THE MONGE PROBLEM IN GEODESIC SPACES

STEFANO BIANCHINI AND FABIO CAVALLETTI

Abstract. We address the Monge problem in metric spaces with a geodesic distance: (X, d) is a Polish non branching geodesic space. We show that we can reduce the transport problem to 1-dimensional transport problems along geodesics. We introduce an assumption on the transport problem π which implies that the conditional probabilities of the first marginal on each geodesic are continuous. It is known that this regularity is sufficient for the construction of an optimal transport map.

1. Introduction

This paper concerns the Monge transportation problem in geodesic spaces, i.e. metric spaces with a geodesic structure. Given two Borel probability measures $\mu, \nu \in \mathcal{P}(X)$, where (X, d) is a locally compact Polish space, i.e. a separable complete locally compact metric space, we study the minimization of the functional

$$\mathcal{I}(T) = \int d(x, T(x))\mu(dy)$$

where T varies over all Borel maps $T : X \to X$ such that $T_\sharp \mu = \nu$ and d is a distance that makes (X, d) a non branching geodesic space.

Before giving an overview of the paper and of the existence result, we recall which are the main results concerning the Monge problem.

In the original formulation given by Monge in 1781 the problem was settled in $\mathbb{R}^n$, with the cost given by the Euclidean norm and the measures μ, ν were supposed to be absolutely continuous and supported on two disjoint compact sets. The original problem remained unsolved for a long time. In 1978 Sudakov [13] claimed to have a solution for any distance cost function induced by a norm: an essential ingredient in the proof was that if $\mu \ll \mathcal{L}^d$ and $\mathcal{L}^d$-a.e. $\mathbb{R}^d$ can be decomposed into convex sets of dimension k, then the conditional probabilities are absolutely continuous with respect to the $\mathcal{H}^k$ measure of the correct dimension. But it turns out that when $d > 2$, $0 < k < d-1$ the property claimed by Sudakov is not true. An example with $d = 3$, $k = 1$ can be found in [11] and [1].

The Euclidean case has been correctly solved only during the last decade. L. C. Evans and W. Gangbo in [8] solved the problem under the assumptions that $\operatorname{spt}\mu \cap \operatorname{spt}\nu = \emptyset$, $\mu, \nu \ll \mathcal{L}^d$ and their densities are Lipschitz function with compact support. The first existence results for general absolutely continuous measures μ, ν with compact support have been independently obtained by L. Caffarelli, M. Feldman and R.J. McCann in [5] and by N. Trudinger and X.J. Wang in [14]. Afterwards M. Feldman and R.J. McCann [9] extended the results to manifolds with geodesic cost. The case of a general norm as cost function on $\mathbb{R}^d$, including also the case with non strictly convex unitary ball, has been solved first in the particular case of crystalline norm by L. Ambrosio, B. Kirchheim and A. Pratelli in [1], and then in fully generality independently by L. Caravenna in [6] and by T. Champion and L. De Pascale in [7].

1.1. **Overview of the paper.** The presence of 1-dimensional sets (the geodesics) along which the cost is linear is a strong degeneracy for transport problems. This degeneracy is equivalent to the following problem in $\mathbb{R}$: if μ is concentrated on $(-\infty, 0]$, and ν is concentrated on $[0, +\infty)$, then every transference plan is optimal for the 1-dimensional distance cost $|\cdot|$. In fact, every $\pi \in \Pi(\mu, \nu)$ is supported on the set $(-\infty, 0] \times [0, +\infty)$, on which $|x - y| = y - x$ and thus

$$\int |x - y|\pi(dxdy) = -\int x\mu(dx) + \int y\nu(dy).$$

Nevertheless, for this easy case an explicit map $T : \mathbb{R} \to \mathbb{R}$ can be constructed if μ is non atomic: the easiest choice is the monotone map, a minimizer of the quadratic cost $|\cdot|^2$.

The approach suggested by the above simple case is the following:

(1) reduce the problem to transportation problems along distinct geodesics;
(2) show that the disintegration of the marginal μ on each geodesic is continuous;
(3) find a transport map on each geodesic and piece them together.

While the last point can be seen as an application of selection principles in Polish spaces, the first two points are more subtle.

The geodesics used by a given transference plan π to transport mass can be obtained from a set Γ on which π is concentrated. If π wants to be a minimizer, then it certainly chooses the shortest paths: however the metric space can be branching, i.e. geodesics can bifurcate. In this paper we assume that *the space is non branching.*

Under this assumption, a cyclically monotone plan π yields a natural partition R of the transport set $\mathcal{T}_e$, i.e. the set of the geodesics used by π:

- the set $\mathcal{T}$ made of inner points of geodesics,
- the set $a \cup b := \mathcal{T}_e \setminus \mathcal{T}$ of initial points a and end points b.

The non branching assumption and the cyclical monotonicity of Γ imply that the geodesics used by π are a partition on $\mathcal{T}$, but no other conditions can be obtained on $a \cup b$: one can think to the one dimensional torus $\mathbb{T}^1$ with $\mu = \delta_0$ and $\nu = \delta_{1/2}$. We note here that π gives also a direction along each component of R, as the one dimensional example above shows.

Even if we have a natural partition R in $\mathcal{T}$ and $\mu(a \cup b) = 0$, we cannot reduce the transport problem to one dimensional problems: a necessary and sufficient condition is that the disintegration of the measure μ is strongly consistent, which is equivalent to the fact that there exists a μ-measurable quotient map $f : \mathcal{T} \to \mathcal{T}$. In this case, one can write

$$m := f_\sharp \mu, \quad \mu = \int \mu_y m(dy), \quad \mu_y(f^{-1}(y)) = 1,$$

i.e. the conditional probabilities μ_y are concentrated on the counterimages $f^{-1}(y)$ (which are single geodesics). At this point we can obtain the one dimensional problems by partition π w.r.t. the partition $R \times (X \times X)$,

$$\pi = \int \pi_y m(dy), \quad \nu = \int \nu_y m(dy) \quad \nu_y := (P_2)_\sharp \pi_y,$$

and considering the one dimensional problems along the geodesic $R(y)$ with marginals μ_y, ν_y and cost $|\cdot|$, the length on the geodesic. At this point we can study the problem of the regularity of the conditional probabilities μ_y.

The existence of a strongly consistent disintegration relies only on the properties of geodesics in Polish spaces. Moreover, a natural operation on sets can be considered: the translation along geodesics. If A is a subset of $\mathcal{T}$, we denote by A_t the set translated by t in the direction determined by π.

It turns out that the fact that $\mu(a \cup b) = 0$ and the measures μ_y are continuous depends on how the function $t \mapsto \mu(A_t)$ behaves. We can now state the main result.

Theorem 1.1 (Lemma 5.3 and Proposition 5.4)**.** *If $\sharp\{t > 0 : \mu(A_t) > 0\}$ is uncountable for all A Borel such that $\mu(A) > 0$, then $\mu(a \cup b) = 0$ and m-a.e. conditional probability μ_y is continuous.*

This is sufficient to solve the Monge problem, i.e. to find a transport map which has the same cost as π. For a more general setting we refer to [3].

2. Preliminaries

In this section we recall some general facts about projective classes, the Disintegration Theorem for measures, measurable selection principles, geodesic spaces and optimal transportation problems.

2.1. Borel, projective and universally measurable sets.

The *projective class* $\Sigma^1_1(X)$ is the family of subsets A of the Polish space X for which there exist Y Polish and $B \in \mathcal{B}(X \times Y)$ such that $A = P_1(B)$. The *coprojective class* $\Pi^1_1(X)$ is the complement in X of the class $\Sigma^1_1(X)$. The class Σ^1_1 is called *the class of analytic sets*, and Π^1_1 are the *coanalytic sets.*

We will denote by $\mathcal{A}$ the σ-algebra generated by Σ^1_1.

We recall that a subset of X Polish is *universally measurable* if it belongs to all completed σ-algebras of all Borel measures on X: it can be proved that every set in $\mathcal{A}$ is universally measurable.

2.2. **Disintegration of measures.** Given a measurable space $(R,\mathscr{R})$ and a function $r : R \to S$, with S generic set, we can endow S with the *push forward σ-algebra $\mathscr{S}$* of $\mathscr{R}$:

$$Q \in \mathscr{S} \iff r^{-1}(Q) \in \mathscr{R},$$

which could be also defined as the biggest σ-algebra on S such that r is measurable. Moreover, given a measure space $(R,\mathscr{R},\rho)$, the *push forward measure* η is then defined as $\eta := (r_\sharp \rho)$.

Consider a probability space $(R,\mathscr{R},\rho)$ and its push forward measure space $(S,\mathscr{S},\eta)$ induced by a map r. From the above definition the map r is clearly measurable and inverse measure preserving.

Definition 2.1. A disintegration of ρ consistent with r is a map $\rho : \mathscr{R} \times S \to [0,1]$ such that

(1) $\rho_s(\cdot)$ is a probability measure on $(R,\mathscr{R})$, for all $s \in S$,
(2) $\rho_\cdot(B)$ is η-measurable for all $B \in \mathscr{R}$,

and satisfies for all $B \in \mathscr{R}, C \in \mathscr{S}$ the consistency condition

$$\rho\left(B \cap r^{-1}(C)\right) = \int_C \rho_s(B)\eta(ds).$$

A disintegration is strongly consistent with r if for all s we have $\rho_s(r^{-1}(s)) = 1$.

We say that a σ-algebra $\mathcal{A}$ is *essentially countably generated* with respect to a measure m, if there exists a countably generated σ-algebra $\hat{\mathcal{A}}$ such that for all $A \in \mathcal{A}$ there exists $\hat{A} \in \hat{\mathcal{A}}$ such that $m(A \vartriangle \hat{A}) = 0$.

We recall the following version of the theorem of disintegration of measure that can be found on [10], Section 452.

Theorem 2.2 (Disintegration of measure)**.** *Assume that $(R,\mathscr{R},\rho)$ is a countably generated probability space, $R = \cup_{s\in S} R_s$ a decomposition of R, $r : R \to S$ the quotient map and $(S,\mathscr{S},\eta)$ the quotient measure space. Then $\mathscr{S}$ is essentially countably generated w.r.t. η and there exists a unique disintegration $s \to \rho_s$ in the following sense: if ρ_1, ρ_2 are two consistent disintegration then $\rho_{1,s}(\cdot) = \rho_{2,s}(\cdot)$ for $\eta-$a.e. s.*

If $\{S_n\}_{n\in\mathbb{N}}$ is a family essentially generating $\mathscr{S}$ define the equivalence relation:

$$s \sim s' \iff \{s \in S_n \iff s' \in S_n,\ \forall\, n \in \mathbb{N}\}.$$

Denoting with p the quotient map associated to the above equivalence relation and with $(L,\mathscr{L},\lambda)$ the quotient measure space, the following properties hold:

- *$R_l := \cup_{s\in p^{-1}(l)} R_s = (p \circ r)^{-1}(l)$ is ρ-measurable and $R = \cup_{l\in L} R_l$;*
- *the disintegration $\rho = \int_L \rho_l \lambda(dl)$ satisfies $\rho_l(R_l) = 1$, for λ-a.e. l. In particular there exists a strongly consistent disintegration w.r.t. $p \circ r$;*
- *the disintegration $\rho = \int_S \rho_s \eta(ds)$ satisfies $\rho_s = \rho_{p(s)}$, for η-a.e. s.*

In particular we will use the following corollary.

Corollary 2.3. *If $(S,\mathscr{S}) = (X,\mathcal{B}(X))$ with X Polish space, then the disintegration is strongly consistent.*

2.3. **Selection principles.** Given a multivalued function $F : X \to Y$, X, Y metric spaces, the *graph* of F is the set

$$\operatorname{graph}(F) := \big\{(x,y) : y \in F(x)\big\}. \tag{2.1}$$

The *inverse image* of a set $S \subset Y$ is defined as:

$$F^{-1}(S) := \big\{x \in X \ :\ F(x) \cap S \neq \emptyset\big\}. \tag{2.2}$$

For $F \subset X \times Y$, we denote also the sets

$$F_x := F \cap \{x\} \times Y, \quad F^y := F \cap X \times \{y\}. \tag{2.3}$$

In particular, $F(x) = P_2(\operatorname{graph}(F)_x)$, $F^{-1}(y) = P_1(\operatorname{graph}(F)^y)$. We denote by F^{-1} the graph of the inverse function

$$F^{-1} := \big\{(x,y) : (y,x) \in F\big\}. \tag{2.4}$$

We say that F is *$\mathcal{R}$-measurable* if $F^{-1}(B) \in \mathcal{R}$ for all B open. We say that F is *strongly Borel measurable* if inverse images of closed sets are Borel. A multivalued function is called *upper-semicontinuous* if the preimage of every closed set is closed: in particular u.s.c. maps are strongly Borel measurable.

In the following we will not distinguish between a multifunction and its graph. Note that the *domain of* F (i.e. the set $P_1(F)$) is in general a subset of X. The same convention will be used for functions, in the sense that their domain may be a subset of X.

Given $F \subset X \times Y$, a *section* u *of* F is a function from $P_1(F)$ to Y such that $\mathrm{graph}(u) \subset F$. A *cross-section of the equivalence relation* E is a set $S \subset E$ such that the intersection of S with each equivalence class is a singleton. We recall that a set $A \subset X$ is saturated for the equivalence relation $E \subset X \times X$ if $A = \cup_{x \in A} E(x)$.

We recall the following selection principle, Theorem 5.2.1 of [12].

Theorem 2.4. *Let* Y *be a Polish space,* X *a nonempty set, and* $\mathcal{L}$ *a* σ*-algebra of subset of* X*. Every* $\mathcal{L}$*-measurable, closed value multifunction* $F : X \to Y$ *admits an* $\mathcal{L}$*-measurable selection.*

2.4. **Metric setting.** In this section we refer to [4].

Definition 2.5. A length structure on a topological space X is a class $\mathtt{A}$ of admissible paths, which is a subset of all continuous paths in X, together with a map $L : \mathtt{A} \to [0, +\infty]$: the map L is called length of path.

The class $\mathtt{A}$ and the map L must satisfy the natural assumptions that one expects (for shortness we do not list them here).

Given a length structure, we can define a distance

$$d_L(x, y) = \inf \Big\{ L(\gamma) : \gamma : [a, b] \to X, \gamma \in A, \gamma(a) = x, \gamma(b) = y \Big\},$$

that makes (X, d_L) a metric space (allowing d_L to be $+\infty$). The metric d_L is called *intrinsic*.

Definition 2.6. A length structure is said to be *complete* if for every two points x, y there exists an admissible path joining them whose length $L(\gamma)$ is equal to $d_L(x, y) < +\infty$.

In other words, a length structure is complete if there exists a shortest path between two points with finite length.

Intrinsic metrics associated with complete length structure are said to be *strictly intrinsic*. The metric space (X, d) with d strictly intrinsic is called a *geodesic space*. A curve whose length equals the distance between its end points is called *geodesic*.

It follows from Proposition 2.5.9 of [4] that every admissible curve of finite length admits a constant speed parametrization, i.e. γ defined on $[0, 1]$ and $L(\gamma \llcorner [t, t']) = v(t' - t)$, with v velocity. Hence from now on geodesics when parametrized are understood as constant speed geodesics.

Definition 2.7. Let (X, d_L) be a metric space. The distance d_L is said to be *strictly convex* if, for all $r \geq 0$, $d_L(x, y) = r/2$ implies that

$$\{z : d_L(x, z) = r\} \cap \{z : d_L(y, z) = r/2\}$$

is a singleton.

The definition can be restated as: geodesics cannot branch in the interior. More precisely: let $\gamma_1, \gamma_2 : [0, 1] \to X$ be geodesics such that $\gamma_1(I_1) = \gamma_2(I_2)$ for some intervals $I_1, I_2 \subset [0, 1]$. Then either $\gamma_1 \subset \gamma_2$ or $\gamma_2 \subset \gamma_1$. An equivalent requirement is that if $\gamma_1(0) = \gamma_2(0)$, $\gamma_1(1) = \gamma_2(1)$, then such geodesics do not admit a geodesic extension i.e. they are not a part of a longer one. The metric space (X, d) is said non-branching.

From now on

$$(X, d) \text{ is a non-branching geodesic locally compact Polish space.}$$

2.5. **General facts about optimal transportation.** Let (X, Ω, μ) and (Y, Σ, ν) be two probability spaces and $c : X \times Y \to \mathbb{R}^+$ be a $\Omega \times \Sigma$ measurable function. Consider the set of transference plans

$$\Pi(\mu, \nu) := \Big\{ \pi \in \mathcal{P}(X \times Y) : (P_1)_\sharp \pi = \mu, (P_2)_\sharp \pi = \nu \Big\},$$

where $P_i(x_1, x_2) = x_i, i = 1, 2$ and $\mathcal{P}(X \times Y)$ is the set of probability measure on $X \times Y$. Define the functional

$$\begin{array}{rcccl} \mathcal{I} & : & \Pi(\mu, \nu) & \longrightarrow & \mathbb{R}^+ \\ & & \pi & \longmapsto & \mathcal{I}(\pi) := \displaystyle\int_{X \times Y} c\pi. \end{array} \tag{2.5}$$

The *Monge-Kantorovich minimization problem* is to find the minimum of $\mathcal{I}$ over all transference plans.

If we consider a map $T : X \to Y$ such that $T_\sharp \mu = \nu$, the functional 2.5 becomes

$$\mathcal{I}(T) := \mathcal{I}((Id \times T)_\sharp \mu) = \int_X c(x, T(x))\mu(dx).$$

The minimum problem over all T is called *Monge minimization problem.*

The Kantorovich problem admits a (pre) dual formulation: before stating it, we give a definition.

Definition 2.8. A set $\Gamma \subset X \times Y$ is said to be *c-cyclically monotone* if, for any $n \in \mathbb{N}$ and for any family $(x_1, y_1), \dots, (x_n, y_n)$ of points of Γ, the following inequality holds

$$\sum_{i=0}^{n} c(x_i, y_i) \leq \sum_{i=0}^{n} c(x_{i+1}, y_i),$$

with $x_{n+1} = x_1$. A transference plan is said to be *c-cyclically monotone* if it is concentrated on a c-cyclically monotone set.

Consider the set

$$\Phi_c := \Big\{(\varphi, \psi) \in L^1(\mu) \times L^1(\nu) : \varphi(x) + \psi(y) \leq c(x, y)\Big\}. \tag{2.6}$$

Define for all $(\varphi, \psi) \in \Phi_c$ the functional

$$J(\varphi, \psi) := \int \varphi\mu + \int \psi\nu. \tag{2.7}$$

The following is a well known result (see Theorem 5.10 of [16]).

Theorem 2.9 (Kantorovich Duality)**.** *Let X and Y be Polish spaces, let $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$, and let $c : X \times Y \to \mathbb{R}^+ \cup \{+\infty\}$ be lower semicontinuous. Then the following holds:*

(1) *Kantorovich duality:*

$$\inf_{\pi \in \Pi[\mu,\nu]} \mathcal{I}(\pi) = \sup_{(\varphi,\psi)\in\Phi_c} J(\varphi, \psi).$$

Moreover, the infimum on the left-hand side is attained and the right-hand side is also equal to

$$\sup_{(\varphi,\psi)\in\Phi_c\cap C_b} J(\varphi, \psi),$$

where $C_b = C_b(X, \mathbb{R}) \times C_b(Y, \mathbb{R})$.

(2) *If c is real valued and the optimal cost*

$$C(\mu, \nu) := \inf_{\pi\in\Pi(\mu,\nu)} I(\pi)$$

is finite, then there is a measurable c-cyclically monotone set $\Gamma \subset X \times Y$, closed if c is continuous, such that for any $\pi \in \Pi(\mu, \nu)$ the following statements are equivalent:

(a) *π is optimal;*

(b) *π is c-cyclically monotone;*

(c) *π is concentrated on Γ;*

(d) *there exists a c-concave function φ such that π-a.s. $\varphi(x) + \varphi^c(y) = c(x, y)$.*

(3) *If moreover*

$$c(x, y) \leq c_X(x) + c_Y(y), \qquad (c_X, c_Y) \in L^1(\mu) \times L^1(\nu),$$

then there exist a couple of potentials and the optimal transference plan π is concentrated on the set

$$\big\{(x, y) \in X \times Y \,|\, \varphi(x) + \psi(y) = c(x, y)\big\}.$$

Finally if $(c_X, c_Y) \in \mathcal{L}^1(\mu) \times \mathcal{L}^1(\nu)$ then the supremum is attained

$$\sup_{\Phi_c} J = J(\varphi, \varphi^c).$$

We recall also that if c is Borel, then every optimal transference plan π is concentrated on a c-cyclically monotone set [2].

3. Optimal transportation in geodesic spaces

Take $\mu, \nu \in \mathcal{P}(X)$ and consider the optimal transportation problem with cost $c(x,y) = d(x,y)$. In our setting the following holds.

(1) It is possible to restrict the Kantorovich duality just to 1-Lipschitz functions.
(2) For a 1-Lipschitz map the d-transform has a particular form $\varphi^d(x) = -\varphi(x)$.
(3) It follows that the support of the minimizing measures is the *transport set*

$$\Gamma := \Big\{(x,y) \in X \times X : \varphi(x) - \varphi(y) = d(x,y)\Big\}, \tag{3.1}$$

for any potential $\varphi \in \mathrm{Lip}_1(X)$.
(4) The distance cost allows to assume $\mu \perp \nu$ because of the triangular inequality.

In this section we study the transport set Γ. Note that Γ is closed, hence σ-compact.

Definition 3.1 (Transport rays)**.** Define for $x \in X$ the outgoing transport ray

$$G(x) := \big\{y \in X : \varphi(x) - \varphi(y) = d(x,y)\big\} \tag{3.2}$$

and the incoming transport ray

$$G^{-1}(x) := \big\{y \in X : \varphi(y) - \varphi(x) = d(x,y)\big\}. \tag{3.3}$$

Define the set of transport rays through x as the set

$$R(x) := G(x) \cup G^{-1}(x). \tag{3.4}$$

Observe that the multivalued maps G, G^{-1}, and R have σ-compact graph.

Definition 3.2. Define the transport sets

$$\mathcal{T} := P_1\big(\mathrm{graph}(G^{-1}) \setminus \{x = y\}\big) \cap P_1\big(\mathrm{graph}(G) \setminus \{x = y\}\big), \tag{3.5}$$

$$\mathcal{T}_e := P_1\big(\mathrm{graph}(G^{-1}) \setminus \{x = y\}\big) \cup P_1\big(\mathrm{graph}(G) \setminus \{x = y\}\big). \tag{3.6}$$

From the definition of G it is fairly easy to prove that $\mathcal{T}$, $\mathcal{T}_e$ are σ-compact sets. The subscript e refers to the endpoints of the geodesics: clearly we have

$$\mathcal{T}_e = P_1(R \setminus \{x = y\}). \tag{3.7}$$

Since $\pi(\Gamma) = 1$, it is fairly easy to prove that $\pi(\mathcal{T}_e \times \mathcal{T}_e \cup \{x = y\}) = 1$. As a consequence, $\mu(\mathcal{T}_e) = \nu(\mathcal{T}_e)$ and any maps T such that for $\nu \llcorner_{\mathcal{T}_e} = T_\sharp \mu \llcorner_{\mathcal{T}_e}$ can be extended to a map T' such that $\nu = T_\sharp \mu$ with the same cost by setting

$$T'(x) = \begin{cases} T(x) & x \in \mathcal{T}_e \\ x & x \notin \mathcal{T}_e \end{cases} \tag{3.8}$$

Therefore we have only to study the Monge problem in $\mathcal{T}_e$.

Remark 3.3. *Take $y \in G(x)$, and take the points z such that*

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

From the definition of G, $z \in G(x)$ and $y \in G(z)$ or, equivalently, $z \in G^{-1}(y)$. Similarly if we take $y \in G(x)$ and $z \in G(y)$ we get $z \in G(x)$ and y is on the geodesic from x to z. So we can say that if $y \in G(x)$, the set

$$G(x) \cap G^{-1}(y)$$

is the union of the minimizing geodesic connecting x to y. Therefore, by the non-branching assumption, if $x \in \mathcal{T}$, then $R(x)$ is a single geodesic.

Definition 3.4. Define the multivalued endpoint maps $a, b : \mathcal{T}_e \to \mathcal{T}_e$ by:

$$a(y) := \big\{z \in G^{-1}(y) : \nexists x \in X \setminus \{z\}, x \in G^{-1}(z)\big\}, \tag{3.9}$$

$$b(y) := \big\{z \in G(y) : \nexists x \in X \setminus \{z\}, x \in G(z)\big\}. \tag{3.10}$$

We call $a(\mathcal{T}_e)$ the set of initial points and $b(\mathcal{T}_e)$ the set of final points.

Observe that the multivalued maps $a, b : \mathcal{T}_e \longrightarrow \mathcal{T}_e$ have σ-compact graph. Other properties of the end point maps:

(1) $a(x)$, $b(x)$ are singleton or empty when $x \in \mathcal{T}$;
(2) $a(\mathcal{T}) = a(\mathcal{T}_e)$, $b(\mathcal{T}) = b(\mathcal{T}_e)$;
(3) $\mathcal{T}_e = \mathcal{T} \cup a(\mathcal{T}) \cup b(\mathcal{T})$, $\mathcal{T} \cap (a(\mathcal{T}) \cup b(\mathcal{T})) = \emptyset$.

W.l.o.g. we can assume that the μ-measure of final points and the ν-measure of the initial points are 0.

4. Partition of the transport set

Let $\{x_i\}_{i\in\mathbb{N}}$ be a dense sequence in (X, d).

Lemma 4.1. *The sets*

$$\mathcal{Z}_{ijk} := \Big\{ x \in \mathcal{T} \cap \bar{B}_{2^{-j}}(x_i) : L(G(x)), L(G^{-1}(x)) \geq 2^{2-k} \Big\}$$

form a countable covering of $\mathcal{T}$ of σ-compact sets.

Proof. We first prove the measurability. We consider separately the conditions defining Z_{ijk}.

Point 1. The set

$$A_{ij} := \mathcal{T} \cap \bar{B}_{2^{-j}}(x_i)$$

is clearly σ-compact.

Point 2. The set

$$B_k := \big\{ x \in \mathcal{T} : L(G(x)) \geq 2^{2-k} \big\} = P_1\Big(G \cap \big\{ d(x,y) \geq 2^{2-k} \big\} \Big)$$

is σ-compact, being the projection of a σ-compact set. Similarly, the set

$$C_k := \big\{ x \in \mathcal{T} : L(G^{-1}(x)) \geq 2^{2-k} \big\} = P_1\Big(G^{-1} \cap \big\{ d(x,y) \geq 2^{2-k} \big\} \Big)$$

is again σ-compact.

We finally can write

$$\mathcal{Z}_{ijk} = A_{ij} \cap B_k \cap C_k.$$

To show that it is a covering, notice that for all $x \in \mathcal{T}$ it holds

$$\min \big\{ L(G(x)), L(G^{-1}(x)) \big\} \geq 2^{2-k}$$

for some $k \in \mathbb{N}$. □

From $\mathcal{Z}_{ijk}$ we can define a countable covering of $\mathcal{T}_e$ of σ-compact saturated sets just taking

$$\mathcal{T}_{ijk} := R^{-1}(\mathcal{Z}_{ijk}).$$

In the natural way, we can find a countable partition into σ-compact saturated sets by defining

$$\mathcal{Z}_{m,e} := \mathcal{T}_{i_m j_m k_m} \setminus \bigcup_{m'=1}^{m-1} \mathcal{T}_{i_{m'} j_{m'} k_{m'}}, \quad \mathcal{Z}_{0,e} := \mathcal{T}_e \setminus \bigcup_{m\in\mathbb{N}} \mathcal{Z}_{m,e}, \tag{4.1}$$

where

$$\mathbb{N} \ni m \mapsto (i_m, j_m, k_m) \in \mathbb{N}^3$$

is a bijective map. Intersecting the above sets with $\mathcal{T}$, we obtain the countable partition of $\mathcal{T}$ in σ-compact sets

$$\mathcal{Z}_m := \mathcal{Z}_{m,e} \cap \mathcal{T}, \quad m \in \mathbb{N}_0. \tag{4.2}$$

Since

$$R = \Big\{ (x,y) \in X \times X : |\varphi(x) - \varphi(y)| = d(x,y) \Big\}$$

is the graph of an equivalence relation on $\mathcal{T}$, we use this partition to prove the strong consistency of the disintegration induced by R.

On $\mathcal{Z}_m$, $m > 0$, we define the closed values map

$$\mathcal{Z}_m \ni x \mapsto F(x) := R(x) \cap \bar{B}_{2^{-j_m}}(x_{i_m}) \subset \mathcal{Z}_m. \tag{4.3}$$

Lemma 4.2. *The equivalence relation R admits a Borel section: there exists a Borel map $f : \mathcal{T} \to \mathcal{T}$ such that*

(1) $xRf(x)$,
(2) xRy *implies* $f(x) = f(y)$.

Proof. It is enough to consider just one $\mathcal{Z}_m$.

Step 1. First we show that F has σ-compact graph:

$$\mathrm{graph}(F) = \mathcal{Z}_m \times X \cap \big(R \cap X \times B_{2^{-j_m}}(x_{i_m})\big)$$

and $F(x)$ is clearly compact.

Step 2. Let $\mathcal{L}$ the family of saturated closed sets w.r.t. F, i.e. inverse images of closed sets, and let $\mathcal{L}_\sigma$ be the smallest σ-algebra containing $\mathcal{L}$. Note that $\mathcal{L}_\sigma$ is a subset of the Borel σ-algebra. Clearly, by construction, F is $\mathcal{L}_\sigma$-measurable.

Step 3. From Theorem 2.4 there exists a $\mathcal{L}_\sigma$-measurable selection f of F. Clearly the atoms of $\mathcal{L}_\sigma$ are $R(x) \setminus \{a(x), b(x)\}$ for $x \in \mathcal{Z}_m$ and f is constant along $R(x) \setminus \{a(x), b(x)\}$ being $\mathcal{L}_\sigma$-measurable. Moreover $f(x) \in R(x) \setminus \{a(x), b(x)\}$. Hence f is a Borel section. □

The set $S = f(\mathcal{T})$ is a Borel cross-section of R restricted to $\mathcal{T}$: indeed

$$S = \{x : d(f(x), x) = 0\}$$

and $R(x) \cap S = \{f(x)\}$. Having a measurable cross-section we can define the parametrization of $\mathcal{T}$, $\mathcal{T}_e$ by geodesics.

$$S \times \mathbb{R} \ni (y, t) \mapsto g(y, t) := \big\{x : \varphi(y) - \varphi(x) = t\big\}. \tag{4.4}$$

We summarize the properties of the set graph$g$

(1) The set graph$g$ is Borel.
(2) It is the graph of a map with range $\mathcal{T}_e$.
(3) $t \mapsto g(y, t)$ is a d_L 1-Lipschitz G-order preserving for $y \in \mathcal{T}$.

5. Regularity of the disintegration

Let μ be a probability measure on (X, d). This section is divided in two parts.

In the first one we consider the translation of Borel sets by the optimal geodesic flow, we introduce the fundamental regularity assumption (Assumption 1) on the measure η and we show that an immediate consequence is that the set of initial points is negligible. A second consequence is that the disintegration of η w.r.t. the R has continuous conditional probabilities.

5.1. **Evolution of Borel sets.** Let $A \subset \mathcal{T}_e$ be an analytic set and define for $t \in \mathbb{R}$ the *t-evolution A_t of A* by

$$A_t := g\big(g^{-1}(A) + (0, t)\big). \tag{5.1}$$

Lemma 5.1. *The set A_t is analytic.*

We can show that $t \mapsto \mu(A_t)$ is measurable.

Lemma 5.2. *Let A be analytic. The function $t \mapsto \mu(A_t)$ is universally measurable.*

Proof. Since A is analytic, then $g^{-1}(A)$ is analytic, and the set

$$\tilde{A} := \big\{(y, t, \tau) : (y, t - \tau) \in g^{-1}(A)\big\}$$

is easily seen to be again analytic. From Fubini theorem applied to the measure $\mu \times \eta$, $\eta \in \mathcal{P}(\mathbb{R})$, it follows that $t \mapsto \mu((A)_t)$ is η-integrable. Since this holds for all η, by definition $t \mapsto \mu((A)_t)$ is universally measurable for $t \in \mathbb{R}$. □

The next assumption is the fundamental assumption of the paper.

Assumption 1 (Non-degeneracy assumption)**.** For all compact sets A such that $\mu(A) > 0$ the set $\{t \in \mathbb{R}^+ : \mu(A_t) > 0\}$ has cardinality $> \aleph_0$.

An immediate consequence of the Assumption 1 is that the measure μ is concentrated on $\mathcal{T}$.

Lemma 5.3. *If μ satisfies Assumption 1 then*

$$\mu(\mathcal{T}_e \setminus \mathcal{T}) = 0.$$

Proof. If $A \subset a(X)$ is compact, then $A_t \cap A_s = \emptyset$ for $0 \leq s < t$. Hence

$$\sharp\big\{t \in \mathbb{R}^+ : \mu(A_t) > 0\big\} \leq \aleph_0,$$

because of the boundedness of μ. This contradicts the assumption unless $\mu(A) = 0$. □

Once we know that $\mu(\mathcal{T}) = 1$, we can use the Disintegration Theorem 2.2 to write

$$\mu = \int_S \mu_y m(dy), \quad m = f_\sharp \mu, \ \mu_y \in \mathcal{P}(R(y)). \tag{5.2}$$

The disintegration is strongly consistent since the quotient map $f : \mathcal{T} \to \mathcal{T}$ is μ-measurable and $(\mathcal{T}, \mathcal{B}(\mathcal{T}))$ is countably generated.

The second consequence of Assumption 1 is that μ_y is continuous, i.e. $\mu_y(\{x\}) = 0$ for all $x \in X$.

Proposition 5.4. *The conditional probabilities μ_y are continuous for m-a.e. $y \in S$.*

Proof. From the regularity of the disintegration and the fact that $m(S) = 1$, we can assume that the map $y \mapsto \mu_y$ is weakly continuous on a compact set $K \subset S$ of comeasure $< \epsilon$ such that $L(R(y)) > \epsilon$ for all $y \in K$. It is enough to prove the proposition on K.

Step 1. From the continuity of $K \ni y \mapsto \mu_y \in \mathcal{P}(X)$ w.r.t. the weak topology, it follows that the map

$$y \mapsto A(y) := \big\{x \in R(y) : \mu_y(\{x\}) > 0\big\} = \cup_n \big\{x \in R(y) : \mu_y(\{x\}) \geq 2^{-n}\big\}$$

is σ-closed: in fact, if $(y_m, x_m) \to (y, x)$ and $\mu_{y_m}(\{x_m\}) \geq 2^{-n}$, then $\mu_y(\{x\}) \geq 2^{-n}$ by u.s.c. on compact sets.

Hence it is Borel, and by Lusin Theorem (Theorem 5.8.11 of [12]) it is the countable union of Borel graphs: setting in case $c_i(y) = 0$, we can consider them as Borel functions on S and order w.r.t. G,

$$\mu_{y,\mathrm{atomic}} = \sum_{i \in \mathbb{Z}} c_i(y) \delta_{x_i(y)}, \quad x_{i+1}(y) \in G(x_i(y)), \ i \in \mathbb{Z}.$$

Step 2. Define the sets

$$S_{ij}(t) := \Big\{y \in K : x_i(y) = g\big(g^{-1}(x_j(y)) + t\big)\Big\} \cap \mathcal{T}.$$

Since $K \subset S$, to define S_{ij} we are using the graph $g \cap S \times \mathbb{R} \times \mathcal{T}$, which is analytic: hence $S_{ij} \in \Sigma^1_1$.

For $A_j := \{x_j(y), y \in K\}$ and $t \in \mathbb{R}^+$ we have that

$$\begin{aligned}
\mu((A_j)_t) &= \int_K \mu_y((A_j)_t) m(dy) = \int_K \mu_{y,\mathrm{atomic}}((A_j)_t) m(dy) \\
&= \sum_{i \in \mathbb{Z}} \int_K c_i(y) \delta_{x_i(y)}\big(g(g^{-1}(x_j(y)) + t)\big) m(dy) = \sum_{i \in \mathbb{Z}} \int_{S_{ij}(t)} c_i(y) m(dy).
\end{aligned}$$

We have used the fact that $A_j \cap R(y)$ is a singleton.

Step 3. For fixed $i, j \in \mathbb{N}$, again from the fact that $A_j \cap R(y)$ is a singleton

$$S_{ij}(t) \cap S_{ij}(t') = \begin{cases} S_{ij}(t) & t = t' \\ \emptyset & t \neq t' \end{cases}$$

so that

$$\sharp\big\{t : m(S_{ij}(t)) > 0\big\} \leq \aleph_0.$$

Finally

$$\mu((A_j)_t) > 0 \quad \Longrightarrow \quad t \in \bigcup_i \big\{t : m(S_{ij}(t)) > 0\big\},$$

whose cardinality is $\leq \aleph_0$, contradicting Assumption 1. □

6. Solution to the Monge problem

In this section we show that Proposition 5.4 allows to construct an optimal map T. We recall the one dimensional result for the Monge problem that can be found on [15].

Theorem 6.1. *Let μ, ν be probability measures on $\mathbb{R}$, μ with no atoms, and let*

$$G(x) = \mu((-\infty; x)), \quad F(x) = \nu((-\infty; x)),$$

be the distribution functions of μ and ν respectively. Then

(1) *The non decreasing function $T : \mathbb{R} \to \mathbb{R} \cup \{\pm\infty\}$ defined by*

$$T(x) = \sup \{y \in \mathbb{R} : F(y) \leq G(x)\},$$

with the convention $\sup \emptyset = -\infty$, maps μ to ν. Moreover any other non decreasing map T' such that $T'_\sharp \mu = \nu$ coincides with T on the support of μ up to a countable set.

(2) *If $\phi : [0, +\infty] \to \mathbb{R}$ is non decreasing and convex, then T is an optimal transport relative to the cost $c(x, y) = \phi(|x - y|)$. Moreover T is the unique optimal transference map if ϕ is strictly convex.*

Assume that μ satisfies Assumption 1. Then we can disintegrate μ and the optimal transference plan π respect to the ray equivalence relation R and $R \times X$ as in equation 5.2,

$$\mu = \int \mu_y m(dy), \ \pi = \int \pi_y m(dy), \quad \mu_y \text{ continuous}, \ (P_1)_\sharp \pi_y = \mu_y. \tag{6.1}$$

We write moreover

$$\nu = \int \nu_y m(dy) = \int (P_2)_\sharp \pi_y m(dy). \tag{6.2}$$

Note that $\pi_y \in \Pi(\mu_y, \nu_y)$ is d-monotone (and hence optimal, because $R(y)$ is one dimensional) for m-a.e. y. If $\nu(\mathcal{T}) = 1$, then (6.2) is the disintegration of ν w.r.t. R.

Theorem 6.2. *Let $\pi \in \Pi(\mu, \nu)$ be an optimal transference plan, and assume that Assumption 1 holds. Then there exists a Borel map $T : X \to X$ with the same transport cost as π.*

Proof. By means of the map g^{-1}, we reduce to a transport problem on $S \times \mathbb{R}$, with cost

$$c((y, s), (y', t)) = \begin{cases} |t - s| & y = y' \\ +\infty & y \neq y'. \end{cases}$$

It is enough to prove the theorem in this setting under the following assumptions: S compact and $S \ni y \mapsto (\mu_y, \nu_y)$ weakly continuous. We consider here the probabilities μ_y, ν_y on $\mathbb{R}$.

Step 1. From the weak continuity of the map $y \mapsto (\mu_y, \nu_y)$, it follows that for all t the map

$$(y, t) \mapsto H(y, t) := \mu_y((-\infty, t)),$$

is continuous in t and l.s.c. in y, hence l.s.c.. Similarly, the map

$$(y, t) \mapsto F(y, t) := \nu_y((-\infty, t))$$

is easily seen to be l.s.c.. Both are clearly increasing in t.

Step 2. The map T defined as Theorem 6.1 by

$$T(y, s) := \Big(y, \sup\big\{t : F(y, t) \leq H(y, s)\big\}\Big)$$

is Borel. In fact, for A Borel,

$$T^{-1}(A \times [t, +\infty)) = \Big\{(y, s) : y \in A, H(y, s) \geq F(y, t)\Big\} \in \mathcal{B}(S \times \mathbb{R}).$$

Step 3. By the definition of the set G, it follows that along each geodesic $\mu_y(g(y, (-\infty, t))) \geq \nu_y(g(y, (-\infty, t)))$, because in the opposite case G is not d-monotone. Hence we can conclude that $T(s) \geq s$, and $c((y, s), T(y, s)) = P_2(T(y, s)) - s$. $\square$

6.1. **Final Remarks.** So from Theorem 6.2 we know that to solve the Monge problem we need Assumption 1 to hold. As it is clear from its formulation, Assumption 1 is not a direct hypothesis on the geometry of the metric space (X, d). Indeed the evolution defined in the paper is induced by the d-monotone set Γ that is deeply related to the problem.

However it is well known that in a finite dimensional manifold a lower bound on Ricci curvature implies an estimates of how the mass is moved from a point by the exponential map. Similar notion of curvature are present also in metric spaces. In [3] we prove that if the metric measure space (X, d, μ) satisfies the *measure contraction property* (MCP) then Assumption 1 is satisfied and therefore the Monge problem is solved.

A possible direction of a future research is to remove the hypothesis of strictly convexity on the distance cost d. The main problem will be that, since bifurcation of geodesics must be taken into account, the reduction to the one-dimensional case, that is the main ingredient we use to solve the Monge problem, will not hold anymore.

References

[1] L. Ambrosio, B. Kirchheim, and A. Pratelli. Existence of optimal transport maps for crystalline norms. *Duke Math. J.*, 125(2):207–241, 2004.

[2] S. Bianchini and L. Caravenna. On the extremality, uniqueness and optimality of transference plans. preprint.

[3] S. Bianchini and F. Cavalletti. The Monge problem for distance cost in geodesic spaces. *preprint*, 2009.

[4] D. Burago, Y. Burago, and S. Ivanov. *A course in metric geometry*. Graduate studies in mathematics. AMS, 2001.

[5] L. Caffarelli, M. Feldman, and R.J. McCann. Constructing optimal maps for Monge's transport problem as a limit of strictly convex costs. *J. Amer. Math. Soc.*, 15:1–26, 2002.

[6] L. Caravenna. An existence result for the Monge problem in $\mathbb{R}^n$ with norm cost function. *preprint*, 2009.

[7] T. Champion and L. De Pascale. The Monge problem in $\mathbb{R}^d$. *preprint*, 2009.

[8] L.C. Evans and W. Gangbo. Differential equations methods for the Monge-Kantorovich mass transfer problem. *Current Developments in Mathematics*, pages 65–126, 1997.

[9] M. Feldman and R. McCann. Monge's transport problem on a Riemannian manifold. *Trans. Amer. Math. Soc.*, 354:1667–1697, 2002.

[10] D. H. Fremlin. *Measure Theory*, volume 4. Torres Fremlin, 2002.

[11] D.G. Larman. A compact set of disjoint line segments in $\mathbb{R}^3$ whose end set has positive measure. *Mathematika*, 18:112–125, 1971.

[12] A. M. Srivastava. *A course on Borel sets*. Springer, 1998.

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

[14] N. Trudinger and X.J. Wang. On the Monge mass transfer problem. *Calc. Var. PDE*, 13:19–31, 2001.

[15] C. Villani. *Topics in optimal transportation*. Graduate studies in mathematics. AMS, 2003.

[16] C. Villani. *Optimal transport, old and new*. Springer, 2008.