Single-Projection Procedure for Infinite Dimensional Convex Optimization Problems

Hoa T. Bui∗ Regina S. Burachik† Evgeni A. Nurminski‡ Matthew K. Tam§

April 10, 2024

Abstract

In this work, we consider a class of convex optimization problems in a real Hilbert space that can be solved by performing a single projection, i.e., by projecting an infeasible point onto the feasible set. Our results improve those established for the linear programming setting in Nurminski (2015) by considering problems that: (i) may have multiple solutions, (ii) do not satisfy strict complementary conditions, and (iii) possess non-linear convex constraints. As a by-product of our analysis, we provide a quantitative estimate on the required distance between the infeasible point and the feasible set in order for its projection to be a solution of the problem. Our analysis relies on a “sharpness” property of the constraint set; a new property we introduce here.

Keywords. linear programming, polytopes and polyhedral sets, convex programming, Hilbert space, projection method, sharpness property, subtransversality.

MSC2020. 90C05, 90C25, 49J53, 49J52

1 Introduction

Let \( \mathcal{H} \) be a real Hilbert space with inner product \( \langle \cdot, \cdot \rangle \) and induced norm denoted \( \|x\| = \sqrt{\langle x, x \rangle} \). Consider a problem of the form

\[
\min_{x \in A} \langle x^*, x \rangle, \tag{P}
\]

where \( A \subseteq \mathcal{H} \) is a nonempty, closed and convex set and \( \langle x^*, \cdot \rangle \) is a linear function \( (x^* \neq 0) \). Without loss of generality, we assume that \( \|x^*\| = 1 \). In [17], the author shows that if \( \mathcal{H} \) is finite dimensional, \( A \) is a polyhedron and strict complementarity conditions hold, then the linear programming problem (P) has the property:

For every \( x^0 \in \mathcal{H} \), there exists \( \theta_0 > 0 \) such that the projection onto \( A \) of the point \( x^0 - \theta x^* \) (which is a steepest descent step from the initial guess \( x^0 \)) solves (P) for any \( \theta \geq \theta_0 \).

We refer to the aforementioned procedure as the single projection procedure (SPP) for Problem (P). Since the publication of [17], the SPP for solving linear programming in finite dimensions has been considered in [3, 5], as a consequence of the finite convergence property of the alternating projection method performed on a polyhedral set and a closed half space in the case when these sets are not intersecting.

The aims of the present paper can be summarized as follows: (i) extend the SPP to the infinite dimensional setting, (ii) show that the SPP remains valid in more general settings including non-linear convex constraints, non-unique solutions and/or in the absence of the strict complementary property, and (iii) provide quantitative estimates on the value of \( \theta_0 \) needed for the SPP to work. To this end, we show that the SPP is valid for problems satisfying a new property called sharpness (see Figure 1) which we introduce here. In the particular case of problem (P), the sharpness property holds when there is a positive lower bound for the distance between the (unit) vector \(-x^*\) and the normal cone \(N_A(\cdot)\) at every non-optimal point of Problem (P) (see Definition 3.2). In this context, Problem (P) can be solved by the SPP (see Theorem 4.5 and Lemma 4.6) whenever \( \theta \) is sufficiently large and \( x^0 \) is sufficiently far from being optimal in the sense that

\[
\langle x^*, x^0 - \theta x^* \rangle = \langle x^*, x^0 \rangle - \theta < \inf_{x \in A} \langle x^*, x \rangle.
\]
As stated in the introduction, we start this section by setting the theoretical framework and recalling the standard definitions for future use. Finally, Section 5 lists some open questions and discussion.

2 Preliminaries

We start this section by setting the theoretical framework and recalling the standard definitions for future use. As stated in the introduction, \( H \) is a real Hilbert space with inner-product \( \langle \cdot, \cdot \rangle \) and induced norm \( \| \cdot \| \). Given a set \( C \subset H \), the distance from \( C \) to \( x \) is denoted \( d_C(x) := \inf_{z \in C} \| x - z \|. \)

Furthermore, as polyhedral sets are sharp with respect to every unit vector (see Proposition 3.12), our analysis shows that every solvable linear programming problem can be solved by the SPP. As a consequence, this work extends the main results of [17, Theorem 1] to Hilbert spaces. The setting of potentially convex objective functions is dealt with in Theorems 4.10 and 4.13.

The remainder of the present paper is organized as follows. In Section 2, we provide essential results that will be used in subsequent sections. Section 3 formally introduces the new notion of the sharpness property and its connections with some existing geometry properties of sets. Section 4 contains our main results on SPP for solving Problem (P), and its extension to general convex problems in which the objective function is not necessary linear. Finally, Section 5 lists some open questions and discussion.

2 Preliminaries

As mentioned above, our analysis relies on a new notion of “sharpness” for sets, which this work also studies in its own right in Section 3. Namely, a set \( A \) is sharp with respect to \( x^* \) if and only if

\[
\inf_{x \in A \setminus F_A(x^*)} d(x^*, N_A(x)) > 0,
\]

where \( F_A(x^*) \) denotes the face of \( A \) defined by a point \( x^* \) (see Definition 3.1). In order to contextualize sharpness in the broader literature, we explore the property from three perspectives:

- **Sharpness of the epigraph:** When the set \( A \) is the epigraph of a convex function, we provide characterizations of sharpness in terms of its subdifferential. This includes establishing that the epigraph of a function is sharp with respect to the vector \((0_H, -1)\) if and only if it satisfies a global Kurdyka-Lojasiewicz \((KL)\) property with exponent 0 (Proposition 3.14). Moreover, we show that a set \( A \) is sharp with respect to \( x^* \) if and only if the function \( \mathbb{1}_A(\cdot) - \langle x^*, \cdot \rangle \) has the global KL property with exponent 0, where \( \mathbb{1}_A(\cdot) \) is the indicator function of the set \( A \).

- **Dual characterizations:** The analysis in [17] relies on the presence of strict complementary. This condition, which is often not satisfied, is known to imply uniqueness of solutions of the linear optimization problem, and that the interior of the normal cone of the feasible set at this optimal solution contains the vector \(-x^*\). On the other hand, we show in Proposition 3.18 that sharpness holds under much weaker conditions, in particular we allow non-unique optimal solutions. Namely, we require only that the union of the normal vectors of \( A \) at all optimal solutions must contain \(-x^*\). Consequently, although the strict complementary condition can easily fail for general linear programming problems, our sharpness condition holds for every polyhedral set and for every nonzero vector \( x^* \) (see Proposition 3.12).

- **Metric characterizations:** We show that sharpness with respect to \( x^* \) is equivalent to a subtransversality property between the set \( A \) and its supporting hyperplane at \( x^* \) (see Corollary 3.26). As such, the sharpness property can be connected with well-known geometric properties in the literature.

Figure 1: The single-projection procedure extends beyond linear programming. As shown in Figure 1(a), the set \( A \) does not need to be polyhedral for the procedure to work, but the sharpness property at \( x^* \) is required (see Definition 3.2).

\[
\begin{array}{c}
\text{(a) Given a vector } x^0, \text{ we can find } \theta \text{ large enough such that the projection of the vector } x^0 - \theta x^* \text{ onto } A \text{ solves } (P). \text{ This set } A \text{ is sharp w.r.t. } x^*.
\end{array}
\]

\[
\begin{array}{c}
\text{(b) For } x^0 \text{ as in the figure, there is no } \theta > 0 \text{ such that the projection of the vector } x^0 - \theta x^* \text{ onto } A \text{ solves } (P). \text{ This set } A \text{ is not sharp w.r.t. } x^*.
\end{array}
\]
We use the notation $\mathbb{R}_\infty := \mathbb{R} \cup \{\infty\}$. Given $A \subset \mathcal{H}$, we denote by $\text{int} A$, $\text{cl} A$, $\text{bd} A$, its topological interior, closure, and boundary, respectively. Unless specifically mentioned, we consider the strong topology in $\mathcal{H}$. We will denote by $\mathcal{B} := \{ u \in \mathcal{H} : \|u\| \leq 1 \}$ the closed unit ball in $\mathcal{H}$, and by $\mathcal{S} := \{ u \in \mathcal{B} : \|u\| = 1 \}$ the boundary of $\mathcal{B}$. Consequently, the open unit ball is $\mathcal{B} \setminus \mathcal{S} = \{ u \in \mathcal{B} : \|u\| < 1 \}$. Therefore, the open ball of radius $r > 0$ and center $x_0 \in \mathcal{H}$ is $x_0 + r (\mathcal{B} \setminus \mathcal{S})$, and the closed ball of radius $r > 0$ and center $x_0 \in \mathcal{H}$ is $x_0 + r \mathcal{B}$.

We will consider the product space $\mathcal{H}^2 = \mathcal{H} \times \mathcal{H}$ with the max norm $\| \cdot \|_\infty$. Namely, given $(x, y) \in \mathcal{H}^2$, we consider $\|(x, y)\|_\infty := \max\{\|x\|, \|y\|\}$. Given this norm in $\mathcal{H}^2$, it is well known that its dual norm (i.e., the norm in the dual space of $(\mathcal{H}^2, \| \cdot \|_\infty)$) is the norm

$$\|(x, y)\|_* := \|x\| + \|y\|.$$  

For the closed unit ball in $\mathcal{H}^2$ induced by the sum norm, we will use the notation $B^*_H$. Namely,

$$B^*_H := \{(u, v) \in \mathcal{H}^2 : \|u\| + \|v\| \leq 1 \}.$$  

**Definition 2.1.** Let $C \subset \mathcal{H}$ be a nonempty, closed and convex set. Let $x \in \mathcal{H}$. By [2, Theorem 3.16], there exists a unique element in $C$ that minimizes the distance from $C$ to $x$. We denote this element by $P_C(x)$ (it is also called the best approximation to $x$ from $C$). Namely,

$$d_C(x) = \inf_{z \in C} \|z - x\| = \|x - P_C(x)\|.$$  

When convenient, we may also use the notation $d(x, C) := d_C(x)$.

Let $f : \mathcal{H} \to \mathbb{R}_\infty$. The set $\text{dom} f := \{ x \in \mathcal{H} : f(x) < \infty \}$ is the domain (or effective domain) of $f$, and the function $f^* : \mathcal{H} \to \mathbb{R}_\infty$ defined by

$$f^*(x^*) := \sup_{x \in \mathcal{H}} \{ \langle x, x^* \rangle - f(x) \},$$

is the Fenchel conjugate of $f$ at $x^* \in \mathcal{H}$. The epigraph of $f$ is $\text{epi} f := \{ (x, r) \in \mathcal{H} \times \mathbb{R} | f(x) \leq r \}$. We say that $f : \mathcal{H} \to \mathbb{R}_\infty$ is proper if $\text{dom} f \neq \emptyset$. A function $f$ is said to be (strongly) lower semicontinuous (lsc) when its epigraph is (strongly) closed. In the latter situation, we say that $f$ is closed. If $f$ is convex with a closed epigraph, then $f$ is also weakly lsc (i.e., $\text{epi} f$ is closed in the weak topology). Given a convex function $f$, recall that the subdifferential of $f$ is the point-to-set mapping $\partial f : \mathcal{H} \rightrightarrows \mathcal{H}$ defined by

$$\partial f(x) := \begin{cases} \{ x^* \in \mathcal{H} : (\forall y \in \mathcal{H}) \langle y - x, x^* \rangle + f(x) \leq f(y) \} & \text{if } x \in \text{dom} f; \\ \emptyset & \text{if } x \notin \text{dom} f. \end{cases}$$  

Note that for points at the boundary of $\text{dom} f$, the subdifferential may or may not exist (i.e., the set in the first line in (3) may be empty).

**Definition 2.2.** Given a point-to-set map $T : \mathcal{H} \rightrightarrows \mathcal{H}$, we consider the following sets.

(a) The **domain of** $T$ is the set

$$D(T) := \{ x \in \mathcal{H} : T(x) \neq \emptyset \}. $$

(b) The **range of** $T$ is the set

$$R(T) := \{ v \in \mathcal{H} : v \in T(x) \text{ for some } x \in \mathcal{H} \}. $$

(c) The **graph of** $T$ is the set

$$G(T) := \{ (x, v) \in \mathcal{H} \times \mathcal{H} : v \in T(x) \}. $$

For a fixed nonzero vector $u \in \mathcal{H}$, we will denote by $\mathbb{R}_+ u := \{ tu : t \geq 0 \} = \text{cone} [u]$, the cone (and also called the ray) generated by $u$. Given $J$ a nonempty set, and a collection of elements $(u_i)_{i \in J} \subset \mathcal{H}$ indexed by $J$, we denote by $\text{cone} [u_i, i \in J]$ the convex cone generated by the collection. If $J = \{1, \ldots, l\}$ is finite, then

$$\text{cone} [u_1, \ldots, u_l] = \sum_{i=1}^l \text{cone} [u_i] = \sum_{i=1}^l \mathbb{R}_+ u_i, $$

where we are using the fact that $\text{cone} [A \cup B] = \text{cone} [A] + \text{cone} [B]$. Note that in these definitions we are using the notation $\text{cone} [C]$ for the convex cone generated by a set $C$.

**Definition 2.3.** Given a subset $A \subset \mathcal{H}$ and the point $x \in \mathcal{H}$, the point-to-set map $N_A : \mathcal{H} \rightrightarrows \mathcal{H}$ defined by

$$N_A(x) := \begin{cases} \{ x^* \in \mathcal{H} : (\forall a \in A) \langle a - x, x^* \rangle \leq 0 \} & \text{if } x \in A; \\ \emptyset & \text{otherwise}, \end{cases}$$

is called the normal cone of a set $A$ at the point $x$. 

3
We will make use of the following result, which is [2, Proposition 16.35].

**Fact 2.1.** Let \( f : \mathcal{H} \to \mathbb{R}_\infty \) be a convex, proper and lsc function. Denote by \( D := \text{dom } f \) and by \( E := \text{epi } f \). Fix \( x \in D \). Then

\[
N_E(x, f(x)) = \mathbb{R}_+(\partial f(x) \times \{-1\}) \bigcup (N_D(x) \times \{0\}),
\]

where \( \mathbb{R}_+ V := \{ tv : t > 0, v \in V \} \), denotes the positive cone generated by \( V \). Equivalently,

\[
N_E(x, f(x)) = \{(u, \eta) \in \mathcal{H} \times \mathbb{R}_- : \eta < 0 \text{ and } u/(-\eta) \in \partial f(x), \text{ or } \eta = 0 \text{ and } u \in N_D(x)\}. \tag{5}
\]

By [2, Proposition 16.17(i)], for every \( x \in \text{bd} D \), we can either have \( \partial f(x) = \emptyset \) or \( \partial f(x) \) unbounded. In the latter case, i.e., when \( x \in \text{bd} D \) and \( \partial f(x) \neq \emptyset \), we have

\[
\partial f(x) + N_D(x) \subset \partial f(x).
\]

Recall that the indicator function of a subset \( A \subset \mathcal{H} \) is the function \( \mathbb{1}_A : \mathcal{H} \to \mathbb{R}_\infty \) such that \( \mathbb{1}_A(x) = 0 \) when \( x \in A \) and \( \mathbb{1}_A(x) = +\infty \) otherwise. Hence, \( \text{dom } \mathbb{1}_A = A \). If \( A \) is a nonempty, closed and convex set, \( \mathbb{1}_A \) is a proper, lsc and convex function. In this situation, (3) yields

\[
\partial (\mathbb{1}_A(x)) = N_A(x) \text{ for all } x \in \mathcal{H},
\]

and note that \( \partial (\mathbb{1}_A(x)) = \emptyset \) otherwise. In this case, we can use the maximal monotonicity of the map \( \partial (\mathbb{1}_A)(\cdot) \) to deduce that the graph of \( N_A \), given by the set

\[
G(N_A) := \{(x, x^*) \in A \times \mathcal{H} : x^* \in N_A(x)\},
\]

is closed w.r.t. the strong-weak topology (i.e., w.r.t. the strong topology in the first coordinate and w.r.t. the weak topology in the second coordinate). We call this type of closedness *demi-closedness* (note that \( G(N_A) \) is also closed when considering the weak topology in the first coordinate and the strong topology in the second one).

**Remark 2.2.** Fix \( z_0 \in \mathcal{H} \). Consider the function \( \varphi_{z_0} : \mathcal{H} \to \mathbb{R} \) defined as \( \varphi_{z_0}(x) := \|x - z_0\| \) where \( \| \cdot \| \) is any given norm in \( \mathcal{H} \). In this situation, the Fenchel-conjugate of \( \varphi_{z_0} \) is given by \( \varphi_{z_0}^* = \mathbb{1}_{\mathcal{B}_s} + (z_0, \cdot) \) where \( \mathcal{B}_s \) is the dual unit ball, i.e., the unit ball with respect to the norm which is dual to the given norm \( \| \cdot \| \). Moreover, by [2, Example 16.62] (see also [21, Corollary 2.4.16]):

\[
\partial \varphi_{z_0}(x) = \begin{cases} \langle x - z_0, \|x - z_0\| \rangle & \text{if } x \neq z_0, \\ \mathcal{B}_s & \text{if } x = z_0. \end{cases} \tag{7}
\]

Consequently, \( \varphi_{z_0} \) is smooth (i.e., the set \( \partial \varphi_{z_0}(x) \) is a singleton) at every \( x \neq z_0 \) see [2, Proposition 17.32].

**Remark 2.3.** For future use, we recall here a fact involving the subdifferential of a maximum of two norms in the product space. Fix \( z = (z_1, z_2) \in \mathcal{H} \times \mathcal{H} \). Consider the function \( \Theta_z : \mathcal{H} \times \mathcal{H} \to \mathbb{R} \) defined as

\[
\Theta_z(x, y) := \max\{\|x - z_1\|, \|y - z_2\|\} = \max\{\varphi_{z_1}(x), \varphi_{z_2}(y)\},
\]

where we are using the notation introduced in Remark 2.2 in the second equality. Note that

\[
\Theta_z(z_1, z_2) = 0 = \varphi_{z_1}(z_1) = \varphi_{z_2}(z_2). \tag{8}
\]

Denote by \( \partial_1 \) and \( \partial_2 \) the partial subdifferentials w.r.t. the first and second variables, respectively. Define also \( \Theta^1(x, y) := \varphi_{z_1}(x) \) for every \( x, y \in \mathcal{H} \) and \( \Theta^2(x, y) := \varphi_{z_2}(y) \) for every \( x, y \in \mathcal{H} \). So that \( \Theta_z(x, y) = \max\{\Theta^1(x, y), \Theta^2(x, y)\} \). Using Remark 2.2 and the fact that \( \Theta^2 \) does not depend on the first variable, we have that

\[
\partial \Theta^1(z_1, z_2) = \left( \partial_1 \Theta^1(z_1, z_2), \partial_2 \Theta^1(z_1, z_2) \right) = \left( \partial_1 \Theta^1(z_1, z_2), 0 \right) = \left( \partial \varphi_{z_1}(z_1), 0 \right) = \mathcal{B}_s \times \{0\},
\]

and where we used (7) in the last equality. Similarly, we obtain

\[
\partial \Theta^2(z_1, z_2) = \left( \partial_1 \Theta^2(z_1, z_2), \partial_2 \Theta^2(z_1, z_2) \right) = \{0\} \times \mathcal{B}_s.
\]

The classical Theorem of Dubovitskii and Milyutin (see [2, Theorem 18.5]), which computes the subdifferential of a maximum of functions, together with (8), yield

\[
\partial \Theta_z(z_1, z_2) = \text{conv} \{\mathcal{B}_s \times \{0\}, \{0\} \times \mathcal{B}_s\} = \{(u, v) \in \mathcal{H} \times \mathcal{H} : \|u\| + \|v\| \leq 1\} = \text{conv} \{\mathcal{B}_s \times \mathcal{B}_s\}.
\]

(Here, \( \text{conv}(A) \) denotes the *convex hull* of the set \( A \), which is the smallest convex set that contains \( A \); whereas \( \text{conv}(A) \) is the closure of its convex hull.) The second equality can be easily checked, while in the last equality we use the notation introduced in (2).
Definition 2.4. Fix $m \in \mathbb{N}^*$ and $b \in \mathbb{R}^m$. Let $A : \mathcal{H} \to \mathbb{R}^m$ be a bounded linear operator. Recall that the adjoint operator of $A$ is the (bounded) linear map $A^* : \mathbb{R}^m \to \mathcal{H}$ defined by the equality

$$
\langle Ax, u \rangle = \langle x, A^*u \rangle \quad \forall (x, u) \in \mathcal{H} \times \mathbb{R}^m.
$$

(9)

Note that, in (9), we are using the same notation for the inner-products in $\mathcal{H}$ and in $\mathbb{R}^m$. From [7, Remark 16], we have that, when $A$ is bounded, $A^*$ is also bounded and both maps have the same norm.

Definition 2.5. Take $A$ and $b$ as in Definition 2.4. Denote by $B_m := \{e_1, \ldots, e_m\}$ the canonical basis in $\mathbb{R}^m$. A polyhedron or polyhedral set induced by a linear map $A$ and a vector $b$ is the set

$$
C(A, b) := \{ x \in \mathcal{H} : Ax \leq b \} := \{ x \in \mathcal{H} : \langle Ax, e_j \rangle \leq b_j \text{ for all } j = 1, \ldots, m \}.
$$

(10)

Hence, a polyhedral set is a finite intersection of level sets of linear maps. Given $x \in \mathcal{H}$, the set $I(x) := \{ j \in \{1, \ldots, m\} : \langle Ax, e_j \rangle = b_j \}$ identifies the inequality constraints that are active at $x$. We say that a function $f : \mathcal{H} \to \mathbb{R}_\infty$ is polyhedral when its epigraph is a polyhedral subset of $\mathcal{H} \times \mathbb{R}$.

The normal cone to a polyhedral set will have an important role in our analysis, so we recall [15, Corollary 4.1], valid in a locally convex space. Polyhedral sets in these general spaces are defined in a similar manner as in Definition 2.5, where $\mathcal{H}$ is replaced by a locally convex space denoted by $X$.

Proposition 2.4. ([15, Corollary 4.1]) Let $X$ be a locally convex space, and let $C_1, \ldots, C_m$ be polyhedral subsets of $X$ with $C := \bigcap_{i=1}^m C_i \neq \emptyset$. Then, for all $x \in C$, we have

$$
N_C(x) = \sum_{i=1}^m N_{C_i}(x).
$$

We recall Ekeland’s Variational Principle, which holds in the setting of metric spaces.

Lemma 2.5 (Ekeland’s Variational Principle [10, Theorem 1.1]). Let $X$ be a complete metric space, $\psi : X \to \mathbb{R}_\infty$ be lsc, $\bar{w} \in X$ and $\varepsilon > 0$. If

$$
\psi(\bar{w}) < \inf_{w \in X} \psi(x) + \varepsilon,
$$

then, for any $\lambda > 0$, there exists $\bar{w}_\lambda \in X$ such that

1. $d(\bar{w}_\lambda, \bar{w}) < \lambda$;
2. $\psi(\bar{w}_\lambda) \leq \psi(\bar{w})$;
3. $\psi(w) + (\varepsilon/\lambda)d(w, \bar{w}_\lambda) \geq \psi(\bar{w}_\lambda)$ for all $w \in X \setminus \{\bar{w}_\lambda\}$.

The next simple fact will be used in later sections.

Fact 2.6. Let $u, v \in \mathcal{H}$ be such that $\|u\| = \|v\| = 1$. Then $d(u, \text{cone}[v])^2 = 1 - \max(0, \langle u, v \rangle)^2$.

Proof. Denote $t_* = \langle u, v \rangle$. Then,

$$
d(u, \text{cone}[v])^2 = \min_{t \geq 0} \|u - tv\|^2 = \min_{t \geq 0} (1 + t^2 - 2t_*t) = 1 - \max_{t \geq 0} (-t^2 + 2t_*t) = 1 - \max(0, t_*)^2.
$$

\[ \square \]

3 Sharp sets and their characterizations

Figure 1 illustrates that, for the SPP to work, the set $A$ must possess certain geometric properties. In this section, we formally define a property that allows problem (P) to be solved by the SPP. Our focus is on a geometric property associated with the presence of “sharp corners” of a convex subset of a Hilbert space.

3.1 Definition and examples

Definition 3.1. We define the face of a convex set $A \subset \mathcal{H}$ with respect to a nonzero vector $x^*$ as

$$
F_A(x^*) := \{ x \in A : \langle x^*, x \rangle = \sup_A \langle x^*, \cdot \rangle \} = \text{Argmax}_A \langle x^*, \cdot \rangle.
$$

A closed convex subset $F^0 \subset A$ is said to be an exposed face of $A$ if there is $x^* \in \mathcal{H}$ such that $F^0 = F_A(x^*)$.

The next definition will be crucial in our analysis.
Definition 3.2. Let $A \subseteq \mathcal{H}$ be closed and convex, $x^* \in \mathcal{S}$ and \( \alpha > 0 \). We say that $A$ is $\alpha$-sharp with respect to $x^*$ if, for all $x \in A$ such that $x^* \notin N_A(x)$, we have

\[ d(x^*, N_A(x)) \geq \alpha. \]  

The modulus of sharpness of $A$ w.r.t. $x^*$, denoted $\text{sr}[A, x^*]$, is defined as

\[ \text{sr}[A, x^*] := \inf_{x^* \notin N_A(x)} d(x^*, N_A(x)). \]

We say that $A$ is sharp w.r.t. vector $x^*$ if $\text{sr}[A, x^*] > 0$.

Remark 3.1. From the definition, $A$ is $\alpha$-sharp w.r.t. to the vector $x^*$ if and only if $\text{sr}[A, x^*] \geq \alpha$. The definition of sharpness involves taking an infimum over the set $\{ x \in A : x^* \notin N_A(x) \}$. When the latter set is empty, we have that $x^* \in N_A(x)$ for every $x \in A$. This implies that $\langle x^*, x \rangle = \sup_{x \in A} \langle x^*, \cdot \rangle$ for all $x \in A$. Hence, $A = F_A(x^*)$.

In this case, $A$ is trivially sharp w.r.t. $x^*$ by vacuity, and since we are taking the infimum over the empty set in (12), we deduce that $\text{sr}[A, x^*] = +\infty$. We also note that, since $\|x^*\| = 1$ and $0 \in N_A(x^*)$, we always have that $\alpha \leq 1$.

The following result relates faces of sets with the sharpness property in Definition 3.2.

Fact 3.2. Let $A$ be a closed convex set. Fix $x \in A$ and $x^* \in \mathcal{S}$. The following statements are equivalent.

(i) $x^* \in N_A(x)$.

(ii) $x \in F_A(x^*)$.

(iii) $x \in \text{Argmax}_A \langle x^*, \cdot \rangle$.

Consequently, for any $\alpha \geq 0$, we have

\[ \text{sr}[A, x^*] \geq \alpha \iff d(x^*, N_A(x)) \geq \alpha, \forall x \in A \setminus F_A(x^*). \]

Proof. Using the definitions and the notation of Definition 3.2, we can write

\[ x^* \in N_A(x) \iff \langle x^*, y - x \rangle \leq 0, \forall y \in A \iff \langle x^*, x \rangle = \sup_{y \in A} \langle x^*, y \rangle \iff x \in F_A(x^*), \]

which proves the equivalence between (i) and (ii) in the first statement. The equivalence between (ii) and (iii) in the first statement follows from the fact that $F_A(x^*) = \text{Argmax} \langle x^*, \cdot \rangle$. To complete the proof, note that the equivalence between (i) and (ii) implies

\[ x \in A \setminus F_A(x^*) \iff x \in A \text{ and } x^* \notin N_A(x). \]

Therefore,

\[ \text{sr}[A, x^*] \geq \alpha \iff d(x^*, N_A(x)) \geq \alpha, \forall x \in A, x^* \notin N_A(x), \]

\[ \iff d(x^*, N_A(x)) \geq \alpha, \forall x \in A \setminus F(x^*), \]

establishing the last claim.

Remark 3.3. Figure 2 illustrates two-dimensional examples of the sharpness property with respect to a given vector, and its connection with faces of the set $A$. In Figure 2(a), the “rounded” section of the boundary of $A$ approaches $\bar{x}$ smoothly. We note, however, that sharpness can also fail for sets whose boundary is not rounded as in Figure 2. We illustrate this situation in the next example.

\[ f(x) := \begin{cases} 
+\infty & \text{for } x < -1, \\
\frac{(3n^2 + 3n + 1)x + (2n + 1)}{n^2(n + 1)^2} & x \in [-1/n, -1/(n + 1)], n \in \mathbb{N}, \\
0 & x \geq 0.
\end{cases} \]

The boundary of the set $A$ to the left of 0 is determined by a piece-wise linear function which at the points $a_n := -1/n$ attains the value $1/n^3$ and is linear between $a_n$ and $a_{n+1}$ ($n = 1, 2, \ldots$). This function is non-smooth because it has kinks at each $a_n$. It is convex because its slopes monotonically increase to zero. The latter fact also implies that $A$ is not sharp w.r.t $x^* := (0, -1)$. Therefore, a set with nonsmooth boundary need not satisfy the sharpness condition.
Remark 3.5. When a closed and convex set $A$ is bounded, the lack of sharpness implies a continuity property of the restriction of the point-to-set map $N_A(\cdot)$ to the boundary of $A$. Indeed, if $\text{sr}[A, x^*] = 0$, then the definition directly implies the existence of a sequence $(x_n, x_n^*) \subset G(N_A) \cap x^* \notin N_A(x_n)$ and $x_n^* \rightarrow x^*$ strongly. The boundedness of $A$ implies that there exists $\bar{x}$ and a subsequence $(x_{n_k})$ of $(x_n)$ weakly converging to $\bar{x}$. For simplicity, we still call this subsequence $(x_n)$. Since $G(N_A)$ is semi-closed, $\bar{x}, x^* \in G(N_A)$. The latter fact, together with $x^* \notin N_A(x_n)$, implies that $x_n \neq \bar{x}$ for all $n$. Altogether, we have a sequence $(x_n, x_n^*) \subset G(N_A)$ such that

(i) $x_n^* \rightarrow x^*$ strongly with $x^* \notin N_A(x_n)$,

(ii) there exists $\bar{x} \in A$ such that $x_n \rightarrow \bar{x}$ weakly, with $x_n \neq \bar{x}$ for all $n$.

In particular, in Figure 2(a), for any positive number $r > 0$, there is $y^* \in B_r(x^*)$ and $y \in B_r(\bar{x}) \setminus F$ such that $y^* \in N_A(y)$. Note here that $\bar{x}$ is an extreme point of the set $A$ in both Figures 2(a), and (b). However, the point $\bar{x}$ in Figure 2(a) is an exposed point (i.e., there exists a hyperplane $H$ such that $A \cap H = \{\bar{x}\}$), whereas the point $\bar{x}$ in Figure 2(a) is not an exposed point (since the only supporting hyperplane of $A$ at $\bar{x}$ is $F$). In Figure 2(b) it is easy to note that, there is $\alpha \in (0,1)$ such that for every $y \in A$ with $x^* \notin N_A(y)$ we will have that $d(x^*, N_A(y)) \geq \alpha$.

Figure 2: An illustration of the sharpness property. In both figures, $F$ is a supporting hyperplane of $A$ at $\bar{x}$. The set $A$ is not sharp w.r.t. the vector $x^*$; there exist a sequence $(x_n, x_n^*) \subset \text{bd } A \setminus H$ with $x_n^* \neq x_n^*$ $(n \in \mathbb{N})$ such that $(x_n)$ converges to $\bar{x}$, and $(x_n^*)$ converges to $x^*$.

Remark 3.6. From Fact 3.2(ii), we know that $x^* \notin R(N_A)$ if and only if $F_A(x^*) = \emptyset$. We note that this situation cannot hold when $A$ is bounded. Indeed, in this case we can use [2, Corollary 21.25] to obtain $R(N_A) = \mathcal{H}$. Namely, there is no $x^*$ verifying $x^* \notin R(N_A)$. We also note that, in general, the set $R(N_A)$ is neither closed nor convex. However, it is well known (see, e.g., [2, Corollary 21.25] or [6, Theorem 4.4.9]) that $\text{cl } R(N_A)$ is a closed and convex set.

For an unbounded set $A$, we next characterize for which $x^* \notin R(N_A)$ we have that $A$ is sharp w.r.t. $x^*$.

Proposition 3.7. Fix $x^* \in S$ and assume that $A \subset \mathcal{H}$ is a closed and convex set such that $x^* \notin R(N_A)$. Then $A$ is sharp w.r.t. $x^*$ if and only if $x^* \notin \text{bd } R(N_A)$. In this situation, $A$ must be unbounded.

Proof. The fact that $A$ is unbounded follows from Remark 3.6. We show first that, if $x^* \in \text{bd } R(N_A)$ then $A$ is not sharp w.r.t. $x^*$. Indeed, in this case, we can take a sequence $(x_n) \subset A$ such that $x_n \neq x_n^* \in N_A(x_n)$ with $x^* = \lim_n x_n^*$. The fact that $x^* \neq x_n^*$ holds because $x^* \notin R(N_A)$. Hence,

$$\inf_{x \in A, x^* \notin N_A(x)} d(x^*, N_A(x)) \leq \lim_{n \to \infty} d(x^*, x_n^*) = 0,$$

so $A$ is not sharp w.r.t. $x^*$. Conversely, assume that $x^* \notin \text{bd } R(N_A)$. Since we also have that $x^* \notin R(N_A)$, we deduce that $x^* \notin \text{cl } R(N_A)$. Then, there exists $\alpha > 0$ such that

$$\alpha = d(x^*, \text{cl } R(N_A)) = \inf_{v \in R(N_A)} d(x^*, v) \leq \inf_{x \in A, x^* \notin N_A(x)} d(x^*, N_A(x)),$$

so $A$ is sharp w.r.t. $x^*$.

We illustrate the last result in the next two examples. In the first example, we have $x^* \notin R(N_A)$ but $x^* \in \text{bd } R(N_A)$. While in the second example $x^* \notin \text{cl } R(N_A)$.
Example 3.8. If $A := \text{epi} g$ where $g : \mathbb{R} \to \mathbb{R}_\infty$ defined as $g(t) = 1/t$ if $t > 0$ and $g(t) = +\infty$ otherwise. Then it is easy to check that $x^* := (0, -1) \notin R(N_A)$ but $x^* \in \text{bd} R(N_A)$. Hence, we are not in the conditions of Proposition 3.7. Let us check that $A$ is not sharp with respect to $x^*$. Indeed, for $x_n := (n, 1/n)$ we have 
$$x^* = (0, -1) = \lim_{n \to \infty} (-1/n^2, -1),$$
so $\lim_{n \to \infty} d(x^*, N_A(n, 1/n)) = 0$ and hence $A$ is not sharp with respect to $x^*$.

Example 3.9. Take $A = \mathbb{R}_+^2$, the nonnegative orthant in $\mathbb{R}^2$ and $x^* = (1/\sqrt{2}, 1)$. Then $x^* \notin \text{cl} R(N_A)$ and we are in the conditions of Proposition 3.7. It can be verified that $A$ is 1-sharp w.r.t. $x^*$. Indeed, if $x \in \text{int} A$ then $N_A(x) = \{0\}$ so $d(x^*, N_A(x)) = d(x^*, 0) = 1$. If $x = t(1, 0)$ for $t > 0$, then $N_A(x) = \{s(0, -1) : s \geq 0\}$.

Using Lemma 2.6 with $u := x^*$ and $v := (0, -1)$, we obtain $d(x^*, N_A(x)) = 1$. Identical argument shows that $d(x^*, N_A(x)) = 1$ for $x = t(0, 1)$ for $t > 0$. In the latter case we use Lemma 2.6 with $u := x^*$ and $v := (-1, 0)$. Finally, if $x = (0, 0)$ then $N_A(x) = \mathbb{R}_+^2$ so again we have $d(x^*, N_A(x)) = d(x^*, \mathbb{R}_+^2) = 1$.

Proposition 3.10. Suppose $A$ is a nonempty bounded closed convex set. Fix $x^* \in S$ and $\alpha \in (0, 1)$. Then, $A$ is $\alpha$-sharp w.r.t. $x^*$ if and only if, for every $y^* \in H \setminus \{0\}$ such that $F_A(y^*) \cap F_A(x^*) = \emptyset$, we have

$$\|x^* - y^*\| \geq \alpha. \quad (13)$$

Consequently, 

$$\text{sr}[A, x^*] = \inf_{y^* \neq 0} \|x^* - y^*\|. \quad (14)$$

Proof. Note that because the set $A$ is nonempty and bounded, from [2, Corollary 21.25], the set $\text{Argmax}_A (y^*, \cdot)$ is nonempty for any $y^* \in H \setminus \{0\}$, therefore from Fact 3.2, the face $F_A(y^*)$ is also nonempty.

First suppose that the set $A$ is $\alpha$-sharp w.r.t. $x^* \in S$. If $F_A(x^*) \cap F_A(y^*)$ is nonempty for every $y^* \in H \setminus \{0\}$, then equation (13) in Proposition 3.10 holds true immediately. Otherwise, assume that there exists $y^* \in H \setminus \{0\}$ such that $F_A(y^*) \cap F_A(x^*) = \emptyset$. By Fact 3.2, for any $y \in F_A(y^*)$, we have that $y^* \in N_A(y)$. Because $F_A(y^*) \cap F_A(x^*) = \emptyset$, we must have $y \notin F_A(x^*)$, and again Fact 3.2 yields $x^* \notin N_A(y)$. Then, the $\alpha$-sharpness property implies

$$\|x^* - y^*\| \geq d(x^*, N_A(y)) \geq \inf_{y^* \in A, y^* \notin N_A(y)} d(x^*, N_A(y')) = \text{sr}[A, x^*] \geq \alpha, \quad (15)$$

which shows (13).

Conversely, suppose that for every $y^* \in H \setminus \{0\}$ satisfying $F_A(y^*) \cap F_A(x^*) = \emptyset$, it holds that $\|x^* - y^*\| \geq \alpha$. We now show that the set $A$ must be $\alpha$-sharp w.r.t. $x^*$. In fact, suppose to the contrary that the set $A$ is not $\alpha$-sharp w.r.t. $x^*$. Then, there is $y \in A$ such that $x^* \notin N_A(y)$ and

$$d(x^*, N_A(y)) < \alpha.$$

Take $y^* \in N_A(y)$ such that $\|x^* - y^*\| < \alpha$, and take $\varepsilon \in (0, \alpha - \|x^* - y^*\|)$. Note that because $x^* \notin N_A(y)$ and $y^* \in N_A(y)$, we have

$$(x^*, y) < \sup_{z \in A} \langle x^*, z \rangle, \quad \text{and} \quad \langle y^*, y - x \rangle \geq 0, \forall x \in A.$$

Then, for every $x \in F_A(x^*)$, we have $\langle x^*, x \rangle = \sup_{z \in A} \langle x^*, z \rangle$, and

$$\langle y^* - \varepsilon x^*, y - x \rangle = \langle y^*, y - x \rangle - \varepsilon \langle x^*, y - x \rangle$$

$$= \langle y^*, y - x \rangle - \varepsilon (\langle x^*, y \rangle - \sup_{z \in A} \langle x^*, z \rangle) > 0.$$

The above expression yields $\langle y^* - \varepsilon x^*, x \rangle < \sup_{z \in A} \langle y^* - \varepsilon x^*, z \rangle$, which implies $x \notin F_A(y^* - \varepsilon x^*)$. Because $x \in F_A(x^*)$ is chosen arbitrarily, we have $F_A(x^*) \cap F_A(y^* - \varepsilon x^*) = \emptyset$. Now we can use the hypothesis (13) for $y^* - \varepsilon x^*$ instead of $y^*$, to obtain

$$\|(y^* - \varepsilon x^*) - x^*\| \geq \alpha.$$

On the other hand, the definition of $\varepsilon$ yields

$$\|(y^* - x^*) - \varepsilon x^*\| \leq \|(y^* - x^*) + \varepsilon x^*\| = \|(y^* - x^* + \varepsilon\| < \alpha,$$
in contradiction with assumption (13). Therefore, we must have $A$ is $\alpha$-sharp w.r.t. $x^*$.

To complete the proof, we need to establish (14). From (15) we deduce by taking infimum that

$$\operatorname{sr}[A, x^*] \leq \inf_{y^* \neq 0} \frac{\|x^* - y^*\|}{F_A(y^*) \cap F_A(x^*) = \emptyset}.$$ 

To prove the opposite inequality, assume that the inequality above is strict, so there is $\gamma$ such that

$$\operatorname{sr}[A, x^*] < \gamma < \inf_{\hat{y}^* \neq 0} \frac{\|x^* - \hat{y}^*\|}{F_A(\hat{y}^*) \cap F_A(x^*) = \emptyset}.$$ 

By definition of $\operatorname{sr}[A, x^*]$ this means that there exists $x \in A$ such that $x^* \notin N_A(x)$ with $d(x^*, N_A(x)) < \gamma$. Thus, we can find $v \in N_A(x)$ s.t. $\|x^* - v\| < \gamma$. We have two possibilities: either $F_A(v) \cap F_A(x^*) = \emptyset$ or $F_A(v) \cap F_A(x^*) \neq \emptyset$. If $F_A(v) \cap F_A(x^*) = \emptyset$ then using the rightmost expression above we deduce that $\|x^* - v\| > \gamma$, contradicting the definition of $v$. Hence, the only possibility is that $F_A(v) \cap F_A(x^*) \neq \emptyset$. In this case, we will again obtain a contradiction. Indeed, as in the earlier part of the proof, take $\varepsilon \in (0, \gamma - \|x^* - v\|)$. Following the same arguments as above we will arrive at $F_A(x^*) \cap F_A(v - \varepsilon x^*) = \emptyset$. Now we can use the rightmost expression above and the definition of $\varepsilon$ to deduce that

$$\gamma < \|x^* - (v - \varepsilon x^*)\| < \|v - x^*\| + \varepsilon < \gamma,$$

a contradiction. Hence, we must have that (14) holds. This completes the proof.

\textit{Remark 3.11.} If the set $A$ is unbounded, the statements in Proposition 3.10 need not hold. Indeed, if $x^* \notin R(N_A) \cap \text{fact 3.2 yields} F_A(x^*) = \emptyset$ so $F_A(x^*) \cap F_A(y^*) = \emptyset$ for all $y^* \in H$. In this situation, inequality (13) will not hold for every nonzero $y^* \in H$. This situation is illustrated by Example 3.9. In this example, $x^* = (1/\sqrt{2})(1, 1) \notin R(N_A), F_A(x^*) = \emptyset$, and $A$ is $\alpha$-sharp at $x^*$ with $\alpha = 1$. In particular, $F_A(x^*) \cap F_A(y^*) = \emptyset$ if $y^* = x^*$, in which case (13) is clearly false.

We show next that a polyhedron is sharp with respect to every vector $x^* \in S$.

\textbf{Proposition 3.12.} Let $A$ be a polyhedron defined by $A := \{x : Ax \leq b\}$, where $A$ is a bounded linear operator $A : H \to \mathbb{R}^m (m > 0)$, and $b \in \mathbb{R}^m$. Denote by $B_m := \{e_1, \ldots, e_m\}$ the canonical basis in $\mathbb{R}^m$. For any vector $x^* \in S$, let $J(x^*)$ be a collection of subsets of indexes defined as $J(x^*) := \{J \subset \{1, \ldots, m\} : x^* \notin \text{cone}[A^* e_i]_{i \in J}\}$, and

$$a_0 := \min \left\{1, \inf_{J \in J(x^*)} d(x^*, \text{cone}[A^* e_i]_{i \in J}) \right\}, \quad (16)$$

with the convention that $\inf \emptyset = +\infty$. Then, $a_0 > 0$ and $A$ is $a_0$-sharp w.r.t. $x^*$.

\textit{Proof.} We first prove that the constant $a_0$ defined in (16) is positive. If the set $J(x^*)$ is empty, then (16) becomes $a_0 = \min\{1, \inf \emptyset\} = 1 > 0$. So, it suffices to consider the case where $J(x^*) \neq \emptyset$. To this end, let $J \in J(x^*)$ be arbitrary. By definition, $x^* \notin \text{cone}[A^* e_i]_{i \in J}$ and hence

$$d(x^*, \text{cone}[A^* e_i]_{i \in J}) > 0. \quad (17)$$

From inequality (17) and the fact that the set $J(x^*)$ is finite, the constant $a_0$ in (16) must be positive.

To complete the proof, we show that the constant $a_0$ defined in (16) is a lower bound of all constants $\alpha$ such that $A$ is $\alpha$-sharp w.r.t. $x^*$. In other words, we will show that $\operatorname{sr}[A, x^*] \geq a_0 > 0$. Observe that the polyhedral set $A$ is the intersection of $m$ closed half spaces. Namely, using Definition 2.5 we can write

$$A = \bigcap_{i=1}^m \{x : \langle e_i, Ax \rangle \leq \langle e_i, b \rangle\} = \bigcap_{i=1}^m \{x : \langle A^* e_i, x \rangle \leq \langle e_i, b \rangle\}, \quad (18)$$

where we used (9) in the second equality. Because $A$ is a bounded linear map, so is $A^*$, and hence for each $i = 1, \ldots, m$, the set $H_i := \{x : \langle A^* e_i, x \rangle \leq \langle e_i, b \rangle\}$ is a closed half space. Thus, the normal cone operator of $H_i$ at a point $x \in H_i$ is given by

$$N_{H_i}(x) = \begin{cases} \text{cone}[A^* e_i], & \text{if } \langle A^* e_i, x \rangle = \langle e_i, b \rangle ; \\ \{0\}, & \text{otherwise}. \end{cases}$$

We now apply the intersection rule in Proposition 2.4 and (4) to derive the normal cone of $A$ as

$$N_A(x) = \text{cone}[A^* e_i]_{i \in I(x)} = \sum_{i \in I(x)} \text{cone}[A^* e_i], \quad \forall x \in A, \quad (19)$$
where $I(x) := \{ i : \langle x, A^* e_i \rangle = b_i \}$. Consider any $x \in A$ such that $x^* \not\in N_A(x)$. Then, by (19) and the definition of $J(x)$, we have $I(x) \subseteq J(x)$, and

$$d(x^*, N_A(x)) = d(x^*, \text{cone}[A^* e_i]_{i \in I(x)}) \geq \alpha_0,$$

where we used the definition of $\alpha_0$ in the last inequality. Hence,

$$\alpha_0 \leq \inf_{x^* \not\in N_A(x)} d(x^*, N_A(x)) = \text{sr}[A, x^*],$$

which implies the set $A$ is $\alpha_0$-sharp w.r.t. vector $x^*$.

\[\square\]

### 3.2 Sharpness of the epigraph

Consider the following optimization problem

$$\min f(x),$$

where $A$ is nonempty, closed and convex and $f : \mathcal{H} \to \mathbb{R}_\infty$ is proper, convex and lsc. Assume that (20) has solutions and denote its solution set by $S$. In connection with this problem, we consider the following property.

Assume that there exists $\beta > 0$ such that

$$\inf_{x \in S} d(0, \partial f(x)) \geq \beta > 0.$$  

(21)

**Remark 3.13.** We note that (21) holds, for instance, if $f$ is piece-wise linear. Indeed, let $f(x) := \max_{i = 1, \ldots, m} \langle x_i^*, x \rangle + b_i$. Call $J := \{ i, \ldots, m \}$. For every $x \in \mathcal{H}$, define $I(x) := \{ i \in J : \langle f(x), x_i^* \rangle + b_i \}$. From the formula of the subdifferential of a supremum (see, e.g., [2, Theorem 18.5] or [21, Theorem 2.4.18]), we know that $\partial f(x) = \text{co}[x_i^*]_{i \in I(x)}$. Hence, $0 \not\in \partial f(x)$ if and only if $0 \not\in \text{co}[x_i^*]_{i \in I(x)}$. Consider

$$\mathcal{F} := \{ I \subset J : 0 \not\in \text{co}[x_i^*]_{i \in I} \} \subset \mathcal{P}(J),$$

where $\mathcal{P}(J)$ denotes the collection of all possible subsets of $J$. Recall that $\mathcal{P}(J)$ has cardinality $2^m$. So there is a finite number of possible subdifferential sets such that $0 \not\in \partial f(x)$. Moreover, $x \not\in S$ if and only if $I(x) \in \mathcal{F}$. Altogether,

$$\inf_{x \not\in S} d(0, \partial f(x)) = \inf_{0 \not\in \partial f(x)} d(0, \partial f(x)) \geq \min_{I \in \mathcal{F}} d(0, \text{co}[x_i^*]_{i \in I}),$$

where we have a minimum in the rightmost expression because the infimum is taken over the finite set $\mathcal{F}$. Since $\text{co}[x_i^*]_{i \in I}$ is a closed convex cone which doesn’t contain zero, $d(0, \text{co}[x_i^*]_{i \in I}) > 0$ for every $I \in \mathcal{F}$. Therefore, the minimum in the right-hand side of the above expression is attained at some positive value $\beta$. Note that the fact that we have a finite supremum of affine functions is essential. The function in Example 3.4 is an infinite supremum of affine functions, for which (21) does not hold.

The next result characterizes the case in which the epigraph of a function is sharp w.r.t. the vector $(0, -1)$. The latter property turns out to be important in Section 4.

**Proposition 3.14.** Let $f : \mathcal{H} \to \mathbb{R}_\infty$ be a proper, convex and lsc function. The following statements are equivalent.

(i) $\text{epi} f$ is $\alpha$-sharp w.r.t. the vector $z^* := (0, -1)$.

(ii) $\alpha < 1$ and $f$ verifies (21) with parameter $\beta := \frac{\alpha}{\sqrt{1 - \alpha^2}}$.

**Proof.** Let $E := \text{epi} f$, and $D := \text{dom} f$. From Fact 2.1, we have

$$N_E((x, f(x)) = \mathbb{R}_+(\partial f(x) \times \{-1\}) \bigcup (N_D(x) \times \{0\}).$$

(22)

For calculating the sharpness of $E$ at $z^*$, it is enough to consider only points of the form $(x, f(x))$, since otherwise for $(x, y) \in E$, with $y > f(x)$, then $N_E((x, y)) = N_{\text{dom} f}(x) \times \{0\}$, and hence $d(z^*, N_E((x, y))) \geq 1$. Thus, for computing the sharpness of $E$ we need to take the infimum over the set:

$$K := \{(x, f(x)) : (0, -1) \not\in N_E((x, f(x))) \} = \{(x, f(x)) : \partial f(x) \not= \emptyset \text{ and } 0 \not\in \partial f(x), \text{ or } \partial f(x) = \emptyset\},$$

where we are using (22) in the characterization of $K$. If $(x, f(x)) \in K$ and $\partial f(x) = \emptyset$, then (22) gives

$$N_E((x, f(x)) = N_D(x) \times \{0\},$$

so in this case we have

$$d(z^*, N_E((x, f(x)))) = d((0, -1), N_D(x) \times \{0\}) = \inf_{w \in N_D(x)} \sqrt{\|w\|^2 + 1} = 1.$$
Fix now any \((x, f(x)) \in K\) such that \(\partial f(x) \neq \emptyset\) and denote by \(\delta(x) := d(0, \partial f(x))\). Since \(\partial f(x)\) is a nonempty, closed and convex set (see [2, Proposition 20.31]), and \(0 \notin \partial f(x)\), we have that \(\delta(x) > 0\). We first prove that for every \((x, f(x)) \in K\) such that \(\partial f(x) \neq \emptyset\), we have

\[
d(z^*, N_E((x, f(x)))) = \frac{\delta(x)}{\sqrt{\delta(x)^2 + 1}}.
\]  

(24)

Indeed, fix \((x, f(x)) \in K\). Using (22) we have

\[
d(z^*, N_E((x, f(x)))) = \inf_{(u, \eta) \in N_E((x, f(x)))} d(z^*, (u, \eta)) = \min \left\{ \inf_{w \in \partial f(x)} d(z^*, (w, 0)), \inf_{u \in \partial f(x), \tau > 0} d(z^*, (\tau u, -\tau)) \right\}
\]

\[
= \min \left\{ \inf_{w \in \partial f(x)} \sqrt{\|w\|^2 + 1}, \inf_{u \in \partial f(x), \tau > 0} \sqrt{\|u\|^2 + 1 - 2\tau + 1} \right\}
\]

\[
= \min \left\{ 1, \inf_{u \in \partial f(x)} \left[ \min_{\tau > 0} \sqrt{\|u\|^2 + 1 - 2\tau + 1} \right] \right\} = \min \left\{ 1, \inf_{u \in \partial f(x)} \frac{\|u\|}{\sqrt{\|u\|^2 + 1}} \right\}
\]

\[
= \frac{\delta(x)}{\sqrt{\delta(x)^2 + 1}},
\]

where we are using (22) in the second equality and the definition of \(z^*\) in the third one. In the fourth equality we are using the fact that the infimum over \(N_D(x)\) is attained at \(w = 0\). In the fifth equality we are using the fact that the infimum in the expression between square brackets is attained at \(t^* := 1/(\|u\|^2 + 1)\), and in the sixth one the fact that the latter infimum value is smaller than 1 as well as the fact that the function \(g(s) := s/\sqrt{s^2 + 1}\) is increasing over \(\mathbb{R}_+\) and attains its minimum at \(s^* := \delta(x)\). Hence, (24) holds for every \((x, f(x)) \in K\).

Now, we assume that (i) holds. Then by (23) and (24) this means that, for every \((x, f(x)) \in K\), we have

\[
\alpha \leq \inf_{(x, f(x)) \in K} d(z^*, N_E((x, f(x))))
\]

\[
= \min \left\{ \inf_{(x, f(x)) \in K, \partial f(x) = \emptyset} d(z^*, N_E((x, f(x)))), \inf_{(x, f(x)) \in K, \partial f(x) \neq \emptyset} d(z^*, N_E((x, f(x)))) \right\}
\]

\[
= \min \left\{ 1, \inf_{(x, f(x)) \in K, \partial f(x) \neq \emptyset} \frac{\delta(x)}{\sqrt{\delta(x)^2 + 1}} \right\} \leq \frac{\delta(y)}{\sqrt{\delta(y)^2 + 1}} < 1,
\]

for every \((y, f(y)) \in K\) such that \(\partial f(y) \neq \emptyset\). Note that we are using (23) and (24) in the second equality. So \(\alpha < 1\), and the above expression re-writes as

\[
\delta(y) \geq \frac{\alpha}{\sqrt{1 - \alpha^2}}, \forall (y, f(y)) \in K \text{ s.t. } \partial f(y) \neq \emptyset,
\]  

(25)

which means that (ii) holds with parameter \(\alpha \sqrt{1 - \alpha^2}\). Note that we are using the convention that the infimum of the empty set is \(+\infty\), so (ii) automatically holds if \(\partial f(x) = \emptyset\).

Conversely, if condition (ii) holds with parameter \(\beta = \frac{\alpha}{\sqrt{1 - \alpha^2}}\), then (25) holds. The latter rewrites as

\[
\frac{\delta(y)}{\sqrt{\delta(y)^2 + 1}} \geq \alpha, \forall (y, f(y)) \in K \text{ s.t. } \partial f(y) \neq \emptyset,
\]

which, together with (23) and (24), give (i). Therefore, it is enough to establish (24).

We illustrate the results in Proposition 3.14 by two examples in Figure 3. Condition (21) is closely related with the well-known Kurdyka-Łojasiewicz inequality. To make this connection precise, we recall next the necessary definitions.

Definition 3.3 (Kurdyka-Łojasiewicz inequality [4, Section 2.3]). Let \(f: \mathcal{H} \to \mathbb{R}_\infty\), and assume that \(\mathcal{S} := \arg\min f \neq \emptyset\). Fix \(\bar{x} \in \mathcal{S}\). The function \(f\) satisfies the global Kurdyka-Łojasiewicz (KL) property at \(\bar{x}\) if there exists a concave continuously differentiable function \(\varphi: \mathbb{R}_+ \to \mathbb{R}_+\) with \(\varphi(0) = 0\) and \(\varphi' > 0\) such that

\[
\varphi'(f(x) - f(\bar{x})) d(0, \partial f(x)) \geq 1 \quad \forall x \notin \mathcal{S}.
\]  

(26)

In this case, we say that \(\varphi\) is a desingularizing function for \(f\) at \(\bar{x}\). If \(f\) satisfies the global KL property and admits the same desingularizing function \(\varphi\) at every point \(\bar{x} \in \mathcal{S}\), then we say that \(f\) satisfies the global KL property with global desingularizing function \(\varphi\).
The next result establishes the connection between the global KL property and sharpness.

**Corollary 3.15.** Let \( f : \mathcal{H} \to \mathbb{R}_\infty \) be a proper lsc convex function. The following statements are equivalent.

(i) \( \text{epi } f \) is \( \alpha \)-sharp w.r.t. the vector \( z^* := (0, -1) \).

(ii) \( \alpha < 1 \) and \( f \) satisfies the global KL property with global desingularizing function \( \varphi(t) = \frac{t\sqrt{1 - \alpha^2}}{\alpha} \).

**Proof.** The claim of the corollary follows from Proposition 3.14 because the global KL property (ii) is equivalent to condition (b) in Proposition 3.14.

**Corollary 3.16.** Let \( f : \mathcal{H} \to \mathbb{R}_\infty \) be a polyhedral function (see Definition 2.5). Then there exists \( \alpha < 1 \) such that \( f \) satisfies the global KL property with global desingularizing function \( \varphi(t) = \frac{t\sqrt{1 - \alpha^2}}{\alpha} \). In particular, property (21) holds for \( \beta := \frac{\alpha}{\sqrt{1 - \alpha^2}} \).

**Proof.** By Definition 2.5, the epigraph of a polyhedral function is a polyhedral set. By Proposition 3.12, it is sharp with respect to any unit vector, in particular with respect to \( z^*:=(0, -1) \). The two claims now follow from the part (i) implies (ii) in Corollary 3.15, and from the part (i) implies (ii) in Proposition 3.14.

![Figure 3: Illustration of Proposition 3.14](image)

(a) The function \( f(x) = x^2 \) does not have the global KL property, and the set \( \text{epi } f \) is not sharp w.r.t. the vector \( (0, -1) \).

(b) The function \( g(x) = |x| \) has the global KL property, and the set \( \text{epi } g \) is sharp w.r.t. the vector \( (0, -1) \).

Figure 3: Illustration of Proposition 3.14: \( \text{epi } f \) is sharp w.r.t. the vector \( (0, -1) \) if and only if \( f \) has the global KL property. The rationale for these figures is identical to the one given earlier for Figure 2.

In the next result, we establish yet another connection between the sharpness property of a set and the KL property. The result shows that a set is sharp w.r.t. \( x^* \) if and only if the function \( \mathbb{1}_A(\cdot) - \langle x^*, \cdot \rangle \) has the KL property with exponent of 0.

**Proposition 3.17.** Fix \( x^* \in S \) and consider a closed and convex set \( A \). The following statements are equivalent.

(i) The modulus of sharpness of \( A \) is w.r.t. \( x^* \) is \( \alpha > 0 \) or, equivalently, \( \alpha := s\mathbb{1}_A[A, x^*] > 0 \).

(ii) The function \( f(x) := \mathbb{1}_A(x) - \langle x^*, x \rangle \) satisfies the global KL property with global desingularizing function \( \varphi(t) = \frac{t}{\alpha} \).

(iii) The function \( f \) as in (ii) satisfies (21) with \( \beta := \alpha \).

If, under any of the above conditions, we have that \( \alpha < 1 \), then \( \text{epi } f \) is \( (\alpha/\sqrt{1 + \alpha^2}) \)-sharp w.r.t. \( z^* := (0, -1) \).

**Proof.** The definition of \( f \) yields

\[
\partial f(x) = N_A(x) - x^*, \quad \forall x \in \mathcal{H}.
\]

Thus, for any \( x \in A \), \( x^* \notin N_A(x) \) if and only if \( 0 \notin \partial f(x) \). In other words, \( x \notin \text{argmin } f \) if and only if \( x^* \notin N_A(x) \). The equality above implies that

\[
d(0, \partial f(x)) = d(0, N_A(x) - x^*) = d(x^*, N_A(x)),
\]

which, combined with the fact that \( \varphi'(t) = 1/\alpha \) gives, for all \( x \notin \text{argmin } f \), \( \bar{x} \in \text{argmin } f \),

\[
(1/\alpha) d(0, \partial f(x)) = \varphi'(f(x) - f(\bar{x})) d(0, \partial f(x)) = (1/\alpha)d(x^*, N_A(x)),
\]

12
The three equivalences follow directly from the expression above and the definitions. As for the last statement, assume that $\alpha < 1$. Since all (i)-(iii) are equivalent, we can use Proposition 3.14(ii) for $\beta := \alpha$. Indeed, by (iii) we have that $\delta(x) \geq \alpha$. By (24) this implies that, whenever $\partial f(x) \neq \emptyset$ and $0 \notin \partial f(x)$

$$
d(z^*, N_E((x, f(x)))) = \frac{\delta(x)}{\sqrt{\delta(x)^2 + 1}} \geq \frac{\alpha}{\sqrt{1 + \alpha^2}},
$$

where we used the fact that the function $g(s) := s/\sqrt{s^2 + 1}$ is increasing over $\mathbb{R}_+$. By (23) we know that

$$
d(z^*, N_E((x, f(x)))) = 1, \text{ if } \partial f(x) = \emptyset.
$$

Altogether, and using the same set $K$ as in the proof of Proposition 3.14, we can write

$$
\inf_{(x, f(x)) \in K} d(z^*, N_E((x, f(x)))) = \min \left\{ \inf_{(x, f(x)) \in K, \partial f(x) = \emptyset} d(z^*, N_E((x, f(x))), \inf_{(x, f(x)) \in K, \partial f(x) \neq \emptyset} d(z^*, N_E((x, f(x)))) \right\}
$$

$$
\geq \min \left\{ 1, \frac{\alpha}{\sqrt{1 + \alpha^2}} \right\} = \frac{\alpha}{\sqrt{1 + \alpha^2}},
$$

establishing the last claim. \(\square\)

### 3.3 Dual characterizations

As seen in Figure 2, the sharpness property of the set $A$ w.r.t. the vector $x^*$ is related to the condition that $x^*$ belongs to the interior of the set $\cup_{x \in F_A(x^*)} N_A(x)$. The next result explores this connection in general settings.

We recall that the open ball of radius $r > 0$ and center $x_0$ is written as $x_0 + r(B \setminus S)$, and that the corresponding closed ball is $x_0 + rB$.

**Proposition 3.18.** Consider a nonempty closed convex set $A$ of a Hilbert space $H$, a vector $x^* \in S$, and a positive constant $\alpha \in (0, 1)$. Consider the following statements:

(i) The set $A$ is $\alpha$-sharp w.r.t. $x^*$.

(ii) $x^* + \alpha(B \setminus S) \subset \bigcup_{x \in F_A(x^*)} N_A(x)$.

Then, (ii) $\Rightarrow$ (i). If the set $A$ is bounded, then (i) $\Rightarrow$ (ii).

**Proof.** We first prove (ii) $\Rightarrow$ (i). Suppose (ii) holds. We now prove that, for any vector $y \in A$ satisfying

$$
d(x^*, N_A(y)) < \alpha,
$$

we must have $x^* \in N_A(y)$. This statement then implies that inequality (11) must hold for any $x \in A$ with $x^* \notin N_A(x)$.

Suppose to the contrary that there is $y \in A$ such that (28) holds and $x^* \notin N_A(y)$. From the inequality (28), we can choose $y^* \in N_A(y)$ such that $0 < \|x^* - y^*\| < \alpha$. Hence, $y^* \in x^* + \alpha(B \setminus S)$. Because $x^* + \alpha(B \setminus S)$ is an open set, then there exists $r > 0$ such that $y^* + rB \subset x^* + \alpha(B \setminus S)$. Let $t_0 := 1 + \frac{r}{\|y^* - x^*\|}$ and $z^* := x^* + t_0(y^* - x^*)$ (see Figure 4). Using these definitions, we can write

$$
\|z^* - y^*\| = \left\| x^* + \left(1 + \frac{r}{\|y^* - x^*\|}\right)(y^* - x^*) \right\| - y^*
$$

$$
= \left\| \frac{r}{\|y^* - x^*\|}(y^* - x^*) \right\| = r.
$$

This gives

$$
z^* = x^* + t_0(y^* - x^*) \in y^* + rB \subset x^* + \alpha(B \setminus S),
$$

where we used the definition of $r$ in the last inclusion. Now we can use assumption (ii). Namely,

$$
z^* \in x^* + \alpha(B \setminus S) \subset \bigcup_{x \in F_A(x^*)} N_A(x).
$$
Hence, there exists \( z \in F_A(x^*) \) such that \( z^* \in N_A(z) \). Since \( z \in F_A(x^*) \), Fact 3.2 yields \( x^* \in N_A(z) \). Altogether, both of the vectors \( x^* \) and \( z^* \) belong to the normal cone \( N_A(z) \). Using now the fact that \( y \in A \) and \( z^* \in N_A(z) \) we obtain
\[
\langle z^*, y - z \rangle \leq 0.
\]
(30)
Recall that \( x^* \notin N_A(y) \), and \( x^* \in N_A(z) \). Using Fact 3.2 the latter give \( z \in \text{Argmax}_A \langle x^*, \cdot \rangle \) and \( y \notin \text{Argmax}_A \langle x^*, \cdot \rangle \). Therefore, we can write
\[
\langle x^*, y \rangle < \langle x^*, z \rangle = \sup_{\overline{A}} \langle x^*, \cdot \rangle.
\]
(31)
From \( y^* \in N_A(y) \) and \( z \in A \), we also have
\[
\langle y^*, z \rangle \leq \langle y^*, y \rangle.
\]
(32)
We now use inequalities (31) and (32) above, together with the facts that \( z^* = x^* + t_0(y^* - x^*) \) and \( t_0 > 1 \), to derive the following estimation
\[
\langle z^*, y \rangle = \langle t_0y^* + (1 - t_0)x^*, y \rangle = t_0 \langle y^*, y \rangle + (1 - t_0) \langle x^*, y \rangle > t_0 \langle y^*, z \rangle + (1 - t_0) \langle x^*, z \rangle
\]
\[
= \langle t_0y^* + (1 - t_0)x^*, z \rangle = \langle z^*, z \rangle,
\]
which contradicts (30). Therefore, we must have \( x^* \in N_A(y) \). We have shown that, whenever (28) holds, we must have \( x^* \in N_A(y) \). Equivalently, if \( x^* \notin N_A(y) \) then we must have \( d(x^*, N_A(y)) \geq \alpha \). The latter statement implies that \( A \) is \( \alpha \)-sharp w.r.t vector \( x^* \). This completes the proof of (ii) \( \Rightarrow \) (i).

We now prove (i) \( \Rightarrow \) (ii) when the set \( A \) is bounded and \( \alpha \)-sharp w.r.t. \( x^* \). Since the set \( A \) is bounded, we can use the characterization of sharpness established in Proposition 3.10. The latter implies that for all \( y^* \in \mathcal{H} \setminus \{0\} \) such that \( F_A(x^*) \cap F_A(y^*) = \emptyset \) we must have
\[
\|x^* - y^*\| \geq \alpha.
\]
(33)
We now show the inclusion in (ii). Take any \( y^* \in x^* + \alpha(\mathcal{B} \setminus \mathcal{S}) \), and assume that \( y^* \notin \bigcup_{x \in F_A(x^*)} N_A(x) \). This implies that for every \( x \in F_A(x^*) \), we must have \( y^* \notin N_A(x) \). By Fact 3.2, we deduce that for every \( x \in F_A(x^*) \), we must have \( x \notin F_A(y^*) \). In other words, \( F_A(y^*) \cap F_A(x^*) = \emptyset \). By (i) and Proposition 3.10 we deduce that (33) holds. This contradicts the fact that \( y^* \in x^* + \alpha(\mathcal{B} \setminus \mathcal{S}) \). Therefore, \( y^* \in \bigcup_{x \in F_A(x^*)} N_A(x) \), which completes the proof of (i) \( \Rightarrow \) (ii).

Remark 3.19. In [17], a strict complementary condition is used to ensure that SPP solves the linear programming problem min \( \langle x^*, x \rangle \) where \( A \subset \mathcal{H} \) is a polyhedron, \( x^* \in \mathcal{H} \setminus \{0\} \), and \( \dim \mathcal{H} = n < +\infty \). This is established in [17, Theorem 1], whose proof relies on the following two key conditions:

(i) The minimization problem min \( \langle x^*, x \rangle \) has an unique solution \( \bar{x} \in A \); and

(ii) The normal cone \( N_A(\bar{x}) \) has nonempty interior and \( -x^* \in \text{int} N_A(\bar{x}) \) (see Figure 5(a)).

Condition (ii) above implies that there must be at least \( n \) linear independent constraints that are active at \( \bar{x} \), where \( n \) is the dimension of the primal problem. In fact, let \( I(\bar{x}) \) be the set of active constraints at \( \bar{x} \), then the normal cone of \( A \) at \( \bar{x} \) is
\[
N_A(\bar{x}) = \text{cone}[A^* e_{i1}^*]_{i1 \in I(\bar{x})}.
\]
Since \( \text{int} N_A(\bar{x}) \neq \emptyset \), it follows that \( \dim N_A(\bar{x}) = n \). Therefore, \( |I(\bar{x})| \geq n \), and there are \( n \) constraints \( i_1, \ldots, i_n \in I(\bar{x}) \) such that \( A^* e_{i1}^*, \ldots, A^* e_{in}^* \) are linearly independent. Thus, any linear programming problem in \( \mathbb{R}^n \) that satisfies the strict complementary condition must have at least \( n \) linear independent constraints. By Definition 2.5, a linear programming problem can only have finite number of constraints, hence the aforementioned requirement may only work for finite dimensions, and will never hold in infinite dimensional case.
Remark 3.20. Given a convex and closed set $A$, the Fenchel conjugate of $f(\cdot) := 1_A(\cdot)$, denoted by $\sigma_A(\cdot)$, is called the support function of $A$. Recall that $\partial f = N_A$ and $\text{dom} f = A$. Applying (34) to this $f$ and using the definitions, we obtain
\[
\sigma_A(v) = \sup_{y \in A} \langle v, y \rangle.
\]
(36)

Using (34) for this $f$, we obtain
\[
v \in N_A(x) \iff x \in A \text{ and } \langle v, x \rangle = \sigma_A(v) \iff x \in \partial \sigma_A(v) \iff \partial \sigma_A(v) = \text{Argmax}_{y \in A} \langle v, y \rangle = F_A(v),
\]
(37)

where we also used (36) and Fact 3.2 in the rightmost equivalence. The equivalence above also shows that $D(\partial \sigma_A) = R(N_A)$ (see Definition 2.2(a)(b)). These facts will be used in the next result.

\[\]  

Proposition 3.21. Let $A \subset \mathcal{H}$ be a closed and convex set. Fix $x^* \in S \cap R(N_A)$ and $\alpha \in (0,1]$. Consider the following statements:

(i) The set $A$ is $\alpha$-sharp w.r.t. $x^*$.

(ii) $\emptyset \neq \partial \sigma_A(v) \subset \partial \sigma_A(x^*)$ for all $v \in x^* + \alpha(\mathcal{B}\setminus S)$.

Then (ii) implies (i). If $A$ is bounded, then (i) implies (ii). In the latter situation, $\partial \sigma_A(v) \subset F_A(x^*)$ for all $v \in x^* + \alpha(\mathcal{B}\setminus S)$.

Proof. For simplicity, denote $B(\alpha, x^*) := x^* + \alpha(\mathcal{B}\setminus S)$. Note also that the last statement will follow directly from (ii) and the equivalence (37) for $v := x^*$. Indeed, the rightmost expression in (37) with $v := x^*$ gives $F_A(x^*) = \partial \sigma_A(x^*)$.

Now, assume that (ii) holds. We will use Proposition 3.18 to show (i). More precisely, we will show that $B(\alpha, x^*) \subset \bigcup_{x \in F_A(x^*)} N_A(x)$ holds. Then (i) will follow from the part (ii) implies (i) in Proposition 3.18. Indeed, take $v \in B(\alpha, x^*)$ we need to show that there exists $x \in F_A(x^*)$ such that $v \in N_A(x)$. By (ii) we have $\partial \sigma_A(v) \neq \emptyset$ and $\partial \sigma_A(v) \subset \partial \sigma_A(x^*)$. Hence, we can write
\[
x \in \partial \sigma_A(v) \implies x \in \partial \sigma_A(x^*) = F_A(x^*).
\]
On the other hand, by (37) we have

\[ x \in \partial \sigma_A(v) \iff x \in \text{Argmax}_{y \in A} (v, y) \iff v \in N_A(x). \]

Combining the rightmost parts of the two expressions above gives that

\[ v \in B(\alpha, \alpha^*) \quad \text{and} \quad x \in \partial \sigma_A(v) \implies x \in F_A(\alpha^*) \quad \text{and} \quad v \in N_A(x). \]

Therefore, for every \( v \in B(\alpha, \alpha^*) \), there exists \( x \in F_A(\alpha^*) \) s.t. \( v \in N_A(x) \). In other words,

\[ B(\alpha, \alpha^*) \subseteq \bigcup_{x \in F_A(\alpha^*)} N_A(x). \]

Proposition 3.18 (part (ii) implies (i)), now implies that \( A \) is \( \alpha \)-sharp w.r.t. \( \alpha^* \).

Next, assume now that \( A \) is bounded. We will prove that (i) implies (ii). Namely, we will use the boundedness and sharpness of \( A \) and Proposition 3.18 to show that (ii) holds. Fix \( v \in B(\alpha, \alpha^*) \). The nonemptiness of \( \partial \sigma_A(v) \) follows from the fact that the set \( A \) is bounded and \( F_A(v) = \partial \sigma_A(v) \) (from the rightmost expression in (37) for \( v \)). To show that \( \partial \sigma_A(v) \subset \partial \sigma_A(x^*) \) we need to show that \( F_A(v) \subset F_A(x^*) \). Suppose to the contrary that \( F_A(v) \not\subset F_A(x^*) \). Hence, there is \( y \in F_A(v) \setminus F_A(x^*) \). Using Fact 3.2, the latter means that \( v \in N_A(y) \) and \( x^* \not\in N_A(y) \). Since (i) holds, we can use Proposition 3.10 to write

\[ \|x^* - v\| \geq d(x^*, N_A(y)) \geq \alpha, \]

which contradicts the fact that \( v \in B(\alpha, \alpha^*) \). Therefore, we must have \( F_A(v) \subset F_A(x^*) \), and hence (ii) holds.

**Remark 3.22.** We say that a proper lower semicontinuous convex function \( f : \mathcal{H} \ni R_\infty \) is quasi-polyhedral at \( x \in \text{dom } f \) (see [8]) if there exists \( r > 0 \) such that \( \partial f(x) \subset f(\bar{x}) \) for all \( x \in \bar{x} + r \mathcal{B} \). Furthermore, from [8, Proposition 3.4], if \( f \) is continuous at \( \bar{x} \), then \( f \) is quasi-polyhedral at \( \bar{x} \) if and only if \( f \) is conical at \( \bar{x} \); meaning that there is \( r > 0 \) and a sublinear function \( p : \mathcal{H} \to R_\infty \) such that \( f(x) = f(\bar{x}) + p(x - \bar{x}) \) for every \( x \in \bar{x} + r \mathcal{B} \). Hence, Proposition 3.21 shows that \( A \) is sharp w.r.t. \( \alpha^* \) if and only if the function \( \sigma_A \) is quasi-polyhedral at \( \alpha^* \); and if \( \sigma_A \) is continuous at \( \alpha^* \), then \( \sigma_A \) is also conical.

**3.4 Metric characterizations**

The sharpness property has a strong connection with regularity-type properties of sets. In particular, we will show in this section that sharpness with respect to a vector \( \alpha^* \) is equivalent to subtransversality (also known as metric subregularity, see [14, Definition 3.1] and [13]) between the set \( A \) and its supporting hyperplane with respect to \( \alpha^* \). Recall that, given two convex sets \( A, B \) such that \( A \cap B \neq \emptyset \), the pair \( \{A, B\} \) is subtransversal if there is \( \alpha \in (0,1) \) such that

\[ \alpha d(x, A \cap B) \leq \max\{d(x, A), d(x, B)\}, \quad \forall x \in \mathcal{H}. \]

**Remark 3.23.** Note that subtransversality for the sets \( \{A, B\} \) is equivalent to the property

\[ \alpha d(x, A \cap B) \leq d(x, A), \quad \forall x \in B. \quad (38) \]

Indeed, this follows from the fact that \( d(x, A) \leq \max\{d(x, A), d(x, B)\} \) when \( x \not\in B \). The connection between subtransversality and sharpness arises when we specialize (38) for the pair \( \{A, F\} \) with

\[ F := \{x \in \mathcal{H} : \langle x^*, x \rangle = \sup_{y \in A} \langle x^*, y \rangle\}. \quad (39) \]

In this situation, using the definition of \( F_A(x^*) \), (38) becomes

\[ \alpha d(x, F_A(x^*)) \leq d(x, A), \quad \forall x \in F. \]

The next result shows that subtransversality of \( \{A, F\} \) is equivalent to the sharpness of \( A \).

**Theorem 3.24.** Suppose \( A \) is a closed convex set of a Hilbert space \( \mathcal{H} \), \( x^* \in S, F_A(x^*) \neq \emptyset \), and \( \alpha \in (0,1) \). Let \( F \) be as in (39). Assume that

\[ \alpha d(x, F_A(x^*)) \leq d(x, A), \quad \forall x \in F, \quad (40) \]

and define \( \gamma := \alpha \sqrt{1 - \frac{1}{\alpha^*}} \). Then, the set \( A \) is \( \gamma \)-sharp w.r.t. vector \( x^* \). Conversely, define \( \beta := 2\alpha/(1-\alpha) \). If \( A \) is \( \beta \)-sharp w.r.t. \( x^* \), then inequality (40) holds.
\(|(x^*, y^*)| \leq 0.\) Then, by Fact 2.6, \(d(x^*, \text{cone}[y^*]) = 1 \geq \alpha.\) Since \(\alpha\) is always larger than the rightmost expression in (41), in this case (41) holds. Now we show that (41) also holds for the next case. 

**Case 2.** Suppose that \((x^*, y^*) > 0.\) We first show that there is an \(x \in F\) such that \(x - y \in \text{cone}[y^*]\) (see Figure 6), and so \(y\) is a projection of \(x\) onto \(A\). This is equivalent to the existence of \(t \geq 0\) and \(x \in F\) such that \(x - y = ty^*,\) or

\[x = y + ty^* \in F.\]

By the definition of \(F\), the equality above is equivalent to stating that the equation \((x^*, y + ty^*) = \sup_{y \in A} (x^*, \cdot)\) has a solution \(t \geq 0.\) Because \((x^*, y^*) > 0,\) and \(y \in A,\) the constant

\[t := \sup_{w \in A} \frac{(x^*, w) - (x^*, y)}{(x^*, y^*)}\]

is nonnegative and satisfies \((x^*, y + ty^*) = \sup_{w \in A} (x^*, w).\) Therefore, \(x := y + ty^* \in F,\) and \(y = P_A(x).\) Hence, \(x - y \in \text{cone}[y^*] \subset N_A(y).\) Taking to account that \(\|y^*\| = 1,\) we have \(y^* = \frac{x - y}{\|x - y\|}.

Figure 6: There is \(x \in F\) such that \(y \in P_A(x)\) and \(y - x \in \text{cone}[y^*].\)

Let \(z := P_{F_A(x^*)}(x)\) be a projection of \(x\) onto \(F_A(x^*).\) Since \(z \in A\) and \(x - y \in N_A(y),\) the following inequality holds

\[\langle x - y, z - y \rangle \leq 0.\]  

(42)

Additionally, inequality (40) implies

\[\alpha \|x - z\| = \alpha d(x, F_A(x^*)) = \alpha d(x, A \cap F) \leq d(x, A) = \|x - y\|.

Combine the expression above with (42) to obtain

\[
\|x - z\|^2 = \|x - y\|^2 + \|y - z\|^2 + 2 \langle x - y, y - z \rangle \\
\geq \|x - y\|^2 + \|y - z\|^2 \\
\geq \alpha^2 \|x - z\|^2 + \|y - z\|^2.
\]

Therefore, \((1 - \alpha^2) \|x - z\|^2 \geq \|y - z\|^2.\) On the other hand,

\[
\langle x - y, x - z \rangle = \frac{1}{2} \left( \|y - x\|^2 + \|x - z\|^2 - \|y - z\|^2 \right) \\
\geq \frac{1}{2} \left( \|x - z\|^2 + \|x - z\|^2 - (1 - \alpha^2) \|x - z\|^2 \right) \\
= \frac{1}{2} \left( \alpha^2 \|x - z\|^2 + \|x - y\|^2 \right).
\]

From \(y^* = \frac{x - y}{\|x - y\|},\) the inequality above, and Cauchy-Schwartz inequality, we obtain

\[
\langle y^*, x - z \rangle = \left\langle \frac{x - y}{\|x - y\|}, x - z \right\rangle \geq \frac{\alpha^2 \|x - z\|}{2 \|x - y\|} + \frac{1}{2} \frac{\|x - y\|}{\|x - y\|} \geq \alpha,
\]
where the last inequality follows from the fact that the function \( \eta(t) := \frac{\alpha^2 t}{2} + \frac{1}{2t} \) over \( \mathbb{R}_{++} \) attains a minimum at \( t^* := 1/\alpha \) and its minimum value is \( \eta(t^*) = \alpha \). On the other hand, since \( x, z \in F \), by the definition of \( F \) we must have \( \langle x^*, \frac{x - z}{\|x - z\|} \rangle = 0 \). Combining the latter equality with the last inequality yields

\[
\langle y^* - x^*, \frac{x - z}{\|x - z\|} \rangle \geq \alpha,
\]

which, by the definition of dual norm, implies \( \|y^* - x^*\| \geq \alpha \). The latter inequality and Fact 2.6 give

\[
d(x^*, \text{cone}[y^*]) = 1 - \langle y^*, x^* \rangle = 1 - \left( \frac{\|x^*\|^2 + \|y^*\|^2 - \|x^* - y^*\|^2}{2} \right)^2 
\geq 1 - \left( \frac{2 - \alpha^2}{2} \right)^2 = \frac{1}{4} \alpha^2 (4 - \alpha^2),
\]

where we also used the fact that \( \|x^*\| = \|y^*\| = 1 \). Altogether, in both cases 1 and 2, we always have

\[
d(x^*, \text{cone}[y^*]) \geq \alpha \sqrt{1 - \frac{\alpha^2}{4}}, \quad \text{and because } y^* \in S \cap N_A(y) \text{ is chosen arbitrarily, we have }
\]

\[
d(x^*, N_A(y)) = \inf_{y^* \in S \cap N_A(y)} d(x^*, \text{cone}[y^*]) \geq \alpha \sqrt{1 - \frac{\alpha^2}{4}} = \gamma.
\]

Consequently, we have proved that the set \( A \) is \( \gamma \)-sharp w.r.t. vector \( x^* \), where \( \gamma := \alpha \sqrt{1 - \frac{\alpha^2}{4}} > 0 \). To prove the converse implication, we assume by contradiction that the set \( A \) is \( \beta \)-sharp w.r.t. \( x^* \), where \( \beta := 2\alpha/(1 - \alpha) \) and there is \( \bar{x} \in F \) such that

\[
\alpha d(\bar{x}, F_A(x^*)) > d(\bar{x}, A).
\]

The strict inequality implies that there exists \( \bar{\alpha} \in (0, \alpha) \) such that

\[
\alpha d(\bar{x}, F_A(x^*)) > \bar{\alpha} d(\bar{x}, F_A(x^*)) > d(\bar{x}, A). \tag{43}
\]

Consider the closed half-space \( F' := \{ x \in H : \langle x^*, x \rangle \geq \sup_{y \in A} \langle x^*, y \rangle \} \). Observe that, \( F = bd F' \) and

\[
F_A(x^*) = A \cap F = A \cap F',
\]

and for any \( x \in F \), we have \( N_{F'}(x) = \text{cone}[-x^*] \). Set \( \delta := d(\bar{x}, F_A(x^*)) \). Since \( \bar{x} \in F \), (43) implies that \( \delta > 0 \) and \( \bar{x} \notin A \). Let \( \bar{y} := F_A(\bar{x}) \), so \( d(\bar{x}, \bar{y}) = \|\bar{x} - \bar{y}\| > 0 \). With this notation, inequality (43) becomes

\[
\alpha \delta > \bar{\alpha} \delta > \|\bar{x} - \bar{y}\|. \tag{44}
\]

We claim that (44) implies that \( \bar{y} \notin F_A(x^*) \). Indeed, if \( \bar{y} \in F_A(x^*) \) then the definition of \( \delta \) yields

\[
\delta = d(\bar{x}, F_A(x^*)) \leq \|\bar{x} - \bar{y}\| < \alpha \delta < \delta,
\]

a contradiction. Therefore, our claim holds and \( \bar{y} \notin F_A(x^*) \). Take now \( \tilde{y} := P_{F_A(x^*)}(\bar{y}) \). By triangle inequality,

\[
d(\tilde{y}, F_A(x^*)) = \|\tilde{y} - \bar{y}\| \geq \|\bar{y} - \bar{x}\| - \|\bar{x} - \tilde{y}\| \geq d(\bar{x}, F_A(x^*)) - \|\bar{x} - \tilde{y}\| > (1 - \alpha) \delta,
\]

where we used the definition of \( \tilde{y} \) in the first equality, the fact that \( \bar{y} \in F_A(x^*) \) in the second inequality, and (44) in the last inequality. Set \( \tilde{\delta} := (1 - \alpha) \delta \). Using the above expression, and the definitions of \( F' \) and \( \tilde{\delta} \) we derive

\[
d(\tilde{y}, F_A(x^*)) = d(\tilde{y}, A \cap F') > \tilde{\delta}. \tag{45}
\]

With this notation, \( \bar{\alpha} \delta = \frac{\tilde{\alpha} \delta}{(1 - \alpha)} \) and the second inequality in (44) rewrites as

\[
\frac{\tilde{\alpha} \delta}{1 - \alpha} > \|\bar{x} - \tilde{y}\|. \tag{46}
\]

Define the function \( f : H^2 \to \mathbb{R}_\infty \) as

\[
f(x, y) := \|x - y\| + \|F'(x) + \mathbb{1}_A(y), \quad (x, y) \in H^2. \tag{47}
\]
We now use (46) and the Ekeland Variational Principle. Lemma 2.5, for $X := \mathcal{H}^2$ equipped with the max norm $\| (x, y) \| := \max \{ \| x \|, \| y \| \}$, and $\psi := f$. As mentioned in (1), the corresponding dual norm is $\| (x, y) \|_* = \| x \| + \| y \|$. Recall that $(\bar{x}, \bar{y}) \in F' \times A$, so $f(\bar{x}, \bar{y}) = \| \bar{x} - \bar{y} \|$. This fact, combined with (46) and the definition of $f$ gives

$$\inf_{(x, y) \in \mathcal{H}} f(x, y) \geq 0 > f(\bar{x}, \bar{y}) - \frac{\alpha \delta}{1 - \alpha}.$$  

Hence we are in conditions of the Ekeland’s variational principle with $\varepsilon := \frac{\alpha \delta}{1 - \alpha}$ and $\bar{\psi} := (\bar{x}, \bar{y})$. We apply the principle for the choice $\lambda := \delta$, for which there exists $(\tilde{x}, \tilde{y})$ such that

(i) $\| \tilde{x} - \bar{x} \| < \delta$, $\| \tilde{y} - \bar{y} \| < \delta$;

(ii) $f(\tilde{x}, \tilde{y}) = \| \tilde{x} - \tilde{y} \| \leq f(\bar{x}, \bar{y}) = \| \bar{x} - \bar{y} \|.$

Moreover, our choices imply that $\varepsilon/\lambda = \alpha/(1 - \alpha)$. Altogether, condition (iii) in Lemma 2.5 with the max norm implies that for every $H^2 \ni (x, y) \neq (\tilde{x}, \tilde{y})$ we have

$$f(\tilde{x}, \tilde{y}) < f(x, y) + \frac{\alpha \delta}{1 - \alpha} \max \{ \| x - \tilde{x} \|, \| y - \tilde{y} \| \} =: h(x, y).$$  

(48)

The above inequality implies that $(\tilde{x}, \tilde{y}) \in F' \times A$. Consequently, the following statements hold.

(I) By (45) we have that $d(\tilde{y}, A \cap F') > \delta$. Using also (i) and the fact that $d(\bar{x}, A \cap F') = \delta > \delta$, we have $\tilde{x}, \tilde{y} \notin A \cap F'$. Indeed, if we would have $\tilde{x} \in A \cap F'$ then this would imply

$$\tilde{\delta} < d(\tilde{x}, A \cap F') \leq \| \bar{x} - \tilde{x} \|,$$

contradicting (i). A similar expression, mutatis mutandis, can be used to show that $\tilde{y} \notin A \cap F'$.

(II) Because $\tilde{y} \in A$, $\tilde{x} \in F'$, by (I) we must have $\tilde{x} \neq \tilde{y}$. The latter fact and (ii) yield $f(\tilde{x}, \tilde{y}) > 0$.

(III) By (II) and Remark 2.2 we can write the subdifferential of $f$ as the sum of the subdifferentials at $(\tilde{x}, \tilde{y})$

$$\partial f(\tilde{x}, \tilde{y}) = \partial_1 f(\tilde{x}, \tilde{y}) \times \partial_2 f(\tilde{x}, \tilde{y}) = \left( \frac{\tilde{x} - \tilde{y}}{\| \tilde{x} - \tilde{y} \|} + N_{F'}(\tilde{x}), \frac{\tilde{y} - \tilde{x}}{\| \tilde{x} - \tilde{y} \|} + N_A(\tilde{y}) \right),$$

where $\partial_1$ and $\partial_2$ stand for the partial subdifferentials w.r.t. the first and second variables, respectively. Note that each of these partial subdifferentials exist because of the continuity of the first term in (47).

(IV) By (48), it is clear that $(\tilde{x}, \tilde{y})$ is the global minimizer of $h$.

By (IV), we have that $0 \in \partial h(\tilde{x}, \tilde{y})$. To compute the subdifferential of $h$, we note that the second term in (48) is continuous everywhere. Hence, the subdifferential of $h$ can be expressed as the sum of the subdifferential of $f$ plus the subdifferential of the second term in (48). To write down the inclusion $0 \in \partial h(\tilde{x}, \tilde{y})$, we use Remarks 2.2 and 2.3 to write:

$$0 \in \left( \frac{\tilde{x} - \tilde{y}}{\| \tilde{x} - \tilde{y} \|} + N_{F'}(\tilde{x}), \frac{\tilde{y} - \tilde{x}}{\| \tilde{x} - \tilde{y} \|} + N_A(\tilde{y}) \right) + \frac{\alpha \delta}{1 - \alpha} \mathbb{B}_{\mathcal{H}^2}.$$

where the ball in the rightmost term is the one induced by the sum norm as given in (2). The inclusion above and the definition of the dual ball, give

$$d \left( \frac{\tilde{x} - \tilde{y}}{\| \tilde{x} - \tilde{y} \|}, N_{F'}(\tilde{x}) \right) + d \left( \frac{\tilde{x} - \tilde{y}}{\| \tilde{x} - \tilde{y} \|}, N_A(\tilde{y}) \right) = d \left( - \left( \frac{\tilde{x} - \tilde{y}}{\| \tilde{x} - \tilde{y} \|} + \frac{\tilde{y} - \tilde{x}}{\| \tilde{x} - \tilde{y} \|} \right), N_{F'}(\tilde{x}) \times N_A(\tilde{y}) \right) \leq \alpha/(1 - \alpha),$$

where we are using the definition of the sum ball (see (2)) in the first equality. Note also that $\tilde{x} \in F' = \{ x \in \mathcal{H} : \langle x^*, x \rangle \geq \sup_{y \in A} \langle x^*, y \rangle \},$ so

$$N_{F'}(\tilde{x}) = \begin{cases} \text{cone}[\mathbb{x}^*] & \text{if } \tilde{x} \in F = \text{bd } F', \\ \{0\} & \text{otherwise.} \end{cases}$$

This gives, $N_{F'}(\tilde{x}) \subset \text{cone}[\mathbb{x}^*]$. Therefore,

$$d \left( \frac{\tilde{x} - \tilde{y}}{\| \tilde{x} - \tilde{y} \|}, N_A(\tilde{y}) \right) + d \left( \frac{\tilde{x} - \tilde{y}}{\| \tilde{x} - \tilde{y} \|}, \text{cone}[\mathbb{x}^*] \right) \leq d \left( \frac{\tilde{x} - \tilde{y}}{\| \tilde{x} - \tilde{y} \|}, N_A(\tilde{y}) \right) + d \left( \frac{\tilde{y} - \tilde{x}}{\| \tilde{x} - \tilde{y} \|}, N_{F'}(\tilde{x}) \right) \leq \alpha/(1 - \alpha).$$  

(49)
To simplify notation, call \( \omega_0 := \frac{\hat{x} - \hat{y}}{\|\hat{x} - \hat{y}\|} \). By triangle inequality, we have
\[
d(x^*, N_A(\hat{y})) \leq \|x^* - \omega_0\| + d(\omega_0, N_A(\hat{y})).
\] (50)
We claim that
\[
d(\omega_0, \text{cone}[x^*]) \geq \frac{1}{2} \|x^* - \omega_0\|.
\] (51)
Indeed, from Fact 2.6, we have \( d(\omega_0, \text{cone}[x^*])^2 = 1 - \max(0, \langle x^*, \omega_0 \rangle)^2 \). We consider two cases:

1. If \( \langle x^*, \omega_0 \rangle \leq 0 \), then \( d(\omega_0, \text{cone}[x^*]) = 1 \) and by triangle inequality
\[
\|\omega_0 - x^*\| \leq \|\omega_0\| + \|x^*\| = 2 = 2d(\omega_0, \text{cone}[x^*]),
\]
as claimed in (51).

2. If \( \langle x^*, \omega_0 \rangle > 0 \), then
\[
d(\omega_0, \text{cone}[x^*])^2 = 1 - \langle x^*, \omega_0 \rangle^2
\]
\[
= \frac{1}{2} (1 + \langle x^*, \omega_0 \rangle)(2 - 2 \langle x^*, \omega_0 \rangle)
\]
\[
> \frac{1}{2} (2 - 2 \langle x^*, \omega_0 \rangle)
\]
\[
= \frac{1}{2} \left( \|x^*\|^2 + \|\omega_0\|^2 - 2 \langle x^*, \omega_0 \rangle \right)
\]
\[
= \frac{1}{2} \|\omega_0 - x^*\|^2,
\]
where we used \( (1 + \langle x^*, \omega_0 \rangle) > 1 \) in the inequality, and the fact that \( \|x^*\| = \|\omega_0\| = 1 \) in the third equality. The above expression yields
\[
d(\omega_0, \text{cone}[x^*]) > \frac{1}{\sqrt{2}} \|\omega_0 - x^*\| > \frac{1}{2} \|\omega_0 - x^*\|.
\]
In both cases, we proved that (51) holds. From (50), we have
\[
d(x^*, N_A(\hat{y})) \leq \|x^* - \omega_0\| + d(\omega_0, N_A(\hat{y}))
\]
\[
\leq 2d(\omega_0, \text{cone}[x^*]) + d(\omega_0, N_A(\hat{y}))
\]
\[
= [d(\omega_0, \text{cone}[x^*]) + d(\omega_0, N_A(\hat{y}))] + d(\omega_0, \text{cone}[x^*])
\]
\[
\leq \hat{\alpha}/(1 - \alpha) + \hat{\alpha}/(1 - \alpha) = 2\hat{\alpha}/(1 - \alpha) < 2\alpha/(1 - \alpha),
\]
where we used (50) in the first inequality, and (51) in the second one. For the last inequality we used (49) for the expression between square brackets, and for the remaining term we used the fact that disregarding the first term in (49) implies that \( d(\omega_0, \text{cone}[x^*]) \leq \hat{\alpha}/(1 - \alpha) \). From (I), we have \( \hat{y} \notin A \cap F^\dagger = A \cap F = F_A(x^*) \), and so by Fact 3.2 it holds that \( x^* \notin N_A(\hat{y}) \). By the \( \beta \)-sharpness of \( A \) w.r.t. \( x^* \) and the assumption on \( \beta \) we deduce that
\[
d(x^*, N_A(\hat{y})) \geq 2\alpha/(1 - \alpha) = \beta,
\]
contradicting (52). This implies that inequality (40) holds and the proof is complete.

The following result from [16] is a characterization of subtransversality. We will use this result to formally express the connection between subtransversality and sharpness.

**Lemma 3.25 (Subtransversality [16, Theorem 3.1])**. Suppose \( X \) is a normed linear space, \( A, B \subset X \) are closed convex sets and \( A \cap B \neq \emptyset \). The pair \( \{A, B\} \) is subtransversal if and only if there exists a number \( \alpha \in (0, 1) \) such that
\[
ad(x, A \cap B) \leq d(x, A), \quad \forall x \in B \setminus A.
\] (53)

Theorem 3.24 and Lemma 3.25 yield the following result.

**Corollary 3.26**. Suppose \( A \) is a closed convex set of a Hilbert space \( \mathcal{H} \), and \( x^* \in \mathcal{S} \). Then, \( A \) is sharp w.r.t. \( x^* \) if and only if the pair \( \{A, F\} \) is subtransversal, where \( F := \{x \in \mathcal{H} : \langle x^*, x \rangle = \sup_A \langle x^*, \cdot \rangle\} \).
Given a set $B$, Ekeland Variational Principle. In the result below, we denote by our analysis holds in a general Banach space (not necessarily Hilbert). A main tool in our proof is again the

By Theorem 3.24, the set $A$, $\exists$ sufficiently large, then $P_A(b)$ solves Problem (CP). In such situation, instead of solving Problem (CP), we can solve the (hopefully) simpler problem of finding $P_A(b)$. Before establishing the main results of this section, we first find an upper bound on the distance between a point and a set using normal cones.

4 Optimization problems under sharpness condition

We consider now constrained convex problems of the following type.

$$\min_{x \in A} f(x),$$

(\mathrm{CP})

where $f: \mathcal{H} \to \mathbb{R}_\infty$ is a proper, lsc convex function, and $A$ is a closed convex set. We provide in this section sufficient conditions under which Problem (CP) can be solved by SPP. As expected, sharpness plays a crucial role in our analysis. Namely, under the sharpness assumption, if (i) $b \notin A$, and (ii) the difference $(\inf_{x \in A} f) - f(b)$ is sufficiently large, then $P_A(b)$ solves Problem (CP). In such situation, instead of solving Problem (CP), we can solve the (hopefully) simpler problem of finding $P_A(b)$. Before establishing the main results of this section, we first find an upper bound on the distance between a point and a set using normal cones.

4.1 Upper bound on the distance

Given a set $A$ and a point $b \notin A$, how can we estimate the distance $d(b, A)$? We address this question next, our analysis holds in a general Banach space (not necessarily Hilbert). A main tool in our proof is again the Ekeland Variational Principle. In the result below, we denote by $B[\rho, b]$ the closed ball of radius $\rho$ and centre $b$, and by $B(b, \rho)$ the corresponding open ball.

**Theorem 4.1.** Consider a Banach space $X$, a nonempty closed convex set $A$, points $a \in A$, $b \notin A$ with $\rho := \|a - b\|$, and $\varepsilon > 0$. Then, $d(b, A) \leq \|a - b\| - \varepsilon$ if there is $\delta > 0$ such that

$$\inf \{d(x^*, N_A(x)) : \|x^*\| = 1, \langle x^*, b - x \rangle = \|b - x\|, x \in B[\rho, b] \cap B(\delta, a) \cap A \} \geq \frac{\varepsilon}{\delta}. \quad (54)$$

**Proof.** Assume (54) holds for some $\varepsilon, \delta > 0$. For contradictory purposes, assume also that $\varepsilon$ is such that $d(b, A) > \|a - b\| - \varepsilon$. Consider the function $f: X \to \mathbb{R}_\infty$ defined by $f(y) := \|y - b\| + 1_{A}(y) = \varphi(y) + 1_{A}(y)$. Here $\varphi(y) := \|y - b\|$ as in Remark 2.2. Then, the assumption on $\delta$ implies that $f(a) = \|a - b\| < \inf_{A} f + \varepsilon$. Take $0 < \varepsilon' < \varepsilon$, such that

$$\inf_{A} f + \varepsilon > \inf_{A} f + \varepsilon' > f(a).$$

By Ekeland’s Variational Principle (Lemma 2.5) applied to $\tilde{a} := a$, $\psi := f$, $\varepsilon := \varepsilon'$ and $\lambda := \delta$, there exists a vector $\hat{x} \in A \cap B(a, \delta)$ such that

$$f(\hat{x}) \leq f(a), \quad (55)$$

$$f(\hat{x}) \leq f(y) + \frac{\varepsilon'}{\delta} \|y - \hat{x}\|, \quad \forall y \in A. \quad (56)$$

Due to (55) and the fact that $a \in A$, we have that $\hat{x} \in A$ and $\|\hat{x} - b\| \leq \|a - b\| = \rho$, or $\hat{x} \in A \cap B[\rho, b]$. Define $h(y) := \frac{\varepsilon'}{\delta} \|y - \hat{x}\|$. By (56), it follows that $\hat{x}$ is a global minimizer of the sum function $f + h$ and hence the definition of subdifferential yields the inclusion

$$0 \in \partial(f + h)(\hat{x}) = \partial f(\hat{x}) + \partial h(\hat{x})$$

Note that the subdifferential sum formula can be used to differentiate $(f + h)$ because $h$ is continuous everywhere. By Remark 2.2, $\partial h(\hat{x}) = \frac{\varepsilon'}{\delta} B^\ast$, and we obtain

$$0 \in \partial \varphi_b(\hat{x}) + \partial 1_A(\hat{x}) + \partial h(\hat{x}) = \partial \varphi_b(\hat{x}) + N_A(\hat{x}) + \frac{\varepsilon'}{\delta} B^\ast.$$
Therefore,
\[ [\partial \varphi_b(\hat{x}) + N_A(\hat{x})] \cap \left( \begin{array}{c} \varepsilon \delta^* \end{array} \right) \neq \emptyset. \]  \hfill (57)

Furthermore, since \( \hat{x} \in A \) we have that \( \varphi_b(\hat{x}) = \| \hat{x} - b \| \geq d(b, A) > 0 \). By [21, Corollary 2.4.16] and the chain rule, we have

\[ \partial \varphi_b(\hat{x}) = \{ u : (u, \hat{x} - b) = \| \hat{x} - b \|, \| u \| = 1 \}. \]

Altogether, by (57) there is \( x^* \in (-\partial \varphi_b(\hat{x})) \) such that

\[ d(x^*, N_A(\hat{x})) \leq \frac{\varepsilon'}{\delta} < \frac{\varepsilon}{\delta}, \]

where the first inequality holds because, by (57), there exists \( w \in N_A(\hat{x}) \) such that \( (w - x^*) \in \frac{\varepsilon}{\delta} \delta^* \), and the last inequality holds because \( \varepsilon' < \varepsilon \). Noting that (54) holds, and using the facts that \( x^* \in (-\partial \varphi_b(\hat{x})) \), i.e., \( \| x^* \| = 1, (x^*, b - \hat{x}) = \| b - \hat{x} \| \), and \( \hat{x} \in B[\rho, b] \cap B(\delta, a) \cap A \) we can write

\[ \frac{\varepsilon}{\delta} \leq \inf \{ d(x^*, N_A(x)) : \| x^* \| = 1, (x^*, b - x) = \| b - x \|, x \in B[\rho, b] \cap B(\delta, a) \cap A \} < \frac{\varepsilon}{\delta}, \]

which is a contradiction. This implies that we must have \( d(b, A) \leq \| a - b \| - \varepsilon. \)

When the space \( X \) is Hilbert, Theorem 4.1 can be simplified as follows.

**Corollary 4.2.** Suppose \( A \) is a closed convex set of a Hilbert space, \( a, b \in A \), \( b \notin A \) with \( \rho := \| a - b \| \), and \( \varepsilon > 0 \). If there is \( \delta > 0 \) such that the following inequality holds

\[ \inf \left\{ d \left( \begin{array}{c} b - x \hfill \end{array} \frac{\| b - x \|}{\delta} \right), N_A(x) \hfill \right\} : x \in B[\rho, b] \cap B(\delta, a) \cap A \} \geq \frac{\varepsilon}{\delta}, \]  \hfill (58)

then \( d(b, A) \leq \| a - b \| - \varepsilon. \)

**Proof.** In a Hilbert space, the element \( x^* \) we find in the proof of Theorem 4.1 can be taken as \( x^* := (\hat{x} - b)/\| \hat{x} - b \| \), so the expression in (54) becomes (58).

**Remark 4.3.** The aim of Corollary 4.2 is to establish a sufficient condition for \( a \) being “far enough” from being a projection of \( b \) onto \( A \). Note that \( a = P_A(b) \) if and only if \( b - a \in N_A(a) \). Equivalently, \( d \left( \begin{array}{c} b - a \hfill \end{array} \frac{b - a}{\| b - a \|} \right), N_A(a) \) = 0. Hence, to ensure we are far from the latter situation, we require (58) to hold, not merely at \( a \), but at every point \( x \in B[\rho, b] \cap B(\delta, a) \cap A \). We quantify this property by showing that, if (58) holds, then we must have \( \| a - b \| > d(b, A) + \varepsilon \). The latter, in turn, means that the difference \( \| a - b \| - d(b, A) \) is bounded from below by a constant \( \varepsilon \). Note also that the opposite inequality to (58) is the optimality condition for \( \varepsilon \)-projections, where the later define points in \( A \) that are within distance \( d(b, A) + \varepsilon \) from \( b \). Geometrically, (58) ensures that the cosine of the angle between \( b - x \) and a vector in \( N_A(x) \) is always bigger than a positive constant \( \frac{\varepsilon}{\delta} \) > 0, where \( x \in B[\rho, b] \cap B(\delta, a) \cap A \). Figure 7 illustrates an example on estimating the distance from a point to a set.

Since the projection of \( b \) onto \( A \) is an element \( a \in A \) such that \( d(b, A) = \| a - b \| \), an \( \varepsilon \)-projection of \( b \) onto \( A \) can be understood as an element \( a^\varepsilon \in A \) such that \( d(b, A) > \| a^\varepsilon - b \| - \varepsilon \). This motivates the following corollary of Theorem 4.1.
Corollary 4.4. Consider a Banach space $X$, a nonempty closed convex set $A$, $b \notin A$, $a \in A$, $\rho = \|a - b\|$ and $\varepsilon \geq 0$. For any given $\delta > 0$, define the set 

$$C(\delta) := \{(x, x^*) \in X \times X^* : x \in B[\rho, b] \cap B(\delta, a) \cap A \text{ and } \|x^*\| = 1, \langle x^*, b - x \rangle = \|b - x\|\}.$$ 

Assume that $a$ is an $\varepsilon$-projection of $b$, in the sense that $\|a - b\| < d(b, A) + \varepsilon$. Then, for every $\delta > 0$, there exists $(x_0, x_0^*) \in C(\delta)$ such that 

$$d(x_0^*, N_A(x_0)) < \frac{\varepsilon}{\delta}.$$ 

Proof. Assume that the conclusion of the corollary is not true. Namely, assume that there is $\delta > 0$ such that for all $(x, x^*) \in C(\delta)$ we have 

$$d(x^*, N_A(x)) \geq \frac{\varepsilon}{\delta}.$$ 

Hence, 

$$\inf_{(x, x^*) \in C(\delta)} d(x^*, N_A(x)) \geq \frac{\varepsilon}{\delta},$$

which, by definition of $C(\delta)$, the above inequality is exactly (54). By Theorem 4.1, we must have $\|a - b\| \geq d(b, A) + \varepsilon$, contradicting the fact that $a$ is an $\varepsilon$-projection of $b$. \hfill \Box

4.2 Solving problems with SPP: the case with a linear objective

We start this section by considering the following minimization problem with linear objective:

$$\min \quad \langle x^*, x \rangle$$

s.t. 

$$x \in A,$$

where we also used (59) in the first equality. Consider the closed half space $v$ 

Combining this inequality with the inclusion $\langle x^*, x \rangle \leq 0.\}$. From Condition 1, $\langle x^*, v \rangle < \inf_{x^*} \langle x^*, x \rangle$, hence the sets $A$ and $F$ are disjoint, i.e., $A \cap F = \emptyset$. Now, we are going to apply Corollary 4.2. Namely, we will show that (58) holds with lower bound $\alpha/2$. Indeed, we apply this corollary
to the set $F$, $v \in F$, and $y \notin F$ to estimate the distance $d(y, F)$ relative to $\|v - y\|$. Set $\delta := (\alpha/2) \|v - y\|$, and $\rho := \|v - y\|$, we claim that
\[
\inf \left\{ d \left( \frac{y - z}{\|y - z\|}, N_{F}(z) \right) : z \in F \cap B(\delta, v) \cap B(\rho, y) \right\} \geq \alpha/2. \tag{62}
\]
Indeed, take $z \in F \cap B(\delta, v) \cap B(\rho, y)$. We consider two cases.

Case 1. If $z \in \text{int } F$, then $N_{F}(z) = \{0\}$, hence $d \left( \frac{y - z}{\|y - z\|}, N_{F}(z) \right) = 1$.

Case 2. If $z \notin \text{int } F$, then the definition of $F$ implies that $N_{F}(z) = \text{cone}[x^{*}]$. Since $z \in B(\rho, y)$, we use (61) to write
\[
d \left( \frac{y - z}{\|y - z\|}, N_{F}(z) \right) = d \left( \frac{y - z}{\|y - z\|}, \text{cone}[x^{*}] \right) = \inf_{t \geq 0} \left\| \frac{y - z}{\|y - z\|} - tx^{*} \right\| = \frac{\|y - v\|}{\|y - z\|} \inf_{t \geq 0} \left\| \frac{y - z}{\|y - v\|} + \frac{v - z}{\|y - v\|} - tx^{*} \right\| \geq \frac{\|y - v\|}{\|y - z\|},
\]
where in the fourth equality we used the change of variable $t \to \hat{t} := t\|y - z\|/\|y - v\|$, and we used the fact that $\|y - z\| \leq \rho = \|y - v\|$ in the first inequality. As for the last inequality, we use (61) for obtaining the lower bound of the first term. For the second term, recall that $z \in B(\delta, v)$, so $\|v - z\| < \delta = (\alpha/2) \|v - y\|$. These establish the last inequality. Hence, our claim (62) holds. By Theorem 4.1, we have
\[
d(y, F) \leq \|y - v\| - (\alpha/2)\delta = \|y - v\| - (\alpha/2)^{2} \|y - v\| = (1 - (\alpha/2)^{2}) \|y - v\|. \tag{63}
\]
To arrive at a contradiction, we will use Condition 2. Suppose $w \in F$ is the projection of $y$ onto $F$, then $y - w \in N_{F}(w)$. By definition of $F$ we know that $N_{F}(w) = \text{cone}[x^{*}]$. Hence, $x^{*} = (y - w)/\|y - w\|$ and therefore, $d(y, F) = \|y - w\| = \langle x^{*}, y - w \rangle$. Taking into account that $x^{*} \in N_{F}(v) \cap N_{F}(w)$, we obtain $\langle x^{*}, v - w \rangle \leq 0$, and $\langle x^{*}, v - w \rangle \leq 0$, and hence $\langle x^{*}, w - v \rangle = 0$. From the previous equality $d(y, F) = \langle x^{*}, y - w \rangle$, we have $\langle x^{*}, y - v \rangle = (\langle x^{*}, y - w \rangle + \langle x^{*}, v - w \rangle) = \langle x^{*}, y - w \rangle = d(y, F)$. So, the following estimation holds
\[
\inf_{x \in A} \langle x^{*}, x - v \rangle \leq \langle x^{*}, y - v \rangle = d(y, F) = (1 - (\alpha/2)^{2}) \|y - v\| = (1 - (\alpha/2)^{2}) d(v, A),
\]
where we also used (63) in the second inequality, and the definition of $y$ in the last equality. The above expression contradicts condition 2. Therefore, we must have that $P_{A}(v)$ solves problem (P).

As a consequence of Theorem 4.5, if $A$ is $\alpha$-sharp w.r.t. a vector $-x^{*}$, Problem (P) can be solved by projecting onto $A$ an infeasible point $v$ s.t. conditions 1 and 2 hold (see Figure 8). Hence, it is important to be able to construct such vectors. It is clear that condition 1 in Theorem 4.5 follows from condition 2. The next lemma shows that, once we have a vector verifying Condition 1, we can always construct a translation of the vector that verifies Condition 2.

**Lemma 4.6.** With the notation of Theorem 4.5, assume that $v \in \mathcal{H}$ verifies Condition 1 and fix $\alpha \in (0, 1]$. Assume that Condition 2 with parameter $\alpha$ does not hold for $v$. Define
\[
\theta(v) := \inf_{x \in A} \langle x^{*}, x - v \rangle, \quad \text{and } \mu_{0} := \frac{(1 - (\alpha/2)^{2})d(v, A) - \theta(v)}{(\alpha/2)^{2}}. \tag{64}
\]
Then $\theta(v) > 0$ and $\mu_{0} \geq 0$. Moreover, if $\mu > \mu_{0}$, then $u := v - \mu x^{*}$ verifies conditions 1 and 2 from Theorem 4.5.
Figure 8: Illustration of Theorem 4.5: the set $A$ is sharp w.r.t vector $-x^*$. Vector $v$ satisfies conditions 1&2, hence the projection $P_A(v)$ of $v$ onto the set $A$ is the solution of the minimization problem $\inf_A \langle x^*, \cdot \rangle$.

Proof. The fact that $\theta(v) > 0$ is equivalent to the validity of Condition 1 for $v$, so it holds by assumption. The fact that $\mu_0 \geq 0$ is equivalent to the assumption that $v$ fails to verify Condition 2 for the given $\alpha$. Altogether, we have that

$$0 < \theta(v) \leq (1 - (\alpha/2)^2) d(v, A).$$

(65)

We proceed to prove that Conditions 1 and 2 hold for $u$ if $\mu > \mu_0$. Use the definition of $\theta(v)$ to write, for all $x \in A$,

$$\langle x^*, x - u \rangle = \langle x^*, x - (v - \mu x^*) \rangle = \langle x^*, x - v \rangle + \mu \geq \theta(v) + \mu > \mu > 0,$$

where we used the definition of $u$ in the first equality, and the fact that $x^* \in \mathcal{S}$ in the second one. Therefore, $\inf_{x \in A} \langle x^*, x - u \rangle \geq \mu > 0$ and Condition 1 holds for $u$. The above expression also yields

$$\langle x^*, x - u \rangle \geq \theta(v) + \mu > 0,$$

(66)

for all $x \in A$. Let us check now that Condition 2 holds for $u$. Using again the fact that $x^* \in \mathcal{S}$ gives

$$d(u, A) = d(v - \mu x^*, A) = \inf_{x \in A} \|v - \mu x^* - x\| \leq \inf_{x \in A} \|v - x\| + \mu \|x^*\| = d(v, A) + \mu.$$  

(67)

Using the definition of $\mu_0$, we rewrite the inequality $\mu > \mu_0$ as

$$d(v, A) < \frac{(\alpha/2)^2 \mu + \theta(v)}{1 - (\alpha/2)^2}.$$  

(68)

Using (68) in (67) yields

$$d(u, A) \leq d(v, A) + \mu < \frac{(\alpha/2)^2 \mu + \theta(v)}{1 - (\alpha/2)^2} + \mu = \frac{\mu + \theta(v)}{1 - (\alpha/2)^2} \leq \frac{\langle x^*, x - u \rangle}{1 - (\alpha/2)^2},$$

where we also used (66) in the last inequality. Since the above inequality holds for every $x \in A$, we deduce that Condition 2 holds for $u$.

The argument in Lemma 4.6 is the main idea behind the next proposition. Namely, if a vector $v \in \mathcal{H}$ satisfies Condition 1 and not Condition 2, then we translate $v$ by a large enough multiple of $-x^*$, so that Condition 2 holds for the translated vector.

**Proposition 4.7.** Suppose $A$ is a closed convex set of a Hilbert space $\mathcal{H}$, and $x^* \in \mathcal{S}$ such that the set $A$ is $\alpha$-sharp w.r.t. $-x^*$, for some $\alpha \in (0, 1]$. Consider $v \in \mathcal{H}$ such that $\langle x^*, v \rangle < \min_{x \in A} \langle x^*, x \rangle$. Then, the projection of $u := v - \mu x^*$ onto $A$, where $\mu \geq \frac{4 - \alpha^2}{\alpha^2} d(v, A)$, is a solution of $(P)$.

Proof. The proof follows by noting that $\frac{4 - \alpha^2}{\alpha^2} d(v, A) = \frac{1 - (\alpha/2)^2}{(\alpha/2)^2} d(v, A) > \mu_0$, with $\mu_0$ as in Lemma 4.6. Using the lemma, we see that $u$ verifies conditions 1 and 2 in Theorem 4.5. Therefore, the claim follows directly from the theorem.

We illustrate Proposition 4.7 with the following two examples.
Take $\mu x$.

Figure 9: Illustration for Example 4.8: (left) the disjoint sets $A$ and $F$ and (right) the projection of the point $x_0 - \mu x^*$ onto $A$ is the unique solution, $\bar{x}$, of the LP.

**Example 4.8.** Consider Problem (P) with $\mathcal{H} = \mathbb{R}^2$, $x^* = (0, 1)$ and $A := \{x \in \mathbb{R}^2 : Ax \leq b\}$, where $A = \begin{bmatrix} 1 & -1 \\ -1 & -1 \end{bmatrix}$ and $b = (0, 0)$. Our first task is to determine the modulus of sharpness of $A$. It is easy to check that, for every $x \in \text{bd}A$,

$$N_A(x) = \begin{cases} \text{cone}([-1, -1]) =: K_1 & \text{if } x_1 < 0, \\
\text{cone}([(1, -1)] =: K_2 & \text{if } x_1 > 0, \\
\text{cone}([-1, 1, 1, -1]) & \text{if } x_1 = 0,
\end{cases}$$

which can be graphically verified from Figure 9. Note that we have $-x^* \not\in N_A(x)$ if and only if $x_1 \neq 0$. Using (60) in Theorem 4.5, we have

$$\inf_{x \in A, -x^* \notin N_A(x)} d(-x^*, N_A(x)) = \min\{d(-x^*, K_1), d(-x^*, K_2)\} = \sqrt{2}/2 = 2(\sqrt{2}/4).$$

Hence, (60) holds with $\alpha := \sqrt{2}/2$. Consequently, $x_0$ will solve (P) if it verifies conditions 1 and 2 in Theorem 4.5. If only condition 1 holds, then we can use Proposition 4.7 and find $\mu_0$ s.t. $P_A(x_0 - \mu x^*)$ solves the LP for $\mu > \mu_0$. For instance, take $x_0 := (-1, -1/2)$. It is easy to check that $x_0$ verifies condition 1 in Theorem 4.5, but not condition 2, and that $d(x_0, A) = 3\sqrt{2}/4$. With the notation of Proposition 4.7 and $\alpha = \sqrt{2}/4$, we need to take $\mu$ such that

$$\mu > d(x_0, A)(4 - \alpha^2)/\alpha^2 = 21\sqrt{2}/4.$$

Take $\mu = 10 > 21\sqrt{2}/4$. So $u := x_0 - \mu x^* = (-1, (-1/2) - 10) = (-1, -21/2)$ with $P_A(u) = (0, 0)$, the solution of (P). An illustration of this example is shown in Figure 9. \hfill \Diamond

**Example 4.9.** Consider Problem (P) with $\mathcal{H} = \mathbb{R}^3$, $x^* = (0, 1/\sqrt{2}, 1/\sqrt{2})$ and $A := \{x \in \mathbb{R}^2 : Ax \leq b\}$, where $A = -I$ and $b = (0, 0, 0)$. Namely, $A = \mathbb{R}^3$. The solution of (P) is the set $\mathbb{R}^3 \times \{0\} \times \{0\}$ and zero is its optimal value. Again, we first determine the modulus of sharpness of $A$. It can be checked that the only cases in which $-x^* \in N_A(x)$ for $x \in \text{bd}A$ is when $x_1 \geq 0$ and $x_2 = x_3 = 0$. Hence, we need to compute $d(-x^*, N_A(x))$ for $x$ in the following set:

$$T := \{x \in \text{bd}A : x_1 \geq 0 \text{ and } x_2, x_3 \text{ are not simultaneously zero}\}$$

It can be checked that

$$\inf_{x \in T} d(-x^*, N_A(x)) = 1/\sqrt{2},$$

so (60) holds with $\alpha := 1/\sqrt{2}$. Take $x_0 := (1, -1, 0)$ then $d(x_0, A) = 1$, and $x_0$ verifies condition 1 in Theorem 4.5 but not condition 2. With the notation of Proposition 4.7, we need $\mu$ such that

$$\mu > (4 - \alpha^2)/\alpha^2 d(x_0, A) = 7.$$

Take $\mu = 7\sqrt{2}$ so $u := x_0 - \mu x^* = (1, -8, -7)$ with $P_A(u) = (1, 0, 0)$ a solution of (P). \hfill \Diamond
4.3 SPP for a general case

We extend next Theorem 4.5 to a problem where the objective function is an arbitrary convex and lsc function $f : \mathcal{H} \to \mathbb{R}_{\infty}$. Namely, we consider the convex problem

$$\min_{x \in A} f(x),$$

where $A$ is a closed and convex set. For this problem, we will assume that $f$ and $A$ are such that $\partial(f + \mathbb{1}_A) = \partial f + N_A$ over dom $f \cap A$. The latter is true, for instance, when some standard constraint qualification holds (see [2, Corollary 16.38]), e.g., when int $A \cap$ dom $f \neq \emptyset$ or int (dom $f) \cap A \neq \emptyset$. The function $f_A := (f + \mathbb{1}_A)$ will have a crucial role in the next result. Note that

$$\text{epi } f_A = \{(x, t) \in A \times \mathbb{R} : f(x) \leq t \} = (A \times \mathbb{R}) \cap \text{epi } f$$

By imposing a sharpness condition on the set epi $f_A$ w.r.t. the vector $(0, -1)$, we can recover a solution of problem (CP) by using Theorem 4.5 in the extended space $\mathcal{H} \times \mathbb{R}$.

Theorem 4.10. Suppose that the convex problem (CP) has solutions with optimal value $M := \inf_{x \in A} f(x)$ and nonempty set of optimal solutions denoted by $S$. Assume that the following conditions hold.

(i) The set epi $f_A := \tilde{A}$ given in (69) is $\alpha$-sharp w.r.t $z^* := (0, -1)$ for $\alpha \in (0, 1)$.

(ii) Let $(v, t) \in \mathcal{H} \times \mathbb{R}$ be such that

(a) $t < M$; and

(b) $(1 - (\alpha/2)^2)d((v, t), \tilde{A}) < (M - t)$.

In this situation, consider $P_A(v, t) = (w, f(w))$. Then $w \in S$ and hence $f(w) = M$.

Proof. With the notation of (i), problem (CP) is equivalent to the following problem

$$\min_{(x, s) \in \tilde{A}} s,$$

which has the same optimal value as (CP), and linear objective $\psi : \mathcal{H} \times \mathbb{R} \to \mathbb{R}$ defined as $\psi(x, s) := s = ((0, 1), (x, s))$. Note that (EP) is an optimization problem with a linear objective and by (i), its constraint set is $2\alpha$-sharp w.r.t $z^* := (0, -1)$. Take $\tilde{v} := (v, t) \in \mathcal{H} \times \mathbb{R}$ verifying assumptions (a)(b). We claim that this implies that conditions 1 and condition 2 in Theorem 4.5 hold, with $x^* := (0, 1)$, and $A := \tilde{A}$. Indeed, condition (a) re-writes as

$$t = \langle (0, 1), (v, t) \rangle < M = \inf_{A} f(x) = \inf_{(x, s) \in \tilde{A}} \langle (0, 1), (x, s) \rangle,$$

which is Condition 1 in Theorem 4.5 for $\tilde{v} := (v, t)$ and $x^* := (0, 1)$. Condition 2 in Theorem 4.5 follows directly from (b) and the definitions. Therefore, we are in conditions of Theorem 4.5, and $P_A(v, t)$ solves (EP).

By [2, Proposition 29.35], $P_A(v, t) = (w, f(w))$ where $w \in A$ is the unique solution of the inclusion

$$\frac{v - w}{f(w) - t} \in \partial f(w).$$

Note that $t < M \leq f(w)$ so $(f(w) - t) > 0$. Since $P_A(v, t) = (w, f(w))$ solves (EP) this means that $f(w) = M$ and since $w \in A$ we must have $w \in S$. \hfill \Box

Remark 4.11. We give in Proposition 3.14 a necessary and sufficient condition for assumption (i) in Theorem 4.10 to hold for $f_A$.

Remark 4.12. Let $A$ be a closed and convex set and assume that $f_A := f + \mathbb{1}_A$ is polyhedral. By Corollary 3.16, we know that there exists $\alpha < 1$ such that property (21) holds for $\beta := \alpha/\sqrt{1 - \alpha^2}$. The latter fact, combined with Proposition 3.14, imply that epi $f_A := \tilde{A}$ is $\alpha$-sharp w.r.t. $(0, -1)$. Hence, we are in the situation of Theorem 4.10. Therefore, if $(v, t)$ is such that $t < M$ and

$$\left(1 - \frac{\alpha^2}{4}\right)d((v, t), \tilde{A}) < M - t,$$

then $P_A(v, t) = (\bar{x}, f(\bar{x}))$ is such that $\bar{x}$ solves (CP). \hfill \Box
The following result considers problem (CP) on its own, and establishes a sharpness condition under which a projection onto $A$ solves (CP).

**Theorem 4.13.** Suppose that the convex problem (CP) has solutions and denote by $S$ the set of optimal solutions, and $\alpha \in (0,1)$. Assume that the following conditions hold.

(i) \[
\inf_{x \in A | S} d\left(-x^*/\|x^*\|, N_A(x)\right) \geq \alpha > 0, \tag{70}
\]
with the convention $\inf \emptyset = +\infty$ and $\frac{0}{\|0\|} = 0$.

(ii) There is $v \in H$ such that

(a) $f(v) < \inf_{x \in A} f(x)$; and

(b) $(1 - (\alpha/2)^2)d(v, A) < \inf_{x \in A} f(x) - f(v)$. 

In this situation, the projection of $v$ onto $A$ solves problem (CP).

**Proof.** Our assumption on problem (CP) implies that, at a solution $\bar{x} \in S$ there exists $x^* \in \partial f(\bar{x})$ such that $-x^* \in N_A(\bar{x})$. By (i) we know that $x^* \neq 0$, otherwise the left hand side of (70) equals 0. Let $M := \inf_A f$, we consider the sublevel set of $f$ at value $M$

$$F := \{x \in H : f(x) \leq M\}.$$ 

Then it is clear that $A \cap F = S$, the set of optimal solutions of problem (CP). From $x^* \in (-N_A(\bar{x})) \cap \partial f(\bar{x})$, we have

$$\langle x^*, x - \bar{x} \rangle \geq 0, \forall x \in A \quad \text{and} \quad 0 \geq f(x') - M = f(x') - f(\bar{x}) \geq \langle x^*, x' - \bar{x} \rangle, \forall x' \in F, \tag{71}$$

where the first inequality in the rightmost expression follows because $x' \in F$ and $\bar{x} \in S$. In other words, $x^*$ separates the (closed and convex) sets $A$ and $F$. Therefore, $A \cap F = S \subset \{x : \langle x^*, x \rangle = \langle x^*, \bar{x} \rangle\}$. In particular, this implies that

$$\langle x^*, z - \bar{x} \rangle = 0, \quad \tag{72}$$

for all $z \in S$. Take now any $y \in A$, $z \in S$. Use the left hand side of (71) and (72) to deduce that

$$\langle -x^*, y - z \rangle = \langle -x^*, y - \bar{x} \rangle + \langle -x^*, \bar{x} - z \rangle \leq 0,$$

and so

$$-x^* \in N_A(z), \quad \forall z \in S.$$ 

The inclusion above implies that $\langle x^*, z \rangle \leq \langle x^*, y \rangle$, for any $y \in A, z \in S$. In other words, $S \subset \text{argmin}_{A} \langle x^*, \cdot \rangle$. Namely,

$$S \subset \{z \in A : \langle x^*, z \rangle \leq \langle x^*, y \rangle \forall y \in A\} = \{z \in A : \langle -x^*, z \rangle \geq \langle -x^*, y \rangle \forall y \in A\} = F_A(-x^*),$$

and therefore $A \setminus S \supset A \setminus F_A(-x^*)$. Combining the latter inclusion with inequality (70) yields

$$\inf_{x \in A \setminus F_A(-x^*)} d\left(-x^*/\|x^*\|, N_A(x)\right) \geq \alpha. \tag{73}$$

The expression above means that $A$ is $\alpha$-sharp w.r.t. vector $-x^*/\|x^*\|$. Take $v \notin A$ such that $f(v) < M$ and $d(v, A) < \frac{M - f(v)}{1 - (\alpha/2)^2}$. Namely, $v$ verifies condition (ii). By construction, $v \in F$, and using this fact in the rightmost side of (71) yields

$$\langle x^*, v \rangle \leq \langle x^*, \bar{x} \rangle + (f(v) - M) < \langle x^*, \bar{x} \rangle \leq \inf_{x \in A} \langle x^*, x \rangle.$$ 

Hence $\langle x^*, v \rangle < \inf_{x \in A} \langle x^*, x \rangle$ and so $v$ verifies condition 1 in Theorem 4.5. We also have that $x^* \in \partial f(\bar{x})$, so

$$\langle x^*, \bar{x} \rangle - \langle x^*, v \rangle \geq f(\bar{x}) - f(v) = M - f(v) > 0,$$ 

where we also used the fact that $f(\bar{x}) = M$. By definition of $v$, we have

$$d(v, A) < \frac{M - f(v)}{1 - (\alpha/2)^2} \leq \frac{\langle x^*, \bar{x} \rangle - \langle x^*, v \rangle}{1 - (\alpha/2)^2} = \frac{\inf_{x \in A} \langle x^*, x \rangle - \langle x^*, v \rangle}{1 - (\alpha/2)^2},$$

28
where we used (74) in the second inequality and the fact that \( \bar{x} \in S \subseteq \arg\min_A \langle x^*, x \rangle \) in the equality. The expression above implies that \( v \) verifies condition 2 in Theorem 4.5. Since \( v \) satisfies both conditions in the theorem, we deduce that \( P_A(v) \) is a solution of \( \inf_{x \in A} \langle x^*, x \rangle \). Equivalently, \( P_A(v) \in F_A(-x^*) \). To complete the proof, we will show that \( S = F_A(-x^*) \). This will establish that \( P_A(v) \in S \), as wanted. We already know that \( S \subseteq F_A(-x^*) \), so it is enough to show that \( S \supset F_A(-x^*) \). Indeed, assume that there is \( z \in A \setminus S \) such that \( z \in F_A(-x^*) \). The fact that \( z \in F_A(-x^*) \) means that \( \langle x^*, z \rangle = \min_{x \in A} \langle x^*, x \rangle \), or equivalently, \( -x^* \in \partial f(z) \). The latter inclusion gives
\[
0 = d(-x^*/\|x^*\|, N_A(z)) \geq \inf_{x' \in A \setminus S} d(-x^*/\|x^*\|, N_A(x)) \geq \alpha > 0,
\]
where we are using (70) and the fact that \( z \in A \setminus S \) and \( x^* \in \partial f(\bar{x}) \) with \( \bar{x} \in S \) in the first inequality. The above expression entails a contradiction and hence we must have \( S = F_A(-x^*) \). Therefore, our claim is true and \( P_A(v) \in F_A(-x^*) = S \) and thus \( P_A(v) \) solves problem (CP).

## 5 Conclusions and open questions

In this work, we introduce the notation of a sharp set and use it to analyze the single projection procedure for solving convex optimization problem. To conclude, we outline the directions for future work.

1. The paper [3] proved the finite convergence of projection-type methods (e.g., alternating projections method, Douglas Rashford) between a closed half space and a polyhedral set for the cases when the two sets do not intersect. However, the authors in [3] do not provide an estimation of how many steps are required for the convergence. Can Theorem 4.5 be used to estimate the number of steps required for convergence of the projection-type algorithms analyzed in [3]?

2. The same results in Theorem 4.5 still hold for small perturbations of the linear function, namely \( \hat{x}^* \in H \) with \( \|\hat{x}^* - x^*\| \) small enough so that \( \hat{x}^* \in \mathop{\text{int}} \bigcup_{x \in F_A(x^*)} N_A(x) \). This allows inexact projections, reducing the computational effort. This observation is not trivial for nonlinear functions. However, it is worth to notice that SPP may ensure finite termination for nonlinear problems with inexact gradient oracles, provided that we consider an oracle with small error and implement projections with sufficiently high accuracy as a finite operation. The precise conditions for ensuring these properties are the topic of future research.

3. Bundle methods [18] are designed to minimize nonsmooth convex functions. These methods approximate the original function by a suitable piece-wise function, and the iterates are minimizers of these approximations. Since these approximations have a polyhedral epigraph, they will always be sharp sets (i.e., condition (i) in Theorem 4.10 will always hold for some \( \alpha > 0 \)). So, if the modulus of sharpness of the current approximation is known, conditions (a)(b) in Theorem 4.10 could potentially provide a simple way of computing the iterates.

4. From Proposition 3.10, for any polyhedral set, there is \( \alpha > 0 \) such that it is \( \alpha \)-sharp w.r.t. every unit vector. It is interesting to ask whether the converse statement is true, namely: If a set is \( \alpha \)-sharp w.r.t. every unit vector for some \( \alpha > 0 \), then is it true that the set must be polyhedral?

**Acknowledgements**

The research of EAN was supported by the Ministry of Science and Higher Education of the Russian Federation, project 075-02-2022-880 from 31.01.2022. HTB is supported by the Australian Research Council through the Centre for Transforming Maintenance through Data Science (grant number IC180100030). MKT would like to thank HTB for her hospitality during his stay at Curtin University supported by a Small Grant from Faculty of Science and Engineering, Curtin University.

**References**

[1] Bauschke, H. H., The approximation of fixed points of compositions of nonexpansive mappings in Hilbert space. *Journal of Mathematical Analysis and Applications*, 202, 1996, 150–159.

[2] Bauschke, H. H. and Combettes, P. L. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces, second edition*. CMS Books in Mathematics, Springer, 2017.
[3] Behling, R., Bello-Cruz, Y., Santos, L. R. Infeasibility and error bound imply finite convergence of alternating projections. *SIAM Journal on Optimization*, 31(4), 2021, 2863–2892.

[4] Bolte, J., Nguyen, T. P., Peypouquet, J., Suter, B.W. From error bounds to the complexity of first-order descent methods for convex functions *Mathematical Programming Series A*, 165, 2017, 471–507.

[5] Bui, H. T., Loxton, R., Moeneclaey, A. A note on the finite convergence of alternating projections. *Operations Research Letters*, 49(3), 2021, 431–438.

[6] Burachik, R. S., Iusem, A. N. *Set-Valued Mappings And Enlargements Of Monotone Operators* Optimization and Its Applications, Volume 8, Springer, 2008.

[7] Brezis, H. *Functional Analysis, Sobolev Spaces and Partial Differential Equations*. Springer Science & Business Media, 2011.

[8] Chakrabarty, A. K., Shunmugaraj, P. and Zalinescu, C. Continuity Properties for the Subdifferential and ε-Subdifferential of a Convex Function and its Conjugate. *Journal of Convex Analysis*, 14, 2007, 353–342.

[9] Drusvyatskiy, D., Ioffe, A. D. and Lewis, A. S.: Transversality and alternating projections for nonconvex sets. Foundations of Computational Mathematics, 2015, 637-1651.

[10] Ekeland, I. On the variational principle. *Journal of Mathematical Analysis and Applications*, 47, 1974, 353–342.

[11] Hermer, N, Luke, D. R. and Sturm, A. Random function iterations for consistent stochastic feasibility. *Numerical Functional Analysis and Optimization*, 40:4, 2019, 386–420.

[12] Jeyakumar, V.: Generalizations Farkas’ lemma. In: *Encyclopedia of Optimization*, Springer, 2001, 627–631.

[13] Kruger, A. Y., Luke, R., D. and Thao, N.H. About subtransversality of collections of sets. *Set-Valued and Variational Analysis*, Springer, 2017, 701–729.

[14] Li, G. and Pong, T. K. Calculus of the exponent of Lurdyka–Lojasiewicz inequality and its applications to linear convergence of first-order methods. *Foundations of Computational Mathematics*, 18, 2018, 1199–1232.

[15] Ng, K. F. and Song, W.: Fenchel duality in infinite-dimensional setting and its applications. *Nonlinear Analysis*, 55, 2003, 845–858.

[16] Ng, K. F., Yang, W. H. Regularities and their relations to error bounds. *Mathematical Programming Series A*, Springer, 2004, 521–538.

[17] Nurminski, E. Single-projection procedure for linear optimization. *Journal of Global Optimization*, Springer, 2015, 1–16.

[18] Oliveira, W., Sagastizábal, C. and Lemaréchal, C., Convex proximal bundle methods in depth: a unified analysis for inexact oracles. *Mathematical Programming Series B*, 148, 2014, 241–277

[19] Rockafellar, R. T. and Wets, R. J.-B.: *Variational Analysis*. Springer Science & Business Media, 2009.

[20] Schirotzek, W.: *Nonsmooth Analysis*. Springer, Universitext, 2007.

[21] Zalinescu, C. *Convex Analysis in General Vector Spaces*. World Scientific Publishing Co.Inc., River Edge, NJ, 2003.