PROCEEDINGS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 145, Number 7, July 2017, Pages 3127–3136
http://dx.doi.org/10.1090/proc/13415
Article electronically published on December 8, 2016

ON TANGENT CONES IN WASSERSTEIN SPACE

JOHN LOTT

(Communicated by Guofang Wei)

Abstract. If M is a smooth compact Riemannian manifold, let $P(M)$ denote the Wasserstein space of probability measures on M. If S is an embedded submanifold of M, and μ is an absolutely continuous measure on S, then we compute the tangent cone of $P(M)$ at μ.

1. Introduction

In optimal transport theory, a displacement interpolation is a one-parameter family of measures that represents the most efficient way of displacing mass between two given probability measures. Finding a displacement interpolation between two probability measures is the same as finding a minimizing geodesic in the space of probability measures, equipped with the Wasserstein metric W_2 [9, Proposition 2.10]. For background on optimal transport and Wasserstein space, we refer to Villani's book [14].

If M is a compact connected Riemannian manifold with nonnegative sectional curvature, then $P(M)$ is a compact length space with nonnegative curvature in the sense of Alexandrov [9, Theorem A.8], [13, Proposition 2.10]. Hence one can define the tangent cone $T_\mu P(M)$ of $P(M)$ at a measure $\mu \in P(M)$. If μ is absolutely continuous with respect to the volume form dvol_M, then $T_\mu P(M)$ is a Hilbert space [9, Proposition A.33]. More generally, one can define tangent cones of $P(M)$ without any curvature assumption on M, using Ohta's 2-uniform structure on $P(M)$ [11]. Gigli showed that $T_\mu P(M)$ is a Hilbert space if and only if μ is a "regular" measure, meaning that it gives zero measure to any hypersurface which, locally, is the graph of the difference of two convex functions [7, Corollary 6.6]. It is natural to ask what the tangent cones are at other measures.

A wide class of tractable measures comes from submanifolds. Suppose that S is a smooth embedded submanifold of a compact connected Riemannian manifold M. Suppose that μ is an absolutely continuous probability measure on S. We can also view μ as an element of $P(M)$. For simplicity, we assume that $\mathrm{supp}(\mu) = S$.

Theorem 1.1. *We have*

$$T_\mu P(M) = H \oplus \int_{s \in S} P_2(N_s M)\, d\mu(s), \tag{1.1}$$

where

- *H is the Hilbert space of gradient vector fields* $\overline{\mathrm{Im}(\nabla)} \subset L^2(TS, d\mu)$,

Received by the editors March 9, 2016 and, in revised form, August 6, 2016.
2010 *Mathematics Subject Classification.* Primary 51K99.
This research was partially supported by NSF grant DMS-1207654 and a Simons Fellowship.

- *N_sM is the normal space to $S \subset M$ at $s \in S$ and*
- *$P_2(N_sM)$ is the metric cone of probability measures on N_sM with finite second moment, equipped with the 2-Wasserstein metric.*

The homotheties in the metric cone structure on $P_2(N_sM)$ arise from radial rescalings of N_sM. The direct sum and integral in (1.1) refer to computing square distances.

The proof of Theorem 1.1 amounts to understanding optimal transport starting from a measure supported on a submanifold. This seems to be a natural question in its own right which has not been considered much. Gangbo and McCann proved results about optimal transport between measures supported on hypersurfaces in Euclidean space [6]. McCann-Sosio and Kitagawa-Warren gave more refined results about optimal transport between two measures supported on a sphere [8,10]. Castillon considered optimal transport between a measure supported on a submanifold of Euclidean space and a measure supported on a linear subspace [5].

In the setting of Theorem 1.1, a Wasserstein geodesic $\{\mu_t\}_{t\in[0,\epsilon]}$ starting from μ consists of a family of geodesics shooting off from S in various directions. The geometric meaning of Theorem 1.1 is that the tangential component of these directions is the gradient of a function on S. To motivate this statement, in Section 2 we give a Benamou-Brenier-type variational approach to the problem of optimally tranporting a measure supported on one hypersurface to a measure supported on a disjoint hypersurface, through a family of measures supported on hypersurfaces. One finds that the only constraint is the aforementioned tangentiality constraint. The rigorous proof of Theorem 1.1 is in Section 3.

The structure of this paper is as follows. In Section 2 we give a formal derivation of the equation for optimal transport between two measures supported on disjoint hypersurfaces of a Riemannian manifold. The derivation is based on a variational method. In Section 3 we prove Theorem 1.1.

2. Variational approach

Let M be a smooth closed Riemannian manifold. Let S be a smooth closed manifold and let S_0, S_1 be disjoint codimension-one submanifolds of M diffeomorphic to S. Let $\rho_0 \, \mathrm{dvol}_{S_0}$ and $\rho_1 \, \mathrm{dvol}_{S_1}$ be smooth probability measures on S_0 and S_1, respectively. We consider the problem of optimally transporting $\rho_0 \, \mathrm{dvol}_{S_0}$ to $\rho_1 \, \mathrm{dvol}_{S_1}$ through a family of measures supported on codimension-one submanifolds $\{S_t\}_{t\in[0,1]}$. We will specify the intermediate submanifolds to be level sets of a function T, which in turn will become one of the variables in the optimization problem.

We assume that there is a codimension-zero submanifold-with-boundary U of M, with $\partial U = S_0 \cup S_1$. We also assume that there is a smooth submersion $T : U \to [0,1]$ so that $T^{-1}(0) = S_0$ and $T^{-1}(1) = S_1$. For $t \in [0,1]$, put $S_t = T^{-1}(t)$. These are the intermediate hypersurfaces.

We now want to describe a family of measures $\{\mu_t\}_{t\in[0,1]}$ that live on the hypersurfaces $\{S_t\}_{t\in[0,1]}$. It is convenient to think of these measures as fitting together to form a measure on U. Let μ be a smooth measure on U. In terms of the fibering $T : U \to [0,1]$, decompose μ as $\mu = \mu_t dt$ with μ_t a measure on S_t. We assume that $\mu_0 = \rho_0 \, \mathrm{dvol}_{S_0}$ and $\mu_1 = \rho_1 \, \mathrm{dvol}_{S_1}$.

Let V be a vector field on U. We want the flow $\{\phi_s\}$ of V to send level sets of T to level sets. Imagining that there is an external clock, it's convenient to think of

S_t as the evolving hypersurface at time t. Correlating the flow of V with the clock gives the constraint

$$VT = 1. \tag{2.1}$$

Then ϕ_s maps S_t to S_{t+s}.

We also want the flow to be compatible with the measures $\{\mu_t\}_{t\in[0,1]}$ in the sense that $\phi_s^*\mu_{t+s} = \mu_t$. Now $\phi_s^* dT = d\phi_s^* T = d(T+s) = dT$, so it is equivalent to require that ϕ_s^* preserves the measure $\mu = \mu_t dt$. This gives the constraint

$$\mathcal{L}_V \mu = 0. \tag{2.2}$$

In particular, each μ_t is a probability measure.

To define a functional along the lines of Benamou and Brenier [2], put

$$E = \frac{1}{2}\int_U |V|^2\, d\mu = \frac{1}{2}\int_0^1 \int_{S_t} |V|^2\, d\mu_t\, dt. \tag{2.3}$$

We want to minimize E under the constraints $\mathcal{L}_V\mu = 0$, $VT = 1$, $\mu_0 = \rho_0\, \mathrm{dvol}_{S_0}$ and $\mu_1 = \rho_1\, \mathrm{dvol}_{S_1}$. Let ϕ and η be new functions on U, which will be Lagrange multipliers for the constraints. Then we want to extremize

$$\mathcal{E} = \int_U \left[\frac{1}{2}|V|^2\, d\mu + \phi\mathcal{L}_V d\mu + \eta(VT-1)d\mu\right] \tag{2.4}$$

with respect to V, μ, ϕ and η.

We will use the equations

$$\begin{aligned} \int_U \phi\mathcal{L}_V d\mu &= \int_U [\mathcal{L}_V(\phi d\mu) - (\mathcal{L}_V\phi)d\mu] \\ &= -\int_U (V\phi)d\mu + \int_{S_1}\phi(1)d\mu_1 - \int_{S_0}\phi(0)d\mu_0 \end{aligned} \tag{2.5}$$

and

$$\begin{aligned} \int_U \eta VT d\mu &= \int_U [\mathcal{L}_V(T\eta d\mu) - T\mathcal{L}_V(\eta d\mu)] \\ &= -\int_U T\mathcal{L}_V(\eta d\mu) + \int_{S_1}\eta(1)d\mu_1. \end{aligned} \tag{2.6}$$

The Euler-Lagrange equation for V is

$$V - \nabla\phi + \eta\nabla T = 0. \tag{2.7}$$

The Euler-Lagrange equation for μ is

$$\frac{1}{2}|V|^2 - V\phi = 0. \tag{2.8}$$

Varying T gives

$$0 = \mathcal{L}_V(\eta d\mu) = (V\eta)d\mu, \tag{2.9}$$

so the Euler-Lagrange equation for T is

$$V\eta = 0. \tag{2.10}$$

Substituting (2.7) into (2.8) gives $|\nabla\phi|^2 = \eta^2|\nabla T|^2$, so $\eta = \pm\frac{|\nabla\phi|}{|\nabla T|}$. Then (2.7) becomes

$$V = \nabla\phi \mp \frac{|\nabla\phi|}{|\nabla T|}\nabla T. \tag{2.11}$$

Equation (2.1) gives

$$1 = \langle \nabla \phi, \nabla T \rangle \mp |\nabla \phi| \cdot |\nabla T|. \tag{2.12}$$

If the "$\mp$" is "$-$", then the right-hand side of (2.12) is nonpositive, which is a contradiction. Thus

$$1 = \langle \nabla \phi, \nabla T \rangle + |\nabla \phi| \cdot |\nabla T| \tag{2.13}$$

and

$$V = \nabla \phi + \frac{|\nabla \phi|}{|\nabla T|} \nabla T. \tag{2.14}$$

Equation (2.10) becomes

$$V \frac{|\nabla \phi|}{|\nabla T|} = 0, \tag{2.15}$$

which is equivalent to

$$\frac{1}{2} V |V|^2 = 0. \tag{2.16}$$

Equation (2.16) says that V has constant length along its flowlines. The measure μ must still satisfy the conservation law (2.2).

From (2.8), the evolution of ϕ between level sets is given by

$$V\phi = \frac{1}{2}|V|^2 = \frac{1}{2} \frac{|\nabla \phi|}{|\nabla T|}. \tag{2.17}$$

The normal line to a level set S_t is spanned by ∇T. It follows from (2.7) that the tangential part of V is the gradient of a function on S_t:

$$V_{tan} = \nabla_{S_t} \left(\phi \Big|_{S_t} \right). \tag{2.18}$$

The normal part of V is

$$V_{norm} = \frac{\langle V, \nabla T \rangle}{|\nabla T|^2} \nabla T = \frac{1}{|\nabla T|^2} \nabla T, \tag{2.19}$$

as must be the case from (2.1).

The conclusion is that the tangential part of V on S_t is a gradient vector field on S_t, while the normal part of V on S_t is unconstrained.

3. Tangent cones

3.1. **Optimal transport from submanifolds.** Let M be a smooth closed Riemannian manifold. Let $i : S \to M$ be an embedding.

Let $\pi : TM \to M$ be the projection map. Given $\epsilon > 0$, define $E_\epsilon : TM \to TM$ by $E_\epsilon(m, v) = (\exp_m(\epsilon v), d(\exp_m)_{\epsilon v} \epsilon v)$. We define π^S and E^S_ϵ similarly, replacing M by S.

Put $T_S M = i^* TM$, a vector bundle on S with projection map $\pi_{T_S M} : T_S M \to S$. There is an orthogonal splitting $T_S M = TS \oplus N_S M$ into the tangential part and the normal part. Let $\pi_{N_S M} : N_S M \to S$ be the projection to the base of $N_S M$. Given $v \in TS$, let $v^T \in TS$ denote its tangential part and let $v^\perp \in NS$ denote its normal part. Let $p^T : T_S M \to TS$ be the orthogonal projection.

A function $F : S \to \mathbb{R} \cup \{\infty\}$ is *semiconvex* if there is some $\lambda \in \mathbb{R}$ so that for all minimizing constant-speed geodesics $\gamma : [0,1] \to S$, we have

$$F(\gamma(t)) \le tF(\gamma(1)) + (1-t)F(\gamma(0)) - \frac{1}{2}\lambda t(1-t) d_S(\gamma(0), \gamma(1))^2 \tag{3.1}$$

for all $t \in [0,1]$.

Suppose that F is a semiconvex function on S. Then $(s, w) \in TS$ lies in the *subdifferential set* $\nabla^- F$ if for all $w' \in T_s S$,

$$F(s) + \langle w, w' \rangle \le F(\exp_s w') + o(|w'|). \tag{3.2}$$

Define the cost function $c : S \times M \to \mathbb{R}$ by $c(s,x) = \frac{1}{2} d(s,x)^2$. Given $\eta : M \to \mathbb{R} \cup \{-\infty\}$, its c-transform is the function $\eta^c : S \to \mathbb{R} \cup \{\infty\}$ given by

$$\eta^c(s) = \sup_{x \in M} \left(\eta(x) - \frac{1}{2} d^2(s,x) \right). \tag{3.3}$$

Given $\psi : S \to \mathbb{R} \cup \{\infty\}$, its c-transform is the function $\psi^c : M \to \mathbb{R} \cup \{-\infty\}$ given by

$$\psi^c(x) = \inf_{s \in S} \left(\psi(s) + \frac{1}{2} d^2(s,x) \right). \tag{3.4}$$

A function $\psi : S \to \mathbb{R} \cup \{\infty\}$ is *c-convex* if $\psi = \eta^c$ for some $\eta : M \to \mathbb{R} \cup \{-\infty\}$. A function $\eta : M \to \mathbb{R} \cup \{-\infty\}$ is *c-concave* if $\eta = \psi^c$ for some $\psi : S \to \mathbb{R} \cup \{\infty\}$.

From [14, Proposition 5.8], a function $F : S \to \mathbb{R} \cup \{-\infty\}$ is c-convex if and only if $F = (F^c)^c$, i.e., for all $s \in S$,

$$F(s) = \sup_{x \in M} \inf_{s' \in S} \left(F(s') + \frac{1}{2} d^2(s',x) - \frac{1}{2} d^2(s,x) \right). \tag{3.5}$$

The next lemma appears in [7, Lemma 2.9] when $S = M$.

Lemma 3.1. *If $F : S \to \mathbb{R} \cup \{\infty\}$ is a semiconvex function, then there is some $\epsilon > 0$ so that ϵF is c-convex.*

Proof. Clearly

$$\epsilon F(s) \ge \sup_{x \in M} \inf_{s' \in S} \left(\epsilon F(s') + \frac{1}{2} d^2(s',x) - \frac{1}{2} d^2(s,x) \right), \tag{3.6}$$

as is seen by taking $s' = s$ on the right-hand side of (3.6). Hence we must show that for suitable $\epsilon > 0$, for all $s \in S$ we have

$$\epsilon F(s) \le \sup_{x \in M} \inf_{s' \in S} \left(\epsilon F(s') + \frac{1}{2} d^2(s',x) - \frac{1}{2} d^2(s,x) \right). \tag{3.7}$$

For this, it suffices to show that for each $s \in S$, there is some $x \in M$ so that

$$\epsilon F(s) \le \inf_{s' \in S} \left(\epsilon F(s') + \frac{1}{2} d^2(s',x) - \frac{1}{2} d^2(s,x) \right). \tag{3.8}$$

That is, it suffices to show that for each $s \in S$, there is some $x \in M$ so that for all $s' \in S$, we have

$$\epsilon F(s) \le \epsilon F(s') + \frac{1}{2} d^2(s',x) - \frac{1}{2} d^2(s,x), \tag{3.9}$$

i.e.,

$$\epsilon F(s) + \frac{1}{2} d^2(s,x) \le \epsilon F(s') + \frac{1}{2} d^2(s',x). \tag{3.10}$$

We know that F is K-Lipschitz for some $K < \infty$ [14, Theorem 10.8 and Proposition 10.12]. Hence if $v \in \nabla^-_s F$, then $|v| \leq K$. Given s, choose $v \in \nabla^-_s F$ and put $x = \exp_s(\epsilon v) \in M$. Then $d(s,x) \leq \epsilon K$.

Put $G(s') = \epsilon F(s') + \frac{1}{2} d^2(s',x)$. We want to show that $G(s) \leq G(s')$ for all $s' \in S$. Suppose not. Let s' be a minimum point for G; then $G(s') < G(s)$.

We claim first that $s' \in B_{4\epsilon K}(s)$. To see this, if $d(s,s') \geq 4\epsilon K$, then since

$$d(s',x) \geq d(s,s') - d(s,x) \geq d(s,s') - \epsilon K, \tag{3.11}$$

we have

$$\begin{aligned} \frac{1}{2} d^2(s',x) - \frac{1}{2} d^2(s,x) &\geq \frac{1}{2} \left(d(s,s') - \epsilon K \right)^2 - \frac{1}{2} \left(\epsilon K \right)^2 \\ &= \frac{1}{2} (d(s,s') - 2\epsilon K) \cdot d(s,s') \\ &\geq \epsilon K d(s,s') \geq \epsilon (F(s) - F(s')), \end{aligned} \tag{3.12}$$

which contradicts that $G(s') < G(s)$. This proves the claim.

If $10\epsilon K$ is less than the injectivity radius of M, then there is a unique minimizing geodesic from s to x, and its tangent vector at s is ϵv. It follows that $0 \in \nabla^-_s G$. Finally, since $d(s,x) \leq \epsilon K$, we can choose an ϵ (depending on K, S and M) to ensure that G is strictly convex on $B_{4\epsilon K}(s)$, with the latter being a totally convex set. Considering the function G along a minimizing geodesic from s to s', we obtain a contradiction to the assumed strict convexity of G, along with the facts that $0 \in \nabla^-_s G$ and $0 \in \nabla^-_{s'} G$.

Thus G is minimized at s, which implies (3.10). □

Let ν be a compactly-supported probability measure on $T_S M \subset TM$. Let $L < \infty$ be such that the support of ν is contained in $\{v \in T_S M \ : \ |v| \leq L\}$. Put $\mu_\epsilon = \pi_* (E_\epsilon)_* \nu$.

Proposition 3.13. *a. Let f be a semiconvex function on S. Suppose that ν is supported on $\{v \in T_S M \ : \ v^T \in \nabla^- f\}$. Then there is some $\epsilon > 0$ so that the one-parameter family of measures $\{\mu_t\}_{t \in [0,\epsilon]}$ is a Wasserstein geodesic.*
b. Given ν, suppose that for some $\epsilon > 0$, the one-parameter family of measures $\{\mu_t\}_{t \in [0,\epsilon]}$ is a Wasserstein geodesic. Then there is a semiconvex function f on S so that ν is supported on $\{v \in T_S M \ : \ v^T \in \nabla^- f\}$.

Proof. a. For $t > 0$, define $\eta_t : M \to \mathbb{R}$ by $\eta_t = (tf)^c$. From Lemma 3.1, if t is small enough, then tf is c-convex. It follows from [14, Proposition 5.8] that $(\eta_t)^c = tf$.

From [14, Theorem 5.10], if a set $\Gamma_t \subset S \times M$ is such that $\eta_t(x) = tf(s) + \frac{1}{2} d^2(s,x)$ for all $(s,x) \in S \times M$, then any probability measure Π_t with support in Γ_t is an optimal transport plan. We take

$$\Gamma_t = \{(s,x) \in S \times M \ : \ \eta_t(x) = tf(s) + \frac{1}{2} d^2(s,x)\}. \tag{3.14}$$

Now $\eta_t(x) = tf(s) + \frac{1}{2} d^2(s,x)$ if for all $s' \in S$, we have

$$tf(s) + \frac{1}{2} d^2(s,x) \leq tf(s') + \frac{1}{2} d^2(s',x). \tag{3.15}$$

To prove part a. of the proposition, it suffices to show that for all sufficiently small t, equation (3.15) is satisfied for $s, s' \in S$ and $x = \exp_s(tv)$, where $v \in T_s M$ lies in the support of ν and satisfies $v^T \in \nabla^- f$.

Given s and v, we know that $d(s,x) \le tL$. Put $G(s') = tf(s') + \frac{1}{2}d^2(s',x)$. Let s' be a minimum point of G and suppose, to get a contradiction, that $G(s') < G(s)$.

Let $K < \infty$ be the Lipschitz constant of f. We claim first that $s' \in B_{t(2K+2L)}(s)$. To see this, if $d(s,s') \ge t(2K+2L)$, then

$$d(s',x) \ge d(s,s') - d(s,x) \ge d(s,s') - tL \tag{3.16}$$

and

$$\begin{aligned} \frac{1}{2}d^2(s',x) - \frac{1}{2}d^2(s,x) &\ge \frac{1}{2}\left(d(s,s') - tL\right)^2 - (tL)^2 \\ &= \frac{1}{2}(d(s,s') - 2tL) \cdot d(s,s') \\ &\ge tKd(s,s') \ge t(f(s) - f(s')), \end{aligned} \tag{3.17}$$

which is a contradiction and proves the claim.

There is some $\epsilon > 0$ (depending on L, S and M) so that if $t \in [0,\epsilon]$, then we are ensured that there is a unique minimizing geodesic from s to x, and its tangent vector at s is tv. It follows that $0 \in \nabla_s^- G$. Finally, since $d(s,x) \le \epsilon L$, we can choose ϵ (depending on K, L, S and M) to ensure that G is strictly convex on $B_{t(2K+2L)}(s)$, the latter being totally convex. Considering the function G along a minimizing geodesic from s to s', we obtain a contradiction to the assumed strict convexity of G, along with the facts that $0 \in \nabla_s^- G$ and $0 \in \nabla_{s'}^- G$. This proves part a. of the proposition.

Now suppose that $\{\mu_t\}_{t\in[0,\epsilon]}$ is a Wasserstein geodesic. From [14, Theorem 5.10], there is a c-convex function ϵf on S so that if we define its conjugate $(\epsilon f)^c$ using (3.4), then $\{(s, \exp_s(\epsilon v)\}_{(s,v)\in \mathrm{supp}(\nu)}$ is contained in

$$\Gamma_\epsilon = \left\{ (s,x) \in S \times M \;:\; (\epsilon f)^c(x) = \epsilon f(s) + \frac{1}{2}d^2(s,x) \right\}. \tag{3.18}$$

That is, for all $s' \in S$,

$$\epsilon f(s) + \frac{1}{2}d^2(s, \exp_s(\epsilon v)) \le \epsilon f(s') + \frac{1}{2}d^2(s', \exp_s(\epsilon v)). \tag{3.19}$$

Without loss of generality, we can shrink ϵ as desired. Define a curve in S by $s'(u) = \exp_s(-uw')$ where $w' \in T_sS$, u varies over a small interval $(-\delta,\delta)$ and $\exp_s$ denotes here the exponential map for the submanifold S. Let $\{\gamma_u : [0,\epsilon] \to M\}_{u\in(-\delta,\delta)}$ be a smooth one-parameter family with $\gamma_0(t) = \exp_s(tv)$, $\gamma_u(0) = s'(u)$ and $\gamma_u(\epsilon) = \exp_s(\epsilon v)$. Let $L(u)$ be the length of γ_u. Then

$$\epsilon f(s) + \frac{1}{2}d^2(s, \exp_s(\epsilon v)) \le \epsilon f(s'(u)) + \frac{1}{2}L^2(u). \tag{3.20}$$

By the first variation formula,

$$\frac{d}{du}\Big|_{u=0} \frac{1}{2}L^2(u) = \epsilon\langle v^T, w'\rangle. \tag{3.21}$$

It follows that $\epsilon v^T \in \nabla_s^-(\epsilon f)$, so $v^T \in \nabla_s^- f$. $\square$

Remark 3.2. The phenomenon of possible nonuniqueness, in the normal component of the optimal transport between two measures supported on convex hypersurfaces in Euclidean space, was recognized in [6, Proposition 4.3].

Example 3.3. Put $M = S^1 \times \mathbb{R}$. (It is noncompact, but this will be irrelevant for the example.) Let $F \in C^\infty(S^1)$ be a positive function. Put $S = \{(x, F(x)) : x \in S^1\}$. Define $p : S \to S^1 \times \{0\}$ by $p(x, F(x)) = (x, 0)$. Let μ_0 be a smooth measure on S. Put $\mu_1 = p_* \mu_0$. The Wasserstein geodesic from μ_0 to μ_1 moves the measure down along vertical lines. Defining f on S by $f(x, F(x)) = -\frac{1}{2}(F(x))^2$, one finds that $v^T = \nabla f$. Compare with [5, Corollary 2.6].

3.2. **Tangent cones.** If X is a complete length space with Alexandrov curvature bounded below, then one can define the tangent cone $T_x X$ at $x \in X$ as follows. Let Σ'_x be the space of equivalence classes of minimal geodesic segments emanating from x, with the equivalence relation identifying two segments if they form a zero angle at x (which means that one segment is contained in the other). The metric on Σ'_x is the angle. By definition, the space of directions Σ_x is the metric completion of Σ'_x. The tangent cone $T_x X$ is the union of $\mathbb{R}^+ \times \Sigma_x$ and a "vertex" point, with the metric described in [4, §10.9].

If X is finite-dimensional, then one can also describe $T_x X$ as the pointed Gromov-Hausdorff limit $\lim_{\lambda \to \infty} (\lambda X, x)$. This latter description doesn't make sense if X is infinite-dimensional, whereas the preceding definition does.

If M is a smooth compact connected Riemannian manifold, and it has nonnegative sectional curvature, then $P(M)$ has nonnegative Alexandrov curvature and one can talk about a tangent cone $T_\mu P(M)$ [9, Appendix A]. If M does not have nonnegative sectional curvature, then $P(M)$ will not have Alexandrov curvature bounded below. Nevertheless, one can still define $T_\mu P(M)$ in the same way [11, Section 3].

As a point of terminology, what is called a tangent cone here, and in [9], is called the "abstract tangent space" in [7]. The linear part of the tangent cone is called the "tangent space" in [1] and the "space of gradients" or "tangent vector fields" in [7].

A minimal geodesic segment emanating from $\mu \in P(M)$ is determined by a probability measure Π on the space of constant-speed minimizing geodesics

$$\Gamma = \{\gamma : [0,1] \to M \ : \ L(\gamma) = d_M(\gamma(0), \gamma(1))\}, \tag{3.22}$$

which has the property that under the time-zero evaluation $e_0 : \Gamma \to M$, we have $(e_0)_* \Gamma = \mu$ [9, Section 2]. The corresponding geodesic segment is given by $\mu_t = (e_t)_* \Pi$, where $e_t : \Gamma \to M$ is time-t evaluation.

Using this characterization of minimizing geodesic segments, one can describe $T_\mu P(M)$ as follows. With $\pi : TM \to M$ being projection to the base, put

$$P_2(TM)_\mu = \{\nu \in P_2(TM) \ : \ \pi_* \nu = \mu\}, \tag{3.23}$$

where P_2 refers to measures with finite second moment. Given $\nu^1, \nu^2 \in P_2(TM)_\mu$, decompose them as

$$\nu^i = \int_M \nu^i_m \, d\mu(m), \tag{3.24}$$

with $\nu^i_m \in P_2(T_m M)$. Define $W_\mu(\nu^1, \nu^2)$ by

$$W^2_\mu(\nu^1, \nu^2) = \int_M W^2_2(\nu^1_m, \nu^2_m) \, d\mu(m). \tag{3.25}$$

Let Dir_μ be the set of elements $\nu \in P_2(TM)_\mu$ with the property that $\{\pi_*(E_t)_* \nu\}_{t \in [0, \epsilon]}$ describes a minimizing Wasserstein geodesic for some ϵ. Then $T_\mu P(M)$ is isometric to the metric completion of Dir_μ with respect to W_μ [7, Theorem 5.5].

We note that since M is compact, any element of Dir_μ has compact support. This is because for ν-almost all $v \in TM$, the geodesic $\{\exp_{\pi(v)} tv\}_{t\in[0,\epsilon]}$ must be minimizing [9, Proposition 2.10], so $|v| \le \epsilon^{-1} \operatorname{diam}(M)$.

Proof of Theorem 1.1. From Proposition 3.13, Dir_μ is the set of compactly-supported measures $\nu \in P(T_S M) \subset P(TM)$ so that $\pi_* \nu = \mu$ and there is a semiconvex function f on S such that ν has support on $\{v \in T_S M \,:\, v^T \in \nabla^- f\}$. Because μ has full support on S by assumption, $\nabla^- f$ is single-valued at μ-almost all $s \in S$. Equivalently, there is a compactly-supported $\nu^N \in P(N_S M)$, which decomposes under $\pi_{N_S M} \,:\, N_S M \to S$ as $\nu^N = \int_S \nu_s^N \, d\mu(s)$ with $\nu_s^N \in P_2(N_s M)$, so that for all $F \in C(T_S M) = C(TS \oplus N_S M)$, we have

$$\int_{T_S M} F \, d\nu = \int_S \int_{N_s M} F(\nabla^- f(s), w) \, d\nu_s^N(w) \, d\mu(s). \tag{3.26}$$

Given two such measures ν^1, ν^2, it follows that

$$W_\mu^2(\nu^1, \nu^2) = \int_S \langle \nabla^- f^1, \nabla^- f^2 \rangle \, d\mu + \int_S W_2^2(\nu_s^{1,N}, \nu_s^{2,N}) \, d\mu(s). \tag{3.27}$$

Upon taking the metric completion of Dir_μ, the tangential term in (3.27) gives the closure of the space of gradient vector fields in the Hilbert space $L^2(TS, d\mu)$ of square-integrable sections of TS [9, Proposition A.33]. The normal term gives $\int_{s\in S} P_2(N_s M) \, d\mu(s)$, where the metric comes from the last term in (3.27). This proves the theorem. □

Remark 3.4. In Section 2 we considered transports in which the intermediate measures were supported on hypersurfaces. This corresponds to Wasserstein geodesics starting from μ for which the initial velocity, as an element of $T_\mu P(M)$, comes from a section of $T_S M$. In terms of Theorem 1.1, this means that the data for the initial velocity consisted of a gradient vector field $\nabla\phi$ on S and a section $\mathcal{N}$ of $N_S M$, with the element of $P_2(N_s M)$ being the delta measure at $\mathcal{N}(s)$.

3.3. **Gauss map as an optimal transport map.** In this subsection, which is an addendum to the preceding subsections, we give an example of optimal transport coming from the Gauss map of a convex hypersurface in $\mathbb{R}^n$.

Let S be the boundary of a compact convex subset of $\mathbb{R}^n$. We assume that near any point, S is locally the graph of a C^2-regular function. Let $N : S \to S^{n-1}$ be the outward unit normal. Let $\kappa \in C^0(S)$ be the Gaussian curvature function, the product of the principal values. Then $N_*(\kappa \,\mathrm{dvol}_S) = \mathrm{dvol}_{S^{n-1}}$.

The optimal transport plans in $\mathbb{R}^n$ for the cost function $\frac{1}{2}|m_1 - m_2|^2$ are the same as those for the cost function $-\langle m_1, m_2\rangle$. Given $R > 0$, $s \in S$ and $x \in S^{n-1}$, the cost function of the points s and Rx becomes $-R\langle s, x\rangle$. Considering an optimal transport problem between S and $R\cdot S^{n-1}$, the optimal transport plans for the cost function $-R\langle s, x\rangle$ are the same as those for the cost function $-\langle s, x\rangle$. This motivates considering the cost function $c : S \times S^{n-1} \to \mathbb{R}$ given by $c(s, x) = -\langle s, x\rangle$. Here we imagine taking $R \to \infty$ so that S^{n-1} is a "sphere at infinity", not an embedded sphere in $\mathbb{R}^n$, although when we write $\langle s, x\rangle$ we are treating x as a unit vector.

The analog of (3.14) is

$$\Gamma_t = \{(s, x) \in S \times S^{n-1} \,:\, \eta_t(x) = tf(s) - \langle s, x\rangle\}. \tag{3.28}$$

Now $\eta_t(x) = tf(s) - \langle s, x\rangle$ if for all $s' \in S$, we have

$$tf(s) - \langle s, x\rangle \leq tf(s') - \langle s', x\rangle. \tag{3.29}$$

Taking $f = 0$, one sees that for all $s \in S$ we have $(s, N(s)) \in \Gamma_1$, since the convexity of S implies that $\langle s' - s, N(s)\rangle \leq 0$ for all $s' \in S$. Hence N is an optimal transport map from the measure $\kappa \, \mathrm{dvol}_S$ on S, to the measure $\mathrm{dvol}_{S^{n-1}}$ on S^{n-1}.

Remark 3.5. In a different direction, Aleksandrov's problem of realizing a given curvature function was related to optimal transport on a sphere in [12], using a certain cost function; see also [3].

Acknowledgements

The author thanks Cédric Villani for helpful comments, and Robert McCann for references to the literature. The author also thanks the referee for helpful remarks.

References

[1] Luigi Ambrosio and Nicola Gigli, *Construction of the parallel transport in the Wasserstein space*, Methods Appl. Anal. **15** (2008), no. 1, 1–29, DOI 10.4310/MAA.2008.v15.n1.a3. MR2482206

[2] Jean-David Benamou and Yann Brenier, *A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem*, Numer. Math. **84** (2000), no. 3, 375–393, DOI 10.1007/s002110050002. MR1738163

[3] Jérôme Bertrand, *Prescription of Gauss curvature using optimal mass transport*, Geom. Dedicata **183** (2016), 81–99, DOI 10.1007/s10711-016-0147-3. MR3523119

[4] D. Burago, Y. Burago, and S. Ivanov, *A course on metric geometry*, Graduate Studies in Mathematics 33, Amer. Math. Soc., Providence (2001)

[5] Philippe Castillon, *Submanifolds, isoperimetric inequalities and optimal transportation*, J. Funct. Anal. **259** (2010), no. 1, 79–103, DOI 10.1016/j.jfa.2010.03.001. MR2610380

[6] Wilfrid Gangbo and Robert J. McCann, *Shape recognition via Wasserstein distance*, Quart. Appl. Math. **58** (2000), no. 4, 705–737. MR1788425

[7] Nicola Gigli, *On the inverse implication of Brenier-McCann theorems and the structure of* $(\mathcal{P}_2(M), W_2)$, Methods Appl. Anal. **18** (2011), no. 2, 127–158, DOI 10.4310/MAA.2011.v18.n2.a1. MR2847481

[8] Jun Kitagawa and Micah Warren, *Regularity for the optimal transportation problem with Euclidean distance squared cost on the embedded sphere*, SIAM J. Math. Anal. **44** (2012), no. 4, 2871–2887, DOI 10.1137/120865409. MR3023397

[9] John Lott and Cédric Villani, *Ricci curvature for metric-measure spaces via optimal transport*, Ann. of Math. (2) **169** (2009), no. 3, 903–991, DOI 10.4007/annals.2009.169.903. MR2480619

[10] R. J. McCann and M. Sosio, *Hölder continuity for optimal multivalued mappings*, SIAM J. Math. Anal. **43** (2011), no. 4, 1855–1871, DOI 10.1137/100802670. MR2831251

[11] Shin-ichi Ohta, *Gradient flows on Wasserstein spaces over compact Alexandrov spaces*, Amer. J. Math. **131** (2009), no. 2, 475–516, DOI 10.1353/ajm.0.0048. MR2503990

[12] Vladimir Oliker, *Embedding* $\mathbf{S}^n$ *into* $\mathbf{R}^{n+1}$ *with given integral Gauss curvature and optimal mass transport on* $\mathbf{S}^n$, Adv. Math. **213** (2007), no. 2, 600–620, DOI 10.1016/j.aim.2007.01.005. MR2332603

[13] Karl-Theodor Sturm, *On the geometry of metric measure spaces. I*, Acta Math. **196** (2006), no. 1, 65–131, DOI 10.1007/s11511-006-0002-8. MR2237206

[14] Cédric Villani, *Optimal transport*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 338, Springer-Verlag, Berlin, 2009. Old and new. MR2459454

Department of Mathematics, University of California - Berkeley, Berkeley, California 94720-3840

E-mail address: lott@berkeley.edu