Spring School on Variational Analysis
Paseky nad Jizerou, Czech Republic (April 20–24, 2009)

GRADIENT DYNAMICAL SYSTEMS, TAME OPTIMIZATION AND APPLICATIONS

Aris DANIILIDIS

Departament de Matemàtiques
Universitat Autònoma de Barcelona
E–08193, Bellaterra (Cerdanyola del Vallès)

http://mat.uab.cat/~arisd/

Lecture Notes

Abstract. These lectures present an introduction to what is nowadays called *Tame Optimization*, with emphasis to (nonsmooth) Łojasiewicz gradient inequalities and Sard–type theorems. The former topic will be introduced via the asymptotic analysis of dynamical systems of (sub)gradient type; its consequences in the algorithmic analysis (proximal algorithm, gradient-type methods) will also be discussed. The latter topic will be presented as a natural consequence of the structural assumptions made on the function (o-minimality, stratification). Our secondary aim is to provide essential background and material for further research. During the lectures, some open problems will be eventually mentioned.

Contents

1 Trajectories of (sub)gradient systems

We consider the autonomous dynamical system (differential equation)

$$\dot{x}(t) = F(x(t)). \tag{1.1}$$

generated by a locally Lipschitz continuous function $F : \mathbb{R}^n \to \mathbb{R}^n$. We call *solution curve* (also *trajectory* or *orbit*) of the vector field F any C^1 curve $t \longmapsto \gamma(t) \in \mathbb{R}^n$ satisfying (1.1). Existence and uniqueness (if we fix the initial condition $\gamma(0) = x_0 \in \mathbb{R}^n$) of solutions is a classical result in the theory of **O**rdinary **D**ifferential **E**quations (see [41, Section 2.2] *e.g.*).

Unless otherwise stated, we consider *maximal* solutions, meaning that the trajectory $t \mapsto \gamma_{x_0}(t)$ starting at the point x_0 is defined for all t in the maximal interval $[0, T_{x_0})$ ($T_{x_0} \in (0, +\infty]$) for which (1.1) makes sense. Maximal intervals are always right–open and that if $T_{x_0} < +\infty$ then $\{\gamma_{x_0}(t)\}_{t \geq 0}$ is unbounded (see also [41, Section 2.4]). Note that if F is (globally) Lipschitz continuous, then every orbit satisfies $T_{x_0} = +\infty$. The *length* of a C^1 orbit is given by the formula:

$$\text{length}(\gamma) = \int_0^{T_{x_0}} \|\dot{\gamma}(t)\| dt \tag{1.2}$$

We also use the term *integral curve,* especially if we are interested to the image in $\mathbb{R}^n$ of a trajectory rather than to the trajectory itself as a function. The terminology dynamical system suggests an evolution of each point of $\mathbb{R}^n$ by the *flow* generated by F, *i.e.* a function $\Phi(t, x_0)$ which associates to each $x_0 \in \mathbb{R}^n$ and $t_0 \in [0, T_{x_0})$ a point $\gamma_{x_0}(t_0)$, where γ_{x_0} is the unique orbit starting at x_0, that is, $\Phi(0, x_0) = \gamma_{x_0}(0) = x_0$. (Note however that if we relax the assumption on F from local Lipschitz continuity to mere continuity, uniqueness will no longer hold: think of the example $F(x_1, x_2) = (0, \sqrt{|x_2|})$ with initial condition any point of the x_1-axis.)

The flow is often represented by its *phase portrait*, that is, the picture of its integral curves in $\mathbb{R}^n$. Reparametrizing trajectories (*e.g.* length–parametrization), does not change the portrait of the system. More generally, the systems

$$\dot{x}(t) = F(x(t)) \qquad \text{and} \qquad \dot{x}(t) = g(x(t))\, F(x(t)) \tag{1.3}$$

have the same portrait, provided $g : \mathbb{R}^n \to \mathbb{R}$ is a positive smooth function. Intuitively, g corresponds to a change of velocity that the orbits are run through.

Our departure point in these lectures is a particular type of autonomous dynamical system, defined by a (sub)gradient field. To define this system properly, let us first consider a $C^{1,1}$-function $f : \mathbb{R}^n \to \mathbb{R}$ (that is, a differentiable function whose gradient is Lipschitz continuous) and set $F = -\nabla f$. Then (1.1) yields

$$\dot{x}(t) = -\nabla f(x(t)). \tag{1.4}$$

Remark 1 (Reversing the flow)**.** (i). The minus sign in (1.4) is conventional (relating to a minimization rather than a maximization problem). In fact, the system

$$\dot{x}(t) = \nabla f(x(t)). \tag{1.5}$$

is *reversing the flow*, that is, it has the same integral curves as (1.4) with a different orientation.
(ii). Although the differentiability assumption on f seems necessary for the uniqueness of solutions, we shall eventually also consider nonsmooth functions (*convex* or *semiconvex*) for which the so-called subgradient integral curves can be defined in a unique way. These curves are absolutely continuous curves that are solutions of the differential inclusion

$$\dot{x}(t) \in -\partial f(x(t)) \quad \text{(a.e.)} \tag{1.6}$$

where $\partial f(x)$ denotes the set of subgradients (subdifferential) of f at x and where the notation "a.e." stands for "almost everywhere" in the sense of the Lebesgue measure of $\mathbb{R}$. All important features of the asymptotical study of gradient systems remain true in this nonsmooth case. Let us point out though, an important difference stemming from the above remark: Systems (1.6) are of unilateral nature ($-\partial f(x) \neq \partial(-f)(x)$), which means that a subgradient trajectory cannot be reversed on time. (The terminology *semiflow* is thus employed in this case.)

1.1 Elementary properties of gradient systems

A *Lyapunov* function of a dynamical system is any scalar function on $\mathbb{R}^n$ that is strictly decreasing along its integral curves. Although Lyapunov functions might not always exist for the general system (1.1), they do exist for gradient systems. Indeed, given any orbit $\gamma(t)$ (solution of (1.4)) one easily sees that the derivative of the function $t \longmapsto \rho(t) := f(\gamma(t))$ satisfies

$$\rho'(t) = -||\nabla f(\gamma(t))||^2 = -||\dot{\gamma}(t)||^2 \leq 0. \tag{1.7}$$

We shall now use an important consequence of Remark 1(i) combined with the uniqueness of the flow of (1.5): If γ_{x_0} denotes the (unique, maximal) trajectory of (1.4) starting at x_0, it holds

$$\nabla f(x_0) \neq 0 \Longrightarrow \nabla f(\gamma_{x_0}(t)) \neq 0 \quad \text{for all } t \in [0, T_{x_0}). \tag{1.8}$$

In the sequel we denote by S the set of *critical* (or *singular*) points of f, that is, $x_0 \in S$ if and only if $\nabla f(x_0) = 0$. Thus, according to (1.8), unless x_0 is already a singularity, the corresponding trajectory will never pass through S. This shows that the derivative in (1.7) is strictly negative on $[0, T_{x_0})$, thus the function f is a *Lyapunov* function for the system (1.4).

(Level-set parametrization) Assume $x(t)$ is an orbit of (1.4), and set $x(0) = x_0$, $r_0 = f(x_0)$ and $r_\infty = \lim_{t\to T_{x_0}} f(x(t))$. Using (1.8) we easily see that whenever $\nabla f(x_0) \neq 0$ (that is, whenever the orbit $x(t)$ is not reduced to a singleton) the mapping $\rho(t) = f(x(t))$ is a diffeomorphism between $[0, T_{x_0})$ and $(r_\infty, r_0]$, and that the curve $u(r) := x(\rho^{-1}(r))$ satisfies the differential equation

$$\dot{u}(r) = \frac{\nabla f(u(r))}{||\nabla f(u(r))||^2}. \tag{1.9}$$

The existence of a Lyapunov function guarantees that a gradient system does not have periodic (closed) orbits nor limit cycles. Moreover, the ω-limit $\Omega(\gamma)$ of each bounded[1] orbit γ (*i.e.* the set of all limits of sequences $\{\gamma(t_n)\}_n$ with $t_n \nearrow \infty$) consists of singularities, and is either singleton or infinite ([40, page 14] e.g.). Moreover, in view of (1.7), f is constant on $\Omega(\gamma)$ and $\mathrm{dist}(\gamma(t), \Omega(\gamma)) \leq \mathrm{dist}(\gamma(t), S) \to 0$, as t goes to infinity. In this sense, gradient systems are much simpler than general dynamical systems (1.1), and the study of their behavior around singularities results in the study of the asymptotic behavior of their orbits.

1.2 Asymptotic analysis: convergence, length, Palis & De Melo example

We shall now focus on the study of the asymptotic behavior of the orbits of the systems (1.4) and (1.6). Let us introduce some notation. For every $\lambda \in \mathbb{R}$ we set

$$[f \leq \lambda] := \{x \in \mathbb{R}^n : f(x) \leq \lambda\}.$$

The notations $[f < \lambda]$ (or $[\lambda_1 < f \leq \lambda_2]$ and so on) are defined analogously. From now on we assume:

- f is *inf-compact*, that is, $[f \leq \lambda]$ is a compact subset of $\mathbb{R}^n$ for all $\lambda \in \mathbb{R}$.

The above assumption implies in particular that f attains a minimum, which will be assumed to be zero (if this is not the case, we replace f by $\tilde{f} = f - \min f$ and observe that f and $\tilde{f}$ have the same phase portrait). Moreover every trajectory γ_x lies in a compact set, whence $T_x = +\infty$ and its ω-limit set is nonempty. In view of (1.7), f is constant on the ω-limit set of each of its orbits. The following example ([40, page 14]) shows that an ω-limit set can be infinite.

Example 2 (Palis, De Melo)**.** Let $f : \mathbb{R}^2 \to \mathbb{R}$ be defined (in polar coordinates) by

$$f(r\cos\theta, r\sin\theta) = \begin{cases} \exp(\frac{1}{r^2-1}) & \text{if } r < 1 \\ 0 & \text{if } r = 1 \\ \exp(\frac{1}{1-r^2})\sin(\frac{1}{r-1} - \theta) & \text{if } r > 1 \end{cases}$$

Then f is C^∞ and there exists an orbit whose ω-limit set is the unit sphere S^1.

[1]The ω-limit set may be empty for unbounded orbits (think of the example $f(x) = x^3$ and $x_0 < 0$).

The function f of the above example presents many oscillations. It is interesting to visualize its level sets (or its graph in $\mathbb{R}^3$). A similar example of such *Mexican–hat* type function was given in [1, §2]. In both cases, there exist nontrivial trajectories of the corresponding gradient system (1.4) with infinite length.

It is straightforward to see that if an orbit γ_{x_0} of (1.4) has finite length, then it converges to its ω-limit (a singleton in this case). We denote

$$\gamma_\infty := \lim_{t\to\infty} \gamma_{x_0}(t). \tag{1.10}$$

Let us now give an example (see [24, Section 7]) where although a gradient orbit has infinite length, the above limit (1.10) exists.

Example 3 (Convergent gradient orbit of infinite length)**.** Define a function $f : \mathbb{R}^2 \to \mathbb{R}$ in polar coordinates as

$$f(r,\theta) = \begin{cases} e^{-1/r}(1 + r + \sin(\frac{1}{r} + \theta)) & r \neq 0 \\ 0 & r = 0 \end{cases}$$

The graph of f in the plane $\theta = 0$ looks like the graph of Figure 1. Then f is smooth, positive away from

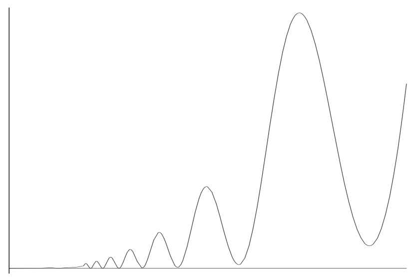


Figure 1:

the origin, with no critical point except at the origin (global minimum of f). The gradient trajectory of f issued from the point $(r,\theta) = ((\frac{3\pi}{2})^{-1}, 0)$ remains close to the spiral given by

$$\begin{cases} r &= \left(\frac{3\pi}{2} + t\right)^{-1} \\ \theta &= -t \end{cases}$$

and thus has infinite length.

We finally mention the following classical result due to Łojasiewicz:

- **(real-analytic case)** Bounded gradient trajectories of real-analytic functions $f : \mathbb{R}^n \to \mathbb{R}$ have finite length.

This is a consequence of the classical (Łojasiewicz) gradient inequality (we give more details in Section 2). In particular, each bounded trajectory γ of an analytic gradient system is converging to its ω-limit γ_∞. Moreover, the so-called *Thom conjecture* [45] for the gradient orbits of real-analytic functions holds true: the *secants* converge towards a fixed direction of the unit sphere (see K. Kurdyka, T. Mostowski and A. Parusinski [33] for the proof)

$$\frac{\gamma(t) - \gamma_\infty}{||\gamma(t) - \gamma_\infty||} \to d_\infty \in S^{n-1}. \tag{1.11}$$

1.3 Convex case: Brezis theorem, Baillon example

The case when the function f is convex and attains its minimum is particularly interesting in view of the important role of convex functions in optimization. Note that differentiability assumptions are not needed for the study of the corresponding integral curves. Indeed, under the assumption that f is convex and merely lower semicontinuous, we consider the (differential inclusion) subgradient system (1.6), where $\partial f : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is the classical convex subdifferential, defined for every $x \in \operatorname{dom} f$ as the set $\partial f(x)$ of all $\zeta \in \mathbb{R}^n$ such that for every $y \in \mathbb{R}^n$

$$f(y) \geq f(x) + \langle \zeta, y - x \rangle \tag{1.12}$$

Recall that for convex functions every critical point $x \in S$ (in the nonsmooth case this means $0 \in \partial f(x)$) is a global minimizer of the function.

Solution curves (*subgradient trajectories*) are then absolutely continuous curves that satisfy the differential inclusion (1.6) *almost everywhere.* Existence and uniqueness follow from Brezis theorem (see [15, Theorem 3.2, p. 57] or [2, Chapter 3.4]). Given a trajectory $\gamma : [0, T) \to \mathbb{R}^n$ of (1.6), for almost all $t \in (0, T)$ we have

$$\frac{d}{dt}(f \circ \gamma)(t) = \langle \dot{\gamma}(t), \zeta \rangle, \quad \text{for all } \zeta \in \partial f(\gamma(t)),$$

and the function $\zeta \mapsto \langle \dot{\gamma}(t), \zeta \rangle$ is constant on $\partial f(\gamma(t))$. Furthermore, if

$$S \equiv \operatorname{argmin} f \ = \ [f = 0] \tag{1.13}$$

where $\operatorname{argmin} f$ stands for the set of global minimizers of f (there is no loss of generality in assuming this), we have the following consequence of (1.6) and (1.12): for *every* $a \in C_0$

$$\frac{1}{2}\frac{d}{dt}||\gamma(t) - a||^2 \leq -f(\gamma(t)) \leq 0 \ \text{ a.e on } (0, +\infty),$$

and therefore the distance mapping $t \mapsto ||\gamma(t) - a||$ is nonincreasing. Since $\operatorname{dist}(\gamma(t), S) \to 0$ as t goes to infinity, we deduce the following result.

- Every subgradient orbit of a (nonnegative) convex function f converges to a global minimizer of f.

Note this result also holds in an infinite dimensional Hilbert space [17, Theorem 4], however the convergence should be taken in the weak topology, unless f is *even* [17, Theorem 5] (see also [39] for a slightly more general statement). Indeed, Baillon [4] shows that for any $\lambda \geq 1$ the function $f_\lambda : \mathbb{R}^2 \to \mathbb{R} \cup \{+\infty\}$

$$f_\lambda(x, y) = \begin{cases} [\arctan(x/y)]^\lambda \sqrt{x^2 + y^2}, & \text{if } x, y \geq 0 \\ +\infty, & \text{elsewhere} \end{cases}$$

is lower semicontinuous and convex and uses it to construct a lower semicontinuous function $\varphi : \ell^2(\mathbb{N}) \to \mathbb{R} \cup \{+\infty\}$ with minimum at 0 and with a bounded gradient trajectory which remains bounded away from 0. In particular, this trajectory does not converge for the norm topology and has infinite length.

A natural question though, is whether or not in finite dimensions subgradient orbits of convex functions have finite length. The rigid structure of convex functions makes natural to think that such orbits should be of finite length. It is rather surprising that *the answer of this question is not yet known* except in some particular cases.

Before we proceed, let us mention the particular case where the set of minimizers in (1.13) has nonempty interior (see [14]).

Theorem 4 (int (argmin f) $\neq \emptyset$)**.** *Let $f : H \to [0, +\infty]$ be a lower semi-continuous convex function such that the set of critical points (in this case, global minimizers) $S = \operatorname{argmin}$ has nonempty interior. Then subgradient orbits have finite length.*
More precisely, assuming $B(0, \delta) \subset S$ for some $\delta > 0$, we obtain the estimation

$$\int_0^T ||\dot{\gamma}(t))||dt \ \leq \ \sqrt{1 + (||\gamma(0)||/\delta)^2}\,(||\gamma(0)|| - ||\gamma(T)||), \quad \textit{for all } T \geq 0.$$

The following result ([13, Section 4.1]) is an extension of Theorem 4 under the assumption that the vector subspace $\operatorname{span}(S)$ generated by $S = \operatorname{argmin} f$, has codimension one in H. We denote by $\operatorname{ri}(S)$ the relative interior of S in $\operatorname{span}(S)$.

Theorem 5. *Let* $f : H \to \mathbb{R} \cup \{+\infty\}$ *be a lower semicontinuous convex function satisfying* $\min f = f(0) = 0$*. For* $S = \operatorname{argmin} f$*, assume that the subspace* $\operatorname{span}(S)$ *has codimension 1 and that the relative interior* $\operatorname{ri}(S)$ *of* C *with respect to* $\operatorname{span}(S)$ *is not empty. If* $x_0 \in \operatorname{dom} f$ *is such that* $\gamma_{x_0}(t)$ *converges (for the norm topology) to* $a \in \operatorname{ri}(S)$ *as* $t \to +\infty$*, then* $\operatorname{length}(\gamma_{x_0}) < +\infty$*.*

In the next section we tackle this problem in the plane, in a more general setting that also encompasses quasiconvex systems.

1.4 Self-contracted curves. Quasiconvex planar systems

We recall that the length of a continuous curve $\gamma : I \to \mathbb{R}^n$ is defined as

$$\operatorname{length}(\gamma) := \sup \left\{ \sum_{i=1}^{k} \operatorname{dist}(\gamma(t_i), \gamma(t_{i+1})) \right\}$$

where the supremum is taken over all the finite subdivisions $\{t_i\}_{i=1}^{k+1}$ of I. (Note that the above definition is equivalent to (1.2) in case γ is C^1.) The key notion in this section is the notion of *self-contracted* curve [24, Definition 1.2], which allows to provide a unified framework for the study of convex and *quasiconvex* gradient systems. Let us define this notion.

Definition 6 (Self-contracted curve)**.** A curve $\gamma : I \to \mathbb{R}^n$ defined in an interval I of $[0, +\infty)$ is called self-contracted, if for every $t_1 \le t_2 \le t_3$, with $t_i \in I$, we have

$$\operatorname{dist}(\gamma(t_1), \gamma(t_3)) \ge \operatorname{dist}(\gamma(t_2), \gamma(t_3)). \tag{1.14}$$

In other words, for every $[a, b] \subset I$, the map

$$t \in [a, b] \mapsto \operatorname{dist}(\gamma(t), \gamma(b))$$

is nonincreasing.

Inequality (1.14) shows that the image of a segment (a, b) by a self-contracted curve γ lies in a ball of radius $\rho := \operatorname{dist}(\gamma(a), \gamma(b))$. Note however that a self-contracted curve might not be (left/right) continuous. A simple example is provided by the following planar self-contracted curve:

$$\gamma(t) = \begin{cases} (t, 1) & \text{if } t \in (-\infty, 0) \\ (0, 0) & \text{if } t = 0 \\ (t, -1) & \text{if } t \in (0, +\infty) \end{cases}$$

Remark 7. Orientation is important in Definition 6. In particular, the curve

$$t \in (a, b) \mapsto \gamma(a + b - t)$$

might not be self-contracted, while $\gamma : (a, b) \to \mathbb{R}^n$ is so. This unilateral aspect can be compared to Remark 1(ii).

We further recall from [24] some elementary properties of self-contracted curves.

- Let $\gamma : I \mapsto \mathbb{R}^n$ be a bounded self-contracted curve and $(a, b) \subset I$. Then, γ has a limit in $\mathbb{R}^n$ whenever $t \in (a, b)$ tends to an endpoint of (a, b). In particular, every self-contracted curve can be extended by continuity to the endpoints of I (possibly equal to $\pm\infty$).

In the sequel, we shall assume that every self-contracted curve $\gamma : I \mapsto \mathbb{R}^n$ is (defined and) continuous at the endpoints of I. The following result is a straightforward consequence of the above.

Corollary 8 (Convergence of bounded self-contracted curves)**.** *Every bounded self-contracted curve* $\gamma : (0, +\infty) \to \mathbb{R}^n$ *converges to some point* $x_0 \in \mathbb{R}^2$ *as* $t \to +\infty$*. Moreover, the function* $t \mapsto \operatorname{dist}(x_0, \gamma(t))$ *is nonincreasing.*

Corollary 8 reveals that the trajectories of a general gradient system

$$\dot\gamma(t) = -\nabla f(\gamma(t)), \qquad \gamma(0) = x_0 \in \mathbb{R}^n$$

might not be self-contracted curves. Indeed, the orbits of Example 2 are bounded but not converging. Nevertheless we have the following results.

- **(convex case)** The orbits of the subgradient system (1.6) of a convex continuous function f are self-contracted curves.
- **(quasiconvex case)** The orbits of the gradient system of a quasiconvex $C^{1,1}$ function are self-contracted curves.

Recall that a function $f : \mathbb{R}^n \to \mathbb{R}$ is called quasiconvex, if its sublevel sets $[f \le \lambda]$ ($\lambda \in \mathbb{R}$) are convex in $\mathbb{R}^n$. If f is differentiable, then it is quasiconvex if and only if for every $x, y \in \mathbb{R}^n$ the following implication holds:

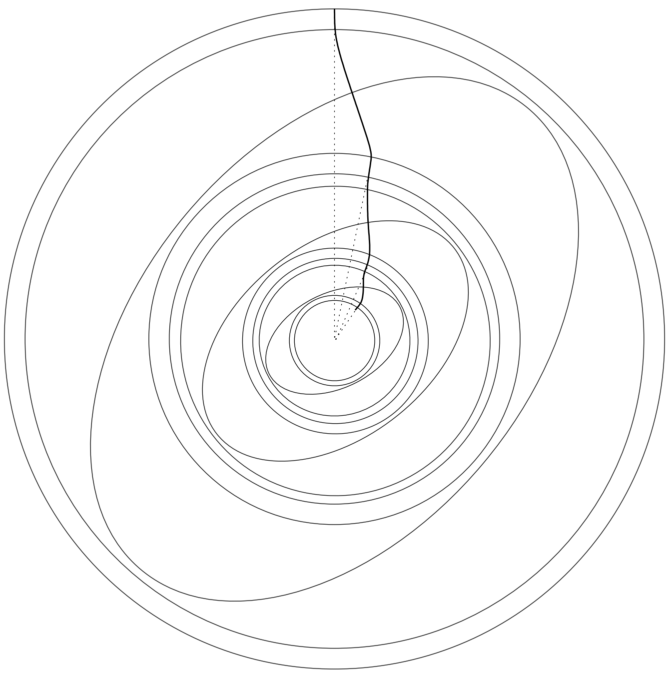
$$\langle \nabla f(x), y - x\rangle > 0 \Rightarrow f(y) \ge f(x)\,.$$

Our main result concerning self-contracted curves is restricted to the 2–dimensional case [24, Theorem 1.3].

Theorem 9 (Planar case)**.** *Every bounded continuous self-contracted planar curve γ is of finite length. More precisely,*

$$\text{length}(\gamma) \le (16\pi + 4)\, D(\gamma)$$

where $D(\gamma)$ is the distance between the endpoints of γ.



As a consequence we obtain the following.

Theorem 10 (Convex gradient system)**.** *Let $f : \mathbb{R}^2 \to \mathbb{R}$ be a smooth convex function with a unique minimum. Then, the trajectories γ of the gradient system (1.4) have a (uniformly) finite length.*

- **(open problem)** It is not known whether Theorem 9 and Theorem 10 hold in higher dimensions.

Let us conclude this section with a final remark.

Remark 11 (Failure of Thom conjecture in the convex case)**.** Theorem 10 guarantees that the orbits of the gradient flow of f have finite length (thus, *a fortiori*, are converging to the global minimum of f). However, in strong contrast to the analytic case, it may happen that each orbit turns around its limit infinitely many times (see counterexample in [24, Section 7.2] – an illustration is presented in Figure 1.4), so Thom conjecture fails in the convex case.

2 Łojasiewicz inequality and generalizations

The results of this section are motivated by a well-known result due to S. Łojasiewicz (see [37]), which asserts that if $f : \mathbb{R}^n \to \mathbb{R}$ is a real-analytic function and $\bar{x} \in f^{-1}(0)$ is a critical point of f, then there exist two constants $\theta \in [1/2, 1)$ and $k > 0$ such that

$$|f(x)|^{\theta} \leq k\,||\nabla f(x)|| \tag{2.1}$$

for all x belonging in a neighborhood U of $\bar{x}$. This result is a cornerstone of the modern theory of *semianalytic geometry* [36] and allows to deduce that all gradient orbits of f that converge to $\bar{x}$ and lie inside U have finite length. The proof is illustrated below:

Let $\gamma : [0, +\infty) \to U$ be a gradient trajectory of f, that is, $\dot{\gamma}(t) = -\nabla f(\gamma(t))$. Then,

$$\begin{aligned} -\left(\frac{k}{1-\theta}\right) \frac{d}{dt}\left[f(\gamma(t))^{1-\theta}\right] &= -k\,\langle \dot{\gamma}(t), \nabla f(\gamma(t)\rangle\, f(\gamma(t))^{-\theta} \\ &= k\,||\nabla f(\gamma(t)||^2\, f(\gamma(t))^{-\theta} \;\geq\; ||\nabla f(\gamma(t)|| = ||\dot{\gamma}(t)||, \end{aligned}$$

yielding (since $f(\gamma_\infty) = f(\bar{x}) = 0$) that

$$\text{length}(\gamma) = \int_0^{+\infty} ||\dot{\gamma}(t)||dt \leq \left(\frac{k}{1-\theta}\right) f(\gamma(0))^{1-\theta} < +\infty.$$

The restriction that the trajectory lies in U is not restrictive. In fact, it can be shown that any bounded trajectory of the gradient flow of a real–analytic function f is eventually trapped inside a convenient ball of its cluster point, that this tail has necessarily a finite length, and finally that the trajectory converges to this cluster point and has bounded length (see [8, Section 4] for an illustration of this standard technique in the more general context of subgradient systems).

Inequality (2.1) has been extended by K. Kurdyka in [31] for C^1 functions belonging to an arbitrary *o-minimal structure* (we give this definition in Section 3), in a way that allows to deduce the finiteness of the lengths of the gradient orbits in this more general context. In [8] and [9], a further extension has been realized to encompass (nonsmooth) functions and orbits of the corresponding subgradient systems. For sake of simplicity in the presentation, we limit ourselves in the smooth case and we fix a C^1 function $f : \mathbb{R}^n \to \mathbb{R}$.

We recall that a value $\bar{r} \in f(\mathbb{R}^n)$ is called *critical value* for f if there exists a critical point $\bar{x} \in \mathbb{R}$ such that $f(\bar{x}) = \bar{r}$. It is called *regular value*, if it is not a critical value. (Note that if $\bar{r}$ is a regular value, then $\mathcal{M} := [f = \bar{r}]$ is a submanifold of $\mathbb{R}^n$ of codimension 1.)

We introduce the following property KŁ($\bar{r}$) for the critical value $\bar{r}$ of f.

Definition 12 (property KŁ($\bar{r}$))**.** We say that the function f satisfies property KŁ($\bar{r}$) if there exists a C^1 function $\psi : (\bar{r}, \bar{r} + \delta) \to (0, \infty)$ with positive derivative and $\lim_{r \to \bar{r}} \psi(r) = 0$ such that

$$||\nabla(\psi \circ f)(x)|| \geq 1, \quad \text{for all } \bar{r} < f(x) < \bar{r} + \delta. \tag{2.2}$$

Remark 13. (i) If r is a regular value of an inf-compact C^1 function f, then KŁ(r) holds.

(ii) If KŁ(r) holds and $r \in f(\mathbb{R}^n)$ is a critical value, then r is an upper isolated critical value, that is, there exists $\delta > 0$ such that the interval $(r, r + \delta)$ is made up of regular values.

The aforementioned result of Kurdyka asserts that every o-minimal function f satisfies property KŁ(r) for every $r \in \mathbb{R}$ (see also Section 3). Thus this is true in particular for real-analytic functions. In fact (2.1) follows from KŁ($\bar{r}$) for $\bar{r} = 0$ and $\psi(r) = r^{1-\theta}$.

Note finally that the functions $\psi \circ f$ and f have the gradient curves on $[\bar{r} < f < \bar{r} + \delta]$. In view of (2.2) it is natural to call ψ a *desingularization* function for f. For convenience, we introduce the following notation.

(Desingularization functions) Given $\bar{r} \in f(\mathbb{R}^n)$ and $\delta > 0$ we set

$$\mathcal{K}(\bar{r}, \bar{r} + \delta) := \left\{ \psi \in C([\bar{r}, \bar{r} + \delta]) \cap C^1(\bar{r}, \bar{r} + \delta) : \ \psi(\bar{r}) = 0, \ \text{ and } \psi'(r) > 0, \forall r \in (\bar{r}, \bar{r} + \delta) \right\}, \tag{2.3}$$

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

As we shall see in Section 2.3, in some particular cases it is possible (and highly convenient) to obtain a desingularization function ψ which is *concave* and defined in the half-line $[0, +\infty)$.

2.1 Defragmented gradient curves

In this subsection we mention an important consequence of (2.2) for the asymptotic behavior of gradient systems. Let γ be a bounded orbit of (1.4) starting at $\gamma(0) = x_0$ and set $r_0 = f(x_0)$. In view of (1.7), the limit $r_\infty = \lim_{t \to \infty} f(\gamma(t))$ exists and the common value r_∞ is necessarily critical. (Note that we do not know yet that the limit of $\gamma(t)$ exists: r_∞ is the value of any ω-limit of γ). Note further that in view of (1.3) one has:

$$\text{length}(\gamma) = \int_0^{+\infty} \|\dot{\gamma}(t)\| dt = \int_{r_\infty}^{r_0} \|\dot{u}(r)\| dr = \int_{r_\infty}^{r_0} \frac{dr}{\|\nabla f(u(r))\|}. \tag{2.4}$$

Note that since the function

$$r \mapsto \frac{1}{\|\nabla f(u(r))\|}$$

is not bounded around r_∞ the above integral may diverge. But this cannot happen if f satisfies KŁ(r_∞). Indeed, in this case using (2.2) and the identity $f(u(r)) = r$ we deduce for some $\delta > 0$ that

$$\int_{r_\infty}^{r_\infty + \delta} \frac{dr}{\|\nabla f(u(r))\|} \leq \int_{r_\infty}^{r_\infty + \delta} \psi'(r)\, dr = \psi(r_\infty + \delta) - \psi(r_\infty) = \psi(r_\infty + \delta),$$

yielding

$$\text{length}(\gamma) \leq \psi(r_\infty + \delta) + \int_{r_\infty + \delta}^{r_0} \frac{dr}{\|\nabla f(u(r))\|} < +\infty. \tag{2.5}$$

Thus γ has a finite length, and converges to its (unique) ω-limit that we denote γ_∞. We have shown the following:

- If r is a (common) value of a point of the ω-limit of a bounded gradient orbit and f satisfies KŁ(r), then the ω-limit reduces to a singleton, the orbit is converging towards this point and its length is bounded.

But formula (2.5) also expresses a uniformity result which is not reflected in the previous statement. To make this precise, let us introduce the notion of *defragmented* (or piecewise) *gradient curve.*

Definition 14 (defragmented gradient curve)**.** A curve $\gamma : [0, T) \to \mathbb{R}^n$ ($T \in (0, \infty]$) is called defragmented gradient curve if there exists a countable partition of $[0, T]$ into (nonempty) intervals I_k such that:

– the restriction $\gamma|_{I_k}$ of γ to each interval I_k is a gradient curve (*i.e.* solution of (1.4)) ;

– 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 gradient orbits satisfy the above definition in a trivial way. It is now easy to obtain another consequence of (2.2).

- if f satisfies KŁ(r), then there exists $\delta > 0$ such that the length of every defragmented gradient curve that lies in $[r \le f \le r + \delta]$ is bounded by the number $\psi(r + \delta) - \psi(r)$.

In the sequel, for $r_1 > r_2$ we shall use the notation

$$\gamma \subset [r_2 \le f \le r_1]$$

to indicate that the image of the (defragmented) gradient curve γ lies in $[r_2 \le f \le r_1]$. We finish this section with an interesting observation ([7, Section 2.3]).

Remark 15 (Reduction to one–dimension)**.** Let ψ be a desingularization function of f as in (2.2). We set $\phi = \psi^{-1} : [0, \psi(\bar{r}+\delta)) \to [\bar{r}, \bar{r}+\delta]$ and we denote by χ_ϕ the gradient curve of the (trivial one-dimensional) system

$$\begin{cases} \dot{\chi}(r) = -\phi'(\chi(r)) \\ \chi(0) = 0 \end{cases}$$

Then length(γ) $\le$ length(χ_ϕ) for every gradient curve $\gamma \subset [\bar{r} \le f \le \bar{r} + \delta]$ of f.

2.2 The Kurdyka-Łojasiewicz inequality: characterizations and applications

In this section we give several characterizations of the property given in Definition 12. The proof of the following characterization is almost straightforward for $C^{1,1}$ inf-compact functions. (This result remains true for nonsmooth semiconvex functions in a Hilbert space, though its proof is not straightforward, see [13, Section 3.3] for details.)

Proposition 16 (local integrability of the inverse minimal gradient norm)**.** *Assume that the interval* $(\bar{r}, \bar{r} + \delta)$ *is made up of regular values and consider the function* $\varphi : (\bar{r}, \bar{r} + \delta) \to (0, +\infty)$ *defined by*

$$\varphi(r) = \max_{f(x)=r} \frac{1}{\|\nabla f(x)\|}.$$

Then KŁ($\bar{r}$) *holds if and only if* φ *is locally integrable around* $\bar{r}$.

We shall also need the notion of a *valley* ([21], [22])

Definition 17 (Valley)**.** For any $\rho > 1$ the ρ-valley $\mathcal{V}_\rho(\cdot)$ of f is defined as follows:

$$\mathcal{V}_\rho(r) = \left\{ x \in f^{-1}(r) : \|\nabla f(x)\| \le \rho \inf_{f(y)=r} \|\nabla f(y)\| \right\}, \qquad \text{for all } r \in (0, \bar{r}]. \tag{2.6}$$

We are ready to announce the main result of this section (this result has been established in [13, Section 3.3] in a more general (nonsmooth, infinite dimensional) framework. We refer to [27] for a discussion on the notion of *metric regularity*.

Theorem 18 (Characterization of KŁ-property)**.** *Let* $f : \mathbb{R}^n \to \mathbb{R}$ *be a* $C^{1,1}$ *inf-compact function. Assume* $\bar{r} = 0$ *is an upper isolated critical value* (*cf.* Remark 13 (ii)) *and let* $r_0 > 0$. *The following are equivalent:*

(i) (Kurdyka–Łojasiewicz inequality) *Property KŁ*(0) *holds with* $\operatorname{dom}\psi \supset [0, r_0)$.

(ii) (uniform length of defragmented gradient curves) *There exists* $M > 0$ *such that for every defragmented gradient curve* $\gamma \subset [0 \le f < r_0]$ *we have* length(γ) $< M$.

(iii) (Talweg on the valley) *For every* $\rho > 1$*, there exists a piecewise* C^1 *curve (discontinuous with countable pieces)* $\theta : (0, r_0) \to \mathbb{R}^n$ *with* finite length *such that* $\theta(r) \in \mathcal{V}_\rho(r)$, *for all* $r \in (0, r_0)$. *(Such a curve is called* talweg.*)*

(iv) (metric regularity) *There exists a* C^1 *function* $\psi : (0, r_0) \to \mathbb{R}_+$ *with* $\lim_{x \to 0^+} \psi(r) = 0$ *and positive derivatives such that*

$$\operatorname{Dist}([f \le r], [f \le s]) \le |\psi(r) - \psi(s)|, \qquad \textit{for all } r, s \in (0, r_0),$$

where $\operatorname{Dist}(A, B)$ *denotes the Hausdorff distance between the compact sets* A *and* B.

Let us mention an application of (iv) to the *proximal algorithm* (see also [3] for functions satisfying the Łojasiewicz inequality). We recall [42] that every $C^{1,1}$ function $f : \mathbb{R}^n \to \mathbb{R}$ is *lower-*C^2 (also known as *semiconvex* or *proximally smooth*) for some $\alpha > 0$ the function

$$\mathbb{R}^n \ni x \longmapsto f(x) + \frac{\alpha}{2}||x||^2$$

is convex. For such functions the so-called *proximal mapping* $\text{prox}_\lambda : \mathbb{R}^n \to \mathbb{R}^n$ ($\alpha^{-1} > \lambda > 0$) is defined by

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

is well-defined and single-valued. It is now easily seen that (iv) of Theorem 18 yields the following result.

Lemma 19 (KŁ–property and proximal mapping). *Under the assumptions of* Theorem 18(iv) *let* $x \in [0 < f < r_0]$ *be such that* $f(\text{prox}_\lambda x) > 0$. *Then*

$$||\text{prox}_\lambda x - x|| \leq \psi(f(x)) - \psi(f(\text{prox}_\lambda x)). \tag{2.7}$$

The above result has an important impact in the asymptotic analysis of the *proximal algorithm.* We recall that, given a sequence of positive parameters $\{\lambda_k\}_{k \geq 1} \subset (0, \alpha^{-1})$ and $x \in \mathbb{R}^n$ the proximal algorithm is defined as follows:

$$y_{k+1} = \text{prox}_{\lambda_k} y_k, \qquad y_0 = x,$$

or in other words

$$\{y_{k+1}\} = \text{argmin} \left\{ f(u) + \frac{1}{2\lambda_k}||u - y_k||^2 \right\}, \qquad y_0 = x.$$

The sequence $\{f(y_k)\}_k$ is decreasing and converges to a real number L. Since the sequence $\{y_k\}_{k \geq k_0}$ evolves in $[L \leq f < f(y_{k_0})]$ Lemma 19 yields

$$\sum_{k=p}^{q} ||y_{k+1} - y_k|| \leq \psi(f(y_{q+1})) - \psi(f(y_p)),$$

for all integers $k_0 \leq p \leq q$, which implies that y_k converges to y_∞. This shows the following result ([13, Section 3.4])

Theorem 20 (strong convergence of the proximal algorithm). *Let* $f : \mathbb{R}^n \to \mathbb{R}$ *be a* $C^{1,1}$ *function which is bounded from below. Let* $x \in \text{dom} f$, $\{\lambda_k\}_k \subset (0, \alpha^{-1})$ *and* $L := \lim_{k \to \infty} f(y_k)$ *and assume that* KŁ(L) *holds true. Then it holds:*

$$||y_\infty - y_k|| \leq \psi(f(y_k)), \qquad \textit{for all } k \geq k_0. \tag{2.8}$$

Remark 21 (Step size). Note that the step-size sequence $\{\lambda_k\}_k$ does not appear explicitly in the estimate (2.8), but instead, it is hidden in the sequence of values $\{f(y_x^k)\}_k$. In practice the choice of the step parameters λ_k is however crucial to obtain the convergence of $\{f(y^k)\}_k$ 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 $\eta \in (0, \alpha^{-1})$.

2.3 Convex case: asymptotic equivalence between continuous and discrete systems

In case of a convex function f, with global minimum value equal to 0, if there exists a desingularizing function $\psi \in \mathcal{K}(0, r_0)$ satisfying KŁ(0), then this function can be taken concave with domain $[0, \infty)$. This is a very important result [13, Section 4.2], which has a striking consequence in the asymptotic equivalence of continuous and discrete systems.

Theorem 22 (KŁ–property – convex case). *Let* $f : H \to \mathbb{R} \cup \{+\infty\}$ *be a* $C^{1,1}$ *convex function with* $\inf f = 0$. *The following statements are equivalent and imply the existence of a minimizer.*

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

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

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

$$||\nabla(\psi \circ f)(x)|| \geq 1, \qquad \text{for all } x \notin [f = 0]. \tag{2.9}$$

To describe the consequence of the above result, let us fix a $C^{1,1}$ convex inf-compact function $f : \mathbb{R}^n \to \mathbb{R}$, any $\beta > 0$ and $x \in \mathbb{R}^n$ and consider any sequence $\{Y_x^k\}$ satisfying

$$\begin{cases} \beta \, ||\nabla f(Y_x^k)|| \, ||Y_x^{k+1} - Y_x^k|| \leq f(Y_x^k) - f(Y_x^{k+1}), \quad k = 1, 2, \ldots \\ Y_x^0 = x \end{cases} \tag{2.10}$$

Remark 23. Condition (2.10) has been called *Primary Descent Condition* in [1, § 3.1]. It is fulfilled by several explicit gradient–like methods, including trust region methods, line–search gradient methods and some Riemannian variants; see [34], [1] for examples and references. It defines a descent sequence, that is, $f(Y_x^k) \geq f(Y_x^{k+1})$, which implies in particular that for bounded sequences, $\{f(Y_x^k)\}$ converges as k goes to infinity.

The following theorem (see [13, Section 4.4]) establishes connections between length boundedness properties of continuous gradient methods and length boundedness of discrete gradient iterations.

Theorem 24 (discrete vs continuous)**.** *Let f be a $C^{1,1}$ convex inf-compact function with $\min f = 0$. The following statements are equivalent:*
(i) (Kurdyka–Łojasiewicz inequality) *There exist $r_0 > 0$ and $\psi \in \mathcal{K}(0, r_0)$ such that*

$$||\nabla(\psi \circ f)(x)|| \geq 1, \qquad \text{for all } x \in [0 < f \leq r_0]. \tag{2.11}$$

(ii) (Length boundedness of piecewise gradient iterates) *For all $\beta > 0$ and $r_0 > 0$, and for all sequences of gradient iterates of the form*

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

with $f(x_0) \leq r_0$, $f(Y_{x_{i+1}}^0) = f(x_{i+1}) \leq f(Y_{x_i}^{k_i})$ and $\{Y_{x_i}^j : j = 0, \ldots, k_i\}$ satisfying (2.10) for all $i \in \mathbb{N}$ we have

$$\sum_{i=0}^{+\infty} \sum_{l=0}^{k_i} ||Y_{x_i}^{l+1} - Y_{x_i}^l|| \leq \frac{1}{\beta} \psi(r_0).$$

(iii) (Length boundedness of defragmented gradient curves) *For every defragmented gradient curve (of the gradient system defined by f) $\chi : [0, +\infty) \to \mathbb{R}^n$ with $f(\gamma(0)) < r_0$, we have*

$$\text{length}(\chi) \leq \psi(r_0)\,.$$

The assumption f is convex seems necessary for the proof of implication (i)⇒(ii) and to assert $f(Y_0^k) \to \inf f$. For this reason Theorem 24 is not stated in a more general setting (as for instance, semiconvex functions in a local version).

- **(open problem)** Under which type of conditions (other than convexity or o-minimality of f) the function ψ of (2.11) can be taken concave?

2.4 A convex counterexample

A natural question is raised in the previous subsection: Do all convex inf-compact functions admit a desingularization function? In other words does KŁ$(\min f)$ hold true for a convex function? In view of Section 2.2, if true, this conjecture would imply that all defragmented gradient curves have bounded length. Notice that the weaker assertion of the uniform boundedness of the lengths of gradient curves is already stated as an open question in Section 1.3 (in Section 1.4 we show it is true in $\mathbb{R}^2$).

In this section we give a negative answer to this conjecture, by constructing a C^2 convex function on $\mathbb{R}^2$ with compact level sets, which fails to satisfy the KŁ–inequality. This construction is very involved [13, Section 4.3] and will be described very roughly.

Step 1. We first show that any sequence of sublevel sets of a convex function that satisfies the KŁ–inequality must comply with a specific property and we build a sequence C_k of nested convex sets for which this property fails.

The following lemma provides a decreasing sequence of convex compact subsets in $\mathbb{R}^2$ which cannot be a sequence of prescribed sublevel sets of a function satisfying the KŁ–inequality.

Lemma 25. *There exists a decreasing sequence of convex compact subsets $\{C_k\}_k$ of $\mathbb{R}^2$ such that:*

(*i*) C_0 *is the unit disk* $D := B(0,1)$ *;*

(*ii*) $C_{k+1} \subset \operatorname{int} C_k$ *for every* $k \in \mathbb{N}$ *;*

(*iii*) $\bigcap_{k\in\mathbb{N}} C_k$ *is the disk* $D_r := B(0,r)$ *for some* $r > 0$ *;*

(*iv*) $\sum_{k=0}^{+\infty} \operatorname{Dist}(C_k, C_{k+1}) = +\infty$.

We first fix a sequence $\{T_n\}_{n\in\mathbb{N}}$, with $T_{n+1} \subset \operatorname{int} T_n$ and $\operatorname{dist}(T_{n+1}, T_n) \approx 1/n^2$ for every $n \in \mathbb{N}$. Then between two successive sets T_n and T_{n+1} we introduce n sets $\{T_{n,k}\}_{k=1}^{n}$ such that $\operatorname{dist}(T_{n,k}, T_{n,k+1}) \approx 1/2n^2$, so that $\sum_{k=1}^{n-1} \operatorname{dist}(T_{n,k-1}, T_{n,k}) \approx 1/2n$. We rename the family $\{T_{n,k}\}_{k,n}$ to $\{C_n\}_{n\in\mathbb{N}}$ and observe that (iv) holds. Figure 2 gives an idea of how these sets are constructed.

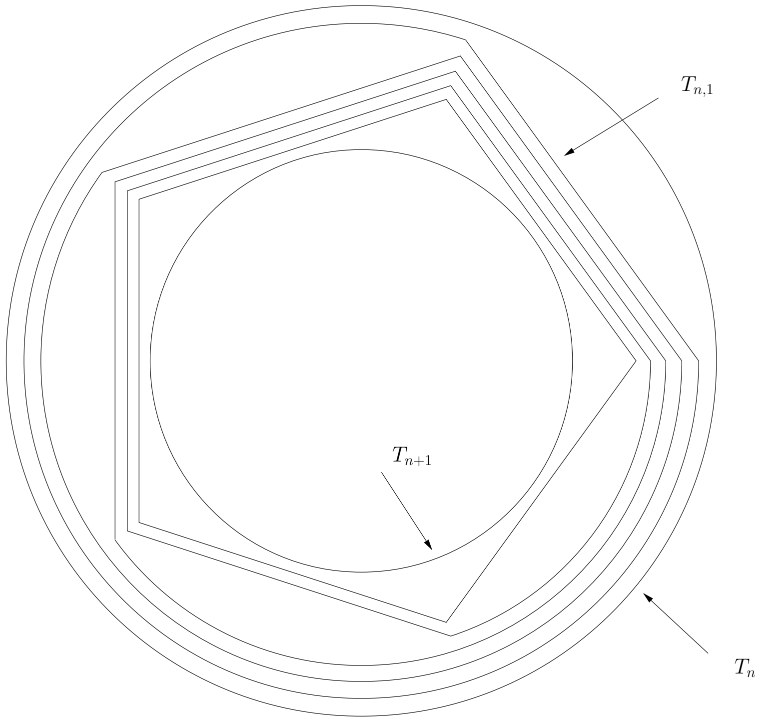


Figure 2: *Construction of the sequence $T_{n,k}$*

Note that in view of Theorem 18 (iv), the above sequence $\{C_k\}_k$ cannot be part of the sublevel sets of any function that satisfies the KŁ–inequality.

Step 2. We show that there exists a (nonsmooth) convex function which admits C_k as sublevel sets.

This part relies on the use of support functions[2] and a result of Kannai [30] (see also [46]). Let us recall this result: Let $\{C_k\}_{k\in\mathbb{N}}$ be a decreasing sequence of convex compact subsets of $\mathbb{R}^2$ such that $C_{k+1} \subset \operatorname{int} C_k$ ($\operatorname{int} C_k$ stands for the interior of C_k in $\mathbb{R}^2$). Set

$$K_k = \max_{||x^*||=1} \frac{\delta_{C_{k-1}}(x^*) - \delta_{C_k}(x^*)}{\delta_{C_k}(x^*) - \delta_{C_{k+1}}(x^*)}.$$

Then for every real sequence $\{\lambda_k\}_{k\in\mathbb{N}}$ satisfying

$$0 < K_k(\lambda_k - \lambda_{k+1}) \le \lambda_{k-1} - \lambda_k \quad \text{for every } k \ge 1,$$

there exists a continuous convex function f such that for every $k \in \mathbb{N}$, $\{f \le \lambda_k\} = C_k$. Moreover, $\lim_{k\to\infty} \lambda_k = \min f$ and, for any $k \ge 0$ and $\lambda \in [\lambda_{k+1}, \lambda_k]$, we have

$$\{f \le \lambda\} = \left(\frac{\lambda - \lambda_{k+1}}{\lambda_k - \lambda_{k+1}}\right) C_k + \left(\frac{\lambda_k - \lambda}{\lambda_k - \lambda_{k+1}}\right) C_{k+1} \tag{2.12}$$

(*i.e.,* the level-sets of f are convex interpolations of the two nearest prescribed level-sets). Now applying this to the nested sequence of Lemma 25 we obtain a convex inf-compact function with at least one defragmented subgradient curve of infinite length.

One can now slightly modify the above level sets in order to obtain a C^2 convex function with the same properties. This part is very technical (see details in [13, Section 4.3]) and will be omitted.

Remark 26. Note that the lengths of (sub)gradient curves of f are uniformly bounded (*cf.* Section 1.4). Thus the above counterexample also shows that uniform boundedness of the lengths of the subgradient curves (starting from a prescribed level set $[f = r_0]$) does not imply uniform boundedness of the lengths of the piecewise subgradient curves γ lying in $[\min f < f < r_0]$.

2.5 The semialgebraic case

A set defined by finitely-many polynomial inequalities is called *basic semialgebraic*; finite unions of such sets are called *semialgebraic set.* Such sets comprise a rich class that is stable under many mathematical operations. Semialgebraic sets are often easy to recognize, even without an explicit representation as a union of basic sets, as a consequence of the Tarski–Seidenberg principle, which states that the projection of a semialgebraic set is semialgebraic. For example, the feasible region of any semidefinite program is semialgebraic. A function (respectively, multivalued map) is called *semialgebraic*, if its graph is a semialgebraic set. For a function, it is also equivalent to saying that its epigraph

$$\operatorname{epi} f := \{(x,\beta) \in \mathbb{R}^n \times \mathbb{R} : f(x) \le \beta\}$$

is semialgebraic.

Checking the semialgebraicity of a set in practice is often easy. A formal approach to semialgebraicity can be found in [18, Chapter 2.1.2] or [38, Chapter 3.3]. In particular, denoting by $\mathbb{R}[x]$ the ring of real polynomials on $\mathbb{R}^n$, let us consider the following three rules for defining *first-order formulas* (in the language of the ordered field $\mathbb{R}$). Call $\mathcal{F}$ the set of all such formulas.

– for every $p \in \mathbb{R}[x]$, we have $[p > 0] \in \mathcal{F}$ and $[p = 0] \in \mathcal{F}$;

– if $\Phi_1, \Phi_2 \in \mathcal{F}$ then $\Phi_1 \vee \Phi_2$, $\Phi_1 \wedge \Phi_2$ and $\neg\Phi_1$ all belong to $\mathcal{F}$;

– if $\Phi(x,y) \in \mathcal{F}$ then $\forall x \Phi(x,y) \in \mathcal{F}$ and $\exists x \Phi(x,y) \in \mathcal{F}$.

Note that the first two rules correspond to the (0-order) definition of semialgebraic set we have just given, while the third one allows to consider first order formulas. Another way to conceive the Tarski–Seidenberg principle is to claim that every first-order formula in the language of ordered fields defines a semialgebraic set. In other words, the Tarski–Seidenberg principle is a *quantifier elimination principle.*

[2]For any convex set $C \subset \mathbb{R}^n$, the support function of C is defined as $\delta_C(x^*) = \sup_{x\in C}\langle x, x^*\rangle$ for all $x^* \in \mathbb{R}^n$.

As an illustration of how this works in practice let us prove that the operator $\hat{\partial} f : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ (Fréchet subdifferential of f) is semialgebraic. Recall that $\hat{\partial} f(x)$ is the set of all $x^* \in \mathbb{R}^n$ satisfying

$$\liminf_{y \to x, y \neq x} \frac{f(y) - f(x) - \langle x^*, y - x \rangle}{\|y - x\|} \;\geq\; 0. \tag{2.13}$$

To prove that $\operatorname{Graph} \hat{\partial} f$ is semialgebraic, set $A = \operatorname{epi} f$, $\Gamma = \operatorname{Gr} f$ and $D = \operatorname{dom} f$, which are all semialgebraic sets. According to (2.13) the graph $\operatorname{Graph} \hat{\partial} f$ of the Fréchet subdifferential $\hat{\partial} f(x)$ is the set of $(x, x^*) \in \mathbb{R}^n \times \mathbb{R}^n$ such that

$$\{\forall \varepsilon > 0, \exists \delta > 0, \forall (y, \beta) \in (B(x, \delta) \times \mathbb{R}) \cap A \Rightarrow (y, \beta - \langle x^*, y - x \rangle + \varepsilon \|y - x\|) \in A\},$$

where $B(x, \delta)$ denotes the open ball of center x and radius $\delta > 0$. Since the above first order formula involves only semialgebraic sets (namely, the sets $B(x, \delta), \mathbb{R}$ and A), it follows that $\operatorname{Graph} \hat{\partial} f$ is semialgebraic.

Let us gather below some elementary properties of semialgebraic sets (e.g. [6], [5], [18]):

- Semialgebraic sets are closed under finite union, intersection and complementarity (by definition).

- If A is semialgebraic then so are its closure $\operatorname{cl} A$, its interior $\operatorname{int} A$, and its boundary $\operatorname{bd} A$ (stability under projection).

- Given a semialgebraic set S, the distance $d_S(x) := \inf \{\|x - a\| : a \in S\}$ is a semialgebraic function.

- *Path connectedness*: Any semialgebraic set has a finite number of connected components. Each component is semialgebraic and semialgebraically path connected, that is, every two points can be joined by a continuous semialgebraic path that lies entirely in the set.

- *Curve selection lemma*: If A is a semialgebraic subset of $\mathbb{R}^n$ and $a \in \operatorname{bd} A$, then there exists an analytic path $z : (-1, 1) \to \mathbb{R}^n$, satisfying $z(0) = a$ and $z((0, 1)) \subset A$.

- The image or the preimage of a semialgebraic set by a semialgebraic function (respectively, semialgebraic multivalued map) is semialgebraic.

- *Monotonicity lemma* Take $\alpha < \beta$ in $\mathbb{R}$. If $\varphi : (\alpha, \beta) \to \mathbb{R}$ is a semialgebraic function, then there is a partition $t_0 := \alpha < t_1 < \ldots < t_{l+1} := \beta$ of (α, β), such that $\varphi|_{(t_i, t_{i+1})}$ is C^∞ and either constant or strictly monotone, for $i \in \{0, \ldots, l\}$.

Moreover φ admits a *Puiseux development* at $t = \alpha$, that is, there exist $\delta > 0$, integers $k, l \in \mathbb{Z}$ with $k > 0$ and sequence $\{a_n\}_{n \geq l} \subset \mathbb{R}$ such that

$$\varphi(t) = \sum_{n \geq l} a_n (t - \alpha)^{n/k}, \quad \text{for all } t \in (\alpha, \alpha + \delta)\,.$$

- *Łojasiewicz factorization lemma*: Let $K \subset \mathbb{R}^n$ be a compact set and $f, g : K \to \mathbb{R}$ be two continuous semialgebraic functions (thus K has to be semialgebraic). If $f^{-1}(0) \subset g^{-1}(0)$, then there exist $c > 0$ and a positive integer β such that $|g(x)|^\beta \leq c\,|f(x)|$ for all $x \in K$.

Let us now illustrate the power of semialgebraic techniques by giving a direct proof that every C^1 convex (bounded from below) function $f : \mathbb{R}^n \to \mathbb{R}$ satisfies KŁ$(\min f)$. In fact the next result holds in a more general case (f nonsmooth, subanalytic). We refer to [8] for details.

Theorem 27 (Łojasiewicz inequality in the convex case)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a C^1 convex semialgebraic function with* $\operatorname{arg\,min} f \neq \emptyset$. *For any bounded set K there exists an exponent $\theta \in [0, 1)$ and $\varrho > 0$ such that for all $x \in K$*

$$|f(x) - \min f|^\theta \leq \varrho\, \|\nabla f(x)\|. \tag{2.14}$$

In particular KŁ$(\min f)$ holds for some $\delta > 0$ and for the function

$$\psi(r) = (r - \min f)^{1 - \theta}, \qquad \textit{for all } r \in (\min f, \min f + \delta).$$

Proof. We may assume that K is compact and semialgebraic (by taking a closed ball containing K). Set $S = \arg\min f$. Then for all $x \in K$, we have $d_S(x) = 0 \iff |f(x) - \min f| = 0$, so by the Łojasiewicz factorization lemma, there exist $\beta > 1$ and $c > 0$ such that

$$[d_S(x)] \le c^{1/\beta} |f(x) - \min f|^{1/\beta}. \tag{2.15}$$

Using convexity we obtain $|f(x) - f(a)| \le \|\nabla f(x)\| \, \|x - a\|$, for all a in S. Taking the infimum over all $a \in S$ we deduce

$$|f(x) - \min f| \le \|\nabla f(x)\| \, d_S(x), \tag{2.16}$$

and therefore

$$|f(x) - \min f| \le c^{1/\beta} \|\nabla f(x)\| \cdot |f(x) - \min f|^{1/\beta}.$$

By setting $\theta = 1 - \beta^{-1}$, and $\varrho = c^{1/\beta}$ we obtain (2.14). $\diamond$

3 Tame variational analysis

The qualitative properties of semialgebraic mappings are shared by a much bigger class, namely the class consisting of the so-called *definable mappings in an o-minimal structure* over $\mathbb{R}$ (or simply *definable mappings*). A slightly more general notion is that of a *tame* mapping, being a mapping whose graph has a definable intersection with every "bounded box". O-minimal structures correspond in some sense to an axiomatization of some of the prominent geometrical properties of semialgebraic geometry ([26], [19]) and particularly of the stability under projection. A similar axiomatization has been proposed by Shiota in [44].

3.1 Definable functions, o-minimal structures

Let us recall the definition of an o-minimal structure (see [26] for example).

Definition 28 (o-minimal structure)**.** An o-minimal structure on the ordered field $\mathbb{R}$ is a sequence of Boolean algebras $\mathcal{O} = \{\mathcal{O}_n\}_{n\ge1}$ such that for each $n \in \mathbb{N}$

(i) $A \in \mathcal{O}_n \Longrightarrow A \times \mathbb{R} \in \mathcal{O}_{n+1}$ and $\mathbb{R} \times A \in \mathcal{O}_{n+1}$;

(ii) $A \in \mathcal{O}_{n+1} \Longrightarrow \Pi(A) \in \mathcal{O}_n$
($\Pi : \mathbb{R}^{n+1} \to \mathbb{R}^n$ denotes the canonical projection onto $\mathbb{R}^n$)

(iii) $\mathcal{O}_n$ contains the family of algebraic subsets of $\mathbb{R}^n$, that is, the sets of the form

$$\{x \in \mathbb{R}^n : p(x) = 0\},$$

where $p : \mathbb{R}^n \to \mathbb{R}$ is a polynomial function ;

(iv) $\mathcal{O}_1$ consists exactly of the finite unions of intervals and points.

An important example of o-minimal structure is the collection of *semialgebraic sets.* Indeed, properties (i),(iii) and (iv) of Definition 28 are straightforward, while (ii) is a consequence of the Tarski–Seidenberg principle.

A subset A of $\mathbb{R}^n$ is called *definable* (in the o-minimal structure $\mathcal{O}$) if it belongs to $\mathcal{O}_n$. Given any $S \subset \mathbb{R}^n$ a mapping $F : S \to \mathbb{R}$ is called *definable* in $\mathcal{O}$ if its graph is a definable subset of $\mathbb{R}^n \times \mathbb{R}$.

A major part of the interest in dealing with definable objects consists of their remarkable stability properties (which are essentially the ones stated for semialgebraic sets in Section 2.5, see also [19]). These properties rely on the projection stability assumption. In particular, any o-minimal C^1 function f has a finite number of critical values and satisfies property KŁ$(\bar{r})$ at each one of them [31], that is, there exists a desingularization function $\psi \in \mathcal{K}(\bar{r}, \bar{r} + \delta)$ satisfying (2.2). In [9] this result is improved as follows:

Theorem 29 (KŁ-inequality – uniform desingularization). *Let $f : U \to \mathbb{R}_+$ be a definable differentiable function, where U is a definable submanifold of $\mathbb{R}^n$ (not necessarily bounded). Let us denote by $C_1, \ldots, C_m$ the (necessarily finite) connected components of $(\nabla f)^{-1}(\{0\})$ and by $r_1, \ldots, r_m$ the corresponding critical values [3]. Then there exist a continuous definable function $\psi : [0, \varepsilon_0) \to \mathbb{R}_+$ which is C^1 on $(0, \varepsilon_0)$ with $\psi(0) = 0$, and relatively open neighborhoods V_i of C_i in U for each $i \in \{1, \ldots, m\}$, such that for all $x \in V_i$ we have*

$$||\nabla[\psi \circ (f - r_i)](x)|| \geq 1.$$

Let us finally recall from Section 2.1 that as a consequence of KŁ($\bar{r}$) any bounded trajectory γ of the gradient system (1.4) of a C^1-definable function has finite length, and is thus converging to $\gamma_\infty = \lim_{t\to\infty} \gamma(t)$.

- **(open problem)** It is not known if the "gradient conjecture" of R. Thom (convergence of secants) remains true in the o-minimal case. It is also unknown whether or not an appropriate reformulation of it would hold in the case of a semiconvex subanalytic function (nonsmooth setting).

3.2 Stratification vs Clarke subdifferential

An important property of definable sets is that of *Whitney stratification* ([26, §4.2]): every definable set can be written as a finite disjoint union of manifolds (*strata*) that fit together in a regular way (Whitney property). The Whitney property can be seen as a *normal regularity* condition on the stratification ([28, Definition 5]). In fact a slightly more general formulation holds true.

Theorem 30 (Whitney $\mathcal{C}^k$ stratification). *For any $k \in \mathbb{N}$ and any definable subsets $X_1, \ldots, X_l$ of $\mathbb{R}^n$, we can write $\mathbb{R}^n$ as a disjoint union of finitely many definable $\mathcal{C}^k$ manifolds $\{\mathcal{M}_i\}_i$ (that is, $\mathbb{R}^n = \dot\bigcup_{i=1}^{I} \mathcal{M}_i$) so that each X_j is a finite union of some of the $\mathcal{M}_i$'s. Moreover, the induced stratification $\{\mathcal{M}_i^j\}_i$ of X_j has the Whitney property that is, for any sequence $\{x_\nu\}_\nu \subset \mathcal{M}_i^j$ converging to $x \in \mathcal{M}_{i_0}^j$ we have*

$$\limsup_{v\to\infty} N_{\mathcal{M}_i^j}(x_\nu) \subset N_{\mathcal{M}_{i_0}^j}(x)\,. \tag{3.1}$$

The *dimension* $\dim(X)$ of a definable set X is defined as the dimension of the manifold of highest dimension of its stratification. This dimension is well defined and independent of the stratification of X. See [18, Section 3.3] (for semialgebraic sets), [26], [19].

Our central idea is to relate objects from two distinct mathematical sources: variational analysis and differential geometry. Specifically, given a definable locally Lipschitz function, we derive a lower bound on the norms of Clarke subgradients at a given point in terms of the *Riemannian gradient* with respect to the stratum containing that point. This direct consequence of a *projection formula* has as a corollary a Morse-Sard type theorem for Clarke critical points.

Let U be a nonempty open subset of $\mathbb{R}^n$ and $f : U \to \mathbb{R}$ a Lipschitz continuous function. Let further D_f denote the set of its points of differentiability. (The Rademacher theorem asserts that D_f is of full measure in U, thus dense there.) The *Clarke subdifferential* of f at x_0 is defined as follows:

$$\partial^\circ f(x_0) \;=\; \mathrm{co}\{\lim_{x_n\to x_0} \nabla f(x_n) : \{x_n\}_{n\geq 1} \subset D_f \diagdown \mathcal{N}\}, \tag{3.2}$$

where $\mathcal{N}$ is a null subset of $\mathbb{R}^n$ and co(S) denotes the convex hull of S. We recall that $\partial^\circ f(x_0)$ is a nonempty convex compact subset of $\mathbb{R}^n$ and that if f is of class C^1 (or more generally, strictly differentiable at x_0) then $\partial^\circ f(x_0) = \{\nabla f(x_0)\}$. A point $x_0 \in U$ is called *Clarke critical*, if $0 \in \partial^\circ f(x_0)$. We say that $y_0 \in f(U)$ is a *Clarke critical value* if the level set $f^{-1}(y_0)$ contains at least one Clarke critical point.

Let us assume that f is locally Lipschitz on U and its graph Graph f admits a C^k-Whitney stratification, that is,

$$\mathrm{Graph}\, f = \cup_i S_i$$

where S_i are definable C^k-submanifolds of $\mathbb{R}^n \times \mathbb{R}$. Denote by X_i the projection of S_i onto $\mathbb{R}^n$. Then, since f is locally Lipschitz, the tangent space $T_{S_i}(u)$ of S_i at any $u = (x, f(x)) \in S_i$ is transversal to $\{0\}_n \times \mathbb{R}$ yielding that $\{X_i\}_i$ is a stratification of U (the domain of f).

[3] Every differentiable o-minimal function f is constant on each connected component of its critical set $S := \nabla f^{-1}(0)$.

For any $x \in U$, we denote by X_x (respectively, S_x) the stratum of $\mathcal{X}$ (respectively of $\mathcal{S}$) containing x (respectively $(x, f(x))$). The manifolds X_i are here endowed with the metric induced by the canonical Euclidean scalar product of $\mathbb{R}^n$. Using the inherited Riemannian structure of each stratum X_i of $\mathcal{X}$, for any $x \in X_i$, we denote by $\nabla_R f(x)$ the gradient of $f|_{X_i}$ (restriction of f to the stratum X_x) at x with respect to the stratum $X_i, \langle \cdot, \cdot \rangle$.

If a stratum S_x is of full dimension (that is, $\dim S_x = \dim X_x = n$), then f is C^k on X_x and the Riemann gradient $\nabla_R f(x)$ coincides with the (usual) gradient of f at x. Denote by U_1 the union of all strata X_i of full dimension, and note that $U_1 \subset D_f$ and $U \diagdown U_1$ is a null set (as a finite union of manifolds of lower dimension). Assume now that $\{x_\nu\}_\nu \subset U_1$ and $\{x_\nu\} \to x_0$ and let $p \in \mathbb{R}^n$ be an accumulation point of $\{\nabla f(x_\nu)\}_\nu$. Then we may assume (taking a subsequence if necessary) that $\{(\nabla f(x_\nu), -1)\} \to (p, -1)$ and that $\{(x_\nu, f(x_\nu)\}_\nu$ lies in the same stratum, say S. Let us denote by S_0 the stratum of $(x_0, f(x_0))$. Since $(\nabla f(x_\nu), -1) = (\nabla_R f(x_\nu), -1) \in N_S((x_\nu, f(x_\nu)))$, the Whitney property (3.1) yields that $(p, -1) \in N_{S_0}((x_0, f(x_0)))$. Taking convex hull we obviously remain in the linear space $N_{S_0}(x_0)$ (recall that S_0 is a manifold), therefore $(\partial^\circ f(x_0), -1) \subset N_{S_0}(x_0)$. This shows the following.

Proposition 31 (Projection formula). *Let $f : U \subset \mathbb{R}^n \to \mathbb{R}$ be a locally Lipschitz function and assume that* Graph f *admits a Whitney stratification $\mathcal{S} = (S_i)_{i\in I}$. Then for all $x \in U$ we have*

$$\operatorname{Proj}_{T_x X_x} \partial^\circ f(x) \,=\, \{\nabla_{\mathrm{R}} f(x)\}, \tag{3.3}$$

where $\operatorname{Proj}_{\mathcal{V}} : \mathbb{R}^n \to \mathcal{V}$ *denotes the orthogonal projection on the vector subspace $\mathcal{V}$ of $\mathbb{R}^n$.*

A more general statement appears in [9, Proposition 4]. See also [28, Theorem 7] for a further extension to definable multifunctions. Note also that the analogous result for the Fréchet subdifferential is straightforward and does not depend on the stratification: we only need to assume that f is partly smooth at x (see [35] for the definition) and consider the tangent space of the corresponding manifold there (see [23] e.g.). Finally, the above technique (use of Whitney stratifications) has been employed in [32] for the smooth case, in order to establish the Łojasiewicz inequality for C^1 subanalytic functions.

3.3 Sard-type theorem for (nonsmooth) tame functions

An easy consequence of the projection formula (Proposition 31) is that for all $x \in U$ and $x^* \in \partial^\circ f(x)$, we have $||\nabla_R f(x)|| \le ||x^*||$. This allows to establish a Morse–Sard type theorem for definable functions (see [9] for a more general result), using the classical Brown-Sard theorem ([16], [43]).

Theorem 32 (Morse-Sard theorem for definable functions). *For any locally Lipschitz definable function $f : U \subset \mathbb{R}^n \to \mathbb{R}$, the set of Clarke critical values of f is finite.*

Let us note though that the above result is not true for any possible variational notion of criticality: for instance, for any continuous function $f : U \subset \mathbb{R}^n \to \mathbb{R}$ we can define the concept of *broadly critical point* as follows:

Definition 33. The point $x \in \mathbb{R}^n$ is called broadly critical for the continuous function f if

$$0 \in \bigcap_{\varepsilon>0} \overline{\mathrm{co}} \left\{ \bigcup_{x \in B(x_0,\varepsilon)} \hat{\partial} f(x) \right\}$$

where $\hat{\partial} f(x)$ is the Fréchet subdifferential defined in (2.13).

Remark 34. Using Caratheodory's theorem we have the following equivalent definition: A point x is broadly critical if for every $\varepsilon > 0$ there exist $\{x_i\}_{i\in\{1,\ldots,n+1\}} \subset B(x, \varepsilon)$ such that

$$0 \in \mathrm{co} \left\{ \hat{\partial} f(x_i) : i \in \{1, \ldots, n+1\} \right\}.$$

This notion coincides with Clarke criticality whenever f is locally Lipschitz. But in the continuous case one has the following result [12, Section 4].

- There exists a continuous subanalytic function $f : \mathbb{R}^3 \to \mathbb{R}$ which is not constant on a segment made up of broadly critical points.

3.4 Applications

Tame optimization is a rapidly developing domain of modern variational analysis. Let us mention briefly, without further details, a couple of recent results obtained by this combination of techniques (vatiational analysis, geometry).

- (Semismoothness) Every locally Lipschitz definable function $F : \mathbb{R}^n \to \mathbb{R}^n$ is semismooth ([10]).
- (Continuity of tame multifunctions) If $T : \mathcal{X} \rightrightarrows \mathbb{R}^m$ is a closed-valued definable multifunction, where $\mathcal{X} \subset \mathbb{R}^n$ is definable, then T fails to be continuous in a set of dimension at most $(\dim \mathcal{X} - 1)$ ([25]).
- (Genericity of nonexistence of first integrals) Let M be a C^1 compact submanifold of $\mathbb{R}^n$ and $\epsilon > 0$. For the C^1 topology, the set of vector fields in M that do not admit Lipschitz continuous first integrals which are *essentially o–minimal* with respect to a given definable ϵ–approximation of M is generic ([20]).
- (Genericity of partial smoothness) Generically, the optimal solution of a linear optimization over a fixed tame compact convex feasible region is unique and lies on a unique manifold, around which the feasible region is "partly smooth". Furthermore, second-order optimality conditions hold, guaranteeing smooth behavior of the optimal solution under small perturbations to the objective ([11]).

Let us finally quote the interesting survey [29] which contains more or less the State-of-the-Art and other applications.

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] Aubin, J.-P. & Cellina, A., *Differential Inclusions*, Grundlehren der Mathematischen Wissenschaften **264** (Springer, 1984)

[3] Attouch, H. & Bolte, J. On the convergence of the proximal algorithm for nonsmooth functions involving analytic features, *Math. Programming* **116** (2009), 5–16.

[4] 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.

[5] Benedetti, R. & Risler, J.-J., *Real Algebraic and Semialgebraic Sets*, Hermann, Éditeurs des sciences et des Arts, (Paris, 1990).

[6] Bierstone, E. & Milman, P., Semianalytic and subanalytic sets, *IHES Publ. Math.* **67** (1988), 5–42.

[7] Bolte, J., *Sur quelques principes de convergence en optimisation*, Habilitation Thesis, Paris 6, 2009.

[8] Bolte, J., Daniilidis, A. & Lewis, A., The Łojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems, *SIAM J. Optim.* **17** (2007), 1205–1223.

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

[10] Bolte, J., Daniilidis, A. & Lewis, A., Tame functions are semismooth, *Math. Programming (Series B)* **116** (2009), 115–127.

[11] Bolte, J., Daniilidis, A. & Lewis, A., Generic optimality conditions for semialgebraic convex programs, preprint 18p, 2009.

[12] Bolte, J., Daniilidis, A., Lewis, A. & Shiota, M., Clarke critical values of subanalytic Lipschitz continuous functions, *Ann. Polon. Math.* **87** (2005), 13–25 (Volume dedicated to the memory of S. Łojasiewicz).

[13] Bolte, J., Daniilidis, A., Ley, O. & Mazet, L., Characterizations of Łojasiewicz inequalities: subgradient flows, talweg, convexity, *Trans. Amer. Math. Soc.* (to appear).

[14] 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.

[15] 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.

[16] Brown, A., Functional dependence, *Trans. Amer. Math. Soc.* **38** (1935), 379–394.

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

[18] Coste, M., *An Introduction to Semialgebraic Geometry*, Instituti Editoriali e poligrafici internazionali (Università di Pisa, 1999), available electronically at `http://perso.univ-rennes1.fr/michel.coste/`

[19] Coste, M, *An Introduction to o-minimal Geometry*, Instituti Editoriali e poligrafici internazionali (Università di Pisa, 2002), available electronically at `http://perso.univ-rennes1.fr/michel.coste/`

[20] Cresson, J., Daniilidis, A & Shiota, M., On the first Integral Conjecture of René Thom, *Bull. Sci. Math.* **132** (2008), 625–631.

[21] D'Acunto, D., *Sur les courbes intégrales du champ de gradient*, 72 p., PhD Thesis, (Université de Savoie, 2001).

[22] D'Acunto, D. & Kurdyka, K., Bounds for gradient trajectories and geodesic diameter of real algebraic sets, *Bull. London Math. Soc.* **38** (2006), 951–965.

[23] Daniilidis, A., Hare, W. & Malick, J., Geometrical interpretation of the predictor-corrector type algorithms in structured optimization problems, *Optimization* **55** (2006), 481–503. (Special issue for D. Pallaschke).

[24] Daniilidis, A., Ley, O. & Sabourau, S., Asymptotic behaviour of self-contracted planar curves and gradient orbits of convex functions, preprint 19 p., (UAB 31/2008).

[25] Daniilidis, A. & Pang, J., Continuity of set-valued mappings revisited in the light of tame geometry, preprint 21 p. (CRM 2009).

[26] Van den Dries, L. & Miller, C., Geometries categories and o-minimal structures, *Duke Math. J.* **84** (1996), 497-540.

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

[28] Ioffe, A., Critical values of set-valued maps with stratifiable graphs. Extensions of Sard and Smale-Sard theorems, *Proc. Amer. Math. Soc.* **136** (2008), 3111–3119.

[29] Ioffe, A., An invitation to Tame Optimization, *SIAM J. Optim.* **19** (2009), 1894–1917.

[30] Kannai, Y., Concavifiability and constructions of concave utility functions, *J. Math. Econom.* **4** (1977), 1–56.

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

[32] Kurdyka, K. & Parusinski, A., w_f -stratification of subanalytic functions and the Lojasiewicz inequality, *C. R. Acad. Sci. Paris Sér. I Math.* **318** (1994), 129–133.

[33] Kurdyka, K., Mostowski, T. & Parusinski, A., Proof of the gradient conjecture of R. Thom, *Annals of Mathematics* **152** (2000), 763-792.

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

[35] Lewis, A.S., Active sets, nonsmoothness and sensitivity, *SIAM J. Optim.* **13** (2003), 702–725.

[36] Łojasiewicz, S., Ensembles semi-analytiques, preprint 112 p. (IHES, 1965).
Available at `http://perso.univ-rennes1.fr/michel.coste/Lojasiewicz.pdf`

[37] Łojasiewicz, S., Sur la géométrie semi- et sous-analytique, *Ann. Inst. Fourier* **43** (1993), 1575-1595.

[38] Marker, D., *Model Theory: An Introduction*, Graduate Texts in Mathematics **217** (Springer, 2002).

[39] Okochi, H., A note on asymptotic strong convergence of nonlinear contraction semigroups, *Proc. Japan Acad.* **56** (1980), 83–84.

[40] Palis, J. & De Melo, W., *Geometric theory of dynamical systems. An introduction,* (Translated from the Portuguese by A. K. Manning), Springer-Verlag, New York-Berlin, 1982.

[41] Perko, L., *Differential Equations and Dynamical Systems* (3^{rd} Edition), Texts in Applied Mathematics **7** (Springer, 2001).

[42] Rockafellar, T., & Wets, R., *Variational Analysis* (Springer, 1998).

[43] Sard, A., The measure of the critical values of differentiable maps, *Bull. Amer. Math. Soc.* **48** (1942), 883–890.

[44] Shiota, M, *Geometry of Subanalytic and Semialgebraic Sets*, Progress in Mathematics **150**, Birkhäuser, (Boston, 1997).

[45] Thom, R., Problèmes rencontrés dans mon parcours mathématique: un bilan, *IHES Publ. Math.* **70** (1989), 199–214.

[46] Torralba, D., *Convergence épigraphique et changements d'échelle en analyse variationnelle et optimisation*, Thèse, Université de Montpellier 2, 1996.