

## A BENAMOU–BRENIER APPROACH TO BRANCHED TRANSPORT\*

L. BRASCO<sup>†</sup>, G. BUTTAZZO<sup>‡</sup>, AND F. SANTAMBROGIO<sup>§</sup>

**Abstract.** The problem of branched transportation aims to describe the movement of masses when, due to concavity effects, they have the impulse to travel together as much as possible, because the cost for a path of length  $\ell$  covered by a mass  $m$  is proportional to  $m^\alpha \ell$  with  $0 < \alpha < 1$ . The optimization of this criterion let branched structures appear and is suitable to applications like road systems, blood vessels, river networks, etc. Several models have been employed in the literature to present this transport problem, and the present paper looks at a dynamical model similar to the celebrated Benamou–Brenier formulation of Kantorovich optimal transport. The movement is represented by a path  $\rho_t$  of probabilities connecting an initial state  $\mu_0$  to a final state  $\mu_1$  and satisfying the continuity equation  $\partial_t \rho + \operatorname{div}_x q = 0$  together with a velocity field  $v$  (with  $q = \rho v$  being the momentum). The transportation cost to be minimized is nonconvex and finite on atomic measures:  $\int_0^1 (\int_\Omega \rho^{\alpha-1} |q| d\#(x)) dt$ .

**Key words.** optimal transport, branched transport, continuity equation, functionals on spaces of measures, optimal networks

**AMS subject classifications.** 49J45, 49Q20, 49Q10, 90B18, 60K30

**DOI.** 10.1137/10079286X

**1. Introduction.** The optimal mass transportation theory corresponds to the study of transporting a given mass distribution  $\mu_0$  on  $\Omega$  (that we assume to be a compact and convex subset of  $\mathbb{R}^d$ ) into a final configuration  $\mu_1$ , by minimizing the total transportation cost, the latter being suitably defined: clearly,  $\mu_0$  and  $\mu_1$  are required to satisfy the mass balance condition  $\int_\Omega d\mu_0 = \int_\Omega d\mu_1$ . From now on, we will assume that they are normalized to be probability measures. The cost for moving a unit mass from a position  $x$  to a position  $y$  is taken equal to  $c(x, y)$ , a function a priori given, which determines the nature of the problem and provides the total minimal cost

$$(1.1) \quad C(\mu_0, \mu_1) = \min \left\{ \int_{\Omega \times \Omega} c(x, y) d\gamma(x, y) : \gamma \in \Gamma(\mu_0, \mu_1) \right\},$$

where  $\Gamma(\mu_0, \mu_1)$  is the class of admissible *transport plans*, i.e., probabilities on the product space  $\Omega \times \Omega$  having first and second marginals given by  $\mu_0$  and  $\mu_1$ , respectively.

---

\*Received by the editors April 20, 2010; accepted for publication (in revised form) February 17, 2011; published electronically April 19, 2011. This work was supported by the *Agence Nationale de la Recherche* via the research project OTARIE, the *Université Franco-Italienne* via the mobility program *Galilée* “Allocation et Exploitation et Evolution Optimales des Ressources: réseaux, points et densités, modèles discrets et continus,” and the *Università di Pisa* through the program *Cooperazione Accademica Internazionale* “Optimal transportation and related topics.”

<http://www.siam.org/journals/sima/43-2/79286.html>

<sup>†</sup>Dipartimento di Matematica e Applicazioni, Università di Napoli “Federico II,” Complesso di Monte S. Angelo, Via Cintia, 80126 Napoli, Italy (lorenzo.brasco@unina.it). This author was partially supported by the ERC Advanced grant 226234 “Analytic Techniques for Geometric and Functional Inequalities.”

<sup>‡</sup>Dipartimento di Matematica, Università di Pisa, Largo B. Pontecorvo 5, 56127 Pisa, Italy (buttazzo@dm.unipi.it).

<sup>§</sup>Département de Mathématiques, Bât. 425, Faculté des Sciences, Université Paris-Sud 11, 91405 Orsay Cedex, France (filippo.santambrogio@math.u-psud.fr).

The cases  $c(x, y) = |x - y|^p$  with  $p \geq 1$  in particular have been studied, and the cost  $C(\mu_0, \mu_1)$  in (1.1) provides, through the relation

$$W_p(\mu_0, \mu_1) = (C(\mu_0, \mu_1))^{1/p},$$

the so-called *Wasserstein distance*  $W_p$ , which metrizes the weak\* convergence on the space of probabilities  $\mathcal{P}(\Omega)$ . A very extensive literature on the subject is available; we simply mention the books [2, 26, 25] where one can find a complete list of references.

Thanks to the fact that the space  $\mathcal{W}_p(\Omega)$  of probability measures endowed with these distances turns out to be a *geodesic space*, dynamical models for optimal transportation are of particular interest. Being a geodesic space means that the distance between two points is always equal to the infimum of the lengths of the curves connecting these points, and that this infimum is actually a minimum:

$$W_p(\mu_0, \mu_1) = \min \left\{ \int_0^1 |\rho'_t|_{W_p} dt : \rho \in \text{Lip}([0, 1]; \mathcal{W}_p(\Omega)), \rho_0 = \mu_0, \rho_1 = \mu_1 \right\},$$

where  $|\rho'|_{W_p}$  is the *metric derivative* of the measure-valued Lipschitz curve  $\rho$ , defined as

$$|\rho'_t|_{W_p} = \lim_{h \rightarrow 0} \frac{W_p(\rho_{t+h}, \rho_t)}{|h|}$$

(we refer the reader to [2] for more details).

Since the curves connecting two points of this space are actually curves of measures, they can be described through the so-called *continuity equation*: it is well known (see [2, Theorem 8.3.1]) that for every Lipschitz or absolutely continuous curve  $\rho_t$  in the space  $\mathcal{W}_p(\Omega)$  ( $p > 1$  for simplicity), there exists a map  $q$  from  $[0, 1]$  into the space of vector-valued measures, such that  $q_t \ll \rho_t$  (hence  $q_t = v_t \cdot \rho_t$ , with  $v$  being the velocity vector), which represents the flux  $q = \rho v$  and satisfies

$$(1.2) \quad \partial_t \rho + \text{div}_x q = 0 \quad \text{and} \quad \|v_t\|_{L^p(\rho_t)} = |\rho'_t|_{W_p}.$$

(The degenerate case  $p = 1$  is a little bit more involved since  $q_t \ll \rho_t$  is no longer guaranteed and the  $L^1$ -norm has to be replaced by the mass of the measure  $q_t$ ; see [1].)

On the other hand, every time that we have a pair  $(\rho, q)$  satisfying  $\partial_t \rho + \text{div}_x q = 0$  with  $q \ll \rho$ , so that  $q_t = v_t \cdot \rho_t$ , we can infer that  $|\rho'_t|_{W_p} \leq \|v_t\|_{L^p(\rho_t)}$ . This means that one can minimize the functional

$$(1.3) \quad \mathcal{A}_p(\rho, q) := \begin{cases} \int_0^1 \left( \int_{\mathbb{R}^d} |v_t|^p d\rho_t \right) dt & \text{if } q \ll \rho \text{ and } q_t = v_t \cdot \rho_t, \\ +\infty & \text{otherwise,} \end{cases}$$

which is nothing but the integral in time of the kinetic energy when  $p = 2$ , and the cost in (1.1) can be recovered through the equality

$$C(\mu_0, \mu_1) = \min \left\{ \int_0^1 \left( \int_{\mathbb{R}^d} \left| \frac{dq_t}{d\rho_t}(x) \right|^p d\rho_t(x) \right) dt : \begin{array}{l} \partial_t \rho + \text{div}_x q = 0, \\ \rho_0 = \mu_0, \rho_1 = \mu_1 \end{array} \right\}.$$

The problem above is the one that was proposed by Benamou and Brenier in [3] as a dynamical version of optimal transportation. It has the advantage that it is the minimization of a convex functional of  $\rho$  and  $q$ , under linear constraints.

Other variants of mass transportation problems have been studied and can be expressed in this way by considering in (1.3) other *convex* functions of the pair  $(\rho, q)$ . Recently, Dolbeault, Nazaret, and Savaré introduced in [18] new classes of distances over  $\mathcal{P}(\mathbb{R}^d)$  based on the minimization of the following functional (where  $\lambda$  is a given reference measure on  $\mathbb{R}^d$ , and  $\rho$  and  $q$  are identified with their densities w.r.t.  $\lambda$ ):

$$\int_0^1 \left( \int_{\mathbb{R}^d} \Phi(\rho, q) d\lambda \right) dt, \quad \text{where } \Phi(\rho, q) = \frac{|q|^p}{h(\rho)^{p-1}} = \left( \frac{|q|}{h(\rho)} \right)^p h(\rho), \quad p \geq 1,$$

which are connected to the nonlinear mobility continuity equation

$$\partial_t \rho + \operatorname{div}_x (h(\rho) v) = 0.$$

(A treatment of the limiting case  $h(\rho) \equiv 1$ , corresponding to consider  $\Phi(\rho, q) = |q|^p$ , can also be found in [13].) If the function  $h$  is concave (for example,  $h(\rho) = \rho^\beta$ , with  $\beta \in [0, 1]$ ), then this problem turns out to be convex as well. The main interest that motivated Dolbeault et al. to study these distances lies in the possible applications to diffusion equations of the type of the nonlinear mobility continuity equation  $\partial_t \rho + \operatorname{div}_x (h(\rho) v) = 0$  above, where the vector field  $v$  depends on  $\rho$  in such a way that the equation can be interpreted as a gradient flow of a given functional w.r.t. this new family of dynamical distances. Moreover, the equations of the geodesics are similar to a *mean-field game system*; see [20].

In connection to *congestion* effects and crowd motion, other models include penalizations on high densities: in [15] the case

$$\Phi(\rho, q) = \frac{|q|^p}{\rho^{p-1}} + c\rho^2, \quad p \geq 1, \quad c \geq 0,$$

has been considered as a model for crowd motion in a congested situation (for instance, in case of panic). This problem as well is convex.

A completely different situation occurs in the opposite case of congestion, when *concentration* effects are present and the mass has the impulse to travel together as much as possible, in order to save part of the cost. This happens very often in many applications, as discovered by Gilbert, who in [19] formulated a mathematical model for the transportation of signals along telephone cables. More recently, Gilbert’s model has been refined and considered in the framework of mass transportation, under the name of *branched transport*, to emphasize the fact that transport rays may bifurcate. All of these models have in common the fact that the cost for a mass  $m$  moving on a path of length  $\ell$  is proportional to  $m^\alpha \ell$  ( $0 < \alpha < 1$ , so that  $(m_1 + m_2)^\alpha < m_1^\alpha + m_2^\alpha$ ). In [4, 5, 6, 7, 22], for every  $0 < \alpha < 1$  a transportation cost from  $\rho_0$  to  $\rho_1$  is considered through the suitable use of probabilities defined on spaces of curves in  $\Omega$ , with [22] (the so-called *irrigation patterns model*) dealing with the case of a single source  $\rho_0 = \delta_{x_0}$ . See section 4 for a glance at the details of these models and their formulations. On the other hand, the model of [28] can be seen as the natural extension of the original Gilbert model and uses vector measures having prescribed divergence  $\rho_0 - \rho_1$  (see also [27]): these vector measures are the continuous generalization of the finite weighted and oriented graphs that were present in Gilbert’s original formulation.

A first attempt to obtain a dynamical formulation of branched transportation through curves of measures was made in [10] and later refined in [11, 12]: in these

papers the starting point is the geodesic formulation of the Wasserstein distance, where the length functional is modified to consider an energy of the type

$$\int_0^1 g(\rho_t) |\rho'_t|_{W_p} dt.$$

The weight function  $g$  is a local term of the moving mass, forcing the mass to concentrate and thus giving rise to branching phenomena.

These models are not satisfactory yet, because they are in general not equivalent to those by Gilbert [19], Xia [27, 28], Maddalena and colleagues [22, 23], or Bernot and colleagues [4, 5, 6, 7]. An attempt to perform some modifications in the functionals defined on curves of measures so as to obtain equivalence with the other models has been made in [11, 12], where on the other hand some quite involved distinction between moving mass and still mass has been done. In all of these models, the branched transportation is studied avoiding the Benamou–Brenier approach consisting of the minimization of a suitable cost  $\mathcal{F}(\rho, q)$  under the constraint of the continuity equation  $\partial_t \rho + \operatorname{div}_x q = 0$ , which we believe *is the most natural for these kinds of problems*. The only approach to dynamical branched transportation using the continuity equation is, as far as we know, that of [9]. Yet, to prove semicontinuity and hence existence, even in this model, the problem is reduced to the minimization of a functional of the form

$$\int \theta^\alpha d\mathcal{H}^1(x, t)$$

(which is the motivation of Xia in [28]), and the dynamical features are not completely exploited.

In the present paper we follow a more direct approach: for all pairs  $(\rho, q)$  verifying the continuity equation, with  $\rho_0 = \mu_0$  and  $\rho_1 = \mu_1$ , we define a functional  $\mathcal{F}(\rho, q)$  and show that this functional is both lower semicontinuous and coercive w.r.t. a suitable convergence on  $(\rho, q)$ , and this directly provides the existence of an optimal dynamical path. The paper is organized as follows:

- (i) in section 2 we give the precise setting and state the main results;
- (ii) section 3 is devoted to the proofs giving the existence of an optimal path  $\rho_t$ ;
- (iii) in section 4 we show that our model is equivalent to the other models of branched transportation available in the literature, comparing it to the traffic plan model of [4], which is one of the most flexible (and anyway equivalent to the others, as shown in [6, Chapter 9]);
- (iv) in the appendix we deal with some inequalities involving Wasserstein distances and branched distances, that is, distances over the space of probabilities given by the minima of some branched transportation problems. These inequalities have already been studied in [24] and [17], but some very precise issues concerning  $d_\alpha$  and  $W_{1/\alpha}$  are very close to the topics of this paper and deserve to be examined here. New and simpler proofs are provided.

**2. Problem setting and main results.** In this section we fix the notation and state the main results of the paper. In what follows,  $\Omega$  will denote a given subset of  $\mathbb{R}^d$ , where all the mass dynamics will take place; for the sake of simplicity we assume that  $\Omega$  is convex and compact. The space  $\mathcal{P}(\Omega)$  of all Borel probabilities on  $\Omega$  can then be endowed with the weak\* convergence, which is metrized by the Wasserstein distances (see the introduction). In the following, we will also use the notation  $\mathcal{M}(\Omega; \mathbb{R}^d)$  to indicate the space of  $\mathbb{R}^d$ -valued Radon measures over  $\Omega$ , while  $\mathcal{L}^k$  will indicate the  $k$ -dimensional Lebesgue measure.

The main objects to consider will be the pairs  $(\rho, q)$  with

$$(2.1) \quad \rho \in C([0, 1]; \mathcal{P}(\Omega)), \quad q \in L^1([0, 1]; \mathcal{M}(\Omega; \mathbb{R}^d))$$

satisfying the *continuity equation* formally written as follows (here  $\nu$  stands for the outer normal vector to  $\partial\Omega$ ):

$$(2.2) \quad \begin{cases} \partial_t \rho + \operatorname{div}_x q = 0 & \text{in } [0, 1] \times \Omega, \\ q \cdot \nu = 0 & \text{on } [0, 1] \times \partial\Omega. \end{cases}$$

Its precise meaning is given in the sense of distributions; that is,

$$(2.3) \quad \int_0^1 \left[ \int_{\Omega} \partial_t \phi(t, x) d\rho_t(x) + \int_{\Omega} D_x \phi(x, t) \cdot dq_t(x) \right] dt = 0$$

for every smooth function  $\phi$  with  $\phi(0, x) = \phi(1, x) = 0$ .

DEFINITION 2.1. We denote by  $\mathfrak{D}$  the set of all pairs  $(\rho, q)$  satisfying (2.1) and (2.3). Moreover, given  $\mu_0, \mu_1 \in \mathcal{P}(\Omega)$ , we define the set  $\mathfrak{D}(\mu_0, \mu_1)$  of admissible configurations connecting  $\mu_0$  to  $\mu_1$  as

$$\mathfrak{D}(\mu_0, \mu_1) = \{(\rho, q) \in \mathfrak{D} : \rho_0 = \mu_0, \rho_1 = \mu_1\}.$$

The velocity vector  $v$  can be defined as the Radon–Nikodym derivative of the vector measure  $q$  w.r.t.  $\rho$ :

$$v = \frac{dq}{d\rho}.$$

Among all pairs  $(\rho, q) \in \mathfrak{D}$  satisfying the continuity equation above, we consider a cost function  $\mathcal{F}(\rho, q)$  of the form

$$(2.4) \quad \mathcal{F}(\rho, q) = \int_0^1 F(\rho_t, q_t) dt,$$

where  $F$  is defined through

$$F(\rho, q) := \begin{cases} G_{\alpha}(|v|^{1/\alpha} \cdot \rho) & \text{if } q = v \cdot \rho, \\ +\infty & \text{if } q \text{ is not absolutely continuous w.r.t. } \rho, \end{cases}$$

and  $G_{\alpha}$  ( $0 < \alpha < 1$ ) is a functional defined on measures of the kind studied by Bouchitté and Buttazzo in [8]:  $G_{\alpha}(\lambda) = +\infty$  if  $\lambda$  is not purely atomic, while

$$G_{\alpha}(\lambda) = \int_{\Omega} |\lambda(\{x\})|^{\alpha} d\#(x) = \sum_{i \in \mathbb{N}} |\lambda_i|^{\alpha} \quad \text{if } \lambda = \sum_{i \in \mathbb{N}} \lambda_i \delta_{x_i}$$

( $\#$  stands for the counting measure). In this way our functional  $\mathcal{F}$  becomes

$$\mathcal{F}(\rho, q) = \int_0^1 \left[ \int_{\Omega} |v_t(x)| \rho_t(\{x\})^{\alpha} d\#(x) \right] dt = \int_0^1 \left[ \sum_{i \in \mathbb{N}} |v_{t,i}| \rho_{t,i}^{\alpha} \right] dt, \quad (\rho, q) \in \mathfrak{D},$$

and the dynamical model for branched transport that we consider is

$$(2.5) \quad \mathcal{B}_{\alpha}(\mu_0, \mu_1) := \min_{(\rho, q) \in \mathfrak{D}(\mu_0, \mu_1)} \mathcal{F}(\rho, q).$$

Our main goal is to show that the minimization problem (2.5) above admits a solution. This will be obtained through the direct methods of the calculus of variations, consisting of proving lower semicontinuity and coercivity of the problem under consideration, w.r.t. a suitable convergence.

*Remark 2.2.* It is easy to see that the weak\* convergence of the pairs  $(\rho, q)$  does not directly imply the lower semicontinuity in (2.5), since the functional is not jointly convex. On the other hand, if  $(\rho^n, q^n) \in \mathfrak{D}$  and we assume

$$(\rho_t^n, q_t^n) \rightharpoonup (\rho_t, q_t) \quad \text{for } \mathcal{L}^1 \text{ a.e., } t \in [0, 1],$$

then a simple application of Fatou’s lemma would lead to the desired semicontinuity property of  $\mathcal{F}$  (because one could prove that  $F$  is a lower semicontinuous functional on measures, as a consequence of the semicontinuity of  $G_\alpha$  and of the convexity of  $(x, y) \mapsto |x|^p/y^{p-1}$ ).

In order to prove a semicontinuity result in the easiest possible way, we will introduce a convergence that is stronger than the weak convergence of measures on  $[0, 1] \times \Omega$ , but weaker than the weak convergence for every fixed time  $t$ . This convergence will be compatible with the compactness we can infer from our variational problem.

**DEFINITION 2.3.** *A sequence  $(\rho^n, q^n)$   $\tau$ -converges to  $(\rho, q)$  if  $(\rho^n, q^n) \rightharpoonup (\rho, q)$  in the sense of measures and*

$$(2.6) \quad \text{the maps } t \mapsto F(\rho_t^n, q_t^n) \text{ are equi-integrable.}$$

**THEOREM 2.4** (coercivity). *Let  $(\rho^n, q^n)$  be a sequence such that  $\mathcal{F}(\rho^n, q^n) \leq C$ ; then, up to a time reparametrization,  $(\rho^n, q^n)$  is  $\tau$ -compact.*

**THEOREM 2.5** (lower semicontinuity). *Let  $(\rho^n, q^n) \in \mathfrak{D}$  be a sequence which  $\tau$ -converges to  $(\rho, q)$ . Then*

$$\mathcal{F}(\rho, q) \leq \liminf_{n \rightarrow \infty} \mathcal{F}(\rho^n, q^n).$$

As a consequence we obtain the following existence result.

**THEOREM 2.6** (existence). *For every  $\mu_0, \mu_1 \in \mathcal{P}(\Omega)$ , the minimization problem (2.5) admits a solution.*

*Remark 2.7.* We point out that, for some choices of the data  $\mu_0, \mu_1$  and the exponent  $\alpha$ , the statement of Theorem 2.6 could be empty, because the functional  $\mathcal{F}$  could be constantly  $+\infty$  on every admissible path  $(\rho, q)$  joining  $\mu_0$  to  $\mu_1$ . This issue will be solved in section 4, where the equivalence to other variational models for branched transportation will be proved. Since for these models finiteness of the minima has been widely investigated, we can infer, for instance, that if  $\alpha > 1 - 1/d$ , then every pair  $\mu_0$  and  $\mu_1$  can be joined by a path of finite energy. On the other hand, if  $\alpha \leq 1 - 1/d$ ,  $\mu_0 = \delta_{x_0}$ , and  $\mu_1$  is absolutely continuous w.r.t.  $\mathcal{L}^d$ , then there are no finite energy paths connecting them.

**3. Proofs.** A preliminary inequality to all the proofs is the following: if  $q \ll \rho$ , then  $q_t = v_t \cdot \rho_t$ , and

$$(3.1) \quad \begin{aligned} F(\rho_t, q_t) &= \sum_i \rho_t(\{x_i\})^\alpha |v_t(x_i)| = \sum_i \left( \rho_t(\{x_i\}) |v_t(x_i)|^{1/\alpha} \right)^\alpha \\ &\geq \left( \sum_i \rho_t(\{x_i\}) |v_t(x_i)|^{1/\alpha} \right)^\alpha = \|v_t\|_{L^{1/\alpha}(\rho_t)}^\alpha \end{aligned}$$

due to the subadditivity of the function  $x \mapsto x^\alpha$ . This inequality and its consequences will be discussed in the appendix as well. In particular it also follows that

$$(3.2) \quad \int_0^1 F(\rho_t, q_t) dt \geq \int_0^1 \|v_t\|_{L^1(\rho_t)} dt = \int_0^1 |q_t|(\Omega) dt = |q|([0, 1] \times \Omega).$$

**3.1. Proof of Theorem 2.4.** Due to the fact that the functional  $\mathcal{F}$  is 1-homogeneous in the velocity, it is clear that reparametrizations in time do not change the values of  $\mathcal{F}$ . By reparametrization, we mean replacing a pair  $(\rho, q)$  with a new pair  $(\tilde{\rho}, \tilde{q})$  of the form  $\tilde{\rho}_t = \rho_{\varphi(t)}$ ,  $\tilde{q}_t = \varphi'(t)q_{\varphi(t)}$  (which equivalently means that  $\tilde{q}$  is the image measure of  $q$  through the inverse of the map  $(t, x) \mapsto (\varphi(t), x)$ ). Thanks to this invariance, if  $(\rho^n, q^n)$  is such that  $\mathcal{F}(\rho^n, q^n) \leq C$ , then one can define a new pair  $(\tilde{\rho}^n, \tilde{q}^n)$ , with

$$F(\tilde{\rho}_t^n, \tilde{q}_t^n) = \mathcal{F}(\tilde{\rho}^n, \tilde{q}^n) = \mathcal{F}(\rho^n, q^n) \leq C \text{ for every } t,$$

which in particular implies that this new sequence  $(\tilde{\rho}^n, \tilde{q}^n)$  satisfies condition (2.6). After that, we need to prove compactness for the weak convergence of measures on  $[0, 1] \times \Omega$ , a fact that requires only bounds on the total variation of  $\tilde{\rho}^n$  and  $\tilde{q}^n$ . The bound on  $\tilde{\rho}^n$  is straightforward, since for every  $t$  the measure  $\tilde{\rho}_t^n$  is a probability, while for  $\tilde{q}^n$ , which is absolutely continuous w.r.t.  $\tilde{\rho}^n$ , it is enough to use (3.2) in order to bound the total variation of  $q$  by  $C$ .

This allows us to extract a subsequence  $(\tilde{\rho}_t^{n_k}, \tilde{q}_t^{n_k})$  that converges weakly to a pair  $(\rho, q)$ . The only nontrivial point is that we a priori restricted our attention to pairs  $(\rho, q)$ , where  $\rho \in C([0, 1]; \mathcal{P}(\Omega))$  and  $q \in L^1([0, 1]; \mathcal{M}(\Omega; \mathbb{R}^d))$ , so that we need to prove that  $\rho$  is continuous and that  $q$  is of the form  $\int q_t dt$ . Yet, the inequality (3.1) applied to the pairs  $(\tilde{\rho}^n, \tilde{q}^n)$  proves a uniform bound on the  $L^{1/\alpha}$  norm of the velocities, which implies that the curves  $\tilde{\rho}^n$  are uniformly Lipschitz continuous according to the distance  $W_{1/\alpha}$ , and this property is inherited by the limit measure  $\rho$ .

For the decomposition of  $q$ , just use the inequality (3.1), thus obtaining a uniform bound on  $\|v_t^n\|_{L^{1/\alpha}(\rho_t^n)}$ , which a fortiori gives a uniform bound on the Benamou–Brenier functional

$$\mathcal{A}_{1/\alpha}(\rho^n, q^n) = \int_0^1 \|v_t^n\|_{L^{1/\alpha}(\rho_t^n)}^{1/\alpha} dt.$$

This functional being lower semicontinuous, we can deduce the same bound at the limit: this in particular implies that  $q$  is absolutely continuous w.r.t.  $\rho$ , with an  $L^{1/\alpha}$  density. Since  $\rho$  is a measure on  $[0, 1] \times \Omega$ , which is of the form  $\int \rho_t dt$ , the same disintegration will be true for  $q$ .

This means that we have actually found an admissible pair  $(\rho, q)$ , which is the  $\tau$ -limit of  $(\tilde{\rho}_t^{n_k}, \tilde{q}_t^{n_k})$ , and the proof is complete.

**3.2. Proof of Theorem 2.5.** We consider here a sequence  $(\rho^n, q^n)$ , where  $q^n = v^n \cdot \rho^n$  (otherwise the functional  $\mathcal{F}$  would not be finite-valued), satisfying the continuity equation and such that  $(\rho^n, q^n)$   $\tau$ -converges to  $(\rho, q)$ .

First, we define a sequence of measures  $\mathbf{m}^n$  on  $[0, 1] \times \Omega$  through

$$\mathbf{m}^n = \int \left( \sum_i \rho_t^n(\{x_{i,t}\})^\alpha |v_t^n(x_{i,t})| \delta_{x_{i,t}} \right) dt,$$

where the points  $x_{i,t}$  are the atoms of  $q_t^n$  (i.e., the atoms of  $\rho_t^n$ , where the velocity  $v_t^n$  does not vanish). We notice that  $\mathcal{F}(\rho^n, q^n) = \mathbf{m}^n([0, 1] \times \Omega)$ .

In order to prove lower semicontinuity of  $\mathcal{F}$ , we first observe that thanks to condition (2.6) in the definition of  $\tau$ -convergence, the energies  $\mathcal{F}(\rho^n, q^n)$  are equi-bounded. This uniform bound implies the convergence  $\mathbf{m}^n \rightharpoonup \mathbf{m}$ , up to the extraction of a subsequence (not relabeled). It is clear that, on this subsequence, we have

$$\lim_{n \rightarrow \infty} \mathcal{F}(\rho^n, q^n) = \lim_{n \rightarrow \infty} \mathbf{m}^n([0, 1] \times \Omega) = \mathbf{m}([0, 1] \times \Omega);$$

then, in order to prove the desired semicontinuity property, it is enough to get some proper lower bounds on  $\mathbf{m}$ .

Notice that we have  $\mathbf{m}^n = \int \mathbf{m}_t^n dt$ , with  $\mathbf{m}_t^n = F(\rho_t^n, q_t^n) \cdot \mathcal{L}^1 \llcorner [0, 1]$ ; that is, the marginal of  $\mathbf{m}^n$  on the time variable is a measure with an equi-integrable  $L^1$  density. This implies that the same disintegration holds true for the limit measure  $\mathbf{m}$ ; i.e., we have  $\mathbf{m} = \int \mathbf{m}_t dt$ .

Let us fix  $M > 0$  and a closed set  $Q$ . Take the function

$$\chi_M(x) := (1 - M \operatorname{dist}(x, Q))_+, \quad x \in \Omega,$$

where  $(\cdot)_+$  stands for the positive part: observe that  $\chi_M$  is positive, takes the value 1 on  $Q$ , is  $M$ -Lipschitz, and vanishes outside a  $1/M$ -neighborhood of  $Q$ . We also fix  $\varepsilon > 0$ . Then, thanks to the equi-integrability of the maps  $t \mapsto F(\rho_t^n, q_t^n)$ , we have that there exists  $\delta > 0$  such that for every  $A \subset [0, 1]$  with  $\mathcal{L}^1(A) < \delta$ , there results

$$\sup_{n \in \mathbb{N}} \int_A F(\rho_t^n, q_t^n) dt < \varepsilon.$$

Correspondingly, we choose a time interval  $[a, b]$ , with  $(b - a) < \delta$ . Indicating by  $1_E$  the characteristic function of a generic set  $E$  (i.e., the function that takes the value 1 on  $E$  and 0 elsewhere), we consider  $\phi(t, x) = \chi_M(x)^\alpha 1_{[a, b]}(t)$ , which is upper semicontinuous on  $[0, 1] \times \Omega$ . Then we have

$$\begin{aligned} \int \phi(t, x) d\mathbf{m}(t, x) &\geq \limsup_{n \rightarrow \infty} \int \phi(t, x) d\mathbf{m}^n(t, x) \\ &= \limsup_{n \rightarrow \infty} \int_a^b \left( \sum_i \rho_t^n(\{x_i\})^\alpha |v_t^n(x_i)| \chi_M(x_i)^\alpha \right) dt, \end{aligned}$$

where the points  $x_i$  are, as before, the atoms of  $q^n$  (and we omitted the dependence on  $n$  and  $t$ ).

We then decompose the product  $\rho_t^n(\{x_i\})^\alpha \chi_M(x_i)^\alpha$  as  $(\rho_t^n(\{x_i\}) \chi_M(x_i))^{\alpha-1} \cdot (\rho_t^n(\{x_i\}) \chi_M(x_i))$  (where  $\rho_t^n \chi_M > 0$ ). Notice that  $\rho_t^n(\{x_i\}) \chi_M(x_i) \leq \int \chi_M d\rho_t^n$ . Then we can estimate the right-hand side in the previous inequality as

$$\begin{aligned} &\int_a^b \left( \sum_i \rho_t^n(\{x_i\})^\alpha |v_t^n(x_i)| \chi_M(x_i)^\alpha \right) dt \\ &\geq \int_a^b \left[ \left( \int \chi_M d\rho_t^n \right)^{\alpha-1} \times \left( \sum_i \rho_t^n(\{x_i\}) |v_t^n(x_i)| \chi_M(x_i) \right) \right] dt \\ &= \int_a^b \left( \int \chi_M d\rho_t^n \right)^{\alpha-1} \left( \int \chi_M d|q_t^n| \right) dt. \end{aligned}$$

We go on by estimating from above  $\int \chi_M d\rho_t^n$ : we have

$$\int_\Omega \chi_M(x) d\rho_t^n(x) \leq \int_\Omega \chi_M(x) d\rho_s^n(x) + M W_1(\rho_t^n, \rho_s^n),$$



which is a consequence of the definition of  $W_1$  by duality with 1-Lipschitz functions (see [26, Theorem 1.14]). To estimate the  $W_1$  distance we use  $W_1 \leq W_{1/\alpha}$  and the following fact:

$$W_{1/\alpha}(\rho_t^n, \rho_s^n) \leq \int_s^t |(\rho_z^n)'|_{w_{1/\alpha}} dz \leq \int_s^t \|v_z^n\|_{L^{1/\alpha}(\rho_z)} dz.$$

Then, applying inequality (3.1), we have

$$\int_{\Omega} \chi_M(x) d\rho_t^n(x) \leq \int_{\Omega} \chi_M(x) d\rho_a^n(x) + M \int_a^b F(\rho_z^n, q_z^n) dz \quad \text{for every } t \in [a, b].$$

The equi-integrability of  $t \mapsto F(\rho_t^n, q_t^n)$  finally gives

$$\int_{\Omega} \chi_M(x) d\rho_t^n(x) \leq \int_{\Omega} \chi_M(x) d\rho_a^n(x) + M\varepsilon \quad \text{for every } t \in [a, b].$$

In this way we have

$$\begin{aligned} \int_a^b \left( \int \chi_M d\rho_t^n \right)^{\alpha-1} \left( \int \chi_M d|q_t^n| dt \right) dt &\geq \left( \int \chi_M d\rho_a^n + M\varepsilon \right)^{\alpha-1} \int_a^b \left( \int \chi_M d|q_t^n| \right) dt \\ &= \left( \int \chi_M d\rho_a^n + M\varepsilon \right)^{\alpha-1} \int \phi^{1/\alpha} d|q^n|. \end{aligned}$$

Hence, we may go on with

$$\begin{aligned} \int \phi d\mathbf{m} &\geq \limsup_{n \rightarrow \infty} \left[ \left( \int \chi_M d\rho_a^n + M\varepsilon \right)^{\alpha-1} \int \phi^{1/\alpha} d|q^n| \right] \\ &\geq \left( \int \chi_M d\rho_a + M\varepsilon \right)^{\alpha-1} \int \phi^{1/\alpha} d|q|. \end{aligned}$$

In the last inequality, the second factor has been dealt with in the following way: suppose  $|q^n| \rightharpoonup \sigma$ ; then we have  $\sigma \geq |q|$ . Moreover,  $\phi \geq \tilde{\phi}$ , where  $\tilde{\phi}(t, x) := \chi_M(x)1_{(a,b)}(t)$ , and this last function is lower semicontinuous and positive, so that

$$\liminf_{n \rightarrow \infty} \int \phi^{1/\alpha} d|q^n| \geq \liminf_{n \rightarrow \infty} \int \tilde{\phi}^{1/\alpha} d|q^n| \geq \int \tilde{\phi}^{1/\alpha} d\sigma \geq \int \tilde{\phi}^{1/\alpha} d|q| = \int \phi^{1/\alpha} d|q|,$$

since the boundaries  $t = a$  and  $t = b$  are negligible for  $|q|$ .

After that, we can divide by  $(b - a)$  (keeping  $M$  fixed for a while) and pass to the limit as  $b \rightarrow a$ . This gives, for  $\mathcal{L}^1$  a.e.,  $a \in [0, 1]$ ,

$$\int \chi_M(x)^\alpha d\mathbf{m}_a(x) \geq \left( \int \chi_M(x) d\rho_a(x) + M\varepsilon \right)^{\alpha-1} \int \chi_M(x) d|q_a|(x).$$

Observe that, by choosing  $\varepsilon = 1/M^2$  from the beginning, if we let  $M \rightarrow \infty$  now and use that  $\chi_M$  monotonically converges to  $1_Q$ , by dominated convergence w.r.t.  $\mathbf{m}_a, \rho_a$ , and  $|q_a|$ , we end up with

$$(3.3) \quad \mathbf{m}_a(Q) \geq \rho_a(Q)^{\alpha-1} |q_a|(Q).$$

In the last term the convention  $0 \cdot \infty = 0$  is used (if  $|q_a|(Q) = 0$ ). This inequality is proved for closed sets, but by regularity of the measures it is not difficult to prove it for arbitrary sets. Actually, if  $S \subset \Omega$  is an arbitrary Borel set, then we can write

$$\mathbf{m}_a(S) \geq \mathbf{m}_a(Q) \geq \rho_a(Q)^{\alpha-1} |q_a|(Q) \geq \rho_a(S)^{\alpha-1} |q_a|(Q)$$

for every  $Q \subset S$  closed and take a sequence of closed sets  $Q_k$  such that  $|q_a|(Q_k) \rightarrow |q_a|(S)$ , since  $|q_a|$  is, for  $\mathcal{L}^1$  a.e.,  $a \in [0, 1]$ , a finite (and hence regular) measure on the compact set  $\Omega$ . We now want to prove the following:

- (i)  $q \ll \rho$ ;
- (ii)  $q_a = v_a \cdot \rho_a$  is atomic for  $\mathcal{L}^1$  a.e.,  $a \in [0, 1]$  (i.e.,  $\rho_a$  is atomic on  $\{v_a \neq 0\}$ );
- (iii)  $\mathbf{m}_a(\Omega) \geq F(\rho_a, q_a)$  for  $\mathcal{L}^1$  a.e.,  $a \in [0, 1]$ .

This would conclude the proof.

The first statement follows from the inequality (3.1): first observe that the curves  $\rho^n$  are equicontinuous, thanks to (2.6), so that they converge uniformly in time. This gives the following decomposition for the limit measure  $\rho = \int \rho_t dt$ . Then we observe that the equi-integrability of  $t \mapsto F(\rho_t^n, q_t^n)$  is equivalent (see [14]) to the existence of a convex nondecreasing superlinear map  $\vartheta$  such that

$$\sup_{n \in \mathbb{N}} \int_0^1 \vartheta(F(\rho_t^n, q_t^n)) dt < +\infty.$$

Observing that the map  $(x, y) \mapsto \vartheta(y/x)x$  is convex and positively 1-homogeneous, the functional

$$\mathcal{A}_\vartheta(\rho, q) = \begin{cases} \int_{[0,1] \times \Omega} \vartheta\left(\left|\frac{dq}{d\rho}\right|\right) d\rho & \text{if } q \ll \rho, \\ +\infty & \text{otherwise,} \end{cases}$$

is well defined and lower semicontinuous. Then we proceed exactly as in the proof of Theorem 2.4, using the functional  $\mathcal{A}_\vartheta$  in place of the Benamou–Brenier one  $\mathcal{A}_{1/\alpha}$ : this guarantees that  $q$  have an  $L^1$  density w.r.t.  $\rho$ . As a consequence, since  $\rho$  is a measure on  $[0, 1] \times \Omega$ , which disintegrates w.r.t. the Lebesgue measure on  $[0, 1]$ , the same will be true for  $q$  and we can write  $q_t = v_t \cdot \rho_t$ .

For the second statement, take the inequality  $\mathbf{m}_a(S) \geq \rho_a(S)^{\alpha-1} |q_a|(S)$ , which is valid for any Borel set  $S$ , and apply it to sets which are contained in the Borel set  $V_\varepsilon := \{x \in \Omega : |v_a(x)| > \varepsilon\}$ . For those sets, we have easily that  $\mathbf{m}_a(S) \geq \varepsilon \rho_a(S)^\alpha$ . This means that the measure  $\lambda := \varepsilon^{1/\alpha} \rho_a \llcorner V_\varepsilon$  satisfies the inequality  $\lambda(S)^\alpha \leq \mathbf{m}(S)$  for every Borel set  $S \subset V_\varepsilon$ . Since  $\mathbf{m}$  is a finite measure, this implies that  $\lambda$  is atomic (see Lemma 3.1 below). If the same is performed for every  $\varepsilon = 1/k$ , this proves that  $\rho_a$  is purely atomic on the set  $\{x : |v_a(x)| \neq 0\}$ ; that is,  $q_a = v_a \cdot \rho_a$  is purely atomic.

Once we know that  $q_a$  is atomic, we can infer that

$$F(\rho_a, q_a) = \sum_i \rho_a(\{x_i\})^{\alpha-1} |q_a|(\{x_i\}),$$

and we need only consider  $Q = \{x_i\}$  in (3.3) and add up:

$$\mathbf{m}_a(\Omega) \geq \sum_i \mathbf{m}_a(\{x_i\}) \geq F(\rho_a, q_a),$$

which finally concludes the proof.

LEMMA 3.1. *Take two finite positive measures  $\lambda$  and  $\mu$  on a domain  $\Omega$ , and  $\alpha \in (0, 1)$ . Suppose that the inequality  $\lambda(S)^\alpha \leq \mu(S)$  is satisfied for every Borel set  $S \subset \Omega$ . Then  $\lambda$  is purely atomic.*

*Proof.* Consider a regular grid on  $\Omega$  of step  $1/k$ , for  $k \in \mathbb{N}$ , and build a measure  $\lambda_k$  by putting, in every cell of the grid, all the mass of  $\lambda$  in a single point of the cell. This measure  $\lambda_k$  is atomic, and we have

$$G_\alpha(\lambda_k) = \sum_i \lambda(S_i)^\alpha \leq \sum_i \mu(S_i) = \mu(\Omega) < +\infty,$$

where the  $S_i$  are the cells of the grid. If we let  $k$  go to  $\infty$ , then the step of the grid goes to zero, and we obviously have  $\lambda_k \rightarrow \lambda$ . On the other hand, the functional  $G_\alpha$  is lower semicontinuous (see [8]), and this implies  $G_\alpha(\lambda) \leq \liminf_{k \rightarrow \infty} G_\alpha(\lambda_k) \leq \mu(\Omega) < +\infty$ . In particular,  $\lambda$  is atomic, thus proving the assertion.  $\square$

**3.3. Proof of Theorem 2.6.** In order to prove existence, one need only take a minimizing sequence and apply Theorem 2.4 to get a new minimizing sequence that is  $\tau$ -converging: this new sequence is obtained through reparametrization (which does not change the value of  $\mathcal{F}$ ) and by extracting a subsequence. Since the constraints in the problem are linear (i.e.,  $\rho_i = \mu_i$  for  $i = 0, 1$  and the continuity equation), the limit  $(\rho, q)$  will satisfy the same constraints as well. The semicontinuity proved in Theorem 2.5 allows us to obtain the existence of a solution.

**4. Equivalence with previous models.** In this section we prove the equivalence of problem (2.5) to the other previous formulations of branched transport problems existing in literature. In particular, as a reference model we will take the one presented in [6], in which the energy is defined as

$$E_\alpha(Q) = \int_{\mathcal{C}} \int_0^1 |\sigma(t)|_Q^{\alpha-1} |\sigma'(t)| dt dQ(\sigma),$$

where  $\mathcal{C} = C([0, 1]; \Omega)$ ,  $Q$  is a probability measure over  $\mathcal{C}$  and concentrated on the set  $\text{Lip}([0, 1]; \Omega)$  (*traffic plan*), and for every  $x \in \Omega$ , the quantity  $|x|_Q$  is the *multiplicity* of  $x$  w.r.t.  $Q$ , defined by

$$|x|_Q = Q(\{\sigma \in \mathcal{C} : x \in \sigma([0, 1])\}).$$

Given  $\mu_0, \mu_1 \in \mathcal{P}(\Omega)$ , the corresponding minimum problem is then given by

$$d_\alpha(\mu_0, \mu_1) = \min_{Q \in TP(\mu_0, \mu_1)} E_\alpha(Q),$$

where  $TP(\mu_0, \mu_1)$  is the set of traffic plans with prescribed time marginals at  $t = 0, 1$ ; that is,

$$TP(\mu_0, \mu_1) = \{Q \in \mathcal{C} : Q \text{ concentrated on } \text{Lip}([0, 1]; \Omega), (e_i)_\# Q = \mu_i, i = 0, 1\},$$

and  $e_t : \mathcal{C} \rightarrow \Omega$  is the *evaluation map* at time  $t$ , given by  $e_t(\sigma) = \sigma(t)$  for every  $\sigma \in \mathcal{C}$ .

*Remark 4.1.* We recall that this model is completely equivalent to that developed by Xia (see [28] for the presentation of the model and [6, Chapter 9] for the equivalence), which is based on a relaxation procedure, starting from an energy defined on finitely atomic probability measures  $\mu_0$  and  $\mu_1$ . In particular, thanks to this relaxed formulation, we get that for every  $\mu_0$  and  $\mu_1$ , there exist two sequences  $\mu_0^n$  and  $\mu_1^n$  of finitely atomic probability measures, weakly converging to  $\mu_0$  and  $\mu_1$ , respectively, such that

$$(4.1) \quad d_\alpha(\mu_0^n, \mu_1^n) \rightarrow d_\alpha(\mu_0, \mu_1).$$

We also need to consider a slight modification of the functional  $E_\alpha$  above, introduced in [7]:

$$C_\alpha(Q) = \int_{\mathcal{C}} \int_0^1 |(\sigma(t), t)|_Q^{\alpha-1} |\sigma'(t)| dt dQ(\sigma),$$

where now the *synchronized multiplicity*  $|(x, t)|_Q$  is

$$|(x, t)|_Q = Q(\{\sigma \in \mathcal{C} : \sigma(t) = x\}).$$

This second multiplicity accounts for the quantity of curves passing at the same time through the same point, while the one used in the definition of  $E_\alpha$  considered all the curves passing eventually through the point: in this sense, the model corresponding to the energy  $C_\alpha$  is more dynamical in spirit. As a straightforward consequence of the definition of the two multiplicities, we get

$$(4.2) \quad |\sigma(t)|_Q \geq |(\sigma(t), t)|_Q,$$

so that  $E_\alpha(Q) \leq C_\alpha(Q)$ . Concerning the comparison between the minimization of  $E_\alpha$  and  $C_\alpha$ , we recall the following result (see [7, Theorem 5.1]).

**THEOREM 4.2.** *Let  $\mu_0, \mu_1 \in \mathcal{P}(\Omega)$ , with  $\mu_0$  a finite sum of Dirac masses. Then for every  $\alpha \in [0, 1]$  we get*

$$\min_{Q \in TP(\mu_0, \mu_1)} E_\alpha(Q) = \min_{Q \in TP(\mu_0, \mu_1)} C_\alpha(Q).$$

We are now in a position to state and prove a result giving the equivalence between our model and the one relative to the energy  $E_\alpha$ .

**THEOREM 4.3.** *For every  $\alpha \in (0, 1)$  and  $\mu_0, \mu_1 \in \mathcal{P}(\Omega)$  we get*

$$(4.3) \quad \mathcal{B}_\alpha(\mu_0, \mu_1) = \min_{Q \in TP(\mu_0, \mu_1)} E_\alpha(Q) = d_\alpha(\mu_0, \mu_1).$$

*Proof.* We start proving the inequality  $\mathcal{B}_\alpha(\mu_0, \mu_1) \geq d_\alpha(\mu_0, \mu_1)$ . Clearly, if  $\mathcal{B}_\alpha(\mu_0, \mu_1) = +\infty$ , then there is nothing to prove; otherwise, take  $(\rho, q)$  optimal, which implies, by the way, that  $q = v \cdot \rho$  and that  $q$  is atomic. Thanks to the *superposition principle* (see [2, Theorem 8.2.1]) we can construct a probability measure  $Q \in \mathcal{C}$  such that  $\rho_t = (e_t)_\# Q$  and  $Q$  is concentrated on absolutely continuous integral curves of  $v$ , in the sense that

$$\int_{\mathcal{C}} \left| \sigma(t) - \sigma(0) - \int_0^t v_s(\sigma(s)) ds \right| dQ(\sigma) = 0 \quad \text{for every } t \in [0, 1].$$

Using this information, together with the fact that  $E_\alpha \leq C_\alpha$  and exchanging the order of integration, we get

$$\begin{aligned} E_\alpha(Q) &\leq C_\alpha(Q) = \int_{\mathcal{C}} \int_0^1 |(\sigma(t), t)|_Q^{\alpha-1} |\sigma'(t)| dt dQ(\sigma) \\ &= \int_0^1 \int_{\mathcal{C}} |(\sigma(t), t)|_Q^{\alpha-1} |\sigma'(t)| dQ(\sigma) dt \\ &= \int_0^1 \int_{\mathcal{C}} |(\sigma(t), t)|_Q^{\alpha-1} |v_t(\sigma(t))| dQ(\sigma) dt \\ &= \int_0^1 \int_{\Omega} |(x, t)|_Q^{\alpha-1} |v_t(x)| d\rho_t(x) dt. \end{aligned}$$

Then we observe that, by virtue of the fact that  $\rho_t = (e_t)_\#Q$ , there holds

$$|(x, t)|_Q = Q(\{\sigma \in \mathcal{C} : \sigma(t) = x\}) = \rho_t(\{x\}),$$

so that we can rewrite the last integral as

$$\int_0^1 \int_\Omega \rho_t(\{x\})^{\alpha-1} |v_t(x)| d\rho_t(x) dt = \int_0^1 \int \rho_t(\{x\})^{\alpha-1} d|q_t|(x) dt = \int_0^1 \sum_{i \in \mathbb{N}} |v_{t,i}| \rho_{t,i}^\alpha dt,$$

which then gives

$$d_\alpha(\mu_0, \mu_1) = \min_{Q \in TP(\mu_0, \mu_1)} E_\alpha(Q) \leq \mathcal{F}(\rho, q) = \mathcal{B}_\alpha(\mu_0, \mu_1).$$

In order to prove the reverse inequality, we first prove that

$$(4.4) \quad \mathcal{B}_\alpha(\mu_0, \mu_1) \leq \min_{Q \in TP(\mu_0, \mu_1)} C_\alpha(Q).$$

Take  $Q \in TP(\mu_0, \mu_1)$  optimal for  $C_\alpha$ . Then we know that there exists a pair  $(\rho, q)$ , which is a solution of the continuity equation, with  $\rho_t = (e_t)_\#Q$  and  $q_t = v_t \cdot \rho_t$ . The velocity  $v$  may be chosen as

$$v_t(x) = \int \sigma'(t) dQ^{t,x}(\sigma),$$

where  $Q^{t,x}$  is the disintegration of  $Q$  w.r.t. the evaluation function  $e_t$  (see [21] for this representation formula of the velocity field  $v$ ). This means that each  $Q^{t,x}$  is a probability measure concentrated on the set  $\{\sigma \in \mathcal{C} : \sigma(t) = x\}$  and  $Q = \int Q^{t,x} d\rho_t(x)$ . Therefore, arguing as before,

$$\begin{aligned} C_\alpha(Q) &= \int_0^1 \int_{\mathcal{C}} |(\sigma(t), t)|_Q^{\alpha-1} |\sigma'(t)| dQ(\sigma) dt \\ &= \int_0^1 \int_\Omega |(x, t)|_Q^{\alpha-1} \left( \int |\sigma'(t)| dQ^{t,x}(\sigma) \right) d\rho_t(x) dt \\ &\geq \int_0^1 \int_\Omega |(x, t)|_Q^{\alpha-1} |v_t(x)| d\rho_t(x) dt \\ &= \int_0^1 \int_\Omega \rho_t(\{x\})_Q^{\alpha-1} |v_t(x)| d\rho_t(x) dt, \end{aligned}$$

which gives the desired inequality (4.4) since, even if we do not know that  $q_t$  or  $\rho_t$  are atomic, we can restrict the last integral to the set of atoms of  $\rho$ .

To summarize, up to now we have shown

$$d_\alpha(\mu_0, \mu_1) \leq \mathcal{B}_\alpha(\mu_0, \mu_1) \leq \min_{Q \in TP(\mu_0, \mu_1)} C_\alpha(Q),$$

and the equality holds whenever  $\mu_0$  is a finite sum of Dirac masses, thanks to Theorem 4.2. In order to conclude, it is enough to notice that thanks to Remark 4.1, we may take two sequences  $\mu_0^n$  and  $\mu_1^n$  of finitely atomic probability measures such that  $\mu_0^n \rightharpoonup \mu_0$ ,  $\mu_1^n \rightharpoonup \mu_1$  and

$$d_\alpha(\mu_0^n, \mu_1^n) \rightarrow d_\alpha(\mu_0, \mu_1),$$

thus getting

$$d_\alpha(\mu_0, \mu_1) \leq \mathcal{B}_\alpha(\mu_0, \mu_1) \leq \liminf_{n \rightarrow \infty} \mathcal{B}_\alpha(\mu_0^n, \mu_1^n) \leq \lim_{n \rightarrow \infty} d_\alpha(\mu_0^n, \mu_1^n) = d_\alpha(\mu_0, \mu_1),$$

and hence concluding the proof.  $\square$

*Remark 4.4.* Observe that in Theorem 4.3, we not only proved the equality of the minima, but we also provided a natural way to pass from a minimizer of our formulation à la Benamou–Brenier to a minimizer of the traffic plans model and back. The two problems are thus equivalent in the sense that they describe the same kind of energy and the same optimal structures of branched transport: the simple equality of the minima (4.3) is just a consequence of this more important fact.

*Remark 4.5.* In the previous proof, we used the equivalence between the models corresponding to  $E_\alpha$  and  $C_\alpha$ , which was the content of Theorem 4.2: as we said, this result has been established in [7] under the assumption that the starting measure  $\mu_0$  is finitely atomic. It is based on the fact that, under this assumption, an optimal traffic plan  $Q$  for  $E_\alpha$  can be *synchronized* (see [7, Proposition 4.10]); i.e., a time reparametrization leads to a  $\tilde{Q}$  such that

$$|\sigma(t)|_{\tilde{Q}} = |(\sigma(t), t)|_{\tilde{Q}},$$

and thanks to the reparametrization invariance of  $E_\alpha$ , we have  $E_\alpha(Q) = E_\alpha(\tilde{Q}) = C_\alpha(\tilde{Q})$ . The synchronization result in [7] is likely to be extendable to more general situations, without any restriction on  $\mu_0$  and  $\mu_1$ ; yet, we decided to use Theorem 4.2 in the form proved in [7] since it was sufficient for our scope. Observe that in any case the proof of Theorem 4.3 gives at least the equality of the minima

$$\min_{Q \in TP(\mu_0, \mu_1)} E_\alpha(Q) = \min_{Q \in TP(\mu_0, \mu_1)} C_\alpha(Q)$$

for every  $\mu_0, \mu_1 \in \mathcal{P}(\Omega)$ .

**Appendix. The distances  $d_\alpha$  and  $W_{1/\alpha}$ .** This last section is devoted to estimates between the distance  $d_\alpha$  induced by the branched transport and the Wasserstein distances  $W_p$ . In particular, in [24] the following estimate is proved for  $\alpha > 1 - 1/d$ :

$$d_\alpha \leq C W_1^{d(\alpha-1)+1}.$$

As far as lower bounds on  $d_\alpha$  are concerned, the most trivial is  $d_\alpha \geq W_1$  but [17, Theorem 8.1] also proves  $d_\alpha \geq W_{1/\alpha}$ , which is slightly better. Moreover, for scaling reasons (w.r.t. the mass) it is not possible to replace  $W_{1/\alpha}$  by  $W_p$  with  $p > 1/\alpha$  in this last inequality.

In this paper we already needed to estimate some branched transport cost in terms of  $W_{1/\alpha}$  distances and metric derivatives. In this section we prove the inequalities

$$W_{1/\alpha} \leq d_\alpha \leq C W_{1/\alpha}^{d(\alpha-1)+1} \quad \forall \alpha \in (1 - 1/d, 1].$$

These inequalities are just particular cases of the aforementioned ones, but the proofs we will provide are different and somehow simpler.

The first inequality will be approached through the formulation of branched transport that we gave in this paper, but the main tool (i.e., inequality (3.1)) is essentially the same as that in [17] and [23]. What is different is the way to extend this idea to generic measures, i.e., nonatomic ones.

THEOREM A.1. For every  $\mu_0, \mu_1 \in \mathcal{P}(\Omega)$  we get

$$(A.1) \quad W_{1/\alpha}(\mu_0, \mu_1) \leq d_\alpha(\mu_0, \mu_1).$$

*Proof.* We first observe that thanks to the results of the previous section, for every  $\mu_0, \mu_1 \in \mathcal{P}(\Omega)$  we get

$$d_\alpha(\mu_0, \mu_1) = \int_0^1 \left[ \int_\Omega |v_t(x)| \rho_t(\{x\})^\alpha d\#(x) \right] dt$$

for a suitable  $(\rho, q)$  admissible in the formulation (2.5), with  $q = \rho v$ . Moreover, using inequality (3.1) once more, the right-hand side in the previous expression can be estimated as

$$\int_0^1 \left[ \int_\Omega |v_t(x)| \rho_t(\{x\})^\alpha d\#(x) \right] dt \geq \int_0^1 \|v_t\|_{L^{1/\alpha}(\rho_t)} dt$$

and, finally, using the fact that  $(\rho, q)$  is a solution of the continuity equation, we can infer (see [2, Theorem 8.3.1]) that

$$|\rho'_t|_{W_{1/\alpha}} \leq \|v_t\|_{L^{1/\alpha}(\rho_t)} \quad \text{for } \mathcal{L}^1 \text{ a.e., } t \in [0, 1],$$

so that

$$d_\alpha(\mu_0, \mu_1) \geq \int_0^1 |\rho'_t|_{W_{1/\alpha}} dt \geq W_{1/\alpha}(\mu_0, \mu_1),$$

where in the last inequality we just estimated the length of a curve by the distance between its endpoints. Thus we have obtained (A.1), concluding the proof.  $\square$

In order to prove the other inequality, we first have to introduce some notation: we set  $Q = [0, 1]^d$  and  $Q_L = [0, L]^d$ , and for every  $j \in \mathbb{N}$  we consider the following subset of multi-indices:

$$B_j = \{z \in \mathbb{N}^d : \|z\|_\infty \leq 2^j - 1\}.$$

We observe that  $\#(B_j) = 2^{jd}$ ; then we make a partition of the cube  $Q_L$  by dyadic cubes having edge length  $L/2^j$ , i.e.,

$$Q_L = \bigcup_{i=1}^{2^{jd}} Q_j^i := \bigcup_{z \in B_j} \frac{LQ + Lz}{2^j}.$$

For every  $\mu \in \mathcal{P}(\Omega)$  such that  $\Omega \subset Q_L$ , its *dyadic approximation* is given by

$$a_j(\mu) = \sum_{i=1}^{2^{jd}} \mu_j^i \delta_{x_j^i},$$

where  $\mu_j^i = \mu(Q_j^i)$  and  $x_j^i$  is the center of  $Q_j^i$ . We shall always assume that  $\Omega \subset Q_L$  for a suitable  $L$ ; then the following estimate is well known (see [6, Proposition 6.6]).

PROPOSITION A.2. Let  $\alpha \in (1 - 1/d, 1]$ . Then for every  $\mu \in \mathcal{P}(\Omega)$  we have

$$(A.2) \quad d_\alpha(a_j(\mu), \mu) \leq \frac{2^{(d(1-\alpha)-1)j} L\sqrt{d}}{2^{1-d(1-\alpha)} - 1} \frac{1}{2}.$$

The main tool (if one wants to estimate  $d_\alpha$  from above by a power of  $W_{1/\alpha}$ ) is to show that the distance  $d_\alpha$  between two dyadic approximations can be estimated in terms of their  $1/\alpha$ -Wasserstein distance: this is the content of the next result.

LEMMA A.3. *Let  $\alpha \in (1 - 1/d, 1]$ . Then for every  $\mu_0, \mu_1 \in \mathcal{P}(\Omega)$  we get*

$$(A.3) \quad d_\alpha(a_j(\mu_0), a_j(\mu_1)) \leq C W_{1/\alpha}(a_j(\mu_0), a_j(\mu_1)) 2^{jd(1-\alpha)},$$

with  $C$  depending only on  $d$  and  $\alpha$ .

*Proof.* Let us consider an optimal transport  $\gamma_j$  between  $a_j(\mu_0)$  and  $a_j(\mu_1)$  for the cost  $c(x, y) = |x - y|^{1/\alpha}$ . That is,  $\gamma_j \in \mathcal{P}(\Omega \times \Omega)$ , and it is of the form

$$\gamma_j = \sum_{i,k=1}^{2^{jd}} M_j(i, k) \delta_{x_j^i} \otimes \delta_{x_j^k},$$

with the  $2^{jd} \times 2^{jd}$  matrix  $\{M_j(i, k)\}_{i,k}$  belonging to the convex set  $\mathfrak{M}$  given by

$$\mathfrak{M} = \left\{ \{a_{i,k}\}_{i,k} : a_{i,k} \geq 0, \sum_{i=1}^{2^{jd}} a_{i,k} = \mu_1(Q_j^k), \sum_{k=1}^{2^{jd}} a_{i,k} = \mu_0(Q_j^i) \right\}.$$

We know by optimality that  $\{M_j(i, k)\}_{i,k}$  can be taken to belong to  $\text{Ext}(\mathfrak{M})$ , the set of extremal points of  $\mathfrak{M}$ , which consists of the so-called *acyclic matrices* (see [16]). They are those matrices belonging to  $\mathfrak{M}$  such that the following property holds:

$$\prod_{r=1}^s a_{i_r, k_r} a_{i_r, k_{r+1}} = 0$$

for every  $2 \leq s \leq 2^{jd}$  and every set of indices  $i_1 < \dots < i_s \in \{1, \dots, 2^{jd}\}$ ,  $k_1 < \dots < k_j \in \{1, \dots, 2^{jd}\}$  (the convention  $i_{2^{jd}+1} = i_1$  and  $k_{2^{jd}+1} = k_1$  is used). This implies in particular that

$$(A.4) \quad \#\{(i, k) : M_j(i, k) \neq 0\} \leq 2 \cdot 2^{jd},$$

that is,  $\{M_j(i, k)\}_{i,k}$  has at most  $2 \cdot 2^{jd}$  nonzero entries: in other terms, this optimal transport plan  $\gamma_j$  does not move more than  $2 \cdot 2^{jd}$  atoms. Setting  $|x_j^i - x_j^k| = \ell_{i,k}$ , we then get

$$W_{1/\alpha}(a_j(\mu_0), a_j(\mu_1)) = \left( \sum_{i,k=1}^{2^{jd}} M_j(i, k) \ell_{i,k}^{1/\alpha} \right)^\alpha,$$

and using (A.4) and Jensen's inequality,

$$\begin{aligned} d_\alpha(a_j(\mu_0), a_j(\mu_1)) &\leq \sum_{i,k=1}^{2^{jd}} M_j(i, k)^\alpha \ell_{i,k} = \sum_{i,k=1}^{2^{jd}} \left( M_j(i, k) \ell_{i,k}^{\frac{1}{\alpha}} \right)^\alpha \\ &\leq \left( \sum_{i,k=1}^{2^{jd}} M_j(i, k) \ell_{i,k}^{\frac{1}{\alpha}} \right)^\alpha (\#\{(i, k) : M_j(i, k) \neq 0\})^{1-\alpha} \\ &\leq C W_{1/\alpha}(a_j(\mu_0), a_j(\mu_1)) 2^{jd(1-\alpha)}, \end{aligned}$$

which concludes the proof.  $\square$



THEOREM A.4. *Let  $\alpha \in (1 - 1/d, 1]$ . Then we get*

$$(A.5) \quad d_\alpha(\mu_0, \mu_1) \leq C W_{1/\alpha}(\mu_0, \mu_1)^{d(\alpha-1)+1},$$

with a constant  $C$  depending only on  $d, \alpha$ , and the diameter of  $\Omega$ .

*Proof.* Using the triangular inequality, (A.2), and (A.3), we get for every  $j \in \mathbb{N}$

$$\begin{aligned} d_\alpha(\mu_0, \mu_1) &\leq d_\alpha(\mu_0, a_j(\mu_0)) + d_\alpha(a_j(\mu_0), a_j(\mu_1)) + d_\alpha(a_j(\mu_1), \mu_1) \\ &\leq C 2^{(d(1-\alpha)-1)j} + d_\alpha(a_j(\mu_0), a_j(\mu_1)) \\ &\leq C 2^{(d(1-\alpha)-1)j} + C W_{1/\alpha}(a_j(\mu_0), a_j(\mu_1)) 2^{jd(1-\alpha)}, \end{aligned}$$

and

$$\begin{aligned} W_{1/\alpha}(a_j(\mu_0), a_j(\mu_1)) &\leq W_{1/\alpha}(a_j(\mu_0), \mu_0) + W_{1/\alpha}(\mu_0, \mu_1) + W_{1/\alpha}(a_j(\mu_0), \mu_1) \\ &\leq C 2^{-j} + W_{1/\alpha}(\mu_0, \mu_1), \end{aligned}$$

which finally gives

$$\begin{aligned} d_\alpha(\mu_0, \mu_1) &\leq C 2^{(d(1-\alpha)-1)j} + C W_{1/\alpha}(\mu_0, \mu_1) 2^{jd(1-\alpha)} \\ &= C 2^{(d(1-\alpha)-1)j} (1 + W_{1/\alpha}(\mu_0, \mu_1) 2^j). \end{aligned}$$

It is now sufficient to choose the index  $j$  in such a way that

$$\frac{\text{diam}(\Omega)}{2^j} \leq W_{1/\alpha}(\mu_0, \mu_1) \leq \frac{\text{diam}(\Omega)}{2^{j-1}},$$

which in turn yields

$$2^{(d(1-\alpha)-1)j} (1 + W_{1/\alpha}(\mu_0, \mu_1) 2^j) \leq C W_{1/\alpha}(\mu_0, \mu_1)^{d(\alpha-1)+1},$$

thus giving the thesis.  $\square$

*Remark A.5.* Notice that the very same  $\mu_0$  and  $\mu_1$  of Example 6.19 in [6] show that the exponent  $d(\alpha - 1) + 1$  in inequality (A.5) cannot be improved.

*Remark A.6.* As we briefly mentioned previously, observe that the distances  $d_\alpha$  and  $W_{1/\alpha}$  have exactly the same scaling w.r.t. the mass. Moreover, the Wasserstein distances and the  $d_\alpha$  can both be extended to positive measures of mass  $m$ , not necessarily equal to 1. Using the scaling properties of  $W_{1/\alpha}$  and  $d_\alpha$ , in conjunction with the previous inequalities, it is easy to see that the dependence on  $m$  and the diameter of  $\Omega$  of the constant  $C$  appearing in (A.5) is

$$C(m, \Omega) \simeq m^{(1-\alpha)(N-1)} \text{diam}(\Omega)^{d(1-\alpha)}.$$

REFERENCES

[1] L. AMBROSIO, *Lecture notes on optimal transport problems*, in *Mathematical Aspects of Evolving Interfaces*, Lecture Notes in Math. 1812, Springer-Verlag, Berlin, 2003, pp. 1–52.  
 [2] L. AMBROSIO, N. GIGLI, AND G. SAVARÉ, *Gradient Flows in Metric Spaces and in the Space of Probability Measures*, 2nd ed., Lectures Math. ETH Zürich, Birkhäuser Verlag, Basel, 2008.  
 [3] J. D. BENAMOU AND Y. BRENIER, *A computational fluid mechanics solution to the Monge–Kantorovich mass transfer problem*, Numer. Math., 84 (2000), pp. 375–393.  
 [4] M. BERNOT, V. CASELLES, AND J. M. MOREL, *Traffic plans*, Publ. Mat., 49 (2005), pp. 417–451.  
 [5] M. BERNOT, V. CASELLES, AND J. M. MOREL, *The structure of branched transportation networks*, Calc. Var. Partial Differential Equations, 32 (2008), pp. 279–371.

- [6] M. BERNOT, V. CASELLES, AND J. M. MOREL, *Optimal Transportation Networks. Models and Theory*, Lecture Notes in Math. 1955, Springer-Verlag, Berlin, 2009.
- [7] M. BERNOT AND A. FIGALLI, *Synchronized traffic plans and stability of optima*, ESAIM Control Optim. Calc. Var., 14 (2008), pp. 864–878.
- [8] G. BOUCHITTÉ AND G. BUTTAZZO, *New lower semicontinuity results for nonconvex functionals defined on measures*, Nonlinear Anal., 15 (1990), pp. 679–692.
- [9] G. BOUCHITTÉ AND P. SEPPECHER, in preparation.
- [10] A. BRANCOLINI, G. BUTTAZZO, AND F. SANTAMBROGIO, *Path functionals over Wasserstein spaces*, J. Eur. Math. Soc. (JEMS), 8 (2006), pp. 415–434.
- [11] L. BRASCO, *Curves of minimal action over metric spaces*, Ann. Mat. Pura Appl. (4), 189 (2010), pp. 95–125.
- [12] L. BRASCO AND F. SANTAMBROGIO, *An equivalent path functional formulation of branched transportation problems*, Discrete Contin. Dyn. Syst., 29 (2011), pp. 845–871.
- [13] Y. BRENIER, *Extended Monge-Kantorovich theory*, in Optimal Transportation and Applications, Lecture Notes in Math. 1813, Springer-Verlag, Berlin, 2003, pp. 91–121.
- [14] G. BUTTAZZO, *Semicontinuity, Relaxation and Integral Representation in the Calculus of Variations*, Pitman Res. Notes Math. Ser. 207, Longman Scientific & Technical, Harlow, 1989.
- [15] G. BUTTAZZO, C. JIMENEZ, AND E. OUDET, *An optimization problem for mass transportation with congested dynamics*, SIAM J. Control Optim., 48 (2009), pp. 1961–1976.
- [16] J. L. DENNY, *The support of discrete extremal measures with given marginals*, Michigan Math. J., 27 (1980), pp. 59–64.
- [17] G. DEVILLANOVA AND S. SOLIMINI, *Elementary properties of optimal irrigation patterns*, Calc. Var. Partial Differential Equations, 28 (2007), pp. 317–349.
- [18] J. DOLBEAULT, B. NAZARET, AND G. SAVARÉ, *A new class of transport distances between measures*, Calc. Var. Partial Differential Equations, 34 (2009), pp. 193–231.
- [19] E. N. GILBERT, *Minimum cost communication networks*, Bell System Tech. J., 46 (1967), pp. 2209–2227.
- [20] J.-M. LASRY AND P.-L. LIONS, *Mean field games*, Jpn. J. Math., 2 (2007), pp. 229–260.
- [21] S. LISINI, *Characterization of absolutely continuous curves in Wasserstein spaces*, Calc. Var. Partial Differential Equations, 28 (2007), pp. 85–120.
- [22] F. MADDALENA, J. M. MOREL, AND S. SOLIMINI, *A variational model of irrigation patterns*, Interfaces Free Bound., 5 (2003), pp. 391–415.
- [23] F. MADDALENA AND S. SOLIMINI, *Transport distances and irrigation models*, J. Convex Anal., 16 (2009), pp. 121–152.
- [24] J. M. MOREL AND F. SANTAMBROGIO, *Comparison of distances between measures*, Appl. Math. Lett., 20 (2007), pp. 427–432.
- [25] C. VILLANI, *Optimal Transport. Old and New*, Grundlehren Math. Wiss. 338, Springer-Verlag, Berlin, 2009.
- [26] C. VILLANI, *Topics in Optimal Transportation*, Grad. Stud. Math. 58, AMS, Providence, RI, 2003.
- [27] Q. XIA, *Interior regularity of optimal transport paths*, Calc. Var. Partial Differential Equations, 20 (2004), pp. 283–299.
- [28] Q. XIA, *Optimal paths related to transport problems*, Commun. Contemp. Math., 5 (2003), pp. 251–279.