Proximal point method on Finslerian manifolds and the "Effort-Accuracy" Trade off

J.X. da Cruz Neto* P. R. Oliveira† P.A. Soares Jr ‡ A. Soubeyran §

July 5, 2011

Abstract

In this paper we consider minimization problems with constraints. We will show that if the set of constraints is a Finslerian manifold of non positive flag curvature, and the objective function is differentiable and satisfies the property Kurdyka-Lojasiewicz, then the proximal point method is naturally extended to solve that class of problems. We will prove that the sequence generated by our method is well defined and converges to a minimizer point. We show how tools of Finslerian geometry, more specifically non symmetrical metrics, can be used to solve nonconvex constrained problems in Euclidean spaces. As an application, we give one result on the speed of decision and making and costs to change.

Keywords Proximal algorithms; Finslerian manifolds; nonconvex optimization; Kurdyka-Lojasiewicz inequality.

1 Introduction

The "effort-accuracy" trade off. In Behavioral sciences it is well known that optimizing is costly. But how much is the problem? In decision making theory and more generally in cognitive sciences, the famous "effort-accuracy " and "speed-accuracy" trade off (see Payne et al. [1] and Rinkenauer [2]) examine how animals and agents balance between the quality of a decision, how "good enough" it must be, and the time, speed and effort spend to find such a "good enough" decision, some solution to a problem. More generally, biology and behavioral sciences have emphasized the proeminent role of the famous

*Department of Mathematics, Federal University of Piauí, Teresina, Brazil
e-mail: jxavier@ufpi.br

†PESC/COPPE - Programa de Engenharia de Sistemas e Computação, Rio de Janeiro, Brazil
e-mail: poliveir@cos.ufrj.br

‡Partially supported by CAPES. PESC/COPPE - Programa de Engenharia de Sistemas e Computação, Rio de Janeiro, Brazil e-mail:pedrosoares@cos.ufrj.br

§GREQAM-SPORT, Université d'Aix-Marseille II, France
e-mail: antoine.soubeyran@univmed.fr

"exploration- exploitation" trade off, Holland [3] in biology, March [4] in management sciences, and their followers in organizational learning (Levitt-March, [5], Cohen-Levinthal [6], Levinthal-March [7], and an impressive list of others), Fu-Gray [8] and Fu [9], in Psychology...

March [4] says: "A central concern of studies of adaptive processes is the relation between the exploration of new possibilities and the exploitation of old certainties..... Exploration includes things captured by terms such as search, variation, risk taking, experimentation, play, flexibility, discovery, innovation. Exploitation includes such things as refinement, choice, production, efficiency, selection, implementation, execution". The problem is that exploration (search) and exploitation (of the benefits to search) use very often competing resources: the more time and resources the agent spends for exploration, the less he has for exploitation.

The "effort-accuracy" and "speed-accuracy " trade off (SAT) are specific examples of the general "exploration-exploitation" trade off, due effort (resources spend to explore, like mental energy, thinking, deliberation costs, time compression costs, reactivity costs, ...) and accuracy (how efficient is a solution) are specific instances of exploration and exploitation activities.

In Psychology, a lot of empirical studies have been made to examine the "effort-accuracy " trade off. Busemeyer-Townsend [10] offered a dynamic-cognitive approach to human decision making which describes how people make decisions and shows how a person's preferences evolve across time until a decision is reached rather than assuming a fixed state of preference (the decision field theory). Preferences follow a stochastic diffusion process. The model has been used to predict how humans make decisions making under uncertainty, how decisions change under time pressure, how the speed-accuracy trade off works, and how the choice context changes preferences. However, explicit mathematical models seem to be very rare in Economics, Management and Psychology.

In his recent and general "variational rationality" approach, Soubeyran [11, 12] has unified a lot of theories of change, focusing on a new and explicit mathematical model of worthwhile " exploration-exploitation " changes including "effort -accuracy " models as important applications. The last section of our paper will use a specific version of this "worthwhile to change" approach in behavioral sciences (Soubeyran, [11, 12]), see also [13, 14] Attouch-Soubeyran, for a preliminary approach) to modelize "the speed of decision making problem" as an "exploration-exploitation" process in term of proximal algorithms on a Finsler manifold. In the present paper the problem is to minimize some differentiable function on a Finsler manifold. The accuracy (quality) of this solution is the difference $f(x_{k_0}) - f(\tilde{x}) \geq 0$ where $\tilde{x} \in \arg\min \{f(x), x \in M\}$. Efforts, or costs of decision making, are modelized as (infimum) resistance to change which is defined as the desutility of (infimum) costs to change. The case of a Finsler distance, introduced and studied first by Matsumoto [15] and Bao [16], is well adapted to modelize costs to change as the costs to explore a path of potential changes, which increase with the length of such an exploration path. Hence geodesics play a major role to define the infimum of such costs. In this paper we concentrate on the time spend to converge to a critical point x^*, starting from x_{k_0}.

Proximal algorithms on a Finsler manifold To be able to solve complex problems, agents with limited capabilities bracket such problems into subproblems. This is the logic of proximal algorithms where, each step, one has to solve a pertubed optimization problem, supposed to be much easier to solve than the original minimization problem. Attouch-Soubeyran [13, 14] have interpreted these perturbation terms as costs to change. More generally Soubeyran ([11, 12]) has seen such terms as some resistance to change, making a precise link between proximal algorithms and behavioral sciences in term of worthwhile changes which balance, each step, motivation and resistance to change (inertia).

Various practical problems in economics lead to minimization problems. Remarked, that the function appearing in these problems is a typically Finsler metric. In this way, elements from Riemann-Finsler geometry are needed in order to handle the question formulated above.

Let M be a complete Finslerian manifold and let $f : M \to \mathbb{R}$. We will consider the optimization problem

$$\min f(x),\ \ x \in M. \tag{1.1}$$

The proximal point algorithm in Finslerian manifold generates, for a starting point $x_0 \in M$, a sequence $\{x_k\} \subset M$ by the iteration

$$x_{k+1} = \arg\min_{z\in M} \{f(z) + \frac{1}{\lambda_k} C_{x_k}(z)\} \tag{1.2}$$

with $C_{x_k} : M \to \mathbb{R}$ defined by

$$C_{x_k}(z) = (1/2)d^2(x_k, z),$$

where d is the Finslerian distance (to be defined later on), and λ_k is a sequence of positive numbers.

This method was considered, in Riemannian context, by Ferreira and Oliveira [17], in the particular case where M is a Hadamard manifold, $d(x_k, \cdot)$ is a Riemannian distance, $\mathrm{dom} f = M$ and f is convex. They proved that, for each $k \in \mathbb{N}$, the function $f(\cdot) + \frac{1}{2}d^2(x_k, \cdot) : M \to \mathbb{R}$ is 1-coercive and, consequently, that the sequence $\{x_k\}$ is well defined, with x_{k+1} being uniquely determined. Moreover, supposing $\sum_{k=0}^{+\infty} 1/\lambda_k = +\infty$ and that f has a minimizer, the authors proved convergence of the sequence $(f(x_k))$ to the minimum value and convergence of the sequence (x_k) to a minimizer point. With the result of convergence of the sequence generated by the proximal algorithm (1.2) (see Theorem 4.1), from a mathematical point - of - view, our paper generalizes the work of Ferreira and Oliveira [17], using Finslerian distances instead of Riemannian distances.

Several authors have proposed in the last three decades the generalized proximal point algorithm, for certain nonconvex minimization problems. As far as we know the first direct generalization, in the case where M is a Hilbert space, was performed by Fukushima and Mine [19]. In the Riemannian context, Li et al. [18] considered the problem of finding a singularity of a multivalued vector field in a Hadamard manifold and presented a general proximal point method to solve that problem. Papa Quiroz and Oliveira [20] considered the proximal point algorithm for quasiconvex functions (not necessarily convex) and proved full convergence of the sequence $\{x^k\}$ to a minimizer point with M being an Hadamard manifold. Bento et al. [21] considered the proximal point algorithm for C^1-lower type functions and obtained local convergence of the generated sequence to a minimizer, also in the case that M is a Hadamard manifold. With the result of convergence of the sequence generated by our algorithm (1.2) (see Theorem 4.1), our paper generalizes the work of Papa Quiroz and Oliveira [20] and Bento et al. [21], using Finslerian distances instead of Riemannian distances.

So far, in the convergence analysis of the exact proximal point algorithm for solving convex or quasiconvex minimization problems, it was necessary to consider Hadamard type manifolds. This is because the convergence analysis is based on the Fejér convergence to the minimizers set of f and these manifolds, apart from having the same topology and differentiable structures of Euclidean spaces, it also has geometric properties satisfactory to the characterization of Fejér convergence of the sequence. When f is not convex, but M is an Hadamard manifold the convergences of the sequence is derived from the objective function and satisfies a well-known property, an Kurdyka-Lojasiewicz inequality. This inequality has been introduced by Kurdyka [22], for differentiable functions definables in an o-minimal structure defined in $\mathbb{R}^n$ (see section 3), through the following result:

Given $U \subset \mathbb{R}^n$ a bounded open set and $f : U \to \mathbb{R}_+$ a differentiable function definable on an o-minimal structure, there exists $c, \eta > 0$ and a strictly increasing positive function definable $\varphi : \mathbb{R}_+ \to \mathbb{R}$ of class C^1, such that

$$\|\nabla (\varphi \circ f)(x)\| \geq c, \qquad x \in U \cap f^{-1}(0, \eta). \tag{1.3}$$

Note that taking $\varphi(t) = t^{1-\alpha}$, $\alpha \in [0, 1)$, the inequality (1.3) yields

$$\|\nabla f(x)\| \geq c|f(x)|^{\alpha}, \tag{1.4}$$

where $c = 1/(1-\alpha)$, which is know as a Lojasiewicz inequality (see Lojasiewicz [23]). For extensions of Kurdyka-Lojasiewicz inequalities to subanalytic nonsmooth functions (defined in Euclidean spaces) see, for example, Bolte et al. [24], Attouch and Bolte [25]. A more general extension, yet in the context of Euclidean spaces, was developed by Bolte et al. [26] mainly for Clarke's subdifferentiable of a lower semicontinuous function definable in an o-minimal structure. Lageman [27] extended the Kurdyka-Lojasiewicz inequality (1.3) for analytic manifolds and differentiable $\mathcal{C}$-functions in an analytic-geometric category (see Section 3), aiming at establishing an abstract result of convergence of gradient-like methods, see of (Lageman [27, Theorem 2.1.22]). It is important to note that Kurdyka et al. [28] had already established an extension of the inequality (1.4) for analytic manifolds and analytic functions to solve René Thom's conjecture.

In recent years, concepts and techniques of mathematical programming in Euclidean spaces have been extended to Riemannian contexts , not only for theoretical purposes but also to obtain effective algorithms, for example ([29, 17, 30, 31]). Far as we know, in Kristály et al. [32] is presented the first work that models optimization problems in Finsler environments, but effective implementation was in Riemannian manifolds. To extend proximal methods to Finslerian manifolds is the subject of this paper.

This paper is organized as follows. In Section 2, we present the basic tools and properties of Finsler geometry. In Section 3, we recall the Kurdyka-Lojasiewicz property in the Finslerian context and some basic notions on o-minimal structures on $(\mathbb{R}, +, \cdot)$ and analytic-geometric categories. In section 4, the proximal algorithm with Finsler quasi distances is presented and its convergence under mild conditions is established. In Section 5, a model of "exploration-exploitation" on a Finsler manifold is presented. In section 6, we use proximal algorithms with Finsler quasi distances to give an application in term of an effort-accuracy trade off. Finally, in Section 7 we present our conclusions.

2 Elements of Finslerian geometry

In the next subsection, we introduce some fundamental properties and notations of Finslerian manifolds. These basics facts can be found in any introductory book of Finslerian geometry, for example Shen [33] and Bao [16].

2.1 Basic concepts and example

Definition 2.1. *Let M be a connected m-dimensional C^∞ manifold and let*

$$TM = \{(x, y) : x \in M, y \in T_xM\}$$

be its tangent bundle. If a continuous function $F : TM \to \mathbb{R}_+$ satisfies the conditions:

(i) F is C^∞ on $TM \setminus \{0\}$;

(ii) for each $x \in M$, $F_x := F_{|T_xM}$ is such that $F_x(ty) = tF_x(y)$ for all $t > 0$ and $y \in T_xM$, i.e., F_x is positively homogeneous of degree one;

(iii) for each $y \in T_xM\setminus\{0\}$, the symmetric bilinear form g_y on T_xM is positive definite, where

$$g_y(u, v) := \frac{1}{2}\frac{\partial^2}{\partial s \partial t}[F_x^2(y + su + tv)]_{s=t=0}.$$

Then we say that F is a Finsler metric and the pair (M, F) is a Finsler manifold.

In some situations, the Finsler metric F satisfies the criterion $F_x(-y) = F_x(y)$. In that case we have absolute homogeneity instead: $F_x(ty) = |t|\, F_x(y)$ for all $t \in \mathbb{R}$.

Remark 2.1. *From this definition we can derive that:*

$$\begin{aligned} g_y(y,u) &:= \frac{1}{2}\frac{\partial^2}{\partial s \partial t}[F_x^2(y+ty+su)]_{s=t=0}, \\ &= \frac{1}{2}\frac{\partial}{\partial s}[F_x^2(y+su)]_{s=0} \end{aligned} \tag{2.1}$$

and

$$\begin{aligned} g_y(y,y) &:= \frac{1}{2}\frac{\partial^2}{\partial s \partial t}[F_x^2(y+sy+ty)]_{s=t=0} \\ &= \frac{1}{2}\frac{\partial^2}{\partial s \partial t}(1+s+t)^2[F_x^2(y)]_{s=t=0} \\ &= F_x^2(y). \end{aligned} \tag{2.2}$$

An important class of Finsler metrics is that of Riemannian metrics.

Example 2.1. *(Riemannian metric) Let $g = \{g_x\}_{x \in M}$, where g_x is a positive definite symmetric bilinear form in T_xM such that in local coordinates (x^i),*

$$g_{ij}(x) := g_x\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right)$$

are C^∞ functions, g is called a Riemannian metric. This g defines a symmetric Finsler struture on TM by the mechanism

$$F_x(y) := \sqrt{g_x(y,y)}.$$

Every Riemannian manifold (M, g) is therefore a Finsler manifold.

Example 2.2. *(Randers metric) Let $\alpha(y) = \sqrt{a_{ij}(x)y^iy^j}$ be a Riemannian metric and $\beta(y) = b_i(x)y^i$ be a 1-form on a manifold M. Assume that for $x \in M$ and $y \in T_xM$*

$$\|\beta(y)\|_x = \sup_{\alpha(y)=1} \beta(y) < 1,$$

where a_{ij} are the components of the Riemannian metric and b_i are those of a 1-form. A Randers metric is a Finsler structure on TM that has the form

$$F_x(y) := \alpha(y) + \beta(y).$$

Note that, due to presence of the term β, Rander's metrics do not satisfy $F_x(-y) = F_x(y)$ when $b_i \neq 0$. In fact, $F_x(-y) = F_x(y)$ if and only if it is a Riemannian metric.

An interesting illustration of Example 2.2 in $\mathbb{R}^2$, shown in Bao [16, page 20] is as follows.

Example 2.3. *Let $M = \{x \in \mathbb{R}^2 : (x_1)^2 + (x_2)^2 < 2\}$ be the Poincaré disc and*

$$F_x(y) = \frac{1}{1 - \frac{r^2}{4}} \sqrt{y \cdot y} + \frac{r}{(1 - \frac{r^2}{4})(1 + \frac{r^2}{4})} dr(y)$$

Here, $r^2 = (x_1)^2 + (x_2)^2$, $x = (x_1, x_2)$ is any point in M and y is an arbitrary vector in the tangent plane $T_x M$.

Given a Finsler metric F on a manifold M. Let $\gamma : [0,1] \to M$ be a piecewise C^∞ curve. Its integral length is defined as

$$L(\gamma) := \int_0^1 F_{\gamma(t)}(\dot{\gamma}(t))dt. \tag{2.3}$$

For $x, z \in M$, denote by $\Gamma(x,z)$ the set of all piecewise C^∞ curves $\gamma : [0,1] \to M$ such that $\gamma(0) = x$ and $\gamma(1) = z$. This length L induces a function

$$d : M \times M \to \mathbb{R}_+$$

by

$$d(x,z) = \inf_{\gamma \in \Gamma(x,z)} L(\gamma)$$

Of course, we have $d(x,z) \geq 0$, where equality holds if and only if $x = z$, and

$$d(x,z) \leq d(x,p) + d(p,z),$$

for all $x, p, z \in M$, the so-called triangle inequality (for details see Shen [33, page 13])

2.2 Fundamental properties

Let $\pi : TM \setminus \{0\} \to M$, $\pi(x,y) = x$, be the natural projection of TM and $\pi^*(TM)$ be the pull-back tangent bundle TM. Unlike the Levi-Civita connection in Riemann geometry, there is no unique natural connection in the Finsler case. Among these connections on $\pi^*(TM)$, we choose the Chern connection whose coefficients are denoted by Γ^k_{ij} ; see Bao [16, page 38]. This connection induces the curvature tensor, denoted by R; see Bao [16, Chap. 3]. The Chern connection defines the covariant derivative $D_V U$ of a vector field U in the direction $V \in T_x M$. Since, in general, the Chern connection coefficients Γ^k_{ij} in natural coordinates have a directional dependence, we must say explicitly that $D_V U$ is defined with a fixed reference vector. In particular, let $\sigma : [0,r] \to M$ be a smooth curve with velocity field $T = T(t) = \dot{\sigma}(t)$. Suppose that U and W are vector fields defined along σ. We define $D_T U$ with reference vector W as

$$D_T U = \left[\frac{dU_i}{dt} + U^j T^k (\Gamma^k_{ij})_{(\sigma,W)} \right] \frac{\partial}{\partial x^i}_{|_{\sigma(t)}}, \tag{2.4}$$

where $\{\frac{\partial}{\partial x^i}|_{\sigma(t)}\}_{i=1\ldots m}$ is a basis of $T_{\sigma(t)}M$. A C^∞ curve $\sigma : [0, r] \to M$, with velocity $T = \sigma'(t)$ is a (Finslerian) geodesic if

$$D_T \left[\frac{T}{F(T)} \right] = 0$$

with reference vector T. If the Finslerian velocity of the geodesic σ is constant, then

$$\frac{d^2\sigma^i}{dt^2} + \frac{d\sigma^j}{dt}\frac{d\sigma^i}{dt}\Gamma^k_{ij}(\sigma, T) = 0 \qquad i = 1, \ldots, m = \dim M, \tag{2.5}$$

where T denotes the velocity field associated with σ.

A Finsler manifold (M, F), where F is positively (but perhaps not absolutely) homogeneous of degree one, is said to be forward geodesically complete if every geodesic $\sigma : [0, 1] \to M$ parametrized to have constant Finslerian speed, can be extended to a geodesic defined on $(0, \infty)$. The Hopf-Rinow's theorem Bao [16, page 168] gives several characterizations of this completeness. It asserts that any pair of points, say x and z in M can be joined by a (not necessarily unique) minimal geodesic segment. In this paper, all manifolds are assumed to be forward geodesically complete.

For any $x \in M$ and $y \in T_xM$, we may define the exponential map

$$\exp_x : T_xM \to M,$$

$$\exp_x(y) = \sigma(1, x, y)$$

where $\sigma(t, x, y)$ is the unique solution (geodesic) of the second order differential equation (6) which passes through x at $t = 0$ with velocity y. Moreover,

$$d(\exp_x)(0) = id_{T_xM}. \tag{2.6}$$

If U, V and W are vector fields along a curve σ, which has velocity $T = \sigma'$, we have the derivative rule

$$\frac{d}{dt} g_W(U, V) = g_W(D_TU, V) + g_W(U, D_TV),$$

whenever D_TU and D_TV use the reference vector W and one of the following conditions hold:
i) U or V is proportional to W, or
ii) W = T and σ is a geodesic.

Let (M, F) be a Finsler space. For a vector $y \in T_xM \setminus \{0\}$, the Riemannian curvature $R_y : T_xM \to T_xM$ is a self-adjoint linear transformation with respect to g_y. Let $P \subset T_xM$ be a tangent plane. For a vector $y \in P \setminus \{0\}$, the number $K(P, y)$ given by

$$K(P, y) = \frac{g_y(R_y(u), u)}{g_y(y, y)g_y(u, u) - g_y(y, u)g_y(y, u)},$$

where $u \in P$ is such that $P = span\{y, u\}$, is called the flag curvature of the flag $(P, y) \in T_xM$.

When F is a Riemannian metric, then, the flag curvature coincides with the sectional curvature. Let K be the collection of flag curvatures $K(P, y)$, with $y \neq 0$, $P \subset T_xM$. We say that the flag curvature of (M, F) is non positive if $K(P, y) \leq 0$ for any flag (P, y).

Definition 2.2. *A Finsler space (M, F) is called a Hadamard manifold if it is forward geodesically complete, simply connected with non-positive flag curvature.*

Theorem 2.1. *(Theorem of Cartan-Hadamard) Let (M, F) be a forward geodesically complete, simply connected Finsler manifold of nonpositive flag curvature. Then the exponential map $\exp_x$ is a C^1 diffeomorphism from the tangent space T_xM onto the manifold M.*

Proof. The proof is in Bao [16, page 238]. ∎

The next lemma establishes that the distance from a point $x_1 \in M$ to another point $x_2 \in M$ is a relatively close distance from x_2 to x_1.

Lemma 2.1. *Let (M,F) be a Finsler manifold. At every given point $x \in M$, there exists a coordinate neighborhood U containing x, together with a constant $c_0 > 1$ such that*

(i) for all $y \in T_xM$ e $x \in \overline{U}$,

$$F_x(-y) \leq c_0^2 F_x(y)$$

(ii) for all $x_1, x_2 \in U$,

$$\frac{1}{c_0^2} d(x_1, x_2) \leq d(x_2, x_1) \leq c_0^2 d(x_1, x_2).$$

Proof. The proof is in Bao [16, page 146]. ∎

2.3 Gradient on a Finsler manifold

First, let us study the relationship between a Minkowski norm (see Shen [33, Definition 1.2.1]) and its dual norm. Let V be a finite-dimensional vector space and V^* the dual vector space. Given a Minkowski norm F on V, F is a norm in the sense that for any y, $v \in V$ and $\lambda > 0$,

$$F(\lambda y) = \lambda F(y),$$

and

$$F(y + v) \leq F(y) + F(v).$$

Let V^* denote the vector space dual to V. Define

$$F^*(\xi) := sup_{F(y)=1} \xi(y),$$

F^* is a Minkowski norm on V^*.

The next Lemma provides a gradient characterization in the setting of Finsler manifolds.

Lemma 2.2. *Let F be a Minkowski norm on V and F^* be the dual norm on V^*. For any vector $y \in V \setminus \{0\}$, the covector $\xi = g_y(y, \cdot) \in V^*$ satisfies*

$$F(y) = F^*(\xi) = \frac{\xi(y)}{F(y)}$$

For any covector $\xi \in V^* \setminus \{0\}$, there exists a unique vector $y \in V \setminus \{0\}$ such that $\xi = g_y(y, \cdot)$.

Proof. See Shen [33, page 35]. ∎

Given a differentiable function $f : M \to \mathbb{R}$ on a manifold M, the differential $df_x \in V^*$ at a point $x \in M$,

$$df_x = \frac{\partial f}{\partial x^i}(x) dx^i,$$

is a linear functional on T_xM. To connect df_x to a vector $gradf_x \in T_xM$, we need a Minkowski norm on T_xM. Let F be a Finsler metric on M. By definition, F_x is a Minkowski norm on T_xM. Assume that $df_x \neq 0$. Since the indicatrix $S := F^{-1}(1)$ is strongly convex, there is a unique unit vector $s_x \in S_xM := F_x^{-1}(1)$ and a positive number $\lambda_x > 0$ such that

$$W^{\lambda_x} = \{v : df_x(v) = \lambda_x\}$$

is tangent to S_xM at s_x. By Lemma 2.2

$$df_x(v) = \lambda_x g_{s_x}(s_x, v)$$

where

$$F_x^*(df_x) = df_x(s_x) = \lambda_x g_{s_x}(s_x, s_x) = \lambda_x.$$

Define

$$gradf_x := \lambda_x s_x = F_x^*(df_x) s_x.$$

Thus we can write

$$df_x(v) = g_{gradf_x}(gradf_x, v), v \in T_xM. \tag{2.7}$$

Given a compact subset $S \subset M$ define $d_+(x) := d(S, x)$ and $d_-(x) := -d(x, S)$.

Lemma 2.3. *Let M be Finsler Hadamard manifold, $S \subset M$ a compact subset and $d_+(x)$ and $d_-(x)$ distances defined in the last expression. Then*

$$F(grad\, d_+(x)) = 1 \ \ e\ F(grad\, d_-(x)) = 1$$

Proof. See Shen [33, page 44]. ∎

Let M be a Finsler Hadamard manifold. For any $x \in M$ we can define the exponential inverse map

$$\exp_x^{-1} : M \to T_xM$$

which is C^∞ nearby x and C^1 at x.

Remark 2.2. *Note that $d(x,z)$ is the distance from x for z and F_x is a Finsler nonsymmetric metric in T_xM such that $d(x,z) = F_x(\exp_x^{-1}(z)) \neq F_x(\exp_z^{-1}(x)) = d(z,x)$, i.e., the function $z \mapsto d(\cdot :, z)$ is a quasi distance in M.*

3 Kurdyka-Lojasiewicz inequality on Finslerian manifolds

In this subsection we present Kurdyka-Lojasiewicz inequality in the Finslerian context and we recall some basic notions on o-minimal structures on $(\mathbb{R}, +, \cdot)$ and analytic-geometric categories. Our main interest here is to observe that Kurdyka-Lojasiewicz inequality in Finslerian contexts holds. The differentiable case was presented by Lageman [27, Corollary 1.1.25] in the setting of Riemannian manifolds. It is important to note that Kurdyka et al. [28] had already established such inequality for analytic manifolds and analytic functions. For a detailed discussion on o-minimal structures and analytic geometric categories see, for example, van den Dries and Miller [34], and references therein.

Let $f : M \to \mathbb{R}$ be a C^1 function and consider the following set:

$$[\eta_1 < f < \eta_2] := \{x \in M : \eta_1 < f(x) < \eta_2\}, \quad -\infty < \eta_1 < \eta_2 < +\infty.$$

Definition 3.1. *The function f above is said to have the Kurdyka-Lojasiewicz property at $\bar{x} \in domf$ if there exist $\eta \in (0, +\infty]$, a neighbourhood U of $\bar{x}$ and a continuous concave function $\varphi : [0, \eta) \to \mathbb{R}_+$ such that:*

(i) $\varphi(0) = 0$, $\varphi \in C^1(0, \eta)$ *and, for all* $s \in (0, \eta)$, $\varphi'(s) > 0$;

(ii) *for all* $x \in U \cap [f(\bar{x}) < f < f(\bar{x}) + \eta]$, *the Kurdyka-Lojasiewicz inequality holds*

$$\varphi'(f(x) - f(\bar{x}))F(grad\, f(x)) \geq 1. \tag{3.1}$$

We call f a KL function if it satisfies the Kurdyka-Lojasiewicz inequality at each point of $domf$.

Next we show that if $\bar{x}$ is a noncritical point of a C^1 function then the Kurdyka-Lojasiewicz inequality holds in $\bar{x}$.

Lemma 3.1. *Let $f : M \to \mathbb{R}$ be a C^1 function and $\bar{x} \in M$ such that $0 \neq grad\, f(\bar{x})$. Then the Kurdyka-Lojasiewicz inequality holds in $\bar{x}$.*

Proof. Since $\bar{x}$ is a noncritical of f, we have that

$$\delta := F(grad\, f(\bar{x})) > 0.$$

Take $\varphi(t) := t/\delta$, $U := B(\bar{x}, \delta/2)$, $\eta := \delta/2$ and note that, for each $x \in M$,

$$\varphi'(f(x) - f(\bar{x}))F(grad\, f(x)) = F(grad\, f(x))/\delta. \tag{3.2}$$

Now, for each $x \in U \cap [f(\bar{x}) - \eta < f < f(\bar{x}) + \eta]$ arbitrary, note that

$$d(x, \bar{x}) + |f(x) - f(\bar{x})| < \delta.$$

We state that, for each x satisfying the last inequality, it holds

$$F(grad\, f(x)) \geq \delta. \tag{3.3}$$

Let us suppose, by contradiction, that this does not hold. Then, there exist sequences $\{(z_k, grad\, f(z_k)\} \subset \text{graph} grad\, f$ and $\{\delta_k\} \subset \mathbb{R}_{++}$ such that

$$d(z_k, \bar{x}) + |f(z_k) - f(\bar{x})| < \delta_k, \quad \text{and} \quad F(grad\, f(z_k)) \leq \delta_k$$

with $\{\delta_k\}$ converging to zero. Thus, using that $\{(z_k, grad\, f(z_k))\}$ and $\{f(z_k)\}$ converge to $(\bar{x}, 0)$ and $f(\bar{x})$ respectively, and $F(grad\, f)$ is continuous , it follows that $\bar{x}$ is a critical point of f, which proves the statement. Therefore, the result of the lemma follows by combining (3.2) with (3.3). ■

Next, we recall some definitions which refer to o-minimal structures on $(\mathbb{R}, +, \cdot)$, following the notations of Bolte et al. [26].

Definition 3.2. *Let $\mathcal{O} = \{\mathcal{O}_n\}_{n \in \mathbb{N}}$ be a sequence such that each $\mathcal{O}_n$ is a collection of subsets of $\mathbb{R}^n$. $\mathcal{O}$ is said to be an o-minimal structure on the real field $(\mathbb{R}, +, \cdot)$ if, for each $n \in \mathbb{N}$:*

(i) $\mathcal{O}_n$ is a boolean algebra;

(ii) If $A \in \mathcal{O}_n$, then $A \times \mathbb{R} \in \mathcal{O}_{n+1}$ and $\mathbb{R} \times A \in \mathcal{O}_{n+1}$;

(iii) If $A \in \mathcal{O}_{n+1}$, then $\pi_n(A) \in \mathcal{O}_n$, where $\pi_n : \mathbb{R}^{n+1} \to \mathbb{R}^n$ is the projection on the first n coordinates;

(iv) $\mathcal{O}_n$ contains the family of algebraic subsets of $\mathbb{R}^n$;

(v) $\mathcal{O}_1$ consists of all finite unions of points and open intervals.

The elements of $\mathcal{O}$ are said to be *definable* in $\mathcal{O}$. A function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be *definable* in $\mathcal{O}$ if its graph belongs to $\mathcal{O}_{n+1}$. Moreover, according to Coste [35] a function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is said to be definable in $\mathcal{O}$ if the inverse images of $f^{-1}(+\infty)$ is a definable subset of $\mathbb{R}^n$ and the restriction of f to $f^{-1}(\mathbb{R})$ is a definable function with values in $\mathbb{R}$. It is worth noting that an o-minimal structure on the real field $(\mathbb{R}, +, \cdot)$ is a generalization of a semialgebraic set on $\mathbb{R}^n$, i.e., a set which can be written as a finite union of sets of the form

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

with $p_i, q_i,\ i = 1, \ldots, r$, being real polynomial functions. Bolte et al. [26], presented a nonsmooth extension of the Kurdyka-Lojasiewicz inequality for definable functions, but in the case where the function φ, which appears in Definition 3.1, is not necessarily concave. Attouch et al. [36], reconsidered the mentioned extension by noting that φ may be taken concave. For an extensive list of examples of definable sets and functions on an o-minimal structure and properties, see , for example, (van den Dries and Miller [34] and Attouch et al. [36]), and references therein. We limit ourselves to present just the material required for our purposes.

The first elementary class of examples of definable sets is given by the semi-algebraic sets, which we denote by $\mathbb{R}_{alg}$. An other class of examples, which we denoted by $\mathbb{R}_{an}$, is given by restricted analytic functions, i.e., the smallest structure containing the graphs of all $f\ |_{[0,1]^n}$ analytic functions , where $f : \mathbb{R}^n \to \mathbb{R}$ is an arbitrary function that vanishes off identically $[0, 1]^n$.

Fulfilling the same role as the semi-algebraic sets on X on analytic manifolds we have the semi-analytic and sub-analytic sets which we define below, see Bierstone and Milman [37] and van den Dries [38]:

A subset of an analytic manifold is said to be semi-analytic if it is locally described by a finite number of analytic equations and inequalities, while the sub-analytic ones are locally projections of relatively compact semi-analytic sets.

A generalization of semi-analytic and sub-analytic sets, analogous to what was given to semi-algebraic sets in terms of the o-minimal structure, leads us to the analytic-geometric categories which we define below:

Definition 3.3. *An analytic-geometric category $\mathcal{C}$ assigns to each real analytic manifold M a collection of sets $\mathcal{C}(M)$ such that for all real analytic manifolds M, N the following conditions hold:*

(i) *$\mathcal{C}(M)$ is a boolean algebra of subsets of M, with $M \in \mathcal{C}(M)$;*

(ii) *If $A \in \mathcal{C}(M)$, then $A \times \mathbb{R} \in \mathcal{C}(A \times \mathbb{R})$;*

(iii) *If $f : M \to N$ is a proper analytic map and $A \in \mathcal{C}(M)$, then $f(A) \in \mathcal{C}(N)$;*

(iv) *If $A \subset M$ and $\{U_i \mid i \in \Lambda\}$ is an open covering of M, then $A \in \mathcal{C}(M)$ if and only if $A \cap U_i \in \mathcal{C}(U_i)$, for all $i \in \Lambda$;*

(v) *Every bounded set $A \in \mathcal{C}(\mathbb{R})$ has finite boundary, i.e. the topological boundary, ∂A, consists of a finite number of points.*

The elements of $\mathcal{C}(M)$ are called $\mathcal{C}$-sets. In the case where the graph of a continuous function $f : A \to B$ with $A \in \mathcal{C}(M), B \in \mathcal{C}(N)$ is contained in $\mathcal{C}(M \times N)$, then f is called a $\mathcal{C}$-function. All subanalytic subsets and continuous subanalytic maps of a manifold are $\mathcal{C}$-sets and $\mathcal{C}$-functions respectively, in that manifold. We denoted this collection by $\mathcal{C}_{an}$ which represents the "smallest" analytic-geometric category.

The next theorem provides a biunivocal correspondence between o-minimal structures containing $\mathbb{R}_{an}$ and analytic-geometric categories.

Theorem 3.1. *For any analytic-geometric category $\mathcal{C}$ there is an o-minimal structure $\mathcal{O}(\mathcal{C})$ and for any o-minimal structure $\mathcal{O}$ on $\mathbb{R}_{an}$ there is an analytic geometric category $\mathcal{C}(\mathcal{O})$, such that*

(i) $A \in \mathcal{C}(\mathcal{O})$ *if for all* $x \in M$ *it exists an analytic chart* $\phi : U \to \mathbb{R}^n$ *,* $x \in U$*, which maps* $A \cap U$ *onto a set definable in* $\mathcal{O}$*.*

(ii) $A \in \mathcal{O}(\mathcal{C})$ *if can be mapped onto a bounded* $\mathcal{C}$*-set in an Euclidean space by a semialgebraic bijection.*

Furthermore, for $\mathcal{C} = \mathcal{C}(\mathcal{O})$ *we get the back the o-minimal structure* $\mathcal{O}$ *by this correspondence, and for* $\mathcal{O} = \mathcal{O}(\mathcal{C})$ *we get again* $\mathcal{C}$*.*

Proof. See van den Dries and Miller [34] and Lageman [27, Theorem 1.1.3]. ■

As a consequence of the correspondence between o-minimal structures containing $\mathbb{R}_{an}$ and analytic-geometric categories, the associated sets allow us to provide examples of $\mathcal{C}$-sets in $\mathcal{C}(\mathcal{O})$. Furthermore, $\mathcal{C}$-functions are locally mapped to definable functions by analytic charts.

Proposition 3.1. *Let* $f : M \to \mathbb{R}$ *be a* $\mathcal{C}$*-function and* $\phi : U \to \mathbb{R}^n$*,* $U \subset M$ *an analytic local chart. Assume that* $U \subset domf$ *and* $V \subset M$ *is a bounded open set such that* $\overline{V} \subset U$*. If* f *restricted to* U *is a bounded* $\mathcal{C}$*-function, then*

$$f \circ \phi^{-1} : \phi(V) \to \mathbb{R}, \tag{3.4}$$

is definable in $\mathcal{O}(\mathcal{C})$*.*

Proof. See Lageman [27, Proposition 1.1.5.]. ■

Next result provided us with the nonsmooth extension of the Kurdyka-Lojasiewicz properties for $\mathcal{C}$-functions defined on analytic manifolds.

Theorem 3.2. *Let* M *be a analytic Riemannian manifold and* $f : M \to \mathbb{R}$ *a continuous* $\mathcal{C}$*-function. Then,* f *is a KL function. Moreover, the function* φ *which appears in* (3.1) *is definable in* $\mathcal{O}$*.*

Proof. Take $\bar{x} \in M$ a critical point of f and let $\phi : V \to \mathbb{R}^n$ be an analytic local chart with $V \subset M$ a neighbourhood of $\bar{x}$ chosen such that V and $f(V)$ are bounded. Thus, from Proposition 3.1, we have that $f \circ \phi^{-1} : \phi(V) \to \mathbb{R}$ is a definable function in $\mathcal{O}(\mathcal{C})$. Thus, as $\phi(V)$ is a bounded open definable set containing $\bar{z} = \phi(\bar{x})$ and ϕ is definable, applying Theorem 11 of (Bollte et al. [26]) with $U = \phi(V)$ and taking into account the proof of Theorem 4.1 of (Attouch et al. [36]), Kurdyka-Lojasiewicz properties holds at $\bar{z} = \phi(\bar{x})$, i.e., there exists $\eta \in (0, +\infty]$ and a continuous concave function $\Phi : [0, \eta) \to \mathbb{R}_+$ such that:

(i) $\Phi(0) = 0$, $\Phi \in C^1(0, \eta)$ and, for all $s \in (0, \eta)$, $\Phi'(s) > 0$;

(ii) for all $z \in U \cap [h(\bar{z}) < h < h(\bar{z}) + \eta]$, it holds

$$\Phi'(h(z) - h(\bar{z}))F(grad\, h(z)) \geq 1.$$

Since ϕ is a diffeomorphism and using that $z = \phi(x)$, $\bar{z} = \phi(\bar{x})$ and $h = f \circ \phi^{-1}$, from chain rule

$$\Phi'(f(x) - f(\bar{x}))F((\phi_x^*)^{-1} grad\, f(x)) \geq 1, \quad x \in V \cap [0 < f < f(\bar{x}) + \eta],$$

where ϕ_x^* denote the derivative adjunct of the map ϕ.

Take $V' \subset V$ an open set such that $K = \overline{V'}$ is contained in the interior of the set V and $\bar{x} \in V'$. Thus, K is compact set and for each $x \in K$ there exists $C_x > 0$ with

$$F((\phi_x^*)^{-1} w) \leq C_x F(w), \qquad w \in T_x M.$$

Since K is a compact set and $(\phi_x^*)^{-1}$ is a diffeomorphism, there exists a positive constant $C := \sup\{C_x : x \in K\}$ such that

$$F((\phi_x^*)^{-1} w) \leq C F(w), \qquad w \in T_x M,\ x \in K.$$

Hence, for $x \in V' \cap [0 < f < f(\bar{x}) + \eta]$, we have

$$1 \leq \Phi'(f(x) - f(\bar{x}))F((\phi_x^*)^{-1} grad\, f(x)) \leq C\, \Phi'(f(x) - f(\bar{x}))F(grad\, f(x)),$$

and the Kurdyka-Lojasiewicz properties holds at $\bar{x}$ with $\varphi = C\,\Phi$. Therefore, combining arbitrary of $\bar{x}$ with Lemma 3.1 we conclude that f is a KL function. The second part also follows from Theorem 11 of (Bolte et al. [26]) and the proof is concluded. ■

In the next section we prove the main results of this paper. We will show that the algorithm is well defined and that the generated sequence converges.

4 Proximal algorithms

From now on we will fix our area of activity and establish the conditions to ensure proper definition and convergence of our method. Before we call attention to the fact that the Finslerian distance plays a fundamental role in the next section. We proceed now stating a result which we will go to use.

4.1 Convergence

Let $f : M \to \mathbb{R}$ be a analytical function. Being given $x_0 \in M$, the algorithm we are to study is of the form:

$$x_{k+1} = \arg\min\{f(z) + \frac{1}{2\lambda_k} d^2(x_k, z), z \in M\}. \tag{4.1}$$

where d is the distance associated with the manifold M and (λ_k) is a sequence of positive numbers.

Throughout the text from now unless explicit mentions we consider that M is on Finsler Hadamard manifold, $\inf_M f > -\infty$ and for some positive $t_1 < t_2$, $(t_1 < \lambda_k < t_2)$, for all $k \geq 0$.

The following proposition is fundamental to this article and will be widely used.

Proposition 4.1. *Let (x_k) be the sequence concerning (4.1). Then (x_k) is well defined. Moreover:*

$$f(x_{k+1}) + \frac{1}{2\lambda_k} d^2(x_k, x_{k+1}) \leq f(x_k) \tag{4.2}$$

and consequently

$$\sum_{k=0}^{\infty} d^2(x_k, x_{k+1}) < \infty. \tag{4.3}$$

Proof. Since $\inf f > -\infty$, M is a Hadamard manifold and using similar arguments the case Riemannian, see Ferreira and Oliveira [17], we have that for any $r > 0$, $\bar{x} \in M$ the function

$$x \to f(x) + \frac{1}{2r} d^2(\bar{x}, x)$$

is 1-coercive, consequently, the well definedness of the sequence (x_k) with x_{k+1} being uniquely determined. An elementary induction ensures then that (4.2) holds. From of (4.2) and using telescopic sum we get (4.3). ∎

Next lemma, we give an estimate of the gradient norm in terms of distance.

Lemma 4.1. *let (x_k) be the sequence generated by* (4.1) *and $x_{k_0} \in B(\tilde{x}, \rho) \subset U$, where the neighborhood U is given by Lemma 2.1 . Then*

$$F(grad\, f(x_{k_0})) \leq c_0^2 t_1^{-1} d(x_{k_0 - 1}, x_{k_0}),$$

where $c_0 > 1$ e $t_1 < \lambda_{k_0}$.

Proof. Since x_{k_0} is minimizer point of function

$$z \to f(z) + \frac{1}{2\lambda_{k_0 - 1}} d^2(x_{k_0 - 1}, z),$$

we have

$$grad\left(f(x_{k_0}) + \frac{1}{2\lambda_{k_0 - 1}} d^2(x_{k_0 - 1}, x_{k_0})\right) = 0,$$

consequently the differential

$$\mathrm{d}\left(f(x) + \frac{1}{2\lambda_{k_0 - 1}} d^2(x_{k_0 - 1}, x)\right)_{|x = x_{k_0}} = 0.$$

Using the linearity of the differential

$$\mathrm{d}f_{x_{k_0}} = -\mathrm{d}\left(\frac{1}{2\lambda_{k_0-1}}d^2(x_{k_0-1},x)\right)_{|x=x_{k_0}}.$$

Denote $h(x) := \frac{1}{2\lambda_{k_0-1}}d^2(x_{k_0-1},x)$ and $F_x = F$. Thus, by (2.7)

$$\mathrm{d}f_{x_{k_0}}\left(grad\, f(x_{k_0})\right) = g_{grad\, f(x_{k_0})}\left(grad\, f(x_{k_0}), grad\, f(x_{k_0})\right)$$

and

$$\mathrm{d}h_{x_{k_0}}(grad\, f(x_{k_0})) = g_{grad\, h(x_{k_0})}(grad\, h(x_{k_0}), grad\, f(x_{k_0})).$$

Thus

$$\begin{aligned} g_{grad\, f(x_{k_0})}(grad\, f(x_{k_0}), grad\, f(x_{k_0})) &= -g_{grad\, h(x_{k_0})}(grad\, h(x_{k_0}), grad\, f(x_{k_0})) \\ &= g_{grad\, h(x_{k_0})}(grad\, h(x_{k_0}), -grad\, f(x_{k_0})). \end{aligned}$$

Therefore, using the Cauchy-Schwartz inequality ($g_y(y,v) \leq F(y)F(v)$) and Lemma 2.1, we concluded

$$\begin{aligned} F^2(grad\, f(x_{k_0})) &\leq F(grad\, h(x_{k_0}))F(-grad\, f(x_{k_0})) \\ &\leq F(grad\, h(x_{k_0}))c_0^2F(grad\, f(x_{k_0})). \end{aligned}$$

On other hand,

$$grad\, h(x_{k_0}) = \frac{1}{\lambda_{k_0-1}}d(x_{k_0-1},x_{k_0})grad\ d(x_{k_0-1},x_{k_0}).$$

The homogeneity of F implies

$$F(grad\, f(x_{k_0})) \leq c_0^2F(grad\, h(x_{k_0})) = \frac{c_0^2 d(x_{k_0-1},x_{k_0})}{\lambda_{k_0-1}}F(grad\ d(x_{k_0-1},x_{k_0})).$$

taking $S = \{x_{k_0-1}\}$ and $x = x_{k_0}$ in the Lemma 2.3, $F(grad\ d(x_{k_0-1},x_{k_0})) = 1$. Then

$$F(grad\, f(x_{k_0})) \leq c_0^2 t_1^{-1} d(x_{k_0-1},x_{k_0}).$$

■

Next, we present one technical result that will be useful in convergence analysis.

Lemma 4.2. *Let $\{a_k\}$ be a sequence of positive numbers such that*

$$\sum_{k=1}^{+\infty} a_k^2/a_{k-1} < +\infty.$$

Then, $\sum_{k=1}^{+\infty} a_k < +\infty$.

Proof. Take $j \in \mathbb{N}$ fixed. Note that,

$$\sum_{k=1}^{j} a_k = \sum_{k=1}^{j} \frac{a_k}{\sqrt{a_{k-1}}}\sqrt{a_{k-1}} \leq \left(\sum_{k=1}^{j} \frac{a_k^2}{a_{k-1}}\right)^{1/2} \left(\sum_{k=1}^{j} a_{k-1}\right)^{1/2},$$

where the above inequality follows from Cauchy-Schwartz inequality in $\mathbb{R}^j$ with respect to the vectors $\left(a_1/\sqrt{a_0}, \ldots, a_j/\sqrt{a_{j-1}}\right)$ and $\left(\sqrt{a_0}, \ldots, \sqrt{a_{j-1}}\right)$. Thus,

$$\sum_{k=1}^{j} a_k \leq \left(\sum_{k=1}^{j} \frac{a_k^2}{a_{k-1}}\right)^{1/2} \left(\sum_{k=1}^{j} a_{k-1}\right)^{1/2}.$$

Now, adding a_0 to both sides of the last inequality and taking into account that $a_j > 0$, we obtain

$$\sum_{k=1}^{j} a_{k-1} \leq a_0 + \left(\sum_{k=1}^{j} \frac{a_k^2}{a_{k-1}}\right)^{1/2} \left(\sum_{k=1}^{j} a_{k-1}\right)^{1/2}.$$

Therefore, dividing both sides of last inequality by $\left(\sum_{k=1}^{j} a_{k-1}\right)^{1/2}$ and observing that

$$a_0 / \left(\sum_{k=1}^{j} a_{k-1}\right)^{1/2} \leq \sqrt{a_0} \qquad (a_k > 0, k = 0, 1, \ldots),$$

it follows that

$$\left(\sum_{k=1}^{j} a_{k-1}\right)^{1/2} \leq \sqrt{a_0} + \left(\sum_{k=1}^{j} \frac{a_k^2}{a_{k-1}}\right)^{1/2},$$

and the desired result follows by using comparison theorem's. ■

Lemma 4.3. *Let (x_k) be the sequence concerning (4.1), $f : M \to \mathbb{R}$ a C^1 function, $\tilde{x}$ a accumulation point of (x_k) and f satisfies the Kurdyka-Lojasiewicz inequality at $\tilde{x}$. Let $a = 1/2t_2$, $b = c_0^2/t_1$ constants and $\rho > 0$ such that $B(\tilde{x}, \rho) \subset U$, where U is given by Lemma 2.1. Then there exists $k_0 \in \mathbb{N}$ such that*

$$f(\tilde{x}) < f(x_k) < f(\tilde{x}) + \eta, \qquad k \geq k_0, \tag{4.4}$$

$$d(\tilde{x}, x_{k_0}) + 2\sqrt{\frac{f(x_{k_0}) - f(\tilde{x})}{a}} + \frac{b}{a}\varphi(f(x_{k_0}) - f(\tilde{x})) < \rho. \tag{4.5}$$

Moreover,

$$\frac{b}{a}[\varphi(f(x_{k_0}) - f(\tilde{x})) - \varphi(f(x_{k_0+1}) - f(\tilde{x}))] \geq \frac{d^2(x_{k_0}, x_{k_0+1})}{d(x_{k_0-1}, x_{k_0})}. \tag{4.6}$$

In particular, if $x_k \in B(\tilde{x}, \rho)$ for all $k \geq k_0$, then $\sum_{k=k_0}^{+\infty} d(x_k, x_{k+1}) < +\infty$ and thus the sequence (x_k) converges to $\tilde{x}$.

Proof. Let (x_{k_j}) be a subsequence of (x_k) converging to $\tilde{x}$. Now, from (4.2) and of continuity of f, $(f(x_{k_j}))$ converge to $f(\tilde{x})$. Since that by Proposition 4.1 $(f(x_k))$ is a decreasing sequence we obtain that the whole sequence $(f(x_k))$ converges to $f(\tilde{x})$ as k goes to $+\infty$ and, hence

$$f(\tilde{x}) < f(x_k), \qquad k \in \mathbb{N}. \tag{4.7}$$

In particular, there exists $N \in \mathbb{N}$ such that

$$f(\tilde{x}) < f(x_k) < f(\tilde{x}) + \eta, \qquad k \geq N. \tag{4.8}$$

Since (4.7) holds, let us define the sequence (b_k) given by

$$b_k = d(\tilde{x}, x_k) + 2\sqrt{\frac{f(x_k) - f(\tilde{x})}{a}} + \frac{b}{a}\varphi(f(x_k) - f(\tilde{x})).$$

As $d(\cdot, \tilde{x})$ and φ are continuous it follows that 0 is an accumulation point of the sequence $\{b_k\}$ and hence there exists $k_0 := k_{j_0} > N$ such that (4.5) holds. In particular, as $k_0 > N$, from (4.8) it also holds (4.4).

From (4.4) combined with $x_{k_0} \in B(\tilde{x}, \rho)$ (it follows from (4.5)), we have that

$$x_{k_0} \in B(\tilde{x}, \rho) \cap [f(\tilde{x}) < f < f(\tilde{x}) + \eta].$$

So, since $\tilde{x}$ is a point where f satisfies the Kurdyka-Lojasiewicz inequality it follows that $0 \neq grad f(x_{k_0})$. Moreover, from Lemma 4.1,

$$F(grad\, f(x_{k_0})) \leq c_0^2 t_1^{-1} d(x_{k_0-1}, x_{k_0},) = bd(x_{k_0-1}, x_{k_0}),$$

where $b = c_0^2 t_1^{-1}$.

Thus, again using that f verifies the Kurdyka Lojasiewicz inequality at $\tilde{x}$, it follows that

$$\varphi'(f(x_{k_0}) - f(\tilde{x})) \geq \frac{1}{bd(x_{k_0-1}, x_{k_0})}. \tag{4.9}$$

On the other hand, the concavity of the function φ implies that

$$\varphi(f(x_{k_0}) - f(\tilde{x})) - \varphi(f(x_{k_0+1}) - f(\tilde{x})) \geq \varphi'(f(x_{k_0}) - f(\tilde{x}))(f(x_{k_0}) - f(x_{k_0+1})),$$

which, combined with $\varphi' > 0$ and assumption (4.2) yields

$$\varphi(f(x_{k_0}) - f(\tilde{x})) - \varphi(f(x_{k_0+1}) - f(\tilde{x})) \geq \varphi'(f(x_{k_0}) - f(\tilde{x})) a d^2(x_{k_0}, x_{k_0+1}),$$

remember that $a = 1/2t_2 \leq 1/2\lambda_{k_0-1}$. Therefore, (4.6) follows by combining the last inequality with (4.9).

The proof of the latter part follows from (4.6) combined with Lemma 4.2, which concludes the proof of the lemma. ∎

Lemma 4.4. *Let (x_k) be the sequence concerning to* (4.1) *and assume that assumptions of Lemma 4.3 hold . Then, there exists a $k_0 \in \mathbb{N}$ such that*

$$x_k \in B(\tilde{x}, \rho), \qquad k > k_0. \tag{4.10}$$

Proof. The proof is by induction on k. It follows trivially from of (4.2) that sequence $(f(x_k))$ is decreasing and

$$d(x_k, x_{k+1}) \leq \sqrt{\frac{f(x_k) - f(x_{k+1})}{a}}, \qquad k \in \mathbb{N}. \tag{4.11}$$

Moreover, as the function f is continuous, from Lemma 4.3 it follows that there exists $k_0 \in \mathbb{N}$ such that (4.5), (4.4) hold and, hence

$$x_{k_0} \in B(\tilde{x}, \rho), \qquad 0 < f(x_{k_0}) - f(x_{k_0+1}) < f(x_{k_0}) - f(\tilde{x}), \tag{4.12}$$

which, combined with (4.11) ($k = k_0$), give us

$$d(x_{k_0}, x_{k_0+1}) \leq \sqrt{\frac{f(x_{k_0}) - f(\tilde{x})}{a}}. \tag{4.13}$$

Now, from the triangle inequality, combining with the last expression and (4.5), we obtain

$$d(\tilde{x}, x_{k_0+1}) \leq \sqrt{\frac{f(x_{k_0}) - f(\tilde{x})}{a}} + d(\tilde{x}, x_{k_0}) < \rho,$$

which implies that $x_{k_0+1} \in B(\tilde{x}, \rho)$.

Suppose now that (4.10) holds for all $k = k_0 + 1, \ldots, k_0 + j - 1$. In this case, for $k = k_0 + 1, \ldots, k_0 + j - 1$, it holds (4.6) and, consequently

$$\sqrt{d(x_{k-1}, x_k)(b/a)[\varphi(f(x_k) - f(\tilde{x})) - \varphi(f(x_{k+1}) - f(\tilde{x}))]} \geq d(x_k, x_{k+1}). \tag{4.14}$$

Thus, since for $r, s \geq 0$ it holds $r+s \geq 2\sqrt{rs}$, considering, for $k = k_0+1, \ldots, k_0+j-1$

$$r = d(x_{k-1}, x_k), \;\; s = (b/a)[\varphi(f(x_k) - f(\tilde{x})) - \varphi(f(x_{k+1}) - f(\tilde{x}))],$$

from the inequality (4.14), it follows, for $k = k_0 + 1, \ldots, k_0 + j - 1$, that

$$2d(x_k, x_{k+1}) \leq d(x_{k-1}, x_k) + \frac{b}{a}[\varphi(f(x_k) - f(\tilde{x})) - \varphi(f(x_{k+1}) - f(\tilde{x}))].$$

So, adding member to member, with $k = k_0 + 1, \ldots, k_0 + j - 1$, we obtain

$$\sum_{i=k_0+1}^{k_0+j-1} d(x_i, x_{i+1}) + d(x_{k_0+j-1}, x_{k_0+j}) \leq d(x_{k_0}, x_{k_0+1}) + \frac{b}{a}[\varphi(f(x_{k_0+1}) - f(\tilde{x})) - \varphi(f(x_{k_0+j}) - f(\tilde{x}))],$$

from which we obtain

$$\sum_{i=k_0+1}^{k_0+j-1} d(x_i, x_{i+1}) \leq d(x_{k_0}, x_{k_0+1}) + \frac{b}{a}\varphi(f(x_{k_0+1}) - f(\tilde{x})) \leq d(x_{k_0}, x_{k_0+1}) + \frac{b}{a}\varphi(f(x_{k_0}) - f(\tilde{x})) \tag{4.15}$$

where the last inequality follows of the second inequality in (4.12) and because φ is increasing. Now, using the triangle inequality and taking into account that $d(x, z) \geq 0$ for all $x, z \in M$, we have

$$
\begin{aligned}
d(\tilde{x}, x_{k_0+j}) &\leq d(x_{k_0}, x_{k_0+j}) + d(\tilde{x}, x_{k_0}) \\
&\leq d(\tilde{x}, x_{k_0}) + d(x_{k_0}, x_{k_0+1}) + \sum_{i=k_0+1}^{k_0+j-1} d(x_i, x_{i+1}),
\end{aligned}
$$

which, combined with (4.15), yields

$$
d(\tilde{x}, x_{k_0+j}) \leq d(\tilde{x}, x_{k_0}) + 2d(x_{k_0}, x_{k_0+1}) + \frac{b}{a}\varphi(f(x_{k_0}) - f(\tilde{x})).
$$

Therefore, from the last inequality, combined with (4.13) and (4.5), we conclude that $x_{k_0+j} \in B(\tilde{x}, \rho)$, which completes the induction proof. ■

In the following theorem, using our method, we prove the full convergence of the sequence (x_k) to a critical point, for functions which satisfy the Kurdyka-Lojasiewicz property in that point.

Theorem 4.1. *Let U, η and $\varphi : [0, \eta) \to \mathbb{R}_+$ be the objects appearing in the Definition 3.1. Assume that $x_0 \in M$, $\tilde{x} \in M$ is an accumulation point of the sequence (x^k), $\rho > 0$ is such that $B(\tilde{x}, \rho) \subset U$ and f satisfies the Kurdyka-Lojasiewicz inequality at $\tilde{x}$. Then there exists $k_0 \in \mathbb{N}$ such that*

$$
\sum_{k=k_0}^{+\infty} d(x_k, x_{k+1}) < +\infty. \tag{4.16}
$$

Moreover, $f(x_k) \to f(\tilde{x})$, as $k \to +\infty$, and the sequence (x_k) converges to $\tilde{x}$ and $\tilde{x}$ is a critical point of f.

Proof. Note that Lemma 4.4 combined with Lemma 4.3 implies that (4.16) holds and, in particular, that the sequence (x_k) converge to $\tilde{x} \in M$. Thus, using the continuity of f combined with (4.2), it follows $f(x_k) \to f(\tilde{x})$, as $k \to +\infty$. Now, combining (4.16) with Lemma 4.1,

$$
F(gradf(x_k)) \leq bd(x_k, x_{k+1}) \to 0,
$$

for $k = k_0$, it follows that $F(gradf(\tilde{x})) = 0$. Therefore, $\tilde{x}$ is a critical point of f. ■

Theorem 4.2. *Assume hypotheses of Lemma 4.4. Assume further that (x_k) converges to x^* and that f has the Kurdyka-Lojasiewicz property at x^* with $\varphi(s) = cs^{1-\theta}$, $\theta \in [0, 1)$, $c > 0$. Then the following estimations hold*

(i) If $\theta = 0$ then the sequence (x_k) converges in a finite number of steps;

(ii) If $\theta \in (0, \frac{1}{2}]$ then there exist $b_0 > 0$ and $\varsigma \in [0, 1)$ such that

$$d(x_k, x^*) \leq b_0 \varsigma^k;$$

(iii) If $\theta \in (\frac{1}{2}, 1)$ then there exist $\xi > 0$ such that

$$d(x_k, x^*) \leq \xi k^{-\frac{1-\theta}{2\theta-1}}.$$

Proof. The notations are those of Lemma 4.3 and for simplicity we assume that $f(x_k) \to 0$. Then by Proposition 4.1 $f(x^*) = 0$.

(i) If $f(x_k)$ is stationary, then so is (x_k) in view of Proposition 4.1. If (f_k) is not stationary, then the Kurdyka-Lojasiewicz inequality yields for any k sufficiently large, i.e, $c\,F(gradf(x_k)) \geq 1$, on the other hand, since the sequence $x_k \to x^*$ and $F(gradf(x_k))$ is continuous, we obtain that $c\,F(gradf(x_k)) \to 0$ which is a contradiction.

Now to complete the proof of the theorem consider any $k \geq 0$, set

$$D(x_k) := \sum_{i=k+1}^{+\infty} d(x_i, x_{i+1}),$$

which is finite by Theorem 4.1. Now, fixed $k \in \mathbb{N}$, let $j \in \mathbb{N}$, $j > k$, thus by using triangle inequality,

$$d(x_k, x^*) \leq d(x_k, x_j) + d(x_j, x^*) \leq \sum_{i=k+1}^{j} d(x_i, x_{i+1}) + d(x_j, x^*),$$

making $j \to +\infty$ in the last inequality, we have

$$d(x_k, x^*) \leq \sum_{i=k+1}^{+\infty} d(x_i, x_{i+1}) = D(x_k),$$

it is sufficient to estimate $D(x_k)$. From of (4.15) we obtain

$$D(x_k) \leq \varphi(f(x_k))E + D(x_{k-1}) - D(x_k), \tag{4.17}$$

where $E = b/a$. Now the Kurdyka-Lojasiewicz inequality successively yields

$$1 \leq \varphi'(f(x_k))F(gradf(x_k)) = c(1-\theta)f(x_k)^{-\theta}F(gradf(x_k)),$$

thus

$$(f(x_k))^{\theta} \leq c(1-\theta)F(gradf(x_k)) \leq c(1-\theta)b(D(x_{k-1}) - D(x_k))$$

and therefore,

$$(f(x_k))^{1-\theta} \leq [c(1-\theta)b]^{\frac{1-\theta}{\theta}}(D(x_{k-1}) - D(x_k))^{\frac{1-\theta}{\theta}}.$$

Using the previous two inequalities and taking $\Theta = c[c(1-\theta)b]^{\frac{1-\theta}{\theta}}$, we obtain

$$\varphi(f(x_k)) = c(f(x_k))^{1-\theta} \leq \Theta(D(x_{k-1}) - D(x_k))^{\frac{1-\theta}{\theta}}.$$

Therefore (4.17) gives

$$D(x_k) \leq E\Theta(D(x_{k-1}) - D(x_k))^{\frac{1-\theta}{\theta}} + (D(x_{k-1}) - D(x_k)). \tag{4.18}$$

(ii) Due to (4.18) and the facto of $\theta \in (0, 1/2]$ we get that there exists a positive constant c_1 such that

$$D(x_k) \leq c_1(D(x_{k-1}) - D(x_k)),$$

for k sufficiently large, where $c_1 = E\Theta + 1$. This yields

$$D(x_k) \leq \frac{c_1}{1+c_1} D(x_{k-1}).$$

Therefore, using recurrence in k, we obtain

$$d(x_k, x^*) \leq D(x_k) \leq b_0 \varsigma^k,$$

where $\varsigma = c_1/(1+c_1)$ and $b_0 = D(x_{k_0})$ is a positive constant.

(iii) We use identical arguments of Attouch and Bolte in [25, Theorem 2]. First, by (4.18) there exists an integer $n_1 > n_0$ and $c_2 > 0$ such that

$$D(x_k)^{\frac{\theta}{1-\theta}} \leq c_2(D(x_{k-1}) - D(x_k)),$$

for all $k \geq n_1$, where $c_2 = (E\Theta + 1)^{\frac{\theta}{1-\theta}}$. Second, from the fact that $\theta \in (1/2, 1)$ we derive a constant $c_3 > 0$ such that

$$0 < c_3 \leq D(x_k)^{\frac{1-2\theta}{1-\theta}} - D(x_{k-1})^{\frac{1-2\theta}{1-\theta}}.$$

Finally for n greater than n_1,

$$c_3(n - n_1) \leq D(x_n)^{\frac{1-2\theta}{1-\theta}} - D(x_{n_1})^{\frac{1-2\theta}{1-\theta}},$$

then

$$d(x_n, x^*) \leq D(x_n) \leq [c_3(n-n_1) + D(x_{n_1})^{\frac{1-2\theta}{1-\theta}}]^{\frac{1-\theta}{1-2\theta}} \leq \xi n^{-\frac{1-\theta}{2\theta-1}},$$

where $\xi = (c_3+1)^{-\frac{1-\theta}{2\theta-1}} > 0$. ■

5 The ”effort-accuracy” trade off: an ”exploration-exploitation” model on a Finsler manifold

In this section and the last we show how a proximal algorithm on a Finsler manifold can modelize the ”effort-accuracy” trade off, using the recent ”variational approach” of the theories of change (Soubeyran [11, 12]), see also Attouch-Soubeyran [13, 14], for more specific cases).

Finsler distances as costs to change Consider a given period and an agent who spends the energy $E'(x') \in \mathbb{R}_+$ to gather and use a stock $x' = (x^h, h \in H) \in X' = \mathbb{R}^H$ of ressources (means) to produce the output $q(x') = \mathbf{q}[x', E'(x')] \in Q$ and get the benefit $b(x') = \mathbf{b}[q(x')] \in \mathbb{R}_+$.To regenerate, next period, this energy lost to produce an output this period, the agent must eat, rest, spend holidays, have hobbies, make healthy activities..... Doing that, he spends the energy $E_2''(x'') \in \mathbb{R}_+$ to gather and use the stock of resources $x'' = (x^k, k \in K) \in X'' = \mathbb{R}^K$.to produce the energy $E_1''(x'') \in \mathbb{R}_+$. The net energy produced is $E''(x'') = E_1''(x'') - E_2''(x'') \geq 0$. The desutility of the consumed energy is $\delta[E'(x') + E_2''(x'')] = c(x) \in \mathbb{R}_+$, where the vector of stocks of resources used this period is $x = (x', x'') = (x^i, i \in I) \in X = X'.X''$.Then, the payoff of the agent is $g(x) = b(x') - c(x'')$.

An important constraint, almost always neglected in the economic literature is that the agent can conserve (regenerate) his energy as time evolves. Then, the regeneration of vital energy imposes the constraint $E''(x'') - E'(x') \geq E$ where $E > 0$ is an extra amount of energy used, as we will see, for other tasks. This defines a manifold $M = \{x \in X, E''(x'') - E'(x') \geq E > 0\}$.

Production and consumption functions of energy can be quadratic (the more an agent does an activity, the more he produces and consumes energy, at an increasing rate). In this case $\Sigma_{k\in K}(x^k)^2 - \Sigma_{h\in H}(x^h)^2 = E > 0$ defines an hyperboloid. A more realistic example can be given where production functions of energy are increasing, concave, and consumption functions of energy are increasing convex. More generally each activity can both consume and produce some energy.

Starting from $x \in M$, the agent spends efforts $e = (\ e^i, i \in I) \in E = \mathbb{R}_+^I = T_xM$ to change his stocks of resources, passing from x to $y \in M$, such that $y - x = \Psi(x, e) \in \mathbb{R}^I$. To simplify, we will take $y - x = e$, where $e \in E$ is the vector of efforts necessary to pass from x to y. An added energy constraint $E \geq \Sigma_{i\in I} e^i$ must also be satisfied when the agent passes from one period to the next. For simplicity we will suppose that it is always satisfied (along the process this constraint will be more and more easily satisfied because $\Sigma_{i\in I} e_n^i = \Sigma_{i\in I}(x_{n+1}^i - x_n^i)$ converges to zero, $n \longrightarrow +\infty$).

Anchored to x, the costs of these efforts are $F(x, e) \in \mathbb{R}_+$ where $F(x, le) = lF(x, e), l > 0$ (the costs of efforts increase proportionally to effort levels and $F(x, e) = 0 \Longleftrightarrow e = 0$ (no effort implies no cost to do an effort and reciprocally).

We impose a convexity condition $F(x, e + e') \leq F(x, e) + F(x, e')$, for all $e, e' \in E = T_xM$. This means that, if $\delta_x(x, y) = F(x, y - x)$ and if $y - x = e$ and $z - y = e'$, then, $\delta_x(x, z) \leq \delta_x(x, y) + \delta_x(y, z)$. Then, $\delta_x(x, y) \geq 0$ represents, starting from x, a quasi distance between x and y. Fixed costs of moving $\delta_x(x, y) = \delta_x > 0$ for all $y \neq x$ are a special case.

Let $p \in P(x, y)$ be a path of resources $p = \{x_0, x_1, .., .x_h = \varphi(t_h), ..x_n\} \subset M$ from $x = x_0$ to $y = x_n$, where $x_h = \varphi(t_h), a = t_0 < t_1 < .. < t_h < ... < t_n = b, \varphi(a) = x, \varphi(b) = y$. Following this path of change, the total cost to change from x to y is

$\mathbf{F}(x, y, p) = \Sigma_{h=0}^{n-1} F(x_h, e_h)$ where $e_h = x(t_{h+1}) - x(t_h)$. As an idealization, consider paths of change from x to y as smooth curves $\varphi = \{\varphi(t), a \leq t \leq b\} \in$

$P(x,y) \subset M$, parametrized by the map $z = \varphi(t), a \leq t \leq b$. Let $\dot{\varphi}(t) = e(t)$ be the efforts spend to pass from $\varphi(t)$ to $y(t) = x(t) + \dot{\varphi}(t)$ where $L_F[\varphi] = \int_a^b F(\varphi(\tau), \dot{\varphi}(\tau))d\tau$ is the cost to pass from the stock of resources $x = \varphi(a)$ to the stock $y = \varphi(b)$ following the path of change $\varphi \in \Phi$. Then $C(x,y) = \inf\{L_F(\varphi), \varphi \in \Phi\}$ represents the infimum cost to change from x to y. This defines a quasi distance where $C(y,x) \neq C(x,y)$ for some x, y. This formulation defines Finsler quasi distances as costs to change on a manifold. The general concept of costs to change in behavioral sciences (economics, management, psychology) comes from Soubeyran [11, 12]. See also Attouch-Soubeyran [14] and Moreno et al. [39] for an interpretation as regularization terms for proximal algorithms.

An "exploration-exploitation" model In the present model, the problem of the agent is, each period, a version of the famous "exploration-exploitation" problem (see Soubeyran [11, 12], for an extensive formulation). Starting from $x \in M$, the agent must find on the manifold M,

i) some new stock of resources $y \in M$ to exploit them $\mu(x) > 0$ units of time. This is the "exploitation phase" of the period. Its length is $\mu(x) > 0$, and it generates the cumulated exploitation payoff $\mu(x)g(y)$.

ii) a path of change $p \in P(x,y)$ to be able to move from x to y. This is the "exploration phase" which lasts $b - a > 0$ times and costs $\mathbf{F}(x,y,p) \geq 0$. In this case the term exploration is somewhat misleading. Here, it represents both the discovery and the evaluation of the path of resources $p \in P(x,y)$ and a way, starting from x, to build step by step, the stocks of resources y.

Each step, the agent must balance between a costly phase of exploration and a rewarding phase of exploitation, ie a low enough cost to change $\mathbf{F}(x,y,p)$ and a high enough cumulated payoff $\mu(x)g(y)$.

Let $D[\mathbf{F}(x,y,p)] \geq 0$ be, each period, the desutility of costs to change and $U[\mu(x)g(y)]$ be the utility of the net payoff, where $D[\mathbf{F}] \geq 0, D[0] = 0$ and $D[.]$ is not decreasing, $U[Z] \geq 0, U[0] = 0$, $U[.]$ is not decreasing. Then, the net payoff of the agent over the period is $J(x,y,p) = U[\mu(x)g(y)] - D[\mathbf{F}(x,y,p)]$ .In this paper we will suppose $U[Z] = Z$ and $D[\mathbf{F}] = \mathbf{F}^2$.Then, $J(x,y,p) = \mu(x)g(y) - [\mathbf{F}(x,y,p)]^2$.

For a given period, the per unit of time net payoff is $G(x,y,p) = g(y) - \lambda(x)\mathbf{F}(x,y,p)^2$ where $\lambda(x) = 1/\mu(x) > 0$ defines a weight over costs to change.

The "exact exploration-exploitation" problem of the agent is to solve, each step, the optimization problem (see Soubeyran [11, 12]):

find $x_{n+1} \in M, p_{n+1} \in P(x_n, y)$ such that

$G(x_n, x_{n+1}, p_{n+1}) = \sup\{G(x_n, y, p), y \in M, p \in P(x_n, y)\} < +\infty$,

where $G(x_n, y, p) = g(y) - \lambda(x_n)\mathbf{F}(x_n, y, p)^2$, with $\lambda(x_n) > 0$. The inexact case is to solve, each step, this problem approximatively.

In the exact case, the agent can follow, each period, a two steps process:

i) given each new stock of resource $y \in M$, solve

$$C(x_n, y) = \inf\{\mathbf{F}(x_n, y, p), p \in P(x_n, y)\},$$

ie find $p_{n+1} \in P(x_n, y)$ such that $C(x_n, y) = \mathbf{F}(x_n, y, p_{n+1})$,

ii) Define $\Gamma(x_n, y) = g(y) - \lambda(x_n)C(x_n, y)^2$ be the proximal payoff of the agent at ($x_n, \lambda_n = \lambda(x_n)$).Then, find a stock of resources $x_{n+1} \in M$ which solves $\Gamma(x_n, x_{n+1}) = \sup\left\{g(y) - \lambda_n C(x_n, y)^2, y \in M\right\}$.

Then, this "exact exploration-exploitation" problem can be identifed with a proximal problem on a Finsler manifold M: starting from $x_0 \in M$, solve, each step, $\sup\{\Gamma(x_n, y), y \in M\}$. This works as follows. Let $\overline{g} = \sup\{g(y), y \in M\} < +\infty$. Then, $f(y) = \overline{g} - g(y) \geq 0$ represents the unsatisfied need off the agent at y. An equivalent formulation of the problem is to say that the agent wants, each step to minimize his unsatisfaction, taking care of costs to change. The agent solves the sequence of iterations $x_{n+1} \in \arg\min\left\{f(y) + \lambda_n C(x_n, y)^2, y \in M\right\}$, where $C(x, y) \geq 0$ is a Finsler quasi distance on M.

Comments 1. Proximal algorithms can be seen as repeated optimization problems with costs to change. In the bounded rationality perspective (Simon [40, 41]) where it seems difficult for agents to optimize, this poses the famous "recursive problem": how costly it is to know the costs to optimize ?,... This opens the door to inexact proximal algorithm formulations which will be more economizing in term of costs to change (given that optimizing each step will not be an optimizing strategy on the long run !).

Comments 2. in our model repeated optimization is necessary because preferences change each step. Preferences are variable,

$z \geq_x y \Longleftrightarrow \Gamma(x, z) \geq \Gamma(x, y)$ (Soubeyran [12])). Then, the agent enters in a course pursuit where, each step, the agent chooses and builds his optimal stock of resources with respect to his present preference, which gives rise to a new preference and a new optimal stock of resource, which....Variable preferences come from inertia (costs to change, and, each period, a variable length $\mu(x) = 1/\lambda(x) > 0$ of the exploitation phase). Inertia is strong because of high costs to change in the small (costs to change represents a quasi distance). This greatly helps for convergence.

Behavioral traps A stock of resources $x^* \in M$ is a (behavioral) "trap" (see Soubeyran [11, 12]) if, starting from itself, the agent prefers to conserve this stock of resources than to change. Define the proximal payoff of the agent at x^* as $\Gamma(x^*, y) = g(y) - \lambda(x^*)C(x^*, y)^2$. Then $x^* \in M$ is a trap iff

$\Gamma(x^*, x^*) \geq \Gamma(x^*, y)$ for all $y \in M$ ie $g(x^*) \geq g(y) - \lambda(x^*)C(x^*, y)^2$ for all $y \in M$. This trap is strict iff $\Gamma(x^*, x^*) > \Gamma(x^*, y)$ for all $y \in M, y \neq x^*$.

A global maximum $\widehat{x} \in M$ of $g(.)$ is a trap if the agent ignores costs to change. In our case the agent takes care of inertia and learning aspects (both embedded in costs to change). At $\widehat{x}$, unsatisfied needs disappear: $g(\widehat{x}) = \overline{g} = \sup\{g(y), y \in M\}$.

If, as we have shown, the process $\{x_n, n \in N\}$ of stocks of resources converges, $x_n \longrightarrow x^*, n \longrightarrow +\infty$, then, the agent uses, each period, more and more similar stocks of resources which render more and more similar services $q(x_n)$. He enters in an habituation process. Let $\lambda(x^*) = \lambda^* > 0$ and $\lambda(x_n) = \lambda_n > 0$. Suppose that the weight $\lambda(.) : x \in M \longrightarrow \lambda(x) > 0$ is a continuous function.

Then, $\lambda_n \longrightarrow \lambda^*, n \longrightarrow +\infty$.

For each n, $\Gamma(x_n, x_{n+1}) = g(x_{n+1}) - \lambda_n C(x_n, x_{n+1})^2 \geq \Gamma(x_n, y) = g(y) - \lambda_n C(x_n, y)^2$ for all $y \in M$, because $x_{n+1} \in \arg\max \{\Gamma(x_n, y), y \in M\}$.

In this paper, we have given conditions such that the limit x^* is a critical point of $g(.)$, ie a critical point of the unsatisfied needs $f(y) = \overline{g} - g(y) \geq 0$. Does this means that x^* is a trap, ie $\Gamma(x^*, x^*) = g(x^*) - \lambda^* C(x^*, x^*) = g(x^*) \geq \Gamma(x^*, y) = g(y) - \lambda^* C(x^*, y)^2$ for all $y \in M$?

This is true if $g(.)$, $\lambda(.)$ and $C(x, .)$ are continuous functions. In this case $\Gamma(x^*, x^*)^{(1)} = \lim_{n \longrightarrow +\infty} \Gamma(x_n, x_{n+1}) \geq^{(2)} \lim_{n \longrightarrow +\infty} \{\Gamma(x_n, y), y \in M\} =^{(3)} \Gamma(x^*, y)$

Equalities (1) and (3) come from the continuity assumption, while inequality (2) comes from $\Gamma(x_n, x_{n+1}) \geq \Gamma(x_n, y)$, for all $y \in M$.

6 The speed of decision making

Let us show how our "exploration-exploitation" interpretation of the proximal algorithm can give precise results on the famous "effort-accuracy" trade off (see Payne-Bettman-Johnson [1]) where the speed of decision making matters when there are costs to change (then, optimizing is costly).

Majoration of the time spend to converge Let us use Lemma 4.3 to get some nice results in term of the time spend to converge towards a critical point.

Majoration of the length of a convergent path.

Let $\sigma(x_k) = \varphi(f(x_k) - f(\tilde{x}))$. Then, Lemma 4.3 gives

$$\frac{b}{a}[\varphi(f(x_{k_0}) - f(\tilde{x})) - \varphi(f(x_{k_0+1}) - f(\tilde{x}))] \geq \frac{d^2(x_{k_0}, x_{k_0+1})}{d(x_{k_0-1}, x_{k_0})},$$

which implies

$$\frac{d^2(x_{k_0}, x_{k_0+1})}{d(x_{k_0-1}, x_{k_0})} \leq \frac{b}{a}\left[\sigma(x_{k_0}) - \sigma(x_{k_0+1})\right].$$

Let $a_k = d(x_k, x_{k+1})$. Then, the inequality

$$(\Sigma_{k=1}^{j} a_{k-1})^{1/2} \leq (a_0)^{1/2} + (\Sigma_{k=1}^{j} a_k^2 / a_{k-1})^{1/2}$$

implies

$$(\Sigma_{k=k_0+1}^{n} a_{k-1})^{1/2} \leq (a_{k_0})^{1/2} + (\Sigma_{k=k_0+1}^{n} a_k^2 / a_{k-1})^{1/2}.$$

It follows that

$$(\Sigma_{k=k_0+1}^{n} d(x_{k-1}, x_k))^{1/2} \leq (d(x_{k_0}, x_{k_0+1}))^{1/2} + \left[\frac{b}{a}\Sigma_{k=k_0+1}^{n}\left[\sigma(x_k) - \sigma(x_{k+1})\right]\right]^{1/2}$$

i.e.,

$$(\Sigma_{k=k_0+1}^{n} d(x_{k-1}, x_k))^{1/2} \leq (d(x_{k_0}, x_{k_0+1}))^{1/2} + \left[\frac{b}{a}\left[\sigma(x_{k_0+1}) - \sigma(x_n)\right]\right]^{1/2}$$

Let $X_n = \{x_{k_0}, x_{k_0+1}, ..., x_{k_0+n}\} \subset X$ be the "n first steps"

$$\{x_h, h = k_0, k_0 + 1, ...\}$$

of the whole path $\{x_n\}$. Let $L\{X_n\} = \Sigma_{k=k_0+1}^{n} d(x_{k-1}, x_k)$ be its length. Then,

$$L\{X_n\} \leq d(x_{k_0}, x_{k_0+1}) + \frac{b}{a}\left[\sigma(x_{k_0+1}) - \sigma(x_n)\right] + 2\left[d(x_{k_0}, x_{k_0+1})\right]^{1/2}\left[\frac{b}{a}\left[\sigma(x_{k_0+1}) - \sigma(x_n)\right]\right]^{1/2}$$

which implies,

Corollary 6.1. *The time spend to converge along a convergent path is lower than*

$$L\{X_n\} \leq d(x_{k_0}, x_{k_0+1}) + \frac{b}{a}\left[\sigma(x_{k_0+1})\right] + 2\left[d(x_{k_0}, x_{k_0+1})\right]^{1/2}\left[\frac{b}{a}\left[\sigma(x_{k_0+1})\right]\right]^{1/2}.$$

Majoration of the time spend to converge

Let $t(x, y) \geq 0$ be the time spend to move from x to y, where $t(x, y) > 0$ if $y \neq x$. Let $v = d(x, y)/t(x, y) > 0, y \neq x$, be the speed of decision-making. If the speed of decision-making $v > 0$ is constant, the time spend to move is proportional to the distance, $t(x, y) = v^{-1}d(x; y)$. Then, the time spend to move along the n first steps of the path $\{x_k\}$ is $T(X_n) = vL\{X_n\}$.

Corollary 6.2. *Suppose that the speed of decision-making $v > 0$ is constant. Then, the time spend to converge $T(X_{+\infty}) = vL\{X_{+\infty}\}$ is lower than*

$$T(X_{+\infty}) \leq vd(x_{k_0}, x_{k_0+1})\left\{1 + \frac{b}{a}\left[\frac{\sigma(x_{k_0+1})}{d(x_{k_0}, x_{k_0+1})}\right] + 2\left[\frac{b}{a}\frac{\sigma(x_{k_0+1})}{d(x_{k_0}, x_{k_0+1})}\right]^{1/2}\right\},$$

i.e., the relative time spend to converge relative to the first step is lower than

$$\frac{T(X_{+\infty})}{t(x_{k_0}, x_{k_0+1})} \leq \left\{1 + \frac{b}{a}\left[\frac{\sigma(x_{k_0+1})}{d(x_{k_0}, x_{k_0+1})}\right] + 2\left[\frac{b}{a}\right]^{1/2}\left[\frac{\sigma(x_{k_0+1})}{d(x_{k_0}, x_{k_0+1})}\right]^{1/2}\right\}.$$

Comments:

We find a really striking result: the lower the first step rate of improvement $\sigma(x_{k_0+1})/d(x_{k_0}, x_{k_0+1}) = \left[\varphi(f(x_k) - f(\tilde{x}))/d(x_{k_0}, x_{k_0+1})\right] > 0$, the lower be the time spend to converge. This means that how the agent makes the first step matters much. The agent must not try to improve too much at the very beginning to hope to converge quickly to the solution. We confirm the italian sentence " que va piano va sano ", at least at the beginning !

Furthermore, from $a = 1/2t_2$ and $b = c_0^2/t_1$ the lower the ratio $b/a = 2c_0^2 t_2/t_1$, ie the lower c_0 and t_2, and the higher t_1, the lower the time spend to converge relative to the first step. Then, from the inequalities $0 < t_1 \leq \lambda_k \leq t_2 \Longleftrightarrow a = 1/2t_2 \leq 1/2\lambda_k \leq 1/2t_1$, the higher t_1, the lower the weight $1/2\lambda_k$ putted on costs to change, the lower the time spend to converge. Hence, less inertia speeds the time of convergence. A nice result too.

Residual time spend to converge Let us comment the implications of Theorem 4.2 in term of the residual time spend to converge $t(x_k, x^*) = v\, d(x_k, x^*)$.

The case $\theta = 1$ gives convergence in a finite number of steps, a very nice result for decision-making problems.

The condition $\theta = 1$ means that, when $\varphi(s) = cs^{1-\theta}$, the Kurdyka-Lojaseviecz inequality

$$|f(x) - f(x^*)| \leq c\, |grad\, f(x)|^{1/\theta}\,,$$

gives

$$|f(x) - f(x^*)| \leq c\, |grad\, f(x)|\,.$$

Then, if the norm of the gradient $|grad\, f|$ at x is low, the residual gap

$$|f(x) - f(x^*)|$$

to be filled, moving from x to x^*, is low.

Hence, a low norm of the gradient at x gives to the agent the important indication that he has no big gains to hope, moving from x to x^*.

The case $0 < \theta \leq 1/2$ shows that, starting from x_k, the residual time spend to converge is lower than $t(x_k, x^*) = v\, d(x_k, x^*) \leq v\, D(k_0)\tau^k$, where collecting terms from several parts of the papers, $\varsigma = c_1/(1 + c_1)$, $c_1 = E\Theta + 1$, $\Theta = c\,[c((1-\theta)b]^{(1-\theta)/\theta}$, $D(k_0) = \Sigma_{i=k_0+1}^{+\infty} d(x_i, x_{i+1})$ and $E = b/a = 2c_0^2 t_2/t_1$. This shows that the higher t_1, the less inertia, i.e. the lower the weight $1/2\lambda_k$ on costs to change, (using the inequalities $1/2t_2 \leq 1/2\lambda_k \leq 1/2t_1$). Then, the lower ς, the lower the residual time spend to converge. Furthermore, the lower $c > 0 \Longrightarrow$ the lower $\Theta \Longrightarrow$ the lower $c_1 \Longrightarrow$ the lower $\tau \Longrightarrow$ the lower the residual time spend to converge. The lower $c > 0$ means a low improvement $|f(x) - f(x^*)|$ to hope, moving from x to x^* : the lower c, the lower $|f(x) - f(x^*)| \leq c\,|g|^{1/\theta}$ for a given norm $|grad\, f(x)|$ of the gradient.

Similar comments can be made for the case $1/2 < \theta < 1$.

7 Conclusion

We present and analyze the proximal point method on Finslerian manifolds for minimizing differentiable KL functions. We use intrinsic relations between the o-minimal structure and analytic-geometric categories to ensure the existence of KL functions. We derive important theoretical results of convergence (Theorem 4.1) and convergence rate (Theorem 4.2) in the Finslerian context. These results are applied in our "exploration-exploitation" model and have a very interesting interpretation. Starting from x_k the length of is $D(x_k) = \Sigma_{i=k+1}^{+\infty} d(x_i, x_{i+1})$. Our main Theorem (Theorem 4.1) shows that $D(x_k) < +\infty$. As a consequence of the convergence results, we show that our application, an " effort-accuracy trade off" model, is naturally connected with proximal point algorithm and Finsler quasi distances.

References

[1] Payne, J. W., Bettman, J. R. and Johnson, E. J., 1993. *The Adaptive Decision Maker*. Cambridge University Press, Cambridge, UK.

[2] Rinkenauer, G., Osman, A., Ulrich, R., Müller-Gethmann, H., Mattes, S.: On the locus of speed-accuracy trade-off in reaction time: inferences from the lateralized readiness potential. Journal of Experimental Psychology **133**, 261-282 (2004)

[3] Holland, J.H.: Adaptation in natural and artificial systems. University of Michigan Press, Ann Arbor, MI (1975)

[4] March, J. G.: Exploration and exploitation in organizational learning. Organization Science **2**, 71-87 (1991)

[5] Levitt, B., March, J. G.: Organizational learning. Annual Review of Sociology **14**, 319-340 (1988)

[6] Cohen, W. M., Levinthal, D. A.: Absorptive Capacity: A New Perspective on Learning and Innovation. Administrative Science Quarterly **35**, 128-152 (1990)

[7] Levinthal, D. A., March, J. G.: The myopia of learning. Strategic Management Journal **14**, 95-112 (1993)

[8] Fu, W. T., Gray, W. D.: Suboptimal tradeoffs in information seeking. Cognitive Psychology **52**, 195-242 (2006)

[9] Fu, W. T.: A Rational-Ecological Approach to the exploration/exploitation trade-offs: Bounded rationality and suboptimal performance. Oxford University Press, USA (2007)

[10] Busemeyer, J. R., Townsend, J. T.: Decision Field Theory: A dynamic cognition approach to decision making. Psychological Review **100**, 432-459, (1993)

[11] Soubeyran, A.: Variational rationality, a theory of individual stability and change: worthwhile and ambidextry behaviors. Pre-print. GREQAM, Aix Marseillle University, (2009).

[12] Soubeyran A.: Variational rationality and the "unsatisfied man": routines and the course pursuit between aspirations, capabilities and beliefs. Preprint. GREQAM, Aix Marseillle University (2010)

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

[14] Attouch H., Soubeyran A.: Local search proximal algorithms as decision dynamics with costs to move, Set-Valued and Variational Analysis, DOI 10.1007/s11228-010-0139-7 (2010)

[15] Matsumoto, M.: A slope of a mountain is a Finsler surface with respect to a time measure. J. Math. Kyoto Univ. **29** (1), 17–25 (1989)

[16] Bao, D., Chern, S.S., Shen, Z.: An Introduction to Riemann–Finsler Geometry. Graduate Texts in Mathematics, vol. 200. Springer, Berlin (2000)

[17] Ferreira, O., Oliveira, P. R.: Proximal point algorithm on Riemannian manifolds, Optimization **51**, 257–270 (2000)

[18] Li, C., Lopez, G., Martin-Marquez, V.: Monotone vector fields and the proximal point algorithm on Hadamard manifolds, J. London Math. Soc. **79** (2), 663-683. (2009)

[19] Fukushima, M., Mine, H.: A generalized proximal point algorithm for certain nonconvex minimization problems. Int. J. Systems Sci. **12**(8), 989-1000 (1981)

[20] Papa Quiroz, E. A., Oliveira, P. R.: Proximal point methods for quasiconvex and convex functions with Bregman distances on Hadamard manifolds. J. Convex Anal. **16**(1), 49-69 (2009)

[21] Bento, G. C., Ferreira O. P., Oliveira P. R.: Local convergence of the proximal point method for a special class of nonconvex functions on Hadamard manifolds. Nonlinear Analysis **73**, 564-572 (2010)

[22] Kurdyka, K.: On gradients of functions definable in o-minimal structures. Ann. Inst. Fourier **48**(3), 769–783 (1998)

[23] Lojasiewicz, S.: Une propriété topologique des sous-ensembles analytiques réels, in Les Équations aux Dérivé es Partielles. Éditions du centre National de la Recherche Scientifique. Paris 87–89 (1963)

[24] Bolte, J., Daniilidis, A. and Lewis, A.: The Lojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems. SIAM J. Optim. **17** (4) 1205–1223 (2006)

[25] Attouch, H., Bolte, J.: On the convergence of the proximal algorithm for nonsmooth functions involving analytic features. Math. Program. Ser. B **116**(1-2), 5–16 (2009)

[26] Bolte, J., Daniilidis, A., Lewis, A., Shiota, M.: Clarke subgradients of stratifiable functions. SIAM J. Optim. **18**(2), 556–572 (2007)

[27] Lageman, C.: Convergence of gradient-like dynamical systems and optimization algorithms. Dissertation zur Erlangung des naturwissenschaftlichen Doktorgrades der Bayerischen Julius-Maximilians-Universitat Wurzburg (2007)

[28] Kurdyka, K. , Mostowski, T., Parusinski, A.: Proof of the gradient conjecture of R. Thom. Annals of Mathematics. **152**(3), 763-792 (2000)

[29] Da Cruz Neto, J. X., Lima, L. L., Oliveira, P. R.: Geodesic algorithm in Riemannian manifolds. Balkan Journal of Geometry its Applications. **3**(2), 89-100 (1998)

[30] Udriste, C.: Convex functions and optimization methods on Riemannian manifolds, Mathematics and its Applications, 297, Kluwer Academic Publishers,(1994)

[31] Absil, P. A., Mahony, R., Sepulchre, R.: Optimization Algorithms on Matrix Manifolds. Princeton University Press, Princeton (2008)

[32] Kristály, A., Morosanu, G., Róth, Á.: Optimal Placement of a Deposit between Markets: Riemann-Finsler Geometrical Approach. J. Optim. Theory Appl. **139**, 263–276 (2008)

[33] Shen, Z.: Lectures on Finsler Geometry. World Scientific Publishing Co. Pte. Ltd., USA (2001)

[34] van den Dries, L., Miller, C.: Geometric categories and o-minimal structures. Duke Mathematical Journal. **84**(2), 497-540 (1996)

[35] Coste, M.: An introduction to O-minimal geometry. Institut de Recherche Mathématique de Rennes (2000)

[36] Attouch,H., Bolte,J., Redont P., Soubeyran, A.: Proximal Alternating Minimization and Projection Methods for Nonconvex Problems: An Approach Based on the Kurdyka-Lojasiewicz Inequality. Math. Oper. Res. **35**(2), 438–457 (2010)

[37] Bierstone, E., Milman, P. D.: Semianalytic and subanalytic sets. Publicactions Mathématiques. **67**, 5-42 (1988)

[38] van den Dries, L.: O-minimal structures and real analytic geometry. In Current developments in mathematics, 1998 (Cambridge, MA), 105-152. Int. Press, Somerville. MA. (1999)

[39] Moreno, F., Oliveira, P., Soubeyran, A.: A Proximal algorithm with quasi distance. Application to habit's formation, in revision in Optimization. (2010)

[40] Simon, H. A.: A Behavioral model of rational choice. The Quarterly Journal of Economics **69**, 99-118 (1955).

[41] Simon, H. A.: Rationality in Psychology and Economics. The Journal of Business, **59**, No. 4,(1986). Part 2: The Behavioral Foundations of Economic Theory. pp. S209-S224