Subgradient Selector in the Generalized Cutting Plane Method with an Application to Sparse Optimization

Seta Rakotomandimby, Jean-Philippe Chancelier, Michel de Lara, Adrien Le Franc

Subgradient Selector in the Generalized Cutting Plane Method with an Application to Sparse Optimization

Seta Rakotomandimby* Jean-Philippe Chancelier* Michel De Lara*
Adrien Le Franc*

December 27, 2024

Abstract

Duality in convex analysis devotes a prominent role to affine functions, as proper convex lower semicontinuous functions are supremum of such functions. This property is used in the Kelley algorithm, to minimize a proper convex lower semicontinuous function by sequentially approximating it from below by maxima of affine functions (cuts). Affine functions are deduced from a bilinear pairing. In generalized convexity, the usual bilinear form is replaced by some bivariate function c, called coupling. The Moreau-Rockafellar subdifferential of a function is replaced by the c-subdifferential. The Kelley algorithm then becomes the generalized c-cutting plane method to minimize a c-subdifferentiable objective function. In this paper, we prove a convergence result whose scope makes it possible to tackle sparse optimization problems. For this purpose, we introduce a selection of c-subgradients involved in a pointwise locally equicontinuous property, together with the coupling c and the objective function. Under the assumptions of the convergence result, we discuss a necessary condition on the continuity points of the function to be minimized. Finally, we give an example of converging Capra-cutting plane method for the minimization of the pseudonorm ℓ_0 on a compact set.

Keywords cutting plane method; generalized convexity; c-subdifferential; pseudonorm ℓ_0; subgradient selector

1 Introduction

Duality in convex analysis devotes a prominent role to affine functions, as proper convex lower semicontinuous functions are supremum of affine functions. This property is used in the Kelley algorithm, to minimize a proper convex lower semicontinuous function by sequentially approximating it from below by maxima of affine functions (cuts). This is the

*CERMICS, École nationale des ponts et chaussées, IP Paris, France

spirit of so-called cutting plane methods, which sequentially minimize and update maxima of base functions (generalized cuts).

Elementary base functions are the building block of so-called *abstract convexity* where the equivalent of closed convex functions are the suprema of such functions. By contrast, in *generalized convexity*, the focus is put on replacing the usual bilinear form (of duality in convex analysis) by some bivariate function c, that is called coupling, and the Moreau-Rockafellar subdifferential of a function by the c-subdifferential. Abstract and generalized convexity are two (related) ways to extend duality beyond convex analysis. As such, they provide the mathematical framework to extend cutting plane methods beyond the convex case.

The cutting plane method is a staple optimization scheme in integer linear programming [14, 5]. In 1958, Gomory [6] introduced one of the first cutting plane methods to solve integer linear programs. In 1960, Kelley [7] proposed a cutting plane method to minimize convex functions (not necessarily differentiable) over a compact set. Pallaschke and Rolewicz [11, Theorem 9.1.1] generalized Kelley's result to the minimization of so-called Φ-convex functions, where elementary base functions are continuous functions. Rubinov gave two convergence results [13, Propositions 9.2, 9.3] in the abstract convex setting. Each of these convergence results relies on properties which relate elementary base functions (generalized cuts), the objective function and all generalized subgradients. In the usual convex finite dimensional setting — and also for so-called one-sided-linear couplings, as introduced in [2, § 2.2] — all of these assumptions boil down to boundedness of the generalized subdifferential of the objective function.

Now, it has been established in [2] that the ℓ_0 pseudonorm — which counts the number of nonzero entries of a vector — has nonempty generalized subdifferential, for a suitable choice of elementary base functions, induced by a so-called Euclidean Capra-coupling. Unfortunately, the assumptions of results [11, 13] are not straightforwardly satisfied by the Capra coupling and the ℓ_0 pseudonorm, as the Capra-subdifferentials of ℓ_0 have the property of being unbounded. This observation has motivated us to extend the scope of the results in [11, 13]. In particular, we emphasize the choice of suitable subsets of subgradients in the statement of a generalized cutting plane method. With this, it is possible to tackle sparse optimization problems, consisting in minimizing ℓ_0 over a compact set. More precisely, as easy-to-compute formulas for the Capra-subdifferential of ℓ_0 are given in [8], we can design a converging Capra-cutting plane method with the corresponding Capra-cuts.

Contributions. We introduce the notion of dual selector and, in Theorem 2, we propose a convergence result of the generalized cutting plane method for c-subdifferentiable functions that satisfy a pointwise locally equicontinuous property. We discuss the link between this property and the continuity points of the objective function in Proposition 3. Finally, we present a converging Capra-cutting plane method for the minimization of ℓ_0 over a compact subset of the unit sphere in Proposition 8.

Notations and basic definitions. For any couple of integers $(i,j) \in \mathbb{N}^2$ such that $i \leq j$, we denote $[\![i,j]\!] = \{i, i+1, \ldots, j-1, j\}$. The extended real line is $\overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty, -\infty\}$, and we set $\mathbb{R}_{++} =]0, +\infty[x$.

Let $(\mathcal{X}, d)$ be a metric space. We denote by $B(x,\eta) = \{x' \in \mathcal{X} \mid d(x,x') \leq \eta\}$ the *ball centered at x of radius η*. We say that a function $h\colon \mathcal{X} \to \overline{\mathbb{R}}$ is *lower semicontinuous* (lsc) at $x \in \mathcal{X}$ if, for all $\{x^i\}_{i \geq 0} \subset \mathcal{X}$, we have that

$$\lim_{i \to +\infty} x^i = x \implies \liminf_{i \to +\infty} h(x^i) \geq h(x) \ . \tag{1}$$

We say that the function h is lsc if it is lsc at x, for all $x \in \mathcal{X}$.

Notions in generalized convexity [11, 15, 13, 9]. We remind notions of generalized convexity with couplings. Let us consider two nonempty sets $\mathcal{X}$ (primal), $\mathcal{Y}$ (dual) and a finite coupling $c\colon \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$. Let $h\colon \mathcal{X} \to \overline{\mathbb{R}}$ be a (primal) function. The *c-Fenchel-Moreau conjugate* of h is the function $h^c\colon \mathcal{Y} \to \overline{\mathbb{R}}$ defined by

$$h^c(y) = \sup_{x \in \mathcal{X}} \big(c(x,y) - h(x) \big) \ , \ \ \forall y \in \mathcal{Y} \ . \tag{2a}$$

We also recall the *c-subdifferential* of h, which is the set-valued mapping $\partial_c h\colon \mathcal{X} \rightrightarrows \mathcal{Y}$ defined, for any $x \in \mathcal{X}$, by

$$y \in \partial_c h(x) \iff c(x',y) - h(x') \leq c(x,y) - h(x) \ , \ \ \forall x' \in \mathcal{X} \ , \tag{2b}$$

or, equivalently, by

$$h^c(y) = c(x,y) - h(x) \ . \tag{2c}$$

We have the property:

$$-\infty < h(x) < +\infty \text{ and } \partial_c h(x) \neq \emptyset \implies h(x) = \sup_{y \in \mathcal{Y}} \big(c(x,y) - h^c(y) \big) \ . \tag{2d}$$

For any subset $X \subset \mathcal{X}$, we say that h is *c-subdifferentiable* on X if its c-subdifferential $\partial_c h\colon \mathcal{X} \rightrightarrows \mathcal{Y}$ satisfies

$$\partial_c h(x) \neq \emptyset \ , \ \ \forall x \in X \ . \tag{2e}$$

Outline. The paper is organized as follows. In Sect. 2, we give a convergence result for the generalized cutting plane method and we provide a necessary condition on the discontinuity points of the objective function. In Sect. 3, we provide an example of converging Capra-cutting plane method for sparse minimization problems. In Appendix A, we relegate a technical result and a comparison table of our main result with those in [11, 13].

2 Generalized cutting plane method with dual selector

In §2.1, we emphasize the notion of dual selector and then present a new convergence result to solve $\min_X h$, where X is a compact constraint set and h is a function which is c-subdifferentiable on X. In §2.2, we provide a necessary condition on the (sequential) continuity of the function h on the constraint set X, under a pointwise locally equicontinuous property.

2.1 Convergence theorem for c-subdifferentiable functions with c-dual selectors

As discussed in the introduction, we present another generalization of Kelley's convergence result that can be applied to the minimization of finite c-subdifferentiable functions over compact sets. Compared to previous convergence results [11, 13], we explicitly link the forthcoming pointwise locally equicontinuous property (5) with what we call a c-dual selector. This addition will prove relevant to obtain a convergence result for the Capra-cutting plane method to minimize ℓ_0 over a compact set (Sect. 3).

Definition 1 *We consider a nonempty primal set $\mathcal{X}$, a nonempty dual set $\mathcal{Y}$, a finite coupling $c\colon \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ and a finite function $h\colon \mathcal{X} \to \mathbb{R}$. Let $X \subset \mathcal{X}$ be a nonempty subset.*

Suppose that the function $h\colon \mathcal{X} \to \mathbb{R}$ is c-subdifferentiable on X, as defined in (2e). *We say that a set-valued mapping $Y : X \rightrightarrows \mathcal{Y}$ is a c-*dual selector *of $\partial_c h$ on X if*

$$\emptyset \subsetneq Y(x) \subset \partial_c h(x) \ , \ \ \forall x \in X \ . \tag{3a}$$

*We say that a mapping $D : X \to \mathcal{Y}$ is a c-*subgradient selector *of $\partial_c h$ on X if*

$$D(x) \in \partial_c h(x) \ , \ \ \forall x \in X \ . \tag{3b}$$

Having defined notions of selectors, we now state a new convergence result for the generalized cutting plane method.

Theorem 2 *Let $(\mathcal{X}, d)$ be a nonempty metric space, $\mathcal{Y}$ be a nonempty set and $c\colon \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ be a finite coupling such that the function $c(\cdot, y)\colon \mathcal{X} \to \mathbb{R}$ is lsc, for any $y \in \mathcal{Y}$. Let $X \subset \mathcal{X}$ be a nonempty compact subset. Let $h\colon \mathcal{X} \to \mathbb{R}$ be a c-subdifferentiable function on X, as defined in* (2e).

We consider the minimization problem

$$\min_{x \in X} h(x) \ . \tag{4}$$

*Let $Y : X \rightrightarrows \mathcal{Y}$ be a c-*dual selector *of $\partial_c h$ on X, as in Definition 1. Let us assume that the couple $\big(c, Y\big)$ satisfies the following* pointwise locally equicontinuous property *on X:*

$$\forall \varepsilon > 0 \ , \ \ \forall x \in X \ , \ \ \exists \eta > 0 \ , \ \ \forall x' \in B(x, \eta) \cap X \ , \ \ \sup_{y \in Y(B(x,\eta) \cap X)} |c(x', y) - c(x, y)| \le \varepsilon \ , \tag{5}$$

where $Y\big(B(x,\eta)\cap X\big) = \bigcup_{x''\in B(x,\eta)\cap X} Y(x'')$.

Then, for all $x^0 \in X$*, there exist a sequence* $\{x^i\}_{i\geq 0} \subset X$ *and a sequence* $\{z^i\}_{i\geq 1} \subset \mathbb{R}$ *such that, for all* $i \geq 1$*,*

$$(x^i, z^i) \in \mathop{\arg\min}_{(x,z)\in X\times\mathbb{R}} \Big\{ z \in \mathbb{R} : z \geq c\big(x, y^j\big) - c\big(x^j, y^j\big) + h(x^j) \ , \ \ \forall j \leq i-1 \Big\} \ , \tag{6}$$

where $y^i \in Y(x^i)$ *(arbitrarily chosen) for all* $i \geq 0$*, and these sequences satisfy*

- $\{z^i\}_{i\geq 1}$ *increases to* $h^* = \inf_X h$*;*
- $\{x^i\}_{i\geq 0}$ *has a subsequence* $\{x^{\nu(i)}\}_{i\geq 0}$ *converging to an optimal solution of* (4)*, that is, to some* $x^* \in \arg\min_X h$*.*

We denote by $\mathrm{CP}(h, X, c; Y, x^0)$ the set of sequences $\big(\{x^i\}_{i\geq 0}, \{z^i\}_{i\geq 1}\big)$ satisfying (6), for all $i \geq 1$ (these sequences are not unique as the arg min (6) is not necessarily a singleton, and as y^i can be arbitrarily chosen in $Y(x^i)$, for all $i \geq 0$). We say that such sequences $\big(\{x^i\}_{i\geq 0}, \{z^i\}_{i\geq 1}\big)$ are *generated by the cutting plane method* $\mathrm{CP}(h, X, c; Y, x^0)$. It is worth noticing that Theorem 2 gives a convergence result independently of the starting point $x^0 \in \mathcal{X}$ and of the selection of x^i in the arg min (6) (and of the selection of y^i in $Y(x^i)$).

We now give the proof of Theorem 2.

Proof. First, we prove that the function $h + \iota_X$ is lsc, and the existence of an optimal $\bar{x} \in \arg\min_X h$.

As the function $h\colon \mathcal{X} \to \overline{\mathbb{R}}$ is c-subdifferentiable on X, and according to (2d), we have that $h(x) = \sup_{y\in\mathcal{Y}} \big(c(x,y) - h^c(y)\big)$, for all $x \in X$. Thus[1], we can write

$$h + \iota_X = h \dot{+} \iota_X = \sup_{y\in\mathcal{Y}} \big(c(\cdot, y) \dot{+} \big(-h^c(y)\big) \dot{+} \iota_X(\cdot) \big) \ .$$

Under the supremum, the functions $c(\cdot, y)$ are lsc by assumption, the function $\iota_X(\cdot)$ is lsc because X is a closed set, and constant functions $-h^c(y)$ are lsc (even for the values $\pm\infty$). Their lower addition is lsc by [10, p. 22]. Thus, the function $h + \iota_X$ is lsc, as a supremum of lsc functions by [1, Lemma 1.26].

Then, we prove the existence of an optimal $\bar{x} \in \arg\min_X h$. As the set X is nonempty compact and as the function $h + \iota_X$ is lsc, we get that $h^* = \inf_X h = \inf_X (h + \iota_X)$ is attained at some $\bar{x} \in X$, according to [1, Theorem 1.29] as metric spaces are Hausdorff spaces.

Second, let $x^0 \in X$. We show the existence of sequences $\{x^i\}_{i\geq 0}$ and $\{z^i\}_{i\geq 1}$ defined by (6). At each step $i \geq 1$, we define a ("c-polyhedral") function $g^i\colon \mathcal{X} \to \overline{\mathbb{R}}$ by

$$g^i(x) = \max_{j\in[\![0,i-1]\!]} \big\{ c(x, y^j) - c(x^j, y^j) + h(x^j) \big\} \ , \ \ \forall x \in \mathcal{X} \ , \tag{7a}$$

[1] We use Moreau's lower addition extending the usual addition by the formulas $+\infty \dot{+} -\infty = -\infty \dot{+} +\infty = -\infty$, and ι_X denotes the indicator function that takes the value 0 on X and $+\infty$ outside.

where $y^j \in Y(x^j)$ (arbitrarily chosen) for $j \in [\![0, i-1]\!]$. It is easy to see that the problem (6) is equivalent to the minimization problem $\min_{x \in X} g^i(x)$, with the correspondence given by (7b) below.

We are going to show that the arg min in (6) is never empty. For this purpose, we show that the function g^i is lsc. Under the maximum in (7a), the functions $c(\cdot, y)$ are lsc by assumption and constant functions $-c(x^j, y^j) + h(x^j)$ are lsc (here constant finite values). Their addition is lsc by [10, p. 22]. Thus, following [1, Lemma 1.26], we conclude that the function g^i is lsc. Then, as the subset X is compact, we apply again [1, Theorem 1.29], so that the sequences $\{x^i\}_{i \geq 0}$ and $\{z^i\}_{i \geq 1}$ exist and satisfy

$$z^i = g^i(x^i) = \min_{x \in X} g^i(x) \in \mathbb{R} \ , \ \ \forall i \geq 1 \ . \tag{7b}$$

Third, let the sequence $\{\alpha^i\}_{i \geq 0} \subset \mathbb{R}$ be defined by $\alpha^i = \min_{j \in [\![0,i]\!]} h(x^j)$, for all $i \geq 0$. We are going to show that $\{z^i\}_{i \geq 1}$ is nondecreasing with limit value $\overline{z} \in \mathbb{R}$, that $\{\alpha^i\}_{i \geq 0}$ is nonincreasing with limit value $\underline{\alpha} \in \mathbb{R}$, and that

$$z^i \leq \overline{z} \leq h^* \leq \underline{\alpha} \leq \alpha^i \ , \ \ \forall i \geq 1 \ . \tag{7c}$$

On the one hand, we have that $g^i \leq g^{i+1}, \forall i \geq 1$ by definition (7a). Thus, according to the equalities in (7b), the sequence $\{z^i\}_{i \geq 1}$ is nondecreasing. In addition, we have that

$$\overbrace{z^i = g^i(x^i)}^{(7b)} \underbrace{\leq g^i(\bar{x})}_{\text{as } x^i \text{ is min. of } g^i} \overbrace{= \max_{j \in [\![0,i-1]\!]} \left\{ c(\bar{x}, y^j) - c(x^j, y^j) + h(x^j) \right\}}^{(7a)} ,$$

where $$c(\bar{x}, y^j) - h(\bar{x}) \leq c(x^j, y^j) - h(x^j) \ , \ \ \forall j \in [\![0, i-1]\!]$$

by definition of the c-subdifferential (2b) as $y^j \in Y(x^j) \subset \partial_c h(x^j)$ by (3a),

$$\iff c(\bar{x}, y^j) - c(x^j, y^j) + h(x^j) \leq h(\bar{x}) \ , \ \ \forall j \in [\![0, i-1]\!] \ ,$$

(as all four quantities are finite)

$$\iff g^i(\bar{x}) \leq h(\bar{x}) = h^* \qquad \text{(by definition (7a) of } g^i \text{, and by } h^* = \inf_X h = h(\bar{x}))$$

Thus, we have gotten that $z^i \leq h^*$, for all $i \geq 1$. On the other hand, $\alpha^i = \min_{j \in [\![0,i]\!]} h(x^j) \geq \min_{j \in [\![0,i+1]\!]} h(x^j) = \alpha^{i+1}$, for all $i \geq 0$, hence the sequence $\{\alpha^i\}_{i \geq 0}$ is nonincreasing. Furthermore, as $x^j \in X$, for all $j \geq 0$, we get that $h^* = \min_X h \leq \min_{j \in [\![0,i]\!]} h(x^j) = \alpha^i$, for any $i \geq 1$. Thus, we have obtained that $z^i \leq h^* \leq \alpha^i$, where the sequence $\{z^i\}_{i \geq 1} \subset \mathbb{R}$ is nondecreasing and the sequence $\{\alpha^i\}_{i \geq 0} \subset \mathbb{R}$ is nonincreasing, and both are made of finite real numbers. As a consequence, there exist $\overline{z} \in \mathbb{R} \cup \{+\infty\}$ and $\underline{\alpha} \in \mathbb{R} \cup \{-\infty\}$ such that $z^i \uparrow \overline{z}$ and $\alpha^i \downarrow \underline{\alpha}$. Then, we get that $\overline{z} \leq h^* \leq \underline{\alpha}$, as well as $\overline{z} \in \mathbb{R}$ and $\underline{\alpha} \in \mathbb{R}$. We have thus proved (7c).

Fourth, we show that $\overline{z} = \underline{\alpha} = h^*$. As a first step, we prove that

$$j < i \implies c(x^j, y^j) - c(x^i, y^j) \geq \underline{\alpha} - \overline{z} \ . \tag{7d}$$

Indeed, for $(i,j) \in \mathbb{N}^2$ such that $j < i$, we have that

$$\begin{aligned}
& z^i \geq c(x^i, y^j) - c(x^j, y^j) + h(x^j) \;, \\
& \qquad\qquad \text{(according to definition (6) of the sequences } \{x^i\}_{i\geq 0} \text{ and } \{z^i\}_{i\geq 1}) \\
\implies & \overline{z} \geq c(x^i, y^j) - c(x^j, y^j) + \underline{\alpha} \;, \\
& \qquad\qquad \text{(as } \overline{z} \geq z^i \text{, for all } i \geq 1 \text{ and } h(x^j) \geq \min_{j \in [\![0,i]\!]} h(x^j) = \alpha^j \geq \underline{\alpha}) \\
\implies & c(x^j, y^j) - c(x^i, y^j) \geq \underline{\alpha} - \overline{z} \;. \qquad \text{(as the coupling } c \text{ is finite as well as } \underline{\alpha} \text{ and } \overline{z})
\end{aligned}$$

From (7c), we get that $\underline{\alpha} - \overline{z} \geq 0$. We are going to prove that $\underline{\alpha} - \overline{z} = 0$ by contradiction. For this purpose, let us assume that $\underline{\alpha} - \overline{z} > 0$. Let $\{x^{\nu(i)}\}_{i\geq 0} \subset X$ be a subsequence of $\{x^i\}_{i\geq 0}$ converging to some $x \in X$, where $\nu \colon \mathbb{N} \to \mathbb{N}$ is increasing (such converging subsequence exists since the set X is compact in the metric space $\mathcal{X}$). According to the pointwise locally equicontinuous property (5) of the couple $\big(c, Y\big)$ — written at the above $x \in X$ and for $\varepsilon = \frac{\underline{\alpha} - \overline{z}}{4} > 0$ — we have that

$$\exists \eta > 0, \forall x' \in B(x,\eta) \cap X \;, \;\; \forall y \in Y\big(B(x,\eta) \cap X\big) \;, \;\; |c(x', y) - c(x, y)| \leq \frac{\underline{\alpha} - \overline{z}}{4} \;. \tag{7e}$$

As $\lim_{k\to\infty} x^{\nu(k)} = x$, we select a pair $j < i$ such that $x^{\nu(j)} \in B(x,\eta)$ and $x^{\nu(i)} \in B(x,\eta)$. Then, we get that

$$0 < \underline{\alpha} - \overline{z} \leq c(x^{\nu(j)}, y^{\nu(j)}) - c(x^{\nu(i)}, y^{\nu(j)})$$

by (7d), as $\nu(j) < \nu(i)$ follows from $j < i$ since $\nu \colon \mathbb{N} \to \mathbb{N}$ is increasing,

$$\begin{aligned}
& \leq |c(x^{\nu(j)}, y^{\nu(j)}) - c(x, y^{\nu(j)})| + |c(x^{\nu(i)}, y^{\nu(j)}) - c(x, y^{\nu(j)})| \\
& \qquad\qquad \text{(by the triangular inequality)} \\
& \leq \begin{cases} \frac{\underline{\alpha} - \overline{z}}{4} & \text{using (7e) with } x' = x^{\nu(j)} \in B(x,\eta) \cap X \text{ and} \\ & \qquad y = y^{\nu(j)} \in Y(x^{\nu(j)}) \subset Y\big(B(x,\eta) \cap X\big) \\ + & \\ \frac{\underline{\alpha} - \overline{z}}{4} & \text{using (7e) with } x' = x^{\nu(i)} \in B(x,\eta) \cap X \text{ and} \\ & \qquad y = y^{\nu(j)} \in Y(x^{\nu(j)}) \subset Y\big(B(x,\eta) \cap X\big) \end{cases} \\
& \leq \frac{\underline{\alpha} - \overline{z}}{2} \;.
\end{aligned}$$

We obtain that $0 < \underline{\alpha} - \overline{z} \leq \frac{\underline{\alpha} - \overline{z}}{2}$, which is impossible. Thus, we have shown by contradiction that $\underline{\alpha} - \overline{z} = 0$. From (7c), we conclude that $\overline{z} = \underline{\alpha} = h^*$.

Lastly, let $\sigma \colon \mathbb{N} \to \mathbb{N}$ be a mapping such that $h(x^{\sigma(i)}) = \min_{j \in [\![0,i]\!]} h(x^j) = \alpha^i$, for any $i \geq 0$. As $\alpha^i \downarrow \underline{\alpha}$, we get that $\lim_{i\to\infty} h(x^{\sigma(i)}) = \underline{\alpha} = h^*$. We can assume that the sequence $\{x^{\sigma(i)}\}_{i\geq 0}$ converges to some $x^* \in X$ (otherwise we consider a converging subsequence as X is compact). By lower semicontinuity of h, we have that $h(x^*) \leq \lim_{i\to\infty} h(x^{\sigma(i)}) = h^*$. As $h^* = \inf_X h$, we get that $h(x^*) = h^*$, hence that $x^* \in \arg\min_X h$. □

2.2 Necessary continuity of the objective function on the set of constraints

To apply Theorem 2 to solve minimization problems of the form $\min_X h$, we are going to show that the constraint set X cannot be chosen independently of the (possible) discontinuity points of the function h. We discuss this point in Proposition 3. More precisely, we identify a necessary condition on the (sequential) continuity of the function h on the constraint set X, under the pointwise locally equicontinuous property (5).

We remind the notion of (sequential) continuity for finite functions. Let $(\mathcal{X}, d)$ be a metric space. Let $h \colon \mathcal{X} \to \mathbb{R}$ be a function and $X \subset \mathcal{X}$ be a nonempty subset. We say that the function h is *(sequentially) continuous on* X at $x \in X$ if, for all sequences $\{x^i\}_{i \geq 0} \subset X$, we have that

$$\lim_{i \to +\infty} x^i = x \implies \lim_{i \to +\infty} h(x^i) = h(x) \; . \tag{8}$$

We say that the function h is *(sequentially) continuous on* X if it is (sequentially) continuous at any $x \in X$.

Proposition 3 *Let $(\mathcal{X}, d)$ be a nonempty metric space, $\mathcal{Y}$ be a nonempty set and $c \colon \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ be a finite coupling such that the function $c(\cdot, y) \colon \mathcal{X} \to \mathbb{R}$ is lsc, for any $y \in \mathcal{Y}$. Let $X \subset \mathcal{X}$ be a nonempty subset. Let $h \colon \mathcal{X} \to \mathbb{R}$ be a c-subdifferentiable function on X, as defined in* (2e). *Let $Y \colon X \rightrightarrows \mathcal{Y}$ be a c-dual selector of $\partial_c h$ on X, as in Definition 1.*

If the couple $\big(c, Y\big)$ satisfies the pointwise locally equicontinuous property (5) *on X, then we have that*

$$\textit{the function } h \textit{ is (sequentially) continuous on } X \; . \tag{9}$$

Proof.

Let us proceed by contradiction. Similarly to [1, §1.10], we define *the domain of (sequential) continuity* $\operatorname{cont} h$ of the function h by

$$\operatorname{cont} h = \{x \in X : h \text{ is (sequentially) continuous on } X \text{ at } x\} \; . \tag{10}$$

Let $x \in X \setminus \operatorname{cont} h$ be a discontinuity point. By definition (10) of the domain of continuity, there exists a sequence $\{x^i\}_{i \geq 0} \subset X$ converging to x such that

$$\liminf_{i \to +\infty} h(x^i) < h(x) \quad \text{ or } \quad h(x) < \limsup_{i \to +\infty} h(x^i) \; ,$$

according to [12, Exercise 1.12]. As shown in the proof of Theorem (2), the function $h + \iota_X$ is lsc. Thus, we know that $h(x) \leq \liminf_{i \to +\infty} h(x^i)$; as a consequence, we get that

$$h(x) < \limsup_{i \to +\infty} h(x^i) \; .$$

Thus, there exists $\varepsilon > 0$ such that $\varepsilon + h(x) \leq h(x^i)$, for i large enough. For each $i \geq 0$, we consider some $y^i \in Y(x^i) \subset \partial_c h(x^j)$ by (3a). It follows that

$$\begin{aligned} 0 < \varepsilon &\leq h(x^i) - h(x) \; , && \text{(for } i \text{ large enough, by definition of } \varepsilon > 0\text{)} \\ &\leq c(x^i, y^i) - c(x, y^i) \; , && \text{(by (2b) as } y^i \in \partial_c h(x^i)\text{, and the function } h \text{ and } c \text{ are finite)} \\ &\leq \frac{\varepsilon}{2} \; , \;\; \text{for } i \text{ large enough,} \end{aligned}$$

using that $\lim_{i\to+\infty} x^i = x$ and that, by (5), $\exists \eta > 0\ ,\ \ d(x^i, x) \le \eta \implies |c(x^i, y^i) - c(x, y^i)| \le \varepsilon/2$, where $x' = x^i$ and $y = y^i$. Thus, we obtain a contradiction. We deduce that $X \setminus \mathrm{cont}\ h = \emptyset$. Thus, the function h is (sequentially) continuous on X, and this ends the proof. □

Thus, to apply Theorem 2 to $\min_X h$, we have to ensure the following necessary condition between the objective function h and the constraint set X: for any discontinuity point $x \in X$ of h, there is no sequence $\{x^i\}_{i\ge0} \subset X$ converging towards x such that $\limsup_{i\to+\infty} h(x^i) > h(x^*)$.

2.3 Convergence theorem with Lipschitz-like property and c-subgradient selector

With the intent of applying, in Sect. 3, a convergence result for a Capra-cutting plane method to minimize ℓ_0, we propose a corollary of Theorem 2. In Corollary 4, instead of a c-dual selector, we use a c-subgradient selector as in Definition 1. Furthermore, the pointwise locally equicontinuous property (5) is replaced by the stronger (but more practical) Lipschitz-like assumption (11).

Corollary 4 *Let $(\mathcal{X}, d)$ be a nonempty metric space, $\mathcal{Y}$ be a nonempty set and $c\colon \mathcal{X}\times\mathcal{Y} \to \mathbb{R}$ be a finite coupling such that the function $c(\cdot, y)\colon \mathcal{X} \to \mathbb{R}$ is lsc, for any $y \in \mathcal{Y}$. Let $X \subset \mathcal{X}$ be a nonempty compact subset. Let $h\colon X \to \mathbb{R}$ be a c-subdifferentiable function on X, as defined in* (2e).

Let $D : \mathcal{X} \to \mathcal{Y}$ be a c-subgradient selector *of $\partial_c h$ on X, as in Definition 1. Let us assume that the couple $\big(c, Y\big)$ satisfies the following* Lipschitz-like property *on X, for some positive scalar $M > 0$,*

$$|c(x', D(x')) - c(x, D(x'))| \le M d(x, x')\ ,\ \ \forall x, x' \in X\ . \tag{11}$$

Then, for all $x^0 \in X$ there exist a sequence $\{x^i\}_{i\ge0} \subset X$ and a sequence $\{z^i\}_{i\ge1} \subset \mathbb{R}$ such that, for all $i \ge 1$,

$$(x^i, z^i) \in \mathop{\arg\min}_{(x,z)\in X\times\mathbb{R}} \Big\{ z \in \mathbb{R} : z \ge c\big(x, D(x^j)\big) - c\big(x^j, D(x^j)\big) + h(x^j)\ ,\ \ \forall j \le i-1 \Big\}\ , \tag{12}$$

and these sequences satisfy

- *$\{z^i\}_{i\ge1}$ increases to $h^* = \inf_X h$;*
- *$\{x^i\}_{i\ge0}$ has a subsequence $\{x^{\nu(i)}\}_{i\ge0}$ converging to an optimal solution of* (4)*, that is, to some $x^* \in \arg\min_X h$.*

Remark 5 *The Lipschitz-like property* (11) *of the couple $\big(c, D\big)$ implies the pointwise locally equicontinuous property* (5) *of the couple $\big(c, Y\big)$, where $Y(x) = \big\{D(x)\big\}, \forall x \in X$.*

Indeed, let $\varepsilon > 0$ and $x \in X$. For given $\eta > 0$ and $x' \in B(x, \eta)$, we have that

$$
\begin{aligned}
&\sup_{y \in Y(B(x,\eta)\cap X)} |¢(x', y) - ¢(x, y)| \\
&= \sup_{x'' \in B(x,\eta)\cap X} |¢(x', D(x'')) - ¢(x, D(x''))| \quad (\text{as } Y(B(x,\eta)\cap X) = \textstyle\bigcup_{x'' \in B(x,\eta)\cap X}\{D(x'')\}) \\
&\leq \sup_{x'' \in B(x,\eta)} \underbrace{|¢(x', D(x'')) - ¢(x'', D(x''))|}_{\leq M d(x', x'')} + \underbrace{|¢(x'', D(x'')) - ¢(x, D(x''))|}_{\leq M d(x, x'')} , \\
&\qquad\qquad\qquad\qquad\qquad\qquad (\text{according to the Lipschitz-like property (11)}) \\
&\leq M \sup_{x'' \in B(x,\eta)} d(x', x) + d(x, x'') + d(x, x'') \qquad (\text{by the triangular inequality}) \\
&\leq M 3\eta \; . \qquad (\text{as } x'' \in B(x,\eta) \text{ and } x' \in B(x,\eta))
\end{aligned}
$$

Thus (5) *is satisfied for* $\eta = \frac{\varepsilon}{3M}$.

3 Capra-cutting plane method for sparse minimization problems

In this section, we consider the Euclidean space $\mathbb{R}^n$, equipped with the scalar product $\langle \, | \, \rangle$, and the Euclidean norm $\|\cdot\|$. We denote by $S = \{x \in \mathbb{R}^n \mid \|x\| = 1\}$ the Euclidean unit sphere. We define the *support* of a vector $x \in \mathbb{R}^n$ by $\mathrm{supp}(x) = \{k \in [\![1, n]\!] \mid x_k \neq 0\}$. The ℓ_0 *pseudonorm* is the function $\ell_0 \colon \mathbb{R}^n \to [\![1, n]\!]$ defined by

$$
\ell_0(x) = \big|\mathrm{supp}(x)\big| \; , \;\; \forall x \in \mathbb{R}^n \; , \tag{13}
$$

where $|J|$ denotes the cardinality of a subset $J \subseteq [\![1, n]\!]$. Then, we consider the problem of minimizing the pseudonorm ℓ_0 over a compact set $X \subset S$ included in the Euclidean unit sphere[2]:

$$
\min_{x \in X} \ell_0(x) \; . \tag{14}
$$

In §3.1, we give basic definitions and results in Capra-convexity for the pseudonorm ℓ_0, and then we discuss the applicability of previous convergence results (motivating why we need a suitable dual selector for the Capra-cutting plane method). In §3.2, we propose a Capra-cutting plane method for minimizing ℓ_0 over a compact set included in the Euclidean unit sphere.

3.1 Justification of a suitable subgradient selector for Capra-cutting plane method

We remind the following result for the ¢-subdifferential of ℓ_0, which makes ℓ_0 a good candidate to apply Theorem 2.

[2]It is because the pseudonorm ℓ_0 is 0-homogeneous that we restrict the constraints to a subset of the Euclidean unit sphere, thus ignoring the trivial minimum at the origin.

Theorem 6 ([4, Proposition 2]) *We define, as in [4], the* Capra coupling $\cent\colon \mathbb{R}^n\times\mathbb{R}^n \to \mathbb{R}$ *by*

$$\forall y \in \mathbb{R}^n \ , \ \ \cent(x,y) = \begin{cases} \frac{\langle x \mid y \rangle}{\|x\|} \ , & \text{if } x \neq 0 \ , \\ 0 \ , & \text{if } x = 0 \ . \end{cases} \tag{15}$$

Then, the pseudonorm ℓ_0 is $\cent$-subdifferentiable on $\mathbb{R}^n$, that is,

$$\partial_{\cent}\ell_0(x) \neq \emptyset \ , \ \ \forall x \in \mathbb{R}^n \ . \tag{16}$$

Here we discuss how Pallaschke-Rolewicz's convergence result [11, Theorem 9.1.1] cannot be applied without caution to the Capra-cutting plane method for ℓ_0 minimization. We postpone the more technical disscusion on Rubinov's result to the Appendix A.1.

We place ourselves in the case of the Capra-coupling $\cent : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ and the minimization of the pseudonorm ℓ_0 over a compact set $X \subset S \subset \mathbb{R}^n \setminus \{0\}$. Then, the pointwise locally equicontinuous property (5) in Theorem 2 writes as

$$\forall \varepsilon > 0 \ , \ \ \forall x \in X \ , \ \ \exists \eta > 0 \ , \ \ \forall x' \in B(x,\eta) \cap X \ , \ \ \sup_{y \in Y(B(x,\eta)\cap X)} |\cent(x',y) - \cent(x,y)| \leq \varepsilon \ , \tag{17a}$$

for any $\cent$-dual selector $Y : X \rightrightarrows \mathbb{R}^n$ of $\partial_{\cent}\ell_0$, whereas the corresponding assumption — called "locally uniform continuity" in [11, Theorem 9.1.1] — can be stated, using the Capra-coupling $\cent$, as

$$\forall \varepsilon > 0 \ , \ \ \forall x \in X \ , \ \ \exists \eta > 0 \ , \ \ \forall x' \in B(x,\eta) \cap X \ , \ \ \sup_{y \in \mathbb{R}^n} |\cent(x',y) - \cent(x,y)| \leq \varepsilon \ . \tag{17b}$$

We observe that Equation (17b) implies Equation (17a) as $Y(x'') \subset \mathbb{R}^n$ for all $x'' \in X$. In the next Proposition 7, we are showing that Equation (17b) is rarely satisfied, depending on the constraint set $X \subset \mathbb{R}^n$.

Proposition 7 *For any $x \in \mathbb{R}^n \setminus \{0\}$, we have that*

$$\sup_{y \in \mathbb{R}^n} |\cent(x',y) - \cent(x,y)| = +\infty \ , \ \ \forall x' \notin \mathbb{R}_{++}x \ . \tag{18}$$

Proof. Let $x' \in \mathbb{R}^n \setminus \mathbb{R}_{++}x$. We have that

$$\begin{aligned} \sup_{y \in \mathbb{R}^n} |\cent(x',y) - \cent(x,y)| &\geq |\cent(x',\lambda x') - \cent(x,\lambda x')| \ , \ \ \forall \lambda > 0 \ , \\ &= \lambda \Big| \|x'\| - \Big\langle \frac{x}{\|x\|} \mid x' \Big\rangle \Big| \ , \ \ \forall \lambda > 0 \ , \qquad \text{(by (15) and as } x \neq 0) \\ &= \frac{\lambda}{\|x\|} \underbrace{\big(\|x\| \, \|x'\| - \langle x \mid x' \rangle \big)}_{>0} \ , \ \ \forall \lambda > 0 \ . \end{aligned}$$

(by Cauchy-Schwarz inequality and as $x' \notin \mathbb{R}_{++}x$)

It follows that $\sup_{y \in \mathbb{R}^n} |\cent(x',y) - \cent(x,y)| = +\infty$. □

So, it makes it impossible to apply Equation (17b) (and thus [11, Theorem 9.1.1]) for the Capra-cutting plane method for minimizing ℓ_0 over a compact set X, except in the very special case where $B(x,\eta) \cap X \subset \mathbb{R}_{++}x$ for any $x \in X$ and $\eta > 0$ small enough.

The above analysis justifies the choice of a suitable $\cent$-subgradient selector and of a suitable constraint set in the next §3.2.

3.2 Converging Capra-cutting plane algorithm for ℓ_0 minimization

Here, we propose a convergent Capra-cutting plane algorithm, parameterized by a positive scalar $\eta > 0$.

Proposition 8 *Let S be the unit Euclidean sphere and $X \subset S$ be a compact set. Let $\eta > 0$ and the set $R_\eta \subset \mathbb{R}^n$ be defined by*

$$R_\eta = \big\{ x \in \mathbb{R}^n \,\big|\, \min_{\substack{1 \le j \le n \\ x_j \neq 0}} |x_j| \ge \eta \big\} \ . \tag{19}$$

We consider the sparse minimization problem

$$\min_{x \in X \cap R_\eta} \ell_0(x) \ . \tag{20}$$

We introduce the following (well-defined) minimal subgradient norm *$\cent$-subgradient selector $D \colon X \cap R_\eta \to \mathbb{R}^n$ defined by*

$$\big\{ D(x) \big\} = \arg\min_{y \in \partial_{\cent} \ell_0(x)} \|y\| \ , \ \ \forall x \in X \cap R_\eta \ . \tag{21}$$

Then, for any $x^0 \in X \cap R_\eta$ and any sequences $\{x^i\}_{i\ge 0}$ and $\{z^i\}_{i \ge 1}$ generated by the cutting plane method $\mathrm{CP}(\ell_0, X \cap R_\eta, \cent; D, x^0)$ in (6), we have that

- *$\{z^i\}_{i\ge 1}$ increases to $\min_{x \in X \cap R_\eta} \ell_0(x)$;*
- *$\{x^i\}_{i \ge 0}$ has a subsequence $\{x^{\nu(i)}\}_{i\ge 0}$ converging to an optimal solution of (20), that is, to some $x^* \in \arg\min_{x \in X \cap R_\eta} \ell_0(x)$.*

Furthermore, the couple $(\cent, D)$ satisfies the Lipschitz-like property (11) for $M = \frac{\sqrt{1-\eta^2}+1}{\eta^2}$.

Before giving the proof, let us comment on the design of the set R_η in (19). We say that *sparse points* are those $x \in \mathbb{R}^n$ such that $\ell_0(x) < n$. Thus, the set $R_\eta \subset \mathbb{R}^n$ in (19) is designed in such a way that points near the sparse points (those on the two axis in the case $n = 2$) are removed, but not the sparse points themselves. This is illustrated in Figure 1 in the case $n = 2$: points near the two axis (the two axis are the sparse points here) are removed, but not all the sparse points are removed. As we are dealing with the ℓ_0 pseudonorm, it appears that ℓ_0 takes a constant value on each of the connected subsets of R_η.

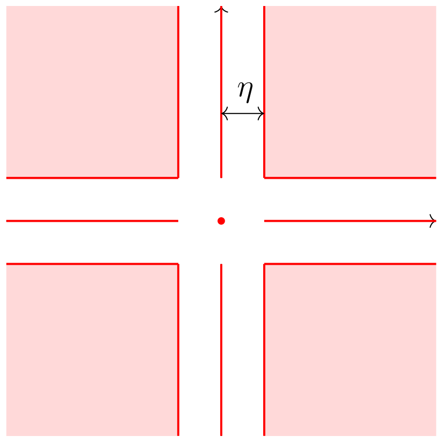

Figure 1: Representation in red of the set R_η in (19) from Proposition 8

We now give the proof of Proposition 8.

Proof. First, let us prove that the $\cent$-subgradient selector D is well-defined by (21), that is, the $\arg\min$ is a singleton. Let $x \in \mathbb{R}^n \setminus \{0\}$, and consider the minimization problem

$$\min_{y \in \partial_{\cent}\ell_0(x)} \|y\|^2 \ . \tag{22}$$

As the Capra-subdifferential $\partial_{\cent}\ell_0(x)$ is a closed convex set [3, Proposition 1], and as the function $\|\cdot\|^2$ is strongly convex, there is a unique optimal solution $y^* \in \mathbb{R}^n$ to the problem (22).

Second, we provide a formula for the $\cent$-subgradient selector D defined in (21). We will need the following notations. Let $y \in \mathbb{R}^n$. For any $L \subset [\![1, n]\!]$, we denote by $y_L \in \mathbb{R}^n$ the vector which coincides with $y \in \mathbb{R}^n$, except for the components outside of L that vanish. We denote by ν a permutation of $[\![1, n]\!]$ such that $|y_{\nu(1)}| \geq \cdots \geq |y_{\nu(n)}|$, and then $\|y\|^{\top}_{(k)} = \big(\sum_{i=1}^{i} |y_{\nu(i)}|^2\big)^{1/2}$ is the so-called *top-(2,k) norm* (following the notations in [3, §2.2.2]).

Now, we remind the formulas of the Capra-subdifferential of ℓ_0 for the Euclidean source norm. Let $x \in \mathbb{R}^n \setminus \{0\}$ and $y \in \partial_{\cent}\ell_0(x)$. Then, denoting $\ell = \ell_0(x)$, we have that, by [8, Theorem 3.1,Equation (27b)],

$$y_{\mathrm{supp}(x)} = \lambda x \ , \quad \text{for a certain } \ \lambda \geq 0 \ , \tag{23a}$$

$$|y_j| \leq \min_{i \in \mathrm{supp}(x)} |y_i| \ , \ \ \forall j \notin \mathrm{supp}(x) \ , \tag{23b}$$

$$|y_{\nu(k+1)}|^2 \geq \big(\|y\|^{\top}_{(k)} + 1\big)^2 - \big(\|y\|^{\top}_{(k)}\big)^2 \ , \ \ \forall k \in [\![0, \ell - 1]\!] \ , \tag{23c}$$

$$|y_{\nu(\ell+1)}|^2 \leq \big(\|y\|^{\top}_{(\ell)} + 1\big)^2 - \big(\|y\|^{\top}_{(\ell)}\big)^2 \quad (\text{when } \ell \neq n) \ . \tag{23d}$$

Notice that $y_{\mathrm{supp}(x)}$ satisfies Equations (23). Indeed, as y satisfies (23a), so does $y_{\mathrm{supp}(x)}$. As y satisfies (23b), we get that $0 = |(y_{\mathrm{supp}(x)})_j| \leq |y_j| \leq \min_{i \in \mathrm{supp}(x)} |y_i| = \min_{i \in \mathrm{supp}(x)} |(y_{\mathrm{supp}(x)})_i|$, for any $j \notin \mathrm{supp}(x)$. As y satisfies (23c), so does $y_{\mathrm{supp}(x)}$, because the inequalities only involve the entries $y_{\nu(k)}$ for $k \in [\![0, \ell]\!]$, hence only involve the (nonzero) entries of $y_{\mathrm{supp}(x)}$. As y satisfies (23d), we get that $0 = |(y_{\mathrm{supp}(x)})_{\nu(\ell+1)}| \leq |y_{\nu(\ell+1)}|^2 \leq \big(\|y\|^{\top}_{(\ell)} + 1\big)^2 - \big(\|y\|^{\top}_{(\ell)}\big)^2 = \big(\|y_{\mathrm{supp}(x)}\|^{\top}_{(\ell)} + 1\big)^2 -$

$\left(\|y_{\mathrm{supp}(x)}\|^{\top}_{(\ell)}\right)^2$, because the inequalities only involve the entries $y_{\nu(k)}$ for $k \in [\![0,\ell]\!]$, hence only involve the (nonzero) entries of $y_{\mathrm{supp}(x)}$. Thus, we get that $y_{\mathrm{supp}(x)} \in \partial_{\cent}\ell_0(x)$.

As[3] $\|y^*\|^2 = \|y^*_{\mathrm{supp}(x)}\|^2 + \|y^*_{-\mathrm{supp}(x)}\|^2$, we deduce that the minimum y^* of (22) satisfies $y^* = y^*_{\mathrm{supp}(x)}$. Thus, according to (23a), we have obtained that

$$y^* = \lambda x \; , \;\text{ for some scalar } \lambda \geq 0 \; . \tag{24}$$

Now, we focus on finding the scalar $\lambda \geq 0$ in (24). We combine (24) and (23c), and we get

$$\begin{aligned}
& \lambda^2 |x_{\nu(k+1)}|^2 \geq \left(\lambda \|x\|^{\top}_{(k)} + 1\right)^2 - \lambda^2 \left(\|x\|^{\top}_{(k)}\right)^2 \; , \;\; \forall k \in [\![0,\ell-1]\!] \; , \\
\iff & \lambda^2 |x_{\nu(k+1)}|^2 - 2\lambda \|x\|^{\top}_{(k)} - 1 \geq 0 \; , \;\; \forall k \in [\![0,\ell-1]\!] \; , \\
\iff & \lambda \geq \frac{\|x\|^{\top}_{(k)} + \sqrt{(\|x\|^{\top}_{(k)})^2 + |x_{\nu(k+1)}|^2}}{|x_{\nu(k+1)}|^2} \; , \;\; \forall k \in [\![0,\ell-1]\!] \; , \\
& \qquad \text{(considering the second order polynom in } \lambda \text{ and keeping the nonnegative solution)} \\
\iff & \lambda \geq \frac{\|x\|^{\top}_{(k)} + \|x\|^{\top}_{(k+1)}}{|x_{\nu(k+1)}|^2} \; , \;\; \forall k \in [\![0,\ell-1]\!] \; , \qquad \text{(as } (\|x\|^{\top}_{(k)})^2 + |x_{\nu(k+1)}|^2 = (\|x\|^{\top}_{(k+1)})^2) \\
\iff & \lambda \geq \frac{\|x\|^{\top}_{(\ell-1)} + \|x\|^{\top}_{(\ell)}}{|x_{\nu(\ell)}|^2} \; , \\
& \qquad \text{(as } \|x\|^{\top}_{(k)} \text{ and } \|x\|^{\top}_{(k+1)} \text{ are nondecreasing in } k \text{, and } |x_{\nu(k+1)}| \text{ is nonincreasing in } k) \\
\iff & \lambda \geq \frac{\sqrt{\|x\|^2 - x_-^2} + \|x\|}{x_-^2} \;\; \text{where we have set } \; x_- = |x_{\nu(\ell)}| \; ,
\end{aligned}$$

as $|x_{\nu(1)}| \geq \cdots \geq |x_{\nu(\ell)}| = x_- \geq |x_{\nu(\ell+1)}| = 0 = \cdots = |x_{\nu(n)}|$, and $(\|x\|^{\top}_{(k)})^2 = \sum_{i=1}^{k} |x_{\nu(i)}|^2$, we have $(\|x\|^{\top}_{(\ell)})^2 = \|x\|^2$ and $(\|x\|^{\top}_{(\ell-1)})^2 = \|x\|^2 - x_-^2$.

As a consequence, we obtain $\lambda^*(x) = \frac{\sqrt{\|x\|^2 - x_-^2} + \|x\|}{x_-^2}$ for the optimal solution $y^* = \lambda^*(x) x$ of the minimization problem (22). Thus, the $\cent$-subgradient selector D, defined in (21), is given by

$$D(x) = \lambda^*(x) x = \frac{\sqrt{\|x\|^2 - x_-^2} + \|x\|}{x_-^2} x \quad \text{ where } \; x_- = \min_{k \in \mathrm{supp}(x)} |x_k| \; . \tag{25}$$

Third, we prove that the Lipschitz-like property (11) is satisfied between the coupling $\cent$ and the $\cent$-subgradient selector $D \colon X \cap R_\eta \to \mathbb{R}^n$ defined in (21). Let $\eta > 0$ and let the set R_η be defined as in (19). For $x, x' \in X \cap R_\eta$, we have that

[3] Here, following notation from Game Theory, we have denoted by $-L$ the complementary subset of L in $\{1, \ldots, n\}$: $L \cup -L = \{1, \ldots, n\}$ and $L \cap (-L) = \emptyset$.

$$
\begin{aligned}
|¢\big(x, D(x)\big) - ¢\big(x', D(x)\big)| &= \Big\langle \frac{x}{\|x\|} - \frac{x'}{\|x'\|} \mid D(x) \Big\rangle \ , && \text{(by definition (15) of the Capra-coupling)} \\
&= \big\langle x - x' \mid D(x) \big\rangle && \text{(as } X \subset S \text{, hence } \|x\| = \|x'\| = 1\text{)} \\
&\leq \|D(x)\| \, \big\|x - x'\big\| && \text{(by Cauchy-Schwarz inequality)} \\
&= \frac{\sqrt{\|x\|^2 - x_-^2} + \|x\|}{x_-^2} \|x\| \, \big\|x - x'\big\| \ , && \text{(by (25))} \\
&\leq \frac{\sqrt{1-\eta^2} + 1}{\eta^2} \big\|x - x'\big\|
\end{aligned}
$$

as $X \subset S$, hence $\|x\| = 1$, and as $x \in R_\eta \iff x_- \geq \eta$ by definition (19) of the set R_η. So, the Lipschitz-like property (11) is satisfied with $M = \frac{\sqrt{1-\eta^2}+1}{\eta^2}$. We conclude using Corollary 4. □

Remark 9 *We have to be cautious when we apply Theorem 2 directly to* $\min_{x \in X} \ell_0(x)$*, without possible additional joint assumptions on* ℓ_0 *and* X*. Indeed, a necessary condition to apply Theorem 2 comes from Proposition 3 and reads as follows: any sequence* $\{x^i\}_{i \geq 0} \subset X$ *converging to a discontinuity point* $x \in X$ *of the lsc function* ℓ_0 *— that is, to a sparse point in* X *— has to satisfy* $\lim_{i \to +\infty} \ell_0(x^i) = \ell_0(x)$*. Thus, by intersecting* X *with* R_η*, we can remove the "faulty" sequences from* X *and we can apply to* $\min_{x \in X \cap R_\eta} \ell_0(x)$ *the convergence result in Theorem 2. Using this restriction framework, the problem of solving* $\min_{x \in X} \ell_0(x)$ *reduces to the problem of finding* $\eta > 0$ *small enough such that*

$$
\big(\arg\min_X \ell_0\big) \cap R_\eta \neq \emptyset \ . \tag{26}
$$

4 Conclusion

In this paper, we have proposed a convergence result (Theorem 2) for the generalized cutting plane method, suitable to the minimization of a c-subdifferentiable function over a compact subset of a metric space. The convergence relies on the notion of c-dual selector and a joint pointwise locally equicontinuous property (5) between the coupling, the objective function and the c-dual selector. In Proposition 3, we have also provided a necessary condition — on the (sequential) continuity of the objective function h on the constraint set X — to satisfy the pointwise locally equicontinuous property. Finally, we have proposed a Capra-cutting plane method that satisfies the convergence assumptions for the problem of minimizing the pseudonorm ℓ_0 over a compact set included in the unit Euclidean sphere. In future works, we intend to develop numerical methods based on the Capra-cutting plane method to solve sparse minimization problems.

A Appendix

A.1 Comparison with Rubinov's result

Here, we discuss the difficulty of applying previous convergence results in [13] to a Capra-cutting plane method for a minimization problem of the form $\min_X \ell_0$.

The sequentially uniformly compact property found in [13, Proposition 9.3] can be stated as follows in the generalized convexity framework (that is, with couplings). Let $Y \subset \mathcal{Y}$ and $c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ be a finite coupling. We say that the couple $\big(c, Y\big)$ satisfies the *sequentially uniformly compact property* if, for any sequence $\{y^k\}_{k\in\mathbb{N}} \subset Y$, there exists a converging subsequence $\{y^{\nu(k)}\}_{k\in\mathbb{N}}$ to $y \in Y$ such that

$$\sup_{x'\in\mathcal{X}} |c(x', y^{\nu(k)}) - c(x', y)| \xrightarrow[k\to+\infty]{} 0 \; . \tag{27}$$

Let $X \subset \mathbb{R}^n$ be a nonempty set of constraints. We are going to show that the above condition (27) does not hold true for the couple $\big(¢, \bigcup_{x\in X} \partial_{¢}\ell_0(x)\big)$. Let $x \in X \setminus \{0\}$. We know by (25) that the sequence $\big\{(1+k)\lambda^*(x)x\big\}_{k\geq 0} \subset \mathbb{R}^n$ is included in $\partial_{¢}\ell_0(x)$. The sequence of functions $\big\{¢(\cdot, (1+k)\lambda^*(x)x)\big\}_{k\geq 0}$ has no uniformly converging subsequence because

$$¢(x, (1+k)\lambda^*(x)x) = (1+k)\lambda^*(x)¢(x, x) = (1+k)\lambda^*(x)\, \|x\| \xrightarrow[k\to\infty]{} +\infty \; .$$

A.2 Summary table of different assumptions in the literature

	base functions regularity	equicontinuous-like properties
[7]	affine	bounded slopes
[13, Proposition 9.2]	concave continuous	bounded directional derivatives
[13, Proposition 9.3]	continuous	sequentially uniformly compact property (27)
[11, Theorem 9.1.1]	continuous	pointwise equicontinuous property (17b)
Theorem 2	lower semicontinuous	pointwise locally equicontinuous property (5)

Table 1: Comparison of the assumptions for convergence results of a cutting plane method

References

[1] H. H. Bauschke and P. L. Combettes. *Convex analysis and monotone operator theory in Hilbert spaces*. CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC. Springer-Verlag, New York, second edition, 2017.

[2] J.-P. Chancelier and M. De Lara. Hidden convexity in the l_0 pseudonorm. *Journal of Convex Analysis*, 28(1):203–236, 2021.

[3] J.-P. Chancelier and M. De Lara. Capra-convexity, convex factorization and variational formulations for the ℓ_0 pseudonorm. *Set-Valued and Variational Analysis*, 30:597–619, 2022.

[4] J.-P. Chancelier and M. De Lara. Constant along primal rays conjugacies and the l_0 pseudonorm. *Optimization*, 71(2):355–386, 2022.

[5] M. Conforti, G. Cornuéjols, and G. Zambelli. *Integer Programming*. Springer Cham, Switzerland, 2014.

[6] R. E. Gomory. Outline of an algorithm for integer solutions to linear programs. *Bulletin of the American Mathematical Society*, 64:275–278, 1958.

[7] J. E. Kelley, Jr. The cutting-plane method for solving convex programs. *Journal of the society for Industrial and Applied Mathematics*, 8(4):703–712, 1960.

[8] A. Le Franc, J.-P. Chancelier, and M. De Lara. The Capra-subdifferential of the ℓ_0 pseudonorm. *Optimization*, 73(4):1229–1251, 2024.

[9] J. E. Martínez-Legaz. Generalized convex duality and its economic applications. In N. Hadjisavvas, S. Komlósi, and S. Schaible, editors, *Handbook of Generalized Convexity and Generalized Monotonicity. Nonconvex Optimization and Its Applications*, volume 76, pages 237–292. Springer-Verlag, New York, 2005.

[10] J. J. Moreau. Fonctionnelles convexes. *Séminaire Jean Leray*, 2:1–108, 1966-1967.

[11] D. Pallaschke and S. Rolewicz. *Foundations of mathematical optimization*, volume 388 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 1997.

[12] R. T. Rockafellar and R. J.-B. Wets. *Variational Analysis*. Springer-Verlag, Berlin, 1998.

[13] A. Rubinov. *Abstract convexity and global optimization*, volume 44 of *Nonconvex Optimization and its Applications*. Kluwer Academic Publishers, Dordrecht, 2000.

[14] A. Schrijver. *Theory of Linear and Integer Programming*. Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley, 1986.

[15] I. Singer. *Abstract Convex Analysis.* Canadian Mathematical Society Series of Monographs and Advanced Texts. John Wiley & Sons, Inc., New York, 1997.