Multi-marginal optimal transport on Riemannian manifolds

Young-Heon Kim† and Brendan Pass‡

February 7, 2014

Abstract

We study a multi-marginal optimal transportation problem on a Riemannian manifold, with cost function given by the average distance squared from multiple points to their barycenter. Under a standard regularity condition on the first marginal, we prove that the optimal measure is unique and concentrated on the graph of a function over the first variable, thus inducing a Monge solution. This result generalizes McCann’s polar factorization theorem on manifolds from 2 to several marginals, in the same sense that a well known result of Gangbo and Swiech generalizes Brenier’s polar factorization theorem on $\mathbb{R}^n$.

1 Introduction

In this paper, we study an optimal transport problem with several marginals. Given Borel probability measures $\mu_1, \mu_2, \ldots, \mu_m$, each compactly supported on some smooth manifold $M$, and a cost function $c : M^m \to \mathbb{R}$, the multi-marginal optimal transport problem of Monge is to minimize

$$\int_M c(x_1, F_2(x_1), \ldots, F_m(x_1))d\mu_1(x_1) \quad (M)$$

among $(m-1)$-tuples of mappings $(F_2, \ldots, F_m)$, such that for each $i$, the map $F_i : M \mapsto M$ pushes the measure $\mu_1$ forward to $\mu_i$; that is, given any Borel $A \subseteq M$, $\mu_1(F_i^{-1}(A)) = \mu_i(A)$.

---

*Y.-H.K. is supported in part by Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grants 371642-09 as well as Alfred P. Sloan research fellowship 2012-2014. B.P. is pleased to acknowledge the support of a University of Alberta start-up grant and National Sciences and Engineering Research Council of Canada Discovery Grant number 412779-2012.

†Department of Mathematics, University of British Columbia, Vancouver BC Canada V6T 1Z2 yhkim@math.ubc.ca

‡Department of Mathematical and Statistical Sciences, 632 CAB, University of Alberta, Edmonton, Alberta, Canada, T6G 2G1 pass@ualberta.ca
The Kantorovich formulation of the multi-marginal optimal transport problem is the minimize
\[ \int_{\mathcal{M}^m} c(x_1, x_2, \ldots, x_m) d\gamma(x_1, x_2, \ldots, x_m) \] (K)
among all probability measures \( \gamma \) on the product space \( \mathcal{M}^m \) which project to the \( \mu_i \); that is, such that \( \gamma(M_i^{-1} \times A \times \mathcal{M}^{m-i}) = \mu_i(A) \) for all Borel \( A \subseteq \mathcal{M} \) and all \( i = 1, 2, \ldots, m \). Under reasonable conditions (e.g. continuity of the cost function and compactness of the supports of the measures), it is fairly straightforward to assert the existence of a minimizer \( \gamma \) in (K). Note that, for any \((F_2, \ldots, F_m)\) satisfying the pushforward constraint in (M), the pushforward measure \( \gamma \) of \( \mu_1 \) by the map \( (\text{Id}, F_2, \ldots, F_M) : \mathcal{M} \to \mathcal{M}^m \) satisfies the marginal constraint in (K) and so (K) is a relaxed version of (M).

When \( m = 2 \), (K) and (M) correspond to the Monge and Kantorovich formulations, respectively, of the classical optimal transportation problem. This problem has been studied extensively over the past 25 years, and has many applications and deep connections to various areas of mathematics. Under a structural condition on the cost function, that is in particular satisfied by the Riemannian distance squared, and a fairly weak regularity condition on the first marginal, it is now well known that the optimal measure is concentrated on the graph \( \{(x_1, F_2(x_1)) : x_1 \in \mathcal{M}\} \) of a map \( F_2 : \mathcal{M} \to \mathcal{M} \). The map \( F_2 \) minimizes (M), and both \( F_2 \) and \( \gamma \) are unique [17] [11] [12] [2] [18].

In recent years, the multi-marginal case, \( m \geq 3 \) has attracted increasing attention, due to emerging applications in areas such as economics [5] [7], physics [9] [3], cyclical monotonicity [10, 14, 15], and systems of elliptic equations [16]. In contrast to the two marginal case, the structure of solutions for general cost functions is not well understood. For costs with certain special properties, however, a number of authors have proven that, like in the two marginal case, the optimal measure \( \gamma \) is concentrated on the graph of a function over \( x_1 \) and is unique [13] [4] [21] [20] [19]. On the other hand, it is known that for certain other cost functions, the optimal measure \( \gamma \) in the \( m \geq 3 \) case may be concentrated on a set of Hausdorff dimension larger than \( n := \text{dim}(\mathcal{M}) \) and be non unique [6] [21]. It remains an open question to determine precisely which costs permit Monge solution and uniqueness results and which do not.

The most important cost in the two marginal case is the quadratic cost, either \( c(x_1, x_2) = |x_1 - x_2|^2 \) on \( \mathcal{M} \subseteq \mathbb{R}^n \), or \( c(x_1, x_2) = d(x_1, x_2)^2 \), where \( d \) denotes the distance on a Riemannian manifold \( \mathcal{M} \). Uniqueness and Monge solution results for these two cost functions constitute famous theorems of Brenier [2] and McCann [18], respectively; these two results are at the heart of optimal transport theory and underlie many of it applications.

The best known result on multi-marginal problems concerns the cost function \( \sum_{i \neq j} |x_i - x_j|^2 \) on Euclidean space, \( \mathcal{M} \subseteq \mathbb{R}^n \). For this cost, Gangbo and Swiech [13] proved uniqueness and existence of Monge solution, amounting to a generalization of Brenier’s theorem from 2 to \( m \geq 3 \) marginals.

The purpose of this paper is to generalize McCann’s result on manifolds to the multi-marginal case, in the same spirit that Gangbo and Swiech’s theorem
generalizes the result of Brenier. Alternatively, we can think of our main result as generalizing Gangbo and Swiech’s result to manifolds, in the same sense that McCann’s theorem generalizes Brenier’s. We will prove uniqueness and Monge solution results for the multi-marginal problem on a compact Riemannian manifold, with cost function

$$c(x_1, x_2, ..., x_m) = \inf_{y \in M} \sum_{i=1}^{m} \frac{d^2(x_i, y)}{2}.$$  \tag{1}$$

Let us note that, while the cost \(\sum_{i \neq j} d(x_i, x_j)^2\) is a direct, algebraic generalization of the Gangbo-Swiech cost, the cost function in (1) seems to be more intimately linked with the underlying geometry of the manifold \(M\). This cost function measures the average distance squared between the points \(x_1, x_2, ..., x_m\) and their barycenter (also known as the Frechet or Karcher mean). Alternatively, we can view \((x_1, \cdot \cdot \cdot, x_n)\) as a point in the Riemannian product \(M \times \cdots \times M\) equipped with the product distance \(\sqrt{\sum_{i=1}^{m} d^2(x_i, y_i)}\) and the cost function in (1) measures the distance from \((x_1, \cdot \cdot \cdot, x_m)\) to the diagonal set \(\{(y, \cdot \cdot \cdot, y)\} \times M\). Straightforward calculations in [19] show that the cost functions of Brenier (when \(m = 2\) on \(\mathbb{R}^n\)), McCann (when \(m = 2\) on a Riemannian manifold) and Gangbo-Swiech (for general \(m\) on \(\mathbb{R}^n\)) are all equal to cost (1) in the appropriate settings, and so (1) for \(m > 2\) on a Riemannian manifold is indeed an extension of these costs. Furthermore, there is an intimate link between optimal transport with cost (1) and the notion of the barycenter of the measures \(\mu_1, \mu_2, ..., \mu_2\) as considered by Agueh and Carlier [1] in the Euclidean case, following the work of Carlier and Ekeland [5] who considered more a general version of the cost function (1) in economic applications (see cost (3) below).

Our main theorem is the following.

**Theorem 1.1.** Assume \(\mu_1\) is absolutely continuous with respect to local coordinates. Then the solution \(\gamma\) to (K) with cost function (1) is concentrated on the graph of a mapping \((F_2, F_3, ..., F_m)\) over the first variable. This mapping is a solution to Monge’s problem (M), and the solutions to both (K) and (M) are unique.

This theorem is, to the best of our knowledge, the first of its kind for multi-marginal problems on manifolds with topology different from \(\mathbb{R}^n\). We also note that one can extend Theorem 1.1 to slightly more general costs, similar to (1): see Theorem 6.1.

The Gangbo-Swiech optimal maps are all compositions of Brenier maps, \(\nabla u_i^* \circ \nabla u_1\), where the \(u_i^*\) and \(u_1\) are convex functions; this structure was further clarified by Agueh and Carlier [1], who showed that the first map, \(\nabla u_1\), is the optimal map pushing \(\mu_1\) forward to the Wasserstein barycenter (see (1)) of the measures \(\mu_1, \cdot \cdot \cdot, \mu_m\); while the second map \(\nabla u_i^*\) is the optimal mapping pushing the barycenter forward to the measure \(\mu_i\) [1]. Our optimizers here take the same form, but using McCann maps rather than Brenier maps to account for
the curved geometry: the optimizers are of the form

\[ F_i(x_1) = \exp_{x_1}(\nabla u_i(x_1)) \left( \exp_{x_1} \left( \nabla u_i \left( \exp_{x_1} \left( \nabla u_1(x_1) \right) \right) \right) \right) \]  

(2)

for \( \frac{d^2}{dx^2} \) concave \( u_1, u_i \); that is, they are compositions of McCann maps (see Proposition 5.1). Therefore, our optimal map is built out of McCann maps, in the same sense that the Gangbo-Swiech maps are built from Brenier maps.

Let us mention that multi-marginal problems with cost functions of the general form

\[ c(x_1, x_2, \ldots, x_m) = \inf_y \sum_{i=1}^{m} c_i(x_i, y) \]  

(3)

arise naturally in matching problems in economics as considered by Carlier and Ekeland [5] and Chiapporri, McCann and Nesheim, [7]. In this setting, solving the multi-marginal problem can be interpreted as finding an equilibrium assignment of \( m \) different groups of workers to contracts \( y \).

Motivated in part by this application, problems of costs of the form (3) were studied by one of us in a recent paper [19], and conditions on the \( c_i \)'s were found that ensured existence and uniqueness of Monge solutions [19]. However, that paper was restricted to Euclidean space, \( M \subseteq \mathbb{R}^n \), and required twist and a type of non-degeneracy conditions on the cost functions \( c_i \); topological obstructions prevent these hypotheses from holding on, for example, a compact manifold. It should be noted, however, that the results in [19] do apply to the quadratic cost on bounded domains in a Hadamard manifold (a simply connected Riemannian manifold with nonpositive sectional curvature); in this case, the Cartan-Hadamard Theorem ensures that \( M \) has the topology of \( \mathbb{R}^n \).

Our method for the proof of Theorem 1.1 is a bit different from the approach in [19], which has its origin in the papers of Carlier and Ekeland [5] and Agueh and Carlier [11]. In that work, it was crucial to have the absolute continuity of the (generalized) barycenter measure with respect to local coordinates, since the optimal maps \( F_i \)'s were constructed from Brenier theory applied to optimal maps from the barycenter measure, say \( \nu \) (see (11)), to the target measure \( \mu_i \), whose existence requires absolute continuity of the source distribution \( \nu \). In the Euclidean setting, the absolute continuity of \( \nu \) has been shown by Agueh and Carlier [11], but its Riemannian extension is still not known except the case of nonpositively curved manifold [19]. In our method, we directly (without relying on absolute continuity of \( \nu \)) show that the solution \( \gamma \) to (K) is concentrated on a graph of map. First, we will show that for \( \mu_1 \) almost every \( x_1 \), there is a unique point \( y \) attaining the minimum in (11), for \( x_2, \ldots, x_m \) such that \( (x_1, x_2, \ldots, x_m) \) is in the support of the optimal measure \( \gamma \) in (11). We will then show, in turn, that for each \( y \), there is at most 1 corresponding point \( (x_1, \ldots, x_m) \) in the support of \( \gamma \); together, these two facts will imply the main result. It is also remarkable that the structure of the maps \( F_i \) as in (2) has been obtained without relying on absolute continuity of \( \nu \), while the maps \( \exp_y \left( \nabla u_i(y) \right) \) in (2) are Brenier maps from \( \nu \) to \( \mu_i \).
In Section 2, we recall the dual problem to (K) and some of its properties. In the third section, we will prove several lemmas which are needed for the proof of Theorem 1.1 while the Theorem itself will be proved in Section 4. In Section 5, we discuss the connection of (K) and (M) with the Wasserstein barycenter of the measures µ₁, ..., µₘ, showing the structure of the maps Fᵢ. In the final section, we extend our results to other functions of the distance on a Riemannian manifold.

2 Duality

Here we recall known results on the Kantorovich dual problem, and make some remarks on the properties of its solution which will be relevant later on.

The dual problem to (K) is to minimize,

\[ \sum_{i} \int_{M} u_i(x_i) d\mu_i(x_i) \]  \hspace{1cm} (D)

among all \( m \)-tuples of functions \((u_1, u_2, \ldots, u_m)\) of functions, with \( u_i \in L^1(\mu_i) \) and

\[ \sum_{i=1}^{m} u_i(x_i) \leq c(x_1, x_2, \ldots, x_m) \] \hspace{1cm} (4)

for \( \otimes_{i=1}^{m} \mu_i \) almost everywhere point \((x_1, x_m)\). We will say that an \( m \)-tuple \((u_1, u_2, \ldots, u_m)\) is c-conjugate if, for all \( i = 1, 2, \ldots, m \), we have

\[ u_i(x_i) = \inf_{x_j \in M, j \neq i} \left[ c(x_1, x_2, \ldots, x_m) - \sum_{j \neq i} u_j(x_j) \right]. \] \hspace{1cm} (5)

It is clear that any c-conjugate \((u_1, u_2, \ldots, u_m)\) satisfies (4), so is a viable competitor in (D).

The following result is well known:

**Theorem 2.1.** There exists a c-conjugate solution \((u_1, u_2, \ldots, u_m)\) to (D). If \( \gamma \) is any optimal measure in the Kantorovich problem, we have

\[ \sum_{i=1}^{m} u_i(x_i) = c(x_1, x_2, \ldots, x_m) \] \hspace{1cm} \( \gamma \)-almost everywhere.

It is also well known that if the cost \( c \) is Lipschitz and \((u_1, u_2, \ldots, u_m)\) is a c-conjugate m-tuple such that if \( u_1 \) is not identically infinity, then \( u_1 \) must be Lipschitz [13]. Our cost function (1) is Lipschitz, since it is defined as the infimum of a family of Lipschitz functions \((x_1, \ldots, x_m) \mapsto \sum_{i=1}^{m} \frac{d^2}{2}(x_i, y)\).
3 Properties of the cost function and consequences of optimality of $\gamma$

In this section, we establish several key lemmas for the proof of Theorem 1.1. Throughout this section, $c(x_1, \ldots, x_m)$ is the cost function as in (1), $\gamma$ denotes an optimal measure of the Kantorovich problem (K), and $(u_1, \ldots, u_m)$ is a $c$-conjugate solution to (D).

Lemma 3.1. Fix $(x_1, \ldots, x_m)$. Then any $y$ which minimizes $y \mapsto \sum_{i=1}^{m} d^2(x_i, y)$ is not in the cut locus of $x_i$ for any $i$.

Proof. Choose a point $y$ in the cut locus of $x_i$ for some $i$; we will show that $y$ cannot minimize $y \mapsto \sum_{i=1}^{m} d^2(x_i, y)$. By Lemma 3.12 in [8], we can find a constant $K$ such that, for all $u \in T_y M$, and $j = 1, 2, \ldots, m$, we have

$$\frac{d^2(x_j, \exp_y u) + d^2(x_j, \exp_y (-u)) - 2d^2(x_j, y)}{|u|^2} \leq K. \quad (6)$$

On the other hand, by Proposition 2.5 in the same paper, we can find some non zero $u \in T_y M$ such that

$$\frac{d^2(x_i, \exp_y u) + d^2(x_i, \exp_y (-u)) - 2d^2(x_i, y)}{|u|^2} \leq -mK. \quad (7)$$

Therefore, we have

$$\sum_{j=1}^{m} d^2(x_j, y) = \sum_{j \neq i}^{m} d^2(x_j, y) + d^2(x_i, y) \geq \frac{-(m-1)K|u|^2}{2} + \frac{1}{2} \sum_{j \neq i}^{m} (d^2(x_j, \exp_y u) + d^2(x_j, \exp_y (-u))) + mK|u|^2 + \frac{1}{2} d^2(x_i, \exp_y u) + d^2(x_i, \exp_y (-u)) > \frac{1}{2} \sum_{j=1}^{m} d^2(x_j, \exp_y u) + \frac{1}{2} \sum_{j=1}^{m} d^2(x_j, \exp_y (-u)).$$

Therefore, either

$$\sum_{j=1}^{m} d^2(x_j, \exp_y u) < \sum_{j=1}^{m} d^2(x_j, y),$$

or

$$\sum_{j=1}^{m} d^2(x_j, \exp_y (-u)) < \sum_{j=1}^{m} d^2(x_j, y);$$

in either case, $y$ cannot minimize $\sum_{j=1}^{m} d^2(x_j, y)$. \qed
Lemma 3.2. The cost function $c$ is everywhere superdifferentiable with respect to $x_1$. That is, for all $(x_1, x_2, ..., x_m) \in M^m$ there exist $p \in T_{x_1}M$ (the supergradient) such that, for small $v \in T_{x_1}M$, we have

$$c(\exp_{x_1} v, x_2, ..., x_m) \leq c(x_1, x_2, ..., x_m) + g(p, v) + o(|v|),$$

where $g$ denotes the metric.

Proof. Choose $y$ minimizing $y \mapsto \sum_i \frac{d^2}{2}(x_i, y)$. By \cite{15}, Proposition 6, the function $x_1 \mapsto d^2(x_1, y)$ is superdifferentiable. Letting $p$ be it’s supergradient, we have, for small $v$

$$\frac{d^2}{2}(\exp_{x_1} v, y) \leq \frac{d^2}{2}(x_1, y) + g(p, v) + o(|v|).$$

Now, by definition, we have

$$c(\exp_{x_1} v, x_2, ..., x_m) \leq \frac{d^2}{2}(\exp_{x_1} v, y) + \sum_{i=2}^m \frac{d^2}{2}(x_i, y) \leq \frac{d^2}{2}(x_1, y) + g(p, v) + o(|v|) + \sum_{i=2}^m \frac{d^2}{2}(x_i, y) = c(x_1, x_2, ..., x_m) + g(p, v) + o(|v|).$$

Therefore, $c$ is superdifferentiable with respect to $x_1$, with supergradient $p$. □

Lemma 3.3. At any point $(x_1, ..., x_m)$ where $c$ is differentiable with respect to $x_1$, there is a unique minimizing $y$ in (1), and moreover,

$$y = \exp_{x_1}(\nabla_{x_1} c(x_1, ..., x_m)). \quad (8)$$

Proof. For any minimizing $y$ in (1), $d^2(x_1, y)$ is differentiable as $y \notin \text{cut}(x_1)$ by Lemma \cite{15}. We then have

$$\nabla_{x_1} c(x_1, ..., x_m) = \nabla_{x_1}(\frac{1}{2} d^2(x_1, y)).$$

This equation implies that $y$ must equal $\exp_{x_1}(\nabla_{x_1} c(x_1, ..., x_m))$; uniqueness follows immediately. □

Remark 1. This lemma may be of independent interest. The minimizing $y$ in (1) is known as the Frechet or Karcher mean of the points $x_1, x_2, ..., x_m$, and can be interpreted as a generalization of the notion of average to a curved space. Uniqueness of Frechet means is linked the curvature of the manifold $M$. For example, Frechet means are easily seen to be unique on a Hadamard manifold, but may be non-unique for certain $x_1$ on spaces with positive sectional curvature; for example, every point on the equator is a Frechet mean of the north and south
poles on the sphere. The preceding lemma yields a sufficient condition for the Frechet mean of the points \(x_1, x_2, \ldots, x_m\) to be unique; namely, that the function \(c\) be differentiable with respect to one of the variables. We will see later that this in fact implies that, if \(\mu_1\) does not charge small sets, the Frechet mean of the points \((x_1, x_2, \ldots, x_m)\) is unique \(\gamma\) almost everywhere, where \(\gamma\) is the optimal measure in \([K]\).

**Lemma 3.4.** Let \((x_1, x_2, \ldots, x_m) \in \text{spt}(\gamma)\), and suppose the potential \(u_1\) is differentiable at \(x_1\). Then the cost \(c\) is differentiable with respect to \(x_1\) at \((x_1, x_2, \ldots, x_m)\) and moreover,

\[
\nabla_{x_1} c(x_1, \ldots, x_m) = \nabla_{x_1} u_1(x_1)
\]

(9)

**Proof.** The proof is based on a now classical argument of McCann [18]. By the properties [11] and [15], Theorem [2] and the fact that \((x_1, \ldots, x_m) \in \text{spt}(\gamma)\), the gradient \(q = \nabla u_1(x_1)\) serves as a subgradient for \(c\) with respect to \(x_1\). By lemma 3.4, \(c\) has a supergradient, \(p\), as well. Therefore, we conclude that \(c\) must be differentiable with respect to \(x_1\), with gradient \(p = q = \nabla u_1(x_1)\). □

Whenever the optimal \(y\) in (11) is unique, we will denote it by \(\overline{y}(x_1, \ldots, x_m)\).

The following proposition implies injectivity of this correspondence from the support of \(\gamma\).

**Lemma 3.5.** Suppose \(x = (x_1, \ldots, x_m)\) and \(\bar{x} = (\bar{x}_1, \ldots, \bar{x}_m)\) are both in \(\text{spt}(\gamma)\) and there exists \(y \in M\) which minimizes both \(z \mapsto \sum_{i=1}^{m} d^2(x_i, z)\) and \(z \mapsto \sum_{i=1}^{m} d^2(\bar{x}_i, z)\). Then \(x = \bar{x}\).

**Proof.** Fix some \(i\). We will show \(x_i = \bar{x}_i\).

Define \(x' = (x_1, \ldots, x_i, \ldots, x_k)\) and \(\bar{x}' = (\bar{x}_1, \ldots, x_i, \ldots, \bar{x}_k)\). Recalling that \(c(x) = \inf_{z \in M} \sum \frac{d^2}{2}(x_i, z)\), we have

\[
c(x') + c(\bar{x}') \leq \left[ \sum_{j \neq i} \frac{d^2}{2}(x_j, y) \right] + \frac{d^2}{2}(\bar{x}_i, y) + \left[ \sum_{j \neq i} \frac{d^2}{2}(\bar{x}_j, y) \right] + \frac{d^2}{2}(x_i, y)
\]

(10)

\[
= c(x) + c(\bar{x}),
\]

where the last equality follows from rearranging the terms and using the assumption

\[
y \in \text{argmin}_{z \in M} \sum_{i=1}^{m} \frac{d^2}{2}(x_i, z) \cap \text{argmin}_{z \in M} \sum_{i=1}^{m} \frac{d^2}{2}(\bar{x}_i, z).
\]

Also, the optimality of \(\gamma\) implies a certain \(c\)-monotonicity property [21]: as \(x, \bar{x} \in \text{spt}\gamma\), this yields,

\[
c(x') + c(\bar{x}') \geq c(x) + c(\bar{x}).
\]

Moreover, from [10], this implies that \(c(x') = [\sum_{j \neq i} \frac{d^2}{2}(x_j, y)] + \frac{d^2}{2}(\bar{x}_i, y)\); that is,
\begin{align*}
y \in \arg\min_{z \in M} \left[ \sum_{j \neq i} \frac{d^2}{2} (x_j, z) + \frac{d^2}{2} (\bar{x}_i, z) \right]
\end{align*}

The Riemannian distance is smooth away from the cut locus, so, by Lemma \ref{lem:3}, each $z \mapsto d^2 / 2 (x_i, z)$ and each $z \mapsto d^2 / 2 (\bar{x}_i, z)$ is differentiable at $z = y$. Therefore, as $y \in \arg\min_z \left[ \sum_j d^2 (x_j, z) \right]$ and $y \in \arg\min_z \left[ \sum_{j \neq i} d^2 (x_j, z) + d^2 (\bar{x}_i, z) \right]$, we have:

\begin{align*}
\sum_j \nabla_y d^2 (x_j, y) &= 0 = \sum_{j \neq i} \nabla_y d^2 (x_j, y) + \nabla_y d^2 (\bar{x}_i, y).
\end{align*}

This implies

\begin{align*}
\nabla_y d^2 (x_i, y) &= \nabla_y d^2 (\bar{x}_i, y)
\end{align*}

It is well known that for $x \notin \text{cut}(y)$, $x = \exp_y (\frac{1}{2} \nabla_y d^2 (x, y))$ and so the above implies

\begin{align*}
x_i &= \exp_y (\frac{1}{2} \nabla_y d^2 (x_i, y)) = \exp_y (\frac{1}{2} \nabla_y d^2 (\bar{x}_i, y)) = \bar{x}_i.
\end{align*}

as desired. \qed

\section{Main result}

We are now ready to prove Theorem \ref{thm:main}.\protect

\begin{proof}
We first consider the assertion that optimal measures must be concentrated on a graph.

For any optimal measure $\gamma$, we need to show that, for $\mu_1$ almost every $x_1$, there is a unique $(x_2, x_3, \ldots, x_m)$ such that $(x_1, x_2, x_3, \ldots, x_m) \in \text{spt}(\gamma)$. As the optimal measure $\gamma$ projects to $\mu_1$, for $\mu_1$ a.e. $x_1$, we must have at least one $(x_2, x_3, \ldots, x_m)$ such that $(x_1, x_2, x_3, \ldots, x_m) \in \text{spt}(\gamma)$. Therefore, it remain only to prove uniqueness of the $(x_2, x_3, \ldots, x_m)$.

As mentioned in Section 2, it is well known that the Kantorovich potential $u_1$ is Lipschitz, and hence differentiable $\mu_1$ almost everywhere \cite{18}. Now, at every point where the potential is differentiable, the cost $c$ is differentiable with respect to $x_1$, by Lemma \ref{lem:3}. Thus from \cite{8} and \cite{9}, we see that there is a unique $y$ minimizing \cite{1}, for all $(x_1, x_2, x_3, \ldots, x_m) \in \text{spt}(\gamma)$; namely

\begin{align*}
y = \exp_{x_1} (\nabla u(x_1)).
\end{align*}

It follows from Lemma \ref{lem:4} that for this $y$, there is at most one $(x_1, x_2, \ldots, x_m) \in \text{spt}(\gamma)$.
with \( y \in \arg\min \sum_i d^2(x_i, y) \), showing the uniqueness of \((x_2, ..., x_m)\). Denote this point

\[ G(y) := (G_1(y), G_2(y), ..., G_m(y)). \]

Then, denoting \( F_i(x_i) = G_i(\exp_{x_i}(\nabla u(x_i))) \), \( F_i \) is a well defined \( \mu_1 \)-almost everywhere map with the desired properties.

It remains to prove uniqueness of the optimal measure \( \gamma \) and the optimal maps \( F_i \); our argument for this is standard in optimal transport theory. Suppose there are two distinct optimizers, \( \gamma \) and \( \bar{\gamma} \); by the argument above, they are concentrated on graphs, \((F_2, F_3, ..., F_m)\) and \((\bar{F}_2, \bar{F}_3, ..., \bar{F}_m)\), respectively. Now, by linearity of the Kantorovitch functional, the interpolant \( \frac{1}{2} \gamma + \frac{1}{2} \bar{\gamma} \) is also so optimal and must also be concentrated on a graph. This immediately implies that \((F_2, F_3, ..., F_m) = (\bar{F}_2, \bar{F}_3, ..., \bar{F}_m)\) almost everywhere, completing the proof.

5 Barycenters in Wasserstein space on Riemannian manifolds

In this section, we discuss the connection of our result with the Wasserstein barycenter of \( \mu_1, ..., \mu_2 \). The barycenter (with equal weights) is defined as the Borel probability measure \( \nu \) on \( M \) which minimizes

\[ \nu \mapsto \sum_{i=1}^m W^2_2(\mu_i, \nu), \quad (11) \]

where \( W^2_2(\mu, \nu) \) denotes the square of the quadratic Wasserstein distance from \( \mu_i \) to \( \nu \). The relevance of the barycenter to the multi-marginal optimal mapping problem \([M]\) was observed and investigated by Carlier and Ekeland \([5]\) and by Agueh and Carlier \([1]\). Existence of the barycenter follows easily from a continuity compactness argument (see \([5]\), Theorem 3). When \( \mu_1 \) is absolutely continuous, uniqueness of the barycenter was shown in \([1]\) for the Euclidean case \( M \subseteq \mathbb{R}^n \), and can be easily generalized to the Riemannian case either from Theorem 3.4.1 in \([22]\) or alternatively, by adapting Proposition 4 in \([5]\), using McCann’s theorem \([18]\) in place of the twist, or generalized Spence-Mirrlees, condition.

Letting \( \gamma \) denote the optimal measure in the multi-marginal problem \([K]\), and \( \bar{y}(x_1, x_2, ..., x_m) \) the minimizer of \( y \mapsto \sum_{i=1}^m d^2(x_i, y) \) (which, by Lemma 3.3, is unique for \( \gamma \) almost all \((x_1, x_2, ..., x_m)\)), a result of Carlier and Ekeland \([5]\) implies that

\[ \nu := \bar{y} \# \gamma \]

is the unique barycenter\([1]\). They also showed that, for each \( i \), the measure \( \gamma_i \)

\[ \begin{align*}
\text{Note that the minimizer of } y & \mapsto \sum_{i=1}^m d^2(x_i, y) \text{ is not necessarily unique for all } (x_1, x_2, ..., x_m) \text{ here, as was assumed by Carlier and Ekeland in } [5]. \text{ However, it is clear by examining their proof that it is sufficient to have uniqueness for } \gamma \text{ almost all } (x_1, x_2, ..., x_m). \end{align*} \]
defined on $M \times M$ by

$$
\gamma_i = (\pi_i, \bar{y}) \# \gamma
$$

is optimal for the two marginal Monge-Kantorovich problem

$$
\inf_{\gamma_i} \int_{M \times M} d^2(x, y) d\gamma_i
$$

with marginals $\mu_i$ and $\nu$. Here $\pi_i(x_1, x_2, ..., x_m) = x_i$ is the canonical projection.

The following proposition further highlights the relationship between the barycenter and the optimal maps in Theorem 1.1.

**Proposition 5.1.** Use the assumption and notation as in Theorem 1.1. Then, the mappings $F_i(x_1)$ in Theorem 1.1 are of the form $G_i \circ G_1^{-1}$, where $G_1^{-1}$ is the optimal map (for the quadratic cost $d^2(x_i, y)$) pushing $\mu_1$ forward to the barycenter $\nu$, and $G_i$ is the optimal map pushing $\nu$ forward to the measure $\mu_i$.

**Remark 5.2.** This generalizes the result of Gangbo and Swiech [13], who showed that when $M \subseteq \mathbb{R}^n$, the optimal maps for $(M)$ take the form $\nabla u_i^* \circ \nabla u_1$, for convex functions $u_i$ and their conjugates $u_i^*$. Agueh and Carlier later realized that the maps $Du_i^*$ were in fact the Brenier (optimal) maps pushing the barycenter forward to $\mu_i$, and $Du_1$ the Brenier map pushing $\mu_1$ forward to the barycenter [1].

**Proof.** Letting $(u_1, u_2, ..., u_m)$ be a c-conjugate solution to the dual problem $(D)$, we have

$$
\begin{align*}
u_i(x_i) &= \inf_{x_j, j \neq i} \left[ c(x_1, x_2, ..., x_m) - \sum_{i \neq j} u_j(x_j) \right] \\
&= \inf_{x_j, j \neq i} \left[ \left( \inf_{k=1}^m \frac{d^2(x_k, y)}{2} \right) - \sum_{j \neq i} u_j(x_j) \right] \\
&= \inf_{y} \left[ \frac{d^2(x, y)}{2} + \sum_{j \neq i} \inf_{x_j} \left[ \frac{d^2(x_j, y) - u_j(x_j)}{2} \right] \right].
\end{align*}
$$

This implies that $\nu_i$ is $\frac{d^2}{2}$-concave. Setting $v_i(y) = -\sum_{j \neq i} \inf_{x_j} \frac{d^2}{2}(x_j, y) - u_j(x_j)$, we have

$$
\nu_i(x_i) + v_i(y) \leq \frac{d^2}{2}(x_i, y)
$$

and we have equality when

$$
x \in S := \{ x = (x_1, ..., x_m) : c(x_1, x_2, ..., x_m) - \sum_{j=1}^m u_j(x_j) \},
$$
that is, \((u_i, v_i)\) solve the Kantorovich dual problem to \((\ref{eq:KantorovichProblem})\).

Now, from the proof of Theorem 1.1 we have \(F_i(x_1) = G_i(\exp_{x_1} (\nabla u_1(x_1)))\) (for a.e \(x_1\) thus, for \(\mu_1\)-a.e \(x_1\) by our assumption that \(\mu_1\) is absolutely continuous with respect to local coordinates). Also, from \((\ref{eq:KantorovichProblem})\), \(u_1\) is \(\frac{d^2}{2}\)-concave, therefore by McCann’s theorem the map \(\exp_{x_1} (\nabla u_1(x_1))\) is optimal. As for \(\mu_1\)-a.e. \(x_1\) we have equality in \((\ref{eq:KantorovichProblem})\) only when \(y = y(x_1, F_2(x_1), \ldots, F_m(x_1))\), it is clear that \(\exp_{x_1} (\nabla u_1(x_1))\#\mu_1 = \bar{y} \# \gamma = \nu\).

It remains to show that the maps \(G_i\) are optimal maps pushing \(\nu\) to \(\mu_i\). Let \(u_i^c(y) := \inf_{x_i \in M} \frac{d^2}{2}(x_i, y) - u_i(x_i)\) be the \(\frac{d^2}{2}\) concave conjugate of \(u_i\). Note that we have \(v_i(y) \leq u_i^c(y)\), and we have equality whenever there is some \(x_i\) achieving equality in \((\ref{eq:KantorovichProblem})\), which happens whenever \(y = \bar{y}(x)\) for some \(x \in S\).

Standard arguments now imply that \(G_i(y) = \exp_y (\nabla u_i^c(y))\) wherever \(u_i^c\) is differentiable; we must show that this is \(\nu\)-almost everywhere. Although the semi-concave function \(u_i^c\) is differentiable almost everywhere with respect to local coordinates, the conclusion is not completely obvious, as we do not know that the measure \(\nu\) does not charge small sets. (This latter fact is known for the Euclidean case, or equivalently for the Gangbo–Świȩch cost as was shown by Agueh and Carlier.)

This difficulty can easily be overcome thanks to Lemma 3.5. Details follow:

Fix \(y \in \bar{y}(S)\). (Notice that for \(\nu\)-almost everywhere points \(y\), we have \(y \in \bar{y}(S)\).) Note that the c-concave function \(u_i^c\) is superdifferentiable everywhere (since the relevant domains, i.e. support of measures, etc, are compact), and that at any point \(x_i\) where

\[ u_i(x_i) + u_i^c(y) = \frac{d^2}{2}(x_i, y), \]

the vector \(\nabla_y \frac{d^2}{2}(x_i, y)\) is in the superdifferential \(\partial u_i^c(y)\). Also, as is well-known for concave and for c-concave functions, we see that at any \(x_i\), if \(\nabla_y \frac{d^2}{2}(x_i, y)\) is an extremal point of the convex set \(\partial u_i^c(y)\), we must have

\[ u_i(x_i) + u_i^c(y) = \frac{d^2}{2}(x_i, y). \]

Furthermore, as for \(y \in \bar{y}(S)\) we have \(v(y) = u_i^c(y)\), we have equality in \((\ref{eq:KantorovichProblem})\) for each such \(x_i\).

In sum, for each \(y \in \bar{y}(S)\) we have shown that for each \(x_i\), if \(\nabla_y \frac{d^2}{2}(x_i, y)\) is extremal in \(\partial u_i^c(y)\), then we have equality in \((\ref{eq:KantorovichProblem})\). But by Lemma 3.5 there is exactly one point \(x_i \in M\) for which equality holds in \((\ref{eq:KantorovichProblem})\). It follows that, at each point \(y \in \bar{y}(S)\), \(\partial u_i^c(y)\) has only one extremal point, hence \(u_i^c\) is differentiable on \(\bar{y}(S)\). Since \(y \in \bar{y}(S)\) for \(\nu\)-a.e. points \(y\), it now follows that

\[ G_i(y) = \exp_y (\nabla u_i^c(y)) \]

for \(\nu\)-a.e. \(y\) and so that

\[ F_i(x_1) = \exp_{\exp_{x_1} (\nabla u_1(x_1))} (\nabla u_i^c(\exp_{x_1} (\nabla u_1(x_1)))) \]

for \(\mu_1\)-a.e. \(x_1\) as desired.
6 Extension to other functions of the distance

In this section we consider the extension of our main theorem to non-quadratic cost functions. Precisely, given \( C^2 \), strictly increasing, strictly convex functions \( f_i : [0, \infty) \to \mathbb{R} \), for \( i = 1, 2, ..., m \), we consider cost functions of the form

\[
c(x_1, x_2, ..., x_m) = \inf_{y \in \mathcal{M}} \sum_{i=1}^{m} f_i(d(x_i, y)).
\]

(17)

Note that the previous part of the paper deals with the case \( f_i(t) = t^2 \).

**Theorem 6.1.** Assume \( \mu_1 \) is absolutely continuous with respect to local coordinates. Then the solution \( \gamma \) to \([K]\) with cost function \([L]\) is concentrated on the graph of a function \((F_2, F_3, ..., F_m)\) over the first variable. This function is a solution to Monge’s problem \([M]\), and the solutions to both \([K]\) and \([M]\) are unique.

**Proof.** The proof is very similar to the proof of Theorem [11] and we only sketch it here, explaining how to deal with the main differences. The main difference is to verify that a minimizer \( y \in \arg\min \sum_{i=1}^{m} f_i(d(x_i, y)) \) is not in the cutlocus of any \( x_i \) (an analogue of Lemma [8]). This follows from a Taylor expansion. Details are given below:

Note that, by Taylor’s theorem, for some \( p \in [d(x_i, \exp_y(u)), d(x_i, y)] \)

\[
f_i(d(x_i, \exp_y(u))) - f_i(d(x_i, y)) = f'_i(d(x_i, y))(d(x_i, \exp_y(u)) - d(x_i, y))
\]

\[
+ f''_i(p) \frac{(d(x_i, \exp_y(u)) - d(x_i, y))^2}{2}.
\]

(18)

Similarly, for some \( q \in [d(x_i, \exp_y(-u)), d(x_i, y)] \)

\[
f_i(d(x_i, \exp_y(u))) - f_i(d(x_i, y)) = f'_i(d(x_i, y))
\]

\[
+ f''_i(q) \frac{(d(x_i, \exp_y(-u)) - d(x_i, y))^2}{2}.
\]

(19)

(20)

(21)

By monotonicity of \( f_i \), we have \( f'_i(d(x_i, y)) > 0 \) and by convexity we have \( f''_i(p), f''_i(q) \geq 0 \). Also, we have (e.g. see [13])

\[
(d(x_i, \exp_y(-u)) - d(x_i, y))^2 \leq C|u|^2 + o(|u|).
\]
Note that near the cutlocus, the factor \( d(x_i, \exp_y(u)) + d(x_i, y) \) is bounded below by \( I \), the injectivity radius of the compact manifold \( M \) (away from the cutlocus, inequality (22) below follows easily from smoothness). Using Lemma 3.12 in [8], and combining (19) and (21), we obtain

\[
 f_i(d(x_i, \exp_y(u))) + f_i(d(x_i, \exp_y(-u))) - 2f_i(d(x_i, y)) \leq K|u|^2 + o(|u|^2) \tag{22}
\]

for some \( K > 0 \). Now, if \( y \in \text{cut}(x_j) \), we can, by Proposition 2.5 in [8], for any \( A > 0 \), choose a \( u \) such that

\[
d^2(x_j, \exp_y(u)) + d^2(x_j, \exp_y(-u)) - 2d^2(x_j, y) \leq -A|u|^2. \tag{23}
\]

Now, the factor \( d(x_i, \exp_y(u)) + d(x_i, y) \) is bounded above by \( 2R \), where \( R \) is the radius of the compact manifold. Therefore, for any \( B > 0 \), combining (19) and (21) and (23), we can find arbitrarily small \( u \) such that

\[
f_i(d(x_j, \exp_y(u))) + f_i(d(x_j, \exp_y(-u))) - 2f_i(d(x_j, y)) \leq -B|u|^2 + o(|u|^2).
\]

Now, choosing \( B \) large enough, and arguing as in Lemma 3.1, we obtain either

\[
\sum_{i=1}^{m} f_i(d(x_i, \exp_y(u))) < \sum_{i=1}^{m} f_i(d(x_i, y))
\]

or

\[
\sum_{i=1}^{m} f_i(d(x_i, \exp_y(-u))) < \sum_{i=1}^{m} f_i(d(x_i, y)).
\]

In either case, \( y \notin \text{argmin} \sum_{i=1}^{m} f_i(d(x_i, y)) \). The remainder of the proof, including analogues of Lemmas 3.3 and 3.5, is very similar to the arguments developed earlier in the paper and is omitted.

References

[1] M. Agueh and G. Carlier. Barycenters in the Wasserstein space. *SIAM J. Math. Anal.*, 43(2):904–924, 2011.

[2] Y. Brenier. Decomposition polaire et rearrangement monotone des champs de vecteurs. *C.R. Acad. Sci. Pair. Ser. I Math.*, 305:805–808, 1987.

[3] Giuseppe Buttazzo, Luigi De Pascale, and Paola Gori-Giorgi. Optimal-transport formulation of electronic density-functional theory. *Phys. Rev. A*, 85:062502, Jun 2012.

[4] G. Carlier. On a class of multidimensional optimal transportation problems. *J. Convex Anal.*, 10(2):517–529, 2003.
[5] G. Carlier and I. Ekeland. Matching for teams. *Econom. Theory*, 42(2):397–418, 2010.

[6] G. Carlier and B. Nazaret. Optimal transportation for the determinant. *ESAIM Control Optim. Calc. Var.*, 14(4):678–698, 2008.

[7] P-A. Chiapporri, R. McCann, and L. Nesheim. Hedonic price equilibria, stable matching and optimal transport; equivalence, topology and uniqueness. *Econom. Theory.*, 42(2):317–354, 2010.

[8] D. Cordero-Erausquin, R. J. McCann, and M. Schmuckenschläger. A Riemannian interpolation inequality a la Borell, Brascamp and Lieb. *Invent. Math.*, 146(2):219–257, 2001.

[9] C. Cotar, G. Friesecke, and C. Klüppelberg. Density functional theory and optimal transportation with Coulomb cost. Preprint available at arXiv:1104.0603.

[10] A. Galichon and N. Ghoussoub. Variational representations for N-cyclically monotone vector fields. Preprint.

[11] W. Gangbo. Habilitation thesis, Universite de Metz, available at http://people.math.gatech.edu/~gangbo/publications/habilitation.pdf, 1995.

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

[13] W. Gangbo and A. Święch. Optimal maps for the multidimensional monge-kantorovich problem. *Comm. Pure Appl. Math.*, 51(1):23–45, 1998.

[14] N. Ghoussoub and B Maurey. Remarks on multi-marginal symmetric monge-kantorovich problems. to appear in *Discrete and Continuous Dynamical Systems-A, special issue on "Optimal Transport and Applications "* (2012).

[15] N. Ghoussoub and A. Moameni. Symmetric monge-kantorovich problems and polar decompositions of vector fields. Preprint, (February 10, 2013) 23pp.

[16] N. Ghoussoub and B. Pass. Decoupling of degiorgi-type systems via multi-marginal optimal transport. Preprint, (February 13, 2013) 9pp.

[17] V. Levin. Abstract cyclical monotonicity and Monge solutions for the general Monge-Kantorovich problem. *Set-Valued Analysis*, 7(1):7–32, 1999.

[18] R. McCann. Polar factorization of maps on Riemannian manifolds. *Geom. Funct. Anal.*, 11:589–608, 2001.
[19] B. Pass. Multi-marginal optimal transport and multi-agent matching problems: uniqueness and structure of solutions. Preprint. Currently available at arXiv:1210.7372.

[20] B. Pass. Uniqueness and monge solutions in the multimarginal optimal transportation problem. *SIAM Journal on Mathematical Analysis*, 43(6):2758–2775, 2011.

[21] B. Pass. On the local structure of optimal measures in the multi-marginal optimal transportation problem. *Calculus of Variations and Partial Differential Equations*, 43:529–536, 2012. 10.1007/s00526-011-0421-z.

[22] B. Pass. Optimal transportation with infinitely many marginals. *Journal of Functional Analysis*, 2012. 10.1016/j.jfa.2012.12.002.