From error bounds to the complexity of first-order descent methods for convex functions

Jérôme Bolte* Trong Phong Nguyen† Juan Peypouquet‡ Bruce W. Suter§

July 20, 2016

Dedicated to Jean-Pierre Dedieu who was of great inspiration to us.

Abstract

This paper shows that error bounds can be used as effective tools for deriving complexity results for first-order descent methods in convex minimization. In a first stage, this objective led us to revisit the interplay between error bounds and the Kurdyka-Łojasiewicz (KL) inequality. One can show the equivalence between the two concepts for convex functions having a moderately flat profile near the set of minimizers (as those of functions with Hölderian growth). A counterexample shows that the equivalence is no longer true for extremely flat functions. This fact reveals the relevance of an approach based on KL inequality. In a second stage, we show how KL inequalities can in turn be employed to compute new complexity bounds for a wealth of descent methods for convex problems. Our approach is completely original and makes use of a one-dimensional worst-case proximal sequence in the spirit of the famous majorant method of Kantorovich. Our result applies to a very simple abstract scheme that covers a wide class of descent methods. As a byproduct of our study, we also provide new results for the globalization of KL inequalities in the convex framework.

Our main results inaugurate a simple methodology: derive an error bound, compute the desingularizing function whenever possible, identify essential constants in the descent method and finally compute the complexity using the one-dimensional worst case proximal sequence. Our method is illustrated through projection methods for feasibility problems, and through the famous iterative shrinkage thresholding algorithm (ISTA), for which we show that the complexity bound is of the form $O(q^k)$ where the constituents of the bound only depend on error bound constants obtained for an arbitrary least squares objective with ℓ^1 regularization.

Key words: Error bounds, convex minimization, forward-backward method, KL inequality, complexity of first-order methods, LASSO, compressed sensing.

*Toulouse School of Economics, Université Toulouse Capitole, Manufacture des Tabacs, 21 allée de Brienne, Toulouse, France. E-mail: jerome.bolte@tse-fr.eu, Effort partially sponsored by the Air Force Office of Scientific Research, Air Force Material Command, USAF, under grant number FA9550-14-1-0056, the FMJH Program Gaspard Monge in optimization and operations research and ANR GAGA

†Toulouse School of Economics, Université Toulouse Capitole, and Department of mathematics, National University of Civil Engineering, 55 Giai Phong, Ha noi, Viet Nam. Email: phongnt@nuce.edu.vn

‡Departamento de Matemática & AM2V, Universidad Técnica Federico Santa María, Avenida España 1680, Valparaíso, Chile. E-mail:juan.peypouquet@usm.cl. Work supported by FONDECYT grant 1140829; Basal Project CMM Universidad de Chile; Millenium Nucleus ICM/FIC RC130003; Anillo Project ACT-1106; ECOS-Conicyt Project C13E03; Conicyt Redes 140183; and MathAmsud Project 15MAT-02.

§Air Force. Research Laboratory / RITB, Rome, NY, United States of America. E-mail: bruce.suter@us.af.mil

1 Overview and main results

A brief insight into the theory of error bounds. Since Hoffman's celebrated result on error bounds for systems of linear inequalities [36], the study of error bounds has been successfully applied to problems in sensitivity, convergence rate estimation, and feasibility issues. In the optimization world, the first natural extensions were made to convex functions by Robinson [59], Mangasarian [51], and Auslender-Crouzeix [8]. However, the most striking discovery came years before in the pioneering works of Łojasiewicz [44, 45] at the end of the fifties: under a mere compactness assumption, the existence of error bounds for arbitrary continuous semi-algebraic functions was provided. Despite their remarkable depth, these works remained unnoticed by the optimization community during a long period (see [48]). At the beginning of the nineties, motivated by numerous applications, many researchers started working along these lines, in quest for quantitative results that could produce more effective tools. The survey of Pang [56] provides a comprehensive panorama of results obtained around this time. The works of Luo [47, 48, 49] and Dedieu [32] are also important milestones in the theory. The recent works [39, 41, 62, 40, 10] provide even stronger quantitative results by using the powerful machinery of algebraic geometry or advanced techniques of convex optimization.

A methodology for complexity of first-order descent methods. Let us introduce the concepts used in this work and show how they can be arranged to devise a new and systematic approach to complexity. Let H be a real Hilbert space, and let $f : H \to (-\infty, +\infty]$ be a proper lower-semicontinuous convex function achieving its minimum $\min f$ so that $\operatorname{argmin} f \neq \emptyset$. In its most simple version, an *error bound* is an inequality of the form

$$\omega\big(f(x) - \min f\big) \geq \operatorname{dist}(x, \operatorname{argmin} f), \tag{1}$$

where ω is an increasing function vanishing at 0 –called here the *residual function*–, and where x may evolve either in the whole space or in a bounded set. *Hölderian* error bounds, which are very common in practice, have a simple power form

$$f(x) - \min f \geq \gamma \operatorname{dist}^p(x, \operatorname{argmin} f),$$

with $\gamma > 0$, $p \geq 1$ and thus $\omega(s) = (\frac{1}{\gamma}s)^{\frac{1}{p}}$. When functions are semi-algebraic on $H = \mathbb{R}^n$ and "regular" (for instance, continuous), the above inequality is known to hold on any compact set [44, 45], a modern reference being [15]. This property is known in real algebraic geometry under the name of *Łojasiewicz inequality.* However, since we work here mainly in the sphere of optimization and follow complexity purposes, we shall refer to this inequality as to the *Łojasiewicz error bound inequality.*

Once the question of computing constants and exponents (here γ and p) for a given minimization problem is settled (see the fundamental works [49, 39, 10, 62]), it is natural to wonder whether these concepts are connected to the complexity properties of first-order methods for minimizing f. Despite the important success of the error bound theory in several branches of optimization, we are not aware of a solid theory connecting the error bounds we consider (as defined in (1)), with the study of the complexity of general descent methods. There are, however, several works connecting error bounds with the convergence rates results of first-order methods (see e.g., [61, 50, 34, 53, 12, 28, 57]). See also the new and interesting work [40] that provides a wealth of error bounds and some applications to convergence rate analysis. An important fraction of these works involves "first-order error bounds"[1] (see [48, 50]) that are different from those we consider here.

[1]That is, involving inequalities of the type $\|\nabla f(x)\| \geq \omega(\operatorname{dist}(x, \operatorname{argmin} f))$

Our answer to the connection between complexity and "zero-order error bounds" will partially come from a related notion, also discovered by Łojasiewicz and further developed by Kurdyka in the semi-algebraic world: the *Łojasiewicz gradient inequality*. This inequality, also called Kurdyka-Łojasiewicz (KL) inequality (see [18]), asserts that for any smooth semi-algebraic function f there is a smooth concave function φ such that

$$\|\nabla\left(\varphi \circ (f - \min f)\right)(x)\| \geq 1$$

for all x in some neighborhood of the set $\operatorname{argmin} f$. Its generalization to the nonsmooth case [17, 19] has opened very surprising roads in the nonconvex world and it has allowed to perform convergence rate analyses for many important algorithms in optimization [4, 21, 35]. In a first stage of the present paper we show, when f is convex, that error bounds are equivalent to nonsmooth KL inequalities provided the residual function has a *moderate behavior* close to 0 (meaning that its derivative blows up at reasonable rate). Our result includes, in particular, all power-type examples like the ones that are often met in practice[2].

Once we know that error bounds provide a KL inequality, one still needs to make the connection with the actual complexity of first-order methods. This is probably the main contribution in this paper: to any given convex objective $f : H \to (-\infty, +\infty]$ and descent sequence of the form

(i) $f(x_k) + a\|x_k - x_{k-1}\|^2 \leq f(x_{k-1})$,

(ii) $\|\omega_k\| \leq b\|x_k - x_{k-1}\|$ where $\omega_k \in \partial f(x_k)$, $k \geq 1$,

we associate *a worst case one dimensional proximal method*

$$\alpha_k = \operatorname{argmin}\left\{\varphi^{-1}(s) + \frac{1}{2\zeta}(s - \alpha_k)^2 : s \geq 0\right\},\ \alpha_0 = \varphi^{-1}(f(x_0)),$$

where ζ is a constant depending explicitly on the triplet of positive real numbers (a, b, ℓ) where $\ell > 0$ is a Lipschitz constant of $\left(\varphi^{-1}\right)'$. Our complexity result asserts, under weak assumptions that the "1-D prox" governs the complexity of the original method through the elementary and natural inequality

$$f(x_k) - \min f \leq \varphi^{-1}(\alpha_k),\ k \geq 0.$$

Similar results for the sequence are provided. These ideas are already present in [16] and [13, Section 3.2]. The function φ^{-1} above –the inverse of a desingularizing function for f on a convenient domain– contains almost all the information our approach provides on the complexity of descent methods. As explained previously, it depends on the precise knowledge of a KL inequality and thus, in this convex setting, of an error bound. The reader familiar with second-order methods might have recognized the spirit of the majorant method of Kantorovich [37], where a reduction to dimension one is used to study Newton's method.

Deriving complexity bounds in practice: applications. Our theoretical results inaugurate a simple methodology: derive an error bound, compute the desingularizing function whenever possible, identify essential constants in the descent method and finally compute the complexity using the one-dimensional worst case proximal sequence. We consider first some classic well-posed problems: finding a point in an intersection of closed convex sets with regular intersection or uniformly convex problems, and we show how complexity of some classical methods can be obtained or recovered. We revisit the *iterative shrinkage*

[2] An absolutely crucial asset of error bounds and KL inequalities in the convex world is their global nature under a mere coercivity assumption – see Section 6.

thresholding algorithm (ISTA) applied to a least squares objective with ℓ^1 regularization [31] and we prove that its complexity is of the form $O(q^k)$ with $q \in (0,1)$ (see [53] for a pioneering work in this direction and also [42] for further geometrical insights). This result contrasts with what was known on the subject [11, 33] and suggests that many questions on the complexity of first-order methods remain open.

Theoretical aspects and complementary results. As explained before, our paper led us to establish several theoretical results and to clarify some questions appearing in a somehow disparate manner in the literature. We first explain how to pass from error bounds to KL inequality in the general setting of Hilbert spaces and vice versa, similar questions appear in [17, 41, 40]. This result is proved by considering the interplay between the contraction semigroup generated by the subdifferential function and the L^1 contraction property of this flow. These results are connected to the geometry of the residual functions ω and break down when error bounds are too flat. This is shown in Section 6 by a dimension 2 counterexample presented in [18] for another purpose.

Our investigations also led us to consider the problem of KL inequalities for convex functions, a problem partly tackled in [18]. We show how to extend convex KL inequalities from a level set to the whole space. We also show that compactness and semi-algebraicity ensure that real semi-algebraic or definable coercive convex functions are automatically KL *on the whole space*. This result has an interesting theoretical consequence in terms of complexity: *abstract descent methods for coercive semi-algebraic convex problems are systematically amenable to a full complexity analysis provided that a desingularizing function –known to exist– is explicitly computable.*

Organization of the paper. Section 2 presents the basic notation and concepts used in this paper, especially concerning elementary convex analysis, error bounds and KL inequalities. Readers familiar with the field can directly skip to Section 3, devoted to the equivalence between KL inequalities and error bounds. We also give some examples where this equivalence is explicitly exploited. Section 4 establishes complexity results using KL inequalities, while Section 5 provides illustrations of our general methodology for the ℓ^1 regularized least squares method and feasibility problems. Finally, Section 6 contains further theoretical aspects related to our main results, namely: some counterexamples to the equivalence between error bounds and KL inequalities, more insight into the relationship between KL inequalities and the length of subgradient curves, globalization of KL inequalities and related questions.

2 Preliminaries

In this section, we recall the basic concepts, notation and some well-known results to be used throughout the paper. In what follows, H is a real Hilbert space and $f : H \to (-\infty, +\infty]$ is proper, lower-semicontinuous and convex. We are interested in some properties of the function f around the set of its minimizers, which we suppose to be nonempty and denote by $\operatorname{argmin} f$ or S. We assume, without loss of generality, that $\min f = 0$.

2.1 Some convex analysis

We use the standard notation from [60] (see also [7, 58] and [52]). The *subdifferential* of f at x is defined as

$$\partial f(x) = \{u \in H : f(y) \geq f(x) + \langle u, y - x\rangle \text{ for all } y \in H\}.$$

Clearly, $\hat{x}$ minimizes f on H if, and only if, $0 \in \partial f(\hat{x})$. The *domain* of the point-to-set operator $\partial f : H \rightrightarrows H$ is $\operatorname{dom} \partial f := \{x \in H : \partial f(x) \neq \emptyset\}$. For $x \in \operatorname{dom} \partial f$, we denote by $\partial^0 f(x)$ the least-norm element of $\partial f(x)$.

The vector $\partial^0 f(x)$ exists and is unique as it is the projection of $0 \in H$ onto the nonempty closed convex set $\partial f(x)$. We have $\|\partial^0 f(x)\| = \operatorname{dist}(0, \partial f(x))$ (when x is not in $\operatorname{dom} \partial f$ we set $\|\partial^0 f(x)\| = +\infty$). We adopt the convention $s \times (+\infty) = +\infty$ for all $s > 0$.
Given $x \in H$, the function f_x, defined by

$$f_x(y) = f(y) + \frac{1}{2}\|y - x\|^2$$

for $y \in H$, has a unique minimizer, which we denote by $\operatorname{prox}_f(x)$. Using Fermat's Rule and the Moreau-Rockafellar Theorem, $\operatorname{prox}_f(x)$ is characterized as the unique solution of the inclusion $x - \operatorname{prox}_f(x) \in \partial f\left(\operatorname{prox}_f(x)\right)$. In particular, $\operatorname{prox}_f(x) \in \operatorname{dom} \partial f \subset \operatorname{dom} f \subset H$. The mapping $\operatorname{prox}_f : H \to H$ is the *proximity operator* associated to f. It is easy to prove that prox_f is Lipschitz continuous with constant 1.

Example 1 If $C \subset H$ is nonempty, closed and convex, the *indicator function* of C is the function $i_C : H \to (-\infty, \infty]$, defined by

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

It is proper, lower-semicontinuous and convex. Moreover, for each $x \in H$, $\partial i_C(x) = N_C(x)$, the *normal cone* to C at x. In turn, $\operatorname{prox}_{i_C}$ is the *projection operator* onto C, which we denote by P_C.

2.2 Subgradient curves

Consider the differential inclusion

$$\begin{cases} \dot{y}(t) \in -\partial f(y(t)), & \text{for almost all } t \text{ in } (0, +\infty) \\ y(0) = x, \end{cases}$$

where $x \in \overline{\operatorname{dom} f}$ and $y(\cdot)$ is an absolutely continuous curve in H. The main properties of this system − for the purpose of this research − are summarized in the following:

Theorem 1 (Brézis [23], Bruck [22]) *For each $x \in \overline{\operatorname{dom} f}$, there is a unique absolutely continuous curve $\chi_x : [0, \infty) \to H$ such that $\chi_x(0) = x$ and*

$$\dot{\chi}_x(t) \in -\partial f\left(\chi_x(t)\right)$$

for almost every $t > 0$. Moreover,

- *i)* $\frac{d}{dt}\chi_x(t^+) = -\partial^0 f(\chi_x(t))$ *for all* $t > 0$;
- *ii)* $\frac{d}{dt} f\left(\chi_x(t^+)\right) = -\|\dot{\chi}_x(t^+)\|^2$ *for all* $t > 0$;
- *iii) For each* $z \in S$, *the function* $t \mapsto \|\chi_x(t) - z\|$ *decreases;*
- *iv) The function* $t \mapsto f(\chi_x(t))$ *is nonincreasing and* $\lim_{t\to\infty} f(\chi_x(t)) = \min f$;
- *v)* $\chi_x(t)$ *converges weakly to some* $\hat{x} \in S$, *as* $t \to \infty$.

The proof of the result above is provided in [23], except for part v), which was proved in [22]. The trajectory $t \mapsto \chi_x(t)$ is called a *subgradient curve.*

2.3 Kurdyka-Łojasiewicz inequality

In this subsection, we present the nonsmooth Kurdyka-Łojasiewicz inequality introduced in [17] (see also [19, 18], and the fundamental works [43, 38]). To simplify the notation, we write $[f < \mu] = \{x \in H : f(x) < \mu\}$ (similar notation can be guessed from the context). Let $r_0 > 0$ and set

$$\mathcal{K}(0, r_0) = \left\{ \varphi \in C^0[0, r_0) \cap C^1(0, r_0),\ \varphi(0) = 0,\ \varphi \text{ is concave and } \varphi' > 0 \right\}.$$

The function f satisfies the *Kurdyka-Łojasiewicz* (KL) *inequality* (or has the KL *property*) locally at $\bar{x} \in \operatorname{dom} f$ if there exist $r_0 > 0$, $\varphi \in \mathcal{K}(0, r_0)$ and $\varepsilon > 0$ such that

$$\varphi'\left(f(x) - f(\bar{x})\right) \operatorname{dist}(0, \partial f(x)) \geq 1$$

for all $x \in B(\bar{x}, \varepsilon) \cap [f(\bar{x}) < f(x) < f(\bar{x}) + r_0]$. We say φ is a *desingularizing function* for f at $\bar{x}$. This property basically expresses the fact that a function can be made sharp by a reparameterization of its values.

If $\bar{x}$ is not a minimizer of f, the KL inequality is obviously satisfied at $\bar{x}$. Therefore, we focus on the case when $\bar{x} \in S$. Since $f(\bar{x}) = 0$, the KL inequality reads

$$\varphi'\left(f(x)\right) \|\partial^0 f(x)\| \geq 1 \tag{2}$$

for $x \in B(\bar{x}, \varepsilon) \cap [0 < f < r_0]$. The function f has the KL property on S if it does so at each point of S.

The *Łojasiewicz gradient inequality* corresponds to the case when $\varphi(s) = cs^{1-\theta}$ for some $c > 0$ and $\theta \in [0, 1)$. Following Łojasiewicz original presentation, (2) can be reformulated as follows

$$\|\partial^0 f(x)\| \geq c'\, f(x)^\theta,$$

where $c' = [(1-\theta)c]^{-1}$. The number θ is the *Łojasiewicz exponent*. If f has the KL property and admits the same desingularizing function φ at *every point*, then we say that φ is a *global* desingularizing function for f.

KL inequalities were developed within the fascinating world of real semi-algebraic sets and functions. For that subject, we refer the reader to the book [15] by Bochnak-Coste-Roy.

We recall the following theorem on the nonsmooth KL inequality (which follows the pioneering works of Łojasiewicz [43] and Kurdyka [38]). It is one of the cornerstones of the present research:

Theorem 2 (Bolte-Daniilidis-Lewis [17]) (*Nonsmooth KL inequality*) *If $f : \mathbb{R}^n \to (-\infty, +\infty]$ is proper, convex, lower-semicontinuous and semi-algebraic*[3]*, then it has the KL property around each point in* $\operatorname{dom} f$.

Under an additional coercivity assumption, a global result is provided in Subsection 6.3.

2.4 Error bounds

Consider a nondecreasing function $\omega : [0, +\infty[\to [0, +\infty[$ with $\omega(0) = 0$. The function f satisfies a local error bound with *residual function* ω if there is $r_0 > 0$ such that

$$(\omega \circ f)(x) \geq \operatorname{dist}(x, S)$$

for all $x \in [0 \leq f \leq r_0]$ (recall that $\min f = 0$). Of particular importance is the case when $\omega(s) = \gamma^{-1} s^{\frac{1}{p}}$ with $\gamma > 0$ and $p \geq 1$, namely:

$$f(x) \geq \gamma \operatorname{dist}(x, S)^p$$

[3]If *semi-algebraic* is replaced by *subanalytic* or *definable*, we obtain the same results.

for all $x \in [0 \le f \le r_0]$.

If f is convex lower semicontinuous, we can extend the error bound beyond $[0 \le f \le r_0]$ by linear extrapolation. More precisely, let $x \in \operatorname{dom} f$ such that $f(x) > r_0$. Then f is continuous on the segment $[x, P_S(x)]$. Therefore, there is $x_0 \in [x, P_S(x)]$ such that $f(x_0) = r_0$. By convexity, we have

$$\frac{f(x) - 0}{\operatorname{dist}(x, S)} \ge \frac{f(x_0) - 0}{\operatorname{dist}(x_0, S)} \ge r_0 \left(\frac{\gamma}{r_0}\right)^{\frac{1}{p}}.$$

It follows that

$$\begin{array}{rcll} f(x) & \ge & \gamma \operatorname{dist}(x, S)^p & \text{for} \quad x \in [0 \le f \le r_0], \\ f(x) & \ge & r_0^{\frac{p-1}{p}} \gamma^{\frac{1}{p}} \operatorname{dist}(x, S) & \text{for} \quad x \notin [0 \le f \le r_0]. \end{array}$$

This entails that

$$f(x) + f(x)^{\frac{1}{p}} \ge \gamma_0 \operatorname{dist}(x, S)$$

for all $x \in H$, where $\gamma_0 = \left(1 + r_0^{\frac{p-1}{p}}\right) \gamma^{\frac{1}{p}}$. This is known in the literature as a global *Hölder-type* error bound (see [39]). Observe that it can be put under the form $\omega(f(x)) \ge \operatorname{dist}(x, S)$ by simply setting $\omega(s) = \frac{1}{\gamma_0}(s + s^{\frac{1}{p}})$. When combined with the Łojasiewicz error bound inequality [44, 45], the above remark implies immediately the following result:

Theorem 3 (Global error bounds for semi-algebraic coercive convex functions) *Let $f : \mathbb{R}^n \to (-\infty, +\infty]$ be proper, convex, lower-semicontinuous and semi-algebraic, and assume that* $\operatorname{argmin} f$ *is nonempty and compact. Then f has a global error bound*

$$f(x) + f(x)^{\frac{1}{p}} \ge \gamma_0 \operatorname{dist}(x, \operatorname{argmin} f), \forall x \in \mathbb{R}^n,$$

where $\gamma_0 > 0$ and $p \ge 1$ is a rational number.

3 Error bounds with moderate growth are equivalent to Łojasiewicz inequalities

In this section, we establish a general equivalence result between error bounds and KL inequalities. Our main goal is to provide a simple and natural way of explicitly computing Łojasiewicz exponents and, more generally, desingularizing functions. To avoid perturbing the flow of our general methodology on complexity, we discuss limitations and extensions of our results later, in Section 6.

As shown in Section 4, KL inequalities allow us to derive complexity bounds for first-order methods. However, KL inequalities with known constants are in general difficult to establish while error bounds are more tractable (see e.g., [39] and references therein). The fact that these two notions are equivalent opens a wide range of possibilities when it comes to analyzing algorithm complexity.

3.1 Error bounds with moderate residual functions and Łojasiewicz inequalities

Moderate residual functions. Error bounds often have a power or Hölder-type form (see e.g. [48, 47, 49, 39, 54, 62]). They can be either very simple $s \to as^p$ or exhibit two regimes, like for instance, $s \to as^p + bs^q$. In any cases, for all concrete instances we are aware of, residual functions are systematically semi-algebraic or of "power-type". In this paper, we introduce a category of functions that allows to encompass these semi-algebraic cases and even more singular ones into a simple and appealing framework.

A function $\varphi : [0, r) \to \mathbb{R}$ in $C^1(0, r) \cap C^0[0, r)$ and vanishing at the origin, has a *moderate behavior (near the origin)* if it satisfies a differential equation of the type

$$s\varphi'(s) \geq c\varphi(s),\ \forall s \in (0, r),$$

where c is a positive constant (observe that by concavity one has necessarily $c \leq 1$). A pretty direct use of the Puiseux Lemma (see [15]) shows:

Lemma 4 *If $\varphi : [0, r) \to \mathbb{R}$ in $C^1(0, r) \cap C^0[0, r)$, vanishes at the origin and is semi-algebraic or subanalytic then it has a moderate behavior.*

The following theorem asserts that if φ has a moderate behavior, f has the global KL property if, and only if, f has a global error bound. Besides, the desingularizing function in the KL inequality and the residual function in the error bound are essentially the same, up to a multiplicative constant. As explained through a counterexample in subsection 6.3, the equivalence breaks down if one argues in a setting where the derivative φ can blow up faster. This result is related to results obtained in [18, 17, 29, 41, 40] and also shares some common techniques.

Theorem 5 (Characterization of Łojasiewicz inequalities for convex functions) *Let $f : H \to (-\infty, +\infty]$ be a proper, convex and lower-semicontinuous, with* $\min f = 0$. *Let $r_0 > 0$, $\varphi \in \mathcal{K}(0, r_0)$, $c > 0$, $\rho > 0$ and $\bar{x} \in \operatorname{argmin} f$.*

(i) [KL inequality implies error bounds] *If $\varphi'(f(x))\, \|\partial^0 f(x)\| \geq 1$ for all $x \in [0 < f < r_0] \cap B(\bar{x}, \rho)$, then* $\operatorname{dist}(x, S) \leq \varphi(f(x))$ *for all $x \in [0 < f < r_0] \cap B(\bar{x}, \rho)$.*

(ii) [Error bounds implies KL inequality] *Conversely, if $s\varphi'(s) \geq c\varphi(s)$ for all $s \in (0, r_0)$ (φ has a moderate behavior), and $\varphi(f(x)) \geq \operatorname{dist}(x, S)$ for all $x \in [0 < f < r_0] \cap B(\bar{x}, \rho)$, then $\varphi'(f(x))\, \|\partial^0 f(x)\| \geq c$ for all $x \in [0 < f < r_0] \cap B(\bar{x}, \rho)$.*

Proof. (i) Recall that the mapping $[0, +\infty) \times \overline{\operatorname{dom} f} \ni (t, x) \to \chi_x(t)$ denotes the semiflow associated to $-\partial f$ (see previous section). Since f satisfies Kurdyka-Łojasiewicz inequality, we can apply Theorem 27 of Section 6, to obtain

$$\|\chi_x(t) - \chi_x(s)\| \leq \varphi(f(\chi_x(t))) - \varphi(f(\chi_x(s))),$$

for each $x \in B(\bar{x}, \rho) \cap [0 < f \leq r_0]$ and $0 \leq t < s$. As established in Theorem 27, $\chi_x(s)$ must converge strongly to some $\tilde{x} \in S$ as $s \to \infty$. Take $t = 0$ and let $s \to \infty$ to deduce that $\|x - \tilde{x}\| \leq \varphi(f(x))$. Thus $\varphi(f(x)) \geq \operatorname{dist}(x, S)$.
(ii) Take $x \in [0 < f < r_0] \cap B(x, \rho)$ and write $y = P_S(x)$. By convexity, we have

$$0 = f(y) \geq f(x) + \langle \partial^0 f(x), y - x \rangle.$$

This implies

$$f(x) \leq \|\partial^0 f(x)\|\, \|y - x\| = \operatorname{dist}(x, S) \|\partial^0 f(x)\| \leq \varphi(f(x)) \|\partial^0 f(x)\|.$$

Since $f(x) > 0$, we deduce that

$$1 \leq \|\partial^0 f(x)\| \frac{\varphi(f(x))}{f(x)} \leq \frac{1}{c} \|\partial^0 f(x)\|\, \varphi'(f(x)),$$

and the conclusion follows immediately. □

In a similar fashion, we can characterize the global existence of a Łojasiewicz gradient inequality.

Corollary 6 (Characterization of Łojasiewicz inequalities for convex functions: global case) *Let $f : H \to (-\infty, +\infty]$ be a proper, convex and lower-semicontinuous, with* $\min f = 0$. *Let $\varphi \in \mathcal{K}(0, +\infty)$ and $c > 0$.*

(i) If $\varphi'(f(x)) \|\partial^0 f(x)\| \geq 1$ for all $x \in [0 < f]$, then $\operatorname{dist}(x, S) \leq \varphi(f(x))$ for all $x \in [0 < f]$.

(ii) Conversely, if $s\varphi'(s) \geq c\varphi(s)$ for all $s \in (0, r_0)$ (φ has moderate behavior), and $\varphi(f(x)) \geq \operatorname{dist}(x, S)$ for all $x \in [0 < f]$, then $\varphi'(f(x)) \|\partial^0 f(x)\| \geq c$ for all $x \in [0 < f]$.

Remark 7 (a) Observe the slight dissymmetry between the conclusions of (i) and (ii) in Theorem 5 and Corollary 6: while a desingularizing function provides directly an error bound in (i), an error bound (with moderate growth) becomes desingularizing after a rescaling, namely $c^{-1}\varphi$.
(b) (Hölderian case) When in (ii) one has $\varphi(s) = \gamma s^{\frac{1}{p}}$ with $p \geq 1$, $\gamma > 0$, then the constant c is given by

$$c = \frac{1}{p}. \tag{3}$$

Analytical aspects linked with the above results, such as connections with subgradient curves and nonlinear bounds, are discussed in a section devoted to further theoretical aspects of the interplay between KL inequality and error bounds. We focus here on the essential consequences we expect in terms of algorithms and complexity. With this objective in mind, we first provide some concrete examples in which a KL inequality with known powers and/or constants can be provided.

3.2 Examples: computing Łojasiewicz exponent through error bounds

The method we use for computing Łojasiewicz exponents is quite simple: we derive an error bound for f with as much information as possible on the constants, and then we use the convexity along with either Theorem 5 or Corollary 6 to compute a desingularizing function together with a domain of desingularization; this technique appears also in [41] a paper which only came to our knowledge during the finalization of our article.

3.2.1 KL inequality for piecewise polynomial convex functions and least squares objective with ℓ^1 regularization

Here, a continuous function $f : \mathbb{R}^n \to \mathbb{R}$ is *piecewise polynomial* if there is a partition of $\mathbb{R}^n$ into finitely many polyhedra[4] $P_1, \ldots, P_k$, such that $f_i = f|_{P_i}$ is a polynomial for each $i = 1, \ldots, k$. The degree of f is defined as $\deg(f) = \max\{\deg(f_i) : i = 1, \ldots, k\}$. We have the following interesting result from Li [39, Corollary 3.6]:

Proposition 8 (Li [39]) *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a piecewise polynomial convex function with* $\operatorname{argmin} f \neq \emptyset$. *Then, for each $r \geq \min f$, there exists $\gamma_r > 0$ such that*

$$f(x) - \min f \geq \gamma_r \operatorname{dist}\left(x, \operatorname{argmin} f\right)^{(\deg(f)-1)^n+1} \tag{4}$$

for all $x \in [f \leq r]$.

Combining Proposition 8 and Corollary 6, the above implies:

[4] Usual definitions allow the subdomains to be more complex

Corollary 9 *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a piecewise polynomial convex function with* $\operatorname{argmin} f \neq \emptyset$*. Then f has the Lojasiewicz property on $[f \leq r]$, with exponent* $\theta = 1 - \dfrac{1}{(\deg(f) - 1)^n + 1}$.

Sparse solutions of inverse problems. Let $f : \mathbb{R}^n \to \mathbb{R}$ be given by

$$f(x) = \frac{1}{2}\|Ax - b\|_2^2 + \mu\|x\|_1,$$

where $\mu > 0$, $b \in \mathbb{R}^m$ and A is a matrix of size $m \times n$. Then f is obviously a piecewise polynomial convex function of degree 2. Since f is also coercive, we have $S = \operatorname{argmin} f \neq \emptyset$. A direct application of Proposition 8 and Corollary 9 gives that $f - \min f$ admits $\theta = \frac{1}{2}$ as a Łojasiewicz exponent.

Yet, in order to derive proper complexity bounds for ISTA we need to identify a computable constant γ_r in (4). For this we shall apply a recent result from Beck-Shtern [10].

First let us recall some basic results on error bounds (see e.g., [36, 63]). In what follows, $\|M\|$ denotes the *spectral* or *operator* norm of a real matrix M.[5]

Definition 1 (Hoffman's error bound) *Given positive integers m, n, r, let $A \in \mathbb{R}^{m\times n}$, $a \in \mathbb{R}^m$, $E = \mathbb{R}^{r\times n}$, $e \in \mathbb{R}^r$. We consider the two polyhedra*

$$X = \{x \in \mathbb{R}^n : Ax \leq a\}, \; Y = \{x \in \mathbb{R}^n : Ex = e\},$$

and we assume that $X \cap Y \neq \emptyset$. There exists a constant $\nu = \nu(A, E) \geq 0$, that only depends on the pair (A, E) and is known as Hoffman's constant *for the pair (A, E), such that*

$$\operatorname{dist}(x, X \cap Y) \leq \nu\|Ex - e\|, \; \forall x \in X. \tag{5}$$

A crucial aspect of Hoffman's error bound is the possibility of estimating the constant ν from the data A, E. We will not enter into these details here, we simply refer the reader to the work of Zălinescu [63] and the references therein.

As suggested by Beck, we shall now apply a very useful result from [10] to derive an error bound for f. Recall that $S = \operatorname{argmin}_{\mathbb{R}^n} f$ is convex, compact and nonempty. For any $x^* \in S$, $f(x^*) \leq f(0) = \frac{1}{2}\|b\|^2$ which implies $\|x^*\|_1 \leq \frac{\|b\|^2}{2\mu}$. Hence $S \subset \{x \in \mathbb{R}^n : \|x\|_1 \leq R\}$ for any fixed $R > \frac{\|b\|^2}{2\mu}$. For such a bound R, one has

$$\begin{aligned}
\min_{\mathbb{R}^n} f &= \min\left\{\frac{1}{2}\|Ax - b\|_2^2 + \mu\|x\|_1 \; : x \in \mathbb{R}^n\right\} \\
&= \min\left\{\frac{1}{2}\|Ax - b\|_2^2 + \mu y \; : (x, y) \in \mathbb{R}^n \times \mathbb{R}, \; \|x\|_1 \leq R, \; y = \|x\|_1\right\} \\
&= \min\left\{\frac{1}{2}\|Ax - b\|_2^2 + \mu y \; : (x, y) \in \mathbb{R}^n \times \mathbb{R}, \; \|x\|_1 - y \leq 0, \; y \leq R\right\} \\
&= \min\left\{\frac{1}{2}\|\tilde{A}\tilde{x} - \tilde{b}\|_2^2 + \langle\tilde{\mu}, \tilde{x}\rangle \; : \; \tilde{x} = (x, y) \in \mathbb{R}^n \times \mathbb{R}, \; M\tilde{x} \leq \tilde{R}\right\}
\end{aligned} \tag{6}$$

[5] It is the largest singular value of M, which is the square root of the largest eigenvalue of the positive-semidefinite square matrix $M^T M$, where M^T is the transpose matrix of M.

where

$$\begin{cases} \bullet\, \tilde{A} = [A, 0_{\mathbb{R}^{m\times 1}}] \in \mathbb{R}^{m\times(n+1)},\ \tilde{b} = (b_1,\dots,b_m,0) \in \mathbb{R}^{m+1}, \\ \bullet\, \tilde{\mu} = (0,\dots,0,\mu) \in \mathbb{R}^{n+1},\ \tilde{R} = (0,\dots,0,R) \in \mathbb{R}^{n+1} \\ \bullet\, M = \begin{bmatrix} E & -1_{\mathbb{R}^{2^n\times 1}} \\ 0_{\mathbb{R}^{1\times n}} & 1 \end{bmatrix} \text{ is a matrix of size } (2^n+1)\times(n+1), \\ \text{where } E \text{ is a matrix of size } 2^n \times n \text{ whose rows are all possible distinct vectors of size } n \\ \text{of the form } e_i = (\pm 1,\dots,\pm 1) \text{ for all } i = 1,\dots,2^n. \text{ The order of the } e_i \text{ being arbitrary.} \end{cases}$$

Set $\tilde{X} := \left\{\tilde{x} \in \mathbb{R}^{n+1} : M\tilde{x} \leq \tilde{R}\right\}$ and observe that the "geometrical complexity" of the problem is now embodied in the matrix M.
It is clear that

$$(x^*, y^*) \in \tilde{S} := \operatorname*{argmin}_{\tilde{x}\in\tilde{X}} \left(\tilde{f}(\tilde{x}) := \frac{1}{2}\|\tilde{A}\tilde{x} - \tilde{b}\|_2^2 + \langle\tilde{\mu}, \tilde{x}\rangle\right) \text{ if and only if } \left(x^* \in S \text{ and } y^* = \|x^*\|_1\right).$$

Using [10, Lemma 2.5], we obtain:

$$\operatorname{dist}^2(\tilde{x}, \tilde{S}) \leq \nu^2 \left(\|\tilde{\mu}\|D + 3GD_A + 2G^2 + 2\right)\left(\tilde{f}(\tilde{x}) - \tilde{f}(\tilde{x}^*)\right),\ \forall \tilde{x} \in \tilde{X}$$

where

- $\tilde{x}^* = (x^*, y^*)$ is any optimal point in $\tilde{S}$,
- ν is the Hoffman constant associated with the couple $(M, [\tilde{A}^T, \tilde{\mu}^T]^T)$ as in Definition 1 above.
- D is the Euclidean diameter of the polyhedron $\tilde{X} = \{(x,y) \in \mathbb{R}^{n+1} : \|x\|_1 \leq y \leq R\}$ and is thus the maximal distance between two vertices. Hence $D = 2R$.
- G is the maximal Euclidean norm of the gradient of $\frac{1}{2}\|. - \tilde{b}\|^2$ over $\tilde{A}(\tilde{X})$, hence, $G \leq R\|A\| + \|b\|$.
- D_A is the Euclidean diameter of the set $\tilde{A}(\tilde{X})$, thus $D_A = \max_{x_i\in X} \|A(x_1 - x_2)\| \leq 2R\|A\|$.

Therefore, we can rewrite the above inequality as follows

$$\operatorname{dist}^2(x, S) + (y - y^*)^2 \leq \kappa_R \left(\frac{1}{2}\|Ax - b\|_2^2 + \mu y - \left(\frac{1}{2}\|Ax^* - b\|_2^2 + \mu\|x^*\|_1\right)\right), \forall (x,y) \in \tilde{X}, \tag{7}$$

where

$$\kappa_R = \nu^2 \left(2R\mu + 6\left(R\|A\| + \|b\|\right)R\|A\| + 2\left(R\|A\| + \|b\|\right)^2 + 2\right). \tag{8}$$

By taking $y = \|x\|_1$, (7) becomes

$$\operatorname{dist}^2(x, S) + (y - y^*)^2 \leq \kappa_R(f(x) - f(x^*)),\ \forall x \in \mathbb{R}^n,\ \|x\|_1 \leq R.$$

We therefore obtain

Lemma 10 (Error bound and KL inequality for the least squares objective with ℓ^1 regularization) *Fix $R > \frac{\|b\|^2}{2\mu}$. Then,*

$$f(x) - f(x^*) \geq 2\gamma_R \operatorname{dist}^2(x, S) \text{ for all } x \text{ in } \mathbb{R}^n \text{ such that } \|x\|_1 \leq R, \tag{9}$$

where

$$\gamma_R = \frac{1}{4\nu^2\big(1+\mu R + (R\|A\|+\|b\|)\,(4R\|A\|+\|b\|)\,\big)}. \tag{10}$$

As a consequence f is a KL function on the ℓ^1 ball of radius R and admits $\varphi(s)=\sqrt{2\gamma_R^{-1}\,s}$ as desingularizing function.

3.2.2 Distances to an intersection: convex feasibility

For $m \geq 2$, one considers closed convex subsets $C_1,\dots,C_m$ of H whose intersection contains a nonempty open ball. This proposition is a quantitative version of [12, Corollary 3.1].

Proposition 11 *Suppose that there is $\bar{x}\in H$ and $R>0$ such that*

$$B(\bar{x},R)\subset \bigcap_{i=1}^{m} C_i. \tag{11}$$

Then,

$$\operatorname{dist}(x,\bigcap_{i=1}^{m} C_i) \leq \left(1+\frac{2\|x-\bar{x}\|}{R}\right)^{m-1} \max\left\{\operatorname{dist}(x,C_i), i=1,\cdots,m\right\}, \ \ \forall x\in H. \tag{12}$$

Proof. We assume $m=2$ in a first stage. Put $C=C_1\cap C_2$, $d=2\max\{\operatorname{dist}(x,C_1),\operatorname{dist}(x,C_2)\}$ and fix $x\in H$. The function $\operatorname{dist}(\cdot,C_2)$ is Lipschitz continuous with constant 1. Thus,

$$|\operatorname{dist}(P_{C_1}(x),C_2)-\operatorname{dist}(x,C_2)| \leq \|x-P_{C_1}(x)\|$$

and so

$$\operatorname{dist}(P_{C_1}(x),C_2)\leq \operatorname{dist}(x,C_1)+\operatorname{dist}(x,C_2)\leq d.$$

By taking $y=\bar{x}+\frac{R}{d}(P_{C_1}(x)-P_{C_2}P_{C_1}(x))$, we deduce that $y\in B(\bar{x},R)\subset C_1\cap C_2$. Now, we construct a specific point $z\in C$ as follows

$$z=\frac{d}{R+d}\,y+\frac{R}{R+d}P_{C_2}P_{C_1}(x).$$

Obviously z is in C_2, and if we replace y in z by $\bar{x}+\frac{R}{d}\big(P_{C_1}(x)-P_{C_2}P_{C_1}(x)\big)$, we obtain

$$z=\frac{d}{R+d}\bar{x}+\frac{R}{R+d}P_{C_1}(x)\in C_1,$$

This implies that $z\in C_1\cap C_2$. Therefore

$$\operatorname{dist}(x,C)\leq \|x-z\|\leq \|x-P_{C_1}(x)\|+\|z-P_{C_1}(x)\|,$$

and, since $\bar{x}\in C_1\cap C_2$,

$$\|z-P_{C_1}(x)\|=\frac{d}{R+d}\|\bar{x}-P_{C_1}(x)\|=\frac{d}{R+d}\|P_{C_1}(\bar{x})-P_{C_1}(x)\|\leq \frac{d}{R+d}\|\bar{x}-x\|.$$

By combining the above results, we have $\operatorname{dist}(x,C)\leq \frac{d}{2}+\frac{d}{R+d}\|x-\bar{x}\|$, which gives

$$\operatorname{dist}(x,C)\leq\left(1+\frac{2\|x-\bar{x}\|}{R}\right)\max\left\{\operatorname{dist}(x,C_1),\operatorname{dist}(x,C_2)\right\}. \tag{13}$$

For arbitrary $m \geq 2$, applying (13) for the two sets C_1 and $\overset{m}{\underset{i=2}{\cap}} C_i$, we obtain

$$\operatorname{dist}(x, \overset{m}{\underset{i=1}{\cap}} C_i) \leq \left(1 + \frac{2\|x - \bar{x}\|}{R}\right) \max\left\{\operatorname{dist}(x, C_1), \operatorname{dist}(x, \overset{m}{\underset{i=2}{\cap}} C_i)\right\}.$$

Repeating the process $(m-1)$ times, we obtain (12). □

A potential function for the barycentric projection method. Let $C := \cap_{i=1}^m C_i$. If $C \neq \emptyset$, finding a point in C is equivalent to minimizing the following convex function over H

$$f(x) = \frac{1}{2} \sum_{i=1}^{m} \alpha_i \operatorname{dist}^2(x, C_i), \tag{14}$$

where $\alpha_i > 0$ for all $i = 1, \ldots, m$ and $\sum_1^m \alpha_i = 1$. As we shall see in the next section, the gradient method applied to f yields the *barycentric projection method* (introduced in [5]; see also [25, 12]). We now provide an error bound for f under assumption (11).

It is clear that $C = \operatorname{argmin} f = \{x \in H : f(x) = 0\}$. Fix any $x_0 \in H$. From Proposition 11, we obtain that f has the following local error bound:

$$\operatorname{dist}(x, C) \leq \left(1 + \frac{2\|x_0 - \bar{x}\|}{R}\right)^{m-1} \left(\frac{2}{\min\limits_{i=1,\ldots,m} \alpha_i}\right)^{\frac{1}{2}} \sqrt{f(x)}, \;\; \forall x \in B(\bar{x}, \|x_0 - \bar{x}\|).$$

Combining with Theorem 5, we deduce that f satisfies the Łojasiewicz inequality on $B(\bar{x}, \|x_0 - \bar{x}\|) \cap [0 < f]$ with desingularizing function $\varphi(s) = \sqrt{\frac{2}{M} s}$, where

$$M = \frac{1}{4}\left(1 + \frac{2\|x_0 - \bar{x}\|}{R}\right)^{2-2m} \min_{i=1,\ldots,m} \alpha_i. \tag{15}$$

A potential function for the alternating projection method. Assume now that $m = 2$, and set $g = i_{C_1} + \frac{1}{2}\operatorname{dist}(\cdot, C_2)^2$ – a function related to the alternating projection method, as we shall see in a Section 5. One obviously has $g(x) \geq \frac{1}{2}\big(\operatorname{dist}^2(x, C_1) + \operatorname{dist}^2(x, C_2)\big)$ for all $x \in H$. From the above remarks, we deduce that

$$\operatorname{dist}(x, C) \leq 2\left(1 + \frac{2\|x_0 - \bar{x}\|}{R}\right)\sqrt{g(x)}, \forall x \in B(\bar{x}, \|x_0 - \bar{x}\|).$$

Hence, g satisfies the Łojasiewicz inequality on $B(\bar{x}, \|x_0 - \bar{x}\|) \cap [0 < g]$ with desingularizing function given by

$$\varphi(s) = \sqrt{\frac{2}{M'}\, s},$$

where

$$M' = \frac{1}{8}\left(1 + \frac{2\|x_0 - \bar{x}\|}{R}\right)^{-2}. \tag{16}$$

4 Complexity for first-order methods with sufficient decrease condition

In this section, we derive complexity bounds for first-order methods with a sufficient decrease condition, under a KL inequality. In what follows, we assume, as before, that $f : H \to (-\infty, +\infty]$ is a proper lower-semicontinuous convex function such that $S = \operatorname{argmin} f \neq \emptyset$ and $\min f = 0$.

4.1 Subgradient sequences

We recall, from [4], that $(x_k)_{k\in\mathbb{N}}$ in H is a *subgradient descent sequence* for $f : H \to (-\infty, +\infty]$ if $x_0 \in \operatorname{dom} f$ and there exist $a, b > 0$ such that:

(**H1**) (Sufficient decrease condition) For each $k \geq 1$,

$$f(x_k) + a\|x_k - x_{k-1}\|^2 \leq f(x_{k-1}).$$

(**H2**) (Relative error condition) For each $k \geq 1$, there is $\omega_k \in \partial f(x_k)$ such that

$$\|\omega_k\| \leq b\|x_k - x_{k-1}\|.$$

We point out that an additional continuity condition $-$ which is not necessary here because of the convexity of f $-$ was required in [4].

It seems that these conditions were first considered in the seminal and inspiring work of Luo-Tseng [50]. They were used to study convergence rates from error bounds. We adopt partly their views and we provide a double improvement: on the one hand, we show how complexity can be tackled for such dynamics, and, on the other hand, we provide a general methodology that will hopefully be used for many other methods than those considered here.

The motivation behind this definition is due to the fact that such sequences are generated by many prominent methods, such as the forward-backward method [50, 4, 35] (which we describe in detail below), many trust region methods [1], alternating methods [4, 21], and, in a much more subtle manner, sequential quadratic methods and a wealth of majorization-minimization methods [20, 55]. In Section 5, we essentially focus on the forward-backward method because of its simplicity and its efficiency. Clearly, many other examples could be worked out.

Remark 12 (Explicit step for Lipschitz continuous gradient) If f is smooth and its gradient is Lipschitz continuous with constant L, then any sequence satisfying:

$$(\textbf{H2'}) \qquad \text{For each } k \geq 1,\ \|\nabla f(x_{k-1})\| \leq b\|x_k - x_{k-1}\|,$$

also satisfies (**H2**).
Indeed, for every $k \geq 1$,

$$\|\nabla f(x_k)\| \leq \|\nabla f(x_{k-1})\| + \|\nabla f(x_k) - \nabla f(x_{k-1})\| \leq b\|x_k - x_{k-1}\| + L\|x_k - x_{k-1}\| = (b + L)\|x_k - x_{k-1}\|.$$

Example 2 (The forward-backward splitting method.) The *forward-backward splitting* or *proximal gradient* method is an important model algorithm, although many others could be considered in the general setting we provide (see [4, 21, 35]). Let $g : H \to (-\infty, +\infty]$ be a proper lower-semicontinuous convex function and let $h : H \to \mathbb{R}$ be a smooth convex function whose gradient is Lipschitz continuous with constant L. In order to minimize $g + h$ over H, the forward-backward method generates a sequence $(x_k)_{k\in\mathbb{N}}$ from a given starting point $x_0 \in H$, and using the recursion

$$x_{k+1} \in \operatorname{argmin} \left\{ g(z) + \langle \nabla h(x_k), z - x_k \rangle + \frac{1}{2\lambda_k} \|z - x_k\|^2 : z \in H \right\} \tag{17}$$

for $k \geq 1$. By the strong convexity, lower-semicontinuity of the argument in the right-hand side and weak topology arguments, the set of minimizers has exactly one element. On the other hand, it is easily seen that (17) is equivalent to

$$x_{k+1} \in \operatorname{argmin} \left\{ g(z) + \frac{1}{2\lambda_k} \|z - (x_k - \lambda_k \nabla h(x_k))\|^2 : z \in H \right\}.$$

Moreover, using the proximity operator defined in Subsection 2.1, the latter can be rewritten as

$$x_{k+1} = \operatorname{prox}_{\lambda_k g} (x_k - \lambda_k \nabla h(x_k)). \tag{18}$$

When $h = 0$, we obtain the *proximal point algorithm* for g. On the other hand, if $g = 0$ it reduces to the classical *explicit gradient method* for h.

We shall see that the forward-backward method generates subgradient descent sequences if the step sizes are properly chosen.

Proposition 13 *Assume now that $0 < \lambda^- \leq \lambda_k \leq \lambda^+ < 2/L$ for all $k \in \mathbb{N}$. Then* (**H1**) *and* (**H2**) *are satisfied for the forward-backward splitting method* (18) *with*

$$a = \frac{1}{\lambda^+} - \frac{L}{2} \quad \text{and} \quad b = \frac{1}{\lambda^-} + L.$$

Proof. Take $k \geq 0$. For the constant a, we use the fundamental inequality provided in [21, Remark 3.2(iii)]:

$$g(x_{k+1}) + h(x_{k+1}) \leq g(x_k) + h(x_k) - \left(\frac{1}{\lambda_k} - \frac{L}{2} \right) \|x_{k+1} - x_k\|^2 \leq g(x_k) + h(x_k) - \left(\frac{1}{\lambda^+} - \frac{L}{2} \right) \|x_{k+1} - x_k\|^2.$$

For b, we proceed as in Remark 12 above. Using the Moreau-Rockafellar Theorem, the optimality condition for the forward-backward method is given by

$$\omega_{k+1} + \nabla h(x_k) + \frac{1}{\lambda_k}(x_{k+1} - x_k) = 0,$$

where $\omega_{k+1} \in \partial g(x_{k+1})$. Using the Lipschitz continuity of ∇h, we obtain

$$\|\omega_{k+1} + \nabla h(x_{k+1})\| \leq \left(\frac{1}{\lambda_k} + L \right) \|x_{k+1} - x_k\| \leq \left(\frac{1}{\lambda^-} + L \right) \|x_{k+1} - x_k\|,$$

as claimed. □

If $f = g + h$ has the KL property, Theorem 14 below guarantees the strong convergence of every sequence generated by the forward-backward method.

Convergence of subgradient descent sequences follows readily from [4] and [21, 35]. Although this kind of result has now became standard, we provide a direct proof for estimating thoroughly the constants at stake.

Theorem 14 (Convergence of subgradient descent methods in a Hilbertian convex setting) *Assume that $f : H \to (-\infty, +\infty]$ is a proper lower-semicontinuous convex function which has the KL property on $[0 < f < \bar{r}]$ with desingularizing function $\varphi \in \mathcal{K}(0, \bar{r})$. We consider a subgradient descent sequence $(x_k)_{k\in\mathbb{N}}$ such that $f(x_0) \le r_0 < \bar{r}$. Then, x_k converges strongly to some $x^* \in \operatorname{argmin} f$ and*

$$\|x_k - x^*\| \le \frac{b}{a}\varphi(f(x_k)) + \sqrt{\frac{f(x_{k-1})}{a}},\ \forall k \ge 1. \tag{19}$$

Proof. Using (**H1**), we deduce that the sequence $(f(x_k))_{k\in\mathbb{N}}$ is nonincreasing, thus $x_k \in [0 \le f < \bar{r}]$. Denote by i_0 the first index $i_0 \ge 1$ such that $\|x_{i_0} - x_{i_0-1}\| = 0$ whenever it exists. If such an i_0 exists, one has $\omega_{i_0} = 0$, and so, $f(x_{i_0}) = 0$. This implies that $f(x_{i_0+1}) = 0$ and thus $x_{i_0+1} = x_{i_0}$ (the sequence is then stationary.) Hence the upper bound holds provided that it has been established for all $k \le i_0 - 1$ in (19). A similar reasoning applies to the case when $f(x_{i_0}) = 0$.

Assume first that $f(x_k) > 0$ and $\|x_k - x_{k-1}\| > 0$ for all $k \ge 1$. Combining (**H1**), (**H2**), and using the concavity of φ we obtain

$$\begin{aligned}\varphi(f(x_k)) - \varphi(f(x_{k+1})) &\ge \varphi'(f(x_k))\,(f(x_k) - f(x_{k+1}))\\ &\ge \frac{a\|x_k - x_{k+1}\|^2}{b\|x_{k-1} - x_k\|}\\ &\ge \frac{a}{b}\frac{\left(2\|x_k - x_{k+1}\|\|x_k - x_{k-1}\| - \|x_{k-1} - x_k\|^2\right)}{\|x_k - x_{k-1}\|}, \forall k \ge 1.\\ &\ge \frac{a}{b}(2\|x_k - x_{k+1}\| - \|x_{k-1} - x_k\|), \forall k \ge 1. \end{aligned} \tag{20}$$

This implies

$$\frac{b}{a}\big(\varphi(f(x_1)) - \varphi(f(x_{k+1}))\big) + \|x_0 - x_1\| \ge \sum_{i=1}^{k} \|x_i - x_{i+1}\|, \forall k \in \mathbb{N},$$

therefore, the series $\sum_{i=1}^{\infty} \|x_i - x_{i+1}\|$ is convergent, which implies, by the Cauchy criterion (H is complete), that the sequence $(x_k)_{k\in\mathbb{N}}$ converges to some point $x^* \in H$. From (**H2**), there is a sequence $\omega_k \in \partial f(x_k)$ which converges to 0. Since f is convex and lower-semicontinuous, the graph of ∂f is closed in $H \times H$ for the strong-weak (and weak-strong) topology. Thus $0 \in \partial f(x^*)$.
Coming back to (20), we also infer

$$\frac{b}{a}(\varphi(f(x_k)) - \varphi(f(x_{k+m}))) + \|x_{k-1} - x_k\| \ge \sum_{i=k}^{k+m} \|x_i - x_{i+1}\|, \forall k, m \in \mathbb{N}.$$

Combining the latter with (**H1**) yields

$$\frac{b}{a}(\varphi(f(x_k)) - \varphi(f(x_{k+m}))) + \sqrt{\frac{f(x_{k-1}) - f(x_k)}{a}} \ge \sum_{i=k}^{k+m} \|x_i - x_{i+1}\|, \forall k, m \in \mathbb{N}.$$

Letting $m \to \infty$, we obtain

$$\frac{b}{a}\varphi(f(x_k)) + \sqrt{\frac{f(x_{k-1}) - f(x_k)}{a}} \geq \|x_k - x^*\|, \forall k \in \mathbb{N},$$

thus

$$\frac{b}{a}\varphi(f(x_k)) + \sqrt{\frac{f(x_{k-1})}{a}} \geq \|x_k - x^*\|, \forall k \in \mathbb{N}.$$

The case when $\|x_k - x_{k-1}\|$ or $f(x_k)$ vanishes for some k follows easily by using the argument evoked at the beginning of the proof. □

Remark 15 When f is twice continuously differentiable and *definable* (in particular, if it is semi-algebraic) it is proved in [13] that $\varphi(s) \geq O(\sqrt{s})$ near the origin. This shows that, in general, the "worst" complexity is more likely to be induced by φ rather than the square root.

4.2 Complexity for subgradient descent sequences

This section is devoted to the study of complexity for first-order descent methods of KL convex functions in Hilbert spaces.

Let $0 < r_0 < \bar{r}$, we shall assume that f has the KL property on $[0 < f < \bar{r}]$ with desingularizing function $\varphi \in \mathcal{K}(0, \bar{r})$ (recall that $\operatorname{argmin} f \neq \emptyset$ and $\min f = 0$.). Whence

$$\varphi'(f(x))||\partial^0 f(x)|| \geq 1$$

for all $x \in [0 < f < \bar{r}]$. Set $\alpha_0 = \varphi(r_0)$ and consider the function $\psi = (\varphi|_{[0,r_0]})^{-1} : [0, \alpha_0] \to [0, r_0]$, which is increasing and convex.

The following assumption will be useful in the sequel:

(**A**) The function ψ' is Lipschitz continuous (on $[0, \alpha_0]$) with constant $\ell > 0$ and $\psi'(0) = 0$.

Intuitively, the function ψ embodies the worst-case "profile" of f. As explained below, the worst-case behavior of descent methods appears indeed to be measured through φ. The assumption (**A**) is definitely weak, since for interesting cases ψ is flat near 0, while it can be chosen affine for large values (see Proposition 30).

We focus on algorithms that generate subgradient descent sequences, thus complying with (**H1**) and (**H2**).

A one-dimensional worst-case proximal sequence. Set

$$\zeta = \frac{\sqrt{1 + 2\ell\, a\, b^{-2}} - 1}{\ell} > 0, \tag{21}$$

where $a > 0$, $b > 0$ and $\ell > 0$ are given in (**H1**), (**H2**) and (**A**), respectively. Starting from α_0, we define the *one-dimensional worst-case proximal sequence* inductively by

$$\alpha_{k+1} = \operatorname{argmin}\left\{\psi(u) + \frac{1}{2\zeta}(u - \alpha_k)^2 : u \geq 0\right\} \tag{22}$$

for $k \geq 0$. Using standard arguments, one sees that α_k is well defined and positive for each $k \geq 0$. Moreover, the sequence can be interpreted through the recursion

$$\alpha_{k+1} = (I + \zeta\psi')^{-1}(\alpha_k) = \text{prox}_{\zeta\psi}(\alpha_k), \tag{23}$$

for $k \geq 0$ and where I is the identity on $\mathbb{R}$. Finally, it is easy to prove that α_k is decreasing and converges to zero. By continuity, $\lim_{k\to\infty} \psi(\alpha_k) = 0$.

The following is one of our main results. It asserts that $(\alpha_k)_{k\in\mathbb{N}}$ is a *majorizing sequence* "à la Kantorovich":

Theorem 16 (Complexity of descent sequences for convex KL functions) *Let $f : H \to (-\infty, +\infty]$ be a proper lower-semicontinuous convex function with* $\operatorname{argmin} f \neq \emptyset$ *and* $\min f = 0$*. Assume further that f has the KL property on $[0 < f < \bar{r}]$. Let $(x_k)_{k\in\mathbb{N}}$ be a subgradient descent sequence with $f(x_0) = r_0 \in (0, \bar{r})$ and suppose that assumption* (**A**) *holds (on the interval $[0, \alpha_0]$ with $\psi(\alpha_0) = r_0$).*

Define the one-dimensional worst-case proximal sequence $(\alpha_k)_{k\in\mathbb{N}}$ as above[6]*. Then, $(x_k)_{k\in\mathbb{N}}$ converges strongly to some minimizer x^* and, moreover,*

$$f(x_k) \leq \psi(\alpha_k), \quad \forall k \geq 0, \tag{24}$$

$$\|x_k - x^*\| \leq \frac{b}{a}\alpha_k + \sqrt{\frac{\psi(\alpha_{k-1})}{a}}, \quad \forall k \geq 1. \tag{25}$$

Proof. For $k \geq 1$, set $r_k := f(x_k)$. If $r_k = 0$ the result is trivial. Assume $r_k > 0$, then one has also $r_j > 0$ for $j = 1, \ldots, k$. Set $\beta_k = \psi^{-1}(r_k) > 0$ and $s_k = \frac{\beta_{k-1}-\beta_k}{\psi'(\beta_k)} > 0$ so that β_k satisfies

$$\beta_k = (1 + s_k\psi')^{-1}(\beta_{k-1}). \tag{26}$$

We shall prove that $s_k \geq \zeta$. Combining the KL inequality and (**H2**), we obtain that

$$b^2\varphi'(r_k)^2\|x_k - x_{k-1}\|^2 \geq \varphi'(r_k)^2\|\omega_k\|^2 \geq 1,$$

where ω_k is as in (**H2**). Using (**H1**) and the formula for the derivative of the inverse function, this gives

$$\frac{a}{b^2} \leq \varphi'(r_k)^2(r_{k-1} - r_k) = \frac{(\psi(\beta_{k-1}) - \psi(\beta_k))}{\psi'(\beta_k)^2}.$$

We now use the descent Lemma on ψ (see, for instance, [58, Lemma 1.30]), to obtain

$$\frac{a}{b^2} \leq \frac{(\beta_{k-1} - \beta_k)}{\psi'(\beta_k)} + \frac{\ell(\beta_{k-1} - \beta_k)^2}{2\psi'(\beta_k)^2} = s_k + \frac{\ell}{2}s_k^2.$$

We conclude that

$$s_k \geq \frac{\sqrt{1 + 2\ell\, a\, b^{-2}} - 1}{\ell} = \zeta. \tag{27}$$

The above holds for every $k \geq 1$ such that $r_k > 0$.

To conclude we need two simple results on the prox operator in one dimension.
Claim 1. *Take $\lambda^0 > \lambda^1$ and $\gamma > 0$. Then*

[6] See (21) and (22).

$$(I+\lambda^0\psi')^{-1}(\gamma) < (I+\lambda^1\psi')^{-1}(\gamma).$$

Proof of Claim 1. It is elementary, set $\delta = (I+\lambda^1\psi')^{-1}(\gamma) \in (0,\gamma)$, one indeed has $(I+\lambda^0\psi')(\delta) = (I+\lambda^1\psi')(\delta) + (\lambda^0-\lambda^1)\psi'(\delta) > \gamma$, and the result follows by the monotonicity of $I+\lambda_0\psi'$.

Claim 2. *Let $(\lambda_k^0)_{k\in\mathbb{N}}, (\lambda_k^1)_{k\in\mathbb{N}}$ two positive sequences such that $\lambda_k^0 \geq \lambda_k^1$ for all $k \geq 0$. Define the two proximal sequences*

$$\beta_{k+1}^0 = (I+\lambda_k^0\psi')^{-1}(\beta_k^0), \quad \beta_{k+1}^1 = (I+\lambda_k^1\psi')^{-1}(\beta_k^1),$$

with $\beta_0^0 = \beta_0^1 \in (0, r_0]$. Then $\beta_k^0 \leq \beta_k^1$ for all $k \geq 0$.

Proof of Claim 2. We proceed by induction, the first step being trivial, we assume the result holds true for $k \geq 0$. We write

$$\beta_{k+1}^0 = (I+\lambda_k^0\psi')^{-1}(\beta_k^0) \leq (I+\lambda_k^0\psi')^{-1}(\beta_k^1) \leq (I+\lambda_k^1\psi')^{-1}(\beta_k^1) = \beta_{k+1}^1,$$

where the first inequality is due to the induction assumption (and the monotonicity of ψ'), while the second one follows from Claim 1.

We now conclude by observing that α_k, β_k are proximal sequences,

$$\alpha_{k+1} = (I+c\psi')^{-1}(\alpha_k), \quad \beta_{k+1} = (I+s_k\psi')^{-1}(\beta_k).$$

Recalling that $s_k \geq \zeta$, one can apply Claim 2 to obtain that $\alpha_k \geq \beta_k$. And thus $\psi(\alpha_k) \geq \psi(\beta_k) = r_k$. The last point follows from Theorem 14. □

Remark 17 (Two complexity regimes) In many cases the function ψ is nonlinear near zero and is affine beyond a given threshold $t_0 > 0$ (see subsection 2.4 or Proposition 30). This geometry reflects on the convergence rate of the estimators as follows:

1. A fast convergence regime is observed when $\alpha_k > t_0$. The objective is cut down by a constant value at each step.
2. When the sequence α_k enters $[0, t_0]$, a slower and restrictive complexity regime appears.

Remark 18 (Complexity with a continuum of minimizers) We draw the attention of the reader that our complexity result *on the sequence* (not only on the values) holds even in the case when there is a continuum of minimizers.

It is obvious from the proof that the following result holds.

Corollary 19 (Stable sets and complexity) Let X be a subset of H. If the set $[0 < f < \bar{r}]$ on which f has the KL property is replaced by a more general set of the form: $\bar{X} = X \cap [0 < f < \bar{r}]$ with the property that $x_k \in \bar{X}$ for all $k \geq 0$, then the same result holds.

The above corollary has the advantage to relax the constraints on the desingularizing function: the smaller the set is, the lower (and thus the better) φ can be[7]. There are thus some possibilities to obtain functions ψ with an improved conditioning/geometry, which could eventually lead to tighter complexity

[7]Desingularizing functions for a given problem (but with different domains) are generally definable in the same o-minimal structure thus their germs are always comparable. This is why the expression "the lower" is not ambiguous in our context.

bounds. On the other hand, the stability condition $x_k \in \bar{X}$, $\forall k \in \mathbb{N}$ is generally difficult to obtain.

We conclude by providing a study of the important case $\psi(s) = \frac{\ell}{2}s^2$. In that case assumption (**A**) holds, and we obtain the following particular instance of Theorem 16:

Corollary 20 *The assumptions and the notation are those of Theorem 16, but we assume further that f has the KL property with $\psi(s) = \frac{\ell}{2}s^2$ on $[0 < f < \bar{r}]$. We set*

$$\sigma = \ell b^{-2}. \tag{28}$$

In that case the complexity estimates given in Theorem 16 take the form

$$f(x_k) \leq \frac{f(x_0)}{(1+2a\sigma)^k}, \quad \forall k \geq 0, \tag{29}$$

$$\|x_k - x^*\| \leq \left[1 + \frac{1}{a\sigma\sqrt{1+\frac{1}{2a\sigma}}}\right] \frac{\sqrt{\frac{1}{a}f(x_0)}}{(1+2a\sigma)^{\frac{k-1}{2}}}, \quad \forall k \geq 1. \tag{30}$$

Proof. First, recall that the one-dimensional worst-case proximal sequence $(\alpha_k)_{k\in\mathbb{N}}$ is given by $\alpha_0 = \varphi(r_0)$, and

$$\alpha_{k+1} = \operatorname{argmin}\left\{\frac{\ell}{2}s^2 + \frac{1}{2\zeta}(s-\alpha_k)^2 : s \geq 0\right\}$$

for all $k \geq 0$, where

$$\zeta = \frac{\sqrt{1+2\ell a b^{-2}}-1}{\ell}.$$

Whence, $\alpha_{k+1} = \frac{\alpha_k}{(1+\ell\zeta)}$, and so,

$$\alpha_k = \frac{\alpha_0}{(1+\ell\zeta)^k}, \quad \forall k \geq 0. \tag{31}$$

From (24), we immediately deduce

$$f(x_k) \leq \frac{f(x_0)}{(1+\ell\zeta)^{2k}}.$$

Finally, since

$$1+\ell\zeta = \sqrt{1+2\ell a b^{-2}} = \sqrt{1+2a\sigma},$$

we obtain (29). For (30), first observe that

$$\frac{b}{a}\alpha_k = \frac{b}{a}\frac{\alpha_0}{(1+\ell\zeta)^k} = \frac{b}{a\sqrt{\ell}}\frac{\sqrt{2f(x_0)}}{(1+\ell\zeta)^k} = \frac{b}{a\sqrt{\ell}}\frac{\sqrt{2f(x_0)}}{(1+2\ell a b^{-2})^{k/2}}, \tag{32}$$

while

$$\sqrt{\frac{\psi(\alpha_{k-1})}{a}} = \sqrt{\frac{\ell\alpha_{k-1}^2}{2a}} = \sqrt{\frac{\ell\alpha_0^2}{2a(1+\ell\zeta)^{2k-2}}} = \sqrt{\frac{1+2\ell a b^{-2}}{2a}}\frac{\sqrt{2f(x_0)}}{(1+2\ell a b^{-2})^{k/2}}. \tag{33}$$

In view of (25), by adding (32) and (33) we obtain:

$$\|x_k - x^*\| \leq \left[\frac{b}{a\sqrt{\ell}} + \sqrt{\frac{1}{2a}+\frac{\ell}{b^2}}\right]\frac{\sqrt{2f(x_0)}}{(1+2\ell a b^{-2})^{k/2}}, \quad \forall k \geq 1. \tag{34}$$

To conclude, observe that

$$
\begin{aligned}
\left[\frac{b}{a\sqrt{\ell}}+\sqrt{\frac{1}{2a}+\frac{\ell}{b^2}}\right] &= \sqrt{\frac{1}{2a}+\frac{\ell}{b^2}}\left[1+\frac{1}{\sqrt{\frac{a\sigma}{2}+a^2\sigma^2}}\right]\\
&= \sqrt{\frac{1+2a\sigma}{2a}}\left[1+\frac{1}{a\sigma\sqrt{1+\frac{1}{2a\sigma}}}\right],
\end{aligned}
$$

and combine this last equality with (34) to obtain the result.

□

Remark 21 (Constants) The constant $\sigma = \ell b^{-2}$ plays the role of a step size as it can be seen in the forthcoming examples. For smooth problems and for the classical gradient method, one has for instance $\sigma = \text{constant} \cdot \frac{1}{L}$ (see Section 5 below).

5 Applications: feasibility problems, uniformly convex problems and compressed sensing

In this section we apply our general methodology to derive complexity results for some keynote algorithms that are used to solve problems arising in compressed sensing and convex feasibility. We shall make a constant use of Corollary 20, so let us keep in mind the notation introduced in Section 4, especially the constants a, b and ℓ.

5.1 Convex feasibility problems with regular intersection

Let $\{C_i\}_{i\in\{1,\ldots,m\}}$ be a family of closed convex subsets of H, for which there exist $R>0$ and $\bar{x}\in H$ with

$$B(\bar{x},R)\subset C := \bigcap_{i=1}^{m} C_i.$$

Barycentric Projection Algorithm. Starting from $x_0 \in H$, this method generates a sequence $(x_k)_{k\in\mathbb{N}}$ by the following recursion

$$x_{k+1} = \sum_{i=1}^{m} \alpha_i P_{C_i}(x_k).$$

where $\alpha_i > 0$ and $\sum_{i=1}^{m} \alpha_i = 1$.

Using the function $f = \frac{1}{2}\sum_{i=1}^{m} \alpha_i \operatorname{dist}^2(\cdot, C_i)$, studied in Subsection 3.2.2, it is easy to check that

$$\nabla f(x) = \sum_{i=1}^{m} \alpha_i (x - P_{C_i}x) = x - \sum_{i=1}^{m} \alpha_i P_{C_i}(x)$$

for all x in H. Thus, the sequence $(x_k)_{k\in\mathbb{N}}$ can be described by the recursion

$$x_{k+1} = x_k - \nabla f(x_k),\ k \ge 0.$$

Moreover, ∇f is Lipschitz continuous with constant $L = 1$. It follows that $(x_k)_{k\in\mathbb{N}}$ satisfies the conditions (**H1**) and (**H2**) with $a = \frac{1}{2}, b = 2$. It is classical to see that for any $\hat{x} \in C$, the sequence $\|x_k - \hat{x}\|$ is decreasing (see, for instance, [58]). This implies that $x_k \in B(\bar{x}, \|x_0 - \bar{x}\|)$ for all $k \geq 0$. As a consequence, f has a global desingularizing function φ on $B(\bar{x}, \|x_0 - \bar{x}\|)$, whose inverse is given by

$$\psi(s) = \frac{M}{2}s^2,\ s \geq 0,$$

where M is given by (15). Using Theorem 20 with $a = \frac{1}{2}$, $b = 2$ and $\ell = M$, we obtain:

Theorem 22 (Complexity of the barycentric projection method for regular intersections) *The barycentric projection sequence $(x_k)_{k\in\mathbb{N}}$ converges strongly to a point $x^* \in C$ and*

$$f(x_k) \ \leq \ \frac{f(x_0)}{\left(1 + \frac{M}{4}\right)^k}, \quad \forall k \geq 0,$$

$$\|x_k - x^*\| \leq \left[1 + \frac{8}{M\sqrt{1 + \frac{4}{M}}}\right] \frac{\sqrt{2f(x_0)}}{\left(1 + \frac{M}{4}\right)^{\frac{k-1}{2}}}\ , \quad \forall k \geq 1,$$

where M is given by (15).

Alternating projection algorithm. We consider here the feasibility problem in the case $m = 2$. The von Neuman's *alternating projection method* is given by the following recursion

$$x_0 \in H, \quad \text{and} \quad x_{k+1} = P_{C_1}P_{C_2}(x_k) \quad \forall k \geq 0.$$

Let $g = i_{C_1} + \frac{1}{2}\operatorname{dist}^2(\cdot, C_2)$ and let M' be defined as in (16) (Subsection 3.2.2). The function $h = \frac{1}{2}\operatorname{dist}^2(\cdot, C_2)$ is differentiable and $\nabla h = I - P_{C_2}$ is Lipschitz continuous with constant 1. We can interpret the sequence $(x_k)_{k\in\mathbb{N}}$ as the forward-backward splitting method[8]

$$x_{k+1} = \operatorname{prox}_{i_{C_1}}(x_k - \nabla h(x_k)) = P_{C_1}(x_k - \nabla h(x_k)),$$

and observe that the sequence satisfies the conditions (**H1**) and (**H2**) with $a = \frac{1}{2}$ and $b = 2$. As before, the fact that $x_k \in B(\bar{x}, \|x_0 - \bar{x}\|)$ for all $k \geq 0$, is standard (see [6]). As a consequence, the function g has a global desingularizing function φ on $B(\bar{x}, \|\bar{x} - x_0\|)$ whose inverse ψ is $\psi(s) = \frac{M'}{2}s^2$, where M' is given by (16). Using Corollary 20 with $a = \frac{1}{2}$, $b = 2$ and $\ell = M'$, we obtain:

Theorem 23 (Complexity of the alternating projection method for regular convex sets) *With no loss of generality, we assume that $x_0 \in C_1$. The sequence generated by the alternating projection method converges to a point $x^* \in C$. Moreover, $x_k \in C_1$ for all $k \geq 1$,*

$$\operatorname{dist}(x_k, C_2) \ \leq \ \frac{\operatorname{dist}(x_0, C_2)}{\left(1 + \frac{M'}{4}\right)^{\frac{k}{2}}}, \quad \forall k \geq 0,$$

$$\|x_k - x^*\| \leq \left[1 + \frac{8}{M'\sqrt{1 + \frac{M'}{4}}}\right] \frac{\operatorname{dist}(x_0, C_2)}{\left(1 + \frac{M'}{4}\right)^{\frac{k-1}{2}}}, \quad \forall k \geq 1,$$

where M' is given by (16).

[8] A very interesting result from Baillon-Combettes-Cominetti [9] establishes that for more than two sets there are no potential functions corresponding to the alternating projection method.

5.2 Uniformly convex problems

Let σ be a positive coefficient. The function f is called *p-uniformly convex, or simply uniformly convex,* if there exists $p \geq 2$ such that:

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

for all $x, y \in H$, $x^* \in \partial f(x)$. It is easy to see that f satisfies the KL inequality on H with $\varphi(s) = p\,\sigma^{-\frac{1}{p}}\, s^{\frac{1}{p}}$ (see [3]). For such a function we have

$$\psi(s) = \frac{\sigma}{p^p} s^p,\ s \geq 0.$$

Fix x_0 in $\operatorname{dom} f$ and set $r_0 = f(x_0)$, $\alpha_0 = \psi(r_0)$. The Lipschitz continuity constant of ψ' is given by $\ell = \frac{(p-1)\sigma}{p^{p-1}}\,\alpha_0^{p-2}$. Choose a descent method satisfying (**H1**), (**H2**), some examples can be found in [4, 35]. Set $\zeta = \frac{\sqrt{1+2\ell\, a\, b^{-2}}-1}{\ell}$. The complexity of the method is measured by the real sequence

$$\alpha_{k+1} = \operatorname{argmin}\left\{\frac{\sigma}{p^p}u^p + \frac{1}{2\zeta}(u-\alpha_k)^2 : u \geq 0\right\},\ k \geq 0.$$

The case $p = 2$ can be computed in closed form (as previously), but in general only numerical estimates are available.

Proposition 8 shows that first-order descent sequences for piecewise polynomial convex functions have a similar complexity structure. This shows that error bounds or KL inequalities capture more precisely the determinant geometrical factors behind complexity than mere uniform convexity.

5.3 Compressed sensing and the ℓ^1-regularized least squares problem

We refer for instance to [24] for an account on compressed sensing and an insight into its vast field of applications. We consider the cost function $f : \mathbb{R}^n \to \mathbb{R}$ given by

$$f(x) = \mu\|x\|_1 + \frac{1}{2}\|Ax - d\|_2^2,$$

where $\mu > 0$, $A \in \mathbb{R}^{m\times n}$ and $d \in \mathbb{R}^m$.

Set $g(x) = \mu\|x\|_1$ and $h(x) = \frac{1}{2}\|Ax - d\|_2^2 = \frac{1}{2}\|Ax - d\|^2$, so that g is proper, lower-semicontinuous and convex, whereas h is convex and differentiable, and its gradient is Lipschitz continuous with constant $L = \|A^T A\|$. Starting from any $x_0 \in \mathbb{R}^n$, the forward-backward splitting method applied to f is known as the *iterative shrinkage thresholding algorithm* [31][9]:

$$\text{(ISTA)} \qquad x_{k+1} = \operatorname{prox}_{\lambda_k\mu\|\cdot\|_1}\left(x_k - \lambda_k(A^T A x_k - A^T d)\right) \quad \text{for } k \geq 0.$$

Here, $\operatorname{prox}_{\lambda_k\mu\|\cdot\|_1}$ is an easily computable piecewise linear object known as the *soft thresholding* operator (see, for instance, [27]). This method has been applied widely in many contexts and is known to have a complexity $O\left(\frac{1}{k}\right)$. We intend to prove here that this bound can be surprisingly "improved" by our techniques.

First, recall that, according to Proposition 13, sequences generated by this method comply with (**H1**) and (**H2**), provided the stepsizes satisfy $0 < \lambda^- \leq \lambda_k \leq \lambda^+ < 2/L$. Recall that the constants a and b can be chosen as

$$a = \frac{1}{\lambda^+} - \frac{L}{2} \quad \text{and} \quad b = \frac{1}{\lambda^-} + L, \tag{35}$$

[9]Connection between ISTA and the forward-backward splitting method is due to Combettes-Wajs [27]

respectively.

Set $R = \max\left(\frac{f(x_0)}{\mu}, 1 + \frac{\|d\|^2}{2\mu}\right)$. We clearly have $R > \frac{\|d\|^2}{2\mu}$, and, using the fact that $(x_k)_{k\in\mathbb{N}}$ is a descent sequence, we can easily verify that $\|x_k\|_1 \leq R$ for all $k \in \mathbb{N}$.

From Lemma 10 we know that the function f has the KL property on $[\min f < f < \min f + r_0] \cap \{x \in \mathbb{R}^n : \|x\|_1 \leq R\}$ with[10] a global desingularizing function φ whose inverse ψ is given by

$$\psi(s) = \frac{\gamma_R}{2} s^2, s \geq 0$$

where γ_R is known to exist and is bounded from above by the constant given in (10).

Remark 24 (Constant step size) If one makes the simple choice of a *constant* step size all throughout the process, namely $\lambda_k = d/L$ with $d \in (0, 2)$, one obtains

$$\zeta = \frac{\sqrt{1 + \frac{d(2-d)}{L(1+d)^2}\gamma_R} - 1}{\gamma_R} \quad \text{and} \quad \alpha_k = \frac{\alpha_0}{\left(1 + \frac{d(2-d)}{L(1+d)^2}\,\gamma_R\right)^{k/2}}, \quad k \geq 0.$$

Combining the above developments with Corollary 19, we obtain the following surprising result:

Theorem 25 (Complexity bounds for ISTA) *The sequence $(x_k)_{k\in\mathbb{N}}$ generated by* ISTA *converges to a minimizer x^* of f, and satisfies*

$$f(x_k) - \min f \;\leq\; \frac{f(x_0) - \min f}{q^k}, \qquad \forall k \geq 0, \tag{36}$$

$$\|x_k - x^*\| \;\leq\; C\frac{\sqrt{f(x_0) - \min f}}{q^{\frac{k-1}{2}}} \qquad \forall k \geq 1, \tag{37}$$

where

$$q = 1 + \frac{2a\gamma_R}{b^2} \quad \textit{and} \quad C = \frac{1}{\sqrt{a}}\left(1 + \frac{1}{ab^{-2}\gamma_R\sqrt{1 + \frac{1}{2ab^{-2}\gamma_R}}}\right).$$

Remark 26 (Complexity and convergence rates for ISTA) (a) While it was known that ISTA has a linear asymptotic convergence rate, see [42] in which a transparent explanation is provided, best known *complexity bounds* were of the type $O(\frac{1}{k})$, see [11, 33]. Much like in the spirit of [42], we show here how geometry impacts complexity –through error bounds/KL inequality– providing thus complementary results to what is usually done in this field.
(b) The estimate of γ_R given in Section 3.2.1 is far from being optimal and more work remains to be done to obtain acceptable/tight bounds. Observe however that the role of an optimal γ_R is absolutely crucial when it comes to complexity (see (36)): a good "conditioning" (γ_R not too small) provides fast convergence, while a bad one[11] comes with "bad complexity".
(c) Assuming that the forward-backward method is performed with a constant stepsize d/L as in Remark 24, the value q appearing in the complexity bounds given by Theorem 25 becomes

$$q = 1 + \frac{d(2-d)}{(d+1)^2 L}\gamma_R.$$

[10]Recall that $r_0 = f(x_0)$.
[11]Bad conditioning are produced by flat objective functions yielding thus small constants γ_R.

This quantity is maximized when $d = 1/2$. In this case, one obtains the optimized estimate:

$$\begin{aligned} f(x_k) - \min f &\leq \frac{f(x_0) - \min f}{\left(1 + \frac{\gamma_R}{3L}\right)^k}, \qquad \forall k \geq 0, \\ \|x_k - x^*\| &\leq \sqrt{\frac{2}{3L}} \left(1 + \frac{6L}{\gamma_R \sqrt{1 + \frac{3L}{\gamma_R}}}\right) \frac{\sqrt{f(x_0) - \min f}}{\left(1 + \frac{\gamma_R}{3L}\right)^{\frac{k-1}{2}}}, \qquad \forall k \geq 1. \end{aligned}$$

6 Error bounds and KL inequalities for convex functions: additional properties

In this concluding section we provide further theoretical perspectives that will help the reader to understand the possibilities and the limitations of our general methodology. We give, in particular, a counterexample to the full equivalence between the KL property and error bounds, and we provide a globalization result for desingularizing functions.

6.1 KL inequality and length of subgradient curves

This subsection essentially recalls a characterization result from [18] on the equivalence between the KL inequality and the existence of a uniform bound for the length of subgradient trajectories verifying a subgradient differential inclusion. Due to the contraction properties of the semi-flow, the result is actually stronger than the nonconvex results provided in [18]. For the reader's convenience, we provide a self-contained proof.

Given $x \in \overline{\operatorname{dom} \partial f}$, we denote by $\chi_x : [0, \infty) \to H$ the unique solution of the differential inclusion

$$\dot{y}(t) \in -\partial f(y(t)), \text{ almost everywhere on } (0, +\infty),$$

with initial condition $y(0) = x$.

The following result provides an estimation on the length of subgradient trajectories, when f satisfies the KL inequality. Given $x \in \overline{\operatorname{dom} f}$, and $0 \leq t < s$, write

$$\operatorname{length}(\chi_x, t, s) = \int_t^s \|\dot{\chi}_x(\tau)\| \, d\tau.$$

Recall that $S = \operatorname{argmin} f$ and that $\min f = 0$.

Theorem 27 (KL and uniform bounds of subgradient curves) *Let $\bar{x} \in S$, $\rho > 0$ and $\varphi \in \mathcal{K}(0, r_0)$. The following are equivalent:*

i) For each $y \in B(\bar{x}, \rho) \cap [0 < f < r_0]$, we have

$$\varphi'(f(y)) \|\partial^0 f(y)\| \geq 1.$$

ii) For each $x \in B(\bar{x}, \rho) \cap [0 < f \leq r_0]$ and $0 \leq t < s$, we have

$$\operatorname{length}(\chi_x, t, s) \leq \varphi\big(f\left(\chi_x\left(t\right)\right)\big) - \varphi\big(f\left(\chi_x\left(s\right)\right)\big).$$

Moreover, under these conditions, $\chi_x(t)$ converges strongly to a minimizer as $t \to \infty$.

Proof. Take $x \in B(\bar{x}, \rho) \cap [0 < f \leq r_0]$ and $0 \leq t < s$. First observe that

$$\varphi(f(\chi_x(t))) - \varphi(f(\chi_x(s))) = \int_s^t \frac{d}{d\tau}\, \varphi(f(\chi_x(\tau)))\, d\tau = \int_t^s \varphi'(f(\chi_x(\tau)))\|\dot{\chi}_x(\tau)\|^2\, d\tau.$$

Since $\chi_x(\tau) \in \operatorname{dom} \partial f \cap B(\bar{x}, \rho) \cap [0 < f < r_0]$ for all $\tau > 0$ (see Theorem 1) and $-\dot{\chi}_x(\tau) \in \partial f(\chi_x(\tau))$ for almost every $\tau > 0$, it follows that

$$1 \leq \|\partial^0(\varphi \circ f)(\chi_x(\tau))\| \leq \varphi'(f(\chi_x(\tau)))\|\dot{\chi}_x(\tau)\|$$

for all such τ. Multiplying by $\|\dot{\chi}_x(\tau)\|$ and integrating from t to s, we deduce that

$$\operatorname{length}(\chi_x, t, s) \leq \varphi(f(\chi_x(t))) - \varphi(f(\chi_x(s))).$$

Conversely, take $y \in \operatorname{dom} \partial f \cap B(\bar{x}, \rho) \cap [0 < f < r_0]$ (if y is not in $\operatorname{dom} \partial f$ the result is obvious). For each $h > 0$ we have

$$\frac{1}{h}\int_0^h \|\dot{\chi}_y(\tau)\|\, d\tau \leq -\frac{\varphi(f(\chi_y(h)) - \varphi(f(y))}{h}.$$

As $h \to 0$, we obtain

$$\|\dot{\chi}_y(0^+)\| \leq \varphi'(f(y))\|\dot{\chi}_y(0^+)\|^2 = \varphi'(f(y))\, \|\partial^0 f(y)\|\, \|\dot{\chi}_y(0^+)\|,$$

and so

$$\|\partial^0(\varphi \circ f)(y)\| \geq 1.$$

Finally, since $\|\chi_x(t) - \chi_x(s)\| \leq \operatorname{length}(\chi_x, t, s)$, we deduce from ii) that the function $t \mapsto \chi_x(t)$ has the Cauchy property as $t \to \infty$. □

6.2 A counterexample: error bounds do not imply KL

In [18, Section 4.3], the authors build a twice continuously differentiable convex function $f : \mathbb{R}^2 \to \mathbb{R}$ which does not have the KL property, and such that $S = \overline{D}(0, 1)$ (the closed unit disk of radius 1). This implies that f does not satisfy the KL inequality whatever choice of desingularizing function φ is made.

Let us show that this function has a smooth error bound. First note that, since S is compact, f is coercive (see, for instance, [60]). Define $\psi : [0, \infty) \to \mathbb{R}_+$ by

$$\psi(s) = \min\{f(x) : \|x\| \geq 1 + s\}.$$

This function is increasing (recall that f is convex) and it satisfies

$$\psi(0) = 0, \tag{38}$$

$$\psi(s) > 0 \text{ for } s > 0, \tag{39}$$

$$f(x) \geq \psi(\operatorname{dist}(x, S)) \text{ for all } x \in [r < f] \tag{40}$$

Let $\hat{\psi}$ be the convex envelope of ψ, that is the greatest convex function lying below ψ. One easily verifies that $\hat{\psi}$ enjoys the same properties (38), (39), (40). The Moreau envelope of the latter:

$$\mathbb{R}_+ \ni s \to \Psi(s) := \hat{\psi}_1(s) = \inf\{\hat{\psi}(\varsigma) + \tfrac{1}{2}(s - \varsigma)^2 : \varsigma \in \mathbb{R}\},$$

is convex, has 0 as a unique minimizer, is continuously differentiable with positive derivative on $\mathbb{R} \setminus \{0\}$, and satisfies $\Psi \leq \psi_1$ (see [7]). Whence,

$$f(x) \geq \Psi(\operatorname{dist}(x, S)) \text{ for all } x \in [0 < f < r].$$

We have proved the following:

Theorem 28 (Error bounds do not imply KL) *There exists a C^2 convex function $f : \mathbb{R}^2 \to \mathbb{R}$, which does not satisfy the KL inequality, but has an error bound with a smooth convex residual function.*

Remark 29 (Hölderian error bounds without convexity) Hölderian error bounds do not necessarily imply Łojasiewicz inequality – not even the KL inequality – for nonconvex functions. The reason is elementary and consists simply in considering a function with non isolated critical values. Given $r \geq 2$, consider the C^{r-1} function

$$f(x) = \begin{cases} x^{2r}\left(2 + \cos\left(\frac{1}{x}\right)\right) & \text{if } x \neq 0, \\ 0 & \text{if } x = 0. \end{cases}$$

It satisfies $f'(0) = 0$ and $f'(x) = 4rx^{2r-1} + 2rx^{2r-1}\cos\left(\frac{1}{x}\right) + x^{2r-2}\sin\left(\frac{1}{x}\right)$ if $x \neq 0$. Moreover, we have $f(x) \geq x^{2r} = \operatorname{dist}(x, S)^{2r}$ for all $x \in \mathbb{R}$. On the other hand, picking $y_k = \frac{1}{2k\pi}$ and $z_k = \frac{1}{2k\pi + 3\frac{\pi}{2}}$, we see that $f'(y_k) = \frac{6r}{(2k\pi)^{2r-1}} > 0$ and $f'(z_k) = \frac{1}{(2k\pi+3\frac{\pi}{2})^{2r-2}}\left(\frac{4r}{2k\pi+3\frac{\pi}{2}} - 1\right) < 0$ for all sufficiently large k. Therefore, there is a positive sequence $(x_k)_{k\in\mathbb{N}}$ converging to zero with $f'(x_k) = 0$ for all k. Hence, f cannot satisfy the KL inequality at 0.

6.3 From semi-local inequalities to global inequalities

We derive here a globalization result for KL inequalities that strongly supports the Lipschitz continuity assumption for the derivative of the inverse of a desingularizing function, an assumption that was essential to derive Theorem 16. The ideas behind the proof are inspired by [18].

Proposition 30 (Globalization of KL inequality – convex case) *Let $f : H \to (-\infty, +\infty]$ be a proper lower semicontinuous convex function such that* $\operatorname{argmin} f \neq \emptyset$ *and* $\min f = 0$*. Assume also that f has the KL property on $[0 < f < r_0]$ with desingularizing function $\varphi \in \mathcal{K}(0, r_0)$. Then, given $r_1 \in (0, r_0)$, the function given by*

$$\phi(r) = \begin{cases} \varphi(r) \text{ when } r \leq r_1 \\ \varphi(r_1) + (r - r_1)\varphi'(r_1) \text{ when } r \geq r_1 \end{cases}$$

is desingularising for f on all of H.

Proof. Let x be such that $f(x) > r_1$. We would like to establish that $\|\partial^0 f(x)\|\phi'(f(x)) \geq 1$, thus we may assume, with no loss of generality, that $\|\partial^0 f(x)\|$ is finite. If there is $y \in [f = r_1]$ such that $\|\partial^0 f(y)\| \leq \|\partial^0 f(x)\|$, then

$$\|\partial^0 f(x)\|\phi'(f(x)) = \|\partial^0 f(x)\|\varphi'(r_1) \geq \|\partial^0 f(y)\|\varphi'(r_1) = \|\partial^0 f(y)\|\varphi'(f(y)) \geq 1.$$

To show that such a y exists, we use the semiflow of ∂f. Consider the curve $t \to \chi_x(t)$ and observe that there exists $t_1 > 0$ such that $f(\chi_x(t_1)) = r_1$, because $f(\chi_x(0)) = f(x) > r_1$, $f(\chi_x(t)) \to \inf f < r_1$ and $f(\chi_x(\cdot))$ is continuous. From [23, Theorem 3.1 (6)], we know also that $\|\partial f^0(\chi_x(t))\|$ is nonincreasing. As a consequence, if we set $y = \chi_x(t_1)$, we obtained the desired point and the final conclusion. □

One deduces easily from the above the following result, which is close to an observation already made in [18]. For an insight into the notion of *definability* of functions, a prominent example being semi-algebraicity, one is referred to [30]. Recall that coercivity of a proper lower-semicontinuous convex function defined on a finite dimensional space is equivalent to the fact that argmin f is nonempty and compact.

Theorem 31 (Global KL inequalities for coercive definable convex functions) *Let $f : \mathbb{R}^n \to \mathbb{R}$ be proper, lower-semicontinuous, convex, definable, and such that* argmin f *is nonempty and compact. Then, f has the KL property on $\mathbb{R}^n$.*

Proof. Take $r_0 > 0$ and use [19] to obtain $\varphi \in \mathcal{K}(0, r_0)$ so that f is KL on $[\min f < f < \min f + r_0]$. Then use the previous proposition to extend φ on $(0, +\infty)$. $\square$

Remark 32 (Complexity of descent methods for definable coercive convex function) The previous result implies that *there always exists a global measure of complexity for first-order descent methods* (**H1**), (**H2**) *of definable coercive convex lower-semicontinuous functions.* This complexity bound is encoded in majorizing sequences computable from a single definable function and from the initial data. These majorizing sequences are of course defined, as in Theorem 16, by

$$\alpha_{k+1} = \operatorname{argmin} \left\{ \varphi^{-1}(u) + \frac{1}{2\zeta}(u - \alpha_k)^2 : u \geq 0 \right\},\ \alpha_0 = \varphi(r_0).$$

or equivalently

$$\alpha_{k+1} = \operatorname{prox}_{\zeta\varphi^{-1}}(\alpha_k),\ \alpha_0 = \varphi(r_0),$$

where ζ is a parameter of the chosen first-order method.

It is a very theoretical result yet conceptually important since it shows that the understanding and the research of complexity is guaranteed by the existence of a global KL inequality and our general methodology.

7 Conclusions

In this paper, we devised a general methodology to estimate the complexity for descent methods which are commonly used to solve convex optimization problems: error bounds can be employed to obtain desingularizing functions in the sense of Łojasiewicz, which, in turn, provide the complexity estimates. These techniques are applied to obtain new complexity results for ISTA in compressed sensing, as well as barycentric and alternating projection method for convex feasibility.

While this work was in its final phase, we discovered the prepublication [40] in which complementary ideas are used to develop error bounds for parametric polynomial systems and to analyze the convergence rate of some first order methods. Numerous interconnections and roads must be investigated at the light of these new discoveries, and we hope to do so in our future research.

Acknowledgements The authors would like to thank Amir Beck, Patrick Combettes, Édouard Pauwels, Marc Teboulle and the anonymous referee for very useful comments.

References

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

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

[3] Attouch H., Bolte J., Redont P., Soubeyran A.: Proximal alternating minimization and projection methods for nonconvex problems. An approach based on the Kurdyka-Łojasiewicz inequality, *Math. Oper. Res. 35, no2, 438–457 (2010).*

[4] Attouch H., Bolte J., Svaiter B.F.: Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized Gauss-Seidel methods, *Math. Program. Ser. A 137, no. 1-2, 91–129 (2013).*

[5] Auslender A.: *Méthodes numériques pour la résolution des problèmes d'optimisation avec contraintes*, PhD Thesis, Université Joseph Fourier Grenoble, France, (1969).

[6] Bauschke H.H., Borwein, J.M., On projection algorithms for solving convex feasibility problems, *SIAM Review 38, no. 3, 367–426 (1996).*

[7] Bauschke H.H., and Combettes P.L.: *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer, New York, (2011).

[8] Auslender A., Crouzeix, J.-P., Global regularity theorems, *Math. Oper. Res. 13, 243–253 (1988).*

[9] Baillon J.-B., Combettes P.L., Cominetti R., There is no variational characterization of the cycles in the method of periodic projections, *Journal of Functional Analysis, 262, no. 1, 400–408 (2012).*

[10] Beck A., Shtern S.: Linearly Convergent Away-Step Conditional Gradient for Non-strongly Convex Functions, http://arxiv.org/abs/1504.05002.

[11] Beck A., Teboulle M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems, *SIAM Journal on Imaging Science, 2, 183–202 (2008).*

[12] Beck A., Teboulle M.: Convergence rate analysis and error bounds for projection algorithms in convex feasibility problem, *Optimization Methods and Software, 18, no. 4, 377–394 (2003).*

[13] Bégout P., Bolte J., Jendoubi M.-A: On damped second-order gradient systems, *J. Differential Equations 259, no 7, 3115–3143 (2015).*

[14] Belousov E.G, Klatte D.: A Frank-Wolfe type theorem for convex polynomial programs, *Comput. Optim. Appl. 22, 37–48 (2002).*

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

[16] Bolte J.: Sur quelques principes de convergence en Optimisation. *Habilitation à diriger des recherches, Université Pierre et Marie Curie, (2008).*

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

[18] Bolte J., Daniilidis A., Ley O., Mazet L.: Characterizations of Łojasiewicz inequalities: subgradient flows, talweg, convexity, *Trans. Amer. Math. Soc. 362, no. 6, 3319–3363 (2010).*

[19] Bolte J., Daniilidis A., Lewis A., Shiota M.: Clarke subgradients of stratifiable functions, *SIAM J. Optim. 18, no. 2, 556–572 (2007).*

[20] Bolte J., Pauwels E.: Majorization-minimization procedures and convergence of SQP methods for semi-algebraic and tame programs. Published online in *Math. Oper. Res. (2015).*

[21] Bolte J., Sabach S., Teboulle M.: Proximal alternating linearized minimization for nonconvex and nonsmooth problems, *Math. Program. Ser. A 146, 1–16 (2013).*

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

[23] Brézis H.: *Opérateurs maximaux monotones et semi-groupes de contractions dans les espace Hilbert*, North-Holland Mathematics studies 5, (North-Holland Publishing Co., (1973).

[24] Candes E. J., Wakin M.B.: An introduction to compressive sampling, *IEEE Signal Processing Magazine, 25, no. 2, 21–30 (2008).*

[25] Combettes P.L.: Inconsistent signal feasibility problems: least-squares solutions in a product space, *IEEE Transactions on Signal Processing, 42(11), 2955–2966, (1994)*

[26] Combettes P.L., Pesquet J.-C.: Proximal Splitting Methods in Signal Processing, *Fixed-Point Algorithms for Inverse Problems in Science and Engineering Springer Optimization and Its Applications, 49, 185–212 (2011).*

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

[28] Cornejo O., Jourani A., Zălinescu C.: Conditioning and Upper-Lipschitz Inverse Subdifferentials in Nonsmooth Optimization Problems, *Journal of Optimization Theory and Applications, 95(1), 127–148, (1997).*

[29] Corvellec J.-N., Motreanu V.V.: Nonlinear error bounds for lower semicontinuous functions on metric spaces. *Math Program, 114 (2), 291–319 (2008).*

[30] Coste M., *An introduction to o-minimal geometry*, RAAG Notes, 81 p., Institut de Recherche Mathématiques de Rennes, (1999).

[31] Daubechies I., Defrise M., De Mol C.: An iterative thresholding algorithm for linear inverse problems with a sparsity constraint, *Comm. Pure Appl. Math., 57, 1413–1457 (2004)*

[32] Dedieu, J.-P: Penalty functions in subanalytic optimization, *Optimization 26, 27–32 (1992).*

[33] Drori, Y.: Contributions to the Complexity Analysis of Optimization Algorithms, PhD Thesis, Tel Aviv (2014).

[34] Ferris M.: Finite termination of the proximal point algorithm, *Math. Programming, 50, 359-366 (1991).*

[35] Frankel P., Garrigos G., Peypouquet J.: Splitting methods with variable metric for KL functions. *J. Optim. Theory Appl. 165, no. 3, 874–900 (2015).*

[36] Hoffman A.J.: On approximate solutions of systems of linear inequalities, *Journal of Research of the National Bureau of Standards* 49(4), 263-265, (1952).

[37] Kantorovich L.V., Akilov G.P.: *Functional Analysis in Normed Spaces*, Pergamon, Oxford, (1964).

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

[39] Li G.: Global error bounds for piecewise convex polynomials, *Math. Program. 137, no. 1-2, Ser. A, 37–64 (2013).*

[40] Li G., Mordukhovich B.S., Nghia T.T.A., Pham T.S: Error Bounds for Parametric Polynomial Systems with Applications to Higher-Order Stability Analysis and Convergence Rates, *ArXiv preprint arXiv:1509.03742, 2015.*

[41] Li G., Mordukhovich B.S., Pham T.S.: New fractional error bound for polynomial systems with applications to Hölderian stability in optimization and spectral theory of tensors, to appear in *Math. Program., Ser. A.*

[42] Liang J., Fadili J., Peyré G.: Local linear convergence of forward-backward under partial smoothness, http://arxiv.org/pdf/1407.5611.pdf

[43] 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).*

[44] Łojasiewicz S.: Division d'une distribution par une fonction analytique de variables réelles, *Comptes de Rendus de Séance, Paris 246, 683–686 (1958).*

[45] Łojasiewicz S.: Sur la problème de la division, *Studia Mathematica 18, 87–136 (1959).*

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

[47] Luo X.D, Luo Z.Q.: Extensions of Hoffman's error bound to polynomial systems, *SIAM J. Optim. 4 383-392 (1994).*

[48] Luo Z.-Q., Pang J. S.: Error bounds for analytic systems and their application, *Math. Program. 67 1-28 (1994).*

[49] Luo Z.-Q., Sturm J. F.: Error bound for quadratic systems, *Applied Optimization, Volume 33, pp 383-404 (2000).*

[50] Luo Z.-Q., Tseng P.: Error bounds and convergence analysis of feasible descent methods: a general approach, *Ann. Oper. Res. 46-47, Issue 1, pp 157–178 (1993).*

[51] Mangasarian O. L.: A condition number for differentiable convex inequalities, *Mathematics of Operations Research 10, 175–179 (1985).*

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

[53] Nedić A., Bertsekas D.: Convergence rate of incremental subgradient algorithms, *Stochastic optimization: Algorithms and applications, S. Uryasev and P. M. Pardalos, (Eds.), Kluwer Academic Publishers, 263–304 (2000).*

[54] Ng K.F., Zheng X.Y.: Global error bound with fractional exponents, *Math. Program. 88, 357-370 (2000).*

[55] Pauwels E.: The value function approach to convergence analysis in composite optimization, arXiv preprint https://arxiv.org/abs/1604.01654.

[56] Pang J.S.: Error bounds in mathematical programming, *Math. Program. 79, 299–332 (1997).*

[57] Peypouquet J.: Asymptotic convergence to the optimal value of diagonal proximal iterations in convex minimization, *J. Convex Anal. 16, no. 1, 277-286 (2009).*

[58] Peypouquet J.: *Convex optimization in normed spaces: Theory, methods and examples*, Springer, Cham, 2015. xiv+124 pp.

[59] Robinson S.M.: An application of error bounds for convex programming in a linear space, *SIAM Journal on Control 13, 271-273 (1975).*

[60] Rockafellar R.T.: *Convex Analysis*, Princeton, (1972).

[61] Rockafellar R.T.: Monotone operators and the proximal point algorithm, *SIAM J. Control Optim., 14, 877–898, (1976).*

[62] Vui H.H.: Global Hölderian error bound for non degenerate polynomials, *SIAM J. Optim., 23, no. 2, 917–933 (2013).*

[63] Zălinescu C.: Sharp estimates for Hoffmans constant for systems of linear inequalities and equalities, *SIAM J. Optim., 14, 517–533 (2003)*