

A Benamou-Brenier approach to branched transport

Lorenzo Brasco^{1,2}

Giuseppe Buttazzo¹

Filippo Santambrogio²

31.03.2010

Abstract

The problem of branched transportation aims to describe the movement of masses when, due to concavity effects, they have the interest to travel together as much as possible, because the cost for a path of length ℓ 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... Several models have been employed in the literature to present this transport problem, and the present paper looks at a dynamical one, similar to the celebrated Benamou-Brenier formulation of Kantorovitch optimal transport. The movement is represented by a path ρ_t of probabilities, connecting an initial state μ_0 to a final state μ_1 , 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 non-convex and finite on atomic measures: $\int_0^1 \left(\int_\Omega \rho^{\alpha-1} |q| d\#(x) \right) dt$.

AMS Subject Classification (2000): 49J45, 49Q20, 49Q10, 90B18, 60K30

Keywords: Optimal transport, branched transport, continuity equation, functionals on spaces of measures, optimal networks.

1 Introduction

The optimal mass transportation theory consists in the study of transporting a given mass distribution μ_0 on Ω (that we assume to be a compact and convex subset of \mathbb{R}^d) into a final configuration μ_1 , by minimizing the total transportation cost, the latter being suitably defined: clearly, μ_0 and μ_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

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

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 μ_0 and μ_1 respectively. The cases $c(x, y) = |x - y|^p$

¹ Dipartimento di Matematica, Università di Pisa, Largo B. Pontecorvo 5 - 56127 Pisa, ITALY brasco@mail.dm.unipi.it, buttazzo@dm.unipi.it

² CEREMADE, UMR CNRS 7534, Université Paris-Dauphine, Pl. de Lattre de Tassigny - 75775 Paris Cedex 16, FRANCE filippo@ceremade.dauphine.fr

with $p \geq 1$ have been particularly 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 wide 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 ρ , defined as (we refer the reader to [2], for more details)

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

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 ρ_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 $q_t = v_t \cdot \rho_t$, v being the velocity vector) which represents the flux $q = \rho v$ and satisfies

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

(the degenerate case $p = 1$ being a little bit more involved, since $q_t \ll \rho_t$ is no more 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 (ρ, 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

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

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

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

The problem above is the one which 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 ρ 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 (ρ, q) . Recently, Dolbeault, Nazaret and Savaré introduced in [17] new classes of distances over $\mathcal{P}(\mathbb{R}^d)$ based on the minimization of

the functional (where λ is a given reference measure on \mathbb{R}^d and ρ and q are identified with their densities w.r.t. λ)

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

which are connected to the non-linear 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 $\beta \in [0, 1]$), this problem turns out to be convex as well. The main interest that motivated Dolbeault et al. to the study of these distances lies in the possible applications to diffusion equations of the type of the non-linear mobility continuity equation $\partial_t \rho + \operatorname{div}_x (h(\rho)v) = 0$ above, where the vector field v depends on ρ in a way such that the equation can be interpreted as a gradient flow of a given functional with respect to these new family of dynamical distances. Moreover, the equations of the geodesics (which are similar to a *mean-field game system*, see [19]) and the conditions for these distances to be finite are being studied in [24].

In connection with *congestion* effects and crowd motion, other models include penalizations on high densities: in [14] 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 case opposite to congestion, when *concentration* effects are present and the mass has the interest 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 [18] formulated a mathematical model for the transportation of signals along telephone cables. More recently the 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 these models have in common the fact that the cost for a mass m moving on a path of length ℓ 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, 21] for every $0 < \alpha < 1$ a transportation cost from ρ_0 to ρ_1 is considered through a suitable use of probabilities defined on spaces of curves in Ω , with [21] (the so called *irrigation patterns model*) dealing with the case of a single source $\rho_0 = \delta_{x_0}$. See Section 4 to have a glance at the details of these models and their formulations. On the other hand, the model of [27] can be seen as the natural extension of the original Gilbert's model and uses vector measures having prescribed divergence $\rho_0 - \rho_1$: 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 considering 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 raise to branching phenomena.

These models are not satisfactory yet, because they are in general not equivalent to those by Gilbert, Xia or Bernot-Caselles-Morel. A tentative 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], 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 in the minimization of a suitable cost $\mathcal{F}(\rho, q)$ under the constraint of the continuity equation $\partial_t \rho + \operatorname{div}_x q = 0$, that we believe *is the most natural for this kind of problems*. The only approach to dynamical branched transportation using the continuity equation is, as far as we know, the one 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 energy of Xia in [27]) and the dynamical features are not completely exploited.

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

- in Section 2 we give the precise setting and state the main results;
- Section 3 is devoted to the proofs giving the existence of an optimal path ρ_t ;
- 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);
- 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 [23] and [16], but some very precise issues concerning d_α and $W_{1/\alpha}$ are very close to the topics of this paper and deserve being treated 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 the following Ω 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 Ω is convex and compact. The space $\mathcal{P}(\Omega)$ of all Borel probabilities on Ω 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 Ω , while \mathcal{L}^k will indicate the k -dimensional Lebesgue measure.

The main objects to be considered will be pairs (ρ, q) with

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

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

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

whose precise meaning is given in the sense of distributions, that is

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

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

Definition 2.1. We denote by \mathfrak{D} the set of all pairs (ρ, 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 μ_0 to μ_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 with respect to ρ :

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

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

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

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

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

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 we consider is

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

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 in proving lower semicontinuity and coercivity of the problem under consideration, with respect to a suitable convergence.

Remark 2.2. We point out that the weak* convergence of the pairs (ρ, 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_{α} and of the convexity of $(x, y) \mapsto |x|^p/y^{p-1}$).

In order to prove in the easiest possible way a semicontinuity result, we will introduce a convergence which is stronger than the weak convergence of measures on $[0, 1] \times \Omega$, but weaker than 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. We say that a sequence (ρ^n, q^n) τ -converges to (ρ, q) if $(\rho^n, q^n) \rightharpoonup (\rho, q)$ in the sense of measures and

$$\sup_{n \in \mathbb{N}, t \in [0, 1]} F(\rho_t^n, q_t^n) < +\infty.$$

Theorem 2.4. Let (ρ^n, q^n) be a sequence such that $\mathcal{F}(\rho^n, q^n) \leq C$, then up to a time reparametrization, (ρ^n, q^n) is τ -compact.

Theorem 2.5. Let $(\rho^n, q^n) \in \mathfrak{D}$ be a sequence τ -converging to (ρ, 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. For every $\mu_0, \mu_1 \in \mathcal{P}(\Omega)$, the minimization problem (2.5) admits a solution.

Remark 2.7. We remark that, for some choices of the data μ_0, μ_1 and of the exponent α , the statement of Theorem 2.6 could be empty, because the functional \mathcal{F} could be constantly $+\infty$ on every admissible path (ρ, q) joining μ_0 to μ_1 . This issue will be solved in Section 4, where the equivalence with other variational models for branched transportation will be proven. 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 μ_0 and μ_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 μ_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

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

due to the sub-additivity 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

$$F(\rho_t, q_t) \geq \|v_t\|_{L^{1/\alpha}(\rho_t)}^\alpha = |q_t|(\Omega). \tag{3.2}$$

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 (ρ, 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 (ρ^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

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

After that, we only need to prove compactness for the weak convergence of measures on $[0, 1] \times \Omega$, a fact which only requires bounds on the mass 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 mass of q by C .

This allows to extract a subsequence $(\tilde{\rho}_t^{n_k}, \tilde{q}_t^{n_k})$ which converges weakly to a pair (ρ, q) . The only nontrivial point is that we a priori restricted our attention to pairs (ρ, 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 ρ 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 ρ .

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. ρ , with an L^p density. Since ρ 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 (ρ, q) which is the τ -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 (ρ^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 (ρ^n, q^n) τ -converges to (ρ, q) .

First of all 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 ρ_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 can assume $\mathcal{F}(\rho^n, q^n)$ to be bounded. This 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, since we have $\mathbf{m}^n([a, b] \times \Omega) = \int_a^b F(\rho_t^n, q_t^n) dt$ and $F(\rho_t^n, q_t^n) \leq C$ (by definition of τ -convergence), we know that the marginal of \mathbf{m}^n on the time variable is a measure with an L^∞

density bounded by the constant C on $[0, 1]$. This bound is uniform and will be then satisfied by the limit measure \mathbf{m} too: in particular, this implies that we can write $\mathbf{m} = \int \mathbf{m}_t dt$.

Let us fix a closed set Q , as well as a time interval $[a, b]$, and 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 χ_M is positive, takes the value 1 on Q , is M -Lipschitz and vanishes outside a $1/M$ -neighborhood of Q . Indicating by 1_E the characteristic function of a generic set E , i.e. the function which 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

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

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 in the end

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

In this way we have

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

Hence, we may go on with

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

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 l.s.c. 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 for a while M fixed) and pass to the limit as $b \rightarrow a$. This gives, for \mathcal{L}^1 -a.e. $a \in [0, 1]$,

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

We let now $M \rightarrow \infty$, so that χ_M monotonically converges to the characteristic function of the set Q , and we have, by dominated convergence w.r.t. \mathbf{m}_a, ρ_a and $|q_a|$,

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

In the last term the convention $0 \cdot \infty = 0$ is used (if $|q_a|(Q) = 0$). This inequality is proven 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, 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 Ω . We want now to prove that

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

This would conclude the proof.

The first statement follows from the inequality (3.1), and the behavior of the Benamou-Brenier functional $\mathcal{A}_{1/\alpha}(\rho^n, q^n)$ as in the proof of Theorem 2.4, which guarantees an $L^{1/\alpha}$ density of q w.r.t. ρ . As a consequence, since ρ 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 $\mathbf{m}_a(S) \geq \varepsilon \rho_a(S)^\alpha$. 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 λ is atomic (see Lemma 3.1 below). If the same is performed for every $\varepsilon = 1/k$, this proves that ρ_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 infer that $F(\rho_a, q_a) = \sum_i \rho_a(\{x_i\})^{\alpha-1} |q_a|(\{x_i\})$ and we only need to consider $Q = \{x_i\}$ in (3.3) and add up:

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

which finally concludes the proof.

Lemma 3.1. *Take two finite positive measures λ and μ on a domain Ω , 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 λ is purely atomic.*

Proof. Consider a regular grid on Ω of step $1/k$, for $k \in \mathbb{N}$, and build a measure λ_k by putting, in every cell of the grid, all the mass of λ in a single point of the cell. This measure λ_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 goes to ∞ , the step of the grid goes to zero and we obviously have $\lambda_k \rightarrow \lambda$. On the other hand, the functional G_α 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, λ is atomic, thus proving the assertion. \square

3.3 Proof of Theorem 2.6

In order to prove existence, one only needs to take a minimizing sequence and apply Theorem 2.4 to get a new minimizing sequence which is τ -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 (ρ, q) will satisfy the same constraints as well. The semicontinuity proven in Theorem 2.5 allows to obtain the existence of a solution.

4 Equivalence with previous models

In this section we prove the equivalence of problem (2.5) with 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 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 with respect to 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 point out that this model is completely equivalent to the one developed by Xia (see [27] 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 μ_0 and μ_1 . In particular, thanks to this relaxed formulation, we get that for every μ_0 and μ_1 , there exist two sequences μ_0^n and μ_1^n of finitely atomic probability measures, weakly converging to μ_0 and μ_1 respectively, and such that

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

We also need to consider a slight modification of the functional E_α 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_α considered all the curves passing eventually through the point: in this sense, the model corresponding to the energy C_α is more dynamical in spirit. As a straightforward consequence of the definition of the two multiplicities, we get

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

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

Theorem 4.2. *Let $\mu_0, \mu_1 \in \mathcal{P}(\Omega)$, with μ_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_α .

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

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

Proof. We first prove the inequality $\mathcal{B}_\alpha(\mu_0, \mu_1) \geq d_\alpha(\mu_0, \mu_1)$. Clearly, if $\mathcal{B}_\alpha(\mu_0, \mu_1) = +\infty$ there is nothing to prove; otherwise, take (ρ, 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

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

Take $Q \in TP(\mu_0, \mu_1)$ optimal for C_α , then we know that there exists a pair (ρ, 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 with respect to the evaluation function e_t (see [20] 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

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

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

Summarizing, 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 equality holds whenever μ_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 μ_0^n and μ_1^n of finitely atomic probability measures such that $\mu_0^n \rightarrow \mu_0$, $\mu_1^n \rightarrow \mu_1$ and

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

thus getting

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

hence concluding the proof. \square

Remark 4.4. Observe that in the previous Theorem, we did not only prove 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.

Appendix: the distances d_α and $W_{1/\alpha}$

This last section is devoted to estimates between the distance d_α induced by the branched transport and the Wasserstein distances W_p . In particular, in [23] the following estimates are proven for $\alpha > 1 - 1/d$ and $p \geq 1$:

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

As far as lower bounds on d_α are concerned, the most trivial one is $d_\alpha \geq W_1$ but [16], 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 go beyond $p = 1/\alpha$ 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 a particular case of those that are already known. We restrict our attention to $p = 1/\alpha$ in the first one and $p \geq 1/\alpha$ in the second one (as we said, they are proven in [16, 23]), but the proof we will provide is different and somehow simpler.

The first inequality will be approached through the formulation of branched transport we gave in this paper, but the main tool (i.e. inequality (3.1)) is essentially in common with [16] and [22]. What is different is the way to extend this idea to generic measures, i.e. non-atomic ones.

Theorem 4.5. *For every $\mu_0, \mu_1 \in \mathcal{P}(\Omega)$ we get*

$$W_{1/\alpha}(\mu_0, \mu_1) \leq d_\alpha(\mu_0, \mu_1). \tag{4.5}$$

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 (ρ, q) admissible in the formulation (2.5), with $q = \rho v$. Moreover, using once more the inequality (3.1) 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 (ρ, q) is solution of the continuity equation, we can infer (see [2], Theorem 8.3.1)

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

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 (4.5), concluding the proof. \square

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

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

and 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 (Proposition 6.6, [6]).

Proposition 4.6. *Let $\alpha \in (1 - 1/d, 1]$, then for every $\mu \in \mathcal{P}(\Omega)$ we have*

$$d_\alpha(a_j(\mu), \mu) \leq \frac{2^{(d(1-\alpha)-1)j}}{2^{1-d(1-\alpha)} - 1} \frac{L\sqrt{d}}{2}. \quad (4.6)$$

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

Lemma 4.7. *Let $\alpha \in (1 - 1/d, 1]$, then for every $\mu_0, \mu_1 \in \mathcal{P}(\Omega)$ we get*

$$d_\alpha(a_j(\mu_0), a_j(\mu_1)) \leq C W_{1/\alpha}(a_j(\mu_0), a_j(\mu_1)) 2^{jd(1-\alpha)}, \quad (4.7)$$

with C depending only on N and α .

Proof. Let us consider an optimal transport γ_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 [15]). 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

$$\#\{(i, k) : M_j(i, k) \neq 0\} \leq 2 \cdot 2^{jd}, \quad (4.8)$$

that is $\{M_j(i, k)\}_{i,k}$ has at most $2 \cdot 2^{jd}$ non-zero entries: in other terms, this optimal transport plan γ_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 (4.8) and Jensen's inequality

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

concluding the proof. □

Theorem 4.8. *Let $\alpha \in (1 - 1/d, 1]$, then for every $p \geq 1/\alpha$, we get*

$$d_\alpha(\mu_0, \mu_1) \leq C W_p(\mu_0, \mu_1)^{d(\alpha-1)+1}, \quad (4.9)$$

with a constant C depending only on d , α and on the diameter of Ω .

Proof. It is enough to show the validity of (4.9) for $p = 1/\alpha$, then the general case will be just a consequence of the monotonicity property of the Wasserstein distances, i.e.

$$W_{1/\alpha} \leq W_p, \quad \text{for every } p \geq 1/\alpha.$$

Using the triangular inequality, (4.6) and (4.7), 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 4.9. As we briefly mentioned, observe that the distances d_α and $W_{1/\alpha}$ have exactly the same scaling with respect to the mass.

Remark 4.10. We point out that the very same μ_0 and μ_1 of Example 6.19 in [6] show that the exponent $d(\alpha - 1) + 1$ in inequality (4.9) cannot be improved.

Acknowledgements. The authors acknowledge the support of *Agence Nationale de la Recherche* via the research project OTARIE, of 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” as well of the *Università di Pisa* through the program *Cooperazione Accademica Internazionale* “Optimal transportation and related topics”. The first author has been partially supported by the European Research Council under FP7, Advanced Grant n. 226234 “Analytic Techniques for Geometric and Functional Inequalities”.

References

- [1] L. Ambrosio: *Lecture notes on optimal transport problems*, in “Mathematical aspects of evolving interfaces” (Funchal, 2000), Lecture Notes in Mathematics **1812**, Springer, Berlin, 2003.

- [2] L. Ambrosio, N. Gigli, 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] J. D. Benamou, Y. Brenier: *A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem.* Numer. Math., **84** (2000), 375–393.
- [4] M. Bernot, V. Caselles, J. M. Morel: *Traffic plans.* Publ. Mat., **49** (2005), 417–451.
- [5] M. Bernot, V. Caselles, J. M. Morel: *The structure of branched transportation networks.* Calc. Var. Partial Differential Equations, **32** (2008), 279–371.
- [6] M. Bernot, V. Caselles, J. M. Morel: *Optimal transportation networks. Models and theory.* Lecture Notes in Mathematics **1955**, Springer-Verlag, Berlin, 2009.
- [7] M. Bernot, A. Figalli: *Synchronized traffic plans and stability of optima.* ESAIM Control Optim. Calc. Var., **14** (2008), 864–878.
- [8] G. Bouchitté, G. Buttazzo: *New lower semicontinuity results for nonconvex functionals defined on measures.* Nonlinear Anal., **15** (1990), 679–692.
- [9] G. Bouchitté, P. Seppecher: paper in preparation.
- [10] A. Brancolini, G. Buttazzo, F. Santambrogio: *Path functionals over Wasserstein spaces.* J. Eur. Math. Soc., **8** (2006), 415–434.
- [11] L. Brasco: *Curves of minimal action over metric spaces.* Ann. Mat. Pura Appl., **189** (2010), 95–125.
- [12] L. Brasco, F. Santambrogio: *An equivalent path functional formulation of branched transportation problems.* Preprint 2009, available at <http://cvgmt.sns.it/people/brasco>.
- [13] Y. Brenier: *Extended Monge-Kantorovich theory.* In “Optimal Transportation and Applications” (Martina Franca, 2001), Lecture Notes in Mathematics **1813**, Springer-Verlag, Berlin, 2003, 91–121.
- [14] G. Buttazzo, C. Jimenez, E. Oudet: *An optimization problem for mass transportation with congested dynamics.* SIAM J. Control Optim., **48** (2009), 1961–1976.
- [15] J. L. Denny: *The support of discrete extremal measures with given marginals.* Michigan Math. J., **27** (1980), 59–64.
- [16] G. Devillanova, S. Solimini: *Elementary properties of optimal irrigation patterns.* Calc. Var. Partial Differential Equations, **28** (2007), 317–349.
- [17] J. Dolbeault, B. Nazaret, G. Savaré: *A new class of transport distances between measures.* Calc. Var. Partial Differential Equations, **34** (2009), 193–231.
- [18] E. N. Gilbert: *Minimum cost communication networks.* Bell System Tech. J., **46** (1967), 2209–2227.
- [19] J.-M. Lasry, P.-L. Lions: *Mean field games.* Japan J. Math., **2** (2007), 229–260.
- [20] S. Lisini: *Characterization of absolutely continuous curves in Wasserstein spaces.* Calc. Var. Partial Differential Equations, **28** (2007), 85–120.

- [21] F. Maddalena, J.M. Morel, S. Solimini: *A variational model of irrigation patterns*. Interfaces Free Bound., **5** (2003), 391–415.
- [22] F. Maddalena, S. Solimini: *Transport distances and irrigation models*. J. Convex Anal., **16** (2009), 121–152.
- [23] J. M. Morel, F. Santambrogio: *Comparison of distances between measures*. Appl. Math. Lett., **20** (2007), 427–432.
- [24] B. Nazaret, F. Santambrogio: *Non-linear mobility distances on probability measures: geodesics via PDE*. Paper in preparation.
- [25] C. Villani: *Optimal transport, old and new*. Grundlehren der Mathematischen Wissenschaften **338**, Springer-Verlag, Berlin, 2009.
- [26] C. Villani: *Topics in optimal transportation*. Graduate Studies in Mathematics **58**, American Mathematical Society, Providence, RI, 2003.
- [27] Q. Xia: *Optimal paths related to transport problems*. Commun. Contemp. Math., **5** (2003), 251–279.