

# GRADIENT FLOWS IN ASYMMETRIC METRIC SPACES AND APPLICATIONS

SHIN-ICHI OHTA AND WEI ZHAO

ABSTRACT. This paper is devoted to the investigation of gradient flows in asymmetric metric spaces (for example, irreversible Finsler manifolds and Minkowski normed spaces) by means of discrete approximation. We study basic properties of curves and upper gradients in asymmetric metric spaces, and establish the existence of a curve of maximal slope, which is regarded as a gradient curve in the non-smooth setting. Introducing a natural convexity assumption on the potential function, which is called the  $(p, \lambda)$ -convexity, we also obtain some regularizing effects on the asymptotic behavior of curves of maximal slope. Applications include several existence results for gradient flows in Finsler manifolds, doubly nonlinear differential evolution equations on infinite-dimensional Funk spaces, and heat flow on compact Finsler manifolds.

## 1. INTRODUCTION

The aim of this article is to develop the theory of gradient flows in *asymmetric* metric spaces (i.e., the symmetry  $d(x, y) = d(y, x)$  is not assumed; see Definition 2.1). Typical and important examples are gradient flows of geodesically convex functions on irreversible Finsler manifolds (or Minkowski normed spaces). The theory of gradient flows has been successfully developed in “Riemannian-like” spaces such as CAT(0)-spaces and RCD-spaces (see, e.g., [1, 2, 10, 11, 20, 24, 30, 31, 38, 43]); nonetheless, the lack of the Riemannian-like structure causes a significant difference and we know much less about gradient flows in “Finsler-like” spaces (see [31, Remark 3.2], [33] and Subsection 4.5 for more details). In this article, based on the recent work [15] on the geometry of asymmetric metric spaces, we investigate gradient flows in asymmetric metric spaces by generalizing the minimizing movement scheme as in [1]. Compared with the preceding studies [6, 36] on asymmetric metric spaces, we will be able to remove some conditions on the space  $(X, d)$  or the potential function  $\phi$  by a more careful analysis (see Remarks 2.26, 3.11 for details). Moreover, the notion of  $(p, \lambda)$ -convexity (discussed in Section 4) seems new and of independent interest even in the symmetric case.

Asymmetric metrics often occur in nature and can be represented as Finsler metrics; a prominent example is the Matsumoto metric describing the law of walking on a mountain slope under the action of gravity (see [19]). Randers metrics appearing as solutions to the Zermelo navigation problem (concerning a Riemannian manifold with “wind” blown on it) provide another important class of irreversible metrics (see [4]). A particular example of the latter metric is given as a “non-symmetrization” of the Klein metric on the  $n$ -dimensional Euclidean unit ball  $\mathbb{B}^n = \{x \in \mathbb{R}^n \mid \|x\| < 1\}$  ( $n \geq 2$ ), called the *Funk metric* (see, e.g., [40, Example 1.3.5]), defined as  $F : \mathbb{B}^n \times \mathbb{R}^n \rightarrow [0, \infty)$  by

$$F(x, v) = \frac{\sqrt{\|v\|^2 - (\|x\|^2\|v\|^2 - \langle x, v \rangle^2)} + \langle x, v \rangle}{1 - \|x\|^2}, \quad x \in \mathbb{B}^n, v \in T_x \mathbb{B}^n = \mathbb{R}^n, \quad (1.1)$$

where  $\|\cdot\|$  and  $\langle \cdot, \cdot \rangle$  denote the Euclidean norm and inner product, respectively. The associated distance function  $d_F$  is written as (see [40, Example 1.1.2])

$$d_F(x_1, x_2) = \log \left( \frac{\sqrt{\|x_1 - x_2\|^2 - (\|x_1\|^2\|x_2\|^2 - \langle x_1, x_2 \rangle^2)} - \langle x_1, x_2 - x_1 \rangle}{\sqrt{\|x_1 - x_2\|^2 - (\|x_1\|^2\|x_2\|^2 - \langle x_1, x_2 \rangle^2)} - \langle x_2, x_2 - x_1 \rangle} \right), \quad x_1, x_2 \in \mathbb{B}^n. \quad (1.2)$$

It is readily seen that  $d_F(x_1, x_2) \neq d_F(x_2, x_1)$  and, for  $\mathbf{0} = (0, \dots, 0)$ ,

$$\lim_{\|x\| \rightarrow 1} d_F(\mathbf{0}, x) = \infty, \quad \lim_{\|x\| \rightarrow 1} d_F(x, \mathbf{0}) = \log 2.$$

---

2020 *Mathematics Subject Classification*. Primary 49J27; Secondary 49J52, 58J60.

*Key words and phrases*. asymmetric metric space, gradient flow, convex function, Finsler manifold, heat flow, Wasserstein space.

The Funk metric  $(\mathbb{B}^n, d_F)$  will be one of the model asymmetric structures we have in mind. We remark that the symmetrization  $d(x_1, x_2) := \{d_F(x_1, x_2) + d_F(x_2, x_1)\}/2$  coincides with the Klein metric.

For functions on an asymmetric metric space  $(X, d)$ , we shall study the associated *curves of maximal slope* in  $X$  (see Subsection 2.5); this conception generalizes gradient curves in the smooth setting. In order to deal with the asymmetry, the *reversibility* of  $(X, d)$ , defined by

$$\lambda_d(X) := \sup_{x \neq y} \frac{d(x, y)}{d(y, x)},$$

will be instrumental. Clearly  $\lambda_d(X) \geq 1$ , and  $\lambda_d(X) = 1$  holds if and only if  $d$  is symmetric. The reversibility may be  $\infty$  for noncompact asymmetric metric spaces. Actually, for the model Funk metric (1.1), a direct calculation yields  $\lambda_{d_F}(\mathbb{B}^n) = \infty$  and  $\lambda_{d_F}(B_0^+(r)) \leq 2e^r - 1$ , where  $B_0^+(r)$  is the forward open ball of radius  $r$  centered at  $\mathbf{0}$ , i.e.,  $B_0^+(r) = \{x \in \mathbb{B}^n \mid d_F(\mathbf{0}, x) < r\}$ . The latter estimate suggests to consider a collection of pointed spaces  $(X, \star, d)$  whose reversibility satisfies  $\lambda_d(B_\star^+(r)) \leq \Theta(r)$  for some non-decreasing function  $\Theta : (0, \infty) \rightarrow [1, \infty)$ . Such spaces are called *forward metric spaces* (see Subsection 2.1) and intensively studied in [15]. For instance, the Gromov–Hausdorff topology and the theory of curvature-dimension condition developed by Lott, Sturm and Villani can be generalized to such spaces. Every forward complete Finsler manifold is a forward metric space (see [15] for details).

In the present article, by generalizing the theory of [1] to forward metric spaces, we are able to obtain some existence and regularity results of curves of maximal slope. Among others, we establish the existence of curves of maximal slope satisfying the energy identity (Theorem 3.30), and show some estimates on the behavior of the potential function and its upper gradient along curves of maximal slope (Theorem 4.17). In the latter result, we make use of the  $(p, \lambda)$ -convexity, defined in Definition 4.1 by

$$\phi(\gamma(t)) \leq (1-t)\phi(\gamma(0)) + t\phi(\gamma(1)) - \frac{\lambda}{p}t(1-t^{p-1})d^p(\gamma(0), \gamma(1))$$

(which slightly differs from the  $(\lambda, p)$ -convexity studied in [37]; see Remark 4.2 for a further account), plays a role. As an application, we have the following in the Finsler setting (see Corollaries 3.33, 4.19 for the precise statements).

**Theorem 1.1.** *Let  $(M, F)$  be a forward complete Finsler manifold and  $\phi \in C^1(M)$ . Then, for any  $x_0 \in M$ , there exists a  $C^1$ -curve  $\xi : [0, T) \rightarrow M$  solving the gradient flow equation*

$$\xi'(t) = \nabla(-\phi)(\xi(t)), \quad \xi(0) = x_0,$$

where  $\lim_{t \rightarrow T} d_F(x_0, \xi(t)) = \infty$  if  $T < \infty$ . If  $\phi$  is  $\lambda$ -geodesically convex for some  $\lambda > 0$ , then  $T = \infty$ ,  $\xi(t)$  converges to a unique minimizer  $\bar{x}$  of  $\phi$  and  $F(\nabla(-\phi)(\xi(t)))$  decreases to 0 as  $t \rightarrow \infty$ .

Another application is concerned with “infinite dimensional Finsler spaces” such as the unit ball in a Hilbert space endowed with the Funk metric (1.2) and a reflexive Banach space  $(\mathcal{B}, \|\cdot\|)$  equipped with a Randers-type metric  $d(x, y) = \|y - x\| + \omega(y - x)$  for some  $\omega \in \mathcal{B}^*$  with  $\|\omega\|_* < 1$ . For such spaces, we prove that curves of maximal slope satisfy a doubly nonlinear differential evolution equation or inclusion, which generalizes some results in [36] (see Subsection 3.6 for details).

Last but not least, we also investigate gradient flows in the Wasserstein space over a compact Finsler manifold. We establish the equivalence between weak solutions to the heat equation and trajectories of the gradient flow for the relative entropy (in the same spirit as [7, 32]), as well as the following existence result of weak solutions to the heat equation as curves of maximal slope, which gives a direct construction rather than the extension by the  $L^2$ -contraction as in [32] (see Subsection 4.4 for a further account).

**Theorem 1.2.** *Let  $(M, F)$  be a compact Finsler manifold endowed with a smooth positive measure  $\mathfrak{m}$ . For any function  $u \in L^2(M)$  bounded above, there exists a weak solution  $(u_t)_{t \geq 0}$  to the heat equation  $\partial_t u_t = \Delta_{\mathfrak{m}} u_t$  with  $u_0 = u$ .*

We stress that there remain important open problems for gradient flows in Finsler-like spaces, even in (symmetric) normed spaces; see Subsection 4.5. We hope that our work motivates further investigations.

This article is organized as follows. In Section 2, we introduce some necessary concepts to analyze curves of maximal slope in forward metric spaces. In Section 3, we take advantage of the Moreau–Yosida

approximation to prove the existence of curves of maximal slope under mild topological assumptions. Section 4 is devoted to the investigation of curves of maximal slope for  $(p, \lambda)$ -convex functions, including some further existence and regularity results and the study of heat flow.

**Acknowledgements.** The authors thank the anonymous referees for valuable comments. The first author was supported by JSPS Grant-in-Aid for Scientific Research (KAKENHI) 19H01786, 22H04942, and the second author was supported by Natural Science Foundation of Shanghai (Nos. 21ZR1418300, 19ZR1411700).

## 2. CURVES AND UPPER GRADIENTS IN ASYMMETRIC METRIC SPACES

**2.1. Forward metric spaces.** First we discuss the basic properties of asymmetric metric spaces. We refer to [6, 15, 21, 22, 36, 39] for related studies (partly with different names).

**Definition 2.1** (Asymmetric metric spaces). Let  $X$  be a nonempty set and  $d : X \times X \rightarrow [0, \infty)$  be a nonnegative function on  $X \times X$ . We call  $(X, d)$  an *asymmetric metric space* if  $d$  satisfies

- (1)  $d(x, y) \geq 0$  for all  $x, y \in X$  with equality if and only if  $x = y$ ;
- (2)  $d(x, z) \leq d(x, y) + d(y, z)$  for all  $x, y, z \in X$ .

The function  $d$  is called a *distance function* or a *metric* on  $X$ .

Since the function  $d$  could be asymmetric (i.e.,  $d(x, y) \neq d(y, x)$ ), there are two kinds of balls. For a point  $x \in X$  and  $r > 0$ , the *forward* and *backward balls* of radius  $r$  centered at  $x$  are defined as

$$B_x^+(r) := \{y \in X \mid d(x, y) < r\}, \quad B_x^-(r) := \{y \in X \mid d(y, x) < r\}.$$

Let  $\mathcal{T}_+$  (resp.  $\mathcal{T}_-$ ) denote the topology induced from forward balls (resp. backward balls). In order to study the relation between  $\mathcal{T}_+$  and  $\mathcal{T}_-$ , the following notion on the reversibility of  $d$  was introduced in [15].

**Definition 2.2** (Pointed forward  $\Theta$ -metric spaces). Let  $\Theta : (0, \infty) \rightarrow [1, \infty)$  be a (not necessarily continuous) non-decreasing function. A triple  $(X, \star, d)$  is called a *pointed forward  $\Theta$ -metric space* if  $(X, d)$  is an asymmetric metric space and  $\star$  is a point in  $X$  such that  $\lambda_d(B_\star^+(r)) \leq \Theta(r)$  for all  $r > 0$ , where

$$\lambda_d(B_\star^+(r)) := \inf \{ \lambda \geq 1 \mid d(x, y) \leq \lambda d(y, x) \text{ for any } x, y \in B_\star^+(r) \}.$$

If we can take a constant function  $\Theta \equiv \theta$  (i.e.,  $\lambda_d(X) \leq \theta$ ), then we call  $(X, d)$  a  *$\theta$ -metric space*.

Note that we have the bounded reversibility  $\lambda_d < \infty$  only on forward balls, thereby the reversibility of backward balls may be infinite (like the Funk metric). For pointed forward  $\Theta$ -metric spaces, the backward topology is weaker than the forward topology as follows (see [15, Theorem 2.6]).

**Theorem 2.3** (Properties of  $\mathcal{T}_+$ ). *Let  $(X, \star, d)$  be a pointed forward  $\Theta$ -metric space. Then,*

- (i)  $\mathcal{T}_- \subset \mathcal{T}_+$  holds and, in particular,  $d$  is continuous in  $\mathcal{T}_+ \times \mathcal{T}_+$  and  $(X, \mathcal{T}_+)$  is a Hausdorff space;
- (ii)  $\mathcal{T}_+$  coincides with the topology  $\widehat{\mathcal{T}}$  induced from the symmetrized distance function

$$\widehat{d}(x, y) := \frac{1}{2} \{d(x, y) + d(y, x)\}.$$

We remark that the topology  $\tau$  considered in [21, 22] coincides with  $\widehat{\mathcal{T}}$ . Precisely,  $\tau$  is induced from both forward and backward balls and associated with  $\widetilde{d}(x, y) := \max\{d(x, y), d(y, x)\}$ , then  $\widehat{d} \leq \widetilde{d} \leq 2\widehat{d}$  yields  $\tau = \widehat{\mathcal{T}}$ . Some more remarks on pointed forward  $\Theta$ -metric spaces are in order.

**Remark 2.4.** (a) A sequence  $(x_i)_{i \geq 1}$  in  $X$  converges to  $x$  with respect to  $\mathcal{T}_+$  if and only if

$$\lim_{i \rightarrow \infty} d(x, x_i) = 0,$$

which implies the convergence  $\lim_{i \rightarrow \infty} d(x_i, x) = 0$  in  $\mathcal{T}_-$  (thanks to  $\lambda_d(B_\star^+(r)) \leq \Theta(r) < \infty$ ). However, the converse does not necessarily hold true (when  $d(\star, x_i) \rightarrow \infty$ ).

- (b) If  $(X, \star, d)$  is a pointed forward  $\Theta$ -metric space, then, for every  $x \in X$ , the triple  $(X, x, d)$  is a pointed forward  $\overline{\Theta}$ -metric space for  $\overline{\Theta}(r) := \Theta(d(\star, x) + r)$ . Moreover, if  $\text{Diam}(X) := \sup_{x, y \in X} d(x, y) < \infty$ , then  $(X, d)$  is a  $\theta$ -metric space with  $\theta := \Theta(\text{Diam}(X))$ .

- (c) One can similarly introduce a *pointed backward  $\Theta$ -metric space*  $(X, \star, d)$  by  $\lambda_d(B_\star^-(r)) \leq \Theta(r)$  for  $r > 0$ . Note that a pointed backward  $\Theta$ -metric space may not be a pointed forward  $\Theta$ -metric space for any  $\Theta$ ; recall the Funk metric in the introduction. Since  $(X, \star, d)$  is a pointed backward  $\Theta$ -metric space if and only if  $(X, \star, \overleftarrow{d})$  is a pointed forward  $\Theta$ -metric space for the *reverse metric*  $\overleftarrow{d}(x, y) := d(y, x)$ , we will focus only on pointed forward  $\Theta$ -metric spaces.

**Notation 2.5** (Forward metric spaces). In the sequel, every pointed forward  $\Theta$ -metric space  $(X, \star, d)$  is endowed with the forward topology  $\mathcal{T}_+$ . Suppressing  $\star$  and  $\Theta$  for the sake of simplicity, we will write  $(X, d)$  and call it a *forward metric space*.

Forward metric spaces possess many fine properties. For example, one can define a generalized Gromov–Hausdorff topology to study the convergence of forward metric spaces, and Gromov’s precompactness theorem remains valid. Besides, optimal transport theory can be developed and the weak curvature-dimension condition in the sense of Lott–Sturm–Villani [18, 41, 42] is stable also in this setting. Furthermore, various geometric and functional inequalities (such as Brunn–Minkowski, Bishop–Gromov, log-Sobolev and Lichnerowicz inequalities) can be established. We refer to [15] for details and further results.

We next recall some concepts related to the completeness (cf. [15, 21, 22, 39]).

**Definition 2.6** (Completeness). Let  $(X, d)$  be an asymmetric metric space.

- (1) A sequence  $(x_i)_{i \geq 1}$  in  $X$  is called a *forward* (resp. *backward*) *Cauchy sequence* if, for each  $\varepsilon > 0$ , there is  $N \geq 1$  such that  $d(x_i, x_j) < \varepsilon$  (resp.  $d(x_j, x_i) < \varepsilon$ ) holds for all  $j \geq i > N$ .
- (2)  $(X, d)$  is said to be *forward* (resp. *backward*) *complete* if every forward (resp. backward) Cauchy sequence in  $X$  is convergent with respect to  $\mathcal{T}_+$ .
- (3) We say that  $(X, d)$  is *forward* (resp. *backward*) *boundedly compact* if every closed set in any forward (resp. backward) bounded ball is compact.

If  $\Theta$  is bounded (i.e.,  $\lambda_d(X) < \infty$ ), then the forward and backward properties are mutually equivalent. However, they are not equivalent when  $\Theta$  is unbounded; e.g., the Funk metric is forward complete but backward incomplete.

**2.2. Forward absolutely continuous curves.** Let  $(X, \star, d)$  be a forward complete pointed forward  $\Theta$ -metric space in this subsection.

**Definition 2.7** (Forward absolutely continuous curves). A curve  $\gamma : I \rightarrow X$  on an interval  $I \subset \mathbb{R}$  is said to be  *$p$ -forward absolutely continuous* for  $p \in [1, \infty]$  (denoted by  $\gamma \in \text{FAC}^p(I; X)$ ) if there is a nonnegative function  $f \in L^p(I)$  such that

$$d(\gamma(s), \gamma(t)) \leq \int_s^t f(r) dr \quad \text{for all } s, t \in I \text{ with } s \leq t. \quad (2.1)$$

We will denote  $\text{FAC}^1(I; X)$  by  $\text{FAC}(I; X)$  and call its element a *forward absolutely continuous curve*.

A standard argument combined with the forward completeness yields the following.

**Lemma 2.8.** *Any curve  $\gamma \in \text{FAC}^p([a, b]; X)$  is forward uniformly continuous, i.e., for any  $\varepsilon > 0$ , there exists  $\delta > 0$  such that  $d(\gamma(s), \gamma(t)) < \varepsilon$  holds for any  $s, t \in [a, b]$  with  $0 \leq t - s < \delta$ . In particular, if  $b < \infty$ , then the left limit  $\gamma(b_-) := \lim_{t \rightarrow b^-} \gamma(t)$  exists. Moreover, if  $\gamma \in \text{FAC}([a, b]; X)$ , then  $\gamma(b_-)$  exists even when  $b = \infty$ .*

We remark that, for  $\gamma \in \text{FAC}^p((a, b); X)$ , the right limit  $\gamma(a_+) := \lim_{t \rightarrow a^+} \gamma(t)$  may not exist. This is the reason why we call  $\gamma$  a *forward* absolutely continuous curve. For example, in the Funk space (1.1), consider the unit speed minimal geodesic  $\gamma : (-\log 2, 0] \rightarrow \mathbb{B}^n$  such that  $\gamma(0) = \mathbf{0}$  and  $\gamma(t)$  converges to  $(-1, 0, \dots, 0)$  in  $\mathbb{R}^n$  as  $t \rightarrow -\log 2$ . Clearly  $\gamma \in \text{FAC}((-\log 2, 0]; \mathbb{B}^n)$  with  $f \equiv 1$  in (2.1); however,  $\gamma$  is not defined at  $t = -\log 2$ . In fact,  $\lim_{t \rightarrow -\log 2} d_F(\mathbf{0}, \gamma(t)) = \infty$ .

Thanks to Lemma 2.8, we always have  $\text{FAC}^p([a, b]; X) = \text{FAC}^p([a, b]; X)$  for  $b < \infty$  and  $\text{FAC}([a, \infty); X) = \text{FAC}([a, \infty); X)$ . Hence, we will mainly consider  $\text{FAC}^p([a, b]; X)$  with  $a \in \mathbb{R}$ . Owing to [36, Proposition 2.2], there always exists a minimal function  $f$  satisfying (2.1) as follows (we also refer to [1, Theorem 1.1.2], [6, Theorem 3.5] and [32, Lemma 7.1]).

**Theorem 2.9** (Forward metric derivative). *Suppose that either  $p = 1$  with  $b \leq \infty$  or  $p \in (1, \infty]$  with  $b < \infty$ . Then, for any curve  $\gamma \in \text{FAC}^p([a, b]; X)$ , the limit*

$$|\gamma'_+|(t) := \lim_{s \rightarrow t} \frac{d(\gamma(\min\{s, t\}), \gamma(\max\{s, t\}))}{|t - s|}$$

exists for  $\mathcal{L}^1$ -a.e.  $t \in (a, b)$  and we have

$$d(\gamma(s), \gamma(t)) \leq \int_s^t |\gamma'_+|(r) \, dr \quad \text{for any } a \leq s \leq t \leq b.$$

Moreover,  $|\gamma'_+|$  belongs to  $L^p([a, b])$  and satisfies  $|\gamma'_+|(t) \leq f(t)$  for  $\mathcal{L}^1$ -a.e.  $t \in (a, b)$  for any  $f$  satisfying (2.1).

We call  $|\gamma'_+|$  the *forward metric derivative* of  $\gamma$ , and  $\mathcal{L}^1$  denotes the 1-dimensional Lebesgue measure. One can also define the *p-backward absolute continuity* by requiring (2.1) for  $s, t \in I$  with  $t \leq s$ . Then, due to the asymmetry, the resulting *backward metric derivative*  $|\gamma'_-|$  may not coincide with  $|\gamma'_+|$ .

**Definition 2.10** (Length). For  $\gamma \in \text{FAC}([a, b]; X)$ , its *length*  $L(\gamma)$  is defined by

$$L(\gamma) := \int_a^b |\gamma'_+|(t) \, dt.$$

Thanks to Theorem 2.9 and Lemma 2.8, it is not difficult to see that

$$L(\gamma) = \sup \left\{ \sum_{i=1}^N d(\gamma(t_{i-1}), \gamma(t_i)) \mid N \in \mathbb{N}, a = t_0 < \dots < t_N = b \right\}.$$

**Definition 2.11** (Lipschitz curves). A curve  $\gamma : [a, b] \rightarrow X$  is said to be *C-Lipschitz* for  $C > 0$  if it satisfies  $d(\gamma(s), \gamma(t)) \leq C(t - s)$  for all  $a \leq s \leq t \leq b$ .

Obviously, a *C-Lipschitz* curve belongs to  $\text{FAC}([a, b]; X)$  (provided  $b < \infty$ ) and  $|\gamma'_+|(t) \leq C$  for  $\mathcal{L}^1$ -a.e.  $t \in (a, b)$ . The next lemma (proved in the same manner as [1, Lemma 1.1.4]) tells that every forward absolutely continuous curve can be viewed as a Lipschitz curve.

**Lemma 2.12** (Reparametrization). *Given  $\gamma \in \text{FAC}([a, b]; X)$  with length  $L := L(\gamma)$ , set*

$$\mathfrak{s}(t) := \int_a^t |\gamma'_+|(r) \, dr, \quad \mathfrak{t}(s) := \min\{t \in [a, b] \mid \mathfrak{s}(t) = s\}.$$

Then  $\mathfrak{s} : [a, b] \rightarrow [0, L]$  is a non-decreasing absolutely continuous function and  $\mathfrak{t} : [0, L] \rightarrow [a, b]$  is a left-continuous increasing function such that  $\mathfrak{s}(a_+) = 0$ ,  $\mathfrak{s}(b_-) = L$  and  $\mathfrak{s} \circ \mathfrak{t}(s) = s$ . Moreover, the curve  $\hat{\gamma} := \gamma \circ \mathfrak{t} : [0, L] \rightarrow X$  is 1-Lipschitz and satisfies  $\gamma = \hat{\gamma} \circ \mathfrak{s}$  and  $|\hat{\gamma}'_+| = 1$   $\mathcal{L}^1$ -a.e. in  $[0, L]$ .

**2.3. Asymmetric metrics on Finsler manifolds.** In this subsection, we discuss the case of Finsler manifolds as a typical example of asymmetric metric spaces.

**2.3.1. Finsler manifolds.** We first recall the basics of Finsler geometry; see [3, 27, 40] for further reading. Let  $M$  be an  $n$ -dimensional connected  $C^\infty$ -manifold without boundary, and  $TM = \bigcup_{x \in M} T_x M$  be its tangent bundle. We call  $(M, F)$  a *Finsler manifold* if a nonnegative function  $F : TM \rightarrow [0, \infty)$  satisfies

- (1)  $F \in C^\infty(TM \setminus \{0\})$ ;
- (2)  $F(cv) = cF(v)$  for all  $v \in TM$  and  $c \geq 0$ ;
- (3) For any  $v \in T_x M \setminus \{0\}$ , the  $n \times n$  symmetric matrix

$$g_{ij}(v) := \frac{1}{2} \frac{\partial^2 [F^2]}{\partial v^i \partial v^j}(v)$$

is positive-definite, where  $v = \sum_{i=1}^n v^i (\partial/\partial x^i)|_x$  in a chart  $(x^i)_{i=1}^n$  around  $x$ .

We remark that  $g_{ij}(v)$  cannot be defined at  $v = 0$  unless  $F$  is Riemannian. For  $v \in T_x M \setminus \{0\}$ , we define a Riemannian metric  $g_v$  of  $T_x M$  by  $g_v(w, \bar{w}) := \sum_{i,j=1}^n g_{ij}(v) w^i \bar{w}^j$ . Euler's homogeneous function theorem yields that  $F^2(v) = g_v(v, v)$  for any  $v \in T_x M \setminus \{0\}$ . The *reversibility* of  $F$  on  $U \subset M$  is defined as

$$\lambda_F(U) := \sup_{v \in TU \setminus \{0\}} \frac{F(-v)}{F(v)}.$$

If  $\lambda_F(M) = 1$  (i.e.,  $F(-v) = F(v)$  for all  $v \in TM$ ), then we say that  $F$  is *reversible*. Note that  $\lambda_F(U) < \infty$  for any compact set  $U \subset M$  thanks to the smoothness of  $F$ .

We define the *dual Finsler metric*  $F^*$  of  $F$  by

$$F^*(\zeta) := \sup_{v \in T_x M \setminus \{0\}} \frac{\zeta(v)}{F(v)}, \quad \zeta \in T_x^* M.$$

By definition, we have

$$\zeta(v) \leq F(v)F^*(\zeta). \quad (2.2)$$

Then the *Legendre transformation*  $\mathfrak{L} : T_x M \rightarrow T_x^* M$  is defined by  $\mathfrak{L}(v) := \zeta$ , where  $\zeta$  is the unique element satisfying  $F(v) = F^*(\zeta)$  and  $\zeta(v) = F^2(v)$ . Note that  $\mathfrak{L} : TM \setminus \{0\} \rightarrow T^*M \setminus \{0\}$  is a diffeomorphism. For  $f \in C^1(M)$ , the *gradient vector field* of  $f$  is defined by  $\nabla f := \mathfrak{L}^{-1}(df)$ . We have  $df(v) = g_{\nabla f}(\nabla f, v)$  provided  $df \neq 0$ . We remark that the gradient  $\nabla$  is nonlinear; indeed,  $\nabla(f + h) \neq \nabla f + \nabla h$  and  $\nabla(-f) \neq -\nabla f$  in general (the latter is due to the irreversibility of  $F$ ). Note also that

$$\limsup_{y \rightarrow x} \frac{f(y) - f(x)}{d_F(x, y)} = F^*(df(x)) = F(\nabla f(x)), \quad (2.3)$$

where  $d_F$  is the Finsler distance function defined below.

**2.3.2. Length structure and absolutely continuous curves.** Let  $\mathcal{A}_\infty([0, 1]; M)$  denote the class of piecewise smooth curves in  $M$  defined on  $[0, 1]$ . Given  $\gamma \in \mathcal{A}_\infty([0, 1]; M)$ , we define its *length* by

$$L_F(\gamma) := \int_0^1 F(\gamma'(t)) dt.$$

Then the associated *distance function*  $d_F : M \times M \rightarrow [0, \infty)$  is defined as

$$d_F(x, y) := \inf\{L_F(\gamma) \mid \gamma \in \mathcal{A}_\infty([0, 1]; M), \gamma(0) = x, \gamma(1) = y\}.$$

Note that  $d_F : M \times M \rightarrow [0, \infty)$  is a continuous function and  $(M, d_F)$  is an asymmetric metric space in the sense of Definition 2.1. Indeed,  $d_F(x, y)$  may not coincide with  $d_F(y, x)$  unless  $F$  is reversible. Observe also that  $\mathcal{T}_+ = \mathcal{T}_-$  is exactly the original topology of  $M$ .

A Finsler manifold is said to be *forward* (resp. *backward*) *complete* if  $(M, d_F)$  is forward (resp. backward) complete in the sense of Definition 2.6 (see [3, Theorem 6.6.1] for a Finsler analogue of the Hopf–Rinow theorem). As for an estimate of the reversibility, one has the following (see [15, Theorem 2.23]).

**Theorem 2.13.** *Let  $(M, F)$  be a forward complete Finsler manifold. Then, for any  $\star \in M$ , the triple  $(M, \star, d_F)$  is a pointed forward  $\Theta$ -metric space for*

$$\Theta(r) := \lambda_F\left(B_\star^+(2r + \lambda_F(B_\star^+(r))r)\right).$$

Since  $M$  is smooth and the reversibility is finite on every compact set, the study of absolutely continuous curves can be largely reduced to the case of Riemannian manifolds or Euclidean spaces. We only briefly explain for thoroughness. According to [5], we say that a curve  $\gamma : [0, 1] \rightarrow M$  is *absolutely continuous* if, for any chart  $\varphi : U \rightarrow \mathbb{R}^n$  of  $M$ , the composition  $\varphi \circ \gamma : \gamma^{-1}(U) \rightarrow \varphi(U) \subset \mathbb{R}^n$  is locally absolutely continuous, i.e., absolutely continuous on any closed subinterval of  $\gamma^{-1}(U)$ . Let  $\mathcal{A}_{ac}([0, 1]; M)$  denote the class of absolutely continuous curves defined on  $[0, 1]$ . For any  $\gamma \in \mathcal{A}_{ac}([0, 1]; M)$ , the derivative  $\gamma'(t)$  exists for  $\mathcal{L}^1$ -a.e.  $t \in [0, 1]$  and we can define

$$L_F(\gamma) := \int_0^1 F(\gamma'(t)) dt.$$

Note that  $L_F(\gamma) < \infty$  and  $F(\gamma') \in L^1([0, 1])$ . Moreover, we have

$$\lim_{\delta \rightarrow 0^+} \frac{d(\gamma(t), \gamma(t + \delta))}{\delta} = F(\gamma'(t)) \quad \text{for } \mathcal{L}^1\text{-a.e. } t \in (0, 1). \quad (2.4)$$

One can see  $\mathcal{A}_{\text{ac}}([0, 1]; M) = \text{FAC}([0, 1]; M)$  by an argument similar to that of [5, Proposition 3.18]. Then, for  $\gamma \in \text{FAC}([0, 1]; M)$ , we find from (2.4) that

$$|\gamma'_+(t)| = F(\gamma'(t)) \quad \text{for } \mathcal{L}^1\text{-a.e. } t \in (0, 1). \quad (2.5)$$

**2.4. Upper gradients.** In this subsection, we introduce upper gradients for functions on asymmetric metric spaces. First, let us consider the case of Finsler manifolds.

**Example 2.14.** Let  $(M, d_F)$  be a forward metric space induced by a forward complete Finsler manifold  $(M, F)$ . Given  $\phi \in C^1(M)$ , for any  $\gamma \in \text{FAC}([a, b]; M)$ ,  $\phi \circ \gamma$  is absolutely continuous and (2.2) yields

$$(\phi \circ \gamma)'(t) = d\phi(\gamma'(t)) \leq F^*(d\phi(\gamma(t)))F(\gamma'(t)) = F(\nabla\phi(\gamma(t)))F(\gamma'(t)) \quad (2.6)$$

for  $\mathcal{L}^1$ -a.e.  $t \in (a, b)$ . Hence, a nonnegative function  $\mathbf{g} : M \rightarrow [0, \infty)$  satisfies  $F(\nabla\phi) \leq \mathbf{g}$  if and only if

$$\phi(\gamma(t_2)) - \phi(\gamma(t_1)) \leq \int_{t_1}^{t_2} \mathbf{g}(\gamma(t))F(\gamma'(t)) dt$$

for all  $\gamma \in \text{FAC}([a, b]; M)$  and  $a \leq t_1 \leq t_2 \leq b$ .

Now, let  $(X, d)$  be a forward complete forward metric space. In what follows, let  $\phi : X \rightarrow (-\infty, \infty]$  denote a *proper* function, i.e., its *proper effective domain*  $\mathfrak{D}(\phi) := \{x \in X \mid \phi(x) < \infty\}$  is nonempty.

**Definition 2.15** (Strong upper gradients). A function  $\mathbf{g} : X \rightarrow [0, \infty]$  is called a *strong upper gradient* for  $\phi$  if, for every curve  $\gamma \in \text{FAC}([a, b]; X)$ ,  $\mathbf{g} \circ \gamma$  is Borel and satisfies

$$\phi(\gamma(t_2)) - \phi(\gamma(t_1)) \leq \int_{t_1}^{t_2} \mathbf{g}(\gamma(t))|\gamma'_+(t)| dt \quad \text{for all } a \leq t_1 \leq t_2 \leq b. \quad (2.7)$$

Notice that, as is natural in view of Example 2.14 (see also Example 2.20), we did not take the absolute value in the left-hand side of (2.7). Therefore, our definition of upper gradients is weaker than [6, Definition 3.6].

If  $\mathbf{g} \circ \gamma |\gamma'_+| \in L^1(a, b)$ , then  $\phi \circ \gamma$  is absolutely continuous (see the remark below) and

$$(\phi \circ \gamma)'(t) \leq \mathbf{g}(\gamma(t))|\gamma'_+(t)| \quad \text{for } \mathcal{L}^1\text{-a.e. } t \in (a, b).$$

**Remark 2.16.** For  $\gamma \in \text{FAC}([a, b]; X)$ , we deduce from (2.7) for  $\gamma$  and its reverse curve that

$$|\phi(\gamma(t_1)) - \phi(\gamma(t_2))| \leq \Theta(d(\star, \gamma(0)) + L(\gamma)) \int_{t_1}^{t_2} \mathbf{g}(\gamma(t))|\gamma'_+(t)| dt$$

for all  $a \leq t_1 \leq t_2 \leq b$ . Hence, if  $\mathbf{g} \circ \gamma |\gamma'_+| \in L^1(a, b)$ , then  $\phi \circ \gamma$  is absolutely continuous and

$$|(\phi \circ \gamma)'(t)| \leq \Theta(d(\star, \gamma(0)) + L(\gamma))\mathbf{g}(\gamma(t))|\gamma'_+(t)| \quad \text{for } \mathcal{L}^1\text{-a.e. } t \in (a, b).$$

**Definition 2.17** (Weak upper gradients). A function  $\mathbf{g} : X \rightarrow [0, \infty]$  is called a *weak upper gradient* for  $\phi$  if, for every curve  $\gamma \in \text{FAC}([a, b]; X)$  satisfying

- (1)  $\mathbf{g} \circ \gamma |\gamma'_+| \in L^1(a, b)$ ;
- (2)  $\phi \circ \gamma$  is  $\mathcal{L}^1$ -a.e. equal to a function  $\varphi$  with finite pointwise variation in  $(a, b)$  (see [1, (1.1.2)]),

we have

$$\varphi'(t) \leq \mathbf{g}(\gamma(t))|\gamma'_+(t)| \quad \text{for } \mathcal{L}^1\text{-a.e. } t \in (a, b). \quad (2.8)$$

Note that a strong upper gradient is a weak upper gradient. A sufficient condition for a weak upper gradient to be a strong upper gradient is as follows.

**Proposition 2.18.** *Let  $\mathbf{g}$  be a weak upper gradient for  $\phi$ . If  $\phi \circ \gamma$  is absolutely continuous for every  $\gamma \in \text{FAC}([a, b]; X)$  with  $\mathbf{g} \circ \gamma |\gamma'_+| \in L^1(a, b)$ , then  $\mathbf{g}$  is a strong upper gradient for  $\phi$ .*

*Proof.* Let  $\gamma \in \text{FAC}([a, b]; X)$  and  $a \leq t_1 \leq t_2 \leq b$ . On the one hand, if  $\mathbf{g} \circ \gamma |_{\gamma'_+} \in L^1(t_1, t_2)$ , then  $\varphi = \phi \circ \gamma$  satisfies (2.8) and hence (2.7) holds. On the other hand, (2.7) is trivial if  $\mathbf{g} \circ \gamma |_{\gamma'_+} \notin L^1(t_1, t_2)$ .  $\square$

**Definition 2.19** (Slopes). The *local* and *global (descending) slopes* of  $\phi$  at  $x \in \mathfrak{D}(\phi)$  are defined by

$$|\partial\phi|(x) := \limsup_{y \rightarrow x} \frac{[\phi(x) - \phi(y)]_+}{d(x, y)}, \quad \mathfrak{l}_\phi(x) := \sup_{y \neq x} \frac{[\phi(x) - \phi(y)]_+}{d(x, y)},$$

respectively, where  $[a]_+ := \max\{a, 0\}$ . For  $x \in X \setminus \mathfrak{D}(\phi)$ , we set  $|\partial\phi|(x) = \mathfrak{l}_\phi(x) := \infty$ .

The local slope  $|\partial\phi|$  represents how fast the function  $\phi$  can decrease. We remark that  $|\partial(-\phi)| \neq |\partial\phi|$  even in symmetric metric spaces.

**Example 2.20.** Let  $(M, d_F)$  and  $\phi$  be as in Example 2.14. It follows from (2.3) that  $F(\nabla(-\phi)(x)) = |\partial\phi|(x)$ . We remark that  $F(\nabla(-\phi))$  may not coincide with either  $F(-\nabla\phi)$  or  $F(\nabla\phi)$ . Note also that  $|\partial\phi| = F(\nabla(-\phi))$  is a strong upper gradient for  $-\phi$  by (2.5) and (2.6).

In general, we have the following (cf. [1, Theorem 1.2.5]).

**Theorem 2.21** (Slopes are upper gradients). *Let  $(X, d)$  be a forward complete forward metric space and  $\phi : X \rightarrow (-\infty, \infty]$  be a proper function.*

- (i)  $|\partial\phi|$  is a weak upper gradient for  $-\phi$ ;
- (ii) If  $\phi$  is lower semicontinuous (with respect to  $\mathcal{T}_+$ ), then  $\mathfrak{l}_\phi$  is lower semicontinuous and a strong upper gradient for  $-\phi$ .

*Proof.* (i) Let  $\gamma \in \text{FAC}([a, b]; X)$  and  $\varphi$  satisfy the assumptions (1), (2) in Definition 2.17 for  $-\phi$ . Set

$$A := \{t \in (a, b) \mid -\phi(\gamma(t)) = \varphi(t), \varphi \text{ is differentiable at } t, |\gamma'_+(t)| \text{ exists}\}.$$

Note that  $(a, b) \setminus A$  is  $\mathcal{L}^1$ -negligible (by Theorem 2.9). To see (2.8), it suffices to consider  $t \in A$  with  $\varphi'(t) > 0$ . Since  $\varphi'(t) > 0$ , we may assume that  $\gamma(s) \neq \gamma(t)$  for  $s (\neq t)$  close to  $t$ . Then we have

$$\begin{aligned} \varphi'(t) &= \lim_{s \rightarrow t^+, s \in A} \frac{\phi(\gamma(t)) - \phi(\gamma(s))}{s - t} \leq \limsup_{s \rightarrow t^+, s \in A} \frac{\phi(\gamma(t)) - \phi(\gamma(s))}{d(\gamma(t), \gamma(s))} \limsup_{s \rightarrow t^+, s \in A} \frac{d(\gamma(t), \gamma(s))}{s - t} \\ &\leq |\partial\phi|(\gamma(t)) |\gamma'_+(t)|. \end{aligned}$$

Therefore,  $|\partial\phi|$  is a weak upper gradient for  $-\phi$ .

(ii) We first prove the lower semicontinuity. It follows from the assumption that  $x \mapsto [\phi(x) - \phi(y)]_+$  is lower semicontinuous. Hence, for any sequence  $x_i \rightarrow x$  and  $y \neq x$ , we have  $x_i \neq y$  for large  $i$  and

$$\liminf_{i \rightarrow \infty} \mathfrak{l}_\phi(x_i) \geq \liminf_{i \rightarrow \infty} \frac{[\phi(x_i) - \phi(y)]_+}{d(x_i, y)} \geq \frac{[\phi(x) - \phi(y)]_+}{d(x, y)}.$$

Then, taking the supremum in  $y \neq x$  furnishes the lower semicontinuity of  $\mathfrak{l}_\phi$ .

Next, we prove that  $\mathfrak{l}_\phi$  is a strong upper gradient for  $-\phi$ . Since  $|\partial\phi| \leq \mathfrak{l}_\phi$  by definition,  $\mathfrak{l}_\phi$  is a weak upper gradient for  $-\phi$ . Therefore, in view of Proposition 2.18, it is sufficient to show the following.

**Claim 2.22.** *For any  $\gamma \in \text{FAC}([a, b]; X)$  with  $\mathfrak{l}_\phi \circ \gamma |_{\gamma'_+} \in L^1(a, b)$ ,  $-\phi \circ \gamma$  is absolutely continuous.*

To this end, for  $\mathbf{t} : [0, L] \rightarrow [a, b]$  given by Lemma 2.12 with  $L = L(\gamma)$ , we set

$$\hat{\gamma} := \gamma \circ \mathbf{t}, \quad \varphi := -\phi \circ \hat{\gamma}, \quad g := \mathfrak{l}_\phi \circ \hat{\gamma}.$$

Since  $L < \infty$  and  $\hat{\gamma}([0, L]) = \gamma([a, b])$ , the triangle inequality yields

$$\lambda := \sup_{s \in [0, L]} \Theta(d(\star, \hat{\gamma}(s))) \leq \Theta(d(\star, \gamma(a)) + L) < \infty.$$

Recalling that  $\hat{\gamma}$  is a 1-Lipschitz curve, we have  $d(\hat{\gamma}(s_1), \hat{\gamma}(s_2)) \leq \lambda |s_1 - s_2|$  for all  $s_1, s_2 \in [0, L]$  (regardless of the order of  $s_1, s_2$ ), thereby

$$[\varphi(s_2) - \varphi(s_1)]_+ \leq \lambda g(s_1) |s_1 - s_2| \quad \text{for all } s_1, s_2 \in [0, L]. \quad (2.9)$$

Thus, we have

$$|\varphi(s_1) - \varphi(s_2)| = [\varphi(s_1) - \varphi(s_2)]_+ + [\varphi(s_2) - \varphi(s_1)]_+ \leq \lambda(g(s_1) + g(s_2)) |s_1 - s_2|. \quad (2.10)$$

Moreover, by the hypothesis  $\iota_\phi \circ \gamma |\gamma'_+| \in L^1(a, b)$  and Lemma 2.12, we find

$$\int_0^L g(s) ds = \int_0^L \iota_\phi(\gamma(\mathbf{t}(s))) ds = \int_a^b \iota_\phi(\gamma(t)) |\gamma'_+(t)| dt < \infty.$$

Therefore, we obtain  $g \in L^1(0, L)$ , which together with (2.10) and [1, Lemma 1.2.6] yields that  $\varphi$  belongs to  $W^{1,1}(0, L)$  with  $|\varphi'| \leq 2\lambda g$  and has a continuous representative.

To see that  $\varphi$  itself is continuous, on the one hand, note that  $\varphi$  is upper semicontinuous by the lower semicontinuity of  $\phi$ . On the other hand, we infer from (2.9) and  $g \in L^1(0, L)$  that

$$\liminf_{\varepsilon \rightarrow 0^+} \frac{1}{2\varepsilon} \int_{-\varepsilon}^{\varepsilon} \varphi(s+r) dr \geq \varphi(s) \quad \text{for all } s \in (0, L).$$

This implies that  $\varphi$  is continuous and, since it lives in  $W^{1,1}(0, L)$ , absolutely continuous. Then we observe from  $-\phi(\gamma(t)) = \varphi(\mathbf{s}(t))$  that  $-\phi \circ \gamma$  is absolutely continuous. This completes the proof of Claim 2.22.  $\square$

**2.5. Curves of maximal slope.** Let  $(X, d)$  be a forward complete forward metric space. For  $p \in [1, \infty]$ , denote by  $\text{FAC}_{\text{loc}}^p((a, b); X)$  the class of *locally  $p$ -forward absolutely continuous curves*  $\xi$  defined on  $(a, b)$ , i.e.,  $\xi|_{[s,t]} \in \text{FAC}^p([s, t]; X)$  for every  $a < s < t < b$ .

**Definition 2.23** (Curves of maximal slope). Let  $\phi : X \rightarrow (-\infty, \infty]$  be a proper function,  $\mathbf{g}$  be a weak upper gradient for  $-\phi$ , and  $p \in (1, \infty)$ . We call  $\xi \in \text{FAC}_{\text{loc}}^1((a, b); X)$  a  *$p$ -curve of maximal slope* for  $\phi$  with respect to  $\mathbf{g}$  if  $\phi \circ \xi$  is  $\mathcal{L}^1$ -a.e. equal to a non-increasing function  $\varphi$  satisfying

$$\varphi'(t) \leq -\frac{1}{p} |\xi'_+|^p(t) - \frac{1}{q} \mathbf{g}^q(\xi(t)) \quad \text{for } \mathcal{L}^1\text{-a.e. } t \in (a, b), \quad (2.11)$$

where  $1/p + 1/q = 1$ . In the case of  $p = 2$ , we simply call  $\xi$  a *curve of maximal slope*.

In fact, equality holds in (2.11) for  $\mathcal{L}^1$ -a.e.  $t \in (a, b)$  as follows.

**Proposition 2.24** (Energy identity). *If  $\xi : (a, b) \rightarrow X$  is a  $p$ -curve of maximal slope for  $\phi$  with respect to  $\mathbf{g}$ , then we have  $\xi \in \text{FAC}_{\text{loc}}^p((a, b); X)$  and  $\mathbf{g} \circ \xi \in L_{\text{loc}}^q(a, b)$  with*

$$|\xi'_+|^p(t) = \mathbf{g}^q(\xi(t)) = -\varphi'(t) \quad \text{for } \mathcal{L}^1\text{-a.e. } t \in (a, b). \quad (2.12)$$

Moreover, if  $\mathbf{g}$  is a strong upper gradient for  $-\phi$ , then  $\varphi = \phi \circ \xi$  is locally absolutely continuous in  $(a, b)$  and satisfies the energy identity

$$\frac{1}{p} \int_s^t |\xi'_+|^p(r) dr + \frac{1}{q} \int_s^t \mathbf{g}^q(\xi(r)) dr = \phi(\xi(s)) - \phi(\xi(t)) \quad \text{for all } a < s < t < b. \quad (2.13)$$

*Proof.* Since  $\varphi$  is non-increasing,  $\varphi'(t)$  is locally integrable. This together with (2.11) implies  $\mathbf{g} \circ \xi \in L_{\text{loc}}^q(a, b)$  and  $|\xi'_+| \in L_{\text{loc}}^p(a, b)$ , and hence  $\xi \in \text{FAC}_{\text{loc}}^p((a, b); X)$  and  $\mathbf{g} \circ \xi |\xi'_+| \in L_{\text{loc}}^1(a, b)$ . For  $\mathcal{L}^1$ -a.e.  $t \in (a, b)$ , since  $\mathbf{g}$  is a weak upper gradient for  $-\phi$ , we have  $-\varphi'(t) \leq \mathbf{g}(\xi(t)) |\xi'_+(t)|$ . Combining this with the Young inequality yields the reverse inequality to (2.11), thereby we obtain (2.12).

If  $\mathbf{g}$  is additionally a strong upper gradient, then  $\phi \circ \xi$  is locally absolutely continuous in  $(a, b)$  (recall Remark 2.16) and we have  $\varphi = \phi \circ \xi$ . Then it follows from the Young inequality and (2.11) that

$$\begin{aligned} \phi(\xi(s)) - \phi(\xi(t)) &\leq \int_s^t \mathbf{g}(\xi(r)) |\xi'_+(r)| dr \leq \frac{1}{p} \int_s^t |\xi'_+|^p(r) dr + \frac{1}{q} \int_s^t \mathbf{g}^q(\xi(r)) dr \\ &\leq - \int_s^t \varphi'(r) dr = \phi(\xi(s)) - \phi(\xi(t)). \end{aligned}$$

This furnishes the energy identity (2.13).  $\square$

**Example 2.25.** Let  $(M, d_F)$  and  $\phi$  be as in Example 2.14. According to (2.5) and Example 2.20, if  $\xi : (a, b) \rightarrow M$  is a  $p$ -curve of maximal slope for  $\phi$  with respect to  $F(\nabla(-\phi)) = F^*(-d\phi)$ , then we have

$$(\phi \circ \xi)'(t) = -F^p(\xi'(t)) = -F^*(-d\phi(\xi(t)))^q. \quad (2.14)$$

This implies  $(\phi \circ \xi)'(t) = -F(\xi'(t))F^*(-d\phi(\xi(t)))$ , and hence  $\xi'(t) = \alpha(t)\nabla(-\phi)(\xi(t))$  holds for some  $\alpha(t) \geq 0$ . Actually, we deduce from (2.14) that

$$\xi'(t) = \begin{cases} F^{\frac{2-p}{p-1}}(\nabla(-\phi)(\xi(t))) \cdot \nabla(-\phi)(\xi(t)) & \text{if } \nabla(-\phi)(\xi(t)) \neq 0, \\ 0 & \text{if } \nabla(-\phi)(\xi(t)) = 0. \end{cases} \quad (2.15)$$

In particular,  $\xi$  is  $C^1$  since  $\phi$  is  $C^1$ , thereby (2.14) holds for all  $t \in (a, b)$ . We may rewrite (2.15) as  $j_p(\xi'(t)) = \nabla(-\phi)(\xi(t))$  by introducing an operator  $j_p : TM \rightarrow TM$  defined by  $j_p(v) := F^{p-2}(v)v$  if  $v \neq 0$  and  $j_p(0) := 0$ . In the case of  $p = 2$ , we obtain the usual gradient flow equation  $\xi'(t) = \nabla(-\phi)(\xi(t))$ . We stress that  $\xi'(t) = -\nabla\phi(\xi(t))$  holds only when  $F$  is reversible.

We conclude the section with a comparison to the setting of [36].

**Remark 2.26.** In [36], they considered a convex function  $\psi : [0, \infty) \rightarrow [0, \infty]$  satisfying some natural conditions (see [36, (2.30)]) and investigated curves  $\xi$  fulfilling

$$\phi'(t) \leq -\psi(|\xi'_+(t)|) - \psi^*(\mathfrak{g}(\xi(t))) \quad (2.16)$$

instead of (2.11), where  $\psi^*$  is the Legendre–Fenchel–Moreau transform of  $\psi$ . Choosing  $\psi(x) = x^p/p$  recovers (2.11). Nonetheless, to establish the corresponding existence theory, they assumed that  $\phi$  is bounded from below (see [36, (2.19b)]), which is unnecessary in the present paper.

### 3. GENERALIZED MINIMIZING MOVEMENTS AND CURVES OF MAXIMAL SLOPE

**3.1. Problem and strategy.** Throughout this section, let  $(X, d)$  be a forward complete forward metric space,  $\phi : X \rightarrow (-\infty, \infty]$  be a proper function, and  $p \in (1, \infty)$ . The main objective of this section is to study the following problem.

**Problem 3.1.** Given an initial datum  $x_0 \in \mathfrak{D}(\phi)$ , does there exist a  $p$ -curve  $\xi : (0, \infty) \rightarrow X$  of maximal slope for  $\phi$  such that  $\lim_{t \rightarrow 0} \xi(t) = x_0$ ?

We shall solve this problem via a discrete approximation. We begin with some definitions and notations.

**Definition 3.2** (Resolvent operator). We define the  $p$ -resolvent operator by, for  $\tau > 0$  and  $x \in X$ ,

$$J_\tau[x] := \operatorname{argmin} \Phi(\tau, x; \cdot),$$

where

$$\Phi(\tau, x; y) := \phi(y) + \frac{d^p(x, y)}{p\tau^{p-1}}, \quad y \in X.$$

That is to say,  $y \in J_\tau[x]$  if and only if  $\Phi(\tau, x; y) \leq \Phi(\tau, x; z)$  for all  $z \in X$ .

Let  $P_{\mathfrak{T}} := \{0 = t_{\mathfrak{T}}^0 < t_{\mathfrak{T}}^1 < \dots < t_{\mathfrak{T}}^k < \dots\}$  be a partition of the time interval  $[0, \infty)$  corresponding to a sequence of positive time steps  $\mathfrak{T} = (\tau_k)_{k \geq 1}$  in the sense that

$$\tau_k = t_{\mathfrak{T}}^k - t_{\mathfrak{T}}^{k-1}, \quad \lim_{k \rightarrow \infty} t_{\mathfrak{T}}^k = \sum_{k=1}^{\infty} \tau_k = \infty.$$

Set  $\|\mathfrak{T}\| := \sup_{k \geq 1} \tau_k$ . We will consider the following recursive scheme:

$$\text{Given } \Xi_{\mathfrak{T}}^0 \in X, \text{ whenever } \Xi_{\mathfrak{T}}^1, \dots, \Xi_{\mathfrak{T}}^{k-1} \text{ are known, take } \Xi_{\mathfrak{T}}^k \in J_{\tau_k}[\Xi_{\mathfrak{T}}^{k-1}]. \quad (3.1)$$

This is a well-known scheme to construct (descending) gradient curves of  $\phi$ . The following example in the case of Minkowski spaces may be helpful to understand the choice of  $J_\tau[x]$  as above.

**Example 3.3.** Let  $(\mathbb{R}^n, F)$  be a *Minkowski space*, i.e., each of its tangent spaces is canonically isometric to  $(\mathbb{R}^n, F)$ , and  $\phi \in C^1(\mathbb{R}^n)$ . For any  $C^1$ -curve  $\gamma : (-\varepsilon, \varepsilon) \rightarrow \mathbb{R}^n$  with  $\gamma(0) = \Xi_{\mathfrak{T}}^k$  and  $\gamma'(0) = v$ , we have

$$\frac{d}{dt} \Big|_{t=0} d(\Xi_{\mathfrak{T}}^{k-1}, \gamma(t)) = \frac{d}{dt} \Big|_{t=0} F(\gamma(t) - \Xi_{\mathfrak{T}}^{k-1}) = \frac{g_{\Xi_{\mathfrak{T}}^k - \Xi_{\mathfrak{T}}^{k-1}}(\Xi_{\mathfrak{T}}^k - \Xi_{\mathfrak{T}}^{k-1}, v)}{F(\Xi_{\mathfrak{T}}^k - \Xi_{\mathfrak{T}}^{k-1})},$$

provided  $\Xi_{\mathfrak{X}}^k \neq \Xi_{\mathfrak{X}}^{k-1}$ . Combining this with  $\frac{d}{dt}|_{t=0}\Phi(\tau_k, \Xi_{\mathfrak{X}}^{k-1}; \gamma(t)) = 0$  by the choice (3.1) of  $\Xi_{\mathfrak{X}}^k$ , we find

$$-d\phi(v) = g_{\Xi_{\mathfrak{X}}^k - \Xi_{\mathfrak{X}}^{k-1}}(w, v) = [\mathfrak{L}(w)](v), \quad \text{where } w = \frac{F^{p-2}(\Xi_{\mathfrak{X}}^k - \Xi_{\mathfrak{X}}^{k-1})}{\tau_k^{p-1}}(\Xi_{\mathfrak{X}}^k - \Xi_{\mathfrak{X}}^{k-1}).$$

Since  $v$  was arbitrary, we arrive at the equation  $w = \mathfrak{L}^{-1}(-d\phi(\Xi_{\mathfrak{X}}^k)) = \nabla(-\phi)(\Xi_{\mathfrak{X}}^k)$ . By the choice of  $w$ , this is equivalent to

$$\frac{\Xi_{\mathfrak{X}}^k - \Xi_{\mathfrak{X}}^{k-1}}{\tau_k} = F^{\frac{2-p}{p-1}}(\nabla(-\phi)(\Xi_{\mathfrak{X}}^k)) \cdot \nabla(-\phi)(\Xi_{\mathfrak{X}}^k),$$

which can be regarded as a discrete version of (2.15).

**Definition 3.4** (Discrete solutions). Given  $\mathfrak{X}, \Xi_{\mathfrak{X}}^0 \in X$  and a sequence  $(\Xi_{\mathfrak{X}}^k)_{k \geq 1}$  solving (3.1), we define a piecewise constant curve  $\bar{\Xi}_{\mathfrak{X}} : [0, \infty) \rightarrow X$  by

$$\bar{\Xi}_{\mathfrak{X}}(0) := \Xi_{\mathfrak{X}}^0, \quad \bar{\Xi}_{\mathfrak{X}}(t) := \Xi_{\mathfrak{X}}^k \quad \text{for } t \in (t_{\mathfrak{X}}^{k-1}, t_{\mathfrak{X}}^k], \quad k \geq 1.$$

We call  $\bar{\Xi}_{\mathfrak{X}}$  a *discrete solution* corresponding to the partition  $P_{\mathfrak{X}}$ .

Under appropriate conditions on  $(X, d)$  and  $\phi$ , we shall solve Problem 3.1 in the following steps:

- Show that the minimization algorithm (3.1) starting from  $x_0$  is solvable;
- Find a sequence  $(P_{\mathfrak{X}_m})_m$  of admissible partitions with  $\|\mathfrak{X}_m\| \rightarrow 0$  such that the discrete solutions  $(\bar{\Xi}_{\mathfrak{X}_m}^k)_{k \geq 1}$  converge to a solution to Problem 3.1 with respect to a suitable topology  $\sigma$  on  $X$ .

**3.2. Topological assumptions.** In the sequel, we always assume that  $\sigma$  is a Hausdorff topology on  $X$ , possibly different from  $\mathcal{T}_{\pm}$ , compatible with  $d$  in the following sense:

- (1)  $\sigma$  is weaker than the forward topology  $\mathcal{T}_+$  induced from  $d$  (i.e.,  $x_i \rightarrow x$  in  $\mathcal{T}_+$  implies  $x_i \rightarrow x$  in  $\sigma$ );
- (2)  $d$  is  $\sigma$ -sequentially lower semicontinuous (i.e., if  $x_i \rightarrow x$  and  $y_i \rightarrow y$  in  $\sigma$ , then  $\liminf_{i \rightarrow \infty} d(x_i, y_i) \geq d(x, y)$ ).

We will denote by  $x_i \xrightarrow{\sigma} x$  the convergence with respect to the topology  $\sigma$ .

**Remark 3.5** (Topology comparison). (a) Recall that  $d$  is  $\mathcal{T}_+$ -continuous by Theorem 2.3, thereby  $\sigma = \mathcal{T}_+$  always satisfies (1) and (2) above.

(b) If  $x_i \xrightarrow{\mathcal{T}_-} x$ , then we deduce from the triangle inequality the following upper semicontinuity:

$$\limsup_{i \rightarrow \infty} d(x_i, y) \leq \lim_{i \rightarrow \infty} \{d(x_i, x) + d(x, y)\} = d(x, y).$$

(c) By the  $\sigma$ -sequential lower semicontinuity of  $d$ , the limit under  $\sigma$  is unique. Indeed, if  $x$  and  $x'$  are  $\sigma$ -limit points of  $(x_i)_{i \geq 1}$ , then  $0 = \liminf_{i \rightarrow \infty} d(x_i, x_i) \geq d(x, x')$  necessarily holds, thereby  $x = x'$ .

We give an example where  $\sigma$  is different from  $\mathcal{T}_{\pm}$  (see [1, Remark 2.3.9] for another example).

**Example 3.6** (Randers-like spaces). Let  $(X, \langle \cdot, \cdot \rangle)$  be a Hilbert space and  $\|x\| := \sqrt{\langle x, x \rangle}$ . Choose  $a \in X$  with  $\|a\| < 1$  and define a function  $d : X \times X \rightarrow [0, \infty)$  by

$$d(x, y) := \|y - x\| + \langle a, y - x \rangle.$$

Then  $(X, d)$  is a  $[(1 + \|a\|)/(1 - \|a\|)]$ -metric space (recall Definition 2.2), and  $\mathcal{T}_+ = \mathcal{T}_-$  coincides with the (strong) topology of  $(X, \langle \cdot, \cdot \rangle)$ . Now, let  $\sigma$  be the weak topology of  $X$ . Since  $\|\cdot\|$  is  $\sigma$ -sequentially lower semicontinuous, so is  $d$ .

**Lemma 3.7.** *Every  $\sigma$ -sequentially compact set  $K \subset X$  is forward complete.*

*Proof.* Given a forward Cauchy sequence  $(x_i)_{i \geq 1}$  in  $K$ , on the one hand, the  $\sigma$ -sequential compactness of  $K$  yields a subsequence  $(x_{i_j})_{j \geq 1}$  of  $(x_i)_{i \geq 1}$  converging to a point  $x' \in K$  in  $\sigma$ . On the other hand, the forward completeness of  $X$  furnishes a point  $x \in X$  such that  $x_i$  converges to  $x$  in  $\mathcal{T}_+$ . Since  $\sigma$  is weaker than  $\mathcal{T}_+$ ,  $x_i$  converges to  $x$  in  $\sigma$  as well. Hence,  $x_{i_j}$  converges to both  $x$  and  $x'$  in  $\sigma$ , and we find from Remark 3.5(c) that  $x = x' \in K$ . Thus,  $K$  is forward complete.  $\square$

A set  $A \subset X$  is said to be *forward bounded* if  $A \subset B_\star^+(r)$  for some  $r > 0$ . We remark that, thanks to  $\lambda_d(B_\star^+(r)) \leq \Theta(r)$  in Definition 2.2,  $A$  is forward bounded if and only if  $\sup_{x,y \in A} d(x,y) < \infty$ .

Now we introduce our main assumptions on  $(X, d)$  and  $\phi$ .

**Assumption 3.8.** (a) **Lower semicontinuity.**  $\phi$  is  $\sigma$ -sequentially lower semicontinuous on forward bounded sets, i.e., if  $\sup_{i,j} d(x_i, x_j) < \infty$  and  $x_i \xrightarrow{\sigma} x$ , then we have  $\liminf_{i \rightarrow \infty} \phi(x_i) \geq \phi(x)$ . (In particular,  $\phi$  is  $\mathcal{T}_+$ -lower semicontinuous.)

(b) **Coercivity.** There exist  $\tau_\star > 0$  and  $x_\star \in X$  such that

$$\Phi_{\tau_\star}(x_\star) := \inf_{y \in X} \Phi(\tau_\star, x_\star; y) = \inf_{y \in X} \left\{ \phi(y) + \frac{d^p(x_\star, y)}{p\tau_\star^{p-1}} \right\} > -\infty.$$

(c) **Compactness.** Every forward bounded set contained in a sublevel set of  $\phi$  is relatively  $\sigma$ -sequentially compact, i.e., if a sequence  $(x_i)_{i \geq 1}$  in  $X$  satisfies  $\sup_i \phi(x_i) < \infty$  and  $\sup_{i,j} d(x_i, x_j) < \infty$ , then it admits a  $\sigma$ -convergent subsequence.

**Remark 3.9** ( $\sigma = \mathcal{T}_+$  case). When  $\sigma = \mathcal{T}_+$ , (a) and (c) above can be rewritten as follows, respectively:

(a')  $\phi$  is  $\mathcal{T}_+$ -lower semicontinuous;

(c') Every forward bounded set in a sublevel set of  $\phi$  is relatively compact in  $X$ .

The next proposition presents one of the simplest situations where Assumption 3.8 holds (cf. [1, Remark 2.1.1]). We remark that  $\sigma$  may be different from  $\mathcal{T}_+$ .

**Proposition 3.10.** *Suppose that every sublevel set of  $\phi$  is compact in  $\mathcal{T}_+$ . Then Assumption 3.8 holds.*

*Proof.* (a) Assume that  $x_i \xrightarrow{\sigma} x$  and the limit  $\alpha := \lim_{i \rightarrow \infty} \phi(x_i)$  exists. To see  $\phi(x) \leq \alpha$ , suppose  $\alpha < \infty$  without loss of generality. For any  $\varepsilon > 0$ ,  $A_\varepsilon := \{y \in X \mid \phi(y) \leq \alpha + \varepsilon\}$  is compact by hypothesis, and hence a subsequence of  $(x_i)_{i \geq 1}$  converges to some  $x' \in A_\varepsilon$  in  $\mathcal{T}_+$ . Then  $x = x' \in A_\varepsilon$  by Remark 3.5(c), and the arbitrariness of  $\varepsilon$  yields  $\phi(x) \leq \alpha$  as desired.

(b) is seen by noticing  $\inf_X \phi > -\infty$ , which follows from (a). (c) is clear by hypothesis.  $\square$

**Remark 3.11.** In [36], Rossi, Mielke and Savaré investigated the *doubly nonlinear evolution equation* (DNE), which is more general than the gradient flow equation. Their topological requirements are close to ours, however, recall from Remark 2.26 that they assumed  $\inf_X \phi > -\infty$ , which is stronger than the coercivity above and can simplify some arguments below. We remark that Chenchiah–Rieger–Zimmer’s [6] is also concerned with the existence of (2-)curves of maximal slope in the asymmetric setting. On the one hand, they assumed the lower semicontinuity of  $d$  only in the second argument. On the other hand, they assumed that the backward convergence implies the forward convergence (see [6, Assumption 4.3]; compare it with [36, Remark 2.9] and Remark 2.4(a)) and used a stronger notion of upper gradient (recall Definition 2.15).

**3.3. Moreau–Yosida approximation.** In this subsection, we will present an existence result of solutions to (3.1). For this purpose, we recall the definition of Moreau–Yosida approximation.

**Definition 3.12** (Moreau–Yosida approximation). For  $\tau > 0$  and  $x \in X$ , the *Moreau–Yosida approximation*  $\Phi_\tau$  is defined as

$$\Phi_\tau(x) := \inf_{y \in X} \Phi(\tau, x; y) = \inf_{y \in X} \left\{ \phi(y) + \frac{d^p(x, y)}{p\tau^{p-1}} \right\}.$$

We also set

$$\tau_\star(\phi) := \sup\{\tau > 0 \mid \Phi_\tau(x) > -\infty \text{ for some } x \in X\}.$$

Note that Assumption 3.8(b) is equivalent to  $\tau_\star(\phi) > 0$ . Moreover, we have the following (cf. [1, Lemma 2.2.1]).

**Lemma 3.13.** *Suppose Assumption 3.8(b). For  $0 < \tau < \tau_\star \leq \tau_\star(\phi)$ , set*

$$\epsilon = \epsilon(p, \tau_\star, \tau) := \frac{\tau_\star^{p-1} - \tau^{p-1}}{2\tau^{p-1}} > 0, \quad C(p, \tau_\star, \tau) := \frac{\mathfrak{C}(p, \epsilon)}{p\tau_\star^{p-1}} > 0,$$

where  $\mathfrak{C}(p, \epsilon)$  is the constant introduced in Lemma A.1. Then we have

$$\Phi_\tau(x) \geq \Phi_{\tau_*}(x_*) - C(p, \tau_*, \tau) d^p(x_*, x), \quad (3.2)$$

$$d^p(x, y) \leq \frac{2p\tau^{p-1}\tau_*^{p-1}}{\tau_*^{p-1} - \tau^{p-1}} \left\{ \Phi(\tau, x; y) - \Phi_{\tau_*}(x_*) + C(p, \tau_*, \tau) d^p(x_*, x) \right\} \quad (3.3)$$

for all  $x, y \in X$ . In particular, sublevel sets of  $\Phi(\tau, x; \cdot)$  are forward bounded.

*Proof.* We deduce from Lemma A.1 (with  $a = d(x, y)$  and  $b = d(x_*, x)$ ) and the triangle inequality that

$$\frac{\tau_*^{p-1} + \tau^{p-1}}{2p\tau^{p-1}\tau_*^{p-1}} d^p(x, y) + C(p, \tau_*, \tau) d^p(x_*, x) \geq \frac{d^p(x_*, y)}{p\tau_*^{p-1}}$$

for any  $x, y \in X$ . By the definition of  $\Phi_{\tau_*}$ , this implies

$$\phi(y) + \frac{\tau_*^{p-1} + \tau^{p-1}}{2p\tau^{p-1}\tau_*^{p-1}} d^p(x, y) + C(p, \tau_*, \tau) d^p(x_*, x) \geq \Phi_{\tau_*}(x_*).$$

Then the first claim (3.2) follows since, for any  $y \in X$ ,

$$\begin{aligned} \Phi(\tau, x; y) &= \phi(y) + \frac{\tau_*^{p-1} + \tau^{p-1}}{2p\tau^{p-1}\tau_*^{p-1}} d^p(x, y) + \frac{\tau_*^{p-1} - \tau^{p-1}}{2p\tau^{p-1}\tau_*^{p-1}} d^p(x, y) \\ &\geq \Phi_{\tau_*}(x_*) - C(p, \tau_*, \tau) d^p(x_*, x) + \frac{\tau_*^{p-1} - \tau^{p-1}}{2p\tau^{p-1}\tau_*^{p-1}} d^p(x, y) \\ &\geq \Phi_{\tau_*}(x_*) - C(p, \tau_*, \tau) d^p(x_*, x). \end{aligned}$$

Observe also that the first inequality corresponds to the second claim (3.3). The forward boundedness of sublevel sets of  $\Phi(\tau, x; \cdot)$  readily follows from (3.3).  $\square$

Now we prove the existence of a solution to (3.1), giving a discrete solution as in Definition 3.4.

**Theorem 3.14** (Existence of discrete solutions). *Suppose Assumption 3.8(a)–(c). Then, for every  $\tau \in (0, \tau_*(\phi))$  and  $x \in X$ , we have  $J_\tau[x] \neq \emptyset$ . In particular, for any  $\Xi_{\mathfrak{I}}^0 \in X$  and partition  $P_{\mathfrak{I}}$  with  $\|\mathfrak{I}\| < \tau_*(\phi)$ , there exists a discrete solution  $\Xi_{\mathfrak{I}}$  corresponding to  $P_{\mathfrak{I}}$ .*

*Proof.* Given  $c > \Phi_\tau(x)$ , consider the sublevel set  $A := \{y \in X \mid \Phi(\tau, x; y) \leq c\}$ . Recall from Lemma 3.13 that  $A$  is forward bounded, and hence  $d$  is bounded on  $A \times A$ . Moreover, for any  $y \in A$ , we have

$$\Phi_\tau(x) - \frac{d^p(x, y)}{p\tau^{p-1}} \leq \phi(y) \leq \Phi(\tau, x; y) \leq c.$$

Thus,  $\phi$  is also bounded on  $A$ .

Next, we show that  $A$  is  $\sigma$ -sequentially compact. For any sequence  $(y_i)_{i \geq 1}$  in  $A$ , since  $\sup_{i,j} d(y_i, y_j)$  and  $\sup_i \phi(y_i)$  are bounded, Assumption 3.8(c) yields a subsequence  $(y_{i_k})_{k \geq 1}$  which is  $\sigma$ -convergent to some  $y_\infty \in X$ . Then, since both  $\phi$  and  $d$  are  $\sigma$ -sequentially lower semicontinuous, we find  $\Phi(\tau, x; y_\infty) \leq c$ , thereby  $y_\infty \in A$ . Hence,  $A$  is  $\sigma$ -sequentially compact.

By the  $\sigma$ -sequential compactness of  $A$  and the  $\sigma$ -sequential lower semicontinuity of  $\phi$  and  $d$ , we can take  $y_* \in A$  with  $\Phi(\tau, x; y_*) = \inf_{y \in A} \Phi(\tau, x; y) = \Phi_\tau(x)$ . This completes the proof.  $\square$

**Remark 3.15.** As in [36, (3.2)], one can also consider a resolvent operator associated with a convex function  $\psi$ :

$$J_\tau[x] := \operatorname{argmin}_{y \in X} \left\{ \phi(y) + \tau\psi\left(\frac{d(x, y)}{\tau}\right) \right\}$$

corresponding to the equation (2.16) in Remark 2.26. See [36, Lemma 3.2] for the existence of discrete solutions in this context.

In the rest of this subsection, we study some further properties of the Moreau–Yosida approximation. For  $x \in X$  and  $\tau > 0$  with  $J_\tau[x] \neq \emptyset$ , we set

$$d_\tau^+(x) := \sup_{y \in J_\tau[x]} d(x, y), \quad d_\tau^-(x) := \inf_{y \in J_\tau[x]} d(x, y).$$

We introduce the following assumption for convenience; note that it is stronger than Assumption 3.8(b).

**Assumption 3.16.** For any  $x \in X$  and  $\tau \in (0, \tau_*(\phi))$ ,  $J_\tau[x] \neq \emptyset$  holds.

**Remark 3.17.** By Theorem 3.14, if Assumption 3.8(a)–(c) hold, then Assumption 3.16 holds as well. See also Remark 3.9, Proposition 3.10 and Remark 4.7 for some situations where Assumption 3.8 holds.

We first discuss some continuity and monotonicity properties (cf. [1, Lemma 3.1.2]).

**Lemma 3.18.** *Suppose Assumption 3.16 in (ii)–(v) below.*

(i) *The function  $(\tau, x) \mapsto \Phi_\tau(x)$  is continuous in  $(0, \tau_*(\phi)) \times X$ .*

(ii) *For any  $x \in X$ ,  $0 < \tau_0 < \tau_1$  and  $y_i \in J_{\tau_i}[x]$  ( $i = 0, 1$ ), we have*

$$\phi(x) \geq \Phi_{\tau_0}(x) \geq \Phi_{\tau_1}(x), \quad d(x, y_0) \leq d(x, y_1), \quad \phi(x) \geq \phi(y_0) \geq \phi(y_1), \quad d_{\tau_0}^+(x) \leq d_{\tau_1}^-(x). \quad (3.4)$$

(iii) *If  $x \in \overline{\mathfrak{D}(\phi)}$ , then  $\lim_{\tau \rightarrow 0} d_\tau^+(x) = 0$ .*

(iv) *For any  $x \in X$ , there exists an at most countable set  $\mathcal{N}_x \subset (0, \tau_*(\phi))$  such that  $d_\tau^-(x) = d_\tau^+(x)$  for all  $\tau \in (0, \tau_*(\phi)) \setminus \mathcal{N}_x$ .*

(v) *If  $\phi$  is  $\mathcal{T}_+$ -lower semicontinuous, then we have, for all  $x \in \overline{\mathfrak{D}(\phi)}$ ,*

$$\lim_{\tau \rightarrow 0} \Phi_\tau(x) = \lim_{\tau \rightarrow 0} \inf_{y \in J_\tau[x]} \phi(y) = \phi(x).$$

Moreover,  $\lim_{\tau \rightarrow 0} \Phi_\tau(x) = \phi(x)$  holds for all  $x \in X$ .

*Proof.* (i) Take  $(\tau, x) \in (0, \tau_*(\phi)) \times X$  and a sequence  $((\tau_i, x_i))_{i \geq 1}$  in  $(0, \tau_*(\phi)) \times X$  converging to  $(\tau, x)$ . On the one hand, for any  $y \in X$ , we have

$$\limsup_{i \rightarrow \infty} \Phi_{\tau_i}(x_i) \leq \limsup_{i \rightarrow \infty} \Phi(\tau_i, x_i; y) = \Phi(\tau, x; y).$$

Taking the infimum in  $y \in X$  yields the upper semicontinuity  $\limsup_{i \rightarrow \infty} \Phi_{\tau_i}(x_i) \leq \Phi_\tau(x)$ . On the other hand, to see the lower semicontinuity, let  $(y_i)_{i \geq 1} \subset \mathfrak{D}(\phi)$  be a sequence such that

$$\lim_{i \rightarrow \infty} \{\Phi(\tau_i, x_i; y_i) - \Phi_{\tau_i}(x_i)\} = 0.$$

Since  $\sup_{i \geq 1} \Phi(\tau_i, x_i; y_i) < \infty$ , we find from (3.3) that  $D := \sup_{i \geq 1} d(x_i, y_i) < \infty$ . Thus, the triangle inequality implies  $\sup_{i \geq 1} d(x, y_i) < \infty$ . It follows from Lemma A.1 that

$$d^p(x, y_i) \leq \{d(x, x_i) + d(x_i, y_i)\}^p \leq (1 + \epsilon)d^p(x_i, y_i) + \mathfrak{C}(p, \epsilon)d^p(x, x_i)$$

for any  $\epsilon > 0$ . Hence, we have

$$\liminf_{i \rightarrow \infty} \Phi_{\tau_i}(x_i) = \liminf_{i \rightarrow \infty} \Phi(\tau_i, x_i; y_i) \geq \liminf_{i \rightarrow \infty} \left\{ \phi(y_i) + \frac{d^p(x, y_i)}{p\tau_i^{p-1}} \right\} - \frac{\epsilon D^p}{p\tau^{p-1}} \geq \Phi_\tau(x) - \frac{\epsilon D^p}{p\tau^{p-1}}.$$

Letting  $\epsilon \rightarrow 0$  furnishes the lower semicontinuity  $\Phi_\tau(x) \leq \liminf_{i \rightarrow \infty} \Phi_{\tau_i}(x_i)$ , which completes the proof.

(ii) The first claim is clear by the definition of  $\Phi_\tau(x)$ , and the second claim follows from

$$\begin{aligned} \phi(y_0) + \frac{d^p(x, y_0)}{p\tau_0^{p-1}} &= \Phi_{\tau_0}(x) \leq \Phi(\tau_0, x; y_1) = \Phi_{\tau_1}(x) + \left( \frac{1}{p\tau_0^{p-1}} - \frac{1}{p\tau_1^{p-1}} \right) d^p(x, y_1) \\ &\leq \phi(y_0) + \frac{d^p(x, y_0)}{p\tau_1^{p-1}} + \left( \frac{1}{p\tau_0^{p-1}} - \frac{1}{p\tau_1^{p-1}} \right) d^p(x, y_1). \end{aligned}$$

Note also that the fourth claim is an immediate consequence of the second claim. Finally, in the third claim, the first inequality is obvious and the second one is a consequence of the second claim as

$$\phi(y_1) + \frac{d^p(x, y_1)}{p\tau_1^{p-1}} \leq \phi(y_0) + \frac{d^p(x, y_0)}{p\tau_1^{p-1}} \leq \phi(y_0) + \frac{d^p(x, y_1)}{p\tau_1^{p-1}}.$$

(iii) For  $x \in \overline{\mathfrak{D}(\phi)}$ ,  $y_\tau \in J_\tau[x]$  and any  $y \in \mathfrak{D}(\phi)$ , we deduce from (3.2) that

$$\infty > \Phi(\tau, x; y) \geq \Phi(\tau, x; y_\tau) \geq \phi(y_\tau) \geq \Phi_\tau(y_\tau) \geq \Phi_{\tau_*}(x_*) - C(p, \tau_*, \tau) d^p(x_*, y_\tau). \quad (3.5)$$

Then, since  $\tau < \tau_*$ ,  $y_* \in J_{\tau_*}[x]$  satisfies  $d(x, y_\tau) \leq d(x, y_*)$  by the second claim in (ii) and we find

$$\Phi(\tau, x; y) \geq \phi(y_\tau) \geq \Phi_{\tau_*}(x_*) - C(p, \tau_*, \tau)(d(x_*, x) + d(x, y_*))^p.$$

Combining this with

$$\lim_{\tau \rightarrow 0} C(p, \tau_*, \tau) = \frac{1}{p\tau_*^{p-1}} < \infty \quad (3.6)$$

from Lemmas 3.13 and A.1, we find

$$\lim_{\tau \rightarrow 0} \tau^{p-1} \inf_{y_\tau \in J_\tau[x]} \phi(y_\tau) = 0. \quad (3.7)$$

Now, for any  $y \in \mathfrak{D}(\phi)$ , we have

$$d_\tau^+(x)^p = \sup_{y_\tau \in J_\tau[x]} p\tau^{p-1}(\Phi(\tau, x; y_\tau) - \phi(y_\tau)) \leq p\tau^{p-1}\phi(y) + d^p(x, y) - p\tau^{p-1} \inf_{y_\tau \in J_\tau[x]} \phi(y_\tau).$$

Then, (3.7) yields  $\limsup_{\tau \rightarrow 0} d_\tau^+(x)^p \leq d^p(x, y)$  for any  $y \in \mathfrak{D}(\phi)$ , and  $\lim_{\tau \rightarrow 0} d_\tau^+(x) = 0$  since  $x \in \overline{\mathfrak{D}(\phi)}$ .

(iv) Given  $x \in X$ , since the function  $\tau \mapsto d_\tau^-(x)$  is non-decreasing, it is continuous except for at most countably many points. Then, at any continuous point  $\tau \in (0, \tau_*(\phi))$ , the last claim in (ii) implies

$$d_\tau^+(x) \leq \lim_{t \rightarrow \tau^+} d_t^-(x) = d_\tau^-(x) \leq d_\tau^+(x).$$

Hence,  $d_\tau^+(x) = d_\tau^-(x)$ .

(v) Given  $x \in X$ , we deduce from (iv) that  $\phi$  is constant on  $J_\tau[x]$  for  $\tau \in (0, \tau_*(\phi)) \setminus \mathcal{N}_x$ . Moreover, thanks to the monotonicity as in (ii), it suffices to show the convergence within  $(0, \tau_*(\phi)) \setminus \mathcal{N}_x$ . Thus, we choose  $y_\tau \in J_\tau[x]$  for each  $\tau \in (0, \tau_*(\phi)) \setminus \mathcal{N}_x$  and consider the convergence of  $\phi(y_\tau)$ .

First, if  $x \in \mathfrak{D}(\phi)$ , then  $\lim_{\tau \rightarrow 0} d_\tau^+(x) = 0$  by (iii) and hence  $y_\tau \rightarrow x$  as  $\tau \rightarrow 0$ . Thus, on the one hand, the  $\mathcal{T}_+$ -lower semicontinuity of  $\phi$  yields

$$\liminf_{\tau \rightarrow 0} \Phi_\tau(x) = \liminf_{\tau \rightarrow 0} \left\{ \phi(y_\tau) + \frac{d^p(x, y_\tau)}{p\tau^{p-1}} \right\} \geq \liminf_{\tau \rightarrow 0} \phi(y_\tau) \geq \phi(x).$$

On the other hand, we have  $\phi(x) \geq \limsup_{\tau \rightarrow 0} \Phi_\tau(x) \geq \limsup_{\tau \rightarrow 0} \phi(y_\tau)$ . Combining these shows the first assertion.

Next, let  $x \notin \overline{\mathfrak{D}(\phi)}$ . On the one hand, clearly  $\liminf_{\tau \rightarrow 0} d(x, y_\tau) > 0$  holds since  $y_\tau \in \mathfrak{D}(\phi)$ . On the other hand, by (3.5), (3.6), the triangle inequality and  $d(x, y_\tau) \leq d(x, y_{\tau_*})$  from (ii), we find

$$\liminf_{\tau \rightarrow 0} \phi(y_\tau) \geq \Phi_{\tau_*}(x_*) - \frac{1}{p\tau_*^{p-1}}(d(x_*, x) + d(x, y_{\tau_*}))^p > -\infty.$$

This is enough to obtain the claim, indeed,

$$\liminf_{\tau \rightarrow 0} \Phi_\tau(x) \geq \liminf_{\tau \rightarrow 0} \phi(y_\tau) + \liminf_{\tau \rightarrow 0} \frac{d^p(x, y_\tau)}{p\tau^{p-1}} = \infty = \phi(x). \quad \square$$

Next we investigate the derivative of the Moreau–Yosida approximation (cf. [1, Theorem 3.1.4], [36, Lemma 4.5]).

**Theorem 3.19** (Derivative of  $\Phi_\tau(x)$ ). *Suppose Assumption 3.16. Then, for any  $x \in X$ , the function  $\tau \mapsto \Phi_\tau(x)$  is locally Lipschitz in  $(0, \tau_*(\phi))$  and*

$$\frac{d}{d\tau} [\Phi_\tau(x)] = -\frac{p-1}{p} \left( \frac{d_\tau^+(x)}{\tau} \right)^p = -\frac{p-1}{p} \left( \frac{d_\tau^-(x)}{\tau} \right)^p \quad (3.8)$$

holds for  $\mathcal{L}^1$ -a.e.  $\tau \in (0, \tau_*(\phi))$ . Moreover, if in addition  $\phi$  is  $\mathcal{T}_+$ -lower semicontinuous, then we have

$$\frac{d^p(x, y_\tau)}{p\tau^{p-1}} + \int_0^\tau \frac{p-1}{p} \left( \frac{d_r^\pm(x)}{r} \right)^p dr = \phi(x) - \phi(y_\tau) \quad (3.9)$$

for all  $\tau \in (0, \tau_*(\phi))$  and  $y_\tau \in J_\tau[x]$ .

*Proof.* For  $0 < \tau_0 < \tau_1 < \tau_*(\phi)$  and any  $y_{\tau_1} \in J_{\tau_1}[x]$ , we have

$$\Phi_{\tau_0}(x) - \Phi_{\tau_1}(x) \leq \Phi(\tau_0, x; y_{\tau_1}) - \Phi(\tau_1, x; y_{\tau_1}) = \frac{\tau_1^{p-1} - \tau_0^{p-1}}{p\tau_0^{p-1}\tau_1^{p-1}} d^p(x, y_{\tau_1}).$$

Note that

$$\tau_1^{p-1} - \tau_0^{p-1} \leq (p-1)\tau_i^{p-2}(\tau_1 - \tau_0) \quad \text{with } i = \begin{cases} 0 & \text{for } 1 < p \leq 2, \\ 1 & \text{for } 2 \leq p < \infty \end{cases}$$

by the concavity ( $1 < p \leq 2$ ) or the convexity ( $2 \leq p < \infty$ ) of  $t^{p-1}$  in  $t > 0$ . This implies

$$\Phi_{\tau_0}(x) - \Phi_{\tau_1}(x) \leq \frac{\tau_1^{p-1} - \tau_0^{p-1}}{p\tau_0^{p-1}\tau_1^{p-1}} (d_{\tau_1}^-(x))^p \leq \frac{(p-1)(\tau_1 - \tau_0)}{p\tau_{1-i}^{p-1}\tau_i} (d_{\tau_1}^-(x))^p.$$

We similarly observe

$$\Phi_{\tau_0}(x) - \Phi_{\tau_1}(x) \geq \frac{\tau_1^{p-1} - \tau_0^{p-1}}{p\tau_0^{p-1}\tau_1^{p-1}} (d_{\tau_0}^+(x))^p \geq \frac{(p-1)(\tau_1 - \tau_0)}{p\tau_{1-i}\tau_i^{p-1}} (d_{\tau_0}^+(x))^p.$$

Combining them furnishes

$$\frac{p-1}{p} \frac{(d_{\tau_0}^+(x))^p}{\tau_{1-i}\tau_i^{p-1}} \leq \frac{\Phi_{\tau_0}(x) - \Phi_{\tau_1}(x)}{\tau_1 - \tau_0} \leq \frac{p-1}{p} \frac{(d_{\tau_1}^-(x))^p}{\tau_{1-i}^{p-1}\tau_i},$$

which shows that  $\tau \mapsto \Phi_\tau(x)$  is locally Lipschitz in  $(0, \tau_*(\phi))$ .

Take  $\tau \in (0, \tau_*(\phi)) \setminus \mathcal{N}_x$  at where  $\Phi_\tau(x)$  is differentiable, with  $\mathcal{N}_x$  from Lemma 3.18(iv). Then, by applying the above estimates to  $\tau_1 \rightarrow \tau^+$  ( $\tau_0 = \tau$ ) and  $\tau_0 \rightarrow \tau^-$  ( $\tau_1 = \tau$ ), we have

$$\frac{p-1}{p} \frac{(d_\tau^+(x))^p}{\tau^p} \leq -\frac{d}{d\tau} [\Phi_\tau(x)] \leq \frac{p-1}{p} \frac{(d_\tau^-(x))^p}{\tau^p}.$$

Therefore, (3.8) follows. Integrating (3.8) on  $[\tau_0, \tau] \subset (0, \tau_*(\phi))$  yields

$$\Phi_\tau(x) + \int_{\tau_0}^{\tau} \frac{p-1}{p} \left( \frac{d_r^\pm(x)}{r} \right)^p dr = \Phi_{\tau_0}(x).$$

Letting  $\tau_0 \rightarrow 0$ , we deduce (3.9) from Lemma 3.18(v) and  $\Phi_\tau(x) = \Phi(\tau, x; y_\tau)$ .  $\square$

Now we estimate the local slope  $|\partial\phi|$  defined in Definition 2.19 (cf. [1, Lemmas 3.1.3, 3.1.5]). Let  $\mathfrak{D}(|\partial\phi|) := \{x \in \mathfrak{D}(\phi) \mid |\partial\phi|(x) < \infty\}$  be the proper effective domain of  $|\partial\phi|$ . Recall that  $q = p/(p-1)$ .

**Lemma 3.20.** *Suppose Assumption 3.16. Then we have, for any  $x \in X$ ,  $\tau \in (0, \tau_*(\phi))$  and  $y_\tau \in J_\tau[x]$ ,*

$$|\partial\phi|^q(y_\tau) \leq \frac{d^p(x, y_\tau)}{\tau^p}.$$

*In particular,  $y_\tau \in \mathfrak{D}(|\partial\phi|)$  holds and  $\mathfrak{D}(|\partial\phi|)$  is dense in  $\overline{\mathfrak{D}(\phi)}$  with respect to  $\mathcal{T}_+$ .*

*Proof.* Assume  $|\partial\phi|(y_\tau) > 0$  without loss of generality and take a sequence  $(y_i)_{i \geq 1}$  converging to  $y_\tau$  with

$$\lim_{i \rightarrow \infty} \frac{\phi(y_\tau) - \phi(y_i)}{d(y_\tau, y_i)} = |\partial\phi|(y_\tau).$$

Observe from the choice of  $y_\tau$  that

$$\phi(y_\tau) - \phi(y_i) \leq \frac{d^p(x, y_i)}{p\tau^{p-1}} - \frac{d^p(x, y_\tau)}{p\tau^{p-1}}.$$

Dividing this inequality by  $d(y_\tau, y_i)$  and using the mean value theorem, we obtain the first claim as

$$|\partial\phi|(y_\tau) \leq \frac{d^{p-1}(x, y_\tau)}{\tau^{p-1}} \limsup_{i \rightarrow \infty} \frac{d(x, y_i) - d(x, y_\tau)}{d(y_\tau, y_i)} \leq \frac{d^{p-1}(x, y_\tau)}{\tau^{p-1}}.$$

Thus  $y_\tau \in \mathfrak{D}(|\partial\phi|)$ , and the density of  $\mathfrak{D}(|\partial\phi|)$  in  $\overline{\mathfrak{D}(\phi)}$  follows from Lemma 3.18(iii).  $\square$

**Lemma 3.21.** *Suppose Assumption 3.16. Then we have*

$$\limsup_{\tau \rightarrow 0} \frac{\phi(x) - \Phi_\tau(x)}{\tau} = \frac{|\partial\phi|^q(x)}{q} \quad \text{for all } x \in \mathfrak{D}(\phi). \quad (3.10)$$

Moreover, there exists a sequence  $\tau_i \rightarrow 0$  such that, for  $y_\tau \in J_\tau[x]$ ,

$$|\partial\phi|^q(x) = \lim_{i \rightarrow \infty} \frac{\phi(x) - \phi(y_{\tau_i})}{\tau_i} = \lim_{i \rightarrow \infty} \frac{d^p(x, y_{\tau_i})}{\tau_i^p} \geq \liminf_{\tau \rightarrow 0} |\partial\phi|^q(y_\tau). \quad (3.11)$$

*Proof.* We first show (3.10). By the Young inequality, for any  $s > 0$ , we have

$$\frac{s^q}{q} = \sup_{t>0} \left( st - \frac{t^p}{p} \right) = \sup_{t>a} \left( st - \frac{t^p}{p} \right) \quad (3.12)$$

for all  $a \in (0, s^{1/(p-1)})$ . This implies

$$\begin{aligned} \limsup_{\tau \rightarrow 0} \frac{\phi(x) - \Phi_\tau(x)}{\tau} &= \limsup_{\tau \rightarrow 0} \left\{ \frac{\phi(x) - \phi(y_\tau)}{\tau} - \frac{d^p(x, y_\tau)}{p\tau^p} \right\} \\ &\leq \limsup_{\tau \rightarrow 0} \left\{ \frac{[\phi(x) - \phi(y_\tau)]_+}{d(x, y_\tau)} \frac{d(x, y_\tau)}{\tau} - \frac{d^p(x, y_\tau)}{p\tau^p} \right\} \\ &\leq \limsup_{\tau \rightarrow 0} \left\{ |\partial\phi|(x) \frac{d(x, y_\tau)}{\tau} - \frac{d^p(x, y_\tau)}{p\tau^p} \right\} \leq \frac{|\partial\phi|^q(x)}{q}, \end{aligned}$$

where we used Lemma 3.18(iii) in the second inequality. If  $|\partial\phi|(x) = 0$ , then we are done with the help of the first claim in (3.4). To show the reverse inequality under  $|\partial\phi|(x) > 0$ , we again use (3.12) to see

$$\begin{aligned} \frac{|\partial\phi|^q(x)}{q} &= \limsup_{y \rightarrow x} \frac{1}{q} \left( \frac{[\phi(x) - \phi(y)]_+}{d(x, y)} \right)^q \\ &= \limsup_{y \rightarrow x} \sup_{0 < \tau < d(x, y)/a} \left\{ \frac{[\phi(x) - \phi(y)]_+}{d(x, y)} \frac{d(x, y)}{\tau} - \frac{d^p(x, y)}{p\tau^p} \right\} \end{aligned}$$

for any  $a \in (0, |\partial\phi|^{1/(p-1)}(x))$ . Hence, for small  $\varepsilon > 0$ ,

$$\frac{|\partial\phi|^q(x)}{q} \leq \sup_{y \neq x} \sup_{0 < \tau < \varepsilon} \left\{ \frac{\phi(x) - \phi(y)}{\tau} - \frac{d^p(x, y)}{p\tau^p} \right\} = \sup_{0 < \tau < \varepsilon} \frac{\phi(x) - \Phi_\tau(x)}{\tau}.$$

Letting  $\varepsilon \rightarrow 0$  shows

$$\frac{|\partial\phi|^q(x)}{q} \leq \limsup_{\tau \rightarrow 0} \frac{\phi(x) - \Phi_\tau(x)}{\tau}$$

and completes the proof of (3.10).

We next prove (3.11). First of all, the last inequality in (3.11) immediately follows from Lemma 3.20. Now, on the one hand, (3.10) furnishes a sequence  $\tau_i \rightarrow 0$  such that

$$\lim_{i \rightarrow \infty} \frac{\phi(x) - \Phi_{\tau_i}(x)}{\tau_i} = \frac{|\partial\phi|^q(x)}{q}. \quad (3.13)$$

On the other hand, we find from the above argument that

$$\frac{|\partial\phi|^q(x)}{q} = \limsup_{\tau \rightarrow 0} \left\{ |\partial\phi|(x) \frac{d(x, y_\tau)}{\tau} - \frac{d^p(x, y_\tau)}{p\tau^p} \right\}.$$

Comparing this with the Young inequality (3.12), taking a subsequence if necessary, we obtain

$$|\partial\phi|^q(x) = \lim_{i \rightarrow \infty} \left( \frac{d(x, y_{\tau_i})}{\tau_i} \right)^p.$$

Finally, substituting this into (3.13) yields

$$|\partial\phi|^q(x) = q \lim_{i \rightarrow \infty} \left\{ \frac{\phi(x) - \phi(y_{\tau_i})}{\tau_i} - \frac{d^p(x, y_{\tau_i})}{p\tau_i^p} \right\} = q \lim_{i \rightarrow \infty} \frac{\phi(x) - \phi(y_{\tau_i})}{\tau_i} - (q-1)|\partial\phi|^q(x).$$

This completes the proof of (3.11).  $\square$

**3.4. A priori estimates for discrete solutions and a compactness result.** We introduce De Giorgi's variational interpolation associated with a solution to the recursive scheme (3.1) (cf. [1, Definition 3.2.1]).

**Definition 3.22** (Variational interpolation). Given a solution  $(\Xi_{\mathfrak{T}}^k)_{k \geq 0}$  to (3.1), we denote by  $\tilde{\Xi}_{\mathfrak{T}} : [0, \infty) \rightarrow X$  its arbitrary interpolation satisfying

$$\tilde{\Xi}_{\mathfrak{T}}(t_{\mathfrak{T}}^k) = \Xi_{\mathfrak{T}}^k, \quad \tilde{\Xi}_{\mathfrak{T}}(t_{\mathfrak{T}}^{k-1} + \delta) \in J_{\delta}[\Xi_{\mathfrak{T}}^{k-1}] \quad \text{for } \delta \in (0, t_{\mathfrak{T}}^k - t_{\mathfrak{T}}^{k-1}).$$

Define a function  $G_{\mathfrak{T}}$  on  $(0, \infty)$  associated with the discrete solution  $(\Xi_{\mathfrak{T}}^k)_{k \geq 0}$  as, for  $\delta \in (0, t_{\mathfrak{T}}^k - t_{\mathfrak{T}}^{k-1}]$ ,

$$G_{\mathfrak{T}}(t_{\mathfrak{T}}^{k-1} + \delta) := \frac{d_{\delta}^{+}(\Xi_{\mathfrak{T}}^{k-1})}{\delta} \geq \frac{d(\Xi_{\mathfrak{T}}^{k-1}, \tilde{\Xi}_{\mathfrak{T}}(t_{\mathfrak{T}}^{k-1} + \delta))}{\delta}.$$

Observe that  $G_{\mathfrak{T}}$  is a Borel function by (3.4). Moreover, we find from Lemma 3.20 that

$$|\partial\phi|^q(\tilde{\Xi}_{\mathfrak{T}}(t)) \leq G_{\mathfrak{T}}^p(t) \quad \text{for } t > 0. \quad (3.14)$$

We also define a piecewise constant function  $|\Xi'_{\mathfrak{T}}|$  on  $(0, \infty)$  by

$$|\Xi'_{\mathfrak{T}}|(t) := \frac{d(\Xi_{\mathfrak{T}}^{k-1}, \Xi_{\mathfrak{T}}^k)}{t_{\mathfrak{T}}^k - t_{\mathfrak{T}}^{k-1}} = \frac{d(\Xi_{\mathfrak{T}}^{k-1}, \Xi_{\mathfrak{T}}^k)}{\tau_k} \quad \text{for } t \in (t_{\mathfrak{T}}^{k-1}, t_{\mathfrak{T}}^k]. \quad (3.15)$$

Then we have the following a priori estimates (cf. [1, Lemma 3.2.2, Remark 3.2.5], [36, Proposition 4.7]). Recall Definition 3.4 for the piecewise constant interpolation  $\bar{\Xi}_{\mathfrak{T}}(t) = \Xi_{\mathfrak{T}}^k$  for  $t \in (t_{\mathfrak{T}}^{k-1}, t_{\mathfrak{T}}^k]$ , and Notation 2.5 that  $(X, \star, d)$  is always a pointed forward  $\Theta$ -metric space.

**Lemma 3.23** (A priori estimates). *Suppose that  $\phi$  is  $\mathcal{T}_+$ -lower semicontinuous and Assumption 3.16 holds, and let  $(\Xi_{\mathfrak{T}}^k)_{k \geq 0}$  be a solution to (3.1).*

(i) *If  $\|\mathfrak{T}\| \in (0, \tau_*(\phi))$ , then, for each couple of integers  $0 \leq k < l$ , we have*

$$\frac{1}{p} \int_{t_{\mathfrak{T}}^k}^{t_{\mathfrak{T}}^l} |\Xi'_{\mathfrak{T}}|^p(t) dt + \frac{1}{q} \int_{t_{\mathfrak{T}}^k}^{t_{\mathfrak{T}}^l} G_{\mathfrak{T}}^p(t) dt + \phi(\Xi_{\mathfrak{T}}^l) = \phi(\Xi_{\mathfrak{T}}^k).$$

(ii) *For any  $x_* \in X$  and  $S, T > 0$ , there exist positive constants  $C_1 = C_1(p, x_*, \tau_*(\phi), S, T)$  and  $C_2 = C_2(\Theta, p, x_*, \tau_*(\phi), S, T)$  such that, if*

$$\phi(\Xi_{\mathfrak{T}}^0) \leq S, \quad d^p(x_*, \Xi_{\mathfrak{T}}^0) \leq S, \quad t_{\mathfrak{T}}^{N-1} \leq T, \quad \|\mathfrak{T}\| \leq \frac{\tau_*(\phi)}{2^{p/(p-1)} p^2},$$

*then we have*

$$d^p(x_*, \Xi_{\mathfrak{T}}^k) \leq C_1, \quad \sum_{i=1}^k \frac{d^p(\Xi_{\mathfrak{T}}^{i-1}, \Xi_{\mathfrak{T}}^i)}{p\tau_i^{p-1}} \leq \phi(\Xi_{\mathfrak{T}}^0) - \phi(\Xi_{\mathfrak{T}}^k) \leq C_1 \quad \text{for all } 1 \leq k \leq N, \quad (3.16)$$

$$\max \left\{ d^p(\bar{\Xi}_{\mathfrak{T}}(t), \tilde{\Xi}_{\mathfrak{T}}(t)), d^p(\tilde{\Xi}_{\mathfrak{T}}(t), \bar{\Xi}_{\mathfrak{T}}(t)) \right\} \leq C_2 \|\mathfrak{T}\|^{p-1} \quad \text{for all } t \in [0, t_{\mathfrak{T}}^N]. \quad (3.17)$$

*Proof.* (i) It follows from (3.9) that, for each  $k+1 \leq i \leq l$ ,

$$\phi(\Xi_{\mathfrak{T}}^{i-1}) - \phi(\Xi_{\mathfrak{T}}^i) = \frac{d^p(\Xi_{\mathfrak{T}}^{i-1}, \Xi_{\mathfrak{T}}^i)}{p\tau_i^{p-1}} + \frac{1}{q} \int_{t_{\mathfrak{T}}^{i-1}}^{t_{\mathfrak{T}}^i} G_{\mathfrak{T}}^p(t) dt = \frac{1}{p} \int_{t_{\mathfrak{T}}^{i-1}}^{t_{\mathfrak{T}}^i} |\Xi'_{\mathfrak{T}}|^p(t) dt + \frac{1}{q} \int_{t_{\mathfrak{T}}^{i-1}}^{t_{\mathfrak{T}}^i} G_{\mathfrak{T}}^p(t) dt.$$

Summing up in  $k+1 \leq i \leq l$  shows the claim.

(ii) Observe from (i) that

$$\sum_{i=1}^k \frac{d^p(\Xi_{\mathfrak{T}}^{i-1}, \Xi_{\mathfrak{T}}^i)}{p\tau_i^{p-1}} = \frac{1}{p} \int_0^{t_{\mathfrak{T}}^k} |\Xi'_{\mathfrak{T}}|^p(t) dt \leq \phi(\Xi_{\mathfrak{T}}^0) - \phi(\Xi_{\mathfrak{T}}^k). \quad (3.18)$$

Note also that

$$\frac{d^p(x_*, \Xi_{\mathfrak{T}}^i)}{p} - \frac{d^p(x_*, \Xi_{\mathfrak{T}}^{i-1})}{p} \leq d(\Xi_{\mathfrak{T}}^{i-1}, \Xi_{\mathfrak{T}}^i) d^{p-1}(x_*, \Xi_{\mathfrak{T}}^i),$$

which follows from the convexity of the function  $t^p$  in  $t > 0$  and the triangle inequality if  $d(x_*, \Xi_{\mathfrak{I}}^i) > d(x_*, \Xi_{\mathfrak{I}}^{i-1})$  (and it is clear if  $d(x_*, \Xi_{\mathfrak{I}}^i) \leq d(x_*, \Xi_{\mathfrak{I}}^{i-1})$ ). We deduce from these inequalities and the Young inequality that, for any  $\varepsilon > 0$  and  $\tau_* := \tau_*(\phi)/p < \tau_*(\phi)$ ,

$$\begin{aligned} \frac{d^p(x_*, \Xi_{\mathfrak{I}}^k)}{p} - \frac{d^p(x_*, \Xi_{\mathfrak{I}}^0)}{p} &\leq \sum_{i=1}^k d(\Xi_{\mathfrak{I}}^{i-1}, \Xi_{\mathfrak{I}}^i) d^{p-1}(x_*, \Xi_{\mathfrak{I}}^i) \\ &\leq \sum_{i=1}^k \left\{ \varepsilon^{p-1} \frac{d^p(\Xi_{\mathfrak{I}}^{i-1}, \Xi_{\mathfrak{I}}^i)}{p\tau_i^{p-1}} + \frac{p-1}{p\varepsilon} \tau_i d^p(x_*, \Xi_{\mathfrak{I}}^i) \right\} \\ &\leq \varepsilon^{p-1} (\phi(\Xi_{\mathfrak{I}}^0) - \phi(\Xi_{\mathfrak{I}}^k)) + \frac{p-1}{p\varepsilon} \sum_{i=1}^k \tau_i d^p(x_*, \Xi_{\mathfrak{I}}^i) \\ &\leq \varepsilon^{p-1} (\phi(\Xi_{\mathfrak{I}}^0) - \Phi_{\tau_*}(x_*)) + \frac{\varepsilon^{p-1}}{p\tau_*^{p-1}} d^p(x_*, \Xi_{\mathfrak{I}}^k) + \frac{p-1}{p\varepsilon} \sum_{i=1}^k \tau_i d^p(x_*, \Xi_{\mathfrak{I}}^i). \end{aligned}$$

Substituting  $\varepsilon = 2^{-1/(p-1)}\tau_*$  and recalling our hypotheses, we obtain

$$\begin{aligned} d^p(x_*, \Xi_{\mathfrak{I}}^k) &\leq 2d^p(x_*, \Xi_{\mathfrak{I}}^0) + p\tau_*^{p-1} (\phi(\Xi_{\mathfrak{I}}^0) - \Phi_{\tau_*}(x_*)) + \frac{2^{p/(p-1)}(p-1)}{\tau_*} \sum_{i=1}^k \tau_i d^p(x_*, \Xi_{\mathfrak{I}}^i) \\ &\leq 2S + p\tau_*^{p-1}S - p\tau_*^{p-1}\Phi_{\tau_*}(x_*) + \frac{2^{p/(p-1)}(p-1)}{\tau_*} \sum_{i=1}^k \tau_i d^p(x_*, \Xi_{\mathfrak{I}}^i). \end{aligned}$$

Then it follows from [1, Lemma 3.2.4] with

$$a_i = d^p(x_*, \Xi_{\mathfrak{I}}^i), \quad A = 2S + p\tau_*^{p-1}S - p\tau_*^{p-1}\Phi_{\tau_*}(x_*), \quad \alpha = \frac{2^{p/(p-1)}(p-1)}{\tau_*}, \quad m = \alpha \|\mathfrak{I}\| \leq \frac{p-1}{p}$$

that

$$a_k \leq B e^{\beta t_{\mathfrak{I}}^{k-1}} \leq B e^{\beta T},$$

where  $B = A/(1-m)$  and  $\beta = \alpha/(1-m)$ . We remark that  $A \geq S > 0$  since

$$\Phi_{\tau_*}(x_*) \leq \phi(\Xi_{\mathfrak{I}}^0) + \frac{d^p(x_*, \Xi_{\mathfrak{I}}^0)}{p\tau_*^{p-1}} \leq S + \frac{S}{p\tau_*^{p-1}}.$$

Note also that  $\Phi_{\tau_*}(x_*)$  depends only on  $p$ ,  $\tau_*(\phi)$  and  $x_*$  (and  $\phi$ ). Thus, we obtain the first claim in (3.16). Moreover, we find from (3.18) that

$$\sum_{i=1}^k \frac{d^p(\Xi_{\mathfrak{I}}^{i-1}, \Xi_{\mathfrak{I}}^i)}{p\tau_i^{p-1}} \leq \phi(\Xi_{\mathfrak{I}}^0) - \phi(\Xi_{\mathfrak{I}}^k) \leq \phi(\Xi_{\mathfrak{I}}^0) - \Phi_{\tau_*}(x_*) + \frac{d^p(x_*, \Xi_{\mathfrak{I}}^k)}{p\tau_*^{p-1}},$$

which together with the first claim implies the second claim in (3.16).

As for (3.17), since  $\tilde{\Xi}_{\mathfrak{I}}(t) \in J_{t-t_{\mathfrak{I}}^{k-1}}[\Xi_{\mathfrak{I}}^{k-1}]$  for  $t \in (t_{\mathfrak{I}}^{k-1}, t_{\mathfrak{I}}^k]$  with  $1 \leq k \leq N$ , we see from the second claim in (3.4), the former claim in (3.16) and Definition 2.2 that

$$\begin{aligned} d(\bar{\Xi}_{\mathfrak{I}}(t), \tilde{\Xi}_{\mathfrak{I}}(t)) &\leq d(\Xi_{\mathfrak{I}}^k, \Xi_{\mathfrak{I}}^{k-1}) + d(\Xi_{\mathfrak{I}}^{k-1}, \tilde{\Xi}_{\mathfrak{I}}(t)) \leq d(\Xi_{\mathfrak{I}}^k, \Xi_{\mathfrak{I}}^{k-1}) + d(\Xi_{\mathfrak{I}}^{k-1}, \Xi_{\mathfrak{I}}^k) \\ &\leq (\Theta(d(\star, x_*) + C_1^{1/p}) + 1) d(\Xi_{\mathfrak{I}}^{k-1}, \Xi_{\mathfrak{I}}^k). \end{aligned}$$

Combining this with the latter claim in (3.16) furnishes

$$d^p(\bar{\Xi}_{\mathfrak{I}}(t), \tilde{\Xi}_{\mathfrak{I}}(t)) \leq (\Theta(d(\star, x_*) + C_1^{1/p}) + 1)^p p \|\mathfrak{I}\|^{p-1} C_1.$$

Furthermore, one can bound  $d^p(\tilde{\Xi}_{\mathfrak{T}}(t), \bar{\Xi}_{\mathfrak{T}}(t))$  by Definition 2.2, thanks to

$$\begin{aligned} d(\star, \bar{\Xi}_{\mathfrak{T}}(t)) &\leq d(\star, x_*) + d(x_*, \Xi_{\mathfrak{T}}^k) \leq d(\star, x_*) + C_1^{1/p}, \\ d(\star, \tilde{\Xi}_{\mathfrak{T}}(t)) &\leq d(\star, \bar{\Xi}_{\mathfrak{T}}(t)) + d(\bar{\Xi}_{\mathfrak{T}}(t), \tilde{\Xi}_{\mathfrak{T}}(t)) \\ &\leq d(\star, x_*) + C_1^{1/p} + (\Theta(d(\star, x_*) + C_1^{1/p}) + 1)(p\|\mathfrak{T}\|^{p-1}C_1)^{1/p} \end{aligned}$$

for  $t \in (t_{\mathfrak{T}}^{k-1}, t_{\mathfrak{T}}^k]$  with  $1 \leq k \leq N$ . This completes the proof.  $\square$

We shall need the following version of Ascoli–Arzelà theorem. Owing to Lemma 3.7, it is shown in the same way as [1, Proposition 3.3.1] and hence we omit the proof.

**Proposition 3.24.** *For  $T > 0$  and a  $\sigma$ -sequentially compact set  $K \subset X$ , let  $\xi_i : [0, T] \rightarrow K$ ,  $i \geq 1$ , be a sequence of (possibly discontinuous) curves such that*

$$\limsup_{i \rightarrow \infty} d(\xi_i(s), \xi_i(t)) \leq \omega(s, t) \quad \text{for all } s, t \in [0, T]$$

for a symmetric function  $\omega : [0, T] \times [0, T] \rightarrow [0, \infty)$  satisfying  $\lim_{(s,t) \rightarrow (r,r)} \omega(s, t) = 0$  for all  $r \in [0, T] \setminus \mathcal{C}$ , where  $\mathcal{C} \subset [0, T]$  is an at most countable set. Then there exist a subsequence  $(\xi_{i_j})_{j \geq 1}$  and a limit curve  $\xi : [0, T] \rightarrow X$  such that  $\xi_{i_j}(t) \xrightarrow{\sigma} \xi(t)$  for all  $t \in [0, T]$  and  $\xi$  is  $\mathcal{T}_+$ -continuous in  $[0, T] \setminus \mathcal{C}$ .

We remark that  $\omega(s, s) > 0$  can occur for  $s \in \mathcal{C}$ . For example, let  $\mu$  be a nonnegative finite measure on  $[0, T]$  and set  $\omega(s, t) = \omega(t, s) := \mu([s, t])$  for  $0 \leq s \leq t \leq T$ . Then  $\mathcal{C}$  is the set of atoms of  $\mu$  (see [1, Remark 3.3.2]). We also need the following variant of the local slope  $|\partial\phi|$ .

**Definition 3.25** (Relaxed slope). We define the *relaxed slope* of  $\phi$  by

$$|\partial^- \phi|(x) := \inf \left\{ \liminf_{i \rightarrow \infty} |\partial\phi|(x_i) \mid x_i \xrightarrow{\sigma} x, \sup_i \{d(x, x_i), \phi(x_i)\} < \infty \right\}$$

for  $x \in \mathfrak{D}(\phi)$ , and  $|\partial^- \phi|(x) := \infty$  otherwise.

Note that, by choosing  $x_i = x$  for all  $i$ ,  $|\partial^- \phi|(x) \leq |\partial\phi|(x)$  always holds.

**Remark 3.26.** If  $|\partial\phi|$  is  $\sigma$ -sequentially lower semicontinuous, then we have  $|\partial\phi|(x) \leq |\partial^- \phi|(x)$  for any  $x \in \mathfrak{D}(\phi)$ , and hence  $|\partial^- \phi| = |\partial\phi|$  holds.

Now, under Assumption 3.8, we have a compactness result (generalizing [1, Corollary 3.3.4]; see also [36, Proposition 4.9]).

**Theorem 3.27** (Compactness). *Suppose that Assumption 3.8(a)–(c) hold. Let  $\Lambda$  be a family of sequences of time steps  $\mathfrak{T}$  corresponding to partitions of  $[0, \infty)$  such that  $\inf_{\mathfrak{T} \in \Lambda} \|\mathfrak{T}\| = 0$ , and let  $\{\Xi_{\mathfrak{T}}^0\}_{\mathfrak{T} \in \Lambda}$  be a family of initial data satisfying*

$$\Xi_{\mathfrak{T}}^0 \xrightarrow{\sigma} x_0 \quad \text{and} \quad \phi(\Xi_{\mathfrak{T}}^0) \rightarrow \phi(x_0) \quad \text{as} \quad \|\mathfrak{T}\| \rightarrow 0, \quad \sup_{\mathfrak{T} \in \Lambda} d(x_0, \Xi_{\mathfrak{T}}^0) < \infty,$$

where  $x_0 \in \mathfrak{D}(\phi)$  is a fixed point. Then there exist a sequence  $(\mathfrak{T}_\alpha)_{\alpha \geq 1}$  in  $\Lambda$  with  $\|\mathfrak{T}_\alpha\| \rightarrow 0$ , a limit curve  $\xi \in \text{FAC}_{\text{loc}}^p([0, \infty); X)$ , a non-increasing function  $\varphi : [0, \infty) \rightarrow \mathbb{R}$ , and  $A \in L_{\text{loc}}^p([0, \infty))$  such that

- (i)  $\bar{\Xi}_{\mathfrak{T}_\alpha}(t) \xrightarrow{\sigma} \xi(t)$  and  $\tilde{\Xi}_{\mathfrak{T}_\alpha}(t) \xrightarrow{\sigma} \xi(t)$  as  $\alpha \rightarrow \infty$  for all  $t \geq 0$ ;
- (ii)  $\varphi(t) = \lim_{\alpha \rightarrow \infty} \phi(\bar{\Xi}_{\mathfrak{T}_\alpha}(t)) \geq \phi(\xi(t))$  for all  $t \geq 0$ , and  $\xi(0) = x_0$ ;
- (iii)  $|\Xi'_{\mathfrak{T}_\alpha}| \rightarrow A$  weakly in  $L_{\text{loc}}^p([0, \infty))$ , and  $A(t) \geq |\xi'_+(t)|$  for  $\mathcal{L}^1$ -a.e.  $t \in (0, \infty)$ ;
- (iv)  $\liminf_{\alpha \rightarrow \infty} G_{\mathfrak{T}_\alpha}^p(t) \geq |\partial^- \phi|^q(\xi(t))$  for all  $t > 0$ .

*Proof.* By hypotheses, we can choose a sequence  $(\mathfrak{T}_\alpha)_{\alpha \geq 1}$  in  $\Lambda$  such that  $\|\mathfrak{T}_\alpha\| \rightarrow 0$  and

$$\|\mathfrak{T}_\alpha\| \leq \frac{\tau_*(\phi)}{2^{p/(p-1)}p^2}, \quad |\phi(\Xi_{\mathfrak{T}_\alpha}^0) - \phi(x_0)| < 1.$$

In order to apply Lemma 3.23 (with the help of Remark 3.17), set  $S := \max\{\phi(x_0) + 1, \sup_{\mathfrak{T} \in \Lambda} d^p(x_0, \Xi_{\mathfrak{T}}^0)\}$ . Then, for  $T > 0$ ,  $N(\alpha)$  with  $t_{\mathfrak{T}_\alpha}^{N(\alpha)-1} \leq T < t_{\mathfrak{T}_\alpha}^{N(\alpha)}$  and  $C := \max\{C_1, C_2\}$  in Lemma 3.23, we have

$$d^p(x_0, \Xi_{\mathfrak{T}_\alpha}^k) \leq C, \quad \sum_{i=1}^k \frac{d^p(\Xi_{\mathfrak{T}_\alpha}^{i-1}, \Xi_{\mathfrak{T}_\alpha}^i)}{p\tau_{\alpha,i}^{p-1}} \leq \phi(\Xi_{\mathfrak{T}_\alpha}^0) - \phi(\Xi_{\mathfrak{T}_\alpha}^k) \leq C \quad \text{for } 1 \leq k \leq N(\alpha), \quad (3.19)$$

$$\max\{d^p(\bar{\Xi}_{\mathfrak{T}_\alpha}(t), \tilde{\Xi}_{\mathfrak{T}_\alpha}(t)), d^p(\tilde{\Xi}_{\mathfrak{T}_\alpha}(t), \bar{\Xi}_{\mathfrak{T}_\alpha}(t))\} \leq C\|\mathfrak{T}_\alpha\|^{p-1} \quad \text{for } t \in [0, t_{\mathfrak{T}_\alpha}^{N(\alpha)}], \quad (3.20)$$

where we put  $\mathfrak{T}_\alpha = (\tau_{\alpha,i})_{i \geq 1}$ . Note that (3.19) implies

$$\sup_{\alpha} \sup_{t \in [0, T]} d^p(x_0, \bar{\Xi}_{\mathfrak{T}_\alpha}(t)) \leq C, \quad \sup_{\alpha} \sup_{t \in [0, T]} \phi(\bar{\Xi}_{\mathfrak{T}_\alpha}(t)) \leq \sup_{\alpha} \phi(\Xi_{\mathfrak{T}_\alpha}^0) \leq S.$$

Now, consider a set

$$\mathcal{H} := \{x \in X \mid d^p(x_0, x) \leq C, \phi(x) \leq S\},$$

and observe  $\bar{\Xi}_{\mathfrak{T}_\alpha}([0, T]) \subset \mathcal{H}$ . It follows from Assumption 3.8(a), (c) that  $\mathcal{H}$  is  $\sigma$ -sequentially compact.

Note that  $\varphi_\alpha(t) := \phi(\bar{\Xi}_{\mathfrak{T}_\alpha}(t))$  is non-increasing by definition. Therefore, by passing to a subsequence if necessary, we may assume that  $\varphi(t) := \lim_{\alpha \rightarrow \infty} \varphi_\alpha(t)$  exists for all  $t \in [0, T]$  and  $\varphi$  is a non-increasing function (see [1, Lemma 3.3.3]). Moreover, by a diagonal argument, one can assume that  $\varphi(t) := \lim_{\alpha \rightarrow \infty} \varphi_\alpha(t)$  exists for all  $t \geq 0$ . We also find from Lemma 3.23(i) and (3.19) that

$$\int_0^T |\Xi'_{\mathfrak{T}_\alpha}|^p(t) dt \leq p\{\phi(\bar{\Xi}_{\mathfrak{T}_\alpha}(0)) - \phi(\bar{\Xi}_{\mathfrak{T}_\alpha}(T))\} \leq pC,$$

which together with the reflexivity of  $L^p([0, T])$  yields a weakly convergent subsequence of  $(|\Xi'_{\mathfrak{T}_\alpha}|)_{\alpha \geq 1}$  in  $L^p([0, T])$ . Again by a diagonal argument, we can assume that  $(|\Xi'_{\mathfrak{T}_\alpha}|)_{\alpha \geq 1}$  converges weakly to some function  $A \in L^p_{\text{loc}}([0, \infty))$ .

Let  $P_{\mathfrak{T}_\alpha}$  be the partition of  $[0, \infty)$  corresponding to  $\mathfrak{T}_\alpha$ . For  $0 \leq s < t$ , set  $s_\alpha := \max\{r \in P_{\mathfrak{T}_\alpha} \mid r \leq s\}$  and  $t_\alpha := \min\{r \in P_{\mathfrak{T}_\alpha} \mid t \leq r\}$ . Then we have

$$d(\bar{\Xi}_{\mathfrak{T}_\alpha}(s), \bar{\Xi}_{\mathfrak{T}_\alpha}(t)) \leq \int_{s_\alpha}^{t_\alpha} |\Xi'_{\mathfrak{T}_\alpha}|(r) dr$$

by the definition (3.15) of  $|\Xi'_{\mathfrak{T}_\alpha}|$ , and hence

$$\limsup_{\alpha \rightarrow \infty} d(\bar{\Xi}_{\mathfrak{T}_\alpha}(s), \bar{\Xi}_{\mathfrak{T}_\alpha}(t)) \leq \int_s^t A(r) dr. \quad (3.21)$$

Therefore, since  $\mathcal{H} \subset \overline{B_\star^+(d(\star, x_0) + C^{1/p})}$ , we can apply Proposition 3.24 to  $\bar{\Xi}_{\mathfrak{T}_\alpha} : [0, T] \rightarrow \mathcal{H}$  with  $\omega(s, t) = \omega(t, s) := \Theta(d(\star, x_0) + C^{1/p}) \int_s^t A(r) dr$  to find a subsequence of  $(\bar{\Xi}_{\mathfrak{T}_\alpha})_{\alpha \geq 1}$   $\sigma$ -converging to a  $\mathcal{T}_+$ -continuous curve  $\xi : [0, T] \rightarrow \mathcal{H}$ . Together with (3.20) and by a diagonal argument, we obtain (i).

It is straightforward from Assumption 3.8(a) and (i) that

$$\varphi(t) = \lim_{\alpha \rightarrow \infty} \phi(\bar{\Xi}_{\mathfrak{T}_\alpha}(t)) \geq \phi(\xi(t)).$$

Moreover, the assumption  $\Xi_{\mathfrak{T}_\alpha}^0 \xrightarrow{\sigma} x_0$  combined with (i) furnishes (as in Remark 3.5(c))

$$d(\xi(0), x_0) \leq \liminf_{\alpha \rightarrow \infty} d(\bar{\Xi}_{\mathfrak{T}_\alpha}(0), \Xi_{\mathfrak{T}_\alpha}^0) = 0,$$

which shows  $\xi(0) = x_0$  and completes the proof of (ii). Next, we deduce from (i) and (3.21) that

$$d(\xi(s), \xi(t)) \leq \liminf_{\alpha \rightarrow \infty} d(\bar{\Xi}_{\mathfrak{T}_\alpha}(s), \bar{\Xi}_{\mathfrak{T}_\alpha}(t)) \leq \limsup_{\alpha \rightarrow \infty} d(\bar{\Xi}_{\mathfrak{T}_\alpha}(s), \bar{\Xi}_{\mathfrak{T}_\alpha}(t)) \leq \int_s^t A(r) dr,$$

which implies that  $|\xi'_+|(t) \leq A(t)$  holds for  $\mathcal{L}^1$ -a.e.  $t \in (0, \infty)$ . Thus we have (iii) and, since  $A \in L^p_{\text{loc}}([0, \infty))$ ,  $\xi \in \text{FAC}^p_{\text{loc}}([0, \infty); X)$ . Moreover, it follows from (i) and (3.14) that, for all  $t > 0$ ,

$$|\partial^- \phi|^q(\xi(t)) \leq \liminf_{\alpha \rightarrow \infty} |\partial \phi|^q(\bar{\Xi}_{\mathfrak{T}_\alpha}(t)) \leq \liminf_{\alpha \rightarrow \infty} G_{\mathfrak{T}_\alpha}^p(t).$$

This yields (iv) and completes the proof.  $\square$

**3.5. Generalized minimizing movements.** In this subsection, we will take advantage of generalized minimizing movements to obtain the existence of curves of maximal slope (recall Definition 2.23).

**Definition 3.28** (Generalized minimizing movements). Let  $p \in (1, \infty)$  and  $x_0 \in X$ . We say that a curve  $\xi : [0, \infty) \rightarrow X$  is a  $p$ -generalized minimizing movement for  $\phi$  starting from  $x_0$  if  $\xi(0) = x_0$  and there is a sequence  $(\mathfrak{T}_\alpha)_{\alpha \geq 1}$  of  $\|\mathfrak{T}_\alpha\| \rightarrow 0$  along with corresponding discrete solutions  $(\bar{\Xi}_{\mathfrak{T}_\alpha})_{\alpha \geq 1}$  such that

$$\lim_{\alpha \rightarrow \infty} \phi(\Xi_{\mathfrak{T}_\alpha}^0) = \phi(x_0), \quad \limsup_{\alpha \rightarrow \infty} d(x_0, \Xi_{\mathfrak{T}_\alpha}^0) < \infty, \quad \bar{\Xi}_{\mathfrak{T}_\alpha}(t) \xrightarrow{\sigma} \xi(t) \text{ for all } t \geq 0.$$

We denote by  $\text{GMM}_p(\phi; x_0)$  the collection of  $p$ -generalized minimizing movements for  $\phi$  starting from  $x_0$ .

We remark that the construction of discrete solutions depends on  $p$ . Note also that, under  $\Xi_{\mathfrak{T}_\alpha}^0 \xrightarrow{\sigma} \xi(0)$ ,  $\xi(0) = x_0$  is equivalent to  $\Xi_{\mathfrak{T}_\alpha}^0 \xrightarrow{\sigma} x_0$  (see Remark 3.5(c)). The following variant of Theorem 3.27 ensures that  $\text{GMM}_p(\phi; x_0)$  is nonempty under mild assumptions.

**Corollary 3.29.** *Suppose that Assumption 3.8(a)–(c) hold. Let  $\Lambda$  be a family of sequences of time steps  $\mathfrak{T}$  corresponding to partitions of  $[0, \infty)$  with  $\inf_{\mathfrak{T} \in \Lambda} \|\mathfrak{T}\| = 0$ . If a family of initial data  $\{\Xi_{\mathfrak{T}}^0\}_{\mathfrak{T} \in \Lambda}$  satisfies*

$$\sup_{\mathfrak{T} \in \Lambda} \phi(\Xi_{\mathfrak{T}}^0) < \infty, \quad \sup_{\mathfrak{T} \in \Lambda} d(x_0, \Xi_{\mathfrak{T}}^0) < \infty,$$

*then there exist a sequence  $(\mathfrak{T}_\alpha)_{\alpha \geq 1}$  in  $\Lambda$  with  $\|\mathfrak{T}_\alpha\| \rightarrow 0$  and  $\xi \in \text{FAC}_{\text{loc}}^p([0, \infty); X)$  such that  $\bar{\Xi}_{\mathfrak{T}_\alpha}(t) \xrightarrow{\sigma} \xi(t)$  for all  $t \geq 0$ . Moreover, if  $\Xi_{\mathfrak{T}_\alpha}^0 \xrightarrow{\sigma} x_0$  and  $\phi(\Xi_{\mathfrak{T}_\alpha}^0) \rightarrow \phi(x_0)$  as  $\alpha \rightarrow \infty$ , then we have  $\xi \in \text{GMM}_p(\phi; x_0)$ .*

*Proof.* Since  $\inf_{\mathfrak{T} \in \Lambda} \|\mathfrak{T}\| = 0$ , we can choose a sequence  $(\mathfrak{T}_\alpha)_{\alpha \geq 1}$  with  $\|\mathfrak{T}_\alpha\| \rightarrow 0$ . By letting  $S := \max\{\sup_{\mathfrak{T} \in \Lambda} \phi(\Xi_{\mathfrak{T}}^0), \sup_{\mathfrak{T} \in \Lambda} d^p(x_0, \Xi_{\mathfrak{T}}^0)\}$ , the existence of  $\xi$  is shown in the same way as in Theorem 3.27. Moreover, if  $\Xi_{\mathfrak{T}_\alpha}^0 \xrightarrow{\sigma} x_0$ , then we find  $\xi(0) = x_0$  and  $\phi(\Xi_{\mathfrak{T}_\alpha}^0) \rightarrow \phi(x_0)$  implies  $\xi \in \text{GMM}_p(\phi; x_0)$ .  $\square$

We remark that, in Theorem 3.27, we assumed  $x_0 \in \mathfrak{D}(\phi)$  to obtain an upper bound of  $\sup_{\mathfrak{T} \in \Lambda} \phi(\Xi_{\mathfrak{T}}^0)$  and apply Lemma 3.23(ii). In Corollary 3.29, however, we assumed  $\sup_{\mathfrak{T} \in \Lambda} \phi(\Xi_{\mathfrak{T}}^0) < \infty$  instead and  $x_0 \in \mathfrak{D}(\phi)$  was removed. Note also that  $\sup_{\mathfrak{T} \in \Lambda} \phi(\Xi_{\mathfrak{T}}^0) < \infty$  and  $\phi(\Xi_{\mathfrak{T}_\alpha}^0) \rightarrow \phi(x_0)$  imply  $x_0 \in \mathfrak{D}(\phi)$ .

In the next two theorems (generalizing [1, Theorems 2.3.3, 2.3.1]), we see that  $\text{GMM}_p(\phi; x_0)$  consists of  $p$ -curves of maximal slope. Recall (2.13) for the energy identity of  $p$ -curves of maximal slope.

**Theorem 3.30.** *Suppose that Assumptions 3.8(a) and 3.16 hold and  $|\partial^- \phi|$  is a strong upper gradient for  $-\phi$ . Let  $x_0 \in \mathfrak{D}(\phi)$  and  $(\bar{\Xi}_{\mathfrak{T}_\alpha})_{\alpha \geq 1}$  be a sequence of discrete solutions with a curve  $\xi : [0, \infty) \rightarrow X$  such that  $\xi(0) = x_0$  and*

$$\|\mathfrak{T}_\alpha\| \rightarrow 0, \quad \phi(\Xi_{\mathfrak{T}_\alpha}^0) \rightarrow \phi(x_0), \quad \sup_{\alpha} d(x_0, \Xi_{\mathfrak{T}_\alpha}^0) < \infty, \quad \bar{\Xi}_{\mathfrak{T}_\alpha}(t) \xrightarrow{\sigma} \xi(t) \text{ for all } t \geq 0.$$

*Then we have  $\xi \in \text{FAC}_{\text{loc}}^p([0, \infty); X)$  and*

- (i)  $\lim_{\alpha \rightarrow \infty} \phi(\bar{\Xi}_{\mathfrak{T}_\alpha}(t)) = \phi(\xi(t))$  for all  $t \geq 0$ ;
- (ii)  $|\Xi'_{\mathfrak{T}_\alpha}| \rightarrow |\xi'_+|$  in  $L^p_{\text{loc}}([0, \infty))$ ;
- (iii)  $|\partial \phi| \circ \bar{\Xi}_{\mathfrak{T}_\alpha} \rightarrow |\partial^- \phi| \circ \xi$  in  $L^q_{\text{loc}}([0, \infty))$ .

*In particular, every curve  $\xi \in \text{GMM}_p(\phi; x_0)$  is a  $p$ -curve of maximal slope for  $\phi$  with respect to  $|\partial^- \phi|$  and we have the energy identity*

$$\frac{1}{p} \int_0^T |\xi'_+|^p(t) dt + \frac{1}{q} \int_0^T |\partial^- \phi|^q(\xi(t)) dt + \phi(\xi(T)) = \phi(x_0) \quad \text{for all } T > 0. \quad (3.22)$$

*Proof.* We obtain from Theorem 3.27 a subsequence  $(\mathfrak{T}_{\alpha_i})_{i \geq 1}$  of  $(\mathfrak{T}_\alpha)_{\alpha \geq 1}$ , denoted by  $(\mathfrak{T}_i)_{i \geq 1}$  for simplicity, satisfying Theorem 3.27(i)–(iv). We remark that, in the proof of Theorem 3.27, Assumption 3.8(c) was used only to find a convergent subsequence and a limit curve  $\xi$  satisfying Theorem 3.27(i). In the current theorem,  $\bar{\Xi}_{\mathfrak{T}_\alpha}(t) \xrightarrow{\sigma} \xi(t)$  is included in the hypotheses and we do not need Assumption 3.8(c).

It follows from Theorem 3.27, Fatou's lemma and Lemma 3.23(i) that

$$\begin{aligned}
& \frac{1}{p} \int_0^T |\xi'_+|^p(t) dt + \frac{1}{q} \int_0^T |\partial^- \phi|^q(\xi(t)) dt + \phi(\xi(T)) \\
& \leq \frac{1}{p} \int_0^T A^p(t) dt + \frac{1}{q} \int_0^T \liminf_{i \rightarrow \infty} G_{\mathfrak{X}_i}^p(t) dt + \lim_{i \rightarrow \infty} \phi(\bar{\Xi}_{\mathfrak{X}_i}(T)) \\
& \leq \liminf_{i \rightarrow \infty} \left\{ \frac{1}{p} \int_0^T |\Xi'_{\mathfrak{X}_i}|^p(t) dt + \frac{1}{q} \int_0^T G_{\mathfrak{X}_i}^p(t) dt + \phi(\bar{\Xi}_{\mathfrak{X}_i}(T)) \right\} \\
& = \lim_{i \rightarrow \infty} \phi(\Xi_{\mathfrak{X}_i}^0) = \phi(x_0).
\end{aligned} \tag{3.23}$$

Furthermore, since  $|\partial^- \phi|$  is a strong upper gradient for  $-\phi$  by assumption, Definition 2.15 furnishes

$$\phi(x_0) = \phi(\xi(0)) \leq \phi(\xi(T)) + \int_0^T |\partial^- \phi|(\xi(t)) |\xi'_+(t)| dt. \tag{3.24}$$

Combining this with (3.23) shows

$$\frac{1}{p} \int_0^T |\xi'_+|^p(t) dt + \frac{1}{q} \int_0^T |\partial^- \phi|^q(\xi(t)) dt \leq \int_0^T |\partial^- \phi|(\xi(t)) |\xi'_+(t)| dt.$$

Therefore, we obtain from the equality condition of the Young inequality that

$$|\xi'_+|^p(t) = |\partial^- \phi|^q(\xi(t)) \quad \text{for } \mathcal{L}^1\text{-a.e. } t \in (0, \infty), \tag{3.25}$$

and all the inequalities in (3.23) and (3.24) are in fact equalities. In particular, (3.23) yields (3.22) and

$$\begin{aligned}
& |\Xi'_{\mathfrak{X}_i}| \rightarrow |\xi'_+| \text{ in } L^p_{\text{loc}}([0, \infty)), \\
& \liminf_{i \rightarrow \infty} G_{\mathfrak{X}_i}^p = |\partial^- \phi|^q \circ \xi \text{ in } L^1_{\text{loc}}([0, \infty)), \quad \lim_{i \rightarrow \infty} \phi(\bar{\Xi}_{\mathfrak{X}_i}(t)) = \phi(\xi(t)) \text{ for all } t \geq 0.
\end{aligned} \tag{3.26}$$

Thus, we have (i) and (ii) for  $(\mathfrak{X}_i)_{i \geq 1}$ .

We next show (iii). On the one hand, observe from Definition 3.25 that  $\liminf_{i \rightarrow \infty} |\partial \phi|(\bar{\Xi}_{\mathfrak{X}_i}(t)) \geq |\partial^- \phi|(\xi(t))$ . On the other hand, we deduce from (3.25), (3.26), (3.15) and Lemma 3.20 that

$$\begin{aligned}
\int_0^T |\partial^- \phi|^q(\xi(t)) dt &= \int_0^T |\xi'_+|^p dt = \lim_{i \rightarrow \infty} \int_0^T |\Xi'_{\mathfrak{X}_i}|^p dt \geq \limsup_{i \rightarrow \infty} \int_0^T |\partial \phi|^q(\bar{\Xi}_{\mathfrak{X}_i}(t)) dt \\
&\geq \liminf_{i \rightarrow \infty} \int_0^T |\partial \phi|^q(\bar{\Xi}_{\mathfrak{X}_i}(t)) dt \geq \int_0^T \liminf_{i \rightarrow \infty} |\partial \phi|^q(\bar{\Xi}_{\mathfrak{X}_i}(t)) dt.
\end{aligned}$$

Hence, we obtain (iii) for  $(\mathfrak{X}_i)_{i \geq 1}$ . Note finally that, since every subsequence of  $(\mathfrak{X}_\alpha)_{\alpha \geq 1}$  includes a further subsequence satisfying (i)–(iii), the original sequence  $(\mathfrak{X}_\alpha)_{\alpha \geq 1}$  necessarily satisfies (i)–(iii).

For  $\xi \in \text{GMM}_p(\phi; x_0)$ , one can apply the above argument and obtain the energy identity (3.22). Moreover, we deduce from the equality in (3.24) as well as (3.25) that

$$\frac{d}{dt} \phi(\xi(t)) = -|\partial^- \phi|(\xi(t)) |\xi'_+(t)| = -\frac{1}{p} |\xi'_+|^p(t) - \frac{1}{q} |\partial^- \phi|^q(\xi(t))$$

for  $\mathcal{L}^1$ -a.e.  $t \in (0, \infty)$ . Therefore,  $\xi$  is a  $p$ -curve of maximal slope for  $\phi$  with respect to  $|\partial^- \phi|$ .  $\square$

See [36, Theorem 3.5] and [6, Theorem 4.21] for the results corresponding to Theorem 3.30 in their settings. Theorem 3.30 cannot be deduced from them (and vice versa), due to the difference of assumptions we explained in Remarks 2.26 and 3.11.

**Theorem 3.31.** *Suppose that Assumptions 3.8(a) and 3.16 hold. If  $|\partial^- \phi|$  is a weak upper gradient for  $-\phi$  and  $\phi$  satisfies the continuity condition:*

$$\sup_{i \geq 1} \{ |\partial \phi|(x_i), d(x, x_i), \phi(x_i) \} < \infty, \quad x_i \xrightarrow{\sigma} x \implies \phi(x_i) \rightarrow \phi(x),$$

then every  $\xi \in \text{GMM}_p(\phi; x_0)$  with  $x_0 \in \mathfrak{D}(\phi)$  is a  $p$ -curve of maximal slope for  $\phi$  with respect to  $|\partial^- \phi|$ .

*Proof.* Let  $(\mathfrak{T}_\alpha)_{\alpha \geq 1}$  be as in Definition 3.28. Similarly to Theorem 3.30, we can choose a subsequence  $(\mathfrak{T}_i)_{i \geq 1}$  of  $(\mathfrak{T}_\alpha)_{\alpha \geq 1}$  satisfying Theorem 3.27(i)–(iv). Setting  $\varphi(t) = \lim_{i \rightarrow \infty} \phi(\bar{\Xi}_{\mathfrak{T}_i}(t))$  as in Theorem 3.27, we infer from (3.23) and Lemma 3.23(i) that

$$\begin{aligned} \frac{1}{p} \int_s^t |\xi'_+|^p(r) dr + \frac{1}{q} \int_s^t |\partial^- \phi|^q(\xi(r)) dr &\leq \varphi(s) - \varphi(t) \quad \text{for all } 0 \leq s \leq t, \\ \frac{1}{p} \int_s^t \liminf_{i \rightarrow \infty} |\Xi'_{\mathfrak{T}_i}|^p(r) dr &\leq \varphi(s) - \varphi(t) < \infty \quad \text{for all } 0 \leq s \leq t. \end{aligned} \quad (3.27)$$

The latter inequality implies that  $\liminf_{i \rightarrow \infty} |\Xi'_{\mathfrak{T}_i}|(t) < \infty$  for  $\mathcal{L}^1$ -a.e.  $t \in (0, \infty)$ . Since  $\varphi$  is non-increasing,  $\varphi'$  exists  $\mathcal{L}^1$ -a.e. and we find from (3.27) that

$$\varphi'(t) \leq -\frac{1}{p} |\xi'_+|^p(t) - \frac{1}{q} |\partial^- \phi|^q(\xi(t)) \quad \text{for } \mathcal{L}^1\text{-a.e. } t \in (0, \infty).$$

As  $|\partial^- \phi|$  is a weak upper gradient by assumption, it suffices to show that  $\varphi$  is  $\mathcal{L}^1$ -a.e. equal to  $\phi \circ \xi$ .

Observe from Lemma 3.20 (as in the proof of Theorem 3.30) that

$$\liminf_{i \rightarrow \infty} |\partial \phi|^q(\bar{\Xi}_{\mathfrak{T}_i}(t)) \leq \liminf_{i \rightarrow \infty} |\Xi'_{\mathfrak{T}_i}|^p(t) < \infty \quad \text{for } \mathcal{L}^1\text{-a.e. } t \in (0, \infty).$$

Thus, for  $\mathcal{L}^1$ -a.e.  $t \in (0, \infty)$ , the assumed continuity condition yields  $\varphi(t) = \lim_{i \rightarrow \infty} \phi(\bar{\Xi}_{\mathfrak{T}_i}(t)) = \phi(\xi(t))$ . This completes the proof.  $\square$

Now, we present an existence result to Problem 3.1. We say that a function  $\phi : X \rightarrow \mathbb{R}$  is of *lower  $p$ -growth* if there are constants  $C, D \geq 0$  such that  $\phi(x) \geq -C - Dd^p(\star, x)$  holds for all  $x \in X$ . For instance, if  $\inf_X \phi > -\infty$ , then  $\phi$  is of lower  $p$ -growth with  $C = -\min\{\inf_X \phi, 0\}$ .

**Corollary 3.32.** *Let  $(X, d)$  be a forward boundedly compact forward metric space,  $p \in (1, \infty)$ , and  $\phi : X \rightarrow \mathbb{R}$  be a continuous function of lower  $p$ -growth. If  $|\partial^- \phi|$  is a weak upper gradient for  $-\phi$ , then, for any  $x_0 \in X$ , there exists a  $p$ -curve  $\xi : [0, \infty) \rightarrow X$  of maximal slope for  $\phi$  with respect to  $|\partial^- \phi|$  with  $\xi(0) = x_0$ . If additionally  $|\partial^- \phi|$  is a strong upper gradient, then  $\xi$  satisfies the energy identity (3.22).*

*Proof.* Let  $\sigma = \mathcal{T}_+$ . Observe from the triangle inequality and Lemma A.1 that

$$\Phi(\tau, x; y) \geq -C - Dd^p(\star, y) + \frac{d^p(x, y)}{p\tau^{p-1}} \geq -C - D\{\mathfrak{C}d^p(\star, x) + (1 + \epsilon)d^p(x, y)\} + \frac{d^p(x, y)}{p\tau^{p-1}}.$$

Hence, for sufficiently small  $\tau > 0$ , we have  $\inf_{y \in X} \Phi(\tau, x; y) > -\infty$  and Assumption 3.8(b) holds. Moreover, Assumption 3.8(a), (c) hold by hypotheses. Therefore, given any sequence of time steps  $(\mathfrak{T}_\alpha)_{\alpha \geq 1}$  with initial data  $(\Xi_{\mathfrak{T}_\alpha}^0)_{\alpha \geq 1}$  such that  $\|\mathfrak{T}_\alpha\| \rightarrow 0$  and  $\Xi_{\mathfrak{T}_\alpha}^0 \rightarrow x_0$ , Corollary 3.29 provides a limit curve  $\xi \in \text{GMM}_p(\phi; x_0)$ . Then the claims follow from Theorems 3.31 and 3.30.  $\square$

We remark that the lower  $p$ -growth was assumed merely for ensuring that Assumption 3.8(b) holds. Another condition implying Assumption 3.8(b) is, e.g., that every sublevel set of  $\phi$  is compact (see Proposition 3.10).

In view of Example 2.25, we also give an existence result for gradient curves in the Finsler case.

**Corollary 3.33.** *Let  $(M, F)$  be a forward complete Finsler manifold and  $p \in (1, \infty)$ . For any  $\phi \in C^1(M)$  and  $x_0 \in M$ , there exists a  $C^1$ -curve  $\xi : [0, T) \rightarrow M$  solving the gradient flow equation*

$$j_p(\xi'(t)) = \nabla(-\phi)(\xi(t)), \quad \xi(0) = x_0,$$

where  $\lim_{t \rightarrow T} d_F(x_0, \xi(t)) = \infty$  if  $T < \infty$ .

*Proof.* Let  $\sigma = \mathcal{T}_+$ , and note that the forward metric space  $(M, d_F)$  is forward boundedly compact by the Hopf–Rinow theorem. Since  $\phi \in C^1(M)$ , it follows from Example 2.20 and Remark 3.26 that  $|\partial^- \phi| = |\partial \phi| = F(\nabla(-\phi))$  is a strong upper gradient for  $-\phi$ . Now, if  $\phi$  is bounded below (or of lower  $p$ -growth), then Corollary 3.32 furnishes a  $p$ -curve  $\xi : [0, \infty) \rightarrow M$  of maximal slope for  $\phi$  with respect to  $|\partial^- \phi|$  with  $\xi(0) = x_0$ , and we conclude the proof by the argument in Example 2.25.

When  $\inf_X \phi = -\infty$ , we replace  $\phi$  with  $\phi_r := \max\{\phi, \inf_{B_{x_0}^+(r)} \phi\}$  for large  $r > 0$  and construct a gradient curve  $\xi$  within  $B_{x_0}^+(r)$ . If  $\xi$  does not reach  $\partial B_{x_0}^+(r)$ , then  $\xi$  is defined on  $[0, \infty)$ . If  $\xi$  reaches

$\partial B_{x_0}^+(r)$  at some  $T_1 \in (0, \infty)$ , then we continue the construction for  $\phi_{2r}$  from  $x_1 := \xi(T_1)$ . Iterating this procedure, since  $\nabla(-\phi_r) = \nabla(-\phi)$  in  $B_{x_0}^+(r)$ , we eventually obtain a  $C^1$ -curve  $\xi : [0, T) \rightarrow M$  satisfying  $j_p(\xi'(t)) = \nabla(-\phi)(\xi(t))$  and  $\lim_{t \rightarrow T} d_F(x_0, \xi(t)) = \infty$  if  $T < \infty$ .  $\square$

**Remark 3.34.** For  $\phi \in C^l(M)$  with  $l \geq 2$ ,  $\nabla(-\phi)$  is only continuous at its zeros while  $C^{l-1}$  at other points (see [9, 27, 32]). Hence, in order to get a higher regularity of  $\xi$ , we need to restrict ourselves to an interval in which  $\nabla(-\phi)(\xi(t)) \neq 0$ . See Corollary 4.19 below for instance.

**3.6. Doubly nonlinear evolution equations.** This subsection is devoted to gradient flows in infinite dimensional Funk and Randers-like spaces. Let  $(\mathcal{H}, \langle \cdot, \cdot \rangle)$  be a Hilbert space and set  $\|\cdot\| := \sqrt{\langle \cdot, \cdot \rangle}$ . On the unit ball  $\mathbb{B} := \{x \in \mathcal{H} \mid \|x\| < 1\}$ , we define an asymmetric distance function  $d : \mathbb{B} \times \mathbb{B} \rightarrow [0, \infty)$  in the same way as (1.2). Similarly to (1.1),  $d$  is associated with the (infinite dimensional) Finsler structure

$$F(x, v) := \lim_{\varepsilon \rightarrow 0^+} \frac{d(x, x + \varepsilon v)}{\varepsilon} = \frac{\sqrt{(1 - \|x\|^2)\|v\|^2 + \langle x, v \rangle^2} + \langle x, v \rangle}{1 - \|x\|^2}, \quad v \in T_x \mathbb{B} \cong \mathcal{H}.$$

We call  $(\mathbb{B}, d)$  a *generalized Funk space* and observe the following (see [1, Remark 1.1.3] for (iii)).

**Proposition 3.35.** *Let  $(\mathbb{B}, d)$  be a generalized Funk space.*

- (i)  $(\mathbb{B}, \mathbf{0}, d)$  is a forward complete pointed forward  $\Theta$ -metric space with  $\Theta(r) = 2e^r - 1$ .
- (ii) Both the forward and backward topologies coincide with the original topology of  $\mathbb{B} \subset \mathcal{H}$ .
- (iii) A curve  $\gamma$  belongs to  $\text{FAC}^p((a, b); \mathbb{B})$  if and only if it is differentiable at  $\mathcal{L}^1$ -a.e.  $t \in (a, b)$  with the derivative  $\gamma'$  satisfying  $F(\gamma, \gamma') \in L^p(a, b)$  and

$$\gamma(t) - \gamma(s) = \int_s^t \gamma'(r) \, dr \quad \text{for any } a < s \leq t < b.$$

Moreover, for  $\gamma \in \text{FAC}^p((a, b); \mathbb{B})$ , we have  $|\gamma'_+|(t) = F(\gamma(t), \gamma'(t))$  for  $\mathcal{L}^1$ -a.e.  $t \in (a, b)$ .

Let  $\mathcal{H}^*$  denote the dual space of  $\mathcal{H}$ . For  $\zeta \in T_x^* \mathbb{B} \cong \mathcal{H}^*$  ( $\cong \mathcal{H}$ ), the dual norm of  $F$  is defined as

$$F^*(x, \zeta) := \sup_{v \in T_x \mathbb{B} \setminus \{0\}} \frac{\langle \zeta, v \rangle}{F(x, v)}.$$

For  $p \in (1, \infty)$ ,  $q = p/(p-1)$  and  $v \in T_x \mathbb{B}$ , define  $\mathfrak{J}_p(x, v) \subset T_x^* \mathbb{B}$  by

$$\mathfrak{J}_p(x, v) := \{\zeta \in T_x^* \mathbb{B} \mid \langle \zeta, v \rangle = F^p(x, v) = F^*(x, \zeta)^q = F(x, v)F^*(x, \zeta)\}.$$

Note that  $\mathfrak{J}_p(x, y)$  is at most a singleton by the differentiability of  $\|\cdot\|$  (see Proposition B.1).

Given  $\phi : \mathbb{B} \rightarrow (-\infty, \infty]$  and  $x \in \mathbb{B}$ , we set

$$\begin{aligned} \partial\phi(x) &:= \left\{ \zeta \in T_x^* \mathbb{B} \mid \liminf_{v \rightarrow 0} \frac{\phi(x+v) - \phi(x) - \langle \zeta, v \rangle}{d(x, x+v)} \geq 0 \right\} \text{ for } x \in \mathfrak{D}(\phi), \\ \partial^\circ\phi(x) &:= \{\zeta \in \partial\phi(x) \mid F^*(x, -\zeta) \leq F^*(x, -\eta) \text{ for all } \eta \in \partial\phi(x)\} \text{ for } x \in \mathfrak{D}(\partial\phi), \\ F^*(-\partial^\circ\phi(x)) &:= \begin{cases} \inf_{\zeta \in \partial\phi(x)} F^*(x, -\zeta) & \text{if } x \in \mathfrak{D}(\partial\phi), \\ +\infty & \text{if } x \notin \mathfrak{D}(\partial\phi), \end{cases} \end{aligned}$$

where  $\partial\phi(x)$  is the *Fréchet subdifferential* of  $\phi$  at  $x$  and  $\mathfrak{D}(\partial\phi) := \{x \in \mathfrak{D}(\phi) \mid \partial\phi(x) \neq \emptyset\}$ . The function  $x \mapsto F^*(-\partial^\circ\phi(x))$  is a weak upper gradient for  $-\phi$  (see Proposition B.2).

The main result of this subsection reads as follows (cf. [1, Proposition 1.4.1, Theorem 2.3.7]).

**Theorem 3.36.** *Let  $(\mathbb{B}, d)$  be a generalized Funk space, and let  $\phi : \mathbb{B} \rightarrow (-\infty, \infty]$  admit a decomposition  $\phi = \phi_1 + \phi_2$  where  $\phi_1$  is a proper, lower semicontinuous, convex function and  $\phi_2$  is of class  $C^1$  in the Fréchet sense.*

- (i) The local slope  $|\partial\phi|$  is a strong upper gradient for  $-\phi$ . Moreover, for every  $x \in \mathbb{B}$ , we have  $|\partial\phi|(x) = F^*(-\partial^\circ\phi(x)) = |\partial^-\phi|(x)$ .
- (ii) If  $\xi : (a, b) \rightarrow \mathbb{B}$  is a  $p$ -curve of maximal slope for  $\phi$  with respect to  $|\partial\phi|$ , then it satisfies the following doubly nonlinear differential equation:

$$\mathfrak{J}_p(\xi(t), \xi'(t)) = -\partial^\circ\phi(\xi(t)) \neq \emptyset \quad \text{for } \mathcal{L}^1\text{-a.e. } t \in (a, b). \quad (3.28)$$

(iii) If Assumption 3.8(b), (c) hold with  $\sigma = \mathcal{T}_+$ , then, for every  $x_0 \in \mathfrak{D}(\phi)$ ,  $\text{GMM}_p(\phi; x_0)$  is nonempty and its element  $\xi$  satisfies (3.28) and the energy identity

$$\int_0^T |\xi'_+(t)|^p dt + \phi(\xi(T)) = \phi(x_0) \quad \text{for all } T > 0.$$

*Proof.* (i) By Proposition B.2, the same argument as in the proof of [1, Corollary 1.4.5] yields that  $\partial\phi = \partial\phi_1 + d\phi_2$  satisfies (B.1) and (B.2), and  $|\partial\phi|$  is a strong upper gradient for  $-\phi$  with  $|\partial\phi|(x) = F^*(-\partial^\circ\phi(x))$ . Moreover,  $|\partial\phi| = |\partial^-\phi|$  follows from (the proof of) [1, Lemma 2.3.6] (with  $\sigma = \mathcal{T}_+$ ).

(ii) Since  $\xi'(t)$  exists  $\mathcal{L}^1$ -a.e. due to Proposition 3.35(iii), Proposition 2.24 combined with (i) implies

$$\frac{d}{dt}\phi(\xi(t)) = -\frac{1}{p}|\xi'_+(t)|^p - \frac{1}{q}F^*(-\partial^\circ\phi(\xi(t)))^q \quad \text{for } \mathcal{L}^1\text{-a.e. } t \in (a, b).$$

On the other hand, for  $\mathcal{L}^1$ -a.e.  $t \in (a, b)$  and any  $\zeta \in \partial^\circ\phi(\xi(t))$ , we have

$$\frac{d}{dt}\phi(\xi(t)) = \lim_{\varepsilon \rightarrow 0^+} \frac{\phi(\xi(t+\varepsilon)) - \phi(\xi(t))}{\varepsilon} \geq \liminf_{\varepsilon \rightarrow 0^+} \frac{\langle \zeta, \xi(t+\varepsilon) - \xi(t) \rangle}{\varepsilon} = -\langle -\zeta, \xi'(t) \rangle.$$

The Young inequality then furnishes  $-\zeta \in \mathfrak{J}_p(\xi(t), \xi'(t))$ , and Proposition B.1(iii) yields (3.28).

(iii) This follows from Corollary 3.29, Theorem 3.30, and (i), (ii) above.  $\square$

We remark that the above argument also applies to the unit balls in general Banach spaces  $\mathcal{B}$  as well as Minkowski normed spaces of infinite dimension. For example, let  $(\mathcal{B}, \|\cdot\|)$  be a reflexive Banach space and  $(\mathcal{B}^*, \|\cdot\|_*)$  be its dual space. Given  $\omega \in \mathcal{B}^*$  with  $\|\omega\|_* < 1$ , we can define an asymmetric metric  $d_\omega$  on  $\mathcal{B}$  by  $d_\omega(x, y) := \|y - x\| + \omega(y - x)$  (cf. Example 3.6). Then the Minkowski normed space  $(\mathcal{B}, d_\omega)$  is a  $[(1 + \|\omega\|_*)/(1 - \|\omega\|_*)]$ -metric space and Proposition 3.35(ii), (iii) hold. Moreover, since

$$F(x, v) = \lim_{\varepsilon \rightarrow 0^+} \frac{d_\omega(x, x + \varepsilon v)}{\varepsilon} = \|v\| + \omega(v),$$

both  $F^*(x, \zeta)$  and  $\mathfrak{J}_p(x, v)$  are independent of  $x$ . Theorem 3.36 remains valid by replacing (3.28) with the doubly nonlinear differential inclusion  $\mathfrak{J}_p(\xi'(t)) \supset -\partial^\circ\phi(\xi(t))$  (if  $\|\cdot\|$  is differentiable, then  $\mathfrak{J}_p(v)$  is at most a singleton and Theorem 3.36 holds as it is).

A generalization of the theory of doubly nonlinear evolution equations (DNE) to asymmetric Finsler-like metrics on Banach spaces was investigated in [36, §8], under the finite reversibility and the lower boundedness of  $\phi_1$  (see [36, (5.1c), Definition 5.10]). We remark that the reversibility of generalized Funk spaces is infinite by Proposition 3.35(i).

#### 4. $(p, \lambda)$ -CONVEXITY

In this section, we study the behavior of curves of maximal slope under a certain convexity assumption on  $\Phi$ , where  $\Phi$  was defined in Definition 3.2 as a combination of the target function  $\phi$  and the distance function. Let  $(X, d)$  be a forward complete forward metric space,  $\phi : X \rightarrow (-\infty, \infty]$  be a proper function, and  $p \in (1, \infty)$  throughout this section.

**4.1. Convexity assumption.** A curve  $\gamma : [0, 1] \rightarrow X$  is called a *minimal geodesic* if  $d(\gamma(s), \gamma(t)) = (t - s)d(\gamma(0), \gamma(1))$  holds for any  $0 \leq s < t \leq 1$ .

**Definition 4.1** ( $(p, \lambda)$ -convexity). For  $\lambda \in \mathbb{R}$ , we say that  $\phi$  is  $(p, \lambda)$ -convex (resp.  $(p, \lambda)$ -geodesically convex) if, for any  $x_0, x_1 \in X$ , there is a curve (resp. a minimal geodesic)  $\gamma : [0, 1] \rightarrow X$  from  $x_0$  to  $x_1$  such that

$$\phi(\gamma(t)) \leq (1 - t)\phi(x_0) + t\phi(x_1) - \frac{\lambda}{p}t(1 - t^{p-1})d^p(x_0, x_1) \quad \text{for all } t \in [0, 1].$$

In the particular case of  $p = 2$ ,  $\phi$  is also called a  $\lambda$ -convex (resp.  $\lambda$ -geodesically convex) function.

**Remark 4.2.** We remark that our definition of  $(p, \lambda)$ -convexity slightly differs from the  $(\lambda, p)$ -convexity in [37, Definition 2.5] (discussed on symmetric metric spaces), which requires

$$\phi(\gamma(t)) \leq (1 - t)\phi(x_0) + t\phi(x_1) - \frac{\lambda}{p}t(1 - t)d^p(x_0, x_1) \quad \text{for all } t \in [0, 1]$$

(see also [1, Remark 2.4.7]). Since  $t \in [0, 1]$ , this is stronger (resp. weaker) than our  $(p, \lambda)$ -convexity for  $\lambda > 0$  and  $p \in (1, 2)$  or  $\lambda < 0$  and  $p \in (2, \infty)$  (resp. for  $\lambda > 0$  and  $p \in (2, \infty)$  or  $\lambda < 0$  and  $p \in (1, 2)$ ). Our  $(p, \lambda)$ -convexity is well-behaved in subdivisions in the sense that combining

$$\phi\left(\gamma\left(\frac{1}{4}\right)\right) \leq \frac{1}{2}\phi(x_0) + \frac{1}{2}\phi\left(\gamma\left(\frac{1}{2}\right)\right) - \frac{\lambda}{p} \frac{1}{2} \left(1 - \frac{1}{2^{p-1}}\right) \left(\frac{d(x_0, x_1)}{2}\right)^p$$

and

$$\phi\left(\gamma\left(\frac{1}{2}\right)\right) \leq \frac{1}{2}\phi(x_0) + \frac{1}{2}\phi(x_1) - \frac{\lambda}{p} \frac{1}{2} \left(1 - \frac{1}{2^{p-1}}\right) d^p(x_0, x_1)$$

implies

$$\phi\left(\gamma\left(\frac{1}{4}\right)\right) \leq \frac{3}{4}\phi(x_0) + \frac{1}{4}\phi(x_1) - \frac{\lambda}{p} \frac{1}{4} \left(1 - \frac{1}{4^{p-1}}\right) d^p(x_0, x_1).$$

Moreover, our usage of  $\Phi$  below in the spirit of [1] seems also advantageous. See Remarks 4.12, 4.18 for further discussions.

Now we introduce an important convexity assumption (in the spirit of [1, Assumption 2.4.5]).

**Assumption 4.3.** Let  $\lambda \in \mathbb{R}$  and put  $\lambda_- := -\min\{\lambda, 0\}$ . We assume that, for any  $x_0, x_1 \in \mathfrak{D}(\phi)$ , there exists a curve  $\gamma : [0, 1] \rightarrow X$  from  $x_0$  to  $x_1$  such that

$$\Phi(\tau, x_0; \gamma(t)) \leq (1-t)\Phi(\tau, x_0; x_0) + t\Phi(\tau, x_0; x_1) - \frac{1}{p} \left(\lambda + \frac{1}{\tau^{p-1}}\right) t(1-t^{p-1})d^p(x_0, x_1) \quad (4.1)$$

for all  $\tau \in (0, \lambda_-^{-1/(p-1)})$  and  $t \in [0, 1]$ , where we set  $\lambda_-^{-1/(p-1)} := \infty$  if  $\lambda_- = 0$ .

In other words,  $\Phi(\tau, x_0; \cdot)$  is  $(p, \lambda + \tau^{1-p})$ -convex for a common curve  $\gamma$  for all  $\tau \in (0, \lambda_-^{-1/(p-1)})$ . We remark that  $\gamma$  is always emanating from  $x_0$  (cf. [1, Assumption 4.0.1], which is a stronger convexity condition requiring the convexity of  $\Phi(\tau, x_0; \cdot)$  between any pair of points).

**Example 4.4.** (a) It is readily seen that a  $(p, \lambda)$ -geodesically convex function  $\phi : X \rightarrow (-\infty, \infty]$  satisfies Assumption 4.3 (regardless of the convexity of the distance function). Indeed, for a minimal geodesic  $\gamma : [0, 1] \rightarrow X$  from  $x_0$  to  $x_1$  along which the  $(p, \lambda)$ -convexity holds, we have

$$\begin{aligned} \Phi(\tau, x_0; \gamma(t)) &= \phi(\gamma(t)) + \frac{d^p(x_0, \gamma(t))}{p\tau^{p-1}} \\ &\leq (1-t)\phi(x_0) + t\phi(x_1) - \frac{\lambda}{p} t(1-t^{p-1})d^p(x_0, x_1) + \frac{t^p}{p\tau^{p-1}} d^p(x_0, x_1) \\ &= (1-t)\Phi(\tau, x_0; x_0) + t\Phi(\tau, x_0; x_1) - \frac{1}{p} \left(\lambda + \frac{1}{\tau^{p-1}}\right) t(1-t^{p-1})d^p(x_0, x_1). \end{aligned}$$

(b) The convexity of the distance function is intimately related to how the space is curved. For a complete, simply-connected Riemannian manifold  $(M, g)$ ,  $d_g^2(x, \cdot)$  is 2-geodesically convex for every  $x \in M$  if and only if the sectional curvature is nonpositive. This is a fact at the origin of the fruitful theory of CAT(0)-spaces (see, e.g., [12]). In the Finsler case, only the flag curvature is not sufficient to control the convexity of the distance function and we need some additional conditions. For example, for forward complete, simply-connected Finsler manifold  $(M, F)$  of Berwald type with nonpositive flag curvature,  $d_F^2(x, \cdot)$  is 0-geodesically convex for every  $x \in M$ . We refer to [40, §15.1] and [26, §5] for this and more general results.

We collect some immediate consequences of Assumption 4.3 in the next lemma.

**Lemma 4.5.** *Suppose that  $\phi$  satisfies Assumption 4.3 for some  $(p, \lambda)$ . Then we have the following.*

- (i)  $\phi$  satisfies Assumption 4.3 for all  $(p, \lambda')$  with  $\lambda' < \lambda$ .
- (ii) For any  $\tau \in (0, \lambda_-^{-1/(p-1)})$  and  $t \in (0, 1]$ , we have

$$\frac{\phi(\gamma(t)) - \phi(x_0)}{t} \leq \phi(x_1) - \phi(x_0) + \frac{t^{p-1} - \lambda\tau^{p-1}(1-t^{p-1})}{p\tau^{p-1}} d^p(x_0, x_1), \quad (4.2)$$

where  $\gamma$  is the curve satisfying (4.1). In particular,

$$d(x_0, \gamma(t)) \leq td(x_0, x_1) \quad \text{for all } t \in [0, 1]. \quad (4.3)$$

(iii) If  $\lambda \geq 0$ , then we have

$$\frac{\phi(\gamma(t)) - \phi(x_0)}{t} \leq \phi(x_1) - \phi(x_0) - \frac{\lambda}{p}(1 - t^{p-1})d^p(x_0, x_1) \quad \text{for all } t \in (0, 1]. \quad (4.4)$$

In particular,  $\phi$  is  $(p, \lambda)$ -convex.

*Proof.* (i) is trivial by definition. In (ii), (4.2) follows from  $\phi(\gamma(t)) \leq \Phi(\tau, x_0; \gamma(t))$  and (4.1) as

$$\begin{aligned} \Phi(\tau, x_0; \gamma(t)) &\leq (1-t)\phi(x_0) + t\phi(x_1) + t \frac{d^p(x_0, x_1)}{p\tau^{p-1}} - \frac{1}{p} \left( \lambda + \frac{1}{\tau^{p-1}} \right) t(1-t^{p-1})d^p(x_0, x_1) \\ &= (1-t)\phi(x_0) + t\phi(x_1) + t \frac{t^{p-1} - \lambda\tau^{p-1}(1-t^{p-1})}{p\tau^{p-1}} d^p(x_0, x_1). \end{aligned} \quad (4.5)$$

We obtain (4.3) by letting  $\tau \rightarrow 0$  in the above inequality. For (iii), the assumption  $\lambda \geq 0$  implies  $\lambda_- = 0$  and hence (4.2) holds for all  $\tau \in (0, \infty)$ , which furnishes (4.4) as  $\tau \rightarrow \infty$ .  $\square$

Next we give an estimate of  $\tau_*(\phi)$  (see Definition 3.12) under Assumption 4.3 (cf. [1, Lemma 2.4.8]).

**Lemma 4.6.** *Suppose that Assumption 4.3 holds for some  $(p, \lambda)$  and there are  $x_* \in \mathfrak{D}(\phi)$  and  $r_* > 0$  such that*

$$m_* := \inf\{\phi(x) \mid d(x_*, x) \leq r_*\} > -\infty. \quad (4.6)$$

- (i) We have  $\tau_*(\phi) \geq \lambda_-^{-1/(p-1)}$ . In particular,  $\tau_*(\phi) = \infty$  if  $\lambda \geq 0$ .
- (ii) If  $\lambda > 0$ , then  $\phi$  is bounded from below. Moreover, if  $\phi$  is  $\mathcal{T}_+$ -lower semicontinuous, then it has a unique minimizer  $\bar{x} \in X$ .

*Proof.* (i) Given  $\tau = (\lambda_- + \varepsilon)^{-1/(p-1)} \in (0, \lambda_-^{-1/(p-1)})$ , we claim that

$$\inf\{\Phi(\tau, x_*; y) \mid d(x_*, y) > r_*\} > -\infty. \quad (4.7)$$

Combining this with (4.6) implies  $\tau_*(\phi) \geq (\lambda_- + \varepsilon)^{-1/(p-1)}$ , and then (i) follows by letting  $\varepsilon \rightarrow 0$ .

In order to show (4.7), take an arbitrary point  $y \in \mathfrak{D}(\phi)$  with  $d(x_*, y) > r_*$ , let  $\gamma : [0, 1] \rightarrow X$  be a curve from  $x_0 = x_*$  to  $x_1 = y$  satisfying (4.1), and set  $y_* := \gamma(r_*/d(x_*, y))$ . Then (4.3) implies  $d(x_*, y_*) \leq r_*$  and hence  $\phi(y_*) \geq m_*$ . We deduce from (4.2) that

$$\begin{aligned} &\frac{d(x_*, y)}{r_*} (\phi(y_*) - \phi(x_*)) \\ &\leq \phi(y) - \phi(x_*) + \frac{1}{p\tau^{p-1}} \left\{ \left( \frac{r_*}{d(x_*, y)} \right)^{p-1} - \lambda\tau^{p-1} \left( 1 - \left( \frac{r_*}{d(x_*, y)} \right)^{p-1} \right) \right\} d^p(x_*, y). \end{aligned}$$

This together with  $\phi(y_*) \geq m_*$  yields

$$\phi(x_*) + c_* d(x_*, y) \leq \phi(y) - \frac{\lambda}{p} d^p(x_*, y) \leq \phi(y) + \frac{\lambda_-}{p} d^p(x_*, y), \quad (4.8)$$

$$\text{where } c_* := \frac{m_* - \phi(x_*) - (\tau^{1-p} + \lambda)r_*^p/p}{r_*}.$$

Combining this with the Young inequality

$$-c_* d(x_*, y) \leq |c_*| d(x_*, y) \leq \frac{\varepsilon}{p} d^p(x_*, y) + \frac{\varepsilon^{-q/p}}{q} |c_*|^q$$

and recalling  $\tau = (\lambda_- + \varepsilon)^{-1/(p-1)}$ , we find

$$\Phi(\tau, x_*; y) = \phi(y) + \frac{\lambda_- + \varepsilon}{p} d^p(x_*, y) \geq \phi(x_*) + c_* d(x_*, y) + \frac{\varepsilon}{p} d^p(x_*, y) \geq \phi(x_*) - \frac{\varepsilon^{-q/p}}{q} |c_*|^q.$$

This implies (4.7) and completes the proof.

(ii) If  $\lambda > 0$ , then for any  $\varepsilon > \lambda$ , the above argument with  $\tau = (\varepsilon - \lambda)^{-1/(p-1)}$  shows

$$\phi(y) + \frac{\varepsilon - \lambda}{p} d^p(x_*, y) \geq \phi(x_*) + c_* d(x_*, y) + \frac{\varepsilon}{p} d^p(x_*, y) \geq \phi(x_*) - \frac{\varepsilon^{-q/p}}{q} (-c_*)^q.$$

Letting  $\varepsilon \rightarrow \lambda$ , we deduce from the above inequality and (4.6) that

$$\inf_X \phi \geq \min \left\{ m_*, \phi(x_*) - \frac{\lambda^{-q/p} (\phi(x_*) - m_* + (\lambda r_*^p)/p)^q}{r_*^q} \right\}.$$

When  $\phi$  is lower semicontinuous, take any minimizing sequence  $(x_i)_{i \geq 1}$ , i.e.,  $\phi(x_i) \rightarrow \inf_X \phi$ . For any  $0 < i \leq j$ , let  $\gamma_{i,j}$  be a curve from  $x_i$  to  $x_j$  satisfying (4.1). It follows from (4.4) (with  $t = 1/2$ ) that

$$2 \left\{ \phi \left( \gamma_{i,j} \left( \frac{1}{2} \right) \right) - \phi(x_i) \right\} \leq \phi(x_j) - \phi(x_i) - \frac{\lambda}{p} (1 - 2^{1-p}) d^p(x_i, x_j),$$

which furnishes

$$\frac{\lambda}{p} (1 - 2^{1-p}) d^p(x_i, x_j) \leq \phi(x_i) + \phi(x_j) - 2 \phi \left( \gamma_{i,j} \left( \frac{1}{2} \right) \right) \leq \phi(x_i) + \phi(x_j) - 2 \inf_X \phi \rightarrow 0$$

as  $i, j \rightarrow \infty$ . Thus,  $(x_i)_{i \geq 1}$  is a forward Cauchy sequence and converges to some  $\bar{x} \in X$ , which is a minimizer of  $\phi$  due to the lower semicontinuity of  $\phi$ . The uniqueness also follows from the above argument, as any minimizing sequence is convergent.  $\square$

Two remarks on the relationship between Assumption 3.8 and (4.6) are in order.

**Remark 4.7.** (a) Under Assumption 4.3, Assumption 3.8(b) is equivalent to (4.6). On the one hand, it follows from Lemma 4.6(i) that (4.6) implies Assumption 3.8(b). On the other hand, if Assumption 3.8(b) ( $\Phi_{\tau_*}(x_*) > -\infty$ ) holds, then we have, for any  $r_* > 0$  and  $x \in \overline{B_{x_*}^+(r_*)}$ ,

$$\phi(x) \geq \Phi_{\tau_*}(x_*) - \frac{d^p(x_*, x)}{p\tau_*^{p-1}} \geq \Phi_{\tau_*}(x_*) - \frac{r_*^p}{p\tau_*^{p-1}}.$$

Hence, (4.6) holds.

(b) Assumption 3.8(a), (c) imply (4.6). Assume on the contrary that there is a sequence  $(x_i)_{i \geq 1}$  in  $\overline{B_{x_*}^+(r_*)}$  with  $\phi(x_i) \rightarrow -\infty$ . By Assumption 3.8(c), we can find a subsequence of  $(x_i)_{i \geq 1}$  which is  $\sigma$ -converging to some point  $x_\infty$ . Then Assumption 3.8(a) furnishes  $\phi(x_\infty) \leq \lim_{i \rightarrow \infty} \phi(x_i) = -\infty$ , which contradicts  $\phi(x_\infty) > -\infty$ .

**4.2. Existence of curves of maximal slope.** Under the convexity as in Assumption 4.3, we can show a global formula for the local slope  $|\partial\phi|$  (recall Definition 2.19), which plays an important role in proving the existence of curves of maximal slope (cf. [1, Theorem 2.4.9], [36, Lemma 5.3], [37, Proposition 2.7]).

**Theorem 4.8** (Global formula of  $|\partial\phi|$ ). *If Assumption 4.3 holds for some  $(p, \lambda)$ , then*

$$|\partial\phi|(x) = \sup_{y \neq x} \left[ \frac{\phi(x) - \phi(y)}{d(x, y)} + \frac{\lambda}{p} d^{p-1}(x, y) \right]_+ \quad \text{for all } x \in \mathfrak{D}(\phi). \quad (4.9)$$

Moreover, when  $\lambda \geq 0$ , we have  $|\partial\phi|(x) = \mathfrak{I}_\phi(x)$  for all  $x \in \mathfrak{D}(\phi)$ .

*Proof.* Let  $x \in \mathfrak{D}(\phi)$ . Clearly we have

$$|\partial\phi|(x) = \limsup_{y \rightarrow x} \frac{[\phi(x) - \phi(y)]_+}{d(x, y)} \leq \sup_{y \neq x} \left[ \frac{\phi(x) - \phi(y)}{d(x, y)} + \frac{\lambda}{p} d^{p-1}(x, y) \right]_+.$$

In order to show the reverse inequality, without loss of generality, we assume that there is  $y \neq x$  such that

$$\phi(x) - \phi(y) + \frac{\lambda}{p} d^p(x, y) > 0.$$

Let  $\gamma$  be a curve from  $x$  to  $y$  satisfying (4.1). Then (4.2) yields that, for any  $\tau \in (0, \lambda_-^{-1/(p-1)})$ ,

$$\frac{\phi(x) - \phi(\gamma(t))}{d(x, \gamma(t))} \geq \left\{ \frac{\phi(x) - \phi(y)}{d(x, y)} + \frac{\lambda \tau^{p-1}(1 - t^{p-1}) - t^{p-1}}{p\tau^{p-1}} d^{p-1}(x, y) \right\} \frac{td(x, y)}{d(x, \gamma(t))}.$$

Combining this with  $d(x, \gamma(t)) \leq td(x, y)$  from (4.3), we find

$$|\partial\phi|(x) \geq \limsup_{t \rightarrow 0} \frac{\phi(x) - \phi(\gamma(t))}{d(x, \gamma(t))} \geq \frac{\phi(x) - \phi(y)}{d(x, y)} + \frac{\lambda}{p} d^{p-1}(x, y) > 0.$$

Now (4.9) follows by taking the supremum in  $y$ .

When  $\lambda \geq 0$ , it is straightforward from (4.9) that  $|\partial\phi|(x) \geq \mathfrak{l}_\phi(x)$ . Since the reverse inequality clearly holds by definition, we have  $|\partial\phi|(x) = \mathfrak{l}_\phi(x)$ .  $\square$

**Corollary 4.9.** *Suppose that Assumption 4.3 holds for some  $(p, \lambda)$  and  $\phi$  is  $\mathcal{T}_+$ -lower semicontinuous. Then  $|\partial\phi|$  is a strong upper gradient for  $-\phi$  and  $\mathcal{T}_+$ -lower semicontinuous.*

*Proof.* It follows from Theorem 2.21 that  $|\partial\phi|$  is a weak upper gradient and  $\mathfrak{l}_\phi$  is a strong upper gradient. Hence, if  $\lambda \geq 0$ , Theorem 4.8 implies that  $|\partial\phi|$  is a strong upper gradient.

Next, suppose that  $\lambda < 0$  and  $\text{Diam}(X) < \infty$ . Observe from (4.9) that

$$\mathfrak{l}_\phi(x) \leq |\partial\phi|(x) - \frac{\lambda}{p} \text{Diam}(X)^{p-1} \quad \text{for all } x \in \mathfrak{D}(\phi). \quad (4.10)$$

Hence, for any  $\gamma \in \text{FAC}([a, b]; X)$ ,  $|\partial\phi| \circ \gamma |\gamma'_+| \in L^1(a, b)$  implies that  $\mathfrak{l}_\phi \circ \gamma |\gamma'_+| \in L^1(a, b)$  and  $-\phi \circ \gamma$  is absolutely continuous by Remark 2.16. Thus, Proposition 2.18 shows that  $|\partial\phi|$  is a strong upper gradient.

Finally, when  $\lambda < 0$  and  $\text{Diam}(X) = \infty$ , we shall again prove that  $-\phi \circ \gamma$  is absolutely continuous for  $\gamma \in \text{FAC}([a, b]; X)$  with  $|\partial\phi| \circ \gamma |\gamma'_+| \in L^1(a, b)$ . Given such a curve  $\gamma$ , we find from Lemma 2.8 that  $\gamma([a, b])$  is compact. Let  $X_0 := \gamma([a, b])$  be the compact forward metric space equipped with the restricted metric  $d$ , and denote by  $\mathfrak{l}_\phi^0$  the corresponding global slope of  $\phi$ . Then, the same argument as above together with (4.10) yields  $\mathfrak{l}_\phi^0 \circ \gamma |\gamma'_+| \in L^1(a, b)$  and  $-\phi \circ \gamma$  is absolutely continuous. Therefore,  $|\partial\phi|$  is always a strong upper gradient for  $-\phi$ .

The lower semicontinuity of  $|\partial\phi|$  can be shown in the same way as Theorem 2.21(ii) by (4.9).  $\square$

Now we present two existence results under Assumption 4.3 (cf. [1, Corollaries 2.4.11, 2.4.12]). Compare them with Corollary 3.32. Recall Remark 3.9 for Assumption 3.8 in the  $\sigma = \mathcal{T}_+$  case.

**Proposition 4.10.** *Suppose Assumption 3.8(a), (c) with  $\sigma = \mathcal{T}_+$  and Assumption 4.3 for some  $(p, \lambda)$ . Then, for every  $x_0 \in \mathfrak{D}(\phi)$ , there exists a  $p$ -curve  $\xi : [0, \infty) \rightarrow X$  of maximal slope for  $\phi$  with respect to  $|\partial\phi|$  with  $\xi(0) = x_0$ . In particular,  $\xi$  satisfies the energy identity (3.22).*

*Proof.* Observe from Remark 4.7 that Assumption 3.8(b) also holds. Then, from an arbitrary sequence of time steps  $(\mathfrak{T}_\alpha)_{\alpha \geq 1}$  such that  $\|\mathfrak{T}_\alpha\| \rightarrow 0$  and a corresponding sequence of initial data  $(\Xi_{\mathfrak{T}_\alpha}^0)_{\alpha \geq 1}$  with  $\Xi_{\mathfrak{T}_\alpha}^0 = x_0$  for all  $\alpha$ , we can find a limit curve  $\xi \in \text{GMM}_p(\phi; x_0)$  by Corollary 3.29. Furthermore, since  $|\partial^- \phi| = |\partial\phi|$  is a strong upper gradient for  $-\phi$  by Remark 3.26 and Corollary 4.9, we deduce from Theorem 3.30 that  $\xi$  is a  $p$ -curve of maximal slope and satisfies the energy identity.  $\square$

**Proposition 4.11.** *Suppose Assumption 3.8(a), (c) and Assumption 4.3 for some  $(p, \lambda)$ . If  $|\partial\phi|$  is  $\sigma$ -sequentially lower semicontinuous on forward bounded subsets of sublevel sets of  $\phi$ , then for every  $x_0 \in \mathfrak{D}(\phi)$ , there exists a  $p$ -curve  $\xi : [0, \infty) \rightarrow X$  of maximal slope for  $\phi$  with respect to  $|\partial\phi|$  satisfying  $\xi(0) = x_0$  and the energy identity (3.22).*

*Proof.* One can show  $|\partial^- \phi| = |\partial\phi|$  in the same way as in Remark 3.26. Moreover, it follows from Assumption 3.8(a) and Corollary 4.9 that  $|\partial\phi|$  is a strong upper gradient for  $-\phi$ . Then we obtain the claim by a similar argument to Proposition 4.10 together with Corollary 3.29 and Theorem 3.30.  $\square$

**Remark 4.12.** When we replace  $t(1 - t^{p-1})$  with  $t(1 - t)$  in (4.1) (as in the  $(\lambda, p)$ -convexity; recall Remark 4.2), we have  $d(x_0, \gamma(t)) \leq t^{2/p}d(x_0, x_1)$  in place of (4.3). Although this does not affect Lemma 4.6 up to suitable modifications (e.g.,  $y_* := \gamma((r_*/d(x_*, y))^{p/2})$ ), however the proof of Theorem 4.8 works only when  $p \leq 2$ . The lack of (4.9) causes problems when one tries to follow the succeeding arguments.

**4.3. Regularizing effects.** In this subsection, we shall show that Assumption 4.3 with  $\lambda > 0$  implies various finer properties, including the exponential convergence to a minimizer of  $\phi$  (Theorem 4.14; cf. [1, Theorem 2.4.14]) as well as a decay estimate of the local slope  $|\partial\phi|$  (Theorem 4.17; cf. [1, Theorem 2.4.15], [37, Proposition 2.7]).

**Lemma 4.13.** *Let  $\phi$  satisfy Assumption 4.3 with  $\lambda > 0$ . Then we have*

$$\phi(x) - \inf_X \phi \leq \frac{|\partial\phi|^q(x)}{q\lambda^{q/p}} \quad \text{for all } x \in \mathfrak{D}(\phi). \quad (4.11)$$

Moreover, if  $\bar{x} \in \mathfrak{D}(\phi)$  is a minimizer of  $\phi$ , then

$$\frac{\lambda}{p} d^p(\bar{x}, x) \leq \phi(x) - \phi(\bar{x}) \quad \text{for all } x \in \mathfrak{D}(\phi). \quad (4.12)$$

*Proof.* For any  $x, y \in \mathfrak{D}(\phi)$  with  $\phi(x) \geq \phi(y)$ , we deduce from (4.9) and the Young inequality that

$$\phi(x) - \phi(y) \leq |\partial\phi|(x)d(x, y) - \frac{\lambda}{p} d^p(x, y) \leq \frac{|\partial\phi|^q(x)}{q\lambda^{q/p}},$$

which implies (4.11). When  $\bar{x}$  is a minimizer of  $\phi$ , (4.4) with  $x_0 = \bar{x}$  and  $x_1 = x \in \mathfrak{D}(\phi)$  yields

$$\frac{\lambda}{p} (1 - t^{p-1}) d^p(\bar{x}, x) \leq \phi(x) - \phi(\bar{x})$$

since  $\phi(\gamma(t)) \geq \phi(\bar{x})$ . Letting  $t \rightarrow 0$  completes the proof.  $\square$

**Theorem 4.14.** *Suppose that Assumption 4.3 holds and  $\phi$  is  $\mathcal{T}_+$ -lower semicontinuous with  $\inf_X \phi > -\infty$ . Then, for any  $p$ -curve  $\xi : [0, \infty) \rightarrow X$  of maximal slope for  $\phi$  with respect to  $|\partial\phi|$ , we have*

$$\phi(\xi(t)) - \inf_X \phi \leq \left\{ \phi(\xi(t_0)) - \inf_X \phi \right\} \cdot \exp(-q \operatorname{sgn}(\lambda) |\lambda|^{q/p} (t - t_0)) \quad \text{for all } t \geq t_0 > 0. \quad (4.13)$$

In particular, if  $\lambda > 0$  and  $\bar{x} \in \mathfrak{D}(\phi)$  is a minimizer of  $\phi$ , then

$$d^p(\bar{x}, \xi(t)) \leq \frac{p}{\lambda} \left\{ \phi(\xi(t_0)) - \inf_X \phi \right\} \cdot \exp(-q\lambda^{q/p} (t - t_0)) \quad \text{for all } t \geq t_0 > 0.$$

*Proof.* Owing to Corollary 4.9,  $|\partial\phi|$  is a strong upper gradient for  $-\phi$ . Thus, we observe from Proposition 2.24 that  $\phi \circ \xi$  is locally absolutely continuous and

$$|\xi'_+|^p(t) = |\partial\phi|^q(\xi(t)) = -(\phi \circ \xi)'(t) \quad \text{for } \mathcal{L}^1\text{-a.e. } t \in (0, \infty).$$

Note that

$$q \operatorname{sgn}(\lambda) |\lambda|^{q/p} \left\{ \phi(\xi(t)) - \inf_X \phi \right\} \leq |\partial\phi|^q(\xi(t)),$$

which follows from (4.11) when  $\lambda > 0$  and is trivial for  $\lambda \leq 0$ . Thus,  $\Delta(t) := \phi(\xi(t)) - \inf_X \phi \geq 0$  satisfies

$$\Delta'(t) = -|\partial\phi|^q(\xi(t)) \leq -q \operatorname{sgn}(\lambda) |\lambda|^{q/p} \Delta(t) \quad \text{for } \mathcal{L}^1\text{-a.e. } t \in (0, \infty).$$

This implies (4.13) by integration. The second assertion is straightforward from (4.12) and (4.13).  $\square$

We can derive from Assumption 4.3 some estimates stronger than those in Subsection 3.3 (cf. [1, Theorem 3.1.6, Remark 3.1.7]).

**Theorem 4.15.** *Suppose that Assumptions 3.16 and 4.3 hold and let  $\tau \in (0, \tau_*(\phi))$ .*

(i) *If  $1 + \lambda\tau^{p-1} > 0$ , then we have, for any  $x \in \mathfrak{D}(\phi)$  and  $y_\tau \in J_\tau[x]$ ,*

$$(1 + \lambda\tau^{p-1}) |\partial\phi|^q(y_\tau) \leq (1 + \lambda\tau^{p-1}) \frac{d^p(x, y_\tau)}{\tau^p} \leq q \frac{\phi(x) - \Phi_\tau(x)}{\tau} \leq \frac{|\partial\phi|^q(x)}{(1 + \lambda\tau^{p-1})^{q/p}}. \quad (4.14)$$

*The last inequality holds even when  $J_\tau[x] = \emptyset$ .*

(ii) *If  $\lambda \geq 0$ , then we have, for any  $x \in \mathfrak{D}(\phi)$  and  $y_\tau \in J_\tau[x]$ ,*

$$\phi(y_\tau) - \inf_X \phi \leq \frac{1}{1 + \tau\lambda^{q/p}(q + \lambda\tau^{p-1})} \left\{ \phi(x) - \inf_X \phi \right\}, \quad \sup_{\tau > 0} \frac{\phi(x) - \Phi_\tau(x)}{\tau} = \frac{|\partial\phi|^q(x)}{q}.$$

*Proof.* (i) The first inequality in (4.14) was shown in Lemma 3.20. In order to prove the second one, set

$$f(t) := (1-t)\phi(x) + t\phi(y_\tau) + \frac{t}{p\tau^{p-1}} \{t^{p-1} - \lambda\tau^{p-1}(1-t^{p-1})\} d^p(x, y_\tau), \quad t \in [0, 1].$$

By (4.5) with  $x_0 = x$  and  $x_1 = y_\tau$ , we obtain

$$f(1) = \Phi(\tau, x; y_\tau) \leq \Phi(\tau, x; \gamma(t)) \leq f(t)$$

for any  $t \in (0, 1)$ . This implies  $f'(1) \leq 0$ , thereby

$$\phi(y_\tau) - \phi(x) + \left( \frac{1}{\tau^{p-1}} + \frac{\lambda}{q} \right) d^p(x, y_\tau) \leq 0.$$

One can rearrange this inequality to see

$$\frac{1 + \lambda\tau^{p-1}}{q} \frac{d^p(x, y_\tau)}{\tau^{p-1}} \leq \phi(x) - \phi(y_\tau) - \frac{d^p(x, y_\tau)}{p\tau^{p-1}} = \phi(x) - \Phi_\tau(x).$$

This is the second inequality in (4.14). The last inequality follows from (4.9) and the Young inequality as

$$\begin{aligned} \frac{\phi(x) - \Phi_\tau(x)}{\tau} &= \left\{ \frac{\phi(x) - \phi(y_\tau)}{d(x, y_\tau)} + \frac{\lambda}{p} d^{p-1}(x, y_\tau) \right\} \frac{d(x, y_\tau)}{\tau} - (1 + \lambda\tau^{p-1}) \frac{d^p(x, y_\tau)}{p\tau^p} \\ &\leq |\partial\phi|(x) \frac{d(x, y_\tau)}{\tau} - (1 + \lambda\tau^{p-1}) \frac{d^p(x, y_\tau)}{p\tau^p} \leq \frac{|\partial\phi|^q(x)}{q(1 + \lambda\tau^{p-1})^{q/p}}. \end{aligned}$$

If  $J_\tau[x] = \emptyset$ , then we replace  $y_\tau$  with  $y \neq x$  and take the supremum in  $y$ .

(ii) Owing to (4.14), we obtain

$$\frac{\phi(x) - \phi(y_\tau)}{\tau} = \frac{\phi(x) - \Phi_\tau(x)}{\tau} + \frac{d^p(x, y_\tau)}{p\tau^p} \geq \left( \frac{1 + \lambda\tau^{p-1}}{q} + \frac{1}{p} \right) |\partial\phi|^q(y_\tau) = \left( 1 + \frac{\lambda\tau^{p-1}}{q} \right) |\partial\phi|^q(y_\tau).$$

This together with (4.11) implies the first assertion as

$$\left\{ \phi(x) - \inf_X \phi \right\} - \left\{ \phi(y_\tau) - \inf_X \phi \right\} \geq \tau \left( 1 + \frac{\lambda\tau^{p-1}}{q} \right) q\lambda^{q/p} \left\{ \phi(y_\tau) - \inf_X \phi \right\}.$$

The second assertion follows from (3.10) and (4.14) as

$$\frac{|\partial\phi|^q(x)}{q} = \limsup_{\tau \rightarrow 0} \frac{\phi(x) - \Phi_\tau(x)}{\tau} \leq \sup_{\tau > 0} \frac{\phi(x) - \Phi_\tau(x)}{\tau} \leq \frac{|\partial\phi|^q(x)}{q}.$$

□

**Lemma 4.16.** *Suppose that Assumption 4.3 holds for some  $(p, \lambda)$  satisfying one of the following: (1)  $p \in (1, 2)$  and  $\lambda \geq 0$ ; (2)  $p = 2$  and  $\lambda \in \mathbb{R}$ ; (3)  $p \in (2, \infty)$  and  $\lambda = 0$ . Given a sequence of time steps  $\mathfrak{T} = (\tau_k)_{k \geq 1}$  with  $\lambda \|\mathfrak{T}\|^{p-1} > -1$ , we set*

$$\lambda_{\mathfrak{T}} := \frac{\log(1 + \lambda \|\mathfrak{T}\|^{p-1})}{\|\mathfrak{T}\|^{p-1}} = \inf_{k \geq 1} \frac{\log(1 + \lambda\tau_k^{p-1})}{\tau_k^{p-1}} \leq \lambda.$$

Then  $e^{\lambda_{\mathfrak{T}}(t_{\mathfrak{T}}^k)^{p-1}} |\partial\phi|(\Xi_{\mathfrak{T}}^k)$  is non-increasing in  $k$  for any  $(\Xi_{\mathfrak{T}}^k)_{k \geq 0}$  solving (3.1).

*Proof.* We consider only the case (1) (the other cases can be seen similarly). Owing to  $p-1 \in (0, 1)$ ,

$$(t_{\mathfrak{T}}^k)^{p-1} = (t_{\mathfrak{T}}^{k-1} + \tau_k)^{p-1} \leq (t_{\mathfrak{T}}^{k-1})^{p-1} + \tau_k^{p-1}.$$

Combining this with (4.14),  $\lambda_{\mathfrak{T}} \geq 0$  and  $e^{\lambda_{\mathfrak{T}}\tau_k^{p-1}} \leq 1 + \lambda\tau_k^{p-1}$ , we obtain

$$e^{\lambda_{\mathfrak{T}}(t_{\mathfrak{T}}^k)^{p-1}} |\partial\phi|(\Xi_{\mathfrak{T}}^k) \leq e^{\lambda_{\mathfrak{T}}(t_{\mathfrak{T}}^{k-1})^{p-1}} e^{\lambda_{\mathfrak{T}}\tau_k^{p-1}} \frac{|\partial\phi|(\Xi_{\mathfrak{T}}^{k-1})}{1 + \lambda\tau_k^{p-1}} \leq e^{\lambda_{\mathfrak{T}}(t_{\mathfrak{T}}^{k-1})^{p-1}} |\partial\phi|(\Xi_{\mathfrak{T}}^{k-1})$$

as desired. □

The above lemma furnishes the following results including a decay estimate of  $|\partial\phi|$  (a kind of regularizing effect).

**Theorem 4.17.** *Suppose that  $\phi$  is  $\mathcal{T}_+$ -lower semicontinuous, and that Assumptions 3.16, 4.3 hold for some  $(p, \lambda)$  satisfying one of the following: (1)  $p \in (1, 2)$  and  $\lambda \geq 0$ ; (2)  $p = 2$  and  $\lambda \in \mathbb{R}$ ; (3)  $p \in (2, \infty)$  and  $\lambda = 0$ . Then, for every  $x_0 \in \mathfrak{D}(\phi)$ , each element  $\xi \in \text{GMM}_p(\phi; x_0)$  is locally Lipschitz in  $(0, \infty)$  and satisfies the following.*

(i) *For any  $t > 0$ , the right forward metric derivative*

$$|\xi'_R|(t) := \lim_{s \rightarrow t^+} \frac{d(\xi(t), \xi(s))}{s - t}$$

*exists,  $\xi(t) \in \mathfrak{D}(|\partial\phi|)$ , and*

$$\frac{d}{dt_+} \phi(\xi(t)) = -|\partial\phi|^q(\xi(t)) = -|\xi'_R|^p(t) = -|\partial\phi|(\xi(t))|\xi'_R|(t)$$

*holds, where  $\frac{d}{dt_+}$  denotes the right derivative.*

(ii)  *$\phi(\xi(t))$  is convex in  $t \geq 0$  if  $\lambda \geq 0$ , and  $e^{\lambda t^{p-1}} |\partial\phi|(\xi(t))$  is non-increasing and right continuous in  $t > 0$ . Moreover, we have*

$$\frac{t}{q} |\partial\phi|^q(\xi(t)) \leq e^{q\lambda - t^{p-1}} \{\phi(x_0) - \Phi_t(x_0)\}, \quad (4.15)$$

$$t |\partial\phi|^q(\xi(t)) \leq (1 + p\lambda_+ C(p, \lambda, t)) e^{-q\lambda t^{p-1}} \left\{ \phi(x_0) - \inf_X \phi \right\}, \quad (4.16)$$

*where  $\lambda_+ := \max\{\lambda, 0\}$  and*

$$C(p, \lambda, t) := \int_0^t s^{p-2} e^{q\lambda s^{p-1}} e^{-q \operatorname{sgn}(\lambda) |\lambda|^{q/p} s} ds.$$

*In particular,  $C(p, 0, t) = t^{p-1}/(p-1)$  and  $C(2, \lambda, t) = t$ .*

*Proof.* Take  $(\mathfrak{X}_\alpha)_{\alpha \geq 1}$  associated with  $\xi \in \text{GMM}_p(\phi; x_0)$  as in Definition 3.28. First, we shall construct a right continuous function  $\mathcal{G}_R$  on  $(0, \infty)$  such that

$$\mathcal{G}_R(t) \geq e^{\lambda t^{p-1}} |\partial\phi|(\xi(t)) \quad \text{for any } t > 0, \quad \mathcal{G}_R(t) = e^{\lambda t^{p-1}} |\partial\phi|(\xi(t)) \quad \text{for } \mathcal{L}^1\text{-a.e. } t \in (0, \infty)$$

(i.e.,  $\mathcal{G}_R$  is the right continuous representative of  $e^{\lambda t^{p-1}} |\partial\phi|(\xi(t))$ ). Define a function  $\mathcal{G}_\alpha$  on  $[0, \infty)$  by

$$\mathcal{G}_\alpha(0) := |\partial\phi|(\Xi_{\mathfrak{X}_\alpha}^0), \quad \mathcal{G}_\alpha(t) := e^{\lambda_{\mathfrak{X}_\alpha} (t_{\mathfrak{X}_\alpha}^k)^{p-1}} |\partial\phi|(\Xi_{\mathfrak{X}_\alpha}^k) \quad \text{for } t \in (t_{\mathfrak{X}_\alpha}^{k-1}, t_{\mathfrak{X}_\alpha}^k],$$

where  $\lambda_{\mathfrak{X}_\alpha}$  is as in Lemma 4.16. Owing to Lemma 4.16,  $\mathcal{G}_\alpha$  is a non-increasing function provided  $\lambda \|\mathfrak{X}_\alpha\|^{p-1} > -1$ . By [1, Lemma 3.3.3], passing to a subsequence if necessary, we can assume that  $\lim_{\alpha \rightarrow \infty} \mathcal{G}_\alpha(t)$  exists for all  $t \geq 0$  and the limit function  $\mathcal{G}$  is non-increasing. Hence,  $\mathcal{G}$  has at most countably many discontinuous points. Since  $\lambda_{\mathfrak{X}_\alpha} \rightarrow \lambda$ , we deduce from Corollary 4.9, Theorem 3.30(iii) and Remark 3.26 that  $\mathcal{G}(t) = e^{\lambda t^{p-1}} |\partial\phi|(\xi(t)) = e^{\lambda t^{p-1}} |\partial\phi|(\xi(t))$  for  $\mathcal{L}^1$ -a.e.  $t \in (0, \infty)$ . Now, define

$$\mathcal{G}_R(t) := \lim_{s \rightarrow t^+} \mathcal{G}(s) \quad \text{for } t \geq 0.$$

Then  $\mathcal{G}_R(t) \leq \mathcal{G}(t)$  for all  $t \geq 0$ , since  $\mathcal{G}$  is non-increasing. Moreover, the almost everywhere continuity of  $\mathcal{G}$  implies that  $\mathcal{G}_R$  is right continuous and  $\mathcal{G}_R(t) = \mathcal{G}(t)$  for  $\mathcal{L}^1$ -a.e.  $t \in (0, \infty)$ , and the lower semicontinuity of  $|\partial\phi|$  yields that  $\mathcal{G}_R(t) \geq e^{\lambda t^{p-1}} |\partial\phi|(\xi(t))$  for all  $t \geq 0$ . We also set  $\mathcal{S}_R(t) := e^{-\lambda t^{p-1}} \mathcal{G}_R(t)$ , which is the right continuous representative of  $|\partial\phi|(\xi(t))$ .

(i) We deduce from Theorem 3.30 and Proposition 2.24 that  $|\xi'_+|^p(t) = |\partial\phi|^q(\xi(t)) = -(\phi \circ \xi)'(t)$  for  $\mathcal{L}^1$ -a.e.  $t \in (0, \infty)$  and the energy identity (3.22) holds. Thus, for any  $t \geq 0$ , we have

$$\limsup_{\delta \rightarrow 0^+} \frac{d(\xi(t), \xi(t + \delta))}{\delta} = \limsup_{\delta \rightarrow 0^+} \frac{1}{\delta} \int_t^{t+\delta} |\partial\phi|^{q/p}(\xi(s)) ds \leq \mathcal{S}_R^q(t). \quad (4.17)$$

Given  $T > 0$ , put  $\mu := \min\{\lambda, 0\} = -\lambda_-$  and  $\eta := \inf_{t \in [0, T]} \phi(\xi(t))$ , and consider the function  $h(t) := e^{q\mu t^{p-1}} \{\phi(\xi(t)) - \eta\}$ . We claim that  $h$  is convex in  $[0, T]$ . Since  $h$  is continuous, it suffices to show that  $h'$  is  $\mathcal{L}^1$ -a.e. equal to a non-decreasing function. Note that, for  $\mathcal{L}^1$ -a.e.  $t \in (0, \infty)$ ,

$$h'(t) = -e^{q\mu t^{p-1}} |\partial\phi|^q(\xi(t)) + p\mu t^{p-2} h(t) \leq 0 \quad (4.18)$$

since  $h(t) \geq 0$  and  $\mu \leq 0$ . Thus,  $h$  is non-increasing. Now, if  $\lambda \leq 0$ , then we find  $\mu = \lambda$  and  $\mu = 0$  unless  $p = 2$ , and hence the latter term of  $h'(t)$  in (4.18) is non-decreasing. As for the former term, recall that  $\mathcal{G}(t) = e^{\lambda t^{p-1}} |\partial\phi|(\xi(t))$  for  $\mathcal{L}^1$ -a.e.  $t \in (0, \infty)$  and that  $\mathcal{G}$  is non-increasing. Therefore, we find that  $h$  is convex. In the case of  $\lambda > 0$ , Lemma 4.5(i) yields that Assumption 4.3 holds also for  $(p, 0)$ . Thus we can apply the above argument, for  $\mu = 0$  in both cases.

The convexity implies that  $h$  is right differentiable and the right derivative  $\frac{d}{dt_+} h$  is non-decreasing. Thus,  $\phi \circ \xi$  is also right differentiable and

$$\frac{d}{dt_+} \phi(\xi(t)) \leq \lim_{s \rightarrow t^+} \frac{d}{dt_+} \phi(\xi(s)).$$

By choosing a sequence  $(s_i)_{i \geq 1}$  with  $s_i \rightarrow t^+$  and  $(\phi \circ \xi)'(s_i) = -|\partial\phi|^q(\xi(s_i)) = -\mathcal{S}_R^q(s_i)$ , we find

$$\frac{d}{dt_+} \phi(\xi(t)) \leq \lim_{i \rightarrow \infty} (\phi \circ \xi)'(s_i) = -\mathcal{S}_R^q(t) \quad \text{for any } t > 0 \quad (4.19)$$

(this also implies that  $\mathcal{S}_R(t) < \infty$  for all  $t > 0$ ). A direct calculation yields

$$\begin{aligned} \frac{d}{dt_+} \phi(\xi(t)) &\geq -\liminf_{s \rightarrow t^+} \left\{ \frac{[\phi(\xi(t)) - \phi(\xi(s))]_+}{d(\xi(t), \xi(s))} \frac{d(\xi(t), \xi(s))}{s-t} \right\} \\ &\geq -\limsup_{s \rightarrow t^+} \frac{[\phi(\xi(t)) - \phi(\xi(s))]_+}{d(\xi(t), \xi(s))} \liminf_{s \rightarrow t^+} \frac{d(\xi(t), \xi(s))}{s-t} \\ &\geq -|\partial\phi|(\xi(t)) \liminf_{s \rightarrow t^+} \frac{d(\xi(t), \xi(s))}{s-t}. \end{aligned} \quad (4.20)$$

Combining this with (4.17) and (4.19) furnishes

$$\mathcal{S}_R(t) \limsup_{s \rightarrow t^+} \frac{d(\xi(t), \xi(s))}{s-t} \leq \mathcal{S}_R^q(t) \leq |\partial\phi|(\xi(t)) \liminf_{s \rightarrow t^+} \frac{d(\xi(t), \xi(s))}{s-t} \leq \mathcal{S}_R(t) \liminf_{s \rightarrow t^+} \frac{d(\xi(t), \xi(s))}{s-t}.$$

Hence,  $|\xi'_R|(t)$  exists and

$$|\xi'_R|(t) = \mathcal{S}_R^{q/p}(t) = |\partial\phi|^{q/p}(\xi(t)) < \infty$$

for all  $t > 0$ , and especially  $\xi(t) \in \mathfrak{D}(|\partial\phi|)$ . Moreover, since equality holds in (4.20), we find

$$\frac{d}{dt_+} \phi(\xi(t)) = -|\partial\phi|(\xi(t)) |\xi'_R|(t) = -|\partial\phi|^q(\xi(t)) = -|\xi'_R|^p(t).$$

Hence, (i) follows. Note also that  $\mathcal{S}_R(t) = |\partial\phi|(\xi(t))$  and  $\mathcal{G}_R(t) = e^{\lambda t^{p-1}} |\partial\phi|(\xi(t))$  hold for all  $t \geq 0$ .

(ii) First of all, if  $\lambda \geq 0$ , then the convexity of  $\phi(\xi(t))$  follows from that of  $h$ . Moreover, since  $\mathcal{G}$  is non-increasing, so is  $\mathcal{G}_R(t) = e^{\lambda t^{p-1}} |\partial\phi|(\xi(t))$ . To see (4.15), assuming  $\lambda \leq 0$ , we have

$$\frac{t}{q} e^{q\lambda t^{p-1}} |\partial\phi|^q(\xi(t)) \leq \frac{1}{q} \int_0^t e^{q\lambda s^{p-1}} |\partial\phi|^q(\xi(s)) ds \leq \frac{1}{q} \int_0^t |\partial\phi|^q(\xi(s)) ds.$$

Combining this with the energy identity (3.22), Hölder inequality and  $\xi(0) = x_0$ , we obtain

$$\begin{aligned} \frac{t}{q} e^{q\lambda t^{p-1}} |\partial\phi|^q(\xi(t)) &\leq \phi(x_0) - \phi(\xi(t)) - \frac{1}{p} \int_0^t |\xi'_+|^p(s) ds \\ &\leq \phi(x_0) - \phi(\xi(t)) - \frac{d^p(\xi(0), \xi(t))}{pt^{p-1}} \\ &\leq \phi(x_0) - \Phi_t(x_0). \end{aligned}$$

This shows (4.15). We can reduce the  $\lambda > 0$  case to the above argument with  $\lambda = 0$  by Lemma 4.5(i).

Finally, we prove (4.16). Since there is nothing to prove if  $\inf_X \phi = -\infty$ , we assume  $\eta := \inf_X \phi > -\infty$ . As  $e^{\lambda t^{p-1}} |\partial\phi|(\xi(t))$  is non-increasing, it follows from (4.13) that

$$\begin{aligned} & e^{q\lambda t^{p-1}} \{\phi(\xi(t)) - \eta\} - \{\phi(x_0) - \eta\} \\ &= \int_0^t -e^{q\lambda s^{p-1}} |\partial\phi|^q(\xi(s)) ds + p\lambda \int_0^t s^{p-2} e^{q\lambda s^{p-1}} \{\phi(\xi(s)) - \eta\} ds \\ &\leq -te^{q\lambda t^{p-1}} |\partial\phi|^q(\xi(t)) + p\lambda_+ \int_0^t s^{p-2} e^{q\lambda s^{p-1}} e^{-q \operatorname{sgn}(\lambda) |\lambda|^{q/p} s} \{\phi(x_0) - \eta\} ds. \end{aligned}$$

Therefore, we have

$$te^{q\lambda t^{p-1}} |\partial\phi|^q(\xi(t)) \leq (1 + p\lambda_+ C(p, \lambda, t)) \{\phi(x_0) - \eta\}$$

as desired. One can see  $C(p, 0, t) = t^{p-1}/(p-1)$  and  $C(2, \lambda, t) = t$  by a direct calculation.  $\square$

Note that, if  $\inf_X \phi > -\infty$ , then the convexity of  $e^{-q\lambda - t^{p-1}} \{\phi(\xi(t)) - \inf_X \phi\}$  for  $\lambda < 0$  can also be seen in the same way as above.

**Remark 4.18.** Continuing the discussion in Remark 4.12, suppose that  $t(1-t^{p-1})$  is replaced with  $t(1-t)$  in (4.1). Then, we have

$$(1 + \lambda\tau^{p-1}) |\partial\phi|^q(y_\tau) \leq \frac{p}{q} \frac{|\partial\phi|^q(x)}{(1 + \lambda\tau^{p-1})^{q/p}}$$

instead of (4.14), provided that (4.9) holds. Thus, we can follow the lines of Lemma 4.16 and Theorem 4.17 only when  $p \leq 2$  (i.e.,  $p/q \leq 1$ ).

We conclude this subsection by presenting a regularity result for gradient curves in the Finsler case.

**Corollary 4.19.** *Let  $(M, F)$  be a forward complete Finsler manifold and  $\phi : M \rightarrow \mathbb{R}$  be a lower semi-continuous,  $\lambda$ -geodesically convex function.*

- (i) *For any  $x_0 \in \mathfrak{D}(\phi)$ , there exists a curve  $\xi : [0, \infty) \rightarrow M$  of maximal slope for  $\phi$  with respect to  $|\partial\phi|$  with  $\xi(0) = x_0$ .*
- (ii) *If  $\lambda > 0$ , then  $\phi$  has a unique minimizer  $\bar{x} \in M$ , and  $|\partial\phi|(\xi(t))$  decreases to 0 and  $\xi(t) \rightarrow \bar{x}$  as  $t \rightarrow \infty$ .*
- (iii) *If  $\lambda \geq 0$  and  $\phi \in C^l(M)$  for some  $l \geq 1$ , then  $\xi$  is  $C^l$  in  $(0, T)$  with  $T := \inf\{t \geq 0 \mid \xi(t) = \inf_X \phi\}$ . Moreover, if  $\lambda > 0$  and  $T < \infty$ , then we have  $\xi(t) = \bar{x}$  for all  $t \geq T$ .*

*Proof.* (i) Thanks to Example 4.4(a), Assumption 4.3 holds for  $(2, \lambda)$ . Since  $(M, d_F)$  is forward boundedly compact by the Hopf–Rinow theorem, we find that Assumption 3.8(a), (c) hold. Then, the existence of a (2-)curve  $\xi$  of maximal slope follows from Proposition 4.10.

(ii) Note that (4.6) and Assumption 3.16 also hold due to Remarks 3.17, 4.7. Then, in view of Lemma 4.6(ii),  $\phi$  has a unique minimizer  $\bar{x} \in M$ . Moreover, on the one hand, Theorem 4.14 yields

$$d_F^2(\bar{x}, \xi(t)) \leq \frac{2}{\lambda} e^{-2\lambda t} \{\phi(x_0) - \phi(\bar{x})\}$$

for any  $t \geq 0$ , thereby  $\xi(t) \rightarrow \bar{x}$  as  $t \rightarrow \infty$ . On the other hand, it follows from Theorem 4.17(ii) that  $|\partial\phi|(\xi(t))$  is decreasing. Thus (4.16) yields

$$0 \leq \lim_{t \rightarrow \infty} |\partial\phi|^2(\xi(t)) \leq \lim_{t \rightarrow \infty} \frac{1 + 2\lambda t}{t} e^{-2\lambda t} \{\phi(x_0) - \phi(\bar{x})\} = 0.$$

(iii) When  $\phi \in C^l(M)$ , we find from Corollary 3.33 that  $\xi$  is  $C^1$  and satisfies  $\xi'(t) = \nabla(-\phi)(\xi(t))$ . Now, when  $\phi(x) > \inf_X \phi$ , we deduce  $F(\nabla(-\phi)(x)) = |\partial\phi|(x) > 0$  from the  $\lambda$ -convexity along a minimal geodesic from  $x$  to some  $y$  with  $\phi(y) < \phi(x)$ . This implies that  $\nabla(-\phi)$  is a  $C^{l-1}$ -vector field around  $\xi(t)$  for any  $t \in (0, T)$  (recall Remark 3.34), and hence  $\xi$  is  $C^l$  in  $(0, T)$  by  $\xi'(t) = \nabla(-\phi)(\xi(t))$ . In the case of  $\lambda > 0$  and  $T < \infty$ , we see that  $\xi(t) = \bar{x}$  for all  $t \geq T$  since  $\phi \circ \xi$  is non-increasing.  $\square$

We remark that the convergences in (ii) are exponentially fast. Note also that the argument in (iii) shows that  $\phi$  is of lower 2-growth for general  $\lambda \in \mathbb{R}$  (or of lower  $p$ -growth if  $\phi$  is  $(p, \lambda)$ -geodesically convex). Indeed, given  $x \in \mathfrak{D}(|\partial\phi|)$  and any  $y \in M$ , the  $\lambda$ -convexity along a minimal geodesic  $\gamma : [0, 1] \rightarrow M$  from  $x$  to  $y$  yields

$$\begin{aligned} \phi(y) &\geq \phi(x) + \lim_{t \rightarrow 0} \frac{\phi(\gamma(t)) - \phi(x)}{t} + \frac{\lambda}{2} d^2(x, y) \geq \phi(x) - |\partial\phi|(x) d(x, y) + \frac{\lambda}{2} d^2(x, y) \\ &\geq \phi(x) - \frac{1}{2} |\partial\phi|^2(x) + \frac{\lambda - 1}{2} d^2(x, y). \end{aligned}$$

**4.4. Heat flow on compact Finsler manifolds.** This subsection is devoted to a study of heat flow on compact Finsler manifolds. It is well known that heat flow can be regarded as gradient flow in the Wasserstein space; we refer to [7, 24, 32, 44] for more details. In what follows, let  $(M, F, \mathfrak{m})$  be a compact Finsler manifold endowed with a smooth positive measure  $\mathfrak{m}$ . Along [7], we shall present a slightly more general framework than [32].

**4.4.1. Finsler structure of the Wasserstein space.** Denote by  $\mathcal{P}(M)$  the collection of Borel probability measures on  $M$ . For  $\mu_0, \mu_1 \in \mathcal{P}(M)$ , the  $L^2$ -Wasserstein distance is defined by

$$d_W(\mu_0, \mu_1) := \inf_{\pi} \left( \int_{M \times M} d_F^2(x, y) \pi(dx dy) \right)^{1/2},$$

where  $\pi \in \mathcal{P}(M \times M)$  runs over all couplings of  $(\mu_0, \mu_1)$ . The fundamental theory of optimal transport yields that the  $L^2$ -Wasserstein space  $(\mathcal{P}(M), d_W)$  is a compact  $\lambda_F(M)$ -metric space and any  $\mu_0, \mu_1 \in \mathcal{P}(M)$  admit a minimal geodesic  $(\mu_t)_{t \in [0, 1]}$  from  $\mu_0$  to  $\mu_1$  (cf. [15, Corollary 4.17]).

Let  $\mathcal{P}_{\text{ac}}(M; \mathfrak{m}) \subset \mathcal{P}(M)$  be the set of measures absolutely continuous with respect to  $\mathfrak{m}$ . Then we have the following standard fact (cf. [25, Theorem 4.10]).

**Proposition 4.20.** *For any  $\mu_0 \in \mathcal{P}_{\text{ac}}(M; \mathfrak{m})$  and  $\mu_1 \in \mathcal{P}(M)$ , there exists a  $\mu_0$ -a.e. unique vector field  $\Psi_{\mu_0}^{\mu_1}$  on  $M$  such that*

- (i)  $\pi = (\text{Id}_M, \exp \Psi_{\mu_0}^{\mu_1})_{\#} \mu_0$  is a unique optimal coupling of  $(\mu_0, \mu_1)$ ;
- (ii)  $\mu_t = [\exp(t\Psi_{\mu_0}^{\mu_1})]_{\#} \mu_0$ ,  $t \in [0, 1]$ , is a unique minimal geodesic from  $\mu_0$  to  $\mu_1$ ;
- (iii)  $d_W^2(\mu_0, \mu_1) = \int_M F^2(\Psi_{\mu_0}^{\mu_1}) d\mu_0$ ;
- (iv) there exists a  $(d_F^2/2)$ -convex function  $\varphi : M \rightarrow \mathbb{R}$  such that  $\Psi_{\mu_0}^{\mu_1} = \nabla\varphi$ ,

where  $f_{\#} \mu$  denotes the push-forward measure of  $\mu$  by  $f$ .

The vector field  $\Psi_{\mu_0}^{\mu_1}$  will be referred to as the *optimal transport vector field*. Observe that  $\Psi_{\mu_0}^{\mu_1} = t\Psi_{\mu_0}^{\mu_1}$  holds for all  $t \in [0, 1]$ . By [26], the function  $\varphi$  in (iv) is Lipschitz and twice differentiable  $\mathfrak{m}$ -a.e.

Now we recall the Finsler structure of  $\mathcal{P}(M)$  introduced in [32] (see [7, 44] for the Riemannian case). For  $\mu \in \mathcal{P}(M)$ , let  $L^2(\mu; TM)$  denote the space of measurable vector fields  $\mathbf{w}$  with the asymmetric norm

$$\|\mathbf{w}\|_{\mu} := \left( \int_M F^2(\mathbf{w}) d\mu \right)^{1/2} < \infty.$$

We similarly define  $L^2(\mu; T^*M)$  as the space of measurable 1-forms  $\zeta$  with

$$\|\zeta\|_{\mu}^* := \left( \int_M F^*(\zeta)^2 d\mu \right)^{1/2} < \infty.$$

We also define

$$\langle \zeta, \mathbf{w} \rangle_{\mu} := \int_M \zeta(\mathbf{w}) d\mu$$

for  $\mathbf{w} \in L^2(\mu; TM)$  and  $\zeta \in L^2(\mu; T^*M)$ , and the *Legendre transformation*  $\mathfrak{L}_{\mu} : L^2(\mu; TM) \rightarrow L^2(\mu; T^*M)$  by applying the pointwise Legendre transformation induced by  $F$  (recall Subsection 2.3). The *tangent* and *cotangent spaces* of  $\mathcal{P}(M)$  at  $\mu$  are defined by

$$T_{\mu} \mathcal{P}(M) := \overline{\{\nabla\varphi \mid \varphi \in C^{\infty}(M)\}}^{\|\cdot\|_{\mu}}, \quad T_{\mu}^* \mathcal{P}(M) := \overline{\{d\varphi \mid \varphi \in C^{\infty}(M)\}}^{\|\cdot\|_{\mu}^*},$$

respectively. We summarize some properties of these spaces (see Appendix C for the proof).

**Proposition 4.21.** *Given  $\mu \in \mathcal{P}(M)$ , we have the following.*

- (i) *The Legendre transformation  $\mathfrak{L}_\mu : L^2(\mu; TM) \rightarrow L^2(\mu; T^*M)$  is a homeomorphism. In particular, its restriction  $\mathfrak{L}_\mu : T_\mu\mathcal{P}(M) \rightarrow T_\mu^*\mathcal{P}(M)$  is also a homeomorphism.*
- (ii) *For any  $\zeta \in L^2(\mu; T^*M)$ , we have  $\|\zeta\|_\mu^* = \sup_{\mathbf{w} \in L^2(\mu; TM) \setminus \{0\}} \langle \zeta, \mathbf{w} \rangle_\mu / \|\mathbf{w}\|_\mu$  and  $\mathbf{w} = \mathfrak{L}_\mu^{-1}(\zeta)$  is a unique element in  $L^2(\mu; TM)$  satisfying  $\|\mathbf{w}\|_\mu^2 = (\|\zeta\|_\mu^*)^2 = \langle \zeta, \mathbf{w} \rangle_\mu$ .*
- (iii) *For every bounded linear functional  $\mathcal{Q}$  on  $L^2(\mu; TM)$ , there exists a unique 1-form  $\zeta \in L^2(\mu; T^*M)$  such that  $\mathcal{Q}(\mathbf{w}) = \langle \zeta, \mathbf{w} \rangle_\mu$  for all  $\mathbf{w} \in L^2(\mu; TM)$ .*
- (iv) *We have  $T_\mu^*\mathcal{P}(M) = \{\zeta \in L^2(\mu; T^*M) \mid \langle \zeta, \mathbf{w} \rangle_\mu = 0 \text{ for all } \mathbf{w} \in \mathbf{Ker}(\mathbf{div})(\mu)\}$ , where*

$$\mathbf{Ker}(\mathbf{div})(\mu) := \{\mathbf{w} \in L^2(\mu; TM) \mid \langle d\varphi, \mathbf{w} \rangle_\mu = 0 \text{ for all } \varphi \in C^\infty(M)\}.$$

- (v) *If  $\mathbf{v}, \mathbf{w} \in T_\mu\mathcal{P}(M)$  satisfy  $\mathbf{v} - \mathbf{w} \in \mathbf{Ker}(\mathbf{div})(\mu)$ , then we have  $\mathbf{v} = \mathbf{w}$ .*

The next theorem is seen in the same manner as [32, Lemma 7.2, Theorem 7.3] or [7, Proposition 2.5]. (In fact, (4.21) does not depend on the metric and can be reduced to any bi-Lipschitz equivalent Riemannian metric.)

**Theorem 4.22.** *Let  $I \subset \mathbb{R}$  be an open interval and  $\mu = (\mu_t)_{t \in I}$  be a continuous curve in  $\mathcal{P}(M)$ .*

- (i) *If  $\mu \in \mathbf{FAC}_{\text{loc}}^2(I; \mathcal{P}(M))$ , then there exists a Borel vector field  $\mathbf{v} : I \times M \rightarrow TM$  with  $\|\mathbf{v}_t\|_{\mu_t} \in L_{\text{loc}}^2(I)$  ( $\mathbf{v}_t(x) := \mathbf{v}(t, x) \in T_x M$ ) such that*
  - (a)  $\mathbf{v}_t \in T_{\mu_t}\mathcal{P}(M)$  for  $\mathcal{L}^1$ -a.e.  $t \in I$ ;
  - (b) *the continuity equation  $\partial_t \mu_t + \mathbf{div}(\mathbf{v}_t \cdot \mu_t) = 0$  holds in the sense of distributions, i.e.,*

$$\int_I \int_M \{\partial_t \varphi(t, x) + \langle d\varphi(t, x), \mathbf{v}_t(x) \rangle\} \mu_t(dx) dt = 0 \quad \text{for all } \varphi \in C_0^\infty(I \times M), \quad (4.21)$$

where  $C_0^\infty(I \times M)$  denotes the set of  $C^\infty$ -functions on  $I \times M$  of compact support.

Moreover, such a vector field  $\mathbf{v}_t$  is unique and satisfies  $\|\mathbf{v}_t\|_{\mu_t} = |\mu_t'|_+(t)$  for  $\mathcal{L}^1$ -a.e.  $t \in I$ .

- (ii) *If  $\mu$  satisfies (4.21) above for some Borel vector field  $(\mathbf{v}_t)_{t \in I}$  with  $\|\mathbf{v}_t\|_{\mu_t} \in L_{\text{loc}}^2(I)$ , then  $\mu$  is locally forward absolutely continuous with  $|\mu_t'|_+(t) \leq \|\mathbf{v}_t\|_{\mu_t}$ .*

The vector field  $\mathbf{v} = (\mathbf{v}_t)_{t \in I}$  in (i) above is called the *tangent vector field* of the curve  $\mu$ .

**Lemma 4.23.** (i) *For any  $\mu_0 \in \mathcal{P}_{\text{ac}}(M; \mathfrak{m})$  and  $\mu_1 \in \mathcal{P}(M)$ , we have  $\Psi_{\mu_0}^{\mu_1} \in T_{\mu_0}\mathcal{P}(M)$ .*

- (ii) *For any  $(\mu_t)_{t \in I} \in \mathbf{FAC}_{\text{loc}}^2(I; \mathcal{P}_{\text{ac}}(M; \mathfrak{m}))$  and its tangent vector field  $(\mathbf{v}_t)_{t \in I}$ , we have, for  $\mathcal{L}^1$ -a.e.  $t \in I$ ,*

$$\frac{1}{h} \Psi_{\mu_t}^{\mu_t+h} \rightarrow \mathbf{v}_t \text{ weakly in } L^2(\mu_t; TM) \text{ as } h \rightarrow 0^+. \quad (4.22)$$

*Proof.* Thanks to Proposition 4.20(iv), one can show (i) via an approximation by smooth functions. (ii) follows from a similar argument to the proof of [7, Lemma 2.7] along with Proposition 4.21(v), Theorem 4.22 and (i) above.  $\square$

4.4.2. *Subdifferentials and gradient flows.* Let  $\phi : \mathcal{P}(M) \rightarrow (-\infty, \infty]$  be a  $\mathcal{T}_+$ -lower semicontinuous function on the Wasserstein space  $(\mathcal{P}(M), d_W)$  with  $\mathfrak{D}(\phi) \subset \mathcal{P}_{\text{ac}}(M; \mathfrak{m})$ .

**Definition 4.24** (Subdifferentials). For  $\mu \in \mathfrak{D}(\phi)$ , a 1-form  $\zeta \in L^2(\mu; T^*M)$  is said to *belong to the subdifferential*  $\partial\phi(\mu)$  if

$$\phi(\nu) - \phi(\mu) \geq \langle \zeta, \Psi_\mu^\nu \rangle_\mu + o(d_W(\mu, \nu)) \quad \text{for all } \nu \in \mathcal{P}(M).$$

We call  $\zeta \in \partial\phi(\mu)$  a *strong subdifferential* if it satisfies

$$\phi((\exp \mathbf{w})_\# \mu) - \phi(\mu) \geq \langle \zeta, \mathbf{w} \rangle_\mu + o(\|\mathbf{w}\|_\mu) \quad \text{for all } \mathbf{w} \in L^2(\mu; TM).$$

**Proposition 4.25.** *Given  $\mu \in \mathfrak{D}(\phi)$ , we have the following.*

- (i) *If  $\partial\phi(\mu) \neq \emptyset$ , then  $|\partial\phi(\mu)| \leq \inf_{\zeta \in \partial\phi(\mu)} \|\zeta\|_\mu^*$ .*
- (ii) *If  $\zeta \in \partial\phi(\mu) \cap T_\mu^*\mathcal{P}(M)$ , then  $\zeta$  is a strong subdifferential.*

(iii) If  $\phi$  is  $\lambda$ -geodesically convex, then  $\zeta \in L^2(\mu; T^*M)$  belongs to  $\partial\phi(\mu)$  if and only if

$$\phi(\nu) - \phi(\mu) \geq \langle \zeta, \Psi_\mu^\nu \rangle + \frac{\lambda}{2} d_W^2(\mu, \nu) \quad \text{for all } \nu \in \mathcal{P}(M). \quad (4.23)$$

*Proof.* Definitions 2.19 and 4.24 directly yield (i) directly. (ii) and (iii) follow from the same arguments as in [7, Lemmas 3.2, 3.5], respectively.  $\square$

**Proposition 4.26.** *Let  $I \subset \mathbb{R}$  be an open interval and  $(\mu_t)_{t \in I} \in \text{FAC}^2(I; \mathfrak{D}(\phi))$  with the tangent vector field  $(\mathbf{v}_t)_{t \in I}$ . If  $\phi$  is  $\lambda$ -geodesically convex and  $\int_I |\partial\phi|(\mu_t) |\mu'_+|(t) dt < \infty$ , then  $t \mapsto \phi(\mu_t)$  is absolutely continuous in  $I$  and we have, for  $\mathcal{L}^1$ -a.e.  $t \in I$ ,*

$$\frac{d}{dt} \phi(\mu_t) = \langle \zeta, \mathbf{v}_t \rangle_{\mu_t} \quad \text{for all } \zeta \in \partial\phi(\mu_t).$$

*Proof.* The absolute continuity of  $\phi(\mu_t)$  follows from Example 4.4(a), Corollary 4.9 and Remark 2.16. It follows from the assumption and Lemma 4.23(ii) that, for  $\mathcal{L}^1$ -a.e.  $t \in I$ ,  $|\partial\phi|(\mu_t) < \infty$ ,  $s \mapsto \phi(\mu_s)$  is differentiable at  $t$ , and (4.22) holds. Then, for any  $\zeta \in \partial\phi(\mu_t)$ , we deduce from (4.23) that

$$\begin{aligned} \phi(\mu_{t+h}) - \phi(\mu_t) &\geq \langle \zeta, \Psi_{\mu_t}^{\mu_{t+h}} \rangle_{\mu_t} + \frac{\lambda}{2} d_W^2(\mu_t, \mu_{t+h}) \\ &= h \left\langle \zeta, \frac{1}{h} \Psi_{\mu_t}^{\mu_{t+h}} - \mathbf{v}_t \right\rangle_{\mu_t} + h \langle \zeta, \mathbf{v}_t \rangle_{\mu_t} + \frac{\lambda}{2} d_W^2(\mu_t, \mu_{t+h}) \\ &= h \langle \zeta, \mathbf{v}_t \rangle_{\mu_t} + o(h) \end{aligned}$$

for  $h > 0$ . This implies

$$\frac{d}{dt_+} \phi(\mu_t) \geq \langle \zeta, \mathbf{v}_t \rangle_{\mu_t}.$$

Moreover, by considering the reverse curve  $\bar{\mu}_t := \mu_{-t}$  with the tangent vector field  $\bar{\mathbf{v}}_t := -\mathbf{v}_{-t}$  for the reverse Finsler structure  $\bar{F}(v) := F(-v)$ , we also obtain

$$-\frac{d}{dt_-} \phi(\mu_t) \geq -\langle \zeta, \mathbf{v}_t \rangle_{\mu_t}.$$

This completes the proof.  $\square$

Now, we introduce gradient flows in the Wasserstein space in the same spirit as [7, Definition 3.7] (slightly weaker than [32, Definition 7.6]), which is compatible with the notion of (2-)curves of maximal slope as we will see in Proposition 4.29 below.

**Definition 4.27** (Gradient flows in the Wasserstein space). A curve  $(\mu_t)_{t>0} \in \text{FAC}_{\text{loc}}^2((0, \infty); \mathcal{P}(M))$  with the tangent vector field  $(\mathbf{v}_t)_{t>0}$  is called a *trajectory of the gradient flow* for  $\phi$  if

$$-\mathfrak{L}_{\mu_t}(\mathbf{v}_t) \in \partial\phi(\mu_t) \quad \text{for } \mathcal{L}^1\text{-a.e. } t \in (0, \infty). \quad (4.24)$$

**Remark 4.28.** Owing to Theorem 4.22(i) and Proposition 4.25(i), we have

$$|\mu'_+|(t) |\partial\phi|(\mu_t) \leq \|\mathbf{v}_t\|_{\mu_t} \|\mathfrak{L}_{\mu_t}(\mathbf{v}_t)\|_{\mu_t}^* = \|\mathbf{v}_t\|_{\mu_t}^2 \in L_{\text{loc}}^1((0, \infty)).$$

Thus, it follows from Proposition 4.26 (provided that  $\phi$  is  $\lambda$ -geodesically convex) that  $t \mapsto \phi(\mu_t)$  is locally absolutely continuous in  $(0, \infty)$  and

$$-\frac{d}{dt} \phi(\mu_t) = \|\mathbf{v}_t\|_{\mu_t}^2 \quad \text{for } \mathcal{L}^1\text{-a.e. } t \in (0, \infty),$$

which implies  $\phi(\mu_t) < \infty$  for all  $t > 0$  and the energy identity

$$\phi(\mu_r) - \phi(\mu_s) = \int_r^s \|\mathbf{v}_t\|_{\mu_t}^2 dt \quad \text{for all } 0 < r < s.$$

**Proposition 4.29.** *Let  $\phi$  be a  $\lambda$ -geodesically convex function such that, for every  $\mu \in \mathfrak{D}(|\partial\phi|)$ , there is  $\zeta \in \partial\phi(\mu)$  with  $\|-\zeta\|_\mu^* = |\partial\phi|(\mu)$ . Then,  $(\mu_t)_{t>0} \in \text{FAC}_{\text{loc}}^2((0, \infty); \mathcal{P}(M))$  is a curve of maximal slope for  $\phi$  with respect to  $|\partial\phi|$  if and only if it is a trajectory of the gradient flow for  $\phi$ .*

*Proof.* First, suppose that  $(\mu_t)_{t>0}$  is a trajectory of the gradient flow for  $\phi$ . Since  $|\partial\phi|$  is a weak upper gradient for  $-\phi$  by Theorem 2.21(i), it follows from Definition 2.17 together with Remark 4.28 that

$$-\|\mathbf{v}_t\|_{\mu_t}^2 = \frac{d}{dt}\phi(\mu_t) \geq -|\partial\phi|(\mu_t)|\mu'_+|(t) \geq -\|\mathbf{v}_t\|_{\mu_t}^2$$

for  $\mathcal{L}^1$ -a.e.  $t \in (0, \infty)$ . This implies

$$\frac{d}{dt}\phi(\mu_t) = -\frac{1}{2}|\mu'_+|^2(t) - \frac{1}{2}|\partial\phi|^2(\mu_t), \quad |\partial\phi|(\mu_t) = \|\mathbf{v}_t\|_{\mu_t} = |\mu'_+|(t). \quad (4.25)$$

Thus,  $(\mu_t)_{t>0}$  is a curve of maximal slope for  $\phi$  with respect to  $|\partial\phi|$ .

Conversely, if  $(\mu_t)_{t>0}$  is a curve of maximal slope with the tangent vector field  $(\mathbf{v}_t)_{t>0}$ , then (4.25) holds (see (2.12)) and we have

$$\int_r^s |\partial\phi|(\mu_t)|\mu'_+|(t) dt = \int_r^s |\mu'_+|^2(t) dt < \infty \quad \text{for all } 0 < r \leq s < \infty.$$

By hypothesis, there exists  $\zeta_t \in \partial\phi(\mu_t)$  with  $\|-\zeta_t\|_{\mu_t}^* = |\partial\phi|(\mu_t)$ . Then Proposition 4.26 yields

$$|\partial\phi|(\mu_t)|\mu'_+|(t) = -\frac{d}{dt}\phi(\mu_t) = -\langle \zeta_t, \mathbf{v}_t \rangle_{\mu_t} \leq \|-\zeta_t\|_{\mu_t}^* \|\mathbf{v}_t\|_{\mu_t} = |\partial\phi|(\mu_t)|\mu'_+|(t),$$

which implies  $-\mathfrak{L}_{\mu_t}(\mathbf{v}_t) = \zeta_t \in \partial\phi(\mu_t)$  as desired.  $\square$

**Remark 4.30.** Although we need only the  $p = 2$  case in the analysis of heat flow, one can generalize the concepts and results in this subsection to the  $L^p$ -Wasserstein distance with  $p > 1$  by suitable modifications. For instance, it follows from [14, Theorem 3.4] that Proposition 4.20 remains valid by replacing  $\nabla\varphi$  with  $F^{q-2}(\nabla\varphi)\nabla\varphi$ , where  $q$  is the conjugate exponent of  $p$ . The  $L^p$ -Wasserstein distance can be used to analyze the gradient flow structure of the  $q$ -heat flow as studied in the symmetric case by Kell [13]; one could apply our method to generalize it to the asymmetric setting.

4.4.3. *Gradient flow for the relative entropy.* The relative entropy  $H_{\mathbf{m}} : \mathcal{P}(M) \rightarrow (-\infty, \infty]$  with respect to  $\mathbf{m}$  is defined by

$$H_{\mathbf{m}}(\mu) := \int_M \rho \log \rho \, d\mathbf{m}$$

if  $\mu = \rho \mathbf{m}$  and  $[\rho \log \rho]_+$  is integrable; otherwise we set  $H_{\mathbf{m}}(\mu) := +\infty$ . In particular,  $\mathfrak{D}(H_{\mathbf{m}}) \subset \mathcal{P}_{\text{ac}}(M; \mathbf{m})$ . We recall some basic properties (see [25, 32]).

- Proposition 4.31.** (i)  $H_{\mathbf{m}}(\mu) \geq -\log \mathbf{m}(M)$  for any  $\mu \in \mathcal{P}(M)$ ;  
(ii)  $H_{\mathbf{m}}$  is  $\mathcal{T}_+$ -lower semicontinuous;  
(iii) There exists some  $K \in \mathbb{R}$  such that  $H_{\mathbf{m}}$  is  $K$ -geodesically convex.

The Sobolev space  $W^{1,1}(M)$  is defined as the closure of  $C^\infty(M)$  with respect to the asymmetric norm  $\|u\|_{W^{1,1}} := \int_M (|u| + F^*(du)) \, d\mathbf{m}$ . The proof of the following key proposition is postponed to Appendix C.

**Proposition 4.32.** For  $\mu = \rho \mathbf{m} \in \mathfrak{D}(H_{\mathbf{m}})$ , the following are equivalent:

- (I)  $|\partial H_{\mathbf{m}}|(\mu) < \infty$ ;  
(II)  $\rho \in W^{1,1}(M)$  with  $d\rho = \rho\zeta$  for some  $\zeta \in L^2(\mu; T^*M)$ .

In this case,  $\zeta \in \partial H_{\mathbf{m}}(\mu) \cap T_\mu^*\mathcal{P}(M)$  and is a unique strong subdifferential with  $|\partial H_{\mathbf{m}}|(\mu) = \|-\zeta\|_\mu^*$ .

The following existence result follows from Proposition 4.10 with the help of Propositions 4.29, 4.32.

**Theorem 4.33.** For any  $\mu \in \mathfrak{D}(H_{\mathbf{m}})$ , there exists a trajectory  $(\mu_t)_{t \geq 0}$  of the gradient flow for  $H_{\mathbf{m}}$  with  $\mu_0 = \mu$ .

For  $u \in W^{1,1}(M)$ , the distributional Laplacian  $\Delta_{\mathbf{m}}u$  is defined as

$$\int_M \varphi \Delta_{\mathbf{m}}u \, d\mathbf{m} := - \int_M \langle d\varphi, \nabla u \rangle \, d\mathbf{m} \quad \text{for all } \varphi \in C^\infty(M).$$

Note that the Laplacian  $\Delta_{\mathbf{m}}$  is a nonlinear operator (unless  $F$  comes from a Riemannian metric).

**Definition 4.34** (Weak solutions to the heat equation). We say that  $u \in L^1_{\text{loc}}((0, \infty); W^{1,1}(M))$  is a *weak solution to the heat equation*  $\partial_t u_t = \Delta_{\mathbf{m}} u_t$  if

$$\int_0^\infty \int_M \{u_t \cdot \partial_t \varphi - \langle d\varphi, \nabla u_t \rangle\} \, \text{d}\mathbf{m} \, dt = 0 \quad \text{for all } \varphi \in C_0^\infty((0, \infty) \times M).$$

**Remark 4.35.** The  $W^{1,1}$ -regularity along with (4.27) below is the minimum regularity to which our characterization result applies. In the Riemannian case, it is well known that such a weak solution has a smooth version (e.g., by using the heat kernel). In the Finsler setting, however, there is no heat kernel due to the nonlinearity and it seems unclear (to the authors) if a solution to the heat equation given in Theorem 4.36(II) automatically enjoys a higher regularity (e.g.,  $C^{1,\alpha}$  as in [9, 32]).

**Theorem 4.36.** *Let  $(\mu_t)_{t \geq 0}$  be a continuous curve in  $(\mathcal{P}(M), d_W)$ . Then the following are equivalent:*

- (I)  $(\mu_t)_{t > 0}$  is a trajectory of the gradient flow for  $H_{\mathbf{m}}$ ;
- (II)  $\mu_t$  is given by  $\mu_t = \rho_t \mathbf{m} \in \mathcal{P}_{\text{ac}}(M; \mathbf{m})$  for  $t > 0$  and  $(-\rho_t)_{t > 0}$  is a weak solution to the heat equation

$$\partial_t(-\rho_t) = \Delta_{\mathbf{m}}(-\rho_t) \tag{4.26}$$

satisfying

$$H_{\mathbf{m}}(\rho_t \mathbf{m}) < \infty \text{ for all } t > 0, \quad \int_{t_0}^{t_1} \int_M \frac{F^2(\nabla(-\rho_t))}{\rho_t} \, \text{d}\mathbf{m} \, dt < \infty \text{ for all } 0 < t_0 < t_1. \tag{4.27}$$

*Proof.* (I)  $\Rightarrow$  (II) Let  $(\mathbf{v}_t)_{t > 0}$  be the tangent vector field of  $(\mu_t)_{t > 0}$ . Recall from Remark 4.28 that  $H_{\mathbf{m}}(\mu_t) < \infty$  holds for all  $t > 0$ , and hence, we have  $\mu_t = \rho_t \mathbf{m} \in \mathfrak{D}(H_{\mathbf{m}})$ . Moreover, we deduce from Proposition 4.21(i) and (4.24) that  $-\mathfrak{L}_{\mu_t}(\mathbf{v}_t) \in \partial H_{\mathbf{m}}(\mu_t) \cap T_{\mu_t}^* \mathcal{P}(M)$  for  $\mathcal{L}^1$ -a.e.  $t \in (0, \infty)$ . Then it follows from Proposition 4.25 that  $-\mathfrak{L}_{\mu_t}(\mathbf{v}_t)$  is a strong subdifferential and

$$|\partial H_{\mathbf{m}}|(\mu_t) \leq \|\mathfrak{L}_{\mu_t}(\mathbf{v}_t)\|_{\mu_t}^* = \|\mathbf{v}_t\|_{\mu_t} < \infty.$$

Hence, Proposition 4.32 yields that  $\rho_t \in W^{1,1}(M)$  and  $-\mathfrak{L}_{\mu_t}(\mathbf{v}_t) = d\rho_t/\rho_t$  for  $\mathcal{L}^1$ -a.e.  $t \in (0, \infty)$ . Note that  $\mathbf{v}_t = \mathfrak{L}_{\mu_t}^{-1}(-d\rho_t/\rho_t) = \nabla(-\rho_t)/\rho_t$  and then the continuity equation (4.21) is exactly the heat equation (4.26). Moreover, thanks to  $\|\mathbf{v}_t\|_{\mu_t} \in L^2_{\text{loc}}((0, \infty))$ , for any  $0 < t_0 < t_1$ , we have

$$\int_{t_0}^{t_1} \int_M \frac{F^2(\nabla(-\rho_t))}{\rho_t} \, \text{d}\mathbf{m} \, dt = \int_{t_0}^{t_1} \int_M F^2(\mathbf{v}_t) \, d\mu_t \, dt = \int_{t_0}^{t_1} \|\mathbf{v}_t\|_{\mu_t}^2 \, dt < \infty.$$

(II)  $\Rightarrow$  (I) Note again that the heat equation (4.26) is equivalent to the continuity equation (4.21) for  $\mu_t$  with  $\mathbf{v}_t := \nabla(-\rho_t)/\rho_t$ . Then the assumption (4.27) implies  $\mathbf{v}_t \in L^2(\mu_t; TM)$  for  $\mathcal{L}^1$ -a.e.  $t \in (0, \infty)$ . Now it follows from Proposition 4.22(ii) that  $(\mu_t)_{t > 0} \in \text{FAC}_{\text{loc}}^2((0, \infty); \mathcal{P}(M))$ . Moreover, the Hölder inequality furnishes

$$\frac{\|\rho_t\|_{W^{1,1}}}{\lambda_F(M)} \leq \|-\rho_t\|_{W^{1,1}} = 1 + \int_M \frac{F(\nabla(-\rho_t))}{\rho_t} \, d\mu_t \leq 1 + \|\mathbf{v}_t\|_{\mu_t} < \infty,$$

thereby  $\rho_t \in W^{1,1}(M)$ . We deduce from Proposition 4.32 that  $\zeta_t := -\mathfrak{L}_{\mu_t}(\mathbf{v}_t) = d\rho_t/\rho_t$  belongs to  $\partial H_{\mathbf{m}}(\mu_t) \cap T_{\mu_t}^* \mathcal{P}(M)$  and is a unique strong subdifferential at  $\mu_t$ . Thus  $\mathbf{v}_t \in T_{\mu_t} \mathcal{P}(M)$ , and  $(\mathbf{v}_t)_{t > 0}$  is the tangent vector field of  $(\mu_t)_{t > 0}$  (recall Proposition 4.22(i)). Since  $-\mathfrak{L}_{\mu_t}(\mathbf{v}_t) = \zeta_t \in \partial H_{\mathbf{m}}(\mu_t)$  for  $\mathcal{L}^1$ -a.e.  $t \in (0, \infty)$ , we conclude that  $(\mu_t)_{t > 0}$  is a trajectory of the gradient flow for  $H_{\mathbf{m}}$ .  $\square$

We remark that, due to the irreversibility of  $F$ ,  $(\rho_t)_{t > 0}$  in Theorem 4.36 is not necessarily a weak solution to the heat equation. Instead, it is a weak solution to the heat equation with respect to the reverse Finsler structure  $\overleftarrow{F}(v) := F(-v)$ . We denote by  $\overleftarrow{\Delta}_{\mathbf{m}}$  the Laplacian for  $(M, \overleftarrow{F}, \mathbf{m})$ .

**Corollary 4.37.** (i) *For any  $u \in L^2(M)$  with  $\sup_M u < \infty$ , there exists a weak solution  $(u_t)_{t \geq 0}$  to the heat equation  $\partial_t u_t = \Delta_{\mathbf{m}} u_t$  with  $u_0 = u$ .*  
(ii) *For any  $u \in L^2(M)$  with  $\inf_M u > -\infty$ , there exists a weak solution  $(u_t)_{t \geq 0}$  to the heat equation  $\partial_t u_t = \overleftarrow{\Delta}_{\mathbf{m}} u_t$  with respect to  $\overleftarrow{F}$  with  $u_0 = u$ .*

*Proof.* (i) Suppose that  $u$  is not constant and set  $\rho := (-u + \sup u)/\|u - \sup u\|_{L^1}$ . Then  $\mu := \rho \mathbf{m} \in \mathcal{P}_{\text{ac}}(M; \mathbf{m})$  and, moreover,  $\mu \in \mathcal{D}(H_{\mathbf{m}})$  (see, e.g., the proof of [27, Lemma 16.2]). Hence, there exists a trajectory  $\mu_t = \rho_t \mathbf{m}$  of the gradient flow for  $H_{\mathbf{m}}$  with  $\mu_0 = \mu$  by Theorem 4.33, and  $(-\rho_t)_{t>0}$  is a weak solution to the heat equation by Theorem 4.36. This completes the proof by letting  $u_t := \sup u - \|u - \sup u\|_{L^1} \cdot \rho_t$ .

(ii) Thanks to  $\overleftarrow{\Delta}_{\mathbf{m}} f = -\Delta_{\mathbf{m}}(-f)$ , this is shown in the same way as (i).  $\square$

**Remark 4.38.** In [32, §3], we constructed a weak solution to the heat equation starting from  $u \in H^1(M)$  as a trajectory of the gradient flow for the energy in the Hilbert space  $L^2(M)$ . Then, by the  $L^2$ -contraction property, we can extend it to a contraction semigroup acting on  $L^2(M)$ . In contrast, Corollary 4.37 provides a direct construction for (bounded)  $u \in L^2(M)$ . Because of the lack of a higher regularity (Remark 4.35) and the Wasserstein contraction (see [33] and Subsection 4.5 below), it seems unclear if these flows starting from  $u \in L^2(M) \setminus H^1(M)$  coincide.

**4.5. Further problems and related works.** In ‘‘Riemannian-like’’ spaces, one can proceed to the *contraction property* asserting that two gradient curves emanating from different points are getting closer with an exponential rate depending on the convexity of  $\phi$ :

$$d(\xi_1(t), \xi_2(t)) \leq e^{-\lambda t} d(\xi_1(0), \xi_2(0))$$

(see, e.g., [1, Theorem 4.0.4(iv)]). This property is, however, known to fail in Finsler-like spaces (see [33]). It is an intriguing open problem if one can obtain any weaker contraction estimate for convex functions on Finsler manifolds or normed spaces. An important apparatus to study the contraction property is the *evolution variational inequality* (see [1, 23]), which also forces the space to be Riemannian.

The contraction property plays a vital role in the study of heat flow. In view of Section 4.4 and [32], heat flow can be regarded as the gradient flow of the relative entropy in the  $L^2$ -Wasserstein space, and the convexity of the relative entropy is equivalent to the associated lower Ricci curvature bound (see [25, 35, 44]). Moreover, the contraction property of heat flow is equivalent to the corresponding gradient estimate (see [16]). Though we know some gradient estimates in the Finsler setting (see [27, 28, 34]), the lack of the Riemannian-like structure prevents us to obtain a contraction property of heat flow (precisely, the nonlinearity of the Finsler Laplacian causes an essential difference). Therefore, generalizing the contraction property to the Finsler setting will make a breakthrough also in the study of heat flow on Finsler manifolds.

We refer to [29] for a recent work on discrete-time gradient flows in Gromov hyperbolic spaces; note that some non-Riemannian Finsler manifolds can be Gromov hyperbolic. Let us also mention another related work [17] concerning *self-contracted curves*, which is available for some Finsler manifolds.

## APPENDIX A. AN AUXILIARY LEMMA

**Lemma A.1.** *Given  $p \in [1, \infty)$  and any  $\epsilon > 0$ , we have*

$$(1 + \epsilon)a^p + \mathfrak{C}(p, \epsilon)b^p \geq (a + b)^p \quad \text{for all } a, b \geq 0,$$

where

$$\mathfrak{C}(1, \epsilon) = 1, \quad \mathfrak{C}(p, \epsilon) = \frac{1 + \epsilon}{((1 + \epsilon)^{1/(p-1)} - 1)^{p-1}} \quad \text{for } p > 1.$$

In particular,  $\lim_{\epsilon \rightarrow \infty} \mathfrak{C}(p, \epsilon) = 1$  for all  $p \geq 1$ .

*Proof.* We assume  $p > 1$  and  $a, b > 0$  without loss of generality. Moreover, by dividing both sides with  $b^p$ , it suffices to consider the case of  $b = 1$ . Put  $f(a) := (a + 1)^p - (1 + \epsilon)a^p$  for  $a > 0$ . Since

$$f'(a) = p(a + 1)^{p-1} - p(1 + \epsilon)a^{p-1}$$

attains 0 only at  $\bar{a} = \{(1 + \epsilon)^{1/(p-1)} - 1\}^{-1}$ , we have

$$f(a) \leq f(\bar{a}) = \frac{1 + \epsilon}{((1 + \epsilon)^{1/(p-1)} - 1)^{p-1}}$$

for all  $a > 0$ . This completes the proof.  $\square$

## APPENDIX B. COMPLEMENTARY RESULTS FOR GENERALIZED FUNK SPACES

Let  $(\mathbb{B}, d)$  be a generalized Funk space as in Subsection 3.6.

**Proposition B.1.** *For any  $x \in \mathbb{B}$ , we have the following.*

- (i)  $F(x, \cdot)$  is convex on  $T_x\mathbb{B} \cong \mathcal{H}$ , and  $F(x, v) \leq d(x, x+v)$  for any  $v \in \mathcal{H}$  with  $x+v \in \mathbb{B}$ .
- (ii)  $F^*(x, \cdot)$  is convex on  $T_x^*\mathbb{B} \cong \mathcal{H}^*$  and weakly\*-lower semicontinuous, i.e., if  $\zeta_i \xrightarrow{*} \zeta$  (weakly\* convergent), then we have  $F^*(x, \zeta) \leq \liminf_{i \rightarrow \infty} F^*(x, \zeta_i)$ .
- (iii) The set  $\mathfrak{J}_p(x, v)$  is at most a singleton for every  $v \in T_x\mathbb{B} \cong \mathcal{H}$ .

*Proof.* (i) This is reduced to the finite-dimensional case. The inequality  $F(x, v) \leq d(x, x+v)$  can be seen from  $d(x, x+\varepsilon v) \leq d(x+tv, x+(t+\varepsilon)v)$  for  $0 \leq t < t+\varepsilon \leq 1$  (by the interpretation of  $d$  as in [40, (1.1)]).

(ii) This is a direct consequence of the definition of  $F^*$ .

(iii) Assume  $v \neq \mathbf{0}$  without loss of generality. Given any  $\zeta_1, \zeta_2 \in \mathfrak{J}_p(x, v)$ , set  $\bar{v} := v/F(x, v)$  and  $\bar{\zeta}_i := \zeta_i/F^*(x, \zeta_i)$ ,  $i = 1, 2$ . It suffices to show  $\bar{\zeta}_1 = \bar{\zeta}_2$ . Since  $f := F(x, \cdot)$  is differentiable at  $\bar{v}$ , for any  $\varepsilon > 0$ , there exists  $\delta > 0$  such that  $f(\bar{v}+w) + f(\bar{v}-w) \leq 2f(\bar{v}) + \varepsilon\|w\| = 2 + \varepsilon\|w\|$  for any  $w \in \mathcal{H}$  with  $\|w\| < \delta$  (see, e.g., [8, Lemma 8.3]). Combining this with  $F^*(x, \bar{\zeta}_i) = \langle \bar{\zeta}_i, \bar{v} \rangle = 1$ , we find

$$\langle \bar{\zeta}_1, w \rangle - \langle \bar{\zeta}_2, w \rangle = \langle \bar{\zeta}_1, \bar{v}+w \rangle + \langle \bar{\zeta}_2, \bar{v}-w \rangle - \langle \bar{\zeta}_1, \bar{v} \rangle - \langle \bar{\zeta}_2, \bar{v} \rangle \leq f(\bar{v}+w) + f(\bar{v}-w) - 2 \leq \varepsilon\|w\|$$

for any  $w$  with  $\|w\| < \delta$ . This implies  $\bar{\zeta}_1 = \bar{\zeta}_2$ .  $\square$

**Proposition B.2.** *Let  $\phi : \mathbb{B} \rightarrow (-\infty, \infty]$  be a proper lower semicontinuous function.*

- (i) We have  $|\partial\phi|(x) \leq F^*(-\partial^\circ\phi(x))$  for all  $x \in \mathbb{B}$ . In particular,  $x \mapsto F^*(-\partial^\circ\phi(x))$  is a weak upper gradient for  $-\phi$ .
- (ii) If  $\phi$  is convex, then  $|\partial\phi|$  is a strong upper gradient and we have

$$|\partial\phi|(x) = \mathfrak{l}_\phi(x) = F^*(-\partial^\circ\phi(x)) \quad \text{for all } x \in \mathbb{B}. \quad (\text{B.1})$$

- (iii) If  $\phi$  is convex, then the graph of  $\partial\phi$  in  $\mathbb{B} \times \mathcal{H}^*$  is strongly-weakly\* closed. Moreover, we have

$$\zeta_i \in \partial\phi(x_i), \quad x_i \rightarrow x, \quad \zeta_i \xrightarrow{*} \zeta \implies \zeta \in \partial\phi(x), \quad \phi(x_i) \rightarrow \phi(x). \quad (\text{B.2})$$

*Proof.* (i) Assume  $x \in \mathfrak{D}(\partial\phi)$  and  $|\partial\phi|(x) > 0$  without loss of generality. For any  $\zeta \in \partial\phi(x)$ , we have

$$|\partial\phi|(x) = \limsup_{v \rightarrow \mathbf{0}} \frac{\phi(x) - \phi(x+v)}{d(x, x+v)} \leq \limsup_{v \rightarrow \mathbf{0}} \frac{\langle -\zeta, v \rangle}{F(x, v)} \leq F^*(x, -\zeta).$$

This shows the former assertion, and the latter one is a consequence of Theorem 2.21(i).

(ii) It follows from Theorem 2.21(ii) and Theorem 4.8 that  $|\partial\phi| = \mathfrak{l}_\phi$  is a strong upper gradient for  $-\phi$ . In view of (i), it remains to show the existence of  $\zeta \in \partial\phi(x)$  with  $F^*(x, -\zeta) \leq \mathfrak{l}_\phi(x)$  for  $x \in \mathfrak{D}(\phi)$  satisfying  $\mathfrak{l}_\phi(x) < \infty$ . Observe from Theorem 4.8 and Proposition B.1(i) that

$$-\mathfrak{l}_\phi(x)F(x, v) \leq \phi(x+v) - \phi(x) \quad \text{for all } v \in \mathcal{H} \text{ with } x+v \in \mathbb{B},$$

i.e., the convex set  $\{(v, r) \in \mathcal{H} \times \mathbb{R} \mid r \geq \phi(x+v) - \phi(x), x+v \in \mathbb{B}\}$  is disjoint from the open convex set  $\{(v, r) \in \mathcal{H} \times \mathbb{R} \mid r < -\mathfrak{l}_\phi(x)F(x, v), x+v \in \mathbb{B}\}$ . Therefore, we can apply a geometric version of the Hahn–Banach theorem to obtain  $\zeta \in \mathcal{H}^*$  and  $\alpha \in \mathbb{R}$  such that

$$-\mathfrak{l}_\phi(x)F(x, v) \leq \langle \zeta, v \rangle + \alpha \leq \phi(x+v) - \phi(x) \quad \text{for all } v \in \mathcal{H} \text{ with } x+v \in \mathbb{B}.$$

Taking  $v = \mathbf{0}$  implies  $\alpha = 0$ . Thus, the first inequality shows  $F^*(x, -\zeta) \leq \mathfrak{l}_\phi(x)$  while the second one means  $\zeta \in \partial\phi(x)$ . This completes the proof.

- (iii) Owing to Proposition 3.35(ii), this claim can be proved in the same way as [1, Proposition 1.4.4].  $\square$

## APPENDIX C. COMPLEMENTARY RESULTS FOR THE WASSERSTEIN SPACE

Let  $(M, F, \mathfrak{m})$  be a compact Finsler manifold with a smooth positive measure as in Subsection 4.4.

*Proof of Proposition 4.21.* (i) and (ii) readily follow from the properties of the Legendre transformation.

(iii) Let  $g$  be an arbitrary Riemannian metric on  $M$ , which is bi-Lipschitz equivalent to  $F$  by the compactness of  $M$ . Then  $L^2(\mu; TM)$  can be regarded as a Hilbert space equipped with the inner product

$$G(\mathbf{v}, \mathbf{w}) := \int_M g(\mathbf{v}, \mathbf{w}) \, d\mu.$$

Hence, for any bounded linear functional  $\mathcal{Q}$  on  $L^2(\mu; TM)$ , the Riesz representation theorem yields unique  $\mathbf{v} \in L^2(\mu; TM)$  such that  $\mathcal{Q}(\mathbf{w}) = G(\mathbf{v}, \mathbf{w})$ . We conclude the proof by setting  $\zeta := g(\mathbf{v}, \cdot)$ .

(iv) Let  $g$  and  $G$  be as above and  $T_\mu^g \mathcal{P}(M)$  be the tangent space of  $\mathcal{P}(M)$  at  $\mu$  induced from  $g$ . Note that  $L^2(\mu; T^*M)$ ,  $L^2(\mu; TM)$  and  $\mathbf{Ker}(\mathbf{div})(\mu)$  are independent of the choice of a metric, and  $T_\mu^g \mathcal{P}(M)$  is identified with  $T_\mu^* \mathcal{P}(M)$  via the Legendre transformation of  $g$ . Then the claim follows from  $L^2(\mu; TM) = T_\mu^g \mathcal{P}(M) \oplus \mathbf{Ker}(\mathbf{div})(\mu)$  in [7, Lemma 2.4], where  $\oplus$  is the orthogonal direct sum with respect to  $G$ .

(v) Owing to (iv), we have  $0 = \langle \mathfrak{L}_\mu(\mathbf{v}), \mathbf{v} - \mathbf{w} \rangle_\mu = \|\mathbf{v}\|_\mu^2 - \langle \mathfrak{L}_\mu(\mathbf{v}), \mathbf{w} \rangle_\mu$ , and hence

$$\|\mathbf{v}\|_\mu^2 = \langle \mathfrak{L}_\mu(\mathbf{v}), \mathbf{w} \rangle_\mu \leq \|\mathbf{v}\|_\mu \|\mathbf{w}\|_\mu.$$

This yields  $\|\mathbf{v}\|_\mu \leq \|\mathbf{w}\|_\mu$ , and similarly  $\|\mathbf{v}\|_\mu \geq \|\mathbf{w}\|_\mu$  holds. Therefore we find  $\|\mathbf{v}\|_\mu^2 = \|\mathbf{w}\|_\mu^2 = \langle \mathfrak{L}_\mu(\mathbf{v}), \mathbf{w} \rangle_\mu$ , and then  $\mathbf{v} = \mathbf{w}$  follows from (ii) above.  $\square$

*Proof of Proposition 4.32.* Note that the equivalence between (I) and (II) can be reduced to a Riemannian metric  $g$  on  $M$ . Hence, it follows from [7, Proposition 4.3] that (I) and (II) are equivalent and  $d\rho = \rho\zeta$  for some  $\zeta \in T_\mu^* \mathcal{P}(M)$ .

Next we prove that  $\zeta$  is a subdifferential at  $\mu \in \mathcal{D}(H_m)$ . Given  $\nu \in \mathcal{P}(M)$ , let  $\Psi_\mu^\nu$  be the unique optimal transport vector field and set  $\mu_t := \exp(t\Psi_\mu^\nu)_\# \mu$ . On the one hand, by the same argument as in [32, Proposition 7.7], we have

$$\lim_{t \rightarrow 0^+} \frac{H_m(\mu_t) - H_m(\mu)}{t} = \int_M \langle d\rho, \Psi_\mu^\nu \rangle \, d\mathbf{m} = \int_M \langle \zeta, \Psi_\mu^\nu \rangle \, d\mu \leq \|\zeta\|_\mu^* \|\Psi_\mu^\nu\|_\mu < \infty.$$

On the other hand, Proposition 4.31(iii) furnishes

$$\frac{H_m(\mu_t) - H_m(\mu)}{t} \leq H_m(\nu) - H_m(\mu) - \frac{K}{2}(1-t)d_W^2(\mu, \nu), \quad t \in (0, 1].$$

Letting  $t \rightarrow 0^+$ , we find

$$H_m(\nu) - H_m(\mu) \geq \langle \zeta, \Psi_\mu^\nu \rangle_\mu + \frac{K}{2}d_W^2(\mu, \nu).$$

Then it follows from Proposition 4.25 that  $\zeta$  is a strong subdifferential and  $|\partial H_m|(\mu) \leq \|-\zeta\|_\mu^* < \infty$ .

To see the reverse inequality, we consider an arbitrary smooth vector field  $\mathbf{w}$  and a map  $T_t : M \rightarrow M$  such that, for every  $x \in M$ ,  $t \mapsto T_t(x)$  is the geodesic with  $\frac{d}{dt}|_{t=0}[T_t(x)] = \mathbf{w}(x)$ . Thus, by the same discussion as in [7, Lemma 4.2], we find that  $\mu_t := (T_t)_\# \mu$  satisfies

$$\lim_{t \rightarrow 0} \frac{H_m(\mu_t) - H_m(\mu)}{t} = - \int_M \rho \operatorname{div}_m(\mathbf{w}) \, d\mathbf{m}, \quad (\text{C.1})$$

where  $\operatorname{div}_m$  denotes the divergence with respect to  $\mathbf{m}$  (see also the calculation after the proof). This implies that

$$\mathcal{Q}(\mathbf{w}) := \int_M \rho \operatorname{div}_m(\mathbf{w}) \, d\mathbf{m} \leq |\partial H_m|(\mu) \|\mathbf{w}\|_\mu$$

is extended to a linear bounded functional on  $L^2(\mu; TM)$ . Then Proposition 4.21(iii) provides a unique 1-form  $\bar{\zeta} \in L^2(\mu; T^*M)$  with  $\mathcal{Q}(\mathbf{w}) = \langle -\bar{\zeta}, \mathbf{w} \rangle_\mu$  and  $\|-\bar{\zeta}\|_\mu^* \leq |\partial H_m|(\mu)$ . Since

$$\int_M \rho \operatorname{div}_m(\mathbf{w}) \, d\mathbf{m} = - \int_M \rho \langle \bar{\zeta}, \mathbf{w} \rangle \, d\mathbf{m}$$

for all smooth vector fields  $\mathbf{w}$ , we obtain  $\rho\bar{\zeta} = d\rho = \rho\zeta$ . Therefore,  $\|-\zeta\|_\mu^* \leq |\partial H_m|(\mu)$  and hence  $|\partial H_m|(\mu) = \|-\zeta\|_\mu^*$  holds as desired.

Finally, owing to (C.1), the uniqueness of  $\zeta$  can be shown in a similar way to [7, Proposition 4.3].  $\square$

For completeness, we give a (standard) calculation of  $\operatorname{div}_{\mathbf{m}}(\mathbf{w})$  needed in the proof of (C.1):

$$\begin{aligned} \int_M \rho \operatorname{div}_{\mathbf{m}}(\mathbf{w}) \, \mathrm{d}\mathbf{m} &= - \int_M \mathbf{w}(\rho) \, \mathrm{d}\mathbf{m} = \lim_{t \rightarrow 0} \int_M \frac{\rho - \rho(T_t)}{t} \, \mathrm{d}\mathbf{m} = \lim_{t \rightarrow 0} \frac{1}{t} \int_M \rho(1 - \det[\mathrm{d}(T_t^{-1})]) \, \mathrm{d}\mathbf{m} \\ &= \lim_{t \rightarrow 0} \frac{1}{t} \int_M \rho \left( 1 - \frac{1}{\det[\mathrm{d}T_t] \circ T_t^{-1}} \right) \, \mathrm{d}\mathbf{m} = \int_M \rho \cdot \frac{\mathrm{d}}{\mathrm{d}t} \Big|_{t=0} \det[\mathrm{d}T_t] \, \mathrm{d}\mathbf{m}. \end{aligned}$$

## REFERENCES

- [1] L. Ambrosio, N. Gigli and G. Savaré, *Gradient flows in metric spaces and in the spaces of probability measures* (second edition), Lectures in Mathematics, Birkhäuser Verlag, Basel, 2008.
- [2] M. Bačák, “Convex analysis and optimization in Hadamard spaces”, Walter de Gruyter & Co., Berlin, 2014.
- [3] D. Bao, S.-S. Chern and Z. Shen, “An introduction to Riemann–Finsler geometry”, GTM **200**, Springer, New York, 2000.
- [4] D. Bao, C. Robles and Z. Shen, *Zermelo navigation on Riemannian manifolds*, J. Differential Geom. **66** (2004), 377–435.
- [5] A.-Y. Burthsccher, *Length structures on manifolds with continuous Riemannian metrics*, New York J. Math. **21** (2015), 273–296.
- [6] I. V. Chenchiah, M. O. Rieger and J. Zimmer, *Gradient flows in asymmetric metric spaces*, Nonlinear Anal. **71** (2009), 5820–5834.
- [7] M. Erbar, *The heat equation on manifolds as a gradient flow in the Wasserstein space*, Ann. Inst. Henri Poincaré Probab. Statist. **46** (2010), 1–23.
- [8] M. Fabian, P. Habala, P. Hájek, V. Montesinos Santalucía, J. Pelant and V. Zizler, “Functional analysis and infinite-dimensional geometry”, CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC, **8**. Springer-Verlag, New York, 2001.
- [9] Y. Ge and Z. Shen, *Eigenvalues and eigenfunctions of metric measure manifolds*, Proc. Lond. Math. Soc. (3) **82** (2001), 725–746.
- [10] N. Gigli and F. Nobili, *A differential perspective on gradient flows on CAT(k)-spaces and applications*, J. Geom. Anal. **31** (2021), 11780–11818.
- [11] J. Jost, *Convex functionals and generalized harmonic maps into spaces of nonpositive curvature*, Comment. Math. Helv. **70** (1995), 659–673.
- [12] J. Jost, “Nonpositive curvature: Geometric and analytic aspects”, Birkhäuser Verlag, Basel, 1997.
- [13] M. Kell, *q-Heat flow and the gradient flow of the Renyi entropy in the p-Wasserstein space*, J. Funct. Anal. (8) **271** (2016), 2045–2089.
- [14] M. Kell, *On interpolation and curvature via Wasserstein geodesics*, Adv. Calc. Var. **10** (2017), 125–167.
- [15] A. Kristály and W. Zhao, *On the geometry of irreversible metric-measure spaces: Convergence, stability and analytic aspects*, J. Math. Pures Appl. (9) **158** (2022), 216–292.
- [16] K. Kuwada, *Duality on gradient estimates and Wasserstein controls*, J. Funct. Anal. **258** (2010), 3758–3774.
- [17] 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.
- [18] J. Lott and C. Villani, *Optimal transport for metric-measure spaces via Ricci curvature*, Ann. of Math. **169** (2009), 903–991.
- [19] M. Matsumoto, *A slope of a mountain is a Finsler surface with respect to a time measure*, J. Math. Kyoto Univ. **29** (1989), 17–25.
- [20] U. F. Mayer, *Gradient flows on nonpositively curved metric spaces and harmonic maps*, Comm. Anal. Geom. **6** (1998), 199–253.
- [21] A. C. G. Menzucci, *On asymmetric distances*, Anal. Geom. Metr. Spaces **1** (2013), 200–231.
- [22] A. C. G. Menzucci, *Geodesics in asymmetric metric spaces*, Anal. Geom. Metr. Spaces **2** (2014), 115–153.
- [23] M. Muratori and G. Savaré, *Gradient flows and evolution variational inequalities in metric spaces. I: Structural properties*, J. Funct. Anal. **278** (2020), 108347, 67 pp.
- [24] S. Ohta, *Gradient flows on Wasserstein spaces over compact Alexandrov spaces*, Amer. J. Math. **131** (2009), 475–516.
- [25] S. Ohta, *Finsler interpolation inequalities*, Calc. Var. Partial Differential Equations **36** (2009), 211–249.
- [26] S. Ohta, *Uniform convexity and smoothness, and their applications in Finsler geometry*, Math. Ann. **343** (2009), 669–699.
- [27] S. Ohta, “Comparison Finsler geometry”, Springer Monographs in Mathematics. Springer, Cham, 2021.
- [28] S. Ohta, *A semigroup approach to Finsler geometry: Bakry–Ledoux’s isoperimetric inequality*, Comm. Anal. Geom. (to appear). Available at [arXiv:1602.00390](https://arxiv.org/abs/1602.00390)
- [29] S. Ohta, *Discrete-time gradient flows in Gromov hyperbolic spaces*, Israel J. Math. (to appear). Preprint (2023). Available at [arXiv:2205.03156](https://arxiv.org/abs/2205.03156)
- [30] S. Ohta and M. Pálfi, *Discrete-time gradient flows and law of large numbers in Alexandrov spaces*, Calc. Var. Partial Differential Equations **54** (2015), 1591–1610.

- [31] 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.
- [32] S. Ohta and K.-T. Sturm, *Heat flow on Finsler manifolds*, Comm. Pure Appl. Math. **62** (2009), 1386–1433.
- [33] S. Ohta and K.-T. Sturm, *Non-contraction of heat flow on Minkowski spaces*, Arch. Ration. Mech. Anal. **204** (2012), 917–944.
- [34] S. Ohta and K.-T. Sturm, *Bochner–Weitzenböck formula and Li–Yau estimates on Finsler manifolds*, Adv. Math. **252** (2014), 429–448.
- [35] M.-K. v. Renesse and K.-T. Sturm, *Transport inequalities, gradient estimates, entropy and Ricci curvature*, Commun. Pure Appl. Math. **58** (2005), 923–940.
- [36] R. Rossi, A. Mielke and G. Savaré, *A metric approach to a class of doubly nonlinear evolution equations and applications*, Ann. Sc. Norm. Super. Pisa Cl. Sci. **7** (2008), 97–169.
- [37] R. Rossi, A. Segatti and U. Stefanelli, *Global attractors for gradient flows in metric spaces*, J. Math. Pures Appl. (9) **95** (2011), 205–244.
- [38] G. Savaré, *Gradient flows and diffusion semigroups in metric spaces under lower curvature bounds*, C. R. Math. Acad. Sci. Paris **345** (2007), 151–154.
- [39] Y. Shen and W. Zhao, *Gromov pre-compactness theorems for nonreversible Finsler manifolds*, Differ. Geom. Appl. **28** (2010), 565–581.
- [40] Z. Shen, “Lectures on Finsler geometry”, World Sci., Singapore, 2001.
- [41] K.-T. Sturm, *On the geometry of metric measure spaces. I*, Acta Math. **196** (2006), 65–131.
- [42] K.-T. Sturm, *On the geometry of metric measure spaces. II*, Acta Math. **196** (2006), 133–177.
- [43] 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.
- [44] C. Villani, “Optimal transportation, old and new”, Grundlehren der mathematischen Wissenschaften **338**, Springer, 2009.

DEPARTMENT OF MATHEMATICS, OSAKA UNIVERSITY, 560-0043 OSAKA, JAPAN, AND RIKEN CENTER FOR ADVANCED INTELLIGENCE PROJECT (AIP), 1-4-1 NIHONBASHI, TOKYO 103-0027, JAPAN

*Email address:* s.ohta@math.sci.osaka-u.ac.jp

SCHOOL OF MATHEMATICS, EAST CHINA UNIVERSITY OF SCIENCE AND TECHNOLOGY, 200237 SHANGHAI, CHINA

*Email address:* szhao\_wei@yahoo.com