# First variation formula in Wasserstein spaces over compact Alexandrov spaces

Nicola Gigli*and Shin-ichi Ohta†

January 29, 2010

**Abstract**

We extend results proven by the second author ([Oh]) for nonnegatively curved Alexandrov spaces to general compact Alexandrov spaces $X$ with curvature bounded below: the gradient flow of a geodesically convex functional on the quadratic Wasserstein space $(\mathcal{P}(X), W_2)$ satisfies the evolution variational inequality. Moreover, the gradient flow enjoys uniqueness and contractivity. These results are obtained by proving a first variation formula for the Wasserstein distance.

## 1 Introduction

This paper should be considered as an addendum to [Oh] of the second author. In [Oh], it is studied the quadratic Wasserstein space $(\mathcal{P}(X), W_2)$ built over a compact Alexandrov space $X$ with curvature bounded below, and proven the existence of Euclidean tangent cones (see also [Gi]). This result is particularly interesting for Alexandrov spaces with a negative curvature bound, as it is known that Wasserstein spaces built over them do *not* admit lower curvature bounds in the sense of Alexandrov.

The existence of such tangent cones has then been used in [Oh] to perform studies of gradient flows of geodesically convex functionals on $(\mathcal{P}(X), W_2)$. In particular, existence of gradient flows of such functionals has been proven via an approach inspired by [Ly] (which differs from the 'purely metric' approach in [AGS] without using tangent cones). For technical reasons, however, uniqueness and contraction of such gradient flows have been obtained only in the case where $X$ has the nonnegative curvature.

In this paper we extend these latter results to general Alexandrov spaces with curvature bounded below possibly by a negative value (Theorem 4.2). A key tool in our approach is a first variation formula for the Wasserstein distance (Theorem 3.4) from which it also easily follows that gradient flows satisfy the evolution variational inequality (Proposition 4.1). See Remarks 3.7, 4.5 for the difference from the argument in [Oh].

Section 2 is devoted to recalling known results on the geometric structure of and gradient flows in $(\mathcal{P}(X), W_2)$. We show the first variation formula in Section 3, and use it to study gradient flows in Section 4.

*Institut für Angewandte Mathematik, Universität Bonn, Endenicher Allee 60, 53115 Bonn, Germany (nicolagigli@googlemail.com)

†Department of Mathematics, Kyoto University, Kyoto 606-8502, Japan (sohta@math.kyoto-u.ac.jp);
Supported in part by the Grant-in-Aid for Young Scientists (B) 20740036.

# 2 Preliminaries

## 2.1 Wasserstein spaces over compact Alexandrov spaces

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. We say that $\gamma$ is *minimal* if it is globally minimizing (i.e., $d(\gamma(s), \gamma(t)) = (|s-t|/l) \cdot d(\gamma(0), \gamma(l))$ for all $s, t \in [0, l]$). If any two points in $X$ are joined by some minimal geodesic, then $X$ is called a *geodesic space*.

Throughout the article, $(X, d)$ will be a compact *Alexandrov space of curvature* $\geq -1$. By this we mean that $(X, d)$ is a geodesic space such that every triangle in $X$ is thicker than a geodesic triangle (with the same side lengths) in the hyperbolic plane $\mathbb{H}^2(-1)$ of constant sectional curvature $-1$ (see [Oh] for the detailed definition, [BGP], [OtSh] and [BBI] for the basic theory). We remark that $(X, d)$ can be infinite-dimensional, so that its local structure may be very wild.

Denote by $\mathcal{P}(X)$ the set of all Borel probability measures on $X$. Given $\mu, \nu \in \mathcal{P}(X)$, we consider the $L^2$-*Wasserstein distance*

$$W_2(\mu, \nu) := \inf_{\pi} \left\{ \int_{X \times X} d(x, y)^2 \, d\pi(x, y) \right\}^{1/2},$$

where $\pi \in \mathcal{P}(X \times X)$ runs over all *couplings* of $\mu$ and $\nu$. Note that $W_2(\mu, \nu)$ is finite and $(\mathcal{P}(X), W_2)$ is compact as $X$ is assumed to be compact. We refer to [AGS] and [Vi] for more on Wasserstein geometry and optimal transport theory.

It is known that, if $(X, d)$ has nonnegative curvature, then so does $(\mathcal{P}(X), W_2)$ ([St2, Proposition 2.10], [LV, Theorem A.8]). Although the analogues implication is false for negative curvature bounds (because it is not a scaling invariant condition, [St2, Proposition 2.10]), we obtain the following generalization of the 2-*uniform smoothness* in Banach space theory.

**Proposition 2.1** ([Oh, Proposition 3.1, Lemma 3.3], [Sa]) *For all $\mu, \nu, \omega \in \mathcal{P}(X)$, all minimal geodesics $\alpha : [0, 1] \longrightarrow \mathcal{P}(X)$ from $\nu$ to $\omega$ and for all $\tau \in [0, 1]$, we have*

$$W_2\big(\mu, \alpha(\tau)\big)^2 \geq (1-\tau) W_2(\mu, \nu)^2 + \tau W_2(\mu, \omega)^2 - S^2 (1-\tau) \tau W_2(\nu, \omega)^2, \qquad (1)$$

*where $S = \sqrt{1 + (\operatorname{diam} X)^2}$.*

In fact, the 2-uniform smoothness (1) holds in $(X, d)$ and descends to $(\mathcal{P}(X), W_2)$ with the same constant $S$. Although $(\mathcal{P}(X), W_2)$ is not an Alexandrov space, it is possible to show the following (see also [Oh, Theorem 3.6]):

**Theorem 2.2** ([Gi, Theorem 3.4, Remark 3.5]) *Given $\mu \in \mathcal{P}(X)$ and unit speed geodesics $\alpha, \beta : [0, \delta) \longrightarrow \mathcal{P}(X)$ with $\alpha(0) = \beta(0) = \mu$, the joint limit*

$$\lim_{s,t \to 0} \frac{s^2 + t^2 - W_2(\alpha(s), \beta(t))^2}{2st}$$

*exists.*

Theorem 2.2 means that $(\mathcal{P}(X), W_2)$ looks like a Riemannian space (rather than a Finsler space), and we can investigate its infinitesimal structure according to the theory of Alexandrov spaces. For $\mu \in \mathcal{P}(X)$, denote by $\Sigma'_\mu[\mathcal{P}(X)]$ the set of all (nontrivial) unit speed minimal geodesics emanating from $\mu$. Given $\alpha, \beta \in \Sigma'_\mu[\mathcal{P}(X)]$, Theorem 2.2 verifies that the *angle*

$$\angle_\mu(\alpha, \beta) := \arccos\left( \lim_{s,t \to 0} \frac{s^2 + t^2 - W_2(\alpha(s), \beta(t))^2}{2st} \right) \in [0, \pi]$$

is well-defined and provides an appropriate (pseudo-)distance structure of $\Sigma'_\mu[\mathcal{P}(X)]$. We define the *space of directions* $(\Sigma_\mu[\mathcal{P}(X)], \angle_\mu)$ as the completion of $(\Sigma'_\mu[\mathcal{P}(X)]/\sim, \angle_\mu)$, where $\alpha \sim \beta$ holds if $\angle_\mu(\alpha, \beta) = 0$. Then the *tangent cone* $(C_\mu[\mathcal{P}(X)], \sigma_\mu)$ is defined as the Euclidean cone over $(\Sigma_\mu[\mathcal{P}(X)], \angle_\mu)$:

$$C_\mu[\mathcal{P}(X)] := \Big(\Sigma_\mu[\mathcal{P}(X)] \times [0, \infty)\Big) \Big/ \Big(\Sigma_\mu[\mathcal{P}(X)] \times \{0\}\Big),$$
$$\sigma_\mu\big((\alpha, s), (\beta, t)\big) := \sqrt{s^2 + t^2 - 2st \cos \angle_\mu(\alpha, \beta)}.$$

We denote the origin of $C_\mu[\mathcal{P}(X)]$ by $o_\mu$ and define

$$\langle (\alpha, s), (\beta, t) \rangle_\mu := 2st \cos \angle_\mu(\alpha, \beta), \qquad |(\alpha, s)|_\mu := s = \sigma_\mu\big(o_\mu, (\alpha, s)\big)$$

for $(\alpha, s), (\beta, t) \in C_\mu[\mathcal{P}(X)]$. The subscript $\mu$ will be omitted if the space under consideration is clearly understood. We sometimes abbreviate like $t \cdot (\alpha, s) := (\alpha, st)$ and identify $\alpha \in \Sigma_\mu[\mathcal{P}(X)]$ with $(\alpha, 1) \in C_\mu[\mathcal{P}(X)]$.

Using this infinitesimal structure, we introduce a class of 'differentiable curves'.

**Definition 2.3** (Right differentiability) We say that a curve $\xi : [0, l) \longrightarrow \mathcal{P}(X)$ is *right differentiable* at $t \in [0, l)$ if there is $\mathbf{v} \in C_{\xi(t)}[\mathcal{P}(X)]$ such that, for any sequences $\{\varepsilon_i\}_{i \in \mathbb{N}}$ of positive numbers tending to zero and $\{\alpha_i\}_{i \in \mathbb{N}}$ of unit speed minimal geodesics from $\xi(t)$ to $\xi(t + \varepsilon_i)$, the sequence $\{(\alpha_i, W_2(\xi(t), \xi(t + \varepsilon_i))/\varepsilon_i)\}_{i \in \mathbb{N}} \subset C_{\xi(t)}[\mathcal{P}(X)]$ converges to $\mathbf{v}$. Such $\mathbf{v}$ is clearly unique if it exists, and then we write $\dot{\xi}(t) = \mathbf{v}$.

In particular, we have $\lim_{\varepsilon \to 0} W_2(\xi(t), \xi(t + \varepsilon))/\varepsilon = |\dot{\xi}(t)|_{\xi(t)}$. We also remark that every minimal geodesic $\alpha : [0, l] \longrightarrow \mathcal{P}(X)$ is right differentiable at all $t \in [0, l)$. This is because Alexandrov spaces are known to satisfy the non-branching property, which is inherited by $(\mathcal{P}(X), W_2)$ (cf. [Vi, Corollary 7.32]). Therefore $\alpha|_{[0,t]}$ is a unique minimal geodesic between $\alpha(0)$ and $\alpha(t)$ for all $t \in (0, l)$.

## 2.2 Gradient flows in Wasserstein spaces

The contents of this subsection will come into play in Section 4. The readers interested only in the first variation formula can skip to Section 3.

Consider a lower semi-continuous function $f : \mathcal{P}(X) \longrightarrow (-\infty, +\infty]$ which is *$K$-convex* for some $K \in \mathbb{R}$ in the sense that

$$f\big(\alpha(\tau)\big) \le (1 - \tau) f\big(\alpha(0)\big) + \tau f\big(\alpha(1)\big) - \frac{K}{2}(1 - \tau)\tau W_2\big(\alpha(0), \alpha(1)\big)^2 \qquad (2)$$

holds along all minimal geodesics $\alpha : [0,1] \longrightarrow \mathcal{P}(X)$ and all $\tau \in [0,1]$. We also suppose that $f$ is not identically $+\infty$, and define

$$\mathcal{P}^*(X) := \{\mu \in \mathcal{P}(X) \,|\, f(\mu) < \infty\}.$$

The $K$-convexity guarantees that minimal geodesics between points in $\mathcal{P}^*(X)$ are again included in $\mathcal{P}^*(X)$, and hence it makes sense to consider $\Sigma_\mu[\mathcal{P}^*(X)]$ as well as $C_\mu[\mathcal{P}^*(X)]$ for $\mu \in \mathcal{P}^*(X)$.

Given $\mu \in \mathcal{P}^*(X)$ and $\alpha \in \Sigma_\mu[\mathcal{P}^*(X)]$, we set

$$D_\mu f(\alpha) := \liminf_{\beta \to \alpha} \lim_{t \to 0} \frac{f(\beta(t)) - f(\mu)}{t},$$

where $\beta : [0,\delta) \longrightarrow \mathcal{P}^*(X)$ is a unit speed geodesic and the convergence $\beta \to \alpha$ is with respect to $\angle_\mu$. Define the *absolute gradient* (called the *local slope* in [AGS]) of $f$ at $\mu \in \mathcal{P}^*(X)$ by

$$|\nabla_{\!-} f|(\mu) := \max\left\{0, \limsup_{\nu \to \mu} \frac{f(\mu) - f(\nu)}{W_2(\mu,\nu)}\right\}.$$

Note that $D_\mu f(\alpha) \le |\nabla_{\!-} f|(\mu)$ for any $\alpha \in \Sigma_\mu[\mathcal{P}^*(X)]$. According to the argument in [PP] and [Ly], we find a negative gradient vector of $f$ at each point in $\mathcal{P}^*(X)$ with finite absolute gradient.

**Lemma 2.4** ([Oh, Lemma 4.2]) *For each $\mu \in \mathcal{P}^*(X)$ with $0 < |\nabla_{\!-} f|(\mu) < \infty$, there exists unique $\alpha \in \Sigma_\mu[\mathcal{P}^*(X)]$ satisfying $D_\mu f(\alpha) = -|\nabla_{\!-} f|(\mu)$. Moreover, for any $\beta \in \Sigma_\mu[\mathcal{P}^*(X)]$, it holds that $D_\mu f(\beta) \ge -|\nabla_{\!-} f|(\mu)\langle \alpha, \beta \rangle_\mu$.*

The second assertion is regarded as a first variation formula for $f$. Using $\alpha$ in the above lemma, we define the *negative gradient vector* of $f$ at $\mu$ as

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

In the case of $|\nabla_{\!-} f|(\mu) = 0$, we simply set $\nabla_{\!-} f(\mu) := o_\mu$.

**Definition 2.5** (Gradient curves) A continuous curve $\xi : [0,l) \longrightarrow \mathcal{P}^*(X)$ which is locally Lipschitz continuous on $(0,l)$ is called a *gradient curve* of $f$ if $|\nabla_{\!-} f|(\xi(t)) < \infty$ for all $t \in (0,\infty)$ and if it is right differentiable with $\dot{\xi}(t) = \nabla_{\!-} f(\xi(t))$ at all $t \in [0,l)$ with $|\nabla_{\!-} f|(\xi(t)) < \infty$. We say that a gradient curve $\xi$ is *complete* if it is defined on $(0,\infty)$.

Again along the discussion in [PP] and [Ly], despite some technical difficulties as $(\mathcal{P}(X), W_2)$ is not an Alexandrov space, we can show the existence of complete gradient curves.

**Theorem 2.6** ([Oh, Theorem 5.11]) *For any $\mu \in \mathcal{P}^*(X)$, there exists a complete gradient curve $\xi : [0,\infty) \longrightarrow \mathcal{P}^*(X)$ of $f$ with $\xi(0) = \mu$.*

**Remark 2.7** Let us make a more detailed comment on the construction of gradient curves. The strategy in [Oh] (following [PP] and [Ly]) is that we first construct a unit speed curve $\eta$ with $(f \circ \eta)'(t) = -|\nabla_{\!-} f|(\eta(t))$ a.e. $t$, then an appropriate reparametrization of $\eta$ provides a gradient curve. Another way is the direct construction comprehensively discussed in [AGS]. In fact, a *generalized minimizing movement* $u : [0,\infty) \longrightarrow \mathcal{P}(X)$ is locally Lipschitz continuous on $(0,\infty)$ and satisfies

$$\lim_{\varepsilon\downarrow 0}\frac{f(u(t+\varepsilon))-f(u(t))}{\varepsilon} = -|\nabla_{\!-} f|\big(u(t)\big)^2 = -\bigg\{\lim_{\varepsilon\downarrow 0}\frac{W_2(u(t),u(t+\varepsilon))}{\varepsilon}\bigg\}^2$$

at every $t\in(0,\infty)$ ([AGS, Theorem 2.4.15]), and therefore the discussion as in [Oh, Lemma 5.5] shows that $u$ is a gradient curve in the sense of Definition 2.5. Moreover, the uniqueness of gradient curves proved below (Theorem 4.2) ensures that both constructions give rise to the same curve.

# 3 A first variation formula

This section contains our main results. These are shown after a series of lemmas.

**Lemma 3.1** *For any minimal geodesic $\alpha : [0,t] \longrightarrow \mathcal{P}(X)$ and $\nu\in\mathcal{P}(X)$, we have*

$$\begin{aligned}&\frac{W_2(\alpha(t),\nu)^2 - W_2(\alpha(0),\nu)^2}{t}\\ &\le \liminf_{s\to 0}\frac{W_2(\alpha(s),\nu)^2 - W_2(\alpha(0),\nu)^2}{s} + \frac{S^2}{t}W_2\big(\alpha(0),\alpha(t)\big)^2,\end{aligned}$$

*where $S=\sqrt{1+(\operatorname{diam}X)^2}$.*

*Proof.* For $s\in(0,t)$, the 2-uniform smoothness (1) shows

$$\begin{aligned}&W_2\big(\alpha(s),\nu\big)^2\\ &\ge \Big(1-\frac{s}{t}\Big)W_2\big(\alpha(0),\nu\big)^2 + \frac{s}{t}W_2\big(\alpha(t),\nu\big)^2 - S^2\Big(1-\frac{s}{t}\Big)\frac{s}{t}W_2\big(\alpha(0),\alpha(t)\big)^2.\end{aligned}$$

Dividing both sides by $s$ yields

$$\begin{aligned}&\frac{W_2(\alpha(t),\nu)^2 - W_2(\alpha(0),\nu)^2}{t}\\ &\le \frac{W_2(\alpha(s),\nu)^2 - W_2(\alpha(0),\nu)^2}{s} + \frac{S^2}{t}\Big(1-\frac{s}{t}\Big)W_2\big(\alpha(0),\alpha(t)\big)^2,\end{aligned}$$

and letting $s$ tend to zero completes the proof. $\square$

**Lemma 3.2** *For any pair of minimal geodesics $\alpha : [0,\delta) \longrightarrow \mathcal{P}(X)$, $\beta : [0,1] \longrightarrow \mathcal{P}(X)$ with $\alpha(0)=\beta(0)=:\mu$, we have*

$$\limsup_{s\to 0}\frac{W_2(\alpha(s),\beta(1))^2 - W_2(\mu,\beta(1))^2}{s} \le -2\langle\dot\alpha(0),\dot\beta(0)\rangle_\mu. \tag{3}$$

*Proof.* For any $s\in(0,\delta)$ and $t\in(0,s^{-1})$, the triangle inequality gives

$$\begin{aligned}\frac{W_2(\alpha(s),\beta(1)) - W_2(\mu,\beta(1))}{s} &\le \frac{W_2(\alpha(s),\beta(st)) + W_2(\beta(st),\beta(1)) - W_2(\mu,\beta(1))}{s}\\ &= \frac{W_2(\alpha(s),\beta(st)) - W_2(\mu,\beta(st))}{s}.\end{aligned}$$

Thus we have, for any $t > 0$,

$$\limsup_{s\to 0} \frac{W_2(\alpha(s),\beta(1)) - W_2(\mu,\beta(1))}{s} \le \sigma\big(\dot\alpha(0), t\dot\beta(0)\big) - t|\dot\beta(0)|$$
$$= \frac{|\dot\alpha|^2 - 2t|\dot\alpha||\dot\beta|\cos\angle(\alpha,\beta)}{\sigma(\dot\alpha, t\dot\beta) + t|\dot\beta|}(0).$$

Letting $t$ go to infinity implies

$$\limsup_{s\to 0} \frac{W_2(\alpha(s),\beta(1))^2 - W_2(\mu,\beta(1))^2}{s}$$
$$= 2W_2\big(\mu,\beta(1)\big) \limsup_{s\to 0} \frac{W_2(\alpha(s),\beta(1)) - W_2(\mu,\beta(1))}{s}$$
$$\le 2|\dot\beta| \frac{-2|\dot\alpha||\dot\beta|\cos\angle(\alpha,\beta)}{2|\dot\beta|}(0) = -2\langle\dot\alpha(0),\dot\beta(0)\rangle.$$

□

We remark that equality holds in (3) if $X$ is nonnegatively curved (cf. [BBI, Theorem 4.5.6]).

**Lemma 3.3** *For any triplet* $\mathbf{v}_1, \mathbf{v}_2, \mathbf{w} \in C_\mu[\mathcal{P}(X)]$*, we have*

$$\langle \mathbf{v}_1, \mathbf{w}\rangle_\mu \le \langle \mathbf{v}_2, \mathbf{w}\rangle_\mu + |\mathbf{w}|_\mu \cdot \sigma_\mu(\mathbf{v}_1, \mathbf{v}_2).$$

*Proof.* We just calculate, putting $\mathbf{v}_i = (\alpha_i, s_i)$ and $\mathbf{w} = (\beta, t)$,

$$|\langle \mathbf{v}_1, \mathbf{w}\rangle - \langle \mathbf{v}_2, \mathbf{w}\rangle|^2 = t^2\big\{s_1\cos\angle(\alpha_1,\beta) - s_2\cos\angle(\alpha_2,\beta)\big\}^2$$
$$= t^2\Big[s_1^2\{1-\sin^2\angle(\alpha_1,\beta)\} + s_2^2\{1-\sin^2\angle(\alpha_2,\beta)\}$$
$$- 2s_1s_2\cos\angle(\alpha_1,\beta)\cos\angle(\alpha_2,\beta)\Big]$$
$$\le t^2\Big[s_1^2 + s_2^2 - 2s_1s_2\big\{\cos\angle(\alpha_1,\beta)\cos\angle(\alpha_2,\beta) + \sin\angle(\alpha_1,\beta)\sin\angle(\alpha_2,\beta)\big\}\Big]$$
$$= t^2\Big\{s_1^2 + s_2^2 - 2s_1s_2\cos\big(\angle(\alpha_1,\beta) - \angle(\alpha_2,\beta)\big)\Big\}$$
$$\le t^2\big\{s_1^2 + s_2^2 - 2s_1s_2\cos\angle(\alpha_1,\alpha_2)\big\}$$
$$= t^2\sigma(\mathbf{v}_1,\mathbf{v}_2)^2.$$

□

Now we are ready to prove our main theorem.

**Theorem 3.4** (First variation formula) *Let* $\xi : [0,\delta) \longrightarrow \mathcal{P}(X)$ *be a curve right differentiable at* 0*, and let* $\beta : [0,1] \longrightarrow \mathcal{P}(X)$ *be a minimal geodesic from* $\mu := \xi(0)$ *to* $\nu$*. Then we have*

$$\limsup_{t\to 0} \frac{W_2(\xi(t),\nu)^2 - W_2(\mu,\nu)^2}{t} \le -2\langle\dot\xi(0),\dot\beta(0)\rangle_\mu.$$

*Proof.* For each small $t > 0$, let $\alpha_t : [0, t] \longrightarrow \mathcal{P}(X)$ be a minimal geodesic from $\mu$ to $\xi(t)$. Then it follows from the right differentiability of $\xi$ that $\dot{\alpha}_t(0)$ converges to $\dot{\xi}(0)$ in $C_\mu[\mathcal{P}(X)]$. We deduce from Lemmas 3.1, 3.2, 3.3 that

$$\frac{W_2(\xi(t),\nu)^2 - W_2(\mu,\nu)^2}{t} = \frac{W_2(\alpha_t(t),\nu)^2 - W_2(\mu,\nu)^2}{t}$$
$$\leq -2\langle \dot{\alpha}_t(0), \dot{\beta}(0)\rangle + \frac{S^2}{t} W_2\big(\mu, \alpha_t(t)\big)^2$$
$$\leq -2\langle \dot{\xi}(0), \dot{\beta}(0)\rangle + 2|\dot{\beta}(0)| \cdot \sigma\big(\dot{\xi}(0), \dot{\alpha}_t(0)\big) + \frac{S^2}{t} W_2\big(\mu, \xi(t)\big)^2.$$

Letting $t$ tend to zero, we complete the proof. $\square$

The following simple lemma (valid for general metric spaces) is useful.

**Lemma 3.5** *Let $\xi, \zeta : [0, \delta) \longrightarrow Y$ be curves in a metric space $(Y, d)$, and $z \in Y$ be a midpoint of $x := \xi(0)$ and $y := \zeta(0)$ (i.e., $d(z, x) = d(z, y) = d(x, y)/2$). Then we have*

$$\limsup_{t\to 0} \frac{d(\xi(t),\zeta(t))^2 - d(x,y)^2}{t}$$
$$\leq 2 \limsup_{t\to 0} \frac{d(\xi(t),z)^2 - d(x,z)^2}{t} + 2 \limsup_{t\to 0} \frac{d(\zeta(t),z)^2 - d(y,z)^2}{t}.$$

*Proof.* The triangle inequality immediately implies

$$\limsup_{t\to 0} \frac{d(\xi(t),\zeta(t))^2 - d(x,y)^2}{t} = 2d(x,y) \limsup_{t\to 0} \frac{d(\xi(t),\zeta(t)) - d(x,y)}{t}$$
$$\leq 2d(x,y) \limsup_{t\to 0} \frac{d(\xi(t),z) + d(\zeta(t),z) - d(x,y)}{t}$$
$$\leq 2d(x,y) \left\{ \limsup_{t\to 0} \frac{d(\xi(t),z) - d(x,z)}{t} + \limsup_{t\to 0} \frac{d(\zeta(t),z) - d(y,z)}{t} \right\}$$
$$= 2 \limsup_{t\to 0} \frac{d(\xi(t),z)^2 - d(x,z)^2}{t} + 2 \limsup_{t\to 0} \frac{d(\zeta(t),z)^2 - d(y,z)^2}{t}.$$

$\square$

Combining this with Theorem 3.4 yields the following first variation formula for the distance between two right differentiable curves.

**Corollary 3.6** *Let $\xi, \zeta : [0, \delta) \longrightarrow \mathcal{P}(X)$ be two curves right differentiable at 0. Put $\mu := \xi(0)$, $\nu := \zeta(0)$, let $\alpha : [0, 1] \longrightarrow \mathcal{P}(X)$ be a minimal geodesic from $\mu$ to $\nu$ and let $\beta(\tau) := \alpha(1 - \tau)$ be its converse curve. Then we have*

$$\limsup_{t\to 0} \frac{W_2(\xi(t),\zeta(t))^2 - W_2(\mu,\nu)^2}{t} \leq -2\langle \dot{\xi}(0), \dot{\alpha}(0)\rangle_\mu - 2\langle \dot{\zeta}(0), \dot{\beta}(0)\rangle_\nu.$$

*Proof.* Apply Lemma 3.5 to $\xi$ and $\zeta$ with $z = \alpha(1/2) = \beta(1/2)$. Then Theorem 3.4 yields the desired estimate. $\square$

**Remark 3.7** If $X$ is nonnegatively curved, then $(\mathcal{P}(X), W_2)$ is an Alexandrov space and hence the right differentiability gives a better control at the level of $\mathcal{P}(X)$ (not only in $C_\mu[\mathcal{P}(X)]$). Such *strong right differentiability* ([Oh, (6.1)]) and the first variation formula along geodesics (Lemma 3.2) immediately lead us to the formula along right differentiable curves (see [Oh, Lemma 6.1]).

# 4 Applications for gradient flows in $\mathcal{P}(X)$

In this section, we use the first variation formula in the previous section to extend results in [Oh, Section 6] where we assumed that $(X, d)$ is nonnegatively curved.

## 4.1 Uniqueness and contraction

As in Subsection 2.2, let $f : \mathcal{P}(X) \longrightarrow (-\infty, +\infty]$ be a lower semi-continuous, $K$-convex function for some $K \in \mathbb{R}$ such that $\mathcal{P}^*(X) = f^{-1}((-\infty, +\infty))$ is nonempty. We first verify the *evolution variational inequality* (see [AGS, (4.0.13)]) as a consequence of first variation formulas Lemma 2.4 and Theorem 3.4.

**Proposition 4.1** (Evolution variational inequality) *Let $\xi : [0, \infty) \longrightarrow \mathcal{P}^*(X)$ be a gradient curve of $f$. Then we have, for any $t \in (0, \infty)$ and $\nu \in \mathcal{P}(X)$,*

$$\limsup_{\varepsilon \downarrow 0} \frac{W_2(\xi(t+\varepsilon), \nu)^2 - W_2(\xi(t), \nu)^2}{2\varepsilon} + \frac{K}{2} W_2\big(\xi(t), \nu\big)^2 + f\big(\xi(t)\big) \le f(\nu).$$

*Proof.* The assertion is clear if $f(\nu) = \infty$, so that we assume $\nu \in \mathcal{P}^*(X)$. We observe from Theorem 3.4 that

$$\limsup_{\varepsilon \downarrow 0} \frac{W_2(\xi(t+\varepsilon), \nu)^2 - W_2(\xi(t), \nu)^2}{2\varepsilon} \le -\langle \dot{\xi}(t), \dot{\beta}(0) \rangle_{\xi(t)},$$

where $\beta : [0, 1] \longrightarrow \mathcal{P}^*(X)$ is a minimal geodesic from $\xi(t)$ to $\nu$. Then Lemma 2.4 and the $K$-convexity (2) of $f$ together imply

$$-\langle \dot{\xi}(t), \dot{\beta}(0) \rangle_{\xi(t)} \le \lim_{\tau \to 0} \frac{f(\beta(\tau)) - f(\xi(t))}{\tau} \le f(\nu) - f\big(\xi(t)\big) - \frac{K}{2} W_2\big(\xi(t), \nu\big)^2.$$

□

A similar argument shows the contraction property of gradient curves.

**Theorem 4.2** (Contraction and uniqueness) *Given any pair of gradient curves $\xi, \zeta : [0, \infty) \longrightarrow \mathcal{P}^*(X)$ of $f$,*

$$W_2\big(\xi(t), \zeta(t)\big) \le e^{-tK} W_2\big(\xi(0), \zeta(0)\big)$$

*holds for all $t \in [0, \infty)$. In particular, each $\mu \in \mathcal{P}^*(X)$ admits a unique complete gradient curve starting from $\mu$.*

*Proof.* Put $h(t) := W_2(\xi(t),\zeta(t))^2$, fix $t \in (0,\infty)$, let $\alpha : [0,1] \longrightarrow \mathcal{P}(X)$ be a minimal geodesic from $\xi(t)$ to $\zeta(t)$ and put $\beta(\tau) := \alpha(1-\tau)$. Then Corollary 3.6 and Lemma 2.4 show that

$$\begin{aligned}\limsup_{\varepsilon\downarrow 0}\frac{h(t+\varepsilon)-h(t)}{\varepsilon} &\le -2\langle\dot\xi(t),\dot\alpha(0)\rangle_{\xi(t)} - 2\langle\dot\zeta(t),\dot\beta(0)\rangle_{\zeta(t)} \\ &\le 2\bigg\{\lim_{\tau\to 0}\frac{f(\alpha(\tau))-f(\alpha(0))}{\tau} + \lim_{\tau\to 0}\frac{f(\beta(\tau))-f(\beta(0))}{\tau}\bigg\} \le -2Kh(t).\end{aligned}$$

We used the $K$-convexity (2) along $\alpha$ in the last inequality. Therefore we obtain $h' \le -2Kh$ for a.e. $t$, and hence $h(t) \le e^{-2tK}h(0)$ by Gronwall's theorem. □

We define the *gradient flow* $G : \mathcal{P}^*(X)\times[0,\infty) \longrightarrow \mathcal{P}^*(X)$ as $\xi(t) := G(\mu,t)$ is the unique gradient curve starting from $\xi(0) = \mu$. Note that $G$ is continuous by virtue of the contraction property.

**Corollary 4.3** *The gradient flow* $G : \mathcal{P}^*(X)\times[0,\infty) \longrightarrow \mathcal{P}^*(X)$ *extends uniquely and continuously to* $G : \overline{\mathcal{P}^*(X)}\times[0,\infty) \longrightarrow \overline{\mathcal{P}^*(X)}$. *Moreover,* $G$ *satisfies the contraction property:*

$$W_2\big(G(\mu,t),G(\nu,t)\big) \le e^{-tK}W_2\big(G(\mu,0),G(\nu,0)\big)$$

*for* $\mu,\nu \in \overline{\mathcal{P}^*(X)}$ *and* $t \in (0,\infty)$*; as well as the semigroup property:*

$$G(\mu,s+t) = G\big(G(\mu,s),t\big)$$

*for* $\mu \in \overline{\mathcal{P}^*(X)}$ *and* $s,t \in [0,\infty)$.

## 4.2 Heat flow as gradient flow

Until here, we only deal with the triangle comparison property of the Wasserstein space. In this last subsection, in order to see that gradient flow of the free energy produces a solution to the Fokker-Planck equation, we use the structure of the underlying space (that was implicitly avoided in [Oh]). This kind of interpretation of evolution equations goes back to celebrated work of Jordan et al. [JKO]. It is recently demonstrated that there is also a remarkable connection with the Ricci flow (see [MT]).

Although some parts also work in Alexandrov spaces, we consider only compact Riemannian manifolds for brevity. Since Theorem 4.2 ensures that our gradient curve coincides with the one constructed as in [AGS] (see Remark 2.7), the realization of solutions to the Fokker-Planck equation as gradient flows of the free energy is a well established fact in the Riemannian setting. Here, however, we present a way of completing the self-contained proof in [Oh, Subsection 6.2] along our notion of gradient curves for thoroughness.

Let $(M,g)$ be a compact Riemannian manifold equipped with the associated Riemannian distance $d$ and the volume measure $\mathrm{vol}_g$. Thanks to McCann's theorem [Mc], we can represent each $\mathbf{v} \in C_\mu[\mathcal{P}(M)]$ as a (measurable) vector field on $M$ which will be again denoted by $\mathbf{v}$ by a slight abuse of notation. Moreover, for $\mathbf{v},\mathbf{w} \in C_\mu[\mathcal{P}(M)]$, we have $\sigma_\mu(\mathbf{v},\mathbf{w})^2 = \int_M |\mathbf{v}(x)-\mathbf{w}(x)|^2\,d\mu(x)$. In other words, Otto's [Ot] Riemannian structure coincides with ours induced from Theorem 2.2.

Due to the compactness of $M$, the Taylor expansion immediately gives the following:

**Lemma 4.4** *For any $h \in C^\infty(M)$ and any geodesic $\alpha : [0,l) \longrightarrow \mathcal{P}(M)$, we have*

$$\int_M h\, d\mu_t = \int_M h\, d\mu + t \int_M \langle \mathbf{v}, \nabla h\rangle\, d\mu + O_h\big(W_2(\mu,\mu_t)^2\big),$$

*where we set $\mu := \alpha(0)$, $\mu_t := \alpha(t)$, $\mathbf{v} := \dot\alpha(0) \in C_\mu[\mathcal{P}(M)]$.*

Let $f : \mathcal{P}(M) \longrightarrow (-\infty, +\infty]$ be as in Subsections 2.2, 4.1. Take $\mu \in \mathcal{P}^*(M)$ and let $\xi : [0,\infty) \longrightarrow \mathcal{P}^*(M)$ be the unique complete gradient curve with $\xi(0) = \mu$. Fix $t > 0$ and recall that $\xi$ is right differentiable with $\dot\xi(t) = \nabla_{\!-} f(\xi(t))$. Therefore we deduce from Lemma 4.4 that, for any $h \in C^\infty(M)$,

$$\lim_{\delta\downarrow 0} \frac{1}{\delta}\bigg\{ \int_M h\, d\mu_{t+\delta} - \int_M h\, d\mu_t \bigg\} = \int_M \langle \nabla_{\!-} f(\mu_t), \nabla h\rangle\, d\mu_t,$$

where we set $\mu_t := \xi(t)$. For each $\delta > 0$, choose some $\nu_\delta \in \mathcal{P}^*(M)$ attaining the infimum of the function

$$\mathcal{P}^*(M) \ni \nu \quad\longmapsto\quad f(\nu) + \frac{W_2(\mu_t,\nu)^2}{2\delta}.$$

Such $\nu_\delta$ indeed exists since $\mathcal{P}(M)$ is compact and $f$ is lower semi-continuous. We also choose a minimal geodesic $\beta_\delta : [0, l_\delta] \longrightarrow \mathcal{P}(M)$ from $\mu_t$ to $\nu_\delta$ where $l_\delta := W_2(\mu_t, \nu_\delta)$. Then we know that $(\beta_\delta, l_\delta/\delta) \in C_{\mu_t}[\mathcal{P}(M)]$ converges to $\nabla_{\!-} f(\mu_t)$ as $\delta$ tends to zero ([Oh, Lemma 6.4]). Thus Lemma 4.4 shows that, for any $h \in C^\infty(M)$,

$$\lim_{\delta\downarrow 0} \frac{1}{\delta}\bigg\{ \int_M h\, d\nu_\delta - \int_M h\, d\mu_t \bigg\} = \int_M \langle \nabla_{\!-} f(\mu_t), \nabla h\rangle\, d\mu_t.$$

Thus we conclude that

$$\lim_{\delta\downarrow 0} \frac{1}{\delta}\bigg\{ \int_M h\, d\mu_{t+\delta} - \int_M h\, d\nu_\delta \bigg\} = 0 \tag{4}$$

holds for all $h \in C^\infty(M)$.

**Remark 4.5** If $X$ has the nonnegative curvature, then the convergence of $(\beta_\delta, l_\delta/\delta)$ to $\nabla_{\!-} f(\mu_t)$ implies $\lim_{\delta\downarrow 0} W_2(\nu_\delta, \mu_{t+\delta})/\delta = 0$ (see [Oh, Lemma 6.4]). Thus the Kantorovich-Rubinstein theorem yields (4) without using Lemma 4.4, moreover, the convergence (4) is uniform for all 1-Lipschitz functions $h$.

For $\mu \in \mathcal{P}(M)$, we define the *relative entropy* as

$$\mathrm{Ent}(\mu) := \begin{cases} \int_M \rho \log \rho\, d\mathrm{vol}_g & \text{if } \mu = \rho \cdot \mathrm{vol}_g, \\ +\infty & \text{otherwise.} \end{cases}$$

Given $V \in C^\infty(M)$, let $f_V : \mathcal{P}(M) \longrightarrow (-\infty, \infty]$ be the associated *free energy*:

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

Note that $f_V$ is lower semi-continuous and the corresponding subset $\mathcal{P}^*(M) \subset \mathcal{P}(M)$ satisfies $\overline{\mathcal{P}^*(M)} = \mathcal{P}(M)$. Furthermore, the $K$-convexity of $f_V$ is known to be equivalent to the lower bound of the *Bakry-Émery tensor*: $\mathrm{Ric} + \mathrm{Hess}\, V \geq K$ ([St1]) (in particular, the $K$-convexity of Ent is equivalent to $\mathrm{Ric} \geq K$, [vRS]). Hence $f_V$ is $K$-convex for some $K \in \mathbb{R}$ by virtue of the compactness of $M$.

The estimate (4) is enough to follow the proof of [Oh, Theorem 6.6] and yields the following:

**Theorem 4.6** *Given $V \in C^\infty(M)$, a gradient curve $\xi = \rho \cdot \mathrm{vol}_g : [0,\infty) \longrightarrow \mathcal{P}^*(M)$ of $f_V$ produces a weak solution $\rho$ to the associated Fokker-Planck equation:*

$$\frac{\partial \rho}{\partial t} = \Delta \rho + \mathrm{div}(\rho \cdot \nabla V).$$

*In particular, the gradient flow of* Ent *coincides with the heat flow.*

To be precise, for any $h \in C^\infty(\mathbb{R} \times M)$ and $0 \leq t_1 < t_2 < \infty$, we have

$$\int_M h_{t_2}\, d\mu_{t_2} - \int_M h_{t_1}\, d\mu_{t_1} = \int_{t_1}^{t_2} \int_M \left\{ \frac{\partial h_t}{\partial t} + \Delta h_t - \langle \nabla h_t, \nabla V \rangle \right\} d\mu_t\, dt,$$

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

**Remark 4.7** At this point, it should be recalled that in the Riemannian setting there are two ways to see the heat flow as gradient flow: as gradient flow of the relative entropy with respect to the distance $W_2$ as we just did, or as gradient flow of the Dirichlet energy with respect to the $L^2$-distance (associated with the volume measure). Recently the second author and Sturm [OhSt] proved that these two approaches coincide also in the Finsler setting.

For finite-dimensional Alexandrov spaces, the construction of the heat kernel via the Dirichlet energy has been performed in [KMS]. It is yet to be proven that such heat kernel coincides with the gradient flow of the entropy with respect to $W_2$ in the genuine Alexandrov setting.

# References


[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), Uspekhi Mat. Nauk **47** (1992), 3–51, 222; English translation: Russian Math. Surveys **47** (1992), 1–58.

[Gi] N. Gigli, On the inverse implication of Brenier-McCann theorems and the structure of $(\mathcal{P}_2(M), W_2)$, Preprint (2009). Available at http://cvgmt.sns.it/people/gigli/

[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.

[LV] J. Lott and C. Villani, Ricci curvature for metric-measure spaces via optimal transport, Ann. of Math. **169** (2009), 903–991.

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

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

[MT] R. J. McCann and P. Topping, Ricci flow, entropy and optimal transportation, to appear in Amer. J. Math.

[Oh] S. Ohta, Gradient flows on Wasserstein spaces over compact Alexandrov spaces, Amer. J. Math. **131** (2009), 475–516.

[OhSt] S. Ohta and K.-T. Sturm, Heat flow on Finsler manifolds, Comm. Pure Appl. Math. **62** (2009), 1386–1433.

[OtSh] Y. Otsu and T. Shioya, The Riemannian structure of Alexandrov spaces, J. Differential Geom. **39** (1994), 629–658.

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

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

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

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

[St1] K.-T. Sturm, Convex functionals of probability measures and nonlinear diffusions on manifolds, J. Math. Pures Appl. **84** (2005), 149–168.

[St2] K.-T. Sturm, On the geometry of metric measure spaces, Acta Math. **196** (2006), 65–131.

[Vi] C. Villani, Optimal transport, old and new, Springer-Verlag, Berlin, 2009.