JOURNAL OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 00, Number 0, Xxxx XXXX, Pages 000–000
S 0894-0347(XX)0000-0

CONSTRUCTING OPTIMAL MAPS FOR MONGE'S TRANSPORT PROBLEM AS A LIMIT OF STRICTLY CONVEX COSTS

LUIS A. CAFFARELLI, MIKHAIL FELDMAN, AND ROBERT J. MCCANN

The Monge-Kantorovich problem is to move one distribution of mass onto another as efficiently as possible, where Monge's original criterion for *efficiency* [19] was to minimize the average distance transported. Subsequently studied by many authors, it was not until 1976 that Sudakov showed solutions to be realized in the original sense of Monge, i.e., as mappings from $\mathbf{R}^n$ to $\mathbf{R}^n$ [23]. A second proof of this existence result formed the subject of a recent monograph by Evans and Gangbo [7], who avoided Sudakov's measure decomposition results by using a partial differential equations approach. In the present manuscript, we give a third existence proof for optimal mappings, which has some advantages (and disadvantages) relative to existing approaches: it requires no continuity or separation of the mass distributions, yet our explicit construction yields more geometrical control than the abstract method of Sudakov. (Indeed, this control turns out to be essential for addressing a gap which has recently surfaced in Sudakov's approach to the problem in dimensions $n \geq 3$; see the remarks at the end of this section.) It is also shorter and more flexible than either, and can be adapted to handle transportation on Riemannian manifolds or around obstacles, as we plan to show in a subsequent work [13]. The problem considered here is the classical one:

Received by the editors February 20, 2000.
2000 *Mathematics Subject Classification.* Primary 49Q20; Secondary 26B10, 28A50, 58E17, 90B06.
Key words and phrases Monge-Kantorovich mass transportation, resource allocation, optimal map, optimal coupling, infinite dimensional linear programming, dual problem, Wasserstein distance.
This research was supported by grants DMS 9714758, 9623276, 9970577, and 9622997 of the US National Science Foundation, and grant 217006-99 RGPIN of the Natural Sciences and Engineering Research Council of Canada. The hospitality of the Max-Planck Institutes at Bonn and Leipzig are gratefully acknowledged by MF and RJM respectively.

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

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

In the set $\mathcal{A}(\mu^+,\mu^-)$ 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 $r \in \mathcal{A}(\mu^+,\mu^-)$ 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 assume there exist constants $\Lambda, \lambda > 0$ such that all $x, y \in \mathbf{R}^n$ satisfy the uniform smoothness and convexity estimates:

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

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. For a further discussion of Monge's problem, its history, and applications, we refer the reader to Evans [5], Evans and Gangbo [7], Gangbo and McCann [15] or Rachev and Rüschendorf [20].

Part of the difficulty of this problem is the degeneracy which results from failure of the norm to be strictly convex (radially). Even in the simplest one-dimensional examples this leads to non-uniqueness of the minimizing map. By contrast, when the transportation cost function $\|y-x\|$ is replaced by a strictly convex function such as $\|y-x\|^p$ with $p > 1$, the problem simplifies considerably, and following ideas of Brenier and others it is possible to show that a unique map solves the problem and possesses nice measurability properties in both the Euclidean [4][15] and Riemannian [18] settings.

In this paper, our key idea for resolving this degeneracy is to first find a change of coordinates which adapts the problem's local geometry so that all transport directions become parallel, and then solve these one-dimensional transportation problems separately before invoking Fubini's theorem to complete the proof. The map we construct in this way might, in principle, be recovered in the $p \to 1$ limit from the unique maps solving the $p > 1$ problems. Although we don't carry out this limit directly, we do use structural features of the optimal maps for $p > 1$ to facilitate several aspects of the proof. This distinguishes our solution from that of Evans and Gangbo, as illustrated by the book-shifting example $f^+ = \chi_{[0,n]}$ and $f^- = \chi_{[1,n+1]}$ on the line $\mathbf{R}^1$ [15]. For this example, the map we construct is the translation $s(x) = x + 1$, whereas Evans and Gangbo would leave the mass common to f^+ and f^- in its place a priori, obtaining the map $s(x) = x$ on $x \in [1, n]$ and $s(x) = x + n$ on $[0, 1]$ as a result. We anticipate that the the ability to deal with overlapping densities f^+ and f^- will be significant in applications. We also point out that a sequel shows the map constructed below is the only optimal map to preserve the ordering of pairs of points with collinear images [12].

We now give a heuristic outline of our existence proof. Following previous authors, we begin by solving a dual problem whose solution defines the set of *transport rays*, according to the terminology of Evans and Gangbo [7]. These rays are determined by the property that the Lipschitz potential $u : \mathbf{R}^n \longrightarrow \mathbf{R}$ from the dual problem decreases along them with maximum admissible rate. As we show below, the optimal mapping s takes each transport ray into itself. We therefore restrict the measures μ^+ and μ^- to each ray, so that mass balance holds for the restrictions, and solve a transportation problem on each ray. These one-dimensional problems are easy to solve. Thus we get an optimal map on each ray, and as the result a map from $\mathbf{R}^n$ into $\mathbf{R}^n$. We show then that this map pushes μ^+ forward to μ^-, and is optimal.

The most delicate step in this procedure involves restricting the measures to rays, and it is here that our approach diverges from Sudakov's. Instead of building on the measure decomposition results of Halmos [16] or Rokhlin [21], we seek a local change of variables in $\mathbf{R}^n$ so that the new coordinate x_n measures distance along each ray, while the remaining $n - 1$ coordinates vary across nearby rays. For the Euclidean norm on $\mathbf{R}^n$ the directions of rays are given by the gradient of Monge's potential u, and thus it is natural to use level sets of u to parametrize rays, i.e., the variables $x_1, \ldots, x_{n-1}$ will be coordinates on a fixed level set of u. This can also be adapted to more general norms, if one defines the gradient of u using the appropriate (Finsler) identification of vectors with one-forms. But we also need certain properties of this change of variables in order to be able to express μ^+ and μ^- in the new coordinates: Indeed, expressing these measures is tantamount to changing variables under the integral, and the change of variables must be Lipschitz continuous to apply the Area formula. However, the typical pattern of rays is too complicated for us to achieve this globally. We therefore decompose the set of all rays into a countable collection of special subsets, chosen so that the rays enjoy a more "regular" structure within each subset while the mass of μ^+ still balances μ^-, and perform a Lipschitz change of variables on each subset separately. Thus the Lipschitz control on directions of rays given by Lemma 16 is absolutely crucial to our proof. The estimate which provides this control is a restriction on the geometry of quadrilaterals in a smooth, uniformly convex Banach space; established in Lemma 14, this estimate holds some independent interest; c.f. Federer [10, §4.8(8)] and Feldman [11, Appendix A] .

The remainder of this paper is organized as follows. In the first section we recall the general duality theory for Monge-type problems introduced by Kantorovich [17], and the construction of optimal maps for transportation costs given by strictly convex functions instead of a norm [4] [15]. The Kantorovich dual problem is solved by taking a limit of such costs, and the section concludes with a criterion for optimality. It is followed by a section which introduces the transport rays and geometry dictated by the Kantorovich solution and criterion for optimality. Several observations by Evans and Gangbo are summarized here, followed by our key new estimate giving Lipschitz control on the directions of transport rays. In the third section we construct the local changes of variables which parallize nearby transport rays, while the fourth section verifies that the traces of f^+ and f^- — weighted by a Jacobian factor accounting for the change of variables — are balanced on each individual ray. Finally, these ingredients are combined in Section 5 to give a proof of our main theorem by constructing a map solving Monge's problem:

Theorem 1 (Existence 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). Then there exists a Borel map $s : \mathbf{R}^n \longrightarrow \mathbf{R}^n$ which solves Monge's problem, in the sense that it minimizes the average distance (2) transported among all maps pushing f^+ forward to f^- (3).*

Remarks added in revision: After the submission of this manuscript, the authors learned of several significant developments. Foremost among them is the concurrent but independent discovery of a existence result for optimal maps in Monge's problem by Trudinger and Wang [24]. Although very similar to our approach, the argument there is streamlined somewhat by their decision to focus exclusively on the Euclidean norm.

Lecture notes subsequently released by Ambrosio [2] contain an excellent summary of progress on Monge's problem cast into the framework of geometric measure theory; they include a derivation for existence and uniqueness of a transport density corresponding to $f^\pm \in L^1(\mathbf{R}^n)$ which parallels certain results of Feldman and McCann [12]. Ambrosio's notes also highlight a logical gap in the solution of Monge's problem proposed by Sudakov. Without Lipschitz control on ray directions, it is impossible to know that the conditional restriction of $f^\pm$ yields an absolutely continuous measure along almost every ray, as required for Sudakov's proof. In two dimensions, disjointness prevents nearby rays from turning too much without bumping into each other, so Lipschitz control is automatic and the gap can be bridged (as long as the norm has a strictly convex unit ball). But a counterexample in $\mathbf{R}^3$ due to Alberti, Kircheim, and Preiss [1] shows an uncountable union of disjoint segments can be constructed, whose midpoints form the support of an absolutely continuously probability measure on $\mathbf{R}^3$. The restriction of this measure to each of the segments yields a Dirac mass at its midpoint, violating the absolute continuity claimed by Sudakov. In this context the geometrical control provided by Lemma 16 is required to preclude such a collection of segments from forming transport rays in Monge's problem. Thus it would seem that Evans and Gangbo [7] contains the the first complete proof of existence for optimal maps between Lipschitz densities $f^\pm$ with disjoint support, while the present manuscript and Trudinger-Wang [24] complete the first 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 [23, p 164].

1. Duality in the limit of strictly convex costs

In this section we recall a problem formulated by Kantorovich as a dual to Monge's problem. We construct its solution, and extract properties germane to our purposes.

Consider $\mathbf{R}^n$ metrized by a norm $\|\cdot\|$ satisfying the uniform smoothness and convexity conditions (4), and denote the associated distance by $d(x,y) := \|x-y\|$. Then the problem asserted by Kantorovich [17] to be dual to Monge's problem is formulated as follows. Let $\mathrm{Lip}_1(\mathcal{X}, d)$ denote the set of functions on $\mathcal{X} \subset \mathbf{R}^n$ which are Lipschitz continuous with Lipschitz constant no greater than one; thus

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

Problem 2 (Kantorovich). *For $f^+, f^- \in L^1(\mathbf{R}^n)$ from Monge's Problem 1, maximize $\hat{K}[v]$ on $Lip_1(\mathbf{R}^n, d)$, where*

$$\hat{K}[v] := \int_{\mathbf{R}^n} (vf^+ - vf^-)\, dx.$$

To solve the Monge and Kantorovich problems, we consider a second pair of dual problems in which the metric $d(x, y)$ is replaced by a more general transportation cost function $c_\varepsilon(x, y)$ on $\mathbf{R}^n \times \mathbf{R}^n$. The Monge problem analogous to (1) then becomes:

Problem 3 (Primal). *Fix two Borel densities $f^+, f^- \geq 0$ in $L^1(\mathbf{R}^n)$ with compact support satisfying the mass balance condition (1). Among Borel maps $r \in \mathcal{A}(\mu^+, \mu^-)$ which pushes the measure $d\mu^+ = f^+(x)dx$ forward to $d\mu^- = f^-(y)dy$ as in (3), find a map $s : \mathbf{R}^n \longrightarrow \mathbf{R}^n$ which minimizes the total transportation cost*

$$I_\varepsilon[r] := \int_{\mathbf{R}^n} c_\varepsilon(x, r(x)) f^+(x) dx. \tag{5}$$

The corresponding dual problem is:

Problem 4 (Dual). *Take f^+, $f^- \in L^1(\mathbf{R}^n)$ as in Problem 3. Denote the support of f^+ by $\mathcal{X}$ and f^- by $\mathcal{Y}$. Among all pairs of continuous functions φ, ψ in*

$$\begin{aligned} J_\varepsilon(\mathcal{X}, \mathcal{Y}) \;:=\; & \{(\varphi, \psi) \in C(\mathcal{X}) \times C(\mathcal{Y}) \mid \\ & \varphi(x) + \psi(y) \geq -c_\varepsilon(x, y) \text{ on } \mathcal{X} \times \mathcal{Y}\} \end{aligned} \tag{6}$$

find a pair $(\varphi_\varepsilon, \psi_\varepsilon)$ minimizing the functional

$$K(\varphi, \psi) := \int_{\mathcal{X}} \varphi f^+\, dx + \int_{\mathcal{Y}} \psi f^-\, dy. \tag{7}$$

The duality assertion $I_\varepsilon[s] = -K(\varphi_\varepsilon, \psi_\varepsilon)$ which relates these two problems holds rather generally; see Rachev and Rüschendorf [20]. However, the Dual Problem 4 takes a fundamentally different form than the Kantorovich problem, due to the fact that the cost $c_\varepsilon(x, y)$ need no longer satisfy a triangle inequality. This generalization is useful, since it permits us to replace the distance function $d(x, y) = \|x-y\|$ by a strictly convex cost function

$$c_\varepsilon(x, y) := h_\varepsilon(x - y) = \|x - y\|^{1+\varepsilon}, \tag{8}$$

for which existence, uniqueness, and a characterization of optimal maps in Monge's problem can be found in Caffarelli [4] and Gangbo and McCann [14] [15]. Noting that $h_\epsilon(x)$ is $C^{1,\varepsilon}(\mathbf{R}^n)$ smooth and strictly convex from Lemma 11 below, we recall the relevant results as follows:

Theorem 2 (Duality and Optimal Maps for Strictly Convex Costs [4][14])**.** *Take $f^+, f^- \in L^1(\mathbf{R}^n)$ and $\mathcal{X} := \operatorname{spt} f^+$ and $\mathcal{Y} := \operatorname{spt} f^-$ as in Problem 3. If the transportation cost $c_\varepsilon(x,y)$ satisfies (8) and (4) then for $\varepsilon > 0$:*

i. *Some pair $(\varphi_\varepsilon, \psi_\varepsilon)$ minimizing $K(\varphi,\psi)$ on $J_\varepsilon(\mathcal{X},\mathcal{Y})$ in the dual problem satisfies*

$$\varphi_\varepsilon(x) = \sup_{y\in\mathcal{Y}}(-c_\varepsilon(x,y) - \psi_\varepsilon(y)); \tag{9}$$

$$\psi_\varepsilon(y) = \sup_{x\in\mathcal{X}}(-c_\varepsilon(x,y) - \varphi_\varepsilon(x)). \tag{10}$$

ii. *The function φ_ε is Lipschitz on $\mathcal{X}$ (as ψ_ε is on $\mathcal{Y}$), with Lipschitz constant dominated by the Lipschitz constant of $c_\varepsilon(x,y)$ on $\mathcal{X}\times\mathcal{Y}$.*

iii. *For a.e. $x \in \mathcal{X}$ there exists a unique $y \in \mathcal{Y}$ such that*

$$\varphi_\varepsilon(x) + \psi_\varepsilon(y) = -c_\varepsilon(x,y). \tag{11}$$

iv. *Define the mapping $s_\varepsilon : \mathcal{X} \to \mathcal{Y}$ by assigning to a.e. $x \in \mathcal{X}$ the unique $y \in \mathcal{Y}$ for which (11) holds. Then s_ε pushes the measure $d\mu^+ = f^+(x)dx$ forward to $d\mu^- = f^-(y)dy$ and is the unique minimizer for the primal Problem 3.*

Our first goal is to extract a pair of functions minimizing $K(\varphi,\psi)$ on $J_0(\mathcal{X},\mathcal{Y})$ from the limit $\varepsilon \to 0$ of this theorem. This follows from a simple compactness result:

Proposition 3 (Limit of Minimizing Pairs)**.** *For some sequence $\varepsilon_j > 0$ which tends to zero, the Dual Problem 4 admits a sequence of pairs $(\varphi_{\varepsilon_j}, \psi_{\varepsilon_j})$ which minimize $K(\varphi,\psi)$ on $J_{\varepsilon_j}(\mathcal{X},\mathcal{Y})$ and converge uniformly on the compact sets $\mathcal{X}$ and $\mathcal{Y}$ respectively to limits $\varphi_{\varepsilon_j} \to \varphi_0$ and $\psi_{\varepsilon_j} \to \psi_0$ as $j \to \infty$. The limit functions φ_0, ψ_0 minimize $K(\varphi,\psi)$ on $J_0(\mathcal{X},\mathcal{Y})$ and satisfy*

$$\varphi_0(x) = \sup_{y\in\mathcal{Y}}(-\|x-y\| - \psi_0(y)); \tag{12}$$

$$\psi_0(y) = \sup_{x\in\mathcal{X}}(-\|x-y\| - \varphi_0(x)). \tag{13}$$

Proof. Fix $x_0 \in \mathcal{X}$ and observe that $K(\varphi,\psi) = K(\varphi - A, \psi + A)$ for each $A \in \mathbf{R}^1$ according to the mass balance condition (1). Thus any pair $(\varphi_\varepsilon, \psi_\varepsilon)$ minimizing $K(\varphi,\psi)$ on $J_\varepsilon(\mathcal{X},\mathcal{Y})$ may be shifted by $A = \varphi_\varepsilon(x_0)$ to ensure $\varphi_\varepsilon(x_0) = 0$.

Now $\mathcal{X}$ and $\mathcal{Y}$ are compact, so for $\varepsilon \in (0,1)$ the costs $c_\varepsilon(x,y) = \|x-y\|^{1+\varepsilon}$ form an equi-Lipschitz family on $\mathcal{X}\times\mathcal{Y}$. The minimizing functions φ_ε and ψ_ε in Theorem 2(ii) also form equi-Lipschitz families on $\mathcal{X}$ and $\mathcal{Y}$ respectively. Moreover $\varphi_\varepsilon(x_0) = 0$, so the functions φ_ε are uniformly bounded on $\mathcal{X}$. Also, $|c_\varepsilon(x,y)| \le C$ for all $(x,y,\epsilon) \in$

$\mathcal{X} \times \mathcal{Y} \times (0,1)$, implying a uniform bound on the ψ_ε in (9). The Ascoli-Arzela theorem then yields a subsequence $\varepsilon_j \to 0$ such that φ_{ε_j} and ψ_{ε_j} converge uniformly on $\mathcal{X}$ and $\mathcal{Y}$ respectively to $\varphi_0 \in C(\mathcal{X})$ and $\psi_0 \in C(\mathcal{Y})$; (12–13) follow from (9–10) and imply that $(\varphi_0, \psi_0) \in J_0(\mathcal{X}, \mathcal{Y})$.

It remains to show that (φ_0, ψ_0) minimizes $K(\varphi, \psi)$ on $J_0(\mathcal{X}, \mathcal{Y})$. For any other pair $(\tilde{\varphi}, \tilde{\psi}) \in J_0(\mathcal{X}, \mathcal{Y})$ and $\varepsilon > 0$, define $\tilde{\varphi}_\varepsilon \equiv \tilde{\varphi}$ and

$$\tilde{\psi}_\varepsilon(y) = \tilde{\psi}(y) + \max_{x \in \mathcal{X}}[-c_\varepsilon(x,y) + c(x,y)].$$

Compactness of $\mathcal{X}$ yields $\tilde{\psi}_\varepsilon \in C(\mathcal{Y})$, while $(\tilde{\varphi}_\varepsilon, \tilde{\psi}_\varepsilon) \in J_\varepsilon(\mathcal{X}, \mathcal{Y})$ from their definition and (6). Moreover, $\tilde{\psi}_\varepsilon \to \tilde{\psi}$ uniformly on $\mathcal{Y}$ as $\varepsilon \to 0$. For each j these competitors satisfy $K(\varphi_{\varepsilon_j}, \psi_{\varepsilon_j}) \leq K(\tilde{\varphi}_{\varepsilon_j}, \tilde{\psi}_{\varepsilon_j})$. Uniform convergence yields $K(\varphi_0, \psi_0) \leq K(\tilde{\varphi}, \tilde{\psi})$ in the limit $j \to \infty$, so the proposition is proved. □

Next we demonstrate equivalence of the Dual Problem 4 in the case $\epsilon = 0$ to the Kantorovich Problem 2 via the triangle inequality; see (14) and (17) especially.

Proposition 4 (Lipschitz Maximizer). *Suppose (φ_0, ψ_0) satisfy (12–13) and minimize $K(\varphi, \psi)$ on $J_0(\mathcal{X}, \mathcal{Y})$. Then there exists $u \in Lip_1(\mathbf{R}^n, d)$ such that*

$$u = -\varphi_0 \quad on \ \mathcal{X}, \quad u = \psi_0 \quad on \ \mathcal{Y}. \tag{14}$$

Moreover, u maximizes $\hat{K}[v]$ on $Lip_1(\mathbf{R}^n, d)$ and satisfies

$$\begin{aligned} u(x) &= \min_{y \in \mathcal{Y}} (u(y) + \|x - y\|) \qquad for\ any\ \ x \in \mathcal{X}; \\ u(y) &= \max_{x \in \mathcal{X}} (u(x) - \|x - y\|) \qquad for\ any\ \ y \in \mathcal{Y}. \end{aligned} \tag{15}$$

Proof. Extend φ_0, ψ_0 to the whole space $\mathbf{R}^n$ using the right-hand sides of (12–13). We show first that $\varphi_0, \psi_0 \in \mathrm{Lip}_1(\mathbf{R}^n, d)$. Indeed, let $x_1, x_2 \in \mathbf{R}^n$. Continuity of ψ_0 on the compact set $\mathcal{Y}$ yields a point $y_1 \in \mathcal{Y}$ where the supremum (12) is attained: $\varphi_0(x_1) = -\|x_1 - y_1\| - \psi_0(y_1)$. Also (12) implies $\varphi_0(x_2) \geq -\|x_2 - y_1\| - \psi_0(y_1)$. Thus

$$\varphi_0(x_1) - \varphi_0(x_2) \leq -\|x_1 - y_1\| + \|x_2 - y_1\| \leq \|x_1 - x_2\|$$

by the triangle inequality. Thus $\varphi_0 \in \mathrm{Lip}_1(\mathbf{R}^n, d)$, and $\psi_0 \in \mathrm{Lip}_1(\mathbf{R}^n, d)$ similarly.

Next we show that $\varphi_0 + \psi_0 = 0$ on $\mathcal{X}$. For any $x \in \mathcal{X}$, (12) yields

$$\varphi_0(x) + \psi_0(x) \geq 0 \text{ on } \mathcal{X}. \tag{16}$$

Suppose for some $z \in \mathcal{X}$ a strict inequality holds: $\varphi_0(z) + \psi_0(z) > 0$. By (12–13) and continuity of φ_0 and ψ_0, there exist $x \in \mathcal{X}$ and $y \in \mathcal{Y}$

such that

$$\varphi_0(z) = -\|z-y\| - \psi_0(y),$$
$$\psi_0(z) = -\|z-x\| - \varphi_0(x).$$

Combined with $\varphi_0(z)+\psi_0(z) > 0$ and $(\varphi_0, \psi_0) \in J_0(\mathcal{X}, \mathcal{Y})$ this implies

$$\|z-y\|+\|z-x\| = -\varphi_0(z)-\psi_0(z)-\varphi_0(x)-\psi_0(y) < -\varphi_0(x)-\psi_0(y) \le \|x-y\|,$$

contradicting the triangle inequality. Thus $\varphi_0 + \psi_0 \le 0$ on $\mathcal{X}$. In conjunction with (16) this yields $\varphi_0 + \psi_0 = 0$ on $\mathcal{X}$ as desired.

Thus, denoting $u = \psi_0$ in $\mathbf{R}^n$ we have shown $u \in \mathrm{Lip}_1(\mathbf{R}^n, d)$ and both parts of (14). Also, (15) follows directly from (12–13). It remains to prove u maximizes $\hat{K}[v]$ in the Kantorovich Problem 2.

Note that

$$\hat{K}[u] = -K[\varphi_0, \psi_0] \tag{17}$$

by (14). Let $v \in Lip_1(\mathbf{R}^n, d)$. Then the pair $\hat{\varphi}, \hat{\psi}$ defined by

$$\hat{\varphi} = -v \quad \text{on} \quad \mathcal{X}, \quad \text{and} \quad \hat{\psi} = v \quad \text{on} \quad \mathcal{Y}$$

belongs to the set $J_0(\mathcal{X}, \mathcal{Y})$ defined in (6): indeed, for $(x, y) \in \mathcal{X} \times \mathcal{Y}$ we have

$$\hat{\varphi}(x) + \hat{\psi}(y) = -v(x) + v(y) \ge -\|x-y\|.$$

Now

$$\hat{K}[u] = -K[\varphi_0, \psi_0] \ge -K[\hat{\varphi}, \hat{\psi}] = \hat{K}[v],$$

where the last equality follows from definition of $\hat{\varphi}, \hat{\psi}$, and the proposition is proved. ☐

Definition 5 (Kantorovich Potentials). *Any function u which maximizes $\hat{K}[v]$ on $Lip_1(\mathbf{R}^n, d)$ may be referred to as a* Kantorovich potential. *Such potentials exist by Propositions 3–4. However, the Kantorovich potentials obtained in this way — via a limit (φ_0, ψ_0) of pairs from Theorem 2(i) — have additional virtues, (14–15). We call such u a* limiting Kantorovich potential *and exploit its existence hereafter.*

Finally, we discuss the connection between the primal and dual problems. For the strictly convex costs (8) this connection is given by Theorem 2, which shows how the primal problem can be solved using a solution to the dual problem. However, for the nonstrictly cost $c_0(x, y)$ the uniqueness assertion of Theorem 2(iii) would fail, so the corresponding map is not well-defined: its direction is clear, but its distance ambiguous. Indeed, when a minimizing pair (φ_0, ψ_0) for $K(\varphi, \psi)$ satisfies (12–13) and $\varphi_0(x)+\psi_0(y) = \|x-y\|$ holds for some $(x, y) \in \mathcal{X} \times \mathcal{Y}$, we shall see $\varphi_0(x) + \psi_0(z) = \|x-z\|$ for all $z \in [x, y] \cap \mathcal{Y}$, meaning all $z = tx + (1-t)y \in \mathcal{Y}$ with $t \in [0, 1]$.

The next lemma exhibits the connection between the primal and dual problems for the cost function $c_0(x,y) = \|x - y\|$. It shows in particular that to obtain an optimal map in the primal problem, it is sufficient to start from a Kantorovich potential u and construct any admissible map consistent with (18). The rest of this paper is devoted to carrying out this program on $\mathbf{R}^n$, suitably normed.

Lemma 6 (Dual Criteria for Optimality)**.** *Fix $u \in Lip_1(\mathbf{R}^n, d)$ and let $s : \mathbf{R}^n \to \mathbf{R}^n$ be a mapping which pushes μ^+ forward to μ^-. If*

$$u(x) - u(s(x)) = \|x - s(x)\| \quad \text{for } \mu^+ \text{ a.e. } x \in \mathcal{X} \tag{18}$$

then:

i. *u is a Kantorovich potential maximizing Problem 2.*
ii. *s is an optimal map in Problem 1.*
iii. *The infimum $I[s]$ in Problem 1 is equal to the supremum $\hat{K}[u]$ in Problem 2.*
iv. *Every optimal map $\hat{s}$ and Kantorovich potential $\hat{u}$ also satisfy (18).*

Proof. For any map $r : \mathbf{R}^n \to \mathbf{R}^n$ pushing forward μ^+ to μ^+ and $v \in \mathrm{Lip}_1(\mathbf{R}^n, d)$ we compute:

$$\begin{aligned} I[r] &= \int_{\mathbf{R}^n} \|x - r(x)\| d\mu^+(x) \\ &\geq \int_{\mathbf{R}^n} [v(x) - v(r(x))] d\mu^+(x) \\ &= \int_{\mathbf{R}^n} v(x) d\mu^+(x) - \int_{\mathbf{R}^n} v(y) d\mu^-(y) \\ &= \hat{K}[v]. \end{aligned} \tag{19}$$

using (3). Thus the minimum value of $I[r]$ on $\mathcal{A}(\mu^+, \mu^-)$ is at least as large as the maximum of $\hat{K}[v]$ on $\mathrm{Lip}_1(\mathbf{R}^n, d)$. On the other hand, our hypothesis (18) produces a case of equality $I[s] = \hat{K}[u]$ in (19). This implies the assertions (i) $\hat{K}[u]$ is a maximum; (ii) $I[s]$ is a minimum; and (iii) $I[s] = \hat{K}[u]$ of the lemma.

Now let $r \in \mathcal{A}(\mu^+, \mu^-)$ and $v \in \mathrm{Lip}_1(\mathbf{R}^n, d)$ be any other optimal map and Kantorovich potential. Then $I[r] = I[s]$ and $\hat{K}[v] = \hat{K}[u]$ combine with (iii) to yield $I[r] = \hat{K}[v]$. But this implies a pointwise equality μ^+ almost everywhere in (19), so the proof of assertion (iv) and hence the lemma, are complete. ☐

2. Transport rays and their geometry

The preceding section reduced the problem of finding an optimal map in Monge's problem to constructing an admissible map which also satisfies (18). We carry out this program on $\mathbf{R}^n$ metrized by the norm $d(x,y) = \|x-y\|$. Our starting point is a Kantorovich potential $u \in \mathrm{Lip}_1(\mathbf{R}^n, \|\cdot\|)$. In this section, we study the geometric meaning of condition (18), and introduce the transport rays and transport sets which are ultimately used to construct an optimal map. We study the properties of transport rays, in particular proving a Lipschitz estimate for how much the direction of nearby rays can vary if none of the rays are too short. The underlying idea is that smoothness and uniform convexity of the norm ball (4) impose geometrical constraints on each quadrilateral whose opposite sides are formed by transport rays. This estimate is much in the spirit of Federer's theorem on Euclidean distance functions [10, §4.8(8)]; see also Feldman [11, Appendix A]

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. Through §§2–5, we fix a limiting Kantorovich potential u — a maximizer in Problem 2 obtained from a limit of solutions to dual problems with strictly convex costs $c_\varepsilon(x,y) = \|x-y\|^{1+\varepsilon}$. Such a potential exists and satisfies (15) by Propositions 3–4 and Definition 5. Note that u has Lipschitz constant one with respect to the distance $d(x,y) = \|x-y\|$. The derivative of any function $\varphi : \mathbf{R}^n \to \mathbf{R}^1$ at $x \in \mathbf{R}^n$ — viewed as a linear functional on the tangent space — is denoted by $D\varphi(x) \in (\mathbf{R}^n)^*$.

Since we want to investigate the geometrical implications of (18) for u, suppose $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ satisfy

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

From the Lipschitz constraint

$$|u(z_1) - u(z_2)| \leq \|z_1 - z_2\| \quad \text{for any} \quad z_1, z_2 \in \mathbf{R}^n, \tag{20}$$

it follows that on the segment connecting x and y the function u is affine and decreasing with the maximum rate compatible with (20). We will call maximal segments $[x,y]$ having these properties the *transport rays*. More precisely:

Definition 7 (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 (20) that any point $z \in R$ satisfies

$$u(z) = u(b) + \|z - b\| = u(a) - \|a - z\|. \tag{21}$$

Definition 8 (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*

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

From these two definitions and the property (15) of u we immediately infer the following lemma, whose obvious proof is omitted.

Lemma 9 (Data is Supported Only on Transport Rays). *$\mathcal{X}\cup\mathcal{Y} \subseteq T_0\cup T_1$.*

To study the properties of rays, 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 $[a, b]$. The basic observation which goes back to Monge is that transport rays do not cross.

Lemma 10 (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.*

Proof. First note the strict convexity of the unit ball $\|x\| \leq 1$ asserted in Lemma 11 implies that equality

$$\|x - y\| + \|y - z\| = \|x - z\|$$

holds if and only if y lies on the segment $[x, z]$.

Since $R_1 \neq R_2$ share the point c, they cannot be collinear: otherwise (21) and the maximality part of Definition 7 would force $R_1 = R_2$. Thus the two rays can only intersect in a single point: $R_1 \cap R_2 = \{c\}$. It remains to prove either $c = a_1 = a_2$ or $c = b_1 = b_2$, where a_k denotes the upper end and b_k the lower end of R_k, $k = 1, 2$.

We shall assume $c \neq b_2$ and argue this forces $c = a_1$. It then follows that $c \neq b_1$ which by symmetry forces $c = a_2$ to complete the proof. The other possibility $c \neq a_2$ is handled similarly, leading to the conclusion $c = b_1 = b_2$ must be the lower end of both rays.

Assuming $c \neq b_2$ means $b_2 \notin R_1$. By (21)

$$u(c) = u(b_2) + \|c - b_2\|, \qquad u(c) = u(a_1) - \|a_1 - c\|,$$

thus

$$u(a_1) - u(b_2) = \|a_1 - c\| + \|c - b_2\| \geq \|a_1 - b_2\|.$$

Strict inequality would violate the Lipschitz condition (20). Thus equality must hold, meaning c lies in the segment $[a_1, b_2]$ as well as $R_1 = [a_1, b_1]$. Since $b_2 \notin R_1$ these two segments, like the two rays, are not collinear. Their sole intersection point is a_1, hence $c = a_1$. By our above remarks this completes the proof: $c \neq b_1$ hence $c = a_2$ is the upper end of both rays. $\square$

Denoting the norm by $N(x) := \|x\|$ and its square by $F(x) := \|x\|^2$, the next lemma highlights some smoothness and strict convexity which follow from (4). From (23) it is clear that the strict convexity is uniform over the sphere ∂B.

Lemma 11 (Norm Smoothness and Strict Convexity)**.** *If the norm $N(x) := \|x\|$ satisfies (4), then $F(x) := \|x\|^2$ is of smoothness class $C^{1,1}(\mathbf{R}^n)$. Moreover, the unit ball $B := \{x \in \mathbf{R}^n \mid N(x) \leq 1\}$ is strictly convex, and*

(22)
$$|DN(x)\, y| < 1, \quad DN(x)\, x = 1 \quad \text{for all } y \neq \pm x \text{ with } \|x\| = \|y\| = 1,$$

where $DN(x)\, y$ denotes the pairing of $DN(x) \in (\mathbf{R}^n)^*$ and $y \in \mathbf{R}^n$.

Proof. Every norm $N(x)$ is convex throughout $\mathbf{R}^n$ and bounded by some multiple of the Euclidean norm: $N(y) \leq L|y|$. Thus both $N(x)$ and its square $F(x) = \|x\|^2$ are continuous functions. The midpoint convexity condition (4) therefore implies convexity of $F(x)$. We shall use the opposite inequality to conclude concavity of $g(x) := F(x) - L^2\Lambda|x|^2$: indeed, for $x, y \in \mathbf{R}^n$, g satisfies the midpoint estimate

$$\begin{aligned} \frac{g(x+y) + g(x-y)}{2} - g(x) &= \frac{F(x+y) + F(x-y)}{2} - F(x) - L^2\Lambda|y|^2 \\ &\leq \Lambda F(y) - L^2\Lambda|y|^2 \\ &\leq 0. \end{aligned}$$

Now recall that the distributional second derivative of a convex function is a non-negative definite matrix of Radon measures $D^2_{ij}F(x)$ [8, §6.3]. Concavity of g implies a pointwise bound on this matrix: $0 \leq D^2F(x) \leq 2L^2\Lambda I$. Thus F belongs to the Sobolev space $W^{2,\infty}(\mathbf{R}^n)$ and is differentiable Lipschitz continuously [6, §5.8.2–3]: $F \in C^{1,1}(\mathbf{R}^n)$.

To address (22), first observe strict convexity of the closed unit ball: given two distinct points $a, b \in B$, their midpoint must lie in the interior of B according to (4):

$$\left\| \frac{a+b}{2} \right\|^2 + \lambda \left\| \frac{a-b}{2} \right\|^2 \leq \frac{\|a\|^2 + \|b\|^2}{2} \leq 1 \tag{23}$$

with $\lambda > 0$. Now the triangle inequality implies

$$\begin{aligned} DN(x)\, y &:= \lim_{t \to 0} \frac{\|x + ty\| - \|x\|}{t} \\ &\leq \|y\| \leq 1 \end{aligned} \tag{24}$$

while homogeneity yields $DN(x)\, x = \|x\| = 1$. Thus the supporting hyperplane to the ball at $x \in \partial B$ consists of those $z \in \mathbf{R}^n$ satisfying $DN(x)\, z = 1$. Strict convexity prevents this hyperplane from touching the ball at more than one point, whence (24) can be sharpened to $DN(x)\, y < 1$ for $y \in B \setminus \{x\}$. Similarly, $DN(x)\,(-y) < 1$ for $-y \in B \setminus \{x\}$, which concludes the proof of (22). ☐

Lemma 12 (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:*

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

Remark 13. *This proof requires a modification of the Euclidean case dealt with by Evans and Gangbo* [7, Lemma 4.1].

Proof of Lemma 12. Choose z_0 in the interior of R. By (21), for some small $r_0 > 0$, we have

$$u(z_0 + te) = u(z_0) + t \quad \text{on} \quad -r_0 \leq t \leq r_0.$$

Rescale u by setting

$$u^r(z) = \frac{u(z_0 + rz) - u(z_0)}{r} \quad \text{for } \ 0 < r \leq r_0.$$

Then u^r satisfies the same Lipschitz condition (20) as u but is centered at $u^r(0) = 0$. Hence for some subsequence $r_k \to 0$ we have $u^{r_k} \to v$ where the convergence is uniform on every compact subset of $\mathbf{R}^n$. Clearly $v \in Lip_1(\mathbf{R}^n, \|\cdot\|)$ and

$$v(te) = t \ \text{ for all } t \in \mathbf{R}^1. \tag{25}$$

We shall now show linearity

$$v(z) = DN(e)\, z \quad \text{for all } z \in \mathbf{R}^n, \tag{26}$$

by exploiting the Lipschitz condition v inherits from u together with the first order Taylor expansion of $N(e + z/t) := \|e + z/t\|$ around e guaranteed by Lemma 11:

$$\frac{v(te) - v(z)}{t} \leq \frac{\|te - z\|}{|t|} = \|e\| - DN(e)\,\frac{z}{t} + o(\frac{1}{t}).$$

Subtracting $v(te)/t = \|e\|$ from both sides, the two limits $t \to \pm\infty$ of this inequality combine to yield (26).

We conclude that

$$\lim_{r\to\infty} \frac{u(rz)}{r} = DN(e)\,z \quad \text{uniformly for } \ z \in B_1(0).$$

This implies that u is differentiable at z_0, with $Du(z_0) = DN(e)$. The remaining assertions of Lemma 12 follow directly from (22). $\square$

The next lemma exploits uniform convexity and smoothness of the norm to produce a quantitative estimate of how far away any two rays must be from crossing. When $F(x) = \|x\|^2$, it states that the sums of the squares of the diagonals AC and BD of any quadrilateral $ABCD$ are controlled by the distance between the midpoints of the shorter pair (in least squares sense) of opposite sides. In particular, no quadrilateral can be folded in such a way that the midpoints of these two sides are brought close together unless both pairs of opposite corners are also driven together — with a particular rate. In the Euclidean or Hilbert space setting, the rate constant $(1 + \Lambda/\lambda)/(1 + \lambda) = 1$ given by the polarization identity is seen to be sharp by folding up a square. Alternately, the estimate (28) can be interpreted as a reverse form of the triangle inequality, which holds for vectors that are sufficiently aligned.

Lemma 14 (Twisted Quadrilateral Non-crossing Estimate). *Let $F : V \longrightarrow \mathbf{R}$ be any function on a vector space V, uniformly smooth and convex enough that for some $\Lambda, \lambda > 0$ and all $x, y \in V$ the following inequalities hold:*

$$\lambda\, F(y) \leq \frac{1}{2}F(x + y) - F(x) + \frac{1}{2}F(x - y) \leq \Lambda\, F(y). \tag{27}$$

If four points $a, b, c, d \in V$ satisfy $F(a-b)+F(c-d) \leq F(a-d)+F(c-b)$ then

$$F\left(\frac{a-c}{2}\right) + F\left(\frac{b-d}{2}\right) \leq \frac{1 + (\Lambda/\lambda)}{1 + \lambda}\, F\left(\frac{a+b}{2} - \frac{c+d}{2}\right). \tag{28}$$

Proof. Applying uniform convexity (27) with both $(x,y) = (a-c, b-d)/2$ and $(x,y) = (b-d, a-c)/2$ and then summing yields

$$(1+\lambda)\left[F\left(\frac{a-c}{2}\right) + F\left(\frac{b-d}{2}\right)\right] \leq F\left(\frac{a-c}{2} + \frac{b-d}{2}\right) + \tilde{F}\left(\frac{a-c}{2} - \frac{b-d}{2}\right), \tag{29}$$

where $\tilde{F}(z) := [F(z) + F(-z)]/2$. The desired inequality will follow if we can show the second last term controls the last one. Applying uniform convexity again yields

$$\lambda \tilde{F}\left(\frac{a-b}{2} - \frac{c-d}{2}\right) \leq \frac{F(a-b)+F(c-d)}{2} - F\left(\frac{a-b}{2} + \frac{c-d}{2}\right), \tag{30}$$

either with or without the tilde. Uniformity of the smoothness gives

$$\frac{F(a-d)+F(c-b)}{2} - F\left(\frac{a-d}{2} + \frac{c-b}{2}\right) \leq \Lambda F\left(\frac{a-d}{2} - \frac{c-b}{2}\right). \tag{31}$$

But the left hand side of (31) dominates the right hand side of (30) by hypothesis, so

$$\lambda \tilde{F}\left(\frac{a-c}{2} - \frac{b-d}{2}\right) \leq \Lambda F\left(\frac{a+b}{2} - \frac{c+d}{2}\right).$$

Together with (29), this completes the proof of (28). □

The next lemma is crucial for the definition of the change to variables in which one variable is along transport rays. The lemma shows that if transport rays intersect a level set of $u(z)$ in their interior points, then directions of rays have a Lipschitz dependence on the point of intersection, provided distances from the point of intersection to endpoints of a ray are uniformly bounded away from zero for all rays. Taking $x = 0$ and $F(y) = F(-y)$ symmetrical implies $\lambda \leq 1 \leq \Lambda$ in (27), so the Lipschitz constant of Lemma 16 is seen to satisfy $C \geq 1$ with equality only in the Hilbert space case.

Definition 15 (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 10) then

$$\nu(z) := \frac{a-b}{\|a-b\|}. \tag{32}$$

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.*

Lemma 12 shows that on transport rays, the direction function $\nu(z)$ is nothing but the gradient of u computed in the Finsler setting.

Lemma 16 (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 Monge's potential $u(y_1) = u(y_2)$, then the ray directions (32) satisfy a Lipschitz bound*

$$\|\nu(y_1) - \nu(y_2)\| \leq \frac{C}{\sigma}\|y_1 - y_2\|, \tag{33}$$

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.

Proof. Let $z_k, x_k \in R_k$ denote the points at distance σ above and below y_k on the ray, so that

$$\begin{aligned} u(z_k) &= u(y_1) + \sigma, && (34)\\ u(x_k) &= u(y_1) - \sigma, && (35)\\ \|z_k - x_k\| &= 2\sigma, && (36)\\ y_k &= (z_k + x_k)/2, && (37)\\ \text{and} \quad \nu(y_k) &= (z_k - x_k)/2\sigma && (38) \end{aligned}$$

for $k = 1, 2$. Thus

$$\|\nu(y_1) - \nu(y_2)\|^2 = \frac{1}{\sigma^2}\left\|\frac{z_1 - x_1}{2} - \frac{z_2 - x_2}{2}\right\|^2, \tag{39}$$

while uniform convexity of the norm (4) gives

$$\begin{aligned} \left\|\frac{z_1 - z_2}{2} + \frac{x_2 - x_1}{2}\right\|^2 \leq\ & \frac{1}{2}\|z_1 - z_2\|^2 + \frac{1}{2}\|x_2 - x_1\|^2 \\ & -\lambda\left\|\frac{z_1 - z_2}{2} - \frac{x_2 - x_1}{2}\right\|^2. \end{aligned} \tag{40}$$

Combining (34–36),

$$\begin{aligned} \|z_1 - x_1\| = & \quad u(z_1) - u(x_2) & \le \|z_1 - x_2\| \\ \|z_2 - x_2\| = & \quad u(z_2) - u(x_1) & \le \|z_2 - x_1\|, \end{aligned}$$

where the Lipschitz condition (20) controls the cross-terms. Lemma 14 therefore applies to $F(x) = \|x\|^2$ and the four points $(a, b, c, d) = (z_1, x_1, z_2, x_2)$, to yield

$$\left\| \frac{z_1 - z_2}{2} \right\|^2 + \left\| \frac{x_1 - x_2}{2} \right\|^2 \le \frac{1 + (\Lambda/\lambda)}{1 + \lambda} \left\| \frac{z_1 + x_1}{2} - \frac{z_2 + x_2}{2} \right\|^2. \tag{41}$$

Combining (39–41) with the identity (37) gives

$$\sigma^2 \|\nu(y_1) - \nu(y_2)\|^2 \le \left(2\frac{1 + (\Lambda/\lambda)}{1 + \lambda} - \lambda \right) \|y_1 - y_2\|^2,$$

to complete the proof. □

Remark 17. *The proof of Lemma 16 uses only the Lipschitz property of u, and not the optimality of u in the Kantorovich Problem 2. Thus its conclusions hold true for any $u \in Lip_1(\mathbf{R}^n, \|\cdot\|)$, if we call each segment $[a, b]$ on which $u(a) - u(b) = \|a - b\|$ a transport ray, and define the direction function ν accordingly.*

3. Measure decomposing change of variables

It is in this section that we construct the change of variables on $\mathbf{R}^n$ which is the heart of our proof. Lemma 16 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 the lemma 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.

We begin by parameterizing the level sets of u using a lemma of Federer [9, §3.2.9]. The key observation is that we only need this parameterization on the interiors of transport rays, where $Du \neq 0$ exists in view of Lemma 12. From now on, it will be convenient to fix a Euclidean structure in $\mathbf{R}^n$. The Euclidean scalar product and associated norm are denoted by $(\cdot, \cdot)$ and $|z| := (z, z)^{1/2}$, while $B_R(z)$ denotes the Euclidean ball of radius R centered at $z \in \mathbf{R}^n$. Of course, a function is Lipschitz in one norm if and only if it is Lipschitz in all norms, though the Lipschitz constants may differ.

Lemma 18 (Bi-Lipschitz Parametrization of Level Sets)**.** *Let* $u : \mathbf{R}^n \to \mathbf{R}^1$ *be a Lipschitz function,* $p \in \mathbf{R}^1$*, and* S_p *the level set* $\{x \in \mathbf{R}^n \mid u(x) = p\}$*. Then the set*

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

has a countable covering consisting of Borel sets $S_p^i \subset S_p$*, 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*

$$V(U(x)) = x \quad \textit{for all} \quad x \in S_p^i. \tag{42}$$

Proof. Note that if $Du(x) \neq 0$ exists, then $Du(x)(\mathbf{R}^n) = \mathbf{R}^1$. Federer [9, §3.2.9] asserts that

$$\{x \in \mathbf{R}^n \mid u \text{ is differentiable at } x \text{ and } Du(x)(\mathbf{R}^n) = \mathbf{R}^1\}$$

has a countable covering consisting of Borel sets E_i such that there exist orthogonal projections $\pi_i : \mathbf{R}^n \longrightarrow \mathbf{R}^{n-1}$ in $O^*(n, n-1)$ and Lipschitz maps

$$\hat{U}_i : \mathbf{R}^n \to \mathbf{R}^1 \times \mathbf{R}^{n-1} \quad \text{and} \quad \hat{V}_i : \mathbf{R}^1 \times \mathbf{R}^{n-1} \to \mathbf{R}^n \tag{43}$$

with

$$\hat{U}_i(x) = (u(x), \pi_i(x)), \quad \text{and} \quad \hat{V}_i[\hat{U}_i(x)] = x \quad \text{for all} \quad x \in E_i.$$

Clearly the sets

$$S_p^i := S_p \cap E_i \tag{44}$$

cover S_p. For any fixed $i \in \mathbf{N}$, define

$$U : \mathbf{R}^n \to \mathbf{R}^{n-1} \quad \text{and} \quad V : \mathbf{R}^{n-1} \to \mathbf{R}^n \tag{45}$$

by

$$U := \pi \circ \hat{U}_i, \;\; \text{where } \pi : \mathbf{R}^1 \times \mathbf{R}^{n-1} \to \mathbf{R}^{n-1} \text{ is the projection } (x_1, X) \to X,$$
$$V(X) = \hat{V}_i(p, X) \quad \text{for all} \quad X \in \mathbf{R}^{n-1}.$$

Clearly U and V are Lipschitz continuous, while $U(x) = \pi_i(x)$ for $x \in S_p^i$, and $V(U(x)) = \hat{V}_i(p, \pi_i(x)) = \hat{V}_i[\hat{U}_i(x)] = x$ establishes (42). □

For each rational level $p \in \mathbf{Q}$ and integer $i \in \mathbf{N}$, we shall extend these coordinates to the transport rays intersecting S_p^i. Taken together, these coordinate charts must parametrize all points $T_1 \subset \mathbf{R}^n$ on transport rays (cf. Definition 8). It is convenient to define them on a countable collection of subsets called *clusters of rays*:

Definition 19 (Ray Clusters). *Fix $p \in \mathbf{Q}$, a Kantorovich potential u, and the Borel cover $\{S_p^i\}_i$ of the level set $S_p := \{x \in \mathbf{R}^n \mid u(x) = p\}$ in Lemma 18. For each $i, j \in \mathbf{N}$ let the* cluster *$T_{pij} := \cup R_z$ denote the union of all transport rays R_z which intersect S_p^i, and for which the point of intersection $z \in S_p^i$ 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 $T_{pij}^0 := \cup_z (R_z^0)$.*

Lemma 20 (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:*

$$\text{Either } (R)^0 \subset T_{pij}, \quad \text{or } (R)^0 \cap T_{pij} = \emptyset. \tag{46}$$

Proof. A transport ray $R = [a, b]$ has positive length by Definition 7. Along it, the Kantorovich potential u is an affine function with nonzero slope, according to Lemma 12. Thus there is some rational number $p \in (u(a), u(b))$, for which R intersects the level set $S_p := \{x \mid u(x) = p\}$. The point x of intersection belongs to one of the covering sets $S_p^i \subset S_p$ of Lemma 18, and lies a positive distance from each end of the ray, so $R \subset T_{pij}$ for some $j \in \mathbf{N}$.

Conversely, if the interior of some other ray R^0 intersects one of the rays R_z comprising the cluster T_{pij}, the non-crossing property of Lemma 10 forces $R = R_z \subset T_{pij}$, to complete the proof of (46). ☐

Definition 21 (Ray Ends). *Denote by $\mathcal{E} \subset T_1$ the set of endpoints of transport rays.*

On each ray cluster T_{pij} we are now ready to define the Lipschitz change of variables which inspired the title of this section:

Lemma 22 (Lipschitz Change of Variables). *Each ray cluster $T_{pij} \subset \mathbf{R}^n$ admits coordinates $G = G_{pij} : T_{pij}^0 \to \mathbf{R}^{n-1} \times \mathbf{R}^1$ with inverse $F = F_{pij} : G(T_{pij}^0) \to \mathbf{R}^n$ 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 $T_{pij}^\lambda := \{x \in T_{pij}^0 \mid \|x - a\|, \|x - b\| > \lambda\}$, where a and b denote the endpoints of the (unique) transport ray R_x;*
iii. *$F(G(x)) = x$ for each $x \in T_{pij}^0$;*
iv. *If a transport ray $R_z \subset T_{pij}$ intersects S_p^i at z, then each interior point $x \in (R_z)^0$ of the ray satisfies*

$$G(x) = (U(z), u(x) - u(z)), \tag{47}$$

where $U : \mathbf{R}^n \to \mathbf{R}^{n-1}$ gives the Lipschitz coordinates (42) on S_p^i.

Remark 23 (Flattening Level Sets). *The final assertion of Lemma 22 implies: (a) F maps the part of the hyperplane $\mathbf{R}^{n-1} \times \{0\}$ which lies within $G(T^0_{pij})$ onto S^i_p; (b) F maps the segment where each "vertical" line $\{X\} \times \mathbf{R}^1$ intersects $G(T^0_{pij})$ onto a transport ray. Thus in the new coordinates $(X, x_n) \in \mathbf{R}^{n-1} \times \mathbf{R}^1$, the level sets of u are flattened: they are parameterized by the variables $X = (x_1, \ldots, x_{n-1})$ while x_n varies along the transport rays.*

Proof. Lemma 10 shows that rays do not cross, while Definition 7 (or Lemma 12) shows that u is an affine function on each ray, with slope as large as permitted by the Lipschitz constraint (20). Thus every point $x \in T^0_{pij}$ lies on a unique transport ray, and this ray intersects the level set S_p in a single point $z \in S^i_p$, so the expression (47) defines a map $G : T^0_{pij} \to \mathbf{R}^{n-1} \times \mathbf{R}^1$ throughout the cluster. It remains to construct the inverse map F on $G(T^0_{pij}) \subset \mathbf{R}^{n-1} \times \mathbf{R}^1$. Let $(X, x_n) \in G(T^0_{pij})$, and let V be the map (42) parametrizing S^i_p. Then the point $V(X) \in S^i_p$ is an interior point of some transport ray R, both of whose endpoints are separated from $V(X)$ by a distance exceeding $1/j$. Let $\nu(\,\cdot\,)$ be the direction function (32) associated with the Kantorovich potential u, and define

$$F(X, x_n) := V(X) + x_n \nu(V(X)). \tag{48}$$

That F inverts G (assertion (iii)) now follows from (42), (47) and the fact that u is affine with maximal slope along the ray R.

To prove F is Lipschitz on $G(T^0_{pij}) \subset \mathbf{R}^{n-1} \times \mathbf{R}^1$, introduce

$$\Lambda := \{X \in \mathbf{R}^{n-1} \mid (X, 0) \in G(T^0_{pij})\}. \tag{49}$$

We first claim the ray direction $\nu \circ V$ is a Lipschitz function of $X \in \Lambda$. Indeed, recalling that $V(X) \in S^i_p$ is separated from the endpoints of $R_{V(X)}$ by a distance greater than $1/j$, we invoke Lemma 16 to conclude that $X, X' \in \Lambda$ satisfy

$$\begin{aligned} \|\nu(V(X)) - \nu(V(X'))\| &\leq jC_1 \|V(X) - V(X')\| && (50)\\ &\leq jC_2 |X - X'| && (51) \end{aligned}$$

because $V : \mathbf{R}^{n-1} \to \mathbf{R}^n$ was Lipschitz in Lemma 18. To complete the proof that F is Lipschitz, it remains only to bound x_n in (48). Since the supports $\mathcal{X}$ and $\mathcal{Y}$ of the original measures were compact, the transport rays $T_{pij} \subset T_1$ lie in a bounded set. It follows from the definition (47) of G that $(X, x_n) \in G(T^0_{pij})$ is also bounded, since u and U are Lipschitz on $\mathbf{R}^n$. Finally, we can extend F to all of $\mathbf{R}^{n-1} \times \mathbf{R}^1$ while preserving the Lipschitz bound (51) using Kirszbraun's theorem [9, §2.10.43], to conclude the proof of assertion 1.

It remains to prove assertion (ii) of the lemma. Let $\lambda > 0$. We first show the direction function $\nu(\,\cdot\,)$ to be Lipschitz on T^λ_{pij}. Being discontinuous at the mutual end of two rays, its Lipschitz constant must depend on λ. Let x, $x' \in T^\lambda_{pij}$ lie on the transport rays R and R'. If $\|x - x'\| \geq \lambda/2$ there is nothing to prove, since

$$\|\nu(x) - \nu(x')\| \leq 2 \leq \frac{4}{\lambda}\|x - x'\|.$$

Therefore, assume $\|x-x'\| < \lambda/2$ and hence $|u(x)-u(x')| \leq \|x-x'\| < \lambda/2$. The point $y' := x' + [u(x) - u(x')]\nu(x')$ then lies on the ray R', since the ends of R' are at least distance λ from x'. Moreover, $u(y') = u(x') + [u(x) - u(x')] = u(x)$, and the distances from from x and y' to the ends of R and R' are at least $\lambda/2$ respectively. Invoking Lemma 16 again yields:

$$\|\nu(x) - \nu(x')\| = \|\nu(x) - \nu(y')\| \leq \frac{2C}{\lambda}\|x - y'\|. \tag{52}$$

Moreover, $x', y' \in R'$ lie on the same transport ray, and $u(x) = u(y')$, so

$$\|x' - y'\| = |u(x') - u(y')| = |u(x') - u(x)| \leq \|x' - x\| \tag{53}$$

combines with the triangle inequality $\|x - y'\| \leq \|x - x'\| + \|x' - y'\|$ to produce the desired bound for $\nu(\,\cdot\,)$ on T^λ_{pij}:

$$\|\nu(x) - \nu(x')\| \leq \frac{4C}{\lambda}\|x - x'\|. \tag{54}$$

Turning to G, we estimate

$$|G(x) - G(x')| \leq |G(x) - G(y')| + |G(y') - G(x')|. \tag{55}$$

Since x' and y' lie on R', definition (47) yields

$$|G(y') - G(x')| = |u(x') - u(y')| = |u(x') - u(x)| \leq \|x - x'\|. \tag{56}$$

Let z and z' be the points where R and R' pierce S^i_p. Since $u(x)-u(z) = u(y') - u(z')$, the same definition gives

$$|G(x) - G(y')| = |U(z) - U(z')|. \tag{57}$$

Setting $\sigma := u(z) - u(x)$, we have $z = x + \sigma\nu(x)$ and $z' = y' + \sigma\nu(x')$. Also $|\sigma|$ is bounded by the diameter of the cluster T^λ_{pij}. Because the coordinates U were Lipschitz, we have

$$|U(z) - U(z')| \leq C_3\|x - y'\| + \sigma C_3\|\nu(x) - \nu(x')\| \tag{58}$$

$$\leq C_3(2 + 4\sigma C\lambda^{-1})\|x - x'\| \tag{59}$$

from (53–54). Now (55–59) imply G is Lipschitz on T^λ_{pij}, to complete the lemma. $\square$

The next step is to address measurability of the sets T_{pij} and $G(T^0_{pij})$. As in Evans and Gangbo [7], this is done with the help of the distance functions to the upper and lower ends of rays:

Lemma 24 (Semicontinuity of Distance to Ray Ends)**.** *At each $z \in \mathbf{R}^n$ define*

$$\alpha(z) := \sup \{\|z-y\| \mid y \in \mathcal{Y},\ u(z)-u(y) = \|z-y\|\}, \tag{60}$$

$$\beta(z) := \sup \{\|z-x\| \mid x \in \mathcal{X},\ u(x)-u(z) = \|z-x\|\}, \tag{61}$$

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

Proof. We prove only the upper semicontinuity of $\alpha(z)$; the proof for $\beta(z)$ is similar. Given any sequence of points $z_n \to z$ for which $\alpha_0 := \lim_n \alpha(z_n)$ exists, we need only show $\alpha_0 \leq \alpha(z)$. It costs no loss of generality to assume $\alpha_0 > -\infty$ and $\alpha(z_n) > -\infty$; moreover, $\alpha(z_n) < \infty$ since the support $\mathcal{Y}$ of the measure μ^- was assumed compact. From (60),

$$\alpha(z_n) - 1/n \leq \|z_n - y_n\| = u(z_n) - u(y_n) \tag{62}$$

for some sequence $y_n \in \mathcal{Y}$. By compactness of $\mathcal{Y}$, a convergent subsequence $y_n \to y \in \mathcal{Y}$ exists. The (Lipschitz) continuity of u yields $\alpha_0 \leq \|z-y\| = u(z)-u(y) \leq \alpha(z)$ from the limit of (62), which proves the lemma. $\square$

Geometrically, the functions α, β have the following meaning: If z lies on a transport ray R, then $\alpha(z)$ and $\beta(z)$ are the distances (in $\|\cdot\|$) from z to the lower and upper end of R respectively; thus at ray ends $z \in \mathcal{E}$, exactly one of these distances vanishes. If $z \in T_0$ is a ray of zero length, then $\alpha(z) = \beta(z) = 0$. If $z \in \mathbf{R}^n \setminus (T_0 \cup T_1)$, then either $\alpha(z) = -\infty$ or $\beta(z) = -\infty$. We combine these functions with our change of variables to show the clusters of ray interiors T^0_{pij} are Borel sets, and to give a much simpler proof than Evans and Gangbo that the ray ends have measure zero [7, Proposition 5.1]. In what follows, n-dimensional Lebesgue measure is denoted $\mathcal{L}^n$.

Lemma 25 (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 $p \in \mathbf{Q}$ and $i, j \in \mathbf{N}$, the cluster T^0_{pij} of ray interiors and its flattened image $G(T^0_{pij})$ from Lemma 22 are Borel.*

Proof. First observe that $T_0 = \{z \in \mathbf{R}^n \mid \alpha(z) = \beta(z) = 0\}$ while $\mathcal{E} = \{z \in \mathbf{R}^n \mid \alpha(z)\beta(z) = 0 \text{ but } \alpha(z) + \beta(z) > 0\}$. Both of these sets are Borel by the upper semicontinuity of α and β shown in Lemma 24.

Therefore, fix $p \in \mathbf{Q}$ and $i, j \in \mathbf{N}$ and recall the Borel set $S_p^i \subset \mathbf{R}^n$ and Lipschitz coordinates $U : \mathbf{R}^n \longrightarrow \mathbf{R}^{n-1}$ on it from Lemma 18. Since U is univalent (i.e., one to one) on S_p^i, it follows from Federer [9, §2.2.10, page 67] that $U(S_p^i)$ is a Borel subset of $\mathbf{R}^{n-1}$. Moreover, the set Λ defined in (49) is given by

$$\Lambda = \{X \in U(S_p^i) \mid \alpha(U^{-1}(X)), \beta(U^{-1}(X)) > 1/j\}$$

according to (47), which with Definition 19 also yields the image

$$G(T_{pij}^0) = \{(X, x_n) \mid X \in \Lambda,\ -\alpha(V(X)) < x_n < \beta(V(X))\}$$

of the ray cluster in flattened coordinates. Here $V = U^{-1}$ is Lipschitz, so $\alpha \circ V$, $\beta \circ V$ are upper semicontinuous in view of Lemma 24. Thus we conclude that both $\Lambda \subset \mathbf{R}^{n-1}$ and $G(T_{pij}^0) \subset \mathbf{R}^{n-1} \times \mathbf{R}^1$ are Borel. Lemma 22(iii) shows the transformation $F = G^{-1}$ back to the original coordinates is well-defined and univalent on $G(T_{pij}^0)$. Since F extends to a Lipschitz function throughout $\mathbf{R}^n$ and $T_{pij}^0 = F(G(T_{pij}^0))$, we conclude, using Federer [9, §2.2.10] again, that T_{pij}^0 is Borel.

To show the ray ends have measure zero, consider the corresponding points $\mathcal{G} \subset \mathbf{R}^{n-1} \times \mathbf{R}^1$ of T_{pij} in the flattened coordinate system:

$$\mathcal{G} = \{(X, -\alpha(V(X))) \mid X \in \Lambda\} \cup \{(X, \beta(V(X))) \mid X \in \Lambda\}.$$

Using upper semicontinuity of $\alpha \circ V$ and $\beta \circ V$ we conclude $\mathcal{G}$ is a Borel set, and $\mathcal{L}^n(\mathcal{G}) = 0$ by Fubini's theorem. Now $\mathcal{E} \cap T_{pij} = F(\mathcal{G})$. Since $F : \mathbf{R}^{n-1} \times \mathbf{R}^1 \to \mathbf{R}^n$ is a Lipschitz map, we can use $\mathcal{L}^n(\mathcal{G}) = 0$ and the Area formula [9, §3.2.3] to conclude that $\mathcal{L}^n(\mathcal{E} \cap T_{pij}) = 0$ (and hence is Lebesgue measurable). By Lemma 20, the clusters $\{T_{pij}\}$ form a countable cover for $\mathcal{E} \subset T_1$, so $\mathcal{L}^n(\mathcal{E}) = 0$ to conclude the proof. □

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_{pij}^0 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 $T_{(k)}^0 \to T_{(k)}^0 \setminus (\cup_{l=1}^{k-1} T_{(l)}^0)$ analogously. We will continue to denote the modified sets by T_{pij} and T_{pij}^0. Note that the structure of the clusters T_{pij} remains the same: for each T_{pij} we have a Borel subset $S_{pij} := T_{pij} \cap S_p$ of $S_p^i \subset \mathbf{R}^n$ on which there are Lipschitz coordinates U, V (42) satisfying

$$V(U(x)) = x \quad \text{for all} \quad x \in S_{pij}. \tag{63}$$

Indeed, since the new cluster is a subset of the old, the former maps U, V will suffice. From the modification procedure it also follows that the ray property (46) holds for the modified sets — which justifies calling them clusters — and that the ray R_z corresponding to each $z \in S_{pij}$ extends far enough on both sides of S_p, (i.e. $\alpha(z), \beta(z) > 1/j$) to define coordinates F, G on T_{pij} satisfying all assertions of Lemma 22; (again, the original maps F and G work for the modified clusters). The measurability Lemma 25 holds for the new clusters, as follows readily from the modification procedure. Thus from now on we assume:

(64) *The clusters of ray interiors T^0_{pij} are disjoint.*

4. Mass balance on rays

Since we intend to solve Monge's problem by constructing a map which moves mass along rays, it is essential to know that f^+ and f^- assign the same amount of mass to each transport ray. In the Euclidean case this is the content of Evans and Gangbo [7, Lemma 5.1]. Here it remains true, but our proof exploits the fact that the optimal maps $s_\varepsilon(x)$ between f^+ onto f^- for the cost $c_\varepsilon(x,y) = \|x-y\|^{1+\varepsilon}$ in Theorem 2(iv) accumulate onto transport rays of the limiting Kantorovich potential.

While individual rays all have mass zero, one can consider arbitrary collections of transport rays instead. The ray ends, having measure zero, are neglected. Thus:

Definition 26 (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 $R^0_z \subseteq A$, 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_{pij}.

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

Lemma 27 (Mass Balance on Rays). *Let $A \subset \mathbf{R}^n$ be a Borel transport set. Then*

$$\int_A f^+(x)\,dx = \int_A f^-(x)\,dx. \tag{65}$$

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

$$\int_{A^+} f^+(x)\,dx \geq \int_{A^+} f^-(x)\,dx. \tag{66}$$

Proof. We will prove inequality (66) for a positive end of a transport set. Equality (65) for transport sets then follows by symmetry.

Let A^+ be a positive end of a transport set. Since $T_0 \cup T_1 \subset conv[\mathcal{X} \cup \mathcal{Y}]$ contains all transport rays by Definitions 7–8, and the supports of $f^\pm$ by Lemma 9, it costs no generality to replace A^+ by its intersection with $T_0 \cup T_1$. Thus we assume $A^+ \subset T_0 \cup T_1$.

Assume first that A^+ is a closed set and that A^+ does not contain any lower ends of rays. Then A^+ is compact since $\mathcal{X} \cup \mathcal{Y}$ is bounded. Recall that our limiting Kantorovich potential u was obtained from (14) and a limit $(\varphi_{\varepsilon_j}, \psi_{\varepsilon_j}) \to (\varphi_0, \psi_0)$ — uniform on $\mathcal{X} \times \mathcal{Y}$ — of potentials minimizing $K(\varphi, \psi)$ on $J_{\varepsilon_j}(\mathcal{X}, \mathcal{Y})$. Here the convex costs $c_{\varepsilon_j}(x, y) = \|x - y\|^{1+\varepsilon_j} \to c_0(x, y)$ uniformly on $\mathcal{X} \times \mathcal{Y}$ as $j \to \infty$ in Proposition 3.

For $r > 0$, let $N_r(A^+) = \{x \in \mathbf{R}^n \mid \mathrm{dist}(x, A^+) < r\}$ denote the r-neighborhood of A^+. Since A^+ is the positive end of a transport set and closed and does not contain any lower ends of rays, it follows that if $y \in A^+$ and $u(x) - u(y) = \|x - y\|$ for some $x \in \mathcal{X}$, then $x \in A^+$. Since $u \in \mathrm{Lip}_1(\mathbf{R}^n, d)$ and A^+, $\mathcal{X} \setminus N_r(A^+)$ are compact sets, it follows that

$$\inf_{y \in A^+,\, x \in \mathcal{X} \setminus N_r(A^+)} [\|x - y\| - u(x) + u(y)] \geq \delta(r) > 0.$$

By (14),

$$\varphi_0(x) + \psi_0(y) \geq -\|x - y\| + \delta(r) \quad \text{for any} \quad y \in A^+,\ x \in \mathcal{X} \setminus N_r(A^+).$$

The uniform convergence mentioned above then yields

$$\varphi_{\varepsilon_j}(x) + \psi_{\varepsilon_j}(y) \geq -c_{\varepsilon_j}(x, y) + \frac{\delta(r)}{2} \quad \text{for any} \quad y \in A^+,\ x \in \mathcal{X} \setminus N_r(A^+),$$

provided $j > j_0(r)$ is sufficiently large. From Theorem 2(iii–iv) it now follows that $s_{\varepsilon_j}(x) \in A^+$ implies $x \in N_r(A^+)$ if $j > j_0(r)$. Here $s_{\varepsilon_j} : \mathcal{X} \longrightarrow \mathcal{Y}$ is the unique optimal map between f^+ and f^- with respect to the cost function $c(x, y) = \|x - y\|^{1+\varepsilon_j}$. Since s_{ε_j} pushes forward μ^+ to μ^-, we obtain $\mu^+[N_r(A^+)] \geq \mu^-(A^+)$. But because A^+ is closed, the limit $r \to 0^+$ yields (66).

If A^+ is merely Borel, then we can replace A^+ by $A^+ \setminus \mathcal{E}$ since $\mathcal{E}$ is a Borel set of $\mathcal{L}^n$ measure zero and $\mu^\pm$ are absolutely continuous with respect to Lebesgue. Thus we assume that $A^+ \subset (T_0 \cup T_1) \setminus \mathcal{E}$. Since

A^+ is Borel, for any $\delta > 0$ there exists a closed set $C = C_\delta \subset A^+$ such that $\mathcal{L}^n[A^+ \setminus C] < \delta$. Denote by $R^+(C)$ the set

$$R^+(C) = C \cup \left(\cup_{z \in C \cap (T_1 \setminus \mathcal{E})}[z, a(z)]\right),$$

where $a(z)$ denotes the upper end of the transport ray R_z. Continuity of $u(\cdot)$ implies that $R^+(C)$ is closed. By definition, $R^+(C)$ is a positive end of a transport set, and does not contain lower ends of rays since $C \cap \mathcal{E} = \emptyset$. Thus

$$\mu^+[R^+(C)] \geq \mu^-[R^+(C)]. \tag{67}$$

Since $A^+ \supset C$ is a positive end of a transport set, $R^+(C) \subset A^+ \cup \mathcal{E}$. Moreover, $\mathcal{L}^n(\mathcal{E}) = 0$, so $\mu^\pm[R^+(C)] \leq \mu^\pm(A^+)$. Finally, $\mathcal{L}^n[A^+ \setminus R^+(C)] < \delta$ and the measures $\mu^\pm$ are absolutely continuous with respect to Lebesgue, so $\mu^\pm[A^+ \setminus R^+(C)] \to 0$ with $\delta \to 0^+$. Thus $\mu^\pm[R^+(C)] \to \mu^\pm(A^+)$, and (67) implies (66). $\square$

5. Construction of the optimal map

This final section is devoted to the proof of Theorem 1 by constructing an optimal map for Monge's problem.

Proof. **Step 1. Localization to clusters of rays.** According to Lemma 6, 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 (18).

Decompose the set $\mathcal{X} \cup \mathcal{Y}$ into the rays T_0 of length zero, clusters of ray interiors T^0_{pij}, and the ray ends $\mathcal{E}$ using Lemmas 9 and 20. The cluster property (46) 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 25, they are neglected here and in the sequel. Thus we can construct an optimal map s separately on each cluster T^0_{pij} and on T_0. Indeed, suppose for each (p, i, j) we have a map $s_{pij} : T^0_{pij} \to T^0_{pij}$ pushing forward $\mu^+_{|T^0_{pij}}$ onto $\mu^-_{|T^0_{pij}}$, and a map $s_0 : T_0 \to T_0$ pushing $\mu^+_{|T_0}$ forward onto $\mu^-_{|T_0}$. Here $\mu_{|A}$ denotes the restriction of measure μ from $\mathbf{R}^n$ to $A \subset \mathbf{R}^n$. The clusters T^0_{pij} and T_0 are disjoint and Borel by (64) and Lemma 25. Thus the map $s : \mathbf{R}^n \to \mathbf{R}^n$ defined by

$$s(x) = \begin{cases} s_0(x) & \text{for} \quad x \in T_0, \\ s_{pij}(x) & \text{for} \quad x \in T^0_{pij}, \end{cases} \tag{68}$$

is well-defined, Borel, and pushes μ^+ forward to μ^-.

Consider s_0 first. Since every subset $A \subset T_0$ is a transport set, Lemma 27 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 proof 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 22 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 [9, §3.2.5] yields

$$\int_{G(T^0_{pij})} \varphi(F(x)) f^\pm(F(x)) J_n F(x)\, dx = \int_{T^0_{pij}} \varphi(z) f^\pm(z) dz \tag{69}$$

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

$$\hat{f}^\pm(x) = \begin{cases} f^\pm(F(x)) J_n F(x) & x \in G(T^0_{pij}); \\ 0 & \text{otherwise.} \end{cases} \tag{70}$$

The characteristic function $\varphi = \chi_{G(T^0_{pij})}$ in (69) shows $\hat{f}^\pm$ is summable; it is obviously non-negative and Borel since Lemma 25 shows $G(T^0_{pij})$ Borel and bounded. Introduce the measures $d\theta^\pm := \hat{f}^\pm(x)dx$. Comparing (3) with (69) gives

$$F_\# \theta^\pm = \mu^\pm_{pij}, \tag{71}$$

meaning the map F pushes $\theta^\pm$ forward to $\mu^\pm_{pij}$. From Lemma 22(ii–iii) we deduce the inverse map G is Borel on T^0_{pij}, and $G(F(y)) = y$ on $G(T^0_{pij})$. With (71) this implies

$$G_\# \mu^\pm_{pij} = \theta^\pm. \tag{72}$$

From (71–72) 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 22(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

$$\Psi^\pm(X,\tau) := \int_\tau^\infty \hat f^\pm(X,x_n)\,dx_n \tag{73}$$

$$= \int_{\mathbf{R}^1} \chi_{(0,\infty)}(x_n-\tau)\hat f^\pm(X,x_n)\,dx_n. \tag{74}$$

Here $\Psi^\pm$ is non-negative and Borel throughout $\mathbf{R}^n$ [22, §8.8], with a continuous non-increasing dependence on τ. For a.e. $X \in \mathbf{R}^{n-1}$ we shall show:

$$\Psi^+(X,\tau) \geq \Psi^-(X,\tau) \tag{75}$$

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

$$\Psi^+(X,-\infty) = \Psi^-(X,-\infty) < \infty \tag{76}$$

as $\tau \to -\infty$. These inequalities are interpreted to mean that at no point τ along a transport ray, can the mass of μ^- above τ exceed the mass of μ^+, though they balance in the limit (76). Note that $\Psi^\pm(X,\tau)$ becomes independent of $|\tau|$ for large $|\tau|$, since $\hat f^\pm$ had compact support in (70). The bound in (76) comes from Fubini's theorem applied to $\hat f^\pm \in L^1(\mathbf{R}^n)$.

Let us first fix $\tau \in \mathbf{R}$, and establish (75) for a.e. $X \in \mathbf{R}^{n-1}$. Consider the sets

$$\Lambda^+ := \{X \in \mathbf{R}^{n-1} \mid \Psi^+(X,\tau) < \Psi^-(X,\tau)\}, \tag{77}$$

$$\Omega^+ := \{(X,x_n) \in G(T^0_{pij}) \mid X \in \Lambda^+, x_n > \tau\}, \tag{78}$$

$$F(\Omega^+) = \{x \in T^0_{pij} \mid G(x) \in \Omega^+\}. \tag{79}$$

Noting that $(Z,z_n) \in \Omega^+$ implies $(Z,x_n) \in \Omega^+$ for every $x_n > z_n$ with $(Z,x_n) \in G(T^0_{pij})$, it is not hard to verify that $F(\Omega^+) \subset T^0_{pij}$ is the positive end of a transport set from (46–47) and Definition 26. Now $\Lambda^+ \subset \mathbf{R}^{n-1}$ is Borel like $\Psi^\pm$, and $\Omega^+ \subset \mathbf{R}^n$ is Borel by Lemma 25. Fubini's theorem, (70), (73) and (77–78) yield

$$\int_{\Lambda^+} [\Psi^+(X,\tau) - \Psi^-(X,\tau)]\,dX = \int_{\Omega^+} [\hat f^+(x) - \hat f^-(x)]\,dx. \tag{80}$$

On the other hand, $F(\Omega)$ is Borel whenever $\Omega \subset G(T^0_{pij})$ is, since F is one to one and continuous on $G(T^0_{pij})$ [9, §2.2.10]. Choosing the characteristic function $\varphi = \chi_{F(\Omega)}$ in (69–70) yields

$$\int_\Omega \hat f^\pm(x)\,dx = \int_{F(\Omega)} f^\pm(z)\,dz < \infty,$$

hence

$$\int_{\Lambda^+} [\Psi^+(X,\tau) - \Psi^-(X,\tau)]\, dX = \int_{F(\Omega^+)} [f^+(z) - f^-(z)]\, dz \geq 0 \tag{81}$$

from (80). Here the last integral is non-negative by Lemma 27, since $F(\Omega^+)$ was the positive end of a transport set. But the first integrand is negative by (77), so we infer $\mathcal{L}^{n-1}(\Lambda^+) = 0$. Thus (75) holds for a.e. $X \in \mathbf{R}^{n-1}$, depending on our fixed τ. Fubini's theorem then shows it holds for a.e. $(X,\tau) \in \mathbf{R}^n$. Therefore, fix $X_0 \in \mathbf{R}^{n-1}$. The continuous dependence of $\Psi^{\pm}(X_0,\tau)$ on τ implies (75) is not violated for $X = X_0$ and any τ, unless it is violated on an interval of positive measure around τ. Using Fubini again, we conclude for a.e. $X \in \mathbf{R}^{n-1}$ that (75) holds for all τ.

To obtain the equality (76), use compactness to fix $\tau < x_n$ for all $(X, x_n) \in G(T^0_{pij})$, so that $\Psi^{\pm}(X,\tau) = \Psi^{\pm}(X,-\infty)$. We need the reverse inequality to (75), so define

$$\begin{aligned} \Lambda^- :=& \quad \{X \in \mathbf{R}^{n-1} \mid \Psi^+(X,\tau) > \Psi^-(X,\tau)\}, \\ \Omega^- :=& \quad \{(X,x_n) \in G(T^0_{pij}) \mid X \in \Lambda^-\}, \\ F(\Omega^-) =& \quad \{x \in T^0_{pij} \mid G(x) \in \Omega^-\}. \end{aligned}$$

Note that this time Ω^- is independent of τ, whence $(Z, z_n) \in \Omega^-$ implies $(Z, x_n) \in \Omega^-$ if $(Z, x_n) \in G(T^0_{pij})$. It follows that $F(\Omega^-) \subset T^0_{pij}$ is a complete transport set (and not merely its positive end). Repeating the same argument as before, Lemma 27 implies

$$\int_{\Lambda^-} [\Psi^+(X,\tau) - \Psi^-(X,\tau)]\, dX = \int_{F(\Omega^-)} [f^+(z) - f^-(z)]\, dz = 0$$

instead of (81). This time the first integrand is strictly positive, so $\mathcal{L}^n(\Lambda^-) = 0$, which completes the proof that mass balance (76) also holds for almost every $X \in \mathbf{R}^{n-1}$.

This balancing of mass (76) is a consistency condition which enables us to solve the one-dimensional transport problem on a.e. vertical line $\{X\} \times \mathbf{R}^1$ separately. We shall use inequality (75) to show the solution maps we construct verify $t_X(x_n) \leq x_n$ on $\mathbf{R}^1$. After that, it will remain only to prove that the resulting map $\hat{s} : \mathbf{R}^{n-1} \times \mathbf{R}^1 \to \mathbf{R}^{n-1} \times \mathbf{R}^1$, defined by $\hat{s}(X, x_n) = (X, t_X(x_n))$, pushes θ^+ forward to θ^-.

Step 4. One-dimensional transport. Fix $X \in \mathbf{R}^{n-1}$ for which (75–76) hold. We will construct a map $t_X(x) \leq x$ on $\mathbf{R}^1$ which pushes $\hat{f}^+(X, x_n)\, dx_n$ forward to $\hat{f}^-(X, x_n)\, dx_n$. Note that this map is not unique: among the possible solutions we are free (and we elect) to choose the unique non-decreasing, lower semicontinuous map. But this choice is arbitrary; the only important thing is that our choices are consistent enough on different rays that we end up with a measurable map on $\mathbf{R}^n$.

We define t_X using the distribution functions $\Psi^\pm(X,\tau)$. 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 (76), there exists some $\zeta \in \mathbf{R}^1$ which satisfies

(82)
$$\Psi^+(X,\tau) := \int_\tau^\infty \hat{f}^+(X,x_n)\,dx_n = \int_\zeta^\infty \hat{f}^-(X,x_n)\,dx_n =: \Psi^-(X,\zeta).$$

Of course ζ need not be unique, since $\Psi^-(X,\,\cdot\,)$ will not decrease strictly where $\hat{f}^-$ vanishes. But the set of all ζ satisfying (82) form a closed segment (or half-line). If we define

$$t_X(\tau) := \inf \{\zeta \in \mathbf{R}^1 \mid \Psi^+(X,\tau) \geq \Psi^-(X,\zeta)\} \tag{83}$$
$$= \sup \{\zeta \in \mathbf{R}^1 \mid \Psi^+(X,\tau) < \Psi^-(X,\zeta)\}, \tag{84}$$

then monotonicity of $\Psi^\pm(X,\,\cdot\,)$ shows t_X non-decreasing, and the equivalence of (83) to (84). Lower semicontinuity of $t_X(\,\cdot\,)$ follows from (83) and continuity of $\Psi^\pm(X,\,\cdot)$, while $t_X(\tau) \leq \tau$ follows from (75). Finally, we claim the map t_X pushes $d\theta_X^+ := \hat{f}^+(X,x_n)\,dx_n$ forward to $d\theta_X^- := \hat{f}^-(X,x_n)\,dx_n$, or equivalently

$$\theta_X^+((t_X)^{-1}(A)) = \theta_X^-(A) \quad \text{for each Borel set } A \subset \mathbf{R}^1. \tag{85}$$

Indeed, when A is a half-line $(-\infty,\zeta)$, then (85) follows directly from the conditions (82–84) defining t_X. But half-lines generate all Borel sets, so (85) is established.

Step 5. $\hat{s}$ pushes θ^+ forward to θ^-. Here $\hat{s} : \mathbf{R}^{n-1} \times \mathbf{R}^1 \to \mathbf{R}^{n-1} \times \mathbf{R}^1$ is defined as $\hat{s}(X,x_n) = (X, t_X(x_n))$, where $t_X(x_n) \leq x_n$ is from Step 4.

First we prove $\hat{s}$ is Borel. It is enough to show that the function $t : \mathbf{R}^{n-1} \times \mathbf{R}^1 \to \mathbf{R}^1$ defined as $t(X,x_n) := t_X(x_n)$ is Borel. For each $\zeta \in \mathbf{R}^1$ set

$$T_\zeta := \{(X,\tau) \in \mathbf{R}^{n-1} \times \mathbf{R}^1 \mid t(X,\tau) > \zeta\}, \qquad \text{and} \tag{86}$$
$$M_\zeta := \{(X,\tau) \in \mathbf{R}^{n-1} \times \mathbf{R}^1 \mid \Psi^+(X,\tau) < \Psi^-(X,\zeta)\}. \tag{87}$$

Observe that M_ζ is Borel since $\Psi^\pm$ are, and $M_\zeta \supseteq M_{\zeta+\varepsilon}$ for $\varepsilon > 0$. We shall prove $T_\zeta = M_\zeta$, to conclude that $t(X,\tau)$ is Borel on $\mathbf{R}^n$.

Indeed, if $(X,\tau) \in T_\zeta$, then $t_X(\tau) > \zeta$ and we must have $(X,\tau) \in M_\zeta$ to avoid contradicting (83). Thus $T_\zeta \subseteq M_\zeta$. Conversely, let $(X,\tau) \in M_\zeta$. Continuity of $\Psi^-(X,\,\cdot\,)$ in (87) shows $(X,\tau) \in M_{\zeta+\varepsilon}$ for some $\varepsilon > 0$. Thus $t(X,\tau) \geq \zeta + \varepsilon$ by (84), and $(X,\tau) \in T_\zeta$ to complete the proof that $T_\zeta = M_\zeta$.

Having shown $\hat{s}$ is Borel throughout $\mathbf{R}^{n-1}\times\mathbf{R}^1$, the fact that $\theta^+(\hat{s}^{-1}A) = \theta^-(A)$ for each Borel $A \subset \mathbf{R}^n$ follows from (85) by Fubini's theorem. Thus $\hat{s}$ pushes θ^+ forward to θ^-. 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 in (68) to yield an optimal map $s : \mathbf{R}^n \to \mathbf{R}^n$ for Problem 1. $\square$

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 with Proceedings of a Centro Internazionale Matematico Estivo Summer School in the Springer-Verlag Lecture Notes in Mathematics Series.

[3] K. Ball, E.A. Carlen, and E.H. Lieb. Sharp uniform convexity and smoothness inequalities for trace norms. *Invent. Math.* 115:463–482, 1994.

[4] L. Caffarelli. Allocation maps with general cost functions. In P. Marcellini et al, editor, *Partial Differential Equations and Applications*, number 177 in Lecture Notes in Pure and Appl. Math., pages 29–35. Dekker, New York, 1996.

[5] L.C. Evans. Partial differential equations and Monge-Kantorovich mass transfer. In R. Bott et al., editors, *Current Developments in Mathematics*, pages 26–78. International Press, Cambridge, 1997.

[6] L.C. Evans. Partial Differential Equations. Graduate Studies in Mathematics 19. American Mathematical Society, Providence, 1998.

[7] L.C. Evans and W. Gangbo. Differential equations methods for the Monge-Kantorovich mass transfer problem. *Mem. Amer. Math. Soc.*, 137:1–66, 1999.

[8] L.C. Evans, R.F. Gariepy. Measure theory and fine properties of functions. *CRC Press*, Boca Raton, 1992.

[9] H. Federer. *Geometric Measure Theory.* Springer-Verlag, New York, 1969.

[10] H. Federer. Curvature Measures. *Trans. Amer. Math. Soc.*, 93:418–491, 1959.

[11] M. Feldman. Variational evolution problems and nonlocal geometric motion. *Arch. Rat. Mech. Anal.*, 146:221–274, 1999.

[12] M. Feldman, R.J. McCann. Uniqueness and transport density in Monge's mass transportation problem. Submitted to *Calc. Var. Partial Differential Equations.*

[13] M. Feldman, R.J. McCann. Monge's transport problem on a Riemannian manifold. Submitted to *Trans. Amer. Math. Soc.*

[14] W. Gangbo and R.J. McCann. Optimal maps in Monge's mass transport problem, *C.R. Acad. Sci. Paris Sér. I Math.* 321:1653–1658, 1995.

[15] W. Gangbo and R.J. McCann. The geometry of optimal transportation. *Acta Math.*, 177:113–161, 1996.

[16] P.R. Halmos. The decomposition of measures. *Duke Math. J.*, 8:386–392, 1941.

[17] L. Kantorovich. On the translocation of masses. *C.R. (Doklady) Acad. Sci. URSS (N.S.)*, 37:199–201, 1942.

[18] R.J. McCann. Polar factorization of maps on Riemannian manifolds. To appear in *Geom. Funct. Anal.*

[19] 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.

[20] S.T. Rachev and L. Rüschendorf. *Mass Transportation Problems.* Probab. Appl. Springer-Verlag, New York, 1998.

[21] V.A. Rokhlin. On the fundamental ideas of measure theory. *Mat. Sbornik (N.S.)*, 25(67):107–150, 1949.

[22] W. Rudin. *Real and Complex Analysis.* McGraw-Hill Book Company, New York, 1987.

[23] V.N. Sudakov. Geometric problems in the theory of infinite-dimensional probability distributions. *Proc. Steklov Inst. Math.*, 141:1–178, 1979.

[24] N.S. Trudinger, X.-J. Wang. On the Monge mass transfer problem. To appear in *Calc. Var. Partial Differential Equations.*

ABSTRACT. Given two densities on $\mathbf{R}^n$ with the same total mass, the Monge transport problem is to find a Borel map $s : \mathbf{R}^n \longrightarrow \mathbf{R}^n$ rearranging the first distribution of mass onto the second, while minimizing the average distance transported. Here distance is measured by a norm with a uniformly smooth and convex unit ball. This paper gives a complete proof of the existence of optimal maps under the technical hypothesis that the distributions of mass be compactly supported. The maps are not generally unique. The approach developed here is new, and based on a geometrical change-of-variables technique offering considerably more flexibility than existing approaches.

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF TEXAS AT AUSTIN, AUSTIN TX 78712-1082 USA.

E-mail address: `caffarel@math.utexas.edu`

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF WISCONSIN, MADISON WI 53706 USA.

E-mail address: `feldman@math.wisc.edu`

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF TORONTO, TORONTO ONTARIO CANADA M5S 3G3.

E-mail address: `mccann@math.toronto.edu`