HÖLDER CONTINUITY FOR OPTIMAL MULTIVALUED MAPPINGS*

R. J. MCCANN† AND M. SOSIO‡

Abstract. Gangbo and McCann showed that optimal transportation between hypersurfaces generally leads to multivalued optimal maps – bivalent when the target surface is strictly convex. In this paper we quantify Hölder continuity of the bivalent map optimizing average distance squared between arbitrary measures supported on Euclidean spheres.

1. Introduction. Let X, Y be two measure spaces, μ, ν two probability measures defined on X and Y, respectively, and c a measurable map from $X \times Y$ to $[0, +\infty]$. Let us denote with $\Gamma = \Gamma(\mu, \nu)$ the set of all the probability measures on $X \times Y$ that have marginals μ and ν. More explicitly, $\gamma \in \Gamma(\mu, \nu)$ if and only if γ is a nonnegative measure satisfying $\gamma(A \times Y) = \mu(A)$, $\gamma(X \times B) = \nu(B)$, for all measurable subsets A of X and B of Y. The minimization problem

$$\inf_{\gamma \in \Gamma(\mu,\nu)} \int_{X \times Y} c(x, y) d\gamma(x, y) \tag{1.1}$$

is known as *Kantorovich's optimal transportation problem*; c is called the *cost function*, and every probability measures in $\Gamma(\mu, \nu)$ is called a *transference plan*. Kantorovich's problem is meant to investigate how a certain mass μ distributed on a domain X is transported to another location (described by ν and Y) at a minimal cost (see [32] for an exhaustive description).

When $X = Y = \mathbf{R}^n$, and the cost function is the Euclidean squared distance, the minimizers of (1.1) are characterized by the existence of a convex function $\psi : \mathbf{R}^n \to \mathbf{R} \cup \{+\infty\}$, whose subdifferential $\partial\psi \subset \mathbf{R}^n \times R^n$ contains the support of every optimal transference plan $\gamma \in \Gamma(\mu, \nu)$(see Brenier [3] for references). This convex function is called *Brenier's potential.* When μ is absolutely continuous with respect to the Hausdorff measure of dimension n, then ψ is differentiable on a set of full μ–measure, and the optimizer γ is unique, and it full mass lies on the graph $\{(x, \nabla\psi(x)) \mid x \in \mathrm{dom}\nabla\psi\}$ of the gradient of ψ. Then μ–a.e. point x must be mapped to the unique destination $y = \nabla\psi$ for transportation to be efficient. Therefore the optimal transference plan is the push-forward of μ by $\mathrm{Id} \times \nabla\psi$, denoted

$$\gamma = (\mathrm{Id} \times \nabla\psi)_\sharp \mu.$$

The measurable map $T = \nabla\psi$ is called an *optimal map.*

Several authors treated the regularity of optimal maps when the cost function is the Euclidean squared distance; among them Caffarelli [5] [6] [7] [8] [9] [10], Delanoë [11], and Urbas [31]. In particular, Caffarelli showed that if the domain Y is convex, $d\mu = f dVol, d\nu = g dVol$, where $dVol$ denotes the Lebesgue measure, and the densities $f, 1/g$ are bounded, then the optimal map is Hölder continuous.

---

* [MS] gratefully acknowledges the University of Toronto for its generous hospitality during various stages of this work, which has been carried out to complete her PhD program at the University of Pavia. [RJM]'s research was supported in part by NSERC grant 217006-08.

†Department of Mathematics, University of Toronto, 40 St. George Street, Toronto, Ontario, Canada, M5S 2E4. e-mail: mccann@math.utoronto.ca

‡Dipartimento di Matematica "F. Casorati", Università degli Studi di Pavia, via Ferrata 1, 27100, Pavia, Italy. e-mail: maria.sosio@unipv.it

In some applications of Optimal Transportation to Physics or Economics, other cost functions are of interest. For example, the problem of the reflector antenna (see Wang [33] and Oliker and Waltman [26]) has been shown to be equivalent to optimal transportation of measures on the Euclidean unit sphere with respect to the cost function $-\log|x-y|$ [34] [16]. Inspired by these works on the reflector antenna, Ma, Trudinger, and Wang found a condition on the cost function, which implies the regularity of the optimal map [25]. It is a structural condition depending upon derivatives up to the order four of the cost function. Following their notation, we name it (A3). It will be stated in Section 5.

Loeper [24], Kim, and McCann [19] [20] clarified the role of (A3) when an optimal map exists and is unique. More precisely, when the cost function is sufficiently smooth and (A3) holds, under suitable convexity hypotheses on the domains, and the absolute continuity of the Lebesgue measure with respect to ν, Loeper was able to prove the Hölder continuity of the optimal map (see Section 5 for a precise statement of the hypothesis). On the other hand, Kim and McCann [20] found a covariant expression of (A3), named (A3s) in their paper, and extended Loeper's results to transportation problems set on a pair of smooth manifolds.

Our paper makes use of Loeper, Kim, and McCann's argument to improve the regularity results obtained by Gangbo and McCann [14] for a transportation problem between boundaries of convex sets, with the Euclidean squared distance cost. Optimal transportation between boundaries of convex sets does not generally lead to a single-valued optimal map, but rather to multivalued mappings. This means that an optimizer $\gamma \in \Gamma(\mu,\nu)$ takes the form

$$\gamma = \sum_{i=1}^{m} \gamma_i, \qquad \gamma_i = (\mathrm{Id} \times t_i)_\sharp \mu_i,$$

where t_i are measurable maps from X to Y, and $\mu = \sum_{i=1}^{m} \mu_i$. This is the case of the Kantorovich problem analysed by Gangbo and McCann, who found a bivalent mapping. The novelty of our paper is the quantification of the continuity in this setting of multi-valued mappings.

Let Ω and Λ be two bounded, strongly convex (in the sense of Section 2), open sets in $\mathbf{R}^{n+1}$, with Borel probability measures μ on $\partial\Omega$ and ν on $\partial\Lambda$. We consider the Monge-Kantorovich problem

$$\inf_{\gamma \in \Gamma(\mu,\nu)} \int_{\mathbf{R}^{n+1} \times \mathbf{R}^{n+1}} |x-y|^2 d\gamma(x,y). \tag{1.2}$$

When μ is absolutely continuous with respect to the Hausdorff measure of dimension n, $(\mathcal{H}^n)$, and Ω is strictly convex, the optimal transference plan is unique, but its support fail to concentrate on the graph of a single map (see Theorem 2.6 of [14]). Gangbo and McCann [14] showed that the unique optimizer $\gamma \in \Gamma(\mu,\nu)$ is supported by two maps, named t^+ and t^-, i.e.

$$\gamma = \gamma_1 + \gamma_2, \quad \gamma_1 = (\mathrm{id} \times t^+)_\sharp \mu_1, \quad \gamma_2 = (\mathrm{id} \times t^-)_\sharp \mu_2,$$

where $\mu = \mu_1 + \mu_2$. This means that the mass at a point $x \in \partial\Omega$ does not always have a unique destination on $\partial\Lambda$, but can be split into two different destinations, $t^+(x)$ and $t^-(x)$, which correspond to the two limits $\nabla\psi(x_k)$ obtained as $x_k \to x$ from outside or inside Ω, respectively. Indeed, while Brenier's potential ψ is tangentially differentiable at $\mathcal{H}^n$–a.e. boundary point $x \in \partial\Omega$, the normal differentiability might fail. This implies that the subdifferential $\partial\psi$ consists of a segments with endpoints $t^+(x)$ and $t^-(x)$ on $\partial\Lambda$ (see Lemma 1.6 of [14]).

Gangbo and McCann proved that t^+ is a homeomorphism between $\partial\Omega$ and $\partial\Lambda$. Moreover, they conjectured Hölder regularity for t^+ on $\partial\Omega \setminus S_0$, where

$$S_0 := \{x \in \partial\Omega \mid n_\Omega(x) \cdot n_\Lambda(t^+(x)) = 0\}$$

represents a part of the "boundary" between the region where the mass splits and the region where it does not. More precisely, if S_2 denotes the region where the mass splits (bivalent region), then S_0 contains those limit points of S_2 at which the split images degenerate to a single image. In the present work, we will prove a slight modification of their conjecture, i.e. that t^+ is locally Hölder continuous on S_2 and on $S_1 = \partial\Omega \setminus (S_0 \cup S_2)$.

The peculiarity of (1.2) is the "hybrid" setting given by combining the choice of the Euclidean squared distance cost with a transportation problem set on embedded hypersurfaces. One of the difficulties we encountered has been to combine the convexity notion deriving from the Euclidean cost with the dimension and the pseudo-Riemannian structure of the manifolds where the measures are supported. Since the Hausdorff dimension of sptμ and sptν is n rather than $n+1$, we are not able to adapt Caffarelli's regularity theory to our problem; (see however [13]). Nevertheless Gangbo and McCann's conjecture about Hölder continuity is reinforced by examples of Ma-Trudinger-Wang [25] and Example 2.4 of Kim-McCann [19]: the authors showed that the Euclidean squared distance cost, in the settings of (1.2), satisfies (A3) on

$$N := \{(x, y) \in \partial\Omega \times \partial\Lambda \mid n_\Omega(x) \cdot n_\Lambda(y) > 0\}.$$

Despite this comforting result, the regularity of t^+ is not immediate. Loeper's results needs to be adapted to our hybrid setting. Moreover, the target measure with respect to t^+, ν_1, which is the portion of mass "transferred" by t^+, does not inherit the hypothesis on ν of having a positive lower bound on its density with respect to the Lebesgue surface measure. This means there are regions in $\partial\Omega$ where the Lebesgue surface measure is not absolutely continuous with respect to ν_1, so one of the required hypotheses of Loeper's argument is not satisfied. We will treat these regions separately with a different argument.

Our paper is organized as follows. In Section 2 we report the main result of Gangbo and McCann's paper [14]; we also discuss the most important statement of this paper and the strategy we are going to adopt to prove it. We will restrict our argument to the case of spherical domains, $\partial\Omega = \partial\Lambda = \mathbf{S}^n$, though we expect that our regularity result could be extended to more general uniformly convex domains with more work. In Section 3 we introduce and clarify some notation. In Section 4 we comment on some questions related to our problem. In Section 5 we adapt Loeper's theory to our transportation problem, restricting his argument to the subset of $\partial\Omega$ where the necessary hypothesis on the measures holds. The regularity result on the remaining subset of $\partial\Omega$ is then derived in Section 6 .

The authors wish to thank Aldo Pratelli and Giuseppe Savaré for a careful reading of the manuscript and helpful suggestions.

2. Preliminaries, strategy, and results. We recall the following definitions from [14]. For a smooth convex domain Ω, *strong* convexity asserts the existence of a positive lower bound for all principal curvatures of $\partial\Omega$.

DEFINITION 2.1. *A pair of Borel measures μ on $\partial\Omega$, ν on $\partial\Lambda$ is said to be* suitable *if* *(i) there exists $\epsilon > 0$ such that $\mu < \frac{1}{\epsilon}\mathcal{H}^n\lfloor_{\partial\Omega}$ and $\nu > \epsilon\mathcal{H}^n\lfloor_{\partial\Lambda}$, and*

(ii) Ω is strongly convex.

If the above hypotheses are satisfied also when the roles of $\mu \leftrightarrow \nu$ and $\Omega \leftrightarrow \Lambda$ are interchanged, we say that the pair (μ, ν) is symmetrically suitable. Under these assumptions on the measures, Gangbo and McCann were able to prove the following optimality results.

THEOREM 2.2. *Fix bounded, strictly convex domains $\Omega, \Lambda \in \mathbf{R}^{n+1}$ with suitable measures μ on $\partial\Omega$ and ν on $\partial\Lambda$. Then the infimum of* (1.2) *is uniquely attained.* Let $N_\Omega(x)$ denote the set of all outward unit normals to $\partial\Omega$ at x. When $N_\Omega(x)$ contains only one element, we denote that unit vector by $n_\Omega(x)$.

PROPOSITION 2.3. *Fix bounded, strictly convex domains $\Omega, \Lambda \in \mathbf{R}^{n+1}$ with suitable measures μ on $\partial\Omega$ and ν on $\partial\Lambda$. Let ψ be the Brenier convex potential. For each $x \in \partial\Omega$ exactly one of the following statements holds:*

(o) $\partial\psi(x) = \{y_1\}$ with $n \cdot q_1 = 0$ for some pair $n \in N_\Omega(x), q_1 \in N_\Lambda(y_1)$;

(i) $\partial\psi(x) = \{y_1\}$ with $n \cdot q_1 > 0$ for all pairs $n \in N_\Omega(x), q_1 \in N_\Lambda(y_1)$;

(ii) $\partial\psi(x) = [y_1, y_2]$, in which case $\partial\Omega$ is differentiable at x and $n_\Omega(x) \cdot q_1 > 0, n_\Omega(x) \cdot q_2 < 0$ for all $q_i \in N_\Lambda(y_i)$, $i = 1, 2$.

DEFINITION 2.4. *Given $\Omega, \Lambda, (\mu, \nu)$, and ψ as in Proposition 2.3, we decompose $\partial\Omega = S_0 \cup S_1 \cup S_2$ into three disjoint sets such that (o) holds for $x \in S_0$, (i) holds for $x \in S_1$, (ii) holds for $x \in S_2$. Moreover we use the extreme images $y_1, y_2 \in \partial\psi$ of the proposition to define an outer map $t^+ : \partial\Omega \to \partial\Lambda$, and an inner map $t^- : S_2 \to \partial\Lambda$ by $t^+(x) = y_1$, and $t^-(x) = y_2$. It is convenient to extend the definition of t^- to $\partial\Omega$ by setting $t^-(x) = t^+(x)$ for $x \in S_0 \cup S_1$, so that $\partial\psi(x) = [t^+(x), t^-(x)]$.*

THEOREM 2.5. *Fix bounded, strictly convex domains $\Omega, \Lambda \in \mathbf{R}^{n+1}$ with symmetrically suitable measures μ on $\partial\Omega$ and ν on $\partial\Lambda$. Then the minimizer γ can be expressed by*

$$\gamma = \gamma_1 + \gamma_2, \quad \gamma_1 = (\mathrm{id} \times t^+)_\sharp \mu_1, \quad \gamma_2 = (\mathrm{id} \times t^-)_\sharp \mu_2,$$

where $\mu_1 := (t^+)^{-1}_\sharp \nu_1$, $\mu_2 := \mu - \mu_1$, and $\nu_1 := \nu\lfloor_{T_2^c}$, with $T_2^c := \partial\Lambda \setminus t^-(S_2)$. Whenever $x \in S_2$, $t^+(x) - t^-(x) \neq 0$ is an outward normal for $\partial\Omega$ at x. Moreover $t^+ : \partial\Omega \to \partial\Lambda$ and $t^-\lfloor_{\bar{S}_2}: \bar{S}_2 \to \bar{T}_2$ are homeomorphisms.

The partition $\partial\Omega = S_0 \cup S_1 \cup S_2$ will play an important role in our paper, so it is essential to understand the meaning of these sets. The mass lying on $S_0 \cup S_1$ is transferred without splitting to a target set on $\partial\Lambda$ by t^+, while the mass lying on S_2 splits into two destinations, which are described by t^+ and t^-. For this reason we will call S_0 the *degenerate set*, S_1 the *non-degenerate univalent set*, and S_2 the *bivalent set*. When the measures (μ, ν) are symmetrically suitable, an analogous decomposition of $\partial\Lambda = T_0 \cup T_1 \cup T_2$ can be introduced (see Definition 3.6 of [14]). In particular T_2 is the bivalent set for the Kantorovich transportation problem (1.2), where (Ω, ν) and (Λ, ν) are exchanged, with (Λ, ν) playing the role of the source.

Our aim is to prove that the map $t^+ : \partial\Omega \longrightarrow \partial\Lambda$ is Hölder continuous on S_1 and S_2. The second author has been able to show that t^+ satisfies bi-Lipschitz estimates when $n = 1$ [29], via an argument relying on the results of Ahmad [1], which cannot be extended to higher dimensions. Here we are developing a different strategy which works for all $n > 1$, when $\partial\Omega, \partial\Lambda = \mathbf{S}^n$. We will proceed in two steps. First we will show that t^+ is Hölder continuous on the preimage $(t^+)^{-1}(T_1) \subset \mathbf{S}^n$ of the set T_1 where

$$\nu_1 > \epsilon\mathcal{H}^n\lfloor_{\partial\Lambda},$$

where ϵ is a constant satisfying Definition 2.1. This lower bound on ν_1 allows us to adapt the argument used by Kim and McCann in [20]. On $(t^+)^{-1}(T_2)$, where the lower bound fails, the regularity of t^+ will be derived from the Hölder continuity of t^+ on S_2. In the end we will be able to obtain the following result.

THEOREM 2.6 (Hölder continuity of multivalued maps outside the degenerate set). *If (μ, ν) are symmetrically suitable measures on $(\mathbf{S}^n, \mathbf{S}^n)$, $n > 1$, then*

$$t^+ \in C_{loc}^{\frac{1}{4n-1}}(S_1) \qquad \textit{and} \qquad t^+ \in C_{loc}^{\frac{1}{4n-1}}(S_2).$$

3. Notation. The notation we are going to use is similar to that of [14] and [19], in particular we refer to Example 2.4 of [19], with $\partial\Omega = \partial\Lambda = \mathbf{S}^n$, $c : \mathbf{S}^n \times \mathbf{S}^n \to \mathbf{R}$, $c(x, y) = |x - y|^2$, $N := \{(x, y) \in \mathbf{S}^n \times \mathbf{S}^n \mid n_{\mathbf{S}^n}(x) \cdot n_{\mathbf{S}^n}(y) > 0\}$, and $\hat{N}(x) := \{y \in \mathbf{S}^n \mid (x, y) \in N\}$. We will always use the variable x for points on the source domain $\partial\Omega = \mathbf{S}^n$, and the variable y for points on the target domain $\partial\Lambda = \mathbf{S}^n$.

Let us recall the usual system of local coordinates for the points of $\mathbf{S}^n$

$$\varphi_i : \mathbf{S}^n \cap \{x \in \mathbf{S}^n | x_i > 0\} \to \mathbf{R}^n, \quad \varphi_i(x) = (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n).$$

Analogously, given $x \in \mathbf{S}^n$ and $y \in \hat{N}(x)$ we can consider a system π_x of local coordinates projecting on the hyperplane perpendicular to x. In this way both x and y can be represented in local coordinates by means of the same map π_x

$$x \xrightarrow{\pi_x} X, \qquad y \xrightarrow{\pi_x} Y,$$

where the capital letters stand for the image of the projection. To simplify the notation, given a function $F : \mathbf{R}^{n+1} \to \mathbf{R}$ and a projection π_{x_0}, whenever $x \in \hat{N}(x_0)$ we will write $F(X)$ to denote $F(\pi_{x_0}^{-1}(X)) = F(x)$. We will therefore write $\psi(X)$, $c(X, Y)$ instead of $\psi(\pi_{x_0}^{-1}(X))$, $c((\pi_{x_0}^{-1}(X), (\pi_{x_0}^{-1}(Y))$. For example, given $x \in \mathbf{S}^n$ and $y \in N(x)$, by mean of π_x we can write

$$c(X, Y) = |X - Y|^2 + (\sqrt{1 - |X|^2} - \sqrt{1 - |Y|^2})^2.$$

In local coordinates, we use the notation $Dc = (\frac{\partial c}{\partial X_1}, \ldots, \frac{\partial c}{\partial X_n}$ and $\bar{D}c = (\frac{\partial c}{\partial Y_1}, \ldots, \frac{\partial c}{\partial Y_n})$ to denote the partial derivatives. The cross partial derivatives $\bar{D}Dc$ at $(x, y) \in N$ define an unambiguous linear map from vectors at y to covectors at x.

Hereafter $d\mathcal{H}^n$ denotes the Hausdorff measure of dimension n, $\mathcal{N}_\rho(B)$ represents the ρ-neighbourhood of a set B, and $[Y_0, Y_1]$ indicate the Euclidean segment whose extreme points are Y_0 and Y_1.

In Section 6 we will use the expression "angle between two vectors z_1 and $z_2 \in \mathbf{R}^{n+1}$" to refer to $\arccos \frac{z_1 \cdot z_2}{|z_1||z_2|}$.

4. Some related questions.

4.1. Relation between the convex potential ψ and the mappings t^+, t^-. Let ψ be the Brenier potential associated to (1.2). It is well known that the subdifferential $\partial\psi$ includes the support $\mathrm{spt}\gamma \subset \mathbf{R}^{n+1} \times \mathbf{R}^{n+1}$ of all minimizers $\gamma \in \Gamma(\mu, \nu)$ for (1.2)(see [3][4] for references). Under the hypothesis of Theorem 2.5, there exists a unique optimizer $\gamma \in \Gamma(\mu, \nu)$ for (1.2), and there exist two continuous maps $t^\pm : \partial\Omega \to \partial\Lambda$, such that

$$\begin{aligned}\{(x, t^+(x))\}_{x \in \mathrm{spt}\mu} \quad \subset \mathrm{spt}\gamma \subset \quad &\{(x, t^+(x))\}_{x \in \partial\Omega} \cup \{(x, t^-(x)))\}_{x \in S_2} \\ &(= \partial\psi \cap (\partial\Omega \times \partial\Lambda)).\end{aligned}$$

So, what is the relation between the optimal mappings t^+, t^-, and the convex potential ψ? Can we derive any regularity for ψ from Theorem 2.6? Gangbo and McCann answered to the first question in Lemma 1.6 of [14]. Indeed the maps t^+ and t^- correspond to the outer and inner trace of $\nabla\psi$, respectively. So we can write the subdifferential of ψ in terms of the optimal mappings: $\partial\psi(x) = [t^+(x), t^-(x)]$ at any boundary point $x \in \partial\Omega$. Moreover, in Corollary 4.4 of [14], Gangbo and McCann proved that, when Ω is bounded andstrongly convex, Λ is bounded and strictly convex, and (μ, ν) are suitable measures on $\partial\Omega, \partial\Lambda$, then ψ is tangentially differentiable along $\partial\Omega$. This answers the second question. From Theorem 2.6 it follows immediately that

$$\psi \in C^{1,\frac{1}{4n-1}}_{loc} \text{ on } S_1 \subset \mathbf{S}^n,$$

i.e. on the non-degenerate univalent set, where $\partial\psi(x) = \{\nabla\psi(x)\} = \{t^+(x)\}$. Notice that the conclusion of Theorem 2.6 does not imply $\psi \in C^{1,\frac{1}{4n-1}}_{loc}$ on S_2, since ψ is not differentiable in the normal direction to the sphere on S_2. Nevertheless, choosing the coordinates of Lemma A.1 of [14], $\frac{\partial\psi}{\partial x_1}$ exists for $i = 2, 3, \dots, n+1$, and

$$\frac{\partial\psi}{\partial x_i}(x) = t^+(x)_i = t^-(x)_i, \text{ for } i = 2, 3, \dots, n+1, \text{ and } x \in \mathbf{S}^n.$$

We conclude that the restriction of ψ to $\mathbf{S}^n$ has a derivative which is Hölder continuous locally on S_1 and S_2.

4.2. The regularity of t^+ on S_0. We do not presently have any regularity result for t^+ on the degenerate set S_0, except continuity from [14], On the contrary, we will see in the statements of Theorem 5.1 and Theorem 6.1 that, on $\mathbf{S}^n \setminus S_0$, close to S_0 the Hölder constant of t^+ provided by our proof may become very big. Moreover, as noticed in Example 2.4 of [19], the nondegeneracy hypothesis (A2) fails on S_0. Therefore, we cannot apply Loeper's argument on S_0. On the other hand we believe the set S_0 to be small. In dimension $n = 1$, with Ω and Λ bounded strictly convex planar domains, Ahmad [1] proved that S_0 consists of at most two points.

4.3. Extending the results to more general domains. Theorem 2.6 can be extended to the problem of transporting a measure on a given Euclidean sphere to a measure on any other Euclidean sphere, possibly with a different centre and radius. Indeed, identities (9) and (10) of [14] indicate how to reduce this more general problem to the case treated in this paper.

Thanks to the results in Example 2.4 of [19], Theorem 5.1 can be extended to the transportation problem where the measures (μ, ν) are supported on $(\partial\Omega, \partial\Lambda)$, with $\Omega, \Lambda \subset \mathbf{R}^{n+1}$ bounded convex domains with C^2-smooth boundaries. We expect that the same extension is possible for Theorem 6.1, but cannot presently provide a proof. Our argument relies crucially on Lemma 6.9, whose proof exploits the peculiar geometric properties of $\mathbf{S}^n$, and cannot be easily extended to more general convex domains.

4.4. Nearly constant measures on $\mathbf{S}^n$. J. Kitagawa and M. Warren [21] proved that when the measures μ, ν are nearly constant on $\mathbf{S}^n$ (in C^1 topology), then the optimizer $\gamma \in \Gamma(\mu, \nu)$ is supported on the graph of a single map.

4.5. Sharp Hölder exponent. The Hölder exponent in Theorem 2.6 is not sharp. It is the same exponent provided by Loeper's argument [24], i.e. $1/(4n-1)$, where n is the dimension of the sphere where μ and ν are supported. Recently, Liu [23] improved Loeper's Hölder exponent to the sharp exponent $1/(2n-1)$.

5. t^+ is Hölder continuous on $(t^+)^{-1}(T_1) \subset \mathbf{S}^n$. In this section we are going to adapt Kim-McCann's version of Loeper's argument (Appendices B,C and D of [20]) to our mapping t^+, which satisfies $(t^+)_\sharp \mu_1 = \nu_1$. Thus, let us recall the regularity conditions (A0),(A1), (A2), and (A3s) from [19] [25] on a cost function $c : \mathbf{S}^n \times \mathbf{S}^n \to \mathbf{R}$

(A0)(Smoothness) *$c \in C^4(N)$, where N has been define in Section 3.*
(A1)(Twist condition) *$c \in C^1(N)$ and for all $x \in \mathbf{S}^n$ the map $y \to -Dc(x, y)$ from $\hat{N}(x) \subset \partial\Lambda$ to $T^*_x(\mathbf{S}^n)$ is injective.*
(A2)(Non-degeneracy) *$c \in C^2(N)$ and for all $(x, y) \in N$ the linear map $\bar{D}Dc : T_y\mathbf{S}^n \to T^*_x\mathbf{S}^n$ is bijective.*
(A3s)(Strictly regular costs) *$c \in C^4(N)$ satisfies (A2) and for every $(x, y) \in N$*

$$sec_{(x,y)}(p \oplus 0) \wedge (0 \oplus \bar{p}) \geq 0 \text{ for all null vectors } p \oplus \bar{p} \in T_{(x,y)}N, \tag{5.1}$$

and equality in (5.1) implies $p = 0$ or $\bar{p} = 0$.

The notation "sec" refers to the sectional curvature of a two–plane. We define it by means of the Riemann curvature tensor $R_{i'j'k'l'}$ induced by the symmetric bilinear form

$$h = \frac{1}{2}\begin{pmatrix} 0 & -\bar{D}Dc \\ -D\bar{D}c & 0 \end{pmatrix} \tag{5.2}$$

on N. If $c \in C^4(N)$, the *sectional curvature* of a two–plane $P \wedge Q$ at $(x, y) \in N$ is given by

$$\sec_{(x,y)} P \wedge Q = \sum_{i'=1}^{2n}\sum_{j'=1}^{2n}\sum_{k'=1}^{2n}\sum_{l'=1}^{2n} R_{i'j'k'l'} P^{i'} Q^{j'} P^{k'} Q^{l'}.$$

We recall also some notions of convexity from Definition 2.5 of [19]. Though we are assuming $\partial\Omega = \partial\Lambda = \mathbf{S}^n$, the following definition holds for more general convex domains.

DEFINITION 5.1. *A subset $W \subseteq N \subseteq \partial\Omega \times \partial\Lambda$ is* geodesically convex *if each pair of points in W is linked by a curve satisfying the geodesic equation on (N, h). We say that $B \subset \partial\Lambda$* appears convex *from $x \in \partial\Omega$ if $\{x\} \times B$ is geodesically convex and $B \subset \hat{N}(x)$. We say $W \subseteq \partial\Omega \times \partial\Lambda$ is* vertically convex *if $\hat{W}(x) := \{y \in \partial\Lambda \mid (x, y) \in W\}$ appears convex from x for each $x \in \partial\Omega$. We say that $A \subset \partial\Omega$* appears convex *from $y \in \partial\Lambda$ if $A \times \{y\}$ is geodesically convex and $A \subset N(y)$. We say $W \subseteq \partial\Omega \times \partial\Lambda$ is* horizontally convex *if $W(y) := \{x \in \partial\Omega \mid (x, y) \in W\}$ appears convex from y for each $y \in \partial\Lambda$. If W is both vertically and horizontally convex, we say it is* bi–convex. The regularity result that we are going to exploit is Theorem D.1 of [20]. We now state in a reductive form, referring to our particular settings, to avoid the introduction of new unnecessary notations.

THEOREM 5.2 (**Simplified version of Theorem D.1 of [20]**). *Assume $c \in C^4(M)$ satisfies (A1),(A2), and (A3s) on the closure of M, where $M \subset \mathbf{S}^n \times \mathbf{S}^\mathbf{n}$ is a bounded domain bi–convex with respect to (5.2). Fix $m > 0$, and let $\rho, \bar{\rho}$ be probability measures on $\mathbf{S}^n$ with Lebesgue densities $d\bar{\rho}/d\mathrm{vol} \geq m$ throughout $\mathbf{S}^n$ and $d\rho/d\mathrm{vol} \in L^\infty(\mathbf{S}^n)$. Then there exists a map $F \in C^{1/\max\{5,4n-1\}}_{loc}(\mathbf{S}^n, \mathbf{S}^n)$ between ρ and $\bar{\rho}$ which is optimal with respect to the transportation cost c.*

Assuming (μ, ν) to be suitable measures on $(\mathbf{S}^n, \mathbf{S}^n)$, in order to apply Kim–McCann's argument we need ν_1 to satisfy

$$\text{there exists } \epsilon_1 \text{ such that } \nu_1 > \epsilon_1 \mathcal{H}^n\lfloor_{\partial\Lambda}. \tag{5.3}$$

From the definition of ν_1 in Theorem 2.2 we see that ν_1 satisfies (5.3) only outside the bivalent set $T_2 \in \partial\Lambda = \mathbf{S}^n$, i.e. outside the set where the image of t^+ is bivalent. This is the reason why we can state a regularity result only on a portion of the source domain, $(t^+)^{-1}(T_1) \subset \mathbf{S}^n$. Hereafter we will assume $n > 1$.

THEOREM 5.1. *Suppose (μ, ν) are symmetrically suitable measures on $(\mathbf{S}^n, \mathbf{S}^n)$ (in particular, from Definition 2.1, there exists $\epsilon > 0$ such that $\nu > \epsilon \mathcal{H}^n\lfloor_{\mathbf{S}^n}$). Then t^+ is locally Hölder continuous on $(t^+)^{-1}(T_1)$, with Hölder exponent at least $\frac{1}{4n-1}$. The local Hölder constant depends on ϵ, n, and tends to infinity when one approaches the boundary of N.*

REMARK 5.2. *Computations that show the explicit dependence of the Hölder constant on the distance of the boundary of N can be found in [29].*

LEMMA 5.3. *The set*

$$N = \{(x, y) \in \mathbf{S}^n \times \mathbf{S}^n \mid n_{\mathbf{S}^n}(x) \cdot n_{\mathbf{S}^n}(y) > 0\}$$

is bi–convex in the sense of Definition 2.5 of [19]. *Proof:* Fix $x_0 \in \mathbf{S}^n$. $\hat{N}(x_0)$ appears convex from x_0 if and only if $Dc(x_0, \hat{N}(x_0))$ is convex in $T^*_{x_0}(\mathbf{S}^n)$ (see Lemma 4.4 of [19]). Suppose

$$Dc(x_0, y_0), Dc(x_0, y_1) \in Dc(x, \hat{N}(x)),$$

where $y_0, y_1 \in \hat{N}(x)$. We are going to show that for every $\theta \in (0, 1)$

$$\theta Dc(x_0, y_1) + (1 - \theta) Dc(x_0, y_0) \in Dc(x_0, \hat{N}(x)). \tag{5.4}$$

Let us consider a system of local coordinates. Given $x_0 \in \mathbf{S}^n$ we project x_0 and $y \in \hat{N}(x_0)$ to the hyperplane perpendicular to $\hat{n}_\Omega(x_0)$ and containing the origin (notice that this choice of local coordinates is well defined since $\hat{n}_\Omega(x_0) \cdot \hat{n}_\Lambda(y_k) > 0$, when $y_k \in \hat{N}(x_0)$, $k = 0, 1$)

$$x_0 \xrightarrow{\pi_{x_0}} 0, \quad y \xrightarrow{\pi_{x_0}} Y \tag{5.5}$$

so that, in local coordinates,

$$x_0 = (0, 1), \quad y = (Y, \sqrt{1 - |Y|^2})$$
$$c(X, Y) = |X - Y|^2 + (\sqrt{1 - |X|^2} - \sqrt{1 - |Y|^2})^2.$$

We easily get

$$\frac{\partial c}{\partial X_i}(0, Y) = -2Y_i.$$

If $v \in T_x(\partial\Omega)$ and v_i are its coordinate with respect to the basis $\frac{\partial}{\partial X_i}$, we can write

$$Dc(v)(x_0, y) = v(c)(x_0, y) = \sum_{i=1}^n v_i \frac{\partial c}{\partial X_i}(0, Y).$$

Hence we can compute

$$\begin{aligned}
\theta Dc(v)(x_0, y_1) + (1-\theta)Dc(v)(x_0, y_0) &= \theta v(c)(x_0, y_1) + (1-\theta)v(c)(x_0, y_0) \\
&= \sum_{i=1}^{n} \left[\theta v_i \frac{\partial c}{\partial X_i}(0, Y_1) + (1-\theta) v_i \frac{\partial c}{\partial X_i}(0, Y_0) \right] \\
&= \sum_{i=1}^{n} 2v_i \left[\theta\left(-Y_{1,i}\right) + (1-\theta)\left(-Y_{0,i}\right) \right] \\
&= \sum_{i=1}^{n} -2v_i(\theta Y_{1,i} + (1-\theta)Y_{0,i}). \qquad (5.6)
\end{aligned}$$

Therefore, for all $\theta \in (0,1)$

$$\theta Dc(x_0, y_1) + (1-\theta)Dc(x_0, y_0) = Dc(x_0, \pi_{x_0}^{-1}(\theta Y_1 + (1-\theta)Y_0) \in Dc(x_0, \hat{N}(x_0)).$$

Since x_0 is an arbitrary point of $\mathbf{S}^n$, we conclude that N is vertically convex. By a similar argument, it is easy to show that N is also horizontally convex. We conclude that N is bi-convex. □

Proof of Theorem 5.1: Fix $(x, y) = (x, t^+(x)) \in N$, with $t^+(x) \in T_1$. Since T_1 is open, and t^+ is continuous, we can choose R and then r small enough that $B_r(y) \subset t^+(B_R(x)) \subset T_1$; as asserted by Trudinger and Wang in [30], since N is bi-convex, taking R and r even smaller, $P = B_R(x) \times B_r(y) \subset N$ is bi-convex (alternatively, we could show directly that P is bi-convex, by means of the same argument used for N in Lemma 5.3). We replace ν_1 with its restriction ν_1' to $B_r(y)$ and we denote $\mu_1' = s^+_{\#}(\nu_1')$. Up to further decreasing R and r, we get us local coordinates over both domains simultaneously (for example through the chart π_x). Let $X = \pi_x(x), Y = \pi_x(y)$, and $P' = \pi_x(B_R(x)) \times \pi_x(B_r(y))$. Since P is bi-convex and the notion of bi-convexity is coordinate invariant (as manifest from Definition 2.5 of [19]), P' is bi-convex with respect to the cost

$$c(X, Y) = |X - Y|^2 + (\sqrt{1 - |X|^2} - \sqrt{1 - |Y|^2})^2, \qquad (5.7)$$

which satisfies (A0). Kim and McCann showed that the cost in the original coordinates satisfies also condition (A2) and (A3s) (see Example 2.4 of [19]), and that the quantities in these conditions have an intrinsic meaning independent of coordinates, since they are geometric quantities (i.e. pseudo-Riemannian curvatures in the case of (A3s) and non-degeneracy of the metric in the case (A2)). This implies that also the cost (5.7) satisfies (A2) and (A3s). Only the constant C_0' of (A3s) will depend on the coordinates. Since we know that the equation $D_X c(X, Y) = D\psi(X)$ has at most two solutions, $Y^+ = t^+(X)$ and $Y^- = t^-(X)$ and only Y^+ lies in P', the cost satisfies (A1) on P'.

At this point we can apply Theorem D.1. of [20] to the cost (5.7) on P', with probability measures μ_1^x and ν_I^x, on $\pi_x(B_R(x))$ and $\pi_x(B_r(y))$ respectively, defined by

$$\mu_1^x := (\pi_x)_\sharp \mu_1', \qquad \nu_1^x := (\pi_x)_\sharp \nu_1'.$$

The source μ_1^x is supported (and bounded above) in $\pi_x(B_R(x))$ and target ν_1^x supported (and bounded below) in $\pi_x(B_r(y))$, We deduce the existence of a locally Hölder continuous optimal map pushing μ_1' forward to ν_1'. By the uniqueness of optimal transport, this map must coincide μ_1'-a.e. with t^+. Since both maps are continuous they agree on the (closed) support of μ_1'. Since sptμ_1' contains a small ball around x, this shows t^+ is locally Hölder at x. □

6. t^+ is locally Hölder continuous where its image is bivalent. In the previous section we established local Hölder continuity for the outer map $t^+ = (s^+)^{-1}$ on the source domain $s^+(T_1) \subset \mathbf{S}^n$, but not on $s^+(T_0 \cup T_2) = S_0 \cup s^+(T_2)$. Our strategy for extending this estimate to $s^+(T_2)$ is described at the end of this paragraph. First note, however, that Gangbo and McCann's *Sole Supplier Lemma*, 2.5 of [14], implies the outer image of the bivalent source is contained in the univalent target $t^+(S_2) \subset T_1$, and similarly $s^+(T_2) \subset S_1$. Since $s^+ : \mathbf{S}^n \longrightarrow \mathbf{S}^n$ is a homeomorphism, from $S_1 \cup S_2 = s^+(T_1) \cup s^+(T_2)$, it follows that the bivalent source $S_2 \subset s^+(T_1)$ belongs to the domain where Hölder continuity of t^+ has already been shown. On this bivalent set S_2, the inner map t^- is related to the outer map $t^+(x) = t^-(x) + \lambda(x)x$ by the geometry of the target. In Proposition 6.2, this relation will be used to deduce (i) Hölder continuity of t^- from that of t^+. This quantifies injectivity (ii) of the inverse map $s^- = (t^-)^{-1}$ (through a bi-Hölder estimate in Proposition 6.6), whose relation to the outer map $s^+(y) = s^-(y) + \omega(y)y$ is then used in Proposition 6.10 to quantify injectivity (iii) of $s^+ = (t^+)^{-1}$ on the bivalent target $T_2 = t^-(S_2)$. This yields the desired local Hölder continuity of t^+ on the source set $s^+(T_2)$ mentioned at the outset.

Let us recall the geometric characterization of t^+ and t^- from Proposition 2.3 and Definition 2.4. Remembering that, on $\mathbf{S}^n$, $n_{\mathbf{S}^n}(x) = x$, we have

- If $x \in S_0$ then $x \cdot t^+(x) = 0$.
- If $x \in S_1$ then $x \cdot t^+(x) > 0$.
- If $x \in S_2$ then $x \cdot t^+(x) > 0$ and $x \cdot t^-(x) < 0$.

We are going to introduce a geometric approach, based on the previous characterization, which allows us to prove the following theorem. Hereafter we will assume $n > 1$.

Theorem 6.1. *If (μ, ν) are symmetrically suitable measures on $(\mathbf{S}^n, \mathbf{S}^n)$, then t^+ is locally Hölder continuous on $(t^+)^{-1}(T_2)$.*

From Lemma 1.6 of [14] we know that t^+ and t^- are related by

$$\forall x \in S_2 \subset \mathbf{S}^n \qquad t^+(x) - t^-(x) = \lambda(x)x,$$

where λ is a continuous positive function on S_2. Given x_0, x_1 in S_2 we then have

$$|t^-(x_1) - t^-(x_0)| \leq |t^+(x_1) - t^+(x_0) - \lambda(x_1)x_1 + \lambda(x_0)x_0|. \tag{6.1}$$

We would like to exploit the regularity of t^+ on $S_2 \subset (t^+)^{-1}(T_1)$, proved in the previous section, to prove that also t^- is Hölder continuous on S_2. For this purpose we also need to estimate the term $\lambda(x_1)x_1 + \lambda(x_0)x_0$. This will be done applying the Mean Value Theorem to a suitable function and utilizing the geometric properties of the target.

Proposition 6.2 (Hölder continuity of t^-).

If $t^+ \in C^\alpha_{loc}(S_2)$ then $t^- \in C^\alpha_{loc}(S_2)$. Let $U \subset S_2$ and $0 < k_U := \min\{-x \cdot t^+(x) \mid x \in U\}$. If C_U^+ bounds the Hölder constant for t^+ on U, then

$$C_U^- := \left(1 + \frac{1}{k_U}\right)(C_U^+ + 2)$$

is the Hölder constant for t^- on U.

Proof: The function $h(y) := d(y, \mathbf{S}^n) = 1 - |y|$ is differentiable on $\Lambda = B_1(0)$ except at $y = 0$. Notice that $h(t^-(x)) = h(t^+(x) - \lambda(x)x) = h(t^+(x)) = 0$ whenever $x \in S_2$. Consider a neighbourhood $U \subset S_2$ and the corresponding k_U, C_U^+ from the statement of Proposition 6.2. Let $x_0, x_1 \in U$, $|x_1 - x_0| < 2$ (we need ∇h to be well defined on the line segment between $t^-(x_0)$ and $t^-(x_1)$, i.e. $0 \notin [t^-(x_0, t^-(x_1)])$. Applying the Mean Value Theorem, we get

$$\begin{aligned} 0 &= h(t^+(x_1) - \lambda(x_1)x_1) - h(t^+(x_0) - \lambda(x_0)x_0) \\ &= \nabla h(u) \cdot (t^+(x_1) - t^+(x_0) - \lambda(x_1)x_1 + \lambda(x_0)x_0), \end{aligned}$$

for some u on the line segment between $t^-(x_0)$ and $t^-(x_1)$. It follows

$$(\lambda(x_1)x_1 - \lambda(x_0)x_0) \cdot \nabla h(u) = (t^+(x_1) - t^+(x_0)) \cdot \nabla h(u). \tag{6.2}$$

We can rewrite (6.2) as

$$\begin{aligned} (t^+(x_1) - t^+(x_0)) \cdot \nabla h(u) + \lambda(x_0)(x_0 - x_1) \cdot \nabla h(u) \\ = (\lambda(x_1) - \lambda(x_0))x_1 \cdot \nabla h(u); \end{aligned}$$

then, using $|\nabla h(u)| = 1$,

$$\begin{aligned} &|\lambda(x_1) - \lambda(x_0)||x_1 \cdot \nabla h(u)| \\ &\leq |(t^+(x_1) - t^+(x_0))| + \lambda(x_0)|x_0 - x_1|. \end{aligned} \tag{6.3}$$

We now state a claim, whose demonstration is postponed to the end of this proof.

Lemma 6.3. *Under the hypotheses of Proposition 6.2, fix $\epsilon \in (0,1)$, such that $\epsilon^2 < \frac{k_U}{2}$. Since t^- is uniformly continuous on $\bar{S}_2$, there exists δ_ϵ, depending on the data through ψ, such that*

$$|x_1 - x_0| < \delta_\epsilon \Rightarrow |t^-(x_1) - t^-(x_0)| < \epsilon.$$

Then, taking x_0, x_1 such that $|x_1 - x_0| < \delta_\epsilon$, we have

$$x_i \cdot \nabla h(u) > \frac{k_U}{2} > 0 \qquad \textit{for } i = 1, 2.$$

Recalling that $\lambda(x) \leq 2$, since $\partial\Omega = \mathbf{S}^n$, by means of Lemma 6.3 we simplify (6.3) to

$$\begin{aligned} &|\lambda(x_1) - \lambda(x_0)| \\ &\leq \frac{2}{k_U} \left[|t^+(x_1) - t^+(x_0)| + \lambda(x_0)|x_0 - x_1|\right] \\ &\leq \frac{2}{k_U}[|t^+(x_1) - t^+(x_0)| + 2|x_1 - x_0|]. \end{aligned} \tag{6.4}$$

Therefore, by (6.1) and (6.4),

$$\begin{aligned} &|t^-(x_1) - t^-(x_0)| \\ &\leq |t^+(x_1) - t^+(x_0)| + \lambda(x_1)|x_1 - x_0| + |\lambda(x_1) - \lambda(x_0)| \\ &\leq |t^+(x_1) - t^+(x_0)| + 2|x_1 - x_0| + |\lambda(x_1) - \lambda(x_0)| \\ &\leq \left(1 + \frac{2}{k_U}\right) |t^+(x_1) - t^+(x_0)| + 2\left(1 + \frac{2}{k_U}\right)|x_1 - x_0|. \end{aligned} \tag{6.5}$$

Combining (6.5) and $t^+ \in C^\alpha(U)$, we conclude

$$\begin{aligned} &|t^-(x_1) - t^-(x_0)| \\ &\leq C_U^+ \left(1 + \frac{2}{k_U}\right) |x_1 - x_0|^\alpha + 2\left(1 + \frac{2}{k_U}\right)|x_1 - x_0|, \end{aligned} \tag{6.6}$$

i.e. t^- is Hölder continuous on S_2 whenever $|x_1 - x_0| < \delta_\epsilon$, with $\epsilon^2 < \frac{k_U}{2}$. We can take $\delta_\epsilon < 1$, so that (6.6) implies

$$|t^-(x_1) - t^-(x_0)| \leq \left(1 + \frac{2}{k_U}\right)\left[C_U^+ + 2\right]|x_1 - x_0|^\alpha$$
$$= C_U^-|x_1 - x_0|^\alpha. \qquad \square$$

Proof of Lemma 6.3: Let $z_i = t^-(x_i), i = 1, 2$. Notice that $\nabla h(u) = -\frac{u}{|u|}$. We have $u = sz_1 + (1-s)z_0$ for some $s \in (0,1)$. Hence, there exists $\xi \in (0, \epsilon)$ such that

$$\begin{aligned} x_1 \cdot u &< -k_U s + (1-s) x_1 \cdot z_0 \\ &= -k_U s + (1-s) x_1 \cdot (z_1 + \xi(z_0 - z_1)) \\ &< -k_U + (1-s)\xi\epsilon < -k_U + \epsilon^2. \end{aligned}$$

Using a similar argument for $x_0 \cdot u$, we conclude that if $\epsilon^2 < \frac{k_U}{2}$ then $x_i \cdot \nabla h(u) > \frac{k_U}{2|u|} > \frac{k_U}{2} > 0$, for $i = 1, 2$. $\square$

Remark 6.4. *Proposition 6.2 admits a converse, i.e. if $t^- \in C^\alpha_{loc}(S_2)$ then $t^+ \in C^\alpha_{loc}(S_2)$. This can be proved with minor changes in the preceding argument.*

Remark 6.5. *By means of Theorem 5.1 and Proposition 6.2, t^- is indeed locally Hölder continuous on S_2 with exponent $\frac{1}{4n-1}$.*

The injectivity (ii) of the inverse map $s^- = (t^-)^{-1}$ on T_2, is an immediate consequence of the local Hölder continuity of t^- on S_2, and it has been included in the following proposition.

Proposition 6.6 (Quantifying injectivity of s^-). *Let $V \subset T_2$. Under the hypotheses of Theorem 6.1 $s^- := (t^-)^{-1}$ satisfies*

$$\forall y_0, y_1 \in V \text{ sufficiently close}, |s^-(y_1) - s^-(y_0)| \geq \hat{C}_V^-|y_1 - y_0|^{4n-1},$$

where

$$\hat{C}_V^- = (C_U^-)^{-1},$$

with $U = s^-(V)$ and $0 < k_V := \min\{-y \cdot s^-(y) \mid y \in V\}$. *Proof:* Since $s^- := (t^-)^{-1}$ is uniformly continuous on $\bar{T}_2$, given $\delta_\epsilon > 0$ there exists $\gamma_{\delta_\epsilon} > 0$ such that, if $|y_1 - y_0| < \gamma_{\delta_\epsilon}$, then $|s^-(y_1) - s^-(y_0)| < \delta_\epsilon$. Supposing $|y_1 - y_0| < \gamma_{\delta_\epsilon}$, we can apply Proposition 6.2 to $x_1 = s^-(y_1), x_0 = s^-(x_0)$ to get

$$|s^-(y_1) - s^-(y_0)| \geq \frac{1}{C_U^-}|y_1 - y_0|^{4n-1}. \qquad \square$$

We now state an elementary Lemma about vectors in $\mathbf{R}^n$.

Lemma 6.7. *Let $u, v \in \mathbf{R}^n$. Suppose the angle between u and v is less than $\frac{\pi}{2} + \alpha$, with $\alpha \in \left[0, \frac{\pi}{2}\right)$. Then $|u + v| \geq |u| \cos\alpha$.* *Proof:* Let $\theta_{u,v}$ denote the angle between u and v. Keeping $|u|$ and $|v|$ fixed, $|u+v|$ can be seen as a function of $\theta_{u,v}$ by mean of

$$|u + v|^2 (\theta_{u,v}) = |u|^2 + |v|^2 + 2|u||v|\cos\theta_{u,v},$$

When $\theta_{u,v} \in \left[0, \frac{\pi}{2} + \alpha\right]$, the function $|u+v|(\theta_{u,v})$ reaches its minimum at $\theta_{u,v} = \frac{\pi}{2} + \alpha$. To our purpose we can take $\theta_{u,v} = \frac{\pi}{2} + \alpha$. For simplicity we assume v parallel to $e_1 \in \mathbf{R}^n$. Let us consider the projection p on the hyperplane perpendicular to e_1 and containing the origin. Then $p(u+v) = p(u) = |u|\cos\alpha$. Since $|p(u+v)| \le |u+v|$, we have the thesis. $\square$

This Lemma turns out to be the key to the proof of step (iii). Under the hypothesis of symmetrically suitable measures, the optimal transportation problem we are studying is symmetric, hence every result that holds for t^+ on $\mathbf{S}^n$ implies an analogous result for s^+ on $\mathbf{S}^n$. In particular, from Lemma 1.6 of [14], for every $y \in T_2$ we can write

$$s^+(y) - s^-(y) = \omega(y)y, \tag{6.7}$$

where ω is a nonnegative function on T_2. Hence

$$|s^+(y_1) - s^+(y_0)| = |s^-(y_1) - s^-(y_0) + \omega(y_1)y_1 - \omega(y_0)y_0|.$$

If we were allowed to apply Lemma 6.7 to the right hand side of the previous equality, with $u = s^-(y_1) - s^-(y_0)$ and $v = \omega(y_1)y_1 - \omega(y_0)y_0$, we would then be able to exploit the regularity of s^- to prove step (iii). Therefore, we need to understand the behaviour of the angle between $s^-(y_1) - s^-(y_0)$ and $\omega(y_1)y_1 - \omega(y_0)y_0$, when y_0 gets close to y_1. From the monotonicity of $\partial\psi$ we have

$$(s^-(y_1) - s^-(y_0)) \cdot (y_1 - y_0) \ge 0 \qquad \forall y_1, y_0 \in T_2,$$

which says that the angle between $s^-(y_1) - s^-(y_0)$ and $y_1 - y_0$ is in $\left[0, \frac{\pi}{2}\right]$. If we can show that the angle between $y_1 - y_0$ and $\omega(y_1)y_1 - \omega(y_0)y_0$ is in $[0, \alpha]$, for a certain $\alpha \in \left[0, \frac{\pi}{2}\right)$, then we can apply Lemma 6.7 to get the desired estimate on $|s^+(y_1) - s^+(y_0)|$.

LEMMA 6.8. *Given $y_0, y_1 \in T_2$ we denote with $\beta(y_0, y_1)$ the angle between $y_1 - y_0$ and $\omega(y_1)y_1 - \omega(y_0)y_0$. If the angle between y_0 and y_1 is equal to γ then*

$$\beta(y_0, y_1) \in \left[0, \frac{\pi - \gamma}{2}\right). \tag{6.8}$$

Proof: The angle between y_1 and $-y_0$ is equal to $\pi - \gamma$, while the angle between y_1 (or $-y_0$) and $y_1 - y_0$ is $\frac{\pi-\gamma}{2}$. Since $\omega(y_0), \omega(y_1) > 0$, $\beta(y_0, y_1) \in \left[0, \frac{\pi-\gamma}{2}\right)$. $\square$

LEMMA 6.9 (Dichotomy). *Fix $y_1 \in T_2$. For every integer $m > 1$ define*

$$\Theta_m(y_1) := \left\{ y \in T_2 \mid \beta(y, y_1) \in \left[\frac{\pi}{2} - \frac{1}{m}, \frac{\pi}{2}\right] \right\}.$$

Unless $\Theta_m(y_1)$ is empty for m sufficiently large, there exist $M > 0$ and $K > 0$ such that

$$|s^+(y_1) - s^+(y)| \ge K|y_1 - y|, \qquad \forall y \in \Theta_m(y_1), \text{ with } m > M. \tag{6.9}$$

Proof: We are interested in the sets $\Theta_m(y_1)$ for m large, so hereafter we assume $m > 50$. Define

$$0 < \varpi_m := \inf\{\omega(y) > 0 \mid y \in \Theta_m(y_1)\}$$

and note $\varpi_m \leq \varpi_{m+1}$ since $\Theta_m(y_1) \supset \Theta_{m+1}(y_1)$. By elementary computations, we have

$$
\begin{aligned}
&|\omega(y_1)y_1 - \omega(y)y| \cos\beta(y, y_1) = \frac{(\omega(y_1)y_1 - \omega(y)y)\cdot(y_1 - y)}{|y_1 - y|} \\
&= \frac{\omega(y_1)y_1 \cdot (y_1 - y) - \omega(y)y \cdot (y_1 - y)}{|y_1 - y|} \\
&\geq \frac{\varpi_m y_1 \cdot (y_1 - y) - \omega(y)y \cdot (y_1 - y)}{|y_1 - y|} \\
&= \frac{\varpi_m |y_1 - y|^2 + (\varpi_m - \omega(y))y \cdot (y_1 - y)}{|y_1 - y|} \\
&\geq \varpi_m |y_1 - y| \qquad\qquad \forall y \in \Theta_m(y_1),
\end{aligned}
\tag{6.10}
$$

where we used the definition of ϖ_m and the trivial inequality $y \cdot y_1 \leq 1$ to show that the term $(\varpi_m - \omega(y))y \cdot (y_1 - y)$ is non-negative. Consider now the two vectors $\omega(y_1)y_1 - \omega(y)y$ and $(\omega(y_1) - \omega(y))y_1$, with $y \in T_2$. Their difference is parallel to $y_1 - y$, so they have the same projection on any hyperplane perpendicular to $y_1 - y$. This projection has length $|\omega(y_1)y_1 - \omega(y)y| \sin\beta(y, y_1)$. Therefore

$$
|\omega(y_1)y_1 - \omega(y)y| \sin\beta(y, y_1) \leq |\omega(y_1) - \omega(y)| \qquad \forall y \in T_2. \tag{6.11}
$$

Putting together (6.10) and (6.11), we obtain an estimate for $\tan\left(\frac{\pi}{2} - \frac{1}{m}\right)$

$$
\tan\left(\frac{\pi}{2} - \frac{1}{m}\right) \leq \tan\beta(y, y_1) \leq \frac{|\omega(y_1) - \omega(y)|}{\varpi_m |y_1 - y|} \qquad \forall y \in \Theta_m(y_1).
$$

As $m \to +\infty$, $\tan\left(\frac{\pi}{2} - \frac{1}{m}\right) \to +\infty$; then for every $N > 0$ there exists $m_N > 50$ such that

$$
|\omega(y_1) - \omega(y)| > N\varpi_m |y_1 - y|, \qquad \forall y \in \Theta_m(y_1), m > m_N. \tag{6.12}
$$

From (6.7) we have, for every $y \in T_2$,

$$
s^+(y_1) - s^+(y) - \omega(y)(y_1 - y) = s^-(y_1) - s^-(y) + (\omega(y_1) - \omega(y))y_1.
$$

We define

$$
\begin{aligned}
A &:= |s^+(y_1) - s^+(y)| + |\omega(y)(y_1 - y)| \\
&\geq |s^-(y_1) - s^-(y) + (\omega(y_1) - \omega(y))y_1|, \qquad y \in T_2.
\end{aligned}
\tag{6.13}
$$

Using $|v - u| \geq |v| - |u| \quad \forall u, v \in \mathbf{R}^{n+1}$, we get two different estimates for A

$$
A \geq |s^-(y_1) - s^-(y)| - |\omega(y_1) - \omega(y)|, \tag{6.14}
$$

$$
A \geq |\omega(y_1) - \omega(y)| - |s^-(y_1) - s^-(y)|. \tag{6.15}
$$

By the symmetry of the problem, using (6.4), we have

$$
|\omega(y_1) - \omega(y)| \leq \frac{2}{k'_m}\left[|s^+(y_1) - s^+(y)| + 2|y_1 - y|\right] \qquad \forall y \in \Theta_m(y_1),
$$

where $0 < k'_m := \inf\{-y \cdot s^-(y) \mid y \in \Theta_m(y_1)\} \le k'_{m+1}$. From (6.14) it follows

$$A \ge |s^-(y_1) - s^-(y)| - \frac{2}{k'_m}\left[|s^+(y_1) - s^+(y)| + 2|y_1 - y|\right]. \tag{6.16}$$

On the other hand, combining (6.12) and (6.15)

$$A \ge N\varpi_m|y_1 - y| - |s^-(y_1) - s^-(y)|, \qquad \forall y \in \Theta_m(y_1), m > m_N. \tag{6.17}$$

We can sum (6.16) and (6.17) to get

$$2A \ge N\varpi_m|y_1 - y| - \frac{2}{k'_m}\left[|s^+(y_1) - s^+(y)| + 2|y_1 - y|\right].$$

From the definition (6.13) of A, this becomes

$$2\left(1 + \frac{1}{k'_m}\right)|s^+(y_1) - s^+(y)| \ge \left(N\varpi_m - \frac{4}{k'_m} - 2\omega(y)\right)|y_1 - y|,$$

for every $y \in \Theta_m(y_1), m > m_N$. Since neither ϖ_m nor k'_m is decreasing as a function of m, taking N large enough ensures $N > \left(\frac{4}{k'_{m_N}} + 4\right)\frac{1}{\varpi_{m_N}}$ to yield a positive constant

$$K = \frac{N\varpi_{m_N} - 4(\frac{1}{k'_{m_N}} + 1)}{2\left(1 + \frac{1}{k'_{m_N}}\right)},$$

such that

$$|s^+(y_1) - s^+(y)| \ge K|y_1 - y|, \qquad \forall y \in \Theta_m(y_1), m > m_N.$$

To conclude we take $M = m_N$. $\square$

The injectivity (iii) of $s^+ = (t^+)^{-1}$ on the bivalent target $T_2 = t^-(S_2)$. follows from Lemma 6.7 and Lemma 6.9.

Proposition 6.10 (Quantifying injectivity of s^+ on the bivalent target). *Let $y_1 \in V \subset T_2$. Under the hypotheses of Theorem 6.1, there exists $\hat{C}^+_V > 0$, depending on $\hat{C}^-_V$, k_V (from Proposition 6.6), and $\bar{\theta}(y_1)$ (from Lemma 6.9), such that, when y_0 is sufficiently close to y_1,*

$$|s^+(y_1) - s^+(y_0)| \ge \hat{C}^+_V|y_1 - y_0|^{4n-1}.$$

Proof: When $y_0 \in \Theta(y_1, \theta)$, with $\theta > \bar{\theta}(y_1)$ we apply Lemma 6.9 and we are done. Otherwise the angle between $s^-(y_1) - s^-(y_0)$ and $\omega(y_1)y_1 - \omega(y_0)y_0$ is smaller than $\frac{\pi}{2} + \bar{\theta}(y_1)$. Applying Lemma 6.7, we obtain

$$\begin{aligned}&|s^+(y_1) - s^+(y_0)|\\ &= |s^-(y_1) - s^-(y_0) + \omega(y_1)y_1 - \omega(y_0)y_0|\\ &\ge |s^-(y_1) - s^-(y_0)|\cos\bar{\theta}(y_1).\end{aligned}$$

Taking y_0, y_1 sufficiently close ($|y_1 - y_0| < \gamma_{\delta_\epsilon}$, from the proof of Proposition 6.6), Proposition 6.6 implies

$$|s^+(y_1) - s^+(y_0)| \ge \cos\bar{\theta}(y_1)\hat{C}^-_V|y_1 - y_0|^{4n-1}. \quad \square$$

Proof of Theorem 6.1: Define $y_i := t^+(x_i) \in V \subset T_2$. If $y_0 \in \Theta(y_1, \theta)$, with $\theta > \bar{\theta}(y_1)$, we have

$$|y_1 - y_0| < K|x_1 - x_0|.$$

Otherwise, by the uniform continuity of t^+, taking x_0 sufficiently close to x_1, we have $|t^+(x_1) - t^+(x_0)| < \gamma_{\delta_\epsilon}$ and we can apply Proposition 6.10 to $y_i = t^+(x_i)$, $i = 1, 0$ to conclude

$$|y_1 - y_0| < \frac{1}{\hat{C}_V^+}|x_1 - x_0|^{\frac{1}{4n-1}}. \qquad \square$$

REFERENCES

[1] N. Ahmad, *The Geometry of Shape Recognition via the Monge-Kantorovich Optimal Problem*, PhD thesis, Brown University, 2004.

[2] A. D. Aleksandrov, *Existence and uniqueness of a convex surface with a given integral curvature*, C. R. (Doklady) Acad. Sci. URSS (N.S.), 35, (1942), pp. 131–134.

[3] Y. Brenier, *Décomposition polaire and réarrangement monotone des champs de vecteurs*, C.R. Acad. Sci. Paris Sér. I Math., 305, (1987), pp. 805–808.

[4] Y. Brenier, *Polar factorization and monotone rearrangement of vector-valued functions*, Comm. Pure Appl. Math., 44, (1991), pp. 375–417.

[5] L. Caffarelli, *Interior $W^{2,p}$ estimates for solutions of the Monge-Ampère equation*, Ann. of Math., 131, (1990), pp. 135–150.

[6] L. Caffarelli, *A localization property of viscosity solutions to the Monge-Ampère equation and their strict convexity*, Ann. of Math., 131, (1990), pp. 129–134.

[7] L. Caffarelli, *Some regularitiy properties of solutions of Monge-Ampère equation*, Comm. Pure Appl. Math., 44, (1991), pp. 965–969.

[8] L. Caffarelli, *Boundary regularity of maps with convex potentials*, Comm. Pure Appl. Math., 45, (1992), pp. 1141–1151.

[9] L. Caffarelli, *The regularity of mappings with a convex potential*, J. Amer. Math. Soc., 5,(1992), pp. 99–104.

[10] L. Caffarelli, *Boundary regularity of maps with convex potentials, II,* Ann. of Math., 144, (1996), pp. 453–496.

[11] P. Delanoë, *Classical solvability in dimension two of the second boundary-value problem associated with the Monge-Ampère operator*, Ann. Inst. H. Poincarè Anal. Non Linèaire, 8, (1991), pp. 443–457.

[12] L. C. Evans, R. F. Gariepy, *Measure Theory and Fine Properties of Functions*, CRC Press, 1992.

[13] A. Figalli, Y.-H. Kim, R. J. McCann, *Continuity and injectivity of optimal maps for non-negatively cross-curved costs*, preprint at http://www.math.toronto.edu/mccann/papers/C1NNCC.pdf.

[14] W. Gangbo and R. J. McCann, *Shape recognition via Wasserstein distance*, Quart. Appl. Math, 58,(2000), pp. 705–737.

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

[16] T. Glimm and V. I. Oliker, *Optical design of single reflector systems and the Monge-Kantorovich mass transfer problem*, J. of Math. Sciences, 117, (2003), pp. 4096–4108.

[17] C. E. Gutiérrez, *The Monge-Ampère Equation*, Birkhäuser, 2001.

[18] H. G. Kellerer, *Duality theorems for marginal problems*, Z. Wahrsch. Verw. Gebiete, 67, (1984), pp. 399–432.

[19] Y.-H. Kim, R. J. McCann, *Continuity, curvature, and the general covariance of optimal transportation*, J. Eur. Math. Soc., 12, (2010), pp. 1009–1040.

[20] Y.-H. Kim, R. J. McCann, *Appendices to original version of Continuity, curvature, and the general covariance of optimal transportation*, Preprint at arXiv:math/0712.3077v1.

[21] J. Kitagawa, M. Warren, *Regularity of optimal transport with Euclidean distance squared cost on the embdedded sphere*, in preparation, http://www.math.princeton.edu/ mww/papers/kw.pdf.

[22] M. Knott and C. S. Smith, *On Hoeffding-Fréchet bounds and cyclic monotone relations*, J. Multivariate Anal., 40, (1992), pp. 328–334.

[23] J. Liu, *Hölder regularity of optimal mappings in optimal transportation*, Calc. of Var. and Part. Diff. Eq., 34, (2009), pp. 435–451.
[24] G. Loeper, *On the regularity of solutions of optimal transportation problems*, Acta Math., 202, (2009), pp. 241–283.
[25] X.-N. Ma, N. Trudinger and X.-J. Wang, *Regularity of potential functions of the optimal transportation problem*, Arch. Rational Mech. Anal., 177, (2005), pp. 151–183.
[26] V. I. Oliker, P. Waltman, *Radially symmetric solutions of a Monge-Ampère equation arising in a reflector mapping problem*, Proc. Int. conf. on Differential Equation and Mathematical Pysics, Birningham, Ala., (1986), pp. 361–374.
[27] R. J. McCann, *Existence and uniqueness of monotone measure-preserving maps*, Duke Mathematical Journal, 80, (1995), pp. 309–323.
[28] R. T. Rockafellar, *Characterization of the subdifferentials of convex functions*, Pacific J. Math., 17, (1966), pp. 497–510.
[29] M. Sosio, *Regularity for Transport and Nonlinear Diffusion Problems*, PhD Thesis, Università degli Studi di Pavia, 2011.
[30] N. Trudinger and X.-J. Wang, *On the second boundary value problem for Monge-Ampere type equations and optimal transportation*, Ann. Scuola Norm. Sup. Pisa, 8, (2009), pp. 1–32.
[31] J. Urbas, *On the second boundary value problem for equations of Monge-Ampère type*, J. Reine Angew. Math., 487, (1997), pp. 115–124.
[32] C. Villani, *Topics in Optimal Transportation*, American Mathematical Society, 2003.
[33] X.-J. Wang, *On the design of a reflector antenna*, Inverse Problems, 12, (1996), pp. 351–375.
[34] X.-J. Wang, *On the design of a reflector antenna II*, Calc. Var., 20, (2004), pp. 329–341.