



**HAL**  
open science

## On the existence of Monge maps for the Gromov-Wasserstein problem

Théo Dumont, Théo Lacombe, François-Xavier Vialard

► **To cite this version:**

Théo Dumont, Théo Lacombe, François-Xavier Vialard. On the existence of Monge maps for the Gromov-Wasserstein problem. Foundations of Computational Mathematics, In press, 10.1007/s10208-024-09643-0 . hal-03818500v3

**HAL Id: hal-03818500**

**<https://hal.science/hal-03818500v3>**

Submitted on 22 May 2024

**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 PROBLEM

THÉO DUMONT<sup>†</sup>, THÉO LACOMBE<sup>†</sup>, AND FRANÇOIS-XAVIER VIALARD<sup>†</sup>

ABSTRACT. The Gromov–Wasserstein problem is a non-convex optimization problem over the polytope of transportation plans between two probability measures supported on two spaces, each equipped with a cost function evaluating similarities between points. Akin to the standard optimal transportation problem, it is natural to ask for conditions guaranteeing some structure on the optimizers, for instance if these are induced by a (Monge) map. We study this question in Euclidean spaces when the cost functions are either given by (i) inner products or (ii) squared distances, two standard choices in the literature. We establish the existence of an optimal map in case (i) and of an optimal 2-map (the union of the graphs of two maps) in case (ii), both under an absolute continuity condition on the source measure. Additionally, in case (ii) and in dimension one, we numerically design situations where optimizers of the Gromov–Wasserstein problem are 2-maps but are not maps. This suggests that our result cannot be improved in general for this cost. Still in dimension one, we additionally establish the optimality of monotone maps under some conditions on the measures, thereby giving insight on why such maps often appear to be optimal in numerical experiments.

KEYWORDS. Non-convex optimization · Optimal transportation · Monge map · Gromov–Wasserstein

MATHEMATICS SUBJECT CLASSIFICATION. 49Q22 · 49Q10 · 90C26

## CONTENTS

<b>1</b>	<b>Introduction</b>	2
1.1	The Gromov–Wasserstein problem	3
1.1.1	Formulation	3
1.1.2	Relation with the optimal transportation problem: a tight bilinear relaxation	4
1.2	Related work	6
1.2.1	Monge maps for the OT problem	6
1.2.2	Gromov–Monge maps for the GW problem	7
1.3	Outline, contributions and perspectives	8
<b>2</b>	<b>Existence of Monge maps for fiber-invariant costs</b>	9
2.1	Statement of the results	9
2.2	Proof of Theorem 2.1	11
2.3	Proof of Theorem 2.4	13
<b>3</b>	<b>Applications to the inner product and squared distance GW problems</b>	14
3.1	The inner product cost	15
3.2	The squared distance cost	16
3.3	Complementary study of GW with squared distance 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	22
3.3.3	A positive result for measures with two components	25
	<b>References</b>	27
	<b>Appendix</b>	31

---

<sup>†</sup>Laboratoire d’Informatique Gaspard Monge, Université Gustave Eiffel, CNRS, F-77454 Marne-la-Vallée, France.  
E-mail addresses: {theo.dumont,theo.lacombe,francois-xavier.vialard}@univ-eiffel.fr.

## 1. INTRODUCTION

Finding correspondences between objects that do not live in the same metric space is a problem of fundamental interest both in applications and in theory appearing in very different fields such as computer vision and shape analysis [Ume88; MS04; MS05; BBM05; Mém07; Mém11], mathematics [Stu12], biology [Dem+20] and machine learning [Red+20; AJ18]. Graph matching [ZD15] is a prominent example of such a problem. A usual situation is when the objects of interest are metric spaces themselves, whose comparison is of practical and theoretical importance [Gro+99]; in this context, the Gromov–Hausdorff distance has been used in different settings [Mém08; MSW21] and its relaxation to  $L^p$  spaces, the Gromov–Wasserstein distance, has been explored in [Stu06; Mém07; Mém08]. Following this line of research, comparing metric measure spaces (i.e. metric spaces endowed with probability measures) using Wasserstein-type distances has attracted a lot of interest [Mém11; Stu12]. Follow-up works over the past decade include for instance [SVP21; DM22], which propose extensions to the case of metric spaces with positive measures (whose total volume is not normalized to 1). The Gromov–Wasserstein distance and its extensions are applied successfully in machine learning [GJB19; Bun+19; Xu+20] and biology [Dem+22].

Following [Mém07], the Gromov–Wasserstein approach consists in seeking for a correspondence between two metric measure spaces, called transport plan, or coupling [San15], that is of lowest distortion. In general, transport plans may not be deterministic: a point can be paired with (or transported to) several points of the other space, and conversely. A natural question is thus to ask for conditions under which *optimal* transport plans for the Gromov–Wasserstein problem can be guaranteed to be deterministic, that is to be supported on the graph of a map from one space to the other. This problem has been put forward by Sturm [Stu12, Challenge 3.6] and by Mémoli and Needham [MN22b, Question 2.14]. However, in contrast to optimal transport, which is a linear programming problem, the formulation of the Gromov–Wasserstein problem falls in the class of non-convex quadratic assignment problems [KB57], whose global optimizers are computationally harder to solve and theoretically harder to characterize. As a consequence, it is not surprising that fewer results are available in the literature, though some particular instances can be computed in polynomial time [Mém+23].

In optimal transport, the fact that the optimization can, under mild conditions, be restricted to the space of maps has been developed a lot since Brenier’s work [Bre87] and further generalized by McCann [McC01]. Brenier’s result essentially states that, in Euclidean spaces, when the cost function is the squared Euclidean distance and the initial measure has a Lebesgue density, the optimal transport plan is a map that is given by the gradient of a convex function. Such results on the structure of optimal plans/maps are of great interest for reducing the optimization set [Mak+20]. They heavily depend on the choice of the cost function, and the same can be said for the Gromov–Wasserstein problem, as detailed in this work.

In this work, we address the question of the existence of optimal maps for the  $L^2$ -Gromov–Wasserstein problem in Euclidean spaces in two particular cases. The first one is when the distortion is measured in terms of Euclidean inner products, in which we establish the existence of optimal maps and detail their structure. The second is when the distortion is measured in terms of squared Euclidean distances, in which we establish the existence of optimal 2-maps (plans supported on the union of the graphs of two maps) in general, and of optimal maps in specific cases. In both cases, we require an absolute continuity condition similar to the one of Brenier’s theorem. We also study the second case (squared distances) in dimension one, which has attracted recent attention [Vay+19; BHS22]. We provide numerical evidence for a counter-example to the existence of optimal maps, suggesting that our result asserting the existence of optimal 2-maps cannot be improved in general. Furthermore, it was numerically observed in [Vay+19] that monotone (non-decreasing or non-increasing) maps between discrete measures in dimension one were often optimal; however, a counter-example to this claim was designed in [BHS22]. Yet, we show that this holds under suitable conditions, shedding light on why the approach developed in [Vay+19] seems to work well in practice. 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](#).

**Notation.** In the following,  $\mathcal{P}(\mathcal{X})$  is the set of probability measures on a Polish space  $\mathcal{X}$ ,  $\text{Leb}_n$  is the  $n$ -dimensional Lebesgue measure on  $\mathbb{R}^n$ ,  $\text{vol}_M$  is the volume measure on a manifold  $M$ ,  $\ll$  denotes absolute continuity,  $T_{\#}\mu$  is the pushforward of a measure  $\mu \in \mathcal{P}(\mathcal{X})$  by a measurable map  $T : \mathcal{X} \rightarrow \mathcal{Y}$ , namely the measure of  $\mathcal{P}(\mathcal{Y})$  defined on Borel sets  $A$  by  $T_{\#}\mu(A) := \mu(T^{-1}(A))$ ,  $\chi_S$  is the indicator function of a set  $S$ ,  $O_n(\mathbb{R})$  is the set of orthogonal matrices of dimension  $n \times n$  with real entries,  $\mathfrak{S}_n$  is the set of permutations,  $\mathbb{1}_n$  is the vector of  $\mathbb{R}^n$  with entries all equal to 1, and “a.e.” means “almost everywhere” or “almost every” depending on the context. Also, to alleviate notation,  $\langle \cdot, \cdot \rangle$  and  $\| \cdot \|$  respectively denote the inner product and the 2-norm  $\| \cdot \|_2$  on  $\mathbb{R}^n$  for any  $n \geq 1$ .

### 1.1. The Gromov–Wasserstein problem.

1.1.1. *Formulation.* The Gromov–Wasserstein (GW) problem, initially introduced by Mémoli [Mém07], can be seen as an extension of the Gromov–Hausdorff distance [Gro+99] to the context of (probability) measure spaces  $(\mathcal{X}, \mu)$  and  $(\mathcal{Y}, \nu)$  equipped with cost functions  $c_{\mathcal{X}} : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$  and  $c_{\mathcal{Y}} : \mathcal{Y} \times \mathcal{Y} \rightarrow \mathbb{R}$  (typically,  $c_{\mathcal{X}}$  and  $c_{\mathcal{Y}}$  can be distance functions on  $\mathcal{X}$  and  $\mathcal{Y}$ , respectively). The GW problem seeks a correspondence  $\pi$  (i.e. a joint law) between  $\mu$  and  $\nu$  that would make the quantity  $\mathbb{E}_{\pi \otimes \pi}[\|c(X, X') - c(Y, Y')\|_{L^p}]$  as small as possible, where  $(X, Y) \sim \pi$  and  $(X', Y') \sim \pi$  are independent coupled random variables. The formal definition is as follows.

**Definition 1.1** (GW problem). Let  $\mathcal{X}$  and  $\mathcal{Y}$  be Polish spaces and  $p \geq 1$ . Given two probability measures  $\mu \in \mathcal{P}(\mathcal{X})$  and  $\nu \in \mathcal{P}(\mathcal{Y})$ , two continuous symmetric cost functions  $c_{\mathcal{X}} : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$  and  $c_{\mathcal{Y}} : \mathcal{Y} \times \mathcal{Y} \rightarrow \mathbb{R}$ , the  $L^p$ -Gromov–Wasserstein problem aims at finding

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

where  $\Pi(\mu, \nu)$  denotes the subset of  $\mathcal{P}(\mathcal{X} \times \mathcal{Y})$  of probability measures that admit  $\mu$  and  $\nu$  as first and second marginals. Such an element  $\pi \in \Pi(\mu, \nu)$  is said to be a *plan* between  $\mu$  and  $\nu$ . If there exists a measurable map  $T : \mathcal{X} \rightarrow \mathcal{Y}$  such that  $\pi = (\text{id}, T)_{\#}\mu$ , then  $T$  is said to be a *map* between  $\mu$  and  $\nu$  (because of the marginal constraints on  $\pi$ , this is equivalent to requiring  $T_{\#}\mu = \nu$ ). Plans  $\pi \in \Pi(\mu, \nu)$  that solve (GW) will be called *optimal correspondence plans* between  $\mu$  and  $\nu$ , and maps such that  $(\text{id}, T)_{\#}\mu$  solves (GW) will be called *optimal correspondence maps*, or *Gromov–Monge maps*, between  $\mu$  and  $\nu$  (see Remark 1.4 for a note on the terminology we adopt).

While the existence of optimal correspondence plans holds under mild assumptions 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 restrict to the case  $\mathcal{X} \subset \mathbb{R}^n$  and  $\mathcal{Y} \subset \mathbb{R}^d$  for two integers  $n \geq d$ , we fix  $p = 2$  and omit the “ $L^2$ ” when referring to the Gromov–Wasserstein problem for the sake of conciseness. We now consider the following two specific instances of (GW):

- (i) the *inner product* case [AJ18; Vay20], where  $c_{\mathcal{X}}$  and  $c_{\mathcal{Y}}$  compare angles in  $\mathbb{R}^n$  and  $\mathbb{R}^d$ , respectively:

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

- (ii) and the *squared distance* case [Stu12; Vay20], where  $c_{\mathcal{X}}$  and  $c_{\mathcal{Y}}$  are the squared Euclidean 2-norms on  $\mathbb{R}^n$  and  $\mathbb{R}^d$ , respectively:

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

The choice (ii) is standard, as it fits the more general setting of  $c_{\mathcal{X}}$  and  $c_{\mathcal{Y}}$  being powers of some distance functions  $d_{\mathcal{X}}$  and  $d_{\mathcal{Y}}$  on  $\mathcal{X}$  and  $\mathcal{Y}$ . In this case, if  $\mu$  and  $\nu$  have full support,  $\text{GW}_p(\mu, \nu) = 0$  is equivalent to the metric measure spaces  $(\mathcal{X}, d_{\mathcal{X}}, \mu)$  and  $(\mathcal{Y}, d_{\mathcal{Y}}, \nu)$  being isomorphic, where by isomorphic we mean that there exists an isometry  $\varphi : (\mathcal{X}, d_{\mathcal{X}}) \rightarrow (\mathcal{Y}, d_{\mathcal{Y}})$  such that  $\varphi_{\#}\mu = \nu$  [Stu12, Lemma 9.2].<sup>1</sup>

<sup>1</sup>Note that in the inner product case,  $\text{GW}(\mu, \nu) = 0$  implies that the metric measure spaces are isomorphic, but the converse is not true—for instance in cases where  $(\mathcal{X}, d_{\mathcal{X}}, \mu)$  and  $(\mathcal{Y}, d_{\mathcal{Y}}, \nu)$  are point clouds in an Euclidean space that simply differ by a translation, meaning  $\mathcal{Y} = T(\mathcal{X})$ ,  $d_{\mathcal{X}} = d_{\mathcal{Y}}$  the Euclidean distance, and  $\nu = T_{\#}\mu$ , where  $T : x \mapsto x + x_0$ .

A subcase of problem (ii) 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 measures supported on  $N$  points each. In this scenario, optimal correspondence plans  $\pi$  can be chosen as permutations  $\sigma$  of  $\{1, \dots, N\}$  (see [ML18, Theorem 1] and [Vay20, Theorem 4.1.2]), and the problem optimizes over the set of such permutations  $\mathfrak{S}_N$ ,

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

which is a particular case of the *Quadratic Assignment Problem*, first introduced in [KB57].

**Remark 1.2** (Assumptions on the cost functions). The Gromov–Wasserstein problem has been carefully studied in [Mém07; Mém11; Stu12] in the context of *metric measure spaces*, where it is therefore required that the cost functions  $c_{\mathcal{X}}$  and  $c_{\mathcal{Y}}$  are distance functions on  $\mathcal{X}$  and  $\mathcal{Y}$ , respectively. It was further observed by Sturm in [Stu12, Section 5] that the definition of GW can be extended not only to powers of distance functions but also to the more general setting of *gauged spaces*, where cost functions are merely required to be symmetric and square-integrable—in particular, they are no longer required to satisfy the triangle inequality. Chowdhury and Mémoli [CM19] then studied *network spaces*, where the symmetry assumption is further removed. Our definition of the GW problem solely requires cost functions to be continuous and symmetric.

1.1.2. *Relation with the optimal transportation problem: a tight bilinear relaxation.* Let us first recall the formulation of the Optimal Transportation (OT) problem, which will play an extensive role in this work.

**Definition 1.3** (OT problem). Let  $\mathcal{X}$  and  $\mathcal{Y}$  be Polish spaces. Given two probability measures  $\mu \in \mathcal{P}(\mathcal{X})$  and  $\nu \in \mathcal{P}(\mathcal{Y})$  and a continuous *cost function*  $c: \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}$ , we consider the problem

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

Plans  $\pi \in \Pi(\mu, \nu)$  that solve (OT) are called *optimal transport plans* between  $\mu$  and  $\nu$ , and maps such that  $(\text{id}, T)_{\#} \mu$  solves (OT) are called *optimal transport maps*, or *Monge maps*, between  $\mu$  and  $\nu$ .

**Remark 1.4** (A note on terminology). Note that while elements of  $\Pi(\mu, \nu)$  will always be called *plans* in this work, optimal plans for (OT) and (GW) will respectively be called optimal *transport* plans and optimal *correspondence* plans. By doing so, we deviate from the terminology of “optimal couplings”, standard in the literature [Stu12; CM19; MN22a], that does not differentiate between optimality for (OT) and optimality for (GW). We make this distinction to avoid confusion in the rest of the paper, as we will be sometimes simultaneously dealing with plans that are optimal for (OT) and others that are optimal for (GW). The same remark holds about *Monge* and *Gromov–Monge* maps.

The minimization problem in (GW) can be interpreted as the minimization of the map  $\pi \mapsto F(\pi, \pi) := \iint k d\pi \otimes \pi$  where  $k((x, y), (x', y')) = |c_{\mathcal{X}}(x, x') - c_{\mathcal{Y}}(y, y')|^2$ , and  $F$  is a symmetric bilinear map. The first-order optimality condition ensures that if  $\pi^*$  minimizes (GW), then it also minimizes  $\pi \mapsto 2F(\pi, \pi^*)$ . Letting  $C_{\pi^*}(x, y) := \int_{\mathcal{X} \times \mathcal{Y}} k((x, y), (x', y')) d\pi^*(x', y')$ , this is up to a factor 2 the linear problem

$$\min_{\pi \in \Pi(\mu, \nu)} \int_{\mathcal{X} \times \mathcal{Y}} C_{\pi^*}(x, y) d\pi(x, y), \quad (1.1)$$

which is nothing but the (OT) problem induced by the cost  $C_{\pi^*}$  on  $\mathcal{X} \times \mathcal{Y}$ . Therefore, any optimal *correspondence* plan for (GW) with costs  $c_{\mathcal{X}}$ ,  $c_{\mathcal{Y}}$  must be an optimal *transportation* plan for (OT) with cost  $C_{\pi^*}$ . A crucial point, proved in [SVP21, Theorem 3] as a generalization of a result by Konno [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  $\int k d\alpha \otimes \alpha \leq 0$  for every such  $\alpha$ , then the converse implication holds: any solution  $\gamma^* \in \Pi(\mu, \nu)$  of the (OT) problem with cost  $C_{\pi^*}$  is also a solution of the (GW) problem, that is

$$F(\pi^*, \pi^*) = F(\gamma^*, \gamma^*) = F(\pi^*, \gamma^*). \quad (1.2)$$

This result is useful in the present article to derive theoretical properties on the minimizers but also for the alternate minimization algorithms that were first proposed in [Mém07] and extended in [SVP21].

The question of the existence and structure of optimal transport maps has been extensively studied in optimal transportation and a review can be found in [Section 1.2.1](#). Since in this case, the solutions of GW are in correspondence with the solutions of an OT problem, the tools and knowledge from optimal transportation theory can be used to derive existence and structure of optimal correspondence maps. In particular, this holds for our two problems of interest ([GW-SD](#)) and ([GW-IP](#)): if  $\alpha$  denotes a finite signed measure on  $\mathcal{X} \times \mathcal{Y} \subset \mathbb{R}^n \times \mathbb{R}^d$  with null marginals, observe that

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

Expanding the remaining integrand involves nine terms. Given that  $\alpha$  has null marginals (in particular, null total mass), we get  $\int \|x\|^2 \|y\|^2 d\alpha \otimes \alpha = 0$  (and similarly for the terms involving  $\|x'\|^2 \|y'\|^2$ ,  $\|x\|^2 \|y'\|^2$  and  $\|x'\|^2 \|y\|^2$ ), as well as  $\int \|x\|^2 \langle y, y' \rangle d\alpha \otimes \alpha = 0$  (and similarly for the other terms). Eventually, the only remaining term is

$$-8 \int \langle x, x' \rangle \langle y, y' \rangle d\alpha \otimes \alpha = -8 \left\| \int xy^\top d\alpha(x, y) \right\|_F^2 \leq 0, \quad (1.3)$$

where  $\|\cdot\|_F$  denotes the Frobenius norm of a matrix. The same holds for non-finite  $\alpha$  by density. The negativity of this term ensures that solutions of ([GW-SD](#)) are exactly the solutions of an OT problem. Computations for ([GW-IP](#)) are similar—actually, they immediately boil down to the same last two equations. More generally, let  $d_{\mathcal{X}}$  and  $d_{\mathcal{Y}}$  be distance functions on  $\mathcal{X}$  and  $\mathcal{Y}$ . When one considers the cost  $(d_{\mathcal{X}}(x, x') - d_{\mathcal{Y}}(y, y'))^2$ , by expanding the square, the only term that matters in the optimization is  $-2d_{\mathcal{X}}(x, x')d_{\mathcal{Y}}(y, y')$ . Whenever it is possible to write  $d_{\mathcal{X}}$  and  $d_{\mathcal{Y}}$  as squared distances in Hilbert spaces, namely  $d_{\mathcal{X}}(x, x') = \|\varphi(x) - \varphi(x')\|_{H_{\mathcal{X}}}^2$  and  $d_{\mathcal{Y}}(y, y') = \|\psi(y) - \psi(y')\|_{H_{\mathcal{Y}}}^2$  for an embedding  $\varphi : \mathcal{X} \rightarrow H_{\mathcal{X}}$  in a Hilbert space  $H_{\mathcal{X}}$  and similarly for  $\mathcal{Y}$ , computation (1.3) holds. Such a property depends on the metric space. When it is satisfied, the metric space is said to be of *negative type*, or the distance to be *Hilbertian*. Several important Riemannian manifolds are of negative type; among them the real hyperbolic space, the sphere, and the Euclidean space. Counter-examples are for instance the hyperbolic space on the quaternions [[FH74](#)] or the 2-Wasserstein space on  $\mathbb{R}^d$  for  $d \geq 3$  [[ANN18](#)]. We refer to [[Lyo13](#)] for a thorough discussion. Another equivalent formulation is to say that  $d_{\mathcal{X}}$  is a *conditionally negative definite kernel* on  $\mathcal{X}$ .

**Definition 1.5** (CND kernel). A function  $k_{\mathcal{X}} : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$  is a *conditionally negative definite (CND) kernel* if it is symmetric and for all  $N \geq 1$ ,  $x_1, \dots, x_N \in \mathcal{X}$  and  $\omega_1, \dots, \omega_N \in \mathbb{R}$  such that  $\sum_{i=1}^N \omega_i = 0$ , one has  $\sum_{i,j=1}^N \omega_i \omega_j k_{\mathcal{X}}(x_i, x_j) \leq 0$ .

Every CND kernel can be written as  $k_{\mathcal{X}}(x, x') = f(x) + f(x') + \frac{1}{2} \|\varphi(x) - \varphi(x')\|_H^2$  for an embedding  $\varphi : \mathcal{X} \rightarrow H$  with  $H$  a Hilbert space and  $f : \mathcal{X} \rightarrow \mathbb{R}$  a function [[Sch38](#)]. As far as the Gromov–Wasserstein functional is concerned, our discussion above shows that  $c_{\mathcal{X}}$  and  $c_{\mathcal{Y}}$  can actually be replaced with CND kernels and that the relaxation still holds. Obviously, due to the square in the  $L^2$ -GW problem, the “sign” of both kernels does not matter. The following proposition sums up the discussion.

**Proposition 1.6** (Tight bilinear relaxation of GW). *Let  $(\mathcal{X}, k_{\mathcal{X}}, \mu)$  and  $(\mathcal{Y}, k_{\mathcal{Y}}, \nu)$  be two Polish spaces endowed each with a conditionally negative definite kernel (or each with a conditionally positive definite one) and a probability measure. Then the bilinear relaxation of GW is tight, in the sense of (1.2). The corresponding kernel  $k((x, y), (x', y'))$  yields a non-positive quadratic form on signed measures with null marginals on  $\mathcal{X} \times \mathcal{Y}$ .*

## 1.2. Related work.

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 the existence of Monge maps, or variations of, can be ensured. Brenier’s theorem [Bre87], stated below, is the best-known of these results.

**Theorem 1.7** (Brenier’s theorem). *Let  $\mathcal{X} = \mathcal{Y} = \mathbb{R}^d$ ,  $c(x, y) = \|x - y\|^2$ , and  $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$  such that the optimal cost between  $\mu$  and  $\nu$  is finite. If  $\mu \ll \text{Leb}_d$ , then there exists a unique (up to a set of  $\mu$ -measure zero) solution  $\pi^*$  of (OT) and it is induced by a map, which is the gradient of a convex function  $f : \mathbb{R}^d \rightarrow \mathbb{R}$ :*

$$\pi^* = (\text{id}, \nabla f)_{\#}\mu.$$

This central result admits a generalization in the manifold setting, initially proposed by McCann [McC01], that we shall use later on:

**Theorem 1.8** ([Vil08, Theorem 10.41]). *Let  $\mathcal{X} = \mathcal{Y} = M$  be a Riemannian manifold with geodesic distance  $d$ , and  $c(x, y) = d(x, y)^2$ . Let  $\mu, \nu \in \mathcal{P}(M)$  with compact support. If  $\mu \ll \text{vol}_M$ , then there exists a unique solution  $\pi^*$  of (OT) and it is induced by a map, which is the gradient of a  $d^2/2$ -convex function  $f : M \rightarrow \mathbb{R}$  (see Remark 1.10):*

$$\pi^* = (\text{id}, \exp_x(\nabla f))_{\#}\mu.$$

This theorem can be extended in a few directions. The condition that  $\mu$  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 be more general than the squared distance function, as long as it satisfies the *twist condition*, defined below. In the following,  $\mathcal{X}$  and  $\mathcal{Y}$  are complete Riemannian manifolds and  $c : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}$  is a continuous cost function, differentiable with respect to  $x$ . We refer to [MG11; Vil08] for more information on the twist condition, to [CMN10; AKM11; McC12] for the introduction of the subtweak condition, and to [Moa16] for that of the  $m$ -twist and generalized twist conditions.

**Proposition 1.9** (Twist condition). *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  $\mu$ -a.e. on its domain. If  $\mu$  and  $\nu$  have finite transport cost, then (OT) admits a unique optimal transport plan  $\pi^*$  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 1.10** ( $c$ -convexity,  $c$ -exponential). The notion of  $c$ -convexity is a generalization of plain convexity, and we refer for instance to Villani for the definition [Vil08, Definition 5.2] as well as some intuition on  $d^2/2$ -convex functions [Vil08, pp. 355–356]. Following Loeper [Loe09, Definition 2.6] and Villani [Vil08, Definition 12.29], we also recall that the  $c$ -exponential map is defined such that  $c\text{-exp}_x(p)$  is a  $y \in \mathcal{Y}$  satisfying  $\nabla_x c(x, y) + p = 0$ , when it exists and is unique. This notion reduces to the usual Riemannian exponential map when  $c = d^2/2$ .

**Remark 1.11** (Local twist condition and uniqueness). Suppose  $\mathcal{X}$  and  $\mathcal{Y}$   $d$ -dimensional and  $\mu$  and  $\nu$  compactly supported. McCann et al. [MPW12] showed that if additionally  $c$  is  $C^2$  and satisfies the weaker condition of *non-degeneracy*  $\det D_{xy}^2 c(x, y) \neq 0$  for all  $(x, y) \in \mathcal{X} \times \mathcal{Y}$ , any solution of (OT) is supported on a  $d$ -dimensional Lipschitz submanifold (instead of  $2d$  in full generality); but with no guarantee of uniqueness. General conditions for the uniqueness of the solutions of the (OT) problem to hold have recently been established in [MR20].

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. Unfortunately, it cannot be satisfied for smooth costs on compact manifolds. Two weaker notions can however be introduced to retain some (but less) structure on optimal plans.

**Proposition 1.12** (Subtwist condition). *We say that  $c$  satisfies the subtwist condition if*

$$\begin{aligned} & \text{for all } y_1 \neq y_2 \in \mathcal{Y}, x \in \mathcal{X} \mapsto c(x, y_1) - c(x, y_2) \text{ has no critical points,} \\ & \text{save at most one global maximum and at most one global minimum.} \end{aligned} \quad (\text{Subtwist})$$

*Suppose that  $c$  satisfies (Subtwist) and is bounded. If  $\mu$  vanishes on each hypersurface, then (OT) admits a unique optimal transport plan  $\pi^*$  and it is supported on the union of a graph (map from  $\mathcal{X}$  to  $\mathcal{Y}$ ) and an anti-graph (map from  $\mathcal{Y}$  to  $\mathcal{X}$ ):*

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

*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  $G$ . We call such a plan a map/anti-map.*

**Proposition 1.13** ( $m$ -twist condition). *We say that  $c$  satisfies a  $m$ -twist condition if*

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

*Suppose that  $c$  is bounded, satisfies ( $m$ -twist), and assume that any  $c$ -concave function is differentiable  $\mu$ -almost surely on its domain. If  $\mu$  has no atom and  $\mu$  and  $\nu$  have finite transport cost, then each optimal plan  $\pi^*$  of (OT) is supported on the graphs of  $k \leq m$  measurable maps, i.e. there exists non-negative functions  $\alpha_i : \mathcal{X} \rightarrow [0, 1]$  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  $\pi^*(S) = \sum_{i=1}^k \int_{\mathcal{X}} \alpha_i(x) \chi_S(x, T_i(x)) d\mu(x)$  for any Borel  $S \subset \mathcal{X} \times \mathcal{Y}$ . We call such a plan a  $m$ -map.*

**Example 1.14** (2-twist for squared distance GW in dimension one). If  $\mathcal{X} = \mathcal{Y} = \mathbb{R}$  and  $c(x, y) = x^2 y^2 + \lambda xy$  for some  $\lambda \neq 0$ , the 2-twist condition holds and there exists an optimal correspondence 2-map, i.e. a plan that is supported on the graph of two maps. As we shall see in Section 3.3, such costs are closely related to the GW problem with squared distance cost (GW-SD) in dimension one.

**Remark 1.15** (Simplifying assumptions). When measures  $\mu$  and  $\nu$  have compact support and  $\mu$  has a density—which are assumptions that we make in the following—the conditions on measures  $\mu$  and  $\nu$  of Propositions 1.9, 1.12 and 1.13 required to apply the twist conditions are satisfied [Vil08, Remark 10.33].

1.2.2. *Gromov–Monge maps for the GW problem.* Besides being an interesting mathematical question in itself and in addition to reducing the computational complexity of the problem, preferring Gromov–Monge maps in the Gromov–Wasserstein setting is interesting for registration purposes via mappings in imaging, shape analysis, machine learning, and simulation-based inference [HGL21; Zha+22]. The interest of restricting the Gromov–Wasserstein problem to the class of mappings is discussed by Mémoli and Needham in [MN22b], where it is shown that this formulation retains some properties of a metric (Theorem 3). Cases where doing so doesn’t change the value of the (GW) optimization problem are discussed in [MN22a].

In sharp contrast with the optimal transportation problem, there are very few results that ensure the existence of a Gromov–Monge map for the GW problem, even in the particular cases considered in this work. The following Propositions 1.16 to 1.18 ensure the existence of a Gromov–Monge map under restrictive conditions, either on the symmetry of the measures  $\mu$  and  $\nu$  or on the rank of a matrix that depends on an optimal correspondence plan, which is nontrivial to check in practice as we usually have no such plan beforehand. For the inner product case, one has the following result by Vayer:

**Proposition 1.16** (Inner product GW, [Vay20, Theorem 4.2.3]). *Let  $n \geq d$  and  $\mu, \nu \in \mathcal{P}(\mathbb{R}^n) \times \mathcal{P}(\mathbb{R}^d)$  two measures of finite second order moment with  $\mu \ll \text{Leb}_n$ . Suppose that there exists a solution  $\pi^*$  of (GW-IP) such that  $M^* := \int yx^\top d\pi^*(x, y)$  is of full rank. Then there exists an optimal correspondence map between  $\mu$  and  $\nu$ , which can be written as  $T = \nabla f \circ M^*$  with  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  convex.*



For the squared distance case, in [Vay+19] is claimed that in the discrete case in dimension one with uniform mass and same number of points  $N$ , the optimal solution of the discrete GW problem (QAP) would either be the identity  $\sigma(i) = i$  or the anti-identity  $\sigma(i) = N + 1 - i$  (Theorem 3.1). However, a counter-example to this assertion has been recently provided by Beinert et al. [BHS22]. To the best of our knowledge, the only positive results on the existence of Gromov–Monge maps for the squared distance cost are the following.

**Proposition 1.17** (Squared distance GW, [Stu12, Theorem 9.21]). *Let  $\mu, \nu \in \mathcal{P}(\mathbb{R}^n)$ . Assume that  $\mu, \nu \ll \text{Leb}_n$  and that both measures are rotationally invariant around their barycenter. Then every solution  $\pi^*$  of (GW-SD) is induced by a map, unique up to composition with rotations, which rearranges the radial distributions of  $\mu$  and  $\nu$  around their barycenter in a monotone non-decreasing way.*

**Proposition 1.18** (Squared distance GW, [Vay20, Proposition 4.2.4]). *Let  $n \geq d$  and  $\mu, \nu \in \mathcal{P}(\mathbb{R}^n) \times \mathcal{P}(\mathbb{R}^d)$  with compact support. Assume that  $\mu \ll \text{Leb}_n$  and that both  $\mu$  and  $\nu$  are centered. Suppose that there exists a solution  $\pi^*$  of (GW-SD) such that  $M^* = \int yx^\top d\pi^*(x, y)$  is of full rank. Then there exists  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  convex such that  $T = \nabla f \circ M^*$  pushes  $\mu$  to  $\nu$ . Moreover, if there exists a differentiable convex function  $F : \mathbb{R} \rightarrow \mathbb{R}$  such that  $\|T(x)\|^2 = F'(\|x\|^2)$   $\mu$ -a.e., then  $T$  is an optimal correspondence map between  $\mu$  and  $\nu$ .*

It is of interest to note that Proposition 1.17 implies that the optimal correspondence map between two Gaussian measures in dimension one is a monotone map (non-decreasing or non-increasing). Yet, in higher dimensions, the optimal correspondence plan in the Gaussian case is not known unless more constraints on the structure of the plan are imposed [SDD21, Theorem 4.1].

**1.3. Outline, contributions and perspectives.** This work is organized in the following way. Section 2 provides a general setting in which the existence of optimal *transport* maps can be shown for optimal transport problems with costs that are invariant on the fibers of some projection. We provide two versions of the result: Theorem 2.1 requires no structure and is fairly general, and Theorem 2.4 imposes a more structured setting and therefore yields more structure for the optimal maps. The latter has the benefit of being more usable in further theoretical or computational analysis. Its proof requires a measurability argument which is addressed in detail in Proposition 2.10. In Section 3, these two results are applied to the GW problem in Euclidean spaces. When the cost functions are inner products (GW-IP), we establish the existence of Gromov–Monge maps (see Theorem 3.2), improving on [Vay20, Theorem 4.2.3] (see Proposition 1.16) by dispensing with the rank assumption, difficult to check in practice. When the cost functions are squared distances (GW-SD), we establish the existence of optimal 2-maps in the general case (see Theorem 3.6). Yet, we show that Gromov–Monge maps can occur as well in some specific settings that depend on the rank of some matrix, akin to Theorem 4.2.3 in [Vay20].

An important question is to know if the result in the squared distance case (GW-SD) is tight, that is if one can prove or disprove the existence of a Gromov–Monge *map* in general, rather than merely a 2-*map*. We address this question in Section 3.3 in the one-dimensional case. We first prove some structural properties of the optimal plan, then conduct a numerical exploration that suggests the tightness of Theorem 3.6 in dimension one. Namely, we numerically exhibit two measures that satisfy the assumptions of our Theorem 3.6 and for which the solution of (GW-SD) is a 2-map but not a map. We also prove a positive result on the optimality of the monotone rearrangements, which partly explains why such plans are often optimal in practice. Notably, this result highlights the importance of the long-range effects of the squared distance cost function. In comparison, Sturm’s result [Stu12] (Proposition 1.17 above) concerns symmetric measures and does not account for the frequent optimality of the monotone rearrangements in dimension one. As a possible extension, it is of interest to know if Sturm’s optimality result is stable with respect to the input measures. We show that this is not the case in dimension one: there exist small perturbations of symmetric measures (namely, the degenerate  $\mu = \nu = \delta_0$ ) for which the monotone rearrangements are not optimal. Note that this result does not prove the instability in non-degenerate symmetric cases. We also exhibit empirically a non-symmetric case with instability.

Several questions and research directions are left open by our work. Among them, we may mention the following ones: (1) Characterize (or at least give examples of) cost functions that yield a proper Gromov–Wasserstein distance (that would notably equal 0 if and only if the metric measure spaces are isomorphic, contrary to the inner product case, see discussion in [Section 1.1.1](#)) and that guarantee the existence of Gromov–Monge maps. Note that such cost functions, although well-behaved in the sense of our study, may not fit every practical application. Indeed, being free to choose the cost regardless of whether it defines a distance, for instance to favor local or long-range effects, may be highly relevant since it directly affects the optimal correspondence plans, as well as the optimization problem. (2) Establish sharper sufficient conditions for the optimality of the monotone rearrangements in dimension one for (GW-SD) than the one we obtain in [Section 3.3.3](#). (3) Assess the stability of optimal correspondence plans or maps with respect to the input measures to understand the impact of discretization, sampling, etc, for computational matters. For instance, when attempting to solve (GW-SD) between two measures with densities on  $\mathbb{R}$  in [Section 3.3](#), we naturally rely on a finite-grid discretization of the real line to run computations and assume that the thus obtained optimal correspondence plan between our discretized measures (which is a 2-map) is a faithful approximation of the optimal correspondence plan between the underlying measures (which is thus expected to be a 2-map, though we have no formal proof of this). (4) Design algorithms that leverage the knowledge our work gives on the structure of the optimizers of (GW-IP) and (GW-SD) (roughly speaking, they can be parameterized by gradients of convex functions), for instance by learning such a map modeled as an Input Convex Neural Network [[AXK17](#)] in the vein of [[Kor+21](#); [KSB22](#)].

## 2. EXISTENCE OF MONGE MAPS FOR FIBER-INVARIANT COSTS

This section provides the main result on the 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  $C_{\pi^*}$  arising from the first-order condition of (GW-SD) and (GW-IP)—see [Section 1.1.2](#).

**2.1. Statement of the results.** We briefly present the main idea, also represented on [Figure 1](#). Let  $\mu, \nu$  be two probability measures supported on a measurable space  $(E, \Sigma_E)$  and consider a measurable map  $\varphi : E \rightarrow B$ , for some measurable space  $(B, \Sigma_B)$ , referred to as the *base space* later on. We shall omit to mention the  $\sigma$ -algebras afterward. Let  $(\mu_u)_{u \in B}$  and  $(\nu_u)_{u \in B}$  denote disintegrations of  $\mu$  and  $\nu$  with respect to  $\varphi$  (see [Appendix C](#) or for instance [[AGS05](#), Theorem 5.3.1] for a definition). Let  $c : E \times E \rightarrow \mathbb{R}$  that is invariant on the fibers of  $\varphi$ , that is  $c(x, y) = \tilde{c}(\varphi(x), \varphi(y))$  for all  $(x, y) \in E \times E$  and some cost function  $\tilde{c} : B \times B \rightarrow \mathbb{R}$ . Solving the OT problem between  $\mu$  and  $\nu$  on  $E \times E$  with cost  $c$  reduces to solving the OT problem between  $\varphi_{\#}\mu$  and  $\varphi_{\#}\nu$  on  $B \times B$  with cost  $\tilde{c}$ . Assuming that there exists a Monge map  $t_B$  between  $\varphi_{\#}\mu$  and  $\varphi_{\#}\nu$ , we build a Monge map  $T$  between  $\mu$  and  $\nu$  by (i) transporting each fiber  $\mu_u$  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  $T_{\#}\mu = \nu$ . This map  $T$  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 emphasize that ensuring the measurability of the map  $T$  is non-trivial and crucial from a theoretical standpoint.

We propose two theorems to formalize this idea: the first one ([Theorem 2.1](#)) guarantees the existence of a Monge map for the (GW) problem in a fairly general setting, 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 ([Theorem 2.4](#)), in particular on the fibers of  $\varphi$ , enables the construction of a Monge map for (GW) with a structure akin to the one of [Theorem 1.8](#). As detailed in [Section 3](#), both (GW-SD) and (GW-IP) fall in the latter setting.

**Theorem 2.1.** *Let  $\mathcal{X}$  and  $\mathcal{Y}$  be two measurable spaces for which there exist 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  $c : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}$  and  $B_{\mathcal{X}}, B_{\mathcal{Y}}$  be two measurable spaces along with measurable maps  $\varphi : \mathcal{X} \rightarrow B_{\mathcal{X}}$  and  $\psi : \mathcal{Y} \rightarrow B_{\mathcal{Y}}$ . Assume that there exists a cost  $\tilde{c} : B_{\mathcal{X}} \times B_{\mathcal{Y}} \rightarrow \mathbb{R}$  such that*

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

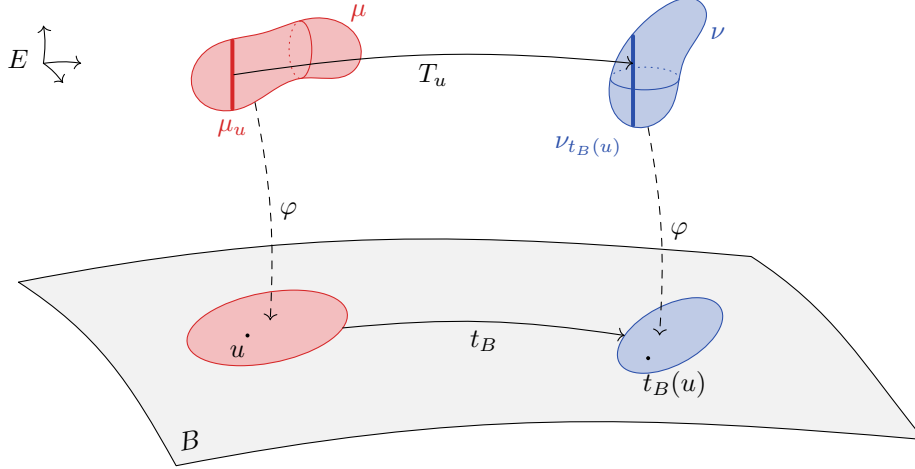


FIGURE 1. Illustration of the construction of a Monge map between  $\mu$  and  $\nu$ : we optimally transport the projections of the measures in  $B$  and then “lift” the resulting map  $t_B$  to  $E$  by sending each fiber  $\mu_u$  onto the fiber  $\nu_{t_B(u)}$ , resulting from the disintegrations of  $\mu$  and  $\nu$  by  $\varphi$ .

and that there exists a Monge map  $t_B : B_{\mathcal{X}} \rightarrow B_{\mathcal{Y}}$  between  $\varphi_{\#}\mu$  and  $\psi_{\#}\nu$  for the cost  $\tilde{c}$ . Assume that there exists a disintegration  $(\mu_u)_{u \in B_{\mathcal{X}}}$  of  $\mu$  with respect to  $\varphi$  such that  $\varphi_{\#}\mu$ -a.e.,  $\mu_u$  is atomless. Then there exists a Monge map  $T$  between  $\mu$  and  $\nu$  for the cost  $c$ . Furthermore, it can be chosen to project onto  $t_B$  through  $(\varphi, \psi)$ , in the 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 2.2** (Atomless assumption). The atomless assumption on the disintegration  $(\mu_u)_u$  is a natural minimal requirement to expect the existence of a map (without specific assumption on the target measure  $\nu$ ) and implies in particular that the fibers  $(\varphi^{-1}(u))_{u \in B_{\mathcal{X}}}$  should not be discrete (at least  $\varphi_{\#}\mu$ -a.e.). Indeed, if for instance  $\mathcal{X} = \mathcal{Y} = B_{\mathcal{X}} = B_{\mathcal{Y}} = \mathbb{R}$  and  $\varphi : x \mapsto \|x\|$ , the fibers of  $\varphi$  are of the form  $\{-u, u\}$ , 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 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 would imply in that case that  $\psi_{\#}\nu$  also has atoms.

**Remark 2.3** (Projection of Monge maps). The map  $T$  projects onto  $t_B$ , in the sense of [Theorem 2.1](#), or equivalently  $\psi \circ T(x) = t_B \circ \varphi(x)$ , for  $\mu$ -a.e.  $x$ . A more general statement, that is “any Monge map  $T$  between  $\mu$  and  $\nu$  projects onto a Monge map between  $\varphi_{\#}\mu$  and  $\psi_{\#}\nu$ ” may not hold in general:  $T$  may send two elements of the same fiber onto two different fibers and therefore not project onto a map downstairs. The statement 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 induced by a map  $t_B$  (e.g. if we can apply Brenier’s [Theorem 1.7](#)). In that case, any  $T$  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 [Theorem 1.8](#), we will assume that the (common) basis  $B = B_{\mathcal{X}} = B_{\mathcal{Y}}$  is a manifold, that *almost all* the fibers of  $\varphi : E \rightarrow B$  are homeomorphic to the *same* manifold  $F$ , and that all source measures of interest  $\mu$ ,  $\mu_u$  and  $\varphi_{\#}\mu$  have a density. 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. Note that if  $\mu(E') = 1$ , then  $\varphi_{\#}\mu$  defines a probability measure on  $B$ . This formalism allows us to state our theorem even when some assumptions only hold  $\mu$ -almost everywhere.

**Theorem 2.4.** *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, when restricted to its image, is measurable as well. Let  $p_B$  and  $p_F$  denote the projections of  $B_0 \times F$  on  $B_0$  and  $F$  respectively, and let  $\varphi := p_B \circ \Phi : E \rightarrow B_0$ . Let  $c : E_0 \times E_0 \rightarrow \mathbb{R}$  and suppose that there exists a cost function  $\tilde{c} : B_0 \times B_0 \rightarrow \mathbb{R}$  satisfying the twist condition such that*

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

*Assume that  $\varphi_{\#}\mu \ll \text{vol}_{B_0}$  and let thus  $t_B$  denote the unique Monge map between  $\varphi_{\#}\mu$  and  $\varphi_{\#}\nu$  for this cost. To alleviate notation, we let  $\mu' := \Phi_{\#}\mu$  and  $\nu' := \Phi_{\#}\nu$  in the following. Suppose that there exists a disintegration  $(\mu'_u)_{u \in B_0}$  of  $\mu'$  by  $p_B$  such that for  $\varphi_{\#}\mu$ -almost every  $u$ ,  $\mu'_u \ll \text{vol}_F$ . Then there exists an optimal map  $T : E \rightarrow E$  between  $\mu$  and  $\nu$  for the cost  $c$ , which can be decomposed as*

$$\Phi \circ T \circ \Phi^{-1}(u, v) = (t_B(u), t_F(u, v)) = (\tilde{c}\text{-exp}_u(\nabla f(u)), \text{exp}_v(\nabla g_u(v))), \quad (2.1)$$

*where  $f : B_0 \rightarrow \mathbb{R}$  is  $\tilde{c}$ -convex, and  $g_u : F \rightarrow \mathbb{R}$  are  $d_F^2/2$ -convex for  $\varphi_{\#}\mu$ -a.e.  $u$ . Note that  $t_F$  could actually be any measurable function that sends each fiber  $\mu'_u$  onto  $\nu'_{t_B(u)}$ .*

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 [\(GW-SD\)](#) as detailed in [Section 3.2](#).

**Example 2.5** (Squared distance GW). Let  $E_0 = \mathbb{R}^d$  and  $E = E_0 \setminus \{0\}$ , let  $B_0 = \mathbb{R}$  and  $F = S^{d-1} = \{x \in E_0 \mid \|x\| = 1\}$  be the  $(d-1)$ -dimensional sphere. For convenience, we also introduce the space  $B = \mathbb{R}_{>0}$ . 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

$$\Phi : E \rightarrow \mathbb{R}_{>0} \times S^{d-1} = B \times F \subset B_0 \times F, \quad x \mapsto (\|x\|, x/\|x\|).$$

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 satisfies the twist condition). Now, if  $\mu$  has a density on  $\mathbb{R}^d$ , so does  $\mu'$  on  $B_0 \times F$  as  $\Phi$  is a diffeomorphism. The coarea formula gives the existence of a disintegration  $(\mu'_u)_{u \in B}$  of  $\mu'$  by  $p_B : (u, v) \mapsto u$  (note that  $p_B\#\mu' = \varphi_{\#}\mu$  also has a density) such that all the  $\mu'_u$  admit a density on  $S^{d-1}$ .

[Theorem 2.4](#) thus applies, ensuring the existence of a structured Monge map between  $\mu$  and (any)  $\nu$  for the cost  $c$ . It decomposes for almost every  $x \in \mathbb{R}^d$  as a Monge map on the basis  $B_0 = \mathbb{R}$  (although it is actually only characterized on the image of  $\varphi$ , that is  $B = \mathbb{R}_{>0}$ ) obtained as the gradient of a convex function  $f$  (there is no need for the exponential map here and  $\nabla f$  is the non-decreasing 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 all the conditions to hold almost everywhere on  $E_0 = \mathbb{R}^d$ , which is important since it allows one to ignore the singularity of  $\varphi$  at  $x = 0$ .

**2.2. Proof of [Theorem 2.1](#).** The proof is composed of 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  $\mu$  and  $\nu$ , 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 that (ii) any such  $\pi$  is an optimal transport plan between  $\mu$  and  $\nu$  for the cost  $c$ . This is formalized by the following lemmas.

**Lemma 2.6** (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  $(\mu_u)_{u \in B_x}$  and  $(\nu_v)_{v \in B_y}$  be disintegrations of  $\mu$  and  $\nu$  by  $\varphi$  and  $\psi$ , respectively. Given  $\tilde{\pi} \in \Pi(\varphi_{\#}\mu, \psi_{\#}\nu)$ , we define

$$\pi := \iint_{B_x \times B_y} (\mu_u \otimes \nu_v) d\tilde{\pi}(u, v),$$

which belongs to  $\Pi(\mu, \nu)$  since  $\tilde{\pi}$  has  $\varphi_{\#}\mu$  and  $\psi_{\#}\nu$  as marginals and, by disintegration,  $\int_{B_{\mathcal{X}}} \mu_u d(\varphi_{\#}\mu) = \mu$  (and same for  $\nu$ ). The relation  $(\varphi, \psi)_{\#}\mu = \tilde{\pi}$  then follows from the fact that by disintegration  $\mu_u$  is exactly supported on  $\varphi^{-1}(u)$  so one has  $\mu_u(\varphi^{-1}(U)) = \delta_U(u)$  for any measurable  $U \subset B_{\mathcal{X}}$  (and same for  $\nu_u$ ).  $\square$

**Lemma 2.7.** *Let  $c : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}$  and  $\tilde{c} : B_{\mathcal{X}} \times B_{\mathcal{Y}} \rightarrow \mathbb{R}$  such that*

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

*Then optimal plans for the base space cost  $\tilde{c}$  are the projections of optimal plans for  $c$ : writing  $\Pi_c^*(\mu, \nu)$  the set of optimal transport plan between  $\mu$  and  $\nu$  for the cost  $c$ , and similarly for  $\Pi_{\tilde{c}}^*(\varphi_{\#}\mu, \psi_{\#}\nu)$ , one has*

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

*Proof.* Let us first remark that the relation between  $c$  and  $\tilde{c}$  implies that for any  $\pi \in \Pi(\mu, \nu)$  and  $\tilde{\pi} = (\varphi, \psi)_{\#}\pi$ , one has  $\langle c, \pi \rangle = \langle \tilde{c}, \tilde{\pi} \rangle$ . Now, let  $\tilde{\pi}^* \in \Pi_{\tilde{c}}^*(\varphi_{\#}\mu, \psi_{\#}\nu)$ ; by [Lemma 2.6](#), there exists a  $\pi \in \Pi(\mu, \nu)$  such that  $(\varphi, \psi)_{\#}\pi = \tilde{\pi}^*$ . Then for any  $\gamma \in \Pi(\mu, \nu)$ , one has  $\langle c, \pi \rangle = \langle \tilde{c}, \tilde{\pi}^* \rangle \leq \langle \tilde{c}, (\varphi, \psi)_{\#}\gamma \rangle = \langle c, \gamma \rangle$ , hence the optimality of  $\pi$ . Conversely, let  $\pi^* \in \Pi_c^*(\mu, \nu)$ ; by [Lemma 2.6](#), 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  $\langle \tilde{c}, (\varphi, \psi)_{\#}\pi^* \rangle = \langle c, \pi^* \rangle \leq \langle c, \gamma \rangle = \langle \tilde{c}, \tilde{\gamma} \rangle$ , hence the optimality of  $(\varphi, \psi)_{\#}\pi^*$ .  $\square$

• *Step 2: Existence of Monge maps between the fibers.* Using [Lemma 2.6](#) 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  $B_{\mathcal{X}} \times B_{\mathcal{Y}}$  and transports each fiber  $\mu_u$  onto  $\nu_{t_B(u)}$  for  $\mu$ -a.e.  $u \in B_{\mathcal{X}}$ . To build a Monge map between  $\mu$  and  $\nu$ , we must show that one can actually transport almost every  $\mu_u$  onto  $\nu_{t_B(u)}$  using a map rather than a plan. For this, we use the following result by Santambrogio [[San15](#), Remark 1.23, Lemma 1.28, Corollary 1.29].

**Proposition 2.8.** *Let  $\alpha, \beta$  be two measures supported on  $\mathbb{R}^d$  and suppose  $\alpha$  atomless. Then:*

- (i) *if  $d = 1$ , there exists a transport map  $\tilde{T}$  that pushes  $\alpha$  onto  $\beta$ . Furthermore, it is the unique optimal map between these measures for the squared distance 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  $\alpha, \beta$ ) 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 2.8](#) 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 squared distance 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  $T_u = \sigma_{\mathcal{Y}}^{-1} \circ \tilde{T}_u \circ \sigma_{\mathcal{X}} : \mathcal{X} \rightarrow \mathcal{Y}$ , which is a (measurable) transport map between  $\mu_u$  and  $\nu_{t_B(u)}$ .

• *Step 3: Building a measurable global map.* Now that we have maps  $(T_u)_u$  between each  $\mu_u$  and  $\nu_{t_B(u)}$ , it may be tempting to simply define a given map  $T : \mathcal{X} \rightarrow \mathcal{Y}$  by  $T(x) = T_{\varphi(x)}(x)$  whenever  $\mu_{\varphi(x)}$  is atomless (which, by assumption, holds  $\mu$ -a.e.). Intuitively, the map  $T$  induces a transport plan  $(\text{id}, T)_{\#}\mu$  satisfying  $(\varphi, \psi)_{\#}(\text{id}, T)_{\#}\mu = (\text{id}, t_B)_{\#}(\varphi_{\#}\mu)$  on  $B_{\mathcal{X}} \times B_{\mathcal{Y}}$  and thus must be optimal according to [Lemma 2.7](#). A crucial point, 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 Fontbona et al. [[FGM10](#)].

**Proposition 2.9** (Measurable selection of maps, [[FGM10](#), Theorem 1.1]). *Let  $(B, \Sigma, m)$  be a  $\sigma$ -finite measure space and consider a measurable function  $B \ni u \mapsto (\mu_u, \nu_u) \in \mathcal{P}(\mathbb{R}^d)^2$ . Let  $c : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$  and assume that for  $m$ -a.e.  $u \in B$ , there is a unique Monge map  $T_u$  between  $\mu_u$  and  $\nu_u$  for the cost  $c$ . Then there exists a measurable function  $(u, x) \mapsto T(u, x)$  such that for  $m$ -almost every  $u$ ,  $T(u, x) = T_u(x)$ ,  $\mu_u$ -a.e.*

We can apply [Proposition 2.9](#) in the case  $d = 1$  to the family of measures  $(\sigma_{\mathcal{X}\#\mu_u}, \sigma_{\mathcal{Y}\#\nu_{t_B(u)}})_{u \in B_{\mathcal{X}}}$ , where the reference measure on  $B_{\mathcal{X}}$  is  $\varphi_{\#\mu}^2$ . This family is indeed measurable: by definition of the measure disintegration, the map  $v \in B_{\mathcal{Y}} \mapsto \nu_v$  is measurable; and as the Monge map  $t_B$  is measurable as well, so is the map  $B_{\mathcal{X}} \ni u \mapsto \sigma_{\mathcal{Y}\#\nu_{t_B(u)}}$  as a composition of measurable maps, and thus so is the map  $u \mapsto (\mu_u, \nu_{t_B(u)})$ . [Proposition 2.9](#) therefore guarantees the existence of a measurable map  $\tilde{T} : B_{\mathcal{X}} \times \mathbb{R} \rightarrow \mathbb{R}$  such that  $\tilde{T}(u, x) = \tilde{T}_u(x)$  for  $\varphi_{\#\mu}$ -almost every  $u$  and  $\sigma_{\mathcal{X}\#\mu}$ -almost every  $x$ . Now, we can define

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

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

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

where we use the disintegration of  $\mu$  with respect to  $\varphi$  and the fact that the  $\mu_u$  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  $T_u = \sigma_{\mathcal{Y}}^{-1} \circ \tilde{T}_u \circ \sigma_{\mathcal{X}}$  defines a transport map between  $\mu_u$  and  $\nu_{t_B(u)}$ . In particular, the image of the fiber  $\varphi^{-1}(u)$  by this map is  $\psi^{-1}(t_B(u)) \subset \mathcal{Y}$ . Since  $\nu_{t_B(u)}$  is supported on  $\psi^{-1}(t_B(u))$ , we get

$$\int_{\mathcal{Y}} \zeta dT_{\#\mu} = \int_{u \in B_{\mathcal{X}}} \int_{y \in \mathcal{Y}} \zeta(y) d\nu_{t_B(u)}(y) d\varphi_{\#\mu}(u) = \int_{v \in B_{\mathcal{Y}}} \int_{y \in \mathcal{Y}} \zeta(y) d\nu_v(y) d\psi_{\#\nu}(v),$$

where we use the fact that  $t_B$  pushes  $\varphi_{\#\mu}$  to  $\psi_{\#\nu}$ ; by disintegration of  $\nu$  with respect to  $\psi$ , we then have  $T_{\#\mu} = \nu$ .

By [Lemma 2.7](#), this map  $T$  is optimal if and only if it satisfies  $(\varphi, \psi)_{\#}(\text{id}, T)_{\#\mu} = (\text{id}, t_B)_{\#}(\varphi_{\#\mu})$ . To show this, let  $\zeta : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}$  be a continuous function with compact support. We have

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

proving the required equality and thus that  $T$  is a Monge map between  $\mu$  and  $\nu$ .

**2.3. Proof of [Theorem 2.4](#).** We recall that we note  $\mu' := \Phi_{\#\mu}$  and  $\nu' := \Phi_{\#\nu}$ . We also denote by  $B$  the image of  $\varphi = p_B \circ \Phi$ , so that  $\mu', \nu'$  are supported on  $B \times F \subset B_0 \times F$ .

• *Step 1: Construction of the structured Monge map.* Given that  $\varphi_{\#\mu} \ll \text{vol}_{B_0}$  and that  $B_0$  is a complete (separable) Riemannian manifold, by [Proposition 1.9](#) and [Remark 1.15](#) there exists a unique optimal transport plan  $\pi_B^*$  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  $t_B = \tilde{c} \text{-exp}_u(\nabla f)$ , with  $f$  being  $\tilde{c}$ -convex on  $B_0$ .

By [Lemma 2.7](#), we know that any transport plan  $\pi \in \Pi(\mu, \nu)$  satisfying  $(\varphi, \varphi)_{\#}\pi = (\text{id}, t_B)_{\#\mu}$  must be optimal. Therefore, if  $\pi$  happens to be induced by a map  $T$ , that is  $\pi = (\text{id}, T)_{\#\mu}$ , we will obtain a Monge map between  $\mu$  and  $\nu$ . To build such a  $T$ , we proceed as in [Section 2.2](#): we define a Monge map  $T_u$  between  $\mu'_u$  and  $\nu'_{t_B(u)}$  for  $\varphi_{\#\mu}$ -a.e.  $u$  and build a global map between  $\mu'$  and  $\nu'$  by (roughly)

<sup>2</sup>Note that we cannot apply [Proposition 2.9](#) 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.

setting  $T(u, x) = T_u(x)$ . As explained in [Section 2.2](#), proving the measurability of such  $T$  requires care.

- *Step 2: Transport between the fibers.* For  $\varphi_{\#}\mu$ -a.e.  $u$ ,  $\mu'_u \ll \text{vol}_F$  and the optimal cost between  $\mu'_u$  and  $\nu'_{t_B(u)}$  is finite by assumption. Whenever  $\mu'_u$  has a density, we can therefore apply [Theorem 1.8](#) between  $\mu'_u$  and  $\nu'_{t_B(u)}$  with the cost  $d_F^2$  to establish the existence of a plan  $\pi_u$  between these fibers that is induced by a map  $T_u : F \rightarrow F$ , which can be expressed as  $T_u(v) = \exp_v(\nabla g_u(v))$  with  $g_u$  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  $\mu'$  onto  $\nu'$  satisfying  $T(u, x) = (t_B(u), T_u(x))$  for  $\varphi_{\#}\mu$ -almost every  $u$  and  $\mu'_u$ -almost every  $x$ . For this, we need an adaptation of [Proposition 2.9](#) to the manifold setting. Namely, we have the following:

**Proposition 2.10** (Measurable selection of maps, manifold case). *Let  $M$  be a complete Riemannian manifold and  $(B, \Sigma, m)$  a measure 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  $\mu_u$  and  $\nu_u$  have a finite transport cost. Let  $T_u$  denote the (unique by [Theorem 1.8](#)) optimal transport map induced by the squared geodesic distance cost  $d_M^2$  on  $M$  between  $\mu_u$  and  $\nu_u$ . Then there exists a measurable function  $(u, x) \mapsto T(u, x)$  such that for  $m$ -almost every  $u$ ,  $T(u, x) = T_u(x)$ ,  $\mu_u$ -a.e.*

This can be shown by adapting the proof of [Theorem 1.1](#) in [\[FGM10\]](#) 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 need specific care in adaptation, is deferred to [Appendix B.2](#).

*Sketch of proof of [Proposition 2.10](#).* The map  $(u, x) \mapsto T(u, x)$  is defined as being the limit of a Cauchy sequence (on a complete metric space) of measurable maps  $(u, x) \mapsto T^{(k)}(u, x)$  which are built as  $L^1$ -approximations of the  $T_u(x)$ , yielding the relation  $T(u, x) = T_u(x)$  for a.e.  $u, x$  when  $k \rightarrow \infty$ . More precisely, given a partition  $(A_{nk})_n$  of  $M$  made of cells of diameters smaller than  $2^{-k}$ , and given  $a_{nk} \in A_{nk}$  any fixed point,  $T^{(k)}$  is precisely defined as  $T^{(k)}(u, x) = a_{nk}$ , where  $a_{nk}$  is the single point such that  $T_u(x) \in A_{nk}$ . Since on each set of the form  $\{(u, x) \mid T_u(x) \in A_{nk}\}$ , one has  $d_M(T^{(k)}(u, x), T_u(x)) \leq 2^{-k}$  (the diameter of the cell), it follows that  $T^{(k)}(u, x) \rightarrow T_u(x)$ , at least for  $m$ -a.e.  $u$  and  $\mu_u$ -a.e.  $x$ .

The key point is to ensure the measurability of the map  $T^{(k)}$ , which boils down to proving that the set  $\{(u, x) \mid T_u(x) \in A_{nk}\}$  is measurable. This is possible by relying on the theory of measurable set-valued maps [\[RW09, Chapters 5 and 14\]](#), that we carefully adapt from the Euclidean to the manifold setting in [Appendix B.1](#).  $\square$

We can apply this proposition with the manifold  $M$  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  $t_F$  satisfying  $t_F(u, v) = T_u(v)$  for  $\varphi_{\#}\mu$ -a.e.  $u$ , and  $\mu'_u$ -a.e.  $v$ , and we can thus define the measurable map  $T(u, x) = (t_B(u), t_F(u, x))$ .

An elementary computation using the pushforward and disintegration definitions as well as the facts that  $T_{u\#}(\mu'_u) = \nu'_{t_B(u)}$  and  $t_{B\#}(\varphi_{\#}\mu) = \varphi_{\#}\nu$  then guarantees that  $T$  sends  $\mu'$  to  $\nu'$ , and  $T_E := \Phi^{-1} \circ T \circ \Phi$  therefore sends  $\mu$  to  $\nu$ . Since  $(\varphi, \varphi)_{\#}(\text{id}, T_E)_{\#}\mu = \pi_B^*$ , by [Lemma 2.7](#),  $T_E$  is an optimal map between  $\mu$  and  $\nu$ .

### 3. APPLICATIONS TO THE INNER PRODUCT AND SQUARED DISTANCE GW PROBLEMS

In this section, we apply the results of [Section 2](#) to our two problems of interest, [\(GW-IP\)](#) and [\(GW-SD\)](#), to obtain the existence of Gromov–Monge maps for these GW problems in suitable settings. The proofs can be sketched in the following way: for both of these problems,

- Let  $\pi^*$  be an optimal plan for the GW problem.
- Consider the associated linearized problem [\(1.1\)](#) at  $\pi^*$ , which is a standard optimal transportation problem. From [Section 1.1.2](#), we know that any solution of this OT problem ( $\pi^*$  being one of them) must be a solution of the initial GW problem.

- Observe that the linearized versions of (GW-IP) and (GW-SD) reduce to optimal transportation problems with costs satisfying the assumptions of [Theorem 2.4](#), thus guaranteeing the existence of structured Gromov–Monge maps.

As such, we only focus on detailing this last point in the following subsections.

**Remark 3.1** (Non-uniqueness for GW). Our results provide the existence of Gromov–Monge maps for the GW problems, but their construction depends on the linearization (that is, of the initial  $\pi^*$  selected). Thus, even though we would have uniqueness of solutions for the linearized problem as given e.g. by the subtwist condition, this does not give uniqueness of the solution to the corresponding GW problem: the GW problem may have several optimizers, each leading to a different linearized problem for which solutions, unique or not, may differ.

**3.1. The inner product cost.** Expanding the integrand in (GW-IP) 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  $\nu$ ), one gets the equivalent problem

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

This problem is not invariant to translations but it is to the action of  $O_n(\mathbb{R}) \times O_d(\mathbb{R})$ . Any optimal correspondence plan  $\pi^*$  is also an optimal transport plan for the linearized problem (1.1) with cost

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

where  $M^* := \int y' x'^{\top} d\pi^*(x', y') \in \mathbb{R}^{d \times n}$ . This linearized cost satisfies the twist condition if and only if  $M^*$  is of full rank, hence in this case  $\pi^*$  is the only solution of the linearized problem and it is induced by a map, and [Theorem 4.2.3](#) from [\[Vay20\]](#) gives a result on the structure of this map. We can actually generalize this to the case where  $M^*$  is arbitrary, as follows.

**Theorem 3.2** (Existence of a Gromov–Monge map for GW with inner product cost). *Let  $n \geq d$  and  $\mu, \nu \in \mathcal{P}(\mathbb{R}^n) \times \mathcal{P}(\mathbb{R}^d)$  two measures with compact supports. Suppose that  $\mu \ll \text{Leb}_n$ . Then there exists an optimal correspondence map  $T : \mathbb{R}^n \rightarrow \mathbb{R}^d$  for (GW-IP), which can be written as*

$$T = O_{\mathcal{Y}}^{\top} \circ T_0 \circ p_{\mathbb{R}^d} \circ O_{\mathcal{X}}, \quad (3.1)$$

where  $(O_{\mathcal{X}}, O_{\mathcal{Y}}) \in O_n(\mathbb{R}) \times O_d(\mathbb{R})$ ,  $p_{\mathbb{R}^d} : \mathbb{R}^n \rightarrow \mathbb{R}^d$  projects on the  $d$  first coordinates, and

$$T_0 : \mathbb{R}^d \rightarrow \mathbb{R}^d, \quad (x_1, \dots, x_d) \mapsto (\nabla f \circ \Sigma(x_1, \dots, x_h), \nabla g_{x_1, \dots, x_h}(x_{h+1}, \dots, x_d)), \quad (3.2)$$

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

To show this, we will need two simple lemmas that essentially state that Brenier’s theorem on optimal transport maps holds when the inner product is perturbed by homeomorphisms. We state them now and prove them in [Appendix A](#).

**Lemma 3.3.** *Let  $\mu, \nu \in \mathcal{P}(E)$  and let  $\psi_1, \psi_2 : E \rightarrow F$  be homeomorphisms. Let  $\tilde{c} : F \times F \rightarrow \mathbb{R}$  and consider the cost  $c(x, y) = c(\psi_1(x), \psi_2(y))$ . Then a map is optimal for the cost  $c$  between  $\mu$  and  $\nu$  if and only if it is of the form  $\psi_2^{-1} \circ T \circ \psi_1$  with  $T : F \times F$  optimal for the cost  $\tilde{c}$  between  $\psi_{1\#}\mu$  and  $\psi_{2\#}\nu$ .*

The second lemma is a simple corollary of the first:

**Lemma 3.4.** *Let  $h \geq 1$  and  $\mu, \nu \in \mathcal{P}(\mathbb{R}^h)$  with compact supports and such that  $\mu \ll \text{Leb}_h$ . 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  $\mu$  and  $\nu$  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  $f : \mathbb{R}^h \rightarrow \mathbb{R}$  convex.*

We now prove [Theorem 3.2](#).



*Proof of Theorem 3.2.* Using a singular value decomposition, we write  $M^* = O_Y^\top \Sigma O_X \in \mathbb{R}^{d \times n}$  with  $(O_X, O_Y) \in O_n(\mathbb{R}) \times O_d(\mathbb{R})$  orthogonal matrices and  $\Sigma \in \mathbb{R}^{d \times n}$  diagonal with non-negative entries. The cost  $C_{\pi^*}$  then becomes

$$C_{\pi^*}(x, y) = -\langle O_Y^\top \Sigma O_X x, y \rangle = -\langle \Sigma O_X x, O_Y y \rangle.$$

Using Lemma 3.3, the problem becomes an optimal transportation problem between  $\mu' := O_X \# \mu$  and  $\nu' := O_Y \# \nu$ . Up to permutation of the rows and columns of  $O_X$  and  $O_Y$ , we can assume  $\sigma_1 \geq \dots \geq \sigma_h > 0$  with  $h := \text{rk}(M^*) \leq d$ , and the problem therefore becomes  $\min_{\tilde{\pi}} \langle c_\Sigma, \tilde{\pi} \rangle$  for  $\tilde{\pi} \in \Pi(\mu', \nu')$ , where  $c_\Sigma(\tilde{x}, \tilde{y}) = -\sum_{i=1}^h \sigma_i \tilde{x}_i \tilde{y}_i =: \tilde{c}(p(\tilde{x}), p(\tilde{y}))$ ,  $p$  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  $\mu'' := p_{\mathbb{R}^d} \# \mu'$  and  $\nu'$  will induce a map  $T = T_0 \circ p_{\mathbb{R}^d}$  optimal between  $\mu'$  and  $\nu'$  (by Lemma 2.7). One can then induce an optimal map between  $\mu$  and  $\nu$  by composing with  $O_X$  and  $O_Y^\top$  (Lemma 3.3), hence Equation (3.1). The existence of such a map  $T_0$ , optimal between  $\mu''$  and  $\nu'$  and satisfying (3.2), follows from the application of Theorem 2.4 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, the cost  $\tilde{c}$  satisfies the twist condition on  $B_0 \times B_0$ ,  $p \# \mu'' \ll \text{Leb}_h$ , and  $\mu''_u \ll \text{Leb}_{d-h}$  as a conditional probability for a.e.  $u$ .

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 3.4 stated above with  $(\psi_1, \psi_2) = (\tilde{\Sigma}, \text{id})$  to obtain the existence of a unique optimal transport plan  $\pi_B^*$  between  $p \# \mu''$  and  $p \# \nu'$  for the cost  $c_\Sigma$ , which 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 3.5** (Recovering Theorem 4.2.3 from [Vay20]). Applying Theorem 3.2 when  $h = d$ , the optimal correspondence map between  $O_X \# \mu$  and  $O_Y \# \nu$  can be written as  $T_0 \circ p_{\mathbb{R}^d}$  with  $T_0 = \nabla f \circ \tilde{\Sigma}$ . The induced optimal correspondence map between  $\mu$  and  $\nu$  is then

$$T = O_Y^\top \circ (\nabla f \circ \tilde{\Sigma} \circ p_{\mathbb{R}^d}) \circ O_X = O_Y^\top \circ (\nabla f \circ \Sigma) \circ O_X = \nabla(f \circ O_Y) \circ O_Y^\top \circ \Sigma \circ O_X = \nabla \tilde{f} \circ M^*,$$

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

**3.2. The squared distance cost.** The (GW-SD) problem is invariant by translation of  $\mu$  and  $\nu$ . With no loss of generality, we assume that both measures are centered. Expanding the integrand gives

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

and the first two terms only depend on  $\mu$  and  $\nu$ , not on  $\pi$ . 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  $\mu$  and  $\nu$  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:

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

Any optimal correspondence plan  $\pi^*$  is also an optimal transport plan for the linearized problem (1.1) with cost

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

where  $M^* := \int y' x'^\top d\pi^*(x', y') \in \mathbb{R}^{d \times n}$ . In the cases where the rank of  $M^*$  is  $d$ , this linearized cost satisfies both the subtweak and the 2-tweak conditions, yielding an optimal 2-map that is also a map/anti-map, by compactness of the support of  $\mu$  and  $\nu$  and when  $\mu$  has a density. Similarly, when the rank of  $M^*$  is  $d - 1$ , the cost satisfies the 2-tweak condition, yielding an optimal 2-map structure. In the case where  $\text{rk } M^* \leq d - 2$ , none of the tweak conditions hold and there is *a priori* no obvious reason for the existence of an optimal correspondence map. Perhaps surprisingly, this property can actually be guaranteed.

**Theorem 3.6** (Existence of a Gromov–Monge map or 2-map for GW with squared distance cost). *Let  $n \geq d$  and  $\mu, \nu \in \mathcal{P}(\mathbb{R}^n) \times \mathcal{P}(\mathbb{R}^d)$  two measures with compact supports. Suppose that  $\mu \ll \text{Leb}_n$ . Let  $\pi^*$  be a solution of (GW-SD) and  $M^* := \int y'x'^\top d\pi^*(x', y')$ . Then:*

- (i) *if  $\text{rk } M^* = d$ , there exists an optimal correspondence 2-map, which is also a map/anti-map;*
- (ii) *if  $\text{rk } M^* = d - 1$ , there exists an optimal correspondence 2-map;*
- (iii) *if  $\text{rk } M^* \leq d - 2$ , there exists an optimal correspondence map, which can be written as*

$$T : \mathbb{R}^n \rightarrow \mathbb{R}^n, \quad T = O_Y^\top \circ T_0 \circ O_X,^3$$

where  $(O_X, O_Y) \in O_n(\mathbb{R})^2$ . Writing any  $x \in \mathbb{R}^n$  as  $x = (x_H, x_\perp) \in \mathbb{R}^h \times \mathbb{R}^{n-h}$  and defining a function  $\Phi$  akin to a partial polar change of coordinates

$$\Phi : \mathbb{R}^n \rightarrow (\mathbb{R}^h \times \mathbb{R}_{\geq 0}) \times S^{n-h-1}, \quad x \mapsto \left( \underbrace{(x_H, \|x_\perp\|^2)}_{x_B}, \underbrace{x_\perp / \|x_\perp\|}_{x_F} \right), \quad (3.3)$$

one can write the map  $T_0$  as

$$T_0 : \mathbb{R}^n \rightarrow \mathbb{R}^n, \quad \Phi \circ T_0(x) = (\tilde{c} \cdot \exp_{x_B}(\nabla f(x_B)), \exp_{x_F}(\nabla g_{x_B}(x_F))),$$

where explicit expressions for  $\tilde{c}$  and  $\tilde{c}\text{-exp}$  are given in (3.6) and (3.7) respectively,  $h = \text{rk } M^* \leq d - 2$ ,  $f : \mathbb{R}^{h+1} \rightarrow \mathbb{R}$  is  $\tilde{c}$ -convex and all the  $g_{x_B} : S^{n-h-1} \rightarrow \mathbb{R}$  are  $d_{S^{n-h-1}}^2/2$ -convex.

Therefore, regardless of the rank of  $M^*$ , there always exists an optimal correspondence 2-map.

*Proof of (i).* We show that in the case where  $M^*$  is of full rank, both the subtwist and the 2-twist conditions are satisfied. For the subtwist condition, remark that the equation

$$0 = \nabla_x C_{\pi^*}(x, y_1) - \nabla_x C_{\pi^*}(x, y_2) = -2(\|y_1\|^2 - \|y_2\|^2)x - 4(M^*)^\top(y_1 - y_2)$$

has no solution in  $x$  if  $\|y_1\| = \|y_2\|$  and a unique one otherwise. For the 2-twist condition, remark that any  $y \in \mathcal{Y}$  satisfying  $-(\|y\|^2 - \|y_0\|^2)x_0 - 2(M^*)^\top(y - y_0) = 0$  is fully determined by  $r := \|y\|^2 - \|y_0\|^2$  as

$$y = y_0 - \frac{1}{2}r(M^*)^{-1}x_0. \quad (3.4)$$

Taking the norm in (3.4) shows that  $r$  is a solution of a polynomial of degree two, and therefore that there are always at most two such points  $y$ . Both subtwist and 2-twist conditions are therefore satisfied. The uniqueness of optimal plans ensured by the subtwist condition then guarantees that the unique optimal transport plan for the linearized cost is induced both by a 2-map and a map/anti-map.

*Proof of (ii).* We show that in the case where  $\text{rk } M^* = d - 1$ , the 2-twist condition is satisfied. Let  $y \in \mathcal{Y}$  such that  $\|y\|^2 x_0 + (M^*)^\top y = \|y_0\|^2 x_0 + (M^*)^\top y_0 =: v$ . Up to singular value decomposition, suppose  $M^*$  rectangular diagonal in  $\mathbb{R}^{d \times n}$  with sorted singular values and write  $\tilde{\Sigma} = \text{diag}(\sigma_1, \dots, \sigma_h) \in \mathbb{R}^{h \times h}$  with  $h := \text{rk}(M^*)$ . Decompose each vector  $z$  of  $\mathbb{R}^n$  or  $\mathbb{R}^d$  as  $z = (z_H, z_\perp)$ , where  $z_H \in \mathbb{R}^h$  and  $z_\perp$  contains the remaining coordinates. The equation becomes:

$$\begin{cases} \|y\|^2 x_H + \tilde{\Sigma} y_H = v_H \\ \|y\|^2 x_\perp = v_\perp. \end{cases}$$

This imposes  $y$  to live in the intersection of a  $(d - 1)$ -dimensional sphere and of a  $(d - r)$ -dimensional affine subspace of  $\mathbb{R}^d$ . As  $r = d - 1$ , the latter subspace is a line and  $y$  therefore belongs to a set of at most 2 points, proving that the 2-twist condition is satisfied.

*Proof of (iii).* The case  $\text{rk } M^* \leq d - 2$  is a consequence of Theorem 2.4 and the proof is as follows. We consider the measure  $\nu$  as a measure on  $\mathbb{R}^n$  with  $d$ -dimensional support. Similarly to the inner product case, using a singular value decomposition, the cost  $C_{\pi^*}$  becomes

$$C_{\pi^*}(x, y) = -\|x\|^2 \|y\|^2 - \langle O_Y^\top \Sigma O_X x, y \rangle = -\|O_X x\|^2 \|O_Y y\|^2 - \langle \Sigma O_X x, O_Y y \rangle.$$

By Lemma 3.3 the problem reduces to  $\min_{\tilde{\pi}} \langle c_\Sigma, \tilde{\pi} \rangle$  for  $\tilde{\pi} \in \Pi(O_X \# \mu, O_Y \# \nu)$ , where  $c_\Sigma(x, y) := -\|x\|^2 \|y\|^2 - \langle \Sigma x, y \rangle$ . Up to permutation of the rows and columns of  $O_X$  and  $O_Y$ , we further assume  $\sigma_1 \geq \dots \geq \sigma_h > 0$ . Writing any  $z \in \mathbb{R}^n$  as  $z = (z_H, z_\perp) \in \mathbb{R}^h \times \mathbb{R}^{n-h}$ ,

$$c_\Sigma(x, y) = -\|x_H\|^2 \|y_H\|^2 - \|x_H\|^2 \|y_\perp\|^2 - \|x_\perp\|^2 \|y_H\|^2 - \|x_\perp\|^2 \|y_\perp\|^2 - \langle \tilde{\Sigma} x_H, y_H \rangle$$

<sup>3</sup>We consider  $\nu$  as a measure on  $\mathbb{R}^n$  with  $d$ -dimensional support.

$$= \tilde{c}(\varphi(x), \varphi(y)), \quad (3.5)$$

where  $\tilde{\Sigma}$  is the same as in the proof of (ii) above,  $\varphi : x \mapsto (x_H, \|x_\perp\|^2)$ , and  $\tilde{c} : (\mathbb{R}^h \times \mathbb{R}) \times (\mathbb{R}^h \times \mathbb{R}) \rightarrow \mathbb{R}$  is defined as

$$\tilde{c}((u, r), (v, s)) \mapsto -\|u\|^2\|v\|^2 - s\|u\|^2 - r\|v\|^2 - rs - \langle \tilde{\Sigma}u, v \rangle. \quad (3.6)$$

Indeed, the cost  $c_\Sigma(x, y)$  depends on the values of  $x_H$  and  $y_H$ , but merely on the squared norm of  $x_\perp$  and  $y_\perp$ . A direct computation proves that  $\tilde{c}$  satisfies the twist condition.

Now, the same argument as in [Example 2.5](#) 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  $F = S^{n-h-1} = \{x \in E_0 \mid \|x_\perp\| = 1\}$ . Using [Theorem 2.4](#), one obtains the existence of a Monge map  $T_0 : \mathbb{R}^n \rightarrow \mathbb{R}^n$  between  $\mu$  and  $\nu$  for the cost  $c_\Sigma$ . It has a structure akin to [Equation \(2.1\)](#) and decomposes into a map on the basis  $B_0 = \mathbb{R}^{h+1}$  and a map on each fiber  $F = S^{n-h-1}$ :

$$\Phi \circ T_0(x) = \left( \tilde{c}\text{-exp}_{(x_H, \|x_\perp\|^2)}(\nabla f(x_H, \|x_\perp\|^2)), \exp_{x_\perp/\|x_\perp\|}(\nabla g_{(x_H, \|x_\perp\|^2)}(x_\perp/\|x_\perp\|)) \right),$$

where  $\Phi$  is defined in [\(3.3\)](#). Note that the  $\tilde{c}$ -exponential map is given in closed form by

$$\tilde{c}\text{-exp}_{(u,r)}(w, t) = \begin{pmatrix} \tilde{\Sigma}^{-1}(w - 2tu) \\ t - \|\tilde{\Sigma}^{-1}(w - 2tu)\|^2 \end{pmatrix} \in \mathbb{R}^h \times \mathbb{R}, \quad (3.7)$$

where  $(u, r) \in \mathbb{R}^h \times \mathbb{R}$  is the base point and  $(w, t) \in \mathbb{R}^h \times \mathbb{R}$  the tangent vector. Lastly, note that the case where  $M^* = 0$  has not been explicitly treated. In this case, the cost is simply  $c(x, y) = -\|x\|^2\|y\|^2 = \tilde{c}(\varphi(x), \varphi(y))$  and the strategy above directly applies.  $\square$

### 3.3. Complementary study of GW with squared distance cost in the one-dimensional case.

The (GW-SD) problem being invariant by translation, we assume that measures  $\mu$  and  $\nu$  below are centered. In the one-dimensional case  $\mathcal{X}, \mathcal{Y} \subset \mathbb{R}$ , the linearized GW problem [\(1.1\)](#) reads, with  $\pi^*$  an optimal correspondence plan:

$$\min_{\pi \in \Pi(\mu, \nu)} \int_{\mathcal{X} \times \mathcal{Y}} (-x^2y^2 - 4mxy) d\pi(x, y), \quad \text{where } m = \int_{\mathcal{X} \times \mathcal{Y}} x'y' d\pi^*(x', y'), \quad (3.8)$$

and for any plan  $\pi \in \Pi(\mu, \nu)$  (not necessarily optimal), we denote by  $m(\pi) := \int xy d\pi(x, y)$  the *covariance* of  $\pi$ . Let us denote the OT cost function associated with [\(3.8\)](#) by  $c_m(x, y) := -x^2y^2 - 4mxy$ , as it only depends on the real parameter  $m$ . Let us apply [Theorem 3.6](#): when  $m \neq 0$ , there exists an optimal transport plan that is a 2-map and a map/anti-map, and when  $m = 0$ , there exists an optimal transport plan that is a 2-map. In the following sections, we study the tightness of this result, asking if there are cases where the optimal plan for this cost can be a map. In the one-dimensional case, the *submodularity* property of the cost function, or Spence–Mirrlees condition, will prove useful. It guarantees the optimality of the *non-decreasing rearrangement*  $\pi_{\text{mon}}^\uparrow$ , that match the quantiles of  $\mu$  and  $\nu$  with each other [[Car12](#); [San15](#); [PC19](#)]. We also define the *non-increasing rearrangement*  $\pi_{\text{mon}}^\downarrow$ , that match the quantiles in reverse, and call these two plans the *monotone rearrangements*.

**Definition 3.7** (Submodular cost). Let  $\mathcal{X}, \mathcal{Y} \subset \mathbb{R}$  be two intervals. A function  $c : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}$  is said to be *submodular* if for all  $(x, y) \in \mathcal{X} \times \mathcal{Y}$  and  $\delta_1, \delta_2 \geq 0$  such that  $(x + \delta_1, y + \delta_2) \in \mathcal{X} \times \mathcal{Y}$ ,

$$c(x, y) + c(x + \delta_1, y + \delta_2) \leq c(x, y + \delta_2) + c(x + \delta_1, y).$$

If  $c$  is twice differentiable, this global condition on the rectangle  $\mathcal{X} \times \mathcal{Y}$  is equivalent to the local condition

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

*Supermodularity* is defined with the reversed inequality.

We now state the well-known consequence of submodularity on the structure of optimal plans [[Car12](#); [McC12](#); [San15](#)].

**Proposition 3.8** (Optimality of monotone rearrangements). *Let  $\mu, \nu \in \mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{Y})$  of finite transport cost. If  $c$  satisfies (Submod), then  $\pi_{\text{mon}}^\uparrow$  is an optimal plan for (OT), unique if the inequality is strict everywhere on  $\mathcal{X} \times \mathcal{Y}$ . Similarly, supermodularity induces the optimality of  $\pi_{\text{mon}}^\downarrow$ .*

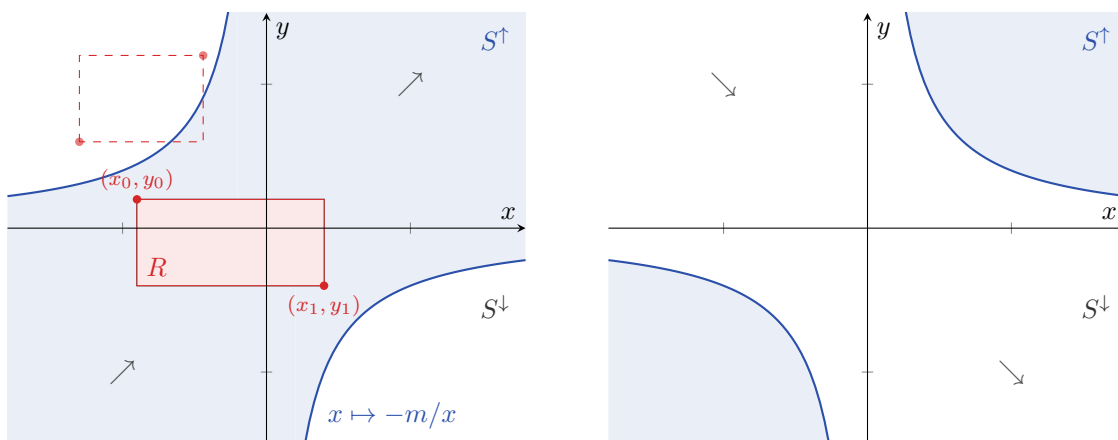


FIGURE 2. Submodularity region  $S^\uparrow$  (blue) and supermodularity region  $S^\downarrow$  for the linearized GW cost  $c_m$  with parameter  $m > 0$  (Left) or  $m < 0$  (Right), with an example of a rectangle  $R \subset S$  (red) defined by a decreasing matching  $(x_0, y_0) \rightarrow (x_1, y_1)$  that allows to obtain the monotonicity of optimal plans on  $S^\uparrow$  (see proof of Proposition 3.9). The same argument does not apply to rectangles in  $S^\downarrow$  (dashed red), see Remark 3.10. Up ( $\nearrow$ ) and down ( $\searrow$ ) arrows are drawn in regions where the monotonicity of optimal correspondence plans is given by Proposition 3.9.

The linearized GW cost with parameter  $m \geq 0$  is submodular on the region  $S^\uparrow := \{(x, y) \mid xy \geq -m\}$  and supermodular on  $S^\downarrow := \{(x, y) \mid xy \leq -m\}$  (see Figure 2 for an illustration), so we cannot directly apply this proposition. In general, for a given cost in dimension one, there are regions of sub and supermodularity and this property gives only few information without knowledge of any particular structure of the regions: to be applicable, submodularity requires a rectangular region. Given the form of the regions of sub and supermodularity in our particular case, we can state the following.

**Proposition 3.9** (Monotonicity of optimal plans on a subdomain). *Let  $m > 0$  and denote by  $\pi^*$  an optimal transportation plan for the cost  $c_m$ . Then,  $[\pi^*]_{|_{S^\uparrow}}$  (the plan restricted to the submodularity region) is the non-decreasing rearrangement between its marginals. Similarly, if  $m < 0$ ,  $[\pi^*]_{|_{S^\downarrow}}$  (the plan restricted to the supermodularity region) is the non-increasing rearrangement. If  $m = 0$ , there are four regions of sub/supermodularity, and on each of these regions, the plan is a monotone rearrangement.*

Figure 2 illustrates the regions, with up ( $\nearrow$ ) and down ( $\searrow$ ) arrows, where the monotonicity of optimal correspondence plans is given by this proposition.

*Proof.* In this proof, we use the fact that if  $\pi^*$  is optimal, then it is also optimal when restricted on a domain between the corresponding marginals. In particular, when  $m > 0$ , the plan  $[\pi^*]_{|_{S^\uparrow}}$  is necessarily optimal for the cost  $c_m$  between its marginals. Consider  $(x_0, y_0)$  and  $(x_1, y_1)$  in  $\text{supp}([\pi^*]_{|_{S^\uparrow}})$  such that  $x_0 < x_1$  and  $y_0 > y_1$ . The particular geometry of the space  $S^\uparrow$  ensures that in this particular configuration, the rectangle  $R = [x_0, x_1] \times [y_1, y_0]$  is necessarily contained in  $S^\uparrow$  (see Figure 2). As a consequence, the strict submodularity of the cost applies on  $R$  to prove that  $[\pi^*]_{|_R}$  is the non-decreasing rearrangement plan, contradicting the configuration. The proof goes similarly when  $m < 0$ . The case  $m = 0$  yields four different regions of sub/supermodularity which are quadrants. Since these regions are unions of rectangles on which the cost is strictly sub/supermodular, the argument above applies.  $\square$

**Remark 3.10** (Rectangular regions). Let us underline that it is not possible to apply the argument to the whole region  $S^\uparrow$  since any two points of  $S^\uparrow$  are not necessarily contained in a rectangle included in  $S^\uparrow$ . The proof above only uses the fact that this property is used for some particular configuration of

points, namely whenever  $x_0 < x_1$  and  $y_0 > y_1$ , which is (somewhat fortunately) exactly the configuration we want to consider when wondering if a map may be non-increasing. For instance, this argument does not apply to  $S^\perp$ , although the property is of course satisfied on every rectangle contained in it.

Before going into further details on our complementary study, we recall the discrete formulation of (OT) in dimension one. Given two sets  $\{x_1, \dots, x_N\}$  and  $\{y_1, \dots, y_M\}$  of elements 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  $U(a,b) := \{\pi \in \mathbb{R}^{N \times M} \mid \pi \mathbb{1}_M = a, \pi^\top \mathbb{1}_N = b\}$  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 (3.8), we denote by  $C_{\text{GW}(m)}$  the cost matrix, that has entries  $(C_{\text{GW}(m)})_{i,j} = -x_i^2 y_j^2 - 4m x_i y_j$  with  $m = \langle C_{xy}, \pi^* \rangle$  and  $(C_{xy})_{i,j} = x_i y_j$ . In our notation, the index of any cost matrix  $C$  is its associated OT cost function.

In the following sections, we study the optimality of the monotone rearrangements  $\pi_{\text{mon}}^\uparrow$  and  $\pi_{\text{mon}}^\downarrow$ . It is worth noting that by submodularity of the map  $(x, y) \mapsto -xy$ , these two plans have respective covariances  $m_{\text{min}}$  and  $m_{\text{max}}$ , where

$$\begin{cases} m_{\text{min}} = \min_{\pi} \langle C_{xy}, \pi \rangle \\ m_{\text{max}} = \max_{\pi} \langle C_{xy}, \pi \rangle \end{cases}, \quad \text{with } (C_{xy})_{i,j} = x_i y_j, \quad (3.9)$$

and that for any correspondence plan  $\pi$ , the value of its covariance  $m(\pi)$  lies in the interval  $[m_{\text{min}}, m_{\text{max}}]$ . We provide in the following a complementary study of the squared distance 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 3.6](#);
- (iii) a proof of the instability of having a monotone rearrangement as an 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 [GitHub](#)<sup>4</sup>.

**3.3.1. Adversarial computation of non-monotone optimal correspondence plans.** Theorem 3.1 of [\[Vay+19\]](#) claims that in the one-dimensional discrete case with  $N = M$  and  $a = b = \mathbb{1}_N$ , the optimal solution of (QAP) is either the non-decreasing rearrangement  $\pi_{\text{mon}}^\uparrow$  or the non-increasing one  $\pi_{\text{mon}}^\downarrow$ . It seems to be the case with 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 recently been exhibited in [\[BHS22\]](#). We further propose a procedure to automatically obtain additional counter-examples, demonstrating empirically that such adversarial measures occupy a non-negligible place in the space of discrete measures. We propose to perform a gradient descent over the space of discrete measures on  $\mathcal{X} \times \mathcal{Y}$  using an objective function that favors the strict sub-optimality of the monotone rearrangements. Let us now detail this procedure.

For  $N \geq 1$ , we consider the set of discrete measures 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  $\pi$  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  $c_{\text{GW}}$  the functional that takes a correspondence plan and returns its cost on the GW problem, we then define  $\mathcal{F} : \mathbb{R}^N \times \mathbb{R}^N \rightarrow \mathbb{R}$  by

$$\mathcal{F}(X, Y) := c_{\text{GW}}(\pi) - \min \{c_{\text{GW}}(\pi_{\text{mon}}^\uparrow), c_{\text{GW}}(\pi_{\text{mon}}^\downarrow)\},$$

where  $\pi = \text{id}(X, Y)$  and  $\pi_{\text{mon}}^\uparrow$  and  $\pi_{\text{mon}}^\downarrow$  are the monotone rearrangements between  $X$  and  $Y$ . This quantifies how well the plan  $\pi$  compares to the two monotone rearrangements. We generate  $N$  points

<sup>4</sup>The link of the code is <https://github.com/theodumont/monge-gromov-wasserstein>.

at random in  $[0, 1]^2$  and then perform a simple gradient descent over the positions of the points  $(X, Y) = (x_i, y_i)_i$  to minimize the objective function  $\mathcal{F}(X, Y)$  over  $(X, Y) \in (\mathbb{R}^N)^2$ . We include an early-stopping threshold  $-\tau < 0$  since when  $\mathcal{F}(\pi)$  becomes negative—i.e. we found an adversarial example—the objective function often decreases exponentially fast. The procedure can be found in [Algorithm 1](#) below. We implemented it using PyTorch’s automatic differentiation [[Pas+19](#)] and used [[Blo+20](#)] to implement a differentiable sorting operator `sort` to compute the monotone rearrangements. Adversarial plans  $\pi_{\text{adv}} = \text{id}(X_{\text{adv}}, Y_{\text{adv}})$  obtained by [Algorithm 1](#) 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  $\mathcal{F}(X_{\text{adv}}, Y_{\text{adv}}) < 0$ , proving the sub-optimality of the latter.

---

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

---

**Parameters:**

- $N \geq 1$ : number of points of the measures
- $N_{\text{iter}} \geq 1$ : maximum number of iterations
- $\eta > 0$ : step size
- $\tau > 0$ : 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:    $\pi_{\text{mon}}^\uparrow \leftarrow \text{id}(\text{sort}(X), \text{sort}(Y))$                     $\triangleright$  `id` is the identity mapping
- 5:    $\pi_{\text{mon}}^\downarrow \leftarrow \text{id}(\text{sort}(X), \text{reverse}(\text{sort}(Y)))$
- 6:    $\pi \leftarrow \text{id}(X, Y)$
- 7:    $\mathcal{F}(X, Y) \leftarrow \text{GW}(\pi) - \min(\text{GW}(\pi_{\text{mon}}^\uparrow), \text{GW}(\pi_{\text{mon}}^\downarrow))$             $\triangleright$  `GW` computes  $c_{\text{GW}}$
- 8:   **if**  $\mathcal{F}(X, Y) \leq -\tau$  **then** stop    $\triangleright$  early stopping
- 9:    $(X, Y) \leftarrow (X, Y) - \eta \nabla \mathcal{F}(X, Y)$                                     $\triangleright$  step of gradient descent
- 10: **end for**
- 11: return  $\pi_{\text{adv}} = \text{id}(X, Y)$

**Output:** a plan  $\pi_{\text{adv}}$  with better GW cost than  $\pi_{\text{mon}}^\uparrow$  and  $\pi_{\text{mon}}^\downarrow$

---

On [Figure 3](#) is displayed an example of adversarial plans obtained following this procedure. It can be observed that during the descent, the plan  $\pi$  has difficulties getting out of what seems to be a saddle point consisting of being (close to) a monotone rearrangement 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  $\mu$ , and two for  $\nu$ , one on each tail.

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 3.11** (Tightness of [Theorem 3.6](#)). There exist measures  $\mu$  and  $\nu$  for which no optimal correspondence plan for (GW-SD) is a map, but always a union of two graphs (that of two maps or one map and one anti-map). This result holds even if  $\mu$  has a density, classical assumption for the existence of an optimal transport map.

In order to approximate numerically the case of a measure that has density with respect to the Lebesgue measure, we convolve our adversarial measures  $\mu_{\text{adv}} = (X_{\text{adv}}, \mathbb{1}_N)$  and  $\nu_{\text{adv}} = (Y_{\text{adv}}, \mathbb{1}_N)$  with a Gaussian measure of standard deviation  $\sigma$  and represent them in Eulerian coordinates  $a_{\text{adv}}$  and  $b_{\text{adv}}$ ; that is we evaluate the closed form densities on a fine enough regular grid. When  $\sigma$  is large, the optimal correspondence plan is probably induced by a monotone rearrangement, as it is the case very frequently empirically; on the contrary, if  $\sigma$  is sufficiently small, i.e. when the measures are very close

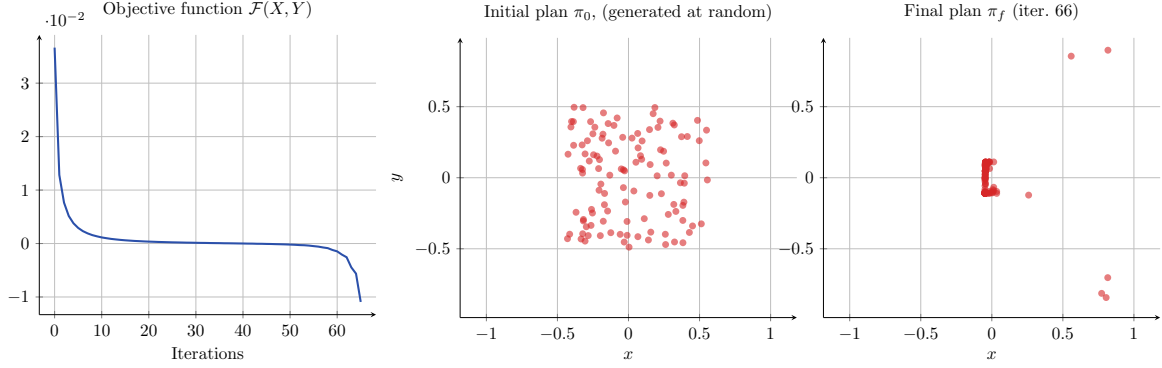


FIGURE 3. Gradient descent results with parameters  $N = 122$ ,  $\eta = 26$ ,  $\tau = 2$ .

to their sum of Dirac measures discrete analogous, the optimal correspondence plan should not be a monotone rearrangement, by construction of  $\mu_{\text{adv}}$  and  $\nu_{\text{adv}}$ .

**Remark 3.12** (Counter-examples from Algorithm 1 sometimes agree with Conjecture 3.11). Because of the adversarial nature of  $\pi_{\text{adv}}$  for the sub-optimality of  $\pi_{\text{mon}}^{\uparrow}$  and  $\pi_{\text{mon}}^{\downarrow}$ , we know that when  $\sigma$  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  $\pi_{\text{adv}}$  will agree with Conjecture 3.11. Surprisingly, it sometimes does, as the numerical experiments below suggest.

In order to find the optimal correspondence plan  $\pi^*$  between  $\mu_{\text{adv}}$  and  $\nu_{\text{adv}}$ , we leverage the fact that  $\pi^*$  is a solution of its associated linearized problem. Therefore, a minimizer of the GW functional is given by

$$\arg \min_{\pi_m^*} \left\{ \text{GW}(\pi_m^*) \mid \pi_m^* \in \arg \min_{\pi \in U(a_{\text{adv}}, b_{\text{adv}})} \langle C_{\text{GW}(m)}, \pi \rangle, m \in [m_{\min}, m_{\max}] \right\}, \quad (3.10)$$

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 (3.9), discretize the interval  $[m_{\min}, m_{\max}]$  with  $N_m$  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 2-map structure. The procedure is described in Algorithm 2.

We display the results on Figure 4, where we plot the optimal correspondence plan  $\pi^*$  in two cases, starting from an adversarial plan:

- (i) with both marginals  $\mu_{\text{adv}}$  and  $\nu_{\text{adv}}$  convolved as to simulate densities;
- (ii) with only the first marginal  $\mu_{\text{adv}}$  convolved and the second marginal  $\nu_{\text{adv}}$  being a discrete measure with uniform mass.

To facilitate the reading of Figure 4, we draw a blue pixel at locations  $x$  on the discretized  $x$ -axis if  $x$  has two images (making  $\pi^*$  a 2-map), and at locations  $y$  on the  $y$ -axis if  $y$  has two pre-images (making  $\pi^*$  a anti-2-map). In both cases, we observe that  $\pi^*$  is not a map but a 2-map instead, similarly to [CMN10, Section 4.5]. Note that in case (ii),  $\nu$  being atomic, there cannot exist a map from  $\nu$  to  $\mu$ , so in both (i) and (ii) we numerically exhibit an instance where there is *a priori* no map from neither  $\mu$  to  $\nu$  nor  $\nu$  to  $\mu$ . 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 optimal plan gives mass to a region where the cost is submodular or supermodular, it has respectively a non-decreasing or non-increasing behavior in this region.

**3.3.2. Empirical instability of the optimality of monotone rearrangements.** The above study demonstrates that there exist probability measures  $\mu$  and  $\nu$  for which property

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

**Algorithm 2** Generating 2-maps from adversarial examples.**Input:** an adversarial plan  $\pi_{\text{adv}} = \text{id}(X_{\text{adv}}, Y_{\text{adv}})$  obtained from [Algorithm 1](#)**Parameters:**

- $\sigma$ : standard deviation of convolution
- $N_x$ : discretization precision
- $N_m$ : discretization precision of the interval  $[m_{\min}, m_{\max}]$

**Algorithm:**

```

1:  $a_{\text{adv}} \leftarrow \text{convolution}(X_{\text{adv}}, \sigma, N_x)$ 
2:  $b_{\text{adv}} \leftarrow \text{convolution}(Y_{\text{adv}}, \sigma, N_x)$  ▷ (optional for  $\nu_{\text{adv}}$ )
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: scores  $\leftarrow []$ 
6: for  $m \in \{m_{\min}, \dots, m_{\max}\}$  do ▷ with  $N_m$  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:  $b \leftarrow$  “ $\pi^*$  is a 2-map”
12: return  $\pi^*, b$ 

```

**Outputs:**

- $\pi^*$ : optimal plan for GW
- $b$ : boolean asserting if  $\pi^*$  is a 2-map

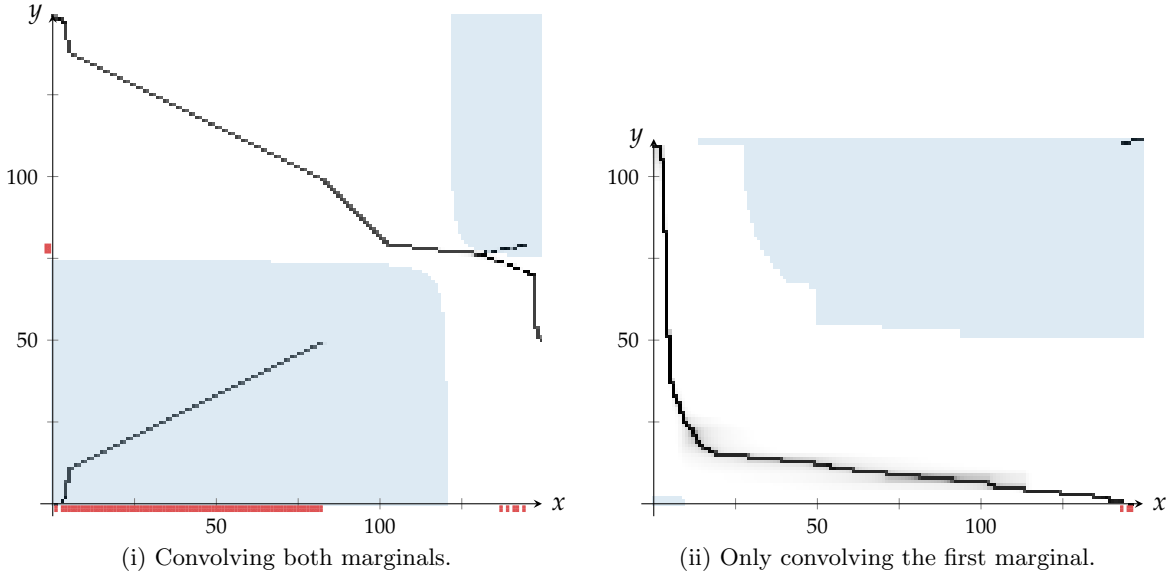


FIGURE 4. Optimal correspondence plan (in log scale) obtained with our procedure, with 2-map and anti-2-map coordinates (red) and submodularity regions (blue). Parameters:  $\sigma = 5.10^{-3}$ ,  $N_x = 150$ ,  $N_m = 2000$ .

does not hold. However, as it is very likely in practice when generating discrete measures at random, one could ask if property  $P$  is at least *stable*, i.e. if when we have  $\mu_0$  and  $\nu_0$  satisfying  $P(\mu_0, \nu_0)$  there is



a small ball around  $\mu_0$  and  $\nu_0$  (for a given distance, say Wasserstein-2) inside which property  $P$  remains valid. A negative answer to this—besides, in the symmetric case—is given by the counter-example by Beinert et al. [BHS22] with an increasing number of points:

**Proposition 3.13** (Instability of the optimality of monotone rearrangements). *There exists two symmetric measures  $\mu, \nu$  on  $\mathbb{R}$  and sequences  $(\mu_n)_n, (\nu_n)_n$  that weakly converge to  $\mu, \nu$  such that optimal plans  $\pi_n$  between  $\mu_n$  and  $\nu_n$  are never a monotone rearrangement.*

*Proof.* We consider  $\mu = \nu = \delta_0$  and the discrete measures  $\mu_n = \frac{1}{n} \sum_{i=1}^n \delta_{x_i}$  and  $\nu_n = \frac{1}{n} \sum_{i=1}^n \delta_{y_i}$  defined as follows for  $n \geq 7$ :

$$x_i := \begin{cases} -1 & \text{for } i = 1 \\ (i - \frac{n+1}{2}) \frac{1}{n^2} & \text{for } i = 2, \dots, n-1 \\ 1 & \text{for } i = n \end{cases} \quad \text{and} \quad y_i := \begin{cases} -1 & \text{for } i = 1 \\ -1 + \frac{1}{n^2} & \text{for } i = 2 \\ (i-2) \frac{1}{n^2} & \text{for } i = 3, \dots, n \end{cases}$$

which is simply the counter-example from Beinert et al. [BHS22] with  $n$  points and  $\varepsilon_n = 1/n^2$ . Since  $n \geq 7$ ,  $\varepsilon_n < 2/(n-3)$  and the identity or anti-identity mappings are not optimal between  $\mu_n$  and  $\nu_n$ . By direct computation,

$$W_2^2(\delta_0, \mu_n) = O(2/n + \varepsilon_n^2 n^2) \xrightarrow{n \rightarrow \infty} 0,$$

and the exact same goes for  $\nu = \delta_0$  and  $\nu_n$ .  $\square$

One can actually obtain non-degenerate (although not symmetric anymore) examples of such measures  $\mu, \nu$ . We start from the counter-example given in [BHS22] with  $N = 7$  points and  $\varepsilon = 10^{-2}$ , that we convolve with a Gaussian measure of standard deviation  $\sigma$  as before. We then plot as a function of  $m \in [m_{\min}, m_{\max}]$  the (true) GW cost of a plan  $\pi_m^*$  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 covariances of optimal correspondence plans, as explained in Section 3.3.1. Hence if  $\sigma$  is small, optimal correspondence plans are not monotone by construction and the minima are not located on the boundary of the domain. On the contrary, when  $\sigma$  is large, a monotone rearrangement is optimal again. 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  $\sigma$ .

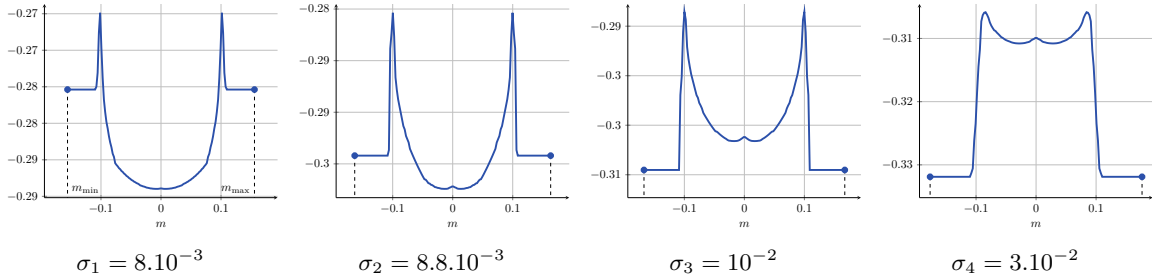


FIGURE 5. Evolution of the graph of  $m \mapsto \text{GW}(\pi_m^*)$  when making  $\sigma$  vary on the counter-example of Beinert et al. [BHS22] with  $N = 7$  points and  $\varepsilon = 10^{-2}$ . Parameters:  $N_x = 100$ ,  $N_m = 150$ .

Looking at Figure 5, it is worth noting that there is an incentive for optimal plans at covariances 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  $\sigma_4$ , the monotone rearrangements are optimal, as their covariances realize the minimum of  $m \mapsto \text{GW}(\pi_m^*)$ ; unlike for  $\sigma_1$  and  $\sigma_2$ , for which the minimum value of the plot is located near zero. Hence there should exist a  $\sigma_0 \in (\sigma_2, \sigma_3)$  for which the convolved measures have both  $\pi_{\text{mon}}^\uparrow, \pi_{\text{mon}}^\downarrow$  and another  $\pi_0$  as optimal correspondence plans; it is direct that property  $P$  does not hold in the neighborhood of these specific measures  $\mu_0$  and  $\nu_0$ .

3.3.3. *A positive result for measures with two components.* In the following,  $\mu_1, \mu_2, \nu_1$  and  $\nu_2$  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

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

Note that  $\mu_1$  and  $\tau_{K\#}\mu_2$  have disjoint supports, and same for  $\nu_1$  and  $\tau_{K\#}\nu_2$ . We have the following result.

**Proposition 3.14** (Optimality of monotone rearrangements for measures with two components). *For  $K$  large enough, the unique optimal plan for the squared distance cost between  $\mu$  and  $\nu$  is given by one of the two monotone rearrangements (non-decreasing or non-increasing).*

**Remark 3.15** (Long-range correspondences). Conditions of [Proposition 3.14](#) illustrate that monotone rearrangements are favored when  $\mu$  and  $\nu$  both contain one or more outliers. The proof actually shows the importance of long-range correspondences or global effects over the local correspondences on the plan. In other words, even though monotone rearrangements may not be optimal locally, 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, pairs of points that are at long distances tend to be put in correspondence. In turn, this correspondence, as shown in the proof, favors monotone rearrangements. This argument gives some insight into the fact that a monotone rearrangement is often optimal and it is made quantitative by [Proposition 3.14](#).

We first prove the following lemma:

**Lemma 3.16.** *In the setting described above, there exists  $K_0 > 0$  such that if  $K \geq K_0$ , every  $\pi$  optimal plan for  $\text{GW}(\mu, \nu)$  can be decomposed as  $\pi = \pi_1 + \pi_2$ , where either:*

- (i)  $\pi_1$  is supported on  $A \times A$  and  $\pi_2$  on  $(A + K) \times (A + K)$  (that is, we separately transport  $\mu_1$  to  $\nu_1$  and  $\tau_{K\#}\mu_2$  to  $\tau_{K\#}\nu_2$ ), or
- (ii)  $\pi_1$  is supported on  $A \times (A + K)$  and  $\pi_2$  on  $A \times (A + K)$  (that is, we separately transport  $\mu_1$  to  $\tau_{K\#}\nu_2$  and  $\mu_2$  to  $\tau_{K\#}\nu_1$ ).

Furthermore, whenever  $t \neq \frac{1}{2}$ , only (i) can occur.

*Proof.* Consider first the case  $t = 1/2$ . To alleviate notation, we introduce  $A_1 := A$  and  $A_2 := A + K$ . We can now decompose any plan  $\pi$  as  $\pi_{11} + \pi_{12} + \pi_{21} + \pi_{22}$  where  $\pi_{ij}$  denotes the restriction of the plan  $\pi$  to the product  $A_i \times A_j$ , for  $i, j \in \{1, 2\}$ . Let us also denote by  $r$  the mass of  $\pi_{12}$ . One has  $0 \leq r \leq 1/2$  and by symmetry, one can choose  $r \leq 1/4$ —otherwise, we switch the roles of  $A_1$  and  $A_2$  for  $\nu$  since the cost is invariant by isometries. Remark that, due to marginal constraints, the total mass of  $\pi_{11}$  and  $\pi_{22}$  is  $1/2 - r$  and the mass of  $\pi_{21}$  is  $r$ . Therefore, it is possible to consider a transport plan  $\tilde{\pi}_{11}$  between the first marginal of  $\pi_{12}$  and the second marginal of  $\pi_{21}$ . Similarly, let  $\tilde{\pi}_{22}$  be a transport plan between the first marginal of  $\pi_{21}$  and the second marginal of  $\pi_{12}$ . We then define a competitor plan  $\tilde{\pi} := \pi_{11} + \tilde{\pi}_{11} + \pi_{22} + \tilde{\pi}_{22}$ . Slightly overloading notation, we introduce

$$\text{GW}(\pi, \gamma) := \int c \, d\pi \otimes \gamma. \tag{3.11}$$

The first step is to get a lower bound on the term  $\text{GW}(\pi, \pi)$ . We expand it 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. (i) 8 terms are of the type  $\text{GW}(\pi_{12}, \pi_{11}) \geq r(1/2 - r)(K^2 - \Delta^2)^2$ . Indeed, since one compares pairs of points  $(x, x')$  and  $(y, y')$  satisfying  $(x, y) \in A_1 \times A_1$  and  $(x', y') \in A_1 \times A_2$ ,  $(x - x')^2$  is upper-bounded by  $\Delta^2$  and  $(y - y')^2$  lower-bounded by  $K^2$ . The bound above then follows after integrating. (ii) The 4 remaining

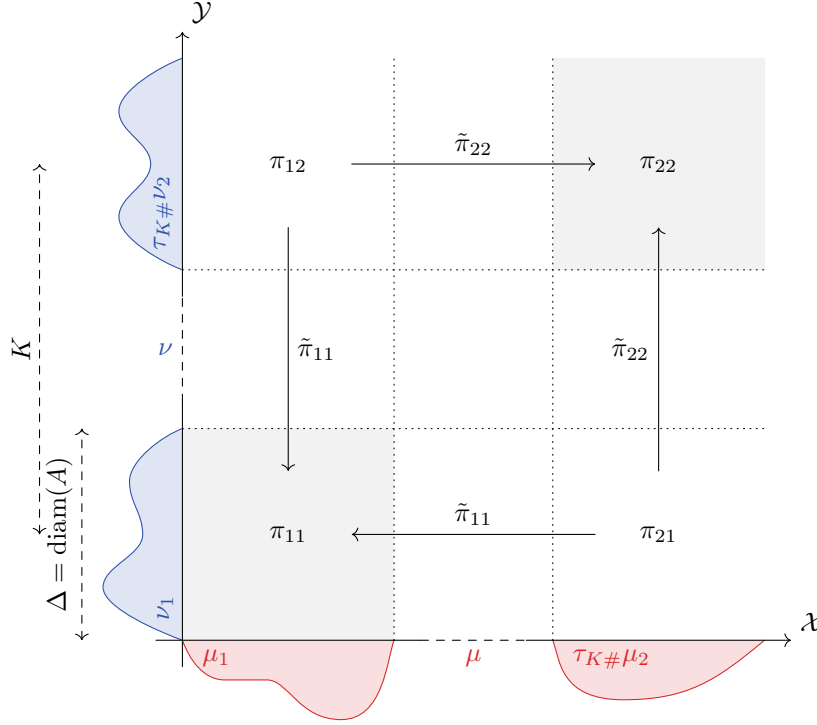


FIGURE 6. Visual sketch of the proof of Lemma 3.16.

terms are of the type  $\text{GW}(\pi_{12}, \pi_{21}) \geq 0$ . Summing everything yields  $R \geq 8r(1/2 - r)(K^2 - \Delta^2)^2$ . We now upper-bound the competitor. Similarly, one has

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

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 last two terms are upper-bounded by  $2r(1/2 - r)\Delta^2$ , since one compares squared distances within  $A_1$ . By elementary inequalities again (see Figure 6), the first two 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

$$\text{GW}(\pi, \pi) - \text{GW}(\tilde{\pi}, \tilde{\pi}) \geq r(8(1/2 - r)(K^2 - \Delta^2)^2 - 4(1/2 - r)\Delta^2 - 2(2K\Delta + \Delta^2)^2). \quad (3.12)$$

Then, since  $1/2 - r \geq 1/4$ , the polynomial function in  $K$  on the right-hand side of Equation (3.12) tends to  $+\infty$  uniformly in  $r \in [0, \frac{1}{4}]$  when  $K$  goes to infinity, and the result follows: there exists  $K > 0$  such that the polynomial function above is non-negative, for instance,  $K := \max(0, K_0)$  where  $K_0$  is the largest root of the polynomial.

The proof in the case  $t > 1/2$  is even simpler: since  $t - r > t - 1/2$ , there is no choice in the matching of the two measures and it is determined by the corresponding masses. One can then directly apply the argument above. The same reasoning applies to the case  $t < 1/2$ .  $\square$

We now prove Proposition 3.14.

*Proof of Proposition 3.14.* Thanks to Lemma 3.16, we know that we can restrict to transportation plans  $\pi = \pi_1 + \pi_2$  where, up to flipping  $\nu$ , we can assume that  $\pi_1$  is supported on  $A \times A$  and  $\pi_2$  on

$(A + K) \times (A + K)$ .<sup>5</sup> Using again the bilinear form  $\text{GW}(\pi, \gamma)$  defined in (3.11), the objective value 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 found  $\pi_2^*$  optimal. Let us minimize in  $\pi_1$  the resulting quadratic problem:

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

We know that if  $\pi_1^*$  is a minimizer of this quantity, it must also be a solution of the *linear* problem

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

which is exactly the optimal transportation problem for the cost

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

Now, using the relation  $((x - x'')^2 - (y - y'')^2)^2 = ((x - y) - (x'' - y''))^2 ((x + y) - (x'' + y''))^2$  and the fact that  $\pi_2^*$  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

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

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

$$\partial_{xy}c(x, y) = -2K^2 + O(K) < 0 \quad \text{for } K \text{ large enough.}$$

The plan  $\pi_1^*$  is optimal for a submodular cost, and by Proposition 3.8 must be the non-decreasing rearrangement between  $\mu_1$  and  $\nu_1$ . By symmetry, so is  $\pi_2^*$ .  $\square$

#### ACKNOWLEDGEMENTS

We thank Facundo Mémoli and Gabriel Peyré for their helpful remarks. We also thank the anonymous reviewers who substantially helped us improve the clarity of the paper. This work was supported by the Bézout Labex (New Monge Problems), funded by ANR, reference ANR-10-LABX-58.

#### 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 (cit. on pp. 9, 35).
- [AJ18] David Alvarez-Melis and Tommi S Jaakkola. *Gromov–Wasserstein alignment of word embedding spaces*. 2018. arXiv: 1809.00013 (cit. on pp. 2, 3).
- [AKM11] Najma Ahmad, Hwa Kil Kim, and Robert J McCann. “Optimal transportation, topology and uniqueness”. In: *Bulletin of Mathematical Sciences* 1.1 (2011), pp. 13–32 (cit. on p. 6).
- [ANN18] Alexandr Andoni, Assaf Naor, and Ofer Neiman. “Snowflake universality of Wasserstein spaces”. In: *Annales Scientifiques de l’École Normale Supérieure*. Vol. 51. Societe Mathematique de France. 2018, pp. 657–700 (cit. on p. 5).
- [AXK17] Brandon Amos, Lei Xu, and J Zico Kolter. “Input convex neural networks”. In: *International Conference on Machine Learning*. PMLR. 2017, pp. 146–155 (cit. on p. 9).

<sup>5</sup>Note: this is where the choice is made, as in the proof of Lemma 3.16, between the non-decreasing and non-increasing rearrangements. Using this convention, the non-decreasing rearrangement is shown to be optimal.

- [BBM05] Alexander C Berg, Tamara L Berg, and Jitendra Malik. “Shape matching and object recognition using low distortion correspondences”. In: *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR’05)*. Vol. 1. IEEE, 2005, pp. 26–33 (cit. on p. 2).
- [BHS22] Robert Beinert, Cosmas Heiss, and Gabriele Steidl. *On assignment problems related to Gromov–Wasserstein distances on the real line*. 2022. arXiv: [2205.09006](https://arxiv.org/abs/2205.09006) (cit. on pp. 2, 8, 20, 21, 24).
- [Blo+20] Mathieu Blondel, Olivier Teboul, Quentin Berthet, and Josip Djolonga. “Fast differentiable sorting and ranking”. In: *International Conference on Machine Learning*. PMLR, 2020, pp. 950–959 (cit. on p. 21).
- [Bre87] Yann Brenier. “Décomposition polaire et réarrangement monotone des champs de vecteurs”. In: *CR Acad. Sci. Paris Sér. I Math.* 305 (1987), pp. 805–808 (cit. on pp. 2, 6, 10, 15, 31).
- [Bun+19] Charlotte Bunne, David Alvarez-Melis, Andreas Krause, and Stefanie Jegelka. *Learning generative models across incomparable spaces*. 2019. arXiv: [1905.05461](https://arxiv.org/abs/1905.05461) (cit. on p. 2).
- [Car12] Guillaume Carlier. “Optimal transportation and economic applications”. In: *Lecture Notes* (2012) (cit. on p. 18).
- [Chi07] David Chiron. “On the definitions of Sobolev and BV spaces into singular spaces and the trace problem”. In: *Communications in Contemporary Mathematics* 9.04 (2007), pp. 473–513 (cit. on p. 34).
- [CM19] Samir Chowdhury and Facundo Mémoli. “The Gromov–Wasserstein distance between networks and stable network invariants”. In: *Information and Inference: A Journal of the IMA* 8.4 (2019), pp. 757–787 (cit. on p. 4).
- [CMN10] Pierre-André Chiappori, Robert J McCann, and Lars P Nesheim. “Hedonic price equilibria, stable matching, and optimal transport: equivalence, topology, and uniqueness”. In: *Economic Theory* 42.2 (2010), pp. 317–354 (cit. on pp. 6, 22).
- [Dem+20] Pınar Demetçi, Rebecca Santorella, Björn Sandstede, William Stafford Noble, and Ritambhara Singh. *Gromov–Wasserstein optimal transport to align single-cell multi-omics data*. 2020. DOI: [10.1101/2020.04.28.066787](https://doi.org/10.1101/2020.04.28.066787) (cit. on p. 2).
- [Dem+22] Pınar Demetçi, Rebecca Santorella, Manav Chakravarthy, Bjorn Sandstede, and Ritambhara Singh. “SCOTv2: single-cell multiomic alignment with disproportionate cell-type representation”. In: *Journal of Computational Biology* 29.11 (2022), pp. 1213–1228 (cit. on p. 2).
- [DM22] Nicolás De Ponti and Andrea Mondino. “Entropy–Transport distances between unbalanced metric measure spaces”. In: *Probability Theory and Related Fields* 184.1 (2022), pp. 159–208 (cit. on p. 2).
- [FGM10] Joaquin Fontbona, Hélène Guérin, and Sylvie Méléard. “Measurability of optimal transportation and strong coupling of martingale measures”. In: *Electronic communications in probability* 15 (2010), pp. 124–133 (cit. on pp. 12, 14, 33, 34).
- [FH74] Jacques Faraut and Khélifa Harzallah. “Distances hilbertiennes invariantes sur un espace homogène”. In: *Annales de l’Institut Fourier*. Vol. 24. 1974, pp. 171–217 (cit. on p. 5).
- [GJB19] Edouard Grave, Armand Joulin, and Quentin Berthet. “Unsupervised alignment of embeddings with wasserstein procrustes”. In: *The 22nd International Conference on Artificial Intelligence and Statistics*. PMLR, 2019, pp. 1880–1890 (cit. on p. 2).
- [Gro+99] Mikhael Gromov, Misha Katz, Pierre Pansu, and Stephen Semmes. *Metric structures for Riemannian and non-Riemannian spaces*. Vol. 152. Springer, 1999 (cit. on pp. 2, 3).
- [HGL21] YoonHaeng Hur, Wenxuan Guo, and Tengyuan Liang. *Reversible Gromov–Monge sampler for simulation-based inference*. 2021. arXiv: [2109.14090](https://arxiv.org/abs/2109.14090) (cit. on p. 7).
- [KB57] Tjalling C Koopmans and Martin Beckmann. “Assignment problems and the location of economic activities”. In: *Econometrica: journal of the Econometric Society* (1957), pp. 53–76 (cit. on pp. 2, 4).
- [Kon76] Hiroshi Konno. “Maximization of a convex quadratic function under linear constraints”. In: *Mathematical programming* 11.1 (1976), pp. 117–127 (cit. on p. 4).

- [Kor+21] Alexander Korotin, Lingxiao Li, Aude Genevay, Justin M Solomon, Alexander Filipov, and Evgeny Burnaev. “Do neural optimal transport solvers work? a continuous wasserstein-2 benchmark”. In: *Advances in Neural Information Processing Systems* 34 (2021), pp. 14593–14605 (cit. on p. 9).
- [KSB22] Alexander Korotin, Daniil Selikhanovych, and Evgeny Burnaev. “Neural optimal transport”. In: *arXiv preprint arXiv:2201.12220* (2022) (cit. on p. 9).
- [Loe09] Grégoire Loeper. “On the regularity of solutions of optimal transportation problems”. In: *Acta mathematica* 202.2 (2009), pp. 241–283 (cit. on p. 6).
- [Lyo13] Russell Lyons. “Distance covariance in metric spaces”. In: *The Annals of Probability* 41.5 (2013), pp. 3284–3305 (cit. on p. 5).
- [Mak+20] Ashok Makkuva, Amirhossein Taghvaei, Sewoong Oh, and Jason Lee. “Optimal transport mapping via input convex neural networks”. In: *International Conference on Machine Learning*. PMLR, 2020, pp. 6672–6681 (cit. on p. 2).
- [McC01] Robert J McCann. “Polar factorization of maps on Riemannian manifolds”. In: *Geometric & Functional Analysis GAFA* 11.3 (2001), pp. 589–608 (cit. on pp. 2, 6).
- [McC12] Robert J McCann. *A glimpse into the differential topology and geometry of optimal transport*. 2012. arXiv: [1207.1867](https://arxiv.org/abs/1207.1867) (cit. on pp. 6, 18).
- [Mém+23] Facundo Mémoli, Axel Munk, Zhengchao Wan, and Christoph Weitkamp. “The ultrametric Gromov–Wasserstein distance”. In: *Discrete & Computational Geometry* (2023), pp. 1–73 (cit. on p. 2).
- [Mém07] Facundo Mémoli. “On the use of Gromov–Hausdorff distances for shape comparison”. In: *Eurographics Symposium on Point-Based Graphics*. The Eurographics Association, 2007 (cit. on pp. 2–4).
- [Mém08] Facundo Mémoli. “Gromov–Hausdorff distances in Euclidean spaces”. In: *2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*. IEEE, 2008, pp. 1–8 (cit. on p. 2).
- [Mém11] Facundo Mémoli. “Gromov–Wasserstein distances and the metric approach to object matching”. In: *Foundations of Computational Mathematics* 11.4 (2011), pp. 417–487 (cit. on pp. 2, 4).
- [MG11] Robert J McCann and Nestor Guillen. “Five lectures on optimal transportation: geometry, regularity and applications”. In: *Analysis and geometry of metric measure spaces: lecture notes of the séminaire de Mathématiques Supérieure (SMS) Montréal* (2011), pp. 145–180 (cit. on p. 6).
- [ML18] Haggai Maron and Yaron Lipman. “(Probably) concave graph matching”. In: *Advances in Neural Information Processing Systems* 31 (2018) (cit. on p. 4).
- [MN22a] Facundo Mémoli and Tom Needham. *Comparison results for Gromov–Wasserstein and Gromov–Monge distances*. 2022. arXiv: [2212.14123](https://arxiv.org/abs/2212.14123) (cit. on pp. 4, 7).
- [MN22b] Facundo Mémoli and Tom Needham. “Distance distributions and inverse problems for metric measure spaces”. In: *Studies in Applied Mathematics* 149.4 (2022), pp. 943–1001 (cit. on pp. 2, 7).
- [Moa16] Abbas Moameni. “A characterization for solutions of the Monge–Kantorovich mass transport problem”. In: *Mathematische Annalen* 365.3 (2016), pp. 1279–1304 (cit. on p. 6).
- [MPW12] Robert J McCann, Brendan Pass, and Micah Warren. “Rectifiability of optimal transportation plans”. In: *Canadian Journal of Mathematics* 64.4 (2012), pp. 924–934 (cit. on p. 6).
- [MR20] Abbas Moameni and Ludovic Rifford. “Uniquely minimizing costs for the Kantorovitch problem”. In: *Annales de la Faculté des sciences de Toulouse: Mathématiques*. Vol. 29. 2020, pp. 507–563 (cit. on p. 6).
- [MS04] Facundo Mémoli and Guillermo Sapiro. “Comparing point clouds”. In: *Proceedings of the 2004 Eurographics/ACM SIGGRAPH symposium on Geometry processing*. 2004, pp. 32–40 (cit. on p. 2).

- [MS05] Facundo Mémoli and Guillermo Sapiro. “A theoretical and computational framework for isometry invariant recognition of point cloud data”. In: *Foundations of Computational Mathematics* 5.3 (2005), pp. 313–347 (cit. on p. 2).
- [MSW21] Facundo Mémoli, Zane Smith, and Zhengchao Wan. *The Gromov–Hausdorff distance between ultrametric spaces: its structure and computation*. 2021. arXiv: [2110.03136](#) (cit. on p. 2).
- [Pas+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: *Advances in Neural Information Processing Systems*. Vol. 32. 2019, pp. 8024–8035 (cit. on p. 21).
- [PC19] Gabriel Peyré and Marco Cuturi. “Computational optimal transport: with applications to data science”. In: *Foundations and Trends in Machine Learning* 11.5-6 (2019), pp. 355–607 (cit. on pp. 6, 18).
- [Red+20] Ivegen Redko, Titouan Vayer, Rémi Flamary, and Nicolas Courty. “CO-optimal transport”. In: *Advances in Neural Information Processing Systems* 33 (2020), pp. 17559–17570 (cit. on p. 2).
- [RW09] R Tyrrell Rockafellar and Roger J-B Wets. *Variational analysis*. Vol. 317. Springer Science & Business Media, 2009 (cit. on pp. 14, 31–33).
- [San15] Filippo Santambrogio. “Optimal transport for applied mathematicians”. In: *Birkhäuser, NY* 55.58-63 (2015), p. 94 (cit. on pp. 2, 6, 12, 18).
- [Sch38] Isaac J Schoenberg. “Metric spaces and positive definite functions”. In: *Transactions of the American Mathematical Society* 44 (1938), pp. 522–536 (cit. on p. 5).
- [SDD21] Antoine Salmona, Julie Delon, and Agnès Desolneux. *Gromov–Wasserstein distances between Gaussian distributions*. 2021. arXiv: [2104.07970](#) (cit. on p. 8).
- [Stu06] Karl-Theodor Sturm. “On the geometry of metric measure spaces”. In: *Acta mathematica* 196.1 (2006), pp. 65–131 (cit. on p. 2).
- [Stu12] Karl-Theodor Sturm. *The space of spaces: curvature bounds and gradient flows on the space of metric measure spaces*. 2012. arXiv: [1208.0434](#) (cit. on pp. 2–4, 8).
- [SVP21] Thibault Séjourné, François-Xavier Vialard, and Gabriel Peyré. “The unbalanced Gromov Wasserstein distance: Conic formulation and relaxation”. In: *Advances in Neural Information Processing Systems* 34 (2021), pp. 8766–8779 (cit. on pp. 2, 4).
- [Ume88] Shinji Umeyama. “An eigendecomposition approach to weighted graph matching problems”. In: *IEEE transactions on pattern analysis and machine intelligence* 10.5 (1988), pp. 695–703 (cit. on p. 2).
- [Van00] Aad W Van der Vaart. *Asymptotic statistics*. Vol. 3. Cambridge university press, 2000 (cit. on p. 34).
- [Vay+19] Titouan Vayer, Rémi Flamary, Nicolas Courty, Romain Tavenard, and Laetitia Chapel. “Sliced Gromov–Wasserstein”. In: *Advances in Neural Information Processing Systems* 32 (2019) (cit. on pp. 2, 8, 20).
- [Vay20] Titouan Vayer. “A contribution to optimal transport on incomparable spaces”. PhD thesis. 2020. arXiv: [2011.04447](#) (cit. on pp. 3, 4, 7, 8, 15, 16).
- [Vil08] Cédric Villani. *Optimal transport: old and new*. Vol. 338. Springer, 2008 (cit. on pp. 6, 7, 33).
- [Xu+20] Hongteng Xu, Dixin Luo, Ricardo Henao, Svati Shah, and Lawrence Carin. *Learning autoencoders with relational regularization*. 2020. arXiv: [2002.02913](#) (cit. on p. 2).
- [ZD15] Feng Zhou and Fernando De la Torre. “Factorized graph matching”. In: *IEEE transactions on pattern analysis and machine intelligence* 38.9 (2015), pp. 1774–1789 (cit. on p. 2).
- [Zha+22] Zhengxin Zhang, Youssef Mroueh, Ziv Goldfeld, and Bharath Sriperumbudur. “Cycle consistent probability divergences across different spaces”. In: *International Conference on Artificial Intelligence and Statistics*. PMLR. 2022, pp. 7257–7285 (cit. on p. 7).

## APPENDIX A. PROOFS OF LEMMAS 3.3 AND 3.4

In this section, we prove [Lemmas 3.3](#) and [3.4](#). Let us first recall the statement of [Lemma 3.3](#):

**Lemma 3.3.** *Let  $\mu, \nu \in \mathcal{P}(E)$  and let  $\psi_1, \psi_2 : E \rightarrow F$  be homeomorphisms. Let  $\tilde{c} : F \times F \rightarrow \mathbb{R}$  and consider the cost  $c(x, y) = c(\psi_1(x), \psi_2(y))$ . Then a map is optimal for the cost  $c$  between  $\mu$  and  $\nu$  if and only if it is of the form  $\psi_2^{-1} \circ T \circ \psi_1$  with  $T : F \times F$  optimal for the cost  $\tilde{c}$  between  $\psi_{1\#}\mu$  and  $\psi_{2\#}\nu$ .*

*Proof.* Remark that the continuity of  $\psi_1, \psi_2$  and their inverse ensures their measurability. We have the following equalities:

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

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  $\psi_1$  and  $\psi_2$ . 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  $\pi^*$  between  $\mu$  and  $\nu$ , and *vice versa*. Writing  $\tilde{\pi}^* = (\text{id}, T)_{\#}(\psi_{1\#}\mu)$ , this plan  $\pi^*$  is given by

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

We now recall the statement of [Lemma 3.4](#):

**Lemma 3.4.** *Let  $h \geq 1$  and  $\mu, \nu \in \mathcal{P}(\mathbb{R}^h)$  with compact supports and such that  $\mu \ll \text{Leb}_h$ . 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  $\mu$  and  $\nu$  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  $f : \mathbb{R}^h \rightarrow \mathbb{R}$  convex.*

*Proof.* As  $\psi_{1\#}\mu$  has a density with respect to the Lebesgue measure since  $\psi_1$  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  $\nabla f$ , where  $\varphi$  is a convex function. Using [Lemma 3.3](#) then yields the result.  $\square$

**Remark A.1** (Hypothesis of [Lemma 3.4](#)). In the proof of [Lemma 3.4](#), we only needed (i)  $\psi_1, \psi_2$  and their inverse to be measurable, (ii)  $\psi_{1\#}\mu$  to have a density with respect to the Lebesgue measure, and (iii)  $\psi_{1\#}\mu$  and  $\psi_{2\#}\nu$  to have compact support. Imposing  $\psi_1$  to be a diffeomorphism and  $\psi_2$  to be a homeomorphism ensures both (i) and (ii) and is natural to expect.

## APPENDIX B. MEASURABLE SELECTION OF MAPS IN THE MANIFOLD SETTING

**B.1. Measurability of set-valued maps.** Let  $X$  and  $U$  be two topological spaces, and let  $\mathcal{B}$  denote the Borel  $\sigma$ -algebra on  $X$ . A set-valued map  $S$  is a map from  $X$  to  $P(U)$  (the set of subsets of  $U$ ) and will be denoted by  $S : X \rightrightarrows U$ . The idea is to introduce a notation that is 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 Rockafellar and Wets [[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  $M$ , which we shall assume in the following. For the sake of completeness, we provide all the proofs and highlight those that require specific care when replacing  $\mathbb{R}^n$  by a manifold. Of importance for our proofs, we define the *pre-image* of a set  $B \subset U$  by  $S$ ,

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



and the *domain* of  $S$ , defined as  $S^{-1}(U) = \{x \in X \mid S(x) \neq \emptyset\}$ . We will often use the following relation: if a set  $A$  can be written as  $A = \bigcup_k A_k$ , then  $S^{-1}(A) = \bigcup_k S^{-1}(A_k)$ . Indeed,

$$x \in S^{-1}(A) \iff S(x) \cap A \neq \emptyset \iff \exists k, S(x) \cap A_k \neq \emptyset \iff \exists k, x \in S^{-1}(A_k) \iff x \in \bigcup_k S^{-1}(A_k).$$

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

$$S^{-1}(O) \in \mathcal{B}.$$

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 B.1** ([RW09, Theorem 14.3.c]). *A closed-valued map  $S : X \rightrightarrows U$  is measurable if and only if the property “ $S^{-1}(B) \in \mathcal{B}$ ” holds either:*

- (a) *for all  $B \subset U$  open (the definition);*
- (b) *for all  $B \subset U$  compact;*
- (c) *for all  $B \subset U$  closed.*

*Proof.*

- (a)  $\implies$  (b): For a compact  $B \subset U$ , let  $B_k = \{x \in U \mid d(x, B) < k^{-1}\}$  for  $k \geq 0$  (which is open). Note that

$$x \in S^{-1}(B) \iff S(x) \cap B \neq \emptyset \iff S(x) \cap B_k \neq \emptyset \text{ for all } k,$$

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, and so is  $S^{-1}(B)$  as a countable intersection of measurable sets.

- (b)  $\implies$  (a): Fix  $O$  an open set of  $U$ . As we assumed  $U$  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)  $\implies$  (c): Immediate.
- (c)  $\implies$  (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  $\mathcal{B}$ .  $\square$

Now, we state a result on operations that preserve the 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 B.2** ([RW09, Proposition 14.11], 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

$$P^{-1}(A \times B) = \{x \mid S_1(x) \times S_2(x) \in A \times B\} = \{x \mid S_1(x) \in A \text{ and } S_2(x) \in B\} = S_1^{-1}(A) \cap S_2^{-1}(B),$$

so that finally,  $P^{-1}(O') = \bigcup_n S_1^{-1}(O_1^{(n)}) \cap S_2^{-1}(O_2^{(n)})$  is measurable as a countable union of (finite) intersection of measurable sets (given that  $S_1$  and  $S_2$  are measurable). Note that this does not require  $S_1$  and  $S_2$  to be closed-valued.

Now, let us focus on the second point, which requires more attention. Thanks to [Proposition B.1](#), it is sufficient to show that  $Q^{-1}(C) \in \mathcal{B}$  for any compact set  $C \subset U$ . In [RW09], this is done by writing

$$\begin{aligned} Q^{-1}(C) &= \{x \mid S_1(x) \cap S_2(x) \cap C \neq \emptyset\} = \{x \mid R_1(x) \cap R_2(x) \neq \emptyset\} \\ &= \{x \mid 0 \in (R_1(x) - R_2(x))\} = (R_1 - R_2)^{-1}(0), \end{aligned}$$

where  $R_j(x) = S_j(x) \cap C$  (which is also closed-valued), and using the fact that the (Minkowski) difference of measurable closed-valued maps is measurable as well [RW09, Proposition 14.11.c]. To

adapt this idea (we cannot consider Minkowski differences in our setting), we introduce the diagonal  $\Delta = \{(u, u) \mid u \in U\} \subset U \times U$ . Now, observe that

$$R_1(x) \cap R_2(x) \neq \emptyset \iff (R_1(x) \times R_2(x)) \cap \Delta \neq \emptyset,$$

that is  $x \in R^{-1}(\Delta)$ , where  $R(x) = R_1(x) \times R_2(x)$ . 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  $\Delta$  is closed,  $R^{-1}(\Delta) = Q^{-1}(C)$  is measurable.  $\square$

**B.2. Proof of Proposition 2.10.** Let us first recall the result.

**Proposition 2.10** (Measurable selection of maps, manifold case). *Let  $M$  be a complete Riemannian manifold and  $(B, \Sigma, m)$  a measure space. Consider a measurable function  $B \ni u \mapsto (\mu_u, \nu_u) \in P(M)^2$ . Assume that for  $m$ -almost every  $u \in B$ ,  $\mu_u \ll \text{vol}_M$  and  $\mu_u$  and  $\nu_u$  have a finite transport cost. Let  $T_u$  denote the (unique by Theorem 1.8) optimal transport map induced by the squared distance cost  $d_M^2$  on  $M$  between  $\mu_u$  and  $\nu_u$ . Then there exists a measurable function  $(u, x) \mapsto T(u, x)$  such that for  $m$ -almost every  $u$ ,  $T(u, x) = T_u(x)$ ,  $\mu_u$ -a.e.*

The proof is essentially an adaptation of the one of Theorem 1.1 from Fontbona et al. [FGM10], with additional care required since we do not have access to a linear structure on the manifold  $M$ . It relies on the measurability of set-valued maps (see [RW09, Chapters 5 and 14] and Appendix B.1 for a summary). In the following, we consider a partition  $(A_{nk})_n$  of  $M$  made of cells with diameter  $\leq 2^{-k}$  and such that  $(A_{n'j})_n$  is a refinement of  $(A_{nk})_n$  in sense that each  $A_{nk}$  is itself partitioned by some  $A_{n'j}$ . The crucial point regarding measurability is the following proposition.

**Proposition B.3.** *The set  $B_{nk} = \{(u, x) \mid T_u(x) \in A_{nk}\}$  is measurable.*

Its proof relies on a core lemma:

**Lemma B.4.** *Let  $F \subset M$  be a closed set. Then the set  $B_F = \{(u, x) \mid T_u(x) \in F\}$  is measurable.*

The key will be to identify this set as the domain of a measurable set-valued map, see Appendix B.1.

*Proof of Lemma B.4.* Observe that

$$B_F = \{(u, x) \mid (\{x\} \times F) \cap \text{gph}(T_u) \neq \emptyset\},$$

where  $\text{gph}(T_u) := \{(x, T_u(x)) \mid x \in M\}$  denotes the topological closure of the graph of the optimal transport map  $T_u$  that pushes  $\mu_u$  onto  $\nu_u$ . Let  $S_1 : (u, x) \mapsto \{x\} \times F$  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 B.2, 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. Establishing the measurability of these two maps can be easily done by adapting the work of [FGM10]. We give details here for the sake of completeness.

- *Measurability of  $S_1$ :* Let  $O \subset M \times M$  be open. Observe that

$$S_1^{-1}(O) = \{(u, x) \mid \{x\} \times F \cap O \neq \emptyset\} = B \times \{x \mid \{x\} \times F \cap O \neq \emptyset\}.$$

Fix  $z \in F$ . There exists  $\varepsilon > 0$  such that  $B(x, \varepsilon) \times \{z\} \subset O$  (since  $O$  is open), and thus  $B(x, \varepsilon) \times F \cap O \neq \emptyset$ , proving that there is a neighborhood of  $x$  included in  $\{x \mid \{x\} \times F \cap O \neq \emptyset\}$  which is thus open, making in turn  $S_1^{-1}(O)$  open (hence measurable), hence 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  $\mu$  and  $\nu$  for the squared distance cost on  $M$ ) is measurable, and

(ii) the map  $\mathfrak{U} : \pi \in P(M^2) \mapsto \text{supp } \pi$  satisfies that  $\mathfrak{U}^{-1}(O)$  is open for any open set  $O \subset P(M^2)$ .

From these two points, we obtain the measurability of  $(U \circ S)^{-1}(O)$ , 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 B.1. Let  $C \subset P(M^2)$  be closed. Then,  $S^{-1}(C) = \{(\mu, \nu) \mid \Pi^*(\mu, \nu) \cap C \neq \emptyset\}$ . Consider a sequence  $(\mu_n, \nu_n)_n$  in  $S^{-1}(C)$  that converges to  $(\mu, \nu)$  for the weak topology. Let  $\pi_n \in \Pi^*(\mu_n, \nu_n) \cap C$ . According to [Vil08, Theorem 5.20],  $(\pi_n)_n$

admits a weak limit  $\pi$  in  $\Pi^*(\mu, \nu)$ ; but since  $C$  is closed,  $\pi \in C$ , so  $(\mu, \nu) \in S^{-1}(C)$ , which is closed (hence measurable), proving the measurability of  $S$ .

(ii) simply follows from the fact that

$$\mathfrak{U}^{-1}(O) = \{\pi \mid \text{supp } \pi \cap O \neq \emptyset\} = \{\pi \mid \pi(O) > 0\},$$

which is open. Indeed, the Portmanteau theorem shows 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  $\mathfrak{U}^{-1}(O)$  is closed, that is  $\mathfrak{U}^{-1}(O)$  is open.  $\square$

*Proof of Proposition B.3.* Using that  $A_{nk}$  can be inner-approximated by a sequence of closed set  $F_j \subset A_{nk}$ , we obtain a sequence of sets  $(B_{F_j})_j$  such that  $\bigcup_j B_{F_j} = A_{nk}$ . By Lemma B.4, the  $(B_{F_j})_j$  are measurable, so is  $A_{nk}$  as the (countable) union of measurable sets.  $\square$

We can now prove our main proposition on measurable selections of maps.

*Proof of Proposition 2.10.* Recall that we assume that  $M = \bigsqcup_n A_{nk}$ . For each  $(n, k)$ , select (in a measurable way) an element  $a_{nk} \in A_{nk}$ . Then, define the map

$$T^{(k)} : (u, x) \mapsto a_{nk}, \quad \text{where } T_u(x) \in A_{nk}.$$

This map is measurable: for any open  $O \subset M$ ,

$$T^{(k), -1}(O) = \{(u, x) \mid T_u(x) \in A_{nk}\} = B_{nk},$$

which is measurable by Proposition B.3. Now, for two maps  $f, g : B \times M \rightarrow M$ , let  $D_1$  denotes the natural  $L^1$  distance on  $M$ , that is

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

This yields a complete metric space [Chi07], and  $(T^{(k)})_k$  is a Cauchy sequence for this distance. Indeed, for  $k \leq j$  two integers, recall that we assumed that  $(A_{nj})_n$  is a refinement of  $(A_{nk})_n$ , yielding

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

where we successively used (i) that for all  $u$ ,  $\int_M \mathbb{1}_{B_{n'j}}(u, x) \, d\mu_u(x) = \nu_u(A_{n'j})$  by construction, (ii) that the diameter of the partition  $A_{nk}$  is less than or equal to  $2^{-k}$  and (iii) that  $\nu_u$  and  $m$  are probability measures. Note that claim (i) is given by

$$(u, x) \in B_{n'j} \iff T_u(x) \in A_{n'j} \iff x \in \mu_u(T_u^{-1}(A_{n'j})) = T_{u\#}\mu_u(A_{n'j}) = \nu_u(A_{n'j}).$$

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

$$\int_M g(x) f(T(x, u)) \, d\mu_u(x) = \int_M g(x) f(T_u(x)) \, d\mu_u(x),$$

for any pair  $f, g : M \rightarrow \mathbb{R}$  of bounded Lipschitz-continuous functions [Van00, 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  $(A_{nk})_n$ . We have for  $m$ -a.e.  $u$ :

$$\left| \int_M g(x) f(T_u(x)) \, d\mu_u(x) - \int_M g(x) f(T(u, x)) \, d\mu_u(x) \right|$$

$$\leq \left| \int_M g(x)f(T_u(x)) d\mu_u(x) - \int_M g(x)f(T^{(k)}(u,x)) d\mu_u(x) \right| + \|g\| \|f\| \int_M d(T^{(k)}(u,x), T(u,x)) d\mu_u(x).$$

Since  $T^{(k)} \rightarrow T$  in  $D_1$ , it implies that up to a subsequence, for  $m$ -a.e.  $u$ , the second term tends to zero:

$$\int_M d(T^{(k)}(u,x), T(u,x)) d\mu_u(x) \xrightarrow{k \rightarrow \infty} 0.$$

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

$$\begin{aligned} \left| \int_M g(x)f(T_u(x)) d\mu_u(x) - \int_M g(x)f(T^{(k)}(u,x)) d\mu_u(x) \right| &\leq \int_M |g(x)| |f(T_u(x)) - f(T^{(k)}(u,x))| d\mu_u(x) \\ &\leq \|g\| \|f\| \int_M d(T_u(x), T^{(k)}(u,x)) d\mu_u(x) \\ &\leq \|g\| \|f\| 2^{-k} \sum_n \nu_u(A_{nk}) \xrightarrow{k \rightarrow \infty} 0, \end{aligned}$$

which concludes the proof.  $\square$

### APPENDIX C. MEASURE DISINTEGRATION

The following definition and theorem are taken from Ambrosio et al. [AGS05].

**Definition C.1** (Measure disintegration). Let  $\mathcal{X}$  and  $\mathcal{Z}$  be two Radon spaces,  $\mu \in \mathcal{P}(\mathcal{X})$  and  $\varphi : \mathcal{X} \rightarrow \mathcal{Z}$  a Borel-measurable function. A family of probability measures  $\{\mu_u\}_{u \in \mathcal{Z}} \subset \mathcal{P}(\mathcal{X})$  is a *disintegration* of  $\mu$  by  $\varphi$  if:

- (i) the function  $u \mapsto \mu_u$  is Borel-measurable;
- (ii)  $\mu_u$  lives on the fiber  $\varphi^{-1}(u)$ , that is for  $\varphi_{\#}\mu$ -a.e.  $u \in \mathcal{Z}$ ,

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

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

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

$$\int_{\mathcal{X}} f(x) d\mu(x) = \int_{\mathcal{Z}} \left( \int_{\varphi^{-1}(u)} f(x) d\mu_u(x) \right) d(\varphi_{\#}\mu)(u).$$

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

$$\mu(B) = \int_{\mathcal{Z}} \mu_u(B) d(\varphi_{\#}\mu)(u).$$

**Theorem C.2** (Disintegration theorem). Let  $\mathcal{X}$  and  $\mathcal{Z}$  be two Radon spaces,  $\mu \in \mathcal{P}(\mathcal{X})$  and  $\varphi : \mathcal{X} \rightarrow \mathcal{Z}$  a Borel-measurable function. There exists a  $\varphi_{\#}\mu$ -a.e. uniquely determined family of probability measures  $\{\mu_u\}_{u \in \mathcal{Z}} \subset \mathcal{P}(\mathcal{X})$  that provides a disintegration of  $\mu$  by  $\varphi$ .