Proximal alternating minimization and projection methods for nonconvex problems. An approach based on the Kurdyka-Łojasiewicz inequality [1]

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

Abstract. We study the convergence properties of an alternating proximal minimization algorithm for nonconvex structured functions of the type: $L(x,y) = f(x)+Q(x,y)+g(y)$, where $f : \mathbb{R}^n \to \mathbb{R}\cup\{+\infty\}$ and $g : \mathbb{R}^m \to \mathbb{R}\cup\{+\infty\}$ are proper lower semicontinuous functions, and $Q : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ is a smooth C^1 function which couples the variables x and y. The algorithm can be viewed as a *proximal* regularization of the usual Gauss-Seidel method to minimize L.

We work in a nonconvex setting, just assuming that the function L satisfies the Kurdyka-Łojasiewicz inequality. An entire section illustrates the relevancy of such an assumption by giving examples ranging from semialgebraic geometry to "metrically regular" problems.

Our main result can be stated as follows: If L has the Kurdyka-Łojasiewicz property, then each bounded sequence generated by the algorithm 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. When specialized to $Q(x,y) = \|x-y\|^2$ and to f, g indicator functions, the algorithm is an alternating projection mehod (a variant of Von Neumann's) that converges for a wide class of sets including semialgebraic and tame sets, transverse smooth manifolds or sets with "regular "intersection. In order to illustrate our results with concrete problems, we provide a convergent proximal reweighted ℓ^1 algorithm for compressive sensing and an application to rank reduction problems.

Key words Alternating minimization algorithms, alternating projections algorithms, proximal algorithms, nonconvex optimization, Kurdyka-Łojasiewicz inequality, o-minimal structures, tame optimization, convergence rate, finite convergence time, gradient systems, sparse reconstruction.

AMS 2000 Subject Classification: 65K10, 90C26, 49J53, 49M27

1 Introduction

Presentation of the algorithm. In this paper, we will be concerned with the convergence analysis of alternating minimization algorithms for (nonconvex) functions $L : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}\cup\{+\infty\}$ of the following type:

$$(\mathcal{H}) \quad \begin{cases} L(x,y) = f(x) + Q(x,y) + g(y), \\ f : \mathbb{R}^n \to \mathbb{R}\cup\{+\infty\},\ g : \mathbb{R}^m \to \mathbb{R}\cup\{+\infty\} \text{ are proper lower semicontinuous,} \\ Q : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R} \text{ is a } C^1 \text{ function,} \\ \nabla Q \text{ is Lipschitz continuous on bounded subsets of } \mathbb{R}^n \times \mathbb{R}^m. \end{cases}$$

Assumption $(\mathcal{H})$ will be needed throughout the paper.

We aim at finding critical points of

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

and possibly solve the corresponding minimization problem.

The specific structure of L allows in particular to tackle problems of the form

$$\min\{f(z) + g(z) : z \in \mathbb{R}^n\}. \tag{2}$$

[1] The first three authors acknowledge the support of the French ANR under grant ANR-08-BLAN-0294-03.

[2] Institut 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).

[3] UPMC Paris 06, Equipe Combinatoire et Optimisation and Inria Saclay (CMAP, Polytechnique) , UMR 7090, case 189, Université Pierre et Marie Curie, 4 Place Jussieu, 75252 Paris cedex 05 (bolte@math.jussieu.fr).

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

It suffices indeed to set $L_\rho(x,y) = f(x) + \frac{\rho}{2}||x-y||^2 + g(y)$, ρ being a positive penalization (or relaxation) parameter, and to minimize L_ρ over $\mathbb{R}^n \times \mathbb{R}^n$. Feasibility problems involving two closed sets are particular cases of the above problem: just specialize f and g to be the indicator functions. (In a convex setting, the square of the euclidean distance may be replaced by a Bregman distance, see [11].)

Minimizing the sum of simple functions or finding a common point to a collection of closed sets is a very active field of research with applications in approximation theory [44], image reconstruction [24, 27], statistics [25, 32], partial differential equations and optimal control [38, 48]. A good reference for problems involving convex instances is [24]: many examples coming from signal processing problems are shown to be rewritable as (1).

The specificity of our approach is twofold. First, we work in a nonconvex setting, just assuming that the function L satisfies the Kurdyka-Łojasiewicz inequality, see [39, 40, 35]. As it has been established recently in [16, 17, 18], this assumption is satisfied by a wide class of nonsmooth functions called functions definable in an o-minimal structure (see Section 4.3). 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 in [6] (see also [7]) and which has proved to be a flexible tool [5] permitting to handle general coupling functions $Q(x,y)$ (for example $Q(x,y) = \|Ax-By\|^2$, with A, B linear operators):

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

$$\begin{cases} x_{k+1} \in \operatorname{argmin}\left\{ L(u, y_k) + \dfrac{1}{2\lambda_k}\|u - x_k\|^2 : u \in \mathbb{R}^n \right\}, \\ y_{k+1} \in \operatorname{argmin}\left\{ L(x_{k+1}, v) + \dfrac{1}{2\mu_k}\|v - y_k\|^2 : v \in \mathbb{R}^m \right\}. \end{cases} \tag{3}$$

The above algorithm can be viewed as a *proximal* regularization of a two block Gauss-Seidel method for minimizing L:

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

Some general results for Gauss-Seidel method, also known as coordinate descent method, can be found for instance in [9, 13]; block coordinate methods for nonsmooth and nonconvex functions have been investigated by many authors (see [47] and references therein). However very few general results ensure that the sequence (x_k, y_k) converges to a global minimizer, even for strictly convex functions. An important fact concerning our approach is that the convergence of algorithm (3) works for any stepsizes λ_k, μ_k greater than a fixed positive parameter which can be chosen arbitrarily large. For such parameters the algorithm is very close to a coordinate descent method. On the other hand, when the stepsizes are not too large, the method is an alternating gradient-like method.

Kurdyka-Łojasiewicz inequalities and tame geometry. Before describing and illustrating our convergence results, let us recall some important facts that have motivated our mathematical approach.

In his pioneering work on real-analytic functions [39, 40], Łojasiewicz provided the basic ingredient, the so-called "Łojasiewicz inequality", that allows to derive the convergence of the bounded trajectories of the steepest descent equation to critical points. Given a real-analytic function $f : \mathbb{R}^n \to \mathbb{R}$ and a critical point $a \in \mathbb{R}^n$, the Łojasiewicz inequality asserts that there exists some $\theta \in \left[\frac{1}{2}, 1\right)$ 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 [16, 17, 4]).

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 as semialgebraic sets and functions [28, 46, 49]. In such a framework the central concept is that of o-minimal structure over $\mathbb{R}$. Basic results are recalled and illustrated in Section 4.3. Following van den Dries [28], functions and sets belonging to such structures are called definable or *tame* ([5]). Extensions of the Łojasiewicz inequality to definable functions and applications to their gradient

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

vector fields have been obtained by Kurdyka [35], while nonsmooth versions have been developed in [18]. The corresponding generalized inequality is here called the Kurdyka-Łojasiewicz inequality (see Definition 7, Section 3.2).

An important motivation for designing alternating algorithms for tame problems relies on these generalized Łojasiewicz inequalities but also on the following facts:

– tame sets and functions allow to model many problems both smooth and nonsmooth: most of standard subsets of matrices are semialgebraic (symmetric positive semidefinite matrices, orthogonal groups, Stiefel manifolds, constant rank matrices) and most criteria involve piecewise polynomial and analytic functions.

– tameness is a highly stable "concept": finite unions or finite intersections of tame sets are tame, compositions of tame mappings are tame, subdifferentials of tame functions are tame (see the paragraph after Definition 13 and also [16, 18]).

Convergence results, tractability of the algorithm and applications. One of our central result (Theorem 9) can be stated as follows: Assume that L has the Kurdyka-Łojasiewicz property at each point. Then either the sequence (x_k, y_k) converges to infinity, or 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 Łojasiewicz exponent).

Let us now give some insight into the interest of algorithm (3).

Verifying that a nonsmooth function has the Kurdyka-Łojasiewicz property *at each point* is often a very easy task. For instance check the semialgebraicity or the analyticity of functions f, g, Q and apply the nonsmooth Łojasiewicz inequalities provided in [16, 18]. Section 4 below shows the preeminence of this inequality and its links to prominent concepts in Optimization:

– uniformly convex functions and convex functions enjoying growth conditions;

– metric regularity and constraint qualification;

– semialgebraic and definable functions.

Specific examples related to feasibility problems are provided, they involve (possibly tangent) real-analytic manifolds, transverse manifolds (see [36]), semialgebraic sets or more generally tame sets.

Our results also apply to *tame convex problems* which are easily identifiable in practice and which provide in turn an important field of applications (see for instance [24]). Many convergence results of various types are available under convexity assumptions [10, 24]. Our results seem new; when L is a tame closed convex function and has at least a minimizer, then the sequences generated by algorithm (3) converge. Besides, when L is semialgebraic or globally subanalytic, convergence rates are necessarily of the form $O(\frac{1}{k^s})$ with $s > 0$.

Computability of the proximal iterates deserves some explanation. For simplicity take $Q(x, y) = \frac{1}{2}||x - y||^2$. A first general remark is the following: when, for instance, f is locally convex up to a square, an adequate choice of the stepsize λ_k makes the computation of the first step of algorithm 3 a *convex*, hence tractable, problem. Computational issues concerning these aspects of the implementation of approximate proximal points are given by Hare-Sagastizábal in [30]; besides, the Nesterov optimal gradient algorithm provides a simple tool for efficiently solving convex problems ([43]). Some nonconvex cases are easily computable and amount to *explicit* or standard computations: projections onto spheres or onto ellipsoids, constant rank matrices, or even one real variable second order equations. A good reference for some of these aspects is [36]. In order to provide a more realistic and flexible tool for solving real-world problems, it would be natural to consider inexact versions of algorithm (3) (see [34, 23] for some work in that direction); this subject is out of the scope of the present paper but it is a matter for future research.

Several applications involving nonconvex aspects are provided (see Section 4.3): rank reduction of correlation matrices, compressive sensing with nonconvex "norms". For this last case our algorithm provides a regularized version of the reweighted l^1 minimization algorithm of Candès-Wakin-Boyd (see Example 2). To the best of our knowledge no convergence results are known for the reweighted algorithms. The regularized version we provide converges even when small constant stepsizes are chosen in the implementation of (3).

In Section 3.4, when specializing algorithm (3) to indicator functions, we obtain an alternating projection algorithm (recently introduced in [6]), which can be seen as a proximal regularization of the von Neumann algorithm (see [44]). Being given two closed subsets C, D of $\mathbb{R}^n$ the algorithm reads

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

where $P_C, P_D : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ are the projection mappings onto 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 [36] (and also [37]), in which similar results were derived.

The paper is organized as follows: 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, and then to establish our main theoretical results. A last section is devoted to examples and applications: various classes of functions satisfying the Kurdyka-Lojasiewicz property are provided and specific examples for which computations can be effectively performed are given.

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 $|| \cdot ||$. Some general references for nonsmooth analysis are [45, 42].

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

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

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

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

Let us recall a few definitions concerning subdifferential calculus.

Definition 1 ([45]) Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a proper lower semicontinuous function.

(i) The domain of f is defined and denoted by $\mathrm{dom}\, f := \{x \in \mathbb{R}^n : f(x) < +\infty\}$.

(ii) For each $x \in \mathrm{dom}\, f$, the Fréchet subdifferential of f at x, written $\hat{\partial} f(x)$, is the set of vectors $x^* \in \mathbb{R}^n$ which satisfy

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

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

(iii) The limiting-subdifferential ([42]), or simply the subdifferential for short, of f at $x \in \mathrm{dom}\, f$, written $\partial f(x)$, is defined as follows

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

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 and closed while the second one is closed [45, th. 8.6 p. 302].
(b)(Closedness of ∂f) Let $(x_k, x_k^*) \in \operatorname{Graph} \partial f$ be a sequence that converges to (x, x^*). By the very definition of $\partial f(x)$, if $f(x_k)$ converges to $f(x)$ then $(x, x^*) \in \operatorname{Graph} \partial f$.
(c) A necessary (but not sufficient) condition for $x \in \mathbb{R}^n$ to be a minimizer of f is

$$\partial f(x) \ni 0. \tag{4}$$

A point that satisfies (4) is called *limiting-critical* or simply critical. The set of critical points of f is denoted by $\operatorname{crit} f$.

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

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

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

Lemma 2 *Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a proper lower semicontinuous function. Let $\bar{x} \in \operatorname{dom} f$ be a noncritical point of f. Then there exists $c > 0$ such that*

$$\|x - \bar{x}\| + \|f(x) - f(\bar{x})\| < c \implies \operatorname{dist}(0, \partial f(x)) \geq c.$$

Proof. On the contrary, there would exist a sequence (c_k) with $c_k > 0$, $c_k \to 0$, and a sequence (x_k) with $\|x_k - \bar{x}\| + \|f(x_k) - f(\bar{x})\| < c_k$ and $\operatorname{dist}(0, \partial f(x_k)) < c_k$. The latter inequality implies the existence of some $x_k^* \in \partial f(x_k)$ with $\|x_k^*\| < c_k$. Owing to the closedness of ∂f we would then have $0 \in \partial f(\bar{x})$, a contradiction. □

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

The following result, though elementary, is central to the paper.

Proposition 3 *Let L satisfy $(\mathcal{H})$. Then for all $(x, y) \in \operatorname{dom} L = \operatorname{dom} f \times \operatorname{dom} g$ we have*

$$\partial L(x, y) = \{\partial f(x) + \nabla_x Q(x, y)\} \times \{\partial g(y) + \nabla_y Q(x, y)\} = \partial_x L(x, y) \times \partial_y L(x, y).$$

Proof. Observe first that we have $\partial L(x, y) = \partial(f(x) + g(y)) + \nabla Q(x, y)$, since Q is continuously differentiable ([45, 8.8(c) Exercice, p. 304]). Further, the subdifferential calculus for separable functions yields ([45, 10.5 Proposition, p. 426]) $\partial(f(x) + g(y)) = \partial f(x) \times \partial g(y)$. Hence the first equality.

Invoking once more ([45, 8.8(c) Exercice]) yields the second equality. □

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:

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

When C is nonempty, the closedness of C and the compactness of the closed unit ball of $\mathbb{R}^n$ 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

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

When $x \notin C$ we set $N_C(x) = \emptyset$.
(ii) The (limiting) normal cone to C at $x \in C$ is denoted by $N_C(x)$ and is defined by

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

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

$$\partial \delta_C = N_C.$$

3 Alternating proximal minimization algorithms

3.1 Convergence to a critical value

Let L satisfy $(\mathcal{H})$. Being given $(x_0, y_0) \in \mathbb{R}^n \times \mathbb{R}^m$, recall that the alternating discrete dynamical system we are to study is of the form: $(x_k, y_k) \to (x_{k+1}, y_k) \to (x_{k+1}, y_{k+1})$

$$\begin{cases} x_{k+1} \in \operatorname{argmin}\, \{L(u, y_k) + \dfrac{1}{2\lambda_k}\|u - x_k\|^2 : u \in \mathbb{R}^n\} & (5) \\ y_{k+1} \in \operatorname{argmin}\, \{L(x_{k+1}, v) + \dfrac{1}{2\mu_k}\|v - y_k\|^2 : v \in \mathbb{R}^m\}, & (6) \end{cases}$$

where $(\lambda_k)_{k\in\mathbb{N}}, (\mu_k)_{k\in\mathbb{N}}$ are positive sequences.
We make the following standing assumption concerning (5), (6):

$$(\mathcal{H}_1) \begin{cases} \inf_{\mathbb{R}^n \times \mathbb{R}^m} L > -\infty, \\ \text{the function } L(\cdot, y_0) \text{ is proper,} \\ \text{for some positive } r_- < r_+ \text{ the sequences of stepsizes } \lambda_k,\ \mu_k \text{ belong to } (r_-, r_+) \text{ for all } k \geq 0. \end{cases}$$

The next lemma, especially point (iii), is of constant use in the sequel.

Lemma 5 *Under assumptions* $(\mathcal{H})$, $(\mathcal{H}_1)$, *the sequences* (x_k), (y_k) *are correctly defined. Moreover*
(i) The following estimate holds

$$L(x_k, y_k) + \frac{1}{2\lambda_{k-1}}\|x_k - x_{k-1}\|^2 + \frac{1}{2\mu_{k-1}}\|y_k - y_{k-1}\|^2 \leq L(x_{k-1}, y_{k-1}) \quad \forall k \geq 1; \tag{7}$$

hence $L(x_k, y_k)$ *does not increase.*
(ii)

$$\sum_{k=1}^{\infty} \left(\|x_k - x_{k-1}\|^2 + \|y_k - y_{k-1}\|^2\right) < +\infty;$$

hence $\lim(\|x_k - x_{k-1}\| + \|y_k - y_{k-1}\|) = 0$.
(iii) For $k \geq 1$ *define* $(x_k^*, y_k^*) = (\nabla_x Q(x_k, y_k) - \nabla_x Q(x_k, y_{k-1}), 0) - \left(\frac{1}{\lambda_{k-1}}(x_k - x_{k-1}), \frac{1}{\mu_{k-1}}(y_k - y_{k-1})\right)$; *we have*

$$(x_k^*, y_k^*) \in \partial L(x_k, y_k). \tag{8}$$

For all bounded subsequence $(x_{k'}, y_{k'})$ *of* (x_k, y_k) *we have* $(x_{k'}^*, y_{k'}^*) \to 0$, $k' \to +\infty$, *hence* $\operatorname{dist}(0, \partial L(x_{k'}, y_{k'})) \to 0$, $k' \to +\infty$.

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

By the very definition of x_k, and Remark 1 c), 0 must lie in the subdifferential at point x_k of the function $\xi \mapsto \frac{1}{2\lambda_{k-1}}\|\xi - k - 1\|^2 + L(\xi, y_{k-1})$ which is equal to $\frac{1}{\lambda_{k-1}}(\xi - x_{k-1}) + \partial_x L(x_k, y_{k-1})$ since the function $\xi \mapsto \frac{1}{2\lambda_{k-1}}\|\xi - x_{k-1}\|^2$ is smooth. Hence

$$0 \in \frac{1}{\lambda_{k-1}}(x_k - x_{k-1}) + \partial_x L(x_k, y_{k-1}), \ \forall k \geq 1. \tag{9}$$

And similarly

$$0 \in \frac{1}{\mu_{k-1}}(y_k - y_{k-1}) + \partial_y L(x_k, y_k), \ \forall k \geq 1. \tag{10}$$

Due to the structure of L we have $\partial_x L(x_k, y_{k-1}) = \partial f(x_k) + \nabla_x Q(x_k, y_{k-1})$ and $\partial_y L(x_k, y_k) = \partial g(y_k) + \nabla_y Q(x_k, y_k)$. Hence we may write with (9, 10)

$$-\tfrac{1}{\lambda_{k-1}}(x_k - x_{k-1}) - (\nabla_x Q(x_k, y_{k-1}) - \nabla_x Q(x_k, y_k), 0) \in \partial f(x_k) + \nabla_x Q(x_k, y_k);$$

$$-\tfrac{1}{\mu_{k-1}}(y_k - y_{k-1}) \in \partial g(y_k) + \nabla_y Q(x_k, y_k).$$

This yields (8) with proposition 3.

If $(x_{k'}, y_{k'})$ is a bounded sequence, then so is $(x_{k'}, y_{k'-1})$, and, by (ii), $(x_{k'}, y_{k'}) - (x_{k'}, y_{k'-1})$ vanishes as $k' \to \infty$. The uniform continuity of $\nabla_x Q$ on bounded subsets then yields the last point of (iii). □

Remark 3 (a) Without additional assumptions, like for instance the convexity of L, the sequence (x_k, y_k) is not *a priori* uniquely defined.
(b) Note also that the result remains valid if L is bounded from below by an affine function.

The next proposition, especially points (ii)(iii), gives the first convergence results about sequences generated by (5,6). Theorems 9 and 11 below make the convergence properties much more precise.

Proposition 6 *Assume that $(\mathcal{H})$, $(\mathcal{H}_1)$ hold. Let (x_k, y_k) be a sequence complying with (5) and (6). Let $\omega(x_0, y_0)$ denote the set (possibly empty) of its limit points. Then*
(i) If (x_k, y_k) is bounded, then $\omega(x_0, y_0)$ is a nonempty compact connected set and

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

(ii) $\omega(x_0, y_0) \subset \operatorname{crit} L$,
(iii) L is finite and constant on $\omega(x_0, y_0)$, equal to $\inf_{k \in \mathbb{N}} L(x_k, y_k) = \lim_{k \to +\infty} L(x_k, y_k)$.

Proof. Item (i) follows by using $\|x_k - x_{k-1}\| + \|y_k - y_{k-1}\| \to 0$ together with some classical properties of sequences in $\mathbb{R}^n$.
(ii) By the very definition of (x_k, y_k) $(k \geq 1)$ we have

$$\begin{aligned} L(x_k, y_{k-1}) + \tfrac{1}{2\lambda_{k-1}}\|x_k - x_{k-1}\|^2 &\leq L(\xi, y_{k-1}) + \tfrac{1}{2\lambda_{k-1}}\|\xi - x_{k-1}\|^2, \quad \forall \xi \in \mathbb{R}^m \\ L(x_k, y_k) + \tfrac{1}{2\mu_{k-1}}\|y_k - y_{k-1}\|^2 &\leq L(x_k, \eta) + \tfrac{1}{2\mu_{k-1}}\|\eta - y_{k-1}\|^2, \quad \forall \eta \in \mathbb{R}^n. \end{aligned}$$

Due to the special form of L and to $0 < r_- \leq \lambda_{k-1} \leq r_+$ and $0 < r_- \leq \mu_{k-1} \leq r_+$, we have

$$f(x_k) + Q(x_k, y_{k-1}) + \tfrac{1}{2r_+}\|x_k - x_{k-1}\|^2 \leq f(\xi) + Q(\xi, y_{k-1}) + \tfrac{1}{2r_-}\|\xi - x_{k-1}\|^2, \quad \forall \xi \in \mathbb{R}^m, \tag{11}$$

$$g(y_k) + Q(x_k, y_k) + \tfrac{1}{2r_+}\|y_k - y_{k-1}\|^2 \leq g(\eta) + Q(x_k, \eta) + \tfrac{1}{2r_-}\|\eta - y_{k-1}\|^2, \quad \forall \eta \in \mathbb{R}^n. \tag{12}$$

Let $(\bar{x}, \bar{y})$ be a point in $\omega(x_0, y_0)$; there exists a subsequence $(x_{k'}, y_{k'})$ of (x_k, y_k) converging to $(\bar{x}, \bar{y})$. Since $\|x_k - x_{k-1}\| + \|y_k - y_{k-1}\| \to 0$ we deduce from (11)

$$\liminf f(x_{k'}) + Q(\bar{x}, \bar{y}) \leq f(\xi) + Q(\xi, \bar{y}) + \frac{1}{2r_-}\|\xi - \bar{x}\|^2 \quad \forall \xi \in \mathbb{R}^m.$$

In particular for $\xi = \bar{x}$ we obtain

$$\liminf f(x_{k'}) \leq f(\bar{x}).$$

Since f is lower semicontinuous we get further

$$\liminf f(x_{k'}) = f(\bar{x}).$$

There is no loss of generality in assuming that the whole sequence $f(x_{k'})$ converges to $f(\bar{x})$

$$\lim f(x_{k'}) = f(\bar{x}). \tag{13}$$

Similarly, using (12), we may assume that $\lim g(y_{k'}) = g(\bar{y})$. Now, as Q is continuous, we have $Q(x_{k'}, y_{k'}) \to Q(\bar{x}, \bar{y})$ and hence $L(x_{k'}, y_{k'}) \to L(\bar{x}, \bar{y})$. But with lemma 5(iii) and using the same notation , we have $(x^*_{k'}, y^*_{k'}) \in \partial L(x_{k'}, y_{k'})$ and $(x^*_{k'}, y^*_{k'}) \to 0$. Owing to the closedness properties of ∂L, we finally obtain $0 \in \partial L(\bar{x}, \bar{y})$.

(iii) For any point $(\bar{x}, \bar{y}) \in \omega(x_0, y_0)$ we have just seen that there is a subsequence $(x_{k'}, y_{k'})$ with $L(x_{k'}, y_{k'}) \to L(\bar{x}, \bar{y})$. Since the sequence $L(x_k, y_k)$ is nonincreasing, we have $L(\bar{x}, \bar{y}) = \inf L(x_k, y_k)$ independent of $(\bar{x}, \bar{y})$. □

3.2 Kurdyka-Łojasiewicz inequality

Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a proper lower semicontinuous function. For $-\infty < \eta_1 < \eta_2 \leq +\infty$, let us set

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

Definition 7 (Kurdyka-Łojasiewicz property) *The function f is said to have the Kurdyka-Łojasiewicz property at $\bar{x} \in \operatorname{dom} \partial f$ if there exist $\eta \in (0, +\infty]$, a neighborhood ([6]) U of $\bar{x}$ and a continuous* concave *function $\varphi : [0, \eta) \to \mathbb{R}_+$ such that:*

- *$\varphi(0) = 0$,*
- *φ is C^1 on $(0, \eta)$,*
- *for all $s \in (0, \eta)$, $\varphi'(s) > 0$,*
- *and for all x in $U \cap [f(\bar{x}) < f < f(\bar{x}) + \eta]$, the Kurdyka-Łojasiewicz inequality holds*

$$\varphi'(f(x) - f(\bar{x}))\, \operatorname{dist}(0, \partial f(x)) \geq 1. \tag{14}$$

Remark 4 (a) S. Łojasiewicz proved in 1963 [39] that real-analytic functions satisfy an inequality of the above type with $\varphi(s) = s^{1-\theta}$ where $\theta \in [\frac{1}{2}, 1)$. A nice proof of this result can be found in the monograph [26]. In a recent paper, Kurdyka [35] has extended this result to differentiable functions definable in an o-minimal structure (see Section 4.3). More recently Bolte, Daniilidis, Lewis and Shiota have extended the Kurdyka-Lojasiewicz inequality to nonsmooth functions in the subanalytic and o-minimal setting [16, 17, 18]; see also [15].
The concavity assumption imposed on the function φ does not explicitly belong to the usual formulation of the Kurdyka-Lojasiewicz inequality. However the examples of the next section show that the inequality holds in many instances with a concave φ.
(b) A proper lower semicontinuous function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ has the Kurdyka-Łojasiewicz property at any noncritical point $\bar{x} \in \mathbb{R}^n$. Indeed, Lemma 2 yields the existence of $c > 0$ such that

$$\operatorname{dist}(0, \partial f(x)) \geq c > 0$$

whenever $x \in B(\bar{x}, c/2) \cap [f(\bar{x}) - c/2 < f < f(\bar{x}) + c/2]$; $\varphi(s) = c^{-1}s$ is then a suitable concave function.
(c) When a function has the Kurdyka-Łojasiewicz property, it is important to have estimations of η, U, φ. We shall see for instance that many 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.

[6]See Remark (4) (c)

3.3 Convergence to a critical point and other convergence results

This section is devoted to the convergence analysis of the proximal minimization algorithm introduced in Section 3.1. It provides the main mathematical results of this paper. Applications and examples are given in Section 4.

Related previous work may be found in [5, 6]; but there, the setting is convex with a quadratic coupling Q, and the mathematical analysis relies on the monotonicity of the convex subdifferential operators. In [4] Łojasiewicz inequality is used to derive the convergence of the usual proximal method but alternating algorithms are not considered.

Let us introduce the notations: $z_k = (x_k, y_k)$, $l_k = L(z_k)$, $\bar{z} = (\bar{x}, \bar{y})$, $\bar{l} = L(\bar{z})$.

Theorem 8 (pre-convergence result) *Assume that L satisfies $(\mathcal{H})$, $(\mathcal{H}_1)$ and has the Kurdyka-Łojasiewicz property at $\bar{z} = (\bar{x}, \bar{y})$. Denote by U, η and $\varphi : [0, \eta) :\to \mathbb{R}$ the objects appearing in (14), relative to L and $\bar{z}$. Let $\rho > 0$ be such that $B(\bar{z}, \rho) \subset U$.*

Let (z_k) be a sequence generated by the alternating algorithm (5, 6), with z_0 as an initial point. Let us assume that

$$\bar{l} < l_k < \bar{l} + \eta, \ \forall k \geq 0, \tag{15}$$

and

$$M\varphi(l_0 - \bar{l}) + 2\sqrt{2r_+}\sqrt{l_0 - \bar{l}} + \|z_0 - \bar{z}\| < \rho \tag{16}$$

with $M = 2r_+(C + 1/r_-)$ where C is a Lipschitz constant for ∇Q on $B(\bar{z}, \sqrt{2}\rho)$.

Then, the sequence (z_k) converges to a critical point of L and the following estimates hold: $\forall k \geq 0$

$$(i) \quad z_k \in B(\bar{z}, \rho) \tag{17}$$

$$(ii) \quad \sum_{i=k+1}^{\infty} \|z_{i+1} - z_i\| \leq M\varphi(l_k - \bar{l}) + \sqrt{2r_+}\sqrt{l_k - \bar{l}}. \tag{18}$$

The meaning of the theorem is roughly the following: a sequence (z_k) that starts in the neighborhood of a point $\bar{z}$ (as given in (16)) and that does not improve $L(\bar{z})$ (as given in (15)) converges to a critical point near $\bar{z}$.

Proof. The idea of the proof is in the line of Łojasiewicz' argument: it consists mainly in proving – and then using – that $\varphi \circ L$ is a Liapunov function with decreasing rate close to $||z_{k+1} - z_k||$.
With no loss of generality one may assume that $L(\bar{z}) = 0$ (replace if necessary L by $L - L(\bar{z})$).
With (7) we have for $i \geq 0$:

$$l_i - l_{i+1} \geq \frac{1}{2r_+}\|z_{i+1} - z_i\|^2. \tag{19}$$

But $\varphi'(l_i)$ makes sense in view of (15), and $\varphi'(l_i) > 0$; hence

$$\varphi'(l_i)(l_i - l_{i+1}) \geq \frac{\varphi'(l_i)}{2r_+}\|z_{i+1} - z_i\|^2.$$

Owing to φ being concave, we have further:

$$\varphi(l_i) - \varphi(l_{i+1}) \geq \frac{\varphi'(l_i)}{2r_+}\|z_{i+1} - z_i\|^2, \ \forall i \geq 0. \tag{20}$$

Let us first check (i) for $k = 0$ and $k = 1$. In view of (16), z_0 lies in $B(\bar{z}, \rho)$.
As to z_1, (19) yields in particular

$$\frac{1}{2r_+}\|z_1 - z_0\|^2 \leq l_0 - l_1 \leq l_0. \tag{21}$$

Hence:

$$\|z_1 - \bar{z}\| \leq \|z_1 - z_0\| + \|z_0 - \bar{z}\| \leq \sqrt{2r_+}\sqrt{l_0} + \|z_0 - \bar{z}\|. \tag{22}$$

Hence z_1 lies in $B(\bar{z},\rho)$ in view of (16).
Let us now prove by induction that $(x_k,y_k)\in B(\bar{z},\rho)$ for all $k\geq 0$.
This being true for $k\in\{0,1\}$, let us assume that it holds up to some $k\geq 1$.
For $0\leq i\leq k$, since $z_i\in B(\bar{z},\rho)$ and $0<l_i<\eta$, we can write the Kurdyka-Lojasiewicz inequality at z_i:

$$\varphi'(l_i)\,\mathrm{dist}\,(0,\partial L(z_i))\geq 1.$$

Recall, with lemma 5 (iii), that

$$(x_i^*,y_i^*)=-(\nabla_x Q(x_i,y_{i-1})-\nabla_x Q(x_i,y_i),0)-\left(\frac{1}{\lambda_{i-1}}(x_i-x_{i-1}),\frac{1}{\mu_{i-1}}(y_i-y_{i-1})\right)$$

is an element of $\partial L(x_i,y_i)$. Hence we have for $1\leq i\leq k$:

$$\varphi'(l_i)\|(x_i^*,y_i^*)\|\geq 1. \tag{23}$$

Let us examine $\|(x_i^*,y_i^*)\|$, for $1\leq i\leq k$. On the one hand

$$\left\|\left(\frac{1}{\lambda_{i-1}}(x_i-x_{i-1}),\frac{1}{\mu_{i-1}}(y_i-y_{i-1})\right)\right\|\leq\frac{1}{r_-}\|z_i-z_{i-1}\|.$$

On the other hand, let us observe

$$\|(x_i,y_{i-1})-(\bar{x},\bar{y})\|^2=\|x_i-\bar{x}\|^2+\|y_{i-1}-\bar{y}\|^2\leq\|z_i-\bar{z}\|^2+\|z_{i-1}-\bar{z}\|^2\leq 2\rho^2.$$

Hence (x_i,y_{i-1}), and $z_i=(x_i,y_i)$, lie in $B(\bar{z},\sqrt{2}\rho)$; which allows us to apply the Lipschitz inequality between these points

$$\|\nabla_x Q(x_i,y_i)-\nabla_x Q(x_i,y_{i-1})\|\leq C\|y_i-y_{i-1}\|\leq C\|z_i-z_{i-1}\|.$$

Hence, for $1\leq i\leq k$

$$\|(x_i^*,y_i^*)\|\leq(C+1/r_-)\|z_i-z_{i-1}\|. \tag{24}$$

Now (23) yields

$$\varphi'(l_i)\geq\frac{1}{(C+1/r_-)}\|z_i-z_{i-1}\|^{-1},\ 1\leq i\leq k.$$

And (20) yields

$$\varphi(l_i)-\varphi(l_{i+1})\geq\frac{1}{M}\frac{\|z_{i+1}-z_i\|^2}{\|z_i-z_{i-1}\|},\ 1\leq i\leq k.$$

We rewrite the previous inequality in the following way

$$\|z_i-z_{i-1}\|^{1/2}\left(M(\varphi(l_i)-\varphi(l_{i+1}))\right)^{1/2}\geq\|z_{i+1}-z_i\|.$$

Hence (recall $ab\leq(a^2+b^2)/2$)

$$\|z_i-z_{i-1}\|+M(\varphi(l_i)-\varphi(l_{i+1}))\geq 2\|z_{i+1}-z_i\|. \tag{25}$$

This inequality holds for $1\leq i\leq k$; let us sum over i

$$\|z_1-z_0\|+M(\varphi(l_1)-\varphi(l_{k+1}))\geq\sum_{i=1}^{k}\|z_{i+1}-z_i\|+\|z_{k+1}-z_k\|.$$

Hence, in view of the monotonicity properties of φ and l_k

$$\|z_1-z_0\|+M\varphi(l_0)\geq\sum_{i=1}^{k}\|z_{i+1}-z_i\|.$$

We finally get

$$\|z_{k+1} - \bar z\| \le \sum_{i=1}^{k} \|z_{i+1} - z_i\| + \|z_1 - \bar z\| \le M\varphi(l_0) + \|z_1 - z_0\| + \|z_1 - \bar z\|;$$

which entails $z_{k+1} \in B(\bar z, \rho)$ in view of (21, 22) and (16). That completes the proof of (i).
Indeed inequality (25) holds for $i \ge 1$; let us sum it for i running from some k to some $K > k$

$$\|z_k - z_{k-1}\| + M(\varphi(l_k) - \varphi(l_{K+1})) \ge \sum_{i=k}^{K} \|z_{i+1} - z_i\| + \|z_{K+1} - z_K\|.$$

Hence

$$\|z_k - z_{k-1}\| + M\varphi(l_k) \ge \sum_{i=k}^{K} \|z_{i+1} - z_i\|.$$

Letting $K \to \infty$ yields

$$\sum_{i=k}^{\infty} \|z_{i+1} - z_i\| \le M\varphi(l_k) + \|z_k - z_{k-1}\|. \tag{26}$$

We conclude with (19), and that will prove point (ii):

$$\sum_{i=k}^{\infty} \|z_{i+1} - z_i\| \le M\varphi(l_k) + \sqrt{2r_+}\sqrt{l_{k-1}} \le M\varphi(l_{k-1}) + \sqrt{2r_+}\sqrt{l_{k-1}}.$$

This clearly implies that (z_k) is a convergent sequence. As a consequence of Proposition 6, we obtain that its limit is a critical point of L. □

This theorem has two important consequences

Theorem 9 (convergence) *Assume that L satisfies $(\mathcal{H})$, $(\mathcal{H}_1)$ and has the Kurdyka-Lojasiewicz property at each point of the domain of f.*

Then:
– either $||(x_k, y_k)||$ tends to infinity
– or $(x_k - x_{k-1}, y_k - y_{k-1})$ is l^1, i.e.

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

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

Proof. Assume that $\|(x_k, y_k)\|$ does not tend to infinity and let $\bar z$ be a limit-point of (x_k, y_k) for which we denote by ρ, η, φ the associated objects as defined in (14). Note that Proposition 6 implies that $\bar z$ is critical and that $l_k = L(x_k, y_k)$ converges to $L(\bar z)$.

If there exists an integer k_0 for which $L(x_{k_0}, y_{k_0}) = L(\bar z)$, it is straightforward to check (recall (7)) that $(x_k, y_k) = (x_{k_0}, y_{k_0})$ for all $k \ge k_0$, so that $(x_{k_0}, y_{k_0}) = \bar z$. We may thus assume that $L(x_k, y_k) > L(\bar z)$.

Since $\max(\varphi(l_k - L(\bar z)), \|z_k - \bar z\|)$ admits 0 as a cluster point, we obtain the existence of $k_0 \ge 0$ such that (16) is fulfilled with z_{k_0} as a new initial point. The conclusion is then a consequence of Theorem 8. □

Remark 5 (a) Several standard assumptions automatically guarantee the boundedness of the sequence (x_k, y_k), hence its convergence:

– *One-sided coercivity implies convergence.* Assuming that f (or g) has compact lower-level sets and that $Q(x, y) = \frac{1}{2}||x - y||^2$, implies that the sequence (x_k, y_k) is bounded (use Lemma 5).

– *Convexity implies convergence.* Assume that f, g are convex and that Q is of the form $Q(x, y) = ||Ax - By||$ where $A : \mathbb{R}^m \to \mathbb{R}^p$ and $B : \mathbb{R}^n \to \mathbb{R}^p$ are linear mappings. If L has at least a minimizer,

then the sequence (x_k, y_k) is bounded (see [5]).
(b) Theorem 9 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, see next sections), but it also relaxes the quadraticity assumption on Q that is required by the alternating minimization algorithm of [5, 6].

The following result should not be considered as a classical result of local convergence: we do not assume here any standard nondegeneracy conditions like for instance uniqueness of the minimizers, second-order conditions or transversality conditions.

Theorem 10 (local convergence to global minima) *Assume that L satisfies $(\mathcal{H})$, $(\mathcal{H}_1)$ and has the Kurdyka-Łojasiewicz property at $(\bar{x}, \bar{y})$, a global minimum point of L. Then there exist ϵ, η such that*

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

implies that the sequence (x_k, y_k) starting from (x_0, y_0) has the finite length property and converges to (x^*, y^*) with $L(x^*, y^*) = \min L$.

Proof. A straightforward application of Theorem 8 yields the convergence of (x_k, y_k) to some (x^*, y^*), a critical point of L with $L(x^*, y^*) \in [\min L, \min L + \eta)$. Now, if $L(x^*, y^*)$ were not equal to $L(\bar{x}, \bar{y})$ then the Kurdyka-Lojasiewicz inequality would entail $\varphi'(L(x^*, y^*) - L(\bar{x}, \bar{y}))\, \mathrm{dist}\,(0, \partial L(x^*, y^*)) \geq 1$, a clear contradiction since $0 \in \partial L(x^*, y^*)$. □

The convergence rate result that follows is motivated by the analysis of problems involving semialgebraic or subanalytic data. These are very common problems (see Section 4).

Theorem 11 (rate of convergence) *Assume that L satisfies $(\mathcal{H})$, $(\mathcal{H}_1)$. Assume further that (x_k, y_k) converges to (x_∞, y_∞) and that L has the Kurdyka-Łojasiewicz property at (x_∞, y_∞) with $\varphi(s) = cs^{1-\theta}$, $\theta \in [0, 1)$, $c > 0$. Then the following estimations hold*

(i) If $\theta = 0$ then 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 $\tau \in [0, 1)$ such that

$$\|(x_k, y_k) - (x_\infty, y_\infty)\| \leq c\, \tau^k.$$

(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 Theorem 10 and for simplicity we assume that $l_k \to 0$. Then $L(x_\infty, y_\infty) = 0$ (prop. 6(iii)).

(i) Assume first $\theta = 0$. If (l_k) is stationary, then so is (x_k, y_k) in view of lemma 5(i). If (l_k) is not stationary, then the Kurdyka-Lojasiewicz inequality yields for any k sufficiently large $c\, \mathrm{dist}\,(0, \partial L(x_k, y_k)) \geq 1$, a contradiction in view of lemma 5(iii).

(ii,iii) Assume $\theta > 0$. 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 9. Since $\Delta_k \geq \sqrt{\|x^k - x_\infty\|^2 + \|y_k - y_\infty\|^2}$, it is sufficient to estimate Δ_k. The following is a rewriting of (26) with these notations

$$\Delta_k \leq M\varphi(l_k) + (\Delta_{k-1} - \Delta_k). \tag{27}$$

The Kurdyka-Łojasiewicz inequality successively yields

$$\varphi'(l_k)\mathrm{dist}\,(0, \partial L(x_k, y_k)) = c(1-\theta) l_k^{-\theta} \mathrm{dist}\,(0, \partial L(x_k, y_k)) \geq 1$$
$$l_k^{\theta} \leq c(1-\theta)\mathrm{dist}\,(0, \partial L(x_k, y_k)).$$

But with (24) we have

$$\operatorname{dist}(0, \partial L(x_k, y_k)) \leq \|(x_k^*, y_k^*)\| \leq (C + 1/r_-)(\Delta_{k-1} - \Delta_k).$$

Combining the previous two inequalities we obtain for some positive K

$$\varphi(l_k) = c l_k^{1-\theta} \leq K(\Delta_{k-1} - \Delta_k)^{\frac{1-\theta}{\theta}}.$$

Finally (27) gives

$$\Delta_k \leq MK(\Delta_{k-1} - \Delta_k)^{\frac{1-\theta}{\theta}} + (\Delta_{k-1} - \Delta_k).$$

Sequences satisfying such inequalities have been studied in [4, Theorem 2]. Items (ii) and (iii) follow from these results. □

3.4 Convergence of alternating projection methods

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

$$L_{C,D}(x, y) = \delta_C(x) + \frac{1}{2}\|x - y\|^2 + \delta_D(y), \; (x, y) \in \mathbb{R}^n, \tag{28}$$

where C, D are two nonempty closed subsets of $\mathbb{R}^n$. Notice that L satisfies $(\mathcal{H})$ and $(\mathcal{H}_1)$ (see Section 3.1) for any $y_0 \in D$. In this specific setting, the proximal minimization algorithm (5, 6) reads

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

Thus we obtain the following *alternating projection algorithm*

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

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

Corollary 12 (convergence of sequences) *Assume that the bifunction $L_{C,D}$ has the Kurdyka-Lojasiewicz property at each point. Then either $\|(x_k, y_k)\| \to \infty$ as $k \to \infty$, or (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 9 while the second follows from a specialization of Theorem 10 to $L_{C,D}$. □

Observe that for λ_k and μ_k large, the method is very close to the von Neuman alternating projection method. Section 4.6 lists a wide class of sets for which convergence of the sequence is ensured.

4 Examples and applications

A first simple theoretical result which deserves to be mentioned is that a generic smooth function satisfies the Łojasiewicz inequality. Recall that a C^2 function $f : \mathbb{R}^n \to \mathbb{R}$ 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 $\mathbb{R}^n$. Morse functions can be shown to be generic in the Baire sense in the space of C^2 functions (see [8]).

Let $\bar{x}$ be a critical point of f, a Morse function. 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

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

whenever $x \in U$. It is then straightforward to see that f complies with (14) with a function φ of the form $\varphi(s) = c\sqrt{s}$, where c is a positive constant.

Let us now come to more practical aspects and to several concrete illustrations of the Kurdyka-Lojasiewicz inequality (14).

4.1 Convex examples

A general smooth convex function may not satisfy the Kurdyka-Łojasiewicz (see [15] for a counterexample), however, in many practical cases convex functions do satisfy this inequality.

Growth condition for convex functions: Consider a convex function f satisfying the following growth condition: $\exists U$ neighborhood of $\bar{x}$, $\eta > 0$, $c > 0$, $r \ge 1$ such that

$$\forall x \in U \cap [\min f < f < \min f + \eta], \ f(x) \ge f(\bar{x}) + cd(x, \operatorname{argmin} f)^r,$$

where $\bar{x} \in \operatorname{argmin} f \neq \emptyset$. Then f complies with (14) at point $\bar{x}$ (for $\varphi(s) = r\, c^{-\frac{1}{r}}\, s^{\frac{1}{r}}$) on $U \cap [\min f < f < \min f + \eta]$ (see [16]).

Uniform convexity: If f is uniformly convex i.e., satisfies

$$f(y) \ge f(x) + \langle x^*, y - x\rangle + K\|y - x\|^p, \ p \ge 1$$

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

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

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

The conclusion follows from the preceding paragraph.

Tame convex functions: Another important class of functions which is often met in practice is the class of convex functions which are definable in an o-minimal structure (e.g. semialgebraic or globally subanalytic convex functions). In a tame setting the Kurdyka-Łojasiewicz inequality does not involve any convexity assumptions, the reader is thus referred to section 4.3 for an insight into this kind of results.

Among the huge literature devoted to convex problems, one may consult [10, 22, 20, 24] and their references.

4.2 Metrically regular equations

Let the mapping $F : \mathbb{R}^n \to \mathbb{R}^m$ be metrically regular at some point $\bar{x} \in \mathbb{R}^n$, namely (cf. [45, 33]) there exist a neighborhood V of $\bar{x}$ in $\mathbb{R}^n$, a neighborhood W of $F(\bar{x})$ in $\mathbb{R}^m$ and a positive k such that

$$x \in V, \ y \in W \ \Rightarrow \ \operatorname{dist}(x, F^{-1}(y)) \le k \operatorname{dist}(y, F(x)).$$

The coefficient k is a measure of stability under perturbation of equations of the type $F(\bar{x}) = \bar{y}$.

Let $C \subseteq \mathbb{R}^m$ and consider the problem of finding a point $x \in \mathbb{R}^n$ satisfying $F(x) \in C$, so that we are concerned with metric regularity in constraint systems.

For the sake of simplicity, we assume that F is C^1 on a neighborhood of $\bar{x}$ and that C is convex, closed, nonvoid.

Solving our constraint system amounts to solving

$$\min \left\{ f(x) := \frac{1}{2} \text{dist}^2(F(x), C) : x \in \mathbb{R}^n \right\},$$

so that f is now the object under study. The function f is differentiable around $\bar{x}$ with gradient

$$\nabla f(x) = D^* F(x)[F(x) - P_C(F(x))],$$

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

The metric regularity of F at $\bar{x}$ entails that $D^*F(\bar{x}) : \mathbb{R}^m \mapsto \mathbb{R}^n$ is one-to-one, with $\|[D^*F(x)]^{-1}\| \leq k$ ([45, Th. 9.43][33, Ch. 3, Th. 3]). Fix some $k' > k$. In view of the continuity of $\|[D^*F(x)]^{-1}\|$ around $\bar{x}$, there exists a neighborhood U of $\bar{x}$ such that: $x \in U \;\Rightarrow\; \|[D^*F(x)]^{-1}\| \leq k'$. We then deduce from the expression above for $\nabla f(x)$

$$x \in U \;\Rightarrow\; \|\nabla f(x))\| \geq \frac{1}{k'} \|F(x) - P_C(F(x))\| = \frac{1}{k'} \text{dist}\,(F(x), C) = \frac{1}{k'} \sqrt{2f(x)}.$$

Hence

$$x \in U,\; f(\bar{x}) < f(x) \;\Rightarrow \|\nabla f(x))\| \geq \frac{1}{k'} \sqrt{2(f(x) - f(\bar{x}))}.$$

In other words, f satisfies the Kurdyka-Łojasiewicz inequality at all point $\bar{x}$ where F is metrically regular, with $\varphi : s \in [0, +\infty[\mapsto k'\sqrt{2s}$.

It is important to observe here that 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 $\mathbb{R}^n$.

For links between metric regularity and the Kurdyka-Lojasiewicz inequality see [15].

4.3 Tame functions

As it was emphasized in the introduction, tame sets and functions provide a vast field of applications of our main results.

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

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

where p_i, q_i are real polynomial functions.

A function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is semialgebraic if its graph is a semialgebraic subset of $\mathbb{R}^{n+1}$. Such a function satisfies the Kurdyka-Łojasiewicz property (see [16, 18]) with $\varphi(s) = cs^{1-\theta}$, for some $\theta \in [0, 1) \cap \mathbb{Q}$ and some $c > 0$. This nonsmooth result generalizes the famous Łojasiewicz inequality for real-analytic functions [39]. Stability properties of semialgebraic functions are numerous (see e.g. [12, 14]), the following few facts might help the reader to understand how they impact Optimization matters:

– finite sums and products of semialgebraic functions are semialgebraic;
– scalar products are semialgebraic;
– indicator functions of semialgebraic sets are semialgebraic;
– generalized inverse of semialgebraic mappings are semialgebraic;
– composition of semialgebraic functions or mappings are semialgebraic;

– functions of the type $\mathbb{R}^n \ni x \to f(x) = \sup_{y\in C} g(x,y)$ (resp. $\mathbb{R}^n \ni x \to f(x) = \inf_{y\in C} g(x,y)$) where g and C are semialgebraic are semialgebraic.

Matrix theory provides a long list of semialgebraic objects (see [36]): positive semidefinite matrices, Stiefel manifolds (spheres, orthogonal group; see e.g. [29]), constant rank matrices... Let us provide now some concrete problems for which our algorithm can be implemented effectively.

Example 1 (Rank reduction of correlation matrices) The following is a standard problem: being given a symmetric matrix A and an integer $d \in \{1,\ldots,n-1\}$, we wish to find a correlation matrix Z of small rank and as close as possible to A. An abstract formulation of the problem leads to the following

$$\min\{F(Z-A) : Z \in \mathbb{S}_n^+,\, \operatorname{diag} Z = I_n,\, \operatorname{rank} Z \le d\},$$

where $F : \mathbb{R}^{n\times n} \to \mathbb{R}$ is a smooth semialgebraic function (*e.g.* a seminorm), $\mathbb{S}_n^+$ is the cone of positive semidefinite n matrices and I_n is the $n \times n$ identity matrix. The reformulation of the rank reduction problem for correlation matrices we use here, is due to Grubišić-Pietersz [32] who transform the original problem into the minimization of a functional on a matrix manifold (for more on optimization on matrix manifolds see [2]) . Let Y_i, $i \in \{1,\ldots,n\}$, denote line i of $Y \in \mathbb{R}^{n\times d}$. Consider the following manifold of $\mathbb{R}^{n\times d}$

$$\mathcal{C} \;=\; \left\{Y \in \mathbb{R}^{n\times d} : Y_i \in S^{i-1} \times \{0\}^{d-i},\, \forall i \in \{1,\ldots,d\};\, Y_i \in S^{d-1},\, \forall i \in \{d+1,\ldots,n\}\right\},$$

where $S^0 = \{1\}$ and S^{i-1} denotes the unit sphere of $\mathbb{R}^i$ for $i \in \{2,\ldots,d\}$. The set $\mathcal{C}$ is a real-algebraic manifold, and, due to its simple structure (it is a product of low dimensional spheres), the projection operator onto $\mathcal{C}$ is simply given as a product of projections on spheres. Denote by Y^T the transpose matrix of Y. Since $\{Z \in \mathbb{S}_n^+ : \operatorname{diag} Z = I_n, \operatorname{rank} Z \le d\} = \{YY^T : Y \in \mathcal{C}\}$ (see [32]), the correlation matrix rank reduction problem can be reformulated as

$$\min\,\{\, G(Y) \;:\; Y \in \mathcal{C}\,\}, \tag{29}$$

where $G(Y) = F(YY^T - A)$. In [32] this problem is addressed by Newton's method or a conjugate gradient algorithm on the manifold $\mathcal{C}$. Newton's method enjoys its usual local quadratic convergence, while theoretical results about the conjugate gradient are sparse in spite of good numerical evidence.

We propose the following relaxation of problem (29)

$$\min \left\{ \delta_{\mathcal{C}}(X) + \frac{\rho}{2} \sum_{\substack{i=1,\ldots,n\\ j=1,\ldots,d}} (X_{ij} - Y_{ij})^2 + G(Y) \;:\; X, Y \in \mathbb{R}^{n\times d} \right\}, \tag{30}$$

where $\rho > 0$ is a penalization parameter. Since the functional $L(X,Y) = \delta_{\mathcal{C}}(X) + \sum_{i,j}(X_{ij} - Y_{ij})^2 + G(Y)$ is semialgebraic, the problem (30) falls into the context of our study. Algorithm (3) applied to (30) generates a sequence (X_k, Y_k) that converges ($\mathcal{C}$ is bounded) to a critical point $(\bar{X}, \bar{Y})$ of L with $\bar{X} \in \mathcal{C}$. The algorithm reads

$$\begin{cases} X_{k+1} \in \operatorname{argmin}\left\{ \dfrac{\rho}{2} \displaystyle\sum_{\substack{i=1,\ldots,n\\ j=1,\ldots,d}} (X_{ij} - Y_{k,ij})^2 + \dfrac{1}{2\lambda_k} \sum_{\substack{i=1,\ldots,n\\ j=1,\ldots,d}} (X_{ij} - X_{k,ij})^2 \;:\; X \in C \right\} \\ Y_{k+1} \in \operatorname{argmin}\left\{ G(Y) + \dfrac{\rho}{2} \displaystyle\sum_{\substack{i=1,\ldots,n\\ j=1,\ldots,d}} (Y_{ij} - X_{k+1,ij})^2 + \dfrac{1}{2\mu_k} \sum_{\substack{i=1,\ldots,n\\ j=1,\ldots,d}} (Y_{ij} - Y_{k,ij})^2 \;:\; Y \in \mathbb{R}^{n\times d} \right\} \end{cases} \tag{31}$$

The second minimization is a proximal step for the function $Y \to \frac{\rho}{2}\sum_{i,j}(X_{k+1,ij} - Y_{ij})^2 + G(Y)$ (which may be convex, for large ρ). The first minimization reduces to n uncoupled minimizations in $\mathbb{R}^d$ which give the n lines $X_{k+1,i}$ of matrix X_{k+1}:

$$X_{k+1,i} \text{ is a minimizer of } \mathbb{R}^d \ni X_i \to \sum_{j=1,\dots,d}\left(X_{ij} - \frac{\rho\lambda_k Y_{k,ij} + X_{k,ij}}{\rho\lambda_k + 1}\right)^2,$$

where X_i is subject to the constraints: $X_i \in S^{i-1}\times\{0\}^{d-i}$ if $i \in \{1,\dots,d\}$, and $X_i \in S^{d-1}$ if $i \in \{d+1,\dots,n\}$.

Each solution line $X_{k+1,i}$ is simply the projection of line $\rho\lambda_k Y_{k,i} + X_{k,i}$ of matrix $\rho\lambda_k Y_k + X_k$ onto the unit sphere of $\mathbb{R}^d$ if $d \le i \le n$, and onto a unit sphere of lower dimension if $2 \le i \le d-1$. The second step is therefore solved in closed form.

Functions definable in an o-minimal structure over $\mathbb{R}$. Introduced in [28] these structures can be seen as an axiomatization of the qualitative properties of semialgebraic sets.

Definition 13 *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 $\mathcal{O}_{n+1}$, $\Pi(A) := \{(x_1,\dots,x_n)\in\mathbb{R}^n : (x_1,\dots,x_n,x_{n+1})\in A\}$ belongs to $\mathcal{O}_n$.

(iv) For all $i \neq j$ in $\{1,\dots,n\}$, $\{(x_1,\dots,x_n)\in\mathbb{R}^n : x_i = x_j\}\in\mathcal{O}_n$.

(v) The set $\{(x_1,x_2)\in\mathbb{R}^2 : x_1 < x_2\}$ belongs to $\mathcal{O}_2$.

(vi) 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 $f : \mathbb{R}^n \to \mathbb{R}\cup\{+\infty\}$) is said to be definable if its graph is a definable subset of $\mathbb{R}^n\times\mathbb{R}^m$ (resp. $\mathbb{R}^n\times\mathbb{R}$).

As announced in the introduction, definable sets and mappings share many properties of semialgebraic objects. Let $\mathcal{O}$ be an o-minimal structure. Then,
- finite sums of definable functions are definable;
- indicator functions of definable sets are definable;
- generalized inverses of definable mappings are definable;
- compositions of definable functions or mappings are definable;
- functions of the type $\mathbb{R}^n \ni x \to f(x) = \sup_{y\in C} g(x,y)$ (resp. $\mathbb{R}^n \ni x \to f(x) = \inf_{y\in C} g(x,y)$) where g and C are definable, are definable.

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

Theorem 14 ([18]) Any proper lower semicontinuous function $f : \mathbb{R}^n \to \mathbb{R}\cup\{+\infty\}$ which is definable in an o-minimal structure $\mathcal{O}$ has the Kurdyka-Łojasiewicz property at each point of $\operatorname{dom}\partial f$. Moreover the function φ appearing in (14) is definable in $\mathcal{O}$.

The concavity of the function φ is not stated explicitly in [18]. The proof of that fact is however elementary: it relies on the following fundamental result known as the monotonicity Lemma (see [28]).

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

Proof of Theorem 14 [concavity of φ] Let x be a critical point of f and φ a Łojasiewicz function of f at x (see (14)). 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 $\varphi'' \leq 0$ or $\varphi'' > 0$ on $(0, r)$ (apply the monotonicity lemma to φ''). If $\varphi'' > 0$ then φ' is increasing. The function φ in inequality (14) can be replaced by $\psi(s) = \varphi'(r)s$ for all $s \in (0, r)$, since $\psi'(s) = \varphi'(r) > \varphi'(s)$. $\square$.

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

Semilinear sets A subset of $\mathbb{R}^n$ is called semilinear if it is a finite union of sets of the form

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

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 $\varphi(s) = cs$ with $c > 0$.

Real semialgebraic sets By Tarski quantifier elimination theorem the class of semialgebraic sets is an o-minimal structure [12, 14].

Globally subanalytic sets (Gabrielov [31]) There exists an o-minimal structure, denoted by $\mathbb{R}_{\text{an}}$, that contains all sets of the form $\{(x, t) \in [-1, 1]^n \times \mathbb{R} : f(x) = t\}$ where $f : [-1, 1]^n \to \mathbb{R}$ ($n \in \mathbb{N}$) 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 $\varphi(s) = s^{1-\theta}$, $\theta \in [0, 1) \cap \mathbb{Q}$.

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

- 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 $g_j : \mathbb{R}^n \to \mathbb{R}$ 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 $G : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}$ be a real analytic function. When C is bounded, the restriction of the following function to each $[-a, a]^n$ ($a > 0$) is globally subanalytic:

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

Log–exp structure (Wilkie, van der Dries) [49, 28] There exists an o-minimal structure containing $\mathbb{R}_{\text{an}}$ and the graph of $\exp : \mathbb{R} \to \mathbb{R}$. 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 $x \to \exp(-\frac{1}{x^2})$.

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 Theorem 11).

Example 2 *[Compressive sensing]* A current active trend in signal recovery aims at selecting a sparse (*i.e.* with many zero components) solution of an underdetermined linear system. So the following problem comes under consideration (see [21, 19, 27])

$$\min\{\|x\|_0 : Ax = b\} \tag{32}$$

where $\|x\|_0$ is the so-called ℓ^0 "norm" which counts the nonzero components of $x \in \mathbb{R}^n$, A is an $m \times n$ matrix ($m < n$) and $b \in \mathbb{R}^m$; the set of contraints $\mathcal{C} = \{x \in \mathbb{R}^n : Ax = b\}$ is supposed nonvoid. Due to its extremely combinatorial nature, problem (32) is untractable and signal theory scientists have to turn to alternatives. We briefly present one of them below, in connection with our work.

In [21], the ℓ^0 norm is approximated by a weighted ℓ^1 norm

$$\min\left\{\sum_{i=1}^{n} w_i \mid x_i \mid : Ax = b\right\}. \tag{33}$$

For fixed positive weights w_i, problem (33) classically boils down to a linear program whose solution may well bear little relation with that of (32). So actually, in this formulation, the weights are somehow part of the unknowns and are to be chosen at best. The reweighted ℓ^1 minimization algorithm used in [21] updates, at step k, the solution candidate x^k and the weight vector $w^k = (w_1^k, \dots, w_n^k)$ according to the following rule. Fix $\varepsilon > 0$:
– w^k being known, solve (33) with w^k in place of w to obtain x^{k+1};
– update each component of w^k to $w_i^{k+1} = 1/(\mid x_i^{k+1} \mid + \varepsilon)$.

We first observe that the reweighted algorithm can be seen as an alternating minimization method. For $(x, w) \in \mathbb{R}^n \times \mathbb{R}^n$ define $f_0(x) = \delta_{\mathcal{C}}(x)$, $Q_0(x, w) = \sum_{i=1}^n (\mid x_i \mid + \varepsilon) w_i$, and $g_0(w) = -\sum_{i=1}^n \log w_i$ if $w_i > 0$ for all i in $\{1, \dots, n\}$, $g_0(w) = +\infty$ otherwise. Set $L_0(x, w) = f_0(x) + Q_0(x, w) + g_0(w)$; it is not difficult to realize that the original reweighted ℓ^1 minimization algorithm is *exactly* the alternate minimization of the function L_0.

In order to apply a *proximal* alternating method to L_0, the problem $\min L_0$ can be first reformulated as a problem with a smooth coupling term. Let us therefore split x into its nonnegative and nonpositive parts: $x'_i = \max(x_i, 0)$, $x'_{i+n} = \max(-x_i, 0)$ for $i \in \{1, \dots, n\}$. Set $\mathcal{C}' = \{x' \in \mathbb{R}^{2n} : [A, -A]x' = b, x'_i \geq 0, \forall i \in \{1, \dots, 2n\}\}$, and for $x' \in \mathbb{R}^{2n}$ define $f(x') = \delta_{\mathcal{C}'}(x')$, $Q(x', w) = \sum_{i=1}^n (x'_i + x'_{i+n} + \varepsilon) w_i$ and $L(x', w) = f(x') + Q(x', w) + g_0(w)$. Then minimizing $L_0(x, w)$ is equivalent to minimizing $L(x', w)$, with the correspondance $x_i = x'_i - x'_{i+n}$.

Since L is definable in the $\log - \exp$ structure, the proximal alternating minimization (3) applied to $L(x', w)$ yields a sequence $({x'}^k, w^k)$ whose behaviour is ruled by Theorem 9. In particular $({x'}^k, w^k)$, which is bounded because L is coercive, converges to a critical point $(\tilde{x}', \tilde{w})$ of L verifying

$$\begin{cases} 0 \in \partial_{x'}\{Q(\tilde{x}', \tilde{w}) + \delta_{\mathcal{C}'}(\tilde{x}')\}, & \text{hence} \quad \tilde{x}' \in \operatorname{argmin}\{Q(x', \tilde{w}) + \delta_{\mathcal{C}'}(x') \ : \ x' \in \mathbb{R}^{2n}\} \\ 0 = \nabla_w\{Q(\tilde{x}', \tilde{w}) + g_0(\tilde{w})\}, & \text{hence} \quad \tilde{w}_i = \dfrac{1}{\tilde{x}'_i + \tilde{x}'_{i+n} + \varepsilon}, \ \forall i \in \{1, \dots, n\} \end{cases}$$

where the minimizing property of $\tilde{x}'$ stated in the first line is a consequence of the convexity of the function $x' \to Q(x', w) + \delta_{\mathcal{C}'}(x')$ for a fixed w. Now it is easy to see that $\tilde{x}'$ must be such that $\min(\tilde{x}'_i, \tilde{x}'_{i+n}) = 0$ for $i \in \{1, \dots, n\}$. Let $\tilde{x} \in \mathbb{R}^n$ be defined by $\tilde{x}_i = \tilde{x}'_i - \tilde{x}'_{i+n}$, then $\tilde{x}$ is a solution of (33) with $w_i = \tilde{w}_i = 1/(|\tilde{x}_i| + \varepsilon)$. Observe also that L is (globally) subanalytic on bounded boxes so that its desingularizing functions are of the form $\varphi(s) = cs^\theta$ with $c > 0$ and $\theta \in (0, 1]$. Theorem 11 on the rate estimation therefore applies to the sequence $({x'}^k, w^k)$. Let us summarize the previous discussion:

A proximal reweighted ℓ^1 algorithm.
1) Fix (x'_0, w_0) in $\mathbb{R}^{2n} \times \mathbb{R}^m$. Let λ_k, μ_k be some sequences of stepsizes such that $\lambda_k, \mu_k \in (r_-, r_+)$ where r_-, r_+ are positive bounds (with $r_- < r_+$).
2) Generate a sequence $({x'}^k, w^k)$ such that

$$\textbf{(PR)} \quad \begin{cases} {x'}^{k+1} = \operatorname{argmin}\left\{\displaystyle\sum_{i=1}^{n} (x'_i + x'_{i+n} + \varepsilon) w_i^k + \frac{1}{2\lambda_k} \sum_{i=1}^{2n} (x'_i - {x'}_i^k)^2 \ : \ x' \in \mathbb{R}_+^{2n}, \ [A, -A]x' = b\right\} \\ w^{k+1} = \operatorname{argmin}\left\{\displaystyle\sum_{i=1}^{n} ({x'}_i^{k+1} + {x'}_{i+n}^{k+1} + \varepsilon) w_i - \sum_{i=1}^{n} \log w_i + \frac{1}{2\mu_k} \sum_{i=1}^{n} (w_i - w_i^k)^2 \ : \ w \in \mathbb{R}^n\right\} \end{cases}$$

Convergence of (PR). The sequence (x'_k, w_k) converges to a point (x', w) and there exist positive constants d, r such that

$$||(x'_k, w_k) - (x', w)|| \leq \frac{d}{k^r},\ \forall k \geq 1.$$

Moreover if we set $x = (x'_1 - x'_{1+n}, \ldots, x'_i - x'_{i+n}, \ldots, x'_n - x'_{2n})$, then x is solution of the reweighted minimization problem (33) with $w_i = 1/(|x_i| + \varepsilon)$.

Observe finally that the first minimization in (PR) is a convex quadratic program and that the second one may be solved in closed form. Indeed w_i^{k+1} is the positive root of the second degree equation

$$\frac{1}{\mu_k} w^2 + \left[({x'}_i^{k+1} + {x'}_{i+n}^{k+1} + \varepsilon) - \frac{w_i^k}{\mu_k} \right] w - 1 = 0.$$

The computational burden of the proximal reweighting algorithm is not much heavier than that of the reweighting algorithm.

4.4 Kurdyka-Łojasiewicz inequality for feasibility problems

Recall the function $L_{C,D}$ defined by (28). Writing down the optimality condition we obtain

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

This implies

$$\text{dist}\,(0, \partial L_{C,D}(x, y)) = \left(\text{dist}^{\,2}(y - x, N_C(x)) + \text{dist}^{\,2}(x - y, N_D(y))\right)^{\frac{1}{2}}. \tag{34}$$

4.4.1 (Strongly) regular intersection

The following result is a reformulation of [36, Theorem 17] where this fruitful concept was considered in relation with alternate projection methods. For the sake of completeness we give a proof avoiding the use of metric regularity and Mordukhovich criterion.

Proposition 15 *Let C, D be two closed subsets of $\mathbb{R}^n$ and $\bar{x} \in C \cap D$. Assume*

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

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

$$\text{dist}\,(0, \partial L_{C,D}(x, y)) \geq c\|x - y\| > 0, \tag{35}$$

whenever $(x, y) \in U \cap [0 < L_{C,D} < +\infty]$. In other words $L_{C,D}$ has the Kurdyka-Łojasiewicz property at $(\bar{x}, \bar{x})$ with $\varphi(s) = \frac{1}{c}\sqrt{2s}$.

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 (34), there exists $(u_k, v_k) \in N_C(x_k) \times N_D(y_k)$ such that

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

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 $d \in -N_D(\bar{x})$. So d satifies both $\|d\| = 1$ and $-d \in [-N_C(\bar{x})] \cap N_D(\bar{x})$, a contradiction. □

Remark 6 (a) Observe that (35) reads

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

(b) for related, but different, results see [37, 41].

4.5 Transverse Manifolds

As pointed out in [36] 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_xM denotes the tangent space to M at x. Two submanifolds of $\mathbb{R}^n$ are called transverse at $x \in M \cap N$ if they satisfy

$$T_xM + T_xN = \mathbb{R}^n.$$

Since the normal cone to M at x is given by $N_M(x) = T_xM^\perp$, we have

$$N_M(x) \cap N_N(x) = N_M(x) \cap [-N_N(x)] = \{0\},$$

whenever M and N are transverse at x. And therefore $L_{M,N}$ satisfies the Kurdyka-Łojasiewicz inequality near x with $\varphi(s) = c^{-1}\sqrt{s}$, $c > 0$.

The constant c can be estimated [36, Theorem 18].

4.6 Tame feasibility

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

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

where G is the graph of the polynomial function $(x, y) \to \frac{1}{2}||x-y||^2$. From the stability property of o-minimal structures (see Definition 13 (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 $L_{C,D}$ satisfies the Łojasiewicz inequality with $\varphi(s) = s^{1-\theta}$ (where $\theta \in (0, 1]$) we obtain

$$\big(\operatorname{dist}^2(y-x, N_C(x)) + \operatorname{dist}^2(x-y, N_D(y))\,\big)^{\frac{1}{2}} \geq \left(\frac{1}{2}\|x-y\|^2\right)^{\theta} \tag{36}$$

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

Remark 7 (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 the same o-minimal structure, we of course have a similar result. Namely,

$$\varphi'\left(\frac{1}{2}\|x-y\|^2\right)\big(\operatorname{dist}^2(y-x, N_C(x)) + \operatorname{dist}^2(x-y, N_D(y))\,\big)^{\frac{1}{2}} \geq 1, \tag{37}$$

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

Although (37) is simply a specialization of a result of [18], this form of the Kurdyka-Łojasiewicz inequality can be very useful in many contexts.

Remark 8 (a) Note that if C, D are definable in the log-exp structure then (37) 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.
(c) Many concrete problems involving definable data and for which the projection operators are easily computable are given in [36].

References

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

[2] Absil, P.-A., Mahony, R. , Sepulchre, R., Optimization Algorithms on Matrix Manifolds, Princeton University Press, 2008.

[3] Aragon, A., Dontchev, A. , Geoffroy, M., Convergence of the proximal point method for metrically regular mappings, ESAIM Proc., **17**, EDP Sci., Les Ulis, (2007), pp 1–8.

[4] Attouch, H., Bolte, J., On the convergence of the proximal algorithm for nonsmooth functions involving analytic features, Math. Program. Ser. B **116** (2009), pp. 5–16.

[5] Attouch, H., Bolte, J., Redont, P., Soubeyran, A., Alternating proximal algorithms for weakly coupled convex minimization problems. Applications to dynamical games and PDE's, J. of Convex Analysis **15** (3) (2008), pp. 485–506.

[6] Attouch, H., Redont P., Soubeyran A., A new class of alternating proximal minimization algorithms with costs to move, SIAM Journal on Optimization **18** (2007), pp. 1061-1081.

[7] Attouch, H., Soubeyran A., Inertia and reactivity in decision making as cognitive variational inequalities, Journal of Convex Analysis **13** (2006), pp. 207–224.

[8] Aubin, J.-P., Ekeland, I., *Applied Nonlinear Analysis*, Wiley-Interscience, 1984.

[9] Auslender, A., *Optimisation, Méthodes Numériques*, Masson, Paris, 1976.

[10] Bauschke, H.H., Borwein, J.M., On projection algorithms for solving convex feasibility problems, SIAM review **38** (3) (1996), pp. 367-426.

[11] H. H. Bauschke, P. L. Combettes and D. Noll, *Joint minimization with alternating Bregman proximity operators*, Pacific Journal of Optimization, vol. 2, n° 3 (2006), pp. 401-424.

[12] Benedetti, R., Risler, J.-J., *Real Algebraic and Semialgebraic Sets*, Hermann, Éditeur des Sciences et des Arts, Paris, 1990.

[13] Bertsekas, D., *Nonlinear Optimisation*, Athena, Belmont, Massachusetts, second edition, 1999.

[14] Bochnak, J., Coste, M., Roy, M.-F., *Real Algebraic Geometry*, Springer, 1998.

[15] Bolte, J., Daniilidis, A., Ley, O., Mazet, L., Characterizations of Łojasiewicz inequalities and applications: subgradient flows, talweg, convexity, Transactions of the AMS, 2009, to appear.

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

[17] Bolte, J., Daniilidis, A., Lewis, A., A nonsmooth Morse-Sard theorem for subanalytic functions, J. Math. Anal. Appl. **321** (2006), pp. 729–740.

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

[19] Bruckstein A., Donoho D., Elad M., From sparse solutions of systems of equations to sparse modeling of signals and images, SIAM Review **51** (1) (2009), pp. 34–81.

[20] Byrne, C., Censor, Y., Proximity function minimization using multiple Bregman projections, with applications to split feasibility and Kullback-Leibler distance minimization, Annals of Operations Research **105** (2001), pp. 77–98.

[21] Candès E., Wakin M., Boyd S., Enhancing sparsity by reweighting ℓ^1 minimization, J. Fourier Anal. Appl. **14** (2008), pp. 877–905.

[22] Combettes, P.L., Signal recovery by best feasible approximation, IEEE transactions on Image Processing **2** (1993), pp. 269–271.

[23] Combettes, P. L., Pennanen, T., Proximal methods for cohypomonotone operators. SIAM J. Control Optim. **43** (2004), pp. 731–742

[24] Combettes, P.L., Wajs, V. Signal recovery by proximal forward-backward splitting, Multiscale Model. Simul. **4** (2005), pp. 1168–1200.

[25] Csiszar, I., Tusnady, G., Information geometry and alternating minimization procedures, Statistics and Decisions (supplement 1) (1984), pp. 205–237.

[26] Denkowska, Z., Stasica, J., Ensembles sous-analytiques à la Polonaise, Travaux en cours **69**, Hermann, Paris 2008, to appear.

[27] Donoho, D. L., Compressed Sensing, IEEE Trans. Inform. Theory **4** (2006), pp. 1289–1306.

[28] 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.

[29] Edelman, A., Arias, A., Smith, S.T., The geometry of algorithms with orthogonality constraints, SIAM J. Matrix Anal. Appl. **20** (2) (1999), pp. 303–353.

[30] Hare, W., Sagastizábal, C., Computing proximal points of nonconvex functions, Math. Program. Ser. B **116** (2009), pp. 221–258.

[31] Gabrielov, A., Complements of subanalytic sets and existential formulas for analytic functions, Inventiones Math. **125** (1996), pp. 1–12.

[32] Grubisić I., Pietersz R., Efficient rank reduction of correlation matrices, Linear Algebra and its Applications **422** (2007), pp. 629–653.

[33] Ioffe, A.D., Metric regularity and subdifferential calculus, (Russian) Uspekhi Mat. Nauk **55** (2000), no. 3 (333), pp. 103–162; translation in Russian Math. Surveys **55** (2000), pp. 501–558.

[34] Iusem, A. N., Pennanen, T., Svaiter, B. F. Inexact variants of the proximal point algorithm without monotonicity. SIAM J. Optim. **13** (2003), no. 4, pp. 1080–1097.

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

[36] Lewis, A.S., Malick, J., Alternating projection on manifolds, Math. Oper. Res. **33** (2008), pp. 216–234.

[37] Lewis, A.S., Luke, D.R., Malick, J., Local Linear Convergence for Alternating and Averaged Nonconvex Projections, Found. Comput. Math. **9** no. 4 (2009), pp. 485-513.

[38] Lions, P.L., On the Schwarz alternating method. III. A variant for nonoverlapping subdomains in Third International Symposium on Domain Decomposition Methods for Partial Differential Equations, T. F. Chan, R. Glowinski, J. Périaux and O. Widlund, eds., SIAM (1990), pp. 202–231.

[39] 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.

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

[41] Luke R. D., Finding best approximation pairs relative to a convex and prox-regular set in a Hilbert space, SIAM J. Optim. Vol. 19, No. 2, pp. 714-739.

[42] Mordukhovich, B., *Variational analysis and generalized differentiation. I. Basic theory*, Grundlehren der Mathematischen Wissenschaften, **330**, Springer-Verlag, Berlin, 2006, xxii+579 pp.

[43] Nesterov, Yu., *Introductory lectures on convex optimization. A basic course.* Applied Optimization, **87**, Kluwer Academic Publishers, Boston, MA, 2004. xviii+236 pp.

[44] von Neumann, J., *Functional Operators*, Annals of Mathematics studies **22**, Princeton University Press, 1950.

[45] Rockafellar, R.T. , Wets, R., *Variational Analysis*, Grundlehren der Mathematischen Wissenschaften, **317**, Springer, 1998.

[46] Shiota, M., *Geometry of subanalytic and semialgebraic sets*, Progress in Mathematics **150**, Birkhäuser, Boston, 1997.

[47] Tseng, P., Convergence of a block coordinate descent method for nondifferentiable minimization, J. Optim. Theory Appl. **109** (2001), pp. 475–494.

[48] Widrow, B., Wallach, E., *Adaptive Inverse Control*, Prentice-Hall, New Jersey, 1996.

[49] 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), pp. 1051–1094.