CONTINUITY FOR THE MONGE MASS TRANSFER PROBLEM IN TWO DIMENSIONS

QI-RUI LI, FILIPPO SANTAMBROGIO, AND XU-JIA WANG

Abstract. In this paper, we prove the continuity of the monotone optimal mapping of the Monge mass transfer problem in two dimensions under certain conditions on the domains and the mass distributions.

1. Introduction

Let Ω and Ω^* be two bounded domains in the Euclidean space $\mathbb{R}^2$. Let $\varrho \in L^1(\Omega)$ and $\varrho^* \in L^1(\Omega^*)$ be two densities satisfying the mass balance condition

$$\int_\Omega \varrho(x)dx = \int_{\Omega^*} \varrho^*(y)dy. \tag{1.1}$$

The Monge mass transfer problem [23] consists in finding a mapping $s : \Omega \to \Omega^*$ which minimises the cost functional

$$s \mapsto \mathcal{C}(s) = \int_\Omega \varrho(x)|s(x) - x|dx \tag{1.2}$$

among all mappings satisfying the measure-preserving condition $s_\# \varrho = \varrho^*$, namely

$$\int_{s^{-1}(E)} \varrho(x)dx = \int_E \varrho^*(y)dy \quad \text{for all Borel sets } E \subset \Omega. \tag{1.3}$$

Since the original work of Monge, this problem has been extensively studied (see [24, 26] for monographs on the subject). Major advances include the introduction of duality by Kantorovich [15, 16], the existence of optimal mappings by Evans and Gangbo [11] for Lipschitz continuous densities ϱ, ϱ^* with disjoint supports, and by Caffarelli, Feldman and McCann [6], Trudinger and the third author [27] for general densities. An earlier proof using probability theory was given by Sudakov [25], in which a gap was filled up by Ambrosio [1]. When the Euclidean norm in (1.2) is replaced by more general norms, the existence of optimal mappings has also been obtained in [2, 8, 9].

2010 *Mathematics Subject Classification.* 35J60, 35B65, 49Q20.

Key words and phrases. Optimal transportation, Regularity.

This work is supported by ARC FL130100118 and by the ANR through the contract ANR- 12-BS01-0014-01 GEOMETRYA. Also, the second author was a guest at ANU at the beginning of this work, and warmly acknowledges the hospitality of the Mathematical Sciences Institute.

Though the optimal mappings can be obtained by different ways, all of them lead to a mapping satisfying a monotonicity condition, and it was proved by Feldman and McCann that this mapping is indeed unique [12] (we only cite [10] for an example where a different transport mapping is selected through a regularization procedure). Naturally one would like to study the regularity of the monotone optimal mapping.

However, the regularity turns out to be a difficult and delicate problem. A first observation is that in general the optimal mapping fails to be more regular than the two densities ϱ, ϱ^* (examples can be found for instance in [7], but the construction is classical). Yet, there can also be a (strong) loss of regularity: in [17] we constructed a counter-example in which the monotone optimal mapping fails to be Lipschitz continuous, even though the densities are positive and smooth and the domains are convex. Further examples are presented in [7], including examples where the Hölder exponent of the mapping is stricyly smaller than that of ϱ, ϱ^*. On the positive side, in [17] we were able to prove that the eigenvalues of the matrix Ds_ε are locally uniformly bounded as $\varepsilon \to 0$, where s_ε is the optimal mapping associated with the cost function $c_\varepsilon(x, y) = \sqrt{\varepsilon^2 + |x - y|^2}$. Note that the cost function c_ε satisfies the conditions in [22] and so the mapping s_ε is smooth, if the densities and domains satisfy proper conditions. The regularity of optimal mappings for general cost functions has been studied by several authors [4, 5, 13, 18, 19, 20, 21, 28], in particular sharp conditions on cost functions were found in [22].

In view of the above discussion, an interesting question is the continuity of the monotone optimal mapping s in Monge's problem. It is known that for any $x \in \Omega$, if $s(x) \neq x$, then the segment from x to $s(x)$ is contained in a transport segment. Hence the set $\{x \in \Omega \mid x \neq s(x)\}$ is contained in the transport set $\mathcal{T}$ [6, 11, 27]. The continuity of the monotone optimal mapping was studied by Fragalà, Gelli, Pratelli in [14]. They proved the continuity of the monotone optimal mapping in $\mathcal{T}^o$ provided the densities ϱ and ϱ^* are continuous and have compact, convex, and strictly separated supports in $\mathbb{R}^2$. Here we denote by $\mathcal{T}$ the union of all transport segments, and by $\mathcal{T}^o$ the subset of $\mathcal{T}$ by taking away all the endpoints of transport segments.

In this paper we remove the strict separation assumption, and prove a continuity result under a different geometric set of assumptions than the ones in [14]. The most natural framework in our case requires that the supports of the two measures ϱ, ϱ^* coincide or, at least, that the support of ϱ is included in that of ϱ^*. In this case, under certain conditions, we obtain the continuity of the optimal mapping in $\mathcal{T}$ and hence in the whole domain $\overline{\Omega}$ as well. To the best of our knowledge, even if restricted to strong geometric assumptions, this is the first full continuity result for the monotone mapping, while the result of [14] can be considered as a partial regularity result (the mapping is continuous outside a negligible set).

Theorem 1.1. *Let Ω and Ω^* be two bounded, convex domains in $\mathbb{R}^2$. Assume that $\varrho \in C(\overline{\Omega})$ and $\varrho^* \in C(\overline{\Omega}^*)$ are strictly positive and satisfy the mass balance condition* (1.1). *Assume also that $\Omega \subset \Omega^*$ and one of the sets $\{\varrho > \varrho^*\}$, $\{\varrho^* > \varrho\}$ is convex. Then the monotone optimal mapping for Monge's problem* (1.2) *is continuous on $\overline{\Omega}$.*

The convexity of Ω and Ω^* is needed for the continuity of the monotone optimal mapping [17, 22]. This condition may be relaxed but cannot be completely dropped. The convexity condition for one of the sets $\{\varrho > \varrho^*\}$, $\{\varrho^* > \varrho\}$ is such that a transport segment does not cross their boundary multiple times. This seems to be a technical condition but, without it, it is difficult to analyse the geometry of transport segments. The main ingredient in proving Theorem 1.1 is the study of curves, that we call $\mathcal{D}$-curves (we refer the reader to Section 3 for precise definitions), composed of double points, i.e. of points which are endpoints of two different transport segments. In particular we will use the following lemma, where the length of a $\mathcal{D}$-curve is actually defined as its diameter (maximal Euclidean distance between two of its points, and not length along the curve).

Lemma 1.2. *For any given $\varepsilon > 0$, there are at most finitely many disjoint $\mathcal{D}$-curves with length greater than ε in $\mathcal{T}$.*

Once the continuity of the optimal mapping s is proved, the uniform continuity of s with respect to the densities ϱ, ϱ^* and domains Ω, Ω^* also follows if the densities and domains satisfy the conditions in Theorem 1.1 in a uniform way. For fixed domains $\Omega \subset \Omega^*$, the conditions includes the modulus of continuity of ϱ, ϱ^* and the convexity of one of the sets $\{\varrho > \varrho^*\}$, $\{\varrho^* > \varrho\}$. From the proof of Theorem 1.1, one sees that to obtain the uniform continuity of s, we also need uniform positivity bounds on the function $f =: \varrho - \varrho^*$, at least on its support and away from the boundary of $\{\varrho > \varrho^*\}$; and the same for $g =: \varrho^* - \varrho$ in the domain $\{\varrho^* > \varrho\}$. As was pointed out above, by the examples in [7], one cannot expect the modulus of continuity of the optimal mapping is better than that of the densities, and so in general the optimal mapping is not Hölder continuous.

For convenience we give some examples of the densities in Figure 1 below which satisfy conditions in Theorem 1.1.

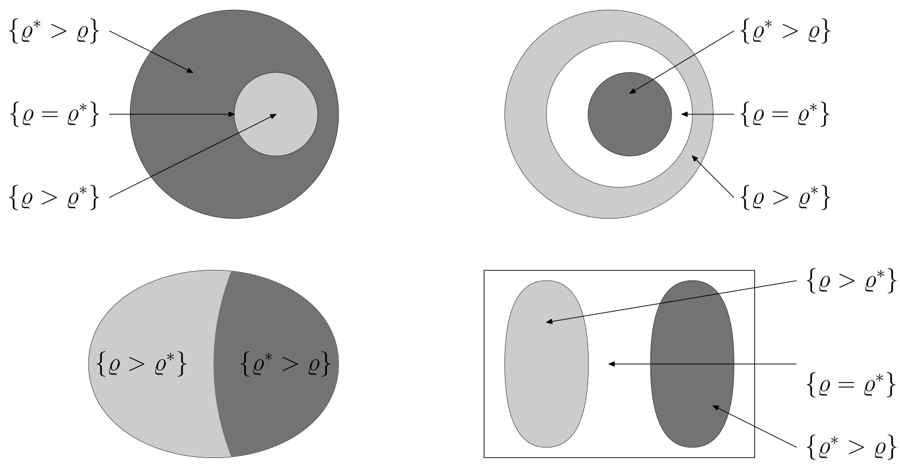

FIGURE 1. Some examples

Our proof uses a similar strategy as that in [14], where the authors proved the continuity of the monotone optimal mapping in $\mathcal{T}^o$. From [6, 27] it is known that the optimal mapping is determined by transport segments. Hence the key ingredient is to understand the distribution of double points. Though not explicitly stated in their paper, they also obtained the result in Lemma 1.2, under the assumption that the densities ϱ and ϱ^* are continuous and have compact, convex, and strictly separated supports in $\mathbb{R}^2$. In this paper we assume that $\Omega \subset \Omega^*$. For the sake of transport segments and double points, the minimisation problem (2.1) is equivalent to (2.4). Hence to understand the distribution of double points, we consider the optimal transportation from Ω to Ω^* with densities $f = (\varrho - \varrho^*)^+$ and $g = (\varrho^* - \varrho)^+$. The difference in our case is that the supports of f and g are not strictly separated, and that causes the main difficulty for this problem. Once Lemma 1.2 is proved, the continuity of the monotone optimal mapping follows.

Therefore in this paper we will focus on the most difficult part of proving the technical Lemma 1.2. The rest of the argument is more or less based on the same idea as that in [14], namely on a thin strip bounded by two transport segments, the monotone optimal mapping can be given by a formula which is essentially one dimensional. We refer the readers to [14] for more detailed discussions on this part.

This paper is arranged as follows. In Section 2 we recall the existence of a monotone optimal mapping and the concepts of transport segment and transport set. In Section 3 we introduce $\mathcal{D}$-curves and discuss their properties. In Section 4 we prove the key point in our argument, i.e. Lemma 4.5. This is the main technical part of the proof. Finally we prove Theorem 1.1 in Section 5.

2. Preliminaries

Let Ω and Ω^* be two bounded convex domains in $\mathbb{R}^2$, and ϱ, ϱ^* be two densities in Ω and Ω^*, satisfying the mass balance condition (1.1). Extend ϱ and ϱ^* to the whole space $\mathbb{R}^2$ such that they are equal to 0 outside Ω and Ω^*. Denote

$$K[\phi] = \int_{\mathbb{R}^2} \phi(x)\big(\varrho(x) - \varrho^*(x)\big)dx, \tag{2.1}$$

and

$$\mathrm{Lip}_1(\mathbb{R}^2) = \Big\{\phi \in C(\mathbb{R}^2) \mid |\phi(x) - \phi(y)| \le |x-y| \ \ \forall\, x, y \in \mathbb{R}^2\Big\}.$$

We have the following well-known result.

Theorem 2.1. *There exists a Lipschitz function u that maximises Kantorovich's dual functional* (2.1) *among functions in $Lip_1(\mathbb{R}^2)$. Moreover, u satisfies*

$$u(x) = \inf_{y\in\Omega^*} \big\{u(y) + |x-y|\big\} \ \ \textit{for any} \ \ x \in \Omega, \tag{2.2}$$

$$u(y) = \sup_{x\in\Omega} \big\{u(x) - |x-y|\big\} \ \ \textit{for any} \ \ y \in \Omega^*. \tag{2.3}$$

A maximiser in the above theorem is called a potential function of the Monge mass transfer problem, or Kantorovich potential.

For convenience we collect some notations below.

- A transport segment ℓ is a maximal line segment $\overline{xy}$, with $x \in \overline{\Omega}$, $y \in \overline{\Omega}^*$ and $x \ne y$, such that
$$|u(x) - u(y)| = |x-y|$$
(which is equivalent to u being affine with slope equal to 1 on the segment $\overline{xy}$) where the maximality means that ℓ is not a proper subset of any line segment with the same property. We will sometimes refer to transport segments as "non-trivial transport segments", with the idea that singletons which are not contained in any transport segment are "trivial" transport segments.
- Denote by $\ell = \ell^x_y$ the transport segment with endpoints x, y such that $u(x) > u(y)$, and call x and y the upper and lower endpoints of ℓ, respectively.

- Set
$$f = (\varrho - \varrho^*)_+ = \max\{\varrho - \varrho^*, 0\},$$
$$g = (\varrho^* - \varrho)_+ = \max\{\varrho^* - \varrho, 0\}.$$
Denote by $\mathcal{S}_f$ and $\mathcal{S}_g$ the supports of f and g, respectively, and by $\mathcal{H}$ the boundary of $\mathcal{S}_g$.
- Note that (2.1) can also be re-written as
$$K[\phi] = \int_{\mathbb{R}^2} \phi(x)(f-g)dx, \tag{2.4}$$
since $f-g = \varrho-\varrho^*$. Hence the minimisation problem (2.1) has the same potential function and the transport segments as the problem (2.4), and to prove Lemma 1.2, it suffices to consider the problem (2.4).
- By approximation (for instance replacing (f, g) with $(f_\varepsilon, g_\varepsilon)$ with $f_\varepsilon(x) = f(x) + \varepsilon d(x, \mathcal{S}_g)$ and $g_\varepsilon = (1+c\varepsilon)g$ for a suitably chosen constant c so that the masses of f_ε and g_ε are equal) we can produce a Kantorovich potential u with the following property: every point of $\Omega^* \setminus \mathcal{H}$ belongs to at least one non-trivial transport segment, which goes from $\Omega \setminus \mathcal{S}_g$ to $\mathcal{S}_g$.
- After the approximation procedure, it is possible that some of the above transport segments ℓ do not enter the region $\{g > 0\}$. In this case the optimal mapping is the identity mapping on them. Such kind of transport segment will be called *degenerate transport segments*, and we will denote by $\mathcal{T}$ the set of points on non-degenerate transport segments (those which have non-empty intersection with $\{g > 0\}$).

In [6, 11, 27] the following existence for optimal mappings was established.

Theorem 2.2. *There exists a measure-preserving mapping s which minimises* (1.2)*. Moreover*
$$u(x) - u(s(x)) = |x - s(x)| \ \textit{ for a.e. } \ x \in \Omega. \tag{2.5}$$

We point out that in the region $\{\varrho \neq \varrho^*\}$, the potential function u is uniquely determined (up to a constant). But in the region $\{\varrho = \varrho^*\}$, u is not necessarily unique. However, by the approximation that we mentioned above, we have selected a particular potential function satisfying Theorems 2.1 and 2.2, and such that every point (outside $\mathcal{H}$) belongs to a non-trivial transport segment.

The optimal transport mapping s obtained in [6, 11, 27] is monotone. This means that, restricted to every transport segment, once an orientation is fixed on such a segment, the mapping s is monotone increasing. Note that such a monotone mapping is unique

[12] and satisfies $s(\mathcal{T}) \subset \mathcal{T}$. Another way of writing the monotonicity condition is the following: the mapping s satisfies

$$\frac{x - x'}{|x - x'|} + \frac{s(x) - s(x')}{|s(x) - s(x')|} \neq 0 \tag{2.6}$$

for all $x \neq x'$ with distinct images $s(x) \neq s(x')$. Note that, when x and x' belong to a same transport segment, this condition boils down to monotonicity, and when x and x' do not belong to a same transport segment, it is not difficult to see that, should we have equality in (2.6), then the corresponding transport segments would meet in an internal point, which is impossible thanks to the following theorem.

Theorem 2.3. *We have the following properties of transport segments and transport sets:*

(i) *If $\ell_k = \ell_{y_k}^{x_k}$ is a sequence of transport segments with $x_k \to x$ and $y_k \to y$, and $x \neq y$, then $\overline{xy}$ is contained in some transport segment ℓ.*

(ii) *Let ℓ_1 and ℓ_2 be two distinct transport segments, then they do not intersect at interior point, namely if $\ell_1 \cap \ell_2 \neq \emptyset$, then $\ell_1 \cap \ell_2$ must be an endpoint of both segments.*

(iii) *$\mathcal{S}_f^o \subset \mathcal{T}$ and $\mathcal{S}_g^o \subset \mathcal{T}$, namely an interior point in $\mathcal{S}_f$ or $\mathcal{S}_g$ is either an interior point or endpoint of a transport segment, where $\mathcal{S}_f^o$ and $\mathcal{S}_g^o$ denote the regions $\{f > 0\}$ and $\{g > 0\}$, respectively.*

Properties in Theorem 2.3 follow from Theorems 2.1 and 2.2 immediately.

To fix notations and in order to present the strategy to prove our continuity result, let us give some names to some important sets. We will set $X := \Omega \setminus \mathcal{S}_g^o$ and call the support of g indifferently $\mathcal{S}_g$ or Y. We will denote by Z_0 the set of points which do not belong to a non-trivial transport segment: we have $Z_0 \subset \mathcal{H}$ (with our notation we have $\mathcal{H} = X \cap Y$). Also, we will denote by Z_1 the union of all the degenerate transport segments, i.e. the union of all the transport segments which have empty intersection with $\mathcal{S}_g^o$. Finally, we call E the set of endpoints of non-trivial transport segments (degenerate or not). With our notations, Ω^* is composed of three parts, which are $\mathcal{T}$, Z_0 and Z_1, and we have $\mathcal{T} \cap (Z_0 \cup Z_1) = \mathcal{T} \cap Z_1 \subset E$ (the only possible intersection of $\mathcal{T}$ with $Z_0 \cup Z_1$ is made of endpoints belonging to more than one transport segment, one which is degenerate and one which is not).

The strategy to prove Theorem 1.1 can be sketched as follows.

- We will prove that $Z_0 \cup Z_1$ is closed. The only difficulties arise in the following cases: either a sequence of transport segments in Z_1 converges to a part of a longer transport segment, which intersects $\mathcal{S}_g^o$, or a sequence of points in Z_0 converges to a point on $\mathcal{H}$ which belongs to a point on a longer transport segment, also intersecting $\mathcal{S}_g^o$. Yet, in both cases the longer segment also enters $\mathcal{S}_f^o$ and we will find a contradiction thanks to Lemma 1.2.

- Inside $\mathcal{T}$ the situation will be described clearly in Section 5. We will show that, close to a point in $\mathcal{T} \setminus E$, the maps associating with every point the upper and lower endpoints of the (unique) transport segment which contains it are continuous, and deduce from this that the set $\mathcal{T} \setminus E$ is open and that the optimal mapping is continuous on it.
- The optimal transport mapping is the identity on Z_0 and points of Z_0 can only be approximated by segments whose length tends to 0 (because of point (i) in Lemma 2.3), so that the map is continuous on Z_0.
- The optimal transport mapping s is also the identity on Z_1, but on Z_1 we have $f = g = 0$, i.e. $\varrho = \varrho^*$. Thanks to this and the continuity of ϱ and ϱ^*, we will prove that for any sequence $x_n \to x$ with $x \in Z_1$ we necessarily have $|s(x_n) - x_n| \to 0$, which proves the continuity at $x \in Z_1$.
- Finally, we are left to prove continuity of s on E. In order to do this, and using the lower bound on ϱ^*, we will prove that for any sequence $x_n \to x$ with x an endpoint of a transport segment, the distance between x_n and the endpoint of the transport segment on which it lies must necessarily tend to 0, and deduce from this that we also have $|s(x_n) - x_n| \to 0$.

In Theorem 1.1, we assume that either $\mathcal{S}_f^o$ or $\mathcal{S}_g^o$ is convex. This convexity condition is essentially used so as to guarantee that any transport segment does not cross $\mathcal{S}_f^o$ or $\mathcal{S}_g^o$ multiple times. In the following let us assume that $\mathcal{S}_g^o$ is convex. The argument below also applies to the case when $\mathcal{S}_f^o$ is convex with very minor changes (in particular, if $\mathcal{S}_f^o$ is supposed to be convex instead of $\mathcal{S}_g^o$, then it would be more convenient to define $\mathcal{H} := \partial\mathcal{S}_f$ instead of $\mathcal{H} := \partial\mathcal{S}_g$). Indeed, the reader can check that most of what we prove in Sections 3 and 4 is completely symmetric in $\mathcal{S}_f$ and $\mathcal{S}_g$, and the arguments could be reversed. When differences between the roles are important, we will underline it. In Section 5, where we prove continuity of the mapping, the role of the two measures is different, but in such a section the key assumption is the convexity of the two domains Ω and Ω^*, with $\Omega \subset \Omega^*$.

In Sections 3 and 4 we will investigate properties of transport segments for the functional (2.4). In Section 5 we use these properties to prove the continuity of the monotone optimal mapping for the original problem (1.2). Note that the optimal mapping for the densities f and g is not continuous in general, as f and g have disjoint supports and the optimal mapping will not be continuous at double points.

3. The set of double points

As in [14], a point is called a double point if there are *at least* two transport segments sharing this point as the common endpoint.

Denote by $\mathcal{D}_X$ and $\mathcal{D}_Y$ the sets of double points in X and in Y, and denote by E_X and E_Y the set of endpoints of transport segments in X and in Y, respectively. We have $E = E_X \cup E_Y$.

Lemma 3.1. *Let $p \in E_X$ be an interior point of X, which is the upper endpoint of a transport segment ℓ_0. Let $p_j \in X$ be a sequence of points converging to p, and ℓ_j be a transport segment containing p_j. If p is not a double point, then $\ell_j \cap X$ converges to $\ell_0 \cap X$. If p is a double point, then $\ell_j \cap X$ converges, up to subsequences, to $\ell \cap X$ for a transport segment ℓ having p as an upper endpoint (but not necessarily ℓ_0).*

Proof. Since p is an interior point of X, the length $|\ell_j|$ has a uniform positive lower bound. Hence we may assume by passing to a subsequence that ℓ_j converges to a segment ℓ'. Obviously, $p \in \ell'$. If p is not a double point, then by Theorem 2.3 (ii), we must have $\ell' = \ell_0$. □

We will also say that a double point is an *oriented double point* whenever all the transport segments sharing it as an endpoint lie in a same half-space at such a point, with no pair of transport segments forming an angle π. Equivalently, this means that p is an oriented double point if there exists a vector v such that for every transport segment ℓ with endpoint p and for every $x \in \ell \setminus \{p\}$ we have $v \cdot (x - p) > 0$. Hence v gives an orientation to all the transport segments at p. With the assumption that Y is convex, all points in $\mathcal{D}_X$ are automatically oriented double points. Note that double points cannot belong to $\mathcal{H}$, namely $\mathcal{D}_X \cap \mathcal{H} = \emptyset$, otherwise the monotonicity of the optimal mapping (or the measure-preserving property) would be violated. Moreover, if $p \in X$ and Y is convex, then the angles between two transport segments sharing p as an endpoint is uniformly far from π as soon as we stay away from Y.

On the contrary, non-oriented double points could exist in $\mathcal{D}_Y$. A typical case of non-oriented double point is the point $p = 0$ whenever the Kantorovich potential is given by $u(x) = |x|$.

For any point $p \in \mathbb{R}^2$ and two line segments ℓ_1 and ℓ_2 which do not lie on a same straight line and share the endpoint p, we denote by $\angle_{p;\ell_1,\ell_2}$ the convex region bounded by the two rays determined by ℓ_1 and ℓ_2. The angle between ℓ_1 and ℓ_2 is then less than π. We will often apply this to the case where p is a double point and ℓ_1 and ℓ_2 are transport segments from this point.

Denote

$$X_{p;\ell_1,\ell_2} = X \cap \angle_{p;\ell_1,\ell_2},$$
$$Y_{p;\ell_1,\ell_2} = Y \cap \angle_{p;\ell_1,\ell_2}.$$

The following statement shows that any interior double point belongs to a $\mathcal{D}$-curve. See also Lemma 5.8 in [14].

Lemma 3.2. *Let ℓ_1, ℓ_2 be two transport segments with the same upper endpoint p. Let $q \in X_{p;\ell_1,\ell_2}$ be an upper endpoint of a transport segment or a point in $Z_0 \subset \mathcal{H}$. Then there is a continuous curve of double points in $X_{p;\ell_1,\ell_2}$ which connects p and q. Moreover, the curve is locally Lipschitz continuous.*

Proof. Denote $\vec{r}_\theta = (\cos\theta, \sin\theta)$ a unit vector in $\mathbb{R}^2$. Choosing the coordinates properly, we assume that $p = 0$, $\ell_1 = \{t\vec{r}_{-\alpha} \mid t \in (0, a_1)\}$, and $\ell_2 = \{(t\vec{r}_\alpha \mid t \in (0, a_2)\}$, for some $\alpha \in (0, \frac{1}{2}\pi)$. Then $q = (y_1, y_2)$ with $y_1 > 0$ and $-\tan\alpha < y_2/y_1 < \tan\alpha$.

Let $\phi(\theta)$ be a function, where $\theta \in (-\alpha, \alpha)$, such that the radial graph of ϕ, $G_\phi := \{\phi(\theta)\vec{r}_\theta \mid \theta \in (-\alpha, \alpha)\}$, is a smooth curve passing through the point q and G_ϕ is contained in $X_{p;\ell_1,\ell_2}$.

For any $t \in (0, 1)$ and any $\theta \in (-\alpha, \alpha)$, there is a transport segment ℓ_θ containing the point $t\phi(\theta)\vec{r}_\theta$. When θ is very close to $-\alpha$ or α, $\ell_\theta \cap X$ is also very close to ℓ_1 or ℓ_2, respectively. If there is no double point on $G_{t\phi}$, then ℓ_θ depends continuously on θ. Hence when we move the point θ from $-\alpha$ to α, there must be a point $p_t = t\phi(\theta)\vec{r}_\theta \in G_{t\phi}$ such that $q \in \ell_{p_t}$. But q is an endpoint point of a transport segment, or a point not belonging to a non-trivial transport segment, and this is impossible by Theorem 2.3 (ii).

Hence there must be a point $p_t \in G_{t\phi}$ such that p_t is a double point and there exist two transport segments $\ell_{1,t}$ and $\ell_{2,t}$ such that $q \in X_{p_t;\ell_{1,t},\ell_{2,t}}$. The property $q \in X_{p_t;\ell_{1,t},\ell_{2,t}}$ implies a bound on the slopes of the chords between the points of $\{p_t \mid t \in (0, y_1)\}$ and implies that this set is a Lipschitz curve. $\square$

Remark 3.1. *The same construction can be applied for points $p, q \in Y$, whenever q is not a double point, or is an oriented double point.*

By Lemma 3.2, we can give the following definition of $\mathcal{D}$-curves.

Definition 3.1. *A $\mathcal{D}$-curve γ in X is a Lipschitz continuous curve such that for any two points $p_1, p_2 \in \gamma$ with $p_i = \gamma(t_i)$ and $t_1 < t_2$, we have $p_2 \in X_{p_1;\ell^+,\ell^-}$, where $\ell^\pm$ are transport segments with upper endpoint p_1. Note that this implies that $u(\gamma(t))$ is a decreasing function of t (see Remark 3.4). We will also say that a $\mathcal{D}$-curve γ ends at p_2 if it is parametrised on $[t_1, t_2]$ and $p_2 = \gamma(t_2)$ (without imposing that $[t_1, t_2]$ is a maximal parametrisation interval, however).*

Similarly one can define a $\mathcal{D}$-curve in Y.

Note that whenever a $\mathcal{D}$-curve γ ends at p_2, it is not necessary that p_2 is a double point. However, we can prove the following:

Lemma 3.3. *If $p \in E_X$ is an interior point of X, then there exists a $\mathcal{D}$-curve ending at p. In particular, $p \in \overline{\mathcal{D}}_X$, the closure of the set $\mathcal{D}_X$,*

Proof. If p is not a double point, there is a unique transport segment ℓ_0 with upper endpoint p, otherwise let us select two transport segments with maximal angle between them, having p as an upper endpoint. This angle is necessarily smaller than π because of the convexity assumption on Y. We will call $\ell^\pm$ these two transport segments, and set $\ell^+ = \ell^- = \ell_0$ in case p is not a double point.

Choose the coordinates such that $p = 0$, $\ell^+ = \{(t, 0) \mid t \in [0, a]\}$ for some $a > 0$, and ℓ^- lies below ℓ^+ (i.e. in the half-space $\{x_2 \le 0\}$).

First we prove that, for sufficiently small $\varepsilon > 0$, a (uniqueness is not important) transport segment passing through $(\varepsilon, \varepsilon^2)$ lies above ℓ^+ (we mean that the part of this segment between $(\varepsilon, \varepsilon^2)$ and its lower endpoint stays in the quadrant $\mathbb{R}_+ \times \mathbb{R}_+$). If it is not the case, then a part of this segment stays in the half-space $\mathbb{R}_- \times \mathbb{R}$. This cannot occur for small ε, otherwise at the limit we would have a transport segment lying in $\mathbb{R}_- \times \mathbb{R}$ and starting from p, which contradicts either the fact that p is not a double point, or the maximality of the angle between the selected transport segments. Let us call p^+ the point $(\varepsilon, \varepsilon^2)$ that we find in this way, such that a transport segment through it lies above ℓ^+.

We now repeat the same construction by looking at transport segments lying below ℓ^-. For this it is enough to repeat the same construction using the segment ℓ^- instead of ℓ^+, and we orientate the coordinate so as to find a point p^- with a transport segment passing through p^- and lying below ℓ^-.

If we connect p^+ to p^- with a curve C, staying inside X, and disjoint from ℓ^+ and ℓ^-, then, as in the proof of Lemma 3.2, there is a double point $\tilde{p} \in C$ and two transport segments $\tilde{\ell}^+$ and $\tilde{\ell}^-$ such that $p \in X_{\tilde{p};\tilde{\ell}^+,\tilde{\ell}^-}$. Lemma 3.2 then implies that there is a $\mathcal{D}$-curve connecting $\tilde{p}$ to p and $p \in \overline{\mathcal{D}}_X$. □

Remark 3.2. *The same construction applies in Y, provided p is not a double point, or is an oriented double point.*

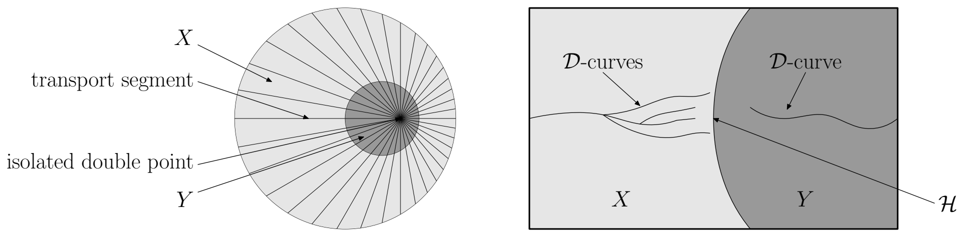

FIGURE 2. Double points and $\mathcal{D}$-curves

We now give a key proposition, valid as usual in the case where Y is convex.

Proposition 3.4. *Every $\mathcal{D}$-curve in X must connect to $\partial X \cap \partial\Omega$.*

Proof. From Lemma 3.3 we know that any $\mathcal{D}$-curve starting from a double point p in the interior of X can be extended "behind" p. Consider now a maximal $\mathcal{D}$-curve, which cannot be extended. Either it has a starting point which is not in the interior of X (in which case the claim is proved), or it has no starting point, i.e. it is parametrized over an open integral. We can see that this last case is not possible, as it is always possible to extend a $\mathcal{D}$-curve parametrized over $]a, b]$ to the whole $[a, b]$. For this, if γ is a parametrization of this curve, it is enough to see that $\lim_{t\to a^+} \gamma(t)$ exists, and that this limit is a double point. If the limit does not exists, then there are at least two accumulation points $p_0 \neq p_1$. But in this case one would have an infinity of double points $p_0^n \to p_0$ and $p_1^n \to p_1$ such that p_0^n is contained in the cone determined by the transport segments at p_1^n and p_1^{n+1} in the cone at p_0^n. But this is impossible, as the transport segments would meet.

Hence a limit $p = \lim_{t\to a^+} \gamma(t)$ exists, and it is a double point, as the transport segments $\ell_t^\pm$ at the double points $\gamma(t)$ converge to two transport segments $\ell^\pm$ passing through p, and these two segments cannot coincide as the angle between ℓ_t^+ and ℓ_t^- is bounded from below. □

Remark 3.3. *The situation is slightly different in Y. Indeed, we can say: every $\mathcal{D}$-curve in Y connects either to $\partial Y \cap \partial\Omega^*$, or to a non-oriented double point in Y.*

Remark 3.4. *Note that, whenever p is the upper endpoint of two transport segments $\ell^{\pm}$ with lower endpoints $q^{\pm}$, then we have $u(x) < u(p)$ for every $x \neq p$ belonging to the triangle with vertices p, q^{+} and q^{-}. This implies that the potential u is monotone decreasing along $\mathcal{D}$-curves, starting from the boundary. Analogously, u is monotone increasing along $\mathcal{D}$-curves in Y, starting from the boundary or from non-oriented double points. Also note that the set of non-oriented double points is exactly composed of the set of local minimum points for u in the interior of Y. Indeed, every local minimum point p for u must be an endpoint of a transport segment (otherwise u is affine on a segment containing p in its relative interior) and must be a non-oriented double point otherwise there would be a $\mathcal{D}$-curve ending at p on which the value of u would be strictly less than $u(p)$. On the other hand, if p is a non-oriented double point, then a neighbourhood of p could be covered with triangles bounded by transport segments, and u would be strictly larger than $u(p)$ in such a neighbourhood. Finally, the neighbourhood V_p of a non-oriented double point p on which p is minimal has a size which is at least comparable to $d(p, \mathcal{H})$.*

Remark 3.5. *Note that it is not true in general that non-oriented double points are all global minima for u, as one can see from the following example of potential u. Consider the strip $\{|x_2| \leq 2\}$ in $\mathbb{R}^2$, divided into four regions: T^+ is the triangle with vertices $p_1 = (-2, 0)$, $p_2 = (1, 0)$ and $p_3 = (0, 2)$, T^- is the triangle with vertices p_1, p_2 and $p_4 = (0, -2)$, A is the part of the strip on the left of $T^+ \cup T^-$, and B the part on the right. Consider $u(x) = |x - p_1|$ for $x \in A$, $u(x) = 2\sqrt{2} - |x - p_3|$ for $x \in T^+$, $u(x) = 2\sqrt{2} - |x - p_4|$ for $x \in T^-$, and $u(x) = 2\sqrt{2} - \sqrt{5} + |x - p_2|$ for $x \in B$. This is a Lip_1 function, where p_1 and p_2 are non-oriented double points but $u(p_1) < u(p_2)$. One can take $Y = \{|x_2| \leq 1\}$ and $X = \{1 \leq |x_2| \leq 2\}$. On the other hand, it does not seem possible to put on X and Y continuous densities such that this function u is their Kantorovich potential.*

Note that a $\mathcal{D}$-curve can start from a point on another $\mathcal{D}$-curve, as shown in Figure 2 above.

4. Proof of Lemma 1.2

In this section we prove Lemma 1.2 under the assumptions of Theorem 1.1. Our idea is as follows. If there exists a sequence of $\mathcal{D}$-curves in $\mathcal{T}$ whose lengths are greater than a given positive constant, then there exists a pair of transport segments in a narrow strip. We will show that there is a sequence of transport segment pairs in the strip such that the angles between these segment pairs increase exponentially. But this is impossible as all these segments fall in a narrow strip and the angles between them cannot exceed a given, small, value.

Let $p_0 \in X$ be a double point, ℓ^+ and ℓ^- in $\mathcal{T}$ be two transport segments from p_0 with lower endpoints q^+ and q^-, respectively. We denote by $|\ell^+|$ and $|\ell^-|$ the length of the segments ℓ^+ and ℓ^-, and by $\theta[\ell^+, \ell^-]$ the angle between the segments ℓ^+ and ℓ^-.

Choose a coordinate system such that ℓ^+ lies on the x_1-axis, ℓ^- in $\{x_2 \le 0\}$, $p_0 = (-a, 0)$, $q^+ = (b, 0)$, and the origin $0 \in \mathcal{H}$, where $a, b > 0$. Then $\ell^+ \cap X$ is contained in $\{x_1 \le 0\}$, $\ell^+ \cap Y$ in $\{x_1 \ge 0\}$. Moreover, ℓ^- is given by $x_2 = \eta(x_1 + a)$, where the slope η is given by $\eta = -\tan\theta$. For the proof of Lemma 1.2, it suffices to consider the case when $\theta < \frac{\pi}{2}$ is small.

We also denote

$$\mathcal{L}_c = \{(x_1, x_2) \in \mathbb{R}^2 \mid x_1 = c\},$$

which is the straight line parallel to the x_2-axis with abscissa c.

Lemma 4.1. *Assume that* $|\ell^+| \le |\ell^-|$. *Then*

$$b = \kappa^+ + o(1), \tag{4.1}$$

where $o(1) \to 0$ *when* $\theta = \theta[\ell^+, \ell^-] \to 0$. *The constant* κ^+ *is determined by*

$$\int_{-a}^{0} (x_1 + a) f(x_1, 0) dx_1 = \int_0^{\kappa^+} (x_1 + a) g(x_1, 0) dx_1. \tag{4.2}$$

Proof. Let us divide our argument into two steps: (i) $b \le \kappa^+ + o(1)$; (ii) $b \ge \kappa^+$. Our idea is to find suitable subsets of $\mathcal{T}$ on which f and g are of mass balance, and then use the continuity of f and g to conclude some one dimensional integral identities with small error terms depending on θ. These identities together with (4.2) consequently imply (i) and (ii).

Step 1: Let us first show that $b \le \kappa^+ + o(1)$ as $\theta \to 0$.

Denote $\triangle = X_{p_0;\ell^+,\ell^-}$, $S_d = \angle_{p_0;\ell^+,\ell^-} \cap \{0 < x_1 < d\}$. Then as $\theta \to 0$, we have

$$\int_\triangle f = \theta \int_{-a}^{0} (x_1 + a) f(x_1, 0) dx_1 + o(\theta), \tag{4.3}$$

$$\int_{S_{\kappa^+}} g = \theta \int_0^{\kappa^+} (x_1 + a) g(x_1, 0) dx_1 + o(\theta). \tag{4.4}$$

The term $o(\theta)$ above is because $\triangle$ and S_{κ^+} lies in a narrow trip $\{(x_1, x_2) \mid -C\theta \le x_2 \le 0\}$, and f, g are continuous on X and Y (hence uniformly continuous). Here C is a positive constant only depending on the diameter of Ω^*. One also easily verifies that $o(\theta)$ in (4.3)

(4.4) can be controlled by $C\theta\omega_f(C\theta)$ and $C\theta\omega_g(C\theta)$ respectively, where ω_f and ω_g are the modulus of continuity of f and g.

Assume to the contrary that $b \geq \kappa^+ + 2\varepsilon_0$ for some $\varepsilon_0 > 0$ (independent of θ). We claim that, for sufficiently small θ, we have

$$\mathcal{S}_g \cap S_{\kappa^+ + \varepsilon_0} \subset s(\triangle), \tag{4.5}$$

where s is the unique monotone optimal mapping sending the density f onto g.

Indeed, if (4.5) is not true, then there is a point $q^* \in \mathcal{L}_{\kappa^+ + \varepsilon_0} \cap \angle_{p_0;\ell^+,\ell^-}$ such that $q^* \notin s(\triangle)$, namely there is a Lebesgue point $p^* \in X \setminus \triangle$ of s such that $s(p^*) = q^*$. Then $\ell^* := \overline{p^* q^*}$ is contained in a transport segment. Considering that this transport segment cannot cross ℓ^+ or ℓ^-, the only possibility, when $\theta[\ell^+, \ell^-]$ is sufficiently small, is that the transport segment ℓ^* is in a direction opposite to ℓ^+ (namely $(p^* - q^*) \cdot (p_0 - q^+) < 0$). Yet, this means that s is not a monotone mapping. Hence (4.5) holds.

By (4.5) and the mass balance condition we then have

$$\begin{aligned} \int_\triangle f &= \int_{s(\triangle)} g \\ &\geq \int_{S_{\kappa^+}} g + C_0 \varepsilon_0 \theta \end{aligned}$$

for some $C_0 > 0$ independent of θ. The constant C_0 depends on the value of g at the point $(0, \kappa^+)$. Notice that we can assume that this point belongs to the open set $\{g > 0\}$ (which is convex), otherwise the inequality $b \leq \kappa^+ + o(1)$ is automatically satisfied (as a consequence of $(0, b) \in \mathcal{S}_g$). Yet, the dependence of C_0 on g is not important in what follows.

By (4.3) and (4.4), we obtain from the above inequality,

$$\int_{-a}^{0} (x_1 + a) f(x_1, 0) dx_1 \geq \int_0^{\kappa^+} (x_1 + a) g(x_1, 0) dx_1 + C_0 \varepsilon_0 + o(1), \tag{4.6}$$

which is in contradiction with (4.2) when θ is sufficiently small.

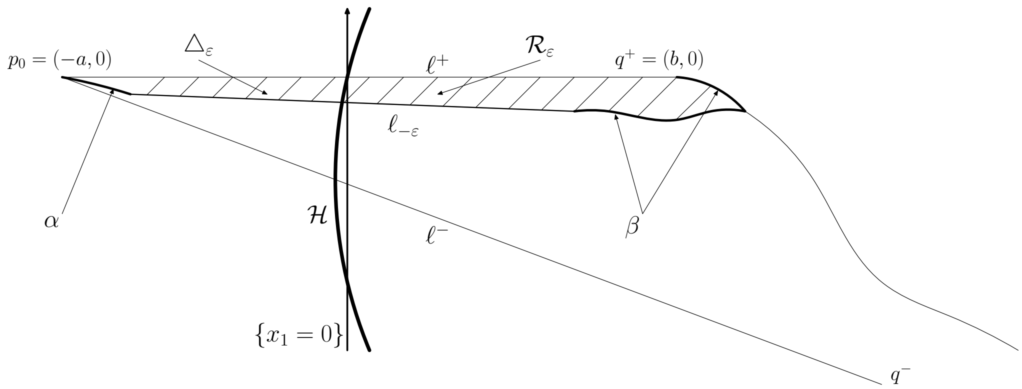

FIGURE 3.

Step 2: We have proved that $b \le \kappa^+ + o(1)$ as $\theta \to 0$. Now we prove that $b \ge \kappa^+$.

First observe that for any sequence of transport segments ℓ_j with upper endpoint $p_j \in X_{p_0;\ell^+,\ell^-}$ and lower endpoint $q_j \in Y$, such that the point $z_j \to 0$, where $z_j = \ell_j \cap \{x_1 = 0\}$ is the intersection of ℓ_j with the x_2-axis, we have, by Theorem 2.3 (i),

$$q_j \in \{x_1 < b + o(1)\} \quad \text{when } j \to \infty. \tag{4.7}$$

Set

$$T_\varepsilon = \{(x_1, x_2) \in \ell_t \mid -\varepsilon < t < 0\},$$

where ℓ_t is the transport segment with upper endpoint in $X_{p_0;\ell^+,\ell^-}$ and lower endpoint in Y such that the intersection of ℓ_t with the x_2-axis is the point $(0, t)$, namely $\ell_t \cap \{x_1 = 0\} = (0, t)$. In other words, T_ε is the area occupied by the transport segments ℓ_t with $-\varepsilon < t < 0$.

The boundary of T_ε consists of four parts: the upper one is the transport segment ℓ^+, the lower one is the transport segment $\ell_{-\varepsilon}$ with $\ell_{-\varepsilon} \cap \{x_1 = 0\} = (0, -\varepsilon)$, the left one is a $\mathcal{D}$-curve $\alpha \subset X_{p_0;\ell^+,\ell^-}$ (by Lemma 3.2, there is a $\mathcal{D}$-curve connecting the upper endpoint of $\ell_{-\varepsilon}$ to p_0) and the right one is contained in Y, which we denote by β. Here one of α and β may be a single point. By Lemma 3.2 and (4.7), we may choose ε small such that

$$\begin{aligned} &\alpha \subset \{-a \le x_1 < -a + \delta\}, \\ &\beta \subset \{x_1 < b + \delta\}, \end{aligned} \tag{4.8}$$

where $\delta \to 0$ as $\varepsilon \to 0$. More precisely, we do not claim that this happens for all ε small, but, for given ε, we can always choose a point in T_ε very close to p_0 and use as a segment $\ell_{-\varepsilon'}$ the transport segment passing though such a point (and we wil have $\varepsilon' < \varepsilon$). See

Figure 3. On the other hand, it is not sure that such a segment arrives close to $x_1 = b$. Suppose hence that we have a sequence of transport segments $\ell_{-\varepsilon}$ with $\varepsilon \to 0$ such that, calling $p_\varepsilon = (p_{1,\varepsilon}, p_{2,\varepsilon})$ and $q_\varepsilon = (q_{1,\varepsilon}, q_{2,\varepsilon})$ their upper and lower endpoints, respectively, we have $\lim_{\varepsilon\to 0} p_{1,\varepsilon} = -a$ and $\lim_{\varepsilon\to 0} q_{1,\varepsilon} = b'$. We can choose the minimal b' for which we can construct such a sequence (hence b' is defined as a $\liminf_{\varepsilon\to 0}$). In general, we only know $b' \le b$).

The boundary $\mathcal{H}$ divides the region T_ε into two parts, the left part $\triangle_\varepsilon$ and the right part $\mathcal{R}_\varepsilon$. By the mass balance condition we have

$$\int_{\triangle_\varepsilon} f = \int_{\mathcal{R}_\varepsilon} g.$$

By the first inclusion in (4.8) we have

$$\int_{\triangle_\varepsilon} f \ge \int_{-a}^{0} |x_2| f(x_1, 0)dx_1 - o(\varepsilon),$$

where x_2 is a function of x_1 which denotes the second coordinate component of the straight line that contains the transport segment $\ell_{-\varepsilon}$, see (4.11) below. Using again the first inclusion in (4.8) we see that $\mathcal{R}_\varepsilon \cap \{x_1 < b\}$ is contained in the region

$$\big\{(x_1, x_2) \in \mathbb{R}^2 \bigm| -a < x_1 < b, -[\frac{\varepsilon}{a} + o(\varepsilon)](x_1 + a) < x_2 < 0\big\}.$$

Also, we will call $\mathcal{Q}_\varepsilon$ the subset or $\mathcal{R}_\varepsilon$ contained in the region $\{x_1 > b'\}$. Therefore we obtain

$$\int_{-a}^{0} |x_2| f(x_1, 0)dx_1 \le \int_{0}^{b'} |x_2| g(x_1, 0)dx_1 + \int_{\mathcal{Q}_\varepsilon} g + o(\varepsilon), \tag{4.9}$$

where the term $o(\varepsilon)$ also appears because $\triangle_\varepsilon$ and $\mathcal{R}_\varepsilon \cap \{x_1 < b'\}$ are contained in the strip $\{(x_1, x_2) \in \mathbb{R}^2 \mid -C\varepsilon \le x_2 \le 0\}$. We want now to estimate the term $\int_{\mathcal{Q}_\varepsilon} g$, which is a sort of error term. To do so, let us fix a number δ_0 and consider the area bounded by ℓ^+, by $\{x_1 = b' - \delta_0\}$, by $\{x_1 = b'\}$, and by $\ell_{-\varepsilon}$ and call it A. Two kind of points are contained in $\mathcal{Q}_\varepsilon$: those whose inverse images through s are contained in A (denote by $\mathcal{Q}^1_\varepsilon$ the set of such points) and those who have at least one inverse image outside of A. Note that, if $A \cap \Delta_\varepsilon = \emptyset$, then the first class of points is empty. Anyway, for this points, by using the fact that s transports f onto g, we have $\int_{Q^1_\varepsilon} g \le \int_{A\cap\Delta_\varepsilon} f \le C\varepsilon\hat{\delta}_0$. For the other points, we notice that they must belong to a transport segment whose slope is at most $C\varepsilon/\delta_0$, and that, by the second inclusion in (4.8), the measure of $\mathcal{Q}_\varepsilon \setminus s(A)$ is at most $C(\varepsilon/\delta_0)(b - b') + o(\varepsilon)$. Hence we get

$$\int_{-a}^{0} |x_2| f(x_1, 0)dx_1 \le \int_{0}^{b'} |x_2| g(x_1, 0)dx_1 + C\varepsilon\delta_0 + C\frac{\varepsilon}{\delta_0}(b - b') + o(\varepsilon).$$

We can decide to choose $\delta_0 = \sqrt{b-b'}$, thus getting

$$\int_{-a}^{0} |x_2| f(x_1,0)dx_1 \le \int_0^{b'} |x_2| g(x_1,0)dx_1 + C\varepsilon\sqrt{b-b'} + o(\varepsilon). \tag{4.10}$$

In our coordinates, $\ell_{-\varepsilon}$ can be expressed in the form

$$\ell_{-\varepsilon} = \big\{(x_1,x_2)\in\mathbb{R}^2 |\ x_2 = \eta_\varepsilon x_1 - \varepsilon, \quad -a_\varepsilon \le x_1 \le b_\varepsilon\big\}, \tag{4.11}$$

where $a_\varepsilon = a + o(1)$, $b_\varepsilon \le b + o(1)$ and $o(1)\to 0$ as $\varepsilon\to 0$. Since $\ell_{-\varepsilon}\subset \angle_{p_0;\ell^+,\ell^-}$, we have $x_2 = \eta_\varepsilon x_1 - \varepsilon \le 0$ for $x_1\in(-a_\varepsilon,0)$. In particular $-\eta_\varepsilon a_\varepsilon - \varepsilon \le 0$, namely $-\eta_\varepsilon/\varepsilon \le 1/a_\varepsilon$. Hence when $x_1<0$, we have

$$|x_2| = \varepsilon(-\frac{\eta_\varepsilon}{\varepsilon}x_1 + 1) \ge \varepsilon(\frac{1}{a_\varepsilon}x_1 + 1).$$

Divide (4.10) by ε. For the left hand side we can write

$$\ge \int_{-a}^{0} (\frac{1}{a}x_1 + 1) f(x_1,0)dx_1 + o(1).$$

Similarly, for the right hand side we have

$$\le \int_0^{b'} (\frac{1}{a}x_1 + 1) g(x_1,0)dx_1 + C\sqrt{b-b'} + o(1).$$

Summarizing, and sending $\varepsilon\to 0$, we get

$$\int_{-a}^{0} (\frac{1}{a}x_1 + 1) f(x_1,0)dx_1 \le \int_0^{b'} (\frac{1}{a}x_1 + 1) g(x_1,0)dx_1 + C\sqrt{b-b'}. \tag{4.12}$$

In the case $b'=b$, this means

$$\int_{-a}^{0} (\frac{1}{a}x_1 + 1) f(x_1,0)dx_1 \le \int_0^{b} (\frac{1}{a}x_1 + 1) g(x_1,0)dx_1. \tag{4.13}$$

Hence by the definition of κ^+ in (4.2), we obtain $b\ge\kappa^+$.

In case $b'<b$, the claim is proven if we prove $b'\ge\kappa^+$. To do so, consider a sequence $\varepsilon_n\to 0$ such that the lower endpoints of $\ell_{-\varepsilon_n}$ converge to $(b',0)$ and first apply the same argument to the segments $\ell_{-\varepsilon_n}$ instead of ℓ^+. For each n we will have a number $b(n)$ corresponding to the lower endpoint of $\ell_{-\varepsilon_n}$, and a number $b'(n)$ corresponding to the length that can be reached by segments approximating $\ell_{-\varepsilon_n}$. Note that $0 \le b(n)-b'(n)\to 0$, because of the definition of b' as the minimal reachable length (hence, $\liminf_n b'(n)\ge b'$, while $b'(n)\le b(n)$ and $\lim_n b(n) = b'$).

Formula (4.12), applied for each n, gives

$$\int_{-a}^{0} (\frac{1}{a}x_1 + 1) f(x_1,0)dx_1 \le \int_0^{b'(n)} (\frac{1}{a}x_1 + 1) g(x_1,0)dx_1 + C\sqrt{b(n)-b'(n)} + o(1), \tag{4.14}$$

where the $o(1)$ tends to 0 as $n \to \infty$, and is due to the change in the reference system passing from ℓ^+ to $\ell_{-\varepsilon_n}$, and to the difference in the upper endpoint of the segments $\ell_{-\varepsilon_n}$, which are not equal to $(-a, 0)$ but converge to such a point.

Letting $n \to \infty$ and using $b(n) - b'(n) \to 0$, we get now

$$\int_{-a}^{0} (\frac{1}{a}x_1 + 1) f(x_1, 0) dx_1 \leq \int_{0}^{b'} (\frac{1}{a}x_1 + 1) g(x_1, 0) dx_1, \tag{4.15}$$

which proveds $b' \geq \kappa^+$. □

We remark that κ^+ is uniquely determined by (4.2), as Y is convex and $g > 0$ in Y. Note that at the end of Section 2, we define the set $X = \overline{\Omega \setminus Y}$, which allows that $f = 0$ in a subset of X. But $f \not\equiv 0$ in $X_{p_0;\ell^+,\ell^-}$, because of the mass balance condition

$$\int_{X_{p_0;\ell^+,\ell^-}} f = \int_{s(X_{p_0;\ell^+,\ell^-})} g,$$

and the right hand side does not vanish.

From the proof of Lemma 4.1, we have

Corollary 4.2. *Denote as above $S_{\kappa^+} = \angle_{p_0;\ell^+,\ell^-} \cap \{0 < x_1 < \kappa^+\}$. We have the estimate,*

$$\left|(S_{\kappa^+} \setminus s(X_{p_0;\ell^+,\ell^-})) \cup (s(X_{p_0;\ell^+,\ell^-}) \setminus S_{\kappa^+})\right| = o(\theta) \tag{4.16}$$

as $\theta = \theta[\ell^+, \ell^-] \to 0$, where s is the monotone optimal mapping transferring mass from f to g, and $|\cdot|$ denotes the Lebesgue measure on $\mathbb{R}^2$.

Proof. The conclusions $b \geq \kappa^+$ and $b \leq \kappa^+ + o(1)$ imply respectively that the sets $S_{\kappa^+} \setminus s(X_{p_0;\ell^+,\ell^-})$ and $s(X_{p_0;\ell^+,\ell^-}) \setminus S_{\kappa^+}$ have small Lebesgue measure in the sense indicated in the statement. □

The way Corollary 4.2 will be used is essentially the following: most of the set $\angle_{p_0;\ell^+,\ell^-}$ after the end of the shortest between the segments ℓ^+ and ℓ^- will be out of $s(X_{p_0;\ell^+,\ell^-})$, and this implies that the transport segments passing through the points of $\angle_{p_0;\ell^+,\ell^-} \setminus S_{\kappa^+}$ should be directed out of the set $s(X_{p_0;\ell^+,\ell^-})$. The use of this fact will be clear from Lemma 4.4.

The following corollary, altough not crucial for the sequel, could be helpful for the readers to understand the geometry of the $\mathcal{D}$-curves.

Corollary 4.3. *If p_0 is a double point, upper endpoint of two transport segments $\ell^\pm \subset \mathcal{T}$, and there is another double point $p_1 \neq p_0$ in $X_{p_0;\ell^+,\ell^-}$ and the angle θ between ℓ^+ and ℓ^- is small enough, then there is a $\mathcal{D}$-curve in $Y_{p_0;\ell^+,\ell^-}$. Moreover, the length of the*

$\mathcal{D}$-curve in $Y_{p_0;\ell^+,\ell^-}$ is greater than c_1 for some positive constant c_1 depending on the distance $|p_0 - p_1|$ and on $d(p_1, \mathcal{H})$.

Proof. By Lemma 3.2, there is a $\mathcal{D}$-curve connecting p_1 to p_0. We can suppose that the transport segments from p_1 enter $\mathcal{S}_g^o$, otherwise, if they stop on the boundary of Y, there is nothing to prove since in this case their lower endpoints are connected to the opposite boundary of Y. Let κ_1^+ be the positive number relative to p_1 defined in Lemma 4.1 (just as κ^+ is the corresponding number relative to p_0, see for example (4.20) below). Let ℓ_1^+ and ℓ_1^- be transport segments with upper endpoint at p_1. Note that $Y_{p_1;\ell_1^+,\ell_1^-}$ is a proper subset of $Y_{p_0;\ell^+,\ell^-}$. We see that κ_1^+ is strictly less than κ^+.

Indeed, under the coordinate system in Lemma 4.1, we have, by (4.16),

$$\frac{\int_{X_{p_0;\ell_1^+,\ell_1^-}} f}{\int_{s(X_{p_0;\ell_1^+,\ell_1^-})} g} = \frac{\int_{-a}^{0}(\frac{1}{a}x_1+1)f(x_1,0)dx_1}{\int_0^{\kappa^+}(\frac{1}{a}x_1+1)g(x_1,0)dx_1} + o(1)$$

and

$$\frac{\int_{X_{p_1;\ell_1^+,\ell_1^-}} f}{\int_{s(X_{p_1;\ell_1^+,\ell_1^-})} g} = \frac{\int_{-a_1}^{0}(\frac{1}{a_1}x_1+1)f(x_1,0)dx_1}{\int_0^{\kappa_1^+}(\frac{1}{a_1}x_1+1)g(x_1,0)dx_1} + o(1)$$

where $-a_1$ (with $0 < a_1 < a$) is the x_1-coordinate of p_1, and $o(1) \to 0$ as $\theta = \theta[\ell^+, \ell^-] \to 0$. Since $a_1 < a$, if $\kappa_1^+ \geq \kappa^+$, we infer that

$$\begin{aligned}
\int_{-a}^{0}(\frac{1}{a}x_1+1)f(x_1,0)dx_1 &> \int_{-a_1}^{0}(\frac{1}{a_1}x_1+1)f(x_1,0)dx_1, \\
\int_0^{\kappa^+}(\frac{1}{a}x_1+1)g(x_1,0)dx_1 &\leq \int_0^{\kappa_1^+}(\frac{1}{a_1}x_1+1)g(x_1,0)dx_1.
\end{aligned}$$

Hence

$$\frac{\int_{X_{p_1;\ell_1^+,\ell_1^-}} f}{\int_{s(X_{p_1;\ell_1^+,\ell_1^-})} g} < \frac{\int_{X_{p_0;\ell^+,\ell^-}} f}{\int_{s(X_{p_0;\ell^+,\ell^-})} g}.$$

But the ratios on both sides are equal to 1, by the mass balance condition.

Hence there is a transport segment whose lower endpoint q_0 lies in $\angle_{p_0;\ell^+,\ell^-} \cap Y^o$, where Y^o is the interior of Y. If θ is small enough, this endpoint cannot be a non-oriented double point. Indeed, for small θ the directions of the transport segments should be almost aligned with that of ℓ^+ or ℓ^-, which prevents from having non-oriented double points. Hence by Remark 3.2, there is a $\mathcal{D}$-curve in $\angle_{p_0;\ell^+,\ell^-} \cap Y$, which connects q_0 to either the boundary of Y or a non-oriented double point. The length of the $\mathcal{D}$-curve is greater than $\kappa^+ - \kappa_1^+$ or than $d(q_0, \mathcal{H})$, which is close to κ_1^+. Both these quantities can be estimated from below in terms of $|p_0 - p_1|$ or $d(p_1, \mathcal{H})$. □

Denote

$$\mathcal{T}_{p_0;\ell^+,\ell^-} = X_{p_0;\ell^+,\ell^-} \cup s(X_{p_0;\ell^+,\ell^-}).$$

The boundary of $\mathcal{T}_{p_0;\ell^+,\ell^-}$ consists of three parts, ℓ^+, ℓ^- and a part in Y. We point out that the part in Y can be a $\mathcal{D}$-curve which connects the lower endpoints of ℓ^+, ℓ^- to a non-oriented double point, and therefore strictly contained in the interior of Y. Analogously, we use the same notation when we fix a point q_0 in Y, with two transport segments r^+ and r^-: we set $\mathcal{T}_{q_0;r^+,r^-} = Y_{q_0;r^+,r^-} \cup s^{-1}(Y_{q_0;r^+,r^-})$ and this set has a boundary composed by three parts: r^+, r^-, and a part in X.

Lemma 4.4. *Let p_0, ℓ^+, ℓ^- be as above. Assume at least one of q^+ and q^- is an interior point of Y. Then there is a transport segment $r \not\subset \mathcal{T}_{p_0;\ell^+,\ell^-}$ such that*

$$\theta[\ell^+, r] \geq N\theta[\ell^+, \ell^-], \tag{4.17}$$

where $N > 1$ is as large as we want, provided $\theta[\ell^+, \ell^-]$ is sufficiently small.

Namely for any large constant N, there is a positive θ_0 depending on N, $d(p_0, \mathcal{H})$, f and g, such that if $\theta[\ell^+, \ell^-] < \theta_0$, then (4.17) *holds.*

Proof. First let us assume that q^+ is an interior point of Y and q^- is not. Then there is a $\mathcal{D}$-curve β_0 connecting q^+ to a point q^* on ∂Y (or to a non-oriented double point) which composes a part of the boundary of $\mathcal{T}_{p_0;\ell^+,\ell^-}$. As $\theta[\ell^+, \ell^-]$ is small, we have $|\ell^+| < |\ell^-|$ in this case. To fix the ideas and the coordinate system, we suppose that ℓ^+ is on the x_1 axes and ℓ^- lies below it, as in Figure 4.

By (4.5) we have $S_{\kappa^+ - \varepsilon_0} \subset \mathcal{T}_{p_0;\ell^+,\ell^-}$ for $\varepsilon_0 > 0$ small. As β_0 is part of the boundary of $\mathcal{T}_{p_0;\ell^+,\ell^-}$, we see that for any point $q \in \beta_0$, there is a transport segment from q, which does not lie in $\mathcal{T}_{p_0;\ell^+,\ell^-}$. Let

$$L = \{q^+ - te_2 = (q_1^+, q_2^+ - t) \mid 0 < t < t_0\}$$

be the line segment parallel to the x_2-axis, with one endpoint q^+ and the other $\tilde{q} = q^+ - t_0 e_2$ on ℓ^-, where $e_2 = (0, 1)$ is a unit vector. Since $\mathcal{T}_{p_0;\ell^+,\ell^-}$ is in a very thin area, from the proof of Lemma 3.2, β_0 is a Lipschitz graph. Hence by the area condition in (4.16), we see that for any point $q_t = q^+ - te_2$, where $t \in (0, \frac{9}{10}t_0)$, there is a point $q_t' \in \beta_0$ on the horizontal line $\{x_2 = q_2^+ - t\}$ and q_t' is very close to q_t.

Let $t^* = \frac{1}{2}t_0$ and let r be the transport segment with lower endpoint $q_{t^*}' \in \beta_0$, which does not lie in $\mathcal{T}_{p_0;\ell^+,\ell^-}$, but lies above β_0 and also above the point q^+.

We claim that (4.17) holds for the transport segment r. Indeed, it suffices to observe that (4.17) is invariant under the coordinate change $x_1' = x_1$ and $x_2' = x_2/h$, where $h = \theta[\ell^+, \ell^-]$. After the change, $\theta[\ell^+, \ell^-] \simeq 1$ but $\theta[\ell^+, r]$ can be as large as we want as the segment r is almost vertical after the change.

From Figure 4, one also easily verifies that

$$N \geq \frac{|q^+ - p_0|}{4|q'_{t^*} - q_{t^*}|}.$$

By Lemma 4.1 and Corollary 4.2, one can see that for any small $\varepsilon > 0$, there is a constant θ^* depending on the modulus of continuity of f, g, and local lower bound of g, such that $|q'_{t^*} - q_{t^*}| < \varepsilon$, provided $\theta[\ell^+, \ell^-] < \theta^*$. Hence N can be arbitrarily large if the angle between ℓ^+ and ℓ^- is sufficiently small.

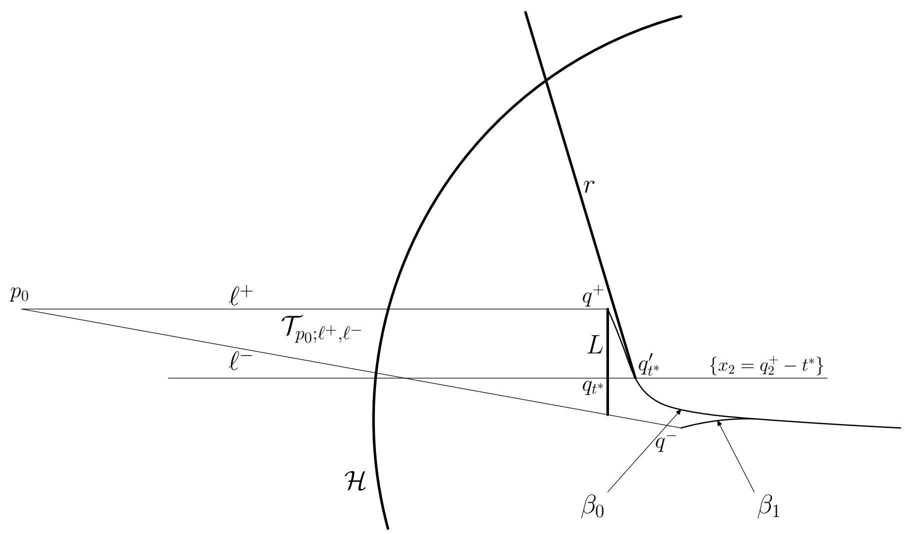

FIGURE 4.

Next we consider the case when both q^+ and q^- are interior points of Y. In this case there are $\mathcal{D}$-curves β_0 and β_1 connecting q^+ and q^- respectively to either the boundary ∂Y or non-oriented double points. There is no loss of generality in assuming that $|\ell^+| \leq |\ell^-|$, so that we can introduce the segment L as above, which connects a point on ℓ^+ to a point on ℓ^-. Then by (4.16), we see that for almost all $t \in (0, t_0)$ (except a very small subset of $(0, t_0)$), there is a point $q'_t \in \beta_0 \cup \beta_1$ on the horizontal line $\{x_2 = q_2^+ - t\}$, and q'_t is very close $q_t = q^+ - te_2$. In particular there exists $t^* \in (\frac{1}{3}t_0, \frac{2}{3}t_0)$ such that $q'_{t^*} \in \beta_0 \cup \beta_1$ and q'_{t^*} is very close to q_{t^*}.

Since q'_{t^*} is a double point and a boundary point of $\mathcal{T}_{p_0;\ell^+,\ell^-}$, there is a transport segment r with lower endpoint q'_{t^*} which does not lie in $\mathcal{T}_{p_0;\ell^+,\ell^-}$. As above one sees that (4.17) holds for the transport segment r. □

Note that (4.17) is equivalent to

$$\theta[\ell^-, r] \geq N\theta[\ell^+, \ell^-], \tag{4.18}$$

for large N, because the angle $\theta[\ell^+, \ell^-]$ is much smaller than $\theta[\ell^+, r]$.

Assume that $q'_{t^*} \in \beta_0$. Then for any point $q \in \beta_0$ between q'_{t^*} and q^*, there is a transport segment r_q with lower endpoint q, lying above the segment r. From Lemma 4.1, the length $|r_q \cap X|$ has a uniform positive lower bound. Hence by (4.17), we have

$$\theta[\ell^+, r_q] \geq N_1\theta[\ell^+, \ell^-], \tag{4.19}$$

with $N_1 \geq \frac{\kappa^+}{\mathrm{diam}(\mathcal{T})}N$.

Lemma 1.2 will follow from the following lemma immediately.

Lemma 4.5. *Let $p_0, p_1 \in \Omega \setminus Z_1$ be two double points with $p_1 \in X_{p_0;\ell^+,\ell^-}$ and the length $|p_0 - p_1| \geq c_0$ for some positive constant c_0. Then the angle $\theta[\ell^+, \ell^-] \geq c_1$ for some $c_1 > 0$, which depends on $c_0, d(p_0, \mathcal{H})$, diam(Ω^*), f and g.*

Proof. Since p_1 is a double point, there exist transport segments ℓ_1^+, ℓ_1^- with the common upper endpoint p_1, and lower endpoints $q_1^+, q_1^- \in \mathcal{T}_{p_0;\ell^+,\ell^-}$.

Corresponding to p_1 and ℓ_1^+, we can introduce κ_1^+ as in Lemma 4.1, namely

$$\int_{-a_1}^{0} (x_1 + a_1) f(x_1, 0)dx_1 = \int_0^{\kappa_1^+} (x_1 + a_1) g(x_1, 0)dx_1. \tag{4.20}$$

where $a_1 = a - |p_0 - p_1|$. Since $Y_{p_1;\ell_1^+,\ell_1^-}$ is a proper subset of $Y_{p_0;\ell^+,\ell^-}$ and g is positive in Y, we have $\kappa_1^+ < \kappa^+$.

Denote $\mathcal{L} = \{(x_1, x_2) \in \mathbb{R}^2 \mid x_1 = -a_1\}$ and $\mathcal{L}^* = \{(x_1, x_2) \in \mathbb{R}^2 \mid x_1 = b_1\}$, where $b_1 = \kappa_1^+ + \delta_0$ for some small $\delta_0 > 0$ such that $b_1 < \kappa^+$.

Since $\kappa_1^+ < \kappa^+$, at least one of q_1^+ and q_1^- is an interior point of Y. Hence by (4.19), there is a double point $q_1 \in \mathcal{L}^*$, which is on a $\mathcal{D}$-curve from either q_1^+ or q_1^- to either the boundary ∂Y or a non-oriented double point, and a transport segment r_1^+ with lower endpoint q_1 (r_1^+ not in $\mathcal{T}_{p_1;\ell_1^+,\ell_1^-}$) such that

$$\theta[\ell_1^+, r_1^+] \geq N\theta[\ell_1^+, \ell_1^-], \tag{4.21}$$

where N can be as large as we want, provided $\theta[\ell_1^+, \ell_1^-]$ is sufficiently small. Since q_1 is a double point, there is another transport segment r_1^- with lower endpoint at q_1, which lies in $\mathcal{T}_{p_1;\ell_1^+,\ell_1^-}$. Hence (4.21) implies

$$\theta[r_1^+, r_1^-] \geq N\theta[\ell_1^+, \ell_1^-]. \tag{4.22}$$

Note that the segment r_1^+ is in $\angle_{p_0;\ell^+,\ell^-}$. Hence both r_1^+ and r_1^- are located in the set $\mathcal{T}_{p_0;\ell^+,\ell^-}$.

Let $p_2^+, p_2^- \in X_{p_0;\ell^+,\ell^-}$ be the upper endpoints of the transport segments r_1^+, r_1^-. For the triple (q_1, r_1^+, r_1^-), we can define κ_1^- as in Lemma 4.1, namely

$$\int_{-\kappa_1^-}^{0} (b_1 - x_1) f(x_1, 0) dx_1 = \int_0^{b_1} (b_1 - x_1) g(x_1, 0) dx_1. \tag{4.23}$$

As f may vanish in a subset of X, the constant κ_1^- may not be unique. However, we can apply the same argument of this proof to another point p_1' in the $\mathcal{D}$-curve connecting p_0 and p_1 such that $|p_0 - p_1'| \simeq \frac{1}{2}|p_0 - p_1|$. By doing so, we can enlarge b_1 a little bit so that there is a unique κ_1^- satisfying (4.23). By (4.20) and (4.23), one easily verifies that $\kappa_1^- < a_1$, provided δ_0 is small. Therefore by (4.19), there exists a double point $p_2 \in \mathcal{L}$, which is on a $\mathcal{D}$-curve from either p_2^+ or p_2^- to p_0 (by Lemma 3.2), and a transport segment ℓ_2^+ with upper endpoint p_2 (ℓ_2^+ not in $\mathcal{T}_{q_1;r_1^+,r_1^-}$) such that

$$\theta[\ell_2^+, r_1^+] \geq N\theta[r_1^+, r_1^-], \tag{4.24}$$

for some N as large as we want, provided $\theta[r_1^+, r_1^-]$ is sufficiently small. Since p_2 is a double point, there is another transport segment ℓ_2^- with upper endpoint p_2, which lies in $\mathcal{T}_{q_1;r_1^+,r_1^-}$. Hence (4.24) implies

$$\theta[\ell_2^+, \ell_2^-] \geq N\theta[r_1^+, r_1^-]. \tag{4.25}$$

Repeating the above procedure, we obtain a sequence of points p_k (all of them lie on the line $\mathcal{L}$) and transport segments ℓ_k^+, ℓ_k^- with upper endpoints p_k, and a sequence of points q_k (all of them lie on the line $\mathcal{L}^*$) and transport segments r_k^+, r_k^- with lower endpoints q_k, such that

$$\begin{aligned} &\theta[r_k^+, r_k^-] \geq N\theta[\ell_k^+, \ell_k^-], \\ &\theta[\ell_{k+1}^+, \ell_{k+1}^-] \geq N\theta[r_k^+, r_k^-], \end{aligned} \tag{4.26}$$

provided the quantities on the right hand side are sufficiently small, where N can be as large as we want. Moreover, $\ell_k^+, \ell_k^-, r_k^+, r_k^- \subset \mathcal{T}_{p_0;\ell^+,\ell^-}$ for all $k \geq 1$.

From the above argument, the segments $\ell_k^+, \ell_k^-, r_k^+, r_k^-$ go across the region bounded by $\mathcal{L}$ and $\mathcal{L}^*$, so their lengths are uniformly bounded from below by a positive number. It is easy to verify that there is a constant $C > 0$ depending only on a, a_1, b_1 and $\operatorname{diam}(\Omega^*)$

such that

$$\begin{aligned} \theta[\ell_k^+, \ell_k^-] &\leq C\theta[\ell^+, \ell^-], \\ \theta[r_k^+, r_k^-] &\leq C\theta[\ell^+, \ell^-]. \end{aligned}$$

Note that a, a_1, b_1 are in fact determined by $c_0, d(p_0, \mathcal{H})$, and densities f, g. Hence by Lemma 4.4, there is a positive constant c_1 depending on $c_0, d(p_0, \mathcal{H}), \mathrm{diam}(\Omega^*)$, f and g, such that if $\theta[\ell^+, \ell^-] < c_1$ then (4.26) holds for a large, fixed constant $N > 1$ (independent of k). Therefore for some large k^*, we have a pair of transport segments $(\ell_{k^*}^+, \ell_{k^*}^-)$ or $(r_{k^*}^+, r_{k^*}^-)$ such that either

$$\theta[\ell_{k^*}^+, \ell_{k^*}^-] > N\theta[\ell^+, \ell^-] \tag{4.27}$$

or

$$\theta[r_{k^*}^+, r_{k^*}^-] > N\theta[\ell^+, \ell^-], \tag{4.28}$$

where N is a constant as large as we want. But since the segments $\ell_{k^*}^+, \ell_{k^*}^-, r_{k^*}^+, r_{k^*}^-$ are contained in $\mathcal{T}_{p_0;\ell^+,\ell^-}$ and their lengths are greater than the distance between the lines $\mathcal{L}$ and $\mathcal{L}^*$, one easily sees that (4.27) and (4.28) are impossible. Hence the angle $\theta[\ell^+, \ell^-]$ cannot be too small, namely $\theta[\ell^+, \ell^-] > c_1$. This completes the proof. □

We can now obtain:

Proof of Lemma 1.2. Take a family of disjoint $\mathcal{D}$-curves in $\mathcal{T}$ of length larger than a given $\varepsilon > 0$. For each of these curves we can apply Lemma 4.5 to two points p_1, p_0 in the curve, with $|p_1 - p_0| \geq \varepsilon$, thus obtaining the existence of a triangle bounded by transport segments, whose area is bounded from below by a given positive number and which contains the curve. These triangles are disjoint, because transport segments do not meet internally. Hence there can only be a finite number of them. □

5. Proof of Theorem 1.1

In this section we denote by s the monotone optimal mapping which sends the density ϱ to the density ϱ^* (and not any more f to g). Note that, in order to prove the continuity of s, we need to choose a precise representative of s. We will see in a while how to define s on $\mathcal{T}$. As for points of Z_0 and Z_1 (see Section 2 for the definitions of these sets), we define s to be the identity on them.

We first prove some of the properties which have been announced in Section 2.

Lemma 5.1. *The set $Z_0 \cup Z_1$ is closed.*

Proof. First, suppose that a sequence of points $x_n \in Z_0 \cup Z_1$ converges to a point outside $Z_0 \cup Z_1$, i.e. a point $x \in \mathcal{T}$. First we suppose that all the points x_n belong to Z_0. From $x_n \in \mathcal{H}$ we deduce $x \in \mathcal{H}$; the point x is in the middle of a non-degenerate transport segment ℓ, which meets $\mathcal{S}_f^o$. We can also suppose that the points x_n stay on a same side of ℓ. We can take a point $x' \in \mathcal{S}_f^o$ on ℓ with $|x' - x| \geq \varepsilon > 0$ and, taking a sequence of points $x'_k \to x'$ staying on the same side and the corresponding transport segments they belong to, we have a sequence of transport segments approaching ℓ, and a subsequence x_{n_k} of the points x_n which are located between them. But the condition $x_{n_k} \in Z_0$ implies that there exist $\mathcal{D}$-curves connecting them to $\partial X \cap \partial\Omega$ and staying between the same transport segments. These curves are disjoint and have length larger than ε, which is a contradiction to Lemma 1.2.

We now suppose that all the points x_n belong to Z_1. They belong to some transport segments ℓ_n on which $\varrho = \varrho^*$. The limit of the segments ℓ_n (which could be a singleton) is contained in a transport segment ℓ. If ℓ is a singleton, then $x \in Z_0$ which contradicts $x \notin Z_0 \cup Z_1$. If $\varrho = \varrho^*$ on ℓ, then $x \in Z_1$, which is also a contradiction. If ℓ meets the set $\mathcal{S}_f^o$, then this means that the uppr endpoints p_n of ℓ_n converge to a point $p \in \ell$ which is not an endpoint, and the part of ℓ where $f > 0$ is between p and the upper endpoint of ℓ. If we denote by ε the distance of p to the upper endpoint of ℓ, then we are in a situation as the one above, with a sequence of disjoint $\mathcal{D}$-curves with length at least ε, and we find again a contradiction to Lemma 1.2. $\square$

Once we know that $Z_0 \cup Z_1$ is closed, we now consider the set $\mathcal{T}$.

Let ℓ be a transport segment in $\mathcal{T}$. We consider a neighbourhood of a point in the relative interior of ℓ. Choose a coordinate system such that $\ell = \{(x_1, x_2) \mid x_2 = 0, -h \leq x_1 \leq h\}$ for some constant $h > 0$. For small $\delta > 0$ and $t \in (-\delta, \delta)$, let ℓ_t denote the transport segment whose intersection with x_2-axis is the point $(0, t)$. Denote

$$T_\delta = \{(x_1, x_2) \in \ell_t \mid -\delta < t < \delta\},$$

which is the area occupied by the transport segments ℓ_t with $|t| < \delta$. In our coordinates, ℓ_t can be expressed as

$$\ell_t = \big\{(x_1, x_2) \in \mathbb{R}^2 \mid x_2 = \eta(t)x_1 + t,\ x_1 \in [a(t), b(t)]\big\},$$

where a and b are respectively the x_1-coordinate values of the upper and lower endpoints of ℓ_t, and η is the slope of ℓ_t.

Lemma 5.2. *For sufficiently small $\delta > 0$, we have $a, b \in C^0(-\delta, \delta)$.*

Proof. We just need to prove the continuity at $t = 0$. By Theorem 2.3 (i) we know that a is upper semi-continuous, namely $\underline{\lim}_{t\to 0} a(t) \geq a(0)$. Assume to the contrary that $\overline{\lim}_{t\to 0} a(t) = a(0) + \varepsilon$ for some $\varepsilon > 0$. Then by the convexity of Ω, there is a sequence $t_k \to 0$ such that the upper endpoints p_k of ℓ_{t_k} are interior points of Ω and we have $p_k \to p \in \ell_0$.

By Lemma 3.3, there is a $\mathcal{D}$-curve connecting p_k to $\partial X \cap \partial\Omega$. This in particular means that, possibly reducing the value of ε, we can also suppose that the points p_k stay far away from the boundary $\mathcal{H}$ (if not, just follow back this $\mathcal{D}$-curve and choose another double point).

Moreover, for every k there is also a double point p'_k with two transport segments $(\ell')^+_k$ and $(\ell')^-_k$ and such that $p_k \in X_{p'_k;(\ell')^+,(\ell')^-}$ and the length $|p_k - p'_k| \geq \varepsilon$. In order to have $p_k \to p \in \ell_0$ there should be infinitely many of these triangles $X_{p'_k;(\ell')^+,(\ell')^-}$, pairwise disjoint. But this means an infinity of disjoint $\mathcal{D}$-curves with length larger than ε, which is a contradiction to Lemma 1.2.

Hence a is continuous. Similarly one can prove that b is continuous once we note that also in this case the length of the $\mathcal{D}$-curves arriving at the lower endpoints of ℓ_{t_k} would be bounded from below. Indeed, if this was not the case, there would be non-oriented double points arbitrarily close to the relative interior of the transport segment ℓ_0, which is impossible. □

Lemma 5.3. *We have $\eta \in C^1(-\delta, \delta)$.*

Proof. First we show $\eta \in C^{0,1}(-\delta, \delta)$. By Lemma 5.2, the length $|\ell_t|$ satisfies $|\ell_t| = 2h + o(1)$ as $t \to 0$. As the transport segments do not intersect with each other at interior points, we see that $|\eta(t) - \eta(t')| \leq |t - t'|/h + o(|t - t'|)$. Hence η is Lipschitz continuous.

Denote

$$T_{t,\varepsilon} = \{(x_1, x_2) \in \ell_{t'} \mid t - \varepsilon < t' < t + \varepsilon\},$$

and denote a point on ℓ_t by $p_{t,r} = (r, \eta(t)r + t)$. The optimal mapping s sends the point $p_{t,r}$ to $p_{t,s_t(r)}$, where $s_t(r)$ is determined by [6, 27]

$$\lim_{\varepsilon\to 0} \frac{1}{\varepsilon} \int_{T_{t,\varepsilon}\cap\{x_1<r\}} \varrho(x_1, x_2)dx_1dx_2 = \lim_{\varepsilon\to 0} \frac{1}{\varepsilon} \int_{T_{t,\varepsilon}\cap\{x_1<s_t(r)\}} \varrho^*(x_1, x_2)dx_1dx_2. \tag{5.1}$$

That is

$$\int_{a(t)}^{r} \big(\eta'(t)x_1 + 1\big)\varrho dx_1 = \int_{a(t)}^{s_t(r)} \big(\eta'(t)x_1 + 1\big)\varrho^* dx_1, \tag{5.2}$$

where ϱ, ϱ^* take values at $(x_1, \eta(t)x_1 + t)$. Letting $r = b(t)$, we obtain

$$\eta'(t) = \frac{\int_{a(t)}^{b(t)} (\varrho^* - \varrho) dx_1}{\int_{a(t)}^{b(t)} x_1(\varrho - \varrho^*) dx_1}. \tag{5.3}$$

Note that the denominator in (5.3) cannot be zero. For, if it is, then the numerator must be zero as well (since we know that η is Lipschitz continuous), and it implies

$$\int_{a(t)}^{b(t)} (x_1 - c)\varrho dx_1 = \int_{a(t)}^{b(t)} (x_1 - c)\varrho^* dx_1 \tag{5.4}$$

for any constant c. But (5.4) cannot hold if we choose c such that $\varrho \geq \varrho^*$ when $x_1 < c$, while $\varrho \leq \varrho^*$ when $x_1 \geq c$. Here, again, we use our geometric assumption on $\{f > 0\} = \{\varrho > \varrho^*\}$ and $\{g > 0\} = \{\varrho < \varrho^*\}$. Since $a(t), b(t) \in C^0$, we deduce from (5.3) that $\eta' \in C^0(-\delta, \delta)$. □

The formula (5.2) and the considerations above allow to deduce the following fact: on each transport segment ℓ, the mapping s is defined by taking the 1D monotone increasing map which transports ϱJ onto $\varrho^* J$, where J is an affine function of the form $x_1 \mapsto \eta'(t)x_1 + 1$. This will be the precise representative of s that we choose. Note in particular that s is the identity on endpoints.

On this kind of transport maps, we need to prove the following lemma.

Lemma 5.4. *Suppose we have a sequence of intervals $[a_n, b_n] \subset \mathbb{R}$, two densities ϱ_n and ϱ_n^* on these intervals, which are equicontinuous and bounded from below and above by uniform constants, a sequence of non-negative affine functions $J_n : [a_n, b_n] \to \mathbb{R}$, and the corresponding monotone transport maps $s_n : [a_n, b_n] \to [a_n, b_n]$ transporting $J_n\varrho_n$ onto $J_n\varrho_n^*$. Then we have*

1. *for any sequence x_n with $|x_n - a_n| \to 0$, we have $|s_n(x_n) - x_n| \to 0$*
2. *if $||\varrho_n - \varrho_n^*||_\infty \to 0$ then $||s_n - id||_\infty \to 0$*

Proof. Both part of the statements can be proven by contradiction. Up to normalizing by a multiplicative factor, we can assume $\max_{[a_n,b_n]} J_n = 1$ for every n and, up to translations, we can suppose $a_n = 0$ for every n. Up to subsequences, we can suppose, using the Ascoli-Arzelà theorem, that $b_n \to b$ and that ϱ_n and ϱ_n^* uniformly converge to some densities ϱ and ϱ^*, as well as the affine factors J_n converge to a limit affine function J (which will also have maximal value equal to 1 and hence be strictly positive on $]0, b[$)

For (1), suppose by contradiction $|s_n(x_n) - x_n| \geq \varepsilon$, together with $x_n \to 0$ (this is a consequence of the assumption $|x_n - a_n| \to 0$) and $s_n(x_n) \to \bar{s}$, which implies $\bar{s} \geq \varepsilon$. We have

$$\int_0^{x_n} \varrho_n J_n = \int_0^{s_n(x_n)} \varrho_n^* J_n. \tag{5.5}$$

Passing to the limit, we have $\int_0^{x_n} \varrho_n J_n \to 0$ because of $x_n \to 0$ and, using pointwise dominated convergence, we have

$$\int_0^{s_n(x_n)} \varrho_n^* J_n \to \int_0^{\bar{s}} \varrho^* J.$$

The lower bounds on ϱ^* (a consequence of the lower bounds on ϱ_n^*) and the strict positivity of J imply $\bar{s} = 0$, which is a contradiction.

For (2) suppose by contradiction that there exists a sequence x_n with $|s_n(x_n) - x_n| \geq \varepsilon$. Passing again to the limit in (5.5) we have

$$\int_0^{x} \varrho J = \int_0^{\bar{s}} \varrho^* J.$$

Yet, in this case we have $\varrho = \varrho^* > 0$, and this implies $\bar{s} = x$, which provides again a contradiction. □

We are now able to prove our main result.

Proof of Theorem 1.1. First, we prove the continuity of s at a point $x \in \mathcal{T} \setminus E$. This point is in the interior of $\mathcal{T}$ thanks to Lemma 5.1 and to Lemma 5.2. Then, we consider T_δ to be the set defined at the beginning of this section. Once we know the continuity of a, b and η', the continuity of s in T_δ, and hence at x, follows from (5.2).

We now need to prove continuity at points of Z_0, at points of Z_1 and at endpoints. In all these points the map s is the identity. For Z_0 we already noticed in Section 2 that the continuity is easy: points of Z_0 can only be approximated by segments whose length tends to 0 (because of point (i) in Lemma 2.3), so that if $x_n \to x \in Z_0$ we necessarily have $|s(x_n) - x_n| \to 0$, and s is continuous on Z_0.

For points of Z_1, we use part (2) of Lemma 5.4: indeed, if $x_n \to x \in Z_1$ and ℓ_n are the transport segments to which belong the points x_n, we necessarily have $\varrho = \varrho^*$ on the transport segment of x, and $||\varrho - \varrho^*||_{L^\infty(\ell_n)} \to 0$. This implies $|s(x_n) - x_n| \to 0$ and proves the continuity of s at x.

For endpoints, we use part (1) of Lemma 5.4. Consider a sequence $x_n \to x$, where x is an upper endpoint of a transport segment. If a_n denotes the upper endpoint of the transport segment of x_n, we must have $|x_n - a_n| \to 0$ (otherwise x cannot be an endpoint).

Then, part (1) of Lemma 5.4 implies $|s(x_n) - x_n| \to 0$ and proves the continuity of s at x. The proof for lower endpoints is analogous.

Therefore Theorem 1.1 is proved. □

Conflict of Interest: The authors declare that they have no conflict of interest.

References

[1] Ambrosio, L.: Lecture notes on optimal transport problems, Mathematical Aspects of Evolving Interfaces, Lecture Notes in Math. **1812** (2003), pp.1–52.

[2] Ambrosio, L.; Kirchheim, B.; Pratelli, A.: Existence of optimal transport maps for crystalline norms, Duke Math. J. **125** (2004), 207–241.

[3] Brenier, Y.: Polar factorization and monotone rearrangement of vector-valued functions, Comm. Pure Appl. Math. **44** (1991), 375–417.

[4] Caffarelli, L.: The regularity of mappings with a convex potential, J. Amer. Math. Soc. **5** (1992), 99–104.

[5] Caffarelli, L.: Boundary regularity of maps with convex potentials II, Ann. of Math. **144** (1996), 453–496.

[6] Caffarelli, L.; Feldman, M.; McCann, R.: Constructing optimal maps for Monge's transport problem as a limit of strictly convex costs, J. Amer. Math. Soc. **15** (2002), 1–26.

[7] Colombo, M.; Indrei, E.: Obstructions to regularity in the classical Monge problem, Math. Res. Lett. **21** (2014), 697–712.

[8] Champion, T.; De Pascale, L.: The Monge problem for strictly convex norms in $\mathbf{R}^d$, J. Eur. Math. Soc. **12** (2010), 1355–1369.

[9] Champion, T.; De Pascale, L.: The Monge problem in $\mathbf{R}^d$, Duke Math. J. **157** (2011), 551–572.

[10] De Pascale, L.; Louet, J.; Santambrogio, F.: The Monge problem with vanishing gradient penalization: vortices and asymptotical profile, J. Math. Pures and Appl., **106** (2016), no. 2, 237–279.

[11] Evans, L.C.; Gangbo, W.: Differential equations methods for the Monge-Kantorovich mass transfer problem, Mem. Amer. Math. Soc. **137** (1999), no. 653, viii+66 pp.

[12] Feldman, M.; McCann, R.: Uniqueness and transport density in Monge's mass transportation problem, Calc. Var. PDE **15**(2002), 81–113.

[13] Figalli, A.; Kim, Y.-M.; McCann, R.: Hölder continuity and injectivity of optimal maps, Arch. Ration. Mech. Anal. **209** (2013), 747–795.

[14] Fragalà, I.; Gelli, M.S.; Pratelli, A.: Continuity of an optimal transport in Monge problem, J. Math. Pures Appl. **84** (2005), 1261–1294.

[15] Kantorovich, L.: On the transfer of masses. Dokl. Acad. Nauk. USSR **37** (1942), 227–229.

[16] Kantorovich, L.: On a problem of Monge. Uspekhi Mat. Nauk. **3** (1948), 225–226.

[17] Li, Q.-R.; Santambrogio, F.; Wang, X.-J.: Regularity in Monge's mass transfer problem. J. Math. Pures Appl. (9) **102** (2014), no. 6, 1015–1040.

[18] Liu, J.K.; Trudinger, N.S.; Wang, X.-J.: Interior $C^{2,\alpha}$ regularity for potential functions in optimal transportation, Comm. Partial Differential Equations **35** (2010), 165–184.

[19] Liu, J.K.; Trudinger, N. S.; Wang, X.-J.: On asymptotic behaviour and $W^{2,p}$ regularity of potentials in optimal transportation, Arch. Ration. Mech. Anal. **215** (2015), 867–905.

[20] Loeper, G.: On the regularity of solutions of optimal transportation problems, Acta Math. **202** (2009), 241–283.

[21] Loeper, G.; Villani, C.: Regularity of optimal transport in curved geometry: the nonfocal case, Duke Math. J. **151** (2010), 431–485.
[22] Ma, X.N.; Trudinger, N.S.; Wang, X.-J.: Regularity of potential functions of the optimal transportation problem, Arch. Rational Mech. Anal. **177** (2005), 151–183.
[23] Monge, G.: Mémoire sur la Theorie des Déblais et des Remblais, In: Historie de l'Académie Royale des Sciences de Paris, avec les Mémories de Mathématique et de Physique pour la Même année, 1781, 666–740.
[24] Santambrogio F.: Optimal Transport for Applied Mathematicians, Progress in Nonlinear Differential Equations and Their Applications 87, Birkhäuser Basel (2015).
[25] Sudakov, V. N.: Geometric problems in the theory of infinite dimensional probability distributions, Proc. Steklov Inst. **141** (1979), 1–178.
[26] Villani, C.: Optimal transport, old and new, Springer-Verlag, Berlin, 2009.
[27] Trudinger, N.S.; Wang, X.-J.: On the Monge mass transfer problem, Calc. Var. PDE **13** (2001), 19–31.
[28] Trudinger, N.S.; Wang, X.-J.: On the second boundary value problem for Monge-Ampère type equations and optimal transportation, Ann. Sc. Norm. Super. Pisa Cl. Sci. (5), **8** (2009), 143–174.

Centre for Mathematics and Its Applications, Australian National University, Canberra, ACT 0200, Australia.

E-mail address: qi-rui.li@anu.edu.au

Laboratoire de Mathématiques d'Orsay, Univ. Paris-Sud, CNRS, Université Paris-Saclay, 91405 Orsay cedex, France

E-mail address: filippo.santambrogio@math.u-psud.fr

Centre for Mathematics and Its Applications, Australian National University, Canberra, ACT 0200, Australia.

E-mail address: xu-jia.wang@anu.edu.au