

**ON THE RATE OF  
ASYMPTOTIC REGULARITY OF  
ITERATIVE METHODS  
FOR NONEXPANSIVE MAPPINGS IN  $CAT(0)$  SPACES  
AND HYPERBOLIC OPTIMIZATION**

Vittorio Colao and Katherine Rossella Foglia

*Communicated by Marek Galewski*

**Abstract.** The Krasnosel’skiĭ–Mann and Halpern iterations are classical schemes for approximating fixed points of nonexpansive mappings in Banach spaces, and have been widely studied in more general frameworks such as  $CAT(\kappa)$  and, more generally, geodesic spaces. Convergence results and convergence rate estimates in these nonlinear settings are already well established. The contribution of this paper is twofold: first, we extend to complete  $CAT(0)$  spaces proof techniques originally developed in the linear setting of Banach and Hilbert spaces, thereby recovering the same asymptotic regularity bounds; second, we introduce a Halpern-type optimizer for hyperbolic optimization as a nonlinear counterpart of the Euclidean HalpernSGD scheme.

**Keywords:** geodesic spaces, Hadamard spaces, metric fixed point theory, optimization, hyperbolic deep learning, Halpern iterates, Krasnosel’skiĭ–Mann iterates.

**Mathematics Subject Classification:** 47H10, 46N10, 53C25, 58C30.

## 1. INTRODUCTION

Deep representation learning in hyperbolic spaces has recently gained significant attention due to the ability of hyperbolic geometry to accommodate tree-like structures and more generally, to effectively model hierarchically structured data [31]. In [19], the authors showed that core neural network operations, such as multinomial logistic regression, feed-forward networks, and recurrent neural networks, can be reformulated in hyperbolic spaces. These developments enable the direct embedding of sequential data and the training of classifiers in hyperbolic spaces. As remarked in [19], until recently it was common practice to embed data in the Euclidean space  $\mathbb{R}^n$  due to the availability of closed-form formulas for distance and inner product, but this is not sufficient. In fact, many datasets reside on manifolds or graphs with inherently non-Euclidean geometry [9]. For instance, no matter how high the dimension, Euclidean space cannot embed arbitrary tree structures with negligibly low distortion, while

a two-dimensional hyperbolic space easily can [37]. Therefore, the so-called hyperbolic neural architectures can potentially lead to drastically more compact models than their Euclidean counterparts [31]. Training these models, however, requires optimizers adapted to the geometric setting, and one of the most widely used is Riemannian stochastic gradient descent (RSGD) [5].

As observed in [12, 18], Gradient Descent (GD) can be seen as a special case of the Krasnosel’skiĭ–Mann (KM) iteration. The same works also point out the existence of alternative schemes, such as the Halpern and Viscosity iterations, which achieve faster convergence rates than KM and GD. These methods have been studied in both linear and nonlinear settings, and can be used to design faster GD-type optimizers for hyperbolic deep neural networks, as already done in the linear case for HalpernSGD [18].

In particular, the rates of asymptotic regularity for both the Krasnosel’skiĭ–Mann and Halpern iteration have been established also in the nonlinear setting. Our contribution in this work is to extend to specific geodesic spaces the proof techniques originally developed in the linear setting, thereby recovering the same rates of convergence. It is worth remarking that part of the results we present in this paper have already been investigated in the literature, although with different approaches and proofs. The aim of this paper is therefore to develop a theoretical understanding of nonlinear optimization and to set the stage for a Hyperbolic version of HalpernSGD.

The paper is organized as follows. Section 2 provides the mathematical background needed for our analysis. Section 3 reviews existing convergence results for KM and Halpern-type iterations in different geometric settings. In Section 4, we present and prove our main results on convergence rates in specific geodesic spaces, including hyperbolic ones. Section 5 provides the theoretical construction of a hyperbolic HalpernGD-type optimizer.

Finally, Section 6 outlines future work and contains our concluding remarks.

## 2. PRELIMINARIES

In this section, we introduce the necessary background on geodesic spaces following [8].

**Definition 2.1.** A metric space  $(X, d)$  is called a geodesic space if for every pair of distinct points  $x, y \in X$ , there exist a real number  $l > 0$  and a continuous map  $\gamma : [0, l] \rightarrow X$  such that

$$\gamma(0) = x, \quad \gamma(l) = y,$$

and

$$d(\gamma(t_1), \gamma(t_2)) = |t_1 - t_2| \quad \text{for all } t_1, t_2 \in [0, l].$$

Such a map  $\gamma$  is called a geodesic path, or more briefly a geodesic, joining  $x$  and  $y$  and, in other words, it is an isometry between the real interval  $[0, l]$  and its image  $\gamma([0, l])$ , which in turn is called a geodesic segment with endpoints  $x$  and  $y$ . In particular, taking  $t_1 = 0$  and  $t_2 = l$ , one obtains  $l = d(x, y)$ . Therefore, by a linear reparametrization, geodesics may equivalently be defined on the interval  $[0, 1]$ . If  $\gamma$  is defined on  $[0, +\infty)$  it is called a geodesic ray; if it is defined on the whole  $\mathbb{R}$  it is called a geodesic line; finally, if it is defined on an interval  $I \subset \mathbb{R}$ , not necessarily closed, and

for every  $t \in I$  there exists  $\varepsilon > 0$  such that the restriction of  $\gamma$  to  $I \cap (t - \frac{\varepsilon}{2}, t + \frac{\varepsilon}{2})$  is a geodesic it is called a local geodesic.

In particular, for a real constant  $r > 0$ ,  $(X, d)$  is said to be an  $r$ -geodesic space if every pair of points  $x, y \in X$  with  $d(x, y) < r$  can be joined by a geodesic. If  $r = \infty$ , an  $\infty$ -geodesic space corresponds to a geodesic space as in Definition 2.1. Important examples of geodesic spaces include:

**Example 2.2.** For  $n \in \mathbb{N}$  and  $\kappa \in \mathbb{R}$ , the model space  $M_\kappa^n$  is defined by

$$M_\kappa^n = \begin{cases} \mathbb{H}^n & \text{if } \kappa < 0, \\ \mathbb{E}^n & \text{if } \kappa = 0, \\ \mathbb{S}^n & \text{if } \kappa > 0, \end{cases}$$

endowed with the metric

$$d_{M_\kappa^n} = \begin{cases} \frac{1}{\sqrt{-\kappa}} d_H & \text{if } \kappa < 0, \\ d_E & \text{if } \kappa = 0, \\ \frac{1}{\sqrt{\kappa}} d_S & \text{if } \kappa > 0, \end{cases}$$

where:

- $\mathbb{E}^n$  is the Euclidean space  $\mathbb{R}^n$  with the usual Euclidean distance  $d_E$ ,
- $\mathbb{S}^n \subset \mathbb{R}^{n+1}$  is the unit sphere with the angular metric

$$d_S(x, y) = \arccos(\langle x, y \rangle),$$

- $\mathbb{H}^n$  is the hyperbolic space  $\mathbb{H}^n := \{u \in \mathbb{R}^{n+1} \mid \langle u, u \rangle_L = -1, u_0 > 0\}$ , where

$$\langle u, v \rangle_L := -u_0 v_0 + \sum_{i=1}^n u_i v_i,$$

and the hyperbolic metric  $d_H$  is given by  $\cosh(d_H(x, y)) = -\langle x, y \rangle_L$ .

With this definition,  $(M_\kappa^n, d_{M_\kappa^n})$  is a geodesic space (see [8, p. 23, Definition 2.10]) and, according to the sign of  $\kappa$ , the geodesics are Euclidean segments, arcs of great circles, or hyperbolic geodesics.

We say that  $(X, d)$  is a uniquely geodesic space if for every pair of points  $x, y \in X$ , there exists exactly one geodesic joining them. In such spaces, the geodesic segment with endpoints  $x$  and  $y$  is denoted by  $[x, y]$ . An important family of examples of uniquely geodesic spaces include (see [8]) the following:

**Example 2.3.** The model spaces  $M_\kappa^n$  provide basic examples in the theory of geodesic spaces. As stated in [8, Proposition 2.11],  $M_\kappa^n$  is uniquely geodesic for  $\kappa \leq 0$ . On the other hand, when  $\kappa > 0$ , that is, in the spherical case,  $M_\kappa^n$  is not uniquely geodesic

in general. Indeed, antipodal points on the sphere can be joined by infinitely many geodesics, namely all the semicircumferences having those points as endpoints. However, if the distance between two points is less than  $\frac{\pi}{\sqrt{\kappa}}$ , then the geodesic joining them is unique.

Let  $(X, d)$  be a geodesic space and  $x, y \in X$ , we say that a point  $z_t \in X$  belongs to a geodesic segment with endpoints  $x$  and  $y$  if there exist  $t \in [0, 1]$  such that:

$$d(z_t, x) = t d(x, y) \quad \text{and} \quad d(z_t, y) = (1 - t) d(x, y). \quad (2.1)$$

Fixing a specific geodesic segment with endpoints  $x$  and  $y$ , we use the notation

$$z_t = (1 - t)x \oplus ty. \quad (2.2)$$

In particular, for  $t = \frac{1}{2}$  the point  $z_t$  is called the midpoint of  $x$  and  $y$ , and is denoted by  $m(x, y)$ .

**Remark 2.4.** If  $(X, d)$  is uniquely geodesic, then there exists a unique geodesic segment with endpoints  $x$  and  $y$ . In this case, there is no ambiguity in writing  $z_t \in [x, y]$ , and it is not necessary to specify a particular geodesic segment in order to guarantee the uniqueness of  $z_t$ .

A further important class of geodesic spaces is given by  $\text{CAT}(\kappa)$  spaces, introduced to generalize curvature conditions to metric spaces. In order to formally introduce them we provide some important definitions (see [8, p. 158]):

**Definition 2.5.** In a geodesic space  $(X, d)$  a geodesic triangle  $\Delta([p, q], [q, r], [r, p])$  is characterized by three distinct points  $p, q, r \in X$ , called the vertices, and by the choices of three geodesic joining them, whose geodesic segments we denote by  $[p, q]$ ,  $[q, r]$ ,  $[r, p]$  and called the sides of the triangle.

Even if the space is not uniquely geodesic, we can fix a single geodesic for each pair of vertices, so that the notation  $[x, y]$  remains unambiguous. In the uniquely geodesic case, this choice is unnecessary, and the triangle may be written simply as  $\Delta(p, q, r)$ . We say that a point  $x$  belongs to  $\Delta([p, q], [q, r], [r, p])$  if it lies on one of its sides and the perimeter is then the sum of the three side lengths, as the notion of distance remains well defined.

**Definition 2.6.** Let  $(X, d)$  be a uniquely geodesic space, for a triangle  $\Delta(p, q, r)$  in  $(X, d)$ , a comparison triangle  $\bar{\Delta}(\bar{p}, \bar{q}, \bar{r})$  in the model space  $M_\kappa^2$  is a triangle such that:  $d_{M_\kappa^2}(\bar{p}, \bar{q}) = d(p, q)$ ,  $d_{M_\kappa^2}(\bar{q}, \bar{r}) = d(q, r)$  and  $d_{M_\kappa^2}(\bar{r}, \bar{p}) = d(r, p)$ .

Similarly, for  $x \in \Delta$ , a comparison point  $\bar{x} \in \bar{\Delta}$  is a point lying on the side of  $\bar{\Delta}$  corresponding to the side of  $\Delta$  to which  $x$  belongs, and has the same distance from the corresponding vertices as  $x$  has from its own. Let  $D_\kappa$  be the following quantity:

$$D_\kappa := \begin{cases} \frac{\pi}{\sqrt{\kappa}} & \text{if } \kappa > 0, \\ \infty & \text{if } \kappa \leq 0. \end{cases}$$

In [8, p. 24, Lemma 2.14], it is proved that in a uniquely geodesic space  $(X, d)$ , for any geodesic triangle  $\Delta(p, q, r) \subset X$  whose perimeter is strictly less than  $2D_\kappa$ , a comparison triangle exists, and it is unique up to isometry of  $M_\kappa^2$ .

Having properly introduced the notions of geodesic triangle and comparison triangle, we can now provide the definition of the  $CAT(\kappa)$  space as in [8, p. 158, Definition 1.1]:

**Definition 2.7.** Let  $(X, d)$  be a geodesic space, let  $\kappa$  be a fixed real number, and let  $\Delta$  be a geodesic triangle in  $X$  with perimeter strictly less than  $2D_\kappa$ . We say that  $\Delta$  satisfies the  $CAT(\kappa)$  inequality if, for all  $x, y \in \Delta$  and the respective comparison points  $\bar{x}, \bar{y} \in \bar{\Delta}$ , it holds that

$$d(x, y) \leq d_{M_\kappa^2}(\bar{x}, \bar{y}). \quad (2.3)$$

$X$  is called a  $CAT(\kappa)$  space if it is  $D_\kappa$ -geodesic and each triangle  $\Delta$  in  $X$  with perimeter strictly less than  $2D_\kappa$  satisfies (2.3).

Note that if  $\kappa \leq 0$ , the inequality (2.3) must hold for every geodesic triangle in  $X$ . A first example of a  $CAT(\kappa)$  space is the model space  $M_\kappa^2$ , while a typical example of a  $CAT(0)$  space is a pre-Hilbert space. An important characterization result is given in the following.

**Theorem 2.8** ([8, Theorem 1.12, p. 165]). *Every  $CAT(\kappa)$  space is also a  $CAT(\kappa')$  space for all  $\kappa' \geq \kappa$ . Moreover, if  $X$  is a  $CAT(\kappa')$  space for every  $\kappa' > \kappa$ , then  $X$  is a  $CAT(\kappa)$  space.*

The previous result implies that every  $CAT(\kappa)$  space with  $\kappa \leq 0$  is also a  $CAT(0)$  space; in particular, this happens for the hyperbolic space  $\mathbb{H}^n$ .

We remark that  $CAT(\kappa)$  spaces are, in general, not complete with respect to their given metric. A well-known complete  $CAT(\kappa)$  space is obtained when  $\kappa = 0$ . In fact, complete  $CAT(0)$  spaces are referred to as Hadamard spaces.

We can now deal with several notions of convexity in the setting of the geodesic spaces.

As in [8], we say that a subset  $C$  of a geodesic space  $X$  is convex if for all  $x, y \in C$ , the image of every geodesic joining them lies in  $C$ . If  $X$  is uniquely geodesic,  $C$  is said to be strongly convex if it is convex and there is no other local geodesic joining  $x$  and  $y$ . The converse implication does not hold in general and a counterexample can be obtained working on the torus endowed with the intrinsic metric. We recall that a metric space  $(X, d)$  satisfies the convexity condition of the metric if for all  $x, y, w \in X$  and for all  $t \in [0, 1]$  it holds:

$$d(w, (1-t)x \oplus ty) \leq (1-t)d(w, x) + td(w, y). \quad (2.4)$$

In particular, as stated in [8, Proposition 2.2, p. 176], if  $(X, d)$  is a  $CAT(0)$  space, the distance function  $d : X \times X \rightarrow \mathbb{R}$  is convex, which means that, given any pair of geodesics  $\gamma : [0, 1] \rightarrow X$  and  $\gamma' : [0, 1] \rightarrow X$ , the following inequality holds for all  $t \in [0, 1]$ :

$$d(\gamma(t), \gamma'(t)) \leq (1-t)d(\gamma(0), \gamma'(0)) + td(\gamma(1), \gamma'(1)). \quad (2.5)$$

For later use, we also recall the notion of convexity for functions on geodesic spaces.

**Definition 2.9.** Let  $(X, d)$  be a geodesic space, let  $C \subseteq X$  be a convex subset, and let  $f : C \rightarrow (-\infty, +\infty]$ . We say that  $f$  is geodesically convex if for every geodesic  $\gamma : [0, 1] \rightarrow C$  and for all  $t \in [0, 1]$  one has

$$f(\gamma(t)) \leq (1-t)f(\gamma(0)) + tf(\gamma(1)).$$

If  $X$  is uniquely geodesic, this is equivalent to  $f((1-t)x \oplus ty) \leq (1-t)f(x) + tf(y)$  for all  $x, y \in C$  and for all  $t \in [0, 1]$ .

Since the classical notions of convergence in normed spaces do not directly extend to this setting, an alternative notion, introduced by Lim in [29] and known as  $\Delta$ -convergence, plays a central role in the analysis.

**Definition 2.10.** Let  $\{x_n\}$  be a bounded sequence in a metric space  $(X, d)$ . First, for all  $x \in X$ , define the quantity

$$r(x, \{x_n\}) = \limsup_{n \rightarrow \infty} d(x, x_n).$$

Then, for  $\{x_n\}$  we can define the asymptotic radius  $r(\{x_n\})$  and the asymptotic center  $A(\{x_n\})$  respectively as follows:

$$\begin{aligned} r(\{x_n\}) &:= \inf_{x \in X} r(x, \{x_n\}), \\ A(\{x_n\}) &:= \{x \in X : r(x, \{x_n\}) = r(\{x_n\})\}. \end{aligned}$$

The sequence  $x_n$  is said to  $\Delta$ -converge to  $x$ , in symbol  $x_n \xrightarrow{\Delta} x$ , if for every subsequence  $\{x_{n_k}\}$  of  $\{x_n\}$ , the point  $x$  is the unique asymptotic center of  $\{x_{n_k}\}$ , that is equivalent to say that  $x$  is the only element to satisfy:

$$\limsup_{k \rightarrow \infty} d(x_{n_k}, x) \leq \limsup_{k \rightarrow \infty} d(x_{n_k}, y) \quad \text{for all } y \in X. \quad (2.6)$$

As noted in [29], the inequality (2.6) does not guarantee the uniqueness of the asymptotic center. However, it is known that if a sequence  $x_n$  in a complete  $CAT(\kappa)$  space has asymptotic radius strictly less than  $\frac{D_\kappa}{2}$ , then it admits a unique asymptotic center (see [15, Proposition 7] for  $\kappa = 0$  case and [17, Proposition 4.1] for  $\kappa > 0$  case). Clearly,  $\Delta$ -convergence is much weaker than strong convergence in a metric space. Moreover, it has been proved in [17] that  $\Delta$ -convergence coincides with weak convergence in Hilbert spaces.

Finally, we recall a metric condition fundamental for what follows.

**Definition 2.11.** Let  $(X, d)$  be a geodesic space and  $C \subset X$ . A mapping  $T : C \rightarrow C$  is said to be nonexpansive if  $d(Tx, Ty) \leq d(x, y)$  for all  $x, y \in C$ .

The study of nonexpansive mappings in  $CAT(\kappa)$  spaces was initiated by Kirk (see, e.g., [23, 24]), and an important class of nonexpansive operators in the linear setting is the so-called forward map, which basically coincides with the step of classical gradient descent algorithm as stated in the following result.

**Theorem 2.12** ([2, 4]). *Let  $H$  be a real Hilbert space, and let  $f : H \rightarrow \mathbb{R}$  be a convex and Fréchet-differentiable function with  $L$ -Lipschitz continuous gradient  $\nabla f$ . Fix  $\eta \in (0, 2/L)$  and define the forward operator  $F : H \rightarrow H$  by*

$$F = I - \eta \nabla f.$$

*Then  $F$  is averaged nonexpansive, that is,  $F = \lambda I + (1 - \lambda)T$  for some  $\lambda \in (0, 1)$  and some nonexpansive mapping  $T : H \rightarrow H$ . Moreover,*

$$\theta^* \in \arg \min_x f(x) \iff \theta^* = F(\theta^*) \iff \theta^* = T(\theta^*),$$

*so that the fixed point set  $\text{Fix}(T)$  coincides with the set of minimizers of  $f$ .*

The nonexpansivity of this operator played a crucial role in neural network hyperparameter optimization and, in particular, was essential for the development of the HalpernSGD optimizer (see [11, 12, 18]).

In the nonlinear setting, more precisely on Hadamard manifolds, the Euclidean nonexpansivity conclusion for the forward operator does not extend automatically. Section 5 provides an explicit counterexample on  $\mathbb{H}^2$ .

However, an important example of a nonexpansive mapping in Hadamard spaces is the proximal mapping (or resolvent), defined as follows.

**Definition 2.13.** Let  $(X, d)$  be a Hadamard space and let  $f : X \rightarrow (-\infty, +\infty]$  be a proper, lower semicontinuous, and convex function. For any  $\lambda > 0$ , the proximal mapping (or resolvent) of  $f$  (see [1, Ch. 2.2 and Ch. 4.2, Lemma 4.2.2]) is defined by

$$J_{\lambda f}(x) := \operatorname{argmin}_{y \in X} \left\{ f(y) + \frac{1}{2\lambda} d(x, y)^2 \right\}, \quad x \in X.$$

This operator can be used to develop an analogue of the HalpernGD optimizer in the hyperbolic setting, as discussed in Section 5.

In [25], the concept of  $\Delta$ -convergence was investigated for nonexpansive mappings in  $CAT(0)$  spaces, showing that such spaces provide a natural framework for this type of convergence. Moreover, by [23, Theorem 18, p. 207], every nonexpansive map defined on a nonempty, closed, bounded and convex subset of a complete  $CAT(0)$  space admits a fixed point. These results justify the choice of the setting for our work.

### 3. KRASNOSEL'SKIĬ-MANN AND HALPERN ITERATIVE METHODS

Iterative methods for approximating fixed points of mappings have a long and rich history, originating in the context of Banach spaces and progressively extending to more general nonlinear spaces such as  $CAT(0)$  spaces and geodesic spaces. Their study has been a flourishing area of research over the past decades, with numerous applications in nonlinear analysis, optimization, and partial differential equations [3, 10, 33, 34]. The Krasnosel'skiĭ-Mann (KM) and Halpern-type iterations stand out as the most classical and extensively investigated schemes. The study of their

convergence behavior has given rise to a vast literature. Beyond the linear setting, however, the intrinsic convex structure of these algorithms makes it more challenging to establish approximation results for fixed points. Consequently, substantial progress has been achieved only in certain specific classes of metric spaces, such as hyperbolic spaces, the Hilbert ball, and Hadamard manifolds.

Let  $X$  be a Banach space, let  $K \subseteq X$  be nonempty and convex, and let  $T : K \rightarrow K$  be a nonexpansive mapping. The Krasnosel'skiĭ–Mann iteration is defined by

$$x_{n+1} = (1 - \lambda_{n+1})x_n + \lambda_{n+1}Tx_n, \quad x_0 \in K, \quad (3.1)$$

where  $(\lambda_n)_{n \geq 1} \subset [0, 1)$ . Its convergence theory in Banach spaces is classical: asymptotic regularity and convergence depend on both the geometry of the space and the control sequence, while strong convergence can be recovered under additional compactness assumptions; see, e.g., [22, 30].

In a uniquely geodesic space, the natural nonlinear analogue is

$$x_{n+1} = (1 - \lambda_{n+1})x_n \oplus \lambda_{n+1}Tx_n, \quad x_0 \in K, \quad (3.2)$$

where  $\oplus$  denotes the convex combination along the unique geodesic joining  $x_n$  and  $Tx_n$ , meaning that  $x_{n+1}$  lies on the geodesic segment  $[x_n, Tx_n]$  (see notation in (2.2)). If  $X$  is a complete  $CAT(0)$  space and  $\text{Fix}(T) \neq \emptyset$ , then (3.2)  $\Delta$ -converges to a fixed point of  $T$  under standard assumptions on  $(\lambda_n)$ , for instance  $\sup_n \lambda_n < 1$  and  $\sum_n \lambda_n = +\infty$  [16]. More generally, analogous results hold in complete  $CAT(\kappa)$  spaces under the radius restriction  $d(x_0, \text{Fix}(T)) \leq \frac{D_\kappa}{4}$  and the condition  $\sum_n \lambda_n(1 - \lambda_n) = +\infty$ ; if the space is boundedly compact, the convergence is strong [21]. In the special case  $\kappa = 0$ , the radius restriction disappears.

We turn now our attention to a closely related scheme which, under suitable assumptions, ensures strong convergence. Let  $X$  be a Banach space, let  $K \subseteq X$  be nonempty, closed and convex, and let  $T : K \rightarrow K$  be nonexpansive with  $\text{Fix}(T) \neq \emptyset$ . For a sequence  $(\alpha_k)_{k \geq 1} \subset (0, 1]$ , the Halpern iteration is defined by

$$z_{n+1} = \alpha_{n+1}u + (1 - \alpha_{n+1})Tz_n, \quad z_0, u \in K. \quad (3.3)$$

Introduced by Halpern [20], this scheme is a classical strongly convergent alternative to the Krasnosel'skiĭ–Mann iteration. In Hilbert spaces, strong convergence is known under the standard conditions

$$(C1) \quad \lim_{n \rightarrow \infty} \alpha_n = 0,$$

$$(C2) \quad \sum_{n=0}^{\infty} \alpha_n = +\infty$$

together with a mild regularity assumption on the control sequence, for instance,

$$(C3) \quad \sum_{n=0}^{\infty} |\alpha_{n+1} - \alpha_n| < \infty$$

(see, e.g., [20, 40]). More generally, the viscosity iteration is given by

$$y_{n+1} = \alpha_{n+1}f(y_n) + (1 - \alpha_{n+1})Ty_n, \quad y_0 \in K, \quad (3.4)$$

where  $f : K \rightarrow K$  is a  $\beta$ -contraction. Choosing  $f \equiv u$ , one recovers the Halpern scheme (3.3) [41].

In a complete uniquely geodesic space, the nonlinear counterpart of (3.3) is

$$z_{n+1} = \alpha_{n+1}u \oplus (1 - \alpha_{n+1})Tz_n, \quad z_0, u \in K. \quad (3.5)$$

If  $X$  is a complete  $CAT(0)$  space and  $(\alpha_n)$  satisfies the above conditions (C1), (C2) and (C3) then (3.5) converges strongly to the metric projection of  $u$  onto  $\text{Fix}(T)$  [36]. This strong convergence is one of the main advantages of Halpern-type methods in the nonlinear setting.

### 3.1. RATE OF ASYMPTOTIC REGULARITY

Having examined the main convergence results for the KM iterative method, we now recall some important results on its rate of asymptotic regularity. This analysis is relevant to the use of such iterations in optimization algorithms for training deep learning models.

Consider now a nonexpansive operator  $T : K \rightarrow K$ , where  $K$  is a bounded and convex subset of a normed space  $X$ .

In [3], it was shown that the Krasnosel'skiĭ–Mann scheme (3.1) yields an asymptotic regularity rate that is independent of both the initial point  $x_0$  and the particular operator  $T$ , depending instead on the diameter of the set  $K$ . This was recently confirmed by Cominetti in [13], who proved that

$$\|x_n - Tx_n\| \leq \frac{\text{diam}(K)}{\sqrt{\pi \sum_{k=1}^n \lambda_k (1 - \lambda_k)}}.$$

Moreover, Bravo and Cominetti later demonstrated that this bound is sharp [7]. It follows that, under the standard assumptions on  $(\lambda_k)$ , the convergence rate remains bounded by  $\mathcal{O}(1/\sqrt{n})$ .

An explicit example of a nonexpansive linear operator is the right-shift operator acting on  $(\ell^1(\mathbb{N}), \|\cdot\|_1)$ . For this operator, the KM iteration starting from  $x^0 = (1, 0, 0, \dots)$  satisfies, for all  $n \geq 0$ ,  $\|x^n - Tx^n\|_1 \geq \frac{1}{\sqrt{n+1}}$  (see [14]).

However, faster rates can be achieved by alternative schemes: for instance, in Hilbert spaces, in [6] a convergence rate of  $\mathcal{O}(1/n)$  was obtained via a Nesterov-type acceleration technique.

As in the case of the Krasnosel'skiĭ–Mann iteration, we review some results from the literature on the rate of asymptotic regularity of the viscosity iteration (3.4) and the Halpern iteration (3.3). When applied to convex bilevel optimization problems, the authors of [35] derived the following estimate in the Euclidean setting  $X = \mathbb{R}^d$ :

$$\|z_n - Tz_n\| \leq \frac{4}{n+1} \|z_0 - P_{\text{Fix}(T)}u\|,$$

where  $P_{\text{Fix}(T)}(\cdot)$  denotes the metric projection onto the fixed-point set  $\text{Fix}(T)$ . Later, Contreras *et al.* [14] showed that this bound extends to general normed linear spaces. In the specific case where  $T$  is a nonexpansive operator defined on a Hilbert space, sharper results are available. Lieder [28] established the improved bound

$$\|z_n - Tz_n\| \leq \frac{2}{n+1} \|z_0 - P_{\text{Fix}(T)}u\|,$$

and furthermore constructed a family of nonexpansive operators on  $H = \mathbb{R}$  for which this inequality holds with equality, thus demonstrating the optimality of the rate. A simple example, highlighting the sharpness of the bound, is given by the rotation operator on the plane  $\mathbb{R}^2$  with angle  $\theta_n = \frac{\pi}{n+1}$ . For  $\lambda_k = \frac{k}{k+1}$  and starting point  $x_0 = (1, 0)$ , one obtains  $\|x_n - Tx_n\|_2 = \frac{2}{n+1}$  for every  $n$  (see [14]). Regarding the asymptotic regularity for both KM and Halpern schemes, the rates known in linear spaces extend to the nonlinear setting of geodesic spaces. In particular, Leuştean and Pinto [27] established that the alternating Halpern–Mann iteration in complete  $CAT(0)$  and more general hyperbolic spaces satisfies an asymptotic regularity bound of order  $O(1/\sqrt{n})$ , thus confirming the rate already known for the Krasnosel’skiĭ–Mann in Banach spaces. On the other hand, Pinto and Pischke [32] proved that the Halpern iteration with a variable anchor point in Hadamard spaces achieves the optimal rate  $O(1/n)$ , in agreement with the estimates derived in Hilbert spaces. These results confirm that the geometry of nonpositively curved metric spaces preserves the asymptotic efficiency of the classical fixed-point algorithms.

#### 4. MAIN RESULT

This section outlines our main research contributions. We first extend the classical analysis of Cominetti [13] on the asymptotic regularity rate of the Krasnosel’skiĭ–Mann (KM) iteration to the nonlinear setting of complete  $CAT(0)$  spaces, establishing the same convergence rate as in Banach spaces (see Subsection 4.1).

We then consider the viscosity iteration, and by adapting the argument of Sabach *et al.* [35], we derive an asymptotic regularity bound in  $CAT(0)$  spaces. As a special case, this result also covers the Halpern iteration, thus recovering in the nonlinear framework, via a proof adapted from the linear setting, estimates already known in the literature.

While the general argument for the KM scheme requires the space  $(X, d)$  to be complete, uniquely geodesic, and to satisfy the convexity condition (2.4) for the metric  $d$ , we will assume  $X$  to be a complete  $CAT(0)$  space. This choice reflects the central role that Hadamard spaces play in the convergence analysis of fixed-point iterations in the literature. Recall that every  $CAT(0)$  space is uniquely geodesic and satisfies the convexity condition (CC) (see [16, Lemma 2.4]).

##### 4.1. ASYMPTOTIC REGULARITY OF THE KRASNOSEL’SKIĬ–MANN ITERATION IN COMPLETE $CAT(0)$ SPACES

We begin this subsection by stating our main result through the following theorem.

**Theorem 4.1.** *Let  $(X, d)$  be a complete  $CAT(0)$  space, and let  $K \subset X$  be nonempty, closed, convex, and bounded. Suppose that  $T : K \rightarrow K$  is nonexpansive. Then, for the Krasnosel’skiĭ–Mann iteration (3.2) generated by a sequence  $\{\lambda_n\} \subset (0, 1)$ , the following inequality holds*

$$d(x_n, Tx_n) \leq \frac{\text{diam}(K)}{\sqrt{\pi \sum_{i=1}^n \lambda_i(1 - \lambda_i)}}, \quad n \geq 1.$$

The proof of Theorem 4.1 is based on a sequence of auxiliary lemmas that extend the argument of Cominetti [13] to the nonlinear metric setting. From this point on, we assume that  $(X, d)$  and  $T$  satisfy the hypotheses of Theorem 4.1, and that  $\text{diam}(K) = 1$ .

**Lemma 4.2.** *Let  $(x_m)$  be the Krasnosel'skiĭ–Mann iterates defined by (3.2), and define the weights*

$$\pi_k^n := \lambda_k \prod_{j=k+1}^n (1 - \lambda_j),$$

with the conventions  $\lambda_0 := 1$  and  $Tx_{-1} := x_0$ . Then, for all  $m \geq 0$  and  $q \in X$ , the following inequality holds:

$$d(x_m, q) \leq \sum_{j=0}^m \pi_j^m d(Tx_{j-1}, q). \quad (4.1)$$

*Proof.* The proof proceeds by induction on  $m$ . For  $m = 0$ , the claim follows immediately:

$$d(x_0, q) = \lambda_0 d(Tx_{-1}, q) = \pi_0^0 d(Tx_{-1}, q) = \sum_{j=0}^0 \pi_j^0 d(Tx_{j-1}, q).$$

Assume that (4.1) holds for  $m - 1$ . By the definition of  $x_m$ , we have

$$d(x_m, q) = d((1 - \lambda_m)x_{m-1} \oplus \lambda_m Tx_{m-1}, q).$$

Applying the convexity condition of the metric (CC) yields

$$d(x_m, q) \leq (1 - \lambda_m) d(x_{m-1}, q) + \lambda_m d(Tx_{m-1}, q).$$

Using the inductive hypothesis on  $d(x_{m-1}, q)$ , we obtain

$$d(x_m, q) \leq (1 - \lambda_m) \sum_{j=0}^{m-1} \pi_j^{m-1} d(Tx_{j-1}, q) + \lambda_m d(Tx_{m-1}, q),$$

and this coincides with

$$d(x_m, q) \leq \sum_{j=0}^m \pi_j^m d(Tx_{j-1}, q),$$

as claimed.  $\square$

**Lemma 4.3.** *Let  $(x_n)$  be the Krasnosel'skiĭ–Mann iterates defined by (3.2). Then, for all  $m \geq 0$  and  $r \geq 1$ , the following inequality holds:*

$$d(x_{m+r}, x_m) \leq \sum_{k=m+1}^{m+r} \pi_k^{m+r} d(Tx_{k-1}, x_m).$$

*Proof.* We proceed by induction on  $r$ . For  $r = 1$ , using the definition of  $x_{m+1}$  by (3.2), the convexity condition of the metric (CC) and the definition of  $\pi_k^n$ , we have

$$\begin{aligned} d(x_{m+1}, x_m) &= d((1 - \lambda_{m+1})x_m \oplus \lambda_{m+1}Tx_m, x_m) \\ &\leq (1 - \lambda_{m+1})d(x_m, x_m) + \lambda_{m+1}d(Tx_m, x_m) \\ &= \sum_{k=m+1}^{m+1} \pi_k^{m+1} d(Tx_{k-1}, x_m). \end{aligned}$$

Now assume the inequality holds for some  $r \geq 1$ . Then, using again the iterates definition and the CC of the metric,

$$\begin{aligned} d(x_{m+r+1}, x_m) &= d((1 - \lambda_{m+r+1})x_{m+r} \oplus \lambda_{m+r+1}Tx_{m+r}, x_m) \\ &\leq (1 - \lambda_{m+r+1})d(x_{m+r}, x_m) + \lambda_{m+r+1}d(Tx_{m+r}, x_m). \end{aligned}$$

Applying the inductive hypothesis to the first term and considering that  $\pi_k^{m+r+1} = (1 - \lambda_{m+r+1})\pi_k^{m+r}$  for  $m+1 \leq k \leq m+r$  and  $\pi_{m+r+1}^{m+r+1} = \lambda_{m+r+1}$ , we obtain

$$\begin{aligned} d(x_{m+r+1}, x_m) &\leq \sum_{k=m+1}^{m+r} \pi_k^{m+r+1} d(Tx_{k-1}, x_m) + \pi_{m+r+1}^{m+r+1} d(Tx_{m+r}, x_m) \\ &= \sum_{k=m+1}^{m+r+1} \pi_k^{m+r+1} d(Tx_{k-1}, x_m), \end{aligned}$$

which proves the statement for  $r+1$ . The claim follows by induction.  $\square$

**Proposition 4.4.** *Let  $(x_n)$  be the Krasnosel'skiĭ-Mann iterates defined by (3.2), and let  $n = m+r$  with  $m \geq 0$ ,  $r \geq 1$  (i.e.,  $0 \leq m < n$ ). Then the following estimate holds:*

$$d(x_n, x_m) \leq \sum_{j=0}^m \sum_{k=m+1}^n \pi_k^n \pi_j^m d(Tx_{j-1}, Tx_{k-1}). \quad (4.2)$$

*Proof.* By Lemma 4.3 and using  $n = m+r$ , we have

$$d(x_n, x_m) = d(x_{m+r}, x_m) \leq \sum_{k=m+1}^{m+r} \pi_k^{m+r} d(Tx_{k-1}, x_m).$$

Fix  $k$  in the summation and apply Lemma 4.2 with  $q = Tx_{k-1}$  and index  $m$ :

$$d(Tx_{k-1}, x_m) \leq \sum_{j=0}^m \pi_j^m d(Tx_{j-1}, Tx_{k-1}).$$

Substituting this into the previous estimate and observing that  $n = m+r$ , we obtain

$$d(x_n, x_m) \leq \sum_{k=m+1}^n \pi_k^n \sum_{j=0}^m \pi_j^m d(Tx_{j-1}, Tx_{k-1}),$$

which coincides with the desired inequality.  $\square$

**Proposition 4.5.** *If for all  $0 \leq m < n$ , we define  $c_{m,n}$  recursively by*

$$c_{m,n} := \sum_{j=0}^m \sum_{k=m+1}^n \pi_j^m \pi_k^n c_{j-1,k-1}, \quad \text{with } c_{-1,n} := 1 \text{ for all } n \geq 0,$$

then, for all  $0 \leq m < n$ ,

$$d(x_n, x_m) \leq c_{m,n}. \quad (4.3)$$

*Proof.* We proceed by induction on  $n$ . Fix  $n > m \geq 0$  and assume that

$$d(x_k, x_j) \leq c_{j,k} \quad \text{for all } 0 \leq j < k \leq n-1.$$

By Proposition 4.4 (i.e., (4.2)),

$$d(x_n, x_m) \leq \sum_{j=0}^m \sum_{k=m+1}^n \pi_j^m \pi_k^n d(Tx_{j-1}, Tx_{k-1}).$$

We split the double sum to treat separately the case  $j = 0$  and the case  $1 \leq j \leq m$ .

$$\begin{aligned} & \sum_{j=0}^m \sum_{k=m+1}^n \pi_j^m \pi_k^n d(Tx_{j-1}, Tx_{k-1}) \\ &= \sum_{k=m+1}^n \pi_0^m \pi_k^n d(Tx_{-1}, Tx_{k-1}) + \sum_{j=1}^m \sum_{k=m+1}^n \pi_j^m \pi_k^n d(Tx_{j-1}, Tx_{k-1}). \end{aligned}$$

Recalling  $Tx_{-1} = x_0$ ,  $c_{-1,n} := 1$  and using that  $d(x_0, Tx_{k-1}) \leq \text{diam}(K) = 1$  we obtain for the first term

$$\sum_{k=m+1}^n \pi_0^m \pi_k^n d(x_0, Tx_{k-1}) \leq \sum_{k=m+1}^n \pi_0^m \pi_k^n c_{-1,k-1}.$$

As for the second term, using the nonexpansiveness of  $T$  and the induction hypothesis,

$$d(Tx_{j-1}, Tx_{k-1}) \leq d(x_{j-1}, x_{k-1}) \leq c_{j-1,k-1}.$$

We then obtain

$$\sum_{j=1}^m \sum_{k=m+1}^n \pi_j^m \pi_k^n d(Tx_{j-1}, Tx_{k-1}) \leq \sum_{j=1}^m \sum_{k=m+1}^n \pi_j^m \pi_k^n c_{j-1,k-1}$$

Summing both terms yields

$$d(x_n, x_m) \leq \sum_{k=m+1}^n \pi_0^m \pi_k^n c_{-1,k-1} + \sum_{j=1}^m \sum_{k=m+1}^n \pi_j^m \pi_k^n c_{j-1,k-1} = c_{m,n},$$

which proves (4.3) and completes the induction.  $\square$

**Corollary 4.6.** *For every  $n \geq 0$ , the Krasnosel'skiĭ–Mann iterates satisfy*

$$d(x_n, Tx_n) \leq \frac{c_{n,n+1}}{\lambda_{n+1}} =: P_n.$$

*Proof.* By definition, we have  $d(x_n, x_{n+1}) = \lambda_{n+1} d(x_n, Tx_n)$ , and by Proposition 4.5 it holds  $d(x_n, x_{n+1}) \leq c_{n,n+1}$ . Dividing both sides by  $\lambda_{n+1}$  completes the proof.  $\square$

**Remark 4.7.** Using the probabilistic interpretation of the coefficients  $\pi_k^n$  (see [13]), one obtains the estimate:

$$\sqrt{\sum_{i=1}^n \lambda_i (1 - \lambda_i)} P_n \leq \frac{1}{\sqrt{\pi}}.$$

We observe that the construction of the weights  $\pi_k^n$  is purely algebraic and does not depend on the geometry of the underlying space. As a consequence, the probabilistic estimate from [13] extends directly to the  $CAT(0)$  setting.

Combining this inequality with Corollary 4.6 immediately yields the bound stated in Theorem 4.1.

#### 4.2. HALPERN/VISCOSITY ITERATION IN COMPLETE $CAT(0)$ SPACES: AN $O(1/k)$ ASYMPTOTIC REGULARITY BOUND

Throughout this subsection, let  $(X, d)$  be a complete  $CAT(0)$  space, and let  $K \subset X$  be nonempty, closed, and convex. Assume that  $T : K \rightarrow K$  is nonexpansive with  $\text{Fix}(T) \neq \emptyset$ , and that  $f : K \rightarrow K$  is a  $\beta$ -contraction for some  $\beta \in [0, 1)$ .

Given a sequence  $(\alpha_k)_{k \geq 1} \subset (0, 1]$ , we consider the viscosity-type Halpern iteration defined by

$$x_{k+1} = \alpha_{k+1} f(x_k) \oplus (1 - \alpha_{k+1}) T(x_k), \quad x_0 \in K \quad (4.4)$$

which is the nonlinear counterpart of (3.4) and for  $f \equiv u$  it coincides with iteration (3.5).

**Lemma 4.8.** *Set  $\bar{x} \in \text{Fix}(T)$ , and set  $z := f(\bar{x})$ . Then the sequence  $(x_k)$  generated by (4.4) is bounded and, for all  $k \geq 0$ ,*

$$d(x_k, \bar{x}) \leq C_{\bar{x}} \quad \text{with} \quad C_{\bar{x}} := \max \left\{ d(x_0, \bar{x}), \frac{1}{1-\beta} d(z, \bar{x}) \right\}.$$

*Proof.* Using the metric convexity in  $CAT(0)$  and the nonexpansiveness of  $T$ , we have

$$\begin{aligned} d(x_{k+1}, \bar{x}) &\leq (1 - \alpha_{k+1}) d(Tx_k, \bar{x}) + \alpha_{k+1} d(f(x_k), \bar{x}) \\ &\leq (1 - \alpha_{k+1}) d(x_k, \bar{x}) + \alpha_{k+1} (\beta d(x_k, \bar{x}) + d(z, \bar{x})), \end{aligned}$$

hence

$$d(x_{k+1}, \bar{x}) \leq d(x_k, \bar{x}) + \alpha_{k+1} d(z, \bar{x}) - \alpha_{k+1} (1 - \beta) d(x_k, \bar{x}).$$

If  $d(x_k, \bar{x}) \geq \frac{1}{1-\beta} d(z, \bar{x})$ , the last term is nonpositive and  $d(x_{k+1}, \bar{x}) \leq d(x_k, \bar{x})$ ; otherwise,  $d(x_{k+1}, \bar{x}) \leq \frac{1}{1-\beta} d(z, \bar{x})$ . By induction, the claim follows.  $\square$

**Lemma 4.9** ([35]). *Let  $M > 0$  and  $\gamma \in (0, 1]$ . Assume that  $(a_k) \subset [0, \infty)$  satisfies  $a_1 \leq M$  and*

$$a_{k+1} \leq (1 - \gamma b_{k+1}) a_k + (b_k - b_{k+1}) c_k, \quad k \geq 1,$$

where

$$b_k := \min \left\{ \frac{2}{\gamma k}, 1 \right\} \quad \text{and} \quad 0 \leq c_k \leq M < \infty \quad \text{for all } k.$$

Then, for all  $k \geq 1$ ,

$$a_k \leq \frac{M J}{\gamma k}, \quad J := \left\lceil \frac{2}{\gamma} \right\rceil.$$

**Remark 4.10.** By the nonexpansiveness of  $T$  and the contraction property of  $f$ , it follows that for every  $k \geq 1$  and  $\bar{x} \in \text{Fix}(T)$ , defining  $z := f(\bar{x})$ , the following inequalities hold:

$$d(Tx_{k-1}, \bar{x}) \leq d(x_{k-1}, \bar{x}), \quad d(f(x_{k-1}), z) \leq \beta d(x_{k-1}, \bar{x}).$$

**Theorem 4.11.** *Assume that  $\alpha_k := \min \left\{ \frac{2}{(1-\beta)k}, 1 \right\}$ , and let  $C_{\bar{x}}$  and  $J$  be the constants defined in Lemmas 4.8 and 4.9, respectively. Then the successive differences  $d(x_k, x_{k-1})$  and the asymptotic regularity residuals  $d(Tx_{k-1}, x_{k-1})$  admit bounds of order  $O(1/k)$ . In particular, the following estimates hold:*

$$d(x_k, x_{k-1}) \leq \frac{2J C_{\bar{x}}}{(1-\beta)k} \quad \text{and} \quad d(Tx_{k-1}, x_{k-1}) \leq \frac{2C_{\bar{x}}(J+2)}{(1-\beta)k} \quad (k \geq 1).$$

*Proof.* By the triangle inequality, and recalling the definitions of the iterates  $x_k$  and  $x_{k+1}$ , we obtain

$$\begin{aligned} & d(x_{k+1}, x_k) \\ & \leq d(x_{k+1}, \alpha_{k+1} f(x_{k-1}) \oplus (1 - \alpha_{k+1}) Tx_{k-1}) \\ & \quad + d(\alpha_{k+1} f(x_{k-1}) \oplus (1 - \alpha_{k+1}) Tx_{k-1}, x_k) \\ & = d(\alpha_{k+1} f(x_k) \oplus (1 - \alpha_{k+1}) Tx_k, \alpha_{k+1} f(x_{k-1}) \oplus (1 - \alpha_{k+1}) Tx_{k-1}) \quad \text{(I)} \\ & \quad + d(\alpha_{k+1} f(x_{k-1}) \oplus (1 - \alpha_{k+1}) Tx_{k-1}, \alpha_k f(x_{k-1}) \oplus (1 - \alpha_k) Tx_{k-1}). \quad \text{(II)} \end{aligned}$$

Using first the convexity inequality (2.5) for  $CAT(0)$  spaces, and then the nonexpansiveness of  $T$  together with the  $\beta$ -contraction of  $f$ , we obtain for term (I):

$$\begin{aligned} \text{(I)} & \leq (1 - \alpha_{k+1}) d(Tx_k, Tx_{k-1}) + \alpha_{k+1} d(f(x_k), f(x_{k-1})) \\ & \leq (1 - \alpha_{k+1} + \alpha_{k+1}\beta) d(x_k, x_{k-1}). \end{aligned}$$

For term (II), using the convexity condition of the metric (CC) and the geodesic property (2.1) (both points lie on the segment  $[f(x_{k-1}), Tx_{k-1}]$ ), we have

$$\text{(II)} \leq (\alpha_k - \alpha_{k+1}) d(f(x_{k-1}), Tx_{k-1}).$$

Substituting the bounds for (I) and (II) into the previous estimate, we obtain

$$\begin{aligned} d(x_{k+1}, x_k) &\leq (1 - \alpha_{k+1}(1 - \beta)) d(x_k, x_{k-1}) \\ &\quad + (\alpha_k - \alpha_{k+1}) d(f(x_{k-1}), Tx_{k-1}). \end{aligned} \quad (4.5)$$

Now let  $\bar{x} \in \text{Fix}(T)$  and set  $z := f(\bar{x})$ . Applying first the triangle inequality, and then the  $\beta$ -contraction property of  $f$  and the nonexpansiveness of  $T$ , we have

$$\begin{aligned} d(f(x_{k-1}), Tx_{k-1}) &\leq d(f(x_{k-1}), z) + d(z, \bar{x}) + d(\bar{x}, Tx_{k-1}) \\ &= d(f(x_{k-1}), f(\bar{x})) + d(f(\bar{x}), \bar{x}) + d(T\bar{x}, Tx_{k-1}) \\ &\leq \beta d(x_{k-1}, \bar{x}) + d(f(\bar{x}), \bar{x}) + d(\bar{x}, x_{k-1}). \end{aligned}$$

Finally, by Lemma 4.8 it holds  $d(x_{k-1}, \bar{x}) \leq C_{\bar{x}}$  and  $d(f(\bar{x}), \bar{x}) \leq (1 - \beta) C_{\bar{x}}$ , hence we obtain

$$d(f(x_{k-1}), Tx_{k-1}) \leq \beta C_{\bar{x}} + (1 - \beta) C_{\bar{x}} + C_{\bar{x}} = 2C_{\bar{x}}.$$

By setting  $a_k := d(x_k, x_{k-1})$ ,  $b_k := \alpha_k$ ,  $\gamma := 1 - \beta$ ,  $M := 2C_{\bar{x}}$  and  $c_k := d(f(x_{k-1}), Tx_{k-1})$ , and substituting into (4.5), we obtain

$$a_{k+1} \leq (1 - \gamma b_{k+1}) a_k + (b_k - b_{k+1}) c_k, \quad (k \geq 1), \quad c_k \leq M.$$

Moreover, by Lemma 4.8,

$$a_1 = d(x_1, x_0) \leq d(x_1, \bar{x}) + d(x_0, \bar{x}) \leq 2C_{\bar{x}} = M.$$

Therefore, by Lemma 4.9,

$$d(x_k, x_{k-1}) \leq \frac{2JC_{\bar{x}}}{(1 - \beta)k}.$$

This completes the proof of the first claim regarding the convergence rate of (4.4).

We now turn to its asymptotic regularity bound. Using the triangle inequality and the fact that  $x_k \in [f(x_{k-1}), Tx_{k-1}]$ , we obtain

$$d(Tx_{k-1}, x_{k-1}) \leq d(x_k, Tx_{k-1}) + d(x_k, x_{k-1}) = \alpha_k d(Tx_{k-1}, f(x_{k-1})) + d(x_k, x_{k-1}).$$

Recalling the definition of  $\alpha_k$ , together with the bounds  $d(f(x_{k-1}), Tx_{k-1}) \leq 2C_{\bar{x}}$  and the first claim  $d(x_k, x_{k-1}) \leq \frac{2JC_{\bar{x}}}{(1 - \beta)k}$ , we deduce

$$d(Tx_{k-1}, x_{k-1}) \leq \frac{2}{k(1 - \beta)} 2C_{\bar{x}} + \frac{2JC_{\bar{x}}}{(1 - \beta)k} = \frac{2C_{\bar{x}}(J + 2)}{(1 - \beta)k}.$$

Hence, the second claim also holds, and both sequences  $d(x_k, x_{k-1})$  and  $d(Tx_{k-1}, x_{k-1})$  decay at rate  $O(1/k)$ .  $\square$

This completes the analysis of the viscosity (Halpern-type) scheme in complete  $CAT(0)$  spaces. Together with the results obtained for the Krasnosel'skiĭ–Mann iteration, this demonstrates that the proof strategies from the linear setting can be carried over to the nonlinear framework, yielding in particular the known  $O(1/\sqrt{k})$  and  $O(1/k)$  asymptotic regularity bounds for the two fundamental fixed-point algorithms.

## 5. HYPERBOLIC HALPERNGD-BASED OPTIMIZER

In this section, we formulate a HalpernGD-type scheme on Hadamard spaces whose specialization to  $\mathbb{H}^n$  yields the Hyperbolic HalpernGD optimizer. The key idea is to replace the Euclidean forward step with the proximal (resolvent) mapping, which remains nonexpansive in the metric setting. The properties of the proximal mapping recalled in Section 2 naturally motivate this construction. In the Euclidean setting, the standard HalpernSGD iteration exploits the nonexpansivity of the forward operator  $F = I - \eta \nabla f$ , whose averaged nature is guaranteed by the Baillon–Haddad Theorem 2.12. However, in nonlinear spaces it does not hold in general, and we provided a counterexample.

5.1. FAILURE OF NONEXPANSIVITY OF THE FORWARD OPERATOR ON  $\mathbb{H}^2$ 

To the best of our knowledge, explicit examples of nonexpansive forward operators in non-Euclidean settings have not been investigated in the literature. The present section aims to fill this gap.

In particular, we will point out that, even on the hyperbolic plane, geodesic convexity and smoothness do not suffice to guarantee the nonexpansivity of the explicit forward step.

Let  $X$  be a general Hadamard manifold, let  $f \in C^1(X)$ , and let  $\eta > 0$ ; then the associated forward operator is defined by

$$F_\eta(x) := \exp_x(-\eta \nabla f(x)), \quad x \in X.$$

Moreover, a function  $f \in C^2(X)$  is geodesically convex (see Theorem 6.2 in [39]) if and only if

$$\text{Hess } f_x(v, v) \geq 0 \quad \forall x \in X, \forall v \in T_x X.$$

**Theorem 5.1.** *There exists a  $C^\infty$  geodesically convex function  $f : \mathbb{H}^2 \rightarrow \mathbb{R}$  with nonempty minimizer set such that, for every  $\eta > 0$ , the map  $F_\eta(x) = \exp_x(-\eta \nabla f(x))$  is not nonexpansive. Equivalently, for every  $\eta > 0$  there exist  $x, y \in \mathbb{H}^2$  such that*

$$d_H(F_\eta(x), F_\eta(y)) > d_H(x, y).$$

*Proof.* We work in the hyperboloid model of  $\mathbb{H}^2$  and use the global Fermi coordinates  $(r, s) \in \mathbb{R}^2$  around the geodesic  $\Gamma = \{(\cosh s, \sinh s, 0) : s \in \mathbb{R}\}$ , given by

$$\Phi(r, s) := (\cosh r \cosh s, \cosh r \sinh s, \sinh r).$$

In these coordinates, the hyperbolic metric is  $\Phi^* g_H = dr^2 + \cosh^2(r) ds^2$ .

Define  $f : \mathbb{H}^2 \rightarrow \mathbb{R}$  by  $f(\Phi(r, s)) := \cosh r - 1$ . Then  $f \geq 0$ , and  $f = 0$  if and only if  $r = 0$ . Hence,  $\arg \min f = \Gamma$ . Moreover,

$$df = \sinh r \, dr, \quad \nabla f = \sinh r \, \partial_r.$$

and, by direct computation, we have:

$$\text{Hess } f(\partial_r, \partial_r) = \cosh r, \quad \text{Hess } f(\partial_r, \partial_s) = 0, \quad \text{Hess } f(\partial_s, \partial_s) = \sinh^2 r \cosh r.$$

Therefore, Hess  $f$  is positive semidefinite, and thus  $f$  is geodesically convex. Fix  $\eta > 0$ . Since  $\frac{\sinh r}{r} \rightarrow +\infty$  as  $r \rightarrow +\infty$ , we may choose  $r > 0$  such that

$$\eta \sinh r > 2r.$$

Let  $a > 0$ , and set  $x := \Phi(r, 0)$ ,  $y := \Phi(r, a)$ . For fixed  $s \in \mathbb{R}$ , consider the curve

$$\gamma(t) := \Phi(r - t\eta \sinh r, s), \quad t \in \mathbb{R}.$$

This is a geodesic, since it is an affine reparametrization of the curves  $u \mapsto \Phi(r - u, s)$  that are unit-speed geodesics orthogonal to  $\Gamma$ . Moreover,

$$\gamma(0) = \Phi(r, s), \quad \gamma'(0) = -\eta \sinh r \partial_r.$$

From  $\nabla f(\Phi(r, s)) = \sinh r \partial_r$ , it follows that

$$F_\eta(\Phi(r, s)) = \exp_{\Phi(r, s)}(-\eta \sinh r \partial_r) = \gamma(1) = \Phi(r - \eta \sinh r, s).$$

Therefore,  $F_\eta(x) = \Phi(r - \eta \sinh r, 0)$  and  $F_\eta(y) = \Phi(r - \eta \sinh r, a)$ .

By the hyperboloid distance formula, it follows that

$$\cosh d_H(x, y) = -\langle \Phi(r, 0), \Phi(r, a) \rangle_L = 1 + \cosh^2 r (\cosh a - 1),$$

and analogously

$$\cosh d_H(F_\eta(x), F_\eta(y)) = 1 + \cosh^2(r - \eta \sinh r) (\cosh a - 1).$$

By the choice of  $r$ , we have

$$r - \eta \sinh r < -r, \quad \text{hence} \quad |r - \eta \sinh r| > r.$$

Since  $\cosh$  is even and strictly increasing on  $[0, +\infty)$ , it follows that

$$\cosh^2(r - \eta \sinh r) > \cosh^2 r.$$

From  $\cosh a - 1 > 0$ , it follows that

$$\cosh d_H(F_\eta(x), F_\eta(y)) > \cosh d_H(x, y).$$

Finally, since  $\operatorname{arcosh}$  is strictly increasing on  $[1, +\infty)$ ,

$$d_H(F_\eta(x), F_\eta(y)) > d_H(x, y).$$

Since  $\eta > 0$  was arbitrary, this proves that for every  $\eta > 0$  there exist  $x, y \in \mathbb{H}^2$  such that

$$d_H(F_\eta(x), F_\eta(y)) > d_H(x, y). \quad \square$$

**Remark 5.2.** The previous counterexample shows that geodesic convexity alone is not sufficient to guarantee nonexpansiveness of the explicit forward step on Hadamard manifolds. Moreover, since  $f$  is  $C^2$ , its Hessian is bounded on every closed geodesic ball; hence  $\nabla f$  is Lipschitz on every bounded convex region. Therefore, local smoothness on bounded convex subsets does not suffice either. Thus, in nonpositively curved geometry, the proximal mapping is not merely a convenient substitute for the forward operator, but in general the natural nonexpansive object.

In this context, the proximal (or resolvent) mapping introduced in Definition 2.13 provides a geometric analogue of the forward step, preserving nonexpansivity in the metric sense.

Following Bačák [1, Ch. 2, §2.2, Prop. 2.2.24, p. 44; Ch. 4, §4.2, Lemma 4.2.2, p. 73], the proximal mapping  $J_{\lambda f}(x)$  is well defined in every Hadamard space  $X$ : by [1, Prop. 2.2.24], the minimizer exists and is unique for each  $x \in X$ ; by [1, Lemma 4.2.2], the operator  $J_{\lambda f}$  is both 1-Lipschitz and firmly nonexpansive. Moreover, as shown in [1, Eq. (2.2.24)],

$$x = J_{\lambda f}(x) \iff x \in \arg \min f,$$

so that the fixed points of  $J_{\lambda f}$  coincide with the minimizers of  $f$ . This equivalence establishes  $J_{\lambda f}$  as the natural nonlinear generalization of the Euclidean forward operator.

Assume in addition that  $\arg \min f \neq \emptyset$ . Replacing the Euclidean forward step with the resolvent  $J_{\lambda f}$  naturally leads to a Hadamard-space analogue of the Halpern iteration; in particular, this yields the desired hyperbolic version when  $X = \mathbb{H}^n$ . For any anchor point  $u \in X$  and any control sequence  $\{\alpha_k\} \subset (0, 1)$ , we obtain

$$x_{k+1} = \alpha_{k+1}u \oplus (1 - \alpha_{k+1})J_{\lambda f}(x_k), \quad x_0 \in X. \quad (5.1)$$

Since  $J_{\lambda f}$  is firmly nonexpansive, it is in particular nonexpansive. Hence, under the standard assumptions on  $(\alpha_k)$ , the sequence  $(x_k)$  is asymptotically regular and converges strongly to the metric projection of  $u$  onto

$$\text{Fix}(J_{\lambda f}) = \arg \min f.$$

Iteration (5.1) thus provides a theoretical foundation for a Hyperbolic Halpern-GD optimizer; indeed, convergence and quantitative results for related Halpern-type proximal point algorithms have already been established in Hilbert spaces and, more generally, in  $CAT(0)$  spaces [26, 38]. It combines the geometric advantages of Hadamard spaces, such as the ability to represent hierarchical or tree-like data structures, with the accelerated convergence properties of Halpern-type iterations. From a practical standpoint, this construction bridges the gap between fixed-point theory in  $CAT(0)$  spaces and modern geometric deep learning. In particular, the convergence guaranteed by the nonexpansivity of  $J_{\lambda f}$  mirrors the behavior of HalpernSGD in Euclidean spaces, while adapting naturally to curved manifolds through the metric formulation of the proximal step.

## 6. CONCLUSION AND FUTURE WORK

Our first main contribution lies in showing how proof strategies originally developed in Banach and Hilbert spaces, where linearity simplifies many arguments, can be successfully reformulated in the purely metric setting of  $CAT(0)$  spaces. This approach not only recovers the classical asymptotic regularity bounds but also clarifies which steps of the convergence analysis depend essentially on the linear structure and which extend naturally to a general geodesic framework. In addition, we introduced a hyperbolic optimization scheme, the Hyperbolic HalpernGD optimizer, obtained by replacing the Euclidean forward step with the proximal (resolvent) mapping in Hadamard spaces. This construction provides a HalpernGD-type algorithm in nonpositively curved geometries, thereby yielding a principled optimizer for deep models operating in Hadamard spaces. As future work, we plan to implement and experimentally evaluate the stochastic counterpart of the Hyperbolic HalpernGD optimizer, comparing its performance with that of the standard Riemannian stochastic gradient descent (RSGD) widely used in hyperbolic deep learning. From a theoretical standpoint, the asymptotic regularity rates suggest that Halpern-type methods should converge faster than Krasnosel'skii–Mann-type iterations, of which gradient descent is a particular instance. Therefore, we expect the Hyperbolic HalpernSGD optimizer to exhibit superior convergence speed and empirical efficiency compared with RSGD, in line with the improvements already observed in the Euclidean setting for HalpernSGD over classical SGD [12, 18]. Ultimately, this line of research aims to bridge fixed-point theory and convex analysis in metric spaces with hyperbolic deep learning, paving the way for faster and more geometrically consistent optimization algorithms on manifolds with curvature.

### Acknowledgements

*This publication was partially funded by the PhD program in Mathematics and Computer Science at the University of Calabria, Cycle XXXVIII, with the support of a scholarship financed by DM 351/2022 (CUP H23C22000440007), within the NRRP funded by the European Union.*

## REFERENCES


- [1] M. Bačák, *Convex Analysis and Optimization in Hadamard Spaces*, Walter de Gruyter, Berlin–Boston, 2014.
- [2] J. Baillon, G. Haddad, *Quelques propriétés des opérateurs angle-bornés et  $n$ -cycliquement monotones*, Israel J. Math. **26** (1977), 137–150.
- [3] J. Baillon, R.E. Bruck, *The rate of asymptotic regularity is  $\mathcal{O}(1/\sqrt{n})$* , [in:] A.G. Kartsatos (ed.), *Theory and Applications of Nonlinear Operators of Accretive and Monotone Type*, Lecture Notes in Pure and Appl. Math. **178**, Dekker, New York, 1996, 51–81.
- [4] H.H. Bauschke, P.L. Combettes, *The Baillon-Haddad Theorem Revisited*, J. Convex Anal. **17** (2010), 781–787.

- 
- [5] S. Bonnabel, *Stochastic gradient descent on Riemannian manifolds*, IEEE Trans. Automat. Control **58** (2013), 2217–2229.
- [6] R.I. Boç, D. Nguyen, *Fast Krasnosel’skiĭ-Mann algorithm with a convergence rate of the fixed point iteration of  $o(1/k)$* , arXiv:2206.09462, 2022.
- [7] M. Bravo, R. Cominetti, *Sharp convergence rates for averaged nonexpansive maps*, Israel J. Math. **227** (2018), 163–188.
- [8] M.R. Bridson, A. Haefliger, *Metric Spaces of Non-Positive Curvature*, Springer, Berlin, 2013.
- [9] M.M. Bronstein, J. Bruna, Y. LeCun, A. Szlam, P. Vandergheynst, *Geometric deep learning: going beyond Euclidean data*, IEEE Signal Process. Mag. **34** (2017), 18–42.
- [10] F.E. Browder, *Nonexpansive nonlinear operators in a Banach space*, Proc. Nat. Acad. Sci. U.S.A. **54** (1965), 1041–1044.
- [11] V. Colao, K.R. Foglia, *On the Convergence of HalpernSGD*, arXiv:2601.18906, 2026.
- [12] V. Colao, K.R. Foglia, A. Giordano, E. Ritacco, W. Spataro, *An optimizer derived from Halpern’s method for enhanced neural network convergence and reduced carbon emissions*, J. Intell. Inf. Syst. **64** (2026), 77–96.
- [13] R. Cominetti, J.A. Soto, J. Vaisman, *On the rate of convergence of Krasnosel’skiĭ-Mann iterations and their connection with sums of Bernoullis*, Israel J. Math. **199** (2014), 757–772.
- [14] J.P. Contreras, R. Cominetti, *Optimal error bounds for non-expansive fixed-point iterations in normed spaces*, Math. Program. **199** (2023), 343–374.
- [15] S. Dhompongsa, W.A. Kirk, B. Sims, *Fixed points of uniformly Lipschitzian mappings*, Nonlinear Anal. **65** (2006), 762–772.
- [16] S. Dhompongsa, B. Panyanak, *On  $\Delta$ -convergence theorems in  $CAT(0)$  spaces*, Comput. Math. Appl. **56** (2008), 2572–2579.
- [17] R. Espínola, A. Fernández-León,  *$CAT(k)$ -spaces, weak convergence and fixed points*, J. Math. Anal. Appl. **353** (2009), 410–427.
- [18] K.R. Foglia, V. Colao, E. Ritacco, *HalpernSGD: A Halpern-Inspired Optimizer for Accelerated Neural Network Convergence and Reduced Carbon Footprint*, [in:] A. Appice, H. Azzag, M.S. Hacid, A. Hadjali, Z.W. Ras (eds.), *Foundations of Intelligent Systems*, Lecture Notes in Comput. Sci. **14670**, Springer, Cham, 2024, 296–305.
- [19] O.-E. Ganea, G. Bécigneul, T. Hofmann, *Hyperbolic neural networks*, Adv. Neural Inf. Process. Syst. **31** (2018), 5345–5355.
- [20] B. Halpern, *Fixed points of nonexpanding maps*, Bull. Amer. Math. Soc. **73** (1967), 957–961.
- [21] J. He, D. Fang, G. López, C. Li, *Mann’s algorithm for nonexpansive mappings in  $CAT(\kappa)$  spaces*, Nonlinear Anal. **75** (2012), 445–452.
- [22] S. Ishikawa, *Fixed points and iteration of a nonexpansive mapping in a Banach space*, Proc. Amer. Math. Soc. **59** (1976), 65–71.

- [23] W.A. Kirk, *Geodesic geometry and fixed point theory*, [in:] *Seminar of Mathematical Analysis (Málaga/Sevilla, 2002/2003)*, Colecc. Abierta **64**, Univ. Sevilla Secr. Publ., Sevilla, 2003, 195–225.
- [24] W.A. Kirk, *Geodesic geometry and fixed point theory II*, [in:] *International Conference on Fixed Point Theory and Applications*, Yokohama Publ., Yokohama, 2004, 113–142.
- [25] W.A. Kirk, B. Panyanak, *A concept of convergence in geodesic spaces*, *Nonlinear Anal.* **68** (2008), 3689–3696.
- [26] L. Leuştean, P. Pinto, *Quantitative results on a Halpern-type proximal point algorithm*, *Comput. Optim. Appl.* **79** (2021), no. 1, 101–125.
- [27] L. Leuştean, P. Pinto, *Rates of asymptotic regularity for the alternating Halpern–Mann iteration*, *J. Math. Anal. Appl.* **515** (2022), 126477.
- [28] F. Lieder, *On the convergence rate of the Halpern-iteration*, *Optim. Lett.* **15** (2021), 405–418.
- [29] T.C. Lim, *Remarks on some fixed point theorems*, *Proc. Amer. Math. Soc.* **60** (1976), 179–182.
- [30] W.R. Mann, *Mean value methods in iteration*, *Proc. Amer. Math. Soc.* **4** (1953), 506–510.
- [31] W. Peng, T. Varanka, A. Mostafa, H. Shi, G. Zhao, *Hyperbolic deep neural networks: a survey*, *IEEE Trans. Pattern Anal. Mach. Intell.* **44** (2022), no. 12, 10023–10044.
- [32] P. Pinto, N. Pischke, *On the Halpern method with adaptive anchoring parameters*, *Math. Comp.*, published electronically in 2026, DOI: 10.1090/mcom/4182.
- [33] R.T. Rockafellar, *Monotone operators associated with saddle-functions and minimax problems*, [in:] F.E. Browder (ed.), *Nonlinear Functional Analysis, Part 1*, Proc. Sympos. Pure Math. **18**, Amer. Math. Soc., Providence, RI, 1970, 241–250.
- [34] R.T. Rockafellar, *Monotone operators and the proximal point algorithm*, *SIAM J. Control Optim.* **14** (1976), 877–898.
- [35] S. Sabach, S. Shtern, *A first order method for solving convex bilevel optimization problems*, *SIAM J. Optim.* **27** (2017), 640–660.
- [36] S. Saejung, *Halpern’s iteration in  $CAT(0)$  spaces*, *Fixed Point Theory Appl.* **2010** (2009), 1–13.
- [37] F. Sala, C. De Sa, A. Gu, C. Ré, *Representation tradeoffs for hyperbolic embeddings*, [in:] J. Dy, A. Krause (eds.), *Proceedings of the 35th International Conference on Machine Learning*, Proc. Mach. Learn. Res. **80**, PMLR, 2018, 4460–4469.
- [38] A. Sipoş, *Abstract strongly convergent variants of the proximal point algorithm*, *Comput. Optim. Appl.* **83** (2022), no. 1, 349–380.
- [39] C. Udrişte, *Convex Functions and Optimization Methods on Riemannian Manifolds*, Math. Appl., Kluwer Acad. Publ., Dordrecht, 1994.
- [40] R. Wittmann, *Approximation of fixed points of nonexpansive mappings*, *Arch. Math. (Basel)* **58** (1992), 486–491.
- [41] H.-K. Xu, *Viscosity approximation methods for nonexpansive mappings*, *J. Math. Anal. Appl.* **298** (2004), 279–291.

Vittorio Colao  
vittorio.colao@unical.it  
 <https://orcid.org/0000-0003-0743-0137>

University of Calabria  
Department of Mathematics and Computer Science  
Ponte P. Bucci, 30B, Arcavacata di Rende (CS), 87036, Italy

Katherine Rossella Foglia (corresponding author)  
katherine.foglia@unical.it  
 <https://orcid.org/0009-0009-0192-2684>

University of Calabria  
Department of Mathematics and Computer Science  
Ponte P. Bucci, 30B, Arcavacata di Rende (CS), 87036, Italy

*Received: October 31, 2025.*

*Revised: March 25, 2026.*

*Accepted: May 8, 2026.*

*Published online: June 23, 2026.*