Monge's transport problem on a Riemannian manifold*†

Mikhail Feldman
Department of Mathematics, University of Wisconsin,
Madison, Wisconsin USA 53706 feldman@math.wisc.edu
and
Robert J. McCann
Department of Mathematics, University of Toronto,
Toronto, Ontario, Canada M5S 3G3 mccann@math.toronto.edu

March 26, 2001

Abstract

Monge's problem refers to the classical problem of optimally transporting mass: given Borel probability measures $\mu^+ \neq \mu^-$, find the measure preserving map $s : M \longrightarrow M$ between them which minimizes the average distance transported. Set on a complete, connected, Riemannian manifold M — and assuming absolute continuity of μ^+ — an optimal map will be shown to exist. Aspects of its uniqueness are also established.

* *2000 Mathematics Subject Classification.* Primary 49Q20, 28A50

† 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.

The mass transportation problem, formulated by Gaspard Monge in 1781, is to move one distribution of mass onto another as efficiently as possible, where Monge's original criterion for *efficiency* was to minimize the average distance transported [15]. For a discussion of the problem, its history, recent results, and applications, we refer the reader to Evans [6], Ambrosio [1] and Villani [20]. The purpose of this paper is to investigate the existence and uniqueness of an optimal solution when the problem is set on a Riemannian manifold. Existence is resolved by extending a method developed in recent work of Caffarelli-Feldman-McCann [3] and Trudinger-Wang [19] on Euclidean space. The uniqueness results parallel Euclidean investigations from our paper [11] and Ambrosio [1]. Previous approaches to the existence question were given in Euclidean space by Sudakov [18] and Evans and Gangbo [7].

Fix a C^3 smooth, geodesically complete, connected Riemannian manifold M. Denote by $d(x,y)$ the geodesic distance distance between points x and $y \in M$. Let μ^+ and μ^- be two Borel measures on M, absolutely continuous with respect to the volume measure on M, i.e., $d\mu^\pm = f^\pm(x) d\,\mathrm{vol}(x)$, where $f^\pm \in L^1(M, dx)$ are Borel functions. We assume that total masses of μ^+ and μ^- are finite and equal:

$$\mu^+(M) = \mu^-(M) < \infty, \tag{1}$$

which also can be written as

$$\int_M f^+(x) d\,\mathrm{vol}(x) = \int_M f^-(x) d\,\mathrm{vol}(x). \tag{2}$$

We assume that f^+ and f^- have compact supports and study the following problem:

Problem 1 *Find a mapping $s : M \to M$ which minimizes the functional*

$$I[r] = \int_M d(x, r(x))\, d\mu^+(x) \tag{3}$$

among all Borel maps $r \in \mathcal{A}(\mu^+, \mu^-)$ which push the measure μ^+ forward to μ^-, meaning each continuous function $\phi \in C(M)$ satisfies

$$\int_M \phi(r(x)) d\mu^+(x) = \int_M \phi(y) d\mu^-(y). \tag{4}$$

Our main result is

Theorem 1 (Existence of Optimal Maps) *Fix two $L^1(M, dx)$ densities $f^+, f^- \geq 0$ with compact support and the same total mass (2). Then there exists a Borel map $s : M \longrightarrow M$ which solves Monge's problem, in the sense that it minimizes the average distance (3) transported among all maps pushing f^+ forward to f^- (4).*

The proof follows an argument developed in the Euclidean setting by Caffarelli-Feldman-McCann [3] and independently Trudinger and Wang [19]. Our starting point is slightly different, however, since we do not begin from a solution to the minimization problem (3) in which the integrand has been replaced by a strictly convex function

d^p $(p > 1)$ of the geodesic distance $d(x, r(x))$ [14]. We instead solve directly a dual problem, whose solution $u : M \to \mathbf{R}^1$, called a *Kantorovich potential*, is a Lipschitz function with Lipschitz constant one. This potential determines a set of *transport rays* — paths from $\operatorname{supp} \mu^+$ to $\operatorname{supp} \mu^-$ — along which u decreases with maximum admissible rate. In the case of a complete Riemannian manifold, the transport rays are length minimizing segments of geodesics.

We show that any optimal map takes each transport ray into itself. Then we restrict the measures μ^+ and μ^- onto each ray so that the mass balance holds, and solve a one-dimensional transportation problem on each ray. Thus we get an optimal map on each ray, hence on M.

In order to restrict the measures onto transport rays, we introduce a measure-decomposing change of variables by defining coordinates $x_1, \ldots, x_n$ on certain subsets of M, such that x_n measures distance along each ray, and the variables $x_1, \ldots, x_{n-1}$ parametrize rays, or, more precisely, parametrize a fixed level set of u. In a smooth setting, these are the *Gaussian normal coordinates* orthogonal to the level set of u. But in general, the lack of smoothness of u will prevent us from constructing these coordinates globally, or even on small open balls. We therefore decompose the set of all rays into a countable collection of special Borel subsets, chosen so that the rays enjoy a more "regular" structure within each subset, and perform a Lipschitz change of variables on each subset separately. These 'coordinate neighborhoods' are not generally open, and therefore cannot be among the coordinate neighborhoods which define the differentiable structure on M.

It is important to know the restrictions of measures on rays which emerge from Fubini's theorem applied in the above coordinates are absolutely continuous with respect to arclength on each ray. For this the change of variables must be Lipschitz locally. This is a subtle point in dimensions $n \geq 3$. Indeed, as discovered by Ambrosio [1], Alberti, Kircheim and Preiss, it is the source of a gap in the original solution to Monge's problem on $\mathbf{R}^n$ proposed by Sudakov [18]. We therefore need to establish a Lipschitz control on the directions of rays passing near each point in M. The proof of this estimate is a main technical difficulty distinguishing the present Riemannian problem from the Euclidean case studied in [3] and [19].

The remainder of the paper is organized as follows. In section 1 we solve the Kantorovich dual problem on a general metric space (which in particular applies to any Riemannian manifold). In section 2 we study the geometry of transport rays on M. In section 3 we prove the Lipschitz control on the directions of transport rays. In sections 4–6 we define the measure decomposing change of variables and briefly describe the construction of optimal map, referring to [3] for details. In section 7 we extend to Riemannian manifolds the results of our work [11] on uniqueness both of the optimal map satisfying a one-dimensional monotonicity condition, and of the *transport density*: we show that the cost-flow density generated along geodesics by any optimal map is an $L^1(M)$ function which does not depend on the specific choice of optimal map. An alternate approach to uniqueness of the Euclidean transport density may be found in Ambrosio [1].

1 Background on dual problems

In this section we recall a problem formulated by Kantorovich as a dual to Monge's problem. We construct its well-known solution, and extract various properties germane to our purposes. This part of the construction is most naturally set in an abstract metric space (i.e., not necessarily a complete Riemannian manifold metrized by the geodesic distance), so for this section alone we work in this more general setting. Throughout the other sections of the paper we revert to a Riemannian manifold.

Let (M, d) be a complete separable metric space, where M is a set, and $d(\cdot,\cdot)$ is a distance on M. Let μ^+, μ^- be two Borel measures on M with compact support, satisfying (1). We formulate Monge's Problem #1 for measures μ^+, μ^- on the metric space (M, d).

The corresponding dual problem is:

Problem 2 (Dual) *Let μ^+, μ^- be the measures in Problem 1. Denote the support of μ^+ by $\mathcal{X}$ and μ^- by $\mathcal{Y}$. Among all pairs of functions $\varphi \in C(\mathcal{X})$, $\psi \in C(\mathcal{Y})$ satisfying*

$$\varphi(x) + \psi(y) \leq d(x, y) \quad \textit{for all} \quad x \in \mathcal{X},\ y \in \mathcal{Y}, \tag{5}$$

find a pair (φ_0, ψ_0) maximizing the functional

$$K(\varphi, \psi) = \int_{\mathcal{X}} \varphi(x)\, d\mu^+(x) + \int_{\mathcal{Y}} \psi(y)\, d\mu^-(y). \tag{6}$$

In fact, the duality assertion $I[s] = K(\phi_0, \psi_0)$ which relates these two problems holds much more generally; see Rachev and Rüschendorf [16]. But when the triangle inequality is satisfied, then — as is also well-known and we shall shortly see — the Dual Problem 3 is equivalent to the problem as stated by Kantorovich [12]: Let $\mathrm{Lip}_1(M, d)$ denote the set of functions on M which are Lipschitz continuous with Lipschitz constant no greater than one; i.e.

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

Problem 3 (Kantorovich) *Maximize $\hat{K}[v]$ on $Lip_1(M, d)$, where*

$$\hat{K}[v] = \int_M v(d\mu^+ - d\mu^-).$$

We begin by demonstrating the equivalence of Problems 2 and 3, which we combine with compactness of $\mathrm{Lip}_1(M, d)$ to deduce that the maximum of (6) is attained. The argument is a modification of a proof from Evans review article [6, Lemma 9.1].

Proposition 2 (Lipschitz Maximizer) *Let (M, d) be a complete separable metric space with two finite Borel measures μ^+, μ^- having compact support $\mathcal{X}$, $\mathcal{Y} \subset M$ and the same mass (1). Then a maximizing pair (φ_0, ψ_0) exists for Problem 2 satisfying*

$$\varphi_0 = u \quad on \ \ \mathcal{X}, \quad \psi_0 = -u \quad on \ \ \mathcal{Y}, \tag{7}$$

with $u \in Lip_1(M,d)$. *Moreover,*

$$\varphi_0(x) = \inf_{y\in\mathcal{Y}}(d(x,y) - \psi_0(y)) \quad \textit{for any } \; x \in \mathcal{X}; \tag{8}$$

$$\psi_0(y) = \inf_{x\in\mathcal{X}}(d(x,y) - \varphi_0(x)) \quad \textit{for any } \; y \in \mathcal{Y}. \tag{9}$$

Proof. Let φ, ψ satisfy (5). For $x \in \mathcal{X}$ define

$$\hat\varphi(x) = \inf_{y\in\mathcal{Y}}(d(x,y) - \psi(y)). \tag{10}$$

Then

$$\hat\varphi(x) + \psi(y) \le d(x,y) \tag{11}$$

$$\hat\varphi(x) \ge \varphi(x) \tag{12}$$

hold for all $x \in \mathcal{X}$ and $y \in \mathcal{Y}$, where (5) was used in the last inequality. Define

$$\hat\psi(y) = \inf_{x\in\mathcal{X}}(d(x,y) - \hat\varphi(x)) \quad \text{for any } \; y \in M. \tag{13}$$

¿From (13) and (11)

$$\hat\varphi(x) + \hat\psi(y) \le d(x,y) \qquad (x \in \mathcal{X}, \; y \in M), \tag{14}$$

$$\hat\psi(y) \ge \psi(y) \qquad (y \in \mathcal{Y}). \tag{15}$$

Now, by (10) and (15),

$$\hat\varphi(x) \ge \inf_{y\in\mathcal{Y}}(d(x,y) - \hat\psi(y)) \qquad (x \in \mathcal{X}).$$

So, by (14), we get for all $x \in \mathcal{X}$

$$\hat\varphi(x) = \inf_{y\in\mathcal{Y}}(d(x,y) - \hat\psi(y)). \tag{16}$$

Now we can extend $\hat\varphi$ from $\mathcal{X}$ to the whole space M using the right-hand side of (16) as the definition of $\hat\varphi(x)$. Thus (16) holds for all $x \in M$.

Next, we prove $\hat\varphi \in \mathrm{Lip}_1(M,d)$ [being an infimum of Lipschitz functions of x]. Indeed, let $x^* \in M$ and $\varepsilon > 0$. By (16), there exists $y^* \in \mathcal{Y}$ such that

$$\hat\varphi(x^*) > d(x^*,y^*) - \hat\psi(y^*) - \varepsilon.$$

For any $x \in M$, (16) yields

$$\hat\varphi(x) \le d(x,y^*) - \hat\psi(y^*),$$

so

$$\hat\varphi(x) - \hat\varphi(x^*) \le d(x,y^*) - d(x^*,y^*) + \varepsilon \le d(x^*,x) + \varepsilon.$$

Since the above inequality holds for any $x^*, x \in M$ and $\varepsilon > 0$, it follows that $\hat{\varphi} \in \mathrm{Lip}_1(M, d)$. Similarly $\hat{\psi} \in \mathrm{Lip}_1(M, d)$.

Now we show that $\hat{\varphi} + \hat{\psi} = 0$ on the compact set $\mathcal{X}$. Let $z \in \mathcal{X}$ be such that $\hat{\varphi}(z) + \hat{\psi}(z) < 0$. By (16), (13) and continuity of $\hat{\varphi}$, $\hat{\psi}$, there exist $x \in \mathcal{X}$, $y \in \mathcal{Y}$ such that

$$\hat{\varphi}(z) = d(z, y) - \hat{\psi}(y),$$

$$\hat{\psi}(z) = d(z, x) - \hat{\varphi}(x).$$

¿From this we get, using $\hat{\varphi}(z) + \hat{\psi}(z) < 0$:

$$d(z, y) + d(z, x) = \hat{\varphi}(z) + \hat{\psi}(z) + \hat{\varphi}(x) + \hat{\psi}(y) < \hat{\varphi}(x) + \hat{\psi}(y) \leq d(x, y),$$

contradicting the triangle inequality. Thus we have $\hat{\varphi} + \hat{\psi} \geq 0$ on $\mathcal{X}$. Combining this with (14) yields $\hat{\varphi} + \hat{\psi} = 0$ on $\mathcal{X}$. Thus, denoting $u = -\hat{\psi}$ we have $u \in Lip_1(M, d)$,

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

Note that

$$\hat{K}[u] = K[\hat{\varphi}, \hat{\psi}] \geq K[\varphi, \psi] \tag{17}$$

by (12) and (15). Thus the maximum in the Kantorovich Problem 3 will be just as large as the maximum in Problem 2, despite the fact that it is taken over a class of functions $\{(v, -v) \mid v \in Lip_1(M, d)\}$ more restricted than (5). In this sense the two problems are equivalent. Moreover, for any maximizing (φ, ψ) in Problem 2, there is another maximizing pair $(\hat{\varphi}, \hat{\psi})$ satisfying all properties asserted in the proposition; indeed $u = -\hat{\psi}$ also maximizes Problem 3.

It remains to prove the existence of a maximizing pair (φ, ψ). Therefore, let $\{v_n\}$ be a maximizing sequence for Problem 3. The mass balance condition (1) ensures that adding a constant to v_n does not change $\hat{K}[v_n]$, so we can fix any $z \in M$ and assume that $v_n(z) = 0$ for all n. Then, since $\mathcal{X}$, $\mathcal{Y}$ are compact and $v_n \in \mathrm{Lip}_1(M, d)$ for all n, the v_n are uniformly bounded on $\mathcal{X} \cup \mathcal{Y}$. Also, the v_n are equi-Lipschitz. The Ascoli-Arzela theorem yields a subsequence $v_{n(k)}$ uniformly convergent on $\mathcal{X} \cup \mathcal{Y}$ to some $v \in \mathrm{Lip}_1(\mathcal{X} \cup \mathcal{Y})$. Clearly $\hat{K}[v] = \lim_k \hat{K}[v_{n(k)}]$ is maximal. Define (φ, ψ) by

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

Then the pair (φ, ψ) maximizes Problem 2, since $K[\varphi, \psi] = \sup_{u \in \mathrm{Lip}_1(M,d)} \hat{K}[u]$. $\square$

Remark 3 (Equivalence of Two Dual Problems) *At (17) we showed the function $u \in Lip_1(M, d)$ defined in the statement of the proposition to be a maximizing solution of Problem 3:*

$$\hat{K}[u] = \sup_{v \in Lip_1(M,d)} \hat{K}[v].$$

In this sense Problems 2 and 3 are equivalent, and we refer to them both as duals to Monge's problem. We call a solution of Problem 3 a Kantorovich potential. *In addition, from (7–9)*

$$\begin{aligned} u(x) &= \min_{y \in \mathcal{Y}}(u(y) + d(x,y)) \quad \textit{for any } \ x \in \mathcal{X}; \\ u(y) &= \max_{x \in \mathcal{X}}(u(x) - d(x,y)) \quad \textit{for any } \ y \in \mathcal{Y}. \end{aligned} \tag{18}$$

The next lemma exhibits the connection between the primal and dual problems. 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 (19). The rest of this paper is devoted to carrying out this program on a space M which is a complete Riemannian manifold with geodesic distance $d(x, y)$.

Lemma 4 (Duality) *Fix $u \in Lip_1(M,d)$ and let $s : M \to M$ be a mapping which pushes μ^+ forward to μ^-. If*

$$u(x) - u(s(x)) = d(x, s(x)) \quad \textit{for } \mu^+ \ \textit{a.e.} \ x \in \mathcal{X} \tag{19}$$

then:

i. u is a Kantorovich potential maximizing Problem 3.

ii. s is an optimal map in Problem 1.

iii. The infimum in Problem 1 is equal to the supremum in Problem 3.

iv. Every optimal map $\hat{s}$ and Kantorovich potential $\hat{u}$ also satisfy (19).

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

$$\begin{aligned} I[r] &= \int_M d(x, r(x)) d\mu^+(x) \\ &\geq \int_M [v(x) - v(r(x))] d\mu^+(x) \\ &= \int_M v(x) d\mu^+(x) - \int_M v(y) d\mu^-(y) \\ &= \hat{K}[v]. \end{aligned} \tag{20}$$

using (4). 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 $\mathrm{Lip}_1(M, d)$. On the other hand, our hypothesis (19) produces a case of equality $I[s] = \hat{K}[u]$ in (20). 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 $v \in Lip_1(M, d)$ 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 (20), so the proof of assertion iv. and hence the lemma, are complete. $\square$

2 Transport rays and their geometry

Section 1 reduced the problem of finding an optimal map in Monge's problem to constructing an admissible map which also satisfies (19). We carry out this program for the complete Riemannian manifold M metrized by the geodesic distance $d(x, y)$. Our starting point is a solution $u \in \mathrm{Lip}_1(M, d)$ of the Kantorovich dual Problem 3. In this section, we study the geometric meaning of condition (19), and introduce the transport rays, which in the case of manifold are segments of distance-minimizing geodesics, and transport sets which are ultimately used to construct an optimal map.

Let M be a complete n-dimensional Riemannian manifold. Let TM and T^*M be the tangent and cotangent bundles of M. We denote the Riemannian metric by $\langle \cdot, \cdot \rangle$, i.e., for $p \in M$ the scalar product on T_pM is $\langle \cdot, \cdot \rangle_p$. We denote by $|\xi|_p = \sqrt{\langle \xi, \xi \rangle_p}$ the associated norm on T_pM. The derivative of the function $\varphi : M \to \mathbf{R}^1$ at $x \in M$ is denoted by $D\varphi(x) \in T^*_x M$. The gradient of φ at x is denoted by $\nabla\varphi(x) \in T_xM$. According to the Hopf-Rinow theorem [5, §7.2.8], the geodesic $d(x, y)$ distance between points x and y makes (M, d) a complete separable metric space.

Fix two measures μ^+ and μ^- defined by non-negative densities $f^+, f^- \in L^1(M)$ with respect to the volume measure on M, and satisfying the mass balance condition (2). Assume that μ^+ and μ^- have compact supports, denoted by $\mathcal{X}$ and $\mathcal{Y} \subset M$ respectively. Through the remainder of the paper, we fix a Kantorovich potential u solving the Kantorovich problem and satisfying (18). Such a u exists by Proposition 2, Remark 3. Note that u has Lipschitz constant one with respect to the geodesic distance $d(x, y)$.

Since we want to investigate the geometrical implications of (19) for u, suppose $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ satisfy

$$u(x) - u(y) = d(x, y).$$

By the Hopf-Rinow Theorem [5, §7.2.8] a minimizing geodesic $\sigma : [0, 1] \to M$ links x to y, and is given by $\sigma(\tau) = \exp_x(\tau\dot\sigma(0))$. Note that, since M is a complete manifold, the curve $\sigma(\tau)$ is defined by the above formula for any $\tau \in \mathbf{R}^1$. ¿From the Lipschitz constraint

$$|u(z_1) - u(z_2)| \leq d(z_1, z_2) \quad \text{for any } \ z_1, z_2 \in M, \tag{21}$$

it follows that on any minimizing geodesics σ from x and y the function u is decreasing with the maximum rate compatible with (21), i.e.,

$$u(z) = u(y) + d(z, y) \qquad \text{for any } \ z \in \sigma([0, 1]).$$

We will call maximal segments of geodesics having these properties the *transport rays.* More precisely:

Definition 5 (Transport Rays) *A* transport ray *R is a geodesic $\sigma : [0, 1] \to M$ joining $\sigma(0) = a$ to $\sigma(1) = b$ and having length $\ell(\sigma) = d(a, b)$ such that*

i. $a \in \mathcal{X}$, $b \in \mathcal{Y}$, $a \neq b$;

ii. $u(a) - u(b) = d(a, b)$*;*

iii. Maximality: for any $t > 0$ *such that* $a_t := \exp_a(-t\dot{\sigma}(0)) \in \mathcal{X}$ *there holds*

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

and for any $t > 0$ *such that* $b_t := \exp_b(t\dot{\sigma}(1)) \in \mathcal{Y}$ *there holds*

$$|u(b_t) - u(a)| < d(a, b_t).$$

We call the points a and b the *upper* and *lower ends* of R, respectively. Since $u(a) - u(b) = d(a, b)$, it follows from ii. and (21) that any point $z \in R$ satisfies

$$u(z) = u(b) + d(z, b) = u(a) - d(a, z). \tag{22}$$

Definition 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*

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

¿From these two definitions and the property (18) of u we immediately infer the following lemma, whose obvious proof is omitted.

Lemma 7 (Data is Supported Only on Transport Rays) $\mathcal{X} \cup \mathcal{Y} \subseteq T_0 \cup T_1$*.*

We also note the following:

Lemma 8 (Transport Set is Compact) *The set* $T_0 \cup T_1$ *is compact.*

Proof. By Hopf-Rinow theorem [5, §7.2.8], it is enough to show that $T_0 \cup T_1$ is closed and bounded. The function $v : M \times M \to \mathbf{R}^1$ defined by $v(x, y) = u(x) - u(y)$ is continuous, and thus v attains maximum $L < \infty$ on the compact set $\mathcal{X} \times \mathcal{Y}$. Note that $L \geq 0$. Indeed, if $\mathcal{X} \cap \mathcal{Y}$ is nonempty, then for any $x \in \mathcal{X} \cap \mathcal{Y}$ we have $(x, x) \in \mathcal{X} \times \mathcal{Y}$ and $v(x, x) = 0$. Otherwise, it follows from Lemma 7 that T_1 is nonempty, i.e., there exists a transport ray R. If $a \in \mathcal{X}, b \in \mathcal{Y}$ are the upper and lower ends of R, then $v(a, b) = d(a, b) > 0$.

If the set

$$A := (T_0 \cup T_1) \setminus (\mathcal{X} \cup \mathcal{Y})$$

is nonempty, then any $z \in A$ lies on a transport ray R_z. Denoting again by a, b the upper and lower ends of R_z, we have

$$d(a, z) + d(z, b) = d(a, b) = v(a, b) \leq L,$$

and thus A lies in the union of L-neighborhoods of the compact sets $\mathcal{X}$ and $\mathcal{Y}$. Thus $T_0 \cup T_1$ is bounded.

Now we prove that $T_0 \cup T_1$ is closed. Let a sequence $z_n \in T_0 \cup T_1$ converge to $z \in M$. If an infinite subsequence of z_n lies in $\mathcal{X} \cup \mathcal{Y}$ then we obtain $z \in \mathcal{X} \cup \mathcal{Y}$ by compactness of $\mathcal{X}$ and $\mathcal{Y}$. Thus we can assume that $z_n \in A = (T_0 \cup T_1) \setminus (\mathcal{X} \cup \mathcal{Y})$. Then each z_n lies in the interior of a transport ray R_n, with upper and lower endpoints a_n, b_n. Extracting a subsequence, we obtain $a_{n_j} \to a \in \mathcal{X}$, $b_{n_j} \to b \in \mathcal{Y}$. Since for each n_j there holds

$$d(z_{n_j}, a_{n_j}) + d(z_{n_j}, b_{n_j}) = d(a_{n_j}, b_{n_j}) = u(a_{n_j}) - u(b_{n_j}),$$

it follows that

$$d(z, a) + d(z, b) = d(a, b) = u(a) - u(b),$$

and thus either z lies on a transport ray, or $z = a = b \in T_0$. In either case $z \in T_0 \cup T_1$. $\square$

To study the properties of rays, let us call a point $z \in M$ an *interior point* of a minimizing geodesic $\sigma : [0, 1] \to M$ from a to b, where $a, b \in M$, if $z = \sigma(t)$ for some $0 < t < 1$. By a slight abuse of notations, we denote by σ the set $\sigma([0, 1])$, and by σ^0 the set of interior points of σ. Similar to the case of $\mathbf{R}^n$, we show that transport rays on M do not cross in interior points:

Lemma 9 (Transport Rays Are Disjoint) *Let R_1, R_2 be transport rays, let $R_1 \neq R_2$ and $R_1 \cap R_2 \neq \emptyset$. Then either*

i. $R_1 \cap R_2 = \{c\}$ and c is either the upper end of both rays, or the lower end of both rays, or else

ii. $R_1 \cap R_2 = \{c_l, c_u\}$ where c_u is the upper end of both rays, and c_l is the lower end of both rays.

In particular, an interior point of a transport ray cannot lie on any other transport ray.

Proof. First note that if

$$d(x, y) + d(y, z) = d(x, z)$$

holds then y lies on a minimizing geodesic from x to z.

Since R_i, $i = 1, 2$, is a transport ray, it follows that R_i is a minimizing geodesic $\sigma_i : [0, 1] \to M$, where the parametrization is chosen so that the function $\tau \to u(\sigma_i(\tau))$ is decreasing on $[0, 1]$.

Suppose $R_1 \neq R_2$ share a point c, and let $c = \sigma_1(\tau_1) = \sigma_2(\tau_2)$, where $\tau_i \in [0, 1]$. Then the vectors $\dot\sigma_1(\tau_1)$, $\dot\sigma_2(\tau_2) \in T_cM$ cannot be collinear: since $R_i \subset \{\exp_c(t\dot\sigma_i(\tau_i)) \mid t \in \mathbf{R}^1\}$ for $i = 1, 2$, the maximality part of Definition 5 would force $R_1 = R_2$.

Suppose $R_1 \cap R_2$ contains at least two points, denote them $\{c_l, c_u\}$, where $u(c_u) > u(c_l)$. Then $u(c_u) = u(c_l) + d(c_l, c_u)$. Since tangent vectors to R_1 and R_2 at c_l are not collinear, it follows that the segments of R_1 and R_2 between c_l and c_u do not coincide.

Then c_u lies in the cut locus of c_l, since R_1 and R_2 are minimizing geodesics. Thus c_l and c_u are endpoints of both R_1 and R_2, and $R_1 \cap R_2 = \{c_l, c_u\}$.

It remains to consider the case when R_1 and R_2 have only one common point, i.e. $R_1 \cap R_2 = \{c\}$. Denote $a_i = \sigma_i(0)$ and $b_i = \sigma_i(1)$ for $i = 1, 2$, i.e., a_i and b_i are the upper and the lower ends of R_i.

We shall assume $c \neq b_2$ and argue this forces $c = a_1$. Since R_1 has positive length, 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 (22)

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

thus

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

Strict inequality would violate the Lipschitz condition (21). Thus equality must hold, meaning c lies on a minimizing geodesic γ from a_1 to b_2. Suppose $c \neq a_1$. Since both curves γ and σ_1 minimize distance between a_1 and c, either they coincide, or c lies within the cut locus of a_1, which implies that $c = b_2$ since c lies on a minimizing geodesic γ from a_1 to b_2. But this contradicts our assumption $c \neq b_2$. The curve segments of γ and σ_1 must therefore coincide between a_1 and c. Then it follows that $\gamma \subset \{\exp_c(t\dot{\sigma}_1(\tau_1)) \mid t \in \mathbf{R}^1\}$, where $\tau_1 \in [0, 1]$ is such that $c = \sigma_1(\tau_1)$. But then, by maximality of R_1 it follows that $\gamma \subset R_1$, and thus $b_2 \in R_1$ which contradicts our assumption. Thus $c = a_1$. $\square$

Lemma 10 (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, if R is parametrized as $\sigma : [0, 1] \to M$ such that $\dot{\sigma}(t) \neq 0$ and the function $t \to u(\sigma(t))$ is decreasing on $[0, 1]$, and $z_0 = \sigma(t_0)$ for $t_0 \in (0, 1)$, then $\nabla u(z_0) = -\dfrac{\dot{\sigma}(t_0)}{|\dot{\sigma}(t_0)|_{z_0}}$.*

Proof. For $c \in M$ define a function $d_c : M \to \mathbf{R}^1$ by $d_c(z) := d(c, z)$.

Choose z_0 in the interior of R. Choose $a, b \in R$ such that $u(a) > u(z_0) > u(b)$ and $d(a, b) < \frac{r_{inj}}{2}$, where $r_{inj} > 0$ is the infimum of the injectivity radii of points of $T_0 \cup T_1$ (note that r_{inj} is positive since distance to cut locus is a continuous and positive function of a point [5, Chap. 13, Prop. 2.9], and $T_0 \cup T_1$ is a compact set). Then the functions $d_a(\cdot)$, $d_b(\cdot)$ are smooth in a neighborhood of z_0. Since R is a minimizing geodesic and u satisfies (21), it follows that

$$d_b(z) \geq u(z) - u(b) \geq d(a, b) - d_a(z) \quad \text{for any} \quad z \in M, \tag{23}$$

where the second inequality is obtained from $u(a) - u(z) \leq d_a(z)$ and $u(a) = u(b) + d(a, b)$. Since R is a transport ray, (23) holds with equalities for $z = z_0$.

Thus u is differentiable at z_0. In addition, $\nabla d_b(z_0) = -\nabla d_a(z_0) = -\frac{\dot\sigma(t_0)}{|\dot\sigma(t_0)|_{z_0}}$, thus $\nabla u(z_0) = -\frac{\dot\sigma(t_0)}{|\dot\sigma(t_0)|_{z_0}}$. □

3 Lipschitz directions of transport rays

In this section we prove a fact which is crucial for the construction of the measure decomposing change of variables in Section 4. We show 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. Since rays are geodesics, the directions of rays can be defined by the unit tangent to the rays vectors at the points of intersection of rays with the level set mentioned above. Thus in order to define how close the directions of rays are, we need to measure a distance between two vectors on TM. We describe this distance in Section 3.1, prove some estimates in Section 3.2, and prove the Lipschitz dependence of the directions of rays in Section 3.3.

3.1 Metric structure on M and TM

In this subsection we collect the facts from Riemannian geometry which we use through the rest of the paper.

We use the following Riemannian metric on TM described in [13, §1.9.12]:

Let $\tau : TM \to M$ be the tangent bundle of M. Let $\pi : T(TM) \to TM$ be the tangent bundle of TM. Then the Levi-Civita connection on M determines the splitting of $T(TM)$ onto the vertical and horizontal bundles

$$\pi = \pi_V \oplus \pi_H : V \bigoplus H \to TM \tag{24}$$

such that each fiber $T_\xi(TM)$ of $T(TM)$, where $\xi \in T_pM$, is written as the direct sum

$$T_\xi(TM) = V_\xi \bigoplus H_\xi,$$

of the vertical and horizontal subspaces, see e.g. [13, SS1.5.9–10]. The vertical subspace V_ξ is the set of vectors tangent to the fiber T_pM at $\xi \in T_pM$. The horizontal subspace H_ξ is the horizontal lift of T_pM to ξ by the Levi-Civita connection. The spaces V_ξ and H_ξ are canonically identified to the fiber T_pM, i.e., the maps

$$K_V : V_\xi \to T_pM, \qquad d\tau(\xi) : H_\xi \to T_pM, \tag{25}$$

are isomorphisms of linear spaces, where the map K_V is the identification of the tangent space to the fiber T_pM at ξ with the fiber. We define the Riemannian metric $\langle\cdot,\cdot\rangle^{TM}$ on TM as follows: On each $T_\xi(TM)$, where $\xi \in T_pM$, the inner product is defined by letting the horizontal and vertical spaces be orthogonal, and taking on each of these spaces the inner product induced from T_pM with the canonical identifications. Explicitly, let $\xi \in T_pM$, and $\alpha, \beta \in T_\xi(TM)$. Then α and β are uniquely represented as

$$\alpha = \alpha_V + \alpha_H, \quad \beta = \beta_V + \beta_H, \quad \text{where} \quad \alpha_V,\ \beta_V \in V_\xi, \ \text{ and } \ \alpha_H,\ \beta_H \in H_\xi. \tag{26}$$

We define the inner product $\langle \cdot, \cdot \rangle_\xi^{TM}$ on $T_\xi(TM)$ by

$$\langle \alpha, \beta \rangle_\xi^{TM} = \langle K_V \alpha_V, K_V \beta_V \rangle_p + \langle d\tau(\xi)\alpha_H, d\tau(\xi)\beta_H \rangle_p. \tag{27}$$

Let d_{TM} denote the geodesic distance on TM with respect to the Riemannian metric (27).

In the following lemma we collect standard properties of geodesic metrics on M and TM which we use in the rest of the paper:

Lemma 11 *Let M be a complete Riemannian manifold. Then:*

i. Let N be another complete Riemannian manifold, and let $F : M \to N$ be a smooth mapping. Then for any compact set $K \subset M$ there exists $C = C(K)$ such that $d_N(F(p), F(q)) \le C d_M(p, q)$ for any $p, q \in K$;

ii. TM with Riemannian metric (27) is a complete Riemannian manifold;

iii. Let $\gamma : [0, T] \to M$, $\tilde\gamma : [0, \tilde T] \to TM$ be piecewise differentiable curves on M and TM respectively, and $\tau[\tilde\gamma] = \gamma$. Then $l_{TM}(\tilde\gamma) \ge l_M(\gamma)$, where $l_{TM}(\cdot)$ and $l_M(\cdot)$ denote length of curves on TM and M respectively. If, in addition, $\tilde\gamma$ is the horizontal lift of γ, then $l_{TM}(\tilde\gamma) = l_M(\gamma)$.

iv. Let $p, q \in M$ and $\xi \in T_pM$, $\eta \in T_qM$. Then $d_{TM}(\xi, \eta) \ge d(p, q)$. If in addition η is the parallel transport of ξ (with respect to Levi-Civita connection on M) along a minimizing geodesic from p to q on M, then $d_{TM}(\xi, \eta) = d(p, q)$.

v. Let $\tilde\gamma : [0, \tilde T] \to TM$ be a piecewise differentiable curve on TM, and $\lambda \in \mathbf{R}^1$. Let $\tilde\gamma_\lambda$ be the curve on TM defined by $t \to \lambda\tilde\gamma(t)$ for $t \in [0, \tilde T]$. Then

$$l_{TM}(\tilde\gamma_\lambda) \le \max(1, |\lambda|) l_{TM}(\tilde\gamma).$$

In particular, for any $\xi, \eta \in TM$ and $\lambda \in \mathbf{R}^1$,

$$d_{TM}(\lambda\xi, \lambda\eta) \le \max(1, |\lambda|) d_{TM}(\xi, \eta).$$

3.2 Estimates of distances in normal coordinates

Now we derive certain estimates which we use later. Let $a \in M$, let $i(a)$ be the injectivity radius of a. We need to estimate, for $q_1, q_2 \in B_{\frac{i(a)}{2}}(a) \subset M$, how $\exp_a^{-1} q_1 - \exp_a^{-1} q_2$ changes when a changes. Precisely, for points p close enough to a we compare $\xi := \exp_a^{-1} q_1 - \exp_a^{-1} q_2 \in T_aM$ with vector η obtained by parallel transport of

$\exp_p^{-1} q_1 - \exp_p^{-1} q_2 \in T_pM$ to T_aM along the (unique) minimizing geodesic from p to a.

Let $a \in M$ and let $r(a) > 0$ be such that $B_r(a)$ is strongly convex for all $0 < r < 2r(a)$ (i.e., for any q_1, q_2 in the closure of $B_r(a)$ there exists a unique minimizing geodesic from q_1 to q_2, and its interior lies in $B_r(a)$); such $r(a) > 0$ exists and depends continuously on a by [4, Theorem 5.14]. Also we can restrict $r(a) \le 1$.

Define the following sets

$$\begin{aligned} \mathcal{B} &= \{(P,Q) \mid P,Q \in T_aM,\ a \in M,\ |P|_a, |Q|_a < r(a)\} \subset TM \oplus TM; \\ \mathcal{C} &= \{(P,a) \mid P \in T_pM,\ a,p \in M,\ p \in B_{r(a)}(a)\} \subset TM \times M \end{aligned} \tag{28}$$

Let $\Phi : \mathcal{B} \to TM$ be a map. We denote by Φ_a the restriction of Φ on $\mathcal{B} \cap (T_aM \oplus T_aM)$.

Define the following maps: ¿From the definition of $r(a)$ it follows that all three are well-defined and smooth:

i. $F : \mathcal{B} \to TM$ defined by $F(P,Q) \equiv F_a(P,Q) = \exp^{-1}_{\exp_a P}(\exp_a Q)$ where $a \in M,\ P,Q \in B_{r(a)}(0) \subset T_aM$;

ii. $\Pi : \mathcal{C} \to TM$ defined by $\Pi(P,a) \equiv \Pi_p(P,a)$, where $\Pi_p(P,a) \in T_aM$ is the parallel translate of the vector $P \in T_pM$ along the unique minimizing geodesic from p to a.

iii. Finally, let $\Phi : \mathcal{B} \to TM$ be defined by $\Phi(P,Q) \equiv \Phi_a(P,Q) = \Pi(F(P,Q),a)$ where $P,Q \in B_{r(a)}(0) \subset T_aM$.

Lemma 12 *The map $\Phi : \mathcal{B} \to TM$ is C^2 smooth, and satisfies the following properties: for any $a \in M$ and $P,Q \in B_{r(a)}(0) \subset T_aM$:*

$$\Phi_a(\mathcal{B}) \equiv \Phi(\mathcal{B} \cap (T_aM \oplus T_aM)) \subset T_aM, \tag{29}$$

$$\Phi(0,Q) = Q, \tag{30}$$

$$\Phi(P,P) = 0 \in T_aM, \tag{31}$$

$$|\Phi(P,Q)|_a = d(\exp_a P, \exp_a Q). \tag{32}$$

Moreover, let $\theta : [0,1] \to M$ be the unique minimizing geodesic from $\exp_a P$ to $\exp_a Q$ and let the curve $\gamma : [0,1] \to T_aM$ from P to Q be defined by $\gamma(t) = \exp_a^{-1}(\theta(t))$ for $t \in [0,1]$. Then $\Phi(P, \gamma([0,1]))$ is a line segment in T_aM from $\Phi(P,P) = 0$ to $\Phi(P,Q)$ (i.e., for $P \in T_aM$ the mapping $\Phi_a(P,\cdot) \circ \exp_a^{-1}$ maps geodesics passing through $\exp_a P$ into straight lines in T_aM passing through the origin).

Proof. ¿From the definition of F it follows that for any $a \in M,\ P \in B_{r(a)}(0) \subset T_aM$

$$\begin{aligned} &F(P,\cdot) : B_{r(a)}(0) \to T_{\exp_a P}M; \\ &F(0,P) = \exp^{-1}_{\exp_a 0}(\exp_a P) = P; \\ &F(P,P) = \exp^{-1}_{\exp_a P}(\exp_a P) = 0 \in T_{\exp_a P}M \end{aligned} \tag{33}$$

Thus for any $P, Q \in B_{r(a)}(0) \subset T_aM$ we have $F(P,Q) \in T_{\exp_a P}M$, and from definition of F there holds $\exp_{\exp_a P} F(P,Q) = \exp_a Q$. This implies that

$$|F(P,Q)|_{\exp_a P} = d(\exp_a P, \exp_a Q) \quad \text{for any} \quad P,\ Q \in B_{r(a)}(0) \subset T_aM. \tag{34}$$

Let $P, Q \in B_{r(a)}(0) \subset T_aM$, and let $\theta : [0,1] \to M$ be the unique minimizing geodesic from $\exp_a P$ to $\exp_a Q$. If the curve $\gamma : [0,1] \to T_aM$ from P to Q is defined by $\gamma(t) = \exp_a^{-1}(\theta(t))$, then $F(P, \gamma([0,1]))$ is a line segment in the space $T_{\exp_a P}M$ from $F(P,P) = 0$ to $F(P,Q)$. Indeed, by the definition of F,

$$F(P, \gamma([0,1])) = \exp^{-1}_{\exp_a P}(\theta([0,1])),$$

which is a line segment since $\theta : [0,1] \to M$ is a geodesic through $\exp_a P$.

Thus the map F has all required properties. It remains, for each $a \in M$, to map the set $F_a(P, B_{r_1}(0)) \subset T_{\exp_a P}M$ back to T_aM.

We use the map $\Pi(\cdot, a) : T_pM \to T_aM$. By properties of the Levi-Civita connection, if $p \in B_{r(a)}(a)$ and $\xi \in T_pM$, then

$$|\Pi(\xi, a)|_a = |\xi|_p. \tag{35}$$

¿From its definition, Φ is smooth as a composition of smooth maps.

The properties of Π and (33) imply (29)–(31): indeed, (29) is obvious, and to get (30)–(31) we compute for $P \in T_aM$

$$\begin{aligned}
\Phi(0,P) &= \Pi\Big(F(0,P), a\Big) = \Pi(P,a) = P; \\
\Phi(P,P) &= \Pi\Big(F(P,P), a\Big) = \Pi(0_{\exp_a P}, a) = 0 \in T_aM;
\end{aligned}$$

Also, (34), (35) and the first line of (33) imply (32):

$$|\Phi(P,Q)|_a = |\Pi(F(P,Q),a)|_a = |F(P,Q)|_{\exp_a P} = d(\exp_a P, \exp_a Q).$$

The remaining assertion, i.e., that geodesics passing through $\exp_a P$ are mapped by $\Phi(P,\cdot) \circ \exp_a^{-1}$ into straight lines through $0 \in T_aM$, follows from the corresponding property of the map F, and from the linearity of the parallel translation map $\Pi_p(\cdot, a) : T_pM \to T_aM$. □

Let $a \in M$ and $P_1, P_2 \in B_{r(a)}(0) \subset T_aM$. By (29), the $\Phi_a(\cdot, P_2)$ and $\Phi_a(P_1, \cdot)$ are mappings from an open subset of the Euclidean space $T_aM \simeq \mathbf{R}^n$ into T_aM. For $k = 1,2$ we denote by $D_k\Phi$ the derivative of $\Phi(P_1,P_2)$ with respect to the variables P_k. Thus for any multi-index $\beta = (\beta_1, \beta_2)$ such that $k = |\beta| = \beta_1 + \beta_2$, the derivative $D_\beta \Phi_a(P_1,P_2) \equiv D_1^{\beta_1} D_2^{\beta_2} \Phi_a(P_1,P_2)$ is a k-linear map from T_aM into T_aM. Since Φ is smooth on $\mathcal{B}$, we have the following estimates of derivatives:

Lemma 13 *Let M be a C^N-smooth complete Riemannian manifold. For any compact set $K \subset M$ there exists $C > 0$ such that for any $a \in K$, any multi-index $\beta = (\beta_1, \beta_2)$ satisfying $k = |\beta| < N$, any $P_1, P_2 \in B_{\frac{1}{2}r(a)} \subset T_aM$, $\xi_1, \ldots, \xi_k \in T_aM$*

$$|D_\beta \Phi_a(P_1, P_2)(\xi_1, \ldots, \xi_k)|_a \leq C \prod_{j=1}^{k} |\xi_j|_a. \tag{36}$$

Proof. Fix $p \in M$. Introduce local coordinates $x : B_{r(p)}(p) \to U \subset \mathbf{R}^n$ on $B_{r(p)} \subset M$, and corresponding coordinates $(x, v) : \tau^{-1}(B_{r(p)}(p)) \subset TM \to U \times \mathbf{R}^n$. The Riemannian structure on M induces a scalar product on each $\{a\} \times \mathbf{R}^n$, where $a \in U$ corresponds to $a \in B_{r(p)}(p) \subset M$. Let $\mathcal{U} \subset U \times \mathbf{R}^n \times \mathbf{R}^n$ be the image of $\mathcal{B}$ under the coordinate mapping. In coordinates Φ is a smooth mapping $\Phi : \mathcal{U} \to U \times \mathbf{R}^n$ of the form $\Phi(a, P, Q) = (a, \varphi(a, P, Q))$, where $\varphi : \mathcal{U} \to \mathbf{R}^n$ is smooth. Let $\mathcal{U}' \subset \mathcal{U}$ be the image of the set $\{(P, Q) \mid P, Q \in T_aM,\ a \in B_{\frac{r(p)}{2}}(p),\ |P|_a, |Q|_a < \frac{1}{2}r(a)\}$ under the coordinate mapping. Fixing a Euclidean structure $\langle \cdot, \cdot \rangle$ on $\mathbf{R}^n$, we obtain existence of C_β such that for any $(a, P_1, P_2) \in \overline{\mathcal{U}'}$, $\xi_1, \ldots, \xi_k \in \mathbf{R}^n$

$$|D_\beta \varphi(a, P_1, P_2)(\xi_1, \ldots, \xi_k)| \leq C_\beta \prod_{j=1}^{k} |\xi_j|, \tag{37}$$

where $|\xi| = \sqrt{\langle \xi, \xi \rangle}$. Indeed, the constant C_β is obtained by taking supremum of the left-hand side of the inequality (37) over the compact set

$$\{(a, P_1, P_2, \xi_1, \ldots, \xi_k) \mid (a, P_1, P_2) \in \overline{\mathcal{U}'},\ |\xi_1| = \ldots = |\xi_k| = 1\}.$$

Now, since the functions $a \to \sup\limits_{|\xi|=1} \dfrac{|\xi|_a}{|\xi|}$ and $a \to \inf\limits_{|\xi|=1} \dfrac{|\xi|_a}{|\xi|}$ are continuous on $B_{r(p)}(p)$, we obtain (36) with C depending continuously on $a \in \overline{B_{\frac{r(p)}{2}}(p)}$. $\square$

Finally we prove the uniform estimates on the distortion of distances in normal coordinates.

Lemma 14 *Let M be a C^3 smooth complete Riemannian manifold. Let K be a compact subset of M. There exists C such that for any $a \in K$ and $P, Q_1, Q_2 \in B_{\frac{1}{2}r(a)}(0) \subset T_aM$*

$$\Big|[\Phi(P, Q_1) - \Phi(P, Q_2)] - (Q_1 - Q_2)\Big|_a \leq C|P|_a|Q_1 - Q_2|_a, \tag{38}$$

and for any $a \in K$ and $P, Q \in B_{\frac{1}{2}r(a)}(0) \subset T_aM$

$$|d(\exp_a P, \exp_a Q) - |P - Q|_a| \leq C|P - Q|_a(|P|_a + |P - Q|_a), \tag{39}$$

Proof. Let $a \in K$, $P, Q_1, Q_2 \in T_aM$, $|P|_a, |Q_1|_a, |Q_2|_a \leq \frac{1}{2}r(a)$. For $k = 1, 2$ we use (30) to obtain

$$\begin{aligned}\Phi(P, Q_k) &= \Phi(0, Q_k) + \int_0^1 D_1\Phi(tP, Q_k)Pdt \\ &= Q_k + \int_0^1 D_1\Phi(tP, Q_k)Pdt, \qquad (40)\end{aligned}$$

where $tP \in B_{\frac{1}{2}r(a)}(0)$ for $0 \leq t \leq 1$. Then we estimate for any $t \in [0, 1]$, using (36):

$$\begin{aligned}|D_1\Phi(tP, Q_1)P &- D_1\Phi(tP, Q_2)P|_a \\ &= \left|\int_0^1 D_1D_2\Phi(tP, \tau Q_1 + (1-\tau)Q_2)(P, Q_1 - Q_2)d\tau\right|_a \\ &\leq C\,|P|_a|Q_1 - Q_2|_a,\end{aligned}$$

and thus using (40) we get (38).

Now we prove (39). Let $a \in K$ and $P, Q \in B_{\frac{1}{2}r(a)}(0) \subset T_aM$. Using (31) we obtain

$$\begin{aligned}\Phi(P, Q) &= \Phi(P, P) + D_2\Phi(P, P)(P - Q) + \int_0^1 D_2^2\Phi(P, \tilde{Q}(t))(P - Q, P - Q)(1 - t)dt \\ &= D_2\Phi(P, P)(P - Q) + \int_0^1 D_2^2\Phi(P, \tilde{Q}(t))(P - Q, P - Q)(1 - t)dt,\end{aligned}$$

where $\tilde{Q}(t) := tP + (1 - t)Q \in B_{\frac{1}{2}r(a)}(0)$ for $0 \leq t \leq 1$. Since, by (30), $D_2\Phi(0, P) = Id_{T_aM}$, we compute

$$\begin{aligned}D_2\Phi(P, P)(P - Q) &= D_2\Phi(0, P)(P - Q) + \int_0^1 D_1D_2\Phi(tP, P)(P, P - Q)dt \\ &= (P - Q) + \int_0^1 D_1D_2\Phi(tP, P)(P, P - Q)dt,\end{aligned}$$

and $tP \in B_{\frac{1}{2}r(a)}(0)$ for $0 \leq t \leq 1$. Thus using (36) with $C = C(T_0 \cup T_1)$ we have

$$|\Phi(P, Q) - (P - Q)|_a \leq C|P - Q|_a(|P|_a + |P - Q|_a),$$

and, using (32), we arrive at (39). □

3.3 Directions of transport rays

¿From now on we assume that M is a C^3 smooth complete connected Riemannian manifold and state the main result of Section 3.

Proposition 15 (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 Kantorovich potential $u(y_1) = u(y_2)$, then the ray directions satisfy a Lipschitz bound*

$$d_{TM}(\nabla u(y_1), \nabla u(y_2)) \leq \frac{C}{\sigma} d(y_1, y_2), \tag{41}$$

where $\sigma := \min_{k=1,2}\{\sigma_0, d(y_k, a_k), d(y_k, b_k)\}$, and the constants C and $\sigma_0 > 0$ depend only on the Riemannian manifold M and the compact set $T_0 \cup T_1$.

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

$$\begin{aligned} u(z_k) &= u(y_1) + \sigma, & (42)\\ u(x_k) &= u(y_1) - \sigma, & (43)\\ d(x_k, z_k) &= 2\sigma, & (44)\\ d(x_k, y_k) &= d(y_k, z_k) = \sigma & (45) \end{aligned}$$

for $k = 1, 2$. We prove first the following:

Lemma 16 *There exists $C, \sigma_0 > 0$ depending only on M, $T_0 \cup T_1$ such that if σ in (42–45) satisfies $0 < \sigma < \sigma_0$, then*

$$d(x_1, x_2),\ d(z_1, z_2) \leq C d(y_1, y_2). \tag{46}$$

Proof. We can assume without loss that

$$d(x_1, x_2) \leq d(z_1, z_2). \tag{47}$$

Otherwise we consider $-u(x)$ instead of $u(x)$ and reverse the directions of rays.

We also can assume that

$$d(y_1, y_2) < \sigma. \tag{48}$$

Indeed, if $d(y_1, y_2) \geq \sigma$ then

$$d(z_1, z_2) \leq d(z_1, y_1) + d(y_1, y_2) + d(y_2, z_2) = d(y_1, y_2) + 2\sigma \leq 3d(y_1, y_2)$$

and the lemma is proved.

By (45), (47–48)

$$\begin{aligned} d(x_1, y_2) &\leq d(x_1, y_1) + d(y_1, y_2) \leq 2\sigma, & (49)\\ d(x_1, x_2) &\leq d(z_1, z_2) \leq d(z_1, y_1) + d(y_1, y_2) + d(y_2, z_2) \leq 3\sigma, & (50) \end{aligned}$$

and so

$$\begin{aligned} d(x_1, z_2) &\leq d(x_1, x_2) + d(x_2, z_2) \leq 5\sigma & (51)\\ d(z_1, x_2) &\leq d(z_1, z_2) + d(z_2, x_2) \leq 5\sigma & (52) \end{aligned}$$

By [4, Theorem 5.14], for every compact subset K of M there exists $r_0 > 0$ depending only on M, K, such that for any $p \in K$, $r < r_0$ the geodesic ball $B_r(p)$ is strongly convex. From this and the fact that the distance to the cut locus is a continuous function on M it follows that for any $p \in M$ there exists $r_1 > 0$, depending continuously on p, such that:

i. $20r_1 \le r_0(M, \overline{B_{r_1}(p)})$

ii. for any $q \in B_{r_1}(p)$ there holds $B_{20r_1}(p) \subset B_{i(q)}(q)$, where $i(q)$ is the injectivity radius of q.

Thus, since $T_0 \cup T_1$ is a compact set, we can find $r_1 = r_1(T_0 \cup T_1) > 0$, such that (i) and (ii) hold with this r_1 for all $p \in T_0 \cup T_1$.

Let

$$20\sigma_0 \le r_1(T_0 \cup T_1). \tag{53}$$

By (42–45) and (47–52), the points $\{x_k\}$, $\{y_k\}$, $\{z_k\}$ for $k = 1, 2$ lie in $B_{\frac{r_1}{2}}(x_1)$.

By our choice of r_1, the map $\exp_{x_1}^{-1} : B_{r_1}(x_1) \to B^t_{r_1}(0) = \exp_{x_1}^{-1}(B_{r_1}(x_1)) \subset T_{x_1}M$ is a diffeomorphism. Here $B^t_r(0) \subset T_{x_1}M$ denotes the ball of radius r and center 0 in the Euclidean space $(T_{x_1}M, \langle\cdot,\cdot\rangle_{x_1})$. Denote for $k = 1, 2$

$$X_k = \exp_{x_1}^{-1} x_k, \quad Y_k = \exp_{x_1}^{-1} y_k, \quad Z_k = \exp_{x_1}^{-1} z_k.$$

Using (44), (45), we get

$$Y_1 = \frac{1}{2}(X_1 + Z_1). \tag{54}$$

In addition,

$$X_1 = 0,$$

and

$$|\exp_{x_1}^{-1} p|_{x_1} = d(p, x_1) \quad \text{for any} \quad p \in B_{r_1}(x_1). \tag{55}$$

By (55) and (42)–(45):

$$|Z_1 - X_1|_{x_1} = d(x_1, z_1) = u(z_1) - u(x_1) = u(z_2) - u(x_1) \le d(x_1, z_2) = |Z_2 - X_1|_{x_1}.$$

Squaring this inequality yields

$$|Z_1 - X_1|^2_{x_1} \le |(Z_2 - Z_1) + (Z_1 - X_1)|^2_{x_1} = |Z_2 - Z_1|^2_{x_1} + 2\langle Z_2 - Z_1, Z_1 - X_1\rangle_{x_1} + |Z_1 - X_1|^2_{x_1},$$

from where

$$\langle Z_2 - Z_1, Z_1 - X_1\rangle_{x_1} \ge -\frac{1}{2}|Z_2 - Z_1|^2_{x_1},$$

and finally, using (54), we get

$$\langle Z_2 - Z_1, Z_1 - Y_1\rangle_{x_1} \ge -\frac{1}{4}|Z_2 - Z_1|^2_{x_1}. \tag{56}$$

We need to obtain a similar estimate of $\langle Z_1 - Z_2, Z_2 - Y_2\rangle_{x_1}$. By (42)–(45)

$$d(x_2, z_2) = u(z_2) - u(x_2) = u(z_1) - u(x_2) \le d(x_2, z_1) \tag{57}$$

However, since $d(x_2, z_2)$ and $d(x_2, z_1)$ are generally not equal to $|Z_2 - X_2|_{x_1}$ and $|Z_1 - X_2|_{x_1}$ respectively, we cannot repeat the previous argument. To estimate this distortion of distances, we will use the map Φ introduced in Lemma 12. Note that from definition of $r_1(T_0 \cup T_1)$ in (53) and of $r(a)$ in (28), it follows that $r_1(T_0 \cup T_1) \le \frac{1}{20} r(a)$ for any $a \in T_0 \cup T_1$. ¿From (44), (45), and (47–52) we obtain using (55)

$$|X_k|_{x_1},\ |Y_k|_{x_1},\ |Z_k|_{x_1} \le 5\sigma. \tag{58}$$

Thus, by (53) and (28), X_k, Y_k, Z_k are in the domain of Φ, and in the estimates below the conditions of Lemma 14 are always satisfied.

Denote

$$\begin{aligned} \hat{Z}_k &= \Phi(X_2, Z_k) \quad \text{for} \quad k = 1, 2, \\ \hat{Y}_2 &= \Phi(X_2, Y_2), \\ \hat{X}_2 &= \Phi(X_2, X_2). \end{aligned}$$

Since X_k, Y_k, $Z_k \in T_{x_1}M$ then, by (30), $\hat{X}_k$, $\hat{Y}_k$, $\hat{Z}_k \in T_{x_1}M$. By (31)

$$\hat{X}_2 = 0.$$

Using (32) with $a = x_1$, we get

$$\begin{aligned} |\hat{Z}_k - \hat{X}_2|_{x_1} &= d(\exp_{x_1} Z_k, \exp_{x_1} X_2) = d(z_k, x_2), \quad \text{for} \quad k = 1, 2, \\ |\hat{Y}_2 - \hat{X}_2|_{x_1} &= d(\exp_{x_1} Y_2, \exp_{x_1} X_2) = d(y_2, x_2). \end{aligned} \tag{59}$$

By (44–45), the point y_2 lies on the minimizing geodesic from x_2 to z_2. Thus Lemma 12 implies that $\hat{Y}_2$ lies on the line segment between $\hat{X}_2$ and $\hat{Z}_2$. Using (44–45) and (59), we conclude that

$$\hat{Y}_2 = \frac{1}{2}(\hat{X}_2 + \hat{Z}_2) \tag{60}$$

From (57) and (59) we obtain

$$|\hat{Z}_1 - \hat{X}_2|_{x_1} \ge |\hat{Z}_2 - \hat{X}_2|_{x_1}.$$

Squaring this inequality, and repeating, with use of (60), the argument which led to (56), we obtain

$$\langle \hat{Z}_1 - \hat{Z}_2, \hat{Z}_2 - \hat{Y}_2 \rangle_{x_1} \ge -\frac{1}{4} |\hat{Z}_2 - \hat{Z}_1|^2_{x_1}. \tag{61}$$

In the following estimate we use the inequality

$$-|P|^2 \ge -(1+\varepsilon)|Q|^2 - \frac{2}{\varepsilon}|P - Q|^2, \quad \text{for any } \ P, Q \in \mathbf{R}^n,\ \varepsilon \in (0, 1], \tag{62}$$

where $|\cdot|$ is a norm in $\mathbf{R}^n$ defined by a scalar product $\langle \cdot, \cdot \rangle$. The (62) is easily checked: by expanding and rearranging terms, we rewrite (62) as

$$(1 - \frac{\varepsilon}{2})|P|^2 + (1 + \frac{\varepsilon}{2} + \frac{\varepsilon^2}{2})|Q|^2 \ge 2\langle P, Q \rangle,$$

and this is true since $(1-\frac{\varepsilon}{2})(1+\frac{\varepsilon}{2}+\frac{\varepsilon^2}{2}) = 1+\frac{\varepsilon^2}{4}-\frac{\varepsilon^3}{4} \geq 1$ if $\varepsilon \in (0,1]$.

We substitute $\hat{Z}_1 - \hat{Z}_2 = (Z_1 - Z_2) + [(\hat{Z}_1 - \hat{Z}_2) - (Z_1 - Z_2)]$ and $\hat{Z}_2 - \hat{Y}_2 = (Z_2 - Y_2) + [(\hat{Z}_2 - \hat{Y}_2) - (Z_2 - Y_2)]$ into the left-hand side of (61). We estimate the right-hand side of (61) from below using (62) with $P = \hat{Z}_1 - \hat{Z}_2$, $Q = Z_1 - Z_2$. Thus we get from (61)

$$\begin{aligned}\langle Z_1 - Z_2, Z_2 - Y_2\rangle_{x_1} \geq\ & -\tfrac{1}{4}(1+\varepsilon)|Z_2 - Z_1|^2_{x_1} \\ & -\tfrac{1}{2\varepsilon}|(Z_1 - Z_2) - (\hat{Z}_1 - \hat{Z}_2)|^2_{x_1} \\ & -|(Z_1 - Z_2) - (\hat{Z}_1 - \hat{Z}_2)|_{x_1}|Z_2 - Y_2|_{x_1} \\ & -|Z_1 - Z_2|_{x_1}|(Z_2 - Y_2) - (\hat{Z}_2 - \hat{Y}_2)|_{x_1} \\ & -|(Z_1 - Z_2) - (\hat{Z}_1 - \hat{Z}_2)|_{x_1}|(Z_2 - Y_2) - (\hat{Z}_2 - \hat{Y}_2)|_{x_1}\end{aligned} \tag{63}$$

for any $\varepsilon \in (0,1]$.

Now we estimate the error terms in (63). We use Lemma 14 with the compact set $K = T_0 \cup T_1$. In the calculations below C will denote different constants depending only on M and $T_0 \cup T_1$.

¿From (58) and (39) it follows that if σ_0 is chosen small depending on M, $T_0 \cup T_1$, and $\sigma \leq \sigma_0$, then:

$$\frac{1}{2}d(y_1, y_2) \leq |Y_1 - Y_2|_{x_1} \leq 2d(y_1, y_2), \tag{64}$$

$$\frac{1}{2}d(z_1, z_2) \leq |Z_1 - Z_2|_{x_1} \leq 2d(z_1, z_2) \tag{65}$$

By (55), $|X_1 - X_2|_{x_1} = d(x_1, x_2)$, and thus by (47), (65)

$$|X_1 - X_2|_{x_1} \leq 2|Z_1 - Z_2|_{x_1}. \tag{66}$$

Using (66), (58), (38) we estimate:

$$|Z_2 - Y_2|_{x_1} \leq |Z_2|_{x_1} + |Y_2|_{x_1} \leq 10\sigma;$$

$$|(Z_1 - Z_2) - (\hat{Z}_1 - \hat{Z}_2)|_{x_1} \leq C|X_1 - X_2|_{x_1}|Z_1 - Z_2|_{x_1} \leq C|Z_1 - Z_2|^2_{x_1};$$

$$|(Z_2 - Y_2) - (\hat{Z}_2 - \hat{Y}_2)|_{x_1} \leq C|X_1 - X_2|_{x_1}|Z_2 - Y_2|_{x_1} \leq C\sigma|Z_1 - Z_2|_{x_1}.$$

Using these estimates and (58) and recalling that $\sigma < 1$ we conclude that (63) implies

$$\langle Z_1 - Z_2, Z_2 - Y_2\rangle_{x_1} \geq -\frac{1}{4}(1 + \varepsilon + C\sigma + C\frac{\sigma^2}{\varepsilon})|Z_2 - Z_1|^2_{x_1} \tag{67}$$

Choosing first $\varepsilon = \frac{1}{2}$, and then reducing if necessary $\sigma_0 > 0$ to achieve $C\sigma_0 + C\dfrac{\sigma_0^2}{\varepsilon} \leq \dfrac{1}{2}$ where $C = C(T_0 \cup T_1)$ is from (67) (and thus σ_0 depends only on $T_0 \cup T_1$), we obtain for $\sigma \leq \sigma_0$

$$\langle Z_1 - Z_2, Z_2 - Y_2\rangle_{x_1} \geq -\frac{1}{2}|Z_2 - Z_1|^2_{x_1}. \tag{68}$$

Now we can finish the proof of the lemma. Combining (56) and (68), we estimate

$$\begin{aligned} |Z_2 - Z_1|_{x_1} |Y_2 - Y_1|_{x_1} &\geq \langle Z_2 - Z_1, Y_2 - Y_1 \rangle_{x_1} \\ &= \langle Z_2 - Z_1, (Y_2 - Z_2) + (Z_2 - Z_1) + (Z_1 - Y_1) \rangle_{x_1} \\ &\geq -\frac{1}{2}|Z_2 - Z_1|^2_{x_1} + |Z_2 - Z_1|^2_{x_1} - \frac{1}{4}|Z_2 - Z_1|^2_{x_1} \\ &= \frac{1}{4}|Z_2 - Z_1|^2_{x_1}. \end{aligned}$$

Thus

$$|Z_2 - Z_1|_{x_1} \leq 4|Y_2 - Y_1|_{x_1} \tag{69}$$

Using (64) and (65) we conclude the proof of Lemma 16. □

Finally, we return to the proof of Proposition 15. We choose σ_0 from Lemma 16 and assume that $\sigma \leq \sigma_0$ in (42–45). We can assume without loss of generality that (47), (48) hold. Then we also have (50–52) and (53). In particular, every pair of points from y_k, z_k, $k = 1, 2$ is connected by a unique minimizing geodesic.

Denote

$$v_k = \sigma \nabla u(y_k) \in T_{y_k} M \quad k = 1, 2. \tag{70}$$

Then $|v_k|_{y_k} = \sigma$. Let $\tilde{v}_2 \in T_{y_1} M$ be the vector obtained by parallel translation of v_2 along the (unique) minimizing geodesic from y_2 to y_1. Then $|\tilde{v}_2|_{y_1} = |v_2|_{y_2} = \sigma$. Also, by Lemma 11(iv),

$$d_{TM}(v_2, \tilde{v}_2) = d(y_1, y_2). \tag{71}$$

Since the map $\exp : TM \to M$ is smooth, it is locally Lipschitz as a map between the metric spaces (TM, d_{TM}) and (M, d), by Lemma 11(i). Let $C = C(T_0 \cup T_1)$ be its Lipschitz constant on the compact set $\{(p, v) \mid p \in T_0 \cup T_1,\ v \in T_p M,\ |v|_p \leq 2\sigma_0\} \subset TM$. Then we get, using (71)

$$d(\exp_{y_1} \tilde{v}_2, \exp_{y_2} v_2) \leq C d_{TM}(v_2, \tilde{v}_2) = C d(y_1, y_2).$$

Note that $\exp_{y_k} v_k = z_k$, $k = 1, 2$. Thus, using the above inequality and Lemma 16, we obtain

$$d(\exp_{y_1} v_1, \exp_{y_1} \tilde{v}_2) \leq d(z_1, z_2) + d(z_2, \exp_{y_1} \tilde{v}_2) \leq C d(y_1, y_2). \tag{72}$$

Let $\mathcal{U} \subset TM$ be the set $\mathcal{U} = \{(p, \xi) \mid p \in T_0 \cup T_1,\ \xi \in T_p M,\ |\xi|_p < 2\sigma_0\}$. By choice of σ, i.e. by (53), the map $F : TM \to M \times M$ defined by $F(p, \xi) = (p, \exp_p \xi)$ is a diffeomorphism from $\mathcal{U}$ to $F(\mathcal{U}) = \{(p, q) \mid p \in T_0 \cup T_1,\ q \in M,\ d(p, q) < 2\sigma_0\}$ (see, e.g. the proof of [5, §3.3.7]). Thus, by Lemma 11(i), the map F^{-1} is Lipschitz on the compact set $\{(p, q) \mid p \in T_0 \cup T_1,\ q \in M,\ d(p, q) \leq \sigma_0\} \subset F(\mathcal{U})$. Thus there exists a constant C depending only on M, $T_0 \cup T_1$ such that for any $p \in T_0 \cup T_1$, $\xi, \eta \in T_p M$, $|\xi|_p, |\eta|_p \leq \sigma_0$ there holds

$$d_{TM}(\xi, \eta) \leq C d(\exp_p \xi, \exp_p \eta).$$

Using this, we estimate

$$d_{TM}(v_1, \tilde{v}_2) \le Cd(\exp_{y_1} v_1, \exp_{y_1} \tilde{v}_2) \le Cd(y_1, y_2),$$

where the last inequality follows from (72), and constants C depend only on M, $T_0 \cup T_1$. Combining this with (70), (71) we get

$$d_{TM}(\sigma\nabla u(y_1), \sigma\nabla u(y_2)) \le d_{TM}(\sigma\nabla u(y_1), \tilde{v}_2) + d_{TM}(\tilde{v}_2, \sigma\nabla u(y_2)) \le Cd(y_1, y_2).$$

Now, using Lemma 11(v) and condition $\sigma \le \sigma_0$, we obtain (41). □

4 Measure decomposing coordinates on M

In this section we define Lipschitz coordinates $(x_1, \ldots, x_n)$ on M such that $n-1$ of the variables are used to parameterize a given level set of the Kantorovich potential u, while x_n measures distance to this set along the transport rays which pierce it. But the conditions of Proposition 15 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.

Lemma 17 (Bi-Lipschitz Parametrization of Level Sets) *Let $u : M \to \mathbf{R}^1$ be a Lipschitz function, $\sigma \in \mathbf{R}^1$, and S_σ the level set $\{x \in M \mid u(x) = \sigma\}$. Then the set*

$$S_\sigma \cap \{x \in M \mid u \textit{ is differentiable at } x \textit{ and } \nabla u(x) \neq 0\}$$

has a countable covering consisting of Borel sets $S_\sigma^i \subset S_\sigma$, such that for each $i \in \mathbf{N}$ there exist Lipschitz maps $U : S_\sigma^i \to \mathbf{R}^{n-1}$ and $V : \mathbf{R}^{n-1} \to M$ (i.e., Lipschitz maps between metric spaces (M, d) and $(\mathbf{R}^{n-1}, |\cdot|)$, where $|\cdot|$ is a Euclidean metric on $\mathbf{R}^{n-1}$), satisfying

$$V(U(z)) = z \quad \textit{for all} \quad z \in S_\sigma^i. \tag{73}$$

Also, the $U(S_\sigma^i)$ are Borel subsets of $\mathbf{R}^{n-1}$.

Proof. We can cover the manifold M by a countable collection of bounded coordinate neighborhoods $\{M_k\}$. Thus it is enough to prove the Lemma for each set $S_\sigma \cap M_k$. Fix k.

Let $\phi : M_k \to \mathbf{R}^n$ be a coordinate mapping. Then ϕ is a diffeomorphism from M_k to $\mathcal{U} = \phi(M_k) \subset \mathbf{R}^n$.

Since $\psi^{-1} : \mathcal{U} \to M_k$ is a diffeomorphism, it follows from the proof of Lemma 11(i) that ψ^{-1} is Lipschitz on $\mathcal{U}$, as a map between subsets of metric spaces $(\mathbf{R}^n, |\cdot|)$ and (M, d). Similarly, ψ is Lipschitz on M_k.

Thus the function $v = u \circ \psi^{-1} : \mathcal{U}_l \to \mathbf{R}^1$ is Lipschitz. Also, since ψ is a diffeomorphism from M'_k to $\mathcal{U}$, it follows that u is differentiable at $z \in M_{kl}$ if and only if v is differentiable at $x = \psi(z) \in \mathcal{U}_l$, and that $\nabla u(z) \neq 0$ for $z \in M_{kl}$ if and only if $Dv(x) \neq 0$ where $x = \psi(z)$.

Extend v to the whole space $\mathbf{R}^n$ as a Lipschitz function using Kirszbraun's theorem [9, §2.10.43]. Denote $\hat{S}_\sigma$ the level set $\{x \in \mathbf{R}^n \ \ v(x) = \sigma\}$.

Now, applying [3, Lemma 18] to the function v, we obtain a countable covering of the set

$$\hat{S}_\sigma \cap \{x \in \mathbf{R}^n \mid v \text{ is differentiable at } x \text{ and } Dv(x) \neq 0\}$$

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

$$\hat{V}(\hat{U}(x)) = x \quad \text{for all} \quad x \in \hat{S}^i_\sigma.$$

Since $\hat{U}$ is univalent (i.e., one to one) on $\hat{S}^i_\sigma \cap \mathcal{U}_l$, it follows from Federer [9, §2.2.10, page 67] that $\hat{U}(\hat{S}^i_\sigma \cap \mathcal{U}_l)$ is a Borel subset of $\mathbf{R}^{n-1}$.

Now the sets $S^i_\sigma = \psi^{-1}(\hat{S}^i_\sigma \cap \mathcal{U}_l)$ are Borel and cover the set

$$S_\sigma \cap \{x \in M \mid u \text{ is differentiable at } x \text{ and } \nabla u(x) \neq 0\} \cap M_{kl},$$

and the maps $U = \hat{U} \circ \psi : S^i_\sigma \to \mathbf{R}^{n-1}$ and $V = \psi^{-1} \circ \hat{V} : \mathbf{R}^{n-1} \to M$ are Lipschitz and satisfy (73). Also, $U(S^i_\sigma) = \hat{U}(\hat{S}^i_\sigma \cap \mathcal{U}_l)$, and thus $U(S^i_\sigma)$ is Borel. □

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 18 (Ray clusters) *Fix $\sigma \in \mathbf{R}^1$, a Kantorovich potential u, and the Borel cover $\{S^i_\sigma\}_i$ of the level set $S_\sigma := \{x \in M \mid u(x) = p\}$ in Lemma 17. 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 $1/j$. 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 $T_{\sigma ij}$ we are now ready to define the Lipschitz change of variables which inspired the title of this section:

Lemma 19 (Lipschitz Change of Variables) *Each ray cluster $T_{\sigma ij} \subset M$ admits coordinates $G = G_{\sigma ij} : T^0_{\sigma ij} \to \mathbf{R}^{n-1} \times \mathbf{R}^1$ with inverse $F = F_{\sigma ij} : G(T^0_{\sigma ij}) \to M$ satisfying:*

1. *F is a Lipschitz mapping between $G(T^0_{\sigma ij}) \subset \mathbf{R}^{n-1} \times \mathbf{R}^1$ and M;*
2. *for each $\lambda > 0$, G is Lipschitz on $T^\lambda_{\sigma ij} := \{x \in T^0_{\sigma ij} \mid d(a, x), d(b, x) > \lambda\}$, where a and b denote the endpoints of the (unique) transport ray R_x;*

3. $F(G(x)) = x$ *for each* $x \in T^0_{\sigma ij}$;

4. 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*

$$G(x) = (U(z), u(x) - u(z)), \tag{74}$$

where $U : M \to \mathbf{R}^{n-1}$ *gives the Lipschitz coordinates (73) on* S^i_σ.

Remark 20 (Flattening Level Sets) *The final assertion of Lemma 19 implies: (a)* F *maps the part of the hyperplane* $\mathbf{R}^{n-1} \times \{0\}$ *which lies within* $G(T^0_{\sigma ij})$ *onto* S^i_σ*; (b)* F *maps the segment where each "vertical" line* $\{X\} \times \mathbf{R}^1$ *intersects* $G(T^0_{\sigma ij})$ *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 9 shows that rays do not cross in the interior points, while Lemma 10 shows that u increases with rate one along each ray. Thus every point $x \in T^0_{\sigma ij}$ lies on a unique transport ray, and this ray intersects the level set S_σ in a single point $z \in S^i_\sigma$, so the expression (74) defines a map $G : T^0_{\sigma ij} \to \mathbf{R}^{n-1} \times \mathbf{R}^1$ throughout the cluster. It remains to construct the inverse map F on $G(T^0_{\sigma ij}) \subset \mathbf{R}^{n-1} \times \mathbf{R}^1$. Let $(X, x_n) \in G(T^0_{\sigma ij})$, and let V be the map (73) parametrizing S^i_σ. Then the point $V(X) \in S^i_\sigma$ is an interior point of some transport ray R, both of whose endpoints are separated from $V(X)$ by a distance exceeding $1/j$. Define

$$F(X, x_n) := \exp_{V(X)}[x_n \nabla u(V(X))]. \tag{75}$$

That F inverts G (assertion 3) now follows from (73), (74), Lemma 10, and the fact that a ray is a geodesic.

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

$$\Lambda := \{X \in \mathbf{R}^{n-1} \mid (X, 0) \in G(T^0_{\sigma ij})\}. \tag{76}$$

We first claim the scaled ray direction $\frac{1}{j}\nabla u \circ V : \Lambda \to TM$ is a Lipschitz function. Indeed, recalling that $V(X) \in S^i_\sigma$ is separated from the endpoints of $R_{V(X)}$ by a distance greater than $1/j$, we invoke Proposition 15 to conclude that $X, X' \in \Lambda$ satisfy

$$\begin{aligned} d_{TM}\left(\nabla u(V(X)), \nabla u(V(X'))\right) &\leq jC_1\, d(V(X), V(X')) \\ &\leq jC_2\, |X - X'| \end{aligned} \tag{77}$$

because $V : (\mathbf{R}^{n-1}, |\cdot|) \to (M, d)$ was Lipschitz in Lemma 17.

Since $T_0 \cup T_1$ is a compact set, it follows from the definition (74) of G that $(X, x_n) \in G(T^0_{\sigma ij})$ is also bounded, since u and U are Lipschitz on M and S^i_σ respectively. Since $|\nabla u| \equiv 1$ on $T^0_{\sigma ij}$, it follows that $\{x_n \nabla u(V(X)) \mid (X, x_n) \in G(T^0_{\sigma ij})\}$ is a bounded subset of TM. Let $(X, x_n), (X', x'_n) \in G(T^0_{\sigma ij})$. Since $\exp : TM \to M$ is a smooth map, then (75) and Lemma 11(i) yield

$$\begin{aligned} d(F(X,x_n), F(X',x'_n)) &\leq C d_{TM}(x_n \nabla u(V(X)), x'_n \nabla u(V(X'))) \\ &\leq C[d_{TM}(x_n \nabla u(V(X)), x_n \nabla u(V(X'))) \\ &\quad + d_{TM}(x_n \nabla u(V(X')), x'_n \nabla u(V(X')))]. \end{aligned} \tag{78}$$

¿From boundedness of $|x_n|$, Lemma 11(v), and (77)

$$\begin{aligned} d_{TM}(x_n \nabla u(V(X)), x_n \nabla u(V(X')) &\leq jC d_{TM}\left(\nabla u(V(X)), \nabla u(V(X'))\right) \\ &\leq jC\, |X - X'|. \end{aligned} \tag{79}$$

To estimate the last term in (78), we note that, connecting $x_n \nabla u(V(X')$ and $x'_n \nabla u(V(X')$ by a straight line in $T_{V(X')}M$ and using that $|\nabla u(V(X')|_{V(X')} = 1$, we get

$$d_{TM}(x_n \nabla u(V(X'), x'_n \nabla u(V(X')) = |x_n - x'_n|.$$

Combining this with (78), (79) we conclude that F is Lipschitz.

It remains to prove assertion 2 of the lemma. Let $\lambda > 0$. We first show the function $\nabla u : M \to TM$ to be Lipschitz on $T^\lambda_{\sigma ij}$. Being discontinuous at the mutual end of two rays, its Lipschitz constant must depend on λ. Let $x,\, x' \in T^\lambda_{\sigma ij}$ lie on the transport rays R and R'. Assume first that $d(x,x') \geq \lambda/2$. Let $\xi \in T_x M$ be obtained by the parallel translation of $\nabla u(x') \in T_{x'}M$ along the minimizing geodesic from x to x'. Then $|\xi|_x = |\nabla u(x')|_{x'} = 1$, and $d_{TM}(\nabla u(x'), \xi) = d(x,x')$ by Lemma 11(iv). Thus

$$\begin{aligned} d_{TM}(\nabla u(x), \nabla u(x')) \leq d_{TM}(\nabla u(x), \xi) + d_{TM}(\nabla u(x'), \xi) &\leq 2 + d(x,x') \\ &\leq \left(\frac{4}{\lambda} + 1\right) d(x,x'). \end{aligned}$$

Therefore, assume $d(x,x') < \lambda/2$ and hence $|u(x) - u(x')| \leq d(x,x') < \lambda/2$. The point $y' := \exp_{x'}\Big([u(x) - u(x')]\nabla u(x')\Big)$ 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. Thus Proposition 15 yields:

$$d_{TM}\left(\nabla u(x), \nabla u(y')\right) \leq \frac{C}{\lambda} d(x,y'). \tag{80}$$

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

$$d(x',y') = |u(x') - u(y')| = |u(x') - u(x)| \leq d(x',x). \tag{81}$$

Turning to G, we estimate

$$|G(x) - G(x')| \leq |G(x) - G(y')| + |G(y') - G(x')|. \tag{82}$$

Since x' and y' lie on R', definition (74) yields

$$|G(y') - G(x')| = |u(x') - u(y')| = |u(x') - u(x)| \leq d(x, x'). \tag{83}$$

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

$$|G(x) - G(y')| = |U(z) - U(z')|. \tag{84}$$

Setting $\delta := u(z) - u(x)$, we have $z = \exp_x[\delta \nabla u(x)]$ and $z' = \exp_{y'}[\delta \nabla u(y')]$. Also $|\delta|$ is bounded by the diameter of the cluster $T^\lambda_{\sigma ij}$. Because the coordinates U were Lipschitz, we have

$$\begin{aligned} |U(z) - U(z')| &\leq C\, d(\exp_x[\delta \nabla u(x)], \exp_{y'}[\delta \nabla u(y')]) \\ &\leq C\, d_{TM}\, (\nabla u(x), \nabla u(y')) \\ &\leq \frac{C}{\lambda} d(x, x') \end{aligned} \tag{85}$$

from (80), (81), and Lemmas 11(v) and 11(i), and C denotes different constants. Now (82–85) imply G is Lipschitz on $T^\lambda_{\sigma ij}$, to complete the lemma. □

Lemma 21 (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 $T_1 \subset M$. Moreover, each T_{pij} and transport ray R satisfy:*

$$\textit{Either } (R)^0 \subset T_{pij}, \quad \textit{or } (R)^0 \cap T_{pij} = \emptyset. \tag{86}$$

Proof. A transport ray R is a geodesic $\gamma : [0, 1] \to M$, and has positive length by Definition 5. Let $a = \gamma(0)$ and $b = \gamma(1)$ be the lower and upper ends of R. By Lemma 10, $|u(b) - u(a)| = \text{length}(\gamma)$. 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 $S^i_p \subset S_p$ of Lemma 17, 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 9 forces $R = R_z \subset T_{pij}$, to complete the proof of (86). □

Definition 22 (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_{pij} and $G(T^0_{pij})$. Introduce the distance functions to the upper and lower ends of rays:

Lemma 23 (Semicontinuity of Distance to Ray Ends) *At each $z \in \mathbf{R}^n$ define*

$$\alpha(z) := \sup \{d(y,z) \mid y \in \mathcal{Y},\ u(z) - u(y) = d(y,z)\}, \tag{87}$$
$$\beta(z) := \sup \{d(x,z) \mid x \in \mathcal{X},\ u(x) - u(z) = d(x,z)\}, \tag{88}$$

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 (87),

$$\alpha(z_n) - 1/n \leq d(y_n, z_n) = u(z_n) - u(y_n) \tag{89}$$

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 $\alpha_0 \leq d(y,z) = u(z) - u(y) \leq \alpha(z)$ from the limit of (89), which proves the lemma. □

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 on M 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 $z \in M \setminus (T_0 \cup T_1)$, 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 that the ray ends have measure zero. In what follows, n-dimensional Lebesgue measure is denoted $\mathcal{L}^n$, and the volume measure on M is denoted Vol.

Lemma 24 (Measurability of Clusters / Negligibility of Ray Ends) *The ray ends $\mathcal{E} \subset T_1$ form a Borel set of measure zero: Vol$(\mathcal{E}) = 0$. The rays of length zero $T_0 \subset M$ 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 19. In particular the sets $T^0_{\sigma ij}$ and $G[T^0_{\sigma ij}]$ are Borel.*

Proof. First observe that $T_0 = \{z \in M \mid \alpha(z) = \beta(z) = 0\}$ while $\mathcal{E} = \{z \in M \mid \alpha(z)\beta(z) = 0 \text{ but } \alpha(z) + \beta(z) > 0\}$. Both of these sets are Borel by the upper semicontinuity of α and β shown in Lemma 23.

Therefore, fix $\sigma \in \mathbf{R}^1$ and $i, j \in \mathbf{N}$ and recall the Borel set $S^i_\sigma \subset M$ and Lipschitz coordinates $U : S^i_\sigma \longrightarrow \mathbf{R}^{n-1}$ on it from Lemma 17. Let $B \subset S^i_\sigma$ be Borel. By Lemma 17 the sets $U(S^i_\sigma)$ and $U(B)$ are Borel. Moreover, the set Λ defined in (76) is given by

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

according to (74). Similarly, introduce a set

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

Now Definition 18 and Lemma 19(#2) yield that for any $\lambda > 0$

$$G(T^\lambda_{\sigma ij}) = \{(X, x_n) \mid X \in \Lambda,\ -\alpha(V(X)) + \lambda < x_n < \beta(V(X)) - \lambda\},$$
$$G(T^\lambda_{\sigma ij}(B)) = \{(X, x_n) \mid X \in \Lambda(B),\ -\alpha(V(X)) + \lambda < x_n < \beta(V(X)) - \lambda\},$$

where $T^\lambda_{\sigma ij}(B) := T^\lambda_{\sigma ij} \cap T_{\sigma ij}(B)$. Since $V = U^{-1}$ is Lipschitz, the functions $\alpha \circ V$, $\beta \circ V$ are upper semicontinuous in view of Lemma 23. Thus we conclude that both $\Lambda(B) \subset \mathbf{R}^{n-1}$ and $G(T^\lambda_{\sigma ij}(B)) \subset \mathbf{R}^{n-1} \times \mathbf{R}^1$ are Borel. Lemma 19 (#2) shows the transformation G is Lipschitz on $T^\lambda_{\sigma ij}$. Since a Lipschitz map from a subset of any metric space into $\mathbf{R}^1$ can be extended to a Lipschitz map of the whole metric space into $\mathbf{R}^1$ (see, e.g., [9, sect. 2.10.44]), it follows that G extends from $T^\lambda_{\sigma ij}$ to a Lipschitz map $\hat{G} : M \to \mathbf{R}^{n-1} \times \mathbf{R}^1$ (note that, in general, $\hat{G} \neq G$ on $T^0_{\sigma ij} \setminus T^\lambda_{\sigma ij}$). Thus we see that $T^\lambda_{\sigma ij}(B) = \hat{G}^{-1}[G(T^\lambda_{\sigma ij}(B))]$ is a Borel subset of M. Since $T^0_{\sigma ij}(B) = \cup_{k=1}^\infty T^{\frac{1}{k}}_{\sigma ij}(B)$, the set $T^0_{\sigma ij}(B)$ is Borel.

To show the ray ends have measure zero, we use Lemma 21. Fix $p \in \mathbf{Q}$ and $i, j \in \mathbf{N}$, and consider the points $\mathcal{G} \subset \mathbf{R}^{n-1} \times \mathbf{R}^1$ of T_{pij} corresponding to the ray ends 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. By Lemma 19 (#1), F can be extended as a Lipschitz mapping to the closure of $G(T^0_{pij})$ in $\mathbf{R}^{n-1} \times \mathbf{R}^1$, the set $\overline{G(T^0_{pij})}$. Clearly, $\mathcal{G} \subset \overline{G(T^0_{pij})}$. Now $\mathcal{E} \cap T_{pij} = F(\mathcal{G})$. Cover $\mathcal{E} \cap T_{pij}$ by a countable collection of coordinate neighborhoods $M_k \subset M$. It is enough to prove that $\mathrm{Vol}(\mathcal{E} \cap T_{pij} \cap M_k) = 0$ for each k, l. Fix k and l. Let $\phi : M_k \to \mathbf{R}^n$ be a coordinate mapping. Then ϕ is a diffeomorphism from M_k to $\mathcal{U} = \phi(M_k) \subset \mathbf{R}^n$. Thus it follows from the proof of Lemma 11(i) that ϕ is Lipschitz on M_k.

Thus $\hat{F} : \mathcal{G} \cap F^{-1}(M_k) \to \mathbf{R}^n$, defined as $\hat{F} = \psi \circ F$, is Lipschitz. By Kirszbraun's theorem [9, §2.10.43], we can extend $\hat{F}$ to a Lipschitz map $\hat{F} : \mathbf{R}^{n-1} \times \mathbf{R}^1 \to \mathbf{R}^n$. Now we can use $\mathcal{L}^n(\mathcal{G}) = 0$ and the Area formula [9, §3.2.3] to conclude that $\mathcal{L}^n[\hat{F}(\mathcal{G} \cap F^{-1}(M_k))] = 0$. From $\mathcal{E} \cap T_{pij} = F(\mathcal{G})$ and the fact that $\phi : M_k \to \mathcal{U}$ is a diffeomorphism we get $\phi(\mathcal{E} \cap T_{pij} \cap M_k) = \hat{F}(\mathcal{G} \cap F^{-1}(M_k))$. Thus $\mathrm{Vol}[\mathcal{E} \cap T_{pij} \cap M_k] = \mathcal{L}^n[\phi(\mathcal{E} \cap T_{pij} \cap M_k)] = \mathcal{L}^n[\hat{F}(\mathcal{G} \cap F^{-1}(M_k))] = 0$ (and hence $\mathcal{E} \cap T_{pij} \cap M_k$ a measurable set). Since the sets M_k form a countable cover for $T_{pij} \cap \mathcal{E}$, and, by Lemma 21, the clusters $\{T_{pij}\}$ form a countable cover for $\mathcal{E} \subset T_1$, we have $\mathrm{Vol}(\mathcal{E}) = 0$ 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^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 rational clusters of rays T_{pij}, indexed by $p \in \mathbf{Q}$ and $i, j \in \mathbf{N}$, to be disjoint, and so that they still cover all rays. 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 $S_{pij} := T_{pij} \cap S_p$ of $S^i_p \subset M$ on which there are Lipschitz coordinates U, V (73) satisfying

$$V(U(x)) = x \quad \text{for all} \quad x \in S_{pij}. \tag{90}$$

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 (86) 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 19; (again, the original maps F and G work for the modified clusters). The measurability Lemma 24 holds for the new clusters, as follows readily from the modification procedure. Thus we have:

The (modified) clusters of ray interiors T^0_{pij}, where $p \in \mathbf{Q}$ and $i,j \in \mathbf{N}$, are disjoint and cover all transport rays (91)

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

$$\begin{aligned} \mathrm{j}(A) &= \cup_{z \in A \cap T^0_1} R^0_z, \\ \mathrm{j}^+(A) &= \mathrm{j}(A) \cap \{y \mid u(y) \geq \sigma\}, \\ \mathrm{j}^-(A) &= \mathrm{j}(A) \cap \{y \mid u(y) < \sigma\} \end{aligned} \tag{92}$$

where R^0_z is the relative interior of the unique transport ray through z. Thus, $\mathrm{j}(A)$ is the smallest transport set containing $A \cap T^0_1$ 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 25 *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 $B \cap T^0_1$ is Borel, we can replace B by $B \cap T^0_1$, i.e., assume that $B \subset S_\sigma \cap T^0_1$. We have

$$\mathrm{j}(B) = \bigcup_{i,j=1}^{\infty} T^0_{\sigma ij}(B \cap S^i_\sigma).$$

Since both B and S^i_σ are Borel, we use Lemma 24 to conclude the proof. □

5 Detailed mass balance

Definition 26 (Transport Sets) *A set $A \subset M$ 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 segment $[z, a)$ of the transport ray R_z whenever $z \in A \cap (T_1 \setminus \mathcal{E})$ and a denotes the upper end of R_z.*

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

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

Lemma 27 (Detailed mass balance) *Let $A \subset M$ be a Borel transport set. Then*

$$\int_A f^+(x)\, dx = \int_A f^-(x)\, dx. \tag{93}$$

More generally, if a Borel set $A^+ \subset M$ forms the positive end of a transport set, then

$$\int_{A^+} f^+(x)\, dx \geq \int_{A^+} f^-(x)\, dx. \tag{94}$$

Proof. The first statement, in the case of the Euclidean space is [7, Lemma 5.1]. In fact, Evans and Gangbo also show (94) in the proof of [7, Lemma 5.1], in the case of the Euclidean space (precisely, they show (94) for A^+ of a particular form, but the proof works for general positive end of a transport set). The geometry of Euclidean space enters the proof [7, Lemma 5.1] only through the facts that rays do not cross in the interior points and the set of rays endpoints has measure zero. We have shown these properties in Lemmas 9 and 24 for the case of manifold, so both the equality (93) and the inequality (94) follow immediately. □

6 Construction of optimal map

Once we defined measure-decomposing Lipschitz change of variables on the rays clusters Lemma 19 and proved its properties Lemmas 21, 23, 24 and (91), and detailed mass balance Lemma 27, construction of optimal map follows [3, sect. 5].

A sketch of the construction is following:

Since $\mu^+[M \setminus \mathcal{X}] = 0$ and $\mathcal{X} \subset T_0 \cup T_1$ by Lemma 7, we need only define an optimal map s on $T_0 \cup T_1$.

Set $s = id$ on T_0.

Since $T_1 = \cup T_{pij}$, where $p \in \mathbf{Q}$ and $i, j \in \mathbf{N}$, and (91) holds, it is enough to define s on each T_{pij} separately.

Fix T_{pij}, and let F, G be the corresponding maps from Lemma 19. By Lemma 19 the map F is Lipschitz in $\mathbf{R}^n$, one to one on $G(T^0_{pij})$, and $F(G(T^0_{pij})) = T^0_{pij}$. Then Area formula for the case of Riemannian manifolds [9, §3.2.46, 3.2.5] yields

$$\int_{G(T^0_{pij})} \varphi(F(x)) f^{\pm}(F(x)) J_n F(x)\, dx = \int_{T^0_{pij}} \varphi(z) f^{\pm}(z) d\,\mathrm{vol}(z) \tag{95}$$

for any summable $\varphi : M \to \mathbf{R}^1$. Here $J_n F$ denotes the n-dimensional Jacobian of F computed with respect to the Riemannian inner product.

Define $\hat{f}^{\pm} : \mathbf{R}^{n-1} \times \mathbf{R}^1 \to \mathbf{R}^1$ by

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

¿From Lemma 27 one can derive that $\hat{f}^+$ and $\hat{f}^-$ are in mass balance on each vertical line $\tau \to (X, \tau)$, where $X \in \mathbf{R}^{n-1}$, $\tau \in \mathbf{R}^1$. Precisely, let us introduce the distribution function

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

Then for a.e. $X \in \mathbf{R}^{n-1}$

$$\Psi^+(X, \tau) \geq \Psi^-(X, \tau) \geq 0 \tag{96}$$

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

$$\Psi^+(X, -\infty) = \Psi^-(X, -\infty) < \infty. \tag{97}$$

In order to construct an optimal map $s_{pij} : M \to M$ for measures $\mu^{\pm}_{|T_{pij}}$ it is enough by Lemma 4 to construct s_{pij} pushing forward $\mu^+_{|T_{pij}}$ onto $\mu^-_{|T_{pij}}$ and acting down transport rays. To get such s_{pij}, it is enough, by Lemma 19 and (95), to construct $\hat{s} : \mathbf{R}^{n-1} \times \mathbf{R}^1 \to \mathbf{R}^{n-1} \times \mathbf{R}^1$ pushing forward $\hat{f}^+ dx$ onto $\hat{f}^- dx$ and acting down the vertical lines, i.e. of the form $\hat{s}(X, \tau) = (X, t_X(\tau)$ where $X \in \mathbf{R}^{n-1}$, $\tau \in \mathbf{R}^1$ and $t_X : \mathbf{R}^1 \to \mathbf{R}^1$ satisfies $t_X(\tau) \leq \tau$, and then define $s_{pij} = F \circ \hat{s}_{pij} \circ G$.

Thus it remains to construct the map t_X for each $X \in \mathbf{R}^{n-1}$, i.e., solve one-dimensional transportation problem on each vertical line. This is possible by (97). Using (96), (97), one can show that t_X can be defined by

$$t_X(\tau) := \inf \{\zeta \in \mathbf{R}^1 \mid \Psi^+(X, \tau) \geq \Psi^-(X, \zeta)\} \tag{98}$$

$$= \sup \{\zeta \in \mathbf{R}^1 \mid \Psi^+(X, \tau) < \Psi^-(X, \zeta)\}. \tag{99}$$

Details and proofs for the above construction are the same as in the Euclidean case [3, §5].

7 Uniqueness of monotonic optimal map, and transport density

Measure-decomposing change of variables defined in Section 4 allows us to extend to the case of manifolds the results of [11]. We will only state the results and give some remarks about proofs, since the proofs follow closely those of [11], using Sections 4, 5 of the present manuscript.

We first address the question of uniqueness of optimal map. As in Euclidean space, the optimal map is nonunique. The source of the nonuniqueness is nonuniqueness in the one-dimensional transfer problem. We show that this is the only source of nonuniqueness. Precisely, we show uniqueness of optimal map satisfying a one-dimensional monotonicity condition:

Theorem 28 (Uniqueness of optimal maps) *Let densities $f^+, f^- \geq 0$ on M satisfy same conditions as in Theorem 1. Among Borel maps $s : M \longrightarrow M$ solving Monge's problem, in the sense that they minimize the average distance (3) transported among all maps pushing f^+ forward to f^- (4), there exists a unique optimal map $s \in \mathcal{A}(\mu^+, \mu^-)$ satisfying the following monotonicity condition:*

Let $x_1, x_2 \in M$ be such that four points x_1, $s(x_1)$, x_2, $s(x_2)$ lie on one minimizing geodesic $\gamma : [0, T] \to M$. For p on γ denote by $t(p)$ the unique $t \in [0, T]$ such that $\gamma(t) = p$. Then

$$[t(x_1) - t(x_2)][t(s(x_1)) - t(s(x_2))] \geq 0. \tag{100}$$

Proof. The difference between the monotonicity condition stated above from the condition in [11, Theorem 1.2] is that now we require some condition only for points x_1, x_2 such that x_1, $s(x_1)$, x_2, $s(x_2)$ lie on one minimizing geodesic.

Thus for the proof of Theorem 28 we follow the argument of [11, Section 3], skipping Lemma 3.1 of [11, Section 3] since this lemma considers the case when x_1, $s(x_1)$, x_2, $s(x_2)$ are not in one line. □

While the optimal map satisfying monotonicity condition is unique, there are other optimal maps, which are not monotonic along transport rays. Now we compare different optimal maps, and show that they generate same mass (or cost) flow.

Let u be a Kantorovich potential for Problem 1, satisfying (18). By Theorem 1 and Lemma 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 $\nabla u(z)$. It remains to study the rate of optimal mass transfer through a point of M. 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 segment of transport ray 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 $d(x, s(x))$. The *transport cost density* $a(z)$ defined in (101) 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. We quantify this definition by choosing a particular sequence of open neighborhoods $\mathcal{D}_R(z)$ shrinking to z and setting

$$a(z) = \lim_{R \to 0+} \frac{\text{cost of transportation through } \mathcal{D}_R(z) \text{ of mass flow generated by } s}{\mathrm{vol}[\mathcal{D}_R(z)]}, \tag{101}$$

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 Section §4. 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 $z \in T_1^0$, $R > 0$

$$\mathcal{D}_R(z) := \mathrm{j}[B_R(z) \cap S_{z,u}] \cap \{y \mid u(z) - R \le u(y) \le u(z) + R\}, \tag{102}$$

where the map j is defined by (92), and $B_R(z) = \{y \in M \mid d(y, z) < R\}$. By Corollary 25, $\mathcal{D}_R(z)$ is a Borel subset of M.

Let us compute 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 $u(x)-u(y) = d(x, y)$, 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 (92). Then calculation given in [11, Section 4] shows that this expression can be rewritten as

$$\mu^+[\mathrm{j}^+(A)] - \mu^-[\mathrm{j}^+(A)]. \tag{103}$$

Note that (103) depends only on the Kantorovich potential u, and no longer on the map s.

Now the cost of transport of the mass (103) 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 $B \in M$ is

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

Thus, recalling (101), we define the transport cost density of the flow generated by any optimal map s at the point $z \in T_1^0$ as

$$a(z) := \lim_{R\to 0+} \frac{\int_{-\infty}^{\infty} \Big\{\mu^+\Big[\mathrm{j}^+[\mathcal{D}_R(z) \cap u^{-1}(t)]\Big] - \mu^-\Big[\mathrm{j}^+[\mathcal{D}_R(z) \cap u^{-1}(t)]\Big]\Big\}dt}{\mathrm{vol}[\mathcal{D}_R(z)]}, \tag{104}$$

where $\mathcal{D}_R(z)$ is defined by (102).

Theorem 29 (Existence, uniqueness, and properties of transport density) *Fix a Kantorovich potential u satisfying (18).*

i. The limit (104) exists a.e. on T_1^0.

ii. There exists $a \in L^1(M)$, *called the* transport cost density, *with the following properties:*

$$a \geq 0 \text{ on } M, \quad a \equiv 0 \text{ on } M \setminus T_1, \tag{105}$$

and $a(z)$ *is equal to the right-hand side of (104) for* $\mathcal{L}^n$ *a.e.* $z \in T_1^0$. *In addition,* $a(\,\cdot\,)$ *satisfies the equation*

$$-div\,(a\,\nabla u) = f^+ - f^- \quad \text{in } M \tag{106}$$

in the weak sense, meaning for any test function $\varphi \in C^1(M)$

$$\int_M a\,\langle \nabla u, \nabla \varphi \rangle\, d\,vol(z) = \int_M (f^+ - f^-)\varphi\, d\,vol(z). \tag{107}$$

Moreover, for any measurable transport set $A \subset M$ *and* $\varphi \in C^1(M)$

$$\int_A a\,\langle \nabla u, \nabla \varphi \rangle\, d\,vol(z) = \int_A (f^+ - f^-)\varphi\, d\,vol(z). \tag{108}$$

iii. A function $a \in L^1(M)$ *satisfying (108) for all measurable transport sets* A *and* $\varphi \in C^1(M)$ *is uniquely determined by the constraints (105).*

The proof follows the argument of [11, Section 4].

References

[1] L. Ambrosio. Lecture notes on optimal transport problems. *Preprint.*

[2] 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.

[3] L. Caffarelli, M. Feldman, R.J. McCann. Constructing optimal maps for Monge's transport problem as a limit of strictly convex costs. Preprint.

[4] J. Cheeger, D.G. Ebin. Comparison Theorems in Riemannian Geometry. North-Holland Publishing Co., Amsterdam-Oxford, 1975.

[5] M.P. Do Carmo. Riemannian geometry. Birkhauser Boston, Inc., Boston, MA, 1992.

[6] 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.

[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*, 1992.

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

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

[11] M. Feldman, R.J. McCann. Uniqueness and transport density in Monge's mass transportation problem. Preprint.

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

[13] W. Klingenberg Riemannian geometry. de Gruyter Studies in Mathematics, 1. Walter de Gruyter & Co., Berlin-New York, 1982. 396 pp.

[14] R.J. McCann. Polar factorization of maps on Riemannian manifolds. To appear in *Geom. Funct. Anal.* Also preprint #41 (1999) at http://www.mis.mpg.de/.

[15] 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.

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

[17] M. Spivak A comprehensive introduction to differential geometry. Vol. I. Second edition. Publish or Perish, Inc., Wilmington, Del., 1979. xiv+668 pp

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

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

[20] C. Villani. Topics in mass transportation. *In preparation.*