Calculus of Variations

Convergence of semi-convex functions in CAT(1)-spaces

Hedvig Gál[1] · Miklós Pálfia[1,2]

Abstract

We generalize the results of Kuwae–Shiya and Bačák on Mosco convergence established for CAT(0)-spaces to the CAT(1)-setting, so that Mosco convergence implies convergence of resolvents which in turn imply convergence of gradient flows for lower-semicontinuous semi-convex functions. Our techniques utilize weak convergence in CAT(1)-spaces and also cover asymptotic relations of sequences of such spaces introduced by Kuwae-Shiya, including Gromov–Hausdorff limits.

1 Introduction

The theory of convex functions in CAT(1)-spaces has received a considerable attention in the last 20 years or so. The theory first started in the papers of Jost [13, 14] and Mayer [20] establishing, among others, Crandall–Liggett-type [10] generation theorems for nonlinear semigroups corresponding to semi-convex functions in the CAT(0)-setting. The theory has been further expanded in the monographs [2, 5] where the theory of gradient flows has been studied in great detail. This followed by further investigations into convex optimization theory in the CAT(0)-setting [5]. Somewhat more recently Ohta and the second author has developed a theory of discrete [24] and continuous-time [25] gradient flows for lower semi-continuous semi-convex functions in CAT(1)-spaces. Note that it is enough to study the CAT(1)-setting, since for $\kappa > 0$, the metric $d(\cdot, \cdot)$ of a CAT(κ)-space can be rescaled to $\frac{1}{\sqrt{k}}d(\cdot, \cdot)$ so that it becomes a CAT(1)-space with this new metric. Using the *Moreau-Yosida* resolvent $J_\tau^\phi(x)$ for a lower-semicontinuous λ-convex function $\phi : X \to (-\infty, \infty]$ and $\tau > 0$, one can construct

---

Communicated by Juergen Jost.

✉ Miklós Pálfia
miklos.palfia@uni-corvinus.hu

Hedvig Gál
hedvig.gal@stud.uni-corvinus.hu

[1] Department of Mathematics, Corvinus University of Budapest, Fővám tér 8, Budapest 1093, Hungary

[2] Bolyai Institute, Interdisciplinary Excellence Centre, University of Szeged, Szeged 6720, Hungary

Springer

the gradient flow as

$$S(t)x_0 = \lim_{n \to \infty} \left(J^{\phi}_{t/n} \right)^n (x_0) \tag{1}$$

for a starting point $S(0)x_0 := x_0$ and $t > 0$. This curve $\xi(t) = S(t)x_0$ is characterized as the unique solution [22] to the *Evolution Variational Inequality*

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

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

Variational convergence for sequences of functions in CAT(0)-spaces has been considered in the meantime, see Jost [13–15] and then Kuwae–Shioya [17] which introduced the concept of *Mosco convergence* for sequences of convex functions in a CAT(0)-space extending the pioneering works of Mosco [21] and Dal Maso [11]. See also [3] for a very detailed approach to the Banach and Hilbert settings. In particular following [14] it has been established in [17], that Mosco convergence of lower-semicontinuous convex functions implies the convergence of the corresponding Moreau-Yosida resolvents and in turn convergence of their gradient flows provided that all functions are bounded from below. This latter requirement has been relaxed in [6] and established the convergence through (1) also in the case of limits of sequences of metric spaces. They considered the general notion of an *asymptotic relation* of metric spaces which includes, among others, Gromov-Hausdorff limits. An essential feature of these approaches is the concept of *weak convergence* in a CAT(0)-space, where closed and bounded sets turn out to be weakly sequentially compact [6]. The concept of weak convergence and compactness carries over to the CAT(1)-setting as has been investigated by a number of authors along with the study of (semi-)convex functions [16, 26].

The purpose of this paper is to generalize the convergence of resolvents and gradient flows of Mosco convergent sequences of lower-semicontinuous semi-convex functions to the CAT(1)-setting from the CAT(0) with convex functions case [5, 6, 17]. In order to utilize weak convergence, thus to make sense of Mosco convergence, we need to assume some diameter bounds on our spaces in the most general cases of arbitrary lower-semicontinuous semi-convex functions. Under somewhat stronger assumptions of uniform lower boundedness or Lipschitz continuity of our function sequence we are able to prove convergence without diameter restrictions. Also we extend to the case of asymptotic relations, by generalizing the required weak convergence and compactness results to the CAT(1)-setting which may be of independent interest. In [6] the author utilizes local and global slope estimates for lower semicontinuous functions in the CAT(0)-setting established in [2]. The second author with Ohta in [25] established more direct estimates of error terms of minimizing movements related to gradient flows for semi-convex functions compared to those in [2]. We utilize these estimates in this paper in order to control error terms of minimizing movements converging to gradient curves. This approach is more direct than the one available for CAT(0)-spaces in [6], thus provides a simpler approach. Also we rely on Lipschitz estimates of [18] established for Moreau-Yosida resolvents locally in CAT(1)-spaces. The local nature of our analysis is one of the essential differences between the CAT(0)- and CAT(1)-setting.

Before moving onto the next section we fix some notations used in this paper. Given a real number $\kappa > 0$, then M^2_κ is obtained from the sphere S^2 by rescaling the distance function by the constant $1/\sqrt{k}$. For the definition of CAT(κ)-spaces we refer to the textbook [9]. A subset A of a CAT(κ)-space (X, d) is said to be λ-convex for $\lambda > 0$ if for any two points $x, y \in A$ with $d(x, y) < \lambda$ can be joined by a unique minimal geodesic denoted by $[x, y]$ contained in A. In this paper all CAT(κ)-spaces are assumed to be complete.

2 Weak convergence in CAT(1)-spaces

We begin with recalling the concept of *weak convergence* in a CAT(1) space [13], see also Definition 27 in [26]. In the definition we make use of $P_Z(x)$ which denotes the unique nearest point projection of the point $x \in X$ onto a closed π-convex set $Z \subseteq X$ with $d(Z, x) := \inf_{z \in Z} d(z, x) < \pi/2$. Under these conditions $P_Z(x)$ is a retraction with the obtuse-angle property, see for example Exercise 2.6(1) in [9].

Definition 2.1 (*Weak convergence*) Let (X, d) be a CAT(1) space. A net $x_i \in X$ included in $B_x(\pi/2)$ with $x \in X$ *weakly converges* to the point x, that is $x_i \overset{w}{\to} x$, if for any π-convex geodesic segment $\gamma \subseteq X$ with $\gamma(0) = x$, $P_\gamma(x_i)$ strongly converges to x.

The following result characterizes weak convergence, this lemma can be found in [6] in the case of CAT(0)-spaces. In the CAT(1)-case the price to be paid in the formulation is the required boundedness implied by the required convexity of geodesic lines.

Lemma 2.1 *Let (X, d) be a CAT(1) space. Let $x_n, x \in X$ such that $d(x_n, x) < \pi/2$ for all $n \in \mathbf{N}$. Then $x_n \to x$ if and only if $x_n \overset{w}{\to} x$ and $d(x_n, y) \to d(x, y)$ for some $y \in X$ such that $d(x, y) < \pi/2$.*

Proof The " $\Rightarrow$ " is trivial, so we only need to prove the " $\Leftarrow$ " implication. So let $x_n \overset{w}{\to} x$ and $d(x_n, y) \to d(x, y)$ for some $y \in X$ such that $d(x, y), d(x_n, x) < \pi/2$. We can assume that $d(y, x) > 0$, otherwise $d(x_n, x) \to d(x, x) = 0$ implying already that $x_n \to x$. Then, by Exercise 2.6(1) [9] we have that the Alexandrov-angle $\angle_{P_{[x,y]}(x_n)}(x_n, y) \geq \pi/2$ which we call the *obtuse-angle property*, thus by the spherical law of cosines we have

$$
\begin{aligned}
\cos d(x_n, y) &\leq \cos d(x_n, P_{[x,y]}(x_n)) \cos d(y, P_{[x,y]}(x_n)) \\
&\quad + \sin d(x_n, P_{[x,y]}(x_n)) \sin d(y, P_{[x,y]}(x_n)) \cos \angle_{P_{[x,y]}(x_n)}(x_n, y) \\
&\leq \cos d(x_n, P_{[x,y]}(x_n)) \cos d(y, P_{[x,y]}(x_n)).
\end{aligned}
$$

Now, weak convergence implies convergence of projections $P_{[x,y]}(x_n) \to x$. Moreover, since $P_{[x,y]}(x_n) \to x$, we may assume $P_{[x,y]}(x_n) \neq y$. Thus, the assumption $d(x_n, y) \to d(x, y)$ with the above estimate implies

$$
\cos d(x_n, P_{[x,y]}(x_n)) \to 1, \tag{2}
$$

thus $x_n \to x$ as well.   □

We also have the following fact which ensures that small enough bounded sets are weakly sequentially compact, thus providing a Banach-Alaoglu property.

Lemma 2.2 (cf. Lemma 28 in [26]) *Let (X, d) be a complete CAT(1)-space. Any sequence $(p_n)_{n \in \mathbb{N}}$ of points in X included in a π-convex set has a subsequence which converges weakly to a point in X.*

We will make use of the following result later.

Lemma 2.3 *Let (X, d) be a CAT(1) space with $\mathrm{diam}(X) < \pi$. Then if $(C_i)_{i \in I}$ is a non-increasing family of closed convex sets in X for an index set I, we have $\cap_{i \in I} C_i \neq \emptyset$.*

Proof See for example [16, Theorem 2.5.]   □

3 Ekeland principle and Mosco continuity

Let (X, d) be a CAT(1) space in this section. We denote the extended real numbers by $\overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, \infty\}$.

Definition 3.1 (*semi-continuity*) A function $\phi : X \to \overline{\mathbb{R}}$ with its effective domain $D_\phi := \{y \in X | \phi(y) < \infty\}$ is *lower semi-continuous* (lsc) at $x_0 \in D_\phi \subseteq X$, if

$$\liminf_{x \to x_0} \phi(x) \geq \phi(x_0).$$

Similarly we define weak lsc at $x_0 \in D_\phi \subseteq X$ if

$$\liminf_{x \xrightarrow{w} x_0} \phi(x) \geq \phi(x_0).$$

A function $\phi : X \to \overline{\mathbb{R}}$ is called lower semi-continuous on D_ϕ, if ϕ is lsc at any point in D_ϕ. It is well known that a function $\phi : X \to \overline{\mathbb{R}}$ is lsc if and only if all sublevel sets are closed, if and only if its epigraph is closed.

Definition 3.2 (*semi-convexity*) A function $\phi : X \to \overline{\mathbb{R}}$ is *semi-convex*, more precisely λ-*convex* for a $\lambda \in \mathbb{R}$, if

$$\phi\big(\gamma(t)\big) \leq (1 - t)\phi\big(\gamma(0)\big) + t\phi\big(\gamma(1)\big) - \frac{\lambda}{2}(1 - t)t d^2\big(\gamma(0), \gamma(1)\big) \tag{3}$$

along geodesics $\gamma : [0, 1] \to X$.

More generally ϕ is *quasi-convex* if its sublevel sets are convex.

Lemma 3.1 (Ohta's lemma cf. Lemma 5 in [26]) *Any geodesic* $\gamma : [0, 1] \to X$ *with* $\gamma(0), \gamma(1) \in \overline{B}_z((\pi - \epsilon)/2)$ *for an* $\epsilon > 0$ *satisfies*

$$d^2(\gamma(t), z) \leq (1 - t)d^2(\gamma(0), z) + t d^2(\gamma(1), z) - \frac{\kappa}{2}t(1 - t)d^2(\gamma(0), \gamma(1))$$

for any $t \in [0, 1]$ *and* $\kappa = (\pi - \epsilon)\tan(\epsilon/2)$.

Lemma 3.2 *Let* $\mathrm{diam}(X) < \pi$. *Let* $f : X \to (-\infty, \infty]$ *be a convex lsc function. Then* f *is bounded below.*

Proof Assume on the contrary that $\inf_X f = -\infty$. Then the sub-level sets $X_k := \{x \in X : f(x) \leq -k\}$ for $k \in \mathbb{N}$ are nonempty and are bounded, closed and convex. The sequence $\{X_k\}_{k \in \mathbb{N}}$ is non-increasing, therefore by Lemma 2.3 they have nonempty intersection containing a point y. Then by construction $f(y) = -\infty$ which is a contradiction. $\square$

Theorem 3.3 (Theorem 3.5. [16]) *Let* $\mathrm{diam}(X) < \pi$. *Then closed convex sets are weakly closed.*

Lemma 3.4 (Proposition 3.8. [16]) *Let* $\mathrm{diam}(X) < \pi$. *Let* $f : X \to (-\infty, \infty]$ *be a quasiconvex lsc function. Then* f *is weakly lsc. In particular* $x \to d^2(a, x)$ *is weakly lsc on* $B_a(\pi/2)$.

Proof Assume on the contrary that

$$\liminf_{x \xrightarrow{w} x_0} f(x) < f(x_0).$$

This means that there exists a subsequence x_{n_k}, and index $k_0 \in \mathbb{N}$, and a $\delta > 0$ such that $f(x_{n_k}) < f(x_0) - \delta$ for all $k > k_0$. Using lsc and quasiconvexity of f, we get

$$f(y) \leq f(x_0) - \delta$$

for all $y \in \overline{\mathrm{co}}\{x_{n_k} : k > k_0\}$. This contradicts $x_0 \in \overline{\mathrm{co}}\{x_{n_k} : k > k_0\}$ by Lemma 3.2. [16].

Fact 4 in [26] is the quasiconvexity of $d(a, \cdot)$. □

We need the following result of Kendall, which assures the existence of jointly convex functions on CAT(1) spaces with convex geometry.

Theorem 3.5 (Yokota's Theorem A in [26]) *Let* $\mathrm{diam}(X) < \pi$. *There exists a jointly κ-convex lsc function* $\Phi : X \times X \to [0, \infty)$ *vanishing on the diagonal set* $\Delta := \{(x, x) | x \in X\}$ *for some* $\kappa > 0$.

Lemma 3.6 (Ekeland principle, cf. [12]) *Given* $x_0 \in X$ *and a lsc function* $f : X \to (-\infty, \infty]$ *that is bounded below, there exist* $\alpha, \beta \geq 0$ *such that*

$$f(x) \geq -\alpha d(x, x_0) - \beta$$

for all $x \in X$.

Definition 3.3 (*Mosco convergence*) A sequence of lsc functions $f_n : X \to \overline{\mathbb{R}}$ said to converge to $f : X \to \overline{\mathbb{R}}$ in the sense of Mosco if, for any $x \in X$, we have

(M1) $f(x) \leq \liminf_{n \to \infty} f_n(x_n)$ whenever $x_n \xrightarrow{w} x$,
(M2) there exists an $(y_n) \subseteq X$, such that $y_n \to x$ and $f_n(y_n) \to f(x)$.

Proposition 3.7 (*Ekeland principle for a Mosco convergent sequence of lsc semi-convex functions, bounded case) Let* X *be a CAT(1)-space with* $\mathrm{diam}(X) < \pi$. *Given* $x_0 \in X$ *and a sequence of lsc λ-convex functions* $f_n : X \to (-\infty, \infty]$ *that is Mosco converging to* $f : X \to (-\infty, \infty]$, *there exist* $\alpha, \beta \geq 0$ *such that*

$$f_n(x) \geq -\alpha d(x, x_0) - \beta$$

for all $x \in X$ *and* $n \in \mathbb{N}$.

Proof Assume that the assertion is false; that is, for any $k \in \mathbb{N}$, there exists $n_k \in \mathbb{N}$ and $x_k \in X$ such that

$$f_{n_k}(x_k) + k[d(x_k, x_0) + 1] < 0.$$

First assume that $n_k \to \infty$ as $k \to \infty$. Since we have $\mathrm{diam}(X) < \pi$, by Lemma 2.2 we can select a weakly convergent subsequence from x_k still denoted by x_k with weak limit denoted by $\overline{x} \in X$. Then by the Mosco convergence of f_n we have

$$f(\overline{x}) \leq \liminf_{k \to \infty} f_{n_k}(x_k) \leq \liminf_{k \to \infty} -k[d(x_k, x_0) + 1]$$
$$= -\limsup_{k \to \infty} k[d(x_k, x_0) + 1]$$
$$= -\infty$$

which is a contradiction.

What is left to check is the case when n_k remains bounded, that is after some index $k_0 \in \mathbb{N}$ we have $n_k = n_{k_0}$ for all $k \geq k_0$. Let x_0 denote the circumcenter of X which exists by Lemma 3.3 [8] and satisfies $d(x_0, x) < \frac{\pi}{2}$ uniformly for all $x \in X$. Consider the lsc function

$g(x) := f_{n_{k_0}}(x) + cd^2(x_0, x)$. Since $f_{n_{k_0}}$ is λ-convex, we claim that there exists a large enough $c > 0$ so that $g(x)$ is convex. Indeed, using Lemma 3.1 we get

$$(1 - t)g(\gamma(0)) + tg(\gamma(1)) - g(\gamma(t)) \geq \left(\frac{\lambda}{2} + c\frac{\kappa}{2}\right) t(1 - t)d^2(\gamma(0), \gamma(1))$$

$$\geq 0$$

if $c > 0$ is large enough. Pick such a $c > 0$. Then by Lemma 3.4 $g : X \to (-\infty, \infty]$ is weakly lsc, thus for a weakly convergent subsequence of x_k still denoted by x_k with weak limit $\overline{x} \in X$ we get

$$\begin{aligned}
g(\overline{x}) &\leq \liminf_{k \to \infty} f_{n_{k_0}}(x_k) + cd^2(x_0, x_k) \\
&\leq \liminf_{k \to \infty} -k[d(x_k, x_0) + 1] + cd^2(x_0, x_k) \\
&\leq -\limsup_{k \to \infty} k[d(x_k, x_0) + 1] + \liminf_{k \to \infty} cd^2(x_0, x_k) \\
&\leq -\limsup_{k \to \infty} k[d(x_k, x_0) + 1] + \text{const.} \\
&= -\infty
\end{aligned}$$

which is a contradiction.   □

For a potential function, we always consider a lower semi-continuous function $\phi : X \to \overline{\mathbb{R}}$. The *effective domain* of ϕ is defined as

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

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 the *Moreau–Yosida resolvent* 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 can consider two kinds of conditions on ϕ:

Assumption 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 [2, 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 [2, Corollary 2.2.2]).

Remark 3.1 If $\mathrm{diam}(X) < \infty$ 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 $\tau_*(\phi) = \infty$, in particular, (1) holds.

The following assures nonemptyness and unicity of $J_\tau^\phi(z)$:

Remark 3.2 For λ-convex ϕ by Proposition 3.26 in [18] $J_\tau^\phi(z)$ is a unique point in the same neighborhoods of z as in Lemma 3.1 for all $K + 2\lambda\tau > 0$, where $K > 0$. The K-convexity of $x \mapsto d^2(z,x)$ holds under $\mathrm{diam}(X) < \frac{\pi}{2} - \epsilon$ with $K := (\pi - 2\epsilon)\tan(\epsilon)$.

The following is one of the main results of this section.

Theorem 3.8 *Let* $\mathrm{diam}(X) < \pi$ *and* $f_n : X \to (-\infty, \infty]$ *a sequence of lsc λ-convex functions that is Mosco converging to* $f : X \to (-\infty, \infty]$*. Then*

$$\lim_{n \to \infty} (f_n)_\tau(x) = f_\tau(x)$$

and

$$\lim_{n \to \infty} J_\tau^{f_n}(x) = J_\tau^f(x)$$

for any small enough $\tau > 0$ and $x \in D(f)$.

Proof From Proposition 3.7 we have for some $\alpha, \beta \geq 0$:

$$f_n(J_\tau^{f_n}(x)) \geq -\alpha d(J_\tau^{f_n}(x), x) - \beta.$$

From the definition of $J_\tau^{f_n}(x)$ and taking $y_n \to x$ as in (M2), we have

$$f_n(y_n) + \frac{1}{2\tau} d^2(x, y_n) \geq f_n(J_\tau^{f_n}(x)) + \frac{1}{2\tau} d^2(x, J_\tau^{f_n}(x))$$

where $y_n \to x$, which combined with the above estimate yields

$$f_n(y_n) + \frac{1}{2\tau} d^2(x, y_n) + \alpha d(J_\tau^{f_n}(x), x) + \beta \geq \frac{1}{2\tau} d^2(x, J_\tau^{f_n}(x)).$$

That is,

$$0 \geq d^2(x, J_\tau^{f_n}(x)) - 2\tau\alpha d(J_\tau^{f_n}(x), x) - 2\tau(f_n(y_n) + \beta) - d^2(x, y_n).$$

Since $d^2(x, y_n) \to 0$ and $f_n(y_n) \to f(x) < \infty$ by (M2), for small enough $\tau > 0$ the above forces

$$d(x, J_\tau^{f_n}(x)) < \epsilon \tag{4}$$

for all large enough $n \in \mathbb{N}$, given an arbitrary $\epsilon \leq \frac{\pi}{2}$.

Now we are in position to prove $\lim_{n \to \infty} J_\tau^{f_n}(x) = J_\tau^f(x)$. Suppose on the contrary that there exists a subsequence $j_k \in J_\tau^{f_{n_k}}(x)$ such that $d(j_k, J_\tau^f(x))$ does not converge to 0. Since $\mathrm{diam}(X) < \pi$, there exists a subsequence of $\{j_k\}_{k \in \mathbb{N}}$ still denoted by j_k with weak limit $c \in X$.

Since $f_n \to f$ in the sense of Mosco, there exists a sequence $y_n \to J_\tau^f(x)$ and $f_n(y_n) \to f(J_\tau^f(x))$. Then using the weak lsc property of $d^2(x, \cdot)$ we get

$$
\begin{aligned}
\limsup_{k \to \infty}(f_{n_k})_\tau(x) &\leq \limsup_{k \to \infty} f_{n_k}(y_{n_k}) + \frac{1}{2\tau}d^2(x, y_{n_k}) \\
&= f(J_\tau^f(x)) + \frac{1}{2\tau}d^2(x, J_\tau^f(x)) \\
&\leq f(c) + \frac{1}{2\tau}d^2(x, c) \\
&\leq \liminf_{k \to \infty} f_{n_k}(j_k) + \frac{1}{2\tau}d^2(x, j_k) \\
&= \liminf_{k \to \infty} f_{n_k}(J_\tau^{f_{n_k}}(x)) + \frac{1}{2\tau}d^2(x, J_\tau^{f_{n_k}}(x)) \\
&= \liminf_{k \to \infty}(f_{n_k})_\tau(x)
\end{aligned}
\tag{5}
$$

which yields $c \in J_\tau^f(x)$. Furthermore using (M1) and (M2) with $z_n \to c$ and $f_n(z_n) \to f(c)$ we have

$$
\begin{aligned}
\limsup_{k \to \infty} \frac{1}{2\tau}d^2(x, j_k) &\leq \limsup_{k \to \infty} -f_{n_k}(j_k) + f_{n_k}(z_{n_k}) + \frac{1}{2\tau}d^2(x, z_{n_k}) \\
&= -\liminf_{k \to \infty} f_{n_k}(j_k) + f(c) + \frac{1}{2\tau}d^2(x, c) \\
&\leq -f(c) + f(c) + \frac{1}{2\tau}d^2(x, c) \\
&= \frac{1}{2\tau}d^2(x, c) \\
&\leq \liminf_{k \to \infty} \frac{1}{2\tau}d^2(x, j_k)
\end{aligned}
$$

which together with (4) and Lemma 2.1 prove the strong convergence

$$
j_k \to c
$$

which contradicts the assumption that $d(j_k, J_\tau^f(x))$ does not converge to 0. Then using (5) we get

$$
\begin{aligned}
\lim_{n \to \infty}(f_n)_\tau(x) &= \lim_{n \to \infty} f_n(J_\tau^{f_n}(x)) + \frac{1}{2\tau}d^2(x, J_\tau^{f_n}(x)) \\
&= f(J_\tau^f(x)) + \frac{1}{2\tau}d^2(x, J_\tau^f(x)) = f_\tau(x)
\end{aligned}
$$

finishing the proof of the second part of the assertion.   □

Remark 3.3 If $J_\tau^f(x)$ is not unique in the above Theorem 3.8, then it still follows from the same proof that all weak cluster points of $J_\tau^{f_n}(x)$ are in fact strong cluster points and are in $J_\tau^f(x)$.

The following is a variant of the above result using different assumptions.

Theorem 3.9 Let $f_n : X \to (-\infty, \infty]$ be a sequence of L-Lipschitz functions that is Mosco converging to $f : X \to \mathbb{R}$. Then

$$
\lim_{n \to \infty}(f_n)_\tau(x) = f_\tau(x)
$$

and

$$\lim_{n\to\infty} J_\tau^{f_n}(x) = J_\tau^f(x)$$

for any small enough $\tau > 0$.

Proof From the definition of $J_\tau^{f_n}(x)$, we have

$$f_n(x) \geq f_n(J_\tau^{f_n}(x)) + \frac{1}{2\tau}d^2(x, J_\tau^{f_n}(x))$$

which combined with the L-Lipschitz property yields

$$d^2(x, J_\tau^{f_n}(x)) \leq 2\tau[f_n(x) - f_n(J_\tau^{f_n}(x))],$$

$$d(x, J_\tau^{f_n}(x)) \leq 2\tau \frac{f_n(x) - f_n(J_\tau^{f_n}(x))}{d(x, J_\tau^{f_n}(x))}$$

$$\leq 2\tau L$$

From here for small enough $\tau > 0$ we obtain that

$$d(x, J_\tau^{f_n}(x)) < \epsilon$$

for all large enough $n \in \mathbb{N}$, given an arbitrary $\epsilon \leq \frac{\pi}{2}$. We can follow the proof of the previous Theorem 3.8 from (5) onward to obtain our assertions. □

The following is yet another variant of the above, it generalizes Proposition 5.12. in [17] which proves this in CAT(0)-spaces.

Theorem 3.10 *Let $f_n : X \to (-\infty, \infty]$ be a uniformly lower bounded sequence of lsc functions that is Mosco converging to $f : X \to (-\infty, \infty]$. Then*

$$\lim_{n\to\infty} (f_n)_\tau(x) = f_\tau(x)$$

and

$$\lim_{n\to\infty} J_\tau^{f_n}(x) = J_\tau^f(x)$$

for any small enough $\tau > 0$ and $x \in D(f)$.

Proof Without loss of generality assume the uniform lower bound $f_n, f \geq 0$. Since $f \not\equiv +\infty$, we have $f_\tau(x) < +\infty$ and thus $f(J_\tau^f(x)) < +\infty$. By (M2) there exists an $\{y_n\}_{n\in\mathbb{N}} \subseteq X$, such that $y_n \to x$ and $f_n(y_n) \to f(x)$. Then by definition we have

$$f_n(y_n) + \frac{1}{2\tau}d^2(x, y_n) \geq f_n(J_\tau^{f_n}(x)) + \frac{1}{2\tau}d^2(x, J_\tau^{f_n}(x)),$$

which yields

$$d^2(x, J_\tau^{f_n}(x)) \leq 2\tau[f_n(y_n) - f_n(J_\tau^{f_n}(x))] + d^2(x, y_n)$$

$$\leq 2\tau f_n(y_n) + d^2(x, y_n) \to 2\tau f(x)$$

as $n \to \infty$. Choosing small enough $\tau > 0$ this eventually yields that

$$d(x, J_\tau^{f_n}(x)) < \epsilon$$

for all large enough $n \in \mathbb{N}$, given an arbitrary $\epsilon \leq \frac{\pi}{2}$. We can follow the proof of the previous Theorem 3.8 from (5) onward to obtain our assertions. □

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

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

and set

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

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

$$x_\tau^0 := x_0 \text{ and recursively choose } x_\tau^k \in J_{\tau_k}^\phi(x_\tau^{k-1}) \text{ for each } k \in \mathbb{N}. \tag{6}$$

We call $\{x_\tau^k\}_{k \in \mathbb{N}}$ a *discrete solution* of the variational scheme (6) associated with the partition $\mathscr{P}_\tau$, which is thought of as a *discrete-time gradient curve* for the potential function ϕ. The following a priori estimates (see [2, Lemma 3.2.2]) will be useful in the sequel. We remark that these estimates are easily obtained if ϕ is bounded below.

Lemma 3.11 (A priori estimates) *Let $\phi : X \to (-\infty, \infty]$ satisfy Assumption 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 $\mathscr{P}_\tau$ and an associated discrete solution $\{x_\tau^k\}_{k \in \mathbb{N}}$ of (6) satisfy*

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

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

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

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

$$d^2(x_0, x_\tau^k) \leq \left(\sum_{l=1}^k d(x_\tau^{l-1}, x_\tau^l) \right)^2 \leq \sum_{l=1}^k \frac{d^2(x_\tau^{l-1}, x_\tau^l)}{\tau_l} \cdot \sum_{l=1}^k \tau_l \leq 2C t_\tau^k. \tag{7}$$

The following result ensures the existence of the gradient curve.

Theorem 3.12 (gradient curve, cf. [25]) *Fix an initial point $x_0 \in D(\phi)$ and consider discrete solutions $\{x_{\tau_i}^k\}_{k \in \mathbb{N}}$ with $x_{\tau_i}^0 = x_0$ associated with a sequence of partitions $\{\mathscr{P}_{\tau_i}\}_{i \in \mathbb{N}}$ such that $\lim_{i \to \infty} |\tau_i| = 0$. Then the piecewisely interpolated curve $\bar{x}_{\tau_i} : [0, \infty) \to X$ converges to a curve $\xi : [0, \infty) \to X$ with $\xi(0) = x_0$ as $i \to \infty$ uniformly on each bounded interval $[0, T]$. In particular, the limit curve ξ does not depend on the choice of the sequence of partitions nor discrete solutions.*

Theorem 3.13 (Error Estimate, cf. [25]) *Let $\lambda \leq 0$ and $\phi : X \to (-\infty, \infty]$ be lsc λ-convex and $|\tau| < \tau_*(\phi)$. Fix $x_0 \in D(\phi)$ and let $\xi(x_0) := \mathcal{G}(t, x_0)$ denote the gradient flow of ϕ. Then we have*

$$d^2(\bar{x}_\tau, \xi(t)) \leq e^{-2\lambda t} \left(\sqrt{1 - K' - \frac{2\lambda}{3}|\tau|} + \sqrt{-\frac{4\lambda}{3}|\tau|} \right)^2 \tag{8}$$

$$|\tau| \{\phi(x_0) - \phi(\bar{x}_\tau(t))\}$$

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

Thanks to the a priori estimate (Lemma 3.11), we have

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

Theorem 3.14 (Contraction property, cf. [25]) *Let ϕ be a lower semi continuous λ-convex function. 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 $t > 0$,*

$$d(\xi(t), \zeta(t)) \leq e^{-\lambda t} d(x_0, y_0). \tag{9}$$

The following inequality is proved in Lemma 3.34. in [18] for all $K + 2\lambda\tau > 0$, where $K > 0$ is the K-convexity of $x \mapsto d^2(a, x)$. The K-convexity of $x \mapsto d^2(z, x)$ holds under $\mathrm{diam}(X) < \frac{\pi}{2} - \epsilon$ with $K := (\pi - 2\epsilon) \tan(\epsilon)$. We have $K > 0$ for example on $\overline{B}_a((\pi - \epsilon)/2)$ by Lemma 3.1, thus $K > 0$ for any $a \in \overline{B}_z((\pi - \epsilon)/4)$ and then

$$d(J_\tau^f(x), J_\tau^f(y)) \leq 4 \frac{2d(x, y) + d(x, J_\tau^f(y)) + d(J_\tau^f(x), y)}{K + 2\lambda\tau} d(x, y) \tag{10}$$

by Lemma 3.34 in [18] for any $x, y \in \overline{B}_z((\pi - \epsilon)/4)$ and fixed z.

Now we are in position to prove the main result of this section regarding the continuity of gradient flows under Mosco convergence. The proof also works in the same way in CAT(0)-spaces providing a new proof of this result.

Theorem 3.15 *Let (X,d) be a CAT(1) space. Assume $f_n : X \to (-\infty, \infty]$ is a sequence of lsc semi-convex functions that is Mosco converging to $f : X \to (-\infty, \infty]$. Let $J_\tau^f(x)$ and $J_\tau^{f_n}(x)$ be the corresponding resolvents such that for all small enough $\tau > 0$*

$$J_\tau^{f_n}(x) \to J_\tau^f(x)$$

as established earlier under the extra assumptions imposed by either Theorems 3.8 or 3.9 or 3.10 and let $S_t(x)$ and $S_t^n(x)$ denote the corresponding gradient flows. Then

$$\lim_{n \to \infty} S_t^n(x) = S_t(x)$$

for any $x \in D(f)$.

Proof By (M2) there exists an $y_n \to x$ with $f_n(y_n) \to f(x)$. Then

$$
\begin{aligned}
d(S_t(x), S_t^n(x)) &\leq d(S_t(x), S_t^n(y_n)) + d(S_t^n(y_n), S_t^n(x)) \\
&\leq d(S_t(x), S_t^n(y_n)) + e^{-\lambda t} d(y_n, x),
\end{aligned} \tag{11}
$$

furthermore

$$
\begin{aligned}
d(S_t(x), S_t^n(y_n)) &\leq d\left(S_t(x), \left(J_{t/k}^f\right)^k (x)\right) \\
&\quad + d\left(\left(J_{t/k}^f\right)^k (x), S_t^n(y_n)\right) \\
&\leq d\left(S_t(x), \left(J_{t/k}^f\right)^k (x)\right) + d\left(\left(J_{t/k}^f\right)^k (x), \left(J_{t/k}^{f_n}\right)^k (y_n)\right) \\
&\quad + d\left(\left(J_{t/k}^{f_n}\right)^k (y_n), S_t^n(y_n)\right).
\end{aligned} \tag{12}
$$

To estimate the second term of (12) we argue as

$$
\begin{aligned}
d & \left(\left(J_{t/k}^{f} \right)^{k}(x), \left(J_{t/k}^{f_n} \right)^{k}(y_n) \right) \\
\leq & \, d \left(J_{t/k}^{f} \left(\left(J_{t/k}^{f} \right)^{k-1}(x) \right), J_{t/k}^{f_n} \left(\left(J_{t/k}^{f} \right)^{k-1}(x) \right) \right) \\
& + d \left(J_{t/k}^{f_n} \left(\left(J_{t/k}^{f} \right)^{k-1}(x) \right), J_{t/k}^{f_n} \left(\left(J_{t/k}^{f_n} \right)^{k-1}(x) \right) \right) \\
& + d \left(J_{t/k}^{f_n} \left(\left(J_{t/k}^{f_n} \right)^{k-1}(x) \right), \left(J_{t/k}^{f_n} \right)^{k}(y_n) \right) \\
= & \, d \left(J_{t/k}^{f} \left(\left(J_{t/k}^{f} \right)^{k-1}(x) \right), J_{t/k}^{f_n} \left(\left(J_{t/k}^{f} \right)^{k-1}(x) \right) \right) \\
& + d \left(J_{t/k}^{f_n} \left(\left(J_{t/k}^{f} \right)^{k-1}(x) \right), J_{t/k}^{f_n} \left(\left(J_{t/k}^{f_n} \right)^{k-1}(x) \right) \right) \\
& + d \left(\left(J_{t/k}^{f_n} \right)^{k}(x), \left(J_{t/k}^{f_n} \right)^{k}(y_n) \right).
\end{aligned}
\tag{13}
$$

First notice that (4) is established in all of Theorems 3.8,3.9,3.10. Then (10) establishes a Lipschitz estimate for resolvents in CAT(1) spaces with small enough radius where (4) ensures boundedness of Lipschitz constants, so applying (13) recursively for each k we obtain an upper bound in which the already established pointwise convergence of resolvents assures this second term also goes to 0 for each fixed k as $n \to \infty$.

We continue by estimating each of the remaining terms on the right hand side of (12). By (8) with $C_0 := \sqrt{1 - K' - \frac{2\lambda}{3}|\tau|} + \sqrt{-\frac{4\lambda}{3}|\tau|}$ and $x_{t/k}^{k} := J_{t/k}^{f}\left(x_{t/k}^{k-1}\right)$, $x_{t/k}^{0} := x$ we have that

$$
d \left(S_t(x), \left(J_{t/k}^{f} \right)^{k}(x) \right) \leq e^{-\lambda t} C_0 \sqrt{\frac{t}{k}} \sqrt{f(x) - f(x_{t/k}^{k})}
\tag{14}
$$

and similarly with $y_{t/k}^{k} := J_{t/k}^{f_n}\left(y_{t/k}^{k-1}\right)$, $y_{t/k}^{0} := y_n$ we also have that

$$
\begin{aligned}
d \left(\left(J_{t/k}^{f_n} \right)^{k}(y_n), S_t^{n}(y_n) \right) & \leq e^{-\lambda t} C_0 \sqrt{\frac{t}{k}} \sqrt{f_n(y_n) - f_n(y_{t/k}^{k})} \\
& \leq e^{-\lambda t} C_0 \sqrt{\frac{t}{k}} \sqrt{f(x) - f(x_{t/k}^{k}) + \epsilon_n},
\end{aligned}
\tag{15}
$$

where to obtain the second inequality for fixed k we used (M2) for $f_n(y_n) \to f(x)$, and (M1) for $f(x_{t/k}^{k}) \leq \liminf_{n\to\infty} f_n(y_{t/k}^{k})$ where $y_{t/k}^{k} \to x_{t/k}^{k}$ as $n \to \infty$ by (13), thus the remainder $\epsilon_n \to 0$.

Combining all these estimates together with the a priori estimate Lemma 3.11 for f, for a given $\epsilon > 0$ first choose a large enough k and then a large enough N to obtain that the right hand side of (11) is less than ϵ for all $n > N$. □

4 Sequences of spaces with an asymptotic relation

Let (X_i, d_i), (X, d) be complete CAT(1)-spaces. Define

$$\mathcal{X} := X \sqcup \left(\bigsqcup_i X_i \right)$$

as a disjoint union.

Definition 4.1 (*Asymptotic relation*, cf. [17]) We call a topology on $\mathcal{X}$ an *asymptotic relation between X_i and X* if

(1) X_i and X are all closed in $\mathcal{X}$, and the restricted topology of $\mathcal{X}$ on each of X_i and X coincides with its original topology;
(2) for any $x \in X$ there exists a net $x_i \in X_i$ converging to x in $\mathcal{X}$;
(3) if $X_i \ni x_i \to x \in X$ and $X_i \ni y_i \to y \in X$ in $\mathcal{X}$, then we have $d_i(x_i, y_i) \to d(x, y)$;
(4) if $X_i \ni x_i \to x \in X$ and $y_i \in X_i$ is a net with $d_i(x_i, y_i) \to 0$, then $y_i \to x$ in $\mathcal{X}$.

In this section we assume that X_i and X have an asymptotic relation. The following definition can be found in [17] which generalizes weak convergence to an asymptotic relation in the CAT(0) setting. We modify it suitably for our CAT(1)-setting.

Definition 4.2 (*Weak convergence*) A net $x_i \in X_i$ *weakly converges* to a point $x \in X$, that is $x_i \overset{w}{\to} x$, if for any net of π-convex geodesic segments $\gamma_i \subseteq X_i$ strongly converging to a π-convex geodesic segment $\gamma \in X$ with $\gamma(0) = x$, such that $d_i(\gamma_i, x_i) := \inf_{z \in \gamma_i} d_i(z, x_i) < \pi/2$, we have that $P_{\gamma_i}(x_i)$ strongly converges to x. Strong convergence of γ_i to γ means that $\gamma_i(t) \to \gamma(t)$ for each $t \in D(\gamma)$.

Similarly to the case of a single CAT(1)-space it is easy to prove that strong convergence implies weak convergence and that weak limit points are unique in small enough metric balls.

Let $D \subset M_\kappa^2$ be a closed Jordan domain whose boundary has finite length, and let X be a metric space. A map $f : D \to X$ is *majorizing* if it is 1-Lipschitz and its restriction to the boundary ∂D is length-preserving. If Γ is a closed curve in X, we say that D majorizes Γ if there is a majorizing map $f : D \to X$ such that the restriction $f|_{\partial D}$ traces out Γ. See Section 8.12 of [1] for the following result.

Theorem 4.1 (Reshetnyak's majorization theorem) *For any closed curve Γ in a* CAT(κ) *space (of length at most $R_\kappa := \frac{\pi}{\sqrt{\kappa}}$ when $\kappa > 0$), there exists a convex region D in M_κ^2, and an associated map f such that D majorizes Γ under f.*

The following three results appeared in [17] as Lemma 5 (3–5) in the CAT(0)-setting and have also been mentioned in [19] without proofs in the CAT(1)-case. We prove them below in the CAT(1)-setting.

Lemma 4.2 *Let $x_i, y_i \in X_i$ and $x, y \in X$ such that $d_i(x_i, y_i), d(y, x) < \pi/2$ for all i. Assume $x_i \overset{w}{\to} x$ and $y_i \to y$. Then we have*

(1) $d(x, y) \leq \liminf_i d_i(x_i, y_i)$,
(2) $d_i(x_i, y_i) \to d(x, y)$ *if and only if $x_i \to x$.*

Proof (1) : Take a net $\hat{x}_i \to x$ such that all geodesic segments $[\hat{x}_i, y_i]$ are π-convex. By the obtuse-angle property the spherical law of cosines gives

$$\cos d_i(x_i, y_i) \leq \cos d_i(x_i, P_{[\hat{x}_i, y_i]}(x_i)) \cos d_i(y_i, P_{[\hat{x}_i, y_i]}(x_i))$$
$$+ \sin d_i(x_i, P_{[\hat{x}_i, y_i]}(x_i)) \sin d_i(y_i, P_{[\hat{x}_i, y_i]}(x_i)) \cos \angle_{P_{[\hat{x}_i, y_i]}(x_i)}(x_i, y_i)$$
$$\leq \cos d_i(x_i, P_{[\hat{x}_i, y_i]}(x_i)) \cos d_i(y_i, P_{[\hat{x}_i, y_i]}(x_i))$$
$$\leq \cos d_i(y_i, P_{[\hat{x}_i, y_i]}(x_i)),$$

which implies

$$d_i(x_i, y_i) \geq d_i(y_i, P_{[\hat{x}_i, y_i]}(x_i)),$$

thus

$$\liminf_i d_i(x_i, y_i) \geq \liminf_i d_i(y_i, P_{[\hat{x}_i, y_i]}(x_i)) = d(y, x).$$

(2) : The implication "$\Leftarrow$" is obvious. We shall prove "$\Rightarrow$". The assumption yields that in the proof of (1) instead of inequalities we have equalities, thus it follows that $\cos d_i(x_i, P_{[\hat{x}_i, y_i]}(x_i)) \to 1$ implying $d_i(x_i, P_{[\hat{x}_i, y_i]}(x_i)) \to 0$. Since $P_{[\hat{x}_i, y_i]}(x_i) \to x$, this implies $x_i \to x$.   □

Lemma 4.3 *Let $x_i \in X_i$ be a net and $\gamma_i, \sigma_i : [0, 1] \to X_i$ geodesic segments such that x_i, γ_i, σ_i are contained in $(\pi/2 - \epsilon)$-convex sets for each index i and a given $\epsilon > 0$ and*

$$\lim_i d_i(\gamma_i(0), \sigma_i(0)) = \lim_i d_i(\gamma_i(1), \sigma_i(1)) = 0. \tag{16}$$

Then we have

$$\lim_i d_i(P_{\gamma_i}(x_i), P_{\sigma_i}(x_i)) = 0.$$

Proof The $(\pi/2 - \epsilon)$-convexity assumption and (16) with Theorem 4.1 assures that

$$\lim_i \sup_{t \in [0,1]} d_i(\gamma_i(t), \sigma_i(t)) = 0. \tag{17}$$

Alternatively to obtain (17) instead of Theorem 4.1 one can also use Ohta's L-convexity [23] available in CAT(1)-spaces. We set $y_i := P_{\gamma_i}(x_i)$, $z_i := P_{\sigma_i}(x_i)$ and take $s_i, t_i \in [0, 1]$ satisfying $\gamma_i(s_i) = y_i$ and $\sigma_i(t_i) = z_i$. Then (17) leads to

$$\lim_i d_i(\sigma_i(s_i), y_i) = \lim_i d_i(\gamma_i(t_i), z_i) = 0.$$

By Exercise 2.6(1) [9] P is a nonexpansive map, thus $d_i(x_i, y_i) \leq d_i(x_i, \gamma_i(t_i))$ and $d_i(x_i, z_i) \leq d_i(x_i, \sigma_i(s_i))$, we have

$$\lim_i |d_i(x_i, y_i) - d_i(x_i, z_i)| = 0,$$
$$\lim_i |d_i(x_i, y_i) - d_i(x_i, \gamma_i(t_i))| = 0.$$

By the spherical law of cosines we have

$$\cos d_i(x_i, \gamma_i(t_i)) \leq \cos d_i(x_i, y_i) \cos d_i(y_i, \gamma_i(t_i))$$

which implies

$$\cos d_i(x_i, \gamma_i(t_i)) - \cos d_i(x_i, y_i) \leq \cos d_i(x_i, y_i)(\cos d_i(y_i, \gamma_i(t_i)) - 1),$$

and also using the trigonometric addition formulas for cos we have that $\cos d_i(x_i, \gamma_i(t_i)) - \cos d_i(x_i, y_i) \to 0$ since $d_i(x_i, y_i) - d_i(x_i, \gamma_i(t_i)) \to 0$. Furthermore by the $(\pi/2 - \epsilon)$-convexity assumption we have

$$\cos d_i(x_i, \gamma_i(t_i)), \cos d_i(x_i, y_i) \geq s$$

for some $s > 0$, so the above implies $0 \geq (\cos d_i(y_i, \gamma_i(t_i)) - 1) \to 0$, equivalently $d_i(y_i, \gamma_i(t_i)) \to 0$ and thus also $d_i(y_i, z_i) \to 0$. □

In the proof of the result below we assume that the spaces X_i, X are separable, however the argument should generalize to nets in non-separable spaces, since the main tool of Cantor's diagonalization process is available there as well, see for example [4]. The proof itself closely follows its CAT(0) variant of Lemma 5.5 in [17].

Lemma 4.4 *Let $\epsilon > 0$ be arbitrary. Any $x_i \in X_i$ net satisfying $d_i(x_i, o_i) < \pi/4 - \epsilon$ for a net $o_i \to o \in X$ has a weakly convergent subnet.*

Proof We may assume that $\{x_i\}$ is a countable sequence. First take a dense countable subset $\{\xi_\nu\}_{\nu \in \mathbb{N}} \subseteq B_o(\pi/4 - \epsilon) \subseteq X$. Let $y_0 \in B_o(\pi/4 - \epsilon)$ and select $y_{0,i} \in B_{o_i}(\pi/4 - \epsilon) \subseteq X_i$ such that $y_{0,i} \to y_0$ strongly. For each $\nu \in \mathbb{N}$, we select a sequence $\xi_{\nu,i} \in B_{o_i}(\pi/4 - \epsilon) \subseteq X_i$ so that $\xi_{\nu,i} \to \xi_\nu$ strongly. Then we have $\gamma_{\nu,i}^0 := [y_{0,i}, \xi_{\nu,i}] \to [y_0, \xi_\nu] =: \gamma_\nu^0$ as $i \to \infty$. Since $d_i(x_i, y_{0,i})$ is bounded, $w_{\nu,i}^0 := P_{[y_{0,i}, \xi_{\nu,i}]}(x_i)$ has a convergent subsequence whose limit denoted by w_ν^0 is a point in $[y_0, \xi_\nu]$. By Cantor's diagonalization process, we can choose a common subsequence of $\{i\}$ independent of $\nu \in \mathbb{N}$ for which $w_{\nu,i}^0 \to w_\nu^0$ and we denote this subsequence with the indices $\{i\}$. Select an arbitrary sequence $0 < \epsilon_m \to 0$. We define $y_{m,i}$ and y_m inductively as follows. We assume that $y_{m,i}$ and y_m are defined so that $y_{m,i} \to y_m$. The sequence $w_{\nu,i}^m := P_{[y_{m,i}, \xi_{\nu,i}]}(x_i)$ has a convergent subsequence, which can again be chosen independent of $\nu \in \mathbb{N}$ and we replace $\{i\}$ with that subsequence and set $w_\nu^m := \lim_i w_{\nu,i}^m$. There exists a number $\nu(m+1) \in \mathbb{N}$ such that

$$d(y_m, w_{\nu(m+1)}^m) > \sup_\nu d(y_m, w_\nu^m) - \epsilon_m.$$

Define $y_{m+1} := w_{\nu(m+1)}^m$ and $y_{m+1,i} := w_{\nu(m+1),i}^m$. Then $y_{m,i} \to y_m$ for each m.

By the obtuse-angle property, the spherical law of cosines implies

$$\cos d_i(y_{m,i}, x_i) \leq \cos d_i(x_i, y_{m+1,i}) \cos d_i(y_{m+1,i}, y_{m,i}).$$

Taking a subsequence of $\{i\}$ again, we can assume that for each m, $d_i(y_{m,i}, x_i)$ converges to some $\lambda_m \in \mathbb{R}$. This implies that λ_m is monotone nonincreasing, thus $d(y_{m+1}, y_m) \to 0$ as well. For any $\nu \in \mathbb{N}$,

$$\lim_i d_i(y_{m,i}, w_{\nu,i}^m) = d(y_m, w_\nu^m) < \epsilon_m + d(y_{m+1}, y_m) =: \epsilon_m' \to 0. \tag{18}$$

There exists $\{\nu(l,i)\}$ such that $\lim_i \xi_{\nu(l,i)} = y_l$. By (18), if $\nu \ll i$, then $d_i(y_{m,i}, w_{\nu,i}^m) \leq \epsilon_m'$. We may assume $\nu(l,i) \ll i$, so that we have

$$\limsup_i d_i(y_{m,i}, w_{\nu,i}^m) \leq \epsilon_m'.$$

Since $[y_{m,i}, \xi_{\nu(l,i),i}] \to [y_m, y_l]$, we can choose a common subsequence $\{i\}$ independent of m, l for which $w_{\nu(l,i),i}^m \in [y_{m,i}, \xi_{\nu(l,i),i}]$ converges to some point $x_{m,l} \in [y_m, y_l]$. By the inequality above

$$d(y_m, x_{m,l}) \leq \epsilon_m'.$$

Applying Lemma 4.3 to $\gamma_i := [y_{m,i}, y_{l,i}]$ and $\sigma_i := [y_{m,i}, y_{v(l,i),i}]$, we have $P_{[y_{m,i}, y_{l,i}]}(x_i) \to x_{m,l}$. Therefore,

$$d(y_l, y_m) \leq d(y_l, x_{l,m}) + d(x_{l,m}, y_m) \leq \epsilon'_m + \epsilon'_l$$

implying that $\{y_m\}$ is Cauchy. Let $x := \lim_m y_m$. By (18), for any geodesic γ emanating from x with an associated sequence of geodesics $\gamma_i \in X_i$ strongly converging to γ we have that $P_{\gamma_i}(x_i) \to x$ strongly. Then it follows that $x_i \overset{w}{\to} x$ completing the proof. $\qquad\square$

Remark 4.1 It is interesting to compare the statement of the above Lemma 4.4 with Lemma 2.2. Can we have the weak compactness of larger sets as well in Lemma 4.4 such as in Lemma 2.2 for a single space? Note that the proof above uses projections in an essential way, on the other hand the asymptotic center approach in [26] can prove Lemma 2.2 in a different way obtaining weak compactness in π-convex sets.

Definition 4.3 (*Mosco convergence for asymptotic relations*) A sequence of lsc functions $f_n : X_n \to \overline{\overline{\mathbb{R}}}$ said to converge to $f : X \to \overline{\overline{\mathbb{R}}}$ in the sense of Mosco if, for any $x \in X$, we have

(M1) $f(x) \leq \liminf_{n\to\infty} f_n(x_n)$ whenever $x_n \in X_n$ and $x_n \overset{w}{\to} x$,
(M2) there exists a sequence $y_n \in X_n$, such that $y_n \to x$ and $f_n(y_n) \to f(x)$.

We prove a version of Proposition 3.7 adapted to the setting of this section.

Proposition 4.5 (Ekeland principle, bounded case) *Let* $\mathrm{diam}(X) < \pi/2 - \epsilon$ *for an* $\epsilon > 0$. *Given* $w_n \in X_n$ *such that* $w_n \to w \in X$ *and a sequence of lsc* λ-*convex functions* $f_n : X_n \to (-\infty, \infty]$ *that is Mosco converging to* $f : X \to (-\infty, \infty]$. *Then there exist* $\alpha, \beta \geq 0$ *such that*

$$f_n(x_n) \geq -\alpha d_n(x_n, w_n) - \beta$$

for all $x_n \in X_n$ *and* $n \in \mathbb{N}$.

Proof Assume that the assertion is false; that is, for any $k \in \mathbb{N}$, there exists $n_k \in \mathbb{N}$ and $x_{n_k} \in X_{n_k}$ such that

$$f_{n_k}(x_{n_k}) + k[d_{n_k}(x_{n_k}, w_{n_k}) + 1] < 0.$$

We first assume that $n_k \to \infty$ as $k \to \infty$. Since we have $\mathrm{diam}(X) < \pi$, we can select a weakly convergent subsequence from x_{n_k} still denoted by x_{n_k} with weak limit denoted by $\overline{x} \in X$. Then by the Mosco convergence of f_n we have

$$\begin{aligned}
f(\overline{x}) &\leq \liminf_{k\to\infty} f_{n_k}(x_{n_k}) \leq \liminf_{k\to\infty} -k[d_{n_k}(x_{n_k}, w_{n_k}) + 1] \\
&\leq -\limsup_{k\to\infty} k[d_{n_k}(x_{n_k}, w_{n_k}) + 1] \\
&= -\infty
\end{aligned}$$

which is a contradiction. The rest of the proof is similar to the proof of Proposition 3.7. $\quad\square$

Theorem 4.6 *Let* $\mathrm{diam}(X) < \pi/2 - \epsilon$ *for an* $\epsilon > 0$ *and* $f_n : X_n \to (-\infty, \infty]$ *a sequence of lsc* λ-*convex functions that is Mosco converging to* $f : X \to (-\infty, \infty]$. *Then*

$$\lim_{n\to\infty} (f_n)_\tau(x_n) = f_\tau(x)$$

and

$$\lim_{n\to\infty} J_\tau^{f_n}(x_n) = J_\tau^f(x)$$

for any small enough $\tau > 0$ and $x \in D(f)$ and $x_n \in X_n$ with $x_n \to x$.

Proof From Proposition 4.5 we have

$$f_n(J_\tau^{f_n}(x_n)) \geq -\alpha d_n(J_\tau^{f_n}(x_n), w_n) - \beta.$$

From the definition of $J_\tau^{f_n}(x_n)$ and (M2), we have

$$f_n(y_n) + \frac{1}{2\tau}d_n^2(x_n, y_n) \geq f_n(J_\tau^{f_n}(x_n)) + \frac{1}{2\tau}d_n^2(x_n, J_\tau^{f_n}(x_n))$$

where $y_n \to x$, which combined with the above estimate yields

$$f_n(y_n) + \frac{1}{2\tau}d_n^2(x_n, y_n) + \alpha d_n(J_\tau^{f_n}(x_n), w_n) + \beta \geq \frac{1}{2\tau}d_n^2(x_n, J_\tau^{f_n}(x_n)).$$

That is,

$$0 \geq d_n^2(x_n, J_\tau^{f_n}(x_n)) - 2\tau\alpha d(J_\tau^{f_n}(x_n), w_n) - 2\tau(f_n(y_n) + \beta) - d_n^2(x_n, y_n).$$

Since $d_n^2(x_n, y_n) \to 0$ and $f_n(y_n) \to f(x) < \infty$ by (M2), for small enough $\tau > 0$ the above forces

$$d_n(x_n, J_\tau^{f_n}(x_n)) < s \tag{19}$$

for all large enough $n \in \mathbb{N}$, given an arbitrary $s \leq \frac{\pi}{4} - \epsilon$.

Now we are in position to prove $\lim_{n\to\infty} J_\tau^{f_n}(x_n) = J_\tau^f(x)$. Pick any subsequence $j_k \in J_\tau^{f_{n_k}}(x_{n_k})$. Since $\mathrm{diam}(X) < \pi/2 - \epsilon$, there exists a subsequence of $\{j_k\}_{k\in\mathbb{N}}$ still denoted by j_k with weak limit $c \in X$.

Since $f_n \to f$ in the sense of Mosco, there exists a sequence $y_n \to J_\tau^f(x)$ and $f_n(y_n) \to f(J_\tau^f(x))$. Then using (1) of Lemma 4.2 we get

$$
\begin{aligned}
\limsup_{k\to\infty}(f_{n_k})_\tau(x_{n_k}) &\leq \limsup_{k\to\infty} f_{n_k}(y_{n_k}) + \frac{1}{2\tau}d_{n_k}^2(x_{n_k}, y_{n_k}) \\
&= f(J_\tau^f(x)) + \frac{1}{2\tau}d^2(x, J_\tau^f(x)) \\
&\leq f(c) + \frac{1}{2\tau}d^2(x, c) \\
&\leq \liminf_{k\to\infty} f_{n_k}(j_k) + \frac{1}{2\tau}d_{n_k}^2(x_{n_k}, j_k) \\
&= \liminf_{k\to\infty} f_{n_k}(J_\tau^{f_{n_k}}(x_{n_k})) + \frac{1}{2\tau}d_{n_k}^2(x_{n_k}, J_\tau^{f_{n_k}}(x_{n_k})) \\
&= \liminf_{k\to\infty}(f_{n_k})_\tau(x_{n_k})
\end{aligned}
\tag{20}
$$

which yields $c \in J_\tau^f(x)$. Furthermore using (M1) and (M2) with $z_n \to c$ and $f_n(z_n) \to f(c)$ we have

$$\limsup_{k\to\infty} \frac{1}{2\tau} d_{n_k}^2(x_{n_k}, j_k) \le \limsup_{k\to\infty} -f_{n_k}(j_k) + f_{n_k}(z_{n_k}) + \frac{1}{2\tau} d_{n_k}^2(x_{n_k}, z_{n_k})$$

$$= -\liminf_{k\to\infty} f_{n_k}(j_k) + f(c) + \frac{1}{2\tau} d^2(x, c)$$

$$\le -f(c) + f(c) + \frac{1}{2\tau} d^2(x, c)$$

$$\le \frac{1}{2\tau} d^2(x, c)$$

$$\le \liminf_{k\to\infty} \frac{1}{2\tau} d_{n_k}^2(x_{n_k}, j_k)$$

which together with (19) and (2) of Lemma 4.2 prove the strong convergence

$$j_k \to c.$$

Then using (20) we get

$$\lim_{n\to\infty} (f_n)_\tau(x_n) = \lim_{n\to\infty} f_n(J_\tau^{f_n}(x_n)) + \frac{1}{2\tau} d_n^2(x_n, J_\tau^{f_n}(x_n))$$

$$= f(J_\tau^f(x)) + \frac{1}{2\tau} d^2(x, J_\tau^f(x)) = f_\tau(x)$$

finishing the proof of the second part of the assertion.   □

At this point the analogs of Theorem 3.9 or Theorem 3.10 for asymptotic relations can be established as well in the same manner. We omit the details.

Theorem 4.7 *Assume $f_n : X_n \to (-\infty, \infty]$ is a sequence of lsc semi-convex functions that is Mosco converging to $f : X \to (-\infty, \infty]$. Let $J_\tau^f(x)$ and $J_\tau^{f_n}(x_n)$ be the corresponding resolvents such that for all small enough $\tau > 0$*

$$J_\tau^{f_n}(x_n) \to J_\tau^f(x)$$

as established earlier in this section and let $S_t(x)$ and $S_t^n(x_n)$ denote the corresponding gradient flows. Then

$$\lim_{n\to\infty} S_t^n(x_n) = S_t(x)$$

for any $t \ge 0$ and $x \in D(f)$ and $x_n \in X_n$ with $x_n \to x$.

Proof By (M2) there exists an $y_n \to x$ with $f_n(y_n) \to f(x)$. Since $y_n, x_n \to x$, thus $d_n(y_n, x_n) \to 0$ implying

$$d_n(S_t^n(y_n), S_t^n(x_n)) \le e^{-\lambda t} d_n(y_n, x_n) \to 0,$$

so it is enough to show that $S_t^n(y_n) \to S_t(x)$. First we claim that

$$\left(J_{t/k}^{f_n}\right)^k (y_n) \to \left(J_{t/k}^f\right)^k (x) \tag{21}$$

for any fixed $k \in \mathbb{N}$. Indeed, $y_n \to x$ and assuming that (21) holds for a $k \ge 0$, Theorem 4.6 proves (21) for $k + 1$ yielding the claim by induction.

Next, by (8) with $C_0 := \sqrt{1 - K' - \frac{2\lambda}{3}|\tau|} + \sqrt{-\frac{4\lambda}{3}|\tau|}$ and $x_{t/k}^k := J_{t/k}^f\left(x_{t/k}^{k-1}\right), x_{t/k}^0 := x$ we have that

$$d\left(S_t(x), \left(J_{t/k}^f\right)^k (x)\right) \le e^{-\lambda t} C_0 \sqrt{\frac{t}{k}} \sqrt{f(x) - f(x_{t/k}^k)} \tag{22}$$

and similarly with $y_{t/k}^k := J_{t/k}^{f_n}\left(y_{t/k}^{k-1}\right)$, $y_{t/k}^0 := y_n$ we also have that

$$
\begin{aligned}
d_n\left(\left(J_{t/k}^{f_n}\right)^k(y_n), S_t^n(y_n)\right) &\le e^{-\lambda t} C_0 \sqrt{\frac{t}{k}} \sqrt{f_n(y_n) - f_n(y_{t/k}^k)} \\
&\le e^{-\lambda t} C_0 \sqrt{\frac{t}{k}} \sqrt{f(x) - f(x_{t/k}^k) + \epsilon_n},
\end{aligned}
\tag{23}
$$

where to obtain the second inequality for fixed k we used (M2) for $f_n(y_n) \to f(x)$, and (M1) for $f(x_{t/k}^k) \le \liminf_{n\to\infty} f_n(y_{t/k}^k)$ where $y_{t/k}^k \to x_{t/k}^k$ as $n \to \infty$ by (21), thus the remainder $\epsilon_n \to 0$.

Now choose sequences $z_n^t, j_n^{t,k} \in X_n$ such that $z_n^t \to S_t(x)$ and $j_n^{t,k} \to \left(J_{t/k}^f\right)^k(x)$ as $n \to \infty$. Then it is enough to show that $d_n(S_t^n(y_n), z_n^t) \to 0$. We estimate as

$$
\begin{aligned}
d_n(S_t^n(y_n), z_n^t) &\le d_n\left(S_t^n(y_n), \left(J_{t/k}^{f_n}\right)^k(y_n)\right) \\
&\quad + d_n\left(\left(J_{t/k}^{f_n}\right)^k(y_n), j_n^{t,k}\right) + d_n\left(j_n^{t,k}, z_n^t\right)
\end{aligned}
\tag{24}
$$

Combining all these estimates together with the a priori Lemma 3.11 for f and an arbitrary $\epsilon > 0$, we first choose a large enough k so that the right hand side of (22) is less than ϵ which implies that there exists a large enough $N_1 > 0$ such that the last term on the right hand side of (24) is less than 3ϵ for all $n > N_1$. Then choose a large enough $N_2 > N_1$ such that for all $n > N_2$ by (21) the second term on the right hand side of (24) is less than 2ϵ. Finally choose a large enough $N_3 > N_2$ such that for all $n > N_3$ the right hand side of (23) is less than 2ϵ making the first term on the right hand side of (24) less than 2ϵ. This implies that $d_n(S_t^n(y_n), z_n^t) < 7\epsilon$, thus the desired $S_t^n(y_n) \to S_t(x)$ follows. $\qquad\square$

Data availability statement Data sharing is not applicable to this article as no new data were created or analyzed in this study.

Declarations

Conflict of interest The authors declare no conflict of interest.

Ethics approval An ethics statement is not applicable because this study is based exclusively on published literature.

References

1. Alexander, S., Kapovitch, V., Petrunin, A.: Alexandrov geometry: preliminary version no. 1 (2019)
2. Ambrosio, L., Gigli, N., Savaré, G.: Gradient Flows in Metric Spaces and in the Space of Probability Measures. Birkhäuser Verlag, Basel (2005)
3. Attouch, H.: Variational Convergence for Functions and Operators. Applicable Mathematics Series, Pitman (Advanced Publishing Program), Boston (1984)
4. Azouzi, Y.: A diagonal principle for nets. arXiv:2304.04189
5. Bačák, M.: Convex Analysis and Optimization in Hadamard Spaces. De Gruyter Series in Nonlinear Analysis and Applications, vol. 22. De Gruyter, Berlin (2014)
6. Bačák, M.: Convergence of nonlinear semigroups under nonpositive curvature. Trans. Am. Math. Soc. **367**(6), 3929–3953 (2015)
7. Bačák, M., Montag, M., Steidl, G.: Convergence of functions and their Moreau envelopes on Hadamard spaces. J. Appl. Theory (C) **224**, 1–12 (2017)
8. Balser, A., Lytchak, A.: Centers of convex subsets of buildings. Ann. Global Anal. Geom. **28**, 201–209 (2005)
9. Bridson, M.R., Haefliger, A.: Metric Spaces of Non-Positive Curvature. Springer, Berlin, Heidelberg (1999)
10. Crandall, M.G., Liggett, T.M.: Generation of semi-groups of nonlinear transformations on general Banach spaces. Am. J. Math. **93**, 265–298 (1971)
11. Dal Maso, G.: An Introduction to Γ-Convergence. Birkhäuser Boston Inc., Boston, MA (1993)
12. Ekeland, I.: Nonconvex minimization problems. Bull. Am. Math. Soc. **1**, 443–474 (1979)
13. Jost, J.: Equilibrium maps between metric spaces. Calc. Var. PDE **2**, 173–204 (1994)
14. Jost, J.: Nonlinear Dirichlet forms, New directions in Dirichlet forms, pp. 1–47. Providence, American Mathematical Soceity (1998)
15. Jost, J.: From linear to convex. J. Convex Anal. **28**(2), 599–612 (2021)
16. Kell, M.: Uniformly convex metric spaces. Anal. Geom. Metric Spaces **2**(1), 359–380 (2014)
17. Kuwae, K., Shioya, T.: Variational convergence over metric spaces. Trans. Am. Math. Soc. **360**(1), 35–75 (2008)
18. Kuwae, K.: Resolvent flows for convex functionals and P-harmonic maps. Anal. Geom. Metr. Spaces **3**, 46–72 (2015)
19. Kuwae, K.: Variational convergence over p-uniformly convex spaces, in geometric aspect of partial differential equations and conservation laws. RIMS Kokyuroku **2013**, 22–39 (1842)
20. Mayer, U.F.: Gradient flows on nonpositively curved metric spaces and harmonic maps. Commun. Anal. Geom. **6**, 199–253 (1998)
21. Mosco, U.: Composite media and asymptotic Dirichlet forms. J. Funct. Anal. **123**(2), 368–421 (1994)
22. Muratori, M., Savaré, G.: Gradient flows and evolution variational inequalities in metric spaces. I: structural properties. J. Funct. Anal. **278**, 108347 (2020)
23. Ohta, S.-I.: Convexities of metric spaces. Geom. Ded. **125**(1), 225–250 (2007)
24. Ohta, S., Pálfia, M.: Discrete-time gradient flows and law of large numbers in Alexandrov spaces. Calc. Var. PDE **54**, 1591–1610 (2015)
25. Ohta, S., Pálfia, M.: Gradient flows and a Trotter–Kato formula of semi-convex functions on CAT(1)-spaces. Am. J. Math. **139**(4), 937–965 (2017)
26. Yokota, T.: Convex functions and barycenter on CAT(1)-spaces of small radii. J. Math. Soc. Jpn. **68**(3), 1297–1323 (2016)