Записки научных
семинаров ПОМИ
Том 390, 2011 г.

L. Brasco

A SURVEY ON DYNAMICAL TRANSPORT DISTANCES

Abstract. In this paper we review some transport models based on the continuity equation, starting with the so-called *Benamou-Brenier formula*, which is nothing but a fluid mechanics reformulation of the Monge-Kantorovich problem with cost $c(x,y) = |x-y|^2$. We discuss some of its applications (gradient flows, sharp functional inequalities...), as well as some variants and generalizations to dynamical transport problems, where interaction effects among mass particles are considered.

§1. Introduction

The increasing interest in Optimal Transport problems in the last years is undoubtedly due, among others, to the fact that they are suitable for applications to a wide range of different areas of Mathematics. Nowadays, more than 200 years after Monge first formulated such a kind of problem, an Optimal Transport problem is usually set as follows: it is given an ambient space X (which could be $\mathbb{R}^N$ as well as a general metric space) and a cost function $c : X \times X \to \mathbb{R} \cup \{+\infty\}$, so that $c(x,y)$ stands for the cost to move a mass particle from x to y. Then, given two positive mass distributions ρ_0 and ρ_1 in X (i.e. positive measures on X having the same mass), one is asked to minimize the total cost functional

$$\gamma \mapsto \int_{X\times X} c(x,y)\, d\gamma(x,y), \tag{1.1}$$

over the set of all *plans* $\Pi(\rho_0,\rho_1)$ which transport ρ_0 on ρ_1 (see next section for the definition): roughly speaking, the term $d\gamma(x,y)$ has to be thought as the quantity of mass located at x which is sent to y. We refer to this problem as *Monge-Kantorovich problem* with cost c.

A powerful tool in Optimal Transport is that of having at our disposal *equivalent formulations*: for example, as it is well-known the problem of

---

Key words and phrases: Optimal Transport; continuity equation; relativistic heat equation; Sobolev inequalities; branched transport.

minimizing (1.1) is equivalent to the following dual problem, involving a maximization in place of a minimization (here we are a little bit imprecise about this duality)

$$\sup \left\{ \int_X u(x)\, d\rho_0(x) + \int_X v(y)\, d\rho_1(y) \,:\, u(x) + v(y) \le c(x,y) \right\}.$$

The starting point of our presentation is one of these equivalent formulations, usually known under the name of *Benamou–Brenier formula.* When $X = \mathbb{R}^N$ and the cost c is given by $c(x,y) = |x-y|^2$, Benamou and Brenier in [5] discovered that for every ρ_0, ρ_1 the corresponding optimal transport problem

$$w_2(\rho_0,\rho_1)^2 := \min \left\{ \int_{\mathbb{R}^N \times \mathbb{R}^N} |x-y|^2\, d\gamma(x,y) \,:\, \gamma \in \Pi(\rho_0,\rho_1) \right\},$$

is equivalent to the following one

$$\min \int_0^1 \int_{\mathbb{R}^N} |v_t(x)|^2 \; d\mu_t(x) \; dt,$$

where the minimum in this case is taken among all pairs (μ_t, v_t), with μ_t curve of measures and v_t time-dependent vector field such that they solve the *continuity equation*

$$\partial_t \mu_t + \mathrm{div}_x(v_t\, \mu_t) = 0,$$

and such that $\mu_i = \rho_i$, $i = 0, 1$.

We will give more details on this equivalence in the next sections: for the moment, observe that such a result could be a little bit surprising at a first glance. Indeed, the first problem is *static*, that is a time-evolutionary description of the transportation is absent and the whole process only depends on the optimal coupling γ: once you know this, then it is implicitly understood that you move mass located at x to y along the segment joining these points. On the contrary, the second problem is *dynamical* and in principle many different trajectories of transportation, driven by the velocity fields v_t, are admissible. the continuity equation expresses the fact that the initial and final measures have the same total mass and no fractions of mass should disappear during the transportation. In particular, the second problem can be regarded as an *Eulerian* point of view on Optimal Transport: we just remark that one more equivalent dynamical formulation of the same problem, but now *Lagrangian* in spirit, is given by

$$\min \left\{ \int_{\mathrm{Lip}([0,1];\mathbb{R}^N)} \int_0^1 |\sigma'(t)|^2 \, dt \, dQ(\sigma) \right\},$$

where the minimum is taken among all probability measures Q over

$$\mathrm{Lip}([0,1];\mathbb{R}^N)$$

such that

$$\int \varphi(\sigma(i)) \, dQ(\sigma) = \int \varphi(x) \, d\rho_i(x), \qquad \text{for every } \varphi \in C(\mathbb{R}^N), \quad i = 0, 1.$$

Here $\mathrm{Lip}([0,1];\mathbb{R}^N)$ stands for the set of Lipschitz curves in $\mathbb{R}^N$, parametrized over the interval $[0,1]$.

With the Benamou–Brenier formula in mind, the aim of this survey is that of reviewing some dynamical transport models appeared in the last years, whose common root is that they can all be formulated as the minimization of an *action functional* under the constraint of the conservation of mass, the latter being expressed through the continuity equation. The kind of problems we want to address are thus of the following type: for every given ρ_0, ρ_1 positive measures on $\mathbb{R}^N$ with the same mass (that we can think to be 1, just for simplicity), we set

$$\mathcal{T}(\rho_0, \rho_1) := \min \left\{ \int_0^1 \mathcal{A}(\mu_t, v_t) \, dt \ : \ \begin{array}{c} \partial_t \mu_t + \mathrm{div}(v_t \mu_t) = 0, \\ \mu_i = \rho_i, \ i = 0, 1 \end{array} \right\}, \tag{1.2}$$

where the velocity vector field v_t has to be thought as a sort of tangent vector to the curve μ_t. Assuming that $\mathcal{A}(s, \cdot)$ is $p-$homogeneous ($p \geq 1$), we then refer to the quantity $\mathcal{T}^{1/p}$ as a *dynamical transport distance*. The case considered by Benamou and Brenier clearly fits in this framework, with the choice

$$\mathcal{A}(\mu_t, v_t) = \int_{\mathbb{R}^N} |v_t(x)|^2 \, d\mu_t(x) \quad (\textit{kinetic energy}).$$

Among these problems, we will also present some recent models introduced by Dolbeault, Nazaret and Savaré in [21], still of the kind (1.2), but with the conservation of mass now being expressed through a *nonlinear mobility continuity equation*, i.e. an equation of the type

$$\partial_t \mu_t + \operatorname{div}_x (v_t\, \theta(\mu_t)) = 0,$$

with θ *mobility function*: the previous equation has to be considered as non dimensionalized. Particularly interesting cases are

$$\theta(s) = s(1-s)_+ \quad \text{or} \quad \theta(s) = s^\beta,$$

with $(\,\cdot\,)_+$ standing for the positive part and $\beta \in (0,1)$. As we will see, these problems will be considerably different from that corresponding to the case of Benamou and Brenier, since in this case geodesics for $\mathcal{T}^{1/p}$ are no more strictly related to geodesics of $\mathbb{R}^N$ and actually an equivalent static formulation of these problems is not known.

Recently, in the paper [13], the same kind of variational model (i.e. minimization of an action under the constraint of the continuity equation) has been proposed in the context of *branched transport*. With this name we refer to optimal transport problems where the infinitesimal cost to move a mass m for a length ℓ is of the type $\varphi(m)\,\ell$, with $\varphi : [0,\infty) \to [0,\infty)$ being *increasing* and *subadditive*. The archetypical choice is $\varphi(m) = m^\alpha$, with $\alpha \in (0,1)$: due to concavity, in order to decrease the total cost, which would be a quantity of the type

$$\sum m^\alpha \ell,$$

it is better to gather the mass as much as possible during the transport. This clearly gives rise to tree-shaped optimal structures of transportation: root systems in a tree and blood vessels in a human body can be seen as concrete applications of this energy-saving principle.

Observe that the extremal choices $\alpha = 0$ and $\alpha = 1$ would correspond to the *Steiner minimal connection problem*[1] and to the usual *Monge problem* (i.e. minimize (1.1) with $c(x,y) = |x-y|$), respectively.

When equipped with a dynamical transport distance $\mathcal{T}^{1/p}$, a space of measures inherits a sort of Riemannian (or Finslerian) manifold structure:

[1]This is a little bit imprecise, since strictly speaking the case $\alpha = 0$ corresponds to finding a set of minimal length connecting the supports of the measures, with the requirement that this Steiner connection has to transport the first measure on the second. This clearly introduces a further constraint on the Steiner connection, namely on its orientation, which is absent from the original formulation. We thank Emanuele Paolini for this observation.

in the case corresponding to w_2, this point of view has been first explored (independently by Benamou and Brenier) by Otto in [36]. However, in these notes we will avoid to explicitely refer to the so-called *Otto's calculus* and we will simply present some examples of how the formal differential calculus based on the dynamical formulation can be exploited to highlight deep connections between Optimal Transport and functional analysis issues, like the characterization of evolution equations as *gradient flows* in spaces of measures ([2, 25, 36, 37]) and the derivation of sharp functional inequalities ([1, 18, 30, 31, 35, 38]).

The whole presentation in this survey will be quite a sketchy one, but as self-contained as possible. We have tried to focus on ideas rather than on rigourous proofs: for these, we provide quite a comprehensive bibliography and precise references where needed. In any case, besides the original research papers, the reader should always refer to the books [2,41]. Moreover, many generalizations are possible for the results and the techniques that we will present: for the sake of simplicity and readability, we have decided to avoid many of these.

The outline of the paper is the following: after recalling some basic facts about Optimal Transport and Wasserstein spaces (Section 2), in Section 3 we come to illustrate the result of Benamou and Brenier and its generalitazion due to Ambrosio, Gigli and Savaré. Particular stress is posed on the fact that this can be equivalently regarded as a convex optimization problem under linear constraints. In the same section, we also show how this formulation can be employed to derive evolution equations as gradient flows of suitable energy functionals and to establish sharp functional inequalities. In Section 4 we describe some variants of the Benamou-Brenier formulation: particularly interesting is the relativistic model already introduced by Brenier himself in [11]. Section 5 is devoted to present the models considered by Dolbeault, Nazaret and Savaré, based on the continuity equation $\partial_t \mu_t + \mathrm{div}_x(v_t\, \theta(\mu_t)) = 0$. Finally, we give in Section 6 a brief account of the Eulerian dynamical formulation for branched transport, as introduced in [13].

§2. Notations and preliminaries on Optimal Transport and Wasserstein spaces

In this section, we recall some well-known facts on Wasserstein spaces and their geometry that we will use throughout the paper. With $\mathscr{L}^k$ and

$\mathscr{H}^k$ we will always denote the k-dimensional Lebesgue and Hausdorff measures respectively, while the notation $\mathscr{P}(X)$ will indicate the space of Borel probability measures over a given metric space X. Also, given a positive Borel measure ρ and a Borel measurable map T, we will denote by $(T)_{\#}\rho$ the *push-forward of ρ through T*, i.e. the measure defined by

$$(T)_{\#}\rho(A) := \int_{T^{-1}(A)} d\rho, \qquad \text{for every Borel set } A.$$

All the materials presented in this section are nowadays standard and can be found in [2,41].

2.1. Wasserstein distances. Let (X,d) be a metric space. Given ρ_0, $\rho_1 \in \mathscr{P}(X)$, we set $\Pi(\rho_0,\rho_1)$ for the collection of all *transport plans* between ρ_0 and ρ_1, i.e. $\gamma \in \Pi(\rho_0,\rho_1)$ if $\gamma \in \mathscr{P}(X \times X)$ and

$$(\pi_x)_{\#}\gamma = \rho_0 \qquad \text{and} \qquad (\pi_y)_{\#}\gamma = \rho_1,$$

with $\pi_x(x,y) = x$ and $\pi_y(x,y) = y$. For every $p \geq 1$, we define the following optimal transport problem

$$\mathcal{C}_p(\rho_0,\rho_1) = \min \left\{ \int_{X \times X} d(x,y)^p \, d\gamma(x,y) \ : \ \gamma \in \Pi(\rho_0,\rho_1) \right\},$$

then $w_p := \mathcal{C}_p^{1/p}$ is a distance on the space of probability measures ρ having finite $p-$th momentum, i.e. such that

$$\int_X d(x,x_0)^p \, d\rho(x) < +\infty,$$

for a certain (and thus any) $x_0 \in X$. We indicate with $\mathcal{W}_p(X)$ the space of these measures endowed with the distance w_p: this is called *Wasserstein space of order p* or simply *$p-$Wasserstein space*. In the following, we will always take $X = \mathbb{R}^N$ or $X = \Omega$ compact convex subset of $\mathbb{R}^N$, equipped with $d(x,y) = |x-y|$ the standard Euclidean distance. The topology induced by the distance w_p is in general stronger than that induced by the *narrow convergence*, defined by duality with continuous and bounded functions. Indeed, it is equivalent to the narrow convergence plus convergence of the $p-$th moments: when we work on a compact set Ω, the two topologies coincide and $\mathcal{W}_p(\Omega) = \mathscr{P}(\Omega)$. Moreover, the distance w_p is lower semicontinuous w.r.t. the narrow convergence, i.e.

$$w_p(\rho_0, \rho_1) \leq \liminf_{n\to\infty} w_p(\rho_0^n, \rho_1^n), \tag{2.1}$$

for every sequence $\{\rho_i^n\}_{n\in\mathbb{N}}$ narrowly converging to ρ_i, $i = 0, 1$.

2.2. Geodesics. Let $I \subset \mathbb{R}$ be a compact interval, then a curve $\mu : I \to X$ is said to be *absolutely continuous with finite p-energy* if there exists $\psi \in L^p(I)$ such that

$$d(\mu_t, \mu_s) \leq \int_s^t \psi(r)\, dr, \text{ for every } s < t.$$

The minimal ψ for which the previous holds coincides with the *metric derivative* of μ w.r.t. the distance d, given by

$$|\mu_t'|_d = \lim_{h\to 0} \frac{d(\mu_{t+h}, \mu_t)}{h}.$$

The set of these curves is denoted with $AC^p(I; X)$: observe that $AC^\infty(I; X)$ coincides with the space of Lipschitz curves.

It is not difficult to see that $\mathcal{W}_p(\mathbb{R}^N)$ is a geodesic space[2], that is

$$w_p(\rho_0, \rho_1) = \min\left\{ \int_0^1 |\mu_t'|_{w_p}\, dt \ :\ \mu \in AC([0,1]; \mathcal{W}_p(\mathbb{R}^N)), \mu_i = \rho_i, i = 0, 1 \right\}.$$

Equivalently, one can consider the minimization of the L^p norm of the metric derivative: while the minimum value is still $w_p(\rho_0, \rho_1)$, now we are selecting a precise minimizer, given by the constant speed geodesic, i.e. a curve μ_t connecting ρ_0 to ρ_1 and such that

$$w_p(\mu_s, \mu_t) = |s - t|\, w_p(\rho_0, \rho_1) \text{ for every } s, t \in [0, 1].$$

We have an explicit formula for these curves in $\mathcal{W}_p(\mathbb{R}^N)$: for every $\rho_0, \rho_1 \in \mathcal{W}_p(\mathbb{R}^N)$, the unique constant speed geodesic connecting them is given by

$$\mu_t = ((1 - t)\pi_x + t\pi_y)_\# \gamma, \tag{2.2}$$

with $\gamma \in \Pi(\rho_0, \rho_1)$ optimal transport plan. When γ is given by a *transport map* T, i.e. $\gamma = (\mathrm{Id} \times T)_\# \rho_0$, the previous can be rewritten as

$$\mu_t = ((1 - t)\mathrm{Id} + tT)_\# \rho_0, \tag{2.3}$$

[2]More generally $\mathcal{W}_p(X)$ is a geodesic space each time the same holds true for X (see [27, Proposition 1 and Theorem 6]

a formula which has been first introduced by McCann in [32] under the name of *displacement interpolation*.

2.3. Convex energy functionals on Wasserstein spaces. The geometry induced on $\mathscr{P}(\mathbb{R}^N)$ by the Wasserstein distances implies that a new kind of convexity has to be taken into account: since $\mathcal{W}_p(\mathbb{R}^N)$ is a geodesic space, the natural notion of convexity is that of convexity along geodesics. We say that a functional $\mathfrak{F}$ is *displacement convex* if for every $\rho_0, \rho_1 \in \mathcal{W}_p(\mathbb{R}^N)$ there exists a constant speed geodesic μ_t such that

$$\mathfrak{F}(\mu_t) \leq (1-t)\,\mathfrak{F}(\rho_0) + t\,\mathfrak{F}(\rho_1), \quad \text{for every } t \in [0,1].$$

More generally $\mathfrak{F}$ is said to be $\Lambda-$displacement convex if

$$\mathfrak{F}(\mu_t) + \Lambda \frac{t(1-t)}{2} w_2(\rho_0,\rho_1)^2 \leq (1-t)\,\mathfrak{F}(\rho_0) + t\,\mathfrak{F}(\rho_1),$$

which, roughly speaking, corresponds to say that the second derivatives of $\mathfrak{F}$ are bounded below by Λ.

We have the following result, first proven in [32]: it characterizes an important class of displacement convex functionals on $\mathcal{W}_p(\mathbb{R}^N)$, the so called *internal energy functionals* (see [2, Proposition 9.3.9] for the proof).

Theorem 2.1. *Let $U : [0,+\infty) \to \mathbb{R}$ be a proper and lower semicontinuous convex functions, verifying $U(0) = 0$. We define the functional $\mathcal{U} : \mathcal{W}_p(\mathbb{R}^N) \to \mathbb{R} \cup \{+\infty\}$ by*

$$\mathcal{U}(\rho) = \begin{cases} \int_{\mathbb{R}^N} U(f(x))\,dx, & \textit{if } \rho = f \cdot \mathscr{L}^N, \\ +\infty, & \textit{otherwise} \end{cases}$$

and we set $\mathcal{U}^*$ for its lower semicontinuous envelope, i.e.

$$\mathcal{U}^*(\rho) = \inf\left\{\liminf_{n\to\infty} \mathcal{U}(\rho^n) \,:\, \rho^n \to \rho \ \textit{in } \mathcal{W}_p(\mathbb{R}^N)\right\}.$$

If U satisfies the following condition

$$\lambda \mapsto \lambda^N U(\lambda^{-N}) \quad \textit{is convex and non increasing,} \tag{2.4}$$

then $\mathcal{U}$ and $\mathcal{U}^*$ are displacement convex.

Important examples of functions U satisfying (2.4) are the following:

(i) $U(s) = s \log s$ and the corresponding functional (the *Boltzmann entropy functional*) is given by

$$\mathcal{U}(\rho) = \int_{\mathbb{R}^N} f(x) \log f(x)\,dx, \qquad \text{if } \rho = f \cdot \mathscr{L}^N, \tag{2.5}$$

set to be $+\infty$ on measures ρ such that $\rho \not\ll \mathscr{L}^N$. Due to the superlinearity of the function $s \mapsto s\log s$, in this case we have that $\mathcal{U}^* = \mathcal{U}$ (see [15]);

(ii) $U(s) = s^\vartheta/(\vartheta - 1)$, for $\vartheta \in [1-1/N, 1) \cup (1, \infty)$. In particular, in the superlinear case, we have that $\mathcal{U} = \mathcal{U}^*$, while in the sublinear case, i.e. when $\vartheta < 1$, we have that $\mathcal{U} \neq \mathcal{U}^*$ and

$$\mathcal{U}^*(\rho) = \frac{1}{\vartheta - 1}\int_{\mathbb{R}^N} f(x)^\vartheta \, dx, \qquad \text{if } \rho = f \cdot \mathscr{L}^N + \rho^\perp,$$

so that the $\mathcal{U}^*$ can assume finite values also on measures having a singular part w.r.t. the Lebesgue measure $\mathscr{L}^N$.

Remark 2.2. If U is smooth, one can easily see that condition (2.4) is equivalent to require that

$$P'(s)s \geq \left(1 - \frac{1}{N}\right) P(s),$$

where $P(s) := \int_0^s \lambda \, U''(\lambda)\, d\lambda = s\, U'(s) - U(s)$.

§3. A Fluid Mechanics reformulation of Wasserstein distances

3.1. ***AC*** **curves in Wasserstein spaces and the continuity equation.** We start with the following crucial result, giving an equivalent characterization of the 2-Wasserstein distance in terms of solutions of the continuity equation.

Benamou-Brenier formula ([5]). *Given $\rho_0, \rho_1 \in \mathscr{P}(\mathbb{R}^N)$ having smooth densities w.r.t. to $\mathscr{L}^N$ and bounded supports, let us set*

$$A(\rho_0, \rho_1) = \{(\mu, v) \,:\, \partial_t \mu_t + \mathrm{div}_x(v_t \mu_t) = 0 \text{ in } I \times \mathbb{R}^N,\ \mu_0 = \rho_0,\ \mu_1 = \rho_1\}.$$

Then the 2-Wasserstein distance between ρ_0 and ρ_1 can be characterized as follows:

$$w_2(\rho_0, \rho_1)^2 = \min_{(\mu, v) \in A(\rho_0, \rho_1)} \int_0^1 \int_{\mathbb{R}^N} |v_t(x)|^2 \, d\mu_t(x)\, dt. \tag{3.1}$$

The previous result has been considerably extended to cover much more general situations: in this sense, we cite the following fundamental result (see [2, Theorem 8.3.1]), which at the same time generalizes the Benamou-Brenier formula and gives a complete characterization of absolutely continuous curves in Wasserstein spaces. continuity equation. In what follows, we take $I = [0, 1]$ for simplicity.

Theorem 3.1 (Ambrosio-Gigli-Savaré). *Let us fix an exponent* $p \in (1, \infty)$*. Let* $\mu : I \to \mathcal{W}_p(\mathbb{R}^N)$ *be a narrowly continuous curve satisfying the continuity equation*

$$\partial_t \mu_t + \mathrm{div}\,(v_t \mu_t) = 0, \qquad in\ \mathbb{R}^N \times I,$$

in the sense of distributions, for some Borel vector field $v : I \times \mathbb{R}^N \to \mathbb{R}^N$ *such that* $\|v_t\|_{L^p(\mathbb{R}^N,\mu_t)}$ *is integrable in time. Then* $\mu_t \in AC(I; \mathcal{W}_p(\mathbb{R}^N))$ *and there holds*

$$|\mu_t'|_{w_p} \leq \|v_t\|_{L^p(\mathbb{R}^N,\mu_t)} \quad for\ \mathscr{L}^1\!-a.e.\ t \in I.$$

On the other hand, for every $\mu \in AC^p(I; \mathcal{W}_p(\mathbb{R}^N))$*, there exists a Borel vector field* $v : I \times \mathbb{R}^N \to \mathbb{R}^N$ *such that*

$$\|v_t\|_{L^p(\mathbb{R}^N,\,\mu_t)} \leq |\mu_t'|_{w_p}, \quad for\ \mathscr{L}^1\!-a.e.\ t \in I,$$

and the continuity equation holds in the sense of distributions.

Proof. Let us start from the first fact. One observes that, if regularity issues are disregarded, the curve μ_t would be of the form $(X_t)_\sharp \mu_0$, with X_t flow map of v_t, i.e.

$$\begin{cases} X_t'(x) &= v_t(X_t(x)) \\ X_0(x) &= x \end{cases}$$

This is the well-known *method of characteristics* for the continuity equation (see [2, Chapter 8]). Then, fixing $s, t \in [0, 1]$, we can estimate the Wasserstein distance between μ_s and μ_t using the transport plan $\gamma_{s,t} = (X_s \times X_t)_\# \mu_0$, i.e.

$$w_p(\mu_s, \mu_t)^p \leq \int_{\mathbb{R}^N} |X_s(x) - X_t(x)|^p \, d\mu_0(x).$$

We then observe the following

$$X_t(x) - X_s(x) = \int_s^t X_r'(x)\, dr = \int_s^t v_r(X_r(x))\, dr,$$

so that using Cauchy-Schwarz and Jensen inequalities, one ends up with

$$w_p(\mu_t,\mu_s)^p \le |t-s|^{p-1}\int_{\mathbb{R}^N}\int_t^s |v_r(X_r(s))|^p\,dr\,d\mu_0(x).$$

It is sufficient to exchange the order of integration and use the definition of push-forward measure, so that the previous can be recast into

$$\frac{w_p(\mu_t,\mu_s)^p}{|t-s|^p} \le \frac{1}{|s-t|}\int_s^t\int_{\mathbb{R}^N} |v_r(x)|^p\,d\mu_r(x)\,dr, \tag{3.2}$$

which gives the desired estimate on the metric derivative of μ_t, taking the limit as s tends to t.

So far, this was the heuristic argument: to establish rigorously the result, one has to simply go on through a smoothing argument, considering $\mu_t^\varepsilon := \mu_t * \varrho_\varepsilon$ and $\phi_t^\varepsilon := (v_t\,\mu_t) * \varrho_\varepsilon$, with ϱ_ε smooth convolution kernel supported on the whole $\mathbb{R}^N$. Then one can see that μ_t^ε solves the continuity equation with the smooth velocity field v_t^ε implicitly defined by $\phi_t^\varepsilon = v_t^\varepsilon\cdot\mu_t^\varepsilon$: in this way, $\mu_t^\varepsilon = (X_t^\varepsilon)_\#\mu_0^\varepsilon$, with X_t^ε flow map of v_t^ε and the calculations above are justified. We then rewrite the right-hand side in (3.2) as follows

$$\int_s^t\int_{\mathbb{R}^N} |v_r^\varepsilon(x)|^p\,d\mu^\varepsilon(x)\,dr = \int_s^t\int_{\mathbb{R}^N}\left|\frac{\phi_r^\varepsilon(x)}{\mu_r^\varepsilon(x)}\right|^p\,\mu_r^\varepsilon(x)\,dx\,dr.$$

Observing that $(\mu,\phi)\mapsto|\phi|^p\mu^{1-p}$ is jointly convex and $1-$homogeneous (see the next subsection), by means of Jensen inequality we have that

$$\int_{\mathbb{R}^N}\left|\frac{\phi_r^\varepsilon(x)}{\mu_r^\varepsilon(x)}\right|^p\,\mu_r^\varepsilon(x)\,dx \le \int_{\mathbb{R}^N}\left|\frac{d\phi_r}{d\mu_r}(x)\right|^p\,d\mu_r(x),$$

which then enables to conclude

$$\begin{aligned}\frac{w_p(\mu_s^\varepsilon,\mu_t^\varepsilon)^p}{|t-s|^p} &\le \frac{1}{|t-s|}\int_{[s,t]\times\mathbb{R}^N}\left|\frac{d\phi_r}{d\mu_r}(x)\right|^p\,d\mu_r(x)\,dr\\ &= \frac{1}{|t-s|}\int_s^t\int_{\mathbb{R}^N}|v_r(x)|^p\,d\mu_r(x)\,dr.\end{aligned}$$

Passing to the limit as ε goes to 0 and using (2.1), one obtains again (3.2).

The first part of the statement is a little bit more involved: here we just sketch the main idea of the constructive argument which can be found in [27] and which differs from the argument used in [2]. According to [27, Theorem 5], given a curve $\mu \in AC^p(I;\mathcal{W}_p(\mathbb{R}^N))$, one can always find a probability measure $Q \in \mathscr{P}(AC^p(I;\mathbb{R}^N))$ such that $\mu_t = (e_t)_\# Q$, where e_t is the *evaluation at time t map*, i.e. $e_t(\sigma) = \sigma(t)$ for every continuous curve σ. Intuitively, this means that one can always realize an AC curve in the Wasserstein space $\mathcal{W}_p$ as a *superposition* of AC curves of the base space $\mathbb{R}^N$. More important, these curves can be taken in such a way that we can control their $p-$average velocities in terms of the w_p-metric derivative of μ, i.e. Q can be constructed so to satisfy

$$\left(\int_{AC^p(I;\mathbb{R}^N)} |\sigma'(t)|^p \, dQ(\sigma)\right)^{1/p} \leq |\mu_t'|_{w_p}, \; t \in I. \tag{3.3}$$

The proof of this fact is lengthy, but the underlying idea is very simple: one starts considering a partition $0 = t_0 < t_1 < \cdots < t_{k+1} = 1$ of the time interval $[0,1]$. Then we discretize the curve μ_t by considering the measures $\mu_{t_0}, \dots, \mu_{t_{k+1}}$ and at each time step t_i we interpolate between μ_{t_i} and $\mu_{t_{i+1}}$ using an optimal plan $\gamma_{i,i+1}$, for the cost $c(x,y) = |x-y|^p$. In this way, we can construct a measure concentrated on linear curves (transport rays), parametrized over $[t_i, t_{i+1}]$. Gluing together this measures, gives a $Q_k \in \mathscr{P}(AC^p(I;\mathbb{R}^N))$, concentrated on piecewise linear curves satisfying (3.3). Then one can show $\{Q_k\}_{k\in\mathbb{N}}$ to be equi-tight: choosing the partitions $\{t_0, \dots, t_{k+1}\}$ to be dyadic and taking the limit as k goes to ∞, one can conclude.

Once we made this construction, we can consider the disintegration of Q with respect to the map e_t, obtaining $Q = \int Q_x^t \, d\mu_t$, where Q_x^t is a Borel probability measure concentrated on the fiber $e_t^{-1}(\{x\}) = \{\sigma : \sigma(t) = x\}$. It is then natural to construct the desired vector field v_t as the average velocity of the curves corresponding to Q, that is

$$v_t(x) := \int_{\{\sigma \,:\, \sigma(t)=x\}} \sigma'(t) \, dQ(\sigma).$$

It is easy to verify that (μ_t, v_t) solves the continuity equation; moreover, thanks to estimate (3.3), we have that $v_t \in L^p(\mu_t)$ and $\|v_t\|_{L^p(\mu_t)} \leq |\mu_t'|_{w_p}$, thus concluding the proof. □

The Benamou–Brenier formula in its general form (i.e. without smoothness assumptions on ρ_0, ρ_1) is then a simple consequence of the previous result.

Corollary 3.2. *For every $\rho_0, \rho_1 \in \mathcal{W}_p(\mathbb{R}^N)$, there holds*

$$w_p(\rho_0, \rho_1)^p = \min \left\{ \int_0^1 \int_{\mathbb{R}^N} |v_t(x)|^p \, d\mu_t(x) \, dt \ : \ \begin{array}{c} \partial_t \mu_t + \mathrm{div}_x(v_t \mu_t) = 0, \\ \mu_0 = \rho_0, \, \mu_1 = \rho_1 \end{array} \right\}. \tag{3.4}$$

Proof. Let us take a curve $\mu \in AC^p(I; \mathcal{W}_p(\mathbb{R}^N))$ connecting ρ_0 to ρ_1 and a velocity vector field v_t such that $\|v_t\|_{L^p(\mu_t)} = |\mu_t'|_{w_p}$, then

$$w_p(\rho_0, \rho_1)^p \leq \int_0^1 |\mu_t'|_{w_p}^p \, dt = \int_0^1 \int_{\mathbb{R}^N} |v_t(x)|^p \, d\mu_t(x) \, dt,$$

so that the minimum value of the action is greater than or equal to $w_p(\rho_0, \rho_1)^p$. Then taking μ_t to be a constant speed geodesic in $\mathcal{W}_p(\mathbb{R}^N)$ we obtain

$$w_p(\rho_0, \rho_1)^p = \int_0^1 |\mu_t'|_{w_p}^p \, dt = \int_0^1 \int_{\mathbb{R}^N} |v_t(x)|^p \, d\mu_t(x) \, dt,$$

thus concluding the proof. □

Remark 3.3. When $p = \infty$, we can completely characterize Lipschitz curves (i.e. AC^∞ curves) in the Wasserstein space $\mathcal{W}_\infty(\mathbb{R}^N)$ in terms of L^∞ vector fields, with $\|v_t\|_{L^\infty(\mu_t)}$ integrable in time and (μ_t, v_t) solution of the continuity equation. The result can be achieved by means of an easy limit argument. On the contrary, in the case $p = 1$, the second part of Theorem 3.1 fails to be true. This is due to the presence of the *teleport* phenomenon, typical of the non-strictly convex cost $|x - y|$: for example, taken the curve

$$\mu_t = (1 - t)\delta_{x_0} + t\delta_{x_1},$$

this is absolutely continuous in $\mathcal{W}_1(\mathbb{R}^N)$, but there can not exist a v_t such (μ_t, v_t) solve the continuity equation and $\|v_t\|_{L^1(\mu_t)} \leq |\mu_t'|_{w_1}$. Observe that μ_t solves the continuity equation

$$\partial_t \mu_t + \mathrm{div}_x \phi_t = 0,$$

with the vector measure $\phi_t = \overrightarrow{\tau} \cdot \mathscr{H}^1 \llcorner \overline{x_1 x_0}$ for every t and $\phi_t \not\ll \mu_t$, the set $\overline{x_1 x_0}$ being the segment joining x_0 to x_1, oriented according to $\overrightarrow{\tau} = (x_1 - x_0)|x_1 - x_0|^{-1}$.

We can substitute $\mathbb{R}^N$ with a convex bounded set $\Omega \subset \mathbb{R}^N$, provided that the continuity equation is interpreted with a homogeneous Neumann condition $\langle v, \eta_\Omega \rangle = 0$ at the boundary $\partial\Omega$, with η_Ω standing for the outer normal versor. This has to be intended in a weak sense, that is

$$\int_\Omega \varphi(1,x)\, d\rho_1(x) - \int_\Omega \varphi(0,x)\, d\rho_0(x) = \int_0^1 \int_\Omega \partial_t \varphi(t,x)\, d\mu_t(x)\, dt + \int_0^1 \int_\Omega \langle \nabla\varphi(t,x), v_t(x)\rangle\, d\mu_t(x)\, dt,$$

for every $\varphi \in C^1([0,1] \times \overline{\Omega})$. Note that from a physical point of view, the homogeneous Neumann boundary condition prevents the necessity of using boundary conditions for μ and let the flow of v stay inside Ω.

Remark 3.4. In the case of a non convex set $\Omega \subset \mathbb{R}^N$, we still obtain a dynamical characterization of the $p-$Wasserstein distance, but with the Euclidean distance replaced by the geodesic one in Ω. For this reason, in what follows we will mainly confines ourselves to work with Ω either a convex subset of $\mathbb{R}^N$ or the whole space.

3.2. Convex optimization reformulation. As pointed out by the proof of Theorem 3.1, the problem (3.4) can be reformulated as a convex optimization problem under linear constraints, simply introducing the *flux* variable $\phi_t = v_t \cdot \mu_t$. If we do so, then the continuity equation now simply rewrites as a linear equation in the variables (μ, ϕ), that is

$$\partial_t \mu_t + \mathrm{div}_x \phi_t = 0.$$

Moreover thanks to the Disintegration Theorem (see [19, Chapter III]), we can identify the curves of measures $t \mapsto \mu_t$ and $t \mapsto \phi_t$ with the measures on $[0,1] \times \mathbb{R}^N$ given by

$$\mu = \int \mu_t\, dt \quad \text{and} \quad \phi = \int \phi_t\, dt. \tag{3.5}$$

In this way, it is then natural to enlarge the class of admissible pairs to (μ, ϕ) measures on $[0,1] \times \mathbb{R}^N$, not necessarily of the form (3.5). For these pairs of measures, the continuity equation with a constraint on the endpoints has to be interpreted as follows

$$\partial_t \mu + \mathrm{div}_x \phi = \rho_0 \otimes \delta_0 - \rho_1 \otimes \delta_1,$$

still in distributional sense, i.e.

$$\begin{aligned}\int_{[0,1]\times\mathbb{R}^N} \partial_t \varphi(t,x)\, d\mu(t,x) + \int_{[0,1]\times\mathbb{R}^N} \nabla_x \varphi(t,x) \cdot d\phi(t,x) \\ = \int_{\mathbb{R}^N} \varphi(1,x)\, d\rho_1(x) - \int_{\mathbb{R}^N} \varphi(0,x)\, d\rho_0(x),\end{aligned} \tag{3.6}$$

for every $\varphi \in C^1_0([0,1]\times\mathbb{R}^N)$. As for the energy functional, we have already observed that in the original formulation the function $|v_t|^p \mu_t = |\phi_t|^p \mu_t^{1-p}$ is jointly convex and $1-$homogeneous, in the variables (μ, ϕ): more precisely, we introduce

$$f_p(x,y) = \begin{cases} |y|^p\, x^{1-p}, & \text{if } x > 0,\, y \in \mathbb{R}^N, \\ 0, & \text{if } x = 0,\, y = 0, \\ +\infty, & \text{otherwise }, \end{cases}$$

which is also l.s.c. in addition to the properties recalled above. Then the corresponding functional defined on measures, which we call the *Benamou–Brenier functional*

$$\mathcal{F}_p(\mu,\phi) = \int_{[0,1]\times\mathbb{R}^N} f_p\left(\frac{d\mu}{dm}, \frac{d\phi}{dm}\right) dm, \tag{3.7}$$

is local, lower semicontinuous and, thanks to the $1-$homogeneity of f_p, does not depend on the choice of the reference measure m. Using this fact and the definition of f_p, the previous can be rephrased as

$$\mathcal{F}_p(\mu,\phi) = \begin{cases} \displaystyle\int_{[0,1]\times\mathbb{R}^N} \left|\frac{d\phi}{d\mu}(t,x)\right|^p d\mu(t,x), & \text{if } \phi \ll \mu, \\ +\infty, & \text{otherwise}, \end{cases}$$

and we consider the problem

$$\min\{\mathcal{F}_p(\mu,\phi) \,:\, \partial_t \mu + \mathrm{div}_x \phi = \rho_0 \otimes \delta_0 - \rho_1 \otimes \delta_1\}, \tag{3.8}$$

where the admissible pairs (μ, ϕ) are now general Radon measures on $[0,1] \times \mathbb{R}^N$. Observe that this is a convex optimization problem, under a linear constraint, with the functional $\mathcal{F}_p$ being both lower semicontinuous and coercive: (3.8) is equivalent to the original problem addressed by Benamou and Brenier.

The only non trivial fact is that in (3.8) we are allowing for *general Radon measures* (μ, ϕ) on the space-time, which are not necessarily curves of measures: however, this is the case for each admissible pair (μ, ϕ) having finite energy, thanks to the continuity equation.

Indeed, at first the latter implies that μ must be a probability measure, disintegrating as $\mu = \int \mu_t \, dt$, i.e. μ can be identified with a curve of probability measures.

To show this, we start disintegrating μ with respect to the time variable, then

$$\mu = \int \mu_t \, d\lambda,$$

with λ positive measure on $[0,1]$. Inserting smooth test functions ψ depending only on t into (3.6) and using $\int\limits_{\mathbb{R}^N} d\rho_0 = \int\limits_{\mathbb{R}^N} d\rho_1 = 1$, one obtains

$$\int\limits_0^1 \psi'(t)\, d\lambda(t) = \psi(1) - \psi(0), \text{ for every } \psi \in C^1([0,1]),$$

thus giving $\lambda = \mathscr{L}^1 \llcorner [0,1]$. As for the measure ϕ, one observes that by the very definition of the functional $\mathcal{F}_p$, we have

$$\mathcal{F}_p(\mu, \phi) < +\infty \Longrightarrow \phi \ll \mu.$$

Hence we also get the disintegration $\phi = \int \phi_t \, dt$, with $\phi_t = v_t \cdot \mu_t$ and and where the vector field v_t is the Radon-Nykodim derivative of ϕ_t w.r.t. μ_t. In conclusion, for each admissible pair (μ, ϕ) having finite energy, we obtain

$$\mathcal{F}_p(\mu, \phi) = \int\limits_0^1 \int\limits_{\mathbb{R}^N} \left| \frac{d\phi_t}{d\mu_t}(x) \right|^p d\mu_t(x)\, dt = \int\limits_0^1 \int\limits_{\mathbb{R}^N} |v_t(x)|^p \, d\mu_t(x)\, dt,$$

thus recovering the objective functional in the Benamou–Brenier formula.

3.3. Dual formulation. Once the problem is formulated as a convex optimization with linear constraints, it is then very natural to investigate its dual formulation. Introducing Lagrange multipliers, the previous optimization problem can be rephrased as a saddle-point problem, i.e.

$$\min_{(\mu,\phi)} \sup_{\varphi} \left\{ \int_0^1 \int \frac{1}{p} \frac{|\phi_t|^p}{\mu_t^{p-1}} - \int_0^1 \int (\partial_t \varphi \, d\mu_t + \nabla\varphi \cdot \phi_t) \right.$$
$$\left. + \int \varphi(1,x)\, d\rho_1(x) - \int \varphi(0,x)\, d\rho_0(x) \right\}.$$

We then exchange the inf and the sup and use the Legendre–Fenchel conjugate f_p^* of $1/p\, f_p$, which is the indicator function of a closed convex set, precisely we have

$$f_p^*(\xi_1, \xi_2) = \begin{cases} 0, & \text{if } \xi_1 + \frac{1}{q}|\xi_2|^q \le 0, \\ +\infty, & \text{otherwise}, \end{cases}$$

where $q = p/(p-1)$. In this way, we have the following equivalence

$$\frac{1}{p} w_p(\rho_0,\rho_1)^p = \min \left\{ \int_0^1 \int \frac{1}{p} \frac{|\phi|^p}{\mu^{p-1}} \,:\, \partial_t \mu_t + \mathrm{div}_x \phi = 0,\ \mu_i = \rho_i,\ i = 0,1 \right\}$$
$$= \sup_{\left\{\psi \,:\, \partial_t \psi + \frac{1}{q}|\nabla_x \psi|^q \le 0\right\}} \int \psi(1,x)\, d\rho_1(x) - \int \psi(0,x)\, d\rho_0(x).$$

Observe that the formal primal-dual optimality conditions are given by

$$\begin{cases} \partial_t \mu_t + \mathrm{div}_x(|\nabla_x \psi_t|^{q-2} \nabla_x \psi_t \cdot \mu_t) = 0, \\ \partial_t \psi_t + \dfrac{1}{q} |\nabla_x \psi_t|^q = 0, \end{cases} \tag{3.9}$$

which means that the velocity of an optimal curve μ (that is, a geodesic in $\mathcal{W}_p(\mathbb{R}^N)$) is given by $v_t = |\nabla_x \psi_t|^{q-2} \nabla_x \psi_t$, with the time-dependent potential ψ_t solving an Hamilton-Jacobi equation.

Remark 3.5. Clearly, the whole derivation of this duality is not rigorous: anyway, the exchange between the inf and the sup can be justified by standard convex duality arguments, while generally to properly give meaning to the primal-dual optimality conditions, the ambient space of the dual problem has to be suitably relaxed and the gradient operator has to be properly defined on this new space (see [16] for more details).

3.4. Evolution equations as gradient flows in Wasserstein spaces. The characterization of AC curves in $\mathcal{W}_p(\mathbb{R}^N)$ in terms of the continuity equation $\partial_t \mu_t + \mathrm{div}_x \phi_t = 0$, in addition to a physical meaning, gives a natural and porwerful calculus in the Wasserstein space. In order to illustrate the basic idea, in what follows for simplicity we confine ourselves to consider the case $p = 2$. We take an internal energy functional

$$\mathcal{U}(\rho) = \int U(f(x))\, dx, \quad \text{if } \rho = f \cdot \mathscr{L}^N,$$

then using formally the continuity equation we can estimate the rate of dissipation of $\mathcal{U}$ along a curve $\mu_t \in AC^2(I; \mathcal{W}_2(\mathbb{R}^N))$. Supposing $\mu_t = f_t \cdot \mathscr{L}^N$ and $\phi_t = v_t f_t \cdot \mathscr{L}^N$, we obtain

$$\begin{aligned}\frac{d}{dt}\mathcal{U}(\mu_t) = \frac{d}{dt}\int U(f_t(x))\, dx &= \int U'(f_t(x))\, \partial_t f_t(x)\, dx \\ &= \int \langle \nabla U'(f_t(x)), v_t(x)\rangle f_t(x)\, dx.\end{aligned}$$

Observe that the previous can be estimated from below by means of Cauchy–Schwarz and Young inequalities as follows

$$\frac{d}{dt}\mathcal{U}(\mu_t) \geq -\frac{1}{2}\int |\nabla U'(f_t(x))|^2 f_t(x)\, dx - \frac{1}{2}\int |v_t(x)|^2 f_t(x)\, dx.$$

Thanks to Theorem 3.1, we can choose the vector field v_t having minimal L^2 norm, i.e. such that $\|v_t\|_{L^2(\mu_t)} = |\mu_t'|_{w_2}$, thus the previous inequality rewrites as follows

$$\frac{d}{dt}\mathcal{U}(\mu_t) \geq -\frac{1}{2}\int |\nabla U'(f_t(x))|^2 f_t(x)\, dx - \frac{1}{2}|\mu_t'|^2_{w_2}. \tag{3.10}$$

Moreover, we can have equality in (3.10), i.e. we maximize the rate of dissipation of the energy $\mathcal{U}$ along the curve μ_t, if and only if we take

$$v_t(x) = -\nabla U'(f_t(x)). \tag{3.11}$$

This means that with this choice, $\mu_t = f_t \cdot \mathscr{L}^N$ is a curve of steepest descent for the functional $\mathcal{U}$ and using the fact that μ_t and $\phi_t = v_t \cdot \mu_t$ are linked through the continuity equation, we obtain that such a curve μ_t solves the following evolution equation

$$\partial_t f_t = \mathrm{div}_x(\nabla U'(f_t)\, f_t). \tag{3.12}$$

Observe that if one regards the v_t of minimal $L^2(\mu_t)$ norm as the tangent vector to μ_t and $\nabla U'(f_t)$ as the gradient of $\mathcal{U}$ with respect to the Wasserstein structure, then (3.11) is exactly a gradient flow equation in $\mathcal{W}_2(\mathbb{R}^N)$ (see also [36] for more details).

Remark 3.6. It is useful to keep in mind that in a Hilbert space X, given a smooth functional $\mathfrak{F} : X \to \mathbb{R}$, for every AC curve σ we have

$$\frac{d}{dt}\mathfrak{F}(\sigma(t)) \geq -\frac{1}{2}|\nabla\mathfrak{F}(\sigma(t))|^2 - \frac{1}{2}|\sigma'(t)|^2, \ t \in I,$$

thanks to Cauchy–Schwarz and Young inequalities. Moreover, equality holds for every t if and only if

$$\sigma'(t) = -\nabla\mathfrak{F}(\sigma(t)), \ t \in I,$$

i.e. the curve σ satisfies the gradient flow equation.

The first significative application of these formal computations is to the case of the Boltzmann entropy functional,

$$\mathcal{U}(\rho) = \begin{cases} \int f(x)\log f(x)\,dx, & \text{if } \rho = f \cdot \mathscr{L}^N, \\ +\infty, & \text{otherwise,} \end{cases}$$

then observing that

$$\nabla U'(f_t) = \frac{\nabla f_t}{f_t}$$

and inserting this into (3.12), we obtain that the gradient flow of $\mathcal{U}$ w.r.t. the 2-Wasserstein distance is given by the heat equation

$$\partial_t f = \Delta f,$$

with homogeneous Neumann conditions if we replace $\mathbb{R}^N$ with a bounded convex set Ω. It is very remarkable that once the metric structure corresponding to w_2 is chosen, solutions of the heat equation can be obtained as curves of steepest descent for the Boltzmann entropy. In the words of the authors of [25], "*this formulation allows us to attach a precise interpretation to the conventional notion that diffusion arises from the tendency of the system to maximize the entropy*". In the very same way, one can show that the gradient flow of an internal energy of the type

$$\mathcal{U}(\rho) = \frac{1}{\vartheta - 1}\int f(x)^\vartheta\,dx, \ \text{ if } \rho = f \cdot \mathscr{L}^N,$$

leads to the evolution equation

$$\partial_t f_t = \Delta f_t^\vartheta,$$

called *porous media equation* when $\vartheta > 1$ and *fast diffusion equation* in the case $\vartheta < 1$.

Remark 3.7. More generally, one could consider a functional containing also a *potential energy* and an *interaction* one, i.e.

$$\mathfrak{F}(\rho) = \int U(f(x))\, dx + \int V(x)\, d\rho(x) + \int W(x-y)\, d\rho(x)\, d\rho(y), \ \rho = f \cdot \mathscr{L}^N,$$

then obtaining that curves of w_2-steepest descent are solutions of the following evolution equation

$$\partial_t f_t = \mathrm{div}_x \Big(\nabla U'(f_t) f_t + \nabla V f_t + (f_t * \nabla W) f_t \Big).$$

The previous heuristic considerations have been put on solid grounds in the last years, especially with the work of Ambrosio, Gigli and Savarè, that in [2] have developed a theory of gradient flows in general metric spaces. In this setting, where no linear or differentiable structures are available, one can define various concepts of gradient flow (in general not equivalent): among them, we just mention the *minimizing movement scheme* and the *evolution variational inequalities (EVI)* approach.

The former consists in fixing a time step τ and defining recursively ρ_n as follows

$$\rho_{n+1} \in \mathrm{argmin} \left\{ \frac{1}{2} \frac{d(\rho, \rho_n)^2}{\tau} + \mathfrak{F}(\rho) \right\}, \tag{3.13}$$

starting from an initial datum ρ_0. If these minimiziation problems admits solutions, then one can interpolate this points, defining the piecewise constant curve

$$\mu_t^\tau = \sum_{n=0}^{\infty} 1_{(n\tau, (n+1)\tau]}(t) \rho_n.$$

If one can show the convergence of this interpolating curves to a same limit curve μ, then μ can be taken to be the gradient flow, starting from ρ_0, of the functional $\mathfrak{F}$ with respect to the metric d. Observe that in a Hilbertian setting, $d(x,y)^2 = \langle x - y, x - y \rangle$ and a minimizer of (3.13) thus satisfies

$$\frac{\rho_{n+1} - \rho_n}{\tau} = -\nabla \mathfrak{F}(\rho_{n+1}),$$

which can be seen as a natural time discretization of the gradient flow equation.

In the (EVI) approach, which gives the strongest concept of gradient flow and requires the functional to be convex along geodesics, an absolutely continuous curve μ can be regarded as a curve of steepest descent for a functional $\mathfrak{F}$ if it satisfies the following differential inequality

$$\frac{1}{2}\frac{d}{dt}d^2(\mu_t,\rho_0) \le \mathfrak{F}(\mu_t) - \mathfrak{F}(\rho_0), \text{ for every } t \text{ and for every } \rho_0. \tag{3.14}$$

Observe that (3.14), roughly speaking, means that the time derivative of μ_t belongs to $-\partial\,\mathfrak{F}(\mu_t)$, i.e. minus the subdifferential of $\mathfrak{F}$: once again, in the case of a Hilbert space, this is exactly the meaning of (3.14).

Remark 3.8. The connection between diffusion equations and curves of steepest descent with respect to the Wasserstein structure has been probably first identified by Otto in his paper [37]. There, the author shows how solutions to the 2−dimensional porous media equation

$$\partial_t f_t = \Delta f_t^2,$$

can be obtained through a minimizing movement scheme, starting from the functional

$$\mathfrak{F}(\rho) = \int f(x)^2\,dx, \quad \text{if } \rho = f\cdot\mathscr{L}^N,$$

and taking as metric the 2−Wasserstein distance: observe that this functional is displacement convex, according to Theorem 2.1. This approach has then been exploited by Jordan, Kinderlehrer and Otto himself in [25], to give the first rigorous justification of the fact that the linear Fokker-Planck equation

$$\partial_t f_t = \Delta f_t + \mathrm{div}_x(f_t\nabla_x V),$$

can be obtained as the 2−Wasserstein gradient flow of the free energy

$$\mathfrak{F}(\rho) = \int f(x)\log f(x)\,dx + \int V(x)f(x)\,dx, \quad \text{if } f = \rho\cdot\mathscr{L}^N.$$

3.5. Sharp functional inequalities. We now illustrate, with a couple of significant examples, how the formal calculus based on the Benamou–Brenier formula can be exploited to derive sharp functional inequalities.

As in the previous subsection, our discussion, despite being quite an informal one, has the valuable aspect that (in principle) can be generalized to any dynamical transport distance. Also in this case, we confine for simplicity to the case $p = 2$, with $\mathscr{L}^N$ as reference measure: similar heuristic considerations can be found in [42, Section V].

We start taking a $\Lambda-$displacement convex functional $\mathfrak{F}$ over $\mathcal{W}_2(\mathbb{R}^N)$: this can be rephrased by saying that for every $\rho_0, \rho_1 \in \mathcal{W}_2(\mathbb{R}^N)$ and every $t \in [0,1]$ we have

$$\frac{\mathfrak{F}(\mu_t) - \mathfrak{F}(\rho_0)}{t} + \Lambda \frac{(1-t)}{2} w_2(\rho_0, \rho_1)^2 \leq \mathfrak{F}(\rho_1) - \mathfrak{F}(\rho_0),$$

with μ_t constant speed geodesic in $\mathcal{W}_2(\mathbb{R}^N)$ connecting ρ_0 to ρ_1. From the previous, taking the limit as $t \to 0^+$, we can obtain

$$\frac{d}{dt}\mathfrak{F}(\mu_t)_{|t=0^+} + \frac{\Lambda}{2} w_2(\rho_0, \rho_1)^2 \leq \mathfrak{F}(\rho_1) - \mathfrak{F}(\rho_0). \tag{3.15}$$

Suppose for example that $\mathfrak{F}$ has the following form

$$\mathfrak{F}(\rho) = \int U(f(x))\, dx + \frac{1}{2}\int |x|^2\, d\rho(x), \quad \text{for } \rho = f \cdot \mathscr{L}^N,$$

i.e. $\mathfrak{F}$ is the sum of an internal energy and a potential one, the latter coinciding with the 2nd moment of a measure. If U satisfies the displacement convexity condition of Theorem 2.1, then $\mathfrak{F}$ is $1-$displacement convex, thanks to the second term. Using again the continuity equation, with a velocity field v_t such that $\|v_t\|_{L^2(\mu_t)} = |\mu_t'|_{w_2}$ (see Theorem 3.1), we can compute the derivative of $\mathfrak{F}$ along the curve $\mu_t = f_t \cdot \mathscr{L}^N$, thus obtaining

$$\begin{aligned} &\frac{d}{dt}_{|t=0^+}\left(\int U(f_t)\, dx + \frac{1}{2}\int |x|^2\, d\mu_t\right) \\ &\geq -\frac{1}{2}\int |\nabla U'(f_t) + x|^2 \rho_0\, dx - \frac{1}{2} w_2(\rho_0, \rho_1)^2, \end{aligned}$$

where we used that μ_t is a constant speed geodesic, so that

$$w_2(\rho_0, \rho_1)^2 = |\mu_t'|^2_{w_2} = \int |v_t(x)|^2\, d\mu_t(x).$$

Inserting this into (3.15), the term with $w_2(\rho_0, \rho_1)^2$ cancels out and we would obtain

$$\mathfrak{F}(\rho_0) - \mathfrak{F}(\rho_1) \leq \frac{1}{2}\int |\nabla U'(\rho_0) + x|^2 \rho_0\, dx,$$

which is valid for every pair ρ_0, ρ_1. In particular, if the functional $\mathfrak{F}$ admits a (unique, thanks to the $1-$convexity) minimizer $\rho_m = f_m \cdot \mathscr{L}^N$ on the space $\{\rho \in \mathcal{W}_2(\mathbb{R}^N) : \rho \ll \mathscr{L}^N\}$, then we have shown that for every $\rho = f \cdot \mathscr{L}^N$ we have

$$\mathfrak{F}(\rho) - \mathfrak{F}(\rho_m) \leq \frac{1}{2}\int |\nabla U'(f(x)) + x|^2 f(x)\, dx, \tag{3.16}$$

and the equality sign holds if and only if $\rho = \rho_m$. Indeed, observe that the Euler–Lagrange equation of $\mathfrak{F}$ is given by

$$\nabla U'(f_m(x)) + x \equiv 0.$$

The quantity $\mathfrak{F}(\rho) - \mathfrak{F}(\rho_m)$ is also called *relative energy* of ρ with respect to ρ_m, while the right-hand side in (3.16) is called *relative entropy production* of ρ with respect to ρ_m (see [1]).

With suitable choices of the function U, we can derive interesting functional inequalities from the energy–entropy production inequality (3.16): taking $U(s) = s \log s$, developing the calculations and substituting f with $g^2/(\int g^2\,dx)$, we would obtain

$$\int_{\mathbb{R}^N} g(x)^2 \log g^2(x)\,dx + c_N \int_{\mathbb{R}^N} g(x)^2\,dx \le 2\int_{\mathbb{R}^N} |\nabla g(x)|^2\,dx + \left(\int_{\mathbb{R}^N} g(x)^2\,dx\right) \log\left(\int_{\mathbb{R}^N} g(x)^2\,dx\right),$$

where $c_N = N/2\,\log(2\pi e)$, i.e. the celebrated *logarithmic Sobolev inequality* of Gross with respect to the Lebesgue measure, in sharp form (see [23]). Thanks to a standard scaling argument, this can also be rewritten as

$$\int_{\mathbb{R}^N} \left(\frac{g(x)}{\|g\|_{L^2}}\right)^2 \log\left(\frac{g(x)}{\|g\|_{L^2}}\right) dx \le \frac{N}{4} \log\left(\frac{2}{\pi e N}\frac{\|\nabla g\|_{L^2}^2}{\|g\|_{L^2}^2}\right). \tag{3.17}$$

Remark 3.9. With this choice of U, it is not difficult to see that the unique minimizer $\rho_m = f_m \cdot \mathscr{L}^N$ of the free energy $\mathfrak{F}$ is given by

$$f_m(x) = (2\pi)^{-N/2}\,\exp(-|x|^2/2),$$

that is ρ_m is the standard Gaussian measure on $\mathbb{R}^N$. Then equality in (3.17) holds if and only if g equals, up to translations and multiplications, the function

$$\varphi(x) = \sqrt{f_m(x)}.$$

We point out a nice paper [4] by Beckner, where some interesting connections between (3.17), Nash's inequality

$$\left(\int_{\mathbb{R}^N} |f(x)|^2\,dx\right)^{1+2/N} \le c_N \int_{\mathbb{R}^N} |\nabla f(x)|^2\,dx \left(\int_{\mathbb{R}^N} |f(x)|\,dx\right)^{4/N}$$

and the isoperimetric inequality are derived.

In the very same way, choosing $U(s) = -Ns^{1-1/N}$ in (3.16), we could obtain the standard Sobolev inequality

$$\|g\|_{L^{2N/(N-2)}(\mathbb{R}^N)} \le c_N \|\nabla g\|_{L^2(\mathbb{R}^N)},$$

with sharp constant c_N, where $N \ge 3$ (see [40] for the classical proof based on symmetrizations). Indeed, in this case the energy-entropy production inequality (3.16) gives

$$-\mathfrak{F}(\rho_m) \le \frac{2(N-1)^2}{(N-2)^2} \int \left|\nabla(f(x)^{(N-2)/2N})\right|^2 \, dx,$$

then subsituting f with $g^{2N/(N-2)}/(\int g\, dx)^{2N/(N-2)}$ we precisely end up with Sobolev inequality. As in the previous case, the extremal functions for this inequality are translates, multiples and dilations of the function

$$\varphi(x) = f_m(x)^{(N-2)/2N},$$

with $\rho_m = f_m \cdot \mathscr{L}^N$ unique minimizer of the relevant functional $\mathfrak{F}$. The existence of such a minimizer, in this case, is not completely trivial: indeed the functional $-N\int f(x)^{1-1/N}\, dx$ is not l.s.c. with respect to the weak convergence and its relaxation is given by

$$\mathcal{U}^*(\rho) = -N \int f(x)^{1-1/N}\, dx, \quad \text{if } \rho = f \cdot \mathscr{L}^N + \rho_s,$$

with $\rho_s \perp \mathscr{L}^N$, due to the sub-linearity of the power $s^{1-1/N}$ (see [15]). The existence of a minimizer ρ_m of the relaxed functional

$$\mathfrak{F}^*(\rho) = \mathcal{U}^*(\rho) + \frac{1}{2}\int |x|^2\, d\rho(x),$$

in $\mathscr{P}(\mathbb{R}^N)$ is straightforward, thanks to the presence of the 2nd moment: then one can show that ρ_m can not be purely singular with respect to $\mathscr{L}^N$ and that its singular part has to be a Dirac delta concentrated in 0. Moreover, its absolutely continuous part f_m has to be supported on the whole $\mathbb{R}^N$ and finally has to satisfy the optimality condition

$$(1-N) f_m(x)^{-1/N} + \frac{|x|^2}{2} \equiv C.$$

Then one can finally show that the minimizer is of the form $\rho_m = f_m \cdot \mathscr{L}^N$, with

$$f_m(x) = \left(C + \frac{1}{N-1}\frac{|x|^2}{2}\right)^{-N}, \tag{3.18}$$

the constant C being chosen so that $\int_{\mathbb{R}^N} f_m(x)\,dx = 1$.

Remark 3.10. The functions (3.18) are usually called *Barenblatt–Prattle profiles*: if on the one hand they are connected to extremals in the Sobolev inequality, on the other hand they describe the long-time behaviour of the relevant gradient flow equation (see the previous subsection), given by the following fast diffusion equation with linear drift term

$$\partial_t f_t = \Delta f_t^{1-1/N} + \operatorname{div}(xf), \tag{3.19}$$

i.e. we have that $f_t \to f_m$ in $L^1(\mathbb{R}^N)$ as t goes to ∞, if f_t is a positive solution of (3.19) with $\int_{\mathbb{R}^N} f_t(x)\,dx = 1$ (see [20, Section 4] for more details).

§4. Some variants

4.1. Dynamical transport with finite speed of propagation. The first variant of (3.4) we want to take into account has been addressed by Brenier in the lecture notes [11]. Let us fix a parameter $c > 0$, which can be thought as a maximal admissible speed of propagation (for example, the speed of sound or the speed of light in a given medium). If we set

$$h(z) = \begin{cases} \left(1 - \sqrt{1 - \frac{|z|^2}{c^2}}\right) c^2, & |z| \leq c, \\ +\infty, & \text{otherwise}\ , \end{cases}$$

it is very natural to consider the action functional

$$\int_0^1 \mathcal{A}(\mu_t, v_t)\,dt := \int_0^1 \int h(v_t(x))\,d\mu_t(x)\,dt,$$

with the pair (μ, v) still solving the continuity equation: observe that this is the integral in time of a *relativistic* kinetic energy. More precisely, as in the case of w_2, we set

$$H(x,y) = \begin{cases} h\left(\frac{y}{x}\right) x, & \text{if } x > 0, y \in \mathbb{R}^N, \\ 0, & \text{if } x = 0, y = 0, \\ +\infty, & \text{otherwise}, \end{cases}$$

and observe that this is still a $1-$homogeneous and jointly convex function. In this way, the functional

$$\mathcal{H}(\mu, \phi) = \int_{[0,1]\times\mathbb{R}^N} H\left(\frac{d\mu}{dm}, \frac{d\phi}{dm}\right) dm,$$

is l.s.c. with respect to the $*$-weak convergence on $[0,1]\times\mathbb{R}^N$ and its integral representation does not depend on the choice of the reference measure m. Then again for every $\rho_0,\rho_1\in\mathscr{P}(\mathbb{R}^N)$ the following minimization problem is well-posed

$$W_{2,c}(\rho_0,\rho_1)^2 := \min\left\{\mathcal{H}(\mu,\phi)\ :\ \partial_t\mu+\operatorname{div}_x\phi=\rho_0\otimes\delta_0-\rho_1\otimes\delta_1\right\}. \quad (4.1)$$

Indeed, observe that the problem has enough coercivity properties with respect to the $*$-weak convergence: using the continuity equation as in the previous section, we have that every admissible μ is a probability, disintegrating as $\mu=\int\mu_t\,dt$, while using Jensen inequality, if $\phi\ll\mu$ we get

$$\mathcal{H}(\mu,\phi)=\int h\left(\left|\frac{d\phi}{d\mu}\right|\right)\,d\mu\ge h\left(\int d|\phi|\right)\ge a\int d|\phi|-b,$$

for suitable positive constants a,b, not depending on ϕ. In this way, we see that taken a sequence of admissible pairs $\{(\mu^n,\phi^n)\}_{n\in\mathbb{N}}$ having equi-bounded energies, we obtain that $\mu^n\overset{*}{\rightharpoonup}\mu$ and $\phi^n\overset{*}{\rightharpoonup}\phi$, up to subsequences. Moreover, these limit measures (μ,ϕ) still solve the continuity equation

$$\partial_t\mu+\operatorname{div}_x\phi=\rho_0\otimes\delta_0-\rho_1\otimes\delta_1,$$

in the sense of distributions.

Remark 4.1. Clearly $\mathcal{H}(\mu,\phi)<+\infty$ not only gives that $\phi\ll\mu$, but it also implies an upper bound on the velocity of μ, i.e. $|d\phi_t/d\mu_t|\le c$. Due to this bound on the speed of propagation, now it could happen that for some measures ρ_0 and ρ_1 the previous problem does not admit any configuration with finite energy: for example, taking $\rho_0=\delta_{x_0}$ and $\rho_1=\delta_{x_1}$, with $|x_1-x_0|>c$, and choosing the time interval $[0,1]$, it is easy to see that $H(\mu_t,\phi_t)=+\infty$ for every admissible pair (μ,ϕ).

Once again, it is interesting to investigate the dual formulation of (4.1): noting that the Legendre–Fenchel transform of h is

$$h^*(\xi)=c^2\left(\sqrt{\frac{|\xi|^2}{c^2}+1}-1\right),$$

we get that the transform of H is given by

$$H^*(\xi_1,\xi_2)=\begin{cases}0, & \text{if } \xi_1+c^2\left[\sqrt{\frac{|\xi_2|^2}{c^2}+1}-1\right]\le 0,\\ +\infty, & \text{otherwise.}\end{cases}$$

Then, similarly to the previous case, we can obtain that (4.1) has the following dual formulation

$$\sup\left\{\int \varphi(1,x),\ d\rho_1(x) - \int \varphi(0,x)\,d\rho_0(x) : \partial_t\varphi + c^2\left[\sqrt{\frac{|\nabla\varphi|^2}{c^2}+1}-1\right]\le 0\right\}.$$

In particular, the formal optimality conditions for the "geodesics" of $W_{2,c}$ are given by

$$\begin{cases} \partial_t\mu_t + \operatorname{div}_x\left(\mu_t \dfrac{c\nabla_x\psi_t}{\sqrt{c^2+|\nabla_x\psi_t|^2}}\right) & = & 0, \\ \partial_t\psi_t + c^2\left[\sqrt{\dfrac{|\nabla_x\psi_t|^2}{c^2}+1}-1\right] & = & 0, \end{cases}$$

so that now a geodesic is driven by the velocity field

$$v_t = c\nabla_x\psi_t\,(\sqrt{c^2+|\nabla_x\psi_t|^2})^{-1},$$

with the potential ψ_t still solving a Hamilton-Jacobi equation.

Remark 4.2. When $c\to\infty$, we have

$$c^2\left[\sqrt{\frac{|\nabla\varphi|^2}{c^2}+1}-1\right]\simeq \frac{1}{2}|\nabla\varphi|^2,$$

so that we are back to the case of the 2-Wasserstein distance.

Using the same formal calculus as in the Wasserstein case, we can guess which kind of interesting evolution equations corresponds to gradient flows with respect to the dynamical transport distance $\widetilde{w}_2$: for example, taking the Boltzmann entropy functional (2.5), we get

$$\frac{d}{dt}\mathcal{U}(\mu_t) = \int \partial_t f_t(\log f_t + 1)\,dx = \int \left\langle \frac{\nabla f_t}{f_t}, v_t\right\rangle f_t\,dx,$$

and then an application of Young inequality yields

$$\frac{d}{dt}\mathcal{U}(\mu_t) \ge -\int h^*\left(\frac{\nabla f_t}{f_t}\right) f_t\,dx - \int h(v_t) f_t\,dx.$$

Assuming that a result analogous to Theorem 3.1 holds also in the case of $W_{2,c}$, with $\int h(v_t)\,d\mu_t$ replacing the term $\int |v_t|^p\,d\mu_t$, we can assume the existence of a velocity field v_t such that $\int h(v_t)\,f_t\,dx = |\mu_t'|^2_{W_{2,c}}$.

Remembering the cases of equality in Young inequality, then $\mu_t = f_t \cdot \mathscr{L}^N$ is a curve of steepest descent if and only if

$$v_t = -\nabla h^*\left(\frac{\nabla f_t}{f_t}\right) = -\frac{c\nabla f_t}{\sqrt{|\nabla f_t|^2 + c^2 f_t^2}}.$$

In this way, we obtain that such a curve μ_t solves the *relativistic heat equation*

$$\partial_t f_t = \operatorname{div}_x\left(\frac{c f_t |\nabla f_t|}{\sqrt{c^2 f_t^2 + |\nabla f_t|^2}}\right). \tag{4.2}$$

The latter corresponds to a diffusion equation where the speed of propagation is bounded and tends to saturate as the gradient of f becomes unbounded: the classical heat equation can be recovered in the limit as c goes to ∞. The previous interpretation of (4.2) as gradient flow of the entropy functional w.r.t. $W_{2,c}$ has been rigorously proven by McCann and Puel, using a minimizing movement scheme in the spirit of the work of Jordan, Kinderlehrer and Otto (see [33]).

Finally, the reader can consult [3] for a discussion of equation (4.2) and some related regularity results.

4.2. Optimal Transport with penalization on high concentrations. In connection with *congestion* effects and crowd motions, other variants of the Benamou-Brenier functional including penalizations on high densities have been considered in [16]. In this context, the two prototypical examples are the following:

$$\mathcal{A}(\mu,\phi) = \mathcal{F}_p(\mu,\phi) + k \int_{[0,1]\times\mathbb{R}^N} \left|\frac{d\mu}{dm}(t,x)\right|^2 dm(t,x), \tag{4.3}$$

with $p \geq 1$ and $k > 0$, and

$$\mathcal{B}(\mu,\phi) = \begin{cases} \mathcal{F}_p(\mu,\phi), & \text{if } \left|\dfrac{d\mu}{dm}\right| \leq M, \\ +\infty, & \text{otherwise,} \end{cases} \tag{4.4}$$

with $M > 0$, where the reference measure is $m = \mathscr{L}^1 \otimes \mathscr{L}^N$ and both functionals are set to be $+\infty$ if μ is not of the form $\mu = \int \mu_t \, dt$, with $\mu_t \ll \mathscr{L}^N$. Observe that in (4.3), curves of diffused measures are favoured not only because the functional has a finite value only on them, thanks to the choice of m, but also because the function $x \mapsto x^2$ is super-additive, that is

$$x_1^2 + x_2^2 < (x_1 + x_2)^2,$$

so that the masses have the interest to split as much as possible during the transport, in order to lower the value of the energy functional. In (4.4) the penalization is even stronger, because it imposes an L^∞ bound on particle densities. Clearly, (4.3) and (4.4) are no more independent on the choice of m, but they are still l.s.c. and coercive with respect to the $*-$weak convergence, so that for both of them the problem of finding a minimizer under the constraint of the continuity equation is well-posed (see [16] for more details and some interesting numerical simulations).

§5. The case of a non linear mobility function

5.1. Some motivations. All the transport models considered so far were based on the continuity equation $\partial_t \mu_t + \mathrm{div}_x(\phi_t) = 0$, possibly with homogeneus Neumann boundary condition, expressing conservation of mass. All of them contained the physical assumption that the flux variable was of the form $\phi_t = v_t \cdot \mu_t$. This assumption is typical in fluid mechanics, but when considering more general situations, usually this is not appropriate. For example, in mathematical biology, some models for *chemotaxis* assume the flux ϕ_t to be of the form

$$\phi_t = v_t \cdot \mu_t \left(1 - \frac{\mu_t}{\gamma}\right), \tag{5.1}$$

where $\gamma \geq 1$ denotes a maximal density for the particles distribution μ_t at time t and $0 \leq \mu_t \leq \gamma$. Observe that setting

$$\widetilde{v}_t = v_t \left(1 - \frac{\mu_t}{\gamma}\right),$$

we have $\partial_t \mu_t + \mathrm{div}_x(\widetilde{v}_t \mu_t) = 0$ and we recover the usual continuity equation, but now with the mass-dependent velocity field $\widetilde{v}_t$ which tends to 0 as the concentration of the particles is near the critical threshold γ. In this way, assumption (5.1) is pertinent to a model for chemotaxis where particles are supposed to move avoiding overcrowding effects (see [24], model (M3a), for example).

More generally, in a number of mathematical models coming from biology, physics and chemistry (see the Introduction of [17] and the references therein), the flux is assumed to have the form $\phi_t = v_t \cdot \theta(\mu_t)$, the function

θ being called *mobility function*. It is thus natural to consider the following generalization of problem (3.1),

$$\min\left\{\int_0^1\int |v_t(x)|^p\theta(f_t(x))\,dx\,dt\ :\ \begin{array}{c}\partial_t f_t+\mathrm{div}_x(v_t\,\theta(f_t))=0,\\ f_i=g_i,\ i=0,1\end{array}\right\},$$

assuming for a moment that the admissible curves are of the form $\mu_t = f_t\cdot\mathscr{L}^N$. This type of problems has been introduced in [21] by Dolbeault, Nazaret and Savaré: further studies can be found also in [17, 28] (where the case of the mobility function corresponding to (5.1) is considered).

5.2. Mathematical framework. From a mathematical point of view, these models consist in replacing the Benamou-Brenier function $|\phi|^p\mu^{1-p}$, by the more general function

$$(\mu,\phi)\mapsto|\phi|^p\theta(\mu)^{1-p},$$

with $\theta:[0,\infty)\to[0,\infty)$ concave increasing function. The concavity of θ guarantees that the previous function is still jointly convex, but now it is no more 1-homogeneous, so that the corresponding action functional

$$(\mu,\phi)\mapsto\int_0^1\int\frac{|\phi_t|^p}{\theta(\mu_t)^{p-1}},$$

will now also depend on the choice of the reference measure m. Particularly interesting choices for θ are power functions, i.e.

$$\theta(t)=t^\beta,\ \text{with } 0<\beta<1.$$

In what follows, we will mainly confine to consider this case. Then we take as reference measure $m=\nu\otimes(\mathscr{L}^1\llcorner[0,1])$, with ν positive Radon measure on $\mathbb{R}^N$, and the resulting action considered in [21] is given by

$$\mathcal{S}_{p,\beta;\nu}(\mu,\phi)=\int_0^1\int_{\mathbb{R}^N}s_{p,\beta}\left(\frac{d\mu_t}{d\nu}(x),\frac{d\phi_t}{d\nu}(x)\right)d\nu(x)\,dt,$$

set to be $+\infty$ if $\phi_t\not\ll\nu$ for $t\in A$ with $\mathscr{L}^1(A)>0$. Here $s_{p,\beta}(x,y)=|y|^p\,x^{\beta(1-p)}$ or more precisely

$$s_{p,\beta}(x,y)=\begin{cases}|y|^p\,x^{\beta(1-p)}, & \text{if } x>0,\ y\in\mathbb{R}^N,\\ 0, & \text{if } x=0,\ y=0,\\ +\infty, & \text{otherwise.}\end{cases}$$

Observe in particular that

$$\mathcal{S}_{p,\beta;\nu}(\mu,\phi) < +\infty \Longrightarrow \phi_t \ll \nu \text{ for } \mathscr{L}^1\text{-a.e. } t,$$

and the functional can assume finite values also if μ_t has a singular part with respect to ν: this is crucial in order to obtain the l.s.c. of $\mathcal{S}_{p,\beta}$. Indeed, observe that the *recession function* (see [15]) of $s_{p,\beta}$ is given by

$$s^{\infty}_{p,\beta}(x,y) = \lim_{t\to+\infty} \frac{s_{p,\beta}(tx,ty)}{t} = \begin{cases} 0, & \text{if } y=0,\\ +\infty, & \text{otherwise.}\end{cases}$$

In this way, the problem of minimizing $\mathcal{S}_{p,\beta}$ under the constraint of the continuity equation turns out to be convex as well and we have the following result.

Theorem 5.1. *For every $\rho_0,\rho_1 \in \mathscr{P}(\mathbb{R}^N)$ the following problem*

$$\mathscr{W}_{p,\beta;\nu}(\rho_0,\rho_1)^p := \min\{\mathcal{S}_{p,\beta;\nu}(\mu,\phi) \,:\, \partial_t\mu + \mathrm{div}_x\phi = \delta_0\otimes\rho_0 - \delta_1\otimes\rho_1\}, \quad (5.2)$$

admits at least a solution, provided there exists an admissible pair $(\widetilde{\mu},\widetilde{\phi})$ having finite energy.

Proof. Also in this case, the coercivity of the problem is guaranteed by the convexity of the functional and by the continuity equation. Indeed, for every admissible pair (μ,ϕ) having finite energy, using Jensen inequality and the fact that $\phi_t \ll \nu$, we get

$$\begin{aligned}\mathcal{S}_{p,\beta;\nu}(\mu,\phi) &= \int_0^1\int_{\mathbb{R}^N}\left[\left|\frac{d\phi_t}{d\nu}(x)\right|\left(\frac{d\mu_t}{d\nu}(x)\right)^{-\beta}\right]^p\left(\frac{d\mu_t}{d\nu}(x)\right)^{\beta} d\nu(x)\,dt\\ &\ge \left(\int_0^1\int_{\mathbb{R}^N}\left(\frac{d\mu_t}{d\nu}(x)\right)^{\beta} d\nu(x)\,dt\right)^{1-p}\left(\int_0^1\int_{\mathbb{R}^N} d|\phi_t|(x)\,dt\right)^{p}.\end{aligned}$$

Let us suppose for simplicity that $\nu(\mathbb{R}^N) < +\infty$ (this hypothesis can be easily removed), then using once again Jensen inequality in the first term on the right-hand side, the previous estimate implies

$$|\phi|([0,1]\times\mathbb{R}^N) \le \mathcal{S}_{p,\beta;\nu}(\mu,\phi)^{1/p}\left(\int_0^1 \nu\left(\mathbb{R}^N\right)^{1-\beta}\mu_t(\mathbb{R}^N)^{\beta}\,dt\right)^{(p-1)/p}. \quad (5.3)$$

Finally, observe that $\mu_t(\mathbb{R}^N) = 1$ for every t, thanks to the continuity equation: then (5.3) gives the desired coercivity on the variable ϕ, with respect to the $*$-weak convergence on $[0,1]\times\mathbb{R}^N$.

This implies that every sequence $\{(\mu^n,\phi^n)\}_{n\in\mathbb{N}}$ such that $\mathcal{S}_{p,\beta;\nu}(\mu^n,\phi^n) \le C$ is $*$-weak convergent and the limit measures (μ,ϕ) solve the continuity equation. It remains to check that μ and ϕ are still of the form $\mu = \int \mu_t\,dt$ and $\phi = \int \phi_t\,dt$, because in the definition of $\mathcal{S}_{p,\beta;\nu}$ we a priori restricted the class of competitors to curves of measures, rather than general measures on $[0,1]\times\mathbb{R}^N$. However, the disintegration $\mu = \int \mu_t\,dt$ comes as always from the fact that μ solves the continuity equation (see Section 3), while thanks to the semicontinuity of the energy we have

$$\mathcal{S}_{p,\beta;\nu}(\mu,\phi) \le \liminf_{n\to\infty} \mathcal{S}_{p,\beta;\nu}(\mu^n,\phi^n) \le C,$$

thus giving the desired disintegration for ϕ, with $\phi_t \ll \nu$. $\square$

The following result characterizes the basic topological properties of the dynamical transport distance $\mathcal{W}_{p,\beta;\nu}$ defined by (5.2): see [21, Theorems 5.5 and 5.7] for the proofs.

Theorem 5.2. *The application $\mathscr{W}_{p,\beta;m} : \mathscr{P}(\mathbb{R}^N)\times\mathscr{P}(\mathbb{R}^N) \to [0,+\infty]$ is a pseudo-distance. Given $\eta \in \mathscr{P}(\mathbb{R}^N)$, if we define*

$$\mathcal{M}_{p,\beta;\nu}(\eta) = \{\rho \in \mathscr{P}(\mathbb{R}^N) \,:\, \mathscr{W}_{p,\beta;\nu}(\eta,\rho) < +\infty\},$$

this is a complete metric space, when endowed with the distance $\mathscr{W}_{p,\beta;\nu}$. On the space $\mathcal{M}_{p,\beta;\nu}(\eta)$, the convergence with respect to $\mathscr{W}_{p,\beta;\nu}$ is stronger than the $*-$weak one.

Remark 5.3. This new family of "distances" interpolate between the usual Wasserstein ones, corresponding to the choice $\beta = 1$ (so that $\mathcal{S}_{p,1;\nu} = \mathcal{F}_p$), and the dual Sobolev ones, corresponding to $\beta = 0$, for which $\mathcal{S}_{p,0;\nu}$ takes the form (here we choose $\nu = \mathscr{L}^N$)

$$\int_0^1 \int_{\mathbb{R}^N} \left|\frac{d\phi_t}{d\mathscr{L}^N}(x)\right|^p dx\,dt, \tag{5.4}$$

set to be $+\infty$ if ϕ is not of the form $\phi = \int \phi_t\,dt$ with $\phi_t \ll \mathscr{L}^N$. Observe that the problem corresponding to (5.4) is equivalent to

$$\min\left\{\int_{\mathbb{R}^N} |\Phi(x)|^p\,dx \,:\, \operatorname{div}\Phi = \rho_0 - \rho_1\right\},$$

which is just the dual formulation of an elliptic problem involving the $q-$Laplace operator, with $q = p/(p-1)$.

5.3. Geodesics. For the dynamical transport distance $\mathcal{W}_{p,\beta;\nu}$ still holds a characterization of AC curves, analogous to that of Theorem 3.1. Moreover, as one can easily guess, the space of measures endowed with this metric is a geodesic one (see [21, Corollary 5.18]).

Theorem 5.4. *For every $\eta \in \mathscr{P}(\mathbb{R}^N)$, the space $\mathcal{M}_{p,\beta;\nu}(\eta)$ is a geodesic space, that is for every $\rho_0, \rho_1 \in \mathcal{M}_{p,\beta;\nu}(\eta)$ there exists a constant speed geodesic μ_t connecting them and such that*

$$\mathscr{W}_{p,\beta;\nu}(\mu_t,\mu_s) = |t-s|\mathscr{W}_{p,\beta;\nu}(\rho_0,\rho_1), \quad \textit{for every } s,t \in [0,1].$$

It is interesting to investigate, still at a formal level, the conditions for a curve μ to be a geodesic in $(\mathcal{M}_{p,\beta;\nu}(\eta), \mathscr{W}_{p,\beta;\nu})$. One can proceed as in the previous sections, rewriting the variational problem defining $\mathscr{W}_{p,\beta;m}$ as a saddle-point problem and exchanging the inf and the sup. When $\nu = \mathscr{L}^N$, we obtain that (μ_t, ϕ_t) is a geodesic if

$$\mu_t = f_t \cdot \mathscr{L}^N \quad \text{and} \quad \phi_t = \nabla_x \psi_t \, f_t^\beta \cdot \mathscr{L}^N,$$

with the pair (f_t, ψ_t) solving

$$\begin{cases} \partial_t f_t + \operatorname{div}_x(f_t^\beta \, \nabla_x \psi_t) &= 0 \\ \\ \partial_t \psi_t + \dfrac{\beta}{2} f_t^{\beta-1} \, |\nabla_x \psi_t|^2 &= 0 \end{cases} \tag{5.5}$$

which is a kind of *mean field games* system (see [26]): the main difference with the systems considered by Lasry and Lions in [26] is that (5.5) comes with a constraint on f_0 and f_1, while usually in mean field games this system is forward in time with respect to f (i.e. f_0 is prescribed) and *backward in time* with respect to ψ (i.e. ψ_1 is prescribed).

Observe that in both the equations of (5.5) the two variables are coupled: this is in contrast with the Wasserstein case, corresponding to (3.9), where the Hamilton-Jacobi equation can be solved independently of f_t.

5.4. Gradient flows issues. The distances $\mathscr{W}_{p,\beta;\nu}$ have a lot of interesting features and a wide collection of open questions are connected with them: first, we point out that these distances are interesting in the study of diffusion equations of the type

$$\partial_t \mu_t + \operatorname{div}_x(\theta(\mu_t) \, |\xi|^{q-2} \xi), \quad \xi = -\nabla \left(\frac{\partial \mathfrak{F}}{\partial \mu} \right),$$

where $\partial\mathfrak{F}/\partial\mu$ is the first variation of a given functional $\mathfrak{F}$ and $q = p/(p-1)$. Indeed, at least formally, these equations can be interpreted as gradient flows of $\mathfrak{F}$ with respect to $\mathscr{W}_{p,\beta;\nu}$. For example, taking $\nu = \mathscr{L}^N$ and the internal energy functional

$$\mathcal{U}(\rho) = \frac{1}{(2-\beta)(1-\beta)} \int f(x)^{2-\beta}\, dx, \quad \text{if } \rho = f \cdot \mathscr{L}^N, \tag{5.6}$$

we see that the gradient flow with respect to the distance $\mathscr{W}_{2,\beta;\mathscr{L}^N}$ is given by the heat equation. The formal derivation of this fact can be done exactly in the same way as in the previous section, i.e. differentiating the energy $\mathcal{U}$ along a curve μ_t, using the continuity equation and observing that for a curve of steepest descent we have precisely $\phi_t = \nabla f_t\, f_t^{-\beta}$.

However, in this new context, the picture is undoubtedly less clear with respect to the Wasserstein case. The main difficulty in establishing rigorously these results is the lack of informations on the geodesics of the space of probabilities endowed with the distance $\mathscr{W}_{2,\beta;\mathscr{L}^N}$: indeed, these are not directly related to geodesics of the base space $\mathbb{R}^N$ through a formula like (2.3). In particular, it is not easy to guess the conditions for the displacement convexity in this context. This question has started to be investigated in the recent paper [17]. There in particular the following sufficient conditions have been established for the geodesic convexity of an internal energy functional, in the case of a convex bounded set Ω, with reference measure $\nu = \mathscr{L}^N \llcorner \Omega$.

Theorem 5.5 (Generalized displacement convexity). *Let us consider the internal energy functional*

$$\mathcal{U}(\rho) = \int_\Omega U(f(x))\, dx, \qquad \text{if } \rho = f \cdot \mathscr{L}^N,$$

with $U : [0,+\infty) \to [0,+\infty)$ *smooth, convex and such that*

$$\lim_{s\to+\infty} U(s)s^{-1} = +\infty.$$

We extend $\mathcal{U}$ *to the whole space of probability measures* $\mathscr{P}(\Omega)$ *by setting* $\mathcal{U}(\rho) = +\infty$ *if* $\rho \not\ll \mathscr{L}^N$. *We then define*

$$P(s) = \int_0^s U''(r) r^\beta\, dr \qquad \text{and} \qquad H(s) = \beta \int_0^s U''(r) r^{2\beta-1}\, dr.$$

Suppose that there results

$$P'(s)s^\beta \geq \left(1 - \frac{1}{N}\right) H(s) \geq 0, \quad \textit{for every } s > 0. \tag{5.7}$$

Then for every $\rho_0, \rho_1 \in \mathscr{P}(\Omega)$ such that $\mathscr{W}_{p,\beta;\mathscr{L}^N}(\rho_0, \rho_1) < +\infty$ and with finite energy $\mathcal{U}$, there exists a constant speed geodesic μ_t connecting ρ_0 to ρ_1 and such that

$$\mathcal{U}(\mu_t) \leq (1-t)\mathcal{U}(\rho_0) + t\mathcal{U}(\rho_1), \qquad \textit{for every } t \in [0,1].$$

Remark 5.6. Observe that in the case $\beta = 1$, corresponding to the usual Wasserstein distance, we have $P(s) = H(s)$ and (5.7) coincides with the usual condition for the displacement convexity (see Section 2, Remark 2.2). Also observe that $U(s) = s \log s$ still verifies (5.7), while

$$U(s) = \frac{1}{\vartheta - 1} s^\vartheta,$$

verifies (5.7) if and only if $\vartheta \geq 2 - \beta(1 + 1/N)$, with $\vartheta \neq 1$. In particular the internal energy functional (5.6) is displacement convex in this generalized sense and one can rigorously establish that the gradient flow of (5.6) with respect to $\mathcal{W}_{2,\beta;\mathscr{L}^N}$ is given by the heat equation (see [21, Theorem 5.29]).

§6. Branched transport problems

As already recalled in the introduction, this kind of problems received a certain interest in the last years and various variational formulations have been proposed. First of all, we want to recall some of them: the presentation will be rather sketchy and informal, for more details and results the reader should consult the monography [8]. In what follows, Ω will always be a compact and convex subset of $\mathbb{R}^N$.

6.1. Some models. The first model to be proposed has been the one by Gilbert ([22]) in the '60s, dealing with finitely discrete sources and destinations, i.e.

$$\rho_0 = \sum_{i=1}^k a_i \delta_{x_i} \quad \text{ and } \quad \rho_1 = \sum_{j=1}^m b_j \delta_{y_j},$$

such that $\sum_{i=1}^k a_i = \sum_{j=1}^m b_j = 1$. The admissible transportation structures are represented by *weighted oriented graphs* $\mathfrak{g}$ sastisfying Kirchhoff's

Law for circuits, whose total cost is given by

$$M_\alpha(\mathfrak{g}) = \sum_{e_h \in \mathfrak{g}} m_h^\alpha \, \mathscr{H}^1(e_h), \tag{6.1}$$

where e_h are the edges of the graph $\mathfrak{g}$ and m_h is the weight associated with the edge e_h. In what follows, we will refer to an energy of the type (6.1) as a *Gilbert-Steiner energy.*

The model introduced by Xia in [43] is based on a relaxation procedure, starting from the previous energy M_α defined for ρ_0 and ρ_1 finitely atomic probability measures. This can be seen as a natural extension to general measures of Gilbert's model and the resulting energy has the following integral expression

$$M^*_\alpha(\Phi) = \begin{cases} \int m(x)^\alpha \, d\mathscr{H}^1(x), & \text{if } \Phi = m \overrightarrow{\tau} \mathscr{H}^1 \llcorner \Sigma, \\ +\infty, & \text{otherwise.} \end{cases}$$

defined over vector measures[3] $\Phi \in \mathcal{M}(\Omega;\mathbb{R}^N)$. Here Σ is a $1-$rectifiable set and the vector field $\overrightarrow{\tau}$ is an orientation, belonging to the approximated tangent space to Σ, while m is the multiplicity. In particular, this is still a Gilbert-Steiner energy. Then in this framework the branched transport problem is formulated as

$$d_\alpha(\rho_0, \rho_1) := \min\{M^*_\alpha(\Phi) \, : \, \mathrm{div}\, \Phi = \rho_0 - \rho_1\}. \tag{6.2}$$

Observe that this is quite close in spirit to the models presented so far, except for the fact that the model is completely static and a dynamical description of the branched transport is missing: in particular, we could say that this is a sort of Eulerian model. We recall the following result.

Theorem 6.1 (Xia). *Let $\alpha \in (1-1/N, 1]$ and $\rho_0, \rho_1 \in \mathscr{P}(\Omega)$, then the minimization problem defining $d_\alpha(\rho_0, \rho_1)$ does admit a solution with finite energy. Moreover $d_\alpha : \mathscr{P}(\Omega) \times \mathscr{P}(\Omega) \to [0, \infty)$ defines a distance on $\mathscr{P}(\Omega)$ which metrizes the $*-$weak convergence and such that $(\mathscr{P}(\Omega), d_\alpha)$ is a geodesic space.*

Remark 6.2. The condition on the exponent α is sharp: for example, when $\alpha \leq 1 - 1/N$, it is not possible to find a Φ with finite M^*_α energy, transporting $\rho_0 = \delta_{x_0}$ to $\rho_1 = |\Omega|^{-1} \mathscr{L}^N \llcorner \Omega$.

[3]A related model based on 1-rectifiable currents, instead of vector measures, has been studied by Paolini and Stepanov in [39].

On the other hand, some Lagrangian descriptions have been proposed by many authors (Bernot, Caselles, Figalli, Maddalena, Morel, Solimini). Here we just recall the one by Bernot, Caselles and Morel ([6,7]), which is the more flexible: for the others, one can consult [9,29]. We consider the following transport cost

$$E_\alpha(Q) = \int_{\mathcal{L}} \int_0^1 [\sigma(t)]_Q^{\alpha-1} \, |\sigma'(t)| dt \, dQ(\sigma), \; Q \in \mathscr{P}(\mathcal{L}),$$

where $\mathcal{L} = \mathrm{Lip}([0,1];\Omega)$ and the *multiplicity* $[\,\cdot\,]_Q$ is defined by

$$[x]_Q = Q(\{\sigma \,:\, \sigma([0,1]) \ni x\}), \; x \in \Omega,$$

which represents the total cumulated transiting mass at the point x. In particular, the term

$$\sigma \mapsto \int_0^1 [\sigma(t)]_Q^{\alpha-1} \, |\sigma'(t)| \, dt,$$

is a sort of weighted-length functional, with the weight depending on the mass travelling on the curve (we have in particular that $[\sigma(t)]_Q^{\alpha-1} \geq 1$). We remark that, dimensionally, the energy E_α is a Gilbert-Steiner one. In such a setting, the transport problem is formulated as

$$\mathcal{E}_\alpha(\rho_0, \rho_1) := \min\{E_\alpha(Q) \,:\, Q \in \mathscr{P}(\mathcal{L}), \; (e_i)_\# Q = \rho_i, \; i = 0, 1\}, \quad (6.3)$$

where $e_t : \mathcal{L} \to \Omega$ is given by $e_t(\sigma) = \sigma(t)$, the *evaluation at time t map.*

Remark 6.3. Observe that the functional E_α is invariant under time reparametrizations, thanks to the definition of the multiplicity.

The following important fact is proven in [8, Chapter 9].

Theorem 6.4 (Bernot–Caselles–Morel)**.** *The two models corresponding to* (6.2) *and* (6.3) *are equivalent.*

Here by equivalent we mean that that the two models describe the same kind of energy and the same optimal structures of branched transport: the simple equality of the minima is just a consequence of this more important fact.

6.2. A Benamou–Brenier formula in the branched setting. In [13] this transport problem has been settled in the framework of dynamical transport distance. Let us introduce the following local l.s.c. functional defined on measure by (see [10])

$$g_\alpha(\rho) = \int_\Omega \rho(\{x\})^\alpha \, d\#(x),$$

where $\#$ is the counting measure and g_α is set to be $+\infty$ if ρ is not a purely atomic measure, i.e. the sum of countably many Dirac masses. Then we consider the action functional

$$\mathcal{G}_\alpha(\mu,\phi) := \int_0^1 g_\alpha\left(\left|\frac{d\phi_t}{d\mu_t}(x)\right|^{1/\alpha}\mu_t(x)\right) dt, \tag{6.4}$$

wih the convention $\mathcal{G}_\alpha = +\infty$ if $\phi_t \not\ll \mu_t$ for a non negligible set of times. About the functional $\mathcal{G}_\alpha$, some words are in order: first, observe that this energy has exactly the dimensions of a Gilbert-Steiner one, indeed considering the term $d\phi_t/d\mu_t$ as a velocity, we have

$$\mathcal{G}_\alpha(\mu,\phi) \simeq \sum \frac{d\ell}{dt} m^\alpha dt = \sum m^\alpha d\ell.$$

Second, the term inside the functional g_α is exactly the integrand of the Benamou-Brenier functional and the finiteness of the energy implies that ϕ_t (not in any case μ_t) has to be purely atomic for each time t, that is

$$\mathcal{G}_\alpha(\mu,\phi) < +\infty \Longrightarrow \phi_t \text{ is atomic.}$$

This implies that only the mass that is effectively moving has to be atomic.

Our model of branched transport is then given by

$$\mathfrak{B}_\alpha(\rho_0,\rho_1) = \min\left\{\mathcal{G}_\alpha(\mu,\phi) \; : \; \begin{array}{c} \partial_t\mu_t + \mathrm{div}_x\phi_t = 0, \\ \mu_i = \rho_i,\, i = 0,1 \end{array}\right\}, \tag{6.5}$$

and in [13], Theorem 2, we have proven the following existence result.

Theorem 6.5. *Given $\rho_0, \rho_1 \in \mathscr{P}(\Omega)$, the minimization problem* (6.5) *admits a solution, provided there exist an admissible pair (μ,ϕ) with finite energy.*

Proof. The proof uses the direct methods in the calculus of variations: we give a sketch of it, referring the reader to [13] for more details. First of all, observe that in (6.5) the functional to be minimized is not convex, but rather concave with respect to μ: lower semicontinuity is then a non trivial fact (see Remark below). On the other hand, convexity is hidden in the functional: indeed, with respect to the velocity variable $v_t = d\phi_t/d\mu_t$, $\mathcal{G}_\alpha$ is convex and 1−homogeneous. This implies that the functional $\mathcal{G}_\alpha$ can be regarded as a sort of length functional and it is in particular invariant under time reparametrizations. More precisely, given an admissible pair (μ, ϕ) and a strictly increasing time reparametrization $\mathfrak{t} : [0,1] \to [0,1]$, we can define the new pair

$$\widetilde{\mu}_s = \mu_{\mathfrak{t}(s)} \quad \text{and} \quad \widetilde{\phi}_s = \mathfrak{t}'(s)\phi_{\mathfrak{t}(s)},$$

and we have $\mathcal{G}_\alpha(\mu,\phi) = \mathcal{G}_\alpha(\widetilde{\mu},\widetilde{\phi})$, with $(\widetilde{\mu},\widetilde{\phi})$ still solving the continuity equation. Moreover, the following basic inequality holds true: let (μ,ϕ) with $\mathcal{G}_\alpha(\mu,\phi) < +\infty$ and such that $\partial_t \mu_t + \mathrm{div}_x \phi_t = 0$, then using the subadditivity of the map $s \mapsto s^\alpha$ and Theorem 3.1, we obtain

$$\begin{aligned} g_\alpha\left(\left|\frac{d\phi_t}{d\mu_t}\right|^{1/\alpha}\mu_t\right) &= \sum_{i(t)} |v_t(x_{i(t)})| \mu_t(x_i(t))^\alpha \\ &\geq \left(\sum_{i(t)} |v_t(x_{i(t)})|^{1/\alpha}\mu_t(x_{i(t)})\right)^\alpha \\ &= \|v_t\|_{L^{1/\alpha}(\mu_t)} \geq |\mu_t'|_{w_{1/\alpha}}, \end{aligned} \tag{6.6}$$

where we indicated with $x_{i(t)}$ the atoms of ϕ_t. Also, using the simple inequality $\|v_t\|_{L^{1/\alpha}(\mu_t)} \geq \|v_t\|_{L^1(\mu_t)}$, we obtain a bound on the total variation of ϕ, i.e.

$$\mathcal{G}_\alpha(\mu,\phi) \geq \int_0^1 \|v_t\|_{L^1(\mu_t)}\,dt = \int_{[0,1]\times\Omega} d|\phi|. \tag{6.7}$$

Taking an infimizing sequence $\{(\mu^n,\phi^n)\}_{n\in\mathbb{N}}$ with equi-bounded energy $\mathcal{G}_\alpha$, we start reparametrizing each pair in such a way that

$$g_\alpha\left(\left|\frac{d\widetilde{\phi}_t}{d\widetilde{\mu}_t}\right|^{1/\alpha}\widetilde{\mu}_t\right) \equiv \mathcal{G}_\alpha(\widetilde{\mu}^n,\widetilde{\phi}^n) = \mathcal{G}_\alpha(\mu^n,\phi^n),$$

that is we choose the reparametrizations $\mathfrak{t}_n : [0,1] \to [0,1]$ implicitely defined by

$$\mathcal{G}_\alpha(\mu^n,\phi^n)\,s = \int_0^{\mathfrak{t}_n(s)} g_\alpha\left(\left|\frac{d\phi_r^n}{d\mu_r^n}\right|^{1/\alpha}\mu_r^n\right) dr, \;\; s\in[0,1].$$

In this way, (6.6) implies that $\{\widetilde{\mu}^n\}_{n\in\mathbb{N}}$ is equi-Lipschitz in $\mathcal{W}_{1/\alpha}(\Omega)$, so that (up to subsequences) $\widetilde{\mu}^n_t \rightharpoonup \mu_t$ uniformly in time, for some Lipschitz curve μ. In the same way, (6.7) implies that $\{\widetilde{\phi}^n\}_{n\in\mathbb{N}}$ is weakly converging to ϕ, as Radon measures on $[0,1]\times\Omega$. Moreover the pair (μ,ϕ) is still solving the continuity equation. We then have to show that $\phi \ll \mu$, which implies that ϕ disintegrates as $\phi = \int \phi_t\, dt$, and that $\mathcal{G}_\alpha$ is l.s.c. along this (possibly reparametrized) infimizing sequence.

The first fact is a simple consequence of the l.s.c. of the Benamou-Brenier functional: indeed, we have

$$\int\limits_{[0,1]\times\Omega} \left|\frac{d\phi}{d\mu}\right|^{1/\alpha} d\mu \leq \liminf_{n\to\infty} \int\limits_{[0,1]\times\Omega} \left|\frac{d\widetilde{\phi}^n}{d\widetilde{\mu}^n}\right|^{1/\alpha} d\widetilde{\mu}^n$$

and the right-hand side is finite, again thanks to (6.6) and the reparametrization we have chosen. But the finiteness of this functional then implies that $\phi \ll \mu$, thus $\phi = \int \phi_t\, dt$ with $\phi_t \ll \mu_t$.

The proof of the l.s.c. property is a litte bit more involved and we just give some hints, referring the interested reader to [13] for more details: we start considering the familiy of positive measures $\{\mathfrak{m}^n\}_{n\in\mathbb{N}}$ defined on the compact space $[0,1]\times\Omega$ by

$$\mathfrak{m}^n = \int_0^1 \sum |\widetilde{v}^n_t(x_{i(t)})|\, \widetilde{\mu}_t(\{x_{i(t)}\})^\alpha\, \delta_{x_{i(t)}}\, dt,$$

whose mass coincide with our $\alpha-$energy by definition, i.e.

$$\mathfrak{m}^n([0,1]\times\Omega) = \mathcal{G}_\alpha(\widetilde{\mu}^n, \widetilde{\phi}^n) = \mathcal{G}_\alpha(\mu^n, \phi^n).$$

Since we are assuming that these energies are equi-bounded, we then obtain that $\mathfrak{m}^n \rightharpoonup \mathfrak{m}$, for a certain positive measure $\mathfrak{m}$ on $[0,1]\times\Omega$, so that

$$\lim_{n\to\infty} \mathcal{G}_\alpha(\mu^n,\phi^n) = \lim_{n\to\infty} \mathcal{G}_\alpha(\widetilde{\mu}^n,\widetilde{\phi}^n) = \lim_{n\to\infty} \mathfrak{m}^n([0,1]\times\Omega) = \mathfrak{m}([0,1]\times\Omega).$$

Moreover, the measures $\mathfrak{m}^n$ are such that their time marginals are given by

$$g_\alpha\left(\left|\frac{d\widetilde{\phi}^n_t}{d\widetilde{\mu}^n_t}\right|^{1/\alpha} \widetilde{\mu}^n_t\right) \cdot \mathscr{L}^1 \llcorner [0,1],$$

that is they are absolutely continuous w.r.t. the 1−dimensional Lebesgue measure and with equi-bounded, in L^∞ again thanks to our reparametrization, densities. This implies that the limit measure $\mathfrak{m}$ admits the disintegration $\mathfrak{m} = \int \mathfrak{m}_t \, dt$. To conclude, it is then enough to show that

$$\mathfrak{m}_t(\Omega) \geq g_\alpha \left(\left| \frac{d\phi_t}{d\mu_t} \right|^{1/\alpha} \mu_t \right), \qquad \text{for } \mathscr{L}^1-\text{a.e. } t \in [0,1].$$

This is done showing that for every Borel set $S \subset \Omega$ we have $\mathfrak{m}_t(S) \geq |\phi_t|(S)\mu_t(S)^{\alpha-1}$ and that ϕ_t is atomic for $\mathscr{L}^1-$a.e. $t \in [0,1]$. In this way, on taking $S = \Omega$ and using these two informations, one would conclude. □

Some words about the semicontinuity properties of the energy $\mathcal{G}_\alpha$ are in order.

Remark 6.6. It is not difficult to see that the previous proof shows the l.s.c. of the functional $\mathcal{G}_\alpha$ with respect to the following kind of convergence (in [13] the terminology τ-*convergence* is used):

$$(\mu^n, \phi^n) \overset{\tau}{\rightharpoonup} (\mu, \phi) \Longleftrightarrow \begin{cases} (\tau_1) & (\mu^n, \phi^n) \rightharpoonup (\mu, \phi) \text{ as measures on} [0,1] \times \Omega \\ (\tau_2) & (\mu^n, \phi^n) \text{ solve the continuity equation} \\ (\tau_3) & t \mapsto g_\alpha \left(\left| \frac{d\phi^n_t}{d\mu^n_t} \right|^{1/\alpha} \mu^n_t \right) \text{ are equi-integrable} \end{cases}$$

On the other hand, we remark that the $*$-weak convergence only of the pairs (μ, ϕ) as measures on $[0,1] \times \Omega$ does not directly imply the lower semicontinuity of $\mathcal{G}_\alpha$, since the functional is not jointly convex. Finally, it is clear that, assuming the following stronger convergence

$$(\mu^n_t, \phi^n_t) \rightharpoonup (\mu_t, \phi_t), \quad \text{for } \mathscr{L}^1-\text{a.e. } t \in [0,1], \tag{6.8}$$

then the desired semicontinuity property of $\mathcal{G}_\alpha$ would have resulted from a simple application of Fatou's Lemma: anyway, the problem has not enough coercivity properties to guarantee such a strong convergence.

After the existence result of Theorem 6.5, one may be not content and ask whether the dynamical Eulerian description (6.5) is equivalent to the ones already existing or not: the answer is positive, as shown in [13, Theorem 4.2].

Theorem 6.7 (Equivalent descriptions)**.** *Let* $\rho_0, \rho_1 \in \mathscr{P}(\Omega)$, *then the minimization problem* (6.3) *is equivalent to* (6.5)

Proof. One fundamental tool in the proof is the *superposition principle* (see [2, Theorem 8.2.1]), which enables to pass from curves of measures to measures on the space of curves. This also permits to compare the transiting mass and the velocity terms in our approach, i.e.

$$\mu_t(\{x\}) \quad \text{and} \quad v_t(x),$$

with the corresponding quantities in the Lagrangian approach, given by

$$[\sigma(t)]_Q \quad \text{and} \quad |\sigma'(t)|.$$

Indeed, if (μ, ϕ) is optimal for (6.5), then $\phi = v \cdot \mu$ and (μ, v) solve the continuity equation $\partial_t \mu_t + \mathrm{div}_x(\mu_t v_t) = 0$. Thanks to the aforementioned superposition principle, the curve μ can be realized as a superposition of integral curves of the velocity filed v, that is $\mu_t = (e_t)_\# Q$ with Q concentrated on the solutions of $\sigma'(t) = v_t(\sigma(t))$. In this way, we obtain the following estimate for the transiting mass

$$\mu_t(\{x\}) = Q(\{\sigma : \sigma(t) = x\}) \leq Q(\{\sigma : \sigma([0,1]) \ni x\}),$$

and exchanging the order of integration in the definition of E_α and using the definition of push-forward we arrive at

$$\begin{aligned}\int_{\mathcal{L}} \int_0^1 [\sigma(t)]_Q^{\alpha-1} \, |\sigma'(t)| \, dt \, dQ(\sigma) &= \int_0^1 \int_{\mathcal{L}} [\sigma(t)]_Q^{\alpha-1} \, |v_t(\sigma(t))| \, dQ(\sigma) \, dt \\ &\leq \int_0^1 \int_{\Omega} \mu_t(\{x\})^{\alpha-1} \, |v_t(x)| \, d\mu_t(x) \, dt.\end{aligned}$$

This enables to give the first estimate

$$\mathfrak{B}_\alpha(\rho_0, \rho_1) \geq \mathcal{E}_\alpha(\rho_0, \rho_1).$$

The inequality sign is due to the fact that the term $\mu_t(\{x\})$ is local in time, while $[\sigma(t)]_Q$ is not: in particular, the functional E_α counts the mass passing from a certain point x *at any time*, a quantity which in principle could be strictly larger than the mass transiting from x at a given time t. In general we then have $\mu_t(\{x\})^{\alpha-1} \geq |x|_Q^{\alpha-1}$. The question of equivalence is then connected to the question of existence of optimal traffic plans which are *synchronized* , i.e. such that

$$Q(\{\widetilde{\sigma} : \widetilde{\sigma}([0,1]) \ni \sigma(t)\}) = Q(\{\widetilde{\sigma} : \widetilde{\sigma}(t) = \sigma(t)\}).$$

One knows this to be true when the initial measure ρ_0 is a finite sum of Dirac masses, thanks to the results contained in [9]: in this case, taking such a traffic plan Q, we construct a curve μ just setting $\mu_t := (e_t)_\# Q$. Defining $\phi_t = v_t \cdot \mu_t$ with the velocity field v_t given by

$$v_t(x) = \int_{\{\sigma\,:\,\sigma(t)=x\}} \sigma'(t)\, dQ^t_x(\sigma),$$

where $\{Q^t_x\}_{x\in\Omega}$ is the disintegration of Q with respect to the function e_t (we used this argument in the proof of Theorem 3.1), one would obtain

$$\mathcal{G}_\alpha(\mu,\phi) \le E_\alpha(Q) = \mathcal{E}_\alpha(\rho_0,\rho_1),$$

thus concluding the proof of the equivalence, at least when ρ_0 is finitely atomic. For the general case, one can use an approximation argument: indeed we have $\mathcal{E}_\alpha(\rho_0,\rho_1) = d_\alpha(\rho_0,\rho_1)$ and thanks to the relaxed formulation of Xia's model, one can always find sequences of finitely atomic measures $\rho_i^n \rightharpoonup \rho_i$, $i = 0,1$, such that $d_\alpha(\rho_0^n,\rho_1^n) \to d_\alpha(\rho_0,\rho_1)$ (see [13] for more details). □

A joint application of Theorems 6.5, 6.7 and 6.4 easily implies the following.

Corollary 6.8. *The distance d_α is a dynamical transport distance, with*

$$d_\alpha(\rho_0,\rho_1) = \min\left\{ \int_0^1 g_\alpha\left(\left|\frac{d\phi_t}{d\mu_t}\right|^{1/\alpha}\mu_t\right) dt \;:\; \begin{array}{c}\partial_t\mu_t + \mathrm{div}_x\phi_t = 0,\\ \mu_i = \rho_i,\ i=0,1\end{array}\right\}.$$

It is interesting to notice that, as enligthen by the previous formulation, d_α shares some aspects with the case of w_1 (the 1−homogeneity with respect to the velocity, for example), but also with the case of $w_{1/\alpha}$, which comes from the strictly convex cost $c(x,y) = |x-y|^{1/\alpha}$. In particular, the teleport phenomenon (see Remark 3.3) does not occur and moreover

$$d_\alpha \simeq \sum m^\alpha \ell \quad \text{and} \quad w_{1/\alpha} \simeq \left(\sum m\ell^{1/\alpha}\right)^\alpha,$$

that is the two quantities have the same scaling. Actually we can say even more on the comparison between these two distances: indeed, they are topologically equivalent.

Proposition 6.9 (Proof)**.** *Let $\alpha \in (1-1/N,1)$, then*

$$w_{1/\alpha}(\rho_0,\rho_1) \le d_\alpha(\rho_0,\rho_1) \le C\, w_{1/\alpha}(\rho_0,\rho_1)^{N(\alpha-1)+1},\ \ \rho_0,\rho_1 \in \mathscr{P}(\Omega), \tag{6.9}$$

with C depending only on α, diam(Ω) *and N.*

Proof. The first inequality is just an easy consequence of (6.6), while the second one can be proven by approximating the measures with Dirac masses centered on dyadic cubes and then using the triangular inequality, together with the following basic fact: given two discrete measures $\eta_0, \eta_1 \in \mathscr{P}(\Omega)$ having n and s atoms respectively, then there exists an optimal transport plan $\gamma \in \Pi(\eta_0, \eta_1)$ that does not move more that $n+s-1$ atoms (see [13] for more details). □

The exponent $N(\alpha-1)+1$ is strictly less than 1 and it is sharp. Indeed, one can construct (see [34, Example 0.1]) two sequences of discrete probability measures $\{\rho_0^n\}_{n\in\mathbb{N}}$ and $\{\rho_1^n\}_{n\in\mathbb{N}}$ such that

$$d_\alpha(\rho_0^n, \rho_1^n) = \frac{c}{n^{N(\alpha-1)+1}} \quad \text{and} \quad w_{1/\alpha}(\rho_0^n, \rho_1^n) = \frac{c}{n}.$$

Remark 6.10. We remark that in the inequality

$$d_\alpha(\rho_0, \rho_1) \leq C\, w_{1/\alpha}(\rho_0, \rho_1)^{N(\alpha-1)+1},$$

we can replace the distance $w_{1/\alpha}$ with w_1, keeping the same exponent $N(\alpha-1)+1$. This latter (sharper) inequality, which implies the former thanks to the fact that $w_1 \leq w_{1/\alpha}$, has been proven by Morel and Santambrogio (see [34]).

Remark 6.11. An alternative dynamical formulation of branched transport, still based on curves of measures, can be found in [12]. Here the use of the continuity equation is avoided and the problem is set as a kind of minimal length problem: branched transport is described through a curve of measures minimizing a weighted-length functional, the weight being given by the same functional g_α as before, i.e. one considers the problem

$$\min \left\{ \int_0^1 g_\alpha(\mu_t)\, |\mu_t'|_{w_p}\, dt \ :\ \mu \in \mathrm{Lip}([0,1]; \mathcal{W}_p(\Omega)),\ \mu_i = \rho_i,\ i = 0,1 \right\}.$$

Anyway, this model does not describe a Gilbert-Steiner energy and is not equivalent to the others: in [14] it is shown how this weighted-length functional has to be modified, in order to obtain the equivalence with the models presented in this section.

Finally, we point out that it could be interesting to know if some connections with evolution equations are possible also in the case of branched transport: this certainly requires a better understanding of the geometry of the space $(\mathscr{P}(\Omega), d_\alpha)$, especially of its geodesics, which as in the case of the distances considered by Dolbeault, Nazaret and Savaré, are not displacement interpolations.

Acknowledgements. The author gratefully acknowledges some useful conversations with Nicola Gigli, Stefano Lisini, Bruno Nazaret and Filippo Santambrogio. This work, as well as the participation to the OTARIE Workshop "*Optimization and stochastic methods in spatially distribute information*" held in St. Petersburgh, was supported by the European Research Council under FP7, Advanced Grant n. 226234 "Analytic Techniques for Geometric and Functional Inequalities".

References

1. M. Agueh, N. Ghoussoub, and X. Kang, *Geometric inequalities via a general comparison principle for interacting gases.* — Geom. Funct. Anal. **14** (2004), 215–244.
2. L. Ambrosio, N. Gigli, and G. Savaré, *Gradient flows in metric spaces and in the space of probability measure.* Second edition. Lectures in Mathematics ETH Zürich, Birkhäuser Verlag, Basel (2008).
3. F. Andreu, V. Caselles, and J. M. Mazón, *Some regularity results on the 'relativistic' heat equation.* — J. Differential Equations **245** (2008), 3639–3663.
4. W. Beckner, *Geometric proof of Nash's inequality.* — Internat. Math. Res. Notices (1998), 67–71.
5. J.-D. Benamou, Y. Brenier, *A computational Fluid Mechanics solution to the Monge–Kantorovich mass transfer problem.* — Numer. Math. **84** (2000), 375–393.
6. M. Bernot, V. Caselles, J.-M. Morel, Traffic plans, Publ. Mat., **49** (2005), 417–451.
7. M. Bernot, V. Caselles, and J.-M. Morel, *The structure of branched transportation networks.* — Calc. Var. Partial Differential Equations **2** (2008), 279–317.
8. M. Bernot, V. Caselles, and J.-M. Morel, *Optimal transportation networks. Models and theory.* Lecture Notes in Math. **1955**, Springer-Verlag, Berlin (2009).
9. M. Bernot, A. Figalli, *Synchronized traffic plans and stability of optima.* — ESAIM Control Optim. Calc. Var. **14** (2008), 864–878.
10. G. Bouchitté, G. Buttazzo, *New lower semicontinuity results for nonconvex functionals defined on measures.* — Nonlinear Anal. **15** (1990), 679–692.
11. Y. Brenier, *Extended Monge–Kantorovich theory.* In: *Optimal transportation and applications*, 91–121, Lecture Notes in Math., **1813**, Springer, Berlin (2003)
12. A. Brancolini, G. Buttazzo, and F. Santambrogio, *Path functionals over Wasserstein spaces.* — J. Eur. Math. Soc. **8** (2006), 415–434.
13. L. Brasco, G. Buttazzo, F. and Santambrogio, *A Benamou–Brenier approach to branched transport.* — Submitted (2010), available at `http://cvgmt.sns.it/people/brasco`.
14. L. Brasco, F. Santambrogio, *An equivalent path functional formulation of branched transportation problems.* — to appear on Discr. Cont. Dyn. Syst. (2010), available at `http://cvgmt.sns.it/people/brasco`.

15. G. Buttazzo, *Semicontinuity, relaxation and integral representation in the calculus of variations.* Pitman Research Notes in Mathematics Series **207**, Longman Scientific & Technical, Harlow (1989).
16. G. Buttazzo, C. Jimenez, and E. Oudet, *An optimization problem for mass transportation with congested dynamics.* — SIAM J. Control Optim. **48** (2009), 1961–1976.
17. J. A. Carrillo, S. Lisini, G. Savaré, and D. Slepcěv, *Nonlinear mobility continuity equations and generalized displacement convexity.* — J. Funct. Anal. **258** (2010), 1273–1309.
18. D. Cordero-Erausquin, B. Nazaret, C. and Villani, *A mass-transportation approach to sharp Sobolev and Gagliardo–Nirenberg inequalities.* — Adv. Math. **182** (2004), 307–332.
19. C. Dellacherie, and P.-A. Meyer, *Probabilités et potentiel.* (French) Chapitres I à IV. Édition entièrement refondue. Publications de l'Institut de Mathématique de l'Université de Strasbourg, No. XV. Actualités Scientifiques et Industrielles, **1372**, Hermann, Paris (1975).
20. M. Del Pino, J. Dolbeault, *Best constants for Gagliardo–Nirenberg inequalities and applications to nonlinear diffusions.* — J. Math. Pures Appl. **81** (2002), 847–875.
21. J. Dolbeault, B. Nazaret, and G. Savaré, *A new class of transport distances between measures.* — Calc. Var. Partial Differential Equations **34** (2009), 193–231.
22. E. N. Gilbert, Minimum cost communication networks, Bell System Tech. J. **46** (1967), 2209–2227.
23. L. Gross, *Logarithmic Sobolev inequalities.* — Amer. J. Math. **97** (1975) 1061–1083.
24. T. Hillen, K. J. Painter, *A user's guide to PDE models for chemotaxis.* — J. Math. Biol. **58** (2009), 183–217.
25. R. Jordan, D. Kinderlehrer, F. and Otto, *The variational formulation of the Fokker–Planck equation.* — SIAM J. Math. Anal. **29** (1998), 1–17.
26. J. M. Lasry, P.-L. Lions, *Mean field games.* — Japan J. Math. **2** (2007), 229–260.
27. S. Lisini, *Characterization of absolutely continuous curves in Wasserstein spaces.* — Calc. Var. Partial Differential Equations **28** (2007), 85–120.
28. S. Lisini, A. Marigonda, *On a class of modified Wasserstein distances induced by concave mobility functions defined on bounded intervals.* — Manuscripta Math. **133** (2010), 197–224.
29. F. Maddalena, J. M. Morel, and S. Solimini, *A variational model of irrigation patterns.* — Interfaces Free Bound. **5** (2003), 391–415.
30. F. Maggi, C. Villani, *Balls have the worst best Sobolev inequalities.* II. *Variants and extensions.* — Calc. Var. Partial Differential Equations **31** (2008), 47–74.
31. F. Maggi, C. Villani, *Balls have the worst best Sobolev inequalities.* — J. Geom. Anal. **15** (2005), 83–121.
32. R. McCann, A convexity principle for interacting gases, Adv. Math. **128** (1997), 153–179.
33. R. McCann, M. Puel, *Constructing a relativistic heat flow by transport time steps.* — Ann. Inst. H. Poincaré Anal. Non Linéaire **26** (2009), 2539–2580.
34. J.-M. Morel, F. Santambrogio, *Comparison of distances between measures.* — Appl. Math. Lett. **20** (2007), 427–432.

35. B. Nazaret, *Best constant in Sobolev trace inequalities on the half space.* — Nonlinear Anal. **65** (2006), 1977–1985.
36. F. Otto, *The geometry of dissipative evolution equation: the porous medium equation.* — Comm. Partial Differential Equations **26** (2001), 101–174.
37. F. Otto, *Dynamics of labyrinthine pattern formation in magnetic fluids: a mean-field theory.* — Arch. Rational Mech. Anal. **141** (1998), 63–103.
38. F. Otto, C. Villani, *Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality.* — J. Func. Anal. **173** (2000), 361–400.
39. E. Paolini, E. Stepanov, *Optimal transportation networks as flat chains.* — Interfaces Free Bound. **8** (2006), 393–436.
40. G. Talenti, *Best constant in Sobolev inequality.* — Ann. Mat. Pura Appl. **110** (1976), 353–372.
41. C. Villani, *Optimal transport. Old and new.* Die Grundlehren der Mathematischen Wissenschaften **338**. Springer-Verlag, Berli, (2009).
42. C. Villani, *Optimal transportation, dissipative PDE's and functional inequalities.* In: *Optimal transportation and applications*, 53–89, Lecture Notes in Math. **1813**, Springer, Berlin (2003).
43. Q. Xia, *Optimal paths related to transport problems.* — Commun. Contemp. Math. **5** (2003), 251–279.

Laboratoire d'Analyse, Topologie et Probabilites
Universite Aix-Marseille I
39, Rue Frederic Joliot Curie
13453 Marseille Cedex 13, France
E-mail: brasco@cmi.univ-mrs.fr

Поступило 15 сентября 2011 г.