Gradient flows on Wasserstein spaces over compact Alexandrov spaces*†

Shin-ichi OHTA‡
Department of Mathematics, Faculty of Science,
Kyoto University, Kyoto 606-8502, JAPAN
e-mail: sohta@math.kyoto-u.ac.jp

Abstract

We establish the existence of Euclidean tangent cones on Wasserstein spaces over compact Alexandrov spaces of curvature bounded below. By using this Riemannian structure, we formulate and construct gradient flows of functions on such spaces. If the underlying space is a Riemannian manifold of nonnegative sectional curvature, then our gradient flow of the free energy produces a solution of the linear Fokker-Planck equation.

1 Introduction

Our main object in the article is the (quadratic) Wasserstein space $(\mathcal{P}(X), d_2^W)$ (also called the Kantorovich-Rubinstein space) over a compact metric space (X, d). The Wasserstein space $(\mathcal{P}(X), d_2^W)$ is by definition the space of probability measures on X equipped with a certain distance structure d_2^W which metrizes the weak topology of $\mathcal{P}(X)$. Recently, it has turned out that there are strong connections between the structures of the Wasserstein space $\mathcal{P}(X)$ and the underlying space X, and that the geometry of $\mathcal{P}(X)$ provides a powerful tool for studying the structure of X. One of the most interesting examples is an approach to a synthetic lower Ricci curvature bound for general metric measure spaces (see [RS], [S2], [S3], [LV2] and [LV1], and also [CMS], [Oh2] and [Oh3] for related works). There the convexity of the entropy on $\mathcal{P}(X)$ plays the role of the lower Ricci curvature bound of X.

Our first main result (Theorem 3.6) concerns the infinitesimal structure of the Wasserstein space $\mathcal{P}(X)$ over a compact Alexandrov space X of curvature bounded below. We are inspired by the fact that $\mathcal{P}(X)$ is an Alexandrov space of nonnegative curvature if and only if so is X. This fact implies that $\mathcal{P}(X)$ has Euclidean tangent

*Mathematics Subject Classification (2000): 28A33, 53C23, 58J35.
†Keywords: Wasserstein space, Alexandrov space, gradient flow, Fokker-Planck equation, free energy.
‡Partly supported by the JSPS fellowship for research abroad.

cones and it gives a rigorous justification of Otto's formal Riemannian structure on $\mathcal{P}(\mathbb{R}^n)$ ([Ot]). Our theorem extends this fact, namely it asserts that $\mathcal{P}(X)$ has Euclidean tangent cones if X is a compact Alexandrov space with a possibly negative lower curvature bound. In this case, $\mathcal{P}(X)$ is not an Alexandrov space, but satisfies a kind of '2-uniform smoothness' (3.1) which is a concept coming from the geometry of Banach spaces (see [Oh4]). The 2-uniform smoothness can be regarded as a generalization of the nonnegative curvature in the sense of Alexandrov, and the error term is getting smaller as we are scaling up the space, thus we obtain that tangent cones exist and are Euclidean. This discussion proves the usefulness of the view of the geometry of Banach spaces in metric geometry.

Our Riemannian structure on $\mathcal{P}(X)$ enables us to consider gradient flows of lower semi-continuous and K-convex functions on $\mathcal{P}(X)$. Such a gradient flow exists and is complete (Theorem 5.11). Although the existence follows from results in the recent remarkable book [AGS] by Ambrosio, Gigli and Savaré, we will present a different construction. It extends a discussion in [Ly2] (see also [PP]) and seems to be of independent interest. If X is of nonnegative curvature, then we further obtain the uniqueness and the contraction property of gradient flows (Theorem 6.2).

In the particular case where the underlying space X is a Riemannian manifold of nonnegative sectional curvature, we show that our gradient flow of the free energy produces a solution of the linear Fokker-Planck equation (Theorem 6.6). This generalizes a celebrated work by Jordan, Kinderlehrer and Otto [JKO] on Euclidean spaces. As a corollary, the gradient flow of the relative entropy starting from a Dirac measure describes the heat kernel (Corollary 6.7). These results provide effective instruments in analysis, probability theory and geometry on Riemannian manifolds (see [OV] for a somewhat related formal discussion). We refer to the outstanding lecture notes [V2] by Villani for related results. After this paper was finished, I also learned of closely related work by Savaré [Sa].

The article is organized as follows: Section 2 contains reviews on Alexandrov spaces and Wasserstein spaces. We show the existence of Euclidean tangent cones on a Wasserstein space in Section 3. Section 4 concerns properties of lower semi-continuous and K-convex functions. Section 5 is devoted to gradient flows on a Wasserstein space. Finally, we discuss the nonnegatively curved case in Section 6.

Here are several conventions and notations throughout the article.

Conventions and notations. • All Riemannian manifolds are supposed to be smooth, connected, complete and boundaryless.

• We denote by $\lim_{\varepsilon \to 0+}$ the limit as ε tends to zero from the right.

• We denote by $\theta_\alpha(\varepsilon)$ a certain function which depends only on α and satisfies $\lim_{\varepsilon \to 0+} \theta_\alpha(\varepsilon) = 0$.

Acknowledgements. I would like to express my gratitude to Karl-Theodor Sturm for his suggestions and encouragement, and to Alexander Lytchak for fruitful discussions. This work was done while I was visiting Institut für Angewandte Mathematik, Universität Bonn. I am grateful to the institute for its hospitality. I also thank the referees for valuable remarks which lead to several crucial revisions.

2 Preliminaries

Let (X,d) be a metric space. A rectifiable curve $\gamma : [0,l] \longrightarrow X$ is called a *geodesic* if it is locally minimizing and has a constant speed, i.e., parametrized proportionally to the arclength. If a geodesic $\gamma : [0,l] \longrightarrow X$ satisfies $\mathrm{length}(\gamma) = d(\gamma(0),\gamma(l))$, then we say that it is *minimal.* A metric space (X,d) is said to be *geodesic* if any two points in X are connected by a minimal geodesic. For $x \in X$ and $r > 0$, we denote by $B(x,r)$ and $\overline{B}(x,r)$ open and closed balls with center x and radius r, respectively. We sometimes write d_X instead of d in order to emphasize which space is under consideration.

2.1 Alexandrov spaces

We first review Alexandrov spaces, in other words, metric spaces of sectional curvature bounded below. Standard references are [ABN], [BGP] and [BBI].

For $\kappa \in \mathbb{R}$, we denote by $\mathbb{M}^2(\kappa)$ a simply-connected, 2-dimensional Riemannian manifold of constant sectional curvature κ. Given three points $x,y,z \in X$ (provided that $d(x,y)+d(y,z)+d(z,x) < 2\pi/\sqrt{\kappa}$ if $\kappa > 0$), we can take corresponding points $\tilde{x},\tilde{y},\tilde{z} \in \mathbb{M}^2(\kappa)$ (which are unique up to an isometry) such that

$$d_{\mathbb{M}^2(\kappa)}(\tilde{x},\tilde{y}) = d_X(x,y), \quad d_{\mathbb{M}^2(\kappa)}(\tilde{y},\tilde{z}) = d_X(y,z), \quad d_{\mathbb{M}^2(\kappa)}(\tilde{z},\tilde{x}) = d_X(z,x).$$

We denote by $\gamma_{\tilde{x}\tilde{y}} : [0,1] \longrightarrow \mathbb{M}^2(\kappa)$ the unique minimal geodesic from $\tilde{x}$ to $\tilde{y}$.

Definition 2.1 (Alexandrov spaces) Let (X,d) be a geodesic metric space and $\kappa \in \mathbb{R}$. We say that (X,d) is an *Alexandrov space of curvature* $\geq \kappa$ if, for any three points $x,y,z \in X$ (provided that $d(x,y)+d(y,z)+d(z,x) < 2\pi/\sqrt{\kappa}$ if $\kappa > 0$), any minimal geodesic $\gamma : [0,1] \longrightarrow X$ from y to z and for any $\lambda \in [0,1]$, we have

$$d_X\big(x,\gamma(\lambda)\big) \geq d_{\mathbb{M}^2(\kappa)}\big(\tilde{x},\gamma_{\tilde{y}\tilde{z}}(\lambda)\big). \tag{2.1}$$

In the particular case $\kappa = 0$, the inequality (2.1) is rewritten as

$$d_X\big(x,\gamma(\lambda)\big)^2 \geq (1-\lambda)d_X(x,y)^2 + \lambda d_X(x,z)^2 - (1-\lambda)\lambda d_X(y,z)^2. \tag{2.2}$$

It is easy to see the following:

• If (X,d) is an Alexandrov space of curvature $\geq \kappa$, then, given three points $x,y,z \in X$ (provided that $d(x,y)+d(y,z)+d(z,x) < 2\pi/\sqrt{\kappa}$ if $\kappa > 0$) and two minimal geodesics $\gamma,\eta : [0,1] \longrightarrow X$ from x to y and from x to z, respectively, we have

$$d_X\big(\gamma(\lambda),\eta(\tau)\big) \geq d_{\mathbb{M}^2(\kappa)}\big(\gamma_{\tilde{x}\tilde{y}}(\lambda),\gamma_{\tilde{x}\tilde{z}}(\tau)\big) \tag{2.3}$$

for all $\lambda,\tau \in [0,1]$. (This property is also adopted as a definition of an Alexandrov space.)

• If (X,d) is an Alexandrov space of curvature $\geq \kappa$, then it is an Alexandrov space of curvature $\geq \kappa'$ for all $\kappa' \leq \kappa$.

• If (X,d) is an Alexandrov space of curvature $\geq \kappa$, then, given a positive constant $c > 0$, the scaled metric space $(X, c \cdot d)$ is an Alexandrov space of curvature $\geq \kappa/c^2$. Therefore every Alexandrov space can be regarded as an Alexandrov space of curvature ≥ -1 by scaling its distance by a positive constant if necessary.

Here are some fundamental examples of Alexandrov spaces.

Example 2.2 (i) A Riemannian manifold is an Alexandrov space of curvature $\geq \kappa$ if and only if its sectional curvature is greater than or equal to κ everywhere.

(ii) For a compact convex domain $\Omega \subset \mathbb{R}^n$, let $X = \partial\Omega$ be equipped with the length metric d induced from the standard metric of $\mathbb{R}^n$. Then (X, d) is an Alexandrov space of nonnegative curvature.

(iii) Let (M, g) be a Riemannian manifold of sectional curvature $\geq \kappa$ and G be a compact group acting on M by isometries. Then the quotient space M/G equipped with the quotient metric is an Alexandrov space of curvature $\geq \kappa$.

(iv) If a sequence of Alexandrov spaces of curvature $\geq \kappa$ is convergent with respect to the Gromov-Hausdorff distance, then its limit space is also an Alexandrov space of curvature $\geq \kappa$.

In the remainder of the subsection, we briefly discuss the infinitesimal structure of X. Fix a point $x \in X$. We define $\Sigma'_x X$ as the set of unit speed minimal geodesics $\gamma : [0, \delta] \longrightarrow X$ with $\gamma(0) = x$ equipped with the equivalence relation such that $\gamma \sim \eta$ holds if we have $\gamma(t) = \eta(t)$ for all $t \in [0, \varepsilon]$ for some $\varepsilon > 0$. For $\gamma, \eta \in \Sigma'_x X$, consider the function $h(s,t) := \angle\widetilde{\gamma(s)}\,\tilde{x}\,\widetilde{\eta(t)}$, where $\tilde{x}, \widetilde{\gamma(s)}, \widetilde{\eta(t)} \in \mathbb{M}^2(\kappa)$ and $\angle\widetilde{\gamma(s)}\tilde{x}\widetilde{\eta(t)}$ stands for the angle between $\gamma'_{\tilde{x}\widetilde{\gamma(s)}}(0)$ and $\gamma'_{\tilde{x}\widetilde{\eta(t)}}(0)$ at $\tilde{x}$. Then the curvature condition (2.3) guarantees that the function h is monotone non-increasing in both s and t, and hence we can define the *angle* between γ and η by

$$\angle_x(\gamma, \eta) := \lim_{s,t \to 0+} h(s,t) = \lim_{s,t \to 0+} \angle\widetilde{\gamma(s)}\,\tilde{x}\,\widetilde{\eta(t)}.$$

In particular, the limit $\lim_{\varepsilon \to 0+} h(s\varepsilon, t\varepsilon)$ always exists and is independent of the choices of $s, t > 0$. It means that X has infinitesimally a 'Hilbertian' structure (this is not the case of non-Hilbertian Banach spaces). Note also that it follows from (2.3) that

$$d_X\big(\gamma(s), \eta(t)\big) \leq d_{\mathbb{M}^2(\kappa)}\big(\bar{\gamma}(s), \bar{\eta}(t)\big), \tag{2.4}$$

where $\bar{\gamma}$ and $\bar{\eta}$ are unit speed geodesics in $\mathbb{M}^2(\kappa)$ with $\bar{\gamma}(0) = \bar{\eta}(0)$ and $\angle(\bar{\gamma}'(0), \bar{\eta}'(0)) = \angle_x(\gamma, \eta)$. The angle $\angle_x$ is independent of the choices of γ and η in their equivalence classes, and is a natural distance function on $\Sigma'_x X$. We define the *space of directions* $(\Sigma_x X, \angle_x)$ at x as the completion of $(\Sigma'_x X, \angle_x)$.

Define the *tangent cone* $(C_x X, \sigma_x)$ as the Euclidean cone over $(\Sigma_x X, \angle_x)$, namely

$$C_x X := \big(\Sigma_x X \times [0, \infty)\big) / \sim,$$

where $(\gamma, 0) \sim (\eta, 0)$, and

$$\sigma_x\big((\gamma, s), (\eta, t)\big) := \sqrt{s^2 + t^2 - 2st \cos \angle_x(\gamma, \eta)}$$

for $(\gamma, s), (\eta, t) \in C_x X$. We denote by o_x the origin $(*, 0) \in C_x X$. If (X, d) is a Riemannian manifold, then $(\Sigma_x X, \angle_x)$ and $(C_x X, \sigma_x)$ coincide with the unit tangent sphere and the tangent space at x, respectively.

The structures of spaces of directions and tangent cones are well understood for finite dimensional Alexandrov spaces.

Proposition 2.3 (cf. [BBI, Theorem 10.9.3, Corollaries 10.9.5, 10.9.6]) *Let (X, d) be a complete, finite Hausdorff dimensional Alexandrov space of curvature $\geq \kappa$. Then, at every $x \in X$, the scaled pointed metric space $(X, c \cdot d, x)$ converges to $(C_x X, \sigma_x, o_x)$ as c diverges to infinity in the sense of the pointed Gromov-Hausdorff convergence. In particular, the space of directions $(\Sigma_x X, \angle_x)$ is an Alexandrov space of curvature ≥ 1 (provided that $\dim X \geq 2$), and the tangent cone $(C_x X, \sigma_x)$ is an Alexandrov space of curvature ≥ 0.*

However, in the infinite dimensional case, their structures can be more complicated. In fact, Halbeisen [H] constructed an instance of an infinite dimensional, complete Alexandrov space of nonnegative curvature containing a point at which the tangent cone is not an inner metric space. Here we say that a metric space is *inner* if the distance between arbitrary two points coincides with the infimum of the lengths of curves connecting them. Geodesic metric spaces are clearly inner.

2.2 Wasserstein spaces

In this subsection, we recall Wasserstein spaces. It is the set of probability measures on a metric space equipped with a reasonable distance structure derived from the distance structure of the underlying metric space. This concept has many connections with and applications in various fields of mathematics for which we refer to [V1], [V2] and the references therein. We will restrict ourselves to compact metric spaces and it allows us to ignore some delicate points arising in the noncompact case. See [V1] for a general theory.

Let (X, d) be a compact metric space. Denote by $\mathcal{P}(X)$ the set of all Borel probability measures on X. Given two probability measures $\mu, \nu \in \mathcal{P}(X)$, a probability measure $q \in \mathcal{P}(X \times X)$ is called a *coupling* of μ and ν if it satisfies

$$q(A \times X) = \mu(A), \qquad q(X \times A) = \nu(A)$$

for every Borel set $A \subset X$ (i.e., the marginals of q are μ and ν). For instance, the product measure $\mu \times \nu$ is a coupling of μ and ν.

Definition 2.4 (Wasserstein spaces) The (*quadratic*) *Wasserstein space* over (X, d) is a metric space $(\mathcal{P}(X), d_2^W)$ equipped with the distance structure d_2^W defined by

$$d_2^W(\mu, \nu) := \inf_q \left\{ \int_{X \times X} d_X(x, y)^2 \, dq(x, y) \right\}^{1/2}$$

for $\mu, \nu \in \mathcal{P}(X)$. Here the infimum is taken over all couplings $q \in \mathcal{P}(X \times X)$ of μ and ν.

We remark that $d_2^W(\mu, \nu)$ is finite since X is bounded. A coupling $q \in \mathcal{P}(X \times X)$ of μ and ν is said to be *optimal* if it realizes the distance $d_2^W(\mu, \nu)$. The underlying metric space X is isometrically embedded into $\mathcal{P}(X)$ by identifying a point $x \in X$ with the Dirac measure $\delta_x \in \mathcal{P}(X)$ at x. The Wasserstein space $(\mathcal{P}(X), d_2^W)$ is a compact metric space since X is compact. The Wasserstein distance d_2^W metrizes the weak topology of $\mathcal{P}(X)$: A sequence $\{\mu_i\}_{i\in\mathbb{N}} \subset \mathcal{P}(X)$ converges to $\mu \in \mathcal{P}(X)$ with respect to d_2^W if and only if $\{\mu_i\}_{i\in\mathbb{N}}$ converges to μ weakly.

If (X,d) is geodesic, then so is $(\mathcal{P}(X), d_2^W)$ and geodesics in $\mathcal{P}(X)$ can be written by using probability measures on the family of geodesics in X. Let $\Gamma(X)$ be the set of minimal geodesics, say $\gamma : [0,1] \longrightarrow X$, in X and define the evaluation map $e_\lambda : \Gamma(X) \longrightarrow X$ by $e_\lambda(\gamma) := \gamma(\lambda)$ for each $\lambda \in [0,1]$. We regard $\Gamma(X)$ as a subset of the set of Lipschitz maps $\mathrm{Lip}([0,1], X)$ equipped with the uniform topology. Note that e_λ is continuous with respect to the uniform topology.

Proposition 2.5 ([LV2, Proposition 2.10]) *Let (X,d) be a compact, geodesic metric space. Then, for any $\mu, \nu \in \mathcal{P}(X)$ and any minimal geodesic $\alpha : [0,1] \longrightarrow \mathcal{P}(X)$ between them, there exists a Borel probability measure $\Pi \in \mathcal{P}(\Gamma(X))$ such that $(e_0 \times e_1)_*\Pi$ is an optimal coupling of μ and ν and that we have $(e_\lambda)_*\Pi = \alpha(\lambda)$ for all $\lambda \in [0,1]$.*

In the Riemannian case, McCann's significant work provides a more precise description. For a Riemannian manifold (M,g), let us denote by $\mathcal{P}^{ac}(M) \subset \mathcal{P}(M)$ the subset consisting of measures which are absolutely continuous with respect to the Riemannian volume element m.

Theorem 2.6 ([M, Theorems 8, 9]) *Let (M,g) be a compact Riemannian manifold. Then, for any $\mu \in \mathcal{P}^{ac}(M)$ and $\nu \in \mathcal{P}(M)$, there exists a function $\psi : M \longrightarrow \mathbb{R}$ satisfying the following:*

(i) *There exists a function $\phi : M \longrightarrow \mathbb{R}$ such that*

$$\psi(x) = \inf_{y\in M}\{d_M(x,y)^2/2 - \phi(y)\} \tag{2.5}$$

holds for all $x \in M$.

(ii) *The map $\Psi(x) := \exp_x[-\operatorname{grad}\psi(x)]$, $x \in M$, satisfies $\Psi_*\mu = \nu$.*

(iii) *The map Ψ is an optimal transportation from μ to ν, that is, we have*

$$\left\{\int_M d_M\big(x, \Psi(x)\big)^2\, d\mu(x)\right\}^{1/2} = d_2^W(\mu,\nu). \tag{2.6}$$

Moreover, Ψ is the unique map (up to a change on a μ-null measure set) satisfying $\Psi_*\mu = \nu$ and (2.6).

The condition (2.5) is called the *c-concavity* with respect to the cost function $c(x,y) = d(x,y)^2/2$, and it is indeed equivalent to the concavity of the function $\psi(x) - |x|^2/2$ in Euclidean spaces. Moreover, the condition (2.5) implies that ψ is a Lipschitz function. Hence ψ is differentiable and the gradient vector $\operatorname{grad}\psi$ makes sense almost everywhere. In view of Proposition 2.5, if we define a map $F : M \longrightarrow \Gamma(M)$ by $[F(x)](t) := \exp_x[-t\operatorname{grad}\psi(x)]$, then the measure $\Pi := F_*\mu \in \mathcal{P}(\Gamma(M))$ produces a (unique) minimal geodesic from μ to ν. See [V2] for more comprehensive discussion.

2.3 Entropy and Ricci curvature

Let (M, g) be a compact Riemannian manifold and m be the associated volume element. For $\mu \in \mathcal{P}^{ac}(M)$, the *relative entropy* of μ with respect to m is defined by

$$\mathrm{Ent}_m(\mu) := \int_M \rho \log \rho \, dm \ \in (-\infty, \infty], \tag{2.7}$$

where the function ρ stands for the density of μ, i.e., $\mu = \rho \cdot m$. We also define $\mathrm{Ent}_m(\mu) := \infty$ for $\mathcal{P}(M) \setminus \mathcal{P}^{ac}(M)$. The following are well-known facts (cf. [S2, Lemma 4.1]).

Lemma 2.7 (i) *The relative entropy* Ent_m *satisfies* $\mathrm{Ent}_m(\mu) \ge -\log m(M)$ *for all* $\mu \in \mathcal{P}(M)$ *and equality holds if and only if* $\mu = m(M)^{-1} \cdot m$.

(ii) *The relative entropy* Ent_m *is lower semi-continuous on* $\mathcal{P}(M)$.

(iii) *The set* $\mathcal{P}^*(M) := \{\mu \in \mathcal{P}(M) \mid \mathrm{Ent}_m(\mu) < \infty\}$ *is dense in* $\mathcal{P}(M)$.

It is known that there is a strong connection between the behavior of the relative entropy and the Ricci curvature.

Theorem 2.8 ([RS, Theorem 1.1]) *A compact Riemannian manifold* (M, g) *satisfies* $\mathrm{Ric}_M \ge K$ *for* $K \in \mathbb{R}$ *if and only if the relative entropy* Ent_m *is* K*-convex on* $\mathcal{P}(M)$.

The K-convexity of Ent_m means that it is K-convex on every minimal geodesic. See Section 4 for the precise definition. This theorem allows us to adopt the K-convexity of the relative entropy as a synthetic lower Ricci curvature bound for general metric measure spaces. See [S2], [S3], [LV2], [LV1], [Oh2] and [Oh3] for recent progress around this fascinating topic.

More generally, given a smooth function $V \in C^\infty(M)$, we define the associated *free energy* of $\mu \in \mathcal{P}(M)$ by

$$f(\mu) := \mathrm{Ent}_m(\mu) + \int_M V \, d\mu. \tag{2.8}$$

Note that, if $\mu = \rho \cdot m \in \mathcal{P}^{ac}(M)$, then

$$\begin{aligned} f(\mu) &= \int_M \rho \log \rho \, dm + \int_M V\rho \, dm = \int_M \rho \log(\rho \cdot e^V) \, dm \\ &= \int_M (\rho \cdot e^V) \log(\rho \cdot e^V) e^{-V} \, dm. \end{aligned}$$

Hence $f(\mu)$ can also be regarded as the relative entropy of μ with respect to the measure $e^{-V} \cdot m$. The free energy f is clearly lower semi-continuous and the following generalization of Theorem 2.8 holds.

Theorem 2.9 ([S1, Theorem 1.3], [LV2, Theorem 7.3]) *Let* (M, g) *be a compact Riemannian manifold and* $V \in C^\infty(M)$*. Then we have* $\mathrm{Ric}_M + \mathrm{Hess}\, V \ge K$ *for* $K \in \mathbb{R}$ *if and only if the free energy* (2.8) *is* K*-convex on* $\mathcal{P}(M)$.

Here $\mathrm{Hess}\, V$ stands for the Hessian of V, and the inequality $\mathrm{Ric}_M + \mathrm{Hess}\, V \ge K$ is read as $\mathrm{Ric}_M(v,v) + \mathrm{Hess}\, V(v,v) \ge K|v|^2$ for all $v \in TM$. The compactness of M ensures the existence of such a constant $K \in \mathbb{R}$.

3 The structure of Wasserstein spaces

In his remarkable paper [Ot], Otto introduced a formal Riemannian structure on Wasserstein spaces over Euclidean spaces. This structure can be rigorously introduced using the fact that the Wasserstein space is an Alexandrov space of nonnegative curvature if and only if so is the underlying metric space (see [S2, Proposition 2.10] and [LV2, Proposition A.9]). However, it is also known that the Wasserstein space is not an Alexandrov space (even for a negative κ) if the underlying metric space does not have the nonnegative curvature (see [S2, Proposition 2.10]). In order to overcome this difficulty, we introduce a 'non-Hilbertian' extension of the nonnegative curvature in the sense of Alexandrov. This extension has a flavor of the geometry of Banach spaces.

3.1 A generalized 2-uniform smoothness

Let (X,d) be a geodesic metric space and consider the following inequality: Given three points $x,y,z \in X$, a minimal geodesic $\gamma : [0,1] \longrightarrow X$ from y to z and $\lambda \in [0,1]$,

$$d\big(x,\gamma(\lambda)\big)^2 \ge (1-\lambda)d(x,y)^2 + \lambda d(x,z)^2 - S^2(1-\lambda)\lambda d(y,z)^2, \tag{3.1}$$

where $S \ge 1$ is a fixed constant. We say that a geodesic metric space (X,d) *satisfies* (3.1) if there is a constant $S \ge 1$ such that the inequality (3.1) holds for all $x,y,z \in X$, $\gamma : [0,1] \longrightarrow X$ and $\lambda \in [0,1]$. This inequality generalizes (2.2) which amounts to the case of $S = 1$, and it can also be regarded as a nonlinear analogue of the 2-*uniform smoothness* in Banach space theory. For instance, L_p-spaces with $p \in [2,\infty)$ satisfy (3.1) with $S = \sqrt{p-1}$. We refer to [Oh1] and [Oh4] for works in this direction.

Proposition 3.1 *A compact geodesic metric space* (X,d) *satisfies* (3.1) *with a constant* S *if and only if the Wasserstein space* $(\mathcal{P}(X), d_2^W)$ *satisfies* (3.1) *with the same constant* S.

Proof. The 'if' part is obvious because X is isometrically embedded into $\mathcal{P}(X)$. We assume that (X,d) satisfies (3.1) with some constant $S \ge 1$. Fix three probability measures $\mu_0, \mu_1, \nu \in \mathcal{P}(X)$ and a minimal geodesic $\alpha : [0,1] \longrightarrow \mathcal{P}(X)$ with $\alpha(0) = \mu_0$ and $\alpha(1) = \mu_1$. By Proposition 2.5, there exists a probability measure $\Pi \in \mathcal{P}(\Gamma(X))$ with $\mu_\tau := (e_\tau)_*\Pi = \alpha(\tau)$ for $\tau \in [0,1]$ as well as $(e_0 \times e_1)_*\Pi$ is an optimal coupling.

Given $\lambda \in [0,1]$, we fix an optimal coupling $q \in \mathcal{P}(X \times X)$ of ν and μ_λ. Now we consider disintegrations of Π and q by using μ_λ, that is,

$$d\Pi = d\Pi_\lambda^w d\mu_\lambda(w), \qquad dq = dq^w d\mu_\lambda(w),$$

where $\Pi_\lambda^w \in \mathcal{P}(\Gamma(X))$ is concentrated on $\{\gamma \in \Gamma(X) \,|\, \gamma(\lambda) = w\}$ and $q^w \in \mathcal{P}(X)$ for μ_λ-a.e. $w \in X$. For such a point $w \in X$, a curve $\gamma \in \operatorname{supp} \Pi_\lambda^w$ and for a point $x \in \operatorname{supp} q^w$, it follows from (3.1) on X that

$$d(x,w)^2 \ge (1-\lambda)d\big(x,\gamma(0)\big)^2 + \lambda d\big(x,\gamma(1)\big)^2 - S^2(1-\lambda)\lambda d\big(\gamma(0),\gamma(1)\big)^2. \tag{3.2}$$

For $a = 0, 1$, define the (not necessarily optimal) coupling $q_a \in \mathcal{P}(X \times X)$ of ν and μ_a by

$$q_a := \int_X \big(q^w \times [(e_a)_* \Pi_\lambda^w]\big) d\mu_\lambda(w).$$

Then we obtain, by integrating (3.2) on $(X \times \Gamma(X)) \times X$ with respect to the measure $(dq^w(x) d\Pi_\lambda^w(\gamma)) d\mu_\lambda(w)$,

$$\begin{aligned} d_2^W(\nu,\mu_\lambda)^2 &= \int_{X\times X} d(x,w)^2 dq(x,w) \\ &\ge (1-\lambda)\int_{X\times X} d(x,y)^2 dq_0(x,y) + \lambda \int_{X\times X} d(x,z)^2 dq_1(x,z) \\ &\quad - S^2(1-\lambda)\lambda \int_{\Gamma(X)} d\big(\gamma(0),\gamma(1)\big)^2 d\Pi(\gamma) \\ &\ge (1-\lambda) d_2^W(\nu,\mu_0)^2 + \lambda d_2^W(\nu,\mu_1)^2 - S^2(1-\lambda)\lambda d_2^W(\mu_0,\mu_1)^2. \end{aligned}$$

Therefore $(\mathcal{P}(X), d_2^W)$ satisfies (3.1) with the constant S. □

Remark 3.2 The analogue of Proposition 3.1 for the reverse inequality of (3.1) (in other words, the 2-uniform convexity),

$$d\big(x,\gamma(\lambda)\big)^2 \le (1-\lambda)d(x,y)^2 + \lambda d(x,z)^2 - C^{-2}(1-\lambda)\lambda d(y,z)^2, \tag{3.3}$$

does not hold true. In fact, it is known that $\mathcal{P}(\mathbb{R}^2)$ does not satisfy (3.3) with $C = 1$ while $\mathbb{R}^2$ does (see [AGS, Example 7.3.3]).

3.2 The 2-uniform smoothness of Alexandrov spaces

We shall prove that a general Alexandrov space satisfies the 2-uniform smoothness (3.1) locally. By scaling the distance if necessary, without loss of generality, we can assume that the lower curvature bound is -1.

Lemma 3.3 *Let (X,d) be an Alexandrov space of curvature ≥ -1. Then, for any three distinct points $x, y, z \in X$, minimal geodesic $\gamma : [0,1] \longrightarrow X$ from y to z and for any $\lambda \in [0,1]$, we have*

$$\begin{aligned} d\big(x,\gamma(\lambda)\big)^2 &\ge (1-\lambda)d(x,y)^2 + \lambda d(x,z)^2 \\ &\qquad - \Big\{1 + \sup_{\tau\in[0,1]} d\big(x,\gamma(\tau)\big)^2\Big\} \cdot (1-\lambda)\lambda d(y,z)^2. \end{aligned} \tag{3.4}$$

In particular, if (X,d) is bounded, then it satisfies (3.1) *with* $S = \{1 + (\operatorname{diam} X)^2\}^{1/2}$.

Proof. Note that it is sufficient to prove (3.4) for infinitesimally thin triangles, for it implies the required concavity of the function $d(x,\gamma(\cdot))^2$. That is to say, for all $\lambda \in (0,1)$, we will show

$$\frac{4}{d(\gamma(\lambda-\varepsilon),\gamma(\lambda+\varepsilon))^2}\left\{\frac{1}{2}d\big(x,\gamma(\lambda-\varepsilon)\big)^2+\frac{1}{2}d\big(x,\gamma(\lambda+\varepsilon)\big)^2-d\big(x,\gamma(\lambda)\big)^2\right\}$$
$$\le 1+\sup_{\tau\in[0,1]} d\big(x,\gamma(\tau)\big)^2+\theta_{x,\gamma}(\varepsilon). \tag{3.5}$$

We remark that $\theta_{x,\gamma}(\varepsilon)$ is independent of the choice of $\lambda \in (0,1)$.

We assume $x \notin \gamma$ because, if $x \in \gamma$, then the left hand side of (3.5) is equal to 1. Given $\lambda \in (0,1)$ and small $\varepsilon > 0$, put $w := \gamma(\lambda)$, $y_\varepsilon := \gamma(\lambda-\varepsilon)$ and $z_\varepsilon := \gamma(\lambda+\varepsilon)$. We also define the minimal geodesics $\gamma_-,\gamma_+ : [0,1] \longrightarrow X$ from w to y_ε and from w to z_ε by $\gamma_-(t) := \gamma(\lambda - t\varepsilon)$ and $\gamma_+(t) := \gamma(\lambda + t\varepsilon)$, respectively, and fix a minimal geodesic $\eta : [0,1] \longrightarrow X$ from w to x. Since the function $\cosh t$ is convex, we observe

$$\begin{aligned}\cosh d(x,y_\varepsilon)-\cosh d(x,w) &\ge \{d(x,y_\varepsilon)-d(x,w)\}\sinh d(x,w)\\ &= \{d(x,y_\varepsilon)^2-d(x,w)^2\}\frac{\sinh d(x,w)}{d(x,y_\varepsilon)+d(x,w)}\\ &= \frac{1}{2}\{d(x,y_\varepsilon)^2-d(x,w)^2\}\big(1+\theta_{x,\gamma}(\varepsilon)\big)\frac{\sinh d(x,w)}{d(x,w)}.\end{aligned}$$

It follows from the inequality (2.4) for a triangle $\triangle wxy_\varepsilon$ together with the hyperbolic cosine formula that

$$\begin{aligned}\cosh d(x,y_\varepsilon) \le{}& \cosh d(x,w)\cosh d(y_\varepsilon,w)\\ &- \sinh d(x,w)\sinh d(y_\varepsilon,w)\cos\angle_w(\eta,\gamma_-).\end{aligned}$$

These together yield that

$$\begin{aligned}&\frac{1+\theta_{x,\gamma}(\varepsilon)}{2}\frac{\sinh d(x,w)}{d(x,w)}\{d(x,y_\varepsilon)^2-d(x,w)^2\}\\ &\le \cosh d(x,y_\varepsilon)-\cosh d(x,w)\\ &\le \cosh d(x,w)\cosh d(y_\varepsilon,w)-\sinh d(x,w)\sinh d(y_\varepsilon,w)\cos\angle_w(\eta,\gamma_-)\\ &\qquad -\cosh d(x,w)\\ &= \cosh d(x,w)\left\{\cosh\left(\frac{d(y_\varepsilon,z_\varepsilon)}{2}\right)-1\right\}\\ &\qquad -\sinh d(x,w)\sinh\left(\frac{d(y_\varepsilon,z_\varepsilon)}{2}\right)\cos\angle_w(\eta,\gamma_-).\end{aligned}$$

Similarly, it holds that

$$\begin{aligned}&\frac{1+\theta_{x,\gamma}(\varepsilon)}{2}\frac{\sinh d(x,w)}{d(x,w)}\{d(x,z_\varepsilon)^2-d(x,w)^2\}\\ &\le \cosh d(x,w)\left\{\cosh\left(\frac{d(y_\varepsilon,z_\varepsilon)}{2}\right)-1\right\}\\ &\qquad -\sinh d(x,w)\sinh\left(\frac{d(y_\varepsilon,z_\varepsilon)}{2}\right)\cos\angle_w(\eta,\gamma_+).\end{aligned}$$

Note that $\angle_w(\eta, \gamma_-) + \angle_w(\eta, \gamma_+) = \pi$ by the definitions of γ_+ and γ_-, and it implies $\cos\angle_w(\eta, \gamma_-) + \cos\angle_w(\eta, \gamma_+) = 0$. Thus we have

$$\begin{aligned}&\big(1+\theta_{x,\gamma}(\varepsilon)\big)\left\{\frac{1}{2}d(x,y_\varepsilon)^2+\frac{1}{2}d(x,z_\varepsilon)^2-d(x,w)^2\right\}\\&\le \frac{2d(x,w)\cosh d(x,w)}{\sinh d(x,w)}\left\{\cosh\left(\frac{d(y_\varepsilon,z_\varepsilon)}{2}\right)-1\right\},\end{aligned}$$

and hence

$$\begin{aligned}&\frac{4}{d(y_\varepsilon,z_\varepsilon)^2}\left\{\frac{1}{2}d(x,y_\varepsilon)^2+\frac{1}{2}d(x,z_\varepsilon)^2-d(x,w)^2\right\}\\&\le \big(1+\theta_{x,\gamma}(\varepsilon)\big)\frac{2d(x,w)\cosh d(x,w)}{\sinh d(x,w)}\frac{\cosh(d(y_\varepsilon,z_\varepsilon)/2)-1}{\{d(y_\varepsilon,z_\varepsilon)/2\}^2}\\&= \big(1+\theta_{x,\gamma}(\varepsilon)\big)\frac{d(x,w)\cosh d(x,w)}{\sinh d(x,w)}\\&\le \frac{d(x,w)}{\sinh d(x,w)}\{1+d(x,w)\sinh d(x,w)\}+\theta_{x,\gamma}(\varepsilon)\\&\le 1+d(x,w)^2+\theta_{x,\gamma}(\varepsilon)=1+d\big(x,\gamma(\lambda)\big)^2+\theta_{x,\gamma}(\varepsilon).\end{aligned}$$

Therefore we obtain (3.5) and complete the proof. $\square$

It is important for later use to estimate the error term in (3.4) (relative to (2.2)) using only $d(y,z)$.

Lemma 3.4 *Let (X,d) be an Alexandrov space of curvature ≥ -1. Given $x,y,z\in X$, a minimal geodesic $\gamma:[0,1]\longrightarrow X$ from y to z and $\lambda\in[0,1]$, if $d(y,z)\le 1$, then we have*

$$\begin{aligned}d\big(x,\gamma(\lambda)\big)^2\ge&\{1-d(y,z)^{1/2}\}^2\cdot\{(1-\lambda)d(x,y)^2+\lambda d(x,z)^2\}\\&-\{1+4d(y,z)\}\cdot(1-\lambda)\lambda d(y,z)^2.\end{aligned}\tag{3.6}$$

Proof. If $2d(y,z)^{1/2}\ge\sup_{\tau\in[0,1]}d(x,\gamma(\tau))$, then (3.6) immediately follows from (3.4).

In case of $2d(y,z)^{1/2}\le\sup_{\tau\in[0,1]}d(x,\gamma(\tau))$, we observe

$$d(x,y)\ge\sup_{\tau\in[0,1]}d\big(x,\gamma(\tau)\big)-d(y,z)\ge 2d(y,z)^{1/2}-d(y,z)^{1/2}=d(y,z)^{1/2}.$$

We used the assumption $d(y,z)\le 1$ in the second inequality. Thus we find, since $d(x,y)\ge d(y,z)^{1/2}\ge d(y,z)$,

$$\begin{aligned}d\big(x,\gamma(\lambda)\big)^2&\ge\{d(x,y)-d(y,z)\}^2\ge\{d(x,y)-d(y,z)^{1/2}\cdot d(x,y)\}^2\\&=\{1-d(y,z)^{1/2}\}^2d(x,y)^2.\end{aligned}$$

A similar discussion yields $d(x,\gamma(\lambda))^2\ge\{1-d(y,z)^{1/2}\}^2d(x,z)^2$. Therefore we obtain

$$d\big(x,\gamma(\lambda)\big)^2\ge\{1-d(y,z)^{1/2}\}^2\cdot\{(1-\lambda)d(x,y)^2+\lambda d(x,z)^2\}.$$

$\square$

3.3 Tangent cones on Wasserstein spaces over Alexandrov spaces

By integrating the inequality (3.6), we obtain a similar inequality in $\mathcal{P}(X)$. However, we need to take care of the relation between a geodesic α in $\mathcal{P}(X)$ and a family of geodesics in X, say $\Pi \in \mathcal{P}(\Gamma(X))$, which produces α (in the sense of Proposition 2.5). In fact, some $\gamma \in \operatorname{supp}\Pi$ may be long even when α is short.

Lemma 3.5 *Let (X, d) be a compact Alexandrov space of curvature ≥ -1 and set $D := \operatorname{diam} X$. Then, for any minimal geodesic $\alpha : [0, 1] \longrightarrow \mathcal{P}(X)$, $0 \leq s \leq t \leq 1$ with $t - s \leq D^{-1}$, $\nu \in \mathcal{P}(X)$ and for any $\lambda \in [0, 1]$, we have*

$$\begin{aligned} &d_2^W\big(\nu, \alpha((1-\lambda)s + \lambda t)\big)^2 \\ &\geq \{1 - (t-s)^{1/2}D^{1/2}\}^2 \cdot \big\{(1-\lambda)d_2^W\big(\nu, \alpha(s)\big)^2 + \lambda d_2^W\big(\nu, \alpha(t)\big)^2\big\} \\ &\quad - \{1 + 4(t-s)D\} \cdot (1-\lambda)\lambda d_2^W\big(\alpha(s), \alpha(t)\big)^2. \end{aligned}$$

Proof. The proof is similar to the proof of Proposition 3.1, so that we give only an outline. By Proposition 2.5, we find $\Pi \in \mathcal{P}(\Gamma(X))$ such that $(e_\tau)_*\Pi = \alpha(\tau)$ for $\tau \in [0, 1]$ and that $(e_0 \times e_1)_*\Pi$ is an optimal coupling. We remark that every geodesic $\gamma \in \operatorname{supp}\Pi$ has a length at most D. Fix $\lambda \in [0, 1]$ and an optimal coupling $q \in \mathcal{P}(X \times X)$ of ν and $\mu_\lambda := \alpha((1-\lambda)s + \lambda t)$. We consider disintegrations of Π and q by using μ_λ, i.e., $d\Pi = d\Pi_\lambda^w d\mu_\lambda(w)$ and $dq = dq^w d\mu_\lambda(w)$. Then, for μ_λ-a.e. $w \in X$, $\gamma \in \operatorname{supp}\Pi_\lambda^w$ and for $x \in \operatorname{supp} q^w$, it follows from Lemma 3.4 that

$$\begin{aligned} d(x, w)^2 \geq &\{1 - (t-s)^{1/2}D^{1/2}\}^2 \cdot \big\{(1-\lambda)d\big(x, \gamma(s)\big)^2 + \lambda d\big(x, \gamma(t)\big)^2\big\} \\ &- \{1 + 4(t-s)D\} \cdot (1-\lambda)\lambda d\big(\gamma(s), \gamma(t)\big)^2. \end{aligned}$$

Note that $d(\gamma(s), \gamma(t)) \leq (t-s)D \leq 1$ by assumption. By integrating this inequality with respect to $(dq^w(x)d\Pi_\lambda^w(\gamma))d\mu_\lambda(w)$, we obtain the required inequality. $\square$

Now we are ready to prove one of our main results.

Theorem 3.6 (Tangent cones on $\mathcal{P}(X)$) *Let (X, d) be a compact Alexandrov space of curvature ≥ -1. Then, at any $\mu \in \mathcal{P}(X)$, the following hold:*

(i) *Given two unit speed minimal geodesics $\alpha, \beta : [0, \delta] \longrightarrow \mathcal{P}(X)$ with $\alpha(0) = \beta(0) = \mu$ and $\delta > 0$, the limit*

$$\sigma_\mu\big((\alpha, s), (\beta, t)\big) := \lim_{\varepsilon \to 0+} \frac{1}{\varepsilon} d_2^W\big(\alpha(s\varepsilon), \beta(t\varepsilon)\big)$$

exists for all $s, t \geq 0$.

(ii) *For α and β as in* (i)*, the quantity*

$$\frac{1}{2st}\big\{s^2 + t^2 - \sigma_\mu\big((\alpha, s), (\beta, t)\big)^2\big\}$$

is independent of the choices of $s, t > 0$.

Proof. Throughout the proof, we set $D := \operatorname{diam} X$. We assume $0 < s, t \le 1$ without loss of generality.

(i) We shall show that the function $\varepsilon \longmapsto d_2^W(\alpha(s\varepsilon), \beta(t\varepsilon))/\varepsilon$ is 'almost' non-increasing. Take $\varepsilon \in (0, \min\{\delta, D^{-1}\})$ and $\lambda \in (0,1)$. Then it follows from Lemma 3.5 with $\nu = \beta(t\varepsilon)$ and $(s,t) = (0, s\varepsilon/\delta)$ that

$$\begin{aligned}
&d_2^W\big(\beta(t\varepsilon), \alpha(\lambda s\varepsilon)\big)^2 \\
&\ge \{1 - (s\varepsilon D/\delta)^{1/2}\}^2 \cdot \big\{(1-\lambda) d_2^W\big(\beta(t\varepsilon), \alpha(0)\big)^2 + \lambda d_2^W\big(\beta(t\varepsilon), \alpha(s\varepsilon)\big)^2\big\} \\
&\quad - (1 + 4s\varepsilon D/\delta) \cdot (1-\lambda)\lambda d_2^W\big(\alpha(0), \alpha(s\varepsilon)\big)^2 \\
&= (1-\lambda)(t\varepsilon)^2 + \lambda d_2^W\big(\beta(t\varepsilon), \alpha(s\varepsilon)\big)^2 - (1-\lambda)\lambda(s\varepsilon)^2 + \varepsilon^2 \theta_{D/\delta}(\varepsilon).
\end{aligned} \tag{3.7}$$

Similarly, we have

$$\begin{aligned}
&d_2^W\big(\alpha(\lambda s\varepsilon), \beta(\lambda t\varepsilon)\big)^2 \\
&\ge (1-\lambda) d_2^W\big(\alpha(\lambda s\varepsilon), \beta(0)\big)^2 + \lambda d_2^W\big(\alpha(\lambda s\varepsilon), \beta(t\varepsilon)\big)^2 \\
&\quad - (1-\lambda)\lambda d_2^W\big(\beta(0), \beta(t\varepsilon)\big)^2 + \varepsilon^2 \theta_{D/\delta}(\varepsilon) \\
&= (1-\lambda)(\lambda s\varepsilon)^2 + \lambda d_2^W\big(\alpha(\lambda s\varepsilon), \beta(t\varepsilon)\big)^2 - (1-\lambda)\lambda(t\varepsilon)^2 + \varepsilon^2 \theta_{D/\delta}(\varepsilon).
\end{aligned} \tag{3.8}$$

Combining (3.7) and (3.8), we obtain

$$\begin{aligned}
&d_2^W\big(\alpha(\lambda s\varepsilon), \beta(\lambda t\varepsilon)\big)^2 \\
&\ge (1-\lambda)\lambda^2(s\varepsilon)^2 - (1-\lambda)\lambda(t\varepsilon)^2 \\
&\quad + \lambda\big\{(1-\lambda)(t\varepsilon)^2 + \lambda d_2^W\big(\alpha(s\varepsilon), \beta(t\varepsilon)\big)^2 - (1-\lambda)\lambda(s\varepsilon)^2\big\} + \varepsilon^2 \theta_{D/\delta}(\varepsilon) \\
&= \lambda^2 d_2^W\big(\alpha(s\varepsilon), \beta(t\varepsilon)\big)^2 + \varepsilon^2 \theta_{D/\delta}(\varepsilon),
\end{aligned}$$

and hence

$$\frac{1}{\lambda\varepsilon} d_2^W\big(\alpha(\lambda s\varepsilon), \beta(\lambda t\varepsilon)\big) \ge \frac{1}{\varepsilon} d_2^W\big(\alpha(s\varepsilon), \beta(t\varepsilon)\big) + \theta_{D/\delta}(\varepsilon).$$

This implies that the limit $\sigma_\mu((\alpha, s), (\beta, t)) := \lim_{\varepsilon \to 0+} d_2^W(\alpha(s\varepsilon), \beta(t\varepsilon))/\varepsilon$ exists.

(ii) We take $\varepsilon \in (0, \min\{\delta, D^{-1}\})$ and $\lambda \in (0,1)$, and compare $\sigma_\mu((\alpha, s), (\beta, t))$ and $\sigma_\mu((\alpha, \lambda s), (\beta, t))$. On one hand, we observed in (3.7) that

$$\begin{aligned}
&d_2^W\big(\alpha(\lambda s\varepsilon), \beta(t\varepsilon)\big)^2 \\
&\ge (1-\lambda)(t\varepsilon)^2 + \lambda d_2^W\big(\alpha(s\varepsilon), \beta(t\varepsilon)\big)^2 - (1-\lambda)\lambda(s\varepsilon)^2 + \varepsilon^2 \theta_{D/\delta}(\varepsilon).
\end{aligned}$$

By dividing both sides by ε^2 and letting ε tend to zero, this implies

$$\sigma_\mu\big((\alpha, \lambda s), (\beta, t)\big)^2 \ge (1-\lambda)t^2 + \lambda \sigma_\mu\big((\alpha, s), (\beta, t)\big)^2 - (1-\lambda)\lambda s^2. \tag{3.9}$$

On the other hand, it follows from (3.8) that

$$\lambda^2 \sigma_\mu\big((\alpha, s), (\beta, t)\big)^2 \ge (1-\lambda)(\lambda s)^2 + \lambda \sigma_\mu\big((\alpha, \lambda s), (\beta, t)\big)^2 - (1-\lambda)\lambda t^2,$$

and hence

$$\sigma_\mu\big((\alpha,\lambda s),(\beta,t)\big)^2 \le (1-\lambda)t^2+\lambda\sigma_\mu\big((\alpha,s),(\beta,t)\big)^2-(1-\lambda)\lambda s^2. \tag{3.10}$$

Therefore the equality holds in (3.9) and (3.10), and it yields

$$\frac{1}{2\lambda st}\big\{(\lambda s)^2+t^2-\sigma_\mu\big((\alpha,\lambda s),(\beta,t)\big)^2\big\}=\frac{1}{2st}\big\{s^2+t^2-\sigma_\mu\big((\alpha,s),(\beta,t)\big)^2\big\}.$$

This completes the proof. □

As in the case of Alexandrov spaces, for $\mu\in\mathcal{P}(X)$, we define $\Sigma'_\mu[\mathcal{P}(X)]$ as the set of all unit speed geodesics emanating from μ equipped with the equivalence relation such that $\alpha\sim\beta$ if they coincide near μ. Then, by taking Theorem 3.6 into account, we can define the angle $\angle_\mu(\alpha,\beta)\in[0,\pi]$ between $\alpha,\beta\in\Sigma'_\mu[\mathcal{P}(X)]$ by

$$\cos\angle_\mu(\alpha,\beta):=\frac{1}{2}\big\{2-\sigma_\mu\big((\alpha,1),(\beta,1)\big)^2\big\}.$$

The angle $\angle_\mu$ is a pseudo-distance function on $\Sigma'_\mu[\mathcal{P}(X)]$ and the *space of directions* $(\Sigma_\mu[\mathcal{P}(X)],\angle_\mu)$ at μ is the completion of $(\Sigma'_\mu[\mathcal{P}(X)]/\{\angle_\mu=0\},\angle_\mu)$. We define the *tangent cone* $(C_\mu[\mathcal{P}(X)],\sigma_\mu)$ at μ as the Euclidean cone over $(\Sigma_\mu[\mathcal{P}(X)],\angle_\mu)$. Here we abused the symbol 'σ_μ', but Theorem 3.6 ensures that it is compatible. We will sometimes identify $\alpha\in\Sigma_\mu[\mathcal{P}(X)]$ with $(\alpha,1)\in C_\mu[\mathcal{P}(X)]$. For $(\alpha,s),(\beta,t)\in C_\mu[\mathcal{P}(X)]$, we define the *inner product* of them by

$$\langle(\alpha,s),(\beta,t)\rangle_\mu:=st\cos\angle_\mu(\alpha,\beta)=\frac{1}{2}\big\{s^2+t^2-\sigma_\mu\big((\alpha,s),(\beta,t)\big)^2\big\}. \tag{3.11}$$

In addition, for later convenience, we define $(C'_\mu[\mathcal{P}(X)],\sigma_\mu)$ as the Euclidean cone over $(\Sigma'_\mu[\mathcal{P}(X)],\angle_\mu)$. Note that $(C_\mu[\mathcal{P}(X)],\sigma_\mu)$ is the completion of $(C'_\mu[\mathcal{P}(X)]/\{\sigma_\mu=0\},\sigma_\mu)$.

Remark 3.7 The proof of Theorem 3.6 is also applicable to the noncompact case by restricting $\mathcal{P}(X)$ to a subset, say $\mathcal{P}_c(X)$, consisting of measures of compact supports. However, $\mathcal{P}_c(X)$ is not complete with respect to d_2^W.

4 K-convex and lower semi-continuous functions

This section is devoted to recalling important properties of K-convex and lower semi-continuous functions on $\mathcal{P}(X)$. Throughout the section, (X,d) is a compact Alexandrov space of curvature bounded below, and a function $f:\mathcal{P}(X)\longrightarrow(-\infty,\infty]$ is always assumed to be

$$\text{nontrivial, } K\text{-convex and lower semi-continuous.} \tag{4.1}$$

Here a function f is said to be *nontrivial* if $\mathcal{P}^*(X)=\{\mu\in\mathcal{P}(X)\,|\,f(\mu)<\infty\}\neq\emptyset$. The *$K$-convexity* of f for $K\in\mathbb{R}$ means that, for any minimal geodesic $\alpha:[0,1]\longrightarrow\mathcal{P}(X)$ and $\lambda\in[0,1]$, we have

$$f\big(\alpha(\lambda)\big)\le(1-\lambda)f\big(\alpha(0)\big)+\lambda f\big(\alpha(1)\big)-\frac{K}{2}(1-\lambda)\lambda d_2^W\big(\alpha(0),\alpha(1)\big)^2.$$

In particular, given $\mu, \nu \in \mathcal{P}^*(X)$, every minimal geodesic between them is contained in $\mathcal{P}^*(X)$. Therefore we can define $\Sigma'_\mu[\mathcal{P}^*(X)]$, $\Sigma_\mu[\mathcal{P}^*(X)]$, $C'_\mu[\mathcal{P}^*(X)]$ and $C_\mu[\mathcal{P}^*(X)]$ in similar manners (see the paragraph following Theorem 3.6), and they are isometrically embedded into $\Sigma'_\mu[\mathcal{P}(X)]$, $\Sigma_\mu[\mathcal{P}(X)]$, $C'_\mu[\mathcal{P}(X)]$ and $C_\mu[\mathcal{P}(X)]$ by inclusions, respectively. We remark that, if we put $\omega := \inf_{\mu \in \mathcal{P}(X)} f(\mu)$, then it follows from the lower semi-continuity of f and the compactness of $\mathcal{P}(X)$ that ω is attained at some point in $\mathcal{P}^*(X)$, and hence $\omega > -\infty$. One of the most important examples of functions satisfying (4.1) is the free energy (as well as the relative entropy) which will be studied in Subsection 6.2.

4.1 Gradient vectors

For $f : \mathcal{P}(X) \longrightarrow (-\infty, \infty]$ satisfying (4.1), define the *absolute gradient* $|\nabla_- f|(\mu) \in [0, \infty]$ of f at $\mu \in \mathcal{P}^*(X)$ by

$$|\nabla_- f|(\mu) := \max\left\{0, \limsup_{\mathcal{P}^*(X)\setminus\{\mu\} \ni \nu \to \mu} \frac{f(\mu) - f(\nu)}{d_2^W(\mu, \nu)}\right\}, \tag{4.2}$$

where the convergence $\nu \to \mu$ is with respect to d_2^W. Note that $|\nabla_- f|(\mu) = 0$ holds if $f(\mu) = \omega = \inf_{\nu \in \mathcal{P}(X)} f(\nu)$.

Fix $\mu \in \mathcal{P}^*(X)$ with $|\nabla_- f|(\mu) < \infty$. For $v = (\alpha, s) \in C'_\mu[\mathcal{P}^*(X)]$, i.e., a unit speed minimal geodesic $\alpha : [0, \delta] \longrightarrow \mathcal{P}^*(X)$ with $\alpha(0) = \mu$ and $s \geq 0$, we define

$$D'_\mu f(v) := \lim_{\varepsilon \to 0+} \frac{f(\alpha(s\varepsilon)) - f(\mu)}{\varepsilon}.$$

Note that the limit above exists, for $f \circ \alpha$ is K-convex, and that $D'_\mu f(v) \geq -s|\nabla_- f|(\mu)$. Moreover, the K-convexity also implies that

$$\begin{aligned} & D'_\mu f(v) \\ & \leq \lim_{\varepsilon \to 0+} \frac{1}{\varepsilon}\left\{\left(1 - \frac{s\varepsilon}{\delta}\right) f(\mu) + \frac{s\varepsilon}{\delta} f\big(\alpha(\delta)\big) - \frac{K}{2}\left(1 - \frac{s\varepsilon}{\delta}\right)\frac{s\varepsilon}{\delta}\delta^2 - f(\mu)\right\} \\ & = \frac{s}{\delta}\big\{f\big(\alpha(\delta)\big) - f(\mu)\big\} - \frac{K}{2} s\delta. \end{aligned} \tag{4.3}$$

We further define the function $D_\mu f : C_\mu[\mathcal{P}^*(X)] \longrightarrow (-\infty, \infty]$ by

$$D_\mu f(v) := \liminf_{C'_\mu[\mathcal{P}^*(X)] \ni w \to v} D'_\mu f(w) \tag{4.4}$$

for $v = (\alpha, s) \in C_\mu[\mathcal{P}^*(X)]$. Clearly we have $D_\mu f(v) \geq -s|\nabla_- f|(\mu)$, $D_\mu f(v) = s \cdot D_\mu f((\alpha, 1))$ and also $D_\mu f(v) \leq D'_\mu f(v)$ if $v \in C'_\mu[\mathcal{P}^*(X)]$. The following lemma means that $D_\mu f$ is 'almost' convex. The convexity is easily verified by taking a scaling limit if the space in question is of finite dimension (see Proposition 2.3). However, this is not our case because $\mathcal{P}(X)$ is obviously infinite dimensional even when X is of finite dimension.

Lemma 4.1 *Fix a point $\mu \in \mathcal{P}^*(X)$ with $|\nabla_- f|(\mu) < \infty$ and $v, w \in C_\mu[\mathcal{P}^*(X)]$ with $D_\mu f(v), D_\mu f(w) < \infty$. Then, for any $\varepsilon > 0$, there exists some $u \in C_\mu[\mathcal{P}^*(X)]$ for which we have*

$$D_\mu f(u) \le \frac{1}{2} D_\mu f(v) + \frac{1}{2} D_\mu f(w) + \varepsilon, \tag{4.5}$$

$$\sigma_\mu(o_\mu, u)^2 \le \frac{1}{2}\sigma_\mu(o_\mu, v)^2 + \frac{1}{2}\sigma_\mu(o_\mu, w)^2 - \frac{1}{4}\sigma_\mu(v, w)^2 + \varepsilon. \tag{4.6}$$

Proof. It suffices to treat the case of $K = -1$. Put $v = (\alpha, s)$ and $w = (\beta, t)$. If $s = 0$ or $t = 0$, then we just take $u = (\beta, t/2)$ or $u = (\alpha, s/2)$, respectively. Thus, without loss of generality, we may assume $s, t > 0$.

We first suppose that $\alpha, \beta \in \Sigma'_\mu[\mathcal{P}^*(X)]$ and $s \ne t$, and show the analogues of (4.5) and (4.6) for $D'_\mu f$ instead of $D_\mu f$. Note that, for small $\lambda \in (0, 1]$,

$$\frac{f(\alpha(\lambda s)) - f(\mu)}{\lambda} = D'_\mu f(v) + \theta(\lambda), \quad \frac{f(\beta(\lambda t)) - f(\mu)}{\lambda} = D'_\mu f(w) + \theta(\lambda).$$

Let $\xi_\lambda : [0, 1] \longrightarrow \mathcal{P}^*(X)$ be a minimal geodesic from $\alpha(\lambda s)$ to $\beta(\lambda t)$ and set $\nu_\lambda := \xi_\lambda(1/2)$. We also choose a minimal geodesic $\zeta_\lambda : [0, d_2^W(\mu, \nu_\lambda)] \longrightarrow \mathcal{P}^*(X)$ from μ to ν_λ and put

$$u_\lambda := \big(\zeta_\lambda, d_2^W(\mu, \nu_\lambda)/\lambda\big) \in C'_\mu[\mathcal{P}^*(X)].$$

Then (4.3) with $\delta = d_2^W(\mu, \nu_\lambda)$ and $s = d_2^W(\mu, \nu_\lambda)/\lambda$ as well as the (-1)-convexity of f implies that

$$\begin{aligned}
D'_\mu f(u_\lambda) &\le \frac{f(\nu_\lambda) - f(\mu)}{\lambda} + \frac{d_2^W(\mu, \nu_\lambda)^2}{2\lambda} \\
&\le \frac{1}{\lambda}\left\{ \frac{1}{2} f\big(\alpha(\lambda s)\big) + \frac{1}{2} f\big(\beta(\lambda t)\big) + \frac{1}{8} d_2^W\big(\alpha(\lambda s), \beta(\lambda t)\big)^2 - f(\mu) \right\} \\
&\quad + \frac{1}{2\lambda}(\lambda s + \lambda t)^2 \\
&\le \frac{f(\alpha(\lambda s)) - f(\mu)}{2\lambda} + \frac{f(\beta(\lambda t)) - f(\mu)}{2\lambda} + \frac{5}{8}\lambda(s+t)^2 \\
&= \frac{1}{2} D'_\mu f(v) + \frac{1}{2} D'_\mu f(w) + \theta(\lambda).
\end{aligned}$$

Thus we obtain (4.5) for $D'_\mu f$ by taking sufficiently small $\lambda > 0$.

Now we prove (4.6), more precisely,

$$\sigma_\mu(o_\mu, u_\lambda)^2 \le \frac{1}{2}(s^2 + t^2) - \frac{1}{4}\sigma_\mu(v, w)^2 + \theta(\lambda). \tag{4.7}$$

Suppose the contrary, namely there is $\varepsilon > 0$ such that, for any $i \in \mathbb{N}$, we find $\lambda_i \in (0, i^{-1}]$ satisfying

$$\sigma_\mu(o_\mu, u_{\lambda_i})^2 \ge \frac{1}{2}(s^2 + t^2) - \frac{1}{4}\sigma_\mu(v, w)^2 + \varepsilon. \tag{4.8}$$

Then we put

$$a_i := \left\{ \frac{\sigma_\mu(o_\mu, u_{\lambda_i})^2}{(s^2+t^2)/2 - \sigma_\mu(v,w)^2/4} \right\}^{1/2}.$$

Note that our assumption $s \neq t$ guarantees $\sigma_\mu(v,w)^2/4 \le (s+t)^2/4 < (s^2+t^2)/2$ and that, for any $i \in \mathbb{N}$,

$$\begin{aligned} a_i^2 &\ge 1 + \frac{\varepsilon}{(s^2+t^2)/2 - \sigma_\mu(v,w)^2/4} \ge 1 + \frac{2\varepsilon}{s^2+t^2}, \\ a_i^2 &= \frac{d_2^W(\mu, \nu_{\lambda_i})^2/\lambda_i^2}{(s^2+t^2)/2 - \sigma_\mu(v,w)^2/4} \le \frac{(s+t)^2}{(s^2+t^2)/2 - \sigma_\mu(v,w)^2/4}. \end{aligned}$$

It follows from Lemma 3.5 that

$$\begin{aligned} d_2^W\big(\nu_{\lambda_i}, \alpha(\lambda_i s)\big)^2 &\ge \big(1+\theta(\lambda_i)\big)\left\{ \frac{a_i - 1}{a_i} d_2^W(\nu_{\lambda_i}, \mu)^2 + \frac{1}{a_i} d_2^W\big(\nu_{\lambda_i}, \alpha(\lambda_i a_i s)\big)^2 \right\} \\ &\qquad - \big(1+\theta(\lambda_i)\big)\frac{a_i-1}{a_i^2} d_2^W\big(\mu, \alpha(\lambda_i a_i s)\big)^2 \\ &= \frac{a_i-1}{a_i}\lambda_i^2 \sigma_\mu(o_\mu, u_{\lambda_i})^2 + \frac{1}{a_i} d_2^W\big(\nu_{\lambda_i}, \alpha(\lambda_i a_i s)\big)^2 \\ &\qquad - (a_i - 1)\lambda_i^2 s^2 + \lambda_i^2 \theta(\lambda_i) \end{aligned}$$

and, similarly,

$$\begin{aligned} d_2^W\big(\nu_{\lambda_i}, \beta(\lambda_i t)\big)^2 \ge \frac{a_i-1}{a_i}\lambda_i^2\sigma_\mu(o_\mu, u_{\lambda_i})^2 + \frac{1}{a_i} d_2^W\big(\nu_{\lambda_i}, \beta(\lambda_i a_i t)\big)^2 \\ - (a_i-1)\lambda_i^2 t^2 + \lambda_i^2\theta(\lambda_i). \end{aligned}$$

Since ν_{λ_i} is a midpoint between $\alpha(\lambda_i s)$ and $\beta(\lambda_i t)$, we deduce that

$$\begin{aligned} \lambda_i^2 \sigma_\mu(v,w)^2 &= d_2^W\big(\alpha(\lambda_i s), \beta(\lambda_i t)\big)^2 + \lambda_i^2\theta(\lambda_i) \\ &= 2 d_2^W\big(\alpha(\lambda_i s), \nu_{\lambda_i}\big)^2 + 2 d_2^W\big(\nu_{\lambda_i}, \beta(\lambda_i t)\big)^2 + \lambda_i^2\theta(\lambda_i) \\ &\ge \frac{a_i-1}{a_i} 4\lambda_i^2 \sigma_\mu(o_\mu, u_{\lambda_i})^2 + \frac{2}{a_i}\left\{ d_2^W\big(\nu_{\lambda_i}, \alpha(\lambda_i a_i s)\big)^2 + d_2^W\big(\nu_{\lambda_i}, \beta(\lambda_i a_i t)\big)^2 \right\} \\ &\qquad - 2(a_i - 1)\lambda_i^2(s^2+t^2) + \lambda_i^2\theta(\lambda_i) \\ &\ge \frac{a_i-1}{a_i} 4\lambda_i^2\sigma_\mu(o_\mu, u_{\lambda_i})^2 + \frac{1}{a_i} d_2^W\big(\alpha(\lambda_i a_i s), \beta(\lambda_i a_i t)\big)^2 \\ &\qquad - 2(a_i-1)\lambda_i^2(s^2+t^2) + \lambda_i^2\theta(\lambda_i). \end{aligned}$$

By the definition of a_i, the right hand side is equal to

$$\begin{aligned} &\frac{a_i-1}{a_i} 4\lambda_i^2\sigma_\mu(o_\mu, u_{\lambda_i})^2 + \lambda_i^2 a_i \sigma_\mu(v,w)^2 \\ &\qquad - (a_i - 1)\lambda_i^2\left\{ \sigma_\mu(v,w)^2 + \frac{4}{a_i^2}\sigma_\mu(o_\mu, u_{\lambda_i})^2 \right\} + \lambda_i^2\theta(\lambda_i) \\ &= \left(\frac{a_i-1}{a_i}\right)^2 4\lambda_i^2\sigma_\mu(o_\mu, u_{\lambda_i})^2 + \lambda_i^2\sigma_\mu(v,w)^2 + \lambda_i^2\theta(\lambda_i). \end{aligned}$$

Thus we see $(a_i - 1)/a_i \cdot \sigma_\mu(o_\mu, u_{\lambda_i}) = \theta(\lambda_i)$. As $a_i^2 - 1 \geq 2\varepsilon/(s^2 + t^2) > 0$ uniformly in $i \in \mathbb{N}$, we have $\sigma_\mu(o_\mu, u_{\lambda_i}) = \theta(\lambda_i)$, it contradicts (4.8). Therefore we obtain (4.7).

For general $\alpha \in \Sigma_\mu[\mathcal{P}^*(X)]$, let us take a sequence $\{\alpha_i\}_{i\in\mathbb{N}} \subset \Sigma'_\mu[\mathcal{P}^*(X)]$ which converges to α and satisfies $\lim_{i\to\infty} D'_\mu f(v_i) = D_\mu f(v)$, where we put $v_i := (\alpha_i, s) \in C'_\mu[\mathcal{P}^*(X)]$. Choose $\{\beta_i\}_{i\in\mathbb{N}} \subset \Sigma'_\mu[\mathcal{P}^*(X)]$ in a similar manner, and put $w_i := (\beta_i, t_i) \in C'_\mu[\mathcal{P}^*(X)]$, where $\{t_i\}_{i\in\mathbb{N}} \subset (0, \infty)$ is a sequence satisfying $\lim_{i\to\infty} t_i = t$, $t_i \leq t$ and $t_i \neq s$. For sufficiently large $i \in \mathbb{N}$, we observe

$$D'_\mu f(v_i) \leq D_\mu f(v) + \frac{\varepsilon}{2}, \qquad D'_\mu f(w_i) \leq D_\mu f(w) + \frac{\varepsilon}{2}, \tag{4.9}$$

$$\sigma_\mu(v_i, w_i)^2 \geq \sigma_\mu(v, w)^2 - 2\varepsilon. \tag{4.10}$$

As $\alpha_i, \beta_i \in \Sigma'_\mu[\mathcal{P}^*(X)]$, the first part of the proof guarantees that there exists some $u \in C'_\mu[\mathcal{P}^*(X)]$ satisfying

$$D'_\mu f(u) \leq \frac{1}{2} D'_\mu f(v_i) + \frac{1}{2} D'_\mu f(w_i) + \frac{\varepsilon}{2},$$
$$\sigma_\mu(o_\mu, u)^2 \leq \frac{1}{2}\sigma_\mu(o_\mu, v_i)^2 + \frac{1}{2}\sigma_\mu(o_\mu, w_i)^2 - \frac{1}{4}\sigma_\mu(v_i, w_i)^2 + \frac{\varepsilon}{2}.$$

Combining these with $D_\mu f(u) \leq D'_\mu f(u)$, (4.9) and (4.10), we obtain

$$D_\mu f(u) \leq \frac{1}{2} D_\mu f(v) + \frac{1}{2} D_\mu f(w) + \varepsilon,$$
$$\sigma_\mu(o_\mu, u)^2 \leq \frac{1}{2}\sigma_\mu(o_\mu, v)^2 + \frac{1}{2}\sigma_\mu(o_\mu, w)^2 - \frac{1}{4}\sigma_\mu(v, w)^2 + \varepsilon.$$

□

The convexity of $D_\mu f$ enables us to find a unique steepest direction of f.

Lemma 4.2 *For any $\mu \in \mathcal{P}^*(X)$ with $0 < |\nabla_- f|(\mu) < \infty$, there exists unique $\alpha \in \Sigma_\mu[\mathcal{P}^*(X)]$ satisfying $D_\mu f(\alpha) = -|\nabla_- f|(\mu)$. Moreover, for any $\beta \in \Sigma_\mu[\mathcal{P}^*(X)]$, we have $D_\mu f(\beta) \geq -|\nabla_- f|(\mu) \cdot \langle \alpha, \beta \rangle_\mu$. Here we identified $\alpha, \beta \in \Sigma_\mu[\mathcal{P}^*(X)]$ with $(\alpha, 1), (\beta, 1) \in C_\mu[\mathcal{P}^*(X)]$, respectively.*

Proof. By the definition of $|\nabla_- f|(\mu)$ and (4.3), we can take a sequence $\{\alpha_i\}_{i\in\mathbb{N}} \subset \Sigma'_\mu[\mathcal{P}^*(X)]$ such that $\lim_{i\to\infty} D_\mu f(\alpha_i) = -|\nabla_- f|(\mu)$. For $i, j \in \mathbb{N}$ and arbitrary $\varepsilon > 0$, Lemma 4.1 shows that there is $u = (\xi, \tau) \in C_\mu[\mathcal{P}^*(X)]$ with

$$D_\mu f(u) \leq \frac{1}{2} D_\mu f(\alpha_i) + \frac{1}{2} D_\mu f(\alpha_j) + \varepsilon, \quad \tau^2 \leq 1 - \frac{1}{4}\sigma_\mu\big((\alpha_i, 1), (\alpha_j, 1)\big)^2 + \varepsilon.$$

Combining these with $D_\mu f(u) \geq -\tau |\nabla_- f|(\mu)$, we observe

$$\frac{1}{2} D_\mu f(\alpha_i) + \frac{1}{2} D_\mu f(\alpha_j) \geq D_\mu f(u) - \varepsilon \geq -\tau |\nabla_- f|(\mu) - \varepsilon$$
$$\geq -\left\{1 - \frac{1}{4}\sigma_\mu\big((\alpha_i, 1), (\alpha_j, 1)\big)^2 + \varepsilon\right\}^{1/2} |\nabla_- f|(\mu) - \varepsilon.$$

As ε is arbitrary and $|\nabla_- f|(\mu) > 0$, we find that $\{\alpha_i\}_{i\in\mathbb{N}}$ is a Cauchy sequence, and hence it converges to some $\alpha \in \Sigma_\mu[\mathcal{P}^*(X)]$. By the choice of $\{\alpha_i\}_{i\in\mathbb{N}}$, α satisfies $D_\mu f(\alpha) = -|\nabla_- f|(\mu)$ and the uniqueness of α also follows from Lemma 4.1.

Next we consider the second assertion. Clearly we can assume $D_\mu f(\beta) < \infty$. For each $i \in \mathbb{N}$, by Lemma 4.1, we find $u_i \in C_\mu[\mathcal{P}^*(X)]$ satisfying

$$D_\mu f(u_i) \le \frac{1}{2} D_\mu f\big((\alpha, i)\big) + \frac{1}{2} D_\mu f(\beta) + i^{-1} = -\frac{i}{2}|\nabla_- f|(\mu) + \frac{1}{2} D_\mu f(\beta) + i^{-1},$$

$$\sigma_\mu(o_\mu, u_i)^2 \le \frac{i^2}{2} + \frac{1}{2} - \frac{1}{4}\sigma_\mu\big((\alpha, i), (\beta, 1)\big)^2 + i^{-1} = \frac{i^2}{4} + \frac{1}{4} + \frac{i}{2}\langle \alpha, \beta\rangle_\mu + i^{-1}.$$

Thus we see

$$D_\mu f(u_i) \ge -\sigma_\mu(o_\mu, u_i)|\nabla_- f|(\mu) \ge -\frac{1}{2}\{i^2 + 1 + 2i\langle \alpha, \beta\rangle_\mu + 4i^{-1}\}^{1/2}|\nabla_- f|(\mu),$$

and hence

$$\begin{aligned} D_\mu f(\beta) &\ge 2D_\mu f(u_i) + i|\nabla_- f|(\mu) - 2i^{-1} \\ &\ge \big[i - \{i^2 + 1 + 2i\langle \alpha, \beta\rangle_\mu + 4i^{-1}\}^{1/2}\big]|\nabla_- f|(\mu) - 2i^{-1} \\ &= \frac{-1 - 2i\langle \alpha, \beta\rangle_\mu - 4i^{-1}}{i + \{i^2 + 1 + 2i\langle \alpha, \beta\rangle_\mu + 4i^{-1}\}^{1/2}}|\nabla_- f|(\mu) - 2i^{-1} \\ &\to -\langle \alpha, \beta\rangle_\mu \cdot |\nabla_- f|(\mu) \end{aligned}$$

as i diverges to infinity. This completes the proof. $\square$

We define the *negative gradient vector* of f at μ with $0 < |\nabla_- f|(\mu) < \infty$ by

$$\nabla_- f(\mu) := \big(\alpha, |\nabla_- f|(\mu)\big) \in C_\mu[\mathcal{P}^*(X)], \tag{4.11}$$

where $\alpha \in \Sigma_\mu[\mathcal{P}^*(X)]$ is the unique element obtained in Lemma 4.2. If $|\nabla_- f|(\mu) = 0$, then we simply define $\nabla_- f(\mu) := o_\mu$. In the smooth case, $\nabla_- f(\mu)$ corresponds to $-\operatorname{grad} f(\mu)$.

4.2 Upper gradients

A nonnegative, Borel function $g : \mathcal{P}^*(X) \longrightarrow [0, \infty]$ is called an *upper gradient* for f if, for every Lipschitz curve $\eta : [0, l] \longrightarrow \mathcal{P}^*(X)$, it holds that

$$\big|f\big(\eta(0)\big) - f\big(\eta(l)\big)\big| \le \int_0^l g\big(\eta(t)\big)|\eta'|(t)\, dt, \tag{4.12}$$

where we set $|\eta'|(t) := \lim_{s\to t} d_2^W(\eta(s), \eta(t))/|s-t|$. We remark that the *metric derivative* $|\eta'|$ exists a.e. on $[0, l]$ (see [AGS, Theorem 1.1.2]). The following lemma is useful.

Lemma 4.3 ([AGS, Corollary 2.4.10]) *The absolute gradient $|\nabla_- f| : \mathcal{P}^*(X) \longrightarrow [0, \infty]$ of f is lower semi-continuous and is an upper gradient for f.*

5 Gradient flows on Wasserstein spaces

In this section, we formulate and construct a gradient flow $G : \mathcal{P}^*(X) \times [0, \infty) \longrightarrow \mathcal{P}^*(X)$ of a function f on $\mathcal{P}(X)$ by using the Riemannian structure of $\mathcal{P}(X)$ established in Section 3. There are at least two strategies for constructing gradient flows by way of discrete approximations. The first one is a direct approximation considered in [JKO] and [AGS] (cf. Subsection 6.1) and, in that context, most results in this section are covered in [AGS, Chapter 2] (see, especially, [AGS, Corollary 2.4.11, Theorem 2.4.15]). However, here we adopt the other one, a two-step approximation, which extends a discussion in [Ly2] and seems to be of independent interest. Compared with [Ly2], we need to be more careful at some points because of the discontinuity of f, so that we will give all proofs for completeness. Throughout the section, let (X, d) be a compact Alexandrov space of curvature bounded below, and let $f : \mathcal{P}(X) \longrightarrow (-\infty, \infty]$ be a function satisfying the condition (4.1).

5.1 Gradient-like curves

In this subsection, we construct a gradient-like curve which will turn out to be a unit speed curve whose reparametrization produces a gradient curve.

For $a \in [\omega, \infty)$, denote a sublevel set of f by $U[a] := f^{-1}([\omega, a])$. Since f is lower semi-continuous and $\mathcal{P}(X)$ is compact, the set $U[a]$ is compact. Note also that $U[a] \supset U[\omega] \neq \emptyset$.

Lemma 5.1 *Given $\mu \in \mathcal{P}^*(X)$ and $C, r > 0$, suppose that $|\nabla_- f|(\nu) \geq C$ holds for all $\nu \in \overline{B}(\mu, r) \cap \mathcal{P}^*(X)$. Then, for each $c \in (0, \min\{Cr, f(\mu) - \omega\}]$, there exists $\nu \in \overline{B}(\mu, r) \cap \mathcal{P}^*(X)$ satisfying*

$$f(\nu) = f(\mu) - c, \quad d_2^W(\mu, \nu) = \operatorname{dist}\big(\mu, U[f(\mu) - c]\big). \tag{5.1}$$

In particular, we have $\operatorname{dist}(\mu, U[f(\mu) - c]) \leq r$.

Proof. First of all, we observe the existence of $\nu \in \mathcal{P}^*(X)$ satisfying the conditions (5.1). By the hypothesis, we know $f(\mu) - c \geq \omega$ and $U[f(\mu) - c] \neq \emptyset$. As $U[f(\mu) - c]$ is compact, we can take $\nu \in U[f(\mu) - c]$ satisfying $d_2^W(\mu, \nu) = \operatorname{dist}(\mu, U[f(\mu) - c])$. Note that $f(\nu) \leq f(\mu) - c$. Let $\alpha : [0, 1] \longrightarrow \mathcal{P}^*(X)$ be a minimal geodesic from μ to ν. Since the function $f \circ \alpha : [0, 1] \longrightarrow [\omega, \infty)$ is K-convex and lower semi-continuous, it is continuous and we find $\lambda \in [0, 1]$ such that $f(\alpha(\lambda)) = f(\mu) - c$. If $\lambda < 1$, then we see

$$d_2^W\big(\mu, \alpha(\lambda)\big) < d_2^W(\mu, \nu) = \operatorname{dist}\big(\mu, U[f(\mu) - c]\big),$$

this is a contradiction. Therefore we have $\lambda = 1$, and hence $f(\nu) = f(\mu) - c$.

It remains to show $d_2^W(\mu, \nu) \leq r$, that is, $U[f(\mu) - c] \cap \overline{B}(\mu, r) \neq \emptyset$. Let $A \subset [0, 1]$ be the maximal subset such that $U[f(\mu) - ca] \cap \overline{B}(\mu, ra) \neq \emptyset$ for each $a \in A$. Note that $0 \in A$ and it suffices to show $1 \in A$. It follows from the compactness of $\mathcal{P}(X)$ and the lower semi-continuity of f that A is closed. Hence we have $a := \sup_{a' \in A} a' \in A$. Now we suppose $a < 1$ and will derive a contradiction. We first consider the case of

$c < Cr$. Take $\nu \in U[f(\mu) - ca] \cap \overline{B}(\mu, ra)$. Then it holds that $|\nabla_- f|(\nu) \ge C > c/r$ by the hypothesis, so that we can choose $\nu' \in \mathcal{P}^*(X) \setminus \{\nu\}$ with $\delta := d_2^W(\nu', \nu)/r \le 1 - a$ and $f(\nu) - f(\nu') \ge (c/r) d_2^W(\nu, \nu')$. We observe

$$d_2^W(\mu, \nu') \le d_2^W(\mu, \nu) + d_2^W(\nu, \nu') \le r(a + \delta),$$
$$f(\nu') \le f(\nu) - (c/r) d_2^W(\nu, \nu') \le f(\mu) - c(a + \delta).$$

Therefore we find $U[f(\mu) - c(a+\delta)] \cap \overline{B}(\mu, r(a+\delta)) \neq \emptyset$, it implies $a + \delta \in A$ and contradicts the maximality of a. Thus we obtain $1 \in A$. In case of $c = Cr$, we take an increasing sequence $\{c_i\}_{i\in\mathbb{N}}$ tending to c. Then we deduce from the above argument that $U[f(\mu) - c_i] \cap \overline{B}(\mu, r) \neq \emptyset$ for every $i \in \mathbb{N}$, and this together with the compactness of $\mathcal{P}(X)$ and the lower semi-continuity of f shows that $U[f(\mu) - c] \cap \overline{B}(\mu, r) \neq \emptyset$. $\square$

Corollary 5.2 *For every $\mu \in \mathcal{P}^*(X)$ and $r > 0$, we have*

$$\inf_{\nu \in \overline{B}(\mu, r) \cap \mathcal{P}^*(X)} |\nabla_- f|(\nu) \le \frac{1}{r}\{f(\mu) - \omega\} < \infty.$$

In particular, $Cr \le f(\mu) - \omega$ automatically holds in Lemma 5.1.

Proof. Suppose the contrary. Then we can apply Lemma 5.1 with $c = f(\mu) - \omega$ and obtain $U[\omega] \cap \overline{B}(\mu, r) \neq \emptyset$. This is a contradiction because it implies

$$0 = \inf_{\nu \in \overline{B}(\mu, r) \cap \mathcal{P}^*(X)} |\nabla_- f|(\nu) > \frac{1}{r}\{f(\mu) - \omega\} \ge 0.$$

$\square$

Now we define a gradient-like curve and observe some straightforward properties.

Definition 5.3 (Gradient-like curves) A 1-Lipschitz curve $\eta : [0, l) \longrightarrow \mathcal{P}^*(X)$ is called a *gradient-like curve* of f if we have $|\nabla_- f|(\eta(t)) > 0$ and

$$f\big(\eta(t)\big) = f\big(\eta(0)\big) - \int_0^t |\nabla_- f|\big(\eta(s)\big)\, ds \tag{5.2}$$

for all $t \in [0, l)$.

Definition 5.4 A curve $\eta : [0, \delta) \longrightarrow \mathcal{P}(X)$ is said to be *right differentiable* at 0 if there is $v \in C_{\eta(0)}[\mathcal{P}(X)]$ such that, for any sequence $\{\varepsilon_i\}_{i\in\mathbb{N}}$ tending to zero and any sequence of unit speed minimal geodesics $\{\alpha_i\}_{i\in\mathbb{N}}$ from $\eta(0)$ to $\eta(\varepsilon_i)$, the sequence $\{(\alpha_i, d_2^W(\eta(0), \eta(\varepsilon_i))/\varepsilon_i)\}_{i\in\mathbb{N}}$ converges to v in $C_{\eta(0)}[\mathcal{P}(X)]$. Then we write $\eta'(0) = v$.

Lemma 5.5 *Let $\eta : [0, l) \longrightarrow \mathcal{P}^*(X)$ be a gradient-like curve of f. Then the following properties hold.*

(i) *We have $|\nabla_- f|(\eta(t)) < \infty$ for a.e. $t \in [0, l)$ and*

$$\lim_{\varepsilon \to 0+} \frac{f(\eta(t+\varepsilon)) - f(\eta(t))}{\varepsilon} = -|\nabla_- f|\big(\eta(t)\big) \tag{5.3}$$

for all $t \in [0, l)$.

(ii) *At every* $t \in [0, l)$ *with* $|\nabla_- f|(\eta(t)) < \infty$*,* η *is right differentiable and* $\eta'(t) = (\alpha, 1)$*, where* $\nabla_- f(\eta(t)) = (\alpha, |\nabla_- f|(\eta(t))) \in C_{\eta(t)}[\mathcal{P}^*(X)]$ *(see* (4.11)*). In particular, the curve* η *has a unit speed.*

(iii) *If* $l < \infty$*, then the limit* $\eta(l) := \lim_{t \to l} \eta(t)$ *exists and it satisfies*

$$f\big(\eta(l)\big) = f\big(\eta(0)\big) - \int_0^l |\nabla_- f|\big(\eta(s)\big)\, ds. \tag{5.4}$$

Proof. (i) The equation (5.2) immediately implies that $|\nabla_- f|(\eta(t)) < \infty$ holds for a.e. $t \in [0, l)$. By the 1-Lipschitz continuity of η, for all $t \in [0, l)$, we see

$$-|\nabla_- f|\big(\eta(t)\big) \le \liminf_{\varepsilon \to 0+} \frac{f(\eta(t+\varepsilon)) - f(\eta(t))}{d_2^W(\eta(t+\varepsilon), \eta(t))} \le \liminf_{\varepsilon \to 0+} \frac{f(\eta(t+\varepsilon)) - f(\eta(t))}{\varepsilon}.$$

Moreover, the lower semi-continuity of $|\nabla_- f|$ (Lemma 4.3) yields that

$$\limsup_{\varepsilon \to 0+} \frac{f(\eta(t+\varepsilon)) - f(\eta(t))}{\varepsilon} = -\liminf_{\varepsilon \to 0+} \frac{1}{\varepsilon} \int_t^{t+\varepsilon} |\nabla_- f|\big(\eta(s)\big)\, ds \le -|\nabla_- f|\big(\eta(t)\big).$$

These show (5.3).

(ii) Fix $t \in [0, l)$ for which $|\nabla_- f|(\eta(t)) < \infty$. We observe that, by (i) and the definition of $|\nabla_- f|$,

$$\begin{aligned} |\nabla_- f|\big(\eta(t)\big) &= \lim_{\varepsilon \to 0+} \frac{f(\eta(t)) - f(\eta(t+\varepsilon))}{\varepsilon} \\ &\le \limsup_{\varepsilon \to 0+} \frac{f(\eta(t)) - f(\eta(t+\varepsilon))}{d_2^W(\eta(t), \eta(t+\varepsilon))} \cdot \liminf_{\varepsilon \to 0+} \frac{d_2^W(\eta(t), \eta(t+\varepsilon))}{\varepsilon} \\ &\le |\nabla_- f|\big(\eta(t)\big) \cdot \liminf_{\varepsilon \to 0+} \frac{d_2^W(\eta(t), \eta(t+\varepsilon))}{\varepsilon}. \end{aligned}$$

Combining this with the 1-Lipschitz continuity of η and $|\nabla_- f|(\eta(t)) > 0$, we find that

$$\lim_{\varepsilon \to 0+} \frac{d_2^W(\eta(t), \eta(t+\varepsilon))}{\varepsilon} = 1. \tag{5.5}$$

Given a decreasing sequence $\{\varepsilon_i\}_{i \in \mathbb{N}} \subset (0, l-t)$ tending to zero and a sequence $\{\alpha_i\}_{i \in \mathbb{N}} \subset \Sigma'_{\eta(t)}[\mathcal{P}^*(X)]$ of minimal geodesics from $\eta(t)$ to $\eta(t+\varepsilon_i)$, put

$$v_i := \big(\alpha_i, d_2^W(\eta(t), \eta(t+\varepsilon_i))/\varepsilon_i\big) \in C'_{\eta(t)}[\mathcal{P}^*(X)].$$

Then, by a discussion similar to Lemma 4.2 (using (5.3) and (5.5)), the sequence $\{v_i\}_{i \in \mathbb{N}}$ is a Cauchy sequence and converges to $(\alpha, 1)$.

(iii) The existence of $\eta(l) = \lim_{t \to l} \eta(t)$ is an immediate consequence of the completeness of $\mathcal{P}(X)$ and the 1-Lipschitz continuity of η. Moreover, the lower semi-continuity of f yields that $f(\eta(l)) \le f(\eta(0)) - \int_0^l |\nabla_- f|(\eta(s))\, ds$. We suppose that there is $\varepsilon > 0$ such that

$$f\big(\eta(l)\big) \le f\big(\eta(0)\big) - \int_0^l |\nabla_- f|\big(\eta(s)\big)\, ds - \varepsilon.$$

For each $t \in [0,l)$, take a unit speed minimal geodesic $\alpha_t : [0, d_2^W(\eta(t), \eta(l))] \longrightarrow \mathcal{P}^*(X)$ from $\eta(t)$ to $\eta(l)$. Then we have, by (4.3),

$$\begin{aligned} |\nabla_- f|\big(\eta(t)\big) &\geq -D'_{\eta(t)} f(\alpha_t) \geq \frac{f(\eta(t)) - f(\eta(l))}{d_2^W(\eta(t), \eta(l))} + \frac{K}{2} d_2^W\big(\eta(t), \eta(l)\big) \\ &\geq \frac{1}{l-t}\left\{ \int_t^l |\nabla_- f|\big(\eta(s)\big)\, ds + \varepsilon \right\} + \frac{\min\{K,0\}}{2}(l-t) \\ &\geq \frac{\varepsilon}{l-t} + \frac{\min\{K,0\}}{2}(l-t). \end{aligned}$$

However, since $\int_0^l \varepsilon/(l-t)\, dt = \infty$, it implies $f(\eta(l)) = -\infty$, a contradiction. Therefore we obtain (5.4). □

Lemma 5.1 and the Ascoli-Arzelà theorem ensure that a gradient-like curve starts from every $\mu \in \mathcal{P}^*(X)$ with $|\nabla_- f|(\mu) > 0$.

Lemma 5.6 *For each $\mu \in \mathcal{P}^*(X)$ with $|\nabla_- f|(\mu) > 0$, we have a gradient-like curve $\eta : [0,l] \longrightarrow \mathcal{P}^*(X)$ of f with $\eta(0) = \mu$ for some $0 < l < \infty$.*

Proof. By the lower semi-continuity of $|\nabla_- f|$ (Lemma 4.3), we find $C_0, l > 0$ such that we have $|\nabla_- f|(\nu) \geq C_0$ for all $\nu \in \overline{B}(\mu, l) \cap \mathcal{P}^*(X)$ just as the assumption in Lemma 5.1. In particular, $U[\omega] \cap \overline{B}(\mu, l) = \emptyset$. Given $k \in \mathbb{N}$ with $k > l^{-1}$, by applying Lemma 5.1 repeatedly, we can choose a maximal sequence $\{\mu_i^k\}_{i=0}^{N(k)} \subset \overline{B}(\mu, l) \cap \mathcal{P}^*(X)$ with $\mu_0^k = \mu$ satisfying

$$\begin{aligned} c_{i-1}^k &:= k^{-1} \cdot \inf_{\overline{B}(\mu_{i-1}^k, k^{-1}) \cap \mathcal{P}^*(X)} |\nabla_- f|, \\ f(\mu_i^k) &= f(\mu_{i-1}^k) - c_{i-1}^k, \\ d_2^W(\mu_{i-1}^k, \mu_i^k) &= \operatorname{dist}\big(\mu_{i-1}^k, U[f(\mu_{i-1}^k) - c_{i-1}^k]\big) \leq k^{-1} \end{aligned}$$

for each $i = 1, 2, \ldots, N(k)$. We remark that the maximality of the sequence means that $d_2^W(\mu, \mu_{N(k)}^k) > l - k^{-1}$. We know $c_{i-1}^k \in [C_0 k^{-1}, f(\mu_{i-1}^k) - \omega]$ by Corollary 5.2. Set $a_0^k := 0$ and $a_i^k := \sum_{j=1}^i d_2^W(\mu_{j-1}^k, \mu_j^k)$ for $i = 1, 2, \ldots, N(k)$, and note that $a_{N(k)}^k \geq d_2^W(\mu, \mu_{N(k)}^k) \geq l - k^{-1}$. We define $A_k := \{a_0^k, a_1^k, \ldots, a_{N(k)}^k\} \cap [0, l]$ and the map $\eta_k : A_k \longrightarrow \overline{B}(\mu, l)$ by $\eta_k(a_i^k) := \mu_i^k$. Observe that every η_k is 1-Lipschitz and recall that $\overline{B}(\mu, l)$ is compact. Thus the Ascoli-Arzelà theorem yields that a subsequence of $\{\eta_k\}$ (again denoted by $\{\eta_k\}$) uniformly converges to a 1-Lipschitz curve $\eta : [0, l] \longrightarrow \overline{B}(\mu, l)$.

We shall show that η is a gradient-like curve of f. For $\varepsilon > 0$, let $k \in \mathbb{N}$ so large as to satisfy $d_2^W(\eta_k(a_i^k), \eta(a_i^k)) \leq \varepsilon$ for all $a_i^k \in A_k$. Now we fix $a \in (0, l)$ and take $i(k)$

satisfying $a^k_{i(k)-1} < a \le a^k_{i(k)}$. Then we have, by the construction of $\mu^k_{i(k)}$,

$$\begin{aligned}
f\big(\eta_k(a^k_{i(k)})\big) &= f(\mu^k_{i(k)}) = f(\mu) - k^{-1}\sum_{j=1}^{i(k)} \inf_{\overline{B}(\mu^k_{j-1},k^{-1})\cap\mathcal{P}^*(X)} |\nabla_- f| \\
&\le f(\mu) - \sum_{j=1}^{i(k)} \Big\{ (a^k_j - a^k_{j-1}) \cdot \inf_{\overline{B}(\eta(a^k_{j-1}),k^{-1}+\varepsilon)\cap\mathcal{P}^*(X)} |\nabla_- f| \Big\} \\
&\le f(\mu) - \int_0^{a^k_{i(k)}} \inf_{\overline{B}(\eta(t),2k^{-1}+\varepsilon)\cap\mathcal{P}^*(X)} |\nabla_- f| \, dt \\
&\le f(\mu) - \int_0^{a} \inf_{\overline{B}(\eta(t),2k^{-1}+\varepsilon)\cap\mathcal{P}^*(X)} |\nabla_- f| \, dt.
\end{aligned}$$

It follows from the monotone convergence theorem and the lower semi-continuity of $|\nabla_- f|$ that

$$\begin{aligned}
&\lim_{\varepsilon\to 0+}\lim_{k\to\infty} \int_0^a \inf_{\overline{B}(\eta(t),2k^{-1}+\varepsilon)\cap\mathcal{P}^*(X)} |\nabla_- f| \, dt \\
&= \int_0^a \lim_{\varepsilon\to 0+}\lim_{k\to\infty} \Big(\inf_{\overline{B}(\eta(t),2k^{-1}+\varepsilon)\cap\mathcal{P}^*(X)} |\nabla_- f| \Big) \, dt \ge \int_0^a |\nabla_- f|\big(\eta(t)\big)\, dt.
\end{aligned}$$

Hence we obtain

$$f\big(\eta(a)\big) \le \liminf_{k\to\infty} f\big(\eta_k(a^k_{i(k)})\big) \le \limsup_{k\to\infty} f\big(\eta_k(a^k_{i(k)})\big) \le f(\mu) - \int_0^a |\nabla_- f|\big(\eta(t)\big)\, dt.$$

Moreover, as $|\nabla_- f|$ is an upper gradient for f by Lemma 4.3, we find

$$f(\mu) - f\big(\eta(a)\big) \le \int_0^a |\nabla_- f|\big(\eta(t)\big)|\eta'|(t)\, dt \le \int_0^a |\nabla_- f|\big(\eta(t)\big)\, dt,$$

and hence

$$f\big(\eta(a)\big) = \lim_{k\to\infty} f\big(\eta_k(a^k_{i(k)})\big) = f(\mu) - \int_0^a |\nabla_- f|\big(\eta(t)\big)\, dt. \tag{5.6}$$

Thus we complete the proof. □

As a by-product of the proof of Lemma 5.6, we obtain the K-convexity of the function $f \circ \eta$. We first recall an easily proved lemma.

Lemma 5.7 *Let $h : [0,l] \longrightarrow \mathbb{R}$ be a continuous function. Given $a_0 = 0 < a_1 < a_2 < \cdots < a_{N-1} < a_N = l$, if $h|_{[a_{i-1},a_i]}$ is convex for all $i = 1, 2, \ldots, N$ and if*

$$\lim_{\varepsilon\to 0+} \frac{h(a_i) - h(a_i - \varepsilon)}{\varepsilon} \le \lim_{\varepsilon\to 0+} \frac{h(a_i+\varepsilon) - h(a_i)}{\varepsilon}$$

holds for all $i = 1, 2, \ldots, N-1$, then h is convex on $[0,l]$.

Proposition 5.8 *Let $\eta : [0,l] \longrightarrow \mathcal{P}^*(X)$ be the gradient-like curve constructed in Lemma* 5.6*. Then $f \circ \eta : [0,l] \longrightarrow \mathbb{R}$ is K-convex, i.e., the function $h(t) := f(\eta(t)) - Kt^2/2$ is convex.*

Proof. All notations are according to the proof of Lemma 5.6. Fix large $k \in \mathbb{N}$, put $I_k := [0, a^k_{N(k)}]$ and extend the function η_k to I_k by $\eta_k((1-\lambda)a^k_{i-1} + \lambda a^k_i) := \alpha^k_i(\lambda)$ for $\lambda \in [0,1]$, where $\alpha^k_i : [0,1] \longrightarrow \mathcal{P}^*(X)$ is an arbitrarily fixed minimal geodesic from μ^k_{i-1} to μ^k_i. Recall that $a^k_i - a^k_{i-1} = d^W_2(\mu^k_{i-1}, \mu^k_i)$ and hence η_k has a unit speed. Define the function $h_k : I_k \longrightarrow \mathbb{R}$ by $h_k(t) := f(\eta_k(t)) - Kt^2/2$. Then the K-convexity of f yields that $h_k|_{[a^k_{i-1}, a^k_i]}$ is convex for all $i = 1, 2, \ldots, N(k)$.

We shall show that

$$\lim_{\varepsilon \to 0+} \frac{h_k(a^k_i) - h_k(a^k_i - \varepsilon)}{\varepsilon} \le \lim_{\varepsilon \to 0+} \frac{h_k(a^k_i + \varepsilon) - h_k(a^k_i)}{\varepsilon} \tag{5.7}$$

holds for each $i = 1, 2, \ldots, N(k) - 1$. Assume the contrary, that is, we have

$$\lim_{\varepsilon \to 0+} \frac{f(\eta_k(a^k_i)) - f(\eta_k(a^k_i - \varepsilon))}{\varepsilon} \ge \lim_{\varepsilon \to 0+} \frac{f(\eta_k(a^k_i + \varepsilon)) - f(\eta_k(a^k_i))}{\varepsilon} + \delta \tag{5.8}$$

for some i and $\delta > 0$. For each small $\varepsilon > 0$, we choose a minimal geodesic $\beta_\varepsilon : [0,1] \longrightarrow \mathcal{P}^*(X)$ from $\eta_k(a^k_i - \varepsilon)$ to $\eta_k(a^k_i + \varepsilon)$. Note that

$$\begin{aligned} d^W_2\big(\mu^k_{i-1}, \beta_\varepsilon(1/2)\big) &\le d^W_2\big(\mu^k_{i-1}, \eta_k(a^k_i - \varepsilon)\big) + \frac{1}{2} d^W_2\big(\eta_k(a^k_i - \varepsilon), \eta_k(a^k_i + \varepsilon)\big) \\ &\le d^W_2(\mu^k_{i-1}, \mu^k_i) - \varepsilon + \varepsilon = d^W_2(\mu^k_{i-1}, \mu^k_i). \end{aligned}$$

It follows from the K-convexity of f that

$$f\big(\beta_\varepsilon(1/2)\big) \le \frac{1}{2} f\big(\eta_k(a^k_i - \varepsilon)\big) + \frac{1}{2} f\big(\eta_k(a^k_i + \varepsilon)\big) - \frac{K}{8} d^W_2\big(\eta_k(a^k_i - \varepsilon), \eta_k(a^k_i + \varepsilon)\big)^2.$$

However, combining this with our assumption (5.8) shows that

$$\begin{aligned} &\liminf_{\varepsilon \to 0+} \frac{f(\eta_k(a^k_i)) - f(\beta_\varepsilon(1/2))}{\varepsilon} \\ &\ge \frac{1}{2} \lim_{\varepsilon \to 0+} \frac{f(\eta_k(a^k_i)) - f(\eta_k(a^k_i - \varepsilon))}{\varepsilon} + \frac{1}{2} \lim_{\varepsilon \to 0+} \frac{f(\eta_k(a^k_i)) - f(\eta_k(a^k_i + \varepsilon))}{\varepsilon} \\ &\ge \delta > 0. \end{aligned}$$

This implies that $f(\beta_\varepsilon(1/2)) < f(\eta_k(a^k_i)) = f(\mu^k_i)$ holds for some small $\varepsilon > 0$, and it contradicts the choice of μ^k_i. Thus we have (5.7) and hence h_k is convex on I_k by Lemma 5.7. Recall that, as we saw in (5.6), $h(a) = \lim_{k \to \infty} h_k(a^k_{i(k)})$, where $i(k)$ is such that $a^k_{i(k)-1} < a \le a^k_{i(k)}$. Therefore, by taking the limit as k goes to infinity, we obtain the convexity of h. □

We verify several corollaries of the K-convexity of $f \circ \eta$ for later use. They enrich Lemma 5.5.

Corollary 5.9 *Let $\eta : [0,l] \longrightarrow \mathcal{P}^*(X)$ be the gradient-like curve constructed in Lemma* 5.6*. Then the following properties hold.*

(i) *The function $f \circ \eta : [0,l] \longrightarrow \mathbb{R}$ is locally Lipschitz on $(0,l)$ and $|\nabla_- f|(\eta(t)) < \infty$ holds for all $t \in (0,l)$.*

(ii) *For any $t \in [0,l)$ and $s \in (t,l)$, we have*

$$|\nabla_- f|\big(\eta(t)\big) \geq |\nabla_- f|\big(\eta(s)\big) + K(s-t).$$

(iii) *For any $t \in [0,l)$, it holds that*

$$|\nabla_- f|\big(\eta(t)\big) = \lim_{\varepsilon \to 0+} \frac{1}{\varepsilon} \int_t^{t+\varepsilon} |\nabla_- f|\big(\eta(s)\big)\, ds.$$

Proof. (i) This immediately follows from the K-convexity of $f \circ \eta$ together with (5.3).

(ii) Take $s \in (t,l)$ at where $f \circ \eta$ is differentiable. The K-convexity of $f \circ \eta$ shows that

$$\lim_{\varepsilon \to 0+} \frac{f(\eta(t+\varepsilon)) - f(\eta(t))}{\varepsilon} + \lim_{\varepsilon \to 0+} \frac{f(\eta(s-\varepsilon)) - f(\eta(s))}{\varepsilon} \leq -K(s-t),$$

and hence $-|\nabla_- f|(\eta(t)) + |\nabla_- f|(\eta(s)) \leq -K(s-t)$ by (5.3). General case follows from the lower semi-continuity of $|\nabla_- f|$.

(iii) If $|\nabla_- f|(\eta(t)) = \infty$, then $t = 0$ and we deduce from the lower semi-continuity of $|\nabla_- f|$ that $\lim_{\varepsilon \to 0+} \int_0^\varepsilon |\nabla_- f|(\eta(s))\, ds = \infty$. If $|\nabla_- f|(\eta(t)) < \infty$, then the lower semi-continuity of $|\nabla_- f|$ and (ii) imply that

$$\begin{aligned}
&\liminf_{\varepsilon \to 0+} \frac{1}{\varepsilon} \int_t^{t+\varepsilon} |\nabla_- f|\big(\eta(s)\big)\, ds \geq |\nabla_- f|\big(\eta(t)\big) \\
&\geq \limsup_{\varepsilon \to 0+} \frac{1}{\varepsilon} \int_t^{t+\varepsilon} \big\{ |\nabla_- f|\big(\eta(s)\big) + K(s-t) \big\}\, ds = \limsup_{\varepsilon \to 0+} \frac{1}{\varepsilon} \int_t^{t+\varepsilon} |\nabla_- f|\big(\eta(s)\big)\, ds.
\end{aligned}$$

□

5.2 Gradient curves

A gradient curve is defined by using the gradient vector (4.11) as follows:

Definition 5.10 (Gradient curves) A continuous curve $\xi : [0,l) \longrightarrow \mathcal{P}^*(X)$ which is locally Lipschitz on $(0,l)$ is called a *gradient curve* of f if we have $|\nabla_- f|(\xi(t)) < \infty$ for all $t \in (0,l)$ and if, at all $t \in [0,l)$ with $|\nabla_- f|(\xi(t)) < \infty$, ξ is right differentiable and

$$\xi'(t) = \nabla_- f\big(\xi(t)\big) \tag{5.9}$$

holds. If $l = \infty$, then we say that the gradient curve ξ is *complete*.

The existence of a complete gradient curve is a consequence of Lemma 5.6.

Theorem 5.11 (Existence and completeness) *Let* (X, d) *be a compact Alexandrov space of curvature bounded below and* f *be a function satisfying* (4.1)*. Then, for every* $\mu \in \mathcal{P}^*(X)$*, there exists a complete gradient curve* $\xi : [0, \infty) \longrightarrow \mathcal{P}^*(X)$ *of* f *with* $\xi(0) = \mu$*.*

Proof. If $|\nabla_- f|(\mu) = 0$, then the constant curve $\xi(t) \equiv \mu$ gives a gradient curve. If $|\nabla_- f|(\mu) > 0$, then, by applying Lemma 5.6 repeatedly, we have a maximal gradient-like curve $\eta : [0, l) \longrightarrow \mathcal{P}^*(X)$ with $\eta(0) = \mu$ and $l > 0$. We remark that $|\nabla_- f|(\eta(t)) > 0$ for all $t \in [0, l)$ and that, if $l < \infty$, then Lemma 5.5(iii) shows that $|\nabla_- f|(\eta(l)) = 0$ (otherwise, we can apply Lemma 5.6 again at $\eta(l)$ and it contradicts the maximality of η). Note also that $f \circ \eta : [0, l) \longrightarrow \mathbb{R}$ is K-convex by Proposition 5.8, and hence $|\nabla_- f|(\eta(t)) < \infty$ for all $t \in (0, l)$ as in Corollary 5.9(i).

As a famous theorem of Peano, we find a solution $\psi : [0, L) \longrightarrow [0, l')$ of the equation

$$\psi(t) = \int_0^t |\nabla_- f| \circ \eta\big(\psi(s)\big)\, ds,$$

where $l' \in (0, l]$ and $l' = l$ if $L < \infty$. (The proof of the existence of ψ is given at the end of this section for completeness (Lemma 5.13).) The function ψ is monotone increasing and bijective, and hence homeomorphic. In addition, ψ is locally Lipschitz on $(0, L)$ by Corollary 5.9(ii). Now we show the analogue of Corollary 5.9(iii). It follows from the lower semi-continuity of $|\nabla_- f|$ and Corollary 5.9(ii) that, for every $t \in [0, L)$ with $|\nabla_- f| \circ \eta(\psi(t)) < \infty$,

$$\begin{aligned}
&\liminf_{\varepsilon \to 0+} \frac{\psi(t+\varepsilon) - \psi(t)}{\varepsilon} = \liminf_{\varepsilon \to 0+} \frac{1}{\varepsilon} \int_t^{t+\varepsilon} |\nabla_- f| \circ \eta\big(\psi(s)\big)\, ds \\
&\geq |\nabla_- f| \circ \eta\big(\psi(t)\big) \geq \limsup_{\varepsilon \to 0+} \frac{1}{\varepsilon} \int_t^{t+\varepsilon} \big\{ |\nabla_- f| \circ \eta\big(\psi(s)\big) + K\big(\psi(s) - \psi(t)\big) \big\}\, ds \\
&= \limsup_{\varepsilon \to 0+} \frac{1}{\varepsilon} \int_t^{t+\varepsilon} |\nabla_- f| \circ \eta\big(\psi(s)\big)\, ds = \limsup_{\varepsilon \to 0+} \frac{\psi(t+\varepsilon) - \psi(t)}{\varepsilon},
\end{aligned}$$

namely

$$\lim_{\varepsilon \to 0+} \frac{\psi(t+\varepsilon) - \psi(t)}{\varepsilon} = |\nabla_- f| \circ \eta\big(\psi(t)\big).$$

Define $\xi(t) := \eta(\psi(t))$ and observe that it is a gradient curve by construction. If $L = \infty$, then ξ is complete. If $L < \infty$, then $l' = l$ and we have

$$\begin{aligned}
l &= \mathrm{length}(\eta) = \mathrm{length}(\xi) = \int_0^L |\nabla_- f|\big(\xi(t)\big)\, dt \\
&\leq L^{1/2} \bigg(\int_0^L |\nabla_- f|\big(\xi(t)\big)^2\, dt \bigg)^{1/2} = L^{1/2} \big\{ f\big(\xi(0)\big) - f\big(\xi(L)\big) \big\}^{1/2}.
\end{aligned}$$

Here the last equality follows from the local Lipschitz continuity of $f \circ \xi = (f \circ \eta) \circ \psi$ on $(0, L)$ and, for a.e. $t \in (0, L)$,

$$(f \circ \xi)'(t) = [(f \circ \eta) \circ \psi]'(t) = (f \circ \eta)'\big(\psi(t)\big) \cdot \psi'(t)$$
$$= -|\nabla_- f|\big(\eta(\psi(t))\big)^2 = -|\nabla_- f|\big(\xi(t)\big)^2. \tag{5.10}$$

Thus we have $l < \infty$ as well as $|\nabla_- f|(\xi(L)) = |\nabla_- f|(\eta(l)) = 0$. Hence we extend ξ by $\xi(t) := \xi(L)$ for $t \in [L, \infty)$ and it is a complete gradient curve. □

The following estimates are straightforward consequences of our construction.

Proposition 5.12 *Let $\xi : [0, \infty) \longrightarrow \mathcal{P}^*(X)$ be the gradient curve of f constructed in Theorem* 5.11*.*

(i) *For any $t > 0$, it holds that*

$$f\big(\xi(t)\big) = f\big(\xi(0)\big) - \int_0^t |\nabla_- f|\big(\xi(s)\big)^2 \, ds.$$

(ii) *For any $t > s \geq 0$, we have*

$$d_2^W\big(\xi(s), \xi(t)\big)^2 \leq (t - s)\{f\big(\xi(s)\big) - f\big(\xi(t)\big)\}.$$

Proof. We have already observed (i) in the proof of Theorem 5.11 (see (5.10)). Moreover, (ii) follows from

$$d_2^W\big(\xi(s), \xi(t)\big)^2 \leq \left(\int_s^t |\nabla_- f|\big(\xi(\tau)\big)\, d\tau\right)^2 \leq (t - s) \int_s^t |\nabla_- f|\big(\xi(\tau)\big)^2 \, d\tau$$
$$= (t - s)\{f\big(\xi(s)\big) - f\big(\xi(t)\big)\}.$$

□

Lemma 5.13 *Let a function $g : [0, l) \longrightarrow \mathbb{R}$ be lower semi-continuous, strictly positive on $[0, l)$ and locally bounded on $(0, l)$. Then there exists a continuous function $\psi : [0, L) \longrightarrow [0, l')$ satisfying $\psi(t) = \int_0^t g(\psi(s))\, ds$ for all $t \in [0, L)$, where $l' \in (0, l]$ and $l' = l$ if $L < \infty$.*

Proof. Since g is lower semi-continuous and strictly positive, it admits a positive lower bound near 0. Therefore we find the maximal $l' \in (0, l]$ such that the function $\phi(s) := \int_0^s (1/g)\, d\tau$ is well-defined on $[0, l')$. In other words, l' is the maximal number satisfying $\int_0^s (1/g)\, d\tau < \infty$ for all $s \in [0, l')$. Note that ϕ is continuous and monotone increasing, and that $\phi' = 1/g$ at Lebesgue points of $1/g$. Now we define $L := \int_0^{l'} (1/g)\, d\tau$ and $\psi := \phi^{-1} : [0, L) \longrightarrow [0, l')$. The local boundedness of g shows that ψ is locally Lipschitz on $(0, L)$. Moreover, we have $\psi'(\phi(\tau)) = 1/\phi'(\tau) = g(\tau)$ for any Lebesgue point $\tau \in (0, l')$ of $1/g$, and hence $\psi'(t) = g(\psi(t))$ for a.e. $t \in [0, L)$. As ψ is continuous on $[0, L)$ and locally Lipschitz on $(0, L)$, we complete the proof. □

6 Nonnegatively curved case

In this section, we study the special case where the underlying space is nonnegatively curved. This additional condition is imposed only for technical reasons, so that it is expected that all results hold for general compact Riemannian manifolds or Alexandrov spaces.

6.1 Strong differentiability and its consequences

In this subsection, let (X,d) be a compact Alexandrov space of nonnegative curvature, and let $f:\mathcal{P}(X)\longrightarrow(-\infty,\infty]$ be a function satisfying the condition (4.1). Then the Wasserstein space $(\mathcal{P}(X),d_2^W)$ is also an Alexandrov space of nonnegative curvature. If we take a curve $\xi:[0,\delta)\longrightarrow\mathcal{P}(X)$ which is right differentiable at 0, then it follows from (2.4) that ξ is *strongly right differentiable*, that is, for any $(\alpha,s)\in C'_{\xi(0)}[\mathcal{P}(X)]$, we have

$$\lim_{\varepsilon\to 0+} d_2^W\big(\xi(\varepsilon),\alpha(s\varepsilon)\big)/\varepsilon=\sigma_{\xi(0)}\big(\xi'(0),(\alpha,s)\big). \tag{6.1}$$

(This is essentially the sole reason why we need the nonnegative curvature.) It implies a kind of first variation formula (see [Ly1, Lemma 9.1]).

Lemma 6.1 *Given distinct $\mu,\nu\in\mathcal{P}(X)$, let $\alpha:[0,l]\longrightarrow\mathcal{P}(X)$ be a unit speed minimal geodesic from μ to ν and $\alpha^-:[0,l]\longrightarrow\mathcal{P}(X)$ be its converse, i.e., $\alpha^-(t):=\alpha(l-t)$. Then, for any curves $\xi,\zeta:[0,\delta]\longrightarrow\mathcal{P}(X)$ right differentiable at 0 with $\xi(0)=\mu$ and $\zeta(0)=\nu$, we have*

$$\limsup_{\varepsilon\to 0+}\frac{h(\varepsilon)-h(0)}{\varepsilon}\le-\langle\xi'(0),(\alpha,1)\rangle_\mu-\langle\zeta'(0),(\alpha^-,1)\rangle_\nu,$$

where we set $h(t):=d_2^W(\xi(t),\zeta(t))$.

Proof. Take a sequence $\{\varepsilon_i\}_{i\in\mathbb{N}}\subset(0,\delta)$ tending to zero. For each $i\in\mathbb{N}$, set $a_i:=d_2^W(\mu,\xi(\varepsilon_i))$ and $b_i:=d_2^W(\nu,\zeta(\varepsilon_i))$ and choose minimal geodesics $\beta_i:[0,a_i]\longrightarrow\mathcal{P}(X)$ from μ to $\xi(\varepsilon_i)$ and $\gamma_i:[0,b_i]\longrightarrow\mathcal{P}(X)$ from ν to $\zeta(\varepsilon_i)$. We put $v_i:=(\beta_i,a_i/\varepsilon_i)\in C'_\mu[\mathcal{P}(X)]$ and $w_i:=(\gamma_i,b_i/\varepsilon_i)\in C'_\nu[\mathcal{P}(X)]$ and observe that

$$\begin{aligned}&\left|\limsup_{\varepsilon\to 0+}\frac{h(\varepsilon)-h(0)}{\varepsilon}-\limsup_{\varepsilon\to 0+}\frac{d_2^W(\beta_i(a_i\varepsilon),\gamma_i(b_i\varepsilon))-d_2^W(\mu,\nu)}{\varepsilon_i\varepsilon}\right|\\&\le\limsup_{\varepsilon\to 0+}\frac{d_2^W(\xi(\varepsilon),\beta_i(a_i\varepsilon/\varepsilon_i))+d_2^W(\zeta(\varepsilon),\gamma_i(b_i\varepsilon/\varepsilon_i))}{\varepsilon}\\&=\sigma_\mu(\xi'(0),v_i)+\sigma_\nu(\zeta'(0),w_i)\to 0\end{aligned}$$

as i diverges to infinity. Similarly, we find

$$\lim_{i\to\infty}\langle v_i,(\alpha,1)\rangle_\mu=\langle\xi'(0),(\alpha,1)\rangle_\mu,\quad\lim_{i\to\infty}\langle w_i,(\alpha^-,1)\rangle_\nu=\langle\zeta'(0),(\alpha^-,1)\rangle_\nu.$$

Thus it suffices to see

$$\limsup_{\varepsilon\to 0+} \frac{d_2^W(\beta_i(a_i\varepsilon),\gamma_i(b_i\varepsilon)) - d_2^W(\mu,\nu)}{\varepsilon_i\varepsilon} \le -\langle v_i,(\alpha,1)\rangle_\mu - \langle w_i,(\alpha^-,1)\rangle_\nu.$$

Now we consider the product space $Y := C'_\mu[\mathcal{P}(X)] \times C'_\nu[\mathcal{P}(X)]$ and the function $D : Y \longrightarrow [0,\infty)$ defined by, for $(v,w) = ((\beta,s),(\gamma,t)) \in Y$,

$$D(v,w) := \limsup_{\varepsilon\to 0+} \frac{d_2^W(\beta(s\varepsilon),\gamma(t\varepsilon)) - d_2^W(\mu,\nu)}{\varepsilon}.$$

The function D can be regarded as a differential of the distance function d_2^W at (μ,ν). The triangle inequality yields that $|D(v,w)| \le s+t$ and

$$|D(v,w) - D(v',w')| \le \sigma_\mu(v,v') + \sigma_\nu(w,w').$$

Thus we have, for any $i \in \mathbb{N}$ and $t \ge 0$,

$$\begin{aligned} D(v_i,w_i) &\le D\big((\alpha,t),(\alpha^-,t)\big) + \sigma_\mu\big(v_i,(\alpha,t)\big) + \sigma_\nu\big(w_i,(\alpha^-,t)\big) \\ &= -2t + \{(a_i/\varepsilon_i)^2 + t^2 - 2\langle v_i,(\alpha,t)\rangle_\mu\}^{1/2} + \{(b_i/\varepsilon_i)^2 + t^2 - 2\langle w_i,(\alpha^-,t)\rangle_\nu\}^{1/2}. \end{aligned}$$

Note that

$$\begin{aligned} &\lim_{t\to\infty} \big[\{(a_i/\varepsilon_i)^2 + t^2 - 2\langle v_i,(\alpha,t)\rangle_\mu\}^{1/2} - t\big] \\ &= \lim_{t\to\infty} \frac{(a_i/\varepsilon_i)^2 - 2t\langle v_i,(\alpha,1)\rangle_\mu}{\{(a_i/\varepsilon_i)^2 + t^2 - 2t\langle v_i,(\alpha,1)\rangle_\mu\}^{1/2} + t} = -\langle v_i,(\alpha,1)\rangle_\mu. \end{aligned}$$

Similarly, we deduce that

$$\lim_{t\to\infty} \big[\{(b_i/\varepsilon_i)^2 + t^2 - 2\langle w_i,(\alpha^-,t)\rangle_\nu\}^{1/2} - t\big] = -\langle w_i,(\alpha^-,1)\rangle_\nu.$$

Hence we obtain

$$\begin{aligned} \limsup_{\varepsilon\to 0+} \frac{d_2^W(\beta_i(a_i\varepsilon),\gamma_i(b_i\varepsilon)) - d_2^W(\mu,\nu)}{\varepsilon_i\varepsilon} &= D(v_i,w_i) \\ &\le -\langle v_i,(\alpha,1)\rangle_\mu - \langle w_i,(\alpha^-,1)\rangle_\nu. \end{aligned}$$

□

Combining this with Lemma 4.2 and arguing as in [Ly2, Section 9.2], we obtain the following:

Theorem 6.2 (Uniqueness and contraction) *Let (X,d) be a compact Alexandrov space of nonnegative curvature, f be a function satisfying the condition (4.1) and let $\xi,\zeta : [0,\infty) \longrightarrow \mathcal{P}^*(X)$ be gradient curves of f. Then, for any $t \in [0,\infty)$, we have*

$$d_2^W\big(\xi(t),\zeta(t)\big) \le e^{-Kt} d_2^W\big(\xi(0),\zeta(0)\big).$$

In particular, for each $\mu \in \mathcal{P}^*(X)$, there exists a unique complete gradient curve of f starting from μ.

Proof. Put $h(t) := d_2^W(\xi(t), \zeta(t))$ and fix $t \in (0, \infty)$. If $h(t) > 0$, let $\alpha : [0, h(t)] \longrightarrow \mathcal{P}^*(X)$ be a minimal geodesic from $\xi(t)$ to $\zeta(t)$, and $\beta : [0, h(t)] \longrightarrow \mathcal{P}^*(X)$ be its converse, that is, $\beta(s) = \alpha(h(t) - s)$. It follows from Lemmas 6.1, 4.2 and (5.9) that

$$\begin{aligned}
&\limsup_{\varepsilon \to 0+} \frac{h(t+\varepsilon) - h(t)}{\varepsilon} \\
&\le -\langle \nabla_- f\big(\xi(t)\big), (\alpha, 1)\rangle_{\xi(t)} - \langle \nabla_- f\big(\zeta(t)\big), (\beta, 1)\rangle_{\zeta(t)} \\
&\le D_{\xi(t)} f(\alpha) + D_{\zeta(t)} f(\beta) \\
&\le \lim_{\varepsilon \to 0+} \frac{f(\alpha(\varepsilon)) - f(\alpha(0))}{\varepsilon} + \lim_{\varepsilon \to 0+} \frac{f(\beta(\varepsilon)) - f(\beta(0))}{\varepsilon}.
\end{aligned}$$

Furthermore, by the K-convexity of f along the geodesic α, we find

$$\lim_{\varepsilon \to 0+} \frac{f(\alpha(\varepsilon)) - f(\alpha(0))}{\varepsilon} + \lim_{\varepsilon \to 0+} \frac{f(\beta(\varepsilon)) - f(\beta(0))}{\varepsilon} \le -Kh(t).$$

If $h(t) = 0$, then we immediately observe from (6.1) that

$$\lim_{\varepsilon \to 0+} \frac{h(t+\varepsilon)}{\varepsilon} = \sigma_{\xi(t)}\big(\nabla_- f(\xi(t)), \nabla_- f(\xi(t))\big) = 0.$$

Thus we have $h'(t) \le -Kh(t)$ for a.e. $t \in (0, \infty)$ and hence $h(t) \le e^{-Kt} h(0)$ by Gronwall's inequality. It completes the proof. $\square$

We define the *gradient flow* $G : \mathcal{P}^*(X) \times [0, \infty) \longrightarrow \mathcal{P}^*(X)$ of f by $G(\mu, t) := \xi(t)$, where $\xi : [0, \infty) \longrightarrow \mathcal{P}^*(X)$ is the unique gradient curve starting from μ. The contraction property allows us to extend the gradient flow G to the closure $\mathcal{P}^*(X)^-$ of $\mathcal{P}^*(X)$.

Corollary 6.3 *The gradient flow $G : \mathcal{P}^*(X) \times [0, \infty) \longrightarrow \mathcal{P}^*(X)$ extends uniquely and continuously to $G : \mathcal{P}^*(X)^- \times [0, \infty) \longrightarrow \mathcal{P}^*(X)^-$ and, for any $\mu, \nu \in \mathcal{P}^*(X)^-$ and $t \in [0, \infty)$, we have*

$$d_2^W\big(G(\mu, t), G(\nu, t)\big) \le e^{-Kt} d_2^W(\mu, \nu).$$

Clearly G satisfies the semigroup property: $G(\mu, s + t) = G(G(\mu, s), t)$ for $\mu \in \mathcal{P}^*(X)^-$ and $s, t \ge 0$.

We next observe the compatibility between our gradient flow and a gradient flow considered in [JKO] and [AGS]. It holds for general Alexandrov spaces, and we further obtain a useful error estimate in our nonnegatively curved situation. Fix $\mu \in \mathcal{P}^*(X)$ and, for each $\delta > 0$, take $\nu_\delta \in \mathcal{P}^*(X)$ attaining the infimum of

$$\mathcal{P}^*(X) \ni \nu \longmapsto f(\nu) + \frac{d_2^W(\mu, \nu)^2}{2\delta}.$$

We remark that such a point ν_δ indeed exists by the compactness of $\mathcal{P}(X)$ and the lower semi-continuity of f. Moreover, we find that

$$d_2^W(\mu, \nu_\delta)^2 \le 2\delta\{f(\mu) - f(\nu_\delta)\} \le 2\delta\{f(\mu) - \omega\}.$$

We choose a minimal geodesic $\beta_\delta : [0, t_\delta] \longrightarrow \mathcal{P}^*(X)$ from μ to ν_δ and put $v_\delta := (\beta_\delta, t_\delta/\delta) \in C'_\mu[\mathcal{P}^*(X)]$, where we set $t_\delta := d_2^W(\mu, \nu_\delta)$.

Lemma 6.4 *If $|\nabla_- f|(\mu) < \infty$, then the sequence $\{v_\delta\}_{\delta>0} \subset C'_\mu[\mathcal{P}^*(X)]$ converges to $\nabla_- f(\mu)$ as δ tends to zero. In particular, we have* $\lim_{\delta\to 0+} d_2^W(G(\mu,\delta), \nu_\delta)/\delta = 0$.

Proof. We first consider the case of $|\nabla_- f|(\mu) = 0$. If $\nu_\delta = \mu$ for some $\delta > 0$, then there is nothing to prove. Otherwise, by the choice of ν_δ, we see

$$f(\mu) \geq f(\nu_\delta) + \frac{d_2^W(\mu,\nu_\delta)^2}{2\delta} = f(\nu_\delta) + \frac{t_\delta \cdot d_2^W(\mu,\nu_\delta)}{2\delta}.$$

This implies

$$\limsup_{\delta\to 0+} \frac{t_\delta}{2\delta} \leq \limsup_{\delta\to 0+} \frac{f(\mu) - f(\nu_\delta)}{d_2^W(\mu,\nu_\delta)} \leq |\nabla_- f|(\mu) = 0.$$

Hence $\{v_\delta\}_{\delta>0}$ converges to $o_\mu \in C_\mu[\mathcal{P}^*(X)]$.

Next we suppose $|\nabla_- f|(\mu) \in (0,\infty)$ and put $\nabla_- f(\mu) = (\alpha, t)$. Note that t_δ is positive in this case. Take a sequence $\{\alpha_i\}_{i\in\mathbb{N}} \subset \Sigma'_\mu[\mathcal{P}^*(X)]$ such that $\lim_{i\to\infty} D'_\mu f(\alpha_i) = -|\nabla_- f|(\mu)$. Then we deduce from the choice of ν_δ that, for each $i \in \mathbb{N}$,

$$\liminf_{\delta\to 0+} \frac{f(\mu) - f(\nu_\delta)}{t_\delta} \geq \liminf_{\delta\to 0+} \frac{f(\mu) - f(\alpha_i(t_\delta))}{t_\delta} = -D'_\mu f(\alpha_i).$$

By letting i diverge to infinity, it yields $\liminf_{\delta\to 0+}\{f(\mu) - f(\nu_\delta)\}/t_\delta \geq |\nabla_- f|(\mu)$. Thus the same discussion as Lemma 4.2 shows that $\{\beta_\delta\}_{\delta>0}$ is a Cauchy sequence and converges to α as δ goes to zero.

It remains to show $\lim_{\delta\to 0+} t_\delta/\delta = t$. On one hand, again by the choice of ν_δ, we find

$$f(\nu_\delta) + \frac{t_\delta^2}{2\delta} \leq f\big(\alpha_i(\delta t)\big) + \frac{(\delta t)^2}{2\delta} = f\big(\alpha_i(\delta t)\big) + \frac{t^2}{2}\delta$$

for fixed $i \in \mathbb{N}$ and sufficiently small $\delta > 0$. On the other hand, it follows from (4.3) that

$$-t = -|\nabla_- f|(\mu) \leq \frac{f(\nu_\delta) - f(\mu)}{t_\delta} - \frac{K}{2} t_\delta.$$

By combining these, it holds that

$$\begin{aligned}
\frac{t_\delta^2}{2\delta^2} &\leq \frac{f(\alpha_i(\delta t)) - f(\nu_\delta)}{\delta} + \frac{t^2}{2} \\
&= \frac{f(\alpha_i(\delta t)) - f(\mu)}{\delta} + \frac{t_\delta}{\delta} \cdot \frac{f(\mu) - f(\nu_\delta)}{t_\delta} + \frac{t^2}{2} \\
&\leq \frac{f(\alpha_i(\delta t)) - f(\mu)}{\delta} + \frac{t_\delta}{\delta}\left(t - \frac{K}{2} t_\delta\right) + \frac{t^2}{2}.
\end{aligned}$$

Thus we have

$$\left(\limsup_{\delta\to 0+}\frac{t_\delta}{\delta}\right)^2 \le \lim_{i\to\infty}\left\{2t\cdot D'_\mu f(\alpha_i)+2t\cdot\limsup_{\delta\to 0+}\frac{t_\delta}{\delta}+t^2\right\} = 2t\cdot\limsup_{\delta\to 0+}\frac{t_\delta}{\delta}-t^2.$$

Therefore we obtain $\limsup_{\delta\to 0+} t_\delta/\delta = t$ and also $\liminf_{\delta\to 0+} t_\delta/\delta = t$ similarly. These imply $\lim_{\delta\to 0+} t_\delta/\delta = t$, and hence $\{v_\delta\}_{\delta>0}$ converges to $\nabla_- f(\mu)$. The second assertion follows from (2.4). $\square$

6.2 The free energy and the Fokker-Planck equation

In this final subsection, we study the Riemannian case and observe that our gradient flow of the free energy produces a solution of the linear Fokker-Planck equation, as was shown by Jordan, Kinderlehrer and Otto [JKO] in the Euclidean setting. See also [V2] for related work. Throughout this subsection, let (M,g) be a compact Riemannian manifold of nonnegative sectional curvature and m be the associated volume element.

We are concerned with the *linear Fokker-Planck equation* in the following form:

$$\frac{\partial\rho_t}{\partial t} = \Delta\rho_t + \operatorname{div}(\rho_t\cdot\operatorname{grad}V), \tag{6.2}$$

where $\Delta = \operatorname{div}\circ\operatorname{grad}$ is the Laplacian and the potential $V\in C^\infty(M)$ is a smooth function on M. The associated free energy $f:\mathcal{P}(M)\longrightarrow(-\infty,\infty]$ is defined by $f(\mu) := \operatorname{Ent}_m(\mu)+\int_M V\,d\mu$ (see (2.8)). Then Lemma 2.7 and Theorem 2.9 show that f satisfies (4.1) for some $K\in\mathbb{R}$, and $\mathcal{P}^*(M)^- = \mathcal{P}(M)$. Thus the gradient flow $G:\mathcal{P}(M)\times[0,\infty)\longrightarrow\mathcal{P}(M)$ is defined on entire $\mathcal{P}(M)$ (see Corollary 6.3). In the particular case $V\equiv 0$, (6.2) corresponds to the *heat equation*:

$$\frac{\partial\rho_t}{\partial t} = \Delta\rho_t, \tag{6.3}$$

and the free energy f is nothing but the relative entropy Ent_m.

Along discussions in [JKO, Theorem 5.1] and [V1, Subsection 8.4.2], we prove that the gradient flow of the free energy is a solution of the linear Fokker-Planck equation. Before it, we recall the Kantorovich-Rubinstein theorem which will be used in the proof.

Theorem 6.5 (cf. [V1, Theorem 1.14]) *For any $\mu,\nu\in\mathcal{P}(M)$, we have*

$$d_1^W(\mu,\nu) = \sup_g\left\{\int_M g\,d\mu - \int_M g\,d\nu\right\},$$

where the supremum is taken over all 1*-Lipschitz functions $g: M\longrightarrow\mathbb{R}$.*

Here d_1^W stands for the L_1-*Wasserstein distance* defined by, just as d_2^W (see Definition 2.4),

$$d_1^W(\mu,\nu) := \inf_q\left\{\int_{M\times M} d_M(x,y)\,dq(x,y)\right\},$$

where $q\in\mathcal{P}(M\times M)$ runs over all couplings of μ and ν. Note that $d_1^W(\mu,\nu)\le d_2^W(\mu,\nu)\le\{\operatorname{diam}M\cdot d_1^W(\mu,\nu)\}^{1/2}$.

Theorem 6.6 *Let (M,g) be a compact Riemannian manifold of nonnegative sectional curvature equipped with the Riemannian volume element m. Fix $V \in C^\infty(M)$ and let f be the associated free energy (2.8). Then, if a curve $\xi : [0,\infty) \longrightarrow \mathcal{P}^*(M)$ is a gradient curve of f, then its density function ρ (i.e., $\xi = \rho \cdot m$) is a weak solution of the linear Fokker-Planck equation (6.2). More precisely, for any $0 \le t_0 < t_1 < \infty$ and smooth function $h \in C^\infty(M \times \mathbb{R})$, we have*

$$\begin{aligned}&\int_M h_{t_1}\,d\mu_{t_1} - \int_M h_{t_0}\,d\mu_{t_0}\\ &= \int_{t_0}^{t_1}\int_M \left\{\frac{\partial h_t}{\partial t} + \Delta h_t - \langle \operatorname{grad} h_t, \operatorname{grad} V\rangle\right\} d\mu_t\,dt,\end{aligned} \tag{6.4}$$

where we set $\mu_t := \xi(t)$ and $h_t := h(\cdot,t)$.

Proof. Fix $t \in (t_0,t_1)$ and, for small $\delta > 0$, take $\nu_\delta^t \in \mathcal{P}^*(M)$ satisfying

$$f(\nu_\delta^t) + \frac{d_2^W(\mu_t,\nu_\delta^t)^2}{2\delta} = \inf_{\nu\in\mathcal{P}^*(M)}\left\{f(\nu) + \frac{d_2^W(\mu_t,\nu)^2}{2\delta}\right\}.$$

By McCann's theorem (Theorem 2.6), we have a Lipschitz function $\psi_\delta : M \longrightarrow \mathbb{R}$ such that a map $\Psi_\delta : M \longrightarrow M$ defined by $\Psi_\delta(x) := \exp_x[\operatorname{grad}\psi_\delta(x)]$ satisfies $(\Psi_\delta)_*\nu_\delta^t = \mu_t$ and

$$d_2^W(\nu_\delta^t,\mu_t)^2 = \int_M d_M\big(x,\Psi_\delta(x)\big)^2\,d\nu_\delta^t(x).$$

For small $\varepsilon > 0$, define the smooth map $\Phi_\varepsilon : M \longrightarrow M$ by $\Phi_\varepsilon(x) := \exp_x[\varepsilon \operatorname{grad} h_t(x)]$ and note that Φ_ε is a diffeomorphism if ε is sufficiently small. Put $\tilde\nu_{\delta,\varepsilon}^t := (\Phi_\varepsilon)_*\nu_\delta^t$. Then the choice of ν_δ^t implies that

$$f(\tilde\nu_{\delta,\varepsilon}^t) + \frac{d_2^W(\mu_t,\tilde\nu_{\delta,\varepsilon}^t)^2}{2\delta} - f(\nu_\delta^t) - \frac{d_2^W(\mu_t,\nu_\delta^t)^2}{2\delta} \ge 0. \tag{6.5}$$

We observe that, by the first variation formula,

$$\begin{aligned}&\limsup_{\varepsilon\to 0+}\frac{d_2^W(\mu_t,\tilde\nu_{\delta,\varepsilon}^t)^2 - d_2^W(\mu_t,\nu_\delta^t)^2}{\varepsilon}\\ &\le \limsup_{\varepsilon\to 0+}\frac{1}{\varepsilon}\int_M \left\{d_M\big(\Psi_\delta(x),\Phi_\varepsilon(x)\big)^2 - d_M\big(\Psi_\delta(x),x\big)^2\right\} d\nu_\delta^t(x)\\ &= -\int_M 2\langle \operatorname{grad} h_t(x), \operatorname{grad}\psi_\delta(x)\rangle\,d\nu_\delta^t(x).\end{aligned}$$

By the expansion and the compactness of M, there is a constant $C \ge 0$ depending on h_t such that

$$h_t\big(\Psi_\delta(x)\big) \le h_t(x) + \langle \operatorname{grad} h_t(x), \operatorname{grad}\psi_\delta(x)\rangle + C d_M\big(x,\Psi_\delta(x)\big)^2.$$

Therefore we obtain

$$\begin{aligned}
&\liminf_{\delta\to0+}\frac{1}{2\delta}\limsup_{\varepsilon\to0+}\frac{d_2^W(\mu_t,\tilde{\nu}^t_{\delta,\varepsilon})^2-d_2^W(\mu_t,\nu^t_\delta)^2}{\varepsilon}\\
&\le -\limsup_{\delta\to0+}\frac{1}{\delta}\int_M\langle\operatorname{grad}h_t(x),\operatorname{grad}\psi_\delta(x)\rangle\,d\nu^t_\delta(x)\\
&\le \liminf_{\delta\to0+}\frac{1}{\delta}\biggl[\int_M\bigl\{h_t(x)-h_t\bigl(\Psi_\delta(x)\bigr)\bigr\}\,d\nu^t_\delta(x)+C\int_M d_M\bigl(x,\Psi_\delta(x)\bigr)^2\,d\nu^t_\delta(x)\biggr]\\
&=\liminf_{\delta\to0+}\frac{1}{\delta}\biggl\{\int_M h_t\,d\nu^t_\delta-\int_M h_t\,d\mu_t\biggr\}.
\end{aligned}\tag{6.6}$$

In the last equality, we used the fact

$$\lim_{\delta\to0+}\frac{1}{\delta}\int_M d_M\bigl(x,\Psi_\delta(x)\bigr)^2\,d\nu^t_\delta(x)=\lim_{\delta\to0+}\frac{1}{\delta}d_2^W(\nu^t_\delta,\mu_t)^2\le 2\lim_{\delta\to0+}\{f(\mu_t)-f(\nu^t_\delta)\}=0.$$

Now we estimate the difference of entropies. Put $\nu^t_\delta=\varsigma_\delta\cdot m$ and $\tilde{\nu}^t_{\delta,\varepsilon}=\tilde{\varsigma}_{\delta,\varepsilon}\cdot m$ for simplicity. For arbitrary $\varphi\in C^\infty(M)$, by the definition of $\tilde{\nu}^t_{\delta,\varepsilon}$, it holds that

$$\int_M\varphi\,d\tilde{\nu}^t_{\delta,\varepsilon}=\int_M\varphi\bigl(\Phi_\varepsilon(x)\bigr)\,d\nu^t_\delta(x)=\int_M\varphi\bigl(\Phi_\varepsilon(x)\bigr)\varsigma_\delta(x)\,dm(x).$$

On the other hand, the change of variables formula (for $y=\Phi_\varepsilon(x)$) yields that

$$\begin{aligned}
\int_M\varphi\,d\tilde{\nu}^t_{\delta,\varepsilon}&=\int_M\varphi(y)\tilde{\varsigma}_{\delta,\varepsilon}(y)\,dm(y)\\
&=\int_M\varphi\bigl(\Phi_\varepsilon(x)\bigr)\tilde{\varsigma}_{\delta,\varepsilon}\bigl(\Phi_\varepsilon(x)\bigr)\det[D\Phi_\varepsilon(x)]\,dm(x).
\end{aligned}$$

Since $\varphi\in C^\infty(M)$ is arbitrary, these together imply that $\tilde{\varsigma}_{\delta,\varepsilon}(\Phi_\varepsilon(x))\det[D\Phi_\varepsilon(x)]=\varsigma_\delta(x)$ holds for a.e. $x\in M$. Combining this with the change of variables formula, we have

$$\begin{aligned}
\operatorname{Ent}_m(\tilde{\nu}^t_{\delta,\varepsilon})&=\int_M\tilde{\varsigma}_{\delta,\varepsilon}(y)\log\tilde{\varsigma}_{\delta,\varepsilon}(y)\,dm(y)\\
&=\int_M\tilde{\varsigma}_{\delta,\varepsilon}\bigl(\Phi_\varepsilon(x)\bigr)\log\tilde{\varsigma}_{\delta,\varepsilon}\bigl(\Phi_\varepsilon(x)\bigr)\det[D\Phi_\varepsilon(x)]\,dm(x)\\
&=\int_M\varsigma_\delta(x)\log\biggl(\frac{\varsigma_\delta(x)}{\det[D\Phi_\varepsilon(x)]}\biggr)\,dm(x)\\
&=\operatorname{Ent}_m(\nu^t_\delta)-\int_M\varsigma_\delta(x)\log\bigl(\det[D\Phi_\varepsilon(x)]\bigr)\,dm(x).
\end{aligned}$$

Similarly, we observe

$$\begin{aligned}
&\int_M V\,d\tilde{\nu}^t_{\delta,\varepsilon}=\int_M V(y)\tilde{\varsigma}_{\delta,\varepsilon}(y)\,dm(y)\\
&=\int_M V\bigl(\Phi_\varepsilon(x)\bigr)\tilde{\varsigma}_{\delta,\varepsilon}\bigl(\Phi_\varepsilon(x)\bigr)\det[D\Phi_\varepsilon(x)]\,dm(x)=\int_M V\bigl(\Phi_\varepsilon(x)\bigr)\varsigma_\delta(x)\,dm(x).
\end{aligned}$$

Thus we see

$$
\begin{aligned}
&\lim_{\varepsilon\to 0+}\frac{1}{\varepsilon}\{f(\nu_\delta^t)-f(\tilde{\nu}_{\delta,\varepsilon}^t)\}\\
&=\lim_{\varepsilon\to 0+}\frac{1}{\varepsilon}\{\mathrm{Ent}_m(\nu_\delta^t)-\mathrm{Ent}_m(\tilde{\nu}_{\delta,\varepsilon}^t)\}+\lim_{\varepsilon\to 0+}\frac{1}{\varepsilon}\bigg\{\int_M V\,d\nu_\delta^t-\int_M V\,d\tilde{\nu}_{\delta,\varepsilon}^t\bigg\}\\
&=\lim_{\varepsilon\to 0+}\int_M\frac{1}{\varepsilon}\log\big(\det[D\Phi_\varepsilon(x)]\big)\,d\nu_\delta^t(x)+\lim_{\varepsilon\to 0+}\frac{1}{\varepsilon}\int_M\big\{V(x)-V\big(\Phi_\varepsilon(x)\big)\big\}\,d\nu_\delta^t(x)\\
&=\int_M\frac{\partial}{\partial\varepsilon}\bigg|_{\varepsilon=0+}\det[D\Phi_\varepsilon(x)]\,d\nu_\delta^t(x)-\int_M\langle\operatorname{grad}h_t,\operatorname{grad}V\rangle\,d\nu_\delta^t\\
&=\int_M\operatorname{trace}(\operatorname{Hess}h_t)\,d\nu_\delta^t-\int_M\langle\operatorname{grad}h_t,\operatorname{grad}V\rangle\,d\nu_\delta^t\\
&=\int_M\Delta h_t\,d\nu_\delta^t-\int_M\langle\operatorname{grad}h_t,\operatorname{grad}V\rangle\,d\nu_\delta^t.
\end{aligned}
$$

Since ν_δ^t converges to μ_t weakly as δ goes to zero, we find

$$\lim_{\delta\to 0+}\lim_{\varepsilon\to 0+}\frac{1}{\varepsilon}\{f(\nu_\delta^t)-f(\tilde{\nu}_{\delta,\varepsilon}^t)\}=\int_M\{\Delta h_t-\langle\operatorname{grad}h_t,\operatorname{grad}V\rangle\}\,d\mu_t. \tag{6.7}$$

These three inequalities (6.5), (6.6) and (6.7) together imply

$$\liminf_{\delta\to 0+}\frac{1}{\delta}\bigg\{\int_M h_t\,d\nu_\delta^t-\int_M h_t\,d\mu_t\bigg\}\ge\int_M\big\{\Delta h_t-\langle\operatorname{grad}h_t,\operatorname{grad}V\rangle\big\}\,d\mu_t.$$

Moreover, the same inequality for $-h$ gives the reverse inequality (with lim sup instead of lim inf). Therefore we obtain

$$\lim_{\delta\to 0+}\frac{1}{\delta}\bigg\{\int_M h_t\,d\nu_\delta^t-\int_M h_t\,d\mu_t\bigg\}=\int_M\big\{\Delta h_t-\langle\operatorname{grad}h_t,\operatorname{grad}V\rangle\big\}\,d\mu_t. \tag{6.8}$$

It follows from Lemma 6.4 and Theorem 6.5 that

$$
\begin{aligned}
&\lim_{\delta\to 0+}\frac{1}{\delta}\bigg\{\int_M h_{t+\delta}\,d\mu_{t+\delta}-\int_M h_t\,d\mu_t\bigg\}\\
&=\lim_{\delta\to 0+}\frac{1}{\delta}\bigg\{\int_M(h_{t+\delta}-h_t)\,d\mu_{t+\delta}+\int_M h_t\,d\mu_{t+\delta}-\int_M h_t\,d\mu_t\bigg\}\\
&=\int_M\frac{\partial h_t}{\partial t}\,d\mu_t+\lim_{\delta\to 0+}\frac{1}{\delta}\bigg\{\int_M h_t\,d\nu_\delta^t-\int_M h_t\,d\mu_t\bigg\}.
\end{aligned}
$$

We consequently obtain (6.4) from (6.8) by integration. □

If we know the uniqueness of the weak solution in any way, then Theorems 5.11 and 6.6 show that it is in fact the gradient curve (in other words, the converse of Theorem 6.6 holds true). In particular, we obtain the following characterization of the heat kernel.

Corollary 6.7 *Let (M, g) be a compact Riemannian manifold of nonnegative sectional curvature and $p : (0, \infty) \times M \times M \longrightarrow [0, \infty)$ be the associated heat kernel. Then, for any point $x \in M$ and $t \in (0, \infty)$, we have $G(\delta_x, t) = p(t, x, \cdot) \cdot m$, where δ_x stands for the Dirac measure at x.*

Proof. Put $\nu_s := p(s, x, \cdot) \cdot m \in \mathcal{P}^*(M)$ for $s > 0$. We first recall that $p(s + t, x, \cdot)$, $t \geq 0$, is indeed a unique weak solution of the heat equation (6.3). Given $y \in M$, $t > 0$ and each small $\delta > 0$, we define $h^\delta(z, \tau) := p(t + \delta - \tau, y, z)$. Then any weak solution $\rho : M \times [0, \infty) \longrightarrow \mathbb{R}$ of the heat equation with $\rho(\cdot, 0) = p(s, x, \cdot)$ a.e. on M satisfies

$$\int_M h^\delta(z, t)\rho(z, t)\, dm(z) - \int_M h^\delta(z, 0)\rho(z, 0)\, dm(z) = 0$$

(see (6.4)). Letting δ go to zero yields that, for a.e. $y \in M$,

$$\rho(y, t) = \int_M p(t, y, z)p(s, x, z)\, dm(z) = p(s + t, x, y).$$

Therefore we deduce from Theorems 5.11 and 6.6 that ν_s is the gradient curve, namely $G(\nu_s, t) = p(s + t, x, \cdot) \cdot m$ for all $t \geq 0$. Letting s tend to zero, we obtain $G(\delta_x, t) = p(t, x, \cdot) \cdot m$. □

By virtue of Theorem 6.2, we recover the well-known contraction property of the heat kernel. See [RS] for more comprehensive results.

Corollary 6.8 *Let (M, g) be a compact Riemannian manifold of nonnegative sectional curvature and $p : (0, \infty) \times M \times M \longrightarrow [0, \infty)$ be the associated heat kernel. If $\mathrm{Ric}_M \geq K$, then, for any $x, y \in M$ and $t \in (0, \infty)$, we have*

$$d_2^W\big(p(t, x, \cdot) \cdot m, p(t, y, \cdot) \cdot m\big) \leq e^{-Kt} d_M(x, y).$$

Remark 6.9 The heat kernel on an Alexandrov space is constructed by Kuwae, Machigashira and Shioya [KMS] and the analogues of Corollaries 6.7 and 6.8 should hold true on Alexandrov spaces. However, its proof may be more involved.

References

[ABN] A. D. Alexandrov, V. N. Berestovskii and I. G. Nikolaev, *Generalized Riemannian spaces*, Russian Math. Surveys **41** (1986), 1–54.

[AGS] L. Ambrosio, N. Gigli and G. Savaré, Gradient flows in metric spaces and in the space of probability measures, Birkhäuser Verlag, Basel, 2005.

[BBI] D. Burago, Yu. Burago and S. Ivanov, A course in metric geometry, American Mathematical Society, Providence, RI, 2001.

[BGP] Yu. Burago, M. Gromov and G. Perel'man, *A. D. Alexandrov spaces with curvatures bounded below*, Russian Math. Surveys **47** (1992), 1–58.

[CMS] D. Cordero-Erausquin, R. J. McCann and M. Schmuckenschläger, *A Riemannian interpolation inequality á la Borell, Brascamp and Lieb*, Invent. Math. **146** (2001), 219–257.

[H] S. Halbeisen, *On tangent cones of Alexandrov spaces with curvature bounded below*, Manuscripta Math. **103** (2000), 169–182.

[JKO] R. Jordan, D. Kinderlehrer and F. Otto, *The variational formulation of the Fokker-Planck equation*, SIAM J. Math. Anal. **29** (1998), 1–17.

[KMS] K. Kuwae, Y. Machigashira and T. Shioya, *Sobolev spaces, Laplacian, and heat kernel on Alexandrov spaces*, Math. Z. **238** (2001), 269–316.

[LV1] J. Lott and C. Villani, *Weak curvature conditions and functional inequalities*, J. Funct. Anal. **245** (2007), 311–333.

[LV2] J. Lott and C. Villani, *Ricci curvature for metric-measure spaces via optimal transport*, to appear in Ann. of Math.

[Ly1] A. Lytchak, *Differentiation in metric spaces*, St. Petersburg Math. J. **16** (2005), 1017–1041.

[Ly2] A. Lytchak, *Open map theorem for metric spaces*, St. Petersburg Math. J. **17** (2006), 477–491.

[M] R. J. McCann, *Polar factorization of maps on Riemannian manifolds*, Geom. Funct. Anal. **11** (2001), 589–608.

[Oh1] S. Ohta, *Convexities of metric spaces*, Geom. Dedicata **125** (2007), 225–250.

[Oh2] S. Ohta, *On the measure contraction property of metric measure spaces*, Comment. Math. Helv. **82** (2007), 805–828.

[Oh3] S. Ohta, *Products, cones, and suspensions of spaces with the measure contraction property*, J. Lond. Math. Soc. **76** (2007), 225–236.

[Oh4] S. Ohta, *Markov type of Alexandrov spaces of nonnegative curvature*, preprint (2006), available at arXiv:0707.0102.

[Ot] F. Otto, *The geometry of dissipative evolution equation: the porous medium equation*, Comm. Partial Differential Equations **26** (2001), 101–174.

[OV] F. Otto and C. Villani, *Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality*, J. Funct. Anal. **173** (2000), 361–400.

[PP] G. Perel'man and A. Petrunin, *Quasigeodesics and gradient curves in Alexandrov spaces*, unpublished preprint (1994).

[RS] M.-K. von Renesse and K.-T. Sturm, *Transport inequalities, gradient estimates, entropy and Ricci curvature*, Comm. Pure Appl. Math. **58** (2005), 1–18.

[Sa] G. Savaré, *Gradient flows and diffusion semigroups in metric spaces under lower curvature bounds*, C. R. Math. Acad. Sci. Paris **345** (2007), 151–154.

[S1] K.-T. Sturm, *Convex functionals of probability measures and nonlinear diffusions on manifolds*, J. Math. Pures Appl. **84** (2005), 149–168.

[S2] K.-T. Sturm, *On the geometry of metric measure spaces*, Acta Math. **196** (2006), 65–131.

[S3] K.-T. Sturm, *On the geometry of metric measure spaces.* II, Acta Math. **196** (2006), 133–177.

[V1] C. Villani, Topics in optimal transportation, American Mathematical Society, Providence, RI, 2003.

[V2] C. Villani, Optimal transport, old and new, to appear.