Approximation for convex functional on non positively curved spaces and the Trotter-Kato product formula

October 18, 2010

Igor Stojković Mathematical Institute Leiden University NielsBohrweg 1, 2333 CM Leiden Netherlands
email: stojkov@ math.liedenuniv.nl

Abstract

We study the validity of Trotter-Kato product formula in the setting of gradient flows on CAT(0) metric spaces. We follow the same strategy as in the Hilbert space case, but instead of the linear structure and inerproduct we have only geodesics and the CAT(0) inequality available. Thus we have to show several other approximation theorems as well , which are of independant interest. A major difficulty in comparison with the linear theory is the lack of an appropriate notion of weak convergence on such spaces. This difficulty is successfully overcome with the aid of ultrafilters technique.

1 Introduction

During the past few decades much of the classical analysis of linear spaces has been extended to a metric spaces setting of one or another kind. One such striking development is the theory of gradient flows on metric spaces. In the monograph [2] authors L.Ambrosio,N.Gigli and G.Savare give a rather extensive treatment of this topic adopting a purely metric point of weiw. In their work these authors, inspired by fundamental ideas of Di Giorgy (such as metric slope and curves of maximal slope) and by incredible discoveries of Jordan–Kinderleiher–Otto [18] and Otto [34], authors of [2]construct a metric counterpart of a large part of the gradient flow theory on Hilbert spaces (which can be found in [4]). One can also see the work of Carillo–McCann–Villani [8] for related topics or Villani [45] and [46] for a very detailed exposition in this direction.

The main aim of this paper is verification of the Trotter–Kato product formula (see [23]) for gradient flows on complete on CAT(0) spaces (i.e. complete geodesic metric spaces which are non positively curved is the sense of Alexandrov). An integral part of the theory of gradient flows and maximal monotone operators on Hilbert spaces are various approximation theorems and in particular the product formula (see [4] chapter IV). Same holds for the theory of accreative operators on Banach spaces, and these results are generalizations of approximation theory for linear C_0–semigroups on Banach spaces (see Engel–Negel [9] Chapter III). The product formula originates from the Trotter's 1959. product formula for the matrix exponentials which asserts that for any pair of complex matrces $A, B \in \mathbb{C}^{d^2}$

$$e^{A+B} = \lim_{k\to+\infty} (e^{\frac{A}{k}} e^{\frac{B}{k}})^k \tag{1.1}$$

The importance of various approximation theorems in different contexts is of theoretical aa well as of applied character. One of the most well known examples on the theoretical side are appropriate approximations of a given contraction semigroup in the proof of the Hille-Yosida theorem-the kee step in the proof of this famous theorem (see [9] for the theory of linear semigroups and [4] for a nonlinear counterpart). It is also a way to show that certain properties are passed to the limit. Moreover one might encounter a situation where a product formula is the only way to construct a semigroup corresponding to the 'sum' of two generators (see [40]). On the applied side there are instances where one has two convex functionals which generate semigroups such that for each of these flows one has nice expressions. While the flow induceed by the sum functional is very hard to handle. This occurs for instance if one wants to impose a convex constrainto an equation (i.e. force solutions to stay in a convex domain) which is acheaved by adding the indicator functional of this domain to the original one (see [23] Section Examples). In such situation the product formula is applied togain insight about the constrained solution which may be very hard to handle directly from the variational inequalities.

It seems that there is no satisfactory version of this theorem at a reasonable level of generally beyond the linear setting – therefore we consider our efforts appropriate. However the product formula in the linear setting is a consequence of several other approximation theorems. In particular a theorem stating that convergence of resolvents implies convergence of semigroups, and these results are of independent interest (see [5] for the theory on Banach spaces). Droping the assumption of linearity of underlying space does not seem to make our aim more easy to reach, therefore our goal does require effort indeed.

CAT(0) spaces have been attracting much attention already since the birth of comparison geometry in 1940's. Being the natural generalization of non-positively curved Riemannian manifolds, CAT(0) spaces have very intersting geometric properties. As a matter a fact the larger class of matric spaces with one sided curvature bound in the sense of Alexandrov are certainly the most studied non-smooth geometric objects up to the current date. The foundations of this theory are now well developed and a detailed exposition of this topic can be found in [6], [7] and [1] among other monographs.

However not only purely geometric aspects of CAT(0) spaces have been studied up to now.

The theory of harmonic maps on Riemannian manifolds has been extended to CAT(0) spaces during the 1990's by various authors. This was initiated by M.Gromov and R.Schoen in light of some applications to the group theory (see [9]). This work was followed by a series of remarkable papers by R.Schoen, Korevaar and J.Jost (see [9], [19], [20] and [21]), and many more by different authors.

There have also been many instances where various authors studied monotone and accreative operator theory and the corresponding semigroups in non-smooth setting. In particular on a Hilbert ball, and among other works we have monographsm [3], [37], [38], [39] by S.Reich et all. Moreover these questions in a Hilbert manifold have been subject of investigations (see [16], and also [17] for a treatment of a 'vawe-like' equation on a Hilbert manifold by Iwamiya and Okochi-these authors make no assumptions about the curvature, but assume instead a lot on the part of the functional. Recently in [24] and [27] Li, Lopez and Martin–Marques treated some problem involving monotone vector fields on Hadamard manifolds in relation with some optimization problems (see also [36]).

A gradient flow generation theorem in CAT(0) spaces was provided by U.Mayer in [30] and this is the first published result on gradient flows beyond linear setting. However we also have the unpublished work of G.Perelman and A.Petrunin [35]. K.T.Sturm developed a large portion of stochastics for CAT(0) valued stochastic processes, such as Markov operators and matingales (see [41], [42], [43] and [44]). Moreover in 2007 Lorotonda shows that the set of positive invertible unitized Hilbert–Schmidt operators on a Hilbert manifold (see references thesis for many more papers in such spirit, such as [14]). Finally let us mention works of S-I.Ohta [32] and [33] in which he introduces Sobolev spaces of functions with CAT(0) targets and considers flows on the Wassesstein spaces over compact CAT(0) spaces respectively.[1]

To conclude this introduction, let us give the content of the remainder of the paper. Section 2 contains some background material. In section 3 we construct the approximation semigroups which we need for proving the Trotter–Kato product formula. We also show two convergence theorems. In particular a theorem which asserts that convergence of resolvents implies convergence of the corresponding semigroups. We will not formulate nor prove these theorems at the same level of generality as it has been done in the linear setting for one should first give a treatment of maximal monotone operators on CAT(0) spaces. However this is a work in progress of the author of this paper. This section corresponds to the context of [5] and the proof of our Lemma 3.10 on Banach spaces can be found in [31]. In Section 4 we prove the Trotter–Koto product formula and we follow the strategy of [23] as far as possable in our setting. However in this paper authors use the weak convergence on Hilbert spaces, and there is no

[1] However we do not claim that our list is complete by any means at all. It is only a short list of relevant literature given to help the reder to orientate.

suitable counterpart of it on general CAT(0) spaces (but see [11] for results in this direction). Instead of using weak convergence and inspired by a succesfull employment of ultrafilters in the context of gradient flows on metric spaces by Lytchak in [29], we exploit the idea in our context (contarary to Lytchak's work our functionals are only lowersemicontinuous, and not locally Lipschitz) and we complete the argument by these means.[2]

2 Preliminaries

In this section we present the basic concepts and facts that we need in the remainder of the exposition.

Definition 2.1 *Let X be a metric space. A curve $x : [a, b] \to X$ is called a geodesic if for all $s, t \in [a, b]$*

$$d(x(s), x(t) = |t - s| d(x(a), x(b)) \tag{2.1}$$

X is a (uniquely) geodesic space if any pair of points x_0 and x_1 can be connected by a (unique) geodesic. By composing with an affine increasing homeomorphism we can always asume our geodesics to be defined on $[0, 1]$.

Definition 2.2 *A geodesic metric space X is called CAT(0) space [3] (named in honour of Carton–Alexandrov–Topoganov) if for any pair of points $x_0, x_1 \in X$ and a (base) point $y \in X$ each geodesic $x : [0, 1] \to X$ connecting x_0 and x_1 satisfies the following (CAT(0)) inequality:*

$$d^2(y, x(t)) \leqslant (1 - t)d^2(y, x_0) + td^2(y, x_1) - t(1 - t)d^2(x_0, x_1) \tag{2.2}$$

It is easy to show that CAT(0) spaces are uniquely geodesic.

Proposition 2.3 *If X is a CAT(0) spaces then it is uniquely geodesic.*

Proof See [6] Chapter II Proposition 1.4 (1). ∎

In the sequal we will use the following notation: if X is a CAT(0) space, $x, y \in X$ and $t \in [0, 1]$, than

$$(1 - t)x \oplus ty \tag{2.3}$$

denotes the point $\gamma(t)$ on the unique geodesic $\gamma : [0, 1] \mapsto X$ which connects x and y.

Example 1. A convex subset of a Hilbert space is a CAT(0) space. However, normed space is CAT(0) only if its norm is generated by an inner product.

[2] During authors visit to Münster, Germany in august 2010, his host Anton Petrumin told the author about the work [29]. Author wishes to thank A.Petrunin for his kind and usefull advise.

[3] Some authors call such spaces nonpositively curved spaces.

2. A C^3 simply connected Riemannian manifold with non-positive sectional curvature equiped with it's Riemannian metric is a CAT(0) space.

3. The set of positive unitized Hilbert–Schmidt operators on a Hilbert space can be given a Riemannian structure that makes it a smooth Hilbert manifold which is metricly complete, geodesic and globaly non–positively curved, i.e. a complete CAT(0) space (see [9] and references for more examples of Hilbert manifolds that are defined as a certain subset of the operator space of a Hilbert space).

4. Eucledean Bruhat Tits buildings are CAT(0) spaces.

5. The unit ball in a Hilbert space equiped with the hyperbolic metric is a CAT(0) space (see [13])

6. Let X be CAT(0) space, let $(M, \mathcal{M}, m)$ be a measure space and consider the se Ω of L^2 maps $f : M \to X$ (i.e. for some hence all points $x \in X$ $\int_M d^2(f(\omega), x)dm(\omega) < +\infty$), and define the distance by

$$d^2(f_1, f_2) := \int_M d^2(f_1(\omega), f_2(\omega))dm(\omega).$$

   Than (Ω, d) is a complete CAT(0) space.

7. The spaces $W^{1,2}_\Phi(\Omega, X)$ introduced in [2] 3.2 Section 6 are CAT(0) spaces as it is shown in this work.

There are several other equivalent characterization of CAT(0) spaces. In order to formulate these we need to introduce two concepts which are basic tools in the study of geometry in non-smooth setting.

The first one is the concept of a comparison triangle which was extensively used by the Russian mathematician Alexandrov and his school in 1940's, though it seems to appear for the first time in the work Australian mathematician Wald. Given any triple of distinct points p, q, r in a metric space X there is a unique up to isometry triangle in $\mathbb{R}^2$ with vertices $\bar{p}, \bar{q}, \bar{r} \in \mathbb{R}^2$ such that

$$d(p, q) = |\bar{p} - \bar{q}|, d(q, r) = |\bar{q} - \bar{r}|, d(p, r) = |\bar{p} - \bar{r}|, \tag{2.4}$$

where $|.|$ denotes the Euclidean norm on $\mathbb{R}^2$. This follows by the triangle inequality for the distance d. given any pair of points $\bar{r}$,$\bar{r}$ in $\mathbb{R}^2$ the line segment $[\bar{q}, \bar{r}]$ is by definition the set of convex combinations of points $\bar{q}$ and $\bar{r}$ in the linear space $\mathbb{R}^2$, i.e.

$$[\bar{q}, \bar{r}] = \{(1 - t)\bar{q} + t\bar{r} \quad t \in [0, 1]\}.$$

More generally given a pair of points q and r in our metric space X and a geodesic $\gamma : [0, a] \to X$ with end points $\gamma(0) = q$ and $\gamma(1) = r$ geodesic segment $[q, r]$ is the image of γ. If X is not uniquely geodesic than $[q, r]$ may indeed depend on the particular choice of a geodesic γ.

If p, q, r are distinct points in X and $\gamma_1, \gamma_2, \gamma_3$ is a choice of geodesics joining them, then the corresponding geodesic triangle denoted by $\Delta = \Delta(p, q, r)$ is the set of points in the union of the images of these three geodesics. The corresponding comparison triangle is usually denoted by $\bar{\Delta} = \bar{\Delta}(\bar{p}, \bar{q}, \bar{r})$. Note that the geometry of $\bar{\Delta}$ depends only on quantities $d(p, q)$, $d(p, r)$ and $d(q, r)$, not on $\gamma_1, \gamma_2, \gamma_3$.

For any triple of points $\bar{p}, \bar{q}, \bar{r} \in \mathbb{R}^2$ (actually in any Hilbert space), and for eacht $t \in [0, 1]$ the point $\bar{x} = (q - t)\bar{q} + t\bar{r} \in [\bar{q}, \bar{r}]$ satisfies the following identity:

$$|\bar{p}-\bar{x}|^2 = |\bar{p}-((1-t)\bar{q}+t\bar{r})|^2 = (1-t)^2|\bar{p}-\bar{q}|^2+t^2|\bar{p}-\bar{r}|^2+t(1-t)|\bar{p}-\bar{q}||\bar{p}-\bar{r}| \quad (2.5)$$

Now we can easily see that a geodesic metric space X is CAT(0) iff for any triple of distinct points $p, q, r \in X$ a choice of geodesic segment $[q, r]$ and a point $x = x(t) \in [q, r]$ the (unique uo to isometry) comparison triangle $\bar{\Delta} = \bar{\Delta}(\bar{p}, \bar{q}, \bar{r})$ for the geodesic triangle $\Delta = \Delta(p, q, r)$ satisfies

$$d(p, x) \leqslant |\bar{p} - \bar{x}(t)| \quad (2.6)$$

where $\bar{x}(t) := (1 - t)\bar{q} + t\bar{r}$.

The second concept the Alexandrov upper angle between geodesics. In a metric space geodesics should be seen as straight lines, and in a flat space (like $\mathbb{R}^n$, are more generally a convex subset of a Hilbert space) if p, q, r are points the angle $\angle p(q, r)$ between line segments $[p, q]$ and $[p, r]$ (the geodesics are unique in this situation) is given by the cosine low

$$|q - r|^2 = |q - p|^2 + |p - r|^2 - 2|q - p||p - r| \cos \angle p(q, r). \quad (2.7)$$

Equivalently

$$\cos \angle(\vec{x}, \vec{y}) = \frac{\langle \vec{x}, \vec{y} \rangle}{|\vec{x}||\vec{y}|} \quad (2.8)$$

for any pair of vectors $\vec{x}$ and $\vec{y}$ in our flat space (and we adopt the convention that angles assume values in $[0, \pi)$).

If the space under consideration is a Riemannian manifold M, than the angle between two geodesic segments with common end point is naturally defined as angle between their velocity vectors at this end point. Both of these vectors are elements of the tangent space T_pM to M at p, and the angle between tham is computed as in (2.7). However we can express this quantity purely in terms of the distance function, and subsequently use that expression as the definition of the (upper or lower) angle between geodesics in any metric space.

Definition 2.4 *Let p be a point in a metric space X, and let $\gamma_1, \gamma_2 : [0, 1 \to X$ be two geodesics emaniting from p, i.e. $\gamma_1(0) = \gamma_2(0) = p$. For each $s, t \in (0, 1]$ let $\bar{\Delta}_{s,t} := \bar{\Delta}(\bar{p}, \overline{\gamma_1(s)}, \overline{\gamma_2(t)})$ denote a comparison triangle for the triangle $\Delta(p, \gamma_1(s), \gamma_2(t)) \subset X$. Let moreover $\alpha_{s,t}$ be the angle in $\bar{\Delta}_{s,t}$ at vertex $\bar{p}$. The Alexandrov upper angle between γ_1 and γ_2 is defined by*

$$\angle_p(\gamma_1, \gamma_2) := \limsup_{s,t \downarrow 0} \alpha_{s,t} \quad (2.9)$$

Equivalently one can define

$$\angle_p(\gamma_1,\gamma_2) := \lim_{\varepsilon\to 0} \sup_{s\in(0,\varepsilon);t\in[0,1]} \alpha_{s,t} \tag{2.10}$$

(for proof see [14], Proposition 1.16). Notice that definition coincides with (2.8 if $X=\mathbb{R}^n$, and also that $\angle_p(\gamma_1,\gamma_2)$ is in fact defined in terms od the distance function via (2.7).

Proposition 2.5 *Let p be point in a metric space X and let $\gamma_1,\gamma_2,\gamma_3$ be geodesics with $\gamma_1(0)=\gamma_2(0)=\gamma_3(0)=p$. Then*

$$\angle_p(\gamma_1,\gamma_3) \leqslant \angle_p(\gamma_1,\gamma_2)+\angle_p(\gamma_2,\gamma_3) \tag{2.11}$$

i.e. Alexandrov angle satisfies the triangle inequality.

Proof See [6]. ∎

We remark that if $p,q,r\in X$ and γ_1,γ_2 are geodesics with $\gamma_1(0)=\gamma_2(0)=p$, $\gamma_1(1)=q$, $\gamma_2(1)=r$, we might write $\angle_p(q,r)$ instead of $\angle_p(\gamma_1,\gamma_2)$ or $\angle([p,q],[p,r])$ provided there is no ambiguity.

The notion of angle gives rise to two other characterizations of CAT(0) spaces, and we state all of these characterizations in the following proposition.

Proposition 2.6 *Let X be a geodesic space. The following conditions are equivalent*

1. *X is a CAT(0) space*
2. *For every geodesic triangle $\Delta(p,q,r)$ in X and every point $x\in[p,q]$, the comparison point $\bar{x}\in[\bar{q},\bar{r}]\subset\Delta(\bar{p},\bar{q},\bar{r})\subset\mathbb{R}^2$ satisfies*

$$d(p,x)\leqslant|\bar{p}-\bar{x}| \tag{2.12}$$

3. *For every geodesic triangle $\Delta(p,q,r)$ in X and every pair of points $x\in[p,q]$, $y\in[p,r]$ distinct from p, the angles at the vertices corresponding to p in the comparison triangles $\bar{\Delta}(\bar{p},\bar{q},\bar{r})$, $\bar{\Delta}(\bar{p},\bar{x},\bar{y})\subset\mathbb{R}^2$ satisfy*

$$\angle_{\bar{p}}(\bar{x},\bar{y})\leqslant\angle_{\bar{p}}(\bar{q},\bar{r}) \tag{2.13}$$

4. *The Alexandrov upper angle between the sides of geodesic triangle in X is not greater than the angle between the corresponding sides of it's comparison triangle in $\mathbb{R}^2$.*

If any of the above conditions hiold, than for each triple of points $p,q,r\in X$ we have

$$d^2(q,r)\geqslant d^2(p,q)+d^2(p,r)-2d(p,q)d(p,r)\cos\angle_p(q,r) \tag{2.14}$$

Proof See [6]. ∎

Remark In exactly the same way one can compare the geometry of triangles in metric space X with other model spaces to define CAT(k) spaces for all $k \in \mathbb{R}$- by requiring angles to be thinner than in the model space. For $k = 1$ the model space is the unit sphere S^2 equiped with the angular metric (see proposition 2.5) given by

$$\cos d(A, B) := \langle A, B\rangle \quad A, B \in S^2 \tag{2.15}$$

where by definition $d(A, B) \in [0, \pi]$. For other $k > 0$ this distance is multiplied by $\frac{1}{x}$.

For $k = -1$ the model space is the hyperbolic space $\mathbb{H}^2$, and one of the possable ways to define it is the following. Consider the symmetric bilinear form on $\mathbb{R}^3$ given by $\langle x|y\rangle := x_1y_1 + x_2y_2 - x_3y_3$. Then $\mathbb{H}^2 = \{x \in \mathbb{R}^3 | \langle x|\rangle = -1, x_3 > 0\}$, equiped with the distance function given by

$$\cosh d(x, y) := -\langle x|y\rangle \tag{2.16}$$

For other $k < 0$ one rescales this distance by $\frac{1}{\sqrt{-k}}$. See [14] Section 2 for a detailed treatment.

However in this thesis we restrict our consideration to CAT(0) spaces. There is also a characterization of CAT(x) spaces based on comparison of quadiliterals (see [14] Proposition 1.11).

Proposition 2.7 *If X is a CAT(0) space then the distance function $d : X \times X \to \mathbb{R}_{\geqslant 0}$ is convex, i.e. given any pair of geodesics $\gamma_1, \gamma_2 : [0, 1] \to X$ the following inequality holds for all $t \in [0, 1]$*

$$d(\gamma_1(t), \gamma_2(t)) \leqslant (1 - t)d(\gamma_1(0), \gamma_2(0)) + td(\gamma_1(1), \gamma_2(1)) \tag{2.17}$$

Proof See [6] Chapter II Proposition 2.2. ∎

The convexity property stated above is clearly weaker than the CAT(0) inequality (2.2). Indeed the same inequality holds in uniformly Banach spaces (which are uniquely geodesic, and the geodesics are precisely the convex combination of pairs of distinct points), but as we already remarked in Example 2.3 Hilbert spaces are the only normed spaces that are also CAT(0).

CAT(0) spaces have another interesting property as Hilbert spaces do. Namely, one can define the nearest point projection onto a closed convex set. Here a subset $A \subset X$ is said to be convex if for any pair of points $x \neq y$ in A the image of any geodesic γ that connects x and y lies in A.

Proposition 2.8 *Let X be a CAT(0) space and let $A \subset X$ be complete for the induced metric. Than*

1. *For every $x \in X$ there exists a unique point $\pi(x) \in A$ such that $d(x, \pi(x)) = d(x, A) := \inf_{y \in A} d(x, y)$.*

2. *If x' belongs to the geodesic segment $[x, \pi(x)]$ then $\pi(x) = \pi(x')$.*

3. *Given $x \notin A$ and $y \in A$, if $y \neq \pi(x)$ then $\angle_{\pi(x)}(x, y) \geqslant \pi/2$.*

4. *The map $x \mapsto \pi(x)$ is a contractive retraction defined on X: $d(\pi(x), \pi(y)) \leqslant d(x, y)$ for $x, y \in X$ and the map $H : X \times [0, 1] \to X$ according to (x, t) be point a distance $(x\pi(x))$ from x on the geodesic segment $[x, \pi(x)]$ is a continuous homotopy from the identity map of X to π.*

Proof See [6]

Remark By the cosine law (2.14), item 3. above implies that for $v \in A$

$$d^2(x, v) - d^2(\pi(x), v) \geqslant d^2(x, \pi(x)) \tag{2.18}$$

In order to state one of the results of this paper we will also use the following general fact about the metric spaces. As usually a topological space is said to be local compact if every point in it has a relatively compact neighborhood.

Theorem 2.9 *(The Hopf–Rinow–theorem) Let X be a length space. If X is complete and locally compact then*

1. *Every closed bounded set is compact*

2. *X is a geodesic space.*

Proof See [6] Chapter I Proposition 3.7.

Next we give some facts from the theory of gradient flows on metric spaces.

Definition 2.10 *Let X be a geodesic space. A subset $A \subset X$ is convex if for each pair of points $x \neq y$ in A for some geodesic segment $[x, y]$ we have $[x, y] \subset A$.*

A map $\varphi : X \to (-\infty, +\infty]$ is called a functional, and it is said to be geodesically convex (and we will abbrevate to convex) if for any pair of points $x_0 \neq x_1$ in X any geodesic $\gamma : [0, 1] \to X$ with $\gamma(0) = x_0$ $\gamma(1) = x_1$ satisfies

$$\varphi(\gamma(t)) \leqslant (1 - t)\varphi(\gamma(0)) + t\varphi(\gamma(1)). \tag{2.19}$$

$D(\varphi) := \{x \in X | \varphi(x) < +\infty\}$ is called the (proper) domain of φ, and φ is said to be proper if $D(\varphi) \neq \emptyset$. φ is lower semicontinuous (shortly denoted l.s.c.) if for each sequence $\{x_n\}$ in X $\lim_n x_n = x$ implies $\varphi(x) \leqslant \liminf_n \varphi(x_n)$.

The sets $L_a := \{x \in X | \varphi(x) \leqslant a\}$ for $a \in \mathbb{R}$ are called level sets.

A curve $u : [a, b] \mapsto X$ is absolutely continuous of order $p \geqslant 1$ denoted $AC^p([a, b]; X)$, if there is a nonnegative function $m \in L^p([a, b]; R)$ such that for $a \leqslant c \leqslant d \leqslant b$

$$d(u(c), u(d)) \leqslant \int_c^d m(s)ds. \tag{2.20}$$

If the above holds for a curve u than its metric derivative

$$|\dot{u}|(t) := \lim_{h \mapsto 0} \frac{d(u(t+h), u(t))}{h} \tag{2.21}$$

is defined $\mathcal{L}^1 a.e.$ on a, b and it is the smallest function m such that (2.20) holds (see the first section in [2] for details). Moreover a curve $u : (a, b) \mapsto X$ is locally absolutely continuous of order $p \geqslant 1$ denoted $AC^p_{loc}((a, b); X$, if it is absolutely continuous of same order on each compact subinterval of (a, b), .

As it is explained in [2] there is an appropriate metric formulation of the gradient flow inclusion $\frac{d}{dt}X(t) \in -\partial\varphi(X(t)), \quad X(0) = x \in \overline{D(\varphi)}$ asociated with a convex l.s.c. functional φ on a Hilbert space H ($\partial\varphi$ denotes the subdifferential of φ). This is the so called Evolution Variational inequalitu (shortly denoted EVI), i.e. one seeks a continuous curve $u : [0, +\infty) \mapsto X$ of class $AC^2_{loc}([0, +\infty); X)$ such that for each $v \in D(\varphi)$

$$\frac{d}{dt} d^2(x(t), v) + \varphi(x(t)) \leqslant \varphi(x) \quad \text{for} \quad \mathcal{L}^1 - a.e. \quad t \in [0, +\infty) \tag{2.22}$$

In [2] authors work with a certain genralized convexity assumption (see Assumption 4.0.1 and also a weaker Assumption 2.4.5 there) which is easily seen to hold if X is a CAT(0) space and φ is geodesicly convex[4]. Next to this assumption, in [2] the functional under consideration is assumed to be bounded from bellow on some ball in the space (see (2.1.2b) there). However this always holds in the CAT(0) setting. This fact is proven in [30] Lemma 1.3. Hence all of the theory presented in [2] applies if the folowing assumption holds:

Assumption 2.11 *X is complete CAT(0) space. $\varphi : X \to (-\infty, +\infty]$ a proper, convex and lower semicontinuous functional.*

Proposition 2.12 *Let Assumption 2.11 hold. Than there for any $x \in X$ there are constants $c, b \in \mathbb{R}$ such that*

$$\varphi(y) \geqslant c + bd(x, y) \quad \forall y \in X \tag{2.23}$$

Proof See [9] Lemma 4.1. ∎

In order to construct the semigroup of solutions of the EVI and prove its contractiveness, we define the resolvents of φ. For $x, y \in X$ and $h \geqslant 0$ we consider functionals

$$\Phi(h, x, y) := \frac{1}{2h} d^2(x, y) + \varphi(y) \tag{2.24}$$

and we look for a minimizer $y \in X$ for $x \in X$ and $h \geqslant 0$ fixed.

$$\mathcal{J}_h x := \text{argmin} \frac{1}{2h} d^2(x, y) + \varphi(y) \quad y \in X \tag{2.25}$$

[4]As a matter a fact these assumptions are modeled upon such situation for the reason that the variational aparatus can be built than.

This 'eliptic' problem has a unique solution, provided the above discussed assumptions hold, and moreover it has the following properties [5]:

Proposition 2.13 *Grant Assumption 2.11. Then the for each $h > 0$ and $x \in X$ the functional $y \mapsto \Phi(h, x, y)$ given by (2.24) admits a unique minimizer, denoted by $\mathcal{J}_h x$. Moreover $\mathcal{J}_h : \overline{D(\varphi)} \to D(\varphi)$ is a contraction each $h > 0$.*

Proof As we already pointed out our Assumption 2.11 insures assumpions in [2], and existance and uniqueness of minimizers are proven in Theorem 4.1.2 (i) there. The contractivity of the resolvents $\mathcal{J}_h : \overline{D(\varphi)} \to D(\varphi)$ under Assumption 2.16 is shown in [30] Lemma 1.12. ∎

For $x \in X$ and $h > 0$ we denote

$$\varphi_h(x) := \inf_{y \in X} \Phi(t, x, y) = \frac{1}{2h} d^2(x, \mathcal{J}_h x) + \varphi(\mathcal{J}_h x) \tag{2.26}$$

The folowing facts are crucial in constructing the solutions to EVI.

Proposition 2.14 *Grant Assumption 2.11. than if $x \in \overline{D(\varphi)}$ and if $\{x_n\}$ is a sequence in $D(\varphi)$ satisfying*

$$\limsup_{n \to +\infty} \Phi(t, x, x_n) \leqslant \varphi_h(x) \tag{2.27}$$

then $x_n \to \mathcal{J}_h x$ as $n \to +\infty$.

Moreover we have for $v \in D(\varphi)$

$$\frac{1}{2h} d^2(\mathcal{J}_h x, v) - \frac{1}{2h} d^2(x, v) + \frac{1}{2h} d^2(x, \mathcal{J}_h x) + \varphi(\mathcal{J}_h x) \leqslant \varphi(v) \tag{2.28}$$

Notice that this is stronger than the minimizer condition for $\mathcal{J}_h x$!

Proof For the first statement see [4] Lemma 4.1.1. and for the second one see [4] Theorem 4.1.2 (ii). ∎

If the metric space under consideration has one sided curvature bounds in the sense of Alexandrov, than one can define the Eucledean tangent cones over the space of directions at each point in space, and also define an apropriate gradient of a convex functional (see [29], [32] and [1]). We will not use these notions, however we do need to use the notion is the 'norm' of the 'negative of the gradient', which is called the matric slope and is defined as follows [6]

Definition 2.15 *Let X be a metric space and let φ be a functional on X. Then for each $x \in D(\varphi)$ the metric slope $|\partial\varphi(x)|$ of φ of X is given by*

$$|\partial\varphi|(x) := \limsup_{y \to x} \frac{(\varphi(x) - \varphi(y))^+}{d(x, y)} \tag{2.29}$$

[5] We have formulated the main results about the gradient flows in the setting of our Assumption 2.11. We remark though that execpt for the contractivness of the resolvents, all the other properties hold under the generalized convexity and nondegenaracy assumptions mentioned above. Resolvents are than still continuous in x.

[6] This definition is due to Di Giorgi.

If our Asumption 2.11 holds then we may substitute lim sup by sup (see [9] Theorem 2). We may indeed have that $|\partial\varphi|(x) = +\infty$ for $x \in D(\varphi)$. Notice that is X is a Hilbert space and φ is convex and C^1, we have that $|\partial\varphi|(x) = |-\nabla\varphi|(x)$, which clarifies the relevance of this definition.

Proposition 2.16 *Let X be a metric space and let a functional φ on X given. Grant Assumption 2.11. Then for $x \in X$ and $h > 0$ we have $\mathcal{J}_h x \in D(|\partial\varphi|)$ and*

$$|\partial\varphi|(\mathcal{J}_h x) \leqslant \frac{d(\mathcal{J}_h x, x)}{h} \tag{2.30}$$

In particular $\overline{D(|\partial\varphi|)} = \overline{D(\varphi)} \subset X$.

If $x \in D(|\partial\varphi|)$ we also have the following estimates

$$|\partial\varphi|^2(\mathcal{J}_h x) \leqslant \frac{d^2(\mathcal{J}_h x, x)}{h} \leqslant 2\frac{\varphi(x) - \varphi_h(x)}{h} \leqslant |\partial\varphi|^2(x) \tag{2.31}$$

Proof For the first statement see [4] Lemma 3.1.3. For the second statement, see [4] Theorem 3.1.6 (2). ∎

Finally we state the theorem about the generation of the flow and its properties in our setting (but we remark that exactly the same theorem is proven in [2] under the generalized convexity and non degeneracy assumpions)

Theorem 2.17 *Grant assumption 2.11 Then we have the following:*

1. *(Convergence and exponential formula) For each $x_0 \in \overline{D(\varphi)}$ and $t > 0$ the limit*

$$\lim_{n\to\infty} \left(\mathcal{J}_{\frac{t}{n}}\right)^n x =: S_t x_0 \tag{2.32}$$

   exists. Moreover $(S_t)_{t\geqslant 0}$ is a semigroup on $\overline{D(\varphi)}$.

2. *(Regularizing effect) $x(t) := S_t x_0$ is a curve of class $AC^2_{loc}((0, +\infty); X)$ with $x(t) \in D(|\partial\varphi|) \subset D(\varphi)$ for each $t > 0$, i.e. for each $\varepsilon > 0$ $I > 0$ $t \mapsto x(t)$ is Lipschitz on $[\varepsilon, T]$ (actually on $[\varepsilon, +\infty)$) and*

$$\begin{aligned} \varphi(x(t)) &\leqslant \varphi_t(x_0) \leqslant \varphi(v) + \frac{1}{2t} d^2(v, x_0) \quad \forall v \in D(\varphi) \\ |\partial\varphi|^2(x(t)) &\leqslant |\partial\varphi|^2(v) + \frac{1}{t^2} d^2(v, x_0) \quad \forall v \in D(|\partial\varphi|) \end{aligned} \tag{2.33}$$

3. *(Uniqueness and EVI) $x(t)$ is unique solution of the EVI*

$$\frac{1}{2}\frac{d}{dt} d^2(x(t), v) + \varphi(x(t)) \leqslant \varphi(v) \quad \mathcal{L}^1 \quad a.e. \quad t > 0 \quad b\forall v \in D(\varphi) \tag{2.34}$$

   among all the locally absolutely continuous curves on $[0, +\infty)$ such that $\lim_{t\downarrow 0} x(t) = x_0$.

4. *(Contraction semigroup) The map* $(t,x) \mapsto S_t(x) : \overline{D(\varphi)} \times [0,\infty) \to \overline{D(\varphi)}$ *(where by definition* $S_0(x) := x \ \forall x \in \overline{D(\varphi)}$*) is a contraction semigroup*

$$d(S_t x, S_t y) \leqslant d(x,y) \quad \forall t \geqslant 0 \quad x,y \in \overline{D(\varphi)} \tag{2.35}$$

5. *(Optimal Error estimate) If* $x_0 \in D(\varphi)$ *then*

$$d^2(S_t x_0, (\mathcal{J}_{t/n})^n x_0) \leqslant \frac{t}{n}(\varphi(x_0) - \varphi_{t/n}(x_0)) \leqslant \frac{t^2}{2n^2}|\partial\varphi|^2(x_0) \tag{2.36}$$

for $t > 0$ *and* $n \in \mathbb{N}$.

6. *The equation*

$$\frac{d}{dt_+}\varphi(x(t)) = -|\partial\varphi|^2(x(t)) = -|\dot{x}|_+^2(t) = -|\partial\varphi|(x(t))|\dot{x}|_+(t) \tag{2.37}$$

is satisfied for every $t > 0$

Proof See [4] Theorem 4.04 for the first five items, and Theorem 2.4.15 for the last item ∎

Remark The function $t \mapsto \varphi(x(t))$ is nonincreasing since by (2.37) we have $\frac{d}{dt}\varphi(x(t)) = -|\partial\varphi|^2(x(t))$ and $t \mapsto x(t)$ is continuous up to 0, hence integrating 2.34 we obtain for any $x_0 \in \overline{D(\varphi)}$ and $t > 0$ and $v \in D(\varphi)$

$$\frac{1}{2t}d^2(x(t),v) - \frac{1}{2t}d^2(x(0),v) + \varphi(x(t)) \leqslant \varphi(v) \tag{2.38}$$

Moreover since we assume that φ is geodesicly convex and X is a CAT(0) space, $\overline{D(\varphi)}$ is a closed convex set in X, and denoting $\pi : X \to \overline{D(\varphi)}$ to be the orthogonal projection defined in Proposition 2.8, the Remark bellow this proposition give that for any $y \in X$ $t > 0$ and $v \in D(\varphi)$

$$\frac{1}{2t}d^2(x(t),v) - \frac{1}{2t}d^2(y,v) + \varphi(x(t)) \leqslant \varphi(v) \tag{2.39}$$

where $x(t)$ is the unique solution of the EVI subject to initial condition $x_0 := \pi(y)$.

Let us also state the classical Trotter product formula for convex functionals on Hilbert spaces for the convinience of the reader. First we need some notation. H is the Hilbert space under consideration and $\varphi_1, ..., \varphi_n$ are proper, convex, lower semicontinuous functionals on H. $D_j := D(\varphi_j)$, $E_j := \overline{D_j}$ $\varphi := \sum_{j=1}^n \varphi_j$, $D := D(\varphi) \neq 0$ by assumption, $E := \bar{D}$. Furthermore P_j denotes the nearest point projection onto E_j, $\left(S_t^j\right)_{t \geqslant 0}$ $j = \overline{1,n}$ denote the semigroups of solutions associated to φ_j's and $\left(\mathcal{J}_h^j\right)_{h \geqslant 0}$ denote the resolvents associated to φ. Finally for $t > 0$ we write $U_t := \mathcal{J}_t^j$ for $j = 1, ...n$ and $F_t := \mathcal{J}_t^n \cdot \mathcal{J}_t^{n-1} \cdot \dots \cdot \mathcal{J}_t^1$, or $U_t := S_t^j$ for $j = 1, ..., n$ and $F_t := (S_t^n \circ P^n) \cdot \left(S_t^{n-1} \circ P^{n-1}\right) \cdot \dots \circ \left(S_t^1 \circ P^1\right)$. Then we have

Theorem 2.18 *(Trotter product formula for convex functionals on Hilbert spaces) Let H be a Hilbert space, and let $\varphi_1, ..., \varphi_n$ be proper convex and lowersemicontinuous functionals such that $\varphi := \sum_{j=1}^{n} \varphi_j \neq +\infty$. Then*

$$\lim_{t \downarrow 0} \left[1 + \frac{\lambda}{t}(1 - F_t) \right]^{-1} x \to \mathcal{J}_\lambda x \quad \text{for } x \in H, \lambda > 0 \tag{2.40}$$

$$\lim_{n \to \infty} \left(F_{\frac{t}{n}} \right)^n x \to S_t x \text{ for } x \in E, t > 0 \tag{2.41}$$

Proof See [8] and [10]. ∎

The remainder of this section is devoted to ultrafilters, ultraproducts and ultralimits. Most of this material is borrowed from [6] and [1], and subsequently addapted to our situation-the functionals that these authors work with are assumed to be locally Lipschitz while we assume only the l.s.c. property.

Definition 2.19 *A non–principal ultrafilter on $\mathbb{N}$ is finitely additive probability measure ω such that each subset $S \subset \mathbb{N}$ is ω-measurable and we have: $\omega(S) \in \{0,1\}$ and $\omega(S) = 0$ if S is finite.*

Such ultrafilter exists (see exercise 5.48 in Chapter I of [6]) though it is not unique. However any such non–principal ultrafilter suffices in our considerations, and we fix one such ω henceforth.

Definition 2.20 *Let $(a_n)_n$ be a sequence in $[-\infty, +\infty]$. We say that $a \in [-\infty, +\infty]$ is the ω-limit of $(a_n)_n$ and we write $\lim_{n\to\omega} a_n = a$ if for eeach neighborhood of $a \in [-\infty, +\infty]$ we have that for ω–a.e. $n \in \mathbb{N}$, a_n is in this neighborhood.*

Lemma 2.21 *For any sequence of numbers $a_k \in [-\infty, +\infty]$ there is unique $l \in [-\infty, +\infty]$ such that $l = \lim_{k\to\omega} a_k$, i.e. for each $\varepsilon > 0$ for ω – a.e. $n \in \mathbb{N}$ $|a_k - l| < \varepsilon$ if $l \in \mathbb{R}$ and $a_k < -\frac{1}{\varepsilon}$ resp. $a_k > \frac{1}{\varepsilon}$ if $l = -\infty$ resp $l = +\infty$. We write then $\lim_{k\to\omega} a_k = l$ and say that l is the ω–limit of a_k. Moreover if $\lim_{k\to\omega} a_k = l_1$, $\lim_{k\to\omega} b_k = l_2$ and $l_1 \in \mathbb{R}$ or $l_2 \in \mathbb{R}$ then $\lim_{k\to\omega} a_k + b_k = l_1 + l_2$ (where $c + \infty = +\infty$ $-\infty + c = -\infty$ for $c \in \mathbb{R}$).*

Proof For uniqueness suppose $l_1 < l_2$ both satisfy $l_1 = \lim_{k\to\omega} a_k$, $l_2 = \lim_{k\to\omega} a_k$. We only treat the case $l_1, l_2 \in \mathbb{R}$ the other cases being analogous. Choose $\varepsilon > 0$ such that $l_1 + \varepsilon < l_2 - \varepsilon$. Then $S_j := \{k \in \mathbb{N} | |a_k - l_j| < \varepsilon\}$ $j = 1, 2$ satisfy $\omega(S_j) = 1$ but $S_1 \cap S_2 = \emptyset$ a contradiction.

For existence define $l := \sup\{s \in \mathbb{R} | \forall \varepsilon > 0\ \omega - -$ a.e. $k \in \mathbb{N}$ satisfy $a_k > s - \varepsilon\}$. Then for any $\varepsilon > 0$ ω, a.e. $n \in \mathbb{N}$ satisfy $a_k > \left(l - \frac{\varepsilon}{2}\right) - \frac{\varepsilon}{2} = l - \varepsilon$. Also $l + \frac{\varepsilon}{2}$ does not satisfy this, which means that there is $\varepsilon > 0$ such that for ω a.e. $n \in \mathbb{N}$ (recall that ω assumes values in $\{0,1\}$) $a_n \leqslant l + \frac{\varepsilon}{2} - \varepsilon'$. Since $\varepsilon' < \frac{\varepsilon}{2}$ must hold as otherwise $l + \frac{\varepsilon}{2} - \varepsilon' < l$, we conclude that l is the ultralimit of $(a_n)_n$.

The linearity of the ultralimit operation is not hard to prove and is left to the reader. ∎

Next we want to define ultralimit of a sequence of pointed metric spaces (X_k, r_k), i.e. X_k is a metric space and $r_k \in X_k$ for $k \in \mathbb{N}$.

Definition 2.22 *Let $(x_k)_{k\in\mathbb{N}}$ and $(y_k)_{k\in\mathbb{N}}$ be sequence such that $x_k, y_k \in X_k$ for each $k \in \mathbb{N}$ and $\sup_{k\in\mathbb{N}} d_{X_k}(x_k, r_k) < +\infty$, $\sup_{k\in\mathbb{N}} d_{X_k}(y_k, r_k) < +\infty$. Define the pseudo–distance*

$$d_\omega((x_k)_k, (y_k)_k) := \lim_{k\to\omega} d_{X_k}(x_k, y_k) \tag{2.42}$$

The linearity of ω–limit of numbers from Lemma 2.22 and the triangle inequality in X_k's yield the triangle inequality for d_ω. This pseudo–distance splits the set of all such sequences into equivalence classes $\sim$, where $(x_k)_k \sim (y_k)_k$ iff $d_\omega((x_k)_k, (y_k)_k) = 0$. The ultralimit of (X_k, μ_k) is the set of these equivalence classes, and we denote it by $X' = \lim_{k\to\omega}(X_k, \mu_k)$; it is clear that this construction defines a metric space. A particular instant of such construction is the ultralimit of a constant sequence $X_k = X$, $r_k = x \in X$, and we call if the ultraproduct of X, denoted $X^\omega = \lim_{k\to\omega} X$. Its elements are thus equivalence classes of bounded sequences in X, and we will be denoting such element by $\dot{x} = [(x_k)]$ or by $\dot{x} = \lim_{k\to\omega} x_k$. It is easy to see that $X \subset X^\omega$. Moreover $X = X^\omega$ holds iff X is a locally compact space(see [1] and [29])

Theorem 2.23 *The ultralimit of any sequence of metric spaces is a complete metric space. The ultralimit of a sequence of intrinsic spaces is a geodesic space. The ultralimit of a sequence of CAT(k) spaces is a CAT(k) space.*

Proof See [6] Chapter I Lemma 5.53, Exercise 5.54 and Chaper II Corollary 3.10(1) respectively. ∎

We also need to define the ultrextension of a proper, l.s.c. and convex functional $\varphi : X \to (-\infty, +\infty]$ when X is a CAT(0) space (but see [1] and [29] for ultralimits of sequences of locally − Lipschits functionals).

Definition 2.24 *Let X be a CAT(0) space, and let $\varphi : X \to (-\infty, +\infty]$ be proper l.s.c. and convex. Then the ultraextension of φ is the mapping*

$$\varphi^\omega : X^\omega \to (-\infty,] + \infty] \qquad \varphi^\omega(x) := \inf\{\lim_{k\to\omega} \varphi(x_k) | (x_k)_k \in \dot{x}\} \tag{2.43}$$

By [30] Lemma 1.3 φ is bounded bellow on each ball hence $\varphi^\omega(\dot{x}) > -\infty$ for each $\dot{x} \in X^\omega$ indeed.

Lemma 2.25 *$\varphi^\omega = \varphi$ on X and φ^ω is geodesicly convex.*

Proof For the first claim pick $x \in X$. Then for any sequence $(x_k)_k$ in the equivalence class of $(x, x, ...)$ we have $\lim_{k\to\omega} d(x_k, x) = 0$ hence for each $\varepsilon > 0$, ω − a.e. $k \in \mathbb{N}$ satisfy $d(x_k, x) < \varepsilon$. As φ is assumed to be l.s.c. for any $\delta > 0$ there is $\varepsilon > 0$ such that $d(y, x) < \varepsilon$ implies $\varphi(y) > \varphi(x) - \delta$. Hence for any $(x_k)_k \in [(x, x, ...)]$ and any $\delta > 0$ we have $\lim_{k\to\omega} \varphi(x_k) \geqslant \varphi(x) - \delta$, and as $\lim_{k\to\omega} \varphi(x) = \varphi(x)$ we must have $\varphi^\omega([x, x, ...]) = \varphi(x)$. For the second claim let $\dot{x^0}, \dot{x^1} \in D(\varphi^\omega)$. Fix $\varepsilon > 0$ and let $(x^0_k)_k \in \dot{x}^0$, $(x^1_k)_k \in \dot{x}^1$ be such that $\varphi^\omega(\dot{x}^j) + \varepsilon > \lim_{k\to\omega} \varphi(x^j_k)$ for $j = 1, 2$. Let moreover for $k \in \mathbb{N}$ $\gamma_k : [0,1] \to X$ be the unique constant speed geodesic joining x^0_k and x^1_k. As $(x^0_k)_k$, $(x^1_k)_k$ are bounded sequences in X, the same holds for $(\gamma_k(t))_{k\in\mathbb{N}}$ for each $t \in [0,1]$, and we have for $0 \leqslant s < t \leqslant 1$.

$$d_\omega([(\gamma_k(s))_k], [(\gamma_k(t))_k] = \lim_{k\to\omega} d(\gamma_k(s), \gamma_k(t))$$

$$= \lim_{k\to\omega} (t-s) d(\gamma_k(0), \gamma_k(1)) = (t-s) d_\omega(\dot{x^0}, \dot{x^1})$$

Hence $t \mapsto [(\gamma_k(t))_k]$ is a geodesic in X^ω and as X^ω is $CAT(0)$ space by Theorem 2.23, it is the unique one. Now we only need to observe that

$$\begin{aligned}\varphi^\omega([(\gamma_k(t))_k]) \leqslant \lim_{k\to\omega} \varphi(\gamma_k(t)) \leqslant &(1-t) \lim_{k\to\omega} \varphi(x^0_k) + t \lim_{k\to\omega} \varphi(x^1_k) \\ \leqslant &(1-t)(\varphi^\omega(\dot{x^0}) + \varepsilon) + t(\varphi^\omega(\dot{x^1}) + \varepsilon)\end{aligned}$$

and conclude since $\varepsilon > 0$ was arbitrarily chosen. ∎

We will also use the following simple fact.

Lemma 2.26 *Let X be a complete metric space and $[(x_k)_k] \in X^\omega$, i.e. $\lim_{k\to\omega} x_k = \dot{x} \in X^\omega$, If for any subsequence $y_r = x_{k_r}$ of $(x_k)_k$ we have $\lim_{r\to\omega} y_r = \dot{x}$ as well, then $\dot{x} \in X$ and $x_k \to x$ in X.*

Proof The sequence $(x_k)_k$ does not converge in X iff it is not Caushy. Suppose that this is the case. Then for some $\varepsilon > 0$ for each $N \in \mathbb{N}$ we can find $k, r \geqslant N$ such that $d(x_k, x_r) \geqslant \varepsilon$ hence we can construct two subsequences $(y^1_r)_r$ and $(y^2_r)_r$ of $(x_k)_k$ such that $d(y^1_r, y^2_r) \geqslant \varepsilon$ for each $r \in \mathbb{N}$. But then $d_\omega([(y^1_r)_r], [(y^2_r)_r]) = \lim_{k\to\omega} d(y^1_r, y^2_r) \geqslant \varepsilon$ which contradicts assumption of the lemma. ∎

3 Construction of approximation semigroups and some convergence theorems

In this section we will first recall the strategy of proving the product formula on a Hilbert space and atempt to clarify it. Than we will point out why it is reasonable to expect that it will work in a CAT(0) space as well. And having done that we will get to work by constructing the tolls that we need in the next section. As we already mentioned in the introduction the results we obtain in this section are of independant interest.

To fix ideas consider two convex, proper, l.s.c functionals φ_1 φ_2 defined on as Hilbert space H, and let us discuss the version of the product formula where both steps are taken by the resolvents, i.e. in Theorem 2.18 we take $F_t := \mathcal{J}^2_t \mathcal{J}^1_t$. Observe the following sequence of facts. First of all for any contraction F on a closed convex subset C of H we can construct the semogroup $\left(S^F_t\right)_{t\geqslant 0}$ of contractions on C whose paths are the solutions of

$$S_t^\rho x = x - \int_0^t \frac{(I-F)}{\rho} S_s^{F,\rho} x dx \quad \text{for } x \in C \tag{3.1}$$

Such semigroups can be constructed by a fixed point argument in a suitable space of curves. Alternatively one can observe that $\frac{I-F}{\rho}$ is a maximal monotone operator operator and use the results of [7]. Next since our nonlinear generator $\frac{F-I}{\rho}$ is Lipschitz and we are in a Hilbert space these solutions can also be obtained by the Euler forward procedure. That is we can solve the discrete version of our equation (3.1) with a uniform time step size τ as follows: for an inital condition x_0 find x_τ such that

$$\frac{x_\tau^\tau - x_0}{\tau} = \frac{F-I}{\rho} x_0$$

which is

$$x_\tau^\tau = x_0 + \frac{\tau}{\rho}(F-I)x_0 \tag{3.2}$$

and define further by induction $x_{k\tau}^\tau := x_{(k-1)\tau}^\tau + \frac{\tau}{\rho}(F-I)x_{(k-1)\tau}^\tau$. Than it is not hard to show that these dectrete solutions converge to the solution of 3.1. We can do this for any contraction F in particular for $F := F_\rho$ for any $\rho \in (0,+\infty)$, and let us write $S_t^\rho := S_t^{\rho,F_\rho}$. Than for $m \in \mathbb{N}$ and $t \in (0,+\infty)$ and the choice $\rho = \tau := t/m$ the above described Euler forward approximation of (3.1) reeds $x_{k\tau}^\tau = (F_{t/m})^k$ and in particular $x_t^{t/m} = (F_{t/m})^m x_0$, i.e. the Trotter approximation of $S_t x_0$. This fact clarifies why we are able to find a good estimate between solutions of (3.1) and the Trotter approximations (see Lemma 3.10 bellow for the precise statement and the proof in our setting).

The classical proof argues further in several steps: one shows that in this situation convergence of the resolvents

$$\mathcal{J}_{\lambda,\rho} := (I + \frac{\lambda}{\rho}(I - U_\rho))^{-1} \quad \rho > 0 \tag{3.3}$$

of the approximation semigroups $(S_t^\rho)_t$ to the resolvent $\mathcal{J}_\lambda$ of $(S_t)_{t\leqslant 0}$ as $\rho \to 0$ for each $\lambda > 0$ pointwise implies convergence of the semigroups, and in light of (3.10) one only has to prove that the resolvents convergence indeed holds (which is done in [23]). Now a carefull reader of the clasical proofs might notice that this part of the argument essentially works due to convexity of the Hilbert norm and the Lipschitz property of U_t. Thus since we have these properties in the CAT(0) setting too, we may hope to be able to do the same trick.[7]

Now that we know what to do, let us construct the approximation semigroups as above but in the CAT(0) setting. We will do so by constructing the resolvents 3.3 first, than the we obtain same type of estimates as in the work of Crandall-Liggett [7], and finally give a mimic of that proof of convergence. Observe

[7] We may however not hope to have a theorem of such generality for gradient flows on spaces with convex metric as even on Banach spaces, already for the exsitance of the flow one needs a lower level sets compactness assumption.

that by construction these approximation semigroups indeed coincide with the classical ones if X is a closed convex subset of a Hilbert space.

If X is a subset of linear space, then identity $(\Big(I+\frac{\lambda}{\rho}(I-F)\Big)^{-1}x=y$ is equivalent to $y=\frac{1}{1+\frac{\lambda}{\rho}}x+\frac{\frac{\lambda}{\rho}}{1+\frac{\lambda}{\rho}}Fy$, that is y is the unique point in X having property to be the point $\gamma(\frac{\lambda/\rho}{1+\lambda/\rho})$ on unit speed geodesic γ connecting a and Fy. Hence we are led to define the mapping

$$G_{\lambda,\rho,x}:X\to X,\quad G_{\lambda,\rho,x}(y):=\frac{1}{1+\frac{\lambda}{\rho}}x\oplus\frac{\frac{\lambda}{\rho}}{1+\frac{\lambda}{\rho}}Fy \tag{3.4}$$

for a contraction $F:X\to X$ (but we omit F in the notation for simplicity).

Lemma 3.1 *Let $F:X\to X$ be a contraction. Than for any $\lambda,\rho>0$ and $x\in X$ the mapping $G_{\lambda,\rho,x}$ is a strict contraction with the Lipschitz constant $l\leqslant\frac{\lambda/\rho}{1+\lambda/\rho}<1$.*

Proof Lemma 2.3 and contractivity of F yield for $y_1,y_2\in X$

$$d(G_{\lambda,\rho,x}(y_1),G_{\lambda,\rho,x}(y_2))\leqslant\frac{\lambda/\rho}{1+\lambda/\rho}d(Fy1,Fy_2)\leqslant\frac{\lambda/\rho}{1+\lambda/\rho}d(y1,y_2)$$

∎

By the Banach fixed point theorem $G_{\lambda,\rho,x}$ has unique fixed point which we denote $\mathcal{J}^F_{\lambda,\rho}x$. Notice that this definition coincides with $\Big(I+\frac{\lambda}{\rho}(I-F)\Big)^{-1}$ when X is subset of a Banach space.

Lemma 3.2 *Let $F:X\to X$ be a contraction. Than for each $\lambda,\rho>0$ the map $\mathcal{J}_{\lambda,\rho}:=\mathcal{J}^F_{\lambda,\rho}:X\to X$ is a contraction.*

Proof Let $z:=\frac{\lambda/\rho}{1+\lambda/\rho}\in(0,1)$ and $x_1,x_2\in X$. Proposition 2.2 gives

$$\begin{aligned}d(\mathcal{J}_{\lambda,\rho}x_1,\mathcal{J}_{\lambda,\rho}x_2)\leqslant&(1-t)d(x_1,x_2)+td(F\mathcal{J}_{\lambda,\rho}x_1,F\mathcal{J}_{\lambda,\rho}x_2)\\ \leqslant&(1-t)d(x_1,x_2)+td(\mathcal{J}_{\lambda,\rho}x_1,\mathcal{J}_{\lambda,\rho}x_2)\end{aligned}$$

hence $d(\mathcal{J}_{\lambda,\rho}x_1,\mathcal{J}_{\lambda,\rho}x_2)\leqslant d(x_1,x_2)$. ∎

Lemma 3.3 *Let $F:X\to X$ be a contraction, and let $\mathcal{J}_{\lambda,\rho}:=\mathcal{J}^F_{\lambda,\rho}$. Than for $\lambda,\rho>0$, $x\in X$ we have*

$$\frac{d(x,\mathcal{J}_{\lambda,\rho})}{\lambda}\leqslant\frac{d(x,Fx)}{\rho} \tag{3.5}$$

Proof Since $\mathcal{J}_{\lambda,\rho}x$ is the unique fixed point of the mapping $G_{\lambda,\rho,x}$, we have that $\lim_{n\to+\infty}(G_{\lambda,\rho,x})^n(x) = \mathcal{J}_{\lambda,\rho}x$. By Lemma 3.1 we estimate

$$\begin{aligned} d(x, \mathcal{J}_{\lambda,\rho}x) &\leqslant \sum_{n=1}^{\infty} d((G_{\lambda,\rho,x})^{n-1}(x), (G_{\lambda,\rho,x})^n(x)) \\ &\leqslant \sum_{n=1}^{\infty} \left(\frac{\lambda/\rho}{1+\lambda/\rho}\right)^{n-1} d(x, G_{\lambda,\rho,x}(x)) \\ &\leqslant \frac{1}{1-\frac{\lambda/\rho}{1+\lambda/\rho}} d(x, G_{\lambda,\rho,x}(x)) = \frac{\lambda}{\rho} d(x, Fx)) \end{aligned}$$

since by definition $d(x, G_{\lambda,\rho,x}(x)) = \frac{\lambda/\rho}{1+\lambda/\rho} d(x, Fx)$. ∎

Lemma 3.4 *(The Resolvent Identity) Let $F : X \to X$ be a contraction, and let $\mathcal{J}_{\lambda,\rho} := \mathcal{J}^F_{\lambda,\rho}$. Than for each $x \in X$, $0 < \mu < \lambda$, $\rho > 0$ the Resolvent Identity holds:*

$$\mathcal{J}_{\lambda,\rho}x = \mathcal{J}_{\mu,\rho} \frac{\mu}{\lambda} x \oplus \frac{\lambda-\mu}{\lambda} \mathcal{J}_{\lambda,\rho}x \tag{3.6}$$

Proof By construction $\mathcal{J}_{\mu,\rho}y$ is the unique point $x \in X$ that has the property that

$$d(y, z) = \frac{\mu/\rho}{1+\mu/\rho} d(y, Fz) \tag{3.7}$$

Hence it is enough to show that (3.4) holds for $z \in \mathcal{J}_{\lambda,\rho}x$. To this aim set $a := d(x, F\mathcal{J}_{\lambda,\rho}x)$, $b := d(\mathcal{J}_{\lambda,\rho}x, F\mathcal{J}_{\lambda,\rho}x) = \frac{1}{1+\lambda/\rho}a = \frac{\rho}{\rho+\lambda}a$, $c := d(y, \mathcal{J}_{\lambda,\rho}x) = (a-b)\frac{\mu}{\lambda} = \frac{\lambda}{\rho+\lambda} \cdot \frac{\mu}{\lambda} a = \frac{\mu}{\rho+\lambda} a$, and compute

$$\frac{d(y, \mathcal{J}_{\lambda,\rho}x)}{d(y, F\mathcal{J}_{\lambda,\rho}x)} = \frac{c}{b+c} = \frac{\frac{\mu}{\rho+\lambda}a}{\frac{\rho+\mu}{\rho+\lambda}a} = \frac{\mu}{\rho+\mu}.$$

∎

Lemma 3.5 *Let $F : X \to X$ be a contraction, and let $\mathcal{J}_{\lambda,\rho} := \mathcal{J}^F_{\lambda,\rho}$. Than for $\lambda, \rho > 0$, $x \in X$ and $n \in \mathbb{N}$*

$$d(\mathcal{J}^n_{\lambda,\rho}x, x) \leqslant \frac{\lambda}{\rho} d(x, Fx) \tag{3.8}$$

Proof Follows by triangle inequality , Lemma 3.2 and Lemma 3.3. ∎

Lemma 3.6 *Let $F : X \to X$ be a contraction, and let $\mathcal{J}_{\lambda,\rho} := \mathcal{J}^F_{\lambda,\rho}$. than for $\rho > 0$, $0 < \mu \leqslant \lambda$ and $n, m \in \mathbb{N}$, $n \geqslant m$ we have*

$$\begin{aligned} d(\mathcal{J}^n_{\mu,\rho}x, \mathcal{J}^m_{\lambda,\rho}x) \leqslant & \sum_{j=1}^{m-1} \alpha^j \beta^{n-1} B(n,j) d(\mathcal{J}^{m-j}_{\lambda,\rho}x, x) + \\ & + \sum_{j=m}^{n} \alpha^m \beta^{j-m} B(j-1, m-1) d(\mathcal{J}^{m-j}_{\mu,\rho}x, x) \end{aligned} \tag{3.9}$$

where $\alpha = \frac{\mu}{\lambda}$, $\beta = \frac{\lambda-\mu}{\lambda}$ and $B(\cdot,\cdot)$ are the binomial coefficients.

Proof Fix ρ, μ, λ, x as above and let $n \geqslant m$. For integers j and k such that $0 \leqslant j \leqslant n$, $0 \leqslant k \leqslant m$ we define $a_{k,j} := d\left(\mathcal{J}^j_{\mu,\rho}x, \mathcal{J}^k_{\lambda,\rho}x\right)$. With the aid of Lemma 3.3 and Lemma 3.4 and Lemma 2.3 we estimate

$$\begin{aligned} a_{k,j} &= d\left(\mathcal{J}^j_{\mu,\rho}x, \mathcal{J}_{\mu,\rho}x\left(\frac{\mu}{\lambda}\mathcal{J}^{k-1}_{\lambda,\rho}x \oplus \frac{\lambda-\mu}{\lambda}\mathcal{J}^k_{\lambda,\rho}x\right)\right) \\ &\leqslant d\left(\mathcal{J}^{j-1}_{\mu,\rho}x, \frac{\mu}{\lambda}\mathcal{J}^{k-1}_{\lambda,\rho}x \oplus \frac{\lambda-\mu}{\mu}\mathcal{J}^k_{\lambda,\rho}x\right) \\ &\leqslant \frac{\mu}{\lambda} d\left(\mathcal{J}^{j-1}_{\mu,\rho}x, \mathcal{J}^{k-1}_{\lambda,\rho}x\right) + \frac{\lambda-\mu}{\mu} d\left(\mathcal{J}^{j-1}_{\mu,\rho}x, \mathcal{J}^k_{\lambda,\rho}x\right) \\ &= \alpha a_{k-1,j} + \beta a_{k,j-1} \end{aligned}$$

This is precisely the same estimate as in [7] Lemma 1.3 and we conclude the the claim. ∎

Lemma 3.7 *Let $F : X \to X$ be a contraction, and let $\mathcal{J}_{\lambda,\rho} := \mathcal{J}^F_{\lambda,\rho}$. Than for $\rho > 0$, $< \mu \leqslant \lambda$, $n, m \in \mathbb{N}$, $n < m$ we have*

$$d\left(\mathcal{J}^n_{\mu,\rho}x, \mathcal{J}^m_{\lambda,\rho}x\right) \leqslant \sum_{j=0}^{n} \alpha^j \beta^{n-j} B(n,j) d\left(\mathcal{J}^{m-j}_{\lambda}x, x\right) \tag{3.10}$$

Proof As in previous lemma we obtain precisely the same estimates as in [7] hence we may conclude our claim. ∎

The following Lemma is from [7] and we state it for the completeness.

Lemma 3.8 *Let $n \geqslant m > 0$ be integers, and α and β positive numbers adding to 1. Then*

1. $\sum_{j=0}^{n} B(n,j)\alpha^j\beta^{n-j}(m-j) \leqslant \sqrt{(n\alpha - m)^2 + n\alpha\beta}$
2. $\sum_{j=m}^{n} B(j-1,m-1)\alpha^m\beta^{j-m}(m-j) \leqslant \sqrt{\frac{m\beta}{\alpha^2} + \left(\frac{m\beta}{\alpha} + m - n\right)^2}$

Proof See [7] Lemma 1.4. ∎

Now we are able to show the Crandall-Liggett theorem in our setting.

Theorem 3.9 *Let X be a complete CAT(0) space and let $F : X \to X$ be a contraction. Then denoting $\mathcal{J}_{\lambda,\rho} := \mathcal{J}^F_{\lambda,\rho}$ and $(S^\rho_t)_{t\geqslant 0} := \left(S^{\rho,F}_t\right)_{t\geqslant 0}$ we have that for each $\rho > 0$ the exponential formula holds:*

$$S^\rho_t x := \lim_n \left(\mathcal{J}_{\frac{t}{n},\rho}\right)^n x \tag{3.11}$$

exists for all $t > 0$ and all $x \in X$. Moreover $(S^\rho_t)_{t\geqslant 0}$ is a c_0–contraction semigroup on X with globally Lipschitz paths, and we have the estimate

$$d\left(S^\rho_t x, S^\rho_s x\right) \leqslant 2\frac{d(x,Fx)}{\rho}|t-s| \tag{3.12}$$

for $x \in X$, $t, s \geqslant 0$.

Proof For $x \in X$, $\rho > 0$ and abbreviate $\mathcal{J}_\lambda := \mathcal{J}_{\lambda,\rho}$ throughout this proof. Let $\lambda \geqslant \mu > 0$, and let $n, m \in \mathbb{N}$ with $n \geqslant m$. Using Lemmas 3.6, 3.5 and 3.8 we estimate

$$\begin{aligned} d\left(\mathcal{J}_\mu^n x, \mathcal{J}_\lambda^m x\right) \leqslant & \sum_{j=0}^{m-1} \alpha^j \beta^{n-j} \mathrm{B}(n,j) d\left(\mathcal{J}_\lambda^{m-j} x, x\right) \\ & + \sum_{j=m}^{n} \alpha^m \beta^{j-m} \mathrm{B}(j-1, m-1) d\left(\mathcal{J}_\mu^{n-1} x, x\right) \\ \leqslant & \lambda \sum_{j=0}^{m-1} \alpha^j \beta^{n-j} \mathrm{B}(n,j)(m-j) \frac{d(x,Fx)}{\rho} + \\ & + \mu \sum_{j=m}^{n} \alpha^m \beta^{j-m} \mathrm{B}(j-1,m-1)(m-j) \frac{d(x,Fx)}{\rho} \qquad (3.13) \\ \leqslant & [\lambda \sqrt{\left(n\frac{\mu}{\lambda} - m\right)^2 + n\frac{\mu}{\lambda}\frac{\lambda-\mu}{\lambda}} \\ & + \sqrt{\frac{\lambda}{\mu^2}\frac{\lambda-\mu}{\lambda} m + \left(\frac{\lambda}{\mu}\frac{\lambda-\mu}{\mu} m + m - n\right)^2}\,] \frac{d(x,Fx)}{\rho} \\ & = [\sqrt{(n\mu - \lambda m)^2 + n\mu(\lambda-\mu)} \\ & + \sqrt{m\lambda(\lambda-\mu) + (m\lambda - n\mu)^2}\,] \frac{d(x,Fx)}{\rho} \end{aligned}$$

For $\mu := \frac{t}{n}$, $\lambda := \frac{t}{m}$ is the inequality above reeds

$$d\left(\mathcal{J}_{t/m}^m x, \mathcal{J}_{t/n}^n x\right) \leqslant 2t \sqrt{\left|\frac{1}{m} - \frac{1}{n}\right|} \frac{d(x,Fx)}{\rho} \qquad (3.14)$$

Hence by completeness the limit

$$S_t^\rho x := \lim_n \left(\mathcal{J}_{t/n}\right)^n x \qquad (3.15)$$

exists for $t > 0$, $x \in X$, and we define S_0^ρ to be the identity map on X. As $\mathcal{J}_{t/n} = \mathcal{J}_{t/n,\rho}$ is a contraction for all $t > 0$ and $n \in \mathbb{N}$, $(S_t^\rho)_{t \geqslant 0}$ is a family of contractions. Moreover choosing $\mu := t/n$, $\lambda := s/n$ in (3.13) and passing to the limit gives the estimate (3.12) of the theorem, and the global Lipschitz property of the paths.

Next we argue the semigroup property. To this aim let $m \in \mathbb{N}$, $t > 0$ and $x \in X$. Since $\mathcal{J}_{t/n}^n = \mathcal{J}_{t/n,\rho}^n$ is contraction and (3.13) holds

$$\begin{aligned} d\left(\left(S_t^\rho\right)^m x, \left(\mathcal{J}_{t/n}^n\right)^m x\right) \leqslant & d\left(S_t^\rho \left(S_t^\rho\right)^{m-1} x, \mathcal{J}_{t/n}^n \left(S_t^\rho\right)^{m-1} x\right) + \\ & + d\left(\mathcal{J}_{t/n}^n \left(S_t^\rho\right)^{m-1} x, \mathcal{J}_{t/n}^n \left(\mathcal{J}_{t/n}^{n(m-1)}\right) x\right) \to 0 \end{aligned}$$

as $n \to \infty$. Hence we have

$$(S_t^\rho)^m x := \lim_n \left(\mathcal{J}_{t/n}^n\right)^m x = \lim_n \mathcal{J}_{\frac{mt}{nm}}^{nm} x = S_{mt}^\rho x \tag{3.16}$$

Next for $l, k, r, m \in \mathbb{N}$ by (3.13)

$$S^\rho_{\frac{l}{k}+\frac{n}{m}} = S^\rho_{\frac{lm+nk}{km}} = \left(S^\rho_{\frac{1}{km}}\right)^{lm+nk}$$

$$= \left(S^\rho_{\frac{1}{km}}\right)^{lm}\left(S^\rho_{\frac{1}{km}}\right)^{nk} = S^\rho_{\frac{lm}{km}} S^\rho_{\frac{nk}{km}} = S^\rho_{l/k} S^\rho_{r/m},$$

i.e. $S^\rho_{t+s} = S^\rho_t S^\rho_s$ for $s, t \in \mathbb{Q}_+$. Finally for arbitrary $s, t > 0$ take two sequences of positive rational numbers $s_n \to s$, $t_n \to t$-we just showed that $S_{t_n} S s_n x = S_{t_n+s_n} x$. Observe moreover that (3.12) implies $S^\rho_{t_n+s_n} \to S^\rho_{t+s}$ for $x \in X$. We may conclude the theat the semigroup property holds by the following estimate:

$$\begin{aligned} d(S^\rho_{t_n} S^\rho_{s_n} x, S^\rho_t S^\rho_s x) \leqslant & d(S^\rho_{t_n} S^\rho_{s_n} x, S^\rho_{t_n} S^\rho_s x) + d(S^\rho_{t_n} S^\rho_s x, S^\rho_t S^\rho_s x) \\ \leqslant & d(S^\rho_{s_n} x, S^\rho_s x) + d(S^\rho_{t_n} S^\rho_s x, S^\rho_t S^\rho_s x) \to 0 \text{ as } n \to +\infty \end{aligned}$$

∎

Remark As it is obvious from (3.1 in the Hilbert these semigroups space case have derivative $\frac{d}{dt} S^\rho_t x = \frac{F-I}{\rho} S^\rho_t x$ and as $\frac{I-F}{\rho}$ is a maximal monotone operator the metric derivative of eeach path is decreasing in time, hence the Lipschitz constant of each path is just $\frac{d(x,Fx)}{\rho}$. We however use the resolvent identity to construct these semigroups, which means that we used the convexity of the metric in our estimates, instead of using the stronger (-1) convexity of its square. Therefore the worse constant 2 at the right side of (3.12) which is immaterial in our considerations. It should be possable to improve this estimate though.

We are going to need the following estimate. As in the previous theorem, we have a constant 2 at the right side of both inequalities bellow (see [4] Theoreme 1.7 and [31] for the linear version of this estimate):

Lemma 3.10 *Let $F : X \to X$ be a contraction and let $(S^\rho_t)_{t\geqslant 0}$ the semigroup constructed in Theorem 3.9 for $\rho > 0$. Then for $t > 0$, $x \in X$ we have $S^\rho_t x = S^1_{t/\rho} x$, and for $m \in \mathbb{N}$*

1. $d(S^1_t x, F^m x) \leqslant 2\sqrt{(m-t)^2 + t}\, d(x, Fx)$
2. $d(S^\rho_t x, F^m x) \leqslant 2\sqrt{\left(m - \frac{t}{\rho}\right)^2 + \frac{t}{\rho}}\, d(x, Fx)$.

Proof The last claim is a direct consequence of the first two.

The first claim follows by the simple observation that for all $\rho, \lambda > 0$, $\mathcal{J}_{\lambda,\rho} = \mathcal{J}_{\frac{\lambda}{\rho},1}$, hence for $x \in X$

$$S^\rho_t x \overset{n\to\infty}{\longleftarrow} \mathcal{J}^n_{\frac{t}{n},\rho} x = \mathcal{J}^n_{\frac{t/\rho}{n},1} x \overset{n\to\infty}{\longrightarrow} S_1\left(\frac{t}{\rho}\right) x$$

For the second claim we need to make more efort. Recall that for each $y \in X$ $\mathcal{J}_\lambda x := \mathcal{J}_{\lambda,1} x = \frac{1}{1+\lambda} y \oplus \frac{\lambda}{1+\lambda} F \mathcal{J}_{\lambda,1} y$ hence by (2.2)

$$d^2(F^m x, \mathcal{J}_{t/n}^n x) \leqslant \frac{1}{1+\frac{t}{n}} d^2(F^m x, \mathcal{J}_{\frac{t}{n}}^{n-1} x) + \frac{\frac{t}{n}}{1+\frac{t}{n}} d^2(F^m x, F \mathcal{J}_{\frac{t}{n}}^{n-1} x)$$

Repeating the argument n times we obtain

$$\begin{aligned} d^2(F^m x, \mathcal{J}_{\frac{t}{n}}^n x) &\leqslant \left(1+\frac{t}{n}\right)^{-n} d^2(F^m x, x) \\ &+ \frac{t}{n} \sum_{k=1}^{n} \left(1+\frac{t}{n}\right)^{-k} d^2(F^m x, F \mathcal{J}_{\frac{t}{n}}^{n-k+1} x) \\ &\leqslant \left(1+\frac{t}{n}\right)^{-n} m^2 d^2(Fx, x) \\ &+ \frac{t}{n} \sum_{k=1}^{n} \left(1+\frac{t}{n}\right)^{-k} d^2(F^{m-1} x, \mathcal{J}_{\frac{t}{n}}^{n-k+1} x) \\ &= \left(1+\frac{t}{n}\right)^{-n} m^2 d^2(x, Fx) \\ &+ \frac{t}{n} \sum_{k=1}^{n} \left(1+\frac{t}{n}\right)^{-(n-k+1)} d^2(F^{m-1} x, \mathcal{J}_{\frac{t}{n}}^{k} x) \end{aligned}$$

For $n \in \mathbb{N}$ define functions

$$f_n(s) := \sum_{k=1}^{n} \left(1+\frac{t}{n}\right)^{-(n-k+1)} 1_{\left(\frac{(k-1)t}{n}, \frac{kt}{n}\right]}(s)$$

$$g_n(s) := \sum_{k=1}^{n} d^2(F^{m-1} x, \mathcal{J}_{\frac{t}{n}}^{k} x) 1_{\left(\frac{(k-1)t}{n}, \frac{kt}{n}\right]}(s)$$

With this notation the above inequality becomes

$$d^2(F^m x, \mathcal{J}_{\frac{t}{n}}^n x) \leqslant \left(1+\frac{t}{n}\right)^{-n} m^2 d^2(x, Fx) + \int_0^t f_n(s) g_n(s) ds$$

We are going to show that $f_n(s) \to e^{-(t-s)}$ and $g_n(s) \to d^2(F^{m-1} x, S_s x)$ for $s \in (0, t]$ and that $\sup_{n\in\mathbb{N}, s\in[0,t)} |g_n(s)||f_n(s)| < \infty$. Than by the dominated convergence theorem we will have

$$d^2(F^m x, S_t x) \leqslant e^{-t} m^2 d^2(x, Fx) + \int_0^t e^{-(t-1)} d^2(F^{m-1} x, S_s x) ds \qquad (3.17)$$

(we abbreviate $S_s := S_s^1$). Let us establish these facts.

Pick $n \in \mathbb{N}$ and $s \in (0,t]$. There is a unique $0 < k_{s,n} \leqslant n$ such that $\frac{(k_{s,n}-1)t}{n} < s \leqslant \frac{k_{s,n}t}{n}$.

Filling in $k_{s,n}$,$\frac{t}{n}$,m,$\frac{s}{m}$ for n,μ,m,λ respectively in (3.13) gives

$$d\left(\mathcal{J}_{\frac{t}{n}}^{k_{s,n}}x, \mathcal{J}_{\frac{s}{m}}^{m}x\right) \leqslant (\sqrt{\left(\frac{k_{s,n}t}{n} - s\right)^2 + \frac{k_{s,n}t}{n}\left(\frac{s}{m} - \frac{t}{n}\right)} + \\ + \sqrt{s\left(\frac{s}{m} - \frac{t}{n}\right) + \left(\frac{k_{s,n}t}{n} - s\right)^2})\frac{2d\left(x, Fx\right)}{\rho} \quad (3.18)$$

and taking limit for $m \to \infty$ gives

$$d\left(\mathcal{J}_{\frac{t}{n}}^{k_{s,n}}x, S_s x\right) \leqslant (\sqrt{(\frac{k_{s,n}t}{n} - s)^2 + \frac{k_{s,n}t}{n}\frac{t}{n}} + \sqrt{s\frac{t}{n} + (\frac{k_{s,n}t}{n} - s)^2}) \quad \frac{2d(x, Fx)}{\rho} \quad (3.19)$$

In light of (3.14) it is not hard to see that $\left\{\mathcal{J}_{\frac{t}{n}}^{k}x | n \in \mathbb{N}\ 0 \leqslant k \leqslant n\right\}$ is a bounded subset of X, hence identity

$$\left|g_n(s) - d^2(F^{m-1}x, S(s)x)\right| = \left|\left(d\left(F^{m-1}x, \mathcal{J}_{\frac{t}{n}}^{k_{s,n}}x\right) + d\left(\mathcal{J}_{\frac{t}{n}}^{k_{s,n}}, S\right)\right)\right. \\ \left. \cdot \left(d\left(F^{m-1}x, \mathcal{J}_{\frac{t}{n}}^{k_{s,n}}x\right) - d\left(F^{m-1}x, S(s)x\right)\right)\right|$$

triangle inequality and (3.19) now easily yield that

$$g_n(s) \to d^2(F^{m-1}x, S(s)x)$$

for $s \in (0,t]$ as well as that g_n is a uniformly bounded sequence of functions. The claims concerning f_n are easy to handle, and we omit the details. Hence we established (3.17). To complete the proof set

$$\varphi_m(s) := \frac{d(F^m x, S_s x)}{2d(x, Fx)} \quad \text{for} \quad m \geqslant 0 \quad \text{and} \quad s \geqslant 0 \quad (3.20)$$

so that (3.17) reeds

$$\varphi_m^2(t) \leqslant e^{-t}m^2 + \int_0^t e^{-|t-s|}\varphi_{m-1}^2(s)ds$$

for $m \geqslant 1$ (F^0 being identity map by definition). We have to show that $\varphi_m(t) \leqslant ((m-t)^2 + t)$ and we will do that by induction.

Well $\varphi_0 \leqslant t$ is just (3.12).

Suppose the claim is true for $m-1$. Than by (3)

$$\varphi_m^2(t) \leqslant e^{-t}m^2 + \int_0^t e^{-(t-s)}((m-1-s)^2 + s)ds$$

hence showing

$$m^2 + \int_0^t e^s((m-1-s)^2 + s)ds = e^t((m-t)^2 + t).$$

sufficies. Well at $t = 0$ (3 holds (i.e. $m^2 = m^2$) and the derivatives of the two functions concide as well:

$$e^t((m-1-t)^2+t) = e^t((m-t)^2+t) - 2(m-t)e^t + e^t = \frac{d}{dt}e^t((m-t)^2+t)$$

Now the proof is complete. ∎

Next we consider a finite sequence $\varphi_1, ..., \varphi_k$ of geodesically convex l.s.c functionals $X \to (-\infty, +\infty]$ such that $\varphi := \sum_{j=1}^k \varphi_j \neq +\infty$. Than Assumption 2.11 holds for each of these functuionals, i.e. each φ_j generates a contraction semigroup $(S_t^j)_{t\geqslant 0}$ on $\overline{D(\varphi_j)}$. Moreover $X_j = \overline{D(\varphi_j)}$ is a closed geodesically convex subset of X and we denote $P_j : X \to X_j$ to be the metric projection onto X_j as in Proposition 2.8. Then for $t \geqslant 0$, $j = 1, 2, ..., k$ and $x, y \in X$

$$d\left(S_t^j P_j x, S_t^j P_j y\right) \leqslant d(P_j x, P_j y) \leqslant d(x, y)$$

Hence the mapping

$$F_t^1 : X \to X \quad F_t := S_t^k P_k \circ \cdots \circ S_t^1 P_1 \tag{3.21}$$

is a contraction for each $t \geqslant 0$.

Further denoting $\mathcal{J}_t^j x := \text{argmin}_{y\in X} \frac{1}{2t} d^2(x,y) + \varphi(y)$ the resolvent of φ^j, we define for $t > 0$

$$F_t^2 : X \to X \quad F_t^2 x := \mathcal{J}_t^k \circ \cdots \circ \mathcal{J}_t^1 x \tag{3.22}$$

which is also a contraction.

Theorem 3.11 *Let either $F_\rho = F_\rho^1$ for each $\rho > 0$ or $F_\rho = F_\rho^2$ for each $\rho > 0$. Let moreover for $\rho > 0$ $(S_t^\rho)_{t\geqslant 0}$ be the semigroup constructed in Theorem 3.9 for the choice $F = F_\rho$, and let $\mathcal{J}_{\lambda,\rho}$ be its resolvents. Then denoting $\mathcal{J}_\lambda$ to be the resolvents of the sum functional $\varphi := \sum_1^n \varphi_j$ convergence of the resolvents*

$$\mathcal{J}_{\lambda,\rho} x \to \mathcal{J}_\lambda x \quad \textit{as } \rho \downarrow 0 \tag{3.23}$$

for $x \in \overline{D(\varphi)}$, implies

$$S_t^\rho x \to S_t x \quad \textit{for} \quad x \in \overline{D(\varphi)} \tag{3.24}$$

and the limit is uniform on compact time intervals.

Proof Fix $T > 0$ $x \in D(|\partial\varphi|)$, and $0 < t \leqslant T$. For each $\lambda > 0$ our assumption implies

$$\frac{d(x, \mathcal{J}_{\lambda,\rho})}{\lambda} \longrightarrow \frac{d(x, \mathcal{J}_\lambda x)}{\lambda} \leqslant |\partial\varphi|(x) \tag{3.25}$$

as $\rho \downarrow 0$ where the inequality holds by by (2.30). Next we have the following estimates:

$$\begin{aligned} d(S_t^\rho x, S_t x) \leqslant & d(S_t^\rho x, S_t^\rho \mathcal{J}_{\lambda,\rho} x) + d(S_t^\rho \mathcal{J}_{\lambda,\rho} x, S_t x) \\ \leqslant & d(x, \mathcal{J}_{\lambda,\rho} x) + d(S_t^\rho \mathcal{J}_{\lambda,\rho} x, S_t x) \end{aligned} \tag{3.26}$$

and

$$
\begin{aligned}
d(S_t^\rho \mathcal{J}_{\lambda,\rho}x, S_t x) \leqslant & d(S_t^\rho \mathcal{J}_{\lambda,\rho}x, (\mathcal{J}_{t/n,\rho})^n \mathcal{J}_{\lambda,\rho}x) \\
& + d((\mathcal{J}_{t/n,\rho})^n \mathcal{J}_{\lambda,\rho}x, (\mathcal{J}_{t/n,\rho})^n) \\
& + d((\mathcal{J}_{t/n,\rho})^n x, (\mathcal{J}_{t/n})^n) + d((\mathcal{J}_{t/n})^n x, S_t x)
\end{aligned} \tag{3.27}
$$

We need to find upper bounds for the four expressions appearing in (3.27). Firstly we have

$$
d(S_t^\rho \mathcal{J}_{\lambda,\rho}x, (\mathcal{J}_{t/n,\rho}x) \leqslant \frac{2t}{\sqrt{n}} d(\mathcal{J}_{\lambda,\rho}x, F_\rho \mathcal{J}_{\lambda,\rho}x) = \frac{2t}{\sqrt{n}} \frac{\rho}{\lambda} d(x, \mathcal{J}_{\lambda,\rho}x) \tag{3.28}
$$

since by construction $\mathcal{J}_{\lambda,\rho}x = \frac{1}{1+\frac{\lambda}{\rho}} x \oplus \frac{\frac{\lambda}{\rho}}{1+\frac{\lambda}{\rho}} F_\rho \mathcal{J}_{\lambda,\rho}x$ hence $d(x, \mathcal{J}_{\lambda,\rho}x) = \frac{\lambda}{\rho} d(\mathcal{J}_{\lambda,\rho}x, F_\rho \mathcal{J}_{\lambda,\rho}x)$. As $\mathcal{J}_{\lambda,\rho}$ is also a contraction, we have for $n \in \mathbb{N}$

$$
d((\mathcal{J}_{\frac{t}{n},\rho})^n \mathcal{J}_{\lambda,\rho}x, (\mathcal{J}_{\frac{t}{n},\rho})^n x) \leqslant d(\mathcal{J}_{\lambda,\rho}x, x) \tag{3.29}
$$

and by estimate (2.36) we have

$$
d\left(\left(\mathcal{J}_{\frac{t}{n}}\right)^n x, S_t x\right) \leqslant \frac{t}{\sqrt{2n}} |\partial\varphi|(x) \tag{3.30}
$$

Well now for this arbitrary $x \in D(|\partial\varphi|)$ $n \in \mathbb{N}$ $t \leqslant T$ and $\lambda > 0$ we have obtained the following estimate

$$
\begin{aligned}
d\left(S_t^\rho x, S_t x\right) \leqslant & 2d(x, \mathcal{J}_{\lambda,\rho}x) + \frac{2t}{\sqrt{n}} d(x, \mathcal{J}_{\lambda,\rho}x) \\
& + d\left(\left(\mathcal{J}_{\frac{t}{n},\rho}\right)^n x, \left(\mathcal{J}_{\frac{t}{n}}\right)^n x\right) + \frac{t}{\sqrt{2n}} |\partial\varphi|(x)
\end{aligned} \tag{3.31}
$$

To show convergence for $x \in D(|\partial\varphi|)$ let $\varepsilon > 0$, and choose $\lambda_0 > 0$ so that $\lambda_0 |\partial\varphi|(x) < \varepsilon$.

By (3.25) there is a $\delta > 0$ such that for $\rho < \delta$

$$
\frac{d\left(x, \mathcal{J}_{\lambda,\rho}x\right)}{\lambda_0} \leqslant \frac{d\left(x, \mathcal{J}_{\lambda_0}x\right)}{\lambda_0} + \varepsilon
$$

hence

$$
d\left(x, \mathcal{J}_{\lambda,\rho}x\right) \leqslant \lambda_0 |\partial\varphi|(x) + \varepsilon\lambda_0 < 2\varepsilon \quad \text{for } \rho < \delta \tag{3.32}
$$

Next fix $n_0 \in \mathbb{N}$ such that

$$
\frac{t}{\sqrt{2n_0}} |\partial\varphi|(x) \leqslant \frac{T}{\sqrt{2n_0}} |\partial\varphi|(x) < \varepsilon \tag{3.33}
$$

Then there is a $\delta_0' < \delta$ such that for $\rho < \delta_0'$

$$
\frac{2t}{\sqrt{n_0}} \frac{\rho}{\lambda_0} d\left(x, \mathcal{J}_{\lambda,\rho}x\right) \leqslant \frac{2T}{\sqrt{n_0}} \frac{\rho_0}{\lambda_0} 2\varepsilon < \varepsilon \tag{3.34}
$$

Finally we estimate for $t \leqslant T$

$$d\left(\left(\mathcal{J}_{t/n_0,\rho}\right)^n x, \left(\mathcal{J}_{t/n_0}\right)^n x\right) \leqslant d\left(\left(\mathcal{J}_{t/n_0,\rho}\right)^{n_0} x, \left(\mathcal{J}_{t/n_0}\right)^{n_0-1} \mathcal{J}_{t/n_0}x\right) + d\left(\left(\mathcal{J}_{t/n_0,\rho}\right)^{n_0-1} \mathcal{J}_{t/n_0}x, \left(\mathcal{J}_{t/n_0}\right)^{n_0-1} \mathcal{J}_{t/n_0}x\right) \tag{3.35}$$

hence the assumption (3.23) and induction give existence of $\delta_0 \leqslant \delta_0' < \delta$ such that for $\rho < \delta_0$

$$d\left(\left(\mathcal{J}_{t/n_0,\rho}\right)^{n_0} x, \left(\mathcal{J}_{t/n_0}\right)^{n_0} x\right) < \varepsilon \tag{3.36}$$

Now (3.31), (3.33), (3.34), (3.35) and (3.36) give

$$\lim_{\rho\to 0} S_t^\rho x = S_t x \text{ for } t \geqslant 0 \text{ and } x \in D(|\partial\varphi|) \tag{3.37}$$

For uniform convergence $S_t^\rho x \to S_t x$ for $t \leqslant T$ and $x \in D(|\partial\varphi|)$, pick $\tau \in (0,T)$ and estimate

$$\begin{aligned} d(S_t x, S_t^\rho x) \leqslant & d\left(S_t x, S_t^\rho \mathcal{J}_{\lambda,\rho}x\right) + d\left(S_t^\rho \mathcal{J}_{\lambda,\rho}x, S_t^\rho x\right) \\ \leqslant & d\left(S_t x, S_\tau x\right) + d\left(S_\tau x, S_\tau^\rho \mathcal{J}_{\lambda,\rho}x\right) + \\ & d\left(S_\tau^\rho \mathcal{J}_{\lambda,\rho}x, S_t^\rho x\right) + d\left(\mathcal{J}_{\lambda,\rho}x, x\right) \\ \leqslant & |t-\tau||\partial\varphi|(x) + d\left(S_\tau x, S_\tau^\rho \mathcal{J}_{\lambda,\rho}x\right) + \\ & + |t-\tau|\frac{d(\mathcal{J}_{\lambda,\rho}x, F_\rho \mathcal{J}_{\lambda,\rho}x)}{\rho} + d(\mathcal{J}_{\lambda,\rho}x, x) \end{aligned} \tag{3.38}$$

Now observe the following facts:for small $\lambda > 0$ we have $d(\mathcal{J}_{\lambda,\rho}x, x) \leqslant \lambda|\partial\varphi|(x) + \varepsilon$ whereε is arbitrarily small, further $\frac{d(\mathcal{J}_{\lambda,\rho}x, F_\rho \mathcal{J}_{\lambda,\rho}x)}{\rho} = \frac{d(x, \mathcal{J}_{\lambda,\rho}x)}{\lambda}$ and for fixed λ this term is bounded in ρ, and finally we already argued that for any $\tau \leqslant T$ quantity $d(S_t x, S_t^\rho \mathcal{J}_{\lambda,\rho}x)$ is small for small ρ. Well now we can finish the argument by compactness of $[0,T]$ and conclude that for $x \in D(|\partial\varphi|)$ convergence is uniform on $[0,T]$.

At last for $x \in \overline{D(\varphi)} = \overline{D(|\partial\varphi|)}$ pick $\varepsilon > 0$, and $x \in D(|\partial\varphi|)$ so that $d(x,y) < \varepsilon$ and estimate

$$\begin{aligned} d\left(S_t^\rho x, S_t x\right) \leqslant & d\left(S_t^\rho x, S_t^\rho y\right) + d\left(S_t^\rho y, S_t y\right) + d\left(S_t y, S_t x\right) \\ \leqslant & 2\varepsilon + d\left(S_t^\rho y, S_t y\right) \end{aligned} \tag{3.39}$$

The proof is now completed. ∎

We conclude this section with the folowing theorem.

Theorem 3.12 *Under the same assumptions as in Theorem 3.11 the extension of Trotter product formula holds:*

$$\left(F_{\frac{t}{n}}\right)^n x \longrightarrow S_t x \tag{3.40}$$

and the limit is uniform on compact time intervals.

Proof Assume first that $x \in D(|\partial\varphi|)$. Fix $T > 0$ and let $t \leqslant T$. We estimate

$$d\left(S_t x, \left(F_{\frac{t}{n}}\right)^n x\right) \leqslant d\left(S_t x, S_t^{\frac{t}{n}} x\right) + d\left(S_t^{\frac{t}{n}} x, \left(F_{\frac{t}{n}}\right)^n x\right) \tag{3.41}$$

and handle the two terms separately. For $\rho > 0$, $n \in \mathbb{N}$ and $\lambda > 0$ we have

$$\begin{aligned} d\left(S^\rho_{n\rho} x, (F_\rho)^n x\right) \leqslant & d\left(S^\rho_{n\rho} x, S^\rho_{n\rho} \mathcal{J}_{\lambda,\rho} x\right) + \\ & + d\left(S^\rho_{n\rho} \mathcal{J}_{\lambda,\rho} x, (F_\rho)^n \mathcal{J}_{\lambda,\rho} x\right) + d\left((F_\rho)^n \mathcal{J}_{\lambda,\rho} x, (F_\rho)^n x\right) \\ \leqslant & 2d(x, \mathcal{J}_{\lambda,\rho} x) + d\left(S^\rho_{n\rho} \mathcal{J}_{\lambda,\rho} x, (F_\rho)^n \mathcal{J}_{\lambda,\rho} x\right) \end{aligned}$$

Next for $\rho = \frac{t}{n}$ and $\lambda > 0$ we have by Lemma 3.10 that

$$\begin{aligned} d\left(S^\rho_{n\rho} \mathcal{J}_{\lambda,\rho} x, (F_\rho)^n \mathcal{J}_{\lambda,\rho} x\right) \leqslant & 2\sqrt{\left(n - \frac{n\rho}{\rho}\right)^2 + \frac{n\rho}{\rho}} d\left(\mathcal{J}_{\lambda,\rho} x, F_\rho \mathcal{J}_{\lambda,\rho} x\right) \\ = & 2\sqrt{n}\rho \frac{d(x, \mathcal{J}_{\lambda,\rho} x)}{\lambda} \\ \leqslant & 2\frac{T}{\sqrt{n}} \frac{d(x, \mathcal{J}_{\lambda,\rho} x)}{\lambda} \end{aligned}$$

Choose $\varepsilon > 0$ and pick $1 > \lambda_0 > 0$ so that $d(x, \mathcal{J}_{\lambda_0} x) \leqslant \lambda_0 |\partial\varphi|(x) < \varepsilon$ (the first inequality holds by (2.31)). Observe that for $\rho > 0$ small enough $2d(x, \mathcal{J}_{\lambda,\rho} x) \leqslant 2d(x, \mathcal{J}_{\lambda_0} x) + \varepsilon < 2\varepsilon$ holds due to our assumption, and than for n large enough the same holds for $\rho = \frac{t}{n}$ for any $t \in [0, T]$. For this now fixed $x > 0$ we have $\frac{d(x, \mathcal{J}_{\lambda, t/n} x)}{\lambda_0} \to \frac{d(x, \mathcal{J}_{\lambda_0} x)}{\lambda_0} < +\infty$ hence for any $t \in [0, T]$ one can find n_0 so that for $n \geqslant n_0$, $\frac{T}{\sqrt{n}} \frac{d(x, \mathcal{J}_{\lambda_0, t/n} x)}{\lambda_0} < \varepsilon$. well now (3.41) and Theorem 3.11 yield (3.40) uniformly on compact time intervals for $x \in D(|\partial\varphi|)$. For arbitrary $x \in \overline{D(\varphi)}$ let $y \in D(|\partial\varphi|)$ be such that $d(x, y) < \varepsilon$ to estimate

$$d\left(S_t x, F^n_{t/n} x\right) \leqslant 2d(x, y) + S\left(S_t y, F^n_{t/n} y\right)$$

and conclude by the previous. ∎

4 The Trotter product formula

In this section we argue the convergence of the resolvents and our final results. We will use the same notation as in the previous section, in particular (3.21) and (3.22). We will also use notation

$$U^j_t := S^j_t \circ P_j \quad t > 0 \quad j = 1, ..., n \tag{4.1}$$

or

$$U^j_t := \mathcal{J}^j_t \quad t > 0 \quad j = 1, ..., n \tag{4.2}$$

(unless we specify). We need some auxillary lemmas first.

Lemma 4.1 *Let $z \in D(\varphi)$. Set $z_0(t) = z$, and $z_j(t) := U_t^j \circ \cdots \circ U_t^1 z$ for $j = \overline{1,n}$ and for $t > 0$. Then there is a constant $c^* > 0$ such that for $j = 1, 2, ..., n$*

$$d(z_j(t), z) \leqslant c\sqrt{t} \tag{4.3}$$

Proof For $j = \overline{1,n}$ we have

$$d(z_j(t), z) \leqslant d(z_j(t), U_t^j z) + d(U_t^j z, z) \leqslant d(z_{j-1}(t), z) + d(U_t^j z, z)$$

hence by induction we only need to estimate $d(U_t^j z, z)$. Since by definition $D(\varphi) \subset D(\varphi_j)$ for each j (2.38) gives

$$d^2(U_t^j z, z) + 2t\varphi_j(U_t^j z) \leqslant 2t\varphi_j(z)$$

and by (2.23)

$$d^2\left(U_t^j z, z\right) + 2tbd\left(U_t^j z, z\right) + 2t(c - \varphi_j(z)) \leqslant 0 \tag{4.4}$$

holds, where $b, c \in \mathbb{R}$ depend only on z. Hence we must have that $d\left(U_t^j z, z\right)$ is bounded by the positive root of the quadratic equation in (4.4), i.e.

$$d\left(U_t^j z, z\right) \leqslant \frac{-2tb + \sqrt{4t^2b^2 - 8t(c - \varphi_j(z))}}{2} \tag{4.5}$$

and since there are finitely many j's to be considered we obtain exastance of c^* such that (4.3) holds. ∎

Let us now fix $x \in X$ and $\lambda \geqslant 0$ in what follows, and let us define for $t > 0$

$$x_0(t) := \mathcal{J}_\lambda x \tag{4.6}$$
$$x_j(t) := U_t^j x_{j-1}(t) \quad \text{for} j = 1, 2, ..., n \tag{4.7}$$

Observe that $x_n(t) = F_t x_0(t)$ hence by construction (recall (3.4)) and Lemma 3.1) $x_0(t) = \frac{1}{1+\lambda/t} x \oplus \frac{\lambda/t}{1+\lambda/t} x_n(t)$ thus also

$$d(x, x_n(t)) = \frac{1 + \lambda/t}{\lambda/t} d(x, x_0(t)) = \frac{t + \lambda}{\lambda} d(x, x_0(t)). \tag{4.8}$$

for any $t > 0$. Let us also rewrite (2.2) for arbitrary base point $v \in X$ as

$$d^2(v, x_0(t)) - \frac{1}{1 + \lambda/t} d^2(v, x) + \frac{\lambda/t}{1 + \lambda/t} \frac{1}{1 + \lambda/t} d^2(x, x_n(t)) \leqslant \frac{\lambda/t}{1 + \lambda/t} d^2(v, x_n(t)) \tag{4.9}$$

If F_t is given by (3.21), than by (2.39) we have

$$\frac{1}{2t} d^2(x_j(t), v) - \frac{1}{2t} d^2(x_{j-1}(t), v) + \varphi_j(x_j(t)) \leqslant \varphi_j(v) \quad \forall v \in D(\varphi_j) \tag{4.10}$$

for $t > 0$ and $j = 1, 2, ..., n$. Adding these variational inequalities and we obtain

$$\frac{1}{2t} d^2(x_n(t), v) - \frac{1}{2t} d^2(x_0(t), v) + \sum_{j=1}^{n} \varphi_j(x_j(t)) \leqslant \varphi(v) \tag{4.11}$$

for each for each $t > 0$ and $v \in D(\varphi)$. On the other hand if F_t is given by (3.22) we have a stronger estimate (2.28) for each φ_j, and summing these inequalities gives

$$\frac{1}{2t} d^2(x_n(t), v) - \frac{1}{2t} d^2(x_0(t), v) + \frac{1}{2t} \sum_{j=1}^{n} d^2(x_j(t), x_{j-1}(t)) + \sum_{j=1}^{n} \varphi_j(x_j(t)) \leqslant \varphi(v) \tag{4.12}$$

for each $t > 0$ and $v \in D(\varphi)$. In either case the inequality (4.9) and identity (4.8) can be combined with (4.11) and after some basic algebra we obtain that inequality

$$\frac{1}{2\lambda} d^2(v, x_0(t)) - \frac{1}{2\lambda} d^2(x, v) + \frac{1}{2\lambda} d^2(x, x_0(t)) + \sum_{j=1}^{n} \varphi_j(x_j(t)) \leqslant \varphi(v) \tag{4.13}$$

holds for each $v \in D(\varphi)$. Moreover if F_t is given by (3.22) we have

$$\begin{aligned} \varphi(v) \geqslant \frac{1}{2\lambda} d^2(v, x_0(t)) - \frac{1}{2\lambda} d^2(x, v) + \frac{1}{2\lambda} d^2(x, x_0(t)) \\ + \frac{1}{2t} \sum_{j=1}^{n} d^2(x_j(t), x_{j-1}(t)) + \sum_{j=1}^{n} \varphi_j(x_j(t)) \end{aligned} \tag{4.14}$$

for each $v \in D(\varphi)$.

Proposition 4.2 *For $j = 0, 1, ..., n$, $\{x_j(t)\}_{t \leqslant \varepsilon}$ is a bounded subset of X for some $\varepsilon > 0$. Moreover $\{\varphi_j(x_j(t)) | t \leqslant \varepsilon,\ j = 1, 2, ..., n\}$ is a bounded subset of $\mathbb{R}$.*

Proof We will use (4.13) to prove both claims so that it holds for either of two definitions in (4.1) of F_t. To this aim fix a $z \in D(\varphi)$, and recall the "curves" $z_j(t) = U_t^j \circ \cdots \circ U_t^1 z$ from Lemma 4.1. The contractivity property of U_t^j's implies

$$d(x_j(t), z_j(t)) \leqslant d(x_{j-1}(t), z_{j-1}(t)) \leqslant \cdots d(x_0(t), z) \tag{4.15}$$

hence

$$\limsup_{t \downarrow 0} d(x_j(t), z_j(t)) - d(x_0(t), z) \leqslant 0. \tag{4.16}$$

Hence if we show that $\{x_0(t)\}_{t \leqslant \varepsilon}$ is a bounded subset, than (4.3) and (4.15) guarantee that $\{x_j(t)\}_{t \leqslant \varepsilon}$ is bouded too for $j = 1, 2, ..., n$. By (2.23) we can find constants $b, c \in \mathbb{R}$ such that for all j's and $t > 0$, $\varphi_j(x_j(t)) \geqslant c + bd(x, x_j(t))$. Moreover estimate

$$d^2(x_0(t), v) \geqslant (d(x_0(t), x) - d(x, v))^2$$

$$= d^2(x_0(t),x) - 2d(x_0(t),x)d(x,v) + d^2(x,v)$$

and (4.13) give (with $c^* := nc$)

$$\frac{1}{2\lambda}d^2(x_0(t),x) - \frac{1}{2\lambda}d(x_0(t),x)d(x,v) + c^* + b\sum_{j=1}^{n} d(x,x_j(t)) \leqslant \varphi(v) \quad (4.17)$$

for each $v \in D(\varphi)$. This identity clearly implies that $x_0(t)$ is bounded if $b \geqslant 0$, so assuming that $b < 0$ does not reduce level of generality in the first claim. Next for each $j \in 1, ...n$ we have the following estimate:

$$\begin{aligned} d(x,x_j(t)) \leqslant & d(x,U_t^j x) + d(U_t^j x, U_t^j U_t^{j-1} x) + d(U_t^j U_t^{j-1} x, U_t^j U_t^{j-1} U_t^{j-2} x) \\ & + \cdots + d(U_t^j \circ \cdots \circ U_t^1 x, x_j(t)) \\ \leqslant & \sum_{k=1}^{j} d(x,U_t^k x) + d(x,x_0(t)). \end{aligned} \quad (4.18)$$

We want to show that the quantities $d(x,U_t^k x)$ are bounded as $t \to 0$ for each $k = 1, ...n$. Well if $U_t^j = \mathcal{J}_t^j$ for all j's pick $y \in D(|\partial\varphi_j|)$ and observe that

$$\begin{aligned} d(x,U_t^j x) &\leqslant d(x,y) + d(\mathcal{J}_t^j y, y) + d(\mathcal{J}_t^j x, \mathcal{J}_t^j y) \\ &\leqslant 2d(x,y) + d(\mathcal{J}_t^j y, y) \end{aligned}$$

which is bounded as $t \to 0$ by (2.31) and we used contractivness of resolvents (see Proposition 2.13). For the other choice of U_t^j this boundedness folows by continuity of gradient flow curves up to zero (c.f. Theorem 2.17). Hence returning to (4.17), due to (4.18) we can conclude that

$$\frac{1}{\lambda}d^2(x_0(t),x) - \alpha d(x_0(t),x) \leqslant 0 \quad \text{for } t \text{ small enough} \quad (4.19)$$

for some constants $\alpha, \beta \in \mathbb{R}$ depending only on v and x and not on t, which guarantees that $x_0(t)$ is bounded as $t \to 0$. as we already noticed (4.16) implies that $\{x_j(t)\}_{t\leqslant\varepsilon}$ is also bounded and (2.23) implies that $\{\varphi_j(x_j(t))\}_{t\leqslant\varepsilon}$ is bounded bellow for all j . Finally (4.13)) implies a posteriori that $\{\varphi_j(x_j(t))\}_{t\geqslant\varepsilon}$ is bounded above too for all j under consideration. ∎

Lemma 4.3 *Let $z \in \overline{D(\varphi)}$. Then for $j = 1, ..., n$*

$$d^2(x_j(t),z) - d^2(x_0(t),z) \longrightarrow 0 \text{ as } t \downarrow 0 \quad (4.20)$$

Moreover

$$d(x_0(t),x_n(t)) \longrightarrow 0 \text{ as } t \downarrow 0. \quad (4.21)$$

Proof Let us show (4.21) first, which follows directly from boundedness of $x_0(t)$ for small $t > 0$ cf Proposition 4.2). Indeed as $x_0(t) = \frac{1}{1\oplus\lambda/t}x + \frac{\lambda/t}{1+\lambda/t}x_n(t)$ we have $d(x_0(t),x_n(t)) = \frac{t}{\lambda}d(x,x_0(t)) \to 0$ as $t \downarrow 0$.

To prove (4.20) pick $z \in \overline{D(\varphi)}$ and $1 \leqslant j \leqslant n$. On one hand we have

$$\begin{aligned} d(x_j(t), z) - d(x_0(t), z) \leqslant & d(U_t^j z, z) + d(U_t^j U_t^{j-1} z, U_t^j z) \\ & + d(U_t^j U_t^{j-1} U_t^{j-2} z, U_t^j U_t^{j-1} z) + \cdots \\ & + d(U_t^j \cdots U_t^1 z, x_j(t)) - d(x_0(t), z) \\ \leqslant & \sum_{k=1}^{j} d(U_t^k z, z) + d(z, x_0(t)) - d(x_0(t), z) \\ = & \sum_{k=1}^{j} d(U_t^k z, z) \end{aligned} \tag{4.22}$$

and on the other hand by contractivity of U_t^j we have

$$\begin{aligned} d(x_j(t), z) - d(x_0(t), z) & \geqslant d(x_n(t), U_t^n \cdots U_t^{j+1} z) - d(x_0(t), z) \\ & \geqslant d(x_n(t), z) - d(z, U_t^n \cdots U_t^{j+1} z) - d(x_0(t), z) \\ & \geqslant -d(x_n(t), x_0(t)) - [d(z, U_t^n z) \\ & + d(U_t^n z, U_t^n U_t^{n-1} z) + \cdots + d(U_t^n \cdots U_t^{j+2} z, U_t^n \cdots U_t^{j+1} z)] \\ & \geqslant -d(x_n(t), x_0(t)) - \sum_{k=j+1}^{n} d(z, U_t^k z). \end{aligned} \tag{4.23}$$

We already argued in Proposition 4.2 that $\sum_{k=1}^{n} d(U_t^k z, z) \to 0$ for any $z \in \overline{D(\varphi)} \subset \overline{D(\varphi^k)}$ $k = \overline{1,n}$, therefore (4.21) and Proposition 4.2 yield (4.20). ∎

Now we are able to prove our main results. It turns out that the version of Trotter–Kato product formula where all each of n steps is resolvent holds on CAT(0) space. For the version with steps take by the semigroups we need a condition however (see Remark bellow Theorem 4.5).

Theorem 4.4 *Let X be a complete CAT(0) space, and let $\varphi^1, ..., \varphi^n : X \to (-\infty, +\infty]$ be proper, l.s.c. convex functionals on X such that $\varphi := \sum_{j=1}^{n} \varphi_j$ is proper too. Then the extension of the Trotter–Kato product formula holds:*

$$\left(\mathcal{J}^n_{\frac{t}{k}} \circ \mathcal{J}^{n-1}_{\frac{t}{k}} \circ \cdots \circ \mathcal{J}^1_{\frac{t}{k}} \right)^k x \longrightarrow S_t x \text{ as } k \to +\infty \text{ for } x \in \overline{D(\varphi)} \tag{4.24}$$

and the convergence is uniform on compact time intervals.

Proof Recall the resolvents $\mathcal{J}_{\lambda,t}$ for $\lambda > 0$, $t > 0$ constructed in Section 3. By Theorem 3.12 we only need to show that for $x \in D(\varphi)$ and $\lambda > 0$, convergence $\mathcal{J}_{\lambda,t} x \to \mathcal{J}_\lambda x$ as $t \downarrow 0$ holds. We fix such x and λ throughout the remainder of this proof. It is enough to show that $x_0(t_r) := \mathcal{J}_{\lambda,t_r} x \to \mathcal{J}_\lambda x$ as $t \to \infty$ for arbitrary sequence $t_r \downarrow 0$ (recall (4.7), and let us fix one such sequence.

Recall the ultraextensions φ_j^ω of φ_j for $j = \overline{1,n}$ from Section 2. relative to arbitrary but fixed ultrafilter on $\mathbb{N}$. Denote for $j = \overline{1,n}$ ultralimit $\dot{x}^j :=$

$[(x_j(t_t))_r] \in X^\omega$. This is well defined by Proposition 4.2, and note moreover that by the same proposition $\dot{x}^j \in D(\varphi_j^\omega)$ for $j = \overline{1,n}$. Next by (4.14) we have for $v \in D(\varphi) \neq \emptyset$

$$\begin{aligned} 2t_r\varphi(v) &\geqslant d^2(x_n(t_r), v) - d^2(x_0(t_r), v) \\ + \sum_{j=1}^{n} d^2(x_j(t_r), x_{j-1}(t_r)) &+ 2t_r \sum_{j=1}^{n} \varphi(x_j(t_r)) \end{aligned} \tag{4.25}$$

and Proposition 4.2 and (4.20) now yield that

$$\limsup_{r\to\infty} \sum_{j=1}^{n} d^2(x_j(t_t), x_{j-1}(t_r)) \leqslant 0 \tag{4.26}$$

hence $\lim_{r\to\omega} d(x_j(t_r), x_{j-1}(t_r)) = 0$ holds as well. In particular $\dot{x}^0 = \dot{x}^1 = \cdots = \dot{x}^n$. Returning to the inequalities (4.10) (we actally even have the stronger inequality (2.28 here) and passing to ultra–limit there for each ultralimit $\lim_{j\to\omega} v_j = \dot{v} \in X^\omega$, and subsequantly taking infimum at the right (with a slight notation abuse by writing y instead of $[y, y, y, ...]$ for $y \in X$), we conclude that

$$\frac{1}{2t_r} d_\omega^2(x_j(t_r), \dot{v}) - \frac{1}{2t_r} d_\omega^2(x_{j-1}(t_r), \dot{v}) + \varphi_j(x_j(t_r)) \leqslant \varphi_j^\omega(\dot{v}) \tag{4.27}$$

holds for each $\dot{v} \in D(\varphi_j^\omega)$, $j = \overline{1,n}$ and $r \in \mathbb{N}$. Summing (4.27) for $j = \overline{1,n}$, where we denote $\psi^\omega := \sum_{j=1}^n \varphi_j^\omega$ (which is convex on X^ω since it is a sum of convex functionals by Lemma 2.25) [8] we obtain

$$\frac{1}{2t_r} d_\omega^2(x_n(t_r), \dot{v}) - \frac{1}{2t_r} d_\omega^2(x_0(t_r), \dot{v}) + \sum_{j=1}^{n} \varphi_j(x_j(t_r)) \leqslant \psi^\omega(\dot{v}) \tag{4.28}$$

for each $\dot{v} \in D(\psi^\omega) \supset D(\varphi) \neq \emptyset$. Since $X \subset X^\omega$ and X^ω is a CAT(0) space too, the curves $s \mapsto (1-s)x \oplus s x_n(t_r)$ are geodesics in X^ω for each $r \in \mathbb{N}$ and we can do the same estimate as we did to obtain (4.14) but this time with a base point $\dot{v} \in D(\psi^\omega)$. What we get is the following ineuality:

$$\frac{1}{2\lambda} d_\omega^2(x, x_0(t_r)) - \frac{1}{2\lambda} d_\omega^2(x, \dot{v}) + \sum_{j=1}^{n} \varphi_j(x_j(t_r)) \leqslant \psi^\omega(\dot{v}) \tag{4.29}$$

for each $\dot{v} \in D(\psi^\omega)$. Since $d_\omega^2(x, x_0(t_r)) = d^2(x, x_0(t_r))$ for all $r \in \mathbb{N}$ we can take the ulta–limit for $r \to \omega$ in (4.29) and recalling the definition of ultra–extension of φ_j's conclude that

$$\frac{1}{2\lambda} d_\omega^2(x, \dot{x}^0) - \frac{1}{2\lambda} d_\omega^2(x, \dot{v}) + \psi^\omega(\dot{x}^0) \leqslant \psi^\omega(\dot{v}) \tag{4.30}$$

[8] Clearly $\psi^\omega \leqslant \varphi^\omega$ holds here, but it is not clear to the author weather $\psi^\omega = \varphi^\omega$ holds, hence a modification of the straightforward argument is necessary.

for $\dot{v} \in D(\psi^\omega)$. Now as the sequence $t_r \downarrow 0$ was arbitrary the same conclusion (4.30) holds for any subsequence of $\{t_r\}$. That is for any subsequence of $\{x_0(t_r)\}_r$. Therefore Lemma 2.26 implies that $\dot{x}_0$ and that $x_0(t_r)$ onverges to $\dot{x}_0$ in X, provided we show that the functional $\dot{\omega} \mapsto \frac{1}{2\lambda} d^2_\omega(x, \dot{\omega}) + \psi^\omega(\dot{\omega})$ admits at most one minimum point in X^ω. But this uniqueness follows by convexity of ψ^ω and stict convexity of $\dot{\omega} \mapsto \frac{1}{2\lambda} d^2_\omega(x, \dot{\omega})$ - if there were two minimizers $\dot{w}^0 \neq \dot{w}^1$ then their middle point $\dot{z} := \frac{1}{2}\dot{w}_0 \oplus \frac{1}{2}\dot{w}_1$ would satisfy

$$\begin{aligned}\frac{1}{2\lambda} d^2_\omega(x, \dot{z}) + \psi^\omega(\dot{z}) \leqslant & \frac{1}{2}\frac{1}{2\lambda} d^2_\omega(x, \dot{\omega}^0) + \frac{1}{2}\frac{1}{2\lambda} d^2_\omega(x, \dot{\omega}^1) - \frac{1}{2}\frac{1}{2}\frac{1}{2\lambda} d^2_\omega(\dot{\omega}^0, \dot{\omega}^1) \\ & + \frac{1}{2}\frac{1}{2\lambda}\psi^\omega(\dot{w}^0) + \frac{1}{2}\frac{1}{2\lambda}\psi^\omega(\dot{w}^1) < \frac{1}{2\lambda} d^2_\omega(x, \dot{w}^0) + \psi^\omega(\dot{w}^0)\end{aligned}$$

which is a contradiction. Now we have completed the proof. ∎

Remark Notice that in the above proof once we can show that for any sequence $t_r \downarrow 0$ $[x_0(t_r)_r] = [x_1(t_r)_r] = \cdots = [x_n(t_r)_r]$ we can claim the product formula formula. We will use this in the following theorem.

Theorem 4.5 *Let X a complete CAT(0) space, and let $\varphi_1, ..., \varphi_n : X \to (-\infty, +\infty]$ be proper, l.s.c. convex functionals such that $\varphi := \sum_{j=1}^n \varphi_j \not\equiv +\infty$. Assume that for $j = 1, ..., n$ the functional φ_j has the property that $\overline{D(\varphi_j^\omega)}^{X^\omega} = (\overline{D_j})^\omega$. Then the extension of the Trotter–Kato formula holds:*

$$\left(S^n_{\frac{t}{k}} \circ P_n \circ S^{n-1}_{\frac{t}{k}} \circ P_{n-1} \circ \cdots S^1_{\frac{t}{k}} \circ P_1\right)^k \longrightarrow S_t x \text{ as } k \to \infty \tag{4.31}$$

for $x \in \overline{D(\varphi)}$ and the convergence is uniform on compact time intervals. In particular it holds if $\varphi_1, ..., \varphi_{n-1}$ are bounded from above on each ball in X.

Remark As D_j is the domain of convex functional φ_j its closure $\overline{D_j}^X$ is a complete CAT(0) space we can consider $(\overline{D_j})^\omega$, and moreover it is a subspace of X^ω (the only diference between this space viewed as an ultraproduct versus viewing it as a subspace of X^ω is that in the second case some classes will have more representing sequences-namely sequences that are not in $(\overline{D_j})$. It is easy to see that $D(\varphi_j^\omega) \subset (\overline{D_j})^\omega$, but it is not clear to the author weather the oposite inclusion holds in general. However a reasonable situation when it does hold is when φ_j is bounded above on each bounded subset of X.

Proof In light of the proof of Theorem 4.4 we only need to prove that for a fixed $x \in \overline{D(\varphi)}$ and a sequence $t_r \downarrow 0$ and prove that $[(x_0(t_r))_r] = [(x_1(t_r))_r] = \cdots = [(x_n(t_r))_r]$ in X^ω where these sequances are defined by (4.7). By the remark stated before this theorem this yields the claim which we want to prove. By (2.38) we have for each $k \in \mathbb{N}$ and $j = \overline{1, n}$

$$d^2(x_j(t_r), v) - d^2(P_j x_{j-1}(t_r), v) + 2t_r \varphi_j(x_j(t_r)) \leqslant 2t_r \varphi_j(v)$$

for all $v \in D(\varphi_j) \neq \emptyset$. Hence

$$\limsup_{t\to\infty} d^2(x_j(t_r), v) - d^2(P_j x_{j-1}(t_r), v) \leqslant 0 \text{ for } v \in \overline{D(\varphi_j)} \tag{4.32}$$

hence also

$$d_\omega(\dot{x}^j, \dot{v}) - d_\omega(\dot{z}_j, v) \leqslant 0 \tag{4.33}$$

for each $\dot{v} \in \overline{D(\varphi_j^\omega)}$ where we denote $\dot{x}^j := [(x_j(t_r))_r]$ and $\dot{z}_j = [(P_j x_{j-1}(t_r))_r] \in X^\omega$. On the other hand by (2.18) for any $v \in \overline{D(\varphi_j)} \neq \emptyset$

$$d^2(x_{j-1}(t_r), v) - d^2(P_k x_{j-1}(t_r), v) \geqslant d^2(x_{j-1}(t_r), P_j x_{j-1}(t_r)) \tag{4.34}$$

Now (4.32) and (4.34) together imply that

$$\limsup_{k\to\infty} d^2(x_{j-1}(t_r), v) - d^2(P_j x_{j-1}(t_r), v) \leqslant 0 \quad \forall v \in \overline{D(\varphi_j)} \neq \emptyset \tag{4.35}$$

and this together with (4.34) yields

$$d^2(x_{j-1}(t_r), P_j x_{j-1}(t_r)) \longrightarrow 0 \text{ as } r \to \infty \tag{4.36}$$

and in particular $\dot{x}_{j-1} = \dot{z}_j$. Now as $\dot{z}_j \in (\overline{D}_j)^\omega$ by definition, the assumption $\overline{D(\varphi_j^\omega)} = (\overline{D}_j)^\omega$ implies that we can take $\dot{v} = \dot{z}_j = \dot{x}_{j-1}$ so that by (4.33) we must have $\dot{x}_j = \dot{x}_{j-1}$ for all j. The proof is now completed. ∎

Remark We can also take combined steps of semigroups and resolvents. A sufficient assumption in that case that is the following: for all j where we take $S^j_{t/k}$ in the product formula instead of $\mathcal{J}^j_{t/k}$ let $\overline{D(\varphi_j^\omega)} = \overline{(D(\varphi_j)}^\omega$ hold, which in particular holds if those φ_j's are bounded from above on each ball in X. Moreover we can replace each $\mathcal{J}^j_{t/k}$ by $\left(\mathcal{J}^j_{\frac{t}{r_j k}}\right)^{r_j}$ for any $r_j \in \mathbb{N}$. This claimfollows by the same arguments and using the corresponding variational inequality for $\left(\mathcal{J}^j_{\frac{t}{r_j}}\right)^{r_j}$.

There is one other instance when the product formula holds.

Theorem 4.6 *Let X be a complete CAT(0) space, and let $\varphi_1, ..., \varphi_n : X \to (-\infty, +\infty]$ be proper, l.s.c. and convex and let $\varphi := \sum_{j=1}^n \varphi_j \neq +\infty$. Assume moreover that for one $j \in \{1, ..., n\}$ the bounded parts of the level sets of φ_j are relatively compact in X. Then the product formula holds:*

$$\left(U^n_{\frac{t}{k}} \circ \cdots \circ U^1_{\frac{t}{k}}\right)^k x \longrightarrow S_t x \quad x \in \overline{D(\varphi)} \tag{4.37}$$

the convergence being uniform on compact time intervals.

Proof We only need to show that $x_0(t) = \mathcal{J}_{\lambda,t}x \to \mathcal{J}_\lambda x$ for $x \in D(\varphi)$ and $\lambda > 0$. Let $j_0 \in \{1, ..., n\}$ be such that φ_j has the stated property. Then as $\{x_{j_0}(t)\}_{t\geqslant 0}$ is bounded by Proposition 4.2, for each sequence $t_r \downarrow 0$, the sequence $\{x_{j_0}(t)\}$ has a convergent subsequence $\{x_{j_0}(t)\}_l$ in X. But then by Lemma 4.3 we also have that $\{x_{j_0}(t)\}_l$ must converge to the same limit for all $j \in \{1...., \}$ and (4.13) together with the lower semicontinuity of φ_j's yields that $x_0(t_{k_l}) \to \mathcal{J}_\lambda x$ as $l \to \infty$, hence also $x_0(t_r) \to \mathcal{J}_\lambda x$ as $r \to +\infty$. ∎

Remark If X is locally compact then due to Hopf–Rinow Theorem 2.9 (4.37) always holds, provoded the conditions of our theorem are satisfied.

5 Examples

As we already explained in the previous section, out Theorem 4.6 includes the classical i.e. Hilbert space situation. The best known example within this theory is the heat equation functional

$$\begin{aligned} \varphi(U) &:= \int_\Omega |\nabla U|^2 dx \qquad && \text{if } u \in W^{1,2}(\Omega) \\ \varphi(u) &:= +\infty && \text{if } u \in L^2(\Omega) \setminus W^{1,2}(\Omega) \end{aligned} \tag{5.1}$$

where Ω is a domain in $\mathbb{R}^n$. Then $\varphi := \sum_{j=1}^n \varphi_j$ where $\varphi_j(u) = \int (\partial_j u)^2 dx$ defined on a suitable subset of $L^2(\Omega)$. These functionals are convex, proper and lower semicontinuous and as $L^2(\Omega)$ is Hilbert space our theory can be applied.

In the general setting of a CAT(0) space X without any additional structure, the most natural convex and lower semicontinuous functionals are

$$\varphi(x) := \frac{1}{2} d^2(x, z) \quad x \in X \tag{5.2}$$

where $z \in X$ is a fixed point; moreover one can conider any positive combination of such functionals. So let us fix distinct points $z_1, ..., z_n$ in our CAT(0) space X, and positive numbers $\alpha_1, ..., \alpha_n > 0$, and define functionals

$$\varphi_j(x) := \frac{\alpha_j}{2} d^2(x, z_j) \quad j = 1, ...n \quad x \in X \tag{5.3}$$

$$\varphi(x) := \sum_{j=1}^n \frac{\alpha_j}{2} d^2(x, z_j) \quad x \in X \tag{5.4}$$

Each of these functionals is clearly continuous, and the CAT(0) condition is clearly even stronger then the convexity along geodesics.

Moreover such functionals are clearly bounded on balls, so any version of the product formula holds. Furthermore one can consider the functionals $x \mapsto d^p(x, x_j)$ where $p > 1$. There are also convex, continuous, and bounded on balls, hence any version of the product formula holds here as well.

A more complex example is when one considers a gradient flow associated to a convex l.s.c. functional φ on CAT(0) space X but wants to constrain it to a closed convex subset $C \subset X$. In this context convexity is understood in the sence of the CAT(0) metric of our space. For example if X is a Hilbert ball, such convex subset may not be convex in the ussual linear sense. Likewise we can consider a gradient flow on the space of positive unitized Hilbert-Schmidt operators equipped with the trace inner product (c.f. [24]). In order to construct such flow one can add the indicator functional

$$\varphi_C(x) := 1 \quad \text{if} \quad x \in C, \quad \varphi_C(x) := +\infty \quad \text{if} \quad x \in C^c \tag{5.5}$$

to φ , i.e. consider the flow associated to $\varphi^1 := \varphi + \varphi_C$. Than it can be very hard to get any idea about the paths of the constrained flow. But applying the product formula gives more insight as the resolvents associated to the indicator functional φ_C are easily seen to be just the projections onto C, and its emigroup is just identity map on C. We reffer to [25] and [28] for some concrete examples in the linear setting.

References

[1] Alexander, S.; Kapovich, V. Petrunin, A. "Alexandrov Geometry", book in preparation.

[2] Ambrosio, L.; Gigli, N.; Savare, G. "Gradient flows on metric spaces and in the space of probability measures" Lectures in Mathematics ETH Zürich, Birkheuser Verlag, Basel, 2005. ISBN: 978–3–7643–2428–7.

[3] Aharonov, D.; Elin, M.; Reich, S.; Shoikhet, D. Parametric representations of semi-complete vector fields on the unit balls in C^n and in Hilbert space Atti Accad. Naz. Lincei Cl. Sci. Fis. Mat. Natur. Rend. Lincei (9) Mat. Appl. 10 (1999), no. 4, 229–253

[4] Brezis, H. "Opérateurs maximus monotones et semi–groupes de contractions dans les espaces de Hilbert" (French) North Holand Matematics Studies, No, 5. Notes de Mathematics (50). North–Holland Publishing Co. Amsterdam–London; American–Elsevier Publishing Co., Unc. New York, 10973.

[5] Brezis, H.; Pazy, A. "Convergence and Approximation of semigroups of nonlinear operators in Banach spaces J. Functional Analysis 9)1972), 63–74.

[6] Bridson, M.; Haefliger, A. "Metric spaces of non–positive curvature" Grundlefiress der Mathematischen Wissenschaften [Fundamental Principles of Mathematical sciences], 319. Springer–Verlag, Berlin, 1999. ISBN 3–540–64324–9.

[7] Burago, D.; Burago, Y.; Ivanov, S. "A course in metric geometry", Gradute Studies in Mathematics, 33. American Mathematical Society, Providence, RI, 2001 ISBN: 0–8218–2129–6.

[8] Carillo, I.A.; McCann, R.J.; Villani, C. :Contractions in the 2–Wasserstein Length Space and Thermalization of Gronkolov Media", Arch. Rational Mech, Anal. (2005).

[9] Clement, P. "An Introduction to Gradient Flows on Metric Spaces", Report MI 2009-09 Mathematical Institute, Leiden University 2009. www.math.leidenuniv.nl

[10] Crandall, M,G; Liggett, T.M. Generation of semi–group of nonlinear transformations on Banach spaces" Amer. J. Math. 93, 1971, 265–298.

[11] Dhompongsa, P; Panyanak, B On Δ-convergence theorems in CAT(0) spaces. Comput.Math.Appl. 56 (2008) No. 10

[12] Engel, K-J.; Nagel, R. "One-parameter semigroups for linear evolution equations", Graduate Texts in Mathematics, 194. Springer–Verlag, New-York 2000. ISBN: 0–387–98463–1.

[13] Goebel, K; Reich, S. Uniform convexity, hyperbolic geometry, and nonexpansive mappings. Monographs and Textbooks in Pure and Applied Mathematics, 83. . Marcel Dekker, Inc., New York, 1984. ix+170 pp. ISBN: 0-8247-7223-7 58C30

[14] Gorach, G.; Maestripieri, A.L. "Differential and Metrical Structure of Positive Operators", Positivity 3: 297–315 (1999).

[15] Gromov, M.; Schoen, R. "Harmonic Maps into Singular Spaces and p–adic Superrigidity for Latteces in Groups of Rank One", Inst. Hautes Etudes Sci. Publ. Math. No. 76 (1992), 165–246.

[16] Israel, M..; Reich, S. A note on nonlinear semigroups with large sets of fixed points. Global analysis–analysis on manifolds, 154–157, Teubner-Texte Math., 57, Teubner, Leipzig, 1983

[17] Iwamiya, T.; Okochi, H. "Monotonicity, resolvents andYasida approximations of operators on Hilbert manifolds" Nonlinear Analysis 54 (2003) 205–214.

[18] Jordan, R.; Kinderlehrer, D.; Otto, F. "The variational formulation of the Fokker–Planck equation", SIAM J. Math. Anal. 29 (1998) No. 1, 1–17.

[19] Jost, J. "Convex functionals and generalized harmonic maps into spaces of non positive curvature" Comment. Math, Helvetici. 70(1995) 659–673.

[20] Jost, J. "Equlibrium maps between metric spaces" Calc. Var. 2 (1994) 173–204.

[21] Jost, J. "Nonpositive Curvature: Geometric and Analytic Aspects", Lectures in Mathematics ETH Zürich, Birkhäuser Verlag, Basel–Boston–Berlin, ISBN 3–7645–5736–3 (Basel) ISBN 0–8176–5736–3 (Boston).

[22] Karevoar, Schoen, R. "Sobolev spaces and harmonic maps for metric space targets" Comm. Anal. Geom. 1 (1993), No. 3, 561–659.

[23] Kato, R.; Masuda, K. "Trotter's product formula for nonlinear semigroups generated by the subdifferentials of convex functionals" J. Math. Soc. Japan 30 (1978) No. 1, 169–178.

[24] Larotonda, G. "Nonpositive curvature; A geometrical approach to Hilbert–Schmidt operators" Diff. Geom. Appl. 25 (2007) 679–700.

[25] Lewy, H. On variational problem with inequalities on the boundary. Journal of mathematics and mechanics, Vol. 17, No. 9 (1968)

[26] Li, C.; Lopez, G.; Martn–Márquez, V. "Monotone vector fields and the aproximal point algorithm on Hademond manifolds", J. London Math. Soc. (2007).

[27] Li, C.; Lopez, G.; Martn–Márquez, V. ; Wang, J-H. "Resolvents of set Valued Monotone Vector Field in Hademond Manifolds", [re[rint.

[28] Lions, J.L.; Stampacchia, G Variational inequalities. Communications in pure and aplied mathematics, Vol. XX, 493-519 (1967)

[29] Lytchak, A. "Open map theorem for metric spaces" Algebra i Analiz 17 (2005) No. 3 139–159. English translation in St. metric spaces and harmonic mapis" Comm. Anal. Geom. 6 (1998) No. 2, 199–253.

[30] Mayer, U. "Gradient flows on nonpositively curved" Comm. Appl. Geom. 6(1998), No. 2, 199–253.

[31] Miyadera, I.; Ohary, S. "Approximation of semigroups of nonlinear operators", Tehoku Math. J. (2) 22 (1970), 24–27.

[32] Ohta, S-I. "Cheeger Type Sobolev Spaces for Metric Space Targets" Potential Analysis 20: 149–175, 2004.

[33] Ohta, S-I. "Gradient flows on Wasserstein spaces over compact Alexandrov spaces".

[34] Otto, F. "The geometry of dissipative evolution equations: The perous medium equation" Comm. Part. Diff. Eq. 26 (2001), 101–174.

[35] Perelman, G; Petrinin, A. Quasigeodesics and Gradient curves In Alexandrov spaces. Unpublished preprint. 1994

[36] Rapcsk, T. Sectional curvature in nonlinearoptimization. J. Global Optim. 40 (2008) 375-388

[37] Reich, S. The asymptotic behavior of a class of nonlinear semigroups in the Hilbert ball. J. Math. Anal. Appl. 157 (1991), no. 1, 237–242.

[38] Reich, S; Shafrir, I. Nonexpansive iterations in hyperbolic spaces. Nonlinear Anal. 15 (1990), no. 6, 537–558.

[39] Reich, S; Shoikhet, D. Semigroups and generators on convex domains with the hyperbolic metric. Atti Accad. Naz. Lincei Cl. Sci. Fis. Mat. Natur. Rend. Lincei (9) Mat. Appl. 8 (1997), no. 4, 231–250.

[40] Stojkovic, I. Wasserstein-2analysis ofthe Non-symmetric Fokker-Planck equation and the Trotter-Kato product formula. Report MI 2010-13 Mathematical Institute, Leiden University, The Netherlands. 2010.

[41] Sturm, K-T. "A semigroups approach to Harmonic maps" Potential Analysis (2005) 23: 225–277.

[42] Sturm, K-T.; ...., T. "Expections and martingales in metric spaces" Stochastics 80 (2008), No. 1, 1–17.$\bar{p}$

[43] Sturm, K-T. "Nonlinear Markov operators, discrete heat flow and harmonic maps between singular spaces" Potential Analysis 16: 305–340, 2002.

[44] Sturm, K-T. "Nonlinear martingale theory for processes with values in metric spaces of nonpositive curvature". The Annals of Probability 2002, Vol. 30, No. 3, 1195–1222.

[45] Villani, C. "Optimal transport: old and new" Grundlehren der Mathematichen Wissenschaften, 338, Springer-Verlag, Berlin, 2009. xxii + 973 p.p. ISBN: 978–3–540–71049–3.

[46] Villani, C. "Topics in Optimal Transportation" Graduate Studies in Mathematics Volume 58 American Mathematical Society, Providence, Rhode Island, ISBN: 0–8218–3312–X.