CHARACTERIZATIONS OF ŁOJASIEWICZ INEQUALITIES AND APPLICATIONS

JÉRÔME BOLTE, ARIS DANIILIDIS, OLIVIER LEY & LAURENT MAZET

Abstract. The classical Łojasiewicz inequality and its extensions for partial differential equation problems (Simon) and to o-minimal structures (Kurdyka) have a considerable impact on the analysis of gradient-like methods and related problems: minimization methods, complexity theory, asymptotic analysis of dissipative partial differential equations, tame geometry. This paper provides alternative characterizations of this type of inequalities for nonsmooth lower semicontinuous functions defined on a metric or a real Hilbert space. In a metric context, we show that a generalized form of the Łojasiewicz inequality (hereby called the Kurdyka-Łojasiewicz inequality) relates to metric regularity and to the Lipschitz continuity of the sublevel mapping, yielding applications to discrete methods (strong convergence of the proximal algorithm). In a Hilbert setting we further establish that asymptotic properties of the semiflow generated by $-\partial f$ are strongly linked to this inequality. This is done by introducing the notion of a piecewise subgradient curve: such curves have uniformly bounded lengths if and only if the Kurdyka-Łojasiewicz inequality is satisfied. Further characterizations in terms of *talweg* lines —a concept linked to the location of the less steepest points at the level sets of f— and integrability conditions are given. In the convex case these results are significantly reinforced, allowing in particular to establish the asymptotic equivalence of discrete gradient methods and continuous gradient curves. On the other hand, a counterexample of a convex C^2 function in $\mathbb{R}^2$ is constructed to illustrate the fact that, contrary to our intuition, and unless a specific growth condition is satisfied, convex functions may fail to fulfill the Kurdyka-Łojasiewicz inequality.

2000 *Mathematics Subject Classification. Primary* 26D10; *Secondary* 03C64, 37N40, 49J52, 65K10.

Key words and phrases. Łojasiewicz inequality, gradient inequalities, metric regularity, subgradient curve, gradient method, convex functions, global convergence, proximal method.

Contents

1. Introduction

The Łojasiewicz inequality is a powerful tool to analyze convergence of gradient-like methods and related problems. Roughly speaking, this inequality is satisfied by a C^1 function f, if for some $\theta \in [\frac{1}{2}, 1)$ the quantity

$$|f - f(\bar{x})|^{\theta} \, \|\nabla f\|^{-1}$$

remains bounded away from zero around any (possibly critical) point $\bar{x}$. This result is named after S. Łojasiewicz [33], who was the first to establish its validity for the classes of real–analytic and C^1 subanalytic functions. At the same time, it has been known that the Łojasiewicz inequality would fail for C^{∞} functions in general (see the classical example of the function $x \longmapsto \exp(-1/x^2)$, if $x \neq 0$ and 0, if $x = 0$ around the point $\bar{x} = 0$).

A generalized form of this inequality has been introduced by K. Kurdyka in [29]. In the framework of a C^1 function f defined on a real Hilbert space $[H, \langle \cdot, \cdot \rangle]$, and assuming for simplicity that $\bar{f} = 0$ is a critical value, this generalized inequality (that we hereby call the Kurdyka–Łojasiewicz inequality, or in short, the KŁ–inequality) states that

$$||\nabla(\varphi \circ f)(x)|| \geq 1, \tag{1}$$

for some continuous function $\varphi\colon [0,r) \to \mathbb{R}$, C^1 on $(0,r)$ with $\varphi' > 0$ and all x in $[0 < f < r] := \{y \in H : 0 < f(y) < r\}$. The class of such functions φ will be further denoted by $\mathcal{K}(0,\bar{r})$, see (8). Note that the Łojasiewicz inequality corresponds to the case $\varphi(t) = t^{1-\theta}$.

In finite-dimensional spaces it has been shown in [29] that (1) is satisfied by a much larger class of functions, namely, by those that are definable in an o-minimal structure [15], or even more generally by functions belonging to analytic-geometric categories [21]. In the meantime the original Łojasiewicz result was used to derive new results in the asymptotic analysis of nonlinear heat equations [40] and damped wave equations [26]. Many results related to partial differential equations followed, see the monograph of Huang [27] for an insight. Other fields of application of (1) are nonconvex optimization and nonsmooth analysis. This was one of the motivations for the nonsmooth KŁ–inequalities developed in [8, 9]. Due to its considerable impact on several field of applied mathematics: minimization and algorithms [1, 5, 8, 30], asymptotic theory of differential inclusions [38], neural networks [24], complexity theory [37] (see [37, Definition 3] where functions satisfying a KŁ–type inequality are called gradient dominated functions), partial differential equations [40, 26, 27], we hereby tackle the problem of characterizing such inequalities in an nonsmooth infinite-dimensional setting and provide further clarification in several application aspects. Our framework is rather broad (infinite dimensions, nonsmooth functions), nevertheless, to the best of our knowledge, most of the present results are also new in a smooth finite-dimensional framework: readers who feel unfamiliar with notions of nonsmooth and variational analysis may, at a first stage, consider that all functions involved are differentiable and replace subdifferentials by usual derivatives and subgradient systems by smooth ones.

A first part of this work (Section 2) is devoted to the analysis of metric versions of the KŁ–inequality. The underlying space H is only assumed to be a complete metric space (without any linear structure), the function $f\colon H \to \mathbb{R} \cup \{+\infty\}$ is lower semicontinuous and possibly real-extended valued and the notion of a gradient is replaced by the variational notion of a strong-slope [18, 6]. Indeed, introducing the multivalued mapping $F(x) = [f(x), +\infty)$ (whose graph is the epigraph of f), the KŁ–inequality (1) appears to be equivalent to the metric regularity of $F : H \rightrightarrows \mathbb{R}$ on an adequate set, where $\mathbb{R}$ is endowed with the metric $d_\varphi(r,s) = |\varphi(r) - \varphi(s)|$. This fact is strongly connected to famous classical results in this area (see [19, 35, 28, 39] for example) and in particular to the notion of ρ-metric regularity introduced in [28] by A. Ioffe. The particularity of our result is due to the fact that F takes its values in a totally ordered set which is not the case in the general theory. Using results on global error-bounds of Azé-Corvellec [6] and Zorn's lemma, we establish indeed that some global forms of the KŁ-inequality and metric regularity are both equivalent to the "Lipschitz continuity" of the sublevel mapping

$$\left\{ \begin{array}{lll} \mathbb{R} & \rightrightarrows & H \\ r & \mapsto & [f \leq r] := \{x \in H : f(x) \leq r\}, \end{array} \right.$$

where $(0, r) \subset (0, +\infty)$ is endowed with d_φ and the collection of subsets of H with the "Hausdorff distance". As it is shown in a section devoted to applications (Section 3.4), this reformulation is particularly adapted for the analysis of proximal methods involving nonconvex criteria: these results are in the line of [14, 5].

In the second part of this work (Section 3), H is a proper real Hilbert space and f is assumed to be a semiconvex function, *i.e.* f is the difference of a proper lower semicontinuous convex function and a function proportional to the canonical quadratic form. Although this assumption is not particularly restrictive, it does not aim at full generality. Semiconvexity is used here to provide a convenient framework in which the formulation and the study of subdifferential evolution equations are simple and elegant [2, 17]. Using the Fréchet subdifferential (see Definition 8), the corresponding subgradient dynamical system indeed reads

$$\left\{ \begin{array}{l} \dot{x}(t) + \partial f(x(t)) \ni 0, \text{ a.e. on } (0, +\infty), \\ x(0) \in \operatorname{dom} f \end{array} \right. \tag{2}$$

where $x(\cdot)$ is an *absolutely continuous curve* called *subgradient curve*. Relying on several works [17, 34, 11], if f is semiconvex, such curves exist and are unique. The asymptotic properties of the semiflow associated to this evolution equation are strongly connected to the KŁ-inequality. This can be made precise by introducing the following notion: for $T \in (0, +\infty]$, a piecewise absolutely continuous curve $\gamma \colon [0, T) \to H$ (with countable pieces) is called a *piecewise subgradient curve* if γ is a solution to (2) where in addition $t \mapsto (f \circ \gamma)(t)$ nonincreasing (see Definition 15 for details). Consider all piecewise subgradient curves lying in a "KŁ–neighborhood", *e.g.* a slice of level sets. Under a compactness assumption and a condition of Sard type (automatically satisfied in finite dimensions if f belongs to an o-minimal class), their lengths are uniformly bounded if and only if f satisfies the KŁ–inequality in its nonsmooth form (see [9]), that is, for all $x \in [0 < f < r]$,

$$||\partial(\varphi \circ f)(x)||_- := \inf\{||p|| : p \in \partial(\varphi \circ f)\} \geq 1,$$

where $\varphi\colon (0, r) \to \mathbb{R}$ is C^1 function bounded from below such that $\varphi' > 0$ (see (8)). A byproduct of this result (through not an equivalent statement, as we show in Section 4.3 –see Remark 37 (c)) is the fact that bounded subgradient curves have finite lengths and hence converge to a generalized critical point.

Further characterizations are given involving several aspects among which, an integrability condition in terms of the inverse function of the minimal subgradient norm associated to each level set $[f = r]$ of f, as well as connections to the following *talweg* selection problem: Find a piecewise absolutely continuous curve $\theta\colon (0, r) \to H$ with finite length such that

$$\theta(r) \in \Big\{ x \in [f = r] \,:\, ||\partial(\varphi \circ f)(x)||_- \leq \\ \leq R \inf_{y \in [f=r]} ||\partial(\varphi \circ f)(y)||_- \Big\}, \quad \text{with } R > 1.$$

The curve θ is called a *talweg*. Early connections between the KŁ-inequality and this old concept can be found in [29], and even more clearly in [16]. Indeed, under mild assumptions the existence of such a selection curve θ characterizes the KŁ-inequality. The proof relies strongly on the property of the semiflow associated to $-\partial f$. Recent developments of the metric theory of "gradient" curve [3] open the way to a more general approach of these characterizations, and hopefully to new applications in the line of [3, 18].

The analysis of the convex case (that is, f is a convex function) in Section 4, reveals interesting phenomena. In this case, the KŁ-inequality, whenever true on a slice of level sets, will be true on the whole space H (globalization) and, in addition, the involved function φ can be taken to be concave (Theorem 29). This is always the case if a specific growth assumption near the set of minimizers of f is assumed. On the other hand, arbitrary convex functions do not satisfy the KŁ–inequality: this is a straightforward consequence of a classical counterexample, due to J.-B. Baillon [7], of the existence of a convex function f in a Hilbert space, having a subgradient curve which is not strongly converging to $0 \in \arg\min f$. However, surprisingly, even smooth finite-dimensional coercive convex functions may fail to satisfy the KŁ-inequality, and this even in the case that the lengths of their gradient curves are uniformly bounded. Indeed, using the above mentioned characterizations and results from [41], we construct a counterexample of a C^2 convex function whose set of minimizers is compact and has a nonempty interior (Section 4.3).

As another application we consider abstract *explicit* gradient schemes for convex functions with a Lipschitz continuous gradient. A common belief is that the analysis of gradient curves and their explicit discretization used in numerical optimization are somehow disconnected problems. We hereby show that this is not always the case, by establishing that the piecewise gradient iterations are uniformly bounded if and only if the piecewise subgradient curves are so. This aspect sheds further light on the (theoretical) stability of convex gradient-like methods and the interest of relating the KŁ–inequality to the asymptotic study of subgradient-type methods.

Notation. (Multivalued mappings) Let X, Y be two metric spaces and $F\colon X \rightrightarrows Y$ be a multivalued mapping from X to Y. We denote by

$$\operatorname{Graph} F := \{(x, y) \in X \times Y : y \in F(x)\} \tag{3}$$

the *graph* of the multivalued mapping F (subset of $X \times Y$) and by

$$\operatorname{dom} F := \{x \in X : \exists y \in Y, \, (x, y) \in \operatorname{Graph} F\} \tag{4}$$

its *domain* (subset of X).

(Single–valued functions) Given a function $f\colon X \longrightarrow \mathbb{R} \cup \{+\infty\}$ we define its *epigraph* by

$$\operatorname{epi} f := \{(x, \beta) \in X \times \mathbb{R} : f(x) \leq \beta\}. \tag{5}$$

We say that the function f is *proper* (respectively, *lower semicontinuous*) if the above set is nonempty (respectively, closed). Let us recall that the domain of the function f is defined by

$$\operatorname{dom} f := \{x \in X : f(x) < +\infty\}.$$

(Level sets) Given $r_1 \leq r_2$ in $[-\infty, +\infty]$ we set

$$[r_1 \leq f \leq r_2] := \{x \in X : r_1 \leq f(x) \leq r_2\}.$$

When $r_1 = r_2$ (respectively $r_1 = -\infty$), the above set will be simply denoted by $[f = r_1]$ (respectively $[f \leq r_2]$).

(Strong slope) Let us recall from [18] (see also [28], [6]) the notion of *strong slope* defined for every $x \in \operatorname{dom} f$ as follows:

$$|\nabla f|(x) = \limsup_{y \to x} \frac{(f(x) - f(y))^+}{d(x, y)}, \tag{6}$$

where for every $a \in \mathbb{R}$ we set $a^+ = \max\{a, 0\}$.

If $[X, ||\cdot||]$ is a Banach space with (topological) dual space $[X^*, ||\cdot||_*]$ and f is a C^1 finite-valued function then

$$|\nabla f|(x) = ||\nabla f(x)||_*,$$

for all x in X, where $\nabla f(\cdot)$ is the differential map of f.

(Hausdorff distance) We define the *distance* of a point $x \in X$ to a subset S of X by

$$\text{dist}\,(x, S) := \inf_{y \in S} d(x, y),$$

where d denotes the distance on X. The *Hausdorff distance* $\text{Dist}(S_1, S_2)$ of two subsets S_1 and S_2 of X is given by

$$\text{Dist}(S_1, S_2) := \max \left\{ \sup_{x \in S_1} \text{dist}\,(x, S_2),\ \sup_{x \in S_2} \text{dist}\,(x, S_1) \right\}. \tag{7}$$

Let us denote by $\mathcal{P}(X)$ the collection of all subsets of X. In general $\text{Dist}(\cdot, \cdot)$ can take infinite values and does not define a distance on $\mathcal{P}(X)$. However if $K(X)$ denotes the collection of nonempty compact subsets of X, then $\text{Dist}(\cdot, \cdot)$ defines a proper notion of distance on $K(X)$. In the sequel we deal with multivalued mappings $F\colon X \rightrightarrows Y$ enjoying the following property

$$\text{Dist}\,(F(x), F(y)) \leq k\ d(x, y)$$

where k is a positive constant. For simplicity such functions are called Lipschitz continuous, although $[\mathcal{P}(Y), \text{Dist}\,]$ is not a metric space in general.

(Desingularization functions) Given $\bar{r} \in (0, +\infty]$, we set

$$\mathcal{K}(0, \bar{r}) := \big\{ \phi \in C([0, \bar{r})) \cap C^1(0, \bar{r})\ :\ \phi(0) = 0, \\ \text{and } \phi'(r) > 0, \forall r \in (0, \bar{r}) \big\}, \tag{8}$$

where $C([0, \bar{r}])$ (respectively, $C^1(0, \bar{r})$) denotes the set of continuous functions on $[0, \bar{r}]$ (respectively, C^1 functions on $(0, \bar{r})$).

Finally throughout this work, $B(x, r)$ will stand for the usual open ball of center x and radius $r > 0$ and $\bar{B}(x, r)$ will denote its closure. If H is a Hilbert space, its inner product will be denoted by $\langle \cdot, \cdot \rangle$ and the corresponding norm by $|| \cdot ||$.

2. KŁ–inequality is a metric regularity condition

Let X, Y be two *complete* metric spaces, $F\colon X \rightrightarrows Y$ a multivalued mapping and $(\bar{x}, \bar{y}) \in \text{Graph}\, F$. Let us recall from [28, Definition 1 (loc)] the following definition.

Definition 1 (metric regularity of multifunctions)**.** Let $k \in [0, +\infty)$.

(i) The multivalued mapping F is called k-metrically regular at $(\bar{x}, \bar{y}) \in$ Graph F, if there exist $\varepsilon, \delta > 0$ such that for all $(x, y) \in B(\bar{x}, \varepsilon) \times B(\bar{y}, \delta)$ we have

$$\text{dist}\,(x, F^{-1}(y)) \leq k\, \text{dist}\,(y, F(x)). \tag{9}$$

(ii) Let V be a nonempty subset of $X \times Y$. The multivalued mapping F is called k-metrically regular on V, if F is metrically regular at $(\bar{x}, \bar{y})$ for every $(\bar{x}, \bar{y}) \in \operatorname{Graph} F \cap V$.

2.1. **Metric regularity and global error bounds.** The following theorem is an essential result: it will show that Kurdyka-Łojasiewicz inequality and metric regularity are equivalent concepts (see Corollary 4 and Remark 5). The equivalence [(ii)⇔(iii)] is due to Azé-Corvellec (see [6, Theorem 2.1]).

Theorem 2. *Let X be a complete metric space, $f\colon X \longrightarrow \mathbb{R} \cup \{+\infty\}$ a proper lower semicontinuous function and $r_0 > 0$. The following assertions are equivalent:*

(i) The multivalued mapping

$$F : \left\{ \begin{array}{ccc} X & \rightrightarrows & \mathbb{R} \\ x & \longmapsto & [f(x), +\infty) \end{array} \right.$$

is k-metrically regular on $[0 < f < r_0] \times (0, r_0)$;

(ii) For all $r \in (0, r_0)$ and $x \in [0 < f < r_0]$

$$\operatorname{dist}(x, [f \le r]) \;\le\; k\,(f(x) - r)^+; \tag{10}$$

(iii) For all $x \in [0 < f < r_0]$

$$|\nabla f|(x) \ge \frac{1}{k}.$$

Proof. The equivalence of (ii) and (iii) follows from [6, Theorem 2.1] and is based on Ekeland variational principle. Definition 1 (metric regularity of multifunctions) yields the following restatement for (i):

$(i)_1$ For every $(\bar{x}, \bar{r}) \in \operatorname{Graph} F$ with $\bar{x} \in [0 < f < r_0]$ and $\bar{r} \in (0, r_0)$, there exist $\varepsilon > 0$ and $\delta > 0$ such that

$$\begin{aligned} (x, r) \in (B(\bar{x}, \varepsilon) \cap [0 < f < r_0]) \times [(\bar{r} - \delta, \bar{r} + \delta) \cap (0, r_0)] \Longrightarrow \\ \Longrightarrow \operatorname{dist}(x, [f \le r]) \;\le\; k\,(f(x) - r)^+. \end{aligned} \tag{11}$$

Clearly $(i) \Rightarrow (i)_1$. Now, in order to prove $(i)_1 \Rightarrow (i)$, consider $(\bar{x}, \bar{r}) \in \operatorname{Graph} F \cap [0 < f < r_0] \times (0, r_0)$. Take ε and δ positive given by $(i)_1$ such that $0 < \bar{r} - \delta < \bar{r} + 2\delta < r_0$, $\varepsilon \le k(r_0 - \bar{r} - 2\delta)$ and f is positive in $B(\bar{x}, \varepsilon)$ (f is lower semicontinuous so $[f > 0]$ is open). For any $(x, r) \in B(\bar{x}, \varepsilon) \times (\bar{r} - \delta, \bar{r} + \delta)$, we have $r \in (0, r_0)$ and $f(x) > 0$. Thus if $f(x) < r_0$ by $(i)_1$ we have

$$\operatorname{dist}(x, [f \le r]) \le k(f(x) - r)^+ = k \operatorname{dist}(r, F(x)).$$

If $f(x) \geq r_0$, then

$$\begin{aligned}\operatorname{dist}(x,[f\leq r]) &\leq \operatorname{dist}(x,\bar{x}) + \operatorname{dist}(\bar{x},[f\leq r]) \leq \\ &\leq \varepsilon + k\,(f(\bar{x})-r)^+ \leq \\ &\leq \varepsilon + k\delta \leq \\ &\leq k(r_0-\bar{r}-\delta) \leq \\ &\leq k(r_0-r) \leq \\ &\leq k(f(x)-r)^+ = k\operatorname{dist}(r,F(x)).\end{aligned}$$

Thus $(i)_1 \Rightarrow (i)$.

It is now straightforward to see that $(ii) \Longrightarrow (i)$, thus it remains to prove that $(i)_1 \Longrightarrow (ii)$. To this end, fix any $k' > k$, $r_1 \in (0,r_0)$ and $x_1 \in [f = r_1]$. We shall prove that

$$\operatorname{dist}(x_1,[f\leq s]) \leq k'(r_1-s),$$

for all $s \in (0, r_1]$.

Claim 1. Let $r \in (0, r_0)$ and $x \in [f = r]$. Then there exist $r^- < r$ and $x^- \in [f = r^-]$ such that

$$d(x,x^-) \leq k'(r-r^-) \tag{12}$$

with

$$\operatorname{dist}(x,[f\leq s]) \leq k'(r-s), \quad \text{for all } s\in[r^-,r].$$

Proof of Claim 1. Apply $(i)_1$ at $(x,r) \in \operatorname{Graph} F$ to obtain the existence of $\rho \in (0,r)$ such that $\operatorname{dist}(x,[f\leq s]) \leq k(r-s)$ for all $s \in [\rho, r]$. Since $k' > k$ there exists $x^- \in [f \leq \rho]$ satisfying

$$d(x,x^-) < \frac{k'}{k}\operatorname{dist}(x,[f\leq\rho]),$$

which in view of (11) yields

$$d(x,x^-) < k'\,(r-\rho).$$

To conclude, set $r^- = f(x^-) \leq \rho$ and observe that for any $s \in [r^-, \rho]$ we have

$$\operatorname{dist}(x,[f\leq s]) \leq d(x,x^-) \leq k'(r-\rho) \leq k'(r-s) = k'(f(x)-s).$$

This completes the proof of the claim. □

Let $\mathcal{A}$ be the set of all families $\{(x_i,r_i)\}_{i\in I} \subset [f\leq r_1]\times\mathbb{R}$ containing (x_1, r_1) such that

– (P_1) $f(x_i) = r_i$ for all $i \in I$ and $r_i \neq r_j$, for $i \neq j$.

– (P_2) If $i, j \in I$ and $r_i < r_j$ then $d(x_j, x_i) \leq k'\,(r_j - r_i)$.

– (P_3) For $r^* = \inf\{r_i : i \in I\}$ and for $s \in (r^*, r_1]$ we have:

$$\operatorname{dist}(x_1, [f \le s]) \le k'(r_1 - s).$$

The set $\mathcal{A}$ is nonempty (it contains the one–element family $\{(x_1, r_1)\}$) and can be ordered by the inclusion relation (that is, $\mathcal{J}_1 \preceq \mathcal{J}_2$ if, and only if, $\mathcal{J}_1 \subset \mathcal{J}_2$). Under this relation $\mathcal{A}$ becomes a totally ordered set: every totally ordered chain in $\mathcal{A}$ has an upper bound in $\mathcal{A}$ (its union). Thus, by Zorn lemma, there exists a maximal element $\mathcal{M} = \{(x_i, r_i)\}_{i\in I}$ in $\mathcal{A}$.

Claim 2. Any maximal element $\mathcal{M} = \{(x_i, r_i)\}_{i\in I}$ of $\mathcal{A}$ satisfies

$$r^* = \inf_{i\in I} r_i \le 0. \tag{13}$$

Proof of the Claim 2. Let us assume, towards a contradiction, that (13) is not true, *i.e.* $r^* > 0$. Let us first assume that there exists $j \in I$ such that $r^* = r_j$. Define $r^- := r_j^- < r_j$ and $x_j^- = x^- \in [f = r^-]$ as specified in Claim 1 and consider the family $\mathcal{M}_1 = \mathcal{M} \cup \{(x^-, r^-)\}$. Then $\mathcal{M}_1$ clearly complies with (P_1). To see that $\mathcal{M}_1$ satisfies (P_2), simply observe that for each $i \in I$,

$$d(x^-, x_i) \le d(x^-, x_j) + d(x_j, x_i) \le k'(r_i - r^-).$$

Let $s \in [r^-, r_j]$. By using the properties of the couple (x^-, r^-), one obtains

$$\begin{aligned}\operatorname{dist}(x_1, [f \le s]) &\le \operatorname{dist}(x_1, x_j) + \operatorname{dist}(x_j, [f \le s]) \le \\ &\le k'(r_1 - r_j) + k'(r_j - s) \le k'(r_1 - s).\end{aligned}$$

This means that $\mathcal{M}_1 \in \mathcal{A}$ which is contradicts the maximality of $\mathcal{M}$.
Thus it remains to treat the case when the infimum r^* is not attained. Let us take any decreasing sequence $\{r_{i_n}\}_{n\ge1}$, $i_n \in I$ satisfying $r_{i_1} = r_1$ and $r_{i_n} \searrow r^*$. For simplicity the sequences $\{r_{i_n}\}_n$ and $\{x_{i_n}\}_n$ will be denoted, respectively, by $\{r_n\}_n$ and $\{x_n\}_n$. Applying (P_2) we obtain

$$d(x_n, x_{n+m}) \le k'\,(r_n - r_{n+m}). \tag{14}$$

It follows that $\{x_n\}_{n\ge1}$ is a Cauchy sequence, thus it converges to some x^*. Taking the limit as $m \to +\infty$ we deduce from (14) that $d(x_n, x^*) \le k'\,(r_n - r^*)$, for all $n \in \mathbb{N}^*$. For any $i \in I$, there exists n such that $r_n < r_i$ and therefore

$$\operatorname{dist}(x^*, x_i) \le d(x^*, x_n) + d(x_n, x_i) \le k'(r_i - r^*) \le k'(r_i - f(x^*)), \tag{15}$$

where the last inequality follows from the lower semicontinuity of f. Set $f(x^*) = \rho^* \le r^*$ and $\mathcal{M}_1 = \mathcal{M} \cup \{(x^*, \rho^*)\}$. Since the infimum is not attained in $\inf\{r_i : i \in I\}$ the family $\mathcal{M}_1$ satisfies (P_1). Further by using (15), we see that $\mathcal{M}_1$ complies also with (P_2). Take $s \in [\rho^*, r^*]$. Since

$x^* \in [f \leq s]$, we have

$$\operatorname{dist}(x_1, [f \leq s]) \leq \operatorname{dist}(x_1, x^*) \leq k'(r_1 - r^*) \leq k'(r_1 - s).$$

Hence $\mathcal{M}_1$ belongs to $\mathcal{A}$ which contradicts the maximality of $\mathcal{M}$. □

The desired implication follows easily by taking the limit as k' goes to k. This completes the proof. □

Remark 3 (Sublevel mapping and Lipschitz continuity)**.** It is straightforward to see that statement (ii) above is equivalent to the "Lipschitz continuity" (see (7)) of the sublevel set application

$$\begin{cases} (0, r_0) & \rightrightarrows & X \\ r & \longmapsto & [f \leq r] \end{cases}$$

for the Hausdorff "metric" given in (7). Note that F^{-1} is exactly the sublevel mapping given above, and thus in this context the Lipschitz continuity of F^{-1} is equivalent to the Aubin property of F^{-1}, see [20, 28].

2.2. **Metric regularity and KŁ inequality.** As an immediate consequence of Theorem 2 and Remark 3, we have the following result.

Corollary 4 (KŁ-inequality and sublevel set mapping)**.** *Let* $f\colon X \longrightarrow \mathbb{R} \cup \{+\infty\}$ *be a lower semicontinuous function defined on a complete metric space* X *and let* $\varphi \in \mathcal{K}(0, r_0)$ *(see (8)). The following assertions are equivalent:*

(i) the multivalued mapping

$$\begin{cases} X & \rightrightarrows & \mathbb{R} \\ x & \mapsto & [(\varphi \circ f)(x), +\infty) \end{cases}$$

is k*-metrically regular on* $[0 < f < r_0] \times (0, \varphi(r_0))$*;*

(ii) for all $r_1, r_2 \in (0, r_0)$

$$\operatorname{Dist}([f \leq r_1], [f \leq r_2]) \leq k\, |\varphi(r_1) - \varphi(r_2)|;$$

(iii) for all $x \in [0 < f < r_0]$

$$|\nabla(\varphi \circ f)|(x) \geq \frac{1}{k}.$$

It might be useful to observe the following:

Remark 5 (Change of metric)**.** Let $\varphi \in \mathcal{K}(0, r_0)$ and assume that it can be extended continuously to an increasing function still denoted $\varphi\colon \mathbb{R}_+ \to \mathbb{R}_+$. Set $d_\varphi(r, s) = |\varphi(r) - \varphi(s)|$ for any $r, s \in \mathbb{R}_+$ and assume that $\mathbb{R}_+$ is endowed with the metric d_φ. Endowing $\mathbb{R}_+$ with this new metric, assertions (i), (ii) and (iii) can be reformulated very simply:smallskip (i ') The multivalued mapping

$$\left\{ \begin{array}{lll} X & \rightrightarrows & \mathbb{R}_+ \\ x & \mapsto & [f(x), +\infty) \end{array} \right.$$

is k-metrically regular on $[0 < f < r_0] \times (0, r_0)$.

(ii') The sublevel mapping

$$\mathbb{R}_+ \ni r \mapsto [f \leq r],$$

is k Lipschitz continuous on $(0, r_0)$.

(iii') For all $x \in [0 < f < r_0]$

$$|\nabla_\varphi f|(x) \geq \frac{1}{k},$$

where $|\nabla_\varphi f|$ denotes the strong slope of the restricted function $\bar{f}\colon [0 < f] \to [\mathbb{R}_+, d_\varphi]$.

Given a lower semicontinuous function $f\colon X \longrightarrow \mathbb{R} \cup \{+\infty\}$ we say that f is *strongly slope-regular*, if for each point x in its domain $\operatorname{dom} f$ one has

$$|\nabla f|(x) = |\nabla(-f)|(x). \tag{16}$$

Note that all C^1 functions are strongly slope-regular according to the above definition.

Proposition 6 (Level mapping and Lipschitz continuity)**.** *Assume $f\colon X \to \mathbb{R}$ is continuous and strongly slope-regular. Then any of the assertions*
(i)–(iii) of Theorem 2 is equivalent to the fact that the level set application

$$\left\{ \begin{array}{lll} \mathbb{R} & \rightrightarrows & X \\ r & \mapsto & [f = r] \end{array} \right.$$

is Lipschitz continuous on $(0, r_0)$ with respect to the Hausdorff metric.

Proof. The result follows by applying Theorem 2 twice. (Details are left to the reader.) □

Let us finally state the following important corollary.

Corollary 7 (KŁ-inequality and level set mapping)**.** *Let $f\colon X \longrightarrow \mathbb{R}$ be a continuous function which is strongly slope-regular on $[0 < f < r_0]$ and let $\varphi \in \mathcal{K}(0, r_0)$ (recall (8)). Then the following assertions are equivalent:*

(i) $\varphi \circ f$ is k-metrically regular on $[0 < f < r_0] \times (0, \varphi(r_0))$;
(ii) for all $r_1, r_2 \in (0, r_0)$

$$\operatorname{Dist}([f = r_1], [f = r_2]) \leq k\, |\varphi(r_1) - \varphi(r_2)|;$$

(iii) for all $x \in [0 < f < r_0]$

$$|\nabla(\varphi \circ f)|(x) \geq \frac{1}{k}.$$

Proof. It follows easily by combining Theorem 2 with Proposition 6. □

3. KŁ–inequality in Hilbert spaces

From now on, we shall work on a real Hilbert space $[H, \langle\cdot,\cdot\rangle]$. Given a vector x in H, the norm of x is defined by $||x|| = \sqrt{\langle x, x\rangle}$ while for any subset C of H, we set

$$||C||_- = \text{dist}\,(0, C) = \inf\{||x|| : x \in C\} \in \mathbb{R} \cup \{+\infty\}. \tag{17}$$

Note that $C = \emptyset$ implies $||C||_- = +\infty$.

3.1. **Elements of nonsmooth analysis.** Let us first recall the notion of Fréchet subdifferential (see [13, 36]).

Definition 8 (Fréchet subdifferential)**.** Let $f : H \to \mathbb{R} \cup \{+\infty\}$ be a real-extended-valued function. We say that $p \in H$ is a (Fréchet) subgradient of f at $x \in \text{dom}\, f$ if

$$\liminf_{y \to x,\, y \neq x} \frac{f(y) - f(x) - \langle p, y - x\rangle}{||y - x||} \geq 0.$$

We denote by $\partial f(x)$ the set of Fréchet subgradients of f at x and set $\partial f(x) = \emptyset$ for $x \notin \text{dom}\, f$. Let us now define the notion of critical point in variational analysis.

Definition 9 (critical point/values)**.** (i) A point $x_0 \in H$ is called *critical* for the function f, if $0 \in \partial f(x_0)$.

(ii) The value $r \in f(H)$ is called a critical value, if $[f = r]$ contains at least one critical point.

In this section we shall mainly deal with the class of *semiconvex* functions. Let us give the corresponding definition. (The reader should be aware that the terminology is not yet completely fixed in this area, so that the notion of semiconvex function may vary slightly from one author to another.)

Definition 10 (semiconvexity)**.** A proper lower semicontinuous function f is called *semiconvex* (or convex up to a square) if for some $\alpha > 0$ the function

$$x \longmapsto f(x) + \frac{\alpha}{2}||x||^2$$

is convex.

Remark 11. (i) For each $x \in H$, $\partial f(x)$ is a (possibly empty) closed convex subset of H and $\partial f(x)$ is nonempty for $x \in \operatorname{int} \operatorname{dom} f$.

(ii) It is straightforward from the above definition that the multivalued operator $x \longmapsto \partial f(x) + \alpha x$ is (maximal) monotone (see [42, Definition 12.5] for the definition).

(iii) For general properties of semiconvex functions, see [2]. Let us mention that Definition 10 is equivalent to the fact that

$$f(y) - f(x) \geq \langle p, y - x \rangle - \alpha ||x - y||^2, \tag{18}$$

for all $x, y \in H$ and all $p \in \partial f(x)$ (where $\alpha > 0$).
(iii) According to Definition 10, semiconvex functions are contained in several important classes of (nonsmooth) functions, as for instance ϕ-convex functions [17], weakly convex functions [4] and primal–lower–nice functions [34]. Although an important part of the forthcoming results is extendable to these more general classes, we shall hereby sacrifice extreme generality in sake of simplicity of presentation.

Given a real-extended-valued function f on H, we define the *remoteness* (*i.e.*, distance to zero) of its subdifferential ∂f at $x \in H$ as follows:

$$||\partial f(x)||_- = \inf_{p \in \partial f(x)} ||p|| = \operatorname{dist}(0, \partial f(x)). \tag{remoteness}$$

Remark 12 (minimal norm)**.** (i) If $\partial f(x) \neq \emptyset$, the infimum in the above definition is achieved since $\partial f(x)$ is a nonempty closed convex set. If we define $\partial^0 f(x)$ as the projection of 0 on the closed convex set $\partial f(x)$ we of course have

$$||\partial f(x)||_- = ||\partial^0 f(x)||. \tag{19}$$

Some properties of $H \ni x \mapsto ||\partial f(x)||_-$ are given in Section 5 (Annex).(ii) If f is a semiconvex function, then $||\partial f(x)||_-$ coincides with the notion of strong slope $|\nabla f|(x)$ introduced in (6), see Lemma 42 (Annex).

3.2. **Subgradient curves: basic properties.** Let $f \colon H \to \mathbb{R} \cup \{+\infty\}$ be a proper lower semicontinuous semiconvex function. The purpose of this subsection is to recall the main properties of the trajectories (subgradient curves) of the corresponding differential inclusion:

$$\begin{cases} \dot{\chi}_x(t) \in -\partial f(\chi_x(t)) & \text{a.e. on } (0, +\infty), \\ \\ \chi_x(0) = x \in \operatorname{dom} f. \end{cases}$$

The following statement aggregates useful results concerning existence and uniqueness of solutions. These results are essentially known even for a more general class of functions (see [34, Theorem 2.1, Proposition 2.14, Theorem 3.3] for instance for the class of primal–lower–nice functions). It should also be noticed that the integration of measurable curves of the form $\mathbb{R} \ni t \to \gamma(t) \in H$ relies on Bochner integration/measurability theory (basic properties can be found in [11]).

Theorem 13 (subgradient curves)**.** *For every $x \in \operatorname{dom} f$ there exists a unique absolutely continuous curve (called trajectory or subgradient curve) $\chi_x : [0, +\infty) \to H$ that satisfies*

$$\begin{cases} \dot{\chi}_x(t) \in -\partial f(\chi_x(t)) & \text{a.e. on } (0, +\infty), \\ \\ \chi_x(0) = x \in \operatorname{dom} f. \end{cases} \tag{20}$$

Moreover the trajectory satisfies:

(i) $\chi_x(t) \in \operatorname{dom} \partial f$ *for all* $t \in (0, +\infty)$.

(ii) *For all $t > 0$, the right derivative $\dot{\chi}_x(t^+)$ of χ_x is well defined and equal to*
$$\dot{\chi}_x(t^+) = -\partial^0 f(\chi_x(t)).$$
In particular $\dot{\chi}_x(t) = -\partial^0 f(\chi_x(t))$, for almost all t.

(iii) *The mapping $t \mapsto ||\partial f(\chi_x(t))||_-$ is right-continuous at each $t \in (0, +\infty)$.*

(iv) *The function $t \longmapsto f(\chi_x(t))$ is nonincreasing and continuous on $[0, +\infty)$. Moreover, for all $t, \tau \in [0, +\infty)$ with $t \leq \tau$, we have*
$$f(\chi_x(t)) - f(\chi_x(\tau)) \geq \int_t^\tau ||\dot{\chi}_x(u)||^2 \, du \,,$$
and equality holds if $t > 0$.

(v) *The function $t \longmapsto f(\chi_x(t))$ is Lipschitz continuous on $[\eta, +\infty)$ for any $\eta > 0$. Moreover*
$$\frac{d}{dt} f(\chi_x(t)) = -||\dot{\chi}_x(t)||^2 \;\; \text{a.e on } (\eta, +\infty).$$

Proof. The only assertion that does not appear explicitly in [34] is the continuity of the function $f \circ \chi_x$ at $t = 0$ when $x \in \operatorname{dom} f \setminus \operatorname{dom} \partial f$, but this is an easy consequence of the fact that f is lower semicontinuous, χ_x is (absolutely) continuous and $f \circ \chi_x$ is decreasing. For the rest of the assertions we refer to [34]. □

The following result asserts that the semiflow mapping associated with the differential inclusion (20) is continuous. This type of result can be established by standard techniques and therefore is essentially known (see [11, 34] for example). We give here an outline of proof (in case that f is semiconvex) for the reader's convenience.

Theorem 14 (continuity of the semiflow)**.** *For any semiconvex function f the semiflow mapping*

$$\left\{ \begin{array}{lcl} \mathbb{R}_+ \times \operatorname{dom} f & \to & H \\ (t,x) & \mapsto & \chi_x(t) \end{array} \right.$$

is (norm) continuous on each subset of the form $[0,T] \times (B(0,R) \cap [f \le r])$ where $T, R > 0$ and $r \in \mathbb{R}$.

Proof. Let us fix $x, y \in \operatorname{dom} f$ and $T > 0$. Then for almost all $t \in [0,T]$, there exist $p(\chi_x(t)) \in \partial f(\chi_x(t))$ and $q(\chi_y(t)) \in \partial f(\chi_y(t))$ such that

$$\begin{aligned} \frac{d}{dt}||\chi_x(t) - \chi_y(t)||^2 &= 2\langle \chi_x(t) - \chi_y(t), \dot\chi_x(t) - \dot\chi_y(t)\rangle = \\ &= -2\langle \chi_x(t) - \chi_y(t), p(\chi_x(t)) - q(\chi_y(t))\rangle. \end{aligned}$$

It follows by (18) that

$$\frac{d}{dt}||\chi_x(t) - \chi_y(t)||^2 \le 2\alpha||\chi_x(t) - \chi_y(t)||^2,$$

which implies (using Grönwall's lemma) that for all $0 \le t \le T$ we have

$$||\chi_x(t) - \chi_y(t)||^2 \le \exp(2\alpha T)||x-y||^2. \tag{21}$$

For any $0 \le t \le s \le T$, using Cauchy–Schwartz inequality and Theorem 13 we deduce that

$$\begin{aligned} ||\chi_x(s) - \chi_x(t)|| &\le \int_t^s ||\dot\chi_x(\tau)||d\tau \le \\ &\le \sqrt{s-t}\,\sqrt{\int_s^t ||\dot\chi_x(\tau)||^2 d\tau} \le \sqrt{s-t}\,\sqrt{f(x)}. \end{aligned} \tag{22}$$

The result follows by combining (21) and (22). □

Let us introduce the notions of a *piecewise absolutely continuous curve* and of a *piecewise subgradient curve*. This latter notion, due to its robustness, will play a central role in our study.

Definition 15. Let $a, b \in [-\infty, +\infty]$ with $a < b$.
(Piecewise absolutely continuous curve) A curve $\gamma\colon (a,b) \to H$ is said to be *piecewise absolutely continuous* if there exists a countable partition of (a,b) into intervals I_k such that the restriction of γ to each I_k is absolutely continuous.
(Length of a curve) Let $\gamma\colon (a,b) \to H$ be a piecewise absolutely continuous curve. The length of γ is defined by

$$\text{length}\,[\gamma] := \int_a^b ||\dot\gamma(t)||\,dt.$$

(Piecewise subgradient curve) Let $T \in (0, +\infty]$. A curve $\gamma\colon [0, T) \to H$ is called a piecewise subgradient curve for (20) if there exists a countable partition of $[0, T]$ into (nontrivial) intervals I_k such that:

– the restriction $\gamma|_{I_k}$ of γ to each interval I_k is a subgradient curve;

– for each disjoint pair of intervals I_k, I_l, the intervals $f(\gamma(I_k))$ and $f(\gamma(I_l))$ have at most one point in common.

Note that piecewise subgradient curves are piecewise absolutely continuous. Observe also that subgradient curves satisfy the above definition in a trivial way.

3.3. **Characterizations of the KŁ-inequality.** In this section we state and prove one of the main results of this work. Let $f\colon H \to \mathbb{R} \cup \{+\infty\}$ and $\bar{x} \in [f = 0]$ be a critical point. Throughout this section the following assumptions will be used:

– There exist $\bar{r}, \bar{\epsilon} > 0$ such that

$$x \in \bar{B}(\bar{x}, \bar{\epsilon}) \cap [0 < f \leq \bar{r}] \implies 0 \notin \partial f(x) \tag{23}$$

(0 is a locally upper isolated critical value).

– There exist $\bar{r}, \bar{\epsilon} > 0$ such that

(24) $\bar{B}(\bar{x}, \bar{\epsilon}) \cap [f \leq \bar{r}]$ is (norm) compact (local sublevel compactness).

Remark 16. (i)The first condition can be seen as a Sard-type condition.
(ii) Assumption (24) is always satisfied in finite-dimensional spaces, but is also satisfied in several interesting cases involving infinite-dimensional spaces. Here are two elementary examples.

(ii)$_1$ The (convex) function $f : \ell^2(\mathbb{N}) \to \mathbb{R}$ defined by

$$f(x) = \sum_{n \geq 1} n^2 x_i^2$$

has compact lower level sets.

(ii)$_2$ Let $g : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ be a proper lower semicontinuous semiconvex function and let $\Phi\colon L^2(\Omega) \to \mathbb{R} \cup \{+\infty\}$ be as follows [10]

$$\Phi(x) = \begin{cases} \frac{1}{2}\int_\Omega ||\nabla x||^2 + \int_\Omega g(x) \text{ if } x \in H^1(\Omega) \\ +\infty \text{ otherwise.} \end{cases}$$

The above function is a lower semicontinuous semiconvex function and the sets of the form $[\Phi \leq r] \cap B(\bar{x}, R)$ are relatively compact in $L^2(\Omega)$ (use the compact embedding theorem of $H^1(\Omega) \hookrightarrow L^2(\Omega)$).

As shown in Theorem 18, Kurdyka-Łojasiewicz inequality can be characterized in terms of boundedness of the length of "worst (piecewise absolutely continuous) curves", that is those defined by the points of less steepest descent.

Definition 17 (Talweg/Valley)**.** Let $\bar{x} \in [f = 0]$ be a critical point of f and assume that (23) holds for some $\bar{r}, \bar{\epsilon} > 0$. Let D be any closed bounded set that contains $B(\bar{x}, \bar{\epsilon}) \cap [0 < f \leq \bar{r}]$. For any $R > 1$ the R-valley $\mathcal{V}_R(\cdot)$ of f around $\bar{x}$ is defined as follows:

$$\mathcal{V}_R(r) = \left\{ x \in [f = r] \cap D : \, ||\partial f(x)||_- \, \leq \, R \inf_{y \in [f=r] \cap D} ||\partial f(y)||_- \right\}, \tag{25}$$

for all $r \in (0, \bar{r}]$.

A selection $\theta \colon (0, \bar{r}] \to H$ of $\mathcal{V}_R$, *i.e.* a curve such that $\theta(r) \in \mathcal{V}_R(r), \forall r \in (0, \bar{r}]$, is called an R-talweg or simply a talweg.

We are ready to state the main result of this work.

Theorem 18 (Subgradient inequality – local characterization)**.** *Let $f \colon H \to \mathbb{R} \cup \{+\infty\}$ be a lower semicontinuous semiconvex function and $\bar{x} \in [f = 0]$ be a critical point. Assume that there exist $\bar{\epsilon}, \bar{r} > 0$ such that* (23) *and* (24) *hold.*

Then, the following statements are equivalent:

(i) **[Kurdyka-Łojasiewicz inequality]** *There exist $r_0 \in (0, \bar{r})$, $\epsilon \in (0, \bar{\epsilon})$ and $\varphi \in \mathcal{K}(0, r_0)$ such that*

$$||\partial(\varphi \circ f)(x)||_- \geq 1, \qquad \textit{for all } x \in \bar{B}(\bar{x}, \epsilon) \cap [0 < f \leq r_0]. \tag{26}$$

(ii) **[Length boundedness of subgradient curves]** *There exist $r_0 \in (0, \bar{r})$, $\epsilon \in (0, \bar{\epsilon})$ and a strictly increasing continuous function $\sigma \colon [0, r_0] \to [0, +\infty)$ with $\sigma(0) = 0$ such that for all subgradient curves χ_x of (20) satisfying $\chi_x([0, T)) \subset \bar{B}(\bar{x}, \epsilon) \cap [0 < f \leq r_0]$ ($T \in (0, +\infty]$) we have*

$$\int_0^T ||\dot{\chi}_x(t)|| dt \leq \sigma(f(x)) - \sigma(f(\chi_x(T))).$$

(iii) **[Piecewise subgradient curves have finite length]** *There exist $r_0 \in (0, \bar{r})$, $\epsilon \in (0, \bar{\epsilon})$ and $M > 0$ such that for all piecewise subgradient curves $\gamma : [0, T) \to H$ of* (20) *satisfying $\gamma([0, T)) \subset \bar{B}(\bar{x}, \epsilon) \cap [0 < f \leq r_0]$ ($T \in (0, +\infty]$) we have*

$$\text{length}[\gamma] := \int_0^T ||\dot{\gamma}(\tau)|| d\tau < M.$$

(iv) **[Talwegs of finite length]** *For every $R > 1$, there exist $r_0 \in (0, \bar{r})$, $\epsilon \in (0, \bar{\epsilon})$, a closed bounded subset D containing $B(\bar{x}, \epsilon) \cap [0 < f \leq r_0]$ and a*

piecewise absolutely continuous curve $\theta : (0, r_0] \to H$ *of* finite length *which is a selection of the valley* $\mathcal{V}_R(r)$, *that is,*

$$\theta(r) \in \mathcal{V}_R(r), \text{ for all } r \in (0, r_0].$$

(v) [**Integrability condition**] *There exist* $r_0 \in (0, \bar{r})$ *and* $\epsilon \in (0, \bar{\epsilon})$ *such that the function*

$$u(r) = \frac{1}{\inf\limits_{x \in \bar{B}(\bar{x},\epsilon) \cap [f=r]} ||\partial f(x)||_-}, \quad r \in (0, r_0]$$

is finite-valued and belongs to $L^1(0, r_0)$.

Remark 19. (i) As it appears clearly in the proof, statement (iv) can be replaced by (iv') "*There exist* $R > 1$, $r_0 \in (0, \bar{r})$, $\epsilon \in (0, \bar{\epsilon})$, a closed bounded subset D containing $B(\bar{x}, \epsilon) \cap [0 < f \leq r_0]$ and a piecewise absolutely continuous curve $\theta\colon (0, r_0] \to H$ of *finite length* which is a selection of the valley $\mathcal{V}_R(r)$, that is,

$$\theta(r) \in \mathcal{V}_R(r), \text{ for all } r \in (0, r_0]''.$$

(ii) The compactness assumption (24) is only used in the proofs of $(iii) \Rightarrow (ii)$ and $(ii) \Rightarrow (iv)$. Hence if this assumption is removed, we still have:

$$(iv) \Longrightarrow (iv') \Longrightarrow (v) \Longleftrightarrow (i) \Longrightarrow (ii) \Longrightarrow (iii).$$

(iii) Note that (i) implies condition (23). This follows immediately from the chain rule (see Annex, Lemma 43).

Proof of Theorem 18. [**(i)⇒(ii)**] Let ϵ, r_0, φ be as in (i) such that (26) holds. Let further χ_x be a subgradient curve of (20) for $x \in [0 < f \leq r_0]$ and assume that $\chi_x([0, T)) \subset \bar{B}(\bar{x}, \epsilon) \cap [0 < f \leq r_0]$ for some $T > 0$.

Let us first assume that $x \in \operatorname{dom} \partial f$. Since φ is C^1 on $(0, r_0)$, by Theorem 13(v) and Lemma 43 (Annex) we deduce that the curve $t \mapsto \varphi(f(\chi_x(t))$ is absolutely continuous with derivative

$$\frac{d}{dt}(\varphi \circ f \circ \chi_x)(t) = -\varphi'(f(\chi_x(t))||\dot{\chi}_x(t)||^2 \text{ a.e. on } (0, T).$$

Integrating both terms on the interval $(0, T)$ and recalling (26), $\chi_x(0) = x$ we get

$$\begin{aligned} \varphi(f(x)) - \varphi(f(\chi_x(T))) &= -\int_0^T \frac{d}{dt}(\varphi \circ f \circ \chi_x)(t) dt \\ &= \int_0^T \varphi'(f(\chi_x(t))||\dot{\chi}_x(t)||^2 dt \geq \int_0^T ||\dot{\chi}_x(t)|| dt. \end{aligned}$$

Thus (ii) holds true for $\sigma := \varphi$ and for all subgradient curves starting from points in $\operatorname{dom}\partial f$. Let now $x \in \operatorname{dom} f \diagdown \operatorname{dom}\partial f$ and fix any $\delta \in (0,T)$. Since $\chi_x([\delta,T]) \subset \operatorname{dom}\partial f$ we deduce from the above that

$$\int_\delta^T ||\dot\chi_x(t)||dt \le \sigma(f(\chi_x(\delta)) - \sigma(f(\chi_x(T))).$$

Thus the result follows by taking $\delta \searrow 0^+$ and using the continuity of the mapping $t \longmapsto f(\chi_x(t))$ at 0 (Theorem 13(ii)).

[(ii)⇒(iii)] Let γ be a piecewise subgradient curve as in (iii) and let I_k be the associated partition of $[0,T]$ (*cf.* Definition 15). Let $\{a_k\}$ and $\{b_k\}$ be two sequences of real numbers such that int $I_k = (a_k, b_k)$. Since the restriction $\gamma|_{I_k}$ of γ onto I_k is a subgradient curve, applying (ii) on (a_k, b_k) we get

$$\text{length}\ [\gamma|_{I_k}] \le \sigma(f(\gamma(a_k))) - \sigma(f(\gamma(b_k))).$$

Let m be an integer and $I_{k_1}, \dots, I_{k_m}$ a finite subfamily of the partition. We may assume that these intervals are ordered as follows $0 \le a_{k_1} \le b_{k_1} \le \cdots \le a_{k_m} \le b_{k_m}$. Hence

$$\sum_1^m [\sigma(f(\gamma(a_{k_i}))) - \sigma(f(\gamma(b_{k_i})))] \le \sigma(f(\gamma(a_{k_1}))) \le \sigma(r_0).$$

Thus the family $\{\sigma(f(\gamma(a_k))) - \sigma(f(\gamma(b_k)))\}$ is summable, hence using the definition of Bochner integral (see [11])

$$\text{length}\ [\gamma] \ = \ \sum_{k\in\mathbb{N}} \text{length}\ [\gamma|_{I_k}] \ \le \ \sigma(r_0).$$

[(iii)⇒(ii)] Let ϵ, r_0 be as in (iii), pick any $0 \le r' < r \le r_0$ and denote by $\Gamma_{r',r}$ the (nonempty) set of piecewise subgradient curves $\gamma\colon [0,T) \to H$ (where $T \in (0,+\infty]$) such that

$$\gamma([0,T)) \subset \bar B(\bar x, \epsilon) \cap [r' < f \le r].$$

Note that, by Theorem 13(iv) and Proposition 41(iii), $T = +\infty$ is possible only when $r' = 0$. Set further

$$\psi(r', r) := \sup_{\gamma\in\Gamma_{r',r}} \text{length}[\gamma] \qquad \text{and} \qquad \sigma(r) := \psi(0, r).$$

Note that (iii) guarantees that ψ and σ have finite values. We can easily deduce from Definition 15 that

$$\psi(0, r') + \psi(r', r) = \psi(0, r). \tag{27}$$

Thus for each $x \in \bar{B}(\bar{x}, \epsilon) \cap [0 < f \leq r_0]$ and $T > 0$ such that $\chi_x([0,T]) \subset B(\bar{x}, \epsilon) \cap [0 < f \leq r_0]$, we have

$$\int_0^T ||\dot{\chi}_x(\tau)||d\tau + \sigma(f(\chi_x(T)) \leq \sigma(f(x)). \tag{28}$$

Since the function σ is nonnegative and increasing it can be extended continuously at 0 by setting $\sigma(0) = \lim_{t\downarrow 0} \sigma(t) \geq 0$. Since the property (28) remains valid if we replace $\sigma(\cdot)$ by $\sigma(\cdot) - \sigma(0)$, there is no loss of generality to assume $\sigma(0) = 0$.

To conclude it suffices to establish the continuity of σ on $(0, r_0]$. Fix $\tilde{r}$ in $(0, r_0)$ and take a subgradient curve $\chi\colon [0,T) \to H$ satisfying $\chi([0,T)) \subset \bar{B}(\bar{x}, \epsilon) \cap [f \leq r_0]$, where $T \in (0, +\infty]$. Set $f(\chi(0)) = r$ and $\lim_{t\to T} f(\chi(t)) = r'$ and assume that $\tilde{r} \leq r' \leq r \leq r_0$.

From Theorem 13(iv) and Proposition 41(iii) (Annex), we deduce that $T < +\infty$ so that $\chi([0,T]) \subset \bar{B}(\bar{x}, \epsilon) \cap [r' \leq f \leq r]$. Using assumption (23) together with Theorem 13 (i),(v), we deduce that the absolutely continuous function $f \circ \chi\colon [0,T] \to [r', r]$ is invertible and

$$\begin{aligned} \frac{d}{d\rho}[f\circ\chi]^{-1}(\rho) &= \frac{-1}{||\dot{\chi}([f\circ\chi]^{-1}(\rho)||^2} \geq \\ &\geq \frac{-1}{\inf\limits_{x\in\bar{B}(\bar{x},\epsilon)\cap[\tilde{r}\leq f\leq r_0]} ||\partial f(x)||_-^2} := -K, \end{aligned} \tag{29}$$

for almost all $\rho \in (r, r')$. By Proposition 41(iii) (Annex) we get that $K < +\infty$ and therefore the function $\rho \longmapsto [f\circ\chi]^{-1}(\rho)$ is Lipschitz continuous with constant K on $[r', r]$. Using the Cauchy-Schwarz inequality and Theorem 13(iv) we obtain

$$\begin{aligned} \text{length}\,[\chi] \;&=\; \int_0^T ||\dot{\chi}|| \;\leq\; \sqrt{T}\sqrt{\int_0^T ||\dot{\chi}||^2} = \\ &= \sqrt{[f\circ\chi]^{-1}(r) - [f\circ\chi]^{-1}(r')}\sqrt{\int_0^T ||\dot{\chi}||^2} \leq \\ &\leq \sqrt{K(r-r')}\sqrt{r-r'} = \sqrt{K}(r-r'). \end{aligned}$$

This last inequality implies that each piecewise subgradient curve $\gamma\colon [0,T) \to H$ such that $\gamma([0,T)) \subset \bar{B}(\bar{x}, \epsilon) \cap [r' \leq f \leq r]$ satisfies

$$\text{length}\,[\gamma] \;\leq\; \sqrt{K}(r - r'),$$

thus using (27) we obtain $\sigma(r) - \sigma(r') \leq \sqrt{K}(r - r')$, which yields the continuity of σ.

[(ii)⇒(iv)] Let us assume that (ii) holds true for ϵ and r_0. In a first step we establish the existence of a closed bounded subset D of $[0 < f \le r_0]$ satisfying

$$x \in D,\ t \ge 0,\ f(\chi_x(t)) > 0 \ \Rightarrow\ \chi_x(t) \in D. \tag{30}$$

Let $r_0 \ge r_1 > 0$ be such that $\sigma(r_1) < \epsilon/3$ and let us set

$$D := \{y \in \bar{B}(\bar{x}, \epsilon) \cap [0 < f \le r_1] : \exists x \in \bar{B}(\bar{x}, \epsilon/3) \cap [0 < f \le r_1], \\ \exists t \ge 0 \text{ such that } \chi_x(t) = y\}.$$

Let us first show that D enjoys property (30). It suffices to establish that

$$x \in \bar{B}(\bar{x}, \epsilon/3) \cap [0 < f \le r_1],\ t \ge 0,\ f(\chi_x(t)) > 0 \Rightarrow \chi_x(t) \in D.$$

To this end, fix $x \in \bar{B}(\bar{x}, \epsilon/3) \cap [0 < f \le r_1]$. By continuity of the flow, we observe that $\chi_x(t) \in \bar{B}(\bar{x}, \epsilon)$ for small $t > 0$ and for all $t \ge 0$ such that $\chi_x([0, t]) \subset \bar{B}(\bar{x}, \epsilon)$ with $f(\chi_x(t)) > 0$, assumption (ii) yields

$$\begin{aligned} ||\chi_x(t) - \bar{x}|| &\le ||\chi_x(t) - x|| + ||x - \bar{x}|| \le \\ &\le \int_0^t ||\dot{\chi}_x(\tau)|| d\tau + \epsilon/3 \le \\ &\le \sigma(r_1) + \epsilon/3 \le 2\epsilon/3. \end{aligned} \tag{31}$$

Thus D satisfies (30) and $\bar{B}(\bar{x}, \epsilon/3) \cap [f \le r_1] \subset D$.

Let us now prove that D is (relatively) closed in $[0 < f \le r_1]$. Let $y_k \in D$ be a sequence converging to y such that $f(y) \in (0, r_1]$. Then there exist sequences $\{x_n\}_n \subset \bar{B}(\bar{x}, \epsilon/3) \cap [0 < f \le r_1]$ and $\{t_n\}_n \subset \mathbb{R}_+$ such that $\chi_{x_n}(t_n) = y_n$. Since f is lower semicontinuous, there exists $n_0 \in \mathbb{N}$ and $\eta > 0$ such that $f(y_n) > \eta$ for all $n \ge n_0$. By Theorem 13(ii),(iv), (23) and Proposition 41(iii) (Annex), we obtain for all $n \ge n_0$

$$0 <\ t_n \inf_{z \in [\eta \le f \le r_1] \cap \bar{B}(\bar{x}, \epsilon)} ||\partial f(z)||_-^2 \ \le\ \int_0^{t_n} ||\dot{\chi}_{x_n}(t)||^2 dt \ \le\ f(x_n)\ \le\ r_1.$$

The above inequality shows that the sequence $\{t_n\}_n$ is bounded. Using a standard compactness argument we therefore deduce that, up to an extraction, $x_n \to \tilde{x}$ and $t_n \to \tilde{t}$ for some $\tilde{x} \in \bar{B}(\bar{x}, \epsilon/3) \cap [f \le r_1]$ and $\tilde{t} \in \mathbb{R}_+$. Theorem 14 (continuity of the semiflow) implies that $y = \chi_{\tilde{x}}(\tilde{t})$ and consequently that $f(\tilde{x}) \ge f(y) > 0$, yielding that $y \in D$. This shows that D is (relatively) closed in $[0 < f \le r_0]$.

Now we build a piecewise absolute continuous curve in the valley. According to the notation of Proposition 41 (Annex) we set

$$s_D(r) := \inf\{||\partial f(x)||_- : x \in D \cap [f = r]\},$$

so that for any $R > 1$ the R-valley around $\bar{x}$ (cf. Definition 17) is given by

$$\mathcal{V}_R(r) := \{x \in [f = r] \cap D : ||\partial f(x)||_- \ \leq \ R\, s_D(r)\}.$$

If $\bar{B}(\bar{x}, \epsilon/3) \cap [f = r] = \emptyset$ for all $0 < r \leq r_1$, there is nothing to prove. Otherwise, there exists $0 < r_2 \leq r_1$ and $x_2 \in \bar{B}(\bar{x}, \epsilon/3) \cap [f = r_2] \subset D$. From Theorem 13 and Proposition 41(iii) (Annex), we deduce that $\chi_{x_2}(t) \in [f = f(\chi_{x_2}(t))] \cap D \cap \mathrm{dom}\, \partial f$ for all $t \geq 0$ such that $[f \circ \chi_{x_2}](t) > 0$ and that the *inverse* function $[f \circ \chi_{x_2}]^{-1}(\cdot)$ is defined on an interval containing $(0, r_2)$. In other words the set $[f = r] \cap D \cap \mathrm{dom}\, \partial f$ is nonempty for each $r \in (0, r_2)$, which in turn implies that the valley is nonempty for small positive values of r, *i.e.* $\mathcal{V}_R(r) \neq \emptyset$ for all $r \in (0, r_2)$. With no loss of generality we assume that $\mathcal{V}_R(r_2) \neq \emptyset$.

Let further $R' \in (1, R)$ and $x \in [f = r_2] \cap D$ be such that $||\partial f(x)||_- \leq R'\, s_D(r_2)$ (therefore, in particular, $x \in \mathcal{V}_R(r_2)$). Take $\rho \in (R', R)$. Since the mapping $t \longmapsto ||\partial f(\chi_x(t)||_-$ is right–continuous (*cf.* Theorem 13(iii)), there exists $t_0 > 0$ such that $||\partial f(\chi_x(t)||_- < \rho s_D(r_2)$ for all $t \in (0, t_0)$. On the other hand $t \longmapsto s_D(f(\chi_x(t))$ is lower semicontinuous (*cf.* Proposition 41–Annex), hence there exists $t_1 \in (0, t_0)$ such that $R\, s_D(f(\chi_x(t)) > \rho\, s_D(r_2)$, for all $t \in (0, t_2)$. Using the continuity of the mapping $\chi_x(\cdot)$ and the stability property (30), we obtain the existence of $t_2 > 0$ such that

$$\chi_x(t) \in \mathcal{V}_R(f(x(t)) \text{ for all } t \in [0, t_2). \tag{32}$$

By using arguments similar to those of [(iii)⇒(ii)] we define the following absolutely continuous curve:

$$(f \circ \chi_x(t_2), r_2] \ni r \longmapsto \theta(r) = \chi_x([f \circ \chi_x]^{-1}(r)) \in D \cap [f = r].$$

By Proposition 46 based on Zorn's Lemma (see Annex), we obtain a piecewise subgradient curve that we still denote by θ, defined on $(0, r_2]$, satisfying $\theta(r) \in \mathcal{V}_R(r)$ for all $r \in (0, r_2]$. Assumption (iii) now yields

$$\text{length } [\theta] < M < +\infty,$$

completing the proof of the assertion.

[(iv)⇒(v)] Fix $R > 1$ and let ϵ, r_0 and $\theta \colon (0, r_0] \to H$ be as in (iv). Applying Lemma 43 (Annex), we get

$$\frac{d}{dr}(f \circ \theta)(r) = 1 = \langle \dot{\theta}(r), p(r) \rangle \text{ a.e on } (0, r_0], \quad \text{ for all } p(r) \in \partial f(\theta(r)).$$

Using the Cauchy-Schwartz inequality together with the fact that $D \cap [f = r] \supset \bar{B}(\bar{x}, \epsilon) \cap [f = r]$, we obtain

$$R\, ||\dot{\theta}(r)|| \ \geq \ u(r) \ = \ \frac{1}{\inf_{x \in \bar{B}(\bar{x}, \epsilon) \cap [f=r]} ||\partial f(x)||_-}\,,$$

for almost all $r \in (0, r_0]$. Since θ has finite length we deduce that $u \in L^1((0, r_0)$.

[(v)⇒(i)] Let ϵ, r_0 and u be as in (v). From Proposition 41 (Annex) we deduce that u is finite-valued and upper semicontinuous. Applying Lemma 44 (Annex) we obtain a continuous function $\bar{u}\colon (0, r_0] \to (0, +\infty)$ such that $\bar{u}(r) \geq u(r)$ for all $r \in (0, r_0]$. We set

$$\varphi(r) = \int_0^r \bar{u}(s) ds.$$

It is directly seen that $\varphi(0) = 0$, $\varphi \in C([0, r]) \cap C^1(0, r_0)$ and $\varphi'(r) > 0$ for all $r \in (0, r_0)$. Let $x \in \bar{B}(\bar{x}, \epsilon) \cap [f = r]$ and $q \in \partial(\varphi \circ f)(x)$. From Lemma 43 (Annex) we deduce $p := \frac{q}{\varphi'(r)} \in \partial f(x)$, and therefore

$$||q|| \,=\, \varphi'(r)\,||\frac{q}{\varphi'(r)}|| \,\geq\, u(r)\,||p|| \,\geq 1.$$

The proof is complete. □

Under a stronger compactness assumption Theorem 18 can be reformulated as follows.

Theorem 20 (Subgradient inequality – global characterization)**.** *Let $f\colon H \to \mathbb{R} \cup \{+\infty\}$ be a lower semicontinuous semiconvex function. Assume that there exists $r_0 > 0$ such that*

$$[f \leq r_0] \textit{ is compact and } 0 \notin \partial f(x),\ \forall x \in [0 < f < r_0].$$

Then the following propositions are equivalent

(i) **[Kurdyka-Łojasiewicz inequality]** *There exists a $\varphi \in \mathcal{K}(0, r_0)$ such that*

$$||\partial(\varphi \circ f)(x)||_- \geq 1, \qquad \textit{for all } x \in [0 < f < r_0].$$

(ii) **[Length boundedness of subgradient curves]** *There exists an increasing continuous function $\sigma\colon [0, r_0) \to [0, +\infty)$ with $\sigma(0) = 0$ such that for all subgradient curves $\chi_x(\cdot)$ (where $x \in [0 < f < r_0]$) we have*

$$\int_0^T ||\dot{\chi}_x(t)||\, dt \,\leq\, \sigma(f(x)) - \sigma(f(\chi_x(T))),$$

whenever $f(\chi_x(T)) > 0$.

(iii) **[Piecewise subgradient curves have bounded length]** *There exists $M > 0$ such that for all piecewise subgradient curves $\gamma\colon [0, T) \to H$ such that $\gamma([0, T)) \subset [0 < f < r_0]$ we have*

$$\text{length}[\gamma] < M.$$

(iv) **[Talwegs of finite length]** *For all $R > 1$, there exists a piecewise absolutely continuous curve (with countable pieces) $\theta\colon (0, r_0) \to \mathbb{R}^n$ with* finite length *such that*

$$\theta(r) \in \left\{ x \in [f = r] \colon ||\partial f(x)||_- \le R \inf_{y \in [f=r]} ||\partial f(y)||_- \right\}, \qquad \text{for all } r \in (0, r_0).$$

(v) **[Integrability condition]** *The function $u\colon (0, r_0) \to [0, +\infty]$ defined by*

$$u(r) = \frac{1}{\inf\limits_{x \in [f=r]} ||\partial f(x)||_-}, \qquad r \in (0, r_0)$$

is finite-valued and belongs to $L^1(0, r_0)$.

(vi) **[Lipschitz continuity of the sublevel mapping]** *There exists $\varphi \in \mathcal{K}(0, r_0)$ such that*

$$\mathrm{Dist}([f \le r], [f \le s]) \le |\varphi(r) - \varphi(s)| \qquad \text{for all } r, s \in (0, r_0).$$

Proof. The proof is similar to the proof of Theorem 18 and will be omitted. The equivalence between (i) and (vi) is a consequence of Corollary 4. □

3.4. **Application: convergence of the proximal algorithm.** In this subsection we assume that the function $f\colon H \to \mathbb{R} \cup \{+\infty\}$ is *semiconvex* (*cf.* Definition 10). Let us recall the definition of the proximal mapping (see [42, Definition 1.22], for example).

Definition 21 (proximal mapping)**.** Let $\lambda \in (0, \alpha^{-1})$. Then the proximal mapping $\mathrm{prox}_\lambda\colon H \to H$ is defined by

$$\mathrm{prox}_\lambda(x) := \mathrm{argmin} \left\{ f(y) + \frac{1}{2\lambda}||y - x||^2 \right\}, \quad \forall x \in H.$$

Remark 22. The fact that prox_λ is well-defined and single-valued is a consequence of the semiconvex assumption: indeed this assumption implies that the auxiliary function appearing in the aforementioned definition is strictly convex and coercive (see [42], [14] for instance).

Lemma 23 (Subgradient inequality and proximal mapping)**.** *Assume that $f\colon H \to \mathbb{R} \cup \{+\infty\}$ is a semiconvex function that satisfies (i) of Theorem 20. Let $x \in [0 < f < r_0]$ be such that $f(\mathrm{prox}_\lambda x) > 0$. Then*

$$||\mathrm{prox}_\lambda x - x|| \le \varphi(f(x)) - \varphi(f(\mathrm{prox}_\lambda x)). \tag{33}$$

Proof. Set $x^+ = \mathrm{prox}_\lambda(x)$, $r = f(x)$, and $r^+ = f(x^+)$. It follows from the definition of x^+ that $0 < r^+ \le r < r_0$. In particular, for every $u \in [f \le r^+]$ we have

$$||x^+ - x||^2 \le ||u - x||^2 + 2\lambda[f(u) - r^+] \le ||u - x||^2.$$

Therefore by Corollary 4 (Lipschitz continuity of the sublevel mapping) we obtain

$$||x^+ - x|| = \text{dist}\,(x, [f \leq r^+]) \leq \text{Dist}\,([f \leq r], [f \leq r^+]) \leq \varphi(r) - \varphi(r^+).$$

The proof is complete. □

The above result has an important impact in the asymptotic analysis of the *proximal algorithm* (see forthcoming Theorem 24). Let us first recall that, given a sequence of positive parameters $\{\lambda_k\} \subset (0, \alpha^{-1})$ and $x \in H$ the proximal algorithm is defined as follows:

$$Y_x^{k+1} = \text{prox}_{\lambda_k} Y_x^k, \qquad Y_x^0 = x,$$

or in other words

$$\{Y_x^{k+1}\} = \text{argmin}\left\{ f(u) + \frac{1}{2\lambda_k}||u - Y_x^k||^2 \right\}, \qquad Y_x^0 = x.$$

If we assume in addition that $\inf f > -\infty$, then for any initial point x the sequence $\{f(Y_x^k)\}$ is decreasing and converges to a real number L_x.

Theorem 24 (strong convergence of the proximal algorithm)**.** *Let $f\colon H \to \mathbb{R} \cup \{+\infty\}$ be a semiconvex function which is bounded from below. Let $x \in \text{dom}\, f$, $\{\lambda_k\} \subset (0, \alpha^{-1})$ and $L_x := \lim_{k\to\infty} f(Y_x^k)$ and assume that there exists $k_0 \geq 0$ and $\varphi \in \mathcal{K}(0, f(Y_x^{k_0}) - L_x)$ such that*

$$(34) \qquad ||\partial(\varphi \circ [f(\cdot) - L_x])(x)||_- \geq 1, \quad \textit{for all } x \in [L_x < f \leq f(Y_x^{k_0})].$$

Then the sequence $\{Y_x^k\}$ converges strongly to Y_x^∞ and

$$(35) \qquad ||Y_x^\infty - Y_x^k|| \leq \varphi(f(Y_x^k) - L_x), \quad \textit{for all } k \geq k_0.$$

Proof. Since the sequence $\{Y_x^k\}_{k\geq k_0}$ evolves in $L_x \leq f < f(Y_x^{k_0})$, Lemma 23 applies. This yields

$$\sum_{k=p}^{q} ||Y_x^{k+1} - Y_x^k|| \leq \varphi(f(Y_x^{q+1}) - L_x) - \varphi(f(Y_x^p) - L_x),$$

for all integers $k_0 \leq p \leq q$. This implies that Y_x^k converges strongly to Y_x^∞ and that inequality (35) holds. □

Remark 25 (Step-size)**.** "Surprisingly" enough the step-size sequence $\{\lambda_k\}$ does not appear explicitly in the estimate (35), but it is instead hidden in the sequence of values $\{f(Y_x^k)\}$. In practice the choice of the step-size parameters λ_k is however crucial to obtain the convergence of $\{f(Y^k)\}$ to a critical value; standard choices are for example sequences satisfying $\sum \lambda_k = +\infty$ or $\lambda_k \in [\eta, \alpha^{-1})$ for all $k \geq 0$ where $\eta \in (0, \alpha^{-1})$, see [14] for more details.

4. Convexity and KŁ-inequality

In this section, we assume that $f\colon H \to \mathbb{R}\cup\{+\infty\}$ is a lower semicontinuous proper convex function such that $\inf f > -\infty$. Changing f in $f-\inf f$, we may assume that $\inf f = 0$. Let us also denote the set of minimizers of f by

$$C := \operatorname{argmin} f = [f = 0].$$

When C is nonempty, we may assume with no loss of generality that $0 \in C$.

In this convex setting Theorem 13 can be considerably reinforced; related results are gathered in Section 4.1. We also recall well-known facts ensuring that subgradient curves have finite length and provide a new result in that direction (see Theorem 28). In Section 4.2, we give some conditions which ensure that f satisfies the KŁ-inequality and we show that the conclusions of Theorem 20 can somehow be globalized. In section 4.3 we build a counterexample of a C^2 convex function in $\mathbb{R}^2$ which does not satisfy the KŁ-inequality. This counterexample also reveals that the uniform boundedness of the lengths of subgradient curves is a strictly weaker condition than condition (iii) of Theorem 18, which justifies further the introduction of piecewise subgradient curves.

4.1. Lengths of subgradient curves for convex functions.

The following lemma gathers well known complements to Theorem 13 when f is convex.

Lemma 26. *Let $f\colon H \to \mathbb{R}\cup\{+\infty\}$ be a lower semicontinuous proper convex function such that $0 \in C = [f = 0]$. Let $x_0 \in \operatorname{dom} f$.*

(i) If $a \in C$, then

$$\frac{d}{dt}||\chi_{x_0}(t) - a||^2 \leq -2f(\chi_{x_0}(t)) \leq 0 \quad \text{a.e on } (0, +\infty).$$

and therefore $t \mapsto ||\chi_{x_0}(t) - a||$ is nonincreasing.

(ii) The function $t \mapsto f(\chi_{x_0}(t))$ is nonincreasing and converges to $0 = \min f$ as $t \to +\infty$.

(iii) The function $t \in [0, +\infty) \longmapsto ||\partial f(\chi_{x_0}(t)||_-$ is nonincreasing.

(iv) The function $t \mapsto f(\chi_{x_0}(t))$ is convex and belongs to $L^1([0, +\infty))$: for all $T > 0$,

$$\int_0^T f(\chi_{x_0}(t))dt = \frac{1}{2}||x_0||^2 - \frac{1}{2}||\chi_{x_0}(T)||^2 \leq \frac{1}{2}||x_0||^2. \tag{36}$$

(v) For all $T > 0$,

$$\int_0^T ||\dot{\chi}_{x_0}(t)||dt \leq \left(\int_0^{+\infty} f(\chi_{x_0}(t))dt\right)^{1/2} (\log T)^{1/2}. \tag{37}$$

Proof. The proofs of these classical properties can be found in [11, 12]. □

R. Bruck established in [12] that subgradient trajectories of convex functions are always weakly converging to a minimizer in $C = \operatorname{argmin} f$ whenever the latter is nonempty. However, as shown later on by J.-B. Baillon [7], strong convergence does not hold in general.

To the best of our knowledge, the problem of the characterization of length boundedness of subgradient curves for convex functions is still open (see [11, Open problems, pp.167]). In the present framework, the following result of H. Brézis [10, 11] is of particular interest.

Theorem 27 (Uniform boundedness of trajectory lengths [10])**.** *Let $f\colon H \to \mathbb{R} \cup \{+\infty\}$ be a lower semicontinuous proper convex function such that $0 \in C = \operatorname{argmin} f = [f = 0]$. We assume that C has nonempty interior. Then, for all $x_0 \in \operatorname{dom} f$, $\chi_{x_0}(\cdot)$ has finite length. More precisely, if $B(0, \rho) \subset C$, we have, for all $T \geq 0$,*

$$\int_0^T ||\dot{\chi}_{x_0}(t))||dt \leq \frac{1}{2\rho}(||x_0||^2 - ||\chi_{x_0}(T)||^2).$$

Proof. We assume that $B(0, \rho) \subset C$ for some $\rho > 0$ and consider $x_0 \in \operatorname{dom} f \backslash C$ (otherwise there is nothing to prove). Let $t \geq 0$ such that $\chi_{x_0}(t) \notin C$ and $\dot{\chi}_{x_0}(t)$ exists. By convexity, we get

$$\langle -(\chi_{x_0}(t) - \rho u), \dot{\chi}_{x_0}(t)\rangle \geq f(\chi_{x_0}(t)) - f(\rho u) > 0$$

for all u in the unit sphere of H. As a consequence $-\langle \chi_{x_0}(t), \dot{\chi}_{x_0}(t)\rangle > \rho||\dot{\chi}_{x_0}(t)||$. Therefore $\int_0^T ||\dot{\chi}_{x_0}(t)||dt \leq \frac{1}{2\rho}(||x_0||^2 - ||\chi_{x_0}(T)||^2)$. □

The following result is an extension of Theorem 27 under the assumption that the vector subspace $\operatorname{span}(C)$ generated by C, has codimension 1 in H. We denote by $\operatorname{ri}(C)$ the relative interior of C in $\operatorname{span}(C)$.

Theorem 28. *Let $f\colon H \to \mathbb{R} \cup \{+\infty\}$ be a lower semicontinuous proper convex function such that $0 \in C = \operatorname{argmin} f = [f = 0]$. Assume that C generates a subspace of codimension 1 and that the relative interior $\operatorname{ri}(C)$ of C in $\operatorname{span}(C)$ is not empty. If $x_0 \in \operatorname{dom} f$ is such that $\chi_{x_0}(t)$ converges (strongly) to $a \in \operatorname{ri}(C)$ as $t \to +\infty$, then $\operatorname{length}[\chi_{x_0}] < +\infty$.*

Proof. Let us denote by a the limit point of $\chi(t) := \chi_{x_0}(t)$ as t goes to infinity. By assumption a belongs to $\operatorname{ri}(C)$, so that there exists $\delta > 0$ such that $\bar{B}(a, \delta) \cap \operatorname{span}(C) \subset C$. Let $T > 0$ be such that $\chi(t) \in B(a, \delta)$ for all $t \geq T$. Write $\operatorname{span}(C) = \{x \in H : \langle x, x^*\rangle = 0\}$ with $x^* \in H$. We claim that the function $[T, +\infty) \ni t \mapsto h(t) = \langle x^*, \chi(t)\rangle$ has a constant sign. Let us argue by contradiction and assume that there exist $T < t_1 < t_2$ such that $h(t_1) < 0 < h(t_2)$. Hence there exists $t_3 \in (t_1, t_2)$ such that $h(t_3) = 0$. Since

$\chi(t) \in B(a,\delta)$, this implies $\chi(t_3) \in C$ and thus by the uniqueness theorem for subgradient curves (Theorem 13), we have $\chi(t) = \chi(t_3)$ for all $t \geq t_3$ which is a contradiction. Note also that if $h(t_0) = 0$ for some $t_0 \geq T$, then χ has finite length. Indeed applying once more Theorem 13, we deduce that $\chi(t) = \chi(t_0)$ for all $t \geq t_0$, hence

$$\int_0^{+\infty} ||\dot{\chi}|| = \int_0^{t_0} ||\dot{\chi}|| \leq \sqrt{t_0}\sqrt{\int_0^{t_0} ||\dot{\chi}||^2} < +\infty.$$

Assume that h is positive (the case h negative can be treated similarly) and define the following function

$$\tilde{f}(x) = \begin{cases} 0 & \text{if } \langle x, x^* \rangle < 0 \text{ and } x \in \bar{B}(a,\delta) \\ f(x) & \text{if } \langle x, x^* \rangle \geq 0 \text{ and } x \in \bar{B}(a,\delta) \\ +\infty & \text{otherwise.} \end{cases}$$

One can easily check that $\tilde{f}$ is proper, lower semicontinuous, convex and that $\operatorname{argmin} \tilde{f}$ has non empty interior. Note also that $\partial \tilde{f}(x) = \partial f(x)$ for all $x \in B(a,\delta)$ such that $\langle x, x^* \rangle > 0$. The conclusion follows from the previous result and the fact that $\dot{\chi}(t) + \partial \tilde{f}(\chi(t)) \ni 0$ a.e. on $(T, +\infty)$. □

4.2. **KŁ-inequality for convex functions.** The following result shows that if f is convex, then the function φ of Theorem 18(i) can be assumed to be concave and defined on $[0,\infty)$.

Theorem 29 (Subgradient inequality – convex case)**.** *Let $f\colon H \to \mathbb{R} \cup \{+\infty\}$ be a lower semicontinuous proper convex function which is bounded from below (recall that $\inf f = 0$). The following statements are equivalent:*

(i) There exist $r_0 > 0$ and $\varphi \in \mathcal{K}(0, r_0)$ such that

$$||\partial(\varphi \circ f)(x)||_- \geq 1, \quad \textit{for all } x \in [0 < f \leq r_0].$$

*(ii) There exists a **concave** function $\psi \in \mathcal{K}(0,\infty)$ such that*

$$||\partial(\psi \circ f)(x)||_- \geq 1, \quad \textit{for all } x \notin [f = 0]. \tag{38}$$

Proof. The implication (ii)$\Longrightarrow$(i) is obvious. To prove (i)$\Longrightarrow$(ii) let us first establish that the function

$$r \in (0, +\infty) \longmapsto u(r) = \frac{1}{\inf\limits_{x \in [f=r]} ||\partial f(x)||_-}$$

is finite-valued and nonincreasing. Let $0 < r_2 < r_1$ and let us show that $u(r_2) \geq u(r_1)$. To this end we may assume with no loss of generality that $u(r_1) > 0$ (and therefore that $[f = r_1] \cap \operatorname{dom} \partial f$ is nonempty). Take $\epsilon > 0$ and let $x_1 \in [f = r_1]$ and $p_1 \in \partial f(x_1)$ such that $u(r) \leq \frac{1}{||p_1||} + \epsilon$. Since the continuous function $t \mapsto f(\chi_{x_1}(t))$ tends to $\inf f = 0$ as t goes to infinity (see [32] for instance), there exists $t_2 > 0$ such that $f(\chi_{x_1}(t_2)) = r_2$. From Lemma 26 (iii), we obtain

$$\frac{1}{||\partial f(\chi_{x_1}(t_2)||_-} \geq \frac{1}{||p_1||} \geq u(r_1) - \epsilon,$$

which yields $u(r_2) \geq u(r_1)$. By (i) the function u is finite-valued on $(0, r_0)$, thus, since u is nonincreasing, it is also finite-valued on $(0, +\infty)$.

It is easy to see that [(i)⇒(v)] of Theorem 18 holds without the compactness assumption (24) (see Remark 19). It follows that $u \in L^1(0, r_0)$ and by Lemma 44 (Annex) that there exists a decreasing continuous function $\tilde{u} \in L^1(0, r_0)$ such that $\tilde{u} \geq u$. Reproducing the proof of $(v) \Rightarrow (i)$ of Theorem 18 we obtain a strictly increasing, concave, C^1 function

$$\psi(r) := \int_0^r \tilde{u}(s)ds$$

for which (38) holds for all $x \in [0 < f < r_0]$. Fix $\bar{r} \in (0, r_0)$ and take ψ as above. Applying (38) and using the fact that $u(r)$ is decreasing we obtain

$$1 \leq \psi'(\bar{r})u(\bar{r})^{-1} \leq \psi'(\bar{r})u(r)^{-1} \leq \psi'(\bar{r})||p||,$$

for all $p \in \partial f(x)$, $x \in [\bar{r} \leq f]$ and $r \in (\bar{r}, +\infty)$ such that $u(r) > 0$. This shows that the function $\Psi : \mathbb{R}_+ \to \mathbb{R}_+$ defined by

$$\Psi(r) := \begin{cases} \psi(r) & \text{if } r \leq \bar{r}, \\ \psi(\bar{r}) + \psi'(\bar{r})(r - \bar{r}) & \text{otherwise.} \end{cases}$$

satisfies the required properties. □

A natural question arises: when does a convex function f satisfy the KŁ–inequality? In finite-dimensions a quick positive answer can be given whenever f belongs to an o-minimal structure (convexity then becomes superflous). The following result gives an alternative criterion when f is not extremely "flat" around its set of minimizers. More precisely, we assume the following growth condition:

(39)
$$\begin{cases} \text{There exists } m\colon [0, +\infty) \to [0, +\infty) \text{ and } S \subset H \text{ such that} \\ m \text{ is continuous, increasing, } m(0) = 0,\ f \geq m(\operatorname{dist}(\cdot, C)) \text{ on } S \cap \operatorname{dom} f \\ \text{and } \displaystyle\int_0^\rho \frac{m^{-1}(r)}{r} dr < +\infty \text{ (for some } \rho > 0). \end{cases}$$

Theorem 30 (growth assumptions and Kurdyka-Łojasiewicz inequality)**.** *Let $f\colon H \to \mathbb{R}\cup\{+\infty\}$ be a lower semicontinuous proper convex function satisfying (39) and let us assume $0 \in C := \operatorname{argmin} f$. Then the KŁ–inequality holds,* i.e.

$$||\partial(\varphi\circ f)(x)||_- \geq 1, \textit{ for all } x \in S\setminus \operatorname{argmin} f,$$

with

$$\varphi(r) = \int_0^r \frac{m^{-1}(s)}{s}ds.$$

Proof. Let $x \in S\cap \operatorname{dom}\partial f$ and a be the projection of x onto the convex subset $C = \operatorname{argmin} f$. Using the convex inequality we have

$$f(x) - f(a) \leq \langle \partial^0 f(x), x-a\rangle \leq \text{dist }(0,\partial f(x)) \text{ dist }(x,C) \leq$$

$$\leq \text{dist }(0,\partial f(x))\, m^{-1}(f(x)-f(a)).$$

Using the chain rule (see Lemma 43) an the fact that $f(a) = 0$, we obtain dist $(0, \partial(\varphi\circ f)(x)) \geq 1$ where φ is as above (note that $\varphi \in \mathcal{K}(0,\rho)$). □

Remark 31. Assume that H isfinite-dimensional, and let S be a compact convex subset of H which satisfies $S\cap C \neq \emptyset$. Then there exists a convex continuous increasing function $m\colon \mathbb{R}_+ \to \mathbb{R}_+$ with $m(0) = 0$ such that $f(x) \geq m(\text{dist}(x,C))$ for all $x \in S$.

Sketch of the proof. With no loss of generality we assume that $0 \in S \cap C$. Using the Moreau-Yosida regularization (see [11] for instance), we obtain the existence of a finite-valued convex continuous function $g\colon H \to \mathbb{R}$ such that $f \geq g$ and $\operatorname{argmin} f = \operatorname{argmin} g$. Set $\alpha = \max\{\text{dist }(x,C) : x \in S\}$ and $m_0(s) = \min\{g(x) : x \in S,\ \text{dist }(x,C) \geq s\} \in \mathbb{R}_+$ for all $s \in [0,\alpha]$. Let $0 \leq s_1 < s_2 \leq \alpha$, and let $x_2 \in S$ be such that dist $(x_2,C) \geq s_2$ and $0 < g(x_2) = m(s_2)$. Using the convexity of g and the fact that $0 \in \operatorname{argmin} g \cap S$, we see that there exists $\lambda \in (0,1)$ such that $g(\lambda x_2) < g(x_2)$, $\lambda x_2 \in S$ (recall that S is convex and contains 0), and dist $(\lambda x_2, C) \geq s_1$. This shows that the function m_0 is finite-valued increasing on $[0,\alpha]$ and satisfies $m_0(\text{dist }(x,C)) \leq g(x) \leq f(x)$ for any $x \in S$. Applying Lemma 45 (Annex) to m_0, we obtain a smooth increasing finite-valued function m such that $0 < m(s) \leq m_0(s)$ for $s \in [0,\alpha]$ with $m(0) = 0$. The conclusion follows by extending m to an increasing continuous function on $\mathbb{R}_+$. □

Example 32. Take $0 < \alpha < 1$. If $m(r) = \exp(-1/r^\alpha)$ and $m(0) = 0$, then for $0 \leq s \leq \rho < 1$ we have $m^{-1}(s) = 1/(-\log s)^{1/\alpha}$ and

$$\int_0^\rho \frac{m^{-1}(s)}{s}ds < +\infty.$$

Therefore any convex function which is minorized by the function $x \mapsto \exp(-1/\mathrm{dist}(x,C)^\alpha)$ in some neighborhood of $C = \operatorname{argmin} f$ satisfies the KŁ–inequality.

4.3. **A smooth convex counterexample to the KŁ–inequality.** In this section we construct a C^2 convex function on $\mathbb{R}^2$ with compact level sets that fails to satisfy the KŁ–inequality. This counterexample is constructed as follows:

- we first note that any sequence of sublevel sets of a convex function that satisfies the KŁ–inequality must comply with a specific property;
- we build a sequence T_k of nested convex sets for which this property fails;
- we show that there exists a smooth convex function which admits T_k as sublevel sets.

The last part relies on the use of support functions and on a result of Torralba [41]. For any closed convex subset T of $\mathbb{R}^n$, we define its support function by $\sigma_T(x^*) = \sup_{x\in T}\langle x, x^*\rangle$ for all $x^* \in \mathbb{R}^n$. Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be a convex function and $x^* \in \mathbb{R}^n$. Fenchel has observed, see [23], that the function $\lambda \mapsto \sigma_{[f\leq\lambda]}(x^*)$ is concave and nondecreasing. The following result asserts that this fact provides somehow a sufficient condition to rebuild a convex function starting from a countable family of nested convex sets.

Theorem 33 (Convex functions with prescribed level sets [41])**.** *Let $\{T_k\}$ be a nonincreasing sequence of convex compact subsets of $\mathbb{R}^n$ such that int $T_k \supset T_{k+1}$ for all $k \geq 0$. For every $k > 0$ we set:*

$$K_k = \max_{||x^*||=1} \frac{\sigma_{T_{k-1}}(x^*) - \sigma_{T_k}(x^*)}{\sigma_{T_k}(x^*) - \sigma_{T_{k+1}}(x^*)} \in (0, +\infty).$$

Then for every strictly decreasing sequence $\{\lambda_k\}$, starting from $\lambda_0 > 0$ and satisfying

$$0 < K_k(\lambda_k - \lambda_{k+1}) \leq \lambda_{k-1} - \lambda_k, \text{ for each } k > 0,$$

there exists a continuous convex function f such that

$$T_k = [f \leq \lambda_k], \quad \text{for every } k \in \mathbb{N}$$

and being maximal with this property.

Remark 34. (i) If $\{\lambda_k\}$ is as in the above theorem and $x^* \in \mathbb{R}^n \backslash \{0\}$, we have

$$\lambda_k - \lambda_{k+1} \leq \frac{\lambda_0 - \lambda_1}{\sigma_{T_0}(x^*) - \sigma_{T_1}(x^*)}(\sigma_{T_k}(x^*) - \sigma_{T_{k+1}}(x^*)).$$

Since the sum $\sum(\sigma_{T_k}(x^*) - \sigma_{T_{k+1}}(x^*))$ converges, so does the sum $\sum(\lambda_k - \lambda_{k+1})$, yielding the existence of the limit $\lim \lambda_k$. Since f is the

greatest function admitting $\{T_k\}$ as prescribed sublevel sets, we obtain $\min f = \lim \lambda_k$.
(ii) Let $k \geq 0$ and $\lambda \in [\lambda_{k+1}, \lambda_k]$. The function f satisfies further

$$[f \leq \lambda] = \left(\frac{\lambda - \lambda_{k+1}}{\lambda_k - \lambda_{k+1}}\right) T_k + \left(\frac{\lambda_k - \lambda}{\lambda_k - \lambda_{k+1}}\right) T_{k+1}, \tag{40}$$

see [41, Remark 5.9].

The following lemma provides a decreasing sequence of convex compact subsets in $\mathbb{R}^2$ which can not be a sequence of prescribed sublevel sets of a function satisfying the KŁ–inequality (see the *conclusion* part at the end of the proof of Theorem 36).

Lemma 35. *There exists a decreasing sequence of compact subsets $\{T_k\}_k$ in $\mathbb{R}^2$ such that:*
(i) T_0 *is the unit disk* $D := B(0, 1)$*;*
(ii) $T_{k+1} \subset$ *int* T_k *for every* $k \in \mathbb{N}$*;*
(iii) $\bigcap_{k\in\mathbb{N}} T_k$ *is the disk* $D_r := B(0, r)$ *for some* $r > 0$*;*
(iv) $\sum_{k=0}^{+\infty} \mathrm{Dist}(T_k, T_{k+1}) = +\infty$.

Proof. We proceed by constructing the boundaries ∂T_k of T_k for each integer k. Let $C_{2,3}$ denote the circle of radius 1 and let us define recursively a sequence of closed convex curves $C_{n,m}$ for $n \geq 3$ and $1 \leq m \leq n + 1$; we assume that $C_{n-1,n}$ is the circle of radius $R_n > 0$. Let $\{\mu_n\}$ be a sequence in $(0, 1)$ that will be chosen later in order to satisfy (iii). Then, for $1 \leq m \leq n$, let us define $C_{n,m}$ to be the union of the segments:

– $\left[\mu_n^m R_n \exp\left(2i\pi(\frac{j}{n})\right), \mu_n^m R_n \exp\left(2i\pi(\frac{j+1}{n})\right)\right]$ for $0 \leq j \leq m-1$ (here i stands for the imaginary unit) and the circle-arc:
– $\mu_n^m R_n \exp(i\theta)$ for $2\pi \frac{m}{n} \leq \theta \leq 2\pi$.

In other words, $C_{n,m}$ consists of the first m edges of a regular convex n-gonon inscribed in a circle of radius $\mu_n^m R_n$ and a circle-arc of the same radius to close the curve. We then set

$$R_{n+1} = \mu_n^{n+1} R_n \cos\left(\frac{\pi}{n}\right)$$

and define $C_{n,n+1}$ to be the circle of radius R_{n+1}. Figure 1 illustrates the curves $C_{4,5}$ and $C_{5,m}$ for $m = 1, \ldots, 6$.

Ordering $\{(n, m) :\geq 3, 1 \leq m \leq n + 1\}$ lexicographically we define succesively the convex subset T_k to be the convex envelope of the set $C_{n,m}$. By construction (i) and (ii) are satisfied. Item (iii) holds if $\lim R_n > 0$ which

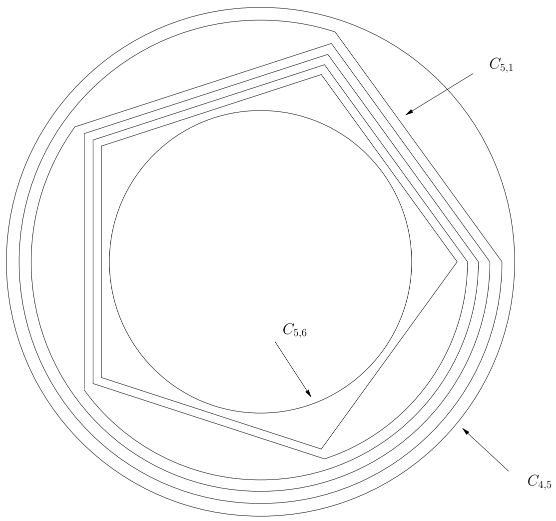

FIGURE 1. *The curves* $C_{4,5}, C_{5,1}$ *to* $C_{5,6}$

is equivalent to the fact that the infinite product $\Pi_{n=3}^{+\infty} \mu_n^{n+1} \cos(\pi/n)$ does not converge to 0. This can be achieved by taking $\mu_n = 1 - 1/n^3$. Let $r > 0$ be the limit of $\{R_n\}$. The intersection of the convex sets T_n is the disk of radius r.

Take $n \geq 3$. Considering the middle of the segment

$$\left[\mu_n R_n, \mu_n R_n \exp\left(\frac{2i\pi}{n}\right)\right]$$

in $C_{n,1}$ and the point $R_n \exp(\frac{i\pi}{n}) \in C_{n-1,n}$, we obtain $\mathrm{Dist}(C_{n,1}, C_{n-1,n}) = R_n(1 - \mu_n \cos(\pi/n))$. If $2 \leq m \leq n$, considering the middle of

$$\left[\mu_n^m R_n \exp\Big(\frac{2i\pi(m-1)}{n}\Big), \mu_n^m R_n \exp\Big(\frac{2i\pi m}{n}\Big)\right]$$

in $C_{n,m}$ and the point $\mu_n^{m-1} R_n \exp\big(\frac{i\pi(2m-1)}{n}\big) \in C_{n,m-1}$, we get $\mathrm{Dist}(C_{n,m}, C_{n,m-1}) = \mu_n^{m-1} R_n(1 - \mu_n \cos(\pi/n))$. Finally considering the points $\mu_n^n R_n \in C_{n,n}$ and $\mu_n^{n+1} \cos(\pi/n) R_n \in C_{n,n+1}$, we obtain

$$\mathrm{Dist}(C_{n,n}, C_{n,n+1}) = \mu_n^n R_n (1 - \mu_n \cos(\pi/n)).$$

Thus

$$\mathrm{Dist}(C_{n,1}, C_{n-1,n}) + \sum_{m=2}^{n+1} \mathrm{Dist}(C_{n,m}, C_{n,m-1}) =$$
$$= \sum_{m=1}^{n+1} \mu_n^{m-1} R_n (1 - \mu_n \cos \frac{\pi}{n}) \sim nr\frac{\pi^2}{2n^2} = \frac{\pi^2 r}{2n}.$$

Hence (iv) holds. □

For $\theta \in \mathbb{R}/2\pi\mathbb{Z}$, set $n(\theta) = (\cos\theta, \sin\theta)$ and $\tau(\theta) = (-\sin\theta, \cos\theta)$. We say that a closed C^2 *curve* C in $\mathbb{R}^2$ is convex if its curvature has constant sign. If moreover the curvature never vanishes, then there exists a C^1 *parametrization* $c\colon \mathbb{R}/2\pi\mathbb{Z} \to C$ of C, called *parametrization of C by its normal*, such that the unit tangent vector at $c(\theta)$ is $\tau(\theta)$. In this case $n(\theta)$ is the outward normal to the convex envelope of C at $c(\theta)$. Moreover, c is C^∞, whenever C is so. In this case, we denote by $\rho_c(\theta)$ the curvature radius of c at $c(\theta)$ and we have

$$\dot{c}(\theta) = \rho_c(\theta)\tau(\theta).$$

Let us denote by T the convex envelope of C. Using the fact that n defines the outward normals to T, we get

$$\langle c(\theta), n(\theta)\rangle = \max_{x\in T}\langle x, n(\theta)\rangle = \sigma_T(n(\theta)), \quad \forall \theta \in \mathbb{R}/2\pi\mathbb{Z}.$$

Theorem 36 (convex counterexample)**.** *There exists a C^2 convex function $f\colon \mathbb{R}^2 \to \mathbb{R}$ such that $\min f = 0$ which does not satisfy the KL–inequality and whose set of minimizers is compact with nonempty interior. More precisely, for each $r > 0$ and for each desingularization function $\varphi \in \mathcal{K}(0, r)$ we have*

$$\inf\{\|\nabla(\varphi \circ f)(x)\| : x \in [0 < f < r]\} = 0.$$

Remark 37. (i) It can be seen from the forthcoming proof that $\operatorname{argmin} f$ is the closed disk centered at 0 of radius r, and that f is actually C^∞ on the complement of the circle of radius r.

(ii) The fact that f is C^2 shows that KŁ–inequality is not related to the smoothness of f. Besides, it seems clear from the proof that a C^k (k arbitrary) counterexample could be obtained.

(iii) Since $\operatorname{argmin} f$ has nonempty interior, Theorem 27 shows that the lengths of subgradient curves are uniformly bounded. Using the notation and the results of Theorem 20, we see that the function f shows that the uniform boundedness of the lengths of the subgradient curves (starting from a given level set $[f = r_0]$) does not yield the uniform boundedness of the lengths of the piecewise subgradient curves γ lying in $[\min f < f < r_0]\}$.

Proof of Theorem 36. Let M, N be topological finite-dimensional manifolds. In this proof, a mapping $F\colon M \to N$ is said to be *proper* if for each compact subset K of N, $F^{-1}(K)$ is a compact subset of M.

Smoothing the sequence T_k. Let us consider a sequence of convex compact sets $\{T_k\}$ as in Lemma 35. Set $C_k = \partial T_k$ and consider a positive sequence ϵ_k such that $\sum \epsilon_k < +\infty$ with $\epsilon_k + \epsilon_{k+1} < \mathrm{Dist}(T_k, T_{k+1}) = \mathrm{Dist}(C_k, C_{k+1})$ for each integer k. The ϵ_k-neighborhood of C_k can be seen to be disjoint from the $\epsilon_{k'}$-neighborhood of $C_{k'}$ whenever $k \neq k'$. We can deform C_k into a C^∞ convex closed curve $\widetilde{C}_k$ whose curvature never vanishes, lying in the ϵ_k-neighborhood of C_k. This smooth deformation can be achieved by letting C_k evolve under the mean-curvature flow during a very short time, see [22] for the smoothing aspects and [25, 43] for the positive curvature results. We set $\widetilde{T}_k$ to be the closed convex envelope of $\widetilde{C}_k$. This process yields a decreasing sequence of compact convex sets $\{\widetilde{T}_k\}$, that satisfies the conditions of Lemma 35. We note that the circle of radius 1 has non-zero curvature and we set $C_0 = \widetilde{C}_0$. Since $\mathrm{Dist}(\widetilde{T}_k, \widetilde{T}_{k+1}) \geq \mathrm{Dist}(T_k, T_{k+1}) - (\epsilon_k + \epsilon_{k+1})$ and $\sum \epsilon_k < +\infty$, condition (iv) holds. With no loss of generality we may therefore assume that for each $k \geq 0$ the curve ∂T_k is smooth and can be parametrized by its normal.

Let K_k be as in Theorem 33, let λ_0 and λ_1 be such that $\lambda_0 > \lambda_1$. We define λ_k recursively by

$$K_k(\lambda_k - \lambda_{k+1}) = \frac{1}{2}(\lambda_{k-1} - \lambda_k). \tag{41}$$

Because of (41), Theorem 33 yields a continuous convex function $f : T_0 \to \mathbb{R}$ such that $T_k = [f \leq \lambda_k]$. Since f is the greatest function with this property, we deduce that $\min f = \lim \lambda_k$ and $\operatorname{argmin} f = \cap_{k\in\mathbb{N}} T_k$.

Smoothing the function f on $\mathbb{R}^n \setminus \operatorname{argmin} f$. We can easily extend f outside T_0 into a smooth convex function. Let us examine the restriction of f to T_0. Since ∂T_k can be parametrized by its normal, we denote by $c_k : \mathbb{R}/2\pi\mathbb{Z} \to \mathbb{R}^2$ this parametrization. Let us fix $k \in \mathbb{N}$. Let θ be in $\mathbb{R}/2\pi\mathbb{Z}$. Using Remark 34 (b), we obtain

$$\begin{aligned}
&\max_{x\in[f\leq\lambda]} \langle x, n(\theta)\rangle = \\
&\quad= \left(\frac{\lambda - \lambda_{k+1}}{\lambda_k - \lambda_{k+1}}\right) \max_{x\in T_k} \langle x, n(\theta)\rangle + \left(\frac{\lambda_k - \lambda}{\lambda_k - \lambda_{k+1}}\right) \max_{x\in T_{k+1}} \langle x, n(\theta)\rangle = \\
&\quad= \left(\frac{\lambda - \lambda_{k+1}}{\lambda_k - \lambda_{k+1}}\right) \langle c_k(\theta), n(\theta)\rangle + \left(\frac{\lambda_k - \lambda}{\lambda_k - \lambda_{k+1}}\right) \langle c_{k+1}(\theta), n(\theta)\rangle = \\
&\quad= \left\langle \left(\frac{\lambda - \lambda_{k+1}}{\lambda_k - \lambda_{k+1}}\right) c_k(\theta) + \left(\frac{\lambda_k - \lambda}{\lambda_k - \lambda_{k+1}}\right) c_{k+1}(\theta), n(\theta) \right\rangle.
\end{aligned}$$

Using (40) once more we obtain

$$\left(\frac{\lambda-\lambda_{k+1}}{\lambda_k-\lambda_{k+1}}\right)c_k(\theta)+\left(\frac{\lambda_k-\lambda}{\lambda_k-\lambda_{k+1}}\right)c_{k+1}(\theta)\in[f\le\lambda]. \tag{42}$$

Since the above maximum is achieved in $[f=\lambda]$, it follows that

$$f\left(\left(\frac{\lambda-\lambda_{k+1}}{\lambda_k-\lambda_{k+1}}\right)c_k(\theta)+\left(\frac{\lambda_k-\lambda}{\lambda_k-\lambda_{k+1}}\right)c_{k+1}(\theta)\right)=\lambda. \tag{43}$$

Let us define $G\colon \mathbb{R}\times\mathbb{R}/2\pi\mathbb{Z}\to\mathbb{R}^2$ by

$$G(\lambda,\theta)=\left(\frac{\lambda-\lambda_{k+1}}{\lambda_k-\lambda_{k+1}}\right)c_k(\theta)+\left(\frac{\lambda_k-\lambda}{\lambda_k-\lambda_{k+1}}\right)c_{k+1}(\theta).$$

The map G is clearly C^∞. Since $\dfrac{\partial G}{\partial\lambda}=\dfrac{c_k(\theta)-c_{k+1}(\theta)}{\lambda_k-\lambda_{k+1}}$, we have

$$\begin{aligned}\left\langle\frac{\partial G}{\partial\lambda},n(\theta)\right\rangle&=\left\langle\frac{c_k(\theta)-c_{k+1}(\theta)}{\lambda_k-\lambda_{k+1}},n(\theta)\right\rangle=\\&=\frac{\langle c_k(\theta),n(\theta)\rangle-\langle c_{k+1}(\theta),n(\theta)\rangle}{\lambda_k-\lambda_{k+1}}\\&=\frac{\max_{x\in T_k}\langle x,n(\theta)\rangle-\max_{x\in T_{k+1}}\langle x,n(\theta)\rangle}{\lambda_k-\lambda_{k+1}}>0.\end{aligned}$$

On the other hand

$$\frac{\partial G}{\partial\theta}=\left(\left(\frac{\lambda-\lambda_{k+1}}{\lambda_k-\lambda_{k+1}}\right)\rho_{c_k}(\theta)+\left(\frac{\lambda_k-\lambda}{\lambda_k-\lambda_{k+1}}\right)\rho_{c_{k+1}}(\theta)\right)\tau(\theta). \tag{44}$$

Since $\rho_{c_k}>0$ and $\rho_{c_{k+1}}>0$, G is a local diffeomorphism on $(\lambda_{k+1}-\delta, \lambda_k+\delta)\times\mathbb{R}/2\pi\mathbb{Z}$ for any $\delta>0$ sufficiently small. In view of (42), we have $G(\lambda,\theta)\in[\lambda_{k+1}\le f\le\lambda_k]$ for $\lambda_{k+1}\le\lambda\le\lambda_k$ and $G(\lambda,\theta)\in[\lambda_{k+1}<f<\lambda_k]$ for $\lambda_{k+1}<\lambda<\lambda_k$. Since the map $\widetilde{G}:[\lambda_{k+1},\lambda_k]\times\mathbb{R}/2\pi\mathbb{Z}\to[\lambda_{k+1}\le f\le\lambda_k]$ defined by $\widetilde{G}(\lambda,\theta)=G(\lambda,\theta)$ is proper, $\widetilde{G}$ is a covering map from $[\lambda_{k+1},\lambda_k]\times\mathbb{R}/2\pi\mathbb{Z}$ to $[\lambda_{k+1}\le f\le\lambda_k]$. The set $[\lambda_{k+1}\le f\le\lambda_k]$ is connected, thus $\widetilde{G}$ is onto. Using (42) and $G(\lambda_k,\theta)=c_k(\theta)$, one sees that (λ_k,θ) is the only antecedent of $c_k(\theta)$ by $\widetilde{G}$ and, since $[\lambda_{k+1},\lambda_k]\times\mathbb{R}/2\pi\mathbb{Z}$ is connected, $\widetilde{G}$ is injective. Thus $\widetilde{G}$ is a C^∞ diffeomorphism (see [31, Proposition 2.19]). By (42), this implies that the restriction of f to $[\lambda_{k+1}\le f\le\lambda_k]$ is C^∞. Using (42), we know that the level line $[f=\lambda]$ (for $\lambda_{k+1}\le\lambda\le\lambda_k$) is parametrized by $G(\lambda,\theta)$ for $\theta\in\mathbb{R}/2\pi\mathbb{Z}$; if c_λ denotes this parametrization, then $c_k=c_{\lambda_k}$. Besides, by (44), c_λ is a parametrization by the normal and ρ_{c_λ} is a convex combination of ρ_{c_k} and $\rho_{c_{k+1}}$, hence $\rho_{c_\lambda}>0$.

Let us compute ∇f at $c_\lambda(\theta)$. Equation (42) yields

$$1 = \langle \nabla f(G(\lambda,\theta)), \frac{\partial G}{\partial \lambda}(\lambda,\theta)\rangle.$$

Besides we also know that the normal to $[f = \lambda]$ at $c_\lambda(\theta)$ is $n(\theta)$. Since the gradient $\nabla f(G(\lambda,\theta))$ and the normal $n(\theta)$ are linearly dependent, we obtain

$$\nabla f(c_\lambda(\theta)) = \frac{\lambda_k - \lambda_{k+1}}{\langle c_{\lambda_k}(\theta) - c_{\lambda_{k+1}}(\theta), n(\theta)\rangle} n(\theta). \tag{45}$$

Note that this expression does not depend on $\lambda \in [\lambda_{k+1} - \lambda_k]$.

Before going further let us observe/recall two facts.

– First using the aforementioned result of Fenchel [23], we deduce from the convexity of f that the function

$$\lambda \mapsto \langle c_\lambda(\theta), n(\theta)\rangle = \sigma_{[f\le\lambda]}(n(\theta)) \text{ is concave and increasing.} \tag{46}$$

– Let λ and λ' be such that $\lambda_{k+1} \le \lambda \le \lambda' \le \lambda_k$. We have :

$$c_\lambda(\theta) = \left(\frac{\lambda - \lambda_{k+1}}{\lambda' - \lambda_{k+1}}\right) c_{\lambda'}(\theta) + \left(\frac{\lambda' - \lambda}{\lambda' - \lambda_{k+1}}\right) c_{\lambda_{k+1}}(\theta), \tag{47}$$

$$c_{\lambda'}(\theta) = \left(\frac{\lambda' - \lambda}{\lambda_k - \lambda}\right) c_{\lambda_k}(\theta) + \left(\frac{\lambda_k - \lambda'}{\lambda_k - \lambda}\right) c_\lambda(\theta). \tag{48}$$

(*Smoothing f around* $[f = \lambda_k]$.) We have seen that the function f is C^∞ on the complement of the union of the level lines $[f = \lambda_k]$ for $k \in \mathbb{N}$. In order to go further we need to modify f around each $[f = \lambda_k]$.

Consider a positive sequence $\{\epsilon_k\}$ such that $\sum_i \epsilon_i < +\infty$ and $\epsilon_k + \epsilon_{k+1} < \mathrm{Dist}(T_k, T_{k+1}) = \mathrm{Dist}([f = \lambda_k], [f = \lambda_{k+1}])$ for eachinteger k. Let us assume that there exists a sequence $f_k \colon \mathbb{R}^2 \to \mathbb{R}$ of convex functions such that:

(P1) $f_0 = f$;
(P2) $f_k = f_{k-1}$ outside an ϵ_k-neighborhood of $[f = \lambda_k]$;
(P3) f_k is C^∞ in $[f > \lambda_{k+1}]$;
(P4) $\|\nabla f_k\|$ is bounded in $[f \le \lambda_k]$ by the maximum of $\|\nabla f\|$ in $[\lambda_k \le f \le \lambda_{k-1}]$.

Let us choose $k \geq 1$ and λ, λ' such that $\lambda_{k+1} \leq \lambda \leq \lambda_k \leq \lambda' \leq \lambda_{k-1}$. Then by (41) and (45) we have:

$$\|\nabla f(c_\lambda(\theta))\| = \frac{\lambda_k - \lambda_{k+1}}{\langle c_{\lambda_k}(\theta) - c_{\lambda_{k+1}}(\theta), n(\theta)\rangle} \leq$$

$$\leq \frac{1}{2} \frac{\lambda_{k-1} - \lambda_k}{\langle c_{\lambda_{k-1}}(\theta) - c_{\lambda_k}(\theta), n(\theta)\rangle} =$$

$$= \frac{1}{2} \|\nabla f(c_{\lambda'}(\theta)\|.$$

Hence

$$\max_{[\lambda_{k+1} \leq f \leq \lambda_k]} \|\nabla f\| \leq \frac{1}{2} \max_{[\lambda_k \leq f \leq \lambda_{k-1}]} \|\nabla f\|. \tag{49}$$

Combining with (P4), the above implies that the sequence $(f_k)_{k\in\mathbb{N}}$ is uniformly Lipschitz continuous. Applying Ascoli compactness theorem we obtain that f_k converge to a continuous function $\tilde{f}$ which is convex. From (P2) and (P3), we obtain successively that $\tilde{f}$ has the same set of minimizers as f, f is C^∞ outside $\operatorname{argmin} \tilde{f}$, $[\tilde{f} = \lambda_k]$ is in the ϵ_k-neighborhood of $[f = \lambda_k]$. Moreover (49) and (P4) imply that $\|\nabla \tilde{f}(x)\|$ goes to zero as x approaches $\operatorname{argmin} \tilde{f}$, hence $\tilde{f}$ is globally C^1. Note also, that the sequence of level sets $[\tilde{f} \leq \lambda_k]$ satisfies the hypothesis (iv) of Lemma 35. As shown in the conclusion, $\tilde{f}$ provides a C^1 counterexample to the KŁ–inequality.

Let us define such a sequence $\{f_k\}$ by induction. Assume that f_{k-1} is defined. In order to construct f_k, it suffices to proceed in the ϵ_k-neighborhood of $[f = \lambda_k]$. Let $\epsilon > 0$ such that $[\lambda_k - 2\epsilon \leq f \leq \lambda_k + 2\epsilon]$ is in the ϵ_k-neighborhood of $[f = \lambda_k]$. Let us consider a C^∞ function $\mu_- : [-2\epsilon, 2\epsilon] \to \mathbb{R}$ which satisfies the following properties:

1. μ_- is nonincreasing,
2. $\mu_-'' \geq 0$,
3. $\mu_-(\lambda) = -\lambda/\epsilon$ on $[-2\epsilon, -\epsilon/2]$,
4. $\mu_-(\lambda) = 0$ on $[\epsilon/2, 2\epsilon]$.

Let us then define $\mu_+(\lambda) := \lambda/\epsilon + \mu_-(\lambda)$ and $\mu_0 = 1 - (\mu_- + \mu_+)$. The function μ_+ satisfies

1′. μ_+ is nondecreasing,
2′. $\mu_+'' = \mu_-'' \geq 0$,
3′. $\mu_+(\lambda) = 0$ on $[-2\epsilon, -\epsilon/2]$,
4′. $\mu_+(\lambda) = \lambda/\epsilon$ on $[\epsilon/2, 2\epsilon]$.

Set $c_- = c_{\lambda_k-\epsilon}$, $c_0 = c_{\lambda_k}$, $c_+ = c_{\lambda_k+\epsilon}$ and

$$M_-(\theta) = \langle c_-(\theta), n(\theta)\rangle = \max_{x\in[f\le\lambda_k-\epsilon]}\langle x, n(\theta)\rangle,$$

$$M_0(\theta) = \langle c_0(\theta), n(\theta)\rangle = \max_{x\in[f\le\lambda_k]}\langle x, n(\theta)\rangle,$$

$$M_+(\theta) = \langle c_+(\theta), n(\theta)\rangle = \max_{x\in[f\le\lambda_k+\epsilon]}\langle x, n(\theta)\rangle.$$

For $(\lambda,\theta)\in[-2\epsilon,2\epsilon]\times\mathbb{R}/2\pi\mathbb{Z}$, we define:

$$H(\lambda,\theta) = \mu_-(\lambda)c_-(\theta) + \mu_0(\lambda)c_0(\theta) + \mu_+(\lambda)c_+(\theta).$$

Then H is a C^∞ map and for any $\lambda\in[-\epsilon,\epsilon]$, we have $\mu_-(\lambda),\mu_0(\lambda)$ and $\mu_+(\lambda)$ in $[0,1]$. Since $H(\lambda,\theta)$ is a convex combination of points in $[f\le\lambda_k+\epsilon]$, we deduce $H(\lambda,\theta)\in[f\le\lambda_k+\epsilon]$ and $H(\lambda,\theta)\in[f<\lambda_k+\epsilon]$ whenever $\lambda<\epsilon$ and $\mu_+(\lambda)<1$. Since

$$\langle H(\lambda,\theta), n(\theta)\rangle = \mu_-(\lambda)M_-(\theta) + \mu_0(\lambda)M_0(\theta) + \mu_+(\lambda)M_+(\theta) \ge M_-(\theta),$$

we get $H(\lambda,\theta)\in[f\ge\lambda_k-\epsilon]$, and $H(\lambda,\theta)\in[f>\lambda_k-\epsilon]$ whenever $\lambda>\epsilon$, $\mu_-(\lambda)<1$. It follows that

$$\frac{\partial H}{\partial\lambda} = \mu_-'(\lambda)c_-(\theta) + \mu_0'(\lambda)c_0(\theta) + \mu_+'(\lambda)c_+(\theta).$$

Since $\mu_0' = -\mu_-' - \mu_+'$, items 1 and $1'$ entail

$$\begin{aligned}\Big\langle \frac{\partial H}{\partial\lambda}, n(\theta)\Big\rangle &= \mu_+'(\lambda)\langle c_+(\theta)-c_0(\theta), n(\theta)\rangle - \mu_-'(\lambda)\langle c_0(\theta)-c_-(\theta), n(\theta)\rangle\\ &= \mu_+'(\lambda)(M_+(\theta)-M_0(\theta)) - \mu_-'(\lambda)(M_0(\theta)-M_-(\theta)) > 0.\end{aligned}$$

On the other hand

$$\frac{\partial H}{\partial\theta} = \big(\mu_-(\lambda)\rho_{c_-}(\theta) + \mu_0(\lambda)\rho_{c_0}(\theta) + \mu_+(\lambda)\rho_{c_+}(\theta)\big)\,\tau(\theta), \tag{50}$$

so that $\Big\langle \frac{\partial H}{\partial\theta}, n(\theta)\Big\rangle = 0$ and $\Big\langle \frac{\partial H}{\partial\theta}, \tau(\theta)\Big\rangle > 0$ for $\lambda\in]-\epsilon',\epsilon'[$ with $\epsilon'>\epsilon$. Thus H is a local diffeomorphism on $]-\epsilon',\epsilon'[\times\mathbb{R}/2\pi\mathbb{Z}$. The map $\widetilde{H}\colon[-\epsilon,\epsilon]\times\mathbb{R}/2\pi\mathbb{Z}\to[\lambda_k-\epsilon\le f\le\lambda_k+\epsilon]$ defined by $\widetilde{H}(\lambda,\theta)=H(\lambda,\theta)$ is proper, therefore $\widetilde{H}$ is a covering map from $[-\epsilon,\epsilon]\times\mathbb{R}/2\pi\mathbb{Z}$ to $[\lambda_k-\epsilon\le f\le\lambda_k+\epsilon]$. Since $[\lambda_k-\epsilon\le f\le\lambda_k+\epsilon]$ is connected, $\widetilde{H}$ is onto. Besides, since $c_+(\theta)\in[f=\lambda_+\epsilon]$, (ϵ,θ) is the only antecedent of $c_+(\theta)$ by H, $\widetilde{H}$ is injective by connectedness of $[-\epsilon,\epsilon]\times\mathbb{R}/2\pi\mathbb{Z}$. $\widetilde{H}$ is therefore a C^∞ diffeomorphism from $[-\epsilon,\epsilon]\times\mathbb{R}/2\pi\mathbb{Z}$ into $[\lambda_k-\epsilon\le f\le\lambda_k+\epsilon]$.

We then define f_k to be f_{k-1} outside of $[\lambda_k-\epsilon\le f\le\lambda_k+\epsilon]$ and by $f_k(H(\lambda,\theta))=\lambda_k+\lambda$ in $[\lambda_k-\epsilon\le f\le\lambda_k+\epsilon]$. When $\lambda\in[\lambda_k-\epsilon,\lambda_k-\epsilon/2]$, Properties 3, 3′ and equation (47) yield

$$\begin{aligned} H(\lambda-\lambda_k,\theta) &= -\frac{\lambda-\lambda_k}{\epsilon}c_-(\theta)+\Big(1+\frac{\lambda-\lambda_k}{\epsilon}\Big)c_0(\theta) \\ &= \frac{\lambda_k-\lambda}{\lambda_k-(\lambda_k-\epsilon)}c_-(\theta)+\frac{\lambda-(\lambda-\epsilon)}{\lambda_k-(\lambda_k-\epsilon)}c_0(\theta) \\ &= c_\lambda(\theta). \end{aligned}$$

Thus $f_k = f = f_{k-1}$ in $[\lambda_k-\epsilon \le f \le \lambda_k-\epsilon/2]$ and for similar reasons $f_k = f_{k-1}$ in $[\lambda_k+\epsilon/2 \le f \le \lambda_k+\epsilon]$. The "gluing" of f_{k-1} and f_k is therefore C^∞ along $[f=\lambda_k-\epsilon]$ and $[f=\lambda_k+\epsilon]$. Hence, f_k satisfies (P3).

Let us compute ∇f_k in $[\lambda_k-\epsilon \le f \le \lambda_k+\epsilon]$. By definition of f_k, $1=\Big\langle \nabla f_k(H(\lambda,\theta)), \dfrac{\partial H}{\partial \lambda}\Big\rangle$. Besides $H(\lambda-\lambda_k,\theta)$ is a parametrization of the level line $[f_k=\lambda]$ by its normal (see (50)), hence $\nabla f_k(H(\lambda,\theta)) = \alpha n(\theta)$ with $\alpha>0$. Using both formulae, we finally get

$$\begin{aligned} &\nabla f_k(H(\lambda,\theta)) = \\ &\quad= \frac{1}{\mu_+'(\lambda)\langle c_+(\theta)-c_0(\theta), n(\theta)\rangle - \mu_-'(\lambda)\langle c_0(\theta)-c_-(\theta), n(\theta)\rangle}\, n(\theta). \end{aligned}$$

From the definition of μ_+, $\mu_+'(\lambda)-\mu_-'(\lambda)=1/\epsilon$. Besides, for $\lambda\in[-\epsilon,-\epsilon/2]$ we have

$$\frac{\epsilon}{\langle c_0(\theta)-c_-(\theta), n(\theta)\rangle} = \|\nabla f(c_{\lambda+\lambda_k}(\theta))\|,$$

while for $\lambda\in[\epsilon/2,\epsilon]$ we get

$$\frac{\epsilon}{\langle c_+(\theta)-c_0(\theta), n(\theta)\rangle} = \|\nabla f(c_{\lambda+\lambda_k}(\theta))\|.$$

Hence by (46):

$$\|\nabla f_k(H(\lambda,\theta))\| \le \|\nabla f(c_{\lambda_k+\epsilon}(\theta))\|.$$

(P4) is therefore satisfied.

The last assertion we need to establish is the convexity of f_k. By construction, it suffices to prove that the Hessian Q_{f_k} of f is nonnegative in $[\lambda_k-\epsilon \le f \le \lambda_k+\epsilon]$. Let us denote by Q_H the Hessian of H (observe that Q_H takes its values in $\mathbb{R}^2$). For $-\epsilon\le\lambda\le\epsilon$, we have $\lambda+\lambda_k = f_k(H(\lambda,\theta))$, thus

$$\begin{aligned} 0 = Q_{f_k}(H(\lambda,\theta))(DH(\lambda,\theta)(\cdot), DH(\lambda,\theta)(\cdot)) +& \\ + \langle \nabla f_k(H(\lambda,\theta)), Q_H(\lambda,\theta)(\cdot,\cdot)\rangle& \end{aligned}$$

where DH denotes the differential map of H. To prove that Q_{f_k} is nonnegative, it suffices to prove that $\langle \nabla f_k(H(\lambda,\theta)), Q_H(\lambda,\theta)(\cdot,\cdot)\rangle \le 0$. We have

$$\begin{aligned}\frac{\partial^2 H}{\partial \lambda^2} &= \mu_-''(\lambda)c_-(\theta) + \mu_0''(\lambda)c_0(\theta) + \mu_+''(\lambda)c_+(\theta)\\ &= \mu_-''(\lambda)(c_-(\theta) - c_0(\theta)) + \mu_+''(\lambda)(c_+(\theta) - c_0(\theta))\\ &= \mu_+''(\lambda)\big((c_+(\theta) - c_0(\theta)) - (c_0(\theta) - c_-(\theta))\big),\end{aligned}$$

where the last equality is due to item $2'$. On the other hand

$$\Big\langle \nabla f_k(H(\lambda,\theta)), \frac{\partial^2 H}{\partial \lambda^2}\Big\rangle =$$

$$= \mu_+''(\lambda)\|\nabla f_k(H(\lambda,\theta)\|\big(\langle c_+(\theta) - c_0(\theta), n(\theta)\rangle - \langle c_0(\theta) - c_-(\theta), n(\theta)\rangle\big)$$

which is nonpositive because of (46). Besides we have

$$\frac{\partial^2 H}{\partial \lambda \partial \theta} = \big(\mu_-'(\lambda)\rho_{c_-}(\lambda) + \mu_0'(\lambda)\rho_{c_0}(\lambda) + \mu_+'(\lambda)\rho_{c_+}(\lambda)\big)\tau(\theta),$$

thus $\Big\langle \nabla f_k(H(\lambda,\theta)), \dfrac{\partial^2 H}{\partial \lambda \partial \theta}\Big\rangle = 0$. Finally

$$\frac{\partial^2 H}{\partial \theta^2} = \big(\mu_-(\lambda)\rho_{c_-}(\theta) + \mu_0(\lambda)\rho_{c_0}(\theta) + \mu_+(\lambda)\rho_{c_+}(\theta)\big)(-n(\theta)) + \big(\cdots\big)\tau(\theta),$$

hence the quantity

$$\Big\langle \nabla f_k(H(\lambda,\theta)), \frac{\partial^2 H}{\partial \theta^2}\Big\rangle =$$

$$= -\big(\mu_-(\lambda)\rho_{c_-}(\theta) + \mu_0(\lambda)\rho_{c_0}(\theta) + \mu_+(\lambda)\rho_{c_+}(\theta)\big)\|\nabla f_k(H(\lambda,\theta))\|$$

is negative since all the μ and ρ are nonnegative. Hence Q_{f_k} is nonnegative and the function f_k is convex.

C^2 *smoothing.* For $\lambda \in (\min \tilde f, \lambda_0]$, define

$$h(\lambda) = (\lambda - \min \tilde f)(1 + \max_{[\lambda \le \tilde f \le \lambda_0]} \|Q_{\tilde f}\|)^{-1}.$$

Since $\tilde f$ is C^∞ in $[\min \tilde f < \tilde f]$, h is a continuous, positive, increasing function. Then there exists $\psi \in C^\infty(\mathbb{R}, \mathbb{R}_+)$ which vanishes on $(-\infty, \min \tilde f]$, increases on $(0, +\infty)$ and for $\lambda \in (\min \tilde f, \lambda_0]$, $0 < \psi(\lambda) \le h(\lambda)$ (see Lemma 45). Let g be the primitive of ψ with $g(\min \tilde f) = 0$. The function g is a strictly increasing convex C^∞-function on $[\min \tilde f, +\infty)$. The function $\bar f = g \circ \tilde f$ is therefore a C^1 convex function. Moreover $\bar f$ is C^∞ at each point outside the boundary of argmin f. For $x \in \operatorname{argmin} f$, we have

$$\frac{\nabla \bar f(x+h) - \nabla \bar f(x)}{\|h\|} = \frac{g'(\tilde f(x+h))\nabla \tilde f(x+h)}{\|h\|} =$$

$$= \frac{g'(\tilde f(x) + o(\|h\|))o(1)}{\|h\|} =$$

$$= \frac{o(\|h\|)}{\|h\|} = o(1).$$

Thus $Q_{\bar f}(x) = 0$. On the other hand

$$\begin{aligned}\|Q_{\bar f}(x+h)\| &\le g'(\tilde f(x+h))\|Q_{\tilde f}(x+h)\| + g''(\tilde f(x+h))\|\nabla \tilde f(x+h)\|^2\\ &\le h(\tilde f(x+h))\|Q_{\tilde f}(x+h)\| + o(1)\\ &\le (f(x+h) - f(x)) + o(1) = o(1).\end{aligned}$$

Thus $Q_{\bar f}$ is continuous at x and thus $\bar f$ is C^2.

Conclusion. Let us prove finally that $\bar f$ does not satisfy the KŁ–inequality. Towards a contradiction, let us assume that there exist $R > \inf \bar f = \min \bar f$, a continuous function $\varphi\colon [\min \bar f, R) \to \mathbb{R}_+$ which satisfies $\varphi(\min \bar f) = 0$, φ is C^1 on $(\min \bar f, R)$ with $\varphi' > 0$, such that we have

$$\|\nabla(\varphi \circ \bar f)(x)\| \ge 1, \quad \forall x \in [\min f < f < R].$$

Applying Theorem 20 [(i)⇔(vi)], we obtain

$$\operatorname{Dist}([\bar f \le g(\lambda_k)], [\bar f \le g(\lambda_{k+1})]) \le \varphi(g(\lambda_k)) - \varphi(g(\lambda_{k+1})).$$

and, as a consequence,

$$\sum_{k=0}^{+\infty} \operatorname{Dist}([\tilde f \le \lambda_k], [\tilde f \le \lambda_{k+1}]) = \sum_{k=0}^{+\infty} \operatorname{Dist}([\bar f \le g(\lambda_k)], [\bar f \le$$

$$\le g(\lambda_{k+1})]) \le \varphi(g(\lambda_0)).$$

This contradicts the fact that $\sum \operatorname{Dist}(T_k, T_{k+1}) = +\infty$. □

4.4. **Asymptotic equivalence for discrete and continuous dynamics.** In this part we assume that $f\colon H \to \mathbb{R}$ is a $C^{1,1}$ *convex* function, that is, continuously differentiable with gradient ∇f Lipschitz continuous. Let L be a Lipschitz constant of ∇f.

Fix $\beta > 0$ and $x \in \mathbb{R}^n$ and consider any sequence $\{Y_x^k\}$ satisfying

$$(51) \quad \begin{cases} \beta \,||\nabla f(Y_x^k)||\; ||Y_x^{k+1} - Y_x^k|| \;\leq\; f(Y_x^k) - f(Y_x^{k+1}), \quad k = 1, 2, \dots \\ \\ Y_x^0 = x \end{cases}$$

This condition has been considered in [1] for nonconvex functions defined in finite-dimensional spaces. It is easily seen that (51) is a descent sequence, that is, $f(Y_x^k) \geq f(Y_x^{k+1})$, which implies in particular that $\{f(Y_x^k)\}$ converges as k goes to infinity.

Condition (51) is fulfilled by several explicit gradient–like methods, including trust region methods, line–search gradient methods and some Riemannian variants; see [1] for examples and references.

The following theorem establishes connections between length boundedness properties of continuous gradient methods and length boundedness of discrete gradient iterations.

Theorem 38 (discrete vs continuous)**.** *Let f be a $C^{1,1}$ convex function with compact sublevel sets such that* $\min f = 0$*. Let us denote by L a Lipschitz constant of ∇f. Then the following statements are equivalent:*

(i) [**Kurdyka-Łojasiewicz inequality**] *There exist $r_0 > 0$ and $\varphi \in \mathcal{K}(0, r_0)$ such that*

$$(52) \qquad ||\nabla(\varphi \circ f)(x)|| \geq 1, \qquad \textit{for all } x \in [0 < f \leq r_0].$$

(ii) [**Length boundedness of piecewise gradient iterates**] *For all $\beta > 0$ and all $R > 0$, there exists $\mathcal{L}(\beta) > 0$ such that for any sequence of gradient iterates of the form*

$$Y_{x_0}^0, Y_{x_0}^1, \dots, Y_{x_0}^{k_0}, Y_{x_1}^0, \dots Y_{x_1}^{k_1}, \dots$$

with $f(x_0) < R$, $f(Y_{x_{i+1}}^0) = f(x_{i+1}) \leq f(Y_{x_i}^{k_i})$ and $\{Y_{x_i}^j : j = 0, \dots, k_i\}$ satisfying (51) for all $i \in \mathbb{N}$ we have

$$\sum_{i=0}^{+\infty} \sum_{l=0}^{k_i} ||Y_{x_i}^{l+1} - Y_{x_i}^l|| \;\leq\; \mathcal{L}(\beta).$$

(iii) [**Length boundedness of piecewise gradient curves**] *For every $R > 0$ there exists $\mathcal{L} > 0$ such that*

$$\text{length}\ (\gamma) \leq \mathcal{L},$$

for all piecewise subgradient curves $\gamma : [0, +\infty) \to H$ with $f(\gamma(0)) < R$.

Proof. Let us first prove that (i)⇒(ii). By Theorem 29[(i) ⇒(ii)] (subgradient inequality – convex case) we may assume that φ is concave, defined on $(0, +\infty)$ and (52) holds for all $x \in [0 < f]$. We now proceed in the spirit of [1]. Let $\beta > 0$, $x \in [0 < f]$ and let $Y_x^0, \ldots, Y_x^k$ be a (finite) sequence of gradient–type iterations that satisfies (51). For simplicity we set $Y_x^j = Y^j$ for all $j \in \{0, \ldots, k\}$, so that

$$f(Y^j) - f(Y^{j+1}) \;\geq\; \beta \,||\nabla f(Y^j)||\;||Y^{j+1} - Y^j||.$$

Multiplying both parts with $\varphi'(f(Y^j))$ and applying (i) we get

$$\varphi'(f(Y^j))[f(Y^j) - f(Y^{j+1})] \;\geq\; \beta \,||Y^{j+1} - Y^j||.$$

Since φ is concave we have

$$\varphi(f(Y^{j+1})) \leq \varphi(f(Y^j)) + \varphi'(f(Y^j))\,[f(Y^{j+1}) - f(Y^j)],$$

and therefore

$$\varphi(f(Y^j)) - \varphi(f(Y^{j+1})) \;\geq\; \beta \,||Y^{j+1} - Y^j||.$$

Adding the above inequalities for $j = 0, \ldots, k$ we obtain

$$\varphi(f(Y^0)) - \varphi(f(Y^k)) \;\geq\; \beta \sum_{j=0}^{k} ||Y^{j+1} - Y^j||. \tag{53}$$

Let us now consider a sequence of the form $\{Y_{x_0}^0, Y_{x_0}^1, \ldots, Y_{x_0}^{k_0}, Y_{x_1}^0, \ldots, Y_{x_1}^{k_1}, \ldots\}$ as in (ii). Then applying (53) to each subsequence $\{Y_{x_i}^j, j = 0, \ldots, k_i\}$ we deduce

$$\sum_{i=0}^{+\infty} \sum_{l=0}^{k_i} ||Y_{x_i}^{l+1} - Y_{x_i}^l|| \;<\; \frac{1}{\beta}\varphi(f(Y_{x_0}^0)) \;\leq\; \frac{1}{\beta}\varphi(R),$$

which proves the assertion.

The equivalence (i)$\Longleftrightarrow$(iii) follows from Theorem 18 and Theorem 29. To complete the proof it suffices to establish that (ii) implies the assertion (iv) of Theorem 18 (valley selection of finite length) (in fact we prove (iv') with $R = 2$). So let us assume that (ii) holds and let $r_0 > m$. We aim to construct a piecewise absolutely continuous curve $\theta : (0, r_0] \to \mathbb{R}^n$ of finite length that satisfies

$$\theta(r) \in \mathcal{V}_2(r) :=$$
$$:= \left\{ x \in [f = r] : \quad ||\nabla f(x)|| \;\leq\; 2 \inf_{y \in [f=r]} ||\nabla f(y)|| \right\}, \; \forall r \in (0, r_0].$$

We shall use the explicit gradient method described in Subsection 5.2. Let $x_0 \in \mathcal{V}_2(r_0)$ be such that

$$||\nabla f(x_0)|| \;\leq\; \frac{3}{2} \inf_{y \in f^{-1}(r_0)} ||\nabla f(y)||,$$

and consider the C^1 curve

$$\left[0, \frac{1}{3L}\right) \ni t \longmapsto x_0(t) := x_0 - t\nabla f(x_0).$$

Set $t_0 = \sup A_0$ where

$$A_0 := \left\{ t \in \left(0, \frac{1}{3L}\right) \;:\; \left[\begin{array}{ll} f \circ x_0 & \text{strictly decreasing on } [0,t], \\ x_0(\tau) \in \mathcal{V}_2(f(x_0(\tau))) & \text{for } \tau \in [0,t]. \end{array} \right. \right\}.$$

Clearly A_0 is nonempty and $0 < t_0 \le (3L)^{-1}$. Set $r_1 = f(x_0(t_0)) < r_0$ and take $x_1 \in \mathcal{V}_2(r_1)$ such that

$$||\nabla f(x_1)|| \le \frac{3}{2} \inf_{y \in [f = r_1]} ||\nabla f(y)||.$$

Proceeding by induction we obtain a sequence $\{(t_k, r_k, x_k)\}$ where $\{r_k\} \subset [0, r_0]$ is strictly decreasing, $x_n(t) := x_n - t\nabla f(x_n)$ with $f(x_n) = r_n$ and

$$||\nabla f(x_n)|| \le \frac{3}{2} \inf_{y \in [f = r_n]} ||\nabla f(y)||.$$

Let us denote by r_∞ the limit of $\{r_k\}$ and let us assume, towards a contradiction, that $r_\infty > 0$. Set

$$s(r) := \inf_{x \in f^{-1}(r)} ||\partial f(x)||_- \quad \text{and} \quad s_\infty = \liminf_{n\to\infty} s(r_n) \;=\; \lim_{n\to\infty} s(r_n)$$

(note that convexity of f guarantees that $s(r_1) \le s(r_2)$ whenever $r_1 \le r_2$) and observe that $r_\infty > 0$ implies that $s_\infty > 0$ (use the compactness of the sublevel set $[f \le r_0]$). Let $n_0 \in \mathbb{N}$ be such that $s(r_n) \le \frac{5}{4} s_\infty$ for all $n \ge n_0$. For $n \ge n_0$ and $t \in [0, t_n)$, Proposition 48 (Annex) yields

$$||\nabla f(x_n(t))|| \le (Lt+1)\, ||\nabla f(x_n)||,$$

which implies

$$||\nabla f(x_n(t))|| \le (Lt+1)\, ||\nabla f(x_n)|| \le \frac{3}{2}(Lt+1) s(r_n) \le \frac{15}{8}(Lt+1) s_\infty.$$

A sufficient condition to have $x_n(t) \in \mathcal{V}_2(f(x_n(t)))$ is therefore

$$\frac{15}{8}(Lt+1)s_\infty \le 2s_\infty \iff 0 \le t \le (15L)^{-1}. \tag{54}$$

Similarly we can estimate the rate of decrease of $f(x_n(t))$. Since

$$\frac{d}{dt} f(x_n(t)) = -\langle \nabla f(x_n), \nabla f(x_n(t)) \rangle,$$

the condition $\frac{d}{dt} f(x_n(t)) < 0$ is satisfied whenever

$$||\nabla f(x_n)||^2 > ||\nabla f(x_n)||\; ||\nabla f(x_n(t)) - \nabla f(x_n)||$$

But since ∇f is Lipschitz continuous, $\|\nabla f(x_n(t)) - \nabla f(x_n)\| \le Lt\|\nabla f(x_n)\|$. Thus the condition is satisfied if

$$\|\nabla f(x_n)\|^2 > Lt\|\nabla f(x_n)\|^2$$

This last inequality is equivalent to $t < L^{-1}$, which implies in particular that for all $n \in \mathbb{N}$ such that $s(r_n) \le \frac{5}{4} s_\infty$, we have

$$t_n \ge (15L)^{-1}.$$

In this case Proposition 48 (Annex) yields

$$f(x_n(t_n)) \le f(x_n) + \left(\frac{Lt_n^2}{2} - t_n\right) ||\nabla f(x_n)||^2 \le$$

$$\le r_n + \frac{9}{4}\left(\frac{Lt_n^2}{2} - t_n\right) s(r_n)^2 \le$$

$$\le r_n + \frac{9}{4}\left(\frac{Lt_n^2}{2} - t_n\right)\left(\frac{5}{4} s_\infty\right)^2.$$

Thus in order to have $f(x_n(t)) < r_\infty$, it suffices to require

$$t_n - \frac{Lt_n^2}{2} > \frac{64}{225}\left(\frac{r_n - r_\infty{}^2}{s_\infty}\right).$$

Using the fact that $(3L)^{-1} \ge t_n \ge (15L)^{-1}$, we see that

$$t_n - \frac{Lt_n^2}{2} \ge (15L)^{-1} - (18L)^{-1} = (90L)^{-1}.$$

Since $(r_\infty - r_n)/s_\infty$ tends to zero, we have that $f(x_n(t_n)) < r_\infty$ for n sufficiently large, which is a contradiction.

We thus conclude that $\{r_k\} \to r_\infty = 0$ and $(0, r_0] = \bigcup_n (r_{n+1}, r_n]$. We define $\theta\colon (0, r_0] \to H$ as follows: $\theta(r) := x_n([f \circ x_n]^{-1}(r))$ whenever $r \in (r_{n+1}, r_n]$. Clearly θ defines a piecewise absolutely continuous curve. To see that θ has finite length it suffices to observe that the sequence $\{x_n\}_n$ is a sequence of gradient iterates that satisfies (51). Using Remark 49 and the fact that the step–sizes in the construction of the x_n's do not exceed $(3L)^{-1}$ we infer that

$$\frac{5}{6}\, ||x_{n+1} - x_n||\, ||\nabla f(x_n)|| \le f(x_n) - f(x_{n+1}).$$

Hence the curve θ has a finite length. This completes the proof. □

Remark 39. The assumption that f is convex has been used to apply Theorem 29 (cf. concavity of φ which seems to be crucial for the proof of implication (i)⇒(ii)) and to assert that $f(Y_0^k) \to \inf f$. These are the reasons for which Theorem 38 is not stated for general semiconvex functions (in a local version). It would therefore be interesting to figure out under which type of conditions (other than convexity or o-minimality of f) the function φ of (52) can be taken concave.

5. Annex

In this Annex section we give several technical results which are needed in the text.

5.1. Technical results.

Proposition 40 (closed graph of the subdifferential)**.** *Let $f : H \to \mathbb{R} \cup \{+\infty\}$ be a lower semicontinuous semiconvex function. Let $\{x_k\}$ and $\{p_k\}$ be two sequences in H such that $p_k \in \partial f(x_k)$, x_k converges strongly to x and p_k converges weakly to p. Then as $k \to +\infty$ we obtain*

$$\begin{cases} f(x_k) \to f(x) \\ p \in \partial f(x) \end{cases}$$

Proof. This is a standard property. For a proof (in the more general setting of primer–lower–nice functions) we refer the reader to [34]. □

Proposition 41 (slope functions and semicontinuity)**.** *Let $f\colon H \to \mathbb{R}\cup \{+\infty\}$ be a lower semicontinuous semiconvex function.*

(i) The extended-real-valued function

(slope at x)
$$H \ni x \longmapsto ||\partial f(x)||_- := \inf_{p\in\partial f(x)} ||p||$$

is lower semicontinuous.

(ii) Take $r_0 \in \mathbb{R}$ and let D be a nonempty compact subset of $[f \leq r_0]$. Then the function

(minimal slope of the r level-line)
$$(-\infty, r_0] \ni r \longmapsto s_D(r) := \inf_{x\in[f=r]\cap D} ||\partial f(x)||_-$$

is lower semicontinuous.

(iii) Assume that (23) *and* (24) *hold for some $\bar{r}, \bar{\epsilon} > 0$. If $0 < r_1 \leq r_2 \leq \bar{r}$, then there exists $\eta_{r_1,r_2} > 0$ such that*

$$\inf_{x\in[r_1\leq f\leq r_2]\cap \bar{B}(\bar{r},\bar{\epsilon})} ||\partial f(x)||_- \geq \eta_{r_1,r_2} > 0.$$

Proof. (ii) Take $r \in (-\infty, r_0]$ and let $\{r_k\} \subset (-\infty, r_0]$ be a sequence such that $r_k \to r$ and $\liminf_k s_D(r_k) < +\infty$. Fix $\eta > 0$ and let $(x_k, p_k) \in \operatorname{graph}\partial f$ be such that $f(x_k) = r_k$, $p_k \in \partial f(x_k)$ and $||p_k|| < s_D(r_k) - \eta$. Using a standard compactness argument together with the fact that $\liminf_k s_D(r_k) < +\infty$ we can assume, with no loss of generality, that x_k converges (strongly) to $x \in D$ and that p_k converges weakly to p. Using

Proposition 40, we obtain that $(x,p) \in \operatorname{graph} \partial f$ and $f(x) = r$. The conclusion follows from the (weak) lower semicontinuity of the norm. Indeed

$$\liminf_{k\to+\infty} s_D(r_k) - \eta \geq \liminf_{k\to+\infty} ||p_k|| \geq ||p|| \geq s_D(r).$$

The proof of (i) and (iii) involve similar arguments. □

Lemma 42 (strong slope)**.** *Let f be a proper lower semicontinuous semiconvex function. Then for all x in H*

$$||\partial f(x)||_- = |\nabla f|(x).$$

Proof. Let $x \in H$ and $p = \partial^0 f(x)$ the projection of 0 on $\partial f(x)$. By (18), for any $y \in H$, we have

$$\frac{(f(x) - f(y))^+}{||y-x||} \leq \left(-\left\langle p, \frac{y-x}{||y-x||} \right\rangle + \alpha ||y-x||^2 \right)^+ \leq$$
$$\leq (||p|| + \alpha||y-x||^2)^+.$$

By taking the limsup as $y \to x$, we get $|\nabla f|(x) \leq ||p|| = ||\partial f(x)||_-$. To prove the opposite inequality, we consider the subgradient trajectory χ_x. If x is a critical point of f, then $0 = ||\partial f(x)||_- \geq |\nabla f|(x)$. Otherwise, $\chi_x(t) \neq x$ for all $t > 0$. By Theorem 13(iv), we have

$$\frac{(f(x) - f(\chi_x(t)))^+}{||x - \chi_x(t)||} \geq \frac{1}{||x - \chi_x(t)||} \int_0^t ||\partial f(\chi_x(\tau))||_-^2 \, d\tau.$$

Taking the limsup as $t \downarrow 0$ and using the continuity of the semiflow and Theorem 13(ii), (iii) we obtain the desired result. □

Lemma 43 (chain rules)**.** *Let $f \colon H \to \mathbb{R} \cup \{+\infty\}$ be a extended-real-valued function.*

(i) Let $\varphi : (0,1) \to \mathbb{R}$ be a C^1 function. Then

$$\partial(\varphi o f)(x) = \varphi'(f(x))\partial f(x), \text{ for all } x \in [0 < f < 1].$$

(ii) Let $\gamma : (0,1) \to H$ be a C^1 curve. For all $t \in (0,1)$, we have

$$\partial(f \circ \gamma)(t) \supset \{\langle \dot\gamma(t), p(t)\rangle : p(t) \in \partial f(\gamma(t))\}.$$

Proof. For the proof see [42] for example. □

Lemma 44 (continuous integrable majorant)**.** *Let $u \colon (0, r_0] \to \mathbb{R}_+$ be an upper semicontinuous function such that $u \in L^1(0, r_0)$. Then there exists a* continuous *function $w \colon (0, r_0] \to \mathbb{R}_+$ such that $w \geq u$ and $w \in L^1(0, r_0)$. If moreover u is assumed to be nonincreasing, w can be chosen to be decreasing.*

Proof. With no loss of generality we assume $r_0 = 1$. Replacing if necessary $u(\cdot)$ by the function $u(\cdot)+1$ we may also assume that $u \geq 1$. Let $a_k > 0$ be a strictly decreasing sequence such that $a_0 = 1$ and $(0,1] = \bigcup_{k\in\mathbb{N}}[a_{k+1}, a_k]$. Let us assume that there exists a sequence of continuous functions $w_k\colon [a_{k+1}, a_k] \to \mathbb{R}$ such that $w_k \geq u$ on $[a_{k+1}, a_k]$ and $\int_{a_{k+1}}^{a_k} w_k \leq \int_{a_{k+1}}^{a_k} u + \frac{1}{(k+1)^2}$. To establish the existence of w, we proceed by induction on k. Fix $k \geq 1$ and assume that w is defined on $[a_k, 1]$ with $w \geq u$, w continuous and

$$\int_{a_k}^1 w \leq \int_{a_k}^1 u + \sum_{i=1}^k \frac{2}{i^2}.$$

There is no loss of generality to assume $w_k(a_k) \leq w(a_k)$ (the case $w_k(a_k) > w(a_k)$ can be treated analogously). Let us define

$$0 < \epsilon_k = \frac{w_k(a_k)(a_k - a_{k+1})}{(k+1)^2\ w(a_k)\ \max_{[a_{k+1},a_k]} w_k} < a_k - a_{k+1},$$

and let us consider the functions

$$\lambda_k : [a_k - \epsilon_k, a_k] \to \left[1, \frac{w(a_k)}{w_k(a_k)}\right]$$

defined by

$$\lambda_k(r) = \frac{1}{\epsilon_k}\left((a_k - r) + (r - (a_k - \epsilon_k))\frac{w(a_k)}{w_k(a_k)}\right).$$

The function w can be now extended to $[a_{k+1}, 1]$ by setting

$$w(r) = \begin{cases} w_k(r), & \text{if } r \in [a_{k+1}, a_k - \epsilon_k), \\ \lambda_k(r) w_k(r), & \text{if } r \in [a_k - \epsilon_k, a_k] \\ w(r), & \text{if } r \in (a_k, 1]. \end{cases}$$

It is easily seen that the function w is continuous (by definition of λ_k), it satisfies $w \geq u$ on $[a_{k+1}, a_k]$ (thus on $(a_{k+1}, 1]$) and moreover

$$\begin{aligned} \int_{a_{k+1}}^1 w &= \int_{a_{k+1}}^{a_k - \epsilon_k} w_k + \int_{a_k - \epsilon_k}^{a_k} \lambda_k w_k + \int_{a_k}^1 w \\ &\leq \int_{a_{k+1}}^{a_k} u + \frac{1}{(k+1)^2} + \epsilon_k \frac{w(a_k)}{w_k(a_k)} \max_{[a_{k+1},a_k]} w_k + \int_{a_k}^1 u + \sum_{i=1}^k \frac{2}{i^2} \\ &\leq \int_{a_{k+1}}^1 u + \frac{2}{(k+1)^2} + \sum_{i=1}^k \frac{2}{i^2}. \end{aligned}$$

This proves the existence of a continuous function w that satisfies the required properties.

To complete the proof it suffices to prove the existence of such a sequence $\{w_k\}$. To this end, fix $k \in \mathbb{N}^*$ and set

$$u^\epsilon(r) = \sup_{\rho \in [a_{k+1}, a_k]} \left\{ u(\rho) - \frac{||r-\rho||^2}{2\epsilon} \right\}.$$

It is easily seen that u^ϵ is continuous, $u(r) \le u^\epsilon(r) \le \max_{\rho \in [a_{k+1}, a_k]} u := M_k < +\infty$ and $\lim_{\epsilon \to 0} u^\epsilon(r) = u(r)$ for all $r \in [a_{k+1}, a_k]$ (see [42], for example). Note that the upper semicontinuity of u on the compact set $[a_{k+1}, a_k]$ guarantees that M_k is finite. Applying the Lebesgue domination convergence theorem we conclude that u^ϵ converges to u in the norm topology of $L^1(a_{k+1}, a_k)$. Thus there exists $\epsilon_0 > 0$ such that

$$\int_{[a_{k+1}, a_k]} u^{\epsilon_0} \le \int_{[a_{k+1}, a_k]} u + \frac{1}{(k+1)^2}.$$

Thus the function $w_k := u^{\epsilon_0}$ satisfies the requirements stated above. This completes the proof of the first part of the statement. The case where u is assumed decreasing, can be treated with similar (and occasionally simpler) arguments. □

Lemma 45. *Let $h \in C^0((0, r_0], \mathbb{R}_+^*)$ be an increasing function, then there exists a function $\psi \in C^\infty(\mathbb{R}, \mathbb{R}_+)$ such that $\psi = 0$ on $\mathbb{R}_-$, $0 < \psi(s) \le h(s)$ for all $s \in (0, r_0)$, and ψ is increasing on $(0, r_0)$.*

Proof. Let us extend the definition of h by 0 on $\mathbb{R}_-$ and $h(r_0)$ for $s > r_0$. Consider $\phi \in C^\infty(\mathbb{R}, \mathbb{R}_+)$ with $[0,1]$ as support and $\int_{\mathbb{R}} \phi = 1$. Then we define ψ by $\psi = \phi * h$; *i.e.* $\psi(s) = \int_{\mathbb{R}} \phi(t) h(s-t) dt$. It is then straightforward to verify that ψ satisfies the expected properties. □

Proposition 46 (Piecewise absolutely continuous selections)**.** *Let $r_0 > 0$ and $\mathcal{V} : (0, r_0] \rightrightarrows H$ be a set-valued mapping with nonempty values. Assume that for each $r \in (0, r_0]$ there exists $\epsilon_r \in (0, r)$ and an absolutely continuous curve $\theta_r \colon (r - \epsilon_r, r] \to H$ such that*

$$\theta_r(s) \in \mathcal{V}(s) \text{ for all } s \text{ in } (r - \epsilon_r, r].$$

Then there exist a countable partition $\{I_n\}_{n \in \mathbb{N}}$ of $(0, r_0]$ into intervals I_n of nonempty interior and a selection $\theta \colon (0, r_0] \to \mathbb{R}^n$ of $\mathcal{V}$ such that θ is absolutely continuous on each I_n.

Proof. Let Ω be the set of couples $(\alpha : I_\alpha \subset (0, r_0] \to \mathbb{R}^n,\ \{I_{\alpha,j}\}_{j \in J_\alpha})$ where $\{I_{\alpha,j}\}_{j \in J_\alpha}$ is a countable partition of I_α into (disjoint) intervals $I_{\alpha,j},\ j \in J_\alpha$ with nonempty interior such that:

(a) for each $j \in J_\alpha$, α is absolutely continuous on $I_{\alpha,j}$,

(b) for each $r \in I_\alpha$, $\alpha(r) \in \mathcal{V}(r)$.

We define a partial order $\preccurlyeq$ on Ω by

$$\alpha_1 \preccurlyeq \alpha_2 \quad \Leftrightarrow \quad \forall j \in J_{\alpha_1},\ \exists k \in J_{\alpha_2},\ I_{\alpha_1,j} \subset I_{\alpha_2,k} \\ \text{and } \alpha_1(r) = \alpha_2(r) \text{ for all } r \in I_{\alpha_1}.$$

Note that $(\Omega, \preccurlyeq)$ is nonempty partially ordered. Let us check that each totally ordered subset of Ω has an upper bound in Ω. To this end, let

$$\omega = \{(\alpha_l, \{I_{\alpha_l,j}\}_{j\in J_{\alpha_l}})\}_{l\in\mathcal{L}}$$

be a totally ordered subset of Ω. For each $r \in \bigcup_{l\in\mathcal{L}} I_{\alpha_l}$ define $\alpha(r)$ by

$$\alpha(r) := \alpha_l(r),$$

whenever $r \in I_l$, and set $I_\alpha = \bigcup_{l\in L} I_{\alpha_l}$. Since ω is totally ordered, the mapping $\alpha\colon I_\alpha \to \mathbb{R}^n$ is well defined and (b) is clearly satisfied. For $l \in L$ and $j \in J_l$, set $J_l := J_{\alpha_l}$, $I_{\alpha_l,j} = I_{l,j}$ and $D := \{(m,k) : m \in L,\ k \in J_m\}$. For each $(l,j) \in D$, let us define

$$M_{l,j} := \bigcup_{(m,k)\in D,\ I_{l,j}\subset I_{m,k}} I_{m,k}. \tag{55}$$

Observe that $I_\alpha = \bigcup_{(l,j)\in D} M_{l,j}$ and that each $M_{l,j}$ is an interval with nonempty interior.

Let us prove that for all $(l,j),(l',j') \in D$, we have either $M_{l',j'} = M_{l,j}$ or $M_{l',j'} \cap M_{l,j} = \emptyset$. In order to establish this result, let us beforehand show that for all (l,j), (l',j') in D such that $I_{l,j} \cap I_{l',j'} \neq \emptyset$, we have $M_{l,j} = M_{l',j'}$. Indeed, since ω is totally ordered, we have for instance $I_{l',j'} \subset I_{l,j}$ and so $M_{l,j} \subset M_{l',j'}$. Conversely, take $(m,k) \in D$ such that $I_{m,k} \supset I_{l',j'}$. Since $I_{m,k} \cap I_{l,j} \neq \emptyset$, we have either $I_{m,k} \subset I_{l,j}$ or $I_{m,k} \supset I_{l,j}$, in any case we see (*cf.* definition (55)) that $I_{m,k} \subset M_{l,j}$ and thus $M_{l',j'} \subset M_{l,j}$.

If $M_{l,j} \cap M_{l',j'} \neq \emptyset$, take r in the intersection, and observe that by definition there exist (m,k) and (m',k') in D such that $I_{m,k} \supset I_{l,j}$ with $r \in I_{m,k}$ and $I_{m',k'} \supset I_{l',j'}$ with $r \in I_{m',k'}$. Using the previous remark, we obtain that $M_{m,k} = M_{l,j}$ and $M_{m',k'} = M_{l',j'}$. But since $I_{m,k} \cap I_{m',k'} \neq \emptyset$, we also have $M_{m,k} = M_{m',k'}$ and thus $M_{l,j} = M_{l',j'}$.

Let us define an equivalence relation $\simeq$ on D by

$$(l,j) \simeq (l',j') \Leftrightarrow M_{l,j} = M_{l',j'}.$$

This equivalence relation defines a partition of D into equivalence classes. By the axiom of choice we can pick one and only one element in each equivalence class and this defines a nonempty subset D' of D. By construction we have $I_\alpha = \bigcup_{(l,j)\in D'} M_{l,j}$ and $M_{l,j} \cap M_{l',j'} = \emptyset$ for each $(l,j) \neq (l',j')$ in D'. Besides since each $M_{l,j}$ (for $(l,j) \in D'$) has a nonempty interior, we see that D' is a countable set. This shows that $(\alpha, \{M_{l,j},\ (l,j) \in D'\})$ is in Ω with in addition $\alpha \geq \alpha_l$ for all $l \in \mathcal{L}$.

Applying Zorn's lemma to Ω, we obtain the existence of a maximal element $(\theta\colon I_\theta \to \mathbb{R}^n, \{I_{\theta,j},\ j \in J_\theta\})$. Arguing by contradiction, we see immediately that $I_\theta = (0, r_0]$. □

5.2. **Explicit gradient method.** We recall the following useful result

Lemma 47 (Descent lemma). *Let f be a $C^{1,1}$ function (that is, ∇f is L-Lipschitz continuous). Then*

$$f(y) \leq f(x) + \langle \nabla f(x), y - x\rangle + \frac{L}{2}||y - x||^2.$$

Proof. Set $x(t) = x + t(y - x)$ and notice that

$$\begin{aligned} f(y) - f(x) &= \int_0^1 \frac{d}{dt} f(x(t))dt = \\ &= \langle \nabla f(x), y - x\rangle + \int_0^1 \langle \nabla f(x(t)) - \nabla f(x), y - x\rangle dt. \end{aligned}$$

The assertion follows easily. □

Given $x \in H$, let us consider the following recursion rule

$$x^+ := X(t, x) = x - t\nabla f(x),\ t > 0. \tag{56}$$

Choosing a starting point x^0 in H, and $\lambda_k > 0$ a sequence of step size, the explicit gradient method writes

$$x^{k+1} = X(\lambda_k, x^k).$$

A part of the convergence analysis of this method (and some of its variants) is based on the following elementary results.

Proposition 48. *Let f be a $C^{1,1}$ function, $x \in H$, $t \in [0, 2L^{-1})$ and x^+ be given by* (56)*. Then*

(i) $(1 - \frac{Lt}{2})\, ||x^+ - x||\, ||\nabla f(x)|| \leq f(x) - f(x^+)$*;*

(ii) $||\nabla f(x^+)|| \leq (Lt + 1)\, ||\nabla f(x)||$.

Proof. Assertion (i) follows directly from Lemma 47 while assertion (ii) is a consequence of the fact that ∇f is Lipschitz continuous on $[x, x(t)]$ of constant L. □

Remark 49. Condition (51) of Section 4.4 corresponds of course to the inequality (i) above.

Acknowledgements

The first two authors acknowledge support of the ANR grant ANR-05-BLAN-0248-01 (France). The second author acknowledge support of the MEC grant MTM2005-08572-C03-03 (Spain). During the preparation of this work, several research visits of the co-authors have been realized, respectively to the CRM (Mathematical Research Center in Barcelona), the University Autonomous of Barcelona, the University of Paris 6 and the University of Tours. In each case the concerned author wishes to acknowledge their hosts for hospitality.

References

[1] Absil, P.-A., Mahony, R. & Andrews, B., Convergence of the iterates of descent methods for analytic cost functions. *SIAM J. Optim.* **16** (2005), 531–547.

[2] Albano, P., Cannarsa, P., Singularities of semiconcave functions in Banach spaces, *Stochastic analysis, control, optimization and applications*, Systems Control Found. Appl., 171–190 (Birkhäuser Boston, 1999).

[3] Ambrosio, L., Gigli, N., Savaré, G. *Gradient flows in metric spaces and in the space of probability measures*, Lectures in Mathematics ETH Zurich. Birkhauser Verlag, Basel, 2005.

[4] Aussel, D., Daniilidis, A. & Thibault, L., Subsmooth sets: functional characterizations and related concepts, *Trans. Amer. Math. Soc.* **357** (2005), 1275–1301.

[5] Attouch, H., Bolte, J. On the convergence of the proximal algorithm for nonsmooth functions involving analytic features, to appear in Math. Prog.

[6] Aze, D. & Corvellec, J.-N., Characterizations of error bounds for lower semicontinuous functions on metric spaces, *ESAIM Control Optim. Calc. Var.* **10** (2004), 409–425.

[7] Baillon, J.-B., Un exemple concernant le comportement asymptotique de la solution du problème $du/dt + \partial\varphi(u) \ni 0$, *J. Funct. Anal.* **28** (1978), 369–376.

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

[9] Bolte, J., Daniilidis, A., Lewis, A. & Shiota, M., Clarke subgradients of stratifiable functions, *SIAM J. Optimization* **18** (2007), 556–572

[10] Brézis, H., Monotonicity methods in Hilbert spaces and some applications to nonlinear partial differential equations, *Contributions to nonlinear functional analysis (Proc. Sympos., Math. Res. Center, Univ. Wisconsin, Madison, Wis., 1971),* (Academic Press, New York, 1971), 101–156.

[11] Brézis, H., Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert (French), *North-Holland Mathematics Studies* **5**, (North-Holland Publishing Co., 1973).

[12] Bruck, Jr., R. E., Asymptotic convergence of nonlinear contraction semigroups in Hilbert space, *J. Funct. Anal.* **18** (1975), 15–26.

[13] Clarke, F. H., Ledyaev, Yu., Stern, R. I., Wolenski, P. R., *Nonsmooth Analysis and Control Theory*, Graduate texts in Mathematics **178**, (Springer-Verlag, New-York, 1998).

[14] COMBETTES, P. & PENNANEN, T., Proximal methods for cohypomonotone operators, *SIAM J. Control Optim.* **43** (2004), 731–742.

[15] COSTE, M., *An Introduction to o-minimal Geometry*, RAAG Notes, 81 pages, Institut de Recherche Mathématiques de Rennes, November 1999.

[16] D'ACUNTO, D., On Talweg lines of polynomial and analytic functions, Working paper.

[17] DEGIOVANNI, M., MARINO, A., TOSQUES, M., Evolution equations with lack of convexity, *Nonlinear Analysis* **9** (1985), 1401-1443.

[18] DE GIORGI, E., MARINO, A. & TOSQUES, M., Problems of evolution in metric spaces and maximal decreasing curve, *Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Natur.* **68** (1980), 180–187.

[19] DONTCHEV, A. L., LEWIS, A. S., ROCKAFELLAR, R. T., The radius of metric regularity, *Trans. Amer. Math. Soc.* **335** (2002), 493–517.

[20] DONTCHEV, A. L., QUINCAMPOIX, M., ZLATEVA, N., Aubin criterion for metric regularity, *J. Convex Anal.* **13** (2006), 281–297.

[21] VAN DEN DRIES, L. & MILLER, C., Geometric categories and o-minimal structures, *Duke Math. J.* **84** (1996), 497-540.

[22] EVANS, L. C. & SPRUCK, J., Motion of level sets by mean curvature. III, *J. Geom. Anal.* **2** (1992) 121–150.

[23] FENCHEL, W., *Convex Cones, Sets and Functions*, Mimeographed lecture note Princeton University, 1951.

[24] FORTI, M., NISTRI, P., QUINCAMPOIX, M., Convergence of Neural Networks for Programming Problems via a Nonsmooth Łojasiewicz Inequality, *IEEE Trans. on Neural Networks*, **17** (2006), 1471–1486.

[25] GAGE, M. & HAMILTON, R. S., The heat equation shrinking convex plane curves, *J. Differential Geom.* **23** (1986) 69–96.

[26] HARAUX, A., A hyperbolic variant of Simon's convergence theorem. Evolution equations and their applications in physical and life sciences (Bad Herrenalb, 1998), *Lecture Notes in Pure and Appl. Math.* **215** (2001), 255-264 (Dekker, New York).

[27] HUANG, S.-Z. *Gradient inequalities. With applications to asymptotic behavior and stability of gradient-like systems,* Mathematical Surveys and Monographs, 126, American Mathematical Society, Providence, RI, 2006.

[28] IOFFE, A., Metric regularity and Subdifferential Calculus, *Russian Math. Surveys* **55** (2000), 501–558

[29] KURDYKA, K., On gradients of functions definable in o-minimal structures, *Ann. Inst. Fourier* **48** (1998), 769-783.

[30] LAGEMAN, C., Convergence of gradient-like dynamical systems and optimization algorithms, *PhD Thesis, University of Würzburg,* (2007), 205 pp.

[31] LEE, J. M., *Introduction to smooth manifolds*, Graduate Texts in Mathematics, 218. Springer-Verlag, New York, 2003. xviii+628 pp.

[32] LEMAIRE, B., An Asymptotical Variational Principle Associated with the Steepest Descent Method for a Convex Function, *J. Convex Anal.* **3** (1996), 63–70.

[33] ŁOJASIEWICZ, S., "Une propriété topologique des sous-ensembles analytiques réels.", in: *Les Équations aux Dérivées Partielles*, pp. 87–89, Éditions du centre National de la Recherche Scientifique, Paris 1963.

[34] MARCELLIN, S. & THIBAULT, L., Evolution problems associated with primal lower nice functions, *J. Convex Anal.* **13** (2006), 385–421.

[35] Mordukhovich, B. Complete characterization of openness, metric regularity and Lipschitzian properties of multifunctions. *Trans. Amer. Math. Soc.* **340** (1993), 1–35.
[36] Mordukhovich, B. *Variational analysis and generalized differentiation. I. Basic theory*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], 330. Springer-Verlag, Berlin, 2006. xxii+579 pp.
[37] Nesterov, Y., Polyak, B. T., Cubic regularization of Newton method and its global performance. *Math. Program. 108*, (2006), no. 1, Ser. A, 177–205.
[38] Nistri, P. Quincampoix, M. On the properties of solutions to a differential inclusion associated with a nonsmooth constrained optimization problem, *Proceedings of the 44th IEEE Conference on Decision and Control and the European Control Conference 2005*, Seville, Spain, December 12-15 2005.
[39] Penot, J.-P., Metric regularity, openness and Lipschitzian behaviour of multifunctions. *Nonlinear Analysis,* **13** (1989), 629–643.
[40] Simon, L., Asymptotics for a class of non-linear evolution equations, with applications to geometric problems, *Ann. Math.* **118** (1983), 525-571.
[41] Torralba, D., *Convergence épigraphique et changements d'échelle en analyse variationnelle et optimisation*, 160pp., PhD Thesis, (Université de Montpellier 2, 1996).
[42] Rockafellar, R.T. & Wets, R., *Variational Analysis*, Grundlehren der Mathematischen, Wissenschaften, Vol. **317** , (Springer, 1998).
[43] Zhu, X. *Lectures on mean curvature flows*, AMS/IP Studies in Advanced Mathematics **32**, American Mathematical Society, 2002.

Jérôme Bolte
UPMC Univ Paris 06-Equipe Combinatoire et Optimisation (UMR 7090), Case 189
Université Pierre et Marie Curie
4 Place Jussieu, F–75252 Paris Cedex 05.
INRIA Saclay, CMAP, Ecole Polytechnique, 91128 Palaiseau, France
E-mail address: bolte@math.jussieu.fr
http://www.ecp6.jussieu.fr/pageperso/bolte

Aris Daniilidis
Departament de Matemàtiques, C1/308, Universitat Autònoma de Barcelona
E–08193 Bellaterra (Cerdanyola del Vallès), Spain.
Laboratoire de Mathématiques et Physique Théorique
Université François Rabelais, Tours, France
E-mail address: arisd@mat.uab.es
http://mat.uab.es/~arisd

Olivier Ley
Laboratoire de Mathématiques et Physique Théorique (CNRS UMR 6083)
Fédération Denis Poisson
Faculté des Sciences et Techniques, Université François Rabelais
Parc de Grandmont, F–37200 Tours, France
E-mail address: ley@lmpt.univ-tours.fr
http://www.phys.univ-tours.fr/~ley

Laurent Mazet
Université Paris-Est,
Laboratoire d'Analyse et Mathématiques Appliquées, UMR 8050
UFR des Sciences et Technologie, Département de Mathématiques
61 avenue du Général de Gaulle 94010 Créteil cedex, France.
E-mail address: laurent.mazet@univ-paris12.fr
http://perso-math.univ-mlv.fr/users/mazet.laurent/