Fundação Getulio Vargas
Escola de Matemática Aplicada

Guilherme Henrique Macieira de Araújo

Circumcentering outer-approximate projections and reflections for the convex feasibility problem

Rio de Janeiro
2021

Guilherme Henrique Macieira de Araújo

Circumcentering outer-approximate projections and reflections for the convex feasibility problem

Dissertação submetida à Escola de Matemática Aplicada como requisito parcial para a obtenção do grau de Mestre em Modelagem Matemática

Orientador: Roger Behling
Co-Orientador: Luiz Rafael Santos

Rio de Janeiro
2021

Abstract

Recently, circumcenter schemes were applied to solving general convex feasibility problems. In order to overcome costly computations of projections and reflections onto convex sets, we present a variant of the circumcentered-reflection method which employs outer-approximate projections, inspired by Fukushima. With a very practical appeal, this notion relies on separating hyperplanes and is considered in our hybrid method for finding a point in the intersection of finitely many convex sets. We derive convergence in general, linear convergence under an error bound condition, and present successful numerical experiments.

List of Figures

3.1 Illustration of CRM, MAP and DRM for the intersection between an affine U and a convex K. 60
7.1 Illustration of CARM and AMAP for the intersection between an affine U and a convex K. 76
8.1 Experiments with epigraph of a quadratic function and an affine subspace, without error bound - CARM, AMAP, CRM and MAP 89
8.2 Experiments with epigraph of a quadratic function and an affine subspace, without error bound - CARM, AMAP and MAP 90
8.3 Experiments with epigraph of a quadratic function and an affine subspace, with error bound - CARM, AMAP, CRM and MAP 91
8.4 Experiments with epigraph of a quadratic function and an affine subspace, with error bound - CARM, AMAP and MAP 91
8.5 Experiments with epigraph of a quadratic function and an affine subspace, without error bound - CARM and AMAP 92

List of Tables

8.1 Statistics of the experiments without error bound (in number of iterations) 90
8.2 Statistics of the experiments with error bound (in number of iterations) . 92

Contents

1 Introduction . . . 6
2 Preliminary concepts . . . 8
2.1 Notation . . . 8
2.2 Convex sets . . . 9
2.3 Projectors onto convex sets . . . 13
2.4 Separation theorems . . . 15
2.5 Convex functions . . . 18
2.6 Nonexpansive operators . . . 23
2.7 Fixed-point nonexpansive algorithm convergence . . . 26
2.8 Fejér monotonicity . . . 29
2.9 Non-differentiability of convex functions: directional derivatives, subgradients and subdifferential . . . 31
2.10 Best approximation mappings . . . 38
2.11 Linear convergence . . . 46
2.12 Circumcenters . . . 46
3 Review of projection methods for the convex feasibility problem . . . 48
3.1 Pierra's product space reformulation . . . 48
3.2 Method of Alternating Projections - MAP . . . 49
3.3 Douglas-Rachford Method - DRM . . . 52
3.4 Circumcentered Reflection Method - CRM . . . 55
3.4.1 CRM for the best approximation problem in affine subspaces . . . 55
3.4.2 Block-wise CRM for the affine best approximation problem . . . 57
3.4.3 CRM for the convex feasibility problem . . . 59
4 Approximate projections . . . 62
4.1 Outer-approximate projections . . . 62
4.2 Alternative definitions of approximate projections . . . 63
5 Circumcentered Approximate Projection Method - CARM . . . 67
5.1 CARM for general convex intersection . . . 71
6 Approximate Method of Alternating Projections - AMAP . . . 73
7 Comparing CARM and AMAP . . . 76
7.1 Myopic comparison . . . 76
7.2 CARM and AMAP under error bound . . . 78
8 Numerical experiments . . . 85
8.1 Computing outer-approximate projections . . . 85
8.1.1 Toy example with explicit calculations . . . 86
8.2 CARM for the CFP given by the epigraph of a quadratic function and an affine subspace . . . 88
9 Final remarks . . . 93

1 Introduction

We consider the convex feasibility problem (CFP) of finding a point in a nonempty set given by the intersection of finitely many closed convex sets. We are going to reduce CFP to seeking a point common to a closed convex set and an affine subspace, referred to as CFP-red. This reduction in the number of target sets embeds CFP by means of Pierra's product space reformulation [59]. Also, feasibility sets of general convex optimization problems are standardly split into convex inequalities and affine equalities, which define a closed convex set and an affine subspace, respectively.

Projection-based methods are usually utilized for solving CFP. Widely known are the Method of Alternating Projections (MAP), the Douglas-Rachford method (DRM) and the Cimmino method [26]. Recently, the the Circumcentered-Reflection method (CRM) has been developed as a powerful new tool for solving CFP, outperforming MAP and DRM. It was introduced in [18, 19] and further enhanced in [2, 12, 14–16, 20, 22, 29, 30, 52, 58]. CRM was shown in [22] to converge to a solution of CFP-red and it was proven in [2] that linear convergence is obtained in the presence of an error bound condition.

Nevertheless, these algorithms rely on the computation of exact orthogonal projections. One has to bear in mind though that, context depending, this task alone might be as hard as the original problem itself. In this regard, we present in this work a version of CRM employing *outer-approximate projections*.

The outer-approximate projections we are going to consider, formally given in Definition 4.1, are more tractable than the exact ones while still enjoying some of their properties. This notion is based on the paper [38] by Fukushima from 1983. We prove that employing these outer-approximate projections within CRM preserves the convergence results from [22] and, under a Lipschitzian error bound, linear convergence as in [2] is maintained as well. In this view, the results shown here are a paramount of the aforementioned works [2, 22]. Our method will be referred to as Circumcentered-Approximate Reflection Method (CARM). Despite featuring outer approximations, CARM performs competitively to CRM in our numerical experiments.

This paper is organized as follows. In Section 2, we recall concepts such as convexity, nonexpansivity, Fejér monotonicity, Q and R-linear convergence and the characterization of circumcenters, while stating and proving facts relevant to the problems and methods of our interest. In Section 3, we state Pierra's product space reformulation and review MAP, DRM and CRM, while proving key convergence results and presenting variants. In Section 4, we present our notion of outer-approximate projection as well as other notions of approximating orthogonal projections. In Section 5, we introduce CARM and prove its convergence. For the sake of completeness, we carry out a similar study for a corresponding outer-approximate MAP (AMAP) in Section 6, in view of numerical comparisons with CARM. Section 7 features a theoretical comparison between CARM and AMAP and their behavior under an error bound condition. In the presence of an error bound, both CARM and AMAP converge linearly to a solution of CFP-red with

CARM having a better linear rate than AMAP. Section 8 contains exploratory numerical experiments, considering problems where exact projections are available and comparing CARM with CRM (as well as MAP and AMAP). In these experiments, CARM performs successfully. Section 9 presents concluding remarks.

2 Preliminary concepts

The aim of this section is to familiarize the reader with the concepts of convexity, projections, fixed-point operators and circumcenters, as well as ancillary theory.

This dissertation is intended to be self-contained, except for the use of some well-known results in calculus and analysis. By facts, we denote propositions, theorems, lemmas and corollaries that have been first presented and proven in previously published works; a small subset of their proofs, mostly of auxiliary technical results, will be omitted, in which case the original proof will be referred to.

It follows that all statements denoted as propositions, theorems, lemmas and corollaries are new to this work. For ease of reading and understanding, however, definitions will be called as such regardless of their authorship, though context is sufficient for determining whether they are new to this work or not.

While this work's definitions, facts and propositions will be presented with respect to $\mathbb{R}^n$, a great deal can be generalized to a Hilbert space $\mathcal{H}$ with extended-reals-valued functions (see [6,60], for reference). In this more general setting, it is often required that functions be lower continuous and proper. The function $f : \mathcal{X} \to [-\infty, +\infty]$, where $\mathcal{X}$ is nonempty, is lower continuous if, for every net $(x_a)_{a\in A}$ in $\mathcal{X}$, $x_a \to x \Rightarrow f(x) \leq \liminf (x_a)$; it is proper if there exists at least one $x \in \mathcal{X}$ such that $f(x) < +\infty$ and $f(x) > -\infty$ for all $x \in \mathcal{X}$. For more on this, see [60, Sections 7 and 30] and [6, Chapter 1 and 9], among others.

First, we fix notation and recollect some basic facts.

2.1 Notation

Throughout the paper, $\langle \cdot, \cdot \rangle$ stands for the Euclidean inner product and its induced norm $\|\cdot\| := \sqrt{\langle \cdot, \cdot \rangle}$. The nonnegative integer numbers will be denoted by $\mathbb{N}$, the nonnegative real numbers by $\mathbb{R}_+$ and the positive real numbers by $\mathbb{R}_{++}$. The distance between a point $x \in \mathbb{R}^n$ and a set $X \subset \mathbb{R}^n$ is given by $\operatorname{dist}(x, X) := \inf_{y\in X} \|x - y\|$. We denote the set of fixed points of the operator T by $\operatorname{Fix} T$, where $\operatorname{Fix} T := \{x \mid T(x) = x\}$. By $\to$ we represent a single-valued relationship, while $\rightrightarrows$ represents a set-valued relationship. $(x^k)_{k\in\mathbb{N}}$ denotes the sequence $\{x^0, x^1, \ldots\}$. By $B(x; r)$ we denote an open ball with center x and radius r. By $\triangle[A, B, C]$ we denote a triangle with vertices A, B and C, and by $\angle[A, B, C]$ we denote the angle formed by the sides AB and BC. When we take a limit $x \to 0+$, we mean that x approaches 0 from the positive side, while $x \to 0-$ means that x approaches 0 from the negative side. By $f(t) = o(t)$ we denote a function $f : \mathbb{R}_+ \to \mathbb{R}$ with the property that $f(t)/t \to 0$ when $t \to 0+$. By $\operatorname{cl} X$ we denote the closure of the set X and by $\operatorname{bdry} X$ we denote the boundary of X.

2.2 Convex sets

We begin by reminding the fundamental notion of convex sets.

Definition 2.1 (Convex sets)**.** A subset K of $\mathbb{R}^n$ is a *convex set* if for all $\alpha \in [0,1]$, for all $x, y \in K$ we have that $\alpha x + (1-\alpha)y \in K$. In particular, $\mathbb{R}^n$ and $\varnothing$ are convex.

Many properties about convex sets can be derived straight from its definition, e.g. that the finite sum of convex sets is convex. While we will not state all properties of convex sets, we will state a few in particular which will be used in further proofs and arguments.

Though these facts are very well-established and are likely to be found in any convex optimization textbook, notation, hypotheses and order of presentation (consequently, previously proven facts) may vary slightly depending on the source. For the sake of internal consistency, we will follow mostly along the lines of [44] for facts on convex sets and functions.

The first fact we will present is that the intersection of convex sets is also a convex set.

Fact 2.2 (Intersection of convex sets is convex)**.** *Let $K_i \subset \mathbb{R}^n$, $i \in I$ (where I is a possibly infinite set), be convex sets. Then the intersection $K = \cap_{i \in I} K_i$ is also a convex set.*

Proof. Let $x \in K$ and $y \in K$. From the definition of intersection, $x \in K_i$ and $y \in K_i$ for all $i \in I$. Since the sets K_i are convex, $\alpha x + (1-\alpha)y \in K_i$ for any $\alpha \in [0,1]$ and all $i \in I$. Thus, by the definition of intersection, $\alpha x + (1-\alpha)y \in K$. Hence, K is convex. □

One characterization of convex sets is that a set is convex if, and only if, it contains all convex combinations of its points. This statement, as well as the definition of convex combination, is presented below.

Fact 2.3 (Convex combination characterization of convex sets)**.** *A set $K \subset \mathbb{R}^n$ is convex if, and only if, for any $p \in \mathbb{N}$, $x^i \in K$ and $\alpha_i \in [0,1]$, $i = 1, \ldots, p$, such that $\sum_{i=1}^p \alpha_i = 1$, the* convex combination *$\sum_{i=1}^p \alpha_i x^i \in K$.*

Proof. If K contains every convex combinations of its points, in particular it contains the convex combinations of any two points (that is, for $p = 2$). Thus, K is convex by Definition 2.1.

Assume now that K is convex. The proof is done by induction with respect to the number of points $p \in \mathbb{N}$. For given $p \in \mathbb{N}$ and $x^i \in K$, $\alpha_i \in [0,1]$, $i = 1, \ldots, p$ such that $\sum_{i=1}^p \alpha_i = 1$, we define $x = \sum i = 1^p \alpha_i x^i$. For the base step, if $p = 1$, we have that $\alpha_1 = 1$ and, thus, $x = x^1 \in K$. Now, suppose any convex combination of $j \geq 1$ points of K belongs to K, and consider the case $p = j + 1$. If $\alpha_{j+1} = 1$, then $\alpha_i = 0$ for all

$i = 1, \ldots, j$. In this case, $x = x^{j+1} \in K$. Let $\alpha_{j+1} \in [0, 1)$. Since $1 - \alpha_{j+1} > 0$, we can write

$$\begin{aligned} x = \sum_{i=1}^{j+1} \alpha_i x^i &= (1 - \alpha_{j+1}) \sum_{i=1}^{j} \frac{\alpha_i}{1 - \alpha_{j+1}} x^i + \alpha_{j+1} x^{j+1} \\ &= (1 - \alpha_{j+1}) y + \alpha_{j+1} x^{j+1}, \end{aligned} \tag{2.1}$$

where

$$y = \sum_{i=1}^{j} \beta_i x^i, \quad \beta_i = \frac{\alpha_i}{1 - \alpha_{j+1}} > 0, \quad i = 1, \ldots, j.$$

Since $1 = \sum_{i=1}^{j+1} \alpha_i = \sum_{i=1}^{j} \alpha + \alpha_{j+1}$, it follows that

$$\begin{aligned} \sum_{i=1}^{j} \beta_i &= (1 - \alpha_{j+1})^{-1} \sum_{i=1}^{j} \\ &= (1 - \alpha_{j+1})^{-1} (1 - \alpha_{j+1}) = 1. \end{aligned}$$

Thus, y is a convex combination of j points of K. By the induction hypothesis, $y \in K$. Now, (2.1) shows that x is a convex combination of two points of K, in particular, y and x^{j+1}. Since K is convex, we get that $x \in K$, which completes the proof. □

Also regarding convex combinations, a well-known fact known as Carathéodory's theorem states that any convex combination of points of a set in $\mathbb{R}^n$ can be expressed as a convex combination of at most $n + 1$ points of the set.

Fact 2.4 (Carathéodory theorem)**.** *Let $x \in \mathbb{R}^n$ be a convex combination of points of the set $X \subset \mathbb{R}^n$. Then there exist $x^i \in X$ e $\alpha_i \in R_+, i = 1, \ldots, n + 1$ such that $x = \sum_{i=1}^{n+1} \alpha_i x^i, \sum_{i=1}^{n+1} \alpha_i = 1$.*

Proof. By assumption, there exists $p \in \mathbb{N}$ such that

$$x = \sum_{i=1}^{p} \beta_i x^i, x^i \in X, \beta_i \in \mathbb{R}_+, i = 1, \ldots, p, \sum_{i=1}^{p} \beta_i = 1. \tag{2.2}$$

We will show that if $p > n + 1$, then x can be written as a convex combination of $p - 1$ points of X.

If in (2.2) we have $\beta_i = 0$ for some $i \in \{1, \ldots, p\}$, x already is a convex combination of $p - 1$ points of X (we can eliminate from (2.2) the corresponding point x^i.) Suppose then that $\beta_i > 0$ for all $i = 1, \ldots, p$. Since $p > n + 1$, we have that whichever $p - 1 > n$ elements of $\mathbb{R}^n$ are linearly dependent. Thus, $\{x^i - x^p, i = 1, \ldots, p - 1\}$ are linearly dependent, that is, there exist $\gamma_i \in \mathbb{R}, i = 1, \ldots, p - 1$ (at least one not equal to 0) such that

$$0 = \sum_{i=1}^{p-1} \gamma_i \left(x^i - x^p \right). \tag{2.3}$$

We can assume that $\gamma_j > 0$ for some $j \in \{1, \ldots, p-1\}$ (if $\gamma_i \leq 0$ for all $i = 1, \ldots, p-1$, we can multiply (2.3) by -1 to change signs). We define $\gamma_p >= -\sum_{i=1}^{p-1} \gamma_i$, from which it follows that

$$\sum_{i=1}^{p} \gamma_i = \sum_{i=1}^{p-1} \gamma_i + \gamma_p = 0, \tag{2.4}$$

and, furthermore,

$$\begin{aligned}
\sum_{i=1}^{p} \gamma_i x^i &= \sum_{i=1}^{p-1} \gamma_i x^i + \gamma_p x^p \\
&= \sum_{i=1}^{p-1} \gamma_i x^i - \left(\sum_{i=1}^{p-1} \gamma_i \right) x^p \\
&= \sum_{i=1}^{p-1} \gamma_i \left(x^i - x^p \right) = 0, \qquad (2.5)
\end{aligned}$$

where the last inequality follows from (2.3).

Let $j \in \{1, \ldots, p\}$ be any index satisfying

$$\frac{\gamma_j}{\beta_j} = \max_{i=1,\ldots,p} \frac{\gamma_i}{\beta_i}.$$

We observe that $\gamma_j > 0$, since there exists $i \in \{1, \ldots, p\}$ such that $\gamma_i > 0$ ($\beta_i > 0$ for all $i \in \{1, \ldots, p\}$). Define

$$q_i := \beta_i - \frac{\beta_j}{\gamma_j} \gamma_i, i = 1, \ldots, p.$$

Note that this definition implies that $q_j = 0$. For an index i such that $\gamma_i \leq 0$, we have that $q_i \geq \beta_i > 0$. For an index i such that $\gamma_i > 0$, by the definition of the index j, we have that

$$\frac{\beta_i}{\gamma_i} \geq \frac{\beta_j}{\gamma_j} > 0, i = 1, \ldots, p,$$

which implies $q_i \geq 0$. Therefore, $q_i \geq 0 = q_j$ for all indices $i = 1, \ldots, p$. Moreover,

$$\sum_{i=1, i \neq j}^{p} q_i = \sum_{i=1}^{p} q_i = \sum_{i=1}^{p} \beta_i - \frac{\beta_j}{\gamma_j} \sum_{i=1}^{p} \gamma_i = \sum_{i=1}^{p} \beta_i = 1,$$

where we used (2.4) and (2.2). We also have that

$$x = \sum^{p} \beta_i x^i = \sum_{i=1}^{p} q_i x^i + \frac{\beta_j}{\gamma_j} \sum_{i=1}^{p} \gamma_i x^i = \sum_{i=1}^{p} q_i x^i = \sum_{i=1, i \neq j}^{p} q_i x^i,$$

in which we used (2.5).

Repeating this argument $p - (n+1)$ times yields the desired result. $\square$

Now, we introduce the concept of convex hull (or convex closure) of a set, which is the smallest convex set that contains that set. In the subsequent fact, we characterize the convex hull of a given set as the set of all convex combinations of that set's points.

Definition 2.5 (Convex hull (or convex closure))**.** Let $X \subset \mathbb{R}^n$ be any set. The *convex hull (or convex closure)* of X, denoted by $\operatorname{conv} X$, is the smallest convex set in $\mathbb{R}^n$ which contains X (or, equivalently, the intersection of all convex sets in $\mathbb{R}^n$ which contain X).

Fact 2.6 (Convex hull is the set of all convex combinations)**.** *Let $X \subset \mathbb{R}^n$ be any set. Then $\operatorname{conv} X$ is the set of all convex combinations of points of X.*

Proof. We define $C \subset \mathbb{R}^n$ as the set of all convex combinations of points of X, that is,

$$C := \{x \in \mathbb{R}^n \mid \exists p \in \mathbb{N} \text{ s.t. } x = \sum_{i=1}^{p} \alpha_i x^i, \sum_{i=1}^{p} \alpha_i = 1, x^i \in X, \alpha_i \in \mathbb{R}_+, i = 1, \dots, p\}.$$

The set $\operatorname{conv} X$ is convex. Thus, by Fact 2.3, $\operatorname{conv} X$ contains all convex combinations of points of $\operatorname{conv} X$ and, therefore, of X, since $X \subset \operatorname{conv} X$. We then have that $C \subset \operatorname{conv} X$.

Clearly, $X \subset C$. Therefore, if C is convex, we must have that $\operatorname{conv} X \subset \operatorname{conv} C \subset C$. We will now prove that C is convex.

Let $z^1, z^2 \in C$, that is,

$$z^1 = \sum_{i=1}^{p_1} \alpha_i x^i, x^i \in K, \alpha_i \in \mathbb{R}_+, \sum_{i=1}^{p_1} \alpha_i = 1, p_1 \in \mathbb{N},$$
$$z^2 = \sum_{i=1}^{p_2} \beta_i y^i, y^i \in K, \beta_i \in \mathbb{R}_+, \sum_{i=1}^{p_2} \beta_i = 1, p_2 \in \mathbb{N}.$$

For any $\gamma \in [0, 1]$,

$$\gamma z^1 + (1-\gamma) z^2 = \sum_{i=1}^{p_1} \gamma \alpha_i x^i + \sum_{i=1}^{p_2} (1-\gamma) \beta_i y^i.$$

Clearly, $\gamma \alpha_i \geq 0$ for $i = 1, \dots, p_1$ and $(1-\gamma)\beta_i \geq 0$ for $i = 1, \dots, p_2$. Furthermore,

$$\sum_{i=1}^{p_1} \gamma \alpha_i + \sum_{i=1}^{p_2} (1-\gamma)\beta_i = \gamma + (1-\gamma) = 1,$$

which shows that the point $\gamma z^1 + (1-\gamma) z^2$ is a convex combination of points of X, in particular, of the points

$$\{x^i, i = 1, \dots, p_1\} \cup \{y^i, i = 1, \dots, p_2\}.$$

Thus, $\gamma z^1 + (1-\gamma) z^2 \in C$, from which we conclude that C is convex. $\square$

We present further facts regarding the convex hull of bounded and compact sets.

Fact 2.7 (Convex hull of bounded and compact sets)**.** *If $X \subset \mathbb{R}^n$ is a bounded set, then* $\operatorname{conv} X$ *is bounded. If $X \subset \mathbb{R}^n$ is a compact set, then* $\operatorname{conv} X$ *is compact.*

Proof. Let X be bounded. Take any $x \in \operatorname{conv} X$. By Fact 2.3, there exist $x^i \in X$, $\alpha_i \in \mathbb{R}_+$, $i = 1, \ldots, p$ such that $x = \sum_{i=1}^p \alpha_i x^i$, where $\sum_{i=1}^p \alpha_i = 1$. From the triangle inequality,

$$\|x\| \leq \sum_{i=1}^p |\alpha_i| \|x^i\| \leq \sup_{y \in X} \|y\| \sum_{i=1}^p \alpha_i = \sup_{y \in X} \|y\| < +\infty,$$

since X is bounded. Thus, $\operatorname{conv} X$ is bounded.

Now, assume that X is also closed, thus compact. Let $(y^k)_{k \in \mathbb{N}}$ be a sequence such that $y^k \in \operatorname{conv} K$ for all $k \in \mathbb{N}$ and $\lim_{k \to \infty} y^k = y$. From the Carathéodory theorem (Fact 2.4), y^k can be represented as a convex combination of $n+1$ points of K:

$$y^k = \sum_{i=1}^{n+1} \alpha_{k,i} x^{k,i}, \quad \sum_{i=1}^{n+1} \alpha_{k,i} = 1, x^{k,i} \in K, \alpha_{k,i} \in \mathbb{R}_+, i = 1, \ldots, n+1.$$

From the hypothesis that K is bounded, we have that all sequences $(x^{k,i})_{k \in \mathbb{N}}, i = 1, \ldots, n+1$ are bounded. Evidently, the sequences $(\alpha_{k,i})_{k \in \mathbb{N}}, i = 1, \ldots, n+1$ are also bounded. Let x^i and α_i, $i = 1, \ldots, n+1$, be any cluster points of those sequences that correspond to the same subsequence of indices $(k_j), j \to \infty$. Since K is closed, we have that $x^i \in K$,$i = 1, \ldots, n+1$. It is also clear that $\alpha_i \geq 0$, $i = 1, \ldots, n+1$, $\sum_{i=1}^{n+1} \alpha_i = 1$. Therefore,

$$\lim_{k \to \infty} y^k = \lim_{j \to \infty} y^{k_j} = \lim_{j \to \infty} \left(\sum_{i=1}^{n+1} \alpha_{k_j,i} x^{k_j,i} \right) = \sum_{i=1}^{n+1} \alpha_i x^i \in \operatorname{conv} X,$$

which shows that $\operatorname{conv} X$ is closed, hence compact. □

2.3 Projectors onto convex sets

We start this subsection by reminding the definition of projection and reflection operators.

Definition 2.8 (Projection and reflection operators)**.** Given a nonempty subset $X \subseteq \mathbb{R}^n$, the projection operator (or projector) onto S is the operator $P_X : \mathbb{R}^n \rightrightarrows X$ defined at each $x \in \mathbb{R}^n$ by

$$P_X(x) := \{p \in X : \|x - p\| = \operatorname{dist}(x, S)\},$$

where $p \in P_X(x)$ is called a projection of x onto X or a best approximation of x from X.

The reflection operator (or reflector) with respect to X is the operator $R_S : \mathbb{R}^n \Rightarrow \mathbb{R}^n$ given by

$$R_X := 2P_X - \mathrm{Id}$$

Generally speaking, P_X is a set-valued operator. However, for convex sets, as if of our interest, we have that P_X is a nonempty, single-valued mapping that associates each point $x \in \mathbb{R}^n$ with its unique nearest point in X, as shown by the following fact, known as the projection theorem.

Fact 2.9 (Projection theorem)**.** *Let $K \subseteq \mathbb{R}^n$ be nonempty, closed and convex. Then K is a Chebyshev set (that is, every point $x \in K$ has an unique projection p) and*

$$p = P_K(x) \iff p \in C \text{ and } \langle y - p, x - p\rangle \leq 0 \text{ for all } y \in C$$

Proof. Let $p, q \in P_K(x)$. Since K is closed, $P_K(x)$ is nonempty. Since K is convex, we have that $\frac{(p+q)}{2} \in K$. By the parallelogram law and the definition of projection, we have

$$\begin{aligned}\|p - q\|^2 &= \|(p - x) - (q - x)\|^2 = 2\|p - x\|^2 + 2\|q - x\|^2 - \|p + q - 2x\|^2 \\ &= 4\,\mathrm{dist}(x, K)^2(x) - 4\left\|\frac{p+q}{2} - x\right\|^2 \leq 0,\end{aligned}$$

which implies $p = q$, thus proving the uniqueness of the projection.

Let $x \in \mathbb{R}^n$ and $p = P_K(x)$. Pick $y \in K$ and define $z_\lambda := (1 - \lambda)p + \lambda y \in K$ for all $\lambda \in (0, 1)$. Then

$$\begin{aligned}\|x - p\|^2 \leq \|x - z_\lambda\|^2 &= \|x - p - \lambda(y - p)\|^2 \\ &= \|x - p\|^2 + \lambda^2\|y - p\|^2 - 2\lambda\langle y - p, y - p\rangle.\end{aligned}$$

Since $\lambda > 0$, we deduce $\langle s-p, c-p\rangle \leq \frac{\lambda}{2}\|y-p\|^2$. Letting $\lambda \to 0$ then yields $\langle x-p, y-p\rangle \leq 0$.

Now, suppose $p \in K$ satisfies $\langle x - p, y - p\rangle \leq 0$. Then, for all $y \in S$, we have

$$0 \geq 2\langle y - p, x - p\rangle = \|y - p\|^2 + \|x - p\|^2 - \|y - x\|^2 \Longrightarrow \|y - x\| \geq \|x - p\|.$$

Therefore $p = P_K(x)$. □

Note that the uniqueness of the projection implies the uniqueness of the reflection.

If X is closed and convex, the orthogonal projection of any point $x \in \mathbb{R}^n$ onto X is well-defined and is denoted throughout the text by $P_X(x)$. As for the reflection of x through X, it is given by $R_X(x) := (2P_X - \mathrm{Id})(x)$, where Id stands for the identity.

We end this subsection with a corollary of the projection theorem regarding affine subspaces; a nonempty set $U \subseteq \mathbb{R}^n$ is an *affine subspace* if $\lambda x_1 + (1 - \lambda)x_2 \in U$ for all $\lambda \in \mathbb{R}$ and $x_1, x_2 \in U$.

Fact 2.10 (Affine projectors)**.** *Let $U \subseteq \mathbb{R}^n$ be a closed affine subspace. Then the following hold:*

(i) *Let x, p be in $\mathbb{R}^n$. Then $p = P_U(x)$ if and only if $p \in U$ and, for all $y, z \in U$,*

$$\langle y - z, x - p\rangle = 0.$$

(ii) *P_U is an affine operator.*

Proof. (i) Let $y \in U$ and $z \in U$. By the definition of an affine subspace, $2P_U(x-y) = 2P_C x + (1-2)y \in U$ and Fact 2.9 therefore yields $\langle y - P_U(x), x - P_U(x)\rangle \leq 0$ and $\langle y - P_U(x), x - P_U(x)\rangle = \langle (2P_U(x-y)) - P_U(x), x - P_U(x)\rangle \leq 0$. Altogether, $\langle y - P_U(x), x - P_U(x)\rangle = 0$. Likewise, $\langle z - P_U(x), x - P_U(x)\rangle = 0$. By subtraction, $\langle y - z, x - P_U(x)\rangle = 0$.

Conversely, it is clear that $p \in U$, $\langle y - z, x - p\rangle = 0$ for all $y \in U$ and $z \in U$ implies the right-hand side of Fact 2.9.

(ii) Let $y \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$, and set $z = \alpha x + (1-\alpha)y$ and $p = \alpha P_U(x) + (1-\alpha)P_U(y)$. We derive from (i) and the definition of affine subspace that $p \in U$. Now fix u and v in U. Then we also derive from (i) that $\langle u - v, z - p\rangle = \alpha \langle u - v, x - P_U(x)\rangle + (1-\alpha)\langle u - v, y - P_U(y)\rangle = 0$. Altogether, it follows from (i) that $p = P_C(z)$.

□

2.4 Separation theorems

An important notion regarding nonempty convex sets with empty intersection is the existence of a hyperplane that separates both sets. This notion is essential to the definition of the outer-approximate projection, which we will present in Definition 4.1, and to prove several other key statements regarding convex sets.

We begin by presenting the definition of separating hyperplanes.

Definition 2.11 (Separating hyperplanes)**.** Let $X_1, X_2 \subset \mathbb{R}^n$ be nonempty sets and $H(a,c) = \{x \in \mathbb{R}^n \mid \langle a, x\rangle = c\}$ a hyperplane, where $a \in \mathbb{R}^n \setminus \{0\}, c \in \mathbb{R}$. We say that the hyperplane $H(a,c)$ *separates* the sets X_1 and X_2 if

$$\left\langle a, x^1 \right\rangle \leq c \leq \left\langle a, x^2 \right\rangle \text{ for all } x^1 \in X_1, \text{ for all } x^2 \in X_2.$$

We say that $H(a,c)$ *strictly separates* X_1 and X_2 it the inequalities above are strict.

The following two facts state the existence of hyperplanes that separate a point in the closure a set and a point in the boundary of a set, respectively, from that set.

Fact 2.12 (Minkowski's lemma)**.** *Let $K \subset \mathbb{R}^n$ a nonempty convex set. If $x \notin \operatorname{cl} K$, then there exist $a \in \mathbb{R}^n \setminus \{0\}$ and $c \in \mathbb{R}$ such that*

$$\langle a, x\rangle = c, \langle a, y\rangle > c \textit{ for all } y \in K.$$

Proof. Since K is convex, from Definition 2.1 it can be easily shown that $\operatorname{cl} K$ is closed and convex. Let $\bar{x}$ be the projection of x onto $\operatorname{cl} K$, whose existence and uniqueness are

assured by Fact 2.9. Define $a := x - \bar{x}$ ($a \neq 0$ since $x \notin \operatorname{cl} K$) and $c := \langle \bar{x} - x, x \rangle$. By definition, $\langle a, x \rangle = c$. Let $y \in K$. We have that

$$\langle a, y \rangle = \langle \bar{x} - x, y \rangle \geq \langle \hat{x} - x, \hat{x} \rangle \,, \tag{2.6}$$

where the inequality is valid for all $y \in \operatorname{cl} K$ (and therefore for all $y \in K$) by the projection theorem (Fact 2.9). Note that we have

$$\begin{aligned} \langle \bar{x} - x, \bar{x} \rangle &= \|x - \bar{x}\|^2 + \langle \bar{x} - x, x \rangle \\ &= \|x - \bar{x}\|^2 + c > c, \end{aligned}$$

where we used that $\|\hat{x} - x\| > 0$ (from the hypothesis that $x \notin \operatorname{cl} K$). Combining the last relationship with (2.6), we obtain the desired result. □

Fact 2.13 (Supporting hyperplane existence)**.** *Let $K \subset \mathbb{R}^n$ be a nonempty convex set. If $x \in \operatorname{bdry} K$, then there exist $a \in \mathbb{R}^n \setminus \{0\}$ and $\in \mathbb{R}$ such that*

$$\langle a, x \rangle = c, \langle a, y \rangle \geq c \text{ for all } y \in K.$$

The hyperplane $H(a, c) = \{x \in \mathbb{R}^n \mid \langle a, x \rangle = c\}$, which separates $x \in \operatorname{bdry} K$ from K, is called the supporting hyperplane *of K at $x \in \operatorname{bdry} K$.*

Proof. It can be easily shown that $x \in \operatorname{bdry} K = \operatorname{bdry} \operatorname{cl} K$. Thus, we can choose a sequence $(x^k)_{k \in \mathbb{N}} \to x$ such that $x^k \notin \operatorname{cl} K$ for all k.

By Fact 2.12, every point x^k can be strictly separated from K, that is, for all k there exist $\alpha^k \in \mathbb{R}^n \setminus \{0\}$ and $c_k \in \mathbb{R}$ such that

$$\left\langle a^k, x^k \right\rangle = c^k, \left\langle a^k, y \right\rangle > c^k \text{ for all } y \in K. \tag{2.7}$$

Taking a subsequence, if necessary, we can state that $\left(a^k / \left\|a^k\right\|\right) \to a \neq 0$ as $k \to \infty$. From this, it follows

$$\lim_{k \to \infty} c_k / \left\|a^k\right\| = \lim_{k \to \infty} \left\langle a^k, x^k \right\rangle / \left\|a^k\right\| = \langle a, x \rangle$$

Defining $c := \langle a, x \rangle$, dividing the second relationship in (2.7) by $\left\|a^k\right\| > 0$, and taking the limit when $k \to \infty$, we get that for all $y \in K$ we have that

$$\langle a, y \rangle \geq c,$$

which concludes the proof. □

With the help of the previous facts, we can present the main fact of this subsection, known as the hyperplane separation theorem, in its general and strict (which requires stronger hypotheses) versions.

Fact 2.14 (Hyperplane separation theorems)**.** (i) *(Separation theorem) Let K_1, K_2 be convex nonempty subsets of $\mathbb{R}^n$ such that $K_1 \cap K_2 = \emptyset$. Then there exist $a \in \mathbb{R}^n \setminus \{0\}$ and $c \in \mathbb{R}$ such that*

$$\left\langle a, x^1 \right\rangle \leq c \leq \left\langle a, x^2 \right\rangle \text{ for all } x^1 \in K_1, \text{ for all } x^2 \in K_2.$$

(ii) *(Strict separation theorem) Let K_1, K_2 be closed, convex and nonempty subsets of $\mathbb{R}^n$. Suppose that one of them is bounded, thus compact. Then $K_1 \cap K_2 = \emptyset$ if, and only if, there exist $a \in \mathbb{R}^n \setminus \{0\}$ and $c \in \mathbb{R}$ such that such that*

$$\left\langle a, x^1 \right\rangle < c < \left\langle a, x^2 \right\rangle \text{ for all } x^1 \in K_1, \text{ for all } x^2 \in K_2.$$

Proof. (i) It can be easily show that the set $K = K_2 - K_1$ is convex and $0 \notin K$ since $K_1 \cap K_2 = \emptyset$. In this situation, there are two possibilities: either $0 \notin \operatorname{cl} K$, or $0 \in \operatorname{bdry} K$. To separate 0 from K, we use Fact 2.12 for this first case and Fact 2.13 for the second. We conclude that there exists $a \in \mathbb{R}^n \setminus \{0\}$ such that

$$\langle a, x \rangle \geq 0 \text{ for all } x \in K,$$

that is, $\langle a, x^2 - x^1 \rangle \geq 0$ for all $x^2 \in K_2, x^1 \in K_1$, or yet,

$$\left\langle a, x^2 \right\rangle \geq \left\langle a, x^1 \right\rangle \text{ for all } x^2 \in K_2, \text{ for all } x^1 \in K_1.$$

In particular, the function $\langle a, \cdot \rangle$ is bounded from below in K_2 and from above in K_1. From the relationship above, we also get that

$$\gamma_2 := \inf_{x^2 \in K_2} \left\langle a, x^2 \right\rangle \geq \sup_{x^1 \in K_1} \left\langle a, x^1 \right\rangle := \gamma_1.$$

Defining $c := (\gamma_1 + \gamma_2)/2$, we have that $\gamma_2 \geq c \geq \gamma_1$ and

$$\text{for all } x^1 \in K_1, \left\langle a, x^1 \right\rangle \leq \sup_{x^1 \in K_1} \left\langle a, x^1 \right\rangle = \gamma_1 \leq c,$$

$$\text{for all } x^2 \in K_2, \left\langle a, x^2 \right\rangle \leq \inf_{x^2 \in K_2} \left\langle a, x^2 \right\rangle = \gamma_1 \geq c,$$

which completes the proof for this item.

(ii) If we have strict separation but there exists $x \in K_1 \cap K_2$, we quickly get to the contradiction $\langle a, x \rangle < c < \langle a, x \rangle$.

Suppose now that K_2 is bounded and $K_1 \cap K_2 = \emptyset$. Since K_1 is closed and convex, the projection operator $P_{K_1} : \mathbb{R}^n \to K_1$ is well-defined by Fact 2.9; in particular, from Fact 2.27, we have that it is continuous. Therefore, the function

$$\psi : \mathbb{R}^n \to \mathbb{R}_+, \psi(x) = \operatorname{dist}(x, K_1) = \|x - P_{K_1}(x)\|$$

is continuous. Therefore, since K_2 is compact, the problem

$$\min \psi(x) \text{ subject to} x \in K_2$$

has a global solution by the Weierstrass theorem, which we denote by a^2. We denote $a^1 = P_{K_1}(a^2)$. By Fact 2.9,

$$\left\langle a^2 - a^1, x \right\rangle \leq \left\langle a^2 - a^1, a^1 \right\rangle = c_1 \text{ for all } x \in K_1. \tag{2.8}$$

For all $x \in K_2$, we have that

$$\begin{aligned} \|x - a^1\| \geq \operatorname{dist}(x, K_1) &\geq \min_{x \in K_2} \{\operatorname{dist}(x, K_1)\} = \psi(a^2) \\ &= \operatorname{dist}(a^2, D_1) = \|a^2 - P_{K_1}(a^2)\| = \|a^2 - a^1\|, \end{aligned}$$

which implies that $a^2 = P_{K_2}(a^1)$. Using Fact 2.9 again, we get that

$$\left\langle a^2 - a^1, x \right\rangle \geq \left\langle a^2 - a^1, a^2 \right\rangle = c_2 \text{ for all } x \in K_2. \tag{2.9}$$

Define $a := a^2 - a^1 \neq 0$ ($K_1 \cap K_2 = \emptyset$ implies that $a^1 \neq a^2$) and $c := (c_1 + c_2)/2$. Note that $c_2 > c > c_1$ since

$$c_2 - c_1 = \|a^1 - a^2\|^2 > 0.$$

The result now follows from (2.8) and (2.9).

□

2.5 Convex functions

We begin this subsection by reminding the definition of convex functions.

Definition 2.15 (Convex functions)**.** Let $K \in \mathbb{R}^n$ be a convex set. Then $f : K \to \mathbb{R}$ is a *convex function* if for all $\alpha \in (0, 1)$, for all $x, y \in K$ we have that $f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha) f(y)$.

Convex functions have many useful characterizations, some of which we now describe.

Fact 2.16 (Epigraph characterization of convexity)**.** *f is convex if and only if its epigraph*

$$\operatorname{epi}(f) := \{(x, c) \in K \times \mathbb{R} : f(x) \leq c\}$$

is a convex set.

Proof. Assume that f is a convex function. Suppose $(x, c_1), (y, c_2) \in \operatorname{epi}(f)$, then $f(x) \leq c_1, f(y) \leq c_2$. Since f is convex, for any $\alpha \in [0,1]$, $f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)f(y) \leq \alpha c_1 + (1-\alpha)c_2$. This implies that $\alpha(x, c_1) + (1-\alpha)(y, c_2) \in \operatorname{epi}(f)$. Hence, $\operatorname{epi}(f)$ is a convex set.

Assume that $\operatorname{epi}(f)$ is a convex set. Let $x, y \in \mathbb{R}^n$. Since $(x, f(x)), (y, f(y)) \in \operatorname{epi}(f)$, by convexity of the epigraph set we have that, for any $\alpha \in [0,1], (\alpha x + (1-\alpha)y, \alpha(x) + (1-\alpha)f(y)) \in \operatorname{epi}(f)$. By definition, $f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)f(y)$. Therefore, the function f is convex. □

Fact 2.17 (Level set characterization of convexity)**.** *Let $K \subseteq \mathbb{R}^n$ be convex and $f : K \to \mathbb{R}$ convex. then the level set for any $t \in \mathbb{R}$*

$$L_{f,K}(t) = \{x \in K : f(x) \leq t\}$$

is a convex set.

Proof. Let $t \in \mathbb{R}$ and be arbitrary but fixed. If $L_{f,K}(t) = \emptyset$, the result follows trivially. Let $x, y \in L_{f,K}(t)$, thus $f(x) \leq t$ and $f(y) \leq t$. Let $\alpha \in (0,1)$ be arbitrary but fixed. By convexity of K, $\alpha x + (1-\alpha)y \in K$. By convexity of f, $f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)f(y) \leq \alpha t + (1-\alpha)t = t$. Hence, $(\alpha x + (1-\alpha)y) \in L_{f,K}(t)$. □

Fact 2.18 (One-dimensional characterization)**.** *Let $K \subseteq \mathbb{R}^n$ be convex. $f : K \to \mathbb{R}^n$ is convex if and only if its restriction on any line, that is, the function*

$$g(t) := f(x + th)$$

is convex on for any $x \in K$ and $h \in \mathbb{R}$.

Proof. Follows straightforwardly from Definition 2.15. □

Fact 2.19 (First-order characterization of differentiable convex functions)**.** *Assume $f : K \to \mathbb{R}$ is differentiable, then f is convex if and only if K is convex and for any $x, y \in K$*

$$f(y) \geq f(x) + \langle f'(x), x - y \rangle .$$

Proof. Let f be convex. For any $x, y \in K$ and $\alpha \in (0,1]$, defining $d := y - x$, from Definition 2.15 we have

$$f(x + \alpha d) = f(\alpha y + (1-\alpha)x) \leq \alpha f(y) + (1-\alpha)f(x),$$

from which we get

$$f(y) - f(x) \geq \lim_{\alpha \to 0+} \frac{f(x + \alpha d) - f(x)}{\alpha} = \langle f'(x), d \rangle = \langle f'(x), y - x \rangle ,$$

thus obtaining Fact 2.19. To prove the reciprocal, consider $z = (1-\alpha)x+\alpha y$ and observe that

$$f(x) \geq f(z) + \langle f'(z), x - z\rangle$$

and

$$f(y) \geq f(z) + \langle f'(z), y - z\rangle .$$

Multiplying the first equation by $(1-\alpha)$ and the second one by α, then adding both yields

$$(1-\alpha)f(x) + \alpha f(y) \geq f\left((1-\alpha)f(x) + \alpha f(y)\right).$$

□

Fact 2.20 (Second-order characterization of differentiable convex functions)**.** *Assume $f : \Omega \to \mathbb{R}$ is twice-differentiable, with $\Omega \subset \mathbb{R}^n$ an open and convex set. Then f is convex if its Hessian matrix is positive semidefinite for all $x \in \Omega$, that is,*

$$\langle f''(x)d, d\rangle \geq 0 \text{ for all } x \in \Omega, \text{ for all } d \in \mathbb{R}^n.$$

Proof. Due to Fact 2.19, it is sufficient to prove the equivalence of the first and second-order characterization. Fix any $x \in \Omega$ and $d \in \mathbb{R}^n$. Since Ω is open, $x + \alpha d \in \Omega$ for all $\alpha > 0$ sufficiently small. From Fact 2.19,

$$f(x + \alpha d) - f(x) \geq \alpha \langle f'(x), d\rangle .$$

Since f is twice-differentiable,

$$\begin{aligned} 0 &\leq f(x + \alpha d) - f(x) - \alpha \langle f'(x), d\rangle \\ &= \frac{\alpha^2}{2} \langle f''(x)d, d\rangle + o(\alpha^2). \end{aligned}$$

Dividing by $\alpha^2 > 0$ and taking the limit $\alpha \to 0+$, we obtain

$$f(x + \alpha d) - f(x) \geq \alpha \langle f'(x), d\rangle .$$

Now let $x, y \in \Omega$ be arbitrary. From the mean value theorem, there exists $\alpha \in (0, 1)$ such that

$$f(y) - f(x) - \langle f'(x), y - x\rangle = \frac{1}{2} \langle f''(x + \alpha(y - x))(y - x), y - x\rangle \geq 0,$$

where the second inequality comes from Fact 2.20 □

The characterizations above are all useful, in certain contexts, to establish the convexity of a function. Using these characterizations, we can establish certain useful properties of convex sets. Next, we prove that convex functions are continuous in open convex sets.

Fact 2.21 (Continuity of convex functions)**.** *Let $\Omega \in \mathbb{R}^n$ be an open convex set and $f : \Omega \to \mathbb{R}$ a convex function. Then f is locally Lipschitz-continuous in Ω. In particular, f is continuous in Ω.*

Proof. Let $\hat{x} \in \Omega$ be arbitrary. Let U be a box around $\hat{x} \in \mathbb{R}^n$, that is, $U = \{x \in \mathbb{R}^n \mid -\delta \leq x^i - \bar{x}^i \leq \delta, i = 1, \ldots, n\}, \delta > 0$. It can be shown that U is the convex hull of 2^n points whose coordinates are $\hat{x}^i + \delta$ or $\hat{x}^i - \delta$, $i = 1, \ldots, n$ in all possible combinations. Let V be the set of the 2^n vertices of U, such that $U = \operatorname{conv} V$. Define $\beta := \max_{v \in V} f(x)$, which exists since V is finite.

Since Ω is open, there exists $\delta > 0$ (which depends on $\hat{x}$) such that $U \subset \Omega$. From Fact 2.17, we have that $L_{f,\Omega}(\beta)$ is convex. Since $V \in L_{f,\Omega}(\beta)$, we have that

$$U = \operatorname{conv} V \subset \operatorname{conv} L_{f,\Omega}(\beta) = L_{f,\Omega}(\beta). \tag{2.10}$$

Let $x \in U$ such that $0 < \|x - \hat{x}\| < \delta$. Define

$$\alpha := \frac{\|x - \hat{x}\|}{\delta} \in (0, 1), \quad d := \frac{\delta(x - \hat{x})}{\|x - \hat{x}\|} = \frac{(x - \hat{x})}{\alpha}.$$

It can be easily seen that $\hat{x} + d, \hat{x} - d \in U$, since $\|d\| = \delta$, and

$$x = \hat{x} + \alpha d = \alpha(\hat{x} + d) + (1 - \alpha)\hat{x}.$$

By the convexity of f,

$$\begin{aligned} f(x) &\leq \alpha f(\hat{x} + d) + (1 - \alpha) f(\hat{x}) \\ &\leq \alpha\beta + (1 - \alpha) f(\hat{x}), \end{aligned} \tag{2.11}$$

since $\hat{x} + d \in U \subset L_{f,\Omega}(\beta)$ (2.10). Similarly,

$$\hat{x} = x - \alpha d = x + \alpha(\hat{x} - d) - \alpha x,$$

which implies

$$\hat{x} = \frac{1}{1 + \alpha} x + \frac{\alpha}{1 + \alpha}(\hat{x} - d),$$

and by the convexity of f,

$$\begin{aligned} f(\hat{x}) &\leq \frac{1}{1 + \alpha} f(x) + \frac{\alpha}{1 + \alpha} f(\hat{x} - d) \\ &\leq \frac{f(x) + \alpha\beta}{1 + \alpha}, \end{aligned}$$

since $\hat{x} - d \in U \subset L_{f,\Omega}(\beta)$ (2.10). From (2.11), we find that

$$\begin{aligned} -\alpha(\beta - f(\hat{x})) &\leq f(x) - f(\hat{x}) \\ &\leq \alpha(\beta - f(\hat{x})), \end{aligned}$$

where the second inequality comes from (2.11). Therefore,

$$\begin{aligned} |f(x) - f(\hat{x}) &\leq \alpha(\beta - f(\hat{x})) \\ &= \frac{\beta - f(\hat{x})}{\delta}\|x - \hat{x}\| \\ &= L\|x - \hat{x}\|, \end{aligned}$$

where $L = \frac{\beta - f(\hat{x})}{\delta}$ depends on $\hat{x}$ and the choice of δ, but not on $x \in \{x \in \mathbb{R}^n \mid 0 < \|x - \hat{x}\| < \delta\}$.

In particular, this shows that f is continuous at $\hat{x}$. Since $\hat{x} \in \Omega$ was arbitrary, we have that f is continuous in Ω.

Let $\delta > 0$ such that $B(\hat{x}, 2\delta) \in \Omega$. By the continuity of f in $B(\hat{x}, 2\delta)$ and the Weierstrass theorem, there exists two numbers c_1 and c_2 such that

$$c_1 \leq f(x) \leq c_2 \text{ for all } x \in B(\hat{x}, 2\delta). \tag{2.12}$$

Let $x, y \in B(\hat{x}, \delta)$, $x \neq y$,

$$z = y + \frac{\delta}{\|y - x\|}(y - x).$$

Since $\|z - y\| = \delta$ and $\|y - x\| \leq \delta$, by the triangular inequality we have that $z \in B(\hat{x}, 2\delta)$. Furthermore, we have that

$$\begin{aligned} y &= \left(1 + \frac{\delta}{\|y - x\|}\right)^{-1}\left(z + \frac{\delta}{\|y - x\|}\right) \\ &= \frac{\|y - x\|}{\delta + \|y - x\|}z + \frac{\delta}{\delta + \|y - x\|}x. \end{aligned}$$

Denoting

$$\lambda = \frac{\|y - x\|}{\delta + \|y - x\|} \leq \delta^{-1}\|y - x\|, \tag{2.13}$$

by the convexity of f we have that $f(y) \leq (1 - \lambda)f(x) + \lambda f(x)$, where utilizing (2.12) and (2.13) yields

$$\begin{aligned} f(y) - f(x) &\leq \lambda\left(f(z) - f(x)\right) \\ &\leq \delta^{-1}(c_2 - c_1)\|y - x\|. \end{aligned}$$

Changing x and y's role in the analysis above, we obtain

$$f(x) - f(y) \leq \delta^{-1}(c_2 - c_1)\|y - x\|.$$

We conclude that for all $\delta > 0$ sufficiently small, there exists $L > 0$ (which depends on δ) such that

$$|f(y) - f(x)| \leq L\|y - x\| \text{ for all } x, y \in B(\hat{x}, \delta).$$

Thus, f is locally Lipschitz-continuous. $\square$

In particular, if $f : K \to \mathbb{R}$ is convex with respect to the convex set $K \subset \mathbb{R}^n$, then f is continuous in the interior of K.

When working with convex sets, it is sometimes helpful to work with the maximum of convex functions. As further explained in Section 8, when working with convex sets described by convex inequalities, redefining the set in terms of the maximum of convex functions reduces a problem from n dimensions to 1. The following fact, relevant for the aforementioned context, states that the supremum of convex functions is also convex.

Fact 2.22 (Convexity of the supremum of convex functions)**.** *Let $K \subset \mathbb{R}^n$ be a convex set and $f_i : K \to \mathbb{R}$, $i \in I$ (I being a possibly infinite set) convex functions. If I is infinite, also assume that there exists $\beta \in \mathbb{R}$ such that $f_i(x) \leq \beta$ for all $x \in K$ and $i \in I$. Then the function*

$$f : K \to \mathbb{R}, f(x) = \sup_{i \in I} f_i(x)$$

is convex.

Proof. We will show that the epigraph of f is a convex set. Let $c \in \mathbb{R}$ be arbitrary. We have that

$$\begin{aligned}
\operatorname{epi}(f) &= \{(x, c) \in K \times \mathbb{R} \mid f(x) \leq c\} \\
&= \{(x, c) \in K \times \mathbb{R} \mid f_i(x) \leq c \forall i \in I\} \\
&= \cap_{i \in I} \{(x, c) \in K \times \mathbb{R} \mid f_i(x) \leq c\} \\
&= \cap_{i \in I} \operatorname{epi}(f_i).
\end{aligned}$$

By convexity of f_i and Fact 2.16, the epigraphs $\operatorname{epi}(f_i)$, for $i \in I$, are also convex. Thus, their intersection is a convex set by Fact 2.2. Using Fact 2.16 once more, the convexity of $\operatorname{epi}(f)$ implies the convexity of f. □

2.6 Nonexpansive operators

A classical theorem in fixed-point theory, known as Banach's contraction principle (first presented in [4]), establishes that a Lipschitz-continuous operator with constant in $[0, 1)$ (also known as a *contraction mapping*) converges to an unique fixed point. However, such condition is hardly attained in most practical settings. Nonexpansivity, which is equivalent to Lipschitz-continuity with constant 1, is much more easily obtained, though not sufficient to ensure convergence in most cases.

The following notions provide variants of nonexpansivity, both weaker and stronger, which help us establish convergence results for fixed-point algorithms in the absence of contractions.

Definition 2.23 (Notions of nonexpansivity)**.** Let X be a nonempty subset of $\mathbb{R}^n$ and let $T : X \to \mathbb{R}^n$. Then T is

(i) firmly nonexpansive if

$$\forall x \in X, \forall y \in X, \quad \|Tx - Ty\|^2 + \|(\mathrm{Id} - T)x - (\mathrm{Id} - T)y\|^2 \leqslant \|x - y\|^2,$$

or, equivalently,

$$\langle x - y, T(x) - T(y)\rangle \geq \|T(x) - T(y)\|^2, \quad \forall x, y \in X$$

(ii) nonexpansive if

$$\forall x \in X, \forall y \in X, \quad \|Tx - Ty\| \leqslant \|x - y\|;$$

(iii) strictly nonexpansive if

$$\forall x \in X, \forall y \in X, \quad x \neq y \quad \Rightarrow \quad \|Tx - Ty\| < \|x - y\|;$$

(iv) firmly quasinonexpansive if

$$\forall x \in X, \forall y \in \operatorname{Fix} T, \quad \|Tx - y\|^2 + \|Tx - x\|^2 \leqslant \|x - y\|^2;$$

(v) quasinonexpansive if

$$\forall x \in X, \forall y \operatorname{Fix} T, \quad \|Tx - y\| \leqslant \|x - y\|;$$

(vi) and strictly quasinonexpansive if

$$\forall x \in X \setminus \operatorname{Fix} T, \forall y \in \operatorname{Fix} T, \quad \|Tx - y\| < \|x - y\|.$$

Note that (i) implies (ii) and (iv), (ii) implies (v), (iv) implies (vi) and (iii) implies (vi).

Another notion intimately related to nonexpansivity is that of *α-averagedness*. Following its definition, we present some notable results relating these two concepts.

Definition 2.24 (α-averaged operator)**.** Let $X \subseteq \mathbb{R}^n$ be nonempty, let $T : X \to \mathbb{R}^n$ be nonexpansive, and let $\alpha \in (0, 1)$. Then T is *α-averaged* if there exists a nonexpansive operator $R : S \to \mathbb{R}^n$ such that $T = (1 - \alpha)\,\mathrm{Id} + \alpha R$.

Fact 2.25 (Nonxepansivity and α-averagedness relationships)**.** *Let $S \subseteq \mathbb{R}^n$ be nonempty and let $T : X \to \mathbb{R}^n$. The following hold:*

(i) *If T is α-averaged for $\alpha \in (0,1)$, then*

$$\|T(x) - T(y)\|^2 \leq \|x - y\|^2 - \frac{1-\alpha}{\alpha}\|(\mathrm{Id} - T)(x) - (\mathrm{Id} - T)(y)\|^2, \quad \forall x, y \in X$$

(ii) *T is firmly nonexpansive if and only if $2T - \mathrm{Id}$ is nonexpansive.*

(iii) *If T α-averaged with $\alpha \in (0, \frac{1}{2}]$, then T is firmly nonexpansive*

Proof. (i) Since T is α-averaged, there exists a nonexpansive operator $R : S \to \mathbb{R}^n$ such that $T = (1-\alpha)\,\mathrm{Id} + \alpha R$. Thus, we have that

$$\|T(x) - T(y)\|^2 = (1-\alpha)^2\|x-y\|^2 + 2\alpha(1-\alpha)\langle x-y, R(x)-R(y)\rangle + \alpha^2\|R(x)-R(y)\|^2$$

and

$$\begin{aligned}\frac{1}{\alpha}\|(\mathrm{Id} - T)(x) - (\mathrm{Id} - T)(y)\|^2 &= \alpha\|(\mathrm{Id} - R)(x) - (\mathrm{Id} - R)(y)\|^2 \\ &= \alpha\left(\|x-y\|^2 - 2\langle x-y, R(x)-R(y)\rangle + \|R(x)-R(y)\|^2\right)\end{aligned}$$

Adding both equalities yields

$$\|T(x) - T(y)\|^2 + \frac{1-\alpha}{\alpha}\|(\mathrm{Id} - T)(x) - (\mathrm{Id} - T)(y)\|^2 = (1-\alpha)\|x-y\|^2 + \alpha\|R(x)-R(y)\|^2,$$

from which the nonexpansivity of R gives us the desired result.

(ii) Follows from the definitions.

(iii) Follows immediately from (i).

□

One important result of α-averagedness is that the convex combination of α-averaged operators is an α-averaged operator, whose averagedness parameter is the convex combination of the averagedness parameters of the original operators.

Fact 2.26 (Convex combination of α-averaged operators is α-averaged)**.** *Let X be a nonempty subset of $\mathbb{R}^n$, let $(T_i)_{i\in I}$ be a finite family of nonexpansive operators from X to $\mathbb{R}^n$, let $(\omega_i)_{i\in I}$ be real numbers in (0,1) such that $\sum_{i\in I}\omega_i = 1$, and let $(\alpha_i)_{i\in I}$ be real numbers in (0,1) such that, for every $i \in I, T_i$ is a_i-averaged, and set $\alpha = \sum_{i\in I}\omega_i\alpha_i$. Then $\sum_{i\in I}\omega_i T_i$ is α-averaged.*

Proof. For every i $\in I$, there exists a nonexpansive operator $R_i : X \to \mathbb{R}^n$ such that $T_i = (1-\alpha_i)\,\mathrm{Id} + \alpha_i R_i$. Define $R = \sum_{i\in I}\left(\frac{\omega_i\alpha_i}{\alpha}\right)R_i$. This operator R can be shown to be nonexpansive, from which it follows that

$$\sum_{i\in I}\omega_i T_i = \sum_{i\in\mathbb{N}}\omega_i\,(1-\alpha_i)\,\mathrm{Id} + \sum_{i\in I}\omega_i\alpha_i R_i = (1-\alpha)\,\mathrm{Id} + \alpha R,$$

which means that T is α-averaged. □

Now, we establish the firm nonexpansivity of the projection and the nonexpansivity of the reflection, which will be of tremendous importance in establishing convergence of projection and reflection-based algorithms.

Fact 2.27 (Firm nonexpansivity of projection)**.** *Let $X \subseteq \mathbb{R}^n$ be nonempty, closed and convex. Then the projector, P_X, is firmly nonexpansive and the reflector, R_X, is nonexpansive.*

Proof. Let $x, y \in \mathbb{R}^n$. By applying the projection theorem to $P_X(x)$ and $P_X(y)$, respectively, we obtain

$$\langle P_X(y) - P_X(x), x - P_X(x)\rangle \leq 0 \text{ and } \langle P_X(x) - P_X(y), y - P_X(y)\rangle \leq 0.$$

The addition of these inequalities gives

$$\langle P_X(x) - P_X(y), P_X(x) - P_X(y) - (x - y)\rangle \leq 0,$$

which implies that P_X is firmly nonexpansive. The fact that R_X is nonexpansive now follows from Fact 2.25. □

Remind that nonexpansivity implies Lipschitz-continuity. Hence, projections and reflections are continuous.

2.7 Fixed-point nonexpansive algorithm convergence

With the following facts, we show how α-averagedness is sufficient for assuring convergence of fixed-point algorithms in closed and convex subsets. These results are relevant not only for their direct application, but also to help put into perspective, later on, how much less straightforward it is to establish convergence in the absence of stronger notions of nonexpansivity.

First, we present a fact that the convergence of α-averaged operators with a nonempty set of fixed points.

Fact 2.28 (Convergence of α-averaged iterations)**.** *Let K be a nonempty, closed and convex subset of $\mathbb{R}^n$ and let $T : K \to K$ be an α-averaged operator with $\operatorname{Fix} T \neq \emptyset$. Given any $x^0 \in K$, set*

$$x^{k+1} = T\left(x^k\right), \qquad \textit{for } k = 0, 1, 2 \ldots.$$

Then $\left(x^k\right)_{k\in\mathbb{N}}$ converges to a point $x^* \in \operatorname{Fix} T$.

Proof. Let $x \in \operatorname{Fix} T$. since T is α-averaged, by Fact 2.25 we have

$$\left\|x^{k+1}-x\right\|^2+\frac{1-\alpha}{\alpha}\left\|x^k-x^{k+1}\right\|^2 \leq\left\|x^k-x\right\|^2 \quad \text{for all } k=0,1,2,\ldots .$$

It then follows that $\left(\|x^k-x\|\right)_{k=1}^{\infty}$ is non-increasing, $\left(x^k\right)_{k \in \mathbb{N}}$ is bounded and $x^k-x^{k+1} \rightarrow 0$. Now, as a bounded sequence, $\left(x^k\right)_{k \in \mathbb{N}}$ has a cluster point x^*. Let $\left(x^{k_n}\right)_{n \in \mathbb{N}}$ be a subsequence such that $x_{k_n} \rightarrow x^*$.

We therefore have that

$$(\operatorname{Id}-T)\left(x^*\right)=\lim_{n \rightarrow \infty}(\operatorname{Id}-T)\left(x^{k_n}\right)=\lim_{n \rightarrow \infty}\left(x^{k_n}-x^{k_{n+1}}\right)=0,$$

which shows that $x^* \in \operatorname{Fix} T$. Since $(\|x_k-x^*\|)_{k-1}^{\infty}$ is non-increasing and $\left\|x^{k_n}-x^*\right\| \rightarrow 0$, it follows that $\|x_k-x^*\| \rightarrow 0$, which completes the proof. □

The following two facts will help prove the subsequent, more important result.

Fact 2.29. *Suppose $T: \mathbb{R}^n \rightarrow \mathbb{R}^n$ is an α-averaged operator for some $\alpha \in (0,1)$. Let $x^0, y^0 \in \mathbb{R}^n$ and denote $x^k=T^k\left(x^0\right)$ and $y^k=T^k\left(y^0\right)$ for $k=1,2,\ldots$. Then, for all $n=0,1,2,\ldots$,*

$$\frac{1-\alpha}{\alpha} \sum_{k=0}^n\left\|(\operatorname{Id}-T)\left(x^k\right)-(\operatorname{Id}-T)\left(y^k\right)\right\|^2 \leq\left\|x^0-y^0\right\|^2-\left\|x^{n+1}-y^{n+1}\right\|^2 .$$

In particular, $(\operatorname{Id}-T)\left(x^k\right)-(\operatorname{Id}-T)\left(y^k\right) \rightarrow 0$ as $k \rightarrow \infty$.

Proof. Since T is α -averaged, Fact 2.25(i) implies that

$$\left\|x_{k+1}-y_{k+1}\right\|^2+\frac{1-\alpha}{\alpha}\left\|(\operatorname{Id}-T)\left(x_k\right)-(\operatorname{Id}-T)\left(y_k\right)\right\|^2 \leq\left\|x_k-y_k\right\|^2, \text{ for all } k \in \mathbb{N},$$

which we can sum telescopically to obtain

$$\left\|x_{n+1}-y_{n+1}\right\|^2+\frac{1-\alpha}{\alpha} \sum_{k=0}^n\left\|(\operatorname{Id}-T)\left(x_k\right)-(\operatorname{Id}-T)\left(y_k\right)\right\|^2 \leq\left\|x_0-y_0\right\|^2, \text{ for all } n \in \mathbb{N}.$$

The series in the previous equation therefore converges as $n \rightarrow \infty$, from which the result follows. □

Fact 2.30 (Césaro summation)**.** *Let $\left(x^k\right)_{k \in \mathbb{N}} \subseteq \mathbb{R}^n$ be a convergent sequence with limit $x^* \in \mathbb{R}^n$. Then*

$$\frac{1}{k} \sum_{n=1}^k x^n \rightarrow x^* \text{ as } k \rightarrow \infty$$

Proof. Let $\varepsilon > 0$. Since $\left(x^k\right)_{k\in\mathbb{N}}$ converges to x^*, there exists $k_0 \in \mathbb{N}$ such that $\left\|x^k - x^\star\right\| < \frac{\varepsilon}{2}$ for all $k \geq k_0$. Furthermore, there exists $k_1 \in \mathbb{N}$ such that

$$\frac{1}{k}\sum_{i=1}^{k_0-1}\left\|x^i - x^\star\right\| < \frac{\varepsilon}{2} \text{ for all } k \geq k_1.$$

Consequently, for all $k \geq \max\{k_0, k_1\}$, we have

$$\left\|\frac{1}{k}\sum_{i=1}^{k} x^i - x^*\right\| \leq \frac{1}{k}\sum_{i=1}^{k_0-1}\left\|x^i - x^\star\right\| + \frac{1}{k}\sum_{i=k_0}^{k}\left\|x^i - x^\star\right\| < \frac{\varepsilon}{2} + \frac{k-k_0}{k}\frac{\varepsilon}{2} \leq \varepsilon,$$

which proves the result. □

We can now establish the other convergence result of this subsection, which states that an α-averaged operator's set of fixed points is empty if, and only if, its iterations explode towards infinity regardless of the point in which the algorithm was initialized.

Fact 2.31 (Asymptotic behavior of averaged iterations)**.** *Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be an α-averaged operator for some $\alpha \in (0,1)$. Then*

$$\operatorname{Fix} T = \emptyset \quad \Leftrightarrow \quad \left\|T^k(x)\right\| \to \infty \textit{ for any } x \in \mathbb{R}^n.$$

Proof. We will prove the result by proving its contrapositive: $\operatorname{Fix} T \neq \emptyset$ if and only if there exists an $x \in \mathbb{R}^n$ such that $\left\|T^k(x)\right\| \neq \infty$. First, take any $x \in \mathbb{R}^n$ and suppose that $\operatorname{Fix} T \neq \emptyset$. Then the sequence $T^k(x)$ is convergent by Fact 2.28; in particular, it is bounded.

For the reverse implication, suppose there exists an $x \in \mathbb{R}^n$ such that $x^k = T^k(x)$ contains a bounded subsequence. The sequence $\left(x^k\right)$ then possesses a convergent subsequence $x_{k_n} \to x^*$. Since T is α-averaged, applying Fact 2.29 with $x^0 = x$ and $y^0 = x^*$ yields

$$(\operatorname{Id} - T)\left(x_k\right) - (\operatorname{Id} - T)T^k\left(x^*\right) \to 0,$$

which implies

$$(\operatorname{Id} - T)T^{k_n}\left(x^*\right) \to (\operatorname{Id} - T)\left(x^*\right).$$

Also, applying Fact 2.29 with $x_0 = x^*$ and $y_0 = T\left(x^*\right)$ yields

$$\begin{aligned}\frac{1-\alpha}{\alpha}\sum_{j=1}^{k}&\left\|(\mathrm{Id} - T)T^{j-1}\left(x^*\right) - (\operatorname{Id} - T)T^j\left(x^*\right)\right\|^2 \\ &\leq \left\|(\operatorname{Id} - T)\left(x^*\right)\mathrm{I}^2 - \right\|(\operatorname{Id} - T)T^{\mathrm{k}}\left(x^*\right)\|^2.\end{aligned}$$

Taking the limit along the subsequence (k_n) gives

$$\begin{aligned}\frac{1-\alpha}{\alpha}\sum_{j=1}^{\infty}&\|(\mathrm{Id}-T)T^{j-1}(x^*)-(\mathrm{Id}-T)T^{j}(x^*)\|^2\\ &\leq \|(\mathrm{Id}-T)(x^*)\|^2-\|(\mathrm{Id}-T)(x^*)\|^2=0,\end{aligned}$$

which shows that all terms in the summation are identically zero. Consequently, we have

$$(\mathrm{Id}-T)T^j\left(x^*\right)=(\mathrm{Id}-T)\left(x^*\right),$$

which implies

$$(\mathrm{Id}-T)\left(x^k\right)\to(\mathrm{Id}-T)\left(x^*\right).$$

Denoting $x^0:=x$, we observe that

$$x^0=\sum_{n=0}^{k-1}(\mathrm{Id}-T)\left(x^n\right)+x^k,$$

leading to

$$\frac{1}{k}x^0=\frac{1}{k}\sum_{n=0}^{k-1}(\mathrm{Id}-T)\left(x^n\right)+\frac{1}{k}x^k.$$

Taking the limit along the subsequence x_{k_n} and using Cesàro summation (Fact 2.30) gives

$$0\cdot x^0=(\mathrm{Id}-T)\left(x^*\right)+0\cdot x^*,$$

from which we have that

$$x^*=T\left(x^*\right).$$

Therefore, $x^*\in\operatorname{Fix}T$ implies $\operatorname{Fix}T\neq\emptyset$. □

2.8 Fejér monotonicity

Besides nonexpansivity and α-averagedness, another helpful property for establishing convergence of fixed-point algorithm sequences is what is called *Fejér monotonicity*. However, instead of being a property of the operators, this is a property of the sequences themselves, established with respect to particular set.

As stated in the following definition, a sequence is Fejér monotone with respect to a set if for every point in that set, the next iteration of the sequence never gets farther from that point.

Definition 2.32 (Fejér monotonicity)**.** Let X be a nonempty subset of $\mathbb{R}^n$ and let $\left(x^k\right)_{k\in\mathbb{N}}$ be a sequence in $\mathbb{R}^n$. Then $\left(x^k\right)_{k\in\mathbb{N}}$ is *Fejér monotone with respect to* X if

$$(\forall x\in X)(\forall k\in\mathbb{N}),\left\|x^{k+1}-x\right\|\leq\left\|x^k-x\right\|.$$

The following are some basic properties of Fejér monotone sequences.

Fact 2.33 (Fejér monotonicity properties)**.** *Let X be a nonempty subset of $\mathbb{R}^n$ and let $\left(x^k\right)_{k\in\mathbb{N}}$ be a sequence in $\mathbb{R}^n$. Suppose that $\left(x^k\right)_{k\in\mathbb{N}}$ is Fejér monotone with respect to $X \subset \mathbb{R}^n$. Then,*

(i) *$\left(x^k\right)_{k\in\mathbb{N}}$ is bounded;*

(ii) *for every $x \in X$, $\left(\left\|x^k - x\right\|\right)_{k\in\mathbb{N}}$ converges;*

(iii) *$\left(\operatorname{dist}\left(x^k, X\right)\right)_{k\in\mathbb{N}}$ is decreasing and converges.*

Proof. (i) Let $x \in X$. Then Definition 2.32 implies that $(x_n)_{n\in\mathbb{N}}$ lies in $B\left(x; \|x^0 - x\|\right)$.

(ii) Clear from Definition 2.32.

(iii) Taking the infimum in Definition 2.32 over $x \in X$ yields $\operatorname{dist}(x^{n+1}, X) \leq \operatorname{dist}(x^n, X)$ for all $n \in \mathbb{N}$.

□

The theorem below establishes an important convergence result for Fejér monotone sequences.

Fact 2.34. *Let X be a nonempty subset of $\mathbb{R}^n$ and let $\left(x^k\right)_{k\in\mathbb{N}}$ be a sequence in $\mathbb{R}^n$. Suppose that $\left(x^k\right)_{k\in\mathbb{N}}$ is Fejér monotone with respect to X and that a cluster point of $\left(x^k\right)_{k\in\mathbb{N}}$ belongs to X. Then $\left(x^k\right)_{k\in\mathbb{N}}$ converges to a point in X.*

Proof. From Fact 2.33(i), $\left(x^k\right)_{k\in\mathbb{N}}$ is bounded. By assumption, $\left(x^k\right)_{k\in\mathbb{N}}$ has a cluster point $x \in X$; therefore, there exists a subsequence $\left(x^{k_n}\right)_{n\in\mathbb{N}}$ which converges to x. A bounded sequence is convergent if and only if it has an unique cluster point. Therefore, it suffices to show that $x \in X$ is the only cluster point of the sequence.

Suppose $\left(x^k\right)_{k\in\mathbb{N}}$ has another cluster point y; therefore, there exists a subsequence $\left(x^{k_m}\right)_{m\in\mathbb{N}}$ which converges to y. It can be easily seen that $y \notin C$ would contradict either the fact that $x \in X$ is a cluster point or that $\left(x^k\right)_{k\in\mathbb{N}}$ is Fejér monotone.

Suppose then that $y \in C$. By Fact 2.33(ii), $\left(\left\|x^k - x\right\|\right)_{k\in\mathbb{N}}$ and $\left(\left\|x^k - y\right\|\right)_{k\in\mathbb{N}}$ converge. Since $2\left\langle x^k, x - y\right\rangle = \left\|x^k - y\right\|^2 - \left\|x^k - x\right\|^2 + \|x\|^2 - \|y\|^2$ for all $n \in \mathbb{N}$, $\left(\left\langle x^k, x - y\right\rangle\right)_{k\in\mathbb{N}}$ also converges, to a limit point which we will call L. Consequently, the subsequences $\left(\left\langle x^{k_n}, x - y\right\rangle\right)_{n\in\mathbb{N}}$ and $\left(\left\langle x^{k_m}, x - y\right\rangle\right)_{m\in\mathbb{N}}$ must also converge to L. Taking the limits of $n \to \infty$ and $m \to \infty$ on the respective subsequences gives us that $\langle x, x - y\rangle = L$ and $\langle y, x - y\rangle = L$. Therefore, we must have that $x = y$.

Since $\left(x^k\right)_{k\in\mathbb{N}}$ is bounded and has an unique cluster point $x \in X$, it converges to $x \in X$. $\square$

2.9 Non-differentiability of convex functions: directional derivatives, subgradients and subdifferential

This subsection deals with the class of non-differentiable convex functions, which are often present in optimization problems. In order to cope with the loss of differentiability, one could work with *directional derivatives*, which always exist for convex functions, or with a generalization of the derivative, which is called the *subdifferential*.

As par for this work, we will only deal with functions in $\mathbb{R}^n$, following Section 3.4 of [44]. We recommend Chapters 16 and 17 of [6] and Section 23 of [60] for a more general framework on this subject.

We begin by presenting the definition of a directional derivative.

Definition 2.35 (Directional derivative)**.** Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex and let $d \in \mathbb{R}^n$. The directional derivative of f at $x \in \mathbb{R}^n$ in the direction d is

$$f'(x;d) = \lim_{\alpha\to 0+} \frac{f(x+\alpha d) - f(x)}{\alpha},$$

provided that this limit exists in $[-\infty, +\infty]$. The right and left derivatives of f at x are given by

$$f'_+(x) = f'(x;1) = \lim_{\alpha\to 0+} \frac{f(x+\alpha) - f(x)}{\alpha}$$

and

$$f'_-(x) = -f'(x;-1) = \lim_{\alpha\to 0-} \frac{f(x+\alpha) - f(x)}{\alpha},$$

respectively, provided the limits exist.

We follow by presenting properties of the directional derivative of a convex function.

Fact 2.36 (Directional derivative of a convex function)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex. Then, for every $x \in \mathbb{R}^n$, f is differentiable in each direction $d \in \mathbb{R}^n$ and*

$$f(x+\alpha d) \geq f(x) + \alpha f'(x;d) \text{ for all } \alpha > 0.$$

Proof. We'll show that the function $\phi : \mathbb{R}_{++} \to \mathbb{R}$ given by $\phi(\alpha) = \frac{f(x+\alpha y)}{\alpha}$ for $x \in \mathbb{R}^n, y \in \mathbb{R}$ is increasing.

Fix α and β in $\mathbb{R}_{++}$ such that $0 < \alpha < \beta$, and set $\lambda = \alpha/\beta$ and $z = x + \beta y$. By Definition 2.15, $f(x+\alpha y) = f(\lambda z+(1-\lambda)x) \leqslant \lambda f(z)+(1-\lambda)f(x) = f(x)+\lambda(f(z)-f(x))$, hence $\phi(\alpha) \leqslant \phi(\beta)$.

Thus, the function $\Phi(\alpha) := \phi(\alpha) - \phi(0)$ is monotonous and limited (from below, in particular); therefore, the limit $f'(x;d) = \lim_{\alpha \to 0+} \Phi(\alpha)$ exists and $\Phi(\alpha) \geq f'(x,d)$ for all $\alpha > 0$. From the definitions of $\Phi(\alpha), \phi(\alpha)$ and $f'(x,d)$, the previous inequality implies $f(x + \alpha d) \geq f(x) + \alpha f'(x;d)$ for all $\alpha > 0$. □

In the following fact, we state that the directional derivatives of a convex function are upper semicontinuous.

Fact 2.37 (Upper semicontinuity of convex function directional derivatives)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex. Then for whichever two sequences $(x^k) \to x$, $(d^k) \to d$ as $k \to \infty$, we have that*

$$\limsup_{k \to \infty} f'(x^k; d^k) \leq f'(x;d).$$

Proof. From Fact 2.36, we have that for all $\varepsilon > 0$, there exists $\alpha > 0$ such that

$$\frac{f(x + \alpha d) - f(x)}{\alpha} \leq f'(x;d) + \varepsilon.$$

Thus, by the continuity of f (Fact 2.21), for all k sufficiently large, we have that

$$\frac{f(x^k + \alpha d) - f(x^k)}{\alpha} \leq f'(x;d) + \varepsilon.$$

From Fact 2.36 it also follows that

$$f'(x^k; d^k) \leq \frac{f(x^k + \alpha d) - f(x^k)}{\alpha}.$$

Thus,

$$f'(x^k; d^k) \leq f'(x;d) + \varepsilon,$$

which yields

$$\limsup_{k \to \infty} f'(x^k; d^k) \leq f'(x;d) + \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, this implies the desired result. □

As a corollary of this fact, it can also be proved that if f is differentiable in $\mathbb{R}^n$, then the derivative of f is continuous in $\mathbb{R}^n$.

Now, we present the generalization of the derivative for convex functions.

Definition 2.38 (Subgradients and subdifferential of a convex function)**.** Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex. $y \in \mathbb{R}^n$ is a *subgradient* of f at x if $\langle y, z - x \rangle + f(x) \leq f(y)$ for all $z \in \mathbb{R}^n$. The set of all subgradients of f at x is called the *subdifferential* of f at x and is denoted by

$$\partial f(x) = \{y \in \mathbb{R}^n \mid \langle y, z - x \rangle + f(x) \leq f(y) \text{ for all } z \in \mathbb{R}^n\}.$$

If $\partial f(x) \neq \emptyset$, then f is subdifferentiable at x.

By writing $z = x + \alpha d$, where $d \in \mathbb{R}^n$ and $\alpha > 0$, a subgradient of f can be rewritten in terms of a directional derivative of f. The following fact establishes this equivalence.

Fact 2.39 (Subdifferential in terms of directional derivatives)**.** *The following definitions for the subdifferential of a convex function are equivalent:*

(i) $\partial f(x) = \{y \in \mathbb{R}^n \mid \langle y, z - x\rangle + f(x) \leq f(y)$ *for all* $z \in \mathbb{R}^n\}$*;*

(ii) $\partial f(x) = \{y \in \mathbb{R}^n \mid \langle y, d\rangle \leq f'(x; d)$ *for all* $d \in \mathbb{R}^n\}$*.*

Proof. Let $y \in \partial f(x)$ with $\partial f(x)$ defined as in (i), and let $d \in \mathbb{R}^n$ and $\alpha > 0$ such that $z = x + \alpha d$. Then $y \in \partial f(x)$ if and only if $\frac{f(x+\alpha d)-f(x)}{\alpha} \geq \langle y, d\rangle$ for all $d \in \mathbb{R}^n$ and for all $\alpha > 0$. Thus, we must have that $f'(x; d) = \lim_{\alpha \to 0+} \frac{f(x+\alpha d)-f(x)}{\alpha} \geq \langle y, d\rangle$, where the limit exists by Fact 2.36. Therefore, $y \in \partial f(x)$ if and only if $f'(x; d) \geq \langle y, d\rangle$ for all $d \in \mathbb{R}^n$, which means that definitions (i) and (ii) are equivalent. □

Some important properties of the subdifferential of a convex function are presented below.

Fact 2.40 (Subdifferential of a convex function)**.** *Let* $f : \mathbb{R}^n \to \mathbb{R}$ *be convex. Then for every* $x \in \mathbb{R}^n$*, the set* $\partial f(x)$ *is convex, compact and nonempty. Furthermore, for all* $d \in \mathbb{R}^n$*, we have that*

$$f'(x; d) = \max_{y \in \partial f(x)} \langle y, d\rangle \,. \tag{2.14}$$

Proof. By Fact 2.39, $\partial f(x) = \{y \in \mathbb{R}^n \mid \langle y, d\rangle \leq f'(x; d)$ for all $d \in \mathbb{R}^n\} = \cap_{d \in \mathbb{R}^n}\{y \in \mathbb{R}^n \mid \langle y, d\rangle \leq f'(x; d)\}$. Since the intersection of closed halfspaces is convex and closed, we have that $\partial f(x)$ is convex and closed as well. We will now show that it is bounded. If $\partial f(x) \neq \emptyset$, it is trivially bounded. Otherwise, suppose there exists a sequence $(y^k)_{k \in \mathbb{N}} \subset \partial f(x)$ such that $\|y^k\| \to \infty$. For all k, let $d^k = \frac{y^k}{\|y^k\|}$. We can assume, without loss of generality, that $(d^k)_{k \in \mathbb{N}} \to d$. By Fact 2.39(ii), we have that

$$\|y^k\| = \left\langle y^k, d^k \right\rangle \leq f'(x; d^k).$$

Therefore,

$$\limsup_{k \to \infty} \|y^k\| \leq \limsup_{k \to \infty} f'(x; d^k) \leq f'(x; d) < \infty,$$

where the second inequality comes from Fact 2.37. This contradicts the hypothesis that $\|y^k\| \to \infty$ as $k \to \infty$. Hence, $\partial f(x)$ is bounded.

Now, we will prove (2.14) whilst simultaneously verifying that $\partial f(x) \neq \emptyset$. Fix an arbitrary $d \in \mathbb{R}^n$. From the analysis above, we have that

$$f'(x; d) \leq \max_{y \in \partial f(x)} \langle y, d\rangle \,, \tag{2.15}$$

where the maximum exists by Weierstrass since $\partial f(x)$ is compact (we take the maximum as being $-\infty$ if $\partial f(x) = \emptyset$). We define the two following sets:

$$D_1 = \{(z, c) \in \mathbb{R}^n \times \mathbb{R} \mid c > f(z)\},$$

and

$$D_2 = \{(z, c) \in \mathbb{R}^n \times \mathbb{R} \mid c = f(x) + \alpha f'(x; d), z = x + \alpha d, \alpha > 0\}.$$

D_2 is clearly convex; D_1 can also be easily shown to be convex, since it is epigraph of f minus its boundary. If $D_1 \cap D_2 \neq \emptyset$, we would have that

$$f(x + \alpha d) < f(x) + \alpha f'(x; d)$$

for some $\alpha > 0$, contradicting Fact 2.36. By the separation theorem, there exists a certain $(u, \beta) \in (\mathbb{R}^n \times \mathbb{R}) \setminus \{0\}$ such that

$$\langle u, z\rangle + \beta c \leq \langle u, x + \alpha d\rangle + \beta(f(x) + \alpha f'(x; d)) \tag{2.16}$$

for all $z \in \mathbb{R}^n$, for all $\alpha > 0$ and for all $c > f(z)$.

If we had $\beta = 0$, we would have

$$\langle u, z\rangle \leq \langle u, x + \alpha d\rangle \text{ for all } z \in \mathbb{R}^n,$$

which can only happen when $u = 0$. Since $(u, \beta) \neq 0$, we conclude that $\beta \neq 0$. Suppose that $\beta > 0$. Choosing $z = x$ and $\alpha = 0$ in (2.16), we have that $\beta c \leq \beta f(x)$ for all c such that $c > f(x)$, which is a contradiction. We then conclude that $\beta < 0$. Dividing both sides of (2.16), we obtain

$$c + \langle u/\beta, z - x\rangle \geq f(x) + \alpha f'(x; d) + \alpha \langle u/\beta, d\rangle \tag{2.17}$$

for all $z \in \mathbb{R}^n$, for all $\alpha > 0$ and for all $c > f(z)$. Taking the limits $\alpha \to 0+$ and $c \to f(z)+$ we have that

$$f(z) \geq f(x) - \langle u/\beta, z - x\rangle \text{ for all } z \in \mathbb{R}^n,$$

that is, $-u/\beta \in \partial f(x)$ by the definition of subgradient. In particular, we just showed that $\partial f(x) \neq \emptyset$. In (2.17), taking $\alpha = 1$ and $z = x$ and taking limit $c \to f(z)+$, we obtain

$$0 \geq f'(x; d) + \langle u/\beta, d\rangle .$$

We then have that

$$-\langle u\beta, d\rangle \geq f'(x; d), \quad -u/\beta \in \partial f(x),$$

where taking (2.15) into account yields

$$\max_{y \in \partial f(x)} = f'(x; d) = -\langle u/\beta, d\rangle , \quad -u/\beta \in \partial f(x).$$

□

The following fact establishes an important relationship between subdifferentials and derivatives; when a convex function is differentiable on a given point, its only subdifferential at that point is the derivative itself.

Fact 2.41 (Subdifferential of a differentiable convex function)**.** *A convex function $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable on $x \in \mathbb{R}^n$ if, and only if, the set $\partial f(x)$ is single-valued, in which case $\partial f(x) = \{f'(x)\}$.*

Proof. Let f be differentiable at x. From Fact 2.19,

$$f(z) \geq f(x) + \langle f'(x), z - x\rangle \text{ for all } z \in \mathbb{R}^n.$$

Now let $y \in \partial f(x)$. For all $d \in \mathbb{R}^n$,

$$f(x) + \langle y, d\rangle \leq f(x + d) = f(x) + \langle f'(x), d\rangle + o(\|d\|).$$

Thus,

$$\langle y - f'(x), d\rangle \leq o(\|d\|).$$

Taking

$$d^k = \frac{y - f'(x)}{k\|y - f'(x)\|},$$

we have that $d^k \to 0$ as $k \to 0$ and

$$\frac{\|y - f'(x)}{k} \leq o(\frac{1}{k}),$$

which is only possible when $y - f'(x) = 0$. Therefore, we conclude that $\partial f(x) = \{f'(x))\}$. Let $\partial f(x) = \{y\}$. By Fact 2.40 ((2.14), in particular), we have that

$$f'(x; d) = \langle y, d\rangle \text{ for all } d \in \mathbb{R}^n.$$

Choosing d as the elements of the canonical basis of $\mathbb{R}^n$, it is easy to see that $y^i = \partial f(x)/\partial x^i$, $i = 1, \ldots, n$. We have that $f'(x; d)$ is a linear function in d of the form $\langle f'(x), d\rangle$, which implies the differentiability of f at x. □

Given the previously stated importance of working with the maximum of convex functions, we present properties of the directional derivative and subdifferential in this context. Note that this time we require compacity for the convex set in which the functions are maximized.

Fact 2.42 (Directional derivative and subdifferential of the maximum of convex functions)**.** *Let Z be a compact set. Suppose $\psi : \mathbb{R}^n \times Z \to \mathbb{R}$ be a continuous function such that $\psi(\cdot, z) : \mathbb{R}^n \to \mathbb{R}$ be convex for all $z \in Z$. Let $f : \mathbb{R}^n \to \mathbb{R}$ be given by $f(x) = \max_{z \in Z} \psi(x, z)$. Then:*

(i) *The function f is convex in $\mathbb{R}^n$ and for all $x \in \mathbb{R}^n$, we have that*

$$f'(x;d) = \max_{\bar{z}\in Z(x)} \psi'(x,\bar{z};d) \text{ for all } d \in \mathbb{R}^n$$

where

$$Z(x) = \{\bar{z} \in Z \mid f(x) = \psi(x,\bar{z}) = \max_{z\in Z} \psi(x,z)\}.$$

In particular, if $Z(x)=\{\bar{z}\}$ and $\psi(\cdot,z)$ is differentiable in $x \in \mathbb{R}^n$, then f is differentiable in x and $f'(x) = \psi'_x(x,\bar{z})$.

(ii) *If $\psi(\cdot,z)$ is differentiable for all $z \in Z$ and $\psi'_x(x,\cdot)$ is continuous in Z for all $x \in \mathbb{R}^n$, then*

$$\partial f(x) = \operatorname{conv}\{\psi'_x(x,\bar{z}) \mid \bar{z} \in Z(x)\}, x \in \mathbb{R}^n.$$

Proof. (i) By the Weierstrass theorem, $Z(x) \neq \emptyset$ and $f(x)$ is finite for all $x \in \mathbb{R}^n$. Since Z is compact, from Fact 2.22, f is convex.

By the definition of f, for all $z \in Z(x)$ we have that $f(x) = \psi(x,\bar{z})$ and $f(x+\alpha d) \geq \psi(x+\alpha d,\bar{z})$. Therefore,

$$\frac{f(x+\alpha d) - f(x)}{\alpha} \geq \frac{\psi(x+\alpha d,\bar{z}) - \psi(x,\bar{z})}{\alpha} \text{ for all } d \in \mathbb{R}^n, \text{ for all } \alpha > 0.$$

Letting $\alpha \to 0+$, we conclude that

$$f'(x;d) \geq \phi'(x,\bar{z};d) \text{ for all } \bar{z} \in Z(x), \text{ for all } d \in \mathbb{R}^n.$$

Thus,

$$f'(x;d) \geq \sup_{\bar{z}\in Z(x)} \psi'(x,\bar{z};d) \text{ for all } d \in \mathbb{R}^n. \tag{2.18}$$

Let $d \in \mathbb{R}^n$ be arbitrary and $(\alpha^k)_{k\in\mathbb{N}} \to 0+$. For all k, define $x^k := x + \alpha^k d$ and $\bar{z}^k \in Z(x^k)$. Since $(\bar{z}^k) \subset Z$ and Z is compact, we can assume (taking a subsequence, if necessary) that $(\bar{z}^k) \to \bar{z} \in Z$. We have that

$$\psi(x^k,\bar{z}^k) = \max_{z\in Z} \psi(x^k,z) \geq \psi(x^k,z) \text{ for all } z \in Z.$$

Taking the limit $k \to \infty$, by the continuity of ψ we find that

$$\psi(x,\bar{z}) \geq \psi(x,z) \text{ for all } z \in Z,$$

which implies that $\bar{z} \in Z(x)$. By Fact 2.36,

$$\begin{aligned} f'(x;d) &\leq \frac{f(x+\alpha^k d) - f(x)}{\alpha^k} \\ &= \frac{\psi(x+\alpha^k d,\bar{z}^k) - \psi(x,\bar{z})}{\alpha^k} \\ &\leq \frac{\psi(x+\alpha^k d,\bar{z}^k) - \psi(x,\bar{z}^k)}{\alpha^k}. \end{aligned}$$

Taking the limit when $k \to \infty$, we get that

$$f'(x;d) \leq \psi'(x,\bar{z};d), \tag{2.19}$$

where $\bar{z} \in Z(x)$. By combining (2.18) with (2.19), we obtain equality in (2.18), which proves the first part of the statement of (i).

Moreover, if we have that $Z(x) = \{\bar{z}\}$ and $\psi(\cdot,\bar{z})$ is differentiable at x, it follows that

$$f'(x;d) = \psi'(x,\bar{z};d) = \langle \psi'_x(x,\bar{z}), d\rangle \text{ for all } d \in \mathbb{R}^n,$$

which implies that f is differentiable at x and $f'(x) = \psi'_x(x,\bar{z})$.

(ii) For all $u \in \mathbb{R}^n$ and all $\bar{z} \in Z(x)$, we have

$$\begin{aligned} f(u) &= \max_{z\in Z} \psi(u,z) \geq \psi(u,\bar{z}) \\ &\geq \psi(x,\bar{z}) + \langle \psi'_x(x,\bar{z}), u-x\rangle \\ &= f(x) + \langle \psi'_x(x,\bar{z}), u-x\rangle\,, \end{aligned}$$

where the second inequality comes from the convexity of $\psi(\cdot,z)$ and Fact 2.19. The relationship above shows that

$$\psi'_x(x,\bar{z}) \in \partial f(x) \text{ for all } \bar{z} \in Z(x).$$

Since $\partial f(x)$ is convex by Fact 2.40, we then have that

$$\operatorname{conv}\{\psi'_x(x,\bar{z}) \mid \bar{z} \in Z(x)\} \subset \operatorname{conv} \partial f(x) = \partial f(x).$$

We observe now that

$$Z(x) = Z \cap \{z \mid \psi(x,z) = f(x)\},$$

where the first set of the intersection is compact and the second one is closed (by continuity of ψ) for all fixed $x \in \mathbb{R}^n$. Thus, $Z(x)$ is compact for all $x \in \mathbb{R}^n$. Since $\psi'_x(\cdot,\bar{z})$ is continuous, we have that $\{\psi'_x(x,\bar{z}) \mid \bar{z} \in Z(x)\}$ is compact. By Fact 2.7, we conclude that $\operatorname{conv}\{\psi'_x(x,\bar{z}) \mid \bar{z} \in Z(x)\}$ is compact. Suppose that there exists $y \in \partial f(x) \setminus \operatorname{conv}\{\psi'_x(x,\bar{z}) \mid \bar{z} \in Z(x)\}$. By the strict hyperplane separation theorem (Fact 2.14(ii)), there exist $a \in \mathbb{R}^n \setminus 0$ and $c \in \mathbb{R}$ such that

$$\langle a, y\rangle > c > \langle a, \psi'_x(x,\bar{z})\rangle \text{ for all } \bar{z} \in Z(x).$$

However, this implies

$$\langle y, a\rangle > \max_{\bar{z}\in Z(x)} \langle \psi'_x(x,\bar{z}), a\rangle = \max_{\bar{z}\in Z(x)} \psi'(x,\bar{z};a) = f'(x;a),$$

where the first inequality follows from the differentiability of $\psi(\cdot, \bar{z})$ at x and last one was proved in item (i). Since $y \in \partial f(x)$, we have found a contradiction with Fact 2.40. Therefore,

$$\partial f(x) \subset \operatorname{conv}\{\psi'_x(x, \bar{z}) \mid \bar{z} \in Z(x)\},$$

which concludes the proof.

□

2.10 Best approximation mappings

The concept of *best approximation mappings* was first introduced by Behling, Santos and Cruz in [20], and its inclusion in this work is to support the theory presented in Section 3.4.2, which was introduced in the same paper. Thus, these facts are presented here, for the sake of self-containment, as they were in the original paper. While we will only refer to this paper for the following facts, we note that the concept has been further explored in [13].

We begin by presenting the definition of this Friedrichs angle. Other than its relevance to best approximation mappings, it was shown in [5] that the cosine of this angle is the rate of linear converge for the Douglas-Rachford method (DRM) Section 3.3 and the lower bound for the linear convergence rate of the Circumcentered-Reflection method (CRM) Section 3.4 in [18] in an affine setting.

Definition 2.43 (Friedrichs angle)**.** The cosine of the Friedrichs angle between affine subspaces V and W with nonempty intersection is given by

$$c_F := \sup\left\{\langle v, w\rangle \mid v \in \hat{V} \cap (\hat{V} \cap \hat{W})^{\perp}, w \in \hat{W} \cap (\hat{V} \cap \hat{W})^{\perp}, \|v\| \leq 1, \|w\| \leq 1\right\}.$$

Now, we present the definition of best approximation mapping.

Definition 2.44 (Best approximation mapping)**.** Let $V \neq \varnothing$ be an affine subspace in $\mathbb{R}^n$. We say that $G_V : \mathbb{R}^n \to \mathbb{R}^n$ is a *best approximation mapping* with respect to V or V-BAM, for short if

(i) $P_V(G_V(z)) = P_V(z)$, for all $z \in \mathbb{R}^n$; and

(ii) there exists a constant $r_V \in [0, 1)$ such that $\|G_V(z) - P_V(z)\| \leq r_V \|z - P_V(z)\|$, for all $z \in \mathbb{R}^n$.

Note that the projection operator P_V is a V-BAM. Indeed, if $G_V = P_V$, for any $z \in \mathbb{R}^n$ and all $r_V \geq 0$, we have $P_V(G_V(z)) = P_V(P_V(z)) = P_V(z)$ and $\|G_V(z) - P_V(z)\| = \|P_V(z) - P_V(z)\| = 0 \leq r_V \|z - P_V(z)\|$. In general, it is easy to see that $G_V = (1 - \alpha)\operatorname{Id} + \alpha P_V$ with $0 < \alpha < 2$ is a V-BAM with $r_V = |1 - \alpha| \in [0, 1)$. Nonetheless, Definition 2.44 allows for non-affine mappings.

The following fact states that a best approximation mapping with respect to a set converges linearly to a projection onto that set.

Fact 2.45 (V-BAM convergence)**.** *Let G_V be a V-BAM with constant $r_V \in [0,1)$. For any $z \in \mathbb{R}^n$ and $\ell \in \mathbb{N}$, $P_V\left(G_V^\ell(z)\right) = P_V(z)$ and $\left(G_V^k(z)\right)_{k\in\mathbb{N}}$ converges to $P_V(z)$ with linear rate r_V.*

Proof. Let $z \in \mathbb{R}^n$ and $\ell \in \mathbb{N}$. Then, Definition 2.44(i) implies

$$P_V\left(G_V^\ell(z)\right) = P_V\left(G_V\left(G_V^{\ell-1}(z)\right)\right) = P_V\left(G_V^{\ell-1}(z)\right) = \cdots = P_V\left(G_V(z)\right) = P_V(z).$$

Moreover,

$$\begin{aligned}\left\|G_V^k(z) - P_V(z)\right\| &= \left\|G_V\left(G_V^{k-1}(z)\right) - P_V\left(G_V^{k-1}(z)\right)\right\| \\ &\le r_V\left\|G_V^{k-1}(z) - P_V\left(G_V^{k-1}(z)\right)\right\| \le \cdots \le r_V^{k-1}\left\|G_V(z) - P_V\left(G_V(z)\right)\right\| \\ &= r_V^{k-1}\left\|G_V(z) - P_V(z)\right\| \\ &\le r_V^k\left\|z - P_V(z)\right\|,\end{aligned}$$

proving the proposition. □

The following technical fact is an auxiliary result regarding adjacent angles.

Fact 2.46. *Let $u, v \in \mathbb{R}^n$ be nonzero vectors forming an angle $\gamma \in \left[0, \frac{\pi}{2}\right]$. If a nonzero vector $w \in \mathbb{R}^n$ forms an angle $\beta \in [0,\pi]$ with v and the angle ϕ between w and u is such that $\phi \in \left[0, \frac{\beta}{2}\right]$, then $\gamma \ge \frac{\beta}{2}$.*

Proof. See the proof for [20, Proposition 2]. □

The next fact is the most important of this subsection, as it states that the composition of two best approximation mappings, with respect to two disjoint affine subspaces, is a best approximating mapping with respect to the intersection.

Fact 2.47 (Composition of two best approximation mappings)**.** *Let us consider two affine subspaces V and W of $\mathbb{R}^n$ with nonempty intersection $V \cap W$. Then, the composition of a V-BAM and a W-BAM is a $(V \cap W)$-BAM.*

Proof. Let $G_V : \mathbb{R}^n \to \mathbb{R}^n$ and $G_W : \mathbb{R}^n \to \mathbb{R}^n$ be a V-BAM and W-BAM with constants $0 \le r_V < 1$ and $0 \le r_W < 1$, respectively.

In order to prove Definition 2.44(i) for the composition $G := G_W \circ G_V$ with regards to $V \cap W$, we are going to combine Pythagoras equations with properties of projections. Note that we have to prove that $P_{V\cap W}(G(z)) = P_{V\cap W}(z)$ for all $z \in \mathbb{R}^n$. Let us take an arbitrary, but fixed, $z \in \mathbb{R}^n$ and set $\hat{z} := P_{V\cap W}(z)$. The definition of $\hat{z}$ implies that $\hat{z} \in V \cap W$. In particular, $\hat{z} \in V$ and we have

$$\|z - \hat{z}\|^2 = \|z - P_V(z)\|^2 + \|P_V(z) - \hat{z}\|^2. \tag{2.20}$$

Since $P_{V\cap W}\left(P_V(z)\right) \in V$, we can write

$$\left\|z - P_{V\cap W}\left(P_V(z)\right)\right\|^2 = \left\|z - P_V(z)\right\|^2 + \left\|P_V(z) - P_{V\cap W}\left(P_V(z)\right)\right\|^2 .$$

Using the fact that $\|z - \hat{z}\| \leq \|z - P_{V\cap W}\left(P_V(z)\right)\|$ and subtracting the second from the first equation yields $\|P_V(z) - \hat{z}\| \leq \|P_V(z) - P_{V\cap W}\left(P_V(z)\right)\|$. By uniqueness of projections onto closed convex sets, we conclude that

$$\hat{z} = P_{V\cap W}\left(P_V(z)\right).$$

The fact that both $\hat{z}$ and $P_{V\cap W}\left(G_V(z)\right)$ lie in V allows us to derive further Pythagoras relations

$$\left\|G_V(z) - \hat{z}\right\|^2 = \left\|G_V(z) - P_V\left(G_V(z)\right)\right\|^2 + \left\|P_V\left(G_V(z)\right) - \hat{z}\right\|^2$$

and

$$\left\|G_V(z) - P_{V\cap W}\left(G_V(z)\right)\right\|^2 = \left\|G_V(z) - P_V\left(G_V(z)\right)\right\|^2 + \left\|P_V\left(G_V(z)\right) - P_{V\cap W}\left(G_V(z)\right)\right\|^2 .$$

Since G_V is a V-BAM, it holds that $P_V\left(G_V(z)\right) = P_V(z)$ and the previous equations reduce to

$$\left\|G_V(z) - \hat{z}\right\|^2 = \left\|G_V(z) - P_V(z)\right\|^2 + \left\|P_V(z) - \hat{z}\right\|^2 \tag{2.21}$$

and

$$\left\|G_V(z) - P_{V\cap W}\left(G_V(z)\right)\right\|^2 = \left\|G_V(z) - P_V(z)\right\|^2 + \left\|P_V(z) - P_{V\cap W}\left(G_V(z)\right)\right\|^2 . \tag{2.22}$$

The fact that $\hat{z} = P_{V\cap W}\left(P_V(z)\right)$ implies that $\|P_V(z) - \hat{z}\| \leq \|P_V(z) - P_{V\cap W}\left(G_V(z)\right)\|$. This inequality, (2.21) and **??** give us $\|G_V(z) - \hat{z}\| \leq \|G_V(z) - P_{V\cap W}\left(G_V(z)\right)\|$. Then, as $P_{V\cap W}\left(G_V(z)\right)$ is uniquely defined and $\hat{z} \in V \cap W$, we must have $P_{V\cap W}\left(G_V(z)\right) = \hat{z}$. We continue with similar arguments, now regarding W. By Pythagoras we get

$$\left\|G_V(z) - P_{V\cap W}\left(G_V(z)\right)\right\|^2 = \left\|G_V(z) - P_W\left(G_V(z)\right)\right\|^2 + \left\|P_W\left(G_V(z)\right) - P_{V\cap W}\left(G_V(z)\right)\right\|^2$$

and

$$\begin{aligned}\left\|G_V(z) - P_{V\cap W}\left(G_W\left(G_V(z)\right)\right)\right\|^2 = &\left\|G_V(z) - P_W\left(G_V(z)\right)\right\|^2 \\ &+ \left\|P_W\left(G_V(z)\right) - P_{V\cap W}\left(G_W\left(G_V(z)\right)\right)\right\|^2 .\end{aligned}$$

From the facts that $P_{V\cap W}\left(G_V(z)\right) = \hat{z}$, $P_W\left(G_W\left(G_V(z)\right)\right) = P_W\left(G_V(z)\right)$ (since G_W is a W-BAM) and $G(z) = G_W\left(G_V(z)\right)$ we can rewrite the equations above as

$$\left\|G_V(z) - \hat{z}\right\|^2 = \left\|G_V(z) - P_W(G(z))\right\|^2 + \left\|P_W(G(z)) - \hat{z}\right\|^2 \tag{2.23}$$

and

$$\|G_V(z) - P_{V\cap W}(G(z))\|^2 = \|G_V(z) - P_W(G(z))\|^2 + \|P_W(G(z)) - P_{V\cap W}(G(z))\|^2 . \tag{2.24}$$

From the definition of projection, it follows that $\|G_V(z) - \hat{z}\| \leq \|G_V(z) - P_{V\cap W}(G_V(z))\|$; this, combined with (2.23) and (2.24), leads to $\|P_W(G(z)) - \hat{z}\| \leq \|P_W(G(z)) - P_{V\cap W}(G(z))\|$. We can derive two additional Pythagoras relations

$$\|G(z) - \hat{z}\|^2 = \|G(z) - P_W(G(z))\|^2 + \|P_W(G(z)) - \hat{z}\|^2 \tag{2.25}$$

and

$$\|G(z) - P_{V\cap W}(G(z))\|^2 = \|G(z) - P_W(G(z))\|^2 + \|P_W(G(z)) - P_{V\cap W}(G(z))\|^2 .$$

The fact that $\|P_W(G(z)) - \hat{z}\| \leq \|P_W(G(z)) - P_{V\cap W}(G(z))\|$, together with the previous two equations, yields $\|G(z) - \hat{z}\| \leq \|G(z) - P_{V\cap W}(G(z))\|$. Hence, $P_{V\cap W}(G(z)) = \hat{z} = P_{V\cap W}(z)$ which fulfills condition (i) of Definition 2.44 for G with regards to $V \cap W$.

Let us now address item (ii). We have to prove that there exists a nonnegative constant $0 \leq r_V \cap W < 1$ so that, for all $z \in \mathbb{R}^n$,

$$\|G(z) - P_{V\cap W}(z)\| \leq r_{V\cap W} \|z - P_{V\cap W}(z)\| . \tag{2.26}$$

Again, let $z \in \mathbb{R}^n$ be arbitrary, fixed and $\hat{z} = P_{V\cap W}(z)$. If $z = \hat{z}$, (2.26) is fulfilled for any nonnegative constant as $G(z) = P_{V\cap W}(z)$. In fact, G_V being a V-BAM, together with $z = \hat{z}$, gives us

$$\|G_V(z) - \hat{z}\| = \|G_V(z) - P_V(z)\| \leq r_V \|z - P_V(z)\| = r_V\|z - \hat{z}\| = 0.$$

Thus, $G_V(z) = \hat{z}$. On the other hand, G_W is a W-BAM, so

$$\begin{aligned}\|G(z) - z\| &= \|G(z) - \hat{z}\| = \|G_W(G_V(z)) - \hat{z}\| = \|G_W(\hat{z}) - \hat{z}\| = \|G_W(\hat{z}) - P_W(\hat{z})\| \\ &\leq r_W \|\hat{z} - P_W(\hat{z})\| = 0.\end{aligned}$$

This means that, if $z = \hat{z}, G(z) = z$ and the left-hand side of (2.26) is equal to zero and this inequality holds for any nonnegative constant $r_{V\cap W}$.

Therefore, from now on, assume $z \neq \hat{z}$. We will construct r_{VnW} upon the constants $r_V \in [0, 1), r_W \in [0, 1)$, and $c_F \in [0, 1)$, the latter the cosine of the Friedrichs angle $\theta_F \in \left(0, \frac{\pi}{2}\right]$ between V and W.

It will be key to look at the angle α between vectors $z - \hat{z}$ and $P_V(z) - \hat{z}$. Note first that $\alpha \in \left[0, \frac{\pi}{2}\right]$, since from (2.20), $\angle[z, P_V(z), \hat{z}] = 90°$. Also, of course,

$$\cos\alpha = \frac{\|P_V(z) - \hat{z}\|}{\|z - \hat{z}\|} \tag{2.27}$$

Moreover, by using (2.21), the V-BAM hypothesis $\|G_V(z) - P_V(z)\| \leq r_V \|z - P_V(z)\|$, (2.20) and that $\cos\alpha \leq 1$, we conclude that

$$\begin{aligned}
\|G_V(z) - \hat{z}\|^2 &= \|G_V(z) - P_V(z)\|^2 + \|P_V(z) - \hat{z}\|^2 \\
&\leq r_V^2 \|z - P_V(z)\|^2 + \|P_V(z) - \hat{z}\|^2 \\
&= r_V^2 \left(\|z - \hat{z}\|^2 - \|P_V(z) - \hat{z}\|^2\right) + \|P_V(z) - \hat{z}\|^2 \\
&= r_V^2 \|z - \hat{z}\|^2 + \left(1 - r_V^2\right) \|P_V(z) - \hat{z}\|^2 \\
&= r_V^2 \|z - \hat{z}\|^2 + \left(1 - r_V^2\right) \cos^2\alpha \|z - \hat{z}\|^2 \qquad (2.28) \\
&\leq \|z - \hat{z}\|^2 \qquad (2.29)
\end{aligned}$$

Now, we split our analysis in two cases.

Case 1: $\alpha \in \left[\frac{\theta_F}{2}, \frac{\pi}{2}\right]$.

In this case, $\cos\alpha \leq \cos\frac{\theta_F}{2}$. This, combined with (2.28) and the fact that $\cos\frac{\theta_F}{2} = \sqrt{\frac{1+c_F^2}{2}}$ provides

$$\begin{aligned}
\|G_V(z) - \hat{z}\|^2 &\leq r_V^2 \|z - \hat{z}\|^2 + \left(1 - r_V^2\right) \cos^2\alpha \|z - \hat{z}\|^2 \\
&\leq r_V^2 \|z - \hat{z}\|^2 + \left(1 - r_V^2\right) \cos^2\frac{\theta_F}{2} \|z - \hat{z}\|^2 \\
&= \left(r_V^2 + \left(1 - r_V^2\right) \frac{1 + c_F^2}{2}\right) \|z - \hat{z}\|^2. \qquad (2.30)
\end{aligned}$$

Since $r_V^2 < 1$ and $\frac{1+c_F^2}{2} < 1$, we have $(1 - r_V^2)\frac{1+c_F^2}{2} < (1 - r_V^2)$. Then,

$$r_V^2 + \left(1 - r_V^2\right) \frac{1 + c_F^2}{2} < r_V^2 + \left(1 - r_V^2\right) = 1.$$

Since G_W is a W-BAM, we have

$$P_W\left(G_W\left(G_V(z)\right)\right) = P_W\left(G_V(z)\right)$$

and

$$\|G_W\left(G_V(z)\right) - P_W\left(G_V(z)\right)\| \leq r_W \|G_V(z) - P_W\left(G_V(z)\right)\|, \quad r_W \in [0, 1).$$

So, we can write $\|G(z) - P_W(G(z))\| \leq \|G_V(z) - P_W(G(z))\|$, which combined with (2.23) and (2.25) gives us

$$\|G(z) - \hat{z}\| \leq \|G_V(z) - \hat{z}\|.$$

Hence, this inequality and (2.30) imply that

$$\|G(z) - \hat{z}\| \leq r_1 \|z - \hat{z}\|,$$

with $r_1 \in [0,1)$, given by

$$r_1 := \sqrt{r_V^2 + (1 - r_V^2)\,\frac{1 + c_F^2}{2}}. \tag{2.31}$$

Case 2: $\alpha \in \left[0, \frac{\theta_F}{2}\right)$

In this case we initially consider the triangle $\triangle[G_V(z), \hat{z}, P_W\left(G_V(z)\right)]$. Since G_V is a V-BAM, $P_V\left(G_V(z)\right) = P_V(z)$. We will be particularly interested in the angle ϕ between $G_V(z) - \hat{z}$ and $P_V(z) - \hat{z}$, when these vectors are nonzero. The vector $P_V(z) - \hat{z}$ is automatically nonzero, because of $\alpha < \frac{\pi}{2}$ and (2.27). If the vector $G_V(z) - \hat{z}$ is zero, we get the desired result as shown below.

Suppose $G_V(z) = \hat{z}$, then $G(z) = G_W\left(G_V(z)\right) = G_W(\hat{z})$ and it is easy to verify that $G_W(\hat{z}) = \hat{z}$. Indeed,

$$\|G_W(\hat{z}) - \hat{z}\| = \|G_W(\hat{z}) - P_W(\hat{z})\| \leq r_W\,\|\hat{z} - P_W(\hat{z})\| = \|\hat{z} - \hat{z}\| = 0.$$

So, $G(z) = \hat{z}$ and the left-hand side of (2.26) is equal to zero and this inequality is fulfilled for any nonnegative constant $r_{V\cap W}$.

Assume for the rest of the proof that $G_V(z) \neq \hat{z}$. Thus,

$$\cos\phi = \frac{\|P_V(z) - \hat{z}\|}{\|G_V(z) - \hat{z}\|} \geq \frac{\|P_V(z) - \hat{z}\|}{\|z - \hat{z}\|} = \cos\alpha,$$

where the inequality is due to (2.30). Therefore, $0 \leq \phi \leq \alpha$ and, consequently, $\phi \in \left[0, \frac{\theta_F}{2}\right)$.

We consider now the triangle $\triangle[P_V(z), \hat{z}, P_W\left(G_V(z)\right)]$. If the vertexes $\hat{z}$ and $P_W\left(G_V(z)\right)$ coincide, we get the following bound:

$$\begin{aligned}
\|G(z) - \hat{z}\| &= \|G_W\left(G_V(z)\right) - \hat{z}\| \\
&= \|G_W\left(G_V(z)\right) - P_W\left(G_V(z)\right)\| \\
&\leq r_W\,\|G_V(z) - P_W\left(G_V(z)\right)\| \\
&\leq r_W\,\|G_V(z) - \hat{z}\| \\
&\leq r_W\|z - \hat{z}\|
\end{aligned} \tag{2.32}$$

where we used, respectively, the definition of G, the current assumption $P_W\left(G_V(z)\right) = \hat{z}$, the hypothesis that G_W is a W-BAM, the fact that $\hat{z} \in W$ and (2.30).

For the rest of the proof, assume also that $P_W\left(G_V(z)\right) \neq \hat{z}$ and define β, the angle between the nonzero vectors $P_V(z) - \hat{z}$ and $P_W\left(G_V(z)\right) - \hat{z}$. It is easy to see that the former belongs to $\hat{V}$ and the latter belongs to $\hat{W}$, where $\hat{V}$ and $\hat{W}$ are the subspaces given by $V - \hat{z}$ and $W - \hat{z}$, respectively. Also, recall that $P_{V\cap W}\left(P_V(z)\right) = \hat{z}$ and therefore $P_V(z) - \hat{z} \in (\hat{V} \cap \hat{W})^{\perp}$.

We rewrite (2.23) using the W-BAM property $P_W(G(z)) = P_W\left(G_W\left(G_V(z)\right)\right) = P_W\left(G_V(z)\right)$ as

$$\|G_V(z) - \hat{z}\|^2 = \|G_V(z) - P_W\left(G_V(z)\right)\|^2 + \|P_W\left(G_V(z)\right) - \hat{z}\|^2 \tag{2.33}$$

and Pythagoras can be employed as

$$\begin{aligned}\|G_V(z) - P_{V\cap W}\left(P_W\left(G_V(z)\right)\right)\|^2 = &\|G_V(z) - P_W\left(G_V(z)\right)\|^2 \\ &+ \|P_W\left(G_V(z)\right) - P_{V\cap W}\left(P_W\left(G_V(z)\right)\right)\|^2 .\end{aligned}$$

On the one hand, $\|P_W\left(G_V(z)\right) - P_{V\cap W}\left(P_W\left(G_V(z)\right)\right)\| \leq \|P_W\left(G_V(z)\right) - \hat{z}\|$. On the other hand, $\|G_V(z) - \hat{z}\| \leq \|G_V(z) - P_{V\cap W}\left(P_W\left(G_V(z)\right)\right)\|$ because we have already seen that $P_{V\cap W}\left(G_V(z)\right) = \hat{z}$. Hence, $P_{V\cap W}\left(P_W\left(G_V(z)\right)\right) = \hat{z}$ and $P_W\left(G_V(z)\right) - \hat{z} \in (\hat{V} \cap \hat{W})^{\perp}$.

We can then use the definition of the cosine of the Friedrichs angle Definition 2.43 θ_F between V and W and get

$$\cos\beta = \left\langle \frac{P_W\left(G_V(z)\right) - \hat{z}}{\|P_W\left(G_V(z)\right) - \hat{z}\|}, \frac{P_V(z) - \hat{z}}{\|P_V(z) - \hat{z}\|} \right\rangle \leq c_F = \cos\theta_F$$

which provides $\beta \in [\theta_F, \pi]$.

By now we have the nonzero vectors $G_V(z) - \hat{z}, P_V(z) - \hat{z}$ and $P_W\left(G_V(z)\right) - \hat{z}$. The vectors $G_V(z) - \hat{z}$ and $P_V(z) - \hat{z}$ form angle $\phi \in \left[0, \frac{\theta_F}{2}\right)$, vectors $P_V(z) - \hat{z}$ and $P_W\left(G_V(z)\right) - \hat{z}$ form angle $\beta \in [\theta_F, \pi]$. Let γ be the angle between vectors $G_V(z) - \hat{z}$ and $P_W\left(G_V(z)\right) - \hat{z}$. Obviously, by Pythagoras, $\gamma \in \left[0, \frac{\pi}{2}\right]$ and

$$\cos\gamma = \frac{\|P_W\left(G_V(z)\right) - \hat{z}\|}{\|G_V(z) - \hat{z}\|}$$

More than that, from Fact 2.46, we conclude that $\gamma \in \left[\frac{\beta}{2}, \frac{\pi}{2}\right]$. In particular, we get $\gamma \geq \frac{\beta}{2} \geq \frac{\theta_F}{2}$ and hence

$$\cos\gamma \leq \cos\frac{\theta_F}{2} = \sqrt{\frac{1 + c_F^2}{2}}$$

Then, enforcing similar arguments as in Case 1, we obtain

$$\begin{aligned}
\|G(z)-\hat{z}\|^2 &= \|G_W(G_V(z))-\hat{z}\|^2 \\
&= \|G_W(G_V(z))-P_W(G_V(z))\|^2 + \|P_W(G_V(z))-\hat{z}\|^2 \\
&\leq r_W^2\,\|G_V(z)-P_W(G_V(z))\|^2 + \|P_W(G_V(z))-\hat{z}\|^2 \\
&= r_W^2\left(\|G_V(z)-P_W(G_V(z))\|^2 + \|P_W(G_V(z))-\hat{z}\|^2\right) \\
&\quad + \left(1-r_W^2\right)\|P_W(G_V(z))-\hat{z}\|^2 \\
&= r_W^2\,\|G_V(z)-\hat{z}\|^2 + \left(1-r_W^2\right)\|P_W(G_V(z))-\hat{z}\|^2 \\
&= r_W^2\,\|G_V(z)-\hat{z}\|^2 + \left(1-r_W^2\right)\cos^2\gamma\,\|G_V(z)-\hat{z}\|^2 \\
&\leq \left(r_W^2 + \left(1-r_W^2\right)\frac{1+c_F^2}{2}\right)\|G_V(z)-\hat{z}\|^2 \\
&\leq \left(r_W^2 + \left(1-r_W^2\right)\frac{1+c_F^2}{2}\right)\|z-\hat{z}\|^2 \qquad (2.34)
\end{aligned}$$

The first line corresponds to the definition of G, the second is by Pythagoras and the third holds because G_W is a W-BAM. The fourth line is a rearrangement of terms, followed by Pythagoras in the fifth. Then, are employed respectively. At last, we used (2.30).

Analogously to the proof that $r_1 = \sqrt{r_V^2 + (1-r_V^2)\frac{1+c_F^2}{2}}$ is strictly smaller than 1, we can see that

$$r_2 := \sqrt{r_W^2 + (1-r_W^2)\,\frac{1+c_F^2}{2}} < 1$$

Finally, we can gather Cases 1 and 2. From (2.31), (2.33) and (2.34), we have $\|G(z)-\hat{z}\| \leq r_{V\cap W}\|z-\hat{z}\|$ for all $z \in \mathbb{R}^n$, with $r_{V\cap W} \in [0,1)$ given by $r_{V\cap W} := \max\{r_1, r_W, r_2\} = \max\{r_1, r_2\}$. □

The following fact generalizes the previous one for a finite composition of best approximation mappings.

Fact 2.48 (Finite composition of best approximation mappings)**.** *Let us consider an indexed family of ℓ affine subspaces $W = \{W_1, W_2, \ldots, W_\ell\}$ of $\mathbb{R}^n$ with nonempty intersection S_ℓ. Assume that each $G_{W_j} : \mathbb{R}^n \to \mathbb{R}^n$, $j = 1, \ldots, \ell$, is W_j-BAM. Then, $G := G_{W_\ell} \circ \cdots \circ G_{W_2} \circ G_{W_1}$ is a S_ℓ-BA M.*

Proof. The proof follows by an induction argument on ℓ, the number of affine subspaces. If $\ell = 1$, we have $G = G_{W_1}$ and then G is a S_ℓ -BAM. Assume the result for a fixed ℓ. Let $\widehat{W} := W \cup \{W_{\ell+1}\}$, where $W_{\ell+1}$ is an affine subspace such that it has nonempty intersection $S_{\ell+1}$ with S_ℓ, and let $G_{W_{\ell+1}}$ be a $W_{\ell+1}$ -BAM. Employing Fact 2.47 with S_ℓ

and $W_{\ell+1}$ playing the role of V and W, respectively, and $\widehat{G}$ and $G_{W_{\ell+1}}$ playing the role of G_V and G_W, respectively, we get that $\widehat{G} := G \circ G_{W_{\ell+1}}$ is a $S_{\ell+1}$-BAM. □

2.11 Linear convergence

Next, we recall the definition of Q-linear and R-linear convergence.

Definition 2.49 (Q-linear and R-linear convergence)**.** Let $\left(y^k\right)_{k\in\mathbb{N}} \subset \mathbb{R}^n$ be a convergent sequence to y^*. Assume that $y^k \neq y^*$ for all $k \in \mathbb{N}$. Define

$$q := \limsup_{k\to\infty} \frac{\left\|y^{k+1} - y^*\right\|}{\left\|y^k - y^*\right\|}, \quad \text{and} \quad r := \limsup_{k\to\infty} \left\|y^k - y^*\right\|^{1/k}.$$

Then, the convergence of $\left(y^k\right)_{k\in\mathbb{N}}$ is

(i) Q-superlinear if $q = 0$;

(ii) Q-linear if $q \in (0, 1)$;

(iii) Q-sublinear if $q \geq 1$;

(iv) R-superlinear if $r = 0$;

(v) R-linear if $r \in (0, 1)$;

(vi) R-sublinear if $r \geq 1$.

The values q, r are called asymptotic constants of $\left(y^k\right)_{k\in\mathbb{N}}$.

Q-linear convergence is well-known to imply R-linear convergence with the same asymptotic constant, but the converse statement is not necessarily true [57].

2.12 Circumcenters

Now, let us present the formal definition of circumcenter, as it pertains to the CRM and CARM methods.

Definition 2.50 (Circumcenters)**.** Let $x, y, z \in \mathbb{R}^n$ be given. The *circumcenter* $\operatorname{circ}\{x, y, z\} \in \mathbb{R}^n$ is a point satisfying

(i) $\|\operatorname{circ}\{x, y, z\} - x\| = \|\operatorname{circ}\{x, y, z\} - y\| = \|\operatorname{circ}\{x, y, z\} - z\|$;

(ii) $\operatorname{circ}\{x, y, z\} \in \operatorname{aff}\{x, y, z\} := \{w \in \mathbb{R}^n \mid w = x + \alpha(y - x) + \beta(z - x), \alpha, \beta \in \mathbb{R}\}$.

The point $\operatorname{circ}\{x, y, z\}$ is well and uniquely defined if the cardinality of the set $\{x, y, z\}$ is one or two. In the case in which the three points are all distinct, $\operatorname{circ}\{x, y, z\}$ is well and uniquely defined only if x, y and z are not collinear. The computation of $\operatorname{circ}\{x, y, z\}$ arises from the solution of a 2×2 linear system of equations. For more general concepts, definitions and results on circumcenters see [12,18,19,22]. In particular, the circumcenter of a set of $m > 3$ points is studied in [14,19].

For a finite collection of nonempty closed convex sets, we can define the following circumcenter operator:

Definition 2.51 (Circumcenter operator)**.** Let $\mathcal{B} := (Y_1, Y_2, \ldots, Y_q)$ be a collection of ordered nonempty closed convex sets in $\mathbb{R}^n$, where $q \geq 1$ is a fixed integer. The circumcenter of $\mathcal{B}$ at the point $z \in \mathbb{R}^n$ is denoted by $C_{\mathcal{B}}(z)$ and defined by the following properties:

(i) $\|z - C_{\mathcal{B}}(z)\| = \|R_{Y_1}(z) - C_{\mathcal{B}}(z)\| = \cdots = \left\|R_{Y_q} \cdots R_{Y_2} R_{Y_1}(z) - C_{\mathcal{B}}(z)\right\|$; and

(ii) $C_{\mathcal{B}}(z) \in \operatorname{aff}\left\{z, R_{Y_1}(z), R_{Y_2} R_{Y_1}(z), \ldots, R_{Y_q} \cdots R_{Y_2} R_{Y_1}(z)\right\}$

3 Review of projection methods for the convex feasibility problem

We formally define the convex feasibility problem (CFP):

$$\text{Find } x^* \in X := \bigcap_{i=1}^{m} X_i, \tag{3.1}$$

where $\{X_i\}_{i=1}^m$ are closed convex sets.

In this section, we will present three families of projection methods that are employed to solve this problem: MAP, DRM and CRM. The first two of were first established in the mid 20th century, with the latter being a more novel method which will be the focus of our subsequent work.

While the CFP is defined for m-sets, many of the most commonly used projection methods for solving this problem were developed for solving two-set problems. While some of them can be generalized and/or expanded for m sets, this is not the case for some. However, using the following reformulation, a two-set algorithm can be used for solving an m-set problem.

3.1 Pierra's product space reformulation

The classical way to apply two-set projection-based algorithms to problems for three or more sets is to use Pierra's product space reformulation [59], which reduces CFP (3.1) to the following two-set feasibility problem, which we call CFP-red:

$$\text{find } z^* \in K \cap U, \tag{3.2}$$

where the constraint sets K and U are given by

$$K := X_1 \times X_2 \times \cdots \times X_m \quad \text{and} \quad U := \{(x, x, \ldots, x) \in \mathbb{R}^{mn} \mid x \in \mathbb{R}^n\}$$

For the feasibility problem in a convex setting, the individual sets $X_1, \ldots, X_m$ are nonempty, closed and convex sets, properties which are inherited by K. Meanwhile, U, which is sometimes called the diagonal subspace, is a closed affine subspace. The fact that z^* must lie on U imposes that the solution must satisfy $x_1 = x_2 = \cdots = x_m$.

The original m-set feasibility problem (3.1) is equivalent to the two-set problem (3.2) by means of this relation:

$$x^* \in \bigcap_{i=1}^{m} X_i \subseteq \mathbb{R}^n \quad \Longleftrightarrow \quad z^* = (x^*, x^*, \ldots, x^*) \in K \cap U \subseteq \mathbb{R}^{nm} \tag{3.3}$$

Assuming projections are available in the original problem for the sets $X_1, \ldots, X_m$, we can obtain projectors for the reformulated problem by the following proposition:

Fact 3.1 (Product-space projectors)**.** *Let $x = (x_1, \ldots, x_m) \in \mathbb{R}^{nm}$. The projectors onto the sets K and U at x are given by*

$$P_K(x) = P_{X_1}(x_1) \times \cdots \times P_{X_m}(x_m) \quad \textit{and} \quad P_U(x) = \left(\frac{1}{m}\sum_{i=1}^{m} x_i, \ldots, \frac{1}{m}\sum_{i=1}^{m} x_i\right)$$

Proof. The fact that P_K is indeed a projector for x onto K can be shown directly from the definition of projection Definition 2.8. For P_U, since U is an affine subspace, we can use the fact that the projection theorem holds with equality for projections onto affine subspaces Fact 2.10, from which we prove the statement. □

With this in mind, while considering the methods which we will now present, we can reduce any m-set feasibility problem in $\mathbb{R}^n$ to a two-set problem in $\mathbb{R}^{nm}$. However, whenever possible, we will present a generalized m-set problem of a two-set algorithm that does not rely on this product-space reformulation.

3.2 Method of Alternating Projections - MAP

The *Method of Alternating Projections* (MAP), whose development dates back to von Neumann [62] and Halperin [40], is an algorithm used for solving the two-set CFP given by

$$\text{Find } z^* \in K_1 \cap K_2$$

with K_1, K_2 being closed convex sets and $K_1 \cap K_2 \neq \emptyset$. Note that this setting is more general than in (3.2), since one of the sets is not required to be an affine subspace.

The MAP algorithm is given by

$$x^{k+1} = P_{K_2} P_{K_1}\left(x^k\right), \quad k = 0, 1, 2, \ldots$$

MAP can solve an m-set CFP by employing Pierra's product space reformulation; in Section 6, we present a more general variant of MAP employing outer-approximate projections, which uses CFP-red to solve general convex feasibility problems. However, among its variants, MAP has one that can solve an m-set feasibility problem without relying on the product space reformulation. This variant is called the *Method of Cyclic Projections* (MCP), whose scheme is

$$x^{k+1} = P_{K_m} P_{K_{m-1}} \cdots P_{K_2} P_{K_1}\left(x^k\right), \quad k = 0, 1, 2, \ldots, \tag{3.4}$$

with $K_1, K_2, \ldots, K_m$ being closed convex sets and $\cap_{i=1}^{m} K_i \neq \emptyset$. Note that when $m = 2$, the MCP is equivalent to MAP.

The first steps towards proving MCP's convergence is to show that the MCP operator is α-averaged.

Fact 3.2 (α-averagedness of the MCP operator)**.** *Let $K_1, \ldots, K_m \subseteq \mathbb{R}^n$ be nonempty, closed and convex set. Then the operator $P_{K_m} P_{K_{m-1}} \cdots P_{K_1}$ is α-averaged with $\alpha = 1 - 2^{-m}$*

Proof. We will prove this by induction.

First, note that by Fact 2.27, P_{K_i} is $\frac{1}{2}$-averaged for $i = 0, 1, \ldots, m$. Therefore, we can write $P_{K_i} = \frac{1}{2}\,\mathrm{Id} + \frac{1}{2} R_{K_i}$, where R_{K_i} is the nonexpansive reflection operator associated with P_{K_i}.

For the base case, we will prove that $P_{K_2} P_{K_1}$ is $1 - 2^2 = 1 - \frac{1}{4}$-averaged.

Note that

$$\begin{aligned}
P_{K_2} P_{K_1} &= \frac{1}{2} P_{K_1} + \frac{1}{2} R_{K_2} P_{C1} \\
&= \frac{1}{4}\,\mathrm{Id} + \frac{1}{4} R_{K_1} + \frac{1}{2} R_{K_2} P_{C1} \\
&= \frac{1}{4}\,\mathrm{Id} + \frac{3}{4}\left(\frac{1}{3} R_{K_1} + \frac{2}{3} R_{K_2} P_{C1}\right) \\
&= \frac{1}{4}\,\mathrm{Id} + \left(1 - \frac{1}{4}\right)\left(\frac{1}{3} R_{K_1} + \frac{2}{3} R_{K_2} P_{K_1}\right)
\end{aligned}$$

Since R_{K_1} and $R_{K_2} P_{K1}$ are nonexpansive, by Fact 2.26 their convex combination is nonexpansive. Thus, we have that $P_{K_2} P_{K_1}$ is $1 - \frac{1}{4}$ averaged.

Now, for the induction step, assume that $P_{K_n} P_{K_{n-1}} \cdots P_{K_1}$ is $1 - 2^n$-averaged for $n > 2$. We want to show that $P_{K_{n+1}} P_{K_n} \cdots P_{K_1}$ is $1 - 2^{(n+1)}$-averaged. Denote $Q = P_{K_n} P_{K_{n-1}} \cdots P_{K_1}$, which we can also write as $Q = \frac{1}{2^n} Id + (1 - \frac{1}{2^n}) R_Q$. We then have that

$$\begin{aligned}
P_{K_{n+1}} Q &= \frac{1}{2} Q + \frac{1}{2} R_{K_{n+1}} Q \\
&= \frac{1}{2^{n+1}}\,\mathrm{Id} + \frac{2^n - 1}{2^{n+1}} R_Q + \frac{1}{2} R_{K_{n+1}} Q \\
&= \frac{1}{2^{n+1}}\,\mathrm{Id} + \frac{2^{n+1} - 1}{2^{n+1}}\left(\frac{2^n - 1}{2^{n+1} - 1} R_Q + \frac{2^n}{2^{n+1} - 1} R_{C_{n+1}} Q\right) \\
&= \frac{1}{2^{n+1}}\,\mathrm{Id} + \left(1 - \frac{1}{2^{n+1}}\right)\left(\frac{2^n - 1}{2^{n+1} - 1} R_Q + \frac{2^n}{2^{n+1} - 1} R_{C_{n+1}} Q\right)
\end{aligned}$$

Since R_Q and $R_{K_{n+1}} Q$ are nonexpansive, by Fact 2.26 their convex combination is nonexpansive. Thus, we have that $P_{K_{n+1}} Q = P_{K_{n+1}} P_{K_n} \cdots P_{K_1}$ is $1 - 2^{(n+1)}$-averaged. The result then follows. $\square$

Next, we establish the set of fixed points of the MCP operator.

Fact 3.3 (Fixed points of the MCP operator)**.** *Let $K_1, K_2, \ldots, K_m \subseteq \mathbb{R}^n$ be closed and convex sets with nonempty intersection. Then*

$$\mathrm{Fix}\left(P_{K_m} P_{K_{r-1}} \cdots P_{K_1}\right) = K_1 \cap \ldots \cap K_r$$

Proof. The inclusion $\operatorname{Fix}\left(P_{K_m}P_{K_{m-1}}\cdots P_{K_1}\right) \supseteq \cap_{i=1}^m K_i$ is immediate, so we need only consider the other inclusion. To this end, let $x \in \operatorname{Fix}\left(P_{K_m}P_{K_{m-1}}\cdots P_{K_1}\right)$ and let $y \in \cap_{i=1}^m K_i$. Since P_{K_i} is firmly nonexpansive by Fact 2.27 we have

$$\begin{aligned}
\|x-y\|^2 &\geq \|(\operatorname{Id}-P_{K_1})\,x\|^2 + \|P_{K_1}(x)-y\|^2 \\
&\geq \left(\|(\operatorname{Id}-P_{K_1})\,x\|^2 + \|(\operatorname{Id}-P_{K_2})\,P_{K_1}(x)\|^2\right) + \|P_{K_2}P_{K_1}(x)-y\|^2 \geq \cdots \\
&\geq \left(\|(\operatorname{Id}-P_{K_1})\,x\|^2 + \cdots + \left\|(\operatorname{Id}-P_{K_m})\left(P_{K_{m-1}}\cdots P_{K_1}\right)(x)\right\|^2\right) \\
&\quad + \left\|\left(P_{K_m}P_{K_{m-1}}\cdots P_{K_1}\right)(x)-y\right\|^2
\end{aligned}$$

Since $\left(P_{K_m}P_{K_{m-1}}\cdots P_{K_1}\right)(x) = x$ by assumption, it follows that

$$\|(\operatorname{Id}-P_{K_1})\,x\| = \cdots = \left\|(\operatorname{Id}-P_{K_m})\left(P_{K_{m-1}}\cdots P_{K_1}\right)(x)\right\| = 0$$

which implies that $x \in \cap_{i=1}^m K_i$, as claimed. □

As established in Section 2.7, we can now state convergence of the method of cyclic projections.

Fact 3.4 (Method of Cyclic Projections)**.** *Let $K_1, K_2, \ldots, K_m \subseteq \mathbb{R}^n$ be closed and convex sets with nonempty intersection. Given an initial point $x_0 \in \mathbb{R}^n$, consider the iteration generated by*

$$x_{k+1} = \left(P_{K_m}P_{K_{m-1}}\cdots P_{K_1}\right)(x_k)\,, \qquad \text{for } k = 0, 1, 2, \ldots$$

Then $(x_k)_{k\in\mathbb{N}}$ converges to a point $x^\star \in \cap_{i=1}^m K_i$

Proof. The operator $T = P_{K_m}P_{K_{m-1}}\cdots P_{K_1}$ is $(1-2^{-m})$ -averaged by Fact 3.2 with $\operatorname{Fix} T = \cap_{i=1}^m K_i \neq \emptyset$ by Fact 3.3. The result immediately follows by applying Fact 2.28. □

Another m-set MAP variant, which employs the product-space formulation, is the Method of Averaged Projections. Its convergence can be proved as a corollary of the MCP.

Fact 3.5 (Method of Averaged Projections)**.** *Suppose $K_1, K_2, \ldots, K_m \subseteq \mathbb{R}^n$ are closed and convex sets with nonempty intersection. Let $K = \cap_{i=1}^m K_i$ and let U be an affine subspace. Given a point $x^0 \in \mathbb{R}^n$, the method of averaged projections is the fixed point iteration governed by*

$$x^{k+1} = T(x_k)\,, \quad \text{where } T = \frac{P_{K_m} + \cdots + P_{K_1}}{m}$$

Then, the sequence $\left(x^k\right)_{k\in\mathbb{N}}$ converges to a point in $\cap_{i=1}^m K_i$.

Proof. Set $z^0 := (x^0, x^0, \ldots, x^0) \in U$ and observe that

$$z^{k+1} = P_U P_K \left(z^k\right) = \left(T\left(x^k\right), T\left(x^k\right), \ldots, T\left(x^k\right)\right), \qquad \text{for } k = 0, 1, 2, \ldots$$

where $z^k = \left(x^k, x^k, \ldots, x^k\right) \in U$. Fact 3.4 therefore asserts that $\left(z^k\right)_{k\in\mathbb{N}}$ converges to some $z^\star = (x^\star, x^\star, \ldots, x^\star) \in K \cap U$ and the claimed result thus follows in view of (3.3). □

Despite dating back to 1950, MAP remains a contemporary topic of research. In the past 20, many works such as [3, 7–10, 17, 21, 24, 25, 28, 32–34, 36, 41–43, 46–51, 61] (a far from exhaustive selection) have delved into more intricate aspects and variations of MAP.

3.3 Douglas-Rachford Method - DRM

Introduced in 1956 by Douglas and Rachford [31], the *Douglas-Rachford method* is another classic method which remains relevant to this day, much like MAP.

Given two nonempty, closed and convex sets $K_1, K_2 \subseteq \mathbb{R}^n$, the DR algorithm is the fixed point iteration generated by the Douglas-Rachford operator, T_{K_1,K_2}, defined by

$$x^{k+1} = T_{K_1,K_2}\left(x^k\right), \quad T_{K_1,K_2} := \frac{\mathrm{Id} + R_{K_2} R_{K_1}}{2}$$

In the following facts, we verify that the DR operator is nonexpansive (1/2-averaged, in particular) and establish its set of fixed points.

Fact 3.6 (Nonexpansivity of the DR operator)**.** *Let K_1 and K_2 be two nonempty, closed and convex subsets of $\mathbb{R}^n$. Then the DR operator T_{K_1,K_2} is $\frac{1}{2}$-averaged and, thus, firmly nonexpansive.*

Proof. Since K_1 and K_2 are nonempty, closed and convex, the reflectors R_{K_1} and R_{K_2} are nonexpansive by Fact 2.27 and, consequently, so too is their composition. We therefore have that $T_{K_1,K_2} = \left(1 - \frac{1}{2}\right)\mathrm{Id} + \frac{1}{2} R_B R_A$ with $R_B R_A$ nonexpansive; i.e., $T_{A,B}$ is $\frac{1}{2}$ -averaged. The remaining assertion follows from Fact 2.25(iii). □

Fact 3.7 (Fixed points of the DR operator)**.** *Let $K_1, K_2 \subseteq \mathbb{R}^n$ be nonempty, closed and convex, and let T_{K_1,K_2} be the DR operator. Then*

$$\operatorname{Fix} T_{K_1,K_2} \neq \emptyset \iff K_1 \cap K_2 \neq \emptyset$$

and, moreover, $P_{K_1}(x) \in K_1 \cap K_2$ for all $x \in \operatorname{Fix} T$

Proof. Observe that $K_1 \cap K_2 \subseteq \operatorname{Fix}(R_{K_2} R_{K_1}) = \operatorname{Fix} T_{K_1,K_2}$. Moreover, $x \in \operatorname{Fix}(R_{K_2} R_{K_1})$ if and only if

$$x = (2P_{K_2} - \text{ Id }) (2P_{K_1} - \text{ Id }) (x) = 2P_{K_2} (2P_{K_1}(x) - x) - 2P_{K_1}(x) + x$$

that is, $x \in \operatorname{Fix} T_{K_1,K_2}$ if and only if $P_{K_2} (2P_{K_1}(x) - x) = P_{K_1}(x)$. Consequently, we have $P_{K_1} (\operatorname{Fix} T_{K_1,K_2}) = K_1 \cap K_2$ and the result follows. □

Using the facts in Section 2.7, we can now prove the Douglas-Rachford method's convergence.

Fact 3.8 (Douglas-Rachford method)**.** *Let $K_1, K_2 \subseteq \mathbb{R}^n$ be nonempty, closed and convex sets, and consider the DR operator T_{K_1,K_2}. Given any initial point $x^0 \in \mathbb{R}^n$ consider the iteration generated by*

$$x^{k+1} = T_{K_1,K_2} \left(x^k\right), \qquad \text{for } k = 0, 1, 2, \ldots$$

Then either exactly one of the following statements is true:

(i) *$K_1 \cap K_2 \neq \emptyset$ and $\left(x^k\right)_{k \in \mathbb{N}}$ converges to a point $x^* \in \operatorname{Fix} T_{K_1,K_2}$, with $P_{K_1}(x^*) \in K_1 \cap K_2$*

(ii) *$K_1 \cap K_2 = \emptyset$ and $\left\|x^k\right\| \to +\infty$*

Proof. The operator T_{K_1,K_2} is $\frac{1}{2}$ -averaged by Fact 3.6. Suppose first that $A \cap B \neq \theta$ Then, according to Fact 3.7. we have $\operatorname{Fix} T_{K_1,K_2} \neq \emptyset$ and therefore we may apply Fact 2.28 to deduce that

$$x^k \to x^* \in \operatorname{Fix} T_{K_1,K_2}$$

Furthermore, Fact 3.7 yields that $P_{K_1}(x^*) \in K_1 \cap K_2$ and thus (i) is proved. Consider now the case when $K_1 \cap K_2 = \emptyset$. Again, by applying Fact 3.7 we deduce that $\operatorname{Fix} T_{K_1,K_2} = \emptyset$ and hence (ii) immediately follows from Fact 2.31. □

Similarly to MAP, the DRM also has a cyclic variant which can be used for m-sets without using the product space reformulation. This variant is known as the *Cyclic Douglas-Rachford Method* [23], introduced in 2014 by Borwein and Tam, whose operator is defined as follows:

$$T_{[K_1,K_2,\ldots,K_m]} := T_{K_r,K_1} T_{K_{m-1},K_m} \cdots T_{K_2,K_3} T_{K_1,K_2}$$

Notably, when $m = 2$, the Cyclic DRM operator does not match the usual two-set DRM operator; that is,

$$T_{[K_1,K_2]} = T_{K_2,K_1} T_{K_1,K_2} \neq T_{K_1,K_2}.$$

An alternative version known as the *Cyclically Anchored Douglas-Rachford*, however, reduces to the usual DRM operator when working with two sets. Presented by Bauschke, Noll and Pham in [11], its operator for m sets is given by

$$T_{K_1,[K_2,\ldots,K_m]} := T_{K_1,K_m} T_{K_1,K_{m-1}} \cdots T_{K_1,K_2},$$

where K_1 is called the anchor set. Furthermore, when $K_1 = \mathbb{R}^n$, the operator coincides with the $(m-1)$-set Method of Cyclic Projections operator (3.4); that is, $T_{\mathbb{R}^n,[K_2,\ldots,K_m]} = P_{K_m} P_{K_{m-1}} \cdots P_{K_2}$.

In two-set problems, the Douglas-Rachford operator can also be generalized as an α-averaged version in Eckstein and Bertsekas' [35] variant known as the *Generalized Douglas-Rachford Method* (GDR), whose operator

$$x^{k+1} = T_{K_1,K_2,\alpha}\left(x^k\right) := (1-\alpha)x^k + \alpha R_{K_2} R_{K_1}\left(x^k\right), \quad \alpha \in (0,1).$$

is equal to the standard DR operator when $\alpha = \frac{1}{2}$.

Another two-set variant is the *Relaxed Averaged Alternating Reflections* (RAAR) [54] method, whose operator is an average of the DR operator and the projection onto the first set:

$$x^{k+1} = (1-\beta)P_{K_1}\left(x^k\right) + \beta T_{K_1,K_2}\left(x^k\right), \quad \beta \in (0,1).$$

Beyond these variants, the Douglas-Rachford method has given way to multiple variations and applications. For a comprehensive survey on the DRM, see [53], as well as [1] for a thorough review on convex and nonconvex applications.

The next subsection deals with the main algorithm of our interest, which also happens to be inspired by the Douglas-Rachford method.

3.4 Circumcentered Reflection Method - CRM

The *Circumcentered Reflection Method* is a circumcentered version of the Douglas-Rachford method, first introduced in [18, 19] and further enhanced in [2, 12, 14–16, 20, 22, 29, 30, 52, 58].

Starting from a point in U, the CRM operator is the circumcenter Definition 2.50 of the triangle formed by the starting point, its reflection through the convex set K and the composite reflection through K and then through U. Formally, the CRM scheme reads as

$$z^{k+1} := C\left(z^k\right) = \text{ circumcenter } \left\{z^k, R_K\left(z^k\right), R_U R_K\left(z^k\right)\right\}, \text{ with } z^k \in U,$$

where the circumcenter operator's properties are given by Definition 2.51.

CRM was originally developed as tool for solving the *best approximation problem*, consisting of projecting onto the intersection of affine subspaces. It was from this following foundation that CRM was later shown to solve convex feasibility problems as well.

3.4.1 CRM for the best approximation problem in affine subspaces

Let $\mathcal{B} := \{U_i\}_{i\in \mathrm{I}}$ be a family of finitely many affine subspaces in $\mathbb{R}^n$ with $I := \{1, 2, \ldots, q\}$ and q fixed. Assume that the intersection of the family, denoted by $U_{\mathcal{B}} := \bigcap_{i\in \mathrm{I}} U_i$, is nonempty. The *best approximation problem* we are interested in is given by

$$\text{Find } z^* \in U_{\mathcal{B}} : \|z - z^*\| = \min_{s\in U_{\mathcal{B}}} \|z - s\|.$$

It is well known that this problem has a unique solution z^* and that $z^* = P_{U_{\mathcal{B}}}(z)$ if, and only if, $z - z^* \in U_{\mathcal{B}}^{\perp}$, i.e., $\langle z - z^*, s\rangle = 0$ for all $s \in U_{\mathcal{B}}$.

We begin by characterizing and defining CRM for the intersection of affine subspaces, proof of which benefits from the following auxiliary linear operator. Due to its length and reliance on other facts not presented in this work, we refer to the original paper for the proof.

Fact 3.9. *Consider a collection of affine subspaces $\mathcal{B} = (U_1, U_2, \ldots, U_q)$ with nonempty intersection $U_{\mathcal{B}} := \bigcap_{i=1}^{q} U_i$. Denote $U_z := \operatorname{aff}\{x, R_{U_1}(x), R_{U_2}R_{U_1}(x), \ldots, R_{U_m}\cdots R_{U_2}R_{U_1}(x)\}$. Let $A : \mathbb{R}^n \to \mathbb{R}^n$, with*

$$A(x) := \frac{1}{m}\sum_{i=1}^{m} A_i(x), \text{ where } A_1 := \frac{1}{2}\left(\mathrm{Id} + P_{U_1}\right) \text{ and } A_i := \frac{1}{2}\left(\mathrm{Id} + P_{U_i}R_{U_{i-1}}\ldots R_{U_1}\right)$$

Then:

(i) *A is linear and $\frac{1}{2}$ -averaged, i.e., firmly nonexpansive;*

(ii) $\operatorname{Fix} A := \{x \in \mathbb{R}^n \mid A(x) = x\} = U_{\mathcal{B}}$;

(iii) *for all* $x \in \mathbb{R}^n, A(x) \in U_z$;

(iv) *there exists* $r_A \in [0,1)$ *such that for all* $x \in \mathbb{R}^n$ *it holds that* $\|A(x) - P_{U_{\mathcal{B}}}(x)\| \leq r_A\|x - P_{U_{\mathcal{B}}}(x)\|$;

(v) *let* $x \in \mathbb{R}^n$ *be given. Then,* $P_{U_{\mathcal{B}}}\left(A^k(x)\right) = P_{U_{\mathcal{B}}}(x)$ *for all* $k \in \mathbb{N}$;

(vi) *let* $x \in \mathbb{R}^n$ *be given. The sequence* $\left\{A^k(x)\right\}_{k\in\mathbb{N}}$ *converges to* $P_{U_{\mathcal{B}}}(x)$ *with linear rate* r_A.

Proof. See the proof for Lemma 2.1 in [19]. □

With the help of this α-averaged operator, we can state good definition and characterization of CRM for intersecting affine subspaces.

Fact 3.10 (Good definition and characterization of CRM for intersecting affine subspaces)**.** *Consider a collection of affine subspaces* $\mathcal{B} = (U_1, U_2, \ldots, U_q)$ *with nonempty intersection* $U_{\mathcal{B}} := \bigcap_{i=1}^q U_i$. *Denote* $U_z := \operatorname{aff}\left\{z, R_{U_1}(z), R_{U_2}R_{U_1}(z), \ldots, R_{U_q}\cdots R_{U_2}R_{U_1}(z)\right\}$. *For any* $z \in \mathbb{R}^n, C_{\mathcal{B}}(z)$ *exists, is unique and fulfills:*

(i) *for any* $s \in U_{\mathcal{B}}$, $C_{\mathcal{B}}(z) = P_{U_z}(s)$;

(ii) $P_{U_{\mathcal{B}}}\left(C_{\mathcal{B}}(z)\right) = P_{U_{\mathcal{B}}}(z)$.

Proof. (i) Let $z \in \mathbb{R}^n$ and $s \in U_{\mathcal{B}}$ be arbitrary but fixed.

Since $s \in U_{\mathcal{B}}$, the definition of reflections and the Pythagoras theorem imply the optimality condition

$$\|z - s\| = \|R_{U_1}(z) - s\| = \|R_{U_2}R_{U_1}(z) - s\| = \cdots = \left\|R_{U_q}\cdots R_{U_2}R_{U_1}(z) - s\right\|. \tag{3.5}$$

Using the Pythagoras theorem on the $\triangle[z - s, P_{U_z}(s) - s, z - P_{U_z}(s)]$, we find that

$$\|z - s\|^2 = \|P_{U_z}(s) - s\|^2 + \|z - P_{U_z}(s)\|^2.$$

Doing the same for similar triangles, replacing z with composite reflections, for $i = 1, \ldots, m$ we have

$$\|R_{U_i}\cdots R_{U_2}R_{U_1}(z) - s\|^2 = \|P_{U_z}(s) - s\|^2 + \|R_{U_i}\cdots R_{U_2}R_{U_1}(z) - P_{U_z}(s)\|^2.$$

Combining these equalities with (3.5) yields

$$\begin{aligned}\|z - P_{U_z}(s)\| = \|R_{U_1}(z) - P_{U_z}(s)\| &= \|R_{U_2}R_{U_1}(z) - P_{U_z}(s)\| \\ &= \cdots \\ &= \left\|R_{U_q}\cdots R_{U_2}R_{U_1}(z) - P_{U_z}(s)\right\|.\end{aligned}$$

Thus, $P_{U_z}(s)$ is a circumcenter w.r.t. $z, R_{U_1}(z), R_{U_2}R_{U_1}(z), \ldots, R_{U_q}\cdots R_{U_2}R_{U_1}(z)$ in $U_{\mathcal{B}}$. The uniqueness of the circumcenter follows directly from the uniqueness of centers of closed balls in $\mathbb{R}^{\dim(U_{\mathcal{B}})}$, where $\dim(U_{\mathcal{B}}) \leq \min\{q, n\}$.

(ii) From the fact that $C(x) = P_{U_z}(s)$, proved in (i), and the facts that $A(z) \in U_{\mathcal{B}}$ and $\|A(z) - P_{U_{\mathcal{B}}}(z)\| \leq r_A \|z - P_{U_{\mathcal{B}}}(z)\|$ from Fact 3.9(iii) and (iv), respectively, we have

$$\|C(z) - P_{U_{\mathcal{B}}}(z)\| \leq \|A(z) - P_{U_{\mathcal{B}}}(z)\| \leq r_A \|z - P_{U_{\mathcal{B}}}(z)\|.$$

Once $P_{U_{\mathcal{B}}}(C(z)) = P_{U_{\mathcal{B}}}(z)$ is proven, $P_{U_{\mathcal{B}}}(C_B) = P_{U_{\mathcal{B}}}$ will follow easily by induction.

Using (i), we can use the Pythagoras theorem on suitable triangles to obtain

$$\|P_{U_{\mathcal{B}}}(C(z)) - C(z)\|^2 = \|P_{U_{\mathcal{B}}}(C(z)) - z\|^2 - \|C(z) - z\|^2, \text{ and}$$
$$\|P_{U_{\mathcal{B}}}(x) - C(z)\|^2 = \|P_{U_{\mathcal{B}}}(z) - z\|^2 - \|C(z) - z\|^2,$$

respectively. These equalities combined with $\|P_{U_{\mathcal{B}}}(z) - z\| \leq \|P_{U_{\mathcal{B}}}(C(x)) - z\|$ imply that

$$\|P_{U_{\mathcal{B}}}(z) - C(z)\| \leq \|P_{U_{\mathcal{B}}}(C(z)) - C(z)\|,$$

thus proving $P_{U_{\mathcal{B}}}(C_B) = P_{U_{\mathcal{B}}}(x)$. □

Fact 3.11 (Convergence of CRM for intersecting affine subspaces)**.** *Let $z \in \mathbb{R}^n$ be given. Then, the CRM sequence $\left\{C^k(z)\right\}$ converges to $P_{U_{\mathcal{B}}}(z)$ with the linear rate $r_A \in [0, 1)$.*

Proof. Given $z \in \mathbb{R}^n$ and for all $k \in \mathbb{N}$, Fact 3.10 provides

$$\begin{aligned}\left\|C^{k+1}(z) - P_{U_{\mathcal{B}}}(z)\right\| &= \|C\left(C^k(z)\right) - P_{U_{\mathcal{B}}}(z)\left(C^k(z)\right)\| \\ &\leq r_A \left\|C^k(z) - P_{U_{\mathcal{B}}}(z)\left(C^k(z)\right)\right\| \\ &= r_A \left\|C^k(z) - P_{U_{\mathcal{B}}}(z)\right\|\end{aligned}$$

which implies

$$\left\|C^k(z) - P_{U_{\mathcal{B}}}(z)\right\| \leq r_A^k \|z - P_{U_{\mathcal{B}}}(z)\|$$

□

The next step in the development of CRM came in the form of a block-wise version, first presented in [20].

3.4.2 Block-wise CRM for the affine best approximation problem

For this subsection, recall the concept of best approximation mapping (BAM) and related facts, presented in Section 2.10.

Referring to the proof of a previously stated fact on BAMs, we prove that the circumcenter operator is a BAM.

Fact 3.12 (Circumcenter operator is a BAM)**.** *Consider a block of affine subspaces* $\mathcal{B} = \{U_1, U_2, \ldots, U_q\}$ *with* $S_{\mathcal{B}} = \cap_{i=1}^{q} U_i$ *nonempty. Then, there exists a constant* $r_{\mathcal{B}} \in [0, 1)$ *so that*

$$\|C_{\mathcal{B}}(z) - P_{S_{\mathcal{B}}}(z)\| \leq r_{\mathcal{B}} \|z - P_{S_{\mathcal{B}}}(z)\|$$

for all $z \in \mathbb{R}^n$*. Moreover,* $P_{S_{\mathcal{B}}}(C_{\mathcal{B}}(z)) = P_{S_{\mathcal{B}}}(z)$

Proof. See the proof for Fact 3.10(ii). □

The following terminology will help define a new circumcenter scheme.

Definition 3.13 (Block partition)**.** We say that $\mathbf{B} = \{\mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_p\}$ is a collection of blocks in crescent index order (with cardinality p) for the ordered affine subspaces $U_1, U_2, \ldots, U_m$ if we can write $\mathcal{B}_1 = \{U_{q_0+1}, U_{q_0+2}, \ldots, U_{q_1}\}$, $\mathcal{B}_2 = \{U_{q_1+1}, U_{q_1+2}, \ldots, U_{q_2}\}$, $\ldots$, $\mathcal{B}_p = \left\{U_{q_{p-1}+1}, U_{q_{p-1}+2}, \ldots, U_{q_p}\right\}$, with $q_0 = 0$ and $q_p = m$. We assume that every block $\mathcal{B}_i$ has size $q_i \geq 1, i = 1, \ldots, p$.

We now define the block-wise circumcentered-reflection operator, and subsequently prove that it is also a BAM.

Definition 3.14 (Block-wise circumcentered-reflection)**.** Let $\mathbf{B} = \{\mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_p\}$ be a collection of blocks for the affine subspaces $U_1, U_2, \ldots, U_m$ and assume that crescent index order is taken for both blocks and subspaces. Then, for a point $z \in \mathbb{R}^n$ we define the block-wise circumcentered-reflection step $C_{\mathbf{B}}(z)$ by

$$C_{\mathbf{B}}(z) := C_{\mathcal{B}_p} \circ C_{\mathcal{B}_{p-1}} \circ \cdots \circ C_{\mathcal{B}_2} \circ C_{\mathcal{B}_1}(z)$$

Fact 3.15 (Block-wise operator is a BAM)**.** *Let* $C_{\mathbf{B}}$ *be the block-wise circumcentered-reflection operator regarding* $\mathbf{B}$*. Then, there exists a constant* $r_{\mathbf{B}} \in [0, 1)$ *so that*

$$\|C_{\mathbf{B}}(z) - P_S(z)\| \leq r_{\mathbf{B}} \|z - P_S(z)\|$$

for all $z \in \mathbb{R}^n$*. Moreover,* $P_S(C_{\mathbf{B}}(z)) = P_S(z)$ *and the convergence of* $\left(C_{\mathbf{B}}^k(z)\right)_{k \in \mathbb{N}}$ *is linear to the unique solution* $P_S(z)$*,*

$$\lim_{k \to \infty} C_{\mathbf{B}}^k(z) = P_S(z)$$

Furthermore, the global Q-linear rate is $r_{\mathbf{B}} \in [0, 1)$*, i.e., for all* $k \in \mathbb{N}$*,*

$$\left\|C_{\mathbf{B}}^k(z) - P_S(z)\right\| \leq r_{\mathbf{B}}^k \|z - P_S(z)\|$$

Proof. Due to Definition 3.14 $C_{\mathbf{B}}$ is a composition of circumcenter operators, which of each is a BAM (Fact 2.48) and thus, by Fact 3.12 , it is itself a BAM. The claims on the sequence $\left(C_{\mathbf{B}}^k(z)\right)_{k \in \mathbb{N}}$ follow then from Fact 2.45. □

Since the definition of best approximation mapping is essentially a relaxation of a projection operator, the block-wise CRM operator can be seen as an approximate projection in that sense.

The block-wise CRM has been shown in [20, Lemma 3, Lemma 4, Theorem 4] to converge in one step for the intersection of hyperplanes. Next, we will show that is also converges in one step for the intersection of a hyperplane and an affine subspace, which paved the way for the use of CRM as a tool for the convex feasibility problem.

3.4.3 CRM for the convex feasibility problem

CRM was shown to solve CFP-red (3.2), that is,

$$\text{Find } z^* \in K \cap U$$

with $K \subset \mathbb{R}^n$ being a closed convex set, $U \subset \mathbb{R}^n$ an affine subspace and $K \cap U$ nonempty.

CRM's convergence was initially proved for the intersections of affine subspaces and hyperplanes. From this foundation, its convergence for a general convex intersection (in a product space setting) followed. Having previously established the well-definedness and characterization of CRM, we can prove that it converges in one step for a hyperplane and an affine subspace.

Fact 3.16 (One step convergence of CRM for a hyperplane and an affine subspace)**.** *Let $H, U \subset \mathbb{R}^n$ be a hyperplane and an affine subspace, respectively. If $H \cap U$ is nonempty, then*

$$P_{H\cap U}(z) = \ \textit{circumcenter}\ \{z, R_H(z), R_U R_H(z)\}$$

for any $z \in U$.

Proof. Let $z \in U$. If z lies also in H, the result follows trivially. Therefore, assume $z \notin H$.

Note that if $P_U R_H(z)$ were to coincide with z, then $\operatorname{aff}\{z, R_H(z), R_U R_H(z)\} = \operatorname{aff}\{R_H(z), R_U R_H(z)\}$, because $z = P_U R_H(z) = \frac{1}{2}(R_H(z) + R_U R_H(z))$. Since the only point equidistant to $R_H(z)$ and $R_U R_H(z)$ in $\operatorname{aff}\{R_H(z), R_U R_H(z)\}$ is $P_U R_H(z) = z$, this would mean that $z = \operatorname{circ}\{z, R_H(z), R_U R_H(z)\}$ and, in particular, $z = R_H(z)$, implying that $z \in H$, a contradiction. Since z and $P_U R_H(z)$ are distinct, the line connecting these points, denoted here by L_z, is well-defined. Note that L_z lies entirely in U and crosses the segment $\overline{R_H(z) R_U R_H(z)} := \{w \in \mathbb{R}^n : w = tR_H(z) + (1-t)R_U R_H(z), t \in [0,1]\}$ at its midpoint, namely $P_U R_H(z)$ perpendicularly, thus being a bisector of the segment $\overline{R_H(z) R_U R_H(z)}$. Therefore, $\tilde{z} := \operatorname{circ}\{z, R_H(z), R_U R_H(z)\}$ has to lie in L_z and, consequently, in U. From the fact that circumcenter is the point where the bisectors of the triangle of vertices $z, R_H(z)$ and $R_U R_H(z)$ intersect, we also have that $\overline{zP_H(z)} \perp \overline{\tilde{z}P_H(z)}$; this, combined with the fact that H is a hyperplane, implies that $\tilde{z} \in H$. Thus, $\tilde{z} \in H \cap U$. Finally, from Fact 3.10 we have that $P_{H\cap U}(z) = P_{H\cap U}(\tilde{z})$. Hence, $P_{H\cap U}(z) = \tilde{z}$ which completes the proof. $\square$

With this result as a starting point, CRM was proven in [22] to converge for general convex intersection. In Section 5, we will present CARM, which is a more general version of CRM employing approximate projections (and reflections). Thus, the proofs and statements about the convergence of CARM are also valid for CRM.

In [22], CRM was shown to outperform both MAP and DRM. The following image and theorem, first in the same paper, help illustrate the comparison.

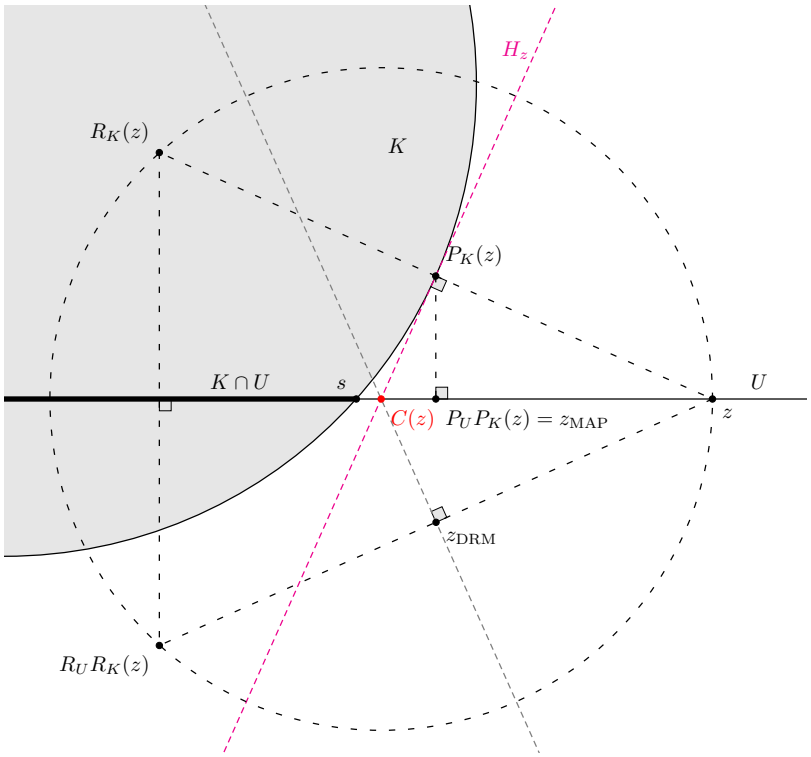

Fig. 3.1: Illustration of CRM, MAP and DRM for the intersection between an affine U and a convex K.

Fact 3.17 (Comparing CRM, MAP and DRM)**.** *Assume $K, U \subset \mathbb{R}^n$ as in Fact 3.10 and let $z \in U$ be given. Also recall the notation $z_{\mathrm{MAP}} := P_U P_K(z)$, $z_{\mathrm{DRM}} := \frac{1}{2}\left(\mathrm{Id} + R_U R_K\right)(z)$ and $C(z) := \mathrm{circumcenter}\left\{z, R_K(z), R_U R_K(z)\right\}$ Then, for any $s \in K \cap U$ we have*

(i) $\|C(z) - s\| \leq \|z_{\mathrm{MAP}} - s\| \leq \|z_{\mathrm{DRM}} - s\|$,

(ii) $\mathrm{dist}(C(z), K \cap U) \leq \mathrm{dist}\left(z_{\mathrm{MAP}}, K \cap U\right) \leq \mathrm{dist}\left(z_{\mathrm{DRM}}, K \cap U\right)$.

Proof. See the proof for Theorem 2 in [22]. □

A similar comparison will be made for the approximate variants of CRM and MAP only, as no suitable DRM equivalent exists.

Before we present CARM, we first need to discuss what we mean by approximate projections and why the particular definition we use for CARM was chosen instead of other existing approximations for orthogonal projections.

4 Approximate projections

We recall that the main objective of this work is to adapt an existing projection method by employing approximate projections in place of the exact (orthogonal) projections. For this purpose, one of the many proposed methods of approximating projections had to be chosen. Each of them relinquishes or modifies certain properties of the orthogonal projections, in order to circumvent the unavailability of said projections or to avoid problems derived from some of these properties. These projections have been given names such as *approximate*, *inexact* or *relaxed*, with some inconsistencies as to the use of these adjectives.

First, we present the definition of approximate projection chosen for our work.

4.1 Outer-approximate projections

The notion of an *outer-approximate projection* was introduced by Fukushima in [38, Lemma 4], whose definition we present below. In subsequent sections, we will use the terms outer-approximate projection and approximate projection interchangeably.

Definition 4.1 (Outer-approximate projector)**.** Let $B \subseteq \mathbb{R}^n$ be a nonempty set and $X \subseteq \mathbb{R}^n$ a closed and convex set. A function $\tilde{P}_X : B \to \mathbb{R}^n$ is called an *outer-approximate projector* defined on B and associated to X if the following properties are satisfied

(i) there exists $\gamma \in [0, 1)$ such that

$$\operatorname{dist}\left(\tilde{P}_X(x), X\right) \leq \gamma \operatorname{dist}(x, X), \text{ for all } x \in B;$$

(ii) for any $x \in B$, it holds that

$$\left\langle x - \tilde{P}_X(x), y - \tilde{P}_X(x) \right\rangle \leq 0, \text{ for all } y \in X.$$

The point $\tilde{P}_X(x)$ will be called *outer-approximate projection* of $x \in B$ onto X. Definition 4.1 guarantees, in view of property (ii), the existence of a hyperplane $H_X(x)$ separating x from X such that $\tilde{P}_X(x)$ is the orthogonal projection of x onto $H_X(x)$. In light of this, we can interpret property (i) as meaning that the distance of the hyperplane $H_X(x)$ to X is at least a fixed percentage of the distance from x to X. This percentage is given by γ; if $\gamma = 0$, the outer-approximate projection is precisely the exact orthogonal projection. If, for instance, $\gamma = 0.3$, then $\tilde{P}_X(x)$ advances at least 70% towards X by means of the distance of x to X. We emphasize that a percentage of improvement on the distance, *per se*, is not sufficient for having an outer-approximate projection; the existence of the separating hyperplane is crucial, too.

In short, it can be said that the outer-approximate projection preserves the sense of approaching the target set from the "outside", that is, by means of a hyperplane

separating the original point being projected and the set it is being projected towards, as well as establishing a minimum gain in distance relative to the original point, keeping the idea that a projection should bring a point somewhat closer to the target set, whenever the point does not belong to the set. Unfortunately, this approach sacrifices the nonexpansivity of the projection and it does not require the approximate projection to lie on the target set.

A major advantage of this definition, as will be seen in Section 8, is the existence of an algorithm which uses subgradients Section 2.9 to obtain outer-approximate projections.

Now, we will present alternative definitions of "approximate" (in a general sense) projections and comment briefly on how they differ from the one we chose to employ.

4.2 Alternative definitions of approximate projections

The following two definitions were taken from [33]. This paper also attempts to avoid the intractability of exact projections, (mostly focusing on nonconvex sets, however). Instead of providing a method for finding a more computational-friendly approximate projection, Drusvyatskyi and Lewis allow, to a certain degree, errors on the computation of the exact projections, still obtaining convergence of an alternating projections scheme with non-exact projections onto one of the sets.

The first approximation was named by the authors as an *inexact* projection, in the sense that it establishes a bound for computing errors in the calculation of the projection whilst not imposing further conditions.

Definition 4.2 (Drusvyatskyi and Lewis' inexact projection)**.** Let $B \subseteq \mathbb{R}^n$ be a nonempty set and $X \subseteq \mathbb{R}^n$ a closed set. For some $\varepsilon \geq 0$, $\tilde{P}_X(x)$ is a DL-inexact projection from $x \in B$ onto X if $\|\tilde{P}_X(x) - P_X(x)\| \leq \varepsilon \operatorname{dist}(x, K)$.

This definition allows for employing of a slightly miscalculated projection. It assumes the exact projection is known, but that it's exact computation might be inefficient or erroneous to a certain extent. Thus, it is not suitable for replacing projections when they are unknown or unavailable.

The second definition, presented below, was called an *approximate* projection. While similar in intent to the inexact projection, this approximation is now required to be on the target set.

Definition 4.3 (Drusvyatskyi and Lewis' approximate projection)**.** Let $B \subseteq \mathbb{R}^n$ and $X \subseteq \mathbb{R}^n$ be closed sets. Let $z \in X$ and $y \in B$ such that $y \in P_B(z)$. For some $\varepsilon \geq 0$, $\tilde{P}_X(y)$ is an approximate projection of y onto X if $\|\tilde{P}_X(y) - P_X(y)\| \leq \varepsilon \|y - z\|$.

In [39], Gonçalves, Gonçalves and Oliveira propose a variation on the Levenberg-Marquardt method using a type of feasible inexact projection which they call an ε-projection. These projections can be obtained iteratively, using the conditional gradient

(Frank-Wolfe) method [37, 45], and can replace exact projections when costly or unavailable. In that sense, they work more similarly to Fukushima's outer-approximate projections than either of Drusvyatskyi and Lewis' projections. The definition of the (inexact) ε-projection is given below.

Definition 4.4 (ε-projection)**.** Let K be a nonempty closed convex set contained in an open set $\Omega \subset \mathbb{R}^n$. Let $x \in \mathbb{R}^n$ and $\varepsilon \geq 0$ be given. $P_K(x, \varepsilon)$ is an *ε-projection* of x onto K when

$$P_K(x, \varepsilon) \in K \text{ and } \langle x - P_K(x, \varepsilon), y - P_K(x, \varepsilon)\rangle \leq \varepsilon, \text{ for all } y \in K.$$

The ε-projection has some interesting properties. Not only does the ε-projection lie on the target set, we also have that the orthogonal projection is recovered if $\varepsilon = 0$. In this characterization, the variational inequality of the projection theorem is only approximated and not preserved, as with the outer-approximate projection; however, this approximation on the inequality, combined with the fact that $P_K(x, \varepsilon) \in K$, allows for a pseudo-nonexpansivity of the operator $P_K(\cdot, \cdot)$ which we will refer to as the ε-nonexpansivity property.

Fact 4.5 (ε-nonexpansivity)**.** *For any $x, y \in \mathbb{R}^n$ and $\varepsilon \geq 0$, we have*

$$\|P_K(x, \varepsilon) - P_K(y)\| \leq \|x - y\| + \sqrt{\varepsilon}$$

Proof. It follows from Fact 2.9 and Definition 4.4 that

$$\langle x - P_K(x), P_K(x, \varepsilon) - P_K(x)\rangle \leq 0, \quad \langle x - P_K(x, \varepsilon), P_K(x) - P_K(x, \varepsilon)\rangle \leq \varepsilon,$$

which combined yield

$$\langle x - P_K(x) - x + P_K(x, \varepsilon), P_K(x, \varepsilon) - P_K(x)\rangle \leq \varepsilon$$

or, equivalently,

$$\|P_K(x) - P_K(x, \varepsilon)\| \leq \sqrt{\varepsilon}.$$

Therefore, using the triangle inequality and Fact 2.27, we have

$$\|P_K(x, \varepsilon) - P_K(y)\| \leq \|P_K(x, \varepsilon) - P_K(x)\| + \|P_K(x) - P_K(y)\| \leq \sqrt{\varepsilon} + \|x - y\|.$$

□

Finally, we discuss *relaxed projections*, which are the averages of a point and its projection.

Definition 4.6 (Relaxed projection)**.** Let A be a nonempty closed subset of $\mathbb{R}^n$, $x \in \mathbb{R}^n$ and $\lambda \in (0, 1)$. $\hat{P}_A(x; \lambda)$ is a λ-relaxed projection of x on A if

$$\hat{P}_A(x; \lambda) = (1 - \lambda)x + \lambda P_A(x).$$

One could also present a similar definition with $\lambda \in [0,1]$, in which case we would have that $\hat{P}_A(x;1) = P_A(x)$ and $\hat{P}_A(x;0) = x$. The notion of relaxed projection can also be expanded to accept $\lambda \in (1,2)$, in which case the relaxed projections are averages of the projection and its associated reflection. In this extended definition, relaxed projections with $\lambda \in (0,1)$ are called *underrelaxed projections*, while those associated with $\lambda \in (1,2)$ are called *overrelaxed projections*. With this in mind, unless stated otherwise, by relaxed projections we mean underrelaxed projections (as defined in Definition 4.6).

Unlike the outer-approximate and ε-projections, which can replace exact projections when they are unavailable, relaxed projections rely on the existence of exact projections to be determined, thus being unfit for this role. Their main utility comes when replacing exact projections in situations where convergence may be faster or larger regions of convergence can be obtained from not taking a full projection step, such as in the Method of Alternating Relaxed Projections (MARP) [17], which employs a scheme similar to MAP's, though replacing the exact projections with relaxed ones.

Relaxed projections can also be seen as a special case of an outer-approximate projection, as proved below.

Proposition 4.7 (Relaxed projection is an outer-approximate projection)**.** *Let $B \subseteq \mathbb{R}^n$ be a nonempty set and let $X \subseteq \mathbb{R}^n$ be a closed convex set. Let $x \in B$, $\lambda \in (0,1)$ and $\hat{P}_X(x;\lambda)$ be as defined in Definition 4.6. Then $\hat{P}_X(x;\lambda)$ is an outer-approximate projection, as defined in Definition 4.1.*

Proof. First, let's prove that $\hat{P}_X(x;\lambda)$ satisfies Definition 4.1(i). Using the triangular inequality and the fact that $P_X(x) \in X$, we have that

$$\begin{aligned}
\operatorname{dist}^2(\hat{P}_X(x;\lambda), X) &\leq \|\hat{P}_X(x;\lambda) - P_X(x)\|^2 + \operatorname{dist}(P_X(x), X) \\
&= \|(1-\lambda)x - \lambda P_X(x) - x\|^2 \\
&= (1-\lambda)^2 \|x - P_X(x)\|,
\end{aligned}$$

which implies that

$$\operatorname{dist}(\hat{P}_X(x;\lambda), X) \leq (1-\lambda)\operatorname{dist}(x, X).$$

Since $(1-\lambda) \in (0,1)$, letting $\gamma = 1-\lambda$ establishes (i) is satisfied.

To prove Definition 4.1(ii), from Definition 4.6 and Fact 2.9, we have

$$\begin{aligned}
\left\langle x - \hat{P}(x;\lambda), y - \hat{P}(x;\lambda) \right\rangle &= \langle \lambda x - \lambda P_X(x), y - (1-\lambda)x - \lambda P_X(x) \rangle \\
&= \lambda \langle x - P_X(x), y - P_X(x) \rangle + \lambda \langle x - P_X(x), -(1-\lambda) - (1-\lambda)P_X(x) \rangle \\
&= \lambda \underbrace{\langle x - P_X(x), y - P_X(x) \rangle}_{\leq 0} + \underbrace{\lambda(\lambda-1)}_{<0} \|x - P_X(x)\|^2 \leq 0,
\end{aligned}$$

thus proving (ii) is satisfied.

Since (i) and (ii) are satisfied, the statement is proved. □

While these alternative definitions are interesting and useful in the adequate contexts, the properties and the existence of an algorithm for their computation made the outer-approximate projections the more well-suited approximation for our intended objective.

In the next section, we introduce our main algorithm, which modifies CRM Section 3.4 by employing outer-approximate projections (and reflections) onto the convex set in a product space environment, while the projections and reflections onto the affine subspace (which have closed form) remain exact.

5 Circumcentered Approximate Projection Method - CARM

In this section, we define the Circumcentered-Approximate Reflection Method (CARM) and show that it solves CFP-red (3.2), that is,

$$\text{find } z^* \in K \cap U,$$

with $K \subset \mathbb{R}^n$ a closed convex set, $U \subset \mathbb{R}^n$ an affine subspace and $K \cap U \neq \emptyset$.

Let us assume that an outer-approximate projector $\tilde{P}_K$ onto K is available. According to Definition 4.1, we have a fixed scalar $\gamma \in [0, 1)$ and a set $B \subset \mathbb{R}^n$. We will consider B as $\mathbb{R}^n$. Thus, the correspondent *outer-approximate reflection* $\tilde{R}_K : \mathbb{R}^n \to \mathbb{R}^n$ is given by $\tilde{R}_K = 2\tilde{P}_K - \mathrm{Id}$.

CARM initializes in U and along its iterates employs outer-approximate reflections through K and exact reflections through U. In fact, CARM iterates as

$$z^{k+1} = \tilde{C}\left(z^k\right) := \operatorname{circ}\left\{z^k, \tilde{R}_K\left(z^k\right), R_U\tilde{R}_K\left(z^k\right)\right\}, \text{ with } z^k \in U.$$

Recall that the circumcenter $\tilde{C}(z^k)$ is the point equidistant to $z^k, \tilde{R}_K\left(z^k\right)$ and $R_U\tilde{R}_K\left(z^k\right)$ lying in $\operatorname{aff}\{z^k, \tilde{R}_K\left(z^k\right), R_U\tilde{R}_K\left(z^k\right)\}$; see Definition 2.50. We refer to $\tilde{C}(\cdot)$ as the *outer-approximate circumcenter operator*. The consistency of the iteration above will be proved in the sense that, when $z^0 \in U$, the sequence is well-defined and remains in U.

Connections between CARM and CRM make this section's structure similar to [22, Section 2], though the outer-approximate projections and reflections have to be carefully incorporated in the new analysis; indeed, CRM is a special case of CARM when the outer-approximate projections employed are the exact orthogonal projections (that is, when $\gamma = 0$).

Our first results concern the well-definedness of the outer-approximate circumcenter operator in U and the characterization of the domain. We are going to see that

$$U \subset \operatorname{dom}(\tilde{C}) := \left\{z \in \mathbb{R}^n \mid \tilde{C}(z) = \operatorname{circ}\left\{z, \tilde{R}_K(z), R_U\tilde{R}_K(z)\right\}\right\}$$

and $\tilde{C}(z) \in U$, whenever $z \in U$.

Lemma 5.1 (Well-definedness and characterization of the CARM iteration)**.** *Let $K, U \subset \mathbb{R}^n$, where K is a closed convex set and U is an affine subspace and assume $K \cap U \neq \emptyset$. Then, for all $z \in U$, the CARM iteration $\tilde{C}(z) := \operatorname{circ}\left\{z, \tilde{R}_K(z), R_U\tilde{R}_K(z)\right\}$ is well-defined and we have $\tilde{C}(z) \in U$. Furthermore, $\tilde{C}(z) = P_{\tilde{H}_z \cap U}(z)$, where $\tilde{H}_z := \left\{x \in \mathbb{R}^n \mid \left\langle x - \tilde{P}_K(z), z - \tilde{P}_K(z)\right\rangle = 0\right\}$ if $z \notin K$ and $\tilde{H}_z := K$, otherwise.*

Proof. Let $z \in U$. If $z \in K$, the result is trivial. Suppose then that $z \in U \setminus K$. It follows that $P_{\tilde{H}_z}(z) = \tilde{P}_K(z)$ and, consequently, $R_{\tilde{H}_z}(z) = \tilde{R}_K(z)$. Note that $\tilde{H}_z \cap U \neq \emptyset$.

Let $y \in K \cap U$. $y \in K$ and Definition 4.1(ii) imply $\left\langle z - \tilde{P}_K(z), y - \tilde{P}_K(z) \right\rangle \leq 0$. Define $f : [0,1] \to \mathbb{R}$ by $f(t) := \left\langle z - \tilde{P}_K(z), tz + (1-t)y - \tilde{P}_K(z) \right\rangle$. Then

$$f(0) = \left\langle z - \tilde{P}_K(z), y - \tilde{P}_K(z) \right\rangle \leq 0, \text{ and}$$
$$f(1) = \left\langle z - \tilde{P}_K(z), z - \tilde{P}_K(z) \right\rangle = \left\| z - \tilde{P}_K(z) \right\|^2 > 0.$$

Since f is continuous, there exists a scalar $\bar{t} \in [0,1)$ such that

$$f(\bar{t}) = \left\langle z - \tilde{P}_K(z), \bar{t}z + (1-\bar{t})y - \tilde{P}_K(z) \right\rangle = 0,$$

implying $\bar{t}z + (1-\bar{t})y \in \tilde{H}_z$.

On the other hand, since U is an affine subspace and $z, y \in U$, we have $\bar{t}z + (1-\bar{t})y \in U$ as well. Altogether, $\bar{t}z + (1-\bar{t})y \in \tilde{H}_z \cap U$. Therefore,

$$\tilde{C}(z) = \operatorname{circ}\left\{z, \tilde{R}_K(z), R_U \tilde{R}_K(z)\right\} = \operatorname{circ}\left\{z, R_{\tilde{H}_z}(z), R_U R_{\tilde{H}_z}(z)\right\} = P_{\tilde{H}_z \cap U}(z),$$

where the last equality follows by employing Fact 3.16. □

Next, we characterize the fixed point set of the outer-approximate circumcenter operator, as well as the fixed point sets of the outer-approximate projectors and reflectors.

Lemma 5.2 (Fixed points of $\tilde{P}_K, \tilde{R}_K$ and $\tilde{C}$)**.** *Assume $K, U \subset \mathbb{R}^n$ as in Lemma 5.1 and consider* $\operatorname{Fix} \tilde{P}_K := \{z \in \mathbb{R}^n \mid \tilde{P}_K(z) = z\}$, $\operatorname{Fix} \tilde{R}_K := \{z \in \mathbb{R}^n \mid \tilde{R}_K(z) = z\}$ *and* $\operatorname{Fix} \tilde{C} := \{z \in \operatorname{dom}(\tilde{C}) \mid \tilde{C}(z) = z\}$. *Then,*

(i) $\operatorname{Fix} \tilde{P}_K = \operatorname{Fix} \tilde{R}_K = K$;

(ii) $\operatorname{Fix} \tilde{C} = K \cap U$.

Proof. (i) Note that $z = \tilde{R}_K(z) = 2\tilde{P}_K(z) - z$ implies that $z = \tilde{P}_K(z)$. Therefore, $\operatorname{Fix} \tilde{P}_K = \operatorname{Fix} \tilde{R}_K$. Clearly, if $z \in K$ then $z \in \operatorname{Fix} \tilde{P}_K$. Conversely, take $z \in \operatorname{Fix} \tilde{P}_K$, i.e., $\tilde{P}_K(z) = z$. By Definition 4.1(i), we have that $\operatorname{dist}(\tilde{P}_k(z), K) \leq \gamma \operatorname{dist}(z, K)$, where $\gamma \in [0,1)$. If $\gamma = 0$, then $z = \tilde{P}_K(z) = P_K(z)$ and $z \in K$ trivially. Otherwise, if $\gamma \in (0,1)$, from $z = \tilde{P}_K(z)$ it follows that $\operatorname{dist}(z, K) \leq \gamma \operatorname{dist}(z, K)$. This is only true if $\operatorname{dist}(z, K) = 0$, which means that $z \in K$. Therefore, that $\operatorname{Fix} \tilde{P}_K = K$.

(ii) It is easy to see that, if $z \in K \cap U$ then $z \in \operatorname{Fix} \tilde{R}_K \cap \operatorname{Fix} R_U$ and thus $z \in \operatorname{Fix} \tilde{C}$. On the other hand, take $z \in \operatorname{Fix} \tilde{C}$. By the definition of $\tilde{C}(z)$, we have that $\tilde{C}(z) = z$ implies $z = \tilde{R}_K(z) = R_U \tilde{R}_K(z)$. Since $\operatorname{Fix} \tilde{R}_K = K$ by (i), $z = \tilde{R}_K(z)$ implies $z \in K$. Then, $z = R_U(z)$, which yields that $z \in U$. Therefore, $z \in K \cap U$.

□

Note that the outer-approximate projection parameter γ being strictly less than 1 in Definition 4.1 is key to deriving the previous result.

We now derive the firm quasinonexpansiveness property (Definition 2.23(iv)) of the outer-approximate circumcenter operator $\tilde{C}$ restricted to U.

Lemma 5.3 (CARM is firmly quasinonexpansive)**.** *Assume $K, U \subset \mathbb{R}^n$ as in Lemma 5.1. Then, for any $z \in U$ and $s \in K \cap U$*

$$\|\tilde{C}(z) - s\|^2 \leq \|z - s\|^2 - \|z - \tilde{C}(z)\|^2. \tag{5.1}$$

Moreover,

$$\operatorname{dist}^2(\tilde{C}(z), K \cap U) \leq \operatorname{dist}^2(z, K \cap U) - \|z - \tilde{C}(z)\|^2.$$

Proof. Let $z \in U$. Define $\tilde{H}_z := \left\{h \in \mathbb{R}^n \mid \left\langle h - \tilde{P}_K(z), z - \tilde{P}_K(z)\right\rangle = 0\right\}$ and let $\tilde{H}_z^+$ be the halfspace generated by $\tilde{H}_z$ such that $z \notin \operatorname{int} \tilde{H}_z^+$.

Clearly, $K \subseteq \tilde{H}_z^+$, thus $K \cap U \subseteq \tilde{H}_z^+ \cap U$. If $z \in \tilde{H}_z$, then $z \in K \cap U$ and, by Lemma 5.2, $\tilde{C}(z) = z$, from which the result easily follows.

Let $z \notin \tilde{H}_z$. Since $\tilde{R}_K(z) = R_{\tilde{H}_z}(z)$, $\tilde{C}(z) = \operatorname{circ}\left\{z, R_{\tilde{H}_z}(z), R_U R_{\tilde{H}_z}(z)\right\}$. From Fact 3.16, we have that $\tilde{C}(z) = P_{\tilde{H}_z \cap U}(z) \in U$. Note that $z \notin \tilde{H}_z$ implies that $z \notin K$. It also easily follows that $z \notin \tilde{H}_z^+$, which combined with the fact that $\tilde{H}_z$ is the boundary of $\tilde{H}_z^+$ yield that $P_{\tilde{H}_z \cap U}(z) = P_{\tilde{H}_z^+ \cap U}(z)$.

Considering that for any $s \in K \cap U, P_{\tilde{H}_z^+ \cap U}(s) = s$, we have

$$\begin{aligned}
\|\tilde{C}(z) - s\|^2 &= \left\|P_{\tilde{H}_z \cap U}(z) - s\right\|^2 \\
&= \left\|P_{\tilde{H}_z^+ \cap U}(z) - P_{\tilde{H}_z^+ \cap U}(s)\right\|^2 \\
&\leq \|z - s\|^2 - \left\|\left[z - P_{\tilde{H}_z^+ \cap U}(z)\right] - \left[s - P_{\tilde{H}_z^+ \cap U}(s)\right]\right\|^2 \\
&= \|z - s\|^2 - \|z - \tilde{C}(z)\|^2,
\end{aligned}$$

where we use the firm nonexpansiveness of the projection (Fact 2.27) to obtain the inequality.

Finally, plug $P_{K \cap U}(z)$ in for s in (5.1) so as to get

$$\begin{aligned}
\operatorname{dist}^2(C(z), K \cap U) &\leq \|\tilde{C}(z) - P_{K \cap U}(z)\|^2 \leq \|z - P_{K \cap U}(z)\|^2 - \|z - \tilde{C}(z)\|^2 \\
&= \operatorname{dist}^2(z, K \cap U) - \|z - \tilde{C}(z)\|^2.
\end{aligned}$$

Then, $\tilde{C}$ is firmly nonexpansive. □

We can now establish the central theorem of this work, stating convergence of CARM. Although the proof follows some lines of [22, Theorem 1], one has to be careful when dealing with the fixed points of the outer-approximate reflection operator $\tilde{R}_K$. In this regard, Lemma 5.2 plays a special role.

Theorem 5.4 (Convergence of CARM)**.** *Assume $K, U \subset R^n$ as in Lemma 2 and let $z \in U$ be given. Then, the CARM sequence $\left(\tilde{C}^k(z)\right)_{k\in\mathbb{N}}$ is well-defined, contained in U and converges to a point in $K \cap U$.*

Proof. Lemma 5.1 gives us that $\left(\tilde{C}^k(z)\right)_{k\in\mathbb{N}}$ is well-defined and belongs to U. From Lemma 5.3, we have, for any $z \in U, s \in K \cap U$ and $\ell \in \mathbb{N}$

$$\begin{aligned}
\left\|z^{\ell+1} - s\right\|^2 &= \left\|\tilde{C}^{\ell+1}(z) - s\right\|^2 \\
&= \left\|\tilde{C}\left(\tilde{C}^{\ell}(z)\right) - s\right\|^2 \\
&\leq \left\|\tilde{C}^{\ell}(z) - s\right\|^2 - \left\|\tilde{C}^{\ell}(z) - \tilde{C}\left(\tilde{C}^{\ell}(z)\right)\right\|^2 \\
&= \left\|\tilde{C}^{\ell}(z) - s\right\|^2 - \left\|\tilde{C}^{\ell}(z) - \tilde{C}^{\ell+1}(z)\right\|^2 \\
&= \left\|z^{\ell} - s\right\|^2 - \left\|z^{\ell} - z^{\ell+1}\right\|^2.
\end{aligned}$$

Hence, for all $\ell \in \mathbb{N}$

$$\left\|z^{\ell} - z^{\ell+1}\right\|^2 \leq \left\|z^{\ell} - s\right\|^2 - \left\|z^{\ell+1} - s\right\|^2.$$

Summing from $\ell = 0$ to m, we have

$$\sum_{\ell=0}^{m} \left\|z^{\ell} - z^{\ell+1}\right\|^2 \leq \sum_{\ell=0}^{m} \left(\left\|z^{\ell} - s\right\|^2 - \left\|z^{\ell+1} - s\right\|^2\right) = \left\|z^0 - s\right\|^2 - \left\|z^{m+1} - s\right\|^2 \leq \left\|z^0 - s\right\|^2.$$

Taking limits as $m \to +\infty$, we get the summability of the associated series and so, $z^k - z^{k+1}$ converges to 0 as $k \to +\infty$.

Moreover, by (5.1) in Lemma 5.3, the sequence $\left(z^k\right)_{k\in\mathbb{N}}$ is bounded. That is,

$$\left\|z^k - s\right\| = \left\|\tilde{C}^k(z) - s\right\| \leq \left\|\tilde{C}^{k-1}(z) - s\right\| \leq \cdots \leq \|z - s\|.$$

This also shows that the sequence is Fejér monotone with respect to $K \cap U$.

Let $\hat{z}$ be any cluster point of the sequence $\left(z^k\right)_{k\in\mathbb{N}}$ and denote $(z^{i_k})_{k\in\mathbb{N}}$ an associated convergent subsequence to $\hat{z}$. Note further that the fact $z^{i_k} - z^{i_k+1} \to 0$ implies $z^{i_k+1} \to \hat{z}$. We claim that $\hat{z} \in K \cap U$. since U is closed and $\left(z^k\right)_{k\in\mathbb{N}}$ is contained in U, we must have $\hat{z} \in U$. By the definition of X we have that

$$\left\|z^{i_k} - z^{i_k+1}\right\| = \left\|z^{i_k} - \tilde{C}\left(z^{i_k}\right)\right\| = \left\|\tilde{C}\left(z^{i_k}\right) - \tilde{R}_K\left(z^{i_k}\right)\right\| = \left\|z^{i_k+1} - \tilde{R}_K\left(z^{i_k}\right)\right\|.$$

So, $z^{ik+1} - \tilde{R}_K\left(z^{ik}\right)$ converges to 0. It follows from the continuity of the reflection onto K and taking limits as $k \to +\infty$ in the last equality that $\hat{z} = \tilde{R}_K(\hat{z})$, that is, $\hat{z} \in \operatorname{Fix} \tilde{R}_K$. From Lemma 5.2(i), $\hat{z} \in K$, proving the claim.

Finally, we have that $\left(z^k\right)_{k\in\mathbb{N}}$ is bounded, contained in U, Fejér monotone with respect to $K \cap U$ and all its cluster points are in $K \cap U$. Applying Fact 2.34 establishes the desired result. □

5.1 CARM for general convex intersection

Now, we'll show that CARM can be used for finding a point in the intersection of finitely many closed convex sets using Pierra's product space reformulation [59]. Consider CFP-red (3.2), that is,

$$\text{find } z^* \in K \cap U,$$

where $K = K_1 \times K_2 \times \cdots \times K_m \subset \mathbb{R}^{nm}$ is the Cartesian product of m closed and convex sets $K_i \in \mathbb{R}^n$ and $U := \{(x, x, \ldots, x) \in \mathbb{R}^{nm} \mid x \in \mathbb{R}^n\}$ is the diagonal set, which can be easily shown to be a subspace of $\mathbb{R}^{nm}$.

Considering m arbitrary vectors $x^{(i)}$ in $\mathbb{R}^n$, with $i = 1, \ldots, m$, we can build an arbitrary point $z = \left(x^{(1)}, x^{(2)}, \ldots, x^{(m)}\right) \in \mathbb{R}^{nm}$. The projection onto U is given by

$$P_U(z) = \frac{1}{m}\left(\sum_{i=1}^{m} x^{(i)}, \sum_{i=1}^{m} x^{(i)}, \ldots, \sum_{i=1}^{m} x^{(i)}\right)$$

The following proposition establishes an outer-approximate projection for K, following the properties given in Definition 4.1.

Proposition 5.5 (Outer-approximate projection on product space)**.** *Assume K and U as above. Let $z = \left(x^{(1)}, x^{(2)}, \ldots, x^{(m)}\right) \in \mathbb{R}^{nm}$ be an arbitrary point in $\mathbb{R}^{nm}$ given by m arbitrary vectors $x^{(i)} \in K_i \subset \mathbb{R}^n$. For each set K_i, let $\tilde{P}_{K_i}$ be an approximate projector associated with the parameter $\gamma_i \in [0, 1)$ as per Definition 4.1(i) and let $\gamma_{\max} := \max\{\gamma_1, \ldots, \gamma_m\}$. Then,*

$$\tilde{P}_K(z) = (\tilde{P}_{K_1}(x^{(1)}), \ldots, \tilde{P}_{K_m}(x^{(m)}))$$

is an outer-approximate projection that satisfies Definition 4.1 with the associated parameter $\gamma_{\max}$.

Proof. (i) Note that, from the definition of distance and Definition 4.1(i), we have

$$\begin{aligned}
\operatorname{dist}^2(\tilde{P}_K(z), K) &= \sum_{i=1}^{m} \operatorname{dist}^2(\tilde{P}_{K_i}(x^{(i)}), K_i) \\
&\leq \sum_{i=1}^{m} \gamma_i^2 \operatorname{dist}^2(x^{(i)}, K_i) \\
&\leq \gamma_{max}^2 \sum_{i=1}^{m} \operatorname{dist}^2(x^{(i)}, K_i) = \gamma_{max}^2 \operatorname{dist}^2(z, K),
\end{aligned}$$

Thus, we have that $\operatorname{dist}(\tilde{P}_K(z), K) \leq \gamma_{max} \operatorname{dist}(z, K)$, satisfying property (i).

(ii) Let $y = (y^{(1)}, \ldots, y^{(m)}) \in \mathbb{R}^{nm}$ be an arbitrary point in $\mathbb{R}^{nm}$, where $y^{(i)} \in K_i \subset \mathbb{R}^n$. Following the definition of inner product, we have that

$$\left\langle z - \tilde{P}_K(z), y - \tilde{P}_K(z) \right\rangle = \sum_{i=1}^{m} \left\langle x^{(i)} - \tilde{P}_K\left(x^{(i)}\right), y^{(i)} - \tilde{P}_K\left(x^{(i)}\right) \right\rangle \leq 0,$$

in which we use the fact that, by Definition 4.1(ii), the m terms of the sum are lesser or equal than 0. Thus, property (ii) is also satisfied.

□

Following this proposition, we can establish convergence of CARM for general convex intersection.

Theorem 5.6 (Convergence of CARM for general convex intersection)**.** *Assume $K, U \subset \mathbb{R}^n$ as above and let $x^0 \in \mathbb{R}^n$ be given. Then, taking as initial point $z^0 := (x^0, x^0, \ldots, x^0) \in U$, we have that the sequence $\left\{z^k\right\} \subset \mathbb{R}^{nm}$ generated by*

$$z^{k+1} := \operatorname{circ}\left\{z^k, \tilde{R}_K\left(z^k\right), R_U \tilde{R}_K\left(z^k\right)\right\}$$

is well defined, entirely contained in U and converges to a point $z^* = (x^*, x^*, \ldots, x^*) \in \mathbb{R}^{nm}$, where $x^* \in \bigcap_{i=1}^m K_i$.

Proof. The result follows immediately from Theorem 5.4, with $\tilde{R}_K$ being given by $\tilde{P}_K$ as defined in Proposition 5.5. □

We have come to the favorable conclusion that CARM converges to a solution of (3.2) when initializing in U, which generalizes Theorems 1 and 3 from [22]. Following this line of generalizing results from [22], we now look at alternating projections employing outer-approximate projections onto one of the sets.

6 Approximate Method of Alternating Projections - AMAP

In this section, the problem under consideration is given by

$$\text{Find } z^* \in K_1 \cap K_2$$

with K_1, K_2 being closed convex sets and $K_1 \cap K_2 \neq \emptyset$. Of course, this problem is more general than CFP-red ((3.2)), as a subspace U of $\mathbb{R}^n$ is closed and convex.

We are going to establish convergence of AMAP for finding a point in the intersection of the two closed convex sets K_1 and K_2. We understand that such a result can be derived from known frameworks for inexact alternating projections; nonetheless, we present it for the sake of completeness. Other approximation techniques concerning MAP can be found, for instance, in [55, 56].

We will initialize AMAP on K_2 and replace MAP's orthogonal projections onto K_1 by outer-approximate ones. Let us describe AMAP formally. For a given point $z \in \mathbb{R}^n$, we generate a sequence $\left(x^k\right)_{k\in\mathbb{N}}$ contained in K_2 such that $x^0 := P_{K_2}(z)$ and

$$x^{k+1} = P_{K_2}\tilde{P}_{K_1}\left(x^k\right), \quad k = 0, 1, 2, \ldots$$

MAP is a special case of AMAP and we are able to carry out a convergence analysis for AMAP resembling MAP's. We start the analysis with a Fejér monotonicity statement.

Proposition 6.1 (Fejér monotonicity of AMAP sequence)**.** *Suppose $K_1, K_2 \subseteq \mathbb{R}^n$ are closed convex sets and $K_1 \cap K_2 \neq \emptyset$. Let $x^0 \in K_2$ be any point in K_2. Then the sequences $\left(x^k\right)_{k\in\mathbb{N}}$ and $\left(\tilde{P}_{K_1}\left(x^k\right)\right)_{k\in\mathbb{N}}$ generated by the AMAP algorithm is Fejér monotone with respect to $K_1 \cap K_2$.*

Proof. Let $\bar{x}$ be any point in the intersection $K_1 \cap K_2$, and let $k \in \mathbb{N}$. From the definition of outer-approximate projection, it easily follows that

$$\left\|\tilde{P}_{K_1}\left(x^k\right) - \bar{x}\right\|^2 \leq \left\|x^k - \bar{x}\right\|^2 - \left\|x^k - \tilde{P}_{K_1}\left(x^k\right)\right\|^2 \leq \left\|x^k - \bar{x}\right\|^2. \tag{6.1}$$

Similarly, we also have that

$$\left\|x^{k+1} - \bar{x}\right\|^2 \leq \left\|\tilde{P}_{K_1}\left(x^k\right) - \bar{x}\right\|^2 - \left\|\tilde{P}_{K_1}\left(x^k\right) - x^{k+1}\right\|^2 \leq \left\|\tilde{P}_{K_1}\left(x^k\right) - \bar{x}\right\|^2. \tag{6.2}$$

From (6.1) and (6.2), it follows that $\left\|x^{k+1} - \bar{x}\right\| \leq \left\|x^k - \bar{x}\right\|$ and $\left\|\tilde{P}_{K_2}\left(x^{k+1}\right) - \bar{x}\right\| \leq \left\|\tilde{P}_{K_2}\left(x^k\right) - \bar{x}\right\|$.

Since k and $\bar{x}$ are arbitrary, the sequences $\left(x^k\right)_{k\in\mathbb{N}}$ and $\left(\tilde{P}_{K_1}\left(x^k\right)\right)_{k\in\mathbb{N}}$ are Fejér monotone with respect to $K_1 \cap K_2$. □

The following proposition establishes the domain of the set of the cluster points of the sequence generated by AMAP.

Proposition 6.2 (AMAP cluster points)**.** *Suppose $K_1, K_2 \subseteq \mathbb{R}^n$ are closed convex sets and $K_1 \cap K_2 \neq \emptyset$. Let $x^0 \in K_2$ be any point in K_2. Then every cluster point of the sequence $\left(x^k\right)_{k\in\mathbb{N}}$ generated by the AMAP algorithm belongs to $K_1 \cap K_2$.*

Proof. Since the sequence $\left(x^k\right)_{k\in\mathbb{N}}$ is Fejér monotone with respect to $K_1 \cap K_2$, it is bounded. Therefore, it must have at least one cluster point.

Let x^* be a cluster point for the sequence $\left(x^k\right)_{k\in\mathbb{N}}$. Since K_2 is closed, and $x^k \in K_1$ $\forall k \in \mathbb{N}$, then $x^* \in K_2$.

Rearranging (6.1) and (6.2), we have that

$$\left\|\tilde{P}_{K_1}\left(x^k\right) - x^k\right\| \leq \left\|x^k - \bar{x}\right\|^2 - \left\|\tilde{P}_{K_1}\left(x^k\right) - \bar{x}\right\|^2, \text{ and} \tag{6.3}$$

$$\left\|x^{k+1} - \tilde{P}_{K_1}\left(x^k\right)\right\| \leq \left\|\tilde{P}_{K_1}\left(x^k\right) - \bar{x}\right\|^2 - \left\|x^{k+1} - \bar{x}\right\|^2. \tag{6.4}$$

From (6.1) and (6.2), we find that the sequence

$$\left\|x^0 - \bar{x}\right\|, \left\|\tilde{P}_{K_1}\left(x^0\right) - \bar{x}\right\|, \left\|x^1 - \bar{x}\right\|, \left\|\tilde{P}_{K_1}\left(x^1\right) - \bar{x}\right\|, \ldots$$

is decreasing; thus, it converges. Therefore, taking the limit on the left side of (6.3) and (6.4) yields that $\left\|\tilde{P}_{K_1}\left(x^k\right) - x^k\right\|$ and $\left\|x^{k+1} - \tilde{P}_{K_1}\left(x^k\right)\right\|$ converge to zero.

Note that

$$\begin{aligned}
\operatorname{dist}(x^k, K_1) &= \left\|P_{K_1}(x^k) - x^k\right\| \\
&= \left\|\tilde{P}_{K_1}\left(x^k\right) - x^k\right\| + \operatorname{dist}(\tilde{P}_{K_1}\left(x^k\right), K_1) \\
&\leq \left\|\tilde{P}_{K_1}\left(x^k\right) - x^k\right\| + \gamma \operatorname{dist}(x^k, K_1) \\
&\leq \frac{1}{1-\gamma}\left\|\tilde{P}_{K_1}\left(x^k\right) - x^k\right\|,
\end{aligned}$$

where $\gamma \in [0, 1)$ is the outer-approximate projection parameter associated to K_1. From the fact that $\left\|\tilde{P}_{K_1}\left(x^k\right) - x^k\right\| \to 0$, we must have that $\operatorname{dist}(x^k, K_1) \to 0$.

Since a subsequence of $(x_k)_{k\in\mathbb{N}}$ converges to x^* and $\operatorname{dist}(x^k, K_1) \to 0$, x^* must be a cluster point of K_1. Thus, $x^* \in K_1$, since K_1 is closed and therefore must contain all of its cluster points. Hence, we conclude that $x^* \in K_1 \cap K_2$. □

Following the previous results, we can establish our main convergence result for AMAP.

Theorem 6.3 (AMAP convergence)**.** *Suppose $K_1, K_2 \subseteq \mathbb{R}^n$ are closed convex sets and $K_1 \cap K_2 \neq \emptyset$. Let $x^0 \in K_2$ be any point in K_2. Then the sequence $\left(x^k\right)_{k\in\mathbb{N}}$ generated by the AMAP algorithm converge to a point $x^* \in K_1 \cap K_2$.*

Proof. From Propositions 6.1 and 6.2, we have that $\left(x^k\right)_{k\in\mathbb{N}}$ is Fejér monotone with respect to $K_1 \cap K_2$ and that every cluster point of $\left(x^k\right)_{k\in\mathbb{N}}$ belongs to $K_1 \cap K_2$. The result follows from applying Fact 2.34. $\square$

7 Comparing CARM and AMAP

In this section, we compare CARM and AMAP for finding a point in the intersection of a closed convex set K and an affine subspace U, both belonging to $\mathbb{R}^n$. Our study is divided in two subsections. In the first, we show that for a point in the affine set, the CARM iteration is always closer to the solution set than the AMAP iteration. In the second subsection, the methods are compared under an error bound assumption. The error bound guarantees that both CARM and AMAP achieve linear convergence, with CARM having a better asymptotic rate than AMAP.

7.1 Myopic comparison

We look at CARM and AMAP acting on an arbitrary point z in the affine set U. As previously done for CRM and MAP in Section 3.4.3, we will present a similar picture, in Figure 7.1, to illustrate how CARM and AMAP compare in a geometrical sense. In this figure, we also feature the point generated by CRM, in order to establish a reference to that method as well.

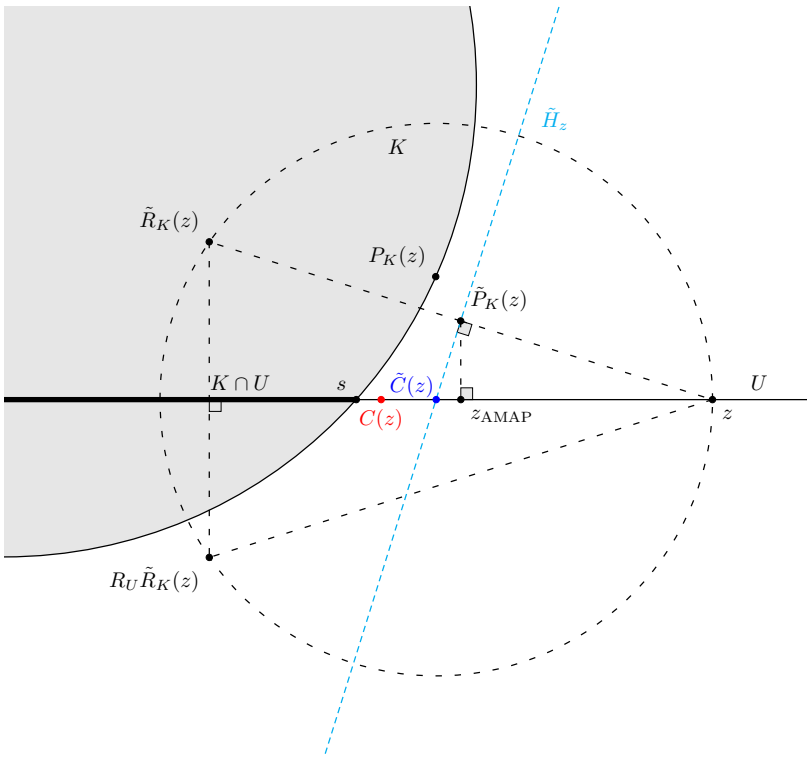

Fig. 7.1: Illustration of CARM and AMAP for the intersection between an affine U and a convex K.

We prove that CARM brings you closer to $K \cap U$ than AMAP. This result generalizes [22, Theorem 2] (excluding DRM, due to the lack of an equivalent outer-approximate projection-based variant) and is stated formally in the following theorem.

Theorem 7.1 (CARM versus AMAP)**.** *Assume $K, U \subset \mathbb{R}^n$ as in Lemma 5.1 and let $z \in U$ be given. Also recall the notation $z_{\text{AMAP}} := P_U \tilde{P}_K(z)$ and $\tilde{C}(z) := \operatorname{circ}\left\{z, \tilde{R}_K(z), R_U \tilde{R}_K(z)\right\}$. Then, for any $s \in K \cap U$ we have*

(i) $\|\tilde{C}(z) - s\| \leq \|z_{\text{AMAP}} - s\|$;

(ii) $\operatorname{dist}(\tilde{C}(z), K \cap U) \leq \operatorname{dist}(z_{\text{AMAP}}, K \cap U)$.

Proof. Assume $z \in U$ and let $s \in K \cap U$ be arbitrary but fixed. If $z \in K$ the result follows trivially, because then $P_U \tilde{P}_K(z) = z$, $\frac{1}{2}\left(\operatorname{Id} + R_U \tilde{R}_K\right)(z) = z$ and $\tilde{C}(z) = z$ due to Lemma 5.2. So, let $z \in U \backslash K$. From Lemma 5.1, we have that $\tilde{C}(z)$ is the projection of z onto the intersection of U and the hyperplane with normal $z - \tilde{P}_K(z)$ passing through $\tilde{P}_K(z)$. Therefore, the triangles $\triangle[z, \tilde{P}_K(z), \tilde{C}(z)]$ and $\triangle[z, P_U \tilde{P}_K(z), \tilde{P}_K(z)]$ have right angles $\angle[z, \tilde{P}_K(z), \tilde{C}(z)]$ and $\angle[z, P_U \tilde{P}_K(z), \tilde{P}_K(z)]$ (the latter since U is a subspace), respectively. Hence,

$$\|\tilde{C}(z) - z\| \geq \left\|\tilde{P}_K(z) - z\right\| \geq \left\|P_U \tilde{P}_K(z) - z\right\|. \tag{7.1}$$

Now, note that $P_U \tilde{P}_K(z)$, z and $\tilde{C}(z)$ are collinear, since both $P_U \tilde{P}_K(z)$ and $\tilde{C}(z)$ lie in the semi-line starting at z and passing through $P_U \tilde{R}_K(z) = \frac{1}{2}\left(\tilde{R}_K(z) + R_U \tilde{R}_K(z)\right)$. In fact, this semi-line contains the circumcenter $\tilde{C}(z)$ as it is a bisector of the isosceles triangle $\triangle[z, \tilde{R}_K(z), R_U \tilde{R}_K(z)]$. Since P_U is a linear operator Fact 2.10 $z \in U$, we have that

$$\begin{aligned} z_{\text{AMAP}} = P_U \tilde{P}_K(z) &= P_U\left(\frac{1}{2}\left(z + \tilde{R}_K(z)\right)\right) = \frac{1}{2} P_U\left(z + \tilde{R}_K(z)\right) \\ &= \frac{1}{2}\left(P_U(z) + P_U \tilde{R}_K(z)\right) = \frac{1}{2}\left(z + P_U \tilde{R}_K(z)\right) \end{aligned}$$

Thus, z_{AMAP} is a convex combination of z and $P_U \tilde{R}_K(z)$ with parameter $\frac{1}{2}$. Therefore, we have more than just collinearity of $z, \tilde{C}(z)$ and z_{AMAP}. Both $\tilde{C}(z)$ and z_{AMAP} lie on the semi-line starting at z and passing through $P_U \tilde{R}_K(z)$. This means that z cannot lie between $\tilde{C}(z)$ and z_{AMAP} and, in view of (7.1), the only remaining possibility is that z_{AMAP} is a convex combination of $\tilde{C}(z)$ and z, that is, there exists a parameter $r \in [0, 1]$ such that

$$z_{\text{AMAP}} = r\tilde{C}(z) + (1 - r)z$$

and

$$z_{\text{AMAP}} - \tilde{C}(z) = (1 - r)(z - \tilde{C}(z)) \text{ and } z - z_{\text{AMAP}} = r(z - \tilde{C}(z)).$$

r must be strictly larger than zero; otherwise z would be in $K \cap U$. Thus,

$$z_{\text{AMAP}} - C(z) = \frac{1-r}{r}\left(z - z_{\text{AMAP}}\right) \tag{7.2}$$

This equation will be properly combined with the following inner product manipulations

$$\begin{aligned}
\left\langle z - z_{\text{AMAP}}, \tilde{C}(z) - s \right\rangle &= \left\langle z - \tilde{P}_K(z), \tilde{C}(z) - s \right\rangle + \left\langle \tilde{P}_K(z) - z_{\text{AMAP}}, \tilde{C}(z) - s \right\rangle \\
&= \left\langle z - \tilde{P}_K(z), \tilde{C}(z) - \tilde{P}_K(z) \right\rangle + \underbrace{\left\langle z - \tilde{P}_K(z), \tilde{P}_K(z) - s \right\rangle}_{\geq 0} \\
&\quad + \underbrace{\left\langle \tilde{P}_K(z) - z_{\text{AMAP}}, \tilde{C}(z) - s) \right\rangle}_{=0} \\
&\geq 0
\end{aligned} \tag{7.3}$$

The first under-brace equality follows as a consequence of the property outlined in Definition 4.1(ii). The last under-brace remark holds since $\tilde{P}_K(z) - z_{\text{AMAP}}$ is orthogonal to the subspace U and both $\tilde{C}(z)$ and s are in U. Now, inequality (7.3) together with equation (7.2) give us

$$\langle z_{\text{AMAP}} - C(z), s - C(z) \rangle = \frac{1-r}{r} \langle z - z_{\text{AMAP}}, s - C(z) \rangle \leq 0$$

which, due to the cosine rule, leads to $\|\tilde{C}(z) - s\| \leq \|z_{\text{AMAP}} - s\|$ for any $s \in K \cap U$. Taking this into account and letting $\hat{z}, \tilde{z} \in K \cap U$ realize the distance of $\tilde{C}(z), z_{\text{AMAP}}$ to $K \cap U$, respectively, we have

$$\operatorname{dist}(\tilde{C}(z), K \cap U) = \|\tilde{C}(z) - \hat{z}\| \leq \|\tilde{C}(z) - \tilde{z}\| \leq \|z_{\text{AMAP}} - \tilde{z}\| = \operatorname{dist}(z_{\text{AMAP}}, K \cap U)$$

proving the first inequalities in items (i) and (ii). □

7.2 CARM and AMAP under error bound

Now, inspired by the results in [2], we show that CARM achieves better convergence rates than AMAP under an error bound assumption (EB). We are going to prove that the EB condition, defined below, ensures linear convergence of the sequences $\left(z^k\right)_{k\in\mathbb{N}}$ and $\left(x^k\right)_{k\in\mathbb{N}}$ generated by AMAP and CARM, respectively.

Assumption EB (Error bound condition)**.** *Let $K \subset \mathbb{R}^n$ be a closed and convex set and $U \subset \mathbb{R}^n$ an affine subspace. Assume that $K \cap U \neq \emptyset$ and there exists $\omega \in (0,1)$ such that $\operatorname{dist}(x, K) \geq \omega \operatorname{dist}(x, K \cap U)$ for all $x \in U$ and $\operatorname{dist}(y, U) \geq \omega \operatorname{dist}(y, K \cap U)$ for all $y \in K$.*

To begin our convergence analysis, a corollary of the following proposition will prove that, for both methods, the sequences $\left(\operatorname{dist}\left(z^k, K \cap U\right)\right)_{k\in\mathbb{N}}$ and $\left(\operatorname{dist}\left(x^k, K \cap U\right)\right)_{k\in\mathbb{N}}$ both converge linearly to 0.

Proposition 7.2. *Assume that K, U satisfy EB. Consider $\tilde{T}, \tilde{C} : \mathbb{R}^n \to \mathbb{R}^n$ as in $\tilde{T} = P_U \circ \tilde{P}_K$, $\tilde{C}(\cdot) = \operatorname{circ}\left(\cdot, \tilde{R}_K(\cdot), R_U\left(\tilde{R}_K(\cdot)\right)\right)$. Then, for all $x \in U$,*

$$\left(1 - \omega^2(1-\gamma)^2\right) \operatorname{dist}^2(x, K \cap U) \geq \operatorname{dist}^2(\tilde{T}(x), K \cap U) \geq \operatorname{dist}^2(\tilde{C}(x), K \cap U) \tag{7.4}$$

with ω as in Assumption EB.

Proof. Note that

$$\begin{aligned}
\|\tilde{T}(x) - P_{K\cap U}(x)\|^2 &= \left\|P_U\left(\tilde{P}_K(x)\right) - P_{K\cap U}(x)\right\|^2 \\
&\leq \left\|\tilde{P}_K(x) - P_{K\cap U}(x)\right\|^2 - \left\|P_U\left(\tilde{P}_K(x)\right) - \tilde{P}_K(x)\right\|^2 \\
&\leq \left\|\tilde{P}_K(x) - P_{K\cap U}(x)\right\|^2 \leq \|x - P_{K\cap U}(x)\|^2 - \left\|x - \tilde{P}_K(x)\right\|^2 \\
&\leq \|x - P_{K\cap U}(x)\|^2 - (1-\gamma)^2 \operatorname{dist}^2(x, K) \\
&\leq (1 - \omega^2(1-\gamma)^2) \operatorname{dist}^2(x, K \cap U)
\end{aligned} \tag{7.5}$$

for all $x \in U$, where we use the firm quasinonexpansivity of the projection onto U (Definition 2.23(iv)) and the fact that $P_{K\cap U} \in U$ to obtain the first inequality; the firm quasinonexpansivity of the outer-approximate projection onto K (which can easily be derived from Definition 4.1(ii)), the fact that $P_{K\cap U} \in K$, and $K = \operatorname{Fix} \tilde{P}_K$ from Lemma 5.2(i) give us the third inequality; finally, we employ Definition 4.1(i) and Assumption EB to obtain the last two inequalities. From (7.5), we get

$$\begin{aligned}
\left\|\tilde{C}(x) - P_{K\cap U}(\tilde{C}(x))\right\|^2 &\leq \left\|\tilde{C}(x) - P_{K\cap U}(\tilde{T}(x))\right\|^2 \\
&\leq \left\|\tilde{T}(x) - P_{K\cap U}(\tilde{T}(x))\right\|^2 \leq \left\|\tilde{T}(x) - P_{K\cap U}(x)\right\|^2 \\
&\leq \|x - P_{K\cap U}(x)\|^2 - \omega^2(1-\gamma)^2 \operatorname{dist}^2(x, K \cap U) \\
&= \left(1 - \omega^2(1-\gamma)^2\right) \|x - P_{K\cap U}(x)\|^2,
\end{aligned} \tag{7.6}$$

using the definition of $P_{K\cap U}$ in the first and third inequalities and Theorem 7.1(i) in the second inequality. (7.6) follows immediately from (7.4), after rewriting the appropriate terms in terms of distance, in view of the definition of $\tilde{P}_{K\cap U}$. □

With the corollary below, we can establish an upper bound for the asymptotic constants of Q-linear convergence for both CARM and AMAP.

Corollary 7.3 (Upper bound for asymptotic constants of CARM and AMAP under EB)**.** *Let* $\left(z^k\right)_{k\in\mathbb{N}}$ *and* $\left(x^k\right)_{k\in\mathbb{N}}$ *be the sequences generated by AMAP and CARM starting at any* $z^0 \in U$ *and any* $x^0 \in \mathbb{R}^n$*, respectively. If* K, U *satisfy Assumption EB then the sequences* $\left(\operatorname{dist}\left(z^k, K\cap U\right)\right)_{k\in\mathbb{N}}$ *and* $\left(\operatorname{dist}\left(x^k, K\cap U\right)\right)_{k\in\mathbb{N}}$ *converge Q-linearly to* 0*, and the asymptotic constants are bounded above by* $\sqrt{1-\omega^2(1-\gamma)^2}$*, with* ω *as in Assumption EB.*

Proof. Since $\left(z^k\right)_{k\in\mathbb{N}} \subset U$ and $z^{k+1} = T\left(z^k\right)$, from the first inequality in (7.4) we get

$$\left(1-\omega^2(1-\gamma)^2\right)\operatorname{dist}^2\left(z^k, K\cap U\right) \geq \operatorname{dist}^2\left(z^{k+1}, K\cap U\right),$$

from which we obtain

$$\frac{\operatorname{dist}\left(z^{k+1}, K\cap U\right)}{\operatorname{dist}\left(z^k, K\cap U\right)} \leq \sqrt{1-\omega^2(1-\gamma)^2} \tag{7.7}$$

Considering the second inequality in (7.4) and Lemma 5.1, following the same steps gives us

$$\frac{\operatorname{dist}\left(x^{k+1}, K\cap U\right)}{\operatorname{dist}\left(x^k, K\cap U\right)} \leq \sqrt{1-\omega^2(1-\gamma)^2} \tag{7.8}$$

The inequalities in (7.7) and (7.8) imply the result. □

We remark that the result for AMAP still holds when U is any closed and convex set, but for CARM we need U to be an affine subspace, as it might diverge otherwise.

We'll now show that, under Assumption EB, CARM achieves a linear rate with an asymptotic constant better than the one given in Corollary 7.3.

Proposition 7.4 (Improved upper bound for asymptotic constant of CARM under EB)**.** *Let* $\left(x^k\right)_{k\in\mathbb{N}}$ *be the sequence generated by CARM starting at any* $x^0 \in \mathbb{R}^n$*. If* K, U *satisfy Assumption EB then the sequence* $\left(\operatorname{dist}\left(x^k, K\cap U\right)\right)_{k\in\mathbb{N}}$ *converges to 0 with the asymptotic constant bounded above by* $\sqrt{\frac{1-\omega^2(1-\gamma)^2}{1+\omega^2(1-\gamma)^2}}$*, where* ω *is as in Assumption EB.*

Proof. Take $y^* \in K \cap U$ and $x \in U$. Note that

$$\begin{aligned}
\operatorname{dist}^2(x, K) = \|x - P_K(x)\|^2 &\leq \frac{1}{(1-\gamma)^2} \left\|x - \tilde{P}_K(x)\right\|^2 \\
&\leq \frac{1}{(1-\gamma)^2} \left(\|x - y^*\|^2 - \left\|\tilde{P}_K(x) - y^*\right\|^2\right) \\
&= \frac{1}{(1-\gamma)^2} \left(\|x - y^*\|^2 - \left\|\tilde{P}_K(x) - \tilde{P}_K\left(y^*\right)\right\|^2\right) \\
&\leq \frac{1}{(1-\gamma)^2} \left(\|x - y^*\|^2 - \left\|P_U\left(\tilde{P}_K(x)\right) - P_U\left(\tilde{P}_K\left(y^*\right)\right)\right\|^2 - \left\|P_U\left(\tilde{P}_K(x)\right) - \tilde{P}_K(x)\right\|^2\right) \\
&= \frac{1}{(1-\gamma)^2} \left(\|x - y^*\|^2 - \left\|P_U\left(\tilde{P}_K(x)\right) - y^*\right\|^2 - \left\|P_U\left(\tilde{P}_K(x)\right) - \tilde{P}_K(x)\right\|^2\right)
\end{aligned} \tag{7.9}$$

using the firm quasinonexpansiveness of the outer-approximate projection and the fact that $y^* \in K$ in second inequality, and the firm nonexpansiviness of the orthogonal projection onto U in the last inequality.

From Lemma 5.1, we have that $\tilde{C}(x) = P_{\tilde{H}_x \cap U}(x)$ with

$$\tilde{H}_x := \left\{y \in \mathbb{R}^n \mid \left\langle y - \tilde{P}_K(x), x - \tilde{P}_K(x)\right\rangle \leq 0\right\} \supset K.$$

Hence, $\left\langle y^* - \tilde{C}(x), x - \tilde{C}(x)\right\rangle \leq 0$ and since $x, P_U\left(\tilde{P}_K(x)\right)$ and $\tilde{C}(x)$ are collinear by (7.2), we get $\left\langle y^* - \tilde{C}(x), P_U\left(\tilde{P}_K(x)\right) - \tilde{C}(x)\right\rangle \leq 0$. Thus,

$$\left\|P_U\left(\tilde{P}_K(x)\right) - y^*\right\|^2 \geq \left\|\tilde{C}(x) - y^*\right\|^2 + \left\|\tilde{C}(x) - P_U\left(\tilde{P}_K(x)\right)\right\|^2 \tag{7.10}$$

Now, (7.9) and (7.10) imply

$$\begin{aligned}
\operatorname{dist}^2(x, K) &\leq \frac{1}{(1-\gamma)^2} \left(\|x - y^*\|^2 - \left\|\tilde{C}(x) - y^*\right\|^2 - \left\|\tilde{C}(x) - P_U\left(\tilde{P}_K(x)\right)\right\|^2 - \left\|P_U\left(\tilde{P}_K(x)\right) - \tilde{P}_K(x)\right\|^2\right) \\
&= \frac{1}{(1-\gamma)^2} \left(\|x - y^*\|^2 - \left\|\tilde{C}(x) - y^*\right\|^2 - \left\|\tilde{C}(x) - \tilde{P}_K(x)\right\|^2\right) \\
&\leq \frac{1}{(1-\gamma)^2} \left(\|x - y^*\|^2 - \left\|\tilde{C}(x) - y^*\right\|^2 - \operatorname{dist}^2(\tilde{C}(x), K)\right) \\
&\leq \frac{1}{(1-\gamma)^2} \left(\|x - y^*\|^2 - \operatorname{dist}^2(\tilde{C}(x), K \cap U) - \operatorname{dist}^2(\tilde{C}(x), K)\right)
\end{aligned}$$

using the definition of the distance in the last two inequalities. Now, taking $y^* = P_{K\cap U}(x)$ and using the error bound condition for x and $C(x)$, we obtain

$$\begin{aligned}\omega^2 \operatorname{dist}^2(x, K\cap U) &\leq \frac{1}{(1-\gamma)^2}\operatorname{dist}^2(x,K)\\ &\leq \frac{1}{(1-\gamma)^2}\left(\operatorname{dist}^2(x,K\cap U) - \operatorname{dist}^2(\tilde{C}(x),K\cap U) - \omega^2(1-\gamma)^2\operatorname{dist}^2(\tilde{C}(x),K\cap U)\right)\\ &= \frac{1}{(1-\gamma)^2}\operatorname{dist}^2(x,K\cap U) - \frac{1}{(1-\gamma)^2}\left(1+\omega^2(1-\gamma)^2\right)\operatorname{dist}^2(\tilde{C}(x),K\cap U)\end{aligned}$$

which we rearrange to get

$$\left(1+\omega^2(1-\gamma)^2\right)\operatorname{dist}^2(\tilde{C}(x),K\cap U) \leq \left(1-\omega^2(1-\gamma)^2\right)\operatorname{dist}^2(x,K\cap U).$$

Since $x^{k+1} = \tilde{C}\left(x^k\right)$, we have

$$\frac{\operatorname{dist}\left(x^{k+1},K\cap U\right)}{\operatorname{dist}\left(x^k,K\cap U\right)} \leq \sqrt{\frac{1-\omega^2(1-\gamma)^2}{1+\omega^2(1-\gamma)^2}}$$

which implies the result. □

Propositions 7.2 and 7.4 do not outright ensure that the sequences $\left(x^k\right)_{k\in\mathbb{N}}$ and $\left(z^k\right)_{k\in\mathbb{N}}$ themselves converge linearly; a sequence $\left(y^k\right)_{k\in\mathbb{N}} \subset \mathbb{R}^n$ may converge to a point $y \in M \subset \mathbb{R}^n$, in such a way that $\left(\operatorname{dist}\left(y^k,M\right)\right)_{k\in\mathbb{N}}$ converges linearly to 0 but $\left(y^k\right)_{k\in\mathbb{N}}$ itself converges sublinearly. However, the following fact establishes that this can not occur when $\left(y^k\right)_{k\in\mathbb{N}}$ is Fejér monotone with respect to M and assures that the sequences converge R-linearly.

Fact 7.5. *Consider $M \subset \mathbb{R}^n$, $\left(y^k\right)_{k\in\mathbb{N}} \subset \mathbb{R}^n$. Assume that $\left(y^k\right)_{k\in\mathbb{N}}$ is Fejér monotone with respect to M, and that $\left(\operatorname{dist}\left(y^k,M\right)\right)_{k\in\mathbb{N}}$ converges R -linearly to 0. Then $\left(y^k\right)_{k\in\mathbb{N}}$ converges R -linearly to some point $y^* \in M$, with asymptotic constant bounded above by the asymptotic constant of $\left(\operatorname{dist}\left(y^k,M\right)\right)_{k\in\mathbb{N}}$.*

Proof. Fix $k \in \mathbb{N}$ and note that the Fejér monotonicity hypothesis implies that, for all $j \geq k$,

$$\left\|y^j - P_M\left(y^k\right)\right\| \leq \left\|y^k - P_M\left(y^k\right)\right\| = \operatorname{dist}\left(y^k,M\right). \tag{7.11}$$

By Fact 2.33(i), $\left(y^k\right)_{k\in\mathbb{N}}$ is bounded. Take any cluster point $\bar{y}$ of $\left(y^k\right)_{k\in\mathbb{N}}$. Taking limits with $j \to \infty$ in (7.11) along a subsequence $\left(y^{k_j}\right)_{j\in\mathbb{N}}$ of $\left(y^k\right)_{k\in\mathbb{N}}$ converging to $\bar{y}$,

we get that $\left\|\bar{y} - P_M\left(y^k\right)\right\| \leq \operatorname{dist}\left(y^k, M\right)$. Since $\lim_{k\to\infty} \operatorname{dist}\left(y^k, M\right) = 0$, we conclude that $\left(P_M\left(y^k\right)\right)_{k\in\mathbb{N}}$ converges to $\bar{y}$, so that there exists a unique cluster point, say y^*. Therefore, $\lim_{k\to\infty} y^k = y^*$, and hence $\left\|y^* - P_M\left(y^k\right)\right\| \leq \operatorname{dist}\left(y^k, M\right)$. Since $y^* = \lim_{k\to\infty} P_M\left(y^k\right)$, we conclude that $y^* \in M$. Observe further that

$$\begin{aligned}\left\|y^k - y^*\right\| &\leq \left\|y^k - P_M\left(y^k\right)\right\| + \left\|P_M\left(y^k\right) - y^*\right\| \\ &= \operatorname{dist}\left(y^k, M\right) + \left\|y^* - P_M\left(y^k\right)\right\| \leq 2\operatorname{dist}\left(y^k, M\right) \end{aligned} \tag{7.12}$$

Taking kth-root and then $\limsup$ with $k \to \infty$ in (7.12) and using the R-linearity hypothesis,

$$\begin{aligned}\limsup_{k\to\infty} \left\|y^k - y^*\right\|^{1/k} &\leq \limsup_{k\to\infty} 2^{1/k} \operatorname{dist}\left(y^k, M\right)^{1/k} \\ &= \limsup_{k\to\infty} \operatorname{dist}\left(y^k, M\right)^{1/k} < 1.\end{aligned}$$

establishing both that $\left(y^k\right)_{k\in\mathbb{N}}$ converges R-linearly to $y^* \in M$ and the statement on the asymptotic constant. □

With this result in hand, we prove that the AMAP and CARM sequences converge R-linearly under Assumption EB and give bounds for their asymptotic constants.

Theorem 7.6 (Convergence of CARM and AMAP under error bound)**.** *Consider a closed and convex set $K \subset \mathbb{R}^n$ and an affine subspace $U \subset \mathbb{R}^m$. Assume that K, U satisfy Assumption EB. Let $\left(z^k\right)_{k\in\mathbb{N}}, \left(x^k\right)_{k\in\mathbb{N}}$ be the sequences generated by AMAP and CARM, respectively, starting from arbitrary points $z^0 \in \mathbb{R}^n, x^0 \in U$. Then both sequences $\left(z^k\right)_{k\in\mathbb{N}}$ and $\left(x^k\right)_{k\in\mathbb{N}}$ converge R-linearly to points in $K \cap U$, and the asymptotic constants are bounded above by $\sqrt{1-\omega^2(1-\gamma)^2}$ for AMAP, and by $\sqrt{\frac{1-\omega^2(1-\gamma)^2}{1+\omega^2(1-\gamma)^2}}$ for CARM, with ω as in Assumption EB.*

Proof. In view of Theorems 5.4 and 6.3, both sequences are Fejér monotone with respect to $K \cap U$ and converge to points in $K \cap U$. By Corollary 7.3, both sequences $\left(\operatorname{dist}\left(z^k, K \cap U\right)\right)_{k\in\mathbb{N}}$ and $\left(\operatorname{dist}\left(x^k, K \cap U\right)\right)_{k\in\mathbb{N}}$ are Q-linearly convergent to 0, and henceforth R-linearly convergent to 0. Corollary 7.3 shows that the asymptotic constant of the sequence $\left(\operatorname{dist}\left(z^k, K \cap U\right)\right)_{k\in\mathbb{N}}$ is bounded above by $\sqrt{1-\omega^2(1-\gamma)^2}$, and Proposition 7.4 establishes that the asymptotic constant of the sequence $\left(\operatorname{dist}\left(x^k, K \cap U\right)\right)_{k\in\mathbb{N}}$ is bounded above by $\sqrt{1-\omega^2(1-\gamma)^2}/\sqrt{1+\omega^2(1-\gamma)^2}$. Finally, by Fact 7.5, both $\left(z^k\right)_{k\in\mathbb{N}}$ and $\left(x^k\right)_{k\in\mathbb{N}}$ are R-linearly convergent, with the announced bounds for their asymptotic constants. □

Theorem 7.6 provides a substantially better upper bound for the asymptotic constant of the CARM sequence than the one for the AMAP sequence, as the CARM bound reduces the AMAP one by a factor of $\sqrt{1+\omega^2(1-\gamma)^2}$.

8 Numerical experiments

8.1 Computing outer-approximate projections

In this section, we perform numerical experiments with CARM split in two subsections. In the first subsection, we compare CARM's performance to CRM's in problems where projections are available. In the second, we employ CARM for obtaining a feasible point of a convex optimization problem.

In order to use CARM, first we need to establish how to compute outer-approximate projections. We consider the closed convex set $K := \{x \in \mathbb{R}^n \mid g(x) \leq 0\}$, where $g : \mathbb{R}^n \to \mathbb{R}$ given by $g(x) := \max\{g_1(x), \ldots, g_m(x)\}$ is the maximum of the convex but not necessarily differentiable functions $g_1, \ldots, g_m : \mathbb{R}^n \to \mathbb{R}$.

The following proposition, based on [38], provides a special outer-approximate projector by means of subgradients of g, i.e., of elements of the subdifferential $\partial g(x)$ of g at x. This gives an explicit formula for computing an outer-approximate projection with low computational expense, provided that the computation of a subgradient of g is not too expensive.

Proposition 8.1 (Subgradient formula for outer-approximate projections)**.** *Suppose that $g : \mathbb{R}^n \to \mathbb{R}$ is a convex function and that there exists some $\hat{x} \in \mathbb{R}^n$ with $g(\hat{x}) < 0$ (Slater condition). Let $v : \mathbb{R}^n \to \mathbb{R}^n$ denote a mapping with $v(x) \in \partial g(x)$ for any $x \in \mathbb{R}^n$. Moreover, let the mapping $\tilde{p} : \mathbb{R}^n \to \mathbb{R}^n$ be defined pointwise, with $\tilde{p}(x)$ being the unique solution of the problem.*

$$\min_p \|p - x\| \quad s.t. \quad g(x) + v(x)^\top (p - x) \leq 0$$

Then, the following assertions hold:

(i) *For any nonempty compact set $B \subset \mathbb{R}^n$, the mapping $\tilde{p}$ is an outer-approximate projector defined on B associated to the set $\Omega := \{x \in \mathbb{R}^n \mid g(x) \leq 0\}$.*

(ii) *For any $x \in \mathbb{R}^n$, the solution $\tilde{p}(x)$ of (10) is given by*

$$\tilde{p}(x) = \begin{cases} x - \frac{g(x)}{\|v(x)\|^2} v(x) & \text{if } g(x) > 0, \\ x & \text{if } g(x) \leq 0. \end{cases}$$

With this outer-approximate projection available, we can now perform numerical experiments using CARM. We remark that in many cases, such as when projecting onto the epigraph of a linear function, the subgradients will yield the exact projection. In our analysis, we will only consider problems in which the approximate projection provided by Proposition 8.1 differs from the exact one.

As a first basic application, we'll see how CARM performs when projecting from the real line onto the norm of the epigraph of a quadratic function. Before the actual

numerical experiments, we present a one-dimensional toy version and present explicit calculations for the outer-approximate projections and iterations of CARM and AMAP.

8.1.1 Toy example with explicit calculations

Consider $f : \mathbb{R}_+ \to \mathbb{R}_+$ given by $f(x) = x^2$, with $\operatorname{epi} f = \{(x, y) \in \mathbb{R}^2_+ \mid x^2 \leq y\}$. Since $f(x)$ is convex, so is its epigraph. Finding points in $\operatorname{epi} f$ is equivalent to solving $g(x, y) = x^2 - y \leq 0$.

Let $a^0 \in \mathbb{R}, a^0 > 0$ be arbitrary but fixed. Setting $K = g(x, y) \leq 0$, $U = \{(a, 0) \in \mathbb{R}^2 \mid a > 0\}$ and starting in $x^0 = (a^0, 0) \in U$, we have by Theorems 5.4 and 6.3 that sequences generated by AMAP and CARM, initialized in x^0, will converge to a point $z^* \in K \cap U$; trivially, $z^* = (0, 0)$, is the only solution.

Note that g is convex, differentiable and has a Slater point in $(0, 1)$, since $g(0, 1) < 0$. Using calculus, we find that $\partial g(x, y) = \{(2x, -1)\}$, which is well-defined for all $(x, y) \in \mathbb{R}^2$. Therefore, we can use Proposition 8.1 to obtain an outer-approximate projection $\tilde{p}(x)$. We will describe the general step of the CARM iteration for this problem.

Let $(x_k)_{k\in\mathbb{N}}$ be the sequence generated by CARM, starting from $x^0 = (a^0, 0)$. Since the sequence must be contained in U (Lemma 5.1), we have that, for any $n \in \mathbb{N}$, $x^k = (a^k, 0)$ and $x_{n+1} = \tilde{C}(x_n)$. For $x = (a, 0)$, $a > 0$, we have that

$$
\begin{aligned}
\tilde{p}(x) &= x - \frac{g(x)}{\|v(x)\|^2} v(x) \\
&= (a, 0) - \frac{(a^2 - 0)}{\|(2a, -1)\|^2}(2a, -1) \\
&= \frac{(4a^3 + a, 0) - (2a^3, -a^2)}{4a^2 + 1} \\
&= \left(\frac{2a^3 + a}{4a^2 + 1}, \frac{a^2}{4a^2 + 1}\right).
\end{aligned}
$$

Note that $P_U \tilde{p}(x) = \left(\frac{2a^3+a}{4a^2+1}, 0\right)$, which happens to be the AMAP step.

From Lemma 5.1, since $\tilde{p}(x)$ is an outer-approximate projection and $x \notin K$, we know that $\tilde{C}(x) = P_{\tilde{H}_x \cap U}(x)$, where $\tilde{H}_x := \{z \in \mathbb{R}^n \mid \langle z - \tilde{p}_K(x), x - \tilde{p}_K(x)\rangle = 0\}$. Since $\tilde{C}(x) \in U$, it must be of the form $(b, 0), b \in \mathbb{R}^n_+$. Since $\tilde{C}(x) \in \tilde{H}_x$, we must have that

$\left\langle \tilde{C}(x) - \tilde{p}(x), x - \tilde{p}(x) \right\rangle = 0$. Thus, we have that

$$\begin{aligned}
&\left\langle (b,0) - \left(\frac{2a^3+a}{4a^2+1}, \frac{a^2}{4a^2+1}\right), (a,0) - \left(\frac{2a^3+a}{4a^2+1}, \frac{a^2}{4a^2+1}\right) \right\rangle \\
=&\left\langle \left(\frac{4a^2b+b}{4a^2+1}, 0\right) - \left(\frac{2a^3+a}{4a^2+1}, \frac{a^2}{4a^2+1}\right), \left(\frac{4a^3+a}{4a^2+1}, 0\right) - \left(\frac{2a^3+a}{4a^2+1}, \frac{a^2}{4a^2+1}\right) \right\rangle \\
=&\left\langle \left(\frac{-2a^3+4a^2b-a+b}{4a^2+1}, \frac{-a^2}{4a^2+1}\right), \left(\frac{2a^3}{4a^2+1}, \frac{-a^2}{4a^2+1}\right) \right\rangle \\
=&\frac{(-2a^3+4a^2b-a+b)(2a^3)+a^4}{(4a^2+1)^2} \\
=&\frac{-a^3(4a^2+1)(a-2b)}{(4a^2+1)^2} = 0.
\end{aligned}$$

which gives us that $b = \frac{a}{2}$. Thus, $\tilde{C}(x) = \left(\frac{a}{2}, 0\right)$.

Therefore, the sequence $(x^k_{CARM})_{k\in\mathbb{N}}$ generated by CARM starting on $x^0 = (a^0, 0)$ is such that $x^k_{CARM} = \left(\frac{a^0}{2^k}\right)$, for all $k \in \mathbb{N}$. For k sufficiently large, this sequence clearly converges to $(0,0)$, which is the only solution.

Comparatively, the sequence generated by AMAP $(x^k_{AMAP})_{k\in\mathbb{N}}$, if also starting on $x^0 = (a^0, 0)$, would have the k-th element be $x^k_{AMAP} = \left(\frac{2a^3+a}{4a^2+1}, 0\right)$. Thus, it can be easily seen that AMAP's sequence also converges to $(0,0)$, albeit slower than CARM's.

In this setting, exact projections are available and can be obtained by solving a minimization problem.

Let $(x,s) \in \operatorname{epi}(f)$ and $(z,r) \in \mathbb{R}^2_+ \setminus \operatorname{epi}(f)$. The projection of (z,r) onto $\operatorname{epi}(f)$ is given by the pair $(\bar{x}, \bar{s})$ which solves

$$\min_{x,s:f(x)\le s} \frac{1}{2}\|x-z\|^2 + \frac{1}{2}\|s-r\|^2.$$

To solve this problem using KKT conditions, we write the Lagrangian

$$\mathcal{L}(x,s,\mu) = \frac{1}{2}\|x-z\|^2 + \frac{1}{2}\|s-r\|^2 + \mu(f(x)-s).$$

Plugging in $f(x) = x^2$ and $\partial f(x) = 2x$, we take partial derivatives with respect to x, s and μ, respectively, to get the conditions:

$$\begin{aligned}
x - z + 2x\mu &= 0, \\
s - r - \mu &= 0, \\
x^2 - s &= 0,
\end{aligned}$$

noting that $(z,r) \notin \operatorname{epi}(f)$ implies $\mu > 0$.

After some calculations, these conditions yield the polynomial

$$4\mu^3 + (4r+4)\mu^2 + (4r+1)\mu + r - z^2 = 0.$$

After calculating μ (the largest positive root of the polynomial above), the projection of (z, r) onto $\operatorname{epi}(f)$ is the solution $(\bar{x}, \bar{s})$, where

$$\bar{x} = \frac{1}{1+2\mu} z, \quad \bar{s} = \mu + r.$$

Using this projection, we can calculate the CRM and MAP iterations. In the following experiment, we will see how CARM fares against CRM, AMAP and MAP in a more general version of this example.

8.2 CARM for the CFP given by the epigraph of a quadratic function and an affine subspace

The problem under consideration for our numerical experiments is the CFP of finding a point in the intersection of the epigraph of a quadratic function and an affine subspace given by one of the axis on a n-dimensional real coordinate space.

Formally, let $K_\alpha = \operatorname{epi}(f_\alpha)$, where $f_\alpha : \mathbb{R}^n \to \mathbb{R}$ is given by $f_\alpha(x) = \langle \alpha x, x \rangle, \alpha \in \mathbb{R}_+$, and let $U_{A,b} : \{x \in \mathbb{R}^n \mid Ax = b\}$, where $A \in \mathbb{R}^{n+1}$ is the matrix $A = [\underbrace{0, \ldots, 0}_{n \text{ times}}, 1]$ and $b \in \mathbb{R}_+$. Clearly, K_α is a closed convex set and $U_{A,b}$ is an affine subspace. Then the problem is question can be written as the CFP

$$\text{Find } x^* \in S := K_\alpha \cap U_{A,b}.$$

We split our tests in two categories, which differ in one important aspect.

For both categories, we randomly generate 100 instances of subspaces $U_{A,b}$, where n is fixed as 200, and 100 instances of K_α, in which α is sampled from an uniform distribution between 0 and 10. For the first category, we fix $b = 0$, which implies the absence of an error bound. For the second category, b is sampled from the absolute values of a Normal distribution with mean 0 and standard deviation 5. Graphically, this means that the affine subspace $U_{A,b}$ is the $n+1$ axis shifted upwards, which implies the existence an error bound as defined in Section 7.2. In both cases, this is sufficient to assure that S is nonempty. In view of Theorems 5.4, 6.3 and 7.6, we are assured that CARM and AMAP will converge with or without an error bound.

The following aspects are true for both EB and non-EB tests. Each instance is run for 10 initial random points, summing up to 1000 individual tests in both categories. Each initial point $z^0 \in \mathbb{R}^{n+1}$ sampled from the standard normal distribution is assured to have norm between 5 and 15 and is accepted as long as $z^0 \notin S$. Then, z^0 is projected it onto $U_{A,b}$ to begin each method.

Let $\left(z^k\right)$ be any of the four sequences that we monitor: $\left(z^k_{CARM}\right)$, $\left(z^k_{AMAP}\right)$, $\left(z^k_{CRM}\right)$ and $\left(z^k_{MAP}\right)$, that is, CARM, AMAP, CRM and MAP sequences, respectively. We chose the gap distance with tolerance $\varepsilon := 10^{-6}$ as our stopping criteria, given by

$$\left\|P_{U_{A,b}}\left(z^k\right) - P_{K_\alpha}\left(z^k\right)\right\| < \varepsilon,$$

which we consider a reasonable measure of infeasibility. Note this calculation does not add any extra cost, since the projections computed to measure the gap distance can be utilized on the next iteration.

First, we will see in Figure 8.1 the performance profile of the four algorithms without an error bound.

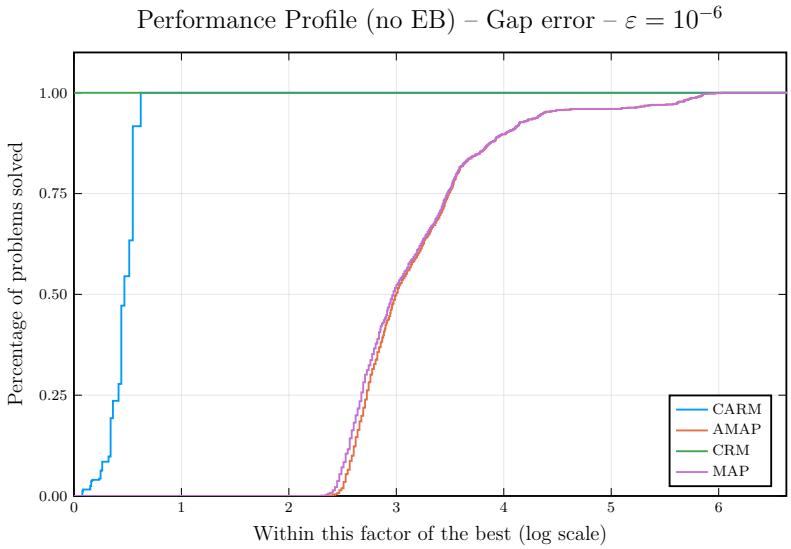

Fig. 8.1: Experiments with epigraph of a quadratic function and an affine subspace, without error bound - CARM, AMAP, CRM and MAP

In the absence of an error bound, CARM perform exceptionally well, with a performance comparable to CRM's, which outperforms all the other algorithms. AMAP and MAP perform very similarly, with AMAP performing only slightly worse, both requiring at least twice the number of iterations required by CRM to in every problem and also being considerably outperformed by CARM.

Removing CRM from the comparison, in Figure 8.2 we can clearly see that CARM was considerably better than MAP and AMAP in all problems.

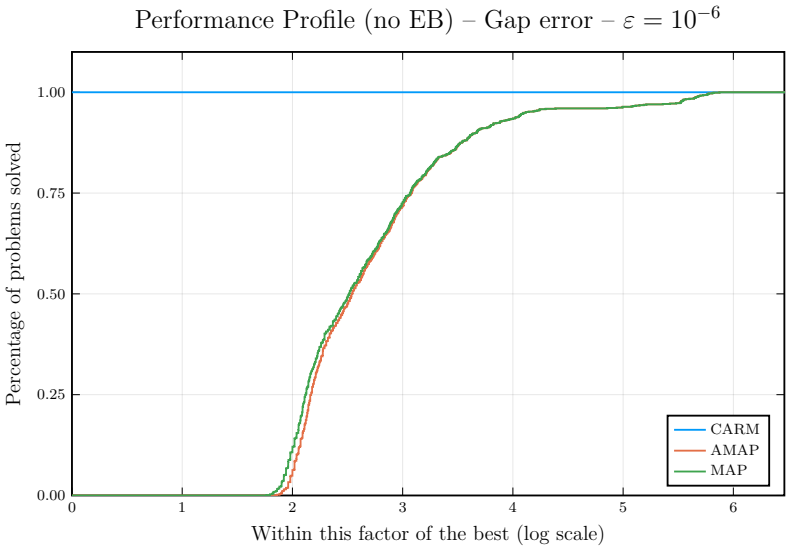

Fig. 8.2: Experiments with epigraph of a quadratic function and an affine subspace, without error bound - CARM, AMAP and MAP

To sum these results, we present descriptive statistics.

	mean	min	median	max
CARM	19.093	18	19.0	20
AMAP	158.957	67	111.0	1171
CRM	13.932	13	14.0	18
MAP	157.083	64	109.0	1174

Table 8.1: Statistics of the experiments without error bound (in number of iterations)

In the experiments with error bound, CRM remains the best performing method, in particular, outperforming MAP. This results is consistent with [22], as K is the epigraph of a (proper) convex function and f satisfies the assumptions A1, A2, A3 and A4 shown in Section 4 of that paper to assure that CRM converges linearly, while MAP converges sublinearly. In Figure 8.3 we can see how the presence of an error bound affects the performance of the algorithms.

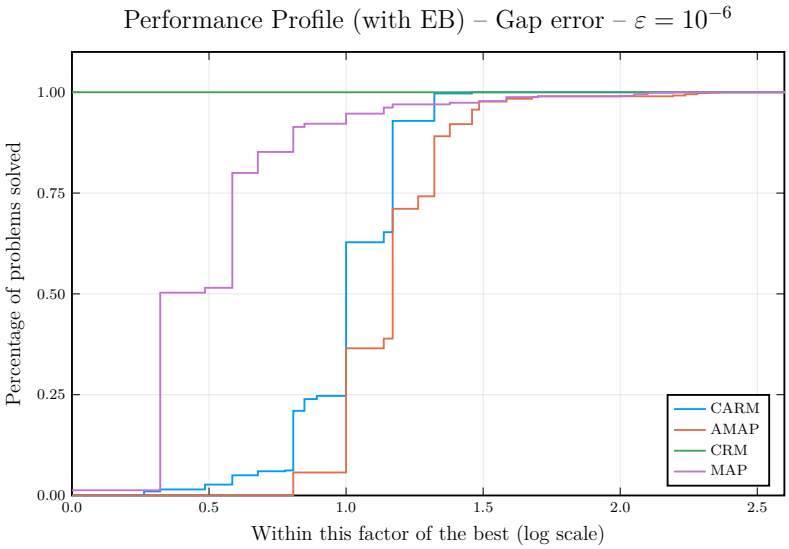

Fig. 8.3: Experiments with epigraph of a quadratic function and an affine subspace, with error bound - CARM, AMAP, CRM and MAP

While MAP outperforms CARM in both cases, this comparison requires a closer look. Once again, we exclude CRM to take a clearer look into how CARM compares to MAP and AMAP.

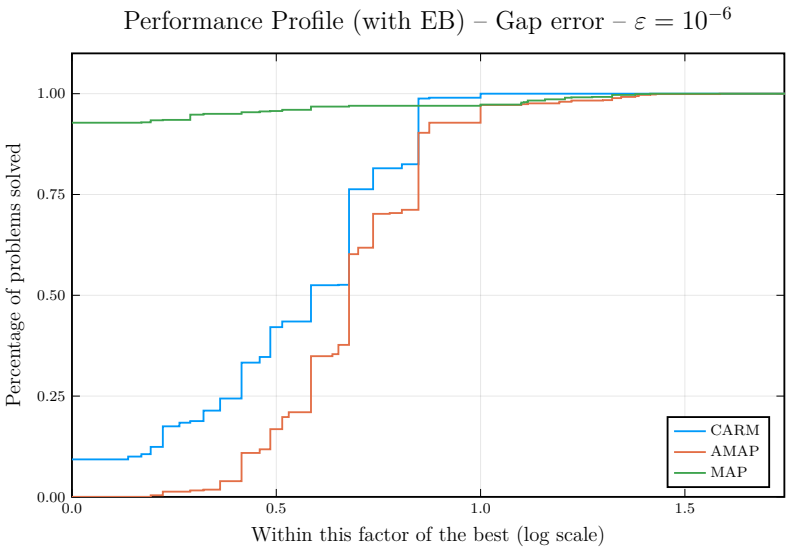

Fig. 8.4: Experiments with epigraph of a quadratic function and an affine subspace, with error bound - CARM, AMAP and MAP

While generally MAP outperforms CARM, CARM converged in at least as few iterations as MAP in 93 out of 1000 problems. We now remove MAP from the comparison to better visualize how CARM fares when compared with AMAP.

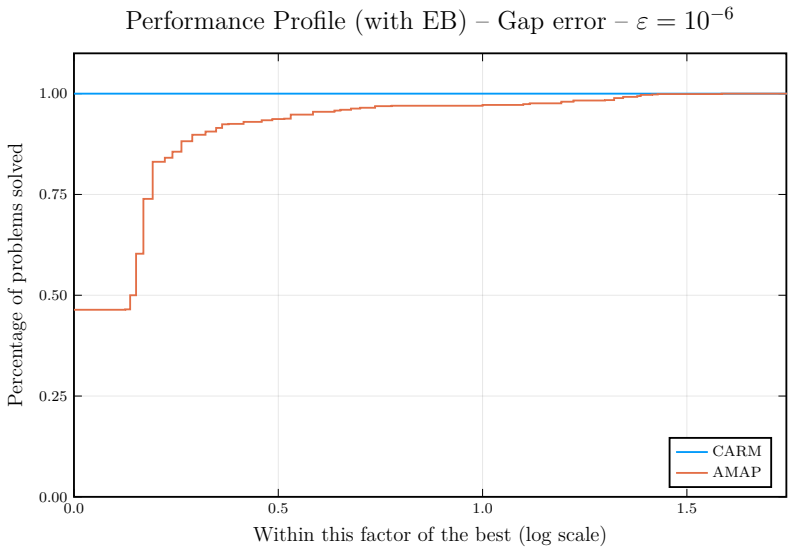

Fig. 8.5: Experiments with epigraph of a quadratic function and an affine subspace, without error bound - CARM and AMAP

The figure above makes it clear that CARM also outperforms AMAP handily with an error bound. We conclude this examination by presenting descriptive statistics for this error bound setting.

	mean	min	median	max
CARM	8.4	5	8.0	13
AMAP	9.492	7	9.0	36
CRM	4.15	4	4.0	7
MAP	6.265	4	5.0	32

Table 8.2: Statistics of the experiments with error bound (in number of iterations)

9 Final remarks

We have extended results concerning the circumcentered-reflection method (CRM) by employing outer-approximate projections within a method we called CARM. In our numerical experiments, we see that CARM performs competitively with CRM. In the future, we hope to provide more interesting numerical experiments, such as when exact projections are unavailable, and more specific applications. In particular, we hope that CARM would perform well when compared with Combettes' Extrapolated Method of Parallel Subgradient Projections (EMOPSP) [27] in an image recovery application.

We would also like to investigate the use overrelaxations onto outer-approximate projections as means of accelerating CARM and AMAP. Preliminary exploratory tests have been favorable, in that sense. Though it would require more complex calculations and perhaps stronger hypotheses, we also believe that outer-approximate projections could be used for projection methods outside of a product space formulation, where the availability of exact projections onto the affine subspace helps very much with proving the convergence.

References

1. F. J. Aragón Artacho, R. Campoy, and M. K. Tam, *The Douglas–Rachford algorithm for convex and nonconvex feasibility problems*, Math Meth Oper Res, (2019), https://doi.org/10.1007/s00186-019-00691-9.

2. R. Arefidamghani, R. Behling, J.-Y. Bello-Cruz, A. N. Iusem, and L.-R. Santos, *The circumcentered-reflection method achieves better rates than alternating projections*, arXiv, (2020), https://arxiv.org/abs/2007.14466.

3. C. Badea, S. Grivaux, and V. Müller, *The rate of convergence in the method of alternating projections*, St Petersburg Math. J., 23 (2012), pp. 413–434, https://doi.org/10.1090/S1061-0022-2012-01202-1.

4. S. Banach, *Sur les opérations dans les ensembles abstraits et leur application aux équations intégrales*, Fundamenta Mathematicae, 3 (1922), pp. 133–181, http://eudml.org/doc/213289.

5. H. H. Bauschke, J.-Y. Bello-Cruz, T. T. A. Nghia, H. M. Phan, and X. Wang, *The rate of linear convergence of the Douglas–Rachford algorithm for subspaces is the cosine of the Friedrichs angle*, J. Approx. Theory, 185 (2014), pp. 63–79, https://doi.org/10.1016/j.jat.2014.06.002.

6. H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, CMS Books in Mathematics, Springer International Publishing, Cham, second ed., 2017, https://doi.org/10.1007/978-3-319-48311-5.

7. H. H. Bauschke, M. N. Dao, D. Noll, and H. M. Phan, *Proximal point algorithm, Douglas-Rachford algorithm and alternating projections: A case study*, J. Convex Anal., 23 (2016), pp. 237–261.

8. H. H. Bauschke, F. Deutsch, and H. Hundal, *Characterizing arbitrarily slow convergence in the method of alternating projections*, Int. Trans. Oper. Res., 16 (2009), pp. 413–425, https://doi.org/10.1111/j.1475-3995.2008.00682.x.

9. H. H. Bauschke, F. R. Deutsch, H. Hundal, and S.-H. Park, *Accelerating the Convergence of the Method of Alternating Projections*, Trans. Am. Math. Soc., 355 (2003), pp. 3433–3461, https://doi.org/10.2307/1194848.

10. H. H. Bauschke, D. R. Luke, H. M. Phan, and X. Wang, *Restricted Normal Cones and the Method of Alternating Projections: Theory*, Set-Valued Var. Anal., 21 (2013), pp. 431–473, https://doi.org/10.1007/s11228-013-0239-2.

11. H. H. Bauschke, D. Noll, and H. M. Phan, *Linear and strong convergence of algorithms involving averaged nonexpansive operators*, J. Math. Anal. Appl., 421 (2015), pp. 1–20, https://doi.org/10.1016/j.jmaa.2014.06.075.

12. H. H. Bauschke, H. Ouyang, and X. Wang, *On circumcenters of finite sets in Hilbert spaces*, Linear Nonlinear Anal., 4 (2018), pp. 271–295, https://arxiv.org/abs/1807.02093.

13. H. H. Bauschke, H. Ouyang, and X. Wang, *Best approximation mappings in Hilbert spaces*, arXiv, (2020), https://arxiv.org/abs/2006.02644.

14. H. H. Bauschke, H. Ouyang, and X. Wang, *Circumcentered methods induced by isometries*, Vietnam J. Math., 48 (2020), https://arxiv.org/abs/1908.11576.

15. H. H. Bauschke, H. Ouyang, and X. Wang, *On circumcenter mappings induced by nonexpansive operators*, Pure Appl. Funct. Anal., to appear (2020), https://arxiv.org/abs/1811.11420.

16. H. H. Bauschke, H. Ouyang, and X. Wang, *On the linear convergence of circumcentered isometry methods*, Numer Algor, (2020), https://doi.org/10.1007/s11075-020-00966-x, https://arxiv.org/abs/1912.01063.

17. H. H. Bauschke, H. M. Phan, and X. Wang, *The method of alternating relaxed projections for two nonconvex sets*, (2013), https://arxiv.org/abs/arXiv:1305.4296.

18. R. Behling, J.-Y. Bello-Cruz, and L.-R. Santos, *Circumcentering the Douglas–Rachford method*, Numer. Algorithms, 78 (2018), pp. 759–776, https://doi.org/10.1007/s11075-017-0399-5.

19. R. Behling, J.-Y. Bello-Cruz, and L.-R. Santos, *On the linear convergence of the circumcentered-reflection method*, Oper. Res. Lett., 46 (2018), pp. 159–162, https://doi.org/10.1016/j.orl.2017.11.018.

20. R. Behling, J.-Y. Bello-Cruz, and L.-R. Santos, *The block-wise circumcentered–reflection method*, Comput Optim Appl, 76 (2020), pp. 675–699, https://doi.org/10.1007/s10589-019-00155-0.

21. R. Behling, J.-Y. Bello-Cruz, and L.-R. Santos, *Infeasibility and error bound imply finite convergence of alternating projections*, arXiv, (2020), https://arxiv.org/abs/2008.03354.

22. R. Behling, J.-Y. Bello-Cruz, and L.-R. Santos, *On the Circumcentered-Reflection Method for the Convex Feasibility Problem*, Numer. Algorithms, (2020), https://doi.org/10.1007/s11075-020-00941-6.

23. J. M. Borwein and M. K. Tam, *A Cyclic Douglas–Rachford Iteration Scheme*, J. Optim. Theory Appl., 160 (2014), pp. 1–29, https://doi.org/10.1007/s10957-013-0381-x.

24. H. T. Bui, R. Loxton, and A. Moeini, *A Note on the Finite Convergence of Alternating Projections*, ArXiv200806314 Math, (2020), https://arxiv.org/abs/2008.06314.

25. A. Cegielski and A. Suchocka, *Incomplete alternating projection method for large inconsistent linear systems*, Linear Algebra Its Appl., 428 (2008), pp. 1313–1324, https://doi.org/10.1016/j.laa.2007.05.045.

26. G. Cimmino, *Calcolo approssimato per le soluzioni dei sistemi di equazioni lineari*, Ric. Sci., 9 (1938), pp. 326–333.

27. P. Combettes, *The convex feasibility problem in image recovery*, in Advances in Imaging and Electron Physics, Elsevier, 1996, pp. 155–270, https://doi.org/10.1016/s1076-5670(08)70157-5, https://doi.org/10.1016/s1076-5670(08)70157-5.

28. P. Diaconis, K. Khare, and L. Saloff-Coste, *Stochastic alternating projections*, Illinois J. Math., 54 (2010), pp. 963–979, https://doi.org/10.1215/ijm/1336568522.

29. N. Dizon, J. Hogan, and S. B. Lindstrom, *Circumcentering Reflection Methods for Nonconvex Feasibility Problems*, arXiv, (2019), https://arxiv.org/abs/1910.04384.

30. N. Dizon, J. Hogan, and S. B. Lindstrom, *Centering Projection Methods for Wavelet Feasibility Problems*, arXiv, (2020), https://arxiv.org/abs/2005.05687.

31. J. Douglas and H. H. Rachford Jr., *On the numerical solution of heat conduction problems in two and three space variables*, Trans. Am. Math. Soc., 82 (1956), pp. 421–421, https://doi.org/10.1090/S0002-9947-1956-0084194-4.

32. D. Drusvyatskiy, A. D. Ioffe, and A. S. Lewis, *Transversality and Alternating Projections for Nonconvex Sets*, Found Comput Math, 15 (2015), pp. 1637–1651, https://doi.org/10.1007/s10208-015-9279-3.

33. D. Drusvyatskiy and A. S. Lewis, *Inexact alternating projections on nonconvex sets*, arXiv, (2018), https://arxiv.org/abs/1811.01298.

34. D. Drusvyatskiy, G. Li, and H. Wolkowicz, *A note on alternating projections for ill-posed semidefinite feasibility problems*, Math. Program., 162 (2016), pp. 537–548, https://doi.org/10.1007/s10107-016-1048-9.

35. J. Eckstein and D. P. Bertsekas, *On the douglas—rachford splitting method and the proximal point algorithm for maximal monotone operators*, Mathematical Programming, 55 (1992), pp. 293–318, https://doi.org/10.1007/bf01581204, https://doi.org/10.1007/bf01581204.

36. R. Escalante and M. Raydan, *Alternating Projection Methods*, Fundamentals of Algorithms, Society for Industrial and Applied Mathematics, Philadelphia, PA, Jan. 2011, https://doi.org/10.1137/9781611971941.

37. M. Frank and P. Wolfe, *An algorithm for quadratic programming*, Naval Research Logistics Quarterly, 3 (1956), pp. 95–110, https://doi.org/https://doi.org/10.1002/nav.3800030109, https://onlinelibrary.wiley.com/doi/abs/10.1002/nav.3800030109, https://arxiv.org/abs/https://onlinelibrary.wiley.com/doi/pdf/10.1002/nav.3800030109.

38. M. Fukushima, *An Outer Approximation Algorithm for Solving General Convex Programs*, Operations Research, 31 (1983), pp. 101–113, https://doi.org/10.1287/opre.31.1.101.

39. D. S. Gonçalves, M. L. N. Gonçalves, and F. R. Oliveira, *Levenberg-marquardt methods with inexact projections for constrained nonlinear systems*, (2019), https://arxiv.org/abs/arXiv:1908.06118.

40. I. Halperin, *The product of projection operators*, Acta Univ. Szeged. Acta Sci. Math., 23 (1962), pp. 96–99.

41. L. M. Hernández-Ramos, R. Escalante, and M. Raydan, *Unconstrained Optimization Techniques for the Acceleration of Alternating Projection Methods*, Numer. Funct. Anal. Optim., 32 (2011), pp. 1041–1066, https://doi.org/10.1080/01630563.2011.591954.

42. R. Hesse, D. R. Luke, and P. Neumann, *Alternating Projections and Douglas-Rachford for Sparse Affine Feasibility*, IEEE Trans. Signal Process., 62 (2014), pp. 4868–4881, https://doi.org/10.1109/TSP.2014.2339801.

43. H. S. Hundal, *An alternating projection that does not converge in norm*, Nonlinear Analysis: Theory, Methods & Applications, 57 (2004), pp. 35–61, https://doi.org/10.1016/j.na.2003.11.004.

44. A. Izmailov and M. Solodov, *Otimização - volume 1*, vol. 1, IMPA, third ed., 2014.

45. M. Jaggi, *Revisiting Frank-Wolfe: Projection-free sparse convex optimization*, in Proceedings of the 30th International Conference on Machine Learning, S. Dasgupta and D. McAllester, eds., vol. 28 of Proceedings of Machine Learning Research, Atlanta, Georgia, USA, 17–19 Jun 2013, PMLR, pp. 427–435, http://proceedings.mlr.press/v28/jaggi13.html.

46. E. Kopecká and S. Reich, *A note on the von Neumann alternating projections algorithm*, J. Nonlinear Convex Anal, 5 (2004), pp. 379–386.

47. E. Kopecká and S. Reich, *Another note on the von Neumann alternating projections algorithm*, J. Nonlinear Convex Anal, (2010).

48. E. Kopecká and S. Reich, *Alternating projections and orthogonal decompositions*, J. Nonlinear Convex Anal, 12 (2011), pp. 155–159.

49. E. Kopecká and S. Reich, *A note on alternating projections in Hilbert space*, J. Fixed Point Theory Appl., 12 (2012), pp. 41–47, https://doi.org/10.1007/s11784-013-0097-4, https://arxiv.org/abs/1702.07239.

50. A. Y. Kruger and N. H. Thao, *Regularity of collections of sets and convergence of inexact alternating projections*, J. Convex Anal., 23 (2016), pp. 823–847, https://arxiv.org/abs/1501.04191.

51. A. S. Lewis, D. R. Luke, and J. Malick, *Local Linear Convergence for Alternating and Averaged Nonconvex Projections*, Found Comput Math, 9 (2009), pp. 485–513, https://doi.org/10.1007/s10208-008-9036-y.

52. S. B. Lindstrom, *Computable Centering Methods for Spiraling Algorithms and their Duals, with Motivations from the theory of Lyapunov Functions*, arXiv, (2020), https://arxiv.org/abs/2001.10784.

53. S. B. Lindstrom and B. Sims, *Survey: Sixty Years of Douglas–Rachford*, arXiv, (2018), https://arxiv.org/abs/1809.07181.

54. D. R. Luke, *Relaxed averaged alternating reflections for diffraction imaging*, Inverse Problems, 21 (2005), pp. 37–50, https://doi.org/10.1088/0266-5611/21/1/004.

55. D. R. Luke, *Local linear convergence of approximate projections onto regularized sets*, Nonlinear Analysis: Theory, Methods & Applications, 75 (2012), pp. 1531–1546, https://doi.org/10.1016/j.na.2011.08.027.

56. D. R. Luke, *Prox-Regularity of Rank Constraint Sets and Implications for Algorithms*, J Math Imaging Vis, 47 (2013), pp. 231–238, https://doi.org/10.1007/s10851-012-0406-3.

57. J. M. Ortega and W. C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*, Classics in Applied Mathematics, SIAM, Philadelphia, PA, first ed., Jan. 2000.

58. H. Ouyang, *Circumcenter Operators in Hilbert Spaces*, master's Thesis, University of British Columbia, Okanagan, CA, 2018, https://doi.org/10.14288/1.0371095.

59. G. Pierra, *Decomposition through formalization in a product space*, Math. Program., 28 (1984), pp. 96–115, https://doi.org/10.1007/BF02612715.

60. R. T. Rockafellar, *Convex Analysis (Princeton Landmarks in Mathematics and Physics)*, Princeton University Press, dec 1996, https://www.xarg.org/ref/a/0691015864/.

61. M. K. Tam, *The Method of Alternating Projections*, PhD thesis, Nov. 2012.

62. J. von Neumann, *Functional Operators, Volume 2: The Geometry of Orthogonal Spaces.*, no. 22 in Annals of Mathematics Studies, Princeton University Press, Princeton, 1950, https://doi.org/10.2307/j.ctt1bc543b.