Math. Program., Ser. A (2015) 149:329–359
DOI 10.1007/s10107-014-0759-z

FULL LENGTH PAPER

Set intersection problems: supporting hyperplanes and quadratic programming

C. H. Jeffrey Pang

Received: 7 January 2013 / Accepted: 28 January 2014 / Published online: 7 February 2014

Abstract We study how the supporting hyperplanes produced by the projection process can complement the method of alternating projections and its variants for the convex set intersection problem. For the problem of finding the closest point in the intersection of closed convex sets, we propose an algorithm that, like Dykstra's algorithm, converges strongly in a Hilbert space. Moreover, this algorithm converges in finitely many iterations when the closed convex sets are cones in $\mathbb{R}^n$ satisfying an alignment condition. Next, we propose modifications of the alternating projection algorithm, and prove its convergence. The algorithm converges superlinearly in $\mathbb{R}^n$ under some nice conditions. Under a conical condition, the convergence can be finite. Lastly, we discuss the case where the intersection of the sets is empty.

Keywords Dykstra's algorithm · Best approximation problem · Alternating projections · Quadratic programming · Supporting hyperplanes · Superlinear convergence

Mathematics Subject Classification (1991) 90C30 · 90C59 · 47J25 · 47A46 · 47A50 · 52A20 · 41A50 · 41A65 · 46C05 · 49J53 · 65K10

Contents

C. H. J. Pang (✉)
Department of Mathematics, National University of Singapore,
Block S17 08-11, 10 Lower Kent Ridge Road, Singapore 119076, Singapore
e-mail: matpchj@nus.edu.sg

1 Introduction

For finitely many closed convex sets $K_1, \ldots, K_r$ in a Hilbert space X, the *set intersection problem* (SIP) is stated as:

$$\text{(SIP):} \quad \text{Find } x \in K := \bigcap_{i=1}^{r} K_i, \text{ where } K \neq \emptyset. \tag{1.1}$$

One assumption on the sets K_i is that projecting a point in X onto each K_i is a relatively easy problem.

A popular method of solving the SIP is the *Method of Alternating Projections* (MAP), where one iteratively projects a point through the sets K_i to find a point in K. Another problem related to the SIP is the *Best Approximation Problem* (BAP): Find the closest point to x_0 in K, that is,

$$\begin{aligned} \text{(BAP):} \quad & \min_{x \in X} \quad \|x - x_0\| \\ & \text{s.t.} \quad x \in K := \bigcap_{i=1}^{r} K_i. \end{aligned} \tag{1.2}$$

for closed convex sets K_i, $i = 1, \ldots, r$. One can easily construct an example in $\mathbb{R}^2$ involving a circle and a line such that the MAP converges to a point in K that is not $P_K(x_0)$. Fortunately, Dykstra's algorithm [7,15] reduces the problem of finding the projection onto K to the problem of projecting onto K_i individually by adding correction vectors after each iteration. It was rediscovered in [22] using mathematical programming duality. For more on the background and recent developments of the MAP and its variants, we refer the reader to [3,9,16], as well as [14, Chapter 9] and [10, Subsubsection 4.5.4].

We quote [13], where it is mentioned that the MAP has found application in at least ten different areas of mathematics, which include: (1) solving linear equations; (2) the Dirichlet problem which has in turn inspired the "domain decomposition" industry; (3) probability and statistics; (4) computing Bergman kernels; (5) approximating multivariate functions by sums of univariate ones; (6) least change secant updates; (7) multigrid methods; (8) conformal mapping; (9) image restoration; (10) computed tomography. See also [12] for more information.

One problem of the MAP and Dykstra's algorithm is slow convergence. A few acceleration methods were explored. The papers [8,18,21] explored the acceleration of the MAP using a line search in the case where K_i are linear subspaces of X. See [13] for a survey. One can easily rewrite the SIP as a *Convex Inequality Problem* (CIP):

$$\text{(CIP):}\quad \text{Find } x \in X \text{ satisfying } g(x) \le 0,$$

where $g : X \to \mathbb{R}^r$ is such that each $g_i : X \to \mathbb{R}$, where $i = 1, \ldots, r$, is convex: Just set $g_i(x)$ to be the distance from x to K_i. In the case where $X = \mathbb{R}^n$ and each $g_i(\cdot)$ is differentiable with Lipschitz gradient, the papers [19,20] proved a superlinear convergent algorithm for the CIP. They make use of the subgradients of $g(\cdot)$ to define separating hyperplanes to the feasible set, and make use of quadratic programming to achieve superlinear convergence. Another related work is [26], where the interest is on problems where r, the number of closed convex sets K_i, is large. In the case where $g_i(\cdot)$ were not assumed to be convex in the CIP, Fletcher and Leyffer [17] designed SQP algorithms that are globally convergent and handles the case of infeasible problems.

We elaborate on the quadratic programming approach. Given $x_1 \in X$ and the projection $x_2 = P_{K_1}(x_1)$, provided $x_2 \neq x_1$, a standard result on supporting hyperplanes gives us $K_1 \subset \{x \mid \langle x_1 - x_2, x\rangle \le \langle x_1 - x_2, x_2\rangle\}$. The aim of this work is make use of the supporting hyperplanes generated in the projection process to accelerate the convergence to a point in K. A relaxation of (1.2) is

$$\begin{aligned} \min_{x\in X} \quad & \|x - x_0\|^2 \\ \text{s.t.} \quad & \langle a_i, x\rangle \le b_i \text{ for } i = 1, \ldots, k, \end{aligned} \tag{1.3}$$

where each constraint $\langle a_i, x\rangle \le b_i$ corresponds to a supporting hyperplane obtained by the projection operation onto one of the sets K_i, where $1 \le i \le r$. Even though there is no nice solution for the minimizer of (1.3), we can make use of quadratic programming to find this minimizer in practice. Let $S = \text{span}\{a_i \mid i = 1, \ldots, k\}$, and let $V = \{v_1, \ldots, v_{k'}\}$ be a set of orthonormal vectors spanning S, where $k' \le k$. We can write $a_i = \sum_{j=1}^{k'} \alpha_{i,j} v_j$, $x_0 = y_0 + z_0$ and $x = \sum_{j=1}^{k'} \lambda_j v_j + z$, where $y_0 \in S$ and $z_0, z \in S^\perp$. Then (1.3) can be rewritten as

$$\begin{aligned} \min_{\lambda\in\mathbb{R}^{k'}, z\in S^\perp} \quad & \left\| \sum_{j=1}^{k'} \lambda_j v_j - y_0 \right\|^2 + \|z_0 - z\|^2 \\ \text{s.t.} \quad & \left\langle \sum_{j=1}^{k'} \alpha_{i,j} v_j, \sum_{j=1}^{k'} \lambda_j v_j \right\rangle \le b_i \text{ for } i = 1, \ldots, k. \end{aligned} \tag{1.4}$$

Therefore (1.3) can be easily solved using convex quadratic programming, especially when k and k' are small. (See for example [31, Chapter 16].)

The quadratic programming formulation (1.3) gathers information from the supporting hyperplanes to many of the closed convex sets K_i, and so is a good approximation to (1.2); the intersection of the halfspaces defined by the supporting hyperplanes can produce a set that is a better approximation of K than each K_i taken singly. Hence there is good reason to believe that (1.3) can achieve better convergence than simple variants of the MAP. As Fig. 1 illustrates, the supporting hyperplanes can provide a good outer estimate of the intersection K. Furthermore, as more constraints are added in the quadratic programming formulation (1.3), it is possible to use warm starts from previous iterations to accelerate convergence. In this paper, we shall only pursue the

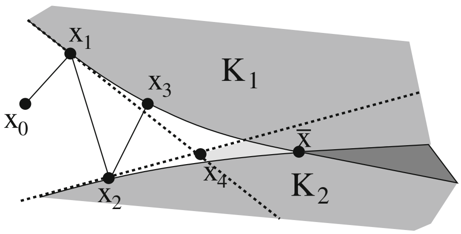

Fig. 1 The method of alternating projections on two convex sets K_1 and K_2 in $\mathbb{R}^2$ with starting iterate x_0 arrives at x_3 in three iterations. But the point x_4 generated by the cutting planes of K_1 and K_2 at x_1 and x_2 respectively is much closer to the point $\bar{x}$, especially when the boundary of K_1 and K_2 have fewer second order effects and when the angle between the boundary of K_1 and K_2 is small. On the other hand, the point x_3 is ruled out by the supporting hyperplane of K_2 passing through x_2

idea of supplementing the MAP with supporting hyperplanes and quadratic programming, but not on the details of the quadratic programming subproblem.

We remark that the idea using supporting halfspaces to approximate the set K was also considered before. In [33], Pierra suggested an extrapolation to find a point in the polyhedron produced by two projections, but not by using QP, and this idea was further studied in [5]. In [11], cutters were defined based on the property that the halfspaces generated contain the intersection of the sets, and studied as a generalization of the projection operation. In [6], the hyperplanes were used to simplify the projections onto the sets K_i rather than accelerating convergence.

1.1 Contributions of this paper

In this paper, we prove theoretical properties of the alternating projection method supplemented with the insight on supporting hyperplanes. Sections 3–6 are mostly independent of each other.

First, we propose Algorithm 3.1 for the Best Approximation Problem (1.2) in Sect. 3. We prove norm convergence, and the finite convergence of Algorithm 3.1 with (3.1b) when $K_l \subset \mathbb{R}^n$ have a local conic structure and satisfy a normal condition. We also show that the normal condition cannot be dropped.

In Sect. 4, we propose modifications of the alternating projection algorithm for the set intersection problem (1.1), and prove their convergence. We also prove the superlinear convergence of a modified alternating projection algorithm in $\mathbb{R}^2$.

In Sect. 5, we prove the most striking result of this paper, which is the superlinear convergence of an algorithm for the set intersection problem (1.1) in $\mathbb{R}^n$ under theoretically reasonable conditions. Instead of assuming smoothness of the sets similiar to [19,20], we make the much weaker assumption of local metric inequality (commonly known as linear regularity; See Definition 5.9), but at the expense of having to solve large QPs. (Note also that if the size of the QPs were too small, then there is no superlinear convergence in the case of a line intersecting a plane in $\mathbb{R}^3$. See Example 4.7.) The convergence can be finite if there is a local conic structure at the limit point. The proofs of superlinear convergence are quite different from the proof in Sect. 4.

Lastly, in Sect. 6, we discuss the behavior of Algorithm 3.1 in the case when the intersection of the closed convex sets is empty.

1.2 Notation

We write down some of the relatively standard notation in convex analysis that will be used for the later sections. For more details, we refer the reader to [23,34].

$\mathbb{B}_r(x)$ The closed ball with radius r and center x. If $r = 1$ and $x = 0$, then we simply write $\mathbb{B}$.
$P_C(\cdot)$ The projection operation onto a set $C \subset X$, where X is a Hilbert space.

Let X be a Hilbert space. For a closed convex set $C \subset X$ and a point $x \in C$,

$N_C(x)$ The *normal cone* $N_C(x)$ at x is the set $\{v \mid \langle v, y - x\rangle \leq 0 \text{ for all } y \in C\}$.
$T_C(x)$ The *tangent cone* $T_C(x)$ at x is the set of all vectors $v \in X$ such that there exists sequences $x_k \to x$ and $t_k \searrow 0$ such that $\frac{x_k - x}{t_k} \to v$.

We say that a set $C \subset X$ is a *cone* if $kC = C$ for all $k > 0$ and $0 \in C$. A cone that is convex is a *convex cone*. For a convex cone $C \subset X$,

C^- The *negative polar cone* C^- is defined by $\{v \mid \langle v, x\rangle \leq 0 \text{ for all } x \in C\}$.

We also recall the definition of linear and superlinear convergence.

Definition 1.1 (*Superlinear convergence*) For a sequence of points $\{x_i\}$ converging to $\bar{x}$, we say that the convergence is *linear* if

$$\limsup_{i\to\infty} \frac{\|x_{i+1} - \bar{x}\|}{\|x_i - \bar{x}\|} \in (0, 1),$$

and *superlinear* if

$$\limsup_{i\to\infty} \frac{\|x_{i+1} - \bar{x}\|}{\|x_i - \bar{x}\|} = 0.$$

2 Some useful results

In this section, we recall or prove some useful results that will be useful in two or more of the sections later. The reader may wish to skip this section and come back to refer to the results as needed.

The result below shows that supporting hyperplanes near a point in a convex set behave well.

Theorem 2.1 (Supporting hyperplane near a point) *Suppose $C \subset \mathbb{R}^n$ is convex, and let $\bar{x} \in C$. Then for any $\epsilon > 0$, there is a $\delta > 0$ such that for any point $x \in [\mathbb{B}_\delta(\bar{x}) \cap C]\backslash\{\bar{x}\}$ and supporting hyperplane A of C with unit normal $v \in N_C(x)$ at the point x, we have $\frac{d(\bar{x}, A)}{\|x - \bar{x}\|} \leq \epsilon$.*

Since $d(\bar{x}, A) = -\langle v, \bar{x} - x\rangle$, the conclusion of this result can be replaced by $0 \leq \frac{-\langle v, \bar{x} - x\rangle}{\|\bar{x} - x\|} \leq \epsilon$ instead.

Proof We refer to Fig. 2. For a given $\epsilon_1 > 0$, there is some $\delta > 0$ such that if $x \in \mathbb{B}_\delta(\bar{x}) \cap C$ and $v \in N_C(x)$ is a unit vector, then by the outer semicontinuity

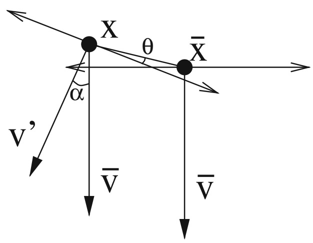

Fig. 2 Diagram in the proof of Theorem 2.1

of the normal cone mapping (see [35, Proposition 6.6]), there is some unit vector $\bar{v} \in N_C(\bar{x})$ such that $\|v - \bar{v}\| < \epsilon_1$. This means that the angle between v and $\bar{v}$ is at most $2\sin^{-1}(\epsilon_1/2)$.

One can easily check that $x - \bar{x}$ is not a multiple of $\bar{v}$. Consider the two dimensional affine space that contains the vector $\bar{v}$ and the points x and $\bar{x}$, and project the point $x+v$ onto this affine space. Let this projection be $x + v'$. It is easy to check that the angle between v' and $\bar{v}$, marked as α in Fig. 2, is bounded from above by $2\sin^{-1}(\epsilon_1/2)$. (The lines with arrows at both ends passing through x and $\bar{x}$ respectively represent the intersection of supporting hyperplanes with the two dimensional affine space.)

The angle θ in Fig. 2 is an upper bound on the angle between $x-\bar{x}$ and the supporting hyperplane A, and is easily checked to satisfy $\theta \leq \alpha$. We thus have

$$\frac{d(\bar{x}, A)}{\|x - \bar{x}\|} \leq \sin\theta \leq \sin\alpha \leq \sin\left(2\sin^{-1}(\epsilon_1/2)\right).$$

So for a given $\epsilon > 0$, if ϵ_1 were chosen to be such that $\sin\left(2\sin^{-1}(\epsilon_1/2)\right) < \epsilon$, then we are done. □

Next, we recall Moreau's Theorem, and remark on how it will be used. For more on Moreau's Theorem, see for example [23, Theorem III.3.2.5] .

Theorem 2.2 [29] (Moreau's Decomposition Theorem) *Suppose $C \subset X$ is a closed convex cone in a Hilbert space X. Then for any $x \in X$, we can write $x = P_C(x) + P_{C^-}(x)$, and moreover, $\langle P_C(x), P_{C^-}(x)\rangle = 0$.*

The following result will be used in Theorems 3.5 and 5.12.

Proposition 2.3 (Projection onto cones) *Suppose $C \subset X$ is a closed convex cone in a Hilbert space X. Then the supporting hyperplane formed by projecting a point y onto C contains the origin.*

Proof By Moreau's Theorem, the projection $P_C(y)$ satisfies

$$y = P_C(y) + P_{C^-}(y) \quad \text{and} \quad \langle P_C(y), P_{C^-}(y)\rangle = 0.$$

The supporting halfspace produced by projecting y onto C would be

$$\{x \mid \langle x, y - P_C(y)\rangle \leq \langle P_C(y), y - P_C(y)\rangle\},$$

which equals $\{x \mid \langle x, P_{C^-}(y)\rangle \leq 0\}$. It is clear that the origin is in the supporting hyperplane. □

3 Convergence for the best approximation problem

In this section, we discuss algorithms for the Best Approximation Problem (1.2). We describe Algorithm 3.1, and show strong convergence to the closest point in the intersection of the closed convex sets (Theorem 3.3). Furthermore, in the finite dimensional case where the sets have a local conic structure (3.5), Algorithm 3.1 with (3.1b) converges in finitely many iterations (Theorem 3.5) under a normal condition (3.6). We give an example to show that the condition (3.6) cannot be dropped.

For each $n \in \mathbb{N}$, let $[n]$ denote "n mod r"; that is,

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

We present our algorithm for this section.

Algorithm 3.1 (Best approximation) *For a point x_0 and closed convex sets K_l, $l = 1, 2, \ldots, r$, of a Hilbert space X, find the closest point to x_0 in $K := \cap_{l=1}^r K_l$.*

Step 0: Let $i = 1$.
Step 1: Choose $J_i \subset \{1, \ldots, r\}$. Some examples of J_i are

$$J_i = \{[i]\}, \tag{3.1a}$$
$$\text{and } J_i = \{1, \ldots, r\}. \tag{3.1b}$$

For $j \in J_i$, define $x_i^{(j)} \in X$, $a_i^{(j)} \in X$ and $b_i^{(j)} \in \mathbb{R}$ by

$$x_i^{(j)} = P_{K_j}(x_{i-1}),$$
$$a_i^{(j)} = x_{i-1} - x_i^{(j)},$$
$$\text{and } b_i^{(j)} = \left\langle a_i^{(j)}, x_i^{(j)} \right\rangle.$$

Define the set $F_i \subset X$ by

$$F_i := \left\{ x \mid \left\langle a_l^{(j)}, x \right\rangle \leq b_l^{(j)} \text{ for all } l = 1, \ldots, i \text{ and } j \in J_l \right\}. \tag{3.2}$$

Let $x_i = P_{F_i}(x_0)$.
Step 2: Set $i \leftarrow i + 1$, and go back to step 1.

When J_i is chosen using (3.1a) and $x_i^{([i])} \in K_{[i]}$ so that $x_i^{([i])} = P_{K_{[i]}}(x_{i-1})$, then $a_i^{([i])} = 0$ and $b_i^{([i])} = 0$, and the algorithm stalls for one step. These values of $a_i^{([i])}$ and $b_i^{([i])}$ are still valid, though any implementation should treat this case separately. When the algorithm stalls for r iterations in a row, then we have found the closest point from x_0 to $\bigcap_{l=1}^r K_l$.

Remark 3.2 (Projecting to sets with greater second order behavior) In Step 1 of Algorithm 3.1, one needs to choose J_i. When the size of the quadratic programs are small and easy to solve, it would be ideal to choose J_i so that $|J_i| = 1$. The cyclic choice in (3.1a) is a natural choice. But as remarked in Fig. 1, one factor in our strategy is the second order behavior of the sets K_l. Another strategy is to record the distances in the most recent projections to the set K_l, and choose J_i to contain the index where the highest distance was recorded. In the case where one of the sets K_l is a subspace (and has fewer second order effects), the computations would be focused on the other sets. However, one may want to ensure that all sets are projected onto every once in a while so that Algorithm 3.1 is not fooled in regions where the boundary is locally but not globally affine. Possible strategies are:

$$\text{There exists } p \text{ such that for all } \bar{i},\ \bigcup_{i=\bar{i}}^{\bar{i}+p} J_i = \{1, \ldots, r\},$$

$$\text{or}\quad \text{For each } l = 1, \ldots, r, \text{ there are infinitely many } J_i \text{ containing } l.$$

The following theorem addresses the convergence of Algorithm 3.1. This theorem can be compared to the Boyle-Dykstra Theorem [7], which establishes the convergence of Dykstra's algorithm [15].

Theorem 3.3 (Strong convergence of Algorithm 3.1) *For any starting point x_0, the sequences $\{x_i\}$ produced by Algorithm 3.1 using (3.1a) or (3.1b) converge strongly to $P_K(x_0)$.*

Proof We shall only prove the result for the choice (3.1a), since the proof for (3.1b) is similar. By considering a translation if necessary, we can let x_0 be 0. We can also assume that $0 \notin K$. The iterates x_i satisfy $\|x_i\| \leq d(0, K)$, so $\{x_i\}$ has a weak cluster point z. Since x_i is the closest point from 0 to F_i, and

$$F_{i+1} \subset F_i \quad \text{for all } i, \tag{3.3}$$

we see that $\|x_i\|$ is an increasing sequence, so $M := \lim_{i\to\infty} \|x_i\|$ exists.

Step 1: z is actually a strong cluster point. It is clear that $\lim_{i\to\infty} \|x_i\| \geq \|z\|$. We only need to prove that

$$\|z\| = \lim_{i\to\infty} \|x_i\|, \tag{3.4}$$

since this condition together with the weak convergence of the subsequence of x_i implies the strong convergence to z. Suppose instead that $\lim_{i\to\infty} \|x_i\| > \|z\|$. Then there is some k such that $\|x_k\| > \|z\|$. By (3.3), we have, for all $i > k$,

$$\langle x_k, x_i\rangle \geq \langle x_k, x_k\rangle = \|x_k\|^2 > \|x_k\|\|z\| \geq \langle x_k, z\rangle,$$

contradicting z being a weak cluster point of $\{x_i\}$. Therefore z is a strong cluster point of $\{x_i\}$.

Step 2: Any z is in K. Suppose on the contrary that $z \notin K$. Then there is some l^* such that $z \notin K_{l^*}$, or $P_{K_{l^*}}(z) \neq z$. Algorithm 3.1 generates a hyperplane that separates z

from K_{l^*}. The halfspace $\{x \mid \langle a_z, x\rangle \le b_z\}$ separates z and K, where for $y \in X$, a_y and b_y are defined by

$$a_y = y - P_{K_{l^*}}(y), b_z = \left\langle y - P_{K_{l^*}}(y), P_{K_{l^*}}(y)\right\rangle.$$

The distance D from 0 to the intersection of halfspaces

$$\left\{x \mid \langle -z, x\rangle \le -\|z\|^2 \text{ and } \langle a_z, x\rangle \le b_z\right\}$$

would satisfy $D > \|z\|$.

Next, the variables a_y and b_y depend continuously on the parameter y, at $y = z$. This means that if x_i is sufficiently close to z and $[i] = l^*$, then the distance $d(0, x_{i+1})$ would be sufficiently close to D. This would mean that $\|x_i\| > \|z\| \ge \epsilon$, where $\epsilon > 0$ is some constant, for i large enough, which is a contradiction to (3.4). Thus $z \in K$ as needed.

Step 3 : $z = P_K(x_0)$. To see this, observe that $z \in K$ implies that $d(0, K) \le \|z\|$. The fact that $\|z\| = \lim_{i\to\infty} \|x_i\|$ from step 1 gives $d(0, K) = \|z\|$.

Thus we are done. □

Remark 3.4 (Reducing number of supporting hyperplanes in defining F_i) In the proof of Theorem 3.3, step 1 relies on the fact that $F_{i+1} \subset F_i$ for all i in the choice of F_i in (3.2). If $X = \mathbb{R}^n$, then step 1 of the proof would be unnecessary, but the sequence $\{\|x_i - x_0\|\}$ needs to be increasing in order for step 2 to work. This can be enforced by adding the halfspace with normal $(x_0 - x_{i-1})$ through x_{i-1} in constructing F_i. To ensure that each quadratic programming problem that needs to be solved is easy, the polyhedron F_i can be chosen such that the number of inequalities that define F_i is small. One can take only the active halfspace in solving the projection problem $x_i = P_{F_i}(x_0)$, or by aggregating some of the active halfspace to one active halfspace when building up the polyhedron F_i.

When the K_j satisfies a local conical property (3.5) at the limit $\bar{x}$ of Algorithm 3.1 in $\mathbb{R}^n$, the convergence is actually finite.

Theorem 3.5 (Finite convergence for conical problems in $\mathbb{R}^n$) *For Algorithm* 3.1 *with* (3.1b)*, suppose that* $X = \mathbb{R}^n$. *Convergence is guaranteed by Theorem* 3.3. *Suppose* $P_K(x_0) = \bar{x}$*, and* K_j *are such that*

$$\text{For some } \bar{\epsilon} > 0, [K_j - \bar{x}] \cap \bar{\epsilon}\mathbb{B} = T_{K_j}(\bar{x}) \cap \bar{\epsilon}\mathbb{B} \text{ for } j = 1, \dots, r, \tag{3.5}$$

(in other words, the sets K_j *are conical in a neighborhood of* $\bar{x}$*) and*

$$x_0 - \bar{x} \in \text{int}\big(N_K(\bar{x})\big). \tag{3.6}$$

Then Algorithm 3.1 *with* (3.1b) *converges to* $\bar{x}$ *in finitely many iterations.*

Proof We can assume $\bar{x} = 0$. Suppose on the contrary that the convergence to 0 requires infinitely many iterations. We seek a contradiction. Let $\{x_i\}$ be the sequence generated by Algorithm 3.1, and let $\{\tilde{x}_i\}$ be a subsequence such that $\|\tilde{x}_i - 0\| < \bar{\epsilon}$ for all i, and $\lim_{i\to\infty} \frac{\tilde{x}_i}{\|\tilde{x}_i\|}$ exists, say $\tilde{x}$.

Step 1: $\tilde{x}$ lies in $T_K(0)$. Suppose on the contrary that $\tilde{x} \notin T_K(0)$. Then $\tilde{x} \notin K_j$ for some j. Assume without loss of generality that $j = 1$. Let $P_{T_{K_1}(0)}(\tilde{x}) = z$, and $\tilde{x} - z \in N_{K_1}(0)$. Let $v_i = \frac{\tilde{x}_i}{\|\tilde{x}_i\|} - P_{T_{K_1}(0)}\left(\frac{\tilde{x}_i}{\|\tilde{x}_i\|}\right)$ and $v = \tilde{x} - z$. By the continuity of the projection, we must have $v_i \to v$. Since the hyperplane $\{x \mid \langle x, v\rangle = \langle z, v\rangle\}$ separates $\tilde{x}$ from K_1 and $\langle z, v\rangle = 0$ by Moreau's Theorem, we have $\langle \tilde{x}, v\rangle > 0$.

Let y be any point in $\bar{\epsilon}\mathbb{B}$, taking into account (3.5) and $\bar{x} = 0$. By Moreau's Theorem (See Proposition 2.3), the supporting hyperplane produced by projecting y onto K_1 contains 0 on its boundary. By the design of Algorithm 3.1, we must have $\langle \tilde{x}_{i+1}, v_i\rangle \leq 0$, which gives

$$\left\langle \frac{\tilde{x}_{i+1}}{\|\tilde{x}_{i+1}\|}, v_i \right\rangle \leq 0.$$

Taking limits, we get $\langle \tilde{x}, v\rangle \leq 0$, which contradicts $\langle \tilde{x}, v\rangle > 0$ earlier.

Step 2: $\tilde{x}$ cannot lie in $T_K(0)$. Suppose otherwise. Then the condition (3.6) implies that if $\tilde{x} \in K$ and i is large enough, then $d(x_0, 0) < d(x_0, \tilde{x}_i)$, contradicting that $d(x_0, 0) > d(x_0, \tilde{x}_i)$ in the choice of $\tilde{x}_i$.

The statements proved in Steps 1 and 2 are clearly contradictory, which ends our proof. □

In view of the above result, we would expect Algorithm 3.1 (especially with (3.1b)) to converge quickly to the closest point under condition (3.6).

The number of iterations needed before convergence depends on, among other things, the $\bar{\epsilon}$. In the case where K_j are cones and (3.6) does not hold, step 2 in the proof of Theorem 3.5 may fail, and there may be no finite convergence. We give an example.

Example 3.6 (No finite convergence) Consider $X = \mathbb{R}^3$. Consider the rays

$$r_1 = \mathbb{R}_+(1, -1, -1)^T \quad \text{and} \quad r_2 = \mathbb{R}_+(-1, -1, -1)^T.$$

For a vector $v \in \mathbb{R}^3$, let θ_1 be the angle r_1 makes with v, and let θ_2 be similarly defined. Let K_1 and K_2 be the ice cream cones defined by

$$K_i = \{v \mid \cos(\theta_i) \leq 1/\sqrt{3}\} \quad \text{for } i = 1, 2.$$

Let $x_0 = (0, 0, 1)^T$. A few consequences are immediate.

(1) The ray $\mathbb{R}_+(0, -1, 0)^T$ is on the boundaries of K_1, K_2 and $K := K_1 \cap K_2$.
(2) $P_K(x_0) = 0$.

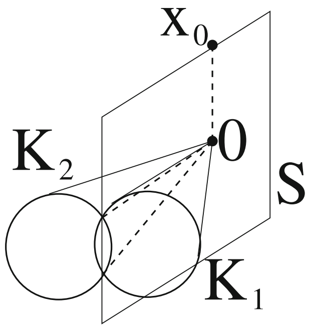

Fig. 3 This figure illustrates Example 3.6, where K_1 and K_2 are *two cones* in $\mathbb{R}^3$ that are mirror reflections of each other about the subspace S. Whenever a point in S outside $K_1 \cap K_2$ is projected onto K_1 and K_2, the supporting hyperplanes intersect in a *line* in S containing the origin 0. The projection of x_0 onto this *line* will never be the origin, so Algorithm 3.1 cannot converge in finitely many iterations

(3) There is only one unit vector in $N_{K_1}\big((0,-1,0)^T\big)$, say u. Let the subspace S be $\{x \mid (1,0,0)x = 0\}$. Then

$$[\mathbb{R}_+\{u\} + \{(0,-1,0)^T\}] \cap S = (0,-1,0)^T.$$

A similar statement holds when K_1 is replaced by K_2.

We now show that the convergence of Algorithm 3.1 with (3.1b) is infinite. One might find Fig. 3 helpful. By symmetry, the iterates x_i lie in S. If Algorithm 3.1 with (3.1b) converges in finitely many iterations, then property (3) would imply that the next to last iterate is of the form $(0,-\alpha,0)^T$, where $\alpha > 0$, and that cannot happen. In the case where $x_0 = (0,\epsilon,1)^T$, where $\epsilon > 0$ is arbitrarily small, we will still get finite convergence to 0, but the number of iterations needed will be arbitrarily large as $\epsilon \searrow 0$.

4 Convergence for the set intersection problem

In this section, we analyze a modified alternating projection algorithm (Algorithm 4.1). The global convergence of this algorithm is proved in Theorem 4.5. The insight on using supporting hyperplanes and quadratic programming to accelerate convergence allows us to obtain local superlinear convergence in $\mathbb{R}^2$ (Theorem 4.6), although Algorithm 4.1 in its current form does not converge superlinearly in $\mathbb{R}^3$ (Example 4.7). A locally superlinearly convergent algorithm will be presented and analyzed in Sect. 5 using very different methods.

We shall analyze the following algorithm.

Algorithm 4.1 (Modified MAP) *For a point x_0 and closed convex sets K_1 and K_2 of a Hilbert space X, find a point in $K := K_1 \cap K_2$.*

Step 0: Set $i = 1$.
Step 1: Choose $J_i \subset \{1,2\}$. Some examples are

$$J_i = \{[i]\}, \tag{4.1a}$$

$$\text{and } J_i = \{1,2\}. \tag{4.1b}$$

Step 2: For $j \in J_i$, define $x_i^{(j)} \in X$, $a_i^{(j)} \in X$ and $b_i^{(j)} \in \mathbb{R}$ by

$$\begin{aligned} x_i^{(j)} &= P_{K_j}(x_{i-1}), \\ a_i^{(j)} &= x_{i-1} - x_i^{(j)}, \\ \text{and } b_i^{(j)} &= \left\langle a_i^{(j)}, x_i^{(j)} \right\rangle. \end{aligned}$$

Define the set $F_i \subset X$ by

$$F_i := \begin{cases} \left\{ x \mid \left\langle a_l^{([l])}, x \right\rangle \leq b_l^{([l])} \quad \text{for } l = i-1, i \right\} & \text{if } J_i = \{[i]\} \quad \text{and} \quad i > 1, \\ \left\{ x \mid \left\langle a_1^{(1)}, x \right\rangle \leq b_1^{(1)} \right\} & \text{if } J_i = \{[i]\} \quad \text{and} \quad i = 1, \\ \left\{ x \mid \left\langle a_i^{(j)}, x \right\rangle \leq b_i^{(j)} \quad \text{for } j = 1, 2 \right\} & \text{if} \quad J_i = \{1, 2\}. \end{cases}$$

Let $x_i = P_{F_i}(x_i^{([i])})$.
Step 3: Set $i \leftarrow i + 1$, and go back to step 1.

As mentioned in Remark 3.2, there are good reasons for choosing J_i to be such that $|J_i| = 1$ but not cyclic, but the construction of F_i has to be amended accordingly. It may turn out that x_i could be in K_1 already, so $P_{K_1}(x_i)$ will not give a new supporting hyperplane. In this case, we can just use the supporting hyperplane obtained from previous iterations. When $J_i = \{1, 2\}$, we can check that x_i lies in the plane containing x_{i-1}, $x_i^{(1)}$ and $x_i^{(2)}$, and that $x_i = P_{F_i}(x_i^{([i])}) = P_{F_i}(x_{i-1})$. We shall prove the superlinear convergence of this case in $\mathbb{R}^2$ in Theorem 4.6.

We now recall some results on Fejér monotonicity to prove convergence of Algorithm 4.1. We take our results from [10, Theorem 4.5.10 and Lemma 4.5.8].

Definition 4.2 (*Fejér monotone sequence*) Let X be a Hilbert space, let $C \subset X$ be a closed convex set and let $\{x_i\}$ be a sequence in X. We say that $\{x_i\}$ is *Fejér monotone with respect to C* if

$$\|x_{i+1} - c\| \leq \|x_i - c\| \quad \text{for all } c \in C \quad \text{and} \quad i = 1, 2, \ldots$$

Theorem 4.3 (Properties of Fejér monotonicity) *Let X be a Hilbert space, let $C \subset X$ be a closed convex set and let $\{x_i\}$ be a Fejér monotone sequence with respect to C. Then*

(1) $\{x_i\}$ is bounded and $d(C, x_{i+1}) \leq d(C, x_i)$.
(2) $\{x_i\}$ has at most one weak cluster point in C.
(3) If $\operatorname{int}(C) \neq \emptyset$, then $\{x_i\}$ converges in norm.

Lemma 4.4 (Attractive property of projection) *Let X be a Hilbert space and let $C \subset X$ be a closed convex set. Then $P_C : X \to X$ is* 1-attracting with respect to C*: For every $x \notin C$ and $y \in C$, we have*

$$\|P_C(x) - x\|^2 \leq \|x - y\|^2 - \|P_C(x) - y\|^2.$$

We now prove the convergence of Algorithm 4.1. We note that Algorithm 4.1 can be easily extended to the case of $r > 2$ closed convex sets, and the corresponding extension of Theorem 4.5 will still be true.

Theorem 4.5 (Convergence of Algorithm 4.1) *Suppose K_1 and K_2 are closed convex sets in a Hilbert space X such that $K := K_1 \cap K_2 \neq \emptyset$. Then the iterates in Algorithm 4.1 with either* (4.1a) *or* (4.1b) *are such that x_i converges weakly to some $z \in K$. The convergence is strong if either* $\mathrm{int}(K) \neq \emptyset$ *or* $X = \mathbb{R}^n$.

Proof We shall first prove convergence when J_i is chosen by (4.1a).

The sequences $\{x^{(1)}_{2i+1}\}_i$ and $\{x^{(2)}_{2i}\}_i$ lie in K_1 and K_2 respectively. Construct the sequence $\{\tilde{x}_i\}$ such that

$$\tilde{x}_i = \begin{cases} x_j & \text{if} \quad i = 2j \\ x^{(1)}_{2j+1} & \text{if} \quad i = 4j+1 \\ x^{(2)}_{2j+2} & \text{if} \quad i = 4j+3. \end{cases}$$

Note that $\{\tilde{x}_i\}$ lines up the points in $\{x_i\}$ and $\{x_i^{([i])}\}$ in the order in which they were produced in Algorithm 4.1.

Step 1: $\{\tilde{x}_i\}$ is Fejér monotone with respect to K. Since $K \subset K_1$, $K \subset K_2$ and $K \subset F_i$ for all i, the projections P_{K_1}, P_{K_2} and P_{F_i} are nonexpansive. So

$$\begin{aligned} &\|x_i^{([i])} - y\| = \|P_{K_{[i]}}(x_{i-1}) - y\| \leq \|x_{i-1} - y\| \\ \text{and} \quad &\|x_i - y\| = \|P_{F_i}(x_i^{(i)}) - y\| \leq \|x_i^{([i])} - y\| \text{ for all } y \in K \text{ and } i \geq 1. \end{aligned}$$

This means that $\{\tilde{x}_i\}$ is a Fejér monotone sequence with respect to K.

Step 2: $\{\tilde{x}_i\}$ is asymptotically regular, i.e.,

$$\lim_{i\to\infty} \|\tilde{x}_i - \tilde{x}_{i+1}\| = 0.$$

Fix any $\bar{y} \in K$. Applying Lemma 4.4, we get

$$\begin{aligned} &\|x_i^{([i])} - x_{i-1}\|^2 = \|P_{K_{[i]}}(x_{i-1}) - x_{i-1}\|^2 \leq \|x_{i-1} - \bar{y}\|^2 - \|x_i^{([i])} - \bar{y}\|^2 \\ \text{and} \quad &\|x_i - x_i^{([i])}\|^2 \leq \|x_i^{([i])} - \bar{y}\|^2 - \|x_i - \bar{y}\|^2 \quad \text{for all } i \geq 1. \end{aligned}$$

This tells us that $\|\tilde{x}_i - \tilde{x}_{i-1}\|^2 \leq \|\tilde{x}_{i-1} - \bar{y}\|^2 - \|\tilde{x}_i - \bar{y}\|^2$ for all $i \geq 1$. Since $\{\tilde{x}_i\}$ is Fejér monotone with respect to K, $\|\tilde{x}_i - \bar{y}\|^2$ is a decreasing sequence. We thus have the asymptotic regularity of $\{\tilde{x}_i\}$.

Step 3: Wrapping up. By Theorem 4.3(1), the sequence $\{\tilde{x}_i\}$ is bounded. So $\{\tilde{x}_i\}$ has a weakly convergent subsequence, say $\{\tilde{x}_{i_k}\}_k$. By the asymptotic regularity of $\{\tilde{x}_i\}$, the sequence $\{\tilde{x}_{i_k+1}\}_k$ has the same limit as $\{\tilde{x}_{i_k}\}_k$, so we can take a different subsequence if necessary and assume that infinitely many of the i_k are odd. We can choose yet another subsequence of $\{\tilde{x}_{i_k}\}$ if necessary so that all terms are in either K_1, or all terms

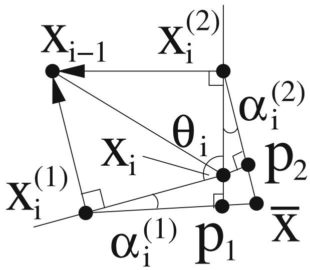

Fig. 4 Diagram for the proof of Theorem 4.6

are in K_2. For the sake of argument, assume that all terms lie in K_1. So the weak limit of $\{\tilde{x}_{i_k}\}_k$, say x, lies in K_1. By the asymptotic regularity of $\{\tilde{x}_{i_k}\}_k$ and considering $\{\tilde{x}_{i_k+2}\}_k$, we see that $x \in K_2$. So the weak cluster point must lie in K. By Theorem 4.3(2), we conclude that $\{\tilde{x}_i\}$ converges to a point in K. The last sentence of the result follows from Theorem 4.3(3).

For the case of using (4.1b), the steps are very similar, so we only give an outline: One proves that the sequences $\{x_i\}$ and $\{x_i^{(j)}\}$ are Fejér monotone with respect to K for $j = 1, 2$. Next, the sequence $x_0, x_1^{(j)}, x_1, x_1^{(j)}, x_2, \ldots$ is asymptotically regular, which implies that the sequences $\{x_i\}$ and $\{x_i^{(j)}\}$ have the same weak cluster points. Since j is arbitrary, the weak cluster points must lie in K, and by Theorem 4.3(2), such a weak cluster point is unique. □

The problem of whether the MAP can converge strongly in a Hilbert space has only been recently resolved to be negative in [24], so it remains to be seen how Theorem 4.5 can be strengthened.

We now move on to the fast local convergence of Algorithm 4.1. Even though the result below is only valid for $\mathbb{R}^2$ and a result establishing superlinear convergence for $\mathbb{R}^n$ is presented in Sect. 5, Theorem 4.6 has value because the proof is simpler than and very different from the proof in Sect. 5, and the assumptions needed are quite different.

Theorem 4.6 (Superlinear convergence in $\mathbb{R}^2$) *Suppose K_1 and K_2 are closed convex sets in $\mathbb{R}^2$ such that*

(1) Algorithm 4.1 with (4.1b) converges to a point $\bar{x}$ such that $\partial N_{K_1}(\bar{x}) \cap \partial[-N_{K_2}(\bar{x})] = \{0\}$, and
(2) There is some iterate x_i such that $x_i \notin \operatorname{int}(K_j)$ for $j = 1, 2$.

Then the sequence $\{x_i\}$ thus produced converges locally superlinearly to $\bar{x}$.

Proof We refer to Fig. 4. Let $\alpha_i^{(1)}$ be the angle between $x_i - x_i^{(1)}$ and $\bar{x} - x_i^{(1)}$, and let $\alpha_i^{(2)}$ be similarly defined. As $i \to \infty$, the points $x_i^{(1)}$ and $x_i^{(2)}$ converge to $\bar{x}$, so Theorem 2.1 says that the angles $\alpha_i^{(1)}$ and $\alpha_i^{(2)}$ converge to zero.

Let θ_i be the angle between $x_i^{(1)} - x_i$ and $x_i^{(2)} - x_i$ as marked. Since $\partial N_{K_1}(\bar{x}) \cap \partial[-N_{K_2}(\bar{x})] = \{0\}$, the angle θ_i is bounded from below by $\bar{\theta} > 0$. It is also easy to check that if $x_i \notin \operatorname{int}(K_j)$ for $j = 1, 2$, then the same property holds for all i afterward.

The points p_1 and p_2 are obtained by projecting x_i onto the line segments $[x_i^{(1)}, \bar{x}]$ and $[x_i^{(2)}, \bar{x}]$. To show that $\{x_i\}$ converges superlinearly to $\bar{x}$, it suffices to show that

$$\lim_{i\to\infty} \frac{\|x_i - \bar{x}\|}{\|x_{i-1} - x_i\|} = 0, \tag{4.2}$$

since $\|x_{i-1} - \bar{x}\| \geq \|x_{i-1} - x_i\|$. Let $L_i = \|x_{i-1} - x_i\|$. By the sine rule, the distance $\|x_i^{(1)} - x_i\|$ equals $L_i \sin \gamma_i^{(1)}$, where $\gamma_i^{(1)}$ is some angle in the interval $[0, \pi - \theta_i]$. The distance $\|p_1 - x_i\|$ can be calculated to be bounded above by $L_i \sin \alpha_i^{(1)} \sin \gamma_i^{(1)}$, while the distance $\|p_2 - x_i\|$ is easily computed to be bounded from above by $L_i \sin \alpha_i^{(2)} \sin \gamma_i^{(2)}$, where $\gamma_i^{(2)}$ is similarly defined. The distance $\|x_i - \bar{x}\|$ is easily seen to be the diameter of the circumcircle of the cyclic quadrilateral with vertices x_i, $\bar{x}$, p_1 and p_2. The angle between $p_1 - x_i$ and $p_2 - x_i$ is easily calculated to be $\pi - \theta_i + \alpha_i^{(1)} + \alpha_i^{(2)}$. (Note that $x_i^{(2)}$, x_i and p_1 need not be collinear.) The distance of $\|p_1 - p_2\|$ can be estimated by

$$\begin{aligned} \|p_1 - p_2\| &\leq \|p_1 - x_i\| + \|p_2 - x_i\| \\ &\leq \sin(\min\{\pi/2, \pi - \theta_i\})[\sin \alpha_i^{(1)} + \sin \alpha_i^{(2)}]L_i. \end{aligned}$$

The value $\|x_i - \bar{x}\|$ can be obtained by the sine rule to be

$$\frac{\|p_1 - p_2\|}{\sin(\pi - \theta_i + \alpha_i^{(1)} + \alpha_i^{(2)})},$$

so we have

$$\|x_i - \bar{x}\| \leq \frac{\sin(\min\{\pi/2, \pi - \theta_i\})}{\sin(\pi - \theta_i + \alpha_i^{(1)} + \alpha_i^{(2)})} L_i \left[\sin \alpha_i^{(1)} + \sin \alpha_i^{(2)}\right].$$

Thus to prove that (4.2), it suffices to prove that

$$\lim_{i\to\infty} \frac{\sin(\min\{\pi/2, \pi - \theta_i\})}{\sin([\pi - \theta_i] + \alpha_i^{(1)} + \alpha_i^{(2)})} \left[\sin \alpha_i^{(1)} + \sin \alpha_i^{(2)}\right] = 0. \tag{4.3}$$

We have shown that $\liminf_{i\to\infty} \theta_i \geq \bar{\theta} > 0$. The limit (4.3) holds because the limits of $\alpha_i^{(1)}$ and $\alpha_i^{(2)}$ are zero and $\theta_i \in [\bar{\theta}, \pi]$ for all i. Hence we are done. □

The superlinear convergence in Theorem 4.6 does not extend to $\mathbb{R}^3$ however, even when K_1 and K_2 are linear subspaces.

Example 4.7 (No superlinear convergence in $\mathbb{R}^3$ for Algorithm 4.1) We give an example of subspaces K_1 and K_2 in $\mathbb{R}^3$ such that $\partial N_{K_1}(\bar{x}) \cap \partial[-N_{K_2}(\bar{x})] = \{0\}$ but there is no superlinear convergence to $\bar{x}$ in Algorithm 4.1 using (4.1b) for some starting point. Consider K_1 and K_2 defined by

$$\begin{aligned} K_1 &= \mathbb{R}(1,0,1)^T, \\ \text{and } K_2 &= \{(x,y,0)^T \mid x,y \in \mathbb{R}\}. \end{aligned}$$

For the starting point $x_0 = (4,-1,0)^T$, we compute the iterates of Algorithm 4.1. We calculate

$$\begin{aligned} x_1^{(1)} &= P_{K_1}(x_0) = (2,0,2)^T, \\ x_1^{(2)} &= x_0, \\ x_1 &= \left(\frac{2}{5},\frac{4}{5},0\right)^T, \\ x_2^{(1)} &= P_{K_1}(x_1) = \left(\frac{1}{5},0,\frac{1}{5}\right)^T, \\ x_2^{(2)} &= x_1, \\ \text{and } x_2 &= \left(\frac{16}{85},\frac{-4}{85},0\right)^T = \frac{4}{85}x_0. \end{aligned} \tag{4.4}$$

To verify that x_1 and x_2 are the correct iterates, we can check that x_0, $x_1^{(1)}$, $x_1^{(2)}$ and x_1 lie in the plane $\{x \mid (1,2,0)x = 2\}$, and that x_1, $x_2^{(1)}$, $x_2^{(2)}$ and x_2 lie in the plane $\{x \mid (4,-1,0)x = 4/5\}$. Another condition helpful for the verification is that

$$\left\langle x_i - x_i^{(1)}, x_{i-1} - x_i^{(1)} \right\rangle = 0 \quad \text{for } i = 1,2.$$

From (4.4), we see that the convergence to zero of Algorithm 4.1 using (4.1b) is linear and not superlinear. But the rate of convergence for our choice of starting iterate is $\frac{4}{85}$ for every four projections, which is more than twice as fast of the rate of $\frac{1}{4}$ for every four projections for the usual MAP.

We show that if there were more supporting hyperplanes used in approximating K, then we get finite convergence to zero for this example. The projection of x_i onto $x_{i+1}^{(1)}$ generates the supporting hyperplanes

$$\begin{aligned} &\{x \mid (2,-1,2)x = 0\} \quad \text{if } i \text{ is even,} \quad \text{and} \\ &\{x \mid (1,4,-1)x = 0\} \quad \text{if } i \text{ is odd.} \end{aligned}$$

The projection of any point of the form $(t,0,t)^T$, where $t > 0$, onto the set

$$\left\{ x \mid \begin{pmatrix} 2 & -1 & 2 \\ 1 & 4 & -1 \\ 0 & 0 & 1 \end{pmatrix} x \leq 0 \right\}$$

is equal to the zero vector, which is the only point in K.

5 Superlinear convergence for the set intersection problem

Our main result in this section is Theorem 5.12, where we prove the superlinear convergence of an algorithm for the set intersection problem (1.1) when the normal cones at the point of intersection are pointed cones satisfying appropriate alignment conditions.

We first describe our algorithm for this section.

Algorithm 5.1 (Mass projection algorithm) *For a starting iterate x_0 and closed convex sets $K_l \subset \mathbb{R}^n$, where $1 \leq l \leq r$, find a point in $K := \cap_{l=1}^r K_l$.*

Step 0: Set $i = 1$, and let $\bar{p}$ be some positive integer.
Step 1: Choose $J_i = \{1, \ldots, r\}$.
Step 2: For $j \in J_i$, define $x_i^{(j)} \in \mathbb{R}^n$, $a_i^{(j)} \in \mathbb{R}^n$ and $b_i^{(j)} \in \mathbb{R}$ by

$$\begin{aligned} x_i^{(j)} &= P_{K_j}(x_{i-1}), \\ a_i^{(j)} &= x_{i-1} - x_i^{(j)}, \\ \text{and } b_i^{(j)} &= \left\langle a_i^{(j)}, x_i^{(j)} \right\rangle. \end{aligned}$$

Define the set $\tilde{F}_i \subset \mathbb{R}^n$ by

$$\tilde{F}_i := \left\{ x \mid \left\langle a_l^{(j)}, x \right\rangle \leq b_l^{(j)} \text{ for } 1 \leq j \leq r,\ \max(1, i - \bar{p}) \leq l \leq i \right\}.$$

Let $x_i = P_{\tilde{F}_i}(x_{i-1})$.
Step 3: Set $i \leftarrow i + 1$, and go back to step 1.

The modifications in Algorithm 5.1 from Algorithm 4.1 are that we set $X = \mathbb{R}^n$, the number of sets r is arbitrary, and the set $\tilde{F}_i$ approximating K is created using more of the previous separating halfspaces produced earlier.

Algorithm 5.1 produces a sequence $\{x_i\}$ Fejér monotone with respect to K and converging to a point $\bar{x} \in K$. The proof is an easy adaptation of that of Theorem 4.5.

We need to recall a few well known properties of convex sets in $\mathbb{R}^n$ to proceed. These properties may be found in standard convex analysis texts, for example [23,34].

Definition 5.2 (*Properties of convex sets*) The *linearity space* $\text{lin}(C)$ of a convex set $C \subset \mathbb{R}^n$ is the set of all directions y such that for every $x \in C$, the line $\{x\} + \mathbb{R}\{y\}$ is contained in C. In the case where C is a convex cone, $\text{lin}(C) = C \cap (-C)$.

A convex cone is said to be *pointed* if $\text{lin}(C) = \{0\}$, or in other words, C does not contain a line.

The affine hull $\text{aff}(C)$ of C is the smallest affine space containing C. We shall write $\text{lin}(\text{aff}(\cdot))$ as $\text{lin} \circ \text{aff}(\cdot)$.

We recall a well known fact about convex cones. See for example [34, Page 65].

Proposition 5.3 (Convex cone decomposition) *A closed convex cone $C \subset \mathbb{R}^n$ can be written as the direct sum $C = L \oplus [L^\perp \cap C]$, where $L = \text{lin}(C)$, and $L^\perp \cap C$ is a pointed convex cone.*

As a consequence of Proposition 5.3, we have the following result on the normal cones of convex sets.

Proposition 5.4 (Linearity spaces of normals of convex sets) *Suppose* $C \subset \mathbb{R}^n$ *is a convex set. Then for any* $x \in C$, $[\mathrm{lin} \circ \mathrm{aff}(C)]^\perp = \mathrm{lin}(N_C(x))$. *In particular,* $\mathrm{lin} \circ \mathrm{aff}(C) \cap N_C(x)$ *is a pointed convex cone.*

Proof $v \in \mathrm{lin}(N_C(x)) \iff \pm v \in N_C(x) \iff \langle v, x - c\rangle = 0$ for all $c \in C \iff v \in [\mathrm{lin} \circ \mathrm{aff}(C)]^\perp$.

To see that $\mathrm{lin}\circ\mathrm{aff}(C)\cap N_C(x)$ is a pointed cone, apply $[\mathrm{lin}\circ\mathrm{aff}(C)]^\perp = \mathrm{lin}(N_C(x))$ to Proposition 5.3. □

In our quest to prove Theorem 5.12, we need the next three lemmas to prove the intermediate result Proposition 5.8.

Lemma 5.5 (Intermediate estimate) *Suppose* v_1 *and* v_2 *are nonzero vectors in* $\mathbb{R}^n$ *such that* $\frac{\|v_1 - v_2\|}{\|v_2\|} \leq \beta$. *Then* $\left\| \frac{v_1}{\|v_1\|} - \frac{v_2}{\|v_2\|} \right\| \leq 2\beta$.

Proof We have

$$\begin{aligned} \left\| \frac{v_1}{\|v_1\|} - \frac{v_2}{\|v_2\|} \right\| &\leq \left\| \frac{v_1}{\|v_2\|} - \frac{v_2}{\|v_2\|} \right\| + \left\| \frac{v_1}{\|v_1\|} - \frac{v_1}{\|v_2\|} \right\| \\ &\leq \beta + \frac{\|v_1\| \big| \|v_1\| - \|v_2\| \big|}{\|v_1\|\|v_2\|} \\ &\leq \beta + \frac{\|v_1 - v_2\|}{\|v_2\|} \leq 2\beta. \end{aligned}$$

□

For the following result, recall that the cone $K^+ := -K^-$ is the *positive polar cone* (or more commonly, the *dual cone*) of K.

Lemma 5.6 (Pointed cone) *For a closed pointed convex cone* $K \subset \mathbb{R}^n$*, there is a unit vector* d *in* K^+*, the positive polar cone of* K*, and some* $c > 0$ *such that* $\mathbb{B}_c(d) \subset K^+$. *Next, for any unit vector* $v \in K$*, we have* $d^T v \geq c$.

Moreover, suppose $\lambda_i \geq 0$*, and* v_i *are unit vectors in* K *for all* i*, and* $\sum_{i=1}^\infty \lambda_i v_i$ *converges to* $\bar{v}$. *Clearly,* $\bar{v} \in K$. *Then* $\|\sum_{i=1}^\infty \lambda_i v_i\| \geq c \sum_{i=1}^\infty \lambda_i$*, which also implies that* $\sum_{i=1}^\infty \lambda_i$ *is finite.*

Proof It is well known that the positive polar cone (or dual) of a closed pointed convex cone has an interior (see for example [34, Corollary 14.6.1]), so the first sentence follows. For the unit vector $v \in K$, we have $(d - cv) \in K^+$, which gives $\langle d - cv, v\rangle \geq 0$, from which the first part follows.

Next,

$$\left\| \sum_{i=1}^\infty \lambda_i v_i \right\| \geq \left\langle d, \sum_{i=1}^\infty \lambda_i v_i \right\rangle \geq c \sum_{i=1}^\infty \lambda_i,$$

and the second part follows. □

Lemma 5.7 (Limit estimates involving pointed cones) *Suppose $\{v_i\}_i$ are unit vectors in $\mathbb{R}^n$ and $\{\lambda_i\}_i$ is a sequence of nonnegative numbers such that the cluster points of $\{v_i\}_i$ belong to a closed pointed convex cone $K \subset \mathbb{R}^n$. Then*

(1) If $\sum_{i=1}^{\infty} \lambda_i = \infty$, then cluster points of $\left\{ \frac{\sum_{i=1}^{j} \lambda_i v_i}{\| \sum_{i=1}^{j} \lambda_i v_i \|} \right\}_{j=1}^{\infty}$ belong to K.

(2) Take $c > 0$ to be the constant in Proposition 5.6. If $\sum_{i=1}^{\infty} \lambda_i v_i$ is convergent and there are unit vectors $\tilde{v}_i \in K$ such that $\|v_i - \tilde{v}_i\| \leq \epsilon$, then

$$\left\| \frac{\sum_{i=1}^{\infty} \lambda_i v_i}{\left\| \sum_{i=1}^{\infty} \lambda_i v_i \right\|} - \frac{\sum_{i=1}^{\infty} \lambda_i \tilde{v}_i}{\left\| \sum_{i=1}^{\infty} \lambda_i \tilde{v}_i \right\|} \right\| \leq \frac{2}{c}\epsilon. \tag{5.1}$$

Proof **Statement (1):** Since the cluster points of $\{v_i\}$ belong to K, for any $\epsilon > 0$, we can find I_ϵ such that for all $i \geq I_\epsilon$, there exists $\tilde{v}_i \in K$ such that $\|v_i - \tilde{v}_i\| < \epsilon$. Then

$$\left\| \sum_{i=1}^{j} \lambda_i v_i - \sum_{i=1}^{j} \lambda_i \tilde{v}_i \right\| = \left\| \sum_{i=1}^{j} \lambda_i (v_i - \tilde{v}_i) \right\| \leq \sum_{i=I_\epsilon}^{j} \lambda_i \epsilon + \sum_{i=1}^{I_\epsilon - 1} 2\lambda_i.$$

Next, Proposition 5.6 implies that $\left\| \sum_{i=1}^{j} \lambda_i \tilde{v}_i \right\| \geq c \sum_{i=1}^{j} \lambda_i$. So

$$\frac{\left\| \sum_{i=1}^{j} \lambda_i v_i - \sum_{i=1}^{j} \lambda_i \tilde{v}_i \right\|}{\left\| \sum_{i=1}^{j} \lambda_i \tilde{v}_i \right\|} \leq \frac{\sum_{i=I_\epsilon}^{j} \lambda_i \epsilon + \sum_{i=1}^{I_\epsilon - 1} 2\lambda_i}{c \sum_{i=1}^{j} \lambda_i}.$$

Proposition 5.5 gives

$$\left\| \frac{\sum_{i=1}^{j} \lambda_i v_i}{\left\| \sum_{i=1}^{j} \lambda_i v_i \right\|} - \frac{\sum_{i=1}^{j} \lambda_i \tilde{v}_i}{\left\| \sum_{i=1}^{j} \lambda_i \tilde{v}_i \right\|} \right\| \leq 2 \frac{\sum_{i=I_\epsilon}^{j} \lambda_i \epsilon + \sum_{i=1}^{I_\epsilon - 1} 2\lambda_i}{c \sum_{i=1}^{j} \lambda_i}.$$

The RHS of the above can be made arbitrarily small since ϵ can be made arbitrarily small and j can be made arbitrarily big. The term $\frac{\sum_{i=1}^{j} \lambda_i \tilde{v}_i}{\left\| \sum_{i=1}^{j} \lambda_i \tilde{v}_i \right\|}$ belongs to K, so Statement (1) holds.

Statement (2): First, since $\sum_{i=1}^{\infty} \lambda_i \tilde{v}_i$ is convergent, Proposition 5.6 implies

$$\left\| \sum_{i=1}^{\infty} \lambda_i \tilde{v}_i \right\| \geq c \sum_{i=1}^{\infty} \lambda_i,$$

which also implies that $\sum_{i=1}^{\infty} \lambda_i$ is finite, and

$$\left\| \sum_{i=1}^{\infty} \lambda_i v_i - \sum_{i=1}^{\infty} \lambda_i \tilde{v}_i \right\| = \left\| \sum_{i=1}^{\infty} \lambda_i (v_i - \tilde{v}_i) \right\| \leq \sum_{i=1}^{\infty} \lambda_i \epsilon.$$

Then

$$\frac{\left\|\sum_{i=1}^{\infty}\lambda_i v_i - \sum_{i=1}^{\infty}\lambda_i \tilde{v}_i\right\|}{\left\|\sum_{i=1}^{\infty}\lambda_i \tilde{v}_i\right\|} \leq \frac{\epsilon}{c}.$$

By Proposition 5.5, we get the conclusion (5.1) as needed. □

The following result shows that under certain conditions, the directions from which the iterates converge to the limit must lie inside the normal cone of K at the limit.

Proposition 5.8 (Approach of iterates to $\bar{x}$) *Consider the problem of finding a point $x \in K$, where $K = \cap_{l=1}^{r} K_l$ and $K_l \subset \mathbb{R}^n$ are closed convex sets. Suppose Algorithm 5.1 produces a sequence $\{x_i\}$ that converges to a point $\bar{x} \in K$ and is Fejér monotone with respect to K. Assume that:*

(1) If $\sum_{l=1}^{r} v_l = 0$ for some $v_l \in N_{K_l}(\bar{x})$, then $v_l = 0$ for all $l = 1, \ldots, r$.

Then provided none of the x_i equals $\bar{x}$, we have

$$\lim_{i\to\infty} \frac{\|P_{N_K(\bar{x})}(x_i - \bar{x})\|}{\|x_i - \bar{x}\|} = 1. \tag{5.2}$$

Proof By the way Algorithm 5.1 is designed, the KKT conditions for the problem of projecting x_{i-1} onto the polyhedron to obtain x_i give

$$x_i = x_{i-1} - \sum_{l=1}^{r} \sum_{k=\max(1,i-\bar{p})}^{i} [\lambda_l^{(i,k)} v_l^k + w_l^{(i,k)}],$$

where $\lambda_l^{(i,k)} v_l^k + w_l^{(i,k)}$ is a multiple of the vector $a_k^{(l)} = x_{k-1} - P_{K_l}(x_{k-1})$, $w_l^{(i,k)} \in [\text{lin} \circ \text{aff}(K_l)]^\perp$, v_l^k is a unit vector in $\text{lin} \circ \text{aff}(K_l) \cap N_{K_l}(P_{K_l}(x_{k-1}))$, and $\lambda_l^{(i,k)} \geq 0$. (The relationship $[\text{lin} \circ \text{aff}(K_l)]^\perp = \text{lin}(N_{K_l}(P_{K_l}(x_{k-1})))$ follows from Proposition 5.4.) For $j > i$, we can write $x_{i-1} - x_j$ as

$$\begin{aligned} x_{i-1} - x_j &= \sum_{s=i}^{j}\sum_{l=1}^{r} \sum_{k=\max(1,s-\bar{p})}^{s} [\lambda_l^{(s,k)} v_l^k + w_l^{(s,k)}] \\ &= \sum_{l=1}^{r}\left[\left[\sum_{s=i}^{j} \sum_{k=\max(1,s-\bar{p})}^{s} \lambda_l^{(s,k)} v_l^k\right] + \tilde{w}_l^{(i-1,j)}\right], \end{aligned}$$

where $\tilde{w}_l^{(i-1,j)} \in [\text{lin} \circ \text{aff}(K_l)]^\perp$. Let $\tilde{v}_l^{(i,j)} \in \mathbb{R}^n$ be the vector

$$\tilde{v}_l^{(i,j)} := \frac{\sum_{s=i}^{j} \sum_{k=\max(1,s-\bar{p})}^{s} \lambda_l^{(s,k)} v_l^k}{\left\|\sum_{s=i}^{j} \sum_{k=\max(1,s-\bar{p})}^{s} \lambda_l^{(s,k)} v_l^k\right\|}. \tag{5.3}$$

Claim 1: All cluster points of $\{v_l^k\}_{k=1}^\infty$ *lie in* $\text{lin}\circ\text{aff}(K_l)\cap N_{K_l}(\bar{x})$ *for* $l=1,\ldots,r$. In view of the fact that the normal cone mapping $N_C(\cdot):C\rightrightarrows\mathbb{R}^n$ has closed graph (see [35, Proposition 6.6]), all cluster points of $\{v_l^k\}_{k=1}^\infty$ lie in $N_{K_l}(\bar{x})$. Since $v_l^k\in\text{lin}\circ\text{aff}(K_l)$, all cluster points of $\{v_l^k\}_{k=1}^\infty$ lie in $\text{lin}\circ\text{aff}(K_l)$. Thus this claim is proved.

Claim 2: For each i, *the infinite sum*

$$z_{l,i}:=\sum_{s=i}^{\infty}\sum_{k=\max(1,s-\bar{p})}^{s}\lambda_l^{(s,k)}v_l^k \tag{5.4}$$

exists as a limit for $l=1,\ldots,r$. *Hence* $\lim_{j\to\infty}\tilde{v}_l^{(i,j)}$ *exists.*

Suppose on the contrary that $z_{l,i}$ does not exist as a limit for some l, $1\leq l\leq r$. It follows that

$$\sum_{s=i}^{\infty}\sum_{k=\max(1,s-\bar{p})}^{s}\lambda_l^{(s,k)}=\infty, \tag{5.5}$$

because if the sum in (5.5) were finite, $z_{l,i}$ would exist as a limit. Note that the cone $\text{lin}\circ\text{aff}(K_l)\cap N_{K_l}(\bar{x})$ is pointed. Using Claim 1 and Lemma 5.7(1), the subsequence $\{\tilde{v}_l^{(i,j)}\}_{j=1}^\infty$ has cluster points in $\text{lin}\circ\text{aff}(K_l)\cap N_{K_l}(\bar{x})$. Let

$$\alpha_{i,j}:=\max_{1\leq l\leq r}\left\{\left\|\sum_{s=i}^{j}\sum_{k=\max(1,s-\bar{p})}^{s}\lambda_l^{(s,k)}v_l^k\right\|,\|\tilde{w}_l^{(i-1,j)}\|\right\}. \tag{5.6}$$

We have $\lim_{j\to\infty}\alpha_{i,j}=\infty$ since (5.5) holds for some l, $\lambda_l^{(s,k)}\geq 0$, and by Lemma 5.6, there is a constant m dependent only on $\text{lin}\circ\text{aff}(K_l)\cap N_{K_l}(\bar{x})$ such that

$$\left\|\sum_{s=i}^{j}\sum_{k=\max(1,s-\bar{p})}^{s}\lambda_l^{(s,k)}v_l^k\right\|\geq m\sum_{s=i}^{j}\sum_{k=\max(1,s-\bar{p})}^{s}\lambda_l^{(s,k)}.$$

Consider the equation

$$\frac{1}{\alpha_{i,j}}[x_{i-1}-x_j]=\sum_{l=1}^{r}\Bigg[\underbrace{\frac{1}{\alpha_{i,j}}\left(\sum_{s=i}^{j}\sum_{k=\max(1,s-\bar{p})}^{s}\lambda_l^{(s,k)}v_l^k\right)}_{t_{l,i,j}}+\underbrace{\frac{1}{\alpha_{i,j}}\tilde{w}_l^{(i-1,j)}}_{t'_{l,i,j}}\Bigg]. \tag{5.7}$$

It is clear that the LHS converges to zero as $j\to\infty$. We can, by the definition of $\alpha_{i,j}$ in (5.6), choose a subsequence such that the limits $t_{l,i}:=\lim_{j\to\infty}t_{l,i,j}$ and $t'_{l,i}:=\lim_{j\to\infty}t'_{l,i,j}$, where $t_{l,i,j}$ and $t'_{l,i,j}$ are defined in (5.7), exist and are not all zero for $1\leq l\leq r$. This would contradict Condition (1), ending the proof of Claim 2.

In view of Claim 2, define

$$\tilde{v}_l^{(i)} := \lim_{j\to\infty} \tilde{v}_l^{(i,j)}. \tag{5.8}$$

Define the matrix $A^{(i,j)} \in \mathbb{R}^{n\times r}$ whose lth column is $\tilde{v}_l^{(i,j)}$. We can write

$$A^{(i,j)}\gamma^{(i,j)} = \sum_{l=1}^{r}\sum_{s=i}^{j}\sum_{k=\max(1,s-\bar{p})}^{s} \lambda_l^{(s,k)} v_l^k, \tag{5.9}$$

where $\gamma^{(i,j)} \in \mathbb{R}^r$ is such that $\gamma_l^{(i,j)} := \|\sum_{s=i}^{j}\sum_{k=\max(1,s-\bar{p})}^{s} \lambda_l^{(s,k)} v_l^k\|$ for $l = 1,\dots,r$. Let $A^{(i)} := \lim_{j\to\infty} A^{(i,j)}$ and $\gamma^{(i)} \in \mathbb{R}^r$ be such that

$$\gamma_l^{(i)} := \|z_{l,i}\| = \left\|\sum_{s=i}^{\infty}\sum_{k=\max(1,s-\bar{p})}^{s} \lambda_l^{(s,k)} v_l^k\right\|.$$

Then

$$A^{(i)}\gamma^{(i)} = \sum_{l=1}^{r} z_{l,i} = \sum_{l=1}^{r}\sum_{s=i}^{\infty}\sum_{k=\max(1,s-\bar{p})}^{s} \lambda_l^{(s,k)} v_l^k.$$

Let

$$\begin{aligned}
\mathcal{A} := \{A \in \mathbb{R}^{n\times r} \mid &\text{ The } l\text{th column of } A \text{ is a} \\
&\text{unit vector in } \operatorname{lin}\circ\operatorname{aff}(K_l)\cap N_{K_l}(\bar{x}) \text{ for } 1 \le l \le r\},\\
L := \bigcap_{l=1}^{r} &\operatorname{lin}\circ\operatorname{aff}(K_l),\\
\text{and } \beta := \inf &\left\{\frac{\|P_L(A\gamma)\|}{\|\gamma\|} \mid A \in \mathcal{A} \text{ and } \gamma \in \mathbb{R}^r_+\backslash\{0\}\right\}.
\end{aligned} \tag{5.10}$$

In view of the fact that $[S_1+S_2]^\perp = S_1^\perp \cap S_2^\perp$ for linear subspaces S_1 and S_2, L also equals

$$L = \left[\sum_{l=1}^{r}[\operatorname{lin}\circ\operatorname{aff}(K_l)]^\perp\right]^\perp. \tag{5.11}$$

Claim 3: $\beta > 0$. Suppose otherwise. Then there are sequences of matrices $\tilde{A}^{(i)} \in \mathcal{A}$ and unit vectors $\tilde{\gamma}^{(i)} \in \mathbb{R}^r$ such that $\tilde{\gamma}^{(i)} \geq 0$ and $P_L(\tilde{A}^{(i)}\tilde{\gamma}^{(i)}) \to 0$ as $i \nearrow \infty$. By taking cluster points of $\tilde{A}^{(i)}$ and $\tilde{\gamma}^{(i)}$, we obtain $P_L(\tilde{A}\tilde{\gamma}) = 0$ for some $\tilde{A} \in \mathcal{A}$ and $\tilde{\gamma} \neq 0$, where $\tilde{\gamma} \geq 0$.

We now check that this contradicts Condition (1). Since $P_L(\tilde{A}\tilde{\gamma}) = 0$, we can find some $\tilde{w} \in L^\perp$ so that $\tilde{A}\tilde{\gamma} + \tilde{w} = 0$. We can write $\tilde{w}$ as $\tilde{w} = \sum_{l=1}^{r} \tilde{w}_l$, where

$\tilde{w}_l \in \text{lin} \circ \text{aff}(K_l)^\perp$. Letting $\tilde{A}_l$ be the lth column of $\tilde{A}$, we see that $\tilde{A}_l\tilde{\gamma}_l + \tilde{w}_l \in [\text{lin} \circ \text{aff}(K_l) \cap N_{K_l}(\bar{x})] + \text{lin} \circ \text{aff}(K_l)^\perp = N_{K_l}(\bar{x})$. Since $\tilde{A}_l\tilde{\gamma}_l + \tilde{w}_l$ are not all zero, this concludes our proof of Claim 3.

Claim 4: $\lim_{i\to\infty}[\inf_{A\in\mathcal{A}} \|A - A^{(i)}\|] = 0$. The lth column of $A^{(i)}$ is the unit vector $\tilde{v}_l^{(i)}$ as defined in (5.3) and (5.8), and each v_l^k lies in $\text{lin} \circ \text{aff}(K_l) \cap N_{K_l}\big(\mathbb{B}_\delta(\bar{x})\big)$, where $\delta = \|x_{k-1} - \bar{x}\|$. Since $\{x_i\}$ converges to $\bar{x}$ and is Fejér monotone, for any $\delta > 0$, we can find i' large enough so that $\|x_i - \bar{x}\| < \delta$ for all $i > i'$. This would mean that for all $\epsilon > 0$, we can find i large enough so that each v_l^k in the sum (5.3) satisfies $\|v_l^k - \bar{v}_l^k\| < \epsilon$ for some unit vector $\bar{v}_l^k \in \text{lin} \circ \text{aff}(K_l) \cap N_{K_l}(\bar{x})$.

Let

$$\hat{v}_l^{(i)} := \frac{\sum_{s=i}^{j} \sum_{k=\max(1,s-\bar{p})}^{s} \lambda_l^{(s,k)} \bar{v}_l^k}{\left\| \sum_{s=i}^{j} \sum_{k=\max(1,s-\bar{p})}^{s} \lambda_l^{(s,k)} \bar{v}_l^k \right\|}.$$

Recall that $\tilde{v}_l^{(i)}$ is the lth column of $A^{(i)}$. By Lemma 5.7(2), there is a constant m dependent only on $\text{lin} \circ \text{aff}(K_l) \cap N_{K_l}(\bar{x})$ such that

$$\|\tilde{v}_l^{(i)} - \hat{v}_l^{(i)}\| \leq \epsilon m.$$

Since $\epsilon \searrow 0$ as $i \nearrow \infty$, we can see that the conclusion to Claim 4 holds.

Claim 5: To prove that the conclusion (5.2) *holds, it suffices to prove*

$$\lim_{i\to\infty} \left(\inf_{A\in\mathcal{A}} \frac{\|A\gamma^{(i)} - A^{(i)}\gamma^{(i)}\|}{\|x_{i-1} - \bar{x}\|} \right) = 0. \tag{5.12}$$

The vector $\gamma^{(i)}$ has nonnegative components, and $x_{i-1} - \bar{x} = A^{(i)}\gamma^{(i)} + \sum_{l=1}^r \tilde{w}_l^{(i)}$ for some $\tilde{w}_l^{(i)} \in [\text{lin} \circ \text{aff}(K_l)]^\perp$. Condition (1) and [35, Theorem 6.42] imply that

$$N_K(\bar{x}) = \sum_{l=1}^{r} N_{K_l}(\bar{x}). \tag{5.13}$$

Then $A\gamma^{(i)} + \sum_{l=1}^r \tilde{w}_l^{(i)}$ would lie in $N_K(\bar{x})$ for any $A \in \mathcal{A}$ by (5.13). Moreau's Theorem implies that for any vector $v \in \mathbb{R}^n$,

$$v = P_{T_K(\bar{x})}(v) + P_{N_K(\bar{x})}(v) \quad \text{and} \quad \|P_{T_K(\bar{x})}(v)\|^2 + \|P_{N_K(\bar{x})}(v)\|^2 = \|v\|^2,$$

If (5.12) holds, then

$$\begin{aligned}\lim_{i\to\infty} \frac{\|P_{T_K(\bar{x})}(x_{i-1} - \bar{x})\|}{\|x_{i-1} - \bar{x}\|} &= \lim_{i\to\infty} \frac{\|[x_{i-1} - \bar{x}] - P_{N_K(\bar{x})}(x_{i-1} - \bar{x})\|}{\|x_{i-1} - \bar{x}\|} \\ &= \lim_{i\to\infty} \left(\min_{v\in N_K(\bar{x})} \frac{\|[x_{i-1} - \bar{x}] - v\|}{\|x_{i-1} - \bar{x}\|} \right)\end{aligned}$$

$$\text{(by the definition of } P_{N_K(\bar{x})}(\cdot)) \\ \leq \lim_{i\to\infty}\left(\inf_{A\in\mathcal{A}}\frac{\|A^{(i)}\gamma^{(i)}-A\gamma^{(i)}\|}{\|x_{i-1}-\bar{x}\|}\right) = 0.$$

(In the last formula, we chose $v=\inf_{A\in\mathcal{A}} A\gamma^{(i)}+\sum_{l=1}^{r}\tilde{w}_l^{(i)}$.) By Moreau's Theorem, (5.2) holds, ending the proof of Claim 5.

Claim 6: (5.12) *holds.* Since $\gamma^{(i)}\geq 0$, it is clear from the definition of β and Claim 4 that if the $\gamma^{(i)}$'s are nonzero, then

$$\liminf_{i\to\infty}\frac{\|P_L(A^{(i)}\gamma^{(i)})\|}{\|\gamma^{(i)}\|}\geq\beta. \tag{5.14}$$

In the case where $\gamma^{(i)}$ are zero, the numerator in (5.12) is zero, so things are straightforward. So we shall look only at the subsequence for which $\gamma^{(i)}$ are nonzero. (We do not relabel.) For the denominator, we have

$$\begin{aligned}\|x_{i-1}-\bar{x}\| &= \left\|A^{(i)}\gamma^{(i)}+\sum_{l=1}^{r}\tilde{w}_l^{(i)}\right\| \\ &\geq \left\|P_L\left(A^{(i)}\gamma^{(i)}+\sum_{l=1}^{r}\tilde{w}_l^{(i)}\right)\right\| = \|P_L(A^{(i)}\gamma^{(i)})\|.\end{aligned}$$

The last equality holds because P_L is linear and, in view of $\tilde{w}_l^{(i)}\in[\text{lin}\circ\text{aff}(K_l)]^{\perp}$ and (5.11), $P_L(\tilde{w}_l^{(i)})=0$ for all l. Then Claim 4 and (5.14) implies

$$0\leq\lim_{i\to\infty}\left(\inf_{A\in\mathcal{A}}\frac{\|A\gamma^{(i)}-A^{(i)}\gamma^{(i)}\|}{\|P_L(A^{(i)}\gamma^{(i)})\|}\right)\leq\lim_{i\to\infty}\frac{\inf_{A\in\mathcal{A}}\|A-A^{(i)}\|}{\beta}=0,$$

from which (5.12) follows easily, ending the proof of Claim 6.

By applying Claim 5 to Claim 6, we prove the result at hand. □

Next, we give conditions for estimating the distance to the point of convergence using the distance to the respective sets. We recall the definition of local linear regularity.

Definition 5.9 (*Local metric inequality*) We say that a collection of closed sets K_l, $l=1,\dots,r$ satisfies the *local metric inequality* at $\bar{x}$ if there are $\beta>0$ and $\delta>0$ such that

$$d(x,\cap_{l=1}^{r}K_l)\leq\beta\max_{1\leq l\leq r}d(x,K_l)\text{ for all }x\in\mathbb{B}_\delta(\bar{x}). \tag{5.15}$$

In this paper, we shall only consider the case where K_l are all convex. The term linear regularity is used in two different ways in [27, after (15)] and [28, Proposition 2.3], so we refrain from using the term here. A concise summary of further studies on the local metric inequality appears in [27], who in turn referred to [4,25,30,32] on the topic of local metric inequality and their connection to metric regularity. Definition 5.9 is sufficient for our purposes. The local metric inequality is useful for proving the linear convergence of alternating projection algorithms [2,28]. See [3] for a survey.

With the additional assumption of local metric inequality, we have the following result.

Lemma 5.10 (Estimates under local metric inequality) *Let $K_l \subset \mathbb{R}^n$, where $1 \leq l \leq r$, be closed convex sets. Suppose a sequence $\{x_i\}$ converges to the point $\bar{x} \in K := \cap_{l=1}^r K_l$, $\{K_l\}_{l=1}^r$ satisfies the local metric inequality at $\bar{x}$, and*

$$\lim_{i\to\infty} \frac{\|P_{N_K(\bar{x})}(x_i - \bar{x})\|}{\|x_i - \bar{x}\|} = 1. \tag{5.16}$$

Then there is a $\beta > 0$ such that

$$\|x_i - \bar{x}\| \leq \beta \max_{1\leq l\leq r} d(x_i, K_l) \text{ for all } i \text{ large enough.} \tag{5.17}$$

Proof By Moreau's Theorem, we have

$$\begin{aligned} \|P_{T_K(\bar{x})}(x_i - \bar{x})\|^2 &= \|x_i - \bar{x}\|^2 - \|P_{N_K(\bar{x})}(x_i - \bar{x})\|^2 \\ \Rightarrow \lim_{i\to\infty} \frac{\|P_{T_K(\bar{x})}(x_i - \bar{x})\|^2}{\|x_i - \bar{x}\|^2} &= \lim_{i\to\infty} \left(1 - \frac{\|P_{N_K(\bar{x})}(x_i - \bar{x})\|^2}{\|x_i - \bar{x}\|^2}\right) = 0. \end{aligned} \tag{5.18}$$

Let $\tilde{x}_i$ be such that $\tilde{x}_i - \bar{x} = P_{N_K(\bar{x})}(x_i - \bar{x})$, and $x_i - \tilde{x}_i = P_{T_K(\bar{x})}(x_i - \bar{x})$. Formulas (5.16) and (5.18) give us

$$\lim_{i\to\infty} \frac{\|\tilde{x}_i - \bar{x}\|}{\|x_i - \bar{x}\|} = 1 \quad \text{and} \quad \lim_{i\to\infty} \frac{\|\tilde{x}_i - x_i\|}{\|x_i - \bar{x}\|} = 0. \tag{5.19}$$

Since $\tilde{x}_i - \bar{x} \in N_K(\bar{x})$, we have $d(\tilde{x}_i, K) = \|\tilde{x}_i - \bar{x}\|$. So, by the Lipschitzness of the projection operation, we have

$$\begin{aligned} d(\tilde{x}_i, K) - \|\tilde{x}_i - x_i\| &\leq d(x_i, K) \leq d(\tilde{x}_i, K) + \|\tilde{x}_i - x_i\| \\ \Rightarrow \|\tilde{x}_i - \bar{x}\| - \|\tilde{x}_i - x_i\| &\leq d(x_i, K) \leq \|\tilde{x}_i - \bar{x}\| + \|\tilde{x}_i - x_i\|. \end{aligned} \tag{5.20}$$

The formulas (5.19) and (5.20) give $\lim_{i\to\infty} \frac{d(x_i,K)}{\|x_i-\bar{x}\|} = 1$. Together with the definition of local metric inequality (5.15), we can obtain what we need. □

Local metric inequality follows from Condition (1) in Proposition 5.8. We paraphrase the result from [28], where the authors remarked that the theorem is well known. For example, a globalized version appears in the survey [1, Theorem 3.7] without attribution.

Lemma 5.11 (Condition for local metric inequality) *Suppose $\bar{x} \in K$, where $K = \cap_{l=1}^r K_l$ and $K_l \subset \mathbb{R}^n$ for $1 \leq l \leq r$, and that Condition (1) of Proposition 5.8 holds. Then $\{K_l\}_{l=1}^r$ satisfies the local metric inequality at $\bar{x}$.*

Proof In [28, Section 3], it was proved that if Condition (1) of Proposition 5.8 holds, then there is a constant $\kappa \geq 0$ such that

$$d\Big(x, \bigcap_i (K_i - z_i)\Big) \leq \kappa \sqrt{\sum_i d^2(x, K_i - z_i)} \text{ for all } (x, z) \text{ near } (\bar{x}, 0),$$

This is easily seen to be stronger than the conclusion since we only need $z_i = 0$ for $1 \leq i \leq r$. □

We state the key result of this section.

Theorem 5.12 (Superlinear convergence) *Consider the problem of finding a point* $x \in K$, *where* $K = \cap_{l=1}^r K_l$ *and* $K_l \subset \mathbb{R}^n$. *Suppose Algorithm 5.1 produces a sequence* $\{x_i\}$ *that converges to a point* $\bar{x} \in K$. *Suppose also that the conditions in Proposition 5.8 hold, i.e.,*

(1) If $\sum_{l=1}^r v_l = 0$ *for some* $v_l \in N_{K_l}(\bar{x})$, *then* $v_l = 0$ *for all* $l = 1, \ldots, r$.

If $\bar{p}$ *in Algorithm 5.1 is sufficiently large, then we have*

$$\limsup_{i\to\infty} \frac{\|x_{i+\bar{p}} - \bar{x}\|}{\|x_i - \bar{x}\|} = 0. \tag{5.21}$$

Moreover, for that choice of $\bar{p}$, *if*

$$\text{for some } \bar{\epsilon} > 0, [K_l - \bar{x}] \cap \bar{\epsilon}\mathbb{B} = T_{K_l}(\bar{x}) \cap \bar{\epsilon}\mathbb{B} \text{ for all } l = 1, \ldots, r, \tag{5.22}$$

then the convergence of $\{x_i\}$ *to* $\bar{x}$ *is finite.*

Proof In Algorithm 5.1, let $l_i \in \{1, \ldots, r\}$ be such that

$$l_i \in \arg\max_{1\leq l\leq r} \|x_i - P_{K_l}(x_i)\| = \arg\max_{1\leq l\leq r} d(x_i, K_l).$$

Let v_i^* be the unit vector $v_i^* := \frac{x_i - P_{K_{l_i}}(x_i)}{\|x_i - P_{K_{l_i}}(x_i)\|}$. In other words, v_i^* is the unit vector of the hyperplane that separates x_i from K_{l_i}.

Without loss of generality, suppose that $\bar{x} = 0$. From Lemma 5.11, we deduce that $\{K_l\}_{l=1}^r$ satisfies the local metric inequality at $\bar{x}$. Suppose $\beta > 0$ is chosen such that (5.17) holds.

The sphere $S^{n-1} := \{w \in \mathbb{R}^n \mid \|w\| = 1\}$ is compact. Suppose $\bar{p}$ is such that we can cover S^{n-1} with $\bar{p}$ balls of radius $\frac{1}{4\beta}$.

Next, among the vectors $\{v_i^*, v_{i+2}^*, \ldots, v_{i+\bar{p}}^*\}$, there must exist j and k such that $i \leq j < k \leq i + \bar{p}$, and v_j^* and v_k^* belong to the same ball of radius $\frac{1}{4\beta}$ covering S^{n-1}. We thus have $\|v_j^* - v_k^*\| \leq \frac{1}{2\beta}$. We can assume, using Theorem 2.1, that i is large enough so that

$$\langle v_j^*, x_k\rangle \leq \epsilon\|x_j\|. \tag{5.23}$$

On the other hand, if i is large enough, we can apply Lemma 5.10 to get

$$\begin{aligned}\langle v_j^*, x_k\rangle &= \langle v_k^*, x_k\rangle + \langle v_j^* - v_k^*, x_k\rangle \\ &\geq d(x_k, K_{l_k}) - \frac{1}{2\beta}\|x_k\| \\ &\geq \frac{1}{2\beta}\|x_k\|. \end{aligned} \tag{5.24}$$

The methods in Theorem 4.5 can be easily adapted to prove that the sequence $\{x_i\}$ is Fejér monotone with respect to K. The inequalities (5.23) and (5.24), and the Fejér monotonicity of $\{x_i\}$ combine to give

$$\|x_{i+\bar{p}}\| \leq \|x_k\| \leq 2\beta\epsilon\|x_j\| \leq 2\beta\epsilon\|x_i\|.$$

As the factor ϵ can be made arbitrarily close to 0, we proved (5.21).

Next, under the added condition (5.22), the formula (5.23) becomes $\langle v_j^*, x_k\rangle \leq 0$ instead by an application of Moreau's Theorem (see Proposition 2.3), and the same steps show us that $\frac{1}{2\beta}\|x_k\| \leq 0$, which forces $x_k = 0$, or $x_k = \bar{x}$. □

Even though the choice of $\bar{p}$ in the proof of Theorem 5.12 is impractical, Theorem 5.12 gives justification that the idea of supporting hyperplanes and quadratic programming can lead to fast convergence.

5.1 Alternative estimates

We close this section with a result that might be helpful for estimating the distance of an iterate to the limit $\bar{x}$.

Lemma 5.13 (Alternative estimate) *Let $K := \cap_{l=1}^r K_l$, where K_l are closed convex sets in $\mathbb{R}^n$ for $1 \leq l \leq r$.*

(1) Let hyperplanes $H_j := \{x \mid \langle a_j, x\rangle = b_j\}$ and points $a_j \in \mathbb{R}^n$ be such that $\|a_j\| = 1$, each H_j is a supporting hyperplane to some K_{l_j}, and $K_{l_j} \subset \{x \mid \langle a_j, x\rangle \leq b_j\}$ for $j = 1, \ldots, J$.
(2) Let $\tilde{x}_j \in \mathbb{R}^n$ be such that $\tilde{x}_j \in H_j \cap K_{l_j}$ for $j = 1, \ldots, J$
(3) Choose $x^* \in \mathbb{R}^n$ so that x^* lies on the all hyperplanes H_j.
(4) Let $\bar{x} \in K$, and let $L := \max_j \|\tilde{x}_j - \bar{x}\|$.
(5) Let $\epsilon > 0$ be such that $-\epsilon \leq \frac{\langle a_j, \bar{x} - \tilde{x}_j\rangle}{\|\bar{x} - \tilde{x}_j\|} \leq 0$ for all $j = 1, \ldots, J$.

Let the matrix $A \in \mathbb{R}^{n\times J}$ be such that the jth column of A is a_j, and let its smallest singular value be $\underline{\sigma}$. We assume $\underline{\sigma} > 0$. Let α be such that

$$\|M\|_{\infty,2} \leq \alpha\|M\|_{2,2} \text{ for all } M \in \mathbb{R}^{n\times J}, \tag{5.25}$$

where $\|M\|_{p,q} := \sup_{v\neq 0} \frac{\|Mv\|_q}{\|v\|_p}$. Let S be span$\{a_1, \ldots, a_J\}$. Then

$$\|P_S(x^* - \bar{x})\| \leq L\epsilon\alpha\underline{\sigma}^{-1}. \tag{5.26}$$

Proof Since x^* is in H_j, we have $\langle a_j, x^* \rangle = b_j$. By Conditions (3) and (4), we get

$$-\epsilon L \leq -\epsilon \|\bar{x} - \tilde{x}_j\| \leq \langle a_j, \bar{x} - \tilde{x}_j \rangle \leq 0.$$

Since $\langle a_j, \tilde{x}_j \rangle = b_j = \langle a_j, x^* \rangle$, we have

$$0 \leq \langle a_j, x^* - \bar{x} \rangle \leq \epsilon L. \tag{5.27}$$

By standard linear least squares, we have

$$\begin{aligned} \|P_S(x^* - \bar{x})\|_2 &= \|A(A^T A)^{-1} A^T (x^* - \bar{x})\|_2 \\ &\leq \|A(A^T A)^{-1}\|_{\infty,2} \|A^T (x^* - \bar{x})\|_\infty \end{aligned}$$

By (5.27), we have $\|A^T(x^* - \bar{x})\|_\infty \leq \epsilon L$. Furthermore, using standard properties of the singular value decomposition, we have

$$\|A(A^T A)^{-1}\|_{\infty,2} \leq \alpha \|A(A^T A)^{-1}\|_{2,2} = \alpha \underline{\sigma}^{-1}.$$

The required bound follows immediately. □

To apply Lemma 5.13 to Algorithm 5.1, note that Condition (4) follows from properties of the projection, while Condition (5) is an attempt to apply Theorem 2.1. Lemma 5.13 is closer to the spirit of Theorem 4.6. However, the term $\underline{\sigma}^{-1}$ is hard to control, so we have not had success in applying Lemma 5.13 so far.

6 Infeasibility

We now discuss the case where the $K := \cap_{l=1}^r K_l = \emptyset$. For any algorithm producing a sequence $\{x_i\}$ in the hope of converging to a limit $\bar{x} \in K$, there are three possibilities:

(1) An infinite sequence cannot be produced because the intersection of the halfspaces is an empty set at some point.
(2) The sequence $\{x_i\}$ contains a cluster point $\bar{x}$.
(3) The sequence $\{x_i\}$ does not contain a cluster point $\bar{x}$.

We first show that case 2 is not possible for Algorithm 3.1 in the case of strong cluster points.

Theorem 6.1 (No cluster point) *For Algorithm 3.1 using (3.1b), in the case where $K = \emptyset$, the sequence $\{x_i\}$ cannot contain a strong cluster point.*

Proof Suppose on the contrary that $\{x_i\}$ contains a strong cluster point, say $\tilde{x}$. Since $\tilde{x} \notin K$, we assume without loss of generality that $\tilde{x} \notin K_1$. Then let $z := P_{K_1}(\tilde{x})$, and $v = \tilde{x} - z$. Let $a_x = x - P_{K_1}(x)$ and $b_x = \langle a_x, P_{K_1}(x) \rangle$. By elementary properties of the projection, we have $\langle a_{\tilde{x}}, \tilde{x} \rangle > b_{\tilde{x}}$. The parameters a_x and b_x depend continuously on x. By the description of Algorithm 3.1, we have

$$\langle a_{x_i}, x_{i+1} \rangle \leq b_{x_i}.$$

As we take limits as $i \to \infty$, we get $\langle a_{\tilde{x}}, \tilde{x} \rangle \leq b_{\tilde{x}}$. This is a contradiction. □

One can easily check that Case 3 can happen. Consider the sets K_1 and K_2 defined by

$$K_1 = \{(x, y) \in \mathbb{R}^2 \mid y \geq e^{-x}\},$$
$$\text{and } K_2 = \{(x, y) \in \mathbb{R}^2 \mid y \leq -e^{-x}\}.$$

If x_0 is chosen to be the origin in Algorithm 3.1, then the iterates x_i cannot converge to a limit by Theorem 6.1, and therefore must move in the direction of the positive x axis. We understand more about such behavior with the result below.

Theorem 6.2 (Recession directions) *If $\{x_i\}$ is a sequence of iterates for Algorithm 3.1 using (3.1b) in the case where $K = \emptyset$ and $X = \mathbb{R}^n$, then any cluster point of $\{\frac{x_i}{\|x_i\|}\}$ must lie in $R(K_l)$, the recession cone of K_l, for all $l = 1, \ldots, r$.*

Proof Let $\{\frac{\tilde{x}_i}{\|\tilde{x}_i\|}\}$ be a subsequence of $\{\frac{x_i}{\|x_i\|}\}$ which has a limit v. We show that such a limit has to lie in $R(K_l)$. Seeking a contradiction, suppose that $v \notin R(K_l)$.

We show that there is a unit vector $w \in \mathbb{R}^n$ and $M \in \mathbb{R}$ such that $\langle w, c \rangle \leq M$ for all $c \in K_l$ and $\langle w, v \rangle > 0$. Take any point $y \in K_l$. Since $v \notin R(K_l)$, there is some $\gamma \geq 0$ such that $y + \gamma v \in K_l$, but $y + \gamma' v \notin K_l$ for all $\gamma' > \gamma$. It follows that there exists a unit vector $w \in N_{K_l}(y + \gamma v)$ such that $\langle w, v \rangle > 0$, and we can take $M = \langle w, y + \gamma v \rangle$. Since $\langle w, v \rangle > 0$, we shall assume that $\langle w, \tilde{x}_i \rangle > M$ for all i.

Let $c_i := P_{K_l}(\tilde{x}_i)$, and let u_i be the unit vector in the direction of $\tilde{x}_i - c_i$. We write $\tilde{x}_i - c_i = \alpha_i u_i$. We have

$$\begin{aligned} \langle u_i, c_i \rangle &= \langle u_i, \tilde{x}_i - \alpha_i u_i \rangle \\ &= \langle u_i, \tilde{x}_i - \alpha_i \rangle. \end{aligned}$$

Also

$$\begin{aligned} \alpha_i \langle w, u_i \rangle &= \langle w, \tilde{x}_i \rangle - \langle w, c_i \rangle \\ &\geq \langle w, \tilde{x}_i \rangle - M. \end{aligned}$$

Since $\alpha_i \langle w, u_i \rangle = \langle w, \tilde{x}_i - c_i \rangle > M - M = 0$, we have $\langle w, u_i \rangle > 0$, and hence $\alpha_i \geq \langle w, \tilde{x}_i \rangle - M$. Therefore,

$$\langle u_i, c_i \rangle \leq \langle u_i, \tilde{x}_i \rangle - \langle w, \tilde{x}_i \rangle + M.$$

By the workings of Algorithm 3.1, we have $\langle u_i, \tilde{x}_i \rangle > \langle u_i, c_i \rangle$ and $\langle u_i, \tilde{x}_j \rangle \leq \langle u_i, c_i \rangle$ for all $j > i$. This gives $\langle u_i, \tilde{x}_j - \tilde{x}_i \rangle \leq 0$, which gives $\langle u_i, v \rangle \leq 0$.

Let u be a cluster point of $\{u_i\}$. We can consider subsequences so that $\lim_{i \to \infty} u_i$ exists. For any point $c \in K_l$, we have

$$\begin{aligned}\langle u, c\rangle &= \lim_{i\to\infty} \langle u_i, c\rangle \\ &\leq \liminf_{i\to\infty} \langle u_i, c_i\rangle \\ &\leq \liminf_{i\to\infty} [\langle u_i, \tilde{x}_i\rangle - \langle w, \tilde{x}_i\rangle + M] \\ &= \liminf_{i\to\infty} \|\tilde{x}_i\| \left(\left\langle u_i, \frac{\tilde{x}_i}{\|\tilde{x}_i\|}\right\rangle - \left\langle w, \frac{\tilde{x}_i}{\|\tilde{x}_i\|}\right\rangle\right) + M \\ &= \liminf_{i\to\infty} \|\tilde{x}_i\| [\langle u_i, v\rangle - \langle w, v\rangle] + M \\ &= -\infty,\end{aligned}$$

which is absurd. The contradiction gives $v \in R(K_l)$. □

7 Conclusion

In this paper, we focus on the theoretical properties of using supporting hyperplanes and quadratic programming to accelerate the method of alternating projections and its variants. It appears that as long as a separating hyperplane is obtained for K and the quadratic programs are not too big, it is a good idea to solve the associated quadratic program to obtain better iterates. Other issues to consider in a practical implementation would be to either remove or combine loose constraints so that the size of the intermediate quadratic programs do not get too big. The ideas in [26] for example can be useful. It remains to be seen whether the theoretical properties in this paper translate to effective algorithms in practice.

Acknowledgments We thank Boris Mordukhovich for asking the question of whether the alternating projection algorithm can achieve superlinear convergence at the ISMP in 2012 during Russell Luke's talk, which led to the idea of considering supporting hyperplanes. We also thank the referees and associate editor for their helpful suggestions and additional references. The author gratefully acknowledges his startup grant at NUS.

References

1. Bauschke, H.H.: Projection algorithms: results and open problems, inherently parallel algorithms. In: Butnariu, D., Censor, Y., Reich, S. (eds.) Feasibility and Optimization and their Applications, pp. 11–22. Elsevier, Amsterdam (2001)
2. Bauschke, H.H., Borwein, J.M.: On the convergence of von Neumann's alternating projection algorithm for two sets. Set-Valued Anal. **1**, 185–212 (1993)
3. Bauschke, H.H., Borwein, J.M.: On projection algorithms for solving convex feasibility problems. SIAM Rev. **38**, 367–426 (1996)
4. Bauschke, H.H., Borwein, J.M., Li, W.: Strong conical hull intersection property, bounded linear regularity, Jameson's property (G), and error bounds in convex optimization. Math. Program. Ser. A **86**(1), 135–160 (1999)
5. Bauschke, H.H., Combettes, P.L., Kruk, S.G.: Extrapolation algorithm for affine-convex feasibility problems. Numer. Algorithms **41**, 239–274 (2006)
6. Bregman, L.M., Censor, Y., Reich, S., Zepkowitz-Malachi, Y.: Finding the projection of a point onto the intersection of convex sets via projections onto half-spaces. J. Approx. Theory **124**, 194–218 (2003)
7. Boyle, J.P., Dykstra, R.L.: A method for finding projections onto the intersection of convex sets in Hilbert spaces. In: Advances in Order Restricted Statistical Inference, Lecture notes in Statistics, pp. 28–47. Springer, New York (1985)

8. Bauschke, H.H., Deutsch, F., Hundal, H.S., Park, S.-H.: Accelerating the convergence of the method of alternating projections. Trans. Am. Math. Soc. **355**(9), 3433–3461 (2003)
9. Birgin, E.G., Raydan, M.: Dykstra's algorithm and robust stopping criteria. In: Floudas, C.A., Pardalos, P.M. (eds.) Encyclopedia of Optimization, 2nd edn, pp. 828–833. Springer, USA (2009)
10. Borwein, J.M., Zhu, Q.J.: Techniques of Variational Analysis. Springer, NY (2005). CMS Books in Mathematics
11. Cegielski, A., Censor, Y.: Opial-type theorems and the common fixed point problem. In: Bauschke, H.H., Burachik, R., Combettes, P.L., Elser, V., Luke, D.R., Wolkowicz, H. (eds.) Fixed-Point Algorithms for Inverse Problems in Science and Engineering, pp. 155–183. Springer, Berlin (2011)
12. Deutsch, F.: The angle between subspaces of a Hilbert space. In: Singh, S.P. (ed.) Approximation Theory Spline Functions and Applications, pp. 107–130. Kluwer, The Netherlands (1995)
13. Deutsch, F.: Accelerating the convergence of the method of alternating projections via a line search: a brief survey. In: Butnariu, D., Censor, Y., Reich, S. (eds.) Inherently Parallel Algorithms in Feasibility and Optimization and their Applications, pp. 203–217. Elsevier, Amsterdam (2001)
14. Deutsch, F.: Best Approximation in Inner Product Spaces. Springer (2001). CMS Books in Mathematics
15. Dykstra, R.L.: An algorithm for restricted least-squares regression. J. Am. Stat. Assoc. **78**, 837–842 (1983)
16. Escalante, R., Raydan, M.: Alternating Projection Methods. SIAM (2011)
17. Fletcher, R., Leyffer, S.: Filter-type algorithms for solving systems of algebraic equations and inequalities. In: di Pillo, G., Murli, A. (eds.) High Performance Algorithms and Software for Nonlinear Optimization, pp. 265–284. Kluwer, Dordrecht (2003)
18. Gearhart, W.B., Koshy, M.: Acceleration schemes for the method of alternating projections. J. Comput. Appl. Math. **26**, 235–249 (1989)
19. García-Palomares, U.M.: A superlinearly convergent projection algorithm for solving the convex inequality problem. Oper. Res. Lett. **22**, 97–103 (1998)
20. García-Palomares, U.M.: Superlinear rate of convergence and optimal acceleration schemes in the solution of convex inequality problems. In: Butnariu, D., Censor, Y., Reich, S. (eds.) Inherently Parallel Algorithms in Feasibility and Optimization and their Applications, pp. 297–305. Elsevier, Amsterdam (2001)
21. Gubin, L.G., Polyak, B.T., Raik, E.V.: The method of projections for finding the common point of convex sets. USSR Comput. Math. Math. Phys. **7**(6), 1–24 (1967)
22. Han, S.P.: A successive projection method. Math. Program. **40**, 1–14 (1988)
23. Hiriart-Urruty, J.-B., Lemaréchal, C.: Convex Analysis and Minimization Algorithms I & II. Springer, Berlin (1993). Grundlehren der mathematischen Wissenschaften, vols. 305 and 306
24. Hundal, H.S.: An alternating projection that does not converge in norm. Nonlinear Anal. **57**(1), 35–61 (2004)
25. Ioffe, A.D.: Metric regularity and subdifferential calculus. Russ. Math. Surv. **55**(3), 501–558 (2000)
26. Kiwiel, K.C.: Block-iterative surrogate projection methods for convex feasibility problems. Linear Algebra Appl. **215**, 225–259 (1995)
27. Kruger, A.Y.: About regularity of collections of sets. Set-Valued Anal. **14**, 187–206 (2006)
28. Lewis, A.S., Luke, D.R., Malick, J.: Local linear convergence for alternating and averaged nonconvex projections. Found. Comput. Math. **9**(4), 485–513 (2009)
29. Moreau, J.-J.: Décomposition orthogonale d'un espace hilbertien selon seux cônes mutuellement polaires. C. R. Acad. Sci. Paris **255**, 238–240 (1962)
30. Ngai, H.V., Théra, M.: Metric inequality, subdifferential calculus and applications. Set-Valued Anal. **9**, 187–216 (2001)
31. Nocedal, J., Wright, S.J.: Numerical Optimization, 2nd edn. Springer, New York (2006)
32. Ng, K.F., Yang, W.H.: Regularities and their relations to error bounds. Math. Program. Ser. A **99**, 521–538 (2004)
33. Pierra, G.: Decomposition through formalization in a product space. Math. Program. **28**, 96–115 (1984)
34. Rockafellar, R.T.: Convex Analysis. Princeton University Press, Princeton (1970)
35. Rockafellar, R.T., Wets, R.J.-B.: Variational Analysis, Grundlehren der mathematischen Wissenschaften, vol. 317. Springer, Berlin (1998)