SIAM J. OPTIM.
Vol. 34, No. 2, pp. 1646–1678

SINGLE-PROJECTION PROCEDURE FOR INFINITE DIMENSIONAL CONVEX OPTIMIZATION PROBLEMS*

HOA T. BUI†, REGINA S. BURACHIK‡, EVGENI A. NURMINSKI§, AND MATTHEW K. TAM¶

Abstract. We consider a class of convex optimization problems in a 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 complementarity conditions, and (iii) possess nonlinear 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.

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

MSC codes. 90C05, 90C25, 49J53, 49J52

DOI. 10.1137/22M1530173

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

$$\text{(P)} \qquad \min_{x \in A} \langle x^*, x \rangle ,$$

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, the linear programming problem (P) has the following 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 the subject for study by the authors in [3, 6], as a consequence of the finite convergence property of the alternating projection method

*Received by the editors October 21, 2022; accepted for publication (in revised form) January 22, 2024; published electronically May 3, 2024.
https://doi.org/10.1137/22M1530173

Funding: The first author's research was supported by the Australian Research Council through the Centre for Transforming Maintenance through Data Science (grant IC180100030). The third author's research was supported by the Ministry of Science and Higher Education of the Russian Federation, project 075-02-2022-880 from 31.01.2022. The fourth author would like to thank the first author for her hospitality during his stay at Curtin University, supported by the Travel Grant from Faculty of Science and Engineering, Curtin University.

†ARC Training Centre for Transforming Maintenance through Data Science, and Curtin Centre for Optimisation and Decision Science, Curtin University, Australia (bth.hoa92@gmail.com).
‡STEM Mathematics, University of South Australia, Australia (regina.burachik@unisa.edu.au).
§Center for Research and Education in Mathematics (CREM), Far-Eastern Federal University, Vladivostok, Russia (nurminskiy.ea@dvfu.ru).
¶School of Mathematics and Statistics, University of Melbourne, Australia (matthew.tam@unimelb.edu.au).

(a) Given a vector x^0, we can find θ large enough such that the projection of the vector $x^0 - \theta x^*$ onto A solves (P). This set A is sharp w.r.t. x^*.

(b) For x^0 as in the figure, there is no $\theta > 0$ such that the projection of the vector $x^0 - \theta x^*$ onto A solves (P). This set A is not sharp w.r.t. x^*.

FIG. 1.1. *The single-projection procedure extends beyond linear programming. As shown in Figure* 1.1(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).

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 nonlinear convex constraints, nonunique solutions, and/or in the absence of the strict complementarity property; and (iii) provide quantitative estimates on the value of θ_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 1.1), 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 nonoptimal 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 θ is *sufficiently large* and x^0 is *sufficiently far from being optimal* in the sense that

$$\left\langle x^*, x^0 - \theta x^* \right\rangle = \left\langle x^*, x^0 \right\rangle - \theta < \inf_{x \in A} \left\langle x^*, x \right\rangle.$$

Furthermore, as polyhedral sets are sharp with respect to every unit vector (see Proposition 3.14), 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 Theorem 4.13 and its corollary.

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. Roughly speaking, 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 vector x^* (see Definition 3.1). In order to contextualize sharpness in the broad 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 $(0_{\mathcal{H}}, -1)$ if and only if it satisfies a global *Kurdyka–Łojasiewicz (KL)* property with exponent 0 (Proposition 3.15). 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 complementarity. 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 3.20 that sharpness holds under much weaker conditions; in particular, we allow nonunique optimal solutions. Namely, we require only that the interior of the union of the normal vectors of A at all optimal solutions must contain $-x^*$. Consequently, although the strict complementarity 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 3.14).
- *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 3.28). As such, the sharpness property can be connected with well-known geometric properties in the literature.

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 geometric properties of sets. Section 4 contains our main results on the SPP for solving problem (P) and its extension to general convex problems in which the objective function is not necessarily linear. 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, $\mathcal{H}$ is a real Hilbert space with inner product $\langle\cdot,\cdot\rangle$ and induced norm $\|\cdot\|$. Given a set $C \subset \mathcal{H}$, the distance from C to x is denoted by $d_C(x) := \inf_{z\in C}\|x - z\|$.

We use the notation $\mathbb{R}_\infty := \mathbb{R} \cup \{\infty\}$. Given $A \subset \mathcal{H}$, we denote by int$A$, cl$A$, and bd$A$ its topological interior, closure, and boundary, respectively. Unless specifically mentioned, we consider the strong topology in $\mathcal{H}$. We will denote by $\mathcal{B} := \{u \in \mathcal{H} : \|u\| \le 1\}$ 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

$$\|(x, y)\|_* := \|x\| + \|y\|. \tag{2.1}$$

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

$$B^*_{\mathcal{H}^2} := \{(u, v) \in \mathcal{H}^2 : \|u\| + \|v\| \le 1\}. \tag{2.2}$$

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 $f\colon \mathcal{H} \to \mathbb{R}_\infty$. The set $\operatorname{dom} f := \{x \in \mathcal{H} : f(x) < \infty\}$ is the *domain* (or *effective domain*) of f. We say that $f\colon \mathcal{H} \to \mathbb{R}_\infty$ is proper if $\operatorname{dom} f \neq \emptyset$. Suppose f is proper and the function $f^*\colon \mathcal{H} \to \mathbb{R}_\infty$ 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 $\operatorname{epi} f := \big\{(x, r) \in \mathcal{H} \times \mathbb{R} : f(x) \le r\big\}$. 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., $\operatorname{epi} f$ is closed in the weak topology). Given a convex function f, recall that the *subdifferential* of f is the point-to-set mapping $\partial f\colon \mathcal{H} \rightrightarrows \mathcal{H}$ defined by

$$\partial f(x) := \begin{cases} \{x^* \in \mathcal{H} : (\forall y \in \mathcal{H})\ \langle y - x, x^* \rangle + f(x) \le f(y)\} & \text{if } x \in \operatorname{dom} f, \\ \emptyset & \text{if } x \notin \operatorname{dom} f. \end{cases} \tag{2.3}$$

Note that for points at the boundary of $\operatorname{dom} f$, the subdifferential may or may not exist (i.e., the set in the first line in (2.3) 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* $\mathrm{R}(T) := \{v \in \mathcal{H} : \exists x \in \mathcal{H},\ v \in T(x)\}$.
(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 $\mathbb{R}_+(u) := \{tu : t \ge 0\} = \operatorname{cone}[u]$ the *cone* (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 $\operatorname{cone}[u_i, i \in J]$ the convex cone generated by the collection. If $J = \{1, \ldots, l\}$ is finite, then

$$\operatorname{cone}[u_1, \ldots, u_l] = \sum_{i=1}^{l} \operatorname{cone}[u_i] = \sum_{i=1}^{l} \mathbb{R}_+(u_i), \tag{2.4}$$

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

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

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

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

FACT 2.4 (see [2, Proposition 16.35]). *Let $f\colon \mathcal{H} \to \mathbb{R}_\infty$ be a convex, proper, and lsc function. Denote by $D := \operatorname{dom} f$ and by $E := \operatorname{epi} f$. Fix $x \in D$. Then, $N_E((x, f(x)) =$*

$\mathbb{R}_{++}(\partial f(x)\times\{-1\})\bigcup(N_D(x)\times\{0\})$, where $\mathbb{R}_{++}V:=\{t\,v\,:\,t>0, v\in V\}$ denotes the positive cones generated by a set V. Equivalently,

$$
\begin{aligned}
N_E((x,f(x)) = \{(u,\eta)\in\mathcal{H}\times\mathbb{R} :\ \eta<0 \text{ and } u/(-\eta)\in\partial f(x)\}\\
\cup\{(u,0)\in\mathcal{H}\times\mathbb{R} :\ u\in N_D(x)\}.
\end{aligned}
\tag{2.5}
$$

By [2, *Proposition* 16.17(i)], *for every $x\in \mathrm{bd}D$, we can either have $\partial f(x)=\emptyset$ or $\partial f(x)$ unbounded. In the latter case, i.e., when $x\in\mathrm{bd}D$ 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, $\mathrm{dom}\,\mathbb{1}_A=A$. If A is a nonempty, closed and convex set, $\mathbb{1}_A$ is a proper, lsc, and convex function. In this situation, (2.3) yields

$$
\partial(\mathbb{1}_A(x))=N_A(x)\quad \text{ for all } x\in\mathcal{H},
\tag{2.6}
$$

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 *demiclosedness* (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 2.5. 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 φ_{z_0} is given by $\varphi^*_{z_0}=\mathbb{1}_{\mathcal{B}_*}+\langle z_0,\cdot\rangle$, where $\mathcal{B}_*$ is the closed 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]),

$$
\partial\varphi_{z_0}(x)=\begin{cases}\{(x-z_0)/\|x-z_0\|\} & \text{if } x\neq z_0,\\ \mathcal{B}_* & \text{if } x=z_0.\end{cases}
\tag{2.7}
$$

Consequently, φ_{z_0} is smooth at every $x\neq z_0$ (see [2, Proposition 17.32]).

Remark 2.6. For future use, we recall here a fact involving the subdifferential of a maximum of two norms in the product space. Fix $\hat z=(\hat z_1,\hat z_2)\in\mathcal{H}\times\mathcal{H}$. Consider the function $\Theta_{\hat z}:\mathcal{H}\times\mathcal{H}\to\mathbb{R}$ defined as

$$
\Theta_{\hat z}(x,y):=\max\{\|x-\hat z_1\|,\|y-\hat z_2\|\}=\max\{\varphi_{\hat z_1}(x),\varphi_{\hat z_2}(y)\},
$$

where we are using the notation introduced in Remark 2.5 in the second equality. Note that

$$
\Theta_{\hat z}(\hat z_1,\hat z_2)=0=\varphi_{\hat z_1}(\hat z_1)=\varphi_{\hat z_2}(\hat z_2).
\tag{2.8}
$$

Denote by ∂_1 and ∂_2 the partial subdifferentials w.r.t. the first and second variable, respectively. Define also $\Theta^1(x,y):=\varphi_{\hat z_1}(x)$ for every $x,y\in\mathcal{H}$ and $\Theta^2(x,y):=\varphi_{\hat z_2}(y)$ for every $x,y\in\mathcal{H}$ so that $\Theta_{\hat z}(x,y)=\max\{\Theta^1(x,y),\Theta^2(x,y)\}$. Using Remark 2.5 and the fact that Θ^1 does not depend on the second variable, we have that

$$
\begin{aligned}
\partial\Theta^1(\hat z_1,\hat z_2)&=\big(\partial_1\Theta^1(\hat z_1,\hat z_2),\partial_2\Theta^1(\hat z_1,\hat z_2)\big)=\big(\partial_1\Theta^1(\hat z_1,\hat z_2),0\big)\\
&=(\partial\varphi_{\hat z_1}(\hat z_1),0)=\mathcal{B}_*\times\{0\},
\end{aligned}
$$

where we used (2.7) in the last equality. Similarly, we obtain $\partial\Theta^2(\hat{z}_1,\hat{z}_2) = (\partial_1\Theta^2(\hat{z}_1,\hat{z}_2), \partial_2\Theta^2(\hat{z}_1,\hat{z}_2)) = \{0\} \times \mathcal{B}_*$. The classical theorem of Dubovitskii and Milyutin (see [2, Theorem 18.5]), which computes the subdifferential of a maximum of functions, together with (2.8), yields

$$\begin{aligned}\partial\Theta_{\hat{z}}(\hat{z}_1,\hat{z}_2) &= \overline{\text{conv}}\{(B_* \times \{0\}) \cup (\{0\} \times B_*)\} = \{(u,v) \in \mathcal{H} \times \mathcal{H} : \|u\| + \|v\| \leq 1\} \\ &= \mathcal{B}^*_{\mathcal{H}^2}.\end{aligned}$$

(Here, conv (A) denotes the *convex hull* of the set A, which is the smallest convex set that contains A, whereas $\overline{\text{conv}}(A)$ 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 (2.2).

DEFINITION 2.7. *Fix $m \in \mathbb{N}^*$ (i.e., $m \in \mathbb{N}$ and $m \neq 0$) and $b \in \mathbb{R}^m$. Let $\mathbb{A} : \mathcal{H} \to \mathbb{R}^m$ be a bounded linear operator. Recall that the* adjoint operator of $\mathbb{A}$ *is the (bounded) linear map $\mathbb{A}^* : \mathbb{R}^m \to \mathcal{H}$ defined by the equality*

$$\langle \mathbb{A}x, u\rangle = \langle x, \mathbb{A}^* u\rangle \quad \forall (x,u) \in \mathcal{H} \times \mathbb{R}^m. \tag{2.9}$$

Note that in (2.9) we are using the same notation for the inner products in $\mathcal{H}$ and in $\mathbb{R}^m$. From [5, Remark 16], we have that, when $\mathbb{A}$ is bounded, $\mathbb{A}^*$ is also bounded and both maps have the same norm.

DEFINITION 2.8. *Take $\mathbb{A}$ and b as in Definition 2.7. Denote by $\mathcal{B}_m := \{e_1, \dots, e_m\}$ the canonical basis in $\mathbb{R}^m$. A* polyhedron or polyhedral set *induced by a linear map $\mathbb{A}$ and a vector b is the set*

$$C(\mathbb{A}, b) := \{x \in \mathcal{H} : \mathbb{A}x \leq b\} := \{x \in \mathcal{H} : \langle \mathbb{A}x, e_j\rangle \leq b_j \text{ for all } j = 1, \dots, m\}. \tag{2.10}$$

Hence, a polyhedral set is a finite intersection of level sets of linear maps. Given $x \in \mathcal{H}$, the set $I(x) := \{j \in \{1, \dots, m\} : \langle \mathbb{A}x, e_j\rangle = b_j\}$ 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 manner similar to that in Definition 2.8, where $\mathcal{H}$ is replaced by a locally convex space denoted by X.

PROPOSITION 2.9 (see [15, Corollary 4.1]). *Let X be a locally convex space, and let $C_1, \dots, C_m$ be polyhedral subsets of X with $C := \cap_{i=1}^m C_i \neq \emptyset$. 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 2.10 (Ekeland's variational principle [10, Theorem 1.1]). *Let X be a complete metric space, let $\psi : X \to \mathbb{R}_\infty$ be lsc, let $\bar{w} \in X$, and let $\varepsilon > 0$. If $\psi(\bar{w}) < \inf_{w \in X} \psi(x) + \varepsilon$, then, for any $\lambda > 0$, there exists $\hat{w}_\lambda \in X$ such that the following hold:*

(i) $d(\hat{w}_\lambda, \bar{w}) < \lambda$;
(ii) $\psi(\hat{w}_\lambda) \leq \psi(\bar{w})$;
(iii) $\psi(w) + (\varepsilon/\lambda) d(w, \hat{w}_\lambda) > \psi(\hat{w}_\lambda)$ *for all* $w \in X \setminus \{\hat{w}_\lambda\}$.

The next simple fact will be used in later sections.

FACT 2.11. *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, $d(u, \text{cone}[v])^2 = \min_{t\geq 0} \|u - tv\|^2 = \min_{t\geq 0}(1 + t^2 - 2t_*t) = 1 - \max_{t\geq 0}(-t^2 + 2t_*t) = 1 - \max(0, t_*)^2$. ◻

3. Sharp sets and their characterizations. Figure 1.1 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 $x^* \neq 0$ *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 $A \subset \mathcal{H}$ be closed and convex, let $x^* \in \mathcal{S}$, and let $\alpha > 0$. We say that A is* α-sharp with respect to x^* *if, for all $x \in A$ such that $x^* \notin N_A(x)$, we have*

$$d\left(x^*, N_A(x)\right) \geq \alpha. \tag{3.1}$$

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

$$sr[A, x^*] := \inf_{x\in A, x^*\notin N_A(x)} d(x^*, N_A(x)). \tag{3.2}$$

Then, A is sharp w.r.t. vector x^* *if* $sr[A, x^*] > 0$.

Remark 3.3. From the definition, A is α-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 $\langle x^*, x\rangle = \sup_A \langle x^*, \cdot\rangle$ for all $x \in A$. Hence, $A = F_A(x^*)$. In this case, A is trivially sharp w.r.t. x^* by vacuity, and from (3.2), with the convention that $\inf \emptyset = +\infty$, we deduce that $\text{sr}[A, x^*] = +\infty$. We also note that, since $\|x^*\| = 1$ and $0 \in N_A(x)$ for any $x \in A$, we always have that $\alpha \leq 1$ if the set $\{x \in A : x^* \notin N_A(x)\}$ is nonempty.

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

FACT 3.4. *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

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

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 $F_A(x^*) = \operatorname{Argmax}\langle x^*, \cdot\rangle$. To complete the proof, note that the equivalence between (i) and (ii) implies that $x \in A \setminus F_A(x^*) \Longleftrightarrow x \in A$ and $x^* \notin N_A(x)$. Therefore,

$$\begin{aligned} \operatorname{sr}[A, x^*] \geq \alpha \quad &\Longleftrightarrow \quad d(x^*, N_A(x)) \geq \alpha \quad \forall x \in A, x^* \notin N_A(x), \\ &\Longleftrightarrow \quad d(x^*, N_A(x)) \geq \alpha \quad \forall x \in A \setminus F(x^*), \end{aligned}$$

establishing the last claim. ∎

Remark 3.5. Figure 3.1 illustrates simple two-dimensional examples of the sharpness property with respect to a given vector and its connection with faces of the set A. In Figure 3.1(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 3.1. We illustrate this situation in the next example.

Example 3.6. Consider the set $A \subset \mathbb{R}^2$ as the epigraph of the convex function $f : \mathbb{R} \to \mathbb{R}_\infty$ defined as follows: $f(x) := +\infty$ if $x < -1$, $f(x) := -\frac{(3n^2+3n+1)x+(2n+1)}{n^2(n+1)^2}$ if $x \in [-1/n, -1/(n+1)]$, $n \in \mathbb{N}$, and $f(x) := 0$ if $x \geq 0$. The boundary of the set A to the left of 0 is determined by a piecewise 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} $(n = 1, 2, \dots)$. This function is nonsmooth 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 3.7. When the 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 $\operatorname{sr}[A, x^*] = 0$, then the definition directly implies the existence of a sequence $(x_n, x_n^*) \subset G(N_A)$ with $x^* \notin N_A(x_n)$ and s.t. $x_n^* \to x^*$ strongly. The boundedness of A implies that there exist $\bar{x}$ 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 demiclosed, $(\bar{x}, x^*) \in G(N_A)$. 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 $(x_n, x_n^*) \subset G(N_A)$ s.t. the following hold:

(i) $x_n^* \to x^*$ strongly; $x^* \notin N_A(x_n)$.

(ii) There exists $\bar{x} \in A$ s.t. $x_n \to \bar{x}$ weakly, with $x_n \neq \bar{x}$ for all n.

In particular, in Figure 3.1(a), for any positive number $r > 0$, there exist $y^* \in B_r(x^*)$ 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 Figures 3.1(a) and (b). However, the point $\bar{x}$ in Figure 3.1(b) is an *exposed*

(a) The set A is not sharp w.r.t.x^*: there exist a sequence $(x_n, x_n^*) \subset \operatorname{bd} A \times \mathcal{H}$ with $x_n^* \in N_A(x_n)$, $x_n^* = x^*$ $(n \in \mathbb{N})$ such that (x_n) converges to $\bar{x}$, and (x_n^*) converges to x^*.

(b) The set A is sharp w.r.t. x^*: there exists $r > 0$ such that, for any $y^* \in B_r(x^*)$ and $x \in A$ with $y^* \in N_A(x)$, it must hold that $x^* \in N_A(x)$.

FIG. 3.1. *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^* in Figure* 3.1(a). *In Figure* 3.1(b)*, there exists no sequence (x_n) outside the face $A \cap F$, such that $x^* \in \operatorname{cl} \bigcup_n N_A(x_n)$. See Remark* 3.5.

point (i.e., there exists a hyperplane H such that $A \cap H = \{\bar{x}\}$), whereas the point $\bar{x}$ in Figure 3.1(a) is not an exposed point (since the only supporting hyperplane of A at $\bar{x}$ is F). In Figure 3.1(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 $d(x^*, N_A(y)) \geq \alpha$.

Remark 3.8. From Fact 3.4(i)–(ii), we know that $x^* \notin \mathrm{R}(N_A)$ 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 $\mathrm{R}(N_A) = \mathcal{H}$. Namely, there is no x^* verifying $x^* \notin \mathrm{R}(N_A)$. We also note that, in general, the set $\mathrm{R}(N_A)$ can neither be closed nor convex. However, it is well known (see, e.g., [2, Corollary 21.25] or [7, Theorem 4.4.9]) that $\mathrm{cl}\ \mathrm{R}(N_A)$ is a closed and convex set.

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

PROPOSITION 3.9. *Fix $x^* \in \mathcal{S}$, and assume that $A \subset \mathcal{H}$ is a nonempty, closed, and convex set such that $x^* \notin R(N_A)$. Then, A is sharp w.r.t. x^* if and only if $x^* \notin \mathrm{bd}\,\mathrm{R}(N_A)$. In this situation, A must be unbounded.*

Proof. The fact that A is unbounded follows from Remark 3.8. We show first that if $x^* \in \mathrm{bd}\ \mathrm{R}(N_A)$, then A is not sharp w.r.t. x^*. Indeed, in this case, we can take a sequence $(x_n) \subset A$ such that $x^* \neq x_n^* \in N_A(x_n)$ with $x^* = \lim_n x_n^*$. The fact that $x^* \neq x_n^*$ holds because $x^* \notin \mathrm{R}(N_A)$. 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 $x^* \notin \mathrm{bd}\,\mathrm{R}(N_A)$. Since we also have that $x^* \notin \mathrm{R}(N_A)$, we deduce that $x^* \notin \mathrm{cl}\ \mathrm{R}(N_A) =: R$. Then, there exists $\alpha > 0$ such that $\alpha = d(x^*, R) = \inf_{v \in \mathrm{R}(N_A)} d(x^*, v) \leq \inf_{x \in A,\, x^* \notin N_A(x)} d(x^*, N_A(x))$, so A is sharp w.r.t. x^*. ∎

We illustrate the last result in the next two examples. In the first example, we have that $x^* \notin \mathrm{R}(N_A)$ but $x^* \in \mathrm{bd}\ \mathrm{R}(N_A)$, while in the second example, $x^* \notin \mathrm{cl}\ \mathrm{R}(N_A)$.

Example 3.10. If $A := \mathrm{epi}\, g$, where $g : \mathbb{R} \to \mathbb{R}_\infty$ is defined as $g(t) = 1/t$ if $t > 0$ and $g(t) = +\infty$ otherwise, then it is easy to check that $x^* := (0, -1) \notin \mathrm{R}(N_A)$ but $x^* \in \mathrm{bd}\,\mathrm{R}(N_A)$. Hence, we are not in the conditions of Proposition 3.9. Let us check that A is not sharp w.r.t. x^*. Indeed, for $x_n := (n, 1/n)$ we have $(-1/n^2, -1) \in N_A(x_n)$ and $x^* = (0, -1) = \lim_{n \to \infty}(-1/n^2, -1)$, so $d(x^*, N_A(n, 1/n)) = 0$, and hence A is not sharp w.r.t. x^*.

Example 3.11. Let $A = \mathbb{R}^2_+$ be the nonnegative orthant in $\mathbb{R}^2$, and let $x^* = (1/\sqrt{2})(1, 1)$. Then, $x^* \notin \mathrm{cl}\ \mathrm{R}(N_A)$ and we are in the conditions of Proposition 3.9. It can be verified that A is 1-sharp w.r.t. x^*. Indeed, if $x \in \mathrm{int} A$, 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 Fact 2.11 with $u := x^*$ and $v := (0, -1)$, we obtain $d(x^*, N_A(x)) = 1$. An identical argument shows that $d(x^*, N_A(x)) = 1$ for $x = (0, t)$ for $t > 0$. In the latter case, we use Fact 2.11 with $u := x^*$ and $v := (-1, 0)$. Finally, if $x = (0, 0)$, then $N_A(x) = \mathbb{R}^2_-$, so again we have $d(x^*, N_A(x)) = d\left(x^*, \mathbb{R}^2_-\right) = 1$.

We show in the next result that when the set A is bounded, the constant $\mathrm{sr}[A, x^*]$ is exactly the infimum of the distance between x^* and every vector y^* such that the face $F_A(y^*)$ does not intersect with the face $F_A(x^*)$. This observation holds true even in the infinite dimensional setting.

PROPOSITION 3.12. *Suppose A is a nonempty, bounded, and closed convex set. Fix $x^* \in \mathcal{S}$ and $\alpha \in (0,1]$. Then, A is α-sharp w.r.t. x^* if and only if, for every $y^* \in \mathcal{H} \setminus \{0\}$ such that $F_A(y^*) \cap F_A(x^*) = \emptyset$, we have*

$$\|x^* - y^*\| \geq \alpha. \tag{3.3}$$

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

Proof. Note that because the set A is nonempty and bounded, from [2, Corollary 21.25], the set $\mathrm{Argmax}_A \langle y^*, \cdot \rangle$ is nonempty for any $y^* \in \mathcal{H} \setminus \{0\}$; therefore, from Fact 3.4, the face $F_A(y^*)$ is also nonempty.

First, suppose that the set A is α-sharp w.r.t. $x^* \in \mathcal{S}$. If $F_A(x^*) \cap F_A(y^*)$ is nonempty for every $y^* \in \mathcal{H} \setminus \{0\}$, then (3.3) in Proposition 3.12 holds true immediately. Otherwise, assume that there exists $y^* \in \mathcal{H} \setminus \{0\}$ such that $F_A(y^*) \cap F_A(x^*) = \emptyset$. By Fact 3.4, 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 3.4 yields $x^* \notin N_A(y)$. Then, the α-sharpness property implies that

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

which shows (3.3). The second statement follows directly by taking infimum in the expression above.

Conversely, suppose that for every $y^* \in \mathcal{H} \setminus \{0\}$ 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 α-sharp w.r.t. x^*. In fact, suppose to the contrary that the set A is not α-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

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

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

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

The above expression yields $\langle y^* - \varepsilon x^*, x \rangle < \sup_{z \in A} \langle y^* - \varepsilon x^*, z \rangle$, which implies that $x \notin F_A(y^* - \varepsilon x^*)$. 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 (3.3) for $y^* - \varepsilon x^*$ instead of y^* to obtain $\|(y^* - \varepsilon x^*) - x^*\| \geq \alpha$. On the other hand, the definition of ε yields

$$\|(y^* - x^*) - \varepsilon x^*\| \leq \|y^* - x^*\| + \|\varepsilon x^*\| = \|y^* - x^*\| + \varepsilon < \alpha,$$

in contradiction with assumption (3.3). Therefore, we must have that A is α-sharp w.r.t. x^*. ∎

Remark 3.13. If the set A is unbounded, the statements in Proposition 3.12 need not hold. Indeed, if $x^* \notin \mathrm{R}(N_A)$, then Fact 3.4 yields $F_A(x^*) = \emptyset$ so $F_A(x^*) \cap F_A(y^*) = \emptyset$ for all $y^* \in \mathcal{H}$. In this situation, inequality (3.3) will not hold for every nonzero $y^* \in \mathcal{H}$. This situation is illustrated by Example 3.11. In this example, $x^* = (1/\sqrt{2})(1,1) \notin \mathrm{R}(N_A)$, $F_A(x^*) = \emptyset$, and A is α-sharp at x^* with $\alpha = 1$. In particular, $F_A(x^*) \cap F_A(y^*) = \emptyset$ if $y^* = x^*$, in which case (3.3) is clearly false.

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

PROPOSITION 3.14. *Let A be a polyhedron defined by $A := \{x : \mathbb{A}x \le b\}$, where $\mathbb{A}$ is a bounded linear operator $\mathbb{A} : \mathcal{H} \to \mathbb{R}^m$ $(m > 0)$, and $b \in \mathbb{R}^m$. Denote by $\mathcal{B}_m := \{e_1, \ldots, e_m\}$ the canonical basis in $\mathbb{R}^m$. For any vector $x^* \in \mathcal{S}$, let $J(x^*)$ be a collection of subsets of indexes defined as $J(x^*) := \{J \subset \{1, \ldots, m\} : x^* \notin \operatorname{cone}[\mathbb{A}^* e_i]_{i\in J}\}$, and*

$$\alpha_0 := \min\left\{1, \inf_{J\in J(x^*)} d\left(x^*, \operatorname{cone}[\mathbb{A}^* e_i]_{i\in J}\right)\right\}, \tag{3.4}$$

with the convention that $\inf \emptyset = +\infty$. *Then,* $\alpha_0 > 0$ *and A is α_0-sharp w.r.t. x^*.*

Proof. We first prove that the constant α_0 defined in (3.4) is positive. If the set $J(x^*)$ is empty, then (3.4) becomes $\alpha_0 = \min\{1, \inf \emptyset\} = 1 > 0$. So, it suffices to consider the case where $J(x^*) \neq \emptyset$. To this end, let $J \in J(x^*)$ be arbitrary. By definition, $x^* \notin \operatorname{cone}[\mathbb{A}^* e_i]_{i\in J}$ and hence

$$d\left(x^*, \operatorname{cone}[\mathbb{A}^* e_i]_{i\in J}\right) > 0. \tag{3.5}$$

From inequality (3.5) and the fact that the set $J(x^*)$ is finite, the constant α_0 in (3.4) must be positive.

To complete the proof, we show that the constant α_0 defined in (3.4) is a lower bound of all constants α such that A is α-sharp w.r.t. x^*. In other words, we will show that $\operatorname{sr}[A, x^*] \ge \alpha_0 > 0$. Observe that the polyhedral set A is the intersection of m closed half spaces. Namely, using Definition 2.8 we can write

$$A = \bigcap_{i=1}^{m} \{x : \langle e_i, \mathbb{A}x\rangle \le \langle e_i, b\rangle\} = \bigcap_{i=1}^{m} \{x : \langle \mathbb{A}^* e_i, x\rangle \le \langle e_i, b\rangle\}, \tag{3.6}$$

where we used (2.9) in the second equality. Because $\mathbb{A}$ is a bounded linear map, so is $\mathbb{A}^*$, and hence for each $i = 1, \ldots, m$, the set $H_i := \{x : \langle \mathbb{A}^* e_i, x\rangle \le \langle e_i, b\rangle\}$ is a closed half space. Thus, the normal cone operator of H_i at a point $x \in H_i$ is given by $N_{H_i}(x) = \operatorname{cone}[\mathbb{A}^* e_i]$ if $\langle \mathbb{A}^* e_i, x\rangle = \langle e_i, b\rangle$, and $N_{H_i}(x) = \{0\}$ otherwise. We now apply the intersection rule in Proposition 2.9 and (2.4) to derive the normal cone of A as

$$N_A(x) = \operatorname{cone}[\mathbb{A}^* e_i]_{i\in I(x)} = \sum_{i\in I(x)} \operatorname{cone}[\mathbb{A}^* e_i] \quad \forall x \in A, \tag{3.7}$$

where $I(x) := \{i : \langle x, \mathbb{A}^* e_i\rangle = b_i\}$. Consider any $x \in A$ such that $x^* \notin N_A(x)$. Then, by (3.7) and the definition of $J(x)$, we have $I(x) \in J(x)$, and

$$d\left(x^*, N_A(x)\right) = d\left(x^*, \operatorname{cone}[\mathbb{A}^* e_i]_{i\in I(x)}\right) \ge \alpha_0,$$

where we used the definition of α_0 in the last inequality. Hence,

$$\alpha_0 \le \inf_{\substack{x\in A \\ x^*\notin N_A(x)}} d(x^*, N_A(x)) = \operatorname{sr}[A, x^*],$$

which implies that the set A is α_0-sharp w.r.t. vector x^*. ∎

3.2. Sharpness of the epigraph. Consider the following optimization problem:

$$\min_{x\in A} f(x), \tag{3.8}$$

where A is nonempty, closed, and convex and $f : \mathcal{H} \to \mathbb{R}_\infty$ is proper, convex, and lsc. Assume that (3.8) has solutions, and denote its solution set by $\mathbb{S}$. In connection with this problem, we consider the following property. Assume that there exists $\beta > 0$ such that

$$\inf_{x\notin \mathbb{S}} d(0, \partial f(x)) \ge \beta > 0. \tag{3.9}$$

We note that (3.9) holds, for instance, if f is piecewise linear. Indeed, let $f(x) := \max_{i=1,\dots,m} \langle x_i^*, x\rangle + b_i$. Call $J := \{i, \dots, m\}$. For every $x \in \mathcal{H}$, define $I(x) := \{i \in J : f(x) = \langle x_i^*, x\rangle + b_i\}$. 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 $\partial f(x) = \mathrm{co}[x_i^*]_{i\in I(x)}$. Hence, $0 \notin \partial f(x)$ if and only if $0 \notin \mathrm{co}[x_i^*]_{i\in I(x)}$. Consider

$$\mathcal{F} := \{I \subset J : 0 \notin \mathrm{co}[x_i^*]_{i\in I}\} \subset \mathcal{P}(J),$$

where $\mathcal{P}(J)$ denotes the collection of all possible subsets of J, which has cardinality 2^m. So there is a finite number of possible subdifferential sets such that $0 \notin \partial f(x)$. Moreover, $x \notin \mathbb{S}$ if and only if $I(x) \in \mathcal{F}$. Altogether,

$$\inf_{x\notin\mathbb{S}} d(0, \partial f(x)) = \inf_{0\notin\partial f(x)} d(0, \partial f(x)) \geq \min_{I\in\mathcal{F}} d(0, \mathrm{co}[x_i^*]_{i\in I}),$$

where we have a minimum in the rightmost expression because the infimum is taken over the finite set $\mathcal{F}$. Since $\mathrm{co}[x_i^*]_{i\in I}$ is a closed convex set which doesn't contain zero, $d(0, \mathrm{co}[x_i^*]_{i\in I}) > 0$ for every $I \in \mathcal{F}$. Therefore, the minimum in the right-hand side of the above expression is attained at some positive value b. Note that the fact that we have a finite supremum of affine functions is essential. The function in Example 3.6 is an infinite supremum of affine functions, for which (3.9) does not hold.

The next result characterizes the case in which the epigraph of a function is sharp w.r.t. the vector $(0_{\mathcal{H}}, -1)$. The latter property turns out to be important in the last section.

Proposition 3.15. *Let $f : \mathcal{H} \to \mathbb{R}_\infty$ be a proper, convex, and lsc function. The following statements are equivalent:*

(i) $\mathrm{epi}\, f$ *is α-sharp w.r.t. the vector $z^* =: (0_{\mathcal{H}}, -1)$.*

(ii) $\alpha < 1$ *and f verifies* (3.9) *with parameter $\beta := \frac{\alpha}{\sqrt{1-\alpha^2}}$.*

Proof. Let $E := \mathrm{epi}\, f$ and $D := \mathrm{dom}\, f$. From Fact 2.4, we have

$$N_E((x, f(x))) = \mathbb{R}_{++}(\partial f(x) \times \{-1\}) \bigcup (N_D(x) \times \{0\}). \tag{3.10}$$

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 $N_E((x, y)) = N_{\mathrm{dom}\, f}(x) \times \{0\}$, 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:

$$\begin{aligned} K &:= \{(x, f(x)) : (0, -1) \notin N_E((x, f(x))\} \\ &= \{(x, f(x)) : \partial f(x) \neq \emptyset \text{ and } 0 \notin \partial f(x), \text{ or } \partial f(x) = \emptyset\}, \end{aligned}$$

where we are using (3.10) in the characterization of K. If $(x, f(x)) \in K$ and $\partial f(x) = \emptyset$, then (3.10) gives $N_E((x, f(x)) = N_D(x) \times \{0\}$, so in this case we have

$$d\left(z^*, N_E((x, f(x)))\right) = d((0_{\mathcal{H}}, -1), N_D(x) \times \{0\}) = \inf_{w\in N_D(x)} \sqrt{\|w\|^2 + 1} = 1. \tag{3.11}$$

Fix now any $(x, f(x)) \in K$ such that $\partial f(x) \neq \emptyset$, and denote $\delta(x) := d(0, \partial f(x))$. 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

$$d\left(z^*, N_E((x, f(x)))\right) = \frac{\delta(x)}{\sqrt{\delta(x)^2 + 1}}. \tag{3.12}$$

Indeed, fix $(x, f(x)) \in K$. Using (3.10), we have

$$\begin{aligned} &d(z^*, N_E((x, f(x)))) \\ &\quad = \inf_{(u,\eta)\in N_E((x,f(x))}d(z^*,(u,\eta)) \\ &\quad = \min\left\{\inf_{w\in N_D(x)} d(z^*,(w,0)), \inf_{u\in\partial f(x),\, t>0} d(z^*,(tu,-t))\right\} \\ &\quad = \min\left\{\inf_{w\in N_D(x)} \sqrt{\|w\|^2+1}, \inf_{u\in\partial f(x),\, t>0} \sqrt{t^2(\|u\|^2+1)-2t+1}\right\} \\ &\quad = \min\left\{1, \inf_{u\in\partial f(x)}\left[\inf_{t>0}\sqrt{t^2(\|u\|^2+1)-2t+1}\right]\right\} = \min\left\{1, \inf_{u\in\partial f(x)}\frac{\|u\|}{\sqrt{\|u\|^2+1}}\right\} \\ &\quad = \frac{\delta(x)}{\sqrt{\delta(x)^2+1}}, \end{aligned}$$

where we are using (3.10) 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 we are using 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, (3.12) holds for every $(x, f(x)) \in K$.

Now, we assume that (i) holds. Then by (3.11) and (3.12) this means that, for every $(x, f(x)) \in K$, we have

$$\begin{aligned} \alpha &\le \inf_{(x,f(x))\in K} d(z^*, N_E((x, f(x)))) \\ &= \min\left\{\inf_{(x,f(x))\in K,\partial f(x)=\emptyset} d(z^*, N_E((x, f(x)))), \inf_{(x,f(x))\in K,\partial f(x)\neq\emptyset} d(z^*, N_E((x, f(x))))\right\} \\ &= \min\left\{1, \inf_{(x,f(x))\in K,\partial f(x)\neq\emptyset}\frac{\delta(x)}{\sqrt{\delta(x)^2+1}}\right\} \le \frac{\delta(y)}{\sqrt{\delta(y)^2+1}} < 1 \end{aligned}$$

for every $(y, f(y)) \in K$ such that $\partial f(y) \neq \emptyset$. Note that we are using (3.11) and (3.12) in the second equality. So $\alpha < 1$, and the above expression rewrites as

$$\delta(y) \ge \frac{\alpha}{\sqrt{1-\alpha^2}} \ \forall (y, f(y)) \in K \text{ s.t. } \partial f(y) \neq \emptyset, \tag{3.13}$$

which means that (ii) holds with parameter $\frac{\alpha}{\sqrt{1-\alpha^2}}$. 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 (3.13) holds. The latter rewrites as

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

which, together with (3.11) and (3.12), gives (i). ∎

We illustrate the results in Proposition 3.15 by two examples in Figure 3.2. Condition (3.9) is closely related with the well-known Kurdyka–Łojasiewicz inequality. To make this connection precise, we recall next the necessary definitions.

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

(b) The function $g(x) = |x|$ has the global KL property, and the set epi g is sharp w.r.t. vector (0,−1).

FIG. 3.2. *Illustration of Proposition* 3.15*:* epi f *is sharp w.r.t. 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* 3.1.

DEFINITION 3.16 (Kurdyka–Łojasiewicz inequality [4, section 2.3]). *Let* $f\colon \mathcal{H} \to \mathbb{R}_\infty$*, and assume that* $\mathbb{S} := \operatorname{argmin} f \neq \emptyset$*. Fix* $\bar{x} \in \mathbb{S}$*. The function* f *satisfies the* global Kurdyka–Łojasiewicz (KL) property *at* $\bar{x}$ *if there exists a concave continuously differentiable function* $\varphi\colon \mathbb{R}_+ \to \mathbb{R}_+$ *with* $\varphi(0) = 0$ *and* $\varphi' > 0$ *such that*

$$\varphi'(f(x) - f(\bar{x}))\, d(0, \partial f(x)) \geq 1 \quad \forall x \notin \mathbb{S}. \tag{3.14}$$

In this case, we say that φ *is a* desingularizing function *for* f *at* $\bar{x}$*. If* f *satisfies the global KL property and admits the same desingularizing function* φ *at every point* $\bar{x} \in \mathbb{S}$*, then we say that* f *satisfies the* global KL property *with* global desingularizing function φ*.*

The next result establishes the connection between the global KL property and sharpness.

COROLLARY 3.17. *Let* $f\colon \mathcal{H} \to \mathbb{R}_\infty$ *be a proper lsc convex function. The following statements are equivalent:*

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

(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 3.15 because the global KL property (ii) is equivalent to condition (ii) in Proposition 3.15. ∎

COROLLARY 3.18. *Let* $f\colon \mathcal{H} \to \mathbb{R}_\infty$ *be a polyhedral function (see Definition* 2.8*). 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* (3.9) *holds for* $\beta := \frac{\alpha}{\sqrt{1-\alpha^2}}$*.*

Proof. By Definition 2.8, the epigraph of a polyhedral function is a polyhedral set. By Proposition 3.14, it is sharp with respect to any unit vector, in particular with respect to $z^* := (0_{\mathcal{H}}, -1)$. The two claims now follow from the fact that part (i) implies (ii) in Corollary 3.17 and the fact that part (i) implies (ii) in Proposition 3.15. ∎

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; i.e., the function admits a linear global desingularizing function.

PROPOSITION 3.19. *Consider* $x^* \in \mathcal{S}$ *and a nonempty closed and convex set* A*. The following statements are equivalent:*

(i) *The modulus of sharpness of* A *w.r.t.* x^* *is* $\alpha > 0$*; equivalently,* $\alpha := sr[A, x^*] > 0$*.*

(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* (3.9) *with* $\beta := \alpha$.

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

Proof. The definition of f yields $\partial f(x) = N_A(x) - x^*$ for all $x \in \mathcal{H}$. Thus, for any $x \in A$, $x^* \notin N_A(x)$ if and only if $0 \notin \partial f(x)$. Therefore, $x \notin \operatorname{argmin} f$ if and only if $x^* \notin N_A(x)$. The equivalence above implies that

$$d(0, \partial f(x)) = d(0, N_A(x) - x^*) = d(x^*, N_A(x)), \tag{3.15}$$

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

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

The three equivalences follow directly from the expression above and the definitions. As for the last statement, assume that $\alpha < 1$. Since (i)–(iii) are all equivalent, we can use Proposition 3.15(ii) for $\beta := \alpha$. Indeed, by (iii) we have that $\delta(x) \geq \alpha$. By (3.12), 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 (3.11), we know that $d\,(z^*, N_E((x, f(x)))) = 1$ if $\partial f(x) = \emptyset$. Altogether, and using the same set K as in the proof of Proposition 3.15, we can write

$$\begin{aligned}
&\inf_{(x,f(x))\in K} d\,(z^*, N_E((x, f(x)))\\
&= \min\left\{ \inf_{\substack{(x,f(x))\in K,\\ \partial f(x)=\emptyset}} d\,(z^*, N_E((x, f(x)))\,, \inf_{\substack{(x,f(x))\in K,\\ \partial f(x)\neq\emptyset}} d\,(z^*, N_E((x, f(x))) \right\}\\
&\geq \min\left\{1, \frac{\alpha}{\sqrt{1+\alpha^2}}\right\} = \frac{\alpha}{\sqrt{1+\alpha^2}},
\end{aligned}$$

establishing the last claim. ∎

3.3. Dual characterizations. As seen in Figure 3.1, the sharpness property of the set A w.r.t. vector x^* is related to the condition that x^* belongs to the interior of the set $\bigcup_{x\in F_A(x^*)} N_A(x)$. 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 $x_0 + r\,(\mathcal{B}\setminus\mathcal{S})$ and that the corresponding closed ball is $x_0 + r\,\mathcal{B}$.

PROPOSITION 3.20. *Consider a nonempty closed convex set* A *of a Hilbert space* $\mathcal{H}$, *a vector* $x^* \in \mathcal{S}$, *and a positive constant* $\alpha \in (0, 1]$. *Consider the following statements:*

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

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

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

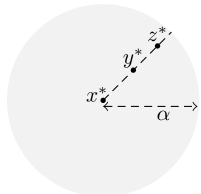


FIG. 3.3. *Here, y^* belongs to the open segment (x^*, z^*).*

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

$$d(x^*, N_A(y)) < \alpha, \tag{3.16}$$

this implies that $x^* \in N_A(y)$. This statement then implies that inequality (3.1) 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 (3.16) holds and $x^* \notin N_A(y)$. From the inequality (3.16), we can choose $y^* \in N_A(y)$ such that $0 < \|x^* - y^*\| < \alpha$. Hence, $y^* \in x^* + \alpha(\mathcal{B} \setminus \mathcal{S})$. Because $x^* + \alpha(\mathcal{B} \setminus \mathcal{S})$ is an open set, there exists $r > 0$ such that $y^* + r\mathcal{B} \subset x^* + \alpha(\mathcal{B} \setminus \mathcal{S})$. Let $t_0 := 1 + \frac{r}{\|y^* - x^*\|}$ and $z^* := x^* + t_0(y^* - x^*)$ (see Figure 3.3). Using these definitions, we can write

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

This gives $z^* = x^* + t_0(y^* - x^*) \in y^* + r\mathcal{B} \subset x^* + \alpha(\mathcal{B} \setminus \mathcal{S})$, where we used the definition of r in the last inclusion.

Now we can use assumption (ii). Namely, $z^* \in x^* + \alpha(\mathcal{B} \setminus \mathcal{S}) \subset \bigcup_{x \in F_A(x^*)} N_A(x)$. Hence, there exists $z \in F_A(x^*)$ such that $z^* \in N_A(z)$. Since $z \in F_A(x^*)$, Fact 3.4 yields $x^* \in N_A(z)$. Altogether, both of the vectors x^* and z^* belong to the normal cone $N_A(z)$. From the fact that $y \in A$ and $z^* \in N_A(z)$,

$$\langle z^*, y - z \rangle \leq 0. \tag{3.17}$$

Recall that $x^* \notin N_A(y)$, and $x^* \in N_A(z)$. Using Fact 3.4, the latter gives $z \in \operatorname{Argmax}_A \langle x^*, \cdot \rangle$ and $y \notin \operatorname{Argmax}_A \langle x^*, \cdot \rangle$. Therefore, we can write

$$\langle x^*, y \rangle < \langle x^*, z \rangle = \sup_A \langle x^*, \cdot \rangle . \tag{3.18}$$

From $y^* \in N_A(y)$ and $z \in A$, we also have $\langle y^*, z \rangle \leq \langle y^*, y \rangle$. We now use this inequality and (3.18), together with the fact that $z^* = x^* + t_0(y^* - x^*)$ and $t_0 > 1$, to derive the following estimation:

$$\begin{aligned} \langle z^*, y \rangle &= \langle t_0 y^* + (1 - t_0)x^*, y \rangle \\ &= t_0 \langle y^*, y \rangle + (1 - t_0) \langle x^*, y \rangle > t_0 \langle y^*, z \rangle + (1 - t_0) \langle x^*, z \rangle \\ &= \langle t_0 y^* + (1 - t_0)x^*, z \rangle = \langle z^*, z \rangle , \end{aligned}$$

which contradicts (3.17). Therefore, we must have $x^* \in N_A(y)$. We have shown that, whenever (3.16) 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 α-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 α-sharp w.r.t. x^*. Since the set A is bounded, we can use the characterization of sharpness established in Proposition 3.12. 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

$$\|x^* - y^*\| \geq \alpha. \tag{3.19}$$

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 3.4, 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 3.12, we deduce that (3.19) 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 3.21. In [17], a strict complementarity condition is used to ensure that the SPP solves the linear programming problem $\min_{x \in A} \langle x^*, x \rangle$, where $A \subset \mathcal{H}$ is a polyhedron, $x^* \in \mathcal{H} \setminus \{0_{\mathcal{H}}\}$, and $\mathcal{H}$ is a finite dimensional space with $\dim \mathcal{H} = n$. This is established in [17, Theorem 1], whose proof relies on the following two key conditions:

a. The minimization problem $\min_{x \in A} \langle x^*, x \rangle$ has a unique solution $\bar{x} \in A$.
b. The normal cone $N_A(\bar{x})$ has nonempty interior and $-x^* \in \operatorname{int} N_A(\bar{x})$ (see Figure 3.4(a)).

Condition 2 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 $N_A(\bar{x}) = \operatorname{cone}[\mathbb{A}^* e_i^*]_{i \in I(\bar{x})}$. Since $\operatorname{int} N_A(\bar{x}) \neq \emptyset$, 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 $\mathbb{A}^* e_{i_1}^*, \ldots, \mathbb{A}^* e_{i_n}^*$ are linearly independent. Thus, any linear programming problem in $\mathbb{R}^n$ that satisfies the strict complementarity condition must have at least n linear independent constraints. By Definition 2.8, a linear programming problem can only have a finite number of constraints, and hence the aforementioned requirement may only work for finite dimensions and will never hold in infinite dimensional case.

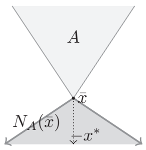
By contrast, for the sharpness condition to hold, neither condition 1 nor condition 2 above is required. Instead, Proposition 3.20 shows that sharpness simply requires $-x^*$ to belong to the interior of the union of all normal cones at every optimal solution (see Figure 3.4(b), where condition 1 does not hold). Thus, the sharpness property requires a much weaker version of condition 2. Furthermore, Proposition 3.14 proves that a polyhedron in an arbitrary Hilbert space is sharp with respect to every unit vector; and we will show later in section 4.2 that the sharpness property is enough to ensure that the minimization problem $\min_{x \in A} \langle x^*, x \rangle$ can be solved by the SPP.

We next establish the connection between the sharpness property and the subdifferential operator of the Fenchel conjugate of the indicator function at x^*. Recall that, for any proper lsc convex function f, the Fenchel–Young characterization of the subdifferential is given as follows:

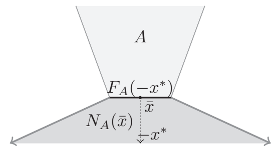
$$x^* \in \partial f(x) \Longleftrightarrow \langle x^*, x \rangle = f(x) + f^*(x^*) \Longleftrightarrow x \in \partial f^*(x^*). \tag{3.20}$$

In particular,

$$x^* \in \partial f(x) \Longrightarrow x^* \in \operatorname{dom} f^* \quad \text{and} \quad x \in \operatorname{dom} f. \tag{3.21}$$



(a) Strict complementarity condition in [17] holds: the normal cone $N_A(\bar{x})$ has nonempty interior and $-x^* \in N_A(\bar{x})$.



(b) Strict complementarity condition in [17] fails; the set $\bigcup_{x\in F_A(-x^*)} N_A(x)$ still has nonempty interior; and sharpness condition holds.

FIG. 3.4. *Comparison between sharpness condition and strict complementarity condition.*

Remark 3.22. Given a nonempty convex and closed set A, the Fenchel conjugate of $f(\cdot) := \mathbb{1}_A(\cdot)$, denoted by $\sigma_A(\cdot)$, is called the *support function of A.* Recall that $\partial f = N_A$ and $\operatorname{dom} f = A$. Applying (3.20) to this f and using the definitions, we obtain

$$f^*(v) = \sigma_A(v) = \sup_{y\in A} \langle v, y\rangle \,. \tag{3.22}$$

Using (3.20) for this f, we obtain

$$\begin{aligned} v \in N_A(x) &\Longleftrightarrow x\in A \text{ and } \langle v,x\rangle = \sigma_A(v) \Longleftrightarrow x \in \partial\sigma_A(v) \\ &\Longleftrightarrow \partial\sigma_A(v) = \operatorname{Argmax}_A \langle v,\cdot\rangle = F_A(v), \end{aligned} \tag{3.23}$$

where we also used (3.22) and Fact 3.4 in the rightmost equivalence. The equivalence above also shows that $D(\partial\sigma_A) = \mathrm{R}(N_A)$ (see Definition 2.2(a)–(b)). These facts will be used in the next result.

PROPOSITION 3.23. *Consider a nonempty closed convex set A of a Hilbert space $\mathcal{H}$. Fix $x^* \in \mathcal{S}\cap R(N_A)$ and $\alpha \in (0,1]$. Let $\sigma_A(\cdot)$ be as defined in* (3.22)*. Consider the following statements:*

(i) *The set A is α-sharp w.r.t. x^*.*
(ii) $\emptyset \neq \partial\sigma_A(v) \subset \partial\sigma_A(x^*)$ *for all* $v \in x^* + \alpha(\mathcal{B}\setminus\mathcal{S})$.

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 $v \in x^* + \alpha(\mathcal{B}\setminus\mathcal{S})$.*

Proof. For simplicity, denote $B(\alpha, x^*) := x^* + \alpha(\mathcal{B}\setminus\mathcal{S})$. The last statement will follow directly from (ii) and the equivalence (3.23) for $v := x^*$. Indeed, the rightmost expression in (3.23) with $v := x^*$ gives $F_A(x^*) = \partial\sigma_A(x^*)$. Assuming that (ii) holds, we will use Proposition 3.20 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 fact that part (ii) implies (i) in Proposition 3.20. Indeed, taking $v \in B(\alpha,x^*)$, 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 $x \in \partial\sigma_A(v) \Longrightarrow x \in \partial\sigma_A(x^*) = F_A(x^*)$. On the other hand, by (3.23) we have

$$x \in \partial\sigma_A(v) \Longleftrightarrow x \in \operatorname{Argmax}_{y\in A} \langle v,y\rangle \Longleftrightarrow v \in N_A(x).$$

Combining the rightmost parts of the two expressions above gives the following:

$$\text{If } v \in B(\alpha,x^*) \quad\text{and}\quad x\in\partial\sigma_A(v), \quad\text{then}\quad x\in F_A(x^*) \quad\text{and}\quad v\in N_A(x).$$

Therefore, for every $v \in B(\alpha,x^*)$, there exists $x \in F_A(x^*)$ s.t. $v \in N_A(x)$. In other words, $B(\alpha,x^*) \subset \bigcup_{x\in F_A(x^*)} N_A(x)$. Proposition 3.20 (part (ii) implies (i)) now implies that A is α-sharp w.r.t. x^*. Assuming now that A is bounded, we will prove that (i) implies (ii). Namely, we will use the boundedness and sharpness of A and Proposition 3.20 to show that (ii) holds. Fix $v \in B(\alpha, x^*)$. 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 (3.23) 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 3.4, the latter means that $v \in N_A(y)$ and $x^* \notin N_A(y)$. Since (i) holds, we can use Proposition 3.12 to write $\|x^* - v\| \geq d(x^*, N_A(y)) \geq \alpha$, which contradicts the fact that $v \in B(\alpha, x^*)$. Therefore, we must have that $F_A(v) \subset F_A(x^*)$, and hence (ii) holds. ∎

Remark 3.24. We say that a proper lsc convex function $f : \mathcal{H} \rightrightarrows \mathbb{R}_\infty$ is *quasi-polyhedral at* $\bar{x} \in \operatorname{dom} f$ (see [8]) if there exists $r > 0$ such that $\partial f(x) \subset \partial f(\bar{x})$ 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 exist $r > 0$ and a sublinear function $p : \mathcal{H} \to \mathbb{R}_\infty$ such that $f(x) = f(\bar{x}) + p(x - \bar{x})$ for every $x \in \bar{x} + r\mathcal{B}$. Hence, Proposition 3.23 shows that A is sharp w.r.t. x^* if and only if the function σ_A is quasi-polyhedral at x^*; and if σ_A is continuous at x^*, then σ_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 x^* is equivalent to the subtransversality property (also known as metric subregularity; see [14, Definition 3.1] and [13]) between the set A and its supporting hyperplane w.r.t. x^*. 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 3.25. Note that the subtransversality property for the pair $\{A, B\}$ is equivalent to the property

$$\alpha d(x, A \cap B) \leq d(x, A) \quad \forall x \in B. \tag{3.24}$$

Indeed, this follows from the fact that $d(x, A) \leq \max\{d(x, A), d(x, B)\}$ when $x \notin B$. The connection between subtransversality and sharpness arises when we specialize (3.24) for the pair $\{A, F\}$ with $F := \{x \in \mathcal{H} : \langle x^*, x\rangle = \sup_{y \in A} \langle x^*, y\rangle\}$. In this situation, using the definition of $F_A(x^*)$, (3.24) 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 3.26. *Consider a nonempty closed convex set A of a Hilbert space $\mathcal{H}$, a vector $x^* \in \mathcal{S}$ such that $F_A(x^*) \neq \emptyset$, and $\alpha \in (0,1)$. Define the set $F := \{x \in \mathcal{H} : \langle x^*, x\rangle = \sup_A \langle x^*, \cdot\rangle\}$. Assume that*

$$\alpha\, d(x, F_A(x^*)) \leq d(x, A) \quad \forall x \in F, \tag{3.25}$$

and define $\gamma := \alpha\sqrt{1 - \frac{1}{4}\alpha^2}$. Then, the set A is γ-sharp w.r.t. vector x^*. Conversely, define $\beta := 2\alpha/(1-\alpha)$. If A is β-sharp w.r.t. x^*, then inequality (3.25) *holds.*

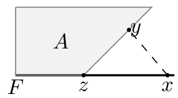


FIG. 3.5. *There is $x \in F$ such that $y \in P_A(x)$ and $y - x \in \operatorname{cone}[y^*]$.*

Proof. We first assume that inequality (3.25) holds, and we now show that the set A is γ-sharp w.r.t. x^*, where $\gamma := \alpha\sqrt{1 - \frac{1}{4}\alpha^2}$. Note that if the set $A \setminus F_A(x^*)$ is empty, then $F_A(x^*) = A$. In this case, inequality (3.25) holds trivially for any $\gamma \in (0, 1)$ in place of α. So it is enough to assume that $A \setminus F_A(x^*)$ is not empty. By Fact 3.4, in this case there is $y \in A$ such that $x^* \notin N_A(y)$. Suppose that $N_A(y) \neq \{0\}$ since otherwise it always holds that $d(x^*, N_A(y)) = \|x^*\| = 1 \geq \alpha$ for all $\alpha \in (0, 1)$. So we can take $y^* \in \mathcal{S} \cap N_A(y)$, i.e., $\|y^*\| = 1$. We consider two cases.

Case 1. Suppose that $\langle x^*, y^* \rangle \leq 0$. Then, by Fact 2.11, $d(x^*, \operatorname{cone}[y^*]) = 1 \geq \alpha$.

Case 2. Suppose that $\langle x^*, y^* \rangle > 0$. We first show that there is an $x \in F$ such that $x - y \in \operatorname{cone}[y^*]$ (see Figure 3.5), 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 claim above is equivalent to stating that the equation $\langle x^*, y + ty^* \rangle = \sup_A \langle x^*, \cdot \rangle$ has a solution $t \geq 0$. Because $\langle x^*, y^* \rangle > 0$, and $y \in A$, the constant $t := \frac{\sup_{w \in A} \langle x^*, w \rangle - \langle x^*, y \rangle}{\langle x^*, y^* \rangle}$ is nonnegative and satisfies $\langle x^*, y + ty^* \rangle = \sup_{w \in A} \langle x^*, w \rangle$. Therefore, $x := y + ty^* \in F$, and $y = P_A(x)$. Hence, $x - y \in \operatorname{cone}[y^*] \subset N_A(y)$. Taking into account that $\|y^*\| = 1$, we have $y^* = \frac{x-y}{\|x-y\|}$.

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

$$\langle x - y, z - y \rangle \leq 0. \tag{3.26}$$

Additionally, inequality (3.25) implies that $\alpha \|x - z\| = \alpha d(x, F_A(x^*)) = \alpha d(x, A \cap F) \leq d(x, A) = \|x - y\|$. Combine the expression above with (3.26) to obtain

$$\begin{aligned} \|x - z\|^2 &= \|x - y\|^2 + \|y - z\|^2 + 2\langle x - y, y - z \rangle \geq \|x - y\|^2 + \|y - z\|^2 \\ &\geq \alpha^2 \|x - z\|^2 + \|y - z\|^2. \end{aligned}$$

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

$$\begin{aligned} \langle x - y, x - z \rangle &= \tfrac{1}{2}\left(\|y - x\|^2 + \|x - z\|^2 - \|y - z\|^2\right) \\ &\geq \tfrac{1}{2}\left(\|y - x\|^2 + \|x - z\|^2 - (1 - \alpha^2)\|x - z\|^2\right) \\ &= \tfrac{1}{2}\left(\alpha^2 \|x - z\|^2 + \|x - y\|^2\right). \end{aligned}$$

From $y^* = \frac{x-y}{\|x-y\|}$, the inequality above, and the Cauchy–Schwarz inequality, we obtain

$$\left\langle y^*, \frac{x - z}{\|x - z\|} \right\rangle = \left\langle \frac{x - y}{\|x - y\|}, \frac{x - z}{\|x - z\|} \right\rangle \geq \frac{\alpha^2}{2} \frac{\|x - z\|}{\|x - y\|} + \frac{1}{2} \frac{\|x - y\|}{\|x - z\|} \geq \alpha,$$

where the last inequality follows from the fact that $\eta(t) := \frac{\alpha^2 t}{2} + \frac{1}{2t}$ over $\mathbb{R}_{++}$ attains a minimum at $t^* := 1/\alpha$ and its minimum value is $\eta(t^*) = \alpha$. On the other hand, since both $x, z \in F$, by the definition of F we must have $\left\langle x^*, \frac{x-z}{\|x-z\|} \right\rangle = 0$. Combining the latter equality with the last inequality yields $\left\langle y^* - x^*, \frac{x-z}{\|x-z\|} \right\rangle \geq \alpha$, which, by the definition of dual norm, implies that $\|y^* - x^*\| \geq \alpha$. The latter inequality and Fact 2.11 give

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

where we also used the fact that $\|x^*\| = \|y^*\| = 1$. Altogether, in Cases 1 and 2, we always have $d(x^*, \text{cone}[y^*]) \geq \alpha\sqrt{1 - \frac{\alpha^2}{4}}$, and because $y^* \in \mathcal{S} \cap N_A(y)$ is chosen arbitrarily, we have

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

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

$$\alpha d(\bar{x}, F_A(x^*)) > \hat{\alpha} d(\bar{x}, F_A(x^*)) > d(\bar{x}, A). \tag{3.27}$$

Consider the closed half-space $F' := \{x \in \mathcal{H} : \langle x^*, x \rangle \geq \sup_{y \in A} \langle x^*, y \rangle\}$. Observe that $F = \text{bd}\, F'$, 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 $\bar{x} \in F$, inequality (3.27) implies that $\delta > 0$ and $\bar{x} \notin A$. Let $\bar{y} := P_A(\bar{x})$, so $d(\bar{x}, A) = \|\bar{x} - \bar{y}\| > 0$. With this notation, inequality (3.27) becomes

$$\alpha\delta > \hat{\alpha}\delta > \|\bar{x} - \bar{y}\|. \tag{3.28}$$

We claim that (3.28) implies that $\bar{y} \notin F_A(x^*)$. Indeed, if $\bar{y} \in F_A(x^*)$, then the definition of δ 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 $\widetilde{y} := P_{F_A(x^*)}(\bar{y})$. By the triangle inequality,

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

where we used the definition of $\widetilde{y}$ in the first equality, the fact that $\widetilde{y} \in F_A(x^*)$ in the second inequality, and (3.28) in the last inequality. Set $\hat{\delta} := (1 - \alpha)\delta$. Using the above expression and the definitions of F' and $\hat{\delta}$, we derive

$$d(\bar{y}, F_A(x^*)) = d(\bar{y}, A \cap F') > \hat{\delta}. \tag{3.29}$$

With this notation, $\hat{\alpha}\delta = \frac{\hat{\alpha}\hat{\delta}}{(1-\alpha)}$ and the second inequality in (3.28) rewrites as

$$(\hat{\alpha}\hat{\delta})/(1 - \alpha) > \|\bar{x} - \bar{y}\|. \tag{3.30}$$

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

$$f(x, y) := \|x - y\| + \mathbb{1}_{F'}(x) + \mathbb{1}_A(y), \quad (x, y) \in \mathcal{H}^2. \tag{3.31}$$

We now use (3.30) and the Ekeland variational principle, Lemma 2.10, for $X := \mathcal{H}^2$ equipped with the max norm $\|(x, y)\| := \max\{\|x\|, \|y\|\}$, and $\psi := f$. As mentioned in (2.1), 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 (3.30) and the definition of f, gives $\inf_{\mathcal{H}\times\mathcal{H}} f(x,y) \geq 0 > f(\bar{x},\bar{y}) - \frac{\hat{\alpha}}{1-\alpha}\hat{\delta}$. Hence, we are in conditions of Ekeland's variational principle with $\varepsilon := \frac{\hat{\alpha}\hat{\delta}}{(1-\alpha)}$ and $\bar{w} := (\bar{x},\bar{y})$. We apply the principle for the choice $\lambda := \hat{\delta}$, for which there exists $(\hat{x},\hat{y})$ such that the following hold:

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

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

Moreover, our choices imply that $\varepsilon/\lambda = \hat{\alpha}/(1-\alpha)$. Altogether, condition (iii) in Lemma 2.10 with the max norm implies that for every $\mathcal{H}^2 \ni (x,y) \neq (\hat{x},\hat{y})$ we have

$$f(\hat{x},\hat{y}) < f(x,y) + \frac{\hat{\alpha}}{(1-\alpha)} \max\{\|x-\hat{x}\|, \|y-\hat{y}\|\} =: h(x,y). \tag{3.32}$$

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

(I) By (3.29), we have that $d(\bar{y}, A\cap F') > \hat{\delta}$. Using also (i) and the fact that $d(\bar{x}, A\cap F') = \delta > \hat{\delta}$, we have $\hat{x},\hat{y} \notin A\cap F'$. Indeed, if we would have $\hat{x} \in A\cap F'$, then this fact would imply that $\hat{\delta} < d(\bar{x}, A\cap F') \leq \|\bar{x}-\hat{x}\|$, contradicting (i). A similar expression, mutatis mutandis, can be used to show that $\hat{y} \notin A\cap F'$.

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

(III) By (II) and Remark 2.5, we write the subdifferential of f as the sum of the subdifferentials at $(\hat{x},\hat{y})$,

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

where ∂_1 and ∂_2 stand for the partial subdifferentials w.r.t. the first and second variables, respectively. Each of the partial subdifferentials exists due to the continuity of the first term in (3.31).

(IV) By (3.32), it is clear that $(\hat{x},\hat{y})$ is the global minimizer of h.

By (IV), we have that $0 \in \partial h(\hat{x},\hat{y})$. To compute the subdifferential of h, we note that the second term in (3.32) 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 (3.32). To write down the inclusion $0 \in \partial h(\hat{x},\hat{y})$, we use Remarks 2.5 and 2.6 to write $0 \in \left(\frac{\hat{x}-\hat{y}}{\|\hat{x}-\hat{y}\|} + N_{F'}(\hat{x}), \frac{\hat{y}-\hat{x}}{\|\hat{x}-\hat{y}\|} + N_A(\hat{y})\right) + \frac{\hat{\alpha}}{(1-\alpha)}\mathbb{B}^*_{\mathcal{H}^2}$, where the ball in the rightmost term is the one induced by the sum norm as given in (2.2). The inclusion above and the definition of the dual ball give

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

where we are using the definition of the sum ball (see (2.2)) in the first equality. Note also that $\hat{x} \in F' = \{x \in \mathcal{H} : \langle x^*, x\rangle \geq \sup_{y\in A}\langle x^*, y\rangle\}$, so $N_{F'}(\hat{x}) = \operatorname{cone}[-x^*]$ if $\hat{x} \in F = \operatorname{bd} F'$, and $N_{F'}(\hat{x}) = \{0\}$ otherwise. This gives $N_{F'}(\hat{x}) \subset \operatorname{cone}[-x^*]$. Therefore,

$$\begin{aligned} d\left(\frac{\hat{x}-\hat{y}}{\|\hat{x}-\hat{y}\|}, N_A(\hat{y})\right) + d\left(\frac{\hat{x}-\hat{y}}{\|\hat{x}-\hat{y}\|}, \operatorname{cone}[x^*]\right) &\leq d\left(\frac{\hat{x}-\hat{y}}{\|\hat{x}-\hat{y}\|}, N_A(\hat{y})\right) + d\left(\frac{\hat{y}-\hat{x}}{\|\hat{x}-\hat{y}\|}, N_{F'}(\hat{x})\right) \\ &\leq \hat{\alpha}/(1-\alpha). \end{aligned} \tag{3.33}$$

To simplify notation, call $\omega_0 := \frac{\hat{x}-\hat{y}}{\|\hat{x}-\hat{y}\|}$. By the triangle inequality, we have

$$d(x^*, N_A(\hat{y})) \leq \|x^* - \omega_0\| + d\left(\omega_0, N_A(\hat{y})\right). \tag{3.34}$$

We claim that

$$d\left(\omega_0, \operatorname{cone}[x^*]\right) \geq \frac{1}{2}\left\|x^* - \omega_0\right\|. \tag{3.35}$$

Indeed, from Fact 2.11, we have $d\left(\omega_0, \operatorname{cone}[x^*]\right)^2 = 1 - \max\left(0, \langle x^*, \omega_0\rangle\right)^2$. We consider two cases:

1. If $\langle x^*, \omega_0\rangle \leq 0$, then $d\left(\omega_0, \operatorname{cone}[x^*]\right) = 1$, and by the triangle inequality, $\|\omega_0 - x^*\| \leq \|\omega_0\| + \|x^*\| = 2 = 2\, d\left(\omega_0, \operatorname{cone}[x^*]\right)$, as claimed in (3.35).
2. If $\langle x^*, \omega_0\rangle > 0$, then

$$\begin{aligned} d\left(\omega_0, \operatorname{cone}[x^*]\right)^2 &= 1 - \langle x^*, \omega_0\rangle^2 = 1/2\left(1 + \langle x^*, \omega_0\rangle\right)\left(2 - 2\langle x^*, \omega_0\rangle\right) \\ &> 1/2\left(2 - 2\langle x^*, \omega_0\rangle\right) = 1/2\left(\|x^*\|^2 + \|\omega_0\|^2 - 2\langle x^*, \omega_0\rangle\right) \\ &= 1/2\left\|\omega_0 - x^*\right\|^2, \end{aligned}$$

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 $d\left(\omega_0, \operatorname{cone}[x^*]\right) > \frac{1}{\sqrt{2}}\left\|\omega_0 - x^*\right\| > \frac{1}{2}\left\|\omega_0 - x^*\right\|$. In both cases, we proved that (3.35) holds. From (3.34), we have

$$\begin{aligned} d(x^*, N_A(\hat{y})) &\leq \|x^* - \omega_0\| + d\left(\omega_0, N_A(\hat{y})\right) \leq 2d\left(\omega_0, \operatorname{cone}[x^*]\right) + d\left(\omega_0, N_A(\hat{y})\right) \\ &= \left[d\left(\omega_0, \operatorname{cone}[x^*]\right) + d\left(\omega_0, N_A(\hat{y})\right)\right] + d\left(\omega_0, \operatorname{cone}[x^*]\right) \\ &\leq \hat{\alpha}/(1-\alpha) + \hat{\alpha}/(1-\alpha) = 2\hat{\alpha}/(1-\alpha) < 2\alpha/(1-\alpha), \end{aligned} \tag{3.36}$$

where we used (3.34) in the first inequality and (3.35) in the second one. For the last inequality, we used (3.33) for the expression between square brackets, and for the remaining term we used the fact that disregarding the first term in (3.33) implies that $d\left(\omega_0, \operatorname{cone}[x^*]\right) \leq \hat{\alpha}/(1-\alpha)$. From (I), we have $\hat{y} \notin A \cap F' = A \cap F = F_A(x^*)$, and so by Fact 3.4 it holds that $x^* \notin N_A(\hat{y})$. By the β-sharpness of A w.r.t. x^* and the assumption on β, we deduce that $d(x^*, N_A(\hat{y})) \geq 2\alpha/(1-\alpha) = \beta$, contradicting (3.36). This implies that inequality (3.25) holds. ∎

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 3.27 (see [16, Theorem 3.1], Subtransversality). *Suppose X is a normed linear space, $A, B \subset X$ are nonempty 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*

$$\alpha d(x, A \cap B) \leq d(x, A) \quad \forall x \in B \setminus A. \tag{3.37}$$

Theorem 3.26 and Lemma 3.27 yield the following result.

Corollary 3.28. *Consider a nonempty closed convex set A of a Hilbert space $\mathcal{H}$, and a vector $x^* \in \mathcal{S}$. Then, A is sharp w.r.t. x^* if and only if the pair $\{A, F\}$ is subtransversal, where $F := \{x \in \mathcal{H} : \langle x^*, x\rangle = \sup_A \langle x^*, \cdot\rangle\}$.*

Proof. Suppose $\{A, F\}$ is subtransversal. Then, from Lemma 3.27, we have that (3.37) holds for some $\alpha \in (0,1)$. By Theorem 3.26, the set A is α'-sharp w.r.t. x^*, where $\alpha' := \alpha\sqrt{1 - 1/4\alpha^2} \in (0,1)$. Conversely, assume that A is α-sharp w.r.t. x^* for some $\alpha \in (0,1)$. Then, by Theorem 3.26, we have $\alpha' d(x, A \cap F) \leq d(x, A)$, for all $x \in F$, with $\alpha' := \frac{\alpha}{2+\alpha} \in (0,1)$. Note that

$$2\alpha'/(1-\alpha') = 2\frac{\alpha/(2+\alpha)}{1-\alpha/(2+\alpha)} = 2\frac{\alpha}{2+\alpha-\alpha} = \alpha.$$

Because $\alpha' \in (0,1)$, and from Lemma 3.27, we conclude that the pair $\{A, F\}$ is subtransversal. ∎

4. Optimization problems under sharpness condition. We consider now constrained convex problems of the following type:

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

where $f : \mathcal{H} \to \mathbb{R}_\infty$ is a proper lsc convex function and A is a nonempty closed convex set. We provide in this section sufficient conditions under which the SPP can solve problem (CP). As expected, the sharpness condition 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 a 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 ρ and center b and by $B(\rho, b)$ 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, $d(b, A) \le \|a-b\| - \varepsilon$ if there is $\delta > 0$ such that*

$$\text{(4.1)} \quad \inf\{d(x^*, N_A(x)) : \|x^*\| = 1,\ \langle x^*, b-x\rangle = \|b-x\|,\ x \in B[\rho, b] \cap B(\delta, a) \cap A\} \ge \varepsilon/\delta.$$

Proof. Assume (4.1) holds for some $\varepsilon, \delta > 0$. For contradictory purposes, assume also that ε is such that $d(b, A) > \|a-b\| - \varepsilon$. Consider the function $f : X \to \mathbb{R}_\infty$ defined by $f(y) := \|y-b\| + \mathbb{1}_A(y) = \varphi_b(y) + \mathbb{1}_A(y)$. Here, $\varphi_b(y) := \|y-b\|$ as in Remark 2.5. Then, the assumption on ε 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 2.10) applied to $\bar{w} := a$, $\psi := f$, $\varepsilon := \varepsilon'$, and $\lambda := \delta$, there exists a vector $\hat{x} \in A \cap B_\delta(a)$ such that

$$\text{(4.2)} \qquad f(\hat{x}) \le f(a),$$

$$\text{(4.3)} \qquad f(\hat{x}) \le f(y) + \varepsilon'/\delta\,\|y-\hat{x}\| \quad \forall y \in A.$$

Due to (4.2) and $a \in A$, we have that $\hat{x} \in A$ and $\|\hat{x}-b\| \le \|a-b\| = \rho$, or $\hat{x} \in A \cap B[\rho, b]$. Define $h(y) := \frac{\varepsilon'}{\delta}\|y-\hat{x}\|$. By (4.3), 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 2.5, $\partial h(\hat{x}) = \frac{\varepsilon'}{\delta}\mathbb{B}^*$, and we obtain $0 \in \partial\varphi_b(\hat{x}) + \partial\mathbb{1}_A(\hat{x}) + \partial h(\hat{x}) = \partial\varphi_b(\hat{x}) + N_A(\hat{x}) + \frac{\varepsilon'}{\delta}\mathbb{B}^*$. Therefore,

$$\text{(4.4)} \qquad [\partial\varphi_b(\hat{x}) + N_A(\hat{x})] \cap (\tfrac{\varepsilon'}{\delta}\mathbb{B}^*) \ne \emptyset.$$

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 $\partial\varphi_b(\hat{x}) = \{x^* : \langle x^*, \hat{x} - b\rangle = \|\hat{x} - b\|, \|x^*\| = 1\}$. Altogether, by (4.4) 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 (4.4), there exists $u \in N_A(\hat{x})$ such that $(u - x^*) \in \frac{\varepsilon'}{\delta}\mathbb{B}^*$, and the last one holds because $\varepsilon' < \varepsilon$. Noting that (4.1) holds, and using the fact that $x^* \in (-\partial\varphi_b(\hat{x}))$, i.e., $\|x^*\| = 1, \langle x^*, b - \hat{x}\rangle = \|b - \hat{x}\|$, and $\hat{x} \in B[\rho, b] \cap B(\delta, a) \cap A$, we can write

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

which is a contradiction. ∎

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, $a \in A$, $b \notin A$ with $\rho := \|a - b\|$, and $\varepsilon > 0$. If there is $\delta > 0$ such that the following inequality holds,*

$$\inf\left\{d\left(\frac{b-x}{\|b-x\|}, N_A(x)\right) : x \in B[\rho, b] \cap B(\delta, a) \cap A\right\} \geq \varepsilon/\delta, \tag{4.5}$$

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 (4.1) becomes (4.5). ∎

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, $d\big(\frac{b-a}{\|b-a\|}, N_A(a)\big) = 0$. Hence, to ensure we are far from the latter situation, we require (4.5) 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 (4.5) 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 ε. Note also that the opposite inequality to (4.5) is the optimality condition for ε-projections, where the later define points in A that are within distance $d(b, A) + \varepsilon$ from b. Geometrically, (4.5) 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 4.1 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 ε-projection of b onto A can be understood as an element $a' \in A$ such that $d(b, A) > \|a' - b\| - \varepsilon$. This motivates the following generalization of Theorem 4.1 to the case of an ε-projection.

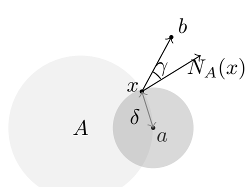


FIG. 4.1. *The distance from b to the set A is bounded above by $\|b - a\| - \delta\alpha$ (so $\varepsilon = \delta\alpha$). Here, γ is the smallest angle between vectors $x - b$ and normal cone $N_A(x)$ for $x \in A$, $\alpha = \cos(\gamma)$, and δ is the size of the neighborhood around a; i.e., we are taking $x \in A \cap B_\delta(a)$. Note here that we only consider $x \in A \cap B_\delta(a)$ such that $\|x - b\| \leq \|b - a\|$.*

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 ε-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 $d(x_0^*, N_A(x_0)) < \varepsilon/\delta$.

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 the definition of $C(\delta)$, is exactly (4.1). By Theorem 4.1, we must have $\|a - b\| \geq d(b, A) + \varepsilon$, contradicting the fact that a is an ε-projection of b. ∎

4.2. Solving problems with SPP: The case with a linear objective. We start this section by considering the following minimization problem with a linear objective:

$$\text{(P)} \qquad \min_{x \in A} \langle x^*, x\rangle ,$$

where $x^* \in \mathcal{S}$ and A is a closed convex set. Denote by $\mathbb{S}$ the set of solutions of problem (P). We assume that $\mathbb{S} \neq \emptyset$. The optimality conditions for problem (P) imply that $\mathbb{S} = \{x : 0 \in x^* + N_A(x)\}$. The following theorem shows that problem (P) can be solved by projecting an infeasible point onto the feasible region A if the set A is sharp. We will use the following fact, which is a consequence of Fact 2.11. For any p, q nonzero vectors, we have

$$d(p, \text{cone}[q]) = d(q, \text{cone}[p]). \tag{4.6}$$

THEOREM 4.5. *Consider a nonempty closed convex set A of a Hilbert space $\mathcal{H}$ and a vector $x^* \in \mathcal{S}$. Suppose that A is α-sharp w.r.t. $-x^*$ for some $\alpha \in (0, 1]$. Suppose also that $v \in \mathcal{H}$ satisfies following conditions:*
1. *$\langle x^*, v\rangle < \inf_{x\in A} \langle x^*, x\rangle$;*
2. *$(1 - (\alpha/2)^2)d(v, A) < \inf_{x\in A} \langle x^*, x - v\rangle$.*

Then, the projection of v onto A is a solution of problem (P).

Proof. Suppose to the contrary that there exists $v \in \mathcal{H}$ satisfying conditions 1 and 2 such that the projection of v onto A is not a solution of (P). Since A is α-sharp w.r.t. $-x^*$, we have

$$\inf_{x\in A, -x^* \notin N_A(x)} d(-x^*, N_A(x)) \geq \alpha. \tag{4.7}$$

Take $y = P_A(v)$; then $v - y \in N_A(y)$. Because y is not a solution of the convex problem (P), we must have $0 \notin x^* + N_A(y)$. From (4.7) and the fact that $-x^* \notin N_A(y)$, we have $d(-x^*, N_A(y)) \geq \alpha > 0$. Combining this inequality with the inclusion $v - y \in N_A(y)$ yields

$$\begin{aligned} d\left(\frac{y - v}{\|y - v\|}, \text{cone}[x^*]\right) &= d(x^*, \text{cone}[y - v]) = d(-x^*, \text{cone}[v - y]) \\ &\geq d(-x^*, N_A(y)) \geq \alpha, \end{aligned} \tag{4.8}$$

where we also used (4.6) in the first equality. Consider the closed half space $F := \{x \in \mathcal{H} : \langle x^*, x - v\rangle \leq 0\}$. From condition 1, $\langle x^*, v\rangle < \inf_{x\in A} \langle x^*, x\rangle$, and 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 (4.5) 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 $\|v - y\|$. Setting $\delta := (\alpha/2)\|v - y\|$, and $\rho := \|v - y\|$, we claim that

$$\inf\left\{ d\left(\tfrac{y-z}{\|y-z\|}, N_F(z)\right) : z \in F \cap B(\delta, v) \cap B[\rho, y]\right\} \geq \alpha/2. \tag{4.9}$$

Indeed, take $z \in F \cap B(\delta, v) \cap B[\rho, y]$. We consider two cases.

Case 1. If $z \in \operatorname{int} F$, then $N_F(z) = \{0\}$, and hence $d\left(\tfrac{y-z}{\|y-z\|}, N_F(z)\right) = 1$.

Case 2. If $z \notin \operatorname{int} F$, then the definition of F implies that $N_F(z) = \operatorname{cone}[x^*]$. Since $z \in B[\rho, y]$, we use (4.8) to write

$$\begin{aligned}
d\left(\frac{y-z}{\|y-z\|}, N_F(z)\right) &= d\left(\frac{y-z}{\|y-z\|}, \operatorname{cone}[x^*]\right)\\
&= \inf_{t\geq 0}\left\|\frac{y-z}{\|y-z\|} - tx^*\right\| = \frac{\|y-v\|}{\|y-z\|} d\left(\frac{y-z}{\|y-v\|}, \operatorname{cone}[x^*]\right)\\
&\geq d\left(\frac{y-z}{\|y-v\|}, \operatorname{cone}[x^*]\right) = d\left(\frac{y-v}{\|y-v\|} + \frac{v-z}{\|y-v\|}, \operatorname{cone}[x^*]\right)\\
&= \inf_{t\geq 0}\left\|\frac{y-v}{\|y-v\|} + \frac{v-z}{\|y-v\|} - tx^*\right\| \geq \inf_{t\geq 0}\left\|\frac{y-v}{\|y-v\|} - tx^*\right\| - \frac{\|v-z\|}{\|y-v\|}\\
&= d\left(\frac{y-v}{\|y-v\|}, \operatorname{cone}[x^*]\right) - \frac{\|v-z\|}{\|y-v\|} \geq \alpha - \alpha/2 = \alpha/2,
\end{aligned}$$

where in the fourth equality we used the change of variables $t \to \tilde{t} := t\|y-z\|/\|y-v\|$, and we used the fact that $\|y - z\| \leq \rho = \|y - v\|$ in the first inequality. As for the last inequality, we use (4.8) 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 (4.9) holds. By Theorem 4.1, we have

$$d(y, F) \leq \|y - v\| - (\alpha/2)\delta = \|y - v\| - (\alpha/2)^2\|y - v\| = (1 - (\alpha/2)^2)\|y - v\|. \tag{4.10}$$

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 the definition of F, we know that $N_F(w) = \operatorname{cone}[x^*]$. 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 $\langle x^*, w - v\rangle \leq 0$, 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 $\langle x^*, y - v\rangle = \langle x^*, y - w\rangle + \langle x^*, w - v\rangle = \langle x^*, y - w\rangle = d(y, F)$. So, the following estimation holds:

$$\inf_{x\in A} \langle x^*, x - v\rangle \leq \langle x^*, y - v\rangle = d(y, F) \leq (1 - (\alpha/2)^2)\|y - v\| = (1 - (\alpha/2)^2) d(v, A),$$

where we also used (4.10) 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 α-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 4.2). 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.

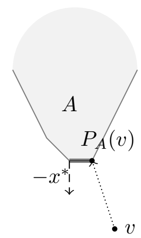


FIG. 4.2. *Illustration of Theorem* 4.5*: the set A is sharp w.r.t vector* $-x^*$*. Vector v satisfies conditions* 1 *and* 2*, and hence the projection* $P_A(v)$ *of v onto the set A is the solution of the minimization problem* $\inf_A \langle x^*, \cdot\rangle$.

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* α *does not hold for* v*. Define*

$$\theta(v) := \inf_{x\in A} \langle x^*, x - v\rangle, \quad \textit{and } \mu_0 := \frac{(1 - (\alpha/2)^2)d(v, A) - \theta(v)}{(\alpha/2)^2}. \tag{4.11}$$

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*.*

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 α. Altogether, we have that

$$0 < \theta(v) \leq (1 - (\alpha/2)^2)d(v, A). \tag{4.12}$$

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

$$\langle x^*, x - u\rangle \geq \theta(v) + \mu > 0 \tag{4.13}$$

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

$$d(u, A) = d(v - \mu x^*, A) = \inf_{x\in A} \|v - \mu x^* - x\| \leq \inf_{x\in A} \|v - x\| + \mu\|x^*\| = d(v, A) + \mu. \tag{4.14}$$

Using the definition of μ_0, we rewrite the inequality $\mu > \mu_0$ as

$$d(v, A) < \frac{(\alpha/2)^2\mu + \theta(v)}{(1 - (\alpha/2)^2)}. \tag{4.15}$$

Using (4.15) in (4.14) 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 (4.13) 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. *Consider a nonempty closed convex set A of a Hilbert space $\mathcal{H}$ and vector $x^* \in \mathcal{S}$ such that the set A is α-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 μ_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.

Example 4.8. Consider problem (P) with $\mathcal{H} = \mathbb{R}^2$, $x^* = (0,1)$, and $A := \{x \in \mathbb{R}^2 : \mathbb{A}x \leq b\}$, where $\mathbb{A} = \begin{bmatrix} 1 & -1 \\ -1 & -1 \end{bmatrix}$ and $b = (0,0)$. Our first task is to determine the modulus of sharpness of A. It is easy to check that, for every $x \in \mathrm{bd}A$, we have $N_A(x) = \mathrm{cone}[(-1,-1)] =: K_1$ if $x_1 < 0$, and $N_A(x) = \mathrm{cone}[(1,-1)] =: K_2$ if $x_1 > 0$, and $N_A(x) = \mathrm{cone}[(-1,-1),(1,-1)]$ if $x_1 = 0$, which can be graphically verified from Figure 4.3. Note that we have $-x^* \notin N_A(x)$ if and only if $x_1 \neq 0$. Using (4.7) in Theorem 4.5, we have

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

Hence, (4.7) 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 μ_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 μ 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 4.3.

Example 4.9. Consider problem (P) with $\mathcal{H} = \mathbb{R}^3$, $x^* = (0, 1/\sqrt{2}, 1/\sqrt{2})$, and $A := \{x \in \mathbb{R}^2 : \mathbb{A}x \leq b\}$, where $\mathbb{A} = -I$ and $b = (0,0,0)$. Namely, $A = \mathbb{R}^3_+$. The solution of (P) is the set $\mathbb{R}_+ \times \{0\} \times \{0\}$, 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 $x \in \mathrm{bd}A$ 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: $T :=$

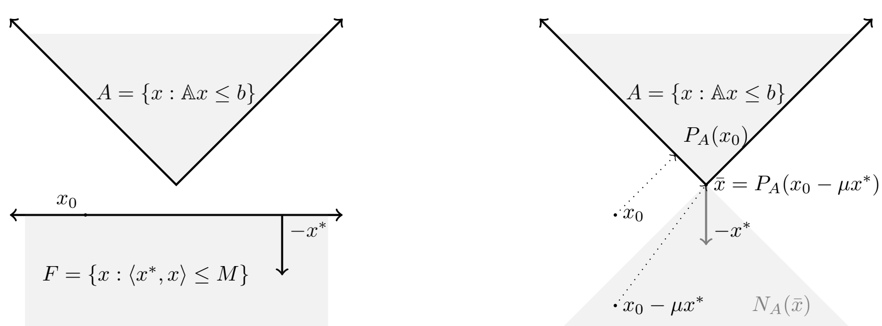


FIG. 4.3. *Illustration for Example* 4.8*: (left) the disjoint sets A and F and (right) the projection of the point $x_0 - \mu x^*$ onto A is the unique solution, $\bar{x}$, of the LP.*

$\{x \in \operatorname{bd} A : x_1 \geq 0$ and x_2, x_3 are not simultaneously zero$\}$. It can be checked that $\inf_{x \in T} d(-x^*, N_A(x)) = 1/\sqrt{2}$, so (4.7) 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 μ 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).

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

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

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 $\operatorname{dom} f \cap A$. The latter is true, for instance, when some standard constraint qualification holds (see [2, Corollary 16.38]), e.g., when $\operatorname{int} A \cap \operatorname{dom} f \neq \emptyset$ or $\operatorname{int}(\operatorname{dom} f) \cap A \neq \emptyset$. The function $f_A := (f + \mathbb{1}_A)$ will have a crucial role in the next result. Note that

$$\operatorname{epi} f_A = \{(x, t) \in A \times \mathbb{R} : f(x) \leq t\} = (A \times \mathbb{R}) \cap \operatorname{epi} f. \tag{4.16}$$

By imposing a sharpness condition on the set $\operatorname{epi} f_A$ w.r.t. the vector $(0_{\mathcal{H}}, -1)$, 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 the convex problem* (CP) *has solutions with optimal value* $M := \inf_{x \in A} f(x)$ *and a nonempty set of optimal solutions denoted by* $\mathbb{S}$. *Assume that the following conditions hold:*

- (i) *The set* $\operatorname{epi} f_A =: \tilde{A}$ *given in* (4.16) *is* α*-sharp w.r.t* $z^* := (0_{\mathcal{H}}, -1)$ *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 $P_{\tilde{A}}(v, t) = (w, f(w))$. *Then,* $w \in \mathbb{S}$, *and hence* $f(w) = M$.

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

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

which has the same optimal value as (CP) and a linear objective $\psi : \mathcal{H} \times \mathbb{R} \to \mathbb{R}$ defined as $\psi(x, s) := s = \langle (0_{\mathcal{H}}, 1), (x, s) \rangle$. Note that (EP) is an optimization problem with a linear objective and by (i), its constraint set is 2α sharp w.r.t. $z^* := (0_{\mathcal{H}}, -1)$. Take $\tilde{v} := (v, t) \in \mathcal{H} \times \mathbb{R}$ verifying assumptions (a)–(b). We claim that this implies that conditions 1 and 2 in Theorem 4.5 hold, with $x^* := (0_{\mathcal{H}}, 1)$, and $A := \tilde{A}$. Indeed, condition (a) rewrites 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 $x^* := (0_{\mathcal{H}}, 1)$. Condition 2 in Theorem 4.5 follows directly from (b) and the definitions. Therefore, we are in conditions of Theorem 4.5, and $P_{\tilde{A}}(v, t)$ solves (EP). By [2, Proposition 29.35], $P_{\tilde{A}}(v, t) = (w, f(w))$, where $w \in A$ is the unique solution of the inclusion $\dfrac{v - w}{(f(w) - t)} \in \partial f(w)$. Note that $t < M \leq f(w)$ so $(f(w) - t) > 0$. Since $P_{\tilde{A}}(v, t) = (w, f(w))$ solves (EP), this means that $f(w) = M$ and since $w \in A$ we must have $w \in \mathbb{S}$. ∎

Remark 4.11. We give in Proposition 3.15 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 3.18, we know that there exists $\alpha < 1$ such that property (3.9) holds for $\beta := \alpha/\sqrt{1-\alpha^2}$. The latter fact, combined with Proposition 3.15, imply that $\operatorname{epi} f_A =: \tilde{A}$ is α-sharp w.r.t. $(0_{\mathcal{H}}, -1)$. Hence, we are in the situation of Theorem 4.10. Therefore, if (v,t) is such that $t < M$ and $\left(1-\alpha^2/4\right) d((v,t),\tilde{A}) < M - t$, then $P_{\tilde{A}}(v,t) = (\bar{x}, f(\bar{x}))$ is such that $\bar{x}$ solves (CP).

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 a solution, and denote by* $\mathbb{S}$ *the set of optimal solutions, and* $\alpha \in (0,1)$. *Assume that the following conditions hold:*

(i) *With the convention* $\inf \emptyset = +\infty$ *and* $\frac{0}{\|0\|} = 0$,

$$\inf_{x \in A \setminus \mathbb{S}, x^* \in \partial f(y), y \in \mathbb{S}} d\left(-x^*/\left\|x^*\right\|, N_A(x)\right) \geq \alpha > 0. \tag{4.17}$$

(ii) *There is* $v \in \mathcal{H}$ *such that the following hold:*
   (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 $\bar{x} \in \mathbb{S}$, 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 (4.17) equals 0. Letting $M := \inf_A f$, we consider the sublevel set of f at value M: $F := \{x \in \mathcal{H} : f(x) \leq M\}$. Then, it is clear that $A \cap F = \mathbb{S}$, the set of optimal solutions of problem (CP). From $x^* \in (-N_A(\bar{x})) \cap \partial f(\bar{x})$, we have

$$\langle x^*, x - \bar{x}\rangle \geq 0 \, \forall x \in A \quad \text{and} \quad 0 \geq f(x') - M = f(x') - f(\bar{x}) \geq \langle x^*, x' - \bar{x}\rangle \ \forall x' \in F, \tag{4.18}$$

where the first inequality in the rightmost expression follows because $x' \in F$ and $\bar{x} \in \mathbb{S}$. In other words, x^* separates the (closed and convex) sets A and F. Therefore, $A \cap F = \mathbb{S} \subset \{x : \langle x^*, x\rangle = \langle x^*, \bar{x}\rangle\}$. In particular, this implies that $\langle x^*, z - \bar{x}\rangle = 0$ for all $z \in \mathbb{S}$. Take now any $y \in A$, $z \in \mathbb{S}$. Use the left-hand side of (4.18) and $\langle x^*, z - \bar{x}\rangle = 0$ to deduce that $\langle -x^*, y - z\rangle = \langle -x^*, y - \bar{x}\rangle + \langle -x^*, \bar{x} - z\rangle \leq 0$, and so $-x^* \in N_A(z)$, for all $z \in \mathbb{S}$. The inclusion above implies that $\langle x^*, z\rangle \leq \langle x^*, y\rangle$ for any $y \in A$, $z \in \mathbb{S}$. In other words, $\mathbb{S} \subset \operatorname{argmin}_A \langle x^*, \cdot\rangle$. Namely,

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

and therefore $A \setminus \mathbb{S} \supset A \setminus F_A(-x^*)$. Combining the latter inclusion with inequality (4.17) yields

$$\inf_{x \in A \setminus F_A(-x^*)} d\left(-x^*/\left\|x^*\right\|, N_A(x)\right) \geq \alpha. \tag{4.19}$$

The expression above means that A is α-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 (4.18) yields $\langle x^*, v\rangle \leq$

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

$$\langle x^*, \bar{x}\rangle - \langle x^*, v\rangle \ge f(\bar{x}) - f(v) = M - f(v) > 0, \tag{4.20}$$

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

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

where we used (4.20) in the second inequality and the fact that $\bar{x} \in \mathbb{S} \subset \operatorname{argmin}_A \langle x^*, \cdot\rangle$ 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 $\mathbb{S} = F_A(-x^*)$. This will establish that $P_A(v) \in \mathbb{S}$, as wanted. We already know that $\mathbb{S} \subset F_A(-x^*)$, so it is enough to show that $\mathbb{S} \supset F_A(-x^*)$. Indeed, assume that there is $z \in A \setminus \mathbb{S}$ 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, $-x^* \in N_A(z)$. The latter inclusion gives $0 = d(-x^*/\|x^*\|, N_A(z)) \ge \inf_{\substack{x\in A\setminus\mathbb{S} \\ x^*\in\partial f(y), y\in\mathbb{S}}} d(-x^*/\|x^*\|, N_A(x)) \ge \alpha > 0$, where we are using (4.17) and the fact that $z \in A \setminus \mathbb{S}$ and $x^* \in \partial f(\bar{x})$ with $\bar{x} \in \mathbb{S}$ in the first inequality. The above expression entails a contradiction, and hence we must have $\mathbb{S} = F_A(-x^*)$. Therefore, our claim is true and $P_A(v) \in F_A(-x^*) = \mathbb{S}$, and thus $P_A(v)$ solves problem (CP). ∎

5. Conclusions and open questions.

In this work, we introduce the notion 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 $\hat{x}^\star \in \mathcal{H}$ with $\|\hat{x}^\star - x^*\|$ small enough so that $\hat{x}^\star \in \operatorname{int} \bigcup_{x\in F_A(x^*)} N_A(x)$. This allows inexact projections, reducing the computational effort. This observation is not trivial for nonlinear functions. However, it is worth noticing that the 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 piecewise 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 3.12, for any polyhedral set, there is $\alpha > 0$ such that it is α-sharp w.r.t. every unit vector. It is interesting to ask whether the converse statement is true, namely: if a set is α-sharp w.r.t. every unit vector for some $\alpha > 0$, then is it true that the set must be polyhedral?

REFERENCES

[1] H. H. Bauschke, *The approximation of fixed points of compositions of nonexpansive mappings in Hilbert space*, J. Math. Anal. Appl., 202 (1996), pp. 150–159.

[2] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, 2nd ed., CMS Books Math./Ouvrages Math. SMC, Springer, Cham, 2017.

[3] R. Behling, Y. Bello-Cruz, and L.-R. Santos, *Infeasibility and error bound imply finite convergence of alternating projections*, SIAM J. Optim., 31 (2021), pp. 2863–2892, https://doi.org/10.1137/20M1358669.

[4] J. Bolte, T. P. Nguyen, J. Peypouquet, and B. W. Suter, *From error bounds to the complexity of first-order descent methods for convex functions*, Math. Program., 165 (2017), pp. 471–507.

[5] H. Brezis, *Functional Analysis, Sobolev Spaces and Partial Differential Equations*, Springer, New York, 2011.

[6] H. T. Bui, R. Loxton, and A. Moeini, *A note on the finite convergence of alternating projections*, Oper. Res. Lett., 49 (2021), pp. 431–438.

[7] R. S. Burachik and A. N. Iusem, *Set-Valued Mappings and Enlargements of Monotone Operators*, Springer Optim. Appl. 8, Springer, New York, 2008.

[8] A. K. Chakrabarty, P. Shunmugaraj, and C. Zalinescu, *Continuity properties for the subdifferential and ε-subdifferential of a convex function and its conjugate*, J. Convex Anal., 14 (2007), pp. 479–514.

[9] D. Drusvyatskiy, A. D. Ioffe, and A. S. Lewis, *Transversality and alternating projections for nonconvex sets*, Found. Comput. Math., 15 (2015), pp. 637–1651.

[10] I. Ekeland, *On the variational principle*, J. Math. Anal. Appl., 47 (1974), pp. 353–342.

[11] N. Hermer, D. R. Luke, and A. Sturm, *Random function iterations for consistent stochastic feasibility*, Numer. Funct. Anal. Optim., 40 (2019), pp. 386–420.

[12] V. Jeyakumar, *Generalizations Farkas' lemma*, in Encyclopedia of Optimization, Springer, Berlin, 2001, pp. 627–631.

[13] A. Y. Kruger, R. D. Luke, and N. H. Thao, *About subtransversality of collections of sets*, Set-Valued Var. Anal., 25 (2017), pp. 701–729.

[14] G. Li and T. K. Pong, *Calculus of the exponent of Kurdyka-Łojasiewicz inequality and its applications to linear convergence of first-order methods*, Found. Comput. Math., 18 (2018), pp. 1199–1232.

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

[16] K. F. Ng and W. H. Yang, *Regularities and their relations to error bounds*, Math. Program., 99 (2004), pp. 521–538.

[17] E. Nurminski, *Single-projection procedure for linear optimization*, J. Global Optim., 66 (2016), pp. 95–110.

[18] W. Oliveira, C. Sagastizábal, and C. Lemaréchal, *Convex proximal bundle methods in depth: A unified analysis for inexact oracles*, Math. Program. Ser. B, 148 (2014), pp. 241–277.

[19] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*, Springer, New York, 2009.

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

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