Chapter 6

Metric projection methods

The alternating projection method is an algorithm for computing a point in the intersection of some convex sets. The common feature of these methods is that they use compositions of metric projections onto closed convex subsets of a Hilbert space or a Banach space. The most impact is given 1933 by a paper of J. von Neumann (see [30] for a reprint of this paper) which describes the method in Hilbert spaces. Metric projection method in a finite-dimensional space were analyzed for the first time by S. Kaczmarz [26] and G. Cimmino [13]. A breakthrough of these methods was the application in the first tomography scanner by G.N. Hounsfield [24].

In this chapter we consider these methods, present some extensions and variants and document the progress made during the last years. We restrict ourselves mainly to problems formulated in Hilbert spaces.

6.1 Alternating projection method – an introduction

Let $\mathcal{X}$ be a Hilbert space and let $C_1, \dots, C_m$ sets in $\mathcal{H}$. The **cycling projection method** consists in the construction of a sequence $(x^k)_{k\in\mathbb{N}_0}$ generated by

$$x^0 \longmapsto x^1 := P_{C_1}(x^0) \longmapsto x^2 := P_{C_2}(x^1) \longmapsto \cdots \longmapsto x^m := P_{C_m}(x^{m-1}) \longmapsto$$
$$\longmapsto x^{m+1} := P_{C_1}(x^m) \longmapsto x^{m+2} := P_{C_2}(x^{m+1}) \longmapsto \cdots \quad (6.1)$$

where x^0 is used as a starting point of the iteration. We know that at least in the case that $\mathcal{X}$ is a Hilbert space each metric projection P_{C_i} is nonexpansive. In the case $m = 2$ we call the cycling method an **alternating projection method**.

As we see, we use the sets $C_1, \dots, C_m$ in a cyclic way. Clearly, the convergence of the sequence $(x^k)_{k\in\mathbb{N}_0}$ is in the focus of our considerations in this chapter. The following questions are of main interest:

(1) Does $(x^k)_{k\in\mathbb{N}}$ converge (strongly, weakly) to a point z ?

(2) Belongs the point z to the set $C := \cap_{i=1}^m C_i$?

(3) Is the point z in (1) a specific point of C ?

(4) How depends the point z in (1) on the starting point x^0 ?

(5) What is the result when C is empty?

(6) How quick converge the sequence $(x^n)_{n\in\mathbb{N}}$ (when it converges)?

Let $\mathcal{H}$ be a Hilbert space, consider the case $m = 2$ and let $C := C_1, D := C_2$ nonempty closed convex subsets of $\mathcal{H}$. Then the metric projections P_C, P_D are well defined. The procedure (6.1) may be written down as follows:

$$x^{2n+2} := P_D(x^{2n+1}),\ x^{2n+1} := P_C(x^{2n}),\ n \in \mathbb{N}_0,\ x^0 := x \in \mathcal{X} \text{ given}. \tag{6.2}$$

This iteration generates a sequence of points $x^k \in \mathcal{H}$. If this sequence converges we expect that it converge to a point z in $C \cap D$. Thus, a necessary condition is that $C \cap D$ should nonempty. Indeed, if C, D are closed subspaces, convergence can be proved; see Section 6.2. But the method of alternating projection is also useful when the sets C, D do not intersect. In this case we should expect that the sequences $(x^{2n+1})_{n\in\mathbb{N}_0}, (x^{2n})_{n\in\mathbb{N}_0}$ converge to points $x^* \in C, y^* \in D$ respectively, which satisfy $\|x^*-y^*\| = \text{dist}(C, D)$. Unfortunately, the convergence of the sequence $(x^k)_{k\in\mathbb{N}}$ does not hold in general; see Section 6.5. In this case only weak convergence can be proved; see [11]. There is a counterexample which shows that convergence in norm does not hold in general; see [25].

In the case that the problem is considered in a Banach space which is not a Hilbert space the convergence analysis of (6.2) is much more difficult and many questions are still open; see [1, 27]. The main reason is that the metric projections P_C, P_D are not nonexpansive in general.

The sequence $(z^k := x^{2k})_{k\in\mathbb{N}_0}$ generated by (6.2) is the orbit of the operator $T : P_D P_C$: $z^k = T^k(x^0), k \in \mathbb{N}$.[1] As we know the metric projections P_C, P_D are well-studied nonexpansive mappings. As a consequence, the composite mapping T is nonexpansive too. This fact is the motivation to consider orbits of nonexpansive mappings in the next chapter.

Due to its conceptual simplicity and elegance of the projection method, it is not surprising that this method has been rediscovered and generalized many times. There are many variations on and extensions of the basic alternating projection method. We sketch a few of them. All variants are considered in Hilbert spaces.

Dykstra's (1980) algorithm is a method that computes a point in the intersection of convex sets; see [8, 18]. It differs from the alternating projection method in that there are intermediate steps. The key difference between Dykstra's algorithm and the standard alternating projection method occurs when there is more than one point in the intersection of the two sets. In this case, the alternating projection method gives some arbitrary point in this intersection, whereas Dykstra's algorithm gives a specific point: the projection of the initial point x^0 onto the intersection. Dykstra's algorithm may be extended to the case of projecting onto more than two sets.

When the convex sets $C_1, \dots, C_m$ are affine subsets in an euclidean space $\mathbb{R}^n$ then this variant is nowadays called the **ART-method**. Kaczmarz studied such a method in 1937 (see [26]) and Cimmino in 1938 (see [13]) too. This method was successfully applied in computerized tomography where one has to solve a (mostly over-determined) linear system of equations; see Section 5.9.

A big class of new methods may be designed by the **method of relaxation**. It can be observed that averaging of the projections may generate mappings with better behavior.

[1]Mostly, we use for the composition $f \circ g$ of f, g the short notation fg.

Each of the variations above may be transformed into a **randomized method**: in each step the number i of the projection P_{C_i} used to update the approximation is picked at random. Thus, we have a to choose a „probability measure“ on the different projections.

6.2 Alternating projection method on subspaces

Let $\mathcal{H}$ be a Hilbert space and let U, V be closed linear subspaces of $\mathcal{H}$. Clearly, $U \cap V$ is a closed linear subspace of $\mathcal{H}$ too. It is easily seen by geometric inspection that the projection P_U, P_V do not commute in general. The **alternate projection method** consists in the following iteration:

$$x^{2n+2} := P_V(x^{2n+1}), \; x^{2n+1} := P_U(x^{2n}), \; n \in \mathbb{N}_0, \; x^0 := x \in \mathcal{H} \text{ given}. \qquad (6.3)$$

Theorem 6.1 (von Neumann, 1933). *Let $\mathcal{H}$ be a Hilbert space space and let U, V be closed linear subspaces of $\mathcal{H}$. Consider the sequence $(x^k)_{k\in\mathbb{N}_0}$ defined by the alternate projection method* (6.3). *Then the sequence $(x^k)_{k\in\mathbb{N}}$ converges in norm to $P_{U\cap V}(x^0)$.*

Proof:
We follow [7].
Let $W := U \cap V$. We have $x^{2n+1} \in U, x^{2n+2} \in V, n \in \mathbb{N}_0$. We conclude from Theorem 3.3 (5) that $x^k - x^{k+1}$ is orthogonal to x^{k+1} and hence by the law of Pythagoras

$$\|x^n\|^2 = \|x^{n+1}\|^2 + \|x^n - x^{n+1}\|^2, \; n \in \mathbb{N}. \qquad (6.4)$$

Therefore,

$$(\|x^n\|)_{n\in\mathbb{N}} \text{ is decreasing, nonnegative and hence convergent.} \qquad (6.5)$$

Let us prove by induction the following assertion:

$$\|x^k - x^l\|^2 \le \|x^k\|^2 - \|x^l\|^2, \; 1 \le k = l - n, k, l \in \mathbb{N}, \text{ for all } n \in \mathbb{N}_0. \qquad (6.6)$$

Clearly (6.6) holds for $n = 0$ and also for $n = 1$ by (6.4). So assume (6.6) holds true for some $n \ge 1$. Take $k, l \in \mathbb{N}$ with $1 \le k = l - (n+1)$.
Case 1: n is even. Then $n + 1 = l - k$ is odd.
If l is odd, then both $x^{k+1} = P_U(x^k)$ and $x^l = P_U(x^{l-1})$ belong to U, whereas $x^k - x^{k+1} = (I - P_U)x^k$ belongs to $U^\perp$; see Theorem 3.3 (5). Hence

$$\langle x^k - x^{k+1} | x^{k+1} - x^l \rangle = 0. \qquad (6.7)$$

If l is even, we argue similar with U replaced by V and we derive (6.7) once again. Using (6.7), (6.4) and the induction hypothesis, we now obtain

$$\begin{aligned} \|x^k - x^l\|^2 &= \|x^k - x^{k+1}\|^2 + \|x^{k+1} - x^l\|^2 = \|x^k\|^2 - \|x^{k+1}\|^2 + \|x^{k+1} - x^l\|^2 \\ &\le \|x^k\|^2 - \|x^{k+1}\|^2 + \|x^{k+1}\|^2 - \|x^l\|^2 = \|x^k\|^2 - \|x^l\|^2 \end{aligned}$$

Case 2: n is odd. Then $n + 1 = l - k$ is even.
By a similar argumentation as above we obtain

$$\langle x^k - x^l | x^l - x^{l-1} \rangle = 0. \qquad (6.8)$$

This implies $\|x^k - x^{l-1}\|^2 = \|x^k - x^l\|^2 + \|x^l - x^{l-1}\|^2$. With (6.4), the induction hypothesis and (6.8) we conclude

$$\begin{aligned}\|x^k - x^l\|^2 &= \|x^k - x^{l-1}\|^2 + \|x^l - x^{l-1}\|^2 = \|x^k - x^{l-1}\|^2 - \|x^{l-1}\|^2 + \|x^l\|^2 \\ &\le \|x^k\|^2 - \|x^{l-1}\|^2 + \|x^l\|^2 - \|x^{l-1}\|^2 = \|x^k\|^2 - \|x^l\|^2\end{aligned}$$

Now, the induction proof is complete.
By (6.5) and (6.6), the sequence $(x^n)_{n\in\mathbb{N}}$ is a Cauchy sequence and hence convergent. Let $z := \lim_n x^n$. Since x^{2n+1} belongs to U, x^{2n+2} belongs to V we conclude $z \in W := U \cap V$. Therefore, by the continuity of P_W,

$$\lim_n P_W(x^n) = P_W(z) = z. \tag{6.9}$$

Choose $n \in \mathbb{N}$, $t \in \mathbb{R}$, and set $u := (1-t)P_W(x^n) + tP_W(x^{n+1})$. Then $u \in W$ and therefore $P_U(u) = P_V(u) = u$. Moreover, we have $x^{n+1} \in \{P_U(x^n), P_V(x^n)\}$. Since P_U, P_V are nonexpansive, we obtain

$$\|x^{n+1} - u\| \le \|x^n - u\|.$$

After squaring and simplifying, this inequality turns into

$$(1-2t)\|P_W(x^{n+1}) - P_W(x^n)\|^2 + \|P_{W^\perp}(x^n)\|^2 \le \|P_{W^\perp}(x^n)\|^2. \tag{6.10}$$

Since n and t where chosen arbitrarily, we conclude

$$P_W(x^n) = P_W(x^{n+1}),\ n \in \mathbb{N}. \tag{6.11}$$

Combining (6.9) and (6.11) we obtain $\lim_n x^n = P_W(x^0)$. ∎

Algorithm 6.1 Alternate projection method for subspaces

Given two closed linear subspaces U, V of a Hilbert space $\mathcal{H}$ and a starting vector $x^0 \in \mathcal{H}$. Moreover given an accuracy parameter $\varepsilon > 0$. This algorithm computes a sequence $x^0, x^1, , \dots, x^N$ such that $x^N \in U, x^{N+1} \in V$ with $\|x^N - x^{N-1}\| \le \varepsilon$ and $\|x^{N+1} - x^N\| \le \varepsilon$; see Theorem 6.1.

(1) $k := 0$; $u := v := \theta$.

(2) $x^{k+1} := P_U(x^k)$ and $x^{k+2} := P_V(x^{k+1})$.

(3) If $\|x^{k+1} - u\| \le \varepsilon$ and $\|x^{k+2} - v\| \le \varepsilon$ set $N := k-1$ and go to line (6).

(4) Set $u := x^{k+1}, v := x^{k+2}$.

(5) Set $k := k+1$ and go to line (2).

(6) STOP with an approximation $x^N \in U$, $x^{N+1} \in V$.

By the Theorem of Banach Steinhaus we conclude from Theorem 6.1 that $\lim_n (P_U P_V)^n$ converges in norm to $P_{V\cap U}$. Since $U \cap V = V \cap U$, by interchanging U and V in the proof of Theorem 6.1 we also have $\lim_n (P_U P_V)^n = P_{U\cap V}$. That is, the projections can be applied in either order.

It is straightforward to extend Theorem 6.1 to closed affine subspaces. The idea is geometric: translate the affine spaces so they become subspaces, perform the method of alternating projections and then reverse the tranlation.

Corollary 6.2. *Let $\mathcal{H}$ be a Hilbert space space and let C, D be closed affine subspaces of $\mathcal{H}$ such that $C \cap D$ is nonempty. Then $\lim_n (P_C P_D)^n x = P_{C\cap D} x$ for all $x \in \mathcal{H}$.*

Proof:
Let $z \in C \cap D$. Then $U := C - z, V := D - z$ are closed subspaces. Applying Theorem (6.1) twice we obtain for each $x \in \mathcal{H}$

$$(P_C P_D)(x) = P_C(P_D(x)) = P_C(P_U(x-z)+z) = P_V(P_U(x-z)+z-z)+z = P_V(P_U(x-z))+z .$$

Continuing we obtain

$$(P_C P_D)^n(x) = (P_U P_V)^n(x-z)) + z .$$

By Theorem (6.1)

$$\lim_n (P_C P_D)^n(x) = \lim_n (P_U P_V)^n(x-z) + z = P_{C\cap D}(x) .$$

■

Theorem 6.3 (Halperin, 1962). *Let $\mathcal{H}$ be a Hilbert space space and let $U_1, \dots, U_m$ be closed linear subspaces of $\mathcal{H}$, let $P_i := P_{U_i}, i = 1, \dots, m$, and let $U := \cap_{i=1}^m U_i$. Then*

$$\lim_n (P_m \cdots P_1)^n x = P_U x \text{ for all } x \in \mathcal{H} .$$

Proof:
This result was proved by Halperin in [22]. We transfer the proof to the next chapter when we consider some general results for nonexpansive mappings. ■

The results of Theorem 6.3 and Corollary 6.2 are not attainable in the context of Banach spaces since metric projections (when they exist) are not linear nonexpansive orthogonal projections but we have an alternative result.

Theorem 6.4 (Bruck and Reich, 1977). *Let $\mathcal{X}$ be a uniformly convex Banach space space and let $P_1 \dots, P_m$ linear continuous mappings with*

$$\|P_i\| = 1, P_i P_i = P_i, i = 1, \dots, m .$$

Let $T := P_m \cdots P_1$. Then the sequence $(T^n)_{n\in\mathbb{N}}$ converges to $T^\infty \in \mathcal{B}(\mathcal{X})$ where T^∞ is a projection onto $\cap_{i=1}^m ran(P_i)$ with norm 1.

Proof:
For the proof we refer to [12]. ■

Remark 6.5. *Theorem 6.4 holds true if we replace the uniformly convex Banach space $\mathcal{X}$ by a uniformly smooth Banach space; see [4]. Remember the „duality“ concerning uniform convexity and uniform smoothness; see Theorem 4.18.* □

6.3 Alternating projection method for subspaces: order of convergence

Let $\mathcal{H}$ be a Hilbert space and let $U_1, \dots, U_m$ be closed linear subspaces of $\mathcal{H}$. As we know, each metric projection $P_i := P_{U_i}$ is a linear orthogonal projection. Set $U := \cap_{i=1}^m U_i$. From Theorem 6.3 we know

$$\lim_n (P_m \cdots P_1)^n x = P_U(x) \text{ for all } x \in \mathcal{H}. \tag{6.12}$$

It is important for applications to know how fast the algorithm of alternating projections converges. This question can be analyzed in terms of the angle between linear subspaces ([20], [17]). Before we introduce these quantities, let us collect some results concerning the interplay of two projections on subspaces.

Lemma 6.6. *Let $\mathcal{H}$ be a Hilbert space and let V, W be closed linear subspaces of $\mathcal{H}$. Let $P_V, P_W, P_{V\cap W}$ be the associated orthogonal projections. Then:*

(1) $P_V P_{V\cap W} = P_{V\cap W} = P_{V\cap W} P_V$.

(2) $P_V P_{(V\cap W)^\perp} = P_{V\cap (V\cap W)^\perp} = P_{(V\cap W)^\perp} P_V$.

(3) $P_V P_W = P_W P_V$ *if and only if* $P_V P_W = P_{V\cap W}$.

(4) $P_V P_W = \theta$ *if and only if* $P_W P_V = \theta$.

(5) $P_V P_W = P_W$ *if and only if* $P_W P_V = P_W$.

(6) $P_W P_V - P_{V\cap W} = P_W P_V P_{(V\cap)^\perp}$.

Proof:
Ad (1) We have $V \cap W \subset V$. Then the result follows from the reduction principle; see Theorem 3.4.
Ad (2) Since $P_V P_{V\cap W} = P_{V\cap W} P_V$ (see (1)) and $P_{(V\cap W)^\perp} = P_{V\cap W}$ we obtain $P_V P_{(V\cap W)^\perp} = P_{V\cap(V\cap W)^\perp}$ as follows: Let $x \in \mathcal{H}$ and $z = P_V P_{(V\cap W)^\perp}(x)$. Then $z = P_V P_{(V\cap W)^\perp}(z)$ and

$$P_{V\cap(V\cap W)^\perp}(x) = P_V P_{(V\cap W)^\perp} P_{V\cap(V\cap W)^\perp}(x) = P_V P_{(V\cap W)^\perp}(x) = z$$

Ad (3) If $z = P_V P_W x$ then $z = P_W P_V z$ and we obtain $z \in V \cap W$, $P_{V\cap W}(z) = z$. Now, with (a)

$$z = P_{V\cap W} z = P_{V\cap W} P_V P_W x = P_V P_W P_{V\cap W} x = P_{V\cap W} x.$$

Conversely, if $z = P_{V\cap W} x$ then $z \in V \cap W$ and therefore $P_V P_W z = z$. Then with (a)

$$z = P_V P_W z = P_V P_W P_{V\cap W} x = P_{V\cap W} x.$$

If $P_V P_W = P_{V\cap W}$ then $P_W P_V = (P_V P_W)^* = (P_{V\cap W})^* = P_{V\cap W} = P_V P_W$.
Ad $(c), (d)$ It is enough to prove (d).

$$P_V P_W = P_W \iff P_W P_V = (P_V P_W)^* = (P_W)^* = P_W.$$

■

Definition 6.7 (Friederichs, 1937). *Let* U_1, U_2 *be closed linear subspaces of a Hilbert space* $\mathcal{H}$; *set* $U := U_1 \cap U_2$. *The* **Friedrichs-angle** *between the subspaces* U_1 *and* U_2 *is defined to be the angle in* $[0, \pi/2]$ *whose cosine is given by*

$$c(U_1, U_2) := \sup \left\{ \langle u, v \rangle : u \in U_1 \cap U^\perp \cap \overline{B}_1, v \in U_2 \cap U^\perp \cap \overline{B}_1 \right\} .$$

□

Definition 6.8 (Diximier, 1949). *Let* U_1, U_2 *be closed subspaces of a Hilbert space* $\mathcal{H}$. *The* **Dixmier-angle** *between the subspaces* U_1 *and* U_2 *is defined to be the angle in* $[0, \pi/2]$ *whose cosine is given by*

$$c_0(U_1, U_2) := \sup \left\{ \langle u, v \rangle : u \in U_1 \cap \overline{B}_1, v \in U_2 \cap \overline{B}_1 \right\} .$$

□

Lemma 6.9. *Let* U_1, U_2 *be closed linear subspaces of a Hilbert space* $\mathcal{H}$; *set* $U := U_1 \cap U_2$. *Then we have:*

(1) If $U_1 \cap U_2 = \{\theta\}$ *then* $c_0(U_1, U_2) = c(U_1, U_2)$.

(2) $0 \le c(U_1, U_2) \le c_0(U_1, U_2) \le 1$.

(3) $c_0(U_1, U_2) = c_0(U_2, U_1)$, $c(U_1, U_2) = c(U_2, U_1)$.

(4) $c(U_1, U_2) = c_0(U_1 \cap U^\perp, U_2 \cap U^\perp)$.

(5) $|\langle x, y \rangle| \le c_0(U_1, U_2)\|x\|\|y\|$, $x \in U_1, y \in U_2$. *(„Cauchy Schwarz inequality")*

Proof:
These assertions are easy to verify. ■

Lemma 6.10. *Let* U_1, U_2 *be closed linear subspaces of a Hilbert space* $\mathcal{H}$; *set* $U := U_1 \cap U_2$. *Then we have:*

(1) $c_0(U_1, U_2) = \|P_{U_1} P_{U_2}\| = \|P_{U_2} P_{U_1}\| = \|P_{U_2} P_{U_1} P_{U_2}\|^{\frac{1}{2}} = \|P_{U_1} P_{U_2} P_{U_1}\|^{\frac{1}{2}}$.

(2) $P_{U_1} P_U = P_{U_2} P_U = P_U$, $P_{U_1} P_{U^\perp} = P_{U_1 \cap U^\perp}, P_{U_2} P_{U^\perp} = P_{U_2 \cap U^\perp}$.

(3) $P_{U_2} P_{U_1} - P_U = P_{U_2} P_{U_1} P_{U^\perp}$.

(4) $c(U_1, U_2) = \|P_{U_2} P_{U_1} - P_U\| = \|P_{U_2} P_{U_1} P_{U^\perp}\|$.

Proof:
Ad (1) We have

$$\begin{aligned} c_0(U_1, U_2) &= \sup \left\{ \langle u, v \rangle : u \in U_1 \cap \overline{B}_1, v \in U_2 \cap \overline{B}_1 \right\} \\ &= \sup \left\{ \langle P_{U_1} x, P_{U_2} y \rangle : x \in \overline{B}_1, y \in \overline{B}_1 \right\} \\ &= \sup \left\{ \langle x, P_{U_1} P_{U_2} y \rangle : x \in \overline{B}_1, y \in \overline{B}_1 \right\} \\ &= \|P_{U_1} P_{U_2}\| . \end{aligned}$$

The second equality follows from

$$c_0(U_1, U_2) = c_0(U_2, U_1) = \|P_{U_2} P_{U_1}\| = \|(P_{U_2} P_{U_1})(P_{U_2} P_{U_1})^*\|^{\frac{1}{2}} = \|P_{U_2} P_{U_1} P_{U_2}\|^{\frac{1}{2}} .$$

Ad (9) Using (7),(8) we obtain

$$\begin{aligned} c(U_1,U_2) &= c_0(U_1\cap U^\perp, U_2\cap U^\perp) = \|P_{U_1\cap U^\perp}P_{U_2\cap U^\perp}\| \\ &= \|P_{U_1}P_{U^\perp}P_{U_2}P_{U^\perp}\| = \|P_{U_1}P_{U_2}P_{U^\perp}\| \\ &= \|P_{U_1}P_{U_2}(I-P_U)\| = \|P_{U_1}P_{U_2}-P_U\| \end{aligned}$$

■

Corollary 6.11. *Let* U_1, U_2 *be closed linear subspaces of a Hilbert space* $\mathcal{H}$; *set* $U := U_1\cap U_2$. *Then*

$$\|(P_{U_2}P_{U_1})^n - P_U\| = \|(P_{U_2}P_{U_1}P_{U^\perp})^n\| \le c(U_1,U_2)^n\,,\ n\in\mathbb{N}\,. \tag{6.13}$$

Proof:
It is enough to prove $(P_{U_2}P_{U_1})^n - P_U = (P_{U_2}P_{U_1}P_{U^\perp})^n, n\in\mathbb{N}$. We do this by induction. The case $n=1$ is already clear. Let the identity hold for $n\in\mathbb{N}$. Then

$$\begin{aligned} (P_{U_2}P_{U_1})^{n+1} - P_U &= (P_{U_2}P_{U_1})^{n+1} - P_{U_2}P_{U_1}P_U \\ &= P_{U_2}P_{U_1}((P_{U_2}P_{U_1})^n - P_U) \\ &= P_{U_2}P_{U_1}(P_{U_2}P_{U_1}P_{U^\perp})^n \\ &= (P_{U_2}P_{U_1})(P_{U^\perp})(P_{U_2}P_{U_1}P_{U^\perp})^n \\ &= (P_{U_2}P_{U_1}P_{U^\perp})^{n+1}\,. \end{aligned}$$

In the third line we have used that P_{U_i} and $P_{U^\perp}$ commute. ■

The bound for the norm $\|(P_{U_2}P_{U_1})^n - P_U\|$ obtained in Corollary 6.11 is not sharp.

Theorem 6.12 (Aronszajn, 1955). *Let* $\mathcal{H}$ *be a Hilbert space and let* U_1, U_2 *be closed subspaces of* $\mathcal{H}$; $U := U_1\cap U_2$. *We set* $P_1 := P_{U_1}, P_2 := P_{U_2}$. *Then*

$$\|(P_2P_1)^n - P_U\| \le c(U_1,U_2)^{2n-1}\,,\ n\in\mathbb{N}\,. \tag{6.14}$$

Proof:
See [2]. We omit the proof since we present immediately a result which shows that in (6.14) equality holds. ■

Theorem 6.13 (Kayalar and Weinert, 1965). *Let* $\mathcal{H}$ *be a Hilbert space and let* U_1, U_2 *be closed subspaces of* $\mathcal{H}$; $U := U_1\cap U_2$. *Then*

$$\|(P_2P_1)^n - P_U\| = c(U_1,U_2)^{2n-1}\,,\ n\in\mathbb{N}\,. \tag{6.15}$$

Proof:
We set $P_1 := P_{U_1}, P_2 := P_{U_2}$ and $Q_i := P_iP_{U^\perp}, i=1,2$. Then $Q_i = P_{U_i\cap U^\perp} = P_{U_1\cap U^\perp}P_{U_i}, i = 1,2$, due to the reduction principle ($U_i\cap U^\perp \subset U_i$); see Theorem 3.4. From the proof of Corollary 6.11 we know that $(P_{U_2}P_{U_1})^n - P_U = (P_{U_2}P_{U_1}P_{U^\perp})^n, n\in\mathbb{N}$, holds. Then for $n\in\mathbb{N}$

$$\begin{aligned} \|(P_{U_2}P_{U_1})^n - P_U\|^2 &= \|(P_{U_2}P_{U_1}P_{U^\perp})^n\|^2 = \|(P_{U_2}P_{U_1}P_{U^\perp})^n((P_{U_2}P_{U_1}P_{U^\perp})^*)^n\| \\ &= \|(Q_2Q_1)^n(Q_1Q_2)^n\| = \|(Q_2Q_1)^n((Q_2Q_1)^*)^n\| \\ &= \|(Q_2Q_1Q_2)^{2n-1}\| = \|(Q_2Q_1Q_1Q_2)^{2n-1}\| \\ &= \|((Q_2Q_1)(Q_2Q_1)^*)^{2n-1}\| = \|Q_2Q_1\|^{2(2n-1)} \\ &= c(U_1,U_2)^{2(2n-1)} \end{aligned}$$

■

From (6.15) we know that the sequence $(P_2P_1)^n$ converges „quickly" to P if $c(U_1,U_2) < 1$; the rate of convergence is then of a geometrical progression. Notice that we have the following equivalences (see the Theorem 6.17 below):

$$c(U_1,U_2) < 1 \iff U_1 + U_2 \text{ is closed } \iff U_1^\perp + U_2^\perp \text{ is closed }. \qquad (6.16)$$

If $U_1 + U_2$ is not closed then the convergence of $(P_2P_1)^n$ to P may be „arbitrary" slow; see [6].

Theorem 6.14. *Let $\mathcal{H}$ be a Hilbert space and let U_1, U_2 be closed subspaces of $\mathcal{H}$; $U := U_1 \cap U_2$. Then the following statements are equivalent:*

(a) $c(U_1,U_2) < 1$.

(b) $U_1 + U_2$ *is closed.*

(c) $U_1 \cap U^\perp + U_2 \cap U^\perp$ *is closed.*

(d) $U_1^\perp + U_2^\perp$ *is closed.*

Proof:
Ad $(1) \iff (2)$ Since $c(U_1,U_2) = c_0(U_1 \cap U^\perp, U_2 \cap U^\perp)$ and $U_1 \cap U^\perp \cap U_2 \cap U^\perp = \{\theta\}$, it follows that $c(U_1,U_2) < 1$ holds if and only if $U_1 \cap U^\perp + U_2 \cap U^\perp$ is closed.
Ad $(2) \iff (3)$ Let $Y := U_1 \cap U^\perp$ and $Z := U_2 \cap U^\perp$. We must show that $Y+Z$ is closed if and only if $U_1 + U_2$ is closed.
Obviously, $Y + Z \subset (U_1 + U_2) \cap U^\perp$. Let $x \in (U_1 + U_2) \cap U^\perp$. Then $x = y + z$ for some $y \in U_1, z \in U_2$. Since $x \in U^\perp$ we have $P_U z = \theta$. Due to

$$x = x - P_U x = y - P_U y + z - P_U z$$

we have $x \in U_1 \cap U^\perp + U_2 \cap U^\perp = Y + Z$. Thus, we have shown

$$Y + Z = (U_1 + U_2) \cap U^\perp .$$

Obviously, $Y + Z + U \subset U_1 + U_2 + U = U_1 = U_2$. Let $x \in U_1 + U_2$. Then $x = y + z$ for some $y \in U_1, z \in U_2$, and

$$x = P_U x + P_{U^\perp} x = P_U x + P_{U^\perp} y + P_{U^\perp} z = P_U x + (y - P_U y) + U_2 \cap U^\perp .$$

This shows $x \in U + U_1 \cap U^\perp + U_2 \cap U^\perp = Y + Z + U_1 + U$. Thus, we have shown

$$U_1 + U_2 = Y + Z + U .$$

Now suppose $Y + Z$ is closed. Then from $U_1 + U_2 = Y + Z + U$ and the fact that $Y + Z \subset U^\perp$, it follows that $U_1 + U_2$ is closed. Conversely, if $U_1 + U_2$ is closed, the relation $Y + Z = (U_1 + U_2) \cap U^\perp$ implies that $Y + Z$ is closed.
The proof of $(3) \iff (4)$ is left to the reader. ■

Lemma 6.15. *Let $\mathcal{H}$ be a Hilbert space and let V, W be closed linear subspaces of $\mathcal{H}$. Let $P_V, P_W, P_{V\cap W}$ be the associated orthogonal projections. Then:*

(a) $P_V P_W - P_{V\cap W} = P_V P_W P_{(V\cap W)^\perp}$.

Proof:
Ad (b) We have $P_{(V\cap W)^\perp} P_W = P_W P_{(V\cap W)^\perp}$ if and only $P_{V\cap W} P_W = P_W P_{V\cap W}$. Since $V\cap W \subset W$ and $W^\perp \subset (V\cap W)^\perp$ it is easy to see that $P_{V\cap W} P_W = P_W P_{V\cap W}$.

■

Lemma 6.16. *Let $\mathcal{H}$ be a Hilbert space and let V, W be closed linear subspaces of $\mathcal{H}$. Let $P_V, P_W, P_{V\cap W}$ be the associated orthogonal projections. Then:*

(a) *If $V \subset W^\perp$ then $V + W$ is closed and $P_{V+W} = P_V + P_W$.*

(b) *If $P_V P_W = P_W P_V$ then $P_{\overline{V+W}} = P_V + P_W - P_V P_W$.*

Proof:
Ad (a) We have
$$\langle P_V P_W x|y\rangle = \langle P_W x|P_V y\rangle = 0 \text{ for all } x, y \in \mathcal{H} .$$
This implies $P_V P_W = \theta$ and $P_W P_V = \theta$. We obtain $(P_V + P_W)^2 = P_V^2 + P_W^2 = P_V + P_W$. This shows that $P_V + P_W$ is a projection.
Let $x \in V, y \in W$. Then $x + y \in V + W$ and $P_V x = x, P_W y = y, P_V y = \theta, P_W y = y$. We conclude $(P_V + P_W)(x + y) = x + y$.
Let $z \in (V + W)^\perp$. Then
$$\langle (P_V + P_W)z|w\rangle = \langle z|(P_V + P_W)w\rangle = \langle z|P_V w + P_W)w\rangle \text{ for all } w \in \mathcal{H} .$$
Therefore, $P_V + P_W)(z) = \theta$. Alltogether, we have that $P_V + P_W = P_{V+W}$.
If $(v^n + w^n)_{n\in\mathbb{N}}$ is a sequence in $V + W$ $(v^n \in V, w^n \in W, n \in \mathbb{N},)$ which converges to $z \in \mathcal{H}$ then $P_{V^\perp}(v^n + w^n) = P_{V^\perp}(w^n) = w^n$ since $W \subset V^\perp$. Therefore $(w^n)_{n\in\mathbb{N}}$ converges to $w := P_{V^\perp} z \in W$ and $(v^n)_{n\in\mathbb{N}}$ converges to $z - w$. Since V is closed, $z - w \in V$. ■

Theorem 6.17. *Let $\mathcal{H}$ be a Hilbert space and let V, W be closed subspaces of $\mathcal{H}$. Then the following statements are equivalent:*

(a) $c(V, W) < 1$.

(b) *$V + W$ is closed.*

Proof:
We show that, without loss of generality, we may assume that $V\cap W = \{\theta\}$. If $V\cap W \neq \{\theta\}$ then we may write
$$V = V_1 \oplus V \cap W \,,\; W = W_1 \oplus V \cap W$$
where
$$V_1 := \{u \in V : \langle u|z\rangle = 0 \text{ for all } z \in V \cap W\} ,\; W_1 := \{v \in W : \langle v|z\rangle = 0 \text{ for all } z \in V \cap W\} .$$
Then
$$V + W = (V_1 \oplus V \cap W) + W_1 \oplus V \cap W = (V_1 + W_1) \oplus V \cap W .$$
Thus, $V + W$ is closed if and only if $V_1 + W_1$ is closed, where $V_1 \cap W_1 = \{\theta\}$. Thus we may prove the equivalence under the additional assumption that $V \cap W = \{\theta\}$. Then

$c(V,W) < 1$ if and only if $c_0(V,W) < 1$.
Ad $(a) \implies (b)$ Consider tha mapping

$$T : V \times W \ni (v,w) \longmapsto v + w \in V + W .$$

T is linear, surjective and injective since $V \cap W = \{\theta\}$. We show that T is bounded.
$V \times W$ is a Banach space endowed with the norm $\|(v,w)\| := (\|v\|^2 + \|w\|^2)^{\frac{1}{2}}$. Let $v \in V, w \in W$. Then

$$\begin{aligned} \|T(v,w)\|^2 &= \langle v+w|v+w\rangle \;=\; \|v\|^2 + 2\langle v|w\rangle + \|w\|^2 \\ &\le \|v\|^2 + 2\|v\|\|w\| + \|w\|^2 \;\le\; \|v\|^2 + \|v\|^2 + \|w\|^2 + \|w\|^2 \\ &= 2\|(v,w)\|^2 \end{aligned}$$

This shows $T(v,w)\| \le \sqrt{2}\|(v,w)\|$, $(v,w) \in V \times W$. By the open mapping theorem, T has a bounded inverse, that is, there exists $d > 0$ such that

$$d\|(v,w)\| \le \|T(v,w)\| = \|v+w\| \le \sqrt{2}\|(v,w)\| , \; v \in V, w \in W .$$

Consequentely,

$$d^2(\|v\|^2 + \|w\|^2) = d^2\|(v,w)\|^2 \le \|v+w\|^2 = 2\|v\| + 2\langle v|w\rangle + \|w\|^2 .$$

Hence,

$$(c^2 - 1)(\|v\|^2 + \|w\|^2) \le 2\langle v|w\rangle , \; v \in V, w \in W .$$

and

$$\frac{1}{2}c^2 - 1 \le \langle v|w\rangle , \; v \in V \cap \overline{B}_1, w \in W \cap \overline{B}_1 .$$

This implies

$$1 - \frac{1}{2}c^2 \ge \langle v,w\rangle , \; v \in V \cap \overline{B}_1, w \in W \cap \overline{B}_1 .$$

Notice that W is a linear subspace. Then

$$c(V,W) \le 1 - \frac{1}{2}c^2 < 1 .$$

Ad $(b) \iff (a)$ We have $c \in [0,1)$ with

$$\langle v|w\rangle \le c\|v\|\|w\| , \; v \in V \cap \overline{B}_1, w \in W \cap \overline{B}_1 .$$

Then

$$-2c\|v\|\|w\| \le 2\langle v|w\rangle \le 2c\|v\|\|w\| , \; v \in V \cap \overline{B}_1, w \in W \cap \overline{B}_1 . \tag{6.17}$$

Since V, W are convex and closed these sets are weakly closed.
Let $(v^k)_{k\in\mathbb{N}}$ be a sequence in V, $(w^k)_{k\in\mathbb{N}}$ a sequence in W such that the sequence $(z^k)_{k\in\mathbb{N}}$ with $z^k = v^k + w^k, k \in \mathbb{N}$, converges to some $z \in \mathcal{H}$. We have to show that z can decomposed as follows: $z = v + w$ with $v \in V, w \in W$.
First we show that $(v^k)_{k\in\mathbb{N}}, (w^k)_{k\in\mathbb{N}}$ are bounded sequnces. We have

$$\begin{aligned} \|z^k\|^2 &= \|v^k + w^k\|^2 \;=\; \|v^k\|^2 + \|w^k\|^2 + 2\langle v^k|w^k\rangle \\ &\ge \|v^k\|^2 + \|w^k\|^2 - 2c\|v^k\|\|w^k\| \\ &= (c\|v^k\| - \|w^k\|)^2 + (1-c)\|w^k\|^2 \end{aligned}$$

and therefore

$$\|z^k\|^2 = \|v^k + w^k\|^2 \geq (c\|v^k\| - \|w^k\|)^2 + (1-c)\|w^k\|^2 .$$

Interchanging v^k and w^k we also obtain

$$\|z^k\|^2 = \|v^k + w^k\|^2 \geq (c\|w^k\| - \|v^k\|)^2 + (1-c)\|v^k\|^2 .$$

We see from these inequalities that both sequences $(v^k)_{k\in\mathbb{N}}, (w^k)_{k\in\mathbb{N}}$ must be bounded. Since the unit ball is weakly sequential compact, passing to subsequences if necessary, we may assume that there exist $v \in V, w \in W$ with

$$v = w - \lim v^k \,,\; w = w - \lim_k w^k .$$

Thus,

$$v + w = w - \lim_k (v^k + w^k) = w - \lim_k z^k = \lim_k z^k = z .$$

■

Theorem 6.18 (Smith, Solmon, and Wagner, 1984). *Let $\mathcal{H}$ be a Hilbert space, let $U_1, \dots, U_m$ be closed subspaces, let $U := \cap_{i=1}^m$, and let $P_i := P_{U_i}, i = 1, \dots, m$. Then*

$$\|(P_m \cdots P_1)^n x - P_U x\| \leq c^n \|x - P_U(x)\| \,,\; x \in \mathcal{H}, n \in \mathbb{N} \tag{6.18}$$

where

$$c = \left(1 - \prod_{j=1}^{m-1}(1 - c_j^2)\right)^{\frac{1}{2}}$$

and $c_i = c(U_i, \cap_{j=i+1}^m U_j), i = 1, \dots, m-1 .$

Proof:
Let $T := P_m \cdots P_1$. Let $x \in \mathcal{H}$ and $y = P_U x$. We have to prove $\|T^n x - y\|^2 \leq c^n \|x - y\|^2$. We follow [19].
Since $y \in U$ and T is the identity on U, the inequality to be proved can also be written as

$$\|T^n v\|^2 \leq c^n \|v\|^2 \,,\; n \in \mathbb{N} \text{ where } v := x - y . \tag{6.19}$$

Observe, $v \in U^\perp, T^n v \in U^\perp, n \in \mathbb{N}$. Now, it is sufficient to prove

$$\|Tv\|^2 \leq c\|v\|^2 \text{ for } v \in U^\perp . \tag{6.20}$$

In fact, if (6.20) holds, then it follows that

$$\|T^n v\|^2 \leq c\|T^{n-1} v\|^2 \leq c^2 \|T^{n-2} v\|^2 \leq \cdots \leq c^n \|v\|^2$$

and (6.19) holds. (6.20) will be proved by induction on m .
If $m = 1$ the conclusion is clear. Let $U' := U_m \cap \cdots \cap U_2$ and $T' := P_m \cdots P_2$. For any $v \in U^\perp$ write $v = w + v^1$ with $w \in U_1$ and $v^1 \in U_1^\perp$. Then $P_U v = T' w$. Write $w = w' + w''$ with $w' \in U'$ and $w'' \in U'^\perp$. Then $T' w = w' + T' w''$ and since

$$\langle T' w'' | w' \rangle = \langle w'' | P_2 \cdots P_m w' \rangle = \langle w'' | w' \rangle = 0,$$

$T'w''$ and w' are orthogonal so that

$$\|T'w\|^2 = \|w'\|^2 + \|T'w''\|^2 .$$

By the inductive hypothesis,

$$\|T'w''\|^2 \leq \big(1 - \prod_{i=2}^{m-1}(1-c_i^2)\big)\|w''\|^2 .$$

Now we obtain

$$\begin{aligned}
\|T'w\|^2 &\leq \|w'\|^2 + \big(1 - \prod_{i=2}^{m-1}(1-c_i^2)\big)\|w''\|^2 \\
&= \|w'\|^2 + \big(1 - \prod_{i=2}^{m-1}(1-c_i^2)\big)(\|w\|^2 - \|w'\|^2) \\
&= \big(1 - \prod_{i=2}^{m-1}(1-c_i^2)\big)\|w\|^2 + \prod_{i=2}^{m-1}(1-c_i^2)\|w'\|^2 \qquad (6.21)
\end{aligned}$$

On the other hand, as $w = v - v^1$ with $v \in U^\perp$ and $v^1 \in U_1^\perp$ it follows that for $z \in U$

$$\langle w|z\rangle = \langle v - v^1|z\rangle = \langle v|z\rangle - \langle v^1|z\rangle = 0 .$$

This shows that $w \in U_1$ is orthogonal to $U = U_1 \cap U'$. Moreover, as $w' = w - w''$ with w orthogonal to $U = U_1 \cap U'$ and $w'' \in U'^\perp$, it follows that for $z \in U$

$$\langle w'|z\rangle = \langle w - w''|z\rangle = \langle w|z\rangle - \langle w''|z\rangle = 0 .$$

Hence, $w' \in U'$ and w' is orthogonal to $U = U_1 \cap U'$. Now, it follows that

$$\|w'\|^2 = \langle w'|w'\rangle = \langle w - w''|w'\rangle = \langle w|w'\rangle \leq c(U_1, U')\|w\|\|w'\| .$$

Thus, $\|w'\| \leq c_1^2\|w\|$. Replacing this last expression in (6.21) we have

$$\begin{aligned}
\|T'w\|^2 &\leq \big(1 - \prod_{i=2}^{m-1}(1-c_i^2)\big)\|w\|^2 + \prod_{i=2}^{m-1}(1-c_i^2)c_1^2\|w\|^2 \\
&= \big(1 - \prod_{i=1}^{m-1}(1-c_i^2)\big)\|w\|^2
\end{aligned}$$

Finally, as $Tv = T'w$ and $\|w\| \leq \|v\|$, it follows that (6.20) holds. Now the proof is complete. ■

As we know, the rate of the alternate projection method on closed linear subspaces is governed by the norm of the operator $(P_m \cdots P_1)^n - P_U$. Indeed,

$$\|(P_m \cdots P_1)^n x - P_U x\| \leq \mathrm{dev}(n, m)\|x\| \text{ for all } x \in \mathcal{H}, \qquad (6.22)$$

where

$$\mathrm{dev}(n, m) := \|(P_m \cdots P_1)^n - P_U\| \qquad (6.23)$$

is the smallest constant which works in (6.22). We have seen alternate useful ways of expressing the constant dev(n, m):

$$\text{dev}(n,m) = \|(P_m \cdots P_1)^n - P_U\| = \|(P_m \cdots P_2 P_1 P_{U^\perp})^n\| = \|(Q_m \cdots Q_1)^n\| \qquad (6.24)$$

where $Q_i = P_i P_{U^\perp}, i = 1, \dots, m$. From these identities we obtain (crude) upper bounds on dev(n, m):

$$\text{dev}(n,m) \le \|P_m \cdots P_2 P_1 P_{U^\perp}\|^n = \|Q_m \cdots Q_1\|^n . \qquad (6.25)$$

In particular, when $m = 2$, we see from this that

$$\text{dev}(n,m) \le \|P_2 P_1 P_{U^\perp}\|^n = c(U_1, U_2)^n . \qquad (6.26)$$

But we know already that dev$(n, 2) = c(U_1, U_2)^{2n-1}$ (Kayalar,Weinert). From Theorem 6.13 we know for $m \ge 2$

$$\text{dev}(n,m) \le c^n \text{ where } c = \left(1 - \prod_{i=1}^{m-1}(1 - c_i^2\right)^{\frac{1}{2}}, c_i := c(U_i, \cap_{j=i+1}^m, i = 1, \dots, m-1 . \qquad (6.27)$$

In Theorem 6.18 the angle between two subspaces is used iteratively but it is possible to define an angle for m subspaces. To motivate the definition of such an angle, let us give an alternatively presentation of the Friedrichs angle.

Lemma 6.19. *Let U_1, U_2 be closed subspaces of a Hilbert space $\mathcal{H}$; set $U := U_1 \cap U_2$. Then the Friedrichs-angle $c(U_1, U_2)$ between U_1 and U_2 is given by*

$$c(U_1, U_2) \quad = \quad \sup\left\{ \frac{2\langle u|v\rangle}{\|u\|^2 + \|v\|^2} : u \in U_1 \cap U^\perp, v \in U_2 \cap U^\perp, (u, v) \ne (\theta, \theta) \right\} .$$

Proof:
Let s denote the supremum in the identity above. For every admissible pair (u_1, u_2) with $\|u_1\|^2 + \|u_2\|^2 \ne 0$ we have

$$\frac{2\langle u_1|u_2\rangle}{\|u_1\|^2 + \|u_2\|^2} \le \frac{\langle u_1|u_2\rangle}{\|u_1\|\|u_2\|} \le \frac{|\langle u_1|u_2\rangle|}{\|u_1\|\|u_2\|} = c(U_1, U_2) .$$

Therefore $s \le c(U_1, U_2)$.
For the reverse inequality, let $\varepsilon > 0$. Then there exist $u_1 \in U_1 \cap U^\perp \cap \overline{B}_1$ and $u_2 \in U_2 \cap U^\perp \cap \overline{B}_1$ such that $c(U_1, U_2) < |\langle u_1|u_2\rangle| + \varepsilon$. We obtain

$$s \ge \frac{2\langle u_1|u_2\rangle}{\|u_1\|^2 + \|u_2\|^2} = |\langle u_1|u_2\rangle| > c(U_1, U_2) - \varepsilon .$$

Since ε is arbitrary, we obtain $s = c(U_1, U_2)$. □

Definition 6.20. *Let $\mathcal{H}$ be a Hilbert space and let $U_1, \dots, U_m$ be closed subspaces of $\mathcal{H}$; set $U := \cap_{i=1}^m U_i$. The* **Friedrichs number** *$c(U_1, \dots, U_m)$ is defined as*

$$c(U_1, \dots, U_m) \quad = \quad \sup\left\{ \frac{1}{m-1} \frac{\sum_{j,k=1, j\ne k}^m \langle u_j|u_k\rangle}{\sum_{j=1}^m \langle u_j|u_j\rangle} : u_i \in U_i \cap U^\perp, i = 1, \dots, m, \sum_{i=1}^m \|u_i\|^2 \ne 0 \right\} .$$

□

The analysis of the convergence of the projection methods on linear subspaces $U_1, \dots, U_m$ for $m > 2$ is ongoing. The focus of such investigations is on the decomposition of the operator $(P_{U_1} \dots P_{U_1})^n - P_{\cap_{i=1,\dots,m} U_i}$ into a product of projections on certain combinations of the subspaces U_i and relating this decomposition to certain angles. We have already seen such a decomposion in the result of Kayalar and Weinert. A more elaborated decomposition can be found in [14] and [5].

6.4 Acceleration techniques

In this section, we propose acceleration methods for projecting onto the intersection of finitely many affine spaces. These strategies can be applied to general feasibility problems where not only affine spaces are involved, as long as there is more than one affine space.

The observation which is the basis of the accelerating steps is that each projection onto an affine subspace identifies a hyperplane of codimension 1 containing the intersection.

Let us begin with the description of an algorithm for the case of linear subspaces. Let $U_1, \dots, U_m$ be closed linear subspaces in the Hilbert space $\mathcal{H}$ and let $U := \cap_{i=1}^m U_i$. Suppose we have a current point $\hat{x}$. Then we produce the projection $\hat{x}^+ := P_{U_i}(\hat{x})$. This pair $\hat{x}, \hat{x}^+$ identifies a hyperplane

$$\hat{H} := \{x \in \mathcal{H} : \langle \hat{y} | x \rangle = \hat{a}\} \text{ where } \hat{y} := \hat{x} - \hat{x}^+, \hat{a} := \langle \hat{y} | \hat{y} \rangle .$$

Clearly, when $\hat{y} = \theta$ then $\hat{x}^+ = \mathcal{H}$ and $U \subset U_i \subset \hat{x}^+$. Otherwise

6.5 Alternating projection method and Fejér monotone sequences

Let $\mathcal{H}$ be a Hilbert space and let C, D be nonempty closed convex subsets of $\mathcal{H}$. Clearly, $C \cap D$ is a closed convex subset of $\mathcal{H}$. We consider the alternate projection method

$$x^{2n+2} := P_D(x^{2n+1}), \ x^{2n+1} := P_C(x^{2n}), \ n \in \mathbb{N}_0, \ x^0 := x \in \mathcal{X} \text{ given}. \tag{6.28}$$

In this way we obtain a sequence $(x^n)_{n\in\mathbb{N}}$ in $\mathcal{H}$.

A simple example using a halfspace and a line in $\mathbb{R}^2$ shows that the method of alternating projections may not converge to the projection of the starting point onto on the intersection of the sets. To analyze the behavior of the method (we follow mainly [9]) we start from the following observation:

$$\|x^{n+1} - u\| \le \|x^n - u\| \text{ for all } n \in \mathbb{N}, u \in C \cap D. \tag{6.29}$$

Clearly, this result follows from the nonexpansivity of P_C, P_D.

Definition 6.21. *Let $\mathcal{H}$ be a Hilbert space , let $(x^n)_{n\in\mathbb{N}_0}$ be a sequence in $\mathcal{H}$ and let F be a nonempty closed convex subset of $\mathcal{H}$. Then $(x^n)_{n\in\mathbb{N}}$ is called* **Fejér-montone** *with respect to F if*

$$\|x^{n+1} - u\| \le \|x^n - u\| \textit{ for all } n \in \mathbb{N}_0 \textit{ and all } u \in F.$$

□

Here is a huge list of properties of a Fejér monotone sequence. They are very helpful to analyze the alternate projection methods.

Lemma 6.22. *Let $\mathcal{H}$ be a Hilbert space and let F be a nonempty closed convex subset in $\mathcal{H}$. If $(x^n)_{n\in\mathbb{N}_0}$ is a Fejér monotone sequence with respect to F then the following hold:*

(1) $dist(x^{n+1}, F) \le dist(x^n, F)$ for all $n \in \mathbb{N}_0$.

(2) For every $u \in F$, the sequences $(\|x^n - u\|)_{n\in\mathbb{N}_0}$, $(\|x^n\|^2 - 2\langle x^n|u\rangle)_{n\in\mathbb{N}_0}$ converge.

(3) The sequence $(x^n)_{n\in\mathbb{N}_0}$ is bounded and the set of strong cluster points of $(x^n)_{n\in\mathbb{N}_0}$ is bounded.

(4) The sequence $(P_F(x^n))_{n\in\mathbb{N}_0}$ converges strongly to a point in F.

(5) If $int(F) \neq \emptyset$, then $(x^n)_{n\in\mathbb{N}_0}$ converges strongly to a point in F.

(6) If w^1, w^2 are weak cluster points of $(x^n)_{n\in\mathbb{N}_0}$ then $w^1 - w^2 \in (F - F)^\perp$.

(7) The sequence $(x^n)_{n\in\mathbb{N}_0}$ has weak cluster points in $\mathcal{H}$ but it has at most one weak cluster point in F.

(8) The sequence $(x^n)_{n\in\mathbb{N}_0}$ converges weakly to some point in F iff all weak cluster points of $(x^n)_{n\in\mathbb{N}_0}$ belong to F.

(9) Every weak cluster point of $(x^n)_{n\in\mathbb{N}_0}$ that belongs to F must be $\lim_n P_F(x^n)$.

(10) If all weak cluster points of $(x^n)_{n\in\mathbb{N}_0}$ belong to F, then $(x^n)_{n\in\mathbb{N}_0}$ converges weakly to $\lim_n P_F(x^n)$.

(11) If at least one strong cluster point x^* of $(x^n)_{n\in\mathbb{N}_0}$ belongs to F then $x^* = \lim_n x^n$.

(12) If F is a closed affine subspace of $\mathcal{H}$, then $P_F(x^n) = P_F(x^0)$ for all $n \in \mathbb{N}_0$.

(13) If F is a closed affine subspace of $\mathcal{H}$ and if every weak cluster point of $(x^n)_{n\in\mathbb{N}_0}$ belongs to F then $P_F(x^0) = w - \lim_n x^n$.

Proof:
Ad (1) A trivial consequence of the defining property.
Ad (2) $(\|x^n - u\|)_{n\in\mathbb{N}_0}$ converges since it is a decreasing sequence for all $u \in F$. This implies the convergence of $(\|x^n\|^2 - 2\langle x^n|u\rangle)_{n\in\mathbb{N}_0}$ for all $u \in F$ since $\|x^n - u\|^2 = \|x^n\|^2 - 2\langle x^n|u\rangle + \|u\|^2$
Ad (3) Let $u \in F$ and $r := \|x^0\| + \|u\|$. Clearly, $x^n \in \overline{B}_r$ for all $n \in \mathbb{N}_0$; see (2). Now, it is clear that the set of strong cluster points is contained in $\overline{B}_r$ too.
Ad (4) Let $m, n \in \mathbb{N}, m > n$. Applying the parallelogram identity, using the best approximation property of P_F and the defining Fejér property we obtain

$$\begin{aligned}
\|P_F(x^m) - P_F(x^n)\|^2 &= 2\|P_F(x^m) - x^m\|^2 + 2\|P_F(x^n) - x^m\|^2 \\
&\quad -4\|\frac{1}{2}(P_F(x^m) + P_F(x^n)) - x^m\|^2 \\
&\le 2\|P_F(x^m) - x^m\|^2 + 2\|P_F(x^n) - x^m\|^2 - 4\|P_F(x^m) - x^m\|^2 \\
&= 2\|P_F(x^n) - x^m\|^2 - 2\|P_F(x^m) - x^m\|^2 \\
&\le 2\|P_F(x^n) - x^n\|^2 - 2\|P_F(x^m) - x^m\|^2 \\
&= 2(\mathrm{dist}(x^n, F) - \mathrm{dist}(x^m, F))
\end{aligned}$$

Since $\operatorname{dist}(x^n, F) - \operatorname{dist}(x^m, F)$ goes to zero as m, n go to ∞, $(P_F(x^n))_{n\in\mathbb{N}}$ is a Cauchy sequence, hence it converges strongly to a point in F.
Ad (5) Suppose that $B_t(u) \subset F$ for some $u \in F, t > 0$. For any x^{n+1} with $x^{n+1} \neq x^n$, we have

$$\|x^{n+1} - (u - t(x^{n+1} - x^n)\|x^{n+1} - x^n\|^{-1})\|^2 \leq \|x^n - (u - t(x^{n+1} - x^n)\|x^{n+1} - x^n\|^{-1})\|^2$$

and hence

$$2t\|x^{n+1} - x^n\| \leq \|x^n - u\|^2 - \|x^{n+1} - u\|^2 .$$

This inequality holds also for the case $x^{n+1} = x^n$. We obtain

$$2t\|x^m - x^n\| \leq \|x^n - u\|^2 - \|x^m - u\|^2 , \ m, n \in \mathbb{N}, m > n ,$$

and by (2) $(x^n)_{n\in\mathbb{N}}$ is a Cauchy sequence.
Ad (6) Let $u^1, u^2 \in F$. Then $l_i := \lim_n \|x^n - u^i\|, i = 1, 2,$ exist due to (2). We have

$$\|x^n - u^1\|^2 = \|x^n - u^2\|^2 + \|u^1 - u^2\|^2 + 2\langle x^n - u^2|u^2 - u^1\rangle . \tag{6.30}$$

Suppose that w^1, w^2 are weak cluster points $(x^n)_{n\in\mathbb{N}}$. Then there exist subsequences $(x^{n_k})_{k\in\mathbb{N}}$, $(x^{n_l})_{l\in\mathbb{N}}$ such that $w^1 = w - \lim_k x^{n_k}$, $w^2 = w - \lim_l x^{n_l}$. Then from (6.30) we obtain

$$l_1^2 = l_2^2 + \|u^1 - u^2\|^2 + 2\langle w^1 - u^2|u^2 - u^1\rangle , \ l_1^2 = l_2^2 + \|u^1 - u^2\|^2 + 2\langle w^2 - u^2|u^2 - u^1\rangle .$$

Subtracting these two equations yields $0 = \langle w^2 - w^1|u^2 - u^1\rangle = \|w^2 - w^1\|$.
Ad (7) Clearly, since the sequence $(x^n)_{n\in\mathbb{N}}$ is bounded it has weak cluster points. From (6) we conclude that there is at most one weak cluster point in F.
Ad (8) If all sequential weak cluster points lie in F then $(x^n)_{n\in\mathbb{N}}$ converges weakly to a point of F due to (7). On the other hand, if $(x^n)_{n\in\mathbb{N}}$ converges to a point in F then all weak cluster points of $(x^n)_{n\in\mathbb{N}}$ belong to F.
Ad (9) Let $w = w - \lim_k x^{n_k}, w \in F$. We know $\hat{x} = \lim_k P_C(x^{n_k})$; see (4). Then from Kolmogorov's criterion we obtain $\langle x^{n_k} - P_C(x^{n_k})|w - P_C(x^{n_k})\rangle \leq 0$. Taking the limit $n_k \to \infty$ we obtain $\langle w - \hat{x}|w - \hat{x}\rangle \leq 0$. Hence $w = \hat{x} = \lim_n P_C(x^n)$.
Ad (10) See (8) and (9).
Ad (11) Let w be a weak cluster point and let $w = w - \lim_k x^{n_k}$. Suppose that x is another cluster point of $(x^n)_{n\in\mathbb{N}}$, $x \neq w$. Let $x = \lim_l x^{n_l}$. Let $\varepsilon = \frac{1}{2}\|w - x\| > 0$. Then there is a $N \in \mathbb{N}$ with $\|x^{n_k} - w\| < \varepsilon, \|x^{n_l} - x\| < \varepsilon$ for all $k, l \geq N$. Let $n_l > n_k \geq N$. Then

$$2\varepsilon = \|x - w\| \leq \|x - x^{n_l}\| + \|x^{n_l} - w\| < 2\varepsilon$$

due to the Fejéer monotonicity. This is a contradiction. Therefore $(x^n)_{n\in\mathbb{N}}$ converges to w.
Ad (12) Let $n \in \mathbb{N}, t \in \mathbb{R}$. Set $y_t := tP_C(x^0) + (1-t)P_C(x^n)$. Since F is affine we have $y_t \in F$. Then

$$\begin{aligned}
t^2\|P_F(x^n) - P_F(x^0)\|^2 &= \|P_F(x^n) - y_t\|^2 \\
&\leq \|x^n - P_F(x^n)\|^2 + \|P_F(x^n) - y_t\|^2 \\
&= \|x^n - y_t\|^2 \ \leq \ \|x^0 - y_t\|^2 \\
&\leq \|x^0 - P_F(x^0)\|^2 + \|P_F(x^0) - y_t\|^2 \\
&= \operatorname{dist}(x^0, F)^2 + (1-t)^2\|P_F(x^n) - P_F(x^0\|^2
\end{aligned}$$

Therefore, $(2t-1)\|P_F(x^n)-P_F(x^0)\|^2 \le \text{dist}(x^0,F)^2$. since t was arbitrary in $\mathbb{R}$ we have $\|P_F(x^n)-P_F(x^0\|=0$.
Ad (13) From the assumption we conclude that there exists x with $x = w-\lim_n x^n$; since F is weakly closed, $x \in F$. Then, since P_F is weakly continuous (see (11) in Theorem 3.3), $x = P_F(x) = w - \lim P_F(x^n) = P_F(x^0)$. ■

Theorem 6.23. *Let $\mathcal{H}$ be a Hilbert space space and let C, D be nonempty closed convex subsets of $\mathcal{H}$. Let $x^0 \in \mathcal{H}$ and let the sequence $(x^n)_{n\in\mathbb{N}_0}$ be defined by*

$$x^{2n+2} := P_D(x^{2n+1}),\ x^{2n+1} := P_C(x^{2n}),\ n \in \mathbb{N}_0,\ x^0 := x \in \mathcal{X} \text{ given}. \tag{6.31}$$

We have:

(a) If $C \cap D \neq \emptyset$ then the sequence $(x^n)_{n\in\mathbb{N}_0}$ converges weakly to a point of $C \cap D$.

(b) If int$(C \cap D) \neq \emptyset$ then the sequence $(x^n)_{n\in\mathbb{N}_0}$ converges strongly to a point of $C \cap D$.

Proof:
Ad (a) Let $y \in C \cap D$. Then, for any $x \in \mathcal{H}$, we have

$$\|P_C(x)-y\| = \|P_C(x)-P_C(y)\| \le \|x-y\|,\ \|P_D(x)-y\| = \|P_D(x)-P_D(y)\| \le \|x-y\|.$$

since x^{k+1} is either $P_C(x^k)$ or $P_D(x^k)$ we have $\|x^{k+1}-y\| \le \|x^k-y\|$. This shows that the sequence $(x^n)_{n\in\mathbb{N}_0}$ is Fejér monotone with respect to $C \cap D$. By (3) in Lemma 6.22 the sequence $(x^n)_{n\in\mathbb{N}_0}$ is bounded. Therefore, $(x^n)_{n\in\mathbb{N}_0}$ has weak cluster points.
Let x be such a weak cluster point and suppose that $x = w - \lim_k x^{n_k}$. Taking a subsequence again if necessary we may assume that $(x^{n_k})_{k\in\mathbb{N}}$ belongs to either C or D. We assume $x^{n_k} \in C, k \in \mathbb{N}$. Then $(P_D(x^{n_k})_{k\in\mathbb{N}}$ converges weakly to x too. Indeed, we know from (c) in Theorem 3.1

$$\|P_D(x^{n_k})-x^{n_k}\|^2 \le \|x^{n_k}-u\|^2 - \|P_D(x^{n_k})-u\|^2,\ k \in \mathbb{N},$$

where $U \in D$. Hence with $x^{n_k+1} = P_D(x^{n_k}), k \in \mathbb{N}$,

$$\|x^{n_k+1}-x^{n_k}\|^2 \le \|\|x^{n_k}-u\|^2 - \|x^{n_k+1}-u\|^2,\ k \in \mathbb{N},$$

and since $(x^n)_{n\in\mathbb{N}}$ is Fejér monotone we may conclude $\lim_k(x^{n_k+1}-x^{n_k}=0$. This implies $w-\lim_k x^{n_k+1} = x$. Since D is weakly closed we obtain $x \in D$. Now, we have $x \in C \cap D$. Thus, we have shown that all weak cluster points of $(x^n)_{n\in\mathbb{N}_0}$ belong to $C \cap D$.
since $(x^n)_{n\in\mathbb{N}_0}$ has at most one weak cluster point in $C \cap D$ by (4) in Lemma 6.22 the sequence $(x^n)_{n\in\mathbb{N}_0}$ converges to a point of $C \cap D$.
Ad (b) When int$(C \cap D) \neq \emptyset$ it follows from (7) in Lemma 6.22 that $(x^n)_{n\in\mathbb{N}_0}$ converges in norm. ■

The result in Theorem 6.23 is the result of Bregman in 1965; see [11]. In general, we cannot expect that the sequence in (a) of Theorem 6.23 converges to a specific point of $C \cap D$. Notice that this is the case when C, D are affine subspaces with nonempty intersection.

Example 6.24. *Whether the alternating projection algorithm would converge in norm under the assumption* $C \cap D \neq \emptyset$ *only, was a long-standing open problem. Recently Hundal [25] constructed an example showing that this is not the case. Here is the formulation of the example.*
Consider in l_2 *the standard „basis“* $\{e^k : k \in \mathbb{N}\}$. *Define* $\nu : [0, \infty) \longrightarrow l_2$ *by*

$$\nu(r) := \exp(-100r^3)e^1 + \cos(\frac{1}{2}(r - [r])\pi)e^{[r]+2} + \sin(\frac{1}{2}(r - [r])\pi)e^{[r]+3}\,,\ r \in [0, \infty)\,,$$

where $[r]$ *denotes the integer part of* r. *Further define*

$$C := \{e^1\}^\perp\,,\ D := co(\{\nu(r) : r \geq 0\})\,.$$

Then we have $C \cap D \neq \emptyset$ *and the alternating projection algorithm constructs a sequence starting from* $x^0 := \nu(1)$ *which does not converge in norm. The proof ot this fact is lengthly and we omit the proof.* □

6.6 Averaged midpoint projection method

We consider the following iteration:

$$\text{Given } x^0 \in \mathcal{H} \text{ set } x^{n+1} := \big(\frac{1}{2}P_D + \frac{1}{2}P_C\big)(x^n)\,,\ n \in \mathbb{N}_0\,. \qquad (6.32)$$

x^0 is used as a starting point of the iteration. This method is called the **averaged midpoint iteration**. We obtain two sequences: $(u^n)_{n\in\mathbb{N}}$ with $u^n := P_C(x^n) \in C, n \in \mathbb{N}$ $(w^n)_{n\in\mathbb{N}}$ with $w^n := P_D(x^n) \in D, n \in \mathbb{N}$.

This type of projection methods has been analyzed by A. Auslender [3] (weak convergence) and S. Reich [28, 29]. In [7] the counterexample of Hundal is reformulated for the case considered here.

Let us a more general situation. Let $\mathcal{H}$ be a Hilbert space space and let $C_1, \dots, C_m$ be nonempty closed convex subsets of $\mathcal{H}$. Here we use the trick to embed the problem in a product space and to apply the theorem above.
Let

$$\mathcal{H}^m := \{(x = (x_1, \dots, x_m) : x_i \in \mathcal{H}, i = 1, \dots, m\}$$

be the product space of m copies of $\mathcal{H}$. This is a Hilbert space with the inner product

$$\langle x|y\rangle := \sum_{i=1}^m w_i\langle x_i|y_i\rangle$$

where the weigth numbers w_i are all positive. Define

$$C := C_1 \times \cdots \times C_m\,,\ D := \{(x_1, \dots, x_m) \in \mathcal{H}^m : x_1 = x_2 = \cdots = x_m\}\,.$$

Then C, D are nonempty closed convex subsets of $\mathcal{H}^m$ and

$$x \in \cap_{i=1}^m C_i \iff (x, \dots, x) \in C \cap D\,.$$

Applying the projection algorithm to the sets C, D leads to the following method in the product space .

$$x^{n+1} := (\sum_{i=1}^{m} b_i P_{C_i})(x^n) \text{ where } b_i = \frac{w_i}{w},, i = 1, \ldots, m, w := \sum_{j=1}^{m} w_j \tag{6.33}$$

Theorem 6.25. *Let $\mathcal{H}$ be a Hilbert space space and let $F_1, \ldots, F_m$ be nonempty closed convex subsets of $\mathcal{H}$. Suppose that $\cap_{i=1}^m F_i \neq \emptyset$. Let the sequence $(x^n)_{n\in\mathbb{N}_0}$ be defined by the iteration* (6.33). *Then the sequence $(x^n)_{n\in\mathbb{N}_0}$ converges weakly to a point of $\cap_{i=1}^m F_i$.*

Proof:
Follows from Theorem 6.23. ■

When the interior of $\cap_{i=1}^m C_i$ is nonempty we also have that the iteration (6.33) converges in norm. However, since D does not have interior point this conclusion cannot be derived from Theorem 6.23. Rather it has to be proved by directly showing that the approximation sequence is Fejér monotone with respect to $\cap_{i=1}^m C_i$.

Theorem 6.26. *Let $\mathcal{H}$ be a Hilbert space space and let $U_1, \ldots, U_m$ be nonempty closed convex subsets of $\mathcal{H}$. Suppose that int$(\cap_{i=1}^m U_i) \neq \emptyset$. Let the sequence $(x^n)_{n\in\mathbb{N}_0}$ be defined by the iteration* (6.33). *Then the sequence $(x^n)_{n\in\mathbb{N}_0}$ converges strongly to a point of $\cap_{i=1}^m U_i$.*

Proof:
Let $y \in \cap_{i=1}^m U_i$. Then

$$\begin{aligned} \|x^{k+1} - y\| &= \left\| \Big(\sum_{i=1}^m b_i P_{U_i}\Big)(x^k) - y \right\| = \left\| \sum_{i=1}^m b_i (P_{U_i}(x^k) - P_{U_i}(y)) \right\| \\ &\leq \sum_{i=1}^m b_i \|P_{U_i}(x^k) - P_{U_i}(y)\| \leq \sum_{i=1}^m b_i \|x^k - y\| == \|x^k - y\| . \end{aligned}$$

Thus, the sequence $(x^n)_{n\in\mathbb{N}_0}$ is Fejér monotone with respect to $\cap_{i=1}^m U_i$. Now, the result follows from Theorem 6.23 and 6.25. ■

6.7 Dykstra's algorithm

In this section we want to solve the feasibility problem again:

$$\text{Given subsets } C, D \text{ of a Hilbert space } \mathcal{H} \text{ find a point in } x \in C \cap D \tag{6.34}$$

We describe the algorithm of Dykstra [18][2] to solve this problem. The basic idea is again von Neumann's iteration but more complex calculations are made. Here is Dykstra's algorithm:

[2]Do not confuse this algorihm with Dijkstra's algorithm in graph theory

Algorithm 6.2 Dyksta's algorithm

Given subsets C, D of a Hilbert space $\mathcal{H}$ and a starting point $x := x^0$. This algorithm computes a sequence $(u^n)_{n\in\mathbb{N}_0}$ in C and a sequence $(w^n)_{n\in\mathbb{N}_0}$ in D and sequences $(p^n)_{n\in\mathbb{N}_0}, (q^n)_{n\in\mathbb{N}_0}$ as follows:

(1) $p^0 := q^0 := \theta, w^0 := x, n := 1$.

(2) $u^n := P_C(w^{n-1} + p^{n-1}),\ p^n := w^{n-1} + p^{n-1} - u^n$.

(3) $w^n := P_D(u^n + q^{n-1}),\ q^n := u^n + pqn - 1 - w^n$.

(4) $n := n + 1$ and go to line (2).

This method is called **Dykstra's algorithm**. It was rediscovered by Han [23] applying duality arguments. Dykstra [18] suggested an algorithm which solves the problem for closed convex cones. Boyle and Dykstra [10]showed that Dykstra's algorithm solves the problem for general closed convex sets in a Hilbert space . In the specific situation that C, D are closed subspaces this algorithm coincides with the alternating method von Neumann. A parallel version of the algorithm was developed by Gaffke and Mathar [21].

Actually, we obtain four sequences from Algorithm 6.7. $(u^n)_{n\in\mathbb{N}}$ with $u^n \in C, n \in \mathbb{N}$ $(w^n)_{n\in\mathbb{N}}$, with $w^n) \in D, n \in \mathbb{N}$, $(p^n)_{n\in\mathbb{N}}, (q^n)_{n\in\mathbb{N}}$. What we will observe in a first step of the analysis of the algorithm is the fact that $p^n, q^n, n \in \mathbb{N}$, are outer normals with respect to C, D respectively: $p^n \in N(u^n, C), q^n \in N(w^n, D)$.

There are two cases to consider:

Consistence $C \cap D \neq \emptyset$.

Inconsistence $C \cap D = \emptyset$.

As we know, in the consistent case the alternating projection methods converges weakly to a point of the intersection $C \cap D$. If this intersection is no singleton this point is no specific point of the intersection. The algorithm of Dykstra which is the alternate projection method with some correction terms in each iteration step converges weakly to a projection onto the intersection of the starting point of the iteration.

In the inconsistent case one may assume that the method of alternate projection converges at least weakly to two points $u \in C, v \in D$ which realize the distance $\text{dist}(C, D)$ between C and D.

In the following we use the following data for the problem (6.34):

- $\mathcal{H}$ is a Hilbert space (of dimension ≥ 2).
- C, D are nonempty closed convex subsets of $\mathcal{H}$.
- $v := P_{\overline{D-C}}(\theta)$.
- $\delta := \text{dist}(C, D)$.
- $C_* := \{u \in C : \text{dist}(u, D) = \delta\},\ D_* := \{w \in D : \text{dist}(w, C) = \delta\}$.

Lemma 6.27.

(1) The infimum $\delta = \inf_{u \in C, w \in D} \|u - w\|$ *is attained iff* $v \in D - C$.

(2) The infimum $\delta = \inf_{u \in C, w \in D} \|u - w\|$ *is attained if* $D - C$ *is closed.*

(3) $\|v\|^2 \le \delta^2 \le \langle v, w - u \rangle$ *for all* $u \in C, w \in D$.

(4) $C_* = Fix(P_C \circ P_D) D_* = Fix(P_D \circ P_C)$.

(5) C_*, D_* *are closed convex sets.*

(6) If δ *is attained then* $C_*.D_*$ *are nonempty.*

(7) If C_* *or* D_* *is nonempty, then* δ *is attained and we have for* $u \in C_*, w \in D_*$

$$P_D(u) = u + v, P_C(w) = w - v, C_* + v = D_*, C_* = C \cap (D - v), D_* = (C + v) \cap D .$$

(8) If δ *is attained then*

$$\langle u - u' | v \rangle \le 0, u \in C, u' \in C_*, \langle w - w' | - v \rangle \le 0, w \in D, w' \in D_*,$$

and

$$\langle u' - u'' | v \rangle = \langle w' - w'' | v \rangle, u', u'' \in C_*, w', w'' \in D_* .$$

(9) If δ *is attained then* $P_{D_*}(x) = P_{C_*}(x) + v$.

(10) If δ *is attained and if* C, D *are affine closed sets then* C_*, D_* *are also affine sets and* $v \in C^\perp \cap D^\perp$.

Proof:
Ad (1)

■

Lemma 6.28. *Suppose that* $(u^n)_{n \in \mathbb{N}}, (w^n)_{n \in \mathbb{N}}$ *are sequences in* C *and* D *respectively and assume that* $\lim_n \|u^n - w^n\| = \delta$. *Then* $\lim_n u^n - w^n = v$ *and every weak cluster point of* $(u^n)_{n \in \mathbb{N}}, (w^n)_{n \in \mathbb{N}}$ *belong to* C_* *and* D_* *respectively. Consequentely, if* δ *is attained then* $\lim_n \|u^n\| = \lim_n \|w^n\| = \infty$.

Proof:

■

To explain how the algorithm works let us consider a small example.

Example 6.29. *Consider in the euclidean space* $\mathbb{R}^2$ *the sets*

$$C := \{(x, y) : y \le 0\}, \; D := \{(x, y) : x + y \le 0\} .$$

With the starting vector $z^0 := w^0 := (1, 1.5)$ *the first two steps of the alterante projection method and the method of Dykstra produce identical iterates:* $u^1 = (1, 0), w^1 = (0.5, -0.5)$. *The method of alternating projections stops with the vector* $z := w^1 = (0.5, -0.5) \in C \cap D$. *But we see that* $z \ne P_{C \cap D}(z^0)$.

The method of Dykstra does not stop here. We compute

Step 1 $u^1 = P_C(w^0) = (1,0), p^1 = (0,1.5), w^1 = P_D(u^1+q^0) = (0.5,-0.5), q^1 = (0.5, 0.5)$.

Step 2 $u^2 = P_C(w^1+p^1) = (0.5,0), p^2 = (0,1), w^2 = P_D(u^2+q^1) = (0.25,-0.25), q^2 = (0.75, 1.25)$.

Step 3 $u^3 = P_C(w^2+p^2) = (0.25,0), p^3 = (0,0.75), w^3 = P_D(u^2+q^1) = (xx,xx), q^2 = (xx, xx)$.

One can see that the approximations u^n *converge on the* x*-axis to* $(0,0)$ *and the approximations* w^n *converge on the line* $x+y=0$ *to* $(0,0)$. *Notice* $(0,0) = P_{C\cap D}(z^0)$. □

Now we want to analyze the convergence properties of the algorithm, especially we want to find a result which makes it possible to formulate a sound **stopping rule**. Here is the main theorem concerning the „power“ of the algorithm.

Theorem 6.30. *Let* $\mathcal{H}$ *be a Hilbert space , let* C, D *be nonempty closed convex subsets of* $\mathcal{H}$ *and let* $x \in \mathcal{H}$. *We set* $x^0 := x, v := P_{\overline{D-C}}(\theta)$, $\delta := dist(C,D)$ *and* $C_* := \{u \in C : dist(u,D) = \delta\}, D_* := \{w \in D : dist(w,C) = \delta\}$. *Let the sequences* $(u^n)_{n\in\mathbb{N}_0}$, $(w^n)_{n\in\mathbb{N}_0}$, $(p^n)_{n\in\mathbb{N}_0}$, $(q^n)_{n\in\mathbb{N}_0}$ *constructed by the Algorithm 6.7. Then the following statements hold:*

(a) $v = \lim_n w^n - u^n = \lim_n w^n - u^{n+1}$ *where* $v := P_{\overline{D-C}}(\theta)$ *and* $\|v\| = dist(C,D)$.

(b) $\lim_n \frac{1}{n}u^n = \lim_n \frac{1}{n}w^n = 0, \lim_n \frac{1}{n}p^n = v, \lim_n \frac{1}{n}w^n = -v$.

(c) *If* $dist(C,D)$ *is not attained, then* $\lim_n \|u^n\| = \lim_n \|w^n\| = \infty$.

(d) *If* $dist(C,D)$ *is attained, then* $\lim_n u^n = P_{C_*}(x), \lim_n w^n = P_{D_*}(x)$ *where*

$$C_* := \{u \in C : dist(u,D) = \delta\}, D_* := \{w \in D : dist(w,C) = \delta\}$$

are nonempty closed convex sets with $C_* + v = D_*$.

Now, the important information is a statement concerning the convergence of Dykstra's algorithm:

The case $C \cap D \neq \emptyset$ each sequence $(u^n)_{n\in\mathbb{N}}$, $(w^n)_{n\in\mathbb{N}}$ converges strongly to a point in $C \cap D$; see [18].

The case that C, D **are a cones** each sequence $(u^n)_{n\in\mathbb{N}}$, $(w^n)_{n\in\mathbb{N}}$ converges strongly to $P_{C\cap D}(x^0)$; see [18].

The case of two subspace s each sequence $(u^n)_{n\in\mathbb{N}}$, $(w^n)_{n\in\mathbb{N}}$ converges strongly to $P_{C\cap D}(x^0)$, as outer normal vectors we may use the null vector; see [10].

The case that C **or** D **is a cone** each sequence $(u^n)_{n\in\mathbb{N}}$, $(w^n)_{n\in\mathbb{N}}$ converges strongly to $P_{C\cap D}(x^0)$; see [10].

The case $C \cap D \neq \emptyset$ each sequence $(u^n)_{n\in\mathbb{N}}$, $(w^n)_{n\in\mathbb{N}}$ converges strongly to $P_{C\cap D}(x^0)$; see [21].

The case $C \cap D = \emptyset$ The difference sequence $(u^n - w^n)_{n\in\mathbb{N}}$ converges strongly to $\text{dist}(C,D)$ and if $\text{dist}(C,D)$ is attained then $(u^n)_{n\in\mathbb{N}}$, $(w^n)_{n\in\mathbb{N}}$ converge to $u \in C$ and $w \in D$ respectively with $\lim_n \|u^n - w^n\| = \|v\| = \text{dist}(C,D)$ where $v \in P_{\overline{D-C}}(\theta)$; see [8].

6.8 The Douglas Rachford reflection method

Let $\mathcal{K}$ be a Hilbert space and let K_1, K_2 be nonempty closed convex subsets of $\mathcal{K}$. Then the reflector of K_1, K_2 is $R_{K_1} := 2P_{K_1} - I$ and $R_{K_2} := 2P_{K_2} - I$ respectively. These two mappings are well defined and Fan be used to construct the **Douglas Rachford reflection**

$$T_{K_2,K_1} := \frac{1}{2}(I + P_{K_2}P_{K_1}).$$

since we want to study the relation of T_{K_2,K_1} with the problem to construct a point in $K_1 \cap K_2$ we present the following Lemma.

Lemma 6.31. *Let $\mathcal{K}$ be a Hilbert space and let K_1, K_2 be nonempty closed convex subsets of $\mathcal{K}$. Then $x \in Fix(T_{K_2,K_1})$ if and only if $P_{K_1}(x) = P_{K_1}P_{K_2}(x)$.*

Proof:

$$\begin{aligned} x \in \mathrm{Fix}(T_{K_2,K_1}) &\iff 2x = x + P_{K_2}P_{K_1}(x) \iff x = R_{K_2}R_{K_1}(x) \\ &\iff x = 2P_{K_2}R_{K_1}(x) - R_{K_1}(x) \iff x = 2P_{K_2}R_{K_1}(x) - 2P_{K_1}(x) + x \\ &\iff P_{K_1}(x) = P_{K_2}R_{K_1}(x) \end{aligned}$$

■

From the result of Lemma 6.31 we conlude that $x \in \mathrm{Fix}(T_{K_2,K_1})$ implies $P_{K_1}(x) \in K_1 \cap K_2$. This is the key for developing an method to find points in $K_1 \cap K_2$: if $x \in \mathrm{Fix}(T_{K_2,K_1})$ then $P_{K_1}(x)$ belongs to $K_1 \cap K_2$.

Using the above considerations we consider the following iteration:

$$\text{Given } x^0 \in \mathcal{H} \text{ set } x^{n+1} := T_{D,C}(x^n) = \frac{1}{2}\big(I + R_D R_C\big)(x^n)\,,\ n \in \mathbb{N}_0\,. \tag{6.35}$$

x^0 is used as a starting point of the iteration. This iteration is called the **Douglas Rachford reflection method**. We obtain a sequence $(u^n)_{n\in\mathbb{N}}$ in $\mathcal{H}$. We hope that this sequence converges to a fixed point of $T_{D,C}$. Here are results concerning the convergence poperties of the Douglas Rachford iteration.

The case $C \cap D \neq \emptyset$ The sequence $(u^n)_{n\in\mathbb{N}}$ converges weakly to a fixed point u of $T_{D,C}$ and $P_C(u) \in C \cap D$.

The case $C \cap D = \emptyset$ The sequence $(u^n)_{n\in\mathbb{N}}$ diverges with $\lim_n \|u^n\| = \infty$.

The case of two subspaces The sequence $(u^n)_{n\in\mathbb{N}}$ converges strongly to a fixed point u of $T_{D,C}$ and $P_C(u) \in C \cap D$.

6.9 The averaged alternating reflection method

6.10 The Douglas Rachford reflection method: order of convergence

6.11 The alternating projection method for finding commom solutions of variational inequalities

6.12 The ART-algorithm

6.13 Some variants of the projection method

6.14 Randomized alternate projection method

6.15 Appendix:

6.16 Conclusions and comments

6.17 Exercises

1.) Let $\mathcal{H}$ be a Hilbert space and let H be a hyperplane of $\mathcal{H}$ with $\theta \in H$. Then the reflector $P_H := 2P_H - I$ is a surjective linear isometry and satisfies $R_H = R_H^* = P_H^{-1}$.

2.) Let $\mathcal{H}$ be a Hilbert space, let H be a hyperplane of $\mathcal{H}$ with $\theta \in H$ and let C be a nonempty closed convex subset of $\mathcal{H}$. Set $D := R_H(C)$. Then the following hold true:

   (1) D is closed and convex.

   (2) $P_D = R_H \circ P_C \circ R_H$.

   (3) $(P_H \circ P_C)\big|_H = \frac{1}{2}(P_C + P_H)$.

3.) Let $\mathcal{H}$ be a Hilbert space and let U_1, U_2 be closed subspace s of $\mathcal{H}$. Show that $P_{U_1}P_{U_2} = P_{U_2}P_{U_1}$ if and only if $P_{U_1}P_{U_2} = P_{U_1 \cap U_2}$.

4.) Let $\mathcal{H}$ be a Hilbert space and let U_1, U_2 be closed subspace s of $\mathcal{H}$. Show that $P_{U_1}P_{U_2} = \theta$ if and only if $U_1 \perp U_2$.

5.) Let $\mathcal{H}$ be a Hilbert space and let U_1, U_2 be closed subspace s of $\mathcal{H}$. Suppose $P_{U_1}P_{U_2} = P_{U_2}P_{U_1}$. Show

$$P_{\overline{U_1+U_2}} = P_{U_1} + P_{U_2} - P_{U_1}P_{U_2}.$$

6.) Let $\mathcal{H}$ be a Hilbert space and let U_1, U_2 be closed subspace s of $\mathcal{H}$. If $U_1 \subset U_2^{\perp}$ then $U_1 + U_2$ is closed and $P_{U_1+U_2} = P_{U_1} + P_{U_2}$.

7.) Let $\mathcal{H}$ be a Hilbert space and let U_1, U_2 be closed subspace s of $\mathcal{H}$. Show that $P_{U_1}P_{U_2} = P_{U_2}P_{U_1}$ if and only if $U_1 = U_1 \cap U_2 + U_1 \cap U_2^\perp$.

8.) Let $\mathcal{H}$ be a Hilbert space and let U_1, U_2 be closed subspace s of $\mathcal{H}$. Suppose $\mathcal{H} = U_1 \oplus U_2$. Show:

$$\mathcal{H} = U_1^\perp + U_2^\perp \,,\; c_0(U_1, U_2) = c_0(U_1^\perp, U_2^\perp) < 1 \,.$$

9.) Find in $\mathcal{H} := \mathbb{R}^3$ endowed with the l_2-norm closed subspace s U_1, U_2 such that

$$c_0(U_1, U_2) = 1 \,,\; c_0(U_1 \perp, U_2^\perp) = 0 \,.$$

10.) Let $\mathcal{H}$ be a Hilbert space and let U_1, U_2 be closed subspace s of $\mathcal{H}$. Show that $c(U_1, U_2) = 0$ if and only if $P_{U_1}P_{U_2} = P_{U_2}P_{U_1}$.

11.) Let $\mathbb{R}^2$ endowed with the l_2-norm and let

$$C := S_{\frac{1}{2}} := \{x = (x_1, x_2) \in \mathbb{R}^2 : |x_1|^{\frac{1}{2}} + |x_2|^{\frac{1}{2}} = 1\} \,.$$

(Unit sphere in the space $\mathbb{R}^2$ endowed with the metriF

$$d((x_1, x_2), (y_1, y_2)) := |x_1 - y_1|^{\frac{1}{2}} + |x_2 - y_2|^{\frac{1}{2}}) \,.$$

Compute for $(u, v) \in A := [0, 1] \times [0, 1]$ a best approximation $P_C(u, v)$.

12.) Let $\mathcal{X}$ be a Banach space and let U, V closed linear subspace s of $\mathcal{X}$.

(a) Show: If the dimension of U is finite then $U + V$ is closed.

(b) Now consider for $\mathcal{X}$ the space c_0 of the real sequences converging to zero. c_0 is endowed with the supremum norm. Let

$$\begin{aligned} U &:= \{(x^n)_{n\in\mathbb{N}} \in c_0 : x^n = nx^{n-1}, n \text{ even}\}, \\ V &:= \{(x^n)_{n\in\mathbb{N}} \in c_0 : x^n = 0, n \text{ odd}\}. \end{aligned}$$

Show: $U + V$ is dense in c_0, but not closed.

13.) Let $\mathcal{H}$ be a Hilbert space and let U, V be linear subspace s with $U \cap V = \{\theta\}$. We say that a sharpened Cauchy Schwarz inequality holds if

$$\langle u|v\rangle \le \gamma \|u\| \|v\| \,,\; u \in U, v \in V,$$

with $0 \le \gamma < 1$ is true. Then.

(a) A sharpened Cauchy Schwarz inequality holds if $\dim \mathcal{H} < \infty$.

(b) If $\dim \mathcal{H} = \infty$ no sharpened Cauchy Schwarz inequality holds in general.

14.) Let $\mathcal{H}$ be a Hilbert space and let U_1, U_2 are closed linear subspace s; $U := U_1 \cap U_2$. Then the following conditions are equivalent:

(a) $P_{U_1^\perp}(U_2) := \{P_{U_1^\perp}(v) : v \in U_2\}$ is closed.

(b) $U_1 + U_2$ is closed.

(c) $P_{U_2^\perp}(U_2) := \{P_{U_2^\perp}(v) : v \in U_1\}$ is closed.

15.) Let $\mathcal{H}$ be a Hilbert space and let U_1, U_2 are closed linear subspace s. Set $U := U_1 \cap U_2$. Then the following conditions are equivalent:

(a) $U_1 + U_2$ is closed.

(b) $U_1 \cap U^\perp + U_2 \cap U^\perp$ is closed.

(c) $U_1^\perp + U_2^\perp$ is closed.

Bibliography

[1] Y.I. Alber. Generalized projection operators in Banach spaces: properties and applications. *Funkt. Differential Equations Israel*, 1:1–21, 1994.

[2] L. Aronszajn. Theory of reproducing kernels. *Trans. Amer. Math. Soc.*, 68:337–404, 1950.

[3] A. Auslender. Methodes numerique pour la resolution des problemes d'optimisation avec constraints, 1969.

[4] C. Badea and Yu.I. Lyubich. Geometric, spectral and asymptotic properties of averaged products of projections in Banach spaces. *Studia Math.*, 201:21–35, 2010.

[5] C. Badea, S. Grivaux and V. Müller. A generalization of the friedrichs angle and the method of alternating projections. *C. R. Acad. Sci. Paris, Ser. I*, 348:53–56, 2010.

[6] H. H. Bauschke, F. Deutsch, and H. Hundal. Characterizing arbitrarily slow convergence in the method of alternating projections. *Int. Trans-Oper. Res.*, pages 413–425, 2009.

[7] H. H. Bauschke, E. Matouskova, and S. Reich. Projection and proximal point methods: convergence results and counterexamples. *Nonlinear Analysis*, 56:715–738, 2004.

[8] H.H. Bauschke and J.M. Borwein. Dykstra's alternating projection algorithm for two sets. *J. of Approximation Theora*, 79:418–443, 1994.

[9] J.M. Borwein and M. Tam. Theory and applications of convex and non-convex feasibility problems, 2015. Spring school on variational Analysis VI Paseky and Jizerou.

[10] J.P. Boyle and R.L. Dykstra. A method for finding projections onto the intersection of convex sets in hilbert spaces. In *Advances in order restricted statistical inference*, pages 28–47. Iowa City, Springer, 1985.

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

[12] R.E. Bruck and S. Reich. Nonexpansive projections and resolvents of accretive operators in Banach spaces. *Houston J. Math.*, 3:459–470, 1977.

[13] G. Cimmino. Calcolo approssimato per le soluzioni dei sistemi di equazioni lineari. *Ric. Sci. (Roma)*, 9:326–333, 1938.

[14] F. Deutsch and H.S. Hundal. The rate of convergence for the method of alternating projections, II. *J. Math. Anal. Appl.*, 205:381–405, 1997.

[15] F. Deutsch and H.S. Hundal. The rate of convergence for the cyclic projections algorithm I: angles between convex sets. *J. Approximation Theory*, 142:36–55, 2006.

[16] F. Deutsch and H.S. Hundal. The rate of convergence for the cyclic projections algorithm II: Norms of nonlinear operators. *J. of Approximation Theory*, 142:56–82, 2006.

[17] J. Dixmier. Etudes sur les varietes et les operateurs de julia avec quelques applications. *Bull. Soc. Math. France*, 77:11–101, 1949.

[18] R.L. Dykstra. An algorithm for restricted least squares regression. *J. Amer. Statist. Assoc.*, 78:837–842, 1983.

[19] R. Escalante and M. Raydan. *Alternating projection methods.* 2011.

[20] K. Friedrichs. On certain inequalities and characteristic value problems for analytic functions and for functions of two variables. *Trans. Amer. Math. Soc.*, 41:321–364, 1937.

[21] N. Gaffke and R. Mathar. A cyclic projection algorithm via duality. *Metrika*, 36:29–54, !989.

[22] I. Halperin. The product of projection operators. *Acta Sc. Math. (Szeged)*, 23:96–99, 1962.

[23] S.P. Han. A successive projection method. *Math. Programming*, 40:1–14, 1988.

[24] G.N. Hounsfield. Computerized transverse axial scanning (tomography): P i. description of the system. *British J. Radiol.*, 46:1016–1022, 1973.

[25] H.S. Hundal. An alternating projection that does not converge in norm. *Nonlinear analysis*, 57:35–61, 2004.

[26] S. Kaczmarz. Angenäherte Auflösung von Systemen linearer Gleichungen. *Bull. Acad. Polon.*, A35:355–357, 1937.

[27] J.-P. Penot and R. Ratsimahalo. Characteritations of metric projections in Banach spaces and applications. 1998. Preprint.

[28] S. Reich. A limit theorem for projections. *Linear and Multilinear Algebra*, 13:281–290, 1983.

[29] S. Reich. Averaged mappings in the Hilbert ball. *J. of Math. Anal. and Appl.*, 109:199–206, 1985.

[30] J. von Neumann. *The geometry of orthogonal spaces, II.* Princeton University Press, 1950. Results first distributed 1933.