EXISTENCE OF OPTIMAL MAPS IN THE REFLECTOR-TYPE PROBLEMS

WILFRID GANGBO[1] AND VLADIMIR OLIKER[2]

Abstract. In this paper, we consider probability measures μ and ν on a d-dimensional sphere in $\mathbf{R}^{d+1}, d \geq 1$, and cost functions of the form $c(\mathbf{x}, \mathbf{y}) = l(\frac{|\mathbf{x}-\mathbf{y}|^2}{2})$ that generalize those arising in geometric optics where $l(t) = -\log t$. We prove that if μ and ν vanish on $(d-1)$-rectifiable sets, if $|l'(t)| > 0$, $\lim_{t \to 0^+} l(t) = +\infty$, and $g(t) := t(2-t)(l'(t))^2$ is monotone then there exists a unique optimal map T_o that transports μ onto ν, where optimality is measured against c. Furthermore, $\inf_{\mathbf{x}} |T_o\mathbf{x} - \mathbf{x}| > 0$. Our approach is based on direct variational arguments. In the special case when $l(t) = -\log t$, existence of optimal maps on the sphere was obtained earlier in [8] and [22] under more restrictive assumptions. In these studies, it was assumed that either μ and ν are absolutely continuous with respect to the d-dimensional Haussdorff measure, or they have disjoint supports. Another aspect of interest in this work is that it is in contrast with the work in [7] where it is proved that when $l(t) = t$ then existence of an optimal map fails when μ and ν are supported by Jordan surfaces.

Mathematics Subject Classification. 49, 35J65.

1. INTRODUCTION

In Euclidean space $\mathbf{R}^{d+1}$ consider a reflector system consisting of a point source $\mathcal{O}$ radiating with intensity $I(\mathbf{x})$ in directions $\mathbf{x} \in \mathbf{X}$, where $\mathbf{X}$ is a closed set on a d-dimensional unit sphere $\mathbf{S^d} \hookrightarrow \mathbf{R}^{d+1}$ centered at $\mathcal{O}$, and a smooth perfectly reflecting hypersurface R, star-shaped relative to $\mathcal{O}$, which intercepts and reflects the light rays with directions from $\mathbf{X}$; see Figure 1. Assuming the geometric optics approximation and applying the classical reflection law to determine the set of reflected directions $\mathbf{Y} \subset \mathbf{S^d}$ (after one reflection), we obtain an associated with R "reflector map" $\xi : \mathbf{X} \to \mathbf{Y}$. Assuming that no energy is lost in the process, one can apply the energy conservation law to calculate the intensity distribution $L(\mathbf{y})$ produced on $\mathbf{Y}$. The reflector problem consists in solving the inverse problem in which the source $\mathcal{O}$, the sets $\mathbf{X}, \mathbf{Y}$ and the intensities I and L are given in advance and the reflector R needs to be determined. That is, R should be such that $\xi(\mathbf{X}) \supseteq \mathbf{Y}$ and

$$L(\xi(\mathbf{x}))|J(\xi)(\mathbf{x})| = I(\mathbf{x})$$

for all $\mathbf{x}$ in the interior of $\mathbf{X}$; here $J(\xi)$ denotes the Jacobian determinant of ξ.

Problems of this type arise often in applications, for example, in design of reflector antennas [23]. In various forms the reflector problem has been considered by many authors and numerous papers by engineers (at least since early 60's [12] until now [22]) are devoted to this subject. It was introduced in electrical engineering and optics independently of the mass transport problem. Because of the strong nonlinear constraints appearing in the problem, progress has been slow and many theoretical and computational issues still remain open. The problem continues to attract considerable attention because of its practical importance and mathematical subtleties. It may be pointed out that a version of the reflector problem appears on the famous list of unsolved problems proposed by Yau [24] in 1993.

Analytically, the reflector problem considered in this paper can be formulated as a nonlinear second order elliptic partial differential equation of Monge-Ampère type on a subset of $\mathbf{S^d}$. In such form it has been studied by Oliker and Waltman [16], Caffarelli and Oliker [3], Wang [21], Guan and Wang [10], Caffarelli, Kochengin and Oliker [4], Oliker [17], and other authors.

Recently, Glimm and Oliker [8] and, independently, Wang [22] have shown that if the function I (resp. L) is treated as the density of a measure μ (resp. ν) that are absolutely continuous with respect to the volume measure on $\mathbf{S^d}$, then the reflector problem can be studied as a variational problem in the framework of Monge-Kantorovich theory, that is, a problem of finding an optimal map minimizing the transport cost of transferring μ onto ν with the cost function $-\log(1 - \mathbf{x} \cdot \mathbf{y})$. In contrast with other cost functions considered usually in the Monge-Kantorovich theory, this cost function may assume infinite values. Consequently, in order to overcome this difficulty, a geometric condition requiring the supports of μ and ν to be disjoint was imposed in [8] and [22] to establish existence and uniqueness of optimal maps. Without imposing the condition that the supports of the measures are disjoint, existence and uniqueness of optimal maps was also obtained in [8]. However, the proof is indirect as it relies on existence of weak solutions in the reflector problem established earlier in [3, 17].

The contribution of this study is twofold. First of all, we obtain existence and uniqueness of optimal maps T_o for a class of cost functions that may be infinite. This class includes the logarithmic cost function of the reflector problem as a special case. The cost functions are precisely of the form $c(\mathbf{x}, \mathbf{y}) = l(\frac{|\mathbf{x} - \mathbf{y}|^2}{2})$ where $|l'(t)| > 0$, $\lim_{t \to 0^+} l(t) = +\infty$, and $g(t) := t(2 - t)(l'(t))^2$ is monotone. Furthermore, we prove that $\inf_{\mathbf{x}} |T_o \mathbf{x} - \mathbf{x}| > 0$, which we view as an intermediary step in the study of the regularity of the map T_o. Secondly, our approach is variational and direct; the supports of measures μ and ν are allowed to overlap and it is merely required that these measures vanish on $(d - 1)$-rectifiable subsets. The precise statement can be found in Theorem 4.6.

Let us recall the main principles which ensure existence of optimal maps, which apparently, have been explicitly pointed out for the first time in [5] and later exploited by many authors. Assume we are given a cost function $c : \mathbf{R}^{d+1} \times \mathbf{R}^{d+1} \to \mathbf{R}$ and two probability measures μ and ν on $\mathbf{R}^{d+1}$, say, absolutely continuous with respect to Lebesgue measure. Existence and uniqueness of an optimal map transporting μ onto ν against c is ensured if $\nabla_{\mathbf{x}} c(\mathbf{x}, \cdot)$ is a one-to-one map of $\mathbf{R}^{d+1}$ into itself. Note that if $c(\mathbf{x}, \mathbf{y}) = |\mathbf{x} - \mathbf{y}|^2$ then $\nabla_{\mathbf{x}} c(\mathbf{x}, \cdot)$ has this property.

It is shown in [7] that if instead, μ and ν are supported by $\mathbf{S^d}$, then existence of an optimal map transporting μ onto ν against $c(\mathbf{x}, \mathbf{y}) = |\mathbf{x} - \mathbf{y}|^2$, fails. One of the points in our study here is that we start with a cost function $c(\mathbf{x}, \mathbf{y}) = l(|\mathbf{x} - \mathbf{y}|^2/2)$ such that $\nabla_{\mathbf{x}} c(\mathbf{x}, \cdot)$ fails to be a one-to-one map of $\mathbf{R}^{d+1}$ into itself. However, if $\mathbf{x} \in \mathbf{S^d}$ and $\mathbf{a} \in \mathbf{R}^{d+1}$ is distinct from the origin, we observed that there exists a unique $\mathbf{y} \in \mathbf{S^d}$ satisfying the equation $\nabla_{\mathbf{S^d}}^{\mathbf{x}} c(\mathbf{x}, \mathbf{y}) = \mathbf{a}$. Here, $\nabla_{\mathbf{S^d}}^{\mathbf{x}} c(\mathbf{x}, \cdot)$ is the tangential gradient of c on $\mathbf{S^d}$ with respect to $\mathbf{x}$. The injectivity of $\nabla_{\mathbf{S^d}}^{\mathbf{x}} c(\mathbf{x}, \cdot)$ yields existence of an optimal map that transports μ onto ν, if μ and ν are supported by $\mathbf{S^d}$ and vanish on $(d - 1)$-rectifiable sets. We use the Kantorovich theory as a tool to give a transparent explanation to the existence of a unique solution in the reflector problem, under sharp assumptions. Our main results are stated in Theorem 4.6. We refer the reader to a variant of the reflector problem involving two reflectors considered in a paper by Glimm and Oliker [9]. We also refer the reader to a recent study by Ahmad [2] in the plane, motivated by [7].

The paper is essentially self-contained. It is organized as follows. In order to motivate out subsequent considerations, we begin with a review of the reflector problem in Section 2. In Section 3 we review and extend
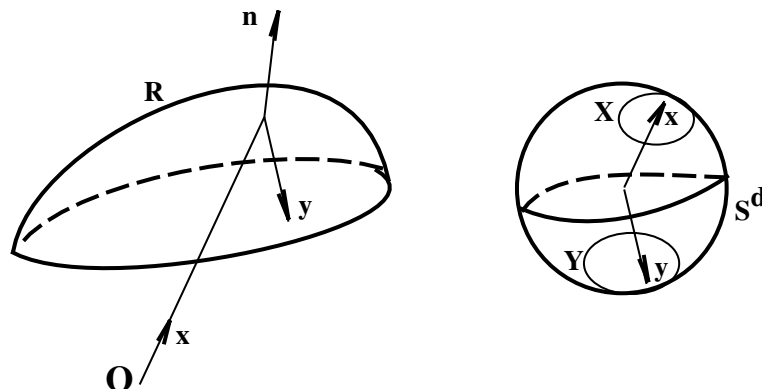
FIGURE 1. The reflector problem.

some results from the Monge-Kantorovich theory. Our main results establishing existence and uniqueness of optimal maps are in Section 4.

2. A REVIEW OF THE REFLECTOR PROBLEM

Let $\mathbf{X}, \mathbf{Y}, I, L$ and R be as in the introduction. If $\mathbf{n}$ is the unit normal field on R, then the incident direction $\mathbf{x}$ and the reflected direction $\mathbf{y}$ are related by the reflection law

$$\mathbf{y} = \mathbf{x} - 2(\mathbf{x} \cdot \mathbf{n})\mathbf{n}. \tag{1}$$

Thus, the hypersurface R defines the *reflector* map $\xi : \mathbf{x} \to \mathbf{y}$ which maps the "input aperture" $\mathbf{X} \subset \mathbf{S^d}$ onto the "output aperture" $\mathbf{Y} \subset \mathbf{S^d}$; see Figure 1. The intensity of the light reflected in direction $\mathbf{y} = \xi(\mathbf{x})$ is given by $I(\mathbf{x})/|J(\xi(\mathbf{x}))|$.

Suppose now that the closed sets $\mathbf{X}$ and $\mathbf{Y}$ on $\mathbf{S^d}$ are given as well as nonnegative integrable functions I on $\mathbf{X}$ and L on $\mathbf{Y}$ representing, respectively, the intensity of the source and the desired intensity on the far-region $\mathbf{Y}$. *The reflector problem is to determine a reflector R such that the map ξ defined by R maps $\mathbf{X}$ onto $\mathbf{Y}$ and satisfies the equation*

$$L(\mathbf{y}) = I(\xi^{-1}(\mathbf{y}))|J(\xi^{-1}(\mathbf{y}))|, \quad \mathbf{y} \in \mathrm{Int}\mathbf{Y}; \tag{2}$$

see [15, 16]. Note that this is an equation on the output aperture $\mathbf{Y}$ rather than on the input aperture $\mathbf{X}$. One could also set it up on $\mathbf{X}$ [14], but (2) is more convenient for our purposes here.

It was shown in [15, 16] that there exists a scalar quasi-potential $p : \mathbf{Y} \to (0, \infty)$ from which the reflector R can be recovered and in terms of which the equation (2) when $J(\xi^{-1}) \neq 0$ is the following equation of Monge-Ampère type

$$L(\mathbf{y}) = I(\xi^{-1}(\mathbf{y}))\frac{|\det[Hess(p) + (p - \rho)e]|}{\rho^n \det(e)}, \quad \mathbf{y} \in \mathrm{Int}\mathbf{Y}, \tag{3}$$

where e is the standard metric on $\mathbf{S^d}$, $\rho = (p^2 + |\nabla p|^2)/2p$, and $Hess(p)$, ∇p are computed in the metric e. In terms of p the position vector of R is given by

$$\mathbf{r}(\mathbf{y}) = -\nabla p(\mathbf{y}) - (p(\mathbf{y}) - \rho(\mathbf{y}))\mathbf{y} : \mathbf{Y} \to \mathbf{R}^{d+1}, \tag{4}$$

while

$$\xi^{-1}(\mathbf{y}) = \frac{\mathbf{r}(\mathbf{y})}{\rho(\mathbf{y})}.$$

Note that $|\mathbf{r}| = \rho$.

A close examination of (4) shows that it describes R as an envelope of a family of paraboloids of revolution $P(\mathbf{y})$ tangent to R, parametrized by their axes $\mathbf{y} \in \mathbf{Y}$ and with common focus $\mathcal{O}$. For each $\mathbf{y}$, $p(\mathbf{y})$ is the focal parameter of $P(\mathbf{y})$. This observation was used in [3] for the weak formulation of the reflector problem where a class of convex reflectors corresponding to positively elliptic solutions of (3) was introduced. Reflectors corresponding to negatively elliptic solutions of (3) can be introduced and analyzed by similar methods [8, 22]. Such reflectors, however, are only piecewise concave (relative to the origin $\mathcal{O}$). For brevity we discuss here only convex reflectors which we now define.

Let $\mathbf{Y}$ be a subset on $\mathbf{S^d}$ and $p' : \mathbf{Y} \to (0, \infty)$ a bounded function. Let $\{P(\mathbf{y})\}$ be a family of paraboloids of revolution with axes of direction $\mathbf{y} \in \mathbf{Y}$, common focus $\mathcal{O}$ and polar radii

$$\rho_{\mathbf{y}}(\mathbf{x}) = \frac{p'(\mathbf{y})}{1 - \mathbf{x} \cdot \mathbf{y}}, \quad \mathbf{x} \in \mathbf{S^d} \setminus \{\mathbf{y}\}. \tag{5}$$

The closed convex hypersurface R given by $\mathbf{r}(\mathbf{x}) = \rho(\mathbf{x})\mathbf{x}$, $\mathbf{x} \in \mathbf{S^d}$, with

$$\rho(\mathbf{x}) = \inf_{\mathbf{y} \in \mathbf{Y}} \rho_y(\mathbf{x}), \quad \mathbf{x} \in \mathbf{S^d}, \tag{6}$$

is called a **reflector defined by the family** $\{P(\mathbf{y})\}_{\mathbf{y} \in \mathbf{Y}}$ (with the light source at $\mathcal{O}$).

Let R be a reflector, z a point on R, P a paraboloid of revolution with focus at $\mathcal{O}$ and B the convex body bounded by P. If $R \subset B$ and $z \in P \bigcap R$ then P is called **supporting to R at** z.

For any $\mathbf{y} \in \mathbf{S^d}$ a reflector R has a supporting paraboloid with axis $\mathbf{y}$. The corresponding continuous function giving the focal parameters of all such supporting paraboloids $P(\mathbf{y})$, $\mathbf{y} \in \mathbf{S^d}$, of R is called the *focal function* and denoted by p. The natural question of characterization of focal functions of closed convex reflectors was partially answered in [17].

The **reflector map** (possibly multivalued) generalizing (1) and denoted again by ξ is defined for $\mathbf{x} \in \mathbf{S^d}$ as

$$\xi(\mathbf{x}) = \{\mathbf{y} \in \mathbf{S^d} \mid P(\mathbf{y}) \text{ is supporting to } R \text{ at } \rho(\mathbf{x})\mathbf{x}\}. \tag{7}$$

It follows from (5) and (6) that, equivalently, the reflector map can be defined as

$$\xi(\mathbf{x}) = \{\mathbf{y} \in \mathbf{S^d} \mid \log \rho(\mathbf{x}) - \log p(\mathbf{y}) = -\log(1 - \mathbf{x} \cdot \mathbf{y})\}. \tag{8}$$

The total amount of energy transferred from the source $\mathcal{O}$ in a given set of directions $\omega \subset \mathbf{Y}$ is best described with the use of the map ξ^{-1} defined on any subset $\omega \subset \mathbf{Y}$ by setting

$$\xi^{-1}(\omega) = \bigcup_{\mathbf{y} \in \omega} \xi^{-1}(\mathbf{y}).$$

It is shown in [3, 17] that for any Borel subset $\omega \subset \mathbf{Y}$ the set $\xi^{-1}(\omega)$ is Lebesgue measurable on $\mathbf{S^d}$. Thus, if $I \geq 0$ is the intensity of the source then the total amount of energy transferred by R to the set $\omega \subset \mathbf{Y}$ is given by the "energy" function

$$G(R, \omega) = \int_{\xi^{-1}(\omega)} I(\mathbf{x}) \mathrm{d}\sigma(\mathbf{x}), \tag{9}$$

where $\mathrm{d}\sigma$ is the standard d-volume form on $\mathbf{S^d}$. It is assumed here and everywhere below that I is extended from $\mathbf{X}$ to the entire $\mathbf{S^d}$ by setting $I(\mathbf{x}) \equiv 0 \ \forall \mathbf{x} \in \mathbf{S^d} \setminus \mathbf{X}$. The function $G(R, \omega)$ is a Borel measure on $\mathbf{Y}$ (not necessarily absolutely continuous) [3, 17].

For a given nonnegative and integrable function L on $\mathbf{Y}$ we say that a closed convex reflector R is a **weak solution** of the reflector problem if

$$G(R, \omega) = \int_{\omega} L(\mathbf{y}) \mathrm{d}\sigma(\mathbf{y}) \quad \text{for any Borel set} \quad \omega \subset \mathbf{Y}. \tag{10}$$

Put

$$\mu[A] = \int_A I(\mathbf{x}) \mathrm{d}\sigma(\mathbf{x}), \qquad \nu[B] = \int_B L(\mathbf{y}) \mathrm{d}\sigma(\mathbf{y}) \tag{11}$$

for $A \subset \mathbf{X}$ and $B \subset \mathbf{Y}$ Borel sets. An obvious necessary condition for existence of a weak solution to the reflector problem is that the total energy of the source and the total energy on the output aperture are in balance, that is,

$$\mu[\mathbf{X}] = \nu[\mathbf{Y}]. \tag{12}$$

Excluding the trivial case when either $\mu[\mathbf{X}]$ or $\nu[\mathbf{Y}]$ is zero, it may be assumed, without loss of generality, that measures satisfying (12) are normalized so that they are probability measures.

The following existence result was established by Caffarelli and Oliker in [3] (and reproduced partly in [4] and [17]).

Theorem 2.1. *Let $\mathbf{X}$ and $\mathbf{Y}$ be closed sets on $\mathbf{S^d}$ (possibly coinciding with $\mathbf{S^d}$) and $I \geq 0$, $L \geq 0$ two probability densities on $\mathbf{X}$ and $\mathbf{Y}$, respectively. Then there exists a reflector which is a weak solution of the reflector problem.*

The proof is obtained in two steps. First the problem is solved in the case when the right hand side in (10) is a finite sum of Dirac masses. In this case a constructive minimization procedure together with an *a priori* two-sided C^0 estimate of ρ is used to obtain the weak solution, which is also unique. The general problem is solved by approximating the right hand side in (10) by finite sums of Dirac masses and obtaining the solution R as a limit of a sequence of special solutions $R_k, k = 1, 2, ...,$ constructed on the previous step. The measures $G(R_k, \cdot)$ are weakly continuous and converge to $G(R, \cdot)$. Consequently, R is indeed a weak solution of the reflector problem.

The procedure we have just described proves, in fact, that the reflector problem admits a solution if the right hand side in (10) is any nonnegative probability measure on $\mathbf{Y}$, possibly with a singular part.

Existence of regular solutions was studied by Wang [21] and Guan and Wang [10].

In the framework of Monge-Kantorovich theory the reflector problem was studied by Glimm and Oliker [8] and Wang [22]. The following result was proved in [8], Theorem 4.1.

Theorem 2.2. *Let R be a weak solution of the reflector problem with the reflector map ξ. Then $\xi_\# \mu = \nu$ (that is, ξ pushes μ forward to ν) and it is a minimizer of the problem*

$$\inf_{\xi'} \left\{ \int_{\mathbf{X}} -\log(1 - \mathbf{x} \cdot \xi'(\mathbf{x})) \mathrm{d}\mu(\mathbf{x}) \mid \xi'_\# \mu = \nu \right\}. \tag{13}$$

Furthermore, any other minimizer of (13) is equal to ξ almost everywhere on the set $\{\mathbf{x} \in \mathbf{S^d} \mid I(\mathbf{x}) \neq 0\}$.

Remark 2.3. Theorems 2.1 and 2.2, together, imply existence of minimizers to the problem (13). In addition, if $I > 0$ on $\mathbf{X}$ and $\mathbf{X}$ is connected, these results imply that, except for a set of measure zero, any minimizer of (13) (with μ and ν as in (11)) is a reflector map associated with a closed convex reflector in $\mathbf{R}^{d+1}$. Such reflector is unique up to a constant multiple of the function $\rho(\mathbf{x})$ in (6); see [8].

On the other hand, the minimization problem (13) is a variant of the Monge problem on $\mathbf{S^d}$ (see the beginning of Sect. 3, below). By a different method, in the framework of Monge-Kantorovich theory, the existence and uniqueness of minimizers to (13) was proved in [8] and [22] under the additional condition that $\mathrm{spt}(\mu) \cap \mathrm{spt}(\nu) = \emptyset$. Furthermore, if ρ and p are the functions in (8) then $u_o(\mathbf{x}) = \log \rho(\mathbf{x})$ and $v_o(\mathbf{y}) = -\log p(\mathbf{y})$ maximize

$$(u, v) \rightarrow \int_{\mathbf{X}} u(\mathbf{x}) \mathrm{d}\mu(\mathbf{x}) + \int_{\mathbf{Y}} v(\mathbf{y}) \mathrm{d}\nu(\mathbf{y})$$

over the set of pairs $(u, v) \in C(\mathbf{X} \times \mathbf{Y})$ satisfying $u(\mathbf{x}) + v(\mathbf{y}) \leq -\log(1 - \mathbf{x} \cdot \mathbf{y})$ for all $(\mathbf{x}, \mathbf{y}) \in \mathbf{X} \times \mathbf{Y}$.

3. Background on the Monge-Kantorovich theory

As observed in Section 2, the reflector problem can be stated as a variant of the problem of Monge with the cost function $c(\mathbf{x}, \mathbf{y}) = -\log(1 - \mathbf{x} \cdot \mathbf{y})$, $\mathbf{x}, \mathbf{y} \in \mathbf{S^d}$. In this section we recall some facts of the Monge-Kantorovich theory that will be needed to study an analogue of the reflector problem with general cost functions.

We first fix some notation. For a set $Z \subset \mathbf{R}^{d+1}$ we denote by $\mathcal{P}(Z)$ the set of Borel probability measures on Z. As usual, if Z is a closed subset of $\mathbf{R}^{d+1}$ and $\gamma \in \mathcal{P}(Z)$ then the support of γ is the smallest closed set $\operatorname{spt}\gamma \subset Z$ such that $\gamma[\operatorname{spt}\gamma] = \gamma[Z] = 1$.

Suppose, $\mathbf{X}, \mathbf{Y} \subset \mathbf{R}^{d+1}$ are closed sets. If $\mu \in \mathcal{P}(\mathbf{X})$, $\nu \in \mathcal{P}(\mathbf{Y})$, we denote by $\Gamma(\mu, \nu)$ the set of joint measures γ on $\mathbf{R}^{d+1} \times \mathbf{R}^{d+1}$ that have μ and ν as their *marginals*: $\mu[U] = \gamma[U \times \mathbf{R}^{d+1}]$ and $\gamma[\mathbf{R}^{d+1} \times U] = \nu[U]$ for Borel $U \subset \mathbf{R}^{d+1}$. In fact, if $\gamma \in \Gamma(\mu, \nu)$ then $\operatorname{spt}\gamma \subset \mathbf{X} \times \mathbf{Y}$.

Assume that we are given two probability measures μ and ν on $\mathbf{R}^{d+1}$. Let $\Gamma(\mu, \nu)$ be the set of joint measures γ on $\mathbf{R}^{d+1} \times \mathbf{R}^{d+1}$ that have μ and ν as their marginals. Kantorovich's problem is to minimize the *transport cost*

$$\mathcal{C}(\gamma) := \int c(\mathbf{x}, \mathbf{y}) \, d\gamma(\mathbf{x}, \mathbf{y}) \tag{14}$$

for some given c among joint measures γ in $\Gamma(\mu, \nu)$, to obtain

$$\inf_{\gamma \in \Gamma(\mu, \nu)} \mathcal{C}[\gamma]. \tag{15}$$

Let $\mathcal{T}(\mu, \nu)$ be the set of Borel maps $T : \mathbf{R}^{d+1} \to \mathbf{R}^{d+1}$ that push μ forward to ν : $\mu[T^{-1}(B)] = \nu[B]$ for all Borel sets $B \subset \mathbf{R}^{d+1}$. The Monge problem is to minimize

$$\mathcal{I}[T] = \int_{\mathbf{R}^{d+1}} c(\mathbf{x}, T\mathbf{x}) d\mu(\mathbf{x})$$

over the set $\mathcal{T}(\mu, \nu)$.

There is a natural embedding which associates to $T \in \mathcal{T}(\mu, \nu)$ a $\gamma_T := (\mathbf{id} \times T)_{\#}\mu \in \Gamma(\mu, \nu)$, where $\mathbf{id} : \mathbf{R}^{d+1} \to \mathbf{R}^{d+1}$ is the identity map. Since $\mathcal{C}[\gamma_T] = \mathcal{I}[T]$ we conclude that

$$\inf_{\gamma \in \Gamma(\mu, \nu)} \mathcal{C}[\gamma] \leq \inf_{T \in \mathcal{T}(\mu, \nu)} \mathcal{I}[T]. \tag{16}$$

Throughout this section we use the notation $\mathbf{R} \cup \{+\infty\} = \bar{\mathbf{R}}$ and assume that $c : \mathbf{R}^{d+1} \times \mathbf{R}^{d+1} \to \bar{\mathbf{R}}$. We endow $\bar{\mathbf{R}}$ with the usual topology so that $c \in C(\mathbf{R}^{d+1} \times \mathbf{R}^{d+1}, \bar{\mathbf{R}})$ means that

$$\lim_{(\mathbf{x}, \mathbf{y}) \to (\bar{\mathbf{x}}, \bar{\mathbf{y}})} c(\mathbf{x}, \mathbf{y}) = c(\bar{\mathbf{x}}, \bar{\mathbf{y}}).$$

In particular, if $c(\bar{\mathbf{x}}, \bar{\mathbf{y}}) = +\infty$ then $c(\mathbf{x}, \mathbf{y})$ tends to $+\infty$ as $(\mathbf{x}, \mathbf{y})$ tends to $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$.

Definition 3.1. A subset $S \subset \mathbf{R}^{d+1} \times \mathbf{R}^{d+1}$ is said to be c-cyclically monotone if for every natural number n and every $\{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^n \subset S$ we have

$$\sum_{i=1}^n c(\mathbf{x}_i, \mathbf{y}_i) \leq \sum_{i=1}^n c(\mathbf{x}_i, \mathbf{y}_{\sigma(i)}),$$

for all permutation σ of n letters.

This notion of c-cyclical monotonicity was introduced by Knott and Smith [13]. When $c(\mathbf{x}, \mathbf{y}) = |\mathbf{x} - \mathbf{y}|^2$, c-cyclical monotonicity is simply called cyclical monotonicity [18].

Proposition 3.2. *Assume that* $\mathbf{X}, \mathbf{Y} \subset \mathbf{R}^{d+1}$ *are closed sets, that* $\mu \in \mathcal{P}(\mathbf{X})$, $\nu \in \mathcal{P}(\mathbf{Y})$ *and that* $c \geq 0$ *is lower semicontinuous on* $\mathbf{X} \times \mathbf{Y}$. *Then,*

(i) there is at least one optimal measure $\gamma_o \in \Gamma(\mu, \nu)$.

(ii) Suppose that in addition $c \in C(\mathbf{X} \times \mathbf{Y}, \bar{\mathbf{R}})$. *Unless* $\mathcal{C} \equiv +\infty$ *throughout* $\Gamma(\mu, \nu)$, *there is a c-cyclically monotone set* $S \subset \mathbf{R}^{d+1} \times \mathbf{R}^{d+1}$ *containing the support of all optimal measures in* $\Gamma(\mu, \nu)$.

Part (i) of Proposition 3.2 can be found in [11] (Th. 2.19). Note that that proof is interesting only in the case $\mathcal{C} \not\equiv +\infty$. Part (ii) was established in [1] and [6] for $c \in C(\mathbf{X} \times \mathbf{Y}, \mathbf{R})$. One can readily adapt the proof of Theorem 2.3 and Corollary 2.4 of [6] to cost functions $c \in C(\mathbf{X} \times \mathbf{Y}, \bar{\mathbf{R}})$.

Definition 3.3. Suppose that $\psi : \mathbf{R}^{d+1} \to \mathbf{R} \cup \{-\infty\}$ is not identically $-\infty$. Then
(i) ψ is said to be *c-concave* if there exists a set $\mathcal{A} \subset \mathbf{R}^{d+1} \times \mathbf{R}$ such that

$$\psi(\mathbf{x}) = \inf_{(\mathbf{y}, \lambda) \in \mathcal{A}} c(\mathbf{x}, \mathbf{y}) + \lambda, \qquad (\mathbf{x} \in \mathbf{X}). \tag{17}$$

(ii) The *c-superdifferential* $\partial^c \psi$ of ψ consists of the pairs $(\mathbf{x}, \mathbf{y}) \in \mathbf{R}^{d+1} \times \mathbf{R}^{d+1}$ for which

$$c(\mathbf{x}, \mathbf{y}) - \psi(\mathbf{x}) \leq c(\mathbf{v}, \mathbf{y}) - \psi(\mathbf{v}), \tag{18}$$

for all $\mathbf{v} \in \mathbf{R}^{d+1}$.
(iii) We define $\partial^c \psi(\mathbf{x}) \subset \mathbf{R}^{d+1}$ to be the set of $\mathbf{y}$ such that $(\mathbf{x}, \mathbf{y}) \in \partial^c \psi(\mathbf{x})$. If $E \subset \mathbf{R}^{d+1}$, $\partial^c \psi(E)$ is the union of the $\partial^c \psi(\mathbf{x})$ such that $\mathbf{x} \in E$.
(iv) The c–transform of ψ is the function $\mathbf{y} \to \psi^c(\mathbf{y}) = \inf_{\mathbf{x} \in \mathbf{X}} \{c(\mathbf{x}, \mathbf{y}) - \psi(\mathbf{x})\}$.

Remark 3.4. Suppose that $\psi : \mathbf{R}^{d+1} \to \mathbf{R} \cup \{-\infty\}$ is not identically $-\infty$ and is given by (17). We have
(i) $\psi^c(\mathbf{y}) \geq -\lambda > -\infty$ if $(\mathbf{y}, \lambda) \in \mathcal{A}$ where $\mathcal{A}$ is the set in (17). Hence, $\psi^c \not\equiv -\infty$.
(ii) $\psi^{cc} = \psi$.

Proof. The proofs of these remarks are well documented when $c : \mathbf{R}^{d+1} \times \mathbf{R}^{d+1} \to \mathbf{R}$. We verify that the same proofs apply when c may take the value $+\infty$.

If $(\mathbf{y}, \lambda) \in \mathcal{A}$ then $-\lambda \leq c(\mathbf{x}, \mathbf{y}) - \psi(\mathbf{x})$ for all $\mathbf{x} \in \mathbf{X}$. Hence $\psi^c(\mathbf{y})$, the infimum of $c(\mathbf{x}, \mathbf{y}) - \psi(\mathbf{x})$ over $\mathbf{X}$, is not smaller than $-\lambda$. This proves (i).

The inequality $\psi^{cc} \geq \psi$ which holds for general functions is readily checked. It remains to prove that when (17) holds, then $\psi^{cc} \leq \psi$. Fix $\mathbf{x} \in \mathbf{X}$ and let $\{(\mathbf{y}_n, \lambda_n)\}_{n=1}^{\infty} \subset \mathcal{A}$ be such that

$$\psi(\mathbf{x}) = \lim_{n \to +\infty} c(\mathbf{x}, \mathbf{y}_n) + \lambda_n.$$

By (i), $\lambda_n \geq -\psi^c(\mathbf{y}_n)$ and so,

$$\psi(\mathbf{x}) \geq \limsup_{n \to +\infty} c(\mathbf{x}, \mathbf{y}_n) - \psi^c(\mathbf{y}_n) \geq \psi^{cc}(\mathbf{x}).$$

This proves (ii). □

4. Existence and uniqueness of optimal maps

Throughout this section, we assume that
(A1) $c \in C^1(\mathbf{R}^{d+1} \times \mathbf{R}^{d+1} \setminus \Delta)$.
(A2) $c : \mathbf{R}^{d+1} \times \mathbf{R}^{d+1} \to [0, +\infty]$ *is lower semicontinuous.*
(A3) *for any* $\mathbf{x}_o \in \mathbf{R}^{d+1}$, $c(\mathbf{x}_o, \mathbf{x}_o) = +\infty$.

We are interested in probability measures μ, ν for which $\mathcal{C} \not\equiv +\infty$ throughout $\Gamma(\mu, \nu)$. Assume that $c \in C(\mathbf{R}^{d+1} \times \mathbf{R}^{d+1} \setminus \Delta, \mathbf{R})$, where $\Delta := \{(\mathbf{x}, \mathbf{x}) \mid \mathbf{x} \in \mathbf{R}^{d+1}\}$ denotes the diagonal. Proposition 4.1 provides a

sufficient condition which ensures that $\mathcal{C} \not\equiv +\infty$ throughout $\Gamma(\mu, \nu)$. Before stating that proposition, let us introduce the sets

$$S(a, b) = \{\mathbf{x} = (x_1, \cdots, x_d, x_{d+1}) \mid a \leq x_{d+1} \leq b\},$$

for $-1 \leq a \leq b \leq 1$.

Proposition 4.1. *Suppose that $\mu, \nu \in \mathcal{P}(\mathbf{S^d})$ are Borel measures which vanish on $(d-1)$-rectifiable sets. Then there exists $\gamma \in \Gamma(\mu, \nu)$ and $\epsilon > 0$ such that*

$$|\mathbf{x} - \mathbf{y}| \geq \epsilon$$

for all $(\mathbf{x}, \mathbf{y}) \in \mathrm{spt}\gamma$.

Proof. Since μ and ν vanish on $(d-1)$-rectifiable sets, the functions

$$t \to \mu[S(a, t)], \quad t \to \nu[S(a, t)] \quad \text{are continuous.} \tag{19}$$

Case 1. Assume first that there exists $c \in (-1, 1)$ such that

$$\mathrm{spt}\mu \subset S(-1, c), \quad \mathrm{spt}\nu \subset S(c, 1). \tag{20}$$

Thanks to (19) we may choose $\epsilon_1, \epsilon_2 > 0$ such that

$$\mu[S(-1, -1 + \epsilon_1)] = \nu[S(c, c + \epsilon_2)] = 1/2.$$

By (20) $-1 + \epsilon_1 < c$. Set

$$\bar{\gamma} = 2(\mu^- \otimes \nu^- + \mu^+ \otimes \nu^+)$$

where

$$\mu^- = \mu|_{S(-1, -1+\epsilon_1)}, \quad \mu^+ = \mu|_{S(-1+\epsilon_1, c)}, \quad \nu^- = \nu|_{S(c, c+\epsilon_2)}, \quad \nu^+ = \nu|_{S(c+\epsilon_2, 1)}.$$

Note that $\bar{\gamma} \in \Gamma(\mu, \nu)$ and

$$|\mathbf{x} - \mathbf{y}| \geq \min\{c + 1 - \epsilon_1, \epsilon_2\} > 0$$

for all $(\mathbf{x}, \mathbf{y}) \in \mathrm{spt}\gamma$. This proves the proposition in this special case.

Case 2. Assume that $\mathrm{spt}\mu$ and $\mathrm{spt}\nu$ are arbitrary. We use (19) to choose $c \in (-1, 1)$ such that

$$\mu[S(c, 1)] = \nu[S(-1, c)] := m. \tag{21}$$

If $m = 0$ then

$$\mathrm{spt}\mu \subset S(-1, c), \quad \mathrm{spt}\nu \subset S(c, 1)$$

and so, we reduce the discussion to the case 1. Similarly, if $m = 1$ we reduce the discussion to the case 1.
    Assume in the sequel that $0 < m < 1$. Set

$$\mu^- = \mu|_{S(-1, c)}, \quad \mu^+ = \mu|_{S(c, 1)}, \quad \nu^- = \nu|_{S(-1, c)}, \quad \nu^+ = \nu|_{S(c, 1)}.$$

By (21), $\frac{\mu^+}{m}$ and $\frac{\nu^-}{m}$ are probability measures. They satisfy

$$\mathrm{spt}\nu^- \subset S(-1, c), \quad \mathrm{spt}\mu^+ \subset S(c, 1).$$

Having that ν^-, μ^+ satisfy the assumptions of case 1, we may find $\bar{\gamma} \in \Gamma(\frac{\mu^+}{m}, \frac{\nu^-}{m})$ and $\bar{\epsilon} > 0$ such that

$$|\mathbf{x} - \mathbf{y}| \geq \bar{\epsilon} \tag{22}$$

for all $(\mathbf{x}, \mathbf{y}) \in \mathrm{spt}\bar{\gamma}$. Similarly, there exists $\tilde{\gamma} \in \Gamma(\frac{\mu^-}{1-m}, \frac{\nu^+}{1-m})$ and $\tilde{\epsilon} > 0$ such that

$$|\mathbf{x} - \mathbf{y}| \geq \tilde{\epsilon} \tag{23}$$

for all $(\mathbf{x}, \mathbf{y}) \in \mathrm{spt}\tilde{\gamma}$.

Set

$$\gamma = m\bar{\gamma} + (1-m)\tilde{\gamma}.$$

Then $\gamma \in \Gamma(\mu, \nu)$ and by (22)–(23), we have that

$$|\mathbf{x} - \mathbf{y}| \geq \min\{\bar{\epsilon}, \tilde{\epsilon}\} > 0$$

for all $(\mathbf{x}, \mathbf{y}) \in \mathrm{spt}\tilde{\gamma}$. This proves the proposition. $\qquad \square$

Lemma 4.2. *Suppose that c satisfies* (A1)–(A3) *and that $\mathbf{X}, \mathbf{Y} \subset \mathbf{R}^{d+1}$ are closed sets. Suppose that $S \subset \mathbf{X} \times \mathbf{Y}$ is c-cyclically monotone and contains two pairs $(\mathbf{x}_o, \mathbf{y}_o)$, $(\bar{\mathbf{x}}_o, \bar{\mathbf{y}}_o)$ such that $\mathbf{x}_o \neq \bar{\mathbf{x}}_o$ and $\mathbf{y}_o \neq \bar{\mathbf{y}}_o$. Then, there exists a function $F \in C(\mathbf{X} \times \mathbf{Y})$ depending only on the pairs such that*

$$c(\mathbf{x}_o, \mathbf{y}_o) + c(\bar{\mathbf{x}}_o, \bar{\mathbf{y}}_o) + c(\mathbf{x}, \mathbf{y}) \leq F(\mathbf{x}, \mathbf{y}) \tag{24}$$

for all $(\mathbf{x}, \mathbf{y}) \in S$.

Proof. Define

$$\begin{cases} F(\mathbf{x}, \mathbf{y}) &= \min\{c(\mathbf{x}, \mathbf{y}_o) + R_1(\mathbf{y}), c(\mathbf{x}, \bar{\mathbf{y}}_o) + R_2(\mathbf{y})\} \\ R_1(\mathbf{y}) &= \min\{c(\mathbf{x}_o, \mathbf{y}) + c(\bar{\mathbf{x}}_o, \bar{\mathbf{y}}_o), c(\mathbf{x}_o, \bar{\mathbf{y}}_o) + c(\bar{\mathbf{x}}_o, \mathbf{y})\} \\ R_2(\mathbf{y}) &= \min\{c(\mathbf{x}_o, \mathbf{y}) + c(\bar{\mathbf{x}}_o, \mathbf{y}_o), c(\mathbf{x}_o, \mathbf{y}_o) + c(\bar{\mathbf{x}}_o, \mathbf{y})\}. \end{cases} \tag{25}$$

We use that $\mathbf{x}_o \neq \bar{\mathbf{x}}_o$ and (A1)–(A3) to obtain that $R_1, R_2 \in C(\mathbf{Y})$. This, together with the fact that $\mathbf{y}_o \neq \bar{\mathbf{y}}_o$ yields that F is continuous on $\mathbf{X} \times \mathbf{Y}$. If $(\mathbf{x}, \mathbf{y})$ is another element of S then setting

$$(\mathbf{x}_1, \mathbf{y}_1) = (\mathbf{x}_o, \mathbf{y}_o), \quad (\mathbf{x}_2, \mathbf{y}_2) = (\bar{\mathbf{x}}_o, \bar{\mathbf{y}}_o), \quad (\mathbf{x}_3, \mathbf{y}_3) = (\mathbf{x}, \mathbf{y})$$

and using the c-cyclical monotonicity of S, we obtain (24). $\qquad \square$

It is well known that a set is cyclically monotone if and only if it is contained in the subdifferential of a convex function [18]. An analogue of this result was proved by Smith and Knott [20] for general cost functions $c : \mathbf{X} \times \mathbf{Y} \to \mathbf{R}$. The following Lemma 4.3 is a further extension that is needed to deal with cost functions satisfying (A1)–(A3) (and may be $= +\infty$ somewhere). Below, we check first that the proof in [19] extends to such cost functions and then we show that the infimum in (26) can be performed over the subset of $\mathbf{y}$ such that $|\mathbf{x} - \mathbf{y}| > \delta > 0$.

Lemma 4.3. *Suppose that c satisfies* (A1)–(A3) *and $\mathbf{X}, \mathbf{Y} \subset \mathbf{R}^{d+1}$ are compact sets. Suppose that $S \subset \mathbf{X} \times \mathbf{Y}$ is c-cyclically monotone and contains two pairs $(\mathbf{x}_o, \mathbf{y}_o)$, $(\bar{\mathbf{x}}_o, \bar{\mathbf{y}}_o)$ such that $\bar{\mathbf{x}}_o, \bar{\mathbf{y}}_o \notin \{\mathbf{x}_o, \mathbf{y}_o\}$ Then,*

(i) S is contained in the c-superdifferential of a c-concave function $\psi : \mathbf{R}^{d+1} \to \mathbf{R} \cup \{-\infty\}$ such that there exists $\delta > 0$ satisfying

$$\psi(\mathbf{x}) = \inf_{\mathbf{y}} \Big\{ c(\mathbf{x}, \mathbf{y}) - \psi^c(\mathbf{y}) \mid |\mathbf{x} - \mathbf{y}| \geq \delta \Big\}, \qquad (\mathbf{x} \in \mathbf{X}). \tag{26}$$

(ii) If $(\mathbf{x}, \mathbf{y}) \in \partial^c \psi$ then $\psi(\mathbf{x}) > -\infty$ and $|\mathbf{x} - \mathbf{y}| \geq \delta$ for some $\delta > 0$ that depends only on ψ.

Proof. The expression $\psi(\mathbf{x}) = \inf_{\mathbf{y} \in \mathbf{Y}} \Big\{ c(\mathbf{x}, \mathbf{y}) - \psi^c(\mathbf{y}) \Big\}$, $\mathbf{x} \in \mathbf{X}$, is well-known in the literature. The only new and useful fact we want to point out is that $\psi(\mathbf{x})$ will be obtained by minimizing $c(\mathbf{x}, \mathbf{y}) - \psi^c(\mathbf{y})$ not on $\mathbf{Y}$ (as it is usually done), but on $\{y \in \mathbf{Y} \mid |\mathbf{x} - \mathbf{y}| \geq \delta\}$. For completeness, we give the detailed proof below.

Since $\mathbf{X} \times \mathbf{Y}$ is compact, the function F defined in (24) attains its maximum. We use Lemma 4.2 and the fact that c is lower semicontinuous and equals $+\infty$ on Δ to conclude the following: if $S' \subset \mathbf{X} \times \mathbf{Y}$ is c-cyclically monotone and contains $(\mathbf{x}_o, \mathbf{y}_o)$, $(\bar{\mathbf{x}}_o, \bar{\mathbf{y}}_o)$ then there exists $\delta > 0$ such that $|\mathbf{x} - \mathbf{y}| \geq 2\delta$ for every $(\mathbf{x}, \mathbf{y}) \in S'$. In particular, $\mathbf{x} \neq \mathbf{y}$ for $(\mathbf{x}, \mathbf{y}) \in S'$. In particular, there exists $\delta_S > 0$ such that $|\mathbf{x} - \mathbf{y}| \geq 2\delta_S$ for every $(\mathbf{x}, \mathbf{y}) \in S$.

As in [19], we define

$$\psi(\mathbf{x}) = \inf_n \inf_{\{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^n \subset S} \left\{ c(\mathbf{x}, \mathbf{y}_n) + \sum_{j=0}^{n-1} c(\mathbf{x}_{j+1}, \mathbf{y}_j) - \sum_{j=0}^{n} c(\mathbf{x}_j, \mathbf{y}_j) \right\}. \tag{27}$$

Since c is finite on S and nonnegative on $\mathbf{R}^{d+1} \times \mathbf{R}^{d+1}$, we conclude that $\psi(\mathbf{x})$ is well-defined.

The c-cyclical monotonicity of S gives that

$$c(\mathbf{x}_o, \mathbf{y}_n) + \sum_{j=0}^{n-1} c(\mathbf{x}_{j+1}, \mathbf{y}_j) - \sum_{j=0}^{n} c(\mathbf{x}_j, \mathbf{y}_j) \geq 0$$

and so, $\psi(\mathbf{x}_o) \geq 0$. Taking $n = 1$, $\mathbf{x}_1 = \mathbf{x}_o$ and $\mathbf{y}_1 = \mathbf{y}_o$ in (27) gives that $\psi(\mathbf{x}_o) \leq 0$. We conclude that $\psi(\mathbf{x}_o) = 0$ and so, ψ is not identically $-\infty$. This, together with the fact that ψ is clearly of the form (17) yields that ψ is c-concave.

Claim 1. We have that $S \subset \partial^c \psi$.

Fix $(\mathbf{x}, \mathbf{y}) \in S$ and for each m integer, let $\{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^{n_m} \subset S$ be such that

$$\lim_{m \to +\infty} \psi_m = \psi(\mathbf{x}),$$

where

$$\psi_m := c(\mathbf{x}, \mathbf{y}_{n_m}) + \sum_{j=0}^{n_m - 1} c(\mathbf{x}_{j+1}, \mathbf{y}_j) - \sum_{j=0}^{n_m} c(\mathbf{x}_j, \mathbf{y}_j).$$

Setting $(\mathbf{x}_{n_m+1}, \mathbf{y}_{n_m+1}) = (\mathbf{x}, \mathbf{y})$ we have that

$$-c(\mathbf{x}, \mathbf{y}) + \psi_m = c(\mathbf{v}, \mathbf{y}_{n_m+1}) - c(\mathbf{v}, \mathbf{y}) + \sum_{j=0}^{n_m} c(\mathbf{x}_{j+1}, \mathbf{y}_j) - \sum_{j=0}^{n_m+1} c(\mathbf{x}_j, \mathbf{y}_j) \geq \psi(\mathbf{v}) - c(\mathbf{v}, \mathbf{y}).$$

Letting m go to $+\infty$ we conclude that

$$\psi(\mathbf{x}) - c(\mathbf{x}, \mathbf{y}) \geq \psi(\mathbf{v}) - c(\mathbf{v}, \mathbf{y}), \tag{28}$$

which proves claim 1.

Claim 2. Whenever $(\mathbf{x}, \mathbf{y}) \in \partial^c \psi$, we have that $\psi(\mathbf{x}) > -\infty$ and $\mathbf{x} \neq \mathbf{y}$.

Recall that $(\mathbf{x}, \mathbf{y}) \in \partial^c \psi$ is equivalent to (28). Setting $(\mathbf{x}, \mathbf{y}) = (\bar{\mathbf{x}}_o, \bar{\mathbf{y}}_o)$, $\mathbf{v} = \mathbf{x}_o$ in (28), using the facts that $\psi(\mathbf{x}_o) = 0$ and $\mathbf{x}_o \neq \bar{\mathbf{y}}_o$ we obtain that $\psi(\bar{\mathbf{x}}_o)$ is finite. Next, if $(\mathbf{u}_o, \mathbf{v}_o) \in S$, setting $\mathbf{v} = \mathbf{u}_o$ we have that

$$c(\mathbf{x}, \mathbf{y}) - \psi(\mathbf{x}) \leq c(\mathbf{u}_o, \mathbf{y}) - \psi(\mathbf{u}_o). \tag{29}$$

If $\mathbf{y} \neq \mathbf{x}_o$ we set $\mathbf{u}_o = \mathbf{x}_o$ in (29) to obtain the claim. If $\mathbf{y} = \mathbf{x}_o$, we set $\mathbf{u}_o = \bar{\mathbf{x}}_o$ and we use the fact that $\psi(\bar{\mathbf{x}}_o)$ is finite to obtain the claim.

Claim 3. The set $\partial^c \psi$ is c-cyclically monotone.

For the sake of completeness, we reprove this claim although it is a repetition of a known argument in [19]. If $\{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^n \subset \partial^c \psi$, setting $(\mathbf{x}_{n+1}, \mathbf{y}_{n+1}) = (\mathbf{x}_1, \mathbf{y}_1)$ we have that

$$c(\mathbf{x}_i, \mathbf{y}_i) - \psi(\mathbf{x}_i) \leq c(\mathbf{x}_{i+1}, \mathbf{y}_i) - \psi(\mathbf{x}_{i+1}). \tag{30}$$

By claim 2, each term in (30) is finite and so,

$$0 = \sum_{i=1}^n \psi(\mathbf{x}_{i+1}) - \psi(\mathbf{x}_i) \leq \sum_{i=1}^n c(\mathbf{x}_{i+1}, \mathbf{y}_i) - c(\mathbf{x}_i, \mathbf{y}_i),$$

which proves the claim.

We use Lemma 4.2, the facts that $\partial^c \psi$ is c-cyclically monotone, that $(\mathbf{x}_o, \mathbf{y}_o)$, $(\bar{\mathbf{x}}_o, \bar{\mathbf{y}}_o) \in \partial^c \psi$, that $\mathbf{x}_o \neq \bar{\mathbf{x}}_o$ and $\mathbf{y}_o \neq \bar{\mathbf{y}}_o$, to obtain the existence of some $\delta > 0$ such that $|\mathbf{x} - \mathbf{y}| \geq 2\delta$ for all $(\mathbf{x}, \mathbf{y}) \in \partial^c \psi$. By Remark 3.4 (ii) $\psi = (\psi^c)^c$ and so if $\mathbf{x} \in \mathbf{X}$, there exists a sequence $\{\mathbf{y}_n\}_{n=1}^\infty \subset \mathbf{Y}$ such that

$$\psi(\mathbf{x}) = \lim_{n \to +\infty} c(\mathbf{x}, \mathbf{y}_n) - \psi^c(\mathbf{y}_n). \tag{31}$$

Since $\mathbf{Y}$ is compact, we may extract from $\{\mathbf{y}_n\}_{n=1}^\infty$ a subsequence (which we still label $\{\mathbf{y}_n\}_{n=1}^\infty$) that converges to some $\mathbf{y} \in \mathbf{Y}$. Recall that the function ψ^c is upper semicontinuous as an infimum of upper semicontinuous functions and therefore, (31) yields

$$\psi(\mathbf{x}) \geq c(\mathbf{x}, \mathbf{y}) - \psi^c(\mathbf{y}).$$

This proves that $\psi(\mathbf{x}) = c(\mathbf{x}, \mathbf{y}) - \psi^c(\mathbf{y})$ and so, $(\mathbf{x}, \mathbf{y}) \in \partial^c \psi$. Hence $|\mathbf{x} - \mathbf{y}| \geq 2\delta$ and $\mathbf{y}$ is a minimizer in (26). $\square$

Remark 4.4. Suppose that $\mu \in \mathcal{P}(\mathbf{X})$ and $\nu \in \mathcal{P}(\mathbf{Y})$ have no atoms and that γ minimizes $\mathcal{C}$ over $\Gamma(\mu, \nu)$ and that $\mathcal{C}(\gamma) < +\infty$. Then, $\gamma[\Delta] = 0$ and so, $\mathrm{spt}(\gamma) \setminus \Delta$ contains at least one element, say $(\mathbf{x}_o, \mathbf{y}_o)$. Also, $\gamma[E] = 0$ where

$$E = \left(\{\mathbf{x}_o, \mathbf{y}_o\} \times \mathbf{Y} \right) \cup \left(\mathbf{X} \times \{\mathbf{x}_o, \mathbf{y}_o\} \right).$$

Hence, the set $\mathbf{X} \times \mathbf{Y} \setminus (E \cup \Delta))$ is nonempty, and so, it contains an element $(\bar{\mathbf{x}}_o, \bar{\mathbf{y}}_o)$. Note that $\bar{\mathbf{x}}_o, \bar{\mathbf{y}}_o \notin \{\mathbf{x}_o, \mathbf{y}_o\}$.

We now further specialize the set of cost functions under consideration by assuming that

$$c(\mathbf{x}, \mathbf{y}) = \begin{cases} l(\frac{|\mathbf{x} - \mathbf{y}|^2}{2}) & \text{if} \quad \mathbf{x} \neq \mathbf{y} \\ +\infty & \text{if} \quad \mathbf{x} = \mathbf{y} \end{cases} \tag{32}$$

where $l \in C^2(0, +\infty)$ is such that

$$\begin{cases} \lim_{t \to 0^+} l(t) = +\infty \\ |l'(t)| > 0 \quad (t > 0). \end{cases} \tag{33}$$

Define

$$g(t) = t(2 - t)(l'(t))^2. \tag{34}$$

Note that if g is monotone on $(0, 2]$ then $g[\frac{\delta}{2}, 2]$ is a closed interval on which g^{-1} exists and is continuous. Define

$$M(\mathbf{a}, \mathbf{x}) = \left[1 - g^{-1}(|\mathbf{a}|^2) \right] \mathbf{x} - \frac{\mathbf{a}}{l'\left(g^{-1}(|\mathbf{a}|^2)\right)},$$

and the closed set

$$K_\delta := \left\{ (\mathbf{a}, \mathbf{x}) \in \mathbf{R}^{d+1} \times \mathbf{S^d} \mid |\mathbf{a}|^2 \in g\left[\frac{\delta}{2}, 2\right] \right\}.$$

Note that M is continuous on K_δ for all $\delta > 0$.

When $\mathbf{x} \in \mathbf{S^d}$ we next denote by $\nabla^{\mathbf{x}}_{\mathbf{S^d}} c$ the tangential gradient of $\mathbf{x} \to c(\mathbf{x}, \mathbf{y})$ at $\mathbf{x} \in \mathbf{S^d}$ and let $T_{\mathbf{x}}$ be the tangent space to $\mathbf{S^d}$ at $\mathbf{x} \in \mathbf{S^d}$.

Lemma 4.5. *Assume that c is given by (32) such that l satisfies (33). Assume that g is monotone on $(0, 2]$. If $0 \neq \mathbf{a} \in \mathbf{R}^{d+1}$, $\mathbf{x}, \mathbf{y} \in \mathbf{S^d}$ and $\nabla^{\mathbf{x}}_{\mathbf{S^d}} c(\mathbf{x}, \mathbf{y}) = \mathbf{a}$ then $\mathbf{y} = M(\mathbf{a}, \mathbf{x})$.*

Proof. Recall that if $\mathbf{x} \in \mathbf{S^d}$ then the orthogonal projection of $\mathbf{y} \in \mathbf{R}^{d+1}$ onto $T_{\mathbf{x}}$ is $\mathbf{y}_{\parallel} = \mathbf{y} - (\mathbf{x} \cdot \mathbf{y})\mathbf{x}$. Hence if $\mathbf{y} \in \mathbf{S^d}$ then

$$|\mathbf{y}_{\parallel}|^2 + (\mathbf{x} \cdot \mathbf{y})^2 = 1. \tag{35}$$

Setting $\mathbf{a} := \nabla^{\mathbf{x}}_{\mathbf{S^d}} c(\mathbf{x}, \mathbf{y})$ yields that

$$\mathbf{a} = -l'\left(\frac{|\mathbf{x} - \mathbf{y}|^2}{2}\right)\mathbf{y}_{\parallel}. \tag{36}$$

This, together with (35) and the fact that $1 - \mathbf{x} \cdot \mathbf{y} = \frac{|\mathbf{x}-\mathbf{y}|^2}{2}$ for $\mathbf{x}, \mathbf{y} \in \mathbf{S^d}$, yields that

$$|\mathbf{a}|^2 = \left(l'\left(\frac{|\mathbf{x} - \mathbf{y}|^2}{2}\right)\right)^2 (1 - (\mathbf{x} \cdot \mathbf{y})^2) = g(1 - \mathbf{x} \cdot \mathbf{y}).$$

Thus,

$$1 - \mathbf{x} \cdot \mathbf{y} = g^{-1}(|\mathbf{a}|^2). \tag{37}$$

We use (36), (37) and the fact that $1 - \mathbf{x} \cdot \mathbf{y} = \frac{|\mathbf{x}-\mathbf{y}|^2}{2}$ to conclude that

$$\mathbf{y} = (\mathbf{x} \cdot \mathbf{y})\mathbf{x} + \mathbf{y}_{\parallel} = \left(1 - g^{-1}(|\mathbf{a}|^2)\right)\mathbf{x} - \frac{\mathbf{a}}{l'\left(g^{-1}(|\mathbf{a}|^2)\right)} = M(\mathbf{a}, \mathbf{x}). \qquad \square$$

Theorem 4.6. *Assume that c is given by (32) with $l \in C^2(0, +\infty)$ and satisfying (33). Assume that g given in (34) is monotone. Let $\mu, \nu \in \mathcal{P}(\mathbf{S^d})$ be Borel measures on the sphere $\mathbf{S^d}$ that vanish on $(d-1)$-rectifiable sets. Then*

(i) There exists a unique measure γ_o that minimizes $\mathcal{C}$ over $\Gamma(\mu, \nu)$.

(ii) There exists a unique map $T_o : \mathbf{S^d} \to \mathbf{S^d}$ that minimizes $\mathcal{I}$ over $\mathcal{T}(\mu, \nu)$. Furthermore the essential infimum $\inf_{\mathbf{x}} |T_o\mathbf{x} - \mathbf{x}| > 0$. The γ_o is uniquely determined and coincides with the measure $(\mathbf{id} \times T_o)_{\#}\mu$.

(iii) The map T_o is invertible except on a set whose μ measure is null.

Proof of (i): existence of the measure γ_o. By Proposition 4.1, there exists $\gamma \in \Gamma(\mu, \nu)$ and $\epsilon > 0$ such that

$$|\mathbf{x} - \mathbf{y}| \geq \epsilon \tag{38}$$

for all $(\mathbf{x}, \mathbf{y}) \in \mathrm{spt}\gamma$. Since $|\mathbf{x} - \mathbf{y}| \leq 2$ on $\mathbf{S^d}$, (38) and the fact that c satisfies (A1) ensures that c is uniformly bounded on $\mathrm{spt}\gamma$. This proves that $\mathcal{C}[\gamma] < +\infty$. By Proposition 3.2 there is at least one optimal measure $\gamma_o \in \Gamma(\mu, \nu)$ and there is a c-cyclically monotone set $S \subset \mathbf{R}^{d+1} \times \mathbf{R}^{d+1}$ containing the support of all optimal measures in $\Gamma(\mu, \nu)$. By Remark 4.4 the support of γ_o and hence S contains pairs $(\mathbf{x}_o, \mathbf{y}_o)$, $(\bar{\mathbf{x}}_o, \bar{\mathbf{y}}_o)$ such that $\bar{\mathbf{x}}_o, \bar{\mathbf{y}}_o \notin \{\mathbf{x}_o, \mathbf{y}_o\}$. Lemma 4.3 ensures existence of $\delta > 0$ and a c-concave function ψ such that S is contained in the c-superdifferential of ψ and

$$\psi(\mathbf{x}) = \inf_{\mathbf{y}}\left\{c(\mathbf{x}, \mathbf{y}) - \psi^c(\mathbf{y}) \mid |\mathbf{x} - \mathbf{y}| \geq \delta\right\}, \qquad (\mathbf{x} \in \mathbf{X}). \tag{39}$$

Clearly ψ is continuous and so is ψ^c. Hence, $\partial^c\psi(\mathbf{x}), \partial^c\psi^c(\mathbf{y}) \neq \emptyset$ for all $\mathbf{x}, \mathbf{y} \in \mathbf{S^d}$ and

$$(\mathbf{x}, \mathbf{y}) \in \partial^c\psi \iff \psi(\mathbf{x}) + \psi^c(\mathbf{y}) = c(\mathbf{x}, \mathbf{y}). \tag{40}$$

We use again (39) and the assumption that $l \in C^2(0, \infty)$ to conclude that ψ is semiconcave on $\mathbf{S^d}$ and so, it is differentiable everywhere except for a set $N \subset \mathbf{S^d}$ which is $(d-1)$-rectifiable. $\qquad \square$

Proof of (ii): existence and uniqueness of T_o; $\gamma_o = (\mathbf{id} \times T_o)_{\#}\mu$; uniqueness of γ_o. Let $\mathbf{x} \in \mathbf{S^d} \setminus N$ and $\mathbf{y} \in \partial^c \psi(\mathbf{x})$. We use the fact that the function $\mathbf{t} \to \psi(\mathbf{t}) + \psi^c(\mathbf{y}) - c(\mathbf{t}, \mathbf{y})$ is differentiable and attains its maximum at $\mathbf{x}$ to conclude that

$$\nabla_{\mathbf{S^d}}\psi(\mathbf{x}) = \nabla^{\mathbf{x}}_{\mathbf{S^d}}c(\mathbf{x}, \mathbf{y}). \tag{41}$$

This, together with the Lemma 4.5, implies that $\mathbf{y} = M(\nabla_{\mathbf{S^d}}\psi(\mathbf{x}), \mathbf{x})$. Since

$$|\mathbf{x} - \mathbf{y}| \geq \delta, \tag{42}$$

we obtain that $(\nabla_{\mathbf{S^d}}\psi(\mathbf{x}), \mathbf{x}) \in K_\delta$. Hence the map T_o defined by

$$T_o\mathbf{x} := M(\nabla_{\mathbf{S^d}}\psi(\mathbf{x}), \mathbf{x}) \tag{43}$$

is a Borel map. Note that by (42), the map T_o satisfies

$$|T_o\mathbf{x} - \mathbf{x}| \geq \delta$$

for all $\mathbf{x} \in \mathbf{S^d} \setminus N$.

Because μ vanishes on N we have that $\gamma_o[N \times \mathbf{S^d}] = 0$ and so, T_o is defined μ almost everywhere. Note that we have proved that

$$\partial^c\psi \setminus (N \times \mathbf{S^d}) \subset \mathrm{graph}T_o. \tag{44}$$

Since $\gamma_o[N \times \mathbf{S^d}] = 0$ and $\gamma_o \in \Gamma(\mu, \nu)$, we obtain that

$$\int L(\mathbf{x}, \mathbf{y})\mathrm{d}\gamma(\mathbf{x}, \mathbf{y}) = \int L(\mathbf{x}, T_o\mathbf{x})\mathrm{d}\gamma_o(\mathbf{x}, \mathbf{y}) = \int_{\mathbf{S^d}} L(\mathbf{x}, T_o\mathbf{x})\mathrm{d}\mu(\mathbf{x}), \tag{45}$$

for all $L \in C(\mathbf{R}^{d+1} \times \mathbf{R}^{d+1})$. By (45) we have that

$$\gamma_o = (\mathbf{id} \times T_o)_{\#}\mu, \quad T_{\#}\mu = \nu.$$

This proves that γ_o is uniquely determined. We use (45) with $c = L$ to obtain that $\mathcal{C}[\gamma_o] = \mathcal{I}[T_o]$. By (16) we conclude that T_o minimizes $\mathcal{I}$ over $\mathcal{T}(\mu, \nu)$.

Furthermore, if T_1 is another minimizer of $\mathcal{I}$ over $\mathcal{T}(\mu, \nu)$ then $\gamma_o = (\mathbf{id} \times T_1)_{\#}\mu$ and $(\mathbf{x}, T_1\mathbf{x}) \in \mathrm{spt}\gamma_o \subset \partial^c\psi$ for μ almost every $\mathbf{x}$. By (44) we have that $T_1(\mathbf{x}) = T_o\mathbf{x}$ for μ almost every $\mathbf{x}$. This proves that T_o is uniquely determined. $\square$

Proof of (iii): invertibility of the map T_o. The analogue of (39) for ψ^c gives that ψ^c is semiconcave and so, the set $\tilde{N}$ where ψ^c is not differentiable is $(d-1)$-rectifiable. Substituting μ by ν, the above reasoning yields that the map

$$S_o\mathbf{y} = M(\nabla_{\mathbf{S^d}}\psi^c(\mathbf{y}), \mathbf{y}), \qquad (\mathbf{y} \in \mathbf{S^d} \setminus \tilde{N}),$$

is such that $S_{o\#}\nu = \mu$, and

$$\partial^c\psi^c \setminus (\tilde{N} \times \mathbf{S^d}) \subset \mathrm{graph}S_o. \tag{46}$$

We use (44), (46) and the fact that $\partial^c\psi = \partial^c\psi^c$ to conclude that $\mathbf{id} = S_o \circ T_o$ on $\mathbf{S^d} \setminus (N \cup S_o^{-1}(\tilde{N}))$. Since $\tilde{N}$ is $(d-1)$-rectifiable and $S_{o\#}\nu = \mu$ we have that

$$\mu[S_o^{-1}(\tilde{N})] = \nu[\tilde{N}] = 0.$$

Thus, T_o is invertible on $\mathbf{S^d}$ up to a set whose μ measure is null. $\square$

Remark 4.7. If it is assumed only that $l \in C^1(0, \infty)$ and μ is absolutely continuous with respect to the standard measure on $\mathbf{S^d}$ then it is easy to see that ψ, ψ^c are differentiable a.e. on $\mathbf{S^d}$. Indeed, it follows from (39), (40) and the assumption that $l \in C^1(0, \infty)$ that ψ, ψ^c are locally Lipschitz. Then by Rademacher's theorem ψ, ψ^c are differentiable a.e. Note, however, that this result is weaker than the one established in Theorem 4.6.

References

[1] T. Abdellaoui and H. Heinich, Sur la distance de deux lois dans le cas vectoriel. *C.R. Acad. Sci. Paris Sér. I Math.* **319** (1994) 397–400.

[2] N. Ahmad, *The geometry of shape recognition via the Monge-Kantorovich optimal transport problem.* Ph.D. dissertation (2004).

[3] L.A. Caffarelli and V.I. Oliker, Weak solutions of one inverse problem in geometric optics. Preprint (1994).

[4] L.A. Caffarelli and S. Kochengin and V.I. Oliker, On the numerical solution of the problem of reflector design with given far-field scattering data. *Cont. Math.* **226** (1999) 13–32.

[5] W. Gangbo, Quelques problèmes d'analyse non convexe. *Habilitation à diriger des recherches en mathématiques.* Université de Metz (Janvier 1995).

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

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

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

[9] T. Glimm and V.I. Oliker, Optical design of two-reflector systems, the Monge-Kantorovich mass transfer problem and Fermat's principle. *Indiana Univ. Math. J.* **53** (2004) 1255–1278.

[10] Pengfei Guan and Xu-Jia Wang, On a Monge-Ampère equation arising in geometric optics *J. Differential Geometry* **48** (1998) 205–223.

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

[12] B.E. Kinber, On two reflector antennas. *Radio Eng. Electron. Phys.* **7** (1962) 973–979.

[13] M. Knott and C.S. Smith, On the optimal mapping of distributions. *J. Optim. Theory Appl.* **43** (1984) 39–49.

[14] E. Newman and V.I. Oliker, Differential-geometric methods in design of reflector antennas. *Symposia Mathematica* **35** (1994) 205–223.

[15] A.P. Norris and B.S. Westcott, Computation of reflector surfaces for bivariate beamshaping in the elliptic case. *J. Phys. A: Math. Gen* **9** (1976) 2159–2169.

[16] V.I. Oliker and P. Waltman, Radially symmetric solutions of a Monge-Ampere equation arising in a reflector mapping problem. *Proc. UAB Int. Conf. on Diff. Equations and Math. Physics*, edited by I. Knowles and Y. Saito, Springer. *Lect. Notes Math.* **1285** (1987) 361–374.

[17] V.I. Oliker, On the geometry of convex reflectors. *PDE's, Submanifolds and Affine Differential Geometry, Banach Center Publications* **57** (2002) 155–169.

[18] R.T. Rockafellar, *Convex Analysis.* Princeton University Press, Princeton (1970).

[19] L. Rüschendorf, On c-optimal random variables. *Appl. Stati. Probab. Lett.* **27** (1996) 267–270.

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

[21] X.-J. Wang, On design of a reflector antenna. *Inverse Problems* **12** (1996) 351–375.

[22] X.-J. Wang, On design of a reflector antenna II. *Calculus of Variations and PDE's* **20** (2004) 329–341.

[23] B.S. Westcott, *Shaped Reflector Antenna Design.* Research Studies Press, Letchworth, UK (1983).

[24] S.T. Yau, Open problems in geometry, in *Differential Geometry. Part 1: Partial Differential Equations on Manifolds* (Los Angeles, 1990), R. Greene and S.T. Yau Eds., *Proc. Sympos. Pure Math., Amer. Math. Soc.* **54** (1993) 1–28.