

Alternating minimization and projection methods for nonconvex problems ¹

Hedy ATTOUCH ², Jérôme BOLTE ³, Patrick REDONT ², Antoine SOUBEYRAN ⁴.

Abstract We study the convergence properties of alternating proximal minimization algorithms for (nonconvex) functions of the following type:

$$L(x, y) = f(x) + Q(x, y) + g(y)$$

where $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ and $g : \mathbb{R}^m \rightarrow \mathbb{R} \cup \{+\infty\}$ are proper lower semicontinuous functions and $Q : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$ is a smooth C^1 (finite valued) function which couples the variables x and y .

The algorithm is defined by:

$$(x_0, y_0) \in \mathbb{R}^n \times \mathbb{R}^m \text{ given, } (x_k, y_k) \rightarrow (x_{k+1}, y_k) \rightarrow (x_{k+1}, y_{k+1})$$

$$\begin{cases} x_{k+1} \in \operatorname{argmin}\{L(u, y_k) + \frac{1}{2\lambda_k}\|u - x_k\|^2 : u \in \mathbb{R}^n\} \\ y_{k+1} \in \operatorname{argmin}\{L(x_{k+1}, v) + \frac{1}{2\mu_k}\|v - y_k\|^2 : v \in \mathbb{R}^m\} \end{cases}$$

Note that the above algorithm can be viewed as an alternating *proximal* minimization algorithm. Alternating projection algorithms on closed sets are particular cases of the above problem: just specialize f and g to be the indicator functions of the two sets and take $Q(x, y) = \|x - y\|^2$.

The novelty of our approach is twofold: first, we work in a nonconvex setting, just assuming that the function L is a function that satisfies the Kurdyka-Lojasiewicz inequality. An entire section illustrates the relevancy of such an assumption by giving examples ranging from semialgebraic geometry to “metrically regular” problems. Secondly, we rely on a new class of alternating minimization algorithms with costs to move which has recently been introduced by Attouch, Redont and Soubeyran.

Our main result can be stated as follows: Assume that L has the Kurdyka-Lojasiewicz property and that the sequence $(x_k, y_k)_{k \in \mathbb{N}}$ is bounded. Then the trajectory has a finite length and, as a consequence, converges to a critical point of L . This result is completed by the study of the convergence rate of the algorithm, which depends on the geometrical properties of the function L around its critical points (namely the Lojasiewicz exponent). As a striking application, we obtain the convergence of our alternating projection algorithm (a variant of the Von Neumann algorithm) for a wide class of sets including in particular semialgebraic and tame sets, transverse smooth manifolds or sets with “regular” intersection.

Key words Alternating minimization algorithms, alternating projections, Lojasiewicz inequality, convergence rate, o-minimal structures, tame optimization, proximal algorithm.

1 Introduction

In this paper, we will be concerned by the convergence analysis of alternating minimization algorithms for (nonconvex) functions of the following type:

$$L(x, y) = f(x) + Q(x, y) + g(y) \tag{1}$$

where $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ and $g : \mathbb{R}^m \rightarrow \mathbb{R} \cup \{+\infty\}$ are proper lower semicontinuous functions and $Q : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$ is a smooth C^1 (finite valued) function which couples the variables x and y . Alternating projection algorithms on closed sets are particular cases of the above problem: just specialize f and g to be the indicator functions of the two sets and take $Q(x, y) = \|x - y\|^2$.

¹The first three authors acknowledge the support of the French ANR under grant ANR-05-BLAN-0248-01.

²Institut de Mathématiques et de Modélisation de Montpellier, UMR CNRS 5149, CC 51, Université de Montpellier II, Place Eugène Bataillon, 34095 Montpellier cedex 5, France (attouch@math.univ-montp2.fr, redont@math.univ-montp2.fr).

³Inria Futurs (CMAP, Polytechnique) et Equipe Combinatoire et Optimisation, UMR 7090, case 189, Université Pierre et Marie Curie, 4 Place Jussieu, 75252 Paris cedex 05 (bolte@math.jussieu.fr).

⁴GREQAM, UMR CNRS 6579, Université de la Méditerranée, 13290 Les Milles, France (antoine.soubeyran@univmed.fr).

The novelty of our approach is twofold: first, we work in a nonconvex setting, just assuming that the function L is a function that satisfies the Kurdyka-Lojasiewicz inequality, see [22], [23], [20]. As it has been established recently in [10, 11, 12], this assumption is satisfied by a wide class of nonsmooth functions called functions definable in an o-minimal structure (see Section 4). Semialgebraic functions and (globally) subanalytic functions are for instance definable in their respective classes.

Secondly, we rely on a new class of alternating minimization algorithms with costs to move which has recently been introduced by Attouch, Redont and Soubeyran [5] and which has proved to be a flexible tool, see [4]. Our study concerns the convergence of the following algorithm (discrete dynamical system)

$$(x_0, y_0) \in \mathbb{R}^n \times \mathbb{R}^m \text{ given, } (x_k, y_k) \rightarrow (x_{k+1}, y_k) \rightarrow (x_{k+1}, y_{k+1})$$

$$\begin{cases} x_{k+1} \in \operatorname{argmin}\{L(u, y_k) + \frac{1}{2\lambda_k}\|u - x_k\|^2 : u \in \mathbb{R}^n\}, \\ y_{k+1} \in \operatorname{argmin}\{L(x_{k+1}, v) + \frac{1}{2\mu_k}\|v - y_k\|^2 : v \in \mathbb{R}^m\}. \end{cases}$$

Note that the above algorithm can be viewed as an alternating *proximal* minimization algorithm.

Our main result (theorem 12) can be stated as follows: Assume that L has the Kurdyka-Lojasiewicz property and that the sequence $(x_k, y_k)_{k \in \mathbb{N}}$ is bounded. Then the trajectory has a finite length and, as a consequence, converges to a critical point of L . This result is completed by the study of the convergence rate of the algorithm, which depends on the geometrical properties of the function L around its critical points (namely the Lojasiewicz exponent).

There are two rich stories behind this problem. One concerns the alternating minimization algorithms, the other the use of the Lojasiewicz inequality in nonsmooth variational analysis. Our study is precisely at the intersection of these two active fields of research. Let us briefly delineate them and so put to the fore the originality of our approach.

1. Let us recall the classical result (1980) due to Acker and Prestel [2]: let \mathcal{H} be a real Hilbert space and $f : \mathcal{H} \mapsto \mathbb{R} \cup \{+\infty\}$, $g : \mathcal{H} \mapsto \mathbb{R} \cup \{+\infty\}$ two closed convex proper functions. Consider the sequence $(x_k, y_k)_{k \in \mathbb{N}}$ generated by the classical alternating minimization algorithm

$$\begin{cases} x_{k+1} = \operatorname{argmin}\{f(u) + \frac{1}{2}\|u - y_k\|_{\mathcal{H}}^2 : u \in \mathcal{H}\}, \\ y_{k+1} = \operatorname{argmin}\{g(v) + \frac{1}{2}\|x_{k+1} - v\|_{\mathcal{H}}^2 : v \in \mathcal{H}\}. \end{cases}$$

Then, the sequence $(x_k, y_k)_{k \in \mathbb{N}}$ weakly converges to a solution of the joint minimization problem on $\mathcal{H} \times \mathcal{H}$

$$\min \left\{ f(x) + g(y) + \frac{1}{2}\|x - y\|_{\mathcal{H}}^2 : (x, y) \in \mathcal{H} \times \mathcal{H} \right\},$$

if we assume that the minimum point set is nonempty.

Acker and Prestel's theorem provides a natural extension of von Neumann's alternating projection theorem [26] for two closed convex nonempty sets of the Hilbert space \mathcal{H} (take $f = \delta_{C_1}$, $g = \delta_{C_2}$). Note that the two sets may have an empty intersection, in which case the algorithm provide sequences converging to points in the respective sets which are as close to each other as possible. This fact has proved of fundamental importance in the applications of the method to the numerical analysis of ill-posed inverse problems. A rich litterature has been devoted to this subject, one may consult [7], [16], [14].

Indeed, in a series of recent papers [5], [4] a new class of alternating minimization algorithms with costs-to-move has been introduced. The most original feature of these algorithms is the introduction of the costs-to-move terms $\|u - x_k\|^2$, $\|v - y_k\|^2$. In the *modelling* of real world dynamical decision processes (here two agents play alternatively) these terms reflect an anchoring effect: the decision x_{k+1} is anchored to the preceding decision x_k , the term $\|u - x_k\|^2$ is the cost to change when passing from decision x_k to u . Similarly, $\|v - y_k\|^2$ is the cost to change when passing from decision y_k to v . The importance of these terms is crucial when decisions, performances are seen as routines, ways of doing. They take account of inertial effects and frictions [6].

From the algorithmic point of view, the introduction of these costs-to-move terms presents many advantages: They confer to the dynamic some stabilization and robustness properties. Moreover they

naturally appear when discretizing steepest descent dynamical systems for the related joint function (here the function L). As a result, the above algorithms are dissipative and enjoy nice convergent properties. Note that the cost-to-move terms asymptotically vanish and do not affect the type of equilibrium which is finally reached (indeed they allow to select one of them in the case of multiple equilibria).

2. Proximal algorithms can be viewed as implicit discretizations of the continuous steepest descent method. In the convex case, most results concerning the continuous dynamical system (steepest descent) can be converted to the discrete one (proximal algorithm) and vice versa. A deep and general understanding of these connections requires the introduction and the study of the differential inclusion

$$\dot{x}(t) \in -\partial f(x(t)), \quad t \geq 0$$

where ∂f denotes the subdifferential of the lower semicontinuous function $f : \mathcal{H} \mapsto \mathbb{R} \cup \{+\infty\}$.

Besides the modelling aspects, our special interest for the proximal method for functions involving analytic features comes from the recent development of the study of the steepest descent method for such functions. In his pioneering work on real-analytic functions [22], [23], Lojasiewicz provided the basic ingredient, the so-called ‘‘Lojasiewicz inequality’’, that allows to derive the convergence of all the bounded trajectories of the steepest descent to critical points. Given a real-analytic function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ and a critical point $a \in \mathbb{R}^n$ the Lojasiewicz inequality asserts that there exists some $\theta \in [\frac{1}{2}, 1)$ such that the function $|f - f(a)|^\theta \|\nabla f\|^{-1}$ remains bounded around a . Similar results have been developed for discrete gradient methods (see [1]) and nonsmooth subanalytic functions (see [10, 11]).

In the last decades powerful advances relying on an axiomatized approach of real-semialgebraic/real-analytic geometry have allowed to set up a general theory in which the basic objects enjoy the same qualitative properties than semialgebraic sets and functions [17, 29, 30]. In such a framework the central concept is the one of o-minimal structures over \mathbb{R} . Basic results are recalled and illustrated in Section 4.2.5. Following van den Dries [17], functions and sets belonging to such structures are called definable or *tame*⁵. Extensions of Lojasiewicz inequality to definable functions and applications to their gradient vector fields have been obtained by Kurdyka [20], while nonsmooth versions have been developed in [12]. The corresponding generalized inequality is hereby called the Kurdyka-Lojasiewicz inequality (see Definition 4.1, Section 4).

Keeping in mind the close connection between the continuous steepest descent and the proximal algorithms it is then natural to study the convergence of such algorithms for functions with ‘‘tame’’ features. A first step has recently been accomplished by Attouch and Bolte [3] who proved the convergence of the proximal algorithm for nonsmooth functions that satisfy the Lojasiewicz inequality around their generalized critical points.

Let us now come to our situation which involves two decision variables x and y and a bivariate function

$$L(x, y) = f(x) + Q(x, y) + g(y). \quad (2)$$

We are interested in studying splitting (alternate) algorithms whose trajectories converge to critical points of L .

The first key point of our approach is to assume that L satisfies the Kurdyka-Lojasiewicz inequality. The relevancy of this assumption is illustrated by many examples including uniformly convex functions, convex functions enjoying growth conditions, semialgebraic functions, definable functions. Specific examples related to feasibility problems are also described, they involve (possibly tangent) real-analytic manifolds, transverse manifolds (see [21]), semialgebraic sets or more generally tame sets...

This being assumed, it is then natural to apply to L an algorithm which is as close as possible to the proximal algorithm in the product space. The proof of our main result consists in proving that the alternating proximal minimization algorithm is close enough to the product space-proximal algorithm (one has to control the correcting terms introduced by the alternating effect), which makes it share the nice convergence properties of this algorithm.

⁵The word tame actually corresponds to a slight generalization of definable objects.

As a striking application, we obtain the convergence of our alternating projection algorithm, which can be seen as a variant of the von Neumann algorithm (section 5.3). Being given two closed subsets C, D of \mathbb{R}^n the latter is modelled on

$$\begin{cases} x_{k+1} \in P_C \left(\frac{x_k + y_k}{2} \right), \\ y_{k+1} \in P_D \left(\frac{y_k + x_{k+1}}{2} \right), \end{cases}$$

where $P_C, P_D : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ are the projection mappings of C and D . The convergence of the sequences $(x_k), (y_k)$ is obtained for a wide class of sets ranging from semialgebraic or definable sets to transverse manifolds or more generally to sets with a regular intersection. A part of this result is inspired by the recent work of Lewis and Malick on transverse manifolds [21], in which similar results were derived.

The paper is organized as follows: After the introduction, section 2 is devoted to recalling some elementary facts of nonsmooth analysis. This allows us to obtain in section 3 some first elementary properties of the alternating proximal minimization algorithm. Then, in section 4 we introduce the Kurdyka-Łojasiewicz inequality for lower semicontinuous functions and describe various classes of functions satisfying this property. The main convergence result is proved in section 5 together with an estimation of the convergence rate. The last section is devoted to some illustrations and applications of our results.

2 Elementary facts of nonsmooth analysis

The Euclidean scalar product of \mathbb{R}^n and its corresponding norm are respectively denoted by $\langle \cdot, \cdot \rangle$ and $|\cdot|$. General references for nonsmooth analysis are [13, 28, 25].

If $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is a point-to-set mapping its *graph* is defined by

$$\text{Graph } F := \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^m : y \in F(x)\}.$$

Similarly the graph of a real-extended-valued function $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ is defined by

$$\text{Graph } f := \{(x, s) \in \mathbb{R}^n \times \mathbb{R} : s = f(x)\}.$$

Let us recall a few definitions concerning subdifferential calculus.

Definition 1 ([28]) Let $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ be a proper lower semicontinuous function.

- (i) The domain of f is defined and denoted by $\text{dom } f := \{x \in \mathbb{R}^n : f(x) < +\infty\}$.
- (ii) For each $x \in \text{dom } f$, the Fréchet subdifferential of f at x , written $\hat{\partial}f(x)$, is the set of vectors $x^* \in \mathbb{R}^n$ which satisfy

$$\liminf_{\substack{y \neq x \\ y \rightarrow x}} \frac{1}{\|x - y\|} [f(y) - f(x) - \langle x^*, y - x \rangle] \geq 0.$$

If $x \notin \text{dom } f$, then $\hat{\partial}f(x) = \emptyset$.

- (iii) The limiting-subdifferential ([24]), or simply the subdifferential for short, of f at $x \in \text{dom } f$, written ∂f , is defined as follows

$$\partial f(x) := \{x^* \in \mathbb{R}^n : \exists x_n \rightarrow x, f(x_n) \rightarrow f(x), x_n^* \in \hat{\partial}f(x_n) \rightarrow x^*\}.$$

Remark 1 (a) The above definition implies that $\hat{\partial}f(x) \subset \partial f(x)$ for each $x \in \mathbb{R}^n$, where the first set is convex while the second one is closed.

(b)(Closedness of ∂f) Let $(x_k, x_k^*) \in \text{Graph } \partial f$ be a sequence that converges to (x, x^*) . If $f(x_k)$ converges

to $f(x)$ then $(x, x^*) \in \text{Graph } \partial f$.

(c) A necessary (but not sufficient) condition for $x \in \mathbb{R}^n$ to be a minimizer of f is

$$\partial f(x) \ni 0. \quad (3)$$

A point that satisfies (3) is called *limiting-critical* or simply *critical*. The set of critical points of f is denoted by $\text{crit } f$.

If K is a subset of \mathbb{R}^n and x is any point in \mathbb{R}^n , we set

$$\text{dist}(x, K) = \inf\{\|x - z\| : z \in K\}.$$

Recall that if K is empty we have $\text{dist}(x, K) = +\infty$ for all $x \in \mathbb{R}^n$. Note also that for any real-extended-valued function f on \mathbb{R}^n and any $x \in \mathbb{R}^n$, $\text{dist}(0, \partial f(x)) = \inf\{\|x^*\| : x^* \in \partial f(x)\}$.

Lemma 2 *For any sequence $x_k \rightarrow x$ such that $f(x_k) \rightarrow f(x)$, we have*

$$\liminf_{k \rightarrow +\infty} \text{dist}(0, \partial f(x_k)) \geq \text{dist}(0, \partial f(x)). \quad (4)$$

Proof. One can assume that the left hand side of (4) is finite. Let $x_k^* \in \partial f(x_k)$ be a sequence such that $\liminf_{k \rightarrow +\infty} \|x_k^*\| = \liminf_{k \rightarrow +\infty} \text{dist}(0, \partial f(x_k))$. There exists a subsequence $(x_{k'}^*)$ of (x_k^*) such that $\lim_{k' \rightarrow +\infty} \|x_{k'}^*\| = \liminf_{k \rightarrow +\infty} \|x_k^*\|$. Since $(x_{k'}^*)$ is bounded, one may assume that it converges to some x^* . Using Remark 1 (b), we see that $x^* \in \partial f(x)$. Thus $\partial f(x)$ is nonempty and $\lim \|x_{k'}^*\| = \|x^*\| \geq \text{dist}(0, \partial f(x))$. \square

Partial subdifferentiation Let $L : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R} \cup \{+\infty\}$ be a lower semicontinuous function. When y is a fixed point in \mathbb{R}^m the subdifferential of the function $L(\cdot, y)$ at u is denoted by $\partial_x L(u, y)$. Similarly one can define partial subdifferentiation with respect to the variable y . The corresponding operator is denoted by $\partial_y L(x, \cdot)$ (where x is a fixed point of \mathbb{R}^n).

The following elementary result will be useful

Proposition 3 *Let $L : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R} \cup \{+\infty\}$ be a function of the form $L(x, y) = f(x) + Q(x, y) + g(y)$ where $Q : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$ is a finite-valued Fréchet differentiable function and $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$, $g : \mathbb{R}^m \rightarrow \mathbb{R} \cup \{+\infty\}$ are lower semicontinuous functions. Then for all $(x, y) \in \text{dom } L = \text{dom } f \times \text{dom } g$ we have*

$$\partial L(x, y) = \begin{pmatrix} \partial f(x) + \nabla_x Q(x, y) \\ \partial g(y) + \nabla_y Q(x, y) \end{pmatrix}$$

Proof. Observe first that we have $\partial L(x, y) = \partial(f(x) + g(y)) + \nabla Q(x, y)$, since Q is differentiable ([28, 10.10 Exercice, p. 431]). Further, the subdifferential calculus for separable functions yields ([28, 10.5 Proposition, p. 426]) $\partial(f(x) + g(y)) = \partial f(x) \times \partial g(y)$. \square

Normal cones, indicator functions and projections

If C is a closed subset of \mathbb{R}^n we denote by δ_C its indicator function, *i.e.* for all $x \in \mathbb{R}^n$ we set

$$\delta_C(x) = \begin{cases} 0 & \text{if } x \in C, \\ +\infty & \text{otherwise.} \end{cases}$$

The projection on C , written P_C , is the following *point-to-set* mapping:

$$P_C : \begin{cases} \mathbb{R}^n & \rightrightarrows \mathbb{R}^n \\ x & \rightarrow P_C(x) := \text{argmin}\{\|x - z\| : z \in C\}. \end{cases}$$

When C is nonempty, the closedness of C and the compactness of the closed unit ball imply that $P_C(x)$ is nonempty for all x in \mathbb{R}^n .

Definition 4 (Normal cone) Let C be a nonempty closed subset of \mathbb{R}^n .

(i) For any $x \in C$ the Fréchet normal cone to C at x is defined by

$$\hat{N}_C(x) = \{v \in \mathbb{R}^n : \langle v, y - x \rangle \leq o(x - y), \forall y \in C\}.$$

When $x \notin C$ we set $N_C(x) = \emptyset$.

(ii) The (limiting) normal cone to C at x is denoted $N_C(x)$ and is defined by

$$v \in N_C(x) \Leftrightarrow \exists x_k \rightarrow x, v_k \in \hat{N}_C(x_k), v_k \rightarrow v.$$

Remark 2 (a) For $x \in C$ the cone $N_C(x)$ is closed but not necessarily convex.

(b) An elementary but important fact about normal cone and subdifferential is the following

$$\partial \delta_C = N_C.$$

(c) Recall that for a point-to-set mapping $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$, its generalized inverse $F^{-1} : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$ is defined by $x \in F^{-1}(y) \Leftrightarrow y \in F(x)$. Writing down the optimality condition associated to the definition of P_C , we obtain

$$P_C \subset (I + N_C)^{-1},$$

where I is the identity mapping of \mathbb{R}^n .

3 Alternating proximal minimization algorithm

Let $L : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R} \cup \{+\infty\}$ be a proper lower semicontinuous function. Being given $(x_0, y_0) \in \mathbb{R}^n \times \mathbb{R}^m$ we consider the following type of alternating discrete dynamical system: $(x_k, y_k) \rightarrow (x_{k+1}, y_k) \rightarrow (x_{k+1}, y_{k+1})$

$$\begin{cases} x_{k+1} \in \operatorname{argmin} \{L(u, y_k) + \frac{1}{2\lambda_k} \|u - x_k\|^2 : u \in \mathbb{R}^n\} \\ y_{k+1} \in \operatorname{argmin} \{L(x_{k+1}, v) + \frac{1}{2\mu_k} \|v - y_k\|^2 : v \in \mathbb{R}^m\}, \end{cases} \quad (5)$$

where $(\lambda_k)_{k \in \mathbb{N}}, (\mu_k)_{k \in \mathbb{N}}$ are positive sequences.

We make the following standing assumptions

$$\begin{cases} (\mathcal{H}_1) & \inf_{\mathbb{R}^n \times \mathbb{R}^m} L > -\infty, \\ (\mathcal{H}_2) & \text{the function } L(\cdot, y_0) \text{ is a proper function.} \end{cases}$$

Lemma 5 Assuming $(\mathcal{H}_1), (\mathcal{H}_2)$, the sequences $(x_k), (y_k)$ are correctly defined. Moreover:

(i) The following estimate holds

$$L(x_{k+1}, y_{k+1}) \leq L(x_k, y_k) - \frac{1}{2\lambda_k} \|x_{k+1} - x_k\|^2 - \frac{1}{2\mu_k} \|y_{k+1} - y_k\|^2. \quad (7)$$

(ii) First-order optimality conditions read

$$x_{k+1} - x_k + \lambda_k \partial_x L(x_{k+1}, y_k) \ni 0, \quad (8)$$

$$y_{k+1} - y_k + \mu_k \partial_y L(x_{k+1}, y_{k+1}) \ni 0. \quad (9)$$

Proof. Since $\inf L > -\infty$, \mathcal{H}_2 implies that for any $r > 0$, $(\bar{u}, \bar{v}) \in \mathbb{R}^n \times \mathbb{R}^m$ the functions $u \rightarrow L(u, \bar{v}) + \frac{1}{2r} \|u - \bar{u}\|^2$ and $v \rightarrow L(\bar{u}, v) + \frac{1}{2r} \|v - \bar{v}\|^2$ are coercive. An elementary induction ensures then that the sequences are well defined and that (7) holds for all integer k . Hence (i) holds.

The first-order optimality condition [28, Theorem 10.1] and the subdifferentiation formula for a sum of functions [28, Exercise 10.10] yield (8) and (9). \square

Remark 3 Without additional assumptions, like for instance the convexity of L , the sequences x_k, y_k are not *a priori* uniquely defined.

We consider the following additional assumptions:

- Being given $0 < r_- < r_+$, we assume that the sequences of stepsizes λ_k and μ_k belong to (r_-, r_+) for all $k \geq 0$.
- (\mathcal{H}_3) The restriction of L to its domain is a continuous function.

Proposition 6 *Assume that (\mathcal{H}_1), (\mathcal{H}_2), (\mathcal{H}_3) hold. Let (x_k, y_k) be a sequence which complies with (5, 6). Denote by $\omega(x_0, y_0)$ the set (possibly empty) of its limit points. Then*

- (i) $L(x_k, y_k)$ is decreasing,
- (ii) $\sum \|x_{k+1} - x_k\|^2 + \|y_{k+1} - y_k\|^2 < +\infty$,
- (iii) assuming moreover that L is of the form $L(x, y) = f(x) + Q(x, y) + g(y)$, then $\omega(x_0, y_0) \subset \text{crit } L$.
If in addition (x_k, y_k) is bounded then
- (iv) $\omega(x_0, y_0)$ is a nonempty compact connected set, and

$$d((x_k, y_k), \omega(x_0, y_0)) \rightarrow 0 \text{ as } k \rightarrow +\infty,$$

- (v) L is finite and constant on $\omega(x_0, y_0)$.

Proof. Assertion (i) follows immediately from (7), while (ii) is obtained by summation of (7) for $k = 1, \dots, N$:

$$L(x_{N+1}, y_{N+1}) + r_- \left(\sum_{k=1}^n \|x_{k+1} - x_k\|^2 + \|y_{k+1} - y_k\|^2 \right) \leq L(x_1, y_1) < +\infty. \quad (10)$$

Let us establish (iii). Item (ii) implies that $x_{k+1} - x_k$ and $y_{k+1} - y_k$ tend to zero. Using (8) and (9), we see that there exists $(x_k^*, y_k^*) \in (\partial_x L(x_k, y_{k-1}), \partial_y L(x_k, y_k)) = (\partial f(x_k) + \nabla_x Q(x_k, y_{k-1}), \partial g(y_k) + \nabla_y Q(x_k, y_k))$ that converges to zero as k goes to ∞ . If (\bar{x}, \bar{y}) is a limit point of (x_k, y_k) , Lemma 5 and the closure of the graphs of ∂f and ∂g show that $\partial L(\bar{x}, \bar{y}) \ni 0$.

Item (iv) follows by using (ii) together with some classical properties of sequences in \mathbb{R}^n . Assertion (v) is a consequence of (i) and (\mathcal{H}_3). \square

4 Kurdyka-Łojasiewicz inequality

4.1 Definition

Let $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ be a proper lower semicontinuous function that is continuous on its domain. For $0 < \eta_1 < \eta_2 \leq +\infty$, let us set

$$[\eta_1 < f < \eta_2] = \{x \in \mathbb{R}^n : \eta_1 < f < \eta_2\}.$$

We recall the following convenient definition from [3].

Definition 7 (Kurdyka-Łojasiewicz property) The function f is said to have the Kurdyka-Łojasiewicz property at $\bar{x} \in \text{dom } \partial f$ if:

There exist $\eta \in (0, +\infty]$, a neighborhood ⁶ U of \bar{x} and a continuous *concave* function $\phi : [0, \eta) \rightarrow \mathbb{R}_+$ such that:

- $\phi(0) = 0$,
- ϕ is C^1 on $(0, \eta)$,
- for all $s \in (0, \eta)$, $\phi'(s) > 0$.

⁶See Remark (4) (c)

and

$$\phi'(f(x) - f(\bar{x})) \text{dist}(0, \partial f(x)) \geq 1, \text{ whenever } x \text{ belongs to } U \cap [f(\bar{x}) < f < f(\bar{x}) + \eta]. \quad (11)$$

Remark 4 (a) S. Lojasiewicz proved in 1963 [22] that real-analytic functions satisfy an inequality of the above type with $\phi(s) = s^{1-\theta}$ where $\theta \in [\frac{1}{2}, 1)$. A nice proof of this result can be found in the monograph [15]. In a recent paper, Kurdyka has extended this result to functions definable in an o-minimal structure (see the next section).

(b) Let $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ be a proper lower semicontinuous function. For any noncritical point $\bar{x} \in \mathbb{R}^n$ of f , Lemma 2 yields the existence of $\epsilon, \eta > 0$ and c such that

$$\text{dist}(0, \partial f(z)) \geq c > 0$$

whenever $x \in B(\bar{x}, \epsilon) \cap [-\eta < f < \eta]$. In other words f has the Kurdyka-Lojasiewicz property at \bar{x} with $\phi(s) = c^{-1}s$.

(c) When a function has the Kurdyka-Lojasiewicz property, it is important to have nice estimations of η, U, ϕ . We shall see for instance that some convex functions satisfy the above property with $U = \mathbb{R}^n$ and $\eta = +\infty$. The determination of tight bounds for the nonconvex case is a lot more involved.

4.2 Examples of functions having the Kurdyka-Lojasiewicz property

4.2.1 Growth condition for convex functions

Let $U \subset \mathbb{R}^n$ and $\eta \in (0, +\infty]$. Consider a convex function f satisfying the following growth condition:

$$\exists c > 0, r \geq 1, \forall x \in U \cap [\min f < f < \eta], f(x) \geq f(\bar{x}) + cd(x, \text{argmin } f)^r,$$

where $\bar{x} \in \text{argmin } f \neq \emptyset$. Then f complies with (11) (for $\phi(s) = r c^{-\frac{1}{r}} s^{\frac{1}{r}}$) on $U \cap [\inf f < f < \eta]$ (see [10]).

4.2.2 Uniform convexity

If f is uniformly convex i.e., satisfies

$$f(y) \geq f(x) + \langle x^*, y - x \rangle + K\|y - x\|^p,$$

for all $x, y \in \mathbb{R}^n$, $x^* \in \partial f(x)$ then f satisfies the Kurdyka-Lojasiewicz inequality on $\text{dom } f$ for $\phi(s) = p K^{-\frac{1}{p}} s^{\frac{1}{p}}$.

Proof. Since f is coercive and strictly convex, $\text{argmin } f = \{\bar{x}\} \neq \emptyset$. Take $y \in \mathbb{R}^n$. By applying the uniform convexity property at the minimum point \bar{x} , we obtain that

$$f(y) \geq \min f + K\|y - \bar{x}\|^p.$$

The conclusion follows from Section 4.2.1.

4.2.3 Morse functions

A C^2 function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is a *Morse function* if for each critical point \bar{x} of f , the hessian $\nabla^2 f(\bar{x})$ of f at \bar{x} is a nondegenerate endomorphism of \mathbb{R}^n .

Let \bar{x} be a critical point of f . Using the Taylor formula for f and ∇f , we obtain the existence of a neighborhood U of \bar{x} and positive constants c_1, c_2 for which

$$|f(x) - f(\bar{x})| \leq c_1\|x - \bar{x}\|^2, \|\nabla f(x)\| \geq c_2\|x - \bar{x}\|,$$

whenever $x \in U$.

It is then straightforward to see that f complies to (11) with a function ϕ of the form $\phi(s) = c\sqrt{s}$, where c is a positive constant.

4.2.4 Metrically regular nonsmooth equations

The example to follow is inspired by a result of [27] that was used to analyze the complexity of second-order gradient methods.

Let us consider a mapping $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$. The function F is called metrically-regular on a subset C if there exists $k > 0$ such that

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

for all $(x, y) \in C \times F(C)$. The coefficient k is a measure of stability under perturbation of equations of the type $F(\bar{x}) = \bar{y}$. When F is C^1 a famous result, the so-called Lusternik-Graves theorem (see [19]), asserts that F is k -metrically regular on C if and only if $DF(x)$ is *surjective* for all x in C with

$$\|[DF(x)^*]^{-1}\| \leq k, \text{ for all } x \text{ in } C,$$

where $DF(x)$ (resp $DF^*(x)$) denotes the Fréchet derivative of F at x (resp. the adjoint of $DF(x)$).

Fundamental extensions to nonsmooth inclusions have been provided by many authors, see [19, 28] and references therein. For the sake of simplicity we assume here that F is C^1 and k -metrically regular on some subset C of \mathbb{R}^n .

Let us consider the following nonlinear problem: find $x \in C$ such that $F(x) = 0$.

Solving this problem amounts to solving

$$\min\{f(x) := \frac{1}{p}\|F(x)\|^p : x \in C\},$$

where $p \in [1, +\infty)$.

The function f is differentiable at each $x \in \mathbb{R}^n \setminus f^{-1}(0)$ with gradient $\nabla f(x) = \|F(x)\|^{p-2} DF^*(x)F(x)$. By using the metric regularity of F and Lusternik-Graves theorem we have

$$\|\nabla f(x)\| \geq \|[DF^*]^{-1}(x)\|^{-1} \|F(x)\|^{p-1} \geq k^{-1} p^{\frac{p-1}{p}} f(x)^{\frac{p-1}{p}} \text{ for all } x \in C \setminus f^{-1}(0).$$

In other words f has the Kurdyka-Lojasiewicz property at any $x \in \text{int } C$ (the interior of C) with

$$U = C, \eta = +\infty, \text{ and } \phi(s) = p^{\frac{1}{p}} k s^{\frac{1}{p}}, s \in \mathbb{R}.$$

It is important to observe here that $\text{argmin } f$ can be a *continuum* (in that case f is not a Morse function). This can easily be seen by taking $m < n$, $b \in \mathbb{R}^m$, a full rank matrix $A \in \mathbb{R}^{m \times n}$ and $F(x) = Ax - b$ for all x in \mathbb{R}^n .

4.2.5 Tame functions

Semialgebraic functions : Recall that a subset of \mathbb{R}^n is called semialgebraic if it can be written as a finite union of sets of the form

$$\{x \in \mathbb{R}^n : p_i(x) = 0, q_i(x) < 0, i = 1, \dots, p\},$$

where p_i, q_i are real polynomial functions.

A function $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ is semialgebraic if its graph is a semialgebraic subset of \mathbb{R}^{n+1} . Such a function satisfies the Kurdyka-Lojasiewicz property (see [10, 12]) with $\phi(s) = s^{1-\theta}$, $\theta \in [0, 1) \cap \mathbb{Q}$. This nonsmooth result generalizes the famous Łojasiewicz inequality for real-analytic functions [22].

Here are a few examples of semialgebraic functions:

- (a) $\mathbb{R}^n \ni x \rightarrow \|Ax - b\|_m^2 + \|Cx - d\|_{m'}$ where $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, $C \in \mathbb{R}^{m' \times n}$, $d \in \mathbb{R}^{m'}$,
- (b) $\mathbb{R}^n \ni x \rightarrow f(x) = \sup_{y \in C} g(x, y)$ where g and C are semialgebraic.
- (c) $\mathbb{R}^n \times \mathbb{R}^n \ni (x, y) \rightarrow L(x, y) = \frac{1}{2}\|x - y\|^2 + \delta_C(x) + \delta_D(y)$ where C and D are semialgebraic subsets of \mathbb{R}^n .

Functions definable in an o-minimal structure over \mathbb{R} . Introduced in [17] these structures can be seen as an axiomatization of the qualitative properties of semialgebraic sets.

Definition 8 Let $\mathcal{O} = \{\mathcal{O}_n\}_{n \in \mathbb{N}}$ be such that each \mathcal{O}_n is a collection of subsets of \mathbb{R}^n . The family \mathcal{O} is an o-minimal structure over \mathbb{R} , if it satisfies the following axioms:

- (i) Each \mathcal{O}_n is a boolean algebra. Namely $\emptyset \in \mathcal{O}_n$ and for each A, B in \mathcal{O}_n , $A \cup B$, $A \cap B$ and $\mathbb{R}^n \setminus A$ belong to \mathcal{O}_n .
- (ii) For all A in \mathcal{O}_n , $A \times \mathbb{R}$ and $\mathbb{R} \times A$ belong to \mathcal{O}_{n+1}
- (iii) For all A in \mathcal{O}_{n+1} , $\Pi(A) := \{(x_1, \dots, x_n) : (x_1, \dots, x_n, x_{n+1}) \in A\}$ belongs to \mathcal{O}_n .
- (iv) $\{x \in \mathbb{R}^n : x_i = x_j\} \in \mathcal{O}_n$ and $\{x \in \mathbb{R}^n : x_1 < x_2\} \in \mathcal{O}_2$
- (v) The elements of \mathcal{O}_1 are exactly finite unions of intervals.

Let \mathcal{O} be an o-minimal structure. A set A is said to be definable (in \mathcal{O}), if A belongs to \mathcal{O} . A point to set mapping $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ (resp. a real-extended-valued function $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$) is said to be definable if its graph is a definable subset of $\mathbb{R}^n \times \mathbb{R}^m$ (resp. $\mathbb{R}^n \times \mathbb{R}$).

Due to their dramatic impact on several domain in mathematics, these structures are intensively studied. One of the interest of such structures in optimization is due to the following nonsmooth extension of Kurdyka-Łojasiewicz inequality.

Theorem 9 ([12]) Any proper lower semicontinuous function $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ which is definable in an o-minimal structure \mathcal{O} has the Kurdyka-Łojasiewicz property at each point of $\text{dom } \partial f$. Moreover the function ϕ appearing in (11) is definable in \mathcal{O} .

The concavity of the function ϕ is not stated explicitly in [12]. The proof of that fact is however elementary: it relies on the following fundamental result known as the monotonicity Lemma (see [17]).

(Monotonicity Lemma) *Let k be an integer and $f : I \subseteq \mathbb{R} \rightarrow \mathbb{R}$ a function definable in some o-minimal structure. There exists a finite partition of I into intervals I_1, \dots, I_p such that the restriction of f to each I_i is C^k and either strictly monotone or constant.*

Proof of Theorem 9 [concavity of ϕ] Let x be a critical point of f and ϕ a L function of f at x (see (11)). As recalled above such a function ϕ exists and is definable in \mathcal{O} moreover. When applied to ϕ the monotonicity lemma yields the existence of $r > 0$ such that ϕ is C^2 with either $\phi'' \leq 0$ or $\phi'' > 0$ on $(0, r)$ (apply the monotonicity lemma to ϕ''). If $\phi'' > 0$ then ϕ' is increasing. Take $s_0 \in (0, r)$. Since $\phi'(s) \leq \phi'(s_0)$ for all $s \in (0, s_0)$, the function ϕ in inequality (11) can be replaced by $\psi(s) = \phi'(s_0)s$ for all $s \in (0, s_0)$. \square .

The following examples of o-minimal structures illustrate the considerable wealth of such a concept.

Example 1 (a) (semilinear sets) A subset of \mathbb{R}^n is called semilinear if it is a finite union of sets of the form

$$\{x \in \mathbb{R}^n : \langle a_i, x \rangle = \alpha_i, \langle b_i, x \rangle < \beta_i, i = 1, \dots, p\},$$

where $a_i, b_i \in \mathbb{R}^n$ and $\alpha_i, \beta_i \in \mathbb{R}$. One can easily establish that such a structure is an o-minimal structure.

Besides the function ϕ is of the form $\phi(s) = cs$ with $c > 0$.

Example 2 (b) (real semialgebraic sets) By Tarski quantifier elimination theorem the class of semialgebraic sets is an o-minimal structure [8, 9].

Example 3 (c) (Globally subanalytic sets) (Gabrielov [18]) There exists an o-minimal structure, denoted by \mathbb{R}_{an} , that contains all sets of the form $\{(x, t) \in [-1, 1]^n \times \mathbb{R} : f(x) = t\}$ where $f : [-1, 1]^n \rightarrow \mathbb{R}$ ($n \in \mathbb{N}$) is an analytic function that can be extended analytically on a neighborhood of the square $[-1, 1]^n$. The sets belonging to this structure are called *globally subanalytic sets*. As for the semialgebraic class, the function appearing in Kurdyka-Łojasiewicz inequality is of the form $\phi(s) = s^{1-\theta}$, $\theta \in [0, 1) \cap \mathbb{Q}$.

Let us give some concrete examples. Consider a finite collection of real-analytic functions $f_i : \mathbb{R}^n \rightarrow \mathbb{R}$ where $i = 1, \dots, p$.

- The restriction of the function $f_+ = \max_i f_i$ (resp. $f_- = \min_i f_i$) to each $[-a, a]^n$ ($a > 0$) is globally subanalytic.

- Take a finite collection of real-analytic functions g_j and set

$$C = \{x \in \mathbb{R}^n : f_i(x) = 0, g_j(x) \leq 0\}.$$

If C is a bounded subset then it is globally subanalytic. Let now $G : \mathbb{R}^m \times \mathbb{R}^n \rightarrow \mathbb{R}$ be a real analytic function. The restriction of the following function to each $[-a, a]^n$ ($a > 0$) is globally subanalytic:

$$f(x) = \max_{y \in C} G(x, y).$$

Example 4 (d) (log-exp structure) (Wilkie, van der Dries) [30, 17] There exists an o-minimal structure containing \mathbb{R}_{an} and the graph of $\exp : \mathbb{R} \rightarrow \mathbb{R}$. This structure is denoted by $\mathbb{R}_{\text{an,exp}}$. This huge structure contains all the aforementioned structures. One of the surprising specificity of such a structure is the existence of “infinitely flat” functions like $x \rightarrow \exp(-\frac{1}{x^2})$.

Many optimization problems are set in such a structure. When it is possible, it is however important to determine the minimal structure in which a problem is definable.

This can for instance have an impact on the convergence analysis and in particular on the knowledge of convergence rates (see Section 5.2 Theorem 14 for an illustration).

4.3 Feasibility problems

Being given two nonempty closed subsets C, D of \mathbb{R}^n , one introduces the following type of function:

$$L_{C,D}(x, y) := L(x, y) = \frac{1}{2}\|x - y\|^2 + \delta_C(x) + \delta_D(y),$$

where $(x, y) \in \mathbb{R}^n \times \mathbb{R}^n$.

Note that L is a proper lower semicontinuous function that satisfies \mathcal{H}_1 , \mathcal{H}_3 and \mathcal{H}_2 for any $y_0 \in D$.

Writing down the optimality condition we obtain that

$$\partial L(x, y) = \{(x - y + u, y - x + v) : u \in N_C(x), v \in N_D(y)\}.$$

This implies that

$$\text{dist}(0, \partial L(x, y)) = (\text{dist}^2(y - x, N_C(x)) + \text{dist}^2(x - y, N_D(y)))^{\frac{1}{2}}. \quad (12)$$

4.3.1 Noncritical/regular intersection

The following result is a reformulation of [21, Theorem 17] where this fruitful concept was introduced. For the sake of completeness we give a proof avoiding the use of metric regularity and Mordukhovich criterion.

Proposition 10 *Let C, D be two closed subsets of \mathbb{R}^n and $\bar{x} \in C \cap D$. Assume that*

$$-N_C(\bar{x}) \cap N_D(\bar{x}) = \{0\}.$$

Then there exists a neighborhood $U \subset \mathbb{R}^n \times \mathbb{R}^n$ of (\bar{x}, \bar{x}) and a positive constant c for which

$$\text{dist}(0, \partial L_{C,D}(x, y)) \geq c\|x - y\| > 0, \quad (13)$$

whenever $(x, y) \in U \cap [0 < L_{C,D} < +\infty]$.

In other words $L_{C,D}$ has the Kurdyka-Lojasiewicz property at (\bar{x}, \bar{x}) with $\phi(s) = \frac{2}{c}\sqrt{s}$.

Proof We argue by contradiction. Let $(x_k, y_k) \in [0 < L < +\infty]$ be a sequence converging to (\bar{x}, \bar{x}) such that

$$\frac{\text{dist}(0, \partial L_{C,D}(x_k, y_k))}{\|x_k - y_k\|} \leq \frac{1}{k+1}.$$

In view of (12), there exists $(u_k, v_k) \in N_C(x_k) \times N_D(y_k)$ such that

$$\|x_k - y_k\|^{-1}[\|y_k - x_k - u_k\| + \|x_k - y_k - v_k\|] \leq \frac{2}{k+1}.$$

Let $d \in S^{n-1}$ be a cluster point of $\|x_k - y_k\|^{-1}(y_k - x_k)$. Since $\|x_k - y_k\|^{-1}[(y_k - x_k) - u_k]$ converges to zero, d is also a limit-point of $\|x_k - y_k\|^{-1}u_k \in N_C(x_k)$. Due to the closedness property of N_C , we therefore obtain $d \in N_C(\bar{x})$. Arguing similarly with v_k we obtain that $d \in -N_D(\bar{x})$. \square

Remark 5 Observe that (13) reads

$$(\text{dist}^2(y - x, N_C(x)) + \text{dist}^2(x - y, N_D(y)))^{\frac{1}{2}} \geq c\|x - y\|.$$

4.3.2 Transverse Manifolds

As pointed out in [21] a nice example of regular intersection is given by the smooth notion of transversality. Let M be a smooth submanifold of \mathbb{R}^n . For each x in M , $T_x M$ denotes the tangent space to M at x . The normal cone to M at x is given by

$$N_M(x) = T_x M^\perp.$$

Two submanifolds of \mathbb{R}^n are called transverse at $x \in M \cap N$ if they satisfy

$$T_x M + T_x N = \mathbb{R}^n.$$

Due to the previous remark we have $N_M(x) \cap N_N(x) = N_M(x) \cap [-N_N(x)] = \{0\}$. And therefore $L_{M,N}$ satisfies the Kurdyka-Lojasiewicz inequality near x with $\phi(s) = c^{-1}\sqrt{s}$, $c > 0$.

The constant c can be estimated by means of a notion of angle [21, Theorem 18].

4.3.3 Tame feasibility

Assume that C, D are two globally analytic subsets of \mathbb{R}^n . The graph of the $L_{C,D}$ is given by

$$\text{Graph } L_{C,D} = G \cap (C \times D \times \mathbb{R}),$$

where G is the graph of the polynomial function $(x, y) \rightarrow \|x - y\|^2$. From the stability property of o-minimal structures (see Definition 8 (i)-(ii)), we deduce that $L_{C,D}$ is globally subanalytic.

Let $(\bar{x}, \bar{y}) \in C \times D$. Using the fact that the bifunction satisfies the Lojasiewicz inequality with $\phi(s) = s^{1-\theta}$ (where $\theta \in (0, 1]$) we obtain

$$(\text{dist}^2(y - x, N_C(x)) + \text{dist}^2(x - y, N_D(y)))^{\frac{1}{2}} \geq \|x - y\|^{2\theta} \quad (14)$$

for all (x, y) with $x \neq y$ in a neighborhood of (\bar{x}, \bar{y}) .

Remark 6 (a) One of the most interesting features of this inequality is that it is satisfied for possibly tangent sets.

(b) Let us also observe that the above inequality is satisfied if C, D are real-analytic submanifolds of \mathbb{R}^n . This is due to the fact that $B((\bar{x}, \bar{y}), r) \cap [C \times D]$ is globally subanalytic for all $r > 0$.

If C, D and the square norm function $\|\cdot\|^2$ are definable in an arbitrary o-minimal structure, we of course have a similar result. Namely,

$$\phi'(\|x - y\|^2) (\text{dist}^2(y - x, N_C(x)) + \text{dist}^2(x - y, N_D(y)))^{\frac{1}{2}} \geq 1, \quad (15)$$

for all (x, y) in a neighborhood of a critical point of L .

Although (15) is simply a specialization of a result of [12], we feel that this form of Kurdyka-Łojasiewicz inequality can be very useful in many contexts.

Remark 7 (a) Note that if C, D are definable in the log-exp structure then (15) holds. This is due to the fact that the square norm is also definable in this structure.
(b) If C, D are semilinear sets, $L_{C,D}$ is not semilinear but it is however semialgebraic.

5 Convergence results: alternating minimization and alternating projection algorithms

5.1 Lipschitz coupling

This section is devoted to the convergence analysis of the proximal minimization algorithm introduced in Section 3.

Let us assume that L is of the form

$$L(x, y) = f(x) + Q(x, y) + g(y) \quad (16)$$

where $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$, $g : \mathbb{R}^m \rightarrow \mathbb{R} \cup \{+\infty\}$ are proper lower semicontinuous functions and $Q : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$ is a C^1 (finite valued) function with ∇Q being Lipschitz continuous on bounded subsets of $\mathbb{R}^n \times \mathbb{R}^m$.

Lemma 11 (pre-convergence result) *Assume that L satisfies (\mathcal{H}_1) , (\mathcal{H}_2) and has the Kurdyka-Łojasiewicz property at $\bar{z} = (\bar{x}, \bar{y})$. Denote by U, η and $\phi : [0, \eta] \rightarrow \mathbb{R}$ the arguments appearing in (11). Let $\epsilon > 0$ be such that $B(\bar{z}, \epsilon) \subset U$ and set $\rho = \max(\frac{2}{r_-}, 2C^2 + \frac{1}{r_-^2})$, where C is a Lipschitz constant of ∇Q on $B(\bar{z}, \epsilon)$.*

Let (x_k, y_k) be a sequence such that $(x_0, y_0) \in B(\bar{z}, \epsilon) \cap [L(\bar{z}) < L < \eta]$ and $\inf_{k \in \mathbb{N}} L(x_k, y_k) \geq L(\bar{z})$. If

$$2\sqrt{\rho}r_+\phi(L(x_0, y_0) - L(\bar{z})) + \sqrt{r_+(L(x_0, y_0) - L(\bar{z}))} + \sqrt{\|x_0 - \bar{x}\|^2 + \|y_0 - \bar{y}\|^2} < \epsilon,$$

then

(i) The sequence $\{(x_k, y_k)\}$ lies in $B(\bar{z}, \epsilon)$.

(ii) The following estimate holds

$$\sum_{i=k}^{+\infty} \sqrt{\|x_{i+1} - x_i\|^2 + \|y_{i+1} - y_i\|^2} \leq 2(\sqrt{\rho}r_+\phi(l_k) + \sqrt{r_+l_k}) + \sqrt{\|x_k - \bar{x}\|^2 + \|y_k - \bar{y}\|^2},$$

where $l_k = L(x_k, y_k) - L(\bar{z})$.

(ii) The sequence (x_k, y_k) converges to a critical point of L .

Proof. With no loss of generality one may assume that $L(\bar{z}) = 0$ (replace if necessary L by $L - L(\bar{z})$). If $L(x_k, y_k) = L(\bar{z})$ for some $k \geq 0$, the fact that $\inf_{k \in \mathbb{N}} L(x_k, y_k) \geq L(\bar{z})$ together with inequality (7) imply that $x_i = x_k, y_i = y_k$ for all $i \geq k$. Since the indices k for which $\|x_{k+1} - x_k\| + \|y_{k+1} - y_k\| = 0$ have no impact on the asymptotic analysis we can assume that $(x_k, y_k) \in [0 < L < \eta]$ and $\|x_{k+1} - x_k\| + \|y_{k+1} - y_k\| > 0$ for all $k \geq 0$. Assume that $k \geq 1$ and let $x_k^* \in \partial f(x_k), y_k^* \in \partial g(y_k)$ be such that

$$\begin{aligned} x_k &= x_{k-1} - \lambda_k(x_k^* + \nabla_x Q(x_k, y_{k-1})), \\ y_k &= y_{k-1} - \mu_k(y_k^* + \nabla_y Q(x_k, y_k)). \end{aligned}$$

Let us prove by induction that $(x_k, y_k) \in B(\bar{z}, \epsilon)$ for all $k \geq 0$. The result is obvious for $k = 0$. Assume that $k \geq 1$ and that $(x_i, y_i) \in B(\bar{z}, \epsilon)$ for all $i \in \{1, \dots, k\}$.

Let $i \in \{1, \dots, k\}$, and set $l_k = L(x_k, y_k)$. Since $\phi' > 0$, (7) implies that

$$\phi'(l_i)(l_i - l_{i+1}) \geq \frac{\phi'(l_i)}{2r_+} [\|x_{i+1} - x_i\|^2 + \|y_{i+1} - y_i\|^2].$$

By using the concavity of the function ϕ

$$\phi(l_i) - \phi(l_{i+1}) \geq \frac{\phi'(l_i)}{2r_+} [\|x_{i+1} - x_i\|^2 + \|y_{i+1} - y_i\|^2]. \quad (17)$$

For each i set $V_i = (x_i^* + \nabla_x Q(x_i, y_i), y_i^* + \nabla_y Q(x_i, y_i)) \in \partial L(x_i, y_i)$. Using the definition of x_i^* and y_i^* we obtain successively

$$\begin{aligned} \|V_i\|^2 &= \left\| \frac{1}{\lambda_i} (x_{i-1} - x_i) + \nabla_x Q(x_i, y_i) - \nabla_x Q(x_i, y_{i-1}) \right\|^2 + \left\| \frac{1}{\mu_i} (y_i - y_{i-1}) \right\|^2 \\ &\leq \frac{1}{r_-^2} (2\|x_i - x_{i-1}\|^2 + \|y_i - y_{i-1}\|^2) + 2C^2 \|y_i - y_{i-1}\|^2 \\ &\leq \max\left(\frac{2}{r_-^2}, 2C^2 + \frac{1}{r_-^2}\right) [\|x_i - x_{i-1}\|^2 + \|y_i - y_{i-1}\|^2] \end{aligned} \quad (18)$$

Set $\rho = \max\left(\frac{2}{r_-^2}, 2C^2 + \frac{1}{r_-^2}\right)$. The induction assumption implies that $\phi'(l_i) \geq \text{dist}(0, \partial L(x_i, y_i))^{-1}$. Hence inequalities (18) and (17) yield

$$2\sqrt{\rho}r_+[\phi(l_i) - \phi(l_{i+1})] \geq \frac{\|x_{i+1} - x_i\|^2 + \|y_{i+1} - y_i\|^2}{\sqrt{\|x_i - x_{i-1}\|^2 + \|y_i - y_{i-1}\|^2}}. \quad (19)$$

Let us observe a simple fact: if $\frac{a^2}{b} \leq r$ for some real numbers $a, b > 0, r \geq 0$, then $a \leq \frac{1}{2}(b + r)$. This follows indeed from the fact that $2ab \leq a^2 + b^2$. When applying this inequality to (19), we obtain

$$\sqrt{\rho}r_+[\phi(l_i) - \phi(l_{i+1})] + \frac{1}{2}\sqrt{\|x_i - x_{i-1}\|^2 + \|y_i - y_{i-1}\|^2} \geq \sqrt{\|x_{i+1} - x_i\|^2 + \|y_{i+1} - y_i\|^2}. \quad (20)$$

After summation over all $i = 1, \dots, k$, we therefore obtain

$$\begin{aligned} 2\sqrt{\rho}r_+\phi(l_1) + \sqrt{\|x_1 - x_0\|^2 + \|y_1 - y_0\|^2} &\geq 2\sqrt{\rho}r_+[\phi(l_1) - \phi(l_{k+1})] + \sqrt{\|x_1 - x_0\|^2 + \|y_1 - y_0\|^2} \\ &\geq \sum_{i=0}^k \sqrt{\|x_{i+1} - x_i\|^2 + \|y_{i+1} - y_i\|^2}. \end{aligned}$$

By (7), we have $\|x_1 - x_0\|^2 + \|y_1 - y_0\|^2 \leq r_+[l_0 - l_1] \leq r_+ l_0$. We finally obtain

$$2\sqrt{\rho}r_+\phi(l_1) + \sqrt{r_+l_0} \geq \sum_{i=0}^k \sqrt{\|x_{i+1} - x_i\|^2 + \|y_{i+1} - y_i\|^2}.$$

Put $z_i = (x_i, y_i)$. Using successively the triangle inequality and the above inequality

$$\begin{aligned} \|z_{k+1} - \bar{z}\| &\leq \|z_{k+1} - z_0\| + \|z_0 - \bar{z}\| \\ &\leq \sum_{i=0}^k \|z_{i+1} - z_i\| + \|z_0 - \bar{z}\| \\ &\leq 2\sqrt{\rho}r_+\phi(l_1) + \sqrt{r_+l_0} + \|z_0 - \bar{z}\| \\ &\leq \epsilon. \end{aligned}$$

The conclusion follows readily. □

This Lemma has two important consequences

Theorem 12 (convergence of bounded sequences) *Assume that L satisfies (\mathcal{H}_1) , (\mathcal{H}_2) , (\mathcal{H}_3) and has the Kurdyka-Lojasiewicz property at each point.*

If the sequence (x_k, y_k) is bounded then

$$\sum_{k=1}^{+\infty} \|x_{k+1} - x_k\| + \|y_{k+1} - y_k\| < +\infty,$$

and as a consequence (x_k, y_k) converges to a critical point of L .

Proof Let \bar{z} be a limit-point of (x_k, y_k) for which we denote by ϵ, η, ϕ the associated objects as defined in (11). Note that Proposition 6 implies that \bar{z} is critical. By Lemma 5 and (\mathcal{H}_3) , $L(x_k, y_k)$ converges to $L(\bar{z})$. Since $\max(\phi(l_k) - L(\bar{z}), \|(x_k, y_k) - \bar{z}\|)$ admits 0 as a cluster point, we obtain the existence of $k_0 \geq 0$ such that

$$\sqrt{\rho}r_+ \phi(l_{k_0} - L(\bar{z})) + 2\sqrt{r_+ l_{k_0}} + \|(x_{k_0}, y_{k_0}) - \bar{z}\| < \epsilon.$$

The conclusion is then a consequence of Lemma 11. \square

Theorem 13 (local convergence to global minima) *Assume that L satisfies (\mathcal{H}_1) , (\mathcal{H}_2) and has the Kurdyka-Lojasiewicz property at \bar{z} .*

If \bar{z} is a global minimum of L , then there exist ϵ, η such that

$$\|x_0 - \bar{z}\| < \epsilon, \quad \min L \leq L(x_0, y_0) < \min L + \eta$$

implies that (x_k, y_k) has the finite length property and converges to (x^, y^*) with $L(x^*, y^*) = L(\bar{z})$.*

Proof A straightforward application of Lemma 11 yields the convergence result. To see that $L(x_k, y_k)$ converges to $\min L$, it suffices to observe that $L(x^*, y^*)$ is a critical value in $[\min L, \min L + \eta_1]$. \square

Remark 8 Theorem 12 gives new insights into convex alternating methods: first it shows that the finite length property is satisfied by many convex functions (*e.g.* convex definable functions), but it also relaxes the usual quadraticity assumption on Q that is required by usual methods, see [4] and references therein.

5.2 Convergence rate

Theorem 14 (rate of convergence) *Assume that (x_k, y_k) converges to (x^∞, y^∞) and that L has the Kurdyka-Lojasiewicz property at (x^∞, y^∞) with $\phi(s) = cs^{1-\theta}$, $\theta \in [0, 1)$, $c > 0$. Then the following estimations hold*

- (i) *If $\theta = 0$, the sequence $(x_k, y_k)_{k \in \mathbb{N}}$ converges in a finite number of steps.*
- (ii) *If $\theta \in (0, \frac{1}{2}]$ then there exist $c > 0$ and $Q \in [0, 1)$ such that*

$$\|(x_k, y_k) - (x^\infty, y^\infty)\| \leq c Q^k.$$

- (iii) *If $\theta \in (\frac{1}{2}, 1)$ then there exists $c > 0$ such that*

$$\|(x_k, y_k) - (x^\infty, y^\infty)\| \leq c k^{-\frac{1-\theta}{2\theta-1}}.$$

Proof. The notations are those of the previous proof and for simplicity we assume that $l_k \rightarrow 0$. Assume first $\theta = 0$ and take $k \geq 0$. Either $l_k = \inf\{l_i : i \geq k\}$ hence (x_k, y_k) is stationary (cf. proof of Lemma 11), or $\text{dist}(0, \partial L(x_k, y_k)) \geq c$ where c is a positive constant. In that case inequality (7) reads

$$l_{k+1} - l_k < -c'$$

where $c' > 0$. We therefore obtain that $l_{k_0} = \inf_i l_i$ for some integer k_0 .

Assume that $\theta > 0$. For any $k \geq 0$, set $\Delta_k = \sum_{i=k}^{\infty} \sqrt{\|x_{i+1} - x_i\|^2 + \|y_{i+1} - y_i\|^2}$ which is finite by Theorem 12. Since $\Delta_k \geq \sqrt{\|x^k - x^\infty\|^2 + \|y_k - y^\infty\|^2}$, it is sufficient to estimate Δ_k . With no loss of generality we may assume that $\Delta_k > 0$ for all $k \geq 0$. From (20) we obtain after summation

$$\sqrt{\rho} r_+ \phi(l_k) + \frac{1}{2}(\Delta_{k-1} - \Delta_k) \geq \Delta_k.$$

Using the Kurdyka-Lojasiewicz inequality and (18) we obtain

$$\begin{aligned} \phi(l_k) = cl_k^{1-\theta} &\leq c[(c(1-\theta))^{-1} \text{dist}(0, \partial L(x_k, y_k))]^{\frac{1-\theta}{\theta}} \\ &\leq K \sqrt{\|x_k - x_{k-1}\|^2 + \|y_k - y_{k-1}\|^2}^{\frac{1-\theta}{\theta}} \\ &\leq K(\Delta_{k-1} - \Delta_k)^{\frac{1-\theta}{\theta}}, \end{aligned}$$

where K is a positive constant. Combining the above results we obtain

$$K(\Delta_{k-1} - \Delta_k)^{\frac{1-\theta}{\theta}} + \frac{1}{2}(\Delta_{k-1} - \Delta_k) \geq \Delta_k \text{ where } K > 0.$$

Sequences Δ_k satisfying such inequalities have been studied in [3, Theorem 2]. Items (ii) and (iii) follow from these results. \square

5.3 An application to alternating projection

In this section we consider the special but important case of bifunctions of the type

$$L_{C,D}(x, y) = \delta_C(x) + \frac{1}{2}\|x - y\|^2 + \delta_D(y), \quad (x, y) \in \mathbb{R}^n,$$

where C, D are two nonempty closed subsets of \mathbb{R}^n . In this specific setting, the proximal minimization algorithm [(5), (6)] reads

$$\begin{aligned} x_{k+1} &\in \operatorname{argmin} \left\{ \frac{1}{2}\|u - y_k\|^2 + \frac{1}{2\lambda_k}\|u - x_k\|^2 : u \in C \right\} \\ y_{k+1} &\in \operatorname{argmin} \left\{ \frac{1}{2}\|v - x_{k+1}\|^2 + \frac{1}{2\mu_k}\|v - y_k\|^2 : v \in D \right\}, \end{aligned}$$

thus we obtain the following *alternating projection algorithm*

$$\begin{cases} x_{k+1} \in P_C \left(\frac{\lambda_k^{-1} x_k + y_k}{\lambda_k^{-1} + 1} \right) \\ y_{k+1} \in P_D \left(\frac{\mu_k^{-1} y_k + x_{k+1}}{\mu_k^{-1} + 1} \right). \end{cases}$$

The following result illustrates the interest of the above algorithm for feasibility problems.

Theorem 15 (convergence of bounded sequences) *Assume that the bifunction $L_{C,D}$ has the Kurdyka-Lojasiewicz property at each point. Then any bounded sequence (x_k, y_k) converges to a critical point of L .*

(local convergence) *Assume that the bifunction $L_{C,D}$ has the Kurdyka-Lojasiewicz property at (x^*, y^*) and that $\|x^* - y^*\| = \min\{\|x - y\| : x \in C, y \in D\}$.*

If (x_0, y_0) is sufficiently close to (x^, y^*) then the whole sequence converges to a point (x^∞, y^∞) such that $\|x^\infty - y^\infty\| = \min\{\|x - y\| : x \in C, y \in D\}$.*

Proof The first point is due to Theorem 12 while the second follows from a specialization of Theorem 13 to $L_{C,D}$. \square

This variant of the von Neuman alternating projection method yields convergent sequences for a wide class of sets (see Section 4.3.3), *e.g.* when $C, D, \|\cdot\|^2$ are definable in the same o-minimal structure or when the intersection $C \cap D$ is regular.

Note also that convergence rates results of Section 5.2 can be applied to the following fundamental cases

- C, D semilinear, semialgebraic, globally subanalytic or real-analytic submanifolds of \mathbb{R}^n .
- $C \cap D$ regular (*e.g.* *transverse manifolds*).

For this last case the convergence is of linear type, *i.e.* $\|x_k - x_\infty\| \leq c_1 Q_1^k, \|y_k - y_\infty\| \leq c_2 Q_2^k$ where $c_i > 0$ and $Q_i \in (0, 1)$.

References

- [1] ABSIL, P.-A., MAHONY, R. , ANDREWS, B., Convergence of the Iterates of Descent Methods for Analytic Cost Functions, preprint 16p, 2004.
- [2] ACKER, F. , PRESTEL, M.-A. Convergence d'un schéma de minimisation alternée, Ann. Fac. Sci. Toulouse. (5), 2 (1980), 1-9.
- [3] ATTOUCH, H., BOLTE, J. On the convergence of the proximal algorithm for nonsmooth functions involving analytic features, Math. Prog. Series B., in press, (published electronically, 2006).
- [4] ATTOUCH, H., BOLTE, J., REDONT, P., SOUBEYRAN, A., Alternating proximal algorithms for weakly coupled minimization problems. Applications to dynamical games and PDE's (submitted to J. of Convex Analysis).
- [5] H. ATTOUCH P. REDONT AND A. SOUBEYRAN, A new class of alternating proximal minimization algorithms with costs-to-move, SIAM Journal on Optimization, 18 (2007), pp. 1061-1081.
- [6] H. ATTOUCH AND A. SOUBEYRAN, Inertia and Reactivity in Decision Making as Cognitive Variational Inequalities, Journal of Convex Analysis, 13 (2006), pp. 207-224.
- [7] BAUSCHKE, H.H. , BORWEIN, J.M. On projection algorithms for solving convex feasibility problems, SIAM review, 38(3), 1996, 367-426.
- [8] BENEDETTI, R. , RISLER, J.-J., *Real Algebraic and Semialgebraic Sets*, Hermann, Éditeur des Sciences et des Arts, (Paris, 1990).
- [9] BOCHNAK, J., COSTE, M. , ROY, M.-F., *Real Algebraic Geometry*, (Springer, 1998).
- [10] BOLTE, J., DANILIDIS, A. , LEWIS, A., The Lojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems. SIAM J. Optim. 17(2006), no. 4, 1205–1223.
- [11] BOLTE, J., DANILIDIS, A. , LEWIS, A., A nonsmooth Morse-Sard theorem for subanalytic functions. J. Math. Anal. Appl. 321(2006), no. 2, 729–740.
- [12] BOLTE, J., DANILIDIS, A., LEWIS, A., SHIOTA, M., Clarke subgradients of stratifiable functions. SIAM J. Optim. 18(2007), no. 2, 556–572.
- [13] CLARKE, F.H., LEDYAEV, YU., STERN, R.I. , WOLENSKI, P.R., *Nonsmooth analysis and control theory*, Graduate texts in Mathematics **178**, (Springer-Verlag, New-York, 1998).

- [14] COMBETTES, P. L. Signal recovery by best feasible approximation, *IEEE transactions on Image Processing*, 2(2)(1993) 269-271.
- [15] DENKOWSKA, Z., STASICA, J., Ensembles sous-analytiques à la Polonaise, *Travaux en cours* 69, Hermann, Paris 2008, to appear.
- [16] DEUTSCH, F. Best approximation in inner product spaces. Springer, New York, 2001.
- [17] VAN DEN DRIES, L., Tame topology and o-minimal structures. London Mathematical Society Lecture Note Series, 248. Cambridge University Press, Cambridge, 1998. x+180 pp.
- [18] GABRIELOV, A., Complements of subanalytic sets and existential formulas for analytic functions, *Inventiones Math.*, 125, 1996, 1–12.
- [19] IOFFE, A. D. Metric regularity and subdifferential calculus. (Russian) *Uspekhi Mat. Nauk* 55 (2000), no. 3(333), 103–162; translation in *Russian Math. Surveys* 55 (2000), no. 3, 501–558.
- [20] KURDYKA, K., On gradients of functions definable in o-minimal structures, *Ann. Inst. Fourier* 48 (1998), 769-783.
- [21] LEWIS, A.S., MALICK, J., Alternating projection on manifolds, to appear in *Math. Op. Res.*
- [22] ŁOJASIEWICZ, S., “Une propriété topologique des sous-ensembles analytiques réels.”, in: *Les Équations aux Dérivées Partielles*, pp. 87–89, Éditions du centre National de la Recherche Scientifique, Paris 1963.
- [23] ŁOJASIEWICZ, S., Sur la géométrie semi- et sous-analytique, *Ann. Inst. Fourier* 43 (1993), 1575-1595.
- [24] MORDUKHOVICH, B., Maximum principle in the problem of time optimal response with nonsmooth constraints, *J. Appl. Math. Mech.*, 40 (1976), 960–969 ; [translated from *Prikl. Mat. Meh.* 40 (1976), 1014–1023]
- [25] MORDUKHOVICH, B. *Variational analysis and generalized differentiation. I. Basic theory*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], 330. Springer-Verlag, Berlin, 2006. xxii+579 pp.
- [26] J. VON NEUMANN, *Functional Operators*, Annals of Mathematics studies 22, Princeton University Press, 1950.
- [27] NESTEROV, YU., *Accelerating the cubic regularization of Newton’s method on convex problems*, *Math. Program.* 112 (2008), no. 1, Ser. B, 159–181.
- [28] ROCKAFELLAR, R.T. , WETS, R., *Variational Analysis*, Grundlehren der Mathematischen, Wissenschaften, Vol. 317, (Springer, 1998).
- [29] SHIOTA, M, *Geometry of subanalytic and semialgebraic sets*, Progress in Mathematics 150, Birkhäuser, (Boston, 1997).
- [30] WILKIE, A. J., *Model completeness results for expansions of the ordered field of real numbers by restricted Pfaffian functions and the exponential function*, *J. Amer. Math. Soc.* 9 (1996), no. 4, 1051–1094.