ON THE DENSITY THEOREM FOR THE SUBDIFFERENTIAL OF CONVEX FUNCTIONS ON HADAMARD SPACES

M. MOVAHEDI, D. BEHMARDI, S. HOSSEINI

Abstract. In this paper, a dual space for a geodesically complete Hadamard space is introduced. By using this notion we present a new definition of the subdifferential of convex functions on geodesically complete Hadamard spaces. Moreover, some properties of this subdifferential such as a density theorem are proved.

1. introduction

Nondifferentiability appears naturally in different areas of mathematics and arises explicitly in the description of various modern technological systems. Nonsmooth analysis studies the local behavior of nondifferentiable functions and sets lacking smooth boundaries. Generalized gradients or subdifferentials refer to several set-valued replacements for the usual derivative which are used in developing differential calculus for nonsmooth functions.

Nondifferentiable functions are often considered on finite-dimensional or infinite dimensional Banach spaces, where the linear structure plays a central role. Recently, attempts have been made to replace Banach spaces with Riemannian

---

Key words and phrases. Hadamard spaces, Convex functions, Subdifferential.

M. Movahedi

Department of Mathematics, Faculty of Sciences, Alzahra University, Tehran, Iran.

E-mail:m.movahedi@alzahra.ac.ir.com

D. Behmardi

Department of Mathematics, Faculty of Sciences, Alzahra University, Tehran, Iran.

E-mail:d2beh@yahoo.com, behmardi@alzahra.ac.ir

S. Hosseini

Seminar of Applied Mathematics, ETH, Zürich, Switzerland.

E-mail: seyedehsomayeh.hosseini@math.ethz.ch

2010 Mathematics Subject Classification: 53C23, 47H05, 47H07.

manifolds and develop a subdifferential calculus; see [9, 10, 11, 12] and the references therein. It is also worth mentioning paper [15], which presents a definition of the coaccretive subdifferential of a convex function defined on a Hilbert ball. The approach in that paper involves the structure of (B, ρ) as a Hilbert manifold, where ρ is the hyperbolic metric on B; see also [13, page 188].

Unlike Riemannian manifolds, Hadamard spaces are not equipped with a Riemannian metric. Hence, we need new tools to construct a suitable dual space in order to define subdifferential of functions on Hadamard spaces. In 2010, Ahmadi-Kakavandi and Amini in [1] defined a dual space for an Hadamard space using the concept of bound vectors. They defined a pseudometric D on $\mathbb{R} \times \mathcal{X} \times \mathcal{X}$, where $\mathcal{X}$ is an Hadamard space, and considered the pseudometric space $(\mathbb{R} \times \mathcal{X} \times \mathcal{X}, D)$, as a subspace of the pseudometric space $(\mathrm{Lip}(\mathcal{X}, \mathbb{R}), L)$ of all real-valued Lipschitz functions. Then, they defined an equivalence relation on $\mathbb{R} \times \mathcal{X} \times \mathcal{X}$, where the equivalence class of (t, a, b) is

$$[t\overrightarrow{ab}] := \{s\overrightarrow{cd}; \quad t\langle \overrightarrow{ab}, \overrightarrow{xy}\rangle = s\langle \overrightarrow{cd}, \overrightarrow{xy}\rangle \quad (x, y \in \mathcal{X})\}.$$

The dual metric space of $\mathcal{X}$ presented in [1] is as follows,

$$\mathcal{X}^* := \{[t\overrightarrow{ab}]; \quad (t, a, b) \in \mathbb{R} \times \mathcal{X} \times \mathcal{X}\},$$

moreover in that paper a notion of the subdifferential for a proper function on an Hadamard space is defined.

The aim of this paper is to present a new dual for any Hadamard space and our most important result is to prove the density theorem for the subdifferential of a lower semicontinuous convex function on an Hadamard space, which is a generalization of the classical one on Hilbert spaces; see [6]. Our approach is different from the one used in [1] and exploits the notion of geodesics to define the dual space. Indeed, we define $\mathcal{X}^*$ as the disjoint union of the sets $\mathcal{X}^*_x$, where $x \in \mathcal{X}$ and $\mathcal{X}^*_x$ contains all unit speed geodesics of $\mathcal{X}$ starting at x. Moreover, the subdifferential of a function f at a point x is defined as a subset of $\mathcal{X}^*_x$, however this property is not visible in the definition of the subdifferential in [1]. Consequently, this leads us to the claim that the subdifferential of convex functions defined in this paper is an analogue of the concept of the subdifferential of convex functions in Riemannian manifolds and the Hilbert balls.

We assume that $\mathcal{X}$ is a geodesically complete Hadamard space with a metric d. Recall that a geodesic in $\mathcal{X}$ is a curve of constant speed which is locally minimizing. We say $\mathcal{X}$ has non-positive curvature (in the sense of Alexandrov) if every point $p \in \mathcal{X}$ has a neighborhood U with the following properties:
(i) for any two points $x, y \in U$ there is a geodesic $\sigma_x^y : [0,1] \to U$ from x to y of length $d(x,y)$,
(ii) for any triple of points $x, y, z \in U$, we have

$$d^2(z,m) \leq \frac{1}{2}(d^2(z,x) + d^2(z,y)) - \frac{1}{4}d^2(x,y),$$

where σ_x^y is as (i) and $m = \sigma_x^y(1/2)$ is the middle point between x and y.

We say $\mathcal{X}$ is an Hadamard space if $\mathcal{X}$ is complete and the assertions (i) and (ii) above hold for all points $x, y, z \in \mathcal{X}$. Hadamard spaces are uniquely geodesic, i.e., there exists a unique geodesic between any pair of points.

In this paper, we assume that $\mathcal{X}$ is a geodesically complete Hadamard space, that is, every geodesic in $\mathcal{X}$ is a subarc of a geodesic which is parameterized on the whole real line. Let $\mathbb{E}^2$ be the Euclidean space equipped with the metric

$$d_{\mathbb{E}^2}((x_1,x_2),(y_1,y_2)) = ((x_1-y_1)^2 + (x_2-y_2)^2)^{1/2}.$$

A geodesic triangle $\triangle(x,y,z)$ in $\mathcal{X}$ is the union of three points $x, y, z \in \mathcal{X}$ and the geodesic segments joining them. The comparison triangle for $\triangle(x,y,z)$, is a triangle $\triangle(\bar{x},\bar{y},\bar{z})$ in $\mathbb{E}^2$ such that $d(x,y) = d_{\mathbb{E}^2}(\bar{x},\bar{y})$, $d(x,z) = d_{\mathbb{E}^2}(\bar{x},\bar{z})$ and $d(z,y) = d_{\mathbb{E}^2}(\bar{z},\bar{y})$. According to this notation, if a is a point on the geodesic segment joining x, y, then $\bar{a}$ is its comparison point provided that $d(x,a) = d_{\mathbb{E}^2}(\bar{x},\bar{a})$. Also the comparison angle $\angle_{\bar{x}}(\bar{y},\bar{z})$, is the interior angle of the comparison triangle $\triangle(\bar{x},\bar{y},\bar{z})$ at $\bar{x}$.

The first step in defining a subdifferential for a function defined on an Hadamard space $\mathcal{X}$ is to introduce a dual space $\mathcal{X}^*$ for $\mathcal{X}$.

We denote by $\mathcal{X}^*$ the set of all unit speed geodesics of $\mathcal{X}$. In other words, $\mathcal{X}^* = \coprod_{x\in\mathcal{X}} \mathcal{X}_x^*$ where $\mathcal{X}_x^*$ is the set of all unit speed geodesics of $\mathcal{X}$ starting at x.

Consider the map $\langle\ ,\ \rangle : \mathcal{X}_x^* \times \mathcal{X}_x^* \to \mathbb{R}$ defined by

$$\langle \gamma_x^y, \gamma_x^z \rangle = \frac{1}{2}[d^2(x,z) + d^2(x,y) - d^2(y,z)].$$

It is clear that $(\langle \gamma_x^y, \gamma_x^y \rangle)^{(1/2)} = d(x,y)$, for more details see [3].

Let $\gamma_x^y \in X_x^*$, $\sigma_z^w \in X_z^*$ and $D := \mathrm{dom}(\sigma_z^w) = \mathrm{dom}(\gamma_x^y)$. Then we say γ_x^y is parallel to σ_z^w if there exists $C \in \mathbb{R}$ such that $d(\sigma_z^w(t), \gamma_x^y(t)) = C$, for all $t \in D$.

2. The subdifferential of a convex function

In this section, we present a new definition of the subdifferential of a convex function on an Hadamard space. Note that the function $f : \mathcal{X} \to \mathbb{R}$ is called convex if, for any geodesic γ, the composition $f \circ \gamma$ is convex (in the usual sense). Let us start with the definition of the directional derivative for functions on geodesically complete Hadamard spaces.

Definition 2.1. Let $f : \mathcal{X} \to \mathbb{R}$ be a real valued function. The directional derivative of f at $x \in \mathcal{X}$ in the direction $\gamma_x^z \in \mathcal{X}_x^*$ for some $z \in \mathcal{X}$, denoted by $Df(x; \gamma_x^z)$, is defined as

$$Df(x; \gamma_x^z) := \lim_{t \downarrow 0} \frac{f(\gamma_x^z(t)) - f(x)}{t}. \tag{2.1}$$

We will use the following remark in the proof of Theorem 2.4.

Remark 2.2. Assume that $\mathcal{X} = \mathbb{R}$, $x \in \mathbb{R}$ and $f : \mathbb{R} \to \mathbb{R}$, then for every $b \in (x, \infty)$, the directional derivative of f at x in the direction of γ_x^{x+b} is defined by

$$Df(x; \gamma_x^{x+b}) = \lim_{t \downarrow 0} \frac{f(x+t) - f(x)}{t},$$

which is the same as the usual directional derivative of f at x in the direction 1 denoted by $Df(x; 1)$.

Theorem 2.3. *If $f : \mathcal{X} \to \mathbb{R}$ is a convex function on $\mathcal{X}$ and $\gamma_x^z \in \mathcal{X}_x^*$, then*
(i) the function $Q : \mathrm{dom}(\gamma_x^z) \cap (0, \infty) \to \mathbb{R}$ defined by

$$Q(t) = \frac{f(\gamma_x^z(t)) - f(x)}{t},$$

is increasing.
(ii) $Df(x; \gamma_x^z)$ exists and is equal to $\inf_t Q(t)$.
(iii) $Df(x; \gamma_x^x) = 0$.

Proof. (i) Since f is convex, the function $g(t) = f(\gamma_x^z(t))$, defined on $\mathrm{dom}(\gamma_x^z)$, is convex. If $0 < t_1 < t_2$, then we have

$$\frac{g(t_1) - g(0)}{t_1} \leq \frac{g(t_2) - g(0)}{t_2}.$$

This implies that

$$\frac{f(\gamma_x^z(t_1)) - f(x)}{t_1} \leq \frac{f(\gamma_x^z(t_2)) - f(x)}{t_2},$$

which means that Q is increasing.

(ii) Assertion (i) implies that for any decreasing sequence of positive numbers $\{t_n\}$ which converges to zero, the sequence $\{Q(t_n)\}$ is increasing so that $\{Q(t_n)\}$ has a limit which is $Df(x;\gamma_x^z) = \inf_t Q(t)$.

(iii) Note that for every $x \in \mathcal{X}$ and each t, $\gamma_x^x(t) = x$. Hence

$$Df(x;\gamma_x^x) = \lim_{t\downarrow 0} \frac{f(\gamma_x^x(t)) - f(x)}{t} = 0.$$

□

Theorem 2.4. *(Mean value theorem) Suppose that $x, y \in \mathcal{X}$, $f : \mathcal{X} \longrightarrow \mathbb{R}$ is convex. Then, there exists $t_0 \in (0, d(x,y))$ such that*

$$\frac{f(y) - f(x)}{d(x,y)} \leq Df(\gamma_x^y(t_0); \sigma^y_{\gamma_x^y(t_0)}).$$

Proof. Let γ_x^y be the unit speed geodesic joining x to y. Then, $fo\gamma_x^y$ is a real valued convex function on $[0, d(x,y)]$. Using the mean value theorem for convex functions from $\mathbb{R}$ to $\mathbb{R}$, there exist $t_0 \in (0, d(x,y))$ and $z \in \partial fo\gamma_x^y(t_0)$ such that

$$\frac{fo\gamma_x^y(d(x,y)) - fo\gamma_x^y(0)}{d(x,y)} = z,$$

where $\partial fo\gamma_x^y(t_0)$ denotes the subdifferential of the real valued function $fo\gamma_x^y$ at t_0. We set $w = \gamma_x^y(t_0)$, then for the unit speed geodesic σ_w^y,

$$Df(w;\sigma_w^y) = \lim_{t\downarrow 0} \frac{fo\sigma_w^y(t) - fo\sigma_w^y(0)}{t} = Dfo\sigma_w^y(0;1).$$

Since the geodesic connecting w and y is unique, so $\sigma_w^y(t) = \gamma_x^y(t_0 + t)$ for every $t \in [0, d(w,y)]$. Hence, $Dfo\sigma_w^y(0;1) = Dfo\gamma_x^y(t_0;1)$ and $z \leq Dfo\gamma_x^y(t_0;1)$. Therefore,

$$\frac{f(y) - f(x)}{d(x,y)} \leq Df(\gamma_x^y(t_0); \sigma^y_{\gamma_x^y(t_0)}).$$

□

Definition 2.5. Let $f : \mathcal{X} \to \mathbb{R}$ be a convex function. A geodesic $\gamma_x^z \in \mathcal{X}_x^*$ is called the subgradient of f at x if

$$f(y) \geq f(x) + \langle \gamma_x^z, \sigma_x^y \rangle, \ \forall y \in \mathcal{X}, \ \forall \sigma_x^y \in \mathcal{X}_x^*.$$

The set of all subgradients of f at x is called the subdifferential of f at x and is denoted by $\partial f(x)$. The set-valued map $\partial f : \mathcal{X} \to \mathcal{X}^*$ is called the subdifferential of f.

It is worth pointing out that $\partial f(x) \subset \mathcal{X}_x^*$, for every $x \in \mathcal{X}$. A roughly analogous concept of subdifferential is introduced and investigated on the Hilbert ball in [14].

Theorem 2.6. *Let $f : \mathcal{X} \to \mathbb{R}$ be a convex function. $\gamma_x^x \in \partial f(x)$ if and only if x is a minimum point of f.*

Proof. We know that $\langle \gamma_x^x, \sigma_x^y \rangle = 0$, for every $x,\ y \in \mathcal{X}$, and $\sigma_x^y \in \mathcal{X}_x^*$. Hence, if $\gamma_x^x \in \partial f(x)$, then

$$f(y) \geq f(x) + \langle \gamma_x^x, \sigma_x^y \rangle = f(x),\ \forall y \in \mathcal{X},\ \forall \sigma_x^y \in \mathcal{X}_x^*,$$

which means that x is a minimum point of f.

Now assume that x is a minimum point of f, so for every $y \in \mathcal{X},\ f(y) \geq f(x)$. Therefore,

$$f(y) \geq f(x) + \langle \gamma_x^x, \sigma_x^y \rangle = f(x),\ \forall y \in \mathcal{X},\ \forall \sigma_x^y \in \mathcal{X}_x^*,$$

and the proof is complete. □

Theorem 2.7. *Let $f : \mathcal{X} \to \mathbb{R}$ be a convex function. If $Df(x; \sigma_x^y) \geq \langle \gamma_x^z, \sigma_x^y \rangle$, for all $y \in \mathcal{X}$ and $\sigma_x^y \in \mathcal{X}_x^*$, then $\gamma_x^z \in \partial f(x)$.*

Proof. The relations

$$Df(x; \sigma_x^y) \geq \langle \gamma_x^z, \sigma_x^y \rangle,\ f(y) - f(x) \geq \frac{f(\sigma_x^y(s)) - f(x)}{s} \geq Df(x; \sigma_x^y),$$

imply

$$f(y) - f(x) \geq \langle \gamma_x^z, \sigma_x^y \rangle,$$

and hence $\gamma_x^z \in \partial f(x)$. □

Corollary 2.8. *Let $f : \mathcal{X} \to \mathbb{R}$ be a convex function. Then x is a minimum point of f if and only if $Df(x; \gamma_x^z) \geq 0$, for each $\gamma_x^z \in \mathcal{X}_x^*$.*

Proof. If x is a minimum point, then $f(\gamma_x^z(t)) \geq f(x)$ for each $z \in \mathcal{X}$ and $t \in \mathrm{dom}\gamma_x^z$. Hence, $Df(x; \gamma_x^z) \geq 0$. The converse is obvious by Theorem 2.7. □

Lemma 2.9. *For each triple of points* $x,\ y,\ z \in \mathcal{X}$*, there exists* $w \in \mathcal{X}$ *such that* $d(x, y) = d(z, w)$ *and* γ_x^y *is parallel to* σ_z^w*.*

Proof. Since $\mathcal{X}$ is geodesically complete, there is a unit speed geodesic ray γ_x which connects x and y. By Proposition 9.2.28 in [5], there exists a unique unit speed geodesic ray σ_z starting at z, parallel to γ_x. Set $w \in \mathcal{X}$ such that $w = \sigma_z(d(x, y))$. Then, $d(x, y) = d(w, z)$ and γ_x^y is parallel to σ_z^w. Suppose that σ_z^v is another geodesic segment parallel to γ_x^y. Since it is also parallel to σ_z^w and $d(\sigma_z^w(0), \sigma_z^v(0)) = 0$, we have $d(\sigma_z^w(t), \sigma_z^v(t)) = 0$, for each $t \in [0, d(x, y)]$. □

We use the notation $\gamma_x^y \parallel \gamma_z^w$ when γ_x^y is parallel to γ_z^w, for $x, y, z, w \in \mathcal{X}$. Also if $x, y \in \mathbb{E}^2$, then xy is the line segment between x and y.

Definition 2.10.

(i) The function $P_{xy} : \mathcal{X}_x^* \longrightarrow \mathcal{X}_y^*$ defined by $P_{xy}(\gamma_x^w) = \gamma_y^v$ is called the parallel translation of γ_x^w along γ_x^y, in which v is selected such that $d(x, w) = d(y, v)$ and γ_x^w is parallel to γ_y^v.

(ii) To define the sum of γ_x^a and γ_x^b, consider a point c such that $P_{xa}(\gamma_x^b) = \gamma_a^c$. Then, $\gamma_x^a + \gamma_x^b := \gamma_x^c$.

(iii) We define

$$-\gamma_x^y := P_{yx}(\gamma_y^x).$$

$$\gamma_x^a - \gamma_x^b := \gamma_x^a + (-\gamma_x^b).$$

Theorem 2.11. *Suppose that* $\gamma_x^y = P_{ax}(\gamma_a^b)$ *and* $\gamma_x^z = P_{ax}(\gamma_a^c)$*, then*

(i) $d(b, c) = d(y, z)$,

(ii) $\angle_a(b, c) = \angle_x(y, z)$,

(iii) $\langle \gamma_a^b, \gamma_a^c \rangle = \langle \gamma_x^y, \gamma_x^z \rangle$,

(iv) $\langle -\gamma_x^y, \gamma_x^z \rangle = \langle \gamma_x^y, -\gamma_x^z \rangle$.

Proof. Let $\triangle(\bar{a}, \bar{b}, \bar{c})$ and $\triangle(\bar{x}, \bar{y}, \bar{z})$ be the comparison triangles for $\triangle(a, b, c)$ and $\triangle(x, y, z)$, respectively. By definition, $d(\gamma_a^b(t), \gamma_x^y(t)) = d_{\mathbb{E}^2}(\overline{\gamma_a^b(t)}, \overline{\gamma_x^y(t)}) = C$ in which C is constant for each t. We can assume that $\bar{a}\bar{b} \parallel \bar{x}\bar{y}$ and $\bar{a}\bar{c} \parallel \bar{x}\bar{z}$.

This means that, $\angle_{\bar{a}}(\bar{b}, \bar{c})$ and $\angle_{\bar{x}}(\bar{y}, \bar{z})$ are two angles with parallel sides. Then, they are congruent or supplementary. But since $d_{\mathbb{E}^2}(\overline{\gamma_a^b(t)}, \overline{\gamma_x^y(t)})$ is constant for each t, the two angles are congruent.

By a similar argument, we get $\angle_{\bar{a}}(\overline{\gamma_a^b(t)}, \overline{\gamma_a^c(t)}) = \angle_{\bar{x}}(\overline{\gamma_x^y(t)}, \overline{\gamma_x^z(t)})$, for each t. Thus, by definition, $\angle_a(b,c) = \angle_x(y,z)$. Moreover, $\triangle(\bar{a},\bar{b},\bar{c})$ is congruent to $\triangle(\bar{x},\bar{y},\bar{z})$. Then, $d_{\mathbb{E}^2}(\bar{b},\bar{c}) = d_{\mathbb{E}^2}(\bar{y},\bar{z})$ and hence $d(b,c) = d(y,z)$. Now by (i) and the definition of $\langle .,. \rangle$, (iii) is obvious.
(iv) Suppose that $-\gamma_x^z = \gamma_x^{z'}$ and $-\gamma_x^y = \gamma_x^{y'}$. Let $\triangle_1 = \triangle(\overline{x}_1, \overline{y'}, \overline{z})$ and $\triangle_2 = \triangle(\overline{x}_2, \overline{y}, \overline{z'})$ be the comparison triangles for $\triangle(x,y',z)$ and $\triangle(x,y,z')$ respectively. Since $\gamma_x^{y'} \parallel \gamma_y^x$ and $\gamma_x^{z'} \parallel \gamma_z^x$, we can consider $\triangle_1$ and $\triangle_2$ such that $\overline{x}_1\overline{y'}$ is parallel to $\overline{y}\overline{x}_2$ and $\overline{x}_2\overline{z'}$ is parallel to $\overline{z}\overline{x}_1$. Then, $\angle_{\overline{x}_1}(\overline{z},\overline{y'}) = \angle_{\overline{x}_2}(\overline{y},\overline{z'})$. Therefore, $\triangle_1$ and $\triangle_2$ are congruent. Hence, $d_{\mathbb{E}^2}(\overline{z'},\overline{y}) = d_{\mathbb{E}^2}(\overline{y'},\overline{z})$. It means that $d(z',y) = d(y',z)$. Now we have

$$\begin{aligned}\langle -\gamma_x^y, \gamma_x^z \rangle &= \langle \gamma_x^{y'}, \gamma_x^z \rangle = \tfrac{1}{2}[d^2(x,z) + d^2(y',x) - d^2(y',z)] \\ &= \tfrac{1}{2}[d^2(x,z') + d^2(x,y) - d^2(z',y)] = \langle \gamma_x^y, \gamma_x^{z'} \rangle = \langle \gamma_x^y, -\gamma_x^z \rangle.\end{aligned}$$

□

Lemma 2.12. *Let $f : \mathcal{X} \to \mathbb{R}$ be a convex function. Then, $\partial f : \mathcal{X} \to \mathcal{X}^*$ is monotone; that is*

$$\langle \gamma_x^y, \sigma_x^z - P_{yx}(\eta_y^n) \rangle \le 0,\ \forall x,\ y \in \mathcal{X},\ \forall \eta_y^n \in \partial f(y),\ \forall \sigma_x^z \in \partial f(x).$$

Proof. Suppose that $\eta_y^n \in \partial f(y)$ and $\sigma_x^z \in \partial f(x)$. Thus $f(y) - f(x) \ge \langle \gamma_x^y, \sigma_x^z \rangle$ and $f(x) - f(y) \ge \langle \gamma_y^x, \eta_y^n \rangle$. Note that

$$\langle \gamma_y^x, \eta_y^n \rangle = \langle P_{yx}(\eta_y^n), P_{yx}(\gamma_y^x) \rangle = \langle -P_{yx}(\eta_y^n), \gamma_x^y \rangle.$$

Therefore,

$$\langle \gamma_x^y, \sigma_x^z - P_{yx}(\eta_y^n) \rangle \le 0,\ \forall x,\ y \in \mathcal{X},\ \forall \eta_y^n \in \partial f(y),\ \forall \sigma_x^z \in \partial f(x).$$

□

Let S be a nonempty closed convex subset of $\mathcal{X}$ and $\pi_S : \mathcal{X} \longrightarrow S$ be the nearest point map onto S.

Now we need some lemmas to prove the density theorem for the subdifferential of a convex lower semicontinuous function on $\mathcal{X}$.

Lemma 2.13. *Let $f : \mathcal{X} \to \mathbb{R} \cup \{\infty\}$ be a proper, convex and lower semicontinuous function. Suppose that $(e, r_e) \in (\mathrm{epi}(f))^c$ and $X_0 = (x_0, f(x_0)) = \pi_{\mathrm{epi}(f)}(e, r_e)$ with $f(x_0) - r_e = 1$. Then, $\partial f(x_0) \ne \emptyset$.*

Proof. Set $E = (e, r_e)$. By Proposition 2.4 in [4], for each $A = (a, r_a) \in \text{epi}(f)$ not equal to X_0, we have $\angle_{X_0}(E, A) \geq \frac{\pi}{2}$. Consequently $\rho^2(A, X_0) + \rho^2(X_0, E) \leq \rho^2(A, E)$, where ρ is the metric of the space $\mathcal{X} \times \mathbb{R}$ defined as follows:

$$\rho^2((x_1, r_1), (x_2, r_2)) = d^2(x_1, x_2) + (r_2 - r_1)^2.$$

Thus

$$d^2(a, x_0) + d^2(x_0, e) + (f(x_0) - r_a)^2 + (f(x_0) - r_e)^2 \leq d^2(a, e) + (r_e - r_a)^2.$$

Therefore, we can easily find

$$\frac{1}{2}[d^2(a, x_0) + d^2(x_0, e) - d^2(a, e)] \leq (r_a - f(x_0))(f(x_0) - r_e). \tag{2.2}$$

Since $f(x_0) - r_e = 1$, we get

$$\langle \gamma_{x_0}^e, \gamma_{x_0}^a \rangle \leq r_a - f(x_0),$$

for all $a \in \text{dom} f$. Put $r_a = f(a)$. Clearly the above inequality holds for each $a \notin \text{dom} f$. Hence, $\gamma_{x_0}^e \in \partial f(x_0)$. □

It is worth pointing out that since r_a in (2.2) of the Lemma 2.13 can be selected large enough, we get $f(x_0) \geq r_e$.

Remark 2.14. The notation $(1-t)a \oplus tb$ is used for some results on Hilbert balls in [15], on hyperbolic spaces in [8, 14] and on Hadamard spaces in [7], to denote the unique point a_t with the property $d(a, a_t) = td(a, b)$ and $d(a_t, b) = (1-t)d(a, b)$. Now, if (x_0, y_0) and (x_1, y_1) are two points in $\mathcal{X} \times \mathcal{Y}$ and (x, y) is a point on the unique geodesic joining them, then (x, y) is the unique point satisfying the following equations:

$$\rho((x_0, y_0), (x, y)) = t\rho((x_0, y_0), (x_1, y_1)),$$

and

$$\rho((x_1, y_1), (x, y)) = (1-t)\rho((x_0, y_0), (x_1, y_1)),$$

for some $t \in [0, 1]$. Moreover, the point

$$(\gamma_{x_0}^{x_1}(td(x_0, x_1)), \gamma_{y_0}^{y_1}(td(y_0, y_1))) = ((1-t)x_0 \oplus tx_1, (1-t)y_0 \oplus ty_1),$$

has the same property. So

$$(1-t)(x_0, y_0) \oplus t(x_1, y_1) = ((1-t)x_0 \oplus tx_1, (1-t)y_0 \oplus ty_1),$$

for all $t \in [0, 1]$.

Assume that $x, y \in \mathcal{X}$. In the next lemma, we use the notation $[[x, y]]$ for the set $\{\gamma_x^y(t) \ : \ t \ \in \mathrm{dom}\gamma_x^y\}$.

Lemma 2.15. *Let $f : \mathcal{X} \to \mathbb{R} \cup \{\infty\}$ be a proper, convex and lower semicontinuous function. Suppose that $(y_0, r_0) \in (\mathrm{epi}(f))^c$ and $X_0 = (x_0, f(x_0)) = \pi_{\mathrm{epi}(f)}((y_0, r_0))$ and $x_0 \in \mathrm{int}(\mathrm{dom}f)$, where $\mathrm{dom}f = \{x \in \mathcal{X} |\ f(x) < \infty\}$. Then, $r_0 \neq f(x_0)$.*

Proof. Assume by contradiction that $r_0 = f(x_0)$. Put $Y_0 = (y_0, r_0)$. Let r be a positive number that $B(x_0, r) \subseteq \mathrm{dom}f$. There is $\lambda_0 \in [0, 1]$, such that $\gamma_{x_0}^{y_0}(\lambda d(x_0, y_0)) \in B(x_0, r)$ for the unit speed geodesic $\gamma_{x_0}^{y_0}$ and for each $\lambda \in [0, \lambda_0]$. First suppose that there exists $x_1 \in B(x_0, r) \cap [[x_0, y_0]]$ such that $f(x_0) < f(x_1)$. Hence, $x_1 = \gamma_{x_0}^{y_0}(\lambda_1 d(x_0, y_0))$ for some $\lambda_1 \in (0, \lambda_0)$. Put $X_1 = (x_1, f(x_1)) \in \mathrm{epi}f$. Then,

$$\rho^2(X_1, X_0) + \rho^2(X_0, Y_0) \leq \rho^2(X_1, Y_0).$$

Putting $\alpha = (f(x_1) - f(x_0))^2$, we have

$$\rho^2(X_0, X_1) = d^2(x_0, x_1) + \alpha = \lambda_1^2 d^2(x_0, y_0) + \alpha \tag{2.3}$$

$$\rho^2(Y_0, X_1) = d^2(y_0, x_1) + \alpha = (1 - \lambda_1)^2 d^2(x_0, y_0) + \alpha \tag{2.4}$$

and

$$\rho^2(X_0, Y_0) = d^2(x_0, y_0). \tag{2.5}$$

Hence, by (2.3), (2.4) and (2.5), we have

$$\lambda_1^2 d^2(x_0, y_0) + \alpha + d^2(x_0, y_0) \leq (1 - \lambda_1)^2 d^2(x_0, y_0) + \alpha.$$

Thus $\lambda_1^2 + 1 \leq (1 - \lambda_1)^2$, and we get $\lambda_1 \leq 0$, a contradiction. Next, consider the case that $f(x) \leq f(x_0)$, for each $x \in B(x_0, r) \cap [[x_0, y_0]]$. Let

$$Y_n = (1 - \frac{1}{n})X_0 \oplus \frac{1}{n} Y_0 = (y_n, r_n).$$

Hence, by Proposition 2.4 in [4], X_0 is the nearest point of epi(f) to each Y_n and $\{Y_n\}$ is a convergent sequence to X_0. If $y_0 \in B(x_0, r)$, then $f(y_0) \le f(x_0)$. Thus $(y_0, r_0) = (y_0, f(x_0)) \in \text{epi}(f)$ which is a contradiction. Therefore, $y_0 \in (B(x_0, r))^c$. Since by Remark 2.14, $r_n = f(x_0)$ for every n, so a similar argument for each Y_n shows that $y_n \in (B(x_0, r))^c$. This means that $\{y_n\}$ is a sequence in $(B(x_0, r))^c$ which is converging to x_0, thus $x_0 \notin B(x_0, r)$ that is a contradiction. □

Lemma 2.16. *Let $E' \in (\text{epi}(f))^c$ and $X_0 = (x_0, f(x_0)) = \pi_{\text{epi}(f)}(E')$, then there exists $E = (e, r_e) \in (\text{epi}(f))^c$ such that $f(x_0) - r_e = 1$ and $X_0 = \pi_{\text{epi}(f)}(E)$.*

Proof. Let γ be the geodesic joining X_0 to E'. Put $E' = (e', r_{e'})$. First suppose that $f(x_0) - r_{e'} \ge 1$. Since γ is continuous, by the intermediate value theorem, the assertion is obvious.

Next, suppose that $f(x_0) - r_{e'} < 1$. Put

$$l = \rho(X_0, E'),\ \ s = \frac{l}{f(x_0) - r_{e'}}.$$

Let $\overline{\gamma}$ be the extension of γ to $[0, \infty)$ that is the unit speed geodesic ray emanating from X_0. Put $E = \overline{\gamma}(s)$. We claim that E is the desired point. If $E = (e, r_e)$, then one has $E' = (1 - \frac{l}{s})X_0 \oplus \frac{l}{s}E$. By Remark 2.14, $e' = (1 - \frac{l}{s})x_0 \oplus \frac{l}{s}e$ and $r_{e'} = (1 - \frac{l}{s})f(x_0) + \frac{l}{s}r_e$. Hence, $f(x_0) - r_{e'} = \frac{l}{s}(f(x_0) - r_e)$. Therefore,

$$f(x_0) - r_e = \frac{s}{l}(f(x_0) - r_{e'}) = s \times \frac{f(x_0) - r_{e'}}{l} = 1.$$

Now we prove that $\pi_{\text{epi}(f)}(E) = X_0$. Suppose by contradiction that $\pi_{\text{epi}(f)}(E) = X'$ and $X_0 \ne X'$. Then, $\angle_{X_0}(X', E') \ge \frac{\pi}{2}$ and $\angle_{X'}(X_0, E) \ge \frac{\pi}{2}$. Then, the sum of the angles of $\triangle(X', X_0, E)$ is more than π, that is a contradiction. □

The next theorem is a generalization of the density theorem on geodesically complete Hadamard spaces. For density theorem on Hilbert spaces see [6].

Theorem 2.17. *Suppose that f is a proper, convex and lower semicontinuous function. Then* $\text{dom}(\partial f(x))$ *is dense in* $\text{int}(\text{dom} f)$.

Proof. Given $x_0 \in \text{int}(\text{dom} f)$, the point $X_0 = (x_0, f(x_0))$ is a boundary point of epi(f). So there exists a sequence $Y_n = (y_n, r_n)$ in the complement of epi(f), converges to X_0. Since epi(f) is convex and closed in $\mathcal{X} \times \mathbb{R}$, for each Y_n, there exists a unique point $X_n = (x_n, f(x_n)) \in$ epi(f) such that $\pi_{\text{epi}(f)}(Y_n) = X_n$. Moreover,

$$\rho(X_n, X_0) \leq \rho(X_n, Y_n) + \rho(Y_n, X_0) \leq 2\rho(Y_n, X_0),$$

which implies that X_n converges to X_0. Therefore, the sequence $\{x_n\}$ converges to x_0 and for every neighborhood U of x_0, there exists $x_n \in U$. By Lemma 2.16, one can assume that $f(x_n) - r_n = 1$, so by Lemma 2.13, $\partial f(x_n) \neq \emptyset$. □

References

[1] B. Ahmadi-kakavandi, M. Amini, Duality and subdifferential for convex functions on complete CAT(0) metric spaces, Nonlinear Anal. **73** (2010), 3450-3455.

[2] W. Ballmann, Lectures on Spaces of Nonpositive Curvature, Birkhauser Verlag (1995).

[3] I. D. Berg, I. G. Nikolaev, Quasilinearization and curvature of Aleksandrov spaces, Geom Dedicata. **133** (2008), 195-218.

[4] M. Bridson, A. Haefliger, Metric Spaces of Nonpositive Curvature, Springer-Verlag, Berlin (1999).

[5] D. Burago, Y. Burago, S. Ivanov, A Course in Metric Geometry, in: Graduate Studies in Math., vol. 33, Amer. Math. Soc., Providence, RI (2001).

[6] F. H. Clarke, Yu. S. Ledyaev, R. J. Stern, P. R. Wolenski, Nonsmooth Analysis and Control Theory, Springer-Verlag New York Berlin Heidelberg (1998).

[7] S. Dhompongsa, B. Panyanak, On Δ-convergence theorems in CAT(0) spaces, Computers and Mathematics with Applications, **56** (2008), 2572-2579.

[8] K. Goebel, S. Reich, Uniform Convexity, Hyperbolic Geometry, and Nonexpansive Mappings, Marcel Dekker, New York and Basel (1984).

[9] S. Hosseini, M. R. Pouryayevali, Generalized gradients and characterization of epi-Lipschitz sets in Riemannian manifolds, Nonlinear Anal., **74** (2011), 3884–3895.

[10] S. Hosseini, M. R. Pouryayevali, Nonsmooth optimization techniques on Riemannian Manifolds, JOTA, **158** (2013), 328–342.

[11] S. Hosseini, M. R. Pouryayevali, Euler characterization of epi-Lipschitz subsets of Riemannian manifolds, J. Convex. Anal., **20** (2013), No. 1, 67–91.

[12] S. Hosseini, M. R. Pouryayevali, On the metric projection onto prox-regular subsets of Riemannian manifolds, Proc. Amer. Math. Soc. **141** (2013), 233–244.

[13] E. Kopecká, S. Reich, A mean ergodic theorem for nonlinear semigroups on the Hilbert ball, J. Nonlinear Convex Anal. **11** (2010), 185-197.

[14] S. Reich, I. Shafrir, Nonexpansive iterations in hyperbolic spaces, Nonlinear Anal. **15** (1990), 537–558.

[15] I. Shafrir, Coaccretive operators and firmly nonexpansive mappings in the Hilbert ball, Nonlinear Anal. **18** (1992), 637–648.