Sharp Wasserstein estimates for integral sampling and Lorentz summability of transport densities

Filippo Santambrogio*

October 18, 2022

Abstract

We prove some Lorentz-type estimates for the average in time of suitable geodesic interpolations of probability measures, obtaining as a by product a new estimate for transport densities and a new integral inequality in terms of Wasserstein distances and norms of gradients. This last inequality was conjectured in a paper by S. Steinerberger.

1 Introduction

A well-known inequality in the Monge-Kantorovich optimal transport theory (see [24, 21]) is the following: on a convex domain $\Omega \subset \mathbb{R}^d$ (by "domain" we mean a compact set with non-empty interior), for any Lipschitz function $\phi : \Omega \to \mathbb{R}$ and any two probability measures $\mu, \nu \in \mathcal{P}(\Omega)$ we have

$$\left| \int \phi d\mu - \int \phi d\nu \right| \leq ||\nabla \phi||_{L^\infty} W_1(\mu, \nu).$$

This is a consequence of the well-known duality formula (also called *Kantorovich* or *Kantorovich-Rubinstein duality*) $W_1(\mu, \nu) = \sup\{\int \phi d\mu + \int \psi d\nu \,:\, \phi(x) + \psi(y) \leq |x - y|\}$ where one can prove that the minimization can be restricted to pairs (ϕ, ψ) with $\psi = -\phi$ and $\phi \in \mathrm{Lip}_1$.

The above inequality shows a duality result between the distance W_1 and the space of Lipschitz functions, and one could wonder whether other distances W_p are in duality with other Sobolev spaces, or, more generally, how to generalize the above inequality to other norms of the gradient, other Wasserstein distances, and possible coefficients depending on some norms of μ and ν. Indeed, it would be natural to see whether the same duality is true between W_p and the Sobolev space $W^{1,p'}$ (with $p' = p/(p-1)$). Actually, a similar inequality cannot hold. Indeed, not all probability measures belong to the dual space $W^{-1,p}$ (in particular, Dirac masses only belong to the dual of $W^{1,q}$ when functions in $W^{1,q}$ are continuous, i.e. when $q > d$). Yet, similar estimates exist when we add L^∞ assumptions on μ and ν. For instance the following inequality was proven for $p = 2$ in [17], then in [18] for different purposes, and is also discussed in [21], Section 5.2.

*Institut Camille Jordan, Université Claude Bernard - Lyon 1; 43 Boulevard du 11 Novembre 1918, 69622 Villeurbanne cedex, France & Institut Universitaire de France santambrogio@math.univ-lyon1.fr .

Proposition 1.1. *Assume that μ and ν are absolutely continuous probability measures on a convex domain Ω and that their densities are bounded by the same constant C. Then, for all function $\phi \in C^1(\Omega)$, we have the following inequality:*

$$\int_\Omega \phi \, d(\mu - \nu) \;\leq\; C^{1/p'} \, ||\nabla \phi||_{L^{p'}(\Omega)} W_p(\mu, \nu)$$

Proof. Let μ_t be the constant speed geodesic between μ and ν, and let v_t be its velocity field. According to the theory of curves in the Wasserstein space (see [2] or Chapter 5 in [21]) the curve (μ_t, v_t) satisfies the continuity equation $\partial_t \mu_t + \nabla \cdot (v_t \mu_t) = 0$, and $||v_t||_{L^p(\mu_t)} = W_p(\mu, \nu)$. The geodesic convexity of all functionals $\rho \mapsto \int \rho^r$ for any $r > 1$ (see [19]) provides $||\mu_t||_{L^r} \leq \max\{||\mu_0||_{L^r}, ||\mu_1||_{L^r}\}$ and, at the limit $r \to \infty$, we obtain that μ_t is absolutely continuous for all t, and its density is bounded by the same constant C. Therefore:

$$\begin{aligned} \int_\Omega \phi \, d(\mu - \nu) \;&=\; \int_0^1 \frac{d}{dt} \left(\int_\Omega \phi(x) d\mu_t(x) \right) dt \;=\; \int_0^1 \int_\Omega \nabla \phi \cdot v_t \, d\mu_t \, dt \\ &\leq\; \left(\int_0^1 \int_\Omega |\nabla \phi|^{p'} \, d\mu_t \, dt \right)^{1/p'} \left(\int_0^1 \int_\Omega |v_t|^p \, d\mu_t \, dt \right)^{1/p} \\ &\leq\; C^{1/p'} \, ||\nabla \phi||_{L^{p'}(\Omega)} W_p(\mu, \nu), \end{aligned}$$

and the proof is completed. □

This inequality can be seen as a way to evaluate the integration error if replacing a measure μ with another measure ν and it would be tempting to apply them to integral sampling, i.e. when approximate an integral with a finite sum. For the history of similar inequalities which were proven for the sake of integration, we refer to the introduction of [23] and, for instance, to the pioneering paper by Bakhalov [4] which involves indeed the Lipschitz constant of the integrand function ϕ. Sharp and recent improvements exist, and we mention, for instance, [5] for an improvement involving a different and weaker combination of norms of $\nabla \phi$.

However, in this setting we have to compare a diffuse measure μ (often the uniform measure on a given domain, for instance a cube) with a finitely atomic one. From this point of view, estimates which require both measures to be L^∞ are useless. It is on the other hand possible to wonder whether similar estimates can be obtained only supposing boundedness or summability assumptions on one of the measures and accepting the other one to be singular. This is exactly the spirit of some estimates involving transport densities in [9] and [20] and, by the way, the strategy in the latter paper [20] exactly starts from the case of a purely atomic target measure, and then proceeds by approximation.

Another important point in choosing the precise form of the inequality that we would like to prove concerns the choice of the Wasserstein distance (i.e., the exponent p). Indeed, if we aim at applications to the case where ν is a discretization of μ (for instance with $\nu = \sum_{i=1}^N a_i \delta_{y_i}$ where the points y_i are on a regular grid and a_i is equal to the mass given by μ to the corresponding cell on the grid), in many situations all Wasserstein distances $W_p(\mu, \nu)$ are of the same order $N^{-1/d}$, where N is the number of points in the discretization and d the space dimension. In this case, we do not lose anything if we replace W_p with the distance W_∞, defined via

$$W_\infty(\mu, \nu) := \inf\{||x - y||_{L^\infty(\gamma)} \;:\; \gamma \in \Pi(\mu, \nu)\},$$

where the set $\Pi(\mu,\nu)$ is the set of transport plans with marginals μ and ν. When μ is absolutely continuous this quantity also equals $\min\{||T-id||_{L^\infty(\mu)} \ : \ T_\#\mu=\nu\}$. For the supremal optimal transport problem behind the definition of W_∞ we cite the pioneering work [8] which introduced the notion of ∞-cyclical monotonicity which we will use later on, and proved the existence of optimal maps for this supremal transport problem.

Replacing Wasserstein distances with W_∞, the natural question, introduced by Stefan Steinerberger in [23] is then: for which functional space X can we prove an inequality of the form

$$\left|\int \phi d\mu - \int \phi d\nu\right| \leq C(\mu)||\nabla\phi||_X W_\infty(\mu,\nu),$$

valid for every probability measure $\nu \in \mathcal{P}(\Omega)$ and for a suitable class of probability densities $\mu \in \mathcal{P}(\Omega)$ which should include at least the uniform measure. The constant $C(\mu)$ will be allowed to depend on the dimension and on possible norms $(L^p, L^\infty, \dots)$ of μ. Even if we explained the role of the W_∞ distance mentioning the example of atomic measures on regular grid, we insist that a sharp inequality in terms of W_∞ could be then succesfully applied to other forms of approximations of μ, for which the value of such a distance could be slightly higher than $O(N^{-1/d})$, for instance to random approximations as in [15].

A first observation is that we can take $X = L^p$ with $C(\mu) = \frac{p}{p-d}||\mu||_{L^{p'}}$ (the exponent p' being the dual of p, i.e. $p' = p/(p-1)$) for any $p > d$. A quick explanation is the following one: we act as in the proof of Lemma 1.1 using a geodesic μ_t with $\mu_1 = \mu$ and $\mu_0 = \nu$ and obtain

$$\int_\Omega \phi \, d(\mu-\nu) = \int_0^1 \int_\Omega \nabla\phi \cdot v_t \, d\mu_t \, dt \leq ||v||_{L^\infty} \int |\nabla\phi| dM, \tag{1.1}$$

where $M := \int_0^1 \mu_t dt$. If the geodesic is chosen to be geodesic in the space W_∞, then $||v||_{L^\infty} = W_\infty(\mu,\nu)$. The question is to find the summability of M and the strategy can be the very same as used in [20]: in the case of a measure ν which is finitely atomic we have, for any exponent q, the equality $||\mu_t||_{L^q} = t^{-d/q'}||\mu||_{L^q}$, which becomes an inequality for general ν, obtained by approximation via atomic measures and semicontinuity. Hence, we have $||M||_{L^q} \leq ||\mu||_{L^q} \int_0^1 t^{-d/q'}$, and the integral converges and can be explicitly computed for $q' > d$.

Yet, choosing $X = L^p$ for $p > d$ is disappointing as it does not show a sharp space to use, but just a family of spaces, and Steinerberger in [23] conjectured that the sharp space could be a Lorentz space of the form $X = L^{d,1}$ (and a partial result in this direction is also shown in the same paper). This would correspond to proving an $L^{d',\infty}$ estimate for M.

In the next section we will briefly recall the definition and main properties of the Lorentz spaces $L^{p,q}$. Here we just finish this introduction by clarifying that the main goal of the present paper is to prove Steinerberger's conjecture. This will be done via a summability estimate on M which will also imply a similar result for the transport density σ (we do not enter into details here about the definition of the transport density, but we refer to [20]).

The main results of the paper are the following.

Theorem 1.2. *Given a convex compact domain $\Omega \subset \mathbb{R}^d$, in dimension $d \geq 2$, and $\mu,\nu \in \mathcal{P}(\Omega)$, suppose $\mu \in L^{d',1}$. Then, for every $p \in [1,+\infty]$, there exists geodesic in W_p connecting $\mu_0 = \nu$ to*

$\mu_1 = \mu$, *of the form* $\mu_t := ((1-t)\mathrm{id} + tT)_{\#}\mu$ *where* T *is the corresponding optimal transport map, such that, setting* $M := \int_0^1 \mu_t dt$, *we have*

$$||M||_{L^{d',\infty}} \leq C||\mu||_{L^{d',1}},$$

for a constant C *only depending on the dimension* d.

It is not difficult to see that the above theorem is in some sense sharp, since when μ is the uniform measure on the unit ball and $\nu = \delta_0$ we have $M(x) = c|x|^{1-d}$ near the origin, and this function exactly belongs to $L^{d',\infty}$, and not to other spaces $L^{p,q}$ with $p > d'$ or $p = d'$ and $q < \infty$.

Moreover, the above theorem has two corollaries:

Corollary 1.3. *Given a convex domain* $\Omega \subset \mathbb{R}^d$ *for* $d \geq 2$, *a function* $\phi \in C^1(\Omega)$ *and* $\mu, \nu \in \mathcal{P}(\Omega)$, *we do have*

$$\left| \int \phi d\mu - \int \phi d\nu \right| \leq C||\mu||_{L^{d',1}} ||\nabla \phi||_{L^{d,1}} W_\infty(\mu, \nu).$$

Corollary 1.4. *Given a convex compact domain* $\Omega \subset \mathbb{R}^d$ *and* $\mu, \nu \in \mathcal{P}(\Omega)$, *suppose* $\mu \in L^{d',1}$. *Then the transport density* σ *in the Monge problem from* μ *to* ν *belongs to* $L^{d',\infty}$ *and we have*

$$||\sigma||_{L^{d',\infty}} \leq C \mathrm{diam}(\Omega) ||\mu||_{L^{d',1}},$$

We see that Corollary 1.3 exactly answers Steinerberger's conjecture, and that Corollary 1.4 slightly extends Dweik's result in [12] to the limit case $p = d'$ (but requires a loss in the q exponent, from $q = 1$ to $q = \infty$).

2 Few facts about Lorentz spaces

We refer to [16, 7, 6] for the whole theory about Lorentz spaces. Here we just give a brief overview of the main facts that we need to know in the paper, which almost coincides with the appendix in [12].

On a measurable space Ω, given $p, q \in [1, \infty]$, we define the Lorentz space $L^{p,q}(\Omega)$ as the space of measurable functions f on Ω satisfying

$$||f||_{L^{p,q}} = p^{1/q} ||t \mapsto t|\{|f| \geq t\}|^{1/p}||_{L^q(\mathbb{R}+, \frac{dt}{t})} < +\infty.$$

In the case $q = 1$ this gives

$$||f||_{L^{p,1}} = p \int_0^\infty |\{|f| \geq t\}|^{1/p} dt$$

and in the case $q = \infty$ we have

$$||f||_{L^{p,\infty}} = \sup_{t \geq 0} t|\{|f| \geq t\}|^{1/p}.$$

It is easy to see that the set of functions making these quantities finite is indeed a vector space, but these quantities themselves are not a norm, since they only satisfy a weaker version of the triangle inequality, involving a multiplicative constant $C = C(p, q)$, i.e. $||f + g||_{L^{p,q}} \leq C(||f||_{L^{p,q}} + ||g||_{L^{p,q}})$. Yet, it is possible to define another quantity, slightly more involved, which is indeed subadditive and can be used as a norm on these spaces. Using the same notations of the main references for Lorentz spaces we set

$$f^{**}(s) := \sup\{\int_A |f| \ : \ |A| = s\}, \qquad |||f|||_{p,q} := ||s \mapsto s^{1/p} f^{**}(s)||_{L^q(\mathbb{R}+,\frac{ds}{s})}.$$

For $q \in [1,\infty]$ and $p > 1$ we have the following inequalities

$$||f||_{L^{p,q}} \le |||f|||_{p,q} \le \frac{p}{p-1} ||f||_{L^{p,q}},$$

which show that $||f||_{L^{p,q}}$ and $|||f|||_{p,q}$ are equivalent up to multiplicative constants. On the other hand, $|||f|||_{p,q}$ is clearly 1-homogeneous and sub-additive, and is hence a norm. The space $L^{p,q}$ endowed with this norm is a Banach space and its dual can be identified with $L^{p',q'}$. More precisely, for the case $q = 1$, we have

$$\int fg \le ||f||_{L^{p,1}} ||g||_{L^{p',\infty}}.$$

Finally, in order to perform approximation, we underline that $f \mapsto |||f|||_{p,q}$ is clearly lower semicontinuous for the weak L^1 convergence, as a consequence of Fatou's lemma and of stability of lower semicontinuity when taking a sup. In particular, since bounds on $|||f|||_{p,q}$ for $p > 1$ imply bounds on any L^r norm for $r \in (1,p)$, we obtain the following result, that we state only for $q = \infty$ since its proof can be easily detailed and it is the only case which will be used in the sequel.

Proposition 2.1. *Suppose m_n is a sequence of measures on a domain Ω with finite Lebesgue measure. Suppose $m_n \overset{*}{\rightharpoonup} m$ in the sense of weak-* convergence of measures (in duality with bounded continuous functions), and suppose that m_n is absolutely continuous for each n and that, identifying measures and densities, we have $|||m_n|||_{L^{p,\infty}} \le C$, for some exponent $p > 1$. Then $|||m|||_{L^{p,\infty}} \le C$.*

Proof. We observe that the bound on $|||m_n|||_{L^{p,\infty}}$ implies

$$|\{|m_n| > t\}| \le \min\{|\Omega|, Ct^{-p}\}, \qquad \text{hence } \int |m_n|^r dx = r \int_0^\infty t^{r-1} |\{|m_n| > t\}| dt \le C(r,p,|\Omega|)$$

for any exponent $r \in (1,p)$. In particular, m_n is bounded in $L^{(p+1)/2}$; it therefore weakly converges up to subsequences in $L^{(p+1)/2}$ and hence in L^1 to m. For any set A the quantity $m \mapsto \int_A m$ is obviously continuous for this convergence and this provides $\liminf_n m_n^{**}(t) \ge m^{**}(t)$ for every t, and hence, by Fatou, $|||m|||_{L^{p,\infty}} \le \liminf_n |||m_n|||_{L^{p,\infty}}$. □

3 Transport and geodesics in W_p and W_∞

We refer to [21], Chapter 5, for most of the facts below.

We recall the definition of Wasserstein distance W_p for $p \in [1,\infty)$:

$$W_p(\mu,\nu) := \min\left\{ \int_{\Omega\times\Omega} |x-y|^p \, d\gamma(x,y) \ : \ \gamma \in \Pi(\mu,\nu) \right\}^{1/p}. \tag{3.1}$$

We also recall that the space $\mathcal{P}(\Omega)$ endowed with the distance W_p (when Ω is a bounded convex set) is a geodesic space where a geodesic curve μ_t can be obtained by taking an optimal transport plan γ between μ and ν and setting $\mu_t = (\pi_t)_\#\gamma$ where $\pi_t(x, y) = (1-t)x + ty$. In the case where the optimal transport plan γ is induced by a transport map T (so that we have $\gamma = (\mathrm{id}, T)_\#\mu$) we would have $\mu_t = ((1-t)\mathrm{id} + tT)_\#\mu$. This curve of measure solves a continuity equation $\partial_t\mu_t + \nabla\cdot(\mu_t v_t) = 0$ for a velocity field v with $||v_t||_{L^p(\mu_t)} = W_p(\mu, \nu)$ for every t.

Similarly, we set

$$W_\infty(\mu, \nu) := \min\left\{||x-y||_{L^\infty(\gamma)} \ : \ \gamma \in \Pi(\mu, \nu)\right\}, \tag{3.2}$$

and the curve $\mu_t = (\pi_t)_\#\gamma$ is also a geodesic for the distance W_∞ for any optimal γ. In this case we have $\partial_t\mu_t + \nabla\cdot(\mu_t v_t) = 0$ for a velocity field v with $|v_t(x)| \le W_\infty(\mu, \nu)$ for every t, x.

We suppose now that μ is absolutely continuous and ν is purely atomic, i.e. $\nu = \sum_{i=1}^N a_i\delta_{y_i}$. When considering a transport map T such that $T_\#\mu = \nu$ we obtain a partition: we have $T(x) = y_i$ for every $x \in \Omega_i$ where the sets $(\Omega_i)_{i=1,\dots,N}$ form a partition of Ω, with $\Omega_i := T^{-1}(\{y_i\})$. We call $\Omega_i(t)$ the image of Ω_i via the map $x \mapsto (1-t)x + ty_i$ We need the following statement.

Proposition 3.1. *Let μ be an absolutely continuous measure on a domain $\Omega \subset \mathbb{R}^d$, $d \ge 2$, and $\nu = \sum_{i=1}^N a_i\delta_{y_i}$, and consider $p \in [1, +\infty]$. Then there exists an optimal transport plan γ of the form $\gamma = (\mathrm{id}, T)_\#\mu$ such that the sets $\Omega_i(t)$ are essentially disjoint (in the sense that we have $|\Omega_i(t)\cap\Omega_j(t)| = 0$ for $i \ne j$) for every $t \in (0, 1)$.*

Proof. In the case $p < \infty$, the optimal transport plan γ is unique and induced by a map. We need to prove that these sets are essentially disjoint for every $t \in (0, 1)$.

In the case $p = 1$ this is a consequence of the fibration into transport rays: 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 y_i and the one from $x_j \in \Omega_j$ to y_j. The only possibility is that these two rays are actually the same, i.e. that the five points x_i, x_j, z, y_i, y_j are aligned. But this implies that these points belong to one of the lines connecting two atoms y_i and y_j. Since we have finitely many of these lines it is enough to remove a negligible set. Notice that this argument only works for $d > 1$.

In the case $p \in (1, \infty)$ we cannot use transport rays, but the transport cost is of the form $c(x, y) = h(x-y)$ for h strictly convex and we can use c-cyclical monotonicity. Indeed, with $z \in \Omega_i(t)\cap\Omega_j(t)$ we have, for two vectors a, b and the same notations as above, $y_i - x_i = a$, $y_j - x_j = b$, $z = x_j + tb = x_i + ta$, $y_j - x_i = (1-t)b + ta$, and $y_i - x_j = (1-t)a + tb$. The c-cyclical monotonicity condition implies

$$h(y_i - x_i) + h(y_j - x_j) \le h(y_j - x_i) + h(y_i - x_j) \Rightarrow h(a) + h(b) \le h((1-t)b + ta) + h((1-t)a + tb) \tag{3.3}$$

but this last inequality is impossible for strictly convex h, unless $a = b$, which finally implies $y_j = y_i$.

The case $p = \infty$ is a little more delicate. In this case it is no more true, in general, that the optimal transport plan γ is unique and induced by a map. Yet, as it is shown in [8], there exists a particular optimal plan, the one obtained as a limit from the case $p \to \infty$ (i.e. the weak limits of plan γ_p, optimal for the integral cost $|x-y|^p$) which is indeed induced by a transport map T, and also satisfies the following ∞-cyclical monotonocity condition: there exists a full-measure set $\tilde\Omega \subset \Omega$ such that for every $x, x' \in \tilde\Omega$ we have

$$\max\{|T(x) - x|, |T(x') - x'|\} \le \max\{|T(x') - x|, |T(x) - x'|\}$$

(the condition proven in [8] is actually stronger, but here we only need to apply it to pairs of points instead of arbitrary cycles).

This also implies the disjointness of the sets $\Omega_i(t)$ and $\Omega_j(t)$ for $t \in (0,1)$. Indeed, the condition (3.3) becomes now

$$\max\{|a|,|b|\} \le \max\{|(1-t)a+tb|,|(1-t)b+ta|\} \le \max\{(1-t)|a|+t|b|,(1-t)|b|+t|a|\} \le \max\{|a|,|b|\}.$$

For $t \in (0,1)$ the last inequality is strict unless $|a| = |b|$. The previous one is strict unless a and b are colinear and with the same orientation. The two conditions together imply $a = b$ and, again, $y_i = y_j$. □

4 Lorentz estimates and applications

We first prove Theorem 1.2.

Proof. We start from the case where ν is atomic. According to Proposition 3.1, we choose an optimal transport map T such that $T_{\#}\mu = \nu$, $||T - \mathrm{id}||_{L^p(\mu)} = W_p(\mu,\nu)$, and the sets $\Omega_i(t)$ defined in the previous section are disjoint. We then define $T_t = t\mathrm{id} + (1-t)T$ (note that for simplicity we reverse time, in order to have easier expressions later on in the proof). In this way $\mu_t := (T_t)_{\#}\mu$ is a geodesic from $\mu_0 = \nu$ to $\mu_1 = \mu$ (pay attention to the fact that, contrarily to standard notation, this curve starts in ν and arrives at μ; this is needed to ease some notations later on in terms of t vs $(1-t$ for dyadic times t). We set $M := \int_0^1 (T_t)_{\#}\mu$.

Take now a set $A \subset \{\mu > 0\}$ and define $\mu_t^{[A]} := (T_t)_{\#}\mathbb{1}_A$. Defining $A_i := A \cap \Omega_i$, then the sets $A_i(t) := (T_t)(A_i)$ are disjoint, and we set $A(t) := \bigcup_i A_i(t)$. Since the map T_t is a homothety of dilation ratio t on each set A_i and the images of these sets are disjoint, we have $\mu_t^{[A]} = t^{-d}\mathbb{1}_{A(t)}$ and $|A(t)| = t^d|A|$.

Look now at $M[A] := \int_0^1 \mu_t^{[A]} dt$. We have

$$\begin{aligned} M[A](x) = \int_0^1 t^{-d}\mathbb{1}_{A(t)}(x)dt &= \sum_{k=0}^{\infty} \int_{2-(k+1)}^{2^{-k}} t^{-d}\mathbb{1}_{A(t)}(x)dt \\ &= \sum_{k=0}^{\infty} 2^{k(d-1)} \int_{1/2}^{1} t^{-d}\mathbb{1}_{A(2^{-k}t)}(x)dt \\ &= \int_{1/2}^{1} t^{-d} M[A,t](x)dt, \end{aligned}$$

where we set $M[A,t] := \sum_{k=0}^{\infty} 2^{k(d-1)}\mathbb{1}_{A(2^{-k}t)}$. We now define, for fixed $t \in (\frac{1}{2},1)$, the function $K(x) := \sup\{k : x \in A(2^{-k}t)\}$ and we have

$$M[A,t](x) = \sum_{k=0}^{\infty} 2^{k(d-1)}\mathbb{1}_{A(2^{-k}t)}(x) \le \sum_{k=0}^{K(x)} 2^{k(d-1)} = \frac{2^{(d-1)(K(x)+1)}-1}{2^{d-1}-1} \le 2^{(d-1)K(x)+1},$$

where we used in the last inequality $2^{d-1}/(2^{d-1}-1) \le 2$, a consequence of $d \ge 2$.

Hence, for every $s > 0$, writing $s = 2^{(d-1)u+1}$ we have

$$\{M[A,t] > s\} \subset \{K > u\} = \bigcup_{k=\lfloor u \rfloor+1}^{\infty} A(2^{-k}t)$$

and

$$|\{M[A,t] > s\}| \le \sum_{k=\lfloor u \rfloor+1}^{\infty} |A(2^{-k}t)| = \sum_{k=\lfloor u \rfloor+1}^{\infty} 2^{-kd}t^d|A| \le C(d)t^d|A|2^{-(\lfloor u \rfloor+1)d} \le C(d)|A|2^{-du}.$$

We then have, using $\frac{d}{d'} = d-1$ and the definition of u

$$s|\{M[A,t] > s\}|^{1/d'} \le sC(d)|A|^{1/d'}2^{-(d-1)u} = C(d)|A|^{1/d'},$$

which shows

$$||M[A,t]||_{L^{d',\infty}} \le C(d)|A|^{1/d'}.$$

A similar inequality, up to changing the constant $C(d)$, also holds for $|||M[A,t]|||_{L^{d',\infty}}$. We first integrate this inequality as follows

$$M[A] = \int_{1/2}^{1} t^{-d}M[A,t]dt \quad \Rightarrow \quad |||M[A]|||_{L^{d',\infty}} \le C(d)|A|^{1/d'},$$

where we used the Minkowski inequality for norms (which is the reason to introduce the norm $|||\cdot|||_{L^{p,q}}$ and avoid using quasi-norms).

We then write $\mu = \int_0^\infty \mathbb{1}_{\{\mu > r\}}dr$ and apply the previous result to $A = \{\mu > r\}$, together with the linearity of the map $\mu \mapsto \int_0^1 (T_t)_\# \mu$ and, again, the Minkowski inequality for norms, and obtain

$$|||M|||_{L^{d',\infty}} \le C(d) \int_0^\infty |\{\mu > r\}|^{1/d'} = C(d)||\mu||_{L^{d',1}},$$

which proves the claim for ν atomic.

The case of general ν is obtained by approximation using Proposition 2.1. $\square$

Before going on with the consequences of this estimate we observe some facts.

The first is that the use of the countable parameter k was necessary to estimate the measure of a union with the sum of the measures, which could not have been done with an uncountable union. If the goal was only to prove Corollary 1.3 then it was also possible to just use a countable parameter writing $\int \phi d(\mu - \nu) = \sum_k \int \phi d(\mu_{2^{-k}} - \mu_{2^{-(k+1)}})$. Yet, in this way we lose the infinitesimal and differential approach described in (1.1) and we are forced to estimate differences instead of derivatives. This is possible replacing $|\nabla\phi|$ with its maximal function, but then requires the boundedness of the maximal operator on Lorentz space, which is a known fact (see, for instance [22]), but would only make the proof heavier.

Remark 4.1. We then remark that the implication $\mu \in L^{d',1} \Rightarrow M \in L^{d',\infty}$ cannot be improved into $\mu \in L^{d',\infty} \Rightarrow M \in L^{d',\infty}$. Indeed, taking $\mu(x) = c|x|^{1-d}\mathbb{1}_{|x|\le 1}$ (a typical example of function belonging to $L^{d',\infty}$) and $\nu = \delta_0$ provides a counter-example, as we have in this case $M(x) = c|x|^{1-d}|\log x|\mathbb{1}_{|x|\le 1} \notin L^{d',\infty}$.

We now go on with the two main corollaries of Theorem 1.2 that we stated in the introduction. We start with the proof of Corollary 1.3.

Proof. We follow the strategy described in (1.1) along the geodesic curve in W_∞ obtained in the proof of Theorem 1.2. We obtain

$$\int_\Omega \phi\, d(\mu - \nu) = \int_0^1 \int_\Omega \nabla\phi \cdot v_t\, d\mu_t\, dt \le ||v||_{L^\infty} \int |\nabla\phi| dM \le ||v||_{L^\infty} ||\nabla\phi||_{L^{d,1}} ||M||_{L^{d',\infty}},$$

and we then use $||v||_{L^\infty} \le W_\infty(\mu,\nu)$ and apply the estimate of Theorem 1.2 to replace $||M||_{L^{d',\infty}}$ with $C(d)||\mu||_{L^{d',1}}$. □

We now move to the proof of Corollary 1.4.

Proof. The statement is interesting only for $d > 1$, otherwise the norm $L^{d'\infty}$ becomes the weak L^1 norm, which is bounded by the L^1 norm, and the claim is a trivial consequence of $||\sigma||_{L^1} = W_1(\mu,\nu) \le \mathrm{diam}(\Omega)$ and of the fact that μ is supposed to be a probability measure.

For $d \ge 2$ we use Theorem 1.2. Indeed, the transport density σ is defined as $\int_0^1 (\pi_t)_\#(c\cdot\gamma)dt$ where $c(x,y) = |x-y|$ and γ is an optimal transport plan for the cost c. The measure σ does not depend on the choice of γ as soon as one of the measures μ, ν is absolutely continuous (see [14, 1]). On a bounded domain, we have $c \le \mathrm{diam}(\Omega)$ and hence bounds on $M = \int_0^1 (\pi_t)_\#\gamma dt$ translate into bounds on σ, following the very same technique as in [20]. It is then enough to choose the optimal transport plan γ for which the estimate of Theorem 1.2 holds. □

We underline that the counter-example of Remark 4.1 does not show that $\mu \in L^{d',\infty}$ does not imply $\sigma \in L^{d',\infty}$ since σ could be much smaller than M and, in this case, if we do not ignore the smallness of the factor $c(x,y)$, we obtain $\sigma(x) = c(|x|^{1-d} - |x|^{2-d})\mathbb{1}_{B_1}(x)$, and we have $\sigma \in L^{d',\infty}$. Sharp summability results for σ are an open problem and exploiting the presence of the term c (as it is done in [13], for instance) is in general not easy.

Acknowledgements. The author acknowledges the support of the ANR project MAGA (ANR-16-CE40-0014) and of the Lagrange Mathematics and Computation Research Center project on Optimal Transportation. He also acknowledges interesting discussions with S. Dweik and S. Steinerberger.

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 spaces of probability measures.* Lectures in Mathematics, ETH Zurich, Birkhäuser, 2005.

[3] 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.

[4] N. S. Bakhvalov On the approximate calculation of integrals, *Vestnik MGU*, Ser. Mat. Mekh. Astron. Fiz. Khim., 4, 2–18, 1959.

[5] L. Brown, S. Steinerberger On the Wasserstein Distance between Classical Sequences and the Lebesgue Measure *Trans. Amer. Math. Soc.*, 373, 8943–8962, 2020.

[6] R. E. Castillo and H. Rafeiro, *An Introductory Course in Lebesgue Spaces*, Springer International Publishing, 2016.

[7] D. Chamorro *Espacios de Lebesgue y de Lorentz* Volumen 3, Colección de Matemáticas Universitarias, Editorial Amarun, 2020.

[8] T. Champion, L. De Pascale, P. Juutinen The ∞-Wasserstein Distance: Local Solutions and Existence of Optimal Transport Maps, *SIAM J. Math. Anal.*, 40(1), 1–20, 2008.

[9] 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(3), 249–274, 2002.

[10] L. De Pascale, L. C. Evans and A. Pratelli, Integral Estimates for Transport Densities, *Bull. London Math. Soc.*. 36(3), 383–385, 2004.

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

[12] S. Dweik, $L^{p,q}$ estimates on the transport density *Comm. Pure Appl. An.* 18(6), 3001–3009, 2019.

[13] S. Dweik, F. Santambrogio, L^p bounds for boundary-to-boundary transport densities, and $W^{1,p}$ bounds for the BV least gradient problem in 2D, *Calc. Var. PDEs* 58, nr 31, 2019.

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

[15] N. García Trillos, D. Slepčev. On the Rate of Convergence of Empirical Measures in $\infty-$transportation Distance. *Canadian Journal of Mathematics* 67(6), 1358–1383, 2015.

[16] L. Grafakos *Classical Fourier Analysis*, Graduate Texts in Mathematics, Springer, 2014.

[17] G. Loeper, Uniqueness of the solution to the Vlasov–Poisson system with bounded density. *J. Math. Pures Appl.* 86 (1), 68–79, 2006.

[18] B. Maury, A. Roudneff-Chupin and F. Santambrogio A macroscopic crowd motion model of gradient flow type, *Math. Models and Methods in Appl. Sciences* 20(10), 1787–1821, 2010.

[19] R. J. McCann, A convexity principle for interacting gases. *Adv. Math.* 128(1), 153–159, 1997.

[20] F. Santambrogio, Absolute continuity and summability of transport densities: simpler proofs and new estimates, *Calc. Var. Par. Diff. Eq.* 36(3), 343–354, 2009.

[21] F. Santambrogio *Optimal Transport for Applied Mathematicians*, Progress in Nonlinear Differential Equations and Their Applications no 87, Birkhäuser Basel, 2015.

[22] E. Sawyer Boundedness of classical operators on classical Lorentz spaces. *Studia Mathematica* 96(2), 145–158, 1990.

[23] S. Steinerberger On a Kantorovich-Rubinstein Inequality, *J. Math. Anal. Appl.,* 501, 125185, 2021.

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