

# Non-contraction of heat flow on Minkowski spaces

Shin-ichi Ohta\* and Karl-Theodor Sturm†

April 18, 2012

## 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 non-Riemannian Finsler manifolds. As an application, we show that the 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 an inner product for Riemannian manifolds. A Minkowski norm is a generalization of usual norms, and is not necessarily centrally symmetric. We will always assume that a Minkowski norm is strongly convex (and in particular strictly convex, see Subsection 2.1 for the definition).

In Riemannian manifolds, given  $K \in \mathbb{R}$ , it is well-known that the  $K$ -convexity of a function  $f$  along geodesics  $\gamma$  (i.e.,  $(f \circ \gamma)'' \geq K|\dot{\gamma}|^2$  in the weak sense) implies the  $K$ -contraction of the gradient flow of  $f$ , namely

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

holds for all  $t \geq 0$  and  $\xi, \zeta$  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

---

\*Department of Mathematics, Kyoto University, Kyoto 606-8502, Japan (sohta@math.kyoto-u.ac.jp) & Max-Planck-Institut 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).

*f.* In Finsler manifolds or even in strictly convex normed spaces, however, it has been unclear whether the gradient flows of convex functions are contractive (cf. [AGS, Introduction]). To avoid trivial counter-examples, of course strict convexity must be imposed (see Remark 4.1).

The point is that, although the aforementioned first variation 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. In Minkowski spaces, the squared norm is always 2-skew convex, while in general it is only  $K$ -convex for some  $K \geq 0$ . We also construct an explicit example of a convex function which is not 0-skew convex. This negatively answers the above question in [AGS] (see Section 4 for details).

In the second part of the article, we apply our technique to the heat flow on Minkowski spaces. Due to the celebrated work of Jordan et al [JKO], the heat flow on Euclidean spaces can be 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 along Wasserstein geodesics (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  $\text{Ric} \geq K$  ([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], [AGS], [Vi], [Er]). 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 works [Oh3], [OS1], 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 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 the heat flow on Minkowski spaces is contractive or not (see also the fourth comment in [Gi, Section 5]). Our main result gives a complete answer to this question.

**Theorem 1.1** *The heat flow on a Minkowski normed 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 (more generally, the curvature-dimension condition) works well for general Finsler manifolds. Among other characterizations of lower Ricci curvature

bounds for Riemannian manifolds, we recently verified that the Bochner-Weitzenböck formula makes sense for general Finsler manifolds ([OS2]).

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 squared norms of Minkowski spaces in Section 4. In Section 5, we discuss the 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 [OS1] are not directly applicable to noncompact spaces. Section 6 is devoted to a proof of Theorem 1.1. Finally, we consider the skew convexity of distance functions on Finsler manifolds in Appendix.

## 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  $\|\cdot\| : \mathbb{R}^n \rightarrow [0, \infty)$  satisfying the following conditions.

- (1) (*Positive homogeneity*)  $\|cx\| = c\|x\|$  holds for all  $x \in \mathbb{R}^n$  and  $c > 0$ .
- (2) (*Strong convexity*) The function  $\|\cdot\|^2/2$  is twice differentiable on  $\mathbb{R}^n \setminus \{0\}$ , and the symmetric matrix

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

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

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

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 (normed) space*. We remark that the strong convexity implies the strict convexity, i.e.,  $\|x+y\| < \|x\| + \|y\|$  unless  $x$  and  $y$  are linearly dependent. 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

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

This is the best approximation of the norm  $\|\cdot\|$  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  $x$ .

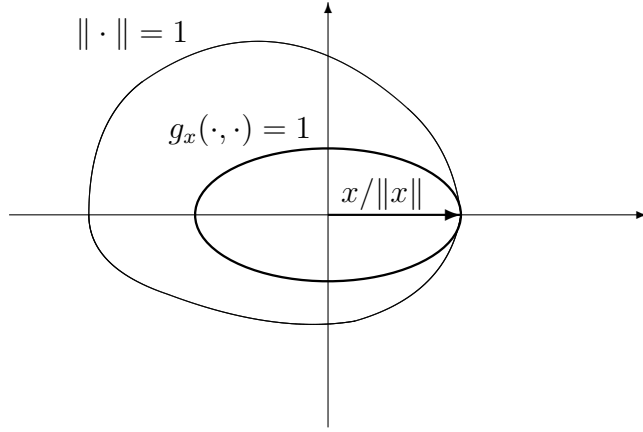


Figure 1

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

$$\begin{aligned} \left\| \frac{x+y}{2} \right\|^2 &\leq \frac{1}{2}\|x\|^2 + \frac{1}{2}\|y\|^2 - \frac{1}{4\mathcal{C}^2}\|x-y\|^2, \\ \left\| \frac{x+y}{2} \right\|^2 &\geq \frac{1}{2}\|x\|^2 + \frac{1}{2}\|y\|^2 - \frac{\mathcal{S}^2}{4}\|x-y\|^2 \end{aligned}$$

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

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

(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.

Denote by  $\|\cdot\|_*$  the dual norm of  $\|\cdot\|$ . Then 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. Moreover,  $\mathcal{L}(x) = (\mathcal{L}_j(x))_{j=1}^n$  can be explicitly written as

$$\mathcal{L}_j(x) = \frac{1}{2} \frac{\partial(\|\cdot\|^2)}{\partial x^j}(x) = \sum_{i=1}^n g_{ij}(x) x^i. \quad (2.5)$$

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  $\nabla f(x) := \mathcal{L}^*(Df(x)) \in T_x \mathbb{R}^n$  (identified with  $\mathbb{R}^n$ ).

**Remark 2.1** We need the strong convexity of the norm to formulate and investigate the skew convexity of functions 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). In addition, the strict convexity will be a necessary condition when one studies the contractivity of gradient flows (see Remark 4.1).

## 2.2 Finsler manifolds

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

For each  $v \in T_x M \setminus \{0\}$ , we define the inner product  $g_v$  on  $T_x M$  according to (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  $T_x M$  as  $v = \sum_{i=1}^n v^i (\partial/\partial x^i)|_x$  and define

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

We denote by  $\mathcal{C}(x)$  and  $\mathcal{S}(x)$  the 2-uniform convexity and smoothness constants of  $F|_{T_x M}$ . For a function  $f : M \rightarrow \mathbb{R}$  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  $\mathcal{L}^* : T_x^* M \rightarrow T_x M$ .

The *distance* from  $x$  to  $y$  is naturally defined as  $d(x, y) := \inf_{\gamma} \int_0^1 F(\dot{\gamma}) dt$ , where  $\gamma : [0, 1] \rightarrow M$  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*  $\gamma : [0, l] \rightarrow M$  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  $\gamma : [0, l] \rightarrow M$  is extended to a geodesic  $\gamma : [0, \infty) \rightarrow M$ . Then, for any  $x, y \in M$ , there is a minimal geodesic from  $x$  to  $y$  by the Hopf-Rinow theorem ([BCS, Theorem 6.6.1]).

Along a geodesic  $\gamma : [0, l] \rightarrow M$ ,  $\gamma(s)$  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  $\xi$  and  $\zeta$  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]):

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

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 the skew convexity of functions on a  $C^\infty$ -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 also 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  $f : M \rightarrow [-\infty, \infty]$  is said to be  $K$ -convex (or *geodesically  $K$ -convex*) for  $K \in \mathbb{R}$  if

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

holds for all geodesics  $\gamma : [0, 1] \rightarrow M$  and  $t \in [0, 1]$ . 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

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

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  $f : M \rightarrow \mathbb{R}$  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  $\gamma : [0, 1] \rightarrow M$  from  $x$  to  $y$  such that

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

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  $g_{\nabla(-f)(\gamma)}$ .

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

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

holds for all gradient curves  $\xi, \zeta$  and  $t \in [0, \infty)$ . Comparing (3.1) with (2.6), we verify that the  $K$ -contractivity is equivalent to the  $K$ -skew convexity.

**Theorem 3.2** *Let  $(M, F)$  be a forward complete Finsler manifold, and let  $f : M \rightarrow \mathbb{R}$  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  $\xi, \zeta : [0, \infty) \rightarrow M$  of  $f$  and set  $l(t) := d(\xi(t), \zeta(t))$ . Given  $t > 0$ , let  $\gamma : [0, 1] \rightarrow M$  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,

$$\begin{aligned} \limsup_{\varepsilon \downarrow 0} \frac{l(t+\varepsilon) - l(t)}{\varepsilon} &\leq \lim_{\varepsilon \downarrow 0} \frac{d(\xi(t+\varepsilon), \gamma(1/2)) - d(\xi(t), \gamma(1/2))}{\varepsilon} \\ &\quad + \lim_{\varepsilon \downarrow 0} \frac{d(\gamma(1/2), \zeta(t+\varepsilon)) - d(\gamma(1/2), \zeta(t))}{\varepsilon} \\ &= -g_{\dot{\gamma}(0)}(\dot{\gamma}(0)/l(t), \dot{\xi}(t)) + g_{\dot{\gamma}(1)}(\dot{\gamma}(1)/l(t), \dot{\zeta}(t)). \end{aligned}$$

By hypothesis, this yields  $l'(t) \leq -Kl(t)$  a.e.  $t$ . Therefore we deduce from Gronwall's theorem that  $d(\xi(t), \zeta(t)) \leq e^{-Kt}d(\xi(0), \zeta(0))$  holds.

To see the converse, suppose that the gradient flow of  $f$  is  $K$ -contractive and take a minimal geodesic  $\gamma : [0, 1] \rightarrow M$ . Dividing  $\gamma$  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

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

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

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

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

□

**Remark 3.3** We can replace the  $C^1$ -regularity in Definition 3.1 with the local semi-convexity as follows (cf., e.g., [Ly], [Oh1] for details). We say that a function  $f : M \rightarrow \mathbb{R}$  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  $\gamma : [0, 1] \rightarrow U$ . Define the *local slope* of  $f$  at  $x \in M$  as

$$|\nabla f|(x) := \limsup_{y \rightarrow x} \frac{\max\{f(x) - f(y), 0\}}{d(x, y)}. \quad (3.2)$$

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  $\gamma$  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, from any initial point, there starts a gradient curve  $\xi$  solving  $\dot{\xi}(t) = \nabla_- f(\xi(t))$  a.e.  $t$ . The  $K$ -skew convexity can be defined by using  $\nabla_- f$  instead of  $\nabla(-f)$  in (3.1), and the analogue of Theorem 3.2 holds by the same argument.

## 4 Skew convexity of squared norms

We study the skew convexity of the squared norm of a Minkowski space, and compare it with the usual convexity along straight lines. The more general case of distance functions on Finsler manifolds will be treated in Appendix.

Let  $(\mathbb{R}^n, \|\cdot\|)$  be a Minkowski space and set  $f(x) := \|x\|^2/2$ . Observe  $\nabla(-f)(x) = -x$ , so that the gradient curve  $\xi$  of  $f$  with  $\xi(0) = x$  is given by  $\xi(t) = e^{-t}x$ . Thus we see that the gradient flow of  $f$  is 1-contractive, which shows that  $f$  is 1-skew convex. This can be proved also by a direct calculation as, for any  $v \in \mathbb{R}^n \setminus \{0\}$ ,

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

**Remark 4.1** The above example obviously requires the strict convexity of the norm. In fact, even the uniqueness of gradient flows fails in non-strictly convex normed spaces. As pointed out by the referee, a typical example is the same function  $f(x) = \|x\|_\infty^2/2$  in the 2-dimensional  $\ell_\infty$ -space  $(\mathbb{R}^2, \|\cdot\|_\infty)$ . Any curve  $\xi(t) = (e^{-t}, h(t))$  satisfying  $|h(t)| \leq e^{-t}$  and  $|h'(t)| \leq e^{-t}$  is a gradient curve of  $f$  in the metric sense of [AGS].

In contrast to the 1-skew convexity of  $f$  above, we can find a norm  $\|\cdot\|$  of  $\mathbb{R}^n$  such that the function  $\tilde{f}(x) := \langle x, x \rangle / 2$  associated with the Euclidean inner product  $\langle \cdot, \cdot \rangle$  is not even 0-skew convex, although  $\tilde{f}$  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

$$g_v(v, \nabla(-\tilde{f})(0)) - g_v(v, \nabla(-\tilde{f})(-v)) = -g_v(v, \mathcal{L}^*(v)) = -\langle \mathcal{L}(v), \mathcal{L}^*(v) \rangle,$$

where we used (2.5) in the second equality. However,  $-\langle \mathcal{L}(v), \mathcal{L}^*(v) \rangle / \|v\|^2$  can be positive. An example is illustrated in Figure 2, where we set  $v = (0, 1) \in \mathbb{R}^2$ . The parallelogram rounded to be strictly convex is the unit sphere of the norm  $\|\cdot\|$ . Note that  $\langle \mathcal{L}(v), v \rangle = \|v\|^2 = 1$ ,  $\langle v, \mathcal{L}^*(v) \rangle = \|\mathcal{L}^*(v)\|^2$  and that  $\langle \mathcal{L}(v), \mathcal{L}^*(v) \rangle < 0$ .

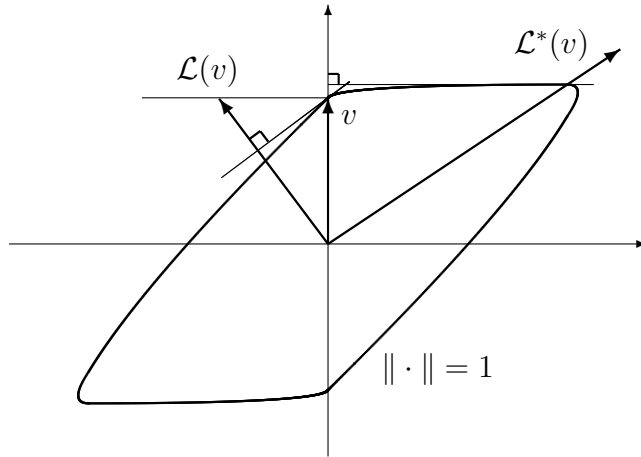


Figure 2

Given any  $K < 0$ , by scaling  $\tilde{f}$  (or the inner product) with sufficiently large  $C > 0$ , the convex function  $C\tilde{f}$  is not  $K$ -skew convex. This observation reveals that the skew convexity has no (obvious) relation with the usual convexity. In addition, via Theorem 3.2, we have seen that the usual convexity does not imply the contractivity in non-Euclidean normed spaces. This answers the question in [AGS] quoted in the introduction.

## 5 Heat flow on Minkowski spaces

In order to apply our technique to the 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  $\mu$  satisfying  $\int_{\mathbb{R}^n} \|x\|^2 d\mu < \infty$  (note that then  $\int_{\mathbb{R}^n} \|-x\|^2 d\mu < \infty$  holds as



well). The subset of absolutely continuous measures with respect to the Lebesgue measure  $dx$  will be denoted by  $\mathcal{P}_2^{\text{ac}}(\mathbb{R}^n) \subset \mathcal{P}_2(\mathbb{R}^n)$ .

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  $\mu$  and  $\nu$  if  $\pi(A \times \mathbb{R}^n) = \mu(A)$  and  $\pi(\mathbb{R}^n \times A) = \nu(A)$  hold for all Borel sets  $A \subset \mathbb{R}^n$ . We define the  $L^2$ -Wasserstein distance from  $\mu$  to  $\nu$  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  $\pi$  runs over all couplings of  $\mu$  and  $\nu$ . A coupling  $\pi$  attaining the infimum above is said to be *optimal*. 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  $\mathcal{C}^{-1}\|y\| \leq \sqrt{g_x(y, y)} \leq \mathcal{S}\|y\|$ . Then, if we denote by  $W_2^{g_x}$  the Wasserstein distance with respect to  $g_x$ , we have  $\mathcal{C}^{-1}W_2(\mu, \nu) \leq W_2^{g_x}(\mu, \nu) \leq \mathcal{S}W_2(\mu, \nu)$ . This relation is sometimes useful to apply known results in the Euclidean case.

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

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

We introduce a Finsler structure of the Wasserstein space along the line of [Ot], see [OS1] for more details in the case of compact Finsler manifolds. We set

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

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

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

(of the space of measurable vector fields  $\Phi$  with  $F_\mu(\Phi) < \infty$ ). Similarly, the cotangent space  $(T_\mu^*\mathcal{P}, F_\mu^*)$  is defined as the completion of  $\hat{T}^*\mathcal{P} := \{\alpha = D\varphi \mid \varphi \in C_c^\infty(\mathbb{R}^n)\}$  with respect to  $F_\mu^*(\alpha) := (\int_{\mathbb{R}^n} \|\alpha\|_*^2 d\mu)^{1/2}$ . We define the Legendre transform  $\mathcal{L}_\mu^* : T_\mu^*\mathcal{P} \rightarrow T_\mu\mathcal{P}$  in the pointwise way that  $\mathcal{L}_\mu^*(D\varphi) = \nabla\varphi$ .

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 function  $h \in L^2(I)$  such that

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

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* (or the *metric speed*)

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

is well-defined for a.e.  $t \in I$  ([AGS, Theorem 1.1.2], [OS1, Lemma 7.1]). We can associate  $(\mu_t)_{t \in I}$  with a Borel vector field  $\Phi$  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.  $t \in I$ ,
- the *continuity equation*  $\partial_t \mu_t + \operatorname{div}(\Phi_t \mu_t) = 0$  in the weak sense that

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

for all  $\psi \in C_c^\infty(I \times \mathbb{R}^n)$  ([AGS, Theorem 8.3.1], [OS1, Theorem 7.3]).

Such a vector field  $\Phi$  is unique up to a difference on a null measure set with respect to  $d\mu_t dt$ , and we have  $F_{\mu_t}(\Phi_t) = |\dot{\mu}_t|$  a.e.  $t \in I$ . We will call  $\Phi$  the *tangent vector field* of the curve  $(\mu_t)_{t \in I}$  and write  $\dot{\mu}_t = \Phi_t$ .

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

$$\int_{\mathbb{R}^n} \alpha(\Phi) d\mu \geq \limsup_{t \rightarrow 0} \frac{Q(\mu_t) - Q(\mu)}{t} \quad (5.2)$$

holds for every minimal geodesic  $(\mu_t)_{t \in [0, 1]}$  with  $\mu_t = (T_t)_\# \mu$  and  $T_t(x) = x + t\Phi(x)$ , and that equality holds in (5.2) if  $\Phi \in \hat{T} \mathcal{P}$  (with  $\lim_{t \rightarrow 0}$  in place of  $\limsup_{t \rightarrow 0}$ ). Such a one-form  $\alpha$  is unique in  $T_\mu^* \mathcal{P}$  up to a difference on a  $\mu$ -null measure set. Thus we write  $DQ(\mu) = \alpha$  and define the *gradient vector* of  $Q$  at  $\mu$  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  $-Q$  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*  $\Delta$  acting on  $u \in H_{\text{loc}}^1(\mathbb{R}^n)$  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  $\Delta$  is a nonlinear operator since the Legendre transform is nonlinear (unless the norm  $\|\cdot\|$  comes from an inner product). We consider the associated *heat equation*  $\partial_t u = \Delta u$  also in the weak form. We have seen in [OS1,

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

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

holds for all  $t \in [0, T]$  and  $\psi \in H_0^1(\mathbb{R}^n)$ .

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, time-dependent differential operator

$$\Delta^{\nabla u} v := \operatorname{div} \left( \sum_{i,j=1}^n g^{ij}(\nabla u) \frac{\partial v}{\partial x^i} \frac{\partial}{\partial x^j} \right), \quad (5.4)$$

where  $(g^{ij})$  stands for 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). 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  $H_{\text{loc}}^2$  in  $x$  and  $C^{1,\alpha}$  on  $(0, \infty) \times \mathbb{R}^n$  ([OS1, Theorems 4.6, 4.9]).

The following lemma allows us to consider  $(u_t \, dx)_{t \geq 0}$  as a curve in  $\mathcal{P}_2(\mathbb{R}^n)$ .

**Lemma 5.3** *Let  $(u_t)_{t \geq 0} \subset H_0^1(\mathbb{R}^n)$  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) *If  $u_0 \, dx \in \mathcal{P}_2(\mathbb{R}^n)$ , then  $u_t \, dx \in \mathcal{P}_2(\mathbb{R}^n)$  for all  $t > 0$ .*
- (iii) (Continuity in  $W_2$ ) *If  $u_0 \, dx \in \mathcal{P}_2(\mathbb{R}^n)$ , then  $\lim_{t \rightarrow 0} W_2(u_0 \, dx, u_t \, dx) = 0$ .*

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

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

where  $\|\cdot\|$  stands for the Euclidean norm and  $C_1, C_2$  depend only on  $\|\cdot\|$ . Thus we obtain

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

where we used the fact  $se^{-s} \leq e^{-s/2}$  in the last inequality and  $C_3$  depends only on  $C_1, C_2$ , the norm  $\|\cdot\|$  and  $n$ .

(iii) It suffices to show that  $u_t dx$  weakly converges to  $u_0 dx$  and

$$\lim_{R \rightarrow \infty} \limsup_{t \rightarrow 0} \int_{\{|x| \geq R\}} |x|^2 u_t(x) dx = 0$$

by Remark 5.1(a) (cf. [Vi, Theorem 6.9]). For the weak convergence, thanks to (i) and [AGS, Remark 5.1.6], it is sufficient to show the convergence for test functions  $f \in C_c^\infty(\mathbb{R}^n)$ . This immediately follows from (5.3), indeed, the Cauchy-Schwarz inequality yields

$$\left| \int_0^\tau \int_{\mathbb{R}^n} f \partial_t u dx dt \right| \leq \Lambda \left( \tau \int_{\mathbb{R}^n} \|Df\|_*^2 dx \right)^{1/2} \left( \int_0^\tau \int_{\mathbb{R}^n} \|\nabla u_t\|^2 dx dt \right)^{1/2} \rightarrow 0$$

as  $\tau$  tends to zero, where  $\Lambda = \sup_{x \in \mathbb{R}^n \setminus \{0\}} \|-x\|/\|x\|$ . The latter condition can be seen similarly to (ii). As  $|x| < 2|x-y|$  if  $|x| \geq R$  and  $|y| < R/2$ , we have

$$\begin{aligned} \int_{\{|x| \geq R\}} |x|^2 u_t(x) dx &\leq \int_{\{|x| \geq R\}} \int_{\{|y| \geq R/2\}} 2(|x-y|^2 + |y|^2) q^u(t, x; 0, y) u_0(y) dy dx \\ &\quad + \int_{\{|x| \geq R\}} \int_{\{|y| < R/2\}} 4|x-y|^2 q^u(t, x; 0, y) u_0(y) dy dx \\ &\leq Ct + 2 \int_{\{|y| \geq R/2\}} |y|^2 u_0(y) dy \rightarrow 0 \end{aligned}$$

as  $t \rightarrow 0$  and then  $R \rightarrow \infty$ . □

### 5.3 Relative entropy and heat flow as its gradient flow

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

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

for  $\mu = \rho dx \in \mathcal{P}_2^{\text{ac}}(\mathbb{R}^n)$ , and  $\text{Ent}(\mu) := \infty$  for  $\mu \in \mathcal{P}_2(\mathbb{R}^n) \setminus \mathcal{P}_2^{\text{ac}}(\mathbb{R}^n)$ . See [JKO, Section 4] (and Remark 5.1(a)) for the fact  $\text{Ent} > -\infty$  on  $\mathcal{P}_2^{\text{ac}}(\mathbb{R}^n)$ . We know that  $\text{Ent}$  is convex along geodesics in  $(\mathcal{P}_2(\mathbb{R}^n), W_2)$  ([Vi, page 908]). 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 [OS1], [AGS] and [Er]. The next lemma corresponds to [OS1, Proposition 7.7], see also [AGS, Theorem 10.4.17] and [Er, Proposition 4.3].

**Lemma 5.4** *For  $\mu = \rho dx \in \mathcal{P}_2^{\text{ac}}(\mathbb{R}^n)$  with  $\rho \in H_0^1(\mathbb{R}^n)$  and  $\text{Ent}(\mu) < \infty$ , the following are equivalent.*

- (I)  $-\text{Ent}$  is differentiable at  $\mu$ .
- (II)  $\|\nabla(-\rho)\|/\rho \in L^2(\mathbb{R}^n, \mu)$ .

Moreover, then  $\nabla(-\rho)/\rho \in T_\mu\mathcal{P}$  and we have

$$\nabla_W(-\text{Ent})(\mu) = \frac{\nabla(-\rho)}{\rho}.$$

*Proof.* (I)  $\Rightarrow$  (II): It suffices to see that  $|\nabla \text{Ent}|(\mu)$  defined as in (3.2) is finite, then [Er, Proposition 4.3] yields (II) because our norm is comparable to a Euclidean norm. Take  $\nu \in \mathcal{P}_2^{\text{ac}}(\mathbb{R}^n)$  with  $\text{Ent}(\nu) < \infty$  and let  $\mu_t = (T_t)_\# \mu$  with  $T_t(x) = x + t\Phi(x)$  be the minimal geodesic from  $\mu_0 = \mu$  to  $\mu_1 = \nu$ . We deduce from the convexity of  $\text{Ent}(\mu_t)$  that

$$\limsup_{t \rightarrow 0} \frac{\text{Ent}(\mu) - \text{Ent}(\mu_t)}{W_2(\mu, \mu_t)} \geq \frac{\text{Ent}(\mu) - \text{Ent}(\nu)}{W_2(\mu, \nu)}.$$

Thanks to the differentiability of  $-\text{Ent}$ , we observe

$$\limsup_{t \rightarrow 0} \frac{\text{Ent}(\mu) - \text{Ent}(\mu_t)}{t} \leq \int_{\mathbb{R}^n} [D(-\text{Ent})(\mu)](\Phi) d\mu \leq F_\mu(\nabla_W(-\text{Ent})(\mu)) \cdot W_2(\mu, \nu).$$

Therefore we have  $|\nabla \text{Ent}|(\mu) \leq F_\mu(\nabla_W(-\text{Ent})(\mu)) < \infty$ .

(II)  $\Rightarrow$  (I): Again due to [Er, Proposition 4.3], we obtain  $|\nabla \text{Ent}|(\mu) < \infty$  as well as  $D(-\rho)/\rho \in T_\mu^*\mathcal{P}$ . For any minimal geodesic  $(\mu_t)_{t \in [0,1]}$  as in (I)  $\Rightarrow$  (II), [Vi, Theorem 23.14] (see also [OS1, Proposition 7.7] for the compact case) shows that

$$\lim_{t \rightarrow 0} \frac{\text{Ent}(\mu) - \text{Ent}(\mu_t)}{t} = \int_{\mathbb{R}^n} \text{div}(\Phi) d\mu \leq \int_{\mathbb{R}^n} \left[ \frac{D(-\rho)}{\rho} \right](\Phi) d\mu,$$

and equality holds if  $\Phi \in \hat{T}\mathcal{P}$ . Thus  $-\text{Ent}$  is differentiable and it holds  $\nabla_W(-\text{Ent})(\mu) = \nabla(-\rho)/\rho$ .  $\square$

The following theorem is a slight modification of [OS1, Theorem 7.8] adapted to non-compact 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

$$\int_{\mathbb{R}^n} \psi \partial_t u dx = - \int_{\mathbb{R}^n} D\psi(\overleftarrow{\nabla}u) dx = \int_{\mathbb{R}^n} D\psi(\nabla(-u)) dx. \quad (5.5)$$

**Theorem 5.5 (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  $\rho_0 dx \in \mathcal{P}_2^{\text{ac}}(\mathbb{R}^n)$ . Then  $\mu_t := \rho_t dx$  is a gradient curve of the relative entropy (in the sense of Definition 5.2).*

(ii) *Conversely, let  $(\mu_t)_{t \geq 0} \subset \mathcal{P}_2^{\text{ac}}(\mathbb{R}^n)$  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  $\rho_t$  is a global solution to the reverse heat equation.*

*Proof.* (i) We first of all remark that  $\text{Ent}(\mu_t) < \infty$  for all  $t > 0$  by the upper Gaussian estimate for fundamental solutions (see the proof of Lemma 5.3(ii)). It follows from the reverse heat equation (5.5) that  $\nabla(-\rho)/\rho$  satisfies the continuity equation (5.1) along the curve  $(\mu_t)_{t \geq 0}$ . More generally, we have

$$\int_{\mathbb{R}^n} \psi_T d\mu_T - \int_{\mathbb{R}^n} \psi_\tau d\mu_\tau = \int_\tau^T \int_{\mathbb{R}^n} \left\{ \partial_t \psi + D\psi \left( \frac{\nabla(-\rho_t)}{\rho_t} \right) \right\} d\mu_t dt \quad (5.6)$$

for all  $\psi \in C_c^\infty([0, \infty) \times \mathbb{R}^n)$  and  $0 < \tau < T < \infty$ . Then we obtain, by choosing a test function  $\psi$  approximating  $\log(\max\{\rho_t, \varepsilon\}) - \log \varepsilon$  on  $[\tau, T] \times \mathbb{R}^n$  and then letting  $\varepsilon \downarrow 0$ ,

$$\int_\tau^T \int_{\mathbb{R}^n} \frac{\|\nabla(-\rho_t)\|^2}{\rho_t^2} d\mu_t dt = \text{Ent}(\mu_\tau) - \text{Ent}(\mu_T) < \infty.$$

Hence  $(\mu_t)_{t>0}$  is absolutely continuous (see [AGS, Theorem 8.3.1]) and, as  $T > 0$  was arbitrary, we find  $\|\nabla(-\rho_t)\|/\rho_t \in L^2(\mathbb{R}^n, \mu_t)$  at a.e.  $t$ . This implies that  $-\text{Ent}$  is differentiable at  $\mu_t$  and  $\nabla_W(-\text{Ent})(\mu_t) = \nabla(-\rho_t)/\rho_t \in T_{\mu_t}\mathcal{P}$  a.e.  $t$  (Lemma 5.4). Combining  $\nabla(-\rho_t)/\rho_t \in T_{\mu_t}\mathcal{P}$  with (5.6), we conclude  $\dot{\mu}_t = \nabla(-\rho_t)/\rho_t$  and thus  $(\mu_t)_{t \geq 0}$  is a gradient curve of  $\text{Ent}$  in the sense of Definition 5.2.

(ii) Note that Lemma 5.4 ensures  $\dot{\mu}_t = \nabla_W(-\text{Ent})(\mu_t) = \nabla(-\rho_t)/\rho_t$  for a.e.  $t$ . Then the continuity equation with  $\Phi = \nabla(-\rho)/\rho$  immediately implies the reverse heat equation.  $\square$

**Remark 5.6** The formula  $\nabla_W(-\text{Ent})(\mu) = \nabla(-\rho)/\rho$  in Lemma 5.4 has an extra importance in the Finsler/Minkowski setting. The reverse heat equation (5.5) is rewritten via the integration by parts as  $\int_{\mathbb{R}^n} \psi \partial_t \rho dx = \int_{\mathbb{R}^n} \Delta^{\nabla(-\rho)} \psi \rho dx$  (in other words,  $\overleftarrow{\Delta} \rho = \Delta^{\nabla(-\rho)} \rho$ , see (5.4)). Then the homogeneity  $g_{\nabla(-\rho)} = g_{\nabla(-\rho)/\rho}$  guarantees that a *formal* calculation with respect to the time-dependent Riemannian structure  $g_{\nabla(-\rho)}$  verifies Theorem 5.5.

## 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 5.7 (First variation formula for  $W_2$  along heat flow)** *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 (5.5) such that  $\mu_t = \rho_t dx \in \mathcal{P}_2^{\text{ac}}(\mathbb{R}^n)$  and  $\nu_t = \sigma_t dx \in \mathcal{P}_2^{\text{ac}}(\mathbb{R}^n)$ , we have*

$$\lim_{\tau \downarrow t} \frac{W_2(\mu_\tau, \nu_\tau)^2 - W_2(\mu_t, \nu_t)^2}{2(\tau - t)} = \int_{\mathbb{R}^n} g_{\dot{\omega}_1}(\dot{\omega}_1, \dot{\nu}_t) d\nu_t - \int_{\mathbb{R}^n} g_{\dot{\omega}_0}(\dot{\omega}_0, \dot{\mu}_t) d\mu_t \quad (5.7)$$

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

*Proof.* Set  $l(t) := W_2(\mu_t, \nu_t)$ , fix  $\delta > 0$  and define  $\tilde{\Gamma}$  as the set of continuous curves  $\xi : [t, t + \delta] \rightarrow \mathbb{R}^n$  endowed with the uniform topology. For  $\tau \in [t, t + \delta]$ , we define the evaluation map  $e_\tau : \tilde{\Gamma} \rightarrow \mathbb{R}^n$  by  $e_\tau(\xi) := \xi(\tau)$ . Then there exist probability measures  $\Pi, \Xi \in \mathcal{P}(\tilde{\Gamma})$  such that  $(e_\tau)_\# \Pi = \mu_\tau$ ,  $(e_\tau)_\# \Xi = \nu_\tau$  for all  $\tau \in [t, t + \delta]$  and that  $\Pi, \Xi$  are concentrated on the set of  $C^1$ -curves  $\xi, \zeta$  solving

$$\dot{\xi}(\tau) = \frac{\nabla(-\rho_\tau)}{\rho_\tau}(\xi(\tau)), \quad \dot{\zeta}(\tau) = \frac{\nabla(-\sigma_\tau)}{\sigma_\tau}(\zeta(\tau))$$

for all  $\tau \in [t, t + \delta]$ , respectively (cf. [AGS, Theorem 8.2.1]). We remark that  $\rho_\tau, \sigma_\tau > 0$  for  $\tau > 0$  by the lower Gaussian estimate.

To see ‘ $\leq$ ’ of (5.7), we disintegrate  $\Pi$  and  $\Xi$  by using  $\mu_t$  and  $\nu_t$  as  $d\Pi = d\Pi_x^t d\mu_t(x)$  and  $d\Xi = d\Xi_y^t d\nu_t(y)$ , where  $\Pi_x^t, \Xi_y^t \in \mathcal{P}(\tilde{\Gamma})$  concentrate on the sets  $e_t^{-1}(x)$  and  $e_t^{-1}(y)$ , respectively. Take the unique minimal geodesic  $\omega : [0, 1] \rightarrow \mathcal{P}_2(\mathbb{R}^n)$  from  $\mu_t$  to  $\nu_t$ , and let  $\pi_t$  be the unique optimal coupling of  $\mu_t$  and  $\nu_t$ . Then we find, for each  $\tau \in [t, t + \delta]$ ,

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

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

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

for  $\Pi$ -a.e.  $\xi$  and  $\Xi$ -a.e.  $\zeta$ . Since  $\rho$  and  $\sigma$  are  $C^1$  on  $(0, \infty) \times \mathbb{R}^n$ , 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  $\varepsilon$  tends to zero, we observe from

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

that

$$\begin{aligned} & \int_{\mathbb{R}^n \times \mathbb{R}^n} \int_{\tilde{\Gamma} \times \tilde{\Gamma}} \frac{\|\zeta(\tau) - \xi(\tau)\|^2 - \|\zeta(t) - \xi(t)\|^2}{2(\tau - t)} d\Pi_x^t(\xi) d\Xi_y^t(\zeta) d\pi_t(x, y) \\ & \leq \left( \int_{\mathbb{R}^n \times \mathbb{R}^n} \int_{\tilde{\Gamma} \times \tilde{\Gamma}} \{\|\zeta(\tau) - \xi(\tau)\|^2 + \|\zeta(t) - \xi(t)\|^2\} d\Pi_x^t(\xi) d\Xi_y^t(\zeta) d\pi_t(x, y) \right)^{1/2} \\ & \quad \times \left( \int_{\tilde{\Gamma}} \frac{\|\zeta(\tau) - \zeta(t)\|^2}{(\tau - t)^2} d\Xi(\zeta) + \int_{\tilde{\Gamma}} \frac{\|\xi(t) - \xi(\tau)\|^2}{(\tau - t)^2} d\Pi(\xi) \right)^{1/2}. \end{aligned}$$

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

$$\begin{aligned} & \frac{1}{(\tau - t)^2} \int_{\tilde{\Gamma}} \|\zeta(\tau) - \zeta(t)\|^2 d\Xi(\zeta) \leq \frac{1}{(\tau - t)^2} \int_{\tilde{\Gamma}} \left( \int_t^\tau \|\dot{\zeta}(s)\| ds \right)^2 d\Xi(\zeta) \\ & \leq \frac{1}{\tau - t} \int_{\tilde{\Gamma}} \int_t^\tau \|\dot{\zeta}(s)\|^2 ds d\Xi(\zeta) = \frac{1}{\tau - t} \int_t^\tau |\dot{\nu}_s|^2 ds. \end{aligned}$$

Therefore we obtain

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

Note that

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

Hence we have

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

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

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

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

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

Thus we find

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

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

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

weakly converges to  $\pi_t$  as  $\tau \downarrow t$  due to [AGS, Lemma 10.2.8]. Precisely, as  $\tilde{\pi}_t^\tau$  is a coupling of  $\mu_t$  and  $\nu_t$ , 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

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

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



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

**Proposition 5.8 (Skew convexity versus Wasserstein contraction)** *For  $K \in \mathbb{R}$ , the following are equivalent.*

- (I) *The relative entropy is  $K$ -skew convex in the sense that, for any  $\mu = \rho dx \in \mathcal{P}_2^{\text{ac}}(\mathbb{R}^n)$  such that  $\rho \in H_0^1(\mathbb{R}^n) \cap C^1(\mathbb{R}^n)$ ,  $\text{Ent}(\mu) < \infty$ ,  $\|\nabla(-\rho)/\rho\| \in L^2(\mathbb{R}^n, \mu)$  and for any  $\nu = \sigma dx \in \mathcal{P}_2^{\text{ac}}(\mathbb{R}^n)$  satisfying the same conditions, it holds*

$$\int_{\mathbb{R}^n} g_{\dot{\omega}_1}(\dot{\omega}_1, \nabla_W(-\text{Ent})(\nu)) d\nu - \int_{\mathbb{R}^n} g_{\dot{\omega}_0}(\dot{\omega}_0, \nabla_W(-\text{Ent})(\mu)) d\mu \leq -KW_2(\mu, \nu)^2, \quad (5.8)$$

where  $\omega : [0, 1] \rightarrow \mathcal{P}_2(\mathbb{R}^n)$  is the minimal geodesic from  $\mu$  to  $\nu$ .

- (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 such that  $\mu_t := \rho_t dx \in \mathcal{P}_2^{\text{ac}}(\mathbb{R}^n)$  and  $\nu_t := \sigma_t dx \in \mathcal{P}_2^{\text{ac}}(\mathbb{R}^n)$ , we have*

$$W_2(\mu_t, \nu_t) \leq e^{-Kt} W_2(\mu_0, \nu_0) \quad \text{for all } t \in [0, \infty). \quad (5.9)$$

We remark that, in the implication (II)  $\Rightarrow$  (I), we need the  $C^1$ -regularity of  $\rho$  and  $\sigma$  to apply (the proof of) Proposition 5.7 at  $t = 0$ . Indeed, as each spatial derivative  $v := \partial\rho/\partial x^k$  ( $k = 1, 2, \dots, n$ ) again solves the linear parabolic equation  $\partial_t v = \Delta^{\nabla(-\rho)} v$  ([OS1, Lemma 4.7]), the upper Gaussian estimate for the fundamental solution ensures that  $\partial\rho_t/\partial x^k$  tends to  $\partial\rho_0/\partial x^k$  locally uniformly.

As a corollary to Proposition 5.8, we obtain the 0-contraction of gradient curves in a special class of *symmetric* measures (compare this with Step 0 in the next section).

**Corollary 5.9 (Non-expansion for Gaussian measures)** *Let  $(\mathbb{R}^n, \|\cdot\|)$  be a symmetric normed space (i.e.,  $\| -x \| = \|x\|$ ). Take two probability 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 normalizing constant  $C > 0$ . Then the gradient curves  $(\mu_t)_{t \geq 0}, (\nu_t)_{t \geq 0}$  of  $\text{Ent}$  starting from them satisfies  $W_2(\mu_t, \nu_t) \leq W_2(\mu_0, \nu_0)$  for all  $t \geq 0$ .

*Proof.* Without loss of generality, we assume  $y = 0$  and  $a \geq b$ . Solving the heat equation, we observe

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

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

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

We can explicitly write as

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

It follows from Lemma 5.4 that

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

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  $\mu_t$ . If  $a > b$ , then we put  $z' = s'z = \{1 - \sqrt{(t+b)/(t+a)}\}^{-1}z$  and deduce  $T_{s'}(x) \equiv z'$  (note that  $s' > 1$ ). Thus we have

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

and  $[D(\|z' - \cdot\|^2)(x)](x) + [D(\|z' - \cdot\|^2)(-x)](-x) \geq 0$  by the convexity of  $\|z' - \cdot\|^2$  (along with the symmetry of  $\|\cdot\|$ ). Therefore we obtain  $\Theta \geq 0$ , and Proposition 5.8 completes the proof.  $\square$

## 6 Non-contraction of heat flow

This section is devoted to a proof of Theorem 1.1. For notational simplicity, we prove this for the reverse norm. That is to say, global solutions to the reverse heat equation (5.3) are not  $K$ -contractive with respect to  $W_2$ .

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

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

Then  $(\omega_s)_{s \in [0, T]}$  is the unique minimal geodesic from  $\mu$  to the Dirac measure  $\delta_O$  at the origin  $O \in \mathbb{R}^n$ , 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 (5.8) is false for any given  $K \in \mathbb{R}$  (i.e.,  $\text{Ent}$  is not  $K$ -skew convex) by choosing suitable  $\rho$ .

We deduce from  $\nabla_W(-\text{Ent})(\omega_s) = \nabla(-\rho^s)/\rho^s$  (Lemma 5.4) that

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

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

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

Thus we have

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

and hence

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

Note that  $W_2(\omega_0, \omega_1)^2 = T^{-2} \int_{\mathbb{R}^n} \|-x\|^2 \rho(x) dx$  and, by putting  $f(x) := \|-x\|^2/2$  (as in Section 4),

$$g_{-x}(-x, \nabla(-\rho)(x)) = [D(-f)(x)](\nabla(-\rho)(x)).$$

We set

$$\Theta(\rho) := \frac{\int_{\text{supp } \rho} [D(-f)(x)](\nabla(-\rho)(x)) dx}{\int_{\mathbb{R}^n} \|-x\|^2 \rho(x) dx} \quad (6.1)$$

and shall demonstrate that  $\Theta(\rho)$  can be positive (Steps 0–2 below) and even arbitrarily large (Step 3) by choosing suitable  $\rho$ , unless  $\|\cdot\|$  is an inner product. This means that (5.8) is false for any  $K \in \mathbb{R}$ , and completes the proof of Theorem 1.1. We start with an explicit example describing the heart of our construction.

**Step 0 (The model case of  $\ell_p^2$  with  $2 < p < \infty$ )** Let  $\|\cdot\|$  be the  $\ell_p$ -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).

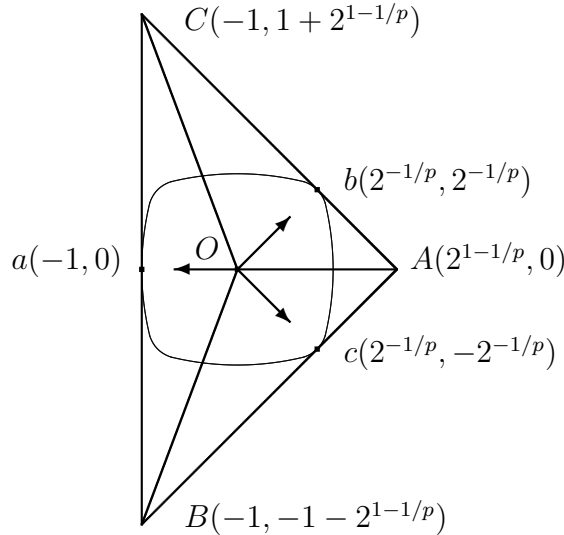


Figure 3

Define the nonnegative function  $\hat{\rho} : \mathbb{R}^2 \rightarrow [0, \infty)$  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

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

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

Now, for large  $R > 1$ , we consider the function  $\hat{\rho}_R(x) := \hat{\rho}(x + (R, 0))$ . Then it follows from (6.2) and  $\nabla(-f)(x) = -x$  that

$$\lim_{R \rightarrow \infty} \int_{\text{supp } \hat{\rho}_R} \left[ D\left(\frac{-f}{R}\right)(x) \right] (\nabla(-\hat{\rho}_R)(x)) dx = (1 + 2^{1-1/p})(2^{1-2/p} - 1)\sigma > 0.$$

Therefore, by taking a smooth approximation of  $\hat{\rho}_R$  (satisfying the conditions imposed on  $\rho$  at the beginning of the section) for sufficiently large  $R$ , we find  $\rho$  satisfying  $\Theta(\rho) > 0$ .

**Step 1 (General two-dimensional case)** The argument in Step 0 shows 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*

$$|\triangle OAB| \cdot c + |\triangle OBC| \cdot a + |\triangle OCA| \cdot b \quad (6.3)$$

*is nonzero, where  $|\triangle OAB|$  denotes the area of  $\triangle OAB$  with respect to the Lebesgue measure. Then there exists a function  $\rho$  for which  $\Theta(\rho)$  as in (6.1) is positive.*

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  $e_1 = (1, 0)$  and  $e_2 = (0, 1)$  by the fundamental theorem of calculus applied to the function  $\hat{\rho}$  defined as in Step 0.) Claim 6.1 is sharp enough for our purpose, as we can certainly verify the following.

**Claim 6.2** *There exists 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  $b'a'$  ( $a'$  and  $b'$  are determined only by  $C'$ ). By letting  $C'$  go to infinity,  $a'$

and  $b'$  can be arbitrarily 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  $\lambda$ . 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 will 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 Euclidean inner product. As in the nonsymmetric 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

$$|\triangle OaB| \cdot a + |\triangle OAb| \cdot b + |\triangle OAB| \cdot c \quad (6.4)$$

is independent of the choice of  $c$  on the arc between  $a$  and  $b$  opposite to  $C$  (since  $\triangle a'b'c$  corresponding to  $\triangle A'B'C$  was independent of the choice 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  $h : [-1, 1] \rightarrow [0, 1]$  by  $\|(t, h(t))\| \equiv 1$ , and compare this with the function  $\tilde{h} : [-1, 1] \rightarrow [0, 1]$  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  $c_0 = (t_0, h(t_0))$ ,  $A_0 = (1, h(t_0))$  and  $B_0 = (-1, h(t_0))$ . Then, on the one hand, 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. On the other hand, since only  $c$  has a nonzero  $y$ -component in (6.4),  $|\triangle OAB| \cdot c$  and  $|\triangle OA_0B_0| \cdot c_0$  have the same  $y$ -component. Similarly, for any  $t' \in (-1, 1)$  and points  $D := (t', \tilde{h}(t'))$ ,  $\tilde{A} := (1, \tilde{h}(t') + (1 - t')\tilde{h}'(t'))$  and  $\tilde{B} := (-1, \tilde{h}(t') - (1 + t')\tilde{h}'(t'))$  corresponding to the ellipse drawn by  $\tilde{h}$ ,  $|\triangle O\tilde{A}\tilde{B}| \cdot D$  has the same  $y$ -component as  $|\triangle OA_0B_0| \cdot D_0$ . Hence we have

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

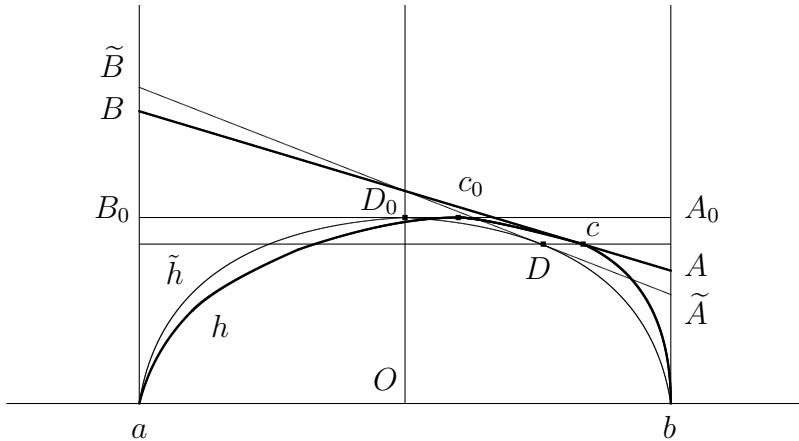


Figure 4

In particular, 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  $|\triangle OAB| = |\triangle O\tilde{A}\tilde{B}|$  and hence  $\tilde{h}'(t') < h'(t)$  (more precisely,  $AB$  and  $\tilde{A}\tilde{B}$  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  $t_0 = 0$ . Furthermore,  $h$  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 2 ( $n$ -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  $P = \{(x, y, 0, \dots, 0) \mid x, y \in \mathbb{R}\}$  for brevity, and sometimes identify this with  $\mathbb{R}^2$ . By Step 1, there is a function  $\rho_R : (\mathbb{R}^2, \|\cdot\||_P) \rightarrow [0, \infty)$  such that  $\int_{\mathbb{R}^2} \rho_R dx = 1$ ,  $\text{supp } \rho_R \subset B((-R, 0), r)$  for some fixed  $r > 0$  and that

$$\lim_{R \rightarrow \infty} \int_{\text{supp } \rho_R} \left[ D\left(\frac{-f}{R}\right)(x) \right] (\nabla(-\rho_R)(x)) dx > 0,$$

where we set  $B(z, r) := \{w \in \mathbb{R}^2 \mid \|w - z\| < r\}$  for  $z \in \mathbb{R}^2$  and  $r > 0$ . Using a smooth cut-off function  $\eta_R : \mathbb{R}^{n-2} \rightarrow [0, \infty)$  such that  $\eta_R \equiv 1$  on  $B(O, \sqrt{R})$ ,  $\text{supp } \eta_R \subset B(O, \sqrt{R} + 1)$  and that  $\sup \|\nabla(-\eta_R)\| < 2$ , define  $\rho : (\mathbb{R}^n, \|\cdot\|) \rightarrow [0, \infty)$  by

$$\rho(x, y) := \left( \int_{\mathbb{R}^{n-2}} \eta_R dz \right)^{-1} \rho_R(x) \eta_R(y)$$

for  $x \in \mathbb{R}^2$  and  $y \in \mathbb{R}^{n-2}$ . Note that  $\nabla(-\rho)(x, y) = \left( \int_{\mathbb{R}^{n-2}} \eta_R dz \right)^{-1} \cdot \nabla(-\rho_R)(x)$  for  $y \in B(O, \sqrt{R}) \subset \mathbb{R}^{n-2}$ . Hence we have  $\Theta(\rho) > 0$  since the effect of the boundary of the cut-off is negligible for large  $R$ . Indeed, we observe that

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

as  $R$  goes to infinity.

**Step 3 (Scaling)** Suppose that there is  $\rho$  with  $\Theta(\rho) > 0$ , and set  $\rho_\varepsilon(x) := \varepsilon^{-n} \rho(\varepsilon^{-1}x)$  for  $\varepsilon > 0$ . Then we have

$$\begin{aligned} \int_{\text{supp } \rho_\varepsilon} [D(-f)(x)] (\nabla(-\rho_\varepsilon)(x)) dx &= \varepsilon^{-(n+1)} \int_{\text{supp } \rho_\varepsilon} [D(-f)(x)] (\nabla(-\rho)(\varepsilon^{-1}x)) dx \\ &= \int_{\text{supp } \rho} [D(-f)(x)] (\nabla(-\rho)(x)) dx \end{aligned}$$

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 infinity as  $\varepsilon$  tends to zero. Thus we complete the proof of Theorem 1.1.

# Appendix: Skew convexity of distance functions on Finsler manifolds

We finally investigate the skew convexity of squared distance functions on Finsler manifolds. This would be of independent interest from the geometric viewpoint. The convexity of distance functions 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 [Sh] and [Oh2] for related work on the usual convexity and concavity along geodesics.

Let  $(M, F)$  be a  $C^\infty$ -Finsler manifold. We introduce some terminologies for which we refer to [BCS]. For a  $C^1$ -vector field  $X$  on  $M$  and tangent vectors  $v, w \in T_x M$  with  $w \neq 0$ , we define the *covariant derivative* of  $X$  by  $v$  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  $\Gamma_{jk}^i$  is the Christoffel symbol. If  $\Gamma_{jk}^i(w)$  depends only on the point  $x$  (i.e., independent of the choice of  $w \in T_x M \setminus \{0\}$ ) 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  $D_{\dot{\gamma}} \dot{\gamma} \equiv 0$ . We will use the following formula borrowed from [BCS, Exercise 10.1.2]:

$$\frac{d}{dt} g_V(V, W) = g_V(D_{\dot{\gamma}}^V V, W) + g_V(V, D_{\dot{\gamma}}^V W) \quad (\text{A.1})$$

for any  $C^1$ -curve  $\gamma$  and  $C^1$ -vector fields  $V, W$  along  $\gamma$  such that  $V \neq 0$ .

A  $C^\infty$ -vector field  $V$  along a geodesic  $\gamma : [0, l] \rightarrow M$  is called a *Jacobi field* if it satisfies the equation  $D_{\dot{\gamma}} \dot{D}_{\dot{\gamma}} V + \mathcal{R}(V, \dot{\gamma}) \dot{\gamma} \equiv 0$ , where  $\mathcal{R} : TM \otimes TM \rightarrow T^*M \otimes TM$  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  $v, w \in T_x M$ , the *flag curvature* is defined by

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

We remark that  $\mathcal{K}(v, w)$  depends not only on the plane in  $T_x M$  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

$$g_V(V, D_W^V D_W^V V - D_W^W D_W^W V) \geq -\delta F(V)^2 F(W)^2 \quad (\text{A.2})$$

for non-vanishing  $C^\infty$ -vector fields  $V, W$  and  $\delta \geq 0$ . This clearly holds with  $\delta = 0$  for Berwald spaces. Therefore  $\delta$  measures how the tangent spaces are distorted as one moves (in  $M$ ) along  $W$ . The *injectivity radius*  $\text{inj}(z)$  at  $z \in M$  is the supremum of  $R > 0$  such that any unit speed geodesic  $\gamma : [0, R] \rightarrow M$  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 A.1** *Let  $(M, F)$  be a forward complete Finsler manifold and suppose that  $\mathcal{K} \leq k$ ,  $\mathcal{S} \leq S$  and (A.2) hold for some  $k \geq 0$ ,  $S \geq 1$  and  $\delta \geq 0$ . 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  $R := \min\{\overleftarrow{\text{inj}}(z), \pi/\sqrt{kS^2 + \delta}\}$ . In particular, if  $\mathcal{K} \leq 0$ , then  $f$  is  $(\sqrt{\delta r} \cot(\sqrt{\delta r}))$ -skew convex in  $\overleftarrow{B}(z, r)$  for  $r \in (0, \min\{\overleftarrow{\text{inj}}(z), \pi/\sqrt{\delta}\})$  regardless  $\mathcal{S}$ .

*Proof.* Fix a unit speed minimal geodesic  $\gamma : [0, l] \rightarrow \overleftarrow{B}(z, r)$  with  $r < R$ , and let  $\sigma : [0, l] \times [0, 1] \rightarrow M$  be the  $C^\infty$ -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:

$$\frac{\partial}{\partial s} [g_{\mathcal{V}}(\mathcal{V}(s, 0), \mathcal{T}(s, 0))] = g_{\mathcal{V}}(\mathcal{V}(s, 0), D_s^{\mathcal{V}} \mathcal{T}(s, 0)). \quad (\text{A.3})$$

We used (A.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 (A.1) that

$$g_{\mathcal{V}}(\mathcal{V}(s, 0), D_s^{\mathcal{V}} \mathcal{T}(s, 0)) = \frac{1}{2} \frac{\partial}{\partial t} [g_{\mathcal{V}}(\mathcal{V}, \mathcal{V})](s, 0) = \frac{1}{2} \frac{\partial [F(\mathcal{V})^2]}{\partial t}(s, 0). \quad (\text{A.4})$$

Again due to (A.1), we observe

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

The second term is nonnegative by the Cauchy-Schwarz inequality. Moreover, by the assumption (A.2), we have

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

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

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

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

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



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

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

The above inequality shows that the function

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

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

$$\frac{1}{2} \frac{\partial[F(\mathcal{V})^2]}{\partial t} = F(\mathcal{V}) \frac{\partial[F(\mathcal{V})]}{\partial t} \leq -\sqrt{kS^2 + \delta}r \cdot \cot(\sqrt{kS^2 + \delta}r(1-t))F(\mathcal{V})^2. \quad (\text{A.5})$$

Combining (A.3), (A.4), (A.5) and  $F(\mathcal{V}(s, 0)) = F(\dot{\gamma}(s)) = 1$ , we conclude

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

This completes the proof. □

Interestingly enough, what appeared in Theorem A.1 is not the 2-uniform convexity constant  $\mathcal{C}$ , but the smoothness constant  $\mathcal{S}$ . Compare this with the usual convexity in [Oh2, Theorem 5.1]. We finally state the Berwald case separately.

**Corollary A.2** *Let  $(M, F)$  be forward complete and of Berwald type and suppose that  $\mathcal{K} \leq k$  and  $\mathcal{S} \leq S$  hold for some  $k \geq 0$  and  $S \geq 1$ . 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  $z \in M$ ,  $r \in (0, \min\{\overleftarrow{\text{inj}}(z), \pi/\sqrt{k}S\})$ . In particular, if  $\mathcal{K} \leq 0$ , then  $f$  is 1-skew convex in  $\overleftarrow{B}(z, \overleftarrow{\text{inj}}(z))$  regardless  $\mathcal{S}$ .*

This recovers the 1-skew convexity of  $f(x) = \| -x \|^2/2$  on Minkowski spaces in Section 4.

## 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.
- [Er] M. Erbar, The heat equation on manifolds as a gradient flow in the Wasserstein space, Ann. Inst. Henri Poincaré Probab. Stat. **46** (2010), 1–23.
- [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.
- [Ly] A. Lytchak, Open map theorem for metric spaces, *St. Petersburg Math. J.* **17** (2006), 477–491.
- [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.
- [OS1] S. Ohta and K.-T. Sturm, Heat flow on Finsler manifolds, *Comm. Pure Appl. Math.* **62** (2009), 1386–1433.
- [OS2] S. Ohta and K.-T. Sturm, Bochner-Weitzenböck formula and Li-Yau estimates on Finsler manifolds, Preprint (2011). Available at [arXiv:1104.5276](https://arxiv.org/abs/1104.5276)
- [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.