Absolute continuity and summability of transport densities: simpler proofs and new estimates

Filippo Santambrogio*

February 17, 2009

Abstract

The paper presents some short proofs for transport density absolute continuity and L^p estimates. Most of the previously existing results which were proven by geometric arguments are re-proved through a strategy based on displacement interpolation and on approximation by discrete measures; some of them are partially extended.

1 Introduction

Monge-Kantorovich theory of optimal transportation deals with the minimization

$$\min\left\{\int_{\Omega\times\Omega}|x-y|\,\gamma(dx,dy)\ :\ \gamma \text{ is a transport plan between } \mu_0 \text{ and } \mu_1\right\}, \tag{1.1}$$

where transport plans are those probabilities on $\Omega\times\Omega$ having μ_0 and μ_1 as marginal measures. In such a theory it is classical to associate to any optimal transport plan γ a positive measure σ on Ω, called transport density, which represents the amount of transport taking place in each region of Ω (Ω is, say, a bounded and convex subset of $\mathbb{R}^d$). This density σ is defined by

$$\langle\sigma,\phi\rangle := \int_{\Omega\times\Omega}\gamma(dx,dy)\int_0^1\phi(\omega_{x,y}(t))|\dot\omega_{x,y}(t)|dt \qquad \text{for all } \phi\in C^0(\Omega) \tag{1.2}$$

where $\omega_{x,y}$ is a curve parametrizing the straight line segment connecting x to y (the same could be generalized to other Riemannian distances than the euclidean one, and this segment should be replaced by a geodesic curve). Alternatively, if we look at the action of σ on sets, we have, for every Borel set A,

$$\sigma(A) := \int_{\Omega\times\Omega}\mathcal{H}^1(A\cap[[x,y]])\,\gamma(dx,dy),$$

where $[[x,y]]$ is the segment joining the two points x and y.

*CEREMADE, UMR CNRS 7534, Université Paris-Dauphine, Place du Maréchal de Lattre de Tassigny, 75775 Paris Cedex 16, FRANCE filippo@ceremade.dauphine.fr .

This positive measure σ is the total variation of a vector measure $\bar{\lambda}$ solving the problem

$$\min\left\{\int_\Omega |\lambda|(dx),\ :\ \lambda \in \mathcal{M}(\Omega;\mathbb{R}^d),\ \nabla\cdot\lambda = \mu_1 - \mu_0\right\}, \tag{1.3}$$

which is the so-called continuous transportation problem proposed by Beckmann in [5].

More precisely, for every transport plan γ we can build a vector measure λ, defined through

$$\langle \lambda, \phi\rangle := \int_{\Omega\times\Omega} \gamma(dx,dy)\int_0^1 \phi(\omega_{x,y}(t))\cdot\dot{\omega}_{x,y}(t)dt, \qquad \text{for all } \phi \in C^0(\Omega;\mathbb{R}^d)$$

and the $\bar{\lambda}$ associated to an optimal γ turns out to be optimal for (1.3). Thanks to our definitions, it is evident that we have $|\bar{\lambda}| \leq \sigma$, while the equality comes from the fact that transport rays cannot cross: if several segments involved by an optimal transport pass through the same point, than they all share the same direction.

One first natural question is whether the transport density σ is absolutely continuous. This would for instance allow to set the problem (1.3) in a L^1 setting instead of using the space $\mathcal{M}(\Omega)$ of finite vector measures on Ω. Notice that, to this aim, it would be sufficient to state that there exists an optimal transport plan γ such that the corresponding σ (or, equivalently, the corresponding λ) is absolutely continuous (it would not be necessary to prove it for every σ).

Actually, the precise relation between$\bar{\lambda}$ and σ is $\bar{\lambda} = \sigma\nabla u$, where u is a Kantorovich potential in the transportation from μ_0 to μ_1. The condition $\sigma << \mathcal{L}^d$ would also allow to write the system

$$\begin{cases} \nabla\cdot(\sigma\nabla u) = \mu_1 - \mu_0 & \text{in } \Omega \\ |\nabla u| \leq 1 & \text{in } \Omega, \\ |\nabla u| = 1 & \text{a.e. on } \sigma > 0, \end{cases} \tag{1.4}$$

without passing through the theory of $\sigma-$tangential gradient (see for instance [14] or [8]).

There are several papers, mainly by De Pascale and Pratelli, Evans and Feldmann and McCann, addressing absolute continuity and more general questions. In [13] the authors show estimates on the dimension of σ in terms of the dimension of μ_0 and μ_1, and they get in particular $\sigma << \mathcal{L}^d$ whenever one of the two source measures μ_0 or μ_1 is absolutely continuous. In the same paper they also give several L^p estimates, which are then strengthened in [14] and in [15], where they finally get the important result

$$\mu_0, \mu_1 \in L^p \Rightarrow \sigma \in L^p \qquad \text{for all } p \in [1, +\infty]. \tag{1.5}$$

Among the other L^p results, [13] proves

$$\mu_0 \in L^p \Rightarrow \sigma \in L^q \qquad \text{for all } q < \min\left\{(2d)', 1 + \frac{p-1}{2}\right\}. \tag{1.6}$$

Estimates on σ may have various applications: for instance, lower bounds could be used to retrieve information on the behavior of u as a solution of (1.4) (and, in order to apply standard elliptic theory, L^∞ estimates as well are needed). Not only, lower bounds would also be useful to prove some density results on the transport set: the techniques developed by Champion and De Pascale in [11] would allow in such a case to derive the existence of an optimal transport map in Monge's problem. To mention, on the other hand, applications of upper bounds and L^p estimates on the transport density, we refer to [10], where these results are used to prove well-posedness for congestion problems and continuous Wardrop equilibria.

Coming back to the simple question of absolute continuity, Feldmann and McCann proved in [16] that there exists, unique, an L^1 transport density for L^1 sources. This is another absolute continuity result and it is coupled with a uniqueness result which is stated as "two L^1 functions which are transport densities, in a certain sense, between the same L^1 sources must coincide". A more complete uniqueness result may be found in the Lecture Notes by Ambrosio [1], where the links between σ and the other formulations of the Monge transport problem are well underlined. The proof of the uniqueness in [1] is based on a decomposition into transport rays and on a one-dimensional result. Here the result reads as "two different optimal transport plans always induce the same measure σ, provided at least one of the two source measures is in L^1". In the same lecture notes, for the absolute continuity proof the reader is addressed to [13].

Yet, we must say that the proofs in [13, 14, 15] and [16] are quite complicated and long. This is natural since they are actually the first pioneering works on transport density; moreover they present much wider results (dimensional estimates, existence of the limit of the cost on a ball, uniqueness. . .).

What we propose here are very simple proofs for a series of results which are partially already known. The starting point is a proof of $\sigma << \mathcal{L}^d$ that arose during the preparation of a course on Optimal Transport at IHP in Paris. One lecture of the course was devoted to divergence-constrained problems and the goal was to show well-posedness in L^1.

The strategy of the proof passes through the absolute continuity of the interpolation measures μ_t. To prove that these measures are absolutely continuous we pass through the discrete case and then get it at the limit. Actually, the absolute continuity of the interpolations is well known, especially in the case of strictly convex cost function (rather than $|x-y|$). In this last case it is explicitly stated in Theorem 8.7 of Part I of the new book on Optimal Transport by Cedric Villani, [19], where a general Lipschitz result on intermediate transport maps (Theorem 8.5) is used. Yet, here, before passing to the limit, the maps are actually piecewise linear and this obviously allows easier computations to be performed. As far as the distance case $|x - y|$ is concerned, there exists a proof of the same absolute continuity estimates which is presented in [7]. Yet, no application to transport densities is presented, even if this is the most natural framework to apply those estimates passing to the limit from the discrete case, since we know that transport density is independent of the particular transport plan which is selected by the approximation.

Later on, it appeared quite easily that the same technique we use here could be used for L^p estimates. This gives easily a stronger result than (1.6). On the contrary, for getting (1.5) we need something more, and precisely we need to guarantee that we can obtain a precise optimal transport plan as a limit of optima in discrete cases, since we need a double approximation and we want to be sure that the same plan is selected. But once this is done (in Section 3.1, thanks to techniques mimicking $\Gamma-$convergence developments, see [12]), we also get $L^p - L^q - L^r$ estimates (Theorem 6), i.e. something in the spirit of "$\mu_0 \in L^p,\ \mu_1 \in L^q \Rightarrow \sigma \in L^r,\ r = r(p,q,d)$". This kind of estimates (higher integrability if one of the measures belongs to a better space L^p) seems to be novel.

The goal of this paper is to prove these results on transport densities just by using some basic facts about optimal transports. It is conceived to be read by somebody who knows the main properties of optimal transport but is not necessarily an expert of transport densities (exactly as the author is). Moreover, for the sake of backgroundlessness, when possible we sticked to the discrete approximation technique instead of using more advanced tools (for instance an alternative proof for (1.5) which is shorter for the reader who is familiar with displacement convexity is only presented in Remark 9 and it is not our main proof).

No one of the ideas of these proofs occupies more than one page, and the reader has not to be scared by this long introduction, nor by the long bibliography. Actually, the only bibliographical references which present specific useful tools for the proofs are [1], [2], [4] and [6], while we refer to [19], [18] and [12] for the general references on transport, displacement convexity and $\Gamma-$convergence; [5, 13, 14, 15, 16] are the main references for transport densities; [3, 7, 9, 17, 19] are cited as papers developing similar techniques in different contests and [5, 8, 10, 11] are presented in view of possible applications of these estimates.

2 One-sided estimates

In all that follows Ω is a compact and convex domain in $\mathbb{R}^d$, μ_0 and μ_1 two probabilities on Ω, and at least one of them will be absolutely continuous. Then, by Theorem 7.3 in [1], there will exist one unique transport density σ associated to those measures, independent of the optimal transport plan γ (the relation between σ and any γ is given in (1.2)).

2.1 Absolute continuity: $\mu_0 << \mathcal{L}^d \Rightarrow \sigma << \mathcal{L}^d$

Theorem 1. *Suppose $\mu_0 << \mathcal{L}^d$, and let σ be the unique transport density associated to the transport of μ_0 onto μ_1. Then $\sigma << \mathcal{L}^d$.*

Proof. Let γ be an optimal transport from μ_0 to μ_1 (recall that we have the right to choose any particular optimal transport plan, if needed, since σ does not depend on the choice of γ) and μ_t the standard interpolation between the two measures: $\mu_t = (\pi_t)_\#\gamma$ where $\pi_t(x,y) = (1-t)x + ty$ (this is the same, in this framework, as $\omega_{x,y}(t)$, when the segments are parametrized at constant speed).

The transport density σ may be easily written as (see [1] or look at the definition (1.2))

$$\sigma = \int_0^1 (\pi_t)_\sharp(c \cdot \gamma)dt,$$

where $c : \Omega \times \Omega \to \mathbb{R}$ is the cost function $c(x,y) = |x-y|$ (hence $c \cdot \gamma$ is a positive measure on $\Omega \times \Omega$).

Since Ω is bounded it is evident that we have

$$\sigma \leq C \int_0^1 \mu_t \, dt. \tag{2.1}$$

To prove that σ is absolutely continuous, it is sufficient to prove that almost every measure μ_t is absolutely continuous, so that, whenever $|A| = 0$, we have $\sigma(A) \leq C \int_0^1 \mu_t(A)dt = 0$.

We will prove $\mu_t << \mathcal{L}^d$ for $t < 1$. First, we will suppose that μ_1 is finitely atomic (the point $(x_i)_{i=1,\dots,N}$ being its atoms). In this case we will choose γ to be any optimal transport plan induced by a transport map T (which exists, since $\mu_0 << \mathcal{L}^d$). Notice that the absolute continuity of σ is an easy consequence of the behavior of the optimal transport from μ_0 to μ_1 (which is composed by N homotheties), but we also want to quantify this absolute continuity, in order to go on with an approximation procedure.

Remember that μ_0 is absolutely continuous and hence there exists a correspondence $\varepsilon \mapsto \delta = \delta(\varepsilon)$ such that

$$|A| < \delta(\varepsilon) \Rightarrow \mu_0(A) < \varepsilon. \tag{2.2}$$

Take now a Borel set A and look at $\mu_t(A)$. The domain Ω is the disjoint union of a finite number of sets $\Omega_i = T^{-1}(\{x_i\})$. We call $\Omega_i(t)$ the images of Ω_i through the map $x \mapsto (1-t)x + tT(x)$. These sets are essentially disjoint. Why? because if a point z belongs to $\Omega_i(t)$ and $\Omega_j(t)$, then two transport rays cross at z, the one going from $x'_i \in \Omega_i$ to x_i and the one from $x'_j \in \Omega_j$ to x_j. The only possibility is that these two rays are actually the same, i.e. that the five points x'_i, x'_j, z, x_i, x_j are aligned. But this implies that z belongs to one of the lines connecting two atoms x_i and x_j. Since we have finitely many of these lines this set is negligible. Notice that this argument only works for $d > 1$ (we will not waste time on the case $d = 1$, since the transport density is always a BV and hence bounded function). Moreover, if we sticked to the optimal transport which is monotone on transport rays, we could have actually proved that these sets are truly disjoint, with no negligible intersection (see the proof of Lemma 4 and the subsequent remark).

Hence we have

$$\mu_t(A) = \sum_i \mu_t(A \cap \Omega_i(t)) = \sum_i \mu_0\left(\frac{A \cap \Omega_i(t) - tx_i}{1-t}\right) = \mu_0\left(\bigcup_i \frac{A \cap \Omega_i(t) - tx_i}{1-t}\right).$$

Since for every i we have

$$\left|\frac{A \cap \Omega_i(t) - tx_i}{1-t}\right| = \frac{1}{(1-t)^d}\,|A \cap \Omega_i(t)|$$

we have

$$\left|\bigcup_i \frac{A \cap \Omega_i(t) - tx_i}{1-t}\right| \leq \frac{1}{(1-t)^d}|A|.$$

Hence it is sufficient to suppose $|A| < (1-t)^d\delta(\varepsilon)$ to get $\mu_t(A) < \varepsilon$. This confirms $\mu_t << \mathcal{L}^d$ and gives an estimate that may pass to the limit.

Take a sequence $(\mu_1^n)_n$ of atomic measures converging to μ_1. The corresponding optimal transport plans γ^n converge to an optimal transport plan γ and μ_t^n converge to the corresponding μ_t. Hence, to prove absolute continuity for the transport density σ associated to such a γ it is sufficient to prove that these μ_t are absolutely continuous.

Take a set A such that $|A| < (1-t)^d\delta(\varepsilon)$. Since the Lebesgue measure is regular, A is included in an open set B such that $|B| < (1-t)^d\delta(\varepsilon)$. Hence $\mu_t^n(B) < \varepsilon$. Passing to the limit, thanks to weak convergence and semicontinuity on open sets, we have

$$\mu_t(A) \le \mu_t(B) \le \liminf_n \mu_t^n(B) \le \varepsilon.$$

This proves $\mu_t << \mathcal{L}^d$ and hence $\sigma << \mathcal{L}^d$. □

Remark 1. Where did we use the optimality of γ? we did it when we said that the $\Omega_i(t)$ are disjoint. For a discrete measure μ_1, it is always true that the measures μ_t corresponding to any transport plan γ are absolutely continuous for $t < 1$, but their absolute continuity may degenerate at the limit if we allow the sets $\Omega_i(t)$ to superpose (since in this case densities sum up and the estimates may depend on the number of atoms).

Remark 2. Notice that we strongly used the equivalence between the two different definitions of absolute continuity, i.e. the $\varepsilon \leftrightarrow \delta$ correspondence on the one hand and the condition on negligible sets on the other. Indeed, to prove that the condition $\mu_t << \mathcal{L}^d$ passes to the limit we need the first one, while to deduce $\sigma << \mathcal{L}^d$ we need the second one, since if we deal with non-negligible sets we have some $(1-t)^d$ factor to deal with...

Remark 3. If we did not know any uniqueness result on σ we could have replaced the statement of Theorem 1 with "there exists a transport density σ which is absolutely continuous". This would have been enough for several aims concerning the variational problem (1.3) or System (1.4)).

Remark 4. Where did we use the optimality of γ? we did it when we said that the $\Omega_i(t)$ are disjoint. For a discrete measure μ_1, it is always true that the measures μ_t corresponding to any transport plan γ are absolutely continuous for $t < 1$, but their absolute continuity may degenerate at the limit if we allow the sets $\Omega_i(t)$ to superpose (since in this case densities sum up and the estimates may depend on the number of atoms).

Remark 5. Last remark: notice that we built an optimal transport plan with the property of having absolutely continuous interpolating measures, and this was the key point in the proof. Notice that this property is not satisfied by any optimal transport plan, since for instance the γ which sends $\mu_0 = \mathcal{L}^2_{|[-2,-1]\times[0,1]}$ onto $\mu_1 = \mathcal{L}^2_{|[1,2]\times[0,1]}$ moving (x,y) to $(-x,y)$ is such that $\mu_{1/2} = \mathcal{H}^1_{|\{0\}\times[0,1]}$. This answers negatively to a natural question raised by L. De Pascale: "is any optimal transport plan from μ_0 to μ_1 approximable through optimal transport plans from μ_0 to atomic measures?".

2.2 Higher summability: $\mu_0 \in L^p \Rightarrow \sigma \in L^p,\ p < d/(d-1)$

From this section on we will often confuse absolutely continuous measures with their densities and write $||\mu||_p$ for $||f||_{L^p(\Omega)}$ when $\mu = f \cdot \mathcal{L}$.

Theorem 2. *Suppose $\mu_0 = f \cdot \mathcal{L}^d$, with $f \in L^p(\Omega)$. The, if $p < d' := d/(d-1)$, the unique transport density σ associated to the transport of μ_0 onto μ_1 belongs to $L^p(\Omega)$ as well, and if $p \ge d'$ it belongs to any space $L^q(\Omega)$ for $q < d'$.*

Proof. Start from the case $p < d'$: following the same strategy (and the same notations) as before, it is sufficient to prove that each measure μ_t (for $t \in [0, 1[$) is in L^p and to estimate their L^p norm. Then we will use

$$||\sigma||_p \le C \int_0^1 ||\mu_t||_p dt,$$

(which is a consequence of (2.1) and of Minkowski inequality), the conditions on p being chosen exactly so that this integral converges.

Consider first the discrete case: we know that μ_t is absolutely continuous and that its density coincides on each set $\Omega_i(t)$ with the density of an homothetical image of μ_0 on Ω_i, the homothety ratio being $(1-t)$. Hence, if f_t is the density of μ_t, we have

$$\begin{aligned} \int_\Omega f_t(x)^p dx &= \sum_i \int_{\Omega_i(t)} f_t(x)^p dx = \sum_i \int_{\Omega_i} \left(\frac{f(x)}{(1-t)^d} \right)^p (1-t)^d dx \\ &= (1-t)^{d(1-p)} \sum_i \int_{\Omega_i} f(x)^p dx = (1-t)^{d(1-p)} \int_\Omega f(x)^p dx. \end{aligned}$$

We get $||\mu_t||_p = (1-t)^{-d/p'} ||\mu_0||_p$, where $p' = p/(p-1)$ is the conjugate exponent of p.

This inequality, which is true in the discrete case, stays true at the limit as well. If μ_1 is not atomic, approximate it through a sequence μ_1^n and take optimal plans γ^n and interpolating measures μ_t^n. Up to subsequences we have $\gamma^n \rightharpoonup \gamma$ (for an optimal transport plan γ) and $\mu_t^n \rightharpoonup \mu_t$ (for the corresponding interpolation); by semicontinuity we have

$$||\mu_t||_p \le \liminf_n ||\mu_t^n||_p \le (1-t)^{-d/p'} ||\mu_0||_p$$

and we deduce

$$||\sigma||_p \le C \int_0^1 ||\mu_t||_p dt \le C ||\mu_0||_p \int_0^1 (1-t)^{-d/p'} dt.$$

The last integral is finite whenever $p' > d$, i.e. $p < d' = d/(d-1)$.

The second part of the statement (the case $p \ge d'$) is straightforward once one considers that any density in L^p also belongs to any L^q space for $q < p$. □

Remark 6. This result improves the one-sided estimate in [13]: the upper bound on the valid exponent p is d' instead of $(2d)'$ and, moreover, we prove that σ belongs to the same L^p space of μ_0 (while in [13] for $\mu_0 \in L^p$ no estimate beyond $L^{1+(p-1)/2}$ was given for σ).

Remark 7. Here and in the following one could use a different strategy, looking for quantitative estimates of the kind of (2.2). Actually, inequalities such as $\sigma(A) \le |A|^\alpha$ imply L^p estimates and one could get this kind of estimates for the measures μ_t by the techniques of Theorem (1). Yet, this approach seems to be weaker than the L^p one and we will not develop it any more.

3 Two-sided estimates

We saw in the previous section that the measures μ_t inherit some regularity (absolute continuity or L^p summability) from μ_0 exactly as it happens for homotheties of ratio $1-t$. This regularity degenerates as $t \to 1$, but we saw two cases where this degeneracy produced no problem: for proving absolute continuity, where the separate absolute continuous behavior of almost all the μ_t was sufficient, and for L^p estimates, provided the degeneracy stays integrable.

It is natural to try to exploit another strategy: suppose both μ_0 and μ_1 share some regularity assumption (e.g., they belong to L^p). Then we can give estimate on μ_t for $t \le 1/2$ starting from μ_0 and for $t \ge 1/2$ starting from μ_1. In this way we have no degeneracy!

This strategy works quite well, but it has an extra difficulty: in our previous estimates we didn't know a priori that μ_t shared the same behavior of piecewise homotheties of μ_0, we got it as a limit from discrete approximations. And, when we pass to the limit, we do not know which optimal transport γ will be selected as a limit of the optimal plans γ^n. This was not important in the previous section, since any optimal γ induces the same transport density σ. Yet, here we would like to glue together estimates on μ_t for $t \le 1/2$ which have been obtained by approximating μ_1, and estimates on μ_t for $t \ge 1/2$ which come from the approximation of μ_0. Should the two approximations converge to two different transport plans, we could not put together the two estimates and deduce anything on σ.

Hence, the main technical issue of this section will be proving that one particular optimal transport plan, i.e. the one which is monotone on transport rays, will be approximable in both directions. We will exhibit, thanks to a variational approximation in the spirit of $\Gamma-$convergence developments (see [12]), a sequence of properly chosen atomic measures, with their corresponding optimal transport plans, which do the job. Yet, the transport plans we will use in the approximation will not be optimal for the cost $\int |x-y| d\gamma$ but for some costs $\int (|x-y| + \varepsilon |x-y|^2) d\gamma$. We need to do this in order to force the selected limit optimal transport to be the monotone one (through a secondary variational problem, say). Anyway, this will not be an issue since these approximating optimal transport will share the same geometric properties that will imply disjointness for the sets $\Omega_i(t)$ will allow for density estimates.

In the whole section all constants C in the estimates could eventually be used to denote larger constants and their value could possibly change from one line to another.

Lastly, even if not precisely stated, the reader will be easily be able to check that all the results of this section stay true for $p = +\infty$ as well.

3.1 Discrete approximation of the ray-monotone optimal transport

For fixed measures $\mu, \nu \in \mathcal{P}(\Omega)$, consider the following family of minimization problems (P_ε):

$$(P_\varepsilon) = \min \left\{ W_1((\pi_1)_\# \gamma, \nu) + \varepsilon C_1(\gamma) + \varepsilon^2 C_2(\gamma) + \varepsilon^{3d+3} \#((\pi_1)_\# \gamma), \, : \, \gamma \in \mathcal{P}(\Omega \times \Omega), \; (\pi_0)_\# \gamma = \mu \right\},$$

where W_1 is the usual Wasserstein distance, i.e. the minimum value of the transport problem for the cost $c(x,y) = |x-y|$, $C_p(\gamma) = \int |x-y|^p \gamma(dx, dy)$ for $p = 1, 2$ and the symbol $\#$ denotes the cardinality of the support of a measure. Concerning W_1, we only need to know that it is a distance (i.e. it is positive, it satisfies the triangle inequality and it vanishes only when the two measures coincide) and that it metrizes the usual weak convergence of probability measures on compact domains (see Chapter 6 in [19]).

Actually, this minimization consists in looking for a transport plan with first marginal equal to μ satisfying the following criteria with decreasing degree of importance: the second marginal must be close to ν, the C_1 cost of transportation should be small, the C_2 as well, and, finally, the second marginal must be atomic with not too many atoms.

This minimization problem has obviously at least a solution (by the direct method, being Ω compact). We call γ_ε such a solution and $\nu_\varepsilon := (\pi_1)_\#\gamma_\varepsilon$ its second marginal. It is straightforward that ν_ε is an atomic measure and that γ_ε is the (unique, if $\mu << \mathcal{L}^d$, since the cost is strictly convex) optimal transport from μ to ν_ε for the cost $C_1 + \varepsilon C_2$. Set

$$\bar{\gamma} = \operatorname{argmin}\left\{C_2(\gamma) \ : \ \gamma \text{ is a } C_1\text{−optimal transport plan from } \mu \text{ to } \nu\right\}. \tag{3.1}$$

This transport plan $\bar{\gamma}$ is unique and it is known to be the unique optimal transport plan from μ to ν which is monotone on transport rays (see for instance [3, 4, 6, 11]; notice that the functional C_2 could have been replaced by any functional $\gamma \mapsto \int \phi(x-y)d\gamma$ for a strictly convex function ϕ).

Lemma 3. *As $\varepsilon \to 0$ we have $\nu_\varepsilon \rightharpoonup \nu$ and $\gamma_\varepsilon \rightharpoonup \bar{\gamma}$.*

Proof. It is sufficient to prove that any possible limit of subsequences coincide with ν or $\bar{\gamma}$, respectively. Let γ_0 be one such a limit and $\nu_0 = (\pi_1)_\#\gamma_0$ the limit of the corresponding subsequence of ν_ε. Moreover, let p_n be any measurable map from Ω to a grid $G^n \subset \Omega$ composed of Cn^d points, with the property $|p_n(x) - x| \le 1/n$. Set $\nu^n := (p_n)_\#\nu$ and notice $\#\nu^n \le Cn^d$, as well as $\nu^n \rightharpoonup \nu$.

First step: $\nu_0 = \nu$. Take γ^n any transport plan from μ to ν^n. By optimality of γ_ε we have

$$W_1(\nu_\varepsilon, \nu) \le W_1(\nu^n, \nu) + \varepsilon C_1(\gamma^n) + \varepsilon^2 C_2(\gamma^n) + C\varepsilon^{3d+3}n^d.$$

Fix n, let ε go to 0 and get

$$W_1(\nu_0, \nu) \le W_1(\nu^n, \nu) \le \frac{1}{n}.$$

Then let $n \to \infty$ and get $W_1(\nu_0, \nu) = 0$, which implies $\nu_0 = \nu$.

Second step: γ_0 is optimal for C_1 from μ to ν. Take any optimal transport plan γ^n (for the C_1 cost) from μ to ν^n. These plans converge to a certain optimal plan $\tilde{\gamma}$ from μ to ν. Then, by optimality, we have

$$\varepsilon C_1(\gamma_\varepsilon) \le W_1(\nu^n, \nu) + \varepsilon C_1(\gamma^n) + \varepsilon^2 C_2(\gamma^n) + C\varepsilon^{3d+3}n^d \le \frac{1}{n} + \varepsilon C_1(\gamma^n) + C\varepsilon^2 + C\varepsilon^{3d+3}n^d.$$

Then take $n \approx \varepsilon^{-2}$ and divide by ε :

$$C_1(\gamma_\varepsilon) \le \varepsilon + C_1(\gamma^n) + C\varepsilon + C\varepsilon^{d+2}.$$

Passing to the limit we get

$$C_1(\gamma_0) \le C_1(\tilde{\gamma}) = W_1(\mu, \nu),$$

which implies that γ_0 is optimal.

Third step: $\gamma_0 = \bar{\gamma}$**.** Take any optimal transport plan γ (for the cost C_1) from μ to ν. Set $\gamma^n = (id \times p_n)_{\#}\gamma$. We have $(\pi_1)_{\#}\gamma^n = \nu^n$. Then we have

$$W_1(\nu_\varepsilon, \nu) + \varepsilon C_1(\gamma_\varepsilon) + \varepsilon^2 C_2(\gamma_\varepsilon) \le W_1(\nu^n, \nu) + \varepsilon C_1(\gamma^n) + \varepsilon^2 C_2(\gamma^n) + C\varepsilon^{3d+3} n^d.$$

Moreover we have

$$C_1(\gamma_\varepsilon) \ge W_1(\mu, \nu_\varepsilon) \ge W_1(\mu, \nu) - W_1(\nu_\varepsilon, \nu),$$
$$C_1(\gamma^n) \le C_1(\gamma) + \int |p_n(y) - y| \gamma(dx, dy) \le C_1(\gamma) + \frac{1}{n} = W_1(\mu, \nu) + \frac{1}{n}.$$

Hence we have

$$(1-\varepsilon) W_1(\nu_\varepsilon, \nu) + \varepsilon W_1(\mu, \nu) + \varepsilon^2 C_2(\gamma_\varepsilon) \le \frac{1}{n} + \varepsilon W_1(\mu, \nu) + \frac{\varepsilon}{n} + \varepsilon^2 C_2(\gamma^n) + C\varepsilon^{3d+3} n^d.$$

Getting rid of the first term (which is positive) in the left hand side, simplifying $\varepsilon W_1(\mu, \nu)$, and dividing by ε^2, we get

$$C_2(\gamma_\varepsilon) \le \frac{1+\varepsilon}{n\varepsilon^2} + C_2(\gamma^n) + C\varepsilon^{3d+1} n^d.$$

Here it is sufficient to take $n \approx \varepsilon^{-3}$ and pass to the limit to get

$$C_2(\gamma_0) \le C_2(\gamma),$$

which is the condition characterizing $\bar{\gamma}$ (C_2-optimality among C_1-minimizers). □

3.2 L^p summabilities

The first tool we need is a uniform L^p estimates of the measures μ_t in terms of the norm of μ_0, when μ_t is an interpolation from μ_0 to μ_1 corresponding to a transport plan γ which is optimal for another cost, different from $|x-y|$. In this case we do not have any transport ray argument, but the result is somehow even stronger under strict convexity assumptions.

Lemma 4. *Let γ be an optimal transport plan between μ_0 and an atomic measure μ_1 for a transport cost $c(x, y) = \phi(y - x)$ where $\phi : \mathbb{R}^d \to \mathbb{R}$ is a strictly convex function. Set as usual $\mu_t = (\pi_t)_{\#}\gamma$. Then we have $||\mu_t||_p \le (1-t)^{-d/p'} ||\mu_0||_p$.*

Proof. The result is exactly the same as in Theorem 2, where the key tool is the fact that μ_t coincides on every set $\Omega_i(t)$ with an homothety of μ_0. The only fact that must be checked again is the disjointness of the sets $\Omega_i(t)$.

To do so, take a point $x \in \Omega_i(t) \cap \Omega_j(t)$. Hence there exist x_i, x_j belonging to Ω_i and Ω_j, respectively, so that $x = (1-t)x_i + ty_i = (1-t)x_j + ty_j$, being y_i and y_j atoms of μ_1. Set $a = y_i - x_i$ and $b = y_j - x_j$.

The $c-$cyclical monotonicity of the support of the optimal γ implies

$$\phi(a) + \phi(b) \le \phi(y_j - x_i) + \phi(y_i - x_j) = \phi(tb + (1-t)a) + \phi(ta + (1-t)b).$$

Yet, if $y_j \neq y_i$ we have $a \neq b$, and strict convexity implies

$$\phi(tb + (1-t)a) + \phi(ta + (1-t)b) < t\phi(b) + (1-t)\phi(a) + t\phi(a) + (1-t)\phi(b) = \phi(a) + \phi(b),$$

which is a contradiction. Hence the sets $\Omega_i(t)$ are disjoint and this implies the bound on μ_t. □

Remark 8. Disjointness of the sets $\Omega_i(t)$ is easier in this strictly convex setting. If the cost is $|x-y|$ this is no more true, but it is anyway true that the two vector a and b should be parallel, i.e. all the points should be aligned, as we pointed out in Theorem 1. If μ does not give mass to lines, than the sets are essentially disjoint. Otherwise one can say that they are truly disjoint if one only looks at the optimal transport which is monotone on transport rays.

Theorem 5. *Suppose that μ_0 and μ_1 are probability measures on Ω, both belonging to $L^p(\Omega)$, and σ the unique transport density associated to the transport of μ_0 onto μ_1. Then σ belongs to $L^p(\Omega)$ as well.*

Proof. Let us consider the optimal transport plan $\bar{\gamma}$ from μ_0 to μ_1 defined by (3.1). We know that this transport plan may be approximated by plans γ_ε which are optimal for the cost $|x-y|+\varepsilon|x-y|^2$ from μ_0 to some discrete atomic measures ν_ε. The corresponding interpolation measures $\mu_t(\varepsilon)$ satisfy the L^p estimate from Lemma 4 and, at the limit, we have

$$||\mu_t||_p \leq \liminf_{\varepsilon \to 0} ||\mu_t(\varepsilon)||_p \leq (1-t)^{-d/p'} ||\mu_0||_p.$$

The same estimate may be performed from the other direction, since the same transport plan $\bar{\gamma}$ may be approximated by optimal plans for the cost $|x-y| + \varepsilon|x-y|^2$ from atomic measures to μ_1. Putting together the two estimates we have

$$||\mu_t||_p \leq \min\left\{(1-t)^{-d/p'}||\mu_0||_p,\, t^{-d/p'}||\mu_1||_p\right\} \leq 2^{d/p'} \max\{||\mu_0||_p, ||\mu_1||_p\}.$$

Integrating these L^p norms we get the bound on $||\sigma||_p$. □

Remark 9. The same result could have been obtained in a strongly different way, thanks to the displacement convexity of the functional $\mu \mapsto ||\mu||_p^p$. This functional is actually convex along geodesics in the space $W_q(\Omega)$ for $q > 1$ (see Theorem 9.3.9 in [2]). This implies that, if we take an optimal transport plan for the cost $c_q(x,y) = |x-y|^q$, the interpolating measures μ_t satisfy $||\mu_t||_p \leq \max\{||\mu_0||_p, ||\mu_1||_p\}$. Then we pass to the limit as $q \to 1$: this gives the result on the interpolating measures corresponding to the optimal plan which is obtained as a limit of the c_q−optimal plans. This plan is, by the way, $\bar{\gamma}$ again. And the integral estimate comes straightforward.

Hence, why did we not present the proof through this approach? the reason is twofold. First, the spirit of the proof we presented (approximation by means of discrete transports) is shared by the other proofs of this paper, and the convexity principle could not have been adapted to the other results (to Theorems 2 and 6, for instance, because, in order to apply it, we need finite L^p norms at both sides). Second, the proof we presented uses directly elementary tools such as the discrete approximation, while Theorem 9.3.9 in [2] requires some more sophisticated tools in optimal transport, such as differentiability of the transport map, concavity properties of its Jacobian... For the goals of this paper, which aims to be as much expository as possible, since almost no new result is presented, we found it better to stick to the kind of proof we presented.

Theorem 6. *Suppose $\mu_0 \in L^p(\Omega)$ and $\mu_1 \in L^q(\Omega)$. For notational simplicity take $p > q$. Then, if $p < d/(d-1)$, the transport density σ belongs to L^p and, if $p \geq d/(d-1)$, it belongs to $L^r(\Omega)$ for all the exponents r satisfying*

$$r < r(p,q,d) := \frac{dq(p-1)}{d(p-1)-(p-q)}.$$

Proof. The first part of the statement (the case $p < d/(d-1)$ is a consequence of Theorem 2. For the second one, using exactly the same argument as before (Theorem 5) we get

$$||\mu_t||_p \leq (1-t)^{-d/p'}||\mu_0||_p; \quad ||\mu_t||_q \leq t^{-d/q'}||\mu_1||_q.$$

We then apply standard Hölder inequality to derive the usual interpolation estimate for any exponent $q < r < p$:

$$||f||_r \leq ||f||_p^{\alpha}||f||_q^{1-\alpha} \quad \text{with } \alpha = \frac{p(r-q)}{r(p-q)}, \text{ and } 1-\alpha = \frac{q(p-r)}{r(p-q)}.$$

This implies

$$||\mu_t||_r \leq C||\mu_t||_p \leq C||\mu_0||_p \text{ for } t < \frac{1}{2}; \quad ||\mu_t||_r \leq C(1-t)^{-\alpha d/p'}||\mu_0||_p^{\alpha}||\mu_1||_p^{1-\alpha} \text{ for } t > \frac{1}{2}.$$

Then, take $r < r(p,q,d)$, so that $\alpha d/p' < 1$ is ensured and hence the L^r norm is integrable, thus giving a bound on $||\sigma||_r$. □

Remark 10. We do not know whether this exponent $r(p,q,d)$ is sharp or not and whether σ belongs or not to $L^{r(p,q,d)}$.

On the contrary, Example 4.15 in [13] shows the sharpness of the bound on p that we set in Theorem 2.

Acknowledgements. The author bothered many people with his "new" proofs on transport density and is deeply indebted to L. Ambrosio for insisting on exploiting this kind of techniques for wider results than the absolute continuity only and to A. Figalli for useful bibliographical remarks. Several personal discussions with G. Buttazzo, G. Carlier, L. De Pascale and C. Villani have been important as well. Moreover, some key points of the proofs have been developed during a stay at PIMS in Vancouver, and the institute and its facilities are warmly thanked. An anonymous referee who carefully read the paper and suggested clarifying remarks is thanked as well.

The author also wants to thank fundings from the french-italian project Galilée "Modèles de réseaux de transport: phénomènes de congestion et de branchement" and from the ANR project OTARIE.

References

[1] L. Ambrosio, Lecture Notes on Optimal Transport Problems, in *Mathematical Aspects of Evolving Interfaces*, 1–52, Springer Verlag, Berlin, 2003.

[2] L. Ambrosio, N. Gigli and G. Savaré, *Gradient Flows in Metric Spaces and in the Space of Probability Measures*, Lectures in Mathematics, Birkhäuser, 2005.

[3] L. Ambrosio, B. Kirchheim and A. Pratelli, Existence of optimal transport maps for crystalline norms, *Duke Math. J.* (125), no. 2, 207–241, 2004.

[4] L. Ambrosio and A. Pratelli. Existence and stability results in the L^1 theory of optimal transportation, in *Optimal transportation and applications*, Lecture Notes in Mathematics (CIME Series, Martina Franca, 2001) 1813, L.A. Caffarelli and S. Salsa Eds., 123-160, 2003.

[5] M. Beckmann, A continuous model of transportation, *Econometrica* 20, 643–660, 1952.

[6] P. Bernard, B. Buffoni, The Monge Problem for supercritical Mañé Potentials on compact Manifolds, *Advances in Mathematics*, 207 n. 2, 691-706, 2006.

[7] S. Bianchini, M. Gloyer, On the Euler-Lagrange equation for a variational problem: the general case II, preprint available online at the page `https://digitallibrary.sissa.it/handle/1963/2551`, 2008.

[8] G. Bouchitté, G. Buttazzo and P. Seppecher, Shape optimization solutions via Monge-Kantorovich equation. *C. R. Acad. Sci. Paris Sér. I Math.* 324, n. 10, 1185–1191, 1997.

[9] A. Brancolini, G. Buttazzo and F. Santambrogio, Path Functionals over Wasserstein spaces, *J. Eur. Math. Soc.*, 8, n. 3, 415–434, 2006.

[10] G. Carlier, C. Jimenez , F. Santambrogio, Optimal transportation with traffic congestion and Wardrop equilibria, *SIAM Journal on Control and Opt.* 47, n. 3, 1330–1350, 2008.

[11] T. Champion and L. De Pascale, The Monge problem for strictly convex norms in $\mathbb{R}^d$, preprint, 2008, available online at the page `http://cvgmt.sns.it/cgi/get.cgi/papers/chadep08/` .

[12] G. Dal Maso: *An Introduction to* Γ*-convergence.* Birkhäuser, Basel, 1992.

[13] L. De Pascale and A. Pratelli, Regularity properties for Monge Transport Density and for Solutions of some Shape Optimization Problem, *Calc. Var. Par. Diff. Eq.* 14, n. 3, pp. 249–274, 2002.

[14] L. De Pascale, L. C. Evans and A. Pratelli, Integral estimates for transport densities, *Bull. of the London Math. Soc.*. 36, n. 3, pp. 383–385, 2004.

[15] L. De Pascale and A. Pratelli, Sharp summability for Monge Transport density via Interpolation, *ESAIM Control Optim. Calc. Var.* 10, n. 4, pp. 549–552, 2004.

[16] M. Feldman and R. McCann, Uniqueness and transport density in Monge's mass transportation problem, *Calc. Var. Par. Diff. Eq.* 15, n. 1, pp. 81–113, 2002.

[17] A. Figalli, N. Juillet, Absolute continuity of Wasserstein geodesics in the Heisenberg group, *J. Funct. Anal.*, 255, n. 1, pp. 133-141, 2008.

[18] R. J. McCann, A convexity principle for interacting gases. *Adv. Math.* 128, no. 1, 153–159, 1997.

[19] C. Villani, *Optimal Transport, Old and New*, Grundlehren der mathematischen Wissenschaften, Vol. 338, Springer, 2009.