EXISTENCE AND UNIQUENESS OF THE GRADIENT FLOW OF THE ENTROPY IN THE SPACE OF PROBABILITY MEASURES

STEFANO BIANCHINI AND ALEXANDER DABROWSKI

Abstract. After a brief introduction on gradient flows in metric spaces and on geodesically convex functionals, we give an account of the proof (following the outline of [3, 7]) of the existence and uniqueness of the gradient flow of the Entropy in the space of Borel probability measures over a compact geodesic metric space with Ricci curvature bounded from below.

Preprint SISSA 17/2014/MATE

Contents

1. Introduction

This paper aims to give an account of some of the main ideas from recent developments on gradient flows in metric measure spaces, examining the special case of the gradient flow of the Entropy functional in the space of probability measures. The results presented in this work are published in several texts, mainly [3, 7, 4, 2]; our aim is to give to the interested reader a single self-contained paper with both

Date: 18 March 2014.

2010 *Mathematics Subject Classification.* 35K90, 60B05, 28A33.

Key words and phrases. Entropy, gradient flows.

This paper resumes the main part of the Bachelor thesis of the second author, discussed in 2013 at the University of Trieste.

the proofs of existence and uniqueness of the gradient flow of the Entropy. We prove technical results when needed, however, to avoid excessive difficulties, in Section 4 we restrict our analysis to the case of a compact metric space.

We assume the reader has some familiarity with standard tools in measure theory; we recall the fundamental ones in Section 2.

In Section 3 we introduce the main concepts of the theory of gradient flows in a purely metric setting. We begin with a generalization to metric spaces of a property of gradient flows in the smooth setting, namely the Energy Dissipation Equality, which relies only on the norm of the differential of the functional and the norm of the derivative of the curve which solves the gradient flow. To make sense of these two concepts in a metric space we introduce the *metric speed* of a curve and the *descending slope* of a functional. We proceed defining K-convexity in geodesic metric spaces and state a useful formula for computing the descending slope and an important weak form of the chain rule for K-convex functionals.

Section 4 is dedicated to the proof of existence of the gradient flow of the Entropy in the space of probability measures over a compact metric space. After defining the fundamental *Wasserstein distance* between probability measures, we introduce the Entropy functional and outline two interesting cases where a solution to a PDE is obtained as a solution of a gradient flow of a functional: the Dirichlet Energy in $L^2(\mathbb{R}^n)$ and the Entropy in the space of probability measures on the torus $\mathbb{T}^n$. We then give the definition of *geodesic* metric space with *Ricci curvature bounded from below*, a concept which will allow us to apply the theory developed for K-convex functionals to the Entropy. We proceed by proving the existence of a curve solving the gradient flow using a discrete approximation scheme and showing its convergence to a curve which satisfies the Energy Dissipation Equality.

Section 5 deals with the uniqueness of the gradient flow of the Entropy. A deeper understanding of the nature of the curve solving the gradient flow is achieved introducing the concept of *push-forward via a plan* and restricting our analysis to plans with *bounded deformation*. After proving some preliminary properties concerning the approximation of the Entropy and the convexity of the squared descending slope of the Entropy, we conclude showing the uniqueness of the gradient flow.

2. Measure theoretic preliminaries

From now on, if not otherwise stated, $(X, \mathfrak{d})$ will be a complete and separable metric space with distance $\mathfrak{d}$. We will indicate as $\mathcal{P}(X)$ the set of Borel probability measures on X.

We recall two concepts we will often use: the push-forward of a measure through a map and narrow convergence of measures.

Definition 2.1. Let X, Y be metric spaces, $\mu \in \mathcal{P}(X)$, $T : X \to Y$ a Borel map. We define the *push-forward* of μ through T as

$$T_*\mu(E) := \mu(T^{-1}(E))$$

for every Borel subset $E \subseteq Y$.

The push-forward of a measure satisfies the following property: for every Borel function $f : Y \to \mathbb{R} \cup \{\infty\}$ it holds

$$\int_Y f \, T_* \mu = \int_X f \circ T \, \mu,$$

where the equality means that if one of the integrals exists so does the other one and their value is the same.

We give a useful weak notion of convergence in $\mathcal{P}(X)$.

Definition 2.2. Given $(\mu_n)_n$ a sequence of measures in $\mathcal{P}(X)$, μ_n *narrowly* converges to $\mu \in \mathcal{P}(X)$, and we write $\mu_n \rightharpoonup \mu$, if for every $\varphi \in C_b(X, \mathbb{R})$ it holds

$$\int_X \varphi \, \mu_n \to \int_X \varphi \, \mu.$$

Notice that if X is compact, weak* convergence and narrow convergence on $\mathcal{P}(X)$ are the same, thanks to the Riesz-Markov-Kakutani Representation Theorem.

We pass to examine absolutely continuous measures.

Definition 2.3. Let λ, μ be measures on a σ-algebra $\mathcal{A}$. λ is *absolutely continuous* with respect to μ, and we write $\lambda \ll \mu$, if for every $E \in \mathcal{A}$ such that $\mu(E) = 0$, it also holds $\lambda(E) = 0$.

The following two classic theorems will be widely used in the last section of this paper.

Theorem 2.4 (Radon-Nikodym)**.** *Let λ, μ be finite measures on X measurable space, $\lambda \ll \mu$. Then there exists a unique $h \in L^1(\mu)$ such that*

$$\lambda = h\mu. \tag{1}$$

To express (1) we also write synthetically

$$\frac{d\lambda}{d\mu} = h.$$

The analogue of Radon-Nikodym Theorem in the space of probability measures is the Disintegration Theorem.

Theorem 2.5 (Disintegration Theorem)**.** *Let $(X, \mathfrak{d}_X), (Y, \mathfrak{d}_Y)$ be complete and separable metric spaces, let $\gamma \in \mathcal{P}(X \times Y)$ and $\pi^1 : X \times Y \to X$ the projection on the first coordinate. Then there exists a $\pi^1_*\gamma$-almost everywhere uniquely determined family of probability measures $\{\gamma_x\}_{x \in X} \subseteq \mathcal{P}(Y)$ such that*

1. *the function $x \mapsto \gamma_x(B)$ is a Borel map for every Borel $B \subseteq Y$,*
2. *for every Borel function $f : X \times Y \to [0, \infty]$ it holds*

$$\int_{X \times Y} f(x, y) \, \gamma(dx, dy) = \int_X \left(\int_Y f(x, y) \, \gamma_x(dy) \right) \, \pi^1_*\gamma(dx). \tag{2}$$

We also express property (2) by writing

$$\gamma = \int_X \gamma_x \, \pi^1_*\gamma(dx).$$

The theorem obviously holds *mutatis mutandis* for the projection π^2 on Y.

For a proof of Theorem 2.5 and a broader view on the topic see [6, Chapter 45]. A simpler proof for vector valued measures can be found in [1, Theorem 2.28].

3. Gradient flows in metric spaces

The following observation is the starting point to extend the notion of gradient flows to metric spaces.

From now on we adopt the subscript notation for curves (i.e. $u_t = u(t)$).

Remark 3.1. Let H be a Hilbert space, $E : H \to \mathbb{R} \cup \{\infty\}$ a Frechét differentiable functional. If $u : [0, \infty) \to \mathbb{R}$ is a gradient flow of E, i.e. $\dot{u}_t = -\nabla E(u_t)$, then

$$\frac{d}{dt}E(u_t) = \langle \nabla E(u_t), \dot{u}_t \rangle = -\frac{1}{2}\|\dot{u}_t\|^2 - \frac{1}{2}\|\nabla E(u_t)\|^2.$$

Integrating with respect to t we obtain

$$E(u_s) - E(u_0) = -\frac{1}{2}\int_0^s \|\dot{u}_t\|^2 \, dt - \frac{1}{2}\int_0^s \|\nabla E(u_t)\|^2 \, dt \quad \forall s > 0.$$

This last equality is called *Energy Dissipation Equality.*

By extending appropriately the concepts of the norm of the derivative of a curve and the norm of the gradient of a functional we can make sense of this last equality even in metric spaces.

We restrict our analysis to a special class of curves.

Definition 3.2. A curve $u : [0, 1] \to X$ is *absolutely continuous* if there exists $g \in L^1(I)$ such that for every $t < s$ it holds

$$\mathfrak{d}(u_t, u_s) \leq \int_t^s g(r) \, dr. \tag{3}$$

For absolutely continuous curves we are able to define a corresponding concept of speed of a curve.

Proposition 3.3. *If u is an absolutely continuous curve, there exists a minimal (in the L^1-sense) g which satisfies* (3)*; this function is given for almost every t by*

$$|\dot{u}_t| := \lim_{s \to t} \frac{\mathfrak{d}(u_s, u_t)}{|s - t|}.$$

The function $|\dot{u}_t|$ is called *metric derivative* or *metric speed* of u.

Proof. Let $(y_n)_n$ be dense in $u(I)$ and define

$$h_n(t) := \mathfrak{d}(y_n, u_t) \quad \forall n \in \mathbb{N}.$$

Let $g \in L^1$ be such that

$$|h_n(t) - h_n(s)| \leq \mathfrak{d}(u_t, u_s) \leq \int_t^s g(r) \, dr \quad \forall n \in \mathbb{N}.$$

Therefore $h_n(t)$ are absolutely continuous for every n, so by the Lebesgue Fundamental Theorem of Calculus there exists $h_n' \in L^1$ such that

$$h_n(t) - h_n(s) = \int_t^s h_n'(r) \, dr.$$

We have that $|h_n'(t)| \leq g(t)$ a.e. and a fairly easy calculation shows that

$$\limsup_{s \to t} \frac{\mathfrak{d}(u_s, u_t)}{|s - t|} \leq \sup_n |h_n'(t)| \leq \liminf_{s \to t} \frac{\mathfrak{d}(u_s, u_t)}{|s - t|},$$

therefore we can take $\sup_n |h_n'(t)|$ as the metric derivative. □

We pass to define the concept which will substitute the norm of the differential of a function in the metric case.

We indicate by $(\,\cdot\,)^+$, $(\,\cdot\,)^-$ the standard positive and negative parts, i.e. $x^+ = \max\{x,0\}$, $x^- = \max\{-x,0\}$.

Definition 3.4. Let $E : X \to \mathbb{R} \cup \{\infty\}$. The *descending slope* of E at x is

$$|D^- E|(x) := \limsup_{y \to x} \frac{\big(E(x) - E(y)\big)^+}{\mathfrak{d}(x,y)}.$$

We are now ready to define gradient flows using the Energy Dissipation Equality (EDE).

For a functional $E : X \to \mathbb{R} \cup \{\infty\}$, we write $D(E) := \{x : E(x) < \infty\}$.

Definition 3.5 (Gradient flow - EDE)**.** Let E be a functional from X to $\mathbb{R} \cup \{\infty\}$ and let $x_0 \in D(E)$. A locally absolutely continuous curve $x : [0,\infty) \to X$ is an *EDE-gradient flow*, or simply a *gradient flow*, of E starting from x_0 if x takes values in $D(E)$ and it holds

$$E(x_s) = E(x_t) - \frac{1}{2}\int_t^s |\dot{x}_r|^2\, dr - \frac{1}{2}\int_t^s |D^- E|^2(x_r)\, dr, \quad \forall s > t,$$

or equivalently

$$E(x_s) = E(x_0) - \frac{1}{2}\int_0^s |\dot{x}_r|^2\, dr - \frac{1}{2}\int_0^s |D^- E|^2(x_r)\, dr, \quad \forall s > 0.$$

3.1. **K-convexity.** A class of functionals with useful properties is that of K-convex functionals.

In $\mathbb{R}^n$ the standard definition is that the distributional derivative of a function $E : \mathbb{R}^n \mapsto \mathbb{R}$ satisfies

$$D^2 E - K I_n \geq 0,$$

where I_n is the $n \times n$-identity matrix. To extend the definition from the smooth setting to metric spaces we will use geodesics.

Definition 3.6. A metric space X is *geodesic* if $\forall x_0, x_1 \in X, \exists g : [0,1] \to X$ such that $g_0 = x_0, g_1 = x_1$ and

$$\mathfrak{d}(g_t, g_s) = |t-s|\mathfrak{d}(x_0,x_1), \quad \forall s,t \in [0,1].$$

Such a g is called *constant speed geodesic* between x_0 and x_1.

It is natural to extend the definition of K-convexity by requiring the K-convexity of the functional along geodesics.

Definition 3.7. Let $(X,\mathfrak{d})$ be a geodesic space, $E : X \to \mathbb{R} \cup \{\infty\}$. E is *K-geodesically convex*, or simply *K-convex*, if $\forall x_0, x_1 \in Y, \exists g : [0,1] \to X$ constant speed geodesic between x_0 and x_1 and for every $t \in [0,1]$ it holds

$$E(g_t) \leq (1-t)E(x_0) + tE(x_1) - \frac{K}{2} t(1-t)\mathfrak{d}^2(x_0,x_1).$$

Remark 3.8. Notice that if E is K-convex then for every $K' \leq K, E$ is K'-convex.

We prove a useful formula for computing the descending slope of K-convex functionals.

Lemma 3.9. *If E is K-convex then*

$$|D^- E|(x) = \sup_{y\neq x}\left(\frac{E(x)-E(y)}{\mathfrak{d}(x,y)} + \frac{K}{2}\mathfrak{d}(x,y)\right)^+. \tag{4}$$

Proof. " $\leq$ ". This inequality holds trivially.
" $\geq$ ". Fix $y \neq x$. Let g be a constant speed geodesic from x to y such that

$$\begin{aligned}\frac{E(x)-E(g_t)}{\mathfrak{d}(x,g_t)} &\geq \frac{t}{\mathfrak{d}(x,g_t)}\left(E(x)-E(y)+\frac{K}{2}(1-t)\mathfrak{d}^2(x,y)\right)\\ &= \frac{E(x)-E(y)}{\mathfrak{d}(x,y)}+\frac{K}{2}(1-t)\mathfrak{d}(x,y).\end{aligned}$$

Therefore as $t \to 0$,

$$\begin{aligned}|D^-E|(x) &\geq \limsup_{t\to 0^+}\left(\frac{E(x)-E(g_t)}{\mathfrak{d}(x,g_t)}\right)^+\\ &\geq \left(\limsup_{t\to 0^+}\left(\frac{E(x)-E(y)}{\mathfrak{d}(x,y)}+\frac{K}{2}(1-t)\mathfrak{d}(x,y)\right)\right)^+\\ &= \left(\frac{E(x)-E(y)}{\mathfrak{d}(x,y)}+\frac{K}{2}\mathfrak{d}(x,y)\right)^+.\end{aligned}$$

We conclude by taking the supremum w.r.t. y. □

For K-convex functionals we have a useful weak form of chain rule.

Theorem 3.10. *Let $E : X \to \mathbb{R}\cup\{\infty\}$ be a K-convex and lower semicontinuous functional. Then for every absolutely continuous curve $x : [0,1] \to X$ such that $E(x_t) < \infty$ for every $t \in [0,1]$, it holds*

$$|E(x_s)-E(x_t)| \leq \int_t^s |\dot{x}_r||D^-E|(x_r)\,dr, \tag{5}$$

with $t < s$.

Proof. We follow the reasoning of [2, Proposition 3.19].

Step 0. By linear scaling we may reduce to the case $t = 0$ and $s = 1$. We may also assume that

$$\int_0^1 |\dot{x}_r||D^-E|(x_r)\,dr < \infty,$$

otherwise the inequality holds trivially. By the standard arc-length reparametrization we may furthermore assume $|\dot{x}_t| = 1$ for almost every t, so x_t is 1-Lipschitz and the function $t \mapsto |D^-E|(x_t)$ is in $L^1([0,1])$.

Step 1. Notice that it is sufficient to prove absolute continuity of the function $t \mapsto E(x_t)$, then the thesis follows from the inequality

$$\begin{aligned}&\limsup_{h\to 0}\frac{E(x_{t+h})-E(x_t)}{h} \leq \limsup_{h\to 0}\frac{\big(E(x_{t+h})-E(x_t)\big)^+}{|h|}\\ &\leq \limsup_{h\to 0}\frac{\big(E(x_{t+h})-E(x_t)\big)^+}{\mathfrak{d}(x_{t+h},x_t)}\limsup_{h\to 0}\frac{\mathfrak{d}(x_{t+h},x_t)}{|h|} \leq |D^-E|(x_t)|\dot{x}_t|\end{aligned}$$

and the fact that for a.c. f it holds

$$f(s) - f(t) = \int_t^s \frac{df}{d\tau} d\tau.$$

Step 2. We define $f, g : [0,1] \to \mathbb{R}$ as

$$\begin{aligned} f(t) &:= E(x_t), \\ g(t) &:= \sup_{s \neq t} \frac{\big(f(t) - f(s)\big)^+}{|t-s|}. \end{aligned}$$

From the fact that $|\dot{x}_t| = 1$ and the trivial inequality $a^+ \leq (a+b)^+ + b^-$ valid for any $a, b \in \mathbb{R}$ we obtain

$$\begin{aligned} g(t) &\leq \sup_{s \neq t} \frac{\big(f(t) - f(s)\big)^+}{\mathfrak{d}(x_t, x_s)} \\ &\leq \left(\sup_{s \neq t} \frac{f(t) - f(s)}{\mathfrak{d}(x_t, x_s)} + \frac{K}{2} \mathfrak{d}(x_t, x_s) \right)^+ + \left(\frac{K}{2} \mathfrak{d}(x_t, x_s) \right)^- . \end{aligned}$$

Since $\{x_t\}_{t \in [0,1]}$ is compact, there exists a $D \in \mathbb{R}^+$ such that $\mathfrak{d}(x_t, x_s) \leq D$; applying then (4) we obtain

$$g(t) \leq |D^- E|(x_t) + \frac{K^-}{2} D.$$

Therefore the thesis is proven if we show that

$$|f(s) - f(t)| \leq \int_t^s g(r)\, dr.$$

Step 3. Fix $M, \varepsilon > 0$ and define $f^M := \min\{f, M\}$, $\rho_\varepsilon : \mathbb{R} \to \mathbb{R}$ a smooth mollifier with support in $[-\varepsilon, \varepsilon]$ and $f_\varepsilon^M, g_\varepsilon^M : [\varepsilon, 1-\varepsilon] \to \mathbb{R}$ such that

$$\begin{aligned} f_\varepsilon^M(t) &:= (f^M * \rho_\varepsilon)(t), \\ g_\varepsilon^M(t) &:= \sup_{s \neq t} \frac{\big(f_\varepsilon^M(t) - f_\varepsilon^M(s)\big)^+}{|s-t|}. \end{aligned}$$

Since f_ε^M is smooth and $g_\varepsilon^M \geq \big|(f_\varepsilon^M)'\big|$,

$$\big|f_\varepsilon^M(s) - f_\varepsilon^M(t)\big| \leq \int_t^s g_\varepsilon^M(r)\, dr.$$

Therefore we have

$$\begin{aligned} g_\varepsilon^M(t) &= \sup_{s \neq t} \frac{1}{|s-t|} \left(\int_0^1 \big(f^M(t-r) - f^M(s-r)\big) \rho_\varepsilon(r)\, dr \right)^+ \\ &\leq \sup_{s \neq t} \frac{1}{|s-t|} \int_0^1 \big(f(t-r) - f(s-r)\big)^+ \rho_\varepsilon(r)\, dr \\ &\leq \sup_{s \neq t} \int_0^1 \frac{\big(f(t-r) - f(s-r)\big)^+}{|(s-r) - (t-r)|} \rho_\varepsilon(r)\, dr \\ &\leq \int_0^1 g(t-r) \rho_\varepsilon(r)\, dr = (g * \rho_\varepsilon)(t), \end{aligned}$$

thus the family $\left\{g_\varepsilon^M\right\}_\varepsilon$ is uniformly integrable in $L^1((0,1))$. In fact, for $A \subset (\varepsilon, 1-\varepsilon)$,

$$\int_A g_\varepsilon^M(t)\,dt \le \int_A (g*\rho_\varepsilon)(t)\,dt = \int_A \int_0^1 g(t-y)\rho_\varepsilon(y)\,dy\,dt$$
$$= \int_0^1 \frac{1}{\varepsilon}\rho\left(\frac{y}{\varepsilon}\right)\int_{A-y} g(t)\,dt\,dy = \int_0^1 \rho(z)\left[\int_{A-\varepsilon z} g(t)\,dt\right]dz \le \omega(|A|),$$

where $\omega(|A|) = \sup_{\mathcal{L}(B)=\mathcal{L}(A)}\{\int_B g\}$.
Since

$$|f_\varepsilon^M(y) - f_\varepsilon^M(x)| \le \omega(|y-x|) \quad \text{and} \quad \lim_{z\to 0}\omega(z) = 0,$$

the family $\left\{f_\varepsilon^M\right\}_\varepsilon$ is equicontinuous in $C([\delta, 1-\delta])$ for every δ fixed. Hence by Arzelà-Ascoli Theorem, up to subsequences, the family $\left\{f_\varepsilon^M\right\}_\varepsilon$ uniformly converges to a function $\tilde{f}^M$ on $(0,1)$ as $\varepsilon \to 0$ for which it holds

$$|\tilde{f}^M(s) - \tilde{f}^M(t)| \le \int_t^s g(r)\,dr.$$

By the fact that $f_\varepsilon^M \to f^M$ in L^1, $f^M = \tilde{f}^M$ on a $A \subseteq [0,1]$ such that $[0,1]\setminus A$ has negligible Lebesgue measure.

Step 4. Now we prove that $f^M = \tilde{f}^M$ everywhere. f^M is lower semicontinuous and $\tilde{f}^M$ is continuous, hence $f^M \le \tilde{f}^M$ in $[0,1]$. Suppose by contradiction that there are $t_0 \in (0,1)$, $c, C \in \mathbb{R}$ such that $f^M(t_0) < c < C < \tilde{f}^M(t_0)$, so there exists $\delta > 0$ such that $\tilde{f}^M(t) > C$ for $t \in [t_0-\delta, t_0+\delta]$. Thus $f^M(t) > C$ for $t \in [t_0-\delta, t_0+\delta]\cap A$, so

$$\int_0^1 g(t)\,dt \ge \int_{[t_0-\delta,t_0+\delta]\cap A} g(t)\,dt \ge \int_{[t_0-\delta,t_0+\delta]\cap A} \frac{C-c}{|t-t_0|}\,dt = +\infty,$$

which is absurd since $g \in L^1(\mathbb{R})$.

Conclusion. Thus we proved that if $g \in L^1((0,1))$,

$$|f^M(s) - f^M(t)| \le \int_t^s g(r)\,dr, \quad \forall t < s, \forall M > 0.$$

Letting $M \to \infty$ the thesis is proven. □

Notice that an application of Young's inequality on (5) gives

$$E(x_s) - E(x_t) \ge -\frac{1}{2}\int_t^s |\dot{x}_r|^2\,dr - \frac{1}{2}\int_t^s |D^-E|^2(x_r)\,dr, \quad \forall t < s, \tag{6}$$

Therefore, a K-convex functional satisfies the Energy Dissipation Equality if we require only a minimum dissipation of E along the curve, in particular

$$E(x_t) \ge E(x_s) + \frac{1}{2}\int_t^s |\dot{x}_r|^2\,dr + \frac{1}{2}\int_t^s |D^-E|^2(x_r)\,dr, \quad \forall t < s. \tag{7}$$

4. The Entropy functional and existence of its gradient flow

For simplicity, from now on we will restric our analysis to a compact metric space $(X, \mathfrak{d})$.

4.1. **The Wasserstein distance.** We can equip the space of probability measures with a natural distance obtained by the minimization problem of Optimal Transport theory.

Definition 4.1. Given $\mu \in \mathcal{P}(X), \nu \in \mathcal{P}(Y)$, we define the set of *admissible plans* from μ to ν as

$$\mathrm{Adm}(\mu,\nu) := \Big\{\gamma \in \mathcal{P}(X \times Y) : \pi^X_* \gamma = \mu, \pi^Y_* \gamma = \nu\Big\},$$

where π^X, π^Y is the projection on X, Y.

Definition 4.2. Given $\mu, \nu \in \mathcal{P}(X)$, the *Wasserstein distance* between μ and ν is

$$W_2(\mu,\nu) := \sqrt{\inf_{\gamma \in \mathrm{Adm}(\mu,\nu)} \int \mathfrak{d}^2(x,y)\, \gamma(dx,dy)}.$$

The space of probability measures $\mathcal{P}(X)$ endowed with the Wasserstein distance inherits many of the properties of the underlying space X. We point out just two of them:

- Given a sequence $(\mu_n)_n$ in $\mathcal{P}(X)$, it holds

$$W_2(\mu_n,\mu) \to 0 \quad \Leftrightarrow \quad \mu_n \rightharpoonup \mu. \tag{8}$$

  Notice that as a consequence sequential narrow compactness and narrow compactness coincide. (In the case of non compact metric spaces, one needs an additional condition on the right hand side of the equivalence (8).)
- If X is geodesic then $\mathcal{P}(X)$ is also geodesic.

For the proofs and a generalization to non-compact spaces see [2, Theorems 2.7 and 2.10].

Before passing to the definition of Entropy, we prove a property of the metric derivative which we will use later on.

Remark 4.3. If μ_t, ν_t are absolutely continuous curves in $(\mathcal{P}(X), W_2)$, then $\forall k \in [0,1], \eta_t := (1-k)\mu_t + k\nu_t$ is absolutely continuous and it holds

$$|\dot{\eta}_t|^2 \le (1-k)|\dot{\mu}_t|^2 + k|\dot{\nu}_t|^2.$$

In fact, it is easy to prove the convexity of the squared Wasserstein distance w.r.t. linear interpolation of measures, i.e.

$$W_2^2\big((1-t)\mu_0 + t\mu_1, (1-t)\nu_0 + t\nu_1\big) \le (1-t)W_2^2(\mu_0,\nu_0) + tW_2^2(\mu_1,\nu_1)$$

for arbitrary $\mu_0, \mu_1, \nu_0, \nu_1 \in \mathcal{P}(X)$. Applying the definition of metric speed, the estimate above follows immediately.

4.2. **Entropy: definition and properties.**

Definition 4.4. The *Entropy functional* $\mathrm{Ent}_m : \mathcal{P}(X) \to [0,\infty]$ relative to $m \in \mathcal{P}(X)$ is defined as

$$\mathrm{Ent}_m(\mu) := \begin{cases} \displaystyle\int_X f \log f \, m & \text{if } \exists f \in L^1(m) : \mu = fm, \\ \infty & \text{otherwise.} \end{cases}$$

Example 4.6 below suggest that the gradient flow of the Entropy functional in measure metric spaces is the natural extension of the heat flow from the smooth setting. As a consequence, it is possible to construct the analogue of the Laplace operator, which is the starting point to construct several tools used in Analysis on measure metric spaces.

A very interesting fact is that gradient flows of certain functionals in carefully selected spaces generate solutions to well known PDEs (see the seminal papers [8], [10]). In this context we show two interesting examples of solutions to a PDE generated by a gradient flow.

Example 4.5. The gradient flow of the Dirichlet Energy functional $D : L^2(\mathbb{R}^n) \to \mathbb{R}$ defined as

$$D(f) := \frac{1}{2}\|\nabla f\|^2_{L^2}$$

produces a solution of the heat equation in $\mathbb{R}^n$.

We sketch the proof assuming that the functions are smooth. Differentiating $D(f)$ along v we obtain

$$\lim_{t\to 0} \frac{1}{2}\frac{\|\nabla(f+tv)\|^2_{L^2} - \|\nabla f\|^2_{L^2}}{t} = \int \langle \nabla f, \nabla v\rangle,$$

which can be rewritten, using Green's first identity, as $-\int v\Delta f$. Therefore we conclude that

$$-\nabla D(f) = \Delta f.$$

Example 4.6. The heat equation is also obtained as a solution of the Entropy gradient flow in $(\mathcal{P}_2(X), W_2)$. We give an informal proof of this fact in the case $X = \mathbb{T}^n$ the n-torus. We refer to [3, Chapters 8-10] and to [10] for a more detailed approach.

Let $f : \mathbb{T}^n \to \mathbb{R}$ be integrable and s.t. $\int_{\mathbb{T}^n} f\, \mathcal{L} = 1$ where $\mathcal{L}$ is the Lebesgue measure on the torus $\mathbb{T}^n$. Define $\mu := f\mathcal{L}$. The natural space of perturbations (tangent vector fields) in the metric space $(\mathcal{P}(\mathbb{T}^n), W_2)$ in the point $\mu = f\mathcal{L}$ are vector fields $v : \mathbb{T}^n \to \mathbb{T}^n$ square integrable w.r.t. μ, corresponding to the perturbations (see Definition 2.31 of [2])

$$\partial_t f = -\mathrm{div}_x(vf).$$

Inserting these perturbations in $\mathrm{Ent}_{\mathcal{L}}(f)$ we obtain

$$\frac{d}{dt}\mathrm{Ent}_{\mathcal{L}}\big((I+tv)_*(f\mathcal{L})\big) = -\int \mathrm{div}_x(vf)\big(\log f + 1\big)\, dx$$
$$= -\int \mathrm{div}_x(vf)\log f\, dx = \int (vf)\cdot\nabla_x \log f\, dx = \int v\cdot\nabla_x f\, dx.$$

To find the norm of the gradient we have to maximize it with respect to v with the restriction $\int |v|^2 f\, dx \le 1$. With a fairly easy calculation we obtain

$$\bar{v} = -\frac{1}{\alpha}\nabla_x \log f \quad \text{where} \quad \alpha = \sqrt{\int f|\nabla_x \log f|^2\, dx}.$$

From

$$\frac{d}{dt}\mathrm{Ent}_{\mathcal{L}}\big((I+t\bar{v})_*(f\mathcal{L})\big) = -\alpha,$$

we conclude that the gradient flow is $\nabla\,\mathrm{Ent}_{\mathcal{L}}(f) = -\nabla_x \log f$, and finally that

$$\partial_t f = -\mathrm{div}_x((\nabla_x \log f) f) = \mathrm{div}_x(\nabla_x f) = \Delta f.$$

In the following two propositions we prove strict convexity w.r.t. linear interpolation and lower semicontinuity of the Entropy.

Proposition 4.7. *The Entropy functional is strictly convex with respect to linear interpolation of measures, i.e. given* $\mu_0, \mu_1 \in D(\mathrm{Ent}_m)$*,*

$$\mathrm{Ent}_m\big((1-t)\mu_0 + t\mu_1\big) \leq (1-t)\,\mathrm{Ent}_m(\mu_0) + t\,\mathrm{Ent}_m(\mu_1), \quad \forall t \in [0,1],$$

and equality holds if and only if $\mu_0 = \mu_1$*.*

Proof. Let $\mu_0 = f_0 m$, $\mu_1 = f_1 m$, $u(z) := z \log z$. Since u is strictly convex, it holds

$$u((1-t)f_0 + tf_1) \leq (1-t)u(f_0) + tu(f_1), \quad \forall t \in [0,1],$$

and equality holds if and only if $f = g$. Integrating we obtain the thesis. □

Proposition 4.8. *The Entropy functional is lower semicontinuous with respect to narrow convergence of measures.*

Proof. Given $\varphi \in C(X)$ let $G_\varphi : \mathcal{P}(X) \to \mathbb{R}$ be such that

$$G_\varphi(\mu) := \int_X \varphi\,\mu - \int_X e^{\varphi-1}\,m.$$

Notice that G_φ is continuous with respect to narrow convergence by definition. Define now

$$F(\mu) := \sup_{\phi \in C(X)} G_\phi(\mu).$$

If $\mu \perp m$, by varying ϕ we can obtain arbitrary large values for $\int \phi\,\mu$ without increasing $\int e^{\phi-1}\,m$ more than 1; therefore $F(\mu) = \infty$.

If $\mu \ll m$, there exists a non-negative $f \in L^1(m)$ s.t. $\mu = fm$. It is easily verified that $1 + \log f$ maximizes F, therefore by a standard approximation technique we have

$$F(\mu) = \int (1+\log f)f - e^{\log f}\,m = \int f \log f\,m,$$

and

$$F(\mu) = \mathrm{Ent}_m(\mu).$$

But F is the supremum of continuous functions, therefore it is l.s.c.. □

We now define boundedness from below of the Ricci curvature, relying on the definition by Sturm and Lott-Villani (see [11] and [9]).

Definition 4.9. Let $(X, \mathfrak{d}, m)$ be a compact geodesic metric space with $m \in \mathcal{P}(X)$. X has *Ricci curvature* bounded from below by $K \in \mathbb{R}$, and we write $(X, \mathfrak{d}, m)$ is a $CD(K,\infty)$ space, if Ent_m is K-convex in $(\mathcal{P}(X), W_2)$, i.e. for every pair of points $\mu_0, \mu_1 \in D(\mathrm{Ent}_m)$, there exists $\mu_t : [0,1] \to \mathcal{P}(X)$ constant speed geodesic between μ_0 and μ_1, such that

$$\mathrm{Ent}_m(\mu_t) \leq (1-t)\,\mathrm{Ent}_m(\mu_0) + t\,\mathrm{Ent}_m(\mu_1) - \frac{K}{2}t(1-t)W_2^2(\mu_0,\mu_1), \quad \forall t \in [0,1].$$

The previous definition will allow us to apply the properties of K-convex functionals to the Entropy.

4.3. **Existence of the gradient flow of the Entropy.** The proof of the existence of a gradient flow of the Entropy functional relies on a variational approach which dates back to De Giorgi (see [5]). We will define a recursive scheme and prove that it converges to a solution of the gradient flow of the Entropy. The recursive scheme is obtained with the following functional.

Definition 4.10. For $\tau > 0$, and $\mu \in \mathcal{P}(X)$ define $J_\tau\mu$ as the probability $\sigma \in \mathcal{P}(X)$ which minimizes

$$\sigma \mapsto \mathrm{Ent}_m(\sigma) + \frac{W_2^2(\sigma,\mu)}{2\tau}.$$

Since $P(X)$ is compact w.r.t. narrow convergence, the existence and uniqueness of the minimizer is obtained through lower semicontinuity and strict convexity of the Entropy.

We prove an important estimate on the curve $t \mapsto J_t\mu$, which is almost the EDE we are looking for.

Lemma 4.11. *Let $\mu \in \mathcal{P}(X)$. Then it holds*

$$\mathrm{Ent}_m(\mu) = \mathrm{Ent}_m(J_t\mu) + \frac{W_2^2(J_t\mu,\mu)}{2t} + \int_0^t \frac{W_2^2(J_r\mu,\mu)}{2r^2}\,dr. \tag{9}$$

Proof. By the definition of $J_t\mu$ we have

$$\begin{aligned}
&\frac{W_2^2(J_t\mu,\mu)}{2t} - \frac{W_2^2(J_t\mu,\mu)}{2s} \\
&\le \mathrm{Ent}_m(J_t\mu) + \frac{W_2^2(J_t\mu,\mu)}{2t} - \left(\mathrm{Ent}_m(J_s\mu) + \frac{W_2^2(J_s\mu,\mu)}{2s}\right) \\
&\le \frac{W_2^2(J_s\mu,\mu)}{2t} - \frac{W_2^2(J_s\mu,\mu)}{2s},
\end{aligned}$$

and passing to the limit we obtain

$$\begin{aligned}
&\lim_{s\to t} \frac{1}{t-s}\left(\mathrm{Ent}_m(J_t\mu) + \frac{W_2^2(J_t\mu,\mu)}{2t} - \left(\mathrm{Ent}_m(J_s\mu) + \frac{W_2^2(J_s\mu,\mu)}{2s}\right)\right) \\
= &-\frac{W_2^2(J_t\mu,\mu)}{2t^2}.
\end{aligned}$$

Since the left hand side is the derivative of

$$r \mapsto \mathrm{Ent}_m(J_r\mu) + \frac{W_2^2(J_r\mu,\mu)}{2r},$$

by integration we obtain

$$\int_0^t \frac{d}{dr}\left(\mathrm{Ent}_m(J_r\mu) + \frac{W_2^2(J_r\mu,\mu)}{2r}\right) dr = -\int_0^t \frac{W_2^2(J_r\mu,\mu)}{2r^2}\,dr.$$

If we show that

$$\lim_{x\to 0^+} \frac{W_2^2(J_x\mu,\mu)}{2x} = 0,$$

we will thus have the thesis. In fact, since $\mathrm{Ent}_m(\mu) < \infty$, then by definition of $J_r\mu$ we have

$$0 \le \mathrm{Ent}_m(J_r\mu) + \frac{W_2^2(J_r\mu,\mu)}{2r} \le \mathrm{Ent}_m(\mu),$$

and the lower semicontinuity of the Entropy yields $\mathrm{Ent}_m(J_r\mu) \ge \mathrm{Ent}_m(\mu)$ as $r \to 0$. $\square$

We are ready to prove the first main result of this chapter, i.e. the existence of an EDE-gradient flow of the Entropy for metric measure spaces with Ricci curvature bounded from below.

Theorem 4.12 (Existence)**.** *If $(X, \mathfrak{d}, m)$ has Ricci curvature bounded from below by K then for every $\tilde{\mu} \in D(\mathrm{Ent}_m)$ there exists a gradient flow of Ent_m starting from $\tilde{\mu}$.*

Proof. The proof will be given in several steps. Notice that in order to prove the EDE, by (7) it is enough to show there exists an absolutely continuous curve $t \mapsto \mu_t$ such that $\mu_0 = \tilde{\mu}$ and

$$\mathrm{Ent}_m(\tilde{\mu}) \geq \mathrm{Ent}_m(\mu_t) + \frac{1}{2}\int_0^t |\dot{\mu}|^2(r)\,dr + \frac{1}{2}\int_0^t |D^- \mathrm{Ent}_m|^2(\mu_r)\,dr, \quad \forall t \geq 0.$$

Step 1. The approximate solution is constructed by defining recursively

$$\begin{aligned} \mu_0^\tau &:= \tilde{\mu}, \\ \mu_{n+1}^\tau &:= J_\tau(\mu_n^\tau). \end{aligned}$$

Then define the curve $t \mapsto \mu^\tau(t)$ as

$$\begin{aligned} \mu^\tau(n\tau) &:= \mu_n^\tau, \\ \mu^\tau(t) &:= J_{t-n\tau}(\mu_n^\tau), \quad \forall t \in (n\tau, (n+1)\tau), \end{aligned}$$

and let

$$|\dot{\mu}^\tau|(t) := \frac{W_2(\mu_n^\tau, \mu_{n+1}^\tau)}{2\tau}, \quad \forall t \in [n\tau, (n+1)\tau).$$

Step 2. We give an estimate on the descending slope of the Entropy. Given $\nu, \sigma \in \mathcal{P}(X)$, since $J_t(\sigma)$ is the minimizer, we have

$$\mathrm{Ent}_m(J_t\sigma) + \frac{W_2^2(\sigma, J_t\sigma)}{2t} \leq \mathrm{Ent}_m(\nu) + \frac{W_2^2(\sigma, \nu)}{2t}.$$

Hence by triangle inequality

$$\begin{aligned} \mathrm{Ent}_m(J_t\sigma) - \mathrm{Ent}_m(\nu) &\leq \frac{1}{2t}\Big(W_2^2(\sigma,\nu) - W_2^2(\sigma, J_t\sigma)\Big) \\ &= \frac{1}{2t}\Big(W_2(\sigma,\nu) - W_2(\sigma, J_t\sigma)\Big)\Big(W_2(\nu,\sigma) + W_2(\sigma, J_t\sigma)\Big) \\ &\leq \frac{W_2(J_t\sigma,\nu)}{2t}\Big(W_2(\nu,\sigma) + W_2(\sigma, J_t\sigma)\Big). \end{aligned}$$

If $J_t\sigma = \nu$ the inequality holds trivially. Otherwise, dividing by $W_2(J_t\sigma, \nu)$ both sides and passing to the limit

$$\begin{aligned} |D^- \mathrm{Ent}_m|(J_t\sigma) &= \limsup_{\nu \to J_t\sigma} \frac{\big(\mathrm{Ent}_m(J_t\sigma) - \mathrm{Ent}_m(\nu)\big)^+}{W_2(J_t\sigma, \nu)} \\ &\leq \limsup_{\nu \to J_t\sigma} \frac{W_2(\nu,\sigma) + W_2(\sigma, J_t\sigma)}{2t} = \frac{W_2(\sigma, J_t\sigma)}{t}. \end{aligned}$$

Step 3. Using now the curve $\mu^\tau(t)$, the definition of its time derivative $|\dot{\mu}^\tau|$ and the previous inequality, we can thus rewrite (9) as

$$\mathrm{Ent}_m(\mu_n^\tau) \geq \mathrm{Ent}_m(\mu_{n+1}^\tau) + \frac{1}{2}(2\tau)|\dot{\mu}^\tau|^2(t) + \frac{1}{2}\int_{n\tau}^{(n+1)\tau} |D^- \mathrm{Ent}_m|^2\big(\mu^\tau(s)\big)\,ds.$$

where $t \in [n, n+1)\tau$. Adding these inequalities from 0 to $T = N\tau$ we obtain

$$\operatorname{Ent}_m(\tilde{\mu}) \geq \operatorname{Ent}_m(\mu_T^\tau) + \frac{1}{2}\int_0^T |\dot{\mu}^\tau|^2(t)\, dt + \frac{1}{2}\int_0^T |D^- \operatorname{Ent}_m|^2\big(\mu^\tau(t)\big)\, dt. \tag{10}$$

Notice that the definition of $|\dot{\mu}^\tau|$ implies that we rescale time as $t \mapsto t/2$, i.e. we take 2τ to pass from μ_n^τ to μ_{n+1}^τ.

The last part of the proof concerns the compactness of the family of curves $\{t \mapsto \mu_t^\tau\}_\tau$ in the space $C(X, \mathcal{P}(X))$ and the lower semicontinuity (with respect to τ in 0) of the right hand side of (10). These results clearly will conclude the proof.

Step 4. We address the convergence of the curve $t \mapsto \mu^\tau(t)$ as $\tau \to 0$. First, since the starting point is in the domain of the Entropy we have by (10) that

$$\int_0^T |\dot{\mu}^\tau|^2(t)\, dt < 2\operatorname{Ent}_m(\tilde{\mu}).$$

By Hölder inequality, for all Borel $A \subset [0, T]$

$$\begin{aligned}\int_A |\dot{\mu}^\tau|(t)\, dt = \int_0^T \chi_A |\dot{\mu}^\tau|(t)\, dt &\leq \sqrt{\int_0^T \chi_A^2\, dt}\sqrt{\int_0^T |\dot{\mu}^\tau|^2(t)\, dt} \\ &\leq \sqrt{\mathcal{L}(A)}\sqrt{2\operatorname{Ent}_m(\tilde{\mu})},\end{aligned}$$

which gives the uniform integrability of $|\dot{\mu}^\tau(t)|$. Since

$$W_2(\mu^\tau(s), \mu^\tau(t)) \leq \int_t^s |\dot{\mu}^\tau|(r)\, dr,$$

the family of curves $(\mu^\tau)_\tau$ is uniformly continuous.
Up to subsequences we can pass to the limit as $\tau_n \searrow 0$, obtaining by Arzelà-Ascoli Theorem that μ^{τ_n} converges uniformly to a curve $t \mapsto \mu(t)$ such that $\mu(0) = \tilde{\mu}$. Since $|\dot{\mu}^\tau(t)|$ is uniformly integrabile, up to subsequences, also $|\dot{\mu}^{\tau_n}(t)|$ L^1-weakly converges to a function g. It follows easily from the definition of metric derivative that $|\dot{\mu}(t)| \leq g(t)$, therefore $\mu(t)$ is locally absolutely continuous.

Step 5. We prove the lower semicontinuity of the right hand side of (10).
By Hölder's inequality we have

$$\int_0^T |\dot{\mu}|^2(t)\, dt \leq \int_0^T g(t)^2\, dt \leq \liminf_n \int_0^T |\dot{\mu}^{\tau_n}|^2(t)\, dt,$$

thus the l.s.c. of the first integral.

Notice that being the supremum of lower semicontinuous functions by formula (4), $|D^- \operatorname{Ent}_m|$ is lower semicontinuous too. Then define for every $k \in \mathbb{N}, \nu \in \mathcal{P}(X)$

$$e_k(\nu) := \inf_{\sigma \in \mathcal{P}(X)} \Big\{ |D^- \operatorname{Ent}_m|^2(\sigma) + k W_2(\nu, \sigma) \Big\}.$$

Notice that $\sup_k e_k(\nu) = |D^- \operatorname{Ent}_m|^2(\nu)$ for all ν.
The infimum of Lipschitz functions bounded from below is a Lipschitz function; therefore e_k is Lipschitz. By uniform convergence

$$\lim_n \int_0^T e_k(\mu^{\tau_n}(t))\, dt = \int_0^T e_k(\mu(t))\, dt.$$

By Fatou Lemma

$$\int e_k(\mu(t))\, dt \leq \liminf_n \int |D^- \operatorname{Ent}_m|^2(\mu^{\tau_n}(t))\, dt,$$

for every $k \in \mathbb{N}$. By Monotone Convergence Theorem we finally have

$$\int |D^- \operatorname{Ent}_m|^2(\mu(t))\, dt = \sup_k \int e_k(\mu(t))\, dt \le \liminf_n \int |D^- \operatorname{Ent}_m|^2(\mu^{\tau_n}(t))\, dt.$$

□

5. Uniqueness of the gradient flow of the Entropy

For the proof of the uniqueness we will follow the argumentation of [4, Section 5] and [7, Section 3].

We start by introducing two new concepts, the push-forward via a plan, and plans with bounded deformation. We prove different interesting auxiliary properties which correlate these two new concepts with the Entropy functional and its descending slope, and the key result in Proposition 5.10. We conclude the section proving the uniqueness of the gradient flow of the Entropy.

5.1. **Plans with bounded deformation and push-forward via a plan.** We extend the notion of push-forward via a map as follows.

Definition 5.1 (Push-forward via a plan)**.** Let $\mu \in \mathcal{P}(X), \gamma \in \mathcal{P}(X^2)$ be such that $\mu \ll \pi^1_* \gamma$. The measures $\gamma_\mu \in \mathcal{P}(X^2)$ and $\gamma_* \mu \in \mathcal{P}(X)$ are defined as

$$\gamma_\mu(dx, dy) := \frac{d\mu}{d\pi^1_*\gamma}(x)\gamma(dx, dy),$$
$$\gamma_*\mu(dy) := \pi^2_*\gamma_\mu(dy).$$

Remark 5.2. Since $\mu \ll \pi^1_*\gamma$, there exists $f \in L^1(\pi^1_*\gamma)$ such that $\mu = f\pi^1_*\gamma$. By Disintegration Theorem, considering $\{\gamma_y\}_{y\in X}$ the disintegration of γ with respect to its second marginal we obtain

$$\gamma_*\mu(dy) = \left(\int_X f(x)\ \gamma_y(dx)\right) \pi^2_*\gamma(dy). \tag{11}$$

The plans which are particularly useful in our analysis belong to the following category.

Definition 5.3. The plan $\gamma \in \mathcal{P}(X^2)$ has *bounded deformation* if $\exists c \in \mathbb{R}^+$ such that $\frac{1}{c}m \le \pi^1_*\gamma, \pi^2_*\gamma \le cm$.

We show now some preliminary properties. The following is a useful estimate.

Proposition 5.4. $\forall \mu, \nu \in \mathcal{P}(X), \forall \gamma \in \mathcal{P}(X^2)$ *such that* $\mu, \nu \ll \pi^1_*\gamma$,

$$\operatorname{Ent}_{\gamma_*\nu}(\gamma_*\mu) \le \operatorname{Ent}_\nu(\mu).$$

Proof. We assume $\mu \ll \nu$, otherwise $\operatorname{Ent}_\nu(\mu) = \infty$ and there is nothing to prove. Then there exists $f \in L^1(\nu)$ such that $\mu = f\nu$ and since $\nu \ll \pi^1_*\gamma$ by hypothesis, $\exists \theta \in L^1(\pi^1_*\theta)$ such that $\nu = \theta\pi^1_*\gamma$. Disintegrating $\gamma_*\nu, \gamma_*\mu$ as in (11) we obtain

$$\gamma_*\mu = \left(\int_X f(x)\theta(x)\ \gamma_y(dx)\right) \pi^2_*\gamma, \tag{12}$$

$$\gamma_*\nu = \left(\int_X \theta(x)\ \gamma_y(dx)\right) \pi^2_*\gamma. \tag{13}$$

It is easily verified that $\gamma_*\mu \ll \gamma_*\nu$. Therefore by Radon-Nikodym Theorem there exists $\eta \in L^1(\gamma_*\mu)$ such that $\gamma_*\mu = \eta\gamma_*\nu$, and considering (12), (13), we have

$$\eta(y) = \frac{\int f\theta\, \gamma_y(dx)}{\int \theta\, \gamma_y(dx)} = \int f \frac{\theta}{\int \theta\, \gamma_y(dx)}\, \gamma_y(dx). \tag{14}$$

Defining

$$\tilde{\gamma} := (\theta \circ \pi^1)\gamma,$$

its disintegration with respect to its second marginal $\gamma_*\nu$ is

$$\tilde{\gamma}_y = \frac{\theta}{\int \theta\, \gamma_y(dx)}\gamma_y,$$

so we can rewrite (14) as

$$\eta(y) = \int f\, \tilde{\gamma}_y(dx).$$

Now let $u(z) := z \log z$. From the convexity of $u(z)$ and Jensen's inequality,

$$u(\eta(y)) \leq \int u(f(x))\, \tilde{\gamma}_y(dx).$$

Integrating both sides with respect to $\gamma_*\nu$ we get

$$\mathrm{Ent}_{\gamma_*\nu}(\gamma_*\mu) = \int u(\eta(y))\, \gamma_*\nu(dy) \leq \int \left(\int u(f(x))\, \tilde{\gamma}_y(dx) \right) \gamma_*\nu(dy),$$

and from Disintegration Theorem,

$$\int \left(\int u(f(x))\, \tilde{\gamma}_y(dx) \right) \gamma_*\nu(dy) = \int u(f(x))\, \nu(dx) = \mathrm{Ent}_\nu(\mu).$$

□

The following formula will be useful in the proof of the next proposition.

Lemma 5.5. *If $\mu, \nu, \sigma \in \mathcal{P}(X)$ and σ is such that there exists $c > 0 : \frac{1}{c}\nu \leq \sigma \leq c\nu$, then it holds*

$$\mathrm{Ent}_\nu(\mu) = \mathrm{Ent}_\sigma(\mu) + \int_X \log\left(\frac{d\sigma}{d\nu}\right) \mu(dx). \tag{15}$$

Proof. From the hypothesis on σ we can deduce there exists $\frac{1}{c} \leq g \leq c$ such that $\sigma = g\nu$. If μ is not absolutely continuous with respect to ν we obtain $\infty = \infty + C$ and (15) holds.

Otherwise if $\mu \ll \nu$ take $\mu = f\nu$; therefore

$$\mu = \frac{f}{g}\sigma$$

and

$$\mathrm{Ent}_\sigma(\mu) + \int_X \log\left(\frac{d\sigma}{d\nu}\right) \mu = \int_X f \log f\, \nu = \mathrm{Ent}_\nu(\mu).$$

□

The push-forward of a measure via a plan with bounded deformation allows us to remain in the domain of the Entropy, as it is proved in the next results.

Proposition 5.6. *If $\mu \in D(\mathrm{Ent}_m)$ and $\gamma \in \mathcal{P}(X^2)$ has bounded deformation, then $\gamma_*\mu \in D(\mathrm{Ent}_m)$.*

Proof. Since γ has bounded deformation there exist $c, C > 0$ such that $cm \leq \pi^1_*\gamma, \pi^2_*\gamma \leq Cm$. Using identity (15) we obtain

$$\begin{aligned}\mathrm{Ent}_m(\gamma_*\mu) =&\ \mathrm{Ent}_{\pi^2_*\gamma}(\gamma_*\mu) + \int_X \log\left(\frac{d\pi^2_*\gamma}{dm}\right)\ \gamma_*\mu \\ \leq&\ \mathrm{Ent}_{\pi^2_*\gamma}(\gamma_*\mu) + \log(C).\end{aligned}$$

From the fact that $\gamma_*(\pi^1_*\gamma) = \pi^2_*(\gamma_{\pi^1_*\gamma}) = \pi^2_*(\gamma)$ and Proposition 5.4,

$$\mathrm{Ent}_{\pi^2_*\gamma}(\gamma_*\mu) = \mathrm{Ent}_{\gamma_*(\pi^1_*\gamma)}(\gamma_*\mu) \leq \mathrm{Ent}_{\pi^1_*\gamma}(\mu).$$

Then using again identity (15)

$$\begin{aligned}\mathrm{Ent}_{\pi^1_*\gamma}(\mu) = \mathrm{Ent}_m(\mu) + \int_X \log\left(\frac{dm}{d\pi^1_*\gamma}\right)\ \mu \\ \leq\ \mathrm{Ent}_m(\mu) + \mu(X)\log\frac{1}{c} = \mathrm{Ent}_m(\mu) - \log c.\end{aligned}$$

In conclusion

$$\mathrm{Ent}_m(\gamma_*\mu) \leq \mathrm{Ent}_m(\mu) - \log c + \log C < \infty.$$

□

The following proposition gives a quite unexpected property of convexity of the Entropy.

Proposition 5.7. *If $\gamma \in \mathcal{P}(X^2)$ has bounded deformation then the map*

$$D(\mathrm{Ent}_m) \ni \mu \mapsto \mathrm{Ent}_m(\mu) - \mathrm{Ent}_m(\gamma_*\mu)$$

is convex with respect to linear interpolation of measures

Proof. Let $\mu_0 = f_0 m, \mu_1 = f_1 m$. Define for every $t \in (0,1)$

$$\begin{aligned}\mu_t &:= (1-t)\mu_0 + t\mu_1, \\ f_t &:= (1-t)f_0 + tf_1.\end{aligned}$$

We compute

$$\begin{aligned}&(1-t)\,\mathrm{Ent}_{\mu_t}(\mu_0) + t\,\mathrm{Ent}_{\mu_t}(\mu_1) \\ &= (1-t)\int_X \frac{f_0}{f_t}\log\left(\frac{f_0}{f_t}\right)\ \mu_t + t\int_X \frac{f_1}{f_t}\log\left(\frac{f_1}{f_t}\right)\ \mu_t \\ &= (1-t)\int_X f_0\log f_0\ m + t\int_X f_1\log f_1\ m - \int_X f_t\log f_t\ m \\ &= (1-t)\,\mathrm{Ent}_m(\mu_0) + t\,\mathrm{Ent}_m(\mu_1) - \mathrm{Ent}_m(\mu_t).\end{aligned}$$

Since $\mu_i \in D(\mathrm{Ent})$ (for $i = 1, 2$) and γ has bounded deformation, from Proposition 5.6 also $\gamma_*\mu_i \in D(\mathrm{Ent})$, so an identical argument with μ_t replaced by $\gamma_*\mu_t$ shows that

$$\begin{aligned}&(1-t)\,\mathrm{Ent}_{\gamma_*\mu_t}(\gamma_*\mu_0) + t\,\mathrm{Ent}_{\gamma_*\mu_t}(\gamma_*\mu_1) \\ &= (1-t)\,\mathrm{Ent}_m(\gamma_*\mu_0) + t\,\mathrm{Ent}_m(\gamma_*\mu_1) - \mathrm{Ent}_m(\gamma_*\mu_t).\end{aligned}$$

By Proposition 5.4 we have $\mathrm{Ent}_{\gamma_*\mu_t}(\gamma_*\mu_i) \leq \mathrm{Ent}_{\mu_t}(\mu_i)$ for $i = 1, 2$, therefore

$$\begin{aligned}&(1-t)\,\mathrm{Ent}_m(\gamma_*\mu_0) + t\,\mathrm{Ent}_m(\gamma_*\mu_1) - \mathrm{Ent}_m(\gamma_*\mu_t) \\ &\leq (1-t)\,\mathrm{Ent}_m(\mu_0) + t\,\mathrm{Ent}_m(\mu_1) - \mathrm{Ent}_m(\mu_t).\end{aligned}$$

Rearranging the terms we finally obtain

$$\begin{aligned}&\operatorname{Ent}_m(\mu_t) - \operatorname{Ent}_m(\gamma_*\mu_t)\\ &\leq (1-t)\operatorname{Ent}_m(\mu_0) + t\operatorname{Ent}_m(\mu_1) - (1-t)\operatorname{Ent}_m(\gamma_*\mu_0) - t\operatorname{Ent}_m(\gamma_*\mu_1).\end{aligned}$$

□

5.2. **Approximability in Entropy and distance.** The following is a technical result which allows us to control the Entropy of a perturbation of a measure.

For $\gamma \in \mathcal{P}(X^2)$, define the *transportation cost*

$$C(\gamma) := \int \mathfrak{d}^2(x,y)\,\gamma(dx,dy).$$

Lemma 5.8. *If $\mu, \nu \in D(\operatorname{Ent}_m)$, there exists a sequence $(\gamma^n)_n$ of plans with bounded deformation such that $\operatorname{Ent}_m(\gamma^n_*\mu) \to \operatorname{Ent}_m(\nu)$ and $C(\gamma^n_\mu) \to W_2^2(\mu,\nu)$ as $n \to \infty$.*

Proof. Let $f, g \in L^1(m)$ be non-negative such that $\mu = fm, \nu = gm$. Pick $\gamma \in \operatorname{Adm}(\mu,\nu)$ s.t.

$$\int \mathfrak{d}^2\,\gamma = \inf_{\gamma' \in \operatorname{Adm}(\mu,\nu)} \int \mathfrak{d}^2\,\gamma'$$

and $\forall n \in \mathbb{N}$ define

$$\begin{aligned}A_n' &:= \Big\{(x,y) \in X^2 : f(x) + g(y) \leq n\Big\},\\ A_n &:= \Big\{(x,y) \in A_n' : \gamma_x(A_n') > \tfrac{1}{2}\Big\},\\ \gamma^n(dx,dy) &:= c_n\left(\gamma_{|A_n}(dx,dy) + \frac{1}{n}(id,id)_* m(dx,dy)\right),\end{aligned}$$

with c_n the normalization constant, i.e. $c_n = \dfrac{1}{\gamma(A_n) + \frac{1}{n}}$. Disintegrating γ we obtain

$$\begin{aligned}\gamma(dx,dy) &= \int_X \big[\gamma_x(dy)\big]\,\mu(dx) = \int_X \big[\gamma_y(dx)\big]\,\nu(dy),\\ \gamma_{|A_n}(dx,dy) &= \int_X \big[\gamma_{x\,|A_n}(dy)\big]\,\mu(dx) = \int_X \big[\gamma_{y\,|A_n}(dx)\big]\,\nu(dy).\end{aligned}$$

Therefore

$$\begin{aligned}\frac{\gamma^n}{c_n} &= \left(\int_X \big[\gamma_{x\,|A_n}(dy)\big]\,\mu(dx)\right) + \frac{1}{n}(id,id)_* m(dx,dy)\\ &= \int_X \left[f(x)\gamma_{x\,|A_n}(dy) + \frac{1}{n}\delta_x(dy)\right]\,m(dx),\end{aligned}$$

and analogously for ν,

$$\frac{\gamma^n}{c_n} = \int_X \left[g(y)\gamma_{y\,|A_n}(dx) + \frac{1}{n}\delta_y(dx)\right]\,m(dy).$$

Then the marginals of γ^n are

$$\begin{aligned}\pi^1_*\gamma^n &= c_n\left(\gamma_x(A_n)f(x) + \frac{1}{n}\right)m(dx),\\ \pi^2_*\gamma^n &= c_n\left(\gamma_y(A_n)g(y) + \frac{1}{n}\right)m(dy).\end{aligned}$$

Since $0 \le f, g \le n$ and $0 \le \gamma_x(A_n), \gamma_y(A_n) \le 1$,

$$\frac{c_n}{n} m \le \pi^1_* \gamma^n, \pi^2_* \gamma^n \le \left(n c_n + \frac{c_n}{n} \right) m,$$

i.e. γ^n has bounded deformation for every n.

By Radon-Nikodym Theorem

$$f_n(x) := \frac{d\mu}{d\pi^1_* \gamma^n} = \frac{f(x)}{c_n(\gamma_x(A_n) f(x) + \frac{1}{n})}, \tag{16}$$

so by definition of push-forward of a measure

$$\begin{aligned} \gamma^n_\mu(dx, dy) &= \frac{d\mu}{d\pi^1_* \gamma^n} \gamma^n = f_n(x) \gamma^n(dx, dy) \\ &= \int_X c_n \left[f_n(x) g(y) \gamma_{y\,|A_n}(dx) + \frac{f_n(x)}{n} \delta_y(dx) \right] m(dy), \end{aligned}$$

and thus

$$\gamma^n_* \mu = \pi^2_* \gamma^n_\mu(dy) = \int_X c_n \left(g(y) \int_{A_n} f_n(x)\, \gamma_y(dx) + \frac{f_n(y)}{n} \right) m(dy).$$

Defining

$$h_n(y) := g(y) \left(\int_{A_n} f_n(x)\, \gamma_y(dx) \right) + \frac{f_n(y)}{n},$$

we can write

$$\mathrm{Ent}_m(\gamma^n_* \mu) = \int_X c_n h_n(y) \log \left[c_n h_n(y) \right] m(dy).$$

We notice that since $c_n \to 1$, $c_n \ge 2/3$ definitely, so

$$c_n \gamma_x(A_n) + \frac{1}{n f(x)} \ge \frac{1}{2} c_n + \frac{1}{n^2} \ge \frac{1}{3},$$

thus by definition (16), $f_n \le 3$ definitely.
Therefore, defined

$$q_n(y) := 3 \left(g(y) + \frac{1}{n} \right),$$

we have that $0 \le h_n \le q_n(y)$ definitely. A calculation shows that $q_n(y) \log q_n(y) \in L^1(m)$ definitely; thus by Dominated Convergence Theorem

$$\begin{aligned} \lim_n \mathrm{Ent}_m(\gamma^n_* \mu) &= \int_X \lim_n \Big(c_n h_n(y) \log \left(c_n h_n(y) \right) \Big) m(dy) \\ &= \int_X g(y) \log g(y)\; m(dy) = \mathrm{Ent}_m(\nu), \end{aligned}$$

since $\lim_n c_n = 1$ and $\lim_n h_n(y) = g(y)$.

We pass to show the convergence of the cost. We can rewrite the cost of γ^n_μ as

$$\begin{aligned} C(\gamma^n_\mu) = \int_{X^2} \mathfrak{d}^2(x, y)\; \gamma^n_\mu(dx, dy) &= \int_{X^2} \mathfrak{d}^2(x, y) f_n\; \gamma^n(dx, dy) \\ &= \int_{X^2} \mathfrak{d}^2 c_n f_n \chi_{A_n}\; \gamma + \frac{1}{n} \int_{X^2} \mathfrak{d}^2 f_n\; (id, id)_* m. \end{aligned}$$

By hypothesis X is compact and we have that $c_n, f_n, \chi_{A_n} \to 1$, therefore there exists $k > 1$ such that $\mathfrak{d}^2 c_n f_n \chi_{A_n} \leq k$ definitely. In conclusion, by Dominated Convergence

$$C(\gamma_\mu^n) \to \int_{X^2} \mathfrak{d}^2 \, \gamma = W_2^2(\mu, \nu).$$

□

5.3. **Convexity of the squared descending slope.** If $(X, \mathfrak{d}, m)$ has Ricci curvature bounded from below by K, from (4) we know that

$$|D^- \operatorname{Ent}_m|(\mu) = \sup_{\nu \in \mathcal{P}(X),\, \nu \neq \mu} \frac{\left(\operatorname{Ent}_m(\mu) - \operatorname{Ent}_m(\nu) + \frac{K}{2} W_2^2(\mu, \nu)\right)^+}{W_2(\mu, \nu)}.$$

We give yet another characterization of $|D^- \operatorname{Ent}_m|$, which relies only on plans with bounded deformation and which we will use in the proof of Proposition 5.10.

Lemma 5.9. *If $(X, \mathfrak{d}, m)$ has Ricci curvature bounded from below by K then*

$$|D^- \operatorname{Ent}_m|(\mu) = \sup_{\gamma} \frac{\left(\operatorname{Ent}_m(\mu) - \operatorname{Ent}_m(\gamma_* \mu) + \frac{K}{2} C(\gamma_\mu)\right)^+}{\left(C(\gamma_\mu)\right)^{1/2}},$$

where the supremum is taken among all $\gamma \in \operatorname{Adm}(\mu, \nu)$ with bounded deformation, and if $C(\gamma_\mu) = 0$ the right hand side is taken 0 by definition.

Proof. We show both inequalities.

" $\geq$ ". We can assume $C(\gamma_\mu) > 0$, $\nu = \gamma_* \mu$ and $K < 0$ (thanks to Remark 3.8). The following inequality is easily proven: if $a, b, c \in \mathbb{R}$ and $0 < b \leq c$, then

$$\frac{(a-b)^+}{\sqrt{b}} \geq \frac{(a-c)^+}{\sqrt{c}}.$$

Substituting

$$\begin{aligned} a &:= \operatorname{Ent}_m(\mu) - \operatorname{Ent}_m(\gamma_* \mu), \\ b &:= -\frac{K}{2} W_2^2(\mu, \gamma_* \mu), \\ c &:= -\frac{K}{2} C(\gamma_\mu), \end{aligned}$$

proves the thesis.

" $\leq$ ". It comes directly from Lemma 5.8.

□

A key ingredient in proving the uniqueness of the flow of the Entropy is the convexity of the squared descending slope of the Entropy, which we now show.

Proposition 5.10. *If (X, d, m) has Ricci curvature bounded from below by K, then the map*

$$\mu \in D(\operatorname{Ent}_m) \mapsto |D^- \operatorname{Ent}_m|^2(\mu)$$

is convex with respect to linear interpolation of measures.

Proof. Recalling that the supremum of convex maps is still convex, and considering Lemma 5.9, we are done if we prove that the map

$$\mu \mapsto \frac{\left(\left(\mathrm{Ent}_m(\mu) - \mathrm{Ent}_m(\gamma_*\mu) + \frac{K^-}{2}C(\gamma_\mu)\right)^+\right)^2}{C(\gamma_\mu)} \tag{17}$$

is convex.
The map

$$\mu \in D(\mathrm{Ent}_m) \mapsto C(\gamma_\mu) = \int_{X\times X} d^2(x,y)\, d\gamma_\mu$$

is linear. Hence, together with the fact that $\mu \mapsto \mathrm{Ent}_m(\mu) - \mathrm{Ent}_m(\gamma_*\mu)$ is convex (Proposition 5.7), also

$$\mu \mapsto \mathrm{Ent}_m(\mu) - \mathrm{Ent}_m(\gamma_*\mu) - \frac{K^-}{2}C(\gamma_\mu)$$

is convex. Taking its positive part we still have a convex function.

Now take

$$a(\mu) := \left(\mathrm{Ent}_m(\mu) - \mathrm{Ent}_m(\gamma_*\mu) - \frac{K^-}{2}C(\gamma_\mu)\right)^+, \; b(\mu) := C(\gamma_\mu),$$

and define also $\psi : [0,\infty)\times[0,\infty) \to \mathbb{R}\cup\{\infty\}$ as

$$\psi(a,b) := \begin{cases} \frac{a^2}{b} & \text{if } b>0, \\ \infty & \text{if } b=0,\ a>0, \\ 0 & \text{if } a=b=0. \end{cases}$$

It is immediately shown that ψ is convex and it is non-decreasing with respect to a. Therefore we obtain

$$\begin{aligned} &\psi\Big(a\big((1-t)\mu_0 + t\mu_1\big), b\big((1-t)\mu_0 + t\mu_1\big)\Big) \\ &\le \psi\Big((1-t)a(\mu_0) + ta(\mu_1), (1-t)b(\mu_0) + tb(\mu_1)\Big) \\ &\le (1-t)\psi\Big(a(\mu_0), b(\mu_0)\Big) + t\psi\Big(a(\mu_1), b(\mu_1)\Big), \end{aligned}$$

thus the convexity of (17). □

We finally have all the tools to prove the uniqueness of the gradient flow generated by the Entropy.

Theorem 5.11 (Uniqueness)**.** *Let (X,d,m) have Ricci curvature bounded from below by K and let $\tilde\mu \in D(\mathrm{Ent}_m)$; then there exists a unique gradient flow of Ent_m in $(\mathcal{P}(X), W_2)$ starting from $\tilde\mu$.*

Proof. Let μ_t, ν_t be gradient flows of Ent_m starting both from $\tilde\mu$. Then

$$\eta_t := \frac{\mu_t + \nu_t}{2}$$

is an absolutely continuous curve (by Remark 4.3) starting from $\tilde\mu$. From the definition of gradient flow,

$$\mathrm{Ent}_m(\tilde\mu) = \mathrm{Ent}_m(\mu_t) + \frac{1}{2}\int_0^t |\dot\mu_s|^2\, ds + \frac{1}{2}\int_0^t |D^-\mathrm{Ent}_m|^2(\mu_s)\, ds,$$

$$\mathrm{Ent}_m(\tilde{\mu}) = \mathrm{Ent}_m(\nu_t) + \frac{1}{2}\int_0^t |\dot{\nu}_s|^2 \, ds + \frac{1}{2}\int_0^t |D^- \mathrm{Ent}_m|^2(\nu_s) \, ds,$$

for every $t \geq 0$.

Adding up these two equalities, by the squared slope convexity (Proposition 5.10), the squared metric speed convexity (Remark 4.3) and the strict convexity of the relative Entropy (Proposition 4.7), we obtain

$$\mathrm{Ent}_m(\tilde{\mu}) > \mathrm{Ent}_m(\eta_t) + \frac{1}{2}\int_0^t |\dot{\eta}_s|^2 \, ds + \frac{1}{2}\int_0^t |D^- \mathrm{Ent}_m|^2(\eta_s) \, ds$$

for every t where $\mu_t \neq \nu_t$. But this contradicts (6); therefore it must be $\mu_t \equiv \nu_t$. □

References

[1] Luigi Ambrosio, Nicola Fusco, and Diego Pallara, *Functions of bounded variation and free discontinuity problems*, Oxford Mathematical Monographs, The Clarendon Press Oxford University Press, New York, 2000.

[2] Luigi Ambrosio and Nicola Gigli, *A users guide to optimal transport*, Lecture notes of C.I.M.E. summer school, 2009.

[3] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré, *Gradient flows in metric spaces and in the space of probability measures*, second ed., Lectures in Mathematics ETH Zürich, Birkhäuser Verlag, Basel, 2008.

[4] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré, *Heat flow and calculus on metric measure spaces with Ricci curvature bounded below—the compact case*, Boll. Unione Mat. Ital. (9) **5** (2012), no. 3, 575–629.

[5] Ennio De Giorgi, Antonio Marino, and Mario Tosques, *Problems of evolution in metric spaces and maximal decreasing curve*, Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Natur. (8) **68** (1980), no. 3, 180–187.

[6] David Heaver Fremlin, *Measure theory*, vol. 4, Torres Fremlin, 2000.

[7] Nicola Gigli, *On the heat flow on metric measure spaces: existence, uniqueness and stability*, Calc. Var. Partial Differential Equations **39** (2010), no. 1-2, 101–120.

[8] Richard Jordan, David Kinderlehrer, and Felix Otto, *The variational formulation of the Fokker-Planck equation*, SIAM J. Math. Anal. **29** (1998), no. 1, 1–17.

[9] John Lott and Cédric Villani, *Ricci curvature for metric-measure spaces via optimal transport*, Ann. of Math. (2) **169** (2009), no. 3, 903–991.

[10] Felix Otto, *The geometry of dissipative evolution equations: the porous medium equation*, Comm. Partial Differential Equations **26** (2001), no. 1-2, 101–174.

[11] Karl-Theodor Sturm, *On the geometry of metric measure spaces. I-II*, Acta Math. **196** (2006), no. 1, 65–177.

(Stefano Bianchini) SISSA, Via Bonomea 265, 34136 Trieste, Italy.
E-mail address: `bianchin@sissa.it`

(Alexander Dabrowski) SISSA, Via Bonomea 265, 34136 Trieste, Italy.
E-mail address: `adabrow@sissa.it`