On the Relationship Between the Kurdyka–Łojasiewicz Property and Error Bounds on Hadamard Manifolds

João Xavier da Cruz Neto[1] · **Ítalo Dowell Lira Melo**[1] ·
Paulo Alexandre Sousa[1] · **João Carlos de Oliveira Souza**[1]

Abstract
This paper studies the interplay between the concepts of error bounds and the Kurdyka–Łojasiewicz (KL) inequality on Hadamard manifolds. To this end, we extend some properties and existence results of a solution for differential inclusions on Hadamard manifolds. As a second contribution, we show how the KL inequality can be used to obtain the convergence of the gradient method for solving convex feasibility problems on Hadamard manifolds. The convergence results of the alternating projection method are also established for cyclic and random projections on Hadamard manifolds and, more generally, CAT(0) spaces.

Keywords Differential inclusion · Kurdyka–Łojasiewicz inequality · Error bounds · Convex feasibility problem · Hadamard manifolds · CAT(0) spaces

Mathematics Subject Classification 65K10 · 90C30 · 49M99 · 65J05

1 Introduction

The Łojasiewicz inequality [39] is a powerful tool to analyze the convergence of some optimization methods, in particular, gradient-like methods. It is the key to obtaining

Communicated by Nicolas Hadjisavvas.

✉ João Carlos de Oliveira Souza
  joaocos.mat@ufpi.edu.br

  João Xavier da Cruz Neto
  jxavier@ufpi.edu.br

  Ítalo Dowell Lira Melo
  italodowell@ufpi.edu.br

  Paulo Alexandre Sousa
  paulosousa@ufpi.edu.br

1 Department of Mathematics, Federal University of Piauí, Teresina, PI, Brazil

convergence to the critical points of all bounded trajectories of the gradient dynamical system. However, it might fail for a C^∞ function without an "adequate" geometric structure; see, for instance, a counterexample due to Palis and De Melo [42, p. 14] showing that the set of cluster points of a bounded trajectory generated by the above gradient dynamical system of a C^∞ function is, in general, far from being a singleton.

A generalized form of this property was introduced by Kurdyka [37] (called Kurdyka–Łojasiewicz inequality, or shortly, KL inequality). In finite-dimensional spaces, it has been shown in [37] that the KL inequality holds for a much larger class of functions, namely, those that are definable in an o-minimal structure, or even more generally, functions belonging to analytic-geometric categories.

Bolte et al. [15] extended the Łojasiewicz inequality to a broad class of non-smooth functions by establishing an analogous inequality in which the gradient of the objective function is replaced by an element of its subdifferential, providing new insights into the convergence aspects of subgradient-type dynamical systems and leading to the conclusion that the Łojasiewicz inequality is more linked to the underlying geometrical structure of the function than to its smoothness. In Bolte et al. [16], a non-smooth extension of the Kurdyka–Łojasiewicz inequality for functions definable in an arbitrary o-minimal structure was obtained.

On the other hand, since Hoffman's celebrated result on error bounds for systems of linear inequalities [32], the study of error bounds has been successfully applied to problems in convergence rate estimation and feasibility problems. Recently, Bolte et al. [18] proved that when the objective function is convex, error bounds are equivalent to non-smooth KL inequalities, provided the residual function in the error bound has a moderate behavior close to 0. Here, moderate behavior means that its derivative blows up at a reasonable rate. This result is significant in determining the complexity of first-order methods.

Error bounds and the Kurdyka–Łojasiewicz inequality are widely used in the literature for analyzing the rate of convergence for various optimization algorithms and for understanding the local variational geometry of the solution set of an optimization problem.

In recent years, there has been a growing interest in optimization on Riemannian manifolds. The study of optimization on Riemannian manifolds and corresponding algorithms began in the 1970s with the remarkable work of Luenberger [40]. One advantage of this study is the possibility of transforming some Euclidean non-convex or constrained problems into Riemannian convex or unconstrained problems by introducing a suitable metric; see, for instance, [24, 25]. With the emergence of optimization problems arising from practical applications posed in a Riemannian setting, interest in this topic has increased significantly over the years. Even though we are not concerned with practical issues at this point, we emphasize that practical applications appear whenever the natural structure of the data is modeled as an optimization problem on a Riemannian manifold, for example, the diffusion tensor and denoising image where the data are given on the Hadamard manifold of positive definite matrices; see, for instance, [14] and references therein. As a result, the number of works dealing with concepts and techniques of nonlinear programming and convex analysis in the Riemannian scenario has also increased.

However, extending results from the Euclidean setting to Hadamard manifolds is natural, but it is not so simple; even a Hadamard manifold is, in some sense, "close" to the Euclidean space. In Kristály et al. [36], it is pointed out some conceptual mistakes within the class of Hadamard manifolds that basically reduced the geometric setting to a Euclidean one, i.e., they proved that some concepts and results established on Hadamard manifolds hold if and only if the sectional curvature is identically zero. It is worth mentioning that this is not the case in our paper. In this direction, Wang et al. [51] and Cruz Neto et al. [28] presented some counter-examples to some results established on Hadamard manifolds and extended from the Euclidean space. This illustrates that an extension from Euclidean space to Hadamard manifolds, in most cases, is not simple, and the geometric aspect of the setting must be taken into account; otherwise, some mistakes may occur. In our paper, this fact appears very clear in the counter-example that the square distance has no globally Lipschitz gradient as in the Euclidean context. In the linear setting, Bolte et al. [18] applied the global Lipschitz property of the gradient of the square distance from a point to a closed and convex set to define a barycentric projection algorithm for solving convex feasibility problems; see [18, Theorems 22 and 23].

Given the considerable impact on several fields of applied mathematics of the aforementioned concepts, we hereby study the relationship between the Kurdyka–Łojasiewicz property and error bounds on Hadamard manifolds. Additionally, we provide an application to the very important convex feasibility problem in the Riemannian context and, more generally, CAT(0) spaces.

The aim of this paper is twofold. First, we prove some properties and existence results on differential inclusion on Hadamard manifolds. These results are used to establish the relationship between the concept of error bounds and Kurdyka–Łojasiewicz inequality. Under convexity and other mild assumptions, we prove that the equivalence between error bounds and Kurdyka–Łojasiewicz inequality holds on Hadamard manifolds. As a second contribution, we study the convergence of methods for solving convex feasibility problems on Hadamard manifolds and, more generally, CAT(0) spaces. We model the convex feasibility problem as a minimization problem using square distances from a point to a closed and convex set. Therefore, by using the convexity and differentiability of the square distance and the fact that the error bounds imply the KL inequality, we obtain convergence rates for the gradient method for solving convex feasibility problems under the boundedly regularity of the sets. Additionally, replacing this last assumption by the Slater condition, we study the convergence of the alternating projection method for solving convex feasibility problems with cyclic and random order of projection on Hadamard manifolds and, more generally, CAT(0) spaces.

Next, we will summarize the main results of this paper:

- We extend some results on the existence of a solution of differential inclusions, originally proposed by Brézis [19], Bruck [21] and Bolte et al. [17], to the context of Hadamard manifolds;
- We extend the equivalence between the concept of error bounds and Kurdyka–Łojasiewicz inequality to Hadamard manifolds. This result was previously established by Bolte et al. [18] in Hilbert spaces;

- Motivated by the question raised by Bento and Melo [13], partially answered by Wang et al. [48, 49], about how to solve convex feasibility problems on Hadamard manifolds, we use the concept of Kurdyka–Łojasiewicz inequality and error bounds to analyze the convergence of the gradient method for solving convex feasibility problems on Hadamard manifolds. In [13], a convergence result is obtained if the intersection of the sets has a non-empty interior (Slater condition). In our analysis, we replace this assumption with the concept of boundedly regularity;
- Bauschke [7] proved strong convergence of the alternating projection method with random projections, provided the sets C_j satisfy the innately boundedly regularity in Hilbert spaces. Replacing this assumption with the Slater condition, we prove the convergence of this method on Hadamard manifolds. In Hilbert spaces, the Slater condition is a sufficient condition for innately boundedly regularity; see Bauschke and Borwein [8]. However, the relationship between these concepts in the Riemannian context is not established yet;
- In Bačák et al. [6], weak convergence of the alternating projection method is obtained in CAT(0) spaces for two closed and convex sets C_1 and C_2. We generalize this result for n closed and convex sets. Strong convergence of the method is obtained in [6] supposing that the sets C_1 and C_2 are boundedly regular. We extend this result to n closed and convex sets, replacing the boundedly regularity assumption by the compactness of one of the sets C_i. In a Hilbert space, the compactness of one of the sets is a sufficient condition for innately boundedly regularity; see [8]. The relationship between these concepts in a nonlinear setting is not known.

This paper is organized as follows. In Sect. 2, we present some basic definitions and results in Riemannian manifolds and CAT(0) spaces. Some results on differential inclusion on Hadamard manifolds are presented in Sect. 3. Section 4 is devoted to studying the relation between the concept of error bounds and Kurdyka–Łojasiewicz inequality on Hadamard manifolds. In Sect. 5, we present two algorithms for solving convex feasibility problems on Hadamard manifolds and, more generally, on CAT(0) spaces. Finally, some conclusions are presented in Sect. 6.

2 Notation and Basic Concepts

Next, we present the standard notations, results and preliminary concepts used throughout the paper.

2.1 Hadamard Manifolds

The standard notations, results and preliminary concepts of Riemannian geometry used throughout the paper can be found, for example, in Sakai [45] and Udriste [46]. In the Hadamard setting, we follow the same notation as in [11–13].

Throughout this paper, we will always assume that M is a finite-dimensional Hadamard manifold, unless the contrary is explicitly stated. We denote by $T_x M$ the

tangent space of M at x, by $TM = \cup_{x \in M} T_x M$ the tangent bundle of M and by $\mathcal{X}(M)$ the space of smooth vector fields over M. The parallel transport along the geodesic γ from x to y is denoted by $P_{xy}^{\gamma} : T_x M \to T_y M$ or, shortly, P_{xy}. The exponential map $\exp_x : T_x M \to M$ is defined by $\exp_x v = \gamma_v(1, x)$, for each $x \in M$.

A subset $C \subset M$ is said to be convex if, for any points p and q in C, the geodesic joining p to q is contained in C, that is, if $\gamma : [a, b] \to M$ is a geodesic such that $\gamma(a) = p$ and $\gamma(b) = q$, then $\gamma((1-t)a + tb) \in C$, for all $t \in [0, 1]$. Let $\Omega \subset M$ be an open convex set and $f : M \to \mathbb{R}$ a continuously differentiable function on Ω. The gradient vector field of f is said to be Lipschitz continuous with constant $L > 0$ on Ω if

$$||\mathrm{grad}\, f(y) - P_{xy}(\mathrm{grad}\, f(x))|| \leq L d(x, y), \ \ \forall x, y \in \Omega.$$

A function $f : M \to \mathbb{R}$ is convex if its restriction to every geodesic in M is a convex function, i.e., if for every geodesic segment $\gamma : [a, b] \to \mathbb{R}$ and every $t \in [0, 1]$,

$$f(\gamma((1-t)a + tb)) \leq (1-t) f(\gamma(a)) + t f(\gamma(b)).$$

Take $p \in M$, a vector $v \in T_p M$ is said to be a subgradient of f at p, if

$$f(q) \geq f(p) + \langle v, \exp_p^{-1}(q) \rangle,$$

for any $q \in M$. The set of all subgradients of f at p, denoted by $\partial f(p)$, is called the subdifferential of f at p.

2.2 CAT(0) Spaces

The standard notations, results and preliminary concepts of metric spaces with curvature used throughout the paper can be found, for example, Bačák [3] and Bridson and Haefliger [20]. In the CAT(0) setting, we follow the same notation as in [4, 6].

Let (X, d) be a metric space. We say that a path $\gamma : [0, 1] \to X$ is a geodesic if $d(\gamma(s), \gamma(t)) = |t - s| d(\gamma(0), \gamma(1))$, for every $t, s \in [0, 1]$. If every two points $x, y \in X$ are connected by a geodesic, then we say that (X, d) is a geodesic space. Throughout this subsection, X denotes a geodesic space. Given a geodesic triangle $\triangle(a, b, c)$ in X, there exists a comparison triangle $\triangle(a', b', c')$ in $\mathbb{R}^2$, that is, three line segments $[a', b'], [a', c']$ and $[b', c']$ such that $d(a, b) = ||a' - b'||, d(a, c) = ||a' - c'||$ and $d(b, c) = ||b' - c'||$. We call (X, d) a CAT(0) space if for every geodesic triangle with vertices $a, b, c \in X$ and $x \in [a, c]$ and $y \in [b, c]$, we have $d(x, y) \leq ||x' - y'||$, where x' and y' are the corresponding comparison points in the comparison triangle $\triangle(a', b', c')$.

There are several equivalent conditions for a geodesic metric space (X, d) to be a CAT(0) space in the Gromov's terminology (this concept also appears in the literature as Hadamard space). One of them is the following inequality, which restricted to Hadamard manifolds is equivalent to the square distance to be strongly convex (see Cruz Neto et al. [24, Corollary 3.1]): for any $x \in X$, any geodesic $\gamma : [a, b] \to X$ and any $t \in [0, 1]$ the following inequality holds

$$d(x, \gamma(t))^2 \le (1 - t)d(x, \gamma(a))^2 + td(x, \gamma(b))^2 - t(1 - t)d(\gamma(a), \gamma(b))^2.$$

This property, in particular, implies that every two points are connected by a unique geodesic. A CAT(0) space is a geodesic metric space of non-positive curvature in the sense of Alexandrov. It includes Hilbert spaces, $\mathbb{R}$-tree, Euclidean Bruhat-Tits buildings, Hadamard manifolds and many other important spaces not included above. Unless stated otherwise, from now on we denote by (H, d) a complete CAT(0) space.

Given $S \subset H$, define the distance function by dist$(x, S) = \inf\{d(x, y) : y \in S\}$, $x \in H$. The metric projection onto S is given by $P_S(x) = \{y \in S : d(x, S) = d(x, y)\}$, $x \in H$.

We now introduce the concepts of convergence in CAT(0) spaces. The notion of weak convergence in CAT(0) spaces was first introduced by Jost [35, Definition 2.7]. Let H be a complete CAT(0) space. Suppose that $\{x^k\} \subset H$ is a bounded sequence and define its asymptotic radius about a given point $x \in H$ as $r(x^k, x) = \limsup_{k \to \infty} d(x^k, x)$, and asymptotic radius as $r(x^k) = \inf_{x \in H} r(x^k, x)$. Further, we say that the point $x \in H$ is the asymptotic center of $\{x^k\}$ if $r(x^k, x) = r(x^k)$. Since H is a complete CAT(0) space, we know that the asymptotic center of $\{x^k\}$ exists and is unique; see [3, Lemma 3.1.1].

We shall say that $\{x^k\} \subset H$ weakly converges to a point $x \in H$ if x is the asymptotic center of each subsequence of $\{x^k\}$. We use the notation $x^k \overset{w}{\to} x$. If there is a subsequence $\{x^{k_j}\}$ of $\{x^k\}$ converging to a point $z \in H$, we say that z is a weak cluster point of $\{x^k\}$. It is well-known that every bounded sequence has a weak cluster point; see [3, Proposition 3.1.2] and if $C \subset H$ is a closed and convex set and $\{x^k\} \subset C$ weakly converges to $x \in H$, then $x \in C$; see [3, Lemma 3.2.1].

We denote strong convergence by "$\to$" instead of "$\overset{w}{\to}$". Clearly, $x^k \to x$ implies $x^k \overset{w}{\to} x$. These concepts of convergence are extensions to CAT(0) spaces of the notion of the weak and strong convergence in Hilbert spaces. It is known that in (finite-dimensional) Hadamard manifolds, a particular instance of CAT(0) spaces, these concepts coincide.

Next, we recall a well-known and useful concept. A sequence $\{x^k\} \subset H$ is called Fejér convergent with respect to a set $S \subset H$ if, for each $y \in S$, we have

$$d(x^{k+1}, y) \le d(x^k, y), \quad \forall k \in \mathbb{N}.$$

3 Differential Inclusion on Hadamard Manifolds

Let X be a Hilbert space. Given a smooth convex function $f : X \to \mathbb{R}$ and $x_0 \in X$, consider the following parabolic problem

$$-\chi'(t) = \nabla f(\chi(t));$$
$$\chi(0) = x \in X$$

for an unknown curve $\chi : [0, +\infty[\to X$. In other words, one looks for a curve starting at x which moves in the direction of the steepest descent of the function f; see this

problem in the Riemannian setting, for instance, in Iusem et al. [34]. We can study a similar parabolic problem after replacing the ∇f by the subgradient ∂f of a non-differentiable function f. More precisely, let $f : X \to \mathbb{R}$ be a convex function, $x \in X$ and consider the problem

$$\begin{aligned} -\chi'(t) &\in \partial f(\chi(t)); \\ \chi(0) &= x. \end{aligned} \tag{1}$$

This problem is called differential inclusion. In this work, we are interested in studying the problem (1) on Hadamard manifolds in order to apply it to prove important facts in the context of optimization.

Let (H, d) be a CAT(0) space and $f : H \to]\infty, +\infty]$ be a lower semicontinuous (in short, lsc) and convex function. Recall that f is lsc if $\{x \in H : f(x) \le \alpha\}$ is closed for each $\alpha \in \mathbb{R}$. The resolvent of f generates a semigroup of non-expansive mappings $(S_t)_{t \ge 0}$ which for a given point $x \in \overline{\mathrm{dom}}\, f$ represents a curve $\chi(t) := S_t x$ moving in the direction of the steepest descent of f. For $\lambda > 0$, the resolvent of f is defined as

$$J_\lambda(x) := \mathrm{argmin}_{y \in H} \left[f(y) + \frac{1}{2\lambda} d^2(x, y) \right],$$

with $J_0(x) := x$ for each $x \in H$. The resolvent mapping of f is non-expansive, that is, $d(J_\lambda(x), J_\lambda(y)) \le d(x, y)$, $x, y \in H$; see [3, Theorem 2.2.22]. In [3], it is shown the existence of a flow associated with the resolvent of f. In other words, if $x \in \overline{\mathrm{dom}}\, f$, then there exists the following limit

$$S_t(x) := \lim_{n \to +\infty} J^n_{\frac{t}{n}}(x), \quad t \in [0, +\infty[,$$

and it defines a non-expansive mapping $S_t : \overline{\mathrm{dom}}\, f \to \overline{\mathrm{dom}}\, f$. In addition, the limit is uniform with respect to t on bounded subintervals of $[0, +\infty[$ and $(S_t)_t$ is a strongly continuous semigroup of non-expansive mappings, that is, for every $x, y \in \overline{\mathrm{dom}}\, f$, one has that $\lim_{t \to 0^+} S_t x = x$, $S_t(S_s x) = S_{t+s} x$, for every $t, s \in [0, +\infty[$, and $d(S_t x, S_t y) \le d(x, y)$, for each $t \in [0, +\infty[$.

Remark 3.1 If the codomain of the convex function is the set of real numbers $\mathbb{R}$, then the function is continuous, and hence, it is lower semicontinuous.

The propositions below present important properties on an absolutely continuous path that will be used in our proofs.

Proposition 3.1 ([3], Theorem 5.1.11) *Let (H, d) be a CAT(0) space and $f : H \to$ $]\infty, +\infty]$ be a lsc and convex function. Assume $x \in \overline{\mathrm{dom}}\, f$ and denote $\chi(t) := S_t x$ for $t \in]0, +\infty[$. Then $t \longmapsto \chi(t)$ is absolutely continuous on $]0, +\infty[$ and satisfies*

$$\frac{1}{2} \frac{\mathrm{d}}{\mathrm{d}t} \left[d^2(y, \chi(t)) \right] + f(\chi(t)) \le f(y), \tag{2}$$

for almost every $t \in]0, +\infty[$ *and every* $y \in \overline{\text{dom}} f$. *Conversely, if an absolutely continuous curve* $z :]0, +\infty[\rightarrow H$ *with* $\lim_{t \rightarrow 0^+} z(t) = x$ *satisfies* (2), *then* $z(t) = S_t x$ *for every* $t \in]0, +\infty[$.

Proposition 3.2 ([23], Proposition 3.7) *Let M be a connected manifold equipped with a Riemannian metric $\langle \cdot, \cdot \rangle$. For any absolutely continuous path $\gamma : [a, b] \rightarrow M$, the derivative γ' exists for almost every $t \in [a, b]$ and $||\gamma'(t)||$ is integrable. In particular,*

$$L(\gamma) = \int_a^b ||\gamma'(t)|| dt.$$

Remark 3.2 It is important to note that a function $F : [a, b] \rightarrow \mathbb{R}$ is absolutely continuous if F' exists for almost every point in $]a, b[$, F' is integrable and

$$F(x) - F(a) = \int_a^x F'(t) dt, \quad \forall a \leq x \leq b.$$

Before we prove the first result of this section, we mention that if M is a Hadamard manifold and $f : M \rightarrow \mathbb{R}$ is a convex function, then for every $x \in M$, $\partial f(x)$ is non-empty, convex and compact (see [46, pp. 74 and 75]). In particular, we have that $\overline{\text{dom}} f = M$. Given any $y \in M$, in order to simplify the notation, we denote by $\rho_y(x) = \frac{1}{2} d^2(x, y)$. It is known that $\rho_y(\cdot)$ is continuously differentiable and $\text{grad} \, \rho_y(x) = -\exp_x^{-1} y$; see for instance [46].

The next result guarantees the existence and uniqueness of subgradient curves for differential inclusion on Hadamard manifolds. It extends to the Riemannian context [19, Theorem 3.2]; see also [17, Theorem 13].

Lemma 3.1 *Let M be a Hadamard manifold and $f : M \rightarrow \mathbb{R}$ be a convex function. For each $x \in M$, there is a unique absolutely continuous curve $\chi : [0, +\infty[\rightarrow M$ such that*

$$\begin{aligned} -\chi'(t) &\in \partial f(\chi(t)); \\ \chi(0) &= x \end{aligned} \tag{3}$$

for almost every $t > 0$.

Proof For each $y \in M$ from Proposition 3.1, it follows that there is a unique absolutely continuous curve $\chi : [0, +\infty[\rightarrow M$ such that $\chi(0) = x$ and

$$\frac{1}{2} \frac{d}{dt} \left[d^2(y, \chi(t)) \right] + f(\chi(t)) \leq f(y),$$

for almost every $t \in]0, +\infty[$. On the other hand, for almost every $t > 0$, we have

$$\frac{1}{2} \frac{d}{dt} \left[d^2(y, \chi(t)) \right] = \frac{d}{dt} \left[\rho_y(\chi(t)) \right] = \langle \chi'(t), \text{grad} \, \rho_y(\chi(t)) \rangle = -\langle \chi'(t), \exp_{\chi(t)}^{-1} y \rangle.$$

Hence, we conclude that $\langle -\chi'(t), \exp_{\chi(t)}^{-1} y \rangle + f(\chi(t)) \leq f(y)$ for every $y \in M$. Therefore, $-\chi'(t) \in \partial f(\chi(t))$. $\qquad \square$

Let $\chi : [0, +\infty[\to M$ be the solution to the differential inclusion (3). In the sequel, we present some properties of such a solution which can be viewed as an extension to the Riemannian setting of its linear version presented in [19]; see also [17, Theorem 13] and [18, Theorem 1]. These properties will be used in the next section in order to obtain the equivalence between the KL inequality and error bounds.

Lemma 3.2 *Let M be a Hadamard manifold and $f : M \to \mathbb{R}$ be a convex function. Let $\chi : [0, +\infty[\to M$ be the solution of (3), then $(f \circ \chi)'(t) = -||\chi'(t)||^2$, for almost every $t > 0$.*

Proof Since $-\chi'(t) \in \partial f(\chi(t))$ for almost every $t > 0$, we deduce that, for almost every $h > 0$

$$\frac{f(\chi(t+h)) - f(\chi(t))}{h} \geq -\left\langle \chi'(t), \frac{\exp_{\chi(t)}^{-1}(\chi(t+h))}{h} \right\rangle$$

and

$$\frac{f(\chi(t+h)) - f(\chi(t))}{h} \leq \left\langle \chi'(t+h), \frac{\exp_{\chi(t+h)}^{-1}\chi(t)}{h} \right\rangle.$$

Given $t \in \mathbb{R}$, where $\chi(t)$ is differentiable, consider $\epsilon > 0$ sufficiently small and the curves $\alpha : (-\epsilon, \epsilon) \to T_{\chi(t)}M$ and $\beta : (-\epsilon, \epsilon) \to M$ defined by $\alpha(h) = \exp_{\chi(t)}^{-1}\chi(t+h)$ and $\beta(h) = \exp_{\chi(t)}\alpha(h) = \chi(t+h)$. Thus, $\alpha(0) = 0$ and $\beta'(0) = \chi'(t)$. Note that

$$\beta'(0) = D(\exp_{\chi(t)})_{\alpha(0)}(\alpha'(0)) = \alpha'(0).$$

Thus, $\alpha'(0) = \chi'(t)$ and $\mathrm{grad}\, \rho_{\chi(t)}(\chi(t+h)) = -\exp_{\chi(t+h)}^{-1}\chi(t)$. On the other hand,

$$-\lim_{h \to 0^+} \left\langle \chi'(t), \frac{\exp_{\chi(t)}^{-1}\chi(t+h)}{h} \right\rangle = -\langle \chi'(t), \alpha'(0) \rangle = -||\chi'(t)||^2,$$

and

$$\lim_{h \to 0^+} \left\langle \chi'(t+h), \frac{\exp_{\chi(t+h)}^{-1}\chi(t)}{h} \right\rangle = \langle \chi'(t), -\mathrm{Hess}\, \rho_{\chi(t)}(\chi(t)) \cdot \chi'(t) \rangle$$

$$\leq -||\chi'(t)||^2.$$

The last inequality follows from the [24, Corollary 3.1]. From the above inequalities, we have

$$\lim_{h \to 0^+} \frac{f(\chi(t+h)) - f(\chi(t))}{h} = -||\chi'(t)||^2.$$

$\square$

We denote by $\partial^0 f(x)$ the least-norm element of $\partial f(x)$, i.e.,

$$||\partial^0 f(x)|| = \inf\{||v|| : \; v \in \partial f(x)\}.$$

Lemma 3.3 *Let M be a Hadamard manifold and $f : M \to \mathbb{R}$ be a convex function. If $\chi : [0, +\infty[\to M$ is the solution of* (3), *then $\chi'(t) = -\partial^0 f(\chi(t))$, for almost every $t > 0$. Furthermore, the function $x \longmapsto ||\partial^0 f(x)||$ is lsc.*

Proof For every $\delta > 0$, it follows from Lemma 3.2 and Remark 3.2 that

$$f(\chi(t + \delta)) - f(\chi(t)) = -\int_t^{t+\delta} ||\chi'(\tau)||^2 d\tau. \qquad (4)$$

Now consider $v \in \partial f(\chi(t))$. From the definition of subgradient, we have

$$\int_t^{t+\delta} ||\chi'(\tau)||^2 d\tau = -[f(\chi(t + \delta)) - f(\chi(t))] \leq -\langle v, \exp_{\chi(t)}^{-1} \chi(t + \delta) \rangle.$$

Hence, we obtain

$$\int_t^{t+\delta} ||\chi'(\tau)||^2 d\tau \leq ||v|| \cdot ||\exp_{\chi(t)}^{-1}(\chi(t + \delta))|| = ||v|| \cdot d(\chi(t + \delta), \chi(t)).$$

On the other hand,

$$d(\chi(t + \delta), \chi(t)) \leq \int_t^{t+\delta} ||\chi'(\tau)|| d\tau.$$

Thus,

$$\frac{\int_t^{t+\delta} ||\chi'(\tau)||^2 d\tau}{\delta} \leq ||v|| \cdot \frac{\int_t^{t+\delta} ||\chi'(\tau)|| d\tau}{\delta}.$$

It follows from Lebesgue's differentiation theorem that $||\chi'(t)||^2 \leq ||v|| \cdot ||\chi'(t)||$ for almost every $t > 0$. Then, $||\chi'(t)|| \leq ||v||$ and we obtain that $\chi'(t) = -\partial^0 f(\chi(t))$. To prove the second part, we recall that $||\partial^0 f(x)||$ is lsc, implying that

$$||\partial^0 f(x)|| \leq \liminf_{k \to +\infty} ||\partial^0 f(x^k)||$$

whenever $x^k \to x$. Let $\{x^{k_j}\}$ be a subsequence such that

$$\lim_{j \to +\infty} ||\partial^0 f(x^{k_j})|| = \liminf_{k \to +\infty} ||\partial^0 f(x^k)||.$$

Using the closedness of the subdifferential, we have $\partial^0 f(x^{k_j}) \to v \in \partial f(x)$; see [13]. Thus,

$$||\partial^0 f(x)|| \le ||v|| = \lim_{j \to +\infty} ||\partial^0 f(x^{k_j})|| = \liminf_{k \to +\infty} ||\partial^0 f(x^k)||$$

which concludes the proof.   □

Next, we define the slope of a function and present some properties; for more details see [3]. The slope of a function will be important to show that the curve $\chi : [0, +\infty[\to M$ is everywhere differentiable from the right, and

$$\chi'(t^+) = -\partial f^0(\chi(t)), \quad \forall t \ge 0.$$

Let (H, d) be a CAT(0) space and $f : H \to]\infty, +\infty]$ be a convex lsc function. Define the slope of f at $x \in \operatorname{dom} f$ as

$$|\partial f|(x) := \limsup_{y \to x} \frac{\max\{f(x) - f(y), 0\}}{d(x, y)}$$

and $\operatorname{dom} |\partial f| := \{x \in H : |\partial f|(x) < +\infty\}$. If $f(x) = +\infty$, we set $|\partial f|(x) = +\infty$. The next theorem presents some important properties of the slope of a function.

Theorem 3.1 ([3], Theorem 5.1.13) *Let (H, d) be a CAT(0) space and $f : H \to]\infty, +\infty]$ be a lsc and convex function. Given $x_0 \in \overline{\operatorname{dom} f}$, put $\chi(t) := S_t x_0$. Then,*

$$|\partial f|(\chi(t)) = \lim_{h \to 0^+} \frac{d(\chi(t+h), \chi(t))}{h},$$

as well as,

$$|\partial f|(\chi(t)) = \lim_{h \to 0^+} \frac{f(\chi(t)) - f(\chi(t+h))}{d(\chi(t+h), \chi(t))},$$

and also,

$$|\partial f|^2(\chi(t)) = \lim_{h \to 0^+} \frac{f(\chi(t)) - f(\chi(t+h))}{h},$$

for every $t \in (0, +\infty)$.

Corollary 3.1 *Let M be a Hadamard manifold and $f : M \to \mathbb{R}$ be a convex function, then $|\partial f|(x) \le ||\partial^0 f(x)||$ for all $x \in M$.*

Proof Given $x, y \in M$ such that $f(x) - f(y) > 0$ and $v \in \partial f(x)$, it follows from the definition of subgradient that

$$f(y) \ge f(x) + \langle v, \exp_x^{-1}(y) \rangle.$$

Thus,

$$f(x) - f(y) \leq -\langle v, \exp_x^{-1}(y) \rangle \leq ||v|| \cdot ||\exp_x^{-1}(y)|| = ||v|| d(x, y).$$

Therefore,

$$\frac{f(x) - f(y)}{d(x, y)} \leq ||v||$$

and using [3, Lemma 5.1.2], we conclude that $|\partial f|(x) \leq ||v||$. In particular, we get $|\partial f|(x) \leq ||\partial^0 f(x)||$. $\qquad \square$

The next result is an extension to the Riemannian setting of its linear version presented in [19]; see also [17, Theorem 13] and [18, Theorem 1].

Lemma 3.4 *Let M be a Hadamard manifold and $f : M \to \mathbb{R}$ be a convex function. If $\chi : [0, +\infty[\to M$ is the solution of (3), then it is everywhere differentiable from the right, and*

$$\chi'(t^+) = -\partial f^0(\chi(t)), \ \forall t \geq 0.$$

Proof Given $t \geq 0$, since $\exp_{\chi(t)} : T_{\chi(t)} \to M$ is a diffeomorphism, in order to show the assertion, it is sufficient to prove that the right-hand derivative of the curve $s \longmapsto \exp_{\chi(t)}^{-1} \chi(s)$ at t exists and coincides with $-\partial^0 f(\chi(t))$. For $h > 0$, we have

$$\frac{\exp_{\chi(t)}^{-1} \chi(t+h) - \exp_{\chi(t)}^{-1} \chi(t)}{h - 0} = \frac{\exp_{\chi(t)}^{-1} \chi(t+h)}{h},$$

and hence, the right-hand derivative at t is equal to $-\partial^0 f(\chi(t))$ if

$$\lim_{h \to 0^+} \left\| \frac{\exp_{\chi(t)}^{-1} \chi(t+h)}{h} + \partial^0 f(\chi(t)) \right\| = 0.$$

First, note that

$$\left\| \frac{\exp_{\chi(t)}^{-1} \chi(t+h)}{h} + \partial^0 f(\chi(t)) \right\|^2 = \frac{1}{h^2} d^2(\chi(t+h), \chi(t)) + ||\partial^0 f(\chi(t))||^2$$

$$+ \frac{2}{h} \langle \exp_{\chi(t)}^{-1} \chi(t+h), \partial^0 f(\chi(t)) \rangle. \qquad (5)$$

From the definition of subgradient, it follows

$$f(\chi(t+h)) - f(\chi(t)) \geq \langle \partial^0 f(\chi(t)), \exp_{\chi(t)}^{-1} \chi(t+h) \rangle.$$

Thus, from Lemma 3.2, Remark 3.2 and using (4) in the above inequality, we have

$$\langle \partial^0 f(\chi(t)), \exp^{-1}_{\chi(t)} \chi(t+h) \rangle \leq - \int_t^{t+h} ||\chi'(s)||^2 ds. \qquad (6)$$

Since the function $x \longmapsto ||\partial^0 f(x)||$ is lsc, for all $\epsilon > 0$ there is $\delta > 0$ such that if $d(y, x) < \delta$, then

$$||\partial^0 f(y)|| > ||\partial^0 f(x)|| - \epsilon.$$

Thus, for h small enough, if $t \leq s \leq t + h$ it follows from (6) together with the first part of Lemma 3.3 that

$$\langle \partial^0 f(\chi(t)), \exp^{-1}_{\chi(t)} \chi(t+h) \rangle \leq -h(||\partial^0 f(\chi(t))|| - \epsilon)^2.$$

Then, we obtain $\frac{1}{h} \langle \partial^0 f(\chi(t)), \exp^{-1}_{\chi(t)} \chi(t+h) \rangle \leq -||\partial^0 f(\chi(t))||^2$. Now, combining (5), Theorem 3.1 and Corollary 3.1, we have $\lim_{h \to 0^+} || \frac{\exp^{-1}_{\chi(t)} \chi(t+h)}{h} + \partial^0 f(\chi(t))|| = 0$. $\qquad \square$

Corollary 3.2 *Let M be a Hadamard manifold and $f : M \to \mathbb{R}$ be a convex function, then $|\partial f|(x) = ||\partial^0 f(x)||$ for all $x \in M$. Furthermore, $(f \circ \chi)(t)$ is everywhere differentiable from the right and $(f \circ \chi)'(t^+) = -||\chi'(t^+)||^2$ for all $t \geq 0$.*

Proof Given $x \in M$, let $\chi : [0, +\infty[\to M$ be the solution of (3). Fix $t \geq 0$ and consider the curve $\alpha : [0, \epsilon[\to T_{\chi(t)}M$ defined by $\alpha(h) = \exp^{-1}_{\chi(t)}(\chi(t+h))$. Then, $\alpha(0) = 0$ and $\chi(t+h) = \exp_{\chi(t)}(\alpha(h))$. Since $\exp_{\chi(t)} : T_{\chi(t)}M \to M$ is a diffeomorphism, it follows from Lemma 3.4 that

$$-\partial^0 f(\chi(t)) = \chi'(t^+) = d(\exp_{\chi(t)})_{\alpha(0)}(\alpha'(0^+)) = \alpha'(0^+).$$

On the other hand, using Theorem 3.1 we obtain

$$||\alpha(0^+)|| = \lim_{h \to 0^+} \frac{\left\| \exp^{-1}_{\chi(t)}(\chi(t+h)) \right\|}{h} = \lim_{h \to 0^+} \frac{d(\chi(t+h), \chi(t))}{h} = |\partial f|(\chi(t)).$$

Thus, $||\partial^0 f(\chi(t))|| = |\partial f|(\chi(t))$. Hence, we conclude that $||\partial^0 f(x)|| = |\partial f|(x)$. Again, from Theorem 3.1, we obtain that the function $(f \circ \chi)(t)$ is everywhere differentiable from the right and

$$(f \circ \chi)'(t^+) = -|\partial f|^2(\chi(t)) = -||\partial^0 f(\chi(t))||^2 = -||\chi'(t^+)||^2.$$

$\qquad \square$

In the next results we assume that $M^* := \operatorname{argmin} f = \{x \in M : f(x) \leq f(z), \ \forall z \in M\} \neq \emptyset$. It is an extension to the Riemannian setting of its linear version presented in [19]; see also [18, Theorem 1].

Lemma 3.5 *Let M be a Hadamard manifold and $f : M \to \mathbb{R}$ be a convex function. If $\chi : [0, +\infty[\to M$ is the solution of (3), then for all $p \in M^*$, $\lim_{t \to +\infty} d(\chi(t), p)$ exists.*

Proof Note that, for each $p \in M^*$ and for almost every $t \in]0, +\infty[$,

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[\frac{1}{2}d^2(\chi(t), p)\right] = \frac{\mathrm{d}}{\mathrm{d}t}[\rho_p(\chi(t))] = -\langle \chi'(t), \exp_{\chi(t)}^{-1}(p)\rangle.$$

On the other hand, since $-\chi'(t) \in \partial f(\chi(t))$ and $p \in M^*$, we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[\frac{1}{2}d^2(\chi(t), p)\right] \le f(p) - f(\chi(t)) \le 0,$$

for almost every $t \in]0, +\infty[$. Thus, we conclude that the function $d^2(\chi(t), p)$ is non-increasing and $d^2(\chi(t), p) \ge 0$, hence $d^2(\chi(t), p)$ converges as $t \to +\infty$. □

Remark 3.3 Given $\epsilon > 0$ and $p \in M^*$ note that if $\chi : [0, +\infty[\to M$ is the solution (3) and $\chi(0) = x \in B(p, \epsilon)$, since $d^2(\chi(t), p)$ is non-increasing, then $\chi(t) \in B(p, \epsilon)$, for all $t \ge 0$.

The next lemma was proved by Bačák [3, Proposition 5.1.12 and Theorem 5.1.16] in the CAT(0) setting. However, we state it in the particular instance of Hadamard manifolds. It extends to a more general setting, the result proposed by [21] in Hilbert spaces.

Lemma 3.6 *Let M be a Hadamard manifold and $f : M \to \mathbb{R}$ be a convex function. Then, the solution $\chi : [0, +\infty[\to M$ of (3) converges to a minimizer of f as $t \to \infty$. Furthermore, $\lim_{t \to +\infty} f(\chi(t)) = \min f$.*

4 KL Inequality and Error Bounds on Hadamard Manifolds

In this section, we study the relationship between the Kurdyka–Łojasiewicz property and the concept of error bounds. Let $f : M \to \mathbb{R}$ be a convex function. It is worth to mention that the Kurdyka–Łojasiewicz property does not require f to be convex; see for instance [12]. To simplify the notation, we write $[f < \mu] = \{x \in M : f(x) < \mu\}$ and, for a given $r > 0$, $K(0, r)$ stands for the set of continuous functions $\varphi : [0, r) \to \mathbb{R}_+$ such that $\varphi(0) = 0$ and it is continuously differentiable in $(0, r)$ with $\varphi'(x) > 0$ for all $x \in (0, r)$.

A function $f : M \to \mathbb{R}$ is said to have the Kurdyka–Łojasiewicz property at a point $\bar{x} \in M$ if there exists $\eta \in (0, +\infty]$, a neighborhood U of $\bar{x}$ and $\varphi \in K(0, \eta)$ (called desingularizing function) such that the Kurdyka–Łojasiewicz inequality holds

$$\varphi'(f(x) - f(\bar{x}))\|\partial^0 f(x)\| \ge 1, \quad \forall x \in U \cap [f(\bar{x}) < f < f(\bar{x}) + \eta]. \qquad (7)$$

This property basically means that a function can be made sharp by a reparameterization of its values. One can check that the Kurdyka–Łojasiewicz property is

automatically satisfied at any non-critical point $\bar{x} \in \text{dom } \partial f$, where f is a lower semi-continuous function; see [26, Lemma 3.6]. This is why some applications focus just on the case where $\bar{x}$ is a critical point. The most fundamental works on this subject are of course due to Łojasiewicz [39] and Kurdyka [37].

Consider a non-decreasing function $\omega : [0, +\infty) \to [0, +\infty)$ with $\omega(0) = 0$. The function f achieves its minimum $\min f$ so that $M^* = \arg \min f \neq \emptyset$ satisfies the local error bound with residual function ω if

$$\omega(f(x) - \min f) \geq \text{dist}(x, M^*),$$

where x may belong to either the whole space or a bounded set. Equivalently, if $\min f = 0$, then there exists $r > 0$ such that

$$(\omega \circ f)(x) \geq \text{dist}(x, M^*), \quad \forall x \in [0 \leq f \leq r].$$

Additionally, we say that $\varphi \in K(0, r)$ has moderate behavior (near the origin) if, for some constant $c > 0$, it satisfies

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

Example 4.1 Let $\varphi : [0, r) \to \mathbb{R}$ given by $\varphi(t) = \alpha t^{\frac{1}{p}}$ with $\alpha > 0$ and $p \geq 1$. One can check that $\varphi \in K(0, r)$ and it has moderate behavior with $c = \frac{1}{p}$.

Example 4.2 Take $\varphi : [0, r) \to \mathbb{R}$ such that $\varphi \in K(0, r)$ and it is semi-algebraic or subanalytic. Then, it has a moderate behavior; see [18, Lemma 4].

Given $x \in \overline{\text{dom } \partial f}$, from Lemma 3.1 there exists a unique absolutely continuous curve $\chi : [0, +\infty) \to M$ such that $\chi(0) = x$ and, for almost every $t > 0$,

$$-\chi'(t) \in \partial f(\chi(t)).$$

The next result provides the equivalence between the KL inequality and the existence of a uniform bound for the lengths of subgradient trajectories verifying a differential inclusion. It generalizes Bolte et al. [17, Theorem 18] and Bolte et al. [18, Theorem 27] to the Riemannian context.

Proposition 4.1 *Let M be a Hadamard manifold and $f : M \to \mathbb{R}$ be a convex function. Suppose that $\bar{x} \in S$, $\rho > 0$ and $\varphi \in K(0, r)$. Then, the following assertions are equivalent:*

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

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

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

$$\int_t^s \|\chi'(\tau)\| d\tau \leq \varphi(f(\chi(t))) - \varphi(f(\chi(s))).$$

Proof Let us prove the assertion (i) $\Rightarrow$ (ii). Take $x \in B(\bar{x}, \rho) \cap [0 < f < r_0]$ and $0 \le t < s$. One has

$$
\varphi(f(\chi(t))) - \varphi(f(\chi(s))) = \int_s^t \frac{\mathrm{d}}{\mathrm{d}\tau}\, \varphi(f(\chi(\tau)))\mathrm{d}\tau
$$

$$
= \int_s^t \varphi'(f(\chi(\tau)))\frac{\mathrm{d}}{\mathrm{d}\tau}(f(\chi(t)))\mathrm{d}\tau. \tag{8}
$$

From Lemmas 3.2 and 3.3, we have that $\frac{\mathrm{d}}{\mathrm{d}\tau}(f(\chi(\tau))) = -||\chi'(\tau)||^2$, for almost every $\tau > 0$, and $\chi'(\tau) = -\partial^0 f(\chi(\tau))$, for almost every $\tau > 0$, respectively. Using these facts in (8), we have

$$
\varphi(f(\chi(t))) - \varphi(f(\chi(s))) = \int_t^s \varphi'(f(\chi(\tau)))||\partial^0 f(\chi(\tau))||||\chi'(\tau)||\mathrm{d}\tau.
$$

From Remark 3.3, we have that $\chi(t) \in B(\bar{x}, \rho)$, for all $t \ge 0$. Furthermore, since $x \in [0 < f < r]$ and from Lemma 3.2, we have that $(f \circ \chi)(t)$ is non-increasing, then $\chi(t) \in [0 < f < r]$, for all $t \ge 0$. Therefore, we have

$$
\chi(t) \in \mathrm{dom}\,\partial f \cap B(\bar{x}, \rho) \cap [0 < f < r], \quad \forall t \ge 0,
$$

and hence,

$$
\varphi'(f(\chi(\tau)))||\partial^0 f(\chi(\tau))|| \ge 1,
$$

for all $\tau > 0$. Thus,

$$
\varphi(f(\chi(t))) - \varphi(f(\chi(s))) = \int_t^s \varphi'(f(\chi(\tau)))||\partial^0 f(\chi(\tau))||||\chi'(\tau)||\mathrm{d}\tau
$$

$$
\ge \int_t^s ||\chi'(\tau)||\mathrm{d}\tau
$$

and the first part is proved.

Now, to prove that $(ii) \Rightarrow (i)$, take $x \in \mathrm{dom}\,\partial f \cap B(\bar{x}, \rho) \cap [0 < f < r]$. For each $h > 0$, we have

$$
\frac{1}{h}\int_0^h ||\chi'(\tau^+)||\,\mathrm{d}\tau = \frac{1}{h}\int_0^h ||\chi'(\tau)||\mathrm{d}\tau \le -\frac{\varphi(f(\chi(h))) - \varphi(f(\chi(0)))}{h}.
$$

Since $\chi_x'(t^+) = -\partial^0 f(\chi_x(t))$ for all $t \ge 0$ and the function $x \longmapsto ||\partial^0 f(x)||$ is lsc, given $\epsilon > 0$ there is $\delta > 0$ such that if $d(y, x) < \delta$, then $||\partial^0 f(y)|| > ||\partial^0 f(x)|| - \epsilon$. Thus, for h small enough,

$$
||\chi'(0^+)|| - \epsilon \le -\frac{\varphi(f(\chi(h))) - \varphi(f(\chi(0)))}{h}.
$$

Letting $h \to 0^+$, we obtain that $||\chi'(0^+)|| - \epsilon \leq -\varphi'(f(x))(f \circ \chi)'(0^+)$, for any $\epsilon > 0$. In particular,

$$||\chi'(0^+)|| \leq -\varphi'(f(x))(f \circ \chi)'(0^+).$$

Therefore, applying Lemma 3.2, we have $||\chi'(0^+)|| \leq \varphi'(f(x))||\chi'(0^+)||^2$. Using Lemma 3.4, we finally obtain

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

and the proof is completed.   □

The next result provides the equivalence between the KL property and global error bounds under the moderate behavior assumption for convex functions. Moreover, it shows that the desingularizing function in the KL inequality and the residual function in the error bounds are essentially the same, up to a multiplicative constant. This result extends Bolte et al. [18, Theorem 5] to the Riemannian context, and it will be used in the next section to obtain the convergence of the gradient method for solving convex feasibility problems.

Theorem 4.1 *Let M be a Hadamard manifold and $f : M \to \mathbb{R}$ be a convex function. Consider $r > 0$, $c > 0$, $\rho > 0$, $\varphi \in K(0, r_0)$ and $\bar{x} \in M^* = \arg\min f$.*

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

$$dist\,(x, M^*) \leq \varphi(f(x)), \quad \forall x \in [0 < f < r_0] \cap B(\bar{x}, \rho).$$

(ii) *(Error bounds with moderate behavior implies the KL inequality) If $s\varphi'(s) \geq c\varphi(s)$ for all $s \in (0, r_0)$ and*

$$dist\,(x, M^*) \leq \varphi(f(x)), \quad x \in [0 < f < r_0] \cap B(\bar{x}, \rho),$$

then

$$\varphi'(f(x))||\partial^0 f(x)|| \geq c, \quad \forall x \in [0 < f < r_0] \cap B(\bar{x}, \rho).$$

Proof (i) From Proposition 4.1 for every $x \in [0 < f < r_0] \cap B(\bar{x}, \rho)$ and $0 \leq t < s$, we have

$$d(\chi(s), \chi(t)) \leq \int_t^s ||\chi'(\tau)||d\tau \leq \varphi(f(\chi(t))) - \varphi(f(\chi(s))), \quad (9)$$

where $\chi : [0, \infty[\to M$ is a subgradient curve such that $\chi(0) = x$. From Lemma 3.6, we have that $\lim_{t \to +\infty} \chi(t) = x^* \in M^*$ and $\lim_{t \to +\infty} (f \circ \chi)(t) = \min f = 0$. Thus, taking $t = 0$ and letting $s \to \infty$ in (9), we obtain $d(x^*, x) \leq \varphi(f(x))$, and so, $dist\,(x, M^*) \leq \varphi(f(x)), \quad \forall\, x \in [0 < f < r_0] \cap B(\bar{x}, \rho)$.

(ii) Take $x \in [0 < f < r_0] \cap B(\bar{x}, \rho)$ and let $y = P_{M^*}(x)$ be the projection of x onto M^*. By convexity of f, we have $0 = f(y) \geq f(x) + \langle \partial^0 f(x), exp_x^{-1} y \rangle$. From Cauchy–Schwarz inequality and the moderate behavior of φ, we have

$$
\begin{aligned}
f(x) \leq \|\partial^0 f(x)\| \cdot d(x, y) &= \|\partial^0 f(x)\| \cdot d(x, M^*) \\
&\leq \|\partial^0 f(x)\| \cdot \varphi(f(x)) \\
&\leq \frac{1}{c} \cdot f(x)\varphi'(f(x)) \cdot \|\partial^0 f(x)\|.
\end{aligned}
$$

Since $f(x) > 0$, then the desired result follows immediately.   □

5 Application to Convex Feasibility Problem

The convex feasibility problem (CFP) is formulated by:

$$
\text{to find a point } x \in C := \bigcap_{i=1}^n C_i,
$$

where each C_i is a subset of a space X, for each $i = 1, \ldots, n$ with $n \geq 2$.

The aim of this section is to consider two classical algorithms for solving CFP on Hadamard manifolds. The first one is the gradient method, which is one of the oldest methods to find a minimizer of differentiable functions in Euclidian spaces. In the Riemannian context, it was first studied by Luenberger [40]; see also Cruz Neto et al. [27]. In the linear setting, Bolte et al. [18] applied the gradient method and the global Lipschitz (with $L = 1$) property of the gradient of the square distance from a point to a closed and convex set to define a barycentric projection algorithm for solving CFP. We show that this Lipschitz property does not hold in general Hadamard manifolds. On the other hand, we use the equivalence between the KL inequality and error bounds to prove that the gradient method can be applied to solve CFP, obtaining some convergence rates under the boundedly linearly regularity of the sets $C_i, i = 1, \ldots, n$. The second method considered is the alternating projection method. We show convergence of this method under the Slater condition for cyclic and random projections on Hadamard manifolds and more general CAT(0) spaces.

Under the Slater condition, Bento and Melo [13] analyzed the convergence of a subgradient method for solving a convex feasibility problem in Riemannian manifolds with nonnegative sectional curvature. Their results were extended to Riemannian manifolds with sectional curvature bounded from below by Wang et al. [48, 49]. Bačák et al. [6] proved the convergence of von Neumann's algorithm (alternating projection method) in complete CAT(0) spaces (also known as Hadamard spaces), which include Hadamard manifolds, among others. The key to their approach is the regularity condition in the intersection of the sets C_i. We say that $A, B \subset X$ are boundedly linearly regular if, for any bounded set $S \subset X$, there exists $\sigma > 0$ such that for $x \in S$, we have

$$
\text{dist}(x, A \cap B) \leq \sigma \max\{\text{dist}(x, A), \text{dist}(x, B)\}.
$$

We say that $A, B \subset X$ are linearly regular if there exists $\sigma > 0$ such that, for $x \in X$, we have

$$\text{dist}(x, A \cap B) \leq \sigma \max\{\text{dist}(x, A), \text{dist}(x, B)\}.$$

This concept is well-known in the CFP literature, and it is also known as (local) error bounds for CFP; see Beck and Teboulle [10]. Error bounds for CFP were introduced by Gubin et al. [30] using the residual function

$$T(x) = \max_{1 \leq i \leq n} \text{dist}\,(x, C_i).$$

Indeed, in this case, one obviously has $T(x) = 0$ if and only if $x \in C = \cap_{i=1}^{n} C_i$. Notice that while $\text{dist}\,(x, C)$ is usually difficult to estimate (otherwise the original problem is trivial), the error bound $T(x)$ is in many cases easily calculated if the distance from a point to each set C_i is known. Thus, error bounds state that we can bound an unknown quantity by a computable quantity.

In the Euclidean context, Beck and Teboulle [10] studied the interplay between the Slater condition and local error bounds. They proved that the Slater condition implies local error bounds; see [10, Lemma 3.1] and [18, Proposition 11]. They also showed that error bounds are central in establishing the (linear) rate of convergence of iterative methods, and without the Slater hypothesis on CFP, projection-type algorithms could in fact behave quite badly. On the one hand, this clarifies the important role played by the Slater condition and, on the other, by the local error bound on CFP.

In the Riemannian context, the connection between these two concepts is not known. Therefore, in this section, we apply two different methodologies to solve a convex feasibility problem under the Slater condition and local error bounds (or boundedly linearly regular).

5.1 Gradient Method on Hadamard Manifolds

Let M be a Hadamard manifold, and consider $f : M \to \mathbb{R}$ given by

$$f(x) = \frac{1}{2} \sum_{i=1}^{n} \alpha_i \text{dist}^2(x, C_i), \tag{10}$$

where $\alpha_i > 0$, for all $i = 1, 2, \ldots, n$, $\sum_{i=1}^{n} \alpha_i = 1$ and $C_i \subset M$, $i = 1, \ldots, n$. If $C := \bigcap_{i=1}^{n} C_i$ is non-empty, then finding a point in C is equivalent to minimizing the function f over M. In this case, it is clear that $C = \arg\min f = \{x \in M : f(x) = 0\}$. This approach was considered in the linear setting (with convex sets) by Bolte et al. [18] and (with non-convex sets) by Luke et al. [41]. In the next results of this subsection, f is given as in (10). We will prove that the gradient method can be used to solve a convex feasibility problem on Hadamard manifolds. In Ferreira et al. [29, Sect. 4], one can find the expression of the gradient of a differentiable function in terms of the Euclidean gradient, for some Hadamard manifolds, as well as in [18, p. 496] is

presented the Euclidean gradient of f given as in (10). In a general Hadamard manifold, it follows from Poly and Raby [43] that $\text{dist}^2(\cdot, C_i)$ is continuously differentiable and $\text{grad dist}^2(x, C_i) = -\exp_x^{-1} p_i$, where p_i is the metric projection of x onto C_i, for each $i = 1, \ldots, n$; see also Barani and Hosseini [5, Example 3.1].

Alternatively, even when the objective function is non-differentiable at some iterations, it remains possible to apply a gradient (sampling) method for minimizing a locally Lipschitz (and hence, almost everywhere differentiable) function. As shown by Hosseini and Uschmajew [33, Theorem 4.4], under mild assumptions, this method finds a solution on Riemannian manifolds. Actually, in a general Riemannian manifold, the scenario of almost everywhere differentiability is natural since, according to [52, Theorem 2], for a complete n-dimensional Riemannian manifold M where $d^2(\cdot, p)$ (with $p \in M$ fixed) is directional-differentiable on all M, then M is diffeomorphic to $\mathbb{R}^n$.

Gradient method
Initialization: Choose $x^0 \in M$;
Stopping rule: Given x^k, if x^k is a critical point of f, then set $x^{k+p} = x^k$ for all $p \in \mathbb{N}$. Otherwise, compute the iterative step;
Iterative step: Take as the next iterate any $x^{k+1} \in M$ such that

$$x^{k+1} = \exp_{x^k}(-t_k \text{grad} f(x^k)), \tag{11}$$

where $\{t_k\}$ is a positive and bounded sequence satisfying

$$f(x^{k+1}) \leq f(x^k) - \alpha t_k \|\text{grad } f(x^k)\|^2, \tag{12}$$

with $\alpha \in (0, 1)$.

Throughout this subsection, we suppose that $\{x^k\}$ is the sequence generated by the gradient method (11) with a stepsize satisfying (12). Note that the above algorithm presents a conceptual stopping rule. From (11), we have

$$t_k \|\text{grad} f(x^k)\| = d(x^{k+1}, x^k). \tag{13}$$

Thus, if $\liminf t_k > 0$ (this is the case for the stepsizes defined in the sequel) and $x^{k+1} = x^k$, then from (13) we have that x^k is a critical point of f. Numerically, for a given tolerance $\epsilon > 0$ due to (13) a natural stopping rule is $d(x^{k+1}, x^k) < \epsilon$ which is equivalent to $\|\text{grad} f(x^k)\| < \epsilon$. Now, let us describe some concrete possibilities for the choice of the stepsize sequence $\{t_k\}$.

Example 5.1 Let $\{t_k\}$ be the sequence defined as follows: given $\delta_1, \delta_2 > 0$ such that $L\delta_1 + \delta_2 < 1$, where L is the Lipschitz constant associated to the gradient map of f. Take $\{t_k\}$ such that

$$t_k \in \left(\delta_1, \frac{2}{L}(1 - \delta_2)\right), \quad \forall k \geq 0. \tag{14}$$

The sequence $\{t_k\}$ as in (14) is called the fixed stepsize rule. Since (14) holds and $L\delta_1 + \delta_2 < 1$, from [27, Theorem 5.1], we have

$$f(x^{k+1}) + \beta_k t_k^2 \|\text{grad } f(x^k)\|^2 \leq f(x^k), \quad \forall k \geq 0, \tag{15}$$

where $\beta_k := \left(\dfrac{1}{t_k} - \dfrac{L}{2}\right) > 0$. Note that, from (14), we have $\beta_k > \beta := \dfrac{L\delta_2}{2(1-\delta_2)} > 0, \quad \forall k \geq 0$.

Example 5.2 Fix $\alpha \in (0, 1)$. Let $\{t_k\}$ be a sequence obtained by $t_k := \max\{2^{-j} : j \in \mathbb{N}\}$ such that

$$f\left(\exp_{x^k}(-2^{-j}\text{grad} f(x^k))\right) \leq f(x^k) - \alpha 2^{-j}\|\text{grad} f(x^k)\|^2\}. \tag{16}$$

This stepsize rule is the so-called Armijo search. Note that, in this case, zero can be a cluster point of the sequence $\{t_k\}$. However, if the gradient map of f, $x \mapsto \text{grad} f(x)$, is a Lipschitz function, then zero is not a cluster point of $\{t_k\}$.

Remark 5.1 In [46, Theorem 4.2], Udriste established the convergence of the gradient method using the stepsizes defined as above; see also Cruz Neto et al [27]. Clearly, in the linear setting, this result can be applied to obtain a solution to the convex feasibility problem through the function given in (10) whose gradient is Lipschitz with constant $L = 1$. However, this is not necessarily true in general Hadamard manifolds. To overcome this drawback, we provide a different approach to obtain the convergence of the method and its convergence rate using the concept of Kurdyka–Łojasiewicz inequality.

Next, we provide a counterexample showing that the gradient of the square distance is not globally Lipschitz mapping in a general Hadamard manifold.

Example 5.3 Let us consider the hyperbolic space as the model of the upper half-plane $\mathbb{H}^2 = \{(x, y) \in \mathbb{R}^2 : y > 0\}$, i.e., the upper half-plane furnished with the metric $\langle u, v \rangle_p = \dfrac{\langle u, v \rangle}{y^2}$, where $u, v \in \mathbb{R}^2$, $p = (x, y)$ with $y > 0$ and $\langle \cdot, \cdot \rangle$ is the usual inner product in $\mathbb{R}^2$. Consider the point $p = (0, 1)$, $d : \mathbb{H}^2 \times \mathbb{H}^2 \to \mathbb{R}$ the distance function and denote by $d(x) = d(x, p)$. Thus, $\rho_p : \mathbb{H}^2 \to \mathbb{R}$ is given by $\rho_p(x) = \frac{1}{2}d^2(x)$. We will show that grad ρ_p is not global Lipschitz. Given $\alpha \in (0, \frac{\pi}{2})$, let us consider the geodesic passing through the point $P_3 = p$ making an angle α with the y-axis. It is not difficult to verify that this geodesic is part of the circle with center in $\left(\frac{\cos\alpha}{\sin\alpha}, 0\right)$ and radius $\frac{1}{\sin\alpha}$.

Now, let us consider the geodesic triangle whose vertices are the points $P_1 = (0, \frac{1}{\sin\alpha})$, $P_2 = (\frac{\cos\alpha}{\sin\alpha}, \frac{1}{\sin\alpha})$ and $P_3 = p$; see Fig. 1. Denote by A the area of this geodesic triangle and note that $A = A_1 + A_2$, where A_1 and A_2 are the areas below and above the straight line connecting P_1 and P_2, respectively. Consider the global parametrization $\varphi : S \to \mathbb{H}^2$, where $\varphi(x, y) = (x, y)$ and $S = \{(x, y) \in \mathbb{R}^2 : y > 0\}$. The area A_1 is given by
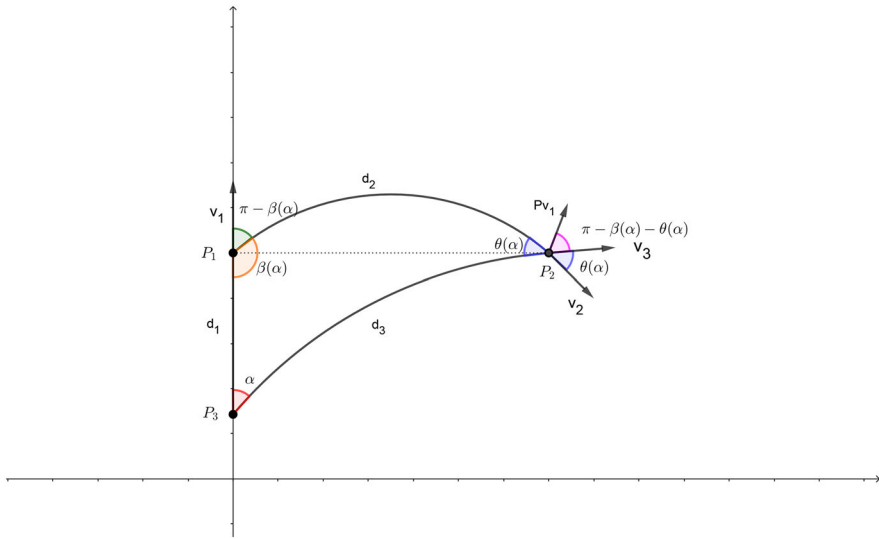
Fig. 1 Square distance is not globally Lipschitz in general Hadamard manifolds

$$\iint_{\varphi^{-1}(B_1)} \frac{1}{y^2}\, \mathrm{d}x\, \mathrm{d}y,$$

where B_1 is the geodesic triangle whose vertices are the points P_1, P_2 and P_3. Observe that

$$B_1 = \left\{ (x, y) : 0 \le x \le \frac{\cos\alpha}{\sin\alpha}, \quad \frac{\sqrt{1 - (\sin\alpha\, x - \cos\alpha)^2}}{\sin\alpha} \le y \le \frac{1}{\sin\alpha} \right\}.$$

Straightforward calculations yield $A_1(\alpha) = -\cos\alpha + \arcsin(\cos\alpha)$. To calculate A_2 it is necessary to know the geodesic joining P_1 to P_2. Again, it is not difficult to verify that such geodesic is part of the circle with center in $\left(\frac{\cos\alpha}{2\sin\alpha}, 0\right)$ and radius $\frac{\sqrt{4+\cos^2\alpha}}{2\sin\alpha}$. Thus, similar to A_1, we obtain that $A_2(\alpha) = \cos\alpha - 2\arcsin(\frac{\cos\alpha}{\sqrt{4+\cos^2\alpha}})$. Then, we conclude $A(\alpha) = \arcsin(\cos\alpha) - 2\arcsin(\frac{\cos\alpha}{\sqrt{4+\cos^2\alpha}})$ and hence,

$$\lim_{\alpha \to 0} A(\alpha) = \frac{\pi}{2} - 2\arcsin\left(\frac{1}{\sqrt{5}}\right). \tag{17}$$

On the other hand, if we denote by $\beta(\alpha)$ and $\theta(\alpha)$ the internal angles of the geodesic triangle in the vertices P_1 and P_2, respectively, it follows from the Gauss–Bonnet Theorem that

$$A(\alpha) = \pi - (\alpha + \beta(\alpha) + \theta(\alpha)). \tag{18}$$

Since $0 < \arcsin\left(\frac{1}{\sqrt{5}}\right) < \frac{\pi}{6}$, for α sufficiently close to 0, we have that $\frac{\pi}{2} - \frac{\pi}{3} < \pi - (\beta(\alpha) + \theta(\alpha)) < \frac{\pi}{2}$. Hence, $\frac{\pi}{2} < \beta(\alpha) + \theta(\alpha) < \frac{5\pi}{6}$. Letting $d_1 = d(P_3, P_1)$, $d_2 = d(P_1, P_2)$ and $d_3 = d(P_3, P_2)$, one has

$$d_1(\alpha) = -\ln(\sin \alpha), \ d_2(\alpha) = \ln\left(\frac{\sqrt{4 + \cos^2 \alpha} + \cos \alpha}{\sqrt{4 + \cos^2 \alpha} - \cos \alpha}\right) \text{ and}$$

$$d_3(\alpha) = \frac{1}{2}\ln\left(\frac{1 + \cos \alpha}{1 - \cos \alpha}\right).$$

Consider the vectors $v_1(\alpha) := \operatorname{grad} \rho_{P_3}(P_1(\alpha))$, $v_2(\alpha) := -\exp^{-1}_{P_2(\alpha)}(P_1(\alpha))$ and $v_3(\alpha) := \operatorname{grad} \rho_{P_3}(P_2(\alpha))$, thereby $d_1 = ||v_1||$ and $d_3 = ||v_3||$. Now consider $Pv_1(\alpha) = P_{P_1(\alpha)P_2(\alpha)}(v_1(\alpha))$ the parallel transport of $v_1(\alpha)$ along the geodesic joining P_1 to P_2, then $d_1 = ||v_1|| = ||Pv_1||$ and

$$\angle(Pv_1, v_2) = \pi - \beta(\alpha), \ \angle(v_3, v_2) = \theta(\alpha), \ \angle(Pv_1, v_3) = \pi - \beta(\alpha) - \theta(\alpha).$$

Therefore, we obtain $||v_3 - Pv_1||^2 = (d_3 - d_1)^2 + 2d_1 d_3 (1 + \cos(\beta(\alpha) + \theta(\alpha)))$. From (17) and (18), when $\alpha \to 0$ we have that $\cos(\beta(\alpha) + \theta(\alpha)))$ converges to $-\frac{4}{5}$. Thus,

$$\lim_{\alpha \to 0} ||\operatorname{grad} \rho_{P_3}(P_2(\alpha)) - P_{P_1(\alpha)P_2(\alpha)}(\operatorname{grad} \rho_{P_3}(P_1(\alpha)))|| = \lim_{\alpha \to 0} ||v_3 - Pv_1||^2(\alpha)$$
$$= +\infty.$$

On the other hand, $\lim_{\alpha \to 0} d(P_1(\alpha), P_2(\alpha)) = \lim_{\alpha \to 0} d_2(\alpha) = \ln\left(\frac{\sqrt{5}+1}{\sqrt{5}-1}\right)$. Therefore, there is no constant $L > 0$ such that

$$||\operatorname{grad} \rho_p(P_2(\alpha)) - P_{P_1(\alpha)P_2(\alpha)}(\operatorname{grad} \rho_p(P_1(\alpha)))|| \le L \cdot d(P_1(\alpha), P_2(\alpha)).$$

This shows that the square distance is not globally Lipschitz in general Hadamard manifolds.

The next result was proved by Attouch et al. [2, Theorem 3.2] in the Euclidean setting. Assuming that the objective function is bounded from below and continuously differentiable with Lipschitz gradient, in particular, they obtained convergence of the gradient method to a critical point of KL functions as long as the sequence generated by the method is bounded. In Remark 3.3, they have pointed out that the conclusion of the theorem remains unchanged if the assumptions that ∇f is Lipschitz continuous on $\mathbb{R}^n$ and f is a KL function are replaced by a locally version of these assumptions, i.e., there exists a closed subset S of $\mathbb{R}^n$ such that ∇f is Lipschitz continuous on $\operatorname{co} S$ and f satisfies the KL inequality at each point of S, where $\operatorname{co} S$ denotes the convex envelope of S.

As an application of the equivalence between the KL property and global error bounds (under the moderate behavior assumption) in Hadamard manifolds, we obtain the convergence of the gradient method for solving a convex feasibility problem as well as its convergence rate. To this end, we denote by $L > 0$ the Lipchitz constant

of the gradient of f in a compact set which contains $C = \bigcap_{i=1}^{n} C_i$ and t_* the lower bound of the sequence $\{t_k\}$ given in Examples 5.1 and 5.2. Recall that from (10) if $x^* \in C$, then $f(x^*) = 0$.

Theorem 5.1 *Assume that $C = \bigcap_{i=1}^{n} C_i$ is non-empty and compact, and the sets C_i are boundedly linearly regular. Then, the following assertions hold:*

(i) *$\{x^k\}$ converges to a point $x^* \in C$;*
(ii) *There exists a constant $0 < q < 1$ such that $f(x^k) \leq f(x^0)q^{k-1}$;*
(iii) *There exist constants $c > 0$ and $0 < \mu < 1$ such that $d(x_k, C) \leq c\mu^{k-1}$.*

Proof (i): Let $r > 0$ be fixed. Since the sets C_i are boundedly linearly regular, we have

$$\operatorname{dist}(x, C) \leq \sigma \max\{\operatorname{dist}(x, C_i), \ i = 1, \ldots, n\}, \quad \forall x \in [0 < f < r].$$

Thus, for every $x \in [0 < f < r]$, we have

$$
\begin{aligned}
\operatorname{dist}(x, C) &\leq \sigma \max\{\operatorname{dist}(x, C_i), \ i = 1, \ldots, n\} \\
&\leq \sigma \sqrt{\operatorname{dist}^2(x, C_1) + \cdots + \operatorname{dist}^2(x, C_n)} \\
&\leq \frac{\sigma}{\alpha} \sqrt{f(x)},
\end{aligned}
\tag{19}
$$

where $\alpha = \left(\dfrac{\min_{i=1,\ldots,n}\{\alpha_i\}}{2} \right)^{\frac{1}{2}}$. From (19), we have that $d(x^k, P_C(x^k)) = \operatorname{dist}(x^k, C) \leq \dfrac{\sigma}{\alpha}\sqrt{f(x^k)}$, and hence $\{d(x^k, P_C(x^k))\}$ is bounded because $\{f(x^k)\}$ is non-increasing and f is bounded from below. Let $\bar{x} \in C$ be fixed. Since C is compact, then $\{d(P_C(x^k), \bar{x})\}$ is also bounded. Thus, using the triangular inequality, we obtain that $\{d(x^k, \bar{x})\}$ is bounded, which means that $\{x^k\}$ is bounded. Let x^* be a cluster point of $\{x^k\}$. Therefore, the first assertion follows from [50, Theorem 3.6 (i)] combined with [50, Remark 3.2 (b)].

Now, we will prove the rates of convergence of the sequences $\{f(x^k)\}$ and $\{x^k\}$. (ii): From (19), we have that f satisfies the local error bound with residual function $\omega(s) = \dfrac{\sigma}{\alpha}\sqrt{s}$ which has moderate behavior. Then, Theorem 4.1 ensures that f satisfies the Kurdyka–Łojasiewicz at every point $x \in [0 < f < r]$ with desingularizing function $\varphi(s) = \dfrac{2\sigma}{\alpha}\sqrt{s}$. Let us denote by $\Delta_k = f(x^k) - f(x^*) \geq 0$. Assume that $\Delta_k > 0$, for all $k \in \mathbb{N}$, otherwise the method stops in a finite number of iterations. Since x^k converges to x^*, there exist $k_0 \in \mathbb{N}$ and a neighborhood $\mathcal{N}(x^*)$ such that $x^k \in \mathcal{N}(x^*) \cap [f(x^*) < f < f(x^*) + \eta] = \mathcal{N}(x^*) \cap [0 < f < \eta]$, for all $k \geq k_0$, where the Kurdyka–Łojasiewicz property holds with desingularizing function $\varphi(s) = \dfrac{2\sigma}{\alpha}\sqrt{s}$. Thus, from (7), we have

$$
\begin{aligned}
1 &\leq (\varphi')^2(\Delta_{k+1})\|\operatorname{grad} f(x^{k+1})\|^2 \leq (\varphi')^2(\Delta_{k+1})\frac{L}{t_k}d^2(x^{k+1}, x^k) \\
&\leq \frac{L}{t_*}(\varphi')^2(\Delta_{k+1})(\Delta_k - \Delta_{k+1}) = \frac{L\sigma^2}{t_*\alpha^2}\frac{(\Delta_k - \Delta_{k+1})}{\Delta_{k+1}},
\end{aligned}
$$

where the second inequality comes from the Lipchitz property of grad f together with $t_k||\text{grad } f(x^k)|| = d(x^{k+1}, x^k)$ and the third inequality follows from (12) combined with the fact that $\{t_k\}$ is lower bounded by $t_* > 0$. Therefore, we obtain $\Delta_{k+1} \leq q\Delta_k$, for all $k \in \mathbb{N}$, where $0 < q = \frac{L\sigma^2}{L\sigma^2 + t_*\alpha^2} < 1$. Applying last inequality $k-1$ times, we prove item (ii).

(iii): The proof of this assertion follows from (19) combined with the estimated rate of the item (ii) having in mind that $f(x^*) = 0$.   □

Remark 5.1 A more general version of the previous result can be obtained by Wang et al. [50, Theorem 3.6] applied to the function (10). Therefore, we obtain convergence of the gradient method for solving a convex feasibility problem in a Riemannian manifold without any assumption on its sectional curvature. On the other hand, to obtain the linear rate of convergence by using [50, Theorem 3.6 (ii)], we have to prove that f given by (10) satisfies the Kurdyka–Łojasiewicz property in a Riemannian manifold without any assumption on its sectional curvature. In Hadamard manifolds, this fact is easily obtained as an application of Theorem 4.1.

5.2 Alternating Projections on Hadamard Spaces

The alternating projection method goes back to von Neumann who considered the case of two closed subspaces. Recall that the projection onto a non-empty set C sends any point x to its nearest point in C denoted by $P_C(x)$. While developing modern operator theory, von Neumann proved that, for $n = 2$ and X a Hilbert space, given starting point $b^0 = x \in X$, both sequences $\{a^k\}$ and $\{b^k\}$ defined by $a^k = P_{C_1}(b^{k-1})$ and $b^k = P_{C_2}(a^k)$ converge to $P_{C_1 \cap C_2}(x)$ in norm; see [47, Theorem 13.7]. The key idea is that often it is easy to project onto the individual C_i and this fact should be helpful in finding the solution to the more complicated problem of projecting onto $C = C_1 \cap C_2$. In 1962, Halperin [31] extended von Neumann's algorithm to more than two subspaces by projecting cyclically (or periodicly).

In 1965, Amemiya and Ando [1] proved that the periodicity assumptions could be dropped, showing the convergence of the method for an arbitrary sequence of indexes, which they called a random selection. However, the convergence obtained by Amemiya and Ando is weak and not in the norm as in von Neumann and Halperin's works. In 1995, Sakai [44] generalized Halperin's results for quasi-periodic sequences. In particular, he obtains the strong convergence of Amemiya and Ando's result when the sequence is quasi-periodic. In 1982, Bruck [22] proved the strong convergence in Hilbert space of the alternating projection method with random projections when one of the sets C_i is compact, provided the projection in the compact set is used infinitely often in the iteration. Bauschke [7] proved strong convergence of the method with random projections provided the sets C_j satisfy an assumption called *innately boundedly regularity*, more precisely, if every closed and convex set C_j satisfies $\max\{d(x^k, C_j) : j \in \{1, \ldots, n\}\} \to 0$ implies $d(x^k, C_1 \cap \ldots \cap C_n) \to 0$, where $\{x^k\}$ is the sequence generated by the alternating projection method.

In 2008, Lewis and Malick [38] considered the alternating projection method for two smooth submanifolds C_1 and C_2 of $\mathbb{R}^n$ which intersect transversally at a point $x^* \in C_1 \cap C_2$, i.e., $T_x C_1 + T_x C_2 = \mathbb{R}^n$, where $T_x C$ stands to the tangent space to C

at x. They proved that if the initial point x^0 is sufficiently close to x^*, then the method converges to a point in $C_1 \cap C_2$ linearly. In particular, if C_1 and C_2 are closed and convex sets of $\mathbb{R}^n$ such that their boundaries bd C_1 and bd C_2 are smooth manifolds and are transversal at $x^* \in C_1 \cap C_2$, then linear convergence of the alternating projection method is obtained. In 2012, Bačák et al. [6] proved that the underlying linear structure of the space is dispensable and von Neumann's algorithm converges in CAT(0) spaces. Under the boundedly regular assumption, they obtained strong convergence of the method.

In this section, we obtain some results on the alternating projection method in CAT(0) spaces and, in particular, Hadamard manifolds, extending some existing ones. We will denote by P_j the projection onto C_j. Let $\{j_k\}$ be a sequence taking values in $\{1, 2, \ldots, n\}$ and x^0 a starting point. Consider the sequence $\{x^k\}$ generated by the alternating projection method, i.e.,

$$x^k = P_{j_k}(x^{k-1}), \tag{20}$$

for all $k \in \mathbb{N}$. Note that $\{j_k\}$ is the sequence that establishes the order of projection onto the sets C_j, $j = 1, \ldots, n$. Such an order can be chosen cyclically or randomly. We say that the sequence $\{j_k\}$ is cyclic if $j_k = (k \bmod n) + 1$, for all $k \in \mathbb{N}$. From now on, we denote by $\{x^k\}$ the sequence generated by the alternating projection method.

In the next results, we show that the alternating projection method converges for any order of projection (i.e., with a random choice of the sequence of projections $\{j_k\}$) on Hadamard manifolds under the Slater condition. We also obtain weak convergence of the method in CAT(0) spaces, replacing the Slater condition by the cyclic projection. Strong convergence is also proved, additionally assuming that at least one of the sets is compact. In particular, we show that the alternating projection method converges to a solution of the CFP on Hadamard manifolds, just assuming cyclic projection.

Theorem 5.2 *Let M be a Hadamard manifold and assume that the intersection $C_1 \cap \cdots \cap C_n$ has an interior point. Then,*

(i) *The sequence $\{x^k\}$ is convergent for any sequence $\{j_k\}$;*
(ii) *It converges to a point $x^* \in \cap C_j$, where $j \in \{1, 2, \ldots, n\}$ is infinitely often in $\{j_k\}$.*

Proof (i): Let a be an arbitrary point belonging to $C_1 \cap \cdots \cap C_m$. Then,

$$
\begin{aligned}
d(a, x^{k+1}) &= d(a, P_{j_{k+1}} \circ \cdots \circ P_{j_1}(x^0)) \\
&= d(P_{j_{k+1}}(a), P_{j_{k+1}} \circ \cdots \circ P_{j_1}(x^0)), \ \ \forall k \in \mathbb{N}.
\end{aligned}
$$

Since the projection mapping is non-expansive (see [20, Proposition 2.4]), we have

$$d(a, x^{k+1}) \leq d(a, P_{j_k} \circ \cdots \circ P_{j_1}(x^0)) = d(a, x^k),$$

which means that $\{x^k\}$ is Fejér convergent to $C_1 \cap \cdots \cap C_n$. Therefore, the sequence $\{d(a, x_k)\}$ converges. Denote by $0 \leq \epsilon = \lim_{n \to +\infty} d(a, x_k)$. If $\epsilon = 0$, then the assertion is proved. Now, suppose that $\epsilon > 0$. Since $\{x^k\}$ is Fejér convergent to $C_1 \cap \cdots \cap C_n$, then it is bounded (see [3, Proposition 3.2.6]) and hence it has a cluster

point. Let y_1 and y_2 be cluster points of $\{x^k\}$. Note that $\{d(a, x_k)\}$ is convergent, so $d(a, y_1) = d(a, y_2)$ for every $a \in C_1 \cap \cdots \cap C_n$ and hence, without loss of generality, we may assume that a is an interior point of $C_1 \cap \cdots \cap C_n$.

Let γ be the unique geodesic such that $\gamma(0) = y_1$ and $\gamma(\epsilon) = a$ and take $0 < \delta < \epsilon$ such that $\gamma(\epsilon + \delta)$ belongs to the interior of $C_1 \cap \cdots \cap C_n$. Thus, $d(y_2, \gamma(\epsilon + \delta)) = d(y_1, \gamma(\epsilon + \delta))$. On the other hand,

$$d(y_1, \gamma(\epsilon + \delta)) = d(y_1, a) + d(a, \gamma(\epsilon + \delta)) = d(y_2, a) + d(a, \gamma(\epsilon + \delta)).$$

This implies that $d(y_2, \gamma(\epsilon + \delta)) = d(y_2, a) + d(a, \gamma(\epsilon + \delta))$ which means that the triangular inequality holds with an equality and hence y_2 belongs to the geodesic γ. Therefore, either $y_1 = y_2$ or $y_2 = \gamma(2\epsilon)$. Suppose that $y_2 = \gamma(2\epsilon)$ and let $z = \gamma(\epsilon + \frac{\delta}{2})$ be a point in $C_1 \cap \cdots \cap C_n$. Using the fact that $d(y_2, a) = d(y_1, a) = \epsilon$, we have $d(y_1, z) = \epsilon + \frac{\delta}{2}$ and $d(z, y_2) = d(z, \gamma(2\epsilon)) = \epsilon - \frac{\delta}{2}$, which implies that $\delta = 0$. This is a contradiction, and hence, we conclude that $y_1 = y_2$. In other words, $\{x^k\}$ has a unique cluster point, so it converges.

(ii): For any $j \in \{1, 2, \ldots, n\}$ which appears infinitely in the alternating projection method, we have a subsequence $\{x^k\}$ belonging to C_j. Let us denote by $\lim_{k \to +\infty} x^k = \lim_{k \to +\infty} P_{j_k} \circ \cdots \circ P_{j_1}(x^0) = x^*$. Since C_j is closed, we have that $x^* \in C_j$, for each $j \in \{1, 2, \ldots, n\}$ which is infinitely often, and hence $x^* \in \cap C_j$. □

Remark 5.1 Actually, the above result remains valid for a countable collection of closed convex sets in a Hadamard manifold. The interplay between the Slater condition and the concept of regularity was studied in the linear setting by Beck and Teboulle [10, Lemma 3.1] for 2 sets and Bolte et al. [18, Proposition 11] for n sets. However, the relation between these concepts in the Riemannian setting is not known.

In order to study the convergence of the method without the Slater condition, we will consider other assumptions, such as cyclic projections or compactness. We mention that Bačák [4, Theorems 3.4 and 3.7] proved the strong convergence of the proximal point method for a function that is written as the sum of convex functions and is Lipschitz continuous throughout the sequence generated by the method in CAT(0) spaces. If this function is taken as the sum of the indicator functions of each C_i, $i = 1, \ldots, n$, then we have strong convergence of the alternating projection method for solving convex feasibility problems. Next, we prove the weak and strong convergence of the alternating projection method for solving convex feasibility problems. Recall that for a given bounded sequence $\{x^k\}$ and a point $x \in H$, the asymptotic radius of $\{x^k\}$ at x is given by $r(x^k, x) = \limsup_{k \to \infty} d(x^k, x)$ and the asymptotic radius of $\{x^k\}$ as $r(x^k) = \inf_{x \in H} r(x^k, x)$. Moreover, we say that a point $x \in H$ is the asymptotic center of $\{x^k\}$ if $r(x^k, x) = r(x^k)$.

Theorem 5.3 *Let X be a complete $CAT(0)$ space and $C_1, \ldots, C_n \subset X$ closed convex sets of X whose intersection is non-empty. Assume that the sequence $\{j_k\}$ is cyclic. Then,*

(i) *The sequence $\{x^k\}$ weakly converges to a point in $C_1 \cap \cdots \cap C_n$;*
(ii) *Moreover, if there exists $j \in \{1, 2, \ldots, n\}$ such that C_j is compact, then $\{x^k\}$ strongly converges.*

Proof (i) Take $a \in C_1 \cap \cdots \cap C_n$. From [3, Theorem 2.1.12], one has

$$d^2(x^{k+1}, a) + d^2(x^k, x^{k+1}) \leq d^2(x^k, a), \quad \forall k \in \mathbb{N}.$$

In particular, for all $k \in \mathbb{N}$, we have

$$d(x^{k+1}, a) \leq d(x^k, a) \quad \text{and} \quad d^2(x^k, x^{k+1}) \leq d^2(x^k, a) - d^2(x^{k+1}, a) \quad (21)$$

and hence, the sequence $\{x^k\}$ is Fejér convergent to the set $C_1 \cap \cdots \cap C_n$ and

$$\lim_{k \to +\infty} d(x^k, x^{k+1}) = 0. \quad (22)$$

Thus, from [3, Proposition 3.2.6], we have that $\{x^k\}$ is bounded, and hence, it has a weak cluster point x^*; see [3, Proposition 3.1.2]. Now consider a subsequence $\{x^{k_m}\}$ such that $x^{k_m} \xrightarrow{w} x^*$ and without loss of generality we may assume that $x^{k_m} \in C_1$ for every $m \in \mathbb{N}$, otherwise we can take another subsequence having in mind that $\{j_k\}$ is cyclic. Thus, we obtain $x^* \in C_1$.

Let p be the period of the sequence $\{j_k\}$. Given $1 < j \leq n$ by definition of $\{j_k\}$, for each $m \in \mathbb{N}$, there exists $r_m \in \mathbb{N}$ with $1 \leq r_m \leq p$ such that $x^{k_m + r_m} \in C_j$. We claim that $x^{k_m + r_m} \xrightarrow{w} x^* \in C_j$, for each $1 < j \leq n$. In fact, to simplify the notation, consider $y^m = x^{k_m}$ and $z^m = x^{k_m + r_m}$ for every $m \in \mathbb{N}$. Given a subsequence $\{z^{m_s}\}$ and $x \in X$, since $r_m \leq p$ for every $m \in \mathbb{N}$, we have

$$d(y^{m_s}, x) - \sum_{j=k_{m_s}}^{k_{m_s} + r_{m_s} - 1} d(x^j, x^{j+1})$$

$$\leq d(z^{m_s}, x)$$

$$\leq \sum_{j=k_{m_s}}^{k_{m_s} + r_{m_s} - 1} d(x^j, x^{j+1}) + d(y^{m_s}, x).$$

From (22), we obtain $r(z^{m_s}, x) = \limsup_{m_s \to \infty} d(z^{m_s}, x) = \limsup_{m_s \to \infty} d(y^{m_s}, x) = r(y^{m_s}, x)$. Therefore, $r(z^{m_s}) = \inf_{x \in X} r(z^{m_s}, x) = \inf_{x \in X} r(y^{m_s}, x) = r(y^{m_s}, z) = r(z^{m_s}, z)$. This proves that $x^{k_m + r_m} \xrightarrow{w} x^*$. Since $x^{k_m + r_m} \in C_j$ for every $m \in \mathbb{N}$ and C_j is closed, it follows that $x^* \in C_j$. This means that $x^* \in C_1 \cap \cdots \cap C_n$ and from [3, Proposition 3.2.6], we obtain that $x^k \xrightarrow{w} x^*$. This proves the first assertion.

(ii) From item (i), we have that $\{x^k\}$ has a unique asymptotic center $x^* \in C_1 \cap \cdots \cap C_n$ which is the weak limit of $\{x^k\}$. From (21), we have that $\{d(x^k, x^*)\}$ converges. Without loss of generality, we may assume that C_n is compact. Since $\{j_k\}$ is cyclic with period p, take $1 \leq r_n \leq p$ such that $\{x^{nk+r_n}\}$ is a subsequence of $\{x^k\}$ which by definition of the projection method belongs to C_n. Since C_n is compact, then $\{x^{nk}\}$ has a subsequence strongly converging to a point $\bar{x}$. By the uniqueness of the asymptotic

center, we have that $\bar{x} = x^*$. Since $\{d(x^k, x^*)\}$ converges, we obtain that $\{x^k\}$ strongly converges to x^*. □

Remark 5.2 Note that item (i) of Theorem 5.3 is an extension from 2 to n closed and convex sets of Bačák et al. [6, Theorem 4.1 (i)]. It follows from the definition that compactness of one set implies boundedly regularity. Thus, we also can see Theorem 5.3 item (ii) as an extension from 2 to n closed and convex sets of [6, Theorem 4.1 (ii)]. On the other hand, compactness does not imply boundedly linearly regularity even in the linear setting; see Bauschke and Borwein [9, Example 3.16]. The boundedly linearly regularity was used in [6, Theorem 4.1 (iii) and (iv)] in order to obtain the rate of convergence of the method.

In a finite-dimensional Hadamard manifold, the compactness hypothesis is unnecessary due to the equivalence between the concepts of weak and strong convergence in this setting. Therefore, we have the following result.

Corollary 5.1 *Let M be a Hadamard manifold. If the intersection $C_1 \cap \cdots \cap C_n$ is non-empty, then the sequence $\{x^k\}$ converges for any cyclic sequence $\{j_k\}$.*

6 Conclusion

In this paper, we have studied the relationship between Kurdyka–Łojasiewicz property and the concept of error bounds on Hadamard manifolds. To this end, we have extended to the Riemannian context some existence results and properties of the differential inclusion problem. As an application, we have applied the gradient method for solving convex feasibility problems on Hadamard manifolds using an approach based on the square distance from a point to a set. We have shown that this function satisfies the local error bound, which implies the Kurdyka–Łojasiewicz inequality to obtain convergence rates of the gradient method under the boundedly linearly regular assumptions of the sets. In the linear setting, the Slater condition is a sufficient condition to obtain linearly regularity. However, in the Riemannian setting, this relationship is not known. Therefore, we have shown the convergence of the alternating projection method for solving convex feasibility problems on Hadamard manifolds, and more generally, CAT(0) spaces, under the Slate condition for finitely many sets with cyclic and random order of projection. While the computational efficiency of the methods and comparisons with other methods have not been discussed, this work provides a solid theoretical foundation for further progress on this topic in the near future.

References

1. Amemiya, I., Andô, T.: Convergence of random products of contractions in Hilbert space. Acta Sci. Math. **26**, 239–244 (1965)

2. Attouch, H., Bolte, J., Svaiter, B.F.: Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting and regularized Gauss-Seidel methods. Math. Program. **137**, 91–129 (2013)
3. Bačák, M.: Convex Analysis and Optimization in Hadamard Spaces. De Gruyter Series in Nonlinear Analysis and Applications, vol. 22. Walter de Gruyter, Berlin (2014)
4. Bačák, M.: Computing medians and means in Hadamard spaces. SIAM J. Optim. **24**, 1542–1566 (2014)
5. Barani, A., Hosseini, S.: Characterization of solution sets of convex optimization problems in Riemannian manifolds. Arch. Math. **114**, 215–225 (2020)
6. Bačák, M., Searston, I., Sims, B.: Alternating projections in CAT(0) spaces. J. Math. Anal. Appl. **385**, 599–607 (2012)
7. Bauschke, H.H.: A norm convergence result on random products of relaxed projections in Hilbert space. Trans. Am. Math. Soc. **347**, 1365–1373 (1995)
8. Bauschke, H.H., Borwein, J.M.: On projection algorithms for solving convex feasibility problems. SIAM Rev. **38**, 367–426 (1996)
9. Bauschke, H.H., Borwein, J.M.: On the convergence of von Neumann's alternating projection algorithm for two sets. Set-Valued Anal. **1**, 185–212 (1993)
10. Beck, A., Teboulle, M.: Convergence rate analysis and error bounds for projection algorithms in convex feasibility problem. Optim. Methods Softw. **18**, 377–394 (2003)
11. Bento, G.C., Bitar, S.D.B., Cruz Neto, J.X., Oliveira, P.R., Souza, J.C.O.: Computing Riemannian center of mass on Hadamard manifolds. J. Optim. Theory Appl. **183**, 977–992 (2019)
12. Bento, G.C., Cruz Neto, J.X., Oliveira, P.R.: A New approach to the proximal point method: convergence on general Riemannian manifolds. J. Optim. Theory Appl. **168**, 743–755 (2016)
13. Bento, G.C., Melo, J.G.: Subgradient method for convex feasibility on Riemannian manifolds. J. Optim. Theory Appl. **152**, 773–785 (2012)
14. Bergmann, R., Persch Steidl, G.: A parallel Douglas–Rachford algorithm for minimizing ROF-like functionals on images with values in symmetric Hadamard manifolds. SIAM J. Imaging Sci. **9**(3), 901–937 (2016)
15. Bolte, J., Daniilidis, A., Lewis, A.: The Łojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems. SIAM J. Optim. **17**(4), 1205–1223 (2007)
16. Bolte, J., Daniilidis, A., Lewis, A., Shiota, M.: Clarke subgradients of stratifiable functions. SIAM J. Optim. **18**(2), 556–572 (2007)
17. Bolte, J., Daniilidis, A., Ley, O., Mazet, L.: Characterizations of Łojasiewicz inequalities: subgradient flows, talweg, convexity. Trans. Am. Math. Soc. **362**, 3319–3363 (2010)
18. Bolte, J., Nguyen, T.P., Peypouquet, J., Suter, B.W.: From error bounds to the complexity of first-order descent methods for convex functions. Math. Program. **165**, 471–507 (2017)
19. Brézis, H.: Opérateurs Maximaux Monotones et Semi-groupes de Contractions Dans les Espace Hilbert. North-Holland Mathematics Studies, vol. 5. North-Holland, Amsterdam (1973)
20. Bridson, M., Haefliger, A.: Metric Spaces of Nonpositive Curvature. Springer, Berlin (1999)
21. Bruck, R.E.: Asymptotic convergence of nonlinear contraction semigroups in Hilbert space. J. Funct. Anal. **18**, 15–26 (1975)
22. Bruck, R.E.: Random products of contractions in metric and Banach Spaces. J. Math. Anal. Appl. **88**, 319–322 (1982)
23. Burtscher, A.Y.: Length structures on manifolds with continuous Riemannian metrics. New York J. Math. **21**, 273–296 (2015)
24. Cruz Neto, J.X., Ferreira, O.P., Lucâmbio Pérez, L.R.: Contribution to the study of monotone vector fields. Acta Math. Hung. **94**, 307–320 (2002)
25. Cruz Neto, J.X., Ferreira, O.P., Lucâmbio Pérez, L.R., Németh, S.Z.: Convex- and monotone-transformable mathematical programming problems and a proximal-like point method. J. Global Optim. **35**(1), 53–69 (2006)
26. Cruz Neto, J.X., Oliveira, P.R., Soares, P.A., Jr., Soubeyran, A.: A Learning how to play Nash, potential games and alternating minimization method for structured non-convex problems on Riemannian manifolds. J. Convex Anal. **20**, 395–438 (2013)
27. Cruz Neto, J.X., Lima, L.L., Oliveira, P.R.: Geodesic algorithms in Riemannian geometry. Balkan J. Geom. Appl. **3**, 89–100 (1998)

28. Cruz Neto, J.X., Melo, I.D., Sousa, P.A., Silva, J.P.: A note on the paper "Proximal point methods for quasiconvex and convex functions with Bregman distances on Hadamard manifolds". J. Convex Anal. **24**, 679–684 (2017)
29. Ferreira, O.P., Louzeiro, M.S., Prudente, L.F.: Gradient method for optimization on Riemannian manifolds with lower bounded curvature. SIAM J. Optim. **29**(4), 2517–2541 (2019)
30. Gubin, L.C., Polyak, B.T., Raik, E.V.: The method of projections for finding the common point of convex sets. USSR Comput. Math. Math. Phys. **7**, 1–24 (1967)
31. Halperin, I.: The product of projection operators. Acta Sci. Math. **23**, 96–99 (1962)
32. Hoffman, A.J.: On approximate solutions of systems of linear inequalities. J. Res. Natl. Bur. Stand. **49**(4), 263–265 (1952)
33. Hosseini, S., Uschmajew, A.: A Riemannian gradient sampling algorithm for nonsmooth optimization on manifolds. SIAM J. Optim. **27**(1), 173–189 (2017)
34. Iusem, A.N., Svaiter, B.F., Cruz Neto, J.X.: Central paths, generalized proximal point methods and Cauchy trajectories in Riemannian manifolds. SIAM J. Control. Optim. **37**, 566–588 (1999)
35. Jost, J.: Equilibrium maps between metric spaces. Calc. Var. Partial Differ. Equ. **2**(2), 173–204 (1994)
36. Kristály, A., Li, C., López-Acedo, G., Nicolae, A.: What do "convexity" imply on Hadamard manifolds? J. Optim. Theory Appl. **170**, 1068–1074 (2016)
37. Kurdyka, K.: On gradients of functions definable in o-minimal structures. Ann. Inst. Fourier **48**, 769–783 (1998)
38. Lewis, A.S., Malick, J.: Alternating projection on manifolds. Math. Oper. Res. **33**(1), 216–234 (2008)
39. Łojasiewicz, S.: Une propriété topologique des sous-ensembles analytiques réels, pp. 87–89. Les Équations aux Dérivées Partielles, Éditions du centre National de la Recherche Scientifique (1963)
40. Luenberger, D.G.: The gradient projection method along geodesics. Manag. Sci. **18**, 620–631 (1972)
41. Luke, D.R., Sabach, S., Teboulle, M.: Optimization on spheres: models and proximal algorithms with computational performance comparisons. SIAM J. Math. Data Sci. **1**, 408–445 (2019)
42. Palis, J., De Melo, W.: Geometric Theory of Dynamical Systems. An Introduction. Translated from the Portuguese by A.K. Manning,. Springer, Berlin (1982)
43. Poly, J.B., Raby, G.: Fonction distance et singularités. Bull. Sci. Math. **108**, 187–195 (1984)
44. Sakai, M.: Strong convergence of infinite products of orthogonal projections in Hilbert space. Appl. Anal. **59**, 109–120 (1995)
45. Sakai, T.: Riemannian Geometry. Translations of Mathematical Monographs, vol. 149. American Mathematical Society, Providence (1996)
46. Udriste, C.: Convex Functions and Optimization Algorithms on Riemannian Manifolds. Mathematics and Its Applications, vol. 297. Kluwer Academic, Dordrecht (1994)
47. von Neumann, J.: Functional Operators. II, The Geometry of Orthogonal Spaces. Annals of Mathematics Studies, vol. 22. Princeton University Press, Princeton (1950)
48. Wang, X.M., Li, C., Yao, J.C.: Subgradient projection algorithms for convex feasibility on Riemannian manifolds with lower bounded curvatures. J. Optim. Theory Appl. **164**, 202–217 (2015)
49. Wang, X.M., Li, C., Wang, J., Yao, J.C.: Linear convergence of subgradient algorithm for convex feasibility on Riemannian manifolds. SIAM J. Optim. **25**, 2334–2358 (2015)
50. Wang, J., Wang, X.M., Li, C., Wang, J., Yao, J.C.: Convergence analysis of gradient algorithms on Riemannian manifolds without curvature constraints and application to Riemannian mass. SIAM J. Optim. **31**, 172–199 (2021)
51. Wang, X.M., Li, C., Yao, J.C.: On some basic results related to affine functions on Riemannian manifolds. J. Optim. Theory Appl. **170**, 783–803 (2016)
52. Wolter, F.E.: Distance function and cut loci on a complete Riemannian manifold. Arch. Math. **32**, 92–96 (1979)