Calc. Var. 15, 81–113 (2002)
Calculus of Variations
DOI (Digital Object Identifier) 10.1007/s005260100119

Mikhail Feldman · Robert J. McCann

Uniqueness and transport density in Monge's mass transportation problem

Received: 18 December 2000 / Accepted: 11 May 2001 /
Published online: 19 October 2001 – © Springer-Verlag 2002

Abstract. Monge's problem refers to the classical problem of optimally transporting mass: given Borel probability measures $\mu^+ \neq \mu^-$ on $\mathbf{R}^n$, find the measure preserving map $s(x)$ between them which minimizes the average distance transported. Here distance can be induced by the Euclidean norm, or any other uniformly convex and smooth norm $d(x,y) = \|x-y\|$ on $\mathbf{R}^n$. Although the solution is never unique, we give a geometrical monotonicity condition singling out a particular optimal map $s(x)$. Furthermore, a local definition is given for the transport cost density associated to each optimal map. All optimal maps are then shown to lead to the same transport density $a \in L^1(\mathbf{R}^n)$.

Mathematics Subject Classification (2000): 49Q20, 28A50

1 Introduction

Let us begin by recalling a modern formulation of Monge's problem in $\mathbf{R}^n$ [6]. First published in 1781, we refer to Evans [8], Rachev and Rüschendorf [23], and Villani [29], for discussions of the problem, its history, and applications.

Problem 1.1 (Monge) Fix a norm $d(x,y) = \|x-y\|$ on $\mathbf{R}^n$, and two compactly supported densities — non-negative Borel functions f^+, $f^- \in L^1(\mathbf{R}^n)$ — satisfying the mass balance condition

$$\int_{\mathbf{R}^n} f^+(x)\,dx = \int_{\mathbf{R}^n} f^-(y)\,dy. \tag{1}$$

M. Feldman: Department of Mathematics, University of Wisconsin, Madison, Wisconsin USA 53706 (e-mail: feldman@math.wisc.edu)

R.J. McCann: Department of Mathematics, University of Toronto, Toronto, Ontario, Canada M5S 3G3 (e-mail: mccann@math.toronto.edu)

The authors gratefully acknowledge the support of grants DMS 0096090 [MF] and 0074037 [MF and RJM] of the U.S. National Science Foundation, and grant 217006-99 [RJM] of the Natural Sciences and Engineering Research Council of Canada. RJM is also pleased to acknowledge the hospitality of the Georgia Institute of Technology, where his work on this project concluded.

In the set $\mathcal{S}(f^+, f^-)$ of Borel maps $r : \mathbf{R}^n \to \mathbf{R}^n$ which push the measure $d\mu^+ = f^+(x)dx$ forward to $d\mu^- = f^-(y)dy$, find a map s which minimizes the cost functional

$$I[r] = \int_{\mathbf{R}^n} \|r(x) - x\| f^+(x)dx. \tag{2}$$

Here $r \in \mathcal{S}(f^+, f^-)$ is sometimes denoted by $r_{\#}\mu^+ = \mu^-$, and means merely that

$$\int_{\mathbf{R}^n} \phi(r(x))f^+(x)dx = \int_{\mathbf{R}^n} \phi(y)f^-(y)dy, \tag{3}$$

holds for each continuous test function ϕ on $\mathbf{R}^n$.

Though the norm $\|x - y\|$ need not be Euclidean, throughout the present manuscript we follow [6] in assuming there exist constants $\Lambda, \lambda > 0$ such that all $x, y \in \mathbf{R}^n$ satisfy the uniform smoothness and convexity estimates:

$$\lambda \|y\|^2 \leq \frac{1}{2}\|x+y\|^2 - \|x\|^2 + \frac{1}{2}\|x-y\|^2 \leq \Lambda \|y\|^2. \tag{4}$$

The estimates (4) assert some uniform convexity and smoothness [3] of the unit ball; they are certainly satisfied if, e.g., the unit sphere $\|x\| = 1$ is a C^2 surface in $\mathbf{R}^n$ with positive principal curvatures. In particular, $\Lambda = \lambda = 1$ makes (4) an identity in the Euclidean case.

Monge's problem has been studied by many authors. In 1976 Sudakov showed solutions to be realized in the original sense of Monge, i.e., as mappings from $\mathbf{R}^n$ to $\mathbf{R}^n$ [28]. A second proof of this existence result formed the subject of a recent monograph by Evans and Gangbo [9], who avoided Sudakov's measure decomposition results by using a partial differential equations approach. Recently a simpler, geometric proof was obtained by Caffarelli, Feldman, McCann [6] and independently by Trudinger and Wang [30] (for the case of the Euclidean norm).

The optimal map for Problem 1.1 is non-unique. In the one-dimensional case, multiple optimal maps can be constructed explicitly for fixed $\mu^+ \neq \mu^-$ on $\mathbf{R}$. We show in this paper that this one-dimensional phenomenon is the only source of non-uniqueness in Monge's problem. We do this by studying the uniqueness of optimal maps and of the flow generated by optimal maps. Our first result is the following:

Theorem 1.2 (Uniqueness of optimal maps) *Fix a norm on $\mathbf{R}^n$ satisfying the uniform smoothness and convexity conditions (4), and two $L^1(\mathbf{R}^n)$ densities $f^+, f^- \geq 0$ with compact support and the same total mass (1). Among Borel maps $s : \mathbf{R}^n \longrightarrow \mathbf{R}^n$ solving Monge's problem, in the sense that they minimize the average distance (2) transported among all maps pushing f^+ forward to f^- (3), there exists a unique optimal map $s \in \mathcal{S}(f^+, f^-)$ satisfying the monotonicity condition*

$$\frac{x_1 - x_2}{\|x_1 - x_2\|} + \frac{s(x_1) - s(x_2)}{\|s(x_1) - s(x_2)\|} \neq 0 \tag{5}$$

for all $x_1 \neq x_2 \in \mathbf{R}^n$ with distinct images $s(x_1) \neq s(x_2)$.

The existence of optimal maps has been shown using various methods, as mentioned above [28] [9] [30] [6]. In fact, all of these approaches lead to (or can be adapted to yield) a map satisfying (5); compare e.g. Trudinger-Wang [30, (18)] with Lemma 3.2 below. Thus the content of Theorem 1.2 is the uniqueness assertion.

In the absence of the restriction (5), Problem 1.1 admits multiple solutions, as can be constructed explicitly in the one-dimensional case for any $f^+ \neq f^- \in L^1(\mathbf{R})$. Lemma 3.1 below shows condition (5) is in fact implied by optimality, unless all four points $x_1, x_2, s(x_1), s(x_2)$ lie on a single line. Thus the restriction does nothing except resolve this one-dimensional degeneracy by ensuring that whenever any pair of points and their images are collinear, then the map s acts monotonically (non-decreasingly) along this line. More precisely, the geometry of optimal mass transport is following. Fix a Kantorovich potential u (see Problem 2.1, Proposition 2.2 and Remark 2.3 below) of problem 1.1. Then u is a Lipschitz function. The potential u determines *transport rays*, i.e. maximal segments with joining supp(f^+) to supp(f^-), along which u decreases linearly with the maximum rate allowed by its Lipschitz constant. The optimal maps constructed in [28], [9], [30], and [6] act down transport rays of u. Indeed [6, Lemma 6] implies that *any* optimal map acts along transport rays of the fixed potential u. However, not every map is monotonic along rays, i.e., satisfies (5). Theorem 1.2 thus shows that condition of monotonicity along transport rays selects a unique optimal map.

Multiple optimal maps in Problem 1.1 are obtained as follows. Restricting measures $\mu^\pm = f^\pm dx$ onto each transport ray as in [6] or [30], one obtains a one-dimensional transportation problem on each ray, which admits nonunique optimal map. Optimal maps on different transport rays can be chosen more or less independently: one need only retain enough consistency to obtain a measurable map $s : \mathbf{R}^n \to \mathbf{R}^n$ by combining different maps on separate rays. Then the map s is optimal [6]. In Sect. 4 we examine what effect the choices of optimal maps on the separate transport rays may cause. We show that the rate of cost of optimal mass transfer through each point of the space does not depend on the particular choice of an optimal map. Note that, since direction of optimal mass transfer is uniquely defined at any point by Theorem 1.2, and cost of transportation per unit mass depends only on direction, the uniqueness of rate of cost at each point is equivalent to uniqueness of rate of mass flux through each point. The quantity which describe a rate of cost of optimal transfer through a point of space is called *transport cost density* (or transport density) and may be introduced heuristically as follows. For a fixed optimal map s, the transport cost density $a(z)$ at a point $z \in \mathbf{R}^n$ is

$$a(z) = \lim_{R \to 0+} \frac{\text{cost of transportation through } \mathcal{D}_R(z) \text{ of mass flow generated by } s}{|\mathcal{D}_R(z)|},$$

where $\mathcal{D}_R(z)$ is a certain domain around z, shrinking nicely to z as $R \to 0+$. See (43) and (40) below for the precise definition of $a(z)$ and $\mathcal{D}_R(z)$. We show that the limit in the definition of $a(z)$ exists at almost every point z of $\mathbf{R}^n$, and that resulting function a belongs to $L^1(\mathbf{R}^n)$ and satisfies, in a weak sense, an equation, which in the Euclidean case has the form

$$-\mathrm{div}\,(aDu) = f^+ - f^-. \tag{6}$$

The transport density and similar functions have been studied by several authors, mostly in relation to the equation (6). In 1952, Beckmann [4] proposed a variational problem for the flow density of minimal transportation cost, but did not relate this transport density to optimal maps. He derived equation (6) formally, by assuming "sectional smoothness" of the transport density. Note that such regularity of a does not generally hold even for smooth $f^{\pm}$. Several authors considered equations similar to (6) motivated by problems of flows through domains (Strang [27], Iri [17]), or by variational problems for vector fields in L^1 or L^∞ (Strang [26], Janfalk [18]), for measures (Bouchitte-Buttazzo-Seppecher [5]), and by variational evolution problems (Evans-Feldman-Gariepy [11], Feldman [13] [14]). Evans and Gangbo [9] considered the transport density as a nonnegative function supported within the collection of transport rays, and satisfying (6). They prove that when $f^{\pm}$ are Lipschitz and disjointly supported, a transport density exists and belongs to L^∞. They used a in their construction of an optimal map, and heuristically interpreted a as the density of flow generated by an optimal map. We make this last interpretation rigorous, and prove existence and uniqueness of a transport density $a \in L^1(\mathbf{R}^n)$ for compactly supported $f^{\pm} \in L^1(\mathbf{R}^n)$, while deducing some further properties of this transport density.

1.1 Epilog

These results were presented at an October 26–28, 2000 workshop on *Mass transport problems, shape optimization, and weak geometrical structures* of the Scuola Normale Superiori in Pisa, Italy, where an earlier version of this manuscript was released. At the same workshop, the authors learned of several notable parallel developments. First, lecture notes were released by Ambrosio [2], which contain an excellent summary of progress on Monge's problem, including an independent derivation for existence and uniqueness of a transport density $a(\,\cdot\,)$ given $f^{\pm} \in L^1(\mathbf{R}^n)$. These notes also highlight a gap in Sudakov's proof for existence of an optimal map. Although this gap can be filled in two dimensions (provided the norm has a strictly convex unit ball), a counterexample in $\mathbf{R}^3$ due to Alberti, Kircheim, and Preiss [1] shows one of his propositions fails in higher dimensions unless additional assumptions are made. Thus it would seem that Evans and Gangbo [9] contains the the first complete proof of existence for optimal maps between Lipschitz densities $f^{\pm}$ with disjoint support, while Caffarelli-Feldman-McCann [6] and Trudinger-Wang [30] contain the first complete proofs for more general $f^{\pm} \in L^1(\mathbf{R}^n)$. Note that all complete proofs require a Euclidean ball, or at least the uniform smoothness and convexity hypothesis (4), which Sudakov explicitly eschews [28, p 164].

In other developments, Stepanov [25] obtained results on differentiability of the transport density along transport rays, and its non-differentiability in orthogonal directions, while DePascale and Pratelli [7] obtained estimates on L^p summability of the transport density $a(\,\cdot\,)$ in terms of $\|f^{\pm}\|_p$ for $p \in [1, \infty]$, and a sharp lower bound on the dimension of the support of a for more singular probability measures $\mu^{\pm}$.

2 Background: existence of optimal maps in Monge's problem

In this section we give a brief survey of the theory of Monge-Kantorovich problem, following [6]. We omit most of proofs in this section, since they can be found in [6].

2.1 Dual problem

First we recall a problem formulated by Kantorovich [19] as a dual to Monge's problem. Let $\mathrm{Lip}_1(\mathbf{R}^n, \|\cdot\|)$ denote the set of functions on $\mathbf{R}^n$ which are Lipschitz continuous with Lipschitz constant no greater than one; i.e.

$$\mathrm{Lip}_1(\mathbf{R}^n, \|\cdot\|) = \left\{ u : \mathbf{R}^n \to \mathbf{R}^1 \mid |u(x) - u(y)| \leq \|x-y\| \text{ for any } x, y \in \mathbf{R}^n \right\}.$$

Problem 2.1 (Kantorovich) Maximize $\hat{K}[v]$ on $Lip_1(\mathbf{R}^n, \|\cdot\|)$, where

$$\hat{K}[v] = \int_{\mathbf{R}^n} v(d\mu^+ - d\mu^-).$$

Proposition 2.2 (Lipschitz maximizer) *Let non-negative Borel functions f^+, $f^- \in L^1(\mathbf{R}^n)$ have compact support $\mathcal{X}$, $\mathcal{Y} \subset \mathbf{R}^n$ and satisfy the mass balance condition (1). Let $\mu^\pm = f^\pm dx$. Then there exists $u \in Lip_1(\mathbf{R}^n, \|\cdot\|)$ which is a maximizing solution of Problem 2.1:*

$$\hat{K}[u] = \sup_{v \in Lip_1(\mathbf{R}^n, \|\cdot\|)} \hat{K}[v].$$

In addition,

$$\begin{aligned} u(x) &= \min_{y \in \mathcal{Y}} (u(y) + \|x-y\|) \quad \textit{for any } x \in \mathcal{X}; \\ u(y) &= \max_{x \in \mathcal{X}} (u(x) - \|x-y\|) \quad \textit{for any } y \in \mathcal{Y}. \end{aligned} \tag{7}$$

Remark 2.3 We call a solution of Problem 2.1 a *Kantorovich potential*.

The next lemma exhibits the connection between the primal and dual problems.

Lemma 2.4 (Duality) *Fix $u \in Lip_1(\mathbf{R}^n, \|\cdot\|)$ and let $s : \mathbf{R}^n \to \mathbf{R}^n$ be a mapping which pushes μ^+ forward to μ^-. If*

$$u(x) - u(s(x)) = \|x - s(x)\| \quad \textit{for } \mu^+ \textit{ a.e. } x \in \mathcal{X} \tag{8}$$

then:

i. u is a Kantorovich potential maximizing Problem 2.1.
ii. s is an optimal map in Problem 1.1.
iii. The infimum in Problem 1.1 is equal to the supremum in Problem 2.1.
iv. Every optimal map $\hat{s}$ and Kantorovich potential $\hat{u}$ also satisfy (8).

2.2 Transport rays and their geometry

Fix two measures μ^+ and μ^- defined by non-negative densities $f^+, f^- \in L^1(\mathbf{R}^n)$ satisfying the mass balance condition (1). Assume that μ^+ and μ^- have compact supports, denoted by $\mathcal{X}$ and $\mathcal{Y} \subset \mathbf{R}^n$ respectively.

Our starting point for constructing an optimal map is a solution $u \in \mathrm{Lip}_1(\mathbf{R}^n, \|\cdot\|)$ of the Kantorovich dual Problem 2.1 satisfying (7). Such a u exists by Proposition 2.2.

Since we want to investigate the geometrical implications of (8) for u, suppose $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ satisfy

$$u(x) - u(y) = \|x - y\|.$$

From the Lipschitz constraint

$$|u(z_1) - u(z_2)| \leq \|z_1 - z_2\| \quad \text{for any } \ z_1, z_2 \in \mathbf{R}^n, \tag{9}$$

it follows that on the segment connecting x and y the function u is affine and decreasing with the maximum rate compatible with (9). We will call maximal segments $[x, y]$ having these properties the *transport rays*. More precisely:

Definition 2.5 (Transport rays) *A* transport ray R *is a segment with endpoints* $a, b \in \mathbf{R}^n$ *such that*

i. $a \in \mathcal{X}$ *,* $b \in \mathcal{Y}$ *,* $a \neq b$*;*
ii. $u(a) - u(b) = \|a - b\|$*;*
iii. Maximality: for any $t > 0$ *such that* $a_t := a + t(a - b) \in \mathcal{X}$ *there holds*

$$|u(a_t) - u(b)| < \| a_t - b\|,$$

and for any $t > 0$ *such that* $b_t := b + t(b - a) \in \mathcal{Y}$ *there holds*

$$|u(b_t) - u(a)| < \| b_t - a\|.$$

We call the points a and b the *upper* and *lower ends* of R, respectively. Since $u(a) - u(b) = \|a - b\|$, it follows from (9) that any point $z \in R$ satisfies

$$u(z) = u(b) + \|z - b\| = u(a) - \|a - z\|. \tag{10}$$

Let us call a point $z \in \mathbf{R}^n$ an *interior point* of a segment $[a, b]$, where $a, b \in \mathbf{R}^n$, if $z = ta + (1 - t)b$ for some $0 < t < 1$. We denote by $[a, b]^0$ the set of interior points of $[a, b]$.

Definition 2.6 (Rays of length zero) *Denote by* T_1 *the set of all points which lie on transport rays. Define a complementary set* T_0*, called the* rays of length zero*, by*

$$T_0 = \{z \in \mathcal{X} \cap \mathcal{Y} \ : \ |u(z) - u(z')| < \|z - z'\| \ \textit{for any } z' \in \mathcal{X} \cup \mathcal{Y},\ z' \neq z\}.$$

We collect some basic properties of transport rays in the following lemma:

Lemma 2.7 (Properties of transport rays) *Let the norm $\|\cdot\|$ satisfy (4). Then:*

i. *Data is Supported Only on Transport Rays: $\mathcal{X} \cup \mathcal{Y} \subseteq T_0 \cup T_1$*
ii. *Transport Rays Are Disjoint: Let two transport rays $R_1 \neq R_2$ share a common point c. Then $R_1 \cap R_2 = \{c\}$ and c is either the upper end of both rays, or the lower end of both rays. In particular, an interior point of a transport ray does not lie on any other transport ray.*
iii. *Differentiability of Kantorovich Potential Along Rays: If z_0 lies in the relative interior of some transport ray R then u is differentiable at z_0. Indeed, setting $e := (a-b)/\|a-b\|$ where a, b are the upper and lower ends of R yields:*

$$|Du(z_0)y| \leq 1 \quad \text{for all } \|y\| = 1, \text{ with equality if and only if } y = \pm e.$$

Here $Du(x) \in (\mathbf{R}^n)^*$ is a derivative of u at $x \in \mathbf{R}^n$, viewed as a linear functional on the tangent space.

We will use the following distance functions to the lower and upper ends of rays:

Lemma 2.8 (Semicontinuity of distance to ray ends) *At each $z \in \mathbf{R}^n$ define*

$$\alpha(z) := \sup \{\|z-y\| \ | \ y \in \mathcal{Y}, \ u(z) - u(y) = \|z-y\|\}, \tag{11}$$

$$\beta(z) := \sup \{\|z-x\| \ | \ x \in \mathcal{X}, \ u(x) - u(z) = \|z-x\|\}, \tag{12}$$

where $\sup \emptyset := -\infty$. Then $\alpha, \beta : \mathbf{R}^n \to \mathbf{R} \cup \{-\infty\}$ are both upper semicontinuous.

Definition 2.9 (Ray directions) *Define a function $\nu : \mathbf{R}^n \to \mathbf{R}^n$ as follows. If z is an interior point of a transport ray R with upper and lower endpoints a, b (note that R is uniquely defined by z in view of Lemma 2.7(ii)) then*

$$\nu(z) := \frac{a-b}{\|a-b\|}. \tag{13}$$

Define $\nu(z) = 0$ for any point $z \in \mathbf{R}^n$ not the interior point of a transport ray. We call $\nu(z)$ the direction function *corresponding to the Kantorovich potential u.*

The next property is crucial for construction of optimal map.

Lemma 2.10 (Ray directions vary Lipschitz continuously) *Let R_1 and R_2 be transport rays, with upper end a_k and lower end b_k for $k = 1, 2$ respectively. If there are interior points $y_k \in (R_k)^0$ where both rays pierce the same level set of the Kantorovich potential, $u(y_1) = u(y_2)$, then the ray directions (13) satisfy a Lipschitz bound*

$$\|\nu(y_1) - \nu(y_2)\| \leq \frac{C}{\sigma} \|y_1 - y_2\|, \tag{14}$$

with constant $C^2 + \lambda = 2(1 + \lambda^{-1}\Lambda)/(1+\lambda)$ depending on the norm (4) and the distance $\sigma := \min\limits_{k=1,2} \{\|y_k - a_k\|, \|y_k - b_k\|\}$ to the ends of the rays.

2.3 Measure decomposing change of variables

It is in this subsection that we construct the change of variables on $\mathbf{R}^n$ which we use to build an optimal map. Lemma 2.10 suggests how these new coordinates must be defined: $n-1$ of the new variables are used to parameterize a given level set of the Kantorovich potential u, while the final coordinate x_n measures distance to this set along the transport rays which pierce it. Thus the effect of this change of variables will be to flatten level sets of u while making transport rays parallel. But the conditions of Lemma 2.10 make clear that we retain Lipschitz control only if we restrict our transformation to clusters of rays in which all rays intersect a given level set of u, and the intersections take place a uniform distance away from both endpoints of each ray. These observations motivate the construction to follow.

Lemma 2.11 (Bi-Lipschitz parametrization of level sets) *Let $u : \mathbf{R}^n \to \mathbf{R}^1$ be a Lipschitz function, $\sigma \in \mathbf{R}^1$, and S_σ the level set $\{x \in \mathbf{R}^n \mid u(x) = \sigma\}$. Then the set*

$$S_\sigma \cap \{x \in \mathbf{R}^n \mid u \textit{ is differentiable at } x \textit{ and } Du(x) \neq 0\}$$

has a countable covering consisting of Borel sets $S^i_\sigma \subset S_\sigma$, such that for each $i \in \mathbf{N}$ there exist Lipschitz coordinates $U : \mathbf{R}^n \to \mathbf{R}^{n-1}$ and $V : \mathbf{R}^{n-1} \to \mathbf{R}^n$ satisfying

$$V(U(x)) = x \quad \textit{for all} \quad x \in S^i_\sigma. \tag{15}$$

For each level $\sigma \in \mathbf{R}^1$ and integer $i \in \mathbf{N}$, we shall extend these coordinates to the transport rays intersecting S^i_p.

Definition 2.12 (Ray clusters) *Fix $\sigma \in \mathbf{R}^1$, a Kantorovich potential u, and the Borel cover $\{S^i_\sigma\}_i$ of the level set $S_\sigma := \{x \in \mathbf{R}^n \mid u(x) = p\}$ in Lemma 2.11. Let $i \in \mathbf{N}$ and let B be a Borel subset of S^i_σ. For each $j \in \mathbf{N}$ let the* cluster *$T_{\sigma ij}(B) := \cup R_z$ denote the union of all transport rays R_z which intersect B, and for which the point of intersection $z \in B$ is separated from both endpoints of the ray by distance greater than $1/j$ in $\|\cdot\|$. The same cluster, but with ray ends omitted, is denoted by $T^0_{\sigma ij}(B) := \cup_z(R^0_z)$. Denote $T_{\sigma ij} := T_{\sigma ij}(S^i_\sigma)$ and $T^0_{\sigma ij} := T^0_{\sigma ij}(S^i_\sigma)$.*

On each ray cluster $T^0_{\sigma ij}$ we define the Lipschitz change of variables:

Lemma 2.13 (Lipschitz change of variables) *Each ray cluster $T_{\sigma ij} \subset \mathbf{R}^n$ admits coordinates $G = G_{\sigma ij} : T^0_{\sigma ij} \to \mathbf{R}^{n-1} \times \mathbf{R}^1$ with inverse $F = F_{\sigma ij} : G(T^0_{\sigma ij}) \to \mathbf{R}^n$ satisfying:*

i. *F extends to a Lipschitz mapping between $\mathbf{R}^{n-1} \times \mathbf{R}^1$ and $\mathbf{R}^n$;*
ii. *for each $\lambda > 0$, G is Lipschitz on $T^\lambda_{\sigma ij} := \{x \in T^0_{\sigma ij} \mid \|x-a\|, \|x-b\| > \lambda\}$, where a and b denote the endpoints of the (unique) transport ray R_x;*
iii. *$F(G(x)) = x$ for each $x \in T^0_{\sigma ij}$;*
iv. *If a transport ray $R_z \subset T_{\sigma ij}$ intersects S^i_σ at z, then each interior point $x \in (R_z)^0$ of the ray satisfies*

$$G(x) = (U(z), u(x) - u(z)), \tag{16}$$

where $U : \mathbf{R}^n \to \mathbf{R}^{n-1}$ gives the Lipschitz coordinates (15) on S^i_σ.

Definition 2.14 (Ray ends) *Denote by $\mathcal{E} \subset T_1$ the set of endpoints of transport rays.*

The next step is to address measurability of the sets $T_{\sigma ij}(B)$ and $G[T^0_{\sigma ij}(B)]$. In what follows, n-dimensional Lebesgue measure is denoted $\mathcal{L}^n$.

Lemma 2.15 (Measurability of clusters / Negligibility of ray ends) *The ray ends $\mathcal{E} \subset T_1$ form a Borel set of measure zero: $\mathcal{L}^n[\mathcal{E}] = 0$. The rays of length zero $T_0 \subset \mathbf{R}^n$ also form a Borel set. Finally, for each $\sigma \in \mathbf{R}^1$, $i, j \in \mathbf{N}$, and Borel $B \subset S^i_\sigma$ the cluster $T^0_{\sigma ij}(B)$ of ray interiors and its flattened image $G[T^0_{\sigma ij}(B)]$ are Borel. Here G is the map from Lemma 2.13. In particular the sets $T^0_{\sigma ij}$ and $G[T^0_{\sigma ij}]$ are Borel.*

Remark 2.16 The statements corresponding to Lemmas 2.13 and 2.15 are formulated in [6] only for clusters T_{pij}, T^0_{pij} with $(p, i, j) \in \mathbf{Q} \times \mathbf{N}^2$. But the proofs work without any changes in the conditions of Lemmas 2.13 and 2.15.

A countable collection of clusters forms a covering of T^0_1:

Lemma 2.17 (Rational clusters cover rays) *The clusters T_{pij} indexed by $p \in \mathbf{Q}$ and $i, j \in \mathbf{N}$ define a countable covering of all transport rays $T_1 \subset \mathbf{R}^n$. Moreover, each T_{pij} and transport ray R satisfy:*

$$\text{Either } (R)^0 \subset T_{pij}, \quad \text{or } (R)^0 \cap T_{pij} = \emptyset. \tag{17}$$

As a particular consequence of this lemma: the set T_1 of all transport rays is Borel, being a countable union of Borel sets T^0_{pij} with $\mathcal{E}$. Also, the sets T_{pij} are Lebesgue measurable, being the union of a Borel set with a subset of a negligible set.

Finally, we can take the clusters T_{pij} of rays to be disjoint. Indeed, enumerate the triples (p, i, j) so the collection of clusters $\{T_{pij}\}$ becomes $\{T_{(k)}\}, k = 1, 2, \ldots$. For $k > 1$ redefine $T_{(k)} \to T_{(k)} \setminus (\cup_{l=1}^{k-1} T_{(l)})$. Redefine $T^0_{(k)} \to T^0_{(k)} \setminus (\cup_{l=1}^{k-1} T^0_{(l)})$ analogously. We will continue to denote the modified sets by T_{pij} and T^0_{pij}. Note that the structure of the clusters T_{pij} remains the same: for each T_{pij} we have a Borel subset $S_{pij} := T_{pij} \cap S_p \subset S^i_p$, and T_{pij} is the cluster $T_{pij}(S_{pij})$ from Definition 2.12. In particular, there are Lipschitz coordinates U, V (15) satisfying $V(U(x)) = x$ for all $x \in S_{pij}$, and maps F, G satisfying all assertions of Lemma 2.13. Indeed, since the new cluster is a subset of the old, the former maps U, V, F, G will suffice. The measurability Lemma 2.15 holds for the new clusters. Thus from now on we assume:

$$\textit{The clusters of ray interiors } T^0_{pij} \textit{ are disjoint.} \tag{18}$$

For future reference, let us point out that the above construction implies the following. Define the following mappings j, j$^\pm$ on subsets of level sets $S_\sigma = u^{-1}(\sigma)$, where $\sigma \in \mathbf{R}^1$: for $A \subset S_\sigma$

$$\begin{aligned} \mathrm{j}(A) &= \cup_{z \in A \cap T^0_1} R^0_z, \\ \mathrm{j}^+(A) &= \mathrm{j}(A) \cap \{y \mid u(y) \geq \sigma\}, \\ \mathrm{j}^-(A) &= \mathrm{j}(A) \cap \{y \mid u(y) < \sigma\} \end{aligned} \tag{19}$$

where R_z^0 is the relative interior of the unique transport ray through z. Thus, $\mathrm{j}(A)$ is the smallest transport set containing $A \cap T_1^0$ for $A \subset S_q$, and $\mathrm{j}^{\pm}(A)$ are the parts of $\mathrm{j}(A)$ which lie above (resp. below) the level set S_σ of u.

Corollary 2.18 *Let $\sigma \in \mathbf{R}^1$, let $S_\sigma := u^{-1}(\sigma)$ be the level set of $u(z)$, and let $B \subset S_\sigma$ be a Borel set. Then the sets $\mathrm{j}(B)$, $\mathrm{j}^{\pm}(B)$ are Borel.*

Proof. Since u is a continuous function, we only need to prove that $\mathrm{j}(B)$ is Borel.

Since $B \cap T_1^0$ is Borel, we can replace B by $B \cap T_1^0$, i.e., assume that $B \subset S_\sigma \cap T_1^0$.

We have

$$\mathrm{j}(B) = \bigcup_{i,j=1}^{\infty} T_{\sigma ij}^0 (B \cap S_\sigma^i).$$

Since both B and S_σ^i are Borel, we use Lemma 2.15 to conclude the proof. □

2.4 Detailed mass balance

Definition 2.19 (Transport sets) *A set $A \subset \mathbf{R}^n$ is called a* transport set *if $z \in A \cap (T_1 \setminus \mathcal{E})$ implies $R_z^0 \subseteq A$, where R_z is the unique transport ray passing through z. It is called the* positive end of a transport set *if A merely contains the interval $[z, a)$ whenever $z \in A \cap (T_1 \setminus \mathcal{E})$ and a denotes the upper end of the transport ray R_z.*

Examples. Any subset $A \subset T_0$ of rays of length zero is a transport set, as are the clusters of rays $T_{\sigma ij}$.

For Borel transport sets, such as T_{pij}^0, the following balance conditions apply.

Lemma 2.20 (Detailed mass balance) *Let $A \subset \mathbf{R}^n$ be a Borel transport set. Then*

$$\int_A f^+(x)\, dx = \int_A f^-(x)\, dx. \tag{20}$$

More generally, if a Borel set $A^+ \subset \mathbf{R}^n$ forms the positive end of a transport set, then

$$\int_{A^+} f^+(x)\, dx \geq \int_{A^+} f^-(x)\, dx. \tag{21}$$

2.5 Construction of the optimal map

In this subsection we construct an optimal map for Problem 1.1.

Step 1. Localization to clusters of rays. According to Lemma 2.4, it is enough to construct a map $s : \mathbf{R}^n \to \mathbf{R}^n$ pushing μ^+ forward to μ^- which only moves mass *down* transport rays: i.e., for any $x \in \mathcal{X}$, the point $s(x)$ must lie below x on the same transport ray R_x, possibly of length zero. Here 'down' and 'below' refer to the constraint $u(x) \geq u(s(x))$ from (8).

Decompose the set $\mathcal{X} \cup \mathcal{Y}$ into the rays T_0 of length zero, clusters of ray interiors T_{pij}^0, and the ray ends $\mathcal{E}$ using Lemmas 2.7(i) and 2.17. The cluster property (17) implies that any such map s satisfies $s(x) \in T^0_{pij}$ almost everywhere on T^0_{pij}, while $s(x) = x$ on T_0. Since the ray ends form a set of measure zero by Lemma 2.15, they are neglected here and in the sequel. Also, the clusters T^0_{pij} and T_0 are disjoint and Borel by (18) and Lemma 2.15. Thus we can construct an optimal map s separately on each cluster T^0_{pij} and on T_0.

Consider s_0 first. Since every subset $A \subset T_0$ is a transport set, Lemma 2.20 shows the identity map pushes $\mu^+_{|T_0}$ forward to $\mu^-_{|T_0}$. Thus we define $s_0(x) = x$ on T_0. The remainder of the subsection is devoted to constructing maps $s_{pij} : T^0_{pij} \to T^0_{pij}$ pushing $\mu^+_{|T^0_{pij}}$ forward to $\mu^-_{|T^0_{pij}}$ which only move mass down transport rays.

Step 2. Change of variables. Fix $p \in \mathbf{Q}$, $i, j \in \mathbf{N}$ and consider T^0_{pij}. Denote $\mu^\pm_{pij} := \mu^\pm_{|T^0_{pij}}$. By Lemma 2.13 the map F is one to one on $G(T^0_{pij})$, and $F(G(T^0_{pij})) = T^0_{pij}$. Since F is Lipschitz, the Area formula [12, §3.2.5] yields

$$\int_{G(T^0_{pij})} \varphi(F(x)) f^\pm(F(x)) J_n F(x)\, dx = \int_{T^0_{pij}} \varphi(z) f^\pm(z) dz \tag{22}$$

for any summable $\varphi : \mathbf{R}^n \to \mathbf{R}^1$. Here $J_n F$ denotes the n-dimensional Jacobian of F. Define $\hat{f}^\pm : \mathbf{R}^{n-1} \times \mathbf{R}^1 \to \mathbf{R}^1$ by

$$\hat{f}^\pm(x) = \begin{cases} f^\pm(F(x)) J_n F(x) & x \in G(T^0_{pij}); \\ 0 & \text{otherwise.} \end{cases} \tag{23}$$

The characteristic function $\varphi = \chi_{G(T^0_{pij})}$ in (22) shows $\hat{f}^\pm$ is summable; it is obviously non-negative and Borel since Lemma 2.15 shows $G(T^0_{pij})$ Borel and bounded. Introduce the measures $d\theta^\pm := \hat{f}^\pm(x) dx$. From (3), (22), (23) we see that

$$F_\# \theta^\pm = \mu^\pm_{pij}, \quad G_\# \mu^\pm_{pij} = \theta^\pm.$$

It then follows that if a map $\hat{s} : \mathbf{R}^{n-1} \times \mathbf{R}^1 \to \mathbf{R}^{n-1} \times \mathbf{R}^1$ pushes θ^+ forward to θ^-, then the composition $s_{pij} = F \circ \hat{s} \circ G$ pushes forward μ^+_{pij} to μ^-_{pij}. In addition, Lemma 2.13(iv) shows that when $\hat{s}$ moves mass down vertical lines, i.e., satisfies $\hat{s}(X, x_n) \in \{X\} \times [-\infty, x_n]$ for any (X, x_n), then s_{pij} moves mass down transport rays. Thus it remains only to construct $\hat{s} : \mathbf{R}^{n-1} \times \mathbf{R}^1 \to \mathbf{R}^{n-1} \times \mathbf{R}^1$ satisfying

$$\hat{s}_\# \theta^+ = \theta^-, \qquad \hat{s}(X, x_n) \in \{X\} \times [-\infty, x_n] \quad \text{for any } (X, x_n) \in \mathbf{R}^{n-1} \times \mathbf{R}^1.$$

Step 3. Restriction to vertical lines. By Fubini's theorem, the functions $\hat{f}^\pm(X, \cdot)$ are summable for a.e. $X \in \mathbf{R}^{n-1}$. Let us introduce the distribution function

$$\Psi^\pm(X, \tau) := \int_\tau^\infty \hat{f}^\pm(X, x_n)\, dx_n. \tag{24}$$

Lemma 2.21 *$\Psi^{\pm}(X,\tau)$ is non-negative and Borel function throughout $\mathbf{R}^{n-1} \times \mathbf{R}^1$, with a continuous non-increasing dependence on τ. In addition, for a.e. $X \in \mathbf{R}^{n-1}$*

$$\Psi^+(X,\tau) \geq \Psi^-(X,\tau) \tag{25}$$

holds for all $\tau \in \mathbf{R}$, with equality

$$\Psi^+(X,-\infty) = \Psi^-(X,-\infty) < \infty \tag{26}$$

as $\tau \to -\infty$.

Properties (25) and (26) are derived from Lemma 2.20.

Step 4. One-dimensional transport. Fix $X \in \mathbf{R}^{n-1}$ for which (25–26) hold. We construct a map $t_X(x) \leq x$ on $\mathbf{R}^1$ which pushes $\hat{f}^+(X,x_n)\,dx_n$ forward to $\hat{f}^-(X,x_n)\,dx_n$ as follows. Fix $\tau \in \mathbf{R}^1$, and recall that $\Psi^{\pm}(X,\,\cdot\,)$ is a continuous, non-increasing function which takes constant values outside a compact set. By (26), there exists some $\zeta \in \mathbf{R}^1$ which satisfies

$$\Psi^+(X,\tau) := \int_\tau^\infty \hat{f}^+(X,x_n)\,dx_n = \int_\zeta^\infty \hat{f}^-(X,x_n)\,dx_n =: \Psi^-(X,\zeta). \tag{27}$$

Of course ζ need not be unique, since $\Psi^-(X,\,\cdot\,)$ will not decrease strictly where $\hat{f}^-$ vanishes. Define

$$t_X(\tau) := \inf\ \{\zeta \in \mathbf{R}^1 \mid \Psi^+(X,\tau) \geq \Psi^-(X,\zeta)\} \tag{28}$$

$$= \sup\ \{\zeta \in \mathbf{R}^1 \mid \Psi^+(X,\tau) < \Psi^-(X,\zeta)\}. \tag{29}$$

Step 5. Construction of optimal map. Define $\hat{s} : \mathbf{R}^{n-1} \times \mathbf{R}^1 \to \mathbf{R}^{n-1} \times \mathbf{R}^1$ as $\hat{s}(X,x_n) = (X, t_X(x_n))$, where $t_X(x_n) \leq x_n$ is from Step 4.

The map $\hat{s}$ is Borel throughout $\mathbf{R}^{n-1} \times \mathbf{R}^1$, and $\hat{s}_\#\theta^+ = \theta^-$. By Step 2 this yields maps $s_{pij} = F \circ \hat{s} \circ G$ on each cluster T^0_{pij} which push $\mu^+_{|T^0_{pij}}$ forward to $\mu^-_{|T^0_{pij}}$ while only moving mass down transport rays. Step 1 combines these maps to yield an optimal map $s : \mathbf{R}^n \to \mathbf{R}^n$ for Problem 1.1.

3 Uniqueness of a raywise monotone optimal map

This section is devoted to establishing the uniqueness of optimal maps under the monotonicity condition (5). It begins with a series of lemmas culminating in the proof of Theorem 1.2. The first lemma asserts any optimal map satisfies condition (5), except possibly along transport rays. Thus the failure of uniqueness in Monge's problem is essentially one dimensional: it is due to indeterminacy along lines.

Lemma 3.1 (Inversions occur only on lines) *Any optimal solution $s \in \mathcal{S}(f^+, f^-)$ to Monge's problem satisfies (after modification on a set of f^+ measure zero) the following property: if $x_1 \neq x_2 \in \mathbf{R}^n$ and $s(x_1) \neq s(x_2)$ are related by*

$$\frac{x_1 - x_2}{\|x_1 - x_2\|} + \frac{s(x_1) - s(x_2)}{\|s(x_1) - s(x_2)\|} = 0, \tag{30}$$

then all four points $x_1, x_2, s(x_1), s(x_2)$ are collinear.

Proof. After modification on a set of f^+ measure zero, any optimal solution $s \in \mathcal{S}(f^+, f^-)$ satisfies Monge's two-point inequality

$$\|x_1 - s(x_2)\| + \|x_2 - s(x_1)\| \geq \|x_1 - s(x_1)\| + \|x_2 - s(x_2)\|; \tag{31}$$

a modern proof may be found e.g., in [16, Theorem 2.3].

Now assume $x_1 \neq x_2$ and $s(x_1) \neq s(x_2)$ satisfy (30). The four points are coplanar because $x_1 - x_2$ parallels $s(x_2) - s(x_1)$ according to (30). Thus the quadrilateral $x_1, x_2, s(x_1), s(x_2)$ is a (convex) trapezoid. Denoting the point where its diagonals intersect by m, we have $\|x_i - s(x_i)\| = \|x_i - m\| + \|m - s(x_i)\|$ for $i = 1, 2$. Furthermore, the triangle inequality yields

$$\|x_1 - s(x_2)\| \leq \|x_1 - m\| + \|m - s(x_2)\| \tag{32}$$

$$\|x_2 - s(x_1)\| \leq \|m - s(x_1)\| + \|x_2 - m\|. \tag{33}$$

Strict convexity of the norm (4) implies the first inequality to be strict unless $s(x_2)$ lies on the diagonal line through x_1, m and $s(x_1)$. Similarly x_2 must lie on the same diagonal or the second inequality will be strict. Either all five points $x_1, x_2, m, s(x_1), s(x_2)$ are collinear, or the sum of (32–33) violates (31) – a contradiction. □

For two distinct functions $0 \leq \hat{f}^{\pm} \in L^1(\mathbf{R})$ with the same total mass on a given transport ray, many maps in $\mathcal{S}(\hat{f}^+, \hat{f}^-)$ verify (30) at some point $x_1 \neq x_2$. We claim only one measure-preserving map satisfies the opposite condition (5). We begin by demonstrating that the map constructed in Sect. 2 above in particular verifies (5).

Lemma 3.2 (Monotonicity along rays by construction) *The map s constructed in Sect. 2.5 satisfies the monotonicity condition (5).*

Proof. Let u be a Kantorovich potential satisfying (7), and s be the optimal map constructed in Sect. 2.5. This map s acts along transport rays of u, i.e., satisfies (8). Also since $\mu^+[\mathbf{R}^n \setminus \mathcal{X}] = 0$ and $|\mathcal{E}| = 0$, where $\mathcal{E}$ is the set of endpoints of transport rays (Definition 2.14 and Lemma 2.15) we can set $s(x) = x$ for all $x \in \mathbf{R}^n \setminus \mathcal{X}$ and $x \in \mathcal{E}$. Since $s(x) = x$ on the set T_0 of rays of length zero, we now have $s(x) = x$ on $\mathbf{R}^n \setminus (\mathcal{X} \cap T_1^0)$.

In order to show that s satisfies (5), it is enough to consider the case when the four points $x_1 \neq x_2$ and $s(x_1) \neq s(x_2)$ are all collinear, according to Lemma 3.1. If $x_1 = s(x_1) \neq s(x_2) = x_2$ then condition (5) holds trivially because $2 \neq 0$. Thus we assume $x_1 \neq s(x_1)$ without loss of generality. Recall that $s(x) = x$ on $\mathbf{R}^n \setminus (\mathcal{X} \cap T_1^0)$, hence $x_1 \in \mathcal{X} \cap T_1^0$. Then x_1 lies in the relative interior R_1^0 of a transport ray R_1 with $s(x_1) \in R_1$ from (8).

If $x_2 \in R_1^0$, then (5) follows from monotonicity of s along transport rays, i.e., from the property $[u(x_1) - u(x_2)][u(s(x_1)) - u(s(x_2))] \geq 0$ for any $x_1, x_2 \in R_1^0$. This last property follows from the fact that the function t_X defined by (28) is nondecreasing, and from Lemma 2.13(iii-iv).

Thus it remains to consider $x_2 \notin R_1^0$.

If $x_2 = s(x_2)$, then equality (30) implies that x_2 lies in the relative interior of R_1, between x_1 and $s(x_1)$ — a contradiction.

The only other possibility is $x_2 \neq s(x_2)$ — in which case x_2 lies in the relative interior R_2^0 of a transport ray R_2. Since $x_2 \notin R_1^0$ we have $R_1^0 \cap R_2^0 \neq \emptyset$ from Lemma 2.7(ii). Furthermore, $s(x_2) \in R_2$ by (8). The disjoint transport rays R_1^0 and R_2^0 must be collinear, so $x_1 - x_2$ and $s(x_1) - s(x_2)$ both point in the same direction: away from R_1^0 and toward R_2^0. In this case (5) is again satisfied since $2 \neq 0$. Lemma 3.2 has therefore been proved. □

The next lemma and corollary pave the way to the proof of Theorem 1.2. They establish uniqueness of raywise monotone maps in the flattened coordinate system of Sect. 2.3.

Lemma 3.3 *Fix two compactly supported densities $0 \leq \theta^{\pm}(x) \in L^1(\mathbf{R}^n, d\mathcal{H}^n)$, satisfying mass balance (1). Write $x = (X, z) \in \mathbf{R}^{n-1} \times \mathbf{R}$ and $\theta^{\pm}_X(z) := \theta^{\pm}(X, z)$. If $s : \mathbf{R}^{n-1} \times \mathbf{R} \longrightarrow \mathbf{R}^{n-1} \times \mathbf{R}$ is a Borel map of the form $s(X, z) = (X, s_X(z))$ and $s \in S(\theta^+, \theta^-)$, then $s_X \in \mathcal{S}(\theta^+_X, \theta^-_X)$ holds for $\mathcal{H}^{n-1}$-a.e. $X \in \mathbf{R}^{n-1}$.*

Proof. Choose $r > 0$ large enough so that both $\theta^{\pm}$ vanish outside $\mathbf{R}^{n-1} \times [-r, r]$. Assume $s_X(z) \in [-r, r]$ without loss of generality (since it holds for for θ^+-a.e. (X, z)), and fix a countable dense subset $\mathcal{V}$ of the continuous test functions $C[-r, r]$ (such as polynomials with rational coefficients). Given $v \in \mathcal{V}$ and a bounded continuous test function $h \in L^\infty(\mathbf{R}^{n-1}) \cap C(\mathbf{R}^{n-1})$, Fubini's theorem combines with $s \in \mathcal{S}(\theta^+, \theta^-)$ to yield

$$\begin{aligned}
&\int_{\mathbf{R}^{n-1}} \left[\int_{\mathbf{R}} v(w)\, \theta^-_X(w) dw \right] h(X) d^{n-1}X \\
&= \int_{\mathbf{R}^n} h(X) v(w)\, \theta^-(X, w)\, d^{n-1}X dw \\
&= \int_{\mathbf{R}^n} h(X) v(s_X(z))\, \theta^+(X, z)\, d^{n-1}X dz \\
&= \int_{\mathbf{R}^{n-1}} \left[\int_{\mathbf{R}} v(s_X(z))\, \theta^+_X(z) dz \right] h(X)\, d^{n-1}X.
\end{aligned}$$

Since $h \in C(\mathbf{R}^{n-1}) \cap L^\infty(\mathbf{R}^{n-1})$ was arbitrary,

$$\int_{\mathbf{R}} v(w)\, \theta^-_X(w) dw = \int_{\mathbf{R}} v(s_X(z))\, \theta^+_X(z) dz \tag{34}$$

holds for μ-a.e. $X \in \mathbf{R}^{n-1}$, and for all v in the countable set $\mathcal{V}$. For such X we have $\theta^{\pm}_X \in L^1(\mathbf{R})$ vanishing outside $[-r, r]$, so (34) extends immediately from the dense subset $\mathcal{V}$ to its uniform closure $C[-r, r]$, and thence to all continuous functions on the line. Thus the Borel map s_X pushes θ^+ forward to θ^-: $s_X \in \mathcal{S}(\theta^+_X, \theta^-_X)$ as desired. □

Combining this lemma with the well known uniqueness of monotone measure preserving maps of the line [20] (or better [21, §A.2]), we recover uniqueness of optimal maps in flattened coordinates.

Corollary 3.4 (Unique optimal maps in flattened coordinates)
Fix $0 \le \theta^{\pm}(X,z) \in L^1(\mathbf{R}^{-1} \times \mathbf{R})$ *and* $s \in \mathcal{S}(\theta^+,\theta^-)$ *as in Lemma 3.3. Suppose another Borel map* $r \in \mathcal{S}(\theta^+,\theta^-)$ *also takes the form* $r(X,z) = (X, r_X(z))$*, and for each* $X \in \mathbf{R}^{n-1}$ *both* $r_X(z)$ *and* $s_X(z)$ *are nondecreasing functions of* $z \in \mathbf{R}$*. Then* $r = s$ *holds outside a subset of* $\mathbf{R}^{n-1} \times \mathbf{R}$ *where* θ^+ *vanishes.*

Proof. The set $N = \{(X,z) \in \mathbf{R}^{n-1} \times \mathbf{R} \mid r_X(z) \neq s_X(z)\}$ is Borel, so Fubini's theorem yields

$$\int_N \theta^+(X,z)\, d^{n-1}X dz = \int_{\mathbf{R}^{n-1}} \left[\int_{\mathbf{R}} \chi_N(X,z)\, \theta^+_X(z) dz \right] d^{N-1}X. \tag{35}$$

For $\mathcal{H}^{n-1}$-a.e. $X \in \mathbf{R}^{n-1}$, Lemma 3.3 asserts $r_X, s_X \in \mathcal{S}(\theta^+_X, \theta^-_X)$ with $\theta^{\pm}_X \in L^1(\mathbf{R}, d\mathcal{H}^1)$. Since r_X and s_X are nondecreasing, it follows that $r_X = s_X$ must hold θ^+_X-a.e., according to [20]. The integrand on the right of (35) therefore vanishes, establishing the corollary. □

Proof of Theorem 1.2. Let u be a Kantorovich potential satisfying (7), and s be the optimal map constructed in Sect. 2.5. This map s acts along transport rays of u, i.e., satisfies (8). Also since $\mu^+[\mathbf{R}^n \setminus \mathcal{X}] = 0$ and $|\mathcal{E}| = 0$, where $\mathcal{E}$ is the set of of endpoints of transport rays (Definition 2.14 and Lemma 2.15) we can set $s(x) = x$ for all $x \in \mathbf{R}^n \setminus \mathcal{X}$ and $x \in \mathcal{E}$.

Let $r : \mathbf{R}^n \to \mathbf{R}^n$ be another optimal map for Problem 1.1 satisfying (5). Note that we do not assume that r acts along the transport rays of u, i.e. that (8) holds for r and u.

Let $\mathcal{N} := \{x \in \mathcal{X} \mid s(x) \neq r(x)\}$. Then $\mathcal{N}$ is a Borel set. In order to prove Theorem 1.2, we need to show that $\mu^+[\mathcal{N}] = 0$.

Since the map s satisfies (8), it follows from Lemma 2.4(iv) that any optimal map and Kantorovich potential also satisfy (8). In particular, the map r and the function u satisfy (8) as well.

Let T_{pij} be the ray clusters associated with u. By Lemma 2.17, $\mathcal{X} \subset T_0 \cup (\cup_{pij} T_{pij})$. Since both r and s satisfy (8) with u, then $s(x) = r(x) = x$ for μ^+ a.e. $x \in T_0$, i.e., $\mu^+[\mathcal{N} \cap T_0] = 0$. Also, since $|\mathcal{E}| = 0$, we have $\mu^+[\mathcal{N} \cap \mathcal{E}] = 0$. Thus it remains to show that $\mu^+[\mathcal{N} \cap T^0_{pij}] = 0$ for all $(p,i,j) \in \mathbf{Q} \times \mathbf{N}^2$.

Fix p, i, j and let F and G be the coordinate maps associated to T^0_{pij}, defined in Lemma 2.13. Let $\theta^{\pm} = \hat{f}^{\pm}(X,\tau) dx' d\tau$ be measures on the coordinate space $\mathbf{R}^{n-1} \times \mathbf{R}^1$, where $\hat{f}^{\pm}$ are the functions (23). As both maps s and r satisfy (8), there holds $\mu^+[s(T^0_{pij}) \setminus T^0_{pij}] = \mu^+[r(T^0_{pij}) \setminus T^0_{pij}] = 0$. Thus we can define mappings $\hat{s}, \hat{r} : G(T^0_{pij}) \to G(T^0_{pij})$ by $\hat{s} = G \circ s \circ F$ and $\hat{r} = G \circ r \circ F$ for θ^+ a.e. point of $G(T^0_{pij})$. From the definition,

$$\theta^{\pm}[F^{-1}(A \cap T^0_{pij})] \equiv \theta^{\pm}[G(A \cap T^0_{pij})] = \mu^{\pm}[A \cap T^0_{pij}].$$

and thus the maps $\hat{s}$ and $\hat{r}$ push forward θ^+ onto θ^-.

Let $\hat{\mathcal{N}} := \{(X, x_n) \in G(T^0_{pij}) \mid \hat{s}(x) \neq \hat{r}(x)\}$. Then $\hat{\mathcal{N}}$ is a Borel set, and, by (22), $\mu^+[\mathcal{N} \cap T^0_{pij}] = \theta^+[\hat{\mathcal{N}}]$. Thus it remains to prove

$$\theta^+[\hat{\mathcal{N}}] = 0. \tag{36}$$

Note that the maps s and r satisfy (8), so $\hat{s}$ and $\hat{r}$ must have the following form. Write $x = (X, z) \in \mathbf{R}^{n-1} \times \mathbf{R}$. Then

$$\hat{s}(X, z) = (X, \hat{s}_X(z)) \text{ and } \hat{r}(X, z) = (X, \hat{r}_X(z)), \tag{37}$$

where for a.e. $X \in \mathbf{R}^{n-1}$ the functions $\hat{s}_X, \hat{r}_X : \mathbf{R}^1 \to \mathbf{R}^1$ satisfy $\hat{s}_X(w) \leq w$ and $\hat{r}_X(w) \leq w$. In addition, condition (5) implies that

$$\text{the functions } \hat{s}_X, \hat{r}_X \text{ are monotone nondecreasing.} \tag{38}$$

Indeed, let there exist $X \in \mathbf{R}^{n-1}$ and $w_1 > w_2$ such that $\hat{r}_X(w_1) < \hat{r}_X(w_2)$. Let $x_1 = F(X, w_1)$, $x_2 = F(X, w_2)$. By Definition 2.12, Lemma 2.7(ii), and Lemma 2.13(iii-iv) it follows that points x_1, x_2, $r(x_1)$, $r(x_2)$ lie on one transport ray R, and

$$x_1 - x_2 = (w_1 - w_2)\nu(x_1), \qquad r(x_1) - r(x_2) = [\hat{r}_X(w_1) - \hat{r}_X(w_2)]\nu(x_1),$$

where $\nu(\,\cdot\,)$ is the direction function introduced in Definition 2.9. Thus the points x_1, x_2 violate (5) for the map r. This and a similar argument for $\hat{s}_X$ prove (38).

Finally, Corollary 3.4 combines with (37–38) to prove (36), hence Theorem 1.2. □

4 Transport density

We continue to work in $\mathbf{R}^n$ metrized by a norm $\|\cdot\|$. Note that we also have a Euclidean structure on $\mathbf{R}^n$ determined by the product structure $(\mathbf{R}^1)^n$ of $\mathbf{R}^n$. Denote by $e \cdot f$ the scalar product of $e, f \in \mathbf{R}^n$, and by $|\cdot|$ the Euclidean norm $|e| = \sqrt{e \cdot e}$. The (Euclidean) gradient of a function $\varphi : \mathbf{R}^n \to \mathbf{R}^1$ at $x \in \mathbf{R}^n$ is denoted by $\nabla\varphi(x) \in \mathbf{R}^n$. We denote by $D\varphi(x) \in (\mathbf{R}^n)^*$ the derivative of φ at x. For $\psi : \mathbf{R}^n \to \mathbf{R}^n$ denote by $\operatorname{div}\psi$ the (Euclidean) divergence of ψ. Through this section $B_R(z)$ denotes closed Euclidean ball with center z and radius R, i.e., $B_R(z) := \{y \in \mathbf{R}^n;\ |y - z| \leq R\}$.

Let u be a Kantorovich potential for Problem 1.1, satisfying (7). By Theorem 1.2 and Lemma 2.4(iv) the direction of optimal mass transfer through any point of T_1^0 is uniquely defined, and is given by the direction function $\nu(z)$ introduced in Definition 2.9. It remains to study the rate of optimal mass transfer through a point of $\mathbf{R}^n$. We define below a corresponding quantity, called the *transport cost density*, and study its properties.

Imagine that as each particle of mass is transported from x to $s(x)$, it deposits a trail of dust uniformly along the line segment joining x to $s(x)$. Imagine furthermore, that the total residue of dust deposited by an individual particle is proportional to the mass of the particle times the trip tariff $\|x - s(x)\|$. The *transport cost density* $a(z)$ defined in (39) gives the cumulative density of dust deposited at $z \in T_1^0$ by all particles of μ^+ as they are transported to μ^- by a map s. Thus $a(z)$ represents a localized contribution of transportation through the point z to the total cost of redistributing μ^+ onto μ^-. We quantify this definition by choosing a particular sequence of open neighborhoods $\mathcal{D}_R(z)$ shrinking to z and setting

$$a(z) = \lim_{R \to 0+} \frac{\text{cost of transportation through } \mathcal{D}_R(z) \text{ of mass flow generated by } s}{|\mathcal{D}_R(z)|}, \tag{39}$$

if the limit exists. Our particular choice of domains $\mathcal{D}_R(z)$ is motivated by the convenience of subsequent arguments, and corresponds to a small cylinder in the flattened coordinates of Sect. 2.3. More precisely, let $S_{z,u}$ be a level set of u containing z, i.e., $S_{z,u} = \{y \mid u(y) = u(z)\}$. Define for $z \in T_1^0$, $R > 0$

$$\mathcal{D}_R(z) := \mathrm{j}[B_R(z) \cap S_{z,u}] \cap \{y \mid u(z) - R \leq u(y) \leq u(z) + R\}, \tag{40}$$

where the map j is defined by (19). By Corollary 2.18, $\mathcal{D}_R(z)$ is a Borel subset of $\mathbf{R}^n$.

The purpose of this section is to show that all optimal maps s in Monge's problem lead to the same transportation cost density $a(z)$ in (39). Furthermore, $a \in L^1(\mathbf{R}^n)$ and solves equation (45) uniquely in a suitable sense. In the Euclidean case, $a(z)$ therefore coincides with the *transport density* of Evans and Gangbo [9], and shows the latter quantity to be unique. These results are collected in Theorem 4.1.

Let us begin by computing the cost of transportation through $\mathcal{D}_R(z)$ of the mass flow generated by an optimal map s. The computation is based on the following observation. Let R be a transport ray and $x, y \in R$ satisfy $u(x) > u(y)$. Then $u(x) - u(y) = \|x - y\|$, and so the cost of transport of unit mass from x to y is $u(x) - u(y)$. It follows that, if $\tau < t$, and a total mass m is distributed within a level set $u^{-1}(t) \cap T_1^0$, then the cost of transportation of this mass along the transport rays to the level set $u^{-1}(\tau) \cap T_1^0$ is $m(t - \tau)$.

Let A be a Borel subset of level set $u^{-1}(t) \cap T_1^0$. Since the map s generates a mass flow down the transport rays of u, the mass flux through A generated by s is $\mu^+\{y \in \mathrm{j}^+(A) \mid s(y) \in \mathrm{j}^-(A)\}$, where the maps $\mathrm{j}^\pm$ are defined by (19). Note that

$$\mu^+\{y \in \mathrm{j}^+(A) \mid s(y) \in \mathrm{j}^-(A)\} = \mu^+[\mathrm{j}^+(A) \cap s^{-1}(\mathrm{j}^-(A))], \tag{41}$$

We now show that the expression (41) does not depend on a choice of an optimal map s. Indeed, for a Borel set $A \subset T_1^0 \cap u^{-1}(t)$ it follows from Corollary 2.18 that $\mathrm{j}(A)$, $\mathrm{j}^\pm(A)$ are Borel. Since $\mathrm{j}(A)$ is a transport set and $\mathrm{j}^+(A)$ is the upper end of a transport set, we have by (8) that there exist sets $\mathcal{N} \subset \mathbf{R}^n$ with $\mu^+[\mathcal{N}] = 0$ such that $s[\mathrm{j}(A)] \setminus \mathcal{N} \subset \mathrm{j}(A)$ and $s^{-1}[\mathrm{j}^+(A)] \setminus \mathcal{N} \subset \mathrm{j}^+(A)$, where the second inclusion holds since $u(y) \geq u(s(y))$ for μ^+ a.e. $y \in \mathbf{R}^n$ by (8). Thus the right-hand side of (41) can be rewritten as follows

$$\begin{aligned} \mu^+[\mathrm{j}^+(A) \cap s^{-1}(\mathrm{j}^-(A))] &= \mu^+[\mathrm{j}^+(A) \setminus s^{-1}(\mathrm{j}^+(A))] \\ &= \mu^+[\mathrm{j}^+(A)] - \mu^+[s^{-1}(\mathrm{j}^+(A))] \\ &= \mu^+[\mathrm{j}^+(A)] - \mu^-[\mathrm{j}^+(A)], \end{aligned} \tag{42}$$

where we used $s_\# \mu^+ = \mu^-$ in the last equality. Note that (42) depends only on the Kantorovich potential u, and no longer on the map s.

Now the cost of transport of the mass (42) along transport rays from $A \subset u^{-1}(t) \cap T_1^0$ to the level set $u^{-1}(t - dt)$ is

$$\Big\{\mu^+[\mathrm{j}^+(A)] - \mu^-[\mathrm{j}^+(A)]\Big\} dt.$$

So the cost of mass transport generated by any optimal map s through a Borel set $B \in \mathbf{R}^n$ is

$$\int_{-\infty}^{\infty} \left\{ \mu^+ \left[\mathrm{j}^+[B \cap u^{-1}(t)] \right] - \mu^- \left[\mathrm{j}^+[B \cap u^{-1}(t)] \right] \right\} dt.$$

Thus we define the transport cost density of the flow generated by any optimal map s at the point $z \in T_1^0$ as

$$a(z) := \lim_{R \to 0+} \frac{\int_{-\infty}^{\infty} \left\{ \mu^+ \left[\mathrm{j}^+[\mathcal{D}_R(z) \cap u^{-1}(t)] \right] - \mu^- \left[\mathrm{j}^+[\mathcal{D}_R(z) \cap u^{-1}(t)] \right] \right\} dt}{|\mathcal{D}_R(z)|}, \tag{43}$$

where $\mathcal{D}_R(z)$ is defined by (40). We show below that for $R > 0$ the integrand in the right-hand side of (43) is an integrable function of t. Thus the question addressed by our final theorem is existence of the limit, and properties of $a(z)$. We make several remarks before giving its proof.

Theorem 4.1 (Existence, uniqueness, and properties of transport density) *Fix a Kantorovich potential u satisfying (7) and let ν be the corresponding direction function from Definition 2.9.*

i. Limit (43) exists a.e. on T_1^0 and does not depend on a choice of optimal map s.
ii. There exists $a \in L^1(\mathbf{R}^n)$, called the transport cost density, *with the following properties:*

$$a \geq 0 \text{ on } \mathbf{R}^n, \quad a \equiv 0 \text{ on } \mathbf{R}^n \setminus T_1, \tag{44}$$

and $a(z)$ is equal to the right-hand side of (43) for $\mathcal{L}^n$ a.e. $z \in T_1^0$. In addition, $a(\,\cdot\,)$ satisfies the equation

$$-div\,(a\,\nu) = f^+ - f^- \quad in\ \mathbf{R}^n \tag{45}$$

in the weak sense, meaning any test function $\varphi \in C^1(\mathbf{R}^n)$ obeys

$$\int_{\mathbf{R}^n} a\,\nu \cdot \nabla\varphi\,dz = \int_{\mathbf{R}^n} (f^+ - f^-)\varphi\,dz. \tag{46}$$

Moreover, for any measurable transport set $A \subset \mathbf{R}^n$ and $\varphi \in C^1(\mathbf{R}^n)$

$$\int_A a\,\nu \cdot \nabla\varphi\,dz = \int_A (f^+ - f^-)\varphi\,dz. \tag{47}$$

iii. A function $a \in L^1(\mathbf{R}^n)$ satisfying (47) for all measurable transport sets A and $\varphi \in C^1(\mathbf{R}^n)$ is uniquely determined by the constraints (44).

Remark 4.2 (Euclidean transport density) In the case when $\|\cdot\|$ is the Euclidean norm $|\cdot|$, we have $\nu(x) = \nabla u(x)$ on T_1^0. Thus the equations (45) and (46) have the form

$$-\mathrm{div}\,(a\nabla u) = f^+ - f^-; \tag{48}$$

$$\int_{\mathbf{R}^n} a\nabla u \cdot \nabla\varphi\,dz = \int_{\mathbf{R}^n} (f^+ - f^-)\varphi\,dz. \tag{49}$$

Remark 4.3 (Vanishing at ray ends) Let us discuss the property (47). For simplicity we consider the case when $\|\cdot\|$ is the Euclidean norm. Then (47) becomes: for any transport set A

$$\int_A a\nabla u \cdot \nabla\varphi\, dz = \int_A (f^+ - f^-)\varphi\, dz \qquad \text{for any} \quad \varphi \in C^1(\mathbf{R}^n). \tag{50}$$

Property (50) can be heuristically interpreted as following: equation (48) holds in $\mathbf{R}^n$, and on the boundary of any transport set A the Neumann-type condition holds

$$a\partial_n u = 0 \qquad \text{on} \;\; \partial A, \tag{51}$$

where $\partial_n u$ is the normal derivative of u. Indeed, (50) implies (51) if ∂A is smooth. Note that (51) means that there is no mass transfer through the boundary of any transport set.

Let us further interpret the condition (51). The discussion is mostly heuristic. Since A is a "cylinder" of transport rays, it is natural to write $\partial A = \Gamma_1 \cup \Gamma_2$, where $\Gamma_1 = \partial A \setminus \mathcal{E}$ is the set of interior points of rays, and $\Gamma_2 = \partial A \cap \mathcal{E}$ is the set of endpoints of rays. By Lemma 2.7(iii), for $x \in \Gamma_1$ the Kantorovich potential u is differentiable at x, and $\nabla u(x)$ is a unit vector in the direction of the transport ray R_x containing x. Thus $\partial_n u(x) = 0$ on Γ_1. So (51) naturally holds on Γ_1. On Γ_2 we cannot expect $\partial_n u = 0$, and thus the meaning of (51) is that $a(\,\cdot\,)$ vanishes (in a certain generalized sense) at the endpoints of transport rays. Note that since all mass transfer occurs within transport rays, there is no mass transfer through endpoints of rays, so the vanishing of the transport density at endpoints of rays is a natural property.

Evans and Gangbo [9] proved that if $f^\pm$ are Lipschitz functions, then $a(\,\cdot\,)$ indeed vanishes on $\mathcal{E}$ in the following sense: the limit of $a(\,\cdot\,)$ along the transport rays is zero at the rays ends. This was an important property for constructing an optimal map in [9]. The vanishing of $a(\,\cdot\,)$ along the rays as the end of a transport ray is approached was also a crucial property for deriving and justifying the law of evolution of a sandpile shape in [11], [13], [14]. Also, in [13] the property (50) was shown for a restricted class of $f^\pm$.

In the general case $f^\pm \in L^1(\mathbf{R}^n)$ one can construct examples in which $a(\,\cdot\,)$ has a positive limit or blows up to $+\infty$ along transport rays at the endpoints of rays (one example is given below). Thus vanishing of transport density on $\mathcal{E}$ along the transport rays holds in general only in the sense of (50).

Example 4.4 (Non-vanishing at ray ends) The example, in $\mathbf{R}^2$, of transport density blowup around $\mathcal{E}$ is the following: $f^+(x) = 2\chi_{B_1(0)}(x)$, $f^-(x) = \dfrac{1}{2|x|^{\frac{3}{2}}}\chi_{B_1(0)}(x)$, where $\chi_{B_1(0)}(\,\cdot\,)$ is the characteristic function of the disk $B_1(0) \subset \mathbf{R}^2$. Then $u(x) = |x|$, transport rays are radii of the disk $B_1(0)$, and $\mathcal{E} = \{0\} \cup \partial B_1(0)$, where the point $x = 0$ is the lower end of all transport rays. The transport density is $a(x) = \left(\dfrac{1}{\sqrt{|x|}} - |x|\right)\chi_{B_1(0)}(x)$, and it blows up along every transport ray at its lower end $x = 0 \in \mathcal{E}$.

Proof of Theorem 4.1. Let us first show that for $R > 0$ the integrand in the right-hand side of (43) is an integrable function of t. Denote $C_R(z) := \mathrm{j}[B_R(z) \cap S_{z,u}]$. By Corollary 2.18, $C_R(z)$ is a Borel subset of $\mathbf{R}^n$. Also, since T_1^0 is a bounded set, $C_R(z)$ is bounded. We have

$$\begin{aligned} A &:= \bigcup_{t\in[-\infty,\infty]} \mathrm{j}^+[\mathcal{D}_R(z) \cap u^{-1}(t)] \times \{t\} \qquad (52) \\ &= \{(y,t) \in \mathbf{R}^n \times \mathbf{R}^1 \ \mid \ u(z) - R \le t \le u(z) + R,\ u(y) \ge t,\ y \in C_R(z)\}, \end{aligned}$$

where we used the convention $\emptyset \times \{t\} = \emptyset$. It follows that A is a Borel subset of $\mathbf{R}^n \times \mathbf{R}^1$, bounded since $A \subset C_R(z) \times [u(z) - R, u(z) + R]$. On $\mathbf{R}^n \times \mathbf{R}^1$ consider the product measure $\mu^+ \times \mathcal{L}^1$. Since μ^+ is a Radon measure, $\mu^+ \times \mathcal{L}^1$ is also a Radon measure. Thus the Borel set $A \subset \mathbf{R}^n \times \mathbf{R}^1$ is $(\mu^+ \times \mathcal{L}^1)$-measurable, and, since A is bounded, $(\mu^+ \times \mathcal{L}^1)(A) < \infty$. Using Fubini's theorem [10, §1.4.1], we conclude that $t \to \mu^+[A_t] \equiv \mu^+\left[\mathrm{j}^+[\mathcal{D}_R(z) \cap u^{-1}(t)]\right]$ is measurable, and

$$\int_{-\infty}^{\infty} \mu^+\left[\mathrm{j}^+[\mathcal{D}_R(z) \cap u^{-1}(t)]\right] dt = (\mu^+ \times \mathcal{L}^1)(A) < \infty. \qquad (53)$$

A similar conclusion holds for the measure μ^-. Thus for $R > 0$ the right-hand side of (43) is well-defined and finite if $|\mathcal{D}_R(z)| > 0$.

Now we show that the limit in (43) exists a.e. in T_1^0, and defines a function which satisfies (44) and (47).

Step 1. Limit in (43) exists a.e. We first obtain a convenient expression for the integral in the right-hand side of (43). By (53)

$$\int_{-\infty}^{\infty} \mu^\pm\left[\mathrm{j}^+[\mathcal{D}_R(z) \cap u^{-1}(t)]\right] dt = \int_A f^\pm(z) dz dt \qquad (54)$$

Let T_{pij}^0 for $(p,i,j) \in \mathbf{Q} \times \mathbf{N}^2$ be the ray clusters introduced in Definition 2.12. Since $A \subset T_1^0 \times \mathbf{R}^1$, we get $A = \bigcup_{pij} A_{pij}$, where the sets $A_{pij} = A \cap (T_{pij}^0 \times \mathbf{R}^1)$ are disjoint (18), and Borel by Lemma 2.15. Thus we can replace the right-hand side of (54) by a sum of integrals over the sets A_{pij}, and in each integral make a Lipschitz change of variables $(z,t) = (F(x),t)$ for $(x,t) \in (G \times Id)(A_{pij})$, where $F = F_{pij}$, $G = G_{pij}$ are the mappings from Lemma 2.13, and $Id : \mathbf{R}^1 \to \mathbf{R}^1$ is the identity map. The set $(G \times Id)(A_{pij})$ is Borel since $(G \times Id)(A_{pij}) = (F \times Id)^{-1}(A_{pij}) \cap [G(T_{pij}^0) \times \mathbf{R}^1]$ and the map F is Lipschitz. Note that $G(T_{pij}^0)$ is Borel by Lemma 2.15. Thus we get from (54)

$$\int_{-\infty}^{\infty} \mu^\pm\left[\mathrm{j}^+[\mathcal{D}_R(z) \cap u^{-1}(t)]\right] dt = \sum_{(p,i,j)\in\mathbf{Q}\times\mathbf{N}^2} \int_{(G\times Id)(A_{pij})} \hat{f}^\pm(x) dx dt, \qquad (55)$$

where the functions $\hat{f}^{\pm} = \hat{f}^{\pm}_{pij}$ are defined by (23). Note that, writing $x = (X, \tau) \in \mathbf{R}^{n-1} \times \mathbf{R}^1$, we get from (16), (52):

$$(G \times Id)(A_{pij}) = \Big\{(X, \tau, t) \in \mathbf{R}^{n-1} \times \mathbf{R}^1 \times \mathbf{R}^1 \ \mid \ (X, t) \in G[\mathcal{D}_R(z) \cap T^0_{pij}], \\ \tau \geq t, \ (X, \tau) \in G(T^0_{pij})\Big\}.$$

Since $\hat{f}^{\pm}$ vanish a.e. in $\mathbf{R}^n \setminus G(T^0_{pij})$, we can integrate in the right-hand side of (55) over the set

$$\Big\{(X, \tau, t) \in \mathbf{R}^{n-1} \times \mathbf{R}^1 \times \mathbf{R}^1 \ \mid \ (X, t) \in G[\mathcal{D}_R(z) \cap T^0_{pij}], \ \ \tau \geq t\Big\}$$

and obtain

$$\int_{-\infty}^{\infty} \mu^{\pm}\Big[\mathrm{j}^{+}[\mathcal{D}_R(z) \cap u^{-1}(t)]\Big]dt \\ = \sum_{(p,i,j) \in \mathbf{Q} \times \mathbf{N}^2} \int_{G[\mathcal{D}_R(z) \cap T^0_{pij}]} \Big(\int_t^{\infty} \hat{f}^{\pm}(X, \tau) d\tau\Big) dX\, dt.$$

We rewrite this using the functions $\Psi^{\pm} = \Psi^{\pm}_{pij}$ defined by (24):

$$\int_{-\infty}^{\infty} \Big\{\mu^{+}\Big[\mathrm{j}^{+}[\mathcal{D}_R(z) \cap u^{-1}(t)]\Big] - \mu^{-}\Big[\mathrm{j}^{+}[\mathcal{D}_R(z) \cap u^{-1}(t)]\Big]\Big\}dt \\ = \sum_{(p,i,j) \in \mathbf{Q} \times \mathbf{N}^2} \int_{G[\mathcal{D}_R(z) \cap T^0_{pij}]} [\Psi^{+}(X, t) - \Psi^{-}(X, t)] dX\, dt, \tag{56}$$

Note that $\Psi^{\pm}$ are Borel, nonnegative, locally summable functions on $\mathbf{R}^{n-1} \times \mathbf{R}^1$.

Denoting

$$\theta[B] := \sum_{(p,i,j) \in \mathbf{Q} \times \mathbf{N}^2} \int_{G(B \cap T^0_{pij})} [\Psi^{+}(X, t) - \Psi^{-}(X, t)] dX\, dt, \tag{57}$$

we see that right-hand side of (56) is $\theta[\mathcal{D}_R(z)]$. Note that (57) is well-defined for any Borel set $B \subset \mathbf{R}^n$. Indeed, $G(B \cap T^0_{pij}) = F^{-1}(B \cap T^0_{pij}) \cap G(T^0_{pij})$ is a Borel set since F is Lipschitz and $G(T^0_{pij})$ is Borel by Lemma 2.15. Thus we need only show that the series on the right-hand side of (57) converges. By (25) each term of the series is nonnegative. Moreover, by (25)

$$\int_{G(B \cap T^0_{pij})} [\Psi^{+}(X, t) - \Psi^{-}(X, t)] dX\, dt \leq \int_{G(T^0_{pij})} [\Psi^{+}(X, t) - \Psi^{-}(X, t)] dX\, dt.$$

Fix p, i, j and denote $\hat{f}^{\pm}_X(\tau) := \hat{f}^{\pm}(X, \tau)$. For a.e. $X \in \mathbf{R}^{n-1}$ we have $\hat{f}^{\pm}_X \in L^1(\mathbf{R}^1)$. Thus by (24) the functions $\Psi^{\pm}$ are continuous on a.e. vertical line $\{X\} \times \mathbf{R}^1$. Also, for a.e. $X \in \mathbf{R}^{n-1}$, the functions $\hat{f}^{\pm}_X$ vanish outside a segment $[c, d]$ of finite length, since $f^\pm$ have compact supports. Thus $\Psi^\pm$ are constant on $\{X\} \times (-\infty, c]$ and on $\{X\} \times [d, \infty)$ for a.e. $X \in \mathbf{R}^{n-1}$. Using (24) and (26), we have

$$\Psi^+(X,\tau) = \Psi^-(X,\tau) \quad \text{for all } \tau \le c \quad \text{and for all } \tau \ge d. \tag{58}$$

We also have $\frac{\partial}{\partial t}\Psi^\pm(X,t) = -f^\pm(X,t)$. Thus, integrating by parts with respect to t on the segment $[c,d]$ (see [12, §2.6.7]) and, using (58) to cancel the boundary terms, we obtain

$$\int_{G(T^0_{pij})} [\Psi^+(X,t) - \Psi^-(X,t)]dX\,dt = \int_{G(T^0_{pij})} t[\hat{f}^+(X,t) - \hat{f}^-(X,t)]dX\,dt.$$

In the last expression we change variables to $z = F(X,t)$. Now, using (23), and noting that by (16) we have $t = u(z) - p$, we get

$$\begin{aligned}\int_{G(T^0_{pij})} t[\hat{f}^+(X,t) - \hat{f}^-(X,t)]dX\,dt &= \int_{T^0_{pij}} [u(z)-p][f^+(z) - f^-(z)]dz \\ &= \int_{T^0_{pij}} u(z)[f^+(z) - f^-(z)]dz,\end{aligned}$$

where we used detailed mass balance (20) on the Borel transport set T^0_{pij} to obtain the second equality. Thus the series with nonnegative terms in the right hand side of (57) is bounded from above by the following convergent series

$$\sum_{(p,i,j)\in\mathbf{Q}\times\mathbf{N}^2} \int_{T^0_{pij}} u(z)[f^+(z)-f^-(z)]dz = \int_{\mathbf{R}^n} u(z)[f^+(z)-f^-(z)]dz, \tag{59}$$

and thus the right hand side of (57) converges.

Define on $\mathbf{R}^n$ an (outer) measure θ as follows: for Borel $B \subset \mathbf{R}^n$ define $\theta[B]$ by (57), and for any other $A \subset \mathbf{R}^n$ define

$$\theta[A] = \inf\{\theta[B] \mid A \subset B, \text{ where } B \text{ is a Borel subset of } \mathbf{R}^n\}. \tag{60}$$

Since, by (57), $\theta[B_1 \cup B_2] \le \theta[B_1] + \theta[B_2]$ for Borel B_1, B_2, it follows that θ is indeed an (outer) measure.

Lemma 4.5 (Absolute continuity and integrability) *θ defined by (57) and (60) is a Radon measure on $\mathbf{R}^n$, absolutely continuous with respect to $\mathcal{L}^n$, and $\theta[\mathbf{R}^n] < \infty$.*

Proof of Lemma 4.5. To see that the measure θ is Borel: let B_1, $B_2 \subset \mathbf{R}^n$ be Borel sets with $\operatorname{dist}(B_1, B_2) > 0$. For any $(p,i,j) \in \mathbf{Q}\times\mathbf{N}^2$ we have $G(B_1 \cap T^0_{pij}) \cap G(B_2 \cap T^0_{pij}) = \emptyset$ since the map G is one-to-one on T^0_{pij}. Thus $\theta[B_1 \cup B_2] = \theta[B_1] + \theta[B_2]$ by (57). Let $A_1, A_2 \subset \mathbf{R}^n$ be any sets satisfying $\operatorname{dist}(A_1, A_2) = 3\delta > 0$. By (60) there exist Borel sets $B^k \subset \mathbf{R}^n$ for $k = 1, 2, \ldots$ such that $A_1 \cup A_2 \subset B^k$ and $\theta[B^k] - \theta[A_1 \cup A_2] < \frac{1}{k}$ for $k = 1, 2, \ldots$. Denote by A_1^δ, A_2^δ the δ-neighborhoods of A_1, A_2. Then A_1^δ, A_2^δ are open sets and $\mathrm{dist}(A_1^\delta, A_2^\delta) \geq \delta > 0$. Then we have using (60)

$$\begin{aligned}\theta[A_1 \cup A_2] = \lim_{k\to\infty} \theta[B^k] &\geq \liminf_{k\to\infty} \theta[(B^k \cap A_1^\delta) \cup (B^k \cap A_2^\delta)] \\ &= \liminf_{k\to\infty} \left(\theta[B^k \cap A_1^\delta] + \theta[B^k \cap A_2^\delta]\right) \\ &\geq \theta[A_1] + \theta[A_2].\end{aligned}$$

Thus the measure θ is Borel by Caratheodory's criterion [12, §2.3.2(9)].

Now the measure θ on $\mathbf{R}^n$ is Borel regular by (60).

Next we show that θ is absolutely continuous with respect to $\mathcal{L}^n$. Indeed, consider first a Borel $B \subset \mathbf{R}^n$ such that $|B| = 0$. Then for each $(p, i, j) \in \mathbf{Q} \times \mathbf{N}^2$ we have $|G(B \cap T^0_{pij})| = 0$. Indeed, for each $\lambda > 0$ the map G is Lipschitz on the Borel set T^λ_{pij}, defined in Lemma 2.13(ii). Thus $|G(B \cap T^\lambda_{pij})| = 0$ for any $\lambda > 0$. We have $T^{\lambda_1}_{pij} \subset T^{\lambda_2}_{pij}$ for any $\lambda_1 > \lambda_2 > 0$, and $T^0_{pij} = \bigcup_{k=1}^{\infty} T^{\frac{1}{k}}_{pij}$. Thus $|G(B \cap T^0_{pij})| = 0$. Then, by (57) $\theta[B] = 0$. Now let $A \subset \mathbf{R}^n$ and $|A| = 0$. Then, since $\mathcal{L}^n$ is Borel regular, there exists a Borel $B \subset \mathbf{R}^n$, such that $A \subset B$ and $|B| = 0$. Then $\theta[B] = 0$, and thus $\theta[A] = 0$.

Finally, since we estimated (57) by (59), we showed that $\theta[A] \leq \int_{\mathbf{R}^n} u(z)[f^+(z) - f^-(z)]dz < \infty$ for any $A \subset \mathbf{R}^n$, which implies that θ is a Radon measure. Lemma 4.5 is proved $\square$

Remark 4.6 (Total cost bounds) For $B = \mathbf{R}^n$, the argument estimating (57) by (59) shows $\theta[\mathbf{R}^n] = \int_{\mathbf{R}^n} u(z)[f^+(z) - f^-(z)]dz$, which coincides with the total transportation cost $K[u] = I[s]$.

Using the absolute continuity proved for the finite Radon measure θ in Lemma 4.5, the Radon-Nikodym theorem [10, §1.6.2] yields a density $0 \leq g \in L^1(\mathbf{R}^n)$ of θ with respect to $\mathcal{L}^n$. Any measurable set $A \subset \mathbf{R}^n$ satisfies

$$\theta[A] = \int_A g(y)dy. \tag{61}$$

and

$$\|g\|_{L^1} = \theta[\mathbf{R}^n] < \infty. \tag{62}$$

Lemma 4.7 (Local transport cost density) *Let $\lambda > 0$ and $(p, i, j) \in \mathbf{Q} \times \mathbf{N}^2$. Let $z \in \mathbf{R}^n$ be a Lebesgue point for the function g, and a point of density 1 for the set T^λ_{pij} of Lemma 2.13(ii). Then for z the limit (43) exists and converges to $g(z)$.*

Proof. By (56) and (57) we can write the right-hand side of (43) as

$$\lim_{R\to 0+} \frac{\theta[\mathcal{D}_R(z)]}{|\mathcal{D}_R(z)|},$$

or, by (61),

$$\lim_{R\to 0+} \frac{1}{|\mathcal{D}_R(z)|} \int_{\mathcal{D}_R(z)} g(y)dy.$$

By [24, §7.9–10], this limit is $g(z)$ if z is a Lebesgue point of g, provided that the family of sets $\mathcal{D}_R(z)$ shrink nicely to z in the following sense: there exist $M, M_1 > 0$ and $R_0 > 0$, depending on z, such that

$$\mathcal{D}_R(z) \subset B_{MR}(z), \qquad |\mathcal{D}_R(z)| \geq \frac{1}{M_1} R^n \quad \text{for all } R \in (0, R_0). \tag{63}$$

According to Lemma 4.8 below, $\mathcal{D}_r(z)$ shrinks nicely to any point z of density 1 for the set T^λ_{pij}. Thus Lemma 4.7 follows from Lemma 4.8. □

The following lemma quantifies the rate at which the cylinders $\mathcal{D}_R(z)$ shrink nicely.

Lemma 4.8 (Cylinders shrink nicely)

i. For any $R > 0$ and $z \in \mathbf{R}^n$ there holds $\mathcal{D}_R(z) \subset B_{MR}(z)$, where $M = (1 + \sup_{\|e\|=1} |e|)$.

ii. Let $\lambda > 0$ and z be a point of density 1 for the set T^λ_{pij}. Then there exist M_1 and $R_0 > 0$ such that

$$|\mathcal{D}_R(z) \cap T^\lambda_{pij}| \geq \frac{1}{M_1} R^n \quad \textit{for all } R \in (0, R_0).$$

Proof. Let $x \in \mathcal{D}_R(z)$. Then, by (40), $|u(x) - u(z)| \leq R$, and there exists a point $y \in B_R(z)$ such that $|u(x) - u(y)| = \|x - y\|$ and $u(y) = u(z)$. Then

$$|x - y| = |u(x) - u(y)| \frac{|x - y|}{\|x - y\|} \leq |u(x) - u(z)| \sup_{\|e\|=1} |e| \leq R \sup_{\|e\|=1} |e|.$$

Thus

$$|x - z| \leq |x - y| + |y - z| \leq (1 + \sup_{\|e\|=1} |e|)R,$$

which proves (i).

Now we address (ii). We first prove

$$B_{\frac{R}{M}}(z) \cap T^\lambda_{pij} \subset \mathcal{D}_R(z) \cap T^\lambda_{pij}. \tag{64}$$

for $R \in (0, R_0)$, where R_0 small enough and M large enough are selected below. Let $x \in B_{\frac{R}{M}}(z) \cap T^\lambda_{pij}$. Let R_x be the unique transport ray containing x. In order to prove (64) it is enough to show that R_x intersects the level set $S_{z,u} = \{\xi \in \mathbf{R}^n \mid u(\xi) = u(z)\}$, and the point y of intersection of R_x and $S_{z,u}$ satisfies $y \in T^0_{pij}$ and $|y - z| < R$.

Since $x \in T^\lambda_{pij}$, for any $\tau \in (-\lambda, \lambda)$ there exists $y \in R_x$ with $u(y) = u(x) + \tau$. Since $x \in B_{\frac{R}{M}}(z)$,

$$|u(x) - u(z)| \leq \|x - z\| \leq |x - z| \sup_{|e|=1} \|e\| \leq \frac{\sup_{|e|=1} \|e\|}{M} R. \tag{65}$$

Thus if $M \geq 2\frac{\sup_{|e|=1} \|e\|}{\lambda}$, then

$$|u(x) - u(z)| \leq \frac{\lambda}{2} R.$$

Thus if $R \leq 1$ there exists $y \in R_x$ with $u(y) = u(z)$, i.e., y is the point of intersection of R_x with $S_{z,u}$. Then $|u(x) - u(y)| = \|x - y\|$ and

$$|x - y| \leq \|x - y\| \sup_{\|e\|=1} |e| = |u(x) - u(y)| \sup_{\|e\|=1} |e|.$$

Since $x \in B_{\frac{R}{M}}(z)$,

$$\begin{aligned} |y - z| &\leq |y - x| + |x - z| \leq |u(x) - u(y)| \sup_{\|e\|=1} |e| + \frac{R}{M} \\ &\leq \frac{(\sup_{\|e\|=1} |e|)(\sup_{|e|=1} \|e\|) + 1}{M} R, \end{aligned}$$

where we used $u(y) = u(z)$ and (65) in the last inequality. Thus $|y - z| \leq \frac{R}{2}$ if $M \geq 2[(\sup_{\|e\|=1} |e|)(\sup_{|e|=1} \|e\|) + 1]$. This implies that $x \in \mathcal{D}_R(z)$. Recall also that $x \in T^\lambda_{pij}$. Thus (64) holds, if $R \leq 1$ and M satisfies all above conditions, i.e. if M is large depending on λ, $\sup_{|e|=1} \|e\|$ and $\sup_{\|e\|=1} |e|$. Fix such an M.

Since z is a point of density 1 for T^λ_{pij}, there exists ρ_0 such that for any $\rho \in (0, \rho_0)$

$$|B_\rho(z) \cap T^\lambda_{pij}| \geq \frac{1}{2} |B_\rho|. \tag{66}$$

Thus choosing $R_0 = \rho_0 M$, applying (66) to $\rho = \frac{R}{M}$ and using (64), we get (ii). This concludes the proof of Lemma 4.8. □

We have $T^0_1 = \bigcup_{(p,i,j) \in \mathbf{Q} \times \mathbf{N}^2} T^0_{pij}$ and $T^0_{pij} = \bigcup_{k=1}^{\infty} T^{\frac{1}{k}}_{pij}$ for each $(p, i, j) \in \mathbf{Q} \times \mathbf{N}^2$, and the sets T^0_{pij}, $T^{\frac{1}{k}}_{pij}$ are Borel. From Lemma 4.7 the limit on the right-hand side of (43) converges to $g(z)$ for a.e. $z \in T^{\frac{1}{k}}_{pij}$, and thus for a.e. $z \in T^0_1$. Since T_1 is a closed set, it follows from the definition of the measure θ that $\theta[\mathbf{R}^n \setminus T_1] = 0$, and so $g \equiv 0$ on $\mathbf{R}^n \setminus T_1 = 0$. In addition, $g \in L^1$ by (62). The endpoints of transport rays $\mathcal{E} = T_1 \setminus T^0_1$ occupy zero volume in Lemma 2.15, so the function $a = g$ satisfies (44) and is equal to the right-hand side of (43) for $\mathcal{L}^n$ a.e. $z \in T^0_1$. This concludes the first assertion of the theorem. In Steps 2 and 3 we address (45–47).

Step 2. Transport cost density on a ray cluster in flattened coordinates.

Lemma 4.9 *Let $\lambda > 0$ and $(p, i, j) \in \mathbf{Q} \times \mathbf{N}^2$. Then for a.e. $z \in T^\lambda_{pij}$*

$$a(z) = \lim_{R \to 0+} \frac{\theta[\mathcal{D}_R(z) \cap T^\lambda_{pij}]}{|\mathcal{D}_R(z) \cap T^\lambda_{pij}|} \tag{67}$$

Proof. Define a measure $\tilde{\theta}$ on $\mathbf{R}^n$ by

$$\tilde{\theta}[A] = \theta[A \setminus T^\lambda_{pij}]$$

for any $A \subset \mathbf{R}^n$. Since T^λ_{pij} is a Borel set, it follows from Lemma 4.5 that $\tilde{\theta}$ is a Radon measure, absolutely continuous with respect to $\mathcal{L}^n$. Thus there exists a density $\tilde{g} \geq 0$ of $\tilde{\theta}$ with respect to $\mathcal{L}^n$. Then $\tilde{g} \in L^1(\mathbf{R}^n)$ since $\tilde{\theta}[\mathbf{R}^n] < \infty$. We have $\tilde{\theta}[T^\lambda_{pij}] = \theta[\emptyset] = 0$, and thus $\tilde{g} = 0$ a.e. in T^λ_{pij}.

Let $z \in T^\lambda_{pij}$ be Lebesgue point of both $g(\,\cdot\,)$ and $\tilde{g}(\,\cdot\,)$, and let z be a point of density 1 for T^λ_{pij} and a point of density 0 for $\mathbf{R}^n \setminus T^\lambda_{pij}$. Note that a.e. $z \in T^\lambda_{pij}$ satisfies these conditions. We show the lemma holds for such z.

Note also that, since $\tilde{g} = 0$ a.e. in T^λ_{pij} and $\tilde{g} \in L^1$, the above conditions imply $\tilde{g}(z) = 0$. In addition, by Lemma 4.7,

$$g(z) = \lim_{R \to 0+} \frac{\theta[\mathcal{D}_R(z)]}{|\mathcal{D}_R(z)|}.$$

Also, by a similarly proof as Lemma 4.7, we get

$$\tilde{g}(z) = \lim_{R \to 0+} \frac{\tilde{\theta}[\mathcal{D}_R(z)]}{|\mathcal{D}_R(z)|}.$$

In addition, since z is a point of density 0 for $\mathbf{R}^n \setminus T^\lambda_{pij}$, we have $|B_R(z) \setminus T^\lambda_{pij}| = o(R^n)$. Also Lemma 4.8(ii) holds since z is a point of density 1 for T^λ_{pij}. Thus we have

$$0 \leq \frac{|\mathcal{D}_R(z) \setminus T^\lambda_{pij}|}{|\mathcal{D}_R(z)|} \leq \frac{|B_{MR}(z) \setminus T^\lambda_{pij}|}{|\mathcal{D}_R(z)|} \leq \frac{o(R^n)}{\frac{1}{M_1} R^n} \to 0 \;\text{ as }\; R \to 0+,$$

so

$$\lim_{R \to 0+} \frac{|\mathcal{D}_R(z)|}{|\mathcal{D}_R(z) \cap T^\lambda_{pij}|} = 1.$$

Now we compute:

$$\frac{\theta[\mathcal{D}_R(z) \cap T^\lambda_{pij}]}{|\mathcal{D}_R(z) \cap T^\lambda_{pij}|} = \frac{\theta[\mathcal{D}_R(z)] - \theta[\mathcal{D}_R(z) \setminus T^\lambda_{pij}]}{|\mathcal{D}_R(z)|} \frac{|\mathcal{D}_R(z)|}{|\mathcal{D}_R(z) \cap T^\lambda_{pij}|}. \tag{68}$$

From the discussion above, the limit as $R \to 0+$ of the right-hand side of (68) exists and is $g(z) - \tilde{g}(z) = g(z) = a(z)$. Thus Lemma 4.9 is proved. □

Lemma 4.10 (Transport cost density in flattened coordinates) *Let* $(p, i, j) \in \mathbf{Q} \times \mathbf{N}^2$*, and let* F, G *be the mappings defined in Lemma 2.13 for* T^0_{pij}*. Then for a.e.* $z \in T^0_{pij}$*,*

$$a(z) = \frac{\hat{a}(G(z))}{J_n F(G(z))}, \tag{69}$$

where $\hat{a} : \mathbf{R}^{n-1} \times \mathbf{R}^1 \to \mathbf{R}^1$ *is defined by*

$$\hat{a}(X, x_n) = \Psi^+(X, x_n) - \Psi^-(X, x_n). \tag{70}$$

Proof. Since $T^0_{pij} = \bigcup_{k=1}^{\infty} T^{\frac{1}{k}}_{pij}$, it is enough to prove that (69) holds for a.e. $z \in T^{\frac{1}{k}}_{pij}$ for every $k \in \mathbf{N}$. Therefore, fix $k \in \mathbf{N}$ and let $\lambda = \frac{1}{k}$. Let $z \in T^{\lambda}_{pij}$ be a point at which (67) holds. Using (57) and the Area formula, we rewrite (67) as follows

$$\begin{aligned} a(z) &= \lim_{R\to 0+} \frac{\int_{G(\mathcal{D}_R(z)\cap T^{\lambda}_{pij})}[\Psi^+(X,x_n)-\Psi^-(X,x_n)]dX\,dx_n}{\int_{G(\mathcal{D}_R(z)\cap T^{\lambda}_{pij})} J_nF(X,x_n)dX\,dx_n} \\ &= \lim_{R\to 0+} \frac{\int_{G(\mathcal{D}_R(z)\cap T^{\lambda}_{pij})} \hat{a}(X,x_n)dX\,dx_n}{\int_{G(\mathcal{D}_R(z)\cap T^{\lambda}_{pij})} J_nF(X,x_n)dX\,dx_n}, \end{aligned} \tag{71}$$

where J_nF is the Jacobian of F. Note that both functions $\hat{a}(\,\cdot\,)$ and $J_nF(\,\cdot\,)$ are locally integrable.

Let z be a point with the following properties: (67) holds, z is a point of density 1 for T^{λ}_{pij}, and $G(z)$ is a Lebesgue point for both $\hat{a}(\,\cdot\,)$ and $J_nF(\,\cdot\,)$. Since the map F is Lipschitz on $\mathbf{R}^n$, and G is Lipschitz on T^{λ}_{pij} by Lemma 2.13, then a.e. $z \in T^{\lambda}_{pij}$ satisfies these conditions.

It follows that Lemma 4.8 holds for z. Since F is Lipschitz (say, with constant L), it follows from Lemma 4.8(i) that $G(\mathcal{D}_R(z) \cap T^{\lambda}_{pij}) \subset B_{LMR}(G(z))$ for any $R > 0$, and from Lemma 4.8(ii) and the Area formula

$$L^n|G(\mathcal{D}_R(z) \cap T^{\lambda}_{pij})| \geq |\mathcal{D}_R(z) \cap T^{\lambda}_{pij}| \geq \frac{1}{M_1}R^n \quad \text{for all } R \in (0, R_0).$$

Thus, by Rudin [24, §7.9–10], since $G(z)$ is a Lebesgue point for both $\hat{a}$ and J_nF,

$$\lim_{R\to 0+} \frac{1}{|G(\mathcal{D}_R(z) \cap T^{\lambda}_{pij})|} \int_{G(\mathcal{D}_R(z)\cap T^{\lambda}_{pij})} \hat{a}(X,x_n)dX\,dx_n = \hat{a}(G(z)) \qquad \text{and}$$

$$\lim_{R\to 0+} \frac{1}{|G(\mathcal{D}_R(z) \cap T^{\lambda}_{pij})|} \int_{G(\mathcal{D}_R(z)\cap T^{\lambda}_{pij})} J_nF(X,x_n)dX\,dx_n = J_nF(G(z)). \tag{72}$$

Note also that by Lemma 2.13(iii), $J_nF(G(z))J_nG(z) = 1$ for a.e. $z \in T^{\lambda}_{pij}$. Thus

$$J_nF(G(z)) \neq 0 \quad \text{for a.e. } z \in T^{\lambda}_{pij}, \tag{73}$$

and we can pass to the limit in (71) using (72). Thus Lemma 4.10 is proved. □

Step 3. Differential equation satisfied by the transport cost density.
Now we prove that that $a(z)$ satisfies (46–47). Let A be a Borel transport set. Fix T^0_{pij} for $(p,i,j) \in \mathbf{Q} \times \mathbf{N}^2$ and consider its coordinate maps U, V, F, G from Lemmas 2.11 and 2.13, and the corresponding functions $\hat{f}^{\pm}$ and $\hat{a}$ defined by (70). By (25), $\hat{a} \geq 0$. By (58), $\hat{a}$ has compact support, and since $\Psi^{\pm} \in L^1_{loc}(\mathbf{R}^{n-1} \times \mathbf{R}^1)$, it follows $\hat{a} \in L^1(\mathbf{R}^{n-1} \times \mathbf{R}^1)$.

Denote $B = S_p^i \cap A \cap T_{pij}^0$, where the subset S_p^i of the level set $S_p = u^{-1}(p)$ is defined in Lemma 2.11. Then B is a Borel set. From Definition 2.12 it follows that $A \cap T_{pij}^0 = \mathrm{j}(B)$. Now from (16)

$$G(A \cap T_{pij}^0) = \Big[U(B) \times \mathbf{R}^1\Big] \cap G(T_{pij}^0) \tag{74}$$

By Lemma 2.11, the set $U(B) = V^{-1}(B)$ is Borel. Thus, using Lemma 2.15, we conclude that the set $G(A \cap T_{pij}^0)$ is Borel.

Let $\varphi \in C^1(\mathbf{R}^n)$. By (23) and Area formula, we get

$$\int_{G(A\cap T_{pij}^0)} \varphi(F(X,x_n))\big[\hat{f}^+(X,x_n) - \hat{f}^-(X,x_n)\big]\, dX\, dx_n = \int_{A\cap T_{pij}^0} \varphi[f^+ - f^-]\, dz. \tag{75}$$

Since $\hat{f}^\pm := 0$ on $\mathbf{R}^n \setminus G(T_{pij}^0)$ from (23), we use (74) to rewrite (75) as

$$\int_{U(B)\times\mathbf{R}^1} \varphi(F(X,x_n))\big[\hat{f}^+(X,x_n) - \hat{f}^-(X,x_n)\big]\, dX\, dx_n = \int_{A\cap T_{pij}^0} \varphi[f^+ - f^-]\, dz. \tag{76}$$

Let c, d be the numbers from (58). Then the functions $\hat{f}_X^\pm(\tau) := \hat{f}^\pm(X,\tau)$ vanish outside the segment $[c,d]$ of finite length for a.e. $X \in \mathbf{R}^{n-1}$. Also, $\hat{f}_X^\pm \in L^1(\mathbf{R}^1)$ for a.e. $X \in \mathbf{R}^{n-1}$. Then, since F is Lipschitz and $\frac{\partial}{\partial x_n}\Psi^\pm = -\hat{f}^\pm$, we can integrate by parts with respect to x_n in the left-hand side of (76) on the segment $[c,d]$ (as in [12, §2.6.7]). Using (58) to cancel the boundary terms, and recalling the definition (70) of $\hat{a}$, we obtain

$$\int_{U(B)\times\mathbf{R}^n} \frac{\partial\varphi(F(X,x_n))}{\partial x_n}\hat{a}(X,x_n)\, dX\, dx_n = \int_{A\cap T_{pij}^0} \varphi(z)[f^+(z) - f^-(z)]\, dz. \tag{77}$$

We shall rewrite the left-hand side of (77) in a more convenient form. From Lemma 2.13(iii-iv) and (13), for any $(X,x_n) \in G(T_{pij}^0)$

$$\frac{\partial F(X,x_n)}{\partial x_n} = \nu(F(X,x_n)).$$

Also, by (24) and (26), it follows that $\Psi^+ = \Psi^-$ a.e. on $\mathbf{R}^n \setminus G(T_{pij}^0)$. We get

$$\hat{a} = \Psi^+ - \Psi^- = 0 \quad \text{a.e. on} \quad \mathbf{R}^n \setminus G(T_{pij}^0). \tag{78}$$

Thus we can integrate over the set $\Big[U(B) \times \mathbf{R}^n\Big] \cap G(T_{pij}^0)$ in the left-hand side of (77), which is the set $G(A \cap T_{pij}^0)$ by (74). Now, using (69–70) and then the Area formula [12, §3.2.5], we get

$$\begin{aligned}
&\int_{U(B)\times\mathbf{R}^1} \frac{\partial\varphi(F(X,x_n))}{\partial x_n}\hat{a}(X,x_n)\, dX\, dx_n \\
&= \int_{G(A\cap T_{pij}^0)} \left(a \sum_{i=1}^n \frac{\partial\varphi}{\partial z_i}\nu^i\right) \circ F(X,x_n) J_n F(X,x_n)\, dX\, dx_n
\end{aligned}$$

$$= \int_{T^0_{pij} \cap A} a\, \nu \cdot \nabla \varphi \, dz.$$

Thus we can write (77) in the form

$$\int_{T^0_{pij} \cap A} a\, \nu \cdot \nabla \varphi \, dz = \int_{T^0_{pij} \cap A} (f^+ - f^-) \varphi \, dz. \tag{79}$$

We sum this equality over $(p, i, j) \in \mathbf{Q} \times \mathbf{N}^2$ and, using (18), obtain (47). Note that (47) implies (46) since the right-hand side of (46) is finite for any $\varphi \in C^1(\mathbf{R}^n)$.

Step 4. Uniqueness of the transport density.
Finally, we address assertion (iii) of the Theorem. Assume $a \in L^1(\mathbf{R}^n)$ satisfies (44–47). By approximation, (47) holds for any Lipschitz φ. In particular, choosing $\varphi(z) = \Phi(u(z))$, where Φ is a Lipschitz function on $\mathbf{R}^1$, and using that $\nu \cdot \nabla u = Du\, \nu = 1$ on T^0_1 by Lemma 2.7(iii), we get

$$\int_A a(y) \Phi'(u(y))\, dy = \int_A [f^+(y) - f^-(y)] \Phi(u(y))\, dy \tag{80}$$

for any Borel transport set A.

Let $R > 0$ and Φ_R be the Lipschitz function

$$\Phi_R(\tau) = \begin{cases} 0 & \text{for } \tau < -R; \\ \dfrac{\tau}{R} + 1 & \text{for } \tau \in [-R, R]; \\ 2 & \text{for } \tau > R; \end{cases}$$

so that $\Phi_R'(\tau) = R^{-1} \chi_{[-R,R]}(\tau)$ is a step function. Let $\lambda > 0$ and $(p, i, j) \in \mathbf{Q} \times \mathbf{N}^2$. Fix $z \in T^{2\lambda}_{pij}$. Insert in (80) the function $\Phi(u(y)) = \Phi_R[u(y) - u(z)]$, and transport set

$$A = \mathrm{j}(B), \quad \text{where} \quad B = S_{z,u} \cap T^\lambda_{pij} \cap B_R(z), \tag{81}$$

where $S_{z,u}$ is the level set $\{y \mid u(y) = u(z)\}$, and j is the map (19). The set B is Borel, and thus $A = \mathrm{j}(B)$ is Borel by Corollary 2.18. We get

$$\frac{1}{R} \int_{\mathcal{D}^\lambda_R(z)} a(y)\, dy = \int_{\mathrm{j}(B)} [f^+(y) - f^-(y)] \Phi_R[u(y) - u(z)]\, dy, \tag{82}$$

where

$$\mathcal{D}^\lambda_R(z) = \mathrm{j}[S_{z,u} \cap T^\lambda_{pij} \cap B_R(z)] \cap \{y \mid u(z) - R \le u(y) \le u(z) + R\}. \tag{83}$$

Note that $\mathcal{D}^\lambda_R(z)$ is a Borel set since the set $\mathrm{j}[S_{z,u} \cap T^\lambda_{pij} \cap B_R(z)] = \mathrm{j}(B)$ is Borel.

Let us rewrite the right-hand side of (82). We have $A = \mathrm{j}(B) \subset T^0_{pij}$, since $B \subset T^0_{pij}$ by (81) and T^0_{pij} is a transport set. Thus we can make on $\mathrm{j}(B)$ the change of variables $y = F(x)$, where $x = (X, t) \in \mathbf{R}^{n-1} \times \mathbf{R}^1$, and U, V, F and G are maps from Lemma 2.13 for T^0_{pij}. Note that, by (16), $t = u(y) - p$. Also, (74) holds. Thus we obtain using (23), the Area formula, (74) and $A \subset T^0_{pij}$:

$$\begin{aligned}
&\int_{\mathrm{j}(B)} [f^+(y) - f^-(y)]\Phi_R[u(y) - u(z)]\, dy \\
&\quad = \int_{G[\mathrm{j}(B)]} [\hat f^+(X,t) - \hat f^-(X,t)]\Phi_R[t + p - u(z)]\, dX dt \\
&\quad = \int_{U(B)\times\mathbf{R}^1} [\hat f^+(X,t) - \hat f^-(X,t)]\Phi_R[t + p - u(z)]\, dX dt,
\end{aligned}$$

where we used $\hat f^\pm = 0$ on $\mathbf{R}^n \setminus G(T^0_{pij})$. Now repeating the argument given after (75) we integrate by parts with respect to t in the last expression, and obtain

$$\begin{aligned}
&\int_{\mathrm{j}(B)} [f^+(y) - f^-(y)]\Phi_R[u(y) - u(z)]\, dy \\
&\quad = \int_{U(B)\times\mathbf{R}^1} \hat a(X,t)\Phi'_R[t + p - u(z)]\, dX dt \\
&\quad = \frac{1}{R}\int_{U(B)\times[p-u(z)-R,p-u(z)+R]} \hat a(X,t)\, dX dt,
\end{aligned}$$

where $\hat a = \Psi^+ - \Psi^-$. Using (78), we can replace the domain of integration in the last expression by $G[\mathcal{D}^\lambda_R(z)] = \Big(U(B)\times[p-u(z)-R, p-u(z)+R]\Big)\cap G(T^0_{pij})$. Now, recalling (82), we recover

$$\int_{\mathcal{D}^\lambda_R(z)} a(y)\, dy = \int_{G(\mathcal{D}^\lambda_R(z))} \hat a(X,t)\, dX dt.$$

Dividing this equality by $|\mathcal{D}^\lambda_R(z)|$ and using the Area formula on the right-hand side, we get

$$\frac{1}{\mathcal{D}^\lambda_R(z)}\int_{\mathcal{D}^\lambda_R(z)} a(y)\, dy = \frac{\int_{G(\mathcal{D}^\lambda_R(z))} \hat a(X,t) dX dt}{\int_{G(\mathcal{D}^\lambda_R(z))} J_n F(X,t) dX dt}. \tag{84}$$

To pass to the limit $R \to 0+$, we need the following analog to Lemma 4.8.

Lemma 4.11 (Cylinders shrink nicely within ray clusters) *Fix $\lambda > 0$.*

i. For any $R > 0$ and $z \in \mathbf{R}^n$, $\mathcal{D}^\lambda_R(z) \subset B_{MR}(z)$ with $M = (1 + \sup_{\|e\|=1} |e|)$.

ii. There exists R_0 depending only on λ such that for any $R \in (0, R_0)$ and $z \in \mathbf{R}^n$, $\mathcal{D}^\lambda_R(z) \subset T^{\frac{\lambda}{2}}_{pij}$.

iii. Let z be a point of density 1 for the set $T^{2\lambda}_{pij}$. Then there exist M and $R_0 > 0$ such that

$$|\mathcal{D}^\lambda_R(z)| \geq \frac{1}{M} R^n \quad \textit{for all } R \in (0, R_0).$$

Proof. Assertion (i) follows from the obvious inclusion $\mathcal{D}_R^\lambda(z) \subset \mathcal{D}_R(z)$ and Lemma 4.8(i). Next choose $R_0 = \frac{\lambda}{2}$. Then (ii) follows readily from definition of $\mathcal{D}_R^\lambda(z)$ and the following property of sets T_{pij}^λ:

Claim 4.12 If $y_1 \in T_{pij}^\lambda$ and y_2 lies on the transport ray R_{y_1}, and $\|y_1 - y_2\| =: \delta < \lambda$, then $y_2 \in T_{pij}^{\lambda-\delta}$.

Proof. Since y_1 is on the $\|\cdot\|$-distance at least λ from the ends of R_{y_1}, it follows from the conditions of Claim that y_2 is on the $\|\cdot\|$-distance at least $\lambda - \delta$ from the ends of R_{y_1}. Claim 4.12 follows. □

Now we prove (iii). The proof is similar to the proof of Lemma 4.8(ii). We will first show

$$B_{\frac{R}{M}}(z) \cap T_{pij}^{2\lambda} \subset \mathcal{D}_R^\lambda(z), \tag{85}$$

for $R \in (0, R_0)$, where a small R_0 and a large M will be chosen below. Let $x \in B_{\frac{R}{M}}(z) \cap T_{pij}^{2\lambda}$. Let R_x be the unique transport ray containing x. In order to prove (85) it is enough to show that R_x intersects the level set $S_{z,u} = \{\xi \in \mathbf{R}^n \mid u(\xi) = u(z)\}$, and the point y of intersection of R_x and $S_{z,u}$ satisfies $y \in T_{pij}^\lambda$ and $|y - z| < R$.

Since $x \in T_{pij}^{2\lambda}$, for any $\tau \in (-\lambda, \lambda)$ there exists $y \in R_x \cap T_{pij}^\lambda$ with $u(y) = u(x) + \tau$. Since $x \in B_{\frac{R}{M}}(z)$, we obtain (65), and then, choosing $M \geq \sup_{|e|=1} 2\|e\|/\lambda$, we get

$$|u(x) - u(z)| \leq \frac{\lambda}{2} R.$$

Thus if $R \leq 1$ there exists $y \in R_x \cap S_{z,u}$ and $y \in T_{pij}^\lambda$ by Claim 4.12. The rest of the proof of (iii) follows the proof of Lemma 4.8(ii). Lemma 4.11 is proved. □

Finally, fix $\lambda > 0$ and $z \in T_{pij}^{2\lambda}$, so we can pass to the limit $R \to 0+$ in (84). By Lemma 4.11(i, iii) and Rudin [24, §7.9–10], the limit on the left-hand side of (84) exists and converges to $a(z)$ whenever z is a Lebesgue point of $a(\,\cdot\,)$ and a point of density 1 for the set $T_{pij}^{2\lambda}$.

Consider the right-hand side of (84). By Lemma 4.11(ii), the map G is Lipschitz on $\mathcal{D}_R^\lambda(z)$ for all $R \in (0, R_0(\lambda))$. Thus we can repeat the proof of Lemma 4.10 with use of Lemma 4.11(i, iii) to conclude that if z is a point of density 1 for the set $T_{pij}^{2\lambda}$ and $G(z)$ is a Lebesgue point for both $\hat{a}$ and $J_n F$, and if $J_n F(G(z)) \neq 0$, then the limit on the right-hand side of (84) exists and converges to $\hat{a}(G(z))/J_n F(G(z))$. Since the map F is Lipschitz on $\mathbf{R}^n$, and G is Lipschitz on $T_{pij}^{2\lambda}$, and (73) holds, a.e. $z \in T_{pij}^{2\lambda}$ satisfies these conditions.

Thus for any $\lambda > 0$, for a.e. $z \in T_{pij}^{2\lambda}$

$$a(z) = \frac{\hat{a}(G(z))}{J_n F(G(z))} = \frac{\Psi^+(G(z)) - \Psi^-(G(z))}{J_n F(G(z))}. \tag{86}$$

Since $T_{pij}^0 = \bigcup_{k=1}^\infty T_{pij}^{\frac{1}{k}}$, the formula (86) holds a.e. on T_{pij}^0. Uniqueness of $a(z)$, satisfying (44–47) follows, and Theorem 4.1 is proved. □

References

1. G. Alberti, B. Kircheim and D. Preiss. Presented in a lecture by Kircheim at the Scuola Normale Superiori workshop, October 27, 2000. See also [2, Remark 6.1].
2. L. Ambrosio. Lecture notes on optimal transport problems. To appear in the CIME Series of Springer Lecture Notes from the Madiera 2000 Summer School on “Mathematical Aspects of Evolving Interfaces”.
3. K. Ball, E.A. Carlen, and E.H. Lieb. Sharp uniform convexity and smoothness inequalities for trace norms. Invent. Math. **115** (1994) 463–482.
4. M. Beckmann. A continuous model of transportation. Econometrica **20** (1952) 643–660.
5. G. Bouchitte, G. Buttazzo, P. Seppecher, Shape optimization solutions via Monge-Kantorovich equation. C. R. Acad. Sci. Paris Ser. I Math. **324** (1997), no. 10, 1185–1191.
6. L. Caffarelli, M. Feldman, R.J. McCann. Constructing optimal maps for Monge's transport problem as a limit of strictly convex costs. To appear in J. Amer. Math. Soc.
7. L. de Pascale and A. Pratelli. Regularity properties for Monge transport density and for solutions of some shape optimization problem. To appear in Calc. Var.
8. L.C. Evans. Partial differential equations and Monge-Kantorovich mass transfer. In R. Bott et al., editor, Current Developments in Mathematics, pages 26–78. International Press, Cambridge, 1997.
9. L.C. Evans and W. Gangbo. Differential equations methods for the Monge-Kantorovich mass transfer problem. Mem. Amer. Math. Soc. **137** (1999) 1–66.
10. L.C. Evans, R.F. Gariepy. Measure theory and fine properties of functions. CRC Press, 1992.
11. L.C. Evans, M. Feldman and R.F. Gariepy. Fast/slow diffusion and collapsing sandpiles. J. Differential Equations **137** (1997) 166–20.
12. H. Federer. Geometric Measure Theory. Springer-Verlag, New York, 1969.
13. M. Feldman. Variational evolution problems and nonlocal geometric motion. Arch. Rat. Mech. Anal. **146** (1999) 221–274.
14. M. Feldman. Growth of a sandpile around an obstacle. In L.A. Caffarelli and M. Milman, editors, Monge Ampère Equation: Applications to Geometry and Optimization (Deerfield Beach, FL, 1997), number 226 in Contemp. Math., pages 55–78. American Mathematical Society, Providence, 1999.
15. M. Feldman, R.J. McCann. Monge transport problem on a Riemannian manifold. Preprint.
16. W. Gangbo and R.J. McCann. The geometry of optimal transportation. Acta Math. **177** (1996) 113–161.
17. M. Iri. Theory of flows in continua as approximation to flows in networks. Survey of mathematical programming (Proc. Ninth Internat. Math. Programming Sympos., Budapest, 1976), Vol. 2, pp. 263–278, North-Holland, Amsterdam-Oxford-New York, 1979.
18. U. Janfalk On a minimization problem for vector fields in L^1. Bull. London Math. Soc. **28** (1996), no. 2, 165–176.
19. L. Kantorovich. On the translocation of masses. C.R. (Doklady) Acad. Sci. URSS (N.S.) **37** (1942) 199–201.
20. R.J. McCann. Existence and uniqueness of monotone measure-preserving maps. Duke Math. J. **80** (1995) 309–323.
21. R.J. McCann. Exact Solutions to the Transportation Problem on the Line. R. Soc. Lond. Proc. Ser. A Math. Phys. Eng. Sci. **455** (1999) 1341–1380.
22. G. Monge. Mémoire sur la théorie des déblais et de remblais. Histoire de l'Académie Royale des Sciences de Paris, avec les Mémoires de Mathématique et de Physique pour la même année, pages 666–704, 1781.
23. S.T. Rachev and L. Rüschendorf. Mass Transportation Problems. Probab. Appl. Springer-Verlag, New York, 1998.
24. W. Rudin. Real and complex analysis. Third edition. McGraw-Hill Book Co., New York, 1987.

25. E. Stepanov. Lecture at the Scuola Normale Superiori workshop, October 26, 2000.
26. G. Strang. L^1 and L^∞ approximation of vector fields in the plane. Nonlinear partial differential equations in applied science (Tokyo, 1982), pages 273–288, North-Holland Math. Stud., 81, North-Holland, Amsterdam-New York, 1983.
27. G. Strang. Maximal flow through a domain. Math. Programming **26** (1983), no. 2, 123–143.
28. V.N. Sudakov. Geometric problems in the theory of infinite-dimensional probability distributions. Proc. Steklov Inst. Math. **141** (1979) 1–178.
29. C. Villani. Topics in mass transportation. In preparation.
30. N.S. Trudinger, X.-J. Wang. On the Monge mass transfer problem. Calc. Var. **13** (2001) 19–31.