THE MONGE PROBLEM IN WIENER SPACE

FABIO CAVALLETTI

Abstract. We address the Monge problem in the abstract Wiener space and we give an existence result provided both marginal measures are absolutely continuous with respect to the infinite dimensional Gaussian measure γ.

1. Introduction

Let $(X, \|\cdot\|)$ be an ∞-dimensional separable Banach space, $\gamma \in \mathcal{P}(X)$ be a non degenerate Gaussian measure over X and $H(\gamma)$ be the corresponding Cameron-Martin space with Hilbertian norm $\|\cdot\|_{H(\gamma)}$. Given two probability measures $\mu, \nu \in \mathcal{P}(X)$, we will prove the existence of a solution for the following Monge minimization problem

$$\min_{T: T_\sharp \mu = \nu} \int_X \|x - T(x)\|_{H(\gamma)} \mu(dx), \tag{1.1}$$

provided μ and ν are both absolutely continuous w.r.t. γ.

Before giving an overview of the paper, we recall the main results on the Monge problem.

In the original formulation given by Monge in 1781 the problem was settled in $\mathbb{R}^d$, with the cost given by the Euclidean norm and the measures μ, ν supposed to be absolutely continuous and supported on two disjoint compact sets. The original problem remained unsolved for a long time. In 1978 Sudakov [17] 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 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 [15].

The Euclidean case has been correctly solved only during the last decade. L. C. Evans and W. Gangbo in [11] solve 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 functions with compact support. The first existence results for general absolutely continuous measures μ, ν with compact support is independently obtained by L. Caffarelli, M. Feldman and R.J. McCann in [8] and by N. Trudinger and X.J. Wang in [18]. M. Feldman and R.J. McCann [12] extend 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, is solved first in the particular case of crystalline norm by L. Ambrosio, B. Kirchheim and A. Pratelli in [2], and then in full generality independently by L. Caravenna in [9] and by T. Champion and L. De Pascale in [10]. The Monge minimization problem for non-branching geodesic metric space is studied in [6], where the existence is proven for spaces satisfying a finite dimensional lower curvature bound, namely $\mathsf{MCP}(K, N)$.

1.1. Overview of the paper.

The approach to this problem is the one of [6]: assume that there exists a transference plan of finite cost, then we can

(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.

Indeed, since the cost function is lower semi-continuous, the existence of transference plan of finite cost implies the existence of an optimal transference plan. This permits to reduce the minimization problem to one dimensional minimization problems. There an explicit map can be constructed provided the first marginal measure μ is continuous (i.e. without atoms), for example choose the monotone minimizer of the quadratic cost $|\cdot|^2$. The third point is an application of selection theorems.

Date: July 27, 2011.

All this strategy has already been implemented in full generality in [6]. Thus to obtain the existence of an optimal transference map we have to show that point (2) of the strategy is fulfilled for μ and ν absolutely continuous with respect to the infinite dimensional Gaussian measure γ.

We recall the main steps of the reduction to geodesics.

The geodesics used by a given transference plan π to transport mass can be obtained from a set Γ on which π is concentrated. It is well-known that every optimal transference plan is concentrated on a $\|\cdot\|_H$-cyclically monotone set. Since the considered norm is non-branching, Γ yields a natural partition R of a subset of the transport set $\mathcal{T}_e$, i.e. the set of points on the geodesics used by π: defining

- 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 cyclical monotonicity of Γ implies that the geodesics used by π are a partition on $\mathcal{T}$. In general in a there are points from which more than one geodesic starts and in b there are points in which more than one geodesic ends, therefore the membership to a geodesic can't be an equivalence relation on the set $a \cup b$. Take as example the unit circle with $\mu = \delta_0$ and $\nu = \delta_{e^{i\pi}}$.

Even if we have a natural partition R of $\mathcal{T}$ and $\mu(a) = 0$, we cannot reduce the transport problem to one dimensional problems: a necessary and sufficient condition is the strongly consistency of the disintegration of μ. The latter is equivalent to the existence of a μ-measurable map $f : \mathcal{T} \to \mathcal{T}$ such that $(x, y) \in R$ if and only if $f(x) = f(y)$, i.e. f is a μ-measurable quotient map of the equivalence relation R.

If this is the case, then

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

i.e. the conditional probabilities μ_y are concentrated on the counterimages $f^{-1}(y)$ (which are single geodesics). In our setting the strong consistency of the disintegration of μ is a consequence of the topological properties of the geodesics of $\|\cdot\|_{H(\gamma)}$ as curves in $(X, \|\cdot\|)$. Finally we obtain the one dimensional problems by partitioning 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 the arc length on the geodesic.

At this point we can study the problem of the regularity of the conditional probabilities μ_y. A natural operation on sets can be considered: the evolution along the transport set. If A is a Borel subset of $\mathcal{T}_e$, we consider the set $T_t(\Gamma \cap A \times X)$ where T_t is the map from $X \times X$ to X that associates to a couple of points its convex combination at time t.

It turns out that the μ-negligibility of the set of initial points and the continuity of measures μ_y they both depend on the behavior of the function $t \mapsto \gamma(T_t(\Gamma \cap A \times X))$.

Theorem 1.1 (Proposition 4.1 and Proposition 4.2)**.** *If for every A with $\mu(A) > 0$ there exist a sequence $t_n \searrow 0$ and a positive constant C such that $\gamma(T_{t_n}(\Gamma \cap A \times X)) \geq C\mu(A)$, then $\mu(a) = 0$ and for m-a.e. y the conditional probabilities μ_y and ν_y are continuous.*

This result implies that the existence of a minimizer of the Monge problem is equivalent to the regularity properties of $t \mapsto \gamma(T_t(\Gamma \cap A \times X))$. Hence the problem is reduced to verify that the Gaussian measure γ satisfies the assumptions of Theorem 1.1.

Let $\mu = \rho_1 \gamma$ and $\nu = \rho_2 \gamma$ and assume that ρ_1 and ρ_2 are bounded. Then we find suitable d-dimensional measures μ_d, ν_d, absolutely continuous w.r.t. the d-dimensional Gaussian measure γ_d, converging to μ and ν respectively, such that (Theorem 6.1) γ_d verifies

$$\gamma_d(T_t(\Gamma_d \cap A \times X)) \geq C\mu_d(A)$$

where the evolution now is induced by the transport problem between μ_d and ν_d (the set Γ_d will be the graph of an optimal map between μ_d and ν_d) and the constant C does not depend on the dimension. Passing to the limit as $d \nearrow \infty$, we prove the same property for γ. Hence the existence result is proved for measures with bounded densities. To obtain the existence result in full generality we observe that the transport set $\mathcal{T}_e$ is a transport set also for transport problems between measures satisfying the uniformity condition stated above (Proposition 7.1 and Proposition 7.2).

The assumption that both μ and ν are a.c. with respect to γ is fundamental. Indeed take as example a diffuse measure μ and $\nu = \delta_x$, then the constant in the evolution estimate induced by the optimal transference plan will depend on the dimension and passing to the limit we lose all the informations on the evolution.

Theorem 1.2 (Theorem 7.3). *Let $\mu, \nu \in \mathcal{P}(X)$ with $\mu, \nu \ll \gamma$. Then there exists a solution T for the Monge minimization problem* (1.1)

$$\min_{T : T_\sharp \mu = \nu} \int \|x - T(x)\|_{H(\gamma)} \mu(dx).$$

Moreover we can find T μ-essentially invertible.

Conditions ensuring the existence of a transference plan of finite transference cost can be found in [13].

To conclude the introduction just few words on the organization of the paper. In Section 2 we recall the basic mathematical results we use: projective set theory, the Disintegration Theorem in the version of [5], selection principles, some fundamental results in optimal transportation theory and the definition and some properties of the abstract Wiener space.

In Section 3 we show, omitting the proof, the construction done in [6] on the Monge problem in a generalized non-branching geodesic space and we show that the Wiener space fits into the general setting.

In Section 4 we prove Theorem 1.1. In Section 5 we prove that the hypothesis of Theorem 1.1 can be proved by a finite-dimensional approximation and Section 6 proves the hypothesis of Theorem 1.1 in the finite dimensional case. Finally in Section 7 we prove Theorem 7.3 and we obtain the existence of an optimal transport map.

2. Preliminaries

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 exists 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*.

The *projective class* $\Sigma^1_{n+1}(X)$ is the family of subsets A of the Polish space X for which there exist Y Polish and $B \in \Pi^1_n(X \times Y)$ such that $A = P_1(B)$. The *coprojective class* $\Pi^1_{n+1}(X)$ is the complement in X of the class Σ^1_{n+1}.

If Σ^1_n, Π^1_n are the projective, coprojective pointclasses, then the following holds (Chapter 4 of [16]):

1. Σ^1_n, Π^1_n are closed under countable unions, intersections (in particular they are monotone classes);
2. Σ^1_n is closed w.r.t. projections, Π^1_n is closed w.r.t. coprojections;
3. if $A \in \Sigma^1_n$, then $X \setminus A \in \Pi^1_n$;
4. the *ambiguous class* $\Delta^1_n = \Sigma^1_n \cap \Pi^1_n$ is a σ-algebra and $\Sigma^1_n \cup \Pi^1_n \subset \Delta^1_{n+1}$.

We will denote by $\mathcal{A}$ the σ-algebra generated by Σ^1_1: clearly $\mathcal{B} = \Delta^1_1 \subset \mathcal{A} \subset \Delta^1_2$.

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. We say that $f : X \to \mathbb{R} \cup \{\pm\infty\}$ is a *Souslin function* if $f^{-1}(t, +\infty] \in \Sigma^1_1$.

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 $\mathscr{R}$:

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

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 $\eta := (r_\sharp \rho)$:

$$\eta(Q) := \rho(r^{-1}(Q)), \qquad \forall Q \in \mathscr{S}.$$

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.

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 respect to* r if for all s we have $\rho_s(r^{-1}(s)) = 1$.

The measures ρ_s are called *conditional probabilities*.

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

We recall the following version of the disintegration theorem that can be found on [14], Section 452 (see [5] for a direct proof).

Theorem 2.2 (Disintegration of measures). *Assume that $(R, \mathscr{R}, \rho)$ is a countably generated probability space, $\{R_s\}_{s\in S}$ a partition 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 $s \mapsto \rho_s$ in the following sense: if ρ_1, ρ_2 are two consistent disintegration then $\rho_{1,s}(\cdot) = \rho_{2,s}(\cdot)$ for η-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:

- *$\hat{R}_l := \cup_{s\in p^{-1}(l)} R_s = (p \circ r)^{-1}(l)$ is ρ-measurable and $R = \cup_{l\in L} \hat{R}_l$;*
- *the disintegration $\rho = \int_L \rho_l \lambda(dl)$ satisfies $\rho_l(\hat{R}_l) = 1$, 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 $\operatorname{graph}(u) \subset F$. We recall the following selection principle, Theorem 5.5.2 of [16], page 198.

Theorem 2.4. *Let X and Y be Polish spaces, $F \subset X \times Y$ analytic, and $\mathcal{A}$ the σ-algebra generated by the analytic subsets of X. Then $P_1(F) \in \mathcal{A}$ and there is an $\mathcal{A}$-measurable section $u : P_1(F) \to Y$ of 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 $A = \cup_{x\in A} E(x)$, or, more clearly, if $x \in A$ and $(x, y) \in E$ then $y \in A$.

The next result is taken from [16], Theorem 5.2.1.

Theorem 2.5. *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 section.*

We will use the following corollary.

Corollary 2.6. *Let $F \subset X \times X$ be $\mathcal{A}$-measurable, X Polish, such that F_x is closed for all $x \in X$ and define the equivalence relation $x \sim y \;\Leftrightarrow\; F(x) = F(y)$. Then there exists a $\mathcal{A}$-section $f : P_1(F) \to X$ such that $(x, f(x)) \in F$ and $f(x) = f(y)$ if $x \sim y$.*

Proof. For all open sets $G \subset X$, consider the sets $F^{-1}(G) = P_1(F \cap X \times G) \in \mathcal{A}$, and let $\mathcal{R}$ be the σ-algebra generated by $F^{-1}(G)$. Clearly $\mathcal{R} \subset \mathcal{A}$.

If $x \sim y$, then

$$x \in F^{-1}(G) \quad \Longleftrightarrow \quad y \in F^{-1}(G),$$

so that each equivalence class is contained in an atom of $\mathcal{R}$, and moreover by construction $x \mapsto F(x)$ is $\mathcal{R}$-measurable.

We thus conclude by using Theorem 2.5 that there exists an $\mathcal{R}$-measurable section f: this measurability condition implies that f is constant on atoms, in particular on equivalence classes. □

2.4. **General facts about optimal transportation.** Let $(X, \mathcal{B}, \mu)$ and $(Y, \mathcal{B}, \nu)$ be two Polish probability spaces and $c : X \times Y \to \mathbb{R}$ be a Borel measurable function. Consider the set of *transference plans*

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

Define the functional

$$\begin{array}{cccc} \mathcal{I} & : & \Pi(\mu,\nu) & \to & \mathbb{R}^+ \\ & & \pi & \mapsto & \mathcal{I}(\pi) := \int c\pi. \end{array} \tag{2.5}$$

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

If we consider a μ-measurable *transport map* $T : X \to Y$ such that $T_\sharp \mu = \nu$, the functional (2.5) becomes

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

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

The Kantorovich problem admits a (pre) dual formulation.

Definition 2.7. A map $\varphi : X \to \mathbb{R} \cup \{-\infty\}$ is said to be *c-concave* if it is not identically $-\infty$ and there exists $\psi : Y \to \mathbb{R} \cup \{-\infty\}$, $\psi \not\equiv -\infty$, such that

$$\varphi(x) = \inf_{y \in Y} \big\{ c(x, y) - \psi(y) \big\}.$$

The *c-transform* of φ is the function

$$\varphi^c(y) := \inf_{x \in X} \{ c(x, y) - \varphi(x) \}. \tag{2.6}$$

The *c-superdifferential* $\partial^c \varphi$ of φ is the subset of $X \times Y$ defined by

$$\partial^c \varphi := \Big\{ (x, y) : c(x, y) - \varphi(x) \leq c(z, y) - \varphi(z) \ \forall z \in X \Big\} \subset X \times Y. \tag{2.7}$$

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 $(x_0, y_0), \ldots, (x_n, y_n)$ of points of Γ, the following inequality holds:

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

where $x_{n+1} = x_0$.

A transference plan is said to be *c-cyclically monotone* if it is concentrated on a σ-compact c-cyclically monotone set.

Consider the set

$$\Phi_c := \Big\{(\varphi,\psi) \in L^1(\mu)\times L^1(\nu) : \varphi(x)+\psi(y) \le c(x,y)\Big\}. \tag{2.8}$$

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

$$J(\varphi,\psi) := \int \varphi\mu + \int \psi\nu. \tag{2.9}$$

The following is a well known result (see Theorem 5.10 of [19]).

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 $c : X\times Y\to[0,+\infty]$ 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 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*
$$c(x,y)\le c_X(x)+c_Y(y), \qquad c_X\ \mu\text{-integrable},\ c_Y\ \nu\text{-integrable},$$
then the supremum is attained:
$$\sup_{\Phi_c} J = J(\varphi,\varphi^c) = \inf_{\pi\in\Pi(\mu,\nu)} \mathcal{I}(\pi).$$

We recall also that if $-c$ is Souslin, then every optimal transference plan π is concentrated on a c-cyclically monotone set [5].

2.5. **Approximate differentiability of transport maps.** The following results are taken from [1] where they are presented in full generality.

Definition 2.10 (Approximate limit and approximate differential)**.** Let $\Omega\subset\mathbb{R}^d$ be an open set and $f:\Omega\to\mathbb{R}^m$. We say that f has an approximate limit (respectively, approximate differential) at $x\in\Omega$ if there exists a function $g:\Omega\to\mathbb{R}^m$ continuous (resp. differentiable) at x such that the set $\{f\neq g\}$ has Lebesgue-density 0 at x. In this case the approximate limit (resp. approximate differential) will be denoted by $\tilde f(x)$ (resp. $\tilde\nabla f(x)$).

Recall that if $f:\Omega\to\mathbb{R}^m$ is $\mathcal{L}^d$-measurable, then it has approximate limit $\tilde f(x)$ at $\mathcal{L}^d$-a.e. $x\in\Omega$ and $f(x)=\tilde f(x)$ $\mathcal{L}^d$-a.e..

Consider $m=d$ and denote with Σ_f the Borel set of points where f is approximately differentiable.

Lemma 2.11 (Density of the push-forward)**.** *Let $\rho\in L^1(\mathbb{R}^d)$ be a nonnegative function and assume that there exists a Borel set $\Sigma\subset\Sigma_f$ such that $\tilde f\llcorner_\Sigma$ is injective and $\{\rho>0\}\setminus\Sigma$ is $\mathcal{L}^d$-negligible. Then $f_\sharp\rho\mathcal{L}^d\ll\mathcal{L}^d$ if and only if $|\det\tilde\nabla f|>0$ for $\mathcal{L}^d$-a.e. on Σ and in this case*

$$f_\sharp(\rho\mathcal{L}^d) = \frac{\rho}{|\det\tilde\nabla f|}\circ\tilde f^{-1}\llcorner_{f(\Sigma)}\mathcal{L}^d. \tag{2.10}$$

We include a regularity result for the Monge minimization problem in $\mathbb{R}^d$ with cost $c_p(x,y)=|x-y|^p$, $p>1$ (Theorem 6.2.7 of [1]):

$$\min_{T:T_\sharp\mu=\nu}\int_{\mathbb{R}^d} c_p(x,T(x))\mu(dx). \tag{2.11}$$

Theorem 2.12. *Assume that* $\mu \in \mathcal{P}^r(\mathbb{R}^d)$, $\nu \in \mathcal{P}(\mathbb{R}^d)$,

$$\mu\Big(\Big\{x \in \mathbb{R}^d : \int c_p(x,y)\nu(dy) < +\infty\Big\}\Big) > 0 \quad \text{and} \quad \nu\Big(\Big\{y \in \mathbb{R}^d : \int c_p(x,y)\mu(dx) < +\infty\Big\}\Big) > 0.$$

If the minimum of (2.5) *is finite, then*

- *i) there exists a unique solution* T_p *for the Monge problem* (2.11)*;*
- *ii) for* μ*-a.e.* $x \in \mathbb{R}^d$ *the map* T_p *is approximately differentiable at* x *and* $\tilde{\nabla} T_p(x)$ *is diagonalizable with nonnegative eigenvalues.*

2.6. **The Abstract Wiener space.** We briefly introduce our setting. The main reference is [7].

Given an infinite dimensional separable Banach space X, we denote by $\|\cdot\|_X$ its norm and X^* denotes the topological dual, with duality $\langle\cdot,\cdot\rangle$. Given the elements $x_1^*, \dots, x_m^*$ in X^*, we denote by $\Pi_{x_1^*,\dots,x_m^*} : X \to \mathbb{R}^m$ the map

$$\Pi_{x_1^*,\dots,x_m^*}(x) := (\langle x, x_1^*\rangle, \dots, \langle x, x_m^*\rangle)\,.$$

Denoted with $\mathcal{E}(X)$ the σ-algebra generated by X^*. A set $C \in \mathcal{E}(X)$ is called *cylindrical* if

$$C = \{x \in X : \Pi_{\{x_i^*\}}(x) \in B\}, \quad B \subset \mathbb{R}^n, \quad \{x_i^*\}_{i\le n} \subset X^*,$$

and we will denote the cylindrical set with $C(B)$ where B is the base of C.

A set E belongs to $\mathcal{E}(X)$ precisely when it has the form

$$E = \{x \in X : \Pi_{\{x_i^*\}}(x) \in B\}, \quad B \in \mathcal{B}(\mathbb{R}^\infty), \quad \{x_i^*\}_{i\in\mathbb{N}} \subset X^*,$$

where $\mathbb{R}^\infty$ is considered with the standard locally convex topology. In our setting $\mathcal{B}(X) = \mathcal{E}(X)$.

Lemma 2.13 (Lemma 2.1.5 of [7])**.** *Let* μ *be a positive Borel measure on* X*. For any set* $A \in \mathcal{B}(X)_\mu$ *(the completion of* $\mathcal{B}(X)$ *w.r.t.* μ*) and any* $\varepsilon > 0$ *there exists a set* $E = C(B)$ *in* $\mathcal{E}(X)$ *with* $B \subset \mathbb{R}^\infty$ *compact in the locally convex topology of* $\mathbb{R}^\infty$*, such that*

$$E \subset A, \qquad \mu(A \setminus E) < \varepsilon.$$

A Borel measure $\gamma \in \mathcal{P}(X)$ is a *non-degenerate centred Gaussian measure* if it is not concentrated on a proper closed subspace of X and for every $x^* \in X^*$ the measure $x^*_\sharp\gamma$ is a centred Gaussian measure on $\mathbb{R}$, that is, the Fourier transform of γ is given by

$$\hat{\gamma}(x^*) = \int_X \exp\{i\langle x^*, x\rangle\}\gamma(dx) = \exp\Big\{-\frac{1}{2}\langle x^*, Qx^*\rangle\Big\}$$

where $Q \in L(X^*, X)$ is the covariance operator. The non-degeneracy hypothesis of γ is equivalent to $\langle x^*, Qx^*\rangle > 0$ for every $x^* \neq 0$. The covariance operator Q is symmetric, positive and uniquely determined by the relation

$$\langle y^*, Qx^*\rangle = \int_X \langle x^*, x\rangle\langle y^*, x\rangle\gamma(dx), \quad \forall x^*, y^* \in X^*.$$

The fact that Q is bounded follows from the Fernique's Theorem, see [7]. This imply that any $x^* \in X^*$ defines a function $x \mapsto x^*(x)$ that belongs to $L^p(X,\gamma)$ for all $1 \le p < \infty$. In particular let us denote by $R_\gamma^* : X^* \to L^2(X,\gamma)$ the embedding $R_\gamma^* x^*(x) := \langle x^*, x\rangle$. The space $\mathscr{H}$ given by the closure of $R_\gamma^* X^*$ in $L^2(X,\gamma)$ is called the *reproducing kernel* of the Gaussian measure. The definition is motivated by the fact that if we consider the operator $R_\gamma : \mathscr{H} \to X$ whose adjoint is R_γ^* then $Q = R_\gamma R_\gamma^*$:

$$\langle y^*, R_\gamma R_\gamma^* x^*\rangle = \langle R_\gamma^* y^*, R_\gamma^* x^*\rangle_{\mathscr{H}} = \int_X \langle x^*, x\rangle\langle y^*, x\rangle\gamma(dx) = \langle y^*, Qx^*\rangle.$$

It can proven that R_γ is injective, compact and that

$$R_\gamma \hat{h} = \int_X \hat{h}(x)x\gamma(dx), \quad \hat{h} \in \mathscr{H}, \tag{2.12}$$

where the integral is understood in the Bochner or Pettis sense.

The space $H(\gamma) = R_\gamma\mathscr{H} \subset X$ is called the Cameron-Martin space. It is a separable Hilbert space with inner product inherited from $L^2(X,\gamma)$ via R_γ:

$$\langle h_1, h_2\rangle_{H(\gamma)} = \langle \hat{h}_1, \hat{h}_2\rangle_{\mathscr{H}}.$$

for all $h_1, h_2 \in H$ with $h_i = R_\gamma \hat{h}_i$ for $i = 1, 2$. Moreover H is a dense subspace of X and by the compactness of R_γ follows that the embedding of $(H(\gamma), \|\cdot\|_{H(\gamma)})$ into $(X, \|\cdot\|)$ is compact. Note that if X is infinite dimensional then $\gamma(H) = 0$ and if X is finite dimensional then $X = H(\gamma)$.

2.7. **Finite dimensional approximations.** Using the embedding of X^* in $L^2(X, \gamma)$ we say that a family $\{x_i^*\} \subset X^*$ is orthonormal if the corresponding family $\{R_\gamma^* x_i^*\}$ is orthonormal in $\mathscr{H}$. In particular starting from a sequence $\{y_i^*\}_{i\in\mathbb{N}}$ whose image under R_γ^* is dense in $\mathscr{H}$, we can obtain an orthonormal basis $R_\gamma^* x_i^*$ of $\mathscr{H}$. Therefore also $h_j = R_\gamma R_\gamma^* x_j^*$ provide an orthonormal basis in $H(\gamma)$.

In the following we will consider a fixed orthonormal basis $\{e_i\}$ of $H(\gamma)$ with $e_i = R_\gamma \hat{e}_i$ for $\hat{e}_i \in R_\gamma^* X^*$.

Proposition 2.14 (Proposition 3.8.12 of [7])**.** *Let γ be a centred Gaussian measure on a Banach space X and $\{e_i\}$ an orthonormal basis in $H(\gamma)$. Define $P_d x := \sum_{i=1}^d \langle \hat{e}_i, x\rangle e_i$. Then the sequence of measures $\gamma_d := P_{d\,\sharp}\gamma \in \mathcal{P}(X)$ converges weakly to γ.*

The measure γ_d defined above is a centred non-degenerate d-dimensional Gaussian measure and, due to the orthonormality of $\{e_i\}_{i\in\mathbb{N}}$, with identity covariance matrix. Note that from (2.12) it follows that $\langle \hat{e}_j, x\rangle = \langle e_i, x\rangle_H$ for all $x \in H$. Hence we will not specify whether the measures γ_d is probability measures on $\mathbb{R}^d$ or on $P_d H$:

$$\gamma_d = \hat{e}_{1\,\sharp}\gamma \otimes \cdots \otimes \hat{e}_{d\,\sharp}\gamma, \qquad \hat{e}_{j\,\sharp}\gamma = \frac{1}{\sqrt{2\pi}} \exp\Big\{ -\frac{x^2}{2} \Big\} \mathcal{L}^1.$$

For every $d \in \mathbb{N}$ we can disintegrate γ w.r.t. the partition induced by the saturated sets of P_d:

$$\gamma = \int \gamma_{y,d}^\perp \gamma_d(dy), \quad \gamma_{y,d}^\perp(P_d^{-1}(y)) = 1 \quad \text{for } \gamma_d - \text{a.e. } y. \tag{2.13}$$

3. Optimal transportation in geodesic spaces

In what follows (X, d, d_L) is a generalized non-branching geodesic space in the sense of [6]. In this Section we retrace, omitting the proof, the construction done in [6] that permits to reduce the Monge problem with non-branching geodesic distance cost d_L, to a family of one dimensional transportation problems. The triple $(X, \|\cdot\|, \|\cdot\|_{H(\gamma)})$ is a generalized non-branching geodesic space in the sense of [6].

Let $\mu, \nu \in \mathcal{P}(X)$ and let $\pi \in \Pi(\mu, \nu)$ be a d_L-cyclically monotone transference plan with finite cost. By inner regularity, we can assume that the optimal transference plan is concentrated on a σ-compact d_L-cyclically monotone set $\Gamma \subset \{d_L(x, y) < +\infty\}$. It is worth noting that due to the lack of regularity of d_L (and of $\|\cdot\|_{H(\gamma)}$) we can't use the existence of optimal potentials (ϕ, ϕ^{d_L}) and therefore of a "fixed" cyclically monotone set. Hence we prefer to consider the cyclically monotone set Γ independent of the transference plan π.

By Lusin Theorem, we can require also that $d_{L \llcorner \Gamma}$ is σ-continuous:

$$\Gamma = \cup_n \Gamma_n, \ \Gamma_n \subset \Gamma_{n+1} \text{ compact}, \quad d_{L \llcorner \Gamma_n} \text{ continuous}. \tag{3.1}$$

Consider the set

$$\Gamma' := \Big\{ (x, y) : \exists I \in \mathbb{N}_0, (w_i, z_i) \in \Gamma \text{ for } i = 0, \ldots, I, \ z_I = y$$

$$w_{I+1} = w_0 = x, \ \sum_{i=0}^{I} d_L(w_{i+1}, z_i) - d_L(w_i, z_i) = 0 \Big\}. \tag{3.2}$$

In other words, we concatenate points $(x, z), (w, y) \in \Gamma$ if they are initial and final point of a cycle with total cost 0. One can prove that $\Gamma \subset \Gamma' \subset \{d_L(x, y) < +\infty\}$, if Γ is analytic so is Γ' and if Γ is d_L-cyclically monotone so is Γ'.

Definition 3.1 (Transport rays)**.** Define the *set of oriented transport rays*

$$G := \Big\{ (x, y) : \exists (w, z) \in \Gamma', d_L(w, x) + d_L(x, y) + d_L(y, z) = d_L(w, z) \Big\}. \tag{3.3}$$

For $x \in X$, the *outgoing transport rays from* x is the set $G(x)$ and the *incoming transport rays in* x is the set $G^{-1}(x)$. Define the *set of transport rays* as the set

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

It is fairly easy to prove that G is still d_L-cyclically monotone, $\Gamma' \subset G \subset \{d_L(x,y) < +\infty\}$ and G and R are analytic sets.

Definition 3.2. Define the *transport sets*

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

$$\mathcal{T}_e := P_1\big(\mathrm{graph}(G^{-1}) \setminus \{x=y\}\big) \cup P_1\big(\mathrm{graph}(G) \setminus \{x=y\}\big). \tag{3.5b}$$

From the definition of G one can prove that $\mathcal{T}$, $\mathcal{T}_e$ are analytic sets. The subscript e refers to the endpoints of the geodesics: we have

$$\mathcal{T}_e = P_1(R \setminus \{x=y\}). \tag{3.6}$$

It follows that we have only to study the Monge problem in $\mathcal{T}_e$: $\pi(\mathcal{T}_e \times \mathcal{T}_e \cup \{x=y\}) = 1$. 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

$$T'(x) = \begin{cases} T(x) & x \in \mathcal{T}_e \\ x & x \notin \mathcal{T}_e. \end{cases} \tag{3.7}$$

By the non-branching assumption, if $x \in \mathcal{T}$, then $R(x)$ is a single geodesic and therefore the set $R \cap \mathcal{T} \times \mathcal{T}$ is an equivalence relation on $\mathcal{T}$ that we will call ray equivalence relation. Notice that the set G is a partial order relation on $\mathcal{T}_e$.

The next step is to study the set $\mathcal{T}_e \setminus \mathcal{T}$.

Definition 3.3. Define the multivalued *endpoint graphs* by:

$$a := \big\{(x,y) \in G^{-1} : G^{-1}(y) \setminus \{y\} = \emptyset\big\}, \tag{3.8a}$$

$$b := \big\{(x,y) \in G : G(y) \setminus \{y\} = \emptyset\big\}. \tag{3.8b}$$

We call $P_2(a)$ the set of *initial points* and $P_2(b)$ the set of *final points*.

Even if a, b are not in the analytic class, still they belong to the σ-algebra $\mathcal{A}$.

Proposition 3.4. *The following holds:*

1. *the sets*
$$a, b \subset X \times X, \quad a(A), b(A) \subset X,$$
belong to the $\mathcal{A}$-class if A analytic;
2. $a \cap b \cap \mathcal{T}_e \times X = \emptyset$;
3. $a(x)$, $b(x)$ *are singleton or empty when* $x \in \mathcal{T}$;
4. $a(\mathcal{T}) = a(\mathcal{T}_e)$, $b(\mathcal{T}) = b(\mathcal{T}_e)$;
5. $\mathcal{T}_e = \mathcal{T} \cup a(\mathcal{T}) \cup b(\mathcal{T})$, $\mathcal{T} \cap (a(\mathcal{T}) \cup b(\mathcal{T})) = \emptyset$.

Finally we can assume that the μ-measure of final points and the ν-measure of the initial points are 0: indeed since the sets $G \cap b(\mathcal{T}) \times X$, $G \cap X \times a(\mathcal{T})$ is a subset of the graph of the identity map, it follows that from the definition of b one has that

$$x \in b(\mathcal{T}) \quad \Longrightarrow \quad G(x) \setminus \{x\} = \emptyset,$$

A similar computation holds for a. Hence we conclude that

$$\pi(b(\mathcal{T}) \times X) = \pi(G \cap b(\mathcal{T}) \times X) = \pi(\{x=y\}),$$

and following (3.7) we can assume that

$$\mu(b(\mathcal{T})) = \nu(a(\mathcal{T})) = 0.$$

3.1. **The Wiener case.** For the abstract Wiener space, it is possible to obtain more regularity for the sets introduced so far. Let $d = \|\cdot\|$ and $d_L = \|\cdot\|_H$: by the compactness of the embedding R_γ of H into X it follows that

(1) $d_L : X \times X \to [0, +\infty]$ is a l.s.c. distance;
(2) $d_L(x, y) \geq Cd(x, y)$ for some positive constant C;
(3) $\cup_{x \in K_1, y \in K_2} \gamma_{[x,y]}$ is d-compact if K_1, K_2 are d-compact and $d_{L|_{K_1 \times K_2}}$ is uniformly bounded.

The set Γ' is σ-compact: in fact, if one restrict to each Γ_n given by (3.1), then the set of cycles of order I is compact, and thus

$$\Gamma'_{n,\bar{I}} := \Big\{ (x, y) : \exists I \in \{0, \ldots, \bar{I}\}, (w_i, z_i) \in \Gamma_n \text{ for } i = 0, \ldots, I, \ z_I = y \\ w_{I+1} = w_0 = x, \ \sum_{i=0}^{I} d_L(w_{i+1}, z_i) - d_L(w_i, z_i) = 0 \Big\}$$

is compact. Finally $\Gamma' = \cup_{n,I} \Gamma'_{n,I}$.

Moreover, $d_{L \llcorner \Gamma'_{n,I}}$ is continuous. If $(x_n, y_n) \to (x, y)$, then from the l.s.c. and

$$\sum_{i=0}^{I} d_L(w_{n,i+1}, z_{n,i}) = \sum_{i=0}^{I} d_L(w_{n,i}, z_{n,i}), \quad w_{n,I+1} = w_{n,0} = x_n, \ z_{n,I} = y_n,$$

it follows also that each $d_L(w_{n,i+1}, z_{n,i})$ is continuous.

Similarly the sets G, R, a, b are σ-compact: assumption (3) and the above computation in fact shows that

$$G_{n,I} := \Big\{ (x, y) : \exists (w, z) \in \Gamma'_{n,I}, d_L(w, x) + d_L(x, y) + d_L(y, z) = d_L(w, z) \Big\}$$

is compact. For a, b, one uses the fact that projection of σ-compact sets is σ-compact.

So we have that Γ, Γ', G, G^{-1}, a and b are σ-compact sets.

3.2. **Strongly consistency of disintegrations.** The strong consistency of the disintegration follows from the next result.

Proposition 3.5. *There exists a μ-measurable cross section $f : \mathcal{T} \to \mathcal{T}$ for the ray equivalence relation R.*

Up to a μ-negligible saturated set $\mathcal{T}_N$, we can assume it to have σ-compact range: just let $S \subset f(\mathcal{T})$ be a σ-compact set where $f_\sharp \mu \llcorner_{\mathcal{T}}$ is concentrated, and set

$$\mathcal{T}_S := R^{-1}(S) \cap \mathcal{T}, \quad \mathcal{T}_N := \mathcal{T} \setminus \mathcal{T}_S, \quad \mu(\mathcal{T}_N) = 0. \tag{3.9}$$

Having the $\mu \llcorner_{\mathcal{T}}$-measurable cross-section

$$\mathcal{S} := f(\mathcal{T}) = S \cup f(\mathcal{T}_N) = (\text{Borel}) \cup (f(\mu\text{-negligible})),$$

we can define the parametrization of $\mathcal{T}$ and $\mathcal{T}_e$ by geodesics.

Using the quotient map f, we obtain a unitary speed parametrization of the transport set.

Definition 3.6 (Ray map)**.** Define the *ray map* g by the formula

$$\begin{aligned} g :=& \Big\{ (y, t, x) : y \in \mathcal{S}, t \in [0, +\infty), x \in G(y) \cap \{d_L(x, y) = t\} \Big\} \\ & \cup \Big\{ (y, t, x) : y \in \mathcal{S}, t \in (-\infty, 0), x \in G^{-1}(y) \cap \{d_L(x, y) = -t\} \Big\} \\ =& \ g^+ \cup g^-. \end{aligned}$$

Proposition 3.7. *The following holds.*

(1) *The restriction $g \cap S \times \mathbb{R} \times X$ is analytic.*
(2) *The set g 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 $y \in \mathcal{T}$.*

(4) *$(t,y) \mapsto g(y,t)$ is bijective on $\mathcal{T}$, and its inverse is*

$$x \mapsto g^{-1}(x) = \big(f(y), \pm d_L(x, f(y))\big)$$

where f is the quotient map of Proposition 3.5 and the positive/negative sign depends on $x \in G(f(y))/x \in G^{-1}(f(y))$.

Another property of d_L-cyclically monotone transference plans.

Proposition 3.8. *For any π d_L-monotone there exists a d_L-cyclically monotone transference plan $\tilde{\pi}$ with the same cost of π such that it coincides with the identity on $\mu \wedge \nu$.*

Coming back to the abstract Wiener space, we have that given $\mu, \nu \ll \gamma$ and given $\pi \in \Pi(\mu,\nu)$ $\|\cdot\|_{H(\gamma)}$-cyclically monotone, we have constructed the transport $\mathcal{T}$ (and $\mathcal{T}_e$), an equivalence relation R on it with geodesics as equivalence classes and the corresponding disintegration is strongly consistent:

$$\mu \llcorner_{\mathcal{T}} = \int_{\mathcal{S}} \mu_y m(dy) \tag{3.10}$$

with $m = f_\sharp \mu$ and $\mu_y(R(y)) = 1$ for m-a.e. $y \in \mathcal{T}$. Using the ray map g one can assume that $\mu_y \in \mathcal{P}(\mathbb{R})$ and

$$\mu \llcorner_{\mathcal{T}} = g_\sharp \int_{\mathcal{S}} \mu_y m(dy).$$

4. Regularity of disintegration

To obtain existence of an optimal transport map it is enough to prove that:

- μ in concentrated on $\mathcal{T}$;
- μ_y is a continuous measure for m-a.e. $y \in \mathcal{S}$.

Indeed at that point, for every $y \in \mathcal{S}$ we consider the unique monotone map T_y such that $T_{y\,\sharp}\mu_y = \nu_y$, then $T(g(y,t)) := T_y(g(y,t))$ is an optimal transport map, see Theorem 6.2 of [6].

Define the map $X \times X \ni (x,y) \mapsto T_t(x,y) := x(1-t) + yt \in X$.

Assumption 1 (Non-degeneracy assumption)**.** The measure γ is said to satisfy Assumption 1 w.r.t. a $\|\cdot\|_{H(\gamma)}$-cyclically monotone set Γ if

i) $\pi(\Gamma) = 1$ with $\pi \in \Pi(\mu,\nu)$ and $\mu,\nu \ll \gamma$;
ii) for each set $A \in \mathcal{E}(X)$ with compact base such that $\mu(A) > 0$ there exist $C > 0$ and $\{t_n\}_{n\in\mathbb{N}} \subset [0,1]$ converging to 0 as $n \to +\infty$ such that

$$\gamma(T_{t_n}(\Gamma \cap A \times X)) \geq C\mu(A) \tag{4.1}$$

for all $n \in \mathbb{N}$.

An immediate consequence of Assumption 1 is that the set of final points is γ-negligible.

Proposition 4.1. *If γ satisfies Assumption* (1) *then*

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

Proof. Let $A = a(\mathcal{T}_e)$ and recall that $\mu = \rho_1\gamma$. Suppose by contradiction $\mu(A) > 0$. By inner regularity and Lemma 2.13 there exist a Borel set $C(B) =: \hat{A} \subset A$, with compact base B, of positive μ-measure and a strictly positive constant $\delta \in \mathbb{R}$ such that $\rho_1(x) \geq \delta$ for all $x \in \hat{A}$. Since $\Gamma \subset \{(x,y) : \|x-y\|_{H(\gamma)} < +\infty\}$, we can moreover assume that

$$\Gamma \cap \hat{A} \times X \subset \{(x,y) : \|x-y\|_{H(\gamma)} \leq M\}$$

for some positive $M \in \mathbb{R}$.

By Assumption 1 there exist $C > 0$ and $\{t_n\}_{n\in\mathbb{N}}$ converging to 0 such that

$$\gamma(T_{t_n}(\Gamma \cap \hat{A} \times X)) \geq C\mu(\hat{A}) \geq \delta C\gamma(\hat{A}).$$

Denote with $\hat{A}_{t_n} = T_{t_n}(\Gamma \cap \hat{A} \times X)$ and define

$$\hat{A}^\varepsilon := \big\{x : \|\hat{A} - x\|_{H(\gamma)} < \varepsilon\big\} = P_1\Big(\big\{(x,y) \in X \times \hat{A} : \|x-y\|_{H(\gamma)} < \varepsilon\big\}\Big).$$

Since $\hat{A} \subset A = a(\mathcal{T}_e)$, $\hat{A}_{t_n} \cap \hat{A} = \emptyset$ for every $n \in \mathbb{N}$. Moreover for $t_n \leq \varepsilon/M$ it holds $\hat{A}^\varepsilon \supset \hat{A}_{t_n}$. So we have for t_n small enough

$$\gamma(\hat{A}^\varepsilon) \geq \gamma(\hat{A}) + \gamma(\hat{A}_{t_n}) \geq (1 + C\delta)\gamma(\hat{A}).$$

Since $\gamma(\hat{A}) = \lim_{\varepsilon \to 0} \gamma(\hat{A}^\varepsilon)$, this is a contradiction. □

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

$$\mu = \int_S \mu_y m(dy), \quad m = f_\sharp \mu, \ \mu_y \in \mathcal{P}(R(y)). \tag{4.2}$$

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 4.2. *If γ satisfies Assumption 1 then the conditional probabilities μ_y are continuous for m_γ-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 $< \varepsilon$. 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 A is Borel.

Step 2. The claim is equivalent to $\mu(P_2(A)) = 0$. Suppose by contradiction $\mu(P_2(A)) > 0$. By Lusin Theorem (Theorem 5.8.11 of [16]) A is the countable union of Borel graphs. Therefore we can take a Borel selection of A just considering one of the Borel graphs, say $\hat{A}$. Clearly $m(P_1(\hat{A})) > 0$ hence by (4.2) $\mu(P_2(\hat{A})) > 0$. Using Lemma 2.13 we can find a Borel subset $\tilde{A} \subset P_2(\hat{A})$ still with positive μ-measure such that $\tilde{A} = C(B)$ with $B \subset \mathbb{R}^\infty$ compact.

By Assumption 1, $\gamma(T_{t_n}(\Gamma \cap \tilde{A} \times X)) \geq C\mu(\tilde{A})$ for some $C > 0$ and $t_n \to 0$. From $T_{t_n}(\Gamma \cap \tilde{A} \times X) \cap (\tilde{A}) = \emptyset$, using the same argument of Proposition 4.1, the claim follows. □

5. An approximation result

Let $P_d : X \to H$ be the projection map of Proposition 2.14 associated to the orthonormal basis $\{e_i\}_{i\in\mathbb{N}}$ of $H(\gamma)$ with $e_i = R_\gamma \hat{e}_i$ for $\hat{e}_i \in R_\gamma^* X^*$ and $P_{d\,\sharp}\gamma = \gamma_d$.

Consider the following measures

$$\mu_d := P_{d\,\sharp}\mu, \qquad \nu_d := P_{d\,\sharp}\nu \tag{5.1}$$

and observe that $\mu_d = \rho_{1,d}\gamma_d$ and $\nu_d = \rho_{2,d}\gamma_d$ with

$$\rho_{i,d}(z) = \int \rho_i(x)\gamma_{z,d}^\perp(dx), \qquad i = 1, 2, \tag{5.2}$$

where $\gamma_{z,d}^\perp$ is defined in (2.13). Recall that $\mu_d \rightharpoonup \mu$ and $\nu_d \rightharpoonup \nu$ as $d \nearrow \infty$.

Denote with $\Pi_o(\mu_d, \nu_d)$ the set of optimal transference plans for the Monge problem between μ_d and ν_d with $\|\cdot\|_{H(\gamma)}$-cost.

Proposition 5.1. *Let $\pi_d \in \Pi_o(\mu_d, \nu_d)$ and let $\pi \in \Pi(\mu, \nu)$ be any weak limit of $\{\pi_d\}_{d\in\mathbb{N}}$. Then $\pi \in \Pi(\mu, \nu)$ is an optimal transport plan for* (1.1).

Proof. Let $\hat{\pi} \in \Pi(\mu, \nu)$ be a transference plan. The following holds true

$$\begin{aligned}\int \|x - y\|_H \hat{\pi}(dxdy) \geq& \int \|P_d(x - y)\|_{H(\gamma)} \hat{\pi}(dxdy) = \int \|x - y\|_{H(\gamma)} ((P_d \otimes P_d)_\sharp \hat{\pi})(dxdy) \\ \geq& \int \|x - y\|_{H(\gamma)} \pi_d(dxdy).\end{aligned}$$

Let $\{d_k\}_{k\in\mathbb{N}}$ be a subsequence such that $\pi_{d_k} \rightharpoonup \pi$ as $d_k \nearrow \infty$. Since $\|\cdot\|_{H(\gamma)}$ is l.s.c. it follows that

$$\int \|x - y\|_{H(\gamma)} \hat{\pi}(dxdy) \geq \liminf_{k\to+\infty} \int \|x - y\|_{H(\gamma)} \pi_{d_k}(dxdy) \geq \int \|x - y\|_{H(\gamma)} \pi(dxdy).$$

Hence the claim follows. □

Since $\rho_{i,d}$ depend only on the first d-coordinates, the measures μ_d, ν_d can be considered also as probability measure on $\mathbb{R}^d$. Clearly for $x, y \in P_d(X)$ the norm $\|\cdot\|_d$ and $\|\cdot\|_{H(\gamma)}$ coincide. Therefore we can study the transport problem with euclidean norm cost $\|x\|_d^2 := \sum_{j=1}^d x_j^2$:

$$\min_{\pi \in \Pi(\mu_d,\nu_d)} \int \|x-y\|_d \pi(dxdy). \tag{5.3}$$

However it is worth noting that when we speak of weak convergence, the measures μ_d, ν_d and γ_d are all thought as probability measures in X.

It is a well-known fact in optimal transportation that (5.3) has a minimizer of the form $(Id, T_d)_\sharp \mu_d$ with T_d μ-essentially invertible and Borel. For each d we choose as optimal map T_d the one obtained gluing the monotone rearrangements over the geodesics and we set $\pi_d := (Id, T_d)_\sharp \mu_d$. Moreover $\Gamma_d := \mathrm{graph}(T_d)$.

The results that we are about to present are true for any weak limit π of the sequence $\{\pi_d\}_{d\in\mathbb{N}}$. Nevertheless to simplify the notation we assume that the whole sequence $\{\pi_d\}_{d\in\mathbb{N}} = \{(Id, T_d)_\sharp \mu_d\}_{d\in\mathbb{N}}$ converges to some π.

Theorem 5.2. *Fix $t \in [0,1]$. Assume that there exists $C > 0$ such that for all $d \in \mathbb{N}$ and $A \subset X$ compact set the following holds true*

$$\gamma_d\big(T_t(\Gamma_d \cap A \times X)\big) \geq C\mu_d(A).$$

Then for all $A \subset X$ γ-measurable

$$\gamma\big(T_t(\Gamma \cap A \times X)\big) \geq C\mu(A), \tag{5.4}$$

where $\Gamma \subset X \times X$ is any $\|\cdot\|_{H(\gamma)}$-cyclically monotone with $\pi(\Gamma) = 1$.

Proof. It follows from Proposition 5.1 that π is an optimal transference plan, hence it is concentrated on a $\|\cdot\|_{H(\gamma)}$-cyclically monotone set Γ.

Step 1. Since $\mu_d \rightharpoonup \mu$ and $\nu_d \rightharpoonup \nu$, for every $\varepsilon > 0$ there exist $K_{1,\varepsilon}$ and $K_{2,\varepsilon}$ compact sets such that $\mu_d(K_{1,\varepsilon}) \geq 1 - \varepsilon/3$ and $\nu_d(K_{2,\varepsilon}) \geq 1 - \varepsilon/3$. Denote with $K_\varepsilon := K_{1,\varepsilon} \times K_{2,\varepsilon}$. For every $d \in \mathbb{N}$ there exists a compact set $\hat\Gamma_d \subset \Gamma_d$ such that $\pi_d(\hat\Gamma_d) \geq 1 - \varepsilon/3$. Consider the compact se $\Gamma_{d,\varepsilon} := \hat\Gamma_d \cap K_\varepsilon$, then $\pi_d(\Gamma_{d,\varepsilon}) \geq 1 - \varepsilon$ and $\Gamma_{d,\varepsilon}$ converges as $d \nearrow \infty$ in the Hausdorff topology, up to subsequences, to a compact set Γ_ε with $\pi(\Gamma_\varepsilon) \geq 1 - \varepsilon$.

Step 2. Let $\Gamma_n \subset \Gamma$ be a compact set such that $\pi(\Gamma_n) \geq 1 - 1/n$. Hence $\pi(\Gamma_\varepsilon \cap \Gamma_n) \geq 1 - \varepsilon - 1/n$. Consider the following sets, open and closed respectively:

$$(\Gamma_\varepsilon \cap \Gamma_n)^\delta := \{x : \|\Gamma_\varepsilon \cap \Gamma_n - x\| < \delta\}, \quad cl(\Gamma_\varepsilon \cap \Gamma_n)^\delta := \{x : \|\Gamma_\varepsilon \cap \Gamma_n - x\| \leq \delta\}.$$

Since $\liminf_d \pi_d(U) \geq \pi(U)$ for every open set $U \subset X$, it follows that for every $\delta > 0$ there exists $d_\delta \in \mathbb{N}$ such that for all $d \geq d_\delta$

$$\pi_d\big((\Gamma_\varepsilon \cap \Gamma_n)^\delta\big) \geq 1 - 2\varepsilon - 1/n.$$

The same inequality holds true for $\pi_d\big(cl(\Gamma_\varepsilon \cap \Gamma_n)^\delta\big)$. Therefore

$$\pi_{d_\delta}\big(\Gamma_{d_\delta,\varepsilon} \cap cl(\Gamma_\varepsilon \cap \Gamma_n)^\delta\big) \geq 1 - 3\varepsilon - 1/n.$$

Take as $\delta = 1/k$ for $k \in \mathbb{N}$ and let $d_k := d_{\delta_k}$. Define the compact set $\Gamma^n_{k,\varepsilon} := \Gamma_{d_k,\varepsilon} \cap cl(\Gamma_\varepsilon \cap \Gamma_n)^{1/k}$, then since $\Gamma^n_{k,\varepsilon} \subset K_\varepsilon$, up to subsequences, $\lim_{k\nearrow\infty} d_H(\Gamma^n_{k,\varepsilon}, \Gamma_{\varepsilon,n}) = 0$ with

$$\Gamma_{\varepsilon,n} \subset \Gamma_\varepsilon \cap \Gamma_n \subset \Gamma, \quad \pi_{d_k}(\Gamma^n_{k,\varepsilon}) \geq 1 - 3\varepsilon - 1/n, \quad \pi(\Gamma_{\varepsilon,n}) \geq 1 - 3\varepsilon - 1/n. \tag{5.5}$$

The inclusion $\Gamma_{\varepsilon,n} \subset \Gamma_\varepsilon \cap \Gamma_n$ can be verified observing that any limit point of sequences of $\Gamma^n_{k,\varepsilon}$ must be contained in $\Gamma_\varepsilon \cap \Gamma_n$.

Step 3. Consider $A = C(B) \in \mathcal{E}(X)$ with $B \in \mathbb{R}^m$ compact set for some fixed $m \in \mathbb{N}$. Since T_t is continuous and $\Gamma^n_{k,\varepsilon} \cap A \times X$ converges in Hausdorff topology to $\Gamma_{\varepsilon,n} \cap A \times X$, it is fairly easy to prove that $T_t(\Gamma^n_{k,\varepsilon} \cap A \times X)$ Kuratowski-converges to $T_t(\Gamma_{\varepsilon,n} \cap A \times X)$. For the definition of Kuratowski-convergence see for instance [4]. Moreover

$$T_t(\Gamma^n_{k,\varepsilon} \cap A \times X) \subset \overline{co}(P_1(K_\varepsilon \cap A \times X) \cap P_2(K_\varepsilon \cap A \times X))$$

and by Proposition A.1.6 of [7], $\overline{co}(P_1(K_\varepsilon \cap A \times X) \cap P_2(K_\varepsilon \cap A \times X))$ is compact. Therefore, by Proposition 4.4.14 of [4], $T_t(\Gamma^n_{k,\varepsilon} \cap A \times X)$ converges also in the Hausdorff topology to $T_t(\Gamma_{\varepsilon,n} \cap A \times X)$.

Step 4. It follows that

$$\gamma(T_t(\Gamma_{\varepsilon,n} \cap A \times X)) \geq \limsup_{k\to+\infty} \gamma_{d_k}(T_t(\Gamma^n_{k,\varepsilon} \cap A \times X)),$$

hence, since $\Gamma^n_{k,\varepsilon}$ is a subset of the graph Γ_{d_k}, it follows that

$$\begin{aligned}\gamma(T_t(\Gamma_{\varepsilon,n} \cap A \times X)) \geq & \limsup_{k\to+\infty} \gamma_{d_k}(T_t(\Gamma^n_{k,\varepsilon} \cap A \times X)) \\ & \geq C \limsup_{k\to+\infty} \mu_{d_k}(P_1(\Gamma^n_{k,\varepsilon}) \cap A) \\ & \geq C \limsup_{k\to+\infty} \mu_{d_k}(A) - C(3\varepsilon - 1/n) \qquad (5.6)\end{aligned}$$

where in the last equation we have used $\mu_d(P_1(\Gamma^n_{k,\varepsilon})) \geq 1 - 3\varepsilon - 1/n$ that follows from (5.5). Since $\mu_{d_k} = P_{d_k\sharp}\mu$ and A has finite dimensional base, the sequence $\{\mu_{d_k}(A)\}_{k\in\mathbb{N}}$ is definitively constant and therefore

$$\gamma(T_t(\Gamma_{\varepsilon,n} \cap A \times X)) \geq C\mu(A) - C(3\varepsilon - 1/n) \qquad (5.7)$$

for all $A \in \mathcal{E}(X)$ with finite dimensional base.

Step 5. Consider $A = C(B) = \{x \in X : \{\ell_i(x)\}_{i\in\mathbb{N}} \in B\}$ with $B \in \mathbb{R}^\infty$ compact set in the locally convex topology of $\mathbb{R}^\infty$ and $\{\ell_i\}_{i\in\mathbb{N}} \subset X^*$. We consider the sequence of compact sets

$$A_d := C(P_d(B)) = \{x \in X : \{\ell_i(x)\}_{i\leq d} \in P_d(B)\}.$$

Clearly A_d is closed with finite-dimensional compact base and $A_d \supset A_{d+1} \supset A$ for every $d \in \mathbb{N}$. Then for every $d \in \mathbb{N}$ from (5.7)

$$\gamma(T_t(\Gamma_{\varepsilon,n} \cap A_d \times X)) \geq C\mu(A_d) - C(3\varepsilon - 1/n) \geq C\mu(A) - C(3\varepsilon - 1/n).$$

Since the first term in the above inequality is decreasing, it follows that

$$\lim_{d\to+\infty} \gamma(T_t(\Gamma_{\varepsilon,n} \cap A_d \times X)) \geq C\mu(A) - C(3\varepsilon - 1/n).$$

Now observe that $\bigcap_{d=1}^\infty T_t(\Gamma_{\varepsilon,n} \cap A_d \times X) = T_t(\Gamma_{\varepsilon,n} \cap A \times X)$: indeed one inclusion is trivial and for the other one observe that if

$$x = (1-t)y_d + tz_d, \quad y_d \in A_d, (y_d, z_d) \in \Gamma_{n,\varepsilon}, \forall d \in \mathbb{N},$$

then from the compactness of $\Gamma_{n,\varepsilon}$, up to subsequences, $y_d \to y$ $z_d \to z$ and observing that $y \in A = \cap_{d\in\mathbb{N}} A_d$ the inclusion, and thus the identity, is proved. Thus (5.7) holds true and

$$\gamma(T_t(\Gamma \cap A \times X)) \geq \gamma(T_t(\Gamma_{\varepsilon,n} \cap A \times X)) \geq C\mu(A) - C(3\varepsilon - 1/n).$$

Letting $\varepsilon \to 0$ and $n \to +\infty$, the claim is proved for every $A \in \mathcal{E}(X)$ with compact base. The extension to γ-measurable sets is now a straightforward application of Lemma 2.13. □

6. Finite dimensional estimate

The next theorem proves that the d-dimensional standard Gaussian measure $\gamma_d = P_{d\sharp}\gamma$ satisfies Assumption 1 for $\Gamma = \mathrm{graph}(T_d) = \Gamma_d$.

Theorem 6.1. *Assume that there exists $C > 0$ such that $\rho_{i,d}(x) \leq C$ for γ_d-a.e. $x \in \mathbb{R}^d$ and $i = 1, 2$. Then the following estimate holds true*

$$\gamma_d\big(T_t(\Gamma_d \cap A \times X)\big) \geq \frac{1}{C}\mu_d(A), \quad \forall t \in [0,1], A \in \mathcal{B}(\mathbb{R}^d).$$

Proof. Observe that the set $T_t(\Gamma_d \cap A \times X)$ is parametrized by the map $T_{d,t} := (1-t)Id + tT_d$.

Step 1. Consider the Monge minimization problem with cost c_p, (2.11), between μ_d and ν_d. It follows from Theorem 2.12 and from the boundedness of $\rho_{i,d}$ that there exists a unique optimal map $T_{p,d}$ approximately differentiable μ_d-a.e.. We will use the following notations: $\rho_i = \rho_{i,d}$ and $T_p = T_{p,d}$. By Lemma 2.11 it follows that

$$\rho_2(T_p(x))|\det \tilde{\nabla} T_p|(x) \prod_{j=1}^d \frac{1}{\sqrt{2\pi}} \exp\Big\{-\frac{T_p(x)_j^2}{2}\Big\} = \rho_1(x) \prod_{j=1}^d \frac{1}{\sqrt{2\pi}} \exp\Big\{-\frac{x_j^2}{2}\Big\}.$$

Since for μ_d-a.e. $x \in \mathbb{R}^d$ $|\det \tilde{\nabla} T_p|(x) > 0$, also $\rho_2(T_p(x)) > 0$ for μ_d-a.e. $x \in \mathbb{R}^d$. Hence the following makes sense μ-a.e.:

$$Jac(T_p)(x) = |\det \tilde{\nabla} T_p|(x) = \frac{\rho_1(x)}{\rho_2(T_p(x))} \exp\Big\{ \sum_{j=1}^d -\frac{1}{2}(x_j^2 - T_p(x)_j^2) \Big\}.$$

Step 2. Let $T_{p,t} := (1-t)Id + tT_p$. From Theorem 2.12, $\det \tilde{\nabla} T_p(x) = \prod_{j=1}^d \lambda_j$ with $\lambda_i > 0$ for $i = 1, \dots, d$. It follows that

$$Jac(T_{p,t})(x) = \det((1-t)Id + t\tilde{\nabla} T_p(x)) = \prod_{j=1}^d \big((1-t) + \lambda_j t\big).$$

Passing to logarithms, we have by concavity

$$\log(Jac(T_{p,t})(x)) \geq t \log(Jac(T_p)(x)) + (1-t)\log(Jac(Id)) = t\log(Jac(T_p)(x)).$$

Hence

$$Jac(T_{p,t})(x) \geq \Big(Jac(T_p)\Big)^t(x) \geq \Big(\frac{\rho_1(x)}{\rho_2(T_p(x))}\Big)^t \exp\Big\{ \sum_{j=1}^d -\frac{1}{2} t(x_j^2 - T_p(x)_j^2) \Big\}. \tag{6.1}$$

Step 3. We have the following

$$\begin{aligned}
&\exp\Big\{ \sum_{j=1}^d -\frac{1}{2}(T_{p,t}(x)_j^2 - x_j^2) \Big\} Jac(T_{p,t})(x) \\
&\qquad \geq \exp\Big\{ \sum_{j=1}^d -\frac{1}{2}(T_{p,t}(x)_j^2 - x_j^2) \Big\} \Big(\frac{\rho_1(x)}{\rho_2(T_p(x))}\Big)^t \exp\Big\{ \sum_{j=1}^d -\frac{1}{2} t(x_j^2 - T_p(x)_j^2) \Big\} \\
&\qquad = \Big(\frac{\rho_1(x)}{\rho_2(T_p(x))}\Big)^t \exp\Big\{ \sum_{j=1}^d -\frac{1}{2}(T_{p,t}(x)_j^2 - x_j^2 + tx_j^2 - tT_p(x)_j^2) \Big\} \\
&\qquad = \Big(\frac{\rho_1(x)}{\rho_2(T_p(x))}\Big)^t \exp\Big\{ \sum_{j=1}^d -\frac{1}{2}\Big(((1-t)x_j + tT_p(x)_j)^2 - ((1-t)x_j^2 + tT_p(x)_j^2\Big) \Big\} \\
&\qquad = \Big(\frac{\rho_1(x)}{\rho_2(T_p(x))}\Big)^t \exp\Big\{ \sum_{j=1}^d -\frac{1}{2}\big(x_j - T_p(x)_j\big)^2 (t^2 - t) \Big\} \\
&\qquad = \Big(\frac{\rho_1(x)}{\rho_2(T_p(x))}\Big)^t \exp\Big\{ -\frac{1}{2} \|x - T_p(x)\|_d^2 (t^2 - t) \Big\}.
\end{aligned}$$

Thus

$$\begin{aligned}
\gamma(T_{p,t}(A)) &= \int_A Jac(T_{p,t})(x) \prod_{j=1}^d \frac{1}{\sqrt{2\pi}} \exp\Big\{ -\frac{1}{2} T_{p,t}(x)_j^2 \Big\} \mathcal{L}^d(dx) \\
&= \int_A Jac(T_{p,t})(x) \exp\Big\{ \sum_{j=1}^d -\frac{1}{2}(T_{p,t}(x)_j^2 - x_j^2) \Big\} \gamma(dx) \\
&\geq \int_A \Big(\frac{\rho_1(x)}{\rho_2(T_p(x))}\Big)^t \exp\Big\{ \frac{1}{2} \|x - T_p(x)\|_d^2 (t - t^2) \Big\} \gamma(dx) \\
&\geq \frac{1}{C^t} \int_A \rho_1(x)^t \gamma(dx) \\
&\geq \frac{1}{C^t} \int_A \rho_1(x)^{t-1} \mu(dx) \\
&\geq \frac{1}{C} \mu(A).
\end{aligned}$$

Step 4. Since $(Id, T_p)_\sharp \mu_d \rightharpoonup (Id, T)_\sharp \mu_d$ as $p \searrow 1$, see Theorem 7.1 of [3], using the techniques of the proof of Theorem 5.2, one can prove that

$$\gamma_d(T_t(\Gamma_d \cap A \times X)) \geq \frac{1}{C}\mu_d(A).$$

□

Remark 6.2. We summarize the results obtained so far. If $\rho_1, \rho_2 \leq C$, then from (5.2) it follows that the densities of μ_d and ν_d enjoy the same property with the same constant C. Identifying μ_d, ν_d and γ_d with the corresponding measures on $\mathbb{R}^d$, we have from Theorem 6.1:

$$\gamma_d\big(T_t(\Gamma_d \cap A \times X)\big) \geq \frac{1}{C}\mu_d(A), \qquad \forall A \in \mathcal{B}(X), t \in [0,1].$$

From Theorem 5.2 we have the same inequality for the ∞-dimensional measures:

$$\gamma\big(T_t(\Gamma \cap A \times X)\big) \geq \frac{1}{C}\mu(A), \qquad \forall A \in \mathcal{B}(X)_\gamma, t \in [0,1].$$

As Proposition 4.1 and Proposition 4.2 show, this estimate implies $\mu(a(\mathcal{T})) = 0$ and the continuity of the conditional probabilities μ_y. Since the optimal finite dimensional map T_d is invertible, following the argument of Theorem 6.1 we can also prove

$$\gamma\big(T_{1-t}(\Gamma \cap X \times A)\big) \geq \frac{1}{C}\nu(A), \tag{6.2}$$

and adapting the proofs of Proposition 4.1 and Proposition 4.2 we can prove that $\nu(b(\mathcal{T})) = 0$ and the continuity of the conditional probabilities ν_y. So we have

$$\mu = \int_{\mathcal{S}} \mu_y m(dy), \quad \nu = \int_{\mathcal{S}} \nu_y m(dy), \qquad \mu_y, \nu_y \text{ continuous for } m - a.e. y \in \mathcal{S}.$$

In the next Section we remove the hypothesis $\rho_1, \rho_2 \leq M$.

7. Solution

Let π be the weak limit of $\pi_d = (Id, T_d)_\sharp \mu_d$ and Γ any $\|\cdot\|_{H(\gamma)}$-cyclically monotone set such that $\pi(\Gamma) = 1$. All the definition of Section 3 are referred to this Γ.

Proposition 7.1. *Let $\mu, \nu \in \mathcal{P}(X)$ be such that $\mu, \nu \ll \gamma$. Then $\mu(a(\mathcal{T})) = \nu(b(\mathcal{T})) = 0$.*

Proof. Let $\mu = \rho_1 \gamma$ and $\nu = \rho_2 \gamma$. We only prove that $\mu(a(\mathcal{T})) = 0$. The other statement follows similarly.

Step 1. Assume by contradiction that $\mu(a(\mathcal{T})) > 0$. Let $A \subset a(\mathcal{T})$ be such that $\mu(A) > 0$ and for every $x \in A$, $\rho_1(x) \leq M$ for some positive constant M. Consider $\gamma \llcorner_{\mathcal{T}}$ and its disintegration

$$\gamma \llcorner_{\mathcal{T}} = \int_{\mathcal{S}} \gamma_y m_\gamma(dy), \quad \gamma_y(\mathcal{T}) = 1, \; m_\gamma - a.e. y \in \mathcal{S}.$$

Consider the initial point map $a : \mathcal{S} \to A$ and the measure $a_\sharp m_\gamma$. Observe that since

$$\forall B \subset A : \mu(B) > 0 \quad \Rightarrow \quad \gamma(R(B) \cap \mathcal{T}) > 0,$$

it follows that $\mu \llcorner_A \ll a_\sharp m_\gamma$. Hence there exists $\hat{A} \subset A$ of positive $a_\sharp m_\gamma$-measure such that the map

$$\hat{A} \ni x \mapsto h(x) := \frac{d\mu \llcorner_A}{da_\sharp m_\gamma}(x)$$

verifies $h(x) \leq M'$ for some positive constan M'.

Step 2. Considering

$$\mu \llcorner_{\hat{A}}, \qquad \hat{\gamma} := \int_{R(\hat{A}) \cap \mathcal{S}} h(a(y)) \gamma_y m_\gamma(dy),$$

we have the claim. Indeed both have uniformly bounded densities w.r.t. γ and $\mathcal{T}_e$ is still a transport set for the transport problem between $\mu\llcorner_{\hat{A}}$ and $\hat{\gamma}$. Indeed for $S \subset \mathcal{S}$

$$\begin{aligned}\mu\llcorner_{\hat{A}}(\cup_{y\in S}R(y)) &= \mu\llcorner_{\hat{A}}(a(S)) \\ &= \int_{a(S)} h(a)(a_\sharp m_\gamma)(da) \\ &= \int_S h(a(y))m_\gamma(dy) = \hat{\gamma}(\cup_{y\in S}R(y)).\end{aligned}$$

Hence we can project the measures, obtain the finite dimensional estimate of Theorem 6.1, obtain the infinite dimensional estimate through Theorem 5.2 and finally by Proposition 4.1 get that $\mu(\hat{A}) = 0$, that is a contradiction with $\mu(\hat{A}) > 0$. In the same way, following Remark 6.2, we obtain that $\nu(b(\mathcal{T})) = 0$. □

It follows that the disintegration formula (3.10) holds true on the whole transportation set:

$$\mu = \int \mu_y m(dy), \qquad \nu = \int \nu_y m(dy).$$

Proposition 7.2. *For m-a.e. $y \in \mathcal{S}$ the conditional probabilities μ_y and ν_y have no atoms.*

Proof. We only prove the claim for μ_y.

Step 1. Suppose by contradiction that there exist a measurable set $\hat{\mathcal{S}} \subset \mathcal{S}$ such that $m(\hat{\mathcal{S}}) > 0$ and for every $y \in \hat{\mathcal{S}}$ there exists $x(y)$ such that $\mu_y(\{x(y)\}) > 0$. Restrict and normalize both μ and ν to $R(\hat{\mathcal{S}})$, and denote them again with μ and ν.

Consider the sets $K_{i,M} := \{x \in X : \rho_i \leq M\}$ for $i = 1, 2$. Note that $\mu(K_{1,M}) \geq 1 - c_1(M)$ and $\nu(K_{2,\delta}) \geq 1 - c_2(M)$ with $c_i(M) \to 0$ as $M \nearrow +\infty$. Hence for M sufficiently large the conditional probabilities of the disintegration of $\mu\llcorner_{K_{1,M}}$ have atoms, therefore we can assume, possibly restricting $\hat{\mathcal{S}}$, that for all $y \in \hat{\mathcal{S}}$ it holds $x(y) \in K_{1,M}$.

Step 2. Define

$$\mu_{y,M} := \mu_y\llcorner_{K_{1,M}}, \qquad \nu_{y,M} := \nu_y\llcorner_{K_{2,M}},$$

and introduce the set

$$D(N) := \Big\{ y \in \hat{\mathcal{S}} : \frac{\mu_{y,M}(R(y))}{\nu_{y,M}(R(y))} \leq N \Big\}.$$

Then for sufficiently large N, $m(D(N)) > 0$. The map $D(N) \ni y \mapsto h(y) := \nu_{y,M}(R(y))/\mu_{y,M}(R(y)) \leq N$ permits to define

$$\hat{\mu} := \int_{D(N)} h(y)\mu_{y,M} m(dy), \qquad \hat{\nu} := \nu\llcorner_{R(D(N))\cap K_{2,M}}.$$

It follows that $\hat{\mu}$ and $\hat{\nu}$ have bounded densities w.r.t. γ and the set $\hat{\mathcal{T}} := \mathcal{T} \cap G(K_{1,\delta}) \cap G^{-1}(K_{2,\delta})$ is a transport set for the transport problem between $\hat{\mu}$ and $\hat{\nu}$.

It follows from Theorem 5.2 and Theorem 6.1 that $\hat{\gamma} := \gamma\llcorner_{\hat{\mathcal{T}}}$ verifies Assumption 1 w.r.t. $G \cap K_{1,M} \times X \cap X \times K_{2,M}$. Therefore from Proposition 4.2 follows that the conditional probabilities $\hat{\mu}_y$ of the disintegration of $\hat{\mu}$ are continuous. Since $\hat{\mu}_y = c(y)\mu_y\llcorner_{\hat{\mathcal{T}}}$ for some positive constant $c(y)$, we have a contradiction. □

It follows straightforwardly the existence of an optimal invertible transport map.

Theorem 7.3. *Let $\mu, \nu \in \mathcal{P}(X)$ absolute continuous w.r.t. γ and assume that there exists $\pi \in \Pi(\mu, \nu)$ such that $\mathcal{I}(\pi)$ is finite. Then there exists a solution for the Monge minimization problem*

$$\min_{T : T_\sharp \mu = \nu} \int_X \|x - T(x)\|_{H(\gamma)} \mu(dx).$$

Moreover we can find T μ-essentially invertible.

Proof. For m-a.e $y \in \mathcal{S}$ μ_y and ν_y are continuous. Since $R(y)$ is one dimensional and the ray map $\mathbb{R} \ni t \mapsto g(t, y)$ is an isometry w.r.t. $\|\cdot\|_{H(\gamma)}$, we can define the non atomic measures $g(y,\cdot)_\sharp \mu_y, g(y,\cdot)_\sharp \nu_y \in \mathcal{P}(\mathbb{R})$. By the one-dimensional theory, there exists a monotone map $T_y : \mathbb{R} \to \mathbb{R}$ such that

$$T_{y\,\sharp}\Big(g(y,\cdot)_\sharp \mu_y\Big) = g(y,\cdot)_\sharp \nu_y.$$

Using the inverse of the ray map, we can define T_y on $R(y)$. Hence for m-a.e. $y \in \mathcal{S}$ we have a $\| \cdot \|_{H(\gamma)}$-cyclically monotone map T_y such that $T_{y\,\sharp}\mu_y = \nu_y$. To conclude define $T : \mathcal{T} \to \mathcal{T}$ such that $T = T_y$ on $R(y)$. Indeed T is μ-measurable, μ-essentially invertible and $T_\sharp \mu = \nu$. For the details, see the proof of Theorem 6.2 of [6]. □

References

[1] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows in metric spaces and in the space of probability measures.* Birkhäuser Verlag, 2005.

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

[3] L. Ambrosio and A. Pratelli. Existence and stability results in the L^1 theory of optimal transportation. In Springer, editor, *Optimal Transportation and Applications*, volume 1813 of *Lecture notes in mathematics*, 2001.

[4] L. Ambrosio and P.Tilli. *Topics on analysis in metric spaces.* Oxford Lecture series in Mathematics and its application, 2004.

[5] S. Bianchini and L. Caravenna. On the extremality, uniqueness and optimality of transference plans. *Bull. Inst. Math. Acad. Sin.(N.S.)*, 4(4):353–454, 2009.

[6] S. Bianchini and F. Cavalletti. The Monge problem for distance cost in geodesic spaces. preprint SISSA 50/2009/M, arXiv:1103.2796, 2009.

[7] V. I. Bogachev. *Gaussian Measures.* Mathematical surveys and monographs. AMS, 1998.

[8] 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.

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

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

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

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

[13] D. Feyel and A. S. Üstunel. Monge-Kantorovich measure transportation and Monge-Ampére equation on Wiener space. *Probab. Theory Related Fields*, 128:347–385, 2004.

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

[15] 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.

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

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

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

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

SISSA, via Bonomea 265, IT-34136 Trieste (ITALY)