Heat flow on Alexandrov spaces

Nicola Gigli[*][†], Kazumasa Kuwada[‡][§], Shin-ichi Ohta[¶][‖]

December 13, 2010

Abstract

We prove that on compact Alexandrov spaces with curvature bounded below the gradient flow of the Dirichlet energy in the L^2-space produces the same evolution as the gradient flow of the relative entropy in the L^2-Wasserstein space. This means that the heat flow is well defined by either one of the two gradient flows. Combining properties of these flows, we are able to deduce the Lipschitz continuity of the heat kernel as well as Bakry-Émery gradient estimates and the Γ_2-condition. Our identification is established by purely metric means, unlike preceding results relying on PDE techniques. Our approach generalizes to the case of heat flow with drift.

Contents

1 Introduction

The heat equation is one of the most important evolutionary PDEs. It is a well known fact in modern analysis that such an equation, say in $\mathbb{R}^n$, can be seen as the gradient flow of the Dirichlet energy

$$\frac{1}{2} \int_{\mathbb{R}^n} |\nabla f|^2 \, dx, \tag{1.1}$$

in the space $L^2(\mathbb{R}^n, \mathcal{L}^n)$. This viewpoint has been extended to the concept of Dirichlet form (see e.g. [5, 12]) and it has grown up into a huge research field in potential analysis and probability theory. More recently, Jordan, Kinderlehrer and Otto [15] understood that the same equation can be seen as the gradient flow of the relative entropy

$$\int_{\mathbb{R}^n} \rho \log \rho \, dx, \tag{1.2}$$

in the L^2-Wasserstein space $(\mathscr{P}_2(\mathbb{R}^n), W_2)$, where $\mathscr{P}_2(\mathbb{R}^n)$ stands for the space of Borel probability measures on $\mathbb{R}^n$ with finite second moment. This intuition, with the further studies of Otto [30], has been one of the fundamental ingredients that drove the research in the field of gradient flows in relation with optimal transportation problems in the past decade (see, e.g., [1] and [40]).

The aim of this paper is to carry on the study of the heat flow as gradient flow of the two very different functionals, (1.1) and (1.2) in the two metric spaces $L^2(\mathbb{R}^n, \mathcal{L}^n)$ and $(\mathscr{P}_2(\mathbb{R}^n), W_2)$, in a non-smooth setting. The point is the following. On the one hand, it is known that these two gradient flows produce the same evolution 'in all the smooth settings', i.e., it has been proved that in $\mathbb{R}^n$, in Riemannian manifolds with Ricci curvature bounded below and in compact Finsler manifolds (for the latter see the work of Sturm and the third author [28]) the gradient flow of the Dirichlet energy with respect to the L^2-distance coincides with the gradient flow of the relative entropy with respect to the Wasserstein distance W_2 (see also [10], [16] for related work on different kinds of spaces). On the other hand, these flows are studied well also in some non-smooth settings. These raise the natural question: do these two notions always coincide?

We remark that a natural abstract setting where one could try to give an answer to this question is the one of metric measure spaces with Ricci curvature bounded below (defined by Sturm [38] and Lott and Villani [26], see Definition 2.4).

Indeed, in order to define either one of the two gradient flows, one usually needs both a metric and a measure on the space considered. Furthermore, it is known that in the Riemannian setting the fact that the heat flow does not lose mass is strictly related to the bound from below on the Ricci curvature, so that one has to assume some kind of lower Ricci curvature bound also in the non-smooth setting.

In this paper, we restrict our attention to the case of finite dimensional, compact Alexandrov spaces $(X, d, \mathcal{H}^n)$ of curvature bounded below, equipped with the Hausdorff measure $\mathcal{H}^n$. An Alexandrov space of curvature $\geq k$ with $k \in \mathbb{R}$ is a metric space of sectional curvature $\geq k$ in the sense of the triangle comparison property (see Definition 2.1). As naturally expected and recently shown in [33] and [41], such spaces actually have Ricci curvature bounded below in the sense of Lott-Sturm-Villani (see Remark 2.5). Our following main result is the first one establishing the equivalence of the two gradient flows in a genuinely non-smooth setting (see Theorem 3.1 for the precise statement):

Theorem 1.1 *For any $f_0 \in L^2(X, \mathcal{H}^n)$ with $f_0 \mathcal{H}^n \in \mathscr{P}(X)$, the gradient flow $(f_t)_{t\in[0,\infty)}$ of the Dirichlet energy with respect to the L^2-distance gives the gradient flow $(f_t\mathcal{H}^n)_{t\in[0,\infty)}$ of the relative entropy with respect to the L^2-Wasserstein distance W_2, and vice versa.*

On Alexandrov spaces, the structure of the heat flow as gradient flow has already been studied in [20] for the Dirichlet energy approach and in [27] by the third author (see also [14] and [36]) for the relative entropy approach. However, up to now it was not clear that these two notions coincide.

We observe that the idea behind our proof of the theorem is completely different from the ones used in the aforementioned smooth settings. Indeed, all the proofs available in the smooth setting had the following structure:

- one studies the gradient flow of the Dirichlet energy with respect to L^2 and writes down the equation that it satisfies;

- one studies the gradient flow of the relative entropy with respect to W_2 and writes down the equation that it satisfies;

- one realizes that the two equations are actually the same and calls into play the uniqueness of the solution of the equation to conclude that the two gradient flows coincide.

Our approach, instead, is completely different and in particular does not pass from the study of the heat equation in the Alexandrov setting. Read back in $\mathbb{R}^n$ our proof gives a new and purely metric way to prove the coincidence of these two gradient flows.

After having proved the identification, we can combine them: their interaction gives fruitful applications. Among them, the Lipschitz continuity (Theorem 4.4) of the heat kernel is derived as a corollary of a Bakry-Émery type gradient estimate (Theorem 4.3), which follows with the aid of a result of the second author [19]. The Lipschitz regularity improves the Hölder regularity established in [20]. We believe it is curious and worth to be underlined that such Lipschitz property immediately follows when one knows the equivalence of the two gradient flows, but is not at all trivial if one sticks to either one of the two approaches alone. Our Lipschitz regularity enables us to deduce the Γ_2-condition (Theorem 4.6) from a Lott-Sturm-Villani Ricci curvature bound. The Γ_2-condition as well as the Wasserstein contraction of heat flow and the Bakry-Émery gradient estimate for heat semigroup are known to be analytic characterizations of a lower Ricci curvature bound (see [35, 2, 23]). We can show the equivalence of these conditions even on Alexandrov spaces (with sharp constants, Theorem 4.8).

It should be remarked that our approach easily generalizes to the case of heat flow with drift, where the drift is given by the gradient of a semiconvex potential V. In other words, the current approach can be used to study the heat flow on weighted Alexandrov spaces of the kind $(X, d, \mathrm{e}^{-V}\mathcal{H}^n)$ (see the end of Section 4).

After having completed the work on this paper, we got aware of a preprint [42] of Zhang and Zhu where the Lipschitz continuity of the heat kernel has been studied, via a completely different argument. In particular, their proof relies on the Lipschitz continuity of harmonic functions due to Petrunin [32]. Our estimate depends only on the 'Ricci curvature bound', while that in [42] involves the 'sectional curvature bound' and the dimension of the space. It is not discussed in [42] whether their approach generalizes to the case of weighted Alexandrov spaces.

The organization of the article is as follows. Section 2 is devoted to preliminaries for Alexandrov geometry and known results on the gradient flow of the Dirichlet energy as well as the gradient flow of the relative entropy on compact Alexandrov spaces. We prove our main theorem in Section 3, and discuss its applications in Section 4.

2   Preliminaries and notations

2.1   Alexandrov spaces

We first review the basics of Alexandrov geometry, see [7], [29] and [6] for details.

Let (X, d) be a metric space. A rectifiable curve $\gamma : [0, l] \to X$ is called a geodesic if it is locally length minimizing and parametrized with constant speed. (Precisely, for any $t_0 \in [0, l]$, there is $\varepsilon > 0$ such that $d(\gamma(s), \gamma(t))/|s-t|$ is constant for all $s, t \in [t_0 - \varepsilon, t_0 + \varepsilon] \cap [0, l]$.) If γ is minimizing between endpoints, then we

call it a minimal geodesic. We say that (X, d) is a geodesic space if any pair of points in X are connected by a minimal geodesic.

For $k \in \mathbb{R}$, we denote by $\mathbb{M}^2(k)$ the simply-connected, two-dimensional space form of constant sectional curvature k. Given three points $x, y, z \in X$, with $d(x, y) + d(y, z) + d(z, x) < 2\pi/\sqrt{k}$ if $k > 0$, we can take a comparison triangle $\triangle \tilde{x}\tilde{y}\tilde{z} \subset \mathbb{M}^2(k)$ such that $d(\tilde{x}, \tilde{y}) = d(x, y)$, $d(\tilde{y}, \tilde{z}) = d(y, z)$ and $d(\tilde{z}, \tilde{x}) = d(z, x)$. Such a triangle is unique up to a difference of isometry.

Definition 2.1 (Alexandrov spaces) For $k \in \mathbb{R}$, a complete geodesic space (X, d) is called an *Alexandrov space of curvature* $\geq k$ if, for any three points $x, y, z \in X$ (with $d(x, y) + d(y, z) + d(z, x) < 2\pi/\sqrt{k}$ if $k > 0$) and any minimal geodesic $\gamma : [0, 1] \to X$ from y to z, we have $d(x, \gamma(t)) \geq d(\tilde{x}, \tilde{\gamma}(t))$ for all $t \in [0, 1]$, where $\triangle \tilde{x}\tilde{y}\tilde{z} \subset \mathbb{M}^2(k)$ is a comparison triangle of $\triangle xyz$ and $\tilde{\gamma} : [0, 1] \to \mathbb{M}^2(k)$ is the unique minimal geodesic from $\tilde{y}$ to $\tilde{z}$.

Example 2.2 (a) A complete Riemannian manifold is an Alexandrov space of curvature $\geq k$ if and only if its sectional curvature is not less than k everywhere.

(b) If (X, d) is an Alexandrov space of curvature $\geq k$, then the scaled metric space $(X, c \cdot d)$ with $c > 0$ is an Alexandrov space of curvature $\geq k/c^2$.

(c) For a convex domain D in the Euclidean space $\mathbb{R}^n$, the boundary ∂D equipped with the length distance is an Alexandrov space of nonnegative curvature.

(d) Let (M, g) be a Riemannian manifold of nonnegative sectional curvature 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 nonnegative curvature.

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

Fix $x \in X$ and let $\hat{\Sigma}_x$ be the set of all unit speed geodesics $\gamma : [0, l] \to X$ with $\gamma(0) = x$. For $\gamma, \eta \in \hat{\Sigma}_x$, thanks to the curvature bound, the joint limit

$$\angle_x(\gamma, \eta) := \arccos \left(\lim_{s,t \to 0} \frac{s^2 + t^2 - d(\gamma(s), \eta(t))^2}{2st} \right)$$

exists and is a pseudo-distance of $\hat{\Sigma}_x$. We define the space of directions $(\Sigma_x, \angle_x)$ at x as the completion of $\hat{\Sigma}_x/\sim$ with respect to $\angle_x$, where $\gamma \sim \eta$ holds if $\angle_x(\gamma, \eta) = 0$. The tangent cone (K_x, d) is the Euclidean cone over $(\Sigma_x, \angle_x)$, i.e.,

$$K_x := \Sigma_x \times [0, \infty)/\Sigma_x \times \{0\},$$

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

The inner product on K_x is defined by $\langle(\gamma,s),(\eta,t)\rangle_x := st\cos\angle_x(\gamma,\eta)$. In Riemannian manifolds, spaces of directions and tangent cones correspond to unit tangent spheres and tangent spaces, respectively.

The Hausdorff dimension of X is an integer or infinity. From here on, *we consider a compact n-dimensional Alexandrov space of curvature $\geq k$ equipped with the n-dimensional Hausdorff measure $\mathcal{H}^n$.* We say that $x \in X$ is a singular point if K_x is not isometric to the Euclidean space $\mathbb{R}^n$, and denote the set of singular points by S_X. We remark that $\mathcal{H}^n(S_X) = 0$ holds and that $S_X \setminus \partial X$ has codimension at least two ([7], [29], see [7] for the definition of ∂X), whereas S_X can be dense in X.

2.2   Dirichlet energy and the associated gradient flow

We introduce the Sobolev space and the Dirichlet energy following [20], and will see that it coincides with other notions of Sobolev spaces. We begin by discussing a C^1-differentiable structure of the set of regular points $X \setminus S_X$ established in [7] and [29]. We remark that Perelman extends this to DC^1-structure (via 'difference of concave functions', see [31]), but the C^1-structure is enough for considering the Sobolev space. There is a weak C^1-atlas $\{(U_\phi, V_\phi, \phi)\}_{\phi\in\Phi}$ in the sense that $U_\phi \subset X$ is an open set, $\phi : U_\phi \to \mathbb{R}^n$ is a bi-Lipschitz embedding, $V_\phi \subset U_\phi$ with $\bigcup_{\phi\in\Phi} V_\phi \supset X \setminus S_X$, and that the coordinate change $\phi_2 \circ \phi_1^{-1}$ is C^1 on $\phi_1(V_{\phi_1} \cap V_{\phi_2} \cap (X \setminus S_X))$ if $V_{\phi_1} \cap V_{\phi_2} \neq \emptyset$ ([29, Theorem 4.2(1)]). Such charts are constructed through the distance function. Precisely, ϕ is introduced as $\phi(x) := (d(x,p_1), d(x,p_2), \ldots, d(x,p_n))$ for suitable $p_1, p_2, \ldots, p_n \in X$, and then V_ϕ is chosen as the set of regular points x such that a minimal geodesic between x and p_i is unique for all i ([29, Section 3]). It is worth mentioning that, for any $\varepsilon > 0$, ϕ can be $(1+\varepsilon)$-bi-Lipschitz by taking smaller U_ϕ (cf. [6, Theorem 10.9.16]). We also remark that it is possible to modify ϕ by taking an average so as to satisfy $V_\phi = U_\phi$ ([29, Section 5]), but it is unnecessary for our discussion (just like [4, Remark 2.9]).

We say that a function f on X is differentiable at a regular point $x \in X \setminus S_X$ if $x \in V_\phi$ and $f \circ \phi^{-1}$ is differentiable at $\phi(x)$ for some $\phi \in \Phi$. Then we can define the gradient vector $\nabla f(x) \in K_x$ by identifying K_x and $\mathbb{R}^n$ through ϕ. (To be precise, by virtue of the first variation formula ([29, Theorem 3.5]), each $d_i := d(\cdot, p_i)$ is differentiable at x with $\nabla d_i(x) = -v_i$, where $v_i \in K_x$ is the tangent vector of the unique, minimal, unit speed geodesic from x to p_i. Then K_x is linearly identified with $\mathbb{R}^n$ as $\sum_i a_i v_i = (-a_i)$.) Moreover, again due to the first variation formula, we obtain the Taylor expansion

$$f\big(\gamma(t)\big) = f(x) + t\langle\nabla f(x), \dot\gamma(0)\rangle_x + o_x(t), \tag{2.1}$$

6

where $\gamma : [0, \delta] \to X$ is a minimal geodesic emanating from x and $o_x(t)$ is independent of the choice of γ (see [4, Lemma 3.4], and note that the remainder term in the first variation formula for $d(\cdot, p_i)$ at x indeed depends only on x and p_i). Another important fact we will use is the Rademacher theorem, namely a Lipschitz function f on X is differentiable $\mathcal{H}^n$-a.e.. This easily follows from the usual Rademacher theorem for $f \circ \phi^{-1}$ (see [4, Corollary 2.14]). It follows from (2.1) that $\sqrt{\langle \nabla f(x), \nabla f(x) \rangle}$ coincides with the local Lipschitz constant $|\nabla_L f|(x)$ given by

$$|\nabla_L f|(x) := \varlimsup_{y \to x} \frac{|f(x) - f(y)|}{d(x, y)}.$$

Based on the notion of gradient vector, we define the Sobolev space and the Dirichlet energy as follows (see [20] and [22] for details). For a function $f : X \to \mathbb{R}$ such that $f \circ \phi^{-1} \in W^{1,2}(\phi(U_\phi))$ for all $\phi \in \Phi$, we introduce the weak gradient vector $\nabla f(x) \in K_x$ for a.e. $x \in V_\phi$ as the element corresponding to the weak gradient vector $\nabla (f \circ \phi^{-1})(\phi(x))$. We define the Sobolev space $W^{1,2}(X)$ and the Dirichlet energy $\mathcal{E}$ by

$$W^{1,2}(X) := \left\{ f \in L^2(X, \mathcal{H}^n) \; \middle| \; \int_X \langle \nabla f, \nabla f \rangle \, d\mathcal{H}^n < \infty \right\},$$

$$\mathcal{E}(f, g) := \int_X \langle \nabla f, \nabla g \rangle \, d\mathcal{H}^n \qquad \text{for } f, g \in W^{1,2}(X).$$

We do not divide $\mathcal{E}$ by 2 for notational simplicity. Note that $\mathcal{E}$ coincides with the energy functional introduced by Korevaar and Schoen [18] ([22, Theorem 6.2], we can reduce the argument to the Euclidean case by using a $(1 + \varepsilon)$-bi-Lipschitz chart). We also remark that the set of Lipschitz functions $C^{\mathrm{Lip}}(X)$ is dense in $W^{1,2}(X)$ with respect to the Sobolev norm $\|f\|_{W^{1,2}}^2 = \|f\|_{L^2}^2 + \mathcal{E}(f, f)$ ([20, Theorem 1.1], the point is that $S_X \setminus \partial X$ has codimension at least two).

Furthermore, if f is a Lipschitz function, then the weak gradient vector $\nabla f(x)$ coincides with the gradient vector as in (2.1) a.e. x, and hence $\sqrt{\langle \nabla f(x), \nabla f(x) \rangle} = |\nabla_L f|(x)$ holds a.e. x. Therefore $\mathcal{E}$ also coincides with Cheeger's energy functional ([8]), because the local Lipschitz constant is the minimal generalized upper gradient ([8, Theorem 6.1]) and Lipschitz functions are dense in both Sobolev spaces (thanks to the weak Poincaré inequality for upper gradients and the volume doubling condition, [8, Theorem 4.24]). Indeed, in our framework, the volume doubling condition directly follows from the Bishop-Gromov volume comparison theorem and the Poincaré inequality is a consequence of [21] and [34], for instance.

By following the general theory of bilinear forms, there exists a selfadjoint operator $(\Delta, D(\Delta))$ on $L^2(X, \mathcal{H}^n)$ associated with $(\mathcal{E}, W^{1,2}(X))$. It is characterized

by the following identity:

$$\mathcal{E}(g, f) = - \int_X g\Delta f \, d\mathcal{H}^n, \quad f \in D(\Delta), \, g \in W^{1,2}(X). \tag{2.2}$$

We call Δ the Laplacian as in the classical case. The semigroup of contractive symmetric linear operators $T_t = e^{t\Delta}$ on $L^2(X, \mathcal{H}^n)$ is defined associated with Δ. For any $f \in L^2(X, \mathcal{H}^n)$, $T_t f$ solves the (linear) heat equation $\partial_t u = \Delta u$ with $u(0, \cdot) = f$ in the sense that $T_t f \in D(\Delta)$ for $t > 0$ and

$$\lim_{t \downarrow 0} \frac{T_t f - f}{t} = \Delta f \quad \text{in } L^2(X, \mathcal{H}^n) \text{ for } f \in D(\Delta),$$

$$\lim_{t \downarrow 0} T_t f = f \quad \text{in } L^2(X, \mathcal{H}^n) \text{ for } f \in L^2(X, \mathcal{H}^n).$$

Note that T_t is Markovian in the sense that $0 \le T_t f \le 1$ holds whenever $0 \le f \le 1$. As shown in [20, Theorem 1.5], there exists a continuous function $(t, x, y) \mapsto p_t(x, y)$ on $(0, \infty) \times X \times X$ satisfying the following properties:

(i) For any $f \in L^2(X, \mathcal{H}^n)$, $t > 0$ and $\mathcal{H}^n$-a.e. $x \in X$,

$$T_t f(x) = \int_X p_t(x, y) f(y) \, \mathcal{H}^n(dy). \tag{2.3}$$

(ii) For any $s, t > 0$ and $x, y \in X$,

$$p_t(x, y) = p_t(y, x), \tag{2.4}$$

$$p_{s+t}(x, y) = \int_X p_s(x, z) p_t(z, y) \, \mathcal{H}^n(dz), \tag{2.5}$$

$$\int_X p_t(x, z) \, \mathcal{H}^n(dz) = 1, \tag{2.6}$$

$$p_t(x, y) > 0.$$

See [20, Theorems 1.4, 1.5(3)] for the continuity of $p_t(x, y)$. The equality (2.6) follows from the fact that $1 \in W^{1,2}(X)$ and $\mathcal{E}(1, 1) = 0$, because X is assumed to be compact. As in the classical case, we call $p_t(x, y)$ the heat kernel. The existence and these properties of $p_t(x, y)$ are deduced from the Poincaré inequality for $\mathcal{E}$ and the volume doubling condition, together with results in [37].

Remark 2.3 The bilinear form $(\mathcal{E}, W^{1,2}(X))$ is a symmetric strongly local regular Dirichlet form (see [20] for it and further details; see [5, 12] for basics on Dirichlet forms). Moreover, T_t enjoys the strong Feller property. As a result, there exists a diffusion process $((X_t)_{t \ge 0}, (\mathbb{P}_x)_{x \in X})$ on the whole space X associated with $(\mathcal{E}, W^{1,2}(X))$ in the sense that $\mathbb{E}_x[f(X_t)] = T_t f(x)$ for $f \in C(X)$, $x \in X$ and $t > 0$.

Denote by $\mathscr{P}(X)$ the set of all Borel probability measures on X. Let us define a positive Borel measure $T_t\nu$ for $\nu \in \mathscr{P}(X)$ and $t \geq 0$ by

$$T_t\nu(dy) := \begin{cases} \left(\int_X p_t(x, y)\,\nu(dx)\right)\mathcal{H}^n(dy) & t > 0, \\ \nu(dy) & t = 0. \end{cases}$$

Thanks to (2.6), $T_t\nu \in \mathscr{P}(X)$ holds. By definition, $T_t\nu$ is absolutely continuous with respect to $\mathcal{H}^n$ for $t > 0$. When $d\nu = f d\mathcal{H}^n$, it holds $dT_t\nu = T_t f d\mathcal{H}^n$. In this paper, we call $T_t\nu$ *the gradient flow of the Dirichlet energy* with the initial condition ν. Indeed, it is easy to see from

$$\frac{1}{2}\mathcal{E}(f + \varepsilon g, f + \varepsilon g) = \frac{1}{2}\mathcal{E}(f, f) - \varepsilon \int_X g\Delta f\, d\mathcal{H}^n + O(\varepsilon^2)$$

that $T_t\nu$ is the gradient flow of $\mathcal{E}/2$ with respect to the L^2-norm.

Before closing this subsection, we review the derivation property of $W^{1,2}(X)$. It is formulated as follows: For $f_1, \ldots, f_k, g \in W^{1,2}(X) \cap L^\infty(X, \mathcal{H}^n)$ and $\Phi : \mathbb{R}^k \to \mathbb{R}$ which is C^1 on the range of $(f_1, \ldots, f_k)$, $\Phi(f_1, \ldots, f_k)$ belongs to $W^{1,2}(X)$ and

$$\langle \nabla\Phi(f_1, \ldots, f_k), \nabla g \rangle = \sum_{j=1}^n \frac{\partial\Phi}{\partial x_j}(f_1, \ldots, f_n)\,\langle \nabla f_j, \nabla g \rangle \quad \mathcal{H}^n\text{-a.e..} \tag{2.7}$$

It directly follows from the definition of $W^{1,2}(X)$ or from the strong locality of the Dirichlet form $(\mathcal{E}, W^{1,2}(X))$ (see [5, Corollary I.6.1.3] and [12, Section 3.2] for the latter).

2.3 Gradient flows in the Wasserstein space

We next introduce the Wasserstein space and a purely metric notion of gradient flows in it. We refer to [1] and [40] for the basic theory as well as the recent diverse developments.

Given $\mu, \nu \in \mathscr{P}(X)$, a probability measure $\pi \in \mathscr{P}(X \times X)$ is called a coupling of μ and ν if $\pi(A \times X) = \mu(A)$ and $\pi(X \times A) = \nu(A)$ hold for all measurable sets $A \subset X$. Then, for $1 \leq p < \infty$, we define the L^p-Wasserstein distance as

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

where π runs over all couplings of μ and ν. In most parts, we work in the quadratic case $p = 2$. The L^2-Wasserstein space $(\mathscr{P}(X), W_2)$ becomes a metric space and inherits the compactness from (X, d). Moreover, $(\mathscr{P}(X), W_2)$ is a geodesic space.

If μ is absolutely continuous with respect to $\mathcal{H}^n$, then a minimal geodesic $(\mu_t)_{t\in[0,1]}$ from μ to any ν is unique and μ_t is also absolutely continuous for all $t \in (0,1)$ (see [4] for a more detailed characterization of μ_t, and [11] for the absolute continuity).

For $\mu \in \mathscr{P}(X)$, we define the relative entropy by

$$\mathrm{Ent}(\mu) := \int_X \rho \log \rho \, d\mathcal{H}^n$$

when $\mu = \rho \mathcal{H}^n$ with $\rho \in L^1(X, \mathcal{H}^n)$, and $\mathrm{Ent}(\mu) := \infty$ otherwise. Set $\mathscr{P}^*(X) := \{\mu \in \mathscr{P}(X) \mid \mathrm{Ent}(\mu) < \infty\}$. Note that Ent is lower semi-continuous with respect to W_2 and satisfies $\mathrm{Ent}(\mu) \geq -\log \mathcal{H}^n(X)$ by Jensen's inequality.

Definition 2.4 (The curvature-dimension condition) For $K \in \mathbb{R}$, we say that $(X, d, \mathcal{H}^n)$ satisfies the *curvature-dimension condition* $\mathsf{CD}(K, \infty)$ if Ent is K-geodesically convex in the sense that any pair $\mu, \nu \in \mathscr{P}(X)$ admits a minimal geodesic $(\mu_t)_{t\in[0,1]}$ from μ to ν such that

$$\mathrm{Ent}(\mu_t) \leq (1-t)\,\mathrm{Ent}(\mu) + t\,\mathrm{Ent}(\nu) - \frac{K}{2}(1-t)tW_2(\mu,\nu)^2$$

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

We remark that the above inequality is obvious if $\mu \notin \mathscr{P}^*(X)$ or $\nu \notin \mathscr{P}^*(X)$. Therefore it is sufficient to consider $\mu, \nu \in \mathscr{P}^*(X)$, and then a minimal geodesic between them is unique.

Remark 2.5 (i) The curvature-dimension condition $\mathsf{CD}(K, \infty)$ for general metric measure spaces is introduced and studied independently in [38] and [26], and known to be equivalent to the lower Ricci curvature bound $\mathrm{Ric} \geq K$ for complete Riemannian manifolds equipped with the Riemannian distance and the volume measure ([35]).

(ii) The condition $\mathsf{CD}(K, N)$ for $N \in (1, \infty)$ is also introduced in [39] and [24]. In general, $\mathsf{CD}(K, N)$ implies $\mathsf{CD}(K, \infty)$. In the Riemannian case, $\mathsf{CD}(K, N)$ is equivalent to $\mathrm{Ric} \geq K$ and $\dim \leq N$.

(iii) It is recently demonstrated in [33] and [41] that n-dimensional Alexandrov spaces of curvature $\geq k$ satisfy $\mathsf{CD}((n-1)k, n)$ (and hence $\mathsf{CD}((n-1)k, \infty)$), as is naturally expected from the relation between the sectional and the Ricci curvatures.

There is a well established theory on the gradient flow of geodesically convex functions as comprehensively discussed in [1]. For later convenience, we recall a

couple of notions in a general form. We say that a curve $(\mu_t)_{t \in I} \subset \mathscr{P}(X)$ on an interval $I \subset \mathbb{R}$ is absolutely continuous if there is $f \in L^1(I)$ such that

$$W_2(\mu_t, \mu_s) \le \int_t^s f(r)\, dr \qquad (2.8)$$

for all $t, s \in I$ with $t \le s$. Note that absolutely continuous curves are continuous. For an absolutely continuous curve $(\mu_t)_{t \in I}$, the metric derivative

$$|\dot{\mu}_t| := \lim_{h \to 0} \frac{d(\mu_t, \mu_{t+h})}{h}$$

is well-defined for a.e. $t \in I$ ([1, Theorem 1.1.2]). Moreover, $|\dot{\mu}_t|$ belongs to $L^1(I)$ and is the minimal function for which (2.8) holds. Given a functional $E : \mathscr{P}(X) \to \mathbb{R} \cup \{+\infty\}$, we consider a gradient flow of E solving "$\dot{\mu}_t = -\nabla E(\mu_t)$" in the following sense. For $\mu \in \mathscr{P}(X)$ with $E(\mu) < \infty$, we define the local slope as

$$|\nabla_- E|(\mu) := \varlimsup_{\nu \to \mu} \frac{\max\{E(\mu) - E(\nu), 0\}}{W_2(\mu, \nu)}. \qquad (2.9)$$

If E is K-geodesically convex, then we have

$$|E(\mu_t) - E(\mu_s)| \le \int_t^s |\dot{\mu}_r| \cdot |\nabla_- E|(\mu_r)\, dr \qquad (2.10)$$

for all $t, s \in I$ with $t < s$ along any absolutely continuous curve $(\mu_t)_{t \in I}$ with values in $\mathscr{P}(X)$ with $E(\mu_t) < \infty$. As a consequence, it holds

$$E(\mu_t) \le E(\mu_s) + \frac{1}{2} \int_t^s |\dot{\mu}_r|^2\, dr + \frac{1}{2} \int_t^s |\nabla_- E|^2(\mu_r)\, dr \qquad (2.11)$$

for all $t < s$. From (2.10) and (2.11), it is natural to give the following definition.

Definition 2.6 (Gradient flows) Let $E : \mathscr{P}(X) \to \mathbb{R} \cup \{+\infty\}$ be a K-geodesically convex functional. We say that an absolutely continuous curve $(\mu_t)_{t \in [0, \infty)}$ in $\mathscr{P}(X)$ is a *gradient flow* of E provided $E(\mu_t) < \infty$ for $t \ge 0$ and

$$E(\mu_t) = E(\mu_s) + \frac{1}{2} \int_t^s |\dot{\mu}_r|^2\, dr + \frac{1}{2} \int_t^s |\nabla_- E|^2(\mu_r)\, dr \qquad (2.12)$$

for all $0 \le t < s$. The equation (2.12) is called the *energy dissipation identity*.

The existence of such a gradient flow comes from the general theory presented in [1, Corollary 2.4.11]. Furthermore, as shown in [27, Theorem 6.2] (see also [14, Theorem 4.2]), the gradient flow produces a contraction semigroup in the sense

that for any $\mu, \nu \in \mathscr{P}(X)$ with $E(\mu) < \infty$ and $E(\nu) < \infty$, the gradient flows $(\mu_t)_{t \in [0,\infty)}, (\nu_t)_{t \in [0,\infty)}$ starting from μ, ν satisfy

$$W_2(\mu_t, \nu_t) \leq \mathrm{e}^{-Kt} W_2(\mu, \nu) \qquad \text{for all } t \geq 0, \tag{2.13}$$

where K is the modulus of convexity of E. In particular, the uniqueness follows from considering $\mu = \nu$. (Though the strategy of the construction in [27] is different from [1], the resulting flow is the same by uniqueness, see [14, Remark 2.7].)

Remark 2.7 (i) In [36, Theorem 7] the contractivity (from the geodesical convexity) is shown on spaces satisfying the *local angle condition*. Alexandrov spaces satisfy this condition.

(ii) The first author [13, Theorem 15] proved the uniqueness of the gradient flow of the relative entropy on general metric measure spaces satisfying $\mathsf{CD}(K, \infty)$, without relying on the contractivity. In this generality, the contractivity seems to fail (e.g., in normed spaces).

In the sequel, we mainly study the gradient flow of the relative entropy. Then (2.13) allows us to continuously and uniquely extend the gradient flow semigroup to the full $\mathscr{P}(X)$ (since the closure of $\mathscr{P}^*(X)$ is $\mathscr{P}(X)$). Such an extension also satisfies (2.13).

3 Identification of the two gradient flows

This section contains our main result:

Theorem 3.1 (Identification of the two gradient flows) *Let (X, d) be a compact n-dimensional Alexandrov space. For any $\nu \in \mathscr{P}(X)$, the gradient flow of the Dirichlet energy starting from ν is the gradient flow of the relative entropy, and vice versa.*

Recall that $\mathsf{CD}(K, \infty)$ holds with $K = (n-1)k$, and we use this, e.g., in the proof of Proposition 3.2 below. The main technical obstacle in the proof of this theorem is to let the L^2 and the W_2 structures 'interact'. Our strategy consists in picking a gradient flow of the Dirichlet energy and in proving that it obeys the energy dissipation identity (2.12) for the relative entropy. We start with a bound on the local slope (2.9) by the Fisher information.

Proposition 3.2 (Bound on the local slope) *Let $\mu = f\mathcal{H}^n \in \mathscr{P}^*(X)$ with $f \in W^{1,2}(X)$. Then it holds*

$$|\nabla_- \operatorname{Ent}|^2(\mu) \leq \int_X \frac{\langle \nabla f, \nabla f \rangle}{f} \, d\mathcal{H}^n.$$

12

Proof. We first assume that f is Lipschitz and bounded away from 0. Then we know that $\langle \nabla f, \nabla f \rangle = |\nabla_L f|^2$ $\mathcal{H}^n$-a.e., so that the conclusion follows from [40, Theorem 20.1] together with $\mathsf{CD}(K, \infty)$. Thus all we need to do is to proceed by approximation. Suppose that $0 < c \leq f$ $\mathcal{H}^n$-a.e. for some $c \in \mathbb{R}$. Since $C^{\mathrm{Lip}}(X)$ is dense in $W^{1,2}(X)$ and $\sqrt{f} \in W^{1,2}(X)$, we can find a sequence $\{g_i\}_{i \in \mathbb{N}}$ of Lipschitz functions such that g_i converges to $\sqrt{f}$ as $i \to \infty$ with respect to the Sobolev norm. Substituting $\max\{C + g_i, \sqrt{c}\}$ for g_i for some $C \in \mathbb{R}$ if necessary, we can assume that $\sqrt{c} \leq g_i$ $\mathcal{H}^n$-a.e. as well as $\|g_i\|_{L^2} = 1$ for all i. Set $f_i := g_i^2$. As f_i is Lipschitz and bounded away from 0, we have

$$|\nabla_- \operatorname{Ent}|^2(f_i \mathcal{H}^n) \leq \int_X \frac{\langle \nabla f_i, \nabla f_i \rangle}{f_i} \, d\mathcal{H}^n.$$

On the one hand, the right-hand side is equal to $4\mathcal{E}(g_i, g_i)$ and converges to $4\mathcal{E}(\sqrt{f}, \sqrt{f}) = \int_X \langle \nabla f, \nabla f \rangle / f \, d\mathcal{H}^n$ by construction. On the other hand, we deduce from [1, Corollary 2.4.10] that

$$|\nabla_- \operatorname{Ent}|(f \mathcal{H}^n) \leq \varliminf_{i \to \infty} |\nabla_- \operatorname{Ent}|(f_i \mathcal{H}^n). \tag{3.1}$$

Combining these, we prove the claim for f. It remains to remove the assumption that f is away from 0. To do this we just consider $f_i := (1 + i^{-1})^{-1}(f + i^{-1})$ and apply (3.1) again. $\qquad\square$

Let us turn to considering the gradient flow of the Dirichlet energy $(T_t \nu)_{t \in [0, \infty)}$. For simplicity of notations, we write $f_t = dT_t \nu / d\mathcal{H}^n$ in the remainder of the section. By (2.3), $f_t \in C(X) \subset L^2(X, \mathcal{H}^n)$ holds for $t > 0$. Then (2.5), (2.4) and (2.3) imply $f_t \in D(\Delta)$. For any $t > 0$, there is $\varepsilon_t > 0$ such that $f_t \geq \varepsilon_t$ holds since $p_t(x, y)$ is positive and continuous. As a well-known fact, the following bound also holds.

Lemma 3.3 (Maximum principle) *Let $f_0 \in L^2(X, \mathcal{H}^n)$. If $f_0 \geq c$ a.e., then $T_t f_0 \geq c$ a.e. for every $t \geq 0$. The same holds for bounds from above.*

Proof. Take $A \subset X$ measurable. Since T_t is Markovian, $T_t 1_A \geq 0$ $\mathcal{H}^n$-a.e.. Thus

$$\int_A (T_t f_0 - c) \, d\mathcal{H}^n = \int_A T_t(f_0 - c) \, d\mathcal{H}^n = \int_X T_t 1_A \cdot (f_0 - c) \, d\mathcal{H}^n \geq 0.$$

Since A is arbitrary, the assertion holds. Bounds from above follow by applying the same argument to $-f_0$. $\qquad\square$

By virtue of Lemma 3.3 with the remark before it, for any $\varepsilon > 0$, there are $c, C > 0$ such that

$$c \leq f_t \leq C \quad \text{for all } t \geq \varepsilon. \tag{3.2}$$

13

Proposition 3.4 (Entropy dissipation) *The function $t \mapsto \mathrm{Ent}(f_t \mathcal{H}^n)$ is locally Lipschitz in $(0, \infty)$ and, moreover, it holds*

$$\frac{d}{dt} \mathrm{Ent}(f_t \mathcal{H}^n) = -\int_X \frac{\langle \nabla f_t, \nabla f_t \rangle}{f_t} \, d\mathcal{H}^n \qquad a.e. \ t.$$

Proof. As the function $e(s) := s \log s$ is C^1 in $[c, C]$ and $t \mapsto f_t$ is locally Lipschitz in $(0, \infty)$ with values in $L^2(X, \mathcal{H}^n)$, we deduce from (3.2) that $t \mapsto \mathrm{Ent}(f_t \mathcal{H}^n)$ is locally Lipschitz in $(0, \infty)$. Applying formulas (2.2) and (2.7) and recalling $\int_X \Delta f_t \, d\mathcal{H}^n = 0$, we obtain

$$\frac{d}{dt} \mathrm{Ent}(f_t \mathcal{H}^n) = \int_X e'(f_t) \Delta f_t \, d\mathcal{H}^n = \int_X \big(\log(f_t) + 1 \big) \Delta f_t \, d\mathcal{H}^n$$
$$= -\int_X \langle \nabla \log(f_t), \nabla f_t \rangle \, d\mathcal{H}^n = -\int_X \frac{\langle \nabla f_t, \nabla f_t \rangle}{f_t} \, d\mathcal{H}^n.$$

$\square$

For the next argument, we briefly recall some properties of the Hamilton-Jacobi semigroup in our context. For $f \in C^{\mathrm{Lip}}(X)$ and $t > 0$, we define $Q_t f : X \to \mathbb{R}$ by

$$Q_t f(x) := \inf_{y \in X} \left[f(y) + \frac{d^2(x, y)}{2t} \right]. \tag{3.3}$$

Also, set $Q_0 f := f$. The following is shown in [3, Theorem 2.5(iv)] and [25, Theorem 2.5(viii)] in the framework of general metric measure spaces supporting the volume doubling condition and the Poincaré inequality for upper gradients.

Proposition 3.5 (Hamilton-Jacobi semigroup) *It holds $Q_t f \in C^{\mathrm{Lip}}(X)$ for every $t \geq 0$, the map $[0, \infty) \ni t \mapsto Q_t f \in C(X)$ is Lipschitz in the uniform norm, and*

$$\frac{d}{dt} Q_t f(x) + \frac{|\nabla_L Q_t f|^2(x)}{2} = 0, \tag{3.4}$$

for almost every t, x.

Remark 3.6 The equation (3.3) has been called the Hopf-Lax formula or the Moreau-Yosida approximation also in the literature. Some properties of $Q_t f$ are known in much more general framework. For instance, it was known to de Giorgi that the Moreau-Yosida approximation leads to the "$\geq$" inequality in (3.4) ([9], see [25, Theorem 2.5 (vi)] also).

Proposition 3.7 (Absolute continuity with respect to W_2) *The curve $t \mapsto T_t\nu$ is absolutely continuous in the space $(\mathscr{P}(X), W_2)$ and its metric speed $|\dot{T_t\nu}|$ satisfies*

$$|\dot{T_t\nu}|^2 \le \int_X \frac{\langle \nabla f_t, \nabla f_t \rangle}{f_t} \, d\mathcal{H}^n, \qquad for \ a.e. \ t. \tag{3.5}$$

Proof. Fix $t, s > 0$. By the Kantorovich duality (cf., e.g., [1, Theorem 6.1.1] and [40, Theorem 5.10]) together with (3.3), we obtain

$$\frac{1}{2}W_2^2\big(T_t\nu, T_{t+s}\nu\big) = \sup_{\varphi \in C^{\mathrm{Lip}}(X)} \left[\int_X (Q_1\varphi) f_{t+s} \, d\mathcal{H}^n - \int_X \varphi f_t \, d\mathcal{H}^n \right]. \tag{3.6}$$

By Proposition 3.5, the map $r \mapsto Q_r\varphi$ from $[0,1]$ to $L^2(X, \mathcal{H}^n)$ is Lipschitz. Moreover, it is differentiable in $L^2(X, \mathcal{H}^n)$ and the derivative is determined by (3.4). The curve $[0,1] \ni r \mapsto f_{t+rs} \in L^2(X, \mathcal{H}^n)$ is Lipschitz as well. Thus $[0,1] \ni r \mapsto (Q_r\varphi) f_{t+rs} \in L^1(X, \mathcal{H}^n)$ is Lipschitz and its derivative can be calculated with the Leibniz rule. Thus we have

$$\int_X (Q_1\varphi) f_{t+s} \, d\mathcal{H}^n - \int_X \varphi f_t \, d\mathcal{H}^n$$
$$= \int_0^1 \frac{d}{dr} \left[\int_X (Q_r\varphi) f_{t+rs} \, d\mathcal{H}^n \right] dr$$
$$= \int_0^1 \int_X \left(-\frac{|\nabla_L Q_r\varphi|^2}{2} f_{t+rs} + s(Q_r\varphi)\Delta f_{t+rs} \right) d\mathcal{H}^n \, dr. \tag{3.7}$$

Using formulas (2.2), (3.2) and the trivial inequality

$$-\langle \nabla g, \nabla \tilde{g} \rangle \le \frac{1}{2s} \langle \nabla g, \nabla g \rangle + \frac{s}{2} \langle \nabla \tilde{g}, \nabla \tilde{g} \rangle, \qquad \mathcal{H}^n\text{-a.e. for } s > 0,$$

we have

$$\int_X (Q_r\varphi)\Delta f_{t+rs} \, d\mathcal{H}^n$$
$$= -\int_X \langle \nabla Q_r\varphi, \nabla f_{t+rs} \rangle \, d\mathcal{H}^n$$
$$= -\int_X \left\langle \nabla Q_r\varphi, \frac{\nabla f_{t+rs}}{f_{t+rs}} \right\rangle f_{t+rs} \, d\mathcal{H}^n$$
$$\le \frac{1}{2s} \int_X \langle \nabla Q_r\varphi, \nabla Q_r\varphi \rangle f_{t+rs} \, d\mathcal{H}^n + \frac{s}{2} \int_X \frac{\langle \nabla f_{t+rs}, \nabla f_{t+rs} \rangle}{f_{t+rs}} \, d\mathcal{H}^n.$$

Plugging this inequality in (3.7) and recalling that $\langle \nabla Q_r\varphi, \nabla Q_r\varphi \rangle = |\nabla_L Q_r\varphi|^2$ $\mathcal{H}^n$-a.e. (since $Q_r\varphi$ is Lipschitz), we obtain

$$\int_X (Q_1\varphi) f_{t+s} \, d\mathcal{H}^n - \int_X \varphi f_t \, d\mathcal{H}^n \le \frac{s^2}{2} \int_0^1 \int_X \frac{\langle \nabla f_{t+rs}, \nabla f_{t+rs} \rangle}{f_{t+rs}} \, d\mathcal{H}^n \, dr.$$

This bound does not depend on φ, thus from (3.6) we deduce

$$W_2^2\big(T_t\nu, T_{t+s}\nu\big) \le s^2 \int_0^1 \int_X \frac{\langle \nabla f_{t+rs}, \nabla f_{t+rs}\rangle}{f_{t+rs}}\, d\mathcal{H}^n\, dr. \tag{3.8}$$

Since we have (3.2) and the Dirichlet energy decreases along the flow $t \mapsto f_t$, we obtain

$$W_2^2\big(T_t\nu, T_{t+s}\nu\big) \le \frac{s^2}{c} \int_0^1 \int_X \langle \nabla f_{t+rs}, \nabla f_{t+rs}\rangle\, d\mathcal{H}^n\, dr \le \frac{s^2}{c} \int_X \langle \nabla f_t, \nabla f_t\rangle\, d\mathcal{H}^n,$$

which gives that the map $t \mapsto T_t\nu \in \mathscr{P}(X)$ is locally Lipschitz. The bound (3.5) follows directly from (3.8) (by recalling the definition of the absolutely continuous curves (2.8)). $\qquad\square$

Now we are ready to prove Theorem 3.1.

Proof of Theorem 3.1. By Propositions 3.7, 3.2 and 3.4, we have

$$\frac{1}{2}|f_t\dot{\mathcal{H}}^n|^2 + \frac{1}{2}|\nabla_- \operatorname{Ent}|^2(f_t\mathcal{H}^n) \le \int_X \frac{\langle \nabla f_t, \nabla f_t\rangle}{f_t}\, d\mathcal{H}^n = -\frac{d}{dt}\operatorname{Ent}(f_t\mathcal{H}^n)$$

a.e. t. As the reverse inequality (2.11) is always true, equality (2.12) holds for all $0 < t \le s$. Since $(f_t\mathcal{H}^n)_{t\in[0,\infty)}$ is continuous also at $t = 0$, it is the gradient flow of the relative entropy. The converse immediately follows from the uniqueness of the gradient flow of the relative entropy. $\qquad\square$

Remark 3.8 (i) When $\nu = f_0\mathcal{H}^n$ with $f_0 \in L^2(X, \mathcal{H}^n)$, we can give a proof of Theorem 3.1 without relying on the positivity improving property (3.2). Indeed, by virtue of the contraction property of both flows, it is possible to use Propositions 3.4, 3.7 only for $0 < c \le f_0 \le C$ and prove the theorem via approximation. It suggests that our argument possibly works in a more general framework where the existence of the density $p_t(x, y)$ does not follow from the theory of Dirichlet forms. In such a case, we could 'construct' $p_t(x, \cdot)$ as the gradient flow starting from the Dirac measure $\delta_x \in \mathscr{P}(X)$ (via the contraction property).

(ii) In the proof of Theorem 3.1, we heavily rely on the fact $\langle \nabla f, \nabla f\rangle = |\nabla_L f|^2$ for $f \in C^{\mathrm{Lip}}(X)$ for which we used the local structure of Alexandrov spaces. Actually, we use $|\nabla_L f|$ in Propositions 3.2, 3.7 and $|\nabla f|$ in Propositions 3.4, 3.7.

4 Applications

In this section, we assume that (X, d) is a compact Alexandrov space satisfying $\mathsf{CD}(K, \infty)$, and prove some applications of Theorem 3.1. It should be stressed

that the 'sectional curvature bound' k in the sense of Alexandrov appears nowhere in the sequel, and the 'Ricci curvature bound' K is essential instead (recall Remark 2.5). Indeed, those results involving K are natural extensions of the corresponding ones on a Riemannian manifold with $\mathrm{Ric} \geq K$. In this sense, the emergence of K instead of k is natural and gives sharper estimates.

Since the gradient flow of the Dirichlet energy $T_t \nu$ is obviously linear and symmetric, we immediately obtain the following:

Theorem 4.1 (Linearity and symmetry) *For $\nu \in \mathscr{P}(X)$, let $(\mu_t^\nu)_{t \geq 0}$ be the gradient flow of the relative entropy on $\mathscr{P}(X)$ with $\mu_0^\nu = \nu$. Then the following hold true.*

(i) *For $\nu_0, \nu_1 \in \mathscr{P}(X)$, $\lambda \in [0,1]$ and $t \geq 0$,*

$$\mu_t^{(1-\lambda)\nu_0 + \lambda\nu_1} = (1-\lambda)\mu_t^{\nu_0} + \lambda\mu_t^{\nu_1}.$$

(ii) *For $f, g \in L^1(X, \mathcal{H}^n)$ with $f, g \geq 0$ and $\|f\|_{L^1} = \|g\|_{L^1} = 1$,*

$$\int_X f \, d\mu_t^{g\mathcal{H}^n} = \int_X g \, d\mu_t^{f\mathcal{H}^n}.$$

We remark that the linearity, but not symmetry, also follows from the gradient flow approach under the local angle condition ([36, Theorem 8]). In general, these properties are completely nontrivial, and the linearity indeed fails in the Finsler setting ([28]).

A new property for $(T_t \nu)_{t \geq 0, \nu \in \mathscr{P}(X)}$ coming from our identification with the gradient flow of the relative entropy is the L^2-Wasserstein contraction (2.13). Together with [19, Corollary 3.4], we obtain the following:

Theorem 4.2 (Contraction for the heat flow) *For $1 \leq p \leq 2$,*

$$W_p(T_t \nu_0, T_t \nu_1) \leq e^{-Kt} W_p(\nu_0, \nu_1) \tag{4.1}$$

holds for every $\nu_0, \nu_1 \in \mathscr{P}(X)$.

Furthermore, by the duality result [19, Theorem 2.2], Theorem 4.2 yields the following Bakry-Émery type L^2-gradient estimate

$$|\nabla_L T_t f|(x) \leq e^{-Kt} T_t(|\nabla_L f|^2)(x)^{1/2} \tag{4.2}$$

for any $f \in C^{\mathrm{Lip}}(X)$ and $x \in X$. By combining (4.2) with the regularity of the heat kernel, we can extend it to $f \in W^{1,2}(X)$ as follows:

Theorem 4.3 (Gradient estimate) *Let $f \in W^{1,2}(X)$ and $t > 0$. Then $T_t f \in C^{\mathrm{Lip}}(X)$ and*

$$|\nabla_L T_t f|(x) \leq \mathrm{e}^{-Kt} T_t(|\nabla f|^2)(x)^{1/2} \tag{4.3}$$

holds for all $x \in X$. In particular,

$$|\nabla T_t f|(x) \leq \mathrm{e}^{-Kt} T_t(|\nabla f|^2)(x)^{1/2} \quad \text{for a.e. } x. \tag{4.4}$$

Proof.   Take $\{f_i\}_{i\in\mathbb{N}} \subset C^{\mathrm{Lip}}(X)$ such that $f_i \to f$ in $W^{1,2}(X)$ as $i \to \infty$. Then (4.2) yields $T_t f_i \in C^{\mathrm{Lip}}(X)$. Let $y \in X$ and $\gamma : [0,l] \to X$ a unit speed minimal geodesic from x to y. Since $|\nabla_L T_t f_i|$ is an upper gradient of $T_t f_i$ (see [8, Proposition 1.11] for instance), we have

$$|T_t f_i(y) - T_t f_i(x)| \leq \int_0^l |\nabla_L T_t f_i|(\gamma(s)) \, ds$$

$$\leq \mathrm{e}^{-Kt} \int_0^l T_t(|\nabla_L f_i|^2)(\gamma(s))^{1/2} \, ds, \tag{4.5}$$

where the second inequality follows from (4.2). Thanks to (2.3) and the boundedness of p_t, $T_t f_i$ converges pointwisely to $T_t f$ as $i \to \infty$. Since $T_t(|\nabla_L f_i|^2)(z) = T_t(|\nabla f_i|^2)(z)$ for $z \in X$, $T_t(|\nabla_L f_i|^2)$ converges pointwisely to $T_t(|\nabla f|^2)$ in a similar manner. Thus, by letting $i \to \infty$ in (4.5), we obtain

$$|T_t f(x) - T_t f(y)| \leq \mathrm{e}^{-Kt} \int_0^l T_t(|\nabla f|^2)(\gamma(s))^{1/2} \, ds. \tag{4.6}$$

The boundedness of p_t yields that there is $C > 0$ satisfying

$$T_t g(z) \leq C \|g\|_{L^1(X, \mathcal{H}^n)} \tag{4.7}$$

for all $g \in L^1(X, \mathcal{H}^n)$ and $z \in X$. Since $l = d(x,y)$, the estimate (4.7) for $g = |\nabla f|^2$ together with (4.6) implies $T_t f \in C^{\mathrm{Lip}}(X)$. In order to show (4.3), choose a sequence $\{y_i\}_{i\in\mathbb{N}}$ in X so that $y_i \to x$ as $i \to \infty$ and

$$\lim_{i\to\infty} \frac{|T_t f(x) - T_t f(y_i)|}{d(x, y_i)} = |\nabla_L T_t f|(x).$$

By the continuity of p_t, $T_t(|\nabla f|^2) \in C(X)$ holds. Thus, applying (4.6) for $y = y_i$, dividing both sides of it by $d(x, y_i)$ and letting $i \to \infty$ yield (4.3). $\square$

As an easy but important consequence of Theorem 4.3, we obtain the Lipschitz continuity of the heat kernel $p_t(x, y)$ as well as that of eigenfunctions. Recall that $-\Delta$ has discrete spectrum consisting of nonnegative eigenvalues with finite multiplicity [20, Corollary 1.1].

Theorem 4.4 (Lipschitz continuity)   (i) *For $\nu \in \mathscr{P}(X)$ and $t > 0$, let $f_t = dT_t\nu/d\mathcal{H}^n$. Then $f_t \in C^{\mathrm{Lip}}(X)$ for $t > 0$. In particular, we have $p_t(x, \cdot) \in C^{\mathrm{Lip}}(X)$ and $T_t f \in C^{\mathrm{Lip}}(X)$ for all $x \in X$ and $f \in L^1(X, \mathcal{H}^n)$.*

(ii) *Let f be an L^2-eigenfunction of Δ. Then $f \in C^{\mathrm{Lip}}(X)$.*

Proof.   Since $f_t \in W^{1,2}(X)$, Theorem 4.3 yields that $f_t = T_{t/2}f_{t/2} \in C^{\mathrm{Lip}}(X)$. For the second assertion, take $\lambda \geq 0$ so that $\Delta f = -\lambda f$. Then $f = \mathrm{e}^{\lambda t}T_t f$ and hence the first assertion yields the conclusion. $\square$

Remark 4.5 (i) It has been known that the heat kernel $p_t(x, y)$ is (locally) Hölder continuous of some fractional exponent, that follows from the parabolic Harnack inequality shown in [20].

(ii) The existence and the continuity of p_t are used in the proof of Theorems 4.3, 4.4 in an essential way (cf. Remark 3.8).

In what follows, we consider two additional applications of the Bakry-Émery gradient estimate (4.4) by employing the Lipschitz continuity of $T_t f$. The first one is the following inequality:

Theorem 4.6 (Γ_2-condition) *Let $f \in D(\Delta)$ with $\Delta f \in W^{1,2}(X)$. Then, for $g \in D(\Delta) \cap L^\infty(X, \mathcal{H}^n)$ with $g \geq 0$ and $\Delta g \in L^\infty(X, \mathcal{H}^n)$, we have*

$$\frac{1}{2}\int_X \Delta g \, \langle \nabla f, \nabla f \rangle \, d\mathcal{H}^n - \int_X g \, \langle \nabla \Delta f, \nabla f \rangle \, d\mathcal{H}^n \geq K \int_X g \, \langle \nabla f, \nabla f \rangle \, d\mathcal{H}^n. \quad (4.8)$$

Remark 4.7 (i) By virtue of the analyticity of T_t and Theorem 4.4, we have $T_t f \in D(\Delta^{m+1})$ and $\Delta^m T_t f \in C^{\mathrm{Lip}}(X)$ for any $f \in L^2(X, \mathcal{H}^n)$, $t > 0$ and $m \geq 0$. Thus there are fairly many f and g's satisfying the condition in Theorem 4.6.

(ii) The inequality (4.8) is nothing but a weak form of the Γ_2-condition

$$\Gamma_2(f, f) := \frac{1}{2}\big\{\Delta(\langle \nabla f, \nabla f \rangle) - 2\langle \nabla f, \nabla \Delta f \rangle \big\} \geq K \langle \nabla f, \nabla f \rangle. \quad (4.9)$$

This inequality is known to be equivalent to (4.4) in an abstract framework (see [2, 23] and references therein, for instance). However, the assumption involves the existence of a nice core $\mathcal{A} \subset D(\Delta)$ and it seems hopeless to verify it on Alexandrov spaces. When X is a complete Riemannian manifold, the inequality (4.9) is equivalent to $\mathrm{Ric} \geq K$. Indeed,

$$\Gamma_2(f, f) = \mathrm{Ric}(\nabla f, \nabla f) + |\operatorname{Hess} f|^2$$

holds by the Bochner identity.

Proof. We first show the claim for $f \in D(\Delta) \cap C^{\mathrm{Lip}}(X)$ with $\Delta f \in D(\Delta) \cap L^\infty(X, \mathcal{H}^n)$. By (4.4), we obtain

$$\int_X g \langle \nabla T_t f, \nabla T_t f \rangle \, d\mathcal{H}^n \leq \mathrm{e}^{-2Kt} \int_X g T_t(\langle \nabla f, \nabla f \rangle) \, d\mathcal{H}^n. \tag{4.10}$$

The derivation property (2.7) yields that, for $t \geq 0$,

$$\begin{aligned}
\int_X g \langle \nabla T_t f, \nabla T_t f \rangle \, d\mathcal{H}^n &= \int_X \langle \nabla(g T_t f), \nabla T_t f \rangle \, d\mathcal{H}^n - \int_X T_t f \langle \nabla g, \nabla T_t f \rangle \, d\mathcal{H}^n \\
&= -\int_X g T_t f \Delta T_t f \, d\mathcal{H}^n - \frac{1}{2} \int_X \langle \nabla g, \nabla (T_t f)^2 \rangle \, d\mathcal{H}^n \\
&= -\int_X g T_t f T_t \Delta f \, d\mathcal{H}^n + \frac{1}{2} \int_X \Delta g (T_t f)^2 \, d\mathcal{H}^n. \tag{4.11}
\end{aligned}$$

Hence we obtain

$$\begin{aligned}
\frac{d}{dt} &\int_X g \langle \nabla T_t f, \nabla T_t f \rangle \, d\mathcal{H}^n \bigg|_{t=0} \\
&= -\int_X g (\Delta f)^2 d\mathcal{H}^n - \int_X g f \Delta^2 f d\mathcal{H}^n + \int_X \Delta g f \Delta f \, d\mathcal{H}^n \\
&= \int_X \langle \nabla(g \Delta f), \nabla f \rangle \, d\mathcal{H}^n + \int_X \langle \nabla(g f), \nabla \Delta f \rangle \, d\mathcal{H}^n - \int_X \langle \nabla g, \nabla(f \Delta f) \rangle \, d\mathcal{H}^n \\
&= 2 \int_X g \langle \nabla \Delta f, \nabla f \rangle \, d\mathcal{H}^n \tag{4.12}
\end{aligned}$$

by using the derivation property again. Since $f \in C^{\mathrm{Lip}}(X)$, $\langle \nabla f, \nabla f \rangle = |\nabla_L f|^2 \in L^\infty(X, \mathcal{H}^n)$ holds. Hence we have

$$\begin{aligned}
\frac{d}{dt} &\left(\mathrm{e}^{-2Kt} \int_X g T_t(\langle \nabla f, \nabla f \rangle) \, d\mathcal{H}^n \right) \bigg|_{t=0} \\
&= \frac{d}{dt} \left(\mathrm{e}^{-2Kt} \int_X (T_t g) \langle \nabla f, \nabla f \rangle \, d\mathcal{H}^n \right) \bigg|_{t=0} \\
&= \int_X \Delta g \langle \nabla f, \nabla f \rangle \, d\mathcal{H}^n - 2K \int_X g \langle \nabla f, \nabla f \rangle \, d\mathcal{H}^n. \tag{4.13}
\end{aligned}$$

Since (4.10) implies

$$\frac{d}{dt} \int_X g \langle \nabla T_t f, \nabla T_t f \rangle \, d\mathcal{H}^n \bigg|_{t=0} \leq \frac{d}{dt} \left(\mathrm{e}^{-2Kt} \int_X g T_t(\langle \nabla f, \nabla f \rangle) \, d\mathcal{H}^n \right) \bigg|_{t=0}, \tag{4.14}$$

we obtain (4.8) by combining (4.14) with (4.12) and (4.13).

Next we consider the case that $f \in D(\Delta)$ with $\Delta f \in W^{1,2}(X)$. Then, by the above discussion, $T_\delta f$ and g satisfy (4.8) for $\delta > 0$ (cf. Remark 4.7(i)). Since $g, \Delta g \in L^\infty(X, \mathcal{H}^n)$ and $\Delta T_\delta f = T_\delta \Delta f$, it suffices to show the claim that $\lim_{\delta \to 0} \langle \nabla T_\delta h, \nabla T_\delta h' \rangle = \langle \nabla h, \nabla h' \rangle$ weakly in $L^1(X, \mathcal{H}^n)$ for $h, h' \in W^{1,2}(X)$. By polarization, we may assume $h = h'$. The spectral decomposition yields

$$\lim_{\delta \to 0} \mathcal{E}(T_\delta h - h, T_\delta h - h) = 0 \tag{4.15}$$

(see [12, Lemma 1.3.3], for instance). Let $\psi \in L^\infty(X, \mathcal{H}^n)$. Then the Schwarz inequality yields

$$\left| \int_X \psi \langle \nabla T_\delta h, \nabla T_\delta h \rangle \, d\mathcal{H}^n - \int_X \psi \langle \nabla h, \nabla h \rangle \, d\mathcal{H}^n \right|$$

$$\leq \left(\left(\int_X \psi^2 |\nabla T_\delta h|^2 d\mathcal{H}^n \right)^{1/2} + \left(\int_X \psi^2 |\nabla h|^2 d\mathcal{H}^n \right)^{1/2} \right) \mathcal{E}(T_\delta h - h, T_\delta h - h)^{1/2}$$

$$\leq 2 \|\psi\|_{L^\infty} \mathcal{E}(h, h)^{1/2} \mathcal{E}(T_\delta h - h, T_\delta h - h)^{1/2}. \tag{4.16}$$

Hence the desired claim follows from (4.15) and (4.16). □

While we proved the implication from (4.1) with $p = 2$ to (4.4) and (4.8), these conditions are equivalent to each other on complete Riemannian manifolds (see [35]). Such an equivalence still holds in our framework under *a priori presence of* $\mathsf{CD}(K, \infty)$, where K can be different from what appears in these conditions. More precisely, we show the following:

Theorem 4.8 (Equivalence of "Ricci curvature bound" inequalities) *Given* $K_0 \in \mathbb{R}$, *the following are equivalent.*

(i) (4.1) *holds for* $\nu_0, \nu_1 \in \mathscr{P}(X)$ *and* $t \geq 0$, *with* $p = 2$ *and* $K = K_0$.

(ii) (4.4) *holds for* $f \in W^{1,2}(X)$ *and* $t \geq 0$ *with* $K = K_0$.

(iii) (4.8) *holds with* $K = K_0$ *for* $f \in D(\Delta)$ *with* $\Delta f \in W^{1,2}(X)$ *and* $g \in D(\Delta) \cap L^\infty(X, \mathcal{H}^n)$ *with* $g \geq 0$ *and* $\Delta g \in L^\infty(X, \mathcal{H}^n)$.

Proof. "(i) $\Rightarrow$ (ii) $\Rightarrow$ (iii)" is already shown in Theorem 4.3 and Theorem 4.6.

For "(iii) $\Rightarrow$ (ii)", it follows from a standard argument of the so-called Γ_2-calculus (see [2, 23] for instance). For completeness, we give a sketch of the proof. Take $g_0 \in L^\infty(X, \mathcal{H}^n)$ with $g_0 \geq 0$ arbitrary and let $g = T_\delta g_0$. Then $g \in D(\Delta) \cap L^\infty(X, \mathcal{H}^n)$ with $g \geq 0$ and $\Delta g \in L^\infty(X, \mathcal{H}^n)$. Let us define $\Psi : [0, t] \to \mathbb{R}$ by

$$\Psi(s) := \int_X g T_s(|\nabla T_{t-s} f|^2) \, d\mathcal{H}^n.$$

By a similar calculation as in (4.11), we can easily prove that Ψ is continuous on $[0,t)$ and C^1 on $(0,t)$. A similar argument as in (4.16) yields that Ψ is continuous at t. Here we use the L^1-strong continuity of T_t (see [5, Proposition I.2.4.2] for instance). A similar calculation as in the proof of Theorem 4.3 together with the assumption in (iii) leads to the inequality $\Psi'(s) \geq 2K_0\Psi(s)$. Hence (4.4) follows by integrating it.

For "(ii) $\Rightarrow$ (i)", we follow the argument in the proof of [19, Theorem 2.2] by means of the Hamilton-Jacobi semigroup introduced in (3.3). Since (4.4) is weaker than (4.2) in the sense that it holds only *almost everywhere*, we need to modify a part of the original argument. It suffices to show (4.1) for those $\nu_0, \nu_1 \in \mathscr{P}(X)$ which are absolutely continuous to $\mathcal{H}^n$. By the Kantorovich duality, we have

$$
\begin{aligned}
\frac{1}{2}W_2(T_t\nu_0, T_t\nu_1)^2 &= \sup_{\varphi \in C^{\mathrm{Lip}}(X)} \left[\int_X Q_1\varphi \, d(T_t\nu_1) - \int_X \varphi \, d(T_t\nu_0) \right] \\
&= \sup_{\varphi \in C^{\mathrm{Lip}}(X)} \left[\int_X T_t(Q_1\varphi) \, d\nu_1 - \int_X T_t\varphi \, d\nu_0 \right], \qquad (4.17)
\end{aligned}
$$

where $Q_1\varphi$ is given by (3.3). Take a geodesic ν_t joining ν_0 and ν_1 in $(\mathscr{P}(X), W_2)$. Recall that ν_s is absolutely continuous to $\mathcal{H}^n$. Let us define Γ as the set of minimal geodesics $\gamma : [0,1] \to X$, and $e_s : \Gamma \to X$ for $s \in [0,1]$ by $e_s(\gamma) := \gamma(s)$. Note that there is a probability measure Π on Γ such that the push-forward of Π by e_s coincides with ν_s for $s \in [0,1]$ and $\int_\Gamma d(e_0(\gamma), e_1(\gamma))^2 \, \Pi(d\gamma) = W_2(\nu_0, \nu_1)^2$ holds (see [26, Section 2], and [4] for a more precise description on Alexandrov spaces). For $\varphi \in C^{\mathrm{Lip}}(X)$, let us define $\Phi(s)$ by

$$
\Phi(s) := \int_X T_t(Q_s\varphi) \, d\nu_s.
$$

Let $M > 0$ be the (global) Lipschitz constant for $[0,1] \ni s \to Q_s\varphi \in C(X)$. By Proposition 3.5, $M < \infty$ holds. Then, by Theorem 4.4 together with (4.2),

$$
\begin{aligned}
\varlimsup_{s' \to s} \frac{|\Phi(s') - \Phi(s)|}{s' - s} &\leq \varlimsup_{s' \to s} \int_\Gamma \left| \frac{T_t(Q_s\varphi)(e_{s'}(\gamma)) - T_t(Q_s\varphi)(e_s(\gamma))}{s' - s} \right| \Pi(d\gamma) \\
&\quad + \varlimsup_{s' \to s} \int_X T_t \left| \frac{Q_{s'}\varphi - Q_s\varphi}{s' - s} \right| d\nu_{s'} \\
&\leq \int_\Gamma |\nabla_L T_t(Q_s\varphi)|(e_s(\gamma)) d(e_0(\gamma), e_1(\gamma)) \, \Pi(d\gamma) + M \\
&\leq e^{-Kt} \left(\int_X T_t(|\nabla_L Q_s\varphi|^2) d\nu_s \right)^{1/2} W_2(\nu_0, \nu_1) + M. \qquad (4.18)
\end{aligned}
$$

22

Thus $\Phi(s)$ is Lipschitz since Proposition 3.5 ensures $|\nabla_L Q_s\varphi| \in L^\infty(X, \mathcal{H}^n)$. By a similar calculation as we did in (4.18), we can verify that we can apply the Leibniz rule for calculating $\Phi'(s)$. Since $T_t(Q_s\varphi) \in C^{\mathrm{Lip}}(X)$, we have $|\nabla_L T_t(Q_s\varphi)| = |\nabla T_t(Q_s\varphi)|$. Hence our assumption in (ii) together with Proposition 3.5 yields

$$
\begin{aligned}
\Phi'(s) &= \int_\Gamma |\nabla T_t(Q_s\varphi)|(e_s(\gamma))d(e_0(\gamma), e_1(\gamma))\,\Pi(d\gamma) - \int_x T_t\left(\frac{|\nabla_L Q_s\varphi|^2}{2}\right)d\nu_s \\
&\leq -\frac{1}{2}\int_x T_t(|\nabla Q_s\varphi|^2)d\nu_s + \mathrm{e}^{-K_0 t}\left(\int_X T_t(|\nabla Q_s\varphi|^2)\,d\nu_s\right)^{1/2} W_2(\nu_0, \nu_1) \\
&\leq \frac{1}{2}\mathrm{e}^{-2K_0 t}W_2(\nu_0, \nu_1)^2.
\end{aligned}
\tag{4.19}
$$

Therefore we obtain

$$
\begin{aligned}
\frac{1}{2}W_2(T_t\nu_0, T_t\nu_1)^2 &= \sup_{\varphi \in C^{\mathrm{Lip}}(X)}(\Phi(1) - \Phi(0)) \\
&= \sup_{\varphi \in C^{\mathrm{Lip}}(X)}\int_0^1 \Phi'(s)ds \leq \frac{1}{2}\mathrm{e}^{-2K_0 t}W_2(\nu_0, \nu_1)^2
\end{aligned}
$$

and the conclusion follows. $\qquad\square$

As the second application of (4.4), we mention that (4.4) together with Theorem 4.4 implies some functional inequalities by means of [17, Theorem 1.3]. Since T_t is Markovian, we can extend T_t to a contraction on $L^\infty(X, \mathcal{H}^n)$. Then we can further extend T_t to a contraction on $L^p(X, \mathcal{H}^n)$ for any $1 \leq p \leq \infty$ by interpolation and the symmetry of T_t. Let us denote the infinitesimal generator of T_t in $L^p(X, \mathcal{H}^n)$ by Δ_p. Let us define $R_\alpha^{(q)}f := |\nabla((\alpha - \Delta_p)^{-q/2}f)|$.

Corollary 4.9 *Let $2 \leq p < \infty$, $q > 1$ and $\alpha > \max\{(-K), 0\}$. Then we have the following:*

(i) *There exists a constant $C_R > 0$ which depends only on p, q and $\max\{(\alpha + K), \alpha\}$ such that*
$$\|R_\alpha^{(q)}f\|_{L^p} \leq C_R\|f\|_{L^p}$$
for $f \in L^p(X, \mathcal{H}^n)$.

(ii) *Suppose $q < 2$. Then there exists $C_{p,q} > 0$ such that*
$$\||\nabla T_t f|\|_{L^p} \leq C_{p,q}\|R_\alpha^{(q)}\|_{L^p \to L^p}\left(\alpha^{q/2} + t^{-q/2}\right)\|f\|_{L^p}$$
for $t > 0$ and $f \in L^p(X, \mathcal{H}^n)$.

23

Proof. It is sufficient to verify the assumption of [17, Theorem 1.3]. Set $\mathcal{A} := C^{\mathrm{Lip}}(X) \cap D(\Delta_2)$. Since we already know (4.4), we only need to show the following claim: $\mathcal{A}$ is dense in $W^{1,2}(X)$ and $f^2 \in D(\Delta_1)$ holds for any $f \in \mathcal{A}$. By Theorem 4.4, $T_t f \in \mathcal{A}$ holds. Thus $\mathcal{A}$ is dense in $W^{1,2}(X)$ since $\mathcal{E}(T_t f - f, T_t f - f) \to 0$ as $t \to 0$ (cf. (4.15)). By [5, Proposition I.2.4.2], it is enough to prove $f^2 \in D(\Delta_2)$ for $f \in \mathcal{A}$. Take $g \in W^{1,2}(X)$ arbitrary. Recall that $|\nabla f|^2 = |\nabla_L f|^2 \in L^\infty(X, \mathcal{H}^n)$. The derivation property yields

$$
\begin{aligned}
|\mathcal{E}(f^2, g)| &= 2 \left| \int_X f \langle \nabla f, \nabla g \rangle \, d\mathcal{H}^n \right| \\
&= 2 \left| \int_X (\langle \nabla f, \nabla(gf) \rangle - g \langle \nabla f, \nabla f \rangle \, d\mathcal{H}^n \right| \\
&= 2 \left| \int_X (f \Delta f - \langle \nabla f, \nabla f \rangle) g \, d\mathcal{H}^n \right| \\
&\le 2 \left(\|f\|_{L^\infty} \|\Delta f\|_{L^2} + \| |\nabla f|^2 \|_{L^2} \right) \|g\|_{L^2}.
\end{aligned}
$$

This estimate means $f^2 \in D(\Delta_2)$ and hence the proof is completed. $\qquad\square$

Finally, we observe that all of our results are generalized to the heat equation with drift, in other words, the Fokker-Planck equation. Given a potential function $V \in C^{\mathrm{Lip}}(X)$, we modify the Dirichlet energy and the relative entropy into

$$
\mathcal{E}^V(f, g) = \int_X \langle \nabla f, \nabla g \rangle \, \mathrm{e}^{-V} d\mathcal{H}^n \qquad f, g \in W^{1,2}(X),
$$

$$
\mathrm{Ent}^V(\mu) = \mathrm{Ent}(\mu) + \int_X V \, d\mu \qquad \mu \in \mathscr{P}(X).
$$

We regard $\mathcal{E}^V$ as a bilinear form on $L^2(X, \mathrm{e}^{-V} \mathcal{H}^n)$. Observe that Ent^V is nothing but the relative entropy with respect to $\mathrm{e}^{-V} \mathcal{H}^n$. Note that the semigroup T_t^V on $L^2(X, \mathrm{e}^{-V} \mathcal{H}^n)$ associated with $\mathcal{E}^V$ solves the following diffusion equation

$$
\frac{d}{dt} u = \Delta u - \langle \nabla V, \nabla u \rangle. \tag{4.20}
$$

Since e^{-V} is bounded and away from 0, $\mathrm{e}^{-V} \mathcal{H}^n$ is equivalent to $\mathcal{H}^n$. Hence $(X, d, \mathrm{e}^{-V} \mathcal{H}^n)$ satisfies the volume doubling condition as well as the Poincaré inequality for upper gradients. Moreover, as $\mathcal{E}$ and $\mathcal{E}^V$ are equivalent, the Poincaré inequality for $\mathcal{E}^V$ is also valid. Therefore a continuous density p_t^V for T_t^V exists. Under the assumption that Ent^V is K-geodesically convex, we can apply the general theory of the gradient flow on $(\mathscr{P}(X), W_2)$ to obtain the gradient flow μ_t of Ent^V. Furthermore, every argument in Sections 3, 4 works verbatim and gives similar results for the equation (4.20). Note that, under $\mathsf{CD}(K, \infty)$ for $(X, d, \mathcal{H}^n)$, the K'-convexity of V implies the $(K + K')$-convexity of Ent^V.

References

[1] L. Ambrosio, N. Gigli, and G. Savaré, *Gradient flows in metric spaces and in the space of probability measures.* Second edition, Birkhäuser Verlag, Basel, 2008.

[2] D. Bakry, *On Sobolev and logarithmic Sobolev inequalities for Markov semigroups*, New trends in stochasitic analysis (Charingworth, 1994), World Sci. Publ. River Edge, NJ, 1997, pp. 43–75.

[3] Z.M. Balogh, A. Engoulatov, L. Hunziker, and O.E. Maasalo, *Functional inequalities and Hamilton-Jacobi equations in geodesic spaces*, Preprint. Available at arXiv:0906.0476

[4] J. Bertrand, *Existence and uniqueness of optimal maps on Alexandrov spaces*, Adv. Math. **219** (2008), 838–851.

[5] N. Bouleau and F. Hirsch, *Dirichlet forms and analysis on Wiener space*, de Gruyter Studies in Mathematics, **14**, Walter de Gruyter & Co., Berlin, 1991.

[6] D. Burago, Yu. Burago, and S. Ivanov, *A course in metric geometry*, Graduate studies in mathematics, **33**, American mathematical society, Providence, RI, 2001.

[7] 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.

[8] J. Cheeger, *Differentiability of Lipschitz functions on metric measure spaces*, Geom. Funct. Anal. **9** (1999), no. 3, 428–517.

[9] E. De Giorgi, A. Marino, and M. Tosques, *Problems of evolution in metric spaces and maximal decreasing curve* (Italian), Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Natur. (8), **68** (1980), pp. 180–187.

[10] S. Fang, J. Shao, and K.-T. Sturm, *Wasserstein space over the Wiener space*, Probab. Theory Related Fields **146** (2010), no. 3-4, 535–565.

[11] A. Figalli and N. Juillet, *Absolute continuity of Wasserstein geodesics in the Heisenberg group*, J. Funct. Anal. **255** (2008), no. 1, 133–141.

[12] M. Fukushima, Y. Oshima, and M. Takeda, *Dirichlet forms and symmetric Markov processes*, de Gruyter Studies in Mathematics, **19**, Walter de Gruyter & Co., Berlin, 1994.

[13] N. Gigli, *On the heat flow on metric measure spaces: existence, uniqueness and stability*, Calc. Var. Partial Differential Equations **39** (2010), 101–120.

[14] N. Gigli and S. Ohta, *First variation formula in Wasserstein spaces over compact Alexandrov spaces*, to appear in Canad. Math. Bull.

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

[16] N. Juillet, *Diffusion by optimal transport in Heisenberg groups*, Preprint. Available at http://www-irma.u-strasbg.fr/~juillet/

[17] H. Kawabi and T. Miyokawa, *The Littlewood-Paley-Stein inequality for diffusion processes on general metric spaces*, J. Math. Sci. Univ. Tokyo **14** (2007), no. 1, 1–30.

[18] N. J. Korevaar and R. M. Schoen, *Sobolev spaces and harmonic maps for metric space targets*, Comm. Anal. Geom. **1** (1993), 561–659.

[19] K. Kuwada, *Duality on gradient estimates and Wasserstein controls*, J. Funct. Anal. **258** (2010), no. 11, 3758–3774.

[20] K. Kuwae, Y. Machigashira, and T. Shioya, *Sobolev spaces, Laplacian and heat kernel on Alexandrov spaces*, Math. Z. **238** (2001), no. 2, 269–316.

[21] K. Kuwae and T. Shioya, *On generalized measure contraction property and energy functionals over Lipschitz maps*, Potential Anal. **15** (2001), no. 1, 105–121.

[22] _____, *Sobolev spaces and Dirichlet spaces over maps between metric spaces*, J. Reine Angew. Math. **555** (2003), 39–75.

[23] M. Ledoux, *The geometry of Markov diffusion generators*, Ann. Fac. Sci. Toulouse Math. (6) **9** (2000), no. 2, 305–366.

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

[25] _____, *Hamilton-Jacobi semigroup in length spaces and applications*, J. Math. Pures Appl. **88**, (2007), 219–229.

[26] _____, *Ricci curvature for metric-measure spaces via optimal transport*, Ann. of Math. **169** (2009), 903–991.

[27] S. Ohta, *Gradient flows on Wasserstein spaces over compact Alexandrov spaces*, Amer. J. Math **131** (2009), no. 2, 475–516.

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

[29] Y. Otsu and T. Shioya, *The Riemannian structure of Alexandrov spaces*, J. Differential Geom. **39** (1994), no. 3, 629–658.

[30] F. Otto, *The geometry of dissipative evolution equations: the porous medium equation*, Comm. Partial Differential Equations **26**, 1–2 (2001), 101–174.

[31] G. Perelman, *DC structure on Alexandrov space with curvature bounded below*, Available at http://www.math.psu.edu/petrunin/

[32] A. Petrunin, *Harmonic functions on Alexandrov spaces and their applications*, Electron. Res. Announc. Amer. Math. Soc. **9** (2003), 135–141.

[33] _____, *Alexandrov meets Lott–Villani–Sturm*, Preprint (2009). Available at arXiv:1003.5948

[34] A. Ranjbar-Motlagh, *Poincaré inequality for abstract spaces*, Bull. Austral. Math. Soc. **71** (2005), no. 2, 193–204.

[35] M.-K. von Renesse and K.-T. Sturm, *Transport inequalities, gradient estimates, entropy and Ricci curvature*, Comm. Pure Appl. Math. **58** (2005), 923–940.

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

[37] K.-T. Sturm, *Analysis on local Dirichlet spaces.III. The parabolic Harnack inequality*, J. Math. Pures Appl. (9) **75** (1996), no. 3, 273–297.

[38] _____, *On the geometry of metric measure spaces. I*, Acta Math. **196** (2006), 65–131.

[39] _____, *On the geometry of metric measure spaces. II*, Acta Math. **196** (2006), 133–177.

[40] C. Villani, *Optimal Transport, old and new*, Grundlehren der mathematischen Wissenschaften collection, 2009.

[41] H.-C. Zhang and X.-P. Zhu *Ricci curvature on Alexandrov spaces and rigidity theorems*, Preprint (2009). Available at arXiv:0912.3190

[42] _____ , *On a new definition of Ricci curvature on Alexandrov spaces*, Preprint (2010). Available at arXiv:1007.4042