AVERTISSEMENT

LIENS

THÈSE

En vue de l'obtention du
DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Université Toulouse I Capitole (UT1 Capitole)

Discipline ou spécialité:
Domaine mathématiques – Mathématiques appliquées

Présentée et soutenue par
NGUYEN Trong Phong
le: 04 juillet 2017

Titre:
Inégalités de Kurdyka-Łojasiewicz et convexité: algorithmes et applications

---

École doctorale:
Mathématiques Informatique Télécommunications (MITT)

Unité de recherche:
TSE–R
Université Toulouse Capitole

Directeur de thèse:
Jérôme BOLTE
Université Toulouse 1 Capitole, France

Rapporteurs:

Jérôme MALICK	Université Grenoble Alpes, France
Tien Son PHAM	Université Da Lat, Viet Nam

Jury:

Jérôme BOLTE	Université Toulouse 1 Capitole, France
Patrick Louis COMBETTES	Université North Carolina State, USA
Jérôme MALICK	Université Grenoble Alpes, France
Edouard PAUWELS	Université Toulouse 3, France
Juan PEYPOUQUET	Université Técnica Federico Santa María, Chili
Aude RONDEPIERRE	INSA, Toulouse, France

Pour ma mère.

Contents

Abstract

This thesis focuses on first order descent methods in the minimization problems. There are three parts. Firstly, we give an overview on local and global error bounds. We try to provide the first bricks of a unified theory by showing the centrality of the Łojasiewicz gradient inequality. In the second part, by using Kurdyka-Łojasiewicz (KL) inequality, we provide new tools to compute the complexity of first-order descent methods in convex minimization. Our approach is completely original and makes use of a one-dimensional worst-case proximal sequence. This result inaugurates a simple methodology: derive an error bound, compute the KL desingularizing function whenever possible, identify essential constants in the descent method and finally compute the complexity using the one-dimensional worst case proximal sequence. Lastly, we extend the extragradient method to minimize the sum of two functions, the first one being smooth and the second being convex. Under Kurdyka-Lojasiewicz assumption, we prove that the sequence produced by the extragradient method converges to a critical point of this problem and has finite length. When both functions are convex, we provide a $O(1/k)$ convergence rate. Furthermore, we show that our complexity result in the second part can be applied to this method. Considering the extragradient method is the occasion to describe exact line search for proximal decomposition methods. We provide details for the implementation of this scheme for the ℓ_1 regularized least squares problem and give numerical results which suggest that combining nonaccelerated methods with exact line search can be a competitive choice.
Key words: Kurdyka-Lojasiewicz inequality, error bounds, complexity of first order method, extragradient, descent method, forward-backward splitting, ℓ^1-regularized least squares.

Résumé

Cette thèse traite des méthodes de descente d'ordre un pour les problèmes de minimisation. Elle comprend trois parties. Dans la première partie, nous apportons une vue d'ensemble des bornes d'erreur et les premières briques d'unification d'un concept. Nous montrons en effet la place centrale de l'inégalité du gradient de Łojasiewicz, en mettant en relation cette inégalité avec les bornes d'erreur. Dans la seconde partie, en usant de l'inégalité de Kurdyka-Łojasiewicz (KL), nous apportons un nouvel outil pour calculer la complexité des méthodes de descente d'ordre un pour la minimisation convexe. Notre approche est totalement originale et utilise une suite proximale "worst-case" unidimensionnelle. Ces résultats introduisent une méthodologie simple : trouver une borne d'erreur, calculer la fonction KL désingularisante quand c'est possible, identifier les constantes pertinentes dans la méthode de descente, et puis calculer la complexité en usant de la suite proximale "worst-case" unidimensionnelle. Enfin, nons étendons la méthode extragradient pour minimiser la somme de deux fonctions, la première étant lisse et la seconde convexe. Sous l'hypothèse de l'inégalité KL, nous montrons que la suite produite par la méthode extragradient converge vers un point critique de ce problème et qu'elle est de longueur finie. Quand les deux fonctions sont convexes, nous donnons la vitesse de convergence $O(1/k)$ qui est classique pour la méthode de gradient. De plus, nous montrons que notre complexité de la seconde partie peut-être appliquée à cette méthode. Considérer la méthode extragradient est l'occasion de décrire la recherche linéaire exacte pour les méthodes de décomposition proximales. Nous donnons des détails pour l'implémentation de ce programme pour le problème des moindres carrés avec régularisation ℓ^1 et nous donnons des résultats numériques qui suggèrent que combiner des méthodes non-accélérées avec la recherche linéaire exacte peut-être un choix performant.
Mots clés: Inégalité Kurdyka-Lojasiewicz, bornes erreur, complexité de méthode d'ordre un, extragradient, méthode descente, moindre carré avec régularisation ℓ^1.

Remerciements

Je tiens à exprimer ma profonde gratitude à mon directeur de thèse Jérôme Bolte, pour son soutien, son orientation et sa patience au fil des années. De lui, j'ai appris la façon de trouver, de penser et de résoudre les problèmes. Il m'a aussi fait découvrir la langue et la culture française. Je me sens chanceux d'avoir eu l'opportunité de travailler avec lui et je suis très reconnaissant envers lui de toutes les portes qu'il m'a ouvertes.

Un grand merci à Patrick Louis Combettes, qui m'a donné l'opportunité d'étudier en France, puis m'a permis de faire la thèse à Toulouse. Je suis également honoré qu'il accepte d'être membre de mon jury de thèse.

Mes remerciements vont aussi à Jérôme Malick et Tien Son Pham, qui m'ont fait l'honneur de rapporter ce manuscrit, ainsi qu'à Edouard Paulwels, Juan Peypouquet, Aude Rondepierre qui ont accepté de faire partie du jury. Particulièrement, je remercie grandement Edouard Pauwels, pour ses aides volontaires, ses conseils, ses commentaires et pour sa collaboration. Je remercie sincèrement Tien Son Pham pour ses conseils et ses discussions amicales.

Je tiens à remercier Juan Peypouquet, Emile Richard et Bruce Suter ; mes autres collaborateurs sur les travaux qui ont contribué à ma thèse. Leurs contributions ont été et continuent d'être cruciales.

Un grand merci à tous les membres de TSE-R (ancien GREMAQ), Toulouse School of Economics de m'avoir offert d'étudier dans un milieu académique.

Ce travail a été financièrement soutenu par le gouvernement Vietnamien, auquel je suis reconnaissant, à travers le projet 911. Je tiens également à remercier mes collègues de l'Université Nationale de génie civil à Hanoi, au Viet Nam, qui m'ont soutenu dans mon travail.

Je souhaite remercier tous mes amis en France, en particulier Eric, Denis-Annie, Ngambou, Tristan, Giang-Binh, Minh-Lien, Tien-Hang, Long-Hoa, Huan-Le, la groupe de mathématiciens Vietnamiens à Toulouse pour leurs attentions et leurs aides.

Ma plus grande gratitude et mon amour appartiennent à toute ma famille : mes parents, mes grandes sœurs, pour leur amour sans fin et leur soutien inconditionnel, les remerciements les plus spéciaux à ma femme Ngoc pour sa présence auprès de moi. Cette thèse est dédiée à mes deux enfants Ngoc Van et Trong Vu, ils sont ma plus grande source d'inspiration et de motivation.

Introduction

The optimization problem

Optimization problems arise naturally in various fields of applications, throughout science and engineering, it can be seen in the framework of signal/image [38, 17], processing and machine learning [116], optimal control [32], mechanics [30], physics [61], economics [67], computational chemistry and biology [55]. General optimization problems we consider have the general form

$$\min_{C} F(x), \tag{P1}$$

where C is a subset of a Hilbert space H and $F\colon H \to (-\infty, +\infty]$ is a real-extended-valued function on H.

Most optimization problems do not have explicit solutions, the goal of an optimization method, algorithm is to generate an approximation sequence which aims at detecting an an optimal solution or some acceptable approximation. If we denote by $(x_k)_{k\in N}$ such a sequence some crucial questions are

1. convergence of $(F(x_k))_{k\in\mathbb{N}}$ to $\min F(x)$,
2. convergence of the sequence $(x_k)_{k\in\mathbb{N}}$ itself to a solution of (P1) (or to a critical point in the absence of convexity),
3. rate of convergences,
4. complexity issues, i.e. how many steps $\bar{k}$ are necessary to build a feasible point $x_{\bar{k}}$ such that $F(x_{\bar{k}}) \leq \text{val } (P1) + \epsilon$.

A limitation of problem (P1) is that the objective function F usually requires some assumptions (convex, smooth, Lipschitz continuous gradient). The applications are thus limited. Therefore, other forms of problem (P1) have been studied, in which the assumptions on F are relaxed. An important case is the case of a composite function

$$\min_{H} \{F(x) = f(x) + g(x)\} \tag{P2}$$

where one function is smooth with Lipschitz continuous gradient and the other is convex or/and simple (in the sense that its proximal operator is computable).

Due to its important applications, this optimization problem has attracted numerous researches. There are many methods for solving this problem in the literature. In this thesis, we focus on the class of first–order methods, which are generally more effective in the large scale setting. The gradient method is one of the oldest optimization algorithms which is due to Cauchy [34], in 1847. The class of first–order method has been further improved by many researchers.

In the modern era some of the most important works are those of Goldstein [57], Polyak and Levitin [75] (the gradient method) and Nesterov [98] (fast gradient method). Later, the proximal method was introduced by Martinet [94], it was extended to the so-called forward-backward splitting method in [29, 115, 38]. This method is nowadays considered as one of most fundamental first–order method (see also the work of Beck and Teboulle [16] for FISTA method).

Many works were made in the case of convex functions but due to major applications in imaging, learning, signal processing, there was a necessity to investigate the case of nonconvex problems. In this research line, the nonsmooth Kurdyka-Łojasiewicz inequality (KL inequality for short, [22]) has opened the road to many theoretical and algorithmic developments. This inequality was introduced and established in [21, 22]. It asserts that, for any definable function F in an arbitrary o-minimal structure over $\mathbb{R}^n$, there exists a smooth, increasing, nonnegative function φ such that

$$d(0, \partial(\varphi \circ (F - \min F)))(x) \geq 1,$$

for all x in some neighborhood of the set $\operatorname{argmin} F$, the function φ is called the desingularizing KL function. The convergence analysis of descent algorithms for the functions satisfying the Kurdyka-Łojasiewicz was investigated recently. The pioneering work, for smooth unconstrained problems, seems to be that of Absil, Mahony and Andrews [1]. After that, Attouch and Bolte [2] obtained the global convergence for the proximal method. Rate of convergence for this method were shown to depend on the exponent of Łojasiewicz gradient inequality. More recently, the result of Attouch, Bolte and Svaiter [4] provided the framework for the analysis of general descent methods, see also [3, 26, 56].

Yet, despite the important success of KL inequality in many algorithms of optimization problem, the connection of this inequality with the study of the complexity of general descent methods was not known, even if F is a convex function. This PhD thesis addresses this issue and provides some simple relationships on the impact of KL inequality on the complexity of first–order descent method.

First-order descent methods

One of the thesis's contribution is to provide a new protocol to calculate the complexity of a general first–order descent method **(FODM)**. For an object function F, we consider a sequence $(x_k)_{k\in\mathbb{N}}$ which satisfies the two following properties

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

This condition guarantees that $(F(x_k))_{k\in\mathbb{N}}$ is a nonincreasing sequence, and that $\|x_k - x_{k-1}\|$ converges to 0.

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

This property ensures that, if $\|x_k - x_{k-1}\| \to 0$ then there exists a subgradient sequence $(\omega_k)_{k\in\mathbb{N}}$ which converges to 0.

It seems that these conditions were first considered in the work of Luo and Tseng [92]. They were used to study convergence rates from error bounds. The motivation behind this definition is due to the fact that such sequences are generated by many methods, such as the forward-backward splitting method [92, 4, 56], many trust region methods [1], or proximal alternating methods [3, 26].

The complexity of some important methods (gradient method, proximal gradient method) of the class of FODM were presented in several works, see [112, 98, 16]. As we mentioned, the geometric approach (KL inequalities, error bounds) are available but they are only used for

convergence rate analysis. Our work gives the links between these concepts to provide new a complexity analysis. More precisely, 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.

Our results: A new approach to complexity

Our result of complexity is based on the notion of KL inequality and the relationship with error bounds. Assume that F is a convex and satisfies KL inequality with the desingularizing function φ. The major novelty in our approach is to introduce a one–dimensional worst case proximal method associate with φ^{-1}. Our method asserts that, the complexity of any method in the FOMD class, initial at $x_0 \in H$ is controlled by the complexity of a known prox method on the one dimensional function φ^{-1}. In other words:

$$F(x_k) - \min F \leq \varphi^{-1}(\alpha_k),\ \forall k \in \mathbb{N},$$

where $(\alpha_k)_{k\in\mathbb{N}}$ is one–dimensional worst case proximal sequence of φ^{-1}.

One sees that this method requires that the desingularizing function of KL inequality is known explicitly or well estimated. However this is not in general an easy task. Another innovative aspect of our work is to use error bounds to calculate such a desingularizing function φ.

An error bound is an inequality of the form

$$\operatorname{dist}(x, \operatorname{argmin} F) \leq \psi\left((F(x) - \min F)\right),$$

where ψ is an increasing function, vanishing at 0 and x may evolve either in the whole space or in a bounded set. The error bound is an important tool in optimization, see e.g. [66, 106, 54, 9]. An important aspect for us is that they are often used to establish the rate of convergence of optimization methods, see e.g., [52, 91, 92, 27, 118, 97, 16, 40, 108]. However, to the best of our knowledge, the connection between error bounds and the complexity of first–order method was never made before our work.

The development of error bound's theory has a long history. The pioneering works due to Hoffman [63], Robinson [113], Mangasarian [93]. The Łojasiewicz function inequality [84] discovered in a geometrical context, has given as well a lot of research directions. The relationships between error bounds and KL inequalities has been presented in several works, the pionnering work seems to be [23]. In general, KL inequality implies the error bound, (see [65, 9, 23]). However, we show that the converse does not hold, even if F is a convex function (see the counterexample in subsection 2.6.2). In order to improve the effectiveness of our complexity, let us give a summary on the relation of error bounds and KL inequalities. We show that, under convexity assumption, in the particular case (for example: $\psi(s) = \tau(s^\alpha + s^\beta)$–Hölder type), they coincide. This result allows us to obtain the complexity of first–order descent methods from error bounds. To better understand our method, we also present a short overview on error bounds, focusing on the Höder type, which are very useful in practice.

A consequence of our result is to provide a method to calculate the complexity of first–order descent method for a convex function which possesses a known error bound:

- Calculate the desingularizing function φ of KL inequality, the inverse of $\psi = \varphi^{-1}$ and the Lipschitz constant of ψ'.
- Calculate the worst-case proximal sequence $(\alpha_k)_{k\in\mathbb{N}}$ in $\mathbb{R}$. The complexity is then

$$F(x_k) - \min F \leq (\psi \circ \cdots \circ \psi)\,(F(x_0)),\ \forall k \in \mathbb{N}.$$

An illustration of our method is to calculate the complexity of the *iterative shrinkage thresholding algorithm* (ISTA) applied to a least squares objective with ℓ^1 regularization [44]. This method was known to have a complexity of the form $O(\frac{1}{k})$. Our result gives that the complexity is actually of the form $O(q^k)$ with $q \in (0,1)$ (but with different multiplicative constant). This is a new result and this suggests that many questions on the complexity of first-order methods remain open.

Extragradient method for minimization

In another work, our method of complexity can be applied to the extragradient algorithm. This method was first presented by Korpelevich [71] for variational inequality problems (VIP), see also [35, 95] for some extensions. In the context of optimization, under an error bound assumption, Luo and Tseng [91] used this method to solve the smooth, convex constrained minimization problem (P1). More precisely, they studied the convergence of the following sequence

$$\begin{cases} y_k = P_C(x_k - s_k \nabla f(x_k)) \\ x_{k+1} = P_C(x_k - s_k \nabla f(y_k)). \end{cases}$$

By using extra proximal operators at each iteration, we extend the extragradient method to the nonsmooth, nonconvex minimization problem (P2). More explicitly, we generate the sequence $(x_k)_{k\in\mathbb{N}}$ by

$$\begin{cases} y_k = \operatorname{prox}_{s_k g}(x_k - s_k \nabla f(x_k)) \\ x_{k+1} = \operatorname{prox}_{\alpha_k g}(x_k - \alpha_k \nabla f(y_k)). \end{cases}$$

Without convexity, by using the assumption on the nonsmooth KL inequality, we prove that the sequence $(x_k)_{k\in\mathbb{N}}$ converges to a critical point of F and we also obtain some convergence rate. When F is convex, firstly, we describe a $1/k$ non-asymptotic rate in terms of the objective function. This is related to classical results of first order methods in convex optimization, see for example the analysis of the forward-backward splitting method in [17]. Furthermore, we also obtain some complexity for this method by applying our general methodology based on error bounds. Lastly, we compare the effectiveness of the extragradient method and of an exact line line search variant we introduce, to those of FISTA and forward-backward splitting methods. The numerics suggest that, both the extragradient and the forward-backward splitting method, when combined with exact line search, constitute promising alternatives to FISTA.

Structure of thesis The thesis contains three parts:

1. [104] A stroll in the jungle of error bounds

   In this part we provide an overview on error bounds. We consider local and global bounds, qualitative and quantitative results. Inspired by the works in [65, 9, 23, 11], we strongly use the characterization of error bounds in term of strong slope. By specifying this result to the convex case, we consider the interplay between error bounds and KL inequalities. In addition, we present some sufficient conditions so that the two concepts are equivalent. We also review the Hölder type error bound, especially their quantitative versions, which play a major role in complexity theory of many optimization methods.

2. [24] From error bounds to the complexity of first-order descent methods for convex functions.

We study KL inequalities in the convex framework and we use our results to provide a protocol for deriving complexity results of first–order descent methods. Our general methodology is illustrated by the ℓ^1 regularized least squares method and by feasibility problems. We also further give theoretical aspects related to our 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 other related questions.

3. [105] Extragradient Method in Optimization: Convergence and Complexity.

   In this last part, we present an extragradient method to tackle the composite problem (P2). We obtain the convergence and finite length property under KL inequality assumption in the nonconvex case. When F is convex, we give both a proof of a sublinear convergence rate and a complexity analysis under semi-algebraic assumptions. This leads to an improved complexity analysis under a KL inequality assumption. Lastly, we describe an exact line search version of the extragradient method and we discuss computational aspects of the line search part. In the context of ℓ^1 penalized least-squares and results from numerical experiment.

Chapter 1

A stroll in the jungle of error bounds

Abstract The aim of this paper is to give a short overview on error bounds and to provide the first bricks of a unified theory. Inspired by the works of [9, 23, 21, 24, 11], we show indeed the centrality of the Łojasiewicz gradient inequality. For this, we review some necessary and sufficient conditions for global/local error bounds, both in the convex and nonconvex case. We also recall some results on quantitative error bounds which play a major role in convergence rate analysis and complexity theory of many optimization methods.
Key words: Error bounds, Kurdyka–Łojasiewicz inequality, Łojasiewicz function inequality, descent method.

1.1 Introduction

Let X be a Banach space. Given a function $f\colon X \to (-\infty, +\infty]$, an error bound is an inequality that bounds the distance from an arbitrary point in a test set to the level set in terms of the function values. More precisely, we shall say that f has an error bound on a set $K \subset X$ if there exists an increasing function $\varphi\colon [0, +\infty) \to [0, +\infty), \varphi(0) = 0$ such that

$$\operatorname{dist}(x, [f \leq 0]) \leq \varphi([f(x)]_+), \forall x \in K. \tag{1.1}$$

When $K = X$ then f is said to possess global error bound; otherwise we say that f has a local error bound.

Error bounds have a lot of applications in many fields. They may be used to establish the rate of convergence of many optimization methods: we can think to descent methods for solving minimization problems [15, 52, 92, 90, 91, 118], to the cyclic projection algorithm [16, 27, 79], to algorithms for solving variational inequalities, see e.g., [117]. In [24, 120] the error bound theory is used to estimate the complexity of a wealth of descent methods for convex problems. Error bounds have also played a major role in the context of metric regularity [6, 66, 72] or within the field of exact penalty functions, see e.g., [46].

Let us now present the two major mathematical results that are structuring the theory of error bounds and along which we will develop our own presentation. Hoffman seems to be the first to provide an error bound in the context of optimization theory. His result concerns affine function system:

Theorem 1.1.1 (Hoffman, 1952). *[63] Let $A, B \in \mathbb{R}^{m\times n}$ be some matrices and a, b are the vectors in $\mathbb{R}^m$. Assume that*

$$S = \{x \in \mathbb{R}^n | Ax \leq a, Bx = b\}.$$

is nonempty. Then, there exists a scalar $c > 0$ such that

$$\operatorname{dist}(x, S) \leq c\left(\|[Ax - a]_+\| + \|Bx - b\|\right), \forall x \in \mathbb{R}^n.$$

Around the same time, a very general and powerful result was provided in [85] for semi-algebraic functions. This result was developed as a positive response to a conjecture of Schwartz in the distribution theory of functions (see [88]). Later, in [62], Hironaka extended this inequality to the case of subanalytic function.

Theorem 1.1.2 (Łojasiewicz, 1959). *[62, 85] Let $\phi, \psi \colon \mathbb{R}^n \longrightarrow \mathbb{R}$ be two continuous subanalytic functions. If $\phi^{-1}(0) \subset \psi^{-1}(0)$ then for each compact, subanalytic set $K \subset \mathbb{R}^n$, there exist a constant $c > 0$ and a integer N such that*

$$c|\psi(x)|^N \leq |\phi(x)|, \forall x \in K.$$

After those pioneering works, the study of error bounds has attracted numerous researches. In 1972, under Slater's condition and a boundedness assumption on the level sets, Robinson [113] extended the result of Hoffman to systems of convex differentiable inequalities. Mangasarian [93] established the same result for the maximum of finitely many differentiable convex functions. Later on, Auslender and Crouzeix [6], extended Mangasarian's result to non-differentiable convex functions. Some other sufficient conditions were also given by Deng in [47, 48], by using in particular Slater's condition on the recession function. In [76], Lewis and Pang gave a characterization of Lipschitz global error bound for convex functions in terms of the directional derivatives. The work of Lewis and Pang was further generalized by Ng and Yang [101], by Wu and Yu in [123, 122], by Klatte and Li in [69]. In a series papers [7, 9, 8, 10, 11], Azé and Corvellec presented some characterizations of error bounds in terms of the strong slope in the context of metric spaces.

The first fundamental works on quantitative error bounds seem to be those of Gwozdziewicz [58] and Kollár [70] for polynomial functions. Inspired by these works many researchers have tried to provide more general types of quantitative error bounds. Li, Mordukhovich and Pham [80] established a local error bound for polynomial function systems in the nonconvex case. Li [77], and Yang [124] obtained some error bounds for polynomial convex functions, the work of Li was extended for piecewise convex polynomial function in [78], which has also improved the result of Li [81]. In [102], Ngai gave some similar results on polynomial function systems. For the quadratic function systems, Luo and Luo [87] seem to be the first to have studied global error bounds for such class, under the assumption of convexity. This work has been improved by Pang and Wang [121] and later by Luo and Sturm [89] who derived a global error bound for such function without assuming convexity.

The connection between error bound and Kurdyka–Łojasiewicz inequality was first settled by Bolte, Daniilidis, Ley and Mazet, in [23]. Later some of these results were improved [24, 11].

This paper is organized as follows:

In Section 3, based on the results in [9, 23, 11], we give a characterization of error bounds, specifying this result in some particular cases. We establish the connection between this result and some other previous sufficient conditions for Lipschitz error bounds.

In Section 4, we review some results on local error bounds and global error bound, respectively. We focus on the class of polynomial functions whose error bounds are of Hölder type, which play a major role in complexity theory of many optimization methods.

1.2 Preliminaries

Let X be a Banach space, X^* be topological dual space and $f\colon X \longrightarrow \mathbb{R}$ be a lower semi-continuous function. For any $\alpha, \beta \in \mathbb{R}$, we set $[f \leq \alpha] = \{x \in X | f(x) \leq \alpha\}$, $[f = \alpha] = \{x \in X | f(x) = \alpha\}$, $[\alpha \leq f \leq \beta] = \{x \in X | \alpha \leq f(x) \leq \beta\}$, and $[\alpha]_+ = \max\{\alpha, 0\}$. For any subset $S \subset X$, denote $\operatorname{dist}(x, S) = \inf_{u \in S} \|x - u\|$, and $\operatorname{bd} S$, $\operatorname{cl} S$, $\operatorname{int} S$ respectively are the boundary, closure and interior set of S. For $x \in X$, $\delta > 0$, set $B_\delta(x) = \{y \in X | \operatorname{dist}(x, y) < \delta\}$.

The Fréchet subdifferential of f at $x \in \operatorname{dom} f$, denote $\partial^F f(x)$, is defined by

$$\partial^F f(x) = \left\{ u \in X^* | \liminf_{y \to x} \frac{f(y) - f(x) - \langle u, y - x \rangle}{\|y - x\|} \geq 0 \right\}.$$

The limiting-subdifferential of f at $x \in \operatorname{dom} f$, written $\partial f(x)$ is defined as follows

$$\partial f(x) = \left\{ u \in X^* | \exists x_k \to x,\ f(x_k) \to f(x),\ u_k \in \partial^F f(x_k) \to u \right\}.$$

And

$$f'(x, d) = \liminf_{t \to 0^+} \frac{f(x + td) - f(x)}{t},$$

is called the derivative of f at x in the direction $d \in X$.

The strong slope of f at x is given by

$$|\nabla f|(x) = \begin{cases} 0 & \text{if } x \text{ is a local minimum point of } f, \\ \limsup_{y \to x} \frac{f(x) - f(y)}{\|x - y\|} & \text{otherwise.} \end{cases}$$

It is easy to see that

- $\|d\| |\nabla f|(x) \geq -f'(x, d),\ \forall (x, d) \in X^2$.
- $|\nabla f|(x) \leq \operatorname{dist}(0, \partial^F f(x)),\ \forall x \in X$.

We recall the chain rule for the strong slope.

Lemma 1.2.1. *[11] Let $-\infty < \alpha < \beta \leq +\infty$ and a function $\varphi\colon]\alpha, \beta[\to \mathbb{R}$ with $\varphi \in C^1(\alpha, \beta)$ and $\varphi'(s) > 0$, $\forall s \in]\alpha, \beta[$. One has*

$$|\nabla(\varphi \circ f)|(x) = \varphi'(f(x)) |\nabla f|(x),\ \forall x \in [\alpha < f < \beta].$$

As mentioned in the introduction, the theory of error bound can be developed, based on the theory of Łojasiewicz on the subanalytic function. Let us recall the definition of such function class.

Definition 1. *[21]*

(i) *A function $f\colon \mathbb{R}^n \to \mathbb{R}$ is real-analytic on $S \subset \mathbb{R}^n$ if it can be represented locally on S by a convergent power series, this means that, for any $\bar{x} = (\bar{x}_1, \ldots, \bar{x}_n) \in S$, there exists a neighbourhood $U(\bar{x})$ such that*

$$f(x) = \sum_{i_1, \ldots, i_n = 0}^{\infty} a_{i_1, \ldots, i_n} (x_1 - \bar{x}_1)^{i_1} \ldots (x_n - \bar{x}_n)^{i_n},\ \forall x \in U(\bar{x}).$$

(ii) *A subset S of $\mathbb{R}^n$ is called semianalytic if for each point $\bar{x} \in \mathbb{R}^n$ admits a neighborhood $U(\bar{x})$ for which $S \cap U(\bar{x})$ is represented in the following form*

$$S \cap U(\bar{x}) = \bigcup_{i=1}^{p} \bigcap_{j=1}^{q} \{x \in \mathbb{R}^n | f_{ij}(x) = 0,\, g_{ij}(x) > 0\},$$

where the functions f_{ij}, $g_{ij} \colon \mathbb{R}^n \to \mathbb{R}$ are real-analytic for all $1 \le i \le p$, $1 \le j \le q$. If the graph of f is a semianalytic set in $\mathbb{R}^{n+1}$, we say that f is a semianalytic function.

(iii) *S is called subanalytic if each point $\bar{x} \in \mathbb{R}^n$, there exist a neighbourhood $U(\bar{x})$ and a bounded semianalytic set $A \subset \mathbb{R}^{n+m}$, (for some $m \in \mathbb{N}^*$) such that $S \cap U(\bar{x})$ is the projection on $\mathbb{R}^n$ of A. The function f is subanalytic if its graph is a subanalytic set in $\mathbb{R}^{n+1}$.*

We give some elementary properties of subanalytic function and subanalytic set, see[21].

1. If S is subanalytic set then so are its boundary $\operatorname{bd} S$, its closure $\operatorname{cl} S$, its interior $\operatorname{int} S$, and its complement set.
2. The class of subanalytic sets is closed under finite union and intersection. The distance function to a subanalytic set is a subanalytic function.
3. The image of a bounded subanalytic set under a subanalytic map is subanalytic. The inverse image of a subanalytic set under a subanalytic map is a subanalytic set.
4. When S is a closed, convex subanalytic set, the Euclidean projector onto S is a subanalytic function.

In this work, we focus on the Hölder-type error bound, which is very common in practice.

Definition 2. *Let f be a function on the Banach space X and K be a subset of X. We say that f admits a*

1. *Hölder-type error bound on K if there exists $\tau > 0$ and $a, b > 0$ such that*

$$\operatorname{dist}(x, [f \le 0]) \le \tau \left([f(x)]_+^a + [f(x)]_+^b \right), \quad \forall x \in K. \tag{1.2}$$

2. *Lipschitz-type (or linear) error bound on K if the inequality (1.2) holds with $a = b = 1$, for all $x \in K$.*

When $K \equiv X$ then f is said to have a global error bound, otherwise we say that f possesses a local error bound.

1.3 Characterization of error bounds

1.3.1 Characterizing error bounds through Kurdyka–Łojasiewicz inequality

The Łojasiewicz gradient inequality was introduced in [86]. Let $f \colon \mathbb{R}^n \to \mathbb{R}$ be an analytic function. For any $\bar{x} \in \operatorname{dom} f$, there exist $\tau > 0$, $\theta \in [0, 1)$ and a neighbourhood $U(\bar{x})$ such that

$$\|\nabla f(x)\| \ge \tau \left| f(x) - f(\bar{x}) \right|^{\theta}, \forall x \in U(\bar{x}).$$

In [73], Kurdyka generalized the above result to the class of C^1 functions whose graphs belong to an o-minimal structure (the definition of o-minimal structure can be seen in [73, Definition 1], [22, Definition 6]), this result was extended to the nonsmooth class by Bolte, Danillidis, Lewis and Shiota [22]. The corresponding generalized Łojasiewicz gradient inequality is called the Kurdyka–Łojasiewicz (KL for short) inequality. In addition, the generalization for the class nonsmooth subanalytic functions has been obtained by Bolte, Danillidis and Lewis in [21]. This has opened the road to many theoretical and algorithmic developments (see [1, 2, 3, 4, 26, 24, 56]). We summarize the above extension by the following theorem.

Theorem 1.3.1. *[22, 21] Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be a lower semicontinuous, definable function in an arbitrary o-minimal structure over $\mathbb{R}$. Then for all $\bar{x} \in \operatorname{dom} f$, there exist $\delta > 0$ and a neighbourhood $U(\bar{x})$ of $\bar{x}$ such that*

$$\varphi'\left(f(x) - f(\bar{x})\right) \operatorname{dist}\left(0, \partial f(x)\right) \geq 1,\ \forall x \in U(\bar{x}) \cap [f(\bar{x}) < f < f(\bar{x}) + \delta],$$

where $\varphi\colon [0,\delta] \to [0,+\infty)$ is an increasing function, which vanishes at zero and $\varphi \in C^0[0,\delta] \cap C^1(0,\delta)$. The class of such functions φ will be denoted by $\mathcal{K}[0,\delta]$.

The connection between error bounds and Kurdyka–Łojasiewicz inequality was established in [23] (see also [24]), this was further improved by Azé and Corvellec [11].

Azé and Corvellec have series of researches on the characterization of global error bounds for lower semicontinuous functions in terms of the strong slope, see [9, 8, 10, 11, 41]. These works are of great help for this section. Let us now give the result in [9], in which, the authors used Ekeland's variational principle to establish the connection between the strong slope and the linear error bound.

Theorem 1.3.2. *[9] Let $f\colon X \to \mathbb{R} \cup \{+\infty\}$ be a lower semicontinuous function, and $-\infty < \alpha < \beta \leq \infty$. Then*

$$\inf_{x \in [\alpha < f < \beta]} |\nabla f|(x) = \inf_{\alpha \leq \gamma < \beta} \left(\inf_{x \in [\gamma < f < \beta]} \frac{f(x) - \gamma}{\operatorname{dist}(x, [f \leq \gamma])} \right).$$

We rewrite the latter theorem as a characterization of linear global error bound, which is a well known result since Ioffe's pioneering works [65].

Theorem 1.3.3. *[9] Let $\tau > 0$, the following assertions are equivalent*

(i) $|\nabla f|(x) \geq \frac{1}{\tau}$, $\forall x \in [\alpha < f < \beta]$.

(ii) $\tau\left(f(x) - \gamma\right) \geq \operatorname{dist}\left(x, [f \leq \gamma]\right)$, $\forall \gamma \in [\alpha, \beta)$, $x \in [\gamma < f < \beta]$.

For any $\varphi \in \mathcal{K}[0, \beta - \alpha]$, thanks to Lemma 1.2.1, we can apply the latter result for the function $x \mapsto \varphi\left(f(x) - \alpha\right)$, therefore we obtain a nonlinear version of Theorem 1.3.3.

Theorem 1.3.4. *Assume that $\varphi \in \mathcal{K}[0, \beta - \alpha)$. The following statements are equivalent*

(i) $\varphi'(f(x) - \alpha)|\nabla f|(x) \geq 1$, $\forall x \in [\alpha < f < \beta]$.

(ii) $\varphi\left(f(x) - \alpha\right) \geq \varphi(\gamma - \alpha) + \operatorname{dist}\left(x, [f \leq \gamma]\right)$, $\forall \gamma \in [\alpha, \beta]$, $\forall x \in [\gamma < f < \beta]$.

This is content of [23, Corollary 4], [11, Theorem 4.2]. In the latter result, if we let γ equal to α in the assertion (ii), then we immediately obtain as a consequence, a sufficient condition for nonlinear global error bound.

Corollary 1.3.5. *We suppose that*

$$\varphi'(f(x)-\alpha)|\nabla f|(x) \geq 1,\ \forall x \in [\alpha < f < \beta],$$

where $\varphi \in \mathcal{K}[0, \beta-\alpha)$. *Then*

$$\varphi\left(f(x)-\alpha\right) \geq \operatorname{dist}\left(x, [f \leq \alpha]\right),\ \forall x \in [\alpha < f < \beta].$$

Generally, the converse of this corollary is false, as shown in [74, Remark 3] (when f is a polynomial function) and in [24, Theorem 28] (when f is convex). However, in some particular cases, this converse may be hold, for example:

- f is an analytic function with an isolated zero, see [58].
- f is a convex function and an additional assumption on φ, see [24], (we also show this result in Theorem 1.3.11).

Recalling $\|d\||\nabla f|(x) \geq -f'(x,d), \forall (x,d) \in X^2$, a consequence of Theorem 1.3.3 is:

Corollary 1.3.6. *For any* $\tau > 0$, *suppose that for each* $x \in [\alpha < f < \beta]$, *there exists a unit vector* $d_x \in X$ *such that*

$$f'(x, d_x) \leq -\frac{1}{\tau}.$$

Then

$$\tau\left(f(x)-\alpha\right) \geq \operatorname{dist}\left(x, [f(x) \leq \alpha]\right),\ \forall x \in [\alpha < f < \beta].$$

This is the content of [101, Theorem 2.5]. A local version of Theorem 1.3.4 is given as follows

Theorem 1.3.7. *[11] Consider the following statements*

(i) *There exists* $\varepsilon > 0$ *such that*

$$\varphi'(f(x)-\alpha)|\nabla f|(x) \geq 1,\ \forall x \in B_\varepsilon(\bar{x}) \cap [\alpha < f < \beta].$$

(ii) *There exists* $\rho > 0$ *such that*

$$\varphi\left(f(x)-\alpha\right) \geq \varphi(\gamma-\alpha) + \operatorname{dist}\left(x, [f(x) \leq \gamma]\right),\ \forall \gamma \in [\alpha, \beta),\ \forall x \in B_\rho(\bar{x}) \cap [\alpha < f < \beta].$$

Then $(i) \Rightarrow (ii)$ *with* $\rho = \varepsilon/2$ *and* $(ii) \Rightarrow (i)$ *with* $\varepsilon = \rho$.

In the statement (ii), by setting $\gamma = \alpha$, we obtain a local version of Corollary 1.3.5.

Corollary 1.3.8. *[11] For any* $\bar{x} \in [f \leq \alpha]$, *suppose that there exists* $\varepsilon > 0$ *such that*

$$\varphi'(f(x)-\alpha)|\nabla f|(x) \geq 1,\ \forall x \in B_{2\varepsilon}(\bar{x}) \cap [\alpha < f < \beta].$$

Then

$$\varphi\left(f(x)-\alpha\right) \geq \operatorname{dist}\left(x, [f(x) \leq \alpha]\right),\ \forall x \in B_\varepsilon(\bar{x}) \cap [\alpha < f < \beta].$$

If we take $\varphi(s) = \tau s^\theta$, $\tau > 0$, $\theta \in [0, 1]$, then this corollary recover the result of Ngai, Thera [119, Corollary 2].

As we mentioned before, the converse of the latter corollary does not always hold. The results of Corollary 1.3.5, Corollary 1.3.8 have appeared in numerous works, for instance, see [58, 74, 103, 119, 110, 111]. This result gives an useful tools for establishing the quantitative error bounds, see [80, 79].

1.3.2 Equivalence in the convex case

In the sequel, we suppose that $f\colon X \to \mathbb{R} \cup \{+\infty\}$ is a proper lower semicontinuous convex function. The following extra-properties are available .

- $\partial^F f(x) = \partial f(x) = \{u \in X^* | \langle u, y - x\rangle \leq f(y) - f(x),\ \forall y \in X\},\ \forall x \in \operatorname{dom} f.$
- $|\nabla f|(x) = \operatorname{dist}(0, \partial f(x)), \forall x \in X.$

In the convex case, Theorem 1.3.2 can be simplified by the following proposition.

Proposition 1.3.9. *[9] For $-\infty < \alpha < \beta \leq +\infty$, the following assertions hold true:*

(i) $|\nabla f|(x) = \sup_{f(z)\leq f(x)} \frac{f(x)-f(z)}{\operatorname{dist}(x,z)}$, *with x is not a minimum point of f.*

(ii) $\inf_{[\alpha<f<\beta]} |\nabla f|(x) \geq \inf_{[f=\alpha]} |\nabla f|(x)$.

(iii) $\inf_{\alpha\leq\gamma<\beta} \left(\inf_{x\in[\gamma<f<\beta]} \frac{f(x)-\gamma}{\operatorname{dist}(x,[f\leq\gamma])}\right) = \inf_{x\in[\alpha<f<\beta]} \frac{f(x)-\alpha}{\operatorname{dist}(x,[f\leq\alpha])}$.

Thanks to Proposition 1.3.9, the convex version of Theorem 1.3.3 is given as follows.

Theorem 1.3.10. *Suppose $-\infty < \alpha < \beta \leq +\infty$ and $\tau > 0$. Consider the following statements*

(i) $\inf\limits_{x\in[f=\alpha]} \operatorname{dist}(0, \partial f(x)) \geq \dfrac{1}{\tau}.$

(ii) $\inf\limits_{x\in[\alpha<f<\beta]} \operatorname{dist}(0, \partial f(x)) \geq \dfrac{1}{\tau}.$

(iii) $\tau\,(f(x) - \alpha) \geq \operatorname{dist}\left(x, [f(x) \leq \alpha]\right),\ \forall x \in [\alpha < f < \beta].$

Then $(i) \Longrightarrow (ii) \Longleftrightarrow (iii)$.

We mention that the assumption (i) in the above theorem is equivalent to the condition $0 \notin \operatorname{cl}\left(\partial f\left(f^{-1}(0)\right)\right)$, which is called *strong Slater's condition* [76, 101].

We now consider the converse of Corollary 1.3.5. Assume that

$$\varphi(f(x) - \alpha) \geq \operatorname{dist}(x, [f \leq \alpha]),\ \forall x \in [\alpha < f < \beta],\ \varphi \in \mathcal{K}(0, \beta - \alpha),$$

which is equivalent to

$$\frac{\varphi(f(x) - \alpha)}{f(x) - \alpha} \frac{f(x) - \alpha}{\operatorname{dist}(x, [f \leq \alpha])} \geq 1,\ \forall x \in [\alpha < f < \beta].$$

Thanks to Proposition1.3.9, the latter inequality implies that

$$\frac{\varphi(f(x) - \alpha)}{f(x) - \alpha} \operatorname{dist}(0, \partial f(x)) \geq 1,\ \forall x \in [\alpha < f < \beta].$$

Thus, if φ satisfies the condition

$$\int_0^{\beta-\alpha} \frac{\varphi(s)}{s} ds < +\infty,$$

then we get

$$\psi'(f(x) - \alpha) \operatorname{dist}(0, \partial f(x)) \geq 1,\ \forall x \in [\alpha < f < \beta],$$

where

$$\psi(s) = \int_0^s \frac{\varphi(t)}{t} dt,\ \forall s > 0.$$

Therefore, when f is convex, the converse of Corollary 1.3.5 is given as following.

Theorem 1.3.11. *Assume that* $\varphi\left(f(x)-\alpha\right) \geq \operatorname{dist}\left(x,[f\leq\alpha]\right), \forall x \in [\alpha < f < \beta]$, *where* $\varphi \in \mathcal{K}[0,\beta-\alpha)$ *and*

$$\int_0^{\beta-\alpha} \frac{\varphi(s)}{s} ds < +\infty. \tag{1.3}$$

Then, we get

$$\psi'(f(x)-\alpha)\operatorname{dist}(0,\partial f(x)) \geq 1, \forall x \in [\alpha < f < \beta], \quad \text{where } \psi(s) = \int_0^t \frac{\varphi(s)}{s} ds.$$

This result has been presented in [24, Theorem 6], [23, Theorem 30]. We remark that when $\varphi(s) = \tau s^{\theta}$, $(\tau,\, \theta > 0)$, then the condition (1.3) holds.

We will show that the Theorem 1.3.10 covers numerous results on Lipschitz global error bounds in the literature.

- In [113], Robinson proved that if f satisfies the Slater condition (there exists $\bar{x}$ such that $f(\bar{x}) < 0$) and the set $[f \leq 0]$ is bounded then f has a Lipschitz global error bound. More generally, in [48], Deng proved the following fact: If there exist $\delta > 0$, $\Delta > 0$ such that

  $$[f \leq -\delta] \neq \emptyset \text{ and } \sup_{[f\leq 0]} \operatorname{dist}(x,[f\leq -\delta]) \leq \Delta,$$

  then

  $$\operatorname{dist}(x,[f\leq 0]) \leq \frac{\Delta}{\delta}[f(x)]_+, \ \forall x \in X.$$

  Let us show that this result is actually a consequence of Theorem 1.3.10. Indeed, take $x \in [f = 0]$ and $u \in \partial f(x)$. For any $\varepsilon > 0$, there exists $z \in [f \leq -\delta]$ such that $\operatorname{dist}(x,z) \leq \Delta + \varepsilon$. Thus, we obtain

  $$\delta \leq f(x) - f(z) \leq \|u\|\|x-z\| \leq \|u\|(\Delta+\varepsilon),$$

  which implies that

  $$\inf_{[f=0]} \operatorname{dist}(0,\partial f(x)) \geq \frac{\delta}{\Delta}.$$

  Combining with Theorem 1.3.10, f has Lipschitz global error bound.

  Note that Deng's result [48, Theorem 1] also covers the one in [47], in which the author start form the assumption that there exist a unit vector u and a constant $\tau > 0$ such that

  $$f^{\infty}(u) = \sup_{t>0} \frac{f(x+tu)-f(x)}{t} \leq -\frac{1}{\tau} \tag{1.4}$$

  to derive that $\operatorname{dist}(x,[f\leq 0]) \leq \tau[f(x)]_+, \forall x \in X$.

- The work of Robinson was also generalized in other directions. More precisely, instead of the boundedness assumption on the set $[f \leq 0]$, in [93], Mangasarian used the asymptotic constraint qualification condition (this means *for any sequence* $(x_k)_{k\in\mathbb{N}} \subset [f = 0]$ *such that* $\lim\|x_k\| = \infty$, *then the zero vector is not a limit point of any sequence* $(u_k)_{k\in\mathbb{N}}$, *with* $u_k \in \partial f(x_k)$) to obtain a Lipschitz global error bound for differentiable convex function. Auslender and Crouzeix in [6] extended the work of Mangasarian to the case nonsmooth convex functions. On the other hand, in [69, Theorem 2], Klatte and Li proved that, a

convex function satisfies the Slater and the asymptotic qualification conditions if and only if

$$\inf_{x\in[f=0]} \operatorname{dist}(0, \partial f(x)) > 0.$$

Therefore, it is clear that the results of Mangasarian [93], Auslender and Crouzeix [6] are the consequences of Theorem 1.3.10.

- In [76], Lewis and Pang characterized Lipschitz error bounds using directional derivatives as follows. Let $f\colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a lower semicontinuous and convex function. They proved that the Lipschitz global error bound holds for f:

$$\operatorname{dist}(x, [f \leq 0]) \leq \tau f(x), \quad \forall x \in \mathbb{R}^n,$$

if and only if

$$f'(\bar{x}, d) \geq \tau\|d\|, \forall \bar{x} \in [f = 0],\ d \in N_{[f\leq 0]}(\bar{x}),$$

where the cone normal is defined by $N_S(\bar{x}) = \{u \in X^* | \langle u, y - \bar{x}\rangle \leq 0, \forall y \in S\}$, $\forall S \subset \mathbb{R}^n$. This result has been obtained by several other researchers, we can mention here the works of Ng and Zheng [100], [101] where they characterized error bounds for lower semicontinous functions.

Consider now [101, Theorem 3.1].

Suppose that X is a reflexive Banach space. Then the following statements are equivalent

(i) $\operatorname{dist}(x, [f \leq 0]) \leq \tau[f(x)]_+$, for all $x \in X$.

(ii) For each $x \in [f = 0]$, we get

$$\inf\{f'(x, d) | d \in N_{[f=0]}(x), \|d\| = 1\} \geq \frac{1}{\tau}.$$

(iii) For each $x \in X\backslash[f \leq 0]$, there exists $d_x \in X, \|d_x\| = 1$, such that

$$f'(x, d_x) \leq -\frac{1}{\tau}.$$

Let us prove that the (iii) above assertions are equivalent to (ii) of Theorem 1.3.10. Assume that the assumption (iii) holds, then for all $x \in X\backslash[f \leq 0]$, we get

$$|\nabla f|(x) \geq -f'(x, d) \geq \frac{1}{\tau}.$$

Conversely, suppose that $|\nabla f|(x) \geq \frac{1}{\tau}, \forall x \in X\backslash[f \leq 0]$. Take any $x \in X$, we get

$$\operatorname{dist}(0, \partial f(x)) \geq \frac{1}{\tau}, \forall x \in X\backslash[f \leq 0],$$

hence there exists $d_x \in X$ such that

$$-\tau = \inf\{\langle x^*, d\rangle | x^* \in X^*, \|x^*\| \leq \tau\} \geq \sup\{\langle u, d\rangle | u \in \partial f(x)\} = f'(x, d).$$

It follows that

$$f'(x, d_x) \leq -\frac{1}{\tau}.$$

Similarly, by setting $\varphi(s) = \tau s^\theta$, $(\tau,\ \theta > 0)$, we can see that the following result of Ng and Zheng [100] is also a consequence of Theorem 1.3.11:

Let X be a reflexive Banach space and $f\colon X \to \mathbb{R}\cup\{+\infty\}$ a continuous function. Suppose that for each $x \in X\backslash S$, there exists $d_x \in X, \|d_x\| = 1$ and $\tau > 0, \theta \in (0,1)$ such that

$$f'(x, d_x) \leq -\tau f^{1-\theta}(x).$$

Then we get

$$\operatorname{dist}(x, [f \leq 0]) \leq \tau [f(x)]_+^{\theta}, \forall x \in X.$$

1.3.3 Qualification conditions and error bounds

1.3.3.1 Slater's condition and error bounds

We recall that if there exists $\bar{x}$ such that $f(\bar{x}) < 0$ then f is said to satisfy the Slater condition. This condition plays an important role for the study of error bounds. The existence of the Lipschitz global error bound usually requires the convexity and the Slater condition. We consider the following example, which shows that for a convex function without the Slater condition, the Lipschitz global error bound may fail to hold.

Example 1. *[76]* $f(x,y) = x + \sqrt{x^2+y^2}$, $(x,y) \in \mathbb{R}^2$.

It is easy to check that the function f is convex, nonnegative on $\mathbb{R}^2$ and $[f = 0] = \{(x,0) | x \leq 0\}$ has empty interior. Take the sequence $(z_k = (-k, 1))_{k\in\mathbb{N}}$ then $f(z_k)$ converges to 0 but $\operatorname{dist}(z_k, [f \leq 0]) = 1, \forall k \in \mathbb{N}$, so that there is not global error bound for S.

As mentioned earlier, the Slater condition was used for the first time by Robinson [113].

Theorem 1.3.12. *[113] Let $f_1, \ldots, f_m$ be convex functions on $\mathbb{R}^n$ and assume that there is $\bar{x}$ such that $f_i(\bar{x}) < 0, \ldots, f_m(\bar{x}) < 0$. Then there exists $\tau > 0$ such that*

$$\operatorname{dist}(x, [f \leq 0]) \leq \tau \|x - \bar{x}\| \sum_{i=1}^{m} [f_i(x)]_+, \forall x \in \mathbb{R}^n,$$

where $f(x) = \max_{i=1,\ldots,m} f_i(x)$.

In additional, when $\{x \in \mathbb{R}^n | f_i(x) \leq 0, (i = 1, \ldots, m)\}$ is bounded then there exists $\tau > 0$ such that

$$\operatorname{dist}(x, [f \leq 0]) \leq \tau \sum_{i=1}^{m} [f_i(x)]_+, \forall x \in \mathbb{R}^n.$$

As a consequence, we immediately deduce that the convex function systems $f_1 \ldots, f_m$ has Lipschitz local error bound.

Luo and Luo [87] used the Slater condition to establish the Lipschitz global error bound for convex quadratic systems, this result has been extended by Pang and Wang [121]. In general, the Slater condition is not sufficient to ensure that the global error bound holds, even if f is a convex function. We consider the following example:

Example 2. *[77] Let $f_1, f_2\colon \mathbb{R}^4 \to \mathbb{R}$ be defined by $f_1(x) = x_1$ and*

$$f_2(x) = x_1^{16} + x_2^8 + x_3^6 + x_1^2 x_2^4 x_3^2 + x_2^2 x_3^4 + x_1^4 x_4^4 + x_1^4 x_2^6 + x_1^2 x_2^6 + x_1^2 + x_2^2 + x_3^2 - x_4,$$

for all $x = (x_1, x_2, x_3, x_4) \in \mathbb{R}^4$. Define $f(x) = \max\{f_1(x), f_2(x)\}$, $\forall x \in \mathbb{R}^4$.

We get the following properties, (see [77]).

(i) f_1, f_2 are convex polynomial functions, therefore f is convex.

(ii) f satisfies the Slater condition.

(iii) For any $\alpha, \beta \in \mathbb{R}$ with $\alpha \leq \beta$, then $\sup_{x \in [f \leq \beta]} \operatorname{dist}(x, [f \leq \alpha]) = +\infty$.

By taking $\alpha = 0$, $\beta = 1$ in the property (iii), we imply that there exists a sequence $(x_k)_{k \in \mathbb{N}} \subset [f \leq 1]$ such that $\operatorname{dist}(x_k, [f \leq 0]) = +\infty$, this show that f can not possess the Hölder global error bound.

However enhancing the assumptions we can derive global error bounds from the Slater like condition. For instance, let f be a lower semicontinuous, convex function on $\mathbb{R}^n$ which satisfies the Slater condition, then f has Lipschitz gloabl error bound if one of the following assertions holds.

1. f can be expressed as maximum of finitely many bounded below convex polynomials function on $\mathbb{R}^n$, i.e: $f(x) = \max_{i=1,\ldots,d} f_i(x)$, $\forall x \in \mathbb{R}^n$, where f_i is a polynomial function on $\mathbb{R}^n$ with $\inf f_i > -\infty$, for all $i = 1, \ldots, d$, see [77, Theorem 4.1].

2. f is a separable function (in the sense that $f(x) = \sum_{i=1}^n f_i(x_i)$ where $x = (x_1, \ldots, x_n)$ and each f_i is a lower semicontinuous function), see[77, Theorem 4.1].

3. f is well-posed (*for any sequence* $\{x_k\}_{k \in \mathbb{N}}$ *for which* $\operatorname{dist}(0, \partial f(x_k)) \to 0$ *then* $f(x_k) \to \inf_X f$), see [76, Corollary 1].

4. f satisfies the asymptotic qualification condition, see [6].

Notice that if f is a convex function, then f satisfies the Slater condition if and only if $0 \notin \partial f\left(f^{-1}(0)\right)$. We can easily see that, if f satisfies the Slater condition and the level set $[f \leq 0]$ is bounded then f possesses the strong Slater condition. Furthermore, in [69], Klatte and Li proved that, for a convex function $f \colon \mathbb{R}^n \to (-\infty, +\infty]$ which satisfies the Slater condition, the following conditions are equivalent:

1. The strong Slater condition holds.

2. The asymptotic qualification condition is satisfied.

3. $\sup_{x \in [f=0]} \inf_{y \in [f<0]} \frac{\|x-y\|}{-f(y)} < +\infty$.

1.3.3.2 Abadie qualification condition and error bounds

We begin this subsection by considering an example:

Example 3. *[82]* For $(x, y) \in \mathbb{R}^2$, take $f_1(x, y) = x + y$, $f_2(x, y) = -x - y$, $f_3(x, y) = (x + y)^2$ and $f(x, y) = (f_1, f_2, f_3)(x, y)$. Then $[f \leq 0] = \{(x, -x) | x \in \mathbb{R}^n\}$ has no interior point, but we can check that $\operatorname{dist}((x, y), [f \leq 0]) \leq 2\|[f(x, y)]_+\|$, for all $(x, y) \in \mathbb{R}^2$.

This means that, the global error bound may be hold without the Slater condition. In [82], Li used the Abadie qualification condition to characterize Lipschitz-type error bound for convex quadratic systems.

Recall that, the tangent cone of $S \subset \mathbb{R}^n$ is defined by

$$T_S(x) = \{y \in \mathbb{R}^n | \langle u, y \rangle \leq 0, \forall u \in N_S(\bar{x})\}.$$

Let us now recall the definition of Abadie's condition.

Definition 3. *[82] We say that the systems $f_1, f_2, \ldots, f_m \colon X \to \mathbb{R}$ satisfies the Abadie condition at $\bar{x} \in S = \{x \in X | f_i(x) \leq 0, i = 1, \ldots, m\}$ if*

$$T_S(\bar{x}) = \{u \in X | \, \langle f_i'(\bar{x}), u\rangle \leq 0, \, \forall i \in I(\bar{x})\},$$

where $I(\bar{x}) = \{i : f_i(\bar{x}) = 0\}$.

If this property holds at every point in S, then we say that the systems $f_1, f_2, \ldots, f_m$ satisfies the Abadie condition on S.

When $X = \mathbb{R}^n$ and $f_1, \ldots, f_m$ are convex functions, we have the two following properties, see [82].

1. The system $f_1, f_2, \ldots, f_m$ satisfies the Abadie condition at $\bar{x} \in S$ if and only if

$$N_S(\bar{x}) = \left\{ \sum_{i \in I(\bar{x})} \lambda_i f_i'(\bar{x}) | \lambda_i \geq 0 \right\}.$$

2. If there exists $x \in S$ such that $f_i(x) < 0$ with f_i is not affine function, for all $i = 1, \ldots, m$ then the systems $f_1, f_2, \ldots, f_m$ satisfies Abadie's condition on S.

Let us now give a necessary and sufficient condition for a convex quadratic system to have a Lipschitz-type global error bound, which was established by Li [82, Theorem 4.2].

Theorem 1.3.13. *[82] Let $f_1, f_2, \ldots, f_m$ be convex quadratic functions on $\mathbb{R}^n$ such that $S = \{x \in \mathbb{R}^n | f_i(x) \leq 0, (i = 1 \ldots, m)\}$ is nonempty. The following statements are equivalent*

(i) *The system $(f_i)_{i=1\ldots,m}$ satisfies the Abadie condition on S.*

(ii) *There exists $\tau > 0$ such that*

$$\operatorname{dist}(x, S) \leq \tau \sum_{i=1}^{m} [f_i(x)]_+, \, \forall x \in \mathbb{R}^n.$$

Later, in [103, Theorem 6], Ngai and Théra extended this result in the Banach space. In which, $f_i \colon X \to \mathbb{R}, i = 1, \ldots, m$ are defined by

$$f_i(x) = \frac{1}{2}\langle A_i x, x\rangle + \langle B_i, x\rangle + c_i,$$

where $A_i \colon X \times X \to \mathbb{R}$ be a symmetric continuous bilinear and semi-definite positive, $B_i \in X^*$ and $c_i \in \mathbb{R}$, for $i = 1, \ldots, m$. In this paper, Ngai and Théra also gave the relation between the Abadie condition and the Lipschitz local error bound for the convex function systems.

Theorem 1.3.14. *[103] Let $f_1, \ldots, f_m$, be convex continuous functions on the neighborhood of $\bar{x} \in S = \{x \in X | f_i(x) \leq 0, \, i = 1, \ldots, m\}$. Set $f = \max_{i=1,\ldots,m} f_i$.*

(i) *If there exist $\tau > 0$, $\varepsilon > 0$ such that*

$$\operatorname{dist}(x, S) \leq \tau \sum_{i=1}^{m} [f_i(x)]_+, \, \forall x \in B_\varepsilon(\bar{x}) \cap K,$$

then the Abadie condition is satisfied on $B_\delta(\bar{x}) \cap S$ for some $\delta > 0$.

(ii) *If $f_i, \ldots, f_m$ are differentiable on $B_\delta(\bar{x})$, then the converse of part (i) holds.*

1.4 Existence and quantitative results

The first result on local error bound was deduced from the result of Hörmander, in his work on the fundamental solution of partial differential equation.

Theorem 1.4.1 (Hömander, 1958)**.** *[64] Let f be a polynomial function on $\mathbb{R}^n$. Under the assumption that $[f \leq 0]$ is nonempty, there exist $\tau > 0$, $a > 0$ and $b \in \mathbb{R}$ such that*

$$\operatorname{dist}(x, [f \leq 0]) \leq \tau \left(1 + \|x\|\right)^b [f(x)]_+^a, \forall x \in \mathbb{R}^n.$$

This "error bound" has an extra factor of $(1 + \|x\|)^b$. One sees that we can remove this extra factor when restricting the error bound to a bounded region, in that case this local error bound for f can be deduced from. Luo and Luo applied the above theorem to obtain the Hölder local error bound for polynomial function systems [87, Theorem 2.2], this result was extended for analytic systems, by Luo and Pang [88, Theorem 2.2]. Recently, Kurdyka and Spondzieja [74, Corollary 10] showed that the exponents a, b in Theorem 1.4.1 can be computed explicitly:

$$b = 2,\ a = \frac{1}{d(6d-3)^{n-1}}.$$

Result of Łojasiewicz A very general local error bound is deduced from the result of Łojasiewicz, Theorem 1.1.2, if we take $\phi(x) = f(x)$ and $\psi(x) = \operatorname{dist}(x, [f \leq 0])$, we get a local error bound result for subanalytic functions, also called Łojasiewicz function inequality. It also includes a special case of the polynomial equation studied by Hörmander.

Theorem 1.4.2. *[85] Let $f \colon \mathbb{R}^n \to \mathbb{R}$ be a continuous subanalytic function. For any compact set $K \subset \mathbb{R}^n$, there exist $\tau > 0, a > 0$ such that*

$$\operatorname{dist}(x, [f \leq 0]) \leq \tau [f(x)]_+^a,\ \forall x \in K.$$

With a direct application of the Theorem 1.4.2 to a subanalytic system, we recover a result of Luo and Pang [88, Theorem 2.2], in which they obtained the similar result for an analytic system:

Theorem 1.4.3. *Let $f_1, f_2, \ldots, f_r$ and $g_1, g_2, \ldots, g_s$ be continuous subanalytic functions on $\mathbb{R}^n$, set*

$$S = \{x \in \mathbb{R}^n | f_i(x) \leq 0, i = 1, \ldots, r; g_j(x) = 0, j = 1, \ldots s\}.$$

Then, for each compact set $K \subset \mathbb{R}^n$, there exist $\tau > 0, a > 0$ such that

$$\operatorname{dist}(x, S) \leq \tau \left(\|[f(x)]_+\| + \|g(x)\|\right)^a,\ \forall x \in K,$$

where $f(x) = (f_1(x), \ldots, f_r(x))$, $g(x) = (g_1(x), \ldots, g_s(x))$.

We mention that in all the above results of error bound, the Hölder exponent is not unknown, even in the result of Luo and Luo [87] for polynomial function systems.

1.4.1 Local error bounds for polynomials

We are now interested in the estimation of exponents within error bounds. First, we present the result of Gwozdziewicz [58], in which a local quantitative error bound for a single real polynomial function with a isolated zero is provided.

For each $n, d \in \mathbb{N}$, we set

$$\kappa(n,d) = (d-1)^n + 1 \text{ and } R(n,d) = \begin{cases} 1 & \text{if } d = 1 \\ d(3d-3)^{n-1} & \text{if } d \geq 2. \end{cases}$$

Theorem 1.4.4. *[58] Let f be a polynomial function on $\mathbb{R}^n$ with degree d. Assume that $x = 0$ is an isolate zero of f, this means $f(0) = 0$ and there is $\delta > 0$ with $f(x) \neq 0$, for all $x \in B_\delta(0)\backslash\{0\}$. There exist positive constants τ, ε such that*

$$\|x\| \leq \tau |f(x)|^{\frac{1}{\kappa(n,d)}},$$

for all $\|x\| \leq \varepsilon$.

A similar result for polynomial function system was given by Kollár in [70].

Theorem 1.4.5. *[70] Let $f_1, \ldots, f_m$ be some polynomial functions on $\mathbb{R}^n$ whose degrees do not exceed d. Set $f(x) = \max_{i=1,\ldots,m} f_i(x)$ for x in $\mathbb{R}^n$. Assume that there is $\delta > 0$ such that $f(x) = 0$ and $f(x) \neq 0$, $\forall x \in B_\delta(0)\backslash\{0\}$. Then there exist τ, ε such that*

$$\|x\| \leq \tau |f(x)|^{\frac{1}{d^n \beta(n-1)}}, \text{ for all } x \text{ such that } \|x\| \leq \varepsilon,$$

where

$$\beta(n-1) = \binom{n-1}{[\frac{n-1}{2}]}.$$

Without the assumption of isolated zero point, Kurdyka and Spodzieja [74, Corollary 4] (see also [111]) obtained an error bound for a polynomial function with the Hölder exponent $a = R^{-1}(n,d)$.

To our knowledge, these are the first general results on error bounds with some estimations of the exponent. Some applications of these above results can be found in [27, 77, 78, 80, 102, 79].

In [80], Li, Mordukhovich and Pham, gave local error bounds for polynomial function systems in the nonconvex case, with exponents explicitly determined by the dimension of the underlying space and the degree of the involved polynomial functions. In this work, they obtained two results, one is based on Łojasiewicz gradient inequality, and the other result is proved with a technique similar to that of Theorem 1.4.3.

Theorem 1.4.6. *[80] . Let $f_1, \ldots, f_r$ and $g_1, \ldots, g_s$ be real polynomial functions on $\mathbb{R}^n$ with degree at most d, and let*

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

Then for each $\bar{x} \in S$ there exist $\tau > 0$, $\varepsilon > 0$ such that

$$\text{dist}(x,S) \leq \tau \left(\sum_{i=1}^{r} [f_i(x)]_+ + \sum_{j=1}^{s} |g_i(x)| \right)^{\frac{1}{R(n+r+s,d+1)}}, \text{ with } \|x - \bar{x}\| \leq \varepsilon. \tag{1.5}$$

Before beginning the proof of the latter theorem, let us recall a result of D'Acunto and Kurdyka [43], which established Łojasiewicz gradient inequality for polynomial function.

Theorem 1.4.7. *[43] Let f be a polynomial function with degree d, suppose that $f(0) = 0$. There exists $c > 0$, $\varepsilon > 0$ such that*

$$\|\nabla f(x)\| \geq \tau |f(x)|^{1 - \frac{1}{R(n,d)}}, \text{ with } \|x\| \leq \varepsilon.$$

Now, we apply this result to establish the Łojasiewicz gradient inequality for maximum of finitely many polynomial functions.

Lemma 1.4.8. *Let $f(x) = \max_{i=1,\dots,r} f_i(x)$ where f_i are polynomial functions on $\mathbb{R}^n$ whose degrees do not exceed d, and $\bar{x} \in \mathbb{R}^n$ with $f(\bar{x}) = 0$. Then, exist $c > 0$, $\varepsilon > 0$ such that*

$$dist\,(0, \partial f(x)) \geq \tau |f(x)|^{1 - \frac{1}{R(n+r-1,d+1)}}, \text{ with } \|x - \bar{x}\| \leq \varepsilon.$$

Proof. Without loss of generality, suppose that $f_i(\bar{x}) = 0$, $i = 1, \dots, r$. For each subset $I = \{i_1, \dots, i_q\} \subset \{1, \dots, r\}$, we define the polynomial function $F_I \colon \mathbb{R}^{n+q-1} \to \mathbb{R}$ as following

$$F_I(x, \lambda) = \begin{cases} \sum\limits_{j=1}^{q-1} \lambda_j f_{i_j}(x) + \left(1 - \sum\limits_{j=1}^{q-1} \lambda_j\right) f_{i_q}(x) & \text{if } q \geq 2 \\ f_{i_1}(x) & \text{if } q = 1, \end{cases}$$

where $\lambda = (\lambda_1, \dots, \lambda_{q-1}) \in \mathbb{R}^{q-1}$. It is clear that F_I has degree at most $d + 1$ and $F(\bar{x}, \lambda) = 0, \forall \lambda \in \mathbb{R}^{q-1}$. Set

$$P = \left\{ \lambda \in \mathbb{R}^{q-1} | \lambda_j \geq 0, \sum_{j=1}^{q-1} \lambda_j \leq 1 \right\}.$$

P is a compact set. For each $\bar{\lambda} \in P$, if $\nabla F_I(\bar{x}, \bar{\lambda}) = 0$, then thanks to Theorem 1.4.7, there exit $\varepsilon_I > 0$, $\tau_I > 0$ such that

$$\|\nabla F(x, \lambda)\| \geq \tau_I |F_I(x, \lambda)|^{1 - \frac{1}{R(n+q-1,d+1)}}, \text{ with } \|\lambda - \bar{\lambda}\| \leq \varepsilon_I, \|x - \bar{x}\| \leq \varepsilon_I. \tag{1.6}$$

In the other case, when $\nabla F_I(\bar{x}, \bar{\lambda}) \neq 0$ then (1.6) immediately holds. By the compactness of P, the inequality (1.6) holds for all $\lambda \in P$. Set

$$\tau = \min \{\tau_I | I \subset \{i, \dots, r\}, I \neq \emptyset\} > 0 \text{ and } \varepsilon = \min \{\varepsilon_I | I \subset \{i, \dots, r\}, I \neq \emptyset\} > 0.$$

Take an arbitrary point $x \in \mathbb{R}^n$ such that $\|x - \bar{x}\| \leq \varepsilon$ and $I(x) = \{i | f_i(x) = f(x)\}$, then there exist $\lambda_i \geq 0, i \in I(x)$ and $\sum\limits_{i \in I(x)} \lambda_i = 1$ such that

$$dist\,(0, \partial f(x)) = \| \sum_{i \in I(x)} \lambda_i \nabla f_i(x) \|.$$

On the other hand, for $i \in I(x)$, we have

$$F_{I(x)}(x, \lambda) = \sum_{i \in I(x)} \lambda_i f_i(x) = f(x)$$

and

$$\|\nabla F_{I(x)}(x, \lambda)\| = \| \sum_{i \in I(x)} \lambda_i \nabla f_i(x) \| = dist\,(0, \partial f(x)).$$

By combining the above inequalities and (1.6), we have the conclusion.

□

We now provide the proof of Theorem 1.4.6

Proof of Theorem 1.4.6 We consider the proof for $\bar{x} \in \text{bd}(S)$. For any $e = (e_i)_{i=1,\ldots,s} \in \{-1,1\}^s$, define the function

$$f_e(x) = \max\{0, f_i(x), \ldots, f_r(x), e_1 g_1(x), \ldots, e_s g_s(x)\}, \forall x \in \mathbb{R}^n.$$

One can see that f_e is the maximum of $r+s+1$ polynomial function with degree not exceed d, and $f_e(\bar{x}) = 0$. Applying Lemma 1.4.8, one obtains $\tau_e > 0$ and $\varepsilon_e > 0$ such that

$$dist\,(0, \partial f_e(x)) \geq \tau_e |f_e(x)|^{1-\frac{1}{R(n+r+s,d+1)}}, \forall \|x - \bar{x}\| \leq \varepsilon_e.$$

Set

$$\tau = \min\{\tau_e | e \in \{-1,1\}^s\} > 0, \quad \varepsilon = \{\varepsilon_e | e \in \{-1,1\}^s\} > 0,$$

and

$$f(x) = \max\{0, f_i(x), \ldots, f_r(x), g_1(x), \ldots, g_s(x), -g_1(x), \ldots, -g_s(x)\}$$

For any x with $\|x - \bar{x}\| \leq \varepsilon$ and $f(x) > 0$, then we can find $e \in \{-1,1\}^s$ such that $f(x) = f_e(x)$ and $dist\,(0, \partial f(x)) = dist\,(0, \partial f_e(x))$. Therefore,

$$\text{dist}\,(0, \partial f(x)) \geq \tau |f(x)|^{1-\frac{1}{R(n+r+s,d+1)}}, \forall \|x - \bar{x}\| \leq \varepsilon.$$

By applying Corollary 1.3.8 with $\varphi(s) = s^{\frac{1}{R(n+r+s,d+1)}}$, $\forall s > 0$, we obtain the conclusion. □

By using the same technique as in[88, Theorem 2.1], [87, Theorem 2.2], one can obtain an error bound whose exponent is different from Theorem 1.4.6.

Theorem 1.4.9. *[80] With the assumptions of Theorem 1.4.6, we have the following local error bound.*

$$\text{dist}(x, S) \leq \tau \left(\sum_{i=1}^{r} [f_i(x)]_+ + \sum_{j=1}^{s} |g_i(x)| \right)^{\frac{2}{R(n+r,2d)}}, \text{ with } \|x - \bar{x}\| \leq \varepsilon.$$

1.4.2 Global error bounds for polynomials

1.4.2.1 Nonconvex case

We begin this subsection by recalling the result of Luo and Sturm [89]. The authors established the global error bound for the zero set of a quadratic function.

Theorem 1.4.10. *[89] Let $f : \mathbb{R}^n \to \mathbb{R}$ be the quadratic function. There exists a constant $\tau > 0$ such that*

$$\text{dist}(x, [f = 0]) \leq \tau(|f(x)| + |f(x)|^{\frac{1}{2}}), \forall x \in \mathbb{R}^n.$$

This result is recovered by the works of [101, Corollary 5], [50, Corollary 2]. Remark that this theorem does not until hold for an arbitrary polynomial,

Example 4. *Let $f(x,y) = (xy-1)^2 + (x-1)^2$, $\forall (x,y) \in \mathbb{R}^2$.*

One has $[f \leq 0] = \{(1,1)\}$. Consider the sequence $(x_k = \frac{1}{k}, y_k = k)_{k \in \mathbb{N}}$, it is easy to check that

$$0 < f(x_k, y_k) = \left(1 - \frac{1}{k}\right)^2 < 1, \forall k \in \mathbb{N} \text{ and } d((x_k, y_k), [f \leq 0]) \to +\infty (k \to +\infty),$$

therefore, f does not possess Hölder global error bound.

However, when f is a polynomial convex, this result was proved by Yang [124], and we present it in Theorem 1.4.15.

Let us now present the characterization of global error bound for semi–algebraic, which is proved by Ha [59].

Suppose that $f\colon \mathbb{R}^n \to (-\infty, +\infty]$ has a Hölder global error bound,

$$\operatorname{dist}(x, [f \leq 0]) \leq \tau \left([f(x)]_+^a + [f(x)]_+^b \right), \quad \forall x \in \mathbb{R}^n. \tag{1.7}$$

We observe easily that for any sequence $(x_k)_{k\in\mathbb{N}} \subset \mathbb{R}^n$, two following assertions hold

(i) If $f(x_k) \to 0$, then $\operatorname{dist}(x_k, [f \leq 0]) \to 0$.

(ii) If $\operatorname{dist}(x_k, [f \leq 0]) \to +\infty$, then $f(x_k) \to +\infty$.

Conversely, in [59], Ha proved that, for a polynomial function which satisfies two above conditions, then it possesses Hölder global error bound. This result was extended for the class of continuous semi-algebraic functions, see [50, Theorem 2]. The definition of the semi-algebraic function is well-known, we can see the one in [50, Definition 1].

Theorem 1.4.11 (Characterization of global error bound for semi-algebraic)**.** *[59, 50] Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be a continuous semi-algebraic function. The following statements are equivalent:*

1. *For any sequence $(x_k)_{k\in\mathbb{N}} \in \mathbb{R}^n \setminus [f \leq 0]$ and $\|x_k\| \to +\infty$, we have:*

   (i) *If $f(x_k) \to 0$ then $\operatorname{dist}(x_k, [f \leq 0]) \to 0$.*

   (ii) *If $\operatorname{dist}(x_k, [f \leq 0]) \to +\infty$ then $f(x_k) \to +\infty$.*

2. *There exist $\tau > 0$ and $a, b > 0$ such that*

$$\operatorname{dist}(x, [f \leq 0]) \leq \tau \left([f(x)]_+^a + [f(x)]_+^b \right), \forall x \in \mathbb{R}^n.$$

Proof. (2) $\Rightarrow$ (1) is obvious, we now prove the implication (1) $\Rightarrow$ (2). The proof is divided into two parts. Using (i), we shall prove that an error bound holds on the neighborhood of $[f \leq 0]$, while by using (ii) we provide a bound for large $\operatorname{dist}(x, [f \leq 0])$.

Assume (i) holds. Let us prove that there exist $\tau_1 > 0, a > 0$ and $r > 0$ such that

$$\operatorname{dist}(x, [f \leq 0]) \leq \tau_1 [f(x)]_+^a, \forall x \in [f \leq r].$$

For $t \in \mathbb{R}$, put $\varphi(t) = \sup\{\operatorname{dist}(x, [f \leq 0]) : f(x) = t\}$. It is a semi-algebraic function. Thanks to (i), there exists $r > 0$ such that $\varphi(t) < \infty$ for all $t \in [0, r]$. We can choose r sufficiently small such that $\varphi(t)$ is continuous and $\varphi(t) \neq 0$ on $(0, r]$. By using Puiseux Lemma:

$$\varphi(t) = \tau t^a + 0(t^a), \ (t \to 0).$$

From the assumption (i), it can be seen that $\tau > 0$, $a > 0$. So there exist $r > 0$ and $\tau_1 > 0$ such that $\varphi(t) \leq \tau_1 t^a$, for all $t \in [f \leq r]$. It means that

$$\operatorname{dist}(x, [f \leq 0]) \leq \tau_1 [f(x)]_+^a, \forall x \in [f \leq r].$$

Using (ii), let us prove that there exist $\tau_2 > 0, b > 0$ and $\delta > 0$ such that

$$\operatorname{dist}(x, [f \leq 0]) \leq \tau_2 [f(x)]_+^b, \forall x \in [\delta < f].$$

This conclusion is clear when f is bounded from above. We assume thus that $\sup_{\mathbb{R}^n} f = \sup_{\mathbb{R}^n} \varphi = +\infty$. It appears that $\varphi(t) > 0$ when t is sufficiently large, so there exist $\tau > 0$ and $b > 0$ such that

$$\varphi(t) = \tau t^b + 0(t^b).$$

This implies that there is $c\tau_2 > 0$, $R > 0$ such that

$$\operatorname{dist}(x, [f \leq 0]) \leq \tau_2 [f(x)]_+^b, \forall x \in [R < f].$$

It is easily seen that (ii) implies the existence of $M > 0$ such that $\operatorname{dist}(x, S) < M$, for all $x \in [r < f < R]$. It gives $\operatorname{dist}(x, [f \leq 0]) \leq \frac{M}{r^\alpha} f(x)^\alpha$. Combining with such inequality on the domain $[f \leq r]$ and $[f \geq R]$, we have the conclusion. □

The implication (2) ⇒ (1) in the latter theorem explains why do we need two exponents $[f(x)]_+^a$ and $[f(x)]_+^b$ in the global error bound (1.7). One is ensures that the inequality (1.7) holds when $\operatorname{dist}(x, [f \leq 0]) \to 0$, and the other keeps such inequality holds when $\operatorname{dist}(x, [f \leq 0]) \to +\infty$. Generally, the exponents are different.

Example 5. *[60] Let $f(x, y) = x^2 + y^4$, $\forall (x, y) \in \mathbb{R}^2$.*

It can be seen that $[f \leq 0] = \{(0, 0)\}$, and

$$\operatorname{dist}((x, y), [f \leq 0]) \leq f^{\frac{1}{4}}(x, y) + f^{\frac{1}{2}}(x, y),\ \forall (x, y) \in \mathbb{R}^2.$$

On the other hands, by taking two sequences $(x_k^1 = k, y_n^1 = 0)_{k \in \mathbb{N}}$ and $(x_k^2 = 0, y_k^2 = 1/k)_{k \in \mathbb{N}}$, this follows that there does not exist $\alpha \in \mathbb{R}$ such that

$$\operatorname{dist}((x, y), [f \leq 0]) \leq \tau [f(x, y)]_+^\alpha,\ \forall (x, y) \in \mathbb{R}^2.$$

By using Theorem 1.4.11, Ha [59] provided a global error bound for polynomial function under a Palais–Smale condition. After that, his result was improved in [50] for continuous semi-algebraic functions.

We recall that, f is said to possess the Palais-Smale condition (PS) at r_0 if any sequence $(x_k)_{k \in \mathbb{N}}$, for which $f(x_k) \to r_0$ and $\operatorname{dist}(0, \partial f(x_k)) \to 0$, then $(x_k)_{k \in \mathbb{N}}$ possesses a converging subsequence.

Theorem 1.4.12. *[59, 50] Let $f \colon \mathbb{R}^n \to \mathbb{R}$ be a continuous semi-algebraic function. Suppose that f satisfies the Palais-Smale condition at each $r > 0$, then there exist constants $\tau > 0$ and $a, b > 0$ such that*

$$\operatorname{dist}(x, [f \leq 0]) \leq \tau \left([f(x)]_+^a + [f(x)]_+^b \right), \forall x \in \mathbb{R}^n.$$

Proof. It is enough to show that f satisfies the two conditions (i) and (ii) in Theorem 1.4.11. First we establish (i). By contradiction, we assume that there exists a sequence $(x_k)_{k \in \mathbb{N}}$ and a constant $\delta > 0$ such that:

$$\|x_k\| \to \infty, f(x_k) \to 0 \text{ and } \operatorname{dist}(x_k, [f \leq 0]) > \delta.$$

Put $X = \{x | f(x) \geq 0\}$, then X is a complete metric space. Applying Ekeland's principle (see [53]), there is a sequence $(y_k)_{k \in \mathbb{N}} \subset X$ such that

$$f(y_k) \leq f(x_k) = \varepsilon_k$$

$$\operatorname{dist}(x_k, y_k) \leq \sqrt{\varepsilon_k}$$

$$f(y_k) \leq f(x) + \sqrt{\varepsilon_k}\,\mathrm{dist}(x, y_k), \forall x \in X.$$

It is clear that $f(y_k) \to 0$ and $\|y_k\| \to +\infty$. We can suppose that $\mathrm{dist}(y_k, [f \leq 0]) \geq \frac{\delta}{2}$, therefore $\forall t \in (0, \frac{\delta}{2})$ and for all $u \in \mathbb{R}^n, \|u\| = 1$ we obtain

$$\frac{f(y_k + tu) - f(y_k)}{t} \geq -\sqrt{\varepsilon_k}.$$

Thus $|\nabla f|(y_k) \leq \sqrt{\varepsilon_k}$. On the other hands, $\|\partial f(y_k)\| \leq |\nabla f|(y_k)$, (see [10, Remark 6.1]), therefore $\partial f(y_k) \to 0$, which is in contradiction with Palais-Smale's condition.

Now, we will prove that f satisfies the condition (ii) of Theorem 1.4.11. By contradiction, suppose that there exists a sequence $(x_k)_{k\in\mathbb{N}} \subset \mathbb{R}^n$ such that:

$$\|x_k\| \to \infty,\ \mathrm{dist}(x_k, [f \leq 0]) \to +\infty \text{ and } f(x_k) \to t \in \mathbb{R}.$$

Set $X = \{x | f(x) \geq 0\}$, X is a complete metric space. Applying Ekeland's principle, there is a sequence $(y_k)_{k\in\mathbb{N}} \subset X$ such that

$$f(y_k) \leq f(x_k) = t_k$$

$$\mathrm{dist}(x_k, y_k) \leq \frac{\mathrm{dist}(x_k, [f \leq 0])}{2}$$

$$f(y_k) \leq f(x) + \frac{2f(x_k)}{\mathrm{dist}(x_k, [f \leq 0])}\,\mathrm{dist}(x, y_k), \forall x \in X$$

Therefore, without loss of generality we can assume that the sequence $f(y_k)$ is convergent, $\|y_k\| \to \infty$ and $\mathrm{dist}(y_k, [f \leq 0]) \to +\infty$, therefore,

$$\|\nabla f(y_k)\| \leq |\nabla f|(y_k) \leq \frac{2f(x_k)}{\mathrm{dist}(x_k, [f \leq 0])} \to 0,$$

contradicting to Palais-Smale's condition. □

1.4.2.2 Convex case

We begin this subsection by giving a result of Facchinei, Pang [54], they assert that a lower semicontinuous convex function, a Hölder-type error bound on a level set can be extended to a global error bound.

Theorem 1.4.13. *[54] Let f be a lower semicontinuous convex function on $\mathbb{R}^n$ with $[f \leq 0]$ nonempty. Suppose that there exist $\delta > 0$ and $\tau > 0$, $\theta > 0$ such that*

$$\mathrm{dist}(x, [f \leq 0]) \leq \tau\left([f(x)]_+ + [f(x)]_+^\theta\right), \quad \forall x \in [f \leq \delta].$$

There exists $\tau' > 0$ such that

$$\mathrm{dist}(x, [f \leq 0]) \leq \tau'\left([f(x)]_+ + [f(x)]_+^\theta\right), \quad \forall x \in \mathbb{R}^n.$$

When we take $\theta = 1$, this means that for a convex function, a Lipschitz error bound on the level set can be extended to a global error bound.

Proof. Let $x \in \mathbb{R}^n$ such that $f(x) > \delta$ and $p = P_{[f\leq 0]}x$. It is clear that $f(p) = 0$. For any $\lambda \in (0,1)$, we denote $x_\lambda = \lambda x + (1-\lambda)p$. It can be seen that $p = P_{[f\leq 0]}x_\lambda$ and $\operatorname{dist}(x_\lambda, [f \leq 0]) = \lambda \operatorname{dist}(x, [f \leq 0])$. By convexity, we get

$$f(x_\lambda) \leq \lambda f(x) + (1-\lambda)f(p) = \lambda f(x).$$

We deduce that

$$\operatorname{dist}(x, [f \leq 0]) \leq \frac{\operatorname{dist}(x_\lambda, [f \leq 0])}{f(x_\lambda)} f(x).$$

On the other hand, by choosing $\lambda = \frac{\delta}{2f(x)}$, we get

$$f(x_\lambda) \leq \lambda f(x) = \frac{\delta}{2} < \delta.$$

Therefore, thanks to the assumption on error bounds, we obtain

$$\operatorname{dist}(x_\lambda, [f \leq 0]) \leq \tau \left(f(x_\lambda) + f^\theta(x_\lambda) \right).$$

It follows that

$$\frac{\operatorname{dist}(x_\lambda, [f \leq 0])}{f(x_\lambda)} < \tau \left(1 + f^{\theta-1}(x_\lambda)\right) < c \left(1 + \left(\frac{\delta}{2}\right)^{\theta-1}\right).$$

Combining the above inequalities, we get

$$\operatorname{dist}(x, [f \leq 0]) \leq \tau \left(1 + \left(\frac{\delta}{2}\right)^{\theta-1}\right) f(x).$$

This means that

$$\operatorname{dist}(x, [f \leq 0]) \leq \tau \left(1 + \left(\frac{\delta}{2}\right)^{\theta-1}\right) \left(f(x) + f^\theta(x)\right), \forall x \in \mathbb{R}^n.$$

□

Combining this result with Theorem 1.4.2, we immediately obtain a result similar to [24, Theorem 3] and [49, Theorem 6].

Theorem 1.4.14. *Let $f_i \colon \mathbb{R}^n \to \mathbb{R}$, $(i = 1, \dots m)$ be continuous, convex and subanalytic functions. Assume that, the set*

$$S = \{x \in \mathbb{R}^n | f_i(x) \leq 0,\ i = 1, \dots m\}$$

is nonempty, compact. Then, there exist τ, $\theta > 0$ such that

$$\operatorname{dist}(x, S) \leq \tau \left([f(x)]_+ + [f(x)]_+^\theta \right),\ \forall x \in \mathbb{R}^n,$$

where $f(x) = \sum_{i=1}^m [f_i(x)]_+$.

We remark that if f_i is coercive then for all $r \in \mathbb{R}$, the set $[f_i \leq r]$ is compact.
We recall now the definition of piecewise convex polynomial functions.

Definition 4. *[81, 78] A continuous function f on $\mathbb{R}^n$ is said to be a piecewise convex polynomial function if there exist finitely many polyhedra $P_1, \ldots, P_k$ with $\cup_{j=1}^k P_j = \mathbb{R}^n$ such that the restriction of f on each P_j, denoted by f_i, is a convex polynomial function. The degree of f, denoted by* $\deg(f)$*, is defined as the maximum of* $\deg(f_j)$.

In [81], Li studied error bounds for a convex piecewise quadratic function. More precisely, let f be a convex piecewise quadratic function. Then, there exists $\tau > 0$ such that

$$\operatorname{dist}(x, [f \leq 0]) \leq \tau \left([f(x)]_+ + \sqrt{[f(x)]_+} \right), \forall x \in \mathbb{R}^n. \tag{1.8}$$

By using Theorem 1.4.5 and Theorem 1.4.13, Li [77] showed that, for a convex polynomial function f on $\mathbb{R}^n$ with degree d, there exists $\tau > 0$ such that

$$\operatorname{dist}(x, [f \leq 0]) \leq \tau \left([f(x)]_+ + [f(x)]_+^{\frac{1}{\kappa(n,d)}} \right), \forall x \in \mathbb{R}^n. \tag{1.9}$$

This result is further improved by Yang [124].

Theorem 1.4.15. *[124] Let f be a polynomial convex with degree d. There exists $\tau > 0$ such that*

$$\operatorname{dist}(x, [f \leq 0]) \leq \tau \left([f(x)]_+ + [f(x)]_+^{\frac{1}{d}} \right), \forall x \in \mathbb{R}^n.$$

The two above results (1.8), (1.9) have been extended by Li ([78]), for general convex piecewise polynomial function.

Theorem 1.4.16. *[78] Let f be a piecewise convex polynomial function on $\mathbb{R}^n$ with degree d. Suppose that one of the following two conditions holds:*

(i) If $\operatorname{dist}(x, [f \leq 0]) \to +\infty$ *then* $f(x) \to +\infty$.

(ii) f is convex.

There exists $c > 0$ such that

$$\operatorname{dist}(x, [f \leq 0]) \leq c \left([f(x)]_+ + [f(x)]_+^{\frac{1}{\kappa(n,d)}} \right), \forall x \in \mathbb{R}^n.$$

Let us now present a global error bound for convex polynomial function systems. In [87], under the Slater condition, Luo and Luo proved that a global Lipschitzian error bound holds for convex quadratic systems. After that, without the Slater condition, Pang and Wang in [121], showed that any systems of convex quadratic has a global error bound.

Theorem 1.4.17. *[121] Let $f_1, f_2 \ldots, f_m$ be convex quadratic functions. Assume that*

$$S = \{x \in \mathbb{R}^n | f_i(x) \leq 0, i = 1, \ldots m\}$$

is not empty, then there exists a positive integer $\operatorname{dist} \leq n + 1$ *and a scalar $c > 0$ such that*

$$\operatorname{dist}(x, S) \leq c \max \left(\|[f(x)]_+\|, \|[f(x)]_+\|^{\frac{1}{2^d}} \right), \forall x \in \mathbb{R}^n,$$

where $f(x) = (f_i(x))_{i=1\ldots,m}$, $\forall x \in \mathbb{R}^n$.

Furthermore, if S contains an interior point, then $d = 0$.

Similarly Theorem 1.3.14, the latter result is extended to the Banach space in [103, Theorem 7], with

$$f_i(x) = \frac{1}{2}\langle A_i x, x\rangle + \langle B_i, x\rangle + c_i,$$

where $A_i \colon X \times X \to \mathbb{R}$ is a symmetric continuous bilinear and semi-definite positive, $B_i \in X^*$ and $c_i \in \mathbb{R}$, for $i = 1, \ldots, m$.

Note that this result does not hold for a general convex polynomial function system, see Example 2. However, in some particular cases, the global error bound hold for such systems.

Theorem 1.4.18. *Let $f_1, \ldots, f_p$ be convex polynomial functions on $\mathbb{R}^n$ whose degrees are at most d. Let $f(x) = \max_{i=1,\ldots,m} f_i(x)$, $\forall x \in \mathbb{R}^n$. Then, the following statements are hold*

1. *[77] If $f_i(x) \geq 0, \forall x \in \mathbb{R}^n, i = 1, \ldots, m$ then there exists $\tau > 0$ such that*

$$\operatorname{dist}(x, [f \leq 0]) \leq \tau \left([f(x)]_+ + [f(x)]_+^{\frac{1}{\kappa(n,d)}} \right), \forall x \in \mathbb{R}^n.$$

2. *[102] If $S = \{x \in K | f(x) \leq 0\}$ is a nonempty compact set, where K is a convex polyhedral in $\mathbb{R}^n$. Then, there exists $\tau > 0$ such that*

$$\operatorname{dist}(x, S) \leq c \left([f(x)]_+ + [f(x)]_+^{\frac{1}{\kappa(n,2d)}} \right), \forall x \in K.$$

3. *[102]Let K is a convex polyhedral in $\mathbb{R}^n$ and $S = \{x \in K | f(x) \leq 0\}$ is nonempty. Assume that, for each $v \in K^\infty$: $\max_{i=1,\ldots,p} f_i^\infty(v) = 0 \Rightarrow f_i^\infty(v) = 0$ (see (1.4)). Then, there exists $\tau > 0$ such that*

$$\operatorname{dist}(x, S) \leq c \left([f(x)]_+ + [f(x)]_+^{\frac{1}{\kappa(n,2d)}} \right), \forall x \in K,$$

where K^∞ is recession cone of K, defined by

$$K^\infty = \{v \in \mathbb{R}^n | x + tv \in K, \forall t > 0, x \in K\}.$$

Chapter 2

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

Abstract This chapter 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.

2.1 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 [63], 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 [113], Mangasarian [93], and Auslender-Crouzeix [6]. However, the most striking discovery came years before in the pioneering works of Łojasiewicz [84, 85] 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 [88]). 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 [106] provides a comprehensive panorama of results obtained around this time. The works of Luo [87, 88, 89] and Dedieu [45] are also important milestones in the theory. The recent works [78, 80, 59, 79, 15]

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

$$\omega\big(f(x) - \min f\big) \geq \operatorname{dist}(x, \operatorname{argmin} f), \tag{2.1}$$

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 [84, 85], a modern reference being [42]. 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 [89, 78, 15, 59]), 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 (2.1)), 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., [92, 97, 16, 40, 108]). See also the new and interesting work [79] 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 [88, 92]) that are different from those we consider here.

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 [23]), asserts that for any smooth semi-algebraic function f there is a smooth concave function φ such that

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

for all x in some neighborhood of the set $\operatorname{argmin} f$. Its generalization to the nonsmooth case [21, 22] 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, 26, 56]. 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].

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

[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 2.6.

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 [20] and [18, 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 [68], 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 thresholding algorithm* (ISTA) applied to a least squares objective with ℓ^1 regularization [44] and we prove that its complexity is of the form $O(q^k)$ with $q \in (0, 1)$ (see [97] for a pioneering work in this direction and also [83] for further geometrical insights). This result contrasts with what was known on the subject [17, 51] 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 [21, 80, 79]. 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 2.6 by a dimension 2 counterexample presented in [23] for another purpose.

Our investigations also led us to consider the problem of KL inequalities for convex functions, a problem partly tackled in [23]. 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.*

2.2 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 $\min f = 0$.

2.2.1 Some convex analysis

We use the standard notation from [114] (see also [14, 109] and [96]). 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 6. 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.

2.2.2 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 2.2.1 (Brézis [28], Bruck [31])**.** *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 [28], except for part v), which was proved in [31]. The trajectory $t \mapsto \chi_x(t)$ is called a *subgradient curve.*

2.2.3 Kurdyka-Łojasiewicz inequality

In this subsection, we present the nonsmooth Kurdyka-Łojasiewicz inequality introduced in [21] (see also [22, 23], and the fundamental works [86, 73]). 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

$$\varphi'\left(f(x)\right) \|\partial^0 f(x)\| \geq 1 \tag{2.2}$$

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, (2.2) 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 [42] by Bochnak-Coste-Roy.

We recall the following theorem on the nonsmooth KL inequality (which follows the pioneering works of Łojasiewicz [86] and Kurdyka [73]). It is one of the cornerstones of the present research:

Theorem 2.2.2 (Bolte-Daniilidis-Lewis [21])**.** (*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 2.6.3.

2.2.4 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:

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

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

If f is convex lower semicontinuous, we can extend the error bound beyond $[0 \leq f \leq r_0]$ 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

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

It follows that

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

This entails that

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

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 [78]). Observe that it can be put under the form $\omega(f(x)) \geq \operatorname{dist}(x, S)$ by simply setting $\omega(s) = \frac{1}{\gamma_0}(s + s^{\frac{1}{p}})$. When combined with the Łojasiewicz error bound inequality [84, 85], the above remark implies immediately the following result:

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

Theorem 2.2.3 (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*

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

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

2.3 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 2.6.

As shown in Section 2.4, 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., [78] and references therein). The fact that these two notions are equivalent opens a wide range of possibilities when it comes to analyzing algorithm complexity.

2.3.1 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. [88, 87, 89, 78, 100, 59]). 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 [42]) shows:

Lemma 2.3.1. *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 2.6.3, 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 [23, 21, 41, 80, 79] and also shares some common techniques.

Theorem 2.3.2 (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 2.6.1 of Section 2.6, 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 2.6.1, $\chi_x(s)$ 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 2.3.3. (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 2.3.4. (a) Observe the slight dissymmetry between the conclusions of (i) and (ii) in Theorem 2.3.2 and Corollary 2.3.3: 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

$$c = \frac{1}{p}. \tag{2.3}$$

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.

2.3.2 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 2.3.2 or Corollary 2.3.3 to compute a desingularizing function together with a domain of desingularization; this technique appears also in [80] a paper which only came to our knowledge during the finalization of our article.

2.3.2.1 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] $P_1, \dots, P_k$, such that $f_i = f|_{P_i}$ is a polynomial for each $i = 1, \dots, k$. The degree of f is defined as $\deg(f) = \max\{\deg(f_i) : i = 1, \dots, k\}$. We have the following interesting result from Li [78, Corollary 3.6]:

Proposition 2.3.5 (Li [78])**.** *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*

$$f(x) - \min f \geq \gamma_r \operatorname{dist}\big(x, \operatorname{argmin} f\big)^{(\deg(f)-1)^n+1} \tag{2.4}$$

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

Combining Proposition 2.3.5 and Corollary 2.3.3, the above implies:

Corollary 2.3.6. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a piecewise polynomial convex function with* $\operatorname{argmin} f \neq \emptyset$*. Then f has the Łojasiewicz 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 2.3.5 and Corollary 2.3.6 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 (2.4). For this we shall apply a recent result from Beck-Shtern [15].

First let us recall some basic results on error bounds (see e.g., [63, 125]). In what follows, $\|M\|$ denotes the *spectral* or *operator* norm of a real matrix M.[5]

Definition 5 (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*

$$\operatorname{dist}(x, X \cap Y) \leq \nu\|Ex - e\|, \; \forall x \in X. \tag{2.5}$$

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

[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.

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 [125] and the references therein.

As suggested by Beck, we shall now apply a very useful result from [15] 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

$$\begin{aligned}
\min_{\mathbb{R}^n} f &= \min\left\{\frac{1}{2}\|Ax-b\|_2^2 + \mu\|x\|_1 \ : x \in \mathbb{R}^n\right\} \\
&= \min\left\{\frac{1}{2}\|Ax-b\|_2^2 + \mu y \ : (x,y) \in \mathbb{R}^n \times \mathbb{R},\ \|x\|_1 \leq R,\ y = \|x\|_1\right\} \\
&= \min\left\{\frac{1}{2}\|Ax-b\|_2^2 + \mu y \ : (x,y) \in \mathbb{R}^n \times \mathbb{R},\ \|x\|_1 - y \leq 0,\ y \leq R\right\} \\
&= \min\left\{\frac{1}{2}\|\tilde{A}\tilde{x}-\tilde{b}\|_2^2 + \langle\tilde{\mu},\tilde{x}\rangle \ :\ \tilde{x} = (x,y) \in \mathbb{R}^n \times \mathbb{R},\ M\tilde{x} \leq \tilde{R}\right\} \qquad (2.6)
\end{aligned}$$

where

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

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 [15, 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 5 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
(2.7)

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

where

$$\kappa_R = \nu^2 \left(2R\mu + 6\left(R\|A\| + \|b\|\right) R\|A\| + 2\left(R\|A\| + \|b\|\right)^2 + 2 \right). \tag{2.8}$$

By taking $y = \|x\|_1$, (2.7) 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 2.3.7. (Error bound and KL inequality for the least squares objective with ℓ^1 regularization) *Fix $R > \frac{\|b\|^2}{2\mu}$. Then,*

$$f(x) - f(x^*) \geq 2\gamma_R \operatorname{dist}^2(x, S) \text{ for all } x \text{ in } \mathbb{R}^n \text{ such that } \|x\|_1 \leq R, \tag{2.9}$$

where

$$\gamma_R = \frac{1}{4\nu^2\left(1 + \mu R + \left(R\|A\| + \|b\|\right)\left(4R\|A\| + \|b\|\right)\right)}. \tag{2.10}$$

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.

2.3.2.2 Distances to an intersection: convex feasibility

For $m \geq 2$, one considers closed convex subsets $C_1, \ldots, C_m$ of H whose intersection contains a nonempty open ball. This proposition is a quantitative version of [16, Corollary 3.1].

Proposition 2.3.8. *Suppose that there is $\bar{x} \in H$ and $R > 0$ such that*

$$B(\bar{x}, R) \subset \bigcap_{i=1}^{m} C_i. \tag{2.11}$$

Then,

$$\operatorname{dist}(x, \bigcap_{i=1}^{m} C_i) \leq \left(1 + \frac{2\|x - \bar{x}\|}{R}\right)^{m-1} \max\left\{\operatorname{dist}(x, C_i), i = 1, \cdots, m\right\},\ \forall x \in H. \tag{2.12}$$

Proof. We assume $m = 2$ in a first stage. Put $C = C_1 \cap C_2$, $d = 2\max\left\{\operatorname{dist}(x, C_1), \operatorname{dist}(x, C_2)\right\}$ 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

$$\operatorname{dist}(x, C) \leq \left(1 + \frac{2\|x - \bar{x}\|}{R}\right) \max\left\{\operatorname{dist}(x, C_1), \operatorname{dist}(x, C_2)\right\}. \tag{2.13}$$

For arbitrary $m \geq 2$, applying (2.13) for the two sets C_1 and $\bigcap_{i=2}^{m} C_i$, we obtain

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

Repeating the process $(m-1)$ times, we obtain (2.12). □

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

$$f(x) = \frac{1}{2}\sum_{i=1}^{m} \alpha_i \operatorname{dist}^2(x, C_i), \tag{2.14}$$

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 [37, 16]). We now provide an error bound for f under assumption (2.11).

It is clear that $C = \operatorname{argmin} f = \{x \in H : f(x) = 0\}$. Fix any $x_0 \in H$. From Proposition 2.3.8, 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 2.3.2, 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

$$M = \frac{1}{4}\left(1 + \frac{2\|x_0 - \bar{x}\|}{R}\right)^{2-2m} \min_{i=1,\dots,m} \alpha_i. \tag{2.15}$$

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 2.5. 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

$$M' = \frac{1}{8}\left(1 + \frac{2\|x_0 - \bar{x}\|}{R}\right)^{-2}. \tag{2.16}$$

2.4 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 $\min f = 0$.

2.4.1 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 [92]. 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 [92, 4, 56] (which we describe in detail below), many trust region methods [1], alternating methods [4, 26], and, in a much more subtle manner, sequential quadratic methods and a wealth of majorization-minimization methods [25, 107]. In Section 2.5, we essentially focus on the forward-backward method because of its simplicity and its efficiency. Clearly, many other examples could be worked out.

Remark 2.4.1 (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 7 (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, 26, 56]). 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

$$x_{k+1} \in \operatorname{argmin}\left\{ g(z) + \langle \nabla h(x_k), z - x_k\rangle + \frac{1}{2\lambda_k}\|z - x_k\|^2 : z \in H \right\} \tag{2.17}$$

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 (2.17) 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 2.2.1, the latter can be rewritten as

$$x_{k+1} = \operatorname{prox}_{\lambda_k g}\left(x_k - \lambda_k \nabla h(x_k)\right). \tag{2.18}$$

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 2.4.2. *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* (2.18) *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 [26, 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 2.4.1 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 2.4.3 below guarantees the strong convergence of every sequence generated by the forward-backward method.

Convergence of subgradient descent sequences follows readily from [4] and [26, 56]. Although this kind of result has now became standard, we provide a direct proof for estimating thoroughly the constants at stake.

Theorem 2.4.3. (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 $f(x_0) \leq r_0 < \bar{r}$. Then, x_k converges strongly to some $x^* \in \operatorname{argmin} f$ and*

$$\|x_k - x^*\| \leq \frac{b}{a}\varphi(f(x_k)) + \sqrt{\frac{f(x_{k-1})}{a}},\ \forall k \geq 1. \tag{2.19}$$

Proof. Using (**H1**), we deduce that the sequence $(f(x_k))_{k\in\mathbb{N}}$ is nonincreasing, thus $x_k \in [0 \leq f < \bar{r}]$. Denote by i_0 the first index $i_0 \geq 1$ 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 $k \leq i_0 - 1$ in (2.19). 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 $k \geq 1$. Combining (**H1**), (**H2**), and using the concavity of φ we obtain

$$\begin{aligned}
\varphi(f(x_k)) - \varphi(f(x_{k+1})) &\geq \varphi'(f(x_k))\,(f(x_k) - f(x_{k+1})) \\
&\geq \frac{a\|x_k - x_{k+1}\|^2}{b\|x_{k-1} - x_k\|} \\
&\geq \frac{a}{b}\frac{\left(2\|x_k - x_{k+1}\|\|x_k - x_{k-1}\| - \|x_{k-1} - x_k\|^2\right)}{\|x_k - x_{k-1}\|}, \forall k \geq 1. \\
&\geq \frac{a}{b}(2\|x_k - x_{k+1}\| - \|x_{k-1} - x_k\|), \forall k \geq 1. \qquad (2.20)
\end{aligned}$$

This implies

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

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 (2.20), we also infer

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

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

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

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 2.4.4. When f is twice continuously differentiable and *definable* (in particular, if it is semi-algebraic) it is proved in [18] 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.

2.4.2 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 2.6.4).

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

A one-dimensional worst-case proximal sequence. Set

$$\zeta = \frac{\sqrt{1 + 2\ell\, a\, b^{-2}} - 1}{\ell} > 0, \tag{2.21}$$

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

$$\alpha_{k+1} = \operatorname{argmin}\left\{\psi(u) + \frac{1}{2\zeta}(u - \alpha_k)^2 : u \geq 0\right\} \tag{2.22}$$

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

$$\alpha_{k+1} = (I + \zeta\psi')^{-1}(\alpha_k) = \operatorname{prox}_{\zeta\psi}(\alpha_k), \tag{2.23}$$

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 2.4.5 (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,*

$$f(x_k) \leq \psi(\alpha_k), \quad \forall k \geq 0, \tag{2.24}$$

$$\|x_k - x^*\| \leq \frac{b}{a}\alpha_k + \sqrt{\frac{\psi(\alpha_{k-1})}{a}}, \quad \forall k \geq 1. \tag{2.25}$$

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

$$\beta_k = (1 + s_k\psi')^{-1}(\beta_{k-1}). \tag{2.26}$$

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

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

[6] See (2.21) and (2.22).

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, [109, 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

$$s_k \geq \frac{\sqrt{1 + 2\ell\, a\, b^{-2}} - 1}{\ell} = \zeta. \tag{2.27}$$

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*

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

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 2.4.3. □

Remark 2.4.6 (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 2.2.4 or Proposition 2.6.4). 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 2.4.7 (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 2.4.8 (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 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 2.4.5:

Corollary 2.4.9. *The assumptions and the notation are those of Theorem 2.4.5, 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*

$$\sigma = \ell b^{-2}. \tag{2.28}$$

In that case the complexity estimates given in Theorem 2.4.5 take the form

$$f(x_k) \leq \frac{f(x_0)}{(1+2a\sigma)^k}, \quad \forall k \geq 0, \tag{2.29}$$

$$\|x_k - x^*\| \leq \left[1 + \frac{1}{a\sigma\sqrt{1+\frac{1}{2a\sigma}}}\right] \frac{\sqrt{\frac{1}{a}f(x_0)}}{(1+2a\sigma)^{\frac{k-1}{2}}}, \quad \forall k \geq 1. \tag{2.30}$$

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,

$$\alpha_k = \frac{\alpha_0}{(1+\ell\zeta)^k}, \ \forall k \geq 0. \tag{2.31}$$

[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.

From (2.24), we immediately deduce

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

Finally, since

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

we obtain (2.29). For (2.30), first observe that

$$\frac{b}{a}\alpha_k = \frac{b}{a}\frac{\alpha_0}{(1+\ell\zeta)^k} = \frac{b}{a\sqrt{\ell}}\frac{\sqrt{2f(x_0)}}{(1+\ell\zeta)^k} = \frac{b}{a\sqrt{\ell}}\frac{\sqrt{2f(x_0)}}{(1+2\ell ab^{-2})^{k/2}}, \tag{2.32}$$

while

$$\sqrt{\frac{\psi(\alpha_{k-1})}{a}} = \sqrt{\frac{\ell\alpha_{k-1}^2}{2a}} = \sqrt{\frac{\ell\alpha_0^2}{2a(1+\ell\zeta)^{2k-2}}} = \sqrt{\frac{1+2\ell ab^{-2}}{2a}}\frac{\sqrt{2f(x_0)}}{(1+2\ell ab^{-2})^{k/2}}. \tag{2.33}$$

In view of (2.25), by adding (2.32) and (2.33) we obtain:

$$\|x_k - x^*\| \leq \left[\frac{b}{a\sqrt{\ell}} + \sqrt{\frac{1}{2a}+\frac{\ell}{b^2}}\right]\frac{\sqrt{2f(x_0)}}{(1+2\ell ab^{-2})^{k/2}}, \quad \forall k \geq 1. \tag{2.34}$$

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 (2.34) to obtain the result.

□

Remark 2.4.10 (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 2.5 below).

2.5 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 2.4.9, so let us keep in mind the notation introduced in Section 2.4, especially the constants a, b and ℓ.

2.5.1 Convex feasibility problems with regular intersection

Let $\{C_i\}_{i\in\{1,\dots,m\}}$ 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 $f=\frac{1}{2}\sum\limits_{i=1}^{m}\alpha_i\,\mathrm{dist}^2(\cdot,C_i)$, studied in Subsection 2.3.2.2, 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

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

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, [109]). 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 (2.15). Using Theorem 2.4.9 with $a=\frac{1}{2}$, $b=2$ and $\ell=M$, we obtain:

Theorem 2.5.1 (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*

$$\begin{aligned}
f(x_k)\ &\leq\ \frac{f(x_0)}{\left(1+\frac{M}{4}\right)^k},\quad \forall k\geq 0,\\
\|x_k-x^*\|&\leq\left[1+\frac{8}{M\sqrt{1+\frac{4}{M}}}\right]\frac{\sqrt{2f(x_0)}}{\left(1+\frac{M}{4}\right)^{\frac{k-1}{2}}},\quad \forall k\geq 1,
\end{aligned}$$

where M is given by (2.15)*.*

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 (2.16) (Subsection 2.3.2.2). 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 [13]). 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 (2.16). Using Corollary 2.4.9 with $a = \frac{1}{2}$, $b = 2$ and $\ell = M'$, we obtain:

Theorem 2.5.2 (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 (2.16)*.*

2.5.2 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, 56]. 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 2.3.5 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.

[8] A very interesting result from Baillon-Combettes-Cominetti [12] establishes that for more than two sets there are no potential functions corresponding to the alternating projection method.

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

We refer for instance to [33] 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 $L = \|A^TA\|$. 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* [44][9]:

$$x_{k+1} = \operatorname{prox}_{\lambda_k\mu\|\cdot\|_1}\big(x_k - \lambda_k(A^TAx_k - A^Td)\big) \quad \text{for } k \geq 0. \tag{ISTA}$$

Here, $\operatorname{prox}_{\lambda_k\mu\|\cdot\|_1}$ is an easily computable piecewise linear object known as the *soft thresholding* operator (see, for instance, [39]). This method has been applied widely in many contexts and is known to have a complexity $O\big(\frac{1}{k}\big)$. We intend to prove here that this bound can be surprisingly "improved" by our techniques.

First, recall that, according to Proposition 2.4.2, 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

$$a = \frac{1}{\lambda^+} - \frac{L}{2} \quad \text{and} \quad b = \frac{1}{\lambda^-} + L, \tag{2.35}$$

respectively.

Set $R = \max\Big(\dfrac{f(x_0)}{\mu}, 1 + \dfrac{\|d\|^2}{2\mu}\Big)$. 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 2.3.7 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 (2.10).

Remark 2.5.3 (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

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

Combining the above developments with Corollary 2.4.8, we obtain the following surprising result:

[9] Connection between ISTA and the forward-backward splitting method is due to Combettes-Wajs [39]

[10] Recall that $r_0 = f(x_0)$.

Theorem 2.5.4 (Complexity bounds for ISTA)**.** *The sequence* $(x_k)_{k\in\mathbb{N}}$ *generated by* ISTA *converges to a minimizer* x^* *of* f*, and satisfies*

$$f(x_k) - \min f \;\leq\; \frac{f(x_0) - \min f}{q^k}, \qquad \forall k \geq 0, \tag{2.36}$$

$$\|x_k - x^*\| \;\leq\; C\frac{\sqrt{f(x_0) - \min f}}{q^{\frac{k-1}{2}}} \qquad \forall k \geq 1, \tag{2.37}$$

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 2.5.5 (Complexity and convergence rates for ISTA)**.** (a) While it was known that ISTA has a linear asymptotic convergence rate, see [83] in which a transparent explanation is provided, best known *complexity bounds* were of the type $O(\frac{1}{k})$, see [17, 51]. Much like in the spirit of [83], 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 2.3.2.1 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 (2.36)): 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 2.5.3, the value q appearing in the complexity bounds given by Theorem 2.5.4 becomes

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

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

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

2.6 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.

2.6.1 KL inequality and length of subgradient curves

This subsection essentially recalls a characterization result from [23] on the equivalence between the KL inequality and the existence of a uniform bound for the length of subgradient

[11] Bad conditioning are produced by flat objective functions yielding thus small constants γ_R.

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 [23]. 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 $0 \le t < s$, 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 2.6.1 (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

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

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

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

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

Proof. Take $x \in B(\bar{x},\rho) \cap [0 < f \le r_0]$ and $0 \le t < s$. 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 2.2.1) and $-\dot{\chi}_x(\tau) \in \partial f(\chi_x(\tau))$ for almost every $\tau > 0$, it follows that

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

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

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

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

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

As $h \to 0$, we obtain

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

and so

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

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

2.6.2 A counterexample: error bounds do not imply KL

In [23, 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, [114]). 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

$$\psi(0) = 0, \tag{2.38}$$

$$\psi(s) > 0 \text{ for } s > 0, \tag{2.39}$$

$$f(x) \geq \psi(\operatorname{dist}(x,S)) \text{ for all } x \in [r < f] \tag{2.40}$$

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 (2.38), (2.39), (2.40). 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 [14]). Whence,

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

We have proved the following:

Theorem 2.6.2 (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 2.6.3 (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.

2.6.3 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 2.4.5. The ideas behind the proof are inspired by [23].

Proposition 2.6.4 (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 [28, 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 [23]. For an insight into the notion of *definability* of functions, a prominent example being semi-algebraicity, one is referred to [19]. Recall that coercivity of a proper lower-semicontinuous convex function defined on a finite dimensional space is equivalent to the fact that $\operatorname{argmin} f$ is nonempty and compact.

Theorem 2.6.5 (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* $\operatorname{argmin} f$ *is nonempty and compact. Then, f has the KL property on $\mathbb{R}^n$.*

Proof. Take $r_0 > 0$ and use [22] 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 $(0, +\infty)$. □

Remark 2.6.6. (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 2.4.5, 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.

Chapter 3

Extragradient method in optimization: Convergence and complexity

Abstract We consider the extragradient method to minimize the sum of two functions, the first one being smooth and the second being convex. Under Kurdyka-Lojasiewicz assumption, we prove that the sequence produced by the extragradient method converges to a critical point of the problem and has finite length. The analysis is extended to the case when both functions are convex. We provide a $1/k$ convergence rate which is classical for gradient methods. Furthermore, we show that the recent *small-prox* complexity result can be applied to this method. Considering the extragradient method is the occasion to describe exact line search for proximal decomposition methods. We provide details for the implementation of this scheme for the ℓ^1 regularized least squares problem and give numerical results which suggest that combining nonaccelerated methods with exact line search can be a competitive choice.

3.1 Introduction

We introduce a new optimization methods for approximating a global minimum of composite objective function F

$$\textbf{(P)} \qquad \min_{x\in\mathbb{R}^n}\{F(x) = f(x) + g(x)\},$$

where f is smooth and g is convex lower semicontinuous. This class of problems is rich enough to encompass many smooth/nonsmooth, convex/nonconvex optimization problems considered in practice. Applications can be found in various fields throughout science and engineering, including signal/image processing [38] and machine learning [116]. Succesful algorithms for these types of problems include for example FISTA method [17] and forward-backward splitting method [39]. The goal of this paper is to investigate to which extent extragradient method can be used to tackle similar problems.

The extragradient method was initially proposed by Korpelevich [71]. It has become a classical method for solving variational inequality problems, finding $\bar{x} \in S$ such that

$$\langle H(x), x - \bar{x}\rangle \geq 0,\ \forall x \in S,$$

where S is a nonempty, closed and convex subset of $\mathbb{R}^n$, $H : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ is a monotone mapping, and $\langle ., . \rangle$ denotes the Euclidean inner product in $\mathbb{R}^n$. The extragradient method, generates a sequence of estimates based on the following recursion, which requires two orthogonal projections onto S at each iteration,

$$\begin{cases} y_k = P_S(x_k - \alpha H(x_k)) \\ x_{k+1} = P_S(x_k - \alpha H(y_k)). \end{cases}$$

After Korpelevich's work, a number of authors extended the extragradient method for variational inequality problems (for example, see [35], [95]). In the context of convex constrained optimization, [92] considered the performances of the extragradient method under error bounds assumptions. In this setting, Luo and Tseng have described asymptotic linear convergence of the extragradient method applied to constrained problems of the form,

$$\min_{x \in C} f(x),$$

where C is a convex set of $\mathbb{R}^n$ and $f(x)$ is a smooth convex function on $\mathbb{R}^n$. To our knowledge, this is the only attempt to analyse the method specifically in an optimization setting.

A distinguished feature of the extragradient method is the use of an additional projected gradient step which can be seen as a guide during the optimization process. Intuitively, this additional iteration allows to *foresee* the geometry of the problem and take into account curvature information, one of the most important bottlenecks for first order methods. Motivated by this observation, our goal is to extend and understand further the extragradient method in the specific setting of **(P)**. Appart from the work of Luo and Tseng, the literature on this topic is quite scarce. For example, the nonconvex case is not considered at all.

We combine the work of [71], [92] and recent extensions for first-order descent methods, (see [2, 4, 26, 24]), and propose the extented extragradient method **(EEG)** to tackle problem **(P)**. The classical extragradient method relies on orthogonal projections. We extend it by considering more general nonsmooth convex functions, the **(EEG)** method is given by the following recursion,

$$\begin{cases} y_k = \text{prox}_{s_k g}(x_k - s_k \nabla f(x_k)) \\ x_{k+1} = \text{prox}_{\alpha_k g}(x_k - \alpha_k \nabla f(y_k)), \end{cases}$$

where s_k, α_k are positive real number. An important challenge in this context is to balance the magnitude of these two parameters to maintain desirable convergence properties. We devise conditions which allow to prove convergence of the method when f is nonconvex. In addition, we describe two different rates of convergence when f is convex.

Following [2, 4, 26, 24] we heavily rely on the Kurdyka-Lojasiewicz (KL) inequality to study the nonconvex setting. The KL inequality [86, 73] has a long history in convergence analysis and nonsmooth optimization. Furthermore, recent generalizations [21, 22] have shown the important versatility of this approach as the inequality holds true for the vast majority of models encountered in practice. This opened the possibility to devise general and abstract convergence results for first order methods [4, 26], which constitute an important ingredient of our analysis. Based on this approach, we derive a general convergence result for the proposed **(EEG)** method.

When the function f is convex, problem **(P)** becomes convex and we may consider global convergence rates. We first describe a $1/k$ nonasymptotic rate in terms of objective function. This is related to classical results from the analysis of first order methods in convex optimization, see for example the analysis of forward–backward splitting method in [17]. Furthermore, we show that the *small-prox* result of [24] also applies to **(EEG)** method which echoes the error bound framework of Luo and Tseng [92] and opens the door to more refined complexity results when further properties of the objective function are available.

As already mentioned, a distinguished aspect of the extragradient method is the use of an additional proximal gradient step at each iteration. The intuition behind this mechanism is the incorporation of curvature information in the optimization process. It is expected that one of the effects of this additional step is to allow talking larger step sizes. With this in mind, the analysis of **(EEG)** method is the occasion to describe an exact line search variant of the method:

$$\begin{cases} y_k = \operatorname{prox}_{s_k g}(x_k - s_k \nabla f(x_k)) \\ \alpha_k = \operatorname{argmin}_{\alpha \in \mathbb{R}} F\left(\operatorname{prox}_{\alpha g}(x_k - \alpha \nabla f(y_k))\right) \\ x_{k+1} = \operatorname{prox}_{\alpha_k g}(x_k - \alpha_k \nabla f(y_k)). \end{cases}$$

Although computing the solution to the exact line search is a nonconvex problem, potentially hard in the general case, we describe an active set method to tackle it for the specific and very popular case of ℓ^1 regularized least squares. In this setting the computational overhead of the exact line search has a magnitude roughly similar to that of a gradient computation (discarding additional logarithmic terms).

On the practical side, we compare the performance of the proposed **(EEG)** method (and its line search variant) to those of FISTA and forward-backward splitting methods on the ℓ^1 regularized least squares problem. The numerical results suggest that in the setting of ill conditioned problems, both **(EEG)** and forward-backward, when combined with exact line search, constitute promising alternatives to FISTA.

Structure of the paper. Section 2 introduces the problem and our main assumptions. We also recall important definitions and notations which will be used throughout the text. Section 3 contains the main convergence results of this paper. More precisely, in Subsection 3.3, we present the convergence and finite length property under KL assumption in the nonconvex case. Subsection 3.4, contains both a proof of sublinear convergence rate and the application of the *small-prox* result for **(EEG)** method leading to improved complexity analysis under Kurdyka-Lojasiewicz inequality assumption. Section 4 describes exact line search for proximal gradient steps in the context of ℓ^1 penalized least-squares and results from numerical experiments.

3.2 The Problem and Some Preliminaries

3.2.1 The Problem

We are interested in solving minimization problems of the form

(P) $$\min_{x \in \mathbb{R}^n} \{F(x) = f(x) + g(x)\},$$

where f, g are extended value functions from $\mathbb{R}^n$ to $(-\infty, +\infty]$. We make the following standing assumptions:

- $\operatorname{argmin} F \neq \emptyset$, and we note $F^* = \min_{\mathbb{R}^n} F$
- g is a lower semi-continuous, convex, proper function.
- f is differentiable with L-Lipschitz continuous gradient, where $L > 0$.

Let us give two classical examples fitting these assumptions.

Constrained minimization. Let C be a closed convex set of $\mathbb{R}^n$. Define g to be the indicator function (i_C) of the set C:

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

Then, the unconstrained minimization of composite function is equivalent to minimize the function f over the set C.

Regularized least squares. The ℓ^1 regularized least squares problem consists in the minimization of the following nonsmooth objective function:

$$F(x) = \frac{1}{2}\|Ax - b\|_2^2 + \lambda\|x\|_1$$

where $A \in \mathbb{R}^{n\times p}$ is a real matrix, $b \in \mathbb{R}^n$ be is real vector, $\lambda > 0$ is a positive real and $\|.\|_1$ denotes the l_1-norm, the sum of coordinates absolute value. Many approaches based on ℓ^1 regularized least squares are very popular in signal processing and statistics.

3.2.2 Nonsmooths analysis

In this subsection, we recall the definitions, notations and some well-known results from nonsmooth analysis which are going to be used throughout the paper. We will use notations from [114] (see also [14]). Let $h : \mathbb{R}^n \to (-\infty, +\infty]$ be a proper, lower-semicontinuous function. For each $x \in \operatorname{dom} h$, the Fréchet subdifferential of h at x, written $\hat{\partial}h(x)$, is the set of vectors $u \in \mathbb{R}^n$ which satisfy

$$\liminf_{y\to x} \frac{h(y) - h(x) - \langle u, y - x\rangle}{\|x - y\|} \geq 0.$$

When $x \notin \operatorname{dom} h$, we set $\hat{\partial}h(x) = \emptyset$. We will use the following set

$$\operatorname{graph}(\hat{\partial}h) = \left\{(x, u) \in \mathbb{R}^n \times \mathbb{R}^n : \ u \in \hat{\partial}h(x)\right\}.$$

The subdifferential of h at $x \in \operatorname{dom} h$ is defined by the following closure process

$$\partial h(x) = \left\{u \in \mathbb{R}^n : \exists\, (x_m, u_m)_{m\in\mathbb{N}} \in \operatorname{graph}(\hat{\partial}h)^{\mathbb{N}},\ x_m \underset{m\to\infty}{\to} x,\ h(x_m) \underset{m\to\infty}{\to} h(x),\ u_m \underset{m\to\infty}{\to} u\right\}.$$

$\operatorname{graph}(\partial h)$ is defined similarly as $\operatorname{graph}(\hat{\partial}h)$. When h is convex, the above definition coincides with the usual notion of subdifferential in convex analysis

$$\partial h(x) = \{u \in \mathbb{R}^n : \ h(y) \geq h(x) + \langle u, y - x\rangle \text{ for all } y \in \mathbb{R}^n\}.$$

Independently from the definition, when h is smooth at x then the subdifferential is a singleton, $\partial h(x) = \{\nabla h(x)\}$.

We can deduce from its definition the following closeness property of the subdifferential: If a sequence $(x_m, u_m)_{m\in\mathbb{N}} \in \operatorname{graph}(\partial h)^{\mathbb{N}}$, converges to (x, u), and $h(x_m)$ converges to $h(x)$ then $u \in \partial h(x)$. The set $\operatorname{crit} h = \{x \in \mathbb{R}^n : 0 \in \partial h(x)\}$ is called the set of critical points of h. In this nonsmooth context, the Fermat's rule remains unchanged: A necessary condition for x to be local minimizer of h is that $x \in \operatorname{crit} h$ [114, Theorem 10.1].

Under our standing assumption, f is a smooth function and we have subdifferential sum rule ([Exercise 10.10][114])

$$\partial(f+h)(x) = \nabla f(x) + \partial h(x). \tag{3.1}$$

We recall a well known important property of smooth functions which have L-Lipschitz continuous gradient, see [99, Lemma 1.2.3].

Lemma 3.2.1 (Descent Lemma (e.g. Lemma 1.2.3 in [99]))**.** *For any $x, y \in \mathbb{R}^n$, we have*

$$f(y) \leq f(x) + \langle y - x, \nabla f(x)\rangle + \frac{L}{2}\|x-y\|^2.$$

For the rest of this paragraph, we suppose that h is a convex function. Given $x \in \mathbb{R}^n$ and $t > 0$, the proximal operator associated to h, which we denote by $\text{prox}_{th}(x)$, is defined as the unique minimizer of function $y \longmapsto\ h(y) + \dfrac{1}{2t}\|y-x\|^2$, i.e:

$$\text{prox}_{th}(x) = \text{argmin}_{y\in\mathbb{R}^n}\, h(y) + \frac{1}{2t}\|y-x\|^2.$$

Using Fermat's Rule, $\text{prox}_{th}(x)$ is characterized as the unique solution of the inclusion

$$\frac{x - \text{prox}_{th}(x)}{t} \in \partial h\left(\text{prox}_{th}(x)\right).$$

We can check that when h is convex then prox_h is Lipschitz continuous with constant 1 (see[14, Proposition 12.27]). As an illustration, let $C \subset \mathbb{R}^n$ be a closed, convex and nonempty set, then prox_{i_C} is the orthogonal projection operator onto C. The following property of the prox mapping will be used in the analysis, see [17, Lemma 1.4].

Lemma 3.2.2. *Let $x \in \mathbb{R}^n$, $t > 0$, and $p = \text{prox}_{th}\, x$, then*

$$h(z) - h(p) \geq \frac{1}{2t}\left(\|x-p\|^2 + \|z-p\|^2 - \|x-z\|^2\right), \forall z \in \mathbb{R}^n.$$

3.2.3 Nonsmooth Kurdyka-Łojasiewicz inequality

In this subsection, we present the nonsmooth Kurdyka-Łojasiewicz inequality introduced in [21] (see also [22, 23], and the fundamental works [86, 73]). We note $[h < \mu] = \{x \in \mathbb{R}^n : h(x) < \mu\}$ and $[\eta < h < \mu] = \{x \in \mathbb{R}^n : \eta < h(x) < \mu\}$. Let $r_0 > 0$ and set

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

Definition 6. *The function h satisfies the* Kurdyka-Łojasiewicz (KL) inequality *(or has the KL* property*) locally at $\bar{x} \in \text{dom}\, f$ if there exist $r_0 > 0$, $\varphi \in \mathcal{K}(r_0)$ and a neighborhood $U(\bar{x})$ such that*

$$\varphi'\left(h(x) - h(\bar{x})\right)\, \text{dist}(0, \partial h(x)) \geq 1 \tag{3.2}$$

for all $x \in U(\bar{x}) \cap [h(\bar{x}) < h(x) < h(\bar{x}) + r_0]$. We say that φ is a desingularizing function *for F at $\bar{x}$. The function h has the KL property on S if it does so at each point of S.*

When h is smooth and $h(\bar{x}) = 0$ then (3.2) can be rewritten as

$$\|\nabla(\varphi \circ h)\| \geq 1, \, \forall x \in U(\bar{x}) \cap [h(\bar{x}) < h(x) < h(\bar{x}) + r_0].$$

This inequality may be interpreted as follows: The function h can be made sharp locally by a reparameterization of its values through a function $\varphi \in \mathcal{K}(r_0)$ for some $r_0 > 0$.

The KL inequality is obviously satisfied at any noncritical point $\bar{x} \in \operatorname{dom} h$ and will thus be useful only for critical points, $\bar{x} \in \operatorname{crit} h$. The *Łojasiewicz gradient inequality* corresponds to the case when $\varphi(s) = cs^{1-\theta}$ for some $c > 0$ and $\theta \in [0, 1)$. The class of functions which satisfy KL inequality is extremely vast. Typical KL functions are semi-algebraic functions, but there exists many extensions, see [21].

If h 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. The following lemma is similar to [26, Lemma 3.6].

Lemma 3.2.3 (KL property)**.** *Let Ω be a compact set and let $h : \mathbb{R}^n \to (-\infty, \infty]$ be a proper and lower semicontinuous function. We assume that h is constant on Ω and satisfies the KL property at each point of Ω. Then there exist $\varepsilon > 0$, $\eta > 0$ and φ such that for all $\bar{x} \in \Omega$, one has*

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

for all $x \in \{x \in \mathbb{R}^n | \operatorname{dist}(x, \Omega) < \varepsilon\} \cap [h(\bar{x}) < h(x) < h(\bar{x}) + \eta]$.

3.3 Extragradient method, Convergence and Complexity

3.3.1 Extragradient method

We now describe our extragradient method dedicated to the minimization of problem (**P**). Recall that the method is defined, given an initial estimate $x_0 \in \mathbb{R}^n$, by the following recursion, for $k \geq 1$,

$$(\mathbf{EEG}) \begin{cases} y_k = \operatorname{prox}_{s_k g}(x_k - s_k \nabla f(x_k)), & (3.3) \\ x_{k+1} = \operatorname{prox}_{\alpha_k g}(x_k - \alpha_k \nabla f(y_k)). & (3.4) \end{cases}$$

where $(s_k)_{k\in\mathbb{N}}, (\alpha_k)_{k\in\mathbb{N}}$ are positive step size sequences. We introduce relevant quantities, $s_- = \inf_{k\in\mathbb{N}} s_k$, $s^+ = \sup_{k\in\mathbb{N}} s_k$, and $\alpha_- = \inf_{k\in\mathbb{N}} \alpha_k$, $\alpha^+ = \sup_{k\in\mathbb{N}} \alpha_k$ for $k \in \mathbb{N}$. Throughout the paper, we will consider the following condition on the two step size sequence,

$$(\mathbf{C}) : 0 < \alpha_-, \, 0 < s_-, \, s^+ < \frac{1}{L} \text{ and } 0 < s_k \leq \alpha_k, \, \forall k \in \mathbb{N}.$$

Depending on the context, additional restrictions will be imposed on the step size sequences.

3.3.2 Basic Properties

We introduce in this subsection two technical properties of sequences produced by(**EEG**) method. They will play a crucial role in the proofs of our main convergence and complexity results. We begin with a descent property.

Proposition 3.3.1 (Descent condition)**.** *For any $k \in \mathbb{N}$, we have*

$$F(x_k) - F(x_{k+1}) \geq \frac{1}{2\alpha_k}\|x_k - x_{k+1}\|^2 + \left(\frac{1}{s_k} - \frac{L}{2} - \frac{1}{2\alpha_k}\right)\|x_k - y_k\|^2 + \left(\frac{1}{2\alpha_k} - \frac{L}{2}\right)\|y_k - x_{k+1}\|^2.$$

Proof. We fix an arbitrary $k \in \mathbb{N}$. Applying Lemma 3.2.2 for (3.3), with $z = x_k$, $p = y_k$, $x = x_k - s_k \nabla f(x_k)$ and $t = s_k$, we obtain

$$\begin{aligned} g(x_k) - g(y_k) &\geq \frac{1}{2s_k} \left(\|x_k - y_k\|^2 + \|x_k - s_k \nabla f(x_k) - y_k\|^2 - \|s_k \nabla f(x_k)\|^2 \right) \\ &= \frac{1}{s_k} \|x_y - y_k\|^2 + \langle y_k - x_k, \nabla f(x_k) \rangle. \end{aligned}$$

Combining with the descent lemma, $f(y_k) \leq f(x_k) + \langle y_k - x_k, \nabla f(x_k) \rangle + \frac{L}{2} \|x_k - y_k\|^2$, we get

$$F(x_k) - F(y_k) \geq \left(\frac{1}{s_k} - \frac{L}{2} \right) \|x_k - y_k\|^2. \tag{3.5}$$

Similarly, applying Lemma 3.2.2 for (3.4), with $z := y_k$, $p := x_{k+1}$ and $x = x_k - \alpha_k \nabla f(y_k)$ we obtain

$$\begin{aligned} g(y_k) - g(x_{k+1}) &\geq \frac{1}{2\alpha_k} \left(\|y_k - x_{k+1}\|^2 + \|x_k - \alpha_k \nabla f(y_k) - x_{k+1}\|^2 - \|y_k - x_k + \alpha_k \nabla f(y_k)\|^2 \right) \\ &= \frac{1}{2\alpha_k} \left(\|y_k - x_{k+1}\|^2 + \|x_k - x_{k+1}\|^2 - \|y_k - x_k\|^2 \right) + \langle x_{k+1} - y_k, \nabla f(y_k) \rangle. \end{aligned}$$

On the other hand, we have from the descent lemma that

$$f(x_{k+1}) \leq f(y_k) + \langle x_{k+1} - y_k, \nabla f(y_k) \rangle + \frac{L}{2} \|y_k - x_{k+1}\|^2.$$

Summing up the last two inequalities, we have

$$F(y_k) - F(x_{k+1}) \geq \frac{1}{2\alpha_k} \left(\|x_k - x_{k+1}\|^2 - \|y_k - x_k\|^2 \right) + \left(\frac{1}{2\alpha_k} - \frac{L}{2} \right) \|y_k - x_{k+1}\|^2. \tag{3.6}$$

Combining inequalities (3.5) and (3.6), we obtain

$$F(x_k) - F(x_{k+1}) \geq \frac{1}{2\alpha_k} \|x_k - x_{k+1}\|^2 + \left(\frac{1}{s_k} - \frac{L}{2} - \frac{1}{2\alpha_k} \right) \|x_k - y_k\|^2 + \left(\frac{1}{2\alpha_k} - \frac{L}{2} \right) \|y_k - x_{k+1}\|^2, \tag{3.7}$$

which concludes the proof □

Remark 3.3.2. *If we combine the constraint that $0 < \alpha_k \leq \frac{1}{L}$ for all $k \in \mathbb{N}$ with condition* **(C)**, *we deduce from Proposition 3.3.1 that, for all $k \in \mathbb{N}$, $\frac{1}{s_k} - \frac{L}{2} - \frac{1}{2\alpha_k} \geq 0$, and*

$$F(x_k) - F(x_{k+1}) \geq \frac{1}{2\alpha_k} \|x_k - x_{k+1}\|^2.$$

Under this condition, we have that **(EEG)** *is a descent method in the sense that it will produce a decreasing sequence of objective value.*

We now establish a second property of sequences produced by (**EEG**) method which is interpreted as a subgradient step property.

Proposition 3.3.3 (Subgradient step)**.** *Assume that $(s_k, \alpha_k)_{k \in \mathbb{N}}$ satisfy condition* **(C)**. *Then for any $k \in \mathbb{N}$, there exists $u_{k+1} \in \partial g(x_{k+1})$ such that*

$$\|u_{k+1} + \nabla f(x_{k+1})\| \leq \frac{L\alpha_k + (1 - Ls_k)^2}{\alpha_k (1 - Ls_k)} \|x_k - x_{k+1}\|.$$

Proof. We write the optimality condition for (3.4),

$$\frac{x_k - x_{k+1}}{\alpha_k} - \nabla f(y_k) \in \partial g(x_{k+1}), \tag{3.8}$$

therefore, there exists $u_{k+1} \in \partial g(x_{k+1})$ such that

$$\frac{x_k - x_{k+1}}{\alpha_k} + \nabla f(x_{k+1}) - \nabla f(y_k) = u_{k+1} + \nabla f(x_{k+1}).$$

This implies that

$$\|u_{k+1} + \nabla f(x_{k+1})\| \leq \frac{\|x_k - x_{k+1}\|}{\alpha_k} + \|\nabla f(x_{k+1}) - \nabla f(y_k)\|.$$

Since ∇f is L-Lipschitz continuous, it follows that

$$\|u_{k+1} + \nabla f(x_{k+1})\| \leq \frac{\|x_k - x_{k+1}\|}{\alpha_k} + L\|x_{k+1} - y_k\|. \tag{3.9}$$

Denote $z_{k+1} = \text{prox}_{s_k g}(x_k - s_k \nabla f(y_k))$, since the $\text{prox}_{s_k g}$ is 1-Lipschitz continuous, we get

$$\begin{aligned} \|y_k - z_{k+1}\| &\leq \|(x_k - s_k \nabla f(y_k)) - (x_k - s_k \nabla f(x_k))\| \\ &\leq L s_k \|x_k - y_k\|, \end{aligned}$$

and therefore

$$\|x_k - z_{k+1}\| \geq \|x_k - y_k\| - \|y_k - z_{k+1}\| \geq (1 - L s_k)\|x_k - y_k\|. \tag{3.10}$$

On the other hand, g is convex, thus in view of the definition of z_{k+1},

$$\left\langle \frac{x_k - z_{k+1}}{s_k} - \nabla f(y_k), x_{k+1} - z_{k+1} \right\rangle \leq g(x_{k+1}) - g(z_{k+1}).$$

Similarly, from (3.8) and convexity of g, we get

$$\left\langle \frac{x_k - x_{k+1}}{\alpha_k} - \nabla f(y_k), z_{k+1} - x_{k+1} \right\rangle \leq g(z_{k+1}) - g(x_{k+1}).$$

Adding the last two inequalities, we obtain

$$\left\langle \frac{x_k - z_{k+1}}{s_k} - \frac{x_k - x_{k+1}}{\alpha_k}, x_{k+1} - z_{k+1} \right\rangle \leq 0,$$

or equivalently

$$\left\langle \frac{x_k - z_{k+1}}{s_k} - \frac{x_k - x_{k+1}}{\alpha_k}, (x_{k+1} - x_k) + (x_k - z_{k+1}) \right\rangle \leq 0.$$

It follows that

$$\frac{\|x_k - z_{k+1}\|^2}{s_k} + \frac{\|x_k - x_{k+1}\|^2}{\alpha_k} \leq \left(\frac{1}{s_k} + \frac{1}{\alpha_k} \right) \langle x_k - z_{k+1}, x_k - x_{k+1} \rangle ,$$

Using Cauchy-Schwarz inequality, we get

$$\frac{\|x_k - z_{k+1}\|^2}{s_k} + \frac{\|x_k - x_{k+1}\|^2}{\alpha_k} \leq \left(\frac{1}{s_k} + \frac{1}{\alpha_k}\right) \|x_k - z_{k+1}\|.\|x_k - x_{k+1}\|.$$

Since from condition **(C)**, $0 < s_k$, this is equivalent to

$$(\|x_k - z_{k+1}\| - \|x_k - x_{k+1}\|) \left(\|x_k - z_{k+1}\| - \frac{s_k\|x_k - x_{k+1}\|}{\alpha_k}\right) \leq 0.$$

This inequality asserts that the product of two terms is nonpositive. Hence one of the terms must be nonpositive and the other one must be nonnegative. From condition **(C)**, we have $\frac{s_k}{\alpha_k} \leq 1$, the last term is bigger than the first one and hence must be nonnegative. This yields

$$\frac{s_k}{\alpha_k}\|x_k - x_{k+1}\| \leq \|x_k - z_{k+1}\| \leq \|x_k - x_{k+1}\|.$$

By combining the latter inequality with (3.10), we get

$$\|x_k - x_{k+1}\| \geq (1 - Ls_k)\|x_k - y_k\|. \tag{3.11}$$

Similarly, from the definitions of y_k, x_{k+1} and the convexity of g, we obtain that

$$\left\langle \frac{x_k - y_k}{s_k} - \nabla f(x_k), x_{k+1} - y_k \right\rangle \leq g(x_{k+1}) - g(y_k),$$

and

$$\left\langle \frac{x_k - x_{k+1}}{\alpha_k} - \nabla f(y_k), y_k - x_{k+1} \right\rangle \leq g(y_k) - g(x_{k+1}).$$

Summing the last two inequalities, we have that

$$\frac{1}{s_k}\|x_{k+1} - y_k\|^2 + \left(\frac{1}{s_k} - \frac{1}{\alpha_k}\right) \langle x_{k+1} - y_k, x_k - x_{k+1}\rangle \leq \langle x_{k+1} - y_k, \nabla f(x_k) - \nabla f(y_k)\rangle.$$

Using the condition $0 < s_k \leq \alpha_k$ and Cauchy-Schwarz inequality, we get

$$\frac{1}{s_k}\|x_{k+1} - y_k\|^2 \leq \left(\frac{1}{s_k} - \frac{1}{\alpha_k}\right) \|x_{k+1} - y_k\|\|x_k - x_{k+1}\| + \|x_{k+1} - y_k\|\|\nabla f(x_k) - \nabla f(y_k)\|.$$

Using Lipschitz continuity of ∇f, we have that

$$\|x_{k+1} - y_k\| \leq \left(1 - \frac{s_k}{\alpha_k}\right) \|x_k - x_{k+1}\| + Ls_k\|x_k - y_k\|.$$

Combining this inequality with (3.11), we obtain

$$\begin{aligned} \|x_{k+1} - y_k\| &\leq \left(1 - \frac{s_k}{\alpha_k} + \frac{Ls_k}{1 - Ls_k}\right) \|x_k - x_{k+1}\| \\ &= \left(\frac{1}{1 - Ls_k} - \frac{s_k}{\alpha_k}\right) \|x_k - x_{k+1}\|. \end{aligned} \tag{3.12}$$

Combining (3.12) with (3.9), we get

$$\|u_{k+1} + \nabla f(x_{k+1})\| \leq \frac{L\alpha_k + (1 - Ls_k)^2}{\alpha_k(1 - Ls_k)}\|x_k - x_{k+1}\|, \tag{3.13}$$

and the result is proved □

Combining Remark 3.3.2 and Proposition 3.3.3 above, we have the following corollary which underlines the fact that (**EEG**) is actually an approximate gradient method in the sense of [4].

Corollary 3.3.4. *Assume that* $(s_k, \alpha_k)_{k\in\mathbb{N}}$ *satisfy the following*

$$(\mathbf{C1}) : (s_k, \alpha_k)_{k\in\mathbb{N}} \text{ satisfy condition } (\mathbf{C}) \text{ and } \alpha_k \leq \frac{1}{L}, \forall k \in \mathbb{N}.$$

Then, for all $k \in \mathbb{N}$

i) $F(x_{k+1}) + \frac{1}{2\alpha_k}\|x_k - x_{k+1}\|^2 \leq F(x_k)$.

ii) *There exists* $\omega_{k+1} \in \partial F(x_{k+1})$ *such that*

$$\|\omega_{k+1}\| \leq b_k\|x_k - x_{k+1}\|,$$

where,

$$0 < b_k := \frac{L\alpha_k + (1 - Ls_k)^2}{\alpha_k(1 - Ls_k)} \leq b := \frac{2}{\alpha_-(1 - s^+L)}.$$

3.3.3 Convergence of extragradient method under KL assumption

In this subsection, we analyse the convergence of (**EEG**) method in the nonconvex setting. The main result is stated in Theorem 3.3.6, which also describes the asymptotic rate of convergence. This result is based on the assumptions that F has the KL property on $\operatorname{crit} F$ and that $(s_k, \alpha_k)_{k\in\mathbb{N}}$ satisfy conditions **(C1)** from Corollary 3.3.4. We will also assume that the sequence $(x_k)_{k\in\mathbb{N}}$ generated by (**EEG**) is bounded. This boundedness assumption is not very restrictive here, since under condition **(C1)**, Corollary 3.3.4 ensures that it is satisfied for any coercive objective function. Similarly to [26, Lemma 3.5], we first give some properties of F on the set of accumulation points of $(x_k)_{k\in\mathbb{N}}$.

Lemma 3.3.5. *Assume that* $(s_k, \alpha_k)_{k\in\mathbb{N}}$ *satisfy condition* **(C1)** *and that* $(x_k)_{k\in\mathbb{N}}$ *is bounded. Let* Ω_0 *be the set of limit points of the sequence* $(x_k)_{k\in\mathbb{N}}$. *It holds that* Ω_0 *is compact and nonempty,* $\Omega_0 \subset \operatorname{crit} F$, $\operatorname{dist}(x_k, \Omega_0) \to 0$ *and* $F(\bar{x}) = \lim_{k\to\infty} F(x_k)$ *for all* $\bar{x} \in \Omega_0$.

Proof. From the boundedness assumption, it is clear that Ω_0 is nonempty. In view of Corollary 3.3.4 **i)**, it follows that $(F(x_k))_{k\in\mathbb{N}}$ is nonincreasing. Furthermore, $F(x_k)$ is bounded from below by F^*, hence there exists $\bar{F} \in \mathbb{R}$ such that $\bar{F} = \lim_{k\to\infty} F(x_k)$. In addition, we have

$$\sum_{k=1}^{m} \|x_{k+1} - x_k\|^2 \leq 2\alpha^+ \left(F(1) - F(m+1)\right),$$

therefore $\sum_{k=1}^{\infty} \|x_{k+1} - x_k\|^2$ converges, thus $(x_{k+1} - x_k) \to 0$. We now fix an arbitrary point $x^* \in \Omega_0$, which means that there exists a subsequence $(x_{k_q})_{q\in\mathbb{N}}$ of $(x_k)_{k\in\mathbb{N}}$ such that $\lim_{q\to\infty} x_{k_q} = x^*$, therefore, by lower semicontinuity of g and continuity of f,

$$g(x^*) \leq \liminf_{q\to\infty} g(x_{k_q}),\ f(x^*) = \lim_{q\to\infty} f(x_{k_q}). \tag{3.14}$$

From the definition of x_{k_q} and condition **(C1)**, we get for all $q \in \mathbb{N}$,

$$\begin{aligned}
&g(x_{k_q}) + \frac{1}{2s_+}\|x_{k_q-1} - x_{k_q}\|^2 + \left\langle x_{k_q} - x_{k_q-1}, \nabla f(y_{k_q-1})\right\rangle \\
&\leq g(x_{k_q}) + \frac{1}{2s_{k_q}}\|x_{k_q-1} - x_{k_q}\|^2 + \left\langle x_{k_q} - x_{k_q-1}, \nabla f(y_{k_q-1})\right\rangle \\
&\leq g(x^*) + \frac{1}{2s_{k_q}}\|x^* - x_{k_q-1}\|^2 + \left\langle x^* - x_{k_q-1}, \nabla f(y_{k_q-1})\right\rangle. \\
&\leq g(x^*) + \frac{1}{2s_-}\|x^* - x_{k_q-1}\|^2 + \left\langle x^* - x_{k_q-1}, \nabla f(y_{k_q-1})\right\rangle.
\end{aligned}$$

Let $q \to \infty$, it follows that $\limsup_{q\to\infty} g(x_{k_q}) \leq g(x^*)$, thus, in view of (3.14), $\lim_{q\to\infty} g(x_{k_q}) = g(x^*)$, therefore $\lim_{q\to\infty} F(x_{k_q}) = F(x^*)$. Since $F(x_k)$ is nonincreasing, $\lim_{q\to\infty} F(x_{k_q}) = \bar{F}$, and we deduce that $F(x^*) = \bar{F}$. Since x^* was arbitrary in Ω_0, it holds that F is constant on Ω_0.

Now, thanks to Corollary 3.3.4 **ii)**, there exist $\omega_{k+1} \in \partial F(x_{k+1})$, such that

$$\|\omega_{k+1}\| \leq b_k\|x_k - x_{k+1}\|.$$

Under condition (**C1**), it holds that b_k remains bounded. Since $\lim_{k\to\infty} x_k - x_{k+1} = 0$, it holds that $\omega_k \to 0$. Combining with the closeness of ∂F, this implies that $0 \in \partial F(x^*)$, hence $x^* \in \operatorname{crit} F$. Since x^* was taken arbitrarily in Ω_0, this means that $\Omega_0 \subset \operatorname{crit} F$. The compactness of Ω_0 is implied by [26, Lemma 3.5]. Combining the boundedness of $(x_k)_{k\in\mathbb{N}}$ and the compactness of Ω_0, we deduce that $\operatorname{dist}(x_k, \Omega_0) \to 0$ which concludes the proof. □

We are now in a position to prove the convergence of the extragradient method in the nonconvex case.

Theorem 3.3.6. *Assume that $(s_k, \alpha_k)_{k\in\mathbb{N}}$ satisfy condition* **(C1)**, *that F has the KL property on $\operatorname{crit} F$ and that $(x_k)_{k\in\mathbb{N}}$ is bounded. Then the sequence $(x_k)_{k\in\mathbb{N}}$ converges to $x^* \in \operatorname{crit} F$, moreover*

$$\sum_{i=1}^{\infty} \|x_k - x_{k+1}\| < \infty.$$

Proof. Note that Lemma 3.3.5 can be applied here and we will use the same notations. We write $\lim_{k\to\infty} F(x_k) = \bar{F}$ and let Ω_0 be the set of limit points of $(x_k)_{k\in\mathbb{N}}$. Combining the KL assumption and Lemma 3.2.3, there exists $\varepsilon > 0$, $\eta > 0$ and a desingularizing function $\varphi \in \mathcal{K}(\eta)$ such that

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

for all $x \in \{x \in \mathbb{R}^n | \operatorname{dist}(x, \Omega_0) < \varepsilon\} \cap [\bar{F} < F(x) < \bar{F} + \eta]$. Denote by i the first index such that $\|x_i - x_{i-1}\| = 0$ or $F(x_i) = \bar{F}$. If such an i exists, one has $\omega_i = 0$ and $x_k = x_i$, for all $k > i$ which shows that the result holds true. For the rest of the proof, we will assume that $\|x_k - x_{k-1}\| > 0$ and $F(x_k) > \bar{F}$ for all $k \geq 1$. From Lemma 3.3.5, we have $\operatorname{dist}(x_k, \Omega_0) \to 0$ and $F(x_k) \to \bar{F}$, therefore there exists $k_0 \in \mathbb{N}$ such that for all $k \geq k_0$,

$$\operatorname{dist}(x_k, \Omega_0) < \varepsilon \text{ and } \bar{F} < F(x_k) < \bar{F} + \eta.$$

Using the concavity of φ and Corollary 3.3.4 we obtain

$$\begin{aligned}
\varphi(F(x_k) - \bar{F}) - \varphi(F(x_{k+1}) - \bar{F}) &\geq \varphi'(F(x_k) - \bar{F})\left(F(x_k) - F(x_{k+1})\right) \\
&\geq \frac{1}{2\alpha_k}\varphi'(F(x_k) - \bar{F})\|x_k - x_{k+1}\|^2, \forall k \geq k_0.
\end{aligned}$$

Since F has KL property on Ω_0, using again Corollary 3.3.4, we get

$$\varphi'(F(x_k) - \bar{F}) \geq \frac{1}{b_k \|x_k - x_{k_1}\|}, \forall k \geq k_0$$

Combining the last two inequalities, we obtain for all $k \geq k_0$,

$$\begin{aligned}
\varphi(F(x_k) - \bar{F}) - \varphi(F(x_{k+1}) - \bar{F}) &\geq \frac{\|x_k - x_{k+1}\|^2}{2\alpha_k b_k \|x_k - x_{k-1}\|} \\
&\geq \frac{1}{2\alpha_k b_k} \frac{\big(2\|x_k - x_{k+1}\|\|x_k - x_{k-1}\| - \|x_{k-1} - x_k\|^2\big)}{\|x_k - x_{k-1}\|} \\
&= \frac{1}{2\alpha_k b_k}(2\|x_k - x_{k+1}\| - \|x_{k-1} - x_k\|) \\
&\geq \frac{1}{2\alpha^+ b}(2\|x_k - x_{k+1}\| - \|x_{k-1} - x_k\|), \qquad (3.15)
\end{aligned}$$

where b is given in Corollary 3.3.4. This implies that

$$2\alpha^+ b\big(\varphi(F(x_{k_0}) - \bar{F}) - \varphi(F(x_{k+1}) - \bar{F})\big) + \|x_0 - x_1\| \geq \sum_{i=k_0}^{k} \|x_i - x_{i+1}\|, \forall k > k_0.$$

Therefore, the series $\sum_{i=k_0}^{\infty} \|x_i - x_{i+1}\|$ is bounded and hence converges. By Cauchy criterion, it follows that the sequence $(x_k)_{k\in\mathbb{N}}$ converges to some point $x^* \in \mathbb{R}^n$. Furthermore, from Lemma 3.3.5, $x^* \in \operatorname{crit} F$ which concludes the proof. □

Remark 3.3.7 (Convergence rate)**.** *When the KL desingularizing function of F is of the form $\varphi(s) = cs^{1-\theta}$, where c is a positive constant and $\theta \in (0, 1]$, then we can estimate the rate of convergence of the sequence $(x_k)_{k\in\mathbb{N}}$, as follows (see Theorem 2, [2]).*

- *$\theta = 0$ then the sequence (x_k) converges in a finite number of steps.*
- *$\theta \in [0, \frac{1}{2}]$ then there exist $C > 0$ and $\tau \in (0, 1)$ such that*

$$\|x_k - x^*\| \leq C\tau^k, \forall k \in \mathbb{N}.$$

- *$\theta \in (\frac{1}{2}, 1)$ then there exist $C > 0$ such that*

$$\|x_k - x^*\| \leq Ck^{-\frac{1-\theta}{2\theta-1}}, \forall k \in \mathbb{N}.$$

3.3.4 The complexity of extragradient in the convex case

Throughout this section, we suppose that the function f is convex and we focus on complexity and non asymptotic convergence rate analysis.

3.3.4.1 Sublinear convergence rate analysis

We begin with a technical Lemma which introduces more restrictive step size conditions.

Lemma 3.3.8. *Assume that $(s_k, \alpha_k)_{k\in\mathbb{N}}$ satisfy the following*

(C2): *$(s_k, \alpha_k)_{k\in\mathbb{N}}$ satisfy condition* **(C)** *and $s_k \leq \frac{1}{2L}$, $\alpha_k \leq \frac{1}{L} - s_k$, $\forall k \in \mathbb{N}$.*

Then for all $k \in \mathbb{N}$,

$$\frac{1}{2\alpha_k}\|x_k - x_{k+1}\|^2 - \frac{L}{2}\|x_{k+1} - y_k\|^2 \geq 0.$$

Proof. First, we note that if $(s_k, \alpha_k)_{k\in\mathbb{N}}$ satisfy **(C2)** then they also satisfy condition **(C1)** and Proposition 3.3.3 applies. Thanks to inequality (3.12) from the proof of Proposition 3.3.3, we get

$$\begin{aligned}\frac{1}{\alpha_k}\|x_k - x_{k+1}\|^2 - L\|x_{k+1} - y_k\|^2 &\geq \frac{1}{\alpha_k}\|x_k - x_{k+1}\|^2 - L\left(\frac{1}{1-Ls_k} - \frac{s_k}{\alpha_k}\right)^2\|x_k - x_{k+1}\|^2 \\ &= \frac{-L\alpha_k^2 + (1 - s_k^2L^2)\alpha_k - Ls_k^2(1-Ls_k)^2}{\alpha_k^2(1-Ls_k)^2}\|x_k - x_{k+1}\|^2.\end{aligned}$$

In addition, it can be checked using elementary calculation that

$$-L\alpha_k^2 + (1 - s_k^2L^2)\alpha_k - Ls_k^2(1-Ls_k)^2 \geq 0,$$

is equivalent to

(3.16)
$$\frac{(1-Ls_k)\left[(1+Ls_k) - \sqrt{(1+Ls_k)^2 - 4L^2s_k^2}\right]}{2L} \leq \alpha_k \leq \frac{(1-Ls_k)\left[(1+Ls_k) + \sqrt{(1+Ls_k)^2 - 4L^2s_k^2}\right]}{2L}.$$

Note that, for $0 \leq b \leq a$ then $a - b \leq \sqrt{a^2 - b^2}$. Using this inequality, with the condition $2Ls_k \leq 1$, we get $(1 + Ls_k) - 2Ls_k \leq \sqrt{(1+Ls_k)^2 - 4L^2s_k^2}$. Thus,

$$\begin{cases}\dfrac{(1-Ls_k)\left[(1+Ls_k) - \sqrt{(1+Ls_k)^2 - 4L^2s_k^2}\right]}{2L} \leq \dfrac{(1-Ls_k)\left[(1+Ls_k) - (1-Ls_k)\right]}{2L} = (1-Ls_k)s_k \leq s_k \\ \dfrac{(1-Ls_k)\left[(1+Ls_k) + \sqrt{(1+Ls_k)^2 - 4L^2s_k^2}\right]}{2L} \geq \dfrac{(1-Ls_k)\left[(1+Ls_k) + (1-Ls_k)\right]}{2L} = \dfrac{1}{L} - s_k.\end{cases}$$

Putting things together, condition **(C2)** implies that

$$\frac{1}{2\alpha_k}\|x_k - x_{k+1}\|^2 - \frac{L}{2}\|x_{k+1} - y_k\|^2 \geq 0, \forall k \in \mathbb{N}.$$

□

With a similar method as in [17], we prove a sublinear convergence rate for $(F(x_k))_{k\in\mathbb{N}}$ in the convex case.

Theorem 3.3.9 (Complexity of extragradient method)**.** *Suppose that $(s_k, \alpha_k)_{k\in\mathbb{N}}$ satisfy condition* **(C2)** *and that f is convex, then, for any $x^* \in \operatorname{argmin} F$, we have*

$$F(x_m) - F(x^*) \leq \frac{1}{2m\alpha_-}\|x_0 - x^*\|^2, \forall m \in \mathbb{N}^*.$$

Proof. We first fix arbirary $k \in \mathbb{N}$ and $x^* \in \operatorname{argmin} F$. Applying Lemma 3.2.2 with $z = x^*$, $p = x_{k+1}$, $x = x_k - \alpha_k \nabla f(y_k)$ and $t = \alpha_k$, we obtain

$$\begin{aligned} g(x^*) - g(x_{k+1}) &\geq \frac{1}{2\alpha_k}\left(\|x^* - x_{k+1}\|^2 + \|x_k - x_{k+1}\|^2 - \|x^* - x_k\|^2\right) + \langle x_{k+1} - x^*, \nabla f(y_k)\rangle \\ &= \frac{1}{2\alpha_k}\left(\|x^* - x_{k+1}\|^2 + \|x_k - x_{k+1}\|^2 - \|x^* - x_k\|^2\right) + \langle x_{k+1} - y_k, \nabla f(y_k)\rangle \\ &+ \langle y_k - x^*, \nabla f(y_k)\rangle \\ &\geq \frac{1}{2\alpha_k}\left(\|x^* - x_{k+1}\|^2 + \|x_k - x_{k+1}\|^2 - \|x^* - x_k\|^2\right) + f(x_{k+1}) - f(y_k) \\ &- \frac{L}{2}\|x_{k+1} - y_k\|^2 + \langle y_k - x^*, \nabla f(y_k)\rangle, \end{aligned}$$

where the last inequality is due to Lemma 3.2.1. It follows that

$$\begin{aligned} F(x^*) - F(x_{k+1}) \geq \frac{1}{2\alpha_k}\left(\|x^* - x_{k+1}\|^2 + \|x_k - x_{k+1}\|^2 - \|x^* - x_k\|^2\right) - \frac{L}{2}\|x_{k+1} - y_k\|^2 \\ + f(x^*) - f(y_k) + \langle y_k - x^*, \nabla f(y_k)\rangle. \end{aligned}$$

Since f is convex, $f(x^*) - f(y_k) + \langle y_k - x^*, \nabla f(y_k)\rangle \geq 0$, and the above inequality implies that

$$F(x^*) - F(x_{k+1}) \geq \frac{1}{2\alpha_k}\left(\|x^* - x_{k+1}\|^2 - \|x^* - x_k\|^2\right) + \frac{1}{2\alpha_k}\|x_k - x_{k+1}\|^2 - \frac{L}{2}\|x_{k+1} - y_k\|^2.$$

Using the fact that $F(x_k)$ is noninreasing and bounded from bellow by $F(x^*)$, it follows from Lemma 3.3.8 that

$$0 \geq F(x^*) - F(x_{k+1}) \geq \frac{1}{2\alpha_k}\left(\|x^* - x_{k+1}\|^2 - \|x^* - x_k\|^2\right) \geq \frac{1}{2\alpha_-}\left(\|x^* - x_{k+1}\|^2 - \|x^* - x_k\|^2\right).$$

Summing this inequality for $k = 0, \cdots, m-1$ gives

$$mF(x^*) - \sum_{k=1}^{m} F(x_k) \geq \frac{1}{2\alpha_-}\left(\|x^* - x_m\|^2 - \|x^* - x_0\|^2\right). \tag{3.17}$$

Coming back to Corollary 3.3.4, it is easy to see that the sequence $(F(x_k))_{k\in\mathbb{N}}$ is nonincreasing, then $\sum_{k=1}^{m} F(x_k) \geq mF(x_m)$. Combining with (3.17), we get

$$m\left(F(x^*) - F(x_m)\right) \geq \frac{1}{2\alpha_-}\left(\|x^* - x_m\|^2 - \|x^* - x_0\|^2\right).$$

It follows that

$$F(x_m) - F(x^*) \leq \frac{1}{2m\alpha_-}\|x^* - x_0\|^2, \forall m \in \mathbb{N}^*.$$

□

3.3.4.2 Small-prox type result under KL property

We now study the complexity of **(EEG)** method when F has, in addition, the KL property on $\operatorname{crit} F$. First, using the convexity of f, Proposition 3.3.1, can be improved by using the following result.

Proposition 3.3.10. *Assume that f is convex and $(s_k, \alpha_k)_{k\in\mathbb{N}}$ satisfy condition* **(C)**, *then for all $k \in \mathbb{N}$, we have*

$$F(x_k) - F(x_{k+1}) \geq c_k \|x_k - x_{k+1}\|^2,$$

where

$$c_k = \frac{1}{\alpha_k} - \frac{L}{2}\left(\frac{1}{1-Ls_k} - \frac{s_k}{\alpha_k}\right)^2.$$

Proof. Fix an arbitrary $k \in \mathbb{N}$. Applying Lemma 3.2.2, with $z = x_k$, $p = x_{k+1}$, $x = x_k - \alpha_k \nabla f(y_k)$ and $t = \alpha_k$, we get

$$\begin{aligned} g(x_k) - g(x_{k+1}) &\geq \frac{1}{\alpha_k}\|x_k - x_{k+1}\|^2 + \langle x_{k+1} - x_k, \nabla f(y_k)\rangle \\ &= \frac{1}{\alpha_k}\|x_k - x_{k+1}\|^2 + \langle x_{k+1} - y_k, \nabla f(y_k)\rangle + \langle y_k - x_k, \nabla f(y_k)\rangle \\ &\geq \frac{1}{\alpha_k}\|x_k - x_{k+1}\|^2 + f(x_{k+1}) - f(y_k) - \frac{L}{2}\|x_{k+1} - y_k\|^2 + \langle y_k - x_k, \nabla f(y_k)\rangle, \end{aligned}$$

where the last inequality follows from Lemma 3.2.1. This implies that

$$F(x_k) - F(x_{k+1}) \geq \frac{1}{\alpha_k}\|x_k - x_{k+1}\|^2 - \frac{L}{2}\|x_{k+1} - y_k\|^2 + f(x_k) - f(y_k) - \langle x_k - y_k, \nabla f(y_k)\rangle.$$

Since f is convex, $f(x_k) - f(y_k) - \langle x_k - y_k, \nabla f(y_k)\rangle \geq 0$, which leads to

$$F(x_k) - F(x_{k+1}) \geq \frac{1}{\alpha_k}\|x_k - x_{k+1}\|^2 - \frac{L}{2}\|x_{k+1} - y_k\|^2. \tag{3.18}$$

From condition **(C)**, Proposition 3.3.3 holds and in particular, inequality (3.12). Combining inequality (3.18) with (3.12), we get the desired result,

$$F(x_k) - F(x_{k+1}) \geq \left[\frac{1}{\alpha_k} - \frac{L}{2}\left(\frac{1}{1-Ls_k} - \frac{s_k}{\alpha_k}\right)^2\right]\|x_k - x_{k+1}\|^2.$$

□

Lemma 3.3.11. *Suppose that s_k, α_k satisfy the following condition*

$$(\mathbf{C3})\begin{cases} s_k, \alpha_k \text{ satisfy } \mathbf{(C)} \text{ condition} \\ s_k \leq \frac{\sqrt{5}-1}{2L}, \text{ and } \alpha_k \leq \frac{2}{L} - 2s_k - (1-Ls_k)Ls_k^2, \forall k \in \mathbb{N}. \end{cases}$$

Then, for all $k \in \mathbb{N}$,

$$\frac{1}{\alpha_k} - \frac{L}{2}\left(\frac{1}{1-Ls_k} - \frac{s_k}{\alpha_k}\right)^2 \geq C := \frac{L^3 s_-^2(1+Ls_-)}{2(2-L^2s_-^2)^2(1-Ls_-)} > 0.$$

Proof. First, one can check that

$$Ls_k \leq 2 - 2Ls_k - (1-Ls_k)L^2s_k^2$$

if and only if

$$Ls_k \in \left[-\frac{\sqrt{5}+1}{2}, \frac{\sqrt{5}-1}{2}\right] \cup [2, +\infty),$$

and the bound $s_k \leq \frac{\sqrt{5}-1}{2}$ is a necessary condition which ensures that there exists α_k which satisfies $s_k \leq \alpha_k \leq \frac{2}{L} - 2s_k - (1 - Ls_k)Ls_k^2$. We now turn to the lower bound under condition **(C3)**. Set

$$\alpha_k^+ = \frac{2}{L} - 2s_k - (1 - Ls_k)Ls_k^2$$
$$Q(u) = u - \frac{1}{2}\left(\frac{1}{1 - Ls_k} - Ls_k u\right)^2,$$

where one can think of u satisfying $u = \frac{1}{L\alpha_k} \in \left[\frac{1}{L\alpha_k^+}, \frac{1}{Ls_k}\right]$. The maximum of $Q(u)$ is attained for $u = \frac{1}{(1-Ls_k)L^2s_k^2} \geq \frac{1}{Ls_k}$ and hence Q is increasing on $\left[\frac{1}{L\alpha_k^+}, \frac{1}{Ls_k}\right]$, and therefore, for all $k \in \mathbb{N}$,

$$\begin{aligned}
\frac{1}{\alpha_k} - \frac{L}{2}\left(\frac{1}{1 - Ls_k} - \frac{s_k}{\alpha_k}\right)^2 &= LQ\left(\frac{1}{L\alpha_k}\right) \geq LQ\left(\frac{1}{L\alpha_k^+}\right) \\
&= L\frac{-(L\alpha_k^+)^2 + 2L\alpha_k^+(1 - Ls_k) - L^2s_k^2(1 - Ls_k)^2}{2(L\alpha_k^+)^2(1 - Ls_k)^2} \\
&= L\frac{-(1 - Ls_k)^2(2 - L^2s_k^2)^2 + 2(1 - Ls_k)^2(2 - L^2s_k^2) - L^2s_k^2(1 - Ls_k)^2}{2(2 - L^2s_k^2)^2(1 - Ls_k)^4} \\
&= L\frac{-(2 - L^2s_k^2)^2 + 2(2 - L^2s_k^2) - L^2s_k^2}{2(2 - L^2s_k^2)^2(1 - Ls_k)^2} \\
&= \frac{L^3s_k^2(1 + Ls_k)}{2(2 - L^2s_k^2)^2(1 - Ls_k)} \geq \frac{L^3s_-^2(1 + Ls_-)}{2(2 - L^2s_-^2)^2(1 - Ls_-)} > 0,
\end{aligned}$$

which is the desired result.

□

We can check that, when condition **(C3)** is satisfied, one has

$$0 < b_k = \frac{L\alpha_k + (1 - Ls_k)^2}{\alpha_k(1 - Ls_k)} \leq B = \frac{6}{\alpha_-}$$

Combining this with Proposition 3.3.3, Proposition 3.3.10 and Lemma 3.3.11, we obtain the following corollary.

Corollary 3.3.12. *Suppose that $(s_k, \alpha_k)_{k\in\mathbb{N}}$ satisfy condition* **(C3)** *and that f is convex, then*

i) $F(x_{k+1}) + C\|x_k - x_{k+1}\|^2 \leq F(x_k)$, $\forall k \in \mathbb{N}$.

ii) *There exists $\omega_{k+1} \in \partial F(x_{k+1})$ such that*

$$\|\omega_{k+1}\| \leq B\|x_k - x_{k+1}\|, \forall k \in \mathbb{N}.$$

where C is given in Lemma 3.3.11 and $B = \frac{6}{\alpha_-}$.

We now consider the complexity for **(EEG)** method under nonsmooth KL inequality. By applying the result of Theorem 2.4.5, we have the complexity of **(EEG)** method in the form of a *small prox* result.

Theorem 3.3.13 (Complexity of **EEG** method)**.** *Assume that* $(s_k, \alpha_k)_{k\in\mathbb{N}}$ *satisfy condition* **(C3)** *and* f *is convex. Then, the sequence* $(x_k)_{k\in\mathbb{N}}$ *converges to* $x^* \in \operatorname{argmin} F$*, and*

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

moreover,

$$F(x_k) - F^* \leq \psi(\beta_k), \quad \forall k \geq 0,$$
$$\|x_k - x^*\| \leq \frac{B}{C}\beta_k + \sqrt{\frac{\psi(\beta_{k-1})}{C}}, \quad \forall k \geq 1,$$

where B *and* C *are given in Corollary 3.3.12.*

3.4 Numerical experiment

In this section, we compare the extragradient method with standard algorithms in numerical optimization: forward-backward and FISTA. We describe the problem of interest, details about exact line search in this context and numerical results.

3.4.1 ℓ^1 regularized least squares

We let $A \in \mathbb{R}^{p\times n}$ be a real matrix, $b \in \mathbb{R}^n$ be a real vector and $\lambda > 0$ be a scalar, all of them given and fxed. Following the notations of the previous section, we define $f\colon x \mapsto \frac{1}{2}\|Ax-b\|_2^2$ and $g\colon x \mapsto \lambda\|x\|_1$ (the sum of absolute values of the entries). With these notations, the optimization problem (**P**) becomes

$$\min_{x\in\mathbb{R}^n} \frac{1}{2}\|Ax - b\|_2^2 + \lambda\|x\|_1. \tag{3.19}$$

Solutions of problem of the form of (3.19) (as well as many extensions) are extensively used in statistics and signal processing [116, 36]. For this problem, we introduce the proximal gradient mapping, a specialization of the proximal gradient step to problem (3.19). This is the main building block of all the algorithms presented in the numerical experiment.

$$\begin{array}{rccc} p\colon & \mathbb{R}^n \times \mathbb{R}_+ & \mapsto & \mathbb{R}^n \\ & (x, s) & \mapsto & S_{s\lambda}(x - s\nabla f(x)) \end{array} \tag{3.20}$$

where S_a ($a \in \mathbb{R}_+$) is the soft-thresholding operator which acts coordinate-wise and satisfies for $i = 1, 2 \ldots, n$

$$[S_a(x)]_i = \begin{cases} 0, & \text{if } |x_i| \leq a \\ x_i - a\operatorname{sign}(x_i), & \text{otherwise.} \end{cases}$$

3.4.2 Exact line search

One intuition behind Extragradient-Method for optimization is the use of an additional iteration as a guide or a scout to provide an estimate of the gradient that better suits the geometry of the problem. This should eventually translate in the possibility of taking larger steps leading to faster convergence. In this section we briefly describe a strategy which allows to perform exact

line search in the context of ℓ^1-regularized least squares. Up to our knowledge, this was not decribed before in the literature. Furthermore, this strategy may be extended to more general least squares problems with nonsmooth regularizes. For the rest of this section, we assume that $x \in \mathbb{R}^n$ is fixed. We heavily rely on the two simple facts:

- The mapping $s \to p(x, s)$ is continuous and piecewise affine.
- The objective function $x \mapsto f(x) + g(x)$ is continuous and piecewise quadratic.

We consider the following function

$$\begin{aligned} q_x \colon \mathbb{R}_+ &\to \mathbb{R} \\ \alpha &\to f(p(x, \alpha)) + g(p(x, \alpha)). \end{aligned}$$

It can be deduced from the properties of f, g and p that q_x is continuous and piecewise quadratic. In classical implementation of proximal splitting methods, the step-size parameter α is a well chosen constant which depends on the problem, or alternatively it is estimated using backtracking. The alternative which we propose is to choose the step-size parameter α minimizing q_x. Since q_x is piecewise quadratic, this only requires to know the values of α for which q_x is not differentiable and the expression of q_x as a quadratic form between these values.

The nonsmooth points of q_x are given by the following set

$$\mathcal{D}_x = \left\{ \frac{x_i}{\frac{\partial f(x)}{\partial x_i} - \lambda}, \frac{x_i}{\frac{\partial f(x)}{\partial x_i} + \lambda} \right\}_{i=1}^{n} \cap \mathbb{R}_+$$

and correspond to limiting values for which coordinates of $p(x, \alpha)$ are null. We assume that the elements of $\mathcal{D}_x$ are ordered nondecreasingly (letting potential ties appear several times). The comments that we have made so far lead to the following.

- $\mathcal{D}_x$ contains no more than $2n$ elements.
- Given x and λ, computing $\mathcal{D}_x$ is as costly as computing ∇f.
- q_x is quadratic between two consecutive elements of $\mathcal{D}_x$.

In order to minimize q_x, the only task that should be performed is to keep track of its value (or equivalently of its quadratic expression) between consecutive elements of $\mathcal{D}_x$. Here, we can use the fact that elements of $\mathcal{D}_x$ corresponds to values of α for which one coordinate of $p(x, \alpha)$ goes to zero or becomes active (non-zero). A careful implementation of the minimization of q_x amounts to sort the values in $\mathcal{D}_x$, screen them in increasing order, keep track of the corresponding quadratic expression and the minimal value. We provide a few details for completeness.

- The vector $d_x(s) := \left(\frac{\partial [p(x,s)]_i}{\partial s} \right)_{i=1}^{n} \in \mathbb{R}^n$ is constant between consecutive elements of $\mathcal{D}_x$. Furthermore the elements of $\mathcal{D}_x$ (counted with multiple ties) corresponds to value of α for which a single coordinate of $d_x(s)$ is modified.
- Suppose that $\alpha_1 < \alpha_2$ are two consecutive elements of $\mathcal{D}_x$. Then for all $\alpha \in [\alpha_1, \alpha_2]$, letting

$d_x(\alpha) = d$ on this segment, we have $p(x, \alpha) = p(x, \alpha_1) + (\alpha - \alpha_1)d$, hence,

$$\begin{aligned}
&\frac{1}{2}\|Ap(x,\alpha) - b\|_2^2 + \lambda\|p(x,\alpha)\|_1 \\
&= \frac{1}{2}\|Ap(x,\alpha_1) - b\|_2^2 + \lambda\|p(x,\alpha_1)\|_1 \\
&\quad + \frac{(\alpha - \alpha_1)}{n}\langle Ad, Ax - b\rangle + \frac{(\alpha - \alpha_1)^2}{2n}\|Ad\|_2^2 + \lambda(\alpha - \alpha_1)\left\langle \bar{d}, d\right\rangle,
\end{aligned}$$

where $\bar{d} \in \mathbb{R}^p$ is a vector which depends on the sign pattern of $p(x, \alpha_1)$ and d.

- For $\alpha = \alpha_2$, the sign pattern of $p(x, \alpha_2)$ and the corresponding value of d and $\bar{d}$ (for the next interval) are modified only at a single coordinate, the same for the three of them. In other words, updating the quadratic expression of q_x at α_2 only requires the knowledge of this coordinate, the value of the corresponding column in A and can be done by computing inner products in $\mathbb{R}^p$. This requires $O(p)$ operations.
- Given these properties, we can perform minimization of q_x by an active set strategy, keeping track only of the sign pattern of $p(x, \alpha)$, the value of $\left\langle \bar{d}, d\right\rangle$, the value of Ad, $Ap(x, \alpha) - b$ and $\|p(x, \alpha)\|_1$ which cost is of the order of $O(p)$. This should not be repeated more than $2n$ times.

Using this active set procedure provides the quadratic expression of q_x for all intervals represented by consecutive values in $\mathcal{D}_x$. From these expressions, it is not difficult to compute the global minimum of q_x. The overall cost of this operation when properly implemented is of the order of $O(np)$ plus the cost of sorting $2n$ elements in $\mathbb{R}$. This is comparable to the cost of computing the gradient of f. Hence in this specific setting, performing exact line search does not add much overhead in term of computational cost compared to existing step-size strategies.

3.4.3 Simulation and results

We generate a matrix A and vector b using the following process.

- Set $n = 400$ and $p = 1600$
- Set $A = DX$ where X has standard Gaussian independent entries and D is a diagonal matrix which i-th diagonal entry is $\frac{1}{i^2}$.
- Choose x_0 in $\mathbb{R}^n$ with the 400 first entries being independent standard Gaussian and the remaining ones are null.
- Set $b = Ax + z$ where $z \in \mathbb{R}^p$ has independant Gaussian entries.
- We choose $\lambda = 0.001$

We compare the following algorithms (L is the Lipschitz constant of f computed from the singular value of A).

- Forward-backward with step-size parameter $1/L$ (see for example [17]).
- Forward-backward with step-size parameter determined by exact line search.

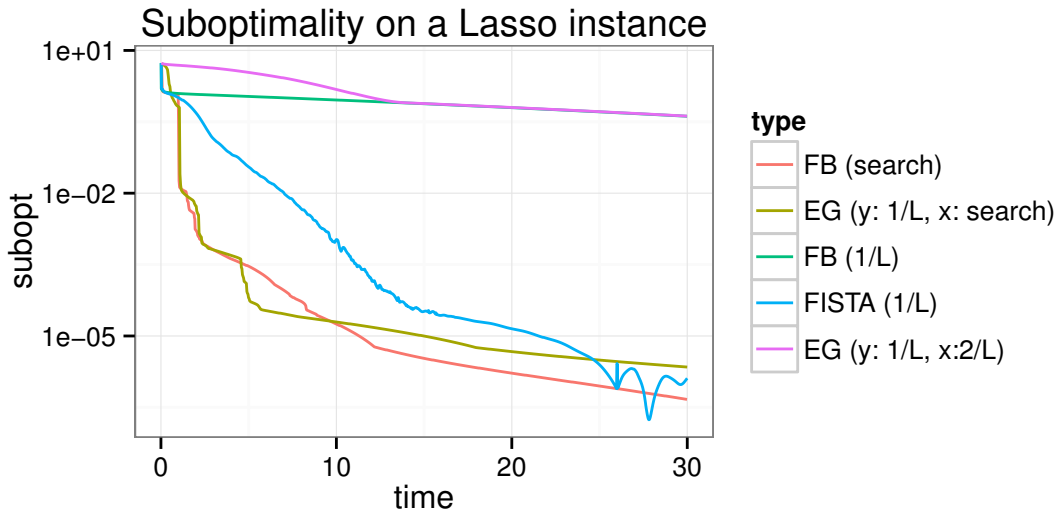

Figure 3.1: Suboptimality $(F(x_k) - F^*)$ as a function of time for simulated ℓ^1 regularized least squares data. FB stands for forward-backward and EG for extragradient. The indications in parenthesis indicate the step sizes strategy that is used.

- Extragradient (discussed in the present paper) with step size parameter $s = 1/L$ and $\alpha = 2/L$.
- Extragradient with step size parameter $s = 1/L$ and α determined by exact line search.
- FISTA as described in [17].

The exact line search active set procedure is implemented in compiled `C` code in order keep a reasonable level of efficiency compared to linear algebra operations which have efficient implementations. All algorithms are initialized at the same point. We keep track of decrease of the objective value, the iteration counter k and the total spent since initialization. The iteration counter is related to analytical complexity while the total time spent is related to the arithmetical complexity (see the introduction in [99] for more details). Comparing algorithms in term of analytical complexity does not reflect the fact that iterations are more costly for some of them compared to others.

Computational times are presented in Figure 3.1 and iteration counters in Figure 3.2. The main comment is that the exact line search procedure improves upon fixed step size parameters while the induced computational overhead remains reasonable. Indeed, both forward-backward and extragradient, when implemented using the exact line search procedure produce results which are comparable to FISTA, a reference in terms of performances for this composite problem. On the other hand, there is not much difference between forward-backward and extragradient, neither for fixed step sizes, nor for exact line search.

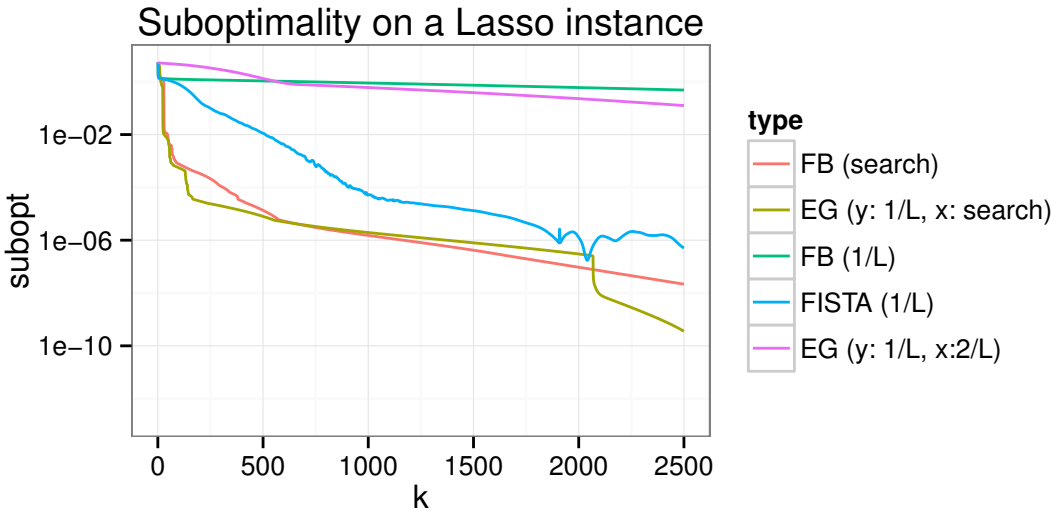

Figure 3.2: Suboptimality $(F(x_k) - F^*)$ as a function of the iteration counter for simulated ℓ^1 regularized least squares data. FB stands for forward-backward and EG for extragradient. The indications in parenthesis indicate the step sizes strategy that is used.

Bibliography

[1] Pierre Antoine Absil, Robert Mahony, and Benjamin Andrews. Convergence of the iterates of descent methods for analytic cost functions. *SIAM Journal on Optimization*, 16(2):531–547, 2005.

[2] Hedy Attouch and Jérôme Bolte. On the convergence of the proximal algorithm for nonsmooth functions involving analytic features. *Mathematical Programming*, 116(1):5–16, 2009.

[3] Hedy Attouch, Jérôme Bolte, Patrick Redont, and Antoine Soubeyran. Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the Kurdyka– Lojasiewicz inequality. *Mathematics of Operations Research*, 35(2):438–457, 2010.

[4] Hedy Attouch, Jérôme Bolte, and Benar Fux Svaiter. Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward–backward splitting, and regularized Gauss–Seidel methods. *Mathematical Programming*, 137(1-2):91–129, 2013.

[5] Alfred Auslender. *Méthodes numériques pour la résolution des problèmes d'optimisation avec contraintes*. Faculté des sciences, Université de Grenoble, 1969.

[6] Alfred Auslender and Jean Pierre Crouzeix. Global regularity theorems. *Mathematics of Operations Research*, 13(2):243–253, 1988.

[7] Dominique Azé. A survey on error bounds for lower semicontinuous functions. In *ESAIM: Proceedings*, volume 13, pages 1–17. EDP Sciences, 2003.

[8] Dominique Azé and Jean Noël Corvellec. On the sensitivity analysis of Hoffman constants for systems of linear inequalities. *SIAM Journal on Optimization*, 12(4):913–927, 2002.

[9] Dominique Azé and Jean Noël Corvellec. Characterizations of error bounds for lower semicontinuous functions on metric spaces. *ESAIM: Control, Optimisation and Calculus of Variations*, 10(3):409–425, 2004.

[10] Dominique Azé and Jean Noël Corvellec. Nonlinear local error bounds via a change of metric. *Journal of Fixed Point Theory and Applications*, 16(1-2):351–372, 2014.

[11] Dominique Azé and Jean Noël Corvellec. Nonlinear error bounds via a change of function. *Journal of Optimization Theory and Applications*, pages 1–24, 2016.

[12] Jean Bernard Baillon, Patrick Louis Combettes, and Roberto Cominetti. There is no variational characterization of the cycles in the method of periodic projections. *Journal of Functional Analysis*, 262(1):400–408, 2012.

[13] Heinz Bauschke and Jonathan Borwein. On projection algorithms for solving convex feasibility problems. *SIAM review*, 38(3):367–426, 1996.

[14] Heinz Bauschke and Patrick Louis Combettes. *Convex analysis and monotone operator theory in Hilbert spaces*. Springer Science & Business Media, 2011.

[15] Amir Beck and Shimrit Shtern. Linearly convergent away-step conditional gradient for non-strongly convex functions. *Mathematical Programming*, pages 1–27, 2015.

[16] Amir Beck and Marc Teboulle. Convergence rate analysis and error bounds for projection algorithms in convex feasibility problems. *Optimization Methods and Software*, 18(4):377–394, 2003.

[17] Amir Beck and Marc Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM journal on imaging sciences*, 2(1):183–202, 2009.

[18] Pascal Bégout, Jérôme Bolte, and Mohamed Ali Jendoubi. On damped second-order gradient systems. *Journal of Differential Equations*, 259(7):3115–3143, 2015.

[19] Jacek Bochnak, Michel Coste, and Marie Françoise Roy. *Real algebraic geometry*, volume 36. Springer Science & Business Media, 2013.

[20] Jérôme Bolte. Sur quelques principes de convergence en optimisation. *Habilitationa diriger des recherches, Université Pierre et Marie Curie*, 2008.

[21] Jérôme Bolte, Aris Daniilidis, and Adrian Lewis. The Łojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems. *SIAM Journal on Optimization*, 17(4):1205–1223, 2007.

[22] Jérôme Bolte, Aris Daniilidis, Adrian Lewis, and Masahiro Shiota. Clarke subgradients of stratifiable functions. *SIAM Journal on Optimization*, 18(2):556–572, 2007.

[23] Jérôme Bolte, Aris Daniilidis, Olivier Ley, and Laurent Mazet. Characterizations of Łojasiewicz inequalities: subgradient flows, talweg, convexity. *Transactions of the American Mathematical Society*, 362(6):3319–3363, 2010.

[24] Jérôme Bolte, Trong Phong Nguyen, Juan Peypouquet, and Bruce Suter. From error bounds to the complexity of first-order descent methods for convex functions. *Mathematical Programming*, pages 1–37, 2015.

[25] Jérôme Bolte and Edouard Pauwels. Majorization-minimization procedures and convergence of SQP methods for semi-algebraic and tame programs. *Mathematics of Operations Research*, 41(2):442–465, 2016.

[26] Jérôme Bolte, Shoham Sabach, and Marc Teboulle. Proximal alternating linearized minimization for nonconvex and nonsmooth problems. *Mathematical Programming*, 146(1-2):459–494, 2014.

[27] Jonathan Borwein, Guoyin Li, and Liangjin Yao. Analysis of the convergence rate for the cyclic projection algorithm applied to basic semialgebraic convex sets. *SIAM Journal on Optimization*, 24(1):498–527, 2014.

[28] Haim Brézis. *Operateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert*, volume 5. Elsevier, 1973.

[29] Haim Brézis and Pierre Louis Lions. Produits infinis de résolvantes. *Israel Journal of Mathematics*, 29(4):329–345, 1978.

[30] Pierre Brousse. *Optimization in mechanics: problems and methods*, volume 34. Elsevier, 2013.

[31] Ronald Bruck. Asymptotic convergence of nonlinear contraction semigroups in hilbert space. *Journal of Functional Analysis*, 18(1):15–26, 1975.

[32] Arthur Earl Bryson. *Applied optimal control: optimization, estimation and control*. CRC Press, 1975.

[33] Emmanuel Candès and Michael Wakin. An introduction to compressive sampling. *IEEE signal processing magazine*, 25(2):21–30, 2008.

[34] Augustin Cauchy. Méthode générale pour la résolution des systemes d'équations simultanées. *Comp. Rend. Sci. Paris*, 25(1847):536–538, 1847.

[35] Yair Censor, Aviv Gibali, and Simeon Reich. The subgradient extragradient method for solving variational inequalities in Hilbert space. *Journal of Optimization Theory and Applications*, 148(2):318–335, 2011.

[36] Scott Shaobing Chen, David Donoho, and Michael Saunders. Atomic decomposition by basis pursuit. *SIAM review*, 43(1):129–159, 2001.

[37] Patrick Louis Combettes. Inconsistent signal feasibility problems: Least squares solutions in a product space. *IEEE Transactions on Signal Processing*, 42(11):2955–2966, 1994.

[38] Patrick Louis Combettes and Jean Christophe Pesquet. Proximal splitting methods in signal processing. In *Fixed-point algorithms for inverse problems in science and engineering*, pages 185–212. Springer, 2011.

[39] Patrick Louis Combettes and Valérie Wajs. Signal recovery by proximal forward-backward splitting. *Multiscale Modeling & Simulation*, 4(4):1168–1200, 2005.

[40] Octavio Cornejo, Abderrahim Jourani, and Constantin Z̲alinescu. Conditioning and upper-lipschitz inverse subdifferentials in nonsmooth optimization problems. *Journal of optimization theory and applications*, 95(1):127–148, 1997.

[41] Jean Noël Corvellec and Viorica Motreanu. Nonlinear error bounds for lower semicontinuous functions on metric spaces. *Mathematical Programming*, 114(2):291, 2008.

[42] Michel Coste. *An introduction to o-minimal geometry*. Istituti editoriali e poligrafici internazionali Pisa, 2000.

[43] Didier D'Acunto and Krzysztof Kurdyka. Explicit bounds for the Łojasiewicz exponent in the gradient inequality for polynomials. *Annales Polonici Mathematici*, 87(1):51–61, 2005.

[44] Ingrid Daubechies, Michel Defrise, and Christine De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Communications on pure and applied mathematics*, 57(11):1413–1457, 2004.

[45] Jean Pierre Dedieu. Penalty functions in subanalytic optimization. *Optimization*, 26(1-2):27–32, 1992.

[46] Jean Pierre Dedieu. Approximate solutions of analytic inequality systems. *SIAM Journal on Optimization*, 11(2):411–425, 2000.

[47] Sien Deng. Computable error bounds for convex inequality systems in reflexive banach spaces. *SIAM Journal on Optimization*, 7(1):274–279, 1997.

[48] Sien Deng. Global error bounds for convex inequality systems in Banach spaces. *SIAM journal on control and optimization*, 36(4):1240–1249, 1998.

[49] Sien Deng. Perturbation analysis of a condition number for convex inequality systems and global error bounds for analytic systems. *Math. Program.*, 83:263–276, 1998.

[50] Si Tiep Dinh, Huy Vui Ha, and Tien Son Pham. Hölder-type global error bounds for non-degenerate polynomial systems. *arXiv preprint arXiv:1411.0859*, 2014.

[51] Yoel Drori. *Contributions to the complexity analysis of optimization algorithms*. PhD thesis, Tel-Aviv University, 2014.

[52] Dmitriy Drusvyatskiy and Adrian Lewis. Error bounds, quadratic growth, and linear convergence of proximal methods. *arXiv preprint arXiv:1602.06661*, 2016.

[53] Ivar Ekeland et al. Nonconvex minimization problems. *Bull. AMS*, 1(3), 1979.

[54] Francisco Facchinei and Jong Shi Pang. *Finite-dimensional variational inequalities and complementarity problems*. Springer Science & Business Media, 2007.

[55] Christodoulos Floudas and Panos Pardalos. *Optimization in computational chemistry and molecular biology: local and global approaches*, volume 40. Springer Science & Business Media, 2013.

[56] Pierre Frankel, Guillaume Garrigos, and Juan Peypouquet. Splitting methods with variable metric for kl functions. *arXiv preprint arXiv:1405.1357*, 2014.

[57] Alan Goldstein. Convex programming in Hilbert space. *Bulletin of the American Mathematical Society*, 70(5):709–710, 1964.

[58] Janusz Gwoździewicz. The Łojasiewicz exponent of an analytic function at an isolated zero. *Commentarii Mathematici Helvetici*, 74(3):364–375, 1999.

[59] Huy Vui Ha. Global Hölderian error bound for nondegenerate polynomials. *SIAM Journal on Optimization*, 23(2):917–933, 2013.

[60] Huy Vui Ha and Tien Son Pham. *Genericity in Polynomial Optimization*, volume 3. World Scientific, 2016.

[61] Alexander Hartmann and Heiko Rieger. *Optimization algorithms in physics*, volume 2. Citeseer, 2002.

[62] Heisuke Hironaka. *Introduction to real-analytic sets and real-analytic maps*. Istituto matematico" L. Tonelli" dell Universita di Pisa, 1973.

[63] Alan Hoffman. On approximate solutions of systems of linear inequalities. *Selected Papers Of Alan J Hoffman: With Commentary*, pages 174–176, 2003.

[64] Lars Hörmander. On the division of distributions by polynomials. *Arkiv för matematik*, 3(6):555–568, 1958.

[65] Aleksandr Davidovich Ioffe. Metric regularity and subdifferential calculus. *Russian Mathematical Surveys*, 55(3):501–558, 2000.

[66] Alexander Davidovich Ioffe. Regular points of lipschitz functions. *Transactions of the American Mathematical Society*, 251:61–69, 1979.

[67] Morton Kamien and Nancy Lou Schwartz. *Dynamic optimization: the calculus of variations and optimal control in economics and management*. Courier Corporation, 2012.

[68] Leonid Kantorovich and Gleb Akilov. Functional analysis in normed spaces, volume 46 of international series of monographs in pure and applied mathematics, 1964.

[69] Diethard Klatte and Wu Li. Asymptotic constraint qualifications and global error bounds for convex inequalities. *Mathematical programming*, 84(1):137–160, 1999.

[70] János Kollár. An effective Łojasiewicz inequality for real polynomials. *Periodica Mathematica Hungarica*, 38(3):213–221, 1999.

[71] GM Korpelevich. The extragradient method for finding saddle points and other problems. *Matecon*, 12:747–756, 1976.

[72] Alexander Kruger. Error bounds and metric subregularity. *Optimization*, 64(1):49–79, 2015.

[73] Krzysztof Kurdyka. On gradients of functions definable in o-minimal structures. *Annales de l'institut Fourier*, 48(3):769–783, 1998.

[74] Krzysztof Kurdyka and Stanisław Spodzieja. Separation of real algebraic sets and the Łojasiewicz exponent. *Proceedings of the American Mathematical Society*, 142(9):3089–3102, 2014.

[75] Evgeny Levitin and Polyak Boris. Constrained minimization methods. *USSR Computational mathematics and mathematical physics*, 6(5):1–50, 1966.

[76] Adrian Lewis and Jong Shi Pang. Error bounds for convex inequality systems. In *Generalized convexity, generalized monotonicity: recent results*, pages 75–110. Springer, 1998.

[77] Guoyin Li. On the asymptotically well behaved functions and global error bound for convex polynomials. *SIAM Journal on Optimization*, 20(4):1923–1943, 2010.

[78] Guoyin Li. Global error bounds for piecewise convex polynomials. *Mathematical Programming*, pages 1–28, 2013.

[79] Guoyin Li, Boris Mordukhovich, TTA Nghia, and Tien Son Pham. Error bounds for parametric polynomial systems with applications to higher-order stability analysis and convergence rates. *Mathematical Programming*, pages 1–34, 2015.

[80] Guoyin Li, Boris Mordukhovich, and Tien Son Pham. New fractional error bounds for polynomial systems with applications to Hölderian stability in optimization and spectral theory of tensors. *Mathematical Programming*, 153(2):333–362, 2015.

[81] Wu Li. Error bounds for piecewise convex quadratic programs and applications. *SIAM Journal on Control and Optimization*, 33(5):1510–1529, 1995.

[82] Wu Li. Abadie's constraint qualification, metric regularity, and error bounds for differentiable convex inequalities. *SIAM Journal on Optimization*, 7(4):966–978, 1997.

[83] Jingwei Liang, Jalal Fadili, and Gabriel Peyré. Local linear convergence of forward–backward under partial smoothness. In *Advances in Neural Information Processing Systems*, pages 1970–1978, 2014.

[84] Stanisław Łojasiewicz. Division d'une distribution par une fonction analytique de variables reelles. *Comptes Rendus Hebdomadaires des Seances de L'Academie des Sciences*, 246(5):683–686, 1958.

[85] Stanisław Łojasiewicz. Sur le probleme de la division. *Studia Mathematica*, 18, 1959.

[86] Stanisław Łojasiewicz. Une propriété topologique des sous-ensembles analytiques réels. *Les équations aux dérivées partielles*, 117:87–89, 1963.

[87] Xiao Dong Luo and Zhi Quan Luo. Extension of Hoffman's error bound to polynomial systems. *SIAM Journal on Optimization*, 4(2):383–392, 1994.

[88] Zhi Quan Luo and Jong Shi Pang. Error bounds for analytic systems and their applications. *Mathematical Programming*, 67(1-3):1–28, 1994.

[89] Zhi Quan Luo and Jos Sturm. Error bounds for quadratic systems. In *High performance optimization*, pages 383–404. Springer, 2000.

[90] Zhi Quan Luo and Paul Tseng. Error bound and convergence analysis of matrix splitting algorithms for the affine variational inequality problem. *SIAM Journal on Optimization*, 2(1):43–54, 1992.

[91] Zhi Quan Luo and Paul Tseng. On the linear convergence of descent methods for convex essentially smooth minimization. *SIAM Journal on Control and Optimization*, 30(2):408–425, 1992.

[92] Zhi Quan Luo and Paul Tseng. Error bounds and convergence analysis of feasible descent methods: a general approach. *Annals of Operations Research*, 46(1):157–178, 1993.

[93] Olvi Mangasarian. A condition number for differentiable convex inequalities. *Mathematics of Operations Research*, 10(2):175–179, 1985.

[94] Bernard Martinet. Brève communication. régularisation d'inéquations variationnelles par approximations successives. *Revue française d'informatique et de recherche opérationnelle, série rouge*, 4(3):154–158, 1970.

[95] Renato Monteiro and Benar Svaiter. Complexity of variants of Tseng's modified forward–backward splitting and Korpelevich's methods for hemivariational inequalities with applications to saddle-point and convex optimization problems. *SIAM Journal on Optimization*, 21(4):1688–1720, 2011.

[96] Boris Mordukhovich. *Variational analysis and generalized differentiation I: Basic theory*, volume 330. Springer Science & Business Media, 2006.

[97] Angelia Nedić and Dimitri Bertsekas. Convergence rate of incremental subgradient algorithms. In *Stochastic optimization: algorithms and applications*, pages 223–264. Springer, 2001.

[98] Yurii Nesterov. A method of solving a convex programming problem with convergence rate o (1/k2). *Soviet Mathematics Doklady*, 27(2):372–376, 1983.

[99] Yurii Nesterov. *Introductory lectures on convex optimization: A basic course*, volume 87. Springer Science & Business Media, 2013.

[100] Kung Fu Ng and Xi Yin Zheng. Global error bounds with fractional exponents. *Mathematical programming*, 88(2):357–370, 2000.

[101] Kung Fu Ng and Xi Yin Zheng. Error bounds for lower semicontinuous functions in normed spaces. *SIAM Journal on Optimization*, 12(1):1–17, 2001.

[102] Huynh Van Ngai. Global error bounds for systems of convex polynomials over polyhedral constraints. *SIAM Journal on Optimization*, 25(1):521–539, 2015.

[103] Huynh van Ngai and Michel Théra. Error bounds for convex differentiable inequality systems in Banach spaces. *Mathematical programming*, 104(2):465–482, 2005.

[104] Trong Phong Nguyen. A stroll in the jungle of error bounds. *preprint arXiv:1704.06938*, 2017.

[105] Trong Phong Nguyen, Edouard Pauwels, Emile Richard, and Bruce W Suter. Extragradient method in optimization: Convergence and complexity. *preprint arXiv:1609.08177*, 2016.

[106] Jong Shi Pang. Error bounds in mathematical programming. *Mathematical Programming*, 79(1-3):299–332, 1997.

[107] Edouard Pauwels. The value function approach to convergence analysis in composite optimization. *Operations Research Letters*, 44(6):790–795, 2016.

[108] Juan Peypouquet. Asymptotic convergence to the optimal value of diagonal proximal iterations in convex minimization. *J. Convex Anal*, 16(1):277–286, 2009.

[109] Juan Peypouquet. *Convex optimization in normed spaces: theory, methods and examples*. Springer, 2015.

[110] Tien Son Pham. The Łojasiewicz exponent of a continuous subanalytic function at an isolated zero. *Proceedings of the American Mathematical Society*, 139(1):1–9, 2011.

[111] Tien Son Pham. An explicit bound for the Łojasiewicz exponent of real polynomials. *Kodai Mathematical Journal*, 35(2):311–319, 2012.

[112] Boris Teodorovich Polyak. Gradient methods for minimizing functionals. *Zhurnal Vychislitel'noi Matematiki i Matematicheskoi Fiziki*, 3(4):643–653, 1963.

[113] Stephen Robinson. An application of error bounds for convex programming in a linear space. *SIAM Journal on Control*, 13(2):271–273, 1975.

[114] Ralph Tyrell Rockafellar. *Convex analysis*. Princeton university press, 1972.

[115] Ralph Tyrrell Rockafellar. Monotone operators and the proximal point algorithm. *SIAM journal on control and optimization*, 14(5):877–898, 1976.

[116] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996.

[117] Paul Tseng. On linear convergence of iterative methods for the variational inequality problem. *Journal of Computational and Applied Mathematics*, 60(1-2):237–252, 1995.

[118] Paul Tseng and Sangwoon Yun. A coordinate gradient descent method for nonsmooth separable minimization. *Mathematical Programming*, 117(1):387–423, 2009.

[119] Huynh Van Ngai and Michel Théra. Error bounds for systems of lower semicontinuous functions in Asplund spaces. *Mathematical Programming*, 116(1):397–427, 2009.

[120] Po Wei Wang and Chih Jen Lin. Iteration complexity of feasible descent methods for convex optimization. *Journal of Machine Learning Research*, 15(1):1523–1548, 2014.

[121] Tao Wang and Jong Shi Pang. Global error bounds for convex quadratic inequality systems. *Optimization*, 31(1):1–12, 1994.

[122] Zili Wu and Jane Ye. On error bounds for lower semicontinuous functions. *Mathematical programming*, 92(2):301–314, 2002.

[123] Zili Wu and Jane Ye. Sufficient conditions for error bounds. *SIAM Journal on Optimization*, 12(2):421–435, 2002.

[124] Wein Hong Yang. Error bounds for convex polynomials. *SIAM Journal on Optimization*, 19(4):1633–1647, 2009.

[125] Constantin Zualinescu. Sharp estimates for Hoffman's constant for systems of linear inequalities and equalities. *SIAM Journal on Optimization*, 14(2):517–533, 2003.