Kyoto University

Gradient flows and a Trotter Kato formula of semi-convex functions on CAT(1)-spaces

by

Shin-ichi Ohta, Miklós Pálfia

November 2014

京都大学理学部数学教室
Department of Mathematics
Faculty of Science
Kyoto University
Kyoto 606-8502, JAPAN

Gradient flows and a Trotter–Kato formula of semi-convex functions on CAT(1)-spaces

Shin-ichi Ohta* & Miklós Pálfia†

Abstract

We generalize the theory of gradient flows of semi-convex functions on CAT(0)-spaces, developed by Mayer and Ambrosio–Gigli–Savaré, to CAT(1)-spaces. The key tool is the so-called "commutativity" representing a Riemannian nature of the space. Our approach combining the semi-convexity of the squared distance function with a Riemannian property of the space seems to be of independent interest, and can be compared with Savaré's work on the local angle condition under lower curvature bounds. Applications include the convergence of the discrete variational scheme to a unique gradient curve, the contraction property and the evolution variational inequality of the gradient flow, and a Trotter–Kato product formula for pairs of semi-convex functions.

1 Introduction

The theory of gradient flows in singular spaces is a field of active research having applications in various fields. For instance, regarding heat flow as gradient flow of the relative entropy in the (L^2-)Wasserstein space, initiated by Jordan, Kinderlehrer and Otto [JKO], is known as a useful technique in partial differential equations (see [Ot], [Vi1], [AGS1], [ASZ] among many others), and has played a crucial role in the recent remarkable development of geometric analysis on metric measure spaces satisfying the (*Riemannian*) *curvature-dimension condition* (see [Vi2], [Gi1], [GKO], [AGS2], [AGS3], [EKS]). One of the recent striking achievements is Gigli's splitting theorem [Gi2] (see also a survey [Gi3]), in which the gradient flow of the Busemann function is used in an impressive way.

We follow the strategy of constructing a gradient flow of a lower semi-continuous, semi-convex function ϕ on a metric space (X, d) via the discrete variational scheme employing the *Moreau–Yosida approximation*:

$$\phi_\tau(x) := \inf_{z \in X} \left\{ \phi(z) + \frac{d^2(x, z)}{2\tau} \right\}, \quad x \in X, \ \tau > 0. \tag{1.1}$$

A point x_τ attaining the above infimum is considered as an approximation of the point $\xi(\tau)$ on the gradient curve ξ of ϕ with $\xi(0) = x$. Here ϕ is said to be *semi-convex* if it is *λ-convex* for some $\lambda \in \mathbb{R}$ meaning that

$$\phi\big(\gamma(t)\big) \leq (1-t)\phi\big(\gamma(0)\big) + t\phi\big(\gamma(1)\big) - \frac{\lambda}{2}(1-t)td^2\big(\gamma(0), \gamma(1)\big)$$

along geodesics $\gamma : [0,1] \longrightarrow X$. Since the approximation scheme is based on the distance function, finer properties of the distance function provide finer analysis of gradient flows. In [Jo], [Ma] (with applications to harmonic maps) and [AGS1] (with applications to the Wasserstein spaces), gradient flows in CAT(0)-*spaces* (non-positively curved metric spaces) are well investigated; see also Bačák's recent book [Ba2]. There the 2-convexity of the squared distance function, that is indeed the definition of CAT(0)-spaces, played essential roles. We shall generalize their theory to CAT(1)-*spaces* (metric spaces of sectional curvature ≤ 1), where the distance functions are only semi-convex. CAT(1)-spaces can have a more complicated global structure than CAT(0)-spaces. For instance, all CAT(0)-spaces are contractible while CAT(1)-spaces may not.

It is known that the direct application of the techniques of CAT(0)-spaces to CAT(1)-spaces does not work. The point is that the K-convexity of the squared distance function with $K < 2$ holds true on some non-Hilbert Banach spaces, on those the behavior of gradient flows is much less understood. We overcome this difficulty by introducing the notion of "commutativity" representing a "Riemannian nature" of the space. Precisely, the key ingredients of our analysis are the following:

(A) The *commutativity*:

$$\lim_{s\downarrow 0} \frac{d^2(\gamma(s), z) - d^2(x, z)}{s} = \lim_{t\downarrow 0} \frac{d^2(\eta(t), y) - d^2(x, y)}{t} \tag{1.2}$$

for geodesics γ and η with $x = \gamma(0) = \eta(0)$, $\gamma(1) = y$ and $\eta(1) = z$ (see (3.1) in the proof of Lemma 3.1);

(B) The semi-convexity of the squared distance function (see Lemma 2.8).

(One can more generally consider some other family of curves along those (A) and (B) hold, as was essentially used to study the Wasserstein spaces in [AGS1]). This approach, reinforcing the semi-convexity with the Riemannian nature of the space, seems to be of independent interest and is in a similar spirit to Savaré's work [Sa] based on the semi-concavity of distance functions and the *local angle condition*, those properties fit the study of spaces with lower curvature bounds. The semi-convexity and semi-concavity are usually studied in the context of Banach space theory and Finsler geometry, see [BCL], [Oh1], [Oh3]. In the CAT(0)-setting, the commutativity follows from the 2-convexity of the squared distance function and the role of the commutativity is implicit. The commutativity is not true in non-Riemannian Finsler manifolds (see Remark 3.2(b)). Actually, the lack of the commutativity is the reason why the first author and Sturm introduced the notion of *skew-convexity* in [OS] to study the contraction property of gradient flows in Finsler manifolds. In contrast, on the Wasserstein space over a Riemannian manifold,

we have a sort of Riemannian structure but the convexity of the squared distance function fails. See [Gi2, §B] for a connection between (1.2) and the *infinitesimal Hilbertianity* which is an "almost everywhere" notion of Riemannian nature.

Modifying the calculation in [AGS1] with the help of the commutativity, we arrive at the key estimate for a λ-convex function ϕ and a K-convex distance (see Lemma 3.1):

$$d^2(x_\tau, y) \leq d^2(x, y) - \lambda \tau d^2(x_\tau, y) + 2\tau\{\phi(y) - \phi(x_\tau)\} - \frac{K}{2}d^2(x, x_\tau),$$

where x_τ is a point attaining the infimum in (1.1). Surprisingly, even with $K < 0$ (as well as $\lambda < 0$), this estimate is enough to generalize the argument of [AGS1]. We show the convergence of the discrete variational scheme to a unique gradient curve (Theorem 4.4), the contraction property (Theorem 4.7) and the evolution variational inequality (Theorem 4.8) of the gradient flow. Moreover, along the lines of [Ma] and [CM], we study the large time behavior of the flow (§4.5) and prove a Trotter–Kato product formula for pairs of semi-convex functions (Theorem 5.4). The latter is a two-fold generalization of the existing results in [CM], [Sto] and [Ba1] for convex functions on CAT(0)-spaces (to be precise, an inequality corresponding to our key estimate with $\lambda = 0$ and $K = 2$ is an assumption of [CM]). We stress that we use only the qualitative properties of CAT(1)-spaces instead of the direct curvature condition. Thus our technique also applies to every metric space satisfying the conditions (A) and (B) above, under a mild coercivity assumption on ϕ (see Case II at the beginning of Section 4).

We finally mention some more related works. Gradient flow in metric spaces with lower sectional curvature bounds (*Alexandrov spaces*) is investigated in [PP], [Ly]. This technique was generalized to the Wasserstein spaces over Alexandrov spaces in [Oh2] and [Sa]. Sturm [Stu] recently studied gradient flows in metric measure spaces satisfying the Riemannian curvature-dimension condition. Discrete-time gradient flow is also an important subject related to optimization theory, for that we refer to [OP] and the references therein.

The organization of the article is as follows. In Section 2, we recall preliminary results on gradient flows in metric spaces from [AGS1], followed by the necessary facts of CAT(1)-spaces. Section 3 is devoted to our key estimate. We apply the key estimate to the study of gradient flows in Section 4, and prove a Trotter–Kato product formula in Section 5.

2   Preliminaries

Let (X, d) be a complete metric space. A curve $\gamma : [0, 1] \longrightarrow X$ is called a *geodesic* if it is locally minimizing and of constant speed. We call γ a *minimal* geodesic if it is globally minimizing, namely $d(\gamma(s), \gamma(t)) = |s - t|d(\gamma(0), \gamma(1))$ for all $s, t \in [0, 1]$. We say that (X, d) is *geodesic* if any two points $x, y \in X$ admit a minimal geodesic between them.

2.1   Gradient flows in metric spaces

We recall basic facts on the construction of gradient curves in metric spaces. We follow the technique called the "minimizing movements" going back to (at least) De Giorgi [DG],

see [AGS1] for more on this theory. We also refer to [Br], [CL] for classical theories on linear spaces.

2.1.1 Discrete solutions

As our potential function, we always consider a lower semi-continuous function $\phi : X \longrightarrow (-\infty, \infty]$ such that

$$D(\phi) := X \setminus \phi^{-1}(\infty) \neq \emptyset.$$

Given $x \in X$ and $\tau > 0$, we define the *Moreau–Yosida approximation*:

$$\phi_\tau(x) := \inf_{z \in X} \left\{ \phi(z) + \frac{d^2(x, z)}{2\tau} \right\}$$

and set

$$J_\tau^\phi(x) := \left\{ z \in X \,\middle|\, \phi(z) + \frac{d^2(x, z)}{2\tau} = \phi_\tau(x) \right\}.$$

For $x \in D(\phi)$ and $z \in J_\tau^\phi(x)$ (if $J_\tau^\phi(x) \neq \emptyset$), it is straightforward from

$$\phi(z) + \frac{d^2(x, z)}{2\tau} \leq \phi(x)$$

that $\phi(z) \leq \phi(x)$ and $d^2(x, z) \leq 2\tau\{\phi(x) - \phi(z)\}$. We consider two kinds of conditions on ϕ.

Assumption 2.1 (1) There exists $\tau_*(\phi) \in (0, \infty]$ such that $\phi_\tau(x) > -\infty$ and $J_\tau^\phi(x) \neq \emptyset$ for all $x \in X$ and $\tau \in (0, \tau_*(\phi))$ (*coercivity*).

(2) For any $Q \in \mathbb{R}$, bounded subsets of the sub-level set $\{x \in X \mid \phi(x) \leq Q\}$ are relatively compact in X (*compactness*).

We remark that, if $\phi_{\tau_*}(x_*) > -\infty$ for some $x_* \in X$ and $\tau_* > 0$, then $\phi_\tau(x) > -\infty$ for every $x \in X$ and $\tau \in (0, \tau_*)$ (see [AGS1, Lemma 2.2.1]). Then, if the compactness (2) holds, we have $J_\tau^\phi(x) \neq \emptyset$ by the lower semi-continuity of ϕ (see [AGS1, Corollary 2.2.2]).

Remark 2.2 If $\operatorname{diam} X < \infty$ and the compactness (2) holds, then the lower semi-continuity of ϕ implies that every sub-level set $\{x \in X \mid \phi(x) \leq Q\}$ is (empty or) compact. Thus ϕ is bounded below and we can take $\tau_*(\phi) = \infty$.

To construct discrete approximations of gradient curves of ϕ, we consider a *partition* of the interval $[0, \infty)$:

$$\mathscr{P}_{\boldsymbol{\tau}} = \{0 = t_{\boldsymbol{\tau}}^0 < t_{\boldsymbol{\tau}}^1 < \cdots\}, \qquad \lim_{k \to \infty} t_{\boldsymbol{\tau}}^k = \infty,$$

and set

$$\tau_k := t_{\boldsymbol{\tau}}^k - t_{\boldsymbol{\tau}}^{k-1} \quad \text{for } k \in \mathbb{N}, \qquad |\boldsymbol{\tau}| := \sup_{k \in \mathbb{N}} \tau_k.$$

We will always assume $|\boldsymbol{\tau}| < \tau_*(\phi)$. Given an initial point $x_0 \in D(\phi)$,

$$x_{\boldsymbol{\tau}}^0 := x_0 \text{ and recursively choose arbitrary } x_{\boldsymbol{\tau}}^k \in J_{\tau_k}^\phi(x_{\boldsymbol{\tau}}^{k-1}) \text{ for each } k \in \mathbb{N}. \tag{2.1}$$

We call $\{x_{\boldsymbol{\tau}}^k\}_{k\in\mathbb{N}}$ a *discrete solution* of the variational scheme (2.1) associated with the partition $\mathscr{P}_{\boldsymbol{\tau}}$, which is thought of as a *discrete-time gradient curve* for the potential function ϕ. The following a priori estimates (see [AGS1, Lemma 3.2.2]) will be useful in the sequel. We remark that these estimates are easily obtained if ϕ is bounded below.

Lemma 2.3 (A priori estimates) *Let $\phi : X \longrightarrow (-\infty, \infty]$ satisfy Assumption 2.1(1). Then, for any $x_* \in X$ and $Q, T > 0$, there exists a constant $C = C(x_*, \tau_*(\phi), Q, T) > 0$ such that, if a partition $\mathscr{P}_{\boldsymbol{\tau}}$ and an associated discrete solution $\{x_{\boldsymbol{\tau}}^k\}_{k\in\mathbb{N}}$ of (2.1) satisfy*

$$\phi(x_0) \leq Q, \quad d^2(x_0, x_*) \leq Q, \quad t_{\boldsymbol{\tau}}^N \leq T, \quad |\boldsymbol{\tau}| \leq \frac{\tau_*(\phi)}{8},$$

then we have for any $1 \leq k \leq N$

$$d^2(x_{\boldsymbol{\tau}}^k, x_*) \leq C, \qquad \sum_{l=1}^{k} \frac{d^2(x_{\boldsymbol{\tau}}^{l-1}, x_{\boldsymbol{\tau}}^l)}{2\tau_l} \leq \phi(x_0) - \phi(x_{\boldsymbol{\tau}}^k) \leq C.$$

In particular, for all $1 \leq k \leq N$, we have $d^2(x_{\boldsymbol{\tau}}^{k-1}, x_{\boldsymbol{\tau}}^k) \leq 2C\tau_k$ and

$$d^2(x_0, x_{\boldsymbol{\tau}}^k) \leq \left(\sum_{l=1}^{k} d(x_{\boldsymbol{\tau}}^{l-1}, x_{\boldsymbol{\tau}}^l)\right)^2 \leq \sum_{l=1}^{k} \frac{d^2(x_{\boldsymbol{\tau}}^{l-1}, x_{\boldsymbol{\tau}}^l)}{\tau_l} \cdot \sum_{l=1}^{k} \tau_l \leq 2Ct_{\boldsymbol{\tau}}^k. \qquad (2.2)$$

2.1.2 Convergence of discrete solutions

From here on, let $\phi : (-\infty, \infty] \longrightarrow X$ be *λ-convex* (also called *λ-geodesically convex*) for some $\lambda \in \mathbb{R}$ in the sense that

$$\phi\big(\gamma(t)\big) \leq (1-t)\phi(x) + t\phi(y) - \frac{\lambda}{2}(1-t)td^2(x, y) \qquad (2.3)$$

for any $x, y \in D(\phi)$ and some minimal geodesic $\gamma : [0, 1] \longrightarrow X$ from x to y. (The existence of a minimal geodesic between two points in $D(\phi)$ is included in the definition, thus in particular $(D(\phi), d)$ is geodesic.) We remark that the compactness (2) in Assumption 2.1 implies the coercivity (1) in this case (see [AGS1, Lemma 2.4.8]; we even have $\tau_*(\phi) = \infty$ if $\lambda \geq 0$).

Fix an initial point $x_0 \in D(\phi)$. Take a sequence of partitions $\{\mathscr{P}_{\boldsymbol{\tau}_i}\}_{i\in\mathbb{N}}$ such that $\lim_{i\to\infty} |\boldsymbol{\tau}_i| = 0$ and associated discrete solutions $\{x_{\boldsymbol{\tau}_i}^k\}_{k\in\mathbb{N}}$ with $x_{\boldsymbol{\tau}_i}^0 = x_0$. Under Assumption 2.1(2), by the compactness argument ([AGS1, Proposition 2.2.3]), a subsequence of the interpolated curves

$$\bar{\boldsymbol{x}}_{\boldsymbol{\tau}_i}(0) := x_0, \qquad \bar{\boldsymbol{x}}_{\boldsymbol{\tau}_i}(t) := x_{\boldsymbol{\tau}_i}^k \quad \text{for } t \in (t_{\boldsymbol{\tau}_i}^{k-1}, t_{\boldsymbol{\tau}_i}^k] \qquad (2.4)$$

converges to a curve $\xi : [0, \infty) \longrightarrow D(\phi)$ point-wise in $t \in [0, \infty)$. In general, under the coercivity and λ-convexity of ϕ (but without the compactness), if a curve ξ is obtained as above (called a *generalized minimizing movement*; see [AGS1, Definition 2.0.6]), then it is locally Lipschitz on $(0, \infty)$ and satisfies $\lim_{t\downarrow 0} \xi(t) = x_0$ as well as the *energy dissipation identity*:

$$\phi\big(\xi(T)\big) = \phi\big(\xi(S)\big) - \frac{1}{2}\int_S^T \{|\dot{\xi}|^2 + |\nabla\phi|^2(\xi)\}\, dt. \qquad (2.5)$$

Here

$$|\dot{\xi}|(t) := \lim_{s \to t} \frac{d(\xi(s), \xi(t))}{|t - s|}$$

is the *metric speed* existing at almost all t, and

$$|\nabla \phi|(x) := \limsup_{y \to x} \frac{\max\{\phi(x) - \phi(y), 0\}}{d(x, y)}$$

is the (descending) *local slope* (see [AGS1, Theorem 2.4.15]). We remark that $|\nabla \phi|$ is lower semi-continuous ([AGS1, Corollary 2.4.10]) and $\lim_{i \to \infty} \phi(\bar{\boldsymbol{x}}_{\boldsymbol{\tau}_i}(t)) = \phi(\xi(t))$ for all $t \geq 0$ ([AGS1, Theorem 2.3.3]). The equation (2.5) can be thought of as a metric formulation of the differential equation $\dot{\xi}(t) = -\nabla \phi(\xi(t))$, thus ξ will be called a *gradient curve* of ϕ starting from x_0. We remark that one does not have uniqueness of gradient curves in this generality (see [AG, Example 4.23] for a simple example in the ℓ_∞^2-space).

Remark 2.4 One can relax the scheme by allowing varying initial points: $x_{\boldsymbol{\tau}_i}^0 \neq x_0$. Then assuming $x_{\boldsymbol{\tau}_i}^0 \to x_0$ and $\phi(x_{\boldsymbol{\tau}_i}^0) \to \phi(x_0)$ yields the same convergence results. In our setting, such a convergence can also follow from the comparison estimate (4.2) (see the proof of Theorem 4.4).

2.2 CAT(1)-spaces

We refer to [BBI] for the basics of CAT(1)-spaces and for more general metric geometry.

Given three points $x, y, z \in X$ with $d(x, y) + d(y, z) + d(z, x) < 2\pi$, we can take corresponding points $\tilde{x}, \tilde{y}, \tilde{z}$ in the 2-dimensional unit sphere $\mathbb{S}^2$ (uniquely up to rigid motions) such that

$$d_{\mathbb{S}^2}(\tilde{x}, \tilde{y}) = d(x, y), \qquad d_{\mathbb{S}^2}(\tilde{y}, \tilde{z}) = d(y, z), \qquad d_{\mathbb{S}^2}(\tilde{z}, \tilde{x}) = d(z, x).$$

We call $\triangle \tilde{x}\tilde{y}\tilde{z}$ a *comparison triangle* of $\triangle xyz$ in $\mathbb{S}^2$.

Definition 2.5 (CAT(1)-spaces) A geodesic metric space (X, d) is called a CAT(1)-*space* if, for any $x, y, z \in X$ with $d(x, y) + d(y, z) + d(z, x) < 2\pi$ and any minimal geodesic $\gamma : [0, 1] \longrightarrow X$ from y to z, we have

$$d\big(x, \gamma(t)\big) \leq d_{\mathbb{S}^2}\big(\tilde{x}, \tilde{\gamma}(t)\big)$$

at all $t \in [0, 1]$, where $\triangle \tilde{x}\tilde{y}\tilde{z} \subset \mathbb{S}^2$ is a comparison triangle of $\triangle xyz$ and $\tilde{\gamma} : [0, 1] \longrightarrow \mathbb{S}^2$ is the minimal geodesic from $\tilde{y}$ to $\tilde{z}$.

It is readily observed from the definition that each pair of points $x, y \in X$ with $d(x, y) < \pi$ is joined by a unique minimal geodesic. Fundamental examples of CAT(1)-spaces are the following.

Example 2.6 (1) A complete, simply connected Riemannian manifold endowed with the Riemannian distance is a CAT(1)-space if and only if its sectional curvature is not greater than 1 everywhere.

6

(2) All CAT(0)-*spaces*, which are defined similarly to Definition 2.5 by replacing $\mathbb{S}^2$ with $\mathbb{R}^2$, are CAT(1)-spaces. Important examples of CAT(0)-spaces are Hadamard manifolds, Hilbert spaces, trees and Euclidean buildings.

(3) Further examples of CAT(1)-spaces include orbifolds obtained as quotient spaces of CAT(1)-manifolds, and spherical buildings. See [BBI, §9.1] for more examples.

Remark 2.7 For general $\kappa \in \mathbb{R}$, CAT(κ)-spaces are defined in the same manner by employing comparison triangles in the 2-dimensional space form of constant curvature κ. If (X, d) is a CAT(κ)-space, then it is also CAT(κ') for all $\kappa' > \kappa$ and the scaled metric space (X, cd) with $c > 0$ is CAT($c^{-2}\kappa$). Therefore considering CAT(1)-spaces covers all CAT(κ)-spaces up to scaling.

As was mentioned in the introduction, the properties of CAT(1)-spaces needed in our discussion are only the semi-convexity of the squared distance function and the commutativity (1.2). The latter is a consequence of the first variation formula. Let us review them.

Lemma 2.8 (Semi-convexity of distance functions) *Let (X, d) be a CAT(1)-space and take $R \in (0, \pi)$. Then there exists $K = K(R) \in \mathbb{R}$ such that the squared distance function $d^2(x, \cdot)$ is K-convex on the open R-ball $B(x, R)$ for all $x \in X$.*

Proof. By the definition of CAT(1)-spaces, it is enough to show the claim in $\mathbb{S}^2$. Then the K-convexity is a direct consequence of the smoothness of $d^2_{\mathbb{S}^2}(\tilde{x}, \cdot)$ on $B(\tilde{x}, \pi)$. □

Clearly $K(R) > 0$ if $R < \pi/2$ (see, e.g., [Oh1] for the precise estimate) and $K(R) < 0$ if $R > \pi/2$. We can define the *angle* between two geodesics γ and η emanating from the same point $\gamma(0) = \eta(0) = x$ by

$$\angle_x(\gamma, \eta) := \lim_{s,t\downarrow 0} \angle \widetilde{\gamma(s)} \tilde{x} \widetilde{\eta(t)},$$

where $\angle \widetilde{\gamma(s)} \tilde{x} \widetilde{\eta(t)}$ is the angle at $\tilde{x}$ of a comparison triangle $\triangle \widetilde{\gamma(s)} \tilde{x} \widetilde{\eta(t)}$ in $\mathbb{S}^2$. By the definition of the angle, we obtain the following (see [BBI, Theorem 4.5.6, Remark 4.5.12]).

Theorem 2.9 (First variation formula) *Let $\gamma : [0, 1] \longrightarrow X$ be a geodesic from x to z, and take $y \in X$ with $0 < d(x, y) < \pi$. Then we have*

$$\lim_{s\downarrow 0} \frac{d(\gamma(s), y) - d(x, y)}{s} = -d(x, z) \cos \angle_x(\gamma, \eta),$$

where $\eta : [0, 1] \to X$ is the unique minimal geodesic from x to y.

3   Key lemma

In this section, let (X, d) be a complete CAT(1)-space and $\phi : X \longrightarrow (-\infty, \infty]$ satisfy the λ-convexity for some $\lambda \in \mathbb{R}$ and Assumption 2.1(1). Note that (2.3) holds along every minimal geodesic since minimal geodesics are unique between points of distance $< \pi$. The next lemma, generalizing [AGS1, Theorem 4.1.2(ii)] to the case where both λ and K can be negative, will be a key tool in the following sections.

Lemma 3.1 (Key lemma) *Let $x \in D(\phi)$ and $\tau \in (0, \min\{\pi^2/(2C), \tau_*(\phi)/8\})$ with $C = C(x, \tau_*(\phi), \phi(x), \tau_*(\phi)/8)$ from Lemma 2.3. Take $x_\tau \in J_\tau^\phi(x)$. Then we have, for any $y \in D(\phi) \cap B(x_\tau, R - d(x, x_\tau))$ with $R < \pi$ and for $K = K(R)$ as in Lemma 2.8,*

$$d^2(x_\tau, y) \leq d^2(x, y) - \lambda\tau d^2(x_\tau, y) + 2\tau\{\phi(y) - \phi(x_\tau)\} - \frac{K}{2}d^2(x, x_\tau)$$

$$\leq d^2(x, y) - \lambda\tau d^2(x_\tau, y) + 2\tau\{\phi(y) - \phi(x_\tau)\}$$
$$+ \max\{0, -K\} \cdot \tau\{\phi(x) - \phi(x_\tau)\}.$$

Proof. Observe that $d^2(x, x_\tau) \leq 2C\tau < \pi^2$ by Lemma 2.3 and the choice of τ. Let $\gamma : [0, 1] \longrightarrow X$ be the minimal geodesic from x_τ to y, and $\eta : [0, 1] \longrightarrow X$ from x_τ to x. For any $s \in (0, 1)$, by the definition of $J_\tau^\phi(x)$ and the λ-convexity of ϕ, we have

$$\phi(x_\tau) + \frac{d^2(x, x_\tau)}{2\tau} \leq \phi(\gamma(s)) + \frac{d^2(x, \gamma(s))}{2\tau}$$

$$\leq (1 - s)\phi(x_\tau) + s\phi(y) - \frac{\lambda}{2}(1 - s)sd^2(x_\tau, y) + \frac{d^2(x, \gamma(s))}{2\tau}.$$

Hence

$$\phi(x_\tau) \leq \phi(y) + \frac{1}{2\tau}\frac{d^2(x, \gamma(s)) - d^2(x, x_\tau)}{s} - \frac{\lambda}{2}(1 - s)d^2(x_\tau, y).$$

Applying the first variation formula (Theorem 2.9) twice, we observe the *commutativity*:

$$\lim_{s\downarrow 0}\frac{d^2(x, \gamma(s)) - d^2(x, x_\tau)}{s} = -2d(x_\tau, x)d(x_\tau, y)\cos\angle_{x_\tau}(\gamma, \eta)$$

$$= \lim_{t\downarrow 0}\frac{d^2(\eta(t), y) - d^2(x_\tau, y)}{t}. \tag{3.1}$$

Notice that η is contained in $B(y, R)$ by the choice of y. Thus it follows from the K-convexity of $d^2(\cdot, y)$ in $B(y, R)$ that

$$\lim_{t\downarrow 0}\frac{d^2(\eta(t), y) - d^2(x_\tau, y)}{t} \leq d^2(x, y) - d^2(x_\tau, y) - \frac{K}{2}d^2(x, x_\tau)$$

$$\leq d^2(x, y) - d^2(x_\tau, y) + \max\{0, -K\} \cdot \tau\{\phi(x) - \phi(x_\tau)\}.$$

This completes the proof. $\qquad\square$

Remark 3.2 (a) Used in the proof of [AGS1, Theorem 4.1.2(ii)] is the direct application of the convexity of ϕ and $d^2(x, \cdot)$ along γ, which implies in our setting

$$\frac{K}{2}d^2(x_\tau, y) \leq d^2(x, y) - \lambda\tau d^2(x_\tau, y) + 2\tau\{\phi(y) - \phi(x_\tau)\} - d^2(x, x_\tau).$$

This coincides with our estimate when $K = 2$. The commutativity was used to move the coefficient $K/2$ from $d^2(x_\tau, y)$ to $d^2(x, x_\tau)$, then one can efficiently estimate $d^2(x_\tau, y) - d^2(x, y)$ (see Theorem 4.1).

(b) As we mentioned in the introduction (see the paragraph including (1.2)), the Riemannian nature of the space (i.e., the angle) is essential in the commutativity (3.1). In fact, on a Finsler manifold (M, F), (1.2) (written using only the distance) implies

$$g_v(v, w) = g_w(v, w) \qquad \text{for all } v, w \in T_x M \setminus \{0\}, \ x \in M.$$

These notations and the basics of Finsler geometry can be found in [OS] for instance. Thus we find, for $v \neq \pm w$,

$$
\begin{aligned}
F^2(v + w) + F^2(v - w) &= g_{v+w}(v + w, v + w) + g_{v-w}(v - w, v - w) \\
&= g_{v+w}(v, v + w) + g_{v+w}(w, v + w) + g_{v-w}(v, v - w) - g_{v-w}(w, v - w) \\
&= g_v(v, v + w) + g_w(w, v + w) + g_v(v, v - w) - g_w(w, v - w) \\
&= 2g_v(v, v) + 2g_w(w, w) = 2F^2(v) + 2F^2(w).
\end{aligned}
$$

This is the parallelogram identity on $T_x M$ and hence F is Riemannian.

The commutativity (1.2) is the essential property connecting the (geodesic) convexity of a function and the contraction property of its gradient flow (see Theorem 4.7). On Finsler manifolds, the contraction property is characterized by the *skew-convexity* which is different from the usual convexity along geodesics (see [OS] for details).

4 Applications to gradient flows

The estimate in Lemma 3.1 is worse than the one in [AGS1, Theorem 4.1.2(ii)] because of the generality that K can be less than 2 and even negative. Nonetheless, as we shall see in this section, Lemma 3.1 is enough to generalize the argumentation in Chapter 4 of [AGS1]. We will give at least sketches of the proofs for completeness.

Our argument covers two cases. In both cases, (X, d) is complete, $\phi : X \longrightarrow (-\infty, \infty]$ is lower semi-continuous, λ-convex and $D(\phi) \neq \emptyset$.

Case I (X, d) *is a* CAT(1)*-space.*

Case II (X, d) *satisfies the commutativity* (3.1) *and the K-convexity of the squared distance function, and ϕ satisfies the coercivity condition (Assumption 2.1(1)). (To be precise, the commutativity, K-convexity and λ-convexity are assumed to hold along the same family of geodesics.)*

We stress that both $\lambda, K \in \mathbb{R}$ can be negative. In Case II, the K-convexity is assumed globally, thus the assertion of Lemma 3.1 holds for any $x, y \in D(\phi)$ and $\tau \in (0, \tau_*(\phi))$. We recall that the coercivity holds if, for instance, the compactness condition (Assumption 2.1(2)) is satisfied. In Case I, the coercivity is guaranteed by restricting ourselves to balls with radii $\leq R < \pi/4$. In these (convex) balls the squared distance function is K-convex with $K = K(2R) > 0$ from Lemma 2.8, and hence $\phi + d^2(x, \cdot)/(2\tau)$ is $(\lambda + K/(2\tau))$-convex. This implies that $J_\tau^\phi(x)$ is nonempty and consists of a single point for $\tau \in (0, -K/(2\lambda))$ even if $\lambda < 0$ (by, for example, [AGS1, Lemma 2.4.8]). By the same reasoning, if $K > 0$ in Case II, then Assumption 2.1(1) is redundant.

To include both cases keeping clarity of the presentation, we discuss under the global K-convexity of the squared distance function and Assumption 2.1(1). Thus we implicitly assume diam $X < \pi/2$ if we are in Case I. This costs no generality for the construction of gradient curves since it is a local problem. We explain how to extend the properties of the gradient flow to the case of diam $X \geq \pi/2$ in §4.6.

4.1 Interpolations

Given an initial point $x_0 \in D(\phi)$ and a partition $\mathscr{P}_{\boldsymbol{\tau}}$ with $|\boldsymbol{\tau}| < \tau_*(\phi)$, we fix a discrete solution $\{x_{\boldsymbol{\tau}}^k\}_{k \in \mathbb{N}}$ of (2.1). Let us also take a point $y \in X$. Similarly to Chapter 4 of [AGS1], we interpolate the discrete data $x_{\boldsymbol{\tau}}^k$, $d(x_{\boldsymbol{\tau}}^k, y)$ and $\phi(x_{\boldsymbol{\tau}}^k)$ as follows (recall (2.4)):

For $t \in (t_{\boldsymbol{\tau}}^{k-1}, t_{\boldsymbol{\tau}}^k]$, $k \in \mathbb{N}$, define

$$\bar{\boldsymbol{x}}_{\boldsymbol{\tau}}(t) := x_{\boldsymbol{\tau}}^k \in J_{\tau_k}^{\phi}(x_{\boldsymbol{\tau}}^{k-1}) \quad (\bar{\boldsymbol{x}}_{\boldsymbol{\tau}}(0) := x_0),$$

$$\bar{\boldsymbol{d}}_{\boldsymbol{\tau}}(t; y) := \left\{ d^2(x_{\boldsymbol{\tau}}^{k-1}, y) + \frac{t - t_{\boldsymbol{\tau}}^{k-1}}{\tau_k} \{ d^2(x_{\boldsymbol{\tau}}^k, y) - d^2(x_{\boldsymbol{\tau}}^{k-1}, y) \} \right\}^{1/2},$$

$$\bar{\boldsymbol{\phi}}_{\boldsymbol{\tau}}(t) := \phi(x_{\boldsymbol{\tau}}^{k-1}) + \frac{t - t_{\boldsymbol{\tau}}^{k-1}}{\tau_k} \{ \phi(x_{\boldsymbol{\tau}}^k) - \phi(x_{\boldsymbol{\tau}}^{k-1}) \}.$$

Recall that $\tau_k = t_{\boldsymbol{\tau}}^k - t_{\boldsymbol{\tau}}^{k-1}$ and note that $\bar{\boldsymbol{\phi}}_{\boldsymbol{\tau}}$ is non-increasing.

Then Lemma 3.1 yields the following discrete version of the *evolution variational inequality* (see [AGS1, Theorem 4.1.4]; we remark that our residual function $\mathscr{R}_{\boldsymbol{\tau}, K}$ is different from that in [AGS1] and depends on K).

Theorem 4.1 (Discrete evolution variational inequality) *Assuming* $|\boldsymbol{\tau}| < \tau_*(\phi)$, *we have*

$$\frac{1}{2} \frac{d}{dt} \left[\bar{\boldsymbol{d}}_{\boldsymbol{\tau}}^2(t; y) \right] + \frac{\lambda}{2} d^2 \left(\bar{\boldsymbol{x}}_{\boldsymbol{\tau}}(t), y \right) + \bar{\boldsymbol{\phi}}_{\boldsymbol{\tau}}(t) - \phi(y) \leq \mathscr{R}_{\boldsymbol{\tau}, K}(t)$$

for almost all $t \in (0, T)$ and all $y \in D(\phi)$, where for $t \in (t_{\boldsymbol{\tau}}^{k-1}, t_{\boldsymbol{\tau}}^k]$

$$\mathscr{R}_{\boldsymbol{\tau}, K}(t) := \left(\frac{t_{\boldsymbol{\tau}}^k - t}{\tau_k} + \frac{\max\{0, -K\}}{2} \right) \{ \phi(x_{\boldsymbol{\tau}}^{k-1}) - \phi(x_{\boldsymbol{\tau}}^k) \}.$$

Proof. Note first that Lemma 3.1 is available merely under $|\boldsymbol{\tau}| < \tau_*(\phi)$ in the current situation. Then we immediately obtain for $t \in (t_{\boldsymbol{\tau}}^{k-1}, t_{\boldsymbol{\tau}}^k)$

$$\frac{1}{2} \frac{d}{dt} \left[\bar{\boldsymbol{d}}_{\boldsymbol{\tau}}^2(t; y) \right] = \frac{d^2(x_{\boldsymbol{\tau}}^k, y) - d^2(x_{\boldsymbol{\tau}}^{k-1}, y)}{2\tau_k}$$

$$\leq -\frac{\lambda}{2} d^2(x_{\boldsymbol{\tau}}^k, y) + \phi(y) - \phi(x_{\boldsymbol{\tau}}^k) + \frac{\max\{0, -K\}}{2} \{ \phi(x_{\boldsymbol{\tau}}^{k-1}) - \phi(x_{\boldsymbol{\tau}}^k) \}$$

$$= -\frac{\lambda}{2} d^2 \left(\bar{\boldsymbol{x}}_{\boldsymbol{\tau}}(t), y \right) + \phi(y) - \bar{\boldsymbol{\phi}}_{\boldsymbol{\tau}}(t) + \frac{t_{\boldsymbol{\tau}}^k - t}{\tau_k} \{ \phi(x_{\boldsymbol{\tau}}^{k-1}) - \phi(x_{\boldsymbol{\tau}}^k) \}$$

$$+ \frac{\max\{0, -K\}}{2} \{ \phi(x_{\boldsymbol{\tau}}^{k-1}) - \phi(x_{\boldsymbol{\tau}}^k) \}.$$

This completes the proof. $\qquad \square$

Taking the limit in Theorem 4.1 as $|\boldsymbol{\tau}| \to 0$ will indeed lead to the (continuous) evolution variational inequality (Theorem 4.8). Another application of Theorem 4.1 is a comparison between two discrete solutions generated from different partitions:

$$\mathscr{P}_{\boldsymbol{\tau}} = \{0 = t_{\boldsymbol{\tau}}^0 < t_{\boldsymbol{\tau}}^1 < \cdots\}, \qquad \mathscr{P}_{\boldsymbol{\sigma}} = \{0 = s_{\boldsymbol{\sigma}}^0 < s_{\boldsymbol{\sigma}}^1 < \cdots\}.$$

We set $\sigma_l := s_{\boldsymbol{\sigma}}^l - s_{\boldsymbol{\sigma}}^{l-1}$ for $l \in \mathbb{N}$. We first observe the following modification of Theorem 4.1 (see [AGS1, Lemma 4.1.6]).

Lemma 4.2 *Suppose $|\boldsymbol{\tau}| < \tau_*(\phi)$ and $\lambda \le 0$. Then we have*

$$\frac{1}{2}\frac{d}{dt}\big[\bar{\boldsymbol{d}}_{\boldsymbol{\tau}}^2(t; y)\big] + \frac{\lambda}{2}\bar{\boldsymbol{d}}_{\boldsymbol{\tau}}^2(t; y) + \lambda \mathscr{D}_{\boldsymbol{\tau}}(t)\bar{\boldsymbol{d}}_{\boldsymbol{\tau}}(t; y) + \bar{\phi}_{\boldsymbol{\tau}}(t) - \phi(y) \le \mathscr{R}_{\boldsymbol{\tau},K}(t) - \frac{\lambda}{2}\mathscr{D}_{\boldsymbol{\tau}}^2(t)$$

for almost all $t \in (0,T)$ and all $y \in D(\phi)$, where for $t \in (t_{\boldsymbol{\tau}}^{k-1}, t_{\boldsymbol{\tau}}^k]$

$$\mathscr{D}_{\boldsymbol{\tau}}(t) := \frac{t_{\boldsymbol{\tau}}^k - t}{\tau_k} d(x_{\boldsymbol{\tau}}^{k-1}, x_{\boldsymbol{\tau}}^k).$$

Proof. This is a consequence of Theorem 4.1 and the inequality

$$d\big(\bar{\boldsymbol{x}}_{\boldsymbol{\tau}}(t), y\big) = d(x_{\boldsymbol{\tau}}^k, y) \le \frac{t - t_{\boldsymbol{\tau}}^{k-1}}{\tau_k} d(x_{\boldsymbol{\tau}}^k, y) + \frac{t_{\boldsymbol{\tau}}^k - t}{\tau_k}\{d(x_{\boldsymbol{\tau}}^{k-1}, y) + d(x_{\boldsymbol{\tau}}^{k-1}, x_{\boldsymbol{\tau}}^k)\}$$

$$\le \bar{\boldsymbol{d}}_{\boldsymbol{\tau}}(t; y) + \mathscr{D}_{\boldsymbol{\tau}}(t),$$

which follows only from the triangle inequality and the convexity of $f(r) = r^2$, $r \in \mathbb{R}$. $\square$

Notice that Lemma 4.2 reduces to Theorem 4.1 when $\lambda = 0$. Applying a version of the Gronwall lemma, we obtain from Lemma 4.2 (or directly from Theorem 4.1 if $\lambda = 0$) a comparison estimate between $\{x_{\boldsymbol{\tau}}^k\}_{k\in\mathbb{N}}$ and $\{y_{\boldsymbol{\sigma}}^l\}_{l\in\mathbb{N}}$ with $y_{\boldsymbol{\sigma}}^0 = y_0 \in D(\phi)$ (see [AGS1, Corollaries 4.1.5, 4.1.7]). For $s \in (s_{\boldsymbol{\sigma}}^{l-1}, s_{\boldsymbol{\sigma}}^l]$, we set

$$\bar{\boldsymbol{d}}_{\boldsymbol{\tau}\boldsymbol{\sigma}}(t, s) := \left\{\bar{\boldsymbol{d}}_{\boldsymbol{\tau}}^2(t; y_{\boldsymbol{\sigma}}^{l-1}) + \frac{s - s_{\boldsymbol{\sigma}}^{l-1}}{\sigma_l}\{\bar{\boldsymbol{d}}_{\boldsymbol{\tau}}^2(t; y_{\boldsymbol{\sigma}}^l) - \bar{\boldsymbol{d}}_{\boldsymbol{\tau}}^2(t; y_{\boldsymbol{\sigma}}^{l-1})\}\right\}^{1/2}.$$

Observe that, for $(t, s) \in (t_{\boldsymbol{\tau}}^{k-1}, t_{\boldsymbol{\tau}}^k] \times (s_{\boldsymbol{\sigma}}^{l-1}, s_{\boldsymbol{\sigma}}^l]$,

$$\bar{\boldsymbol{d}}_{\boldsymbol{\tau}\boldsymbol{\sigma}}^2(t, s) = \left(1 - \frac{t - t_{\boldsymbol{\tau}}^{k-1}}{\tau_k}\right)\left(1 - \frac{s - s_{\boldsymbol{\sigma}}^{l-1}}{\sigma_l}\right) d^2(x_{\boldsymbol{\tau}}^{k-1}, y_{\boldsymbol{\sigma}}^{l-1})$$

$$+ \frac{t - t_{\boldsymbol{\tau}}^{k-1}}{\tau_k}\left(1 - \frac{s - s_{\boldsymbol{\sigma}}^{l-1}}{\sigma_l}\right) d^2(x_{\boldsymbol{\tau}}^k, y_{\boldsymbol{\sigma}}^{l-1})$$

$$+ \left(1 - \frac{t - t_{\boldsymbol{\tau}}^{k-1}}{\tau_k}\right)\frac{s - s_{\boldsymbol{\sigma}}^{l-1}}{\sigma_l} d^2(x_{\boldsymbol{\tau}}^{k-1}, y_{\boldsymbol{\sigma}}^l) + \frac{t - t_{\boldsymbol{\tau}}^{k-1}}{\tau_k}\frac{s - s_{\boldsymbol{\sigma}}^{l-1}}{\sigma_l} d^2(x_{\boldsymbol{\tau}}^k, y_{\boldsymbol{\sigma}}^l).$$

Corollary 4.3 (Comparison between two discrete solutions) *Assume $\lambda \le 0$ and $|\boldsymbol{\tau}|, |\boldsymbol{\sigma}| < \tau_*(\phi)$. Then we have, for almost all $t > 0$,*

$$\frac{d}{dt}\big[\bar{\boldsymbol{d}}_{\boldsymbol{\tau}\boldsymbol{\sigma}}^2(t, t)\big] + 2\lambda\bar{\boldsymbol{d}}_{\boldsymbol{\tau}\boldsymbol{\sigma}}^2(t, t) \le -2\lambda\big(\mathscr{D}_{\boldsymbol{\tau}}(t) + \mathscr{D}_{\boldsymbol{\sigma}}(t)\big)\bar{\boldsymbol{d}}_{\boldsymbol{\tau}\boldsymbol{\sigma}}(t, t)$$

$$+ 2\big(\mathscr{R}_{\boldsymbol{\tau},K}(t) + \mathscr{R}_{\boldsymbol{\sigma},K}(t)\big) - \lambda\big(\mathscr{D}_{\boldsymbol{\tau}}^2(t) + \mathscr{D}_{\boldsymbol{\sigma}}^2(t)\big). \qquad (4.1)$$

11

Moreover, for all $T > 0$,

$$e^{\lambda T} \bar{\boldsymbol{d}}_{\boldsymbol{\tau\sigma}}(T, T) \leq \left\{ d^2(x_0, y_0) + \int_0^T e^{2\lambda t} \left\{ 2\big(\mathscr{R}_{\boldsymbol{\tau}, K}(t) + \mathscr{R}_{\boldsymbol{\sigma}, K}(t)\big) - \lambda\big(\mathscr{D}_{\boldsymbol{\tau}}^2(t) + \mathscr{D}_{\boldsymbol{\sigma}}^2(t)\big) \right\} dt \right\}^{1/2}$$
$$- 2\lambda \int_0^T e^{\lambda t} \big(\mathscr{D}_{\boldsymbol{\tau}}(t) + \mathscr{D}_{\boldsymbol{\sigma}}(t)\big) \, dt. \tag{4.2}$$

Proof. For each fixed s, it follows from Lemma 4.2 that

$$\frac{1}{2} \frac{\partial}{\partial t}\big[\bar{\boldsymbol{d}}_{\boldsymbol{\tau\sigma}}^2(t, s)\big] + \frac{\lambda}{2} \bar{\boldsymbol{d}}_{\boldsymbol{\tau\sigma}}^2(t, s) + \bar{\phi}_{\boldsymbol{\tau}}(t) - \bar{\phi}_{\boldsymbol{\sigma}}(s)$$
$$\leq -\lambda \mathscr{D}_{\boldsymbol{\tau}}(t) \left\{ \left(1 - \frac{s - s_{\boldsymbol{\sigma}}^{l-1}}{\sigma_l}\right) \bar{\boldsymbol{d}}_{\boldsymbol{\tau}}(t; y_{\boldsymbol{\sigma}}^{l-1}) + \frac{s - s_{\boldsymbol{\sigma}}^{l-1}}{\sigma_l} \bar{\boldsymbol{d}}_{\boldsymbol{\tau}}(t; y_{\boldsymbol{\sigma}}^l) \right\} + \mathscr{R}_{\boldsymbol{\tau}, K}(t) - \frac{\lambda}{2} \mathscr{D}_{\boldsymbol{\tau}}^2(t)$$
$$\leq -\lambda \mathscr{D}_{\boldsymbol{\tau}}(t) \bar{\boldsymbol{d}}_{\boldsymbol{\tau\sigma}}(t, s) + \mathscr{R}_{\boldsymbol{\tau}, K}(t) - \frac{\lambda}{2} \mathscr{D}_{\boldsymbol{\tau}}^2(t).$$

Combining this with a similar inequality

$$\frac{1}{2} \frac{\partial}{\partial s}\big[\bar{\boldsymbol{d}}_{\boldsymbol{\tau\sigma}}^2(t, s)\big] + \frac{\lambda}{2} \bar{\boldsymbol{d}}_{\boldsymbol{\tau\sigma}}^2(t, s) + \bar{\phi}_{\boldsymbol{\sigma}}(s) - \bar{\phi}_{\boldsymbol{\tau}}(t) \leq -\lambda \mathscr{D}_{\boldsymbol{\sigma}}(s) \bar{\boldsymbol{d}}_{\boldsymbol{\tau\sigma}}(t, s) + \mathscr{R}_{\boldsymbol{\sigma}, K}(s) - \frac{\lambda}{2} \mathscr{D}_{\boldsymbol{\sigma}}^2(s),$$

we obtain the first inequality (4.1). The second assertion (4.2) is a consequence of (4.1) via a version of the Gronwall lemma (see [AGS1, Lemma 4.1.8]). □

4.2 Convergence of discrete solutions

Corollary 4.3 implies that the discrete solutions $\{x_{\boldsymbol{\tau}}^k\}_{k \in \mathbb{N}}$ converges to a gradient curve as $|\boldsymbol{\tau}| \to 0$ and the limit curve is independent of the choice of the discrete solutions (generalizing [AGS1, Theorem 4.2.2]). Recall §2.1.2 for properties of gradient curves.

Theorem 4.4 (Unique limits of discrete solutions) *Fix an initial point $x_0 \in D(\phi)$ and consider discrete solutions $\{x_{\boldsymbol{\tau}_i}^k\}_{k \in \mathbb{N}}$ with $x_{\boldsymbol{\tau}_i}^k = x_0$ associated with a sequence of partitions $\{\mathscr{P}_{\boldsymbol{\tau}_i}\}_{i \in \mathbb{N}}$ such that $\lim_{i \to \infty} |\boldsymbol{\tau}_i| = 0$. Then the interpolated curve $\bar{\boldsymbol{x}}_{\boldsymbol{\tau}_i} : [0, \infty) \longrightarrow X$ as in §4.1 converges to a curve $\xi : [0, \infty) \longrightarrow X$ with $\xi(0) = x_0$ as $i \to \infty$ uniformly on each bounded interval $[0, T]$. In particular, the limit curve ξ is independent of the choice of the sequence of partitions nor discrete solutions.*

Proof. Since (X, d) is complete, it is sufficient to show $\bar{\boldsymbol{d}}_{\boldsymbol{\tau}_i \boldsymbol{\tau}_j}(t, t) \to 0$ as $i, j \to \infty$ uniformly on $[0, T]$, in the notations of Corollary 4.3. Let $\lambda < 0$ without loss of generality. Take i_0 large enough to satisfy $|\boldsymbol{\tau}_i| \leq \tau_*(\phi)/8$ for all $i \geq i_0$, and put $K' := \min\{0, K\}$.

The integral of $\mathscr{R}_{\boldsymbol{\tau}, K}$ (recall Theorem 4.1 for the definition) is calculated as

$$\int_{t_{\boldsymbol{\tau}}^{k-1}}^{t_{\boldsymbol{\tau}}^k} \mathscr{R}_{\boldsymbol{\tau}, K} \, dt = \left(\frac{1}{2} - \frac{K'}{2}\right) \tau_k \{\phi(x_{\boldsymbol{\tau}}^{k-1}) - \phi(x_{\boldsymbol{\tau}}^k)\}.$$

Similarly, together with the canonical estimate $d^2(x_{\boldsymbol{\tau}}^{k-1}, x_{\boldsymbol{\tau}}^k) \leq 2\tau_k\{\phi(x_{\boldsymbol{\tau}}^{k-1}) - \phi(x_{\boldsymbol{\tau}}^k)\}$, we have

$$\int_{t_{\boldsymbol{\tau}}^{k-1}}^{t_{\boldsymbol{\tau}}^k} \mathscr{D}_{\boldsymbol{\tau}}^2 \, dt \leq \frac{\tau_k}{3} \cdot 2\tau_k\{\phi(x_{\boldsymbol{\tau}}^{k-1}) - \phi(x_{\boldsymbol{\tau}}^k)\}.$$

This also implies

$$\int_0^{t_\tau^k} e^{\lambda t} \mathscr{D}_\tau(t)\,dt \le \left(\int_0^{t_\tau^k} e^{2\lambda t}\,dt\right)^{1/2}\left(\int_0^{t_\tau^k} \mathscr{D}_\tau^2\,dt\right)^{1/2} \le \sqrt{-\frac{1}{2\lambda}}\sqrt{\frac{2}{3}}|\tau|\sqrt{\phi(x_0) - \phi(x_\tau^k)}.$$

Combining these with (4.2), we obtain for $i, j \ge i_0$, $t \le t_{\tau_i}^k \le T$ and $t \le t_{\tau_j}^l \le T$,

$$e^{\lambda t}\bar{d}_{\tau_i \tau_j}(t, t) \le \left\{\int_0^t \{2(\mathscr{R}_{\tau_i, K} + \mathscr{R}_{\tau_j, K}) - \lambda(\mathscr{D}_{\tau_i}^2 + \mathscr{D}_{\tau_j}^2)\}\,ds\right\}^{1/2}$$

$$- 2\lambda \int_0^t e^{\lambda s}\big(\mathscr{D}_{\tau_i}(s) + \mathscr{D}_{\tau_j}(s)\big)\,ds$$

$$\le \left\{\left(1 - K' - \frac{2\lambda}{3}|\tau_i|\right)|\tau_i|\{\phi(x_0) - \phi(x_{\tau_i}^k)\}\right.$$

$$\left. + \left(1 - K' - \frac{2\lambda}{3}|\tau_j|\right)|\tau_j|\{\phi(x_0) - \phi(x_{\tau_j}^l)\}\right\}^{1/2}$$

$$+ \sqrt{-\frac{4\lambda}{3}}\left(|\tau_i|\sqrt{\phi(x_0) - \phi(x_{\tau_i}^k)} + |\tau_j|\sqrt{\phi(x_0) - \phi(x_{\tau_j}^l)}\right).$$

Thanks to the a priori estimate (Lemma 2.3), we have

$$\max\{\phi(x_0) - \phi(x_{\tau_i}^k), \phi(x_0) - \phi(x_{\tau_j}^l)\} \le C = C(x_0, \tau_*(\phi), \phi(x_0), T).$$

Therefore we conclude that $\bar{d}_{\tau_i \tau_j}(t, t)$ tends to 0 as $i, j \to \infty$, uniformly in $t \in [0, T]$. $\quad\square$

By virtue of the uniqueness, we can define the *gradient flow* operator

$$\mathcal{G} : [0, \infty) \times D(\phi) \longrightarrow D(\phi) \tag{4.3}$$

by $\mathcal{G}(t, x_0) := \xi(t)$, where $\xi : [0, \infty) \longrightarrow X$ is the unique gradient curve with $\xi(0) = x_0$ given by Theorem 4.4. Then the semigroup property:

$$\mathcal{G}\big(t, \mathcal{G}(s, x_0)\big) = \mathcal{G}(s + t, x_0) \quad \text{for all } s, t \ge 0$$

also follows from the uniqueness of gradient curves.

One can immediately obtain the following (rough) error estimate from the proof of Theorem 4.4 (compare this with [AGS1, Theorem 4.0.9]).

Corollary 4.5 (An error estimate) *Let $\lambda \le 0$ and $|\tau| < \tau_*(\phi)$, fix $x_0 \in D(\phi)$, and put $\xi(t) := \mathcal{G}(t, x_0)$. Then we have*

$$\bar{d}_\tau^2\big(t; \xi(t)\big) \le e^{-2\lambda t}\left(\sqrt{1 - K' - \frac{2\lambda}{3}|\tau|} + \sqrt{-\frac{4\lambda}{3}|\tau|}\right)^2 |\tau|\left\{\phi(x_0) - \phi\big(\bar{x}_\tau(t)\big)\right\}$$

for all $t > 0$, where $K' := \min\{0, K\}$.

Proof. Taking the limit of (4.2) as $|\boldsymbol{\sigma}| \to 0$ and using the estimates in the proof of Theorem 4.4, we have for $t \in (t_{\boldsymbol{\tau}}^{k-1}, t_{\boldsymbol{\tau}}^k]$

$$
\begin{aligned}
e^{\lambda t} \bar{\boldsymbol{d}}_{\boldsymbol{\tau}}\big(t; \xi(t)\big) &\leq \left\{ \int_0^t (2\mathscr{R}_{\boldsymbol{\tau},K} - \lambda \mathscr{D}_{\boldsymbol{\tau}}^2)\, ds \right\}^{1/2} - 2\lambda \int_0^t e^{\lambda s} \mathscr{D}_{\boldsymbol{\tau}}(s)\, ds \\
&\leq \left\{ \left(1 - K' - \frac{2\lambda}{3}|\boldsymbol{\tau}| \right) |\boldsymbol{\tau}| \{\phi(x_0) - \phi(x_{\boldsymbol{\tau}}^k)\} \right\}^{1/2} \\
&\quad + \sqrt{-\frac{4\lambda}{3}|\boldsymbol{\tau}|} \sqrt{\phi(x_0) - \phi(x_{\boldsymbol{\tau}}^k)} \\
&= \left(\sqrt{1 - K' - \frac{2\lambda}{3}|\boldsymbol{\tau}|} + \sqrt{-\frac{4\lambda}{3}|\boldsymbol{\tau}|} \right) \sqrt{|\boldsymbol{\tau}|} \sqrt{\phi(x_0) - \phi\big(\bar{\boldsymbol{x}}_{\boldsymbol{\tau}}(t)\big)}.
\end{aligned}
$$

This completes the proof. $\qquad\square$

4.3 Contraction property

Coming back to the discrete scheme, we show the following lemma, which readily implies the *contraction property* of $\mathcal{G}$ (Theorem 4.7). Set

$$
\lambda_{\boldsymbol{\tau}} := \frac{\log(1 + \lambda|\boldsymbol{\tau}|)}{|\boldsymbol{\tau}|}
$$

assuming $\lambda|\boldsymbol{\tau}| > -1$ (if $\lambda < 0$), and observe that $\lambda_{\boldsymbol{\tau}} \leq \lambda$ and

$$
\frac{\log(1 + \lambda \tau_k)}{\tau_k} \geq \lambda_{\boldsymbol{\tau}} \tag{4.4}
$$

for all $k \in \mathbb{N}$ (see [AGS1, Lemma 3.4.1]).

Lemma 4.6 (Discrete contraction estimate) *Take $x_0, y_0 \in D(\phi)$ and $\mathscr{P}_{\boldsymbol{\tau}}$ with $|\boldsymbol{\tau}| < \tau_*(\phi)/8$. Assume $\lambda|\boldsymbol{\tau}| > -1$ if $\lambda < 0$. Then we have, for $t \in (t_{\boldsymbol{\tau}}^{k-1}, t_{\boldsymbol{\tau}}^k]$ with $t_{\boldsymbol{\tau}}^k \leq T$,*

$$
\begin{aligned}
e^{2(\lambda_{\boldsymbol{\tau}} t + \lambda_{\boldsymbol{\tau}}^- |\boldsymbol{\tau}|)} d^2\big(\bar{\boldsymbol{x}}_{\boldsymbol{\tau}}(t), \bar{\boldsymbol{y}}_{\boldsymbol{\tau}}(t)\big) &\leq d^2(x_0, y_0) + 2e^{2\lambda_{\boldsymbol{\tau}}^+ t_{\boldsymbol{\tau}}^k} |\boldsymbol{\tau}| \{\phi(y_0) - \phi(y_{\boldsymbol{\tau}}^k)\} \\
&\quad - K' e^{2\lambda_{\boldsymbol{\tau}}^+ t_{\boldsymbol{\tau}}^k} |\boldsymbol{\tau}| \{\phi(x_0) - \phi(x_{\boldsymbol{\tau}}^k) + \phi(y_0) - \phi(y_{\boldsymbol{\tau}}^k)\} \\
&\quad + O_{x_0, y_0, T}(\sqrt{|\boldsymbol{\tau}|}),
\end{aligned}
$$

where $K' := \min\{0, K\}$, $\lambda_{\boldsymbol{\tau}}^- := \min\{0, \lambda_{\boldsymbol{\tau}}\}$ and $\lambda_{\boldsymbol{\tau}}^+ := \max\{0, \lambda_{\boldsymbol{\tau}}\}$.

Proof. The proof is along the line of [AGS1, Lemma 4.2.4]. Applying Lemma 3.1, we have for each $k \in \mathbb{N}$

$$
\begin{aligned}
d^2(x_{\boldsymbol{\tau}}^k, y_{\boldsymbol{\tau}}^{k-1}) &\leq d^2(x_{\boldsymbol{\tau}}^{k-1}, y_{\boldsymbol{\tau}}^{k-1}) - \lambda \tau_k d^2(x_{\boldsymbol{\tau}}^k, y_{\boldsymbol{\tau}}^{k-1}) + 2\tau_k \{\phi(y_{\boldsymbol{\tau}}^{k-1}) - \phi(x_{\boldsymbol{\tau}}^k)\} \\
&\quad - K' \tau_k \{\phi(x_{\boldsymbol{\tau}}^{k-1}) - \phi(x_{\boldsymbol{\tau}}^k)\},
\end{aligned}
$$

14

and

$$d^2(x_{\boldsymbol\tau}^k, y_{\boldsymbol\tau}^k) \le d^2(x_{\boldsymbol\tau}^k, y_{\boldsymbol\tau}^{k-1}) - \lambda\tau_k d^2(x_{\boldsymbol\tau}^k, y_{\boldsymbol\tau}^k) + 2\tau_k\{\phi(x_{\boldsymbol\tau}^k) - \phi(y_{\boldsymbol\tau}^k)\}$$
$$- K'\tau_k\{\phi(y_{\boldsymbol\tau}^{k-1}) - \phi(y_{\boldsymbol\tau}^k)\}.$$

Thus we have

$$(1 + \lambda\tau_k)d^2(x_{\boldsymbol\tau}^k, y_{\boldsymbol\tau}^k) \le d^2(x_{\boldsymbol\tau}^{k-1}, y_{\boldsymbol\tau}^{k-1}) - \lambda\tau_k d^2(x_{\boldsymbol\tau}^k, y_{\boldsymbol\tau}^{k-1}) + 2\tau_k\{\phi(y_{\boldsymbol\tau}^{k-1}) - \phi(y_{\boldsymbol\tau}^k)\}$$
$$- K'\tau_k\{\phi(x_{\boldsymbol\tau}^{k-1}) - \phi(x_{\boldsymbol\tau}^k) + \phi(y_{\boldsymbol\tau}^{k-1}) - \phi(y_{\boldsymbol\tau}^k)\}.$$

Note that

$$|d^2(x_{\boldsymbol\tau}^k, y_{\boldsymbol\tau}^{k-1}) - d^2(x_{\boldsymbol\tau}^{k-1}, y_{\boldsymbol\tau}^{k-1})| \le \{d(x_{\boldsymbol\tau}^k, y_{\boldsymbol\tau}^{k-1}) + d(x_{\boldsymbol\tau}^{k-1}, y_{\boldsymbol\tau}^{k-1})\}d(x_{\boldsymbol\tau}^k, x_{\boldsymbol\tau}^{k-1})$$
$$= O_{x_0,y_0,T}(\sqrt{|\boldsymbol\tau|})$$

by the a priori estimate (Lemma 2.3, (2.2)). Together with $1 - \lambda\tau_k \le (1 + \lambda\tau_k)^{-1}$, this implies

$$(1 + \lambda\tau_k)d^2(x_{\boldsymbol\tau}^k, y_{\boldsymbol\tau}^k) \le \frac{1}{1 + \lambda\tau_k}d^2(x_{\boldsymbol\tau}^{k-1}, y_{\boldsymbol\tau}^{k-1}) + 2\tau_k\{\phi(y_{\boldsymbol\tau}^{k-1}) - \phi(y_{\boldsymbol\tau}^k)\}$$
$$- K'\tau_k\{\phi(x_{\boldsymbol\tau}^{k-1}) - \phi(x_{\boldsymbol\tau}^k) + \phi(y_{\boldsymbol\tau}^{k-1}) - \phi(y_{\boldsymbol\tau}^k)\}$$
$$+ \tau_k \cdot O_{x_0,y_0,T}(\sqrt{|\boldsymbol\tau|}).$$

Multiplying both sides by $e^{\lambda_{\boldsymbol\tau}(2t_{\boldsymbol\tau}^{k-1}+\tau_k)} = e^{\lambda_{\boldsymbol\tau}(2t_{\boldsymbol\tau}^k-\tau_k)}$ yields that, since

$$(1 + \lambda\tau_k)e^{-\lambda_{\boldsymbol\tau}\tau_k}e^{2\lambda_{\boldsymbol\tau}t_{\boldsymbol\tau}^k} \ge e^{2\lambda_{\boldsymbol\tau}t_{\boldsymbol\tau}^k}, \qquad e^{2\lambda_{\boldsymbol\tau}t_{\boldsymbol\tau}^{k-1}}\frac{e^{\lambda_{\boldsymbol\tau}\tau_k}}{1 + \lambda\tau_k} \le e^{2\lambda_{\boldsymbol\tau}t_{\boldsymbol\tau}^{k-1}}$$

by (4.4),

$$e^{2\lambda_{\boldsymbol\tau}t_{\boldsymbol\tau}^k}d^2(x_{\boldsymbol\tau}^k, y_{\boldsymbol\tau}^k) \le e^{2\lambda_{\boldsymbol\tau}t_{\boldsymbol\tau}^{k-1}}d^2(x_{\boldsymbol\tau}^{k-1}, y_{\boldsymbol\tau}^{k-1}) + 2e^{\lambda_{\boldsymbol\tau}(t_{\boldsymbol\tau}^{k-1}+t_{\boldsymbol\tau}^k)}\tau_k\{\phi(y_{\boldsymbol\tau}^{k-1}) - \phi(y_{\boldsymbol\tau}^k)\}$$
$$- K'e^{\lambda_{\boldsymbol\tau}(t_{\boldsymbol\tau}^{k-1}+t_{\boldsymbol\tau}^k)}\tau_k\{\phi(x_{\boldsymbol\tau}^{k-1}) - \phi(x_{\boldsymbol\tau}^k) + \phi(y_{\boldsymbol\tau}^{k-1}) - \phi(y_{\boldsymbol\tau}^k)\}$$
$$+ \tau_k \cdot O_{x_0,y_0,T}(\sqrt{|\boldsymbol\tau|}).$$

Summing up, for $t \in (t_{\boldsymbol\tau}^{k-1}, t_{\boldsymbol\tau}^k]$, we obtain the desired estimate

$$e^{2(\lambda_{\boldsymbol\tau}t+\lambda_{\boldsymbol\tau}^-|\boldsymbol\tau|)}d^2\big(\bar{\boldsymbol{x}}_{\boldsymbol\tau}(t), \bar{\boldsymbol{y}}_{\boldsymbol\tau}(t)\big) \le e^{2\lambda_{\boldsymbol\tau}t_{\boldsymbol\tau}^k}d^2(x_{\boldsymbol\tau}^k, y_{\boldsymbol\tau}^k)$$
$$\le d^2(x_0, y_0) + 2e^{2\lambda_{\boldsymbol\tau}^+ t_{\boldsymbol\tau}^k}|\boldsymbol\tau|\{\phi(y_0) - \phi(y_{\boldsymbol\tau}^k)\}$$
$$- K'e^{2\lambda_{\boldsymbol\tau}^+ t_{\boldsymbol\tau}^k}|\boldsymbol\tau|\{\phi(x_0) - \phi(x_{\boldsymbol\tau}^k) + \phi(y_0) - \phi(y_{\boldsymbol\tau}^k)\}$$
$$+ t_{\boldsymbol\tau}^k \cdot O_{x_0,y_0,T}(\sqrt{|\boldsymbol\tau|}).$$

$\square$

Theorem 4.7 (Contraction property) *Take $x_0, y_0 \in D(\phi)$ and put $\xi(t) := \mathcal{G}(t, x_0)$ and $\zeta(t) := \mathcal{G}(t, y_0)$. Then we have, for any $t > 0$,*

$$d\big(\xi(t), \zeta(t)\big) \le e^{-\lambda t}d(x_0, y_0). \tag{4.5}$$

15

Proof. Take the limit as $|\boldsymbol{\tau}| \to 0$ in Lemma 4.6. Then the claim follows from $\lim_{|\boldsymbol{\tau}| \to 0} \lambda_{\boldsymbol{\tau}} = \lambda$ and the a priori estimate in Lemma 2.3 (which bounds $\phi(x_0) - \phi(x_{\boldsymbol{\tau}}^k)$). $\qquad\square$

The contraction property allows us to take the continuous limit

$$\mathcal{G} : [0, \infty) \times \overline{D(\phi)} \longrightarrow \overline{D(\phi)}$$

of the gradient flow operator in (4.3), which again enjoys the semigroup property as well as the contraction property (4.5).

4.4   Evolution variational inequality

Similarly to [AGS1, Theorem 4.3.2], taking the limit of Theorem 4.1, we obtain the following.

Theorem 4.8 (Evolution variational inequality) *Take $x_0 \in D(\phi)$ and put $\xi(t) := \mathcal{G}(t, x_0)$. Then we have*

$$\limsup_{\varepsilon \downarrow 0} \frac{d^2(\xi(t + \varepsilon), y) - d^2(\xi(t), y)}{2\varepsilon} + \frac{\lambda}{2} d^2\big(\xi(t), y\big) + \phi\big(\xi(t)\big) \leq \phi(y) \qquad (4.6)$$

for all $y \in D(\phi)$ and $t > 0$. In particular,

$$\frac{1}{2} \frac{d}{dt} \big[d^2\big(\xi(t), y\big) \big] + \frac{\lambda}{2} d^2\big(\xi(t), y\big) + \phi\big(\xi(t)\big) \leq \phi(y)$$

for all $y \in D(\phi)$ and almost all $t > 0$.

Proof. By recalling the estimate of the integral of $\mathscr{R}_{\boldsymbol{\tau}, K}$ in the proof of Theorem 4.4, integration in $t \in [S, T]$ of Theorem 4.1 gives

$$\frac{1}{2} \bar{\boldsymbol{d}}_{\boldsymbol{\tau}}^2(T; y) - \frac{1}{2} \bar{\boldsymbol{d}}_{\boldsymbol{\tau}}^2(S; y) + \int_S^T \left\{ \frac{\lambda}{2} d^2\big(\bar{\boldsymbol{x}}_{\boldsymbol{\tau}}(t), y\big) + \bar{\boldsymbol{\phi}}_{\boldsymbol{\tau}}(t) \right\} dt$$

$$\leq (T - S)\phi(y) + \left(\frac{1}{2} - \frac{K'}{2} \right) |\boldsymbol{\tau}| \left\{ \phi\big(\bar{\boldsymbol{x}}_{\boldsymbol{\tau}}(S - |\boldsymbol{\tau}|)\big) - \phi\big(\bar{\boldsymbol{x}}_{\boldsymbol{\tau}}(T)\big) \right\}.$$

Note that $\bar{\boldsymbol{\phi}}_{\boldsymbol{\tau}}$ is uniformly bounded on $[0, T]$ thanks to the a priori estimate (Lemma 2.3). Thus we have

$$\int_S^T \phi \circ \xi \, dt \leq \int_S^T \liminf_{|\boldsymbol{\tau}| \to 0} \bar{\boldsymbol{\phi}}_{\boldsymbol{\tau}} \, dt \leq \liminf_{|\boldsymbol{\tau}| \to 0} \int_S^T \bar{\boldsymbol{\phi}}_{\boldsymbol{\tau}} \, dt$$

by the lower semi-continuity of ϕ and Fatou's lemma. Therefore letting $|\boldsymbol{\tau}| \downarrow 0$ shows the integrated form of the evolution variational inequality:

$$\frac{d^2(\xi(T), y) - d^2(\xi(S), y)}{2} + \int_S^T \left\{ \frac{\lambda}{2} d^2\big(\xi(t), y\big) + \phi\big(\xi(t)\big) \right\} dt \leq (T - S)\phi(y).$$

Dividing both sides by $T - S$ and letting $T - S \downarrow 0$, we obtain the desired inequality by the lower semi-continuity of ϕ ($\phi \circ \xi$ is in fact continuous; see [AGS1, Theorem 2.4.15]). $\square$

4.5 Stationary points and large time behavior of the flow

In this subsection, following the argumentation in [Ma] (on CAT(0)-spaces), we study stationary points and the large time behavior of the gradient flow $\mathcal{G}$. Since the fundamental properties of the flow, for establishing those the CAT(0)-property is used in [Ma, Section 1], is already in hand, we can follow the line of [Ma, Section 2] almost verbatim. We begin with a consequence of the evolution variational inequality (Theorem 4.8) corresponding to [Ma, Lemma 2.8].

Lemma 4.9 *Take $x_0 \in D(\phi)$ and put $\xi(t) := \mathcal{G}(t, x_0)$. Then we have*

$$d^2\big(\xi(T), y\big) \leq e^{-\lambda T} d^2(x_0, y) + 2e^{-\lambda T} \int_0^T e^{\lambda t} \big\{ \phi(y) - \phi\big(\xi(t)\big) \big\} \, dt$$

for all $T > 0$ and $y \in D(\phi)$. In particular, we have

$$d^2\big(\xi(T), y\big) \leq e^{-\lambda T} d^2(x_0, y) - \frac{2(1 - e^{-\lambda T})}{\lambda} \big\{ \phi\big(\xi(T)\big) - \phi(y) \big\},$$

where $(1 - e^{-\lambda T})/\lambda := T$ when $\lambda = 0$.

Proof. Rewrite (4.6) as

$$\limsup_{\varepsilon \downarrow 0} \frac{e^{\lambda(t+\varepsilon)} d^2(\xi(t+\varepsilon), y) - e^{\lambda t} d^2(\xi(t), y)}{2\varepsilon} \leq e^{\lambda t} \big\{ \phi(y) - \phi\big(\xi(t)\big) \big\},$$

which implies the first claim. The second claim readily follows from this and the fact that $\phi(\xi(t))$ is non-increasing in t. □

By the above lemma, one can show a characterization of stationary points of the flow $\mathcal{G}$ in terms of the local slope $|\nabla \phi|$.

Theorem 4.10 *A point $x_0 \in D(\phi)$ satisfies $|\nabla \phi|(x_0) = 0$ if and only if $\mathcal{G}(t, x_0) = x_0$ for all $t > 0$.*

Proof. The "only if" part is a consequence of [Ma, Lemma 2.11] asserting that

$$|\nabla \phi|(x_0) = 0 \quad \text{if and only if} \quad \sup_{x \neq x_0} \frac{\phi(x_0) - \phi(x)}{d^2(x_0, x)} < \infty, \tag{4.7}$$

for which only the λ-convexity of ϕ is used. The "if" part follows from the same relation (4.7) and Lemma 4.9, noticing that $(1 - e^{-\lambda T})/\lambda \geq T$ for $\lambda < 0$. See [Ma, Theorem 2.12] for details. □

It is natural to expect that, if $\phi \circ \xi$ does not diverge to $-\infty$, then $|\nabla \phi| \circ \xi$ tends to 0. This is indeed the case as follows.

Lemma 4.11 *Assume $\lambda \leq 0$, take $x_0 \in D(\phi)$ and put $\xi(t) := \mathcal{G}(t, x_0)$. Then we have*

$$|\nabla \phi|\big(\xi(T)\big) - |\nabla \phi|\big(\xi(S)\big) \leq -\sqrt{2}\lambda \int_S^T |\nabla \phi| \circ \xi \, dt \tag{4.8}$$

for all $0 < S < T$.

Proof. For any $x \in D(\phi)$ and $x_\tau \in J_\tau^\phi(x)$, it follows from the λ-convexity of ϕ that

$$|\nabla\phi|(x_\tau) \leq |\nabla\phi|(x) - \lambda d(x, x_\tau)$$

(see [Ma, Lemma 2.23]). Substituting $d^2(x, x_\tau) \leq 2\tau\{\phi(x) - \phi(x_\tau)\}$ and iterating this estimate, one finds

$$|\nabla\phi|(x_{\boldsymbol{\tau}}^N) \leq |\nabla\phi|(x_0) - \lambda \sum_{k=1}^{N} \sqrt{2\tau_k\{\phi(x_{\boldsymbol{\tau}}^{k-1}) - \phi(x_{\boldsymbol{\tau}}^k)\}}.$$

Applying the Cauchy–Schwarz inequality:

$$\sum_{k=1}^{N} \sqrt{\tau_k\{\phi(x_{\boldsymbol{\tau}}^{k-1}) - \phi(x_{\boldsymbol{\tau}}^k)\}} \leq \sqrt{\sum_{k=1}^{N} \tau_k} \sqrt{\sum_{k=1}^{N} \{\phi(x_{\boldsymbol{\tau}}^{k-1}) - \phi(x_{\boldsymbol{\tau}}^k)\}}$$

and taking the limit as $|\boldsymbol{\tau}| \to 0$, we have

$$|\nabla\phi|\big(\xi(T)\big) - |\nabla\phi|(x_0) \leq -\sqrt{2}\lambda\sqrt{T}\sqrt{\phi(x_0) - \phi\big(\xi(T)\big)} \qquad (4.9)$$

for all $T > 0$, since ϕ and $|\nabla\phi|$ are lower semi-continuous (by [Ma, Proposition 2.25] or [AGS1, Corollary 2.4.10]). By replacing x_0 with $\xi(S)$, the implication from (4.9) to (4.8) is the same as [Ma, Lemma 2.27]. $\square$

Theorem 4.12 (Large time behavior) *Take $x_0 \in D(\phi)$, put $\xi(t) := \mathcal{G}(t, x_0)$ and assume $\lim_{t \to \infty} \phi(\xi(t)) > -\infty$. Then we have $\lim_{t \to \infty} |\nabla\phi|(\xi(t)) = 0$.*

Proof. The proof is done by contradiction with the help of the estimate (4.8) and the right continuity of $|\nabla\phi| \circ \xi$ (see [Ma, Corollary 2.28] or [AGS1, Theorem 2.4.15]). We refer to [Ma, Theorem 2.30] for details. $\square$

The following corollary is immediate, see [Ma, Corollary 2.31].

Corollary 4.13 *Take $x_0 \in D(\phi)$, put $\xi(t) := \mathcal{G}(t, x_0)$ and assume that there is a sequence $\{t_n\}_{n \in \mathbb{N}}$ such that $\lim_{n \to \infty} t_n = \infty$ and $\{\xi(t_n)\}_{n \in \mathbb{N}}$ converges to a point $\bar{x}$. Then $\bar{x}$ is a stationary point of ϕ (in the sense of Theorem 4.10) and $\lim_{t \to \infty} \phi(\xi(t)) = \phi(\bar{x})$.*

In general, $\lim_{t \to \infty} |\nabla\phi|(\xi(t)) = 0$ does not imply the convergence to a stationary point. One needs some compactness condition to find a stationary point, see for instance [Ma, Theorem 2.32].

4.6   The case of $\mathrm{CAT}(1)$-spaces with diameter $\geq \pi/2$

All the results in this section are generalized to complete $\mathrm{CAT}(1)$-spaces (X, d) with $\mathrm{diam}\, X \geq \pi/2$. First of all, given $x_0 \in D(\phi)$, one can restrict the construction of the gradient curve in, say, the open ball $B(x_0, \pi/6)$. Since $B(x_0, \pi/6)$ is (geodesically) convex, the squared distance function in this ball is K-convex with $K = K(\pi/3) > 0$ from Lemma 2.8, and we obtain the gradient curve ξ with $\xi(0) = x_0$. Once $\xi(t)$ hits the

boundary $\partial B(x_0, \pi/6)$ at $t = t_1$, we restart the construction in $B(\xi(t_1), \pi/6)$. We remark that $t_1 \geq (\pi/6)^2/(2C)$ by (2.2). Iterating this procedure gives the gradient curve ξ : $[0, \infty) \longrightarrow X$.

The contraction property and the evolution variational inequality are globalized in a standard way as follows. (Then Theorems 4.10, 4.12 also hold true since they are based only on the evolution variational inequality.)

For the contraction property (Theorem 4.7), if $d(x_0, y_0) \geq \pi/2$, then we consider a minimal geodesic γ from x_0 to y_0 and choose points $z_0 = x_0, z_1, \ldots, z_{m-1}, z_m = y_0$ on γ such that $\max_{1 \leq l \leq m} d(z_{l-1}, z_l) \ll \pi e^{-|\lambda|T}/2$ for given $T > 0$. Applying Theorem 4.7 to adjacent gradient curves $\xi_l := \mathcal{G}(\cdot, z_l)$ shows $d(\xi_{l-1}(t), \xi_l(t)) \leq e^{-\lambda t} d(z_{l-1}, z_l)$ for $t \in [0, T]$. This yields $d(\xi(t), \zeta(t)) \leq e^{-\lambda t} d(x_0, y_0)$ by the triangle inequality.

For the evolution variational inequality (Theorem 4.8), given a minimal geodesic γ : $[0, 1] \longrightarrow X$ from $\xi(t)$ to y, it is easy to see that (4.6) for $y = \gamma(s)$ with small $s > 0$ (so that $d(\xi(t), \gamma(s)) \ll \pi/2$) implies (4.6) itself. Indeed, since

$$\frac{d^2(\xi(t+\varepsilon), y) - d^2(\xi(t), y)}{2\varepsilon}$$
$$\leq \frac{\{d(\xi(t+\varepsilon), \gamma(s)) + d(\gamma(s), y)\}^2 - \{d(\xi(t), \gamma(s)) + d(\gamma(s), y)\}^2}{2\varepsilon}$$
$$= \frac{d^2(\xi(t+\varepsilon), \gamma(s)) - d^2(\xi(t), \gamma(s))}{2\varepsilon} + \frac{d(\xi(t+\varepsilon), \gamma(s)) - d(\xi(t), \gamma(s))}{\varepsilon} d(\gamma(s), y)$$
$$= \frac{d^2(\xi(t+\varepsilon), \gamma(s)) - d^2(\xi(t), \gamma(s))}{2\varepsilon} \left(1 + \frac{2d(\gamma(s), y)}{d(\xi(t+\varepsilon), \gamma(s)) + d(\xi(t), \gamma(s))}\right),$$

we obtain from (4.6) with $y = \gamma(s)$ that

$$\limsup_{\varepsilon \downarrow 0} \frac{d^2(\xi(t+\varepsilon), y) - d^2(\xi(t), y)}{2\varepsilon}$$
$$\leq \left\{-\frac{\lambda}{2} d^2(\xi(t), \gamma(s)) - \phi(\xi(t)) + \phi(\gamma(s))\right\} \left(1 + \frac{d(\gamma(s), y)}{d(\xi(t), \gamma(s))}\right)$$
$$\leq \frac{1}{s} \left\{-\frac{\lambda}{2} s^2 d^2(\xi(t), y) - \phi(\xi(t)) + (1-s)\phi(\xi(t)) + s\phi(y) - \frac{\lambda}{2}(1-s)s d^2(\xi(t), y)\right\}$$
$$= -\frac{\lambda}{2} d^2(\xi(t), y) - \phi(\xi(t)) + \phi(y).$$

5 A Trotter–Kato product formula

This final section is devoted to a further application of our key lemma: a Trotter–Kato product formula for semi-convex functions. See [KM] for the classical setting of convex functions on Hilbert spaces. The Trotter–Kato product formula on metric spaces was established by Stojkovic [Sto] for convex functions on CAT(0)-spaces in terms of ultra-limits (see also a recent result [Ba1] in terms of weak convergence), and by Clément and Maas [CM] for functions satisfying the assertion of our key lemma (Lemma 3.1) with $K = 2$ and $\lambda = 0$ (thus including convex functions on CAT(0)-spaces). We stress that, similarly to the previous section, both the squared distance function and potential functions are allowed to be semi-convex in our argument.

5.1 Setting and the main theorem

Assumption 5.1 Let (X, d) be a complete metric space in either Case I or Case II, and assume additionally $D := \operatorname{diam} X < \infty$. For $i = 1, 2$, we consider a lower semi-continuous, λ_i-convex function $\phi_i : X \longrightarrow (-\infty, \infty]$ $(\lambda_i \in \mathbb{R})$ satisfying $D(\phi_1) \cap D(\phi_2) \neq \emptyset$ and the compactness (Assumption 2.1(2)).

We remark that λ_i can be negative. The sum $\phi := \phi_1 + \phi_2$ is clearly lower semi-continuous, $(\lambda_1 + \lambda_2)$-convex and enjoys Assumption 2.1(2) (with $\tau_*(\phi) = \infty$) since ϕ_i is bounded below (Remark 2.2) and

$$\{x \in X \mid \phi(x) \leq Q\} \subset \{x \in X \mid \phi_1(x) \leq Q - \inf_X \phi_2\}.$$

Given $z_0 \in D(\phi) = D(\phi_1) \cap D(\phi_2)$ and a partition $\mathscr{P}_{\boldsymbol{\tau}}$, we consider the discrete variational schemes for ϕ_1 and ϕ_2 in turn, namely

$$z_{\boldsymbol{\tau}}^0 := z_0, \text{ choose arbitrary } \hat{z}_{\boldsymbol{\tau}}^k \in J_{\tau_k}^{\phi_1}(z_{\boldsymbol{\tau}}^{k-1}) \text{ and then } z_{\boldsymbol{\tau}}^k \in J_{\tau_k}^{\phi_2}(\hat{z}_{\boldsymbol{\tau}}^k) \text{ for } k \in \mathbb{N}. \tag{5.1}$$

If $\phi_1 = \phi_2$, then this scheme reduces to (2.1) for ϕ with respect to the partition:

$$\left\{ 0 < \frac{t_{\boldsymbol{\tau}}^1}{2} < t_{\boldsymbol{\tau}}^1 < \frac{t_{\boldsymbol{\tau}}^1 + t_{\boldsymbol{\tau}}^2}{2} < t_{\boldsymbol{\tau}}^2 < \cdots \right\}.$$

The *Trotter–Kato product formula* asserts that $\{z_{\boldsymbol{\tau}}^k\}_{k \geq 0}$ converges to the gradient curve of ϕ emanating from z_0 in an appropriate sense. This is useful when ϕ_1 and ϕ_2 are easier to handle than their sum ϕ. An additional difficulty (in the discrete scheme) compared with the direct variational approximation for ϕ is that we have a priori no control of $\phi_2(\hat{z}_{\boldsymbol{\tau}}^k) - \phi_2(z_{\boldsymbol{\tau}}^{k-1})$ and $\phi_1(z_{\boldsymbol{\tau}}^k) - \phi_1(\hat{z}_{\boldsymbol{\tau}}^k)$ (both are being nonpositive if $\phi_1 = \phi_2$). Thus we suppose the following condition:

Assumption 5.2 Given $z_0 \in D(\phi)$ and a partition $\mathscr{P}_{\boldsymbol{\tau}}$, set

$$\delta_{\boldsymbol{\tau}}^k(z_0) := \max\{0, \phi_2(\hat{z}_{\boldsymbol{\tau}}^k) - \phi_2(z_{\boldsymbol{\tau}}^{k-1}), \phi_1(z_{\boldsymbol{\tau}}^k) - \phi_1(\hat{z}_{\boldsymbol{\tau}}^k)\}$$

for $k \in \mathbb{N}$ by suppressing the dependence on the choice of $\{\hat{z}_{\boldsymbol{\tau}}^k, z_{\boldsymbol{\tau}}^k\}_{k \in \mathbb{N}}$. Assume that, for any $\varepsilon, T > 0$, there is $\Delta_\varepsilon^T(z_0) < \infty$ such that

$$\sum_{k=1}^N \delta_{\boldsymbol{\tau}}^k(z_0) \leq \Delta_\varepsilon^T(z_0)$$

for any $\mathscr{P}_{\boldsymbol{\tau}}$ with $|\boldsymbol{\tau}| < \varepsilon$, $N \in \mathbb{N}$ with $t_{\boldsymbol{\tau}}^N \leq T$, and for any solution $\{\hat{z}_{\boldsymbol{\tau}}^k, z_{\boldsymbol{\tau}}^k\}_{k \in \mathbb{N}}$ to (5.1). This in particular guarantees that $\hat{z}_{\boldsymbol{\tau}}^k \in D(\phi)$ and $z_{\boldsymbol{\tau}}^k \in D(\phi)$.

Example 5.3 One of the simplest examples satisfying Assumption 5.2 is pairs of Lipschitz functions. If both ϕ_1 and ϕ_2 are L-Lipschitz, then

$$d^2(z_{\boldsymbol{\tau}}^{k-1}, \hat{z}_{\boldsymbol{\tau}}^k) \leq 2\tau_k\{\phi_1(z_{\boldsymbol{\tau}}^{k-1}) - \phi_1(\hat{z}_{\boldsymbol{\tau}}^k)\} \leq 2\tau_k L d(z_{\boldsymbol{\tau}}^{k-1}, \hat{z}_{\boldsymbol{\tau}}^k).$$

Hence $d(z_{\boldsymbol{\tau}}^{k-1}, \hat{z}_{\boldsymbol{\tau}}^k) \leq 2L\tau_k$ and similarly $d(\hat{z}_{\boldsymbol{\tau}}^k, z_{\boldsymbol{\tau}}^k) \leq 2L\tau_k$. Thus we find

$$\sum_{k=1}^{N} \delta_{\boldsymbol{\tau}}^k(z_0) \leq \sum_{k=1}^{N} L\{d(z_{\boldsymbol{\tau}}^{k-1}, \hat{z}_{\boldsymbol{\tau}}^k) + d(\hat{z}_{\boldsymbol{\tau}}^k, z_{\boldsymbol{\tau}}^k)\} \leq 4L^2 t_{\boldsymbol{\tau}}^N.$$

Notice that $\Delta_\varepsilon^T(z_0)$ is taken independently from ε and z_0 in this case. See [CM, Proposition 1.7] for other examples.

Our assumptions are comparable with those in [CM]. (For the sake of simplicity, we do not intend to minimize the assumptions in this section.) To state the main theorem of the section, we introduce the interpolated curve $\bar{\boldsymbol{z}}_{\boldsymbol{\tau}}$ similarly to §4.1:

$$\bar{\boldsymbol{z}}_{\boldsymbol{\tau}}(0) := z_0, \qquad \bar{\boldsymbol{z}}_{\boldsymbol{\tau}}(t) := z_{\boldsymbol{\tau}}^k \quad \text{for } t \in (t_{\boldsymbol{\tau}}^{k-1}, t_{\boldsymbol{\tau}}^k].$$

Theorem 5.4 (A Trotter–Kato product formula) *Let Assumptions* 5.1, 5.2 *be satisfied. Given $z_0 \in D(\phi)$, the curve $\bar{\boldsymbol{z}}_{\boldsymbol{\tau}}$ converges to the gradient curve $\xi := \mathcal{G}(\cdot, z_0)$ of ϕ (constructed in the previous section) as $|\boldsymbol{\tau}| \to 0$ uniformly on each bounded interval $[0, T]$.*

Similarly to the previous section, we will discuss under the global K-convexity of the squared distance function. Thus $\operatorname{diam} X < \pi$ is implicitly assumed in Case I, however, this costs no generality as we explained in §4.6.

5.2   Preliminary estimates

Fix $z_0 \in D(\phi)$, $\mathscr{P}_{\boldsymbol{\tau}}$ with $|\boldsymbol{\tau}| < \varepsilon$ and $\{\hat{z}_{\boldsymbol{\tau}}^k, z_{\boldsymbol{\tau}}^k\}_{k \in \mathbb{N}}$ solving (5.1).

Lemma 5.5   (i) *For each $N \in \mathbb{N}$ with $t_{\boldsymbol{\tau}}^N \leq T$, we have*

$$\max_{i=1,2}\{\phi_i(z_{\boldsymbol{\tau}}^N) - \phi_i(z_0), \phi_i(\hat{z}_{\boldsymbol{\tau}}^N) - \phi_i(z_0)\} \leq \sum_{k=1}^{N} \delta_{\boldsymbol{\tau}}^k(z_0) \leq \Delta_\varepsilon^T(z_0).$$

(ii) *For any $l \leq k$ with $t_{\boldsymbol{\tau}}^k \leq T$, we have*

$$\max\{d(z_{\boldsymbol{\tau}}^{l-1}, \hat{z}_{\boldsymbol{\tau}}^k), d(z_{\boldsymbol{\tau}}^{l-1}, z_{\boldsymbol{\tau}}^k), d(\hat{z}_{\boldsymbol{\tau}}^l, \hat{z}_{\boldsymbol{\tau}}^k), d(\hat{z}_{\boldsymbol{\tau}}^l, z_{\boldsymbol{\tau}}^k)\}$$
$$\leq \sqrt{2(t_{\boldsymbol{\tau}}^k - t_{\boldsymbol{\tau}}^{l-1})} \left\{ \sqrt{\phi_1(z_0) - \inf_X \phi_1 + 2\Delta_\varepsilon^T(z_0)} + \sqrt{\phi_2(z_0) - \inf_X \phi_2 + 2\Delta_\varepsilon^T(z_0)} \right\}.$$

Proof. (i) This is straightforward from the definition of $\delta_{\boldsymbol{\tau}}^k(z_0)$. We know $\phi_1(\hat{z}_{\boldsymbol{\tau}}^k) \leq \phi_1(z_{\boldsymbol{\tau}}^{k-1})$ and hence

$$\phi_1(z_{\boldsymbol{\tau}}^N) = \phi_1(z_0) + \sum_{k=1}^{N}\{\phi_1(z_{\boldsymbol{\tau}}^k) - \phi_1(\hat{z}_{\boldsymbol{\tau}}^k) + \phi_1(\hat{z}_{\boldsymbol{\tau}}^k) - \phi_1(z_{\boldsymbol{\tau}}^{k-1})\} \leq \phi_1(z_0) + \sum_{k=1}^{N} \delta_{\boldsymbol{\tau}}^k(z_0).$$

Similarly we find $\phi_1(\hat{z}_{\boldsymbol{\tau}}^N) \leq \phi_1(z_0) + \sum_{k=1}^{N-1} \delta_{\boldsymbol{\tau}}^k(z_0)$ and

$$\phi_2(z_{\boldsymbol{\tau}}^N) \leq \phi_2(\hat{z}_{\boldsymbol{\tau}}^N) \leq \phi_2(z_0) + \sum_{k=1}^{N} \delta_{\boldsymbol{\tau}}^k(z_0).$$

21

(ii) It follows from the Cauchy–Schwarz inequality that

$$\max\{d(z_{\boldsymbol{\tau}}^{l-1}, \hat{z}_{\boldsymbol{\tau}}^k), d(z_{\boldsymbol{\tau}}^{l-1}, z_{\boldsymbol{\tau}}^k), d(\hat{z}_{\boldsymbol{\tau}}^l, \hat{z}_{\boldsymbol{\tau}}^k), d(\hat{z}_{\boldsymbol{\tau}}^l, z_{\boldsymbol{\tau}}^k)\} \leq \sum_{m=l}^{k}\{d(z_{\boldsymbol{\tau}}^{m-1}, \hat{z}_{\boldsymbol{\tau}}^m) + d(\hat{z}_{\boldsymbol{\tau}}^m, z_{\boldsymbol{\tau}}^m)\}$$

$$\leq \sqrt{\sum_{m=l}^{k} 2\tau_m} \left\{ \sqrt{\sum_{m=l}^{k} \frac{d^2(z_{\boldsymbol{\tau}}^{m-1}, \hat{z}_{\boldsymbol{\tau}}^m)}{2\tau_m}} + \sqrt{\sum_{m=l}^{k} \frac{d^2(\hat{z}_{\boldsymbol{\tau}}^m, z_{\boldsymbol{\tau}}^m)}{2\tau_m}} \right\}$$

$$\leq \sqrt{2(t_{\boldsymbol{\tau}}^k - t_{\boldsymbol{\tau}}^{l-1})} \left\{ \sqrt{\sum_{m=l}^{k}\{\phi_1(z_{\boldsymbol{\tau}}^{m-1}) - \phi_1(\hat{z}_{\boldsymbol{\tau}}^m)\}} + \sqrt{\sum_{m=l}^{k}\{\phi_2(\hat{z}_{\boldsymbol{\tau}}^m) - \phi_2(z_{\boldsymbol{\tau}}^m)\}} \right\}.$$

Note that, by (i),

$$\sum_{m=l}^{k}\{\phi_1(z_{\boldsymbol{\tau}}^{m-1}) - \phi_1(\hat{z}_{\boldsymbol{\tau}}^m)\} \leq \sum_{m=l}^{k}\{\phi_1(z_{\boldsymbol{\tau}}^{m-1}) - \phi_1(z_{\boldsymbol{\tau}}^m) + \delta_{\boldsymbol{\tau}}^m(z_0)\}$$

$$= \phi_1(z_{\boldsymbol{\tau}}^{l-1}) - \phi_1(z_{\boldsymbol{\tau}}^k) + \sum_{m=l}^{k}\delta_{\boldsymbol{\tau}}^m(z_0)$$

$$\leq \phi_1(z_0) - \inf_X \phi_1 + 2\Delta_\varepsilon^T(z_0).$$

Similarly we obtain $\sum_{m=l}^{k}\{\phi_2(\hat{z}_{\boldsymbol{\tau}}^m) - \phi_2(z_{\boldsymbol{\tau}}^m)\} \leq \phi_2(z_0) - \inf_X \phi_2 + 2\Delta_\varepsilon^T(z_0)$. This completes the proof. $\square$

Thanks to (ii) above, as $|\boldsymbol{\tau}| \to 0$, the Ascoli–Arzelà theorem applies and we obtain the uniform convergence of a subsequence of $\bar{z}_{\boldsymbol{\tau}} : [0, T] \longrightarrow X$ to a Hölder continuous curve $\zeta : [0, T] \longrightarrow X$ with $\zeta(0) = z_0$. Our goal is to show that ζ coincides with the gradient curve ξ of ϕ. Then $\bar{z}_{\boldsymbol{\tau}}$ uniformly converges to ξ as $|\boldsymbol{\tau}| \to 0$ without passing to subsequences. Observe that the uniformity can be seen by contradiction; the existence of $\rho > 0$ such that $\sup_{t \in [0,T]} d(\bar{z}_{\boldsymbol{\tau}_i}(t), \xi(t)) \geq \rho$ for all i with $\lim_{i \to 0} |\boldsymbol{\tau}_i| \to 0$ contradicts the uniform convergence of a subsequence of $\{\bar{z}_{\boldsymbol{\tau}_i}\}_{i \in \mathbb{N}}$.

The following key estimate can be thought of as a discrete version of the evolution variational inequality (compare this with [CM, Lemma 2.1]).

Lemma 5.6 *Assume $\lambda_i|\boldsymbol{\tau}| > -1$ for $i = 1, 2$. For any $w \in D(\phi)$ and $k \in \mathbb{N}$ with $t_{\boldsymbol{\tau}}^k \leq T$, we have*

$$e^{(\lambda_1^{\boldsymbol{\tau}} + \lambda_2^{\boldsymbol{\tau}})t_{\boldsymbol{\tau}}^k} d^2(z_{\boldsymbol{\tau}}^k, w) \leq e^{(\lambda_1^{\boldsymbol{\tau}} + \lambda_2^{\boldsymbol{\tau}})t_{\boldsymbol{\tau}}^{k-1}} d^2(z_{\boldsymbol{\tau}}^{k-1}, w) + 2e^{\lambda_2^{\boldsymbol{\tau}} t_{\boldsymbol{\tau}}^{k-1} + \lambda_1^{\boldsymbol{\tau}} t_{\boldsymbol{\tau}}^k} \tau_k\{\phi(w) - \phi(z_{\boldsymbol{\tau}}^k) + \delta_{\boldsymbol{\tau}}^k(z_0)\}$$

$$- K' e^{\lambda_2^{\boldsymbol{\tau}} t_{\boldsymbol{\tau}}^{k-1} + \lambda_1^{\boldsymbol{\tau}} t_{\boldsymbol{\tau}}^k} \tau_k\{\phi(z_{\boldsymbol{\tau}}^{k-1}) - \phi(z_{\boldsymbol{\tau}}^k) + 2\delta_{\boldsymbol{\tau}}^k(z_0)\} + \tau_k \cdot O_{z_0}(\sqrt{\tau_k}),$$

where $K' := \min\{0, K\}$ and

$$\lambda_i^{\boldsymbol{\tau}} := \frac{\log(1 + \lambda_i|\boldsymbol{\tau}|)}{|\boldsymbol{\tau}|}, \quad i = 1, 2.$$

Proof. The proof is based on calculations similar to Lemma 4.6. Applying Lemma 3.1 to the steps $z_{\boldsymbol{\tau}}^{k-1} \to \hat{z}_{\boldsymbol{\tau}}^k$ and $\hat{z}_{\boldsymbol{\tau}}^k \to z_{\boldsymbol{\tau}}^k$, we have

$$(1 + \lambda_1\tau_k)d^2(\hat{z}_{\boldsymbol{\tau}}^k, w) - d^2(z_{\boldsymbol{\tau}}^{k-1}, w) \leq 2\tau_k\{\phi_1(w) - \phi_1(\hat{z}_{\boldsymbol{\tau}}^k)\} - K'\tau_k\{\phi_1(z_{\boldsymbol{\tau}}^{k-1}) - \phi_1(\hat{z}_{\boldsymbol{\tau}}^k)\},$$

$$(1 + \lambda_2\tau_k)d^2(z_{\boldsymbol{\tau}}^k, w) - d^2(\hat{z}_{\boldsymbol{\tau}}^k, w) \leq 2\tau_k\{\phi_2(w) - \phi_2(z_{\boldsymbol{\tau}}^k)\} - K'\tau_k\{\phi_2(\hat{z}_{\boldsymbol{\tau}}^k) - \phi_2(z_{\boldsymbol{\tau}}^k)\}.$$

Thus we find

$$(1 + \lambda_2\tau_k)d^2(z_{\boldsymbol{\tau}}^k, w) \leq d^2(z_{\boldsymbol{\tau}}^{k-1}, w) - \lambda_1\tau_k d^2(\hat{z}_{\boldsymbol{\tau}}^k, w) + 2\tau_k\{\phi(w) - \phi_1(\hat{z}_{\boldsymbol{\tau}}^k) - \phi_2(z_{\boldsymbol{\tau}}^k)\}$$
$$- K'\tau_k\{\phi_1(z_{\boldsymbol{\tau}}^{k-1}) - \phi_1(\hat{z}_{\boldsymbol{\tau}}^k) + \phi_2(\hat{z}_{\boldsymbol{\tau}}^k) - \phi_2(z_{\boldsymbol{\tau}}^k)\}.$$

Note that, by Lemma 5.5(i),

$$|d^2(\hat{z}_{\boldsymbol{\tau}}^k, w) - d^2(z_{\boldsymbol{\tau}}^{k-1}, w)| \leq \{d(\hat{z}_{\boldsymbol{\tau}}^k, w) + d(z_{\boldsymbol{\tau}}^{k-1}, w)\}d(\hat{z}_{\boldsymbol{\tau}}^k, z_{\boldsymbol{\tau}}^{k-1})$$
$$\leq 2D\sqrt{2\tau_k\{\phi_1(z_{\boldsymbol{\tau}}^{k-1}) - \phi_1(\hat{z}_{\boldsymbol{\tau}}^k)\}}$$
$$\leq 2D\sqrt{2\tau_k}\sqrt{\phi_1(z_0) + \Delta_\varepsilon^T(z_0) - \inf_X \phi_1}$$
$$= O_{z_0}(\sqrt{\tau_k}).$$

Moreover,

$$\phi(w) - \phi_1(\hat{z}_{\boldsymbol{\tau}}^k) - \phi_2(z_{\boldsymbol{\tau}}^k) \leq \phi(w) - \phi(z_{\boldsymbol{\tau}}^k) + \delta_{\boldsymbol{\tau}}^k(z_0)$$

and similarly

$$\phi_1(z_{\boldsymbol{\tau}}^{k-1}) - \phi_1(\hat{z}_{\boldsymbol{\tau}}^k) + \phi_2(\hat{z}_{\boldsymbol{\tau}}^k) - \phi_2(z_{\boldsymbol{\tau}}^k) \leq \phi(z_{\boldsymbol{\tau}}^{k-1}) - \phi(z_{\boldsymbol{\tau}}^k) + 2\delta_{\boldsymbol{\tau}}^k(z_0).$$

Combining these yields

$$(1 + \lambda_2\tau_k)d^2(z_{\boldsymbol{\tau}}^k, w) \leq \frac{1}{1 + \lambda_1\tau_k}d^2(z_{\boldsymbol{\tau}}^{k-1}, w) + 2\tau_k\{\phi(w) - \phi(z_{\boldsymbol{\tau}}^k) + \delta_{\boldsymbol{\tau}}^k(z_0)\}$$
$$- K'\tau_k\{\phi(z_{\boldsymbol{\tau}}^{k-1}) - \phi(z_{\boldsymbol{\tau}}^k) + 2\delta_{\boldsymbol{\tau}}^k(z_0)\} + \tau_k \cdot O_{z_0}(\sqrt{\tau_k}).$$

Multiply both sides by $e^{\lambda_2^{\boldsymbol{\tau}} t_{\boldsymbol{\tau}}^{k-1} + \lambda_1^{\boldsymbol{\tau}} t_{\boldsymbol{\tau}}^k} = e^{(\lambda_1^{\boldsymbol{\tau}} + \lambda_2^{\boldsymbol{\tau}})t_{\boldsymbol{\tau}}^{k-1} + \lambda_1^{\boldsymbol{\tau}}\tau_k} = e^{(\lambda_1^{\boldsymbol{\tau}} + \lambda_2^{\boldsymbol{\tau}})t_{\boldsymbol{\tau}}^k - \lambda_2^{\boldsymbol{\tau}}\tau_k}$. Then, recalling (4.4), we obtain the desired estimate. $\qquad\square$

5.3   Proof of Theorem 5.4

By Lemma 5.6,

$$\frac{d^2(z_{\boldsymbol{\tau}}^k, w) - d^2(z_{\boldsymbol{\tau}}^{k-1}, w)}{2\tau_k}$$
$$= \frac{e^{(\lambda_1^{\boldsymbol{\tau}} + \lambda_2^{\boldsymbol{\tau}})t_{\boldsymbol{\tau}}^k}d^2(z_{\boldsymbol{\tau}}^k, w) - e^{(\lambda_1^{\boldsymbol{\tau}} + \lambda_2^{\boldsymbol{\tau}})t_{\boldsymbol{\tau}}^{k-1}}d^2(z_{\boldsymbol{\tau}}^{k-1}, w)}{2e^{(\lambda_1^{\boldsymbol{\tau}} + \lambda_2^{\boldsymbol{\tau}})t_{\boldsymbol{\tau}}^k}\tau_k} + \frac{e^{-(\lambda_1^{\boldsymbol{\tau}} + \lambda_2^{\boldsymbol{\tau}})\tau_k} - 1}{2\tau_k}d^2(z_{\boldsymbol{\tau}}^{k-1}, w)$$
$$\leq e^{-\lambda_2^{\boldsymbol{\tau}}\tau_k}\{\phi(w) - \phi(z_{\boldsymbol{\tau}}^k) + \delta_{\boldsymbol{\tau}}^k(z_0)\} - \frac{K'}{2}e^{-\lambda_2^{\boldsymbol{\tau}}\tau_k}\{\phi(z_{\boldsymbol{\tau}}^{k-1}) - \phi(z_{\boldsymbol{\tau}}^k) + 2\delta_{\boldsymbol{\tau}}^k(z_0)\}$$
$$+ \frac{e^{-(\lambda_1^{\boldsymbol{\tau}} + \lambda_2^{\boldsymbol{\tau}})\tau_k} - 1}{2\tau_k}d^2(z_{\boldsymbol{\tau}}^{k-1}, w) + O(\sqrt{|\boldsymbol{\tau}|})$$
$$= \phi(w) - \phi(z_{\boldsymbol{\tau}}^k) - \frac{K'}{2}\{\phi(z_{\boldsymbol{\tau}}^{k-1}) - \phi(z_{\boldsymbol{\tau}}^k)\} - \frac{\lambda_1 + \lambda_2}{2}d^2(z_{\boldsymbol{\tau}}^{k-1}, w) + (1 - K')\delta_{\boldsymbol{\tau}}^k(z_0)$$
$$+ O(\sqrt{|\boldsymbol{\tau}|}).$$

We used the bound of $\phi(z_{\boldsymbol{\tau}}^k)$ (Lemma 5.5(i)) to estimate the error terms. Denote by $\{x_{\boldsymbol{\tau}}^k\}_{k\geq 0}$ a discrete solution of the variational scheme (2.1) for ϕ with $x_{\boldsymbol{\tau}}^0 = z_0$. We recall from the proof of Theorem 4.1 that, putting $\lambda := \lambda_1 + \lambda_2$,

$$\frac{d^2(x_{\boldsymbol{\tau}}^k, y) - d^2(x_{\boldsymbol{\tau}}^{k-1}, y)}{2\tau_k} \leq \phi(y) - \phi(x_{\boldsymbol{\tau}}^k) - \frac{K'}{2}\{\phi(x_{\boldsymbol{\tau}}^{k-1}) - \phi(x_{\boldsymbol{\tau}}^k)\} - \frac{\lambda}{2}d^2(x_{\boldsymbol{\tau}}^k, y).$$

Applying these inequalities with $w = x_{\boldsymbol{\tau}}^{k-1}$ and $y = z_{\boldsymbol{\tau}}^k$ to

$$d^2(x_{\boldsymbol{\tau}}^N, z_{\boldsymbol{\tau}}^N) = \sum_{k=1}^N \left\{ d^2(x_{\boldsymbol{\tau}}^k, z_{\boldsymbol{\tau}}^k) - d^2(x_{\boldsymbol{\tau}}^{k-1}, z_{\boldsymbol{\tau}}^k) + d^2(x_{\boldsymbol{\tau}}^{k-1}, z_{\boldsymbol{\tau}}^k) - d^2(x_{\boldsymbol{\tau}}^{k-1}, z_{\boldsymbol{\tau}}^{k-1}) \right\},$$

we obtain for N with $t_{\boldsymbol{\tau}}^N \leq T$

$$d^2(x_{\boldsymbol{\tau}}^N, z_{\boldsymbol{\tau}}^N) \leq \sum_{k=1}^N 2\tau_k \left\{ \phi(z_{\boldsymbol{\tau}}^k) - \phi(x_{\boldsymbol{\tau}}^k) - \frac{K'}{2}\{\phi(x_{\boldsymbol{\tau}}^{k-1}) - \phi(x_{\boldsymbol{\tau}}^k)\} - \frac{\lambda}{2}d^2(x_{\boldsymbol{\tau}}^k, z_{\boldsymbol{\tau}}^k) \right\}$$

$$+ \sum_{k=1}^N 2\tau_k \left\{ \phi(x_{\boldsymbol{\tau}}^{k-1}) - \phi(z_{\boldsymbol{\tau}}^k) - \frac{K'}{2}\{\phi(z_{\boldsymbol{\tau}}^{k-1}) - \phi(z_{\boldsymbol{\tau}}^k)\} - \frac{\lambda}{2}d^2(x_{\boldsymbol{\tau}}^{k-1}, z_{\boldsymbol{\tau}}^{k-1}) \right\}$$

$$+ \sum_{k=1}^N 2\tau_k(1 - K')\delta_{\boldsymbol{\tau}}^k(z_0) + t_{\boldsymbol{\tau}}^N \cdot O(\sqrt{|\boldsymbol{\tau}|})$$

$$\leq (2 - K')\sum_{k=1}^N \tau_k\{\phi(x_{\boldsymbol{\tau}}^{k-1}) - \phi(x_{\boldsymbol{\tau}}^k)\} - K'\sum_{k=1}^N \tau_k\{\phi(z_{\boldsymbol{\tau}}^{k-1}) - \phi(z_{\boldsymbol{\tau}}^k)\}$$

$$- \lambda\sum_{k=1}^N \tau_k\{d^2(x_{\boldsymbol{\tau}}^{k-1}, z_{\boldsymbol{\tau}}^{k-1}) + d^2(x_{\boldsymbol{\tau}}^k, z_{\boldsymbol{\tau}}^k)\} + 2(1 - K')|\boldsymbol{\tau}|\Delta_{\varepsilon}^T(z_0)$$

$$+ t_{\boldsymbol{\tau}}^N \cdot O(\sqrt{|\boldsymbol{\tau}|}).$$

Notice that

$$\sum_{k=1}^N \tau_k\{\phi(x_{\boldsymbol{\tau}}^{k-1}) - \phi(x_{\boldsymbol{\tau}}^k)\} \leq \sum_{k=1}^N |\boldsymbol{\tau}|\{\phi(x_{\boldsymbol{\tau}}^{k-1}) - \phi(x_{\boldsymbol{\tau}}^k)\} \leq \left\{ \phi(z_0) - \inf_X \phi \right\}|\boldsymbol{\tau}|.$$

Moreover, since $\phi(z_{\boldsymbol{\tau}}^{k-1}) - \phi(z_{\boldsymbol{\tau}}^k) \geq -2\delta_{\boldsymbol{\tau}}^k(z_0)$,

$$\sum_{k=1}^N \tau_k\{\phi(z_{\boldsymbol{\tau}}^{k-1}) - \phi(z_{\boldsymbol{\tau}}^k)\} \leq \sum_{k=1}^N |\boldsymbol{\tau}|\{\phi(z_{\boldsymbol{\tau}}^{k-1}) - \phi(z_{\boldsymbol{\tau}}^k) + 2\delta_{\boldsymbol{\tau}}^k(z_0)\}$$

$$\leq \left\{ \phi(z_0) - \inf_X \phi + 2\Delta_{\varepsilon}^T(z_0) \right\}|\boldsymbol{\tau}|.$$

Therefore, letting $i \to \infty$ in the sequence $\mathscr{P}_{\boldsymbol{\tau}_i}$ for which $\bar{z}_{\boldsymbol{\tau}_i}$ converges to ζ, we find

$$d^2\big(\xi(T), \zeta(T)\big) \leq -2\lambda\int_0^T d^2(\xi, \zeta)\,dt.$$

This implies that the nonnegative function $f(T) := \int_0^T d^2(\xi, \zeta)\,dt$ satisfies $f' \leq -2\lambda f$ and hence $(e^{2\lambda T}f(T))' \leq 0$. Thus $f \equiv 0$ and we complete the proof of $\zeta = \xi$ and Theorem 5.4 (recall the paragraph following Lemma 5.5). $\qquad\square$

References

[AG]   L. Ambrosio and N. Gigli, A user's guide to optimal transport, Modeling and optimisation of flows on networks, 1–155, Lecture Notes in Math., **2062**, Springer, 2013.

[AGS1]   L. Ambrosio, N. Gigli and G. Savaré, Gradient flows in metric spaces and in the space of probability measures, Birkhäuser Verlag, Basel, 2005.

[AGS2]   L. Ambrosio, N. Gigli and G. Savaré, Calculus and heat flow in metric measure spaces and applications to spaces with Ricci bounds from below, Invent. Math. **195** (2014), 289–391.

[AGS3]   L. Ambrosio, N. Gigli and G. Savaré, Metric measure spaces with Riemannian Ricci curvature bounded from below, Duke Math. J. **163** (2014), 1405–1490.

[ASZ]   L Ambrosio, G. Savaré and L. Zambotti, Existence and stability for Fokker-Planck equations with log-concave reference measure, Probab. Theory Related Fields **145** (2009), 517–564.

[Ba1]   M. Bačák, A new proof of the Lie–Trotter–Kato formula in Hadamard spaces, to appear in Commun. Contemp. Math.

[Ba2]   M. Bačák, Convex analysis and optimization in Hadamard spaces, Walter de Gruyter & Co., Berlin, 2014.

[BCL]   K. Ball, E. A. Carlen and E. H. Lieb, Sharp uniform convexity and smoothness inequalities for trace norms, Invent. Math. **115** (1994), 463–482.

[Br]   H. Brézis, Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert (French), North-Holland Publishing Co., Amsterdam-London; American Elsevier Publishing Co., Inc., New York, 1973.

[BBI]   D. Burago, Yu. Burago and S. Ivanov, A course in metric geometry, American Mathematical Society, Providence, RI, 2001.

[CM]   P. Clément and J. Maas, A Trotter product formula for gradient flows in metric spaces, J. Evol. Equ. **11** (2011), 405–427. Erratum in J. Evol. Equ. **13** (2013), 251–252.

[CL]   M. G. Crandall and T. M. Liggett, Generation of semi-groups of nonlinear transformations on general Banach spaces, Amer. J. Math. **93** (1971), 265–298.

[DG]   E. De Giorgi, New problems on minimizing movements, Boundary value problems for partial differential equations and applications, 81–98, RMA Res. Notes Appl. Math., **29**, Masson, Paris, 1993. (Also in: E. De Giorgi, Selected papers, 699–713, Edited by Luigi Ambrosio et al, Springer-Verlag, Berlin, 2006.)

[EKS]   M Erbar, K. Kuwada and K.-T. Sturm, On the equivalence of the entropic curvature-dimension condition and Bochner's inequality on metric measure spaces, Preprint (2013). Available at arXiv:1303.4382

[Gi1]   N. Gigli, On the heat flow on metric measure spaces: existence, uniqueness and stability, Calc. Var. Partial Differential Equations **39** (2010), 101–120.

[Gi2]   N. Gigli, The splitting theorem in non-smooth context, Preprint (2013). Available at arXiv:1302.5555

[Gi3]   N. Gigli, An overview of the proof of the splitting theorem in spaces with non-negative Ricci curvature, Anal. Geom. Metr. Spaces **2** (2014), 169–213.

[GKO]   N. Gigli, K. Kuwada and S. Ohta, Heat flow on Alexandrov spaces, Comm. Pure Appl. Math. **66** (2013), 307–331.

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

[Jo]   J. Jost, Convex functionals and generalized harmonic maps into spaces of nonpositive curvature, Comment. Math. Helv. **70** (1995), 659–673.

[KM]   T. Kato and K. Masuda, Trotter's product formula for nonlinear semigroups generated by the subdifferentials of convex functionals, J. Math. Soc. Japan **30** (1978), 169–178.

[Ly]   A. Lytchak, Open map theorem for metric spaces, St. Petersburg Math. J. **17** (2006), 477–491.

[Ma]   U. F. Mayer, Gradient flows on nonpositively curved metric spaces and harmonic maps, Comm. Anal. Geom. **6** (1998), 199–253.

[Oh1]   S. Ohta, Convexities of metric spaces, Geom. Dedicata **125** (2007), 225–250.

[Oh2]   S. Ohta, Gradient flows on Wasserstein spaces over compact Alexandrov spaces, Amer. J. Math. **131** (2009), 475–516.

[Oh3]   S. Ohta, Uniform convexity and smoothness, and their applications in Finsler geometry, Math. Ann. **343** (2009), 669–699.

[OP]   S. Ohta and M. Pálfia, Discrete-time gradient flows and law of large numbers in Alexandrov spaces, Preprint (2014). Available at arXiv:1402.1629

[OS]   S. Ohta and K.-T. Sturm, Non-contraction of heat flow on Minkowski spaces, Arch. Ration. Mech. Anal. **204** (2012), 917–944.

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

[PP]   G. Perel'man and A. Petrunin, Quasigeodesics and gradient curves in Alexandrov spaces, Unpublished preprint. Available at http://www.math.psu.edu/petrunin/

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

[Sto]   I. Stojkovic, Approximation for convex functionals on non-positively curved spaces and the Trotter–Kato product formula, Adv. Calc. Var. **5** (2012), 77–126.

[Stu]   K.-T. Sturm, Gradient flows for semiconvex functions on metric measure spaces – existence, uniqueness and Lipschitz continuity, Preprint (2014). Available at arXiv:1410.3966

[Vi1]   C. Villani, Topics in optimal transportation, American Mathematical Society, Providence, RI, 2003.

[Vi2]   C. Villani, Optimal transport, old and new, Springer-Verlag, Berlin, 2009.