# TIME-CONVEXITY OF THE ENTROPY IN THE MULTIPHASIC FORMULATION OF THE INCOMPRESSIBLE EULER EQUATION

## HUGO LAVENANT

ABSTRACT. We study the multiphasic formulation of the incompressible Euler equation introduced by Brenier: infinitely many phases evolve according to the compressible Euler equation and are coupled through a global incompressibility constraint. In a convex domain, we are able to prove that the entropy, when averaged over all phases, is a convex function of time, a result that was conjectured by Brenier. The novelty in our approach consists in introducing a time-discretization that allows us to import a *flow interchange* inequality previously used by Matthes, McCann and Savaré to study first order in time PDE, namely the JKO scheme associated with non-linear parabolic equations.

## 1. INTRODUCTION AND STATEMENT OF THE MAIN RESULT

Since the idea of Arnold [5] to consider the motion of an incompressible and inviscid fluid, described by the Euler equation, as a variational problem, namely a geodesic on the (infinite dimensional) group of orientation and measure-preserving diffeomorphisms (this is formally speaking an instance of the *least action principle*), this variational point of view has turned out to be fruitful. In particular, Brenier introduced relaxations leading to generalized geodesics on the group of measure-preserving maps: translated at a microscopic level, fluid particles are allowed to split and diffuse on the whole space (for a general survey, see for instance [11]). We will concentrate in this paper on one of Brenier's model with a flavor of Eulerian point of view introduced in [9] (see also [10, Section 4], [11, Section 1.5.3] and [1]) which goes as follows.

There are (possibly infinitely) many phases indexed by a parameter $\alpha$ which belongs to some probability space $(\mathfrak{A}, \mathcal{A}, \theta)$. At a fixed time $t$, each phase is described by its density $\rho_t^\alpha$ and its velocity field $\mathbf{v}_t^\alpha$, which are functions of the position $x$. We assume that all the densities are confined in a fixed bounded domain $\Omega$, and up to a normalization constant $\rho_t^\alpha$ can be seen as a probability measure on $\Omega$. The evolution in time of the phase $\alpha$ is done according to the continuity equation

$$\partial_t \rho_t^\alpha + \nabla \cdot (\rho_t^\alpha \mathbf{v}_t^\alpha) = 0, \tag{1.1}$$

where $\nabla\cdot$ stands for the divergence operator. We assume no-flux boundary conditions on $\partial\Omega$, thus the total mass of $\rho^\alpha$ is preserved over time. The different phases are coupled through the incompressibility constraint: at a fixed $t$ the density of all the different phases must sum up to the Lebesgue measure $\mathcal{L}$ (restricted to $\Omega$). In other words, for any $t$ we impose that

$$\int_{\mathfrak{A}} \rho_t^\alpha \, \mathrm{d}\theta(\alpha) = \mathcal{L}. \tag{1.2}$$

Looking at the problem from a variational point of view, we assume that the values of $\rho_t^\alpha$ are fixed for $t = 0$ and $t = 1$ and that the trajectories observed are those solving the following variational problem:

$$\min \left\{ \int_{\mathfrak{A}} \int_0^1 \int_\Omega \frac{1}{2} |\mathbf{v}_t^\alpha(x)|^2 \rho_t^\alpha(x) \, \mathrm{d}x \, \mathrm{d}t \, \mathrm{d}\theta(\alpha) \ : \ (\rho^\alpha, \mathbf{v}^\alpha) \text{ satisfies (1.1) and (1.2)} \right\}. \tag{1.3}$$

From a physical point of view, the functional which is minimized corresponds to the average (over all phases) of the integral over time of the kinetic energy, namely the global *action* of all the phases. Without the

LABORATOIRE DE MATHÉMATIQUES D'ORSAY, UNIV. PARIS-SUD, CNRS, UNIVERSITÉ PARIS-SACLAY, 91405 ORSAY CEDEX, FRANCE

*E-mail address*: hugo.lavenant@math.u-psud.fr.

incompressibility constraint, each phase would evolve independently and follow a geodesic in the Wasserstein space joining $\rho_0^\alpha$ to $\rho_1^\alpha$.

In Brenier's original formulation, the space $(\mathfrak{A}, \mathcal{A}, \theta)$ is the domain $\Omega$ endowed with the Lebesgue measure $\mathcal{L}$. If $h : \Omega \to \Omega$ is a measure-preserving map, "classical" boundary conditions are those where $\rho_0^\alpha$ is the Dirac mass located at $\alpha$ and $\rho_1^\alpha$ is the Dirac mass located at $h(\alpha)$. In a classical solution, each phase $\alpha$ will be of the form $\rho_t^\alpha = \delta_{y^\alpha(t)}$, where $y^\alpha : [0, 1] \to \Omega$ is a curve joining $\alpha$ to $h(\alpha)$. But, even if one starts with "classical" boundary conditions, there are cases where the phase $\alpha$ may split and $\rho^\alpha$ may not be a Dirac mass for any $t \in (0, 1)$, leading to a "non-classical" solution (for examples of such cases, the reader can consult [8, Section 6] or the detailed study [7]).

With formal considerations (see for instance [10, Section 4]), one can be convinced that for each phase $\alpha$, the optimal velocity field is the gradient of a scalar field $\varphi^\alpha$ (i.e. $\mathbf{v}_t^\alpha = \nabla \varphi_t^\alpha$), and that each $\varphi^\alpha$ evolves according to a Hamilton-Jacobi equation

$$\partial_t \varphi_t^\alpha + \frac{|\nabla \varphi_t^\alpha|^2}{2} = -p_t,$$

with a pressure field $p$ that does not depend on $\alpha$ and that arises from the incompressibility constraint. If we look at the Boltzmann entropy of the phase $\alpha$, a lengthy formal computation leads to

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2} \int_\Omega \rho_t^\alpha(x) \ln \rho_t^\alpha(x) \, \mathrm{d}x = \int_\Omega \left[ \Delta p_t(x) + |D^2 \varphi_t^\alpha(x)|^2 \right] \rho_t^\alpha(x) \, \mathrm{d}x,$$

where $|D^2 \varphi_t^\alpha|^2 = \sum_{i,j} (\partial_{ij} \varphi_t^\alpha)^2$ is the square of the Hilbert-Schmidt norm of the Hessian w.r.t. (with respect to) space of $\varphi_t^\alpha$. Thus, if one defines the averaged entropy $H$ as a function of time by

$$H(t) := \int_{\mathfrak{A}} \left( \int_\Omega \rho_t^\alpha(x) \ln \rho_t^\alpha(x) \, \mathrm{d}x \right) \mathrm{d}\theta(\alpha),$$

the previous computation leads to

$$H''(t) = \int_\Omega \Delta p_t + \int_{\mathfrak{A}} \int_\Omega |D^2 \varphi_t^\alpha(x)|^2 \rho_t^\alpha(x) \, \mathrm{d}x \, \mathrm{d}\theta(\alpha) \geqslant \int_{\partial \Omega} \nabla p_t \cdot n$$

where $n$ stands for the outward normal of $\Omega$. At this point, it becomes natural to assume that $\Omega$ is *convex*. Indeed, if this is the case, the acceleration of a fluid particle located on the boundary will be directed toward the interior of $\Omega$ because the particle is constrained to stay in $\Omega$. As the acceleration of the fluid particles is – at least heuristically – equal to $-\nabla p$, it is reasonable to expect that $\nabla p \cdot n \geqslant 0$ on $\partial \Omega$. Therefore, at a formal level, assuming the convexity of $\Omega$ leads to $H'' \geqslant 0$, i.e. to the property that the averaged entropy $H$ is a convex function of time. This was remarked and conjectured by Brenier in [10, section 4], but has received no proof to our knowledge until now. Indeed, the main difficulty lies in the fact that *a priori* the solutions are not regular enough to make the above computation rigorous. The goal of this paper is to give a rigorous statement and to prove this conjecture. The two main statements can informally be stated as follows:

**Theorem.** *Let us assume that $\Omega$ is convex. If $(\rho^\alpha, \mathbf{v}^\alpha)_{\alpha \in \mathfrak{A}}$ is the unique solution of the variational problem* (1.3) *whose total entropy $\int_0^1 H(t) \, \mathrm{d}t$ is finite and minimal compared to all other solutions, then $H$ is a convex function.*

**Theorem.** *Let us assume that $\Omega$ is convex and that the boundary terms are such that $H(0)$ and $H(1)$ are finite. Then there exists a solution of the variational problem* (1.3) *such that $\int_0^1 H(t) \, \mathrm{d}t$ is finite and minimal compared to all other solutions.*

More precisely, see Theorem 2.15 and Theorem 2.14 for the exact assumptions and statements, and Section 5 for the translation in Brenier's parametric setting (see below).

Let us remark that the convexity of the entropy is invisible for classical solutions. Indeed, if $\rho_t^\alpha$ is a Dirac mass for any $\alpha$, then $H(t) = +\infty$. Thus the convexity of the entropy is non trivial only for "strongly" non-classical solutions.

The strategy to prove the convexity of the entropy goes as follows. For a fixed $\alpha$, we see $t \mapsto \rho_t^\alpha$ as a curve in the space of probability measures on $\Omega$. This space (denoted $\mathcal{P}(\Omega)$) can be endowed with the 2-Wasserstein

distance $W_2(\cdot, \cdot)$ coming from optimal transport: the squared distance $W_2^2(\mu, \nu)$ between two measures $\mu$ and $\nu$ is just the optimal transport cost among all transport between $\mu$ and $\nu$. The interest of this distance is that the action $\int_0^1 \int_\Omega \frac{1}{2} |\mathbf{v}_t^\alpha(x)|^2 \rho_t^\alpha(x) \, dx \, dt$ of the curve $\rho^\alpha$ (or at least, the minimal value of the action among all velocity field $\mathbf{v}^\alpha$ satisfying the continuity equation (1.1)) can be seen as the integral w.r.t. time of the square of the *metric derivative* of the curve $t \mapsto \rho_t^\alpha$ in the metric space $(\mathcal{P}(\Omega), W_2)$. In particular, there appears a natural time-discretization of the action: if $N$ is large enough and $\tau := 1/N$ we expect that

$$\inf_{\mathbf{v}^\alpha \text{ satisfying (1.1)}} \int_0^1 \int_\Omega \frac{1}{2} |\mathbf{v}_t^\alpha(x)|^2 \rho_t^\alpha(x) \, dx \, dt = \int_0^1 \frac{1}{2} |\dot{\rho}_t^\alpha|^2 \, dt \simeq \sum_{k=1}^N \frac{\tau}{2} \left( \frac{W_2(\rho_{(k-1)\tau}^\alpha, \rho_{k\tau}^\alpha)}{\tau} \right)^2 .$$

At a discrete level, $\rho^\alpha$, which is a curve valued in $\mathcal{P}(\Omega)$, is approximated by an element of $\mathcal{P}(\Omega)^{N+1}$. The incompressibility constraint will be relaxed in order to allow comparison of the optimizer with any other competitor. If $\mathcal{C}_q : \mathcal{P}(\Omega) \to \mathbf{R}$ penalizes (more and more as $q \to +\infty$) the probability measures that are different from the Lebesgue measure, a discretized version of the Euler variational formulation (1.3) might read

$$\min \left\{ \int_{\mathfrak{A}} \left[ \sum_{k=1}^N \frac{\tau}{2} \left( \frac{W_2(\rho_{(k-1)\tau}^\alpha, \rho_{k\tau}^\alpha)}{\tau} \right)^2 \right] \, d\theta(\alpha) + \sum_{k=0}^N \mathcal{C}_q \left( \int_{\mathfrak{A}} \rho_{k\tau}^\alpha \, d\theta(\alpha) \right) \; : \; \rho_0^\alpha, \rho_1^\alpha \text{ fixed for } \theta\text{-a.e. } \alpha \right\} .$$

Even though it would be possible, we will not write down the optimality conditions of this discretized problem as they contain much more information than needed for our goal. Instead, if we take the minimizer of the discretized problem, we will let the $k$-th component follow the flow of the heat equation (with no-flux boundary conditions) and use the result as a competitor. The key point is that the heat equation is strongly related to the Wasserstein distance: the heat flow is the gradient flow of the functional entropy $\rho \mapsto \int_\Omega \rho \ln \rho$ in the metric space $(\mathcal{P}(\Omega), W_2)$. In particular, and this is called the *Evolution Variational Inequality*, one can estimate the derivative of the Wasserstein distance along the heat flow $\Phi_s$:

$$\frac{d}{ds} \frac{W_2^2(\Phi_s \mu, \nu)}{2} \bigg|_{s=0} \leqslant \int_\Omega \nu \ln \nu - \int_\Omega \mu \ln \mu.$$

This kind of inequality was previously used by Matthes, Mccann and Savaré in [15] under the name of *flow interchange* to tackle first order (in time) PDEs (basically discretizations via the JKO scheme of gradient flows). As far as we know, this kind of technique has never been used for second order PDEs like the Euler equation[1]. With this evolution variational inequality, one can show in a very simple way that the averaged entropy $\int_{\mathfrak{A}} \int_\Omega \rho_{k\tau}^\alpha \ln \rho_{k\tau}^\alpha \, d\theta(\alpha)$ is a (discrete) convex function of $k$. Then, one can expect that the solutions of the discretized problem will converge to those of the original one, and that the convexity of the entropy will be preserved at the limit.

It happens that all the quantities involved do not really depend on the particular dependence of the $\rho^\alpha$ in $\alpha$. Indeed, if one denotes by $\Gamma$ the space of continuous curves valued in the probability measures on $\Omega$ endowed with the Wasserstein distance (in short $\Gamma = C([0, 1], \mathcal{P}(\Omega))$), everything only depends on the image measure of $\theta$ through the map $\alpha \mapsto \rho^\alpha$. The natural object we are dealing with is therefore a probability measure on $\Gamma$, something that one can call (by analogy with [6]) a $W_2$-traffic plan. In a way, the application $\alpha \mapsto \rho^\alpha$ is a parametrization of a $W_2$-traffic plan: that's why we will call Brenier's formulation the parametric one, while we will work in the non parametric setting, dealing directly with probability measures on $\Gamma$. In our setting, most topological properties are easier to handle, and notations are according to us simplified. Even though any probability measure on $\Gamma$ cannot be *a priori* parametrized, we will show that it is the case for the solutions of Euler's variational problem. Therefore, our results can be translated in Brenier's parametric setting.

This paper is organized as follows. In Section 2, we introduce the notations that we will use, we briefly recall some properties of the Wasserstein distance and of the heat equation seen as the gradient flow of the entropy. We state explicitly the variational problem we are interested in and prove the existence of a solution

---

[1]Let us precise that this idea is adapted from an ongoing work [14] with Filippo Santambrogio where the same kind of technique is used to provide regularity for solutions of quadratic Mean Field Games.

(a result which is known since [8]). We give a rigorous statement of the theorems that we prove in the next two sections. Section 3 introduces the discrete problem and proves the convexity of the averaged entropy at the discrete level. This section contains the key ingredient around which all the proof revolves. In Section 4, we show that the solutions of the discrete problems converge to the solution of the original one, and that the convexity of the averaged entropy is preserved when the limit is taken. Though lengthy and technical, this section does not contain profound ideas. Finally, Section 5 is devoted to the proof of the equivalence between the parametric and non parametric formulations.

## 2. NOTATIONS, PRELIMINARY RESULTS AND EXISTENCE OF A SOLUTION TO THE CONTINUOUS PROBLEM

If $X$ is a polish space (complete, metric, separable), the space of Borel probability measures on $X$ will be denoted by $\mathcal{P}(X)$, and $C(X)$ is the space of continuous and bounded functions on $X$ valued in $\mathbf{R}$. The space $\mathcal{P}(X)$ will be endowed with the topology of the weak convergence of measures (i.e. the topology induced by the duality with $C(X)$).

In all the sequel, we will denote by $\Omega$ a closed bounded *convex* subset of $\mathbf{R}^d$ with non empty interior. In particular, $\Omega$ is compact. In order to avoid normalization constants, we assume that the Lebesgue measure of $\Omega$ is $1$. The Lebesgue measure on $\Omega$, which is therefore a probability measure, will be denoted by $\mathcal{L}$.

2.1. **The Wasserstein space.** The space $\mathcal{P}(\Omega)$ of probability measures on $\Omega$ is endowed with the Wasserstein distance: if $\mu$ and $\nu$ are two elements of $\mathcal{P}(\Omega)$, the 2-Wasserstein distance $W_2(\mu, \nu)$ between $\mu$ and $\nu$ is defined by

$$(2.1) \qquad W_2(\mu, \nu) := \sqrt{\min \left\{ \int_{\Omega \times \Omega} |x - y|^2 \, \mathrm{d}\gamma(x, y) \; : \; \gamma \in \mathcal{P}(\Omega \times \Omega) \text{ and } \pi_0 \# \gamma = \mu, \; \pi_1 \# \gamma = \nu \right\}}.$$

In the formula above, $\pi_0$ and $\pi_1 : \Omega \times \Omega \to \Omega$ stand for the projections on respectively the first and second component of $\Omega \times \Omega$. If $T : X \to Y$ is a measurable application and $\mu$ is a measure on $X$, then the image measure of $\mu$ by $T$, denoted by $T \# \mu$, is the measure defined on $Y$ by $(T \# \mu)(B) = \mu(T^{-1}(B))$ for any measurable set $B \subset Y$. It can also be defined by

$$\int_Y a(y) \, \mathrm{d}(T \# \mu)(y) := \int_X a(T(x)) \, \mathrm{d}\mu(x),$$

this identity being valid as soon as $a : Y \to \mathbf{R}$ is an integrable function.

For general results about optimal transport, the reader might refer to [19] or [17]. We recall that $W_2$ defines a metric on $\mathcal{P}(\Omega)$ that metrizes the weak convergence of measures. Therefore, thanks to Prokhorov's Theorem, the space $(\mathcal{P}(\Omega), W_2)$ is a compact metric space. We also recall that $(\mathcal{P}(\Omega), W_2)$ is a geodesic space and if $\gamma \in \mathcal{P}(\Omega \times \Omega)$ is optimal in formula (2.1), then a constant-speed geodesic $\rho : [0, 1] \to \mathcal{P}(\Omega)$ joining $\mu$ to $\nu$ is given by $\rho(t) := \pi_t \# \gamma$ with $\pi_t : (x, y) \in \Omega \times \Omega \to (1 - t)x + ty \in \Omega$ (remark that the convexity of $\Omega$ is important here). Reciprocally, every constant-speed geodesic is of this form (see [17, prop. 5.32]). We also recall that $W_2^2 : \mathcal{P}(\Omega) \times \mathcal{P}(\Omega) \to \mathbf{R}$ is a convex function.

We will consider the entropy (w.r.t. the Lebesgue measure) functional $\mathcal{H}$ on $\mathcal{P}(\Omega)$. It is the functional $\mathcal{H} : \mathcal{P}(\Omega) \to [0, +\infty]$ defined by, for any $\mu \in \mathcal{P}(\Omega)$,

$$(2.2) \qquad \mathcal{H}(\mu) := \begin{cases} \int_\Omega \mu(x) \ln(\mu(x)) \, \mathrm{d}x & \text{if } \mu \text{ is absolutely continuous w.r.t. } \mathcal{L}, \\ +\infty & \text{else.} \end{cases}$$

The fact that $\mathcal{H} \geqslant 0$ on $\mathcal{P}(\Omega)$ is a consequence of Jensen's inequality and of the normalization $\mathcal{L}(\Omega) = 1$. We will also deal with an other functional on $\mathcal{P}(\Omega)$ that we will use to penalize the concentrated measures, namely the $q$-th power of the density. More precisely, if $q > 1$, we denote by $\mathcal{C}_q : \mathcal{P}(\Omega) \to [0, +\infty]$ the congestion functional defined by, for any $\mu \in \mathcal{P}(\Omega)$,

$$(2.3) \qquad \mathcal{C}_q(\mu) := \begin{cases} \int_\Omega \mu(x)^q \, \mathrm{d}x - 1 & \text{if } \mu \text{ is absolutely continuous w.r.t. } \mathcal{L}, \\ +\infty & \text{else.} \end{cases}$$

Again, thanks to Jensen's inequality, we see that $\mathcal{C}_q(\mu) \geqslant 0$ with equality if and only if $\mu = \mathcal{L}$. We recall that a functional is geodesically convex on $(\mathcal{P}(\Omega), W_2)$ if for any two given probability measures, there exists a constant-speed geodesic connecting these two measures along which the functional is convex. A functional will be called convex if it is so w.r.t. the usual affine structure on $\mathcal{P}(\Omega)$. Well known facts about $\mathcal{H}$ and $\mathcal{C}_q$ are summarized in the following proposition (see [17, chap. 7]).

**Proposition 2.1.** *For any $q > 1$, the functionals $\mathcal{H}$ and $\mathcal{C}_q$ are l.s.c. (lower semi-continuous), strictly convex and geodesically convex on $(\mathcal{P}(\Omega), W_2)$.*

Let us underline that the convexity of $\Omega$ is needed to get the geodesic convexity of $\mathcal{H}$ and $\mathcal{C}_q$.

2.2. **Absolutely continuous curves in the Wasserstein space.** If $S$ is a closed subset of $[0, 1]$, $\Gamma_S$ will denote the set of continuous functions on $S$ valued in $\mathcal{P}(\Omega)$ (in practice, we will only consider subsets $S$ that have a finite number of points or that are subintervals of $[0, 1]$). In the case where the index $S$ is omitted, it is assumed that $S = [0, 1]$. This space will be equipped with the distance $d$ of the uniform convergence, i.e.

$$d(\rho^1, \rho^2) := \max_{t \in S} W_2(\rho^1(t), \rho^2(t)).$$

For any closed subset $S'$ of $S$, the application $e_{S'} : \Gamma_S \to \Gamma_{S'}$ is the restriction operator. In the case where $S' = \{t\}$ is a singleton, we will use the notation $e_t := e_{\{t\}}$ and often use the compact writing $\rho_t$ for $e_t(\rho) = \rho(t)$. One can see that $\Gamma_S$ is a polish space, and that it is compact if $S$ contains a finite number of points.

Following [2, Definition 1.1.1], we give ourselves the following definition.

**Definition 2.2.** *We say that a curve $\rho \in \Gamma$ is 2-absolutely continuous if there exists a function $f \in L^2([0, 1])$ such that, for every $0 \leqslant t \leqslant s \leqslant 1$,*

$$W_2(\rho_t, \rho_s) \leqslant \int_t^s f(r) \, \mathrm{d}r.$$

The main interest of this notion lies in the two following theorems that we recall.

**Theorem 2.3.** *If $\rho \in \Gamma$ is a 2-absolutely continuous curve, then the quantity*

$$|\dot{\rho}_t| := \lim_{h \to 0} \frac{W_2(\rho_{t+h}, \rho_t)}{h}$$

*exists and is finite for a.e. $t$. Moreover,*

$$(2.4) \qquad \int_0^1 |\dot{\rho}_t|^2 \, \mathrm{d}t = \sup_{N \geqslant 2} \sup_{0 \leqslant t_1 < t_2 < \ldots < t_N \leqslant 1} \sum_{k=2}^N \frac{W_2^2(\rho_{t_{k-1}}, \rho_{t_k})}{t_k - t_{k-1}}.$$

*Proof.* The first part is just [2, Theorem 1.1.2]. The proof of the representation formula (2.4) can easily be obtained by adapting the proof of [3, Theorem 4.1.6]. □

The quantity $|\dot{\rho}_t|$ is called the metric derivative of the curve $\rho$ and heuristically corresponds to the norm of the derivative of $\rho$ at time $t$ in the metric space $(\mathcal{P}(\Omega), W_2)$. The link between this metric derivative and the continuity equation is the following (and difficult) theorem, whose proof can be found in [2, Theorem 8.3.1] (see also [17, Theorem 5.14]).

**Theorem 2.4.** *Let $\rho \in \Gamma$ be a 2-absolutely continuous curve. Then*

$$(2.5) \qquad \frac{1}{2} \int_0^1 |\dot{\rho}_t|^2 \, \mathrm{d}t = \min \left\{ \int_0^1 \left( \int_\Omega \frac{1}{2} |\mathbf{v}_t|^2 \, \mathrm{d}\rho_t \right) \, \mathrm{d}t \right\},$$

*where the minimum is taken over all families $(\mathbf{v}_t)_{t \in [0,1]}$ such that $\mathbf{v}_t \in L^2(\Omega, \mathbf{R}^d, \rho_t)$ for a.e. $t$ and such that the continuity equation $\partial_t \rho_t + \nabla \cdot (\rho_t \mathbf{v}_t) = 0$ with no-flux boundary conditions is satisfied in a weak sense.*

This result shows that if we are only interested in the minimal value taken by the action of the curve (the r.h.s. (right hand side) of (2.5)), we only need to consider the metric derivative of the curve $\rho$ and we can forget the velocity field $\mathbf{v}$. Therefore we define the action $A : \Gamma \to [0, +\infty]$ by, for any $\rho \in \Gamma$,

$$(2.6) \qquad A(\rho) := \begin{cases} \dfrac{1}{2} \displaystyle\int_0^1 |\dot{\rho}_t|^2 \, \mathrm{d}t & \text{if } \rho \text{ is } 2 - \text{absolutely continuous,} \\ +\infty & \text{else.} \end{cases}$$

Some standard but useful properties of this functional are the following.

**Proposition 2.5.** *The functional $A$ is convex, l.s.c. and its sublevel sets are compact in $\Gamma$.*

*Proof.* To prove that $A$ is convex and l.s.c., we rely on the representation formula (2.4) which shows that $A$ is the supremum of convex continuous functions. Moreover if $\rho \in \Gamma$ is a curve with finite action and $s < t$, then, again with (2.4), one can see that $W_2(\rho_s, \rho_t) \leqslant \sqrt{2A(\rho)}\sqrt{t-s}$. This shows that the sublevel sets of $A$ are uniformly equicontinuous, therefore they are relatively compact thanks to Ascoli-Arzela's theorem. $\square$

2.3. **The heat equation and the Wasserstein space.** Let us denote by $\Phi : [0, +\infty) \times \mathcal{P}(\Omega) \to \mathcal{P}(\Omega)$ the flow of the heat equation with Neumann boundary conditions. In other words, for any $s \geqslant 0$ and any $\mu \in \mathcal{P}(\Omega)$, $\Phi_s(\mu) = u(s)$ where $u$ is the solution (in the sense of distributions) of the Cauchy problem

$$\begin{cases} \dfrac{\partial u}{\partial t} = \Delta u & \text{in } (0, +\infty) \times \mathring{\Omega} \\ \nabla u \cdot n = 0 & \text{on } (0, +\infty) \times \partial \Omega \\ \lim_{t \to 0} u(t) = \mu & \text{in } \mathcal{P}(\Omega) \end{cases}$$

In the equation above, $n$ stands for the outward normal vector to the boundary $\partial \Omega$. As $\Omega$ is convex, it has a Lipschitz boundary, a regularity which is known to be sufficient for this Cauchy problem to be well posed and to admit a unique solution (see for instance [4, Section 7] and [16]). Moreover (see [4, Section 7]), a regularizing effect of the heat flow is encoded in the following estimate (with $C$ a constant that depends only on $\Omega$):

$$\forall \mu \in \mathcal{P}(\Omega), \ \forall s > 0, \ \|\Phi_s \mu\|_{L^\infty} \leqslant C\left(s^{-d/2} + 1\right).$$

In particular, for any $s > 0$ there exists a constant $C_s$ such that for any $\mu \in \mathcal{P}(\Omega)$, we have $\mathcal{H}(\Phi_s \mu) \leqslant C_s$.

The key point in what follows is that the heat flow can be seen as the gradient flow of the entropy functional $\mathcal{H}$ in the metric space $(\mathcal{P}(\Omega), W_2)$. That is (and this was remarked first by [13]), in a very informal way, $\Phi$ flows in the direction where the entropy $\mathcal{H}$ decreases the most. The standard reference about gradient flows in metric spaces is [2], one can also look at the survey [18]. In any case, this seminal point of view explains the three following identities involving the heat flow, the Wasserstein distance, and the entropy.

**Proposition 2.6.** *The Wasserstein distance decreases along the heat flow: if $\mu$ and $\nu \in \mathcal{P}(\Omega)$, and $s \geqslant 0$,*

$$(2.7) \qquad\qquad W_2(\Phi_s \mu, \Phi_s \nu) \leqslant W_2(\mu, \nu).$$

*Moreover, let $\mu \in \mathcal{P}(\Omega)$ with $\mathcal{H}(\mu) < +\infty$. Then the curve $s \mapsto \Phi_s \mu$ is 2-absolutely continuous and the Energy Identity holds: for any $s \geqslant 0$,*

$$(\text{EI}) \qquad\qquad \int_0^s |\dot{\Phi}_r \mu|^2 \, \mathrm{d}r = \mathcal{H}(\mu) - \mathcal{H}(\Phi_s(\mu)).$$

*In addition, for any $\mu, \nu \in \mathcal{P}(\Omega)$ with $\mathcal{H}(\mu) < +\infty$ and any $s \geqslant 0$, the Evolution Variational Inequality holds:*

$$(\text{EVI}) \qquad\qquad \limsup_{h \to 0, \ h > 0} \frac{W_2^2(\Phi_{s+h}\mu, \nu) - W_2^2(\Phi_s \mu, \nu)}{2h} \leqslant \mathcal{H}(\nu) - \mathcal{H}(\Phi_s \mu).$$

One can look at [2, Theorem 11.2.1] to see the generalization of these identities to more general functionals than $\mathcal{H}$ along their gradient flow, provided that assumptions of $\lambda$-convexity along generalized geodesics are satisfied by the functional (and this is the case for $\mathcal{H}$ with $\lambda = 0$ because of our assumption of convexity of $\Omega$). All this properties are also summarized in [15, Theorem 2.4].

2.4. **Statement of the continuous problem.** As explained in the introduction, the object on which we will work, a "$W_2$-traffic plan", is a probability measure on the set of curves valued in $\mathcal{P}(\Omega)$, i.e. an element of $\mathcal{P}(\Gamma)$. Recall that the space $\mathcal{P}(\Gamma)$ is equipped with the topology of weak convergence of measures. If $Q \in \mathcal{P}(\Gamma)$, we need to translate the constraints, namely the fact that the values of the curves at $t = 0$ and $t = 1$ are fixed, and the incompressibility at each time $t$.

Incompressibility means that at each time $t$, the measure $e_t \# Q$ (which is an element of $\mathcal{P}(\mathcal{P}(\Omega))$) when averaged (its mean value is an element of $\mathcal{P}(\Omega)$), is equal to $\mathcal{L}$. We therefore need to define what the mean value of $e_t \# Q$ is.

**Definition 2.7.** *Let $S$ be a closed subset of $[0,1]$ and $t \in S$. If $Q \in \mathcal{P}(\Gamma_S)$, we denote by $m_t(Q)$ the probability measure on $\Omega$ defined by*

$$(2.8) \qquad \forall a \in C(\Omega), \ \int_\Omega a(x) \, \mathrm{d}[m_t(Q)](x) := \int_{\Gamma_S} \left( \int_\Omega a(x) \, \mathrm{d}\rho_t(x) \right) \, \mathrm{d}Q(\rho).$$

We can easily see that, for a fixed $t$, $Q \mapsto m_t(Q)$ is continuous. It is an easy application of Fubini's theorem to show that, if $Q$-a.e. $\rho_t$ is absolutely continuous w.r.t. to $\mathcal{L}$, then $m_t(Q)$ is also absolutely continuous w.r.t. $\mathcal{L}$, and its density is the mean density of the $\rho_t$ w.r.t. $Q$. Incompressibility is then expressed by the fact that $m_t(Q) = \mathcal{L}$ for any $t$.

   To encode the boundary conditions, we just consider a coupling $\gamma \in \mathcal{P}(\Gamma_{\{0,1\}}) = \mathcal{P}(\mathcal{P}(\Omega) \times \mathcal{P}(\Omega))$ between the initial and final values, compatible with the incompressibilty constraint (i.e. $m_0(\gamma) = m_1(\gamma) = \mathcal{L}$), and we impose that $(e_0, e_1)\#Q = \gamma$.

**Definition 2.8.** *Let $\gamma \in \mathcal{P}(\Gamma_{\{0,1\}})$ be a coupling compatible with the incompressibility constraint (i.e. $m_0(\gamma) = m_1(\gamma) = \mathcal{L}$) and $S$ be a closed subset of $[0,1]$ containing $0$ and $1$. The space of incompressible $W_2$-traffic plans is*

$$\mathcal{P}_{\mathrm{in}}(\Gamma_S) := \{Q \in \mathcal{P}(\Gamma_S) \ : \ \forall t \in S, \ m_t(Q) = \mathcal{L}\}.$$

*The space of $W_2$-traffic plans satisfying the boundary conditions is*

$$\mathcal{P}_{\mathrm{bc}}(\Gamma_S) := \{Q \in \mathcal{P}(\Gamma_S) \ : \ (e_0, e_1)\#Q = \gamma\}.$$

*The space of admissible $W_2$-traffic plans is*

$$\mathcal{P}_{\mathrm{adm}}(\Gamma_S) := \mathcal{P}_{\mathrm{in}}(\Gamma_S) \cap \mathcal{P}_{\mathrm{bc}}(\Gamma_S).$$

The following proposition derives directly from the definition.

**Proposition 2.9.** *If $S$ is a closed subset of $[0,1]$ containing $0$ and $1$, the spaces $\mathcal{P}_{\mathrm{in}}(\Gamma_S)$, $\mathcal{P}_{\mathrm{bc}}(\Gamma_S)$ and $\mathcal{P}_{\mathrm{adm}}(\Gamma_S)$ are closed in $\mathcal{P}(\Gamma_S)$.*

   We have now enough vocabulary to state the minimization problem we are interested in, namely to minimize the averaged action over the set of admissible $W_2$-traffic plans. We denote by $\mathcal{A} : \mathcal{P}(\Gamma) \to [0, +\infty]$ the functional defined by, for any $Q \in \mathcal{P}(\Gamma)$,

$$\mathcal{A}(Q) := \int_\Gamma A(\rho) \, \mathrm{d}Q(\rho),$$

where we recall that $A(\rho)$ is the action of the curve $\rho$, see (2.6).

**Definition 2.10.** *The continuous problem is defined as*

(CP) $$\min\{\mathcal{A}(Q) \ : \ Q \in \mathcal{P}_{\mathrm{adm}}(\Gamma)\}.$$

*Any $Q \in \mathcal{P}_{\mathrm{adm}}(\Gamma)$ with $\mathcal{A}(Q) < +\infty$ realizing the minimum will be referred as a solution of the continuous problem.*

   In order to prove the existence of a solution to (CP), we rely on the classical following lemma which is valid if $\Gamma_S$ is replaced by any metric space (see for instance [17, Proposition 7.1] and [2, Remark 5.15]).

**Lemma 2.11.** *Let $S$ be a closed subset of $[0,1]$ and $F : \Gamma_S \to [0, +\infty]$ a l.s.c. positive function. Then the function $\mathcal{F} : \mathcal{P}(\Gamma_S) \to [0, +\infty]$ defined by*

$$\mathcal{F}(Q) = \int_{\Gamma_S} F(\rho) \, \mathrm{d}Q(\rho)$$

*is convex and l.s.c. Moreover, if the sublevel sets of $F$ are compact, so are those of $\mathcal{F}$.*

The existence of a solution to (CP) is then a straightforward application of the direct method of calculus of variations.

**Theorem 2.12.** *There exists at least one solution to* (CP)*.*

*Proof.* The functional $\mathcal{A}$ is l.s.c. and has compact sublevel sets thanks to Proposition 2.5 and Lemma 2.11. Moreover the set $\mathcal{P}_{\mathrm{adm}}(\Gamma)$ is closed. To use the direct method of calculus of variations, we only need to prove that there exists $Q \in \mathcal{P}_{\mathrm{adm}}(\Gamma)$ such that $\mathcal{A}(Q) < +\infty$.

Notice that as $\Omega$ is convex, it is the image of the unit cube of $\mathbf{R}^d$ by a Lipschitz and measure-preserving map (see [12, Theorem 5.4][2]). It is known (see [1, Theorem 3.3] and Proposition 5.1 to translate the result in our setting) that the fact that $\Omega$ is the image of the unit cube by a Lipschitz and measure-preserving map ensures the existence of an admissible $W_2$-traffic plan with finite action. □

Let us mention here some already known results about the continuous problem (the existence of a solution being one of them). In some cases there is no uniqueness in (CP), we refer the reader to [7] for a comprehensive study of one of such cases. In [1], it is shown how, from a pair of measure-preserving plans (i.e. a pair of elements of $\{\mu \in \mathcal{P}(\Omega \times \Omega) \ : \ \pi_0 \# \mu = \mathcal{L} \text{ and } \pi_1 \# \mu = \mathcal{L}\}$), one can build an incompressible coupling $\gamma \in \mathcal{P}_{\mathrm{in}}(\Gamma_{\{0,1\}})$ which is, in some sense, a concatenation of them. Indeed, if $\mu, \nu \in \mathcal{P}(\Omega \times \Omega)$ are two measure-preserving plans, one can consider $(\mu_x)_{x \in \Omega}$ and $(\nu_x)_{x \in \Omega}$ the disintegration of $\mu$ and $\nu$ w.r.t. $\pi_1$, and then define $\gamma \in \mathcal{P}_{\mathrm{in}}(\Gamma_{\{0,1\}})$ by its action on continuous functions $a \in C(\Gamma_{\{0,1\}}) = C(\mathcal{P}(\Omega)^2)$:

$$\int_{\Gamma_{\{0,1\}}} a(\rho_0, \rho_1) \, \mathrm{d}\gamma(\rho_0, \rho_1) := \int_\Omega a(\mu_x, \nu_x) \, \mathrm{d}\mathcal{L}(x).$$

(To understand this construction and check that it is incompressible, one can look at [1] and Section 5 for the translation in the non parametric case, as [1] corresponds to the parametric case with $(\mathfrak{A}, \theta) = (\Omega, \mathcal{L})$). Using this construction, then [1, Proposition 3.4] states that the minimal cost (or more precisely the square root of the minimal value of (CP)) defines a distance on the space of measure-preserving plans.

In this article we are interested in the temporal behavior of the entropy when averaged over all phases.

**Definition 2.13.** *Let $S$ be a closed subset of $[0, 1]$. For any $Q \in \mathcal{P}(\Gamma_S)$, we define the* averaged *entropy* $H_Q : S \to [0, +\infty]$ *by, for any $t \in S$,*

$$H_Q(t) := \int_{\Gamma_S} \mathcal{H}(\rho_t) \, \mathrm{d}Q(\rho).$$

*If $Q \in \mathcal{P}(\Gamma)$, the quantity $\displaystyle\int_0^1 H_Q(t) \, \mathrm{d}t$ will be called the* total *entropy of Q.*

We recall that $\mathcal{H}$ is the entropy of a probability measure w.r.t. $\mathcal{L}$, see (2.2). By lower semi-continuity of $\mathcal{H}$ and Lemma 2.11, we can see that the function (of the variable $t$) $H_Q$ is l.s.c. In the sequel, we will concentrate on the cases where the averaged entropy belongs to $L^1([0, 1])$, i.e. where the total entropy is finite. By doing so, we exclude classical solutions: indeed, for a classical solution $Q \in \mathcal{P}_{\mathrm{adm}}(\Gamma)$, for any $t$ the measure $e_t \# Q$ is concentrated on Dirac masses, for which the entropy is infinite. We denote by $\mathcal{P}_{\mathrm{adm}}^H(\Gamma)$ the set of admissible $W_2$-traffic plans for which the total entropy is finite:

$$\mathcal{P}_{\mathrm{adm}}^H(\Gamma) := \mathcal{P}_{\mathrm{adm}}(\Gamma) \cap \left\{ Q \in \mathcal{P}(\Gamma) \ : \ \int_0^1 H_Q(t) \, \mathrm{d}t < +\infty \right\}$$

The main (and restrictive) assumption that we will consider is that there exists a solution of the continuous problem (CP) in $\mathcal{P}_{\mathrm{adm}}^H(\Gamma)$:

**Assumption 1.** *There exists $Q \in \mathcal{P}_{\mathrm{adm}}^H(\Gamma)$ such that $\mathcal{A}(Q) = \min\{\mathcal{A}(Q') \ : \ Q' \in \mathcal{P}_{\mathrm{adm}}(\Gamma)\}$.*

We will also work with a second assumption which will turn out to be more restrictive than Assumption 1, but which has the advantage of involving only the boundary terms, namely the fact that the initial and final values have finite averaged entropy.

**Assumption 2.** *The coupling $\gamma$ is such that $H_\gamma(0)$ and $H_\gamma(1)$ are finite.*

---

[2]Strictly speaking, in [12], it is required that $\Omega$ has a piecewise $C^1$ boundary, but this assumption is only used to prove that the Minkowski functional of $\Omega$ is Lipschitz. If $\Omega$ is convex, then its Minkowski functional is convex, hence Lipschitz. Thus, one can drop the assumption of a piecewise $C^1$ boundary if $\Omega$ is convex.

In other words, we impose that

$$\int_\gamma \left( \int_\Omega \rho_0 \ln \rho_0 \right) \, \mathrm{d}\gamma(\rho) < +\infty \ \textbf{ and } \ \int_\gamma \left( \int_\Omega \rho_1 \ln \rho_1 \right) \, \mathrm{d}\gamma(\rho) < +\infty.$$

In particular, Assumption 2 implies that $e_0 \# \gamma$ and $e_1 \# \gamma$ are concentrated on measures that are absolutely continuous w.r.t. $\mathcal{L}$: it excludes any classical boundary data.

The two main results of this paper can be stated as follows. Recall that $\Omega$ is assumed to be convex.

**Theorem 2.14.** *Suppose that Assumption 2 holds. Then there exists a solution $Q \in \mathcal{P}_{\mathrm{adm}}(\Gamma)$ of the continuous problem* (CP) *such that $H_Q(t) \leqslant \max(H_\gamma(0), H_\gamma(1))$ for any $t \in [0,1]$.*

In other words, if the initial and final averaged entropy are finite, then there exists a solution of the continuous problem with a uniformly bounded averaged entropy. In particular, Assumption 2 implies Assumption 1.

**Theorem 2.15.** *Suppose that Assumption 1 holds. Then, among all the solutions of the continuous problem* (CP)*, the unique $Q \in \mathcal{P}_{\mathrm{adm}}^H(\Gamma)$ which minimizes the total entropy $\int_0^1 H_Q(t) \, \mathrm{d}t$ is such that $H_Q$ is convex.*

In other words, we are able to prove the convexity of the averaged entropy for the solution which is "the most mixed", i.e. the one for which the total entropy is minimal. This statement contains the fact that the criterion of minimization of the total entropy selects a unique solution among the – potentially infinitely many – solutions of (CP). Let us mention that our proof could be easily adapted to show that the convexity also holds for the solution $Q$ which minimizes $\int_0^1 H_Q(t) a(t) \, \mathrm{d}t$, where $a : [0,1] \to (0, +\infty)$ is any continuous and strictly positive function.

The next two sections are devoted to the proof of these two theorems. As explained in the introduction, we will introduce a discrete (in time) problem (DP) which approximates the continuous one. Without any assumption, we will be able to prove the convexity of the averaged entropy at the discrete level (Theorem 3.2). Then we will show that, under Assumption 1 or Assumption 2, the solutions of the discrete problems converge to a solution of the continuous one (Proposition 4.6). Under Assumption 2, this solution will happen to have a uniformly bounded entropy (Corollary 4.8). Then we will show that, under Assumption 1, this solution will be the one with minimal total entropy (Corollary 4.9) and that its averaged entropy is a convex function of time (Corollary 4.13).

Finally, the uniqueness of such a $Q \in \mathcal{P}_{\mathrm{adm}}^H(\Gamma)$ with minimal total entropy has nothing to do with the discrete problem, it is a simple consequence of the strict convexity of $\mathcal{H}$. We will therefore prove it here to end this section. Indeed, it is a consequence of the following proposition.

**Proposition 2.16.** *Let $Q^1$ and $Q^2 \in \mathcal{P}_{\mathrm{adm}}^H(\Gamma)$ be two distinct admissible $W_2$-traffic plans. Then there exists $Q \in \mathcal{P}_{\mathrm{adm}}^H(Q)$ with*

$$\mathcal{A}(Q) \leqslant \frac{1}{2} \left( \mathcal{A}(Q^1) + \mathcal{A}(Q^2) \right)$$

*and*

$$\int_0^1 H_Q(t) \, \mathrm{d}t < \frac{1}{2} \left( \int_0^1 H_{Q^1}(t) \, \mathrm{d}t + \int_0^1 H_{Q^2}(t) \, \mathrm{d}t \right).$$

*Proof.* As $Q \mapsto H_Q$ is linear, it is not sufficient to consider the mean of $Q^1$ and $Q^2$. Instead, we will need to take means in $\Gamma$. In order to do so, we disintegrate $Q^1$ and $Q^2$ w.r.t. $e_{\{0,1\}} = (e_0, e_1)$. We obtain two families $Q^1_{\rho_0,\rho_1}$ and $Q^2_{\rho_0,\rho_1}$ of $W_2$-traffic plans indexed by $(\rho_0, \rho_1) \in \Gamma_{\{0,1\}} = \mathcal{P}(\Omega)^2$. We define $Q$ by its disintegration w.r.t. $e_{\{0,1\}}$: we set $Q := Q_{\rho_0,\rho_1} \otimes \gamma$ where $Q_{\rho_0,\rho_1}$ is taken to be the image measure of $Q^1_{\rho_0,\rho_1} \otimes Q^2_{\rho_0,\rho_1}$ by the map $(\rho^1, \rho^2) \mapsto (\rho^1 + \rho^2)/2$ (where the $+$ refers to the usual affine structure on $\Gamma$). In other words, for any $a \in C(\Gamma)$,

$$\int_\Gamma a(\rho) \, \mathrm{d}Q(\rho) := \int_{\Gamma_{\{0,1\}}} \left( \int_\Gamma a \left[ \frac{\rho^1 + \rho^2}{2} \right] \, \mathrm{d}Q^1_{\rho_0,\rho_1}(\rho^1) \, \mathrm{d}Q^2_{\rho_0,\rho_1}(\rho^2) \right) \, \mathrm{d}\gamma(\rho_0, \rho_1).$$

As $(e_0, e_1)\#Q^1_{\rho_0,\rho_1}$ and $(e_0, e_1)\#Q^2_{\rho_0,\rho_1}$ are Dirac masses concentrated on $(\rho_0, \rho_1)$, we can easily see that $Q \in \mathcal{P}_{\mathrm{bc}}(\Gamma)$. The incompressibility constraint is straightforward to obtain: for any $a \in C(\Omega)$ and any $t \in [0\,,1]$,

$$\int_\Omega a(x)\,\mathrm{d}[m_t(Q)](x) = \int_{\Gamma_{\{0,1\}}} \left( \int_\Gamma \left[ \int_\Omega a(x) \frac{\mathrm{d}\rho_t^1(x) + \mathrm{d}\rho_t^2(x)}{2} \right] \mathrm{d}Q^1_{\rho_0,\rho_1}(\rho^1)\,\mathrm{d}Q^2_{\rho_0,\rho_1}(\rho^2) \right) \, \mathrm{d}\gamma(\rho_0, \rho_1)$$

$$= \int_{\Gamma_{\{0,1\}}} \left( \int_\Gamma \left[ \int_\Omega a(x) \frac{\mathrm{d}\rho_t^1(x)}{2} \right] \mathrm{d}Q^1_{\rho_0,\rho_1}(\rho^1) + \int_\Gamma \left[ \int_\Omega a(x) \frac{\mathrm{d}\rho_t^2(x)}{2} \right] \mathrm{d}Q^2_{\rho_0,\rho_1}(\rho^2) \right) \, \mathrm{d}\gamma(\rho_0, \rho_1)$$

$$= \frac{1}{2} \int_\Omega a(x)\,\mathrm{d}x + \frac{1}{2} \int_\Omega a(x)\,\mathrm{d}x$$

$$= \int_\Omega a(x)\,\mathrm{d}x.$$

Thus, we have $Q \in \mathcal{P}_{\mathrm{adm}}(\Gamma)$. To handle the action, let us just remark that for any $\rho^1$ and $\rho^2$ in $\Gamma$, by convexity of $A$,

$$A\left( \frac{\rho^1 + \rho^2}{2} \right) \leqslant \frac{1}{2} \left( A(\rho^1) + A(\rho^2) \right).$$

Integrating this inequality w.r.t. to $Q^1_{\rho_0,\rho_1} \otimes Q^2_{\rho_0,\rho_1}$ and then w.r.t. $\gamma$ gives the result. We use the same kind of reasoning for the entropy, but this functional is strictly convex. Hence, for any $t \in [0\,,1]$,

$$\mathcal{H}\left( \frac{\rho_t^1 + \rho_t^2}{2} \right) \leqslant \frac{1}{2} \left( \mathcal{H}(\rho_t^1) + \mathcal{H}(\rho_t^2) \right)$$

with a strict inequality if $\rho_t^1 \neq \rho_t^2$ and if the r.h.s. is finite. Integrating w.r.t. $t$ and w.r.t. $Q^1_{\rho_0,\rho_1} \otimes Q^2_{\rho_0,\rho_1}$ we get,

$$\int_\Gamma \left( \int_0^1 \mathcal{H}\left[ \frac{\rho_t^1 + \rho_t^2}{2} \right] \mathrm{d}t \right) \, \mathrm{d}Q^1_{\rho_0,\rho_1}(\rho^1)\,\mathrm{d}Q^2_{\rho_0,\rho_1}(\rho^2) \leqslant$$

$$\frac{1}{2} \left( \int_\Gamma \left( \int_0^1 \mathcal{H}[\rho_t^1]\,\mathrm{d}t \right) \, \mathrm{d}Q^1_{\rho_0,\rho_1}(\rho^1) + \int_\Gamma \left( \int_0^1 \mathcal{H}[\rho_t^2]\,\mathrm{d}t \right) \, \mathrm{d}Q^2_{\rho_0,\rho_1}(\rho^2) \right),$$

with a strict inequality if $Q^1_{\rho_0,\rho_1} \neq Q^2_{\rho_0,\rho_1}$ and if the r.h.s. is finite. Then, we integrate w.r.t. $\gamma$ and notice that, as $Q^1 \neq Q^2$, then $Q^1_{\rho_0,\rho_1} \neq Q^2_{\rho_0,\rho_1}$ for a $\gamma$-non negligible sets of $(\rho_0, \rho_1)$, and as $Q^1$ and $Q^2 \in \mathcal{P}^H_{\mathrm{adm}}(\Gamma)$, the r.h.s. of the equation above is finite for $\gamma$-a.e. $(\rho_0, \rho_1)$. Using Fubini's theorem, we are led to the announced conclusion. $\qquad\square$

## 3. ANALYSIS OF THE DISCRETE PROBLEM

As we explained before, to tackle the continuous problem (CP), we will introduce a discretized (in time) variational problem that approximates the continuous one. In this section, we give a brief heuristic justification of it, prove its well-posedness, and show that the discrete averaged entropy is convex. In the proof of the latter property, we use the *flow interchange* technique that was previously introduced in [15].

The discrete problem is obtained by performing three different approximations:

- We consider a number of discrete times $N + 1 \geqslant 2$. We will use $\tau := 1/N$ as a notation for the time step. The set $T^N \subset [0\,,1]$ will stand for the set of all discrete times, namely

$$T^N := \{k\tau \ : \ k = 0, 1, \ldots, N\}.$$

  In particular, $\Gamma_{T^N} = \mathcal{P}(\Omega)^{N+1}$. We will work with $W_2$-traffic plans on $\Gamma_{T^N}$, i.e. elements of $\mathcal{P}(\Gamma_{T^N})$. According to the representation of the action (2.4), we expect that for a curve $\rho \in \Gamma$,

$$A(\rho) \simeq \sum_{k=1}^N \frac{W_2^2(\rho_{(k-1)\tau}, \rho_{k\tau})}{2\tau}.$$

- The incompressibility constraint will be relaxed. For any $k \in \{1, 2, \ldots, N-1\}$, we penalize the densities $m_{k\tau}(Q)$ which are away from the Lebesgue measure by adding a term $\mathcal{C}_q(m_{k\tau}(Q))$, where $\mathcal{C}_q$ is defined in (2.3). As explained in Section 2, this term is positive and vanishes if and only if $m_{k\tau}(Q) = \mathcal{L}$, moreover it goes to $+\infty$ as $q \to +\infty$ if $m_{k\tau}(Q) \neq \mathcal{L}$.

- We will also add an entropic penalization, i.e. a discretized version of

$$\lambda \int_0^1 H_Q(t)\, \mathrm{d}t,$$

with $\lambda$ a small parameter. This term explains why we select, at the limit $\lambda \to 0$, the minimizers whose total entropy is minimal. It is crucial because it enables us to show that the averaged entropy of the discrete problem converges pointwisely to the averaged entropy of the continuous problem. This pointwise convergence is necessary to ensure that the averaged entropy of the continuous problem is convex. In particular, the limit $\lambda \to 0$ must be taken after $N \to +\infty$ and $q \to +\infty$.

Let us state formally our discrete minimization problem. We fix $N \geqslant 1$ ($\tau := 1/N$), $q > 1$ and $\lambda > 0$ and define $T^N = \{k\tau \ : \ k = 0, 1, \ldots, N\}$. We denote by $\mathcal{A}^{N,q,\lambda} : \mathcal{P}(\Gamma_{T^N}) \to [0, +\infty]$ the functional defined by, for any $Q \in \mathcal{P}(\Gamma_{T^N})$,

$$\mathcal{A}^{N,q,\lambda}(Q) := \sum_{k=1}^N \int_{\Gamma_{T^N}} \frac{W_2^2(\rho_{(k-1)\tau}, \rho_{k\tau})}{2\tau}\, \mathrm{d}Q(\rho) + \sum_{k=1}^{N-1} \mathcal{C}_q(m_{k\tau}(Q)) + \lambda \sum_{k=1}^{N-1} \tau H_Q\left(k\tau\right).$$

The *Discrete Problem* consists in minimizing this functional under the only constraint that the initial and final values are coupled through $\gamma$, the set of such $W_2$-traffic plans being $\mathcal{P}_{\mathrm{bc}}(\Gamma_{T^N})$ (cf. Definition 2.8):

(DP) $$\min\left\{\mathcal{A}^{N,q,\lambda}(Q) \ : \ Q \in \mathcal{P}_{\mathrm{bc}}(\Gamma_{T^N})\right\}.$$

A solution of the discrete problem is a $Q \in \mathcal{P}_{\mathrm{bc}}(\Gamma_{T^N})$ with $\mathcal{A}^{N,q,\lambda}(Q) < +\infty$ which minimizes $\mathcal{A}^{N,q,\lambda}$.

**Proposition 3.1.** *The discrete problem* (DP) *admits a solution.*

*Proof.* We can see that $\mathcal{A}^{N,q,\lambda}$ is a positive l.s.c. functional. Lower semi-continuity of the discretized action and of the entropic penalization are not difficult to see thanks to Lemma 2.11. Moreover, $Q \mapsto \mathcal{C}_q(m_t(Q))$ is the composition of the linear and continuous map $Q \mapsto m_t(Q)$ and of the l.s.c. map $\mathcal{C}_q$, hence is l.s.c.

As the space $\mathcal{P}(\Gamma_{T^N}) = \mathcal{P}(\mathcal{P}(\Omega)^{N+1})$ is compact, $\mathcal{P}_{\mathrm{bc}}(\Gamma_{T^N})$ is also a compact space, thus it is enough to show that there exists one $Q \in \mathcal{P}_{\mathrm{bc}}(\Gamma_{T^N})$ such that $\mathcal{A}^{N,q,\lambda}(Q) < +\infty$. We take $Q$ to be equal to $\gamma$ on the endpoints, and such that $e_{k\tau}\#Q$ is a Dirac mass concentrated on the Lebesgue measure $\mathcal{L}$ for any $k \in \{1, 2, \ldots, N-1\}$. As $\mathcal{H}(\mathcal{L}) = 0$ and as the incompressibility constraint $m_{k\tau}(Q) = \mathcal{L}$ is satisfied for every $k \in \{0, 1, \ldots, N\}$, we can see that for this $Q$ we have

$$\mathcal{A}^{N,q,\lambda}(Q) = \int_{\Gamma_{\{0,1\}}} \frac{W_2^2(\rho_0, \mathcal{L}) + W_2^2(\mathcal{L}, \rho_1)}{2\tau}\, \mathrm{d}\gamma(\rho).$$

As the Wasserstein distance is uniformly bounded by the diameter of $\Omega$, the r.h.s. of the above equation is finite. The conclusion derives from a straightforward application of the direct method of calculus of variations. $\qquad\square$

One could show that the discrete problem (DP) admits a unique solution (it is basically the same proof as Proposition 2.16), but we will not need it. The key result of this section is the following:

**Theorem 3.2.** *Let $Q \in \mathcal{P}_{\mathrm{bc}}(\Gamma_{T^N})$ be a solution of the discrete problem* (DP). *Then the function $k \in \{0, 1, \ldots, N\} \mapsto H_Q(k\tau)$ is convex, i.e. for every $k \in \{1, 2, \ldots, N-1\}$,*

(3.1) $$H_Q\left(k\tau\right) \leqslant \frac{1}{2}H_Q\left((k-1)\tau\right) + \frac{1}{2}H_Q\left((k+1)\tau\right).$$

*Proof.* As $\mathcal{A}^{N,q,\lambda}(Q)$ is finite we know that for every $k \in \{1, 2, \ldots, N-1\}$, $H_Q(k\tau) < +\infty$ and $m_{k\tau}(Q) \in L^q(\Omega)$. Let us remark that if $H_Q(0) = +\infty$ then there is nothing to prove in equality (3.1) for $k = 1$ (the r.h.s. being infinite); and, equivalently, if $H_Q(1) = +\infty$ there is nothing to prove for $k = N - 1$. So from now on, we fix $k \in \{1, 2, \ldots, N-1\}$ such that $H_Q\left((k-1)\tau\right)$, $H_Q\left(k\tau\right)$ and $H_Q\left((k+1)\tau\right)$ are finite, and it is enough to show (3.1) for such a $k$.

We recall that $\Phi : [0, +\infty) \times \mathcal{P}(\Omega) \to \mathcal{P}(\Omega)$ denotes the heat flow, let us call $\Phi^k : [0, +\infty) \times \Gamma_{T^N} \to \Gamma_{T^N}$ the heat flow acting only on the $k$-th component: for any $s \geqslant 0$, $\rho \in \Gamma_{T^N}$ and $l \in \{0, 1, \ldots, N\}$,

$$\Phi_s^k(\rho)(l\tau) := \begin{cases} \Phi_s(\rho_{l\tau}) & \text{if } l = k, \\ \rho_{l\tau} & \text{if } l \neq k. \end{cases}$$

If $s \geqslant 0$, it is clear that $\Phi_s^k$ leaves unchanged the boundary values, thus $\Phi_s^k \# Q \in \mathcal{P}_{\mathrm{bc}}(\Gamma_{T^N})$, and therefore by optimality of $Q$ we have that

$$(3.2) \qquad \mathcal{A}^{N,q,\lambda}(Q) \leqslant \mathcal{A}^{N,q,\lambda}(\Phi_s^k \# Q).$$

Let us expand this formula. We can see (by definition of $H_Q$) that

$$H_{\Phi_s^k \# Q}(l\tau) = \begin{cases} \displaystyle\int_{\Gamma_{T^N}} \mathcal{H}(\Phi_s[\rho_{l\tau}]) \, \mathrm{d}Q(\rho) & \text{if } l = k, \\ H_Q(l\tau) & \text{if } l \neq k. \end{cases}$$

Concerning the term $m_{l\tau}(Q)$, the linearity of the flow enables us to write

$$m_{l\tau}(\Phi_s^k \# Q) = \begin{cases} \Phi_s(m_{l\tau}[Q]) & \text{if } l = k, \\ m_{l\tau}(Q) & \text{if } l \neq k. \end{cases}$$

Let us underline that the linearity of the heat flow is crucial to handle the congestion term. Our proof would not have worked if we would have wanted to show the convexity (w.r.t. time) of a functional (different from the entropy) whose gradient flow in the Wasserstein space were not linear. We can rewrite (3.2) in the following form (all the terms that do not involve the time $k\tau$ cancel):

$$\int_{\Gamma_{T^N}} \frac{W_2^2(\rho_{(k-1)\tau}, \rho_{k\tau}) + W_2^2(\rho_{k\tau}, \rho_{(k+1)\tau})}{2\tau} \, \mathrm{d}Q(\rho) + \mathcal{C}_q(m_{k\tau}(Q)) + \lambda\tau \int_{\Gamma_{T^N}} \mathcal{H}(\rho_{k\tau}) \, \mathrm{d}Q(\rho)$$

$$\leqslant \int_{\Gamma_{T^N}} \frac{W_2^2(\rho_{(k-1)\tau}, \Phi_s\rho_{k\tau}) + W_2^2(\Phi_s\rho_{k\tau}, \rho_{(k+1)\tau})}{2\tau} \, \mathrm{d}Q(\rho) + \mathcal{C}_q(\Phi_s(m_{l\tau}[Q])) + \lambda\tau \int_{\Gamma_{T^N}} \mathcal{H}(\Phi_s\rho_{k\tau}) \, \mathrm{d}Q(\rho).$$

It is a well known fact that the heat flows decreases the $L^q$ norm, thus $\mathcal{C}_q(\Phi_s(m_{k\tau}[Q])) \leqslant \mathcal{C}_q(m_{k\tau}(Q))$. It is also a well known fact the the heat flow decreases the entropy (it is for example encoded in (EI)), thus

$$\int_{\Gamma_{T^N}} \mathcal{H}(\Phi_s\rho_{k\tau}) \, \mathrm{d}Q(\rho) \leqslant \int_{\Gamma_{T^N}} \mathcal{H}(\rho_{k\tau}) \, \mathrm{d}Q(\rho).$$

Therefore, multiplying by $\tau$ and dividing by $s$, we are left with the following inequality, valid for any $s > 0$:

$$\int_{\Gamma_{T^N}} \frac{W_2^2(\rho_{(k-1)\tau}, \Phi_s\rho_{k\tau}) - W_2^2(\rho_{(k-1)\tau}, \rho_{k\tau})}{2s} \, \mathrm{d}Q(\rho) + \int_{\Gamma_{T^N}} \frac{W_2^2(\Phi_s\rho_{k\tau}, \rho_{(k+1)\tau}) - W_2^2(\rho_{k\tau}, \rho_{(k+1)\tau})}{2s} \, \mathrm{d}Q(\rho) \geqslant 0.$$

The integrand of the first integral is exactly the rate of increase of the function $s \mapsto W_2^2(\rho_{(k-1)\tau}, \Phi_s\rho_{k\tau})/2$ whose $\limsup$ is bounded, when $s \to 0$, by $\mathcal{H}(\rho_{(k-1)\tau}) - \mathcal{H}(\rho_{k\tau})$ according to (EVI). Moreover, as the entropy is positive, the same inequality (EVI) shows that this rate of increase is uniformly (in $s$) bounded from above by $\mathcal{H}(\rho_{(k-1)\tau})$, and the latter is integrable w.r.t. to $Q$. Hence by applying a reverse Fatou's lemma, we see that

$$\int_{\Gamma_{T^N}} [\mathcal{H}(\rho_{(k-1)\tau}) - \mathcal{H}(\rho_{k\tau})] \, \mathrm{d}Q(\rho) \geqslant \limsup_{s \to 0} \int_{\Gamma_{T^N}} \frac{W_2^2(\rho_{(k-1)\tau}, \Phi_s\rho_{k\tau}) - W_2^2(\rho_{(k-1)\tau}, \rho_{k\tau})}{2s} \, \mathrm{d}Q(\rho).$$

We have a symmetric minoration for $\displaystyle\int_{\Gamma_{T^N}} [\mathcal{H}(\rho_{(k+1)\tau}) - \mathcal{H}(\rho_{k\tau})] \, \mathrm{d}Q(\rho)$, hence we end up with

$$0 \leqslant \int_{\Gamma_{T^N}} [\mathcal{H}(\rho_{(k-1)\tau}) - \mathcal{H}(\rho_{k\tau})] \, \mathrm{d}Q(\rho) + \int_{\Gamma_{T^N}} [\mathcal{H}(\rho_{(k+1)\tau}) - \mathcal{H}(\rho_{k\tau})] \, \mathrm{d}Q(\rho)$$

$$= \int_{\Gamma_{T^N}} [\mathcal{H}(\rho_{(k-1)\tau}) + \mathcal{H}(\rho_{(k+1)\tau}) - 2\mathcal{H}(\rho_{k\tau})] \, \mathrm{d}Q(\rho)$$

$$= H_Q((k-1)\tau) + H_Q((k+1)\tau) - 2H_Q(k\tau). \qquad \square$$

## 4. LIMIT OF THE DISCRETE PROBLEMS TO THE CONTINUOUS ONE

In all this section, let us denote by $\bar{Q}^{N,q,\lambda}$ a solution (in fact there exists only one but this is not important) of the discrete problem (DP) with parameters $N$, $q$ and $\lambda$. We want to pass to the limit in the following way:

- By sending $q \to +\infty$, the incompressibility constraint $m_t(Q) = \mathcal{L}$ will be strictly enforced at the discrete times $t \in T^N$.
- Then, we will interpolate geodesically between discrete instants and show that this builds a sequence of $W_2$-traffic plans which converges to a limit $\bar{Q}^\lambda \in \mathcal{P}_{\mathrm{adm}}(\Gamma)$ when $N \to +\infty$. This $\bar{Q}^\lambda$ is expected to be a solution

$$\min \left\{ \mathcal{A}(Q) + \lambda \int_0^1 H_Q(t) \, \mathrm{d}t \ : \ Q \in \mathcal{P}^H_{\mathrm{adm}}(\Gamma) \right\}.$$

- In the end, when $\lambda \to 0$, the $W_2$-traffic plans $\bar{Q}^\lambda$ will converge to the solution $\bar{Q}$ of the original problem with minimal total entropy and $\int_0^1 H_{\bar{Q}^\lambda}(t) \, \mathrm{d}t$ will converge to $\int_0^1 H_{\bar{Q}}(t) \, \mathrm{d}t$. This is the convergence of the total entropy that enables us to get a pointwise convergence of the averaged entropy.

Basically, we are performing three successive $\Gamma$-limits. Let us stress out that the order in which the limits are taken is important, though this importance may be hard to see under the various technical details. In particular taking the limit $\lambda \to 0$ at the end is needed to show that at the limit the selected minimizer of the continuous problem is the one with minimal total entropy (cf. the proof of Proposition 4.9). It may be possible to take the limit $N \to +\infty$ first (instead of sending $q \to +\infty$ first), but then the incompressibility constraint must be handled differently from us.

This section is organized as follows. First we show some kind of $\Gamma - \limsup$, i.e. given continuous curves we build discrete ones whose discrete action and total entropy are close to their continuous counterparts. Then, and thanks to these constructions, we show a uniform bound on $\bar{Q}^{N,q,\lambda}$ that allows us to extract converging subsequences toward a limit $\bar{Q}$, and we show that $\bar{Q}$ is a solution of the continuous problem. Finally, we show that $\bar{Q}$ is the minimizer of $\mathcal{A}$ with minimal total entropy and that its averaged entropy is convex.

### 4.1. **Building discrete curves from continuous ones.** Let us first show a result that will be crucial to handle Assumption 2, namely a procedure to regularize curves in order for the total entropy to be finite.

**Proposition 4.1.** *Under Assumption 2, for any $Q \in \mathcal{P}_{\mathrm{adm}}(\Gamma)$ and for any $\varepsilon > 0$, there exists $Q' \in \mathcal{P}^H_{\mathrm{adm}}(\Gamma)$ such that $\mathcal{A}(Q') \leqslant \mathcal{A}(Q) + \varepsilon$ and $H_{Q'} \in L^\infty([0,1])$.*

*Proof.* Let us fix $Q \in \mathcal{P}_{\mathrm{adm}}(\Gamma)$. The idea is to use the heat flow $\Phi$ to regularize the curves: indeed, we know that if $s > 0$ is fixed, then for any $\rho \in \Gamma$, $\mathcal{H}(\Phi_s \rho_t)$ is bounded independently on $t$ and $\rho$. Moreover, applying uniformly the heat flow decreases the action. Indeed, we recall that at a discrete level the Wasserstein distance decreases along the heat flow: it is Inequality (4.1). Using the representation formula (2.4) for the action, one concludes that for a fixed $s \geqslant 0$,

$$(4.1) \qquad \int_0^1 \frac{1}{2} |\dot{\Phi_s \rho_t}|^2 \, \mathrm{d}t \leqslant \int_0^1 \frac{1}{2} |\dot{\rho}_t|^2 \, \mathrm{d}t.$$

However, by doing this, we lose the boundary values. To recover them, we squeeze the curve $\Phi_s \rho$ into the subinterval $[s, 1-s]$, and then use the heat flow (acting on $\rho_0$) to join $\rho_0$ to $\Phi_s(\rho_0)$ on $[0,s]$ and $\Phi_s(\rho_1)$ to $\rho_1$ on $[1-s,1]$. Formally, for $0 < s \leqslant 1/2$, let us define the regularizing operator $R_s : \Gamma \to \Gamma$ by

$$\forall \rho \in \Gamma, \forall t \in [0,1], \ R_s(\rho)(t) := \begin{cases} \Phi_t(\rho_0) & \text{if } 0 \leqslant t \leqslant s, \\ \Phi_s\left(\rho\left[\frac{t-s}{1-2s}\right]\right) & \text{if } s \leqslant t \leqslant 1-s, \\ \Phi_{1-t}(\rho_1) & \text{if } 1-s \leqslant t \leqslant 1. \end{cases}$$

The continuity of the heat flow allows us to assert that $R_s(\rho)$ is a continuous curve. As the entropy decreases along the heat flow (cf. (EI)), and as $\mathcal{H}(R_s[\rho])$ is uniformly bounded on $[s, 1-s]$ (independently on $\rho$), we can see that there exists a constant $C_s$ depending only on $s$ such that

$$(4.2) \qquad \forall \rho \in \Gamma, \ \forall t \in [0,1], \ \mathcal{H}[R_s(\rho)(t)] \leqslant \max(\mathcal{H}(\rho_0), \mathcal{H}(\rho_1), C_s).$$

To estimate the action of $R_s(\rho)$, we use the estimate (4.1) on $[s, 1-s]$ and the identity (EI) to hold the boundary terms:

$$\begin{aligned}
A(R_s(\rho)) &\leqslant \int_0^s \frac{1}{2}|\dot{\Phi}_t \rho_0|^2 \, \mathrm{d}t + \int_s^{1-s} \frac{1}{2}|\dot{\rho}_{(t-s)/(1-2s)}|^2 \, \mathrm{d}t + \int_{1-s}^1 \frac{1}{2}|\dot{\Phi}_{1-t}\rho_1|^2 \, \mathrm{d}t \\
&= \frac{\mathcal{H}(\rho_0) - \mathcal{H}(\Phi_s[\rho_0])}{2} + \frac{1}{1-2s}\int_0^1 \frac{1}{2}|\dot{\rho}_t|^2 \, \mathrm{d}t + \frac{\mathcal{H}(\rho_1) - \mathcal{H}(\Phi_s[\rho_1])}{2} \\
&= \frac{1}{1-2s}A(\rho) + \frac{1}{2}\left(\mathcal{H}(\rho_0) - \mathcal{H}(\Phi_s[\rho_0]) + \mathcal{H}(\rho_1) - \mathcal{H}(\Phi_s[\rho_1])\right).
\end{aligned}$$

In particular, using the lower semi-continuity of the entropy $\mathcal{H}$ and the continuity w.r.t. $s$ of the heat flow, we see that if $\mathcal{H}(\rho_0)$ and $\mathcal{H}(\rho_1)$ are finite,

$$(4.3) \qquad\qquad \limsup_{s \to 0} A(R_s(\rho)) \leqslant A(\rho).$$

We are now ready to use the regularization operator on the $W_2$-traffic plan $Q$. For a fixed $0 < s \leqslant 1/2$, we define $Q_s := R_s \# Q$. As $R_s$ does not change the boundary points, we still have $(e_0, e_1)\# Q_s = \gamma$. Integrating (4.2) w.r.t. $Q$, we get that

$$\forall t \in [0, 1], \ H_{Q_s}(t) \leqslant H_{Q_s}(0) + H_{Q_s}(1) + C_s = H_\gamma(0) + H_\gamma(1) + C_s,$$

and we know that the r.h.s. is finite because of Assumption 2. Concerning the action, since $\mathcal{H}(\rho_0)$ and $\mathcal{H}(\rho_1)$ are finite for $Q$-a.e. $\rho \in \Gamma$, we can integrate (4.3) w.r.t. $Q$ by using a reverse Fatou's lemma to get

$$\limsup_{s \to 0} \mathcal{A}(Q_s) \leqslant \mathcal{A}(Q).$$

It remains to check the incompressibility. For a fixed $t \in [0, 1]$, we notice that $e_t \# Q_s$ is of the form $(\Phi_r \circ e_{t'})\# Q$ for a some $r \geqslant 0$ and $t' \in [0, 1]$ (for example, $r = t$ and $t' = 0$ if $t \in [0, s]$, and $r = s$ and $t' = (t-s)/(1-2s)$ if $t \in [s, 1-s]$). Thus, by linearity of the heat flow, $m_t(Q_s) = \Phi_r(m_{t'}[Q])$. But $m_{t'}(Q) = \mathcal{L}$ for any $t'$ and the Lebesgue measure is preserved by the heat flow, hence $m_t(Q_s) = \mathcal{L}$.

Therefore, the $Q'$ that we take is just $Q_s$ for $s > 0$ small enough. $\qquad\square$

It is then possible to show how one can build a discrete curve from a continuous one in such a way that the action and the total entropy do not increase too much. This is a standard procedure which would be valid for probability on curves valued in arbitrary geodesic spaces.

**Proposition 4.2.** *Let $Q \in \mathcal{P}_{\mathrm{adm}}^H(\Gamma)$ be an admissible $W_2$-traffic plan with finite total entropy. For any $N \geqslant 1$, we can build a $W_2$-traffic plan $Q_N \in \mathcal{P}_{\mathrm{in}}(\Gamma_{T^N})$ in such a way that*

$$\limsup_{N \to +\infty} \mathcal{A}^{N,q,\lambda}(Q_N) \leqslant \mathcal{A}(Q) + \lambda \int_0^1 H_Q(t) \, \mathrm{d}t.$$

*Proof.* We can assume that $\mathcal{A}(Q) < +\infty$. The idea is to sample each curve on a uniform grid, but not necessarily on $T^N$. Indeed, the key point in this sampling is to ensure that the discrete entropic penalization of the functional $\mathcal{A}^{N,q,\lambda}$ is bounded by $\lambda \int_0^1 H_Q(t) \, \mathrm{d}t$. Let us fix $N \geqslant 1$ and recall that $\tau = 1/N$. We can see that

$$\int_0^\tau \sum_{k=1}^{N-1} H_Q\left(k\tau + s\right) \, \mathrm{d}s = \int_\tau^1 H_Q(t) \, \mathrm{d}t \leqslant \int_0^1 H_Q(t) \, \mathrm{d}t.$$

Therefore, there exists $s_N \in (0, \tau)$ such that

$$\tau \sum_{k=1}^{N-1} H_Q\left(k\tau + s_N\right) \leqslant \int_0^1 H_Q(t) \, \mathrm{d}t.$$

We define the sampling operator $S_N : \Gamma \to \Gamma_{T^N}$ (which samples on the grid $\{k\tau + s_N \ : \ k = 1, 2, \dots, N-1\}$) by

$$\forall \rho \in \Gamma, \forall k \in \{0, 1, \dots, N\}, \ S_N(\rho)(k\tau) = \begin{cases} \rho_0 & \text{if } k = 0, \\ \rho_1 & \text{if } k = N, \\ \rho_{k\tau + s_N} & \text{if } 1 \leqslant k \leqslant N-1. \end{cases}$$

Then we simply define $Q_N := S_N \# Q$. As the initial and final values are left unchanged, it is clear that $(e_0, e_1) \# Q_N = (e_0, e_1) \# Q = \gamma$, i.e. $Q_N \in \mathcal{P}_{\mathrm{bc}}(\Gamma_{T^N})$. By construction, we have that

$$\lambda \sum_{k=1}^{N-1} \tau H_{Q_N}(k\tau) = \lambda \tau \sum_{k=1}^{N-1} H_Q(k\tau + s_N) \leqslant \lambda \int_0^1 H_Q(t) \, \mathrm{d}t.$$

Moreover, as $Q \in \mathcal{P}_{\mathrm{adm}}(\Gamma)$ is incompressible, it is clear that $Q_N$ is incompressible too, hence

$$\sum_{k=1}^{N-1} \mathcal{C}_q(m_{k\tau}[Q_N]) = 0.$$

The last term to handle is the action. Indeed, we have to take care of the fact that we use a translated grid which is not uniform close to the boundaries. After a standard computation (which would be valid in any geodesic space) that we do not detail here, one finds that

$$\sum_{k=1}^{N} \int_{\Gamma_{T^N}} \frac{W_2^2(\rho_{(k-1)\tau}, \rho_{k\tau})}{2\tau} \, \mathrm{d}Q_N(\rho) \leqslant \mathcal{A}(Q) + \int_\Gamma \left( \int_0^{2\tau} \frac{1}{2} |\dot\rho_s|^2 \, \mathrm{d}s \right) \, \mathrm{d}Q(\rho).$$

For every 2-absolutely continuous curve, it is clear that the quantity $\int_0^{2\tau} \frac{1}{2} |\dot\rho_s|^2 \, \mathrm{d}s$ goes to $0$ as $N \to +\infty$ and it is dominated by $A(\rho)$ which is integrable w.r.t. $Q$. Therefore, by dominated convergence,

$$\limsup_{N \to +\infty} \left( \sum_{k=1}^{N} \int_{\Gamma_{T^N}} \frac{W_2^2(\rho_{(k-1)\tau}, \rho_{k\tau})}{2\tau} \, \mathrm{d}Q_N(\rho) \right) \leqslant \mathcal{A}(Q).$$

Gluing all the inequalities we have collected on $Q_N$, we see that $\mathcal{A}^{N,q,\lambda}(Q_N)$ satisfies the desired asymptotic bound. $\qquad\square$

**Corollary 4.3.** *Under Assumption 1 or Assumption 2, there exists $C < +\infty$, such that, uniformly in $N \geqslant 1$, $\lambda \in (0\,,1]$ and $q > 1$, we have*

$$\mathcal{A}^{N,q,\lambda}(\bar{Q}^{N,q,\lambda}) \leqslant C.$$

*Proof.* Indeed, it is enough to take $Q$ any element of $\mathcal{P}_{\mathrm{adm}}^H(\Gamma)$ with finite action (it exists by definition under Assumption 1 and we use Proposition 4.1 under Assumption 2), to construct $Q_N$ as in Proposition 4.2, to define $C := \sup_{N \geqslant 1} \mathcal{A}^{N,q,\lambda}(Q_N)$, and to use the fact that $\mathcal{A}^{N,q,\lambda}(\bar{Q}^{N,q,\lambda}) \leqslant \mathcal{A}^{N,q,\lambda}(Q_N) \leqslant C$. $\qquad\square$

### 4.2. Solution of the continuous problem as a limit of discrete solutions.

To go from $W_2$-traffic plans on discrete curves to $W_2$-traffic plans on continuous ones, we will need an extension operator $E_N : \Gamma_{T^N} \to \Gamma$ that interpolates a discrete curve along geodesics in $(\mathcal{P}(\Omega), W_2)$. More precisely,

**Definition 4.4.** *Let $N \geqslant 1$. If $\rho \in \Gamma_{T^N}$, the curve $E_N(\rho) \in \Gamma$ is defined as the one that coincides with $\rho$ on $T^N$ and such that for any $k \in \{0, 1, \dots, N-1\}$, the restriction of $E_N(\rho)$ to $[k\tau\,,(k+1)\tau]$ is a[3] constant-speed geodesic joining $\rho_{k\tau}$ to $\rho_{(k+1)\tau}$.*

In particular, for any $k \in \{0, 1, 2, \dots, N-1\}$, $|\dot{E_N}(\rho)|$ is constant on $[k\tau\,,(k+1)\tau]$ and equal to $W_2(\rho_{k\tau}, \rho_{(k+1)\tau})/\tau$. Thus, we have the identity

$$\int_{k\tau}^{(k+1)\tau} \frac{1}{2} |\dot{E_N}(\rho)_t|^2 \, \mathrm{d}t = \frac{W_2^2(\rho_{k\tau}, \rho_{(k+1)\tau})}{2\tau},$$

summed over $k \in \{0, 1, \dots, N-1\}$, these identities led to

$$(4.4) \qquad\qquad A(E_N[\rho]) = \sum_{k=1}^{N} \frac{W_2^2(\rho_{(k-1)\tau}, \rho_{k\tau})}{2\tau}.$$

In other words, the action of the extended curve $E_N(\rho)$ is equal to the discrete one of $\rho$.

---

[3]One may worry about the non uniqueness of the geodesic and hence of the fact that the extension operator $E_N$ is ill-defined. However, it is a classical result of optimal transport that the constant-speed geodesic joining two measures is unique as soon as one of the two measures is absolutely continuous w.r.t. $\mathcal{L}$. Moreover, for a traffic plan $Q \in \mathcal{P}(\Gamma_{T^N})$, if $H_Q(t) < +\infty$ for $t \in T^N$, then $Q$-a.e. $\rho$ is absolutely continuous w.r.t. $\mathcal{L}$ at time $t$. Thus as long as we work with $W_2$-traffic pans $Q$ such that $H_Q(k\tau) < +\infty$ for any $k \in \{1, 2, \dots, N-1\}$ (and we leave it to the reader to check that it is the case), the operator $E_N$ is well defined.

We are now ready to show the convergence of $\bar{Q}^{N,q,\lambda}$ to some limit $\bar{Q} \in \mathcal{P}_{\mathrm{adm}}(\Gamma)$. We take three sequences $(N_n)_{n\in\mathbf{N}}$, $(q_m)_{m\in\mathbf{N}}$ and $(\lambda_r)_{r\in\mathbf{N}}$ that converge respectively to $+\infty$, $+\infty$ and $0$. We will not relabel the sequences when extracting subsequences. Moreover, to avoid heavy notations, we will drop the indexes $n, m$ and $r$, and $\lim_{n\to+\infty}$, $\lim_{m\to+\infty}$, $\lim_{r\to+\infty}$ will be denoted respectively by $\lim_{N\to+\infty}$, $\lim_{q\to+\infty}$ and $\lim_{\lambda\to0}$.

**Proposition 4.5.** *Under Assumption 1 or Assumption 2, there exists $\bar{Q} \in \mathcal{P}_{\mathrm{adm}}(\Gamma)$, and families $(\bar{Q}^{N,\lambda})_{N,\lambda} \in \mathcal{P}_{\mathrm{adm}}(\Gamma_{T^N})$, $(\bar{Q}^\lambda)_\lambda \in \mathcal{P}_{\mathrm{adm}}(\Gamma)$ such that (up to extraction)*

$$\lim_{q\to+\infty} \bar{Q}^{N,q,\lambda} = \bar{Q}^{N,\lambda} \text{ in } \mathcal{P}(\Gamma_{T^N}),$$

$$\lim_{N\to+\infty} (E_N \# \bar{Q}^{N,\lambda}) = \bar{Q}^\lambda \quad in \quad \mathcal{P}(\Gamma),$$

$$\lim_{\lambda\to0} \bar{Q}^\lambda = \bar{Q} \quad in \quad \mathcal{P}(\Gamma).$$

*Proof.* We denote by $C$ the constant given by Corollary 4.3.

To prove the existence of $(\bar{Q}^{N,\lambda})_{N,\lambda}$, it is enough to notice that for any $N \geqslant 1$ the space $\mathcal{P}(\Gamma_{T^N})$ is compact and therefore every sequence admits a converging subsequence. By continuity of the Wasserstein distance, we know that

$$\sum_{k=1}^N \int_{\Gamma_{T^N}} \frac{W_2^2(\rho_{(k-1)\tau}, \rho_{k\tau})}{2\tau} \, \mathrm{d}\bar{Q}^{N,\lambda}(\rho) = \lim_{q\to+\infty} \left( \sum_{k=1}^N \int_{\Gamma_{T^N}} \frac{W_2^2(\rho_{(k-1)\tau}, \rho_{k\tau})}{2\tau} \, \mathrm{d}\bar{Q}^{N,q,\lambda}(\rho) \right) \leqslant C.$$

To go on, we use (4.4), namely the fact that $E_N$ transforms the discrete action into the continuous one:

$$\mathcal{A}(E_N \# \bar{Q}^{N,\lambda}) = \int_{\Gamma_{T^N}} A(E_N(\rho)) \, \mathrm{d}\bar{Q}^{N,\lambda}(\rho) = \sum_{k=1}^N \int_{\Gamma_{T^N}} \frac{W_2^2(\rho_{(k-1)\tau}, \rho_{k\tau})}{2\tau} \, \mathrm{d}\bar{Q}^{N,\lambda}(\rho) \leqslant C.$$

We know that the functional $\mathcal{A}$ is l.s.c. and that its sublevel sets are compact. Hence, we get the existence of $(\bar{Q}^\lambda)_\lambda$ such that

$$\lim_{N\to+\infty} (E_N \# \bar{Q}^{N,\lambda}) = \bar{Q}^\lambda$$

in $\mathcal{P}(\Gamma)$ and $\mathcal{A}(\bar{Q}^\lambda) \leqslant C$. Applying exactly the same argument, we can conclude at the existence of $\bar{Q} \in \mathcal{P}(\Gamma)$ with

$$\lim_{\lambda\to0} \bar{Q}^\lambda = \bar{Q}$$

in $\mathcal{P}(\Gamma)$ together with $\mathcal{A}(\bar{Q}) \leqslant C$.

It is easy to show that $(e_0, e_1) \# \bar{Q} = \gamma$ as we have that $(e_0, e_1) \# \bar{Q}^{N,q,\lambda} = \gamma$: this condition passes to the limit and is preserved by $E_N$.

It is slightly more difficult to show the incompressibility. Let us first show that $\bar{Q}^{N,\lambda} \in \mathcal{P}_{\mathrm{in}}(\Gamma_{T^N})$. We fix $N, k$ and $\lambda$. As $\mathcal{A}^{N,q,\lambda}(\bar{Q}^{N,q,\lambda}) \leqslant C$, we see that

$$\left( \int_\Omega |m_{k\tau}(\bar{Q}^{N,q,\lambda})|^q \right)^{1/q} \leqslant (C+1)^{1/q}.$$

As $\mathcal{L}(\Omega) = 1$, the $L^q$ norms are increasing with $q$. Thus, for any $q_0 \leqslant q$, we have

$$\left( \int_\Omega |m_{k\tau}(\bar{Q}^{N,q,\lambda})|^{q_0} \right)^{1/q_0} \leqslant (C+1)^{1/q}.$$

Let us take the limit $q \to +\infty$. We have that $m_{k\tau}(\bar{Q}^{N,q,\lambda})$ converges in $\mathcal{P}(\Omega)$ to $m_{k\tau}(\bar{Q}^{N,\lambda})$. As the $L^{q_0}$ norm is l.s.c. w.r.t. the weak convergence of measures, we can see that

$$\left( \int_\Omega |m_{k\tau}(\bar{Q}^{N,\lambda})|^{q_0} \right)^{1/q_0} \leqslant 1.$$

But now $q_0$ is arbitrary, thus the $L^\infty$ norm of $m_{k\tau}(\bar{Q}^{N,\lambda})$ is bounded by $1$. As we know that $m_{k\tau}(\bar{Q}^{N,\lambda})$ is a probability measure and that $\mathcal{L}(\Omega) = 1$, we deduce that $m_{k\tau}(\bar{Q}^{N,\lambda})$ is equal to $1$ $\mathcal{L}$-a.e. on $\Omega$: it exactly means that $m_{k\tau}(\bar{Q}^{N,\lambda}) = \mathcal{L}$. As $\bar{Q}^{N,\lambda}$ also satisfies the boundary conditions, $\bar{Q}^{N,\lambda} \in \mathcal{P}_{\mathrm{adm}}(\Gamma_{T^N})$.

To show that the incompressibility constraint is satisfied by $\bar{Q}^\lambda$ for every $t$, we proceed as follows: let us consider $t \in [0, 1]$ and $N \geqslant 1$. Let $k \in \{0, 1, \ldots, N-1\}$ such that $k\tau \leqslant t \leqslant (k+1)\tau$. We denote by $s \in [0, 1]$ the

real such that $t = (k + s)\tau$. By definition of $E_N$, if $\rho \in \Gamma_{T^N}$, there exists $\bar{\gamma}$ an optimal transport plan between $\rho_{k\tau}$ and $\rho_{(k+1)\tau}$ (i.e. an optimal $\gamma$ in formula (2.1) with $\mu = \rho_{k\tau}$ and $\nu = \rho_{(k+1)\tau}$) such that $E_N(\rho)(t) = \pi_s \# \bar{\gamma}$ with $\pi_s : (x, y) \mapsto (1 - s)x + sy$. For any $a \in C^1(\Omega)$, we can see that

$$
\left| \int_{\Omega} a \, \mathrm{d}[E_N(\rho)(t)] - \int_{\Omega} a \, \mathrm{d}\rho_{k\tau} \right| = \left| \int_{\Omega \times \Omega} (a[(1 - s)x + sy] - a[x]) \, \mathrm{d}\bar{\gamma}(x, y) \right|
$$

$$
\leqslant \int_{\Omega \times \Omega} s |\nabla a(x)||x - y| \, \mathrm{d}\bar{\gamma}(x, y)
$$

$$
\leqslant \sqrt{\int_{\Omega \times \Omega} |\nabla a(x)|^2 \, \mathrm{d}\bar{\gamma}(x, y)} \sqrt{\int_{\Omega \times \Omega} |x - y|^2 \, \mathrm{d}\bar{\gamma}(x, y)}
$$

$$
\leqslant \|\nabla a\|_{L^\infty} W_2(\rho_{k\tau}, \rho_{(k+1)\tau}).
$$

Therefore, if we estimate the action of $m_t(E_N \# \bar{Q}^{N,\lambda})$ on a $C^1$ function $a$, we find that

$$
\left| \int_{\Omega} a \, \mathrm{d}[m_t(E_N \# \bar{Q}^{N,\lambda})] - \int_{\Omega} a(x) \, \mathrm{d}x \right| = \left| \int_{\Omega} a \, \mathrm{d}[m_t(E_N \# \bar{Q}^{N,\lambda})] - \int_{\Omega} a \, \mathrm{d}[m_{k\tau}(\bar{Q}^{N,\lambda})] \right|
$$

$$
\leqslant \int_{\Gamma_{T^N}} \left| \int_{\Omega} a \, \mathrm{d}[E_N(\rho)(t)] - \int_{\Omega} a \, \mathrm{d}\rho_{k\tau} \right| \, \mathrm{d}\bar{Q}^{N,\lambda}(\rho)
$$

$$
\leqslant \|\nabla a\|_{L^\infty} \int_{\Gamma_{T^N}} W_2(\rho_{k\tau}, \rho_{(k+1)\tau}) \, \mathrm{d}\bar{Q}^{N,\lambda}(\rho)
$$

$$
\leqslant \sqrt{2\tau} \|\nabla a\|_{L^\infty} \sqrt{\int_{\Gamma_{T^N}} \frac{W_2^2(\rho_{k\tau}, \rho_{(k+1)\tau})}{2\tau} \, \mathrm{d}\bar{Q}^{N,\lambda}(\rho)}
$$

$$
\leqslant \sqrt{2C\tau} \|\nabla a\|_{L^\infty}.
$$

Taking the limit $N \to +\infty$ (hence $\tau \to 0$), we know that $m_t(E_N \# \bar{Q}^{N,\lambda})$ converges to $m_t(\bar{Q}^\lambda)$, thus we get

$$
\int_{\Omega} a \, \mathrm{d}[m_t(\bar{Q}^\lambda)] = \int_{\Omega} a(x) \, \mathrm{d}x.
$$

As $a$ is an arbitrary $C^1$ function, we have the equality $m_t(\bar{Q}^\lambda) = \mathcal{L}$ for any $t$, in other words, $\bar{Q}^\lambda \in \mathcal{P}_{\mathrm{in}}(\Gamma)$. As we already know that $\bar{Q}^\lambda \in \mathcal{P}_{\mathrm{bc}}(\Gamma)$, we conclude that $\bar{Q}^\lambda \in \mathcal{P}_{\mathrm{adm}}(\Gamma)$ for any $\lambda > 0$. But $\mathcal{P}_{\mathrm{adm}}(\Gamma)$ is closed, therefore $\bar{Q} \in \mathcal{P}_{\mathrm{adm}}(\Gamma)$. $\qquad\square$

With all the previous work, it is easy to conclude that $\bar{Q}$ is a minimizer of $\mathcal{A}$: we just copy a standard proof of $\Gamma$-convergence.

**Proposition 4.6.** *Under Assumption 1 or Assumption 2, $\bar{Q}$ is a solution of the continuous problem* (CP).

*Proof.* We have already seen that $\mathcal{A}(E_N \# \bar{Q}^{N,\lambda}) \leqslant \liminf_{q \to +\infty} \mathcal{A}^{N,q,\lambda}(\bar{Q}^{N,q,\lambda})$. By lower semi-continuity of $\mathcal{A}$, we deduce that

$$
\mathcal{A}(\bar{Q}) \leqslant \liminf_{\lambda \to 0} \left( \liminf_{N \to +\infty} \left( \liminf_{q \to +\infty} \mathcal{A}^{N,q,\lambda}(\bar{Q}^{N,q,\lambda}) \right) \right).
$$

By contradiction, let us assume that there exists $Q \in \mathcal{P}_{\mathrm{adm}}(\Gamma)$ such that $\mathcal{A}(Q) < \mathcal{A}(\bar{Q})$. If we are under Assumption 2, we can regularize it thanks to Proposition 4.1, and under Assumption 1 we know that we can assume that $Q' \in \mathcal{P}_{\mathrm{adm}}^H(\Gamma)$ and $\mathcal{A}(Q') \leqslant \mathcal{A}(Q)$. In any of these two cases, we can assume that there exists $Q \in \mathcal{P}_{\mathrm{adm}}^H(\Gamma)$ such that $\mathcal{A}(Q) < \mathcal{A}(\bar{Q})$. Thanks to Proposition 4.2, we know that we can construct a sequence $Q_N$ with

$$
\limsup_{N \to +\infty} \mathcal{A}^{N,q,\lambda}(Q_N) \leqslant \mathcal{A}(Q) + \lambda \int_0^1 H_Q(t) \, \mathrm{d}t
$$

Taking the limit $\lambda \to 0$ and using $\mathcal{A}(Q) < \mathcal{A}(\bar{Q})$, we get

$$
\limsup_{\lambda \to 0} \left( \limsup_{N \to +\infty} \mathcal{A}^{N,q,\lambda}(Q_N) \right) < \liminf_{\lambda \to 0} \left( \liminf_{N \to +\infty} \left( \liminf_{q \to +\infty} \mathcal{A}^{N,q,\lambda}(\bar{Q}^{N,q,\lambda}) \right) \right).
$$

Taking $N$ and $q$ large enough and $\lambda$ small enough, one has $\mathcal{A}^{N,q,\lambda}(Q_N) < \mathcal{A}^{N,q,\lambda}(\bar{Q}^{N,q,\lambda})$, which contradicts the optimality of $\bar{Q}^{N,q,\lambda}$. $\hfill\square$

4.3. **Behavior of the averaged entropy of $\bar{Q}$.** Now, we will show that $H_{\bar{Q}} \in L^1([0,1])$ and that $\bar{Q}$ is the minimizer of $\mathcal{A}$ with minimal total entropy. If $Q \in \mathcal{P}(\Gamma_{T^N})$, let us denote by $H_Q^{\text{int}} : [0,1] \to [0,+\infty]$ the piecewise affine interpolation of $H_Q$. More precisely, if $k \in \{0,1,\ldots,N-1\}$ and $s \in [0,1]$, we define

$$H_Q^{\text{int}}((k+s)\tau) := (1-s)H_Q(k\tau) + sH_Q((k+1)\tau).$$

We show the following estimate, which relies on the lower semi-continuity of the entropy:

**Proposition 4.7.** *For any $t \in [0,1]$, we have the following upper bound for $H_{\bar{Q}}(t)$:*

$$H_{\bar{Q}}(t) \leqslant \liminf_{\lambda \to 0}\left(\liminf_{N \to +\infty}\left(\liminf_{q \to +\infty} H_{\bar{Q}^{N,q,\lambda}}^{int}(t)\right)\right).$$

*Proof.* For a fixed $t$, $Q \mapsto H_Q(t)$ is l.s.c. (Lemma 2.11). Thus, for any $k \in \{0,1,2,\ldots,N\}$, we have

$$H_{\bar{Q}^{N,\lambda}}(k\tau) \leqslant \liminf_{q \to +\infty} H_{\bar{Q}^{N,q,\lambda}}(k\tau).$$

Then, to pass to the limit $N \to +\infty$, we will use the fact that the entropy is geodesically convex, i.e. convex along the constant-speed geodesics. Recall that $E_N : \Gamma_{T^N} \to \Gamma$ is the extension operator that interpolates along constant-speed geodesics. Let us take $\rho \in \Gamma_{T^N}$. By geodesic convexity, we have for any $k \in \{0,1,\ldots,N-1\}$ and $s \in [0,1]$

$$\mathcal{H}\left[E_N(\rho)((k+s)\tau)\right] \leqslant (1-s)\mathcal{H}(\rho_{k\tau}) + s\mathcal{H}(\rho_{(k+1)\tau}).$$

Integrating this inequality over $\Gamma_{T^N}$ w.r.t. $\bar{Q}^{N,\lambda}$, we get

$$
\begin{aligned}
H_{E_N\#\bar{Q}^{N,\lambda}}((k+s)\tau) &\leqslant (1-s)H_{\bar{Q}^{N,\lambda}}(k\tau) + sH_{\bar{Q}^{N,\lambda}}((k+1)\tau) \\
&\leqslant \liminf_{q \to +\infty}\left[(1-s)H_{\bar{Q}^{N,q,\lambda}}(k\tau) + sH_{\bar{Q}^{N,q,\lambda}}((k+1)\tau)\right] \\
&= \liminf_{q \to +\infty}\left[H_{\bar{Q}^{N,q,\lambda}}^{\text{int}}((k+s)\tau)\right]
\end{aligned}
$$

We take the limit $N \to +\infty$, followed by $\lambda \to 0$ to get (thanks to the lower semi-continuity of the averaged entropy) the announced inequality. $\hfill\square$

We derive a useful consequence, which implies Theorem 2.14.

**Corollary 4.8.** *Under Assumption 2, the function $H_{\bar{Q}}$ is bounded by $\max(H_\gamma(0), H_\gamma(1))$.*

*Proof.* This is where we use the work of Section 3: thanks to Theorem 3.2, we know that $H_{\bar{Q}^{N,q,\lambda}}$ is convex and therefore bounded by the values at its endpoints which happen to be finite (independently of $N, q$ or $\lambda$):

$$\forall k \in \{0,1,2,\ldots,N\}, \; H_{\bar{Q}^{N,q,\lambda}}(k\tau) \leqslant \max(H_\gamma(0), H_\gamma(1)).$$

Thus the function $H_{\bar{Q}^{N,q,\lambda}}^{\text{int}}$ is also bounded uniformly on $[0,1]$ by $\max(H_\gamma(0), H_\gamma(1))$. Proposition 4.7 allows us to conclude that the same bound holds for $H_{\bar{Q}}$. $\hfill\square$

As we have now proved Theorem 2.14, we will work only under Assumption 1. It remains to show that the $\bar{Q}$ we constructed is the one with minimal total entropy. This is done thanks to the entropic penalization, and is standard in $\Gamma$-convergence theory, the specific structure of the Wasserstein space does not play any role.

**Proposition 4.9.** *For any $Q \in \mathcal{P}_{\text{adm}}^H(\Gamma)$ solution of the continuous problem* (CP)*, we have*

$$\int_0^1 H_{\bar{Q}}(t)\,\mathrm{d}t \leqslant \int_0^1 H_Q(t)\,\mathrm{d}t$$

*Proof.* Let us start with an exact quadrature formula for $H^{\text{int}}_{\bar{Q}^{N,q,\lambda}}$ :

$$\int_\tau^{1-\tau} H^{\text{int}}_{\bar{Q}^{N,q,\lambda}}(t)\,\mathrm{d}t = \frac{\tau}{2}H_{\bar{Q}^{N,q,\lambda}}(\tau) + \tau\sum_{k=2}^{N-2}H_{\bar{Q}^{N,q,\lambda}}(k\tau) + \frac{\tau}{2}H_{\bar{Q}^{N,q,\lambda}}(1-\tau) \leqslant \tau\sum_{k=1}^{N-1}H_{\bar{Q}^{N,q,\lambda}}(k\tau)$$

Then we take successively the limits $q \to +\infty$, $N \to +\infty$ and $\lambda \to 0$, applying Fatou's lemma and using Proposition 4.7 to get

(4.5) 
$$\int_0^1 H_{\bar{Q}}(t)\,\mathrm{d}t \leqslant \liminf_{\lambda\to 0}\left(\liminf_{N\to+\infty}\left(\liminf_{q\to+\infty}\left(\tau\sum_{k=1}^{N-1}H_{\bar{Q}^{N,q,\lambda}}\left[k\tau\right]\right)\right)\right).$$

On the other hand, let us show that the r.h.s. of (4.5) is smaller than the total entropy of any minimizer of (CP). Indeed, assume that this is not the case for some $Q \in \mathcal{P}_{\text{adm}}(\Gamma)$ solution of (CP). In particular, for some $\lambda > 0$ small enough, we have the strict inequality

$$\int_0^1 H_Q(t)\,\mathrm{d}t < \liminf_{N\to+\infty}\left(\liminf_{q\to+\infty}\left(\tau\sum_{k=1}^{N-1}H_{\bar{Q}^{N,q,\lambda}}\left[k\tau\right]\right)\right).$$

Using the fact that $\mathcal{A}(Q) \leqslant \mathcal{A}(\bar{Q}^\lambda)$ by optimality of $Q$, and thanks to the lower semi-continuity of the action,

$$\mathcal{A}(Q) \leqslant \mathcal{A}(\bar{Q}^\lambda) \leqslant \liminf_{N\to+\infty}\left(\liminf_{q\to+\infty}\left(\sum_{k=1}^N\int_{\Gamma_{T^N}}\frac{W_2^2(\rho_{(k-1)\tau},\rho_{k\tau})}{2\tau}\,\mathrm{d}\bar{Q}^{N,q,\lambda}(\rho)\right)\right).$$

Therefore, gluing these two estimates together, we obtain

$$\mathcal{A}(Q) + \lambda\int_0^1 H_Q(t)\,\mathrm{d}t < \liminf_{N\to+\infty}\left(\liminf_{q\to+\infty}\left(\sum_{k=1}^N\int_{\Gamma_{T^N}}\frac{W_2^2(\rho_{(k-1)\tau},\rho_{k\tau})}{2\tau}\,\mathrm{d}\bar{Q}^{N,q,\lambda}(\rho) + \lambda\sum_{k=1}^{N-1}\tau H_{\bar{Q}^{N,q,\lambda}}[k\tau]\right)\right).$$

But if we build the $Q_N$ from $Q$ as in Proposition 4.2, we get, for $N$ and $q$ large enough,

$$\mathcal{A}^{N,q,\lambda}(Q_N) < \sum_{k=1}^N\int_{\Gamma_{T^N}}\frac{W_2^2(\rho_{(k-1)\tau},\rho_{k\tau})}{2\tau}\,\mathrm{d}\bar{Q}^{N,q,\lambda}(\rho) + \lambda\sum_{k=1}^{N-1}\tau H_{\bar{Q}^{N,q,\lambda}}(k\tau) \leqslant \mathcal{A}^{N,q,\lambda}(\bar{Q}^{N,q,\lambda}),$$

which is a contradiction with the optimality of $\bar{Q}^{N,q,\lambda}$. Hence, we have proved that for any $Q \in \mathcal{P}_{\text{adm}}(\Gamma)$ solution of the continuous problem,

(4.6) 
$$\int_0^1 H_{\bar{Q}}(t)\,\mathrm{d}t \leqslant \liminf_{\lambda\to 0}\left(\liminf_{N\to+\infty}\left(\liminf_{q\to+\infty}\left(\tau\sum_{k=1}^{N-1}H_{\bar{Q}^{N,q,\lambda}}\left[k\tau\right]\right)\right)\right) \leqslant \int_0^1 H_Q(t)\,\mathrm{d}t. \qquad \square$$

Now it remains to show that $H_{\bar{Q}}$ is a convex function of time. This will be done by proving that $H_{\bar{Q}}$ is the limit of $H^{\text{int}}_{\bar{Q}^{N,q,\lambda}}$.

**Proposition 4.10.** *Under Assumption 1, for a.e.* $t \in [0,1]$,

$$H_{\bar{Q}}(t) = \lim_{\lambda\to 0}\left(\lim_{N\to+\infty}\left(\lim_{q\to+\infty}\left(H^{int}_{\bar{Q}^{N,q,\lambda}}(t)\right)\right)\right).$$

*Proof.* Taking $Q = \bar{Q}$ in (4.6), we see that, up to extraction,

$$\int_0^1 H_{\bar{Q}}(t)\,\mathrm{d}t = \lim_{\lambda\to 0}\left(\lim_{N\to+\infty}\left(\lim_{q\to+\infty}\left(\tau\sum_{k=1}^{N-1}H_{\bar{Q}^{N,q,\lambda}}\left[k\tau\right]\right)\right)\right).$$

In other words, the integral over time of the discrete averaged entropy converges to the integral of the continuous one. As we know moreover that the discrete averaged entropy is an upper bound for the continuous one (Proposition 4.7), it is not difficult to show that the discrete averaged entropy converges (up to extraction) a.e. to the continuous one. $\qquad \square$

4.4. **From convexity a.e. to true convexity.** Proposition 4.10 is slightly weaker than the result we claimed, as we get information about $H_{\bar{Q}}$ only for a.e. time. The first step toward true convexity is to show that, under Assumption 2, the averaged entropy is everywhere below the line joining the endpoints.

**Proposition 4.11.** *Under Assumption 2, for any $t \in [0,1]$, we have*

$$H_{\bar{Q}}(t) \leqslant (1-t)H_{\bar{Q}}(0) + tH_{\bar{Q}}(1).$$

*Proof.* From Proposition 4.10, we know that $H_{\bar{Q}}$ is a.e. the limit of the functions $H^{\mathrm{int}}_{Q^{N,q,\lambda}}$. Thanks to Theorem 3.2, we can assert that for any $t \in [0,1]$, one has $H^{\mathrm{int}}_{Q^{N,q,\lambda}}(t) \leqslant (1-t)H^{\mathrm{int}}_{Q^{N,q,\lambda}}(0) + tH^{\mathrm{int}}_{Q^{N,q,\lambda}}(1)$. We also know that $H_{\bar{Q}}$ and $H^{\mathrm{int}}_{Q^{N,q,\lambda}}$ coincide for $t=0$ and $t=1$. Therefore, for a.e. $t \in [0,1]$,

$$
\begin{aligned}
H_{\bar{Q}}(t) &= \lim_{\lambda \to 0} \left( \lim_{N \to +\infty} \left( \lim_{q \to +\infty} \left( H^{\mathrm{int}}_{Q^{N,q,\lambda}}(t) \right) \right) \right) \\
&\leqslant \lim_{\lambda \to 0} \left( \lim_{N \to +\infty} \left( \lim_{q \to +\infty} \left( (1-t)H^{\mathrm{int}}_{Q^{N,q,\lambda}}[0] + tH^{\mathrm{int}}_{Q^{N,q,\lambda}}[1] \right) \right) \right) \\
&= (1-t)H_{\bar{Q}}(0) + tH_{\bar{Q}}(1).
\end{aligned}
$$

As $H_{\bar{Q}}$ is l.s.c., we see that the above inequality is valid for any $t \in [0,1]$. $\qquad\square$

Now, if $\bar{Q}$ is the solution of the continuous problem (CP) with minimal total entropy, then its restriction to any subinterval of $[0,1]$ is also optimal: for any $0 \leqslant t_1 < t_2 \leqslant 1$, $e_{[t_1,t_2]}\#\bar{Q}$ is also the solution of the continuous problem (on $[t_1,t_2]$) with boundary conditions $e_{\{t_1,t_2\}}\#\bar{Q}$ with minimal total entropy. This is already known [1, Remark 3.2 and below] and comes from the fact that we can concatenate traffic plans.

**Proposition 4.12.** *Let $0 \leqslant t_1 < t_2 \leqslant 1$. Then for any $Q \in \mathcal{P}_{\mathrm{adm}}(\Gamma_{[t_1,t_2]})$ such that $e_{\{t_1,t_2\}}\#Q = e_{\{t_1,t_2\}}\#\bar{Q}$, we have*

$$\int_\Gamma \left( \int_{t_1}^{t_2} \frac{1}{2}|\dot{\rho}_t|^2 \, \mathrm{d}t \right) \, \mathrm{d}\bar{Q}(\rho) \leqslant \int_{\Gamma_{[t_1,t_2]}} \left( \int_{t_1}^{t_2} \frac{1}{2}|\dot{\rho}_t|^2 \, \mathrm{d}t \right) \, \mathrm{d}Q(\rho).$$

*Moreover, if the inequality above is an equality, then*

$$\int_{t_1}^{t_2} H_{\bar{Q}}(t) \, \mathrm{d}t \leqslant \int_{t_1}^{t_2} H_Q(t) \, \mathrm{d}t.$$

*Proof.* This property relies on the fact that if $Q \in \mathcal{P}_{\mathrm{adm}}(\Gamma_{[t_1,t_2]})$ with $e_{\{t_1,t_2\}}\#Q = e_{\{t_1,t_2\}}\#\bar{Q}$, we can concatenate $Q$ and $\bar{Q}$ together to build a $W_2$-traffic plan $Q' \in \mathcal{P}(\Gamma)$ such that $e_{[0,1]\setminus[t_1,t_2]}\#Q' = e_{[0,1]\setminus[t_1,t_2]}\#\bar{Q}$ and $e_{[t_1,t_2]}\#Q' = e_{[t_1,t_2]}\#Q$. To do that, it is enough to disintegrate the measures $\bar{Q}$ and $Q$ w.r.t. $e_{\{t_1,t_2\}}$ and then to concatenate elements of $\Gamma_{[0,1]\setminus[t_1,t_2]}$ and $\Gamma_{[t_1,t_2]}$ which coincides on $\{t_1,t_2\}$: we leave the details to the reader. $\qquad\square$

Combining the two above propositions, we recover the convexity of $H_{\bar{Q}}$. Let us remark that we rely on the fact that the minimizer of $\mathcal{A}$ with minimal total entropy is unique.

**Corollary 4.13.** *Under Assumption 1 or Assumption 2, for any $0 \leqslant t_1 < t_2 \leqslant 1$ and any $s \in (0,1)$,*

$$H_{\bar{Q}}((1-s)t_1 + st_2) \leqslant (1-s)H_{\bar{Q}}(t_1) + sH_{\bar{Q}}(t_2).$$

*Proof.* If the r.h.s. is infinite, there is nothing to prove. Therefore, we can assume that $H_{\bar{Q}}(t_1)$ and $H_{\bar{Q}}(t_2)$ are finite. By uniqueness of the solution with minimal total entropy (Proposition 2.16), we know that $e_{[t_1,t_2]}\#\bar{Q}$ coincides with the solution of the continuous problem (CP) with minimal total entropy on $[t_1,t_2]$ with boundary conditions $e_{\{t_1,t_2\}}\#\bar{Q}$ (Proposition 4.12). As $H_{\bar{Q}}(t_1)$ and $H_{\bar{Q}}(t_2)$ are finite, Assumption 2 is satisfied for the continuous problem on $[t_1,t_2]$ and therefore we can apply Proposition 4.11 to get

$$H_{\bar{Q}}((1-s)t_1 + st_2) \leqslant (1-s)H_{\bar{Q}}(t_1) + sH_{\bar{Q}}(t_2). \qquad\square$$

## 5. Equivalence with the parametric formulation of the Euler equation

In this section we will explain why our non-parametric formulation is equivalent to Brenier's parametric one. From the way we build it, it is clear that our formulation admits more potential solutions than Brenier's one, so the only technical point will be to show that, if the boundary data are in a parametric form, it is possible to parametrize the *a priori* non-parametric solution of the continuous problem.

Let us take $\mathfrak{A}$ a polish space and consider $\theta \in \mathcal{P}(\mathfrak{A})$ a Borel probability measure on $\mathfrak{A}$. We will assume that we have two families (the initial and the final) $(\rho_i^\alpha)_{\alpha \in \mathfrak{A}}$ and $(\rho_f^\alpha)_{\alpha \in \mathfrak{A}}$ of probabilities measures on $\Omega$ indexed by $\mathfrak{A}$. We denote by $P_{\mathrm{bc}} : \mathfrak{A} \to \Gamma_{\{0,1\}} = \mathcal{P}(\Omega)^2$ the parametrization of the boundary conditions, simply defined by $P_{\mathrm{bc}}(\alpha) = (\rho_i^\alpha, \rho_f^\alpha)$ and assume that it is measurable. We assume that the boundary data satisfy the incompressibility condition, i.e.

$$\int_{\mathfrak{A}} \rho_i^\alpha \, \mathrm{d}\theta(\alpha) = \mathcal{L} \ \text{ and } \ \int_{\mathfrak{A}} \rho_f^\alpha \, \mathrm{d}\theta(\alpha) = \mathcal{L}.$$

Translated in our language, if we set $\gamma := P_{\mathrm{bc}} \# \theta$, we simply impose that $m_0(\gamma) = m_1(\gamma) = \mathcal{L}$.

A measurable family $(\rho_t^\alpha, \mathbf{v}_t^\alpha)_{(\alpha,t) \in \mathfrak{A} \times [0,1]}$ indexed by $\alpha$ and $t$ such that, for $\theta$-a.e. $\alpha$, $(t \mapsto \rho_t^\alpha) \in \Gamma$ and $\mathbf{v}_t^\alpha \in L^2(\Omega, \mathbf{R}^d, \rho_t^\alpha)$ for a.e. $t$, is said to be admissible if

$$\begin{cases} \rho_0^\alpha = \rho_i^\alpha \text{ and } \rho_1^\alpha = \rho_f^\alpha & \text{for } \theta\text{-a.e. } \alpha, \\ \partial_t \rho_t^\alpha + \nabla \cdot (\rho_t^\alpha \mathbf{v}_t^\alpha) = 0 & \text{in a weak sense with no-flux boundary conditions for } \theta\text{-a.e. } \alpha, \\ \int_{\mathfrak{A}} \rho_t^\alpha \, \mathrm{d}\theta(\alpha) = \mathcal{L} & \text{for all } t \in [0,1]. \end{cases}$$

The first equation corresponds to the temporal boundary conditions, the second one is the continuity equation while the last one is the coding of the incompressibility. If $(\rho_t^\alpha, \mathbf{v}_t^\alpha)_{(\alpha,t) \in \mathfrak{A} \times [0,1]}$ is an admissible family, we define its (parametrized) action $\mathcal{A}_P$ by

$$\mathcal{A}_P(\rho, \mathbf{v}) := \int_{\mathfrak{A}} \int_0^1 \int_\Omega \frac{1}{2} |\mathbf{v}_t^\alpha(x)|^2 \, \mathrm{d}\rho_t^\alpha(x) \, \mathrm{d}t \, \mathrm{d}\theta(\alpha)$$

and its parametrized averaged entropy $H_P(\rho, \mathbf{v}) : [0,1] \to \mathbf{R}$ by, for any $t \in [0,1]$,

$$H_P(\rho, \mathbf{v})(t) := \int_{\mathfrak{A}} \mathcal{H}(\rho_t^\alpha) \, \mathrm{d}\theta(\alpha).$$

The first proposition is very simple: it asserts that every parametric family can be seen as an non parametric one. In the sequel, we define the boundary conditions $\gamma \in \mathcal{P}_{\mathrm{in}}(\Gamma_{\{0,1\}})$ for the non-parametric problem by $\gamma := P_{\mathrm{bc}} \# \theta$.

**Proposition 5.1.** *Let* $(\rho_t^\alpha, \mathbf{v}_t^\alpha)_{(\alpha,t) \in \mathfrak{A} \times [0,1]}$ *be an admissible family. Then there exists* $Q \in \mathcal{P}_{\mathrm{adm}}(\Gamma)$ *such that* $\mathcal{A}(Q) \leqslant \mathcal{A}_P(\rho, \mathbf{v})$ *and* $H_Q(t) = H_P(\rho, \mathbf{v})(t)$ *for any* $t \in [0,1]$.

*Proof.* Let $P : \mathfrak{A} \to \Gamma$, defined by $P(\alpha) = (t \mapsto \rho_t^\alpha)$ be the parametrization. We set $Q := P \# \theta$ and leave it to the reader to check that this choice works (Theorem 2.4 might be useful). $\qquad \square$

The reverse proposition is slightly more difficult to prove: it asserts that one can always build a parametric family from a non-parametric $W_2$-traffic plan in such a way that the global action and the total entropy decrease. In particular, it implies together with Proposition 5.1 that (provided that the boundary conditions are in a parametric form) the solution of the continuous problem (CP) with minimal total entropy can be parametrized.

**Proposition 5.2.** *Let* $Q \in \mathcal{P}_{\mathrm{adm}}(\Gamma)$. *Then there exists an admissible family* $(\rho_t^\alpha, \mathbf{v}_t^\alpha)_{(\alpha,t) \in \mathfrak{A} \times [0,1]}$ *such that* $\mathcal{A}_P(\rho, \mathbf{v}) \leqslant \mathcal{A}(Q)$ *and* $H_P(\rho, \mathbf{v})(t) \leqslant H_Q(t)$ *for any* $t \in [0,1]$.

*Proof.* Let us disintegrate $Q$ w.r.t. to $e_{\{0,1\}} = (e_0, e_1)$. We obtain a family $(Q_{\rho_0, \rho_1})_{\rho_0, \rho_1}$ of $W_2$-traffic plans indexed by $(\rho_0, \rho_1) \in \Gamma_{\{0,1\}} = \mathcal{P}(\Omega)^2$. We define the curve $\rho_t^\alpha$ as the average of all the curves in $\Gamma$ w.r.t. to $Q_{\rho_i^\alpha, \rho_f^\alpha}$: for any $t \in [0,1]$ and any $\alpha$ for which $Q_{\rho_i^\alpha, \rho_f^\alpha}$ is defined (and this property holds for $\theta$-a.e. $\alpha$), we set

$$\rho_t^\alpha := m_t \left( Q_{\rho_i^\alpha, \rho_f^\alpha} \right).$$

By definition of disintegration, $e_{\{0,1\}} \# Q_{\rho_i^\alpha, \rho_f^\alpha}$ is a Dirac mass at the point $(\rho_i^\alpha, \rho_f^\alpha)$, thus the boundary conditions are satisfied. The incompressibility condition is just a consequence of the incompressibility of $Q$: for any $a \in C(\Omega)$,

$$
\begin{aligned}
\int_{\mathfrak{A}} \left( \int_{\Omega} a(x) \, \mathrm{d}\rho_t^\alpha(x) \right) \mathrm{d}\theta(\alpha) &= \int_{\mathfrak{A}} \left( \int_\Gamma \left[ \int_\Omega a(x) \, \mathrm{d}\rho_t(x) \right] \mathrm{d}Q_{\rho_i^\alpha, \rho_f^\alpha}(\rho) \right) \mathrm{d}\theta(\alpha) \\
&= \int_{\Gamma_{\{0,1\}}} \left( \int_\Gamma \left[ \int_\Omega a(x) \, \mathrm{d}\rho_t(x) \right] \mathrm{d}Q_{\rho_0, \rho_1}(\rho) \right) \mathrm{d}\gamma(\rho_0, \rho_1) \\
&= \int_\Gamma \left( \int_\Omega a(x) \, \mathrm{d}\rho_t(x) \right) \mathrm{d}Q(\rho) \\
&= \int_\Omega a(x) \, \mathrm{d}x.
\end{aligned}
$$

To handle the action, we use the fact that $A$ is convex and l.s.c. Thus, thanks to Jensen's inequality, for $\theta$-a.e. $\alpha$,

$$
A(\rho^\alpha) \leqslant \int_\Gamma A(\rho) \, \mathrm{d}Q_{\rho_i^\alpha, \rho_f^\alpha}(\rho).
$$

Integrating w.r.t. $\theta$, we end up with

$$
\int_{\mathfrak{A}} A(\rho^\alpha) \, \mathrm{d}\theta(\alpha) \leqslant \mathcal{A}(Q).
$$

We consider only the case $\mathcal{A}(Q) < +\infty$ (else there is nothing to prove). Thus, for $\theta$-a.e. $\alpha$ the quantity $A(\rho^\alpha)$ is finite. By Theorem 2.4, we can find for each $\alpha$ a family $(\mathbf{v}_t^\alpha)_{t \in [0,1]}$ of functions $\Omega \to \mathbf{R}^d$ such that the continuity equation is satisfied, $\mathbf{v}_t^\alpha \in L^2(\Omega, \mathbf{R}^d, \rho_t^\alpha)$ for a.e. $t$ and such that the following identity holds

$$
\int_0^1 \int_\Omega \frac{1}{2} |\mathbf{v}_t^\alpha(x)|^2 \, \mathrm{d}\rho_t^\alpha(x) \, \mathrm{d}t = \int_{\mathfrak{A}} \int_0^1 \frac{1}{2} |\dot{\rho}_t^\alpha|^2 \, \mathrm{d}t.
$$

Therefore, we see that the family $(\rho_t^\alpha, \mathbf{v}_t^\alpha)_{(\alpha, t) \in \mathfrak{A} \times [0,1]}$ is admissible and, integrating the last equality w.r.t. $\theta$, that $\mathcal{A}_P(\rho, \mathbf{v}) \leqslant \mathcal{A}(Q)$.

To get the inequality involving the entropy, we use the fact that the functional $\mathcal{H}$ is convex and l.s.c. on $\mathcal{P}(\Omega)$, thus by Jensen's inequality,

$$
\mathcal{H}\left( m_t \left( Q_{\rho_i^\alpha, \rho_f^\alpha} \right) \right) \leqslant \int_\Gamma \mathcal{H}(\rho_t) \, \mathrm{d}Q_{\rho_i^\alpha, \rho_f^\alpha}(\rho).
$$

Integrating w.r.t. $\theta$ leads to the announced inequality. $\qquad \square$

## ACKNOWLEDGMENTS

## REFERENCES

[1] Luigi Ambrosio and Alessio Figalli. Geodesics in the space of measure-preserving maps and plans. *Archive for rational mechanics and analysis*, 194(2):421–462, 2009.

[2] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient flows in metric spaces and in the space of probability measures*. Birkhäuser Verlag, 2008.

[3] Luigi Ambrosio and Paolo Tilli. *Topics on Analysis in Metric Spaces*. Oxford Lecture Series in Mathematics and Its Applications. 2003.

[4] Wolfgang Arendt. Semigroups and evolution equations: functional calculus, regularity and kernel estimates. *Handbook of Differential Equations: Evolutionary Equations*, 1:1–85, 2002.

[5] Vladimir Arnold. Sur la géométrie différentielle des groupes de Lie de dimension infinie et ses applications à l'hydrodynamique des fluides parfaits. In *Annales de l'institut Fourier*, volume 16, pages 319–361, 1966.

[6] Marc Bernot, Vincent Caselles, and Jean-Michel Morel. Traffic plans. *Publicacions Matemàtiques*, 49(2):417–451, 2005.

[7] Marc Bernot, Alessio Figalli, and Filippo Santambrogio. Generalized solutions for the Euler equations in one and two dimensions. *Journal de mathématiques pures et appliquées*, 91(2):137–155, 2009.

[8] Yann Brenier. The least action principle and the related concept of generalized flows for incompressible perfect fluids. *Journal of the American Mathematical Society*, 2(2):225–255, 1989.

[9] Yann Brenier. Minimal geodesics on groups of volume-preserving maps and generalized solutions of the Euler equations. *Communications on pure and applied mathematics*, 52(4):411–452, 1999.

[10] Yann Brenier. Extended Monge-Kantorovich theory. In *Optimal transportation and applications*, pages 91–121. Springer, 2003.

[11] Sara Daneri and Alessio Figalli. Variational models for the incompressible Euler equations. *HCDTE Lecture Notes. Part II. Nonlinear Hyperbolic PDEs, Dispersive and Transport Equations. AIMS Book Series, Applied Mathematics, to appear*, 2012.

[12] Irene Fonseca and Gareth Parry. Equilibrium configurations of defective crystals. *Archive for rational mechanics and analysis*, 120(3):245–283, 1992.

[13] Richard Jordan, David Kinderlehrer, and Felix Otto. The variational formulation of the Fokker–Planck equation. *SIAM journal on mathematical analysis*, 29(1):1–17, 1998.

[14] Hugo Lavenant and Filippo Santambrogio. Optimal density evolution with congestion: $L^\infty$ bounds via flow interchange techniques and applications to variational Mean Field Games. *arXiv preprint arXiv:1705.05658*, 2017.

[15] Daniel Matthes, Robert J. McCann, and Giuseppe Savaré. A family of nonlinear fourth order equations of gradient flow type. *Communications in Partial Differential Equations*, 34(11):1352–1397, 2009.

[16] Michel Pierre. Uniqueness of the solutions of $u_t - \Delta\phi(u) = 0$ with initial datum a measure. *Nonlinear Analysis: Theory, Methods & Applications*, 6(2):175–187, 1982.

[17] Filippo Santambrogio. Optimal transport for applied mathematicians. *Progress in Nonlinear Differential Equations and their applications*, 87, 2015.

[18] Filippo Santambrogio. {Euclidean, Metric, and Wasserstein} Gradient Flows: an overview , 2016. Preprint avalaible at `http://cvgmt.sns.it/paper/3165/`.

[19] Cédric Villani. *Topics in optimal transportation*. Number 58. American Mathematical Soc., 2003.