+1} := \left( \sum_{i=1}^{N} \lambda_{i}^{(n)} \left[ (1 - \alpha_{i}^{(n)})I + \alpha_{i}^{(n)} P_{i} \right] \right) x_n, \quad \text{for all } n \geq 0.
\]

The \( \alpha_{i}^{(n)} \in [0, 2] \) are called \textit{relaxation parameters} and the \( \lambda_{i}^{(n)} \) are nonnegative \textit{weights} with \( \sum_{i=1}^{N} \lambda_{i}^{(n)} = 1 \); they might well depend on \( n \) and \( i \).

Under additional assumptions on the relaxation parameters and on the weights, numerous (linear, norm, and weak) convergence results have been obtained for the consistent case; see [8] and the references therein for further information.

Moreover, we would like to mention that at least some of the results on the inconsistent case given in the previous sections generalize to the method of \textit{relaxed projections}, which is the method of flexible weighted projections with \( \alpha_{i}^{(n)} \equiv \alpha_{i} \in ]0, 2[ \) and \( \lambda_{i}^{(n-1)} \equiv 1 \).

\textbf{Dykstra’s algorithm}

\textit{Dykstra’s algorithm} is closely related to the method of cyclic projections; in fact, if each \( C_{i} \) is a closed affine subspace, then Dykstra’s algorithm is essentially the method of cyclic projections. It is defined as follows:

Fix an arbitrary \( x_{0} \in H \) and set \( e_{(N+1)} := e_{(N-2)} := \cdots := e_{0} := 0 \). Then define two sequences \((x_n), (e_n)\) by

\[
x_{n+1} := P_{n+1}(x_{n} + e_{n+1} - N), \quad e_{n+1} := (I - P_{n+1})(x_{n} + e_{n+1} - N), \quad \text{for all } n \geq 0.
\]

\textbf{Conjecture 5.8.1} (on Dykstra’s algorithm) If all \( F_{i} \) are empty, then \( \lim_{n} \|x_{n}\| = +\infty \). Otherwise, all \( F_{i} \) are nonempty and

\[
(x_{n})_{[n]} = i \quad \text{converges in norm to } P_{F_{i}}x_{0}, \quad \text{for every } i.
\]

The following fact not only supports this conjecture but also demonstrates the power of Dykstra’s algorithm:

\textbf{Fact 5.8.2} The previous conjecture is true for: (i) the two set case; (ii) the consistent case; and (iii) the affine subspace case.

\textbf{Proof.} (i) is Bauschke and Borwein’s [7, Theorem 3.8], (ii) is Boyle and Dykstra’s [14, Theorem 2]. (iii): Denote the orbit (for the method of cyclic projections) with starting point \( x_{0} \) by \((x_{n}^{+})\). It is not hard to check that \( x_{n}^{+} := x_{(n+1)-N}, \) for all \( n \) and every \( i \). The result now follows from Corollary 5.7.6. \( \square \)
Acknowledgment

It is our pleasure to thank Jon Vanderwerff for numerous stimulating discussions and helpful suggestions.
References


