Existence of Monge maps for the Gromov–Wasserstein distance

Stage de fin d'études – MVA & Mines Paris

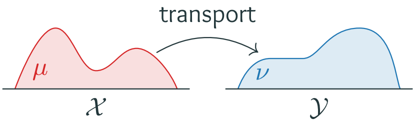
Théo Dumont

October 20th 2022

Laboratoire d'Informatique Gaspard Monge (LIGM)

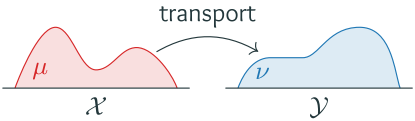
Supervisors: François-Xavier Vialard, Théo Lacombe (LIGM)

1. Introduction



Setup :

- $\mathcal{X}, \mathcal{Y}$ Polish spaces
- probability measures $\mu \in \mathcal{P}(\mathcal{X})$ and $\nu \in \mathcal{P}(\mathcal{Y})$
- cost function $c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$



Monge problem [Monge, 1781]

We consider the following minimization problem:

$$\min_{T_\# \mu = \nu} \int_{\mathcal{X}} c(x, T(x)) \, \mathrm{d}\mu(x) \, .$$

where $T_\# \mu(B) = \mu(T^{-1}(B))$ for all Borel $B \subset \mathcal{Y}$.

- solutions T are called *Monge maps* or *optimal maps*

Monge problem [Monge, 1781]

We consider the following minimization problem:

$$\min_{T_{\#}\mu=\nu} \int_{\mathcal{X}} c(x, T(x))\, \mathrm{d}\mu(x)\,.$$

- problems:
  1. constraints are not linear!
  2. minimum not reached
  3. no maps s.t. $T_{\#}\mu = \nu$

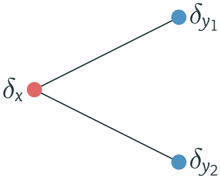
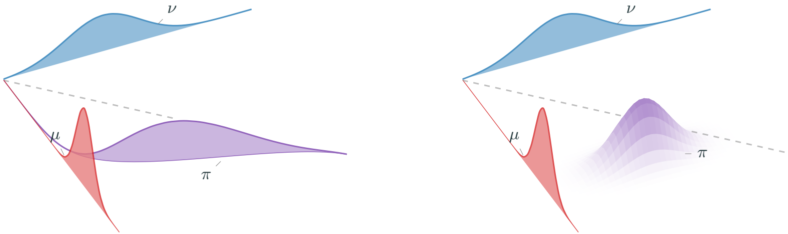


Figure 1: We need to "split" δ_x: no map can do this.

Transport plan [Kantorovich, 1942]

A *transport plan* between μ and ν is a (probability) measure $\pi \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ of marginals μ and ν:

$$\Pi(\mu, \nu) \triangleq \left\{ \pi \in \mathcal{P}(\mathcal{X} \times \mathcal{Y}) \mid P^1_{\#}\pi = \mu, P^2_{\#}\pi = \nu \right\} .$$



π is induced by a map T: $\pi = (\mathrm{id}, T)_{\#}\mu$. π is the product plan $\pi = \mu \otimes \nu$.

Figure 2: Deterministic or non-deterministic transport plans.

Kantorovich problem

We consider the following minimization problem:

$$\min_{\pi \in \Pi(\mu,\nu)} \int_{\mathcal{X}\times\mathcal{Y}} c(x,y)\,\mathrm{d}\pi(x,y)\,. \tag{KP}$$

- always non-empty (contains $\mu \otimes \nu$), existence of a minimizer
- *linear program in π!*
- if $c(x,y) = |x-y|_p^p$, p-Wasserstein distance $W_p(\mu,\nu)^p$

Question

Is the relaxation *tight*? **under some assumptions, yes!**

1. Introduction

1.1. Map solutions of OT

Brenier's theorem [Brenier, 1987]

Let $\mu, \nu \in \mathcal{P}(\mathbb{R}^n)$ and $c(x,y) = |x-y|^2$. If $\mu \ll \mathcal{L}^n$, then there exists a unique solution to (KP) and it is induced by a **map** $T = \nabla f$, with f convex.

- generalize for manifolds $\mathcal{X}$ and $\mathcal{Y}$ and for other cost functions c

Twist condition [Villani, 2008, McCann and Guillen, 2011]

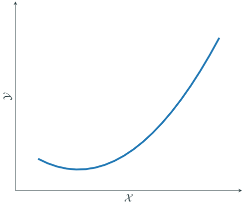
We say that c satisfies the **twist condition** if

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

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

$$\pi^\star = (\mathrm{id}, c\text{-}\exp_x(\nabla f))_{\#}\mu\,.$$

- c-$\exp_x(p)$ is the unique y such that $\nabla_x c(x,y) + p = 0$:
$$c\text{-}\exp_x(p) = (\nabla_x c)^{-1}(x, -p)\,.$$
- usual Riemannian exp when $c(x,y) = d(x,y)^2/2$



Twist condition [Villani, 2008, McCann and Guillen, 2011]

We say that c satisfies the **twist condition** if

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

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

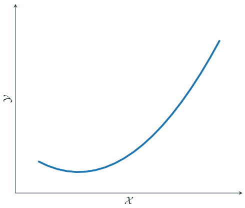
$$\pi^\star = (\mathrm{id}, c\text{-}\exp_x(\nabla f))_{\#}\mu \,.$$

- examples:

		twist
$\lvert x - y\rvert^2$	in $\mathbb{R}^n$	✓
$\langle x, y\rangle$	in $\mathbb{R}^n$	✓
$\langle x, y\rangle$	on S^{n-1}	·

- other formulation:

$$\forall y_1 \neq y_2, \quad x \mapsto c(x, y_1) - c(x, y_2) \text{ has no critical point.}$$



Subtwist condition [Ahmad et al., 2011, Chiappori et al., 2010]

We say that c satisfies the **subtwist condition** if

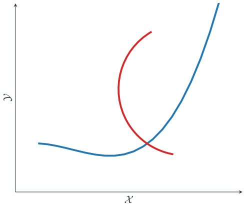
$$\forall y_1 \neq y_2, \quad x \mapsto c(x,y_1) - c(x,y_2) \quad \text{has at most 2 critical points.} \tag{Subtwist}$$

Suppose that c satisfies (Subtwist). Under the *same assumptions than before*, (KP) admits a unique optimal transport plan $\pi^\star$ and it is induced by the **union of a map and an anti-map:**

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

for $G : \mathcal{X} \to \mathcal{Y}$, $H : \mathcal{Y} \to \mathcal{X}$ and $0 \leq \bar{\mu} \leq \mu$ s.t. $\nu - G_\# \bar{\mu}$ vanishes on the range of G.

		twist	subtwist
$\langle x, y \rangle$	on S^{n-1}	$\cdot$	✓



m-twist condition [Moameni, 2016]

We say that c satisfies a ***m*-twist condition** if

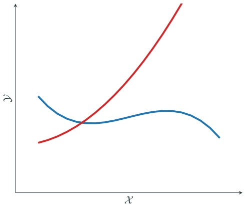
$$\forall x_0 \in \mathcal{X}, y_0 \in \mathcal{Y}, \quad \mathrm{card}\,\{y \mid \nabla_x c(x_0, y) = \nabla_x c\,(x_0, y_0)\} \leq m\,. \tag{m-twist}$$

Suppose that c satisfies (m-twist) and is *bounded*. Under the *same assumptions than before*, each optimal plan $\pi^\star$ of (KP) is supported on the **graphs of** $k \leq m$ **measurable maps** $T_i : \mathcal{X} \to \mathcal{Y}$:

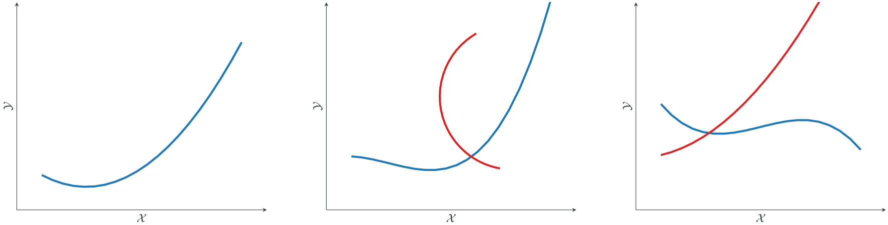
$$\pi^\star = \sum_{i=1}^{k} \alpha_i\,(\mathrm{id}, T_i)_\#\,\mu\,,$$

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

		twist	subtwist	2-twist
$1 - \cos(x - y)$	on $[0, 2\pi)$	·	✓	✓
our cost!	in $\mathbb{R}^n$	·	·	✓



Twist	*Subwist*	*2-twist*
$\Downarrow$	$\Downarrow$	$\Downarrow$
map	map/anti-map	bimap



- all assumptions needed to apply them are satisfied when *μ and ν have compact support* and *μ has a density*

1. Introduction

1.2. Gromov–Wasserstein

sup **formulation**	L^p **formulation**
Set couplings $\mathcal{R}(A,B)$ $\left\{R \subset A \times B \mid P^1(R) = A, P^2(R) = B\right\}$	Transport plans $\Pi(\mu,\nu)$ $\left\{\pi \in \mathcal{P}(\mathcal{X} \times \mathcal{Y}) \mid P^1_{\#}\pi = \mu, P^2_{\#}\pi = \nu\right\}$
Hausdorff distance $\mathrm{H}_{\mathcal{Z}}$ between *sets* A, B $\inf_{R \in \mathcal{R}(A,B)} \left(\sup_{(a,b)\in R} d_{\mathcal{Z}}(a,b)\right)$	**Wasserstein** distance W_p between *measures* μ, ν $\inf_{\pi \in \Pi(\mu,\nu)} \left(\int \lvert x-y \rvert^p \, \mathrm{d}\pi\right)^{1/p}$

Gromov–Wasserstein [Mémoli, 2011]

sup **formulation**	L^p **formulation**
Set couplings $\mathcal{R}(A, B)$ $\left\{R \subset A \times B \mid P^1(R) = A, P^2(R) = B\right\}$	Transport plans $\Pi(\mu, \nu)$ $\left\{\pi \in \mathcal{P}(\mathcal{X} \times \mathcal{Y}) \mid P^1_\# \pi = \mu, P^2_\# \pi = \nu\right\}$
Hausdorff distance $\mathrm{H}_{\mathcal{Z}}$ between *sets* A, B $\inf_{R \in \mathcal{R}(A,B)} \left(\sup_{(a,b) \in R} d_{\mathcal{Z}}(a, b) \right)$	**Wasserstein** distance W_p between *measures* μ, ν $\inf_{\pi \in \Pi(\mu,\nu)} \left(\int \lvert x - y \rvert^p \, \mathrm{d}\pi \right)^{1/p}$
Gromov–Hausdorff distance GH between *metric spaces* $\mathcal{X}, \mathcal{Y}$ $\frac{1}{2} \inf_{R \in \mathcal{R}} \left(\sup_{(x,y),(x',y') \in R} \lvert d_{\mathcal{X}}(x, x') - d_{\mathcal{Y}}(y, y') \rvert \right)$	

Gromov–Wasserstein [Mémoli, 2011]

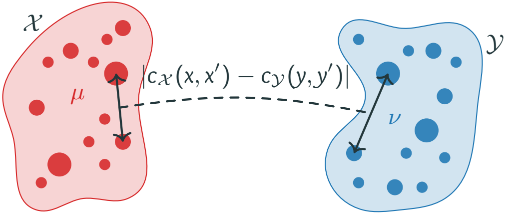
sup **formulation**	L^p **formulation**
Set couplings $\mathcal{R}(A, B)$ $\left\{R \subset A \times B \mid P^1(R) = A, P^2(R) = B\right\}$	Transport plans $\Pi(\mu, \nu)$ $\left\{\pi \in \mathcal{P}(\mathcal{X} \times \mathcal{Y}) \mid P^1_\# \pi = \mu, P^2_\# \pi = \nu\right\}$
Hausdorff distance $H_{\mathcal{Z}}$ between *sets* A, B $\inf_{R \in \mathcal{R}(A,B)} \left(\sup_{(a,b) \in R} d_{\mathcal{Z}}(a, b) \right)$	**Wasserstein** distance W_p between *measures* μ, ν $\inf_{\pi \in \Pi(\mu,\nu)} \left(\int \lvert x - y \rvert^p \, \mathrm{d}\pi \right)^{1/p}$
Gromov–Hausdorff distance GH between *metric spaces* $\mathcal{X}, \mathcal{Y}$ $\frac{1}{2} \inf_{R \in \mathcal{R}} \left(\sup_{(x,y),(x',y') \in R} \lvert d_{\mathcal{X}}(x, x') - d_{\mathcal{Y}}(y, y') \rvert \right)$	**Gromov–Wasserstein** distance GW_p between *measure metric spaces* $\mathbb{X}, \mathbb{Y}$ $\inf_{\pi \in \Pi} \left(\int \lvert d_{\mathcal{X}}(x, x') - d_{\mathcal{Y}}(y, y') \rvert^p \, \mathrm{d}\pi \otimes \pi \right)^{1/p}$

- match measure metric spaces $(\mathcal{X}, d, \mu)$ (*e.g.* point clouds) up to isometry: no notion of transport here, but rather of *correspondence*
- applications in vision, biology...

Gromov–Wasserstein problem

We consider the following quadratic minimization problem:

$$\inf_{\pi \in \Pi(\mu,\nu)} \iint_{\mathcal{X}\times\mathcal{Y}} |c_{\mathcal{X}}(x,x') - c_{\mathcal{Y}}(y,y')|^p \, \mathrm{d}\pi(x,y)\, \mathrm{d}\pi(x',y') \,. \tag{GW}$$



- distance between mm-spaces up to isometry, *i.e.* $\mathrm{GW}(\mathbb{X}, \mathbb{Y}) = 0$ iff $\mathbb{X} = (\mathcal{X}, d_{\mathcal{X}}^q, \mu)$ and $\mathbb{Y} = (\mathcal{Y}, d_{\mathcal{Y}}^q, \nu)$ are *strongly isomorphic* [Mémoli, 2011]

Gromov–Wasserstein problem

We consider the following quadratic minimization problem:

$$\inf_{\pi \in \Pi(\mu,\nu)} \iint_{\mathcal{X}\times\mathcal{Y}} |c_{\mathcal{X}}(x,x') - c_{\mathcal{Y}}(y,y')|^p \, \mathrm{d}\pi(x,y)\, \mathrm{d}\pi(x',y') \,. \tag{GW}$$

- quadratic in π + non-convex $\implies$ much harder than OT
- for $p = 2$, discrete formulation: with $D^{\mathcal{X}}$, $D^{\mathcal{Y}}$ two similarity matrices on $\mathcal{X}$ and $\mathcal{Y}$,

  $$\min_{P \in U(a,b)} \sum_{i,j,i',j'} |D^{\mathcal{X}}_{i,i'} - D^{\mathcal{Y}}_{j,j'}|^p P_{i,j} P_{i',j'} \,.$$

  particular case of the *quadratic assignment problem* (QAP), NP-hard
- bonus: compare measures living in incomparable spaces

Question

What can be said on the existence of Monge maps for the Gromov–Wasserstein distance?

- quadratic...

1. Introduction

1.3. Existing results and contributions

Context

Let $n \geq d$. We consider the GW problem for $\mu, \nu \in \mathbb{R}^n \times \mathbb{R}^d$ in 2 different settings:

1. the *inner product case*, where $c_{\mathcal{X}} = c_{\mathcal{Y}} = \langle \cdot, \cdot \rangle$:

$$\min_{\pi \in \Pi(\mu,\nu)} \iint_{\mathcal{X} \times \mathcal{Y}} \left| \langle x, x' \rangle - \langle y, y' \rangle \right|^2 \mathrm{d}\pi(x,y)\, \mathrm{d}\pi(x',y') \,, \qquad \text{(GW inner prod)}$$

   - *e.g.* on a d-dimensional sphere S^{d-1}

2. the *quadratic case*, where $c_{\mathcal{X}} = c_{\mathcal{Y}} = |\cdot|^2$:

$$\min_{\pi \in \Pi(\mu,\nu)} \iint_{\mathcal{X} \times \mathcal{Y}} \left| |x - x'|^2 - |y - y'|^2 \right|^2 \mathrm{d}\pi(x,y)\, \mathrm{d}\pi(x',y') \,, \qquad \text{(GW quadratic)}$$

   - standard choice for $c_{\mathcal{X}}$ and $c_{\mathcal{Y}}$

$\rightarrow$ both studied in the literature [Alvarez-Melis et al., 2019, Vayer, 2020]

In the following, $n = d$.

Existing results

1. the *inner product case*, where $c_{\mathcal{X}} = c_{\mathcal{Y}} = \langle \cdot, \cdot \rangle$:

[Vayer, 2020]

Let $\mu, \nu \in \mathcal{P}(\mathbb{R}^n)$ of finite second order moment with $\mu \ll \mathcal{L}^n$. Suppose that there exists a solution $\pi^\star$ such that $M^\star = \int y \otimes x \, \mathrm{d}\pi^\star(x, y)$ is of *full rank*. Then there exists an optimal map $T = \nabla f \circ M^\star$ with $f : \mathbb{R}^n \to \mathbb{R}$ convex.

2. the *quadratic case*, where $c_{\mathcal{X}} = c_{\mathcal{Y}} = |\cdot|^2$:

[Sturm, 2020]

Let $\mu, \nu \in \mathcal{P}(\mathbb{R}^n)$ with *density*, *rotationally invariant* around their barycenter. Then optimal transport plans are *induced by a map* which is the monotone increasing rearrangement between the radial distributions of μ and ν.

[Vayer, 2020]

Let $\mu, \nu \in \mathcal{P}(\mathbb{R}^n)$ with compact support. Assume that $\mu \ll \mathcal{L}^n$ and that both μ and ν are centered. Suppose that there exists $\pi^\star$ such that $M^\star = \int y \otimes x \, \mathrm{d}\pi^\star(x, y)$ is of *full rank* and that there exists a differentiable convex $F : \mathbb{R} \to \mathbb{R}$ such that $|T(x)|_2^2 = F'(|x|_2^2)$, then there exists an optimal map $T = \nabla f \circ M^\star$ with f convex.

Contributions

Let $\mu, \nu \in \mathcal{P}(\mathbb{R}^n)$ of *compact support*. Suppose $\mu \ll \mathcal{L}^n$.

1. **Theorem:** The (GW inner prod) problem admits a *map* as a solution.
2. **Theorem:** The (GW quadratic) problem either admits a *map* or a *bimap* as a solution.
3. **Conjecture:** The second claim is *tight*: there exists cases where optimal solutions of (GW quadratic) are *not maps*.

Optimality of the monotone rearrangements $\pi^{\oplus}_{\text{mon}}$ and $\pi^{\ominus}_{\text{mon}}$ in 1D for (GW quadratic):

4. **Algorithm:** There exists measures μ and ν for which the $\pi^{\oplus}_{\text{mon}}$ and $\pi^{\ominus}_{\text{mon}}$ are *not optimal*;
   and having $\pi^{\oplus}_{\text{mon}}$ or $\pi^{\ominus}_{\text{mon}}$ as optimal is *not stable* by perturbations of μ and ν.
5. (**Theorem:** When measures μ and ν are composed of two distant parts, $\pi^{\oplus}_{\text{mon}}$ or $\pi^{\ominus}_{\text{mon}}$ is *optimal*.)

2. Monge maps for GW

Trick: relaxation of GW into OT problem

- (GW) $= \min_\pi F(\pi, \pi)$ with F symmetric bilinear
- *first-order condition*: $\pi^\star$ minimizes (GW) $\implies$ minimizes $\pi \mapsto 2F(\pi, \pi^\star)$:

$$\min_{\pi \in \Pi(\mu,\nu)} \int C_{\pi^\star}(x,y)\,\mathrm{d}\pi(x,y), \qquad \text{with } C_{\pi^\star}(x,y) = \int |c_{\mathcal{X}}(x,x') - c_{\mathcal{Y}}(y,y')|^p \,\mathrm{d}\pi^\star(x',y')$$

- converse implication? [Séjourné et al., 2021]:

Tightness

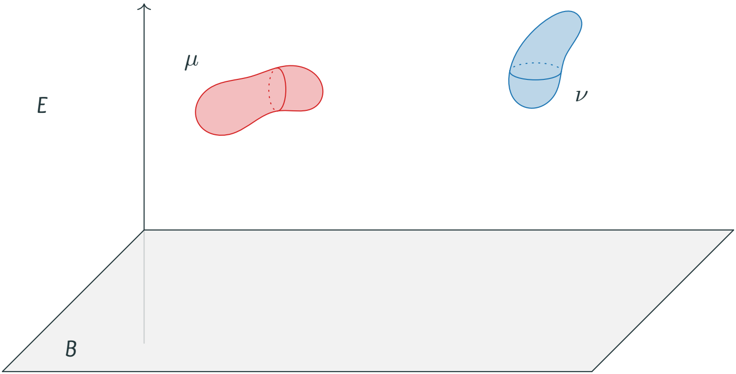
If $\iint |c_{\mathcal{X}}(x,x') - c_{\mathcal{Y}}(y,y')|^p \,\mathrm{d}\alpha \otimes \alpha \leq 0$ for all (signed) measures $\alpha \in \mathcal{M}(\mathcal{X} \times \mathcal{Y})$ with **null marginals**, then the relaxation of GW_2^2 *is tight*.

- twist conditions for our linearized costs?...

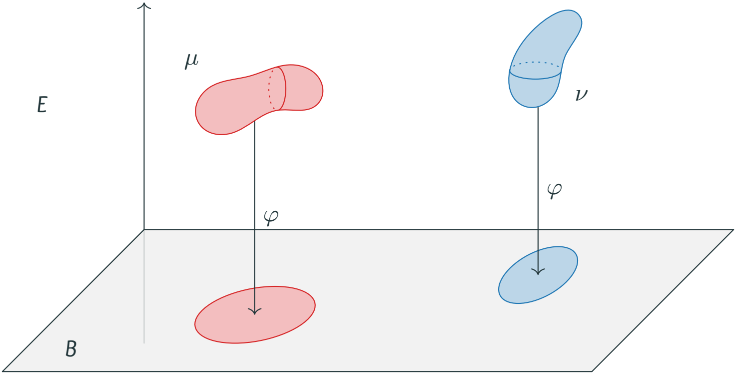
2. Monge maps for GW

2.1. A key lemma

"Let $\mu, \nu \in \mathcal{P}(E)$.



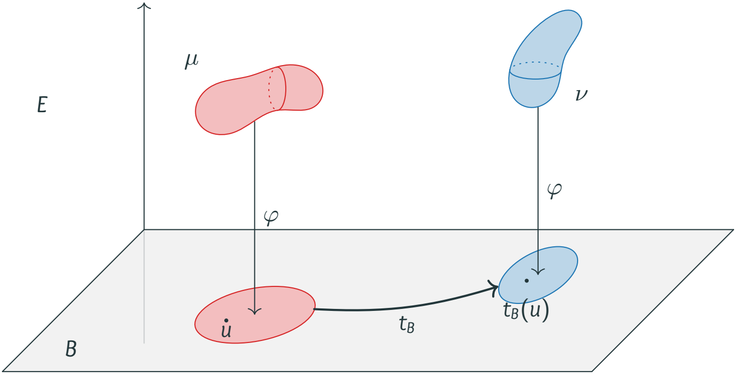
“Let $\mu, \nu \in \mathcal{P}(E)$. If we can send μ and ν in a space B by a function φ,



"Let $\mu, \nu \in \mathcal{P}(E)$. If we can send μ and ν in a space B by a function φ, s.t.

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

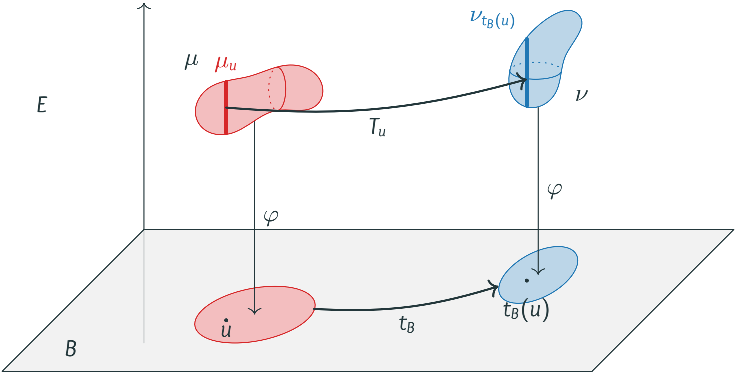
with $\tilde{c}$ a *twisted* cost on B,



"Let $\mu, \nu \in \mathcal{P}(E)$. If we can send μ and ν in a space B by a function φ, *s.t.*

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

with $\tilde{c}$ a *twisted* cost on B, then *we can construct an optimal map between μ and ν*."



Theorem: existence of a Monge map, inner product cost

Let E_0 be a measurable space and B_0 and F be complete Riemannian manifolds. Let $\mu, \nu \in \mathcal{P}(E_0)$ with *compact support*. Assume that there exists a set $E \subset E_0$ s.t. $\mu(E) = 1$ and that there exists a measurable map $\Phi : E \to B_0 \times F$ that is injective and whose inverse on its image is measurable as well. Let $\varphi \triangleq p_B \circ \Phi : E \to B_0$. Let $c : E_0 \times E_0 \to \mathbb{R}$ and suppose that there exists a *twisted* $\tilde{c} : B_0 \times B_0 \to \mathbb{R}$ s.t.

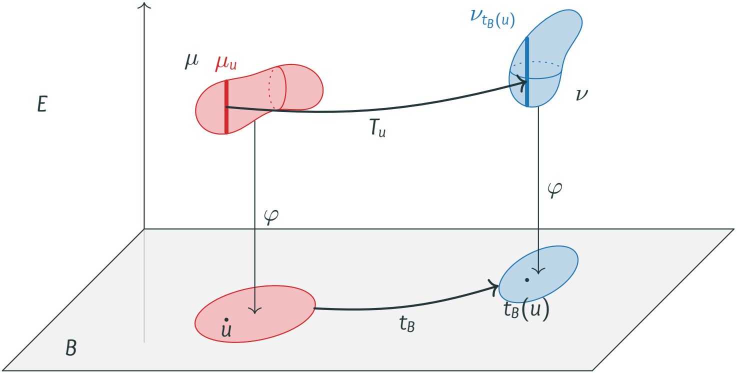
$$c(x, y) = \tilde{c}(\varphi(x), \varphi(y)) \quad \textit{for all } x, y \in E_0 \, .$$

Assume that $\varphi_\# \mu \ll \mathcal{L}_{B_0}$ and let thus t_B denote the unique Monge map between $\varphi_\# \mu$ and $\varphi_\# \nu$ for this cost. Suppose that there exists a disintegration $((\Phi_\# \mu)_u)_u$ of $\Phi_\# \mu$ by p_B s.t. for $\varphi_\# \mu$-a.e. u, $(\Phi_\# \mu)_u \ll \mathrm{vol}_F$.

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

$$\Phi \circ T \circ \Phi^{-1}(u, v) = (t_B(u), t_F(u, v)) = \Big(\underbrace{\tilde{c}\text{-}\exp_u(\nabla f(u))}_{\in B}, \underbrace{\exp_v(\nabla g_u(v))}_{\in \text{ fiber}} \Big) \, ,$$

with $f : B_0 \to \mathbb{R}$ $\tilde{c}$-convex and $g_u : F \to \mathbb{R}$ $d_F^2/2$-convex for $\varphi_\# \mu$-a.e. u.



1. *transport in B:* $\tilde{c}$ satisfies (Twist) on B;
2. *transport the fibers:* choose a map for each couple of fibers $(\mu_u, \nu_{t_B(u)})$
3. is $T(u, x) = T_u(x)$ *measurable*? need theorem! adaptation of [Fontbona et al., 2010] to the manifold setting

Take-home message: $c(x, y) = \tilde{c}(\varphi(x), \varphi(y))$ with $\tilde{c}$ twisted $\Longrightarrow$ map

2. Monge maps for GW

2.2. Application: inner product cost

Let's work on (GW inner prod):

$$\min_{\pi \in \Pi(\mu,\nu)} \iint \left| \langle x, x' \rangle - \langle y, y' \rangle \right|^2 \mathrm{d}\pi(x,y)\,\mathrm{d}\pi(x',y') \qquad \text{(GW inner prod)}$$

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

$\implies$ OT problem with $c(x,y) = -\int \langle x, x' \rangle \langle y, y' \rangle \,\mathrm{d}\pi^\star(x',y') = \cdots = -\langle M^\star x, y \rangle$

$$\text{where } M^\star \triangleq \int y' x'^\top \,\mathrm{d}\pi^\star(x',y') \in \mathbb{R}^{n \times n}$$

rk $M^\star$	$= n$	$\leq n-1$
twist	✓	·
subtwist	✓	·
m-twist, $m \geq 2$	✓	·

1. *a simplification:* up to SVD, suppose $M^\star$ is a diagonal matrix of singular values:

$$M^\star = \begin{pmatrix} \sigma_1 & & & & \\ & \ddots & & & \\ & & \sigma_h & & \\ & & & 0 & \\ & & & & \ddots \\ & & & & & 0 \end{pmatrix}$$

2. *rephrase the cost:*

$$\begin{aligned} c(x,y) &= -\langle M^\star x, y\rangle \\ &= -\sum_{i=1}^{h} \sigma_i x_i y_i \\ &\triangleq \tilde{c}(p(x), p(y)) \qquad \text{with } p \text{ the orthogonal projection on } \mathbb{R}^h. \end{aligned}$$

3. *apply key lemma!*
   - B is $\mathbb{R}^h$
   - fibers are $\mathbb{R}^{n-h}$
   - $\tilde{c}$ is twisted on $\mathbb{R}^h$

$\Rightarrow$ optimal map + structure!

$$\text{for } x = (u,v) \in \mathbb{R}^h \times \mathbb{R}^{n-h}, \quad T(u,v) = (\nabla f \circ M^\star(u), \nabla g_u(v)).$$

$\square$

2. Monge maps for GW

2.3. Application: quadratic cost

Similarly, we work on (GW quadratic) and relax to a classical OT problem with

$$c(x,y) = -|x|^2|y|^2 - 4\langle M^\star x, y\rangle\,.$$

rk $M^\star$	$= n$	$= n-1$	$\leq n-2$
twist	$\cdot$	$\cdot$	$\cdot$
subtwist	✓	$\cdot$	$\cdot$
2-twist	✓	✓	$\cdot$
m-twist, $m \geq 3$	$\cdot$	$\cdot$	$\cdot$

Similarly, we work on (GW quadratic) and relax to a classical OT problem with

$$c(x,y) = -|x|^2|y|^2 - 4\langle M^\star x, y\rangle \,.$$

$\operatorname{rk} M^\star$	$= n$	$= n-1$	$\leq n-2$
twist	·	·	·
subtwist	✓	·	·
2-twist	✓	✓	·
m-twist, $m \geq 3$	·	·	·
	$\Downarrow$	$\Downarrow$	$\Downarrow$
	map/anti-map & bimap	bimap	...

Theorem: quadratic cost

Let $\mu, \nu \in \mathcal{P}(\mathbb{R}^n)$ of compact support. Suppose that μ has a density. Let $\pi^\star$ be an optimal plan and $M^\star \triangleq \int y' x'^\top \, \mathrm{d}\pi^\star(x', y')$. Then:

- ✓ if $\operatorname{rk} M^\star = n$, there is an optimal *bimap* with one map being *one-to-one*,
- ✓ if $\operatorname{rk} M^\star = n - 1$, there is an optimal *bimap*,
- (!!) if $\operatorname{rk} M^\star \leq n - 2$, *there is an optimal map*!

1. *a simplification:* up to SVD, $M^\star$ is diagonal:

$$M^\star = \begin{pmatrix} \sigma_1 & & & & \\ & \ddots & & & \\ & & \sigma_h & & \\ & & & 0 & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix}. \qquad \text{We note } x = (\underbrace{x_1, \ldots, x_h}_{x_H}, \underbrace{x_{h+1}, \ldots, x_n}_{x_\perp}).$$

2. *rephrase the cost:*

$$\begin{aligned} -c(x,y) &= |x|^2|y|^2 + 4\langle M^\star x, y\rangle \\ &= |x_H|^2|y_H|^2 + |x_H|^2|y_\perp|^2 + |x_\perp|^2|y_H|^2 + |x_\perp|^2|y_\perp|^2 + 4\langle \tilde{M}x_H, y_H\rangle \\ &= |x_H|^2|y_H|^2 + |x_H|^2 n(y) + n(x)|y_H|^2 + n(x)n(y) + 4\langle \tilde{M}x_H, y_H\rangle \\ &\triangleq -\tilde{c}(\varphi(x), \varphi(y)), \end{aligned}$$

   with $n : x \mapsto |x_\perp|^2$ and $\varphi : x \mapsto (x_H, |x_\perp|^2)$.

3. *apply key lemma!*
   - B is $\mathbb{R}^h \times \mathbb{R}^+$
   - the fibers are spheres S^{n-h-1}
   - $\tilde{c}$ is twisted on $\mathbb{R}^h \times \mathbb{R}^+$

$\Rightarrow$ optimal map + structure!

for $x \approx (u,v) \in \mathbb{R}^h \times \mathbb{R}^+ \times S^{n-h-1}$, $\quad T(u,v) = (\tilde{c}\text{-}\exp_u(\nabla f(u)), \exp_v(\nabla g_u(v)))$.

$\square$

3. Complementary study of the quadratic cost in 1D

- μ, ν centered
- linearized problem:

$$\min_{\pi \in \Pi(\mu,\nu)} \int (-x^2y^2 - 4mxy)\, \mathrm{d}\pi(x,y)\,, \quad \text{where } m = \int x'y'\, \mathrm{d}\pi^\star(x',y') = \langle C_{xy},\, \pi^\star \rangle\,,$$

and $m \in [m_{\min}, m_{\max}]$ with $m_{\min} = \min_\pi \langle C_{xy},\, \pi \rangle$ and $m_{\max} = \max_\pi \langle C_{xy},\, \pi \rangle$

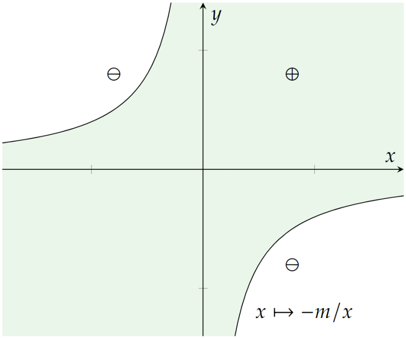
- in 1D, *submodularity* [Carlier, 2008, Santambrogio, 2015]

Submodular cost

Let $\mathcal{X}, \mathcal{Y} \subset \mathbb{R}$. We say that $c \in C^2 : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ is **submodular** if

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

Let $\mu, \nu \in \mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{Y})$ of finite transport cost. If c satisfies (Submod), then $\pi^{\oplus}_{mon}$ *is an optimal plan for* (KP).



- $c(x,y) = -x^2y^2 - 4mxy$ is submodular on the region $S = \{(x,y) \mid xy \geq -m\}$ if $m \geq 0$

- expect increasing on S and decreasing elsewhere?

3. Complementary study of the quadratic cost in 1D

3.1. Computation of non-monotone optimal plans

Sub-optimality of the monotone rearrangements

Theorem [Vayer, 2020]

In the discrete case in dimension 1 with $N = M$ and $a = b = \mathbb{1}_N$, *either* $\pi^{\oplus}_{mon}$ (eq. identity $\sigma(i) = i$) *or* $\pi^{\ominus}_{mon}$ (eq. anti-identity $\sigma(i) = N + 1 - i$) *is optimal* for (GW quadratic).

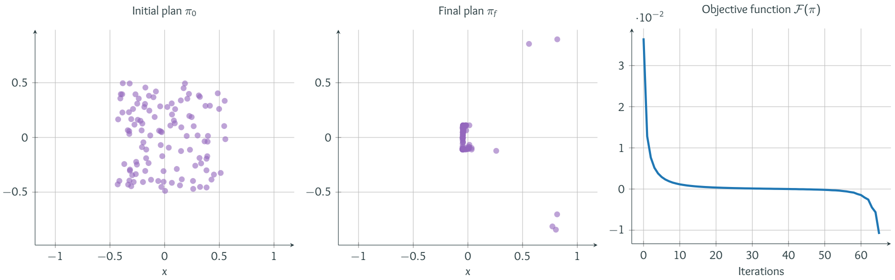
- **empirically:** very often true when generating points at random
- **literature:** *counter-example* by [Beinert et al., 2022] for $N \geq 7$ points
- **here:** procedure to *automatically* obtain additional *counter-examples*

Sub-optimality of the monotone rearrangements

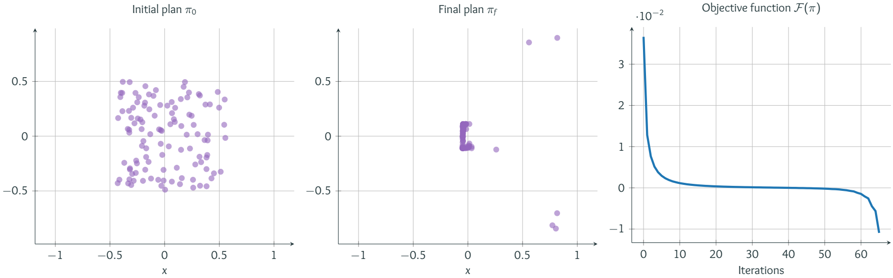
Theorem [Vayer, 2020]

In the discrete case in dimension 1 with $N = M$ and $a = b = \mathbb{1}_N$, *either* $\pi^{\oplus}_{mon}$ (eq. identity $\sigma(i) = i$) *or* $\pi^{\ominus}_{mon}$ (eq. anti-identity $\sigma(i) = N + 1 - i$) *is optimal* for (GW quadratic).

- **empirically:** very often true when generating points at random
- **literature:** *counter-example* by [Beinert et al., 2022] for $N \geq 7$ points
- **here:** procedure to *automatically* obtain additional *counter-examples*



Sub-optimality of the monotone rearrangements



(Left) Objective function $\mathcal{F}$. **(Center)** Initial plan π_0, generated at random. **(Right)** Final plan π_f.

Procedure:

- $\pi = \frac{1}{N}\sum_{i=1}^{N} \delta_{(x_i, y_i)}$
- move away from measures of optimal plans $\pi_{\text{mon}}^{\oplus}$ and $\pi_{\text{mon}}^{\ominus}$ by gradient descent:

$$\mathcal{F}(\pi) \triangleq \underbrace{c_{GW}(\pi)}_{\substack{\text{performance}\\ \text{of } \pi}} - \underbrace{\min\left\{c_{GW}(\pi_{\text{mon}}^{\oplus}),\ c_{GW}(\pi_{\text{mon}}^{\ominus})\right\}}_{\substack{\text{performance of the monotone}\\ \text{rearrangements of the marginals of } \pi}}$$

- results similar to [Beinert et al., 2022]!

What happens? + computation of optimal bimaps

- still, $\pi_{\text{mon}}^{\oplus}$ and $\pi_{\text{mon}}^{\ominus}$ are very often optimal in practice: *what happens?*
- generate measures with N points at random and look at the optimal plan for GW! **how to find it?**
  1. $\pi^\star$ optimal for GW
  2. $\implies$ optimal for *linearized* $\mathrm{GW}(m(\pi^\star))$
  3. consider all linearized $\mathrm{GW}(m)$ for $m \in [m_{\min}, m_{\max}]$ and take all optimal plans $\pi_m^\star$ (easy, linear programs!)
  4. $\pi^\star$ is the one that performs best on GW

What happens? + computation of optimal bimaps

- still, $\pi^{\oplus}_{\text{mon}}$ and $\pi^{\ominus}_{\text{mon}}$ are very often optimal in practice: *what happens?*
- generate measures with N points at random and look at the optimal plan for GW! **how to find it?**
  1. $\pi^\star$ optimal for GW
  2. $\implies$ optimal for *linearized* $\mathrm{GW}(m(\pi^\star))$
  3. consider all linearized $\mathrm{GW}(m)$ for $m \in [m_{\min}, m_{\max}]$ and take all optimal plans $\pi^\star_m$ (easy, linear programs!)
  4. $\pi^\star$ is the one that performs best on GW

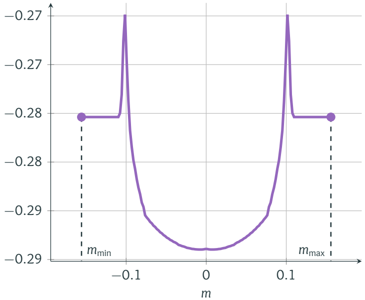


Figure 3: Graph of $m \mapsto \mathrm{GW}(\pi^\star_m)$.

What happens? + computation of optimal bimaps

- **Note:** sum of Diracs $\mu = (X, \mathbb{1}_N) \rightarrow$ density $\mu = (\hat{X}, a)$ by convolution with small Gaussian σ
- evolution of $\pi^\star_m$ as a function of m
  - with N random points: *[all]*, *[zoom]*

- **Note:** sum of Diracs $\mu = (X, \mathbb{1}_N) \to$ density $\mu = (\hat{X}, a)$ by convolution with small Gaussian σ
- evolution of $\pi_m^\star$ as a function of m
  - with N random points: *[all]*, *[zoom]*
  - with counter-examples of before: *[all]*, *[zoom]*
- **Note:** *a priori* no reason to work, and indeed it does not work most of the time

Algorithm 1 Generating bimaps from adversarial examples.

Input: an adversarial plan $\pi_f = \mathrm{id}(X_f, Y_f)$

Parameters: σ, $N_{\Delta x}$, $N_{\Delta m}$

Algorithm:

$a \leftarrow \texttt{convolution}(X_f, \sigma, N_{\Delta x})$

$b \leftarrow \texttt{convolution}(Y_f, \sigma, N_{\Delta x})$ ▷ (optional)

$m_{\min} \leftarrow \min_{\pi \in U(a,b)} \langle C_{xy},\ \pi \rangle$ ▷ solve linear programs

$m_{\max} \leftarrow \max_{\pi \in U(a,b)} \langle C_{xy},\ \pi \rangle$

$\texttt{GW\_scores} \leftarrow \texttt{[]}$

for $m \in \{m_{\min}, \ldots, m_{\max}\}$ **do** ▷ with $N_{\Delta m}$ points

  $\pi_m^\star \leftarrow \arg\min_{\pi \in U(a,b)} \langle C_{\mathrm{GW}(m)},\ \pi \rangle$ ▷ solve linear program

  append $\mathrm{GW}(\pi_m^\star)$ to $\texttt{GW\_scores}$

end for

$\pi^\star \leftarrow \arg\max_\pi \texttt{GW\_scores}$ ▷ take best plan for GW

return $\pi^\star$

Outputs: $\pi^\star$ optimal for GW

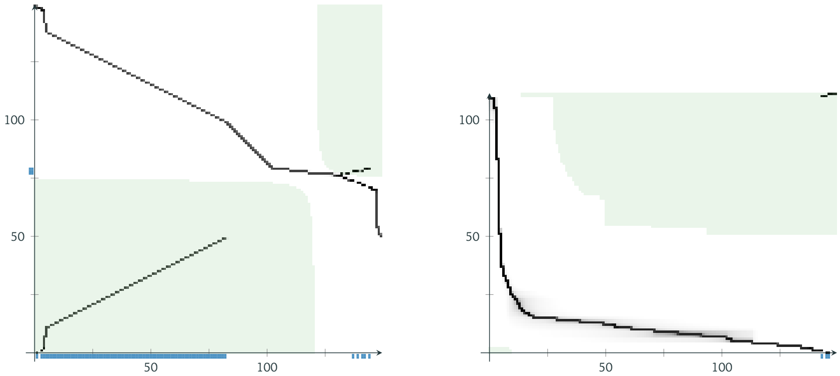


Figure 4: Optimal correspondence plan $\pi^\star$ (in log scale), **(Left)** starting from a plan with both marginals convolved or **(Right)** with only μ convolved. Parameters: $\sigma = 5.10^{-3}$, $N_{\Delta x} = 150$, $N_{\Delta m} = 2000$.

- small bimap region for **(Right)** ?
- "but it's a map from $\mathcal{Y}$ to $\mathcal{X}$!" no, in both cases, no map neither $\mu \to \nu$ nor $\nu \to \mu$

3. Complementary study of the quadratic cost in 1D

3.2. Instability of the optimality of monotone optimal plans

Question

Is having $\pi^{\oplus}_{\mathrm{mon}}$ or $\pi^{\ominus}_{\mathrm{mon}}$ as optimal *stable*?

- minimum are optimal correspondence plans

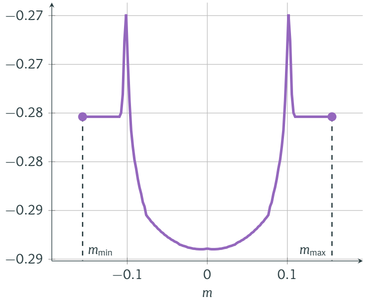


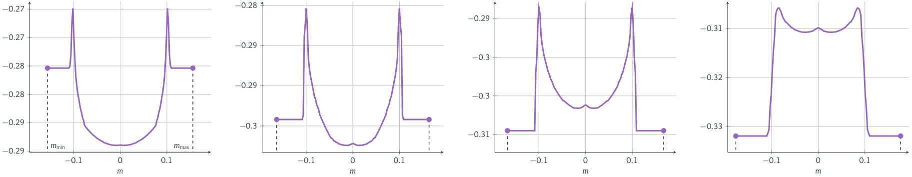
Figure 5: Graph of $m \mapsto \mathrm{GW}(\pi^{\star}_{m})$.

Instability of the optimality of monotone rearrangements

Question

Is having $\pi^{\oplus}_{\text{mon}}$ or $\pi^{\ominus}_{\text{mon}}$ as optimal *stable*?

- minimum are optimal correspondence plans
  - small σ: optimal plan not monotone by construction;
  - large σ: monotone are optimal again.
  - *phase transition:* landscape of $m \mapsto \mathrm{GW}(\pi^\star_m)$ while increasing σ



$\sigma_1 = 8.10^{-3}$ $\sigma_2 = 8.8.10^{-3}$ $\sigma_3 = 10^{-2}$ $\sigma_4 = 3.10^{-2}$

Figure 5: Graphs of $m \mapsto \mathrm{GW}(\pi^\star_m)$ with [Beinert et al., 2022], $N = 7$ points.

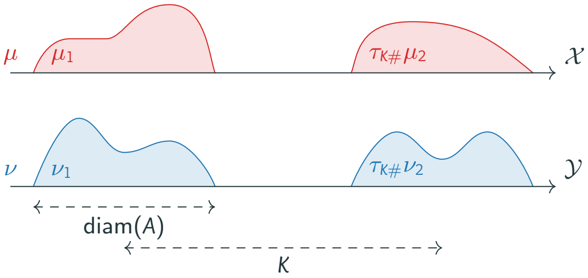
3. Complementary study of the quadratic cost in 1D

3.3. (A positive result for measures with two components)

(A positive result for measures with two components)

Let $\mu_1, \mu_2, \nu_1, \nu_2$ probability measures on $A \subset \mathbb{R}$ compact. Fix $t \in (0,1)$ and $K > \mathrm{diam}(A)$. Let $\tau_K : x \mapsto x + K$. Introduce measures

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



Theorem

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

4. Summary & discussion

Summary & discussion

Contributions

1. **Thm:** always a *map* for (GW inner prod)
2. **Thm:** a *map* or *bimap* for (GW quadratic)
3. **Conj:** this second claim is *tight*
4. **Algo:** *non-optimality* of monotone + *instability* for (GW quadratic)
5. (**Thm:** monotone rearrangement optimal for (GW quadratic) between measures composed of *two distant parts*)

Very soon:

- remove assumption of compact support
- try to show decomposition $\pi_{\text{mon}}^{\oplus} + \pi_{\text{mon}}^{\ominus}$ for (GW quadratic)

Future work:

- quadratic cost:
  - better understanding of the 1d case (maybe simpler)
- inner product cost:
  - is (GW inner prod) computationally tractable?
- other cost functions
  - apply key lemma to other costs $c_{\mathcal{X}}$ and $c_{\mathcal{Y}}$?

Links

- *preprint* on HAL and arxiv: "*On the existence of Monge maps for the Gromov–Wasserstein distance*" https://hal.archives-ouvertes.fr/hal-03818500
- (soon) *code* on GitHub at https://github.com/theodumont/monge-gromov-wasserstein

Thank you!

References i

Ahmad, N., Kim, H. K., and McCann, R. J. (2011).
Optimal transportation, topology and uniqueness.
Bulletin of Mathematical Sciences, 1(1):13–32.

Alvarez-Melis, D., Jegelka, S., and Jaakkola, T. S. (2019).
Towards optimal transport with global invariances.
In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 1870–1879. PMLR.

Beinert, R., Heiss, C., and Steidl, G. (2022).
On assignment problems related to gromov-wasserstein distances on the real line.
arXiv preprint arXiv:2205.09006.

Brenier, Y. (1987).
Décomposition polaire et réarrangement monotone des champs de vecteurs.
CR Acad. Sci. Paris Sér. I Math., 305:805–808.

References ii

Carlier, G. (2008).
Remarks on toland's duality, convexity constraint and optimal transport.
Pacific Journal of Optimization, 4(3):423–432.

Chiappori, P.-A., McCann, R. J., and Nesheim, L. P. (2010).
Hedonic price equilibria, stable matching, and optimal transport: equivalence, topology, and uniqueness.
Economic Theory, 42(2):317–354.

Fontbona, J., Guérin, H., and Méléard, S. (2010).
Measurability of optimal transportation and strong coupling of martingale measures.
Electronic communications in probability, 15:124–133.

Kantorovich, L. (1942).
On the translocation of masses.
Dokl. Akad. Nauk. USSR, 37(7–8):227–229.

McCann, R. J. and Guillen, N. (2011).
Five lectures on optimal transportation: geometry, regularity and applications.
Analysis and geometry of metric measure spaces: lecture notes of the séminaire de Mathématiques Supérieure (SMS) Montréal, pages 145–180.

Mémoli, F. (2011).
Gromov–wasserstein distances and the metric approach to object matching.
Foundations of computational mathematics, 11(4):417–487.

Moameni, A. (2016).
A characterization for solutions of the monge-kantorovich mass transport problem.
Mathematische Annalen, 365(3):1279–1304.

Monge, G. (1781).
Mémoire sur la théorie des déblais et des remblais.
Mem. Math. Phys. Acad. Royale Sci., pages 666–704.

Peyré, G., Cuturi, M., et al. (2019).
Computational optimal transport: With applications to data science.
Foundations and Trends® in Machine Learning, 11(5-6):355–607.

Santambrogio, F. (2015).
Optimal transport for applied mathematicians.
Birkäuser, NY, 55(58-63):94.

Séjourné, T., Vialard, F.-X., and Peyré, G. (2021).
The unbalanced gromov wasserstein distance: Conic formulation and relaxation.
Advances in Neural Information Processing Systems, 34:8766–8779.

Sturm, K.-T. (2020).
The space of spaces: curvature bounds and gradient flows on the space of metric measure spaces.
arXiv preprint arXiv:1208.0434.

Vayer, T. (2020).
A contribution to optimal transport on incomparable spaces.
arXiv preprint arXiv:2011.04447.

Villani, C. (2008).
***Optimal transport: old and new*, volume 338.**
Springer.

Inspiration for figures

- the figure on slides 3 and 4, as well as the ones on slide 5 are adapted from a talk by Lénaïc Chizat;
- the GW figure on slide 13 is adapted from [Peyré et al., 2019];
- all other figures are my own.