WEAK CONVEXITY AND APPROXIMATE SUBDIFFERENTIALS

W. VAN ACKOOIJ, F. ATENAS, C. SAGASTIZÁBAL

ABSTRACT. We explore and construct an enlarged subdifferential for weakly convex functions. The resulting object turns out to be continuous with respect to both the function argument and the enlargement parameter. We carefully analyze connections with other constructs in the literature and extend well-known variational principles to the weakly convex setting. By resorting to the new enlarged subdifferential, we provide an algorithmic pattern of descent for weakly convex minimization. Under minimal assumptions, we show subsequential convergence to a critical point, and links with difference-of-convex algorithms and criticality conditions are also discussed.

AMS subject classification: 65K10, 49J53, 49M05

1. MOTIVATION

Because subdifferentials are crucial tools in Variational Analysis, they have been declined in various forms, to the extent that the search for the "best" subdifferential in the nonconvex setting was compared in [DJ97] to a *safari season in the field of nonsmooth optimization*.

Among the many different gradient generalizations proposed so far, the Mordukhovich subdifferential stands out by its powerful calculus [Mor06]. Early study of these generalizations can be traced back to [Mor76]; see also [Kru85; DLZ85]. Clarke's subdifferential, on the other hand, is appealing from an implementation point of view because of its convexity. Another first-order object in the literature is the regular or Fréchet subdifferential. The outer limit of the regular subdifferential defines the Mordukhovich subdifferential.

When minimizing a function h by some iterative process using first-order information provided by $\partial h(\cdot)$, continuity of the underlying subdifferential arises as a crucial issue. It appears that all the tools above are outer semicontinuous as multifunctions but none is continuous. In an algorithmic scheme, the lack of inner semicontinuity of subdifferentials hinders the definition of criticality certificates. The purpose of such certificates is twofold. First, they allow to stop the iterative process with a solution that is sufficiently close to some critical point. At the same time, they provide asymptotic satisfaction of the criticality condition $0 \in \partial h(\bar{z})$. Namely, if a critical point $\bar{z}$ satisfies the condition for some subdifferential, only inner semicontinuity of the multifunction $\partial h(\cdot)$ ensures that building a sequence $\{g_n \in \partial h(z_n)\} \to 0$ for any sequence $\{z_n\} \to \bar{z}$ will be possible.

Date: Working notes. Manuscript from: January 16, 2024.

In the convex setting and for given $x \in \mathbb{R}^N$, the approximate subdifferential appears as a suitable tool in that respect:

$$\partial_\varepsilon H(x) = \left\{ g \in \mathbb{R}^N : H(z) \geq H(x) + \langle g, z - x \rangle - \varepsilon \text{ for all } z \in \mathbb{R}^N \right\} \text{ for } H \text{ convex.} \tag{1}$$

Indeed, this object not only enjoys full calculus [HL96, Ch.XI], but it is also locally Lipschitz continuous on both $\varepsilon > 0$ and x, whenever the convex function H is Lipschitzian [Hir80]. Additionally, thanks to the approximate subdifferential (1), exact subgradients from past iterates can be *transported*, as follows:

$$g_1 \in \partial H(z_1) \Longrightarrow g_1 \in \partial_\varepsilon H(z_2) \text{ for } \varepsilon = H(z_2) - H(z_1) - \langle g_1, z_2 - z_1 \rangle \,, \tag{2}$$

noting that $\varepsilon \geq 0$ by H's convexity. By this token, a numerical approach can accumulate exact subgradient information along iterations to build special approximate subgradients, satisfying $g_n \in \partial_{\varepsilon_n} H(z_n)$, in such a way that, by driving both g_n and $\varepsilon_n \to 0$, the process convergence is ensured.

In this work we explore a construct akin to (1), in a nonconvex setting. More precisely, we are interested in defining continuous approximate subdifferentials for functions $h : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ that are ρ-weakly convex (ρ-w.c.), i.e., such that

$$\text{the augmented function } H^\rho(\cdot) := h(\cdot) + \frac{\rho}{2}|\cdot|^2 \text{ is convex.} \tag{3}$$

In the notation ρ-w.c., we may drop the dependency on ρ when it is clear from the context. Observe that a ρ-w.c. function is also ρ'-w.c. for any $\rho' > \rho$.

The departing point to define the new ρ-w.c. subdifferential enlargement is the relation

$$\partial H^\rho(z) = \partial h(z) + \rho z, \tag{4}$$

which holds for all $z \in \mathbb{R}^N$ with nonempty (Clarke's) subdifferential $\partial h(z)$. Since the augmented function is convex, the following approximate subgradient extension, depending on the w.c. parameter ρ, seems natural:

$$\partial_\varepsilon^\rho h(z) := \partial_\varepsilon H^\rho(z) - \rho z. \tag{5}$$

Because the ρ-w.c. approximate subdifferential is just a linear translation of the ε-subdifferential in Convex Analysis, all the calculus developed for the latter in [HL96, Ch. XI] is available for the former.

The work is organized as follows. Section 2 starts by illustrating the notion for two simple functions. We continue in § 2.3 with the enlargement proposed by [MS92] for D.C. functions, defined as a difference-of-convex functions. When particularized to a weakly convex h (a very specific structured D.C. function), we show that the ε-subdifferential by [MS92] is exactly (5). This is an interesting result, as our alternative equivalent formulation endows the construct in [MS92] with calculus rules. The section finishes relating (5) with subdifferentials proposed by Goldstein and Bihain, respectively in [Gol77] and [Bih84]. Section 3 is devoted to reformulating for the weakly convex case several well-known variational subgradient principles, such as Ekeland's and Brøndsted-Rockafellar theorem. Following arguments from [Hir80], we also show that the multifunction defined in (5) is Lipschitz continuous on both ε and x. The continuity properties of our approximate subdifferential are exploited from a numerical point of view in Section 4. We propose a new algorithmic pattern of ε-descent for weakly convex minimization, and study its convergence properties. Under minimal requirements (cf. (10) and (11)), satisfied by proximal-like algorithms and by (nonconvex) bundle methods, we first show subsequential convergence to critical points. Global convergence with local linear rate is the topic of § 4.2, provided a couple of additional assumptions, typical in the setting, hold; see (15) and (16) below. Section 5 returns to D.C. literature, specifically, relating the ε-subgradient descent scheme in Section 4 with an algorithm by [AO20]. In particular, we derive a necessary condition for local optimality, expressed in terms of the subdifferential of the original function h. In other works, global and local optimality conditions for D.C. functions are established in terms of the two convex functions whose difference defines h. For the sake of defining optimality certificates, involving the D.C. decomposition is a drawback, notably because the decomposition is not unique. By contrast, by dealing only with $\partial_\varepsilon^\rho h$, our result is more direct, and can be used in practice, thanks to the continuity of the proposed multifunction.

2. Initial properties and relations

We start by explicitly computing approximate subgradients as in (5), for two simple functions, one convex and one weakly convex. Afterwards, we examine relations between our ρ-w.c. subdifferential and other subdifferential enlargements in the literature.

2.1. **Illustrative examples.** We combine various calculus rules for the Convex Analysis ε-subdifferential from Chapter XI in [HL96], applied with $x \in \mathbb{R}$. Below, the ball of radius $\sqrt{r}$ centered at 0 is denoted by $\mathbb{B}_r$ and $\varepsilon, \varepsilon_1, \varepsilon_2 \geq 0$.

– (Ex.XI.1.2.2): if $f(x) = \frac{\rho}{2}x^2 + f_o$, then $\partial_\varepsilon f(\hat{x}) = \rho\hat{x} + \mathbb{B}_{2\rho\varepsilon}$.
– (Prop.XI.1.3.1): $\partial_\varepsilon(\alpha f)(x) = \alpha\partial_{\varepsilon/\alpha} f(x)$ whenever $\alpha > 0$.
– (Thm.XI.3.1.1): $ridom\, f_1 \cap ridom\, f_2 \Longrightarrow \partial_\varepsilon(f_1+f_2)(x) = \bigcup_{\varepsilon_1+\varepsilon_2\leq\varepsilon} \{\partial_{\varepsilon_1} f_1(x) + \partial_{\varepsilon_2} f_2(x)\}$.
– (eq.XI.3.5.5): if $f(x) = \max(f_1(x), f_2(x))$, then

$$\partial_\varepsilon f(x) = \bigcup_{(\alpha,\varepsilon_1,\varepsilon_2)\in S(\varepsilon)} \left\{\partial_{\varepsilon_1}\Big(\alpha f_1\Big)(x) + \partial_{\varepsilon_2}\Big((1-\alpha)f_2\Big)(x)\right\},$$

for $S(\varepsilon) = \left\{(\alpha,\varepsilon_1,\varepsilon_2) : \begin{array}{l} \alpha \in [0,1] \\ \varepsilon_1+\varepsilon_2+f(x)-\alpha f_1(x)-(1-\alpha)f_2(x) \leq \varepsilon \end{array}\right\}$.

2.1.1. *The w.c. approximate subdifferential of the absolute value.* It is known, see for instance Fig.XI.1.2 in [HL96], that the convex function $h(x) = |x|$ for $x \in \mathbb{R}$ has the ε-subdifferential (1)

$$\partial_\varepsilon h(x) = \begin{cases} [-1, -1-\varepsilon/x] & x < -\varepsilon/2 \\ [-1,1] & |x| \leq \varepsilon/2 \\ [1-\varepsilon/x, 1] & x > \varepsilon/2\,. \end{cases}$$

To compare this enlargement with the one provided by (5) for $\rho \geq 0$, consider $H^\rho(x) = f_1(x) + f_2(x)$, for $f_1(x) = |x|$ and $f_2(x) = \frac{\rho}{2}x^2$. By combining the calculus rules above,

$$\partial_\varepsilon H^\rho(\hat{x}) = \rho\hat{x} + \bigcup_{\varepsilon_1+\varepsilon_2\in[0,\varepsilon]} \left\{\partial_{\varepsilon_1}|\cdot|(\hat{x}) + \mathbb{B}_{2\rho\varepsilon_2}\right\}.$$

Therefore,

$$\partial_\varepsilon^\rho h(\hat{x}) = \bigcup_{\varepsilon_1+\varepsilon_2\in[0,\varepsilon]} \left\{\partial_{\varepsilon_1}|\cdot|(\hat{x}) + \mathbb{B}_{2\rho\varepsilon_2}\right\}.$$

In particular, at $\hat{x} = 0$,

$$\partial_\varepsilon^\rho|\cdot|(0) = [-1,1] + \mathbb{B}_{2\rho\varepsilon} = \left[-1-\sqrt{2\rho\varepsilon}, 1+\sqrt{2\rho\varepsilon}\right],$$

while, at $\hat{x} = 1$, since $[1-\varepsilon_1, 1] + \mathbb{B}_{2\rho\varepsilon_2} = [1-\varepsilon_1-\sqrt{2\rho\varepsilon_2}, 1+\sqrt{2\rho\varepsilon_2}]$, it holds that

$$\partial_\varepsilon^\rho|\cdot|(1) = [1-(\varepsilon+\rho/2), 1+\sqrt{2\rho\varepsilon}], \tag{6}$$

whenever $\varepsilon \leq 2$ and $\rho \leq 2\varepsilon$.

Figure 1 shows the multivalued function $\partial_\varepsilon^\rho h(x)$ for $\varepsilon = 1$ and several values of ρ. When $\rho = 0$, the w.c. subdifferential coincides with the Convex Analysis one, as expected.

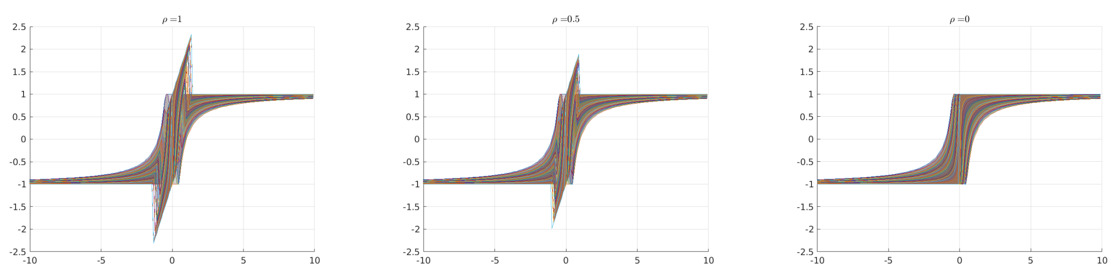

FIGURE 1. Approximate subdifferential (5) of $h(x) = |x|$, for $\varepsilon = 1$ and $\rho \in \{1, 0.5, 0\}$ (left, center, right)

2.1.2. *The w.c. approximate subdifferential of* $\frac{1}{2}|x^2-1|$. The function $h(x) = \frac{1}{2}|x^2-1|$ is weakly convex for $\rho \geq 1$. We let

$$H^\rho(x) = \max(f_1(x), f_2(x)), \text{for } f_1(x) = \frac{1}{2}(\rho+1)x^2 - \frac{1}{2} \text{ and } f_2(x) = \frac{1}{2}(\rho-1)x^2 + \frac{1}{2},$$

and apply the calculus rules. For $\alpha \in (0,1)$, after some algebra we obtain

$$\partial_{\varepsilon_1}(\alpha f_1)(\hat{x}) = \alpha(\rho+1)\hat{x} + \mathbb{B}_{2(\rho+1)\varepsilon_1} \text{ and } \partial_{\varepsilon_2}((1-\alpha)f_2)(\hat{x}) = (1-\alpha)(\rho-1)\hat{x} + \mathbb{B}_{2(\rho-1)(\varepsilon_2)},$$

while

$$S(\varepsilon) = \left\{(\alpha,\varepsilon_1,\varepsilon_2) : \alpha\in[0,1] \text{ and } \varepsilon_1+\varepsilon_2 \leq \left\{\begin{array}{ll} \varepsilon-(1-\alpha)x^2+(1-\alpha) & \text{if } |x|>1 \\ \varepsilon-\alpha(1-x^2) & \text{if } |x|<1 \\ \varepsilon & \text{if } |x|=1 \end{array}\right\}\right\},$$

yielding the expression

$$\partial_\varepsilon^\rho h(\hat{x}) = \bigcup_{(\alpha,\varepsilon_1,\varepsilon_2)\in S(\varepsilon)} \left\{(2\alpha-1)\hat{x} + \mathbb{B}_{2(\rho+1)\varepsilon_1} + \mathbb{B}_{2(\rho-1)\varepsilon_2}\right\}.$$

The corresponding multifunction is shown in Figure 2 for $\varepsilon = 1$ and three values of $\rho \geq 1$.

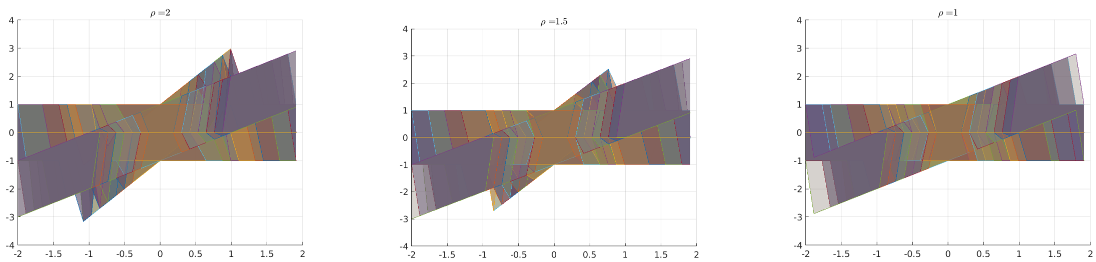

FIGURE 2. Approximate subdifferential (5) of $h(x) = |x^2-1|$, for $\varepsilon = 1$ and $\rho \in \{2, 1.5, 1\}$ (left, middle, right)

2.2. **Properties of the approximate w.c. subdifferential.** Barring the last two relations, the properties below are stated without proof.

Proposition 2.1 (Monotonicity and initial relations)**.** *For a ρ-w.c. function $h : \mathbb{R}^N \to \mathbb{R}\cup\{+\infty\}$ and the set-valued operator defined in (5) the following holds whenever $\varepsilon > 0$.*

(i) $\partial h(x) \subset \partial_\varepsilon^\rho h(x)$ *and* $\partial_\varepsilon^\rho h(x) \neq \emptyset$ *for all* $x \in \mathrm{dom}\, H^\rho$.
(ii) $\rho_1 \leq \rho_2 \Longrightarrow \partial_\varepsilon^{\rho_1} h(x) \subset \partial_\varepsilon^{\rho_2} h(x)$
(iii) $\varepsilon_1 \leq \varepsilon_2 \Longrightarrow \partial_{\varepsilon_1}^\rho h(x) \subset \partial_{\varepsilon_2}^\rho h(x)$
(iv) $g \in \partial_\varepsilon^\rho h(z) \iff h(y) \geq h(z) + \langle g, y-z\rangle - \frac{\rho}{2}|y-z|^2 - \varepsilon$ *for all* $y \in \mathbb{R}^N$.
(v) *At any given z, the approximate w.c. subdifferential defined in (5) coincides with the ε subdifferential of the z-shifted augmented function*

$$H_z^\rho(x) = h(x) + \frac{\rho}{2}|x-z|^2\,, \tag{7}$$

evaluated at $x = z$. Specifically,

$$\partial_\varepsilon H_z^\rho(x) = \partial_\varepsilon^\rho h(x) + \rho(x-z)\,, \tag{8}$$

and, hence, $\partial_\varepsilon^\rho h(z) = \partial_\varepsilon H_z^\rho(z)$.

Proof. The first three items are straightforward. Item (iv) follows from the approximate subgradient inequality for H^ρ. To show (iv), expand $\frac{\rho}{2}|z-w|^2$ and apply items (i) and (vi) from Prop.XI.1.3.1 in [HL96] as follows:

$$\partial_\varepsilon H_z^\rho(x) = \partial_\varepsilon\Big(H^\rho(\cdot) - \rho\,\langle\cdot, z\rangle + \frac{\rho}{2}\|z\|^2\Big)(x) = \partial_\varepsilon H^\rho(x) - \rho z\,.$$

By the definition in (5), the right hand side is equal to $\partial_\varepsilon^\rho h(x) + \rho(x-z)$. Then, (8) follows and taking $x = z$ yields the final assertion. □

We mention in passing a slightly more general definition of weak convexity, considered in [Kru03] and relating back to [Nur73; Nor78]. Instead of the w.c. parameter ρ, the notion depends on certain regulating function $r : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}$, which is why we refer to this concept as r-weak convexity. The aforementioned regulating function satisfies $\frac{r(x,z)}{\|x-z\|} \to 0$ as $x \to z$ and defines non-empty sets

$$G^r(x) = \{x^* \,: h(z) - h(x) - \langle x^*, z-x\rangle + r(x,z) \geq 0\}$$

at any $x \in \mathbb{R}^N$. It turns out ([Kru03, Proposition 1.40]) that for r-weakly convex functions the Fréchet subdifferential is the set G^r, which is is closed, convex and bounded. As a result, also the Mordukhovich and Clarke subdifferentials coincide with $G^r(\cdot)$.

2.3. **Relation with approximate subgradients for difference-of-convex functions.** In [MS92] an approximate subdifferential is proposed for a nonconvex function f, assuming that $f(x) = f_1(x) - f_2(x)$, with both f_1 and f_2 convex. Since a weakly convex function h is a particular D.C. function, with $f_1(x) = H^\rho(x)$ and $f_2(x) = \frac{\rho}{2}|x|^2$, we now state the relation between (5) and that subdifferential, denoted here by $\partial_\varepsilon^{[\text{MS92}]} h$.

By the characterization given in [MS92, Thm. 1] when h is ρ-w.c.

$$\partial_\varepsilon^{[\text{MS92}]} h(x) = \bigcap_{\lambda \geq 0} \left\{ \partial_{\varepsilon+\lambda} H^\rho(x) - \partial_\lambda \left(\frac{\rho}{2}|x|^2 \right) \right\}, \tag{9}$$

where the difference of sets is understood in the sense of the Minkowski sum.

Proposition 2.2 (Interpretation as approximate DC subgradients)**.** *When h is ρ-w.c. the approximate subdifferentials* (5) *and* (9) *coincide.*

Proof. In view of the calculus rules in § 2.1, $\partial_\lambda \left(\frac{\rho}{2}|x|^2 \right) = \rho x + \mathbb{B}_{2\rho\lambda}$. Together with (9),

$$\partial_\varepsilon^{[\text{MS92}]} h(x) = -\rho x + \bigcap_{\lambda \geq 0} \left\{ \partial_{\varepsilon+\lambda} H^\rho(x) - \mathbb{B}_{2\rho\lambda} \right\}.$$

Let $g \in \partial_\varepsilon^{[\text{MS92}]} h(x)$. Then, for each $\lambda \geq 0$,

$$g + \mathbb{B}_{2\rho\lambda} \subseteq \partial_{\varepsilon+\lambda} H^\rho(x) - \rho x.$$

By definition (5), it follows that

$$g + \bigcap_{\lambda \geq 0} \mathbb{B}_{2\rho\lambda} \subseteq \bigcap_{\lambda \geq 0} \partial_{\varepsilon+\lambda}^\rho h(x).$$

Since the intersection in the left-hand side is $\{0\}$, in view of Proposition 2.1(iii), $g \in \partial_\varepsilon^\rho h(x)$. Conversely, suppose $g \in \partial_\varepsilon^\rho h(x)$, and take any $\lambda \geq 0$. From (5) and Proposition 2.1(iii),

$$g + \rho x \in \partial_{\varepsilon+\lambda} H^\rho(x).$$

Since $\rho x \in \partial_\lambda\left(\frac{\rho}{2}|x|^2\right)$, then the conclusion follows directly from (9). □

The equivalence between (5) and (9) provides an explicit expression for the approximate subdifferential [MS92] when the involved difference-of-convex function is weakly convex.

2.3.1. *Interpretation as approximate proximal subdifferential.* The proximal subgradient of h at x, is defined in [RW09, Definition 8.45] as some $g \in \mathbb{R}^N$ such that

$$\exists \rho, \delta > 0, h(z) \geq h(x) + \langle g, z-x\rangle - \frac{1}{2}\|z-x\|^2, \text{ for } z \in x + \mathbb{B}_{\delta^2}.$$

Weak convexity is equivalent to satisfaction of the following inequality, for all $y \in \mathbb{R}^N$,

$$h(y) + \frac{\rho}{2}|y-z|^2 \geq h(z) + \langle g, y-z\rangle$$

by any subgradient $g \in \partial h(z)$ and all $z \in \mathbb{R}^N$ with nonempty (Clarke's) subdifferential, see for instance [Ate+23, Prop. 2.2]. Because of this relation, Clarke subgradients of ρ-w.c. functions coincide with the proximal subgradients [RW09, Ch. 8I]. This remark, combined with Proposition 2.1(iv), justifies identifying $g \in \partial_\varepsilon^\rho h(z)$ with ε-proximal subgradients.

As a by product, our enlargement characterizes approximate optimality. Specifically, because a critical point $\bar{z}$ of a w.c. function h satisfies the inclusion $0 \in \partial h(\bar{z})$,

$$h(y) + \frac{\rho}{2}|y-\bar{z}|^2 \geq h(\bar{z}) \text{ for all } y \in \mathbb{R}^N.$$

For comparison, the approximate criticality condition $0 \in \partial_\varepsilon^\rho h(\bar{z})$ is equivalent to

$$h(y) + \frac{\rho}{2}|y-\bar{z}|^2 \geq h(\bar{z}) - \varepsilon \text{ for all } y \in \mathbb{R}^N.$$

Finally, the w.c. ε-subdifferential (5) can be interpreted in terms of the following ε-regular subdifferential, introduced in Proposition 10.46 in [RW09]:

$$\hat{\partial}_\varepsilon h(x) = \left\{g \in \mathbb{R}^N : h(z) \geq h(x) + \langle g, z-x\rangle - \varepsilon\|z-x\| + o(\|z-x\|)\right\},$$

where $\lim_{t\to 0}\frac{o(t)}{t} = 0$ (when $\varepsilon = 0$ this set is the regular subdifferential). More precisely, if $g \in \hat{\partial}_\varepsilon h(\bar{x})$, then for $z \in \bar{x} + \mathbb{B}_\delta$, g belongs to the set

$$\partial_{\varepsilon\delta}^\rho h(\bar{x}) := \left\{x^* \;: h(x) \geq h(\bar{x}) + \langle x^*, x-\bar{x}\rangle - \varepsilon\delta - \frac{\rho}{2}\|x-\bar{x}\|^2,\ \forall x \in \bar{x} + \mathbb{B}_\delta\right\}.$$

Conversely, if $g \in \partial_\varepsilon^\rho h(\bar{x})$, then for all $z \in \bar{x} + \mathbb{B}_1$, g belongs to a localized version of the ε-regular subdifferential, with $o(t) = -\frac{\rho}{2}t^2$.

2.4. **The Goldstein subdifferential and its enlargement by Bihain.** In [Gol77], another subdifferential enlargement is proposed for nonconvex functions. The Goldstein subdifferential of a function h at $x \in \mathbb{R}^N$ is

$$\partial_{\varepsilon}^{[\text{Gol77}]} h(x) = \operatorname{conv}\left(\bigcup_{z \in x + \mathbf{B}_{\varepsilon^2}} \partial h(z) \right).$$

Contrary to [MS92], the w.c. ε-subdifferential differs from the Goldstein subdifferential, even in the convex case. Indeed, the latter can be strictly contained in the former, as can be seen for the absolute value function example considered in § 2.1. Namely, by (6), for any given $\varepsilon \leq 2$ and $\rho \leq 2\varepsilon$,

$$\partial_{\varepsilon}^{[\text{Gol77}]} |\cdot|(1) = \{1\} \subsetneq [1 - (\varepsilon + \rho/2), 1 + \sqrt{2\rho\varepsilon}] = \partial_{\varepsilon}^{\rho} |\cdot|(1).$$

Along similar lines to (5), [Bih84] defines a (ε, η)-enlargement of the Goldstein subdifferential as follows:

$$\partial_{\varepsilon,\eta}^{[\text{Bih84}]} h(x) = \operatorname{conv}\left(\bigcup_{z \in x + \mathbf{B}_{\varepsilon^2}} \{\partial h(z) + \eta \partial(\|\cdot - x\|)(z)\} \right).$$

This enlargement contains the ε-subdifferential of w.c. functions, a result that follows from the variational principles stated in the next section (see the discussion after Lemma 3.3). Nevertheless, the enlargement of [Bih84] is, in general, larger than (5), even in the convex case. Taking again the absolute-value function, and $\varepsilon \in (0,1)$,

$$\partial_{\varepsilon,\varepsilon}^{[\text{Bih84}]} |\cdot|(1) = [1 - \varepsilon, 1 + \varepsilon] \supsetneq \partial_{\varepsilon}^{\rho} |\cdot|(1) = [1 - \varepsilon, 1],$$

by (6), written with $\rho = 0$. For a general inclusion, that holds for any w.c. function h, we refer to Corollary 3.1.1 below.

3. Variational principles

As shown in (4), there is an explicit relation between subgradients of the w.c. function and of its augmented convex counterpart. Thanks to this relation, we now show how many well-known variational principles for subgradients, stated for convex functions, can be transformed into similar relations that are valid for weakly convex functions.

3.1. **Ekeland's principle.** We state this fundamental result following the formulation in [Pen96, Proposition 1.1].

Theorem 3.1 (Variational principle)**.** *Given a closed proper ρ-w.c. function $h : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$, and constants $\alpha, \varepsilon > 0$, consider $z \in dom(h)$ and $g \in \partial_{\varepsilon}^{\rho} h(z)$. Then, there exists $z_{\varepsilon} \in \mathbb{R}^N$ and $g_{\varepsilon} \in \mathbb{R}^N$, such that*

$$g_{\varepsilon} - \rho z_{\varepsilon} \in \partial h(z_{\varepsilon}).$$

Furthermore, there exists $\gamma \in [-1,1]$, such that

$$\begin{aligned}
\|z_\varepsilon - z\| + \alpha|\langle g, z_\varepsilon - z\rangle| &\leq \sqrt{\varepsilon}\\
\|g_\varepsilon - (1+\alpha\gamma\sqrt{\varepsilon})g\| &\leq \sqrt{\varepsilon}\\
|\langle g_\varepsilon - g, z_\varepsilon - z\rangle| &\leq \varepsilon\\
|\langle g_\varepsilon, z_\varepsilon - z\rangle| &\leq \varepsilon + \alpha^{-1}\sqrt{\varepsilon}\\
|h(z_\varepsilon) - h(z)| &\leq \left(1+\frac{\rho}{2}\right)\varepsilon + \left(\alpha^{-1} + \frac{\rho}{2}\|z\|\right)\sqrt{\varepsilon}.
\end{aligned}$$

Proof. Apply [Pen96, Proposition 1.1] to $g + \rho z \in \partial_\varepsilon H^\rho(z)$, obtaining the existence of a pair $(z_\varepsilon, g_\varepsilon) \in \mathbb{R}^N \times \mathbb{R}^N$ that satisfies the first four estimates, $g_\varepsilon \in \partial H^\rho(z_\varepsilon)$, together with $|H^\rho(z_\varepsilon) - H^\rho(z)| \leq \varepsilon + \alpha^{-1}\sqrt{\varepsilon}$. Furthermore, in view of (3), it follows that $g_\varepsilon - \rho z_\varepsilon \in \partial h(z_\varepsilon)$. From the triangle inequality, it holds that

$$\begin{aligned}
|h(z_\varepsilon) - h(z)| - \frac{\rho}{2}\|z_\varepsilon - z\|^2 - \frac{\rho}{2}|\langle z, z_\varepsilon - z\rangle| &\leq \left| |h(z_\varepsilon) - h(z)| - \left|\frac{\rho}{2}\|z_\varepsilon - z\|^2 + \frac{\rho}{2}\langle z, z_\varepsilon - z\rangle\right| \right|\\
&\leq \left| h(z_\varepsilon) - h(z) + \frac{\rho}{2}\|z_\varepsilon - z\|^2 + \frac{\rho}{2}\langle z, z_\varepsilon - z\rangle \right|\\
&= \left| h(z_\varepsilon) - h(z) + \frac{\rho}{2}\|z_\varepsilon\|^2 - \frac{\rho}{2}\|z\|^2 \right|\\
&= |H^\rho(z_\varepsilon) - H^\rho(z)|\\
&\leq \varepsilon + \alpha^{-1}\sqrt{\varepsilon},
\end{aligned}$$

from which we deduce the fifth estimate. □

3.2. **Lipschitz continuity.** The ε-subdifferential for w.c. functions inherits the locally Lipschitz properties of the analogous object of convex analysis. This type of result can be traced back to [Nur78; Hir80], and represents one of the advantages of well-defined approximate subdifferentials. Indeed, they are not only upper-semicontinuous multifunctions akin to the usual subdifferentials, but they in fact exhibit further stronger continuity properties. In the next result, Δ denotes the Hausdorff distance between two sets. Naturally, the Lipschitz constant with respect to the variable includes the effect of the weak convexity parameter ρ.

Theorem 3.1 (The w.c. ε-subdifferential is a Lipschitzian multifunction)**.** *Let $h : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ be a proper lsc ρ-w.c. function, and $K \subseteq \mathbb{R}^N$ a compact set. Then, there exists a constant $M > 0$ such that, for all $z, z' \in K$ and $\varepsilon, \varepsilon' > 0$,*

$$\Delta\left(\partial_\varepsilon^\rho h(z), \partial_{\varepsilon'}^\rho h(z')\right) \leq \left(\frac{M}{\min\{\varepsilon, \varepsilon'\}} + \rho\right)\|z - z'\| + \frac{M}{\min\{\varepsilon, \varepsilon'\}}|\varepsilon - \varepsilon'|.$$

Proof. Since the augmented function H^ρ is convex and Lipschitz continuous over the compact set K, apply [Hir80, Theorem 3.2] to obtain

$$\Delta\left(\partial_\varepsilon H^\rho(z), \partial_{\varepsilon'} H^\rho(z')\right) \leq \frac{M}{\min\{\varepsilon, \varepsilon'\}}\left(\|z - z'\| + |\varepsilon - \varepsilon'|\right).$$

for some constant $M > 0$. Moreover, the triangle inequality yields

$$\|g' - g\| \leq \|g' + \rho z' - (g + \rho z)\| + \rho\|z - z'\|.$$

Therefore, in view of (5), it holds that

$$\begin{aligned}
&\sup\{\operatorname{dist}\big(g', \partial_\varepsilon^\rho h(z)\big) : g' \in \partial_{\varepsilon'}^\rho h(z')\} \\
&\le \sup\{\operatorname{dist}\big(g' + \rho z', \partial_\varepsilon^\rho h(z) + \rho z\big) : g' \in \partial_{\varepsilon'}^\rho h(z')\} + \rho\|z - z'\| \\
&= \sup\{\operatorname{dist}\big(G', \partial_\varepsilon H^\rho(z)\big) : G' \in \partial_{\varepsilon'} H^\rho(z')\} + \rho\|z - z'\|.
\end{aligned}$$

Therefore, from the definition of Hausdorff distance,

$$\Delta\big(\partial_\varepsilon^\rho h(z), \partial_{\varepsilon'}^\rho h(z')\big) \le \Delta\big(\partial_\varepsilon H^\rho(z), \partial_{\varepsilon'} H^\rho(z')\big) + \rho\|z - z'\|,$$

from which the result readily follows. □

3.3. **Transportation of subgradients of w.c. functions.** The relation (2) expresses subgradients in one point as approximate subdifferential at another point. The inverse relation is given by the following well-known result.

Theorem 3.2 (Brøndsted-Rockafellar)**.** *Let us be given a ρ-w.c. function $h : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$, a point $z \in dom(h)$ and a scalar $\varepsilon \ge 0$. For any $\eta > 0$ and $s \in \partial_\varepsilon^\rho h(z)$, there exist z_η and $s_\eta \in \partial h(z_\eta)$ such that*

$$\|z - z_\eta\| \le \eta \quad \textit{and} \quad \|s - s_\eta\| \le \frac{\varepsilon}{\eta} + \rho\eta\,.$$

Proof. By (5), $g = s + \rho z \in \partial_\varepsilon H^\rho(z)$, and by [HL96, Thm. XI.4.2.1], there exist z_η and $g_\eta \in \partial H^\rho(z_\eta)$ satisfying

$$\|z - z_\eta\| \le \eta \quad \text{and} \quad \|g - g_\eta\| \le \varepsilon/\eta\,.$$

By (4), $g_\eta = s_\eta + \rho z_\eta$ for some $s_\eta \in \partial h(z_\eta)$. The result follows, since

$$\begin{aligned}
\|s - s_\eta\| &= \|g - \rho z - g_\eta + \rho z_\eta\| \\
&\le \|g - g_\eta\| + \rho\|z - z_\eta\| \\
&\le \varepsilon/\eta + \rho\eta
\end{aligned}$$

□

On a similar note, [Rob99] explores a perturbation result for convex functions, where the perturbation has an explicit form. We omit the proof, as it follows by applying Robinson's result to the augmented function H^ρ defined in (3).

Lemma 3.3 (Extension of Theorem 2 in [Rob99])**.** *Let be given a ρ-w.c. function $h : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$, a point $z \in dom(h)$ and a scalar $\varepsilon \ge 0$. For any $\gamma > 0$ and $s \in \partial_\varepsilon^\rho h(z)$, there is a unique perturbation $p = p(\gamma)$ such that*

$$s - \big(\rho\gamma + \frac{1}{\gamma}\big)p \in \partial h(z + \gamma p)\,, \quad \|p\| \le \sqrt{\varepsilon}\,. \quad \square$$

This lemma will prove useful to show the linear rate of convergence of the ε-descent scheme introduced in the next section. The explicit form of the perturbation in Lemma 3.3 can be exploited to relate (5) to the Bihain approximate subdifferential in § 2.4.

Corollary 3.1.1. *For any ρ-w.c. function h and $z \in \mathbb{R}^N$ the approximate w.c. subdifferential (5) is contained in its Bihain's counterpart:*

$$\partial_\varepsilon^\rho h(z) \subseteq \partial_{\varepsilon,\rho+\frac{1}{\varepsilon}}^{[Bih84]} h(z)\,.$$

Proof. Given $s \in \partial_\varepsilon^\rho h(z)$, in view of Lemma 3.3 with $\gamma = \sqrt{\varepsilon}$, there exists $p(\varepsilon)$ such that $\|p(\varepsilon)\| \leq \varepsilon$, and $s - (\rho\sqrt{\varepsilon} + \sqrt{\varepsilon}^{-1})p(\varepsilon) \in \partial(z + \sqrt{\varepsilon}p(\varepsilon))$. Define $z(\varepsilon) = z + \sqrt{\varepsilon}p(\varepsilon)$, so that $\|z(\varepsilon) - z\| \leq \varepsilon$ and $s - (\rho + \varepsilon^{-1})(z(\varepsilon) - z) \in \partial h(z(\varepsilon))$. The result follows, since in any case, $z(\varepsilon) - z \in \partial(\|\cdot - x\|)(z(\varepsilon))$. □

4. Numerical consequences for ε-descent patterns

We are interested in being able to numerically identify and compute a critical point of the mapping h through an iterative procedure. Having a continuous subdifferential multifunction like (5) opens the way to state a minimal set of conditions ensuring convergence in such a setting. To this end, we will consider, akin to [Rob99], but for w.c. functions, how to put in place $\varepsilon-$subgradient descent methods that are convergent; see also [Ate+23].

Our algorithmic pattern defines a sequence $\{x_n\}_{n\geq 0}$ such that any of its cluster points, should there be one, is critical for h. Under additional assumptions, convergence of the sequence itself can be ensured as well. To this aim, the following two minimal assumptions must be satisfied:

- The next iterate is computed using an approximate subgradient as direction:

$$x_{n+1} = x_n - t_n d_n, \quad \text{for } d_n \in \partial_{\varepsilon_n}^\rho h(x_n) \text{ and } t_n \in [t_{\mathtt{low}}, t^{\mathtt{up}}]\,. \tag{10}$$

- The next iterate provides sufficient descent on the objective function as follows:

$$h(x_{n+1}) + \frac{\rho}{2}\|x_{n+1} - x_n\|^2 \leq h(x_n) + m(\langle d_n, x_{n+1} - x_n\rangle - \varepsilon_n), \tag{11}$$

  where $m \in (0,1)$ is an Armijo-like parameter.

Two concrete instances of this abstract algorithmic pattern are:

(1) The proximal point method for a shifted augmented function as in (7), that varies along iterations: given x_n, define the next iterate as $x_{n+1} = (I + t_n \partial H_{x_n}^\rho)^{-1}(x_n)$, that is, perform a proximal step on $H_{x_n}^\rho$ at x_n. Concretely this amounts to:

$$x_{n+1} \in \operatorname{argmin}\left\{H_{x_n}^\rho(w) + \frac{1}{2t_n}\|w - x_n\|^2\right\}.$$

In other words, one needs to compute the unique minimizer of a strictly convex function. The sequence generated this way satisfies (10) and (11) for $m = 1$, and $\varepsilon_n = H_{x_n}^\rho(x_n) - H_{x_n}^\rho(x_{n+1}) - \langle d_n, x_n - x_{n+1}\rangle$, in view of Proposition 2.1(v).

(2) The redistributed bundle method [HS10]: which amounts to applying a bundle method to varying shifted functions $H_{x_n}^{\rho_n}$, for ρ_n an estimate of the w.c. parameter ρ, assumed

unknown. The serious step sequence of the redistributed bundle method satisfies (10) and (11) for $\rho = m\eta_{\min}$, where $\eta_{\min} > 0$ is a lower bound, we refer to [HS10] for details.

4.1. **Convergence of the algorithmic pattern along a subsequence.** We now show that, with our minimal set of assumptions, if the sequence has accumulation points, then it must cluster at a critical point.

Throughout this section, for any given $x \in \mathbb{R}^N$, $e^h_\rho(x)$ denotes the Moreau envelope of h at x with parameter ρ: $e^h_\rho(x) = \inf_y\{h(y) + \frac{\rho}{2}\|y-x\|^2\}$.

Theorem 4.1. *Suppose that $h : \mathbb{R}^N \to \mathbb{R}\cup\{+\infty\}$ is a proper lsc $\rho-$w.c. function, such that* $\inf h > -\infty$. *For any sequence $\{x_n\}$ generated by* (10) *satisfying* (11)*, it holds that:*

(1) *$\{h(x_n)\}$ is a nonincreasing sequence, and $h(x_n) \to \tilde{h} \in \mathbb{R}$, as $n \to +\infty$.*
(2) *$x_{n+1} - x_n \to 0$, $\varepsilon_n \to 0$, $t_n\|d_n\|^2 \to 0$, and $d_n \to 0$.*
(3) *$\tilde{h} = \limsup_n e^h_\rho(x_n)$ and*

$$\liminf_n \left[H^\rho_{x_n}(x_n) - \inf_y H^\rho_{x_n}(y) \right] = 0.$$

Proof. From (10) and (11), it follows for all n,

$$(12)\qquad \begin{aligned} h(x_{n+1}) - h(x_n) &\leq m(\langle d_n, x_{n+1}-x_n\rangle - \varepsilon_n) - \frac{\rho}{2}\|x_{n+1}-x_n\|^2 \\ &= -m(t_n\|d_n\|^2 + \varepsilon_n) - \frac{\rho}{2}\|x_{n+1}-x_n\|^2 \\ &\leq 0, \end{aligned}$$

hence $h(x_{n+1}) \leq h(x_n)$. Then $\{h(x_n)\}$ is a nonincreasing sequence bounded below by $\inf h$, and thus convergent to some $\tilde{h}$, proving item (1). Furthermore, a rearrangement of the previous estimate gives, for any $n > 1$,

$$\begin{aligned} \sum_{n=0}^{k-1} m(t_n\|d_n\|^2 + \varepsilon_n) &\leq \sum_{n=0}^{k-1} h(x_n) - h(x_{n+1}) - \frac{\rho}{2}\|x_{n+1}-x_n\|^2 \\ &= h(x_0) - h(x_k) - \frac{\rho}{2}\sum_{n=0}^{k-1}\|x_{n+1}-x_n\|^2 \end{aligned}$$

Therefore,

$$\sum_{n=0}^{k-1} m(t_n\|d_n\|^2 + \varepsilon_n) + \frac{\rho}{2}\sum_{n=0}^{k-1}\|x_{n+1}-x_n\|^2 \leq h(x_0) - h(x_k) \leq h(x_0) - \inf h.$$

Taking the limit as $k \to +\infty$, it follows that

$$\sum_{n=0}^{+\infty}(t_n\|d_n\|^2 + \varepsilon_n) < +\infty, \quad \sum_{n=0}^{+\infty}\|x_{n+1}-x_n\|^2 < +\infty,$$

and, in particular, $t_n\|d_n\|^2 \to 0$, $\varepsilon_n \to 0$, and $x_{n+1} - x_n \to 0$, as $n \to +\infty$. Since $t_n \geq t_{low}$, then $d_n \to 0$, which proves item (2). To prove the first part of item (3), note that $H^\rho_{x_n}(x_n) = h(x_n) \to \tilde{h}$, and thus

$$\begin{array}{rcl} 0 & = & \liminf_n \left[H^\rho_{x_n}(x_n) - \inf_y H^\rho_{x_n}(y)\right] \\ & = & \liminf_n \left[h(x_n) - \inf_y H^\rho_{x_n}(y)\right] \\ & = & \tilde{h} - \limsup_n \left[\inf_y H^\rho_{x_n}(y)\right] \\ & = & \tilde{h} - \limsup_n \left[\inf_y h(y) + \dfrac{\rho}{2}\|y - x_n\|^2\right] \\ & = & \tilde{h} - \limsup_n e^h_\rho(x_n). \end{array}$$

Next, following the arguments of [CL93, Proposition 1.2], by way of contradiction, suppose the second equality of item (3) is not valid, that is, assume there exists $\delta > 0$, and $n_0 \geq 1$, such that for all $n \geq n_0$,

$$\inf_y H^\rho_{x_n}(y) < H^\rho_{x_n}(x_n) - \delta.$$

Therefore, there exists $y \in \mathbb{R}^N$, such that $H^\rho_{x_n}(y) \leq H^\rho_{x_n}(x_n) - \delta$. From [CL93, Lemma 1.1] applied to $f = H^\rho_{x_n}$, it holds that

$$\|x_{n+1} - y\|^2 \leq t_n\big(t_n\|d_n\|^2 + 2\varepsilon_n\big) + 2t_n\big(H^\rho_{x_n}(y) - H^\rho_{x_n}(x_n)\big) + \|x_n - y\|^2.$$

From item (2), for all sufficiently large n, and without loss of generality, for $n \geq n_0$, $t_n\|d_n\|^2 + 2\varepsilon_n < \delta$. Thus,

$$\|x_{n+1} - y\|^2 \leq t_n\delta - 2t_n\delta + \|x_n - y\|^2 \leq -t_n\delta + \|x_n - y\|^2.$$

Summing over $n = n_0, \ldots, k-1$, it yields

$$\|x_k - y\|^2 - \|x_{n_0} - y\|^2 = \sum_{n=n_0}^{k-1} \|x_{n+1} - y\|^2 - \|x_n - y\|^2 \leq -\delta \sum_{n=n_0}^{k-1} t_n.$$

Since $\{t_n\}$ is bounded from below, then $\sum_{n=n_0}^{+\infty} t_n = +\infty$. In this way, taking the limit as $k \to +\infty$ in

$$0 \leq \|x_k - y\|^2 \leq \|x_{n_0} - y\|^2 - \delta \sum_{n=n_0}^{k-1} t_n,$$

gives a contradiction. □

Without further regularity assumptions, if $\{x_n\}$ is bounded, the sequence subsequentially converges to critical points, as the following result shows.

Proposition 4.2 (Subsequential convergence)**.** *Suppose that the proper lsc ρ-w.c. function $h : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ satisfies $\inf h > -\infty$. Then, any cluster point $\bar{x}$ of the sequence $\{x_n\}$ satisfying* (10)–(11)*, if any, is a critical point of h, that is, $0 \in \partial h(\bar{x})$. Moreover, $\{h(x_n)\}$ converges to the critival value $h(\bar{x})$.*

Proof. In view of Lemma 3.3 and $d_n \in \partial^\rho_{\varepsilon_n} h(x_n)$, there exists $\{p_n\}$ such that

$$d_n - (\rho\gamma + \gamma^{-1})p_n \in \partial h(x_n + \gamma p_n), \text{ and } \|p_n\| \leq \sqrt{\varepsilon_n}. \tag{13}$$

Let $\{x_{n_j}\}$ be a subsequence of $\{x_n\}$ such that $x_{n_j} \to \bar{x}$ as $j \to +\infty$. Taking the limit in (13) throughout the subsequence, Theorem 4.1(2) yields $0 \in \partial h(\bar{x})$. Since h is locally Lipschitzian, then h is continuous around $\bar{x}$. Thus, $h(x_{n_j}) \to h(\bar{x})$ as $j \to +\infty$, and monotonicity of $\{h(x_n)\}$ (Theorem 4.1(1)) implies $h(x_n) \to h(\bar{x})$ as $k \to +\infty$. □

4.2. **Convergence of the algorithmic pattern.** In order to generate a sequence converging globally to a critical point, we need further regularity assumptions. One classical approach is to assume an error-bound for the distance to the set of critical points [Ate+23; LT93]. First, we check that any sequence $\{x_n\}$ generated to satisfy (10)–(11), originates from a descent method in the sense of [Ate+23], and then establish global convergence and local rates of convergence.

Proposition 4.3. *Suppose $h : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ is a proper lsc ρ−w.c. function. Given a sequence $\{x_n\}$ conforming to (10)–(11), there exists a sequence $\{z_n\}$ such that the following statements hold true.*

(1) *There exists $a > 0$ such that for all n,*
$$h(x_{n+1}) + a\big(\|x_{n+1} - x_n\|^2 + \varepsilon_n\big) \le h(x_n).$$
(2) *There exists $b > 0$ and $\{g_n \in \partial h(z_n)\}$, such that for all n,*
$$\|g_n\| \le b\big(\|x_{n+1} - x_n\| + \|x_n - z_n\|\big).$$
(3) *The sequence $\{x_n - z_n\}$ converges to 0.*

Proof. In view of (10), $d_n = \frac{1}{t_n}(x_n - x_{n+1})$. Substitute this identity in (11) to obtain

$$h(x_{n+1}) + \left(\frac{\rho}{2} + \frac{m}{t_n}\right)\|x_{n+1} - x_n\|^2 + m\varepsilon_n \le h(x_n). \tag{14}$$

Taking $a = \min\left\{\frac{\rho}{2} + \frac{m}{t_{\max}}, m\right\}$ yields item (1). Moreover, due to Lemma 3.3, (13) holds. Denoting $g_n = d_n - (\rho\gamma + \gamma^{-1})p_n$, and $z_n = x_n + \gamma p_n$, then $g_n \in \partial h(z_n)$. Since $\varepsilon_n \to 0$, then item (3) follows. Furthermore, for all n,

$$\begin{aligned}\|g_n\| &\le \|d_n\| + \left(\rho\gamma + \frac{1}{\gamma}\right)\|p_n\| \\ &= \frac{1}{t_n}\|x_{n+1} - x_n\| + \frac{1}{\gamma}\left(\rho + \frac{1}{\gamma}\right)\|x_n - z_n\|.\end{aligned}$$

Taking $b = \min\left\{\frac{1}{t_{\min}}, \frac{1}{\gamma}\left(\rho + \frac{1}{\gamma}\right)\right\}$ yields item (2). □

The regularity assumptions ensuring convergence are given below, denoting by $S = (\partial h)^{-1}(0)$ the set of critical points of h.

– h satisfies a subdifferential-based error bound if for any $\bar{h} \geq \inf h$, and $x \in \mathbb{R}^N$ such that $h(x) \leq \bar{h}$, whenever $g \in \partial h(x) \cap \mathbb{B}_\delta$ for some $\delta > 0$, there exists some $\ell > 0$ such that

$$\operatorname{dist}(x, S) \leq \ell \|g\| . \tag{15}$$

– h satisfies the proper separation of isocost surfaces property if there exists $\varepsilon > 0$ such that

$$(\forall x, y \in S)\ \|x - y\| \leq \varepsilon \implies f(x) = f(y) . \tag{16}$$

In the convex case, S is the set of global minimizers, and thus (16) holds trivially. In view of [LT93, Theorem 2.1], (15) holds, for instance, in the strongly convex case, and in dual problems of strongly convex functions with Lipschitz continuous gradient. In a general setting, conditions (15) and (16) imply satisfaction of the Kurdyka-Łojasiewicz inequality [LP18, Theorem 4.1].

Next, we present the main convergence result of this section, assuming the subdifferential-based error bound (15). These convergence results can be derived following [Ate+23, Theorem 4.3]. In order to guarantee boundedness of the iterates, we assume that the function h is level-bounded, namely, for any $\alpha \geq \inf h$, the level set $\{x \in \mathbb{R}^N : h(x) \leq \alpha\}$ is bounded.

Theorem 4.4 (Global convergence and local linear rate)**.** *Suppose $h : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ is a proper lsc ρ−w.c. level-bounded function bounded from below. In addition, assume h satisfies* (15) *and* (16)*. Then, there exists a critical point $\bar{x}$ of h, such that $x_n \to \bar{x}$ and $h(x_n) \downarrow h(\bar{x})$ as $n \to +\infty$. Moreover, there exist $r \in (0,1)$ and $c > 0$, such that for all sufficiently large n,*

$$\|x_n - \bar{x}\| \leq c r^n, \text{ and } h(x_{n+1}) - h(\bar{x}) \leq r\big(h(x_n) - h(\bar{x})\big).$$

Proof. First, level-boundedness implies that $\{x_n\}$ is bounded, since $\{x_n\} \subseteq \{x \in \mathbb{R}^d : h(x) \leq h(x^0)\}$. Furthermore, from the definition in (13) of $\{p_n\}$,

$$\|x_n - z_n\|^2 = \gamma^2 \|p_n\|^2 \leq \gamma^2 \varepsilon_n.$$

For all n, denote $\varphi_n = h(x_n) - h(\bar{x})$. In view of (12),

$$m\varepsilon_n \leq h(x_n) - h(x_{n+1}) = \varphi_n - \varphi_{n+1}. \tag{17}$$

Then, combining the last two estimates, it follows

$$\|x_n - z_n\|^2 \leq \frac{\gamma^2}{m}(\varphi_n - \varphi_{n+1}). \tag{18}$$

Moreover, since $d_n \in \partial^\rho_{\varepsilon_n} h(x_n)$, Proposition 2.1(iv) implies for all n,

$$h(z_n) + \frac{\rho}{2}\|z_n - x_n\|^2 \geq h(x_n) + \langle d_n, z_n - x_n \rangle - \varepsilon_n.$$

Rearranging terms and using the Cauchy-Schwarz inequality, we obtain

$$\begin{aligned}
h(x_n) - h(z_n) &\leq \frac{\rho}{2}\|x_n - z_n\|^2 + \|d_n\|\|x_n - z_n\| + \varepsilon_n \\
&\leq \frac{\rho\gamma^2}{2m}(\varphi_n - \varphi_{n+1}) + \gamma\|d_n\|\sqrt{\frac{1}{m}(\varphi_n - \varphi_{n+1})} + \frac{1}{m}(\varphi_n - \varphi_{n+1}) \\
&\leq \left(\frac{\rho\gamma^2}{2m} + \frac{\gamma}{t_{\min}}\sqrt{\frac{1}{m}\left(\frac{\rho}{2} + \frac{m}{t_{\max}}\right)^{-1}} + \frac{1}{m}\right)(\varphi_n - \varphi_{n+1}),
\end{aligned}$$

where in the second estimate we use (17) and (18), and the third inequality is a consequence of (10) and (14):

$$\begin{aligned}
\|d_n\| &= \frac{1}{t_n}\|x_n - x_{n+1}\| \\
&\leq \frac{1}{t_{\min}}\sqrt{\left(\frac{\rho}{2} + \frac{m}{t_{\max}}\right)^{-1}(\varphi_n - \varphi_{n+1})}.
\end{aligned}$$

In this manner, [Ate+23, Theorem 4.3] applies and the result follows. □

The convergence results ensure that both d_n and ε_n are driven to zero asymptotically. From a numerical point of view, this yields approximate optimality certificates that can be easily checked in practice.

5. Interpreting the subdifferential in light of D.C. approaches

Much like as in Section 4, we are interested in solving the problem

$$\min_{x \in X} h(x), \tag{19}$$

where $X \subseteq \mathbb{R}^N$ is a closed set, and $h : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ is a ρ-w.c. function. At the very least we would like to be able to identify a critical point of the previous problem. Recall that h is a particular D.C. function:

$$h(x) = H^\rho(x) - \frac{\rho}{2}\|x\|^2, \text{ and even } h(x) = H_z^\rho(x) - \frac{\rho}{2}\|x - z\|^2,$$

where any localizer $z \in \mathbb{R}^N$ could be employed to shift the augmented function. A standard method to minimize D.C. functions linearizes the second component function around the current iterate and updates iterates by solving the resulting convex problem. This approach does not fit the scheme proposed in Section 4. We thus consider the extended ε-subgradient scheme in [AO20].

5.1. **Computing critical points.** Following [AO20], a D.C.-like algorithm can be employed to compute critical points of h in the following sense:

$$0 \in \partial h(\bar{x}) + \mathrm{N}_X(\bar{x}), \tag{20}$$

where the latter normal cone could be the Mordukhovich one if X was not assumed to be convex.

More specifically, given $z \in \mathbb{R}^N$, x_{n+1} is defined as an ε_n-solution to

$$\min_{x \in X} H_z^\rho(x) - \frac{\rho}{2}\|x_n - z\|^2 - \rho\,\langle x_n - z, x - x_n\rangle\,.$$

This request amounts to identifying x_{n+1} such that

$$0 \in \partial_{\varepsilon_n} H_z^\rho(x_{n+1}) - \rho(x_n - z) + \mathrm{N}_X(x_{n+1}).$$

In view of (8), the latter inclusion becomes

$$0 \in \partial^\rho_{\varepsilon_n} h(x_{n+1}) + \mathrm{N}_X(x_{n+1}) + \rho\,(x_{n+1} - x_n)\,. \tag{21}$$

With $d_n = \rho(x_n - x_{n+1})$, assuming for simplicity $\mathrm{N}_X(x_{n+1}) = \{0\}$, (e.g., when $X = \mathbb{R}^N$), the just given condition reads $d_n \in \partial^\rho_{\varepsilon_n} h(x_{n+1})$. In light of (10), the descent scheme can be seen as implicit rather than explicit. Furthermore, still in absence of the normal cone, transporting $d_n \in \partial^\rho_{\varepsilon_n} h(x_{n+1})$ results in the alternative inclusion $0 \in \partial^\rho_{\varepsilon'_n} h(x_n)$, for

$$\varepsilon'_n = \varepsilon_n + h(x_{n+1}) - h(x_n) + \frac{\rho}{2}\|x_{n+1}\|^2 - \frac{\rho}{2}\|x_n\|^2 - \rho\,\|x_{n+1} - x_n\|^2\,.$$

It turns out that this D.C. approach converges. As shown in Theorem 2 [AO20], and whenever the sequence $\{x_n\}$ admits accumulation points, driving $\varepsilon_n \to 0$ suffices for the limit to be a critical point. Then, the cluster point $\bar{x}$ satisfies the inclusion

$$\rho(\bar{x} - z) \in \partial H_z^\rho(\bar{x}) + \mathrm{N}_X(\bar{x}),$$

i.e.,

$$\rho(\bar{x} - z) \in \partial h(\bar{x}) + \rho(\bar{x} - z) + \mathrm{N}_X(\bar{x}), \tag{22}$$

yielding (20). Here (22) can be seen to be exactly [BL10, eq. (9)], thus indicating local optimality of $\bar{x}$.

The typical D.C. stopping condition involves $\|x_{n+1} - x_n\|$ being small. If this is indeed so, it is clear how (21) is an approximate optimality condition involving the localized and perturbed subdifferential. The ideal case $x_{n+1} = x_n$ also offers immediate interpretation.

5.2. **Approximate optimality conditions.** When $f : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ is a convex function, $0 \in \partial_\varepsilon f(\bar{x})$ is a sufficient and necessary condition for $\bar{x}$ being a global $\varepsilon-$minimizer of f. In this section, we discuss an analogous condition for $\bar{x}$ being a local minimizer of h. In this case, in particular, $\bar{x}$ is also a critical point: $0 \in \partial h(\bar{x})$. The following result shows a necessary condition for local minimizers.

Proposition 5.1 (Characterization of local minimizers)**.** *Let $h:\mathbb{R}^N\to\mathbb{R}\cup\{+\infty\}$ be a ρ-w.c. function. Suppose that for $\bar{x}\in\mathbb{R}^N$, there exists $\eta>0$ such that $h(\bar{x})\le h(x)$ for all $x\in\bar{x}+\mathbb{B}_{\eta^2}$. Then, for any $\varepsilon>0$,*

$$\mathbb{B}_{2\rho\varepsilon}\subseteq\partial_\varepsilon^\rho h(\bar{x})+\mathbb{B}_{\varepsilon^2/\eta^2} \tag{23}$$

In addition, there exists $\varepsilon'>0$, such that

$$\mathbb{B}_{2\rho\varepsilon}\subseteq\partial_\varepsilon^\rho h(\bar{x})+\partial_{\varepsilon'}\left(\frac{\rho}{2}\|\cdot\|^2\right)(\bar{x}).$$

Proof. First, in view of [HL96, Ch. XI, Example 1.2.2]

$$\partial_\varepsilon\left(\frac{\rho}{2}\|\cdot\|^2\right)(x)=\rho x+\mathbb{B}_{2\rho\varepsilon}.$$

Since $h(x)=H^\rho(x)-\frac{\rho}{2}\|x\|^2$, [BL10, Corollary 4.4] yields for all $\varepsilon>0$:

$$\rho\bar{x}+\mathbb{B}_{2\rho\varepsilon}\subseteq\partial_\varepsilon H^\rho(\bar{x})+\mathbb{B}_{\varepsilon^2/\eta^2}.$$

This last formula can be translated with the help of (5) as $\{0\}+\mathbb{B}_{2\rho\varepsilon}\subseteq\partial_\varepsilon^\rho h(\bar{x})+\mathbb{B}_{\varepsilon^2/\eta^2}$, yielding (23). Using [HL96, Ex.XI.1.2.2] again, $\mathbb{B}_{\varepsilon^2/\eta^2}=\partial_{\bar{\varepsilon}}\left(\frac{\bar{\rho}}{2}\|\cdot\|^2\right)(0)$, for any $\bar{\varepsilon},\bar{\rho}>0$ such that $2\bar{\varepsilon}\bar{\rho}=\frac{\varepsilon^2}{\eta^2}$. Choose $\bar{\rho}=\rho$ and $\bar{\varepsilon}=\frac{\varepsilon^2}{2\rho\eta^2}$, and take any $g\in\partial_{\bar{\varepsilon}}\left(\frac{\rho}{2}\|\cdot\|^2\right)(0)$. By transporting the subgradient to $\bar{x}$, we obtain for all $x\in\mathbb{R}^N$

$$\begin{array}{rcl}\frac{\rho}{2}\|x\|^2 & \ge & \langle g,x\rangle-\bar{\varepsilon}\\ & = & \frac{\rho}{2}\|\bar{x}\|^2+\langle g,x-\bar{x}\rangle-\left(\frac{\rho}{2}\|\bar{x}\|^2+\bar{\varepsilon}-\langle g,\bar{x}\rangle\right).\end{array}$$

Since $g\in\mathbb{B}_{\varepsilon^2/\eta^2}$, then

$$\frac{\rho}{2}\|x\|^2 \quad\ge\quad \frac{\rho}{2}\|\bar{x}\|^2+\langle g,x-\bar{x}\rangle-\left(\frac{\rho}{2}\|\bar{x}\|^2+\bar{\varepsilon}+\frac{\varepsilon}{\eta}\|\bar{x}\|\right).$$

Hence, $\mathbb{B}_{\varepsilon^2/\eta^2}=\partial_{\varepsilon'}\left(\frac{\rho}{2}\|\cdot\|^2\right)(\bar{x})$, where $\varepsilon'=\bar{\varepsilon}+\frac{\rho}{2}\|\bar{x}\|^2+\frac{\varepsilon}{\eta}\|\bar{x}\|$, and the conclusion follows. ☐

Corollary 5.1.1. *Let $h:\mathbb{R}^N\to\mathbb{R}\cup\{+\infty\}$ be a ρ-w.c. function. Suppose that $\bar{x}\in\mathbb{R}^N$ is a local minimum for h. Then there exists $\rho'\ge\rho$ such that*

$$0\in\partial_\varepsilon^{\rho'}h(\bar{x})$$

for any $\varepsilon\ge0$.

Proof. Indeed, in (23), if $\sqrt{2\rho}>\frac{\sqrt{\varepsilon}}{\eta}$, we can decrease $\eta>0$ without loss of generality so that

$$\sqrt{2\rho\varepsilon}=\frac{\varepsilon}{\eta}.$$

If $\sqrt{2\rho}<\frac{\sqrt{\varepsilon}}{\eta}$, we can increase the parameter of weak convexity $\rho>0$ to obtain the same equality, while keeping convexity of H^ρ. Both two balls in (23) are equal, and in view of the Minkowski-Rådström-Hörmander Theorem, e.g., [PU02, Corollary 3.2.2 (ii)], (23) implies $0\in\partial_\varepsilon^\rho h(\bar{x})$ for $\varepsilon>0$. For the trivial case $\varepsilon=0$, any local minimizer is, in particular, a critical point. ☐

The last result offers a nice expression of a necessary condition for determining local optimality, *directly* expressed in terms of the original object h. Usually, global and local optimality conditions for D.C. functions involve the explicit D.C. decomposition. We refer to [Oli20] Theorems 1 and 2 for further information. The interesting difference here is that our local result holds for all $\varepsilon > 0$, whereas typically in D.C. optimality conditions, the full inclusion of the ε-subdifferential of the second component in the first for all $\varepsilon \geq 0$ would entail global optimality. Local optimality in turn would have such an inclusion for some largest $\bar{\varepsilon} \geq 0$.

Concluding remarks. We have proposed a notion of ε-subgradient for weakly convex functions that enjoys favourable continuity properties and full calculus. We have analysed the relationship between this suggestion and alternative ones, and provided variational principles involving this subdifferential. Based on those approximate subgradients, we have introduced an algorithmic pattern of descent that builds a sequence converging to a critical point under mild assumptions. Thanks to the variational principles, subgradients at one point can be transported and expressed as approximate subgradients at another point. By this token, first-order methods for weakly convex minimization can be cast into our algorithmic pattern. Our work gives a unifying perspective on ε-subdifferentials for weakly convex functions and their utility in numerical implementations.

Data statement: there is no data associated with this submission.

REFERENCES

[AO20] W. van Ackooij and W. de Oliveira. “Some brief observations in minimizing the sum of locally Lipschitzian functions”. In: *Optimization Letters* 14.3 (2020), pp. 509–520.

[Ate+23] F. Atenas, C. Sagastizábal, P. J. S. Silva, and M. Solodov. “A Unified Analysis of Descent Sequences in Weakly Convex Optimization, Including Convergence Rates for Bundle Methods”. In: *SIAM Journal on Optimization* 33.1 (2023), pp. 89–115.

[Bih84] A. Bihain. “Optimization of upper semidifferentiable functions”. In: *Journal of Optimization Theory and Applications* 44 (1984), pp. 545–568.

[BL10] I. Bomze and C. Lemaréchal. “Necessary conditions for Local Optimality in Difference of-Convex Programming”. In: *Journal of Convex Analysis* 17.2 (2010), pp. 673–680.

[CL93] R. Correa and C. Lemaréchal. “Convergence of some algorithms for convex minimization”. In: *Mathematical Programming* 62 (1993), pp. 261–275.

[DJ97] V. F. Demyanov and V. Jeyakumar. “Hunting for a smaller convex subdifferential”. In: *Journal of Global Optimization* 10 (1997), pp. 305–326.

[DLZ85] V. Demyanov, C. Lemaréchal, and J. Zowe. “Attempts to approximate a set-valued mapping”. In: *Nondifferentiable Optimization: Motivations and Applications: Proceedings of an IIASA (International Institute for Applied Systems Analysis) Workshop on Nondifferentiable Optimization Held at Sopron, Hungary, September 17-22, 1984*. Springer. 1985, pp. 3–7.

[Gol77] A. A. Goldstein. “Optimization of Lipschitz continuous functions”. In: *Mathematical Programming* 13.1 (Dec. 1977), pp. 14–22.

[Hir80] J.-B. Hiriart-Urruty. "Lipschitz r-continuity of the approximate subdifferential of a convex function". In: *Mathematica Scandinavica* 47.1 (1980), pp. 123–134.

[HL96] J. Hiriart-Urruty and C. Lemaréchal. *Convex Analysis and Minimization Algorithms, vol. II*. 2nd. Grundlehren der mathematischen Wissenschaften 306. Berlin, Heidelberg: Springer-Verlag Berlin Heidelberg, 1996, p. 347.

[HS10] W. Hare and C. Sagastizábal. "A redistributed proximal bundle method for nonconvex optimization". In: *SIAM Journal on Optimization* 20.5 (2010), pp. 2442–2473.

[Kru03] A. Y. Kruger. "On Fréchet subdifferentials". In: *J. Math Sci.* 116.3 (2003), pp. 3325–3358.

[Kru85] A. Y. Kruger. "Properties of generalized differentials". In: *Siberian Mathematical Journal* 26.6 (1985), pp. 822–832.

[LP18] G. Li and T. K. Pong. "Calculus of the exponent of Kurdyka-Łojasiewicz inequality and its applications to linear convergence of first-order methods". In: *Foundations of computational mathematics* 18.5 (2018), pp. 1199–1232.

[LT93] Z.-Q. Luo and P. Tseng. "Error bounds and convergence analysis of feasible descent methods: a general approach". In: *Annals of Operations Research* 46.1 (1993), pp. 157–178.

[Mor06] B. Mordukhovich. *Variational Analysis and Generalized Differentiation I: Basic Theory*. Grundlehren der mathematischen Wissenschaften. Springer Berlin Heidelberg, 2006.

[Mor76] B. Mordukhovich. "Maximum principle in the problem of time optimal response with nonsmooth constraints: PMM vol. 40, number 6, 1976, pp. 1014-1023". In: *Journal of Applied Mathematics and Mechanics* 40.6 (1976), pp. 960–969.

[MS92] J. Martínez-Legaz and A. Seeger. "A formula on the approximate subdifferential of the difference of convex functions". In: *Bulletin of the Australian Mathematical Society* 45.1 (1992), pp. 37–41.

[Nor78] V. Norkin. "Nonlocal minimization algorithms of nondifferentiable functions". In: *Cybern Syst Anal* 14 (1978), pp. 704–707.

[Nur73] E. Nurminski. "The quasigradient method for the solving of the nonlinear programming problems". In: *Cybern Syst Anal* 9 (1973), pp. 145–150.

[Nur78] E. Nurminski. "On epsilon-Differential Mappings and their Applications in Nondifferentiable Optimization". In: (1978).

[Oli20] W. de Oliveira. "The ABC of DC Programming". In: *Set-Valued and Variational Analysis* 28 (2020), pp. 679–706.

[Pen96] J.-P. Penot. "Subdifferential Calculus Without Qualification Assumption". In: *Journal of Convex Analysis* 3 (1996), pp. 207–220.

[PU02] D. Pallaschke and R. Urbański. *Pairs of Compact Convex Sets: Fractional Arithmetic with Convex Sets*. Vol. 548. Mathematics and Its Applications. Springer, Dordrecht, 2002, p. 297.

[Rob99] S. M. Robinson. "Linear convergence of epsilon-subgradient descent methods for a class of convex functions". In: *Mathematical Programming* 86.1 (1999), pp. 41–50.

[RW09] R. Rockafellar and R. J.-B. Wets. *Variational Analysis*. 3rd. Vol. 317. Grundlehren der mathematischen Wissenschaften. Berlin: Springer Verlag Berlin, 2009, p. 734.