



HAL
open science

On the Existence of Monge Maps for the Gromov-Wasserstein Distance

Theo Dumont, Théo Lacombe, François-Xavier Vialard

► **To cite this version:**

Theo Dumont, Théo Lacombe, François-Xavier Vialard. On the Existence of Monge Maps for the Gromov-Wasserstein Distance. 2022. hal-03818500v1

HAL Id: hal-03818500

<https://hal.science/hal-03818500v1>

Preprint submitted on 18 Oct 2022 (v1), last revised 6 Dec 2022 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ON THE EXISTENCE OF MONGE MAPS FOR THE GROMOV–WASSERSTEIN DISTANCE

Théo Dumont, Théo Lacombe and François-Xavier Vialard *

Laboratoire d’Informatique Gaspard Monge,
Univ. Gustave Eiffel, CNRS, LIGM, F-77454 Marne-la-Vallée, France

Abstract

For the L^2 -Gromov–Wasserstein distance, we study the structure of minimizers in Euclidean spaces for two different costs. The first cost is the scalar product for which we prove that it is always possible to find optimizers as Monge maps and we detail the structure of such optimal maps. The second cost is the squared Euclidean distance for which we show that the worst case scenario is the existence of a map/anti-map structure. Both results are direct and indirect consequences of an existence result on costs that are defined by submersions. In dimension one for the squared distance, we show that a monotone map is optimal in some non-symmetric situations, thereby giving insight on why such a map is often found optimal in numerical experiments. In addition, we show numerical evidence for a negative answer to the existence of a Monge map under the conditions of Brenier’s theorem, suggesting that our result cannot be improved in general.

Contents

1	Introduction	2
1.1	The Gromov–Wasserstein problem	2
1.1.1	Formulation	2
1.1.2	Relation with the optimal transportation problem: a tight bi-convex relaxation	3
1.2	Related works	5
1.2.1	Monge maps for the OT problem	5
1.2.2	Monge maps for the GW problem	7
1.3	Outline and Contributions	7
2	Existence of Monge maps for fiber-invariant costs	8
2.1	Statement of the results	8
2.2	Proof of Theorem 2	10
2.3	Proof of Theorem 3	14
3	Applications to the quadratic and inner-product GW problems	15
3.1	The inner-product cost	15
3.2	The quadratic cost	17
3.3	Complementary study of the quadratic cost in the one-dimensional case	18
3.3.1	Adversarial computation of non-monotone optimal correspondence plans	20
3.3.2	Empirical instability of the optimality of monotone rearrangements	24
3.3.3	A positive result for measures with two components	24

*Email addresses: theo.dumont@minesparis.psl.eu, theo.lacombe@univ-eiffel.fr, francois-xavier.vialard@u-pem.fr.

1 Introduction

Finding correspondences between objects that do not live on the same metric space is a problem of fundamental interest both in application and theory, in very different fields such as computer vision [BBM05, Mém11], mathematics [Stu12], biology [DSS⁺20] and machine learning [TRFC20, AMJ18]. The problem of graph matching [ZDIT16] is a prominent example of such a situation.

On the mathematical side, comparing metric measured spaces has attracted interest [Stu06, Stu12, PM22] over the past decade. A common answer to such problems consists in seeking for a map between the two objects that is of low distortion. In the case of measures that live on a metric space, this distortion is measured in terms of the distances. To make the problem well-posed and symmetric, the problem is relaxed to a superposition of deterministic maps, such as in optimal transport and they are called plans or couplings [San15]. In optimal transport, the fact that the optimization can actually be reduced to the space of maps has been developed a lot since Brenier’s work [Bre87] and further generalized [McC01]. Brenier’s result essentially states that for the quadratic cost in Euclidean spaces, the optimal map is given by the gradient of a convex function. Such results on the structure of optimal plans/maps are of great interest in order to reduce the optimization set [MTOL20]. In stark contrast to optimal transport which is a linear programming problem, the formulation of the problems mentioned above falls in the class of quadratic assignment problem [KB57], which is a computationally harder problem. As a consequence, it is not surprising that less results are available in the literature. In fact, the problem of understanding the structure of optimal plans, and when they are actual maps has been proposed by Sturm in [Stu12, Challenge 3.6]. In this work, we address this question in two particular cases in Euclidean spaces. The first one is when the distortion is measured in terms of the scalar product; we show the existence of optimal maps and we detail their structure. The second case is the quadratic squared distance for which the problem seems to have less structure; we show that optimal plans can be chosen to be supported by the union of a graph and an anti-graph of maps. We also study further the one-dimensional case, which has attracted recent attention [Vay20, BHS22]. Indeed, in the latter article, a counter-example is given to the fact that the monotone (increasing or decreasing) mapping is optimal in the discrete case. We improve on these results in two directions. First by showing that this alternative (monotone increasing/decreasing) is actually true under some conditions on the measures and second we provide numerical evidence for a counter-example to the existence of optimal maps between a density and a measure. We refer the reader to Section 1.3 for a detailed account of our contributions, while the background and state-of-the-art are presented respectively in Sections 1.1 and 1.2.

1.1 The Gromov–Wasserstein problem

1.1.1 Formulation. The Gromov–Wasserstein (GW) problem, initially introduced in [Mém11], can be seen as an extension of the Gromov–Hausdorff distance [GKPS99], see also [Stu06] for a similar extension, to the context of measured spaces (X, μ) equipped with a cost function $c_X : X \times X \rightarrow \mathbb{R}$ (typically, c_X can be a distance on X). Given (X, μ) and (Y, ν) equipped with costs c_X, c_Y respectively, and random variables $X, X' \sim \mu$ and $Y, Y' \sim \nu$, the GW problem seeks a correspondence (*i.e.* a joint law) between μ and ν that would make the distribution $c_X(X, X')$ as close as possible to $c_Y(Y, Y')$, in a L^p sense. Formally, it reads

Definition 1. *Let X and Y be Polish spaces and $p \geq 1$. Given two probability measures $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$, two continuous symmetric functions $c_X : X \times X \rightarrow \mathbb{R}$ and $c_Y : Y \times Y \rightarrow \mathbb{R}$, the p -Gromov–Wasserstein problem aims at finding*

$$\text{GW}_p(\mu, \nu) = \inf_{\pi \in \Pi(\mu, \nu)} \left(\int_{X \times Y} \int_{X \times Y} |c_X(x, x') - c_Y(y, y')|^p d\pi(x, y) d\pi(x', y') \right)^{1/p}, \quad (\text{GW})$$

where $\Pi(\mu, \nu)$ denotes the subset of $\mathcal{P}(\mathcal{X} \times \mathcal{Y})$ of probability measures that admits μ (resp. ν) as first (resp. second) marginal. Any π^* minimizing (GW) is said to be an optimal correspondence plan between μ and ν . Whenever π^* can be written as $\pi^* = (\text{id}, T)_\# \mu$ where $T : \mathcal{X} \rightarrow \mathcal{Y}$ measurable satisfies for all Borel A , $T_\# \mu(A) \triangleq \mu(T^{-1}(A)) = \nu(A)$, T is said to be an optimal correspondence map, or a Monge map between μ and ν .

While the existence of optimal correspondence plans holds by compactness arguments as long as the above minimum is not $+\infty$, much less is known about the existence of optimal correspondence maps, even in simple cases.

In this work, we will consider two specific instances of this problem, both assuming that $\mathcal{X} \subset \mathbb{R}^n$ and $\mathcal{Y} \subset \mathbb{R}^d$ for two integers $n \geq d$ and using $p = 2$:

- (i) the *inner product case*, where $c_{\mathcal{X}}$ and $c_{\mathcal{Y}}$ denote the inner products on \mathbb{R}^n and \mathbb{R}^d (both denoted by $\langle \cdot, \cdot \rangle$), respectively:

$$\min_{\pi \in \Pi(\mu, \nu)} \int_{\mathcal{X} \times \mathcal{Y}} \int_{\mathcal{X} \times \mathcal{Y}} |\langle x, x' \rangle - \langle y, y' \rangle|^2 d\pi(x, y) d\pi(x', y'), \quad (\text{GW-IP})$$

which essentially compares distribution of angles in (\mathcal{X}, μ) and (\mathcal{Y}, ν) ;

- (ii) the *quadratic case*, where $c_{\mathcal{X}}$ and $c_{\mathcal{Y}}$ are the squared Euclidean distance on \mathbb{R}^n and \mathbb{R}^d , respectively:

$$\min_{\pi \in \Pi(\mu, \nu)} \int_{\mathcal{X} \times \mathcal{Y}} \int_{\mathcal{X} \times \mathcal{Y}} \|x - x'\|^2 - \|y - y'\|^2 d\pi(x, y) d\pi(x', y'), \quad (\text{GW-Q})$$

where by $\|\cdot\|$ we mean $\|\cdot\|_2$, notation that we keep in the rest of the paper for the sake of clarity. This choice for $c_{\mathcal{X}}$ and $c_{\mathcal{Y}}$ is standard as we have the following property: if $\text{GW}_p(\mu, \nu) = 0$, the measured metric spaces (mms) $(\mathcal{X}, c_{\mathcal{X}}, \mu)$ and $(\mathcal{Y}, c_{\mathcal{Y}}, \nu)$ are *strongly isomorphic*, that is there exists an isometry $\varphi : (\mathcal{X}, d_{\mathcal{X}}) \rightarrow (\mathcal{Y}, d_{\mathcal{Y}})$ such that $\varphi_\# \mu = \nu$, see [Vay20]. A subcase of this problem is given when $\mu = \frac{1}{N} \sum_{i=1}^N \delta_{x_i}$ and $\nu = \frac{1}{N} \sum_{j=1}^N \delta_{y_j}$ are uniform probability distributions supported on N points each. In this scenario, optimal correspondence plans π can be chosen as permutations σ of $\{1, \dots, N\}$ [Vay20, Thm. 4.1.2], and the problem optimizes over the set of such permutations \mathfrak{S}_N ,

$$\min_{\sigma \in \mathfrak{S}_N} \sum_{i,j} \left| \|x_i - x_j\|^2 - \|y_{\sigma(i)} - y_{\sigma(j)}\|^2 \right|^2, \quad (\text{QAP})$$

which is a particular case of the *Quadratic Assignment Problem* (QAP) introduced in [KB57].

1.1.2 Relation with the optimal transportation problem: a tight bi-convex relaxation. Let us first recall the formulation of the Optimal Transportation (OT) problem, also known as the Kantorovich problem, that will play an extensive role in this work.

Definition 2 (Kantorovich problem). *Given two probability measures $\mu \in \mathcal{P}(\mathcal{X})$ and $\nu \in \mathcal{P}(\mathcal{Y})$ and a cost function $c : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R} \cup \{\infty\}$, we consider the problem*

$$\min_{\pi \in \Pi(\mu, \nu)} \int_{\mathcal{X} \times \mathcal{Y}} c(x, y) d\pi(x, y). \quad (\text{OT})$$

A transport plan $\pi \in \Pi(\mu, \nu)$ realizing (OT) is called an optimal transport plan, or optimal coupling. Whenever it can be written as $(\text{id}, T)_\# \mu$ for some map $T : \mathcal{X} \rightarrow \mathcal{Y}$, T is said to be an optimal transport map, or a Monge map between μ and ν for the cost c .

The minimization problem in (GW) can be interpreted as the minimization of the map $\pi \mapsto F(\pi, \pi) \triangleq \iint k d\pi \otimes \pi$ where $k((x, y), (x', y')) = |c_{\mathcal{X}}(x, x') - c_{\mathcal{Y}}(y, y')|^2$, and F is thus a symmetric

bilinear map. By first order condition, if π^\star minimizes (GW), then it also minimizes $\pi \mapsto 2F(\pi, \pi^\star)$. If we let $C_{\pi^\star}(x, y) = \int_{\mathcal{X} \times \mathcal{Y}} k((x, y), (x', y')) d\pi^\star(x', y')$, we obtain the linear problem

$$\min_{\pi \in \Pi(\mu, \nu)} \int_{\mathcal{X} \times \mathcal{Y}} C_{\pi^\star}(x, y) d\pi(x, y), \quad (1)$$

which is nothing but the (OT) problem induced by the cost C_{π^\star} on $\mathcal{X} \times \mathcal{Y}$. Therefore, we obtain that any optimal *correspondence* plan for (GW) with costs c_X, c_Y must be an optimal *transportation* plan for (OT) with cost C_{π^\star} . A crucial point, proved in [SVP21, Thm. 3] as a generalization of [Kon76], is that if k is symmetric negative on the set of (signed) measures on $\mathcal{X} \times \mathcal{Y}$ with null marginals, that is $\iint k d\alpha \otimes \alpha \leq 0$ for all such α , then the converse implication holds: any solution $\gamma^\star \in \Pi(\mu, \nu)$ of the OT problem with cost C_{π^\star} is also a solution of the GW problem, that is $F(\pi^\star, \pi^\star) = F(\gamma^\star, \gamma^\star) = F(\pi^\star, \gamma^\star)$. In fact, when k is symmetric negative, the function $F(\pi, \pi)$ is concave in π and the Gromov–Wasserstein problem falls in the category of concave *minimization* problems on a convex set. As an immediate consequence, in the finite case where the measures are sum of Dirac masses of equal mass, there exists an optimum which is a permutation matrix. In [ML18], this property is indeed used to propose concave relaxation of the Euclidean graph matching problem. Since the solutions of (GW) are in correspondence with the solutions of an OT problem, the tools and knowledge from optimal transportation can be used to derive existence and structure of optimal maps since it has been extensively studied, see Section 1.2.

In particular, this holds for our two problems of interest (GW-Q) and (GW-IP): if α denotes a signed measure on $\mathcal{X} \times \mathcal{Y} \subset \mathbb{R}^n \times \mathbb{R}^d$ with 0 marginals, observe that

$$\begin{aligned} & \int \| |x - x'|^2 - |y - y'|^2 \|^2 d\alpha(x, y) d\alpha(x', y') \\ &= \underbrace{\int |x - x'|^2 d\alpha \otimes \alpha}_{=0} + \underbrace{\int |y - y'|^2 d\alpha \otimes \alpha}_{=0} - 2 \int |x - x'|^2 |y - y'|^2 d\alpha \otimes \alpha \\ &= -2 \int (|x|^2 - 2\langle x, x' \rangle + |x'|^2)(|y|^2 - 2\langle y, y' \rangle + |y'|^2) d\alpha \otimes \alpha. \end{aligned}$$

Developing the remaining factor involve nine terms, but given that α has zero marginals (in particular, zero mass), we obtain that $\int |x|^2 |y|^2 d\alpha \otimes \alpha = 0$ (and similarly for the terms involving $|x'|^2 |y'|^2$, $|x|^2 |y'|^2$ and $|x'|^2 |y|^2$), and also that $\int |x|^2 \langle y, y' \rangle d\alpha \otimes \alpha = 0$ (and similarly for the other terms). Eventually, the only remaining term is

$$-8 \int \langle x, x' \rangle \langle y, y' \rangle d\alpha \otimes \alpha = -8 \left\| \int x \otimes y d\alpha(x, y) \right\|_F^2 \leq 0,$$

where $x \otimes y \in \mathbb{R}^{n \times d}$ is the matrix $(x_i y_j)_{i,j}$, where $x = (x_1, \dots, x_n)$ and $y = (y_1, \dots, y_d)$, and $\|\cdot\|_F$ denotes the Froebenius norm of a matrix. The negativity of this term ensures that solutions of (GW-Q) are exactly the solutions of an OT problem. Computations for (GW-IP) are similar—actually, they immediately boil down to the same last two equalities. More generally, when one considers a cost such as $(d_X(x, x') - d_Y(y, y'))^2$, by expanding the square, the only term that matters in the optimization is $-2d_X(x, x')d_Y(y, y')$. Let us assume that it is possible to write both distances d_X and d_Y as squared distances in Hilbert spaces, namely $d_X(x, x') = \|\varphi(x) - \varphi(x')\|_{H_X}^2$ and $d_Y(y, y') = \|\psi(y) - \psi(y')\|_{H_Y}^2$ for an embedding $\varphi : \mathcal{X} \rightarrow H_X$ in a Hilbert space H_X and similarly for \mathcal{Y} . Then computation (2) holds in this case. Such a property depends on the metric space and when it is satisfied the metric space is said to be of negative type or that the distance is Hilbertian. Another equivalent formulation is to say that d_X is a conditionally negative kernel on \mathcal{X} . We refer to [Lyo13] for a thorough discussion.

Definition 3. A function $k_X : X \times X \rightarrow \mathbb{R}$ is a conditionally negative definite kernel if it is symmetric and for all $N \geq 1$, $x_1, \dots, x_N \in X$ and $\omega_1, \dots, \omega_N \in \mathbb{R}$ such that $\sum_{i=1}^N \omega_i = 0$, $\sum_{i,j \leq N} \omega_i \omega_j k_X(x_i, x_j) \leq 0$.

Every conditionally positive kernel can be written as $k_X(x, x') = f(x) + f(x') - \frac{1}{2} \|\varphi(x) - \varphi(x')\|_H^2$ for an embedding $\varphi : X \rightarrow H$ a Hilbert space, as shown in [Sch38]. With respect to the Gromov-Wasserstein functional, our discussion above shows that in fact c_X can actually be replaced with a kernel which is conditionally negative definite and that the relaxation still holds. To sum up our review of the literature,

Proposition 1. Let (X, k_X, μ) and (Y, k_Y, ν) be two spaces endowed each with a conditionally positive kernel and a probability measure; then the bi-convex relaxation of GW_2^2 is tight. The corresponding kernel $k((x, y), (x', y'))$ is indeed non-positive on signed measures with null marginals on $X \times Y$.

Remark that the problem of minimizing $F(\pi, \gamma)$ is indeed a bi-convex problem since it is linear in each variable π, γ . There are several important Riemannian manifolds which are of negative type, among them the real Hyperbolic space, the sphere and the Euclidean space. Counter-examples are for instance in finite dimension the Hyperbolic space on the quaternions [FH74], and in infinite dimension the L^2 -Wasserstein distance in \mathbb{R}^d for $d \geq 3$ as proven in [ANN18].

1.2 Related works

1.2.1 Monge maps for the OT problem. The (OT) problem has been extensively studied (see [San15, Vil08, PC19] for a thorough introduction) and particular attention has been devoted to situations where existence of Monge maps, or variation of, can be ensured.

Brenier's theorem, stated below, is the most well-known of such cases where the optimal plan is a map.

Theorem 1 (Brenier's theorem). Let $X = Y = \mathbb{R}^d$, $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ such that the optimal cost between μ and ν is finite and $c(x, y) = |x - y|^2$. If $\mu \ll \mathcal{L}^d$, then there exists a unique (up to a set of μ -measure zero) solution of (OT) and it is induced by a map T . This map is characterized by being the unique gradient of a convex function $T = \nabla f$ such that $(\nabla f)_\# \mu = \nu$.

This central result admits a generalization in the manifold setting that we shall use later on.

Proposition 2 ([Vil08, Thm. 10.41], Solution of the Monge problem for the square distance). Let M be a Riemannian manifold, and $c(x, y) = d(x, y)^2$. Let $\mu, \nu \in \mathcal{P}(M)$ such that the optimal cost between μ and ν is finite. If $\mu \ll \text{vol}_M$, then there is a unique solution of the Monge problem between μ and ν and it can be written as

$$y = T(x) = \exp_x(\tilde{\nabla} f(x)),$$

where f is some $d^2/2$ -convex function. The approximate gradient can be replaced by a true gradient if any one of the following conditions is satisfied:

- (a) μ and ν are compactly supported;
- (b) M has nonnegative sectional curvature;
- (c) ν is compactly supported and M has asymptotically nonnegative curvature.

Brenier's theorem can be extended in a few directions. The condition that μ has a density can be weakened to the fact that it does not give mass to sets of Hausdorff dimension smaller than $d - 1$ (e.g. hypersurfaces), and c can actually be a bit more general than being the squared distance function, as long as it satisfies the *twist condition*, that we define now together with its variants. In the following, let $X = Y$ be complete Riemannian manifolds and let $c : X \times Y \rightarrow \mathbb{R}$ be a continuous cost function, differentiable in x . We refer to [MG11, CMN10, Vil08] for more information on the twist condition, to [AKM11, McC12] on the subtweak condition and to [Moa16] on the m -twist and generalized twist conditions.

Proposition 3 (Twist). We say that c satisfies the *twist condition* if

$$\text{for all } x_0 \in \mathcal{X}, \quad y \mapsto \nabla_x c(x_0, y) \in T_{x_0} \mathcal{X} \text{ is injective.} \quad (\text{Twist})$$

Suppose that c satisfies **(Twist)** and assume that any c -concave function is differentiable μ -a.e. on its domain. If μ and ν have finite transport cost, then **(OT)** admits a unique optimal transport plan π^* and it is induced by a map which is the gradient of a c -convex function $f : \mathcal{X} \rightarrow \mathbb{R}$:

$$\pi^* = (\text{id}, c\text{-exp}_x(\nabla f))_* \mu.$$

REMARK 1. Following [MG11, Vil08], we recall that the c -exponential map is defined on the image of $-\nabla_x c$ by the formula $c\text{-exp}_x(p) = (\nabla_x c)^{-1}(x, -p)$, i.e. $c\text{-exp}_x(p)$ is the unique y such that $\nabla_x c(x, y) + p = 0$. This notion particularizes into the usual Riemannian exponential map when $c(x, y) = d(x, y)^2/2$.

REMARK 2. Costs of the form $c(x, y) = h(x - y)$ with h strictly convex, and in particular the costs $c(x, y) = |x - y|^p$ for $p > 1$, do satisfy the twist condition.

The twist condition is equivalent to the fact that for all $y_1 \neq y_2 \in \mathcal{Y}$, the function $x \in \mathcal{X} \mapsto c(x, y_1) - c(x, y_2)$ has no critical point. Remark that on a compact manifold, if the cost is C^1 , this condition can never be satisfied. Note also that the Riemannian distance squared is not C^1 everywhere and one can still prove the existence of Monge map; see Proposition 2. This justifies the introduction of two weaker notions, that turns out to remain sufficient to obtain some (but less) structure on the optimal plans:

Proposition 4 (Subtwist). We say that c satisfies the *subtwist condition* if

$$\text{for all } y_1 \neq y_2 \in \mathcal{Y}, \quad x \in \mathcal{X} \mapsto c(x, y_1) - c(x, y_2) \quad \text{has at most 2 critical points.} \quad (\text{Subtwist})$$

Suppose that c satisfies **(Subtwist)** and assume that any c -concave function is differentiable μ -a.e. on its domain. If μ and ν have finite transport cost, then **(OT)** admits a unique optimal transport plan π^* and it is induced by the union of a map and an anti-map:

$$\pi^* = (\text{id}, G)_\# \bar{\mu} + (H, \text{id})_\# (\nu - G_\# \bar{\mu})$$

for some Borel measurable maps $G : \mathcal{X} \rightarrow \mathcal{Y}$ and $H : \mathcal{Y} \rightarrow \mathcal{X}$ and non-negative measure $\bar{\mu} \leq \mu$ such that $\nu - G_\# \bar{\mu}$ vanishes on the range of G .

Proposition 5 (m -twist). We say that c satisfies a *m -twist (resp. generalized twist) condition* if

$$\text{for all } x_0 \in \mathcal{X}, y_0 \in \mathcal{Y}, \text{ the set } \{y \mid \nabla_x c(x_0, y) = \nabla_x c(x_0, y_0)\} \text{ has at most } m \text{ elements} \quad (m\text{-twist})$$

(resp. is a finite subset of \mathcal{Y}). Suppose that c is bounded, satisfies **(m -twist)** and assume that any c -concave function is differentiable μ -almost surely on its domain. If μ has not atom and μ and ν have finite transport cost, then each optimal plan π^* of **(OT)** is supported on the graphs of $k \in \llbracket m \rrbracket$ (resp. in $\mathbb{N} \cup \{\infty\}$) measurable maps, i.e. there exists a sequence $\{\alpha_i\}_{i=1}^k$ of non-negative functions from \mathcal{X} to $[0, 1]$ and Borel measurable maps $T_i : \mathcal{X} \rightarrow \mathcal{Y}$ such that

$$\pi^* = \sum_{i=1}^k \alpha_i (\text{id}, T_i)_\# \mu,$$

in the sense $\pi^*(S) = \sum_{i=1}^k \int_{\mathcal{X}} \alpha_i(x) 1_S(x, T_i(x)) d\mu$ for any Borel $S \subset \mathcal{X} \times \mathcal{Y}$.

EXAMPLE 1. If $\mathcal{X} = \mathcal{Y} = \mathbb{R}$, and $c(x, y)$ is a second order polynomial in xy with non-zero degree one coefficient, such as $c(x, y) = x^2 y^2 + \lambda xy$ for some $\lambda \neq 0$, the 2-twist condition holds. As we shall see in Section 3.3, such costs are closely related to the quadratic GW problem **(GW-Q)** in dimension 1.

REMARK 3. Notice that although the m -twist condition is a generalization of the twist condition (which is the 1-twist condition since y_0 is always in the set), it is not a generalization of the subtwist condition.

REMARK 4. Following [Vil08, Rem. 10.33], when measures μ and ν have compact support and μ has a density—which belong to our set of assumptions in the following—, all conditions of Propositions 3 to 5 are satisfied.

1.2.2 Monge maps for the GW problem. In sharp contrast with the optimal transportation problem, there are very few results that ensure the existence of a Monge map for the Gromov–Wasserstein problem, even in the particular cases considered in this work.

In the inner product case, [Vay20, Thm. 4.2.3] gives a result on the existence of a Monge map under some assumptions:

Proposition 6 (Inner product cost: optimal map under condition). Let $n \geq d$, $\mu, \nu \in \mathcal{P}(\mathbb{R}^n) \times \mathcal{P}(\mathbb{R}^d)$ two measures of finite second order moment with $\mu \ll \mathcal{L}^n$. Suppose that there exists π^\star solution of (GW-IP) such that $M^\star = \int y \otimes x \, d\pi^\star(x, y)$ is of full rank. Then there exists an optimal map between μ and ν that can be written as $T = \nabla f \circ M^\star$ with $f : \mathbb{R}^d \rightarrow \mathbb{R}$ convex.

For the quadratic case, there is only very little results. In [Vay20] is claimed that in the discrete case in dimension 1 with uniform mass and same number of points N , the optimal solution of (QAP) would be either the identity $\sigma(i) = i$ or the anti-identity $\sigma(i) = N + 1 - i$ (Thm. 4.1.1). However, a counter-example to this claim has been recently provided by [BHS22].

To the best of our knowledge, the only positive results on the existence of Monge maps for the quadratic cost are the following.

Proposition 7 ([Stu12, Thm. 9.21]). Let absolutely continuous probability measures μ_0 and μ_1 on \mathbb{R}^n be given, each of them being rotationally invariant around its barycenter z_0 or z_1 resp., that is, $(U_i)_\# \mu_i = \mu_i$ for each $U \in O(n)$ and $i = 0, 1$ where $U_i(x) \triangleq U(x - z_i) + z_i$. Then every $\pi \in \Pi(\mu_0, \mu_1)$ which minimizes (GW-Q) is induced by a transport map T , unique up to composition with rotations. The transport map is constructed as follows: for $i = 0, 1$, let ν_i be the radial distribution of μ_i around z_i , and let F_i be the respective distribution function, i.e.

$$F_i(r) \triangleq \nu_i([0, r]) \triangleq \mu_i(\bar{B}_r(z_i)).$$

Then the monotone rearrangement $F_1 \circ F_0^{-1} : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ pushes forward ν_0 to ν_1 .

Proposition 8 ([Vay20, Prop. 4.2.4]). Let $\mu, \nu \in \mathcal{P}(\mathbb{R}^n) \times \mathcal{P}(\mathbb{R}^d)$ with compact support, with $n \geq d$. Assume that $\mu \ll \mathcal{L}^n$ and that both μ and ν are centered. Suppose that there exists π^\star solution of (GW-Q) such that $M^\star = \int y \otimes x \, d\pi^\star(x, y)$ is of full rank. Then there exists $f : \mathbb{R}^d \rightarrow \mathbb{R}$ convex such that $T = \nabla f \circ M^\star$ pushes μ to ν . Moreover, if there exists a differentiable convex $F : \mathbb{R} \rightarrow \mathbb{R}$ such that $|T(x)|_2^2 = F'(|x|_2^2)$ μ -a.e., then T is optimal for (GW-Q).

1.3 Outline and Contributions

This work is organized in the following way. Section 2 provides a general setting in which existence of optimal *transport* maps can be shown for cost that are defined by submersions. We provide two versions of the result, one (Theorem 2) which has no structure and is fairly general and one (Theorem 3) which imposes a more structured setting thus recovering more structure in the optimal maps; the latter having the benefit of being more usable in practice. The proof of the second version requires a measurability argument which is addressed in details in Proposition 11. Following the connection between GW and OT problems through the linearization result exposed in Eq. (1), applications of these general results to the Gromov–Wasserstein problems are done in Section 3 for the scalar product cost and the squared distance, both in Euclidean spaces. Finally, Section 3.3 focuses on the one-dimensional

case with quadratic cost and consists in two parts: first, we conduct a numerical exploration in order to assess if our previous structural results are sharp in dimension one; then, we prove a positive result on the optimality of monotone maps, which partly explains why a monotone map is often optimal and highlights the importance of long-range effects of the cost.

2 Existence of Monge maps for fiber-invariant costs

This section provides the main result on existence of Monge maps for OT problems for which the cost satisfies an invariance property. As detailed in Section 3, this property will be satisfied by the transport costs C_{π^*} arising from the first-order condition of (GW-Q) and (GW-IP)—see Section 1.1.2.

2.1 Statement of the results

The idea is the following: let μ, ν be two probability measures supported on a measurable space (E, Σ_E) and consider a measurable map $\varphi : E \rightarrow B$, for some measurable space (B, Σ_B) , we shall omit to mention the σ -algebra afterwards. We sometimes use the name *base space* for the space B . Let $(\mu_u)_{u \in B}$ (resp. $(\nu_u)_{u \in B}$) denote a disintegration of μ (resp. ν) with respect to φ (see Appendix A.3 for a definition of measure disintegration). Consider a cost $c : E \times E \rightarrow \mathbb{R}$ that is invariant on the fibers (that are simply the pre-image of points in the base B by φ) of φ , that is $c(x, y) = \tilde{c}(\varphi(x), \varphi(y))$ for all $x, y \in E$ and some cost function \tilde{c} on $B \times B$. Solving the OT problem between μ and ν for c boils down to the OT problem between $\varphi_{\#}\mu$ and $\varphi_{\#}\nu$ on $B \times B$ for \tilde{c} . If we can ensure that there exists a Monge map t_B between $\varphi_{\#}\mu$ and $\varphi_{\#}\nu$ (for instance, if we can use Theorem 1), we may try to build a Monge map T between μ and ν by (i) transporting each fiber μ_u onto $\nu_{t_B(u)}$ using a map T_u , and (ii) gluing the $(T_u)_{u \in B}$ together to define a measurable map T satisfying $T_{\#}\mu = \nu$ that will be optimal as it coincides with t_B on B and the cost c does not depend on the fibers $(\varphi^{-1}(u))_{u \in B}$. We stress that ensuring the measurability of the map T is non-trivial and crucial from a theoretical standpoint.

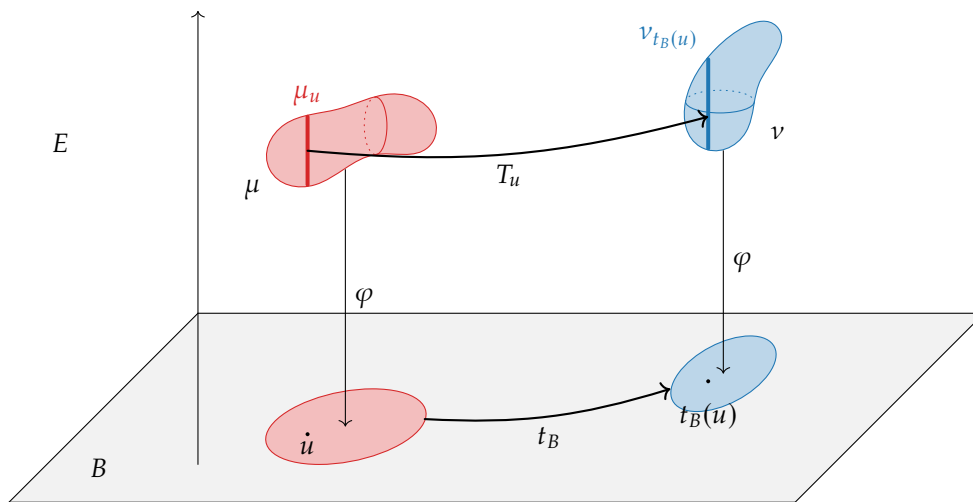


Figure 1: Illustration of the construction of the Monge map between μ and ν : we optimally transport the projections of the measures in B and then “lift” the resulting map t_B to E by sending each fiber μ_u onto the fiber $\nu_{t_B(u)}$, resulting respectively from the disintegrations of μ and ν by φ .

We formalize this idea by the mean of two theorems: the first one guarantees in a fairly general setting the existence of a Monge map for the (GW) problem, but its construction is quite convoluted and there is little to no hope that it can be leveraged in practice, either from a theoretical or computational perspective. Assuming more structure, in particular on the fibers of φ , enables the construction of a

Monge map for (GW) with a structure akin to Proposition 2. As detailed in Section 3, both (GW-Q) and (GW-IP) fall in the latter setting.

Theorem 2. Let \mathcal{X} and \mathcal{Y} be two measurable spaces for which there exists two measurable maps $\Phi_{\mathcal{X}} : \mathcal{X} \rightarrow \mathbb{R}^d$ and $\Phi_{\mathcal{Y}} : \mathcal{Y} \rightarrow \mathbb{R}^d$ that are injective, and whose inverses are measurable. Let $\mu \in \mathcal{P}(\mathcal{X})$ and $\nu \in \mathcal{P}(\mathcal{Y})$ be two probability measures. Let $c : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}$ be a cost function, and B_+, B_- be two measurable spaces along with measurable maps $\varphi : \mathcal{X} \rightarrow B_+$ and $\psi : \mathcal{Y} \rightarrow B_-$. Assume that there exists a cost $\tilde{c} : B_+ \times B_- \rightarrow \mathbb{R}$ such that

$$c(x, y) = \tilde{c}(\varphi(x), \psi(y)) \quad \text{for all } x, y \in \mathcal{X} \times \mathcal{Y}.$$

and that there exists a Monge map $t_B : B_+ \rightarrow B_-$ that transports $\varphi_{\#}\mu$ onto $\psi_{\#}\nu$ for the cost \tilde{c} . Assume that there exists a disintegration $(\mu_u)_{u \in B_+}$ of μ with respect to φ such that $\varphi_{\#}\mu$ -a.e., μ_u is atomless.

Then there exists a Monge map between μ and ν for the cost c . Furthermore, it projects onto t_B through (φ, ψ) , in sense that $(\varphi, \psi)_{\#}(\text{id}, T)_{\#}\mu = (\text{id}, t_B)_{\#}(\varphi_{\#}\mu)$.

The proof of this theorem is provided in Section 2.2.

REMARK 5. The atomless assumption on the disintegration $(\mu_u)_u$ is a natural minimal requirement to expect the existence of map (without specific assumption on the target measure ν) and implies in particular that the fibers $(\varphi^{-1}(u))_{u \in B_+}$ should not be discrete (at least $\varphi_{\#}\mu$ -a.e.). Indeed, if for instance $\mathcal{X} = \mathcal{Y} = B_+ = B_- = \mathbb{R}$ and $\varphi : x \mapsto |x|$, the fibers of φ are of the form $\{-u, u\}$ (for $u \geq 0$), hence the disintegrations $(\mu_u)_{u \geq 0}$ and $(\nu_u)_{u \geq 0}$ are discrete and given by $\mu_u(u)\delta_u + (1 - \mu_u(u))\delta_{-u}$ and $\nu_u(u)\delta_u + (1 - \nu_u(u))\delta_{-u}$, and there is in general no map T_u between two such discrete measures, unless we assume that $\mu_u(u) = \nu_u(u)$ or $1 - \nu_u(u)$, $\varphi_{\#}\mu$ -a.e.

Observe also that $\varphi_{\#}\mu$ may have atoms: as we assume the existence of the Monge map t_B , it implies in that case that $\psi_{\#}\nu$ must also have atoms.

REMARK 6. The “projection” property $(\varphi, \psi)_{\#}(\text{id}, T)_{\#}\mu = (\text{id}, t_B)_{\#}(\varphi_{\#}\mu)$ can also be written $\psi \circ T(x) = t_B \circ \varphi(x)$, for μ -a.e. x . A converse implication, that is “every Monge map between μ and ν projects onto a Monge map between $\varphi_{\#}\mu$ and $\psi_{\#}\nu$ ” may not hold in general. This is however true if we can guarantee that there is a unique optimal transport plan between $\varphi_{\#}\mu$ and $\psi_{\#}\nu$ and that it is of the form $(\text{id}, t_B)_{\#}\mu$ (e.g. if we can apply Theorem 1)—in that case, T necessarily projects onto t_B in the aforementioned sense.

Under additional assumptions, we can build a more structured Monge map. Namely, as our goal is to apply Proposition 2, we will assume that the (common) basis $B = B_+ = B_-$ is a manifold and that *almost all* the fibers of $\varphi : E \rightarrow B$ are homeomorphic to the *same* manifold F , and that every source measure of interest $(\mu, \mu_u, \varphi_{\#}\mu)$ have densities. We also introduce the following convention: if $\mu \in \mathcal{P}(E)$ for some measurable space $E, E' \subset E$, and $\varphi : E' \rightarrow B$, we let $\varphi_{\#}\mu$ be the (non-negative) measure supported on B defined by $\varphi_{\#}\mu(A) = \mu(\varphi^{-1}(A))$ for $A \subset B$ measurable. If $\mu(E') = 1$, note that $\varphi_{\#}\mu$ defines a probability measure on B (i.e. it has mass one). This formalism allows us to state our theorem even when some assumptions only hold λ -a.e.

Theorem 3. Let E_0 be a measurable space and B_0 and F be complete Riemannian manifolds. Let $\mu, \nu \in \mathcal{P}(E_0)$ be two probability measures with compact support. Assume that there exists a set $E \subset E_0$ such that $\mu(E) = 1$ and that there exists a measurable map $\Phi : E \rightarrow B_0 \times F$ that is injective and whose inverse on its image is measurable as well. Let p_B, p_F denote the projections of $B_0 \times F$ on B_0 and F respectively. Let $\varphi \triangleq p_B \circ \Phi : E \rightarrow B_0$. Let $c : E_0 \times E_0 \rightarrow \mathbb{R}$ and suppose that there exists a twisted $\tilde{c} : B_0 \times B_0 \rightarrow \mathbb{R}$ such that

$$c(x, y) = \tilde{c}(\varphi(x), \varphi(y)) \quad \text{for all } x, y \in E_0.$$

Assume that $\varphi_{\#}\mu$ is absolutely continuous w.r.t. the Lebesgue measure on B_0 and let thus t_B denote the unique Monge map between $\varphi_{\#}\mu$ and $\varphi_{\#}\nu$ for this cost. Suppose that there exists a disintegration

$((\Phi_{\#}\mu)_u)_u$ of $\Phi_{\#}\mu$ by p_B such that for $\varphi_{\#}\mu$ -a.e. u , $(\Phi_{\#}\mu)_u$ is absolutely continuous w.r.t. the volume measure on F .

Then there exists an optimal map T between μ and ν for the cost c that can be decomposed as

$$\Phi \circ T \circ \Phi^{-1}(u, v) = (t_B(u), t_F(u, v)) = (\tilde{c}\text{-exp}_u(\nabla f(u)), \text{exp}_v(\nabla h_u(v))), \quad (2)$$

with $f : B_0 \rightarrow \mathbb{R}$ \tilde{c} -convex and $h_u : F \rightarrow \mathbb{R}$ $d_F^2/2$ -convex for $\varphi_{\#}\mu$ -a.e. u . Note that t_F could actually be any function that sends each fiber $(\Phi_{\#}\mu)_u$ onto $(\Phi_{\#}\nu)_{t_B(u)}$ in a measurable way.

The proof of this theorem is provided in Section 2.3. Let us give a simple example that illustrates the role played by our assumptions. This example has connections with (GW-Q) as detailed in Section 3.2.

EXAMPLE 2. Let $E_0 = \mathbb{R}^d$ and $E = E_0 \setminus \{0\}$, let $B_0 = \mathbb{R}$ and $F = S^{d-1} = \{x \in E_0 \mid |x| = 1\}$. For convenience, we also introduce the space $B = \mathbb{R}_+^*$. Consider the cost function $c(x, y) = (|x| - |y|)^2$, so that c only depends on the norm of its entries. The fibers of the map $x \mapsto |x|$ are spheres—with the exception of $x = 0$, which invites us to consider the diffeomorphism

$$\begin{aligned} \Phi : E &\rightarrow \mathbb{R}_+^* \times S^{d-1} = B \times F \subset B_0 \times F \\ x &\mapsto \left(|x|, \frac{x}{|x|} \right). \end{aligned}$$

From this, we can write $c(x, y) = \tilde{c}(\varphi(x), \varphi(y))$ where $\varphi(x) = |x|$ and $\tilde{c}(u, u') = (u - u')^2$ (which is twisted).

Now, if μ has a density on \mathbb{R}^d , so does $\Phi_{\#}\mu$ (as Φ is a diffeomorphism) on $B_0 \times F$. The coarea formula gives the existence of a disintegration $(\mu_u)_{u \in B}$ of $\Phi_{\#}\mu$ by $p_B : (u, v) \mapsto u$ (note that $p_{B\#}(\Phi_{\#}\mu) = \varphi_{\#}\mu$ also has a density) such that all the μ_u admits a density on S^{d-1} .

Our theorem thus applies, ensuring the existence of a structured Monge map between μ and (any) ν for the cost c : it decomposes for almost all $x \in \mathbb{R}^d$ as a Monge map on the basis $B_0 = \mathbb{R}$ (although it is actually only characterized on the image of φ , that is $B = \mathbb{R}_+^*$) obtained as the gradient of a convex function f (there is no need for the exponential map here since ∇f is the increasing mapping between the quantiles of $\varphi_{\#}\mu$ and $\varphi_{\#}\nu$) and a Monge map on each fiber $F = S^{d-1}$, also built from gradients of convex functions (via the exponential map on the sphere).

Note that our theorem only requires assumptions to hold almost everywhere on $E_0 = \mathbb{R}^d$, which is important to allow us to ignore the singularity of φ at $x = 0$.

2.2 Proof of Theorem 2

The proof decomposes in three steps.

Step 1: Existence and optimality of lifts. We know by assumption that there exists a Monge map t_B that is optimal between the pushforward measures $\varphi_{\#}\mu$ and $\psi_{\#}\nu$.

As our goal is to build a Monge map between the initial measures μ and ν , we first show that (i) there exists a transport plan $\pi \in \Pi(\mu, \nu)$ such that $(\varphi, \psi)_{\#}\pi = (\text{id}, t_B)_{\#}\mu$ and (ii) any such π is an optimal transport plan between μ and ν for the cost c . This is formalized by the following lemmas.

Lemma 1 (Existence of a lift). For any transport plan $\tilde{\pi} \in \Pi(\varphi_{\#}\mu, \psi_{\#}\nu)$, there exists a transport plan $\pi \in \Pi(\mu, \nu)$ such that $(\varphi, \psi)_{\#}\pi = \tilde{\pi}$.

Proof. Let $(\mu_u)_{u \in B_+}$ and $(\nu_v)_{v \in B_-}$ be disintegrations of μ and ν by φ and ψ respectively. Given $\tilde{\pi} \in \Pi(\varphi_{\#}\mu, \psi_{\#}\nu)$, we define

$$\pi \triangleq \iint_{B_+ \times B_-} (\mu_u \otimes \nu_v) d\tilde{\pi}(u, v),$$

i.e. trivially sending every fiber μ_u onto every fiber ν_v , while weighting by $\tilde{\pi}$. See [AGS05, Sec. 5.3] for the notation. Then, for any Borel set $A \subset \mathcal{X}$,

$$\begin{aligned} \pi(A \times \mathcal{Y}) &= \iint_{B_+ \times B_-} \mu_u(A) \nu_v(\mathcal{Y}) \, d\tilde{\pi}(u, v) \\ &= \iint_{B_+ \times B_-} \mu_u(A) \, d\tilde{\pi}(u, v) \\ &= \int_{B_+} \mu_u(A) \, d(\varphi_{\#}\mu)(u) && \text{since the first marginal of } \tilde{\pi} \text{ is } \varphi_{\#}\mu \\ &= \mu(A), && \text{(disintegration theorem)} \end{aligned}$$

and similarly for v ; hence $\pi \in \Pi(\mu, \nu)$. Now, let us show that $(\varphi, \psi)_{\#}\pi = \tilde{\pi}$. For U and V Borel sets of B_+ and B_- ,

$$\begin{aligned} ((\varphi, \psi)_{\#}\pi)(U \times V) &= \iint_{U \times V} d((\varphi, \psi)_{\#}\pi)(u, v) \\ &= \iint_{\varphi^{-1}(U) \times \psi^{-1}(V)} d\pi(x, y) \\ &= \iint_{\varphi^{-1}(U) \times \psi^{-1}(V)} \iint_{B_+ \times B_-} d(\mu_u \otimes \nu_v)(x, y) \, d\tilde{\pi}(u, v) \\ &= \iint_{B_+ \times B_-} \left(\int_{\varphi^{-1}(U)} d\mu_u(x) \int_{\psi^{-1}(V)} d\nu_v(y) \right) d\tilde{\pi}(u, v) && \text{by Fubini's theorem} \\ &= \iint_{B_+ \times B_-} \mu_u(\varphi^{-1}(U)) \nu_v(\psi^{-1}(V)) \, d\tilde{\pi}(u, v) \\ &= \iint_{B_+ \times B_-} \delta_U(u) \delta_V(v) \, d\tilde{\pi}(u, v) \\ &= \iint_{U \times V} d\tilde{\pi}(u, v) \\ &= \tilde{\pi}(U \times V). \end{aligned} \quad \square$$

Lemma 2 (Decomposition of optimal plans for the base space cost). Let $c : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}$ and $\tilde{c} : B_+ \times B_- \rightarrow \mathbb{R}$ such that

$$c(x, y) = \tilde{c}(\varphi(x), \psi(y)) \quad \text{for all } x, y \in \mathcal{X} \times \mathcal{Y}.$$

Then

$$\Pi_{\tilde{c}}^*(\varphi_{\#}\mu, \psi_{\#}\nu) = (\varphi, \psi)_{\#}\Pi_c^*(\mu, \nu),$$

where $\Pi_c^*(\mu, \nu)$ denotes the set of optimal transport plan between μ and ν for the cost c , and similarly for $\Pi_{\tilde{c}}^*(\varphi_{\#}\mu, \psi_{\#}\nu)$.

Proof. Let us first remark that for every $\tilde{\pi} \in \Pi(\varphi_{\#}\mu, \psi_{\#}\nu)$ and $\pi \in \Pi(\mu, \nu)$,

$$\text{if } \tilde{\pi} = (\varphi, \psi)_{\#}\pi, \text{ then } \langle c, \pi \rangle = \langle \tilde{c}, \tilde{\pi} \rangle. \quad (3)$$

Indeed, for such a $\tilde{\pi}$

$$\begin{aligned} \iint_{B_+ \times B_-} \tilde{c}(u, v) \, d\tilde{\pi}(u, v) &= \iint_{\mathcal{X} \times \mathcal{Y}} \tilde{c}(\varphi(x), \psi(y)) \, d\pi(x, y) && \text{by definition of the pushforward} \\ &= \iint_{\mathcal{X} \times \mathcal{Y}} c(x, y) \, d\pi(x, y). \end{aligned}$$

c. Let $\tilde{\pi}^* \in \Pi_c^*(\varphi_{\#}\mu, \psi_{\#}\nu)$. By Lemma 1, there exists a $\pi \in \Pi(\mu, \nu)$ such that $(\varphi, \psi)_{\#}\pi = \tilde{\pi}^*$. Then for any $\gamma \in \Pi(\mu, \nu)$,

$$\langle c, \pi \rangle \stackrel{(3)}{=} \langle \tilde{c}, \tilde{\pi}^* \rangle \stackrel{(*)}{\leq} \langle \tilde{c}, (\varphi, \psi)_{\#}\gamma \rangle \stackrel{(3)}{=} \langle c, \gamma \rangle,$$

where $(*)$ follows from the optimality of $\tilde{\pi}^*$. Hence the optimality of π .

d. Let $\pi^* \in \Pi_c^*(\mu, \nu)$. By Lemma 1, for any $\tilde{\gamma} \in \Pi(\varphi_{\#}\mu, \psi_{\#}\nu)$ there exists a $\gamma \in \Pi(\mu, \nu)$ such that $(\varphi, \psi)_{\#}\gamma = \tilde{\gamma}$. We then have

$$\langle \tilde{c}, (\varphi, \psi)_{\#}\pi^* \rangle \stackrel{(3)}{=} \langle c, \pi^* \rangle \stackrel{(*)}{\leq} \langle c, \gamma \rangle \stackrel{(3)}{=} \langle \tilde{c}, \tilde{\gamma} \rangle,$$

where $(*)$ follows from the optimality of π^* . Hence the optimality of $(\varphi, \psi)_{\#}\pi^*$. \square

Step 2: Existence of Monge maps between the fibers. Using Lemma 1 with $\tilde{\pi} = (\text{id}, t_B)_{\#}(\varphi_{\#}\mu)$, we know that we can build an optimal transportation plan $\pi \in \Pi(\mu, \nu)$ that essentially coincides with t_B on $B_+ \times B_-$ and transports each fiber μ_u onto $\nu_{t_B(u)}$ for μ -a.e. $u \in B_+$. In order to build a Monge map between μ and ν , we must show that one can actually transport almost all μ_u onto $\nu_{t_B(u)}$ using a map rather than a plan. For this, we use the following result, see [San15, Rem. 1.23, Lemma 1.28, Cor. 1.29].

Proposition 9. Let α, β be two measures supported on \mathbb{R}^d with α atomless. Then:

- (i) if $d = 1$, there exists a transport map \tilde{T} that pushes α onto β . Furthermore, it is the *unique* optimal map between these measures for the quadratic cost $(x, y) \mapsto |x - y|^2$;
- (ii) there exists a map $\sigma_d : \mathbb{R}^d \rightarrow \mathbb{R}$ (that does not depend on α, β) that is (Borel) measurable, injective, and its inverse is measurable as well.

As we assumed that the ground spaces \mathcal{X} and \mathcal{Y} can be embedded in \mathbb{R}^d using the injective, measurable maps $\Phi_{\mathcal{X}}$ and $\Phi_{\mathcal{Y}}$, we can apply Proposition 9 using $\sigma_{\mathcal{X}} = \sigma_d \circ \Phi_{\mathcal{X}}$ and $\sigma_{\mathcal{Y}} = \sigma_d \circ \Phi_{\mathcal{Y}}$. As $\sigma_{\mathcal{X}}$ is injective, $\sigma_{\mathcal{X}} \# \mu_u$ is atomless on \mathbb{R} , and we can thus consider the *unique* Monge map \tilde{T}_u between $\sigma_{\mathcal{X}} \# \mu_u$ and $\sigma_{\mathcal{Y}} \# \nu_{t_B(u)}$ for the quadratic cost on \mathbb{R} .

From this, as the maps $\sigma_{\mathcal{X}}$ and $\sigma_{\mathcal{Y}}$ are measurable and injective (thus invertible on their image) we can define $T_u = \sigma_{\mathcal{Y}}^{-1} \circ \tilde{T}_u \circ \sigma_{\mathcal{X}} : \mathcal{X} \rightarrow \mathcal{Y}$, that defines a (measurable) transport map between μ_u and $\nu_{t_B(u)}$.

Step 3: building a measurable global map. Now that we have maps $(T_u)_u$ between each μ_u and $\nu_{t_B(u)}$, it may be tempting to simply define a map $T : \mathcal{X} \rightarrow \mathcal{Y}$ by $T(x) = T_{\varphi(x)}(x)$ when $\mu_{\varphi(x)}$ is atomless (which, by assumption, holds μ -a.e.). Intuitively, this map induces a transport plan $(\text{id}, T)_{\#}\mu$ that satisfies $(\varphi, \psi)_{\#}(\text{id}, T)_{\#}\mu = (\text{id}, t_B)_{\#}(\varphi_{\#}\mu)$ on $B_+ \times B_-$ and thus must be optimal according to Lemma 2.

One remaining step, though, is to prove that this map T can be defined in a measurable way. For this, we use the following *measurable selection theorem* due to [FGM10, Thm. 1.1], that reads:

Proposition 10. Let (B, Σ, m) be a σ -finite measure space and consider a measurable function $B \ni u \mapsto (\mu_u, \nu_u) \in \mathcal{P}(\mathbb{R}^d)^2$. Let $c : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ be a cost function, and assume that for m -a.e. $u \in B$, there is a (unique) Monge map T_u between μ_u and ν_u for the cost c .

Then there exists a measurable function $(u, x) \mapsto T(u, x)$ such that m -a.e., $T(u, x) = T_u(x)$, μ_u -a.e.

We can apply this result in the case $d = 1$ to the family of measure $(\sigma_{\mathcal{X}} \# \mu_u, \sigma_{\mathcal{Y}} \# \nu_{t_B(u)})_{u \in B_+}$, where the reference measure on B_+ is $\varphi_{\#}\mu$.¹ We first need to show the measurability of this family of measures. By definition of the disintegration of measures (see for instance [AGS05, Thm. 5.3.1]), the map $u_- \in B_- \mapsto \nu_{u_-}$ is measurable; and the Monge map t_B is measurable as well so is the map $B_+ \ni u \mapsto \sigma_{\mathcal{Y}} \# \nu_{t_B(u)}$ by composition of measurable maps, and thus the map $u \mapsto (\mu_u, \nu_{t_B(u)})$. Therefore,

¹Note that we cannot apply Proposition 10 to the measures $(\mu_u, \nu_{t_B(u)})_u$ and the maps $(T_u)_u$ directly, as T_u may not be the unique Monge map between the measures, a required assumption of the proposition.

Proposition 10 applies and guarantees the existence of a measurable map $\tilde{T} : B_+ \times \mathbb{R} \rightarrow \mathbb{R}$ such that $\tilde{T}(u, x) = \tilde{T}_u(x)$, for $\varphi_{\#}\mu$ almost all u , and $\sigma_{\mathcal{X}}\mu$ almost all x . Now, we can define

$$\begin{aligned} T : \mathcal{X} &\rightarrow \mathcal{Y} \\ x &\mapsto \sigma_{\mathcal{Y}}^{-1} \circ \tilde{T}(\varphi(x), \sigma_{\mathcal{X}}(x)). \end{aligned}$$

This map is measurable as composition of measurable maps. Let us prove that this defines a transport map between μ and ν . For any function $g : \mathcal{Y} \rightarrow \mathbb{R}$ continuous with compact support, we can write

$$\int_{\mathcal{Y}} g(y) dT_{\#}\mu(y) = \int_{\mathcal{X}} g(T(x)) d\mu(x) = \int_{u \in B_+} \int_{x \in \varphi^{-1}(\{u\})} g\left(\sigma_{\mathcal{Y}}^{-1}\left(\tilde{T}_u(\sigma_{\mathcal{X}}(x))\right)\right) d\mu_u(x) d\varphi_{\#}\mu(u),$$

where we use the disintegration of μ w.r.t. φ and the fact that the μ_u are supported on $\varphi^{-1}(\{u\})$, allowing us to write $\tilde{T}(\varphi(x), \sigma_{\mathcal{X}}(x)) = \tilde{T}_u(\sigma_{\mathcal{X}}(x))$ on that fiber ($\varphi_{\#}\mu$ -a.e.).

Now, recall that $T_u : x \mapsto \sigma_{\mathcal{Y}}^{-1}\left(\tilde{T}_u(\sigma_{\mathcal{X}}(x))\right)$ defines a transport map between μ_u and $\nu_{t_B(u)}$. In particular, the image of the fiber $\varphi^{-1}(\{u\})$ by this map is $\psi^{-1}(\{t_B(u)\}) \subset \mathcal{Y}$. Therefore, we get

$$\begin{aligned} \int_{\mathcal{Y}} g(y) dT_{\#}\mu(y) &= \int_{u \in B_+} \int_{y \in \psi^{-1}(\{t_B(u)\})} g(y) d\nu_{t_B(u)} d\varphi_{\#}\mu(u) \\ &= \int_{u \in B_+} \int_{y \in \mathcal{Y}} g(y) d\nu_{t_B(u)} d\varphi_{\#}\mu(u) && \text{as } \nu_{t_B(u)} \text{ is supported on } \psi^{-1}(\{t_B(u)\}) \\ &= \int_{v \in B_-} \int_{y \in \mathcal{Y}} g(y) d\nu_v(y) dt_{B\#}(\varphi_{\#}\mu)(v) && \text{by change of variable } v = t_B(u) \\ &= \int_{v \in B_-} \int_{y \in \mathcal{Y}} g(y) d\nu_v(y) d\psi_{\#}\nu(v) && \text{as } t_B \text{ pushes } \varphi_{\#}\mu \text{ to } \psi_{\#}\nu \\ &= \int_{y \in \mathcal{Y}} g(y) d\nu(y) && \text{as } (\nu_v)_v \text{ is a disintegration of } \nu \text{ by } \psi, \end{aligned}$$

proving that $T_{\#}\mu = \nu$.

By Lemma 2, this map is optimal if and only if it satisfies $(\varphi, \psi)_{\#}(\text{id}, T)_{\#}\mu = (\text{id}, t_B)_{\#}(\varphi_{\#}\mu)$, as t_B is an optimal transportation plan between $\varphi_{\#}\mu$ and $\psi_{\#}\nu$, making $(\text{id}, T)_{\#}\mu$ optimal between μ and ν (hence T a Monge map).

For this, let $g : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}$ be a continuous function with compact support. We have

$$\begin{aligned} \iint_{B_+ \times B_-} g(u, v) d(\varphi, \psi)_{\#}(\text{id}, T)_{\#}\mu(u, v) &= \iint_{\mathcal{X} \times \mathcal{Y}} g(\varphi(x), \psi(y)) d(\text{id}, T)_{\#}\mu(x, y) \\ &= \int_{\mathcal{X}} g(\varphi(x), \psi(T(x))) d\mu(x) \\ &= \int_{u \in B_+} \int_{x \in \varphi^{-1}(\{u\})} g(u, \psi(\sigma_{\mathcal{Y}}^{-1}(T_u(\sigma_{\mathcal{X}}(x)))) d\mu_u(x) d\varphi_{\#}\mu(u) \\ &= \int_{u \in B_+} \int_{y \in \psi^{-1}(\{t_B(u)\})} g(u, t_B(u)) d\nu_{t_B(u)}(y) d\varphi_{\#}\mu(u) \\ &= \int_{u \in B_+} g(u, t_B(u)) d\varphi_{\#}\mu(u) \\ &= \int_{B_+ \times B_-} g(u, v) d(\text{id}, t_B)_{\#}\varphi_{\#}\mu(u, v), \end{aligned}$$

proving the required equality and thus that T is a Monge map between μ and ν .

2.3 Proof of Theorem 3

To alleviate notations, we let $\mu' \triangleq \Phi_{\#}\mu$ and $\nu' \triangleq \Phi_{\#}\nu$ in the following. We also denote by B the image of $\varphi = p_B \circ \Phi$, so that μ', ν' are supported on $B \times F \subset B_0 \times F$.

Step 1: Construction of the structured Monge map. Given that $\varphi_{\#}\mu$ is absolutely continuous w.r.t. the Lebesgue measure on the complete (separable) Riemannian manifold B_0 , by Theorem 1 there exists a unique optimal transport plan π_B^* between $\varphi_{\#}\mu$ and $\varphi_{\#}\nu$ for the cost \tilde{c} and it is induced by a map $t_B : B_0 \rightarrow B_0$ of the form $t_B = \exp_u(\nabla f)$, with f locally Lipschitz and \tilde{c} -convex.

By Lemma 2, we know that any transport plan in $\pi \in \Pi(\mu, \nu)$ that satisfy $(\varphi, \varphi)_{\#}\pi = (\text{id}, t_B)_{\#}\mu$ must be optimal. Therefore, if π happens to be induced by a map T , that is $\pi = (\text{id}, T)_{\#}\mu$, we would obtain a Monge map between μ and ν . To build such a T , we proceed as in Section 2.2: we define a Monge map T_u between $(\mu'_u)_u$ and $(\nu'_{t_B(u)})_u$ (recall that those are the disintegration of $\Phi_{\#}\mu = \mu'$ and $\Phi_{\#}\nu = \nu'$ with respect to p_B), for $\varphi_{\#}\mu$ -a.e. u and build a global map between μ' and ν' by (roughly) setting $T(u, x) = T_u(x)$. As in Section 2.2, proving the measurability of such T requires care.

Step 2: Transport between the fibers. For $\varphi_{\#}\mu$ -a.e. u , μ'_u has a density w.r.t. the volume measure on F and the optimal cost between μ'_u and $\nu'_{t_B(u)}$ is finite by assumption. Whenever μ'_u has a density, we can therefore apply Proposition 2 between μ'_u and $\nu'_{t_B(u)}$ with the cost d_F^2 to obtain that there exists a plan π_u between these fibers that is induced by a map $T_u : F \rightarrow F$ that can be expressed as $T_u(v) = \exp_v(\tilde{\nabla} h_u(v))$ with h_u being $d_F^2/2$ -convex on F .

Step 3: Measurability of the global map. Now that we have built structured maps T_u between corresponding fibers (through t_B), it remains to prove the existence of a measurable map $T : B_0 \times F \rightarrow B_0 \times F$ transporting μ' onto ν' satisfying $T(u, x) = (t_B(u), T_u(x))$ for $\varphi_{\#}\mu$ -almost every u and μ'_u -almost every x .

For this, we need an adaptation of Proposition 10 to the manifold setting. Namely, we have the following:

Proposition 11 (Measurable selection of maps, manifold case). Let M be a complete Riemannian manifold and (B, Σ, m) a measured space. Consider a measurable function $B \ni u \mapsto (\mu_u, \nu_u) \in \mathcal{P}(M)^2$. Assume that for m -almost every $u \in B$, $\mu_u \ll \text{vol}_M$ and μ_u and ν_u have a finite transport cost. Let T_u denote the (unique by Proposition 2) optimal transport map induced by the quadratic cost d_M^2 on M between μ_u and ν_u .

Then there exists a function $(u, x) \mapsto T(u, x)$, measurable w.r.t. $\Sigma \otimes \mathcal{B}(\mathbb{R}^d)$, such that m -a.e.,

$$T(u, x) = T_u(x) \quad \mu_u\text{-a.e.}$$

This proposition can essentially be proved by adapting the proof of [FGM10] to the manifold setting, and most steps adapt seamlessly. We provide a sketch of proof below. A complete proof, where we stress the points that needed specific care in adaptation, is deferred to the appendix.

Sketch of proof of Proposition 11. The proof relies on theory of measurable sets-valued maps [RW09, Ch. 5 and 14]. The main steps are the following:

1. For $k \in \mathbb{N}$, let $(A_{n,k})_n$ be a partition of M of cells with M -volume lesser than or equal to 2^{-kD} (where D denotes the dimension of M) and such that $(A_{n,j})_n$ is a refinement of $(A_{n,k})_n$ for $j \geq k$, that is each $A_{n,k}$ is partitioned by a subset of the $(A_{n,j})_n$. Then, for all n, k , the set $\{(u, x), T_u(x) \in A_{n,k}\}$ is measurable.
2. Consider $a_{n,k} \in A_{n,k}$ for each n, k chosen in a measurable way. Build a sequence of measurable maps $(T^{(k)})_k$ defined by $T^{(k)} : (u, x) \mapsto a_{n,k}$ where $T_u(x) \in A_{n,k}$. This is a Cauchy sequence for the metric $D_1(f, g) \triangleq \int d_M(f(u, x), g(u, x)) d\mu_u(x) dm(u)$ for $f, g : L^1(B \times M \rightarrow M, \mu_u \otimes m)$, that

is the space of functions f such that $\int d_M(f(u, x), z) d\mu_u(x) dm(u) < \infty$ for some $z \in M$. This space is complete [Chi07], so we can consider the (measurable) map $T = \lim_k T^{(k)}$.

3. Prove that we indeed have $T(u, x) = T_u(x)$, roughly using that $T^{(k)}(u, x)$ both approximates $T_u(x)$ (by construction) and $T(u, x)$ (as the limit of the sequence). \square

We can apply this proposition with the manifold being the (common) fiber F on which the $\mu'_u, \nu'_{t_B(u)}$ are supported for $\varphi_{\#}\mu$ -a.e. u , and for which we have access to the (unique) Monge map T_u . It gives the existence of a global map t satisfying $t(u, v) = T_u(v)$ for $\varphi_{\#}\mu$ -a.e. u , and μ'_u -a.e. v , and we can thus define the (measurable) map $T(u, x) = (t_B(u), t(u, x))$.

One then has for any continuous function z with compact support:

$$\begin{aligned}
\int_{B_0 \times F} z(u', v') d(T_{\#}\Phi_{\#}\mu)(u', v') &= \int_{B_0 \times F} z(t_B(u), T_u(v)) d(\Phi_{\#}\mu)(u, v) && \text{(pushforward } T \text{ on } \Phi_{\#}\mu) \\
&= \iint_{B_0 \times F} z(t_B(u), T_u(v)) d(\Phi_{\#}\mu)_u(v) d(\varphi_{\#}\mu)(u) && \text{(disintegration theorem)} \\
&= \iint_{B_0 \times F} z(t_B(u), v') d(g_{u\#}(\Phi_{\#}\mu)_u)(v') d(\varphi_{\#}\mu)(u) && \text{(pushforward } T_u \text{ on } (\Phi_{\#}\mu)_u) \\
&= \iint_{B_0 \times F} z(t_B(u), v') d((\Phi_{\#}\nu)_{t_B(u)})(v') d(\varphi_{\#}\mu)(u) && (g_{u\#}(\Phi_{\#}\mu)_u = (\Phi_{\#}\nu)_{t_B(u)}) \\
&= \iint_{B_0 \times F} z(u', v') d((\Phi_{\#}\nu)_{u'})(v') d(t_B\varphi_{\#}\mu)(u') && \text{(pushforward } t_B \text{ on } \varphi_{\#}(\Phi_{\#}\mu)_u) \\
&= \iint_{B_0 \times F} z(u', v') d((\Phi_{\#}\nu)_{u'})(v') d(\varphi_{\#}\nu)(u') && (t_{B\#}(\varphi_{\#}\mu) = \varphi_{\#}\nu) \\
&= \int_{B_0 \times F} z(u', v') d(\Phi_{\#}\nu)(u', v'), && \text{(disintegration theorem)}
\end{aligned}$$

hence T sends $\Phi_{\#}\mu$ to $\Phi_{\#}\nu$ and $T_E \triangleq \Phi^{-1} \circ T \circ \Phi$ therefore sends μ to ν ; and since

$$(\varphi, \varphi)_{\#}(\text{id}, T_E)_{\#}\mu = (\varphi, \varphi \circ T_E)_{\#}\mu = (\varphi, t_B \circ \varphi)_{\#}\mu = (\text{id}, t_B)_{\#}\varphi_{\#}\mu = \pi_B^{\star},$$

we have that T_E is an optimal map between μ and ν .

3 Applications to the quadratic and inner-product GW problems

3.1 The inner-product cost

We recall the (GW-IP) problem:

$$\min_{\pi \in \Pi(\mu, \nu)} \int_{\mathcal{X} \times \mathcal{Y}} \int_{\mathcal{X} \times \mathcal{Y}} |\langle x, x' \rangle - \langle y, y' \rangle|^2 d\pi(x, y) d\pi(x', y'), \quad (\text{GW-IP})$$

Expanding the integrand and using the fact that $\iint \langle x, x' \rangle^2 d\pi d\pi = \iint \langle x, x' \rangle^2 d\mu d\mu$ is constant (the same goes for the terms that depend on ν), (GW-IP) is equivalent to

$$\min_{\pi \in \Pi(\mu, \nu)} \iint -\langle x, x' \rangle \langle y, y' \rangle d\pi(x, y) d\pi(x', y').$$

This problem is not invariant to translations but it is to the action of $O_n(\mathbb{R}) \times O_d(\mathbb{R})$. Assuming an optimal correspondence plan π^{\star} , this plan is also an optimal transport plan for the linearized problem (1) with cost

$$C_{\pi^{\star}}(x, y) = - \int \langle x, x' \rangle \langle y, y' \rangle d\pi^{\star}(x', y') = \left\langle - \int (y' \otimes x') x d\pi^{\star}(x', y'), y \right\rangle = -\langle M^{\star}x, y \rangle,$$

where $M^\star \triangleq \int y' \otimes x' d\pi^\star(x', y') \in \mathbb{R}^{d \times n}$. This linearized cost satisfies the (Twist) condition if and only if M^\star is of full rank, hence in this case the solution π^\star of (GW-IP) is unique and induced by a map, and [Vay20, Theorem 4.2.3] gives a result on the structure of this map. We can actually generalize this result to the case where M^\star is arbitrary:

Theorem 4 (Existence of an optimal map for the inner product cost). Let $n \geq d$, $\mu, \nu \in \mathcal{P}(\mathbb{R}^n) \times \mathcal{P}(\mathbb{R}^d)$ two measures with compact supports. Suppose that $\mu \ll \mathcal{L}^n$. Then there exists an optimal map for (GW-IP) that can be written as

$$T = O_Y^\top \circ (T_0 \circ p_{\mathbb{R}^d}) \circ O_X, \quad (4)$$

where O_X and O_Y are change-of-basis matrices of \mathbb{R}^n and \mathbb{R}^d , $p_{\mathbb{R}^d} : \mathbb{R}^n \rightarrow \mathbb{R}^d$ is defined by $p_{\mathbb{R}^d}(x_1, \dots, x_n) = (x_1, \dots, x_d)$, and

$$T_0(x_1, \dots, x_d) = (\nabla f \circ \Sigma(x_1, \dots, x_h), \nabla g_{x_1, \dots, x_h}(x_{h+1}, \dots, x_d)) \quad (5)$$

with $h \leq d$, $\Sigma \in \mathbb{R}^{h \times h}$ diagonal with positive entries, $f : \mathbb{R}^h \rightarrow \mathbb{R}$ convex and all $g_{x_1, \dots, x_h} : \mathbb{R}^{d-h} \rightarrow \mathbb{R}$ convex.

In order to show this, we will need two simple lemmas that we state now and prove in Appendix A.1, the second one being a simple corollary of the first:

Lemma 3. Let $\mu, \nu \in \mathcal{P}(E)$ and let $\psi_1, \psi_2 : E \rightarrow F$ be homeomorphisms. Let $c : F \times F \rightarrow \mathbb{R}$ and consider the cost $x, y \mapsto c(\psi_1(x), \psi_2(y))$. Then a map is optimal for the cost c between μ and ν if and only if it is of the form $\psi_2^{-1} \circ T \circ \psi_1$ with T optimal for the cost \tilde{c} between $\psi_{1\#}\mu$ and $\psi_{2\#}\nu$.

Lemma 4 (Brenier with scaled inner product). Let $h \geq 1$ and $\mu, \nu \in \mathcal{P}(\mathbb{R}^h)$ with $\mu \ll \mathcal{L}^h$ with compact supports. Consider the cost $c(x, y) = -\langle \psi_1(x), \psi_2(y) \rangle$ where $\psi_1, \psi_2 : \mathbb{R}^h \rightarrow \mathbb{R}^h$ are diffeomorphisms. Then, there exists a unique optimal transport plan between μ and ν for the cost c , and it is induced by a map $t : \mathbb{R}^h \rightarrow \mathbb{R}^h$ of the form $t = \psi_2^{-1} \circ \nabla f \circ \psi_1$, with f convex.

We are now ready to prove Theorem 4:

Proof of Theorem 4. Using a singular value decomposition, we have $M^\star = O_Y^\top \Sigma O_X \in \mathbb{R}^{d \times n}$ with $O_X, O_Y \in O_n(\mathbb{R}) \times O_d(\mathbb{R})$ orthogonal matrices of each Euclidean space and $\Sigma \in \mathbb{R}^{d \times n}$ diagonal with non-negative coefficients. The cost then becomes $C_{\pi^\star}(x, y) = -\langle O_Y^\top \Sigma O_X x, y \rangle = -\langle \Sigma(O_X x), O_Y y \rangle$. Using Lemma 3, the problem transforms into an optimal transportation problem between $\mu' \triangleq O_X \mu$ and $\nu' \triangleq O_Y \nu$; and choosing O_Y and O_X that sort the singular values in decreasing order, i.e. assuming $\sigma_1 \geq \dots \geq \sigma_h > 0$ with $h \triangleq \text{rk}(M^\star) \leq d$, the problem therefore transforms into $\min_{\tilde{\pi}} \langle c_\Sigma, \tilde{\pi} \rangle$ for $\tilde{\pi} \in \Pi(\mu', \nu')$, where $c_\Sigma(\tilde{x}, \tilde{y}) = -\sum_{i=1}^h \sigma_i \tilde{x}_i \tilde{y}_i \triangleq -\langle p(\tilde{x}), p(\tilde{y}) \rangle_\sigma$, p being the orthogonal projection on \mathbb{R}^h . We reduce to the case where both measures live in the same space by noting that since $c_\Sigma(\tilde{x}, \tilde{y}) = c_\Sigma(p_{\mathbb{R}^d}(\tilde{x}), \tilde{y})$ for all \tilde{x} and \tilde{y} , any map T_0 optimal between $\mu'' \triangleq p_{\mathbb{R}^d \#} \mu'$ and ν' will induce a map $T = T_0 \circ p_{\mathbb{R}^d}$ optimal between μ' and ν' ². One can then recover an optimal map between μ and ν by composing with O_X and O_Y^\top (Lemma 3), hence Eq. (4).

The existence of such a map T_0 optimal between μ'' and ν' satisfying (5) follows from the application of Theorem 3 for $E = E_0 = \mathbb{R}^d = \mathbb{R}^h \times \mathbb{R}^{d-h} = B_0 \times F$ and $\varphi = p$. Indeed, B_0 and F are complete Riemannian manifolds; $\langle \cdot, \cdot \rangle_\sigma$ is twisted on $B_0 \times B_0$; $p_{\#} \mu''$ has a density on B_0 and every $(\mu'')_u$ has a density w.r.t. the Lebesgue measure on F as a conditional probability.

We then make t_B explicit. One has that $c_\Sigma(x, y) = -\langle \tilde{\Sigma} x, y \rangle$, where $\tilde{\Sigma} = \text{diag}(\sigma_i)_{1 \leq i \leq h}$. As $p_{\#} \mu''$ has a density, we can apply Lemma 4 stated above with $(\psi_1, \psi_2) = (\tilde{\Sigma}, \text{id})$ to obtain that there exists

²by Lemma 2 it suffices to check that $(p_{\mathbb{R}^d}, \text{id})_{\#}(\text{id}, T)_{\#} \mu'$ is in $\Pi^\star(p_{\mathbb{R}^d \#} \mu', \nu')$:

$$(p_{\mathbb{R}^d}, \text{id})_{\#}(\text{id}, T)_{\#} \mu' = (p_{\mathbb{R}^d}, T_0 \circ p_{\mathbb{R}^d})_{\#} \mu' = (\text{id}, T_0)_{\#} p_{\mathbb{R}^d \#} \mu'.$$

a unique optimal transport plan π_B^\star between $p_\# \mu''$ and $p_\# \nu'$ for the cost c_Σ and that it is induced by a map $t_B : B \rightarrow B$ of the form $t_B = \nabla f \circ \tilde{\Sigma}$, with f convex. \square

REMARK 7. A special case of our theorem is Theorem 4.2.3 from [Vay20] (Proposition 6 in this work): when $h = d$, the optimal map between $O_X \mu$ and $O_Y \nu$ writes $T_0 \circ p_{\mathbb{R}^d}$ with $T_0 = \nabla f \circ \tilde{\Sigma}$. The induced optimal map between μ and ν is then:

$$\begin{aligned} T &= O_Y^\top \circ (\tilde{T}_0 \circ p_{\mathbb{R}^d}) \circ O_X \\ &= O_Y^\top \circ (\nabla f \circ \tilde{\Sigma} \circ p_{\mathbb{R}^d}) \circ O_X \\ &= O_Y^\top \circ (\nabla f \circ \Sigma) \circ O_X \\ &= \nabla(f \circ O_Y) \circ O_Y^\top \circ \Sigma \circ O_X && \text{since } \nabla(f \circ A) = A^\top \circ \nabla f \circ A \\ &= \nabla \tilde{f} \circ M^\star, \end{aligned}$$

where $\tilde{f} \triangleq f \circ O_Y$ is convex.

3.2 The quadratic cost

We recall the (GW-Q) problem:

$$\min_{\pi \in \Pi(\mu, \nu)} \int_{X \times Y} \int_{X \times Y} \| |x - x'|^2 - |y - y'|^2 \|^2 d\pi(x, y) d\pi(x', y'), \quad (\text{GW-Q})$$

which is invariant by translation of μ and ν . With no loss of generality, we suppose both measures centered. Expanding the integrand provides

$$\| |x - x'|^2 - |y - y'|^2 \|^2 = |x - x'|^4 + |y - y'|^4 - 2|x - x'|^2 |y - y'|^2,$$

and the two first terms only depend on μ and ν , not on π . Expanding the remaining term yields nine terms. Two of them also lead to a constant contribution: $-|x|^2 |y'|^2$ and $-|x'|^2 |y|^2$; four lead to vanishing integrals since μ and ν are centered: $2|x|^2 \langle y, y' \rangle$, $2|x'|^2 \langle y, y' \rangle$, $2|y|^2 \langle x, x' \rangle$ and $2|y'|^2 \langle x, x' \rangle$. The remaining three terms then yield the following equivalent problem:

$$\min_{\pi \in \Pi(\mu, \nu)} \int -|x|^2 |y|^2 d\pi(x, y) + 2 \iint -\langle x, x' \rangle \langle y, y' \rangle d\pi(x, y) d\pi(x', y').$$

Assuming an optimal correspondence plan π^\star , this plan is also an optimal transport plan for the linearized problem (1) with cost

$$C_{\pi^\star}(x, y) = -|x|^2 |y|^2 - 4 \int \langle x, x' \rangle \langle y, y' \rangle d\pi^\star(x', y') = -|x|^2 |y|^2 - 4 \langle M^\star x, y \rangle,$$

where $M^\star \triangleq \int y' \otimes x' d\pi^\star(x', y') \in \mathbb{R}^{d \times n}$. In the cases where the rank of M^\star is d (resp. $d - 1$), this linearized cost satisfies (Subtwist) (resp. (m -twist) with $m = 2$), yielding an optimal map/anti-map (resp. bimap) structure by compactness of the support of μ and ν when μ has a density. In the case where $\text{rk } M^\star \leq d - 2$, nothing can be said and there is *a priori* little hope for the existence of an optimal correspondence map; but not unsurprisingly, we can actually prove it.

Theorem 5 (Existence of an optimal map, bimap or map/anti-map for the quadratic cost). Let $n \geq d$, $\mu, \nu \in \mathcal{P}(\mathbb{R}^n) \times \mathcal{P}(\mathbb{R}^d)$ two measures with compact supports. Suppose that $\mu \ll \mathcal{L}^n$. Let π^\star be a solution of (GW-Q) and $M^\star \triangleq \int y' \otimes x' d\pi^\star(x', y')$. Then:

- if $\text{rk } M^\star = d$, there exists an optimal plan that is induced by a map/anti-map;
- if $\text{rk } M^\star = d - 1$, there exists an optimal plan that is induced by a bimap;

- if $\text{rk } M^\star \leq d - 2$, there exists an optimal plan that is induced by a map that can be written as

$$T = O_{\mathcal{Y}}^\top \circ T_0 \circ O_{\mathcal{X}},$$

where $O_{\mathcal{X}}$ and $O_{\mathcal{Y}}$ are change-of-basis matrices of \mathbb{R}^n and, writing $\Phi(x) \triangleq ((x_u, |x_v|^2), x_v/|x_v|) \triangleq (x_B, x_F)$ for any $x \in \mathbb{R}^n$,

$$\Phi \circ T_0(x) = \left(\tilde{c} - \exp_{x_B}(\nabla f(x_B)), \exp_{x_F}(\nabla g_{x_B}(x_F)) \right)$$

with $h \leq n$, $f : \mathbb{R}^{h+1} \rightarrow \mathbb{R}$ being \tilde{c} -convex and all $g_{x_B} : \mathbb{R}^{n-h} \rightarrow \mathbb{R}$ being $d_{S^{n-h-1}}^2/2$ -convex.

The case $\text{rk } M^\star \leq d - 2$ is a consequence of Theorem 3 and the proof is as follows:

Proof. We consider the measure ν as a measure of \mathbb{R}^n of d -dimensional support. Similarly to the inner product cost, by SVD the cost becomes $c(x, y) = -|x|^2|y|^2 - \langle O_{\mathcal{Y}}^\top \Sigma O_{\mathcal{X}} x, y \rangle = -|O_{\mathcal{X}} x|^2 |O_{\mathcal{Y}} y|^2 - \langle \Sigma(O_{\mathcal{X}} x), O_{\mathcal{Y}} y \rangle$. Using Lemma 3 and similarly to the inner product case, the problem transforms into $\min_{\tilde{\pi}} \langle c_\Sigma, \tilde{\pi} \rangle$ for $\tilde{\pi} \in \Pi(O_{\mathcal{X}} \mu, O_{\mathcal{Y}} \nu)$, where $c_\Sigma(x, y) \triangleq -|x|^2|y|^2 - \langle \Sigma x, y \rangle$. Further assuming $\sigma_1 \geq \dots \geq \sigma_h > 0$ and writing any $z \in \mathbb{R}^n$ as $z = (z_u, z_v) \in \mathbb{R}^h \times \mathbb{R}^{n-h}$,

$$\begin{aligned} c_\Sigma(x, y) &= -|x_u|^2|y_u|^2 - |x_u|^2|y_v|^2 - |x_v|^2|y_u|^2 - |x_v|^2|y_v|^2 - \langle \tilde{\Sigma} x_u, y_u \rangle && \text{with } \tilde{\Sigma} = \text{diag}(\sigma_1, \dots, \sigma_h) \\ &= -|x_u|^2|y_u|^2 - |x_u|^2 n_{y_v} - n_{x_v} |y_u|^2 - n_{x_v} n_{y_v} - \langle \tilde{\Sigma} x_u, y_u \rangle && \text{with } n_{x_v} = |x_v|^2 \text{ and } n_{y_v} = |y_v|^2 \\ &\triangleq \tilde{c}(\varphi(x), \varphi(y)) && \text{with } \varphi : x \mapsto (x_u, |x_v|^2). \end{aligned}$$

The cost $c_\Sigma(x, y)$ therefore only depends of the values of $\varphi(x)$ and $\varphi(y)$. Let us now examine the injectivity of $\nabla_x \tilde{c}(\tilde{x}, \cdot)$ for a fixed $\tilde{x} = \begin{pmatrix} x_u \\ n_{x_v} \end{pmatrix}$. For any $\tilde{y} = \begin{pmatrix} y_u \\ n_{y_v} \end{pmatrix}$:

$$\nabla_x \tilde{c}(\tilde{x}, \tilde{y}) = (w, t) \iff \begin{cases} w = 2(|y_u|^2 + n_{y_v})x_u + \tilde{\Sigma} y_u \\ t = |y_u|^2 + n_{y_v} \end{cases} \iff \begin{cases} y_u = \tilde{\Sigma}^{-1}(w - 2tx_u) \\ n_{y_v} = t - |y_u|^2 \end{cases}$$

hence \tilde{c} satisfies the twist condition.

Now, the same as in Example 2 applies, but this time with $E_0 = \mathbb{R}^h \times \mathbb{R}^{n-h}$, $E = E_0 \setminus (\mathbb{R}^h \times \{0\})$, $B_0 = \mathbb{R}^h \times \mathbb{R}$ and $F = S^{n-h-1} = \{x \in E_0 \mid |x_v| = 1\}$. Is then ensured the existence of a structured Monge map between μ and ν for the cost c : it decomposes for almost all $x \in \mathbb{R}^n$ as a Monge map on the basis $B_0 = \mathbb{R}^{h+1}$ obtained as the gradient of a \tilde{c} -convex function $f : \mathbb{R}^{h+1} \rightarrow \mathbb{R}$ (via the \tilde{c} -exponential map on \mathbb{R}^{h+1}) and a Monge map on each fiber $F = S^{n-h-1}$, also built from gradients of convex functions $h_{(x_u, |x_v|^2)} : S^{n-h-1} \rightarrow \mathbb{R}$ (via the exponential map on the sphere); hence the result. Last, note that the case where $M^\star = 0$ has not been explicitly treated. In this case, the cost is simply $c(x, y) = -|x|^2|y|^2 = \tilde{c}(n_{x_v}, n_{y_v})$ and the strategy above directly applies. \square

3.3 Complementary study of the quadratic cost in the one-dimensional case

Recalling the (GW-IP) is invariant by translation, we assume that measures μ and ν below are centered. In the one-dimensional case $\mathcal{X}, \mathcal{Y} \subset \mathbb{R}$, the linearized quadratic GW problem reads, with π^\star an optimal correspondence plan:

$$\min_{\pi \in \Pi(\mu, \nu)} \int_{\mathcal{X} \times \mathcal{Y}} (-x^2 y^2 - 4mxy) d\pi(x, y), \quad \text{where } m = \int_{\mathcal{X} \times \mathcal{Y}} x' y' d\pi^\star(x', y'), \quad (6)$$

and for any plan $\pi \in \Pi(\mu, \nu)$ (not necessarily optimal), we denote by $m(\pi) = \int xy d\pi(x, y)$ what we call the *correlation* of π . The associated OT cost function $c(x, y) = -x^2 y^2 - 4mxy$ only satisfies the subtweak condition when $m \neq 0$, which does not allow to conclude on the deterministic structure of optimal correspondence plans in the general case. However, in the one-dimensional case one has

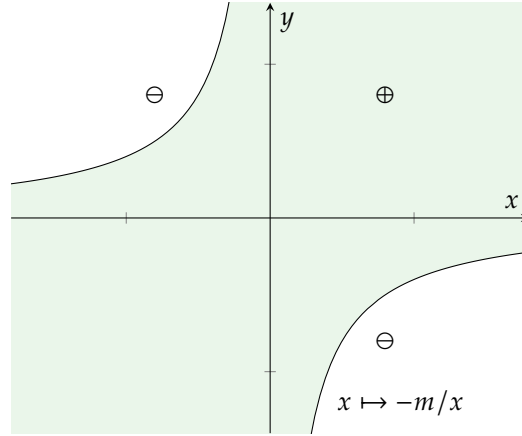


Figure 2: Submodularity (symbol \oplus , in light green) and supermodularity (symbol \ominus) regions for the linearized quadratic GW cost with parameter $m \geq 0$.

at their disposal a useful additional proposition when the cost c is *submodular*, which is sometimes called the Spence–Mirrles condition, that guarantees the optimality of the increasing (resp. decreasing) matching $\pi_{\text{mon}}^{\oplus}$ (resp. $\pi_{\text{mon}}^{\ominus}$) [Car08, San15]:

Proposition 12 (Submodular cost). Let $\mathcal{X}, \mathcal{Y} \subset \mathbb{R}$. We say that a twice-differentiable function $c : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}$ is *submodular* if

$$\text{for all } x, y \in \mathcal{X} \times \mathcal{Y}, \quad \partial_{xy}c(x, y) \leq 0. \quad (\text{Submod})$$

Let $\mu, \nu \in \mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{Y})$ of finite transport cost. If c satisfies (Submod), then $\pi_{\text{mon}}^{\oplus}$ is an optimal plan for (OT), with uniqueness if the inequality is strict.

Similarly, *supermodularity* is defined with the reversed inequality and induces the optimality of $\pi_{\text{mon}}^{\ominus}$.

The linearized quadratic GW cost with parameter $m \geq 0$ is submodular on the region $S = \{(x, y) \mid xy \geq -m\}$ and supermodular elsewhere (see Figure 2 for an illustration); so we cannot directly apply this proposition. Still, it is reasonable to expect that optimal correspondence plans exhibit a monotone increasing structure on S (written \oplus in Figure 2) and a monotone decreasing one elsewhere (written \ominus), and we can actually leverage this type of property to obtain the optimality of the monotone rearrangements in some particular cases (see Sec. 3.3.3).

We also recall the discrete formulation of (OT) in dimension one. Given two sets $\{x_1, \dots, x_N\}$ and $\{y_1, \dots, y_M\}$ of \mathbb{R} and two probability vectors a and b , the (OT) problem between the discrete measures $\mu = \sum_{i=1}^N a_i \delta_{x_i}$ and $\nu = \sum_{j=1}^M b_j \delta_{y_j}$ reads

$$\min_{\pi \in U(a, b)} \langle C, \pi \rangle,$$

where $U(a, b) \triangleq \{\pi \in \mathbb{R}^{N \times M} \mid \pi 1_M = a, \pi^T 1_N = b\}$ is the *transport polytope*, $C = (c(x_i, y_j))_{i, j}$ is the cost matrix and $\langle \cdot, \cdot \rangle$ is the Frobenius inner product. In the case of the linearized problem (6), we denote by $C_{\text{GW}(m)}$ the cost matrix, that has coefficients $(C_{\text{GW}(m)})_{i, j} = -x_i^2 y_j^2 - 4m x_i y_j$ with $m = \langle C_{xy}, \pi^* \rangle$ and $(C_{xy})_{i, j} = x_i y_j$.

In the following sections, we study the optimality of the monotone increasing and decreasing rearrangements $\pi_{\text{mon}}^{\oplus}$ and $\pi_{\text{mon}}^{\ominus}$. It is worth noting that by submodularity of $x, y \mapsto -xy$, these two

correspondence plans have respective correlations m_{\min} and m_{\max} , where

$$\begin{cases} m_{\min} &= \min_{\pi} \langle C_{xy}, \pi \rangle \\ m_{\max} &= \max_{\pi} \langle C_{xy}, \pi \rangle \end{cases}, \quad \text{with } (C_{xy})_{i,j} = x_i y_j, \quad (7)$$

and that for any correspondence plan π , the value of its correlation $m(\pi)$ lies in the interval $[m_{\min}, m_{\max}]$. We provide in the following a complementary study of the quadratic cost in dimension one, namely

- (i) a procedure to find counter-examples to the optimality of the monotone rearrangements;
- (ii) empirical evidence for the tightness of Theorem 5;
- (iii) empirical evidence for the instability of having a monotone rearrangement as optimal correspondence plan;
- (iv) a new result on the optimality of the monotone rearrangements when the measures are composed of two distant parts.

All experiments are reproducible and the code can be found on GitHub³.

3.3.1 Adversarial computation of non-monotone optimal correspondence plans. Theorem 4.1.1 of [Vay20] claims that in the discrete case in dimension 1 with $N = M$ and $a = b = 1_N$, the optimal solution of (QAP) is either the monotone increasing rearrangement $\pi_{\text{mon}}^{\oplus}$ or the monotone decreasing one $\pi_{\text{mon}}^{\ominus}$ (or equivalently the identity $\sigma(i) = i$ or the anti-identity $\sigma(i) = N + 1 - i$); which seems to be the case with a high probability empirically when generating random discrete measures. While this claim is true for $N = 1, 2$ and 3, a counter-example for $N \geq 7$ points has been recently exhibited in [BHS22]. We further propose a procedure to automatically obtain additional counter-examples, demonstrating empirically that such adversarial distributions occupy a non-negligible place in the space of empirical measures. We propose to move away from distributions of optimal plans $\pi_{\text{mon}}^{\oplus}$ and $\pi_{\text{mon}}^{\ominus}$ by performing a gradient descent over the space of empirical distributions with N points using an objective function that favors the strict sub-optimality of the monotone rearrangements; we now detail this procedure.

For $N \geq 1$, we consider the set of empirical distributions over $\mathcal{X} \times \mathcal{Y} = \mathbb{R} \times \mathbb{R}$ with N points and uniform mass, *i.e.* of the form $\pi = \frac{1}{N} \sum_{i=1}^N \delta_{(x_i, y_i)}$. Such plans π can be seen as the identity mapping between vectors $X = (x_1, \dots, x_N)$ and $Y = (y_1, \dots, y_N)$, and we therefore note $\pi = \text{id}(X, Y)$. Denoting by c_{GW} the functional that takes a correspondence plan and returns its cost on the GW problem, we then define \mathcal{F} on $\mathbb{R}^N \times \mathbb{R}^N$ by

$$\mathcal{F}(X, Y) \triangleq c_{\text{GW}}(\pi) - \min \{c_{\text{GW}}(\pi_{\text{mon}}^{\oplus}), c_{\text{GW}}(\pi_{\text{mon}}^{\ominus})\},$$

$$\text{where } \begin{cases} \pi = \text{id}(X, Y) \\ \pi_{\text{mon}}^{\oplus} \text{ and } \pi_{\text{mon}}^{\ominus} \text{ are the monotone rearrangements between } X \text{ and } Y. \end{cases}$$

This quantifies how well the plan π performs when compared to the best of the two monotone rearrangements. We generate N points at random in $[0, 1]^2$ and then perform a simple gradient descent over the positions of the points $(X, Y) = (x_i, y_i)_i$ following the objective

$$\min_{X, Y \in \mathbb{R}^N} \mathcal{F}(X, Y).$$

We include an early-stopping threshold t , since when $\mathcal{F}(\pi)$ becomes negative (*i.e.* we found an slightly adversarial example), the objective function often starts to decrease exponentially fast, exploiting the adversarial behaviour of the plan as much as it can. We found that choosing $t = -2$ gave good results in our experiments. The procedure can be found in Algorithm 1 below. We implemented it using PyTorch's autodiff [PGM⁺19] and used [BTBD20] to implement a differentiable sorting operator to compute the monotone rearrangements. Adversarial plans $\pi_f = \text{id}(X_f, Y_f)$ obtained by Algorithm 1

³link of the code: <https://github.com/theodumont/monge-gromov-wasserstein>.

are not *a priori* optimal for the GW cost between their marginals; but they have at least a better cost than the monotone rearrangements since $\mathcal{F}(X_f, Y_f) < 0$, proving the sub-optimality of the latter.

Algorithm 1 Simple gradient descent over the positions $(x_i)_i$ and $(y_i)_i$.

Parameters:

- N : number of points of the distributions
- N_{iter} : maximum number of iterations
- η : step size
- t : early stopping threshold

Algorithm:

- 1: $X \leftarrow N$ random values in $[0, 1]$, then centered
- 2: $Y \leftarrow N$ random values in $[0, 1]$, then centered
- 3: **for** $i \in \{1, \dots, N_{\text{iter}}\}$ **do**
- 4: $\pi_{\text{mon}}^{\oplus} \leftarrow \text{id}(\text{sort}(X), \text{sort}(Y))$ ▷ id is the identity mapping
- 5: $\pi_{\text{mon}}^{\ominus} \leftarrow \text{id}(\text{sort}(X), \text{sort}(Y)[::-1])$
- 6: $\pi \leftarrow \text{id}(X, Y)$
- 7: $\mathcal{F}(X, Y) \leftarrow \text{GW}(\pi) - \min(\text{GW}(\pi_{\text{mon}}^{\oplus}), \text{GW}(\pi_{\text{mon}}^{\ominus}))$
- 8: **if** $\mathcal{F}(X, Y) < t$ **then stop** ▷ early stopping
- 9: $(X, Y) \leftarrow (X, Y) - \eta \nabla \mathcal{F}(X, Y)$ ▷ step of gradient descent
- 10: **end for**
- 11: return $\pi_f = \text{id}(X, Y)$

Output: a plan π_f with better GW cost than $\pi_{\text{mon}}^{\oplus}$ and $\pi_{\text{mon}}^{\ominus}$

On Figure 3 is displayed an example of adversarial plans obtained following this procedure. It can be observed that during the descent, the plan π has difficulties getting out of what seems to be a saddle point consisting in being the monotone rearrangements between its marginals. Moreover, it is worth noting that the marginals of our typical adversarial plans, such as the one of Figure 3, are often similar to the counter-example proposed in [BHS22], where both measures have their mass concentrated near zero, except for one outlier for ν and two for μ , one on each tail.

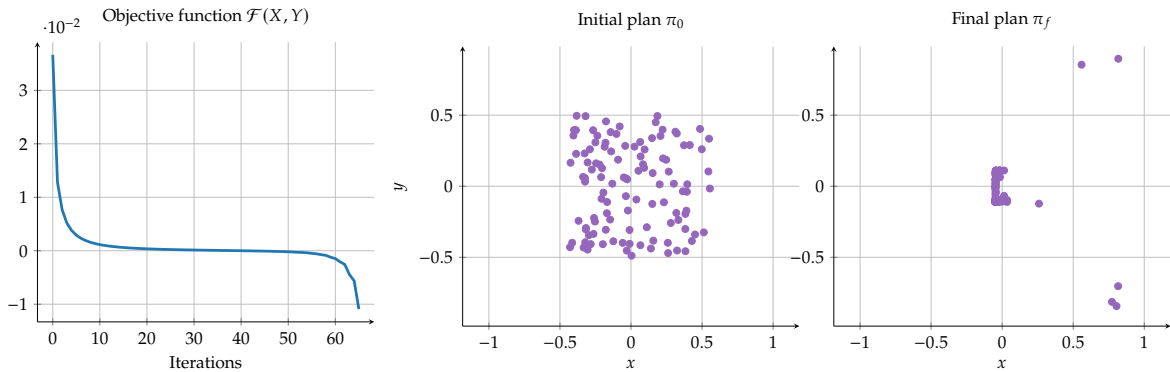


Figure 3: Gradient descent results with parameters $N = 122$, $\eta = 26$, $t = -2$. **(Left)** Evolution of the objective function \mathcal{F} . **(Center)** Initial plan π_0 , generated at random. **(Right)** Final plan π_f (iter. 66).

Furthermore, examining the optimal correspondence plan for these adversarial examples allows to exhibit cases where it is not a map, providing empirical evidence for the following conjecture:

Conjecture 1. Theorem 5 is tight, *i.e.* there exists μ and ν for which optimal correspondence plans for (GW-Q) are not maps but rather a union of two graphs (either that of two maps or that of a map and an anti-map); and this even if μ has a density, classical OT assumption for the existence of an optimal transport map.

In order to approximate numerically the case of a measure which has density w.r.t. the Lebesgue measure, we convolve our distributions $\mu = (X_f, 1_n)$ and $\nu = (Y_f, 1_n)$ with a Gaussian of standard deviation σ and represent it in eulerian coordinates; that is we evaluate the closed form density on a fine enough grid. When σ is large, the optimal correspondence plan for GW is probably induced by a monotone map, as it is the case very frequently empirically; on the contrary, if σ is sufficiently small, *i.e.* when the distributions are very close to their discrete analogous, the optimal correspondence plan should not be a monotone map, by construction of μ and ν .

REMARK 8. Because of the adversarial nature of π_f for the sub-optimality of $\pi_{\text{mon}}^{\oplus}$ and $\pi_{\text{mon}}^{\ominus}$, we know that when σ is sufficiently small, the optimal correspondence plan is not a monotone rearrangement. Still, it could be the case that this optimal plan is a map, but not a monotone one, and there is *a priori* no reason to believe that π_f will agree with Conjecture 1. Surprisingly, it does, as numerical experiments below suggest.

In order to find the optimal correspondence plan π^* between μ and ν , we leverage the fact that π^* is a solution of its associated linearized problem. Therefore, a minimizer of the GW functional is given by

$$\arg \min \left\{ \text{GW}(\pi_m^*) \mid \pi_m^* \in \arg \min_{\pi \in U(a,b)} \langle C_{\text{GW}(m)}, \pi \rangle, m \in [m_{\min}, m_{\max}] \right\}, \quad (8)$$

where $(C_{\text{GW}(m)})_{i,j} = -x_i^2 y_j^2 - 4m x_i y_j$. We therefore compute both m_{\min} and m_{\max} by solving the linear programs in (7), discretize the interval $[m_{\min}, m_{\max}]$ with $N_{\Delta m}$ points, and solve the corresponding linear optimization problem for every value of the parameter m and evaluate the GW cost on each optimal plan for the given parameter m . We then check if the optimal plan exhibits a bimap or a map/anti-map structure. The procedure is described in Algorithm 2.

We display the results on Figure 4, where we plot the optimal correspondence plan π^* in two cases:

- (a) starting from an adversarial plan with both marginals convolved as to simulate densities;
- (b) starting from an adversarial plan with only the first marginal convolved and the second marginal being a sum of Dirac measures.

To facilitate the reading, we draw a blue pixel at a location x on the discretized x -axis (resp. y on the y -axis) each time x (resp. y) has two (disjoint) images (resp. antecedents), making π^* a bimap (resp. a bi-anti-map), or the union of a graph and an anti-graph. In both cases, we observe that π^* is not a map but a bimap instead, similarly to [CMN10, Sec. 4.5]. Note that in case (b), ν being atomic, there cannot be a map from ν to μ , so in both (a) and (b) we numerically exhibit an instance where there is *a priori* no map from either μ to ν nor ν to μ . We also plot the submodularity regions of the linearized GW cost function with parameter $m(\pi^*)$ as an overlay and we observe that when the plan gives mass to a region where the cost is submodular (resp. supermodular), it has a monotone increasing (resp. decreasing) behaviour in this region.

REMARK 9. Although the region on which the optimal plan π^* is a bimap is of small size on Figure 4 right, we cannot expect better due to the form of the adversarial example π_f . Indeed, the bimap behaviour is governed by the outliers of the distributions (see Figure 3), as points in the right tail of μ are encouraged to split in half between points in the right and left tails of ν . As the bimap region only spans the outlier region, it stays of small size when μ and ν have only few outliers.

Algorithm 2 Generating bimap from adversarial examples.**Input:** an adversarial plan $\pi_f = \text{id}(X_f, Y_f)$ obtained from Algorithm 1**Parameters:**

- σ : standard deviation of convolution
- $N_{\Delta x}$: discretization precision
- $N_{\Delta m}$: discretization precision of the interval $[m_{\min}, m_{\max}]$

Algorithm:

- 1: $a \leftarrow \text{convolution}(X_f, \sigma, N_{\Delta x})$
- 2: $b \leftarrow \text{convolution}(Y_f, \sigma, N_{\Delta x})$ ▷ optional (see below)
- 3: $m_{\min} \leftarrow \min_{\pi \in U(a,b)} \langle C_{xy}, \pi \rangle$ ▷ solve linear programs
- 4: $m_{\max} \leftarrow \max_{\pi \in U(a,b)} \langle C_{xy}, \pi \rangle$
- 5: $\text{scores} \leftarrow []$
- 6: **for** $c \in \{m_{\min}, \dots, m_{\max}\}$ **do** ▷ with $N_{\Delta m}$ points
- 7: $\pi_m \leftarrow \arg \min_{\pi \in U(a,b)} \langle C_{\text{GW}(m)}, \pi \rangle$ ▷ solve linear program
- 8: append $\text{GW}(\pi_m)$ to scores
- 9: **end for**
- 10: $\pi^* \leftarrow \arg \max_{\pi} \text{scores}$ ▷ take best plan for GW
- 11: $b \leftarrow \text{"}\pi^* \text{ is a bimap"}$
- 12: return π^*, b

Outputs:

- π^* : optimal plan for GW
- b : boolean asserting if π^* is a bimap

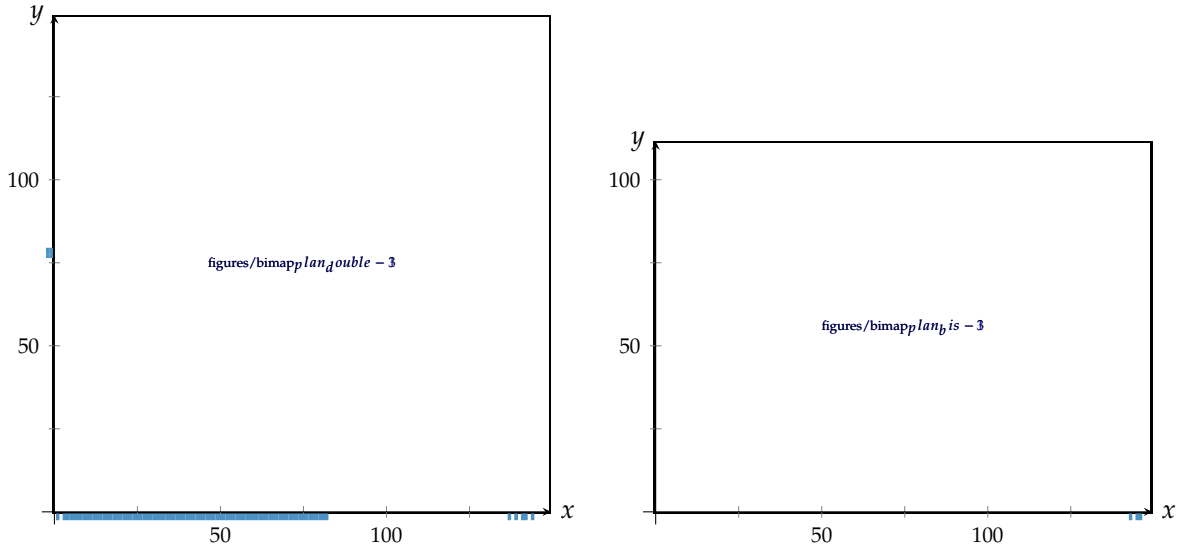


Figure 4: Optimal correspondence plan (in log scale) obtained with our procedure, starting either from a plan with both marginals convolved (**Left**) or with only the first marginal convolved (**Right**); bimap and anti-bimap coordinates (**blue**); submodularity regions (**light green**). Parameters: $\sigma = 5 \cdot 10^{-3}$, $N_{\Delta x} = 150$, $N_{\Delta m} = 2000$.

3.3.2 Empirical instability of the optimality of monotone rearrangements. The above study demonstrates that there exist probability measures μ and ν for which property

$$P(\mu, \nu): \quad \pi_{\text{mon}}^{\oplus} \text{ or } \pi_{\text{mon}}^{\ominus} \text{ is an optimal correspondence plan between } \mu \text{ and } \nu$$

does not hold. However, it is very likely in practice when generating empirical distributions at random; one could ask if property $P(\mu, \nu)$ is at least *stable*, i.e. if when we have μ_0 and ν_0 satisfying $P(\mu_0, \nu_0)$ there is a small ball around μ_0 and ν_0 (for a given distance, say Wasserstein p) inside which property P remains valid. We believe that this is not the case. In order to illustrate this, we start from the counter-example given in [BHS22] with $N = 7$ points and $\varepsilon = 10^{-2}$, that we convolve with a Gaussian of standard deviation σ as before. We then plot as a function of $m \in [m_{\min}, m_{\max}]$ the (true) GW cost of a plan π_m^* , optimal for the linearized GW problem $\pi_m^* \in \arg \min_{\pi} \langle C_{\text{GW}(m)}, \pi \rangle$. The minimum values of this graph are attained by the correlations of optimal correspondence plans, as explained in Section 3.3.1. Hence if σ is small, this optimal plan is not a monotone rearrangement by construction and the minimum are not located on the boundary of the domain. On the contrary, when σ is large, the convolved measures stop being adversarial and the monotone rearrangements start being optimal again. In order to study the phase transition, we plot on Figure 5 the landscape of $m \mapsto \text{GW}(\pi_m^*)$ while gradually increasing the value of σ .

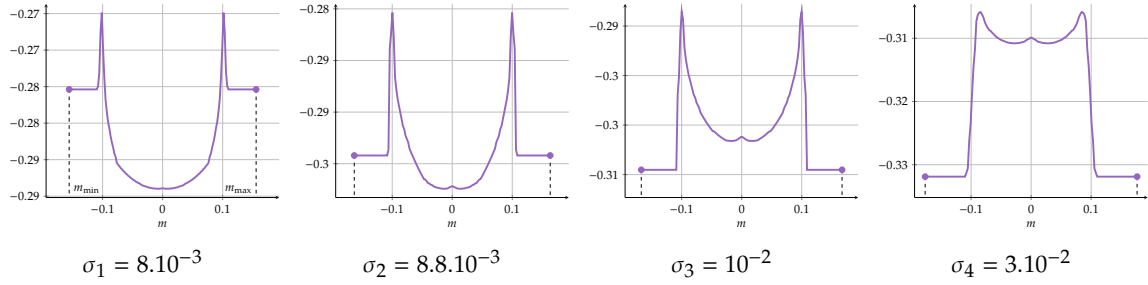


Figure 5: Evolution of the graph of $m \mapsto \text{GW}(\pi_m^*)$ when varying σ on the counter-example of [BHS22] with $N = 7$ points and $\varepsilon = 10^{-2}$. Parameters: $N_{\Delta x} = 100$, $N_{\Delta m} = 150$.

Looking at Figure 5, it is worth noting that there is an incentive for plans of correlation close to m_{\min} or m_{\max} to be the monotone rearrangements, as the horizontal portions of the plot suggest. More importantly, it can be observed that when $\sigma = \sigma_3$ or σ_4 , the monotone rearrangements are optimal, as their correlations realize the minimum of $m \mapsto \text{GW}(\pi_m^*)$; unlike for σ_1 and σ_2 , for which the minimum value of the plot is located near zero. Hence there exists a $\sigma_0 \in (\sigma_2, \sigma_3)$ for which the convolved measures have both $\pi_{\text{mon}}^{\oplus}$, $\pi_{\text{mon}}^{\ominus}$ and another π_0 as optimal correspondence plans. It is direct that property P does not hold in the neighbourhood of these specific measures μ_0 and ν_0 , hence the following result, that we still state as a conjecture since we only provided numerical illustrations of it:

Conjecture 2 (Instability of the optimality of monotone rearrangements). There exists μ, ν two measures on \mathbb{R} such that the optimal plan is supported by the graph of a monotone map, and μ_n, ν_n that weakly converge to μ, ν whose optimal plans are never supported by a monotone map.

3.3.3 A positive result for measures with two components. In the following, μ_1, μ_2, ν_1 and ν_2 are four probability measures supported on a compact interval $A \subset \mathbb{R}$. Denote $\Delta = \text{diam}(A)$, and fix $t \in (0, 1)$ and $K > \Delta$. Let $\tau_K : x \mapsto x + K$ denote the translation by K , and $A + K = \tau_K(A) = \{x + K \mid x \in A\}$. Now, introduce the measures

$$\mu = (1 - t)\mu_1 + t\tau_{K\#}\mu_2 \quad \text{and} \quad \nu = (1 - t)\nu_1 + t\tau_{K\#}\nu_2. \quad (9)$$

Note that μ_1 and $\tau_{K\#}\mu_2$ (resp. ν_1 and $\tau_{K\#}\nu_2$) have disjoint supports. We want to prove the following:

Proposition 13. For K large enough, the unique optimal plan for the quadratic cost between μ and ν is given by one of the two monotone maps (increasing or decreasing).

REMARK 10. The hypothesis of the theorem illustrates that monotone maps are favored when μ and ν both contain a single or more outliers. The proof of the theorem actually shows the importance of long range correspondences or global effect over the local correspondences on the plan. In other words, even though locally, monotone maps may not be optimal, global correspondences favor them. Moreover, these global correspondences have proportionally more weight in the GW functional since the cost is the squared difference of the squared distances. In conclusion, pair of points which are at long distances tend to be put in correspondence. In turn, this correspondence, as shown in the proof, favors monotone matching. Although non-quantitative, this argument gives some insight on the fact that a monotone map is often optimal.

We first prove the following lemma:

Lemma 5. In the setting described above, there exists $K_0 > 0$ such that if $K \geq K_0$, every π optimal plan for $\text{GW}(\mu, \nu)$ can be decomposed as $\pi = \pi_1 + \pi_2$, where either:

1. π_1 is supported on $A \times A$ and π_2 on $(A + K) \times (A + K)$ (that is, we separately transport μ_1 to ν_1 and $\tau_{K\#}\mu_2$ to $\tau_{K\#}\nu_2$), or
2. π_1 is supported on $A \times (A + K)$ and π_2 on $A \times (A + K)$ (that is, we transport μ_1 to $\tau_{K\#}\nu_2$ and μ_2 to $\tau_{K\#}\nu_1$).

Furthermore, whenever $t \neq \frac{1}{2}$, only the first point can occur.

Proof. Consider first the case $t = \frac{1}{2}$. To shorten the notations, we introduce the notations $A_1 = A$ and $A_2 = A + K$. We can now decompose any plan π as $\pi_{11} + \pi_{12} + \pi_{21} + \pi_{22}$ where for instance π_{12} denotes the restriction of the plan π to the product $A_1 \times A_2$. Let us also denote by r the mass of π_{12} , one has $0 \leq r \leq 1/2$ and by symmetry, one can choose that $r \leq 1/4$, otherwise we exchange A_1 and A_2 for the second measure since the cost is invariant to isometries. Remark that, due to marginal constraints, the total mass of π_{11} and π_{22} is $1/2 - r$ and the mass of π_{21} is r . Therefore, it is possible to consider a coupling plan $\tilde{\pi}_{11}$ between the first marginal of π_{12} and the second marginal of π_{21} , and similarly, let $\tilde{\pi}_{22}$ be a coupling plan between the first marginal of π_{21} and the second marginal of π_{12} . We then define a competitor plan $\tilde{\pi} = \pi_{11} + \tilde{\pi}_{11} + \pi_{22} + \tilde{\pi}_{22}$. The first step is to get a lower bound on the term $\text{GW}(\pi, \tilde{\pi})$. Slightly overloading the notations, we introduce

$$\text{GW}(\pi, \gamma) = \int c \, d\pi \otimes \gamma. \quad (10)$$

We expand GW by bilinearity

$$\text{GW}(\pi, \pi) = \sum_{i,j,i',j'} \text{GW}(\pi_{ij}, \pi_{i'j'}) = \sum_{i,j} \text{GW}(\pi_{ii}, \pi_{jj}) + R,$$

where R is the remainder that contains 12 terms from which one can identify two types. 8 terms are of the type $\text{GW}(\pi_{12}, \pi_{11}) \geq r(1/2 - r)(K^2 - \Delta^2)^2$. Indeed, one compares pairs of points (x, x') and (y, y') for $(x, y) \in A_1 \times A_1$ and $(x', y') \in A_1 \times A_2$, therefore $(x - x')^2$ is upper bounded by Δ^2 and $(y - y')^2$ lower bounded by K^2 and the bound above follows after integration against the corresponding measures. The second type is $\text{GW}(\pi_{12}, \pi_{21}) \geq 0$, there are 4 of such terms. We thus have

$$R \geq 8r(1/2 - r)(K^2 - \Delta^2)^2.$$

We now upper-bound the competitor. Similarly, one has

$$\text{GW}(\tilde{\pi}, \tilde{\pi}) = \sum_{i,j} \text{GW}(\pi_{ii}, \pi_{jj}) + \tilde{R}$$

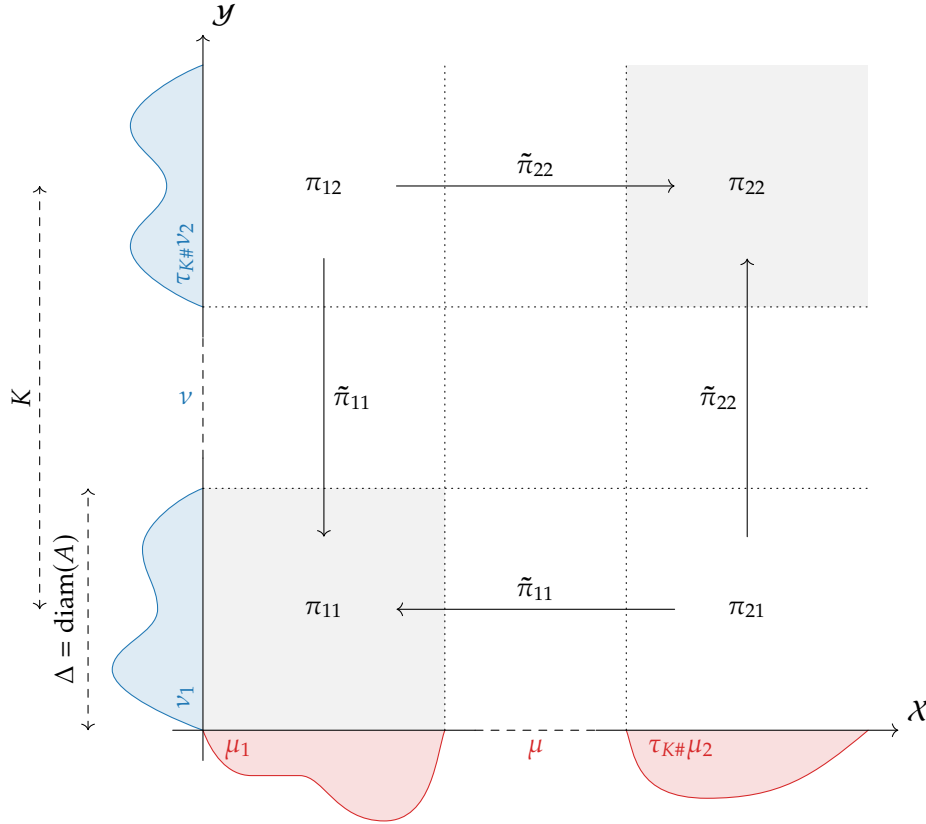


Figure 6: Visual sketch of the proof of Lemma 5.

where $\tilde{R} = 2 \text{GW}(\tilde{\pi}_{11}, \pi_{22} + \tilde{\pi}_{22}) + 2 \text{GW}(\tilde{\pi}_{22}, \pi_{11} + \tilde{\pi}_{11}) + 2 \text{GW}(\pi_{11}, \tilde{\pi}_{11}) + 2 \text{GW}(\pi_{22}, \tilde{\pi}_{22})$. The two last terms can be upper bounded by $2r(1/2 - r)\Delta^2$. Indeed, one compares distance squared of couples of points in A_1 to couple of points in A_1 , so it is upper bounded by Δ^2 . Again by elementary inequalities (see Figure 6), the two first terms can be upper bounded by $r(2K\Delta + \Delta^2)^2$. Note that the total mass of the plan $\pi_{11} + \tilde{\pi}_{11}$ is $1/2$ which explains why $(1/2 - r)$ does not appear. Therefore, the difference between the two values of GW is

$$\text{GW}(\pi, \pi) - \text{GW}(\tilde{\pi}, \tilde{\pi}) \geq r \left(8(1/2 - r)(K^2 - \Delta^2)^2 - 4(1/2 - r)\Delta^2 - 2(2K\Delta + \Delta^2)^2 \right). \quad (11)$$

Then, since $1/2 - r \geq 1/4$ the limit in K of the polynomial function on the r.h.s. of Eq. (11) is $+\infty$ uniformly in $r \in [0, \frac{1}{4}]$, and the result follows; there exists $K > 0$ such that the polynomial function above is nonnegative, for instance $\max(0, K_0)$ where K_0 is the largest root.

The proof in the case $t > 1/2$ (the other is symmetric) is even simpler since $t - r > t - 1/2$ and consequently, there is no choice in the matching of the two measures; it is determined by the corresponding masses. One can directly apply the argument above. \square

We now prove Proposition 13.

Proof of Proposition 13. Thanks to Lemma 5, we know that we can restrict to transportation plans $\pi = \pi_1 + \pi_2$ where, up to flipping v , we can assume that π_1 is supported on $A \times A$ and π_2 on $(A + K) \times (A + K)$.⁴

⁴Note: this is where the choice is made, as in the proof of Lemma 5, between the increasing and the decreasing mappings. Using this convention, the increasing monotone map is shown to be optimal.

Using again the bilinear form $\text{GW}(\pi, \gamma)$ defined in (10), the objective values reached by any transport plan $\pi = \pi_1 + \pi_2$ actually decomposes as

$$\text{GW}(\pi, \pi) = \text{GW}(\pi_1, \pi_1) + 2\text{GW}(\pi_1, \pi_2) + \text{GW}(\pi_2, \pi_2).$$

Now, assume that we have found π_2^* optimal. Let us minimize in π_1 the resulting quadratic problem:

$$\min_{\pi_1} \text{GW}(\pi_1, \pi_1) + 2\text{GW}(\pi_1, \pi_2^*).$$

We know that if π_1^* is a minimizer of this quantity, it must also be a solution of the *linear* problem

$$\min_{\pi_1} \text{GW}(\pi_1, \pi_1^*) + \text{GW}(\pi_1, \pi_2^*).$$

This minimization problem is exactly the optimal transportation problem for the cost

$$\begin{aligned} c(x, y) = \int_{A \times A} ((x - x')^2 - (y - y')^2)^2 d\pi_1^*(x', y') \\ + 2 \int_{(A+K)^2} ((x - x'')^2 - (y - y'')^2)^2 d\pi_2^*(x'', y''). \end{aligned}$$

Now, using the relation $((x - x'')^2 - (y - y'')^2)^2 = ((x - y) - (x'' - y''))^2((x + y) - (x'' + y''))^2$, and that π_2^* is a transportation plan between $\tau_{K\#}\mu_2$ and $\tau_{K\#}\nu_2$ so that we can make a change of variable, observe that

$$\begin{aligned} c(x, y) = \int_{A \times A} ((x - x')^2 - (y - y')^2)^2 d\pi_1^*(x', y') \\ + \int_{A \times A} ((x - y) - (x'' - y''))^2((x + y) - (x'' + y'' + 2K))^2 d(\tau_{-K}, \tau_{-K})\# \pi_2^*(x'', y''). \end{aligned}$$

Now, observe that $\partial_{xy}c(x, y)$ is a polynomial function in K, x, y whose dominant term in K is simply $-2K^2$; recall that A is compact, so that this polynomial function is bounded in x, y . We conclude

$$\partial_{xy}c(x, y) = -2K^2 + O(K) < 0$$

for K large enough, for all $x, y \in A$.

The plan π_1^* is optimal for a submodular cost, and by Proposition 12 must be the increasing matching between μ_1 and ν_1 . By symmetry, so is π_2^* . \square

References

- [AGS05] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient flows: in metric spaces and in the space of probability measures*. Springer Science & Business Media, 2005. 11, 12
- [AKM11] Najma Ahmad, Hwa Kil Kim, and Robert J McCann. Optimal transportation, topology and uniqueness. *Bulletin of Mathematical Sciences*, 1(1):13–32, 2011. 5
- [AMJ18] David Alvarez-Melis and Tommi S Jaakkola. Gromov-wasserstein alignment of word embedding spaces. *arXiv preprint arXiv:1809.00013*, 2018. 2
- [ANN18] Alexandr Andoni, Assaf Naor, and Ofer Neiman. Snowflake universality of wasserstein spaces. In *Annales Scientifiques de l’Ecole Normale Supérieure*, volume 51, pages 657–700. Societe Mathematique de France, 2018. 5
- [BBM05] A.C. Berg, T.L. Berg, and J. Malik. Shape matching and object recognition using low distortion correspondences. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05)*, volume 1, pages 26–33 vol. 1, 2005. 2
- [BHS22] Robert Beinert, Cosmas Heiss, and Gabriele Steidl. On assignment problems related to gromov-wasserstein distances on the real line. *arXiv preprint arXiv:2205.09006*, 2022. 2, 7, 20, 21, 24
- [Bre87] Yann Brenier. Décomposition polaire et réarrangement monotone des champs de vecteurs. *CR Acad. Sci. Paris Sér. I Math.*, 305:805–808, 1987. 2
- [BTBD20] Mathieu Blondel, Olivier Teboul, Quentin Berthet, and Josip Djolonga. Fast differentiable sorting and ranking. In *International Conference on Machine Learning*, pages 950–959. PMLR, 2020. 20
- [Car08] Guillaume Carlier. Remarks on toland’s duality, convexity constraint and optimal transport. *Pacific Journal of Optimization*, 4(3):423–432, 2008. 19
- [Chi07] David Chiron. On the definitions of sobolev and bv spaces into singular spaces and the trace problem. *Communications in Contemporary Mathematics*, 9(04):473–513, 2007. 15
- [CMN10] Pierre-André Chiappori, Robert J McCann, and Lars P Nesheim. Hedonic price equilibria, stable matching, and optimal transport: equivalence, topology, and uniqueness. *Economic Theory*, 42(2):317–354, 2010. 5, 22
- [DSS⁺20] Pinar Demetci, Rebecca Santorella, Björn Sandstede, William Stafford Noble, and Ritambhara Singh. Gromov-wasserstein optimal transport to align single-cell multi-omics data. *BioRxiv*, 2020. 2
- [FGM10] Joaquin Fontbona, Hélène Guérin, and Sylvie Méléard. Measurability of optimal transportation and strong coupling of martingale measures. *Electronic communications in probability*, 15:124–133, 2010. 12, 14, 32, 33, 34
- [FH74] Jacques Faraut and Khélifa Harzallah. Distances hilbertiennes invariantes sur un espace homogène. In *Annales de l’institut Fourier*, volume 24, pages 171–217, 1974. 5
- [GKPS99] Mikhael Gromov, Misha Katz, Pierre Pansu, and Stephen Semmes. *Metric structures for Riemannian and non-Riemannian spaces*, volume 152. Springer, 1999. 2
- [KB57] Tjalling C Koopmans and Martin Beckmann. Assignment problems and the location of economic activities. *Econometrica: journal of the Econometric Society*, pages 53–76, 1957. 2, 3

- [Kon76] Hiroshi Konno. Maximization of a convex quadratic function under linear constraints. *Mathematical programming*, 11(1):117–127, 1976. 4
- [Lyo13] Russell Lyons. Distance covariance in metric spaces. *The Annals of Probability*, 41(5):3284–3305, 2013. 4
- [McC01] Robert J McCann. Polar factorization of maps on riemannian manifolds. *Geometric & Functional Analysis GAFA*, 11(3):589–608, 2001. 2
- [McC12] Robert J McCann. A glimpse into the differential topology and geometry of optimal transport. *arXiv preprint arXiv:1207.1867*, 2012. 5
- [Mém11] Facundo Mémoli. Gromov–wasserstein distances and the metric approach to object matching. *Foundations of computational mathematics*, 11(4):417–487, 2011. 2
- [MG11] Robert J McCann and Nestor Guillen. Five lectures on optimal transportation: geometry, regularity and applications. *Analysis and geometry of metric measure spaces: lecture notes of the séminaire de Mathématiques Supérieure (SMS) Montréal*, pages 145–180, 2011. 5, 6
- [ML18] Haggai Maron and Yaron Lipman. (probably) concave graph matching. *Advances in Neural Information Processing Systems*, 31, 2018. 4
- [Moa16] Abbas Moameni. A characterization for solutions of the monge-kantorovich mass transport problem. *Mathematische Annalen*, 365(3):1279–1304, 2016. 5
- [MTOL20] Ashok Makkuva, Amirhossein Taghvaei, Sewoong Oh, and Jason Lee. Optimal transport mapping via input convex neural networks. In *International Conference on Machine Learning*, pages 6672–6681. PMLR, 2020. 2
- [PC19] Gabriel Peyré and Marco Cuturi. Computational optimal transport: With applications to data science. *Foundations and Trends® in Machine Learning*, 11(5-6):355–607, 2019. 5
- [PGM⁺19] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d’Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., 2019. 20
- [PM22] Nicolò De Ponti and Andrea Mondino. Entropy-transport distances between unbalanced metric measure spaces. *Probability Theory and Related Fields*, 2022. 2
- [RW09] R Tyrrell Rockafellar and Roger J-B Wets. *Variational analysis*, volume 317. Springer Science & Business Media, 2009. 14, 31, 32
- [San15] Filippo Santambrogio. Optimal transport for applied mathematicians. *Birkhäuser, NY*, 55(58-63):94, 2015. 2, 5, 12, 19
- [Sch38] Isaac J Schoenberg. Metric spaces and positive definite functions. *Transactions of the American Mathematical Society*, 44:522–536, 1938. 5
- [Stu06] Karl-Theodor Sturm. On the geometry of metric measure spaces. *Acta mathematica*, 196(1):65–131, 2006. 2
- [Stu12] Karl-Theodor Sturm. The space of spaces: curvature bounds and gradient flows on the space of metric measure spaces. *arXiv preprint arXiv:1208.0434*, 2012. 2, 7

- [SVP21] Thibault Séjourné, François-Xavier Vialard, and Gabriel Peyré. The unbalanced gromov wasserstein distance: Conic formulation and relaxation. *Advances in Neural Information Processing Systems*, 34:8766–8779, 2021. [4](#)
- [TRFC20] Vayer Titouan, Ievgen Redko, Rémi Flamary, and Nicolas Courty. Co-optimal transport. *Advances in Neural Information Processing Systems*, 33:17559–17570, 2020. [2](#)
- [Vay20] Titouan Vayer. A contribution to optimal transport on incomparable spaces. *arXiv preprint arXiv:2011.04447*, 2020. [2](#), [3](#), [7](#), [16](#), [17](#), [20](#)
- [VdV00] Aad W Van der Vaart. *Asymptotic statistics*, volume 3. Cambridge university press, 2000. [34](#)
- [Vil08] Cédric Villani. *Optimal transport: old and new*, volume 338. Springer, 2008. [5](#), [6](#), [7](#), [33](#)
- [ZDIT16] Feng Zhou and Fernando De la Torre. Factorized graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(9):1774–1789, 2016. [2](#)

A Appendix

A.1 Proofs of Lemmas 3 and 4

Proof of Lemma 3. Remark that the continuity of ψ_1 and ψ_2 and their inverse ensures their measurability. We have the following equalities:

$$\begin{aligned} \arg \min_{\pi \in \Pi(\mu, \nu)} \int c(\psi_1(x), \psi_2(y)) \, d\pi(x, y) &= \arg \min_{\pi \in \Pi(\mu, \nu)} \int c(u, v) \, d(\psi_1, \psi_2)_\# \pi(u, v) \\ &= (\psi_1^{-1}, \psi_2^{-1})_\# \arg \min_{\tilde{\pi} \in \Pi(\psi_{1\#}\mu, \psi_{2\#}\nu)} \int c(u, v) \, d\tilde{\pi}(u, v) \end{aligned}$$

since the mapping $(\psi_1^{-1}, \psi_2^{-1})_\#$ is a one-to-one correspondence from $\Pi(\psi_{1\#}\mu, \psi_{2\#}\nu)$ to $\Pi(\mu, \nu)$ by bijectivity of ψ_1 and ψ_2 . This bijectivity ensures that any optimal deterministic transport plan $\tilde{\pi}^*$ between $\psi_{1\#}\mu$ and $\psi_{2\#}\nu$ induces an optimal deterministic transport plan π^* between μ and ν , and *vice versa*. Writing $\tilde{\pi}^* = (\text{id}, T)_\#(\psi_{1\#}\mu)$, this plan π^* is given by

$$\begin{aligned} \pi^* &= (\psi_1^{-1}, \psi_2^{-1})_\# \tilde{\pi}^* \\ &= (\psi_1^{-1}, \psi_2^{-1})_\# (\text{id}, T)_\# \psi_{1\#}\mu \\ &= (\text{id}, \psi_2^{-1} \circ T \circ \psi_1)_\# \mu. \end{aligned} \quad \square$$

Proof of Lemma 4. As $\psi_{1\#}\mu$ has a density w.r.t. the Lebesgue measure since ψ_1 is a diffeomorphism and $\psi_{1\#}\mu$ and $\psi_{2\#}\nu$ have compact support, Brenier’s theorem states that there exists a unique optimal transport plan between $\psi_{1\#}\mu$ and $\psi_{2\#}\nu$ and that it is induced by a map $\nabla\varphi$, where φ is a convex function. Using Lemma 3 then gives the result. \square

REMARK 11 (Discussion on the hypothesis of Lemma 4). In the proof of Lemma 4, we only needed (i) ψ_1, ψ_2 and their inverse to be measurable, (ii) $\psi_{1\#}\mu$ to have a density w.r.t. Lebesgue, and (iii) $\psi_{1\#}\mu$ and $\psi_{2\#}\nu$ to have compact support. Imposing ψ_1 to be a diffeomorphism and ψ_2 to be a homeomorphism ensures both (i) and (ii) and is natural to expect.

A.2 Measurable selection of maps in the manifold setting

A.2.1 Measurability of set-valued maps. Let X, U be two topological spaces, and let \mathcal{A} denote the Borel σ -algebra on X . A set-valued map S is a map from X to $P(U)$ (the set of subsets of U). This will be denoted by $S : X \rightrightarrows U$. The idea is to introduce notations which are consistent with the case where $S(x) = \{u\}$ for all x in X , where we want to retrieve the standard case of maps $X \rightarrow U$. Definitions are taken from [RW09], where measurability is studied when $U = \mathbb{R}^n$. Most results and proofs adapt to a more general setting—in particular when U is a complete Riemannian manifold M . For the sake of completeness, we provide all the proofs, and highlight those that require specific care by replacing \mathbb{R}^n by such a manifold.

Of importance for our proofs, we define:

- The *pre-image* of a set $B \subset U$ is given by

$$S^{-1}(B) = \{x \in X, S(x) \cap B \neq \emptyset\}.$$

- The *domain* of S is $S^{-1}(U)$, that is $\{x \in X, S(x) \neq \emptyset\}$.

We will often use the following relation: if a set A can be written as $A = \bigcup A_k$, then $S^{-1}(A) = \bigcup S^{-1}(A_k)$. Indeed, $x \in S^{-1}(A) \Leftrightarrow S(x) \cap A \neq \emptyset \Leftrightarrow \exists k, S(x) \cap A_k \neq \emptyset \Leftrightarrow \exists k, x \in S^{-1}(A_k) \Leftrightarrow x \in \bigcup_k S^{-1}(A_k)$.

A set-valued map $S : X \rightrightarrows U$ is said to be *measurable* if, for any open set $O \subset U$,

$$S^{-1}(O) \in \mathcal{A}. \quad (12)$$

Note that if S is measurable (as a set-valued map), then its domain must be measurable as well (as an element of \mathcal{A}). We say that $S : X \rightrightarrows U$ is *closed-valued* if $S(x)$ is a closed subset of U for all $x \in X$.

Proposition 14 (Theorem 14.3.c in [RW09]). A closed-valued map S is measurable if and only if $S^{-1}(B) \in \mathcal{A}$ for all $B \subset U$ that are either:

- (a) open (the definition);
- (b) compact;
- (c) closed.

Proof of Proposition 14.

- (a) \Rightarrow (b): For a compact $B \subset U$, let $B_k = \{x \in U, d(x, B) < k^{-1}\}$, $k \geq 0$ (that is open). Note that $x \in S^{-1}(B) \Leftrightarrow S(x) \cap B \neq \emptyset \Leftrightarrow S(x) \cap B_k \neq \emptyset$ for all k because $S(x)$ is a closed set. Hence $S^{-1}(B) = \bigcap_k S^{-1}(B_k)$. All the $S^{-1}(B_k)$ are measurable, so is $S^{-1}(B)$ as a countable intersection of measurable sets.
- (b) \Rightarrow (a): Fix O an open set of U . As we assume U to be a complete separable Riemannian manifold, O can be written as a countable union of compact balls: $O = \bigcup_n \overline{B(x_n, r_n)}$.
- (b) \Rightarrow (c): Immediate.
- (c) \Rightarrow (b): A closed set B can be obtained as a countable union of compact sets by letting $B = \bigcup_n B \cap \overline{B(x_0, n)}$ for some x_0 . Hence $S^{-1}(B) = \bigcup_n S^{-1}(B \cap \overline{B(x_0, n)})$ is in \mathcal{A} . \square

Now, we introduce a proposition on operations that preserve measurability of closed-set valued maps. The proof requires adaptation from the one of [RW09] because the latter uses explicitly the fact that one can compute Minkowski sums of sets (which may not make sense on a manifold).

Proposition 15 (Proposition 14.11 in [RW09], adapted to the manifold case). Let S_1 and $S_2 : X \rightrightarrows U$ be two measurable closed-set valued maps. Then

- $P : x \mapsto S_1(x) \times S_2(x)$ is measurable as a closed-valued map in $U \times U$ (equipped with the product topology).
- $Q : x \mapsto S_1(x) \cap S_2(x)$ is measurable.

Proof. The first point can be proved in the same spirit as the proof proposed by Rockafellar and Wets. Namely, let O' be an open set in $U \times U$. By definition of the product topology, O' can be obtained as $\bigcup_n O_1^{(n)} \times O_2^{(n)}$ where $O_1^{(n)}$ and $O_2^{(n)}$ are open sets in U . Then $P^{-1}(O') = \bigcup_n P^{-1}(O_1^{(n)} \times O_2^{(n)})$. Now, observe that $P^{-1}(A \times B) = \{x, S_1(x) \times S_2(x) \in A \times B\} = \{x, S_1(x) \in A \text{ and } S_2(x) \in B\} = S_1^{-1}(A) \cap S_2^{-1}(B)$, so that finally, $P^{-1}(O') = \bigcup_n S_1^{-1}(O_1^{(n)}) \cap S_2^{-1}(O_2^{(n)})$ that is measurable as a countable union of (finite) intersection of measurable sets (given that S_1, S_2 are measurables). Note that this does not require S_1, S_2 to be closed-valued.

Now, let us focus on the second point, that requires more attention. Thanks to the previous proposition, it is sufficient to show that $Q^{-1}(C) \in \mathcal{A}$ for any compact set $C \subset U$. In [RW09], this is done by writing $Q^{-1}(C) = \{x, S_1(x) \cap S_2(x) \cap C \neq \emptyset\} = \{x, R_1(x) \cap R_2(x) \neq \emptyset\} = \{x, 0 \in (R_1(x) - R_2(x))\} = (R_1 - R_2)^{-1}(\{0\})$, where $R_j(x) = S_j(x) \cap C$ (that is also closed valued), and using the fact that the (Minkowski) difference of measurable closed-valued maps is measurable as well [RW09, Prop. 14.11.c].

To adapt this idea (we cannot consider Minkowski difference in our setting), we introduce the diagonal $\Delta = \{(u, u), u \in U\} \subset U \times U$. Now, observe that $R_1(x) \cap R_2(x) \neq \emptyset \Leftrightarrow (R_1(x) \times R_2(x)) \cap \Delta \neq \emptyset$, that is $x \in R^{-1}(\Delta)$, where $R(x) = R_1(x) \times R_2(x)$. Now, since the maps R_1 and R_2 are measurable closed-valued maps (inherited from S_1, S_2), so is R according to the previous point. And since Δ is closed, $R^{-1}(\Delta) = Q^{-1}(C)$ is measurable. \square

A.2.2 Proof of Proposition 11. The proof is essentially an adaptation of the one of [FGM10], with additional care required due to the fact that we do not have access to a linear structure on M . It relies on measurability of set-valued maps (see [RW09, Ch. 5 and 14] and Appendix A.2.1 for a summary).

The crucial point regarding measurability is the following proposition.

Proposition 16. The set

$$B_{n,k} = \{(u, x), T_u(x) \in A_{n,k}\}. \quad (13)$$

is measurable.

Its proof relies on a core lemma:

Lemma 6. Let $F \subset M$ be a closed set. Then the set

$$B_F = \{(u, x), T_u(x) \in F\}$$

is measurable.

The key will be to identify this set as the domain of a measurable set-valued map, see Appendix A.2.1.

Proof of Lemma 6. Observe that $B_F = \{(u, x), (\{x\} \times F) \cap \text{gph}(T_u) \neq \emptyset\}$, where $\text{gph}(T_u) = \{(x, T_u(x)), x \in M\}$ denotes the topological closure of the graph of the optimal transport map T_u that pushes μ_u onto ν_u . Letting $S_1 : (u, x) \mapsto \{x\} \times F$ and $S_2 : (u, x) \mapsto \text{gph}(T_u)$, so that $B_F = \text{dom}(S)$, where $S(x) = S_1(x) \cap S_2(x)$. According to Proposition 15, given that S_1 and S_2 are closed-valued, if they are measurable, so is S , and so is B_F as the domain of a measurable map. The measurability of these two maps can be easily adapted from the work of [FGM10], we give details for the sake of completeness.

Measurability of S_1 : Let $O \subset M \times M$ be open. The set $S_1^{-1}(O) = \{x, \{x\} \times F \cap O \neq \emptyset\}$ is open (thus measurable): for any $x, z \in S_1^{-1}(O)$, we have $\varepsilon > 0$ such that $B(x, \varepsilon) \times \{z\}$. It proves the measurability of S_1 .

Measurability of S_2 : Given that $u \mapsto (\mu_u, \nu_u)$ is measurable by assumption, and that measurability is preserved by composition, we want to show that (i) the map $S : (\mu, \nu) \mapsto \Pi(\mu, \nu)$ (the set of optimal transport plans between μ and ν) is measurable and (ii) the map $U : \pi \in P(M^2) \mapsto \text{supp } \pi$ satisfies $U^{-1}(O)$ is open for any open set $O \subset P(M^2)$. From these two points, we get that $(U \circ S)^{-1}(O)$ is measurable, thus the measurability of S_2 .

To get (i), observe first that S is closed-valued, so that it is sufficient to prove that $S^{-1}(C)$ is measurable for any closed set $C \subset P(M)^2$ according to Proposition 14. Let $C \subset P(M^2)$ be closed. Then, $S^{-1}(C) = \{(\mu, \nu), \Pi^*(\mu, \nu) \cap C \neq \emptyset\}$, and consider a sequence $(\mu_n, \nu_n)_n$ in $S^{-1}(C)$ that converges to (μ, ν) for the weak topology. Let $\pi_n \in \Pi^*(\mu_n, \nu_n) \cap C$. According to [Vil08, Thm. 5.20], $(\pi_n)_n$ admits a weak limit π in $\Pi^*(\mu, \nu)$, but also since C is closed, $\pi \in C$, so that $(\mu, \nu) \in S^{-1}(C)$ that is closed (hence measurable), proving the measurability of S .

(ii) simply follows from the fact that $U^{-1}(O) = \{\pi, \text{supp } \pi \cap O \neq \emptyset\} = \{\pi, \pi(O) > 0\}$ that is open. Indeed, the Portmanteau theorem gives that if $\pi_n \rightarrow \pi$ (weakly) and $\pi_n(O) = 0$, then $0 = \liminf \pi_n(O) \geq \pi(O) \geq 0$, so $\pi(O) = 0$. The complementary set of $U^{-1}(O)$ is closed, that is $U^{-1}(O)$ is open. \square

Proof of Proposition 16. This follows from the assumption that $A_{n,k}$ can be inner-approximated by a sequence of closed set $F_j \subset A_{n,k}$ and the fact that the B_{F_j} are measurable. \square

We can now prove our main theorem. The proof is clearly inspired from the one of [FGM10], though it requires, in few places, careful adaptation.

Proof of Proposition 11. Recall that we assume that $M = \bigsqcup_n A_{n,k}$. For each n, k , select (in a measurable way) a $a_{n,k}$ in $A_{n,k}$. Then, define the map

$$T^{(k)} : (u, x) \mapsto a_{n,k}, \text{ such that } T_u(x) \in A_{n,k}. \quad (14)$$

This map is measurable. Indeed, the map $\Phi_k : (u, x) \mapsto A_{n,k}$ where $T_u(x) \in A_{n,k}$ is measurable, because $\Phi_k^{-1}(O) = \bigcup_n B_{n,k} \cap O$ that is measurable.

Now, for two maps $f, g : B \times M \rightarrow M$, let D_1 denotes the natural L_1 distance on M , that is

$$D_1(f, g) = \int_B \int_M d(f(u, x), g(u, x)) \, d\mu_u(x) \, dm(u). \quad (15)$$

This yields a complete metric space, and we can observe that $(T^{(k)})_k$ is a Cauchy sequence for this distance. Indeed, for $k \leq j$ two integers, recall that we assume that $(A_{n,j})_n$ is a refinement of $(A_{n,k})_n$, yielding

$$\begin{aligned} D_1(T^{(k)}, T^{(j)}) &= \int_B \int_M d(T^{(k)}(u, x), T^{(j)}(u, x)) \, d\mu_u(x) \, dm(u) \\ &= \int_B \int_M \sum_n \sum_{n': A_{n',j} \subset A_{n,k}} 1_{B_{n',j}}(u, x) \cdot d(a_{n,k}, a_{n',j}) \, d\mu_u(x) \, dm(u) \\ &= \int_B \int_M \sum_n \sum_{n': A_{n',j} \subset A_{n,k}} d(a_{n,k}, a_{n',j}) \, dv_u(A_{n',j}) \, dm(u) \\ &\leq 2^{-k} \end{aligned}$$

where we use that for all u , $\int_{x \in M} 1_{B_{n',j}}(u, x) \, d\mu_u(x) = v_u(A_{n',j})$ by construction (recall that $(u, x) \in B_{n',j} \Leftrightarrow T_u(x) \in A_{n',j} \Leftrightarrow x \in \mu_u(T_u^{-1}(A_{n',j})) = T_u \# \mu_u(A_{n',j})$ and T_u transports μ_u onto v_u), and then that the diameter of the partition $A_{n,k}$ is less than or equal to 2^{-k} and that v_u and m are probability measures.

Now, let T denote the limit of $(T^{(k)})_k$ (that is measurable). It remains to show that $T(u, x) = T_u(x)$, m -a.e. This can be obtained by proving that

$$\int g(x) f(T(x, u)) \, d\mu_u(x) = \int g(x) f(T_u(x)) \, d\mu_u(x), \quad (16)$$

for any pair $f, g : M \rightarrow \mathbb{R}$ of bounded Lipschitz-continuous functions [VdV00, Lemma 2.24].

As in [FGM10], let $\|f\| = \sup_{x \neq y} \frac{|f(x) - f(y)|}{d(x, y)} + \sup_x |f(x)|$. The difference between these two terms can be bounded using the partition $(A_{n,k})_n$. We have for all u m -a.e.:

$$\begin{aligned} &\left| \int g(x) f(T_u(x)) \, d\mu_u(x) - \int g(x) f(T(u, x)) \, d\mu_u(x) \right| \quad (17) \\ &\leq \left| \int g(x) f(T_u(x)) \, d\mu_u(x) - \int g(x) f(T^{(k)}(u, x)) \, d\mu_u(x) \right| + \|g\| \|f\| \int d(T^{(k)}(u, x), T(u, x)) \, d\mu_u(x). \quad (18) \end{aligned}$$

Since $T^{(k)} \rightarrow T$ in D_1 , it implies that up to a subsequence, $\int_x d(T^{(k)}(u, x), T(u, x)) \, d\mu_u(x) \rightarrow 0$ as $k \rightarrow \infty$ for all u m -a.e.

To treat the first term and show that it goes to 0 as $k \rightarrow \infty$ for a subset of B with full m -measure, we write m -a.e. u :

$$\begin{aligned} &\left| \int g(x) f(T_u(x)) \, d\mu_u(x) - \int g(x) f(T^{(k)}(u, x)) \, d\mu_u(x) \right| \\ &\leq \int |g(x)| \left| f(T_u(x)) - f(T^{(k)}(u, x)) \right| \, d\mu_u(x) \\ &\leq \|g\| \|f\| \int d(T_u(x), T^{(k)}(u, x)) \, d\mu_u(x) \\ &\leq \|g\| \|f\| 2^{-k} \sum_n v_u(A_{n,k}). \end{aligned}$$

This concludes the proof. \square

A.3 Measure disintegration

Definition 4 (Measure disintegration). Let \mathcal{X} and \mathcal{Z} be two Radon spaces, $\mu \in P(\mathcal{X})$ and $p : \mathcal{X} \rightarrow \mathcal{Z}$ a Borel-measurable function. A family of probability measures $\{\mu_u\}_{u \in \mathcal{Z}} \subset P(\mathcal{X})$ is a disintegration of μ by p if:

- (i) the function $z \mapsto \mu_z$ is Borel-measurable;
- (ii) μ_u lives on the fiber $p^{-1}(u)$: for $p_*\mu$ -a.e. $u \in \mathcal{Z}$,

$$\mu_u(\mathcal{X} \setminus p^{-1}(u)) = 0,$$

and so $\mu_u(B) = \mu_u(B \cap p^{-1}(u))$ for any Borel $B \subset \mathcal{X}$;

- (iii) for every measurable function $f : \mathcal{X} \rightarrow [0, \infty]$,

$$\int_{\mathcal{X}} f(x) \, d\mu(x) = \int_{\mathcal{Z}} \left(\int_{p^{-1}(u)} f(x) \, d\mu_u(x) \right) d(p_*\mu)(u).$$

In particular, for any Borel $B \subset \mathcal{X}$, taking f to be the indicator function of B ,

$$\mu(B) = \int_{\mathcal{Z}} \mu_u(B) \, d(p_*\mu)(u).$$

Theorem 6 (Disintegration theorem). Let \mathcal{X} and \mathcal{Z} be two Radon spaces, $\mu \in P(\mathcal{X})$ and $p : \mathcal{X} \rightarrow \mathcal{Z}$ a Borel-measurable function. There exists a $p_*\mu$ -a.e. uniquely determined family of probability measures $\{\mu_u\}_{u \in \mathcal{Z}} \subset P(\mathcal{X})$ that provides a disintegration of μ by p .