Desingularizing functions in convex programming and convergence of an abstract gradient descent algorithms

Aïcha Balhag, Abderrahim Jourani

Desingularizing functions in convex programming and convergence of an abstract gradient descent algorithms*

A. Balhag and A. Jourani†

Abstract Over the past years, the idea of the KL property has become increasingly common in the optimisation area. It was used as tool guarantees that every sequence generated by many different descent algorithms enjoys finite length property. When verifying the KL property, one needs to find a desingularizing function, which is usually considered difficult. In light of this, our paper provides a desingularizing function that can be regarded as the smallest among all potential desingularizing functions. We investigate continuity, Lipschitz continuity, differentiability and subdifferentiability properties of this function as well as different characterizations of the so-called Kurdyka-Lojasiewicz-Hoffman constant by introducing a new class of functions with nonsmooth moderate behaviour. Based on these attributes, we provide a convergence analysis and an estimation of the convergence rate of an inexact descent methods for convex differentiable functions that covers a wide class of gradient descent methods. We also delves the complexity property of this method and its relationship with λ-one-dimensional worst-case proximal sequences.

Keywords: Desingularizing function, Error bounds, Kurdyka-Lojasiewicz-Hoffman constant, Convex minimization, Inexact descent methods, Complexity.

Mathematics Subject Classification: 90C06, 90C25, 90C60, 65K05

1 Introduction

Continuous optimization problems are prevalent across diverse domains, including machine learning, signal processing, and data analytic. Recently, there has been a lot of interest in using the Kurdyka-Lojasiewicz (KL) inequality in many applications to solve these problems. Roughly speaking, this inequality states that for a differentiable function f, there is a smooth concave function μ such that the following inequality holds

$$\|\nabla (\mu \circ (f - \min f))(x)\| \geq 1, \tag{1}$$

for all x in a compact neighborhood of the set of critical points of f, see [16]. Its extension to the nonsmooth case [15, 17] has opened up unexpected pathways in the nonconvex world and enabled us to conduct convergence rate for a number of significant optimisation algorithms , see, e.g. [1, 3, 2, 10, 19, 24, 41] and the references therein.
In the KL inequality (1), the desingularizing function μ is a witness to the fact that f is asymptotically well-behaved, in matter of fact, the faster μ' (derivative of μ) tends to infinity at 0, the flatter is f around critical points. However, it is important to note that the desingularizing function is not necessary unique. So, it is natural for us to ponder what the optimal option would be.
Given an extended real-valued function f defined on a Banach space X, such that $m := \inf_X f$ and $S := Argminf$ is nonempty. The function $\mu_f : \mathbb{R}_+ \to \mathbb{R}_+$ given by

$$\mu_f(t) = \sup\{d(x, S) : \quad f(x) - m \leq t\}, \tag{2}$$

*This work was partially supported by the Conseil Régional BFC, FEDER-BFC and EIPHI Graduate School (contract ANR-17-EURE-0002)
†Corresponding author: Abderrahim.Jourani@u-bourgogne.fr
Institut de Mathématiques de Bourgogne, UMR 5584 CNRS, Université Bourgogne, F-21000 Dijon, France
aichabalhag@gmail.com

has been recognized as the canonical conditioner for f in [43, 7]. This value function serves as the most smallest non-decreasing conditioner. In another terms, it is the the smallest function μ_f that satisfies the following error bound inequality:

$$\forall x \in X \quad d(x, S) \leq \mu_f(f(x) - m).$$

It is important to note that error bounds, as indicated in [18], can provide the KL inequality. This implies that a conditioner can be regarded as a desingularizing function, and further μ_f might be the best optimal desingularizing one.
By using either techniques from convex analysis or tools from nonsmooth analysis, we study different properties of μ_f including continuity, Lipschitz continuity, differentiability and subdifferentiability. Indeed, these properties constitute the heart of our study of a gradient descent algorithms that we propose in this work to solve unconstrained smooth optimization problems.

The gradient descent method is certainly among the most fundamental and simple algorithms to solve smooth optimization problems. Actually, many of gradient descent method algorithm convergence results have been analyzed within the broader framework of KL functions and established within an abstract scheme of descent methods as in [2, 18, 28] . In matter of fact, an abstract descent scheme is a set of abstract properties ensuring the convergence of a generic iterative scheme to a stationary point if combined with the KL inequality. It is noteworthy that several specific algorithms may be derived from the abstract scheme. The first abstract descent scheme was considered by Attouch and co-authors in [2] for nonsmooth, nonconvex functions. Later, several authors adopted and developed this method, to be able to analyze and build new and existing algorithms,see for instance, [2, 18, 28] with references therein.
Meanwhile, over the past few decades, the gradient method has been modified in many ways. As a way to speed gradient descent procedures is inertial methods, which provide an alternative strategy for accelerating the rate of convergence. It differs from the usual gradient method by adding an inertial term that is computed by the difference of the two preceding iterations. It may be challenging to estimate the rate of convergence in a nonconvex case, but from a numerical perspective, it is still favourable. One of the noteworthy techniques in the field of inertial gradient descent is Polyak's heavy ball method [44]. The method's name derives from the fact that it may be regarded as an explicit finite difference discretization of the so-called Heavy-ball with friction dynamical see [5, 29]. Another popular inertial method that shows some similarities with the heavy ball method is Nesterov's accelerating gradient method [35], Indeed, both of them can accelerate convergence rate while keeping the cost of each iteration relatively constant. However, while the Heavy-ball method uses gradients based on the current iterate, Nesterov's accelerated gradient method evaluates the gradient at points that are extrapolated by the inertial force.
Many authors modify the abstract scheme proposed in [2] in order to include a lot of inertial algorithms, among which we list the works of Ochs [41] and his co-authors, who studied what is called iPiano (proximal inertial algorithm for nonconvex optimization) and which can be considered as a generalization of the Heavy-Ball method [44, 45]. Similarly, Lazlo in [38] adapted an abstract descent scheme to his proposed Nesterov Gradient type inertial algorithm. We refer the reader to the references [42, 20, 21] among others. The interesting aspect is that the convergence rates of many of these previously mentioned methods depend on the KL inequality rather than the nature of the algorithm. That is, it depends on the desingularizing function μ. For an abstract first-order descent method in convex minimization. The authors in [18] have shown that the inverse μ^{-1} of a desingularization function μ of f on a convenient domain contains almost all the information provided by their approach about the complexity of descent methods. They proved that for a convex objective f and a descent sequence fulfills the following two conditions:

(i) $a\|x_k - x_k\|^2 + \|^2 \leq f(u_k) - f(u_{k+1})$;

(ii) $\|w_k)\| \leq \|b\|x_k - x_{k-1}\|$, where $w_k \in \partial f(x_k)$ for all $k \in \mathbb{N}^*$,

where a and b are positive real numbers, the rate of convergence of the value is

$$|f(x_k) - minf| \leq \mu^{-1}(\alpha_k), \ k \geq 0.$$

where α_k is sequence what's called a worst case one dimensional proximal method that is defined by

$$\alpha_{k+1} = argmin\{\mu^{-1}(\alpha_k) + \frac{1}{2}(s - \alpha_k)^2 : \ s > 0\}, \quad \alpha_0 = \mu^{-1}(f(x_0)).$$

Motivated by this concept, our goal is to unify and extend the frameworks delineated in the previous mentioned works. To this end, after giving different properties of our desingularizing μ_f defined in (2), and through the utilization of the following regularization function of f in Hilbert space $\mathcal{H}$,

$$H : \mathcal{H} \times \mathcal{H} \to \mathbb{R}, H(u, v) = f(u) + \frac{1}{2}\|u - v\|^2,$$

we consider an abstract descent gradient method in a way that generates a link between sequences $(x_k)_{k\in\mathbb{N}}$ in $\mathcal{H}$ and $(z_k)_{k_{\in\mathbb{N}}} := (u_k, v_k)_{k\in\mathbb{N}}$ in $\mathcal{H} \times \mathcal{H}$, through the following hypotheses:

(H_1) For each $k \in \mathbb{N}^*$, for some $a_k > 0$ and $b_k > 0$,

$$a_k\|x_{k+1} - x_k\|^2 + b_k\|x_k - x_{k-1}\|^2 \leq H(z_k) - H(z_{k+1});$$

(H_2) For each $k \in \mathbb{N}^*$, for some $c_k > 0$ and $\varepsilon_k \geq 0$,

$$c_k\|\nabla H(z_k)\| \leq \|x_{k+1} - x_k\| + \|x_k - x_{k-1}\| + \varepsilon_k;$$

(H_3) For each $k \in \mathbb{N}^*$, for every $z = (x, x) \in \mathcal{H} \times \mathcal{H}$, and with fixed $c_1, c_2 \geq 0$,

$$\|z_k - z\| \leq c_1\|x_k - x\| + c_2\|x_{k-1} - x\|;$$

(H_4) The sequences $(a_k)_{k\in\mathbb{N}}$, $(b_k)_{k\in\mathbb{N}}$, $(\varepsilon_k)_{k\in\mathbb{N}}$, and $(c_k)_{k\in\mathbb{N}}$ satisfy

$$\varepsilon_k \notin l^1, \quad \inf_{k\geq 0} c_k a_k > 0, \quad \text{and } \inf_{k\geq 0} c_k b_k > 0.$$

Indeed, for convex differential function f, we analyse the asymptotic behaviour of our abstract algorithm and we prove as in [18] that the convergence rate dependent on the inverse the desingularizing function μ_H.
We present now the paper's structure and highlight its important contributions:
In section 2, we provide some notations and results from (nonsmooth) variational analysis in general Banach space. Following that, we present our main results. In section 3, after giving the definition of the KL property, we exploit a general different feather of μ_f. We prove different properties of μ_f including continuity, Lipschitz continuity as well as subdifferentiability. Actually, our desingularizing function may be neither differentiable nor concave. However, surprisingly, for smooth convex function f, we show in Theorem 3.9, that the function μ_f is differentiable, and further, for all $x \notin Argminf$, the following equality holds

$$\mu_f'(f(x))\|\nabla f(x)\| = 1. \tag{3}$$

We conclude this section by giving different characterization of the so-called Kurdyka-Lojasiewicz-Hoffman constant by introducing a new class of functions with nonsmooth moderate behaviour in the sprit of that considered in [18] and refrences therein.
In the last section, for smooth convex objective function f defined in Hilbert space $\mathcal{H}$, we present our inertial inexact abstract descent gradient schema satisfying the hypotheses $(H_1) - (H_4)$, which is inspired by [38] but extending their setting in order to take into account additive computational errors,which gives a more flexibility in the choice of the parameters. Instead of using KL, we use the equality (3) for μ_H related to the regularization function H. We prove, under certain hypotheses, that the iterates have finite length and strong convergence. Finally, similar to the work in [18], we show that general convergence rates depend also on a parameter α_k which is the sequence of the one dimensional worst case defined by

$$\alpha_{k+1} = argmin\{\mu_H^{-1}(\alpha_k) + \frac{1}{2}(s - \alpha_k)^2 : s > 0\}, \quad \alpha_0 = \mu_H^{-1}(H(z_0)).$$

The methodologies and findings presented in our work open the door to improve convergence results of a broad range of algorithms that adopt the KL property.

2 Notation and preliminaries

In order to make the paper as short as possible, some definitions and the complete wording of the results will not be repeated here, and as needed, will be referenced to [39]-[40] and [25]. Throughout, we shall assume that X is a Banach space endowed with some norm denoted by $\|\cdot\|$ to which we associate the distance function $d_C(\cdot)$ to a set C.
We write $x \xrightarrow{f} x_o$, and $x \xrightarrow{S} x_o$ to express $x \to x_0$ with $f(x) \to f(x_0)$ and $x \to x_0$ with $x \in S$, respectively.
Let f be an extended-real-valued function on X. The limiting Fréchet subdifferential of f at x_0 in x is the set

$$\partial f(x_0) = w^* - seq - \limsup_{\substack{x \xrightarrow{f} x_0 \\ \varepsilon \to 0^+}} \partial_F^{\varepsilon} f(x)$$

where

$$\partial_F^{\varepsilon} f(x) = \{x^* \in X^* : \liminf_{h \to 0} \frac{f(x+h) - f(x) - \langle x^*, h\rangle}{\| h \|} \geq -\varepsilon\}$$

is the $\varepsilon-$Fréchet subdifferential of f at x.
The limiting Fréchet normal cone to a closed set $S \subset X$ at a point $x \in S$ is given by

$$N(S, x) = \partial \delta_S(x)$$

where δ_S denotes the indicator function of S.
In the so-called Asplund spaces, the limiting Fréchet subdifferential takes the following form (which is in fact a characterization of this class of spaces):

$$\partial f(x_0) = w^* - seq - \limsup_{x \xrightarrow{f} x_0} \partial_F f(x)$$

where

$$\partial_F f(x) = \{x^* \in X^* : \liminf_{h \to 0} \frac{f(x+h) - f(x) - \langle x^*, h\rangle}{\| h \|} \geq 0\}$$

is the Fréchet subdifferential of f at x.

If f is an extended-real-valued function on X, the function

$$f^-(x, h) = \liminf_{\substack{u \to h \\ t \downarrow 0}} t^{-1}(f(x + tu) - f(x))$$

is the lower Dini directional derivative of f at x. To this directional derivative, it is associated the so called Dini-subdifferential defined as

$$\partial^- f(x) = \{x^* \in X^* : \langle x^*, h\rangle \leq f^-(x, h) \, \forall h \in X\}.$$

It tunrs out that this subdifferential coincides with the Fréchet subdifferential in finite dimension.
Note that when f is locally Lipschitz at x

$$f^-(x, h) = \liminf_{t \downarrow 0} t^{-1}(f(x + th) - f(x)).$$

If f is locally Lipschitz at x, the function

$$f^0(x, h) = \limsup_{\substack{u \to x \\ t \downarrow 0}} t^{-1}(f(u + th) - f(u))$$

is the Clarke's directional derivative of f at x. The Clarke's subdifferential is then defined by

$$\partial_C f(x) = \{x^* \in X^* : `\langle x^*, h\rangle \leq f^0(x, h) \, \forall h \in X\}.$$

Having in mind this definition for locally Lipschitz functions, the Clarke normal cone to a closed set S at $x_0 \in S$ can be defined as

$$N_C(S, x_0) = \overline{\mathbb{R}_+ \partial_C d_S(x_0)}.$$

Here d_S denotes the distance function to a set S, that is,

$$d_S(x) = \inf_{u \in S} \|u - x\|.$$

As usual, this allows to define the Clarke's subdifferential for any extended real-valued lower semicontinuous f in terms of the Clarke's normal cone to the epigraph $epif$ of the function f

$$\partial_C f(x_0) = \{x^* \in X^* : \quad (x^*, -1) \in N_C(epif, (x_0, f(x_0))\}.$$

If X is Asplund and f is locally Lipschitz at x_0, then

$$\partial_C f(x_0) = cl^* co(\partial f(x_0)).$$

Finally we recall that the function f is Clarke regular at x if

$$f^-(x, h) = f^0(x, h) \, \forall h \in X.$$

Taking into account the previous remarks in finite dimension, the Clarke regularity of f at x is equivalent to say that

$$\partial_F f(x) = \partial_C f(x).$$

To close this section, let us recall that for lower semicontinous convex functions all theses subdifferentials coincide with the Fenchel subdifferential which will be denoted as

$$\partial f(x_0) = \{x^* \in X^* : \, x^*, x - x_0\rangle \leq f(x) - f(x_0) \quad \forall x \in X\}$$

or equivalently

$$\partial f(x_0) = \{x^* \in X^* : \, x^*, h\rangle \leq f'(x_0, h) \quad \forall h \in X\}.$$

Here $f'(x_0, \cdot)$ stands for the directional derivative in the sense of convex analysis, that is,

$$f'(x_0, h) = \lim_{s \downarrow 0} \frac{f(x_0 + sh) - f(x_0)}{s}.$$

3 Kurdyka-Lojasiewicz desingularizing function

3.1 Kurdyka-Lojasiewicz desingularizing function

Let $r > 0$ and consider the sets

$$\mathcal{K}^0(0, r) := \left\{\omega \in \mathcal{C}^0[0, r) : \, \omega \text{ is increasing and } \omega(0) = 0\right\},$$

$$\mathcal{K}(0, r) := \left\{\omega \in \mathcal{C}^0[0, r) \cap \mathcal{C}^1(0, r) : \, \omega(0) = 0, \, \exists \beta > 0; \, \omega(t) \leq \beta \omega'(t) t \quad \forall t \in (0, r)\right\}$$

and

$$\mathcal{K} := \left\{\omega \in \mathcal{C}^0[0, +\infty) \cap \mathcal{C}^1(0, \infty) : \, \omega(0) = 0, \, \exists \beta > 0; \, \omega(t) \leq \beta \omega'(t) t \quad \forall t > 0\right\}.$$

The function f satisfies the local Kurdyka-Łojasiewicz inequality at $\bar{x} \in S$ if there exist $r > 0$, $s > 0$ and $\omega \in \mathcal{K}(0, r)$ such that

$$\omega'(f(x)) d(0, \partial f(x)) \geq 1 \quad \forall x \in B(\bar{x}, s) \cap [0 < f < r], \tag{4}$$

where $[0 < f < r] := \{x \in X : 0 < f(x) < r\}$.

The function f satisfies the global Kurdyka-Łojasiewicz inequality if there exists $\omega \in \mathcal{K}$ such that

$$\omega'(f(x))d(0, \partial f(x)) \geq 1 \quad \forall x \in X \backslash S. \tag{5}$$

The function ω is called a *desingularizing function* for f at $\bar{x}$. One of the important questions is :

How to compute ω?

Its calculus depends on the geometry of f and may be computed in some special situations. One of them is the case (see Theorem 3 in [18]) where f is a semi-algebraic coercive convex function. More precisely, it is established that the function ω is defined by

$$\omega(t) = \alpha(t + t^{\frac{1}{p}})$$

for some $\alpha >$ and a rational number $p \geq 1$.
In 1963, Lojasiewicz [37] proved that any real analytic function $F : \mathcal{H} \to \mathbb{R}$ has the Lojasiewicz property, after that, in 1998 Kurdyka presented a more general construction which applies to differentiable functions definable in an o-minimal structure [35]. The extension to nonsmooth functions has been presented in [3, 12, 16]

Definition 3.1 (The Kurdyka-Lojasiewicz property) *Let $F : \mathcal{H} \to \mathbb{R}$ be a differentiable function. F has the Kurdyka-Lojasiewicz property (for short KL-property) at $\bar{x} \in \mathcal{H}$ if there exist $\eta > 0$, a neighborhood U of $\bar{x}$, and a continuous concave function $\mu \in \mathcal{K}_\eta$ where*

$$\mathcal{K}_\eta = \{\mu \in \mathcal{C}^0[0, \eta) \cap \mathcal{C}^1(0, \eta),\ \mu(0) = 0,\ \mu \text{ is concave and } \mu' > 0\}$$

such that: for all x in the intersection $U \cap \{x \in \mathcal{H} : F(\bar{x}) < F(x) < F(\bar{x}) + \eta\}$, the following inequality holds

$$\mu'(F(x) - F(\bar{x}))\|\nabla F(x)\| \geq 1.$$

The Lojasiewicz inequality or property is a special case of the KL-property when $\mu(s) = s^{1-\theta}$, $\theta \in [\frac{1}{2}, 1]$ It is automatically satisfied for non-critical points, so it is in fact a condition on critical points. We will need the following result which was given in [19].

Lemma 3.2 *Let $F : \mathcal{H} \to \mathbb{R}$ be a differentiable function and let $K \subseteq \mathcal{H}$ be compact. Suppose F is constant on K and has the KL-property at every $\bar{x} \in K$. Then there exists $\varepsilon > 0, \eta > 0$ and a real function $\mu \in \mathcal{K}_\eta$ where*

$$\mathcal{K}_\eta = \{\mu \in \mathcal{C}^0[0, \eta) \cap \mathcal{C}^1(0, \eta),\ \mu(0) = 0,\ \mu \text{ is concave and } \mu' > 0\}$$

such that:

$$\mu'(F(x) - F(\bar{x}))\|\nabla F(x)\| \geq 1.$$

for every $\bar{x}$ and every x such that $dis(x, K) \leq \varepsilon$ and $F(\bar{x}) < F(x) < F(\bar{x}) + \eta$.

3.2 Desingularizing function of f

We may also ask for the existence of such function in the absence of differentiablity. One natural candidate should be (see [43]) the function $\mu_f : \mathbb{R}_+ \to \mathbb{R}_+$ defined by

$$\mu_f(t) = \sup\{d(x, S) : \quad f_+(x) \leq t\}.$$

This function is increasing with $\mu_f(0) = 0$ and

$$d(x, S) \leq \mu_f(f_+(x)) \quad \forall x \in X.$$

Note that when μ_f is increasing, then the last inequality can be formulated as follows :

$$d(x, \{u \in X : (\mu_f \circ f)(u) \leq 0\}) \leq (\mu_f \circ f)((x)) \quad \forall x \notin S.$$

This inequality is nothing else that the so-called Hoffman error bound for the inequality system given by $g := \mu_f \circ f$(see [9] and references therein).

One very important issue of the function μ_f is when does it satisfy the following inequality

$$\limsup_{t \to 0^+} \frac{\mu_f(t)}{t} < \infty? \tag{6}$$

This question is very important in optimization since it characterizes the so called error bound. Indeed, it is not difficult to see that relation (6) is equivalent to say that

$$\exists \alpha > 0, \exists \delta >; \quad d(x, S) \leq \alpha f_+(x) \quad \forall x \in [0 \leq f \leq \delta]. \tag{7}$$

This last one is satisfied in many situation including the case when f is a polyhedral function (see [9] for furher results).

Let $S(t) = \{x \in X : \mu_f(t) = d(x, S),\ f(x) \leq t\}$ be the solution set of the optimization problem

$$\begin{cases} \max d(x, S) \\ f(x) \leq t \end{cases}$$

The function μ_f can be expressed as the supremum of the difference of two convex function whenever f is convex. More precisely, we obtain the following result which is a consequence of Proposition 2.4.3 in [25].

Proposition 3.3 *Let $t \geq 0$ be such that $S(t) \neq \emptyset$. Then*

$$\mu_f(t) = \sup_{x \in X} \{d(x, S) - d(x, S_f(t))\}. \tag{8}$$

Moreover for all $K > 1$ and $t > 0$, $\mu_f(t) = \max_{x \in X}(d(x, S) - Kd(x, S_f(t)))$ and $S(t) = S^K(t)$, where

$$S^K(t) = argmax\{d(x, S) - Kd(x, S_f(t))\}.$$

Recall that f is said to be starshaped at $\bar{x} \in S$ if, for any $x \in X$ and any $s \in [0, 1]$, one has

$$f((1 - s)\bar{x} + sx) \leq (1 - s)f(\bar{x}) + sf(x)$$

it is said to be starshaped on S if it is starshaped at any $\bar{x} \in S$. Note that the starshapeness of f on S ensures automatically the convexity of S.
Under the inf-compactness of f, we will have the following result.

Proposition 3.4 *Suppose that $f : X \to \mathbb{R} \cup \{+\infty\}$ is lower semicontinuous and starshaped on S, where X is a real Banach space, and that S coincides with the set of critical point of f with respect to the Mordukhovich subdifferential. Then μ_f is lower semicontinuous and lower-starshaped at 0 (i.e. $\mu_f(\alpha t) \geq \alpha \mu_f(t)$ for every $t > 0$ and $\alpha \in [0, 1]$) . Therefore, μ_f is subadditive, and $t \in]0, +\infty[\mapsto \frac{\mu_f(t)}{t}$ is decreasing. If moreover, f is inf-compact then μ_f is continuous.*

Proof. The lower semicontinuity is obvious for $t = 0$. So suppose $t > 0$ and let $t_k \to t$ be such that $\liminf_{t' \to t} \mu_f(t') = \lim_{k \to \infty} \mu_f(t_k)$. For $\varepsilon > 0$, pick $x \in X$ such that $\mu_f(t) \leq d(x, S) + \varepsilon$. If $f_+(x) < t$, then for k sufficiently large, $f_+(x) < t_k$, and hence $\mu_f(t) \leq \mu_f(t_k) + \varepsilon$. Thus $\mu_f(t) \leq \lim_{k \to \infty} \mu_f(t_k)$. Now, suppose that $f_+(x) = t$. Then, since S coincides with the set

of critical point of f with respect to the Mordukhovich subdifferential, $0 \notin \partial f(x)$. Corollary 4 in [**?**] asserts the existence of $a > 0$ and $r > 0$ such that

$$d\left(u, \{v \in X : f_+(v) \leq s\}\right) \leq a\left(f_+(u) - s\right)_+ \quad \forall v \in B(x,r),\ s \in]t-r, t+r[.$$

Then $t_k \in]t-r, t+r$ for k sufficiently and there exists x_k, with $f_+(x_k) \leq t_k$, such that

$$\|x - x_k\| \leq a\left(f_+(x) - t_k\right)_+$$

and hence

$$\begin{aligned}\mu_f(t) &\leq d(x,S) + \varepsilon \\ &\leq d(x_k,S) + \|x - x_k\| + \varepsilon \\ &\leq \mu_f(t_k) + a\left(f_+(x) - t_k\right)_+ + \varepsilon.\end{aligned}$$

Since $\left(f_+(x) - t_k\right)_+ \to 0$ and ε is arbitrary, we get $\mu_f(t) \leq \lim_{k\to\infty} \mu_f(t_k)$.
Now, let us show μ_f that lower-starshaped at 0. By contradiction, let $t > 0$ and $\alpha \in]0,1[$. Suppose that $\mu_f(\alpha t) < \alpha\mu_f(t)$. So, one can find $x \in X$ such that

$$\mu_f(\alpha t) < \alpha d(x,S), \quad \text{and} \quad f_+(x) \leq t.$$

Let us take $\varepsilon > 0$, with $\mu_f(\alpha t) < \alpha d(x,S) - \alpha\varepsilon$, and $u \in S$ such that $\|x - u\| \leq d(x,S) + \alpha\varepsilon$. Since

$$d(x,S) \leq (1-\alpha)\|x - u\| + d(\alpha x + (1-\alpha)u, S),$$

then $\alpha\|x - u\| \leq \alpha\varepsilon + d(\alpha x + (1-\alpha)u, S)$ and

$$\mu_f(\alpha t) + \alpha\varepsilon < \alpha d(x,S) \leq \alpha\|x - u\| \leq \alpha\varepsilon + d(\alpha x - (1-\alpha)u, S).$$

As f is starshaped at u, we have

$$f(\alpha x + (1-\alpha)u) \leq \alpha f(x) \leq \alpha t,$$

which implies that $\mu_f(\alpha t) + \alpha\varepsilon < \alpha\varepsilon + \mu_f(\alpha t)$, a contradiction. Hence, μ_f is starshaped at 0. The subadditivity of μ_f comes from the fact that for every $t, s > 0$, we have $\mu_f(t) \geq \frac{t}{t+s}\mu_f(t+s)$ and $\mu_f(s) \geq \frac{s}{t+s}\mu_f(t+s)$. The function $t \to \dfrac{\mu_f(t)}{t}$ is nonincreasing due to fact that for $0 < t \leq s$, we have

$$\mu_f(t) = \mu_f(\frac{t}{s}s) \geq \frac{t}{s}\mu_f(s).$$

⊠

Remark 3.5 *1. Using the monotonicity of $t \to \frac{\mu_f(t)}{t}$, we easily see that*

$$\forall t \in [0,1], \quad t\mu_f(1) \leq \mu_f(t) \text{ and } \forall t \geq 1, \quad \mu_f(t) \leq t\mu_f(1). \tag{9}$$

2. The strict monotony of μ_f combined with the convexity of f does not ensure the concavity nor the convexity of μ_f. To see this, consider the function f defined by

$$f(x) = \begin{cases} x^4 & \text{if } x < 0 \\ 0 & \text{if } x \in [0,1] \\ (x-1)^2 & \text{if } x > 1. \end{cases}$$

Then

$$\mu_f(t) = \max(\sqrt{t}, \sqrt[4]{t})$$

which is not concave nor convex.

Proposition 3.6 *Let f be as in Proposition 3.4 and set $\bar{t} = \sup_{x \in dom f} f(x)$. Suppose in addition that f is continuous on $domf$. Then the following assertions hold:*

1. *If $\bar{t} \in \mathbb{R}$,*
$$\mu_f(t) = \mu_f(\bar{t}) \quad \forall t \geq \bar{t}.$$
2. *For all $t \in [0, \bar{t})$, the set $S(t)$ is nonempty and for all $t \in (0, \bar{t})$, $S \cap S(t) = \emptyset$.*
3. *For all $t_0 \in (0, \bar{t})$, the set-valued mapping $t \mapsto S_f(t) := \{x \in X : f(x) \leq t\}$ is Lipschitz near t_0, that is, there exist real numbers $\gamma > 0$ and $\alpha > 0$ such that*
$$S_f(t) \subset S_f(t') + \gamma|t - t'|\mathbb{B} \quad \forall t, t' \in]t_0 - \alpha, t_0 + \alpha[.$$
Moreover μ_f is locally lipshitz on $(0, \bar{t})$.
4. *Let $t_0 \in (0, \bar{t})$ and $x_0 \in S(t_0)$ at which f is continuous. Then*
   (a) For all $x^* \in \partial d_{S_f(t_0)}(x_0)$, with $x^* \neq 0$, then there exists $t^* \in (0, K)$ such that $x^* \in t^* \partial f(x_0)$.
   (b) $\partial d_S(x_0) \subset (0, K)\partial f(x_0)$.

Proof.
1. This item is obvious.

2. Let $t \in [0, \bar{t})$. By the inf-compactness of the function f, the set $[f \leq t]$ is compact and then there exists $x \in [f \leq t]$ such that, $\mu_f(t) = d(x, S)$ which asserts that $S(t) \neq \emptyset$. We will prove that $S \cap S(t) = \emptyset$. We will argue by contradiction. So let $x \in S \cap S(t)$. Then $f_+(x) = 0$ and hence $f_+(x) < t$. Using the definition of $\bar{t}$, there exists $y \in \text{dom} f$ such that $f(y) > t$. Set
$$\gamma(s) = f(sx + (1 - s)y), \ \forall s \in [0, 1].$$
We have $\gamma(1) < t < \gamma(0)$ and γ is continuous on $[0, 1]$, there exists therefore $s \in (0, 1)$ such that $t = \gamma(s) = f(sx + (1 - s)y)$. It follows that $d(x, S) = \mu_f(t) \geq d(sx + (1 - s)y, S) > 0$ and hence $x \notin S$ and this contradicts our assumption. So that $S \cap S(t) = \emptyset$.

3. To establish this item, let us consider the function
$$\begin{array}{rl} g & : X \times \mathbb{R} \to \mathbb{R} \cup \{+\infty\} \\ & (x, t) \mapsto f(x) - t \end{array} \tag{10}$$
Given $t_0 \in (0, \bar{t})$ and $x_0 \in S(t_0)$, because $S \cap S(t_0) = \emptyset$, we have $0 \notin \partial_x g(x_0, t_0)(= \partial f(x_0))$, where $\partial_x g$ is the partial subgradient of g with respect to x. According to Theorem 7.1 of [34], there exist $r > 0$ and $a > 0$ such that
$$d(x, \{u : g(u, t) \leq 0\}) \leq a g_+(x, t), \ \forall x \in \mathbb{B}(x_0, r), \ \forall t \in \mathbb{B}(t_0, r) \tag{11}$$
Now suppose for contradiction that there exist sequences (t_k) and (t'_k) both converging to t_0, with $t_k \neq t'_k$ for all k, and $x_k \in S_f(t_k)$ with $x_k \notin S_f(t'_k) + k|t_k - t'_k|$. By separation theorem, there exists $x^*_k \in X^*$, with $\|x^*_k\| = 1$, such that
$$\langle x^*_k, x_k \rangle \leq -k|t_k - t'_k| + \inf_{x \in S_f(t'_k)} \langle x^*_k, x \rangle. \tag{12}$$
By the inf-compactness, extracting subsequence if necessary, we may assume that (x_k) converges to x_0 with $f(x_0) \leq t_0$. Relation (11) asserts that for k large enough, there exists $x'_k \in S_f(t'_k)$ such that
$$\|x_k - x'_k\| \leq a g(x_k, t'_k) = a(g(x_k, t'_k) - g(x_k, t_k)) = a|t_k - t'_k|.$$

Using this inequality and relation (12), we obtain

$$k|t_k - t'_k| \leq a|t_k - t'_k|$$

which leads to a contradiction because k is arbitrary.
The local Lipschitzness of μ_f follows from that of the set-valued mapping S_f and relation (8).

4. Let $t_0 \in (0, \bar{t})$, $x_0 \in S(t_0)$ and $x^* \in \partial d_{S_f(t_0)}(x_0)$, with $x^* \neq 0$. Then

$$\|x^*\| \leq 1 \text{ and } \langle x^*, x - x_0\rangle \leq 0 \quad \forall x \in S_f(t_0).$$

As the set-valued mapping S_f is Lipschitz with constant K, we have for all t near t_0 and $x \in S_f(t)$ there exists $b \in \mathbb{B}$ such that

$$x + K|t - t_0|b \in S_f(t_0)$$

and hence

$$\langle x^*, x + K|t - t_0|b - x_0\rangle \leq 0.$$

Thus, because of the convexity of f,

$$\langle x^*, x - x_0\rangle \leq K|t - t_0| \quad \forall (x, t) \in \text{epi}f.$$

So that there exists $t^* \in [-K, K]$ such that $(x^*, -t^*) \in N(\text{epi}f, (x_0, t_0))$. Since f is continuous at x_0 then $f(x_0) = t_0$ and $t^* > 0$ (otherewise $x^* = 0$ and this contradicts our hypothesis on x^*), and hence $x^* \in t^*\partial f(x_0)$. The last inclusion uses the last inclusion and some tools from DC-programming ([31]) namely, because $x_0 \in S(t_0)$, the following inclusion holds

$$\partial d_S(x_0) \subset \partial d_{S_f(t_0)}(x_0).$$

⊠

Now that we have studied the Lipschitz continuity of μ_f, we will focus on its differentiability as well as its subdifferentiability properties. Before doing so, we start with some elementary properties of μ_f:

1. If $x_0 \in S(t_0)$, with $f_+(x_0) < t_0$, then

$$\forall t \in [f_+(x_0), t_0], \quad \mu_f(t) = \mu_f(t_0). \tag{13}$$

2. So that the following inclusion holds

$$\partial d(x, S) \subset \partial_F(\mu_f \circ f)(x) \quad \forall x \in S(t), \quad \forall t \geq 0. \tag{14}$$

3. Consequently

$$d(0, \partial_F(\mu_f \circ f)(x)) \leq 1 \quad \forall x \in S(t), \quad \forall t > 0. \tag{15}$$

Now, we may state the desired result on the differentiability of μ_f.

Proposition 3.7 *Let $f : X \mapsto \mathbb{R}$ be a convex and inf-compact function. Then the following assertions hold true:*

1. *For all $t_0 > 0$, $x_0 \in S(t_0)$ and $x^* \in \partial d_S(x_0)$ we have*

$$\langle x^*, h\rangle \leq \mu_f^-(t_0; f'(x_0, h)) \quad \forall h \in X. \tag{16}$$

2. When X is a Hilbert space then for all $t_0 > 0$ and $x_0 \in S(t_0)$ we have $\partial d_S(x_0) = \{\nabla d_S(x_0)\}$ and

$$1 \leq \mu_f^-(t_0; f'(x_0, \nabla d_S(x_0))), \tag{17}$$

which shows that $f'(x_0, \nabla d_S(x_0)) > 0$. If moreover, f is differentiable on $S(t_0)$, the following relations hold

$$\langle \nabla f(x_0), \nabla d_S(x_0)\rangle > 0 \text{ and } \frac{1}{\langle \nabla f(x_0), \nabla d_S(x_0)\rangle}, \frac{\langle \nabla f(x_0), \nabla d_S(x_0)\rangle}{\|\nabla f(x_0)\|^2} \in \partial_F \mu_f(t_0). \tag{18}$$

Consequently, if $\partial_F(-\mu_f)(t_0) \neq \emptyset$, which is the case whenever μ_f is concave or more generally when $-\mu_f$ is Clarke regular at t_0, then μ_f is differentiable at t_0.

Proof. 1. Let $t_0 > 0$, $x_0 \in S(t_0)$, $h \in X$ and $s_n \downarrow 0$ be such that $\mu_f^-(t_0, f'(x_0, h)) = \lim\limits_{n\to+\infty} \dfrac{\mu_f(t_0 + s_n f'(x_0, h)) - \mu_f(t_0)}{s_n}$. Note that as f is convex, $f'(x_0, h) = \lim\limits_{n\to+\infty} \dfrac{f(x_0 + s_n h) - f(x_0)}{s_n}$. Set $\alpha_n = \frac{f(x_0+s_n h)-f(x_0)}{s_n}$. Since μ_f is locally Lipschitz, it follows that $\mu_f^-(t_0, f'(x_0, h)) = \lim\limits_{n\to+\infty} \dfrac{\mu_f(t_0 + s_n\alpha_n) - \mu_f(t_0)}{s_n} = \lim\limits_{n\to+\infty} \dfrac{\mu_f(f(x_0 + s_n h) - \mu_f(t_0)}{s_n}$. Thus for all $x^* \in \partial d_S(x_0)$ we have

$$\begin{aligned}
\mu_f^-(t_0, f'(x_0, h)) &= \lim_{n\to+\infty} \frac{\mu_f(f(x_0 + s_n h) - \mu_f(t_0)}{s_n} \\
&\geq \lim_{n\to+\infty} \frac{d_S(x_0 + s_n h) - d_S(x_0)}{s_n} \\
&\geq \langle x^*, h\rangle.
\end{aligned}$$

2. As $x_0 \notin S$, $\|\nabla d_S(x_0)\| = 1$ and $1 \leq \mu_f^-(t_0, f'(x_0, \nabla d_S(x_0)))$ (take $h = \nabla d_S(x_0)$ in 1.), which shows, by the monotonicity of μ_f, that $f'(x_0, \nabla d_S(x_0)) > 0$.
Taking into account the inclusion in 1., we can easily see ([36]) that if $\partial_F(-\mu_f)(t_0) \neq \emptyset$ then μ_f is differentiable at t_0.

⊠

Now, may state subdifferential estimates of $-\mu_f$.

Theorem 3.8 *Suppose that $f : X \to \mathbb{R}$ is convex continuous and inf-compact, where X is a real Banach space, and let $t_0 > 0$. Then*

1.

$$\partial(-\mu_f)(t_0) \subset \bigcup_{x_0 \in S(t_0)} \{-t^* \in \mathbb{R} : t^* > 0, \partial d_S(x_0) \subset t^* \partial f(x_0)\}. \tag{19}$$

2.

$$\partial_F(-\mu_f)(t_0) \subset \bigcap_{x_0 \in S(t_0)} \{-t^* \in \mathbb{R} : t^* > 0, \partial d_S(x) \subset t^* \partial f(x_0)\}. \tag{20}$$

Hence if $\partial_F(-\mu_f)(t_0) \neq \emptyset$ and f is differentiable on $S(t_0)$ then μ_f is differentiable at t_0 and for all $x_0 \in S(t_0)$, $\mu_f'(t_0) = \frac{1}{\|\nabla f(x_0)\|}$. If moreover, f is of class $\mathcal{C}^{1,0}$, then so is μ_f on $]0, +\infty[$ and

$$\lim_{t\to 0^+} \mu_f'(t) = +\infty.$$

3. When X is a Hilbert space, the inclusion in relation (19) can be written as

$$\partial(-\mu_f)(t_0) \subset \bigcup_{x_0 \in S(t_0)} \{-t^* \in \mathbb{R} : t^* > 0, \nabla d_S(x_0) \in t^* \partial f(x_0)\}. \tag{21}$$

If moreover f is differentiable on $S(t_0)$, then

$$\bigcup_{x_0 \in S(t_0)} \{-t^* \in \mathbb{R} : t^* > 0, \nabla d_S(x_0) = t^* \nabla f(x_0)\} \subset -\partial_F \mu_f(t_0) \tag{22}$$

and consequently

$$\partial_C \mu_f(t_0) = \partial_F \mu_f(t_0) = \partial \mu_f(t_0), \tag{23}$$

that is, μ_f is Clarke regular at t_0.

4. For all $t \geq 0$,

$$S(t) = M(t) := \{x \in X : \mu_f(t) = d_S(x),\ f(x) = t\}. \tag{24}$$

Proof. Let $t_0 > 0$. Since f is inf-compact, then there exists $k_f >$ and $\gamma > 0$ such that

$$|f(x) - f(x')| \leq k_f \|x - x'\| \quad \forall x, x' \in S(t_0) + \gamma \mathbb{B}.$$

1. Set $w = -\mu_f$. By Proposition 3.6, $S \cap S(t_0) = \emptyset$ and hence for all $x_0 \in S(t_0)$, $0 \notin \partial f(x_0)$. Theorem 7.2 in [34] ensures the following inclusion

$$\partial w(t_0) \subset \bigcup_{x_0 \in S(t_0)} \{-t^* \in \mathbb{R} : t^* \geq 0, 0 \in t^* \partial f(x_0) - \partial d(x_0, S)\}.$$

If $0 \in \partial w(t_0)$, then we get $0 \in \partial d(x_0, S)$, which contradicts the fact that $x_0 \in S(t_0)$, beacause by Proposition 3.6, $S \cap S(t_0) = \emptyset$. Hence,

$$\partial w(t_0) \subset \bigcup_{x_0 \in S(t_0)} \{-t^* \in \mathbb{R} : t^* > 0, 0 \in t^* \partial f(x_0) - \partial d(x_0, S)\}.$$

2. Let $t^* \in \partial_F(-\mu_f)(t_0)$. Then for all $\varepsilon > 0$ there exists $\gamma > \delta > 0$ such that

$$-\mu_f(t) + \mu_f(t_0) - t^*(t - t_0) + \varepsilon |t - t_0| \geq 0 \quad \forall t \in [t_0 - \delta, t_0 + \delta].$$

So that for all $x_0 \in S(t_0)$, we have

$$-d(x, S) + d(x_0, S) - t^*(f(x) - f(x_0)) + \varepsilon k_f \|x - x_0\| \geq 0 \quad \forall x \in B(x_0, \frac{\delta}{k_f}).$$

Thus for all $x^* \in \partial d_S(x_0)$,

$$\langle x^*, x_0 - x \rangle - t^*(f(x) - f(x_0)) + \varepsilon k_f \|x - x_0\| \geq 0 \quad \forall x \in B(x_0, \frac{\delta}{k_f})$$

or equivalently, because $t^* < 0$ and f and S are convex,

$$\langle x^*, x_0 - x \rangle - t^*(f(x) - f(x_0)) + \varepsilon k_f \|x - x_0\| \geq 0 \quad \forall x \in X.$$

As ε is arbitrary

$$\langle x^*, x_0 - x \rangle - t^*(f(x) - f(x_0)) \geq 0 \quad \forall x \in X$$

or equivalently $x^* \in (-t^*)\partial f(x_0)$. So that

$$\partial_F(-\mu_f)(t_0) \subset \bigcap_{x_0 \in S(t_0)} \{-t^* \in \mathbb{R} : t^* > 0,\ \partial d(x_0, S) \subset t^* \partial f(x_0)\}.$$

Now, suppose that $\partial_F(-\mu_f)(t_0) \neq \emptyset$ and f is differentiable on $S(t_0)$. Proposition 3.7 ensures that μ_f is differentiable at t_0 and since f is differentiable on $S(t_0)$, then for all $-t^* \in \partial_F(-\mu_f)(t_0)$ and $x_0 \in S(t_0)$, each $x^* \in \partial d_S(x_0)$ is equal to $t^* \nabla f(x_0)$. This means that $\partial d_S(x_0)$ is a singleton $\{\nabla d_S(x_0)\}$. Since $x_0 \notin S$, $\|\nabla d_S(x_0)\| = 1$. Thus $t^* = \frac{1}{\|\nabla f(x_0)\|}$ and this holds for all $x_0 \in S(t_0)$. This asserts that $\mu_f'(t_0) = \frac{1}{\|\nabla f(x_0)\|}$ for all $x_0 \in S(t_0)$.

3. Let $x_0 \in S(t_0)$ and $t^* > 0$ be such that $\nabla d_S(x_0) = t^* \nabla f(x_0)$. Let $\tau \in \mathbb{R}$ and $h = \tau \nabla d_S(x_0)$. Proposition 3.7 ensures that

$$\langle \nabla d_S(x_0), \tau \nabla d_S(x_0) \rangle \leq \mu_f^-(t_0, \tau \langle \nabla f(x_0), \nabla d_S(x_0) \rangle).$$

Note that $x_0 \notin S$ and $\|\nabla d_S(x_0)\| = 1$. Since $\nabla d_S(x_0) = t^* \nabla f(x_0)$, we obtain

$$t^* \tau \leq \mu_f^-(t_0, \tau) \quad \forall \tau \in \mathbb{R}$$

or equivalently $t^* \in \partial^- \mu_f(t_0) = \partial_F \mu_f(t_0)$. As $-\partial_C \mu_f(t_0) = \partial_C(-\mu_f)(t_0) = co\partial(-\mu_f)(t_0) \subset -\partial_F \mu_f(t_0) \subset -\partial \mu_f(t_0)$ and $\partial_F \mu_f(t_0)$ is convex, we obtain the desired equality.

4. The first items ensures that μ_f is increasing. Taking into account relation (13), we deduce that $S(t) \subset M(t)$.

⊠

Theorem 3.9 *Let f be as in Theorem 3.8. Then*

1. *For all $r > 0$ there exists $\gamma > 0$ such that*

$$\forall t \in]0, r],\ \partial_c \mu_f(t) \subset [\gamma, \frac{\mu_f(t)}{t}[.$$

Consequently,

$$\frac{\mu_f(t) - \mu_f(\bar{t})}{t - \bar{t}} \geq \gamma, \quad \forall t, \bar{t} \in [0, r] \text{ with } t \neq \bar{t}$$

and

$$\forall t \in]0, r], \quad \gamma t \leq \mu_f^-(t, t) \leq \mu_f^0(t, t) \leq \mu_f(t). \tag{25}$$

2. *For all $r > 0$ and $\alpha \in]0, r]$ there exists $\beta \geq 1$ such that*

$$\forall t \in [\alpha, r], \quad \mu_f(t) \leq \beta \mu_f^-(t, t). \tag{26}$$

3. *For all $r > 0$ and $\alpha \in]0, r]$ there exists $c > 0$ such that*

$$d(0, \partial(\mu_f \circ f)(x)) \geq c \quad \forall x \in [\alpha \leq f \leq r]$$

where $[\alpha \leq f \leq r] = \{x \in X : \alpha \leq f(x) \leq r\}$.

4. $\lim\limits_{t \to 0^+} \dfrac{\mu_f(t)}{t} = \limsup\limits_{t \to 0^+} \mu_f^-(t, 1) = \limsup\limits_{t \to 0^+} \mu_f^0(t, 1)$.

Proof. 1. Suppose that this item is not true. Then there exists $r > 0$ such that for $n \in \mathbb{N}^*$ there exists $t_n \in]0, r]$ and $-t_n^* \in \partial(-\mu_f)(t_n)$ such that $t_n^* \to 0^+$. Theorem 3.8 asserts the existence of $x_n \in M(t_n)$, $x_n^* \in \partial d(x_n, S)$, with $\|x_n^*\| = 1$ (because $x_n \notin S$), and $u_n^* \in \partial f(x_n)$ such that

$$x_n^* = t_n^* u_n^*. \tag{27}$$

Now, since f is inf-compact, the sequence (x_n) lives in some compact of X and, extracting subsequence if necessary, we may assume that $x_n \to \bar{x}$. As f is locally Lipschitz at $\bar{x}$, then the sequence (u_n^*) is necessarily bounded, which contradicts relation (27). Indeed, by taking the norm on both sides of (27), one gets $1 = t_n^* \|u_n^*\|$ and since $t_n^* \to 0^+$, then $\|u_n^*\| \to +\infty$.
Let $t, t_0 \in (0, r]$, with $t < \bar{t}$. By the Lebourg mean value theorem, there exists $t' \in (t, \bar{t})$ and $t^* \in \partial_C \mu_f(t')$ such that $\mu_f(t) - \mu_f(\bar{t}) = t^*(t - \bar{t})$. Since $t^* \geq \gamma$, then $\frac{\mu_f(t) - \mu_f(\bar{t})}{t - \bar{t}} \geq \gamma$. As $t, \bar{t}$ play a symetric role, we obtain the desired inequality.

2. Suppose that our assertion is not true. Then there exist $r > 0$ and $\alpha \in]0, r]$ and a sequence $(t_n) \subset [\alpha, r]$

$$\forall n \in \mathbb{N} \quad \mu_f(t_n) > n\mu_f^-(t_n, t_n)$$

and relation (25) ensures

$$\forall n \in \mathbb{N} \quad \mu_f(t_n) > n\gamma t_n.$$

Extracting subsequence if necessary, we may assume that $t_n \to \bar{t} \in [\alpha, r]$. So that

$$\lim_{n \to +\infty} \frac{\mu_f(t_n)}{t_n} = \frac{\mu_f(\bar{t})}{\bar{t}} = +\infty$$

and this contradiction completes the proof of this item.

3. As in the first item, we will argue by contradiction. Suppose there exist $r > 0$ and $\alpha \in]0, r[$ and sequences $(x_n) \subset [\alpha \leq f < r]$ and (x_n^*) such that

$$x_n^* \in \partial(\mu_f \circ f)(x_n) \text{ and } \|x_n^*\| \to 0.$$

Using subdifferential calculus rules, on gets $t_n^* \in \partial \mu_f(f(x_n))$ and $u_n^* \in \partial f(x_n)$ such that $x_n^* = t_n^* u_n^*$. Item 1 ensures the existence of $\beta > 0$ such that $t_n^* \geq \beta$, for all $n \in \mathbb{N}$. Combining this with the fact that $\|x_n^*\| \to 0$, one obtains that $\|u_n^*\| \to 0$. Now, by the inf-compactness of f, one may assume that $x_n \to \bar{x}$. So that $0 \in \partial f(\bar{x})$ and consequently $\bar{x} \in S$. This contradicts the fact that $(x_n) \subset [\alpha \leq f < r]$.

4. Note that since the function $t \mapsto \frac{\mu_f(t)}{t}$ is nonincreasing, the limit $\lim\limits_{t \to 0^+} \dfrac{\mu_f(t)}{t}$ exists in $\mathbb{R} \cup \{+\infty\}$. Relation (25) ensures the inequality $\lim\limits_{t \to 0^+} \dfrac{\mu_f(t)}{t} \geq \limsup\limits_{t \to 0^+} \mu_f^0(t, 1) \geq \limsup\limits_{t \to 0^+} \mu_f^-(t, 1)$. Let us establish the reverse inequality. Let $t > 0$. Since μ_f is continuous on $[0, t]$, Diewert mean value theorem [27], ensures the existence of $\theta \in (0, 1)$ such that

$$\frac{\mu_f(t)}{t} \leq \mu_f^-(\theta t, 1).$$

So that $\lim\limits_{t \to 0^+} \dfrac{\mu_f(t)}{t} \leq \limsup\limits_{t \to 0^+} \mu_f^-(t, 1)$, which completes the proof of the theorem.

⊠

It follows from Theorems 3.8 and 3.9 that μ_f is locally Lipchitzian and increasing on $(0, \bar{t})$ and so it is bijective from $[0, +\infty[$ onto $\mu_f\Big([0, +\infty[\Big)$. We can ask how to compute its inverse. In [43], Penot introduced the function $\varphi_f : [0, +\infty[\to [0, +\infty]$ by

$$\varphi_f(t) = \inf\{f_+(x) : \quad d_S(x) \geq t\}.$$

He shows that μ_f and φ_f are quasi-inverses in the following sense: for any $r, s \in [0, +\infty[$ one has $r \leq \mu_f(s)$, whenever $\varphi_f(r) < s$, and $s \leq \varphi_f(r)$, whenever $\mu_f(s) < r$. The aim of the following result is to show that φ_f is exactly the inverse of μ_f, that is,

$$\mu_f \circ \varphi_f(\tau) = \tau \, \forall \tau \in \mu_f([0, +\infty[) \text{ and } \varphi_f \circ \mu_f(t) = t, \quad \forall t \in [0, +\infty[.$$

Theorem 3.10 *For each $\tau \in \mu_f[0, +\infty[)$, we set $D(\tau) = argmin\{f_+(x) : \quad d_S(x) \geq \tau\}$. Then under the assumption of Theorem 3.9, we have*

1. *For all $t \in [0, +\infty[$, $D(t) \neq \emptyset$ and the function $t \mapsto \frac{\varphi_f(t)}{t}$ is nondecreasing on $]0, +\infty[$.*
2. $$\varphi_f \circ \mu_f(t) = t \quad \forall t \in [0, +\infty[\text{ and } \mu_f \circ \varphi_f(\tau) = \tau \quad \forall \tau \in \mu_f([0, +\infty[) \tag{28}$$
   and hence φ_f is increasing on $\mu_f([0, +\infty[)$.
3. *For all $\tau \in \mu_f([0, +\infty[)$*
   $$\varphi_f(\tau) = \inf\{f_+(x) : \quad d_S(x) = \tau\}, \tag{29}$$
   and hence $D(\tau) = \{u \in X : \varphi_f(\tau) = f_+(u), \quad d_S(u) = \tau\}$.
4. *φ_f is locally Lipschitzian on $\mu_f(]0, +\infty[)$.*
5. *For all $\tau_0 \in \mu_f(]0, +\infty[)$*
   $$\partial\varphi_f(\tau) \subset \bigcup_{x_0 \in D(\tau_0)} \{\tau^* \in \mathbb{R} : \tau^* > 0, \, 0 \in \partial f(x_0) - \tau^* \partial d_S(x_0)\}. \tag{30}$$
6. *For all $\tau \in \mu_f(]0, +\infty[)$*
   $$\partial_F\varphi_f(\tau) \subset \bigcap_{x_0 \in D(\tau_0)} \{\tau^* \in \mathbb{R} : \tau^* > 0, \, 0 \in \partial f(x_0) - \tau^* \partial d_S(x_0)\}. \tag{31}$$
   Moreover if $\partial_F\varphi_f(\tau) \neq \emptyset$ and f is differentiable on $D(\tau)$, then φ_f is differentiable at τ and $\varphi_f'(\tau) = \|\nabla f(x_0)\|$ for all $x_0 \in D(\tau)$ and φ_f is of class $\mathcal{C}^{1,0}$ whenever f is. Hence $\varphi_f'(0) = 0$.

Proof. 1. By definition, φ_f is nondecreasing. Let $\tau \in \mu_f([0, +\infty[)$. Since f is inf-compact, the set $D(\tau) \neq \emptyset$. The monotinicity of $t \to \frac{\varphi_f(t)}{t}$ follows from Proposition 5.1 in [26].

2. By the definitions of μ_f and φ_f, we have

$$\mu_f \circ \varphi_f(\tau) \geq \tau \, \forall \tau \in \mu_f([0, +\infty[), \text{ and } \varphi_f \circ \mu_f(t) \leq t \, \forall t \in [0, +\infty[$$

or equivalently

$$\mu_f \circ \varphi_f(\mu_f(t)) \geq \mu_f(t) \text{ and } \varphi_f \circ \mu_f(t) \leq t \, \forall t \in [0, +\infty[.$$

Suppose that $\varphi_f \circ \mu_f(t) < t$ for some $t \in [0, +\infty[$. As μ_f is increasing, $\mu_f \circ \varphi_f \circ \mu_f(t) < \mu_f(t)$, which contradicts $\mu_f \circ \varphi_f(\tau) \geq \tau$ for $\tau = \mu_f(t)$. We thus obtain the equality

$$\varphi_f \circ \mu_f(t) = t.$$

The strict monotony of φ_f follows from that of μ_f.

3. This follows from the fact that for $\tau \in \mu_f(]0, +\infty[)$ and $u \in D(\tau)$ such that

$$\varphi_f(\tau) = f_+(u)$$

with $d_S(u) > \tau$, then for all $s \in [\tau, d_S(u)]$, it hlods $\varphi_f(s) = \varphi_f(d_S(u)) = \varphi_f(\tau)$. This contradicts the strict monotony of φ_f. The proof of the other items is similar to those of Theorems 3.8 and 3.9.

$\boxtimes$

3.3 Kurdyka-Lojasiewicz-Hoffman constant

The aim of this section is to characterize the so called Kurdyka-Lojasiewicz-Hoffman constant in the spirit of the Hoffman one. There are three types of constants :

- The global Kurdyka-Lojasiewicz-Hoffman constant of the pair (f, μ_f) is defined by

$$\inf_{x \in [f>0]} d(0, \partial(\mu_f \circ f)(x)). \tag{32}$$

- The bounded Kurdyka-Lojasiewicz-Hoffman constant of the pair (f, μ_f) is defined by

$$\inf_{x \in [0<f<r]} d(0, \partial(\mu_f \circ f)(x)), \text{ for some } r > 0. \tag{33}$$

- The local Kurdyka-Lojasiewicz-Hoffman constant of the pair (f, μ_f) at $\bar{x}$ is defined by

$$\inf_{x \in [f>0] \cap B(\bar{x}, \delta)} d(0, \partial(\mu_f \circ f)(x)), \text{ for some } \delta > 0. \tag{34}$$

We introduce the following classes of functions which correspond to the two first inequalities: a function $h :]0, r[\times \mathbb{R}_+ \mapsto \mathbb{R}_+$ belong to the class $\mathcal{K}_r^{\mu_f}$ if it satisfies the following conditions:

1. for all $t \in]0, r[$, the function $s \mapsto h(t, s)$ is positively homogeneous;
2. for all $t \in]0, r[$, $\mu_f(t) \leq h(t, t)$.

For $r = +\infty$, we denote $\mathcal{K}_r^{\mu_f}$ by $\mathcal{K}_\infty^{\mu_f}$. As we will see in the last section, that it is more useful to compute $h(\cdot, \cdot)$ instead of μ_f. First note that, based on Theorem 3.9, we may assert that when, we are far from the origin, then such function satisfying the last two items exists and is equal to $\beta \mu_f^-(\cdot, \cdot)$ for some $\beta \geq 1$.
In what follows, we will give a large class of functions f showing the nonemptiness of $\mathcal{K}_r^{\mu_f}$ in finite dimension. This class is the so-called subanalytic functions (see references [30], [32], [47], [13] for more details).

Proposition 3.11 *Suppose that f is a subanalytic function which is convex and inf-compact. Then for all $r > 0$, the function μ_f is subanalytic on $[0, r]$ and there exist $p \in \mathbb{N}^*$ and $\beta > 0$ such that*

$$\mu_f(t) \leq \beta t^{\frac{1}{p}}, \quad \forall t \in [0, r].$$

So that the function $h :]0, +\infty[\times \mathbb{R}_+ \to \mathbb{R}_+$ defined by

$$h(t, s) = \beta s t^{\frac{1-p}{p}}$$

belong to $\mathcal{K}_r^{\mu_f}$.

Proof. By Proposition 1.3.7 in [47], the function μ_f is subanalytic. To conclude, it remains to use the Lojasiewic inequality (see Theorem 6.4 in [13].

$\boxtimes$

Now, we will give an other class of problems showing the nonemptiness of $\mathcal{K}_r^{\mu_f}$ in Banach spaces. Consider the minimization problem

$$\min_{x\in C} g(Ax) + \langle x^*, x\rangle \tag{35}$$

where C is a closed convex and compact polyhedral subset of X, $A : X \mapsto Y$ is a linear continuous operator between two Banach spaces X and Y with closed range, $x^* \in X^*$ and $g : Y \mapsto \mathbb{R}$ is strongly convex continuous function. The compactness property of C and the local Lipschitzness of g imply immediately the following relation:

$$\exists K > 0; \quad \sup_{x\in C} \sup_{y^*\in\partial g(Ax)} \|y^*\| \leq K. \tag{36}$$

Using the strong convexity of g, one can easily show that the solution set S of the problem (35) is closed convex and compact polyhedral. More precisely, there exists $y \in Y$ and $s \in \mathbb{R}$ such that

$$S = \{x \in C : \ Ax = y, \quad \langle x^*, x\rangle = s\rangle\}. \tag{37}$$

The classical Hoffman error bound (see [33], [34] and references therein) ensures the existence of a constant $a >$, depending only on A, x^* and C, such that

$$d(x, S)^2 \leq a(\|Ax - y\|^2 + (\langle x^*, x\rangle - s)^2)\, \forall x \in C. \tag{38}$$

Combining the last three facts, we obtain the following proposition whose proof is similar to Lemma 14 in [48] where the proof is established in finite dimension and g is differentiable.

Proposition 3.12 *Let $\bar{t} = sup_{x\in C} g(Ax) + \langle x^*, x\rangle$ and define $f : X \mapsto \mathbb{R} \cup \{+\infty\}$ by*

$$f(x) = \begin{cases} g(Ax) - g(A\bar{x}) + \langle x^*, x - \bar{x}\rangle & \text{if } x \in C \\ +\infty & \text{otherwise,} \end{cases}$$

where $\bar{x} \in S$. Then

- *there exists $\alpha > 0$ such that*

$$\mu_f(t) \leq \alpha\sqrt{t} \quad \forall 0 \leq t \leq \bar{t}; \tag{39}$$

- *for all $t \geq \bar{t}$, $\mu_f(t) = \mu_f(\bar{t})$.*

Proof. Note that $\bar{x} \in S$ IFF there exists $y^* \in \partial g(A\bar{x})$ such that

$$\langle A^*y^* + x^*, x - \bar{x}\rangle \geq 0\, \forall x \in C \tag{40}$$

and hence, by (36),

$$\langle x^*, \bar{x} - x\rangle \leq K\|Ax - A\bar{x}\| \quad \forall x \in C. \tag{41}$$

Since g is strongly convex, there exists $\rho > 0$ such that

$$g(y_1) \geq g(y_2) + \langle z^*, y_1 - y_2\rangle + \rho\|y_1 - y_2\|^2, \quad \forall y_1, y_2 \in Y, \forall z^* \in \partial g(y_2). \tag{42}$$

So that for all $x \in C$

$$\begin{aligned} \rho\|Ax - A\bar{x}\|^2 &\leq \langle A^*y^* + x^*, x - \bar{x}\rangle + \rho\|Ax - A\bar{x}\|^2 \quad (\text{ because of (40)}) \\ &= \langle y^*, Ax - A\bar{x}\rangle + \langle x^*, x - \bar{x}\rangle + \rho\|Ax - A\bar{x}\|^2 \\ &\leq g(Ax) - g(A\bar{x}) + \langle x^*, x - \bar{x}\rangle \quad (\text{ because of (42)}) \end{aligned}$$

As $\partial(g\circ A+\langle x^*,\cdot\rangle)(\bar{x})=A^*\partial g(A\bar{x})+x^*$, then $x^*+A^*y^*\in\partial(g\circ A+\langle x^*,\cdot\rangle)(\bar{x})$, where y^* is given by (40). Thus, for all $x\in C$, we obtain

$$\begin{aligned}\langle x^*,x-\bar{x}\rangle &\leq g(Ax)-g(A\bar{x})+\langle x^*,x-\bar{x}\rangle+\langle y^*,A\bar{x}-Ax\rangle\\ &\leq g(Ax)-g(A\bar{x})+\langle x^*,x-\bar{x}\rangle+K\|A\bar{x}-Ax\| \text{ (because of (36)).}\end{aligned}$$

As for all $x\in C$, $g(Ax)-g(A\bar{x})+\langle x^*,x-\bar{x}\rangle\geq 0$, then

$$\begin{aligned}(\langle x^*,x-\bar{x}\rangle)^2 &\leq (g(Ax)-g(A\bar{x})+\langle x^*,x-\bar{x}\rangle+K\|A\bar{x}-Ax\|)^2\\ &\leq (g(Ax)-g(A\bar{x})+\langle x^*,x-\bar{x}\rangle)^2+K^2\|A\bar{x}-Ax\|)^2+2K(g(Ax)-g(A\bar{x})+\langle x^*,x-\bar{x}\rangle)\|A\bar{x}-Ax\|\end{aligned}$$

.

Now, since C is compact, there exists γ_1 such that $\max(\sup_{x\in C}\|A\bar{x}-Ax\|,\sup_{x\in C}g(Ax)-g(A\bar{x})+\langle x^*,x-\bar{x}\rangle)\leq\gamma_1$. So that for all $x\in C$

$$\begin{aligned}(\langle x^*,x-\bar{x}\rangle)^2 &\leq \gamma_1(1+2K)(g(Ax)-g(A\bar{x})+\langle x^*,x-\bar{x}\rangle)+K^2\|A\bar{x}-Ax\|)^2\\ &\leq [\gamma_1(1+2K)+\frac{K^2}{\rho}](g(Ax)-g(A\bar{x})+\langle x^*,x-\bar{x}\rangle).\end{aligned}$$

By (38), we obtain for all $x\in C$

$$\begin{aligned}d(x,S)^2 &\leq a(\|Ax-A\bar{x}\|^2+(\langle x^*,x-\bar{x}\rangle)^2)\\ &\leq a(\frac{1}{\rho}+[\gamma_1(1+2K)+\frac{K^2}{\rho}])(g(Ax)-g(A\bar{x})+\langle x^*,x-\bar{x}\rangle).\end{aligned}$$

Put $\alpha=\sqrt{a(\frac{1}{\rho}+[\gamma_1(1+2K)+\frac{K^2}{\rho}])}$. Then

$$\forall t\in[0,\bar{t}],\quad \mu_f(t)\leq\alpha\sqrt{t}.$$

⊠

As a particular case of the problem (35), we consider the following generalized Lasso problem

$$\min_{x\in\mathbb{R}^n} g(Ax)+J(x) \tag{43}$$

where $g:\mathbb{R}^n\mapsto\mathbb{R}$ is a strongly convex function and A is a real $m\times n$ matrix and $J:\mathbb{R}^n\mapsto\mathbb{R}_+$ is a convex continuous function with compact and polyhedral sublevel sets. We will write this problem in the form of (35). It is obvious that $\gamma:=\inf_{x\in\mathbb{R}^n}g(Ax)\in\mathbb{R}$ and the set $S:=argmin\{g(A\cdot)+J(\cdot)\}$ is not empty. Let $R>g(0)+J(0)-\gamma$. Then ([18])

$$\begin{aligned}\min_{x\in\mathbb{R}^n} g(Ax)+J(x) &= \min\{g(Ax)+J(x):\ J(x)\leq R\}\\ &= \min_{(x,y)\in\mathbb{R}^n\times\mathbb{R}}\{g(Ax)+y: J(x)-y\leq 0,\ y\leq R\}\\ &= \min_{(x,y)\in C}\{g(\tilde{A}(x,y))+\langle x^*,(x,y)\rangle\},\end{aligned}$$

where $\tilde{A}=\begin{pmatrix}A & 0\\ 0 & 0\end{pmatrix}$ is a $(m+1\times(n+1)$-matrix, $x^*=(0,1)\in\mathbb{R}^n\times\mathbb{R}$ and $C=\{(x,y)\in\mathbb{R}^n\times\mathbb{R}:\ J(x)-y\leq 0,\ y\leq R\}$ which is convex compact and polyhedral set. This equivalent formulation allows us to say that the solution set S of the problem (43) is a convex compact and polyhedral set. More precisely, there exist $y\in\mathbb{R}^m$ and $s\in\mathbb{R}$ such that

$$S=\{x\in\mathbb{R}^n: J(x)\leq R,\ Ax=y,\ J(x)=s\}.$$

As a classical examples of these problems, we can also quote the following well known situation of the LASSO and SLOPE problems

$$g(z) = \frac{1}{2}\|z-b\|_2^2, \quad J(x) = \lambda\|x\|_1, \tag{44}$$

or more generally (see [14])

$$g(z) = \frac{1}{2}\|z-b\|_2^2, \quad J(x) = \sum_i^n \lambda_i |x|_{\downarrow i}, \tag{45}$$

where $\lambda_1 > 0$, $\lambda_1 \geq \cdots \lambda_n \geq 0$ and $|x|_{\downarrow 1} \geq \cdots \geq |x|_{\downarrow n}$ are the sorted components of x with respect to the absolute value.

3.3.1 Global characterizations

In this section we compute the Kurdyka-Lojasiewicz-Hoffman constant in terms of an abstract subdifferential. This later one satisfies the following axioms: For any lower semicontinuous function $f : X \to \mathbb{R} \cup \{\infty\}$, and any locally Lipschitz function $g : X \to \mathbb{R} \cup \{\infty\}$ and any $x \in X$:

1. $\partial f(x) \subset X^*$ and $\partial f(x) = \emptyset$ if $f(x) = \infty$;
2. $\partial g(x)$ coincides with the subdifferential in the sense of convex analysis whenever g is convex, that is
$$\partial g(x) = \{x^* \in X^* : \langle x^*, u - x\rangle \leq g(u) - g(x), \quad \forall u \in X\};$$
3. $0 \in \partial f(x)$ whenever x is a local minimum for f;
4. $\partial f(x) = \partial w(x)$ whenever f and w coincide around x;
5. $\partial(f+g)(x) \subset \partial f(x) + \partial g(x)$.

Now, we may state our main theorem in this section on global characterization of the Kurdyka-Lojasiewicz-Hoffman constant.

Theorem 3.13 *Let $f : X \to \mathbb{R} \cup \{\infty\}$ be a lower semi-continuous function and let S be the solution set of the following inequality system*

$$f(x) \leq 0. \tag{46}$$

Suppose that S is nonempty. Then

$$\inf_{x \notin S} d(0, \partial f(x)) \leq \inf_{x \notin S} \frac{f(x)}{d(x,S)}. \tag{47}$$

Moreover, if f is convex and $h \in \mathcal{K}_\infty^{\mu_f}$ then

$$\inf_{x \notin S} h(f(x), d(0, \partial f(x))) \geq \inf_{x \notin S} \frac{(\mu_f \circ f)(x)}{d(x,S)}. \tag{48}$$

Proof. The proof of the first part follows that of [9] or [34]. We will give it for completeness. Suppose that our relation (61) does not hold, that is, there exists $\alpha > 0$ such that

$$\inf_{x \notin S} d(0, \partial f(x)) > \alpha > \inf_{x \notin S} \frac{f(x)}{d(x,S)}. \tag{49}$$

Then there exists $x' \notin S$, such that

$$\alpha d(x', S) > f(x'). \tag{50}$$

Set $\varepsilon = f(x')$ and $\lambda = (\frac{1}{\alpha} + \gamma) f(x')$ where $\gamma > 0$ is such that $\lambda < d(x', S)$. Then

$$f_+(x') \leq \inf_{x \in X} f_+(x) + \varepsilon.$$

By the lower semi-continuity of f, the Ekeland's variational principle ensures the existence of $x \in X$ satisfying

$$\|x - x'\| \leq \lambda, \quad f(x) \leq f(x') \tag{51}$$

$$f_+(x) \leq f_+(u) + \frac{\varepsilon}{\lambda}\|u - x\| \quad \forall u \in X. \tag{52}$$

Note that, by (50)-(51), $x \notin S$. Since f is lower semicontinous it coincides with f_+ in a neighbourhood of x and hence, by (53) and properties1. $-$ 5., we get

$$0 \in \partial f(x) + \frac{1}{(\frac{1}{\alpha} + \gamma)} B_{X^*}$$

and this contradicts relation (49).
Suppose for the second part that there exists $\alpha > 0$ such that

$$\inf_{x \notin S} h(f(x), d(0, \partial f(x))) < \alpha < \inf_{x \notin S} \frac{(\mu_f \circ f)(x)}{d(x, S)}. \tag{53}$$

There exists $u \notin S$ and $u^* \in \partial f(u)$ such that

$$h(f(u), \|u^*\|) < \alpha < \inf_{x \notin S} \frac{(\mu_f \circ f)(x)}{d(x, S)} \tag{54}$$

and hence

$$\alpha d(u, S) < (\mu_f \circ f)(u). \tag{55}$$

As $u \notin S$, Lemma 2.1 in [11] ensures that for all $\varepsilon > 0$ there exist $u_\varepsilon \in S$, $x^*_\varepsilon \in X^*$ and $b^*_\varepsilon \in \mathbb{B}^*$ such that

(*a*) $\|u_\varepsilon - u\| \leq d(u, S) + \varepsilon^2$,

(*b*) $x^*_\varepsilon + \varepsilon b^*_\varepsilon \in (1 + \varepsilon)\partial d(u_\varepsilon, S)$,

(*c*) $\langle x^*_\varepsilon, u - u_\varepsilon \rangle = \|u_\varepsilon - u\|$.

By relation (55) and (*b*), we get

$$\langle \alpha \frac{x^*_\varepsilon + \varepsilon b^*_\varepsilon}{1 + \varepsilon}, u - u_\varepsilon \rangle \leq \alpha d(u, S) \leq (\mu_f \circ f)(u). \tag{56}$$

Since $u^* \in \partial f(u)$ and $h \in \mathcal{K}^{\mu_f}_\infty$, we get

$$f(u) \leq \langle u^*, u - u_\varepsilon \rangle$$

and

$$\mu_f(f(u)) \leq h(f(u), f(u)) \leq h(f(u), \langle u^*, u - u_\varepsilon \rangle).$$

Combining relation (56) with the last one, we get

$$\begin{aligned}\langle\alpha\frac{x_\varepsilon^*+\varepsilon b_\varepsilon^*}{1+\varepsilon},u-u_\varepsilon\rangle &\leq h(f(u),\langle u^*,u-u_\varepsilon\rangle)\\ &\leq h(f(u),\|u^*\|\cdot\|u-u_\varepsilon\|).\end{aligned}$$

Now, by using assertion (c), we obtain

$$\begin{aligned}\alpha\frac{(1-\varepsilon)}{1+\varepsilon}\|u-u_\varepsilon\| &\leq \langle\alpha\frac{x_\varepsilon^*+\varepsilon b_\varepsilon^*}{1+\varepsilon},u-u_\varepsilon\rangle\\ &\leq h(f(u),\|u^*\|\cdot\|u-u_\varepsilon\|).\end{aligned}$$

Thus

$$\alpha\frac{(1-\varepsilon)}{1+\varepsilon}\leq h(f(u),\|u^*\|)$$

and as ε is arbitrary, it follows that $\alpha\leq h(f(u),\|u^*\|)$ and this contradicts relation (54).

⊠

Similar argument leads to the following result.

Theorem 3.14 *Let $r>0$, $f:X\to\mathbb{R}\cup\{\infty\}$ be a lower semicontinuous function and let S be the solution set of the following inequality system*

$$f(x)\leq 0. \tag{57}$$

Suppose that S is nonempty. Then

$$\inf_{x\in[0<f<r]}d(0,\partial f(x))\leq\inf_{x\in[0<f<r]}\frac{f(x)}{d(x,S)}. \tag{58}$$

Moreover, if f is convex and $h\in\mathcal{K}_r^{\mu_f}$ then

$$\inf_{x\in[0<f<r]}h(f(x),d(0,\partial f(x)))\geq\inf_{x\in[0<f<r]}\frac{(\mu_f\circ f)(x)}{d(x,S)}. \tag{59}$$

3.3.2 Local characterizations

Similar argument leads also to the following local result.

Theorem 3.15 *Let $r>0$, $f:X\to\mathbb{R}\cup\{\infty\}$ be a lower semi-continuous function and let S be the solution set of the following inequality system*

$$f(x)\leq 0. \tag{60}$$

Let $\bar{x}\in X$ be such that $f(\bar{x})=0$. Then for all $s>0$,

$$\inf_{x\in[0<f<r]\cap B(\bar{x},2s)}d(0,\partial f(x))\leq\inf_{x\in[0<f<r]\cap B(\bar{x},s)}\frac{f(x)}{d(x,S)}. \tag{61}$$

Moreover, if f is convex and $h\in\mathcal{K}_r^{\mu_f}$ then

$$\inf_{x\in[0<f<r]\cap B(\bar{x},s)}h(f(x),d(0,\partial f(x)))\geq\inf_{x\in[0<f<r]\cap B(\bar{x},s)}\frac{(\mu_f\circ f)(x)}{d(x,S)}. \tag{62}$$

Proof. We duplicate the proof of the previous theorem by taking the precaution of staying in the ball.

⊠

4 Convergence analysis of an inexact descent methods for convex functions

In this section, we present a convergence analysis for an abstract inexact descent methods in convex minimization, and we demonstrate how our desingularizing function μ_f may be used as a tool for obtaining results on the complexity of such algorithms. Let $\mathcal{H}$ be (real) Hilbert space and $f : \mathcal{H} \to \mathbb{R}$ be a *convex function*, which is *continuously differentiable*. Consider the following regularization of f which will play a central role in our study,

$$H : \mathcal{H} \times \mathcal{H} \to \mathbb{R}, \quad H(x, y) = f(x) + \frac{1}{2}\|x - y\|^2, \tag{63}$$

In order not to make the notation more cumbersome, we denote the norm of the product $\mathcal{H} \times \mathcal{H}$ in the same maner as that of the space $\mathcal{H}$ itself, that is,

$$\|(x, y)\| = \sqrt{\|x\|^2 + \|y\|^2}\, \forall x, y \in \mathcal{H}.$$

In the following, consider sequences $(x_k)_{k\in\mathbb{N}}$ in $\mathcal{H}$ and $(z_k)_{k_{\in\mathbb{N}}} := (u_k, v_k)_{k\in\mathbb{N}}$ in $\mathcal{H} \times \mathcal{H}$, and $(\varepsilon_k)_{k\in\mathbb{N}}$ be an $1-$summable sequence of non-negative real numbers. Further, fix $c_1, c_2 \geq 0$ with $c_1 + c_2 \neq 0$, and assume that there are sequences $(a_k)_{k\in\mathbb{N}}$, $(b_k)_{k\in\mathbb{N}}$, and $(c_k)_{k\in\mathbb{N}}$ of non-negative real numbers such that all the previously mentioned sequences are linked together through the following hypotheses:

(H_1) For each $k \in \mathbb{N}^*$, it holds

$$a_k\|x_{k+1} - x_k\|^2 + b_k\|x_k - x_{k-1}\|^2 \leq H(z_k) - H(z_{k+1});$$

(H_2) For each $k \in \mathbb{N}^*$, one has

$$c_k\|\nabla H(z_k)\| \leq \|x_{k+1} - x_k\| + \|x_k - x_{k-1}\| + \varepsilon_k;$$

(H_3) For each $k \in \mathbb{N}^*$, for every $z = (x, x) \in \mathcal{H} \times \mathcal{H}$, one has

$$\|z_k - z\| \leq c_1\|x_k - x\| + c_2\|x_{k-1} - x\|;$$

(H_4) It holds that

$$\inf_{k\geq 0} c_k a_k =: \alpha > 0, \quad \text{and } \inf_{k\geq 0} c_k b_k =: \beta > 0.$$

In the smooth setting, these conditions can be seen as a further extension of the those proposed in [38], and also those used in [41], for nondifferentiable case. Let us examine how our conditions $(H_1) - (H_4)$ are connected to the abstract methods developped in [38, 2, 41, 28].

- Our condition (H_1) and (H_2) encompass those in the paper [38] in which the authors took the sequences considered are in fact constant, namely, $b_k = \varepsilon_k = 0$, $a_k = a > 0$, and $c_k = c > 0$.
- In [2, 28], they considered $z_k = (x_k, x_k)$, where $f(x_k) = H(z_k)$, and also with the parameters $b_k = 0$, $c_1 = \sqrt{2}$ and $c_2 = 0$. Meanwhile, the condition (H_2) in [2, 28] has the form

  $$c_k\|\nabla H(z_k)\| \leq c_k\|x_{k+1} - x_k\| + d_k;$$

  where $d_k = 0$ and $c_k = c > 0$ in [2], and $d_k \geq 0$ in [28].
- The sequence that is being examined in [41] is $z_k = (x_k, x_{k-1})$, making it a special case for our study if f is differentiable.

We also emphasize that in [2, 41], a continuity condition is introduced when the function is lower semicontinuous only. However, the presence of such a condition is unnecessary in our case as the function is continuous.

Remark 4.1 *One can observe that our conditions differ from that of [2, 28]. This distinction arises from their lack of consideration for a two-step algorithm.*

By adopting this intuitive description of descent methods, our framework encompasses several inertial gradient algorithms, such as the following examples:

Example 4.2 ***Polyak-like algorithm***
Let $f : \mathcal{H} \to \mathbb{R}$ be a continuously differentiable convex function, whose gradient is Lipschitz continuous with constant L_f. For all $k > 0$, we consider the following Polyak-like algorithm:

$$x_{k+1} = x_k + \gamma_k (x_k - x_{k-1}) - s\nabla f(x_k), \quad \text{where,} \quad s > 0, \text{ and } \quad \gamma_k \in (0, 1). \tag{64}$$

In this case, one can take $(z_k)_{k\in\mathbb{N}}$ such that $u_k = x_k$ and $v_k = x_k + \sqrt{\frac{\delta\gamma_k}{s}}(x_k - x_{k-1})$ for all $k \geq 1$, with $\delta > 0$. Moreover, we have for every $k > 0$, $H(z_k) = f(u_k) + \frac{\delta\gamma_k}{2s}\|x_k - x_{k-1}\|^2$.

Proposition 4.3 *Suppose that the real number $\delta > 1$ and satisfies*

$$\inf_{k>0} \gamma_k > 0, \quad \sup_{k>0} \gamma_k < \frac{2}{1+\delta} \quad \text{and} \quad 0 < s < \frac{2 - (1+\delta)\sup_{k>0}\gamma_k}{L_f}.$$

Then $(H1) - (H4)$ are satisfied where for all $k > 0$:

- $a_k = \frac{1}{2s}(2 - sL_f - \gamma_k - \delta\gamma_{k+1})$;
- $b_k = \frac{\gamma_k}{2s}(\delta - 1)$;
- $c_k = \left[\max\left(\frac{2}{s}, \frac{(\sqrt{\gamma_k})}{\sqrt{s}}\sqrt{\left(\frac{2\gamma_k}{s} + 3\delta\right)}\right)\right]^{-1}$ *and* $\varepsilon_k = 0$;
- $c_1 = \sqrt{(1 + 2(1 + \sup_{k>0}\gamma_k)^2)}$ *and* $c_2 = \sqrt{\frac{2\delta\sup_{k>0}\gamma_k}{s}}$.

Proof. *Let $k \in \mathbb{N}$, with $k > 0$. First of all, let us show that the hypothesis (H_1) holds. Using Descent Lemma (see[46]), we obtain*

$$f(x_{k+1}) - f(x_k) \leq \langle \nabla f(x_k), x_{k+1} - x_k\rangle + \frac{L_f}{2}\|x_{k+1} - x_k\|^2. \tag{65}$$

Using relation (64), we get

$$f(x_{k+1}) - f(x_k) \leq \frac{1}{s}\gamma_k\langle x_k - x_{k-1}, x_{k+1} - x_k\rangle + \left(\frac{L_f}{2} - \frac{1}{s}\right)\|x_{k+1} - x_k\|^2.$$

Which leads to

$$\begin{aligned} H(z_{k+1}) - H(z_k) &\leq \frac{\gamma_k}{s}\langle x_k - x_{k-1}, x_{k+1} - x_k\rangle + \left(\tfrac{L_f}{2} - \tfrac{1}{s}\right)\|x_{k+1} - x_k\|^2 + \frac{\delta}{2s}\left(\gamma_{k+1}\|x_{k+1} - x_k\|^2 - \gamma_k\|x_k - x_{k-1}\|^2\right). \\ &\leq \frac{\gamma_k}{2s}(1-\delta)\|x_k - x_{k-1}\|^2 + \frac{1}{2s}(\gamma_k + sL_f - 2 + \delta\gamma_{k+1})\|x_{k+1} - x_k\|^2. \end{aligned}$$

Therefore, the condition (H_1) is satisfied with $b_k = \frac{\gamma_k}{2s}(\delta - 1) > 0$ and

$$a_k = \frac{1}{2s}(2 - sL_f - \gamma_k - \delta\gamma_{k+1}) > \frac{1}{2s}\left(2 - (1+\delta)\sup_{k>0}\gamma_k - sL_f\right) > 0.$$

Next, for the hypothesis (H_2), it's simple to confirm that for any $x, y \in H$, one can have

$$\nabla H(x,y) = \left(\nabla f(x) + (x-y), y-x\right). \tag{66}$$

Equation (64) *implies that*

$$\nabla f(x_k) = \frac{1}{s}\left[(x_k - x_{k+1}) + \gamma_k (x_k - x_{k-1})\right], \tag{67}$$

and hence

$$\begin{aligned}
\|\nabla H(z_k)\|^2 &= \| \left(\nabla f(x_k) + \sqrt{\frac{\alpha\gamma_k}{s}}(x_k - x_{k-1}), \sqrt{\frac{\delta\gamma_k}{s}}(x_k - x_{k-1}) \right) \|^2 \\
&\leq \|\nabla f(x_k) + \sqrt{\frac{\delta\gamma_k}{s}}(x_k - x_{k-1})\|^2 + \frac{\delta\gamma_k}{s}\|x_k - x_{k-1}\|^2 \\
&\leq 2\|\nabla f(x_k)\|^2 + \frac{3\delta\gamma_k}{s}\|x_k - x_{k-1}\|^2 \\
&\leq \tfrac{4}{s^2}\|x_{k+1} - x_k\|^2 + \frac{\gamma_k}{s}\left(\frac{2\gamma_k}{s} + 3\delta\right)\|x_k - x_{k-1}\|^2.
\end{aligned}$$

Thus

$$\|\nabla H(z_k)\| \leq \frac{2}{s}\|(x_{k+1} - x_k\| + \sqrt{\frac{\gamma_k}{s}}\sqrt{\left(\frac{2\gamma_k}{s} + 3\delta\right)}\|x_k - x_{k-1}\|.$$

Then (H_2) is fulfilled , with $\varepsilon_k = 0$ and $c_k = \left[\max\left(\frac{2}{s}, \frac{(\sqrt{\gamma_k})}{\sqrt{s}}\sqrt{\left(\frac{2(\gamma_k)}{s} + 3\delta\right)}\right)\right]^{-1}$.
To verify (H_3), note that for every $z = (x,x) \in \mathcal{H} \times \mathcal{H}$, we have

$$\begin{aligned}
\|z_k - z\|^2 &\leq \|x_k - x\|^2 + \|x_k + \sqrt{\frac{\delta\gamma_k}{s}}(x_k - x_{k-1}) - x\|^2 \\
&= \|x_k - x\|^2 + \|(1 + \sqrt{\frac{\delta\gamma_k}{s}})(x_k - x) - \sqrt{\frac{\delta\gamma_k}{s}}(x_{k-1} - x)\|^2 \\
&\leq \left(1 + 2(1+\gamma_k)^2\right)\|x_k - x\|^2 + \frac{2\delta\gamma_k}{s}\|x_{k_1} - x\|^2.
\end{aligned} \tag{68}$$

Thus, the condition (H_3) holds with $c_1 = \sqrt{(1 + 2(1 + \sup_{k>0}\gamma_k)^2)}$ and $c_2 = \sqrt{\frac{2\delta \sup_{k>0}\gamma_k}{s}}$.
Finally, one can easily confirm that (H_4) is true. Indeed, we have

$$c_k > \frac{s}{2} > 0 \quad \textit{and} \quad a_k > \frac{1}{2s}\left(2 - (1+\delta)\sup_{k>0}\gamma_k - sL_f\right) > 0, b_k > \frac{\inf_{k>0}\gamma_k}{2s}(\delta - 1) > 0.$$

Therefore,

$$\inf_{k\geq 0} c_k a_k > 0, \quad \textit{and } \inf_{k\geq 0} c_k b_k > 0.$$

⊠

Example 4.4 ***Nesterov-like algorithm***
Let $f : \mathcal{H} \to \mathbb{R}$ be a continuously differentiable convex function, whose gradient is Lipschitz continuous with constant L_f. For all $k > 0$, we consider the following Nesterov-like algorithm:

$$x_{k+1} = x_k + \gamma_k(x_k - x_{k-1}) - s\nabla f\left(x_k + \gamma_k(x_k - x_{k-1})\right), \quad \textit{where,} \quad s > 0, \textit{ and } \quad \gamma_k \in (0,1). \tag{69}$$

It can be regarded as a general instance of the one in [38], where $\gamma_k = \frac{\beta k}{k+\alpha}$ for all $k > 0$. By selecting $(z_k)_{k\in\mathbb{N}}$ such that $u_k = x_k + \gamma_k(x_k - x_{k-1})$ and $v_k = x_k + (\delta+1)\gamma_k(x_k - x_{k-1})$ for all $k \geq 1$, with $\delta > 0$, we shall see under what condition on s, γ_k and $\delta > 0$, the hypotheses $(H_1) - (H_4)$ are fulfilled.

Proposition 4.5 *Suppose there exists $\delta > L_f$, $\underline{\gamma} > 0$ and $0 < \bar{\gamma} < 1$ such that $0 < \underline{\gamma} < \gamma_k < \bar{\gamma}$ for all $k \geq 0$, and*

$$\frac{1}{\delta} < s \leq \min\left(\frac{1}{L_f}, \frac{1-\bar{\gamma}^2}{1+\delta^2\bar{\gamma}^2}\right).$$

Then $(H1)-(H4)$ are satisfied with the following sequences:

- $a_k = \dfrac{1}{2s}\left(1-\gamma_{k+1}^2 - s\left(1+\delta^2\gamma_{k+1}^2\right)\right)$;
- $b_k = \dfrac{\gamma_k^2}{2s^2}\left(s+s^2\delta^2-1\right)$;
- $c_k = \frac{\sqrt{2}s}{\gamma_k\left(2+\sqrt{3}s\delta\right)}$ *and* $\varepsilon_k = 0$;
- $c_1 = \sqrt{2}\sqrt{\left(1+\bar{\gamma}\right)^2 + \left(1+(1+\delta)\bar{\gamma}\right)^2}$ *and* $c_2 = \sqrt{2}\bar{\gamma}\sqrt{1+(1+\delta)^2}$.

Proof. *Let (z_k) be defined as above. Then, for every $k > 0$, we have $H(z_k) = f(u_k) + \frac{\delta^2\gamma_k^2}{2}\|x_k - x_{k-1}\|^2$. First, let us examine the necessary condition on the parameters that satisfy (H_1). We have*

$$H(z_{k+1}) - H(z_k) = f(u_{k+1}) - f(u_k) + \frac{\delta^2\gamma_{k+1}^2}{2}\|x_{k+1}-x_k\|^2 - \frac{\delta^2\gamma_k^2}{2}\|x_k - x_{k-1}\|^2. \tag{70}$$

By using Descent Lemma (see[46]), we get

$$f(u_{k+1}) - f(u_k) \leq \langle \nabla f(u_k), u_{k+1}-u_k\rangle + \frac{L_f}{2}\|u_{k+1}-u_k\|^2. \tag{71}$$

By equation (69), we have also

$$\nabla f(u_k) = \frac{1}{s}\left[(x_k - x_{k+1}) + \gamma_k\left(x_k - x_{k-1}\right)\right]. \tag{72}$$

By using the fact that $u_k = x_k + \gamma_k\left(x_k - x_{k-1}\right)$, the following inequalities hold:

$$\begin{aligned}
\langle \nabla f(u_k), u_{k+1}-u_k\rangle =& \frac{1}{s}\langle x_k - x_{k+1}, u_{k+1}-u_k\rangle + \frac{\gamma_k}{s}\langle x_k - x_{k-1}, u_{k+1}-u_k\rangle \\
=& \frac{1}{s}\langle x_k - x_{k+1}, \left(1+\gamma_{k+1}\right)\left(x_{k+1}-x_k\right) - \gamma_k\left(x_k - x_{k-1}\right)\rangle \\
&+ \frac{\gamma_k}{s}\langle x_k - x_{k-1}, \left(1+\gamma_{k+1}\right)\left(x_{k+1}-x_k\right) - \gamma_k\left(x_k - x_{k-1}\right)\rangle \\
=& \frac{\gamma_k}{s}\langle x_{k+1}-x_k, x_k - x_{k-1}\rangle + \frac{\gamma_k\left(1+\gamma_{k+1}\right)}{s}\langle x_{k+1}-x_k, x_k-x_{k-1}\rangle - \frac{\left(1+\gamma_{k+1}\right)}{s}\|x_{k+1}-x_k\|^2 \\
&- \frac{\gamma_k^2}{s}\|x_k - x_{k-1}\|^2.
\end{aligned}$$

Since

$$\frac{\gamma_k}{s}\langle x_{k+1}-x_k, x_k-x_{k-1}\rangle \leq \frac{1}{2}\|x_{k+1}-x_k\|^2 + \frac{\gamma_k^2}{2s^2}\|x_k-x_{k-1}\|^2,$$

and

$$\frac{\gamma_k\left(1+\gamma_{k+1}\right)}{s}\langle x_{k+1}-x_k, x_k-x_{k-1}\rangle = \frac{\left(1+\gamma_{k+1}\right)^2}{2s}\|x_{k+1}-x_k\|^2 + \frac{\gamma_k^2}{2s}\|x_k-x_{k-1}\|^2 - \frac{1}{2s}\|u_{k+1}-u_k\|^2,$$

then the following equalities hold:

$$\begin{aligned}\langle \nabla f(u_k), u_{k+1}-u_k\rangle &= \frac{1}{2s}\left(s+(1+\gamma_{k+1})^2-2(1+\gamma_{k+1})\right)\|x_{k+1}-x_k\|^2+\frac{\gamma_k^2}{2s^2}(1-s)\|x_k-x_{k-1}\|^2-\frac{1}{2s}\|u_{k+1}-u_k\|^2\\ &= \frac{1}{2s}\left(s+\gamma_{k+1}^2-1\right)\|x_{k+1}-x_k\|^2+\frac{\gamma_k^2}{2s^2}(1-s)\|x_k-x_{k-1}\|^2-\frac{1}{2s}\|u_{k+1}-u_k\|^2\end{aligned} \tag{73}$$

Returning to the fundamental inequalities (70) *and* (71)*, we deduce that*

$$f(u_{k+1})-f(u_k)\leq \frac{1}{2s}\left(s+\gamma_{k+1}^2-1\right)\|x_{k+1}-x_k\|^2+\frac{\gamma_k^2}{2s^2}(1-s)\|x_k-x_{k-1}\|^2+\frac{1}{2s}\left(sL_f-1\right)\|u_{k+1}-u_k\|^2,$$

which leads to, because $0<s\leq\dfrac{1}{L_f}$*,*

$$H(z_{k+1})-H(z_k)\leq \frac{1}{2s}\left(s\left(1+\delta^2\gamma_{k+1}^2\right)+\gamma_{k+1}^2-1\right)\|x_{k+1}-x_k\|^2+\frac{\gamma_k^2}{2s^2}\left(1-s-s^2\delta^2\right)\|x_k-x_{k-1}\|^2+\frac{1}{2s}\left(sL_f-1\right)\|u_{k+1}-u_k\|^2.$$

Hence

$$H(z_{k+1})-H(z_k)\leq \underbrace{\frac{1}{2s}\left(s\left(1+\delta^2\gamma_{k+1}^2\right)+\gamma_{k+1}^2-1\right)}_{=-a_k}\|x_{k+1}-x_k\|^2+\underbrace{\frac{\gamma_k^2}{2s^2}\left(1-s-s^2\delta^2\right)}_{=-b_k}\|x_k-x_{k-1}\|^2$$

Since $\dfrac{1}{\delta}<s\leq\dfrac{1-\bar{\gamma}}{1+\delta^2\bar{\gamma}}$*, then the sequences* $(a_k)_{k\in\mathbb{N}}$*,* $(b_k)_{k\in\mathbb{N}}$ *are both positive, and condition* (H_1) *is met.*
Now, let's check the fulfillment of (H_2)*. Using equation* (66) *as in the previous example, we have for* $k>0$

$$\nabla H(u_k,v_k)=\left(\nabla f(u_k)+(u_k-v_k),v_k-u_k\right),$$

hence, for every $k>0$

$$\begin{aligned}\|\nabla H(u_k,v_k)\| &\leq \sqrt{\|\nabla f(u_k)+(u_k-v_k)\|^2+\|v_k-u_k\|^2}\\ &\leq \sqrt{2\|\nabla f(u_k)\|^2+2\|(u_k-v_k)\|^2+\|v_k-u_k\|^2}\\ &= \sqrt{2\|\nabla f(u_k)\|^2+\frac{3\delta^2\gamma_k^2}{2}\|x_k-x_{k-1}\|^2}\\ &\leq \sqrt{2}\|\nabla f(u_k)\|+\frac{\sqrt{3}\delta\gamma_k}{\sqrt{2}}\|x_k-x_{k-1}\|.\end{aligned} \tag{74}$$

Combining this last one with equation (72) *yields*

$$\begin{aligned}\|\nabla H(u_k,v_k)\| &\leq \tfrac{\sqrt{2}}{s}\|(x_k-x_{k+1})+\gamma_k(x_k-x_{k-1})\|+\frac{\sqrt{3}\delta\gamma_k}{\sqrt{2}}\|x_k-x_{k-1}\|\\ &\leq \tfrac{\sqrt{2}}{s}\|x_k-x_{k+1}\|+\gamma_k\left(\frac{\sqrt{2}}{s}+\frac{\sqrt{3}\delta}{\sqrt{2}}\right)\|x_k-x_{k-1}\|\\ &\leq \gamma_k\left(\frac{\sqrt{2}}{s}+\frac{\sqrt{3}\delta}{\sqrt{2}}\right)\left[\|x_k-x_{k+1}\|+\|x_k-x_{k-1}\|\right].\end{aligned} \tag{75}$$

Therefore, by taking $c_k=\frac{\sqrt{2}s}{\gamma_k\left(2+\sqrt{3}s\delta\right)}$ *and* $\varepsilon_k=0$ *for all* $k>0$*, our condition* (H_2) *is satisfied. As in the previous example in* (68)*, we note that* (H_3) *can be easily verified with* $c_1=\sqrt{2}\sqrt{(1+\bar{\gamma})^2+\left(1+(1+\delta)\bar{\gamma}\right)^2}$ *and* $c_2=\sqrt{2}\bar{\gamma}\sqrt{1+(1+\delta)^2}$*. Indeed, for every* $z=(x,x)\in\mathcal{H}\times\mathcal{H}$*, and for all* $k>0$*, we have*

$$\|z_k-z\|^2 \leq 2\left(\left(1+\gamma_k\right)^2+\left(1+(1+\delta)\bar{\gamma}\right)^2\right)+\|x_k-x\|^2+2\gamma_k^2\left(1+(1+\delta)^2\right)\|x_{k-1}-x\|^2. \tag{76}$$

Finally, assuming that $0 < \underline{\gamma} < \gamma_k < \bar{\gamma}$, *and* $\frac{1}{\delta} < s \leq \min\left(\frac{1}{L_f}, \frac{1-\bar{\gamma}^2}{1+\delta^2\bar{\gamma}^2}\right)$, *it is straightforward to observe that* (H_4), *is met too.*

⊠

Remark 4.6 *It is worth mentioning that when it comes to choose the sequences* $(u_k)_{k\in\mathbb{N}}$, *and* $(v_k)_{k\in\mathbb{N}}$ *there are several possibilities. As mentioned in [38], the sequence* $(H(z_k))_{k\in\mathbb{N}}$ *can be considered as a discretization of the total energy* E *of the continuous dynamical systems in [4, 22, 23, 28] which is given by*

$$E : [t_0, +\infty) \to \mathbb{R}, \quad E(t) = f(x(t)) + \frac{1}{2}\|\dot{x}(t)\|^2.$$

In fact, the explicit discretization of E *gives for* $k > 0$

$$E_k = f(x_k) + \frac{1}{2}\|x_k - x_{k-1}\|^2 = H\left(x_k, x_{k-1}\right),$$

which is a special case of our approach , and it was the main focus in [41]. Indeed, the choice of $(u_k)_{k\in\mathbb{N}}$ *and* $(v_k)_{k\in\mathbb{N}}$, *depends on the type of discretization of* $E(t)$. *In matter of fact, the various choices of* $(u_k)_{k\in\mathbb{N}}$ *and* $(v_k)_{k\in\mathbb{N}}$ *arise from the fact that one can take each one of them as linear combinations of* x_k *and* x_{k-1}, *which opens the door to numerous possibilities.*

Remark 4.7 *Suppose that* $f : \mathcal{H} \to \mathbb{R}$ *is a convex function, which is continuously differentiable and* $argminf \neq \emptyset$, *then* H *is also continuously differentiable and convex. Besides, we have*

$$argminH = \{(x, y) \in \mathcal{H} \times \mathcal{H} : \nabla H(x, y) = (0, 0)\} = \{(\bar{x}, \bar{x}) \in \mathcal{H} \times \mathcal{H} : \bar{x} \in argmin(f)\}.$$

4.1 Convergence results

In this subsection,, we use the desingularising function μ_H to ensure that the algorithm of interest has a finite length, which means that $\sum_{k=1}^{+\infty} \|x_k - x_{k-1}\| < +\infty$, and so the sequences produced by the procedure in $(H_1)-(H_3)$ exhibit strong convergence to minimum points of f.

Theorem 4.8 *Assume that* $f : \mathcal{H} \to \mathbb{R}$ *is a convex and inf-compact function, which is continuously differentiable,* $argminf \neq \emptyset$ *and* $min_{\mathcal{H}} f = 0$. *Let* H *be the convex function defined in* (63) . *Consider the sequences* $(x_k)_{k_\in\mathbb{N}}$, $(u_k)_{k_\in\mathbb{N}}$, $(v_k)_{k_\in\mathbb{N}}$ *and let* $(z_k)_{k\in\mathbb{N}} = (u_k, v_k)_{k_\in\mathbb{N}}$ *be a sequence that satisfies the conditions* $(H_1), (H_2)$, (H_3) *and* (H_4). *Suppose that* $\inf_{k>0} c_k > 0$, *and that there exists* $r > 0$ *such that* μ_H *is concave on* $]0, r[$ *and* $H(u_0, v_0) < r$. *Then,* $\sum_{k=1}^{+\infty} \|x_k - x_{k-1}\| < +\infty$, *and* (x_k) *converges strongly to some* $\bar{x} \in argminf$. *Furthermore, for all* $k \geq 0$,

$$\|x_k - \bar{x}\| \leq \frac{3}{\min(\alpha, \beta)}\mu_H(H(z_k)). \tag{77}$$

Proof. Using (H_1), one can deduce that the sequence $(H(z_k))_{k_\in\mathbb{N}}$, is nonincreasing, thus $z_k \in [0 \leq H < r]$. Suppose that there exists $\bar{k} > 0$ such that $H(z_{\bar{k}}) = 0$, since $(H(z_k))_{k_\in\mathbb{N}}$, is nonincreasing, then

$$H(z_k) \leq H(z_{\bar{k}}) = 0, \quad \forall k \geq \bar{k},$$

hence

$$H(z_k) = 0, \quad \forall k \geq \bar{k}.$$

which leads, by (H_1), to

$$\|x_{k+1} - x_k\| = 0, \quad \forall k \geq \bar{k},$$

Consequently, $\sum_{k>1} \|x_k - x_{k-1}\| < +\infty$, and the sequence $(x_k))_{k_{\in \mathbb{N}}}$, is constant for all $k \geq \bar{k}$.
Now, let us suppose that for $k > 0$, $H(z_k) > 0$. which means that $z_k \in [0 < H < r]$. Since f is differentiable, then by Theorem 3.9, for all $x \notin ArgminH$, μ_H is differentiable at $H(x)$, besides

$$\mu'_H(H(x))\|\nabla H(x)\| = 1. \tag{78}$$

Using the concavity assumption on μ_H, together with the last equality, we get for $k > 0$

$$\begin{aligned} \mu_H(H(z_k)) - \mu_H(H(z_{k+1}) &\geq \mu'_H(H(z_k))\left(H(z_k) - H(z_{k+1})\right) \\ &= \frac{\left(H(z_k) - H(z_{k+1})\right)}{\|\nabla H(z_k)\|}. \end{aligned}$$

By combining this last inequality with (H_1) and (H_2), we obtain for $k > 0$

$$\frac{c_k a_k \|x_{k+1} - x_k\|^2 + c_k b_k \|x_k - x_{k-1}\|^2}{\|x_{k+1} - x_k\| + \|x_k - x_{k-1}\| + \varepsilon_k} \leq \left[\mu_H(H(z_k)) - \mu_H(H(z_{k+1})\right].$$

By condition (H_4), $\inf_{k\geq 0} c_k a_k =: \alpha > 0$, and $\inf_{k\geq 0} c_k b_k =: \beta > 0$. By taking $a = \min(\alpha, \beta)$, we get

$$\frac{\|x_{k+1} - x_k\|^2 + \|x_k - x_{k-1}\|^2}{\|x_{k+1} - x_k\| + \|x_k - x_{k-1}\| + \varepsilon_k} \leq \frac{1}{a}\left[\mu_H(H(z_k)) - \mu_H(H(z_{k+1})\right].$$

Using the fact that

$$\frac{1}{3}\left(\|x_{k+1} - x_k\| + \|x_k - x_{k-1}\| + \varepsilon_k\right)^2 \leq \|x_{k+1} - x_k\|^2 + \|x_k - x_{k-1}\|^2 + \varepsilon_k^2,$$

one can conclude that for $k > 0$

$$\begin{aligned} \|x_{k+1} - x_k\| + \|x_k - x_{k-1}\| + \varepsilon_k &\leq \tfrac{3}{a}\left[\mu_H(H(z_k)) - \mu_H(H(z_{k+1})\right] + \frac{3\varepsilon_k^2}{\|x_{k+1} - x_k\| + \|x_k - x_{k-1}\| + \varepsilon_k} \\ &\leq \tfrac{3}{a}\left[\mu_H(H(z_k)) - \mu_H(H(z_{k+1})\right] + 3\varepsilon_k. \end{aligned} \tag{79}$$

Hence,

$$\|x_{k+1} - x_k\| + \|x_k - x_{k-1}\| \leq \frac{3}{a}\left[\mu_H(H(z_k)) - \mu_H(H(z_{k+1})\right] + 2\varepsilon_k. \tag{80}$$

Thus, for all $n \in \mathbb{N}^*$

$$\sum_{k=1}^{n} \|x_{k+1} - x_k\| \leq \frac{3}{a}\left[\mu_H(H(z_1)) - \mu_H(H(z_{k+1})\right] + 2\sum_{k=1}^{n} \varepsilon_k.$$

Therefore, by letting $n \to +\infty$, and since $\sum_{k=1}^{+\infty} \varepsilon_k < +\infty$, we conclude that $\sum_{k=1}^{+\infty} \|x_{k+1} - x_k\| < +\infty$, and so, by the Cauchy criterion, the sequence $(x_k)_{k\in\mathbb{N}}$ is convergent. Let $\bar{x} \in \mathcal{H}$ be the limit of (x_k). Using condition (H_3) we have for $z = (\bar{x}, \bar{x})$

$$\lim_{k\leftarrow+\infty} \|z_k - z\| \leq \left(c_1 \lim_{k\leftarrow+\infty} \|x_k - \bar{x}\| + c_2 \lim_{k\leftarrow+\infty} \|x_{k-1} - \bar{x}\|\right) = 0.$$

Hence, $(z_k)_{k\in\mathbb{N}}$ converges to $z = (\bar{x}, \bar{x})$. Now, by conditions (H_2), and (H_4), we have $\inf_{k>0} c_k > 0$, and so one can conclude that

$$\|\nabla H(z)\| = \lim_{k\leftarrow+\infty} \|\nabla H(z_k)\| = 0.$$

Therefore $z = (\bar{x}, \bar{x}) \in argminH$, thus $\bar{x} \in argminf$.
By summing in (79) from $i = k$ to $i = k + m$, we obtain

$$\sum_{i=k}^{k+m} \|x_{i+1} - x_i\| \leq \frac{3}{a} \left(\mu_H(H(z_k)) - \mu_H(H(z_{k+m}))\right).$$

By letting m going to infinity, we get for $k > 0$

$$\|x_k - \bar{x}\| \leq \frac{3}{a}\mu_H(H(z_k)).$$

The proof is then complete.

⊠

4.2 Complexity for inertial gradient descent sequences and λ-one-dimensional worst-case proximal sequence

This subsection delves into the complexities of inertial gradient descent algorithms and its relationship with λ-one-dimensional worst-case proximal sequences as in [18]. Let us begin with a definition and some properties of proximal mapping.

Definition 4.9 *[12] Let $g : \mathcal{H} \to \mathbb{R}$ be a lower semi-continuous convex function and let λ be a positive real. The proximal mapping $prox_{\lambda g} : \mathcal{H} \to \mathcal{H}$ of g is defined by*

$$prox_{\lambda g}(x) = argmin_{u \in \mathbb{H}} \left\{ g(u) + \frac{1}{2\lambda}\|u - x\|^2 \right\}.$$

Let $r > 0$ such that $\mu_H : [0, r] \to \mathbb{R}^+$ is concave. Set $r_0 = \mu_H^{-1}(r)$ and consider the function (see Theorem 3.10) $\varphi_H = \left(\mu_H|_{[0,r]}\right)^{-1} : [0, r_0] \to, [0, r]$ which is increasing and convex. Starting from $\alpha_0 = \mu_H(r_0) > 0$, and for $\lambda > 0$, and $k \geq, 0$ we define the λ- one-dimensional worst-case proximal sequence inductively by

$$\alpha_{k+1} = argmin \left\{ \varphi_H(u) + \frac{1}{2\lambda}(u - \alpha_k)^2 : u \geq 0 \right\} \tag{81}$$

According to Theorem 3.10, for $t \in [0, r_0]$ the function φ_H is exactly $\varphi_H(t) = \inf\{H(x) : \quad d_{S_H}(x) = t\}$, is continuous convex, and $\varphi_H(0) = 0$ then the sequence is well defined and positive for each for $k \geq 0$. Moreover, for all $k \geq 0$, the sequence can be written as

$$\alpha_{k+1} = prox_{\lambda\varphi_H}(\alpha_k)$$

and so α_k is decreasing and converges to zero. For the following Theorem, we will need to following Lemma :

Lemma 4.10 *Let $\varphi : [0, \bar{r}] \to, [0, r]$ be a continuous convex function, and $\lambda_1, \lambda_2 > 0$ such that $\lambda_1 > \lambda_2$, then, for $t > 0$ we have*

$$prox_{\lambda_1\varphi}(t) \leq prox_{\lambda_2\varphi}(t).$$

Theorem 4.11 *Assume that $f : \mathcal{H} \to \mathbb{R}$ is a convex and inf-compact function, which is continuously differentiable. Let H be the convex function defined in (63) , and consider the sequences $(x_k)_{k_\in\mathbb{N}}$, $(u_k)_{k_\in\mathbb{N}}$, $(v_k)_{k_\in\mathbb{N}}$ and $(z_k)_{k_\in\mathbb{N}} = (u_k, v_k)_{k_\in\mathbb{N}}$ satisfying conditions $(H_1), (H_2)$, (H_3), and (H_4). Furthermore, suppose that for all $k > 0$, $\varepsilon_k = 0$, and there exists $r > 0$, such that μ_H is concave on $]0, r[$, and $H(u_0, v_0) := r_0 < r$. Then, $(x_k)_{k_\in\mathbb{N}}$ converges strongly to some $x^* \in argminf$. Moreover, for all $k \geq 0$ we have*

$$\sum_{k>0} c_k\|\nabla H(z_k)\|^2 < +\infty. \tag{82}$$

$$H(z_{k+1}) \leq \varphi_H(\alpha_{k+1}), \tag{83}$$

$$\|x_{k+1} - \bar{x}\| \leq \frac{2c}{\min(\alpha, \beta)} \alpha_{k+1}. \tag{84}$$

Where α_k is the λ-one-dimensional worst-case proximal sequence defined above in (81)*, with* $\lambda = \frac{\min(\alpha,\beta)}{2c}$.

Proof. With some modifications, we will follow the same steps as in [18]. Let us set $r_k = H(z_k)$ and $z = (\bar{x}, \bar{x})$. If there exists $\bar{k} \geq 1$ such that $r_{\bar{k}} = 0$, then one can prove using condition (H_1) that the sequence $(x_k)_{k \geq \bar{k}}$ is constant.
For all $k \geq 0$, let $r_{\bar{k}} > 0$. Set $\beta_k := \mu_H(r_k) = \varphi_H^{-1}(r_k)$ and $s_{k+1} = (\beta_k - \beta_{k+1})\, \mu'_H(r_{k+1})$, which means that

$$\beta_{k+1} = prox_{s_{k+1}\varphi_H}(\beta_k). \tag{85}$$

We need to prove that $\lambda \leq s_k$, where $\lambda = \frac{\min(\alpha,\beta)}{2c}$. In doing so, we use condition (H_2) combined with (78), with $\varepsilon_k = 0$, we get for all $k \geq 0$,

$$c_k = \mu'_H(r_k)\, c_k \|\nabla H(z_k)\| \leq \mu'_H(r_k)\, (\|x_{k+1} - x_k\| + \|x_k - x_{k-1}\|),$$

consequently,

$$c_k \leq \mu'_H(r_k)\, (\|x_{k+1} - x_k\| + \|x_k - x_{k-1}\|),$$

hence

$$c_k^2 \leq 2\mu'_H(r_k)^2 \left(\|x_{k+1} - x_k\|^2 + \|x_k - x_{k-1}\|^2\right). \tag{86}$$

By (H_1), we have

$$\min(\alpha, \beta) \left(\|x_{k+1} - x_k\|^2 + \|x_k - x_{k-1}\|^2\right) \leq c_k a_k \|x_{k+1} - x_k\|^2 + c_k b_k \|x_k - x_{k-1}\|^2 \leq c_k\, (H(z_k) - H(z_{k-1})),$$

which yields to

$$\|x_{k+1} - x_k\|^2 + \|x_k - x_{k-1}\|^2 \leq \frac{c_k}{\min(\alpha, \beta)} (H(z_k) - H(z_{k+1})). \tag{87}$$

Hence, by merging this inequality with equation (86), we conclude that

$$\frac{c_k \min(\alpha, \beta)}{2} \leq \mu'_H(r_k)^2\, (H(z_k) - H(z_{k+1})), \tag{88}$$

and hence for all $n \in \mathbb{N}^*$,

$$\sum_{k=1}^{n} \frac{c_k}{\mu'_H(r_k)^2} \leq \frac{2}{\min(\alpha, \beta)} \sum_{k=1}^{n} [\mu_H(H(z_k)) - \mu_H(H(z_{k+1})],$$

by letting $n \to +\infty$, we conclude that $\sum_{k>0} c_k \|\nabla H(z_k)\|^2 = \sum_{k>0} \frac{c_k}{\mu'_H\left(H(z_k)\right)^2} < +\infty$.
On the other hand, taking into account that $\inf_{k>0} c_k = c > 0$, we obtain from (88),

$$\frac{c \min(\alpha, \beta)}{2} \leq \mu'_H(r_k)^2\, (\varphi_H(\beta_k) - \varphi_H(\beta_{k+1})).$$

Using the convexity of φ_H, and the fact that $\mu'_H(r_k) = \dfrac{1}{\varphi'_H(\beta_k)}$, we have for all $k \geq 0$,

$$\begin{aligned} \frac{c \min(\alpha, \beta)}{2} &\leq \mu'_H(r_k)^2\, (\varphi_H(\beta_k) - \varphi_H(\beta_{k+1})), \\ &\leq \mu'_H(r_k)\, (\beta_k - \beta_{k+1}). \end{aligned}$$

Since μ_H is concave , and r_k is non increasing, then, by the monotonicity of $\mu_H^{'}$, we have $\mu_H'(r_k)^2 \leq \mu_H'(r_{k+1})^2$. Therefore

$$\frac{c\min(\alpha,\beta)}{2} \leq \mu_H'(r_{k+1})(\beta_k - \beta_{k+1}). \tag{89}$$

Hence

$$\frac{c\min(\alpha,\beta)}{2} \leq s_{k+1}.$$

By Lemma 4.10, we conclude that $prox_{s_{k+1}}(\beta_k) \leq prox_{\frac{\min(\alpha,\beta)}{2c}}(\alpha_k)$, that is

$$\beta_{k+1} \leq \alpha_{k+1},$$

which implies that for all $k \geq 0$,

$$H(z_{k+1}) \leq \varphi_H(\alpha_{k+1}),$$

and so, by Theorem 4.8,

$$\|x_{k+1} - \bar{x}\| \leq \frac{3}{\min(\alpha,\beta)}\alpha_{k+1}. \tag{90}$$

Where α_k is $\frac{c\min(\alpha,\beta)}{2}$-one-dimensional worst-case proximal sequence defined above in (81). ⊠

In case when $(\varepsilon_k)_{k\in\mathbb{N}} \neq 0$, one can get the same results (82),(83) and (84), but the sequence α_k will be $\frac{c\min(\alpha,\beta)}{2}$-one-dimensional worst-case proximal sequence defined by

$$\alpha_{k+1} = prox_{\lambda\varphi_H}\left(\alpha_k + \left(\varphi_H^{'}(\alpha_k)\right)^{-1}\varepsilon_k^2\right), \quad \forall k > 0.$$

Proposition 4.12 *Let $(\varepsilon_k)_{k\in\mathbb{N}} \neq 0$. Under assumptions of the last Theorem and if further, $\sum_{k>0}\varepsilon_k < +\infty$, then, we have for $k > 0$,*

$$\sum_{k>0} c_k\|\nabla H(z_k)\|^2 < +\infty \tag{91}$$

$$H(z_{k+1}) \leq \varphi_H(\alpha_{k+1}), \tag{92}$$

$$\|x_{k+1} - \bar{x}\| \leq \frac{2c}{\min(\alpha,\beta)}\alpha_{k+1}. \tag{93}$$

where α_k is $\frac{c\min(\alpha,\beta)}{2}$-one-dimensional worst-case proximal sequence defined for all $k \geq 0$, by

$$\alpha_{k+1} = prox_{\lambda\varphi_H}\left(\alpha_k + \left(\varphi_H^{'}(\alpha_k)\right)^{-1}\varepsilon_k^2\right).$$

Proof. For all $k \geq 0$, let $r_{\bar{k}} > 0$. Set $\beta_k = \varphi_H^{-1}(r_k)$, we take this time $s_{k+1} = \left(\beta_k + \left(\varphi_H^{'}(\beta_k)\right)^{-1}\varepsilon_k^2 - \beta_{k+1}\right)\mu_H'(r_{k+1})$, this means that

$$\beta_{k+1} = prox_{s_{k+1}\varphi_H}\left(\beta_k + \left(\varphi_H^{'}(\beta_k)\right)^{-1}\varepsilon_k^2\right). \tag{94}$$

Following the same step of the proof of last theorem, the inequality (86) becomes,

$$c_k^2 \leq 3\mu_H'(r_k)^2\left(\|x_{k+1} - x_k\|^2 + \|x_k - x_{k-1}\|^2 + \varepsilon_k^2\right). \tag{95}$$

Combining with inequality (87), we obtain

$$\frac{c_k \min(\alpha,\beta)}{3} \leq \mu'_H(r_k)^2 \left(H(z_k) - H(z_{k+1}) + \frac{\varepsilon_k^2}{\min(\alpha,\beta)\, c_k} \right), \tag{96}$$

Consequently, we obtain,

$$\frac{c \min(\alpha,\beta)}{3} \leq \mu'_H(r_k)^2 \left(H(z_k) - H(z_{k+1}) + \frac{\varepsilon_k^2}{\min(\alpha,\beta)\, c} \right).$$

Similar to inequality (89), one can conclude that

$$\begin{aligned} \frac{c \min(\alpha,\beta)}{2} &\leq \mu'_H(r_{k+1}) \left(\beta_k + \varepsilon_k^2 \mu'_H(r_k) - \beta_{k+1} \right) \\ &= \mu'_H(r_{k+1}) \left(\beta_k + \left(\varphi'(\beta_k) \right)^{-1} \varepsilon_k^2 - \beta_{k+1} \right) \\ &= s_{k+1}. \end{aligned}$$

As a result, as demonstrated previously in the proof of Theorem 4.11, the three inequalities are fulfilled where α_k is $\frac{c \min(\alpha,\beta)}{2}$-one-dimensional worst-case proximal sequence defined in (94).

Remark 4.13 *The inequalities in the final proposition do not provide insight into the speed of convergence, except if the convergence sequence α_k converge to the unique minimum of φ_H.*

4.3 Conclusion

In this paper, we have presented different properties of the desingularizing function μ_f, and showed that this function may be neither concave nor convex even if f is convex differentiable. This, in turn, motivates further researchs aimed at understanding conditions on f that lead to μ_f to be concave.
Section 3 mainly focuses on convex differentiable functions in order to exploit the advantageous qualities of μ_H when f is differentiable, but it is crucial to point out that similar results can also be applied to convex nondifferentiable functions. Thus, the pursuit of extending our method to nondifferentiable convex (nonconvex) functions has the potential to improve theory as well as practical applications in optimisation and related domains.

References

[1] H. Attouch and J. Bolte, On the convergence of the proximal algorithm for nonsmooth functions involving analytic features, Mathematical Programming, 116 (2009), pp. 5–16.

[2] H. Attouch, J. Bolte, B.F. Svaiter, Convergence of descent methods for semi-algebraic and tameproblems: proximal algorithms, forward-backward splitting, and regularized Gauss-Seidel methods,Mathematical Programming, 137(1-2), 91-129, 2013.

[3] Attouch, H., Bolte, J., Redont, P., Soubeyran, A.: Proximal alternating minimization and projection methods for nonconvex problems: an approach based on the Kurdyka-Łojasiewicz inequality. Math. Oper.Res. 35(2), 438–457 (2010)

[4] H. Attouch, Z. Chbani, J. Peypouquet, P. Redont, Fast convergence of inertial dynamics and algorithms with asymptotic vanishing viscosity, Mathematical Programming, 168(1-2) Series B, 123-175, 2018

[5] Attouch, H., Goudou, X., Redont, P. (2000). The heavy ball with friction method, I. The continuous dynamical system: global exploration of the local minima of a real-valued function by asymptotic analysis of a dissipative dynamical system. Communications in Contemporary Mathematics, 2(01), 1-34

[6] H. Attouch, J. Peypouquet, P. Redont, A Dynamical Approach to an Inertial Forward-Backward Algorithm for Convex Minimization, SIAM Journal on Optimization, 24(1), 232-256, 2014

[7] H. Attouch, R. J.-B. Wets. Quantitative stability of variational systems: II. A framework for nonlinear conditioning. SIAM J. Optim. 3 (1992), 359-381.

[8] D. Azé, A survey on error bounds for lower semicontinuous functions. Proceedings of 2003 MODE-SMAI Conference, 1–17, ESAIM Proc., 13, EDP Sci., Les Ulis, 2003.

[9] D. Azé, J.-N. Corvellec, Characterizations of error bounds for lower semicontinuous functions on metric spaces, ESAIM: Control, Optimisation and Calculus of Variations, 10 (2004), 409–425.

[10] S. Banert and R. I. Bot, A general double-proximal gradient algorithm for d.c. programming, Mathematical Programming, 178 (2019), pp. 301–326.

[11] A. Barbara, A. Jourani, Error bound characterizations of Guignard's constraint qualification in convex programming, SIAM J. Optim., 32 (2022), 2013–2040.

[12] Bauschke, H.H., Combettes, P.L., Convex analysis and monotone operator theory in hilbert spaces. CMS Books in Mathematics, Springer, New York (2011)

[13] E. Bierstone, P. D. Milman, Semianalytic and subanalyticsets, Publications de l'IHES,67, (1988), 5-42.

[14] Bogdan M., Van Den Berg E., Sabatti C., Su W., Candès E.-J., Slope—adaptive variable selection via convex optimization.The annals of applied statistics, 9 (2015), 1103.

[15] Bolte J., Daniilidis A., Lewis A., The Lojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems, SIAM J. Optim. 17, 1205–1223 (2006).

[16] Bolte, J., Daniilidis, A., Ley, O., Mazet, L., Characterizations of Łojasiewicz inequalities: subgradient flows, talweg, convexity. Trans. Amer. Math. Soc. 362(6), 3319–3363 (2010).

[17] Bolte J., Daniilidis A., Lewis A., Shiota M., Clarke subgradients of stratifiable functions, SIAM J. Optim. 18, no. 2, 556–572 (2007)

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

[19] J. Bolte, S. Sabach, and M. Teboulle, Proximal alternating linearized minimization for nonconvex and nonsmooth problems, Mathematical Programming, 146 (2014),pp. 459–494.

[20] Bonettini, S., Ochs, P., Prato, M., Rebegoldi, S. (2023). An abstract convergence framework with application to inertial inexact forward–backward methods. Computational Optimization and Applications, 84(2), 319-362.

[21] S. Bonettini, M. Prato, S. Rebegoldi, New convergence results for the inexact variable metric forward–backward method, Applied Mathematics and Computation, Volume 392, 2021, 125719, ISSN 0096-3003.

[22] R.I. Bot¸, E.R. Csetnek, S.C. László, Approaching nonsmooth non-convex minimization through second-order proximal-gradient dynamical systems, Journal of Evolution Equations, 18(3), 1291- 1318, 2018

[23] R.I. Bot¸, E.R. Csetnek, S.C. László, An inertial forward-backward algorithm for minimizing the sum of two non-convex functions, Euro Journal on Computational Optimization, 4(1), 3-25, 2016

[24] R. I. Bot, M. N. Dao, and G. Li, Extrapolated proximal subgradient algorithms for nonconvex and nonsmooth fractional programs, arXiv:2003.04124, (2020).

[25] F.H. Clarke (1983), Optimization and nonsmooth analysis, Wiley-Interscience, New-York.

[26] O. Cornejo, A. Jourani, C. Zalinescu, Conditioning and upper-Lipschitz inverse subdifferentials in nonsmooth optimization problems, J. Optim. Th. Appl., 95 (1997), 127-148.

[27] W.E. Diewert W.E., Alternative characterizations of six kinds of quasiconcavity in the nondifferentiable case with applications to nonsmooth programming, in "Generalized concavity in optimization and economics", Schaible S. and Ziemba W.T. (ed.s), Academic Press, (1981), 51-93.

[28] P. Frankel, G. Garrigos, J. Peypouquet, Splitting Methods with Variable Metric for Kurdyka- Lojasiewicz Functions and General Convergence Rates, Journal of Optimization Theory and Applications,165(3), 874-900, 2015.

[29] Goudou, X., Munier, J. (2009). The gradient and heavy ball with friction dynamical systems: the quasiconvex case. Mathematical Programming, 116(1-2), 173-191.

[30] R. M. Hardt. Stratification of real analytic mappings and images. Invent. Math. 28 (1975).

[31] J.-B. Hiriart-Urruty, J. Ponstein, Generalized differentiability/duality and optimization for problems dealing with differences of convex functions. In: Convexity and Duality in Optimization: Proceedings of the Symposium on Convexity and Duality in Optimization Held at the University of Groningen, The Netherlands June 22, 1984, pp. 37–70. Springer, Berlin (1985).

[32] H. Hironaka, Subanalytic sets, Number theory, algebraic geometry and commutative algebra. In honor of Y. Akizuki, Tokyo (1973).

[33] A. D. Ioffe, Regular points of Lipschitz functions, Trans. Amer. Math. Soc., 251 (1979), 61-69.

[34] A. Jourani, Hoffman's error bound, local controllability and sensitivity analysis, SIAM J. Control Optim., 38 (2000), 947-970.

[35] K. Kurdyka, On gradients of functions definable in o-minimal structures. Ann. l'inst. Fourier (Grenoble)48(3), 769–783 (1998).

[36] A. Ya. Kruger, On Fréchet subdifferentials, Journal of Mathematical Sciences, Vol. 116, No. 3, 2003.

[37] S. Łojasiewicz, Une propriété topologique des sous-ensembles analytiques réels, pp. 87–89. Les Équations aux Dérivées Partielles, Éditions du Centre National de la Recherche Scientifique, Paris (1963)

[38] S.C. László, Convergence rates for an inertial algorithm of gradient type associated to a smooth non-convex minimization. Math. Program. 190, 285–329 (2021).

[39] B. S. Mordukhovich, Variational Analysis and Generalized Differentiation. I: Basic Theory, Grundlehren Series (Fundamental Principles of Mathematical Sciences), Vol. 330, 584 pp., Springer-Verlag, Berlin, 2006.

[40] B. S. Mordukhovich, Variational Analysis and Generalized Differentiation. II: Applications, Grundlehren Series (Fundamental Principles of Mathematical Sciences), Vol. 331, 592 pp., Springer-Verlag, Berlin, 2006.

[41] P. Ochs, Y. Chen, T. Brox, T. Pock, T. iPiano: Inertial Proximal Algorithm for Non-convex Optimization,SIAM Journal on Imaging Sciences, 7(2), 1388-1419, 2014

[42] Ochs, P. (2019). Unifying abstract inexact convergence theorems and block coordinate variable metric iPiano. SIAM Journal on Optimization, 29(1), 541-570.

[43] J.P. Penot, Conditioning convex and nonconvex problems. J Optim Theory Appl 90, 535–554 (1996).

[44] B. T. Polyak, Some methods of speeding up the convergence of iteration methods, USSR Computational Mathematics and Mathematical Physics, 4 (1964), pp. 1–17.

[45] B. T. Polyak, Introduction to Optimization. Optimization Software - Inc.Publication Division, New York (1987

[46] Y. Nesterov, Introductory lectures on convex optimization: a basic course, Kluwer Academic Publishers, Dordrecht, 2004

[47] M. Tamm, Subanalytic sets in the calculus of variations. Acta Math., 146 (1981).

[48] P.W. Wang, C.J. Lin, Iteration Complexity of Feasible Descent Methods for Convex Optimization, J. Machine Learning Research, 15 (2014), 1523-1548.

[49] C. Zalinescu, Sharp estimates for Hoffmans constant for systems of linear inequalities and equalities. SIAM J. Optim. 14, (2003), 517–533.