Non-contraction of heat flow on Minkowski spaces

Shin-ichi Ohta*and Karl-Theodor Sturm†

October 20, 2010

Abstract

We study contractivity properties of gradient flows for functions on normed spaces or, more generally, on Finsler manifolds. Contractivity of the flows turns out to be equivalent to a new notion of convexity for the functions. This is different from the usual convexity along geodesics in Finsler manifolds. As an application, we show that heat flow on Minkowski normed spaces other than inner product spaces is not contractive with respect to the quadratic Wasserstein distance.

1 Introduction

The main goal of this article is to prove that, for the heat flow on a Minkowski normed space, no bound for the exponential growth of the L^2-Wasserstein distance exists, unless the space is an inner product space. This is rather surprising, in particular, in view of the fact that the heat flow is the gradient flow in the L^2-Wasserstein space $\mathcal{P}_2$ of the relative entropy and the fact that the latter is known to be a convex function on $\mathcal{P}_2$. In order to find an explanation for this phenomenon, we will first of all study the contraction of the gradient flow of a function on a Finsler manifold. A Finsler manifold is a manifold carrying a Minkowski norm on each tangent space, instead of inner products for Riemannian manifolds. A Minkowski norm is a generalization of usual norms, and is not necessarily centrally symmetric.

In Riemannian manifolds, given $K \in \mathbb{R}$, it is well-known that the K-convexity of a function f along geodesics γ (i.e., $(f \circ \gamma)'' \ge K$ in the weak sense) implies the *K-contraction* of the gradient flow of f, namely

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

holds for all $t \ge 0$ and ξ, ζ solving $\dot{\xi}(t) = \nabla(-f)(\xi(t))$, $\dot{\zeta}(t) = \nabla(-f)(\zeta(t))$. This is obtained via the first variation formulas for the distance $d(\xi(t), \zeta(t))$ and the function f. In Finsler manifolds or even in normed spaces, however, it has been unclear whether the gradient flows of convex functions are contractive (cf. [AGS, Introduction]).

*Department of Mathematics, Kyoto University, Kyoto 606-8502, Japan (sohta@math.kyoto-u.ac.jp) & Max-Planck-Intitut für Mathematik, Vivatsgasse 7, 53111 Bonn, Germany. Supported in part by the Grant-in-Aid for Young Scientists (B) 20740036.

†Institut für Angewandte Mathematik, Universität Bonn, Endenicher Allee 60, 53115 Bonn, Germany (sturm@uni-bonn.de).

The point is that, although these variational formulas do exist also in the Finsler setting, they use different approximate inner products (see the paragraph following Definition 3.1). Keeping this in mind, we introduce a new notion of convexity, called the *skew convexity*, which is equivalent to the usual convexity in Riemannian manifolds. We show that the K-skew convexity of a function on a Finsler manifold is equivalent to the K-contraction of its gradient flow (Theorem 3.2). A difference between the skew convexity and the convexity along geodesics is observed by considering distance functions (Section 4). For instance, the squared norm is 2-skew convex in every Minkowski space, while it is only K-convex for some $K \geq 0$.

In the second part of the article, we apply our technique to heat flow on Minkowski spaces. Due to celebrated work of Jordan et al [JKO], heat flow on Euclidean spaces is regarded as the gradient flow of the relative entropy in the L^2-Wasserstein space. This provides a somewhat geometric interpretation of the non-expansion (0-contraction) of heat flow with respect to the Wasserstein distance, as the relative entropy is known to be convex (also called *displacement convex*, [Mc]). More generally, on Riemannian manifolds, both the K-convexity of the relative entropy and the K-contraction of heat flow are equivalent to the lower Ricci curvature bound $\mathrm{Ric} \geq K$ ([vRS]). Note that the Wasserstein space over a Riemannian manifold possesses a sort of Riemannian structure, for which the first variation formulas are available (see [Ot], [Oh1], [GO], [Sav], [Vi]). We also remark that Gigli [Gi] recently showed the uniqueness of the gradient flow of the relative entropy (with respect to a probability measure) for metric measure spaces such that the relative entropy is K-convex for some $K \in \mathbb{R}$, without relying on the contractivity.

In our previous work [Oh3], [OS], we have extended the equivalence between the Ricci curvature bound and the convexity of the relative entropy, as well as the identification of (nonlinear) heat flow with the gradient flow of the relative entropy with respect to the reverse Wasserstein distance, to (compact) Finsler manifolds. In particular, the relative entropy on any Minkowski space is convex (see also [Vi, page 908]). Then it is natural to ask whether heat flow on Minkowski spaces is contractive or not. Our main result gives a complete answer to this question.

Theorem 1.1 *The heat flow on a Minkowski space $(\mathbb{R}^n, \|\cdot\|)$ is not K-contractive with respect to the reverse L^2-Wasserstein distance for any $K \in \mathbb{R}$, unless $(\mathbb{R}^n, \|\cdot\|)$ is an inner product space.*

Our proof uses a geometric characterization of inner products among Minkowski norms (Claims 6.1, 6.2).

Theorem 1.1 means that the Wasserstein contraction implies that the space must be Riemannian. This makes a contrast with the aforementioned fact that the convexity of the relative entropy (as well as the curvature-dimension condition) works well for general Finsler manifolds. Among other characterizations of lower Ricci curvature bounds for Riemannian manifolds, the Bochner formula/inequality would make sense in some, but not all non-Riemannian Finsler manifolds (work in progress of the authors).

The article is organized as follows. After preliminaries for Minkowski and Finsler geometries, we introduce the skew convexity in Section 3, and study the skew convexity of distance functions in Section 4. In Section 5, we discuss heat flow on Minkowski spaces. We give a detailed explanation on how to identify it with the gradient flow of the relative entropy, because some results in [OS] are not directly applicable to noncompact spaces. Finally, Section 6 is devoted to a proof of Theorem 1.1. We remark that only Subsection 2.1 and Section 5 are necessary for understanding Section 6.

2 Preliminaries

We review the basics of Minkowski spaces and Finsler manifolds. We refer to [BCS] and [Sh] for Finsler geometry, and to [BCS, Chapter 14] for Minkowski spaces.

2.1 Minkowski spaces

In this article, a *Minkowski norm* will mean a nonnegative function $\|\cdot\| : \mathbb{R}^n \longrightarrow [0, \infty)$ satisfying the following conditions.

(1) (Positive homogeneity) $\|cx\| = c\|x\|$ holds for all $x \in \mathbb{R}^n$ and $c \geq 0$.

(2) (Strong convexity) The function $\|\cdot\|^2/2$ is twice differentiable on $\mathbb{R}^n \setminus \{0\}$, and the symmetric matrix

$$\big(g_{ij}(x)\big)_{i,j=1}^n := \left(\frac{1}{2}\frac{\partial^2(\|\cdot\|^2)}{\partial x^i \partial x^j}(x)\right)_{i,j=1}^n \tag{2.1}$$

is measurable in x and uniformly elliptic in the sense that there are constants $\lambda, \Lambda > 0$ such that

$$\lambda \sum_{i=1}^n (a^i)^2 \leq \sum_{i,j=1}^n g_{ij}(x) a^i a^j \leq \Lambda \sum_{i=1}^n (a^i)^2 \tag{2.2}$$

holds for all $x \in \mathbb{R}^n \setminus \{0\}$ and $(a^i) \in \mathbb{R}^n$. In particular, $\|x\| > 0$ for all $x \neq 0$.

We call $(\mathbb{R}^n, \|\cdot\|)$ a *Minkowski space*. Note that the homogeneity is imposed only in positive direction, so that $\|-x\| \neq \|x\|$ is allowed. We also remark that the function $\|\cdot\|^2/2$ is twice differentiable at the origin only in inner product spaces. Given $x \in \mathbb{R}^n \setminus \{0\}$, the matrix (2.1) defines the inner product g_x of $\mathbb{R}^n$ by

$$g_x\big((a^i), (b^j)\big) := \sum_{i,j=1}^n g_{ij}(x) a^i b^j. \tag{2.3}$$

This is the best approximation of the norm in the direction x in the sense that the unit sphere of g_x is tangent to that of $\|\cdot\|$ at $x/\|x\|$ up to the second order (Figure 1). In particular, we have $g_x(x, x) = \|x\|^2$. If the original norm comes from an inner product, then g_x coincides with it for all x.

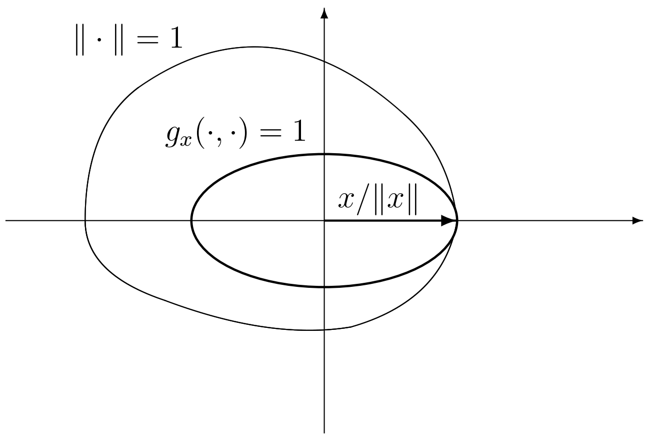

Figure 1

We define the 2-*uniform convexity* and *smoothness constants* $\mathcal{C}, \mathcal{S} \in [1, \infty)$ as the least constants satisfying

$$\left\| \frac{x+y}{2} \right\|^2 \le \frac{1}{2}\|x\|^2 + \frac{1}{2}\|y\|^2 - \frac{1}{4\mathcal{C}^2}\|x-y\|^2,$$
$$\left\| \frac{x+y}{2} \right\|^2 \ge \frac{1}{2}\|x\|^2 + \frac{1}{2}\|y\|^2 - \frac{\mathcal{S}^2}{4}\|x-y\|^2$$

for all $x, y \in \mathbb{R}^n$. In other words, $\mathcal{C}^{-2}$ and $\mathcal{S}^2$ are the moduli of convexity and concavity of $\|\cdot\|^2/2$, respectively. Thanks to (2.2), $\mathcal{C} < \infty$ and $\mathcal{S} < \infty$ hold. Indeed, we know

$$\mathcal{C} = \sup_{x,y \in \mathbb{R}^n \setminus \{0\}} \frac{\|y\|}{g_x(y,y)^{1/2}}, \qquad \mathcal{S} = \sup_{x,y \in \mathbb{R}^n \setminus \{0\}} \frac{g_x(y,y)^{1/2}}{\|y\|} \tag{2.4}$$

(cf. [Oh2, Proposition 4.6]). Note also that $\mathcal{C} = 1$ or $\mathcal{S} = 1$ holds if and only if the norm is an inner product.

We denote by $\|\cdot\|_*$ the dual norm of $\|\cdot\|$. The *Legendre transform* $\mathcal{L} : (\mathbb{R}^n, \|\cdot\|) \longrightarrow (\mathbb{R}^n, \|\cdot\|_*)$ associates x with $\mathcal{L}(x)$ satisfying $\|\mathcal{L}(x)\|_* = \|x\|$ and $[\mathcal{L}(x)](x) = \|x\|^2$. Note that (2.2) ensures that $\mathcal{L}(x)$ is indeed uniquely determined, and it is explicitly written as

$$\mathcal{L}(x) = \left(\sum_{i=1}^n g_{ij}(x) x^i \right)_{j=1}^n . \tag{2.5}$$

The Legendre transform of inverse direction $\mathcal{L}^* : (\mathbb{R}^n, \|\cdot\|_*) \longrightarrow (\mathbb{R}^n, \|\cdot\|)$ is nothing but the inverse map $\mathcal{L}^* = \mathcal{L}^{-1}$ by definition. For a function $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ and $x \in \mathbb{R}^n$ at where f is differentiable, we define the *gradient vector* of f at x by $\nabla f(x) := \mathcal{L}^*(Df(x)) \in T_x\mathbb{R}^n$.

Remark 2.1 We need the strong convexity to formulate and investigate the skew convexity as well as the heat equation, while the characterization of inner products (Claim 6.2) is valid among merely 'convex' Minkowski norms (i.e., its closed unit ball is a closed convex set containing the origin as an inner point).

2.2 Finsler manifolds

Let M be a connected C^∞-manifold without boundary. A nonnegative function $F : TM \longrightarrow [0,\infty)$ is called a C^∞-*Finsler structure* if it is C^∞ on $TM \setminus \{0\}$ ($\{0\}$ stands for the zero section) and if $F|_{T_xM}$ is a Minkowski norm for all $x \in M$. We call (M,F) a C^∞-Finsler manifold. (We will consider only C^∞-structures for simplicity.)

For each $v \in T_xM \setminus \{0\}$, we define the inner product g_v on T_xM like (2.3). That is to say, given a local coordinate $(x^i)_{i=1}^n$ on an open set U containing x, we consider the coordinate of T_xM as $v = \sum_{i=1}^n v^i (\partial/\partial x^i)|_x$ and define

$$g_{ij}(v) := \frac{1}{2}\frac{\partial^2(F^2)}{\partial v^i \partial v^j}(v), \quad g_v\bigg(\sum_{i=1}^n a_i \frac{\partial}{\partial x^i}\Big|_x, \sum_{j=1}^n b_j \frac{\partial}{\partial x^j}\Big|_x\bigg) := \sum_{i,j=1}^n a_i b_j g_{ij}(v).$$

We denote by $\mathcal{C}(x)$ and $\mathcal{S}(x)$ the 2-uniform convexity and smoothness constants of $F|_{T_xM}$, respectively. For a function $f : M \longrightarrow \mathbb{R}$ differentiable at $x \in M$, define the gradient vector of f at x by $\nabla f(x) := \mathcal{L}^*(Df(x))$ via the Legendre transform $\mathcal{L}^* : T_x^*M \longrightarrow T_xM$.

The *distance* from x to y is naturally defined as $d(x,y) := \inf_\gamma \int_0^1 F(\dot\gamma)\,dt$, where $\gamma : [0,1] \longrightarrow M$ runs over all differentiable curves from x to y. We remark that d is nonsymmetric in general, namely $d(y,x) \neq d(x,y)$ may happen. A *geodesic* $\gamma : [0,l] \longrightarrow M$ is a locally length minimizing curve of constant speed (i.e., $F(\dot\gamma)$ is constant). We say that (M,F) is *forward complete* if any geodesic $\gamma : [0,l] \longrightarrow M$ is extended to a geodesic $\gamma : [0,\infty) \longrightarrow M$. Then, for any $x,y \in M$, there is a minimal geodesic from x to y.

Along a geodesic $\gamma : [0,l] \longrightarrow M$, $\gamma(s)$ with $s \in (0,l)$ is called a *cut point* of $\gamma(0)$ if $\gamma|_{[0,s]}$ is minimal and if $\gamma|_{[0,s+\varepsilon]}$ is not minimal for any $\varepsilon > 0$. Suppose that $\gamma(s)$ is not a cut point of $\gamma(0)$ for all $s \in (0,l]$, and let ξ and ζ be differentiable curves with $\xi(0) = \gamma(0)$ and $\zeta(0) = \gamma(l)$. Then we have the following first variation formula ([BCS, Exercise 5.2.4]):

$$\lim_{t \downarrow 0} \frac{d(\xi(t),\zeta(t)) - d(\xi(0),\zeta(0))}{t} = \frac{g_{\dot\gamma(l)}(\dot\gamma(l),\dot\zeta(0)) - g_{\dot\gamma(0)}(\dot\gamma(0),\dot\xi(0))}{l^{-1} \cdot d(\gamma(0),\gamma(l))}. \tag{2.6}$$

As usual in discussing the contraction property, this formula will play a vital role.

It is sometimes useful to consider the *reverse* Finsler structure $\overleftarrow{F}(v) := F(-v)$. We will put an arrow $\leftarrow$ on those associated with $\overleftarrow{F}$, for example, $\overleftarrow{d}(x,y) = d(y,x)$ and $\overleftarrow{\nabla} f = -\nabla(-f)$.

3 Skew convex functions

We introduce skew convex functions on a C^∞-Finsler manifold (M,F), and will see that it is equivalent to the contractivity of their gradient flows. Although we shall work with C^1-functions for simplicity, the same technique is applicable to other classes of functions (e.g., locally semi-convex functions, see Remark 3.3 below).

Let us begin with the standard notion of convexity along geodesics. A function $f : M \longrightarrow [-\infty,\infty]$ is said to be K-*convex* (or *geodesically* K-*convex*) for $K \in \mathbb{R}$ if

$$f\big(\gamma(t)\big) \le (1-t)f\big(\gamma(0)\big) + tf\big(\gamma(1)\big) - \frac{K}{2}(1-t)td\big(\gamma(0),\gamma(1)\big)^2$$

holds for all geodesics $\gamma : [0,1] \longrightarrow M$ and $t \geq 0$. If f is C^2, then this is equivalent to $\partial^2[f \circ \gamma]/\partial t^2 \geq Kd(\gamma(0),\gamma(1))^2$ and to

$$\frac{\partial}{\partial t}\Big[Df\big(\gamma(t)\big)\big(\dot\gamma(t)\big)\Big] \geq Kd\big(\gamma(0),\gamma(1)\big)^2.$$

Now, instead of $Df(\gamma)(\dot\gamma) = -g_{\nabla(-f)(\gamma)}(\nabla(-f)(\gamma),\dot\gamma)$ in the left hand side, we employ $-\mathcal{L}(\dot\gamma)(\nabla(-f)(\gamma)) = -g_{\dot\gamma}(\nabla(-f)(\gamma),\dot\gamma)$ for the skew convexity.

Definition 3.1 (Skew convex functions) Let $f : M \longrightarrow \mathbb{R}$ be a C^1-function. We say that f is K-*skew convex* for $K \in \mathbb{R}$ if, for any pair of distinct points $x, y \in M$, there is a minimal geodesic $\gamma : [0,1] \longrightarrow M$ from x to y such that

$$g_{\dot\gamma(1)}\big(\dot\gamma(1),\nabla(-f)(y)\big) - g_{\dot\gamma(0)}\big(\dot\gamma(0),\nabla(-f)(x)\big) \leq -Kd(x,y)^2. \tag{3.1}$$

Recall that, on a Riemannian manifold (M,g), it holds $g_{\dot\gamma} = g$ and (3.1) indeed implies the K-convexity of f. In the Finsler setting, however, $g_{\dot\gamma}$ is different from $g_{\nabla(-f)(\gamma)}$ (see Subsection 4.1). Compare (3.1) with (2.6).

For a C^1-function $f : M \longrightarrow \mathbb{R}$ and any point $x \in M$, there exists a C^1-curve $\xi : [0,\infty) \longrightarrow M$ satisfying $\xi(0) = x$ and $\dot\xi(t) = \nabla(-f)(\xi(t))$ for all t. We call such ξ a *gradient curve* of f. For $K \in \mathbb{R}$, we say that *the gradient flow of f is K-contractive* if

$$d\big(\xi(t),\zeta(t)\big) \leq e^{-Kt}d\big(\xi(0),\zeta(0)\big)$$

holds for all gradient curves ξ, ζ and $t \in [0,\infty)$.

Theorem 3.2 *Let (M,F) be a forward complete Finsler manifold, and let $f : M \longrightarrow \mathbb{R}$ be a C^1-function. Then the gradient flow of f is K-contractive if and only if f is K-skew convex.*

Proof. We first assume that f is K-skew convex. Fix two gradient curves $\xi, \zeta : [0,\infty) \longrightarrow M$ of f and set $l(t) := d(\xi(t),\zeta(t))$. Given $t > 0$, let $\gamma : [0,1] \longrightarrow M$ be a minimal geodesic from $\xi(t)$ to $\zeta(t)$ such that (3.1) holds. Note that $\gamma(1/2)$ ($\zeta(t)$, resp.) is not a cut point of $\xi(t)$ ($\gamma(1/2)$, resp.). Thus the first variation formula (2.6) shows that, together with the triangle inequality,

$$\begin{aligned} l'(t) &\leq \lim_{\varepsilon\downarrow 0}\frac{d(\xi(t+\varepsilon),\gamma(1/2)) - d(\xi(t),\gamma(1/2))}{\varepsilon} \\ &\quad + \lim_{\varepsilon\downarrow 0}\frac{d(\gamma(1/2),\zeta(t+\varepsilon)) - d(\gamma(1/2),\zeta(t))}{\varepsilon} \\ &= -g_{\dot\gamma(0)}\big(\dot\gamma(0)/l(t),\dot\xi(t)\big) + g_{\dot\gamma(1)}\big(\dot\gamma(1)/l(t),\dot\zeta(t)\big). \end{aligned}$$

By hypothesis, this yields $l'(t) \leq -Kl(t)$. Therefore $d(\xi(t),\zeta(t)) \leq e^{-Kt}d(\xi(0),\zeta(0))$ follows from Gronwall's lemma.

To see the converse, suppose that the gradient flow of f is K-contractive and take a minimal geodesic $\gamma : [0,1] \longrightarrow M$. Dividing γ into $\gamma|_{[0,1/2]}$ and $\gamma|_{[1/2,1]}$ if necessary, we can assume that $\gamma(s)$ is not a cut point of $\gamma(0)$ for all $s \in (0,1]$. Consider gradient curves

$\xi, \zeta : [0, \infty) \longrightarrow M$ of f with $\xi(0) = \gamma(0)$ and $\zeta(0) = \gamma(1)$, and put $l(t) := d(\xi(t), \zeta(t))$ again. Then we deduce from the assumption that

$$\frac{d}{dt}\Big|_{t=0+} \big[e^{Kt} l(t)\big] \le 0.$$

This immediately implies the K-skew convexity, as the first variation formula (2.6) shows

$$\begin{aligned}\frac{d}{dt}\Big|_{t=0+} \big[e^{Kt} l(t)\big] &= Kl(0) + \frac{d}{dt}\Big|_{t=0+} l(t) \\ &= Kl(0) + g_{\dot\gamma(1)}\big(\dot\gamma(1)/l(0), \dot\zeta(0)\big) - g_{\dot\gamma(0)}\big(\dot\gamma(0)/l(0), \dot\xi(0)\big).\end{aligned}$$

We complete the proof. □

Remark 3.3 We can replace the C^1-regularity in Definition 3.1 with the local semi-convexity as follows (cf., e.g., [Oh1] for details). We say that a function $f : M \longrightarrow \mathbb{R}$ is *locally semi-convex* if, for any $x \in M$, there are an open set $U \ni x$ and $K \in \mathbb{R}$ such that $f|_U$ is K-convex along any geodesic $\gamma : [0, 1] \longrightarrow U$ in the weak sense. Define the *local slope* of f at $x \in M$ as

$$|\nabla_{\!-} f|(x) := \limsup_{y \to x} \frac{\max\{f(x) - f(y), 0\}}{d(x, y)}. \tag{3.2}$$

For each $x \in M$ with $|\nabla_{\!-} f|(x) > 0$, there exists a unique unit vector $v \in T_x M$ satisfying $\lim_{t \downarrow 0}\{f(\gamma(t)) - f(x)\}/t = -|\nabla_{\!-} f|(x)$, where γ is the geodesic with $\dot\gamma(0) = v$. We define $\nabla_{\!-} f(x) := |\nabla_{\!-} f|(x) \cdot v$, and $\nabla_{\!-} f(x) := 0$ if $|\nabla_{\!-} f|(x) = 0$. Then there is a gradient curve (solving $\dot\xi(t) = \nabla_{\!-} f(\xi(t))$ at a.e. t) starting from any point. The K-skew convexity can be defined by using $\nabla_{\!-} f$ in (3.1) instead of $\nabla(-f)$, and the analogue of Theorem 3.2 holds by the same argument.

4 Skew convexity of distance functions

We study the skew convexity of the squared distance function. It is closely related to upper bounds of the sectional curvature in the Riemannian case. In our Finsler setting, we need two more quantities to control the distance function. See [Oh2] and [Sh] for related work on the usual convexity and concavity along geodesics.

4.1 In Minkowski spaces

Before considering general Finsler manifolds, we discuss Minkowski spaces. Let $(\mathbb{R}^n, \|\cdot\|)$ be a Minkowski space and set $f(x) := \|-x\|^2/2$. We observe $\nabla(-f)(x) = -x$, so that the gradient curve ξ of f with $\xi(0) = x$ is given by $\xi(t) = e^{-t}x$. Thus we see that the gradient flow of f is always 1-contractive, which shows that f is 1-skew convex. Indeed, for any $v \in \mathbb{R}^n \setminus \{0\}$, we have

$$g_v\big(v, \nabla(-f)(x + v)\big) - g_v\big(v, \nabla(-f)(x)\big) = -g_v(v, v) = -\|v\|^2.$$

In contrast, $\tilde f(x) := \langle x, x\rangle/2$ associated with the Euclidean inner product $\langle\cdot,\cdot\rangle$ of $\mathbb{R}^n$ may not be K-skew convex with respect to $\|\cdot\|$ even for negative K, though $\tilde f$ is convex along straight lines. To see this, we observe $D\tilde{f}(x) = x$ and $\nabla(-\tilde{f})(x) = \mathcal{L}^*(-x)$ by identifying both $T_x\mathbb{R}^n$ and $T_x^*\mathbb{R}^n$ with $\mathbb{R}^n$. Choosing $v = -x \neq 0$, we have

$$g_v\big(v, \nabla(-\tilde{f})(x+v)\big) - g_v\big(v, \nabla(-\tilde{f})(x)\big) = -g_v\big(v, \mathcal{L}^*(v)\big) = -\langle \mathcal{L}(v), \mathcal{L}^*(v)\rangle.$$

We used (2.5) in the last equality. However, $-\langle \mathcal{L}(v), \mathcal{L}^*(v)\rangle/\|v\|^2$ can be arbitrarily large for general $\|\cdot\|$. An example is illustrated in Figure 2. The rounded parallelogram drawn by thick line is the unit sphere of the original norm $\|\cdot\|$. By making the parallelogram longer (as drawn by thin line), we have $-\langle \mathcal{L}(v), \mathcal{L}^*(v)\rangle \to \infty$ while $\|v\| = 1$.

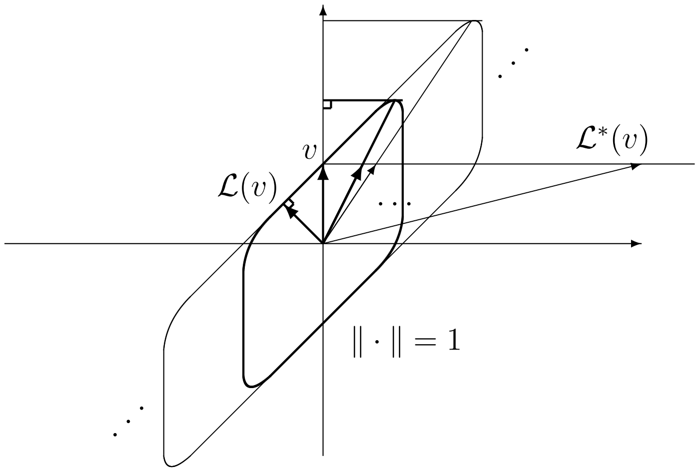

Figure 2

This observation (as well as Theorem 1.1) reveals that the skew convexity has no (obvious) relation with the usual convexity.

4.2 In Finsler manifolds

We need some more terminologies to discuss the Finsler case, see [BCS] for details. For a C^1-vector field X on M and tangent vectors $v, w \in T_xM$ with $w \neq 0$, we define the covariant derivative of X with reference vector w as

$$(D_v^w X)(x) := \sum_{i,j=1}^{n} \left\{ v^j \frac{\partial X^i}{\partial x^j}(x) + \sum_{k=1}^{n} \Gamma_{jk}^i(w) v^j X^k(x) \right\} \frac{\partial}{\partial x^i}\Big|_x,$$

where Γ_{jk}^i is the Christoffel symbol. If $\Gamma_{jk}^i(w)$ depends only on the point x and independent of the choice of $w \in T_xM \setminus \{0\}$ for all $x \in M$, then we say that (M, F) is of *Berwald type*. In a Berwald space, all tangent spaces are isometric to each other. For instance, Riemannian manifolds and Minkowski spaces are of Berwald type.

By using the covariant derivative, the geodesic equation is written in a canonical way as $D_{\dot\gamma}^{\dot\gamma}\dot\gamma \equiv 0$. We will use the following formula borrowed from [BCS, Exercise 10.1.2]:

$$\frac{d}{dt} g_V(V, W) = g_V(D_{\dot\gamma}^V V, W) + g_V(V, D_{\dot\gamma}^V W) \tag{4.1}$$

for any C^1-curve γ and C^1-vector fields V, W along γ.

A C^∞-vector field V along a geodesic $\gamma : [0, l] \longrightarrow M$ is called a *Jacobi field* if it satisfies the equation $D_{\dot\gamma}^{\dot\gamma} D_{\dot\gamma}^{\dot\gamma} V + \mathcal{R}(V, \dot\gamma)\dot\gamma \equiv 0$, where $\mathcal{R} : TM \otimes TM \longrightarrow T^*M \otimes TM$ is the curvature tensor. Similarly to the Riemannian case, the variational vector field of a geodesic variation is a Jacobi field (and vice versa). For linearly independent vectors $v, w \in T_xM$, the *flag curvature* is defined by

$$\mathcal{K}(v, w) := \frac{g_v(\mathcal{R}(v, w)w, v)}{F(v)^2 g_v(w, w) - g_v(v, w)^2}.$$

We remark that $\mathcal{K}(v, w)$ depends not only on the plane in T_xM spanned by v and w (*flag*), but also on the choice of v in it (*flagpole*).

In order to state our theorem, we introduce the condition

$$g_V(V, D_W^V D_W^V V - D_W^W D_W^W V) \ge -\delta F(V)^2 F(W)^2 \tag{4.2}$$

for C^∞-vector fields V, W and $\delta \ge 0$. This clearly holds with $\delta = 0$ for Berwald spaces. Therefore δ measures how the tangent spaces are distorted as one moves (in M) along w. The *injectivity radius* $\mathrm{inj}(z)$ at $z \in M$ is the supremum of $R > 0$ such that any unit speed geodesic $\gamma : [0, R) \longrightarrow M$ with $\gamma(0) = z$ contains no cut point of z. We set $B(x, r) := \{y \in M \mid d(x, y) < r\}$ for $x \in M$ and $r > 0$.

Theorem 4.1 *Let (M, F) be a forward complete Finsler manifold and suppose that $\mathcal{K} \le k$, $\mathcal{S} \le S$ and* (4.2) *hold for some $k \ge 0$, $S \ge 1$ and $\delta \ge 0$. Then the function $f(x) := d(x, z)^2/2$ is $K(k, S, \delta, r)$-skew convex in $\overleftarrow{B}(z, r)$ for all $z \in M$ and $r \in (0, R)$, where we set*

$$K(k, S, \delta, r) := \sqrt{kS^2 + \delta}\, r \cdot \cot(\sqrt{kS^2 + \delta}\, r)$$

and $R := \min\{\overleftarrow{\mathrm{inj}}(z), \pi/\sqrt{kS^2 + \delta}\}$. In particular, if $\mathcal{K} \le 0$, then f is $(\sqrt{\delta} r \cot(\sqrt{\delta} r))$-skew convex in $\overleftarrow{B}(z, r)$ for $r \in (0, \min\{\overleftarrow{\mathrm{inj}}(z), \pi/\sqrt{\delta}\})$ regardless $\mathcal{S}$.

Proof. Fix a unit speed minimal geodesic $\gamma : [0, l] \longrightarrow \overleftarrow{B}(z, r)$ with $r < R$, and let $\sigma : [0, l] \times [0, 1] \longrightarrow M$ be the C^∞-variation such that $\sigma_s := \sigma(s, \cdot)$ is the unique minimal geodesic from $\gamma(s)$ to z. Put $\mathcal{T}(s, t) := \partial_t \sigma(s, t)$ and $\mathcal{V}(s, t) := \partial_s \sigma(s, t)$. Observe that $\dot\gamma(s) = \mathcal{V}(s, 0)$ and $\nabla(-f)(\gamma(s)) = \mathcal{T}(s, 0)$. Hence we need to bound the following:

$$\frac{\partial}{\partial s}\big[g_{\mathcal{V}}\big(\mathcal{V}(s, 0), \mathcal{T}(s, 0)\big)\big] = g_{\mathcal{V}}\big(\mathcal{V}(s, 0), D_s^{\mathcal{V}} \mathcal{T}(s, 0)\big). \tag{4.3}$$

We used (4.1) and the geodesic equation $D_s^{\mathcal{V}} \mathcal{V}(s, 0) \equiv 0$. As $D_s^{\mathcal{V}} \mathcal{T} = D_t^{\mathcal{V}} \mathcal{V}$ (cf. [BCS, Exercise 5.2.1]), we deduce from (4.1) that

$$g_{\mathcal{V}}\big(\mathcal{V}(s, 0), D_s^{\mathcal{V}} \mathcal{T}(s, 0)\big) = \frac{1}{2}\frac{\partial}{\partial t}[g_{\mathcal{V}}(\mathcal{V}, \mathcal{V})](s, 0) = \frac{1}{2}\frac{\partial [F(\mathcal{V})^2]}{\partial t}(s, 0). \tag{4.4}$$

Again due to (4.1), we observe

$$\begin{aligned}\frac{\partial^2[F(\mathcal{V})]}{\partial t^2} &= \frac{\partial}{\partial t}\bigg[\frac{g_{\mathcal{V}}(\mathcal{V}, D_t^{\mathcal{V}}\mathcal{V})}{F(\mathcal{V})}\bigg] \\ &= \frac{g_{\mathcal{V}}(\mathcal{V}, D_t^{\mathcal{V}}D_t^{\mathcal{V}}\mathcal{V}) + g_{\mathcal{V}}(D_t^{\mathcal{V}}\mathcal{V}, D_t^{\mathcal{V}}\mathcal{V})}{F(\mathcal{V})} - \frac{g_{\mathcal{V}}(\mathcal{V}, D_t^{\mathcal{V}}\mathcal{V})^2}{F(\mathcal{V})^3} \\ &= \frac{g_{\mathcal{V}}(\mathcal{V}, D_t^{\mathcal{V}}D_t^{\mathcal{V}}\mathcal{V})}{F(\mathcal{V})} + \frac{F(\mathcal{V})^2 g_{\mathcal{V}}(D_t^{\mathcal{V}}\mathcal{V}, D_t^{\mathcal{V}}\mathcal{V}) - g_{\mathcal{V}}(\mathcal{V}, D_t^{\mathcal{V}}\mathcal{V})^2}{F(\mathcal{V})^3}.\end{aligned}$$

The second term is nonnegative by the Cauchy-Schwarz inequality. Moreover, by the assumption (4.2), we have

$$g_{\mathcal{V}}(\mathcal{V}, D_t^{\mathcal{V}}D_t^{\mathcal{V}}\mathcal{V}) \ge g_{\mathcal{V}}(\mathcal{V}, D_t^{\mathcal{T}}D_t^{\mathcal{T}}\mathcal{V}) - \delta F(\mathcal{V})^2 F(\mathcal{T})^2.$$

Since $\mathcal{V}(s,\cdot)$ is a Jacobi field, it holds $D_t^{\mathcal{T}}D_t^{\mathcal{T}}\mathcal{V} = -\mathcal{R}(\mathcal{V},\mathcal{T})\mathcal{T}$ and hence

$$\begin{aligned}\frac{\partial^2[F(\mathcal{V})]}{\partial t^2} &\ge -\mathcal{K}(\mathcal{V},\mathcal{T})\frac{F(\mathcal{V})^2 g_{\mathcal{V}}(\mathcal{T},\mathcal{T}) - g_{\mathcal{V}}(\mathcal{V},\mathcal{T})^2}{F(\mathcal{V})} - \delta F(\mathcal{V})F(\mathcal{T})^2 \\ &\ge -k\frac{F(\mathcal{V})^2 g_{\mathcal{V}}(\mathcal{T},\mathcal{T}) - g_{\mathcal{V}}(\mathcal{V},\mathcal{T})^2}{F(\mathcal{V})} - \delta F(\mathcal{V})F(\mathcal{T})^2.\end{aligned}$$

As $k \ge 0$, it follows from $\mathcal{S} \le S$ that (recall (2.4))

$$-k\{F(\mathcal{V})^2 g_{\mathcal{V}}(\mathcal{T},\mathcal{T}) - g_{\mathcal{V}}(\mathcal{V},\mathcal{T})^2\} \ge -kF(\mathcal{V})^2 g_{\mathcal{V}}(\mathcal{T},\mathcal{T}) \ge -kS^2F(\mathcal{V})^2F(\mathcal{T})^2.$$

Hence we obtain, together with $F(\mathcal{T}) \le r$,

$$\frac{\partial^2[F(\mathcal{V})]}{\partial t^2} \ge -(kS^2+\delta)r^2F(\mathcal{V}).$$

The above inequality shows that the function

$$\frac{\partial[F(\mathcal{V})]}{\partial t}\sin\big(\sqrt{kS^2+\delta}r(1-t)\big) - F(\mathcal{V})\frac{\partial}{\partial t}\big[\sin\big(\sqrt{kS^2+\delta}r(1-t)\big)\big]$$

is non-decreasing in $t \in [0,1]$, so that it is nonpositive for all t. Thus we have

$$\frac{1}{2}\frac{\partial[F(\mathcal{V})^2]}{\partial t} = F(\mathcal{V})\frac{\partial[F(\mathcal{V})]}{\partial t} \le -\sqrt{kS^2+\delta}r\cdot\cot\big(\sqrt{kS^2+\delta}r(1-t)\big)F(\mathcal{V})^2. \tag{4.5}$$

Combining (4.3), (4.4), (4.5) and $F(\mathcal{V}(s,0)) = F(\dot{\gamma}(s)) = 1$, we conclude

$$\frac{\partial}{\partial s}\big[g_{\mathcal{V}}\big(\mathcal{V}(s,0),\mathcal{T}(s,0)\big)\big] \le -\sqrt{kS^2+\delta}r\cdot\cot(\sqrt{kS^2+\delta}r).$$

This completes the proof. □

Interestingly enough, what appears in Theorem 4.1 is not the 2-uniform convexity constant $\mathcal{C}$, but the smoothness constant $\mathcal{S}$. Compare this with the usual convexity ([Oh2, Theorem 5.1]).

Corollary 4.2 *Assume that (M,F) is forward complete and of Berwald type, and that $\mathcal{K} \le k$ and $\mathcal{S} \le S$ hold for some $k \ge 0$ and $S \ge 1$. Then the function $f(x) := d(x,z)^2/2$ is $(\sqrt{k}Sr\cot(\sqrt{k}Sr))$-skew convex in $\overleftarrow{B}(z,r)$ for all $z \in M$ and $r \in (0,\min\{\overleftarrow{\mathrm{inj}}(z),\pi/\sqrt{k}S\})$. In particular, if $\mathcal{K} \le 0$, then f is 1-skew convex in $\overleftarrow{B}(z,\overleftarrow{\mathrm{inj}}(z))$ regardless $\mathcal{S}$.*

This recovers the 1-skew convexity of $f(x) = \|-x\|^2/2$ on Minkowski spaces.

5 Heat flow on Minkowski spaces

In order to apply our technique to heat flow on Minkowski spaces, we regard it as the gradient flow of the relative entropy with respect to the reverse Wasserstein distance. We refer to [AGS] and [Vi] for Wasserstein geometry as well as the gradient flow theory. Throughout the section, let $(\mathbb{R}^n, \|\cdot\|)$ be a Minkowski space in the sense of Subsection 2.1.

5.1 Wasserstein geometry over Minkowski spaces

Let $\mathcal{P}(\mathbb{R}^n)$ be the set of Borel probability measures on $\mathbb{R}^n$. Define $\mathcal{P}_2(\mathbb{R}^n) \subset \mathcal{P}(\mathbb{R}^n)$ as the set of measures μ such that $\int_{\mathbb{R}^n} \|x\|^2 \, d\mu < \infty$ (note that then $\int_{\mathbb{R}^n} \|-x\|^2 \, d\mu < \infty$ holds as well), where dx stands for the Lebesgue measure. The subset of absolutely continuous measures with respect to dx is denoted by $\mathcal{P}^{\mathrm{ac}}(\mathbb{R}^n) \subset \mathcal{P}(\mathbb{R}^n)$.

Given $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^n)$, a probability measure $\pi \in \mathcal{P}(\mathbb{R}^n \times \mathbb{R}^n)$ is called a *coupling* of μ and ν if $\pi(A \times \mathbb{R}^n) = \mu(A)$ and $\pi(\mathbb{R}^n \times A) = \nu(A)$ hold for all measurable sets $A \subset \mathbb{R}^n$. We define the L^2-*Wasserstein distance* from μ to ν by

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

where π runs over all couplings of μ and ν. We call $(\mathcal{P}_2(\mathbb{R}^n), W_2)$ the L^2-*Wasserstein space* over $(\mathbb{R}^n, \|\cdot\|)$.

Remark 5.1 (a) Thanks to (2.2), our norm is comparable to an inner product. In fact, (2.4) yields $\mathcal{C}^{-1}\|y\| \le \sqrt{g_x(y,y)} \le \mathcal{S}\|y\|$. Then, if we denote by $W_2^{g_x}$ the Wasserstein distance with respect to g_x, we have $\mathcal{C}^{-1} W_2(\mu,\nu) \le W_2^{g_x}(\mu,\nu) \le \mathcal{S} W_2(\mu,\nu)$. This relation is sometimes useful to apply known results in the Euclidean case.

(b) Unlike John's theorem for symmetric norms ($c \le n$), the least constant $c \ge 1$ satisfying $\|y\|^2 \le \langle y, y \rangle \le c\|y\|^2$ for some inner product $\langle \cdot, \cdot \rangle$ and all $y \in \mathbb{R}^n$ can not be bounded only by the dimension n. Consider the norm whose unit sphere is the standard unit sphere, but with the center $(1-\varepsilon, 0, \ldots, 0)$. Letting $\varepsilon \downarrow 0$, we have $c \to \infty$ (and $\mathcal{C}, \mathcal{S} \to \infty$).

For $\mu \in \mathcal{P}_2^{\mathrm{ac}}(\mathbb{R}^n)$ and $\nu \in \mathcal{P}_2(\mathbb{R}^n)$, there exists a semi-convex function φ on an open set $\Omega \subset \mathbb{R}^n$ with $\mu(\Omega) = 1$ such that $\pi := (\mathrm{id}_{\mathbb{R}^n} \times T_1)_\sharp \mu$ provides the unique optimal coupling of μ and ν, where we set $T_t(x) := x + t\nabla\varphi(x)$ for $t \in [0,1]$ (by, e.g., [Vi, Theorem 10.26] with **(locLip)**, **(SC)**, **(H∞)**). Moreover, $\mu_t := (T_t)_\sharp \mu$ is the unique minimal geodesic from $\mu_0 = \mu$ to $\mu_1 = \nu$. Note that φ is twice differentiable a.e. on Ω in the sense of Alexandrov, thus T_t is well-defined and differentiable a.e. on Ω.

We introduce a Finsler structure of the Wasserstein space along the line of [Ot], see [OS] for more details in the case of compact Finsler manifolds. We set

$$\hat{T}\mathcal{P} := \{\Phi = \nabla\varphi \,|\, \varphi \in C_c^\infty(\mathbb{R}^n)\}$$

and define the tangent space $(T_\mu\mathcal{P}, F_\mu)$ at $\mu \in \mathcal{P}_2(\mathbb{R}^n)$ as the completion of $\hat{T}\mathcal{P}$ with respect to the Minkowski norm

$$F_\mu(\Phi) := \left(\int_{\mathbb{R}^n} \|\Phi\|^2 \, d\mu \right)^{1/2}.$$

Similarly, the cotangent space $(T_\mu^*\mathcal{P}, F_\mu^*)$ is defined as the completion of $\hat{T}^*\mathcal{P} := \{\alpha = D\varphi \,|\, \varphi \in C_c^\infty(\mathbb{R}^n)\}$ with respect to $F_\mu^*(\alpha) := (\int_{\mathbb{R}^n} \|\alpha\|_*^2 \, d\mu)^{1/2}$. We define the Legendre transform $\mathcal{L}_\mu^* : T_\mu^*\mathcal{P} \longrightarrow T_\mu\mathcal{P}$ in the pointwise way that $\mathcal{L}_\mu^*(D\varphi) = \nabla\varphi$. The exponential map $\exp_\mu : T_\mu\mathcal{P} \longrightarrow \mathcal{P}_2(\mathbb{R}^n)$ is defined by $\exp_\mu(\Phi) := (\exp\Phi)_\sharp\mu$, where $(\exp\Phi)(x) := x + \Phi(x)$.

We say that a curve $(\mu_t)_{t\in I} \subset \mathcal{P}_2(\mathbb{R}^n)$ on an open interval $I \subset \mathbb{R}$ is (2-)*absolutely continuous* if there is some $h \in L^2(I)$ such that

$$W_2(\mu_t, \mu_\tau) \le \int_t^\tau h(r)\, dr$$

holds for all $t, \tau \in I$ with $t < \tau$. Note that an absolutely continuous curve is continuous. Given an absolutely continuous curve $(\mu_t)_{t\in I}$, the *forward absolute gradient*

$$|\dot{\mu}_t| := \lim_{\tau\to t} \frac{W_2(\mu_{\min\{\tau,t\}}, \mu_{\max\{\tau,t\}})}{|\tau - t|}$$

is well-defined for a.e. $t \in I$ ([AGS, Theorem 1.1.2], [OS, Lemma 7.1]). We can associate $(\mu_t)_{t\in I}$ with a Borel vector field Φ on $I \times \mathbb{R}^n$ (with $\Phi_t(x) := \Phi(t,x) \in T_x\mathbb{R}^n$) satisfying $\Phi_t \in T_{\mu_t}\mathcal{P}$ at a.e. $t \in I$, and solving the *continuity equation* $\partial_t\mu_t + \operatorname{div}(\Phi_t\mu_t) = 0$ in the weak sense that

$$\int_I \int_{\mathbb{R}^n} \{\partial_t\psi + D\psi(\Phi)\}\, d\mu_t dt = 0 \tag{5.1}$$

for all $\psi \in C_c^\infty(I \times \mathbb{R}^n)$ ([AGS, Theorem 8.3.1], [OS, Theorem 7.3]). Such a vector field is unique up to a difference on a null measure set with respect to $d\mu_t dt$ (under the condition $\Phi_t \in T_{\mu_t}\mathcal{P}$), and we have $F_{\mu_t}(\Phi_t) = |\dot{\mu}_t|$ for a.e. $t \in I$. We will call Φ the *tangent vector field* of the curve $(\mu_t)_{t\in I}$ and write $\dot{\mu}_t = \Phi_t$.

Now, consider a function $Q : \mathcal{P}_2(\mathbb{R}^n) \longrightarrow [-\infty, \infty]$. We say that Q is differentiable at $\mu \in \mathcal{P}_2(\mathbb{R}^n)$ if $-\infty < Q(\mu) < \infty$ and if there is $\alpha \in T_\mu^*\mathcal{P}$ such that

$$\int_{\mathbb{R}^n} \alpha(\Phi)\, d\mu = \lim_{t\downarrow 0} \frac{Q(\exp_\mu(t\Phi)) - Q(\mu)}{t} \tag{5.2}$$

holds for all $\Phi = \nabla\varphi \in \hat{T}\mathcal{P}$. Such a one-form α is unique in $T_\mu^*\mathcal{P}$ up to a difference on a μ-null measure set. Thus we write $DQ(\mu) = \alpha$ and define the *gradient vector* of Q at μ by $\nabla_W Q(\mu) := \mathcal{L}_\mu^*(DQ(\mu))$.

Definition 5.2 (Gradient curves in $(\mathcal{P}_2(\mathbb{R}^n), W_2)$) We say that an absolutely continuous curve $(\mu_t)_{t\in[0,T)} \subset \mathcal{P}_2(\mathbb{R}^n)$ with $T \in (0,\infty]$ is a *gradient curve* of Q if $\dot{\mu}_t = \nabla_W(-Q)(\mu_t)$ holds for a.e. $t \in (0,T)$.

We remark that the differentiability of $-Q$ (or, equivalently, Q) at a.e. $t \in (0,T)$ is included in the above definition.

5.2 Nonlinear heat equation and global solutions

We define the (distributional) *Finsler Laplacian* Δ acting on $u \in H^1_{\mathrm{loc}}(\mathbb{R}^n)$ by

$$\int_{\mathbb{R}^n} \psi\Delta u\, dx = -\int_{\mathbb{R}^n} D\psi(\nabla u)\, dx$$

for all $\psi \in C_c^\infty(\mathbb{R}^n)$. Note that Δ is a nonlinear operator since the Legendre transform is nonlinear. We consider the associated *heat equation* $\partial_t u = \Delta u$ also in the weak form. We saw in [OS, Theorem 3.4] that, given $u_0 \in H_0^1(\mathbb{R}^n)$ and $T > 0$, there exists a unique *global solution* $u \in L^2([0,T], H_0^1(\mathbb{R}^n)) \cap H^1([0,T], L^2(M))$ to $\partial_t u = \Delta u$ in the weak sense that

$$\int_{\mathbb{R}^n} \psi \partial_t u \, dx = -\int_{\mathbb{R}^n} D\psi(\nabla u) \, dx \tag{5.3}$$

holds for all $t \in [0,T]$ and $\psi \in H_0^1(\mathbb{R}^n)$. The distributional Laplacian Δu_t is absolutely continuous for each $t \in (0,T]$.

We can also regard $(u_t)_{t\in[0,T]}$ as a weak solution to the heat equation $\partial_t v = \Delta^{\nabla u} v$ associated with the linear second order differential operator

$$\Delta^{\nabla u} v := \operatorname{div}\left(\sum_{i,j=1}^n g^{ij}(\nabla u) \frac{\partial v}{\partial x^i} \frac{\partial}{\partial x^j} \right),$$

where (g^{ij}) is the inverse matrix of (g_{ij}) and $\nabla u(x)$ is replaced with some nonzero vector if $\nabla u(x) = 0$ (in a measurable way in x). By virtue of (2.2), $(g^{ij}(\nabla u))$ is globally uniformly elliptic with respect to the Euclidean inner product. Therefore the classical theory due to Nash [Na], Moser [Mo], Aronson [Ar] and others yields the parabolic Harnack inequality as well as the Gaussian estimates from both sides for fundamental solutions (see also [Sal] for the Riemannian case). Moreover, the continuous version of u is H^2_{loc} in x and $C^{1,\alpha}$ in both t and x ([OS, Theorems 4.6, 4.9]).

Remark 5.3 In the Riemannian case, a second integration by parts allows to rewrite (5.3) as an equation without spatial derivatives of u. In the Finsler case, however, this leads to $\int_{\mathbb{R}^n} \psi \partial_t u \, dx = \int_{\mathbb{R}^n} \Delta^{\nabla u} \psi \, u \, dx$. Thus we do need the (weak) differentiability of u to formulate the heat equation. This causes some difficulties when one intends to start from a purely metric approach of gradient flows.

The following lemma allows us to consider $(u_t \, dx)$ as a curve in $\mathcal{P}_2(\mathbb{R}^n)$.

Lemma 5.4 *Let $(u_t)_{t\ge 0} \subset H_0^1(\mathbb{R}^n)$ be a global solution to the heat equation. Then we have the following.*

- (i) (Mass preserving) *If $u_0 \, dx \in \mathcal{P}(\mathbb{R}^n)$, then $u_t \, dx \in \mathcal{P}(\mathbb{R}^n)$ for all $t > 0$.*
- (ii) (Weak/Narrow continuity) *Suppose $u_0 \, dx \in \mathcal{P}(\mathbb{R}^n)$. Then, for any bounded continuous function f, it holds that $\lim_{t\downarrow 0} \int_{\mathbb{R}^n} f u_t \, dx = \int_{\mathbb{R}^n} f u_0 \, dx$.*
- (iii) *If $u_0 \, dx \in \mathcal{P}_2(\mathbb{R}^n)$, then $u_t \, dx \in \mathcal{P}_2(\mathbb{R}^n)$ for all $t > 0$.*

Proof. (i) This easily follows from the existence of the fundamental solution q^u to the equation $\partial_t v = \Delta^{\nabla u} v$. Precisely, $u_t(x) = \int_{\mathbb{R}^n} q^u(t,x;0,y) u_0(y) \, dy$ and $\int_{\mathbb{R}^n} q^u(t,x;0,y) \, dx = 1$ imply $\int_{\mathbb{R}^n} u_t \, dx = 1$.

(ii) Thanks to (i) and [AGS, Remark 5.1.6], it is sufficient to show the convergence for $f \in C_c^\infty(\mathbb{R}^n)$. This immediately follows from (5.3), indeed, the Cauchy-Schwarz inequality yields

$$\left| \int_0^T \int_{\mathbb{R}^n} f \partial_t u \, dx dt \right| \le \left(T \int_{\mathbb{R}^n} \|Df\|_*^2 \, dx \right)^{1/2} \left(\int_0^T \int_{\mathbb{R}^n} \|\nabla u_t\|^2 \, dx dt \right)^{1/2} \to 0$$

as T tends to zero.

(iii) By virtue of the upper Gaussian bound (cf. [Sal, Corollary 6.2]), we have

$$q^u(t,x;0,y) \le C_1 t^{-n/2} \exp\biggl(-\frac{|x-y|^2}{C_2 t}\biggr),$$

where $|\cdot|$ stands for the Euclidean norm. Thus we obtain

$$\begin{aligned}
&\int_{\mathbb{R}^n} \|x\|^2 u_t(x)\,dx \le \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} 2(\|x-y\|^2+\|y\|^2) q^u(t,x;0,y) u_0(y)\,dydx \\
&\le 2C_1 t^{-n/2} \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} \|x-y\|^2 \exp\biggl(-\frac{|x-y|^2}{C_2 t}\biggr) u_0(y)\,dxdy + 2\int_{\mathbb{R}^n} \|y\|^2 u_0(y)\,dy \\
&\le C_3 t + 2\int_{\mathbb{R}^n} \|y\|^2 u_0(y)\,dy < \infty,
\end{aligned}$$

where C_3 depends only on $\|\cdot\|$. $\square$

5.3 Relative entropy and heat flow as its gradient flow

We define the *relative entropy* (with respect to the Lebesgue measure) by

$$\mathrm{Ent}(\mu) := \int_{\mathbb{R}^n} \rho \log \rho\,dx \in (-\infty,\infty]$$

for $\mu = \rho\,dx \in \mathcal{P}_2^{\mathrm{ac}}(\mathbb{R}^n)$, and $\mathrm{Ent}(\mu) := \infty$ otherwise. See [JKO, Section 4] (and Remark 5.1(a)) for the fact $\mathrm{Ent} > -\infty$ on $\mathcal{P}_2^{\mathrm{ac}}(\mathbb{R}^n)$. We know that Ent is convex along geodesics in $(\mathcal{P}_2(\mathbb{R}^n), W_2)$ ([Vi, page 908], [Oh3, Theorem 1.2]). There is a well established theory on gradient flows of such convex functionals, for which we refer to [AGS]. Here we explain that a global solution to the heat equation gives the gradient flow of the relative entropy, along the lines of [OS] and [AGS]. The next lemma corresponds to [OS, Proposition 7.7] (see also [AGS, Theorem 10.4.17]).

Lemma 5.5 *For $\mu = \rho\,dx \in \mathcal{P}_2^{\mathrm{ac}}(\mathbb{R}^n)$ with $\mathrm{Ent}(\rho\,dx) < \infty$ and $\rho \in H_0^1(\mathbb{R}^n) \cap L^\infty(\mathbb{R}^n)$, $-\mathrm{Ent}$ is differentiable at μ if and only if $\|\nabla(-\rho)\|/\rho \in L^2(\mathbb{R}^n,\mu)$, and then we have*

$$\nabla_W(-\mathrm{Ent})(\mu) = \frac{\nabla(-\rho)}{\rho} \in T_\mu\mathcal{P}.$$

Proof. By a similar calculation to [OS, Proposition 7.7], we obtain that (5.2) holds with $Q = -\mathrm{Ent}$ and $\alpha = D(-\rho)/\rho$. We first show the "if" part, for which it suffices to verify $D(-\rho)/\rho \in T_\mu^*\mathcal{P}$. As $\rho \in H_0^1(\mathbb{R}^n)$, ρ is approximated by some $\{f_k\}_{k\in\mathbb{N}} \subset C_c^\infty(\mathbb{R}^n)$ with respect to the Sobolev norm. For $k \in \mathbb{N}$, take $\varepsilon_k > 0$ such that

$$\int_{\mathbb{R}^n\setminus\Omega_k} \frac{\|D(-\rho)\|_*^2}{\rho}\,dx \le k^{-1}, \qquad \int_{\mathbb{R}^n\setminus\Omega_k} \|D\rho\|_*^2\,d\mu \le k^{-2},$$

where $\Omega_k := \{x \in \mathbb{R}^n \,|\, \|x\| < \varepsilon_k^{-1},\ \rho(x) > \varepsilon_k\}$ (note that $d\mu \le \|\rho\|_\infty\,dx$). Extracting a subsequence of $\{f_k\}$ if necessary, we can assume

$$\sup_{\Omega_k} |\rho - f_k| \le k^{-1},\ \int_{\mathbb{R}^n} \|D(f_k-\rho)\|_*^2\,d\mu \le k^{-2},\ \int_{\Omega_k\cap\{f_k<k^{-1/2}\}} \frac{\|D(-\rho)\|_*^2}{\rho}\,dx \le k^{-1}.$$

Put $h_k := -\log(\max\{f_k, k^{-1/2}\}) - (\log k)/2$ (≥ 0), and note that $Dh_k \in T^*_\mu \mathcal{P}$ and

$$\int_{\mathbb{R}^n \setminus (\Omega_k \cap \{f_k \geq k^{-1/2}\})} \|D(-h_k)\|_*^2 \, d\mu \leq \int_{\mathbb{R}^n \setminus \Omega_k} \left(\frac{\|Df_k\|_*}{\max\{f_k, k^{-1/2}\}} \right)^2 d\mu$$
$$\leq 2k \int_{\mathbb{R}^n \setminus \Omega_k} \{\|D(f_k - \rho)\|_*^2 + \|D\rho\|_*^2\} \, d\mu \leq 4k^{-1}.$$

Moreover, we have

$$\int_{\Omega_k \cap \{f_k \geq k^{-1/2}\}} \left\| \frac{D(-\rho)}{\rho} - Dh_k \right\|_*^2 d\mu$$
$$\leq 2 \int_{\Omega_k} \left(\frac{\|D(f_k - \rho)\|_*}{\max\{f_k, k^{-1/2}\}} \right)^2 d\mu + 2\mathcal{C} \int_{\Omega_k} \|D(-\rho)\|_*^2 \left(\frac{|\max\{f_k, k^{-1/2}\} - \rho|}{\rho \max\{f_k, k^{-1/2}\}} \right)^2 d\mu$$
$$\leq 2\frac{k^{-2}}{k^{-1}} + 2\mathcal{C} \left(\frac{k^{-1}}{k^{-1/2}} \right)^2 \int_{\Omega_k} \frac{\|D(-\rho)\|_*^2}{\rho} \, dx \quad \to 0 \quad \text{as } k \to \infty.$$

We used $\|-x\| \leq \mathcal{C}\sqrt{g_x(-x,-x)} = \mathcal{C}\|x\|$ in the first inequality (for x such that $\rho(x) > \max\{f_k(x), k^{-1/2}\}$). Hence $D(-\rho)/\rho \in T^*_\mu \mathcal{P}$ as required.

The "only if" part is proved similarly. Indeed, if $\|\nabla(-\rho)\|/\rho \notin L^2(\mathbb{R}^n, \mu)$, then we set $\Omega_k := \{x \in \mathbb{R}^n \mid \|x\| < k,\ \rho(x) > k^{-1}\}$ and choose $\{f_k\}_{k \in \mathbb{N}} \subset C_c^\infty(\mathbb{R}^n)$ such that

$$\sup_{\Omega_k} |\rho - f_k| \leq k^{-1} \left(\int_{\Omega_k} \frac{\|D(-\rho)\|_*^2}{\rho} \, dx \right)^{-1/2}, \qquad \int_{\mathbb{R}^n} \|D(f_k - \rho)\|_*^2 \, d\mu \leq k^{-2}.$$

Note that $\{f_k > k^{-1/2}\} \subset \Omega_k$ for large k. Define h_k in the same way as the "if" part, and observe $\lim_{k\to\infty} \int_{\{f_k > k^{-1/2}\}} \|D(-\rho)/\rho - Dh_k\|_*^2 \, d\mu = 0$. Taking $\varphi_k \in C_c^\infty(\{f_k > k^{-1/2}\})$ such that $\nabla\varphi_k$ approximates $F_\mu(\nabla h_k)^{-1} \nabla h_k$ with respect to F_μ, we obtain

$$\lim_{k\to\infty} \lim_{t \downarrow 0} \frac{\mathrm{Ent}(\mu) - \mathrm{Ent}(\exp_\mu(t\nabla\varphi_k))}{t} = \lim_{k\to\infty} \int_{\{f_k > k^{-1/2}\}} \frac{D(-\rho)}{\rho} \left(\frac{\nabla h_k}{F_\mu(\nabla h_k)} \right) d\mu = \infty.$$

Therefore $-\mathrm{Ent}$ is not differentiable at μ. $\square$

The following theorem is a slight modification of [OS, Theorem 7.8] adapted to noncompact spaces. For the sake of simplicity, we are concerned with the reverse heat equation, that is the heat equation with respect to the reverse norm $\|x\|_{\leftarrow} := \|-x\|$. Since $\overleftarrow{\nabla} u = -\nabla(-u)$, we can write it as

$$\int_{\mathbb{R}^n} \psi \partial_t u \, dx = -\int_{\mathbb{R}^n} D\psi(\overleftarrow{\nabla} u) \, dx = \int_{\mathbb{R}^n} D\psi\big(\nabla(-u)\big) \, dx. \tag{5.4}$$

We remark that, due to the Harnack inequality, any nonnegative global L^2-solution u_t is bounded on $\mathbb{R}^n$ for each $t > 0$ (cf. [Sal, Theorem 5.1]).

Theorem 5.6 (Heat flow as gradient flow) (i) *Let $(\rho_t)_{t \geq 0} \subset H_0^1(\mathbb{R}^n)$ be a global solution to the reverse heat equation with $\rho_0 \, dx \in \mathcal{P}_2^{\mathrm{ac}}(\mathbb{R}^n)$ and* $\mathrm{Ent}(\rho_0 \, dx) < \infty$. *Then $\mu_t := \rho_t \, dx$ is a gradient curve of the relative entropy* (*in the sense of Definition* 5.2).

(ii) *Conversely, let* $(\mu_t)_{t\ge 0}\subset \mathcal{P}_2^{\rm ac}(\mathbb{R}^n)$ *be a gradient curve of the relative entropy, put* $\mu_t=\rho_t\,dx$ *and assume that* $\rho_t\in H_0^1(\mathbb{R}^n)$ *for a.e.* t*. Then* ρ_t *is a global solution to the reverse heat equation.*

Proof. (i) It follows from the reverse heat equation (5.4) that $\nabla(-\rho)/\rho$ satisfies the continuity equation (5.1) along the curve $(\mu_t)_{t\ge 0}$, namely

$$\int_0^\infty \int_{\mathbb{R}^n} \{\partial_t\psi\cdot\rho_t + D\psi\big(\nabla(-\rho_t)\big)\}\,dxdt = 0 \tag{5.5}$$

for all $\psi\in C_c^\infty([0,\infty)\times\mathbb{R}^n)$. Then we have, by choosing a test function ψ approximating $\log(\max\{\rho_t,\varepsilon\})-\log\varepsilon$ (as in Lemma 5.5) and then letting $\varepsilon\downarrow 0$,

$$\int_0^T\int_{\mathbb{R}^n}\frac{\|\nabla(-\rho_t)\|^2}{\rho_t}\,dxdt = \mathrm{Ent}(\mu_0)-\mathrm{Ent}(\mu_T)<\infty.$$

Hence $\mathrm{Ent}(\mu_t)<\infty$ and $\|\nabla(-\rho_t)\|/\rho_t\in L^2(\mathbb{R}^n,\mu_t)$ at a.e. t, which implies that $-\mathrm{Ent}$ is differentiable at μ_t and $\nabla_W(-\mathrm{Ent})(\mu_t)=\nabla(-\rho_t)/\rho_t\in T_{\mu_t}\mathcal{P}$ for a.e. t (Lemma 5.5). Combining this with (5.5), we conclude $\dot\mu_t=\nabla(-\rho_t)/\rho_t$ and that (μ_t) is a gradient curve of Ent in the sense of Definition 5.2.

(ii) Note that the "only if" part of Lemma 5.5 ensures $\dot\mu_t=\nabla_W(-\mathrm{Ent})(\mu_t)=\nabla(-\rho_t)/\rho_t$ for a.e. t. This implies the continuity equation (5.5), from which the reverse heat equation is straightforward. $\square$

Remark 5.7 The formula $\nabla_W(-\mathrm{Ent})(\mu)=\nabla(-\rho)/\rho$ in Lemma 5.5 has an extra importance in the Finsler/Minkowski setting. The reverse heat equation (5.4) is rewritten via Remark 5.3 as $\int_{\mathbb{R}^n}\psi\partial_t\rho\,dx=\int_{\mathbb{R}^n}\Delta^{\nabla(-\rho)}\psi\,\rho\,dx$, and the fact $g_{\nabla(-\rho)}=g_{\nabla(-\rho)/\rho}$ guarantees that a formal calculation with respect to the time-dependent Riemannian structure $g_{\nabla(-\rho)}$ verifies Theorem 5.6 (cf. [Oh1, Theorem 6.6]).

5.4 Skew convexity and Wasserstein contraction

To show an analogous result to Theorem 3.2 for the relative entropy, we prove the first variation formula for the Wasserstein distance along heat flow (along the line of [AGS, Section 10.2]).

Proposition 5.8 (First variation formula along heat flow) *Given any global solutions* $(\rho_t)_{t\ge0},(\sigma_t)_{t\ge0}\subset H_0^1(\mathbb{R}^n)$ *to the reverse heat equation* (5.4) *such that* $\mu_t=\rho_t\,dx\in\mathcal{P}_2^{\rm ac}(\mathbb{R}^n)$ *and* $\nu_t=\sigma_t\,dx\in\mathcal{P}_2^{\rm ac}(\mathbb{R}^n)$*, we have*

$$\lim_{\tau\downarrow t}\frac{W_2(\mu_\tau,\nu_\tau)^2-W_2(\mu_t,\nu_t)^2}{2(\tau-t)}=\int_{\mathbb{R}^n}g_{\dot\omega_1}(\dot\omega_1,\dot\nu_t)\,d\nu_t-\int_{\mathbb{R}^n}g_{\dot\omega_0}(\dot\omega_0,\dot\mu_t)\,d\mu_t \tag{5.6}$$

for all $t>0$*, where* $\omega:[0,1]\longrightarrow\mathcal{P}_2(\mathbb{R}^n)$ *is the minimal geodesic from* μ_t *to* ν_t*.*

Proof. Set $l(t):=W_2(\mu_t,\nu_t)$, fix $\delta>0$ and define $\widetilde\Gamma$ as the set of continuous curves $\xi:[t,t+\delta]\longrightarrow\mathbb{R}^n$ endowed with the uniform topology. For $\tau\in[t,t+\delta]$, we define the evaluation map $e_\tau:\widetilde\Gamma\longrightarrow\mathbb{R}^n$ by $e_\tau(\xi):=\xi(\tau)$. Then there exist probability measures

$\Pi, \Xi \in \mathcal{P}(\widetilde{\Gamma})$ such that $(e_\tau)_\sharp \Pi = \mu_\tau$, $(e_\tau)_\sharp \Xi = \nu_\tau$ for all $\tau \in [t, t+\delta]$ and that Π, Ξ are concentrated on the set of absolutely continuous curves ξ, ζ solving

$$\dot{\xi}(\tau) = [\nabla(-\rho_\tau)/\rho_\tau]\big(\xi(\tau)\big), \qquad \dot{\zeta}(\tau) = [\nabla(-\sigma_\tau)/\sigma_\tau]\big(\zeta(\tau)\big),$$

respectively (cf. [AGS, Theorem 8.2.1]).

To see "$\le$" of (5.6), we disintegrate Π and Ξ by using μ_t and ν_t as $d\Pi = d\Pi^t_x d\mu_t(x)$ and $d\Xi = d\Xi^t_y d\nu_t(y)$. Take the unique minimal geodesic $\omega : [0,1] \longrightarrow \mathcal{P}_2(\mathbb{R}^n)$ from μ_t to ν_t, and let π_t be the unique optimal coupling of μ_t and ν_t. Then we find, for each $\tau \in [t, t+\delta]$,

$$l(\tau)^2 \le \int_{\mathbb{R}^n \times \mathbb{R}^n} \int_{\widetilde{\Gamma} \times \widetilde{\Gamma}} \|\zeta(\tau) - \xi(\tau)\|^2 \, d\Pi^t_x(\xi) d\Xi^t_y(\zeta) d\pi_t(x,y).$$

We deduce from the first variation formula (2.6) on the underlying space $(\mathbb{R}^n, \|\cdot\|)$ that

$$\begin{aligned}
&\lim_{\tau \downarrow t} \frac{\|\zeta(\tau) - \xi(\tau)\|^2 - \|\zeta(t) - \xi(t)\|^2}{\tau - t} \\
&= 2\big\{ g_{\zeta(t)-\xi(t)}\big(\zeta(t) - \xi(t), \dot{\zeta}(t)\big) - g_{\zeta(t)-\xi(t)}\big(\zeta(t) - \xi(t), \dot{\xi}(t)\big) \big\} \\
&= 2 g_{\zeta(t)-\xi(t)}\bigg(\zeta(t) - \xi(t), \frac{\nabla(-\sigma_t)}{\sigma_t}\big(\zeta(t)\big) - \frac{\nabla(-\rho_t)}{\rho_t}\big(\xi(t)\big)\bigg)
\end{aligned}$$

for Π-a.e. ξ and Ξ-a.e. ζ. Since ρ and σ are $C^{1,\alpha}$, this convergence is uniform on

$$\Omega_\varepsilon := \{x \in \mathbb{R}^n \mid \|x\| < \varepsilon^{-1}, \ \rho_t(x) > \varepsilon, \ \sigma_t(x) > \varepsilon\}$$

for each $\varepsilon > 0$. In order to see that the effect of $\mathbb{R}^n \setminus \Omega_\varepsilon$ is negligible as ε tends to zero, we observe from

$$\begin{aligned}
&\|\zeta(\tau) - \xi(\tau)\|^2 - \|\zeta(t) - \xi(t)\|^2 \\
&\le (\|\zeta(\tau) - \xi(\tau)\| + \|\zeta(t) - \xi(t)\|)(\|\zeta(\tau) - \zeta(t)\| + \|\xi(t) - \xi(\tau)\|)
\end{aligned}$$

that

$$\begin{aligned}
&\int_{\mathbb{R}^n \times \mathbb{R}^n} \int_{\widetilde{\Gamma} \times \widetilde{\Gamma}} \frac{\|\zeta(\tau) - \xi(\tau)\|^2 - \|\zeta(t) - \xi(t)\|^2}{\tau - t} \, d\Pi^t_x(\xi) d\Xi^t_y(\zeta) d\pi_t(x,y) \\
&\le \bigg(2 \int_{\mathbb{R}^n \times \mathbb{R}^n} \int_{\widetilde{\Gamma} \times \widetilde{\Gamma}} \{\|\zeta(\tau) - \xi(\tau)\|^2 + \|\zeta(t) - \xi(t)\|^2\} \, d\Pi^t_x(\xi) d\Xi^t_y(\zeta) d\pi_t(x,y) \bigg)^{1/2} \\
&\quad \times \bigg(\frac{2}{(\tau - t)^2} \int_{\widetilde{\Gamma}} \|\zeta(\tau) - \zeta(t)\|^2 \, d\Xi(\zeta) + \frac{2}{(\tau - t)^2} \int_{\widetilde{\Gamma}} \|\xi(t) - \xi(\tau)\|^2 \, d\Pi(\xi) \bigg)^{1/2}.
\end{aligned}$$

This is finite uniformly in $\tau \in (t, t+\delta]$, because

$$\begin{aligned}
&\frac{1}{(\tau - t)^2} \int_{\widetilde{\Gamma}} \|\zeta(\tau) - \zeta(t)\|^2 \, d\Xi(\zeta) \le \frac{1}{(\tau - t)^2} \int_{\widetilde{\Gamma}} \bigg(\int_t^\tau \|\dot{\zeta}(s)\| \, ds \bigg)^2 d\Xi(\zeta) \\
&\le \frac{1}{\tau - t} \int_{\widetilde{\Gamma}} \int_t^\tau \|\dot{\zeta}(s)\|^2 \, ds d\Xi(\zeta) = \frac{1}{\tau - t} \int_t^\tau |\dot{\nu}_s|^2 \, ds.
\end{aligned}$$

Therefore we obtain

$$\limsup_{\tau\downarrow t} \frac{l(\tau)^2 - l(t)^2}{2(\tau - t)} \le \int_{\mathbb{R}^n\times\mathbb{R}^n} \int_{\widetilde{\Gamma}\times\widetilde{\Gamma}} g_{y-x}\big(y-x, \dot{\zeta}(t) - \dot{\xi}(t)\big)\, d\Pi_x^t(\xi) d\Xi_y^t(\zeta) d\pi_t(x,y).$$

Note that

$$\begin{aligned}\int_{\mathbb{R}^n\times\mathbb{R}^n} \int_{\widetilde{\Gamma}} g_{y-x}\big(y-x, \dot{\xi}(t)\big)\, d\Pi_x^t(\xi) d\pi_t(x,y) &= \int_{\mathbb{R}^n} \int_{\widetilde{\Gamma}} g_{\dot{\omega}_0(x)}\big(\dot{\omega}_0(x), \dot{\xi}(t)\big)\, d\Pi_x^t(\xi) d\mu_t(x) \\ &= \int_{\mathbb{R}^n} g_{\dot{\omega}_0}(\dot{\omega}_0, \dot{\mu}_t)\, d\mu_t.\end{aligned}$$

Hence we have

$$\limsup_{\tau\downarrow t} \frac{l(\tau)^2 - l(t)^2}{2(\tau - t)} \le \int_{\mathbb{R}^n} g_{\dot{\omega}_1}(\dot{\omega}_1, \dot{\nu}_t)\, d\nu_t - \int_{\mathbb{R}^n} g_{\dot{\omega}_0}(\dot{\omega}_0, \dot{\mu}_t)\, d\mu_t.$$

To see the reverse inequality, we fix $\tau \in (t, t+\delta)$, take the optimal coupling π_τ of μ_τ and ν_τ, and disintegrate Π and Ξ as $d\Pi = d\Pi_x^\tau d\mu_\tau(x)$ and $d\Xi = d\Xi_y^\tau d\nu_\tau(y)$. Observe that

$$l(\tau)^2 - l(t)^2 \ge \int_{\mathbb{R}^n\times\mathbb{R}^n} \int_{\widetilde{\Gamma}\times\widetilde{\Gamma}} \{\|\zeta(\tau) - \xi(\tau)\|^2 - \|\zeta(t) - \xi(t)\|^2\}\, d\Pi_x^\tau(\xi) d\Xi_y^\tau(\zeta) d\pi_\tau(x,y).$$

Since the function

$$[0,1] \ni s \longmapsto \|\{(1-s)\zeta(t) + s\zeta(\tau)\} - \{(1-s)\xi(t) + s\xi(\tau)\}\|^2$$

is convex, the first variation formula (2.6) (at $s=0$) yields that

$$\|\zeta(\tau) - \xi(\tau)\|^2 - \|\zeta(t) - \xi(t)\|^2 \ge 2g_{\zeta(t)-\xi(t)}\big(\zeta(t) - \xi(t), \{\zeta(\tau) - \zeta(t)\} - \{\xi(\tau) - \xi(t)\}\big).$$

Thus we find

$$\begin{aligned}&\frac{l(\tau)^2 - l(t)^2}{2(\tau - t)} \\ &\ge \int_{\mathbb{R}^n\times\mathbb{R}^n} \int_{\widetilde{\Gamma}\times\widetilde{\Gamma}} g_{\zeta(t)-\xi(t)}\bigg(\zeta(t) - \xi(t), \frac{\zeta(\tau) - \zeta(t)}{\tau - t} - \frac{\xi(\tau) - \xi(t)}{\tau - t}\bigg)\, d\Pi_x^\tau(\xi) d\Xi_y^\tau(\zeta) d\pi_\tau(x,y).\end{aligned}$$

Recall that $(\xi(\tau) - \xi(t))/(\tau - t)$ converges to $\dot{\xi}(t) = [\nabla(-\rho_t)/\rho_t](\xi(t))$ uniformly on Ω_ε. Moreover,

$$d\tilde{\pi}_t^\tau := (e_t \times e_t)_\sharp \bigg[\int_{\mathbb{R}^n\times\mathbb{R}^n} d\Pi_x^\tau d\Xi_y^\tau d\pi_\tau(x,y)\bigg] \in \mathcal{P}_2(\mathbb{R}^n \times \mathbb{R}^n)$$

weakly converges to π_t as $\tau \downarrow t$ due to [AGS, Lemma 10.2.8]. Precisely, as $\tilde{\pi}_t^\tau$ is a coupling of μ_t and ν_t, the family $\{\tilde{\pi}_t^\tau\}_{\tau\in(t,t+\delta)}$ is relatively compact (cf. [AGS, Remark 5.2.3]). Combining this with the simple estimate

$$\begin{aligned}&\bigg(\int_{\mathbb{R}^n\times\mathbb{R}^n} \|y - x\|^2\, d\tilde{\pi}_t^\tau(x,y)\bigg)^{1/2} \\ &\le \bigg(\int_{\widetilde{\Gamma}} \|\zeta(t) - \zeta(\tau)\|^2\, d\Xi(\zeta)\bigg)^{1/2} + W_2(\mu_\tau, \nu_\tau) + \bigg(\int_{\widetilde{\Gamma}} \|\xi(\tau) - \xi(t)\|^2\, d\Pi(\xi)\bigg)^{1/2} \\ &\to W_2(\mu_t, \nu_t) \qquad (\tau \downarrow t)\end{aligned}$$

and the uniqueness of the optimal coupling π_t, we see that the limit of $\tilde{\pi}_t^\tau$ must be π_t. Therefore we obtain

$$\liminf_{\tau\downarrow t} \frac{l(\tau)^2 - l(t)^2}{2(\tau - t)} \geq \int_{\mathbb{R}^n\times\mathbb{R}^n} g_{y-x}\left(y - x, \frac{\nabla(-\sigma_t)}{\sigma_t}(y) - \frac{\nabla(-\rho_t)}{\rho_t}(x)\right) d\pi_t(x, y)$$

and complete the proof. □

Now, the following is shown in a similar way to Theorem 3.2.

Proposition 5.9 (Skew convexity versus Wasserstein contraction) *The following are equivalent for each $K \in \mathbb{R}$.*

(I) *The relative entropy is K-skew convex in the sense that, for any $\mu, \nu \in \mathcal{P}_2^{\rm ac}(\mathbb{R}^n)$ at where* $-\,$Ent *is differentiable and for the minimal geodesic $\omega : [0,1] \longrightarrow \mathcal{P}_2(\mathbb{R}^n)$ from μ to ν, it holds*

$$\int_{\mathbb{R}^n} g_{\dot\omega_1}\big(\dot\omega_1, \nabla_W(-\operatorname{Ent})(\nu)\big)\, d\nu - \int_{\mathbb{R}^n} g_{\dot\omega_0}\big(\dot\omega_0, \nabla_W(-\operatorname{Ent})(\mu)\big)\, d\mu \leq -K W_2(\mu,\nu)^2. \quad (5.7)$$

(II) *The reverse heat flow is K-contractive in the sense that, for any global solutions $(\rho_t)_{t\geq 0}, (\sigma_t)_{t\geq 0} \subset H_0^1(\mathbb{R}^n)$ to the reverse heat equation* (5.4) *such that $\mu_t = \rho_t\, dx \in \mathcal{P}_2^{\rm ac}(\mathbb{R}^n)$, $\nu_t = \sigma_t\, dx \in \mathcal{P}_2^{\rm ac}(\mathbb{R}^n)$,* $\operatorname{Ent}(\mu_0) < \infty$ *and that* $\operatorname{Ent}(\nu_0) < \infty$, *we have*

$$W_2(\mu_t, \nu_t) \leq e^{-Kt} W_2(\mu_0, \nu_0) \qquad \textit{for all } t \in [0,\infty). \quad (5.8)$$

As a corollary to Proposition 5.9, we obtain the 0-contraction of gradient curves in a special class of measures (compare this with Step 1 in the next section).

Corollary 5.10 (Non-expansion among Gaussian measures) *Let $(\mathbb{R}^n, \|\cdot\|)$ be a (symmetric) normed space, i.e., $\|-x\| = \|x\|$. Take two measures of Gaussian form*

$$d\mu_0(x) = Ca^{-n/2} \exp\left(-\frac{\|x-y\|^2}{4a}\right) dx, \quad d\nu_0(x) = Cb^{-n/2} \exp\left(-\frac{\|x-z\|^2}{4b}\right) dx$$

for some $a, b > 0$, $y, z \in \mathbb{R}^n$ and the fixed normalizing constant $C > 0$. Then the gradient curves $(\mu_t)_{t\geq 0}$, $(\nu_t)_{t\geq 0}$ of Ent *starting from them satisfies $W_2(\mu_t, \nu_t) \leq W_2(\mu_0, \nu_0)$ for all $t \geq 0$.*

Proof. For brevity, we assume $y = 0$ and $a \geq b$. Solving the heat equation, we observe

$$d\mu_t(x) = C(t+a)^{-n/2} \exp\left(-\frac{\|x\|^2}{4(t+a)}\right) dx,\ d\nu_t(x) = C(t+b)^{-n/2} \exp\left(-\frac{\|x-z\|^2}{4(t+b)}\right) dx.$$

Note also that the unique minimal geodesic $(\omega_s)_{s\in[0,1]}$ from μ_t to ν_t is given by $(T_s)_\sharp \mu_t$, where

$$T_s(x) := (1-s)x + s\left(z + \sqrt{\frac{t+b}{t+a}}x\right).$$

We can explicitly write as

$$d\omega_s(x) = C\big(t + (1-s)a + sb\big)^{-n/2} \exp\left(-\frac{\|x - sz\|^2}{4(t + (1-s)a + sb)}\right) dx.$$

It follows from Lemma 5.5 that

$$\begin{aligned}
&\int_{\mathbb{R}^n} g_{\dot\omega_s}\big(\dot\omega_s, \nabla_W(-\mathrm{Ent})(\omega_s)\big)\, d\omega_s \\
&= \int_{\mathbb{R}^n} g_{T_1(x)-x}\big(T_1(x)-x, \big[\nabla_W(-\mathrm{Ent})(\omega_s)\big](T_s(x))\big)\, d\mu_t(x) \\
&= \frac{1}{2(t+(1-s)a+sb)} \int_{\mathbb{R}^n} g_{T_1(x)-x}\big(T_1(x)-x, T_s(x)-sz\big)\, d\mu_t(x) \\
&= -\frac{(1-s)+s\sqrt{(t+b)/(t+a)}}{2(t+(1-s)a+sb)} \int_{\mathbb{R}^n} g_{T_1(x)-x}\big(x-T_1(x), x\big)\, d\mu_t(x).
\end{aligned}$$

Observe that the coefficient of the last line is non-increasing in s. Hence it suffices to show that $\Theta := \int_{\mathbb{R}^n} g_{T_1(x)-x}(x-T_1(x),x)\, d\mu_t(x)$ (which is independent of s) is nonnegative. If $a=b$, then we find $T_1(x)-x \equiv z$ and $\Theta = 0$ by the symmetry of μ_t. If $a>b$, then we put $z' = s'z = \{1-\sqrt{(t+b)/(t+a)}\}^{-1} z$ and deduce $T_{s'}(x) \equiv z'$. Thus we have

$$g_{T_1(x)-x}\big(x-T_1(x),x\big) = \frac{1}{s'} g_{z'-x}(x-z',x) = \frac{1}{2s'}[D(\|z'-\cdot\|^2)(x)](x),$$

and $[D(\|z'-\cdot\|^2)(x)](x) + [D(\|z'-\cdot\|^2)(-x)](-x) \ge 0$ by the convexity of $\|z'-\cdot\|^2$ (along with the symmetry of $\|\cdot\|$). Therefore we obtain $\Theta \ge 0$, and Proposition 5.9 completes the proof. □

6 Non-contraction of heat flow

This last section is devoted to a proof of Theorem 1.1. For notational simplicity, we prove this for the reverse norm, i.e., global solutions to the reverse heat equation (5.3) (in other words, gradient curves of Ent in $(\mathcal{P}_2(\mathbb{R}^n), W_2)$, recall Theorem 5.6) are not K-contractive with respect to W_2.

Fix $\mu = \rho\, dx \in \mathcal{P}_2^{\mathrm{ac}}(\mathbb{R}^n)$ with $\rho \in H_0^1(\mathbb{R}^n)$, $\mathrm{Ent}(\mu) < \infty$ and $\|\nabla(-\rho)\|/\rho \in L^2(\mathbb{R}^n, \mu)$. For $T > 1$, we set

$$\omega_s = \rho^s dx := (\mathcal{F}_{s/T})_\sharp \mu, \quad \mathcal{F}_{s/T}(x) := \bigg(1 - \frac{s}{T}\bigg) x \quad \text{for } s \in [0,T].$$

Then $(\omega_s)_{s\in[0,T]}$ is the unique minimal geodesic from μ to the Dirac measure δ_O at the origin O, and its tangent vector field $\dot\omega_s$ is simply given by $\dot\omega_s(x) = -x/(T-s)$. Put $\nu = \omega_1$. We will show that (5.7) is false for any $K \in \mathbb{R}$ (i.e., Ent is not K-skew convex) by choosing suitable ρ.

We deduce from $\nabla_W(-\mathrm{Ent})(\omega_s) = \nabla(-\rho^s)/\rho^s$ (Lemma 5.5) that

$$\int_{\mathbb{R}^n} g_{\dot\omega_s}\big(\dot\omega_s, \nabla_W(-\mathrm{Ent})(\omega_s)\big)\, d\omega_s = -\frac{1}{T-s} \int_{\mathrm{supp}\,\rho^s} g_{-x}\big(x, \nabla(-\rho^s)(x)\big)\, dx.$$

It follows from $\rho^s(x) = (T/(T-s))^n \rho(Tx/(T-s))$ and the change of variables formula that

$$\begin{aligned}
\int_{\mathrm{supp}\,\rho^s} g_{-x}\big(x, \nabla(-\rho^s)(x)\big)\, dx &= \int_{\mathrm{supp}\,\rho^s} g_{-x}\bigg(x, \Big(\frac{T}{T-s}\Big)^{n+1} \nabla(-\rho)\Big(\frac{Tx}{T-s}\Big)\bigg)\, dx \\
&= \int_{\mathrm{supp}\,\rho} g_{-x}\big(x, \nabla(-\rho)(x)\big)\, dx.
\end{aligned}$$

Thus we have

$$\int_{\mathbb{R}^n} g_{\dot{\omega}_s}\big(\dot{\omega}_s, \nabla_W(-\operatorname{Ent})(\omega_s)\big)\, d\omega_s = -\frac{1}{T-s}\int_{\operatorname{supp}\rho} g_{-x}\big(x, \nabla(-\rho)(x)\big)\, dx,$$

and hence

$$\frac{d}{ds}\bigg[\int_{\mathbb{R}^n} g_{\dot{\omega}_s}\big(\dot{\omega}_s, \nabla_W(-\operatorname{Ent})(\omega_s)\big)\, d\omega_s\bigg] = -\frac{1}{(T-s)^2}\int_{\operatorname{supp}\rho} g_{-x}(x, \nabla(-\rho)(x))\, dx.$$

Note that $W_2(\omega_0, \omega_1)^2 = T^{-2}\int_{\mathbb{R}^n} \|-x\|^2 \rho(x)\, dx$ and

$$\begin{aligned} g_{-x}\big(x, \nabla(-\rho)(x)\big) &= -[D(-\|-\cdot\|^2/2)(x)]\big(\nabla(-\rho)(x)\big) \\ &= \|-x\| \cdot [D(\|-\cdot\|)(x)]\big(\nabla(-\rho)(x)\big). \end{aligned}$$

We set

$$\Theta(\rho) := \frac{\int_{\operatorname{supp}\rho} \|-x\| \cdot [D(\|-\cdot\|)(x)](\nabla(-\rho)(x))\, dx}{\int_{\mathbb{R}^n} \|-x\|^2 \rho(x)\, dx} \tag{6.1}$$

and shall demonstrate that $\Theta(\rho)$ can be negative (Steps 1–3 below) and even arbitrarily small (Step 4) by choosing suitable ρ, unless $\|\cdot\|$ is an inner product. This means that (5.7) is false for any $K \in \mathbb{R}$, and completes the proof of Theorem 1.1.

Step 1 (The model case of ℓ_p^2 with $2 < p < \infty$) Let $\|\cdot\|$ be the ℓ_p-norm of $\mathbb{R}^2$ such that $2 < p < \infty$. Take the unit vectors $a = (-1, 0)$, $b = (2^{-1/p}, 2^{-1/p})$, $c = (2^{-1/p}, -2^{-1/p})$ and let $\triangle ABC$ be the triangle tangent to the unit sphere of $\|\cdot\|$ at a, b, c. Precisely, $A = (2^{1-1/p}, 0)$, $B = (-1, -1-2^{1-1/p})$ and $C = (-1, 1+2^{1-1/p})$ (Figure 3).

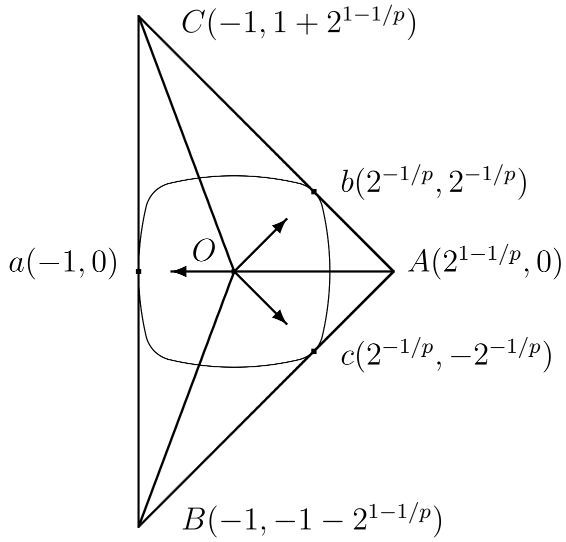

Figure 3

Define the nonnegative function $\hat{\rho} : \mathbb{R}^2 \longrightarrow [0, \infty)$ by $\hat{\rho} := 0$ outside $\triangle ABC$ and by $\hat{\rho}(tx) := (1-t)\sigma$ for a point x on the edges of $\triangle ABC$ and for $t \in [0, 1]$, where the constant $\sigma > 0$ is chosen so that $\int_{\mathbb{R}^2} \hat{\rho}\, dx = 1$. Note that the gradient vector $\nabla(-\hat{\rho})$ is $\sigma \cdot a = (-\sigma, 0)$

inside $\triangle OBC$, $\sigma \cdot b = (2^{-1/p}\sigma, 2^{-1/p}\sigma)$ inside $\triangle OAC$, and $\sigma \cdot c = (2^{-1/p}\sigma, -2^{-1/p}\sigma)$ inside $\triangle OAB$. Hence we have

$$\begin{aligned}\int_{\triangle ABC} \nabla(-\hat{\rho})\, dx &= (1 + 2^{1-1/p}) \cdot (-\sigma, 0) + 2^{1-1/p}(1 + 2^{1-1/p}) \cdot (2^{-1/p}\sigma, 0) \\ &= (1 + 2^{1-1/p})\sigma \cdot (2^{1-2/p} - 1, 0). \end{aligned} \tag{6.2}$$

Note that $2^{1-2/p} - 1 > 0$ since $p > 2$.

Now, for large $T > 1$, we consider the function $\hat{\rho}_T(x) := \hat{\rho}(x + (T, 0))$. Then it follows from (6.2) that

$$\lim_{T\to\infty} \int_{\operatorname{supp}\hat{\rho}_T} \frac{\|-x\|}{T} \cdot [D(\|-\cdot\|)(x)]\big(\nabla(-\hat{\rho}_T)(x)\big)\, dx = -(1 + 2^{1-1/p})(2^{1-2/p} - 1)\sigma < 0.$$

Therefore, by taking a smooth approximation of $\hat{\rho}_T$ for sufficiently large T, we find ρ satisfying $\Theta(\rho) < 0$.

Step 2 (General two-dimensional case) The argument in Step 1 yields the following claim. We will denote by $\mathbf{S}(1)$ the unit sphere of the norm $\|\cdot\|$.

Claim 6.1 *Suppose that a Minkowski space* $(\mathbb{R}^2, \|\cdot\|)$ *admits a triangle* $\triangle ABC$ *such that edges* AB, BC, CA *are tangent to* $\mathbf{S}(1)$ *at points* c, a, b*, respectively, and that the vector*

$$|\triangle OAB| \cdot c + |\triangle OBC| \cdot a + |\triangle OCA| \cdot b \tag{6.3}$$

is nonzero, where $|\triangle OAB|$ *denotes the area of* $\triangle OAB$ *with respect to the Lebesgue measure of* $\mathbb{R}^2$*. Then there exists a function* ρ *for which* $\Theta(\rho)$ *as in* (6.1) *is negative.*

Note that (6.3) is always zero in inner product spaces. (Indeed, for the standard inner product, it holds that $\langle |\triangle OAB| \cdot c + |\triangle OBC| \cdot a + |\triangle OCA| \cdot b, e_i \rangle = 0$ for $e_1 = (1, 0)$, $e_2 = (0, 1)$ by the fundamental theorem of calculus applied to the function ρ defined as in Step 1.) Claim 6.1 is sharp enough for our purpose, as we certainly verify the following.

Claim 6.2 *We can find a triangle* $\triangle ABC$ *satisfying the condition in Claim* 6.1 *unless* $\|\cdot\|$ *is an inner product.*

Although this claim should be a known fact (and there would be a simpler proof), we give a proof for completeness. We first treat the easier case of nonsymmetric norms. Choose a pair $a, b \in \mathbf{S}(1)$ such that $b = -\lambda a$ with $\lambda \neq 1$. If the tangent lines of $\mathbf{S}(1)$ at a and b are not parallel in $\mathbb{R}^2$, then we draw a triangle $\triangle ABC$ in such a way that the edge AB is parallel to ba. As the vectors a and c are linearly independent, (6.3) is not zero. In the other case where the tangent lines of $\mathbf{S}(1)$ at a and b are parallel, we take C' such that OC' is parallel to these tangent lines, and draw the triangle $\triangle A'B'C'$ such that $A'B'$ is parallel to $b'a'$. By letting C' go to infinity, a' and b' can be arbitrary close to a and b, respectively. Then we observe that $|\triangle OA'B'|$ is much smaller than $|\triangle OB'C'|$ and $|\triangle OC'A'|$, and that the ratio $|\triangle OC'A'|/|\triangle OB'C'|$ is close to λ. Thus we have

$$|\triangle OA'B'| \cdot c' + |\triangle OB'C'| \cdot a' + |\triangle OC'A'| \cdot b' \approx |\triangle OB'C'|(1 - \lambda^2) \cdot a \neq 0.$$

Next we consider symmetric norms. We suppose that the sum (6.3) is always zero, and shall see that $\|\cdot\|$ must be an inner product. Take $a, b \in \mathbf{S}(1)$ with $b = -a$ such that $|a| = \sup_{x \in \mathbf{S}(1)} |x|$, where $|\cdot|$ is the Euclidean norm. Then the tangent lines of $\mathbf{S}(1)$ at a and b are perpendicular to ab with respect to the Eulidean inner product. As in the nonsymmentric case, we take C so that OC is parallel to these tangent lines, and consider the triangle $\triangle A'B'C$ for some fixed c. Let C diverge to infinity and denote the limits of A', B' by A, B. Then our hypothesis yields that the vector

$$|\triangle OaB| \cdot a + |\triangle OAb| \cdot b + |\triangle OAB| \cdot c \tag{6.4}$$

is independent of the choice of c on the arc between a and b opposite to C (since $\triangle a'b'C$ was independent of c). We will see that this is the case only for inner products. For simplicity, we assume that $a = (-1, 0)$, $b = (1, 0)$ and that c is in the upper half plane. Define the function $h : [-1, 1] \longrightarrow [0, 1]$ by $\|(t, h(t))\| \equiv 1$, and compare this with the function $\tilde{h} : [-1, 1] \longrightarrow [0, 1]$ such that $\{(t, \tilde{h}(t))\}_{t \in [-1,1]}$ draws (the upper half of) the ellipse having ab and OD_0 as its long and short axes, where $D_0 = (0, \sup h)$ (Figure 4). We first suppose that $\sup h$ is attained at $t_0 > 0$, and put $c_0 = (t_0, h(t_0))$, $A_0 = (1, h(t_0))$, $B_0 = (-1, h(t_0))$. Then clearly the y-components of the vectors $|\triangle OA_0B_0| \cdot c_0$ and $|\triangle OA_0B_0| \cdot D_0$ are the same. Since only c has a nonzero y-component in (6.4), this implies that

$$|\triangle OAB| \cdot c - |\triangle O\widetilde{A}\widetilde{B}| \cdot D \in \mathbb{R} \times \{0\}$$

holds for all $t, t' \in (-1, 1)$, where $c = (t, h(t))$, $D = (t', \tilde{h}(t'))$, $\widetilde{A} = (1, \tilde{h}(t') + (1 - t')\tilde{h}'(t'))$ and $\widetilde{B} = (-1, \tilde{h}(t') - (1 + t')\tilde{h}'(t'))$.

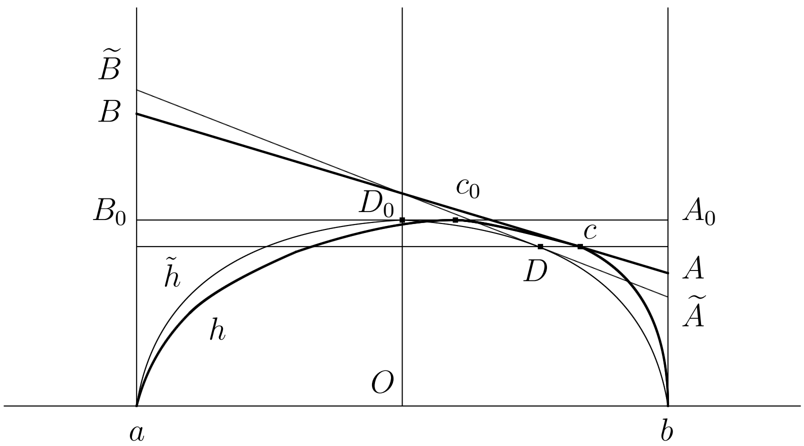

Figure 4

Thus, for any $c = (t, h(t))$ with $t \in (t_0, 1)$ and $t' \in (0, 1)$ with $\tilde{h}(t') = h(t)$, we obtain $|\triangle OAB| = |\triangle O\widetilde{A}\widetilde{B}|$ and hence $\tilde{h}'(t') < h'(t)$ (more precisely, AB and $\widetilde{A}\widetilde{B}$ must intersect on the y-axis). However, this is a contradiction since $\tilde{h}(1) = h(1) = 0$. We similarly derive a contradiction from $t_0 < 0$, so that $t_0 = 0$. Moreover, h must coincide with $\tilde{h}$ everywhere by a similar discussion. Therefore $\|\cdot\|$ is an inner product and we complete the proof of Claim 6.2.

Step 3 (n-dimensional case with $n \geq 3$) Suppose that $(\mathbb{R}^n, \|\cdot\|)$ is not an inner product space. Then there is a two-dimensional subspace $P \subset \mathbb{R}^n$ in which the restriction of $\|\cdot\|$ is not an inner product. We assume $P = \{(x, y, 0, \dots, 0) \,|\, x, y \in \mathbb{R}\}$ for brevity, and sometimes identify this with $\mathbb{R}^2$. By Step 2, there is a function $\rho_T : (\mathbb{R}^2, \|\cdot\||_P) \longrightarrow [0, \infty)$ such that $\int_{\mathbb{R}^2} \rho_T \, dx = 1$, $\operatorname{supp} \rho_T \subset B((-T, 0), r)$ for some fixed $r > 0$ and that

$$\lim_{T \to \infty} \int_{\operatorname{supp} \rho_T} [D(\|-\cdot\|)(x)]\big(\nabla(-\rho_T)(x)\big) \, dx < 0.$$

Using a smooth cut-off function $\eta : \mathbb{R}^{n-2} \longrightarrow [0, \infty)$ such that $\eta \equiv 1$ on $B(O, \sqrt{T})$, $\operatorname{supp} \eta \subset B(O, \sqrt{T} + 1)$ and that $\sup \|\nabla(-\eta)\| < 2$, define $\rho : (\mathbb{R}^n, \|\cdot\|) \longrightarrow [0, \infty)$ by $\rho(x, y) := (\int_{\mathbb{R}^{n-2}} \eta \, dz)^{-1} \rho_T(x) \eta(y)$ for $x \in \mathbb{R}^2$ and $y \in \mathbb{R}^{n-2}$. Note that $\nabla(-\rho)(x, y) = (\int_{\mathbb{R}^{n-2}} \eta \, dz)^{-1} \cdot \nabla(-\rho_T)(x)$ for $y \in B(O, \sqrt{T}) \subset \mathbb{R}^{n-2}$. Hence we have $\Theta(\rho) < 0$ since the effect of the boundary of the cut-off is negligible for large T. Indeed, we observe that

$$|B(O, \sqrt{T} + 1) \setminus B(O, \sqrt{T})| \cdot \left(\int_{\mathbb{R}^{n-2}} \eta \, dz \right)^{-1} = O\big((\sqrt{T})^{n-3} / (\sqrt{T})^{n-2}\big) \to 0$$

as T goes to infinity.

Step 4 (Scaling) Suppose that there is ρ with $\Theta(\rho) < 0$, and set $\rho_\varepsilon(x) := \varepsilon^{-n} \rho(\varepsilon^{-1} x)$ for $\varepsilon > 0$. Then we have

$$\begin{aligned}
&\int_{\operatorname{supp} \rho_\varepsilon} \|-x\| \cdot [D(\|-\cdot\|)(x)]\big(\nabla(-\rho_\varepsilon)(x)\big) \, dx \\
&= \varepsilon^{-(n+1)} \int_{\operatorname{supp} \rho_\varepsilon} \|-x\| \cdot [D(\|-\cdot\|)(x)]\big(\nabla(-\rho)(\varepsilon^{-1} x)\big) \, dx \\
&= \int_{\operatorname{supp} \rho} \|-x\| \cdot [D(\|-\cdot\|)(x)]\big(\nabla(-\rho)(x)\big) \, dx
\end{aligned}$$

and

$$\int_{\mathbb{R}^n} \|-x\|^2 \rho_\varepsilon(x) \, dx = \varepsilon^{-n} \int_{\mathbb{R}^n} \|-x\|^2 \rho(\varepsilon^{-1} x) \, dx = \varepsilon^2 \int_{\mathbb{R}^n} \|-x\|^2 \rho(x) \, dx.$$

Therefore $\Theta(\rho_\varepsilon) = \varepsilon^{-2} \Theta(\rho)$ and it diverges to $-\infty$ as ε tends to zero. Thus we complete the proof of Theorem 1.1.

References

[AGS] 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.

[Ar] D. G. Aronson, Bounds for the fundamental solution of a parabolic equation, Bull. Amer. Math. Soc. **73** (1967), 890–896.

[BCS] D. Bao, S.-S. Chern and Z. Shen, An introduction to Riemann-Finsler geometry, Springer-Verlag, New York, 2000.

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

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

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

[Mc] R. J. McCann, A convexity principle for interacting gases, Adv. Math. **128** (1997), 153–179.

[Mo] J. Moser, A Harnack inequality for parabolic differential equations, Comm. Pure Appl. Math. **17** (1964), 101–134.

[Na] J. Nash, Continuity of solutions of parabolic and elliptic equations, Amer. J. Math. **80** (1958), 931–954.

[Oh1] S. Ohta, Gradient flows on Wasserstein spaces over compact Alexandrov spaces, Amer. J. Math. **131** (2009), 475–516.

[Oh2] S. Ohta, Uniform convexity and smoothness, and their applications in Finsler geometry, Math. Ann. **343** (2009), 669–699.

[Oh3] S. Ohta, Finsler interpolation inequalities, Calc. Var. Partial Differential Equations **36** (2009), 211–249.

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

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

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

[Sal] L. Saloff-Coste, Uniformly elliptic operators on Riemannian manifolds, J. Differential Geom. **36** (1992), 417–450.

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

[Sh] Z. Shen, Lectures on Finsler geometry, World Scientific Publishing Co., Singapore, 2001.

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