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, then our gradient flow of the free energy produces the solution of the linear Fokker-Planck equation, as was demonstrated by Jordan, Kinderlehrer and Otto in the Euclidean setting.

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 is 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 on $\mathcal{P}(X)$ provides a powerful tool for the study of the structure of X. One of the most interesting examples is an approach to the synthetic lower Ricci curvature bound for general metric measure spaces (see [RS], [S2], [S3], [LV1] and [LV2], and also [CMS] and [Oh2] for related works). There the convexity of an entropy on $\mathcal{P}(X)$ plays a role of the lower Ricci curvature bound of X.

Our first main result (Theorem 3.5) concerns the infinitesimal structure of the Wasserstein space $\mathcal{P}(X)$ over a compact Alexandrov space X of curvature bounded below. It is inspired by the fact that $\mathcal{P}(X)$ is an Alexandrov space of nonnegative

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

curvature if and only if so is X. This fact implies that $\mathcal{P}(X)$ has Euclidean tangent 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 [Oh3]). The 2-uniform smoothness can be regarded as a generalization of the nonnegatively curved property 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 the metric geometry.

Our Riemannian structure on $\mathcal{P}(X)$ enables us to consider gradient flows of lower semi-continuous and K-convex functions on $\mathcal{P}(X)$. We establish the existence, completeness, uniqueness and the contraction property of such a gradient flow (Theorems 5.9 and 5.11). This is new even when X is a Riemannian manifold, and one of the most important situations missed in a recent book [AGS] by Ambrosio, Gigli and Savaré. (Their objects in [AGS] are a metric space whose distance function is convex in a sense, as well as the Wasserstein space over a Hilbert space.) Our construction of the gradient flow is different from [AGS] (nor [JKO]), but basically follows and extends the discussion in [Ly2] (see also [PP]) which concerns, among others, locally Lipschitz functions on an Alexandrov space.

In a particular case where the underlying space X is a Riemannian manifold, we show that our gradient flow of the free energy produces the solution of the linear Fokker-Planck equation (Theorem 6.3). This generalizes the celebrated work by Jordan, Kinderlehrer and Otto [JKO] on Euclidean spaces to a much wider class of spaces. As a corollary, the gradient flow of the relative entropy starting from a Dirac measure describes the heat kernel (Corollary 6.4). These results provide a new insight and will become effective instruments in analysis, probability theory and geometry on Riemannian manifolds (see [OV] for a formal discussion). It is worthful to mention that the recent progress on the synthetic lower Ricci curvature bound makes it possible to apply our results to the free energy.

We refer the readers to forthcoming lecture notes by C. Villani [V2] which contains some results related especially to Section 6 in the present article.

The article is organized as follows: Section 2 contains reviews on Alexandrov spaces and Wasserstein spaces. We verify 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 Riemannian 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 also grateful to the institute for its hospitality.

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 the geodesic γ is *minimal*. A metric space (X, d) is said to be *geodesic* if any two points in X are connected by a (not necessarily unique) minimal geodesic. For $x \in X$ and $r > 0$, we denote by $B(x, r)$ and $\overline{B}(x, r)$ the open and closed ball 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. Standard references are [ABN], [BGP] and [BBI]. An Alexandrov space is a metric space whose sectional curvatures are bounded from below by a constant in a certain sense. The definition of an Alexandrov space is based on a simple inequality for geodesic triangles which is indeed equivalent to the lower sectional curvature bound in the Riemannian case. An Alexandrov space has a nice infinitesimal structure, what is called a *space of directions*, as well as a *tangent cone*.

For $\kappa \in \mathbb{R}$, we denote by $\mathbb{M}^2(\kappa)$ a simply-connected, 2-dimensional Riemannian manifold of constant sectional curvature κ. That is, $\mathbb{M}^2(\kappa)$ is a 2-sphere ($\kappa > 0$) or a Euclidean plane ($\kappa = 0$) or a hyperbolic plane ($\kappa < 0$). 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 $\gamma_{\tilde{x}\tilde{y}} : [0, 1] \longrightarrow \mathbb{M}^2(\kappa)$ a 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 a 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:

• The inequality (2.1) with $\lambda = 1/2$ (for all x, y, z and γ) implies (2.1) for all $\lambda \in [0,1]$.

• If (X,d) is an Alexandrov space of curvature $\geq \kappa$, then, given three points $x, y, z \in X$ 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 upto scaling its distance by a positive constant.

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 curvatures are greater than or equal to κ everywhere.

(ii) For a compact convex domain $\Omega \subset \mathbb{R}^n$, let $X = \partial\Omega$ equip 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 concern 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 an 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 a 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 $s, t > 0$. This 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

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

where $\bar{\gamma}$ and $\bar{\eta}$ are given two 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 gives a natural distance structure 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 of $(\Sigma_x X, \angle_x)$, 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 $(*, 0) \in C_x X$. We remark that, in the case of $\kappa = 0$, the Euclidean cosine formula implies that

$$\begin{aligned}
\sigma_x\big((\gamma, s), (\eta, t)\big)^2 &= s^2 + t^2 - 2st\cos\angle_x(\gamma, \eta) \\
&= s^2 + t^2 - 2st\cos\left(\lim_{\varepsilon\to 0+} \angle\widetilde{\gamma(s\varepsilon)}\,\tilde{x}\,\widetilde{\eta(t\varepsilon)}\right) \\
&= \lim_{\varepsilon\to 0+} \frac{1}{\varepsilon^2} d_X\big(\gamma(s\varepsilon), \eta(t\varepsilon)\big)^2.
\end{aligned}$$

A similar equation also holds true for a general $\kappa \in \mathbb{R}$. It means that the distance structure σ_x is compatible with the scaling limit of d_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 $\ge \kappa$. 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 the 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 $\dim X \ge 2$), 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 [Ha] constructed an example 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 a Wasserstein space which is a 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 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 the 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) For $p \in [1, \infty)$, the L_p-*Wasserstein space* over (X, d) is a metric space $(\mathcal{P}(X), d_p^W)$ equipped with a distance structure d_p^W defined by

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

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 $d_p^W(\mu, \nu)$ 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 $d_p^W(\mu, \nu)$. For $p \leq p'$, the Hölder inequality immediately implies $d_p^W(\mu, \nu) \leq d_{p'}^W(\mu, \nu)$. The underlying metric space X is isometrically embedded into $\mathcal{P}(X)$ by identifying a point $x \in X$ with a Dirac measure $\delta_x \in \mathcal{P}(X)$ at x. The Wasserstein space $(\mathcal{P}(X), d_p^W)$ is a compact metric space since X is compact. As we see in the following proposition, the Wasserstein distance d_p^W is one way to metrize the weak topology of $\mathcal{P}(X)$.

Proposition 2.5 (cf. [V1, Theorem 7.12]) *A sequence $\{\mu_i\}_{i \in \mathbb{N}} \subset \mathcal{P}(X)$ converges to $\mu \in \mathcal{P}(X)$ with respect to d_p^W if and only if $\{\mu_i\}_{i \in \mathbb{N}}$ converges to μ weakly.*

There are several equivalent representations of the Wasserstein distance. Here we recall one (for $p = 1$) of them for the later use. Let $\mathrm{Lip}(X)$ be the set of all Lipschitz functions on X and, for $h \in \mathrm{Lip}(X)$,

$$\mathrm{Lip}(h) := \sup_{x, y \in X, \, x \neq y} \frac{|h(x) - h(y)|}{d(x, y)}. \tag{2.5}$$

For $L \in [0, \infty)$, h is said to be *L-Lipschitz* if $\mathrm{Lip}(h) \leq L$.

Theorem 2.6 (Kantorovich-Rubinstein theorem, cf. [V1, Theorem 1.14]) *For $\mu, \nu \in \mathcal{P}(X)$, we have*

$$d_1^W(\mu,\nu) = \sup_{h \in \mathrm{Lip}(X),\, \mathrm{Lip}(h) \le 1} \left\{ \int_X h\, d\mu - \int_X h\, d\nu \right\}.$$

Hereafter, we consider only the quadratic case (i.e., $p = 2$). 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 a 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 2.7 ([LV1, 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 $\alpha : [0,1] \longrightarrow \mathcal{P}(X)$ betwee them, there exists a Borel probability measure $\Pi \in \mathcal{P}(\Gamma(X))$ such 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 information. For a Riemannian manifold (M, g), let us denote by $\mathcal{P}^{ac}(X) \subset \mathcal{P}(X)$ the subset consisting of measures which are absolutely continuous with respect to the Riemannian volume element m.

Theorem 2.8 ([M, Theorem 9]) *Let us take a compact Riemannian manifold (M, g). 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*

$$\psi(x) = \inf_{y \in M} \{ d_M(x,y)^2/2 - \phi(y) \} \tag{2.6}$$

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 a unique optimal transportation from μ to ν, that is, we have*

$$\left\{ \int_M d_M\big(x, \Psi(x)\big)^2\, d\mu(x) \right\}^{1/2} = d_2^W(\mu,\nu) \tag{2.7}$$

and Ψ is a unique map (*upto a change on a μ-null measure set*) *satisfying $\Psi_*\mu = \nu$ and* (2.7).

The condition (2.6) 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 (2.6) implies that ψ is a Lipschitz function. Hence ψ is differentiable and the gradient $\operatorname{grad}\psi$ makes sense almost everywhere. In the view of Proposition 2.7, 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 minimal geodesic between μ and ν.

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

$$\mathrm{Ent}_m(\mu) := \int_M \rho \log \rho \, dm \ \in (-\infty, \infty], \tag{2.8}$$

where the function ρ stands for the absolutely continuous part 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 2.9 (i) *The relative entropy* Ent_m *satisfies*

$$\mathrm{Ent}_m(\mu) \geq -\log m(M)$$

for all $\mu \in \mathcal{P}(M)$ *and the 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* $\mathcal{P}^*(M) := \{\mu \in \mathcal{P}(M) \,|\, \mathrm{Ent}_m(\mu) < \infty\}$ *is dense in* $\mathcal{P}(M)$ *with respect to* d_2^W.

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

Theorem 2.10 ([RS, Theorem 1.1]) *A compact Riemannian manifold* (M, g) *satisfies* $\mathrm{Ric}_M \geq K$ *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. As M is assumed to be compact, such a constant $K \in \mathbb{R}$ actually exists. 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], [LV1], [LV2] and [Oh2] for the 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

$$f(\mu) := \mathrm{Ent}_m(\mu) + \int_M V \, d\mu \ \in (-\log m(M) + \inf V, \infty]. \tag{2.9}$$

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 $e^{-V} \cdot m$. Clearly f is lower semi-continuous and the following generalization of Theorem 2.10 holds.

Theorem 2.11 ([S1, Theorem 1.3], [LV1, Theorem 7.3]) *Let (M, g) be a compact Riemannian manifold and $V \in C^\infty(M)$. Then we have $\mathrm{Ric}_M + \mathrm{Hess}\, V \geq K$ for $K \in \mathbb{R}$ if and only if the free energy (2.9) is K-convex on $\mathcal{P}(M)$.*

Here $\mathrm{Hess}\, V$ stands for the Hessian of V, and the inequality $\mathrm{Ric}_M + \mathrm{Hess}\, V \geq K$ is read as $\mathrm{Ric}_M(v, v) + \mathrm{Hess}\, V(v, v) \geq K|v|^2$ for all $v \in TM$. Again the compactness of M guarantees 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 the Wasserstein space over a Euclidean space. The existence of such a structure is rigorously justified by a 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 [LV1, Proposition A.9]). However, it is also known that the Wasserstein space is not an Alexandrov space any more (even for a negative κ) if the underlying metric space does not have a nonnegative curvature (see [S2, Proposition 2.10]). In order to overcome this difficulty, we introduce a 'non-Hilbertian' extension of the nonnegatively curved property 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 $\lambda \in [0, 1]$,

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

where $S \geq 1$ is a fixed constant. We say that a geodesic metric space (X, d) *satisfies* (3.1) if there is a constant $S \geq 1$ such that the inequality (3.1) holds for all $x, y, z \in X$, minimal geodesic $\gamma : [0, 1] \longrightarrow X$ from y to z and for all $\lambda \in [0, 1]$. 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 the theory of Banach spaces. For instance, L_p-spaces with $p \in [2, \infty)$ satisfy (3.1) with $S = \sqrt{p - 1}$. We refer to [Oh1] and [Oh3] for works in this direction.

Proposition 3.1 *A compact geodesic metric space (X, d) satisfies* (3.1) *with the 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 $S \geq 1$. 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 2.7, there exists a probability measure $\Pi \in \mathcal{P}(\Gamma(X))$ with $\mu_\tau := (e_\tau)_* \Pi = \alpha(\tau)$ for $\tau \in [0, 1]$.

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

$$\Pi_\lambda^w \in \mathcal{P}(\{\gamma \in \Gamma(X) \,|\, \gamma(\lambda) = w\}), \quad q^w \in \mathcal{P}(X) = \mathcal{P}(X \times \{w\}) \subset \mathcal{P}(X \times X)$$

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

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

For $a = 0,1$, define a (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) with respect to $(dq^w(x) d\Pi_\lambda^w(\gamma)) d\mu_\lambda(w)$ on $(X \times \Gamma(X)) \times X$,

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

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

The analogue of Proposition 3.1 for the reverse inequality of (3.1),

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

(in other words, the 2-uniform convexity) does not holds true. In fact, it is known that $\mathcal{P}(\mathbb{R}^2)$ does not satisfy (3.3) with $S = 1$ 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 3.2 *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*

$$d\big(x, \gamma(\lambda)\big)^2 \geq (1-\lambda)d(x,y)^2 + \lambda d(x,z)^2 - \left\{1 + \sup_{\tau\in[0,1]} d\big(x,\gamma(\tau)\big)^2\right\} \cdot (1-\lambda)\lambda d(y,z)^2. \tag{3.4}$$

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, that is, for all $\lambda \in (0,1)$, we will show

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

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 a small $\varepsilon > 0$, put $w := \gamma(\lambda)$, $y_\varepsilon := \gamma(\lambda - \varepsilon)$ and $z_\varepsilon := \gamma(\lambda + \varepsilon)$. We also define minimal geodesics $\gamma_-, \gamma_+ : [0,1] \longrightarrow X$ from w to y_ε and 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

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

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

$$\cosh d(x,y_\varepsilon) \leq \cosh d(x,w) \cosh d(y_\varepsilon, w) - \sinh d(x,w) \sinh d(y_\varepsilon, w) \cos \angle_w(\eta, \gamma_-).$$

These together yield that

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

Similarly, it holds that

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

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

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

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

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

Lemma 3.3 *Let (X,d) be an Alexandrov space of curvature ≥ -1. Given $x,y,z\in X$, 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*

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

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

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

Here we used the assumption $d(y,z) \le 1$ in the second implication. Thus we find

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

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

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

$\square$

Thus the inequality (3.6) tends to (2.2) as $d(y,z)$ is getting small, namely a general Alexandrov space is close to an Alexandrov space of curvature ≥ 0 in a small scale (as was already observed in Proposition 2.3 in a different context).

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 be careful 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 2.7). In fact, some $\gamma \in \operatorname{supp}\Pi$ may be long even when α is short.

Lemma 3.4 *Let (X,d) be a compact Alexandrov space of curvature ≥ -1 and set $D := \operatorname{diam} X$. Then, for any minimal geodesic $\alpha : [0,l] \longrightarrow \mathcal{P}(X)$, $0 \le s \le t \le l$ with $t - s \le D^{-1}$, $\nu \in \mathcal{P}(X)$ and for any $\lambda \in [0,1]$, we have*

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

Proof. The proof is similar to the proof of Proposition 3.1, so we give only an outline. By Proposition 2.7, we find $\Pi \in \mathcal{P}(\Gamma(X))$ such that $(e_\tau)_*\Pi = \alpha(\tau)$ for $\tau \in [0,1]$. 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, that is,

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

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 (3.6) (instead of (3.2)) that

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

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. □

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

Theorem 3.5 (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 $0 < s, t \leq 1$ without loss of generality.

(i) We shall show that the function

$$\varepsilon \longmapsto \frac{1}{\varepsilon} d_2^W\big(\alpha(s\varepsilon), \beta(t\varepsilon)\big)$$

is 'almost' non-decreasing as ε goes to zero. Take $\varepsilon \in (0, \min\{\delta, D^{-1}\})$ and $\lambda \in (0,1)$. Then it follows from Lemma 3.4 with $\nu = \beta(t\varepsilon)$ and $(s,t) = (0, s\varepsilon)$ that

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

Similarly, we have

$$\begin{aligned}
&d_2^W\big(\alpha(\lambda s\varepsilon),\beta(\lambda t\varepsilon)\big)^2\\
&\geq (1-\lambda)d_2^W\big(\alpha(\lambda s\varepsilon),\beta(0)\big)^2+\lambda d_2^W\big(\alpha(\lambda s\varepsilon),\beta(t\varepsilon)\big)^2\\
&\qquad -(1-\lambda)\lambda d_2^W\big(\beta(0),\beta(t\varepsilon)\big)^2+\varepsilon^2\theta(\varepsilon)\\
&=(1-\lambda)(\lambda s\varepsilon)^2+\lambda d_2^W\big(\alpha(\lambda s\varepsilon),\beta(t\varepsilon)\big)^2-(1-\lambda)\lambda(t\varepsilon)^2+\varepsilon^2\theta(\varepsilon). \qquad (3.8)
\end{aligned}$$

Combining (3.7) and (3.8), we obtain

$$\begin{aligned}
&d_2^W\big(\alpha(\lambda s\varepsilon),\beta(\lambda t\varepsilon)\big)^2\\
&\geq (1-\lambda)\lambda^2(s\varepsilon)^2-(1-\lambda)\lambda(t\varepsilon)^2\\
&\qquad +\lambda\big\{(1-\lambda)(t\varepsilon)^2+\lambda d_2^W\big(\alpha(s\varepsilon),\beta(t\varepsilon)\big)^2-(1-\lambda)\lambda(s\varepsilon)^2\big\}+\varepsilon^2\theta(\varepsilon)\\
&=\lambda^2 d_2^W\big(\alpha(s\varepsilon),\beta(t\varepsilon)\big)^2+\varepsilon^2\theta(\varepsilon),
\end{aligned}$$

and hence

$$\frac{1}{\lambda\varepsilon}d_2^W\big(\alpha(\lambda s\varepsilon),\beta(\lambda t\varepsilon)\big)\geq\frac{1}{\varepsilon}d_2^W\big(\alpha(s\varepsilon),\beta(t\varepsilon)\big)+\theta(\varepsilon).$$

This implies that 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.

(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

$$\begin{aligned}
&d_2^W\big(\alpha(\lambda s\varepsilon),\beta(t\varepsilon)\big)^2\\
&\geq(1-\lambda)(t\varepsilon)^2+\lambda d_2^W\big(\alpha(s\varepsilon),\beta(t\varepsilon)\big)^2-(1-\lambda)\lambda(s\varepsilon)^2+\varepsilon^2\theta(\varepsilon).
\end{aligned}$$

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

$$\sigma_\mu\big((\alpha,\lambda s),(\beta,t)\big)^2\geq(1-\lambda)t^2+\lambda\sigma_\mu\big((\alpha,s),(\beta,t)\big)^2-(1-\lambda)\lambda s^2. \qquad (3.9)$$

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

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

and hence

$$\sigma_\mu\big((\alpha,\lambda s),(\beta,t)\big)^2\leq(1-\lambda)t^2+\lambda\sigma_\mu\big((\alpha,s),(\beta,t)\big)^2-(1-\lambda)\lambda s^2. \qquad (3.10)$$

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. $\square$

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 an equivalence relation such that $\alpha \sim \beta$ if they coincide near μ. Define the angle $\angle_\mu(\alpha, \beta) \in [0, \pi]$ between $\alpha, \beta \in \Sigma'_\mu[\mathcal{P}(X)]$ by

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

Then the angle $\angle_\mu$ provides 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 3.5(ii) assures that σ_μ given by Theorem 3.5(i) is coincide with the distance function on the Eulidean cone $C_\mu[\mathcal{P}(X)]$. That is to say, Theorem 3.5(ii) says that, for all $s, t \ge 0$,

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

which is rewritten as

$$\sigma_\mu\big((\alpha, s), (\beta, t)\big)^2 = s^2 + t^2 - 2st \cos \angle_\mu(\alpha, \beta).$$

This corresponds to the Euclidean cosine formula. 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

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

For the later convenience, we also 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 3.6 The proof of Theorem 3.5 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.

3.4 The first variation formula

In this subsection, we recall a kind of first variation formula (see [Ly1, Lemma 9.1]). A curve $\xi : [0, \delta] \longrightarrow \mathcal{P}(X)$ is said to be *differentiable* at 0 if there is $(\alpha, t) \in C_{\xi(0)}[\mathcal{P}(X)]$ 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 $\xi(0)$ to $\xi(\varepsilon_i)$, the sequence $\{(\alpha_i, d_2^W(\xi(0), \xi(\varepsilon_i))/\varepsilon_i)\}_{i \in \mathbb{N}}$ converges to (α, t) in $C_{\xi(0)}[\mathcal{P}(X)]$. In this case, we put $\xi'(0) := (\alpha, t)$. Note that $\xi'(0) = o_{\xi(0)}$ if and only if $\lim_{\varepsilon \to 0+} d_2^W(\xi(0), \xi(\varepsilon))/\varepsilon = 0$ and that, if $\xi'(0) \neq o_{\xi(0)}$, then

$$\lim_{\varepsilon \to 0+} d_2^W\big(\xi(0), \xi(\varepsilon)\big)/\varepsilon = t, \qquad \lim_{i \to \infty} \angle_{\xi(0)}(\alpha_i, \alpha) = 0.$$

Lemma 3.7 (The first variation formula) *Given $\mu, \nu \in \mathcal{P}(X)$, 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 $\xi, \zeta : [0, \delta] \longrightarrow \mathcal{P}(X)$ 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 $\{\varepsilon_i\}_{i \in \mathbb{N}}$ tending to zero and satisfying

$$\lim_{i \to \infty} \frac{h(\varepsilon_i) - h(0)}{\varepsilon_i} = \limsup_{\varepsilon \to 0+} \frac{h(\varepsilon) - h(0)}{\varepsilon}.$$

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 $\zeta(\varepsilon_i)$. Put $v_i := (\beta_i, a_i/\varepsilon_i) \in C'_\mu[\mathcal{P}(X)]$ and $w_i := (\gamma_i, b_i/\varepsilon_i) \in C'_\nu[\mathcal{P}(X)]$ for brevity. Recall that the differentiabilities of ξ and ζ say that v_i and w_i tend to $\xi'(0)$ and $\zeta'(0)$ as i goes to the infinity in $C_\mu[\mathcal{P}(X)]$ and $C_\nu[\mathcal{P}(X)]$, respectively. We also 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 the 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 a product space $Y := C'_\mu[\mathcal{P}(X)] \times C'_\nu[\mathcal{P}(X)]$ and a function $D : Y \longrightarrow [0, \infty)$ defined by, for $v = (\beta, s) \in C'_\mu[\mathcal{P}(X)]$ and $w = (\gamma, t) \in C'_\nu[\mathcal{P}(X)]$,

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

The function D can be regarded as a differential of the distance function $d_2^W : \mathcal{P}(X) \times \mathcal{P}(X) \longrightarrow [0, \infty)$ 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 $t \geq 0$,

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

Note that

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

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

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

$\square$

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 $\mathcal{P}(X)$ which can be found in [AGS], [Ly2] etc. 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 satisfy the following:

A function $f : \mathcal{P}(X) \longrightarrow (-\infty, \infty]$ is nontrivial, K-convex and lower semi-continuous. (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

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

In particular, given $\mu, \nu \in \mathcal{P}^*(X)$, every minimal geodesic between them is contained in $\mathcal{P}^*(X)$ (in other words, the set $\mathcal{P}^*(X)$ is convex 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 a paragraph following Theorem 3.5), 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 is an object in Section 6. The following lemma is a well-known fact on K-convex functions.

Lemma 4.1 *Let $\alpha : [0,l] \longrightarrow \mathcal{P}^*(X)$ be a minimal geodesic. Then $f\circ\alpha$ is differentiable from both sides at every point, and we have*

$$\lim_{\varepsilon\to 0+} \frac{f(\alpha(\varepsilon)) - f(\alpha(0))}{\varepsilon} + \lim_{\varepsilon\to 0+} \frac{f(\alpha(l-\varepsilon)) - f(\alpha(l))}{\varepsilon} \le -Kl.$$

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 $s \ge 0$, 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 $D'_\mu f(v) \ge -s|\nabla_- f|(\mu)$. Moreover, the K-convexity also implies that

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

We further define a function $D_\mu f : C_\mu[\mathcal{P}^*(X)] \longrightarrow \mathbb{R}$ 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 $D_\mu f(v) \ge -s|\nabla_- f|(\mu)$, $D_\mu f(v) = s\cdot D_\mu f((\alpha,1))$ and also $D_\mu f(v) \le D'_\mu f(v)$ if $v \in C'_\mu[\mathcal{P}^*(X)]$. The following lemma means that $D_\mu f$ is 'almost' convex. The convexy is easily verified by taking a scaling limit if the space in question is of finite dimension (see Proposition 2.3). However, it is not the case because $\mathcal{P}(X)$ is obviously infinite dimensional even when X is of finite dimension.

Lemma 4.2 *Fix a point $\mu \in \mathcal{P}^*(X)$ with $|\nabla_- f|(\mu) < \infty$ and $v, w \in C_\mu[\mathcal{P}^*(X)]$. 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 a 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 a sufficiently small $\lambda > 0$.

Now we prove (4.6), more precisely,

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

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

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

It follows from Lemma 3.4 that

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

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

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

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 to (4.8). Therefore we obtain (4.7).

For a 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

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

Note that $D_\mu f(u) \leq D'_\mu f(u)$ and that

$$\sigma_\mu(o_\mu, v) = \sigma_\mu(o_\mu, v_i) = s, \qquad \sigma_\mu(o_\mu, w) = t \geq t_i = \sigma_\mu(o_\mu, w_i).$$

Combining these with (4.9) and (4.10), we obtain

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

□

The convexity of $D_\mu f$ enables us to find the unique steepest direction of f in $\Sigma_\mu[\mathcal{P}^*(X)]$.

Lemma 4.3 *For any $\mu \in \mathcal{P}^*(X)$ with $0 < |\nabla_- f|(\mu) < \infty$, there exists a 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 identify $\alpha, \beta \in \Sigma_\mu[\mathcal{P}^*(X)]$ with $(\alpha, 1), (\beta, 1) \in C_\mu[\mathcal{P}^*(X)]$, respectively.*

Proof. By the definition of $|\nabla_- f|(\mu)$, 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 an arbitrary $\varepsilon > 0$, Lemma 4.2 assures that there is $u = (\xi, \tau) \in C_\mu[\mathcal{P}^*(X)]$ with

$$\begin{aligned} D_\mu f(u) &\leq \frac{1}{2} D_\mu f(\alpha_i) + \frac{1}{2} D_\mu f(\alpha_j) + \varepsilon \leq \frac{1}{2} D'_\mu f(\alpha_i) + \frac{1}{2} D'_\mu f(\alpha_j) + \varepsilon, \\ \tau^2 &\leq 1 - \frac{1}{4}\sigma_\mu\big((\alpha_i, 1), (\alpha_j, 1)\big)^2 + \varepsilon. \end{aligned}$$

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

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

As ε is arbitrary and $|\nabla_- f|(\mu) > 0$, by letting i and j go to the infinity, we obtain

$$\lim_{i,j\to\infty} \sigma_\mu\big((\alpha_i, 1), (\alpha_j, 1)\big) = 0.$$

Thus $\{\alpha_i\}_{i\in\mathbb{N}}$ 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 4.2.

Next we consider the second assertion. For each $i \in \mathbb{N}$, by Lemma 4.2, we find $u_i \in C_\mu[\mathcal{P}^*(X)]$ satisfying

$$\begin{aligned}D_\mu f(u_i) &\leq \frac{1}{2}D_\mu f\big((\alpha, i)\big) + \frac{1}{2}D_\mu f(\beta) + i^{-1} = -\frac{i}{2}|\nabla_- f|(\mu) + \frac{1}{2}D_\mu f(\beta) + i^{-1}, \\ \sigma_\mu(o_\mu, u_i)^2 &\leq \frac{i^2}{2} + \frac{1}{2} - \frac{1}{4}\sigma_\mu\big((\alpha, i), (\beta, 1)\big)^2 + i^{-1} = \frac{i^2}{4} + \frac{1}{4} + \frac{i}{2}\langle \alpha, \beta\rangle_\mu + i^{-1}.\end{aligned}$$

Thus we see

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

and hence

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

as i diverges to the infinity. This completes the proof. □

Thus we can define the (minus) *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 4.3. 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 $|\eta'|(t) := \lim_{s\to t} d_2^W(\eta(s),\eta(t))/|s-t|$ and it exists at a.e. $t \in [0,l]$ since η is Lipschitz (see, e.g., [AGS, Theorem 1.1.2]). See [AGS, §1.2] for more on upper gradients and [Ch] and [HK] for connections with the theory of Sobolev spaces. In order to show that the absolute gradient $|\nabla_- f|$ is an upper gradient for f, we introduce a variant of it. For $r > 0$ and $\mu \in \mathcal{P}^*(X)$, define

$$|\nabla_-^r f|(\mu) := \max\left\{0, \sup_{\nu\in[B(\mu,r)\setminus\{\mu\}]\cap\mathcal{P}^*(X)} \frac{f(\mu)-f(\nu)}{d_2^W(\mu,\nu)}\right\}. \tag{4.13}$$

Recall that $B(\mu,r)$ denotes an open ball (in $\mathcal{P}(X)$) with center μ and radius r. Clearly $|\nabla_-^r f|(\mu) \ge |\nabla_- f|(\mu)$ holds.

Lemma 4.4 (cf. [AGS, Theorem 1.2.5]) *The function $|\nabla_-^r f| : \mathcal{P}^*(X) \longrightarrow [0,\infty]$ is lower semi-continuous and is an upper gradient for f.*

Proof. We first prove the lower semi-continuity. Take a sequence $\{\mu_i\}_{i\in\mathbb{N}} \subset \mathcal{P}^*(X)$ converging to $\mu \in \mathcal{P}^*(X)$. For an arbitrary $\nu \in [B(\mu,r)\setminus\{\mu\}]\cap\mathcal{P}^*(X)$, as $0 < d_2^W(\mu_i,\nu) < r$ for sufficiently large $i \in \mathbb{N}$, it follows from the lower semi-continuity of f that

$$\begin{aligned}
\frac{f(\mu)-f(\nu)}{d_2^W(\mu,\nu)} &\le \liminf_{i\to\infty} \frac{f(\mu_i)-f(\nu)}{d_2^W(\mu_i,\nu)} \\
&\le \liminf_{i\to\infty}\left(\sup_{\nu'\in[B(\mu_i,r)\setminus\{\mu_i\}]\cap\mathcal{P}^*(X)} \frac{f(\mu_i)-f(\nu')}{d_2^W(\mu_i,\nu')}\right) \\
&\le \liminf_{i\to\infty} |\nabla_-^r f|(\mu_i).
\end{aligned}$$

By taking the supremum in ν, we obtain

$$|\nabla_-^r f|(\mu) \le \liminf_{i\to\infty} |\nabla_-^r f|(\mu_i).$$

Therefore $|\nabla_-^r f|$ is lower semi-continuous and, in particular, Borel.

In order to show that $|\nabla_-^r f|$ is an upper gradient for f, let us take a Lipschitz curve $\eta : [0,l] \longrightarrow \mathcal{P}^*(X)$. Without loss of generality, we can suppose that η has a unit speed, $l < r$ and that $\int_0^l |\nabla_-^r f|(\eta(t))\,dt < \infty$. Put $h := f\circ\eta$ and $g := |\nabla_-^r f|\circ\eta$ for brevity. For any $s,t \in [0,l]$, we observe that $d_2^W(\eta(s),\eta(t)) \le |s-t| \le l < r$ and hence

$$h(s)-h(t) \le |\nabla_-^r f|\big(\eta(s)\big)\cdot d_2^W\big(\eta(s),\eta(t)\big) \le g(s)\cdot|s-t|. \tag{4.14}$$

Fix a large $k \in \mathbb{N}$. For $\tau \in (0, l/k)$, it follows from (4.14) that

$$\begin{aligned} h(0) - h(l) &= h(0) - h(\tau) + \sum_{j=1}^{k-1} \left\{ h\left(\tau + \frac{j-1}{k} l\right) - h\left(\tau + \frac{j}{k} l\right) \right\} \\ &\quad + h\left(\tau + \frac{k-1}{k} l\right) - h(l) \\ &\le h(0) - h(\tau) + \frac{l}{k} \sum_{j=0}^{k-1} g\left(\tau + \frac{j}{k} l\right). \end{aligned}$$

Integrating this inequality with respect to $\tau \in (0, l/k)$, we find

$$h(0) - h(l) \le h(0) - \frac{k}{l} \int_0^{l/k} h(t)\, dt + \int_0^l g(t)\, dt.$$

Letting k go to the infinity, we obtain, by the lower semi-continuity of h,

$$h(0) - h(l) \le \int_0^l g(t)\, dt.$$

Similarly, we deduce $h(l) - h(0) \le \int_0^l g(t)\, dt$ and it completes the proof. □

Lemma 4.5 (cf. [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*.*

Proof. We shall observe that, for every $\mu \in \mathcal{P}^*(X)$ and $r > 0$, it holds that

$$|\nabla_- f|(\mu) = \max\left\{0, \sup_{\nu \in [B(\mu,r) \setminus \{\mu\}] \cap \mathcal{P}^*(X)} \left(\frac{f(\mu) - f(\nu)}{d_2^W(\mu,\nu)} + \frac{K}{2} d_2^W(\mu,\nu) \right) \right\} \tag{4.15}$$

(see [AGS, Theorem 2.4.9]). On one hand, we immediately see by definition that

$$\begin{aligned} |\nabla_- f|(\mu) &= \max\left\{0, \limsup_{\nu \to \mu} \frac{f(\mu) - f(\nu)}{d_2^W(\mu,\nu)} \right\} \\ &= \max\left\{0, \limsup_{\nu \to \mu} \left(\frac{f(\mu) - f(\nu)}{d_2^W(\mu,\nu)} + \frac{K}{2} d_2^W(\mu,\nu) \right) \right\} \\ &\le \max\left\{0, \sup_{\nu \in [B(\mu,r) \setminus \{\mu\}] \cap \mathcal{P}^*(X)} \left(\frac{f(\mu) - f(\nu)}{d_2^W(\mu,\nu)} + \frac{K}{2} d_2^W(\mu,\nu) \right) \right\}. \end{aligned}$$

On the other hand, for each fixed $\nu \in [B(\mu, r) \setminus \{\mu\}] \cap \mathcal{P}^*(X)$, the K-convexity of f along a minimal geodesic $\alpha : [0,1] \longrightarrow \mathcal{P}^*(X)$ from μ to ν says that

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

for all $\lambda \in (0, 1]$. Deviding both sides by $d_2^W(\mu, \alpha(\lambda)) = \lambda d_2^W(\mu,\nu)$, we have

$$\frac{f(\mu) - f(\alpha(\lambda))}{d_2^W(\mu, \alpha(\lambda))} \ge \frac{f(\mu) - f(\nu)}{d_2^W(\mu,\nu)} + \frac{K}{2} (1 - \lambda) d_2^W(\mu,\nu).$$

Letting λ tend to zero, we obtain

$$|\nabla_- f|(\mu) \geq \frac{f(\mu) - f(\nu)}{d_2^W(\mu, \nu)} + \frac{K}{2} d_2^W(\mu, \nu).$$

Then (4.15) follows by taking the supremum in ν.

The lower semi-continuity of $|\nabla_- f|$ is derived from the representation (4.15) just as in the proof of Lemma 4.4. It also follows from (4.15) and the definition of $|\nabla_-^r f|(\mu)$ that

$$|\nabla_- f|(\mu) \geq |\nabla_-^r f|(\mu) + \frac{\min\{K, 0\}}{2} r.$$

As every $|\nabla_-^r f|$ is an upper gradient for f, by letting r tend to zero, we see that $|\nabla_- f|$ is also 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 K-convex, lower semi-continuous function f on $\mathcal{P}(X)$ by using the 'Riemannian structure' of $\mathcal{P}(X)$ established in Section 3. Recall that $\mathcal{P}^*(X) = \{\mu \in \mathcal{P}(X) \,|\, f(\mu) < \infty\}$. We will follow known strategies for constructing gradient flows, especially the discussion in [Ly2] (see also [PP]). However, we need to be careful at some points because of the discontinuity of f as well as the infinite dimensionality of $\mathcal{P}(X)$, so we shall give all proofs for completeness. Furthermore, we will obtain some seemingly new estimates (see Propositions 5.7 and 5.12 below). 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 Existence and completeness

In this subsection, we study the existence of a gradient curve with an initial point $\mu \in \mathcal{P}^*(X)$, i.e., a curve $t \longmapsto G(\mu, t)$. To do this, we first construct a 'gradient-like curve' (which will turn out to be a unit speed curve whose reparametrization produces a gradient curve) by way of the discrete approximation.

For $a \in [\omega, \infty)$, define the sublevel set of f by

$$U[a] := f^{-1}([\omega, a]) \subset \mathcal{P}^*(X). \tag{5.1}$$

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*

$$f(\nu) = f(\mu) - c, \quad d_2^W(\mu, \nu) = \operatorname{dist}\big(\mu, U[f(\mu) - c]\big). \tag{5.2}$$

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 (5.2). By the hypothesis, we know $f(\mu) - c \geq \omega$ and it guarantees $U[f(\mu) - c] \neq \emptyset$. As $U[f(\mu) - c]$ is compact, we can take $\nu \in U[f(\mu) - c]$ satisfying

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

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, it is continuous and we find $\lambda \in [0, 1]$ such that $f(\alpha(\lambda)) = f(\mu) - c$. If $\lambda < 1$, then we see

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

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

Now we need 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 a maximal subset such that, for each $a \in A$,

$$U[f(\mu) - ca] \cap \overline{B}(\mu, ra) \neq \emptyset.$$

Note that $0 \in A$ and it suffices to show $1 \in A$. We set $a := \sup_{a' \in A} a'$ and first prove that $a \in A$. If a is an isolated point in A, then clearly $a \in A$. Otherwise, let $\{a_i\}_{i \in \mathbb{N}} \subset A$ be an increasing sequence which converges to a. For each $i \in \mathbb{N}$, as $a_i \in A$, we can take $\nu_i \in \mathcal{P}^*(X)$ such that

$$f(\nu_i) \leq f(\mu) - ca_i, \qquad d_2^W(\mu, \nu_i) \leq ra_i.$$

Since $\mathcal{P}(X)$ is compact, by extracting a subsequence if necessary, the sequence $\{\nu_i\}_{i \in \mathbb{N}}$ converges to some $\nu \in \mathcal{P}(X)$. Then clearly $d_2^W(\mu, \nu) \leq ra$ and the lower semi-continuity of f implies $f(\nu) \leq f(\mu) - ca$. Thus $\nu \in U[f(\mu) - ca] \cap \overline{B}(\mu, ra) \neq \emptyset$ and hence $a \in A$.

We next suppose $a = \sup_{a' \in A} a' < 1$ 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 $|\nabla_{-} f|(\nu) \geq C > c/r$ by the hypothesis, so that we can choose $\nu' \in \mathcal{P}^*(X) \setminus \{\nu\}$ with $d_2^W(\nu', \nu)/r =: \delta \leq 1 - a$ and $f(\nu) - f(\nu') \geq (c/r) d_2^W(\nu, \nu')$. We observe

$$\begin{aligned} d_2^W(\mu, \nu') &\leq d_2^W(\mu, \nu) + d_2^W(\nu, \nu') \leq r(a + \delta), \\ f(\nu') &\leq f(\nu) - (c/r) d_2^W(\nu, \nu') \leq f(\mu) - c(a + \delta). \end{aligned}$$

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 to the maximality of a. Thus we obtain $1 \in A$.

In the case of $c = Cr$, we take an increasing sequence $\{c_i\}_{i \in \mathbb{N}}$ tending to c. Then we know $U[f(\mu) - c_i] \cap \overline{B}(\mu, r) \neq \emptyset$ for every $i \in \mathbb{N}$, and this together with the lower semi-continuity of f shows that $U[f(\mu) - c] \cap \overline{B}(\mu, r) \neq \emptyset$. ☐

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

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

In particular, $Cr \leq f(\mu) - \omega$ automatically holds in the situation 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

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

□

Now we give the definition of a gradient-like curve and observe several straightforward properties.

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

$$f\big(\eta(t)\big) = f\big(\eta(0)\big) - \int_0^t |\nabla_- f|\big(\eta(s)\big)\, ds \tag{5.3}$$

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

Recall that a Lipschitz curve η is said to be 1-Lipschitz if it satisfies $\mathrm{Lip}(\eta) \leq 1$ (see (2.5)).

Lemma 5.4 *Let $\eta : [0, l) \longrightarrow \mathcal{P}^*(X) \setminus \{\mu \in \mathcal{P}^*(X) \,|\, |\nabla_- f|(\mu) = 0\}$ be a gradient-like curve of f. Then the following hold.*

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

$$\lim_{\varepsilon \to 0+} \frac{f(\eta(t+\varepsilon)) - f(\eta(t))}{\varepsilon} = -|\nabla_- f|\big(\eta(t)\big) \tag{5.4}$$

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

(ii) *At every $t \in [0, l)$ with $|\nabla_- f|(\eta(t)) < \infty$, η is differentiable and*

$$\eta'(t) = \nabla_- f\big(\eta(t)\big)/|\nabla_- f|\big(\eta(t)\big)$$

holds. In particular, the curve η has a unit speed. Here we abbreviated

$$\nabla_- f\big(\eta(t)\big)/|\nabla_- f|\big(\eta(t)\big) = (\alpha, 1) \in C_{\eta(t)}[\mathcal{P}^*(X)],$$

where $\nabla_- f(\eta(t)) = (\alpha, |\nabla_- f|(\eta(t))) \in C_{\eta(t)}[\mathcal{P}^*(X)]$ (see (4.11)).

(iii) *If $l < \infty$, then η can be extended to $\eta : [0, l] \longrightarrow \mathcal{P}^*(X)$ as a gradient-like curve. More precisely, the limit $\eta(l) := \lim_{t \to l} \eta(t)$ exists and it satisfies*

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

Proof. (i) The inequality (5.3) 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

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

Here we remark that

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

Moreover, the lower semi-continuity of $|\nabla_- f|$ (Lemma 4.5) 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 (5.4).

(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, 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 $\{\varepsilon_i\}_{i\in\mathbb{N}} \subset (0,\infty)$ 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 the proof of Lemma 4.3 (using (5.4) and (5.5)), the sequence $\{v_i\}_{i\in\mathbb{N}}$ is a Cauchy sequence and converges to $\nabla_- f(\eta(t))/|\nabla_- f|(\eta(t))$.

(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

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

We suppose that there is some $\varepsilon > 0$ 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

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

$\square$

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

Lemma 5.5 *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 4.5), we find $C, r > 0$ such that we have $|\nabla_- f|(\nu) \geq C$ for all $\nu \in \overline{B}(\mu, r) \cap \mathcal{P}^*(X)$, just as the assumption in Lemma 5.1. In particular, $U[\omega] \cap \overline{B}(\mu, r) = \emptyset$. Given $k \in \mathbb{N}$, by applying Lemma 5.1 repeatedly, we can choose a maximal sequence $\{\mu_i^k\}_{i=0}^{N(k)} \subset \overline{B}(\mu, r) \cap \mathcal{P}^*(X)$ 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 $i = 1, 2, \ldots, N(k)$. Note that we know $c_{i-1}^k \in [Ck^{-1}, f(\mu_{i-1}^k) - \omega]$ by Corollary 5.2. We also remark that the maximality of the sequence means that

$$\begin{aligned}
c_{N(k)}^k &> \left(\inf_{\overline{B}(\mu_{N(k)}^k, k^{-1}) \cap \mathcal{P}^*(X)} |\nabla_- f| \right) \cdot \{r - d_2^W(\mu, \mu_{N(k)}^k)\} \\
&= c_{N(k)}^k k \cdot \{r - d_2^W(\mu, \mu_{N(k)}^k)\},
\end{aligned}$$

and hence $d_2^W(\mu, \mu_{N(k)}^k) > r - k^{-1}$. Set $a_0^k := 0$ and $a_i^k := \sum_{j=1}^i d_2^W(\mu_{j-1}^k, \mu_j^k)$ for $i = 1, 2, \ldots, N(k)$. Note that $a_{N(k)}^k \geq d_2^W(\mu, \mu_{N(k)}^k) \geq r - k^{-1}$. We define $A_k := \{a_0^k, a_1^k, \ldots, a_{N(k)}^k\} \cap [0, r]$ and a map $\eta_k : A_k \longrightarrow \overline{B}(\mu, r) \cap U[f(\mu)]$ by $\eta_k(a_i^k) := \mu_i^k$. Note that every η_k is 1-Lipschitz and recall that $\overline{B}(\mu, r) \cap U[f(\mu)]$ is compact. Thus the

Arzela-Ascoli theorem yields that a subsequence of $\{\eta_k\}_{k\in\mathbb{N}}$ (again denoted by $\{\eta_k\}_{k\in\mathbb{N}}$) converges uniformly to a 1-Lipschitz curve $\eta : [0, r] \longrightarrow \overline{B}(\mu, r) \cap U[f(\mu)] \subset \mathcal{P}^*(X)$.

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 $d_2^W(\eta_k(a_i^k), \eta(a_i^k)) \le \varepsilon$ for all $a_i^k \in A_k$. Now we fix $a \in (0, r]$ and take $i(k)$ satisfying $a_{i(k)-1}^k < a \le a_{i(k)}^k$. Then we have, by the construction of $\mu_{i(k)}^k$,

$$\begin{aligned}
&f\bigl(\eta_k(a_{i(k)}^k)\bigr) = f(\mu_{i(k)}^k)\\
&= f(\mu) - k^{-1}\sum_{j=1}^{i(k)} \inf_{\overline{B}(\mu_{j-1}^k, k^{-1})\cap\mathcal{P}^*(X)} |\nabla_- f|\\
&\le f(\mu) - \sum_{j=1}^{i(k)} \left\{(a_j^k - a_{j-1}^k)\cdot \inf_{\overline{B}(\eta(a_{j-1}^k), k^{-1}+\varepsilon)\cap\mathcal{P}^*(X)} |\nabla_- f|\right\}\\
&\le f(\mu) - \int_0^{a_{i(k)}^k} \inf_{\overline{B}(\eta(t), 2k^{-1}+\varepsilon)\cap\mathcal{P}^*(X)} |\nabla_- f|\,dt\\
&\le f(\mu) - \int_0^{a} \inf_{\overline{B}(\eta(t), 2k^{-1}+\varepsilon)\cap\mathcal{P}^*(X)} |\nabla_- f|\,dt.
\end{aligned}$$

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

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

Hence we obtain

$$\begin{aligned}
f\bigl(\eta(a)\bigr) &\le \liminf_{k\to\infty} f\bigl(\eta_k(a_{i(k)}^k)\bigr) \le \limsup_{k\to\infty} f\bigl(\eta_k(a_{i(k)}^k)\bigr)\\
&\le f(\mu) - \int_0^a |\nabla_- f|\bigl(\eta(t)\bigr)\,dt.
\end{aligned}$$

Moreover, as $|\nabla_- f|$ is an upper gradient of f by Lemma 4.5, we find

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

and hence

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

Therefore we complete the proof by setting $l = r$. $\square$

As a by-product of the proof of Lemma 5.5, we obtain the K-convexity of $f \circ \eta$. We first recall an easily proved lemma.

Lemma 5.6 *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*

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

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

Proposition 5.7 *Let $\eta : [0, l] \longrightarrow \mathcal{P}^*(X)$ be the gradient-like curve constructed in Lemma 5.5. 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 5.5. Fix $k \in \mathbb{N}$, 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 a 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

$$\lim_{\varepsilon \to 0+} \frac{h_k(a^k_i) - h_k(a^k_i - \varepsilon)}{\varepsilon} \leq \lim_{\varepsilon \to 0+} \frac{h_k(a^k_i + \varepsilon) - h_k(a^k_i)}{\varepsilon},$$

which is clearly equivalent to

$$\lim_{\varepsilon \to 0+} \frac{f(\eta_k(a^k_i)) - f(\eta_k(a^k_i - \varepsilon))}{\varepsilon} \leq \lim_{\varepsilon \to 0+} \frac{f(\eta_k(a^k_i + \varepsilon)) - f(\eta_k(a^k_i))}{\varepsilon}, \tag{5.7}$$

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

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

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

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

It follows from the K-convexity of f that

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

However, combining this with our assumption (5.8), we find

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

This implies that $f(\beta_\varepsilon(1/2)) < f(\eta_k(a_i^k)) = f(\mu_i^k)$ holds for some small $\varepsilon > 0$, and it contradicts to the choice of μ_i^k. Thus we have (5.7) and hence h_k is convex on I_k by Lemma 5.6. Recall that, as we saw in (5.6),

$$h(a) = \lim_{k\to\infty} h_k\big(a_{i(k)}^k\big),$$

where $i(k)$ is such that $a_{i(k)-1}^k < a \leq a_{i(k)}^k$. Therefore, by taking a limit as k goes to the infinity, we obtain the convexity of h. $\Box$

The gradient curve is defined by using the gradient vector (4.11) as follows.

Definition 5.8 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 differentiable (see a paragraph preceding Lemma 3.7) 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 5.5.

Theorem 5.9 (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 of f. If $|\nabla_- f|(\mu) > 0$, then, by Lemma 5.5, we have a gradient-like curve $\eta : [0,l] \longrightarrow \mathcal{P}^*(X)$ with $\eta(0) = \mu$ and $l > 0$. Moreover, by applying Lemma 5.5 again at $\eta(l)$ and iterating this procedure repeatedly, we can extend η to a maximal gradient-like curve $\eta : [0,l') \longrightarrow \mathcal{P}^*(X)$. We remark that $|\nabla_- f|(\eta(t)) > 0$ for all $t \in [0,l')$ and that, if $l' < \infty$, then Lemma 5.4(iii) shows that $|\nabla_- f|(\eta(l')) = 0$ (otherwise, we can apply Lemma 5.5 again at $\eta(l')$ and it contradicts to the maximality of η). Note also that $f \circ \eta : [0,l') \longrightarrow \mathbb{R}$ is K-convex by (the proof of) Proposition 5.7, and hence $|\nabla_- f|(\eta(t)) < \infty$ for all $t \in (0,l')$.

As the 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 $l'' \le l'$ and $l'' = l'$ if $L < \infty$. Clearly ψ is monotone increasing and bijective, and also locally Lipschitz on $(0, L)$. Moreover, it follows from the lower semi-continuity of $|\nabla_- f| \circ \eta$ and the K-convexity of $f \circ \eta$ that, for every $t \in [0, L)$,

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

namely

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

The proof of the existence of ψ requires a trick since $|\nabla_- f| \circ \eta$ is possibly discontinuous, so we postpone it to the end of the subsection (Lemma 5.10).

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

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

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

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

Thus we have $l' < \infty$ as well as $|\nabla_- f|(\xi(L)) = |\nabla_- f|(\eta(l')) = 0$. Hence we set $\xi(t) := \xi(L)$ for $t \in (L, \infty)$ and it gives a complete gradient curve. $\square$

Lemma 5.10 *Let $h : [0, l) \longrightarrow \mathbb{R}$ be a monotone decreasing and K-convex function with $0 < l \le \infty$. Set*

$$g(t) := \lim_{\varepsilon \to 0+} \frac{h(t) - h(t+\varepsilon)}{\varepsilon} \in [0, \infty]$$

for $t \in [0, l)$ and suppose that g is positive and lower semi-continuous. Then there exists a function $\psi : [0, L) \longrightarrow [0, l')$ satisfying

$$\psi(t) = \int_0^t g\big(\psi(s)\big)\, ds \qquad (5.11)$$

for all $t \in [0, L)$, where $l' \le l$ and $l' = l$ if $L < \infty$.

Proof. We may assume $K = -1$ and $l < \infty$, and it suffices to construct $\psi : [0, \delta] \longrightarrow [0, l)$ satisfying (5.11) for a small $\delta > 0$.

We first treat the case of $g(0) < \infty$. By the (-1)-convexity of h, we observe that $g(t) \le g(0) + t$. Set

$$\delta := \log\left(\frac{l}{g(0)} + 1\right).$$

We fix a sufficiently large $k \in \mathbb{N}$ and put $t_0^k := 0$ and

$$t_i^k := t_{i-1}^k + k^{-1} g(t_{i-1}^k)$$

for $i = 1, 2, \dots, N(k)$. Here $N(k) \in \mathbb{N}$ is taken as the maximal number satisfying $N(k) \le k\delta$. Note that

$$\begin{aligned} t_{N(k)}^k &\le t_{N(k)-1}^k + k^{-1}\{g(0) + t_{N(k)-1}^k\} = k^{-1}g(0) + (1 + k^{-1}) t_{N(k)-1}^k \\ &\le \cdots \le k^{-1} g(0)\{1 + (1 + k^{-1}) + \cdots + (1 + k^{-1})^{N(k)-1}\} \\ &= k^{-1} g(0) \frac{(1 + k^{-1})^{N(k)} - 1}{k^{-1}} = g(0)\{(1 + k^{-1})^{N(k)} - 1\}. \end{aligned}$$

By $N(k) \le k\delta$, $(1 + k^{-1})^k \le e$ and the choice of δ, we find

$$t_{N(k)}^k \le g(0)\{(1 + k^{-1})^{k\delta} - 1\} \le g(0)(e^\delta - 1) \le l.$$

Thus t_i^k's are well-defined. On one hand, for any $t \in [t_{i-1}^k, t_i^k]$ with $i = 1, 2, \dots, N(k) - 1$, we have

$$\begin{aligned} g(t) &\ge g(t_i^k) - (t_i^k - t) \ge g(t_i^k) - (t_i^k - t_{i-1}^k) \\ &= k(t_{i+1}^k - t_i^k) - k^{-1} g(t_{i-1}^k) \ge k(t_{i+1}^k - t_i^k) - k^{-1}\{g(0) + l\}. \end{aligned} \tag{5.12}$$

On the other hand, for any $t \in [t_{i-1}^k, t_i^k]$ with $i = 1, 2, \dots, N(k)$, it holds that

$$\begin{aligned} g(t) &\le g(t_{i-1}^k) + (t - t_{i-1}^k) \le g(t_{i-1}^k) + (t_i^k - t_{i-1}^k) \\ &= k(t_i^k - t_{i-1}^k) + k^{-1} g(t_{i-1}^k) \le k(t_i^k - t_{i-1}^k) + k^{-1}\{g(0) + l\}. \end{aligned} \tag{5.13}$$

Define a function $\psi^k : [0, k^{-1}N(k)] \longrightarrow [0, \infty)$ by

$$\psi^k\big((i-1)k^{-1} + \lambda k^{-1}\big) := t_{i-1}^k + \lambda(t_i^k - t_{i-1}^k)$$

for $i = 1, 2, \dots, N(k)$ and $\lambda \in [0, 1]$. Then we see that $\psi^k(0) = 0$ and, for all $s, t \in [0, k^{-1}N(k)]$,

$$|\psi^k(s) - \psi^k(t)| \le \sup_{\tau \in [0, l]} g(\tau) \cdot |s - t| \le \{g(0) + l\}|s - t|.$$

Hence the Arzela-Ascoli theorem yields a subsequence of $\{\psi^k\}_k$ which converges uniformly to a $(g(0) + l)$-Lipschitz function $\psi : [0, \delta] \longrightarrow [0, \infty)$. We again denote such a convergent sequence by $\{\psi^k\}_k$. The inequality (5.12) deduces that

$$\begin{aligned}\psi(\delta) &= \lim_{k\to\infty} \psi^k\big(k^{-1}N(k)\big) = \lim_{k\to\infty} t^k_{N(k)} = \lim_{k\to\infty} k^{-1}\sum_{i=1}^{N(k)} k(t^k_i - t^k_{i-1})\\ &\le \lim_{k\to\infty}\left\{(t^k_1 - t^k_0) + \int_0^{k^{-1}\{N(k)-1\}} g\big(\psi^k(s)\big)\,ds\right\}\\ &= \lim_{k\to\infty}\int_0^{k^{-1}N(k)} g\big(\psi^k(s)\big)\,ds.\end{aligned}$$

Since g is positive and lower semi-continuous, by replacing l with $l/2$ if necessary, we have $c := \inf_{\tau\in[0,l]} g(\tau) > 0$. We may assume $\sup_{t\in[0,k^{-1}N(k)]}|\psi^k(t)-\psi(t)| \le ck^{-1}$. Then we see, for any $s \in (k^{-1}, k^{-1}N(k))$,

$$\psi^k(s) \ge \psi^k(s-k^{-1}) + ck^{-1} \ge \psi(s-k^{-1}),$$

and hence

$$\begin{aligned}&\lim_{k\to\infty}\int_0^{k^{-1}N(k)} g\big(\psi^k(s)\big)\,ds\\ &\le \lim_{k\to\infty}\int_{k^{-1}}^{k^{-1}N(k)} \big\{g\big(\psi(s-k^{-1})\big) + \psi^k(s) - \psi(s-k^{-1})\big\}\,ds\\ &= \int_0^{\delta} g\big(\psi(s)\big)\,ds.\end{aligned}$$

Similarly, (5.13) implies $\psi(\delta) \ge \int_0^\delta g(\psi(s))\,ds$. Therefore we obtain

$$\psi(\delta) = \int_0^\delta g(\psi(s))\,ds,$$

and (5.11) for a general $t \in [0,\delta)$ can be derived in the same manner.

If $g(0)=\infty$, then it follows from the lower semi-continuity of g that $\inf_{t\in[0,\varepsilon]} g(t) > 0$ holds for a sufficiently small $\varepsilon > 0$. Note also that $g(t) < \infty$ for all $t > 0$. Due to the first part of the proof, we find a solution $\psi_0 : [0, L_0] \longrightarrow [0, l-\varepsilon)$ of (5.11) by replacing g with $g_0(t) := g(t+\varepsilon)$, $t \in [0, l-\varepsilon)$. Similarly, for each $i \in \mathbb{N}$, we have a solution $\psi_i : [0, L_i] \longrightarrow [0, 2^{-i}\varepsilon]$ of (5.11) for $g_i(t) := g(t + 2^{-i}\varepsilon)$, $t \in [0, 2^{-i}\varepsilon]$. We remark that $L_i \le \{\inf_{t\in[0,\varepsilon]} g(t)\}^{-1}\cdot 2^{-i}\varepsilon$, and hence $\sum_{i=1}^\infty L_i < \infty$. Define $\psi : [0, \sum_{i=0}^\infty L_i) \longrightarrow [0,l)$ by

$$\begin{aligned}\psi(t) &:= \psi_j\left(t - \sum_{i=j+1}^\infty L_i\right) + \sum_{i=j+1}^\infty \psi_i(L_i) = \psi_j\left(t - \sum_{i=j+1}^\infty L_i\right) + \sum_{i=j+1}^\infty 2^{-i}\varepsilon\\ &= \psi_j\left(t - \sum_{i=j+1}^\infty L_i\right) + 2^{-j}\varepsilon\end{aligned}$$

for $t \in [\sum_{i=j+1}^\infty L_i, \sum_{i=j}^\infty L_i]$, $j \in \mathbb{N}\cup\{0\}$. This gives a required function and completes the proof. $\square$

5.2 Uniqueness and contraction

The K-convexity of f implies a contraction property of the gradient flow.

Theorem 5.11 (Uniqueness and contraction) *Let (X, d) be a compact Alexandrov space of curvature bounded below, 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*

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

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 $t \in (0, \infty)$. 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 3.7, 4.3 and (5.9) that

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

Furthermore, by Lemma 4.1, we find

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

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

By our construction of a gradient curve through a gradient-like curve, we obtain several extra estimates. Compare these with [JKO] and [V1, Section 8.4].

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

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

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

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

$$d_2^W\bigl(\xi(s), \xi(t)\bigr)^2 \le (t - s)\{f\bigl(\xi(s)\bigr) - f\bigl(\xi(t)\bigr)\}.$$

Proof. The uniqueness in Theorem 5.11 means that ξ coincides with the gradient curve given in the proof of Theorem 5.9. Then (i) has been already observed in the proof of Theorem 5.9 (see (5.10)). Moreover, (ii) follows from

$$\begin{aligned} d_2^W\big(\xi(s),\xi(t)\big)^2 &\le \bigg(\int_s^t |\nabla_- f|\big(\xi(\tau)\big)\,d\tau\bigg)^2 \le (t-s)\int_s^t |\nabla_- f|\big(\xi(\tau)\big)^2\,d\tau \\ &= (t-s)\{f\big(\xi(s)\big)-f\big(\xi(t)\big)\}. \end{aligned}$$

□

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 a unique gradient curve starting from μ. The contraction property (Theorem 5.11) allows us to extend the gradient flow G to the closure $\mathcal{P}^*(X)^-$ of $\mathcal{P}^*(X)$.

Corollary 5.13 (Extension to $\mathcal{P}^*(X)^-$) *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 $s,t\ge 0$.

Remark 5.14 We mention that the existence theorem (Theorem 5.9) essentially follows from [AGS, Theorem 2.3.3]. However, Theorem 5.11 is not a consequence of [AGS, Theorem 4.0.4] because the distance function d_2^W on $\mathcal{P}(X)$ is by no means convex (consider, for example, spheres).

6 Gradient flows of the free energy and the Fokker-Planck equation

In this final section, we study the Riemannian case and verify that our gradient flow of the free energy coincides with the solution of the linear Fokker-Planck equation, as was heuristically shown by Jordan, Kinderlehrer and Otto [JKO] in the Euclidean setting. In particular, our gradient flow of the relative entropy produces the solution of the heat equation and hence, if it starts from a Dirac measure, then the gradient flow describes the heat kernel.

Before proceeding it, we first observe the compatibility between our gradient flow and the gradient flow considered in [JKO] for general Alexandrov spaces.

6.1 Another characterization of gradient flows

Let (X,d) be a compact Alexandrov space of curvature bounded below and $f : \mathcal{P}(X) \longrightarrow (-\infty,\infty]$ be a function satisfying (4.1). Fix $\mu\in\mathcal{P}^*(X)$ and, for each

$\tau > 0$, take $\nu_\tau \in \mathcal{P}^*(X)$ attaining the infimum of

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

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

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

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

Lemma 6.1 *If $|\nabla_- f|(\mu) < \infty$, then the sequence $\{v_\tau\}_{\tau>0} \subset C'_\mu[\mathcal{P}^*(X)]$ converges to $\nabla_- f(\mu)$ as τ tends to zero.*

Proof. We first consider the case of $|\nabla_- f|(\mu) = 0$. By the choice of ν_τ, we see

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

This implies

$$\limsup_{\tau\to 0+} \frac{t_\tau}{2\tau} \le \limsup_{\tau\to 0+} \frac{f(\mu) - f(\nu_\tau)}{d_2^W(\mu,\nu_\tau)} = 0.$$

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

Next we suppose $|\nabla_- f|(\mu) \in (0,\infty)$ and put $\nabla_- f(\mu) = (\alpha, t)$. 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 observe, for each $i \in \mathbb{N}$,

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

by the choice of ν_τ. By letting i diverge to the infinity, it yields

$$\liminf_{\tau\to 0+} \frac{f(\mu) - f(\nu_\tau)}{t_\tau} \ge |\nabla_- f|(\mu) = t.$$

Thus the same discussion as in the proof of Lemma 4.3 (by using Lemma 4.2) yields that $\{\beta_\tau\}_{\tau>0}$ is a Cauchy sequence and converges to α as τ goes to zero.

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

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

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

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

By combining these, it holds that

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

Thus we have

$$\begin{aligned}\left(\limsup_{\tau\to 0+} \frac{t_\tau}{\tau}\right)^2 &\le \liminf_{i\to\infty}\left\{2t \cdot D'_\mu f(\alpha_i) + 2t \cdot \limsup_{\tau\to 0+}\frac{t_\tau}{\tau} + t^2\right\} \\ &= 2t \cdot \limsup_{\tau\to 0+}\frac{t_\tau}{\tau} - t^2,\end{aligned}$$

and hence

$$\left(\limsup_{\tau\to 0+}\frac{t_\tau}{\tau} - t\right)^2 \le 0.$$

Therefore we obtain $\limsup_{\tau\to 0+} t_\tau/\tau = t$ and also $\liminf_{\tau\to 0+} t_\tau/\tau = t$ similarly. These imply $\lim_{\tau\to 0+} t_\tau/\tau = t$ and complete the proof. □

Thanks to the Kantorovich-Rubinstein theorem (Theorem 2.6), we immediately deduce the following.

Lemma 6.2 *For any Lipschitz function $h : X \longrightarrow \mathbb{R}$, we have*

$$\lim_{\tau\to 0+}\frac{1}{\tau}\left\{\int_X h\,d\mu_\tau - \int_X h\,d\nu_\tau\right\} = 0,$$

where we put $\mu_\tau := G(\mu,\tau)$.

Proof. We can assume that h is 1-Lipschitz. Then Theorem 2.6 yields that

$$\left|\int_X h\,d\mu_\tau - \int_X h\,d\nu_\tau\right| \le d_1^W(\mu_\tau,\nu_\tau) \le d_2^W(\mu_\tau,\nu_\tau).$$

By Lemma 6.1, we obtain

$$\lim_{\tau\to 0+}\frac{d_2^W(\mu_\tau,\nu_\tau)}{\tau} = \sigma_\mu\big(\nabla_- f(\mu), \nabla_- f(\mu)\big) = 0.$$

□

6.2 The linear Fokker-Planck equation

Throughout this subsection, let (M, g) be a compact Riemannian manifold and m be the associated volume element.

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

$$\frac{\partial \rho_t}{\partial t} = \Delta \rho_t + \mathrm{div}(\rho_t \cdot \mathrm{grad}\, V), \tag{6.1}$$

where $\Delta = \mathrm{div} \circ \mathrm{grad}$ is the Laplace-Beltrami operator 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

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

(see (2.9)). Then Lemma 2.9 and Theorem 2.11 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 5.13).

In a particular case $V \equiv 0$, (6.1) corresponds to the *heat equation*:

$$\frac{\partial \rho_t}{\partial t} = \Delta \rho_t, \tag{6.2}$$

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

Theorem 6.3 *Let (M, g) be a compact Riemannian manifold equipped with a Riemannian volume element m. Fix $V \in C^\infty(M)$ and let f be the associated free energy* (2.9)*. Then, for any $\mu = \rho \cdot m \in \mathcal{P}^*(M)$, the gradient flow $\mu_t = G(\mu, t)$, $t \in [0, \infty)$, of f gives the unique solution of the linear Fokker-Planck equation* (6.1) *with the initial datum μ. More precisely, the function ρ_t given by $\mu_t = \rho_t \cdot m$ is the unique smooth solution of* (6.1) *on $M \times (0, \infty)$ such that ρ_t converges to ρ strongly in $L_1(M)$ as t goes to zero.*

Proof. The proof is performed along the line of [JKO, Theorem 5.1] (see also [V1, Subsection 8.4.2]). We shall prove that, for an arbitrary smooth function $h \in C^\infty(M \times \mathbb{R})$,

$$\begin{aligned} &\lim_{t \to 0+} \frac{1}{t} \left\{ \int_M h_t \, d\mu_t - \int_M h_0 \, d\mu \right\} \\ &= \int_M \left\{ \frac{\partial}{\partial t}\bigg|_{t=0} h + \Delta h_0 - \langle \mathrm{grad}\, h_0, \mathrm{grad}\, V \rangle \right\} d\mu, \end{aligned} \tag{6.3}$$

where we set $h_t := h(\cdot, t)$ for simplicity. For a small $t > 0$, take $\nu_t \in \mathcal{P}^*(M)$ satisfying

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

By McCann's theorem (Theorem 2.8), we have a Lipschitz function $\psi_t : M \longrightarrow \mathbb{R}$ such that the map $\Psi_t : M \longrightarrow M$ defined by $\Psi_t(x) := \exp_x[\operatorname{grad}\psi_t(x)]$ satisfies $(\Psi_t)_*\nu_t = \mu$ and

$$d_2^W(\nu_t, \mu)^2 = \int_M d_M\big(x, \Psi_t(x)\big)^2\, d\nu_t(x).$$

Note also that Lemma 6.2 implies

$$\begin{aligned}
&\lim_{t\to 0+} \frac{1}{t}\bigg\{ \int_M h_t\, d\mu_t - \int_M h_0\, d\mu \bigg\} \\
&= \lim_{t\to 0+} \frac{1}{t}\bigg\{ \int_M (h_t - h_0)\, d\mu_t + \int_M h_0\, d\mu_t - \int_M h_0\, d\mu \bigg\} \\
&= \int_M \frac{\partial}{\partial t}\bigg|_{t=0} h\, d\mu + \lim_{t\to 0+} \frac{1}{t}\bigg\{ \int_M h_0\, d\nu_t - \int_M h_0\, d\mu \bigg\}.
\end{aligned} \tag{6.4}$$

For a small $\varepsilon > 0$, we define a smooth map $\Phi_\varepsilon : M \longrightarrow M$ by $\Phi_\varepsilon(x) := \exp_x[\varepsilon \operatorname{grad} h_0(x)]$ and remark that Φ_ε is a diffeomorphism for a sufficiently small ε. Put $\tilde{\nu}_{t,\varepsilon} := (\Phi_\varepsilon)_*\nu_t$. Then the choice of ν_t implies that

$$f(\tilde{\nu}_{t,\varepsilon}) + \frac{d_2^W(\mu, \tilde{\nu}_{t,\varepsilon})^2}{2t} - f(\nu_t) - \frac{d_2^W(\mu, \nu_t)^2}{2t} \geq 0. \tag{6.5}$$

We first observe that, by the first variation formula,

$$\begin{aligned}
&\limsup_{\varepsilon\to 0+} \frac{d_2^W(\mu, \tilde{\nu}_{t,\varepsilon})^2 - d_2^W(\mu, \nu_t)^2}{\varepsilon} \\
&\leq \limsup_{\varepsilon\to 0+} \frac{1}{\varepsilon} \int_M \Big\{ d_M\big(\Psi_t(x), \Phi_\varepsilon(x)\big)^2 - d_M\big(\Psi_t(x), x\big)^2 \Big\}\, d\nu_t(x) \\
&= -\int_M 2\langle \operatorname{grad} h_0(x), \operatorname{grad}\psi_t(x)\rangle\, d\nu_t(x).
\end{aligned}$$

By the expansion and the compactness of M, there is a constant $C \geq 0$ depending on h_0 such that

$$h_0\big(\Psi_t(x)\big) \leq h_0(x) + \langle \operatorname{grad} h_0(x), \operatorname{grad}\psi_t(x)\rangle + C d_M\big(x, \Psi_t(x)\big)^2.$$

Therefore we obtain

$$\begin{aligned}
&\liminf_{t\to 0+} \frac{1}{2t} \limsup_{\varepsilon\to 0+} \frac{d_2^W(\mu, \tilde{\nu}_{t,\varepsilon})^2 - d_2^W(\mu, \nu_t)^2}{\varepsilon} \\
&\leq -\limsup_{t\to 0+} \frac{1}{t} \int_M \langle \operatorname{grad} h_0(x), \operatorname{grad}\psi_t(x)\rangle\, d\nu_t(x) \\
&\leq \liminf_{t\to 0+} \frac{1}{t} \bigg[\int_M \Big\{ h_0(x) - h_0\big(\Psi_t(x)\big) \Big\}\, d\nu_t(x) + C \int_M d_M\big(x, \Psi_t(x)\big)^2\, d\nu_t(x) \bigg] \\
&= \liminf_{t\to 0+} \frac{1}{t} \bigg\{ \int_M h_0\, d\nu_t - \int_M h_0\, d\mu \bigg\}.
\end{aligned} \tag{6.6}$$

In the last equality, we used the fact

$$\lim_{t\to 0+} \frac{1}{t} \int_M d_M\big(x, \Psi_t(x)\big)^2 \, d\nu_t(x) = \lim_{t\to 0+} \frac{1}{t} d_2^W(\nu_t, \mu)^2 = 0$$

which follows from Lemma 6.1.

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

$$\int_M \varphi \, d\tilde{\nu}_{t,\varepsilon} = \int_M \varphi\big(\Phi_\varepsilon(x)\big) \, d\nu_t(x) = \int_M \varphi\big(\Phi_\varepsilon(x)\big)\varsigma_t(x) \, dm(x).$$

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

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

Since $\varphi \in C^\infty(M)$ is arbitrary, these together imply that

$$\tilde{\varsigma}_{t,\varepsilon}\big(\Phi_\varepsilon(x)\big) \det[D\Phi_\varepsilon(x)] = \varsigma_t(x) \tag{6.7}$$

holds for a.e. $x \in M$. Combining this with the change of variables formula, we have

$$\begin{aligned}\mathrm{Ent}_m(\tilde{\nu}_{t,\varepsilon}) &= \int_M \tilde{\varsigma}_{t,\varepsilon}(y) \log \tilde{\varsigma}_{t,\varepsilon}(y) \, dm(y) \\ &= \int_M \tilde{\varsigma}_{t,\varepsilon}\big(\Phi_\varepsilon(x)\big) \log \tilde{\varsigma}_{t,\varepsilon}\big(\Phi_\varepsilon(x)\big) \det[D\Phi_\varepsilon(x)] \, dm(x) \\ &= \int_M \varsigma_t(x) \log\left(\frac{\varsigma_t(x)}{\det[D\Phi_\varepsilon(x)]}\right) dm(x) \\ &= \mathrm{Ent}_m(\nu_t) - \int_M \varsigma_t(x) \log\big(\det[D\Phi_\varepsilon(x)]\big) \, dm(x).\end{aligned}$$

Similarly, we observe

$$\begin{aligned}\int_M V \, d\tilde{\nu}_{t,\varepsilon} &= \int_M V(y)\tilde{\varsigma}_{t,\varepsilon}(y) \, dm(y) \\ &= \int_M V\big(\Phi_\varepsilon(x)\big)\tilde{\varsigma}_{t,\varepsilon}\big(\Phi_\varepsilon(x)\big) \det[D\Phi_\varepsilon(x)] \, dm(x) \\ &= \int_M V\big(\Phi_\varepsilon(x)\big)\varsigma_t(x) \, dm(x).\end{aligned}$$

Thus we see

$$
\begin{aligned}
&\lim_{\varepsilon\to 0+}\frac{1}{\varepsilon}\{f(\nu_t)-f(\tilde{\nu}_{t,\varepsilon})\}\\
&=\lim_{\varepsilon\to 0+}\frac{1}{\varepsilon}\{\mathrm{Ent}_m(\nu_t)-\mathrm{Ent}_m(\tilde{\nu}_{t,\varepsilon})\}+\lim_{\varepsilon\to 0+}\frac{1}{\varepsilon}\left\{\int_M V\,d\nu_t-\int_M V\,d\tilde{\nu}_{t,\varepsilon}\right\}\\
&=\lim_{\varepsilon\to 0+}\int_M\frac{1}{\varepsilon}\log\big(\det[D\Phi_\varepsilon(x)]\big)\,d\nu_t(x)+\lim_{\varepsilon\to 0+}\frac{1}{\varepsilon}\int_M\big\{V(x)-V\big(\Phi_\varepsilon(x)\big)\big\}\,d\nu_t(x)\\
&=\int_M\frac{\partial}{\partial\varepsilon}\bigg|_{\varepsilon=0+}\det[D\Phi_\varepsilon(x)]\,d\nu_t(x)-\int_M\langle\operatorname{grad}h_0,\operatorname{grad}V\rangle\,d\nu_t\\
&=\int_M\operatorname{trace}(\operatorname{Hess}h_0)\,d\nu_t-\int_M\langle\operatorname{grad}h_0,\operatorname{grad}V\rangle\,d\nu_t\\
&=\int_M\Delta h_0\,d\nu_t-\int_M\langle\operatorname{grad}h_0,\operatorname{grad}V\rangle\,d\nu_t.
\end{aligned}
$$

Since ν_t converges to μ weakly as t goes to zero, we find

$$
\lim_{t\to 0+}\lim_{\varepsilon\to 0+}\frac{1}{\varepsilon}\{f(\nu_t)-f(\tilde{\nu}_{t,\varepsilon})\}=\int_M\{\Delta h_0-\langle\operatorname{grad}h_0,\operatorname{grad}V\rangle\}\,d\mu. \tag{6.8}
$$

These four inequalities (6.4), (6.5), (6.6) and (6.8) together imply

$$
\begin{aligned}
&\liminf_{t\to 0+}\frac{1}{t}\left\{\int_M h_t\,d\mu_t-\int_M h_0\,d\mu\right\}\\
&\geq\int_M\left\{\frac{\partial}{\partial t}\bigg|_{t=0}h+\Delta h_0-\langle\operatorname{grad}h_0,\operatorname{grad}V\rangle\right\}d\mu.
\end{aligned}
$$

Moreover, the same inequality for $-h$ gives the reverse inequality (with $\limsup$ instead of $\liminf$). Therefore we obtain (6.3).

By integrating (6.3), we deduce that, for any $0\leq t_0<t_1$,

$$
\begin{aligned}
&\int_M h_{t_1}\,d\mu_{t_1}-\int_M h_{t_0}\,d\mu_{t_0}\\
&\quad=\int_{t_0}^{t_1}\left[\int_M\left\{\frac{\partial}{\partial t}\bigg|_{t=s}h+\Delta h_s-\langle\operatorname{grad}h_s,\operatorname{grad}V\rangle\right\}d\mu_s\right]ds.
\end{aligned}
$$

As in the proofs of (a), (b) and (c) in the proof of [JKO, Theorem 5.1], it implies that the function ρ_t is smooth on $M\times(0,\infty)$ and satisfies (6.1). Moreover, ρ_t converges to ρ strongly in $L_1(M)$ as t goes to zero, and such a solution is unique. See [JKO, Theorem 5.1] for more details. □

Applying Theorem 6.3 above to Dirac measures, we obtain the following characterization of the heat kernel.

Corollary 6.4 *Let (M,g) be a compact Riemannian manifold and $p: M\times M\times(0,\infty)\longrightarrow[0,\infty)$ be the associated heat kernel. Then, for any point $x\in M$ and $t\in(0,\infty)$, we have $G(\delta_x,t)=p(\cdot,x,t)\cdot m$, where δ_x stands for the Dirac measure at x.*

Proof. Put $\nu_s := p(\cdot, x, s) \cdot m \in \mathcal{P}^*(M)$ for $s > 0$. By Theorem 6.3, we deduce that $G(\nu_s, t) = p(\cdot, x, s+t) \cdot m$ for all t. Letting s tend to zero, we obtain $G(\delta_x, t) = p(\cdot, x, t) \cdot m$. $\square$

By virtue of Theorem 5.11, we obtain a contraction property of the heat kernel. See [RS] for more general results.

Corollary 6.5 *Let (M, g) be a compact Riemannian manifold and $p : M \times M \times (0, \infty) \longrightarrow [0, \infty)$ 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*

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

Remark 6.6 The heat kernel on Alexandrov spaces is constructed by Kuwae, Machigashira and Shioya [KMS] and the analogues of Corollaries 6.4 and 6.5 should hold true on Alexandrov spaces. However, its proof may be more involved. For instance, McCann's theorem (Theorem 2.8) is not yet generalized to Alexandrov spaces.

For $\mu, \nu \in \mathcal{P}^*(M)$, the K-convexity of the free energy f on a minimal geodesic between μ and ν implies

$$f(\mu) \leq f(\nu) + |\nabla_- f|(\mu) \cdot d_2^W(\mu, \nu) - \frac{K}{2} d_2^W(\mu, \nu)^2. \tag{6.9}$$

If $K > 0$, then we immediately deduce from (6.9) that

$$f(\mu) \leq f(\nu) + \frac{1}{2K} |\nabla_- f|(\mu)^2. \tag{6.10}$$

By putting $c_V := (e^{-V} \cdot m)(M)$ and $\mu = c_V^{-1} e^{-V} \cdot m$ in (6.9), we obtain the *Talagrand inequality*

$$\frac{K}{2} d_2^W(\mu, \nu) - \log c_V \leq f(\nu). \tag{6.11}$$

Moreover, it follows from Theorem 6.3 that, for any $\mu \in \mathcal{P}^*(M)$ and $t > 0$,

$$\begin{aligned}
|\nabla_- f|(\mu_t)^2 &= \lim_{\varepsilon \to 0+} \frac{1}{\varepsilon} \bigg\{ \int_M \rho_t (\log \rho_t + V)\, dm - \int_M \rho_{t+\varepsilon} (\log \rho_{t+\varepsilon} + V)\, dm \bigg\} \\
&= -\int_M \frac{\partial \rho_t}{\partial t} \cdot \{(\log \rho_t + V) + 1\}\, dm \\
&= -\int_M \big\{\Delta \rho_t + \mathrm{div}(\rho_t \cdot \mathrm{grad}\, V)\big\} \cdot \big\{(\log \rho_t + V) + 1\big\}\, dm \\
&= \int_M \langle \mathrm{grad}\, \rho_t + \rho_t \cdot \mathrm{grad}\, V, \mathrm{grad}(\log \rho_t + V) \rangle\, dm \\
&= \int_M \rho_t |\mathrm{grad}(\log \rho_t) + V|^2\, dm,
\end{aligned}$$

where we set $\mu_t = \rho_t \cdot m = G(\mu, t)$. Therefore the inequalities (6.9) and (6.10) yield the *HWI inequality*

$$f(\mu_t) \le f(\nu) + d_2^W(\mu_t, \nu) \cdot \left\{ \int_M \rho_t |\operatorname{grad}(\log \rho_t) + V|^2 \, dm \right\}^{1/2} - \frac{K}{2} d_2^W(\mu_t, \nu)^2 \tag{6.12}$$

and, if $K > 0$, the *logarithmic Sobolev inequality*

$$f(\mu_t) \le f(\nu) + \frac{1}{2K} \int_M \rho_t |\operatorname{grad}(\log \rho_t) + V|^2 \, dm \tag{6.13}$$

for $\mu_t = \rho_t \cdot m = G(\mu, t)$ with $\mu \in \mathcal{P}^*(M)$ and $t > 0$. These give simple (but partial) alternative proofs of theorems in [OV] (see also [LV1]).

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.

[Ch] J. Cheeger, *Differentiability of Lipschitz functions on metric measure spaces*, Geom. Funct. Anal. **9** (1999), 428–517.

[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.

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

[HK] J. Heinonen and P. Koskela, *Quasiconformal maps in metric spaces with controlled geometry*, Acta Math. **181** (1998), 1–61.

[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, *Ricci curvature for metric-measure spaces via optimal transport*, to appear in Ann. of Math.

[LV2] J. Lott and C. Villani, *Weak curvature conditions and Poincaré inequalities*, preprint, 2005: arXiv:math.DG/0506481.

[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*, preprint, 2005.

[Oh2] S. Ohta, *On the measure contraction property of metric measure spaces*, preprint, 2005.

[Oh3] S. Ohta, *Markov type of Alexandrov spaces of nonnegative curvature*, preprint, 2006.

[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*, 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.

[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, in preparation.