SIAM J. OPTIM.
Vol. 18, No. 2, pp. 556–572

CLARKE SUBGRADIENTS OF STRATIFIABLE FUNCTIONS*

JÉRÔME BOLTE†, ARIS DANIILIDIS‡, ADRIAN LEWIS§, AND MASAHIRO SHIOTA¶

Abstract. We establish the following result: If the graph of a lower semicontinuous real-extended-valued function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ admits a Whitney stratification (so in particular if f is a semialgebraic function), then the norm of the gradient of f at $x \in \operatorname{dom} f$ relative to the stratum containing x bounds from below all norms of Clarke subgradients of f at x. As a consequence, we obtain a Morse–Sard type of theorem as well as a nonsmooth extension of the Kurdyka–Łojasiewicz inequality for functions definable in an arbitrary o-minimal structure. It is worthwhile pointing out that, even in a smooth setting, this last result generalizes the one given in [K. Kurdyka, Ann. Inst. Fourier (Grenoble), 48 (1998), pp. 769–783] by removing the boundedness assumption on the domain of the function.

Key words. Clarke subgradient, Łojasiewicz inequality, critical point, nonsmooth analysis, Whitney stratification

AMS subject classifications. Primary, 49J52; Secondary, 26D10, 32B20

DOI. 10.1137/060670080

1. Introduction. Nonsmoothness in optimization seldom occurs in an arbitrary manner, but instead is often well-structured. Such structure can often be exploited in sensitivity analysis and algorithm convergence: Examples include "amenability," "subsmoothness," "prox-regularity" (see [32], for example), and more recently the idea of a "partly smooth" function, where a naturally arising manifold $\mathcal{M}$ contains the minimizer and the function is smooth along this manifold. We quote [24] for formal definitions, examples, and more details. In the past two decades, several researchers have tried to capture this intuitive idea in order to develop algorithms ensuring better convergence results: See, for instance, the pioneer work [23] and also [26], [9] for recent surveys.

In this work we shall be interested in a particular class of well-structured (nonsmooth) functions, namely, functions admitting a Whitney stratification (see section 2 for definitions). Since this class contains in particular the semialgebraic and the subanalytic functions (more generally, functions that are definable in some o-minimal structure over $\mathbb{R}$), the derived results can directly be applied in several concrete optimization problems involving such structures. Our central idea is to relate derivative ideas from two distinct mathematical sources: Variational analysis and differential ge-

*Received by the editors September 18, 2006; accepted for publication (in revised form) January 21, 2007; published electronically May 29, 2007.
http://www.siam.org/journals/siopt/18-2/67008.html

†Equipe Combinatoire et Optimisation (UMR 7090), Case 189, Université Pierre et Marie Curie, 4 Place Jussieu, 75252 Paris Cedex 05, France (bolte@math.jussieu.fr, http://www.ecp6.jussieu.fr/pageperso/bolte/). The research of this author was supported by the CRM, Nagoya University and ANR grant ANR-05-BLAN-60248-01 (France).

‡Departament de Matemàtiques, C1/320, Universitat Autònoma de Barcelona, E-08193 Bellaterra (Cerdanyola del Vallès), Spain (arisd@mat.uab.es, http://mat.uab.es/~arisd). The research of this author was supported by MEC grant MTM2005-08572-C03-03 (Spain) and by ANR grant ANR-05-BLAN-60248-01 (France).

§School of Operations Research and Industrial Engineering, Cornell University, 234 Rhodes Hall, Ithaca, NY 14853 (aslewis@orie.cornell.edu, http://www.orie.cornell.edu/~aslewis). The research of this author was supported in part by National Science Foundation grant DMS-0504032.

¶Department of Mathematics, Nagoya University (Furocho, Chikusa), Nagoya 464-8602, Japan (shiota@math.nagoya-u.ac.jp).

ometry. Specifically, 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 is a direct consequence of the "projection formula" given in Proposition 4 and has as corollaries a Morse–Sard type of theorem for Clarke critical points of lower semicontinuous Whitney stratifiable functions (Corollary 5(ii)) as well as a *global* nonsmooth version of the Kurdyka–Łojasiewicz inequality—which is hereby extended to unbounded domains; see Theorem 11—for lower semicontinuous definable functions (Theorem 14 and Corollary 15). Although these results seem natural, analogous ones fail for the (broader) convex-stable subdifferential (introduced and studied in [4]), unless f is assumed to be locally Lipschitz continuous; see Remark 8 and [3].

Theorems of the Morse–Sard type are central in many areas of analysis, typically describing the size of the set of ill-posed problem instances in a given class. Classical results deal with smooth functions [33], [22], but recent advances deal with a variety of nonsmooth settings [3], [13], [14], [15].

A further long-term motivation of this work is to understand the convergence of minimization algorithms. As one example, in order to treat nonconvex (and nonsmooth) minimization problems, the authors of [4] introduced an algorithm called the "gradient sampling algorithm." The idea behind this algorithm was to sample gradients of nearby points of the current iterate and to produce the next iterate by following the vector of minimum norm in the convex hull generated by the sampled negative gradients. In the case that the function is locally Lipschitz, the above method can be viewed as a kind of ε-Clarke subgradient algorithm for which both theoretical and numerical results are quite satisfactory; see [4]. The convergence of the whole sequence of iterates remains, however, an open question, and this is also the case for many classical subgradient methods for nonconvex minimization; see [19]. We hope that, just as in the smooth case, the nonsmooth Łojasiewicz inequality we develop (cf. (22) in section 4) may help in understanding the global convergence of subgradient methods.

As we outline above, we use a stratification approach to develop our results. Ioffe [14] has recently announced an extension of the work described here, leading to a remarkable and powerful Sard-type result for stratifiable multifunctions (see [15]).

2. Preliminaries. In this section we recall several definitions and results concerning nonsmooth analysis (subgradients, generalized critical points) and stratification theory. For nonsmooth analysis we refer to the comprehensive texts [5], [6], [28], [29], [32].

In what follows the vector space $\mathbb{R}^n$ is endowed with its canonical scalar product $\langle \cdot, \cdot \rangle$.

Nonsmooth analysis. Given an extended-real-valued function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ we denote its domain by $\operatorname{dom} f := \{x \in \mathbb{R}^n : f(x) < +\infty\}$, its graph by

$$\operatorname{Graph} f := \{(x, f(x)) \in \mathbb{R}^n \times \mathbb{R} : x \in \operatorname{dom} f\},$$

and its epigraph by

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

In this work we shall deal with *lower semicontinuous* functions, that is, functions for which $\operatorname{epi} f$ is a closed subset of $\mathbb{R}^n \times \mathbb{R}$. In this setting, we say that $x^* \in \mathbb{R}^n$ is a *Fréchet subgradient* of f at $x \in \operatorname{dom} f$ provided that

$$\liminf_{y \to x, y \neq x} \frac{f(y) - f(x) - \langle x^*, y - x \rangle}{\|y - x\|} \ \ge\ 0. \tag{1}$$

The set of all Fréchet subgradients of f at x is called the *Fréchet subdifferential* of f at x and is denoted by $\hat{\partial} f(x)$. If $x \notin \operatorname{dom} f$, then we set $\hat{\partial} f(x) = \emptyset$.

Let us give a geometrical interpretation of the above definition: It is well known that the gradient of a C^1 function $f : \mathbb{R}^n \to \mathbb{R}$ at $x \in \mathbb{R}^n$ can be defined geometrically as the vector $\nabla f(x) \in \mathbb{R}^n$ such that $(\nabla f(x), -1)$ is normal to the tangent space $\mathrm{T}_{(x,f(x))}\mathrm{Graph}\, f$ of (the C^1 manifold) $\mathrm{Graph}\, f$ at $(x, f(x))$, that is,

$$(\nabla f(x), -1) \perp \mathrm{T}_{(x,f(x))}\mathrm{Graph}\, f.$$

A similar interpretation can be stated for Fréchet subgradients. Let us first define the *(Fréchet) normal cone* of a subset C of $\mathbb{R}^n$ at $x \in C$ by

$$\hat{N}_C(x) = \left\{ v \in \mathbb{R}^n : \limsup_{\substack{y \to x \\ y \in C \setminus \{x\}}} \left\langle v, \frac{y - x}{||x - y||} \right\rangle \leq 0 \right\}. \tag{2}$$

Then it can be proved (see [32, Theorem 8.9], for example) that for a nonsmooth function f we have

$$x^* \in \hat{\partial} f(x) \quad \text{if and only if} \quad (x^*, -1) \in \hat{N}_{\operatorname{epi} f}(x, f(x)). \tag{3}$$

The Fréchet subdifferential extends the notion of a derivative in the sense that if f is differentiable at x, then $\hat{\partial} f(x) = \{\nabla f(x)\}$. However, it is not completely satisfactory in optimization, since $\hat{\partial} f(x)$ might be empty-valued at points of particular interest (think of the example of the function $f(x) = -||x||$, at $x = 0$). Moreover, the Fréchet subdifferential is not a closed mapping, so it is unstable computationally. For this reason we also consider (see [28], [32], for example):

(i) the *limiting* subdifferential $\partial f(x)$ of f at $x \in \operatorname{dom} f$:

$$x^* \in \partial f(x) \iff \exists (x_n, x_n^*) \subset \mathrm{Graph}\, \hat{\partial} f \ : \ \begin{cases} \lim_{n \to \infty} x_n = x, \\ \lim_{n \to \infty} f(x_n) = f(x), \\ \lim_{n \to \infty} x_n^* = x^*, \end{cases} \tag{4}$$

where $\mathrm{Graph}\, \hat{\partial} f := \{(u, u^*) : u^* \in \hat{\partial} f(u)\}$;

(ii) the *singular limiting* subdifferential $\partial^\infty f(x)$ of f at $x \in \operatorname{dom} f$:

(5)

$$y^* \in \partial^\infty f(x) \iff \exists (y_n, y_n^*) \subset \mathrm{Graph}\, \hat{\partial} f,\ \exists t_n \searrow 0^+ : \begin{cases} \lim_{n \to \infty} y_n = x, \\ \lim_{n \to \infty} f(y_n) = f(x), \\ \lim_{n \to \infty} t_n y_n^* = y^*. \end{cases}$$

When $x \notin \operatorname{dom} f$ we set $\partial f(x) = \partial^\infty f(x) = \emptyset$.

The *Clarke subdifferential* $\partial^\circ f(x)$ of f at $x \in \operatorname{dom} f$ is the central notion of this work. It can be defined in several (equivalent) ways; see [5]. The definition below (see [16, Proposition 3.3], [17, Proposition 3.4], or [30, Theorem 8.11]) is the most convenient for our purposes. (For any subset S of $\mathbb{R}^n$ we denote by $\overline{\mathrm{co}}\, S$ the closed convex hull of S.)

DEFINITION 1 (Clarke subdifferential). *The Clarke subdifferential $\partial^\circ f(x)$ of f at x is the set*

$$\partial^\circ f(x) = \begin{cases} \overline{\text{co}}\,\{\partial f(x) + \partial^\infty f(x)\} & \text{if } x \in \operatorname{dom} f, \\ \emptyset & \text{if } x \notin \operatorname{dom} f. \end{cases} \tag{6}$$

Remark 1. The construction (6) does not look very natural at first sight. However, it can be shown that an analogous to (3) formula holds also for the Clarke subdifferential, if $\hat{N}_{\text{epi}\,f}(x, f(x))$ is replaced by the Clarke normal cone, which is the closed convex hull of the limiting normal cone. The latter cone comes naturally from the Fréchet normal cone by closing its graph; and see [32, pp. 305 and 336] for details.

From the above definitions it follows directly that for all $x \in \mathbb{R}^n$, one has

$$\hat{\partial} f(x) \subset \partial f(x) \subset \partial^\circ f(x). \tag{7}$$

The elements of the limiting (respectively, Clarke) subdifferential are called limiting (respectively, Clarke) subgradients.

The notion of a Clarke critical point (respectively, critical value, asymptotic critical value) is defined as follows.

DEFINITION 2 (Clarke critical point). *We say that $x \in \mathbb{R}^n$ is a* Clarke critical point *of the function f if*

$$\partial^\circ f(x) \ni 0.$$

DEFINITION 3 ((asymptotic) Clarke critical value). (i) *We say that $\alpha \in \mathbb{R}$ is a* Clarke critical value *of f if the level set $f^{-1}(\{\alpha\})$ contains a Clarke critical point.*

(ii) *We say that $\lambda \in \mathbb{R} \cup \{\pm\infty\}$ is an* asymptotic Clarke critical value *of f, if there exists a sequence $(x_n, x_n^*)_{n\geq 1} \subset \operatorname{Graph} \partial^\circ f$ such that*

$$\begin{cases} f(x_n) \to \lambda \\ (1 + ||x_n||)\, ||x_n^*|| \to 0. \end{cases}$$

Let us make some observations concerning the above definitions.

Remark 2. (i) Both limiting and Clarke subgradients are generalizations of the usual gradient of smooth functions: Indeed, if f is C^1 around x (or more generally, strictly differentiable at x [32, Definition 9.17]), then we have

$$\partial^\circ f(x) = \partial f(x) = \{\nabla f(x)\}.$$

It should be noted that if f is only Fréchet differentiable at x, then $\partial^\circ f(x) \supset \partial f(x) \supset \{\nabla f(x)\}$, where the inclusions might be strict.

(ii) The singular limiting subdifferential should not be thought as a set of subgradients. Roughly speaking it is designed to detect "horizontal normals" to the epigraph of f. For instance, for the (nonsmooth) function $f(x) = x^{\frac{1}{3}}$ $(x \in \mathbb{R})$ we have $\partial^\infty f(0) = \mathbb{R}_+$. Note that, since the domain of the Fréchet subdifferential is dense in $\operatorname{dom} f$, we always have $\partial^\infty f(x) \ni 0$ for all $x \in \operatorname{dom} f$ (see also [32, Corollary 8.10]); therefore, this latter relation cannot be regarded as a meaningful definition of a critical point.

(iii) To illustrate the definition of the Clarke critical point (Definition 1), let us consider the example of the function $f : \mathbb{R} \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} x & \text{if } x \leq 0, \\ -\sqrt{x} & \text{if } x > 0. \end{cases}$$

Then $\hat{\partial} f(0) = \emptyset$ and $\partial f(0) = \{1\}$. However, since $\partial^\infty f(0) = \mathbb{R}_-$, it follows from (6) that $\partial^\circ f(0) = (-\infty, 1]$, so $x = 0$ is a Clarke critical point.

(iv) It follows from Definition 3 that every Clarke critical value $\alpha \in \mathbb{R}$ is also an asymptotic Clarke critical value (indeed, given $x_0 \in f^{-1}(\{\alpha\})$ with $0 \in \partial^\circ f(x_0)$, it is sufficient to take $x_n := x_0$ and $x_n^* = 0$). Note that in the case that f has a bounded domain $\operatorname{dom} f$, Definition 3(ii) can be simplified in the following way: The value $\lambda \in \mathbb{R} \cup \{\pm\infty\}$ is asymptotically critical if and only if there exists a sequence $(x_n, x_n^*)_{n\geq 1} \subset \operatorname{Graph} \partial^\circ f$ such that $f(x_n) \to \lambda$ and $x_n^* \to 0$.

Stratification results. By the term *stratification* we mean a locally finite partition of a given set into differentiable manifolds, which, roughly speaking, fit together in a regular manner. Let us give a formal definition of a C^p stratification of a set. For general facts about stratifications we quote [27]; more specific results concerning tame geometry can be found in [34], [11], [18], [10], [21].

Let X be a nonempty subset of $\mathbb{R}^n$ and p a positive integer. A C^p *stratification* $\mathcal{X} = (X_i)_{i\in I}$ of X is a locally finite partition of X into C^p submanifolds X_i of $\mathbb{R}^n$ such that for each $i \neq j$

$$\overline{X_i} \cap X_j \neq \emptyset \implies X_j \subset \overline{X_i} \setminus X_i.$$

The submanifolds X_i are called *strata* of $\mathcal{X}$. Furthermore, given a finite collection $\{A_1, \dots, A_q\}$ of subsets of X, a stratification $\mathcal{X} = (X_i)_{i\in I}$ is said to be *compatible with the collection* $\{A_1, \dots, A_q\}$ if each A_i is a locally finite union of strata X_j.

In this work we shall use a special type of stratification (called a *Whitney stratification*) for which the strata are such that their tangent spaces also "fit regularly." To give a precise meaning to this statement, let us first define the *distance* (or *gap*) of two vector subspaces V and W of $\mathbb{R}^n$ by the following standard formula:

$$D(V, W) = \max\left\{ \sup_{v\in V,\, ||v||=1} d(v, W),\ \sup_{w\in W,\, ||w||=1} d(w, V) \right\}.$$

Note that

$$\sup_{v\in V,\, ||v||=1} d(v, W) = 0 \iff V \subset W.$$

Further we say that a sequence $\{V_k\}_{k\in\mathbb{N}}$ of subspaces of $\mathbb{R}^n$ *converges* to the subspace V of $\mathbb{R}^n$ (in short, $V = \lim_{k\to+\infty} V_k$) provided

$$\lim_{k\to+\infty} D(V_k, V) = 0.$$

Notice that in this case the subspaces V_k will eventually have the same dimension (say, d); thus, the above convergence is essentially equivalent to the convergence in the Grassmannian manifold G_d^n.

A C^p stratification $\mathcal{X} = (X_i)_{i\in I}$ of X has the *Whitney-*(a) *property,* if for each $x \in \overline{X_i} \cap X_j$ (with $i \neq j$) and for each sequence $\{x_k\} \subset X_i$ we have

$$\left.\begin{array}{r} \lim_{k\to\infty} x_k = x \\ \text{and} \\ \lim_{k\to\infty} T_{x_k} X_i = \mathcal{T}, \end{array}\right\} \implies T_x X_j \subset \mathcal{T},$$

where $T_x X_j$ (respectively, $T_{x_k} X_i$) denotes the tangent space of the manifold X_j at x (respectively, of X_i at x_k). In what follows we shall use the term *Whitney stratification* to refer to a C^1 stratification with the Whitney-(a) property.

3. Projection formulas for subgradients. In this section we make precise the links between the Clarke subgradients of a lower semicontinuous function whose graph admits a Whitney stratification and the gradients of f (with respect to the strata). As a corollary we obtain a nonsmooth extension of the Morse–Sard theorem for such functions (see Corollary 5).

Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a lower semicontinuous function. We shall deal with *nonvertical* Whitney stratifications $\mathcal{S} = (S_i)_{i \in I}$ of the graph $\operatorname{Graph} f$ of f, that is, Whitney stratifications satisfying for all $i \in I$ and $u \in S_i$ the transversality condition

$$e_{n+1} \notin T_u S_i \qquad (\mathcal{H}),$$

where

$$e_{n+1} = (0, \ldots, 0, 1) \in \mathbb{R}^{n+1}.$$

Remark 3. If f is locally Lipschitz continuous, then it is easy to check that any stratification of $\operatorname{Graph} f$ is nonvertical. This might also happen for other functions (think of the nonlocally Lipschitz function $f(x) = \sqrt{|x|}$: Every stratification of $\operatorname{Graph} f$ should contain the stratum $S_0 = \{(0,0)\}$). However, the example of the function $f(x) = x^{1/3}$ shows that this is not the case for any (continuous stratifiable) function f and any stratification of its graph (consider the trivial stratification consisting of the single stratum $S = \operatorname{Graph} f$ and take $u = (0,0)$).

Let us denote by $\Pi : \mathbb{R}^{n+1} \to \mathbb{R}^n$ the canonical projection on $\mathbb{R}^n$, that is,

$$\Pi(x_1, \ldots, x_n, t) = (x_1, \ldots, x_n).$$

For each $i \in I$ we set

$$X_i = \Pi(S_i) \qquad \text{and} \qquad f_i = f|_{X_i}. \tag{8}$$

Due to the assumption $(\mathcal{H})$ (nonverticality) one has that for all $i \in I$:

(i) X_i is a C^1 submanifold of $\mathbb{R}^n$, and

(ii) $f_i : X_i \to \mathbb{R}$ is a C^1 function.

If, in addition, the function f is continuous, then it can be easily seen that:

(iii) $\mathcal{X} = (X_i)_{i \in I}$ is a Whitney stratification of $\operatorname{dom} f = \Pi(\operatorname{Graph} f)$.

Notation. In what follows, for any $x \in \operatorname{dom} f$, we shall 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_i at x with respect to the stratum $X_i, \langle \cdot, \cdot \rangle$.

PROPOSITION 4 (projection formula). *Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a lower semicontinuous function, and assume that* $\operatorname{Graph} f$ *admits a nonvertical Whitney stratification $\mathcal{S} = (S_i)_{i \in I}$. Then for all $x \in \operatorname{dom} f$ we have*

$$\operatorname{Proj}_{T_x X_x} \partial f(x) \subset \{\nabla_{\mathrm{R}} f(x)\}, \qquad \operatorname{Proj}_{T_x X_x} \partial^\infty f(x) = \{0\}, \tag{9}$$

and

$$\operatorname{Proj}_{T_x X_x} \partial^\circ f(x) \subset \{\nabla_{\mathrm{R}} f(x)\}, \tag{10}$$

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

Proof. We shall use the above notation (and in particular the notation of (8)).

Let us first describe the links between the Fréchet subdifferential $\hat{\partial} f(x)$ and the gradient of $f|_{X_x}$ at a point $x \in \operatorname{dom} f$. For any $v \in T_x X_x$ and any continuously differentiable curve $c : (-\varepsilon, \varepsilon) \to X_x$ ($\varepsilon > 0$) with $c(0) = x$ and $\dot{c}(0) = v$, the function

$$f \circ c \,(:= f_i \circ c)\; : (-\varepsilon, \varepsilon) \to \mathbb{R}$$

is continuously differentiable. In view of [32, Theorem 10.6, p. 427], we have

$$\left\{ \langle x^*, v\rangle : x^* \in \hat{\partial} f(x) \right\} \subset \left\{ \frac{d}{dt} f(c(t))\,|_{t=0} \right\}.$$

Since $\frac{d}{dt} f(c(t))|_{t=0} = \langle \nabla_R f(x), v\rangle$ it follows that

$$\operatorname{Proj}_{T_x X_x} \hat{\partial} f(x) \subset \{\nabla_R f(x)\}. \tag{11}$$

In a second stage we prove successively that

$$\operatorname{Proj}_{T_x X_x} \partial f(x) \subset \{\nabla_R f(x)\} \quad \text{and} \quad \operatorname{Proj}_{T_x X_x} \partial^\infty f(x) \subset \{0\}. \tag{12}$$

To this end, take $p \in \partial f(x)$, and let $\{x_k\} \subset \operatorname{dom} \hat{\partial} f$, $x_k^* \in \hat{\partial} f(x_k)$ be such that $(x_k, f(x_k)) \to (x, f(x))$ and $x_k^* \to p$. Due to the local finiteness property of $\mathcal{S}$, we may suppose that the sequence $\{u_k := (x_k, f(x_k))\}$ lies entirely in some stratum S_i of dimension d.

If $S_i = S_x$, then by (11) we deduce that $\operatorname{Proj}_{T_x X_x}(x_k^*) = \nabla_R f(x_k)$; thus, using the continuity of the projection and the fact that $f|_{X_x}$ is C^1 (so $\nabla_R f(x_k) \to \nabla_R f(x)$), we obtain $\operatorname{Proj}_{T_x X_x}(p) = \nabla_R f(x)$.

If $S_i \neq S_x$, then from the convergence $(x_k, f(x_k)) \to (x, f(x))$ we deduce that $\overline{S_i} \cap S_x \neq \emptyset$ (thus $d = \dim S_i > \dim S_x$). Using the compactness of the Grassmannian manifold G_d^n, we may assume that the sequence $\{T_{u_k} S_i\}$ converges to some vector space $\mathcal{T}$ of dimension d. Then the Whitney-(a) property yields that $\mathcal{T} \supset T_{(x,f(x))} S_x$. Recalling (3), for each $k \geq 1$ we have that the vector $(x_k^*, -1)$ is Fréchet normal to the epigraph $\operatorname{epi} f$ of f at u_k; hence, it is also normal (in the classical sense) to the tangent space $T_{u_k} S_i$. By a standard continuity argument the vector

$$(p, -1) = \lim_{k \to \infty} (x_k^*, -1)$$

must be normal to $\mathcal{T}$ and a fortiori to $T_{(x,f(x))} S_x$. By projecting $(p, -1)$ orthogonally on $T_x X_x + \mathbb{R}\, e_{n+1} \supset T_{(x,f(x))} S_x$, we notice that $(\operatorname{Proj}_{T_x X_x}(p), -1)$ is still normal to $T_{(x,f(x))} S_x$. We conclude that

$$\operatorname{Proj}_{T_x X_x}(p) = \nabla_R f(x); \tag{13}$$

thus, the first part of (12) follows.

Let now any $q \in \partial^\infty f(x)$. By definition there exist $\{y_k\} \subset \operatorname{dom} \hat{\partial} f$, $y_k^* \in \hat{\partial} f(y_k)$, and a positive sequence $t_k \searrow 0^+$ such that $(y_k, f(y_k)) \to (y, f(y))$ and $t_k y_k^* \to q$. As above we may assume that the sequence $\{y_k\}$ belongs to some stratum S_i and that the tangent spaces $T_{u_k} S_i = T_{(x_k, f(x_k))} S_i$ converge to some $\mathcal{T}$. Since $t_k(y_k^*, -1)$ is normal to $T_{u_k} S_i$ we can similarly deduce that $(\operatorname{Proj}_{T_x X_x}(q), 0)$ is normal to $T_{(x,f(x))} S_x$. Since $\operatorname{Proj}_{\mathbb{R}^n \times \{0\}} T_{(x,f(x))} S_x = T_x X_x$ this implies that $\partial^\infty f(x) \subset (T_x X_x)^\perp$, and the second part of (12) is proved. It now follows from (12) and Remark 2(ii) that (9) holds.

In order to conclude let us recall (Definition 1) that $\partial^\circ f(x) = \overline{\text{co}}\,(\partial f(x)+\partial^\infty f(x))$. In view of (12) any element of co $(\partial f(x) + \partial^\infty f(x))$ admits $\nabla_R f(x)$ as a projection onto $T_x X_x$. By taking the closure of the previous set we obtain (10). ∎

Remark 4. The inclusion in (10) may be strict (think of the function $f(x) = -||x||^{1/2}$ at $x = 0$, where $\partial^\circ f(0) = \emptyset$). Of course, whenever $\partial^\circ f(x)$ is nonempty (for example, if f is locally Lipschitz), under the assumptions of Proposition 4 we have

$$\text{Proj}_{T_x X_x} \partial^\circ f(x) = \{\nabla_R f(x)\}.$$

COROLLARY 5. *Assume that f is lower semicontinuous and admits a nonvertical C^p-Whitney stratification. Then:*

(i) *for all $x \in \text{dom}\,\partial^\circ f$ we have*

$$||\nabla_{\mathrm{R}} f(x)|| \;\leq\; ||x^*|| \qquad \text{for all } x^* \in \partial^\circ f(x). \tag{14}$$

(ii) (Morse–Sard theorem) *If $p \geq n$, then the set of Clarke critical values of f has Lebesgue measure zero.*

Proof. Assertion (i) is a direct consequence of (10) of Proposition 4. To prove (ii), set $C := [\partial^\circ f]^{-1}(\{0\}) = \{x \in \mathbb{R}^n : \partial^\circ f(x) \ni 0\}$. Since the set of strata is at most countable, the restrictions of f to each of those yield a countable family $\{f_n\}_{n\in\mathbb{N}}$ of C^p functions. In view of (14), we have that $C \subset \cup_{n\in\mathbb{N}}(\nabla f_n)^{-1}(0)$. The result follows by applying to each C^p-function f_n the classical Morse–Sard theorem [33]. ∎

As we see in the next section, several important classes of lower semicontinuous functions satisfy the assumptions (thus also the conclusions) of Proposition 4 and of Corollary 5.

4. Kurdyka–Łojasiewicz inequalities for o-minimal functions. Let us recall briefly a few definitions concerning o-minimal structures (see, for instance, Coste [7], van den Dries and Miller [11], Ta Lê Loi [35], and references therein).

DEFINITION 6 (o-minimal structure). *An o-minimal structure on* $(\mathbb{R}, +, .)$ *is a sequence of Boolean algebras $\mathcal{O}_n$ of "definable" subsets of $\mathbb{R}^n$ such that for each $n \in \mathbb{N}$:*

(i) if A belongs to $\mathcal{O}_n$, then $A \times \mathbb{R}$ and $\mathbb{R} \times A$ belong to $\mathcal{O}_{n+1}$;

(ii) if $\Pi : \mathbb{R}^{n+1} \to \mathbb{R}^n$ is the canonical projection onto $\mathbb{R}^n$, then for any A in $\mathcal{O}_{n+1}$ the set $\Pi(A)$ belongs to $\mathcal{O}_n$;

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

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

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

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

DEFINITION 7 (definable function). *Given an o-minimal structure $\mathcal{O}$ (over* $(\mathbb{R}, +, .)$*), a function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is said to be* definable *in $\mathcal{O}$ if its graph belongs to $\mathcal{O}_{n+1}$.*

Remark 5 (examples). At first sight, o-minimal structures might appear artificial in optimization. The following fundamental properties (see [11] for the details) might convince the reader that this is not the case.

(i) (Tarski–Seidenberg) The collection of *semialgebraic sets* is an o-minimal structure. Recall that semialgebraic sets are Boolean combinations of sets of the form

$$\{x \in \mathbb{R}^n : \, p(x) = 0, \, q_1(x) < 0, \ldots, \, q_m(x) < 0\},$$

where p and q_i's are polynomial functions on $\mathbb{R}^n$.

(ii) (Gabrielov) There exists an o-minimal structure that contains the sets of the form

$$\{(x,t) \in [-1,1]^n \times \mathbb{R} : f(x) = t\},$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is real analytic around $[-1,1]^n$.

(iii) (Wilkie) There exists an o-minimal structure that contains simultaneously the graph of the exponential function $\mathbb{R} \ni x \mapsto \exp x$ and all semialgebraic sets (respectively, all sets of the structure defined in (ii)).

We insist on the fact that these results are crucial foundation blocks on which o-minimal geometry rests.

Let us finally recall the following elementary but important result: The composition of mappings that are definable in some o-minimal structure remains in the same structure [11, section 2.1]. This is also true for the sum, the inf-convolution, and several other classical operations of analysis involving a finite number of definable objects. Another prominent fact about definable sets is that they admit, for each $k \geq 1$, a C^k–Whitney stratification with finitely many strata (see, for instance, [11, Result 4.8, p. 510]). This remarkable stability, combined with new techniques of finite-dimensional optimization, offers a large field of investigation. Several works have already been developed in this spirit; see, for instance, [1], [3], [12].

Given any o-minimal structure $\mathcal{O}$ and any lower semicontinuous definable function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, the assumptions of Proposition 4 are satisfied. More precisely, we have the following result.

LEMMA 8. *Let $\mathcal{O}$ be an o-minimal structure, $\mathcal{B} := \{B_1, \dots, B_q\}$ be a collection of definable subsets of $\mathbb{R}^n$, and $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a definable lower semicontinuous function. Then for any $p \geq 1$, there exists a nonvertical definable C^p-Whitney stratification $\{S_1, \dots, S_\ell\}$ of* Graph f *yielding (by projecting each stratum $S_i \subset \mathbb{R}^{n+1}$ onto $\mathbb{R}^n$) a C^p-Whitney stratification $\{X_1, \dots, X_\ell\}$ of* dom f *compatible with $\mathcal{B}$.*

Proof. By transforming, using diffeomorphisms preserving verticality, $\mathbb{R}^n$ to $D := \{x \in \mathbb{R}^n : ||x|| < 1\}$ and $\mathbb{R}$ to $(-1,1)$, we may assume without loss of generality that f is defined in $D := \{x \in \mathbb{R}^n : ||x|| < 1\}$ with values in $(-1,1)$. Set $X = \operatorname{Graph} f$ and $A_i = B_i \times (-1,1)$ for $i \in \{1, \dots, q\}$, and let $\pi : X \to D$ denote the restriction to Graph f of the canonical projection of $D \times (-1,1)$ to D. The lemma follows from the canonical stratification of the mapping π according to [34, II.1.17]. □

COROLLARY 9 (Morse–Sard theorem for definable functions). *Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a lower semicontinuous definable function and $p \geq 1$. Then there exists a finite definable C^p-Whitney stratification $\mathcal{X} = (X_i)_{i \in I}$ of* dom f *such that for all $x \in \operatorname{dom} f$*

$$\operatorname{Proj}_{T_x X_x} \partial^\circ f(x) \subset \{\, \nabla_{\mathrm{R}} f(x) \,\}. \tag{15}$$

As a consequence:

(i) *for all $x \in \operatorname{dom} \partial^\circ f$ and $x^* \in \partial^\circ f(x)$, we have $||\nabla_{\mathrm{R}} f(x)|| \leq ||x^*||$;*

(ii) *the set of Clarke critical values of f is finite*;

(iii) *the set of asymptotic Clarke critical values of f is finite.*

Proof. Assertion (i) is a direct consequence of (15). This projection formula follows directly by combining Lemma 8 with Proposition 4. To prove (iii), let f_i be the restriction of f to the stratum X_i. Then assertion (i), together with the fact that the number of strata is finite, implies that the set of the asymptotic Clarke critical values of f is the union (over the finite set I) of the asymptotic critical values of each (definable C^1) function f_i. Thus the result follows from [8, Remarque 3.1.5]. Assertion (ii) follows directly from (iii) (cf. Remark 2(iii)). $\square$

Remark 6. The fact that the set of the asymptotic critical values of a definable differentiable function f is finite has been established in [8, Théorème 3.1.4] (see also [20, Theorem 3.1] for the case that the domain of f is bounded). In [22, Proposition 2] a more general result (concerning functions taking values in $\mathbb{R}^k$) has been established in the semialgebraic case.

We shall now give another application of Proposition 4, namely, a nonsmooth version of the classical Kurdyka–Lojasiewicz inequality ([20, Theorem 1]). Before we proceed, we shall improve the latter in a way that allows us to deal directly with unbounded domains. To this end, we shall need the following proposition.

PROPOSITION 10 (uniform boundedness). *Let $I = [a, +\infty)$ for some $a \in \mathbb{R}$, and let $\mathcal{V}$ be a definable neighborhood of $\{0\} \times I$ in $\mathbb{R}_+ \times I$ and $\phi : \mathcal{V} \to \mathbb{R}_+$ a definable function, continuous throughout $\{0\} \times I$, satisfying $\phi(0, s) = 0$ for all $s \in I$. Then there exist $\varepsilon_0 > 0$ and continuous definable functions $\chi : I \to (0, \varepsilon_0)$ and $\psi : (0, \varepsilon_0) \to [0, +\infty)$ such that ψ is C^1 on $(0, \varepsilon_0)$, $\psi(0) = 0$, and*

$$\psi(t) \geq \phi(t, s) \quad \textit{for all } s \in I, \qquad t \in (0, \chi(s)).$$

Proof. We can clearly assume that $a = 0$. Since $\mathcal{V}$ is a definable neighborhood of $\{0\} \times I$, we may assume there exists a continuous definable function $g : I \to (0, +\infty)$ such that $\{(t, s) \in \mathbb{R}_+ \times I : t \leq g(s)\} \subset \mathcal{V}$. Set

$$\delta(s) := \sup \left\{ \delta \in (0, g(s)) : \quad \phi(t, s) \leq \frac{1}{s+1} \quad \forall t \in [0, \delta) \right\}, \tag{16}$$

and note that $\delta(s)$, being definable, has a finite number of points of discontinuity. Since ϕ is continuous on $\{0\} \times I$ and $\phi(0, s) = 0$ for all $s \in I$, we infer that $\liminf_{s \to \bar{s}} \delta(s) > 0$ for all $\bar{s} \in I$. We deduce that there exists a continuous decreasing and definable function $\chi : I \to (0, +\infty)$ satisfying $\chi(s) \leq \delta(s)$ for all $s \in I$. Set $\varepsilon_0 = \sup \chi(I) = \chi(0) > 0$, and consider the definable function

$$\psi(t) = \max_{s \in [0, \chi^{-1}(t)]} \phi(t, s) \quad \text{for all } t \in [0, \varepsilon_0).$$

By the monotonicity lemma [7, Theorem 2.1] we conclude that ψ is C^1 on $(0, \beta)$ for some $\beta \leq \varepsilon_0$. Truncating χ if necessary (by defining $\tilde{\chi}(s) := \min\{\beta, \chi(s)\}$), we see that there is no loss of generality to assume $\beta = \varepsilon_0$. Note that $\psi(0) = 0$. Let us show that ψ is also continuous at $t = 0$. Let us assume, towards a contradiction, that there exists a sequence $t_n \searrow 0^+$ satisfying $\psi(t_n) > c > 0$. Then for every $n \in \mathbb{N}$ there exists $s_n \in [0, \chi^{-1}(t_n)]$ such that $\phi(t_n, s_n) > c > 0$. If $\{s_n\} \to +\infty$, then since $\delta(s_n) \geq \chi(s_n)$ we would deduce from (16) that $(s_n + 1)^{-1} \geq \phi(t_n, s_n) > c$, which is impossible for large values of n. Thus $\{s_n\}$ is bounded and has a convergent subsequence to some $s \in I$. Using the continuity of ϕ at $(0, s)$ and the fact that $\phi(0, s) = 0$, the contradiction follows. One can easily check that the definable functions ψ and χ satisfy the conclusion of the proposition. $\square$

We now provide the following extension of the Kurdyka–Łojasiewicz inequality ([20, Theorem 1]) for unbounded sets in the smooth case.

THEOREM 11 (Kurdyka–Łojasiewicz inequality). *Let U be a nonempty definable submanifold of $\mathbb{R}^n$ (not necessarily bounded) and $f : U \to \mathbb{R}_+$ be a definable differentiable function. Then there exist a continuous definable function $\psi : [0, \varepsilon_0) \to \mathbb{R}_+$ satisfying $\psi(0) = 0$ and being C^1 on $(0, \varepsilon_0)$ and a continuous definable function $\chi : \mathbb{R}_+ \to (0, \varepsilon_0)$ such that*

$$||\nabla(\psi \circ f)(x)|| \geq 1 \quad \text{for all } 0 < f(x) \leq \chi(||x||). \tag{17}$$

Proof. With no loss of generality we can assume that f is not identically equal to 0 on U.

For each $(t, s) \in (0, +\infty) \times \mathbb{R}_+$ we set

$$F(t, s) := f^{-1}(t) \cap B(0, s) \subset U \qquad \text{and} \qquad m_f(t, s) = \inf\ \{||\nabla f(x)|| : x \in F(t, s)\}. \tag{18}$$

Note that $m_f(t, s) \equiv +\infty$ whenever $F(t, s)$ is empty. If $f^{-1}(0) = \emptyset$, then for every $s \geq 0$ there exists $\delta > 0$ such that for all $t \in (0, \delta)$ we have $F(t, s) = \emptyset$. Thus, the definable function

$$s \mapsto \delta(s) := \sup\{\delta > 0 : F(t, s) = \emptyset \ \ \forall t \in (0, \delta]\} < +\infty$$

is positive (cf. continuity of f), decreasing (since $F(t, s_1) \subset F(t, s_2)$ for $s_1 \leq s_2$), and continuous on $(\bar{s}, +\infty)$ for some $\bar{s} > 0$ (cf. monotonicity lemma [7, Theorem 2.1]). In this case (17) follows trivially by considering the continuous function

$$\chi(s) = \begin{cases} \delta(s)/2 & \text{if } s \geq \bar{s}, \\ \delta(\bar{s})/2 & \text{if } s \leq \bar{s}, \end{cases}$$

and any continuous definable function ψ.

Thus there is no loss of generality to assume that there exists $s_0 \geq 0$ and a decreasing continuous definable function $\rho : [s_0, +\infty) \to (0, +\infty)$ such that $F(t, s) \neq \emptyset$ for all $t \in [0, \rho(s)]$ and all $s \geq s_0$. It follows that for all $s \geq s_0$ and $t \in [0, \rho(s)]$ we have $m_f(t, s) \in \mathbb{R}_+$ and (since $\arg\min f = \{0\}$) $m_f(0, s) = 0$. Using an argument of Kurdyka ([20, Claim, p. 777]) we deduce that the function $t \mapsto m_f(t, s)$ is not identically 0 near the origin, and we set for all $s \geq s_0$

$$g(s) = \sup\ \{\, t_0 \in (0, \rho(s)) :\ m_f(t, s) > 0 \ \forall t \in (0, t_0] \,\} \in (0, +\infty).$$

Then g is decreasing, positive, definable, and thus continuous on $[s_1, +\infty)$ for some $s_1 \geq s_0$. Set $D = \{(t, s) \in \mathbb{R}_+ \times [s_1, +\infty) : t \leq g(s)\}$, and consider the following definable point-to-set mapping $M : D \rightrightarrows U \subset \mathbb{R}^n$, with

$$M(t, s) := \{x \in F(t, s) : ||\nabla f(x)|| \leq \ 2\ m_f(t, s)\}.$$

Using the definable selection lemma (cf. [7, Theorem 3.1]), we obtain a definable mapping $\gamma : D \to \mathbb{R}^n$ such that $\gamma(t, s) \in M(t, s)$ for all $(t, s) \in D$. Note that for each s fixed, the function $(0, g(s)) \ni t \mapsto \gamma(t, s)$ is absolutely continuous and $\frac{\partial}{\partial t}\gamma_i(\cdot, s)$ changes sign only a finite number of times on D for all $i \in \{1, \ldots, n\}$. We set

$$\phi(t, s) = \int_0^t \max_{i \in \{1, \ldots, n\}} \left| \frac{\partial}{\partial t}\gamma_i(\tau, s) \right| \, d\tau$$

for all $(t, s) \in D$. Applying the monotonicity lemma we obtain the integrability of the function

$$\tau \mapsto \max_{i=1,\ldots,n} \left| \frac{\partial}{\partial t}\gamma_i(\tau, s) \right|.$$

Using routine arguments it is easily seen that ϕ is actually definable on D. Moreover, $\phi(t,s) > 0$ whenever $t > 0$ (else the curve $\gamma(\cdot, s)$ would be stationary, which is not possible since $f(\gamma(t,s)) = t$). Note also that $\phi(0,s) = 0$ and $\lim_{t \searrow 0^+} \phi(t,s) = 0$. Considering a stratification of ϕ we deduce that there exists $a \geq s_1$ and a definable neighborhood $\mathcal{V}$ of $\{0\} \times [a, +\infty)$ in D where ϕ is (jointly) continuous. Applying Proposition 10, we obtain $\varepsilon_0 > 0$, a continuous definable function $\chi : [a, +\infty) \to (0, \varepsilon_0)$, and a continuous definable function $\psi : [0, \varepsilon_0) \to \mathbb{R}$, with $\psi(0) = 0$, such that ψ is C^1 on $(0, \varepsilon_0)$ and $\psi(t) \geq \phi(t,s)$ for all $t \in [0, \chi(s)]$.

Fix $s \geq a$. Since $\psi(t) \geq \phi(t,s)$ for $t \in [0, \chi(s)]$ and $\psi(0) = \phi(0,s)$, it follows (see [2, Lemma 1(i)], for example) that for all $t > 0$ sufficiently small

$$\psi'(t) \geq \frac{\partial}{\partial t}\phi(t,s) > 0. \tag{19}$$

For each $s \in [a, +\infty)$ let us define $\varepsilon(s)$ to be the supremum of all $\varepsilon \in (0, \varepsilon_0)$ such that (19) holds true in the interval $(0, \varepsilon)$. It follows that $s \mapsto \varepsilon(s)$ is a positive definable function and is thus continuous on $[b, +\infty)$ for some $b \geq a$. Let us define

$$\tilde{\chi}(s) = \begin{cases} \min\{\chi(s), \varepsilon(s)\} & \text{if } s \geq b, \\ \min\{\chi(b), \varepsilon(b)\} & \text{if } s \in [0, b]. \end{cases}$$

We shall now show that (17) holds for $\tilde{\psi} = (\frac{1}{2\sqrt{n}})\psi$ and for $\tilde{\chi} : \mathbb{R}_+ \to (0, \varepsilon_0)$. Indeed, let $x \in U$ be such that $0 < f(x) \leq \tilde{\chi}(||x||)$ (hence $\nabla f(x) \neq 0$). Set $t = f(x)$ and $s = \max\{||x||, b\}$. Using the definition of γ we obtain

$$||\nabla(\psi \circ f)(x)|| = \psi'(t)||\nabla f(x)|| \geq \frac{1}{2}\psi'(t)||\nabla f(\gamma(t,s))||. \tag{20}$$

On the other hand, since $f(\gamma(t,s)) = t$, we have

$$\frac{d}{dt}f(\gamma(t,s)) = \left\langle \frac{\partial}{\partial t}\gamma(t,s), \nabla f(\gamma(t,s)) \right\rangle = 1$$

for all $(t,s) \in D$; hence

$$\sqrt{n} \max_{i=1,\dots,n} \left| \frac{\partial}{\partial t}\gamma_i(\cdot, s) \right| \; ||\nabla f(\gamma(t,s)|| \geq \left\| \frac{\partial}{\partial t}\gamma(t,s) \right\| \; ||\nabla f(\gamma(t,s)|| \geq 1,$$

and thus

$$||\nabla f(\gamma(t,s))|| \geq \left[\sqrt{n} \max_{i=1,\dots,n} \left| \frac{\partial}{\partial t}\gamma_i(\cdot, s) \right| \right]^{-1} = \left[\sqrt{n} \frac{\partial}{\partial t}\phi(t,s) \right]^{-1}. \tag{21}$$

Since $f(x) \leq \tilde{\chi}(||x||) \leq \varepsilon(s)$, by combining (19), (20), and (21) we finally obtain that

$$||\nabla(\psi \circ f)(x)|| \geq \frac{1}{2\sqrt{n}}\psi'(t) \left[\frac{\partial}{\partial t}\phi(t,s) \right]^{-1} \geq \frac{1}{2\sqrt{n}};$$

that is, (17) holds for $\tilde{\psi} = (\frac{1}{2\sqrt{n}})^{-1}\psi$. $\square$

Remark 7. If in the statement of Theorem 11 the definable set U is not open, then ∇f is understood as the Riemannian gradient of f on U.

We easily obtain the following corollaries.

COROLLARY 12. *Let $f : U \to \mathbb{R}$ be a definable differentiable function, where U is a definable submanifold of $\mathbb{R}^n$ (not necessarily bounded). 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 a relatively open neighborhood V of $f^{-1}(0)$ in U such that*

$$||\nabla(\psi \circ |f|)(x)|| \geq 1$$

for all x in $V \setminus f^{-1}(0)$.

Proof. Let us first assume that f is nonnegative. The result holds trivially if $f^{-1}(0) = \emptyset$, so let us assume $f^{-1}(0) \neq \emptyset$. Take ψ and χ as in Theorem 11, and let $x \in f^{-1}(0)$. It suffices to show that the inequality holds on a ball around x. Take $r \in (0, \varepsilon_0)$ such that $\chi(||x||) > r$. Since χ and f are continuous, there exists $\delta > 0$ such that $y \in B(x, \delta) \cap U$ implies $\chi(||y||) > r > f(y)$. Applying Theorem 11, we conclude that for all $y \in B(x, \delta) \cap U$ inequality (17) holds. When f takes its values in $\mathbb{R}$ (not necessarily in $\mathbb{R}_+$), the conclusion follows easily by considering the submanifolds $\{x \in U : f(x) > 0\}$, $\{x \in U : f(x) < 0\}$ and by applying the monotonicity Lemma. ◻

COROLLARY 13. *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 connected components of $(\nabla f)^{-1}(\{0\})$ and by $c_1, \ldots, c_m$ the corresponding critical values. 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 $x \in V_i \setminus C_i$ we have*

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

Proof. Note that $(\nabla f)^{-1}(\{0\}) \subset \cup_{i=1}^m f^{-1}(c_i)$. For each $i \in \{1, \ldots, m\}$ we apply Corollary 12 to the function $f_i := f - c_i$ on U to obtain a relatively open neighborhood V_i of C_i and $\psi_i : [0, \varepsilon_i) \to \mathbb{R}_+$ such that for all $x \in V_i \setminus f^{-1}(c_i)$

$$||\nabla[\psi_i \circ |f - c_i|](x)|| \geq 1.$$

Set $\varepsilon_0 = \min\{\varepsilon_i : i \in \{1, \ldots, m\}\}$. Since ψ_i are definable functions, shrinking ε_0 if necessary, we may assume (cf. the monotonicity lemma) that $\psi'_{i_0}(t) \geq \psi'_i(t)$ for all $t \in (0, \varepsilon_0)$ and all $i \in \{1, \ldots, m\}$. The conclusion follows by setting $\psi := \psi_{i_0}$ on $[0, \varepsilon_0)$. ◻

We shall now use Corollary 9 to extend Theorem 11 to a nonsmooth setting.

THEOREM 14 (nonsmooth Kurdyka–Łojasiewicz inequality). *Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a lower semicontinuous definable function. There exist $\rho > 0$, a strictly increasing continuous definable function $\psi : [0, \rho) \to (0, +\infty)$ which is C^1 on $(0, \rho)$, with $\psi(0) = 0$, and a continuous definable function $\chi : \mathbb{R}_+ \to (0, \rho)$ such that*

$$||x^*|| \geq \frac{1}{\psi'(|f(x)|)}, \tag{22}$$

whenever $0 < |f(x)| \leq \chi(||x||)$ and $x^* \in \partial^\circ f(x)$.

Proof. Set $U_1 = \{x \in \operatorname{dom} f : f(x) > 0\}$ and $U_2 = \{x \in \operatorname{dom} f : f(x) < 0\}$, and let $X_1, \ldots, X_l$ be a finite definable stratification of $\operatorname{dom} f$ compatible with the (definable) sets U_1 and U_2 such that the definable sets $S_i = \{(x, f(x)) : x \in X_i\}$ are the strata of a nonvertical definable C^p-Whitney stratification of $\operatorname{Graph} f$ (cf. Lemma 8). For each $i \in \{1, \ldots, l\}$ such that $X_i \subset U_1$ we consider the positive

C^1 function $f_i := f|_{X_i}$ on the definable manifold X_i (thus for $x \in X_i$ we have $\nabla f_i(x) = \nabla_R f(x)$ and $f_i(x) = f(x)$), and we apply Theorem 11 to obtain $\varepsilon_i > 0$, a continuous definable function $\chi_i : \mathbb{R}_+ \to (0, \varepsilon_i)$, and a strictly increasing definable C^1-function $\psi_i : (0, \varepsilon_i) \to (0, +\infty)$ such that for all $x \in f^{-1}(0, \chi_i(||x||))$ we have $||\nabla_R f(x)|| \geq [\psi_i'(f(x))]^{-1}$. Similarly, for each $j \in \{1, \dots, l\}$ such that $X_j \subset U_2$ we consider the positive C^1-function $f_j := -f|_{X_i}$ (note that for $x \in X_j$ we have $\nabla f_j(x) = -\nabla_R f(x)$ and $f_j(x) = -f(x)$) to obtain as before a definable function $\chi_j : \mathbb{R}_+ \to (0, \varepsilon_j)$ and a strictly increasing definable C^1-function $\psi_j : (0, \varepsilon_j) \to (0, +\infty)$ such that for all $x \in f^{-1}(0, \chi_i(||x||))$ we have $||\nabla_R f(x)|| \geq [\psi_j'(-f(x))]^{-1}$. Thus for all $i \in \{1, \dots, l\}$ there exist a definable function $\chi_i : \mathbb{R}_+ \to (0, \varepsilon_i)$ and a strictly increasing definable C^1-function $\psi_i : (0, \varepsilon_i) \to \mathbb{R}$ such that

$$||\nabla_R f(x)|| \geq \frac{1}{\psi_i'(|f(x)|)} \qquad \text{for all } x \in f^{-1}(0, \chi_i(||x||)).$$

Set $\chi = \min \chi_i$, $\rho = \min \varepsilon_i$, and let $i_1, i_2 \in \{1, \dots, l\}$. By the monotonicity theorem for definable functions of one variable (see [20, Lemma 2], for example), the definable function

$$(0, \rho) \ni r \quad \mapsto \quad 1/\psi_{i_1}'(r) - 1/\psi_{i_2}'(r)$$

has a constant sign in a neighborhood of 0. Repeating the argument for all couples i_1, i_2 and shrinking ρ if necessary, we obtain the existence of a strictly increasing, positive, definable function $\psi = \psi_{i_0}$ on $(0, \rho)$ of class C^1 that satisfies $1/\psi' \leq 1/\psi_i'$ on $(0, \rho)$ for all $i \in \{1, \dots, l\}$. Evoking Corollary 9(i), we obtain

$$||x^*|| \geq ||\nabla_R f(x)|| \geq \frac{1}{\psi'(|f(x))|},$$

whenever $x \in |f|^{-1}(0, \chi(||x||))$ and $x^* \in \partial^\circ f(x)$. Since ψ is definable and bounded from below, it can be extended continuously to $[0, \rho)$. By eventually adding a constant, we can also assume $\psi(0) = 0$. $\square$

In a similar way to Corollary 13 we obtain the following result.

COROLLARY 15. *Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a lower semicontinuous definable function. Let us denote by $C_1, \dots, C_m$ the connected components of* $(\partial^\circ f)^{-1}(\{0\})$ *and by $c_1, \dots, c_m$ the corresponding critical values* (*cf. Corollary* 9(ii)). *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* dom f *for each $i \in \{1, \dots, m\}$ such that for all $x \in V_i$ we have*

$$||x^*|| \geq \frac{1}{\psi'(|f(x) - c_i|)}, \tag{23}$$

whenever $0 < |f(x) - c_i| \leq \chi(||x||)$ *and* $x^* \in \partial^\circ f(x)$.

The assumption that the function f is definable is important for the validity of (22). It implies in particular that the connected components of the set of the Clarke critical points of f lie in the same level set of f (cf. Corollary 9(ii)). Let us present some examples of C^1-functions for which (22) is not true.

Example 1. (i) Consider the function $f : \mathbb{R} \to \mathbb{R}$, with

$$f(x) = \begin{cases} x^2 \sin \frac{1}{x} & \text{if } x \neq 0, \\ 0 & \text{if } x = 0. \end{cases}$$

Then the set $S = \{x \in \mathbb{R} : f'(x) = 0\}$ meets infinitely many level sets around 0. Consequently, (22) is not fulfilled since the critical value 0 is not isolated. Note also that f provides an example of a nondefinable function whose graph admits a Whitney stratification (in particular f satisfies the conclusion of Proposition 4).

(ii) A nontrivial example is proposed in [31, p. 14], where a C^∞ "Mexican-hat" function has been defined. An example of a similar nature has been given in [1] and will be described below: Let f be defined in polar coordinate on $\mathbb{R}^2$ by

$$f(r,\theta) = \begin{cases} \exp(-\frac{1}{1-r^2})\,[1 - \frac{4r^4}{4r^4+(1-r^2)^4}\,\sin(\theta - \frac{1}{1-r^2})] & \text{if } r \le 1, \\ 0 & \text{if } r > 1. \end{cases}$$

The function f does not satisfy the Kurdyka–Łojasiewicz inequality for the critical value 0; i.e., one cannot find a strictly increasing C^1-function $\psi : (0,\rho) \to (0,+\infty)$, with $\rho > 0$, such that

$$||\nabla(\psi \circ f)(x)|| \ge 1$$

for small positive values of $f(x)$. To see this, let us notice that the proof of [20, Theorem 2] shows that for *any* C^1-function f (not necessarily definable) that satisfies the Kurdyka–Łojasiewicz inequality, the bounded trajectories of the gradient system

$$\dot{x}(t) + \nabla f(x(t)) = 0$$

have a bounded length. However, in the present example, taking as the initial condition $r_0 \in (0,1)$ and θ_0 such that $\theta_0(1-r_0)^2 = 1$, the gradient trajectory $\dot{x}(t) = -\nabla f(x(t))$ must comply with

$$\theta(t) = \frac{1}{1-r(t)^2},$$

where $r(t) \nearrow 1^-$ as $t \to +\infty$ (see [1] for details). The total length of the above curve is obviously infinite, which shows that the Kurdyka–Łojasiewicz inequality (for the critical value 0) does not hold.

Let us finally give an easy consequence of Theorem 14 for the case of subanalytic functions [25].

Corollary 16 (subgradient inequality). *Assume that* $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ *is a lower semicontinuous globally subanalytic function and* $f(x_0) = 0$. *There exist* $\rho > 0$ *and a continuous definable function* $\chi : \mathbb{R}_+ \to (0,+\infty)$ *such that*

$$|f(x)|^\theta \;\le\; \rho\,||x^*||,$$

whenever $0 < |f(x)| \le \chi(||x||)$ *and* $x^* \in \partial^\circ f(x)$.

Proof. In the case that f is globally subanalytic, one can apply [20, Theorem LI] to deduce that the continuous function ψ of Theorem 14 can be taken of the form $\psi(s) = s^{1-\theta}$, with $\theta \in (0,1)$. □

Remark 8. Corollary 9(ii) (and a fortiori Corollary 16) extends [3, Theorem 7] to the lower semicontinuous case. We also remark that the conclusions of Theorem 14 and of Corollary 16 remain valid for any notion of subdifferential that is included in the Clarke subdifferential and thus, in particular, in view of (7), for the Fréchet and the limiting subdifferential. However, let us point out that this is not the case for broader notions of subdifferentials, as, for example, the *convex-stable* subdifferential introduced and studied in [4]. It is known that the convex-stable subdifferential coincides with the Clarke subdifferential whenever the function f is locally Lipschitz continuous, but it is strictly larger in general, creating more critical points. In particular, [3, section 4] constructs an example of a subanalytic continuous function on $\mathbb{R}^3$ that is strictly increasing in a segment lying in the set of its broadly critical points (that is, critical in the sense of the convex-stable subdifferential). Consequently, Theorem 14 and Corollary 16 do not hold for this subdifferential.

Acknowledgments. A part of this work has been done during a visit of the first author at the Autonomous University of Barcelona (July 2005) and at the University of Nagoya (June 2006). The first author thanks J.-P. Dedieu and the second author K. Kurdyka, S. Simon, and T. Lachand-Robert for useful discussions. The first author thanks the C.R.M. (Barcelona) and the University of Nagoya for financial support.

REFERENCES

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

[2] J. Bolte, A. Daniilidis, and A. Lewis, *Tame functions are semismooth*, Math. Program., (to appear).

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

[4] J. V. Burke, A. S. Lewis, and M. L. Overton, *A robust gradient sampling algorithm for nonsmooth, nonconvex optimization*, SIAM J. Optim., 15 (2005), pp. 751–779.

[5] F. H. Clarke, *Optimization and Nonsmooth Analysis*, John Wiley and Sons, New York, 1983 (republished in Classics Appl. Math. 5, SIAM, 1990, p. 308).

[6] F. H. Clarke, Y. Ledyaev, R. Stern, and P. Wolenski, *Nonsmooth Analysis and Control Theory*, Grad. Texts in Math. 178, Springer, New York, 1998.

[7] M. Coste, *An Introduction to o-minimal Geometry*, RAAG Notes, Institut de Recherche Mathématiques de Rennes, 1999.

[8] D. D'Acunto, *Sur les courbes intégrales du champs de gradient*, Thèse de Doctorat, Université de Savoie, Chambery, France, 2001.

[9] A. Daniilidis, W. Hare, and J. Malick, *Geometrical interpretation of the predictor-corrector type algorithms in structured optimization problems*, Optimization, 55 (2006), pp. 481–503.

[10] Z. Denkowska and K. Wachta, *Une construction de la stratification sous-analytique avec la condition* (w), Bull. Pol. Acad. Sci. Math., 35 (1987), pp. 401–405.

[11] L. van den Dries and C. Miller, *Geometric categories and o-minimal structures*, Duke Math. J., 84 (1996), pp. 497–540.

[12] L. M. Graña Drummond and Y. Peterzil, *The central path in smooth convex semidefinite programs*, Optimization, 51 (2002), pp. 207–233.

[13] A. Ioffe, *A Sard Theorem for Tame Nonsmooth Functions*, J. Math. Anal. Appl., to appear.

[14] A. Ioffe, *Tame optimization: State of the art and perspectives*, Plenary talk at The International Conference on Nonlinear Programming with Applications, Shanghai, 2006.

[15] A. Ioffe, *Critical values of set-valued maps with stratifiable graphs*, in Extensions of Sard and Smale-Sard Theorems, Proc. Amer. Math. Soc., to appear.

[16] A. Ioffe, *Approximate subdifferentials and applications* II, Mathematika, 33 (1986), pp. 111–128.

[17] A. Ioffe, *Approximate subdifferentials and applications* III. *The metric theory*, Mathematika, 36 (1989), pp. 1–38.

[18] V. Kaloshin, *A geometric proof of the existence of Whitney stratifications*, Mosc. Math. J., 5 (2005), pp. 125–133.

[19] K. C. Kiwiel, *Methods of Descent for Nondifferentiable Optimization*, Lecture Notes in Math. 1133, Springer-Verlag, Berlin, 1985.

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

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

[22] K. Kurdyka, P. Orro, and S. Simon, *Semialgebraic Sard theorem for generalized critical values*, J. Differential Geom., 56 (2000), pp. 67–92.

[23] C. Lemarechal, F. Oustry, and C. Sagastizabal, *The U-Lagrangian of a convex function*, Trans. Amer. Math. Soc., 352 (2000), pp. 711–729.

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

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

[26] J. Malick and S. Miller, *Newton methods for nonsmooth convex minimization: Connection among U-Lagrangian, Riemannian Newton and SQP methods*, Math. Program., 104 (2005), pp. 609–633.

[27] J. Mather, *Notes in Topological Stability*, lecture notes, Harvard University, 1970.

[28] B. Mordukhovich, *Variational Analysis and Generalized Differentiation. Basic Theory*, Vol. I, Grundlehren Math. Wiss. 330, Springer-Verlag, Berlin, 2006.

[29] B. Mordukhovich, *Variational Analysis and Generalized Differentiation. Applications*, Vol. II, Grundlehren Math. Wiss. 331, Springer-Verlag, Berlin, 2006.

[30] B. Mordukhovich and Y. Shao, *Nonsmooth sequential analysis in Asplund spaces*, Trans. Amer. Math. Soc., 348 (1996), pp. 1235–1280.

[31] J. Palis and W. De Melo, *Geometric Theory of Dynamical Systems. An Introduction*, Springer-Verlag, New York, 1982 (translated from Portuguese by A. K. Manning).

[32] R. T. Rockafellar and R. Wets, *Variational Analysis*, Grundlehren Math. Wiss. 317, Springer-Verlag, Berlin, 1998.

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

[34] M. Shiota, *Geometry of Subanalytic and Semialgebraic Sets*, Progr. Math. 150, Birkhäuser, Boston, 1997.

[35] Ta Lê Loi, *Verdier and strict Thom stratifications in o-minimal structures*, Illinois J. Math., 42 (1998), pp. 347–356.