

# Constrained Optimal Transport Maps

Eloi Tanguy, Agnès Desolneux and Julie Delon

MAP5, Université Paris-Cité

17th February 2025



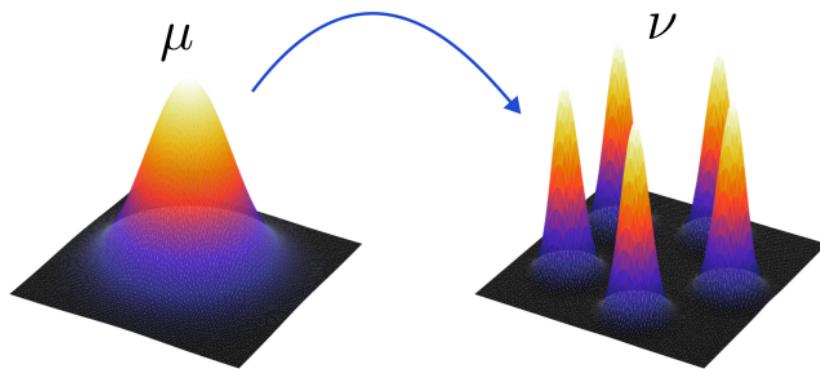
# ① Transporting Measures

## ② Constrained Approximate Transport Maps

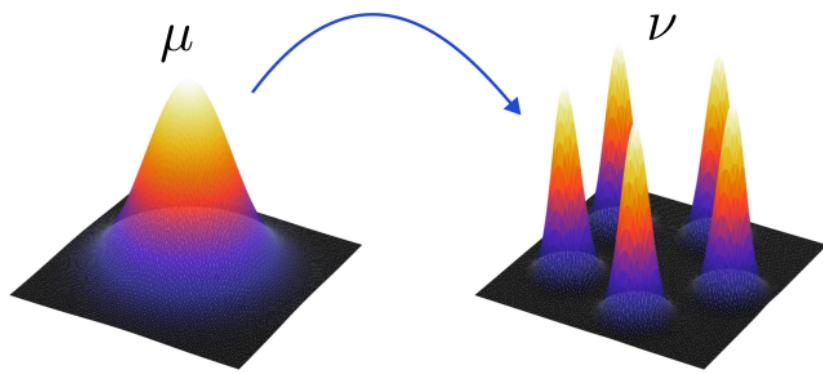
## ③ Zoom on the $L^2$ Case on $\mathbb{R}^d$

## ④ Illustration with Colour Transfer

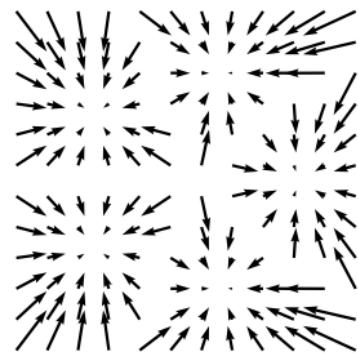
## Push-Forward Measures



## Push-Forward Measures

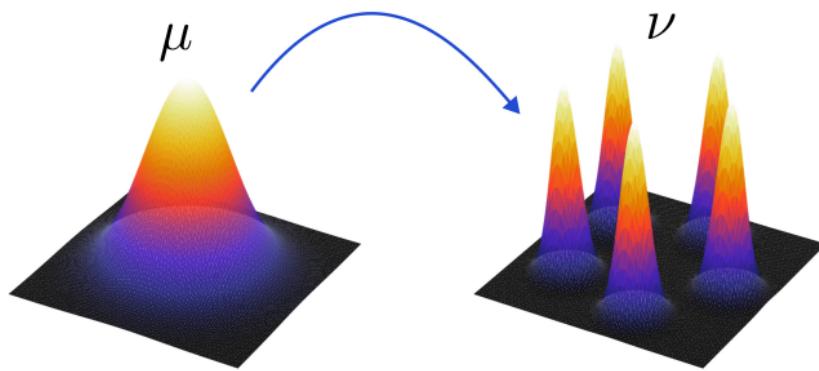
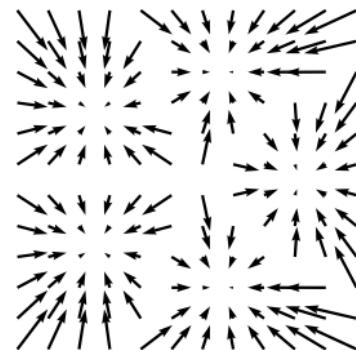
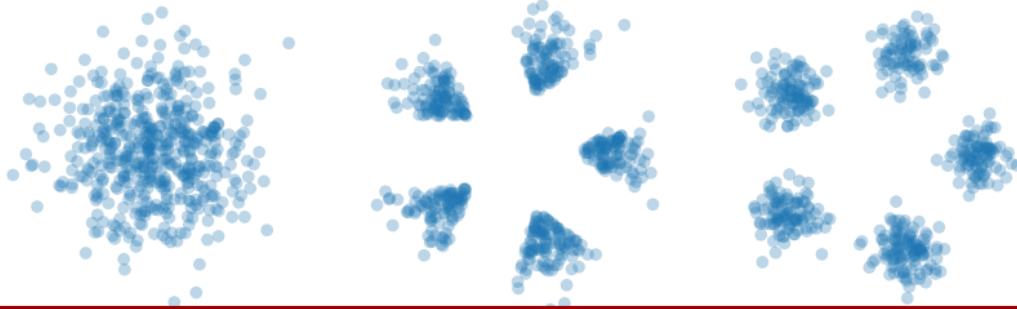


$$f \# \mu := \underset{X \sim \mu}{\text{Law}}[f(X)].$$



## Push-Forward Measures

$$f\#\mu := \underset{X \sim \mu}{\text{Law}}[f(X)].$$

Gaussian  $\mu$ Image  $f\#\mu$ Gaussian Mixture  $\nu$ 

# The Monge Problem

$$\inf_{T: T\#\mu=\nu} \int c(x, T(x)) d\mu(x).$$

# The Monge Problem

$$\inf_{T: T\#\mu=\nu} \int c(x, T(x)) d\mu(x).$$

## Brenier's Theorem

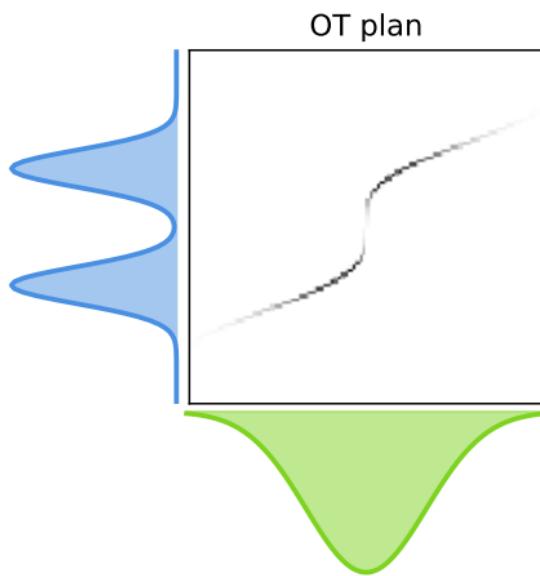
If  $c(x, y) = \|x - y\|_2^2$ ,  
and  $\mu \ll \mathcal{L}^d$ , then  
there is a unique solu-  
tion, and  $T = \nabla \varphi$  with  
 $\varphi$  convex.

# The Monge Problem

$$\inf_{T: T\#\mu=\nu} \int c(x, T(x)) d\mu(x).$$

## Brenier's Theorem

If  $c(x, y) = \|x - y\|_2^2$ ,  
and  $\mu \ll \mathcal{L}^d$ , then  
there is a unique solution,  
and  $T = \nabla \varphi$  with  
 $\varphi$  convex.

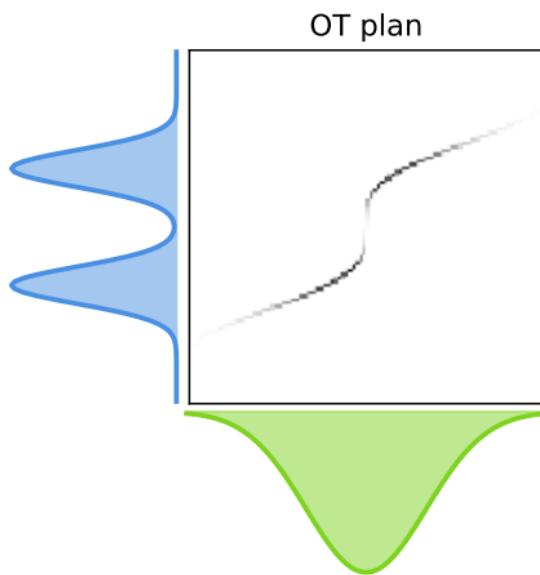


# The Monge Problem

$$\inf_{T: T\#\mu=\nu} \int c(x, T(x)) d\mu(x).$$

## Brenier's Theorem

If  $c(x, y) = \|x - y\|_2^2$ ,  
and  $\mu \ll \mathcal{L}^d$ , then  
there is a unique solu-  
tion, and  $T = \nabla \varphi$  with  
 $\varphi$  convex.



Kantorovich relaxation:  $\mathcal{T}_c(\mu, \nu) := \min_{\pi \in \Pi(\mu, \nu)} \int c(x, y) d\pi(x, y).$

## ① Transporting Measures

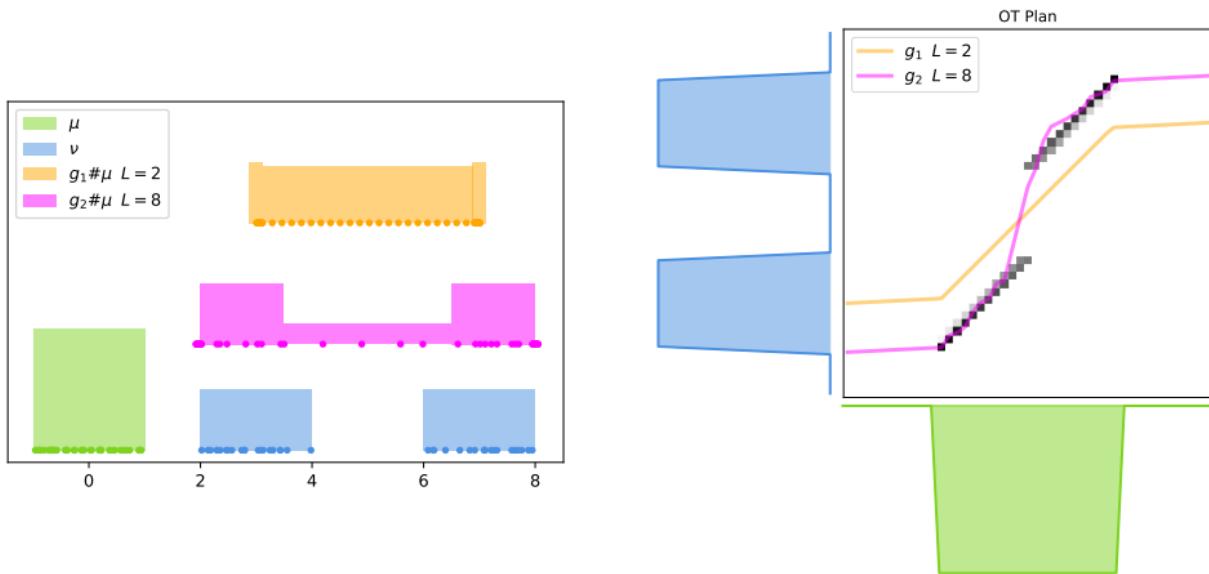
## ② Constrained Approximate Transport Maps

## ③ Zoom on the $L^2$ Case on $\mathbb{R}^d$

## ④ Illustration with Colour Transfer

## Problem statement

$$\mathcal{P} : \operatorname{argmin}_{g \in G} \mathcal{T}_c(g\#\mu, \nu)$$



# Smooth Strongly Convex Nearest Brenier Potentials (Paty 2020 [2])

Case  $G = G_{\ell,L} := \{\nabla\varphi : \ell I \preceq D^2\varphi \preceq L I\}$  and  $c(x,y) = \|x - y\|_2^2$ .

## Smooth Strongly Convex Nearest Brenier Potentials (Paty 2020 [2])

Case  $G = G_{\ell,L} := \{\nabla\varphi : \ell I \preceq D^2\varphi \preceq LI\}$  and  $c(x,y) = \|x - y\|_2^2$ .

$$\mu := \sum_{i=1}^n a_i \delta_{x_i}, \quad \nu := \sum_{j=1}^m b_j \delta_{y_j}.$$

## Smooth Strongly Convex Nearest Brenier Potentials (Paty 2020 [2])

Case  $G = G_{\ell,L} := \{\nabla\varphi : \ell I \preceq D^2\varphi \preceq LI\}$  and  $c(x,y) = \|x - y\|_2^2$ .

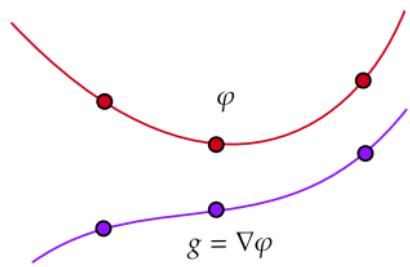
$$\mu := \sum_{i=1}^n a_i \delta_{x_i}, \quad \nu := \sum_{j=1}^m b_j \delta_{y_j}.$$

## Interpolation (Taylor 2017 [3])

$$\exists g = \nabla\varphi \in G_{\ell,L} :$$

$$\forall i, \quad g(x_i) = g_i, \quad \varphi(x_i) = \varphi_i$$

$$\text{iif } \forall i, j, \quad Q_{\ell,L}(x_i, x_j, \varphi_i, \varphi_j, g_i, g_j) \geq 0.$$



## Smooth Strongly Convex Nearest Brenier Potentials (Paty 2020 [2])

Case  $G = G_{\ell,L} := \{\nabla\varphi : \ell I \preceq D^2\varphi \preceq LI\}$  and  $c(x,y) = \|x - y\|_2^2$ .

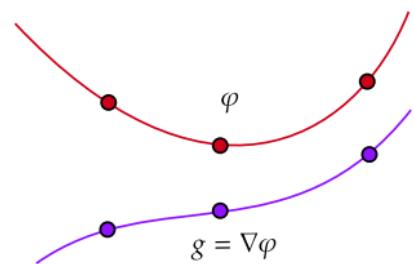
$$\mu := \sum_{i=1}^n a_i \delta_{x_i}, \quad \nu := \sum_{j=1}^m b_j \delta_{y_j}.$$

## Interpolation (Taylor 2017 [3])

$$\exists g = \nabla\varphi \in G_{\ell,L} :$$

$$\forall i, \quad g(x_i) = g_i, \quad \varphi(x_i) = \varphi_i$$

$$\text{iif } \forall i,j, \quad Q_{\ell,L}(x_i, x_j, \varphi_i, \varphi_j, g_i, g_j) \geq 0.$$



$$\operatorname{argmin}_{g \in G_{\ell,L}} W_2^2(g \# \mu, \nu) \longleftrightarrow \operatorname{argmin}_{\substack{\pi \in \mathbb{R}^{n \times m}, \varphi \in \mathbb{R}^n, g \in \mathbb{R}^{n \times d} \\ \pi \geq 0, \pi \mathbf{1} = a, \pi^T \mathbf{1} = b \\ Q_{\ell,L}(x_i, x_j, \varphi_i, \varphi_j, g_i, g_j) \geq 0}} \sum_{i,j} \|g_i - y_j\|_2^2 \pi_{i,j}.$$

# Sufficient Conditions for Existence

$$\mathcal{P} : \operatorname{argmin}_{g \in G} \mathcal{T}_c(g\#\mu, \nu)$$

## Existence if:

- Finite problem value,
- $c : \mathbb{R}^d \times \mathbb{R}^d \longrightarrow \mathbb{R}_+$  lower semi-continuous,
- $c(y_1, y_2) \geq \alpha + \eta(\|y_1 - y_2\|)$  with  $\eta$  non-decreasing and coercive.
- $G$  is a subclass of  $L$ -Lipschitz functions stable by local uniform limit.

Example classes: Neural Networks,  $G_{\ell, L}$ .

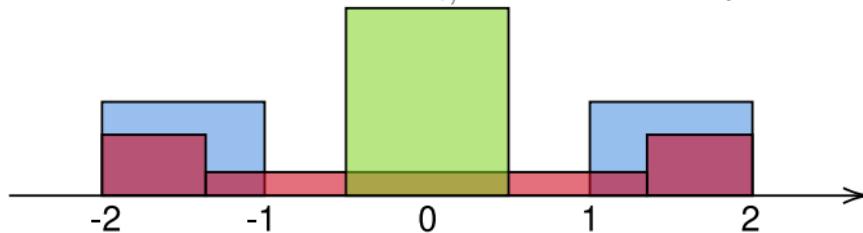
# Sufficient Conditions for Existence

$$\mathcal{P} : \operatorname{argmin}_{g \in G} \mathcal{T}_c(g\#\mu, \nu)$$

**Existence if:**

- Finite problem value,
- $c : \mathbb{R}^d \times \mathbb{R}^d \longrightarrow \mathbb{R}_+$  lower semi-continuous,
- $c(y_1, y_2) \geq \alpha + \eta(\|y_1 - y_2\|)$  with  $\eta$  non-decreasing and coercive.
- $G$  is a subclass of  $L$ -Lipschitz functions stable by local uniform limit.

Example classes: Neural Networks,  $G_{\ell, L}$ . Counter-example:



# SGD for Neural Networks

**Objective:**  $\min_{\theta} \mathcal{T}_c(g_{\theta}\#\mu, \nu).$

**Minibatch version:**

$$\min_{\theta} E(\theta) := \int \mathcal{T}_c(\delta_{g_{\theta}(X^{(n)})}, \delta_{Y^{(m)}}) d\mu^{\otimes n}(X^{(n)}) d\nu^{\otimes m}(Y^{(m)}).$$

$$\theta_{t+1} = \theta_t - \alpha_t \left[ \frac{\partial}{\partial \theta} \mathcal{T}_c(\delta_{g_{\theta}(X^{(n)})}, \delta_{Y^{(m)}}) \right]_{\theta=\theta_t}, \quad X^{(n)} \sim \mu^{\otimes n}, \quad Y^{(n)} \sim \nu^{\otimes m}.$$

# SGD for Neural Networks

**Objective:**  $\min_{\theta} \mathcal{T}_c(g_{\theta}\#\mu, \nu).$

**Minibatch version:**

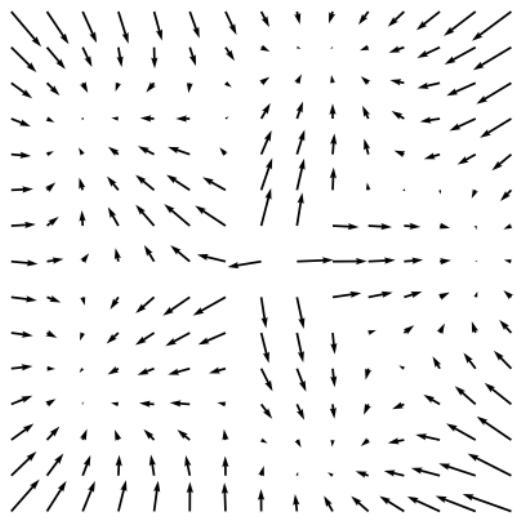
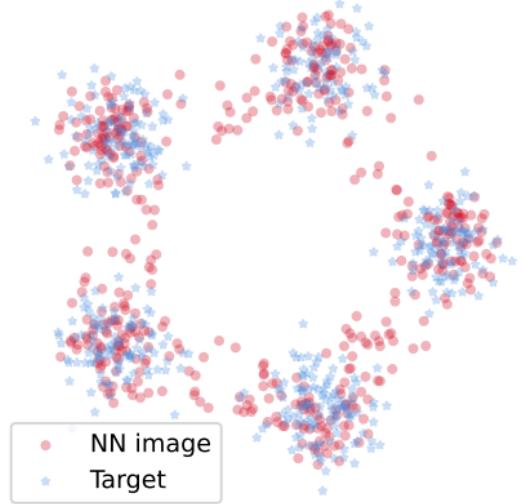
$$\min_{\theta} E(\theta) := \int \mathcal{T}_c(\delta_{g_{\theta}(X^{(n)})}, \delta_{Y^{(m)}}) d\mu^{\otimes n}(X^{(n)}) d\nu^{\otimes m}(Y^{(m)}).$$

$$\theta_{t+1} = \theta_t - \alpha t \left[ \frac{\partial}{\partial \theta} \mathcal{T}_c(\delta_{g_{\theta}(X^{(n)})}, \delta_{Y^{(m)}}) \right]_{\theta=\theta_t}, \quad X^{(n)} \sim \mu^{\otimes n}, \quad Y^{(n)} \sim \nu^{\otimes m}.$$

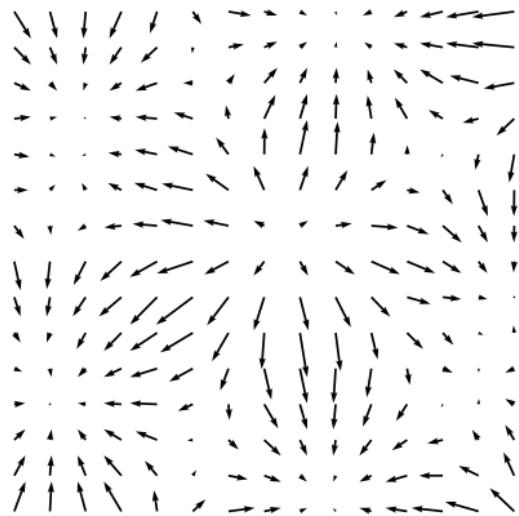
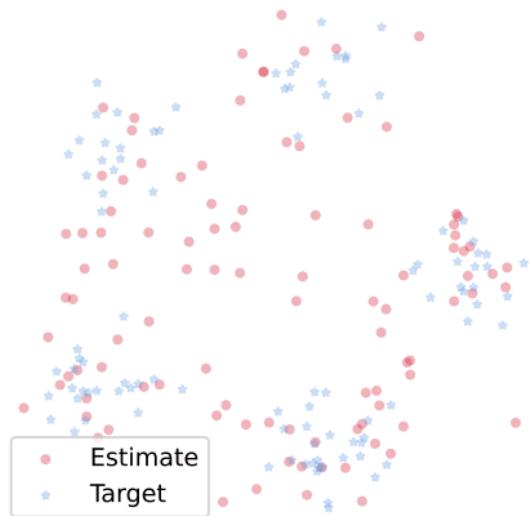
## SGD Convergence, using Bolte-Le-Pauwels [1]

For  $c$  and  $g$  semi-algebraic and  $\mu, \nu$  discrete or AC with semi-algebraic density, almost-surely accumulation points of  $(\theta_t)$  are Clarke critical points of  $E$ .

## Illustration: Neural Network Vector Fields

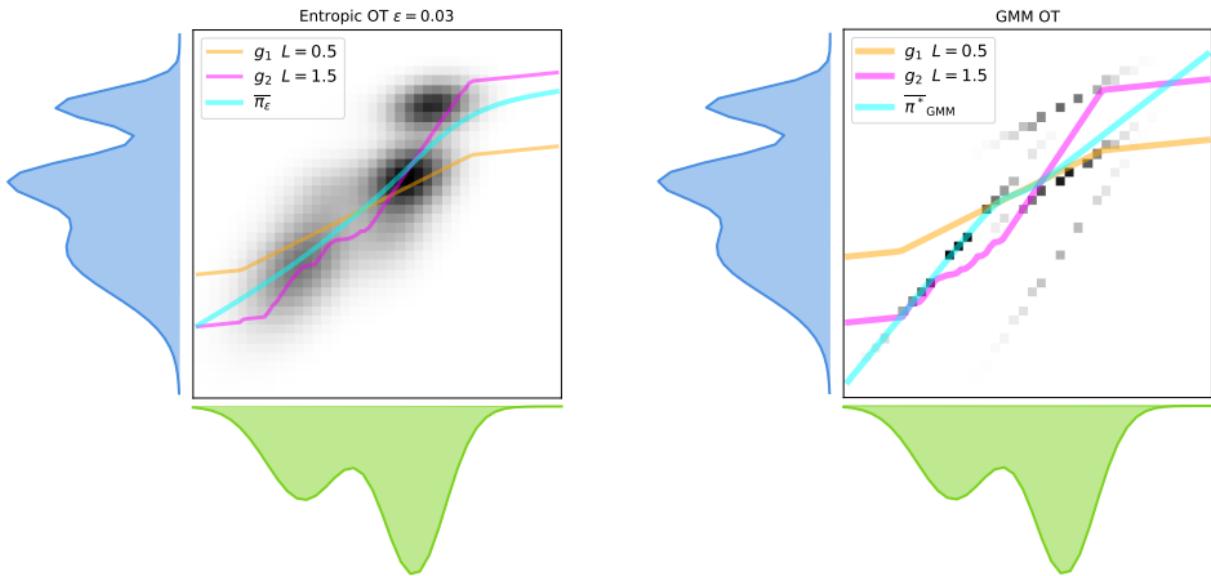


## Illustration: Gradients of Strongly Convex Functions



## Plan Variant 1/2

$$\mathcal{P}_{\text{plan}} : \operatorname{argmin}_{g \in G} \mathcal{T}_C((I, g)\#\mu, \gamma)$$



## Plan Variant 2/2

$$\mathcal{P}_{\text{plan}} : \operatorname{argmin}_{g \in G} \mathcal{T}_C((I, g)\#\mu, \gamma)$$

## Problem Equivalence

$$\mathcal{T}_C((I, g)\#\mu, \gamma) = \mathcal{T}_{c_2}(g\#\mu, \nu)$$

for  $C((x_1, x_2), (y_1, y_2)) = h(c_1(x_1, y_1), c_2(x_2, y_2))$  if:

- $c_1(x, x) = 0,$
- $h(u, v) \geq v,$
- $h(0, v) = v.$

Ex:  $C(\cdot, \cdot) = \|\cdot - \cdot\|_p^{qp}$ ,  $p \in [1, +\infty]$ ,  $q \geq 1$ .

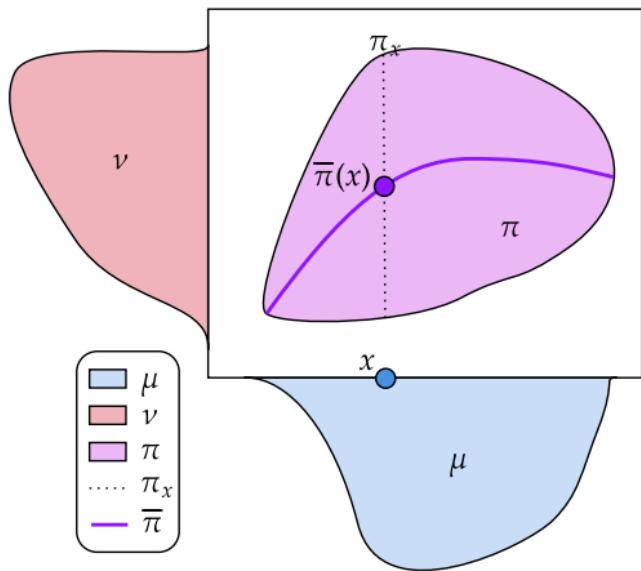
## ① Transporting Measures

## ② Constrained Approximate Transport Maps

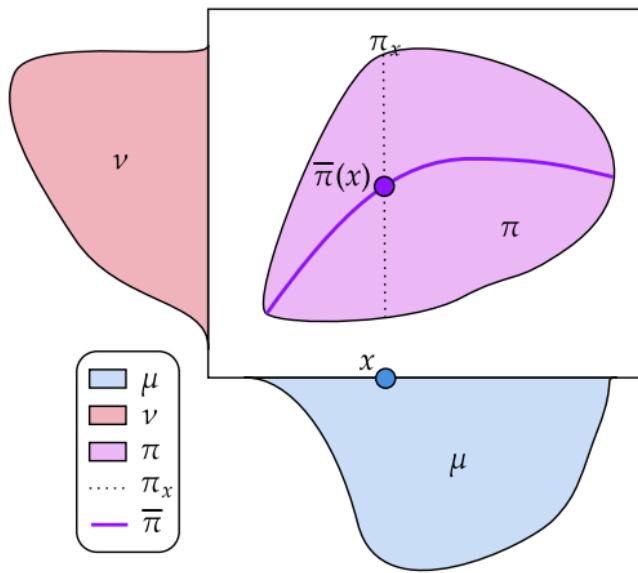
## ③ Zoom on the $L^2$ Case on $\mathbb{R}^d$

## ④ Illustration with Colour Transfer

# Barycentric Projections



## Barycentric Projections



$$\begin{aligned}\bar{\pi}(x) &= \int y d\pi_x(y). \\ \bar{\pi}(x) &= \mathbb{E}_{(X,Y) \sim \pi}[Y | X = x]. \\ \bar{\pi} &= \operatorname{argmin}_{f \in L^2(\mu)} \int \|f(x) - y\|_2^2 d\pi(x, y).\end{aligned}$$

Alternate Formulation for the  $L^2$  cost on  $\mathbb{R}^d$ 

$$\mathcal{P} : \min_{g \in G} \min_{\pi \in \Pi(\mu, \nu)} \int \|g(x) - y\|_2^2 d\pi(x, y)$$

With  $\pi$  fixed:

$$\int \|g(x) - y\|_2^2 d\pi(x, y) = \int \|g(x) - \bar{\pi}(x)\|_2^2 d\mu(x) + K(\bar{\pi}).$$

Alternate Formulation for the  $L^2$  cost on  $\mathbb{R}^d$ 

$$\mathcal{P} : \min_{g \in G} \min_{\pi \in \Pi(\mu, \nu)} \int \|g(x) - y\|_2^2 d\pi(x, y)$$

With  $\pi$  fixed:

$$\int \|g(x) - y\|_2^2 d\pi(x, y) = \int \|g(x) - \bar{\pi}(x)\|_2^2 d\mu(x) + K(\bar{\pi}).$$

**Question:** for  $\pi \in \Pi^*(\mu, \nu)$ , do we have

$$\operatorname{argmin}_{g \in G} \int \|g(x) - \bar{\pi}^*(x)\|_2^2 d\mu(x) \stackrel{?}{=} \operatorname{argmin}_{g \in G} \min_{\pi \in \Pi(\mu, \nu)} \int \|g(x) - y\|_2^2 d\pi(x, y).$$

## Positive Answer in 1D

$$\operatorname{argmin}_{g \in G} \int \|g(x) - \bar{\pi}^*(x)\|_2^2 d\mu(x) \stackrel{?}{=} \operatorname{argmin}_{g \in G} \min_{\pi \in \Pi(\mu, \nu)} \int \|g(x) - y\|_2^2 d\pi(x, y).$$

Equivalence to  $L^2$  projection in 1D for the  $L^2$  cost

If all  $g : \mathbb{R} \rightarrow \mathbb{R}$  are non-decreasing and  $\pi^* \in \Pi^*(\mu, \nu)$ , then

$$\mathcal{P} : \operatorname{argmin}_{g \in G} W_2^2(g \# \mu, \nu) = \operatorname{argmin}_{g \in G} \|g - \bar{\pi}^*\|_{L^2(\mu)}^2.$$

Counter-examples exist in higher dimensions. Generalises Paty 2020 [2].

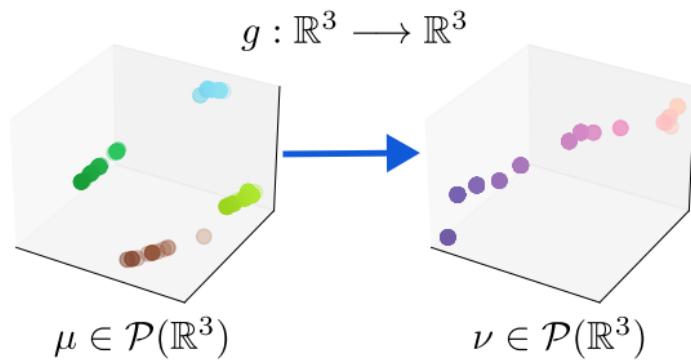
## ① Transporting Measures

## ② Constrained Approximate Transport Maps

## ③ Zoom on the $L^2$ Case on $\mathbb{R}^d$

## ④ Illustration with Colour Transfer

## Colour Transfer



# NN-OT Colour Transfer

Training Source



Evaluation Source



Training Target 1



NN Transfer on Training Source



NN Transfer on Evaluation Source



## NN-OT Colour Transfer with Outlier

Training Source



Evaluation Source



Training Target 3 (outlier + no weight clip)



NN Transfer on Training Source



NN Transfer on Evaluation Source



# NN-OT Colour Transfer with Outlier and Lipschitz Constraint

Training Source



Evaluation Source



Training Target 2 (outlier)



NN Transfer on Training Source



NN Transfer on Evaluation Source



*Thanks*

- [1] Jérôme Bolte, Tam Le, and Edouard Pauwels.  
Subgradient sampling for nonsmooth nonconvex minimization.  
*SIAM Journal on Optimization*, 33(4):2542–2569, 2023.
- [2] François-Pierre Paty, Alexandre d’Aspremont, and Marco Cuturi.  
Regularity as regularization: Smooth and strongly convex brenier potentials in optimal transport.  
In *International Conference on Artificial Intelligence and Statistics*, pages 1222–1232. PMLR, 2020.
- [3] Adrien B Taylor.  
*Convex interpolation and performance estimation of first-order methods for convex optimization.*  
PhD thesis, Catholic University of Louvain, Louvain-la-Neuve, Belgium, 2017.