Thai Journal of Mathematics
Volume 19 Number 2 (2021)
Pages 693–711

http://thaijmath.in.cmu.ac.th

ISSN 1686-0209

Splitting Proximal Algorithms for Convex Optimizations over Metric Spaces with Curvature Bounded Above

Sakan Termkaew[1,2], **Poom Kumam**[1,2,3], **Parin Chaipunya**[1,3,*]

[1] *Department of Mathematics, Faculty of Science, King Mongkut's University of Technology Thonburi (KMUTT), 126 Pracha Uthit Road, Bang Mod, Thung Khru, Bangkok 10140, Thailand*

e-mail : sakan.gotsakan@mail.kmutt.ac.th (S. Termkaew); poom.kum@kmutt.ac.th (P. Kumam); parin.cha@mail.kmutt.ac.th (P. Chaipunya)

[2] *KMUTTFixed Point Research Laboratory, King Mongkut's University of Technology Thonburi (KMUTT), 126 Pracha Uthit Road, Bang Mod, Thung Khru, Bangkok 10140, Thailand*

[3] *Center of Excellence in Theoretical and Computational Science (TaCS-CoE), King Mongkut's University of Technology Thonburi (KMUTT), 126 Pracha Uthit Road, Bang Mod, Thung Khru, Bangkok 10140, Thailand*

[*] *Corresponding author.*

Abstract In this paper, we consider a splitting method combined with proximal methods for minimizing the sum of convex functions, where the proximal operators are defined by the curvature-adapted regularizations. In this paper, using properties of the resolvent on complete CAT(κ) space, we present two theorems showing strong convergence of the splitting proximal algorithm under different hypotheses. The first theorem assumes the local compactness property on the ambient space, while the second relies on the uniform convexity of the objective function. We also apply our main results to solve convex feasibility problems, centroid problems, and particularly the Karcher means. Finally, we include a series of numerical implementations of our algorithms to approximate the Karcher means of some randomly generated datasets fitted to the Lobachevskii plane.

MSC: 90C06; 90C25; 90C48

Keywords: CAT(κ) Space, Convex Optimization, Splitting Proximal Algorithms

Submission date: 11.02.2020 / Acceptance date: 03.06.2020

1. Introduction

The proximal algorithm has been one of the most earliest and successful approximation scheme for convex and generalized convex optimization. The algorithm was first proposed by Moreau [1] and became a keystone by the important study of Rockafellar [2] under the context of a Hilbert space. The method was originally based on the first-order theory of

Published by The Mathematical Association of Thailand.

convex functions, in particular the subdifferential theory, where the resolvent of the subdifferential is applied iteratively. The resolvent was later realized in a minimization-based formulation, which we call the proximal operator. Let us recall that the proximal operator is written solely in terms of the norm. In 2005, Combettes et al. [3] have shown the splitting proximal algorithm on Hilbert spaces for solving convex optimization problems. Later, the splitting methods was developed and applied to solve various optimization problems [4–9]. In 1995 and 1998, Jost [10] and Mayer [11] independently extended the proximal operator to a CAT(0) space by replacing the norm with the distance function. In fact, if (X,d) is a complete CAT(0) space and $f : X \to (-\infty, +\infty]$, then the proximal operator of f is the mapping $J_f : X \multimap X$ defined by

$$\operatorname{prox}_f(x) = \operatorname*{arg\,min}_{y\in X} \left[f(y) + \frac{1}{2} d(x,y)^2 \right] \tag{1.1}$$

for all $x \in X$. If f is proper, (geodesically) convex, and lsc, then J_f is well-defined as a single-valued mapping on the whole space X [10]. It was not until 2013 that this proximal operator was applied to solve convex optimization in complete CAT(0) spaces by Bačák [12]. The subdifferential approach to proximal operators was discussed in [13].

The needs for optimization over nonlinear spaces have been in an increasing trends, e.g. in computational biology and data management, particularly in manifold-valued data and image processing. In such applications, the objective function is usually expressed as the sum of several loss functions, i.e. $f := \sum_{i=1}^m f_i$ where $f_i : X \to (-\infty, +\infty]$ for $i = 1, \ldots, m$. This function can be extraordinarily large and costly to optimize, hence leads to the use of splitting techniques in optimization. In [14], the author proposed a splitting method which replaces the expensive computation of prox_f with a bundle of cheaper ones, i.e. $\operatorname{prox}_{f_i}$ for $i = 1, \ldots, m$. Such a method is called the splitting proximal algorithm, and it was first proved to be strongly convergent to a minimizer of f in [14] under the assumption that each component loss function f_i is Lipschitz continuous and the ambient space is locally compact. The proximal and splitting proximal algorithms were further generalized to complete CAT(κ) spaces for any $\kappa \in \mathbb{R}$ by Ohta [15], with an additional diametric condition for $\kappa > 0$. In these results, the ambient space X was again restricted to the locally compact cases. To the best of our knowledge, the question of possible relaxation of local compactness is still open until today [16].

In 2016, Kimura and Kohsaka [17] introduced a new concept of a proximal operator by replacing the quadratic kernel $t \mapsto \frac{1}{2}t^2(\cdot,\cdot)$, for $t \geq 0$, with the curvature-adapted kernel $\phi : \mathbb{R} \times [0,\infty) \to [0,\infty)$ given by

$$\phi_\kappa(t) := \phi(\kappa; t) := \begin{cases} \tan(t)\sin(t) & \text{if } \kappa > 0, \\ \frac{1}{2}t^2 & \text{if } \kappa = 0, \\ \tanh(t)\sinh(t) & \text{if } \kappa < 0, \end{cases}$$

for each $t \geq 0$. Let (X,d) be a complete CAT(κ) space with $\kappa \in \mathbb{R}$ and $f : X \to (-\infty, +\infty]$ is proper convex lsc function, the novel proximal operator of f is mapping $R_f^\kappa : X \to X$ such that

$$R_f^\kappa(x) := \operatorname*{arg\,min}_{y\in X} \left[f(y) + \phi_\kappa(d(y,x)) \right]$$

for each $x \in X$. This mapping is well-defined as a single-valued mapping under the admissibility condition of the ambient space X [17, 18]. In the case of $\kappa > 0$, the proximal algorithm using this curvature-adapted kernel ϕ_κ was proved to be convergent (in a weaker sense) to a minimizer of f by Kimura and Kohsaka in [17] for constant step-sizes, and later in [19] for variable step-sizes. Independently and unknowingly of the results in [19], Espínola and Nicolae [20] adopted the idea of [17] and developed the proximal algorithm using curvature-adapted kernel in the case $\kappa > 0$ and obtained similar convergence as presented in [19]. In addition to this result, Espínola and Nicolae [20] also applied the splitting technique of [14] to the curvature-adapted kernel and finally proved convergence of the splitting proximal algorithm for this new type of proximal operator for component loss functions f_i being Lipschitz continuous on admissible compact CAT(κ) spaces, again, with $\kappa > 0$.

Recently, Kajimura and Kimura [18] investigated on the proximal algorithms with the operators R_f^κ in CAT(κ) spaces with $\kappa < 0$ and obtained the weak convergence of the scheme. This result therefore completes the convergence behaviors of the proximal algorithms with the operator R_f^κ for all cases of $\kappa > 0$, $\kappa = 0$, and $\kappa < 0$.

Motivated by the above studies, we show in this paper the convergence of splitting proximal algorithms by using the new operator R_f^κ in complete locally compact CAT(κ) spaces where $\kappa < 0$. This result, alongside with [14] and [20], would therefore complete the study of splitting proximal algorithm of curvature-adapted proximals in CAT(κ) spaces. In addition, we also show that if the total cost f is ϕ_κ-uniformly convex, then the strong convergence of the splitting proximal algorithm remains valid without the local compactness condition for any cases of $\kappa \in \mathbb{R}$. In addition, we can also guarantee the sublinear rate of convergence in this latter study. This result extends that of [15] to curvature-adapted proximals and completes similar findings for all cases of $\kappa \in \mathbb{R}$ in the noncompact settings. Finally, we discuss some possible applications where the splitting proximal algorithms may be useful and demonstrate a numerical implementation applied to solve Karcher mean problems, which is used in averaging nonlinear data sets.

2. Preliminaries

In this section, some basic concepts and useful lemmas necessary for the subsequent results are given. Throughout this paper, the set of all positive integers and the set of all real numbers are denoted by $\mathbb{N}$ and $\mathbb{R}$, respectively.

Let (X, d) be a metric space and $x, y \in X$. A *geodesic path* joining x to y is a mapping $\gamma : [0, 1] \to X$ such that $\gamma(0) = x$, $\gamma(1) = y$ and $d(\gamma(t_1), \gamma(t_2)) = d(x, y)|t_1 - t_2|$ for any $t_1, t_2 \in [0, 1]$. We say that X is a *(uniquely) geodesic metric space* if any two points are connected by a (unique) geodesic. If X is a uniquely geodesic, $x, y \in X$, and γ is the geodesic path joining x to y, then we write $[x, y] := \gamma([0, 1])$ to denote the *geodesic segment* of γ. In this case, we also use the notation $(1 - t)x \oplus ty := \gamma(t)$. Also recall that *geodesic triangle* with vertices $x, y, z \in X$, denoted by $\Delta(x, y, z)$, is defined by $[x, y] \cup [y, z] \cup [z, x]$.

Given $\kappa \in \mathbb{R}$, the *model space* M_κ^2 denotes the 2-dimensional complete simply-connected Riemannian manifold with constant sectional curvature κ. On each model space M_κ^2, its Riemannian distance is denoted by d_κ. If $\kappa = 0$, then $M_0^2 = \mathbb{E}^2$ is the Euclidean plane. For other κ, recall that M_κ can always be rescaled into $M_1^2 = \mathbb{S}^2$ for $\kappa > 0$, and into $M_{-1}^2 = \mathbb{H}^2$ for $\kappa < 0$. Here, $\mathbb{S}^2$ and $\mathbb{H}^2$ denote the unit 2-sphere and Lobachevskii plane. Hence, we can concentrate only on the cases $\kappa = 0$, $\kappa = 1$, and $\kappa = -1$. Denoted by D_κ, we mean the diameter of the model space M_κ^2, that is,

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

A *κ-comparison triangle* for a geodesic triangle $\Delta(x_1, x_2, x_3)$ in X is a triangle $\Delta(\overline{x}_1, \overline{x}_2, \overline{x}_3)$ in M^2_κ such that $d(x_i, x_j) = d_\kappa(\overline{x}_i, \overline{x}_j)$ for all $i, j \in \{1, 2, 3\}$. A κ-comparison triangle always exists provided that $d(x_1, x_2) + d(x_2, x_3) + d(x_3, x_1) < 2D_\kappa$ and is unique up to isometries. For any $i, j \in \{1, 2, 3\}$, let γ be the geodesic joining x_i to x_j and $u := \gamma(t)$ for some $t \in [0, 1]$. Then a point $\overline{u} := \overline{\gamma}(t)$ is called the *comparison point* of u, where $\overline{\gamma}$ is the geodesic joining $\overline{x}_i$ to $\overline{x}_j$. A geodesic triangle Δ in X is said to satisfy the CAT(κ) *inequality* if the following inequality holds for all $x, y \in \Delta$:

$$d(x, y) \leq d_\kappa(\overline{x}, \overline{y}),$$

where $\overline{\Delta}$ is the κ-comparison triangle of Δ and $\overline{x}, \overline{y} \in \overline{\Delta}$ are the respective comparison points of x and y. For $\kappa \in \mathbb{R}$, a geodesic space X is called a CAT(κ) *space* if every geodesic triangle satisfies the CAT(κ) inequality. A CAT(κ) space X is called *admissible* if $d(x, y) < D_\kappa/2$ for all $x, y \in X$. It is immediate that every CAT(0) space is admissible.

Remark 2.1 (See [21]). In general, if $\kappa > 0$, then (X, d) is a CAT(κ) space if and only if $(X, \sqrt{\kappa}d)$ is a CAT(1) space. Similarly for $\kappa < 0$, (X, d) is a CAT(κ) space if and only if $(X, \sqrt{-\kappa}d)$ is a CAT(-1) space.

Let (X, d) be a uniquely geodesic metric space. A subset $C \subset X$ is called *convex* if $[x, y] \subset X$ for all $x, y \in C$. and $f : X \to (-\infty, +\infty]$. Recall that the *effective domain* of f is defined by $\operatorname{dom} f := \{x \in X \mid f(x) < +\infty\}$. If $\operatorname{dom} f \neq \varnothing$, we say that f is proper. Moreover, we say that f is *convex* if

$$f((1-t)x \oplus ty) \leq (1-t)f(x) + tf(y)$$

holds for all $x, y \in X$ and $t \in (0, 1)$.

Definition 2.2. Let X be a CAT(κ) space. A function $f : X \to (-\infty, +\infty]$ is said to be uniformly convex if,

$$f((1-t)x \oplus ty) \leq (1-t)f(x) + tf(y) - t(1-t)\varphi(d(x, y)) \tag{2.1}$$

where φ is a function that is non$-$negative and vanishes only at 0, for all $x, y \in X$ and $t \in [0, 1]$.

The rest of this section is devoted to collect useful lemmas for the future usages in the subsequent sections. The following lemmas are already known in the literature.

Lemma 2.3. *[16] Let $(a_k), (b_k)$ and (c_k) be sequences of nonnegative real numbers such that $a_{k+1} \leq a_k - b_k + c_k$ for all $k \in \mathbb{N}$ and $\sum_{k=1}^{\infty} c_k < +\infty$ .Then (a_k) converges and $\sum_{k=1}^{\infty} b_k < +\infty$.*

Lemma 2.4. *[21] Let X be a CAT(-1) space, $x, y, z \in X$ and $\alpha \in [0, 1]$. Then*

$$\begin{aligned} \cosh d\left(\alpha x \oplus (1-\alpha)y, z\right) \sinh d(x, y) \leq\ & \cosh d(x, z) \sinh \alpha d(x, y) \\ & + \cosh d(y, z) \sinh(1-\alpha)d(x, y). \end{aligned}$$

The next simple inequality is one of the key tools used in our main results.

Lemma 2.5. *For $\kappa = \pm 1$, the inequality*

$$\phi_\kappa(t) \geq \frac{t^2}{2}$$

holds for all $t \in [0, D_\kappa/2)$.

Proof. First, we consider the case $\kappa = 1$. Notice that $\phi_1(t) = \frac{t^2}{2}$ at $t = 0$. Since both functions are right continuous at $t = 0$, it is sufficient to show $\frac{d}{dt}\left(\phi_1(t) - \frac{t^2}{2}\right) \geq 0$ for $t \in (0, \pi/2)$. Let $t \in (0, \pi/2)$, we have

$$\begin{aligned} \frac{d}{dt}\left(\phi_1(t) - \frac{t^2}{2}\right) &= \frac{d}{dt}\left(\tan(t)\sin(t) - \frac{t^2}{2}\right) \\ &= \sin(t) + \tan(t)\sec(t) - t \\ &\geq \tan(t)\sec(t) - t. \end{aligned} \tag{2.2}$$

Also observe that

$$\begin{aligned} \frac{d}{dt}(\tan(t)\sec(t) - t) &= \tan^2(t)\sec(t) + \sec^3(t) - 1 \\ &= \sec(t)\left(1 + 2\tan^2(t)\right) - 1 \\ &\geq 0. \end{aligned}$$

Since $\tan(t)\sec(t) - t$ is right continuous and equals to 0 at $t = 0$, the above inequality yields

$$\tan(t)\sec(t) - t \geq 0.$$

Combining this with (2.2), we obtain $\phi_1(t) \geq \frac{t^2}{2}$ for all $t \in \left[0, \frac{\pi}{2}\right)$.

Next, suppose that $\kappa = -1$. Using similar argument as in the previous case, it suffices to show $\frac{d}{dt}\left(\phi_{-1}(t) - \frac{t^2}{2}\right) \geq 0$ for $t > 0$. Indeed, let $t > 0$, we get $\sinh(t) \geq t$ and consequently

$$\begin{aligned} \frac{d}{dt}\left(\phi_{-1}(t) - \frac{t^2}{2}\right) &= \frac{d}{dt}\left(\tanh(t)\sinh(t) - \frac{t^2}{2}\right) \\ &= \sinh(t)\left(1 + \frac{1}{\cosh^2(t)}\right) - t \\ &\geq 0. \end{aligned}$$

The results are thus proved. ■

In view of Lemma 2.5, we immediately obtain the following conclusion.

Proposition 2.6. *Let X be a complete admissible* CAT(κ) *space and $f : X \to (-\infty, +\infty]$. If f is proper lsc and $K\phi_0-$uniformly convex function with some $K > 0$,then f has a unique minimizer.*

Remark 2.7. In general, if $\kappa > 0$, then f is $K\phi_\kappa-$uniformly convex function if and only if f is $\kappa K\phi_0-$uniformly convex function. Similarly $\kappa < 0$, f is $K\phi_\kappa-$uniformly convex function if and only if f is $-\kappa K\phi_0-$uniformly convex function.

We regard the following lemma as a broader aspect of a statement in [22]. For completeness, we will include the proof as though the idea is pretty much the same.

Lemma 2.8. *Let X be a complete CAT(-1) space, $f : X \to (-\infty, +\infty]$ a proper lsc convex function, and $R_{\lambda f}$ a resolvent of λf for $\lambda > 0$. Then the inequality*

$$\lambda(f(R_{\lambda f}x) - f(y)) \leq \left(\frac{1}{\cosh^2 d(R_{\lambda f}x, x)} + 1\right) \times (\cosh d(x,y) - \cosh d(R_{\lambda f}x, x) \cosh d(R_{\lambda f}x, y))$$

holds for all $x, y \in X$ and $\lambda > 0$.

Proof. Let $\lambda > 0$ and $x, y \in X$. Set $z_t = (1-t)R_{\lambda f}x \oplus ty$ for $t \in (0,1)$ and let $D = d(R_{\lambda f}x, y)$. By the definition of R_λ and the convexity of f, we have

$$\begin{aligned}\lambda f(R_{\lambda f}x) &+ \tanh d(R_{\lambda f}x, x) \sinh d(R_{\lambda f}x, x) \\ &\leq \lambda f(z_t) + \tanh d(z_t, x) \sinh d(z_t, x) \\ &\leq t\lambda f(y) + (1-t)\lambda f(R_{\lambda f}x) + \tanh d(z_t, x) \sinh d(z_t, x)\end{aligned}$$

and hence

$$\begin{aligned}\lambda t(f(R_{\lambda f}x) - f(y)) &\leq \tanh d(z_t, x) \sinh d(z_t, x) - \tanh d(R_{\lambda f}x, x) \sinh d(R_{\lambda f}x, x) \\ &= \left(\frac{1}{\cosh d(R_{\lambda f}x, x) \cosh d(z_t, x)} + 1\right) \\ &\quad \times (\cosh d(z_t, x) - \cosh d(R_{\lambda f}x, x)).\end{aligned}$$

For convenience, put $\mathcal{C}_t := \frac{1}{\cosh d(R_{\lambda f}x, x) \cosh d(z_t, x)} + 1$. Multiply both sides of the above inequality with $(\sinh D)/t$ and apply Lemma 2.4, we get

$$\begin{aligned}&\lambda(f(R_{\lambda f}x) - f(y)) \sinh D \\ &\leq \frac{\mathcal{C}_t}{t}(\cosh d(z_t, x) \sinh D - \cosh d(R_{\lambda f}x, x) \sinh D) \\ &\leq \frac{\mathcal{C}_t}{t}(\cosh d(y, x) \sinh tD - \cosh d(R_{\lambda f}x, x)(\sinh D - \sinh(1-t)D)) \\ &= \frac{\mathcal{C}_t}{t} \cdot 2 \sinh\left(\frac{t}{2}D\right) \\ &\quad \times \left(\cosh d(y, x) \cosh\left(\frac{t}{2}D\right) - \cosh d(R_{\lambda f}x, x) \cosh\left(\left(1 - \frac{t}{2}\right)D\right)\right).\end{aligned}$$

Letting $t \downarrow 0$, we obtain

$$\lambda(f(R_{\lambda f}x) - f(y)) \sinh D \leq \left(\frac{1}{\cosh^2 d(R_{\lambda f}x, x)} + 1\right) D \times (\cosh d(x, y) - \cosh d(R_{\lambda f}x, x) \cosh d(R_{\lambda f}x, y)).$$

Since $(\sinh t)/t > 1$ for $t > 0$, we conclude that

$$\lambda(f(R_{\lambda f}x) - f(y)) \leq \left(\frac{1}{\cosh^2 d(R_{\lambda f}x, x)} + 1\right) \times (\cosh d(x, y) - \cosh d(R_{\lambda f}x, x) \cosh d(R_{\lambda f}x, y)),$$

which proves the lemma. ∎

3. Main Results

In this section, we discuss the main results of this paper. Let us recall that we shall show the convergence of splitting proximal algorithms with curvature-adapted proximals and ultimately complete all the cases of $\kappa \in \mathbb{R}$. Our first main result is considered in the setting of complete locally compact admissible CAT(κ) spaces and the component functions f_i's are proper, convex, and lsc. The second main result then simultaneously relaxes the local compactness condition and induce a convergence estimation by adding uniform convexity requirement on the total cost function.

We are now ready to state and prove our first main result.

Theorem 3.1 (Split Proximal Algorithm: Locally Compact Case). *Let X be a complete locally compact admissible* CAT(κ) *space and for each $i = 1, \dots, N$, let $f_i : X \to (-\infty, +\infty]$ be proper, convex, lsc functions which is L-Lipschitz on* dom(f). *Suppose that $f := \sum_{i=1}^{N} f_i$ has a minimizer. Given a starting point $x^0 \in X$ and a sequence (λ_k) of positive real numbers with $\sum_{k=0}^{\infty} \lambda_k = +\infty$ and $\sum_{k=0}^{\infty} \lambda_k^2 < +\infty$. For any $k \in \mathbb{N} \cup \{0\}$ and $i \in \{1, \dots, N\}$, define*

$$x^{kN+i} := R^{\kappa}_{\lambda_k f_i}(x^{kN+i-1}).$$

Then, (x^n) converges to a minimizer of f.

Proof. As discussed in Section 2, it is sufficient to consider only case $\kappa = 0, 1, -1$. The cases where $\kappa = 0$ and $\kappa = 1$ have beed already proved in [16] and [20], respectively. Let us now show the $\kappa = -1$ case. Let $z \in \arg\min_X f$. By Lemma 2.8, we have

$$\begin{aligned}
&\lambda_k(f_i(x^{kN+i}) - f_i(z)) \\
&\quad\le \left(\frac{1}{\cosh^2 d(x^{kN+i}, x^{kN+i-1})} + 1 \right) \\
&\qquad \times \left(\cosh d(x^{kN+i-1}, z) - \cosh d(x^{kN+i}, x^{kN+i-1}) \cosh d(x^{kN+i}, z) \right).
\end{aligned}$$

Since $0 < t/\sinh t < 1$ and $\cosh t > 1$ for all $t > 0$, the above inequality extends to

$$\begin{aligned}
&\lambda_k(f_i(x^{kN+i}) - f_i(z)) \\
&\qquad \le 2(\cosh d(x^{kN+i-1}, z) - \cosh d(x^{kN+i}, x^{kN+i-1}) \cosh d(x^{kN+i}, z)) \\
&\qquad \le 2(\cosh d(x^{kN+i-1}, z) - \cosh d(x^{kN+i}, z))
\end{aligned}$$

which further implies

$$\lambda_k(f_i(x^{kN+i}) - f_i(z)) \le 2(\cosh d(x^{kN+i-1}, z) - \cosh d(x^{kN+i}, z)). \tag{3.1}$$

By the definition of the proximal operator and Lemma 2.5, we have

$$\begin{aligned}
\lambda_k[f_i(x^{kN+i-1}) - f_i(x^{kN+i})] &\ge \tanh d(x^{kN+i}, x^{kN+i-1}) \sinh d(x^{kN+i}, x^{kN+i-1}) \\
&\ge \frac{d(x^{kN+i}, x^{kN+i-1})^2}{2}.
\end{aligned}$$

Since f_i is L-Lipschitz on dom f_i, the above inequality simplifies to

$$d(x^{kN+i}, x^{kN+i-1}) \le 2L\lambda_k. \tag{3.2}$$

By summing up (3.1) for $i = 1, 2, ..., N$ gives

$$\begin{aligned}\lambda_k(f(x^{kN}) - f(z)) &\leq 2(\cosh d(x^{kN}, z) - \cosh d(x^{kN+N}, z)) \\ &\quad + \lambda_k \sum_{i=1}^{N} (f_i(x^{kN}) - f_i(x^{kN+i})) \\ &\leq 2(\cosh d(x^{kN}, z) - \cosh d(x^{kN+N}, z)) \\ &\quad + \lambda_k L \sum_{i=1}^{N} d(x^{kN}, x^{kN+i}), \end{aligned} \tag{3.3}$$

where the rightmost summand can be further estimated by (3.2) as

$$\begin{aligned} d(x^{kN}, x^{kN+i}) &\leq d(x^{kN}, x^{kN+1}) + d(x^{kN+1}, x^{kN+2}) + ... + d(x^{kN+i-1}, x^{kN+i}) \\ &\leq 2L\lambda_k i. \end{aligned}$$

Substitute the estimate into (3.3), we get

$$\lambda_k(f(x^{kN}) - f(z)) \leq 2(\cosh d(x^{kN}, z) - \cosh d(x^{kN+N}, z)) + N(N+1)L^2{\lambda_k}^2. \tag{3.4}$$

Since z is a minimizer of f and by Lemma 2.3, the sequence $(\cosh d(x^{kN}, z))$ is convergent and so does the sequence $(d(x^{kN}, z))$. Moreover, we get that

$$\sum_{k=0}^{\infty} \lambda_k(f(x^{kN}) - f(z)) < +\infty.$$

Hence there must exists a subsequence $(x^{k_l N})$ of (x^{kN}) such that

$$\lim_{l\to\infty} f(x^{k_l N}) = f(z).$$

Since the sequence $(x^{k_l N})$ is bounded, it possesses a subsequence which converges to a point $\hat{z} \in X$. The lower semicontinuity of f ensures that $\hat{z} \in \arg\min_X f$. By the convergence of $(d(x^{kN}, \hat{z}))$ and of $(x^{k_l N})$ to $\hat{z}$, it must be the case that (x^{kN}) converges also to $\hat{z}$. Finally the fact that (x^{kN+i}) converges to $\hat{z}$ for all fixed $i = 1, \dots, N$ follows from the inequality $d(x^{kN+i}, x^{kN+i-1}) \leq 2L\lambda_k$. This shows that the whole sequence (x^n) is convergent to $\hat{z} \in \arg\min_X f$ and the proof is complete. ∎

In the next main result, we relax the local compactness condition in the ambient space X by imposing uniform convexity assumption on the total cost f. The choice of uniformity is chosen as ϕ_κ, accordingly to the upper curvature bound κ. It turns out that such condition can also strengthen the convergence rate of the splitting proximal algorithm. In this case, we can ensure that the convergence is at least of a sublinear rate.

Before going to our next result, let us prove the following useful lemma. Note that this lemma is drawn from its variant appeared in [15].

Lemma 3.2. *Let $(t_k) \subset (0, 1)$, $\sum_{k=0}^{\infty} t_k = +\infty$, $t_k \to 0$ and $M > 0$. Let $a_0 > 0$ and define $a_{k+1} := (1 - t_k)a_k + t_k^2 M, \forall k \in \mathbb{N} \cup \{0\}$, that is*

$$a_{k+1} := \prod_{i=0}^{k} (1 - t_i)a_0 + M \left[\sum_{j=1}^{k} \left(t_{j-1}^2 \prod_{i=j}^{k} (1 - t_i) \right) + t_k^2 \right], \tag{3.5}$$

where the convention $\sum_{i=m}^{n} b_i = 0$ for $n < m$ is used. Then $a_k \to 0$.

Proof. First, we prove $\liminf_{k\to+\infty} a = 0$ by contradiction. Assume that there are $M \geq 0$ and $c > 0$ such that for every $k > M$, we have $a_k > c$ and $Mt_k < \frac{c}{2}$. Then

$$a_{k+1} = a_k + t_k(Mt_k - a_k) \leq a_k - \frac{ct_k}{2},$$

which is a contradiction, since $\sum_{k=0}^{\infty} t_k = +\infty$.

Next we show that $\lim_{k\to+\infty} = 0$. If $a_k > t_k M$ then clearly $a_{k+1} < a_k$. On the other hand, if we have $a_k \leq t_k M$, then

$$a_{k+1} \leq (1 - t_k)t_k M + t_k^2 M = t_k M$$

and we consequently obtain

$$a_{k+1} \leq \max\{a_k, t_k M\}.$$

For any $\ell \geq k$, it follows that

$$a_{\ell+1} \leq \max\{a_k, M \cdot \max\{t_k, t_{k+1}, ..., t_\ell\}\} \leq \max\left\{a_k, M \cdot \sup_{j\geq k} t_j\right\}.$$

Passing to the limits as $\ell, k \to \infty$, we obtain $a_k \to 0$. ∎

Theorem 3.3 (Split Proximal Algorithm: Uniformly Convex Case)**.** *Let X be a complete admissible* CAT(κ) *space and for each $i = 1, \ldots, N$, let $f_i : X \to (-\infty, +\infty]$ be proper, convex, lsc functions which is L-Lipschitz on* dom(f). *Suppose that $f := \sum_{i=1}^{N} f_i$ is $K\phi_\kappa$-uniformly convex for some $K > 0$. Given a starting point $x^0 \in X$ and a sequence (λ_k) of positive real numbers in the interval $(0, 1/K)$ with $\lambda_k \to 0$ and $\sum_{k=0}^{\infty} \lambda_k = +\infty$. For any $k \in \mathbb{N} \cup \{0\}$ and $i \in \{1, \ldots, N\}$, define*

$$x^{kN+i} := R^{\kappa}_{\lambda_k f_i}(x^{kN+i-1}).$$

Then, (x^n) converges to the unique minimizer of f. Moreover, the following estimates hold:

(i) *If $\kappa = 0$, then $d(x^{kN}, z)^2 \leq a_k$ holds with $a_0 := d(x^0, z)^2$ and $a_{k+1} := (1-\lambda_k K)a_k + 2L^2N(N+1)\lambda_k^2$ inductively, that is,*

$$a_{k+1} := \prod_{i=0}^{k} (1 - \lambda_i K)\, a_0 + 2L^2N(N+1)\left[\sum_{j=1}^{k}\left(\lambda_{j-1}^2 \prod_{i=j}^{k}(1 - \lambda_i K)\right) + \lambda_k^2\right].$$

(ii) *If $\kappa > 0$, then $1 - \cos\sqrt{\kappa}d(x^{kN}, z) \leq a_k$ holds with $a_0 := 1 - \cos\sqrt{\kappa}d(x^0, z)$ and $a_{k+1} := \left(1 - \frac{\lambda_k K}{2\alpha}\right) a_k + \frac{1}{2\alpha\kappa}L^2N(N+1)\lambda_k^2$ where $\alpha = 1 + \frac{1}{\cos^2(4L/\sqrt{\kappa})}$ inductively, that is,*

$$a_{k+1} := \prod_{i=0}^{k}\left(1 - \frac{\lambda_i K}{2\alpha}\right) a_0 + \frac{1}{2\alpha\kappa}L^2N(N+1)\left[\sum_{j=1}^{k}\left(\lambda_{j-1}^2 \prod_{i=j}^{k}\left(1 - \frac{\lambda_i K}{2\alpha}\right)\right) + \lambda_k^2\right].$$

(iii) *If $\kappa < 0$, then $\cosh\sqrt{-\kappa}d(x^{kN}, z) - 1 \leq a_k$ holds with $a_0 := \cosh\sqrt{-\kappa}d(x^0, z) - 1$ and $a_{k+1} := (1 - \frac{\lambda_k K}{2})a_k - \frac{1}{2\kappa}L^2N(N+1)\lambda_k^2$ inductively, that is,*

$$a_{k+1} := \prod_{i=0}^{k}\left(1 - \frac{\lambda_i K}{2}\right) a_0 - \frac{1}{2\kappa}L^2N(N+1)\left[\sum_{j=1}^{k}\left(\lambda_{j-1}^2 \prod_{i=j}^{k}\left(1 - \frac{\lambda_i K}{2}\right)\right) + \lambda_k^2\right].$$

Proof. Again, it is sufficient to prove the results for $\kappa = 0, 1, -1$. To retrieve the meaningful estimates exactly as stated in the theorem, note that the Lipschitz constant L changes to $L/\sqrt{\kappa}$ (resp. $L/\sqrt{-\kappa}$) after rescaling the distance $d(\cdot,\cdot)$ with $\sqrt{\kappa}d(\cdot,\cdot)$ for $\kappa > 0$ (resp. $\sqrt{-\kappa}d(\cdot,\cdot)$ for $\kappa < 0$. The results where $\kappa = 0$ was already proved in [15]. By Proposition 2.6 and Remark 2.7, we get f has a unique minimizer. We now prove the remaining cases where $\kappa = 1$ and $\kappa = -1$, respectively. In each cases, we need to show two assertions, i.e. the convergence of the whole sequence (x^n) and the estimates given in terms of (a_k).

Let $\kappa = 1$. By the uniformly convexity of f and the completeness of X, f has a unique minimizer $z \in X$. For any $x \in X$, by dividing (2.1) with $1-t$ and letting $t \to 1$, we get

$$K \tan d(x, y) \sin d(x, y) \leq f(x) - f(y). \tag{3.6}$$

Following the method in the main result of [20] and applying Lemma 2.5, one may obtain the inequality

$$-\cos d(x^{kN+N}, z) \leq -\cos d(x^{kN}, z) - \frac{\lambda_k}{2\alpha}(f(x^{kN}) - f(z)) + \frac{1}{2\alpha}N(N+1)L^2{\lambda_k}^2.$$

Applying (3.6) in the above inequality yields

$$\begin{aligned} -\cos d(x^{kN+N}, z) \leq & -\cos d(x^{kN}, z) - \frac{\lambda_k K}{2\alpha} \tan d(x^{kN}, z) \sin d(x^{kN}, z) \\ & + \frac{1}{2\alpha}N(N+1)L^2{\lambda_k}^2 \\ = & -\cos d(x^{kN}, z) - \frac{\lambda_k K}{2\alpha}\left(\frac{1}{\cos d(x^{kN}, z)} - \cos d(x^{kN}, z)\right) \\ & + \frac{1}{2\alpha}N(N+1)L^2\lambda_k^2 \end{aligned}$$

and so we get

$$1 - \cos d(x^{kN+N}, z) \leq \left(1 - \frac{\lambda_k K}{2\alpha}\right)(1 - \cos d(x^{kN}, z)) + \frac{1}{2\alpha}N(N+1)L^2\lambda_k^2.$$

This shows the estimate $1 - \cos\sqrt{\kappa}d(x^{kN}, z) \leq a_k$ for each $k \in \mathbb{N}$. By using mathematical induction and Lemma 3.2, we have (a_k) converges to 0. This proves the convergence of (x^{kN}) to z.

Now, let $i \in \{1, \ldots N\}$. We have from the definition of the proximal operator and Lemma 2.5 that

$$\begin{aligned} \lambda_k[f_i(x^{kN+i-1}) - f_i(x^{kN+i})] &\geq \tan d(x^{kN+i-1}, x^{kN+i}) \sin d(x^{kN+i-1}, x^{kN+i}) \\ &\geq \frac{d(x^{kN+i-1}, x^{kN+i})^2}{2}. \end{aligned}$$

Since f_i is L-Lipschitz,

$$d(x^{kN+i-1}, x^{kN+i}) \leq 2\lambda_k L.$$

This ensures the convergence of the whole sequence (x^n) to $z = \arg\min_X f$.

Next, let $\kappa = -1$. We shall proceed with similar strategy witht he other case above. Again, by the uniformly convexity of f and the completeness of X, f has a unique minimizer $z \in X$. Let $x \in X$ and dividing (2.1) by $1-t$ and letting $t \to 1$, we have

$$K \tanh d(x, y) \sinh d(x, y) \leq f(x) - f(y). \tag{3.7}$$

By (3.4) in Theorem 3.1, we have

$$\cosh d(x^{kN+N}, z) \leq \cosh d(x^{kN}, z) - \frac{\lambda_k}{2}(f(x^{kN}) - f(z)) + \frac{1}{2}N(N+1)L^2{\lambda_k}^2.$$

In view of (3.7), the above inequality expands to

$$\begin{aligned}\cosh d(x^{kN+N}, z) \leq\ & \cosh d(x^{kN}, z) - \frac{\lambda_k K}{2}\tanh d(x^{kN}, z)\sinh d(x^{kN}, z) \\ & + \frac{1}{2}N(N+1)L^2{\lambda_k}^2. \\ =\ & \cosh d(x^{kN}, z) - \frac{\lambda_k K}{2}\left(\cosh d(x^{kN}, z) - \frac{1}{\cosh d(x^{kN}, z)}\right) \\ & + \frac{1}{2}N(N+1)L^2\lambda_k^2\end{aligned}$$

and we may further obtain

$$\cosh d(x^{kN+N}, z) - 1 \leq \left(1 - \frac{\lambda_k K}{2}\right)(\cosh d(x^{kN}, z) - 1) + \frac{1}{2}N(N+1)L^2\lambda_k^2. \tag{3.8}$$

Hence, the estimate $\cosh\sqrt{-\kappa}d(x^{kN}, z) - 1 \leq a_k$ holds for all $k \in \mathbb{N}$. By mathematical induction and Lemma 3.2, (a_k) converges to 0 so that (x^{kN}) is convergent to z. The convergence of the whole sequence (x^n) to z follows by setting up a similar inequality of the form (3.2). ■

4. Applications

In this section, we consider some viable applications of our main results. Particularly, we shall apply our results from the previous section to solve feasibility and centroid problems. Let us clarify the general ideas behind such problems now. Given nonempty subsets $C_1, \dots, C_N$ in a CAT(κ) space X. The feasibility problem seeks a common point $\overline{x}$ in the intersection $\bigcap_{i=1}^N C_i$. While this intersection is nonempty, solving the feasibility problem is not impossible. On one hand, if $\bigcap_{i=1}^N C_i$ is empty, then one may wish to obtain some negotiation of being reasonably proximal to each C_i. In this case, we consider the centroid problem which finds a center of gravity (in the barycentric sense) between the sets C_i's. On the other hand, the centroid problem with each of C_i's being singleton reduces to the well-known Karcher mean which is greatly useful in the modern theory of information and probability.

The two problems above have a long literature in linear space settings and we would not make a history record here. Let us mention in the scope of CAT(κ) spaces that Bačák et al. [23] introduced the alternating projection method for intersection of two sets, i.e. $N = 2$, in CAT(0) spaces and later extended to CAT(κ) by Choi et al. [24]. On the other hand, the problem of finding a Karcher means of finitely many points by means of the proximal algorithms have been studied in CAT(0) spaces by Bačák [14]. However, the exploration of Karcher means in CAT(κ) spaces began as early as 1990 by Kendall [25] and still in development as in a recent discussion by Kobayashi and Wynn [26]. To the best of our knowledge, there has not been a study of Karcher means with respect to proximal operators as being considered in this paper. Moreover, the extension to center of gravity between sets is still rare.

4.1. Convex Feasibility Problem

In this subsection, we consider the feasibility problem with each C_i's being closed and convex. We call this particular case the convex feasibility problem. Recall that $\overline{x} \in \bigcap_{i=1}^{N} C_i$ if and only if $\sum_{i=1}^{N} \delta_{C_i}(\overline{x}) = 0$, where for a given set $\Omega \subset X$, the function $\delta_\Omega : X \to (-\infty, +\infty]$ is defined by

$$\delta_\Omega(x) := \begin{cases} 0 & \text{if } x \in \Omega, \\ +\infty & \text{otherwise,} \end{cases}$$

for each $x \in X$. The function δ_Ω is lsc if Ω is closed, and is convex if Ω is convex. In fact, when Ω is closed, then δ_Ω is Lipschitz continuous on its effective domain $\operatorname{dom} \delta_\Omega$. If Ω is closed and convex, its proximal $R^{\kappa}_{\delta_\Omega}$ reduces to the metric projection $P_\Omega : X \to \Omega$ given by

$$P_\Omega(x) := \operatorname*{arg\,min}_{y \in \Omega} d(x, y)$$

for all $x \in X$.

With the observations above, we can therefore solve the convex feasibility associated to $C_1, \ldots, C_N$ by minimizing $\sum_{i=1}^{N} \delta_{C_i}$. Hence by letting $f_i := \delta_{C_i}$, the splitting proximal algorithm boils down to the alternating projection method and we obtain the following consequence. This shows that the results in [23] and [24] can be recovered from our main results.

Corollary 4.1 (Alternating Projection Method)**.** *Let (X, d) be a locally compact complete admissible* CAT(κ) *space for some $\kappa \in \mathbb{R}$. If $C_1, \ldots, C_N$ are nonempty closed convex subsets of X whose intersection $\bigcap_{i=1}^{N}$ is nonempty. Then for any initial point $x^0 \in X$, the sequence defined for each $k \in \mathbb{N}$ and $i = 1, \ldots, N$ by*

$$x^{kN+i} := P_{C_i}(x^{kN+i-1})$$

is convergent to an element $\overline{x}$ in $\bigcap_{i=1}^{N} C_i$.

Proof. Letting $f_i := \delta_{C_i}$ for all $i = 1, \ldots, N$ and put $f := \sum_{i=1}^{N} f_i$. Then all f_i's are Lipschitz continuous, proper, and convex and f has a minimizer. Moreover, we have $R^{\kappa}_{f_i} = P_{C_i}$ for each $i = 1, \ldots, N$. The remaining conclusion follows from Theorem 3.1 with any choice of (λ_k). ∎

4.2. Center of Gravity and Karcher Mean

Again, suppose that $C_1, \ldots, C_N$ are nonempty closed convex subsets of a complete CAT(κ) space (X, d). Here, we do not assume that their intersection is nonempty and at the same time reject to strictly assume on the contrary. In the case where $\bigcap_{i=1}^{N} C_i$ is empty, we still allow for some of them to intersect others.

In the previous subsection, we reformulate the sets C_i's with their indicator functions δ_{C_i}'s. This method works very well in the feasibility case but fails to generalize or tolerate in infeasible problems. For this, we replace the indicator δ_Ω (given $\Omega \subset X$) with the point-to-set distance $d(\cdot, \Omega)$ given by

$$d(x, \Omega) = \inf_{y \in \Omega} d(x, y)$$

for each $x \in X$. The value of $d(x, \Omega$ remains 0 if $x \in \operatorname{cl} \Omega$ (cl denotes the closure operator), but replaces $+\infty$ of the indicator with a real value.

Suppose that $C_1,\ldots,C_N$ are closed convex sets in X and then assign positive weights $w_1,\ldots,w_N$ respectively on each of the sets. A point $\overline{x}\in X$ is called a *center of gravity* with weight vector $\mathbf{w}:=(w_1,\ldots,w_N)^\top$ between $C_1,\ldots,C_N$ if it minimizes $\sum_{i=1}^N w_i d(\cdot,C_i)^2$. In this case, we write $\overline{x}\in\mu(C_1,\ldots,C_N;w_1,\ldots,w_N)$. If $\mu(C_1,\ldots,C_N;w_1,\ldots,w_N)$ is singleton, we may replace the inclusion "$\in$" with an equality "$=$". When $\bigcap_{i=1}^N C_i$ is nonempty and each of the sets C_i's is assigned equal weight, then every point in the intersection is a center of gravity. Let us make a remark that even if there exist $j,j'\in\{1,\ldots,N\}$ with $C_j\cap C_{j'}\neq\varnothing$, it is unnecessary to have the center of gravity lying inside such intersection.

For $i=1,\ldots,N$, let $f_i:=w_i d(\cdot,C_i)^2$ and set $f=\sum_{i=1}^N f_i$. We may now apply the proximal operator to f_i's. If $\lambda>0$ and $x\in X$, then

$$R^\kappa_{\lambda f_i}(x)=\operatorname*{arg\,min}_{y\in X}\left[w_i d(y,C_i)^2+\frac{1}{\lambda}\phi_\kappa(d(y,x))\right]=:\mu^\kappa(C_i,x;w_i,1/\lambda)$$

can be thought of as a center of gravity with nonlinear weight ϕ_κ acting on the point x. The splitting proximal algorithm is then to solve alternatively for a center of gravity of a single point and one of C_i's, each one at a time. Let us conclude as follows.

Corollary 4.2 (Approximating Center of Gravity). *Let (X,d) be a compact admissible* CAT(κ) *space, $C_1,\ldots,C_N$ are nonempty closed convex sets in X, and $w_1,\ldots,w_N$ are positive reals. If $x^0\in X$ and for each $k\in\mathbb{N}$ and $i=1,\ldots,N$, we define x^{kN+i} by*

$$x^{kN+i}:=\mu^\kappa(C_i,x^{kN+i-1};w_i,1/\lambda_k),$$

where (λ_k) is a sequence of positive reals with $\sum_{i=1}^\infty\lambda_i=+\infty$ and $\sum_{i=1}^\infty\lambda_i^2<+\infty$. Then, the sequence (x^n) is convergent to a center of gravity of $C_1,\ldots,C_N$ with weights $w_1,\ldots,w_N$, respectively.

Proof. For $i=1,\ldots,N$, let $f_i:=w_i d(\cdot,C_i)^2$ and set $f=\sum_{i=1}^N f_i$. Each f_i is proper, convex and Lipschitz continuous. Apply Theorem 3.1 to conclude the desired convergence. ∎

Next, we consider the case where all C_i's are singleton. Let $G:=\{z_1,\ldots,z_N\}$ be a finte subset of a complete CAT(κ) space. Suppose that $w_1,\ldots,w_N$ are positive reals. Then we define the Karcher mean of G with weight vector $\mathbf{w}:=(w_1,\ldots,w_N)^\top$ to be

$$\mu(G;w_1,\ldots,w_N):=\operatorname*{arg\,min}_{x\in X}\sum_{i=1}^N w_i d(x,z_i)^2.$$

The Karcher mean was considered in [14] under the setting of complete CAT(0) spaces. We generalize this result as a special case of the previous corollary with each C_i's being singletone, we state the following straightforward consequence without a proof. Note that the uniquely existence of a Karcher mean follows from the strong convexity of the objective function when $diam(X)<D_\kappa/2$ (see [27]).

Corollary 4.3 (Approximating Karcher Mean). *Let (X,d) be a compact admissible* CAT(κ) *space, $G:=\{z_1,\ldots,z_N\}$ a finite subset in X, and $w_1,\ldots,w_N$ are positive reals. If $x^0\in X$ and for each $k\in\mathbb{N}$ and $i=1,\ldots,N$, we define x^{kN+i} by*

$$x^{kN+i}:=\mu^\kappa(z_i,x^{kN+i-1};w_i,1/\lambda_k),$$

where (λ_k) is a sequence of positive reals with $\sum_{i=1}^\infty\lambda_i=+\infty$ and $\sum_{i=1}^\infty\lambda_i^2<+\infty$. Then, the sequence (x^n) is convergent to $\mu(G;w_1,\ldots,w_N)$.

5. Numerical Implementations

In this section, we implement the proposed splitting proximal algorithm to approximate the Karcher mean of given datasets which are fitted to the Lobachevskii plane $\mathbb{H}^2$ defined by

$$\mathbb{H}^2 := \{\mathbf{x} = (x_1, x_2, x_3) \in \mathbb{R}^3 \mid \langle \mathbf{x}|\mathbf{x}\rangle = -1\},$$

where $\langle \cdot|\cdot \rangle : \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}$ is the Lorentzian product given by

$$\langle \mathbf{x}|\mathbf{y}\rangle := \mathbf{x}^\top \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix} \mathbf{y}$$

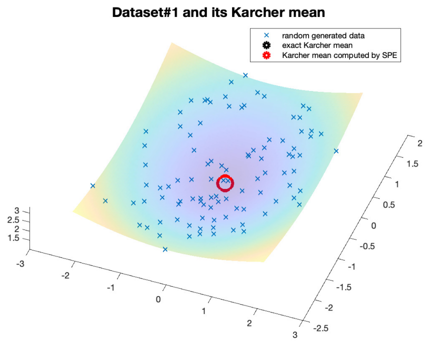
for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^3$. Recall that $\mathbb{H}^2$ is a complete CAT(-1) space when equipped with the distance function d given by

$$\cosh d(\mathbf{x}, \mathbf{y}) := -\langle \mathbf{x}|\mathbf{y}\rangle$$

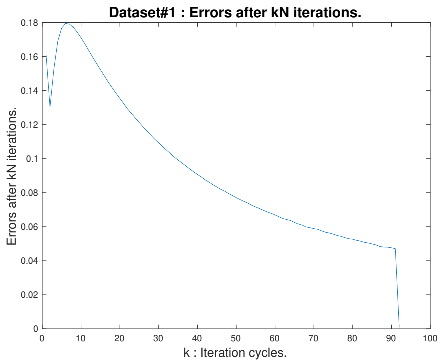
for $\mathbf{x}, \mathbf{y} \in \mathbb{H}^2$.

We use seven different datasets $G_1, \dots, G_7$ which are randomly generated according to the Gaussian distribution, with the same cardinality $|G_i| = N = 100$ for all $i = 1, \dots, 7$. The Karcher means to be computed in this section are given equal weights on each of the points in G_i's. In each of the numerical implementations, we set the start point to $\mathbf{x} = (5, 5, \sqrt{51}) \in \mathbb{H}^2$ in order avoid bias from being already in the cluster of data. In the following numerical implementations of Corollary 4.3, we let $X \subset \mathbb{H}^2$ be compact set containing $G_1, ..., G_7$ and x. The algorithm is implemented with the commercial software `Matlab R2020a` under the Campus-wide license of King Mongkut's University of Technology Thonburi. The hardware is running on the operating system `MacOS Catalina` with a Processor `1.4 GHz Quad-Core Intel Core i5` and a `8 GB 2133 MHz LPDDR3` Memory.

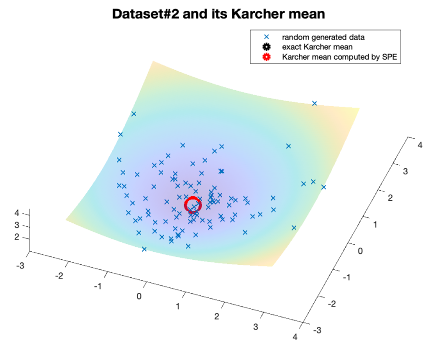
Since the dataset is quite large, we only report the implementation results as presented in Figure 1. Therein, the LHS figures represent the data visualization of G_i (blue 'x' marks) together with the exact Karcher mean (black 'O' marks) and approximated Karcher mean (red 'O' marks) obtained from our Splitting Proximal Algorithm (SPA). The RHS figures show plots of errors after the iteration cycle of kN. Note that the errors presented thereby are computed by the geodesic distance between the actual Karcher mean and the one approximated the SPA. The accepted tolerance in these illustrations is set at tol $= 10^{-5}$.



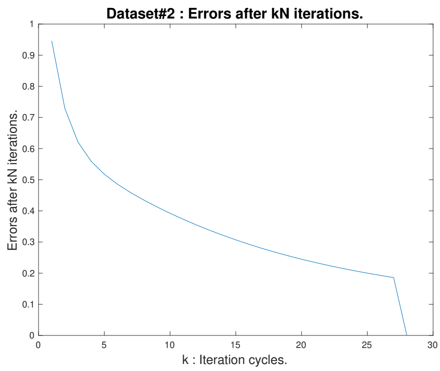
(A) The dataset G_1, its Karcher mean and an approximated one.



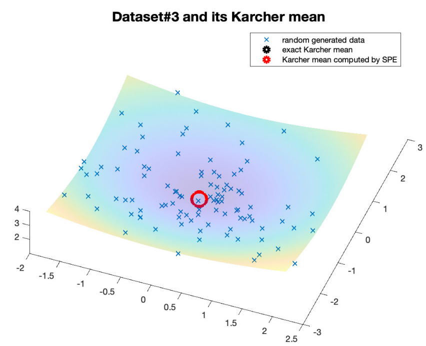
(B) Distance from x^{kN} to the Karcher mean of G_1.



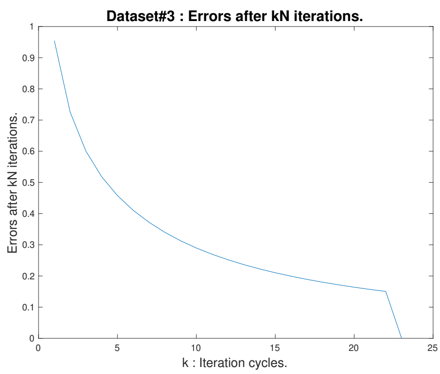
(C) The dataset G_2, its Karcher mean and an approximated one.



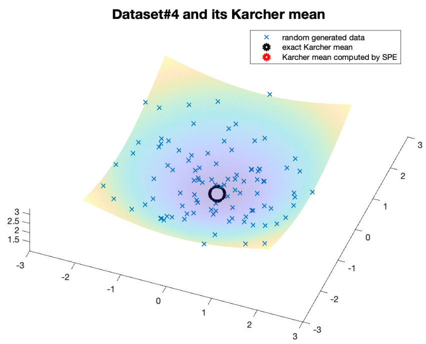
(D) Distance from x^{kN} to the Karcher mean of G_2.



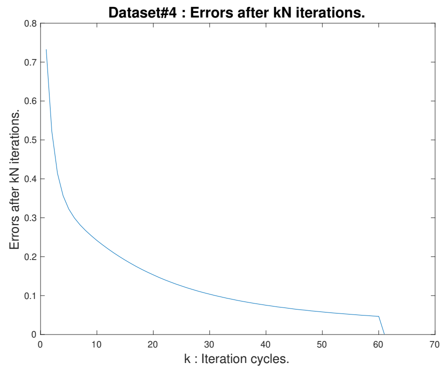
(E) The dataset G_3, its Karcher mean and an approximated one.



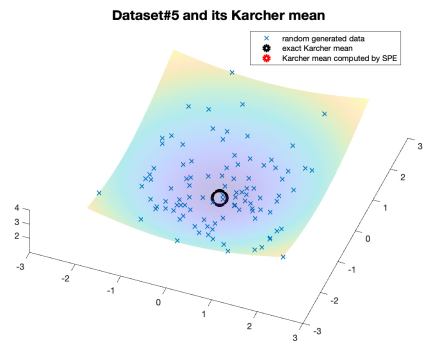
(F) Distance from x^{kN} to the Karcher mean of G_3.



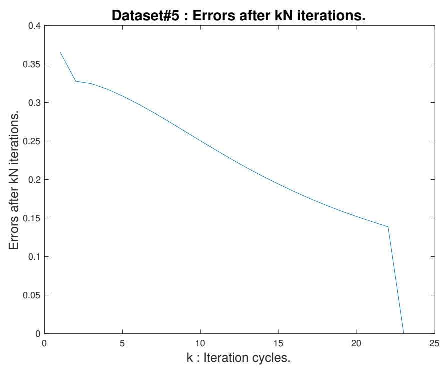
(G) The dataset G_4, its Karcher mean and an approximated one.



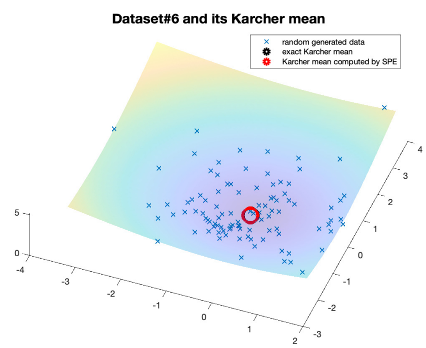
(H) Distance from x^{kN} to the Karcher mean of G_4.



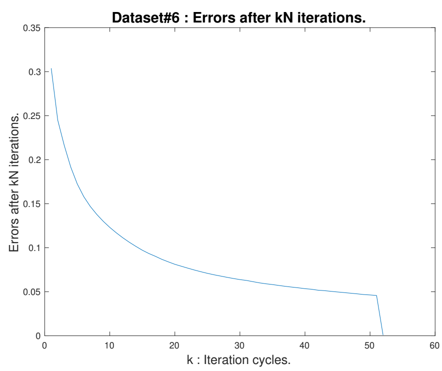
(I) The dataset G_5, its Karcher mean and an approximated one.



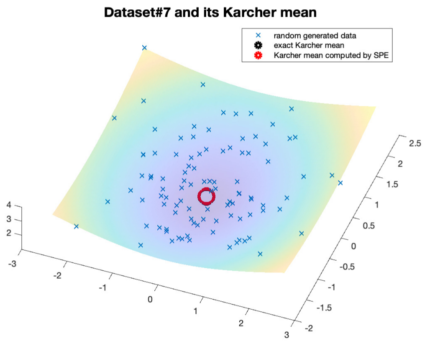
(J) Distance from x^{kN} to the Karcher mean of G_5.



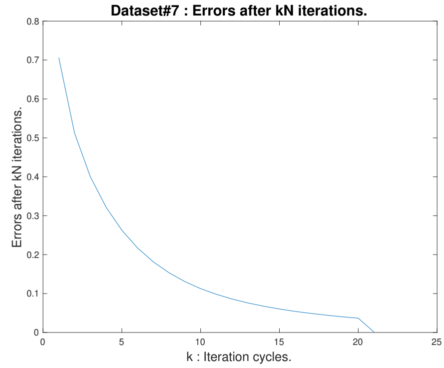
(K) The dataset G_6, its Karcher mean and an approximated one.



(L) Distance from x^{kN} to the Karcher mean of G_6.



(M) The dataset G_7, its Karcher mean and an approximated one.



(N) Distance from x^{kN} to the Karcher mean of G_7.

FIGURE 1. Results of the implementation of SPA to approximate the Karcher means of seven random datasets $G_1, \ldots, G_7$.

As one may notice, the SPA is practically capable of solving Karcher means with significance of 5th decimal place within a reasonable iteration loops. Recall that at each step of SPA, we advance my moving along the geodesic of the current iteration x^n and one of the data point z_i. The *dives* appeared in each error plots near the termination corresponds to when the iteration x^n successfully breached through the large variations caused by scattered data points and finally gets inside the cluster of data.

Acknowledgements

The authors acknowledge the financial support provided by the Center of Excellence in Theoretical and Computational Science (TaCS-CoE), KMUTT. The first author was supported by the Petchra Pra Jom Klao M.Sc. Research Scholarship from King Mongkut's University of Technology Thonburi (Grant No.37/2561).

Aruthor Contributions

Supervision, Poom Kumam; Writing - original draft, Sakan Termkaew; Writing - review & editing, Parin Chaipunya.

References

[1] J.-J. Moreau, "Propriétés des applications "prox"," *C. R. Acad. Sci. Paris*, vol. 256, pp. 1069–1071, 1963.
[2] R. T. Rockafellar, "Monotone operators and the proximal point algorithm," *SIAM J. Control Optimization*, vol. 14, no. 5, pp. 877–898, 1976.
[3] P. L. Combettes and V. R. Wajs, "Signal recovery by proximal forward-backward splitting," *Multiscale Modeling & Simulation*, vol. 4, pp. 1168–1200, jan 2005.
[4] N. Nimana, A. P. Farajzadeh, and N. Petrot, "Adaptive subgradient method for the split quasi-convex feasibility problems," *Optimization*, vol. 65, no. 10, pp. 1885–1898, 2016.

[5] M. Razaviyayn, M. Hong, and Z.-Q. Luo, "A unified convergence analysis of block successive minimization methods for nonsmooth optimization," *SIAM J. Optim.*, vol. 23, no. 2, pp. 1126–1153, 2013.
[6] B. O'Donoghue, E. Chu, N. Parikh, and S. Boyd, "Conic optimization via operator splitting and homogeneous self-dual embedding," *J. Optim. Theory Appl.*, vol. 169, no. 3, pp. 1042–1068, 2016.
[7] B. C. Vũ, "A splitting algorithm for dual monotone inclusions involving cocoercive operators," *Adv. Comput. Math.*, vol. 38, no. 3, pp. 667–681, 2013.
[8] P. Chuasuk, A. Farajzadeh and A. Kaewcharoen, "An iterative algorithm for solving split feasibility problems and fixedpoint problems in p-uniformly convex and smooth banach spaces," *J. Computational Analysis and Applications*, vol. 28, no. 1, pp. 49–66, 2020.
[9] P. L. Combettes and J.-C. Pesquet, "Proximal splitting methods in signal processing," in *Fixed-point algorithms for inverse problems in science and engineering*, vol. 49 of *Springer Optim. Appl.*, pp. 185–212, Springer, New York, 2011.
[10] J. Jost, "Convex functionals and generalized harmonic maps into spaces of nonpositive curvature," *Comment. Math. Helv.*, vol. 70, no. 4, pp. 659–673, 1995.
[11] U. F. Mayer, "Gradient flows on nonpositively curved metric spaces and harmonic maps," *Comm. Anal. Geom.*, vol. 6, no. 2, pp. 199–253, 1998.
[12] M. Bačák, "The proximal point algorithm in metric spaces," *Israel J. Math.*, vol. 194, no. 2, pp. 689–701, 2013.
[13] P. Chaipunya, "Variational inequalities under the global npc condition," *New Trends in Analysis and Geometry*, p. 165, 2020.
[14] M. Bačák, "Computing medians and means in Hadamard spaces," *SIAM J. Optim.*, vol. 24, no. 3, pp. 1542–1566, 2014.
[15] S.-i. Ohta and M. Pálfia, "Discrete-time gradient flows and law of large numbers in Alexandrov spaces," *Calc. Var. Partial Differential Equations*, vol. 54, no. 2, pp. 1591–1610, 2015.
[16] M. Bačák, *Convex analysis and optimization in Hadamard spaces*, vol. 22 of *De Gruyter Series in Nonlinear Analysis and Applications.* De Gruyter, Berlin, 2014.
[17] Y. Kimura and F. Kohsaka, "Spherical nonspreadingness of resolvents of convex functions in geodesic spaces," *J. Fixed Point Theory Appl.*, vol. 18, no. 1, pp. 93–115, 2016.
[18] T. Kajimura and Y. Kimura, "Resolvents of convex functions in complete geodesic metric spaces with negative curvature," *J. Fixed Point Theory Appl.*, vol. 21, no. 1, pp. Paper No. 32, 15, 2019.
[19] Y. Kimura and F. Kohsaka, "Two modified proximal point algorithms for convex functions in Hadamard spaces," *Linear Nonlinear Anal.*, vol. 2, no. 1, pp. 69–86, 2016.
[20] R. Espínola and A. Nicolae, "Proximal minimization in CAT(κ) spaces," *J. Nonlinear Convex Anal.*, vol. 17, no. 11, pp. 2329–2338, 2016.
[21] M. R. Bridson and A. Haefliger, *Metric spaces of non-positive curvature*, vol. 319 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences].* Springer-Verlag, Berlin, 1999.
[22] T. Kajimura and Y. Kimura, "The proximal point algorithm in complete geodesic spaces with negative curvature," *Advances in the Theory of Nonlinear Analysis and its Application*, pp. 162–170, Oct 2019.

[23] M. Bačák, I. Searston, and B. Sims, "Alternating projections in CAT(0) spaces," *J. Math. Anal. Appl.*, vol. 385, no. 2, pp. 599–607, 2012.

[24] B. J. Choi, U. C. Ji, and Y. Lim, "Convergences of alternating projections in CAT(κ) spaces," *Constr. Approx.*, vol. 47, no. 3, pp. 391–405, 2018.

[25] W. S. Kendall, "Probability, convexity, and harmonic maps with small image. I. Uniqueness and fine existence," *Proc. London Math. Soc. (3)*, vol. 61, no. 2, pp. 371–406, 1990.

[26] K. Kobayashi and H. P. Wynn, "Empirical geodesic graphs and CAT(k) metrics for data analysis," *Stat. Comput.*, vol. 30, no. 1, pp. 1–18, 2020.

[27] S.-i. Ohta, "Convexities of metric spaces," *Geom. Dedicata*, vol. 125, pp. 225–250, 2007.