

How does the contraction property fail for convex functions on normed spaces?

Shin-ichi OHTA*

November 25, 2023

Abstract

On Euclidean and Hilbert spaces, Riemannian manifolds, and CAT(0)-spaces, gradient flows of convex functions are known to satisfy the contraction property, which plays a fundamental role in optimization theory and possesses fruitful analytic and geometric applications. On (non-inner product) normed spaces, however, gradient flows of convex functions do not satisfy the contraction property. We give a detailed proof of this characterization of inner products, and discuss a possible form of a weaker contraction property on normed spaces.

1 Introduction

Analysis of gradient flows of convex functions is a fundamental subject in many branches of mathematics, such as optimization theory, operator theory, partial differential equations, geometric analysis, to name a few. Beyond the classical setting of Euclidean and Hilbert spaces (see, e.g., [Br]), there has been a growing interest in gradient flows of convex functions on metric spaces. (The convexity is then understood as the *geodesic convexity* meaning that the restriction of a function to every minimizing geodesic is convex.) Such a generalization was intensively studied on CAT(0)-spaces (nonpositively curved metric spaces in the sense of triangle comparison) in 1990s [Jo1, Jo2, Ma]; we refer to [AGS, Ba1, Ba2] for further reading and more recent developments. Besides CAT(0)-spaces, there are also a number of related studies on CAT(1)-spaces [OP], *Alexandrov spaces of curvature bounded below* as well as *Wasserstein spaces* over them [Ly, Oh1, Pe, Sa], and metric measure spaces satisfying the *Riemannian curvature-dimension condition* $\text{RCD}(K, \infty)$ [St].

On a metric space (X, d) , a gradient curve ξ of a function f is tending to the direction in which f decreases most efficiently (measured by the distance structure d), which is a solution to the equation $\dot{\xi}(t) = \nabla[-f](\xi(t))$ in the smooth setting. One of the most important properties of gradient curves ξ, ζ of a convex function is the *contraction* (or the *non-expansion*) property:

$$d(\xi(t), \zeta(t)) \leq d(\xi(0), \zeta(0)) \quad \text{for all } t \geq 0. \quad (1.1)$$

*Department of Mathematics, Osaka University, Osaka 560-0043, Japan (s.ohata@math.sci.osaka-u.ac.jp); RIKEN Center for Advanced Intelligence Project (AIP), 1-4-1 Nihonbashi, Tokyo 103-0027, Japan

The contraction property plays a vital role from both theoretical and applied viewpoints. It implies the continuous dependence (especially the uniqueness) of gradient curves on their initial points, and gradient curves provide a contraction semigroup (*gradient flow*) which can be analyzed by means of operator theory, partial differential equations, etc.

In the spaces mentioned in the first paragraph, appropriately defined gradient curves of convex functions satisfy the contraction property (1.1). We remark and stress that those spaces are all “Riemannian” in the sense that they exclude all non-Riemannian Finsler manifolds (and all non-inner product normed spaces). In fact, on a normed space $(\mathbb{R}^n, \|\cdot\|)$, we will show that the contraction property (1.1) holds for every convex function if and only if the norm $\|\cdot\|$ comes from an inner product (Theorem 3.1). We remark that, in [OS2, §4], we have already seen an example of a convex function on a normed space whose gradient flow does not have the contraction property. We shall prove that, by testing only for the maximum of linear functions, the contraction property fails for all non-inner product normed spaces.

The lack of the contraction property is one of the main difficulties of analysis of gradient flows in “non-Riemannian” spaces. There is no known quantitative estimate, and even the uniqueness of gradient curves is unclear (see [AGS, p. 4] and [Oh2, Preface(b)] for related remarks). The main aim of this short article is to draw more attention to this fundamental open problem and discuss a possible weaker contraction property available for normed spaces and, more generally, Finsler manifolds (see Remark 3.3). Such a generalized contraction property will be obviously beneficial in optimization theory and Banach space theory, as well as geometric analysis on Finsler manifolds, since heat flow on (measured) Finsler manifolds can be regarded as gradient flow of the relative entropy on the L^2 -Wasserstein space (we refer to [OS1, OZ] for more details). Moreover, solutions of some evolution equations can be regarded as gradient flow in the L^p -Wasserstein space with $p \neq 2$, which is non-Riemannian even when the underlying space is Riemannian (see, for example, [Ke] on q -heat flow). We also refer to [LOZ, Oh3, OZ] and the references therein for further related works concerning gradient flows in some classes of Finsler manifolds.

This article is organized as follows. In Section 2, we review necessary notions and recall the proof of the contraction property in the Euclidean case. Section 3 is devoted to the proof of the failure of the contraction property for normed spaces (Theorem 3.1). We also discuss a possible form of a weaker contraction property for normed spaces inspired by Theorem 3.1.

2 Preliminaries

The canonical coordinates of \mathbb{R}^n are denoted by $(x^i)_{i=1}^n$. We will freely identify points and vectors in \mathbb{R}^n in a canonical way.

2.1 Gradient curves in normed spaces

By a normed space $(\mathbb{R}^n, \|\cdot\|)$, we will mean a possibly asymmetric one; precisely,

- (1) $\|x\| \geq 0$ for all $x \in \mathbb{R}^n$ and $\|x\| = 0$ if and only if $x = 0$;
- (2) $\|cx\| = c\|x\|$ for all $x \in \mathbb{R}^n$ and $c > 0$;

(3) $\|x + y\| \leq \|x\| + \|y\|$ for all $x, y \in \mathbb{R}^n$.

Then, the function $d(x, y) := \|y - x\|$ gives an asymmetric distance function on \mathbb{R}^n . In this generality, a common notion of a *gradient curve* for a locally Lipschitz function $f: \mathbb{R}^n \rightarrow \mathbb{R}$ is a solution $\xi: [0, T) \rightarrow \mathbb{R}^n$ to the *energy dissipation identity*

$$f(\xi(t)) = f(\xi(s)) - \frac{1}{2} \int_s^t |\dot{\xi}|^2(r) \, dr - \frac{1}{2} \int_s^t |\partial f|^2(\xi(r)) \, dr \quad (2.1)$$

for all $0 < s < t < T$, where

$$|\dot{\xi}|(r) := \limsup_{\rho \rightarrow r} \frac{\|\xi(\max\{r, \rho\}) - \xi(\min\{r, \rho\})\|}{|r - \rho|}$$

is the (*forward*) *metric derivative* of ξ and

$$|\partial f|(x) := \limsup_{y \rightarrow x} \frac{\max\{f(x) - f(y), 0\}}{\|y - x\|}$$

is the *local (descending) slope* of f . Since “ \geq ” always holds in (2.1) by the Cauchy–Schwarz and Young inequalities, satisfying “ \leq ” is the essential condition and implies “ $\dot{\xi}(r) = \nabla[-f](\xi(r))$ ” at almost every r . We refer to [AG, AGS] for further reading, and to [OZ] and the references therein for the asymmetric case.

If $\|\cdot\|$ is strictly convex (in the sense that $\|x + y\| = \|x\| + \|y\|$ holds only when $x = ay$ or $y = ax$ for some $a \geq 0$) and f is differentiable, then we can introduce a gradient curve simply as a solution to

$$\dot{\xi}(t) = \nabla[-f](\xi(t)),$$

where $\nabla[-f](x) := \mathcal{L}(-df(x))$ is the *gradient vector* defined via the *Legendre transform* $\mathcal{L}: \mathbb{R}^n \rightarrow \mathbb{R}^n$ associated with the norm $\|\cdot\|$ given by

$$\mathcal{L}(\alpha) := \left(\frac{1}{2} \frac{\partial [\|\cdot\|_*^2]}{\partial \alpha_i}(\alpha) \right)_{i=1}^n,$$

where $\|\cdot\|_*$ is the dual norm on \mathbb{R}^n , namely

$$\|\alpha\|_* := \sup \left\{ \sum_{i=1}^n \alpha_i x^i \mid \|x\| \leq 1 \right\}.$$

(Note that $\|\cdot\|_*^2$ is differentiable since $\|\cdot\|$ is strictly convex.) We remark that $\nabla[-f] \neq -\nabla f$ in general due to the asymmetry of $\|\cdot\|$. In the case where $\|\cdot\|_*$ is twice differentiable on $\mathbb{R}^n \setminus \{0\}$ and $df(x) \neq 0$, the gradient vector is written (with the help of Euler’s homogeneous function theorem) as

$$\nabla[-f](x) = \left(- \sum_{j=1}^n \frac{1}{2} \frac{\partial^2 [\|\cdot\|_*^2]}{\partial \alpha_i \partial \alpha_j} (-df(x)) \frac{\partial f}{\partial x^j}(x) \right)_{i=1}^n. \quad (2.2)$$

We remark that $\|\cdot\|_*^2$ is twice differentiable at the origin 0 only when $\|\cdot\|$ happens to be induced from an inner product, thereby we excluded 0 (see, for example, [Oh2, Proposition 1.7]).

In (2.2),

$$g_{ij}^*(\alpha) := \frac{1}{2} \frac{\partial^2 [\|\cdot\|_*^2]}{\partial \alpha_i \partial \alpha_j}(\alpha) \quad (2.3)$$

(with $\alpha \neq 0$) is the inverse matrix of

$$g_{ij}(\mathcal{L}(\alpha)) := \frac{1}{2} \frac{\partial^2 [\|\cdot\|^2]}{\partial x^i \partial x^j}(\mathcal{L}(\alpha)), \quad (2.4)$$

provided that $(g_{ij}(\mathcal{L}(\alpha)))_{i,j=1}^n$ is positive-definite. By definition, for each $v \in \mathbb{R}^n \setminus \{0\}$, the symmetric matrix $(g_{ij}(v))_{i,j=1}^n$ gives an inner product g_v approximating $\|\cdot\|$ in the direction v up to second order (see [Oh2, §3.1] for a related account on Finsler manifolds).

2.2 Contraction in Euclidean spaces revisited

Before discussing the non-contraction in normed spaces, let us recall how to show the contraction in the Euclidean setting. We denote by $|\cdot|$ and $\langle \cdot, \cdot \rangle$ the standard Euclidean norm and inner product, respectively.

Let $f: \mathbb{R}^n \rightarrow \mathbb{R}$ be a convex C^1 -function and $\xi, \zeta: [0, T] \rightarrow \mathbb{R}^n$ be gradient curves of f . For $t > 0$, we have

$$\begin{aligned} \frac{d}{dt} [|\zeta(t) - \xi(t)|^2] &= 2\langle \zeta(t) - \xi(t), \dot{\zeta}(t) - \dot{\xi}(t) \rangle \\ &= -2\langle \zeta(t) - \xi(t), \nabla f(\zeta(t)) - \nabla f(\xi(t)) \rangle. \end{aligned} \quad (2.5)$$

Put $\gamma(s) := (1-s)\xi(t) + s\zeta(t)$ for $s \in [0, 1]$. Then the convexity of f implies that

$$f(\gamma(s)) \leq (1-s)f(\xi(t)) + sf(\zeta(t)),$$

from which we find

$$\frac{f(\gamma(s)) - f(\xi(t))}{s} \leq \frac{f(\zeta(t)) - f(\gamma(s))}{1-s}.$$

By letting $s \rightarrow 0$ and $s \rightarrow 1$, since $\dot{\gamma}(0) = \dot{\gamma}(1) = \zeta(t) - \xi(t)$, we obtain

$$\langle \nabla f(\xi(t)), \zeta(t) - \xi(t) \rangle \leq f(\zeta(t)) - f(\xi(t)) \leq \langle \nabla f(\zeta(t)), \zeta(t) - \xi(t) \rangle \quad (2.6)$$

(that is, ∇f is monotone). Hence, combining this with (2.5), we have the contraction

$$\frac{d}{dt} [|\zeta(t) - \xi(t)|^2] \leq 0.$$

Moreover, for a K -convex function f with $K \in \mathbb{R}$ in the sense that

$$f((1-s)x + sy) \leq (1-s)f(x) + sf(y) - \frac{K}{2}(1-s)s|y-x|^2$$

for all $x, y \in \mathbb{R}^n$ and $s \in (0, 1)$, we have the K -monotonicity

$$\langle \nabla f(\xi(t)), \zeta(t) - \xi(t) \rangle - \langle \nabla f(\zeta(t)), \zeta(t) - \xi(t) \rangle \leq -K|\zeta(t) - \xi(t)|^2$$

as well as the K -contraction property

$$|\zeta(t) - \xi(t)| \leq e^{-Kt} |\zeta(0) - \xi(0)|. \quad (2.7)$$

In the case of a normed space $(\mathbb{R}^n, \|\cdot\|)$, on the one hand, we have

$$\begin{aligned} \frac{d}{dt} [\|\zeta(t) - \xi(t)\|^2] &= \sum_{j=1}^n \frac{\partial [\|\cdot\|^2]}{\partial x^j} (\zeta(t) - \xi(t)) \cdot (\dot{\zeta}_j(t) - \dot{\xi}_j(t)) \\ &= \sum_{i,j=1}^n \frac{\partial^2 [\|\cdot\|^2]}{\partial x^i \partial x^j} (\zeta(t) - \xi(t)) \cdot (\zeta_i(t) - \xi_i(t)) (\dot{\zeta}_j(t) - \dot{\xi}_j(t)) \end{aligned}$$

in place of (2.5), where we assumed that $\|\cdot\|$ is C^2 on $\mathbb{R}^n \setminus \{0\}$ as well as $\xi(t) \neq \zeta(t)$, and used the homogeneous function theorem in the second line. Substituting $\dot{\xi}(t) = \nabla[-f](\xi(t))$ yields that, with the help of (2.2), (2.3) and (2.4),

$$\begin{aligned} \frac{d}{dt} [\|\zeta(t) - \xi(t)\|^2] &= 2 \sum_{i,j,k=1}^n g_{ij} (\zeta(t) - \xi(t)) \cdot (\zeta_i(t) - \xi_i(t)) \\ &\quad \times \left(g_{jk}^* (-df(\xi(t))) \frac{\partial f}{\partial x^k} (\xi(t)) - g_{jk}^* (-df(\zeta(t))) \frac{\partial f}{\partial x^k} (\zeta(t)) \right). \quad (2.8) \end{aligned}$$

On the other hand, for $\gamma(s) = (1-s)\xi(t) + s\zeta(t)$ as above, we observe

$$\lim_{s \rightarrow 0} \frac{f(\gamma(s)) - f(\xi(t))}{s} = \sum_{i=1}^n \frac{\partial f}{\partial x^i} (\xi(t)) \cdot (\zeta_i(t) - \xi_i(t)),$$

thereby the convexity of f implies

$$\sum_{i=1}^n \left(\frac{\partial f}{\partial x^i} (\xi(t)) - \frac{\partial f}{\partial x^i} (\zeta(t)) \right) \cdot (\zeta_i(t) - \xi_i(t)) \leq 0 \quad (2.9)$$

in place of (2.6). Therefore, if

$$\left(g_{ij}^* (-df(\xi(t))) \right) = \left(g_{ij}^* (-df(\zeta(t))) \right) = \left(g_{ij} (\zeta(t) - \xi(t)) \right)^{-1} \quad (2.10)$$

holds as $n \times n$ matrices, then we deduce from (2.8) and (2.9) the contraction

$$\frac{d}{dt} [\|\zeta(t) - \xi(t)\|^2] \leq 0.$$

When the norm $\|\cdot\|$ comes from an inner product, g_v (defined by $g_{ij}(v)$) coincides with the original inner product for any $v \neq 0$ and (2.10) always holds. On a normed space, however, g_v depends on the direction v and (2.10) does not hold unless $\zeta(t) - \xi(t) = a\nabla[-f](\xi(t)) = b\nabla[-f](\zeta(t))$ for some $a, b > 0$.

In view of (2.10), the difficulty in the case of normed spaces stems from the difference between the directions in question; $\zeta(t) - \xi(t)$ from the first variation of the distance and $\nabla[-f](\xi(t))$ (or $\nabla[-f](\zeta(t))$) from the gradient flow equation. When (2.10) holds in some way, we can connect them and apply the convexity (2.9) of f . We refer to [OP, OS2] for related studies. In [OP] we introduced a condition called the *commutativity* for metric spaces, which corresponds to (2.10) and turned out a Riemannian condition. In [OS2] we investigated a sufficient condition for the contraction, called the *skew convexity*.

3 Non-contraction in normed spaces

Let $(\mathbb{R}^n, \|\cdot\|)$ be a possibly asymmetric normed space in the sense of Subsection 2.1. We will denote by $\mathbf{S} := \{x \in \mathbb{R}^n \mid \|x\| = 1\}$ the unit sphere of the norm.

We remark that it is natural to assume the strict convexity of $\|\cdot\|$. In fact, without strict convexity, gradient curves are not unique and there is no chance to have the contraction property (see [AG, Example 4.23], [OS2, Remark 4.1]).

Theorem 3.1 (Non-contraction in normed spaces). *Suppose that a (possibly asymmetric) strictly convex normed space $(\mathbb{R}^n, \|\cdot\|)$ satisfies the contraction property in the sense that*

$$\|\zeta(t) - \xi(t)\| \leq \|\zeta(0) - \xi(0)\| \quad (3.1)$$

holds for all $t \in [0, T)$ for every convex function $f: \mathbb{R}^n \rightarrow \mathbb{R}$ and any gradient curves $\xi, \zeta: [0, T) \rightarrow \mathbb{R}^n$ for f . Then, the norm $\|\cdot\|$ is necessarily induced from an inner product.

We will see that, in fact, the contraction is tested only for the maximum of two linear functions in our proof.

Proof. We divide the proof into five steps.

Step 1. We first show that $\|\cdot\|$ is differentiable on $\mathbb{R}^n \setminus \{0\}$. For a convex function, the directional differentiability in all directions implies the differentiability (by [Ro, Theorem 25.1]). Thus, assuming the contrary, let $\|\cdot\|$ be not directionally differentiable at $v \in \mathbf{S}$. Then, the closure of the set

$$K := \{x \in \mathbb{R}^n \mid v + \lambda(x - v) \in \mathbf{S} \text{ for some } \lambda > 0\}$$

is not a halfspace (∂K is not a hyperplane) but a closed convex cone with origin at v . Hence, we can modify the coordinates in such a way that $v = (1, 0, \dots, 0) \in \mathbf{S}$ and $\|x\| > 1$ holds for some $\delta > 0$ and all

$$x = (1 + as, s, x^3, \dots, x^n)$$

with $a \in (-\delta, \delta)$ and $(s, x^3, \dots, x^n) \in \mathbb{R}^{n-1} \setminus \{0\}$ (to be precise, the strict inequality $\|x\| > 1$ for all $(s, x^3, \dots, x^n) \in \mathbb{R}^{n-1} \setminus \{0\}$ is guaranteed by the strict convexity).

We consider the convex function

$$f(x) = \max\{-x^1 + \varepsilon x^2, c(-x^1 - \varepsilon x^2)\}$$

for $\varepsilon \in (0, \delta)$ and $c \in (0, 1)$. By the choice of δ , we have

$$\nabla[-f](x) = \begin{cases} (1, 0, \dots, 0) & \text{when } (1 - c)x^1 < (1 + c)\varepsilon x^2, \\ (c, 0, \dots, 0) & \text{when } (1 - c)x^1 > (1 + c)\varepsilon x^2. \end{cases}$$

Put

$$b_c := \frac{1 - c}{1 + c} > 0$$

and observe

$$(1 + c)\varepsilon b_c = 2(1 - c) > (1 - c).$$

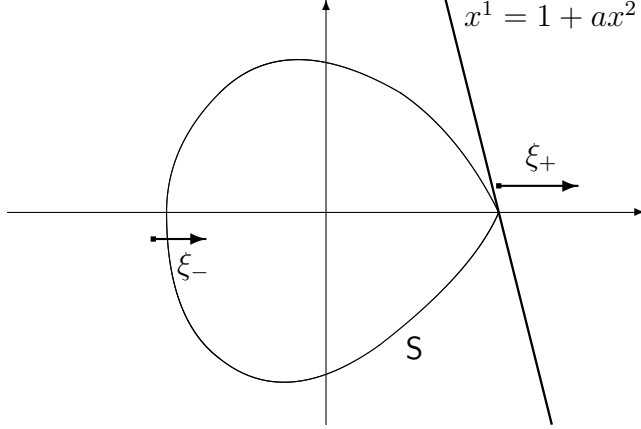


Figure 1: Step 1 for $n = 2$

This implies that

$$\nabla[-f]((1, b_c, 0, \dots, 0)) = (1, 0, \dots, 0), \quad \nabla[-f]((-1, -b_c, 0, \dots, 0)) = (c, 0, \dots, 0).$$

Hence, the gradient curves emanating from $(1, b_c)$ and $(-1, -b_c)$ are given by

$$\begin{aligned} \xi_+(t) &= (1 + t, b_c, 0, \dots, 0) \quad \text{for } t \in [0, 1), \\ \xi_-(t) &= (-1 + ct, -b_c, 0, \dots, 0) \quad \text{for } t \geq 0, \end{aligned}$$

respectively (see Figure 1). Thus, it follows from the assumed contraction property (3.1) that

$$\|(2 + (1 - c)t, 2b_c, 0, \dots, 0)\| \leq \|(2, 2b_c, 0, \dots, 0)\|$$

for $t \in [0, 1)$. However, since $b_c \rightarrow 0$ as $c \rightarrow 1$, we have

$$\|(2 + (1 - c)t, 2b_c, 0, \dots, 0)\| > \|(2, 2b_c, 0, \dots, 0)\|$$

when c is close to 1, a contradiction.

From here on, we assume that $\|\cdot\|$ is differentiable on $\mathbb{R}^n \setminus \{0\}$. Note that $\|\cdot\|$ is then C^1 on $\mathbb{R}^n \setminus \{0\}$ (see [Ro, Corollary 25.5.1]).

Step 2. Next, to prove the symmetry of $\|\cdot\|$, we assume in contrary that the function

$$\phi: \mathbf{S} \longrightarrow (0, \infty), \quad \text{where } -\phi(v) \cdot v \in \mathbf{S},$$

is not constant (and hence its derivative is not identically zero). Thus, by choosing appropriate coordinates, the hyperplane $x^2 = 1$ is tangent to \mathbf{S} at $v := (0, 1, 0, \dots, 0)$ and the hyperplane $x^2 = b$ is tangent to \mathbf{S} at $w := (-a, -b, 0, \dots, 0)$ for some $a, b > 0$. Now, given $c > 0$, the function

$$f(x) = \max\{cb^{-1}x^2, -bx^2\}$$

is convex and we have

$$\nabla[-f](x) = \begin{cases} cw & \text{when } x^2 > 0, \\ bv & \text{when } x^2 < 0. \end{cases}$$

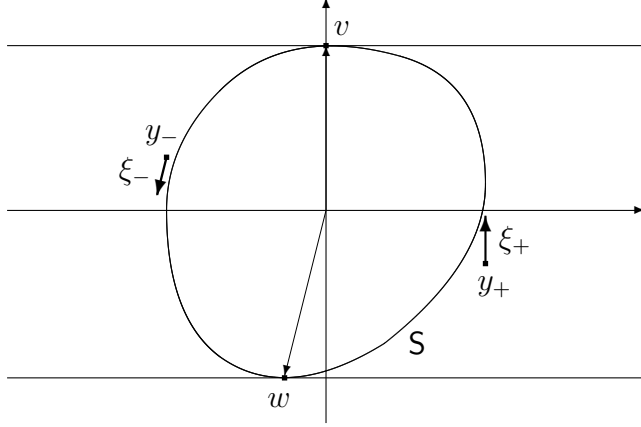


Figure 2: Step 2 for $n = 2$ (S is C^1)

Hence, for $y_{\pm} := (\pm 1, \mp s, 0, \dots, 0)$ with small $s > 0$,

$$\xi_+(t) := y_+ + tbv, \quad \xi_-(t) := y_- + tcw$$

are gradient curves emanating from y_+, y_- , respectively, for small $t \geq 0$ (see Figure 2). Thus, the hypothesized contraction property (3.1) implies that

$$\frac{d^+}{dt} \left[\|\xi_+(t) - \xi_-(t)\| \right]_{t=0} = d[\|\cdot\|]_{y_+ - y_-}(bv - cw) \leq 0.$$

Letting $s \rightarrow 0$, since $\|\cdot\|$ is C^1 on $\mathbb{R}^n \setminus \{0\}$, we obtain

$$d[\|\cdot\|]_{(2,0,\dots,0)}(bv - cw) \leq 0.$$

Moreover, by the contraction property between

$$\zeta_+(t) := (1, s, 0, \dots, 0) + tcw, \quad \zeta_-(t) := (-1, -s, 0, \dots, 0) + tbv,$$

we also find

$$d[\|\cdot\|]_{(2,0,\dots,0)}(cw - bv) \leq 0.$$

Therefore,

$$d[\|\cdot\|]_{(2,0,\dots,0)}(bv - cw) = 0$$

holds for any $c > 0$. This means that S is tangent to $(\lambda, 0, \dots, 0) + \mathbb{R} \cdot (bv - cw)$ for $\lambda > 0$ with $(\lambda, 0, \dots, 0) \in S$ and arbitrary $c > 0$. However, this is possible only when v and w are linearly dependent, so that $a = 0$. This is a contradiction and completes the proof of the symmetry.

Step 3. Given arbitrary $v \in S$, let us modify the coordinates in such a way that $v = (1, 0, \dots, 0)$ and the hyperplane $x^1 = 1$ is tangent to S at v . Then we consider the convex function

$$f(x) = |x^1|$$

and shall show that lines parallel to v and passing through $S \cap f^{-1}(0)$ are tangent to S (this is indeed guessed from the argument in the previous step).

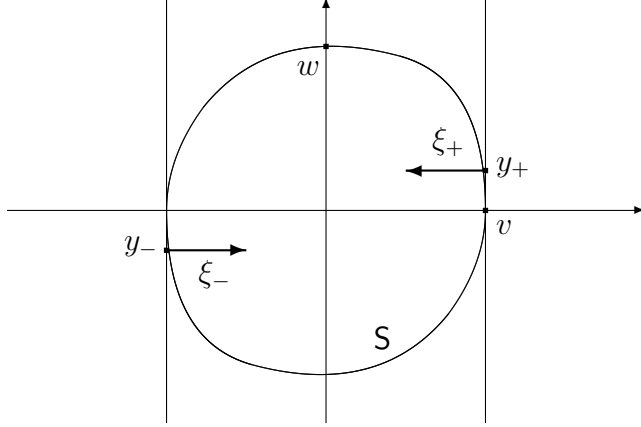


Figure 3: Step 3 for $n = 2$ (S is symmetric)

To this end, take $w \in S \cap f^{-1}(0)$ and put $y_+ := v + sw$ and $y_- := -v - sw$ for $s > 0$. Note that $f(y_{\pm}) = f(\pm v) = 1$ and $\nabla[-f](y_{\pm}) = \nabla[-f](\pm v) = \mp v$, and hence the gradient curves for f emanating from y_{\pm} are given by

$$\xi_+(t) := y_+ - tv, \quad \xi_-(t) := y_- + tv$$

for $t \in [0, 1)$ (see Figure 3). Then we observe from

$$\|\xi_+(t) - \xi_-(t)\| = \|y_+ - y_- - 2tv\| = 2\|v + sw - tv\|$$

that

$$\frac{1}{2} \frac{d}{dt} [\|\xi_+(t) - \xi_-(t)\|]_{t=0} = -d[\|\cdot\|]_{v+sw}(v) = -d[\|\cdot\|]_{(v/s)+w}(v) \rightarrow -d[\|\cdot\|]_w(v)$$

as $s \rightarrow \infty$, where the latter equality is a consequence of the 1-homogeneity of $\|\cdot\|$. Hence, the contraction property (3.1) implies

$$d[\|\cdot\|]_w(v) \geq 0.$$

Moreover, replacing w with $-w$ gives the reverse inequality thanks to the symmetry of $\|\cdot\|$, indeed,

$$d[\|\cdot\|]_{-w}(v) = d[\|\cdot\|]_w(-v) = -d[\|\cdot\|]_w(v).$$

Therefore, we obtain

$$d[\|\cdot\|]_w(v) = 0.$$

This yields that the line $w + \mathbb{R} \cdot v$ is tangent to S at w as we claimed.

Step 4. By virtue of the previous step, we can choose coordinates such that the hyperplanes $x^1 = 1$ and $x^2 = 1$ are tangent to S at $v = (1, 0, \dots, 0)$ and $w = (0, 1, 0, \dots, 0)$, respectively. Now, for $a > 0$, we consider the convex function

$$f(x) = \max\{ax^1, x^2\}.$$

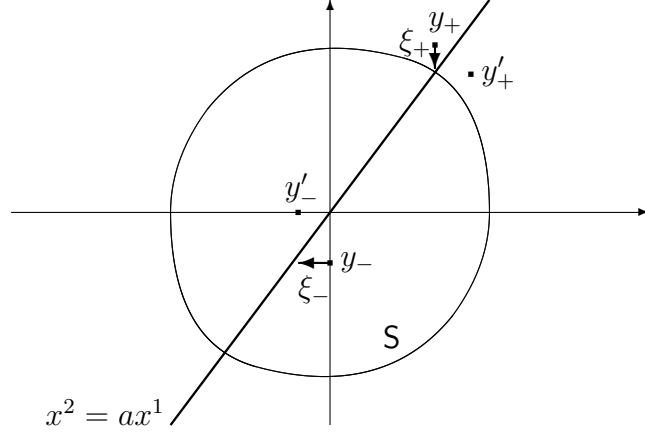


Figure 4: Step 4 for $n = 2$

Note that, by our choice of the coordinates,

$$\nabla[-f](x) = \begin{cases} (-a, 0, \dots, 0) & \text{when } ax_1 > x_2, \\ (0, -1, 0, \dots, 0) & \text{when } ax_1 < x_2. \end{cases}$$

We take $(r, ar, 0, \dots, 0) \in \mathbf{S}$ with $r = r(a) > 0$ and put $y_- := (0, -as, 0, \dots, 0)$, $y_+ := (r, ar + s, 0, \dots, 0)$ for small $s > 0$. Observe that $\nabla[-f](y_-) = (-a, 0, \dots, 0)$, $\nabla[-f](y_+) = (0, -1, 0, \dots, 0)$, and

$$\begin{aligned} \xi_-(t) &:= y_- - (at, 0) = (-at, -as, 0, \dots, 0), \\ \xi_+(t) &:= y_+ - (0, t) = (r, ar + s - t, 0, \dots, 0) \end{aligned}$$

are gradient curves for f for $0 \leq t < \min\{s, s/a\}$ (see Figure 4). Then we have

$$\begin{aligned} \frac{d}{dt} [\|\xi_+(t) - \xi_-(t)\|]_{t=0} &= d[\|\cdot\|]_{(r, ar + (a+1)s)}(a, -1, 0, \dots, 0) \\ &\rightarrow d[\|\cdot\|]_{(r, ar)}(a, -1, 0, \dots, 0) \end{aligned}$$

as $s \rightarrow 0$. Hence, by the contraction property (3.1), we find

$$d[\|\cdot\|]_{(r, ar)}(a, -1, 0, \dots, 0) \leq 0.$$

A similar discussion for $y'_- := (-s, 0)$ and $y'_+ := (r + as, ar)$ yields that, since $\nabla[-f](y'_-) = (0, -1, 0, \dots, 0)$ and $\nabla[-f](y'_+) = (-a, 0, \dots, 0)$,

$$d[\|\cdot\|]_{(r, ar)}(-a, 1, 0, \dots, 0) \leq 0.$$

Therefore, we obtain

$$d[\|\cdot\|]_{(r, ar)}(a, -1, 0, \dots, 0) = 0, \tag{3.2}$$

and the line $(r, ar, 0, \dots, 0) + \mathbb{R} \cdot (a, -1, 0, \dots, 0)$ is tangent to \mathbf{S} at $(r, ar, 0, \dots, 0)$, for any $a > 0$ and $r = r(a) > 0$. One can similarly show (3.2) for $a < 0$ and $r = r(a) > 0$ with $(r, ar, 0, \dots, 0) \in \mathbf{S}$.

Step 5. The outcome of the previous step is that the parallelogram identity holds in the 2-plane $P\langle v, w \rangle$ including v and w . By the way of taking v, w in Step 3, for any $\bar{v}, \bar{w} \in \mathbb{R}^n$, we can choose v, w such that $\bar{v}, \bar{w} \in P\langle v, w \rangle$. Therefore, the parallelogram identity holds for any $\bar{v}, \bar{w} \in \mathbb{R}^n$, and hence $\|\cdot\|$ comes from an inner product. This completes the proof. \square

Comparing Theorem 3.1 with the scaling invariance of the convexity of functions on \mathbb{R}^n , we can derive some immediate corollaries. For this purpose, for a function $\Phi: (0, \infty) \rightarrow (0, \infty)$, we say that a strictly convex normed space $(\mathbb{R}^n, \|\cdot\|)$ satisfies the Φ -contraction property if

$$\|\zeta(t) - \xi(t)\| \leq \Phi(t)\|\zeta(0) - \xi(0)\|$$

holds for all $t \in (0, T)$ for every convex function $f: \mathbb{R}^n \rightarrow \mathbb{R}$ and any gradient curves $\xi, \zeta: [0, T) \rightarrow \mathbb{R}^n$ for f .

Corollary 3.2. *Let $(\mathbb{R}^n, \|\cdot\|)$ be a (possibly asymmetric) strictly convex normed space, which is not an inner product space. If $(\mathbb{R}^n, \|\cdot\|)$ satisfies the Φ -contraction property for some $\Phi: (0, \infty) \rightarrow (0, \infty)$, then we have*

$$\|\zeta(t) - \xi(t)\| \leq \inf_{s>0} \Phi(s) \cdot \|\zeta(0) - \xi(0)\|$$

for all $t \in (0, T)$, for any convex function $f: \mathbb{R}^n \rightarrow \mathbb{R}$ and gradient curves $\xi, \zeta: [0, T) \rightarrow \mathbb{R}^n$ for f . In particular, $(\mathbb{R}^n, \|\cdot\|)$ does not satisfy the Φ -contraction property for any Φ with $\inf_{s>0} \Phi(s) \leq 1$.

Thus, the exponential contraction $\|\zeta(t) - \xi(t)\| \leq e^{-Kt}\|\zeta(0) - \xi(0)\|$ does not hold for any $K < 0$ (compare this with (2.7)).

Proof. For any $\varepsilon > 0$, take $s > 0$ such that $\Phi(s) \leq \inf \Phi + \varepsilon$. Then, given a convex function $f: \mathbb{R}^n \rightarrow \mathbb{R}$, gradient curves $\xi, \zeta: [0, T) \rightarrow \mathbb{R}^n$ for f , and $t \in (0, T)$, we consider the function cf with $c = t/s$. Note that cf is also convex and $\xi_c(t) := \xi(ct), \zeta_c(t) := \zeta(ct)$ are gradient curves for cf . Hence, it follows from the Φ -contraction property that

$$\begin{aligned} \|\zeta(t) - \xi(t)\| &= \|\zeta_c(c^{-1}t) - \xi_c(c^{-1}t)\| \leq \Phi(c^{-1}t)\|\zeta(0) - \xi(0)\| = \Phi(s)\|\zeta(0) - \xi(0)\| \\ &\leq (\inf \Phi + \varepsilon) \cdot \|\zeta(0) - \xi(0)\|. \end{aligned}$$

Since $\varepsilon > 0$ was arbitrary, we obtain the first assertion. The latter assertion is then a consequence of Theorem 3.1. \square

Remark 3.3. (a) In view of Corollary 3.2, as a weaker contraction property on normed spaces, what one could expect is a scaling invariant estimate of the form

$$\|\zeta(t) - \xi(t)\| \leq C\|\zeta(0) - \xi(0)\| \tag{3.3}$$

with some $C > 1$. Then, due to the ‘‘discontinuity’’ at $t = 0$, it seems difficult to apply the method based on a differential inequality (as in Subsection 2.2 in the Euclidean case).

- (b) Note that (3.3) suffices to show the uniqueness of gradient curves. Thus, since the uniqueness fails in non-strictly convex normed spaces, the constant C would depend on the strength of the convexity of the norm $\|\cdot\|$, such as the uniform convexity (see, for example, [Oh2, §1.2.1]).
- (c) Another direction is to consider K -convex functions for $K > 0$. Also in this case, we may need to have a closer look on the uniform convexity (and smoothness) of $\|\cdot\|$.

Acknowledgements. This work was supported in part by JSPS Grant-in-Aid for Scientific Research (KAKENHI) 19H01786, 22H04942.

References

- [AG] L. Ambrosio and N. Gigli, A user’s guide to optimal transport. In: Modelling and optimisation of flows on networks, 1–155. Lecture Notes in Math., **2062**. Springer, Heidelberg, 2013.
- [AGS] L. Ambrosio, N. Gigli and G. Savaré, Gradient flows in metric spaces and in the space of probability measures. Second edition. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, 2008.
- [Ba1] M. Bačák, Convex analysis and optimization in Hadamard spaces. De Gruyter Series in Nonlinear Analysis and Applications, **22**. De Gruyter, Berlin, 2014.
- [Ba2] M. Bačák, Old and new challenges in Hadamard spaces. Jpn. J. Math. **18** (2023), 115–168.
- [Br] H. Brézis, Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert (French). North-Holland Publishing Co., Amsterdam-London; American Elsevier Publishing Co., Inc., New York, 1973.
- [Jo1] J. Jost, Convex functionals and generalized harmonic maps into spaces of nonpositive curvature. Comment. Math. Helv. **70** (1995), 659–673.
- [Jo2] J. Jost, Nonlinear Dirichlet forms. In: New directions in Dirichlet forms, 1–47, AMS/IP Stud. Adv. Math., **8**, Amer. Math. Soc., Providence, RI, 1998.
- [Ke] M. Kell, q -Heat flow and the gradient flow of the Renyi entropy in the p -Wasserstein space. J. Funct. Anal. **271** (2016), 2045–2089.
- [LOZ] N. Lebedeva, S. Ohta and V. Zolotov, Self-contracted curves in spaces with weak lower curvature bound. Int. Math. Res. Not. IMRN **2021** (2021), 8623–8656.
- [Ly] A. Lytchak, Open map theorem for metric spaces. St. Petersburg Math. J. **17** (2006), 477–491.
- [Ma] U. F. Mayer, Gradient flows on nonpositively curved metric spaces and harmonic maps. Comm. Anal. Geom. **6** (1998), 199–253.

- [Oh1] S. Ohta, Gradient flows on Wasserstein spaces over compact Alexandrov spaces. *Amer. J. Math.* **131** (2009), 475–516.
- [Oh2] S. Ohta, Comparison Finsler geometry. Springer Monographs in Mathematics. Springer, Cham, 2021.
- [Oh3] S. Ohta, Discrete-time gradient flows in Gromov hyperbolic spaces. *Israel J. Math.* (to appear). Available at [arXiv:2205.03156](https://arxiv.org/abs/2205.03156)
- [OP] S. Ohta and M. Pálfia, Gradient flows and a Trotter–Kato formula of semi-convex functions on CAT(1)-spaces. *Amer. J. Math.* **139** (2017), 937–965.
- [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, Non-contraction of heat flow on Minkowski spaces. *Arch. Ration. Mech. Anal.* **204** (2012), 917–944.
- [OZ] S. Ohta and W. Zhao, Gradient flows in asymmetric metric spaces and applications. *Ann. Sc. Norm. Super. Pisa Cl. Sci.* (to appear). Available at [arXiv:2206.07591](https://arxiv.org/abs/2206.07591)
- [Pe] A. Petrunin, Semiconcave functions in Alexandrov’s geometry. In: *Surveys in differential geometry*. Vol. XI, 137–201, *Surv. Differ. Geom.*, **11**, Int. Press, Somerville, MA, 2007.
- [Ro] R. T. Rockafellar, *Convex Analysis*. Princeton University Press, Princeton, NJ, 1970.
- [Sa] G. Savaré, Gradient flows and diffusion semigroups in metric spaces under lower curvature bounds. *C. R. Math. Acad. Sci. Paris* **345** (2007), 151–154.
- [St] K.-T. Sturm, Gradient flows for semiconvex functions on metric measure spaces—existence, uniqueness, and Lipschitz continuity. *Proc. Amer. Math. Soc.* **146** (2018), 3985–3994.