BASIC CONVEX ANALYSIS IN METRIC SPACES WITH BOUNDED CURVATURE[*]

ADRIAN S. LEWIS[†], GENARO LÓPEZ-ACEDO[‡], AND ADRIANA NICOLAE[§]

Abstract. Differentiable structure ensures that many of the basics of classical convex analysis extend naturally from Euclidean space to Riemannian manifolds. Without such structure, however, extensions are more challenging. Nonetheless, in Alexandrov spaces with curvature bounded above (but possibly positive), we develop several basic building blocks. We define subgradients via projection and the normal cone, prove their existence, and relate them to the classical affine minorant property. Then, in what amounts to a simple calculus or duality result, we develop a necessary optimality condition for minimizing the sum of two convex functions.

Key words. subdifferential, normal cone, Alexandrov spaces

MSC codes. 65K10, 53C20

DOI. 10.1137/23M1551389

1. Introduction. Extensions of convex analysis tools to Alexandrov spaces find motivating applications, e.g., in averaging phylogenetic trees [6] (in the tree space of Billera, Holmes, and Vogtmann [12], which is an Alexandrov space with nonpositive curvature but not a manifold), restoration of manifold-valued images [10], or a conjecture of Donaldson on the convergence of long time solutions of the Calabi flow in Kähler geometry [41]. We also refer to Bačák's survey [7] for other interesting developments in optimization and analysis in Alexandrov spaces. Zhang and Sra [43] point out the importance of optimization in nonlinear spaces and discuss applications (especially when the space is a Riemannian manifold) in machine learning and theoretical computer science.

The subdifferential is the main analytic tool used to deal with nonsmooth convex functions on Euclidean space. It is an intuitive notion, fulfilling the role of the derivative in first order optimality conditions for such functions [37, 8]. Beyond the traditional setting of linear spaces, however, the question of how to define the subdifferential is less immediate. Rather than relying on the subgradient inequality, the approach we take here follows one standard route in nonconvex variational analysis (see [38, 18, 31]), deriving the subdifferential instead from the notion of the normal cone, an idea we can build through the metric projection.

In the nonlinear setting, the interest to define a suitable notion of subdifferential has found motivation in two facts: on the one hand, the study of the notion of gradient

[†]School of Operations Research and Information Engineering, Cornell University, Ithaca, NY 14853 USA (adrian.lewis@cornell.edu).

[‡]Department of Mathematical Analysis - IMUS, University of Seville, 41012 Seville, Spain (glopez@us.es).

[§]Corresponding author. Department of Mathematics, Babeş-Bolyai University, 400084, Cluj-Napoca, Romania (anicolae@math.ubbcluj.ro).

flow in spaces which are not necessarily endowed with a natural linear or differentiable structure (see [30, 4, 21]) and, on the other hand, the analysis of convergence of some algorithms in Riemannian manifolds (see [39, 26]).

As far as we know, the first notion of subdifferential was given in the nonlinear case by Udriște [42] in the setting of Riemannian manifolds using the subgradient inequality. Important properties of the subdifferential such as its connection to the normal cone and the directional derivative were later extensively analyzed (see, e.g., [27, 28]) and, as a consequence, generalizations of classical first order algorithms have been developed in this framework too. It is worth to mention that, in the case of Riemannian manifolds, the linear structure of the tangent space allows mimicking most of the constructions related to the subdifferential as well as the applications from the Euclidean case.

The case of spaces without a differentiable structure presents additional difficulties, mainly because of the lack of a fully linear structure in the tangent space. In [1, 33], a notion of subdifferential was proposed by first giving different definitions for the dual space and using in the definition of the subgradient inequality instead of the scalar product a quasi-linearization function in terms of distances introduced in [9]. In both cases, the lack of a suitable structure of the dual space and the definition of the scalar product limit the study to the case of nonpositive curvature and essential properties of the subdifferential are left out, such as the existence of subgradients at a continuity point of a convex function or a subdifferential calculus.

In order to elaborate a more coherent theory, in the present work we introduce the concept of subdifferential using normal cones. Since the notion of normal cone can be based on the metric projection whose properties are rich enough in Alexandrov spaces of curvature bounded above (see [5]), we consider this context and briefly discuss in section 2 some of its fundamental properties, together with other notions used in what follows. However, we also impose local compactness in our framework in order to obtain an analogue of the supporting hyperplane theorem from finite-dimensional Hilbert spaces, which allows us to establish the existence of subgradients at a continuity point of a convex function. We use the notion of tangent space and scalar product essentially due to Berestovskiĭ (see [2]) and discussed in detail in [15]. With these notions, Gigli and Nobili [21] introduced recently the concept of minus-subdifferential in connection to the study of gradient flows in Alexandrov spaces of curvature bounded above. This definition and the one we develop here are different (in Remark 4.6, we discuss the relation between these two definitions). Moreover, a characterization of the minus-subdifferential in terms of normal cones is not obvious. Another recent paper [17] uses the notion of tangent space to study monotone vector fields in Alexandrov spaces of nonpositive curvature mentioning as a particular example the subdifferential.

In sections 3 and 4 we introduce the notions of normal cone to a convex set and subdifferential of a convex function as elements of the tangent space and study their basic properties, along with some immediate examples. As pointed out in section 2.2, in the case of smooth Riemannian manifolds, there is a natural identification between the space of tangent vectors at a point and the tangent space as considered here. Consequently, using Remark 3.3(v) for the normal cone and Proposition 4.4 for the subdifferential, our definitions coincide with the ones introduced in [42, 28] in this setting.

To convey our culminating result—Theorem 4.20, which may be seen as a basic calculus or duality theorem—it helps to review the classical case in a Euclidean space X. Given two convex functions $f, g : X \to (-\infty, \infty]$, an archetypal decomposition

question of a kind ubiquitous in modern optimization seeks to characterize points x minimizing the sum $f + g$. A sufficient condition is trivial: if y is a subgradient of f at x, meaning classically that x minimizes the function $z \in X \mapsto f(z) - \langle y, z \rangle$, and the opposite vector $-y$ is a subgradient of g at x, then x minimizes $f + g$. Expressed concisely, the existence of opposite subgradients at x is a sufficient condition for x to minimize the sum. More generally, the sum of any subgradient of f and of g at x is easily seen to be a subgradient at x of $f + g$.

Less obvious is the converse question, a necessary condition. If a point x minimizes the sum $f + g$, must there exist two opposite subgradients, y and $-y$, for the two functions? A counterexample is the sum of the function $f : \mathbb{R} \to (-\infty, \infty]$ defined by

$$
f(x) = \begin{cases} \infty & \text{if } x < 0, \\ -\sqrt{x} & \text{if } x \geq 0 \end{cases}
$$

with the indicator function of the set of nonpositive reals $\delta_{\mathbb{R}_-} : \mathbb{R} \to [0, \infty]$ defined by

$$
\delta_{\mathbb{R}_-}(x) = \begin{cases} 0 & \text{if } x \leq 0, \\ \infty & \text{if } x > 0. \end{cases}
$$

However, if f is continuous at any point where g is finite, then the answer is yes, and indeed, more generally, subgradients of $f + g$ at any point are characterized in that case as sums of subgradients of f and of g there. Our culminating result, Theorem 4.20, is a version of this result extended to Alexandrov spaces.

The existence of the opposite subgradients, y and $-y$, is the most basic form of the Fenchel duality theorem, and many modern optimization algorithms for minimizing the primal objective $f(x) + g(x)$ seek in tandem such a vector y by implicitly maximizing a dual objective $-f^*(y) - g^*(-y)$ involving the conjugates $f^*(y) = \sup_{z \in X} (\langle y, z \rangle - f(z))$ of f and g^* of g. A popular general-purpose example is the alternating directions method of multipliers, surveyed in [14]. For elementary reasons, the dual objective is never larger than the primal, and equality guarantees that the corresponding vectors x and y are primal and dual optimal, a case that holds exactly when y and $-y$ are the requisite opposite subgradients of f and g at x. Whether Theorem 4.20 has an analogous dual interpretation we do not pursue it here. However, in Corollary 4.23, we give an application to a counterpart in Alexandrov spaces of a separation result for two nonempty, convex, closed, and disjoint sets, at least one of which is bounded. As an optimization tool, separation of disjoint convex sets needs no emphasis, either for fundamentals [37, 32] or algorithms [11], whence our broad interest in nonlinear extensions such as Corollary 4.23. A particular and fundamental machine learning example is the "support vector machine," which, in its simplest form, trained on some binary-labeled data points in Euclidean space, classifies new points using a "support vector" determining a separating hyperplane for the two convex hulls [40]. If the data instead lie in a manifold (of low rank matrices, for example) or some more general nonlinear space, then rather than resorting to the lifting procedures standard for support vector machines, we speculate that the pair of opposite tangent elements in our nonlinear separation result, Corollary 4.23, could serve as the classifier. This is a topic of ongoing investigation.

2. Preliminaries. In this section, we give some basic notions and properties of geodesic metric spaces. We refer the reader to [3, 15, 16] for more details.

2.1. Geodesic metric spaces and CAT(κ) spaces. Let (X, d) be a metric space. We denote the open (resp., closed) ball centered at $x \in X$ with radius $r > 0$ by

$B(x, r)$ (resp., $\overline{B}(x, r)$). For $C \subseteq X$, we denote the *diameter* of C by $\text{diam}(C)$. The *metric projection* P_C onto C is the mapping $P_C : X \to 2^C$ defined by

$$P_C(z) = \{y \in C : d(z, y) = \text{dist}(z, C)\} \quad \text{for all } z \in X,$$

where $\text{dist}(z, C) = \inf_{y \in C} d(z, y)$.

A *geodesic* is an isometric mapping $\gamma : [0, l] \subseteq \mathbb{R} \to X$. The image $\gamma([0, l])$ is called a *geodesic segment.* If instead of the interval $[0, l]$ one considers $\mathbb{R}$, then the image of γ is called a *geodesic line.* If every two points in X are joined by a (unique) geodesic segment, then X is called a *(uniquely) geodesic space.* We say that a subset C of a geodesic space is *convex* if for every $x, y \in C$, all geodesic segments with endpoints x and y are contained in C.

Assume that (X, d) is a uniquely geodesic space. For $x, y \in X$, denote the unique geodesic segment with endpoints x and y by $[x, y]$ and, given $t \in [0, 1]$, let $(1 - t)x + ty$ stand for the unique point belonging to $[x, y]$ whose distance to x equals $td(x, y)$. A set $C \subseteq X$ is said to satisfy the *betweenness property* if for every four pairwise distinct points $x, y, z, w \in C$, if $y \in [x, z]$ and $z \in [y, w]$, then $y, z \in [x, w]$. This property was studied in [19, 34, 23].

We say that X has the *geodesic extension property around* $x \in X$ if there exists a positive constant R such that for any distinct $y, z \in B(x, R)$ with $d(y, z) < R$, the geodesic from y to z can be extended beyond z to a geodesic of length R. We will often emphasize the dependence of the constant on the point by denoting it R_x. We say that X has the *geodesic extension property* if every geodesic segment (that is not reduced to a point) is contained in a geodesic line. One can show that if X is complete and satisfies the betweenness property and the geodesic extension property around every $x \in X$, then X has the geodesic extension property.

Let (X, d) be a metric space, and consider on the Cartesian product $X \times \mathbb{R}$ the metric

$$d_2((x_1, y_1), (x_2, y_2)) = \sqrt{d(x_1, x_2)^2 + |y_1 - y_2|^2},$$

where $x_1, x_2 \in X$ and $y_1, y_2 \in \mathbb{R}$.

If X is locally compact, then $X \times \mathbb{R}$ is locally compact as well. At the same time, if X is a (uniquely) geodesic space, then so is $X \times \mathbb{R}$. Moreover, geodesics in $X \times \mathbb{R}$ are given in terms of geodesics in X and in $\mathbb{R}$: if $\gamma : [0, r] \to X$ and $\alpha : [0, s] \to \mathbb{R}$ are geodesics, where $l = \sqrt{r^2 + s^2} > 0$, then $\sigma : [0, l] \to X \times \mathbb{R}$ defined by $\sigma(t) = (\gamma(rt/l), \alpha(st/l))$ for all $t \in [0, l]$ is a geodesic. Conversely, if $l > 0$ and $\sigma : [0, l] \to X \times \mathbb{R}$ is a geodesic from (x, λ) to (y, ω), taking $r = d(x, y)$ and $s = |\lambda - \omega|$, then $\sigma(t) = (\gamma(rt/l), \alpha(st/l))$ for all $t \in [0, l]$, where $\gamma : [0, r] \to X$ and $\alpha : [0, s] \to \mathbb{R}$ are geodesics from x to y and from λ to ω, respectively.

Suppose that (X, d) is a geodesic space. Given $\lambda \in \mathbb{R}$, if X has the geodesic extension property around $x \in X$, then $X \times \mathbb{R}$ also has the geodesic extension property around (x, λ) with the same constant.

A function $f : X \to (-\infty, \infty]$ is called *convex* if for any geodesic $\gamma : [0, l] \to X$ and any $t \in [0, 1]$, $f(\gamma(lt)) \leq (1 - t)f(\gamma(0)) + tf(\gamma(l))$. The (effective) *domain* of f is defined by $\text{dom } f = \{x \in X \mid f(x) < \infty\}$. The *epigraph* of f is $\text{epi } f = \{(x, \lambda) \in \text{dom } f \times \mathbb{R} \mid \lambda \geq f(x)\}$. One can easily see that $\text{epi } f$ is closed if and only if f is lower semicontinuous, which is also equivalent to the fact that its sublevel sets are closed. If f is convex, then $\text{dom } f$ and $\text{epi } f$ are convex.

For $\kappa \in \mathbb{R}$, let M_κ^2 denote the complete, simply connected, 2-dimensional Riemannian manifold of constant sectional curvature κ. We denote the diameter of M_κ^2 by D_κ. In other words, $D_\kappa = \infty$ if $\kappa \leq 0$, while $D_\kappa = \pi/\sqrt{\kappa}$ if $\kappa > 0$.

A *geodesic triangle* is the union of three geodesic segments joining three points. We say that a triangle $\Delta(\overline{x}_1, \overline{x}_2, \overline{x}_3)$ in M_κ^2 is a *comparison triangle* for a geodesic triangle $\Delta(x_1, x_2, x_3)$ if $d(x_i, x_j) = d_{M_\kappa^2}(\overline{x}_i, \overline{x}_j)$ for all $i, j \in \{1, 2, 3\}$.

A metric space is called a CAT(κ) *space* (also known as a space of curvature bounded above by κ in the sense of Alexandrov) if every two points at distance less than D_κ can be joined by a geodesic and geodesic triangles having perimeter less than $2D_\kappa$ are not thicker than the comparison triangles in M_κ^2.

Suppose next that X is a CAT(κ) space. Points in X at distance less than D_κ are joined by a unique geodesic segment, and this segment varies continuously with its endpoints. Moreover, balls of radius smaller than $D_\kappa/2$ are convex. If $\kappa \geq 0$, $X \times \mathbb{R}$ is a CAT(κ) space as well.

Sets of diameter less than D_κ if $\kappa > 0$ satisfy the betweenness property. This shows that if a complete CAT(0) space has the geodesic extension property around every point, then it also has the geodesic extension property.

If the diameter of X is smaller than $D_\kappa/2$ when $\kappa > 0$, then X is 2-*uniformly convex* (see [24, 36]) in the sense that there exists a parameter $K = K(\kappa, \mathrm{diam}(X)) > 0$ such that for all $x, y, z \in X$ and all $t \in [0, 1]$,

$$d(z, (1-t)x + ty)^2 \leq (1-t)\, d(z, x)^2 + t\, d(z, y)^2 - K\, t(1-t)\, d(x, y)^2.$$

The above inequality with $K = 1$ characterizes CAT(0) spaces, in fact.

The Alexandrov angle between two nonconstant geodesics $\gamma : [0, l] \to X$ and $\gamma' : [0, l'] \to X$ issuing at the same point $x = \gamma(0) = \gamma'(0)$ is well-defined and can be determined as $\angle(\gamma, \gamma') = \lim_{t,t' \searrow 0} \angle_{\overline{x}}(\overline{\gamma(t)}, \overline{\gamma'(t')})$, where $\Delta(\overline{\gamma(t)}, \overline{x}, \overline{\gamma'(t')})$ is a comparison triangle in $\mathbb{R}^2$. We also denote it by $\angle_x(y, z)$, where $y \in \gamma((0, l])$ and $z \in \gamma'((0, l'])$.

Let $C \subseteq X$ be a convex set and $x \in X \setminus C$ with $\mathrm{dist}(x, C) < D_\kappa/2$. If $P_C(x) \neq \emptyset$, then $P_C(x)$ is a singleton. If $z = P_C(x)$ and $y \in C$ with $y \neq z$, then $\angle_z(x, y) \geq \pi/2$. If C is additionally complete in the induced metric, then x always has a nearest point in C, and hence $P_C(x)$ is a singleton. We refer the reader to [20, 5] for a more thorough discussion on the behavior of the metric projection in CAT(κ) spaces.

2.2. Tangent spaces. *In what follows, if nothing else is mentioned about the context, we always suppose that (X, d) is a locally compact CAT(κ) space, where $\kappa > 0$. To simplify the exposition and without loss of generality (see Remark 3.3(iv)), we also suppose that X is uniquely geodesic. In addition, assume that for all $x \in X$, X has the geodesic extension property around x with a constant smaller than $D_\kappa/2$. This constant will be denoted by R_x.* (Actually, in the subsequent discussion we only need to impose the geodesic extension property and the existence of a compact neighborhood around a fixed point where we consider the tangent space of X.)

If x is a point in a smooth Riemannian manifold, then there exists a neighborhood of x that is a CAT(κ) space for some suitable $\kappa \in \mathbb{R}$ (see, e.g., [15, Chapter II.1, Appendix]). Moreover, since the injectivity radius in a Riemannian manifold is a continuous function (see [13]), there exists a ball centered at x that is a CAT(κ) space having the geodesic extension property around x with an appropriate constant.

Fix $x \in X$. Denote by $\Theta_x X$ the set of all nonconstant geodesics issuing at x. The Alexandrov angle induces a metric on the set $\Sigma_x X$ of equivalence classes of geodesics in $\Theta_x X$, where two geodesics $\gamma, \eta \in \Theta_x X$ are considered equivalent if $\angle(\gamma, \eta) = 0$. Note that if X is a Riemannian manifold, then $\Sigma_x X$ is isometric to the unit sphere in the tangent space of X at x.

For a geodesic $\gamma \in \Theta_x X$, we denote by $[\gamma]$ its equivalence class. At the same time, when working with an equivalence class in $\Sigma_x X$, we assume that it is represented by

a geodesic in $\Theta_x X$. Since X has the geodesic extension property around x, we can always suppose that a representative of an equivalence class is a geodesic of length at least R_x, where R_x is the constant for the geodesic extension property around x.

Under our assumptions, the metric space $(\Sigma_x X, \angle)$ is complete. We will refer to the elements in $\Sigma_x X$ as *directions* at x and to $(\Sigma_x X, \angle)$ as the *space of directions* at x. The *tangent space* $T_x X$ of X at x is the Euclidean cone over the metric space $(\Sigma_x X, \angle)$ (see [15, Chapter I, Definition 5.6]). More precisely, $T_x X = (\Sigma_x X \times [0, \infty))/\sim$, where for $([\gamma], r), ([\eta], s) \in \Sigma_x X \times [0, \infty)$, $([\gamma], r) \sim ([\eta], s)$ if and only if $(r = s = 0)$ or $(r = s > 0$ and $[\gamma] = [\eta])$. The equivalence class of $([\gamma], 0) \in T_x X$ is called the *origin* of $T_x X$ and will be denoted by o_x. (When dealing with the origin o_x, the direction at x bears no relevance and sometimes, to simplify the writing, we consider the geodesic constantly equal to x.)

Observe that despite the terminology, this construction is, in general, merely a metric cone. However, in the case when X is a Riemannian manifold, this notion coincides with the traditional notion of tangent space. Besides [15, 3], the preliminary part of [25] describes in detail the construction and properties of tangent spaces.

The *metric* d_x on $T_x X$ is defined for $v = ([\gamma], r), w = ([\eta], s) \in T_x X$ by

$$d_x(v, w)^2 = r^2 + s^2 - 2rs \cos \angle(\gamma, \eta),$$

while the *scalar product* of v and w is given by $\langle v, w \rangle = rs \cos \angle(\gamma, \eta)$. The multiplication with a nonnegative scalar $\lambda \geq 0$ is defined for $v = ([\gamma], r)$ by $\lambda v = ([\gamma], \lambda r)$. We say that $v = ([\gamma], r), w = ([\eta], s) \in T_x X$ are *opposite* to each other if $(r = s = 0)$ or $(r = s > 0$ and $\angle(\gamma, \eta) = \pi)$. Observe that, under our assumptions on X, $T_x X$ is a complete CAT(0) space (see [35, 15]).

We finish this section with the following auxiliary result concerning the tangent space of $X \times \mathbb{R}$.

LEMMA 2.1. *Let* $(x, \lambda) \in X \times \mathbb{R}$, *and, for* $i \in \{1, 2\}$, *let* $\sigma_i : [0, l_i] \to X \times \mathbb{R}$ *be a nonconstant geodesic issuing at* (x, λ) *that is written as* $\sigma_i(t) = (\gamma_i(r_i t / l_i), \alpha_i(s_i t / l_i))$ *for all* $t \in [0, l_i]$, *where* $l_i^2 = r_i^2 + s_i^2$, *and* $\gamma_i : [0, r_i] \to X$ *and* $\alpha_i : [0, s_i] \to \mathbb{R}$ *are geodesics. Then the following hold:*

(i) $\langle ([\sigma_1], l_1), ([\sigma_2], l_2) \rangle = \langle ([\gamma_1], r_1), ([\gamma_2], r_2) \rangle + \langle ([\alpha_1], s_1), ([\alpha_2], s_2) \rangle$.

(ii) $[\sigma_1] = [\sigma_2]$ *if and only if* $([\gamma_1] = [\gamma_2], [\alpha_1] = [\alpha_2], r_1/l_1 = r_2/l_2, s_1/l_1 = s_2/l_2)$.

Proof. Note first that for $i \in \{1, 2\}$, $\alpha_i(t) = \lambda + t$ for all $t \in [0, s_i]$ or $\alpha_i(t) = \lambda - t$ for all $t \in [0, s_i]$. If for both $i = 1$ and $i = 2$, α_i has the same form, let $c = 1$. Otherwise, let $c = -1$. Denote also $a_i = r_i/l_i$ and $b_i = s_i/l_i$, where $i \in \{1, 2\}$.

(i) Take $t \in (0, \min\{l_1, l_2\}]$. For the geodesic triangle $\Delta(\sigma_1(t), (x, \lambda), \sigma_2(t))$, consider a comparison triangle $\Delta(\overline{\sigma_1(t)}, \overline{(x, \lambda)}, \overline{\sigma_2(t)})$ in $\mathbb{R}^2$ and denote by $\theta(t)$ its interior angle at $\overline{(x, \lambda)}$. Then $\cos \theta(t) = 1 - d_2(\sigma_1(t), \sigma_2(t))^2/(2t^2)$. As

$$d_2(\sigma_1(t), \sigma_2(t))^2 = d(\gamma_1(a_1 t), \gamma_2(a_2 t))^2 + (b_1 t)^2 + (b_2 t)^2 - 2c b_1 b_2 t^2,$$

it follows that

$$\cos \theta(t) = \frac{1}{2}\left(a_1^2 + a_2^2 - \frac{d(\gamma_1(a_1 t), \gamma_2(a_2 t))^2}{t^2}\right) + c b_1 b_2.$$

Now take a comparison triangle $\Delta(\overline{\gamma_1(a_1 t)}, \overline{x}, \overline{\gamma_2(a_2 t)})$ in $\mathbb{R}^2$ and denote by $\rho(t)$ its interior angle at $\overline{x}$. Then

$$\cos \rho(t) = \frac{1}{2 a_1 a_2}\left(a_1^2 + a_2^2 - \frac{d(\gamma_1(a_1 t), \gamma_2(a_2 t))^2}{t^2}\right).$$

Hence, $\cos\theta(t) = a_1 a_2 \cos\rho(t) + cb_1 b_2$. Taking limit as $t \searrow 0$, we conclude that

$$(2.1) \qquad \cos\angle(\sigma_1, \sigma_2) = a_1 a_2 \cos\angle(\gamma_1, \gamma_2) + cb_1 b_2,$$

from where one obtains the desired equality.

(ii) If $[\gamma_1] = [\gamma_2]$, $[\alpha_1] = [\alpha_2]$, $a_1 = a_2$, and $b_1 = b_2$, then $\cos\angle(\gamma_1, \gamma_2) = 1$, $c = 1$, and (2.1) gives $\cos\angle(\sigma_1, \sigma_2) = a_1^2 + b_1^2 = 1$, and hence $[\sigma_1] = [\sigma_2]$.

Suppose now $[\sigma_1] = [\sigma_2]$. Then $\cos\angle(\sigma_1, \sigma_2) = 1$ and, using again (2.1), we obtain that $1 = a_1 a_2 \cos\angle(\gamma_1, \gamma_2) + cb_1 b_2 \leq a_1 a_2 + b_1 b_2$, so

$$1 \leq (a_1 a_2 + b_1 b_2)^2 = (a_1^2 + b_1^2)(a_2^2 + b_2^2) - (a_1 b_2 - a_2 b_1)^2 = 1 - (a_1 b_2 - a_2 b_1)^2.$$

We conclude that $\cos\angle(\gamma_1, \gamma_2) = 1$, $c = 1$, and $a_1 b_2 = a_2 b_1$. Hence, $[\gamma_1] = [\gamma_2]$ and $[\alpha_1] = [\alpha_2]$. Again taking into account that $a_1^2 + b_1^2 = a_2^2 + b_2^2 = 1$, we obtain $a_1 = a_2$ and $b_1 = b_2$. $\qquad\square$

3. Normal cones. Traditionally, a central idea in smooth analysis is the approximation of a smooth function and a smooth manifold by a linear function and a linear subspace, respectively. In the nonsmooth case, a function is approximated by a family of linear functions, while a set is approximated by cones. Normal cones can be defined in terms of the metric projection. In $\mathbb{R}^n$, given a convex set C and a point $x \in C$, vectors in

$$N_C(x) = \{v \in \mathbb{R}^n \mid x = P_C(x+v)\} = \{v \in \mathbb{R}^n \mid \langle y - x, v \rangle \leq 0 \text{ for all } y \in C\}$$

are called normals to C at x.

The metric projection is in fact a purely metric notion and exhibits a sufficiently regular behavior in nonlinear settings with a rich enough geometry (see [5]). For this reason, it is natural to consider using normal cones to define the notion of subdifferential of a function and get a meaningful subdifferential calculus in nonlinear spaces.

Let $C \subseteq X$ be nonempty and convex, and let $x \in C$.

DEFINITION 3.1. *The normal cone to C at x is given by*

$$N_C(x) = \{([\gamma], r) \in T_x X \mid r = 0 \ or \ (x = P_C(\gamma(t)) \ for \ some \ t > 0)\}.$$

Elements in $N_C(x)$ are called normals to C at x.

In connection to this definition, the following facts are immediate.

Remark 3.2.
(i) If $x = P_C(\gamma(t))$, then $x = P_C(\gamma(s))$ for all $s \in [0, R_x]$.
(ii) $N_C(x)$ is well-defined in the sense that if $\gamma, \eta \in \Theta_x X$ with $\eta \in [\gamma]$ and $x = P_C(\gamma(t))$ for some $t > 0$, then $x = P_C(\eta(s))$ for all $s \in [0, R_x]$.
Suppose there exists $y \in C$ such that $d(y, \eta(R_x)) < d(x, \eta(R_x))$. Then

$$\pi/2 \leq \angle_x(\gamma(t), y) \leq \angle_x(\gamma(t), \eta(R_x)) + \angle_x(\eta(R_x), y) = \angle_x(\eta(R_x), y),$$

from where $d(y, \eta(R_x)) \geq d(x, \eta(R_x))$, a contradiction. We conclude that $x = P_C(\eta(R_x))$, from where $x = P_C(\eta(s))$ for all $s \in [0, R_x]$.

In the next remark, we collect some elementary properties of normal cones.

Remark 3.3.

(i) $N_C(x)$ is closed under multiplication with a nonnegative scalar. Thus, for $r > 0$, $([\gamma], r) \in N_C(x)$ if and only if $([\gamma], 1) \in N_C(x)$.

(ii) If $\mathrm{diam}(C) < D_\kappa$, then $\overline{C}$ is convex and $N_C(x) = N_{\overline{C}}(x)$.

(iii) If $D \subseteq C$ is convex and $x \in D$, then $N_C(x) \subseteq N_D(x)$.

(iv) Let V be a convex neighborhood of x. Then $N_C(x) = N_{C \cap V}(x)$. This allows us to assume that the diameter of the set C is as small as needed.

Indeed, let $r > 0$ such that $B(x, r) \subseteq V$. Using (iii), it is enough to show that if $([\gamma], 1) \in N_{C \cap V}(x)$, then $([\gamma], 1) \in N_C(x)$. Suppose $([\gamma], 1) \in N_{C \cap V}(x)$. Then there exists $t \in (0, r/2)$ such that $x = P_{C \cap V}(\gamma(t))$. For $y \in C \setminus V$,

$$r \leq d(x, y) \leq d(x, \gamma(t)) + d(\gamma(t), y) < r/2 + d(\gamma(t), y),$$

so $d(\gamma(t), y) > r/2 > d(\gamma(t), x)$. This shows that $x = P_C(\gamma(t))$, and we conclude that $([\gamma], 1) \in N_C(x)$.

(v) $N_C(x) = \{([\gamma], r) \in T_x X \mid \langle ([\gamma], r), ([\gamma^y], d(x, y)) \rangle \leq 0 \text{ for all } y \in C\}$, where for $y \in C$, γ^y is the geodesic from x to y.

Indeed, $([\gamma], 1) \in N_C(x)$ if and only if $x = P_C(\gamma(R_x))$ if and only if $\angle(\gamma, \gamma^y) \geq \pi/2$ for all $y \in C$ with $y \neq x$.

(vi) Let (x_n) be a sequence in C converging to x, $l \in (0, D_\kappa/2)$, and, for $n \in \mathbb{N}$, let $\gamma_n : [0, l] \to X$ be a nonconstant geodesic issuing at x_n. If $([\gamma_n], 1) \in N_C(x_n)$ for all $n \in \mathbb{N}$ and (γ_n) converges pointwise to a geodesic γ, then $([\gamma], 1) \in N_C(x)$.

Indeed, $P_C(\gamma_n(l)) = x_n$ for all $n \in \mathbb{N}$. Let $y \in C$. Then

$$d(\gamma(l), x) \leq d(\gamma(l), \gamma_n(l)) + d(\gamma_n(l), x_n) + d(x_n, x)$$
$$\leq d(\gamma(l), \gamma_n(l)) + d(\gamma_n(l), y) + d(x_n, x)$$

for all $n \in \mathbb{N}$. We get that $d(\gamma(l), x) \leq d(\gamma(l), y)$ for all $y \in C$, which shows that $P_C(\gamma(l)) = x$, so $([\gamma], 1) \in N_C(x)$.

We also have the following result.

PROPOSITION 3.4.

(i) *If $x \in \mathrm{int}(C)$, then $N_C(x) = \{o_x\}$.*

(ii) *If $N_C(x) = \{o_x\}$, then $x \in \mathrm{int}(\overline{C})$.*

Proof. (i) Let $x \in \mathrm{int}(C)$. If $([\gamma], 1) \in N_C(x)$, then $x = P_C(\gamma(t))$ for some $t > 0$, which contradicts the fact that $x \in \mathrm{int}(C)$.

(ii) Let $R \in (0, R_x]$ such that $\overline{B}(x, R)$ is compact, and take the set $D = C \cap \overline{B}(x, R)$. As $\{o_x\} = N_C(x) = N_D(x) = N_{\overline{D}}(x)$, we can assume that D is closed and hence compact (otherwise, consider $\overline{D}$ instead of D). Suppose that $x \notin \mathrm{int}(\overline{C})$. Then $x \in D \setminus \mathrm{int}(D)$.

Take a sequence $(x_n) \subseteq X \setminus D$ with $x_n \to x$ and $d(x_n, x) \leq R/2$ for all $n \in \mathbb{N}$. For $n \in \mathbb{N}$, take $p_n = P_D(x_n)$. Observe that $d(x, p_n) \leq d(x, x_n) + d(x_n, p_n) \leq 2d(x, x_n)$ for all $n \in \mathbb{N}$, and hence $p_n \to x$.

Extend the geodesic from p_n to x_n beyond x_n to a point x'_n so that $d(p_n, x'_n) = R/2$. Because $(x'_n) \subseteq \overline{B}(x, R)$, the sequence (x'_n) has a convergent subsequence whose limit we denote by x'. For $n \in \mathbb{N}$, denote by γ_n the geodesic from p_n to x'_n and let γ be the geodesic from x to x'. Since $([\gamma_n], 1) \in N_D(p_n)$ for all $n \in \mathbb{N}$, applying Remark 3.3(vi), we get $([\gamma], 1) \in N_D(x)$, a contradiction. $\square$

The above result can be interpreted as an analogue of the supporting hyperplane theorem from finite-dimensional Hilbert spaces. Indeed, if C is convex and

closed and $x \in \mathrm{bd}(C)$, there exists $([\gamma], 1) \in N_C(x)$, and hence, by Remark 3.3(v), $\langle([\gamma], 1), ([\gamma^y], d(x, y))\rangle \leq 0$ for all $y \in C$.

We finish this section with the following straightforward examples.

Example 3.5.
(i) $N_{\{x\}}(x) = T_x X$.
(ii) If γ and η are geodesics issuing at x, then $([\gamma], 1) \in N_{\mathrm{Im}\,\eta}(x)$ if and only if $([\eta], 1) \in N_{\mathrm{Im}\,\gamma}(x)$.
(iii) Let $z \in X$ and $r \in (0, D_\kappa/2)$. Suppose that $x \in \mathrm{bd}(\overline{B}(z, r))$, and denote by Γ the set of all nonconstant geodesics $\gamma : [0, l] \to X$ issuing at x with the property that $d(z, x) + l = d(z, \gamma(l))$. Then $([\gamma], 1) \in N_{\overline{B}(z,r)}(x)$ if and only if $\gamma \in \Gamma$.
(iv) Given $x \in X$, in the geodesic space $X \times \mathbb{R}$ we have $([\sigma], 1) \in N_{X \times \mathbb{R}_+}(x, 0)$ if and only if $\sigma : [0, l] \to X \times \mathbb{R}$ is defined by $\sigma(t) = (x, -t)$ for all $t \in [0, l]$.

Proof. (iii) If $([\gamma], 1) \in N_{\overline{B}(z,r)}(x)$, then $x = P_{\overline{B}(z,r)}(\gamma(R_x))$. Let $y \in [z, \gamma(R_x)]$ with $d(z, y) = r$. Then $y \in \overline{B}(z, r)$ and $d(\gamma(R_x), y) \geq d(\gamma(R_x), x)$, so

$$d(z, \gamma(R_x)) \leq d(z, x) + d(x, \gamma(R_x)) \leq d(z, y) + d(y, \gamma(R_x)) = d(z, \gamma(R_x)).$$

We conclude that $\gamma \in \Gamma$ and, by uniqueness of geodesics, $x = y$.

Conversely, if $\gamma \in \Gamma$, then for all $y \in \overline{B}(z, r)$,

$$d(z, x) + d(x, \gamma(R_x)) = d(z, \gamma(R_x)) \leq d(z, y) + d(y, \gamma(R_x)),$$

from where $d(\gamma(R_x), x) \leq d(\gamma(R_x), y)$. Thus, $x = P_{\overline{B}(z,r)}(\gamma(R_x))$, so $([\gamma], 1) \in N_{\overline{B}(z,r)}(x)$.

(iv) Let $\sigma : [0, l] \to X \times \mathbb{R}$ be a nonconstant geodesic issuing at $(x, 0)$ that is written as $\sigma(t) = (\gamma(rt/l), \alpha(st/l))$ for all $t \in [0, l]$, where $l^2 = r^2 + s^2$, and $\gamma : [0, r] \to X$ and $\alpha : [0, s] \to \mathbb{R}$ are geodesics. Note that $\alpha(t) = t$ for all $t \in [0, s]$ or $\alpha(t) = -t$ for all $t \in [0, s]$.

If $([\sigma], 1) \in N_{X \times \mathbb{R}_+}(x, 0)$, then $P_{X \times \mathbb{R}_+}(\sigma(t)) = (x, 0)$ for some $t \in (0, l]$. Thus,

$$d_2(\sigma(t), (x, 0)) \leq d_2(\sigma(t), (x, \lambda)),$$

so $|\alpha(st/l)| \leq |\alpha(st/l) - \lambda|$ for all $\lambda \geq 0$ and we get that $\alpha(t) = -t$ for all $t \in [0, s]$. Moreover,

$$d_2(\sigma(t), (x, 0)) \leq d_2(\sigma(t), (\gamma(rt/l), 0)),$$

from where $\gamma(rt/l) = x$ and $r = 0$.

Conversely, if $\sigma : [0, l] \to X \times \mathbb{R}$ is defined by $\sigma(t) = (x, -t)$ for all $t \in [0, l]$, then

$$d_2(\sigma(t), (x, 0))^2 = t^2 \leq d(x, y)^2 + (t + \lambda)^2 = d_2(\sigma(t), (y, \lambda))^2$$

for all $t \in [0, l]$ and all $(y, \lambda) \in X \times \mathbb{R}_+$. This shows that $([\sigma], 1) \in N_{X \times \mathbb{R}_+}(x, 0)$. $\square$

4. The subdifferential via normal cones. One possibility to introduce the subdifferential of a function at a point is to view the subgradients as "slopes" of continuous affine minorants that coincide with the function at that point. Another perhaps more geometric approach is to use the normal cone to the epigraph of the function. As pointed out previously, we consider the second way more suitable for the nonlinear setting due to the important role of the metric projection in this approach.

DEFINITION 4.1. *Let $f : X \to (-\infty, \infty]$ be convex, and let $x \in \operatorname{dom} f$. We say that $([\gamma], r) \in T_x X$ is a subgradient of f at x if $([\sigma], 1) \in N_{\operatorname{epi} f}(x, f(x))$, where $\sigma : [0, R_x] \to X \times \mathbb{R}$ is defined by*

$$\sigma(t) = \left(\gamma \left(rt/\sqrt{r^2 + 1} \right), f(x) - t/\sqrt{r^2 + 1} \right) \quad \text{for all } t \in [0, R_x].$$

The set of all subgradients of f at x forms the subdifferential of f at x, denoted by $\partial f(x)$.

Regarding the previous definition, we observe the following.

Remark 4.2.
  (i) $\partial f(x)$ is well-defined since, by Lemma 2.1(ii), we have that $\widetilde{\sigma} \in [\sigma]$ if and only if there exists $\widetilde{\gamma} \in [\gamma]$ such that

$$\widetilde{\sigma}(t) = \left(\widetilde{\gamma} \left(rt/\sqrt{r^2 + 1} \right), f(x) - t/\sqrt{r^2 + 1} \right) \quad \text{for all } t \in [0, R_x].$$

  (ii) If X is a CAT(0) space with the geodesic extension property, then one can assume that γ is actually a geodesic ray, and so σ can be defined on any interval $[0, l] \subseteq [0, \infty)$.

A key point in understanding the behavior of the subdifferential of convex functions is the fact that such functions that are continuous at a point are locally Lipschitz there.

LEMMA 4.3. *Let (X, d) be a uniquely geodesic space, let $f : X \to (-\infty, \infty]$ be convex, and let $x \in \operatorname{dom} f$. Suppose that X has the geodesic extension property around x. Then the following are equivalent:*
  (i) *f is continuous at x.*
  (ii) *f is bounded on a neighborhood of x.*
  (iii) *f is Lipschitz on a neighborhood of x.*

Proof. We only prove that (ii) implies (iii) since the other implications are obvious. Suppose there exists $r > 0$ such that for all $y \in B(x, r)$, $|f(y)| \le M$ for some $M \ge 0$. We show that f is Lipschitz on $B(x, R)$, where $R = \min\{r/2, R_x/4\}$. To this end, let $y, z \in B(x, R)$ be distinct. Then $d(y, z) < R_x/2$ and we can choose $w \in X$ with $z = (1 - t)y + tw$, where $t = d(y, z)/(d(y, z) + R)$. Note that $d(w, z) = R$. Also, $d(x, w) \le d(x, z) + d(z, w) < r$, so $w \in B(x, r)$. Because f is convex, $f(z) \le (1 - t)f(y) + tf(w)$, and hence

$$f(z) - f(y) \le t(f(w) - f(y)) \le 2tM = 2Md(y, z)\frac{1}{d(y, z) + R} \le \frac{2M}{R}d(y, z).$$

By swapping the roles of x and y in the above argument, we finally get $|f(y) - f(z)| \le (2M/R)d(y, z)$. □

We characterize below the subdifferential of a convex function by means of a variational inequality. This fact will be essential to prove the first order condition for minimizers of convex functions. We return to the setting of a CAT(κ) space (with the additional conditions assumed in section 2.2).

PROPOSITION 4.4. *Let $f : X \to (-\infty, \infty]$ be convex, let $x \in \operatorname{dom} f$, and let $([\gamma], r) \in T_x X$. For $y \in X$, denote by γ^y the geodesic from x to y. Then $([\gamma], r) \in \partial f(x)$ if and only if*

$$(4.1) \qquad \langle ([\gamma], r), ([\gamma^y], d(x, y)) \rangle + f(x) \le f(y) \quad \text{for all } y \in X.$$

Proof. Let $\sigma : [0, R_x] \to X \times \mathbb{R}$ be defined by

$$\sigma(t) = \left(\gamma \left(rt/\sqrt{r^2 + 1} \right), f(x) - t/\sqrt{r^2 + 1} \right) \quad \text{for all } t \in [0, R_x].$$

Taking $u = rR_x/\sqrt{r^2 + 1}$ and $v = R_x/\sqrt{r^2 + 1}$, we have $u^2 + v^2 = R_x^2$ and

$$\sigma(t) = (\gamma(ut/R_x), f(x) - vt/R_x) \quad \text{for all } t \in [0, R_x].$$

By Remark 3.3(v), $([\sigma], 1) \in N_{\mathrm{epi}\, f}(x, f(x))$ if and only if for all $y \in \mathrm{dom}\, f$ and $a \geq 0$,

$$(4.2) \qquad\qquad \langle ([\sigma], R_x), ([\sigma^{y,a}], \delta) \rangle \leq 0,$$

where $\delta = d_2((x, f(x)), (y, f(y) + a))$ and $\sigma^{y,a} : [0, \delta] \to \mathrm{epi}\, f$ is the geodesic from $(x, f(x))$ to $(y, f(y) + a)$.

Denoting $\alpha : [0, v] \to \mathbb{R}$, $\alpha(t) = f(x) - t$, $\gamma^y : [0, d(x, y)] \to X$ the geodesic from x to y, and $\beta : [0, |f(y) + a - f(x)|] \to \mathbb{R}$ the geodesic from $f(x)$ to $f(y) + a$, using Lemma 2.1, inequality (4.2) amounts to

$$\langle ([\gamma], u), ([\gamma^y], d(x, y)) \rangle + \langle ([\alpha], v), ([\beta], |f(y) + a - f(x)|) \rangle \leq 0$$

or, equivalently,

$$(4.3) \qquad\qquad \langle ([\gamma], r), ([\gamma^y], d(x, y)) \rangle + f(x) \leq f(y) + a.$$

Now if $([\gamma], r) \in \partial f(x)$, then taking $a = 0$ in (4.3), we get (4.1). Conversely, if (4.1) holds, then (4.3) holds for all $a \geq 0$, and hence $([\gamma], r) \in \partial f(x)$. □

COROLLARY 4.5. *Let $f : X \to (-\infty, \infty]$ be convex and proper. Then*

$$\mathrm{argmin}\, f = \{ x \in X \mid o_x \in \partial f(x) \}.$$

Proof. Let $x \in X$. Then $x \in \mathrm{argmin}\, f$ if and only if $\langle o_x, ([\gamma^y], d(x, y)) \rangle + f(x) \leq f(y)$ for all $y \in X$. By Proposition 4.4, this is equivalent to $o_x \in \partial f(x)$. □

Remark 4.6. Recently, with the purpose of approaching gradient flows in $\mathrm{CAT}(\kappa)$ spaces from a differential viewpoint, the object minus-subdifferential was introduced in [21, section 3]. In our context and terminology, this definition can be stated as follows: given $f : X \to (-\infty, \infty]$ convex and $x \in \mathrm{dom}\, f$, we say that $([\eta], r) \in T_x X$ belongs to the minus-subdifferential of f at x if

$$(4.4) \qquad -\langle ([\eta], r), ([\gamma^y], d(x, y)) \rangle + f(x) \leq f(y) \quad \text{for all } y \in X,$$

where γ^y is the geodesic from x to y.

Note that if $([\gamma], r), ([\eta], r) \in T_x X$ are opposite to each other and $w \in T_x X$, then we have $\langle ([\gamma], r), w \rangle \leq -\langle ([\eta], r), w \rangle$, with no equality in general. Thus, if $([\eta], r)$ satisfies (4.4), then $([\gamma], r)$ satisfies (4.1), and hence $([\gamma], r) \in \partial f(x)$.

A first step to develop a subdifferential calculus is the description of the subdifferential for functions such as the indicator function, the squared distance to a point, or the distance to a set.

Example 4.7. Let $C \subseteq X$ be nonempty and convex. Consider the indicator function $\delta_C : X \to [0, \infty]$ defined by

$$\delta_C(x) = \begin{cases} 0 & \text{if } x \in C, \\ \infty & \text{if } x \notin C. \end{cases}$$

Then $\partial \delta_C(x) = N_C(x)$ for all $x \in C$.

Proof. Let $x \in C$. Since δ_C is a convex function, by Proposition 4.4, $([\gamma], r) \in \partial\delta_C(x)$ if and only if $\langle([\gamma], r), ([\gamma^y], d(x,y))\rangle \leq \delta_C(y)$ for all $y \in X$, which is equivalent to $\langle([\gamma], r), ([\gamma^y], d(x,y))\rangle \leq 0$ for all $y \in C$, that is, $([\gamma], r) \in N_C(x)$ according to Remark 3.3(v). □

Example 4.8. Suppose additionally that X has diameter smaller than $D_\kappa/2$. Fix $z \in X$. If $x \in X$ with $x \neq z$, denote by Γ_x the set of all nonconstant geodesics $\gamma : [0, l] \to X$ issuing at x with the property that $d(z,x) + l = d(z, \gamma(l))$. By the geodesic extension property around x, $\Gamma_x \neq \emptyset$.

Define $f : X \to \mathbb{R}$ by $f(x) = \frac{1}{2}d(x,z)^2$. Then

$$\partial f(x) = \begin{cases} \{([\gamma], d(z,x)) \in T_x X \mid \gamma \in \Gamma_x\} & \text{if } x \neq z, \\ o_z & \text{if } x = z. \end{cases}$$

Proof. Since X is 2-uniformly convex with parameter $K > 0$, the function f is convex.

Let $x \in X$ and $([\gamma], r) \in \partial f(x)$. We then have that $([\sigma], 1) \in N_{\text{epi } f}(x, f(x))$, where $\sigma : [0, R_x] \to X \times \mathbb{R}$:

$$\sigma(t) = \left(\gamma\left(rt/\sqrt{r^2+1}\right), f(x) - t/\sqrt{r^2+1}\right) \quad \text{for all } t \in [0, R_x].$$

Thus, $P_{\text{epi } f}(\sigma(R_x)) = (x, f(x))$. If $u = R_x/\sqrt{r^2+1}$, then $ur < R_x$. Take $y = (1-a)z + az'$, where $a = 1/(u+1)$ and $z' = \gamma(ur)$. Then, for all $t \in (0,1)$,

$$d_2(\sigma(R_x), (x, f(x)))^2 \leq d_2(\sigma(R_x), (1-t)(x, f(x)) + t(y, f(y)))^2,$$

from where

$$\begin{aligned} d(z', x)^2 + u^2 &\leq d(z', (1-t)x + ty)^2 + (f(x) - u - (1-t)f(x) - tf(y))^2 \\ &\leq (1-t)d(z', x)^2 + td(z', y)^2 - Kt(1-t)d(x,y)^2 \\ &\quad + u^2 - 2tu(f(x) - f(y)) + t^2(f(x) - f(y))^2. \end{aligned}$$

Thus,

$$\begin{aligned} Kt(1-t)d(x,y)^2 &\leq -td(z',x)^2 + t(1-a)^2 d(z',z)^2 - tu\,d(x,z)^2 + tu\,d(y,z)^2 \\ &\quad + t^2(f(x) - f(y))^2. \end{aligned}$$

Dividing by t and then letting $t \searrow 0$, we get

$$\begin{aligned} Kd(x,y)^2 &\leq -d(z',x)^2 + \left((1-a)^2 + ua^2\right)d(z',z)^2 - u\,d(x,z)^2 \\ &\leq \frac{u}{u+1}\left(d(z',x) + d(x,z)\right)^2 - d(z',x)^2 - u\,d(x,z)^2 \\ &= u\left(\frac{1}{u+1}(ur + d(x,z))^2 - ur^2 - d(x,z)^2\right) = -\frac{u^2}{u+1}(r - d(x,z))^2 \leq 0. \end{aligned}$$

From the above inequality, we conclude that $x = y$ and $r = d(x,z)$. If $x = z$, then $r = 0$ and $([\gamma], r) = o_z$. Otherwise, $\gamma \in \Gamma_x$.

Let $x \in X$ with $x \neq z$ and $\gamma \in \Gamma_x$. Taking an extension of γ if needed, denote $r = d(z,x)$, $u = R_x/\sqrt{r^2+1}$, $z' = \gamma(ur)$, and $a = 1/(u+1)$. Then

$$x = (1-a)z + az', \quad d(x, z') = ur, \quad \text{and} \quad d(z, z') = (u+1)r.$$

We prove that $([\gamma], r) \in \partial f(x)$. More precisely, we show that $([\sigma], 1) \in N_{\mathrm{epi}\, f}(x, f(x))$, where $\sigma : [0, R_x] \to X \times \mathbb{R}$:

$$\sigma(t) = \left(\gamma \left(rt/\sqrt{r^2+1} \right), f(x) - t/\sqrt{r^2+1} \right) \quad \text{for all } t \in [0, R_x].$$

Let $(y, \lambda) \in \mathrm{epi}\, f$. Then

$$\begin{aligned}
d_2(\sigma(R_x), (y, \lambda))^2 &= d(z', y)^2 + (f(x) - u - \lambda)^2 \\
&= u^2 + d(z', y)^2 - u\, d(z, x)^2 + 2u\lambda + (f(x) - \lambda)^2 \\
&\geq u^2 + d(z', y)^2 - ur^2 + u\, d(z, y)^2.
\end{aligned}$$

Since $d(z', y) \geq |d(z', z) - d(z, y)| = |(u+1)r - d(z, y)|$, it follows that

$$d(z', y)^2 - ur^2 + u\, d(z, y)^2 \geq u^2 r^2 + (u+1)(r - d(z, y))^2 \geq u^2 r^2,$$

from where

$$d_2(\sigma(R_x), (y, \lambda))^2 \geq u^2 + u^2 r^2 = d(z', x)^2 + u^2 = d_2(\sigma(R_x), (x, f(x)))^2.$$

Hence, $P_{\mathrm{epi}\, f}(\sigma(R_x)) = (x, f(x))$, so $([\sigma], 1) \in N_{\mathrm{epi}\, f}(x, f(x))$.

Finally, we prove that $o_z \in \partial f(z)$, which amounts to showing that $([\sigma], 1) \in N_{\mathrm{epi}\, f}(z, 0)$, where $\sigma : [0, R_x] \to X \times \mathbb{R}$, $\sigma(t) = (z, -t)$ for all $t \in [0, R_x]$. Since $\mathrm{epi}\, f \subseteq X \times \mathbb{R}_+$, this follows by Example 3.5(iv) and Remark 3.3(iii). $\square$

In the above example, the condition that the diameter of X is smaller than $D_\kappa/2$ is used to apply 2-uniform convexity. Thus, when X is a CAT(0) space, its diameter can be unbounded. Moreover, in CAT(0) spaces, the distance function to a convex set is convex and we also have the following example.

Example 4.9. Consider (X, d) a complete, locally compact CAT(0) space with the geodesic extension property. Let $C \subseteq X$ be nonempty, convex, and closed. If $x \in X \setminus C$, denote by Γ_x the set of all geodesics $\gamma : [0, 1] \to X$ issuing at x with the property that $d(P_C(x), x) + 1 = d(P_C(x), \gamma(1))$. By the geodesic extension property, $\Gamma_x \neq \emptyset$.

Define $f : X \to \mathbb{R}$ by $f(x) = \mathrm{dist}(x, C)$. Then

$$\partial f(x) = \begin{cases}
\{([\gamma], 1) \in T_x X \mid \gamma \in \Gamma_x\} & \text{if } x \notin C, \\
\{([\gamma], r) \in N_C(x) \mid r \in [0, 1]\} & \text{if } x \in \mathrm{bd}(C), \\
o_x & \text{if } x \in \mathrm{int}(C).
\end{cases}$$

Proof. Let $x \in X \setminus C$. Denote $z = P_C(x)$ and

$$a = \frac{d(x, z)}{d(x, z) + 1} \in (0, 1).$$

Clearly, $d(x, z) = a/(1 - a)$.

If $([\gamma], r) \in \partial f(x)$, then $([\sigma], 1) \in N_{\mathrm{epi}\, f}(x, f(x))$, where $\sigma : [0, \sqrt{r^2+1}] \to X \times \mathbb{R}$:

$$\sigma(t) = \left(\gamma \left(rt/\sqrt{r^2+1} \right), f(x) - t/\sqrt{r^2+1} \right) \quad \text{for all } t \in [0, \sqrt{r^2+1}].$$

Thus, $P_{\mathrm{epi}\, f}(\sigma(\sqrt{r^2+1})) = (x, f(x))$. Take $y = (1-a)z + a\gamma(r)$. Then, for all $t \in (0, 1)$,

$$d_2(\sigma(\sqrt{r^2+1}), (x, f(x)))^2 \leq d_2(\sigma(\sqrt{r^2+1}), (1-t)(x, f(x)) + t(y, f(y)))^2.$$

As in the proof of Example 4.8, we get that

$$
\begin{aligned}
d(x,y)^2 &\leq (1-a)^2 d(\gamma(r),z)^2 - d(\gamma(r),x)^2 - 2(f(x)-f(y)) \\
&\leq (1-a)^2 d(\gamma(r),z)^2 - d(\gamma(r),x)^2 - 2d(x,z) + 2d(y,z) \\
&= (1-a)^2 d(\gamma(r),z)^2 - d(\gamma(r),x)^2 - 2d(x,z) + 2ad(\gamma(r),z) \\
&\leq (1-a)^2 (d(\gamma(r),x)+d(x,z))^2 - d(\gamma(r),x)^2 - 2d(x,z) \\
&\quad + 2a(d(\gamma(r),x)+d(x,z)) \\
&= (1-a)^2 \left(r+\frac{a}{1-a}\right)^2 - r^2 - 2\frac{a}{1-a} + 2a\left(r+\frac{a}{1-a}\right) \\
&= -a(2-a)(r-1)^2 \leq 0,
\end{aligned}
$$

from where $x=y$ and $r=1$. This shows that $\gamma \in \Gamma_x$.

Now let $\gamma \in \Gamma_x$. We prove next that $([\gamma],1) \in \partial f(x)$. More precisely, we show that $([\sigma],1) \in N_{\mathrm{epi}\,f}(x,f(x))$, where $\sigma:[0,\sqrt{2}] \to X \times \mathbb{R}$:

$$
\sigma(t) = \left(\gamma\left(t/\sqrt{2}\right), f(x)-t/\sqrt{2}\right) \quad \text{for all } t \in [0,\sqrt{2}].
$$

To this end, we prove that for all $(y,\lambda) \in \mathrm{epi}\,f$,

$$
2 = d_2(\sigma(\sqrt{2}),(x,f(x)))^2 \leq d_2(\sigma(\sqrt{2}),(y,\lambda))^2.
$$

Let $(y,\lambda) \in \mathrm{epi}\,f$. Denote $D = d_2(\sigma(\sqrt{2}),(y,\lambda))^2$, $z'=\gamma(1)$, $w=P_C(y)$, and $v=(1-a)w+az'$. Then $x=(1-a)z+az'$, $d(z,z')=1/(1-a)$, $\lambda \geq d(y,w)$, and

$$
\begin{aligned}
d(y,v)^2 &\leq (1-a)d(y,w)^2 + ad(y,z')^2 - a(1-a)d(w,z')^2 \\
&\leq (1-a)d(y,w)^2 + ad(y,z')^2 - a(1-a)d(z,z')^2 \\
&= (1-a)d(y,w)^2 + ad(y,z')^2 - \frac{a}{1-a}.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
D &= d(z',y)^2 + (d(x,z)-1-\lambda)^2 = d(z',y)^2 + \left(\frac{1-2a}{1-a}+\lambda\right)^2 \\
&\geq \frac{1}{a}d(y,v)^2 - \frac{1-a}{a}d(y,w)^2 + \frac{1}{1-a} + \left(\frac{1-2a}{1-a}\right)^2 + 2\frac{1-2a}{1-a}\lambda + \lambda^2.
\end{aligned}
$$

Case I: $a \in (0,1/2)$. Then $1-2a>0$ and we have

$$
\begin{aligned}
D &\geq \frac{1}{a}d(y,v)^2 - \frac{1-a}{a}d(y,w)^2 + \frac{1}{1-a} + \left(\frac{1-2a}{1-a}\right)^2 + 2\frac{1-2a}{1-a}d(y,w) + d(y,w)^2 \\
&= \frac{1}{a}\left(d(y,v)^2 - (d(y,w)-d(x,z))^2\right) + 2\left(d(y,w)-d(x,z)\right)^2 + 2.
\end{aligned}
$$

We show that $|d(y,w)-d(x,z)| \leq d(y,v)$. Taking $p=(1-a)w+az \in C$, we have

$$
\begin{aligned}
d(y,w)-d(x,z) &\leq d(y,p)-d(x,z) \leq d(y,v)+d(v,p)-d(x,z) \\
&\leq d(y,v)+ad(z',z)-d(x,z) = d(y,v).
\end{aligned}
$$

Similarly, taking $q=(1-a)z+aw \in C$, we have

$$
d(x,z)-d(y,w) \leq d(x,q)-d(y,w) \leq ad(z',w)-d(y,w) = d(v,w)-d(y,w) \leq d(v,y).
$$

Thus, $D \geq 2$.

Case II: $a \in [1/2, 1)$. Then $1 - 2a \leq 0$ and we have

$$D \geq \frac{1}{a}d(y,v)^2 - \frac{1-a}{a}\lambda^2 + \frac{1}{1-a} + \left(\frac{1-2a}{1-a}\right)^2 + 2\frac{1-2a}{1-a}\lambda + \lambda^2$$

$$\geq \frac{2a-1}{a}\lambda^2 - 2\frac{2a-1}{1-a}\lambda + \frac{1}{1-a} + \left(\frac{1-2a}{1-a}\right)^2 = \frac{2a-1}{a}\left(\lambda - \frac{a}{1-a}\right)^2 + 2 \geq 2.$$

Consequently, in both cases, $D \geq 2$, which finishes the proof that $([\gamma], 1) \in \partial f(x)$.

Now let $x \in \mathrm{bd}(C)$. Note that $f(x) = 0$.

Suppose that $([\gamma], r) \in \partial f(x)$ with $r > 0$. Then $([\sigma], 1) \in N_{\mathrm{epi}\, f}(x, 0)$, where $\sigma : [0, \sqrt{r^2+1}] \to X \times \mathbb{R}$:

$$\sigma(t) = \left(\gamma\left(rt/\sqrt{r^2+1}\right), -t/\sqrt{r^2+1}\right) \quad \text{for all } t \in [0, \sqrt{r^2+1}].$$

Since $P_{\mathrm{epi}\, f}(\sigma(\sqrt{r^2+1})) = (x, 0)$, it follows that for all $(y, \lambda) \in \mathrm{epi}\, f$,

$$d_2(\sigma(\sqrt{r^2+1}), (x, 0))^2 \leq d_2(\sigma(\sqrt{r^2+1}), (y, \lambda))^2,$$

from where

(4.5) $$d(\gamma(r), x)^2 + 1 \leq d(\gamma(r), y)^2 + (\lambda + 1)^2.$$

Thus, for all $y \in C$, $d(\gamma(r), x) \leq d(\gamma(r), y)$, so $x = P_C(\gamma(r))$, which yields $([\gamma], 1) \in N_C(x)$.

Let $y = (1-t)x + t\gamma(r)$, where $t \in (0, 1)$. Then $f(y) = tr$, $d(\gamma(r), y) = (1-t)r$, and, applying (4.5) for $(y, f(y)) \in \mathrm{epi}\, f$ we obtain

$$r^2 + 1 \leq (1-t)^2 r^2 + (tr + 1)^2.$$

Thus, $r^2(1-t) - r \leq 0$. Letting $t \searrow 0$, this yields $r \leq 1$.

Now take $([\gamma], r) \in N_C(x)$ with $r \in [0, 1]$. Then $P_C(\gamma(r)) = x$. We show that $([\gamma], r) \in \partial f(x)$, i.e, $([\sigma], 1) \in N_{\mathrm{epi}\, f}(x, 0)$, where $\sigma : [0, \sqrt{r^2+1}] \to X \times \mathbb{R}$:

$$\sigma(t) = \left(\gamma\left(rt/\sqrt{r^2+1}\right), -t/\sqrt{r^2+1}\right) \quad \text{for all } t \in [0, \sqrt{r^2+1}].$$

Let $(y, \lambda) \in \mathrm{epi}\, f$. If $d(\gamma(r), y) \geq 1$, then $d(\gamma(r), x) = r \leq 1 \leq d(\gamma(r), y)$. Otherwise, if $d(\gamma(r), y) < 1$, denoting $z = P_C(y)$, we have

$$d(\gamma(r), x)^2 \leq d(\gamma(r), z)^2 \leq (d(\gamma(r), y) + d(y, z))^2 \leq d(\gamma(r), y)^2 + 2d(y, z) + d(y, z)^2$$
$$= d(\gamma(r), y)^2 + (f(y) + 1)^2 - 1 \leq d(\gamma(r), y)^2 + (\lambda + 1)^2 - 1.$$

Therefore, $d_2(\sigma(\sqrt{r^2+1}), (x, 0)) \leq d_2(\sigma(\sqrt{r^2+1}), (y, \lambda))$, so $P_{\mathrm{epi}\, f}(\sigma(\sqrt{r^2+1})) = (x, 0)$ and $([\sigma], 1) \in N_{\mathrm{epi}\, f}(x, 0)$.

Finally, consider $x \in \mathrm{int}(C)$. Again, $f(x) = 0$. If $([\gamma], r) \in \partial f(x)$, as before, one has that $x = P_C(\gamma(r))$. Since $x \in \mathrm{int}(C)$, this implies that $r = 0$. Conversely, to show that $o_x \in \partial f(x)$, one argues as in Example 4.8. $\square$

Remark 4.10. In Example 4.9, if $C = \{z\}$ for some $z \in X$, then $f : X \to \mathbb{R}$, $f(x) = d(x, z)$, and, for $x \neq z$, Γ_x is the set of all geodesics $\gamma : [0, 1] \to X$ issuing at x with the property that $d(z, x) + 1 = d(z, \gamma(1))$. We then have

$$\partial f(x) = \begin{cases} \{([\gamma], 1) \in T_x X \mid \gamma \in \Gamma_x\} & \text{if } x \neq z, \\ \{([\gamma], r) \in T_x X \mid r \in [0, 1]\} & \text{if } x = z. \end{cases}$$

If $C = \overline{B}(z, R)$ for some $z \in X$ and $R > 0$, then $f : X \to \mathbb{R}$, $f(x) = \mathrm{dist}(x, \overline{B}(z, R))$, and, for $x \notin B(z, R)$, Γ_x is the set of all geodesics $\gamma : [0, 1] \to X$ issuing at x with the property that $d(z, x) + 1 = d(z, \gamma(1))$. We then have

$$\partial f(x) = \begin{cases} \{([\gamma], 1) \in T_x X \mid \gamma \in \Gamma_x\} & \text{if } d(x, z) > R, \\ \{([\gamma], r) \in T_x X \mid r \in [0, 1], \gamma \in \Gamma_x\} & \text{if } d(x, z) = R, \\ o_x & \text{if } d(x, z) < R. \end{cases}$$

We show now that, as in the linear setting, the subdifferential of a continuous convex function is nonempty.

THEOREM 4.11. *Consider $f : X \to (-\infty, \infty]$ a convex function, and let $x \in \mathrm{dom}\, f$. If f is continuous at x, then $\partial f(x) \neq \emptyset$.*

Proof. Since f is continuous at x, there exists $\varepsilon \in (0, D_\kappa/2)$ so that f is continuous on $\overline{B}(x, \varepsilon)$.

Suppose $N_{\mathrm{epi}\, f}(x, f(x)) = \{o_{(x, f(x))}\}$, and denote $B = \overline{B}((x, f(x)), \varepsilon)$. Using Remark 3.3(iv), $N_{\mathrm{epi}\, f \cap B}(x, f(x)) = \{o_{(x, f(x))}\}$.

Note that the set $\mathrm{epi}\, f \cap B$ is closed. Indeed, let $(y_n, s_n) \subseteq \mathrm{epi}\, f \cap B$ be a sequence whose limit is (y, s). Clearly, $(y, s) \in B$. Moreover, for all $n \in \mathbb{N}$, $d(y_n, x) \leq d_2((y_n, s_n), (x, f(x))) \leq \varepsilon$, so $y \in \overline{B}(x, \varepsilon)$ and f is continuous at y. As $y_n \to y$, $s_n \to s$, and $f(y_n) \leq s_n$, we get $f(y) \leq s$, and hence $(y, s) \in \mathrm{epi}\, f$.

Applying Proposition 3.4(ii), we conclude that $(x, f(x)) \in \mathrm{int}(\mathrm{epi}\, f \cap B)$, which is a contradiction because $(x, f(x)) \in \mathrm{bd}(\mathrm{epi}\, f)$.

Thus, there exists $([\sigma], 1) \in N_{\mathrm{epi}\, f}(x, f(x))$, where $\sigma : [0, R_x] \to X \times \mathbb{R}$ is a nonconstant geodesic. Write $R_x^2 = r^2 + s^2$ and $\sigma : [0, R_x] \to X \times \mathbb{R}$ as $\sigma(t) = (\gamma(rt/R_x), \alpha(st/R_x))$ for all $t \in [0, R_x]$, where $\gamma : [0, r] \to X$ and $\alpha : [0, s] \to \mathbb{R}$ are geodesics. Then $\alpha(t) = f(x) + t$ for all $t \in [0, s]$ or $\alpha(t) = f(x) - t$ for all $t \in [0, s]$.

For $y \in \mathrm{dom}\, f$ and $a \geq 0$, consider $\gamma^y : [0, d(x, y)] \to X$ the geodesic from x to y and $\beta : [0, |f(y) + a - f(x)|] \to \mathbb{R}$ the geodesic from $f(x)$ to $f(y) + a$. As in the proof of Proposition 4.4, we have

$$\langle ([\gamma], r), ([\gamma^y], d(x, y)) \rangle + \langle ([\alpha], s), ([\beta], |f(y) + a - f(x)|) \rangle \leq 0.$$

Suppose first that $s = 0$. Then $r > 0$ and $\langle ([\gamma], r), ([\gamma^y], d(x, y)) \rangle \leq 0$ for all $y \in \mathrm{dom}\, f$. If $y \in \gamma((0, \min\{\varepsilon, r\}])$, then $y \in \mathrm{dom}\, f$ and $\angle(\gamma, \gamma^y) = 0$, so $0 < rd(x, y) \cos \angle(\gamma, \gamma^y) = \langle ([\gamma], r), ([\gamma^y], d(x, y)) \rangle \leq 0$, a contradiction. Thus, $s > 0$. We distinguish the following two cases.

Case I: $\alpha(t) = f(x) + t$ for all $t \in [0, s]$. Then

$$\langle ([\gamma], r), ([\gamma^y], d(x, y)) \rangle + s(f(y) + a - f(x)) \leq 0$$

for any $y \in \mathrm{dom}\, f$ and $a \geq 0$. Taking $y = x$ and $a = 1$, we obtain again a contradiction.

Case II: $\alpha(t) = f(x) - t$ for all $t \in [0, s]$. Denoting $u = r/s$, we have $r/R_x = u/\sqrt{u^2 + 1}$, $s/R_x = 1/\sqrt{u^2 + 1}$. Hence,

$$\sigma(t) = \left(\gamma(ut/\sqrt{u^2 + 1}), f(x) - t/\sqrt{u^2 + 1} \right) \quad \text{for all } t \in [0, R_x]$$

and we conclude that $([\gamma], u) \in \partial f(x)$. $\qquad \square$

Remark 4.12. If $f : X \to (-\infty, \infty]$ is a convex function and $\partial f(x) \neq \emptyset$ for some $x \in \mathrm{dom}\, f$ (e.g., if f is continuous at x), then f has bounded decay rate in the sense

that there exists a constant $c \geq 0$ such that the function $y \mapsto f(y) + cd(x,y)$ is bounded below.

Indeed, if there exists $([\gamma], c) \in \partial f(x)$, applying Proposition 4.4, we get that

$$f(y) \geq f(x) + \langle ([\gamma], c), ([\gamma^y], d(x,y)) \rangle \geq f(x) - cd(x,y)$$

for all $y \in X$.

In what follows, we suppose additionally that X has diameter smaller than $D_\kappa/2$. We first recall the following result (see, e.g., [15, Chapter II, Corollary 3.6]).

PROPOSITION 4.13 (first variation formula). *Let $x, y \in X$ with $x \neq y$, $\gamma : [0, l] \to X$ be a geodesic issuing at x and η the geodesic from x to y. Then*

$$\lim_{t \searrow 0} \frac{d(\gamma(tl), y) - d(x,y)}{t} = -\frac{1}{d(x,y)} \langle ([\gamma], l), ([\eta], d(x,y)) \rangle.$$

Remark 4.14. Let $x, y \in X$ with $x \neq y$, $\gamma : [0, l] \to X$ be a geodesic issuing at x and η be the geodesic from x to y. Then

$$\langle ([\gamma], l), ([\eta], d(x,y)) \rangle \geq \frac{1}{2} \left(d(x,y)^2 - d(\gamma(l), y)^2 \right).$$

Indeed, since the function $d(y, \cdot)^2$ is convex, we have that for all $t \in (0, 1)$,

$$d(y, \gamma(tl))^2 \leq (1-t)d(y,x)^2 + td(y, \gamma(l))^2,$$

from where

$$\begin{aligned}
(d(y, \gamma(tl)) - d(y,x))\,(d(y, \gamma(tl)) + d(y,x)) &= d(y, \gamma(tl))^2 - d(y,x)^2 \\
&\leq t(d(y, \gamma(l))^2 - d(y,x)^2).
\end{aligned}$$

One obtains the conclusion after dividing by t, letting $t \searrow 0$, and applying Proposition 4.13.

Consider $f : X \to (-\infty, \infty]$ a proper and convex function that has complete sublevel sets.

Remark 4.15. Note first that f is lower semicontinuous and that all its sublevel sets are convex. Moreover, by the Hopf–Rinow theorem, we deduce that all sublevel sets of f are also compact. Thus, f attains its minimum and hence is bounded below.

Indeed, denote $a = \inf f(X)$ and let $(x_n) \subseteq X$ such that $f(x_n) \to a$. Take $y \in \operatorname{dom} f$. If $f(y) = a$, then f attains its minimum at y. Otherwise, $a < f(y)$, so for n sufficiently large, x_n belongs to the compact set $\{x \in X \mid f(x) \leq f(y)\}$. Thus, there exists a subsequence (x_{n_k}) of (x_n) that converges to some $x \in X$. By the lower semicontinuity of f, we obtain that $f(x) = a$.

The *resolvent* of f is the mapping $J_\lambda^f : X \to \operatorname{dom} f$ defined for $\lambda > 0$ by

$$(4.6) \qquad J_\lambda^f(x) := \underset{y \in X}{\operatorname{argmin}} \left(f(y) + \frac{1}{2\lambda}d(x,y)^2 \right) \quad \text{for all } x \in X.$$

Observe that there exists a unique point attaining the minimum in (4.6). Indeed, existence follows because the sublevel sets of the function $y \in X \mapsto f(y) + \frac{1}{2\lambda}d(x,y)^2$ are closed subsets of sublevel sets of the function f and hence are compact. Uniqueness follows by convexity of f and by 2-uniform convexity of X.

In the following proposition, we establish a connection between the resolvent of f and its subdifferential.

PROPOSITION 4.16. *Let $\lambda > 0$, and let $z \in \operatorname{dom} f$. For $x \in X$, denote by γ^x the geodesic from z to x. Given $x \in X$, we have that $z = J_\lambda^f(x)$ if and only if $([\gamma^x], d(x,z)/\lambda) \in \partial f(z)$.*

Proof. Note that, by Proposition 4.4, $([\gamma^x], d(x,z)/\lambda) \in \partial f(z)$ if and only if $([\gamma^x], d(z,x)) \in \partial(\lambda f)(z)$.

Suppose first that $z = J_\lambda^f(x)$, and denote $r = d(z,x)$. We prove that $([\sigma], 1) \in N_{\operatorname{epi}(\lambda f)}(z, \lambda f(z))$, where $\sigma : [0, \sqrt{r^2+1}] \to X \times \mathbb{R}$:

$$\sigma(t) = \left(\gamma^x \left(rt/\sqrt{r^2+1} \right), \lambda f(z) - t/\sqrt{r^2+1} \right).$$

Let $(y, \beta) \in \operatorname{epi}(\lambda f)$. Then $y \in \operatorname{dom} f$, $\beta \geq \lambda f(y)$, and we have

$$\begin{aligned}
d_2(\sigma(\sqrt{r^2+1}), (y,\beta))^2 &= d_2((x, \lambda f(z) - 1), (y,\beta))^2 = d(x,y)^2 + (\lambda f(z) - 1 - \beta)^2 \\
&= 1 + d(x,y)^2 - 2\lambda f(z) + 2\beta + (\lambda f(z) - \beta)^2 \\
&\geq 1 + d(x,y)^2 + 2\lambda(f(y) - f(z)) \\
&\geq 1 + d(z,x)^2 = d_2(\sigma(\sqrt{r^2+1}), (z, \lambda f(z)))^2 \quad \text{by (4.6)}.
\end{aligned}$$

Hence, $P_{\operatorname{epi}(\lambda f)}(\sigma(\sqrt{r^2+1})) = (z, \lambda f(z))$, so $([\sigma], 1) \in N_{\operatorname{epi}(\lambda f)}(z, \lambda f(z))$. This shows that $([\gamma^x], d(z,x)) \in \partial(\lambda f)(z)$.

Suppose now that $([\gamma^x], d(z,x)) \in \partial(\lambda f)(z)$, and let $y \in X$. By Proposition 4.4,

$$\langle ([\gamma^x], d(z,x)), ([\gamma^y], d(z,y)) \rangle \leq \lambda(f(y) - f(z)).$$

Applying Remark 4.14 (with $\gamma = \gamma^y$ and $\eta = \gamma^x$), we get

$$\frac{1}{2} \left(d(z,x)^2 - d(y,x)^2 \right) \leq \langle ([\gamma^y], d(z,y)), ([\gamma^x], d(z,x)) \rangle.$$

The above two inequalities yield that

$$f(z) + \frac{1}{2\lambda} d(x,z)^2 \leq f(y) + \frac{1}{2\lambda} d(x,y)^2.$$

We conclude that $z = J_\lambda^f(x)$. □

Example 4.17. Let $C \subseteq X$ be nonempty, convex, and complete, and let $z \in C$ and $\lambda > 0$. Then $\partial \delta_C(z) = N_C(z)$ and $J_\lambda^{\delta_C} = P_C$. In this case, given $x \in X$, Proposition 4.16 reduces to $z = P_C(x)$ if and only if $([\gamma^x], d(x,z)/\lambda) \in N_C(z)$.

We give next a necessary condition for a point to be a minimizer of the sum of two convex functions starting with the following two lemmas.

LEMMA 4.18. *Let $f : X \to (-\infty, \infty]$ be convex. Given $x \in \operatorname{dom} f$ and $\lambda \geq 0$, define the function $\varphi : X \to (-\infty, \infty]$ by $\varphi(y) = f(y) + \lambda d(y,x)^2$. Then $\partial\varphi(x) = \partial f(x)$.*

Proof. For $y \in X$, denote by γ^y the geodesic from x to y. Let $([\gamma], r) \in \partial f(x)$. We apply Proposition 4.4 and have that for all $y \in X$,

$$\langle ([\gamma], r), ([\gamma^y], d(x,y)) \rangle + \varphi(x) = \langle ([\gamma], r), ([\gamma^y], d(x,y)) \rangle + f(x) \leq f(y) \leq \varphi(y).$$

Therefore, $([\gamma], r) \in \partial\varphi(x)$.

Conversely, suppose that $([\gamma], r) \in \partial\varphi(x)$. Take $y \in X$ with $y \neq x$ and $z = (1-t)x + ty$ for some $t \in (0,1)$. Then

$$\langle ([\gamma], r), ([\gamma^y], d(x,z)) \rangle + f(x) = \langle ([\gamma], r), ([\gamma^z], d(x,z)) \rangle + \varphi(x)$$
$$\leq \varphi(z) = f(z) + \lambda d(z,x)^2,$$

from where

$$t\langle([\gamma], r), ([\gamma^y], d(x, y))\rangle + f(x) \le (1 - t)f(x) + tf(y) + \lambda t^2 d(x, y)^2.$$

After rearranging, dividing by t, and then letting $t \searrow 0$, we obtain that

$$\langle([\gamma], r), ([\gamma^y], d(x, y))\rangle + f(x) \le f(y).$$

Since the above inequality clearly also holds for $y = x$, we deduce that $([\gamma], r) \in \partial f(x)$. □

LEMMA 4.19. *Let $f : X \to (-\infty, \infty]$ be convex, let $x \in \text{dom} f$, and let $R \in (0, R_x)$. Suppose that f is Lipschitz on $B(x, R)$ with Lipschitz constant L. If $y \in B(x, R/2)$ and $([\gamma], r) \in \partial f(y)$, then $r \le L$.*

Proof. Suppose $r > 0$, and let $z = \gamma(R/2)$. Then $z \in B(x, R)$ and we have

$$rR/2 = \langle([\gamma], r), ([\gamma], R/2)\rangle \le f(z) - f(y) \le Ld(z, y) = LR/2,$$

from where $r \le L$. □

THEOREM 4.20. *Let $f, g : X \to (-\infty, \infty]$ be two convex functions that have complete sublevel sets. If $x \in \text{dom} f \cap \text{dom} g$ is a minimum point of $f + g$ and f is continuous at x, then there exist $([\gamma], r) \in \partial f(x)$ and $([\eta], s) \in \partial g(x)$ such that $([\gamma], r)$ and $([\eta], s)$ are opposite to each other.*

Proof. Given $n \in \mathbb{N}$, define $h_n : X \times X \to (-\infty, \infty]$ by

$$h_n(y, z) = f(y) + g(z) + d(y, x)^2 + \frac{n}{2}d(y, z)^2.$$

Denote $\alpha = \inf f(X) > -\infty$ and $\beta = \inf g(X) > -\infty$. For any $a \in \mathbb{R}$,

$$\{(y, z) \in X \times X \mid h_n(y, z) \le a\} \subseteq \{y \in X \mid f(y) \le a - \beta\} \times \{z \in X \mid g(z) \le a - \alpha\}.$$

Since both f and g have compact sublevel sets, we deduce that h_n has compact sublevel sets and so attains its minimum at some point $(y_n, z_n) \in X \times X$. Observe that for all $n \in \mathbb{N}$,

$$(4.7) \quad f(y_n) + g(z_n) + d(y_n, x)^2 + \frac{n}{2}d(y_n, z_n)^2 = h_n(y_n, z_n) \le h_n(x, x) = f(x) + g(x),$$

from where

$$y_n \in \{y \in X \mid f(y) \le f(x) + g(x) - \beta\} \quad \text{and} \quad z_n \in \{z \in X \mid g(z) \le f(x) + g(x) - \alpha\}.$$

We can therefore suppose, after taking subsequences, that (y_n) and (z_n) converge to some $y^* \in X$ and $z^* \in X$, respectively. By (4.7),

$$\frac{n}{2}d(y_n, z_n)^2 \le f(x) + g(x) - \alpha - \beta$$

for all $n \in \mathbb{N}$ and we conclude that $d(y_n, z_n) \to 0$. This yields that $y^* = z^*$. Moreover, using again (4.7) and the lower semicontinuity of the functions f and g, we obtain

$$f(y^*) + g(y^*) + d(y^*, x)^2 \le f(x) + g(x) \le f(y^*) + g(y^*),$$

and hence $y^* = x$.

Because f is continuous at x, we have that $f(y_n) \to f(x)$. Taking limit superior as $n \to \infty$ in (4.7) gives $\limsup_{n \to \infty} g(z_n) \leq g(x)$. Combining this with the lower semicontinuity of g, we get that $g(z_n) \to g(x)$.

For all $n \in \mathbb{N}$, denote by γ_n and η_n the geodesics from y_n to z_n and from z_n to y_n, respectively.

Define $\varphi : X \to (-\infty, \infty]$ by $\varphi(y) = f(y) + d(y,x)^2$. Note that φ is convex and continuous at x and hence Lipschitz on a neighborhood of x by Lemma 4.3. For all $n \in \mathbb{N}$ and all $y \in X$,

$$\varphi(y_n) + g(z_n) + \frac{n}{2}d(y_n,z_n)^2 = h_n(y_n,z_n) \leq h_n(y,z_n) = \varphi(y) + g(z_n) + \frac{n}{2}d(y,z_n)^2,$$

so $y_n = J_{1/n}^{\varphi}(z_n)$. Similarly, we get that $z_n = J_{1/n}^{g}(y_n)$. Proposition 4.16 yields $([\gamma_n], nd(y_n,z_n)) \in \partial\varphi(y_n)$ and $([\eta_n], nd(y_n,z_n)) \in \partial g(z_n)$.

If $y_n = z_n$ for some $n \in \mathbb{N}$, we obtain that $y_n = x$. Thus, $o_x \in \partial\varphi(x)$ and $o_x \in \partial g(x)$. By Lemma 4.18, $o_x \in \partial f(x)$ and the conclusion follows.

Assume next that $y_n \neq z_n$ for all $n \in \mathbb{N}$. For $n \in \mathbb{N}$ sufficiently large, extend γ_n beyond z_n to a geodesic $\overline{\gamma}_n$ of length R_x. Likewise, extend η_n beyond y_n to a geodesic $\overline{\eta}_n$ of length R_x. By the betweenness property, $\mathrm{Im}\,\overline{\gamma}_n \cup \mathrm{Im}\,\overline{\eta}_n$ is a geodesic segment.

Denote $r_n = nd(y_n,z_n)$ for $n \in \mathbb{N}$. As $([\gamma_n], r_n) \in \partial\varphi(y_n)$, using Lemma 4.19, we get that (r_n) is bounded, so it has a convergent subsequence. By local compactness, for $l > 0$ small enough, $(\overline{\gamma}_n(l))$ and $(\overline{\eta}_n(l))$ have convergent subsequences as well. We can assume that $\overline{\gamma}_n(l) \to u \in X$, $\overline{\eta}_n(l) \to v \in X$, and $r_n \to r \geq 0$. Since $y_n \to x$ and $z_n \to x$, we deduce that $x = (1/2)u + (1/2)v$. Denote by γ and η the geodesics from x to u and from x to v, respectively. Note that $([\gamma], r)$ and $([\eta], r)$ are opposite to each other.

We show now that $([\gamma], r) \in \partial f(x)$. For n large enough, take $\sigma_n : [0,l] \to X \times \mathbb{R}$:

$$\sigma_n(t) = \left(\overline{\gamma}_n\left(r_n t/\sqrt{r_n^2+1}\right), \varphi(y_n) - t/\sqrt{r_n^2+1}\right).$$

Note that $\sigma_n(t) \to \sigma(t)$ for all $t \in [0,l]$, where $\sigma : [0,l] \to X \times \mathbb{R}$ is defined by

$$\sigma(t) = \left(\gamma\left(rt/\sqrt{r^2+1}\right), \varphi(x) - t/\sqrt{r^2+1}\right).$$

Because $([\gamma_n], r_n) \in \partial\varphi(y_n)$, we have that $([\sigma_n], 1) \in N_{\mathrm{epi}\,\varphi}(y_n, \varphi(y_n))$. By Remark 3.3(vi), we obtain that $([\sigma], 1) \in N_{\mathrm{epi}\,\varphi}(x, \varphi(x))$, so $([\gamma], r) \in \partial\varphi(x)$. Applying Lemma 4.18, $([\gamma], r) \in \partial f(x)$. Similarly, using the fact that $g(z_n) \to g(x)$, one can show that $([\eta], r) \in \partial g(x)$. $\qquad\square$

Remark 4.21. Note that if $f, g : X \to (-\infty, \infty]$ are proper, convex, and with complete sublevel sets, then $f + g$ attains its minimum since its sublevel sets are closed subsets of sublevel sets of one of the functions and hence are compact.

Observe also that, as pointed out in section 1, in the classical version of Theorem 4.20, f only needs to be continuous at some point in the domain of g which is not necessarily a minimizer of $f + g$.

Remark 4.22. Theorem 4.20 also holds in unbounded CAT(0) spaces. More precisely, suppose that X is a complete, locally compact CAT(0) space with the geodesic extension property, and let $f, g : X \to (-\infty, \infty]$ be two proper, convex, and lower semicontinuous functions. First, observe that even though the sublevel sets of f and g are not necessarily compact, their resolvents are well-defined (see, e.g., [22, 30]). We point out below how one could modify the argument from the proof of

Theorem 4.20 in order to show that if $x \in \operatorname{dom} f \cap \operatorname{dom} g$ is a minimum point of $f+g$, f is continuous at x, and $\partial g(\overline{x}) \neq \emptyset$ for some $\overline{x} \in \operatorname{dom} g$, then there exist $([\gamma], r) \in \partial f(x)$ and $([\eta], s) \in \partial g(x)$ such that $([\gamma], r)$ and $([\eta], s)$ are opposite to each other.

Indeed, note first that, by Remark 4.12, f and g have bounded decay rate. Thus, there exist the constants $c_1, c_2 \geq 0$ and $k_1, k_2 \in \mathbb{R}$ such that $f(y) + c_1 d(y, x) \geq k_1$ and $g(y) + c_2 d(y, \overline{x}) \geq k_2$ for all $y \in X$. Take

$$c = \max\{c_1, c_2\} \quad \text{and} \quad k = \min\{k_1, k_2 - c_2 d(x, \overline{x})\}.$$

Then, applying the triangle inequality, for all $y \in X$ we have

$$f(y) + cd(y, x) \geq k \quad \text{and} \quad g(y) + cd(y, x) \geq k.$$

Given $n \in \mathbb{N}$, define $h_n : X \times X \to (-\infty, \infty]$ by

$$h_n(y, z) = f(y) + d(y, x)^2 + g(z) + d(z, x)^2 + \frac{n}{2} d(y, z)^2.$$

Then

$$h_n(y, z) \geq k - cd(y, x) + d(y, x)^2 + k - cd(z, x) + d(z, x)^2 + \frac{n}{2} d(y, z)^2$$

$$\geq \left(d(y, x) - \frac{c}{2} \right)^2 + \left(d(z, x) - \frac{c}{2} \right)^2 + 2 \left(k - \frac{c^2}{4} \right)$$

for all $y, z \in X$. Let $b = 2(k - c^2/4)$. Given $a \in \mathbb{R}$ with $a \geq b$, denote

$$Y_a = \{y \in X \mid (d(y, x) - c/2)^2 \leq a - b\} \quad \text{and} \quad Z_a = \{z \in X \mid (d(z, x) - c/2)^2 \leq a - b\}.$$

Then Y_a and Z_a are bounded and closed and hence compact. Since

$$\{(y, z) \in X \times X \mid h_n(y, z) \leq a\} \subseteq Y_a \times Z_a,$$

we conclude that the sublevel sets of h_n are compact, so h_n attains its minimum at some $(y_n, z_n) \in X \times X$. We then get that $y_n \in Y_{f(x)+g(x)}$ and $z_n \in Z_{f(x)+g(x)}$ for all $n \in \mathbb{N}$. Therefore, we can suppose, after taking subsequences, that (y_n) and (z_n) converge to some $y^* \in X$ and $z^* \in X$. Similarly as before, one can show that $y^* = z^* = x$, and considering the functions $\varphi, \psi : X \to (-\infty, \infty]$ defined by $\varphi(y) = f(y) + d(y, x)^2$ and $\psi(z) = g(z) + d(z, x)^2$, one writes $h_n(y, z) = \varphi(y) + \psi(z) + \frac{n}{2} d(y, z)^2$ and proceeds as in the proof of Theorem 4.20.

We finish with a consequence of the above result which, in a finite-dimensional Hilbert space, yields the existence of a separating hyperplane for two nonempty, convex, closed, and disjoint sets, at least one of which is bounded. It also provides a version in $\mathrm{CAT}(0)$ spaces of a recent separation result for convex sets proved in [29, Theorem 4.5] in the setting of Hadamard manifolds.

COROLLARY 4.23. *Let (X, d) be a complete, locally compact $\mathrm{CAT}(0)$ space with the geodesic extension property. If $C, D \subseteq X$ are nonempty, convex, closed, and disjoint and at least one of them is bounded, then there exist $x \in C$ and $([\gamma], 1), ([\eta], 1) \in T_x X$ opposite to each other such that*

$$\sup\{\langle ([\gamma], 1), ([\gamma^y], d(x, y)) \rangle \mid y \in D\} < 0$$

and

$$\sup\{\langle ([\eta], 1), ([\gamma^y], d(x, y)) \rangle \mid y \in C\} \leq 0,$$

where γ^y denotes the geodesic from x to y.

Proof. Define $f, g : X \to (-\infty, \infty]$ by $f(x) = \operatorname{dist}(x, D)$ and $g(x) = \delta_C(x)$ for all $x \in X$. These functions are proper and convex. Moreover, f is continuous, while g is lower semicontinuous.

Denote by ρ the distance between the sets C and D. For $n \in \mathbb{N}$, choose $x_n \in C$ and $y_n \in D$ such that $d(x_n, y_n) < \rho + 1/n$. Since one of the sets C or D is bounded, one of the sequences (x_n) or (y_n) is bounded, and hence both sequences are bounded. By the Hopf–Rinow theorem, there exists $x \in C$ such that $d(x, P_D(x)) = \rho$.

Since x is a minimum point of $f+g$, we can use Remark 4.22 and Examples 4.7 and 4.9 to obtain $([\gamma], 1), ([\eta], 1) \in T_x X$ opposite to each other such that $([\gamma], 1) \in \partial f(x)$ and $([\eta], 1) \in N_C(x)$. Applying Proposition 4.4, $\langle ([\gamma], 1), ([\gamma^y], d(x,y)) \rangle \leq -f(x)$ for all $y \in D$, which yields the first desired inequality. The second one follows by using Remark 3.3(v). □

REFERENCES

[1] B. Ahmadi Kakavandi and M. Amini, *Duality and subdifferential for convex functions on complete* CAT(0) *metric spaces*, Nonlinear Anal., 73 (2010), pp. 3450–3455, https://doi.org/10.1016/j.na.2010.07.033.

[2] A. D. Aleksandrov, V. N. Berestovskiĭ, and I. G. Nikolaev, *Generalized Riemannian spaces*, Russian Math. Surveys, 41 (1986), pp. 1–54, https://doi.org/10.1070/RM1986v041n03ABEH003311.

[3] S. Alexander, V. Kapovitch, and A. Petrunin, *Alexandrov Geometry: Foundations*, preprint, https://arxiv.org/abs/1903.08539, 2023.

[4] L. Ambrosio, N. Gigli, and G. Savaré, *Gradient Flows in Metric Spaces and in the Space of Probability Measures*, Lectures Math. ETH Zürich, Birkhäuser, Basel, 2005.

[5] D. Ariza-Ruiz, A. Fernández-León, G. López-Acedo, and A. Nicolae, *Chebyshev sets in geodesic spaces*, J. Approx. Theory, 207 (2016), pp. 265–282, https://doi.org/10.1016/j.jat.2016.02.019.

[6] M. Bačák, *Convex Analysis and Optimization in Hadamard Spaces*, De Gruyter, Berlin, 2014.

[7] M. Bačák, *Old and new challenges in Hadamard spaces*, Jpn. J. Math., 18 (2023), pp. 115–168, https://doi.org/10.1007/s11537-023-1826-0.

[8] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, 2nd ed., Springer, Cham, 2017.

[9] I. D. Berg and I. G. Nikolaev, *Quasilinearization and curvature of Alexandrov spaces*, Geom. Dedicata, 133 (2008), pp. 195–218, https://doi.org/10.1007/s10711-008-9243-3.

[10] R. Bergmann, J. Persch, and G. Steidl, *A parallel Douglas–Rachford algorithm for minimizing ROF-like functionals on images with values in symmetric Hadamard manifolds*, SIAM J. Imaging Sci., 9 (2016), pp. 901–937, https://doi.org/10.1137/15M1052858.

[11] D. Bertsekas, *Convex Optimization Algorithms*, Athena Scientific, Belmont, MA, 2015.

[12] L. J. Billera, S. P. Holmes, and K. Vogtmann, *Geometry of the space of phylogenetic trees*, Adv. in Appl. Math., 27 (2001), pp. 733–767, https://doi.org/10.1006/aama.2001.0759.

[13] N. Boumal, *An Introduction to Optimization on Smooth Manifolds*, Cambridge University Press, Cambridge, UK, 2023.

[14] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, *Distributed optimization and statistical learning via the alternating direction method of multipliers*, Found. Trends Mach. Learn., 3 (2011), pp. 1–122, https://doi.org/10.1561/2200000016.

[15] M. R. Bridson and A. Haefliger, *Metric Spaces of Non-positive Curvature*, Springer-Verlag, Berlin, 1999.

[16] D. Burago, Y. Burago, and S. Ivanov, *A Course in Metric Geometry*, American Mathematical Society, Providence, RI, 2001.

[17] P. Chaipunya, F. Kohsaka, and P. Kumam, *Monotone vector fields and generation of nonexpansive semigroups in complete* CAT(0) *spaces*, Numer. Funct. Anal. Optim., 42 (2021), pp. 989–1018, https://doi.org/10.1080/01630563.2021.1931879.

[18] F. H. Clarke, Y. S. Ledyaev, R. J. Stern, and P. R. Wolenski, *Nonsmooth Analysis and Control Theory*, Springer-Verlag, New York, 1998.

[19] C. R. Diminnie and A. G. White, *Remarks on strict convexity and betweenness postulates*, Demonstrat. Math., 14 (1981), pp. 209–220.

[20] R. Espínola and A. Fernández-León, *CAT(k)-spaces, weak convergence and fixed points*, J. Math. Anal. Appl., 353 (2009), pp. 410–427, https://doi.org/10.1016/j.jmaa.2008.12.015.

[21] N. Gigli and F. Nobili, *A differential perspective on gradient flows on CAT(κ)-spaces and applications*, J. Geom. Anal., 31 (2021), pp. 11780–11818, https://doi.org/10.1007/s12220-021-00701-5.

[22] J. Jost, *Convex functionals and generalized harmonic maps into spaces of nonpositive curvature*, Comment. Math. Helv., 70 (1995), pp. 659–673, https://doi.org/10.1007/BF02566027.

[23] U. Kohlenbach, G. López-Acedo, and A. Nicolae, *A uniform betweenness property in metric spaces and its role in the quantitative analysis of the "Lion-Man" game*, Pacific J. Math., 310 (2021), pp. 181–212, https://doi.org/10.2140/pjm.2021.310.181.

[24] K. Kuwae, *Jensen's inequality on convex spaces*, Calc. Var. Partial Differential Equations, 49 (2014), pp. 1359–1378, https://doi.org/10.1007/s00526-013-0625-5.

[25] U. Lang and V. Schroeder, *Kirszbraun's theorem and metric spaces of bounded curvature*, Geom. Funct. Anal., 7 (1997), pp. 535–560, https://doi.org/10.1007/s000390050018.

[26] C. Li, G. López, and V. Martín-Márquez, *Monotone vector fields and the proximal point algorithm on Hadamard manifolds*, J. Lond. Math. Soc. (2), 79 (2009), pp. 663–683, https://doi.org/10.1112/jlms/jdn087.

[27] C. Li, V. Martín-Márquez, and J.-H. Wang, *Resolvents of set-valued monotone vector fields in Hadamard manifolds*, Set-Valued Var. Anal., 19 (2011), pp. 361–383, https://doi.org/10.1007/s11228-010-0169-1.

[28] C. Li, B. S. Mordukhovich, J. Wang, and J.-C. Yao, *Weak sharp minima on Riemannian manifolds*, SIAM J. Optim., 21 (2011), pp. 1523–1560, https://doi.org/10.1137/09075367X.

[29] M. S. Louzeiro, R. Bergmann, and R. Herzog, *Fenchel duality and a separation theorem on Hadamard manifolds*, SIAM J. Optim., 32 (2022), pp. 854–873, https://doi.org/10.1137/21M1400699.

[30] U. Mayer, *Gradient flows on nonpositively curved metric spaces and harmonic maps*, Comm. Anal. Geom., 6 (1998), pp. 199–253, https://doi.org/10.4310/CAG.1998.v6.n2.a1.

[31] B. S. Mordukhovich, *Variational Analysis and Generalized Differentiation,* I. *Basic Theory*, Springer-Verlag, Berlin, 2006.

[32] B. S. Mordukhovich and N. M. Nam, *Convex Analysis and Beyond. Volume* 1*: Basic Theory*, Springer, Cham, 2022.

[33] M. Movahedi, D. Behmardi, and S. Hosseini, *On the density theorem for the subdifferential of convex functions on Hadamard spaces*, Pacific J. Math., 276 (2015), pp. 437–447, https://doi.org/10.2140/pjm.2015.276.437.

[34] A. Nicolae, *Asymptotic behavior of averaged and firmly nonexpansive mapping in geodesic spaces*, Nonlinear Anal., 87 (2013), pp. 102–115, https://doi.org/10.1016/j.na.2013.03.018.

[35] I. Nikolaev, *The tangent cone of an Aleksandrov space of curvature ≤ k*, Manuscripta Math., 86 (1995), pp. 137–147, https://doi.org/10.1007/BF02567983.

[36] S.-I. Ohta, *Convexities of metric spaces*, Geom. Dedicata, 125 (2007), pp. 225–250, https://doi.org/10.1007/s10711-007-9159-3.

[37] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.

[38] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*, Springer-Verlag, Berlin, 1998.

[39] S. T. Smith, *Optimization techniques on Riemannian manifolds*, in Hamiltonian and Gradient Flows, Algorithms and Control, Fields Inst. Commun. 3, AMS, Providence, RI, 1994, pp. 113–136.

[40] S. Sra, S. Nowozin, and S. J. Wright, *Optimization for Machine Learning*, MIT Press, Cambridge, MA, 2011.

[41] J. Streets, *The consistency and convergence of K-energy minimizing movements*, Trans. Amer. Math. Soc., 368 (2016), pp. 5075–5091, https://doi.org/10.1090/tran/6508.

[42] C. Udrişte, *Convex Functions and Optimization Methods on Riemannian Manifolds*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1994.

[43] H. Zhang and S. Sra, *First-order methods for geodesically convex optimization*, in Proceedings of the 29th Annual Conference on Learning Theory (COLT), 2016, pp. 1617–1638.