J Optim Theory Appl (2014) 162:873–891
DOI 10.1007/s10957-013-0483-5

Proximal Point Method on Finslerian Manifolds and the "Effort–Accuracy" Trade-off

J.X. Cruz Neto · P.R. Oliveira · P.A. Soares Jr. · A. Soubeyran

Received: 5 July 2011 / Accepted: 10 November 2013 / Published online: 22 November 2013

Abstract In this paper, we consider minimization problems with constraints. We 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 Kurdyka-Lojasiewicz property, then the proximal point method can be naturally extended to solve this class of problems. We prove that the sequence generated by our method is well defined and converges to a critical point. We show how tools of Finslerian geometry, specifically non-symmetrical metrics, can be used to solve non-convex constrained problems in Euclidean spaces. As an application, we give one result regarding decision-making speed and costs related to change.

Keywords Proximal algorithms · Finslerian manifolds · Non-convex optimization · Kurdyka-Lojasiewicz inequality

J.X. Cruz Neto (✉)
CCN, DM, Universidade Federal do Piauí, Terezina, PI 64049-550, Brazil
e-mail: jxavier@ufpi.edu.br

P.R. Oliveira
COPPE-Sistemas, Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ 21945-970, Brazil
e-mail: poliveir@cos.ufrj.br

P.A. Soares Jr.
CCN, DM, Universidade Etadual do Piauí, Terezina, PI 64002-150, Brazil
e-mail: pedrosoares@uespi.br

A. Soubeyran
Aix-Marseille School of Economics, Aix-Marseille University, CNRS & EHESS, Marseille, France
e-mail: antoine.soubeyran@gmail.com

1 Introduction

1.1 The "Effort–Accuracy" Trade-off

It is well known in behavioral sciences that optimization is costly, but what is the scale of the problem? In decision-making theory, and more generally in cognitive sciences, the famous "effort–accuracy" and "speed-accuracy" trade-offs (see Payne et al. [1] and Rinkenauer et al. [2]) examine how animals and agents strike a balance between the quality of a solution (how "good enough" it must be) and the time and effort expended to reach such a "good enough" solution. More generally, biology and behavioral sciences have emphasized the prominent role of the "exploration–exploitation" trade-off. For examples see Holland [3] (biology), March [4] (management Sciences), Levitt and March [5], Cohen and Levinthal [6], Levinthal and March [7], and an impressive list of others (organizational learning), and Fu and Gray [8] and Fu [9] (psychology).

In [4], March 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, innovation, discovery. Exploitation includes such things as refinement, choice, production, efficiency, selection, execution, implementation." The problem is that exploration (searching) and exploitation (of the benefits of searching) very often use competing resources; the more time and resources agents spend on exploration, the less they have for exploitation.

The "effort–accuracy" and "speed–accuracy" trade-offs are specific examples of the "exploration and exploitation" trade-off. Because effort (resources expended in exploration and thinking, such as mental energy, deliberation costs, time compression costs, and reactivity costs) and accuracy (how efficient a solution is) are specific instances of exploration and exploitation activities.

In psychology, many empirical studies have examined the "effort–accuracy" trade-off. Busemeyer and Townsend [10] offered a dynamic-cognitive approach to human decision making, which describes how people make decisions and how a person's preferences evolve with time until a decision is reached, rather than having a fixed preference (decision field theory). Preferences follow a stochastic diffusion process. Their model has been used to predict how humans make decisions under uncertainties, how decisions change under time pressures, how the speed–accuracy trade-off works, and how the context of choice 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 unified several theories of change [11, 12]. He focused on a new and explicit mathematical model of worthwhile "exploration–exploitation" changes that can be applied to "effort–accuracy" models. The last section of our paper applies a specific version of this "worthwhile to change" approach to behavioral sciences. A preliminary approach for modeling "the speed of decision making problem" as an "exploration–exploitation" process in terms of proximal algorithms on a Finslerian manifold has been discussed in [11, 12], and also by Attouch and Soubeyran [13, 14].

In the present paper, the aim is to minimize a differentiable function on a Finslerian manifold. The accuracy (quality) of this solution is the difference between the real value function at the starting point and at a minimum point. Efforts, or decision making costs, are modeled as (infimum) resistances to change, defined as the disutilities of (infimum) costs of change. The Finslerian distance is well adapted to modeling the costs of change as the costs of exploring paths of potential changes, and was first introduced and studied by Matsumoto [15] and Bao et al. [16]. The costs increase with the length of exploration paths, so geodesics play a major role in defining their infimums. In this paper we concentrate on the time taken to converge to a critical point x^*, when starting from x_{k_0}.

1.2 Proximal Algorithms on a Finslerian Manifold

To be able to solve complex problems, agents with limited capabilities create subproblems. This is the logic behind proximal algorithms, where each step entails solving a perturbed optimization problem that should be much easier than the original minimization problem. In [13, 14], the authors interpreted these perturbation terms as the costs of change. More generally, in [11, 12] these terms were considered as some resistance to change. This links proximal algorithms and behavioral sciences in terms of worthwhile changes, because each step balances motivation and resistance to change (inertia). Another application has been noted in the market location problem, where the Finslerian metric was observed to arise naturally (see Kristály et al. [17]).

In the last two decades, several authors have proposed the generalized proximal point algorithm in the Riemannian context. In [18], Ferreira and Oliveira considered the particular case for convex functions in a Hadamard manifold. They proved that the sequence of proximal points converges to a minimizer point. In [19], Li et al. considered the problem of finding a singularity of a multivalued vector field in a Hadamard manifold, and presented a general proximal point method. In [20], Papa Quiroz and Oliveira considered this method for quasi-convex functions in a Hadamard manifold, and proved full convergence of the sequence to a minimizer point. In [21], Bento et al. considered this method for C^1-lower type functions in a Hadamard manifold. They showed that the generated sequence locally converged to a minimizer. From a mathematical point of view, our convergence result (see Theorem 4.1) generalizes the results established in [18, 20] and [21], using Finslerian instead of Riemannian distances.

For non quasi-convex functions in a Hadamard manifold, the convergence of the sequence stems from the fact that the objective function satisfies a well-known property: the Kurdyka-Lojasiewicz inequality. This inequality was introduced by Kurdyka in [22] for differentiable functions definable in an o-minimal structure in Euclidean space (see Sect. 3). Extensions of the Kurdyka-Lojasiewicz inequality (in the context of Euclidean spaces) can be found in Bolte et al. [23], Attouch and Bolte [24], and Bolte et al. [25]. Moreover, in [26], Lageman extended this inequality for analytic manifolds and differentiable $\mathcal{C}$-functions in an analytic-geometric category. In this paper (see Theorem 3.1), we extend [26, Theorem 2.1.22].

This paper is organized as follows. In Sect. 2, we give some elementary facts on Finslerian geometry that are required to understand this paper. In Sect. 3, we present the Kurdyka-Lojasiewicz property in the Finslerian context. In Sect. 4, we extend the proximal point algorithm for problems on Finslerian manifolds, perform a convergence analysis, and show a convergence rate result (under a minor assumption). In

Sect. 5, we present a "exploration–exploitation" model on a Finslerian manifold. In Sect. 6, we use Finslerian quasi-distances to give an application in terms of an effort–accuracy trade-off. Finally, in Sect. 7, we discuss the main results of our paper and present some perspectives for future works.

2 Elements of Finslerian Geometry

We now introduce some fundamental properties and notations of Finslerian manifolds. These basic facts can be found in any introductory book on Finslerian geometry, for example, see Shen [27] or [16].

Let M be a connected m-dimensional C^∞ manifold and let $TM = \{(x, y) : x \in M, y \in T_x M\}$ be its *tangent bundle*.

Definition 2.1 A continuous function $F : TM \to \mathbb{R}_+$ is a Finslerian metric iff it satisfies the following conditions.

(i) F is C^∞ on $TM \setminus \{0\}$, and for each $x \in M$, $F_x := F_{|T_x M}$ is such that $F_x(ty) = tF_x(y)$ for all $t > 0$ and $y \in T_x M$ (i.e., F_x is positively homogeneous of degree one).

(ii) For each $y \in T_x M \setminus \{0\}$, the symmetric bilinear form g_y on $T_x M$ is positive definite, where

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

In some situations, the Finslerian metric F satisfies the criterion $F_x(-y) = F_x(y)$. In this case, we have absolute homogeneity, i.e., $F_x(ty) = |t| F_x(y)$ for all $t \in \mathbb{R}$. The pair (M, F) is a *Finslerian manifold*.

Remark 2.1 From Definition 2.1 it can be derived that:

$$g_y(y, u) := \frac{1}{2} \frac{\partial}{\partial s} \big[F_x^2(y + su)\big]_{s=0} \quad \text{and} \quad g_y(y, y) := F_x^2(y).$$

Example 2.1 (Riemannian metric) Let $g = \{g_x\}_{x \in M}$, where g_x is a positive definite symmetric bilinear form in $T_x M$ 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. In this case, g is called a *Riemannian metric*. This g defines a symmetric Finslerian structure on TM by the mechanism $F_x(y) := \sqrt{g_x(y, y)}$. Every Riemannian manifold (M, g) is therefore a Finslerian manifold.

Example 2.2 (Randers metric) Let $\alpha(y) = \sqrt{a_{ij}(x) y^i y^j}$ 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 $y \in T_x M$,

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

where a_{ij} are the components of the Riemannian metric, and b_i are those of a 1-form. A Randers metric is a Finslerian structure on TM that has the form $F_x(y) := \alpha(y) + \beta(y)$. Note that, because of the presence of the term β, Randers 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.

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

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

For $x, z \in M$, let $\Gamma(x, z)$ denote 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 $d : M \times M \to \mathbb{R}_+$, $d(x, z) = \inf_{\gamma \in \Gamma(x,z)} L(\gamma)$. Of course, we have $d(x, z) \geq 0$ (where equality holds if only if $x = z$), and $d(x, z) \leq d(x, p) + d(p, z)$, for all $x, p, z \in M$ (the triangle inequality, [27, p. 13]).

A Finslerian manifold, (M, F), where F is positively (but perhaps not absolutely) homogeneous of degree one, is said to be *forward geodesically complete* if and only if every geodesic $\sigma : [0, 1] \to M$ parameterized to have constant Finslerian speed can be extended to a geodesic defined on $]0, \infty[$. The Hopf–Rinow theorem (see [16, p. 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 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 $K(P, y)$ is called the *flag curvature* of the flag (P, y) in T_xM, and is 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 $P = \operatorname{span}\{y, u\}$. When F is a Riemannian metric, the *flag curvature* coincides with the sectional curvature. Let K be a collection of flag curvatures $K(P, y)$ with $y \neq 0$. We say that the flag curvature of (M, F) is non-positive iff $K(P, y) \leq 0$ for any flag (P, y).

Definition 2.2 A Finslerian space (M, F) is called a Hadamard manifold iff it is forward geodesically complete, simply connected, and has non-positive flag curvature.

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

(i) *for all $y \in T_xM$ and $x \in \overline{U}$, $F_x(-y) \leq c_0^2 F_x(y)$; and*
(ii) *for all $x_1, x_2 \in U$, $c_0^{-2} d(x_1, x_2) \leq d(x_2, x_1) \leq c_0^2 d(x_1, x_2)$.*

Proof See [16, p. 146]. □

2.1 Gradient on a Finslerian Manifold

First, let us study the relationship between a Minkowski norm (see [27, Definition 1.2.1]) and its dual norm. Let V be a finite-dimensional vector space and V^* the dual vector space. Let F be a Minkowski norm on V, where F is a norm in the sense that for any $y, v \in V$ and $\lambda > 0$, $F(y+v) \le F(y) + F(v)$ and $F(\lambda y) = \lambda F(y)$. Define F^* as a Minkowski norm on V^*, where $F^*(\xi) := \sup_{F(y)=1} \xi(y)$.

The following lemma provides a characterization of the gradient in the Finslerian context.

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 $F(y) = F^*(\xi) = \xi(y)/F(y)$. 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 [27, p. 35]. □

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

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

and is a linear functional on $T_x M$. To connect df_x to a vector grad $f_x \in T_x M$ requires a Minkowski norm on $T_x M$. Let F be a Finslerian metric on M. By definition, F_x is a Minkowski norm on $T_x M$. Assume that $df_x \neq 0$. Since the indicatrix $S := F^{-1}(1)$ is strongly convex, there is a unique unit vector $s_x \in S_x M := F_x^{-1}(1)$ and a positive number $\lambda_x > 0$ such that $W^{\lambda_x} = \{v : df_x(v) = \lambda_x\}$ is tangential to $S_x M$ at s_x. Lemma 2.2 states that $\mathrm{d}f_x(v) = \lambda_x g_{s_x}(s_x, v)$, where $F_x^*(df_x) = df_x(s_x) = \lambda_x g_{s_x}(s_x, s_x) = \lambda_x$. Define grad $f_x := \lambda_x s_x = F_x^*(df_x) s_x$. We can write

$$df_x(v) = g_{\text{grad } f_x}(\text{grad } f_x, v), \quad v \in T_x M. \tag{1}$$

Now, given a compact subset $S \subset M$, define

$$d_+(x) := d(S, x) \quad \text{and} \quad d_-(x) := -d(x, S). \tag{2}$$

Lemma 2.3 *Let (M, F) be a Finslerian space, and $d_+(x)$ and $d_-(x)$ be the functions defined in (2). Suppose that for any points $x, z \in M$, there is a shortest unit speed curve from x to z. Then, $F(\text{grad } d_+(x)) = 1$ and $F(\text{grad } d_-(x)) = 1$ hold almost everywhere.*

Proof See [27, Lemma 3.2.3, p. 44]. □

3 Kurdyka-Lojasiewicz Inequality on Finslerian Manifolds

In this section, we present the Kurdyka-Lojasiewicz inequality in a Finslerian context. The differentiable case was presented in [26, Corollary 1.1.25].

Let $f : M \to \mathbb{R}$ be a C^1 function, and consider the set: $[\eta_1 < f < \eta_2] := \{x \in M : \eta_1 < f(x) < \eta_2\}$, where $-\infty < \eta_1 < \eta_2 < +\infty$.

Definition 3.1 The function f defined above is said to satisfy the Kurdyka-Lojasiewicz property at $\bar{x} \in M$ iff 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, i.e.,

$$\varphi'\big(f(x) - f(\bar{x})\big) F\big(\text{grad } f(x)\big) \geq 1. \tag{3}$$

We call f a *KL* function iff it satisfies the Kurdyka-Lojasiewicz inequality at each point of M.

Next lemma, we show that a C^1 function satisfies the Kurdyka-Lojasiewicz property at all non-critical points.

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

Proof Since $\bar{x}$ is a non-critical point of f, we have that $\delta := F(\text{grad } f(\bar{x})) > 0$. Take $\varphi(t) := t/\delta, \eta := \delta/2$ and $U := B(\bar{x}, \delta/2)$. Now, for each $x \in U \cap [f(\bar{x}) - \eta < f < f(\bar{x}) + \eta]$, note that $d(x, \bar{x}) + |f(x) - f(\bar{x})| < \delta$. Then, for each x satisfying the last inequality, it holds that

$$F\big(\text{grad } f(x)\big) \geq \delta.$$

The proof follows by contradiction, assuming that last inequality is not true. □

We use $\mathcal{O} = \{\mathcal{O}_n\}_{n \in \mathbb{N}}$ to represent an algebraic structure; see [25] and van den Dries and Miller [28]. 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 $\mathcal{O}$ iff its graph belongs to $\mathcal{O}_{n+1}$. A generalization of semi-analytic and sub-analytic sets (analogous to that given for semi-algebraic sets in terms of the o-minimal structure) leads us to the analytic-geometric categories $C(M)$ defined by Bierstone and Milman in [29] and van den Dries in [30]. The elements of $\mathcal{C}(M)$ are called $\mathcal{C}$-sets. When the graph of a continuous function $f : A \to B$, $A \in \mathcal{C}(M)$, $B \in \mathcal{C}(N)$ is contained in $\mathcal{C}(M \times N)$, f is called a $\mathcal{C}$-function.

The following result provides an extension of the Kurdyka-Lojasiewicz properties for $\mathcal{C}$-functions defined on analytic Finslerian manifolds.

Theorem 3.1 *Let M be an analytic Finslerian manifold and $f : M \to \mathbb{R}$ be a C^1-function. Assume that f is a $\mathcal{C}$-function. Then, f is a* KL *function. Moreover, the function φ in inequality* (3) *is definable in $\mathcal{O}$.*

Proof Take $\bar{x} \in M$ to be a critical point of f, and let $\phi : V \to \mathbb{R}^n$ be an analytic local chart with $V \subset M$ a neighborhood of $\bar{x}$ chosen such that V and $f(V)$ are bounded. Thus, from [26, Proposition 1.1.5.], we have that $f \circ \phi^{-1} : \phi(V) \to \mathbb{R}$ is a definable function in $\mathcal{O}(\mathcal{C})$. $\phi(V)$ is a bounded open definable set containing $\bar{z}=\phi(\bar{x})$ and ϕ is definable. So we can use [25, Theorem 11] with $U=\phi(V)$, and take into account the proof presented by Attouch et al. in [31, Theorem 4.1], to show that the Kurdyka-Lojasiewicz property holds at $\bar{z}=\phi(\bar{x})$. That is, there exist U, η and $\Phi:[0,\eta[\to\mathbb{R}_+$ as in Definition 3.1. Because ϕ is a diffeomorphism and $z=\phi(x)$, $\bar{z}=\phi(\bar{x})$, and $h=f\circ\phi^{-1}$, we can use the chain rule to show that

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

where ϕ_x^* denotes the derivative adjunct of the map ϕ. Let $V'\subset V$ be 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 a compact set and for each $x\in K$ there exists $C_x>0$ with $F((\phi_x^*)^{-1}w)\leq C_x F(w),\ w\in T_xM$. Because K is a compact set and $(\phi_x^*)^{-1}$ is a diffeomorphism, there exists a constant $C:=\sup\{C_x : x\in K\}$ such that $F((\phi_x^*)^{-1}w)\leq CF(w),\ w\in T_xM,\ x\in K$. Hence, for $x\in V'\cap[0<f<f(\bar{x})+\eta]$, we have

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

and the Kurdyka-Lojasiewicz property holds at $\bar{x}$ with $\varphi=C\,\Phi$. Therefore, as $\bar{x}$ was chosen arbitrarily, it follows from Lemma 3.1 that f is a KL function. It also follows from [25, Theorem 11] that φ is definable in $\mathcal{O}$, and the proof is concluded. □

4 Proximal Point Algorithm

We now establish the conditions that ensure the proper definition and convergence of our method. First, note that the Finslerian distance is fundamental to this section. Let M be a complete Finslerian manifold and $f:M\to\mathbb{R}$ be a KL function, where f is C^1. We consider the optimization problem

$$\min f(x), \quad x\in M.$$

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

$$x_{k+1}\in\arg\min_{z\in M}\left\{f(z)+\frac{1}{2\lambda_k}d^2(x_k,z)\right\}, \tag{4}$$

where d is the Finslerian distance, and $\{\lambda_k\}$ is a sequence of positive numbers.

4.1 Convergence

In the sequel, unless explicitly stated otherwise, we assume that M is on a Finslerian Hadamard manifold, $\inf_M f>-\infty$, and for some positive $t_1<t_2$ ($t_1\leq\lambda_k\leq t_2$), for all $k\geq 0$.

Proposition 4.1 *Let $\{x_k\}$ be the sequence defined in (4). Then $\{x_k\}$ is well defined and*

$$f(x_{k+1}) + \frac{1}{2\lambda_k} d^2(x_k, x_{k+1}) \leq f(x_k). \tag{5}$$

Proof M is a Hadamard manifold and $\inf f > -\infty$. We can use arguments similar to the Riemannian case (see [18]) to show that, for any $k \in \mathbb{N}$, the function of (4) is 1-coercive. Consequently the sequence $\{x_k\}$ is well defined. An elementary induction ensures that (5) holds. □

Remark 4.1 In this case, we say that the sequence is well defined in the sense that there is a solution, but it may not be unique, as in the convex setting.

Next lemma, we present an estimation of the gradient norm in terms of distance.

Lemma 4.1 *Let $\{x_k\}$ be the sequence defined in (4) and $x_{k_0} \in B(\tilde{x}, \rho) \subset U$, where the neighborhood U is given by Lemma 2.1. Then,*

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

where $c_0 > 1$.

Proof From (4), we have

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

Using the linearity of the differential,

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

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

$$g_{\operatorname{grad} f(x_{k_0})}\big(\operatorname{grad} f(x_{k_0}), \operatorname{grad} f(x_{k_0})\big) = g_{\operatorname{grad} h(x_{k_0})}\big(\operatorname{grad} h(x_{k_0}), -\operatorname{grad} f(x_{k_0})\big).$$

Therefore, using the Cauchy–Schwartz inequality $(g_y(y, v) \leq F(y)F(v))$, Lemma 2.1, and the homogeneity of F, we conclude that

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

Taking $S = \{x_{k_0-1}\}$ and $x = x_{k_0}$, Lemma 2.3 states that $F(\operatorname{grad} d(x_{k_0-1}, x_{k_0})) = 1$. □

Now, we present a technical result that we use in our convergence analysis.

Lemma 4.2 *Let $\{a_k\}$ be a sequence such that $a_k > 0$ and $\sum_{k=1}^{+\infty} a_k^2/a_{k-1} < +\infty$. Then, $\sum_{k=1}^{+\infty} a_k < +\infty$.*

Proof For the proof, it is sufficient to use the Cauchy–Schwartz inequality in $\mathbb{R}^j$, with respect to the vectors $(a_1/\sqrt{a_0}, \dots, a_j/\sqrt{a_{j-1}})$ and $(\sqrt{a_0}, \dots, \sqrt{a_{j-1}})$ for fixed $j \in \mathbb{N}$. □

Lemma 4.3 *Let $\{x_k\}$ be the sequence defined in (4). Assume that $f : M \to \mathbb{R}$ is a C^1 function, $\tilde{x}$ is an accumulation point of $\{x_k\}$, and f satisfies the Kurdyka-Lojasiewicz property at $\tilde{x}$. Let $a = 1/2t_2$, $b = c_0^2/t_1$ be constants, and $\rho > 0$ be 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*

$$f(\tilde{x}) < f(x_k) < f(\tilde{x}) + \eta, \quad k \geq k_0, \tag{6}$$

$$d(\tilde{x}, x_{k_0}) + 2\sqrt{\frac{f(x_{k_0}) - f(\tilde{x})}{a}} + \frac{b}{a}\varphi\big(f(x_{k_0}) - f(\tilde{x})\big) < \rho. \tag{7}$$

Moreover,

$$\frac{b}{a}\big[\varphi\big(f(x_{k_0}) - f(\tilde{x})\big) - \varphi\big(f(x_{k_0+1}) - f(\tilde{x})\big)\big] \geq \frac{d^2(x_{k_0}, x_{k_0+1})}{d(x_{k_0-1}, x_{k_0})}. \tag{8}$$

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 $\tilde{x}$ be an accumulation point of $\{x_k\}$. Now, from the inequality (5), the continuity of f, and Proposition 4.1 we have that $\{f(x_k)\}$ converges to $f(\tilde{x})$ and

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

In particular, there exists $N \in \mathbb{N}$ such that inequality (6) holds. Since the last inequality holds, let us define the sequence

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

Note that 0 is an accumulation point of the sequence $\{b_k\}$ and hence there exists $k_0 := k_{j_0} > N$ such that inequality (7) holds.

From inequalities (6) and (7), 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, $0 \neq \operatorname{grad} f(x_{k_0})$. Moreover, Lemma 4.1, a property of the function φ, and inequality (5) yields

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

Therefore, inequality (8) follows from the last inequality. The proof of the latter part follows from inequality (8) combined with Lemma 4.2. □

Lemma 4.4 *Let $\{x_k\}$ be the sequence defined in (4). Under the assumptions of Lemma 4.3, there exists $k_0 \in \mathbb{N}$ such that*

$$x_k \in B(\tilde{x}, \rho), \quad k > k_0. \tag{9}$$

Proof The proof is by induction on k. It follows from inequality (5) that the sequence $\{f(x_k)\}$ is decreasing, and from Lemma 4.3 that there exists $k_0 \in \mathbb{N}$ such that inequalities (7) and (6) hold. Hence,

$$d(x_{k_0}, x_{k_0+1}) \le \sqrt{\frac{f(x_{k_0}) - f(\tilde{x})}{a}}. \tag{10}$$

Using the triangle inequality with the last expression and inequality (7), we find that $x_{k_0+1} \in B(\tilde{x}, \rho)$. Assuming that inequality (9) holds for all $k = k_0+1, \ldots, k_0+j-1$, we have that inequality (8) also holds, and it follows that

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

So, a summation with $k = k_0 + 1, \ldots, k_0 + j - 1$ leads to

$$\begin{aligned}\sum_{i=k_0+1}^{k_0+j-1} d(x_i, x_{i+1}) &\le d(x_{k_0}, x_{k_0+1}) + \frac{b}{a}\varphi\big(f(x_{k_0+1}) - f(\tilde{x})\big) \\ &\le d(x_{k_0}, x_{k_0+1}) + \frac{b}{a}\varphi\big(f(x_{k_0}) - f(\tilde{x})\big).\end{aligned}$$

Now, combining the last inequality with the triangle inequality, we get

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

Therefore, using the last inequality and inequalities (10) and (7), we conclude that $x_{k_0+j} \in B(\tilde{x}, \rho)$, which completes the inductive proof. □

Next theorem, we prove that the sequence defined in (4) converges to a critical point.

Theorem 4.1 *Let U, η, and $\varphi : [0, \eta[\to \mathbb{R}_+$ be as defined in 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 property at $\tilde{x}$. Then,*

$$\sum_{k=0}^{+\infty} d(x_k, x_{k+1}) < +\infty. \tag{11}$$

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

Proof Note that Lemma 4.4 combined with Lemma 4.3 implies that inequality (11) holds and, as a consequence, the sequence $\{x_k\}$ converges to $\tilde{x} \in M$. Thus, using the continuity of f combined with inequality (5), it follows that $f(x_k) \to f(\tilde{x})$, as $k \to +\infty$. Combining inequality (11) with Lemma 4.1, it follows that $F(\operatorname{grad} f(\tilde{x})) = 0$. Therefore, $\tilde{x}$ is a critical point of f. □

Theorem 4.2 *With the same hypotheses as Lemma 4.4, assume that* $\{x_k\}$ *converges to* x^*, *and that* f *satisfies the Kurdyka-Lojasiewicz property at* x^*, *with* $\varphi(s) = cs^{1-\theta}$, $\theta \in [0, 1[$, *and* $c > 0$. *Then*

(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 exists* $\xi > 0$ *such that* $d(x_k, x^*) \leq \xi k^{-\frac{1-\theta}{2\theta-1}}$.

Proof The proof follows from Cruz Neto et al. [32, Theorem 4.4] with appropriate adaptations. □

5 The "Effort–Accuracy" Trade-off: an "Exploration–Exploitation" Model on a Finslerian Manifold

In this and the next section, we show how a proximal algorithm on a Finslerian manifold can model the "effort–accuracy" trade-off, using the recent "variational approach" of change theory (see [11, 12] and [13, 14] for more specific cases).

5.1 Finslerian Distances as Costs of Change

Let us consider a given period and an agent who expends 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 resources (means) to produce the output $q(x') = \mathbf{q}[x', E'(x')] \in Q$ and gain the benefit $b(x') = \mathbf{b}[q(x')] \in \mathbb{R}_+$. To regenerate this energy in the next period, the agent must eat, rest, spend holidays, have hobbies, pursue healthy activities, and so on. Doing so expends 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 disutility of the consumed energy is given by $\delta[E'(x') + E''_2(x'')] = c(x) \in \mathbb{R}_+$, where the vector of stocks of resources used this period is $x = (x', x'') = (x^i, i \in I) \in X = X' \times X''$. Accordingly, $g(x) = b(x') - c(x'')$ is the payoff for the agent.

One important constraint that is usually neglected in economic literature is that the agent can conserve (regenerate) energy over time. Accordingly, the regeneration of vital energy imposes the constraint $E''(x'') - E'(x') \geq E$, where $E > 0$ is the extra energy used for other tasks (which we will discuss later). This defines a manifold $M = \{x \in X : E''(x'') - E'(x') = E > 0\}$.

Energy production and consumption functions can be quadratic (the more an agent performs an activity, the more energy is produced and consumed at an increasing rate). In this case,

$$\sum_{k \in K} (x^k)^2 - \sum_{h \in H} (x^h)^2 = E > 0$$

defines an hyperboloid. A more realistic example can be given where energy production functions are increasing and concave, and energy consumption functions are increasing and convex. More generally, every activity can both consume and produce some energy.

Starting from $x \in M$, the agent expends effort $e = (e^i, i \in I) \in \mathbb{R}^I = T_x M$ to change the stock 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 $e \in R^I$ is the vector of the effort necessary to pass from x to y. An added energy constraint $E \geq \sum_{i \in I} e^i$ must also be satisfied when the agent passes from one period to the next. For simplicity, we will assume that it is always satisfied.

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 $F(x, e) = 0$ if only if $e = 0$ (no effort implies no cost to expend an effort, and vice versa). We impose a convexity condition, $F(x, e + e') \leq F(x, e) + F(x, e')$, for all $e, e' \in T_x M$. Thus, if $\delta_x(x, y) = F(x, y - x)$, $y - x = e$ and $z - y = e'$, then $\delta_x(x, z) \leq \delta_x(x, y) + \delta_x(y, z)$. Now, starting from x, $\delta_x(x, y) \geq 0$ represents a quasi-distance between x and y. The 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 resource path $\{x_0, x_1, \ldots, x_h = \varphi(t_h), , \ldots x_n\} \subset M$ from $x = x_0$ to $y = x_n$, where $x_h = \varphi(t_h), a = t_0 < t_1 < \cdots < t_h < \cdots < t_n = b$, $\varphi(a) = x$, and $\varphi(b) = y$. Following this path of change, the total cost of changing from x to y is $\mathbf{F}(x, y, p) = \sum_{h=0}^{n-1} F(x_h, e_h)$, where $e_h = x(t_{h+1}) - x(t_h)$. As an idealization, consider paths of change from x to y as smooth curves $\Phi = \{\varphi(t) : a \leq t \leq b\} \subset P(x, y) \subset M$, parameterized by the map $z = \varphi(t), a \leq t \leq b$. Let $\dot{\varphi}(t) = e(t)$ be the effort expended to pass from $\varphi(t)$ to $y(t) = \exp_{x(t)} \dot{\varphi}(t)$, 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 of changing from x to y. This defines a quasi-distance where $C(y, x) \neq C(x, y)$ for some x, y. This formulation defines Finslerian quasi-distances as costs of change on a manifold. The general concept of costs of change in behavioral sciences (economics, management, psychology) comes from [11, 12]. See also [14] and Moreno et al. [33] for an interpretation as regularization terms for proximal algorithms.

5.2 An "Exploration–Exploitation" Model

In the present model, the agent's problem in each period is a version of the famous "exploration–exploitation" problem (see [11, 12] for the 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 for $\mu(x) > 0$ 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 is $b - a > 0$ long and costs $\mathbf{F}(x, y, p) \geq 0$. In this case, the term exploration is somewhat misleading. Here, it represents the discovery and

the evaluation of the resource path $p \in P(x, y)$, and a way, step by step starting from x, to build the stocks of resources y.

At each step, the agent must balance a costly exploration phase against a rewarding exploitation phase, i.e., a low enough cost of 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 the disutility of the change in costs for each period, and $U[\mu(x)g(y)]$ be the utility of the net payoff, where $D[\mathbf{F}] \geq 0$,$D[0] = 0$, $D[.]$ is not decreasing, $U[Z] \geq 0$, $U[0] = 0$, and $U[.]$ is not decreasing. Then, the net payoff to 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 assume that $U[Z] = Z$ 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 net payoff per unit of time is $G(x, y, p) = g(y) - \lambda(x)\mathbf{F}(x, y, p)^2$, where a function $\lambda(x) = 1/\mu(x) > 0$ defines a weight applied to costs of change.

The agent's "exact exploration–exploitation" problem is to solve the optimization problem at each step [11, 12]. It is defined: find $x_{n+1} \in M$, $p_{n+1} \in P(x_n, y)$ such that

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

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 this problem approximately at each step. In the exact case, the agent can follow a two-step process in each period:

(i) given each new stock of resources $y \in M$, solve $C(x_n, y) = \inf\{\mathbf{F}(x_n, y, p) : p \in P(x_n, y)\}$, i.e., 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 $\Gamma(x_n, y) = g(y) - \lambda(x_n)C(x_n, y)^2$ as the proximal payoff of the agent at $(x_n, \lambda_n = \lambda(x_n))$. Then, find a stock of resources $x_{n+1} \in M$ that solves $\Gamma(x_n, x_{n+1}) = \sup\{g(y) - \lambda_n C(x_n, y)^2 : y \in M\}$.

Accordingly, this "exact exploration–exploitation" problem can be identified with a proximal problem on a Finslerian manifold M: at each step, starting from $x_0 \in M$, solve $\sup\{\Gamma(x_n, y) : y \in M\}$. 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 agent's unmet need at y. An equivalent formulation of the problem is to say that at each step the agent wants to minimize unmet needs, taking account of the costs of changing. The agent solves the sequence of iterations x_{n+1} in $\arg\min\{f(y) + \lambda_n C(x_n, y)^2 : y \in M\}$, where $C(x, y) \geq 0$ is a Finslerian quasi-distance on M.

Comments 5.1 Proximal point algorithm can be seen as repeated optimization problems with costs associated with changes. From the bounded rationality perspective, see Simon [34, 35], where it seems difficult for agents to optimize, this poses the famous "recursive problem": how costly it is to know the costs of optimization?

Comments 5.2 In our model, repeated optimization is necessary because preferences change at each step. Preferences are variable, $z \geq_x y$ if only if $\Gamma(x, z) \geq \Gamma(x, y)$, (see [12]). The agent enters on a chosen course of action where, at each step, they choose and build their optimal stock of resources with respect to their present preference. This, in turn, gives rise to a new preference and a new optimal stock of resources and the process continues. Variable preferences come from inertia (costs of change and, at each period, a variable exploitation phase length $\mu(x) = 1/\lambda(x) > 0$). Inertia is strong because small changes have high costs (cost of change represents a quasi-distance). This greatly helps convergence.

5.3 Behavioral Traps

A stock of resources $x^* \in M$ is a (behavioral) "trap" [11, 12] if and only if the agent prefers to conserve this stock of resources rather than change. Let us define the proximal payoff of the agent in a point x^* as $\Gamma(x^*, y) = g(y) - \lambda(x^*)C(x^*, y)^2$. Then $x^* \in M$ is a trap if $\Gamma(x^*, x^*) \geq \Gamma(x^*, y)$ for all $y \in M$, i.e., $g(x^*) \geq g(y) - \lambda(x^*)C(x^*, y)^2$ for all $y \in M$. This trap is strict if $\Gamma(x^*, x^*) > \Gamma(x^*, y)$ for all $y \in M, y \neq x^*$. A global maximum $\widehat{x} \in M$ of $g(.)$ is a trap if and only if the agent ignores costs associated with changes. In our case, the agent takes account of inertia and learning aspects (both embedded in costs of change). At $\widehat{x}$, unmet 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 stocks of resources in each period that are increasingly similar, which renders services that are increasingly similar $q(x_n)$. The agent enters into a process of habituation. 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 proposed conditions such that the limit x^* is a critical point of $g(.)$ (i.e., a critical point of the unmet needs $f(y) = \overline{g} - g(y) \geq 0$). Does this mean that x^* is a trap?

$$\Gamma\bigl(x^*, x^*\bigr) = g\bigl(x^*\bigr) - \lambda^* C\bigl(x^*, x^*\bigr) = g\bigl(x^*\bigr) \geq \Gamma\bigl(x^*, y\bigr) = g(y) - \lambda^* C\bigl(x^*, y\bigr)^2 \quad \text{for all } y \in M?$$

This is true if $g(.)$, $\lambda(.)$ and $C(x, .)$ are continuous functions. In this case,

$$\Gamma\bigl(x^*, x^*\bigr) \overset{(1)}{=} \lim_{n \longrightarrow +\infty} \Gamma(x_n, x_{n+1}) \overset{(2)}{\geq} \lim_{n \longrightarrow +\infty} \Gamma(x_n, y) \overset{(3)}{=} \Gamma\bigl(x^*, y\bigr), \quad \text{for each } y \in M.$$

Equalities (1) and (3) in the above equation are derived from the continuity assumption, while inequality (2) is because $\Gamma(x_n, x_{n+1}) \geq \Gamma(x_n, y)$, for all $y \in M$.

6 The Speed of Decision Making

Now, we can demonstrate how our "exploration–exploitation" interpretation of the proximal algorithm can give accurate and precise results for the famous "effort–accuracy" trade-off (see [1]), where the speed of decision making matters and when there are costs associated with change (so optimization is costly).

We can apply Lemma 4.3 to get some results in terms of the time taken to converge to a critical point.

6.1 Majoration of the Length of a Convergent Path

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

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

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

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

implies that

$$\left(\sum_{k=k_0+1}^{n} a_{k-1}\right)^{1/2} \le (a_{k_0})^{1/2} + \left(\sum_{k=k_0+1}^{n} a_k^2 / a_{k-1}\right)^{1/2}.$$

It follows that

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

Let $X_n = \{x_{k_0}, x_{k_0+1}, \ldots, x_{k_0+n}\} \subset X$ be the first n steps $\{x_h : h = k_0, k_0+1, \ldots\}$ of the whole path $\{x_n\}$, and $L\{X_n\} = \sum_{k=k_0+1}^{n} d(x_{k-1}, x_k)$ be its length. Then,

$$\begin{aligned} L\{X_n\} \le{}& d(x_{k_0}, x_{k_0+1}) + \frac{b}{a}\big[\sigma(x_{k_0+1}) - \sigma(x_n)\big] \\ &+ 2\big[d(x_{k_0}, x_{k_0+1})\big]^{1/2}\left[\frac{b}{a}\big[\sigma(x_{k_0+1}) - \sigma(x_n)\big]\right]^{1/2}. \end{aligned}$$

Corollary 6.1 *The time the agent spends to converge to a point along path is less than*

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

6.2 Majoration of the Time to Convergence

Let $t(x, y) \ge 0$ be the time taken to move from a point x to point y, where $t(x, y) > 0$ if $y \ne x$. Let $v = d(x, y)/t(x, y) > 0$, $y \ne x$, be the speed of decision making. If the speed of decision making $v > 0$ is constant, then the time taken to move is proportional to the distance, $t(x, y) = v^{-1} d(x, y)$. Therefore, the time taken to move the first n steps along the path $\{x_k\}$ is $T(X_n) = v^{-1} L\{X_n\}$.

Corollary 6.2 *Assume that the speed of decision making $v > 0$ is constant. Then the time to convergence, $T(X_{+\infty}) = v^{-1}L\{X_{+\infty}\}$, is lower than*

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

i.e., the relative time to convergence as compared to the first step is lower than

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

A really striking result is that the lower the rate of improvement at the first step

$$\sigma(x_{k_0+1})/d(x_{k_0}, x_{k_0+1}) = \left[\varphi\left(f(x_k) - f(\tilde{x})\right)/d(x_{k_0}, x_{k_0+1})\right] > 0,$$

the less time to convergence. This means that the first step is very important. The agent must not try to improve too much at the beginning, which confirms the proverb "more haste, less speed," at least to begin with. Furthermore, for $a = 1/2t_2$, $b = c_0^2/t_1$ the lower the ratio $b/a = 2c_0^2 t_2/t_1$ (i.e., the lower c_0 and t_2, and the higher t_1) the less the time taken to converge. Then, from the inequalities $0 < t_1 \leq \lambda_k \leq t_2$ and $a = 1/2t_2 \leq 1/2\lambda_k \leq 1/2t_1$, a higher t_1 will result in a smaller weight $1/2\lambda_k$ imposed on costs of change, so convergence will take less time. Hence, less inertia speeds up convergence—another satisfying result.

6.3 Residual Time to Convergence

We will now discuss the implications of Theorem 4.2 in terms of the residual time taken to converge $t(x_k, x^*) = v^{-1}d(x_k, x^*)$.

(i) The case $\theta = 0$ gives convergence in a finite number of steps, which is a very satisfying result for decision-making problems;

(ii) The case $0 < \theta \leq 1/2$ shows that, starting from x_k, the residual time taken to converge is less than $t(x_k, x^*) = v^{-1}d(x_k, x^*) \leq v^{-1}D(k_0)\tau^k$. Where $\varsigma = c_1/(1 + c_1)$, $c_1 = E\Theta + 1$, $\Theta = c[c(1-\theta)b]^{(1-\theta)/\theta}$, $D(k_0) = \sum_{i=k_0+1}^{+\infty} d(x_i, x_{i+1})$ and $E = b/a = 2c_0^2 t_2/t_1$. This shows that the higher t_1, the lower inertia (the lower the weight $1/2\lambda_k$ on costs of change, where $1/2t_2 \leq 1/2\lambda_k \leq 1/2t_1$). Then a lower ς results in a shorter residual time to convergence. Furthermore, a lower $c > 0$ results in a lower Θ, which results in a lower c_1, a lower τ, and a lower residual time for convergence. Smaller $c > 0$ means that we can expect little improvement $|f(x) - f(x^*)|$ when moving from x to x^*, because $|f(x) - f(x^*)|^\theta \leq cF(\text{grad } f(x))$. Similar comments can be made for the case $1/2 < \theta < 1$.

7 Conclusions

We have presented and analyzed the proximal point method on Finslerian manifolds for minimizing differentiable KL functions. We used intrinsic relations between the o-minimal structure and analytic-geometric categories to ensure the existence of KL functions. We derived important theoretical results for convergence (Theorem 4.1) and a convergence rate (Theorem 4.2) in the Finslerian context. Applied to our "exploration–exploitation" model, these results have a very interesting interpretation. Our main Theorem (Theorem 4.1) shows that the length of the proximal point path, starting from the present point, is finite. As a consequence of the convergence results, we show that our application—an "effort–accuracy trade-off" model—is naturally connected to proximal point algorithms and Finslerian quasi-distances. This opens the door to inexact proximal algorithm formulations, which will economize more in terms of costs associated with change (given that optimizing at each step will not be a successful optimization strategy in the long run).

References

1. Payne, J.W., Bettman, J.R., Johnson, E.J.: The Adaptive Decision Maker. Cambridge University Press, Cambridge (1993)
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. J. Exp. Psychol. Gen. **133**(2), 261–282 (2004)
3. Holland, J.H.: Adaptation in Natural and Artificial Systems. University of Michigan Press, Ann Arbor (1975)
4. March, J.G.: Exploration and exploitation in organizational learning. Organ. Sci. **2**(1), 71–87 (1991)
5. Levitt, B., March, J.G.: Organizational learning. Annu. Rev. Sociol. **14**, 319–338 (1988)
6. Cohen, W.M., Levinthal, D.A.: Absorptive capacity: a new perspective on learning and innovation. Adm. Sci. Q. **35**(1), 128–152 (1990)
7. Levinthal, D.A., March, J.G.: The myopia of learning. Strateg. Manag. J. **14**, 95–112 (1993)
8. Fu, W.T., Gray, W.D.: Suboptimal tradeoffs in information seeking. Cogn. Psychol. **52**(3), 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, Oxford (2007)
10. Busemeyer, J.R., Townsend, J.T.: Decision field theory: a dynamic cognition approach to decision making. Psychol. Rev. **100**(3), 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, beliefs. Preprint, GREQAM, Aix Marseillle University, (2010)
13. Attouch, H., Soubeyran, A.: Inertia and reactivity in decision making as cognitive variational inequalities. J. Convex Anal. **13**(2), 207–224 (2006)
14. Attouch, H., Soubeyran, A.: Local search proximal algorithms as decision dynamics with costs to move. Set-Valued Var. Anal. **19**(1), 157–177 (2011)
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. Kristály, A., Morosanu, G., Róth, Á.: Optimal Placement of a Deposit between Markets: Riemann–Finsler Geometrical Approach. J. Optim. Theory Appl. **139**(2), 263–276 (2008)
18. Ferreira, O.P., Oliveira, P.R.: Proximal point algorithm on Riemannian manifold. Optimization **51**(2), 257–270 (2002)
19. Li, C., Lopez, G., Martin-Marquez, V.: Monotone vector fields and the proximal point algorithm on Hadamard manifolds. J. Lond. Math. Soc. **79**(2), 663–683 (2009)
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 Anal. **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. Bolte, J., Daniilidis, A., Lewis, A.: The Lojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems. SIAM J. Optim. **17**(4), 1205–1223 (2006)
24. 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)
25. Bolte, J., Daniilidis, A., Lewis, A., Shiota, M.: Clarke subgradients of stratifiable functions. SIAM J. Optim. **18**(2), 556–572 (2007)
26. Lageman, C.: Convergence of gradient-like dynamical systems and optimization algorithms. Dissertation zur Erlangung des naturwissenschaftlichen Doktorgrades der Bayerischen Julius-Maximilians-Universitat Wurzburg (2007)
27. Shen, Z.: Lectures on Finsler Geometry. World Scientific, Singapore (2001)
28. van den Dries, L., Miller, C.: Geometric categories and o-minimal structures. Duke Math. J. **84**(2), 497–540 (1996)
29. Bierstone, E., Milman, P.D.: Semianalytic and subanalytic sets. Publ. Math. Paris **67**, 5–42 (1988)
30. van den Dries, L.: O-minimal structures and real analytic geometry. In: Current Developments in Mathematics, Cambridge, MA, 1998, pp. 105–152. Int. Press, Somerville (1999)
31. 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)
32. Cruz Neto, J.X., Oliveira, P.R., Soares, P.A. Jr, Soubeyran, A.: Learning how to Play Nash, Potential Games and Alternating Minimization Method for Structured Nonconvex Problems on Riemannian Manifolds. J. Convex Anal. **20**(2), 395–438 (2013)
33. Moreno, F., Oliveira, P., Soubeyran, A.: A proximal algorithm with quasi distance. Application to habit's formation. Optimization **61**(2), 1383–1403 (2012)
34. Simon, H.A.: A behavioral model of rational choice. Q. J. Econ. **69**, 99–188 (1955)
35. Simon, H.A.: Rationality in psychology and economics. Part 2: The behavioral foundations of economic theory. J. Bus. **59**(4), S209–S224 (1986)