

Regularizing Optimal Transport Using Regularity Theory

Learning meets Astrophysics

January 17, 2020

FRANÇOIS-PIERRE PATY

francoispierrepaty.github.io

*Based on a joint work with
Alexandre d'Aspremont and
Marco Cuturi (AISTATS 2020)*



Regularizing Optimal Transport Using Regularity Theory

Learning meets Astrophysics

January 17, 2020

FRANÇOIS-PIERRE PATY

francoispierrepaty.github.io

*Based on a joint work with
Alexandre d'Aspremont and
Marco Cuturi (AISTATS 2020)*



Regularizing Optimal Transport Using Regularity Theory

Learning meets Astrophysics

January 17, 2020

FRANÇOIS-PIERRE PATY

francoispierrepaty.github.io

*Based on a joint work with
Alexandre d'Aspremont and
Marco Cuturi (AISTATS 2020)*



A portrait of Gaspard Monge, a French mathematician and physicist. He is depicted from the waist up, wearing a dark blue coat with gold embroidery and a white cravat. He has white hair and is looking slightly to the right. The background is dark and indistinct.

INTRODUCTION



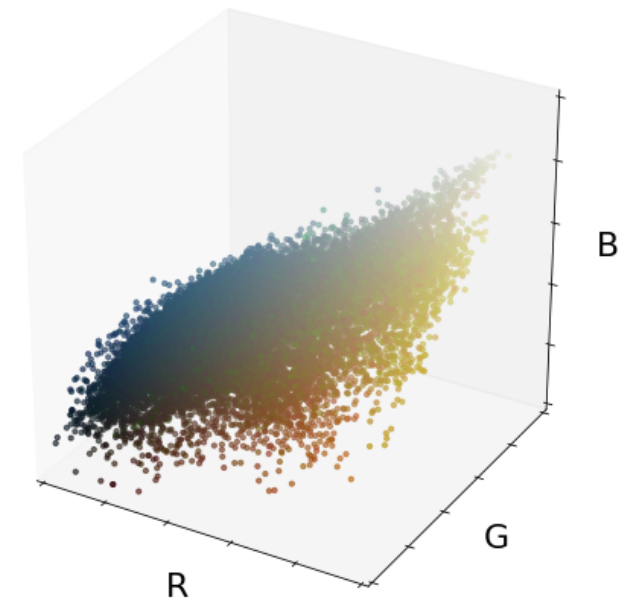
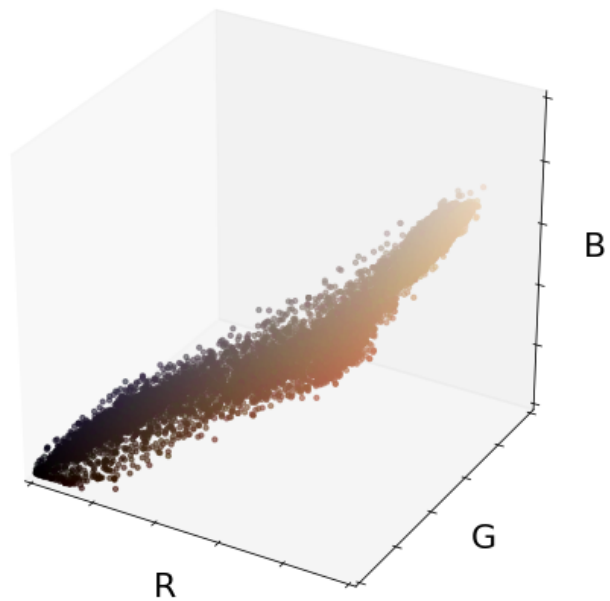


Color Transfer Map



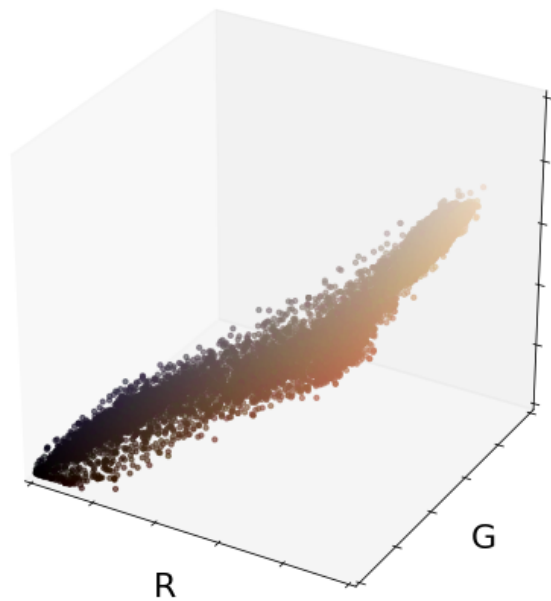


Color Transfer Map

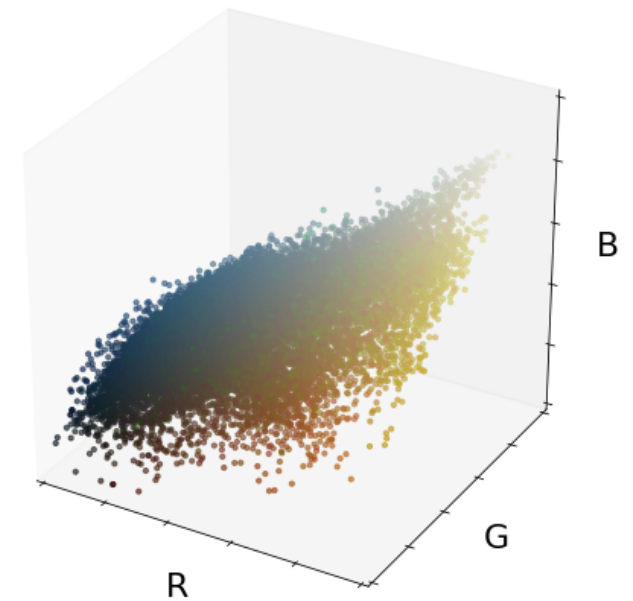




Color Transfer Map

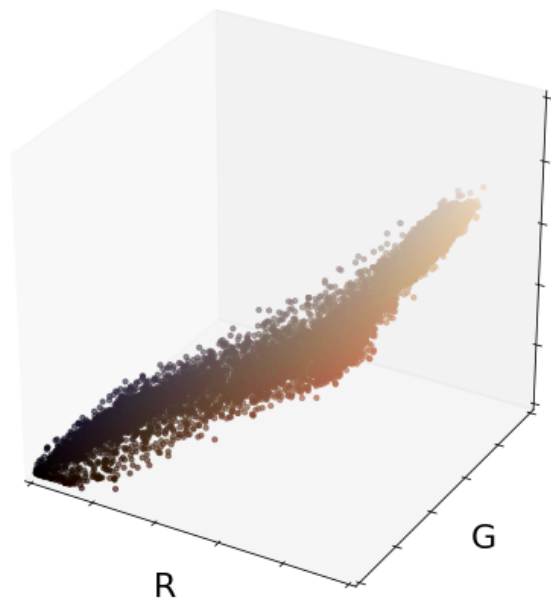


Matching



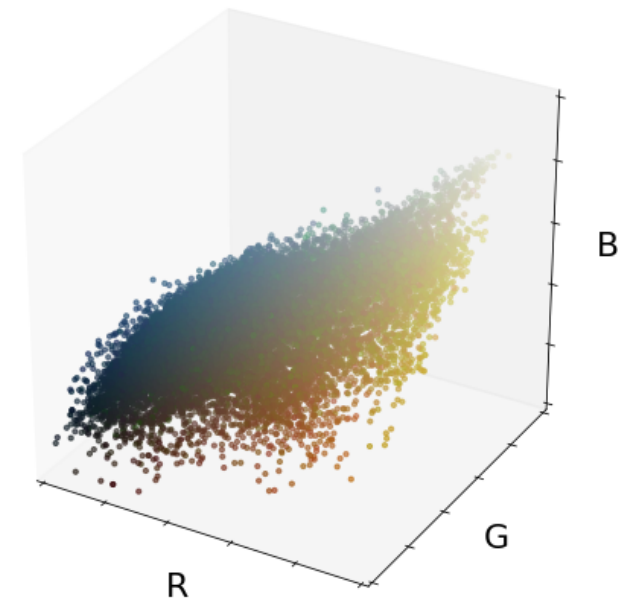


Color Transfer Map

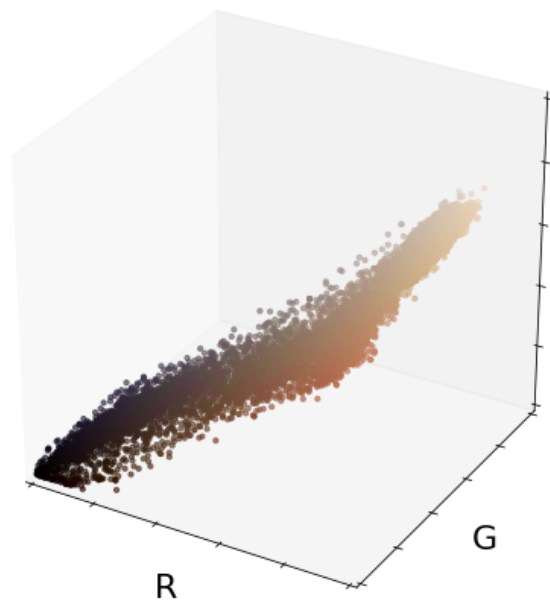


B

Matching



B

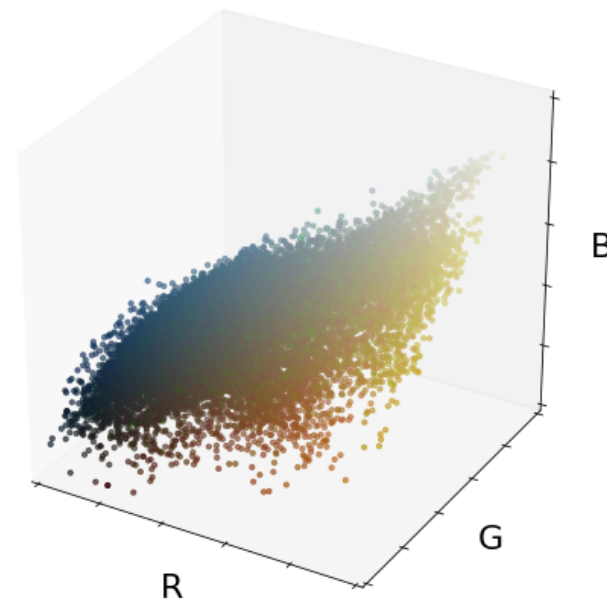


x_1, \dots, x_n

B

Matching

$\sigma \in \mathfrak{S}_n$

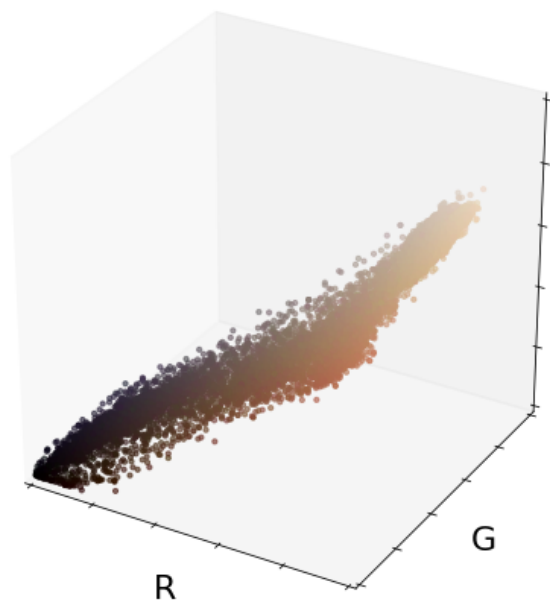


y_1, \dots, y_n

B

R

G

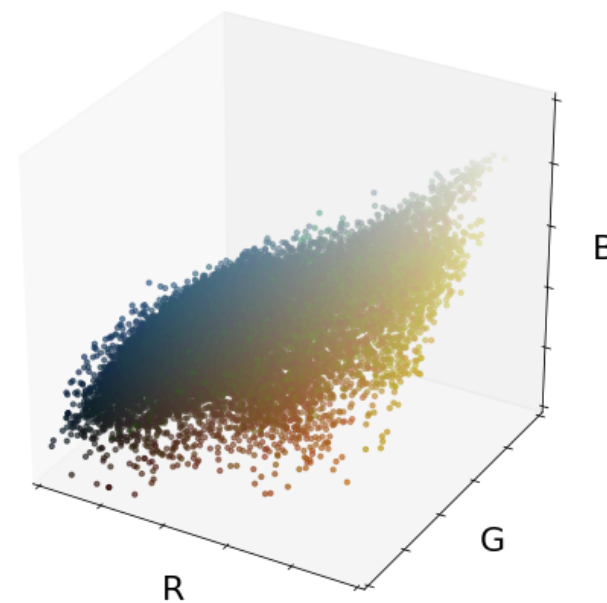


x_1, \dots, x_n

B

Matching

$\sigma \in \mathfrak{S}_n$

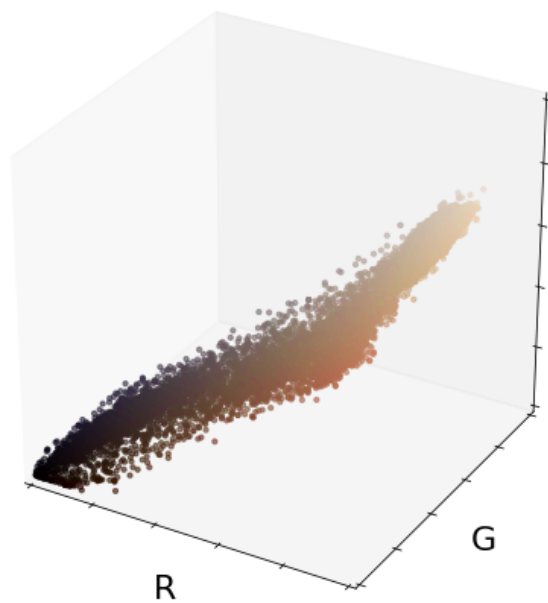


y_1, \dots, y_n

B

R

G

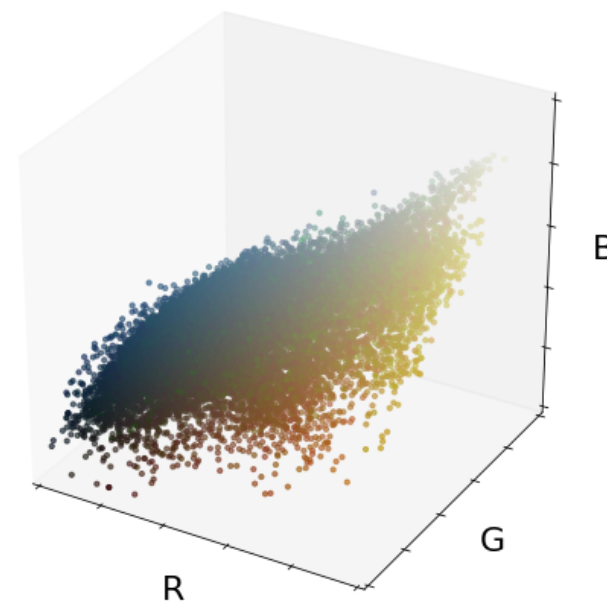


x_1, \dots, x_n

B

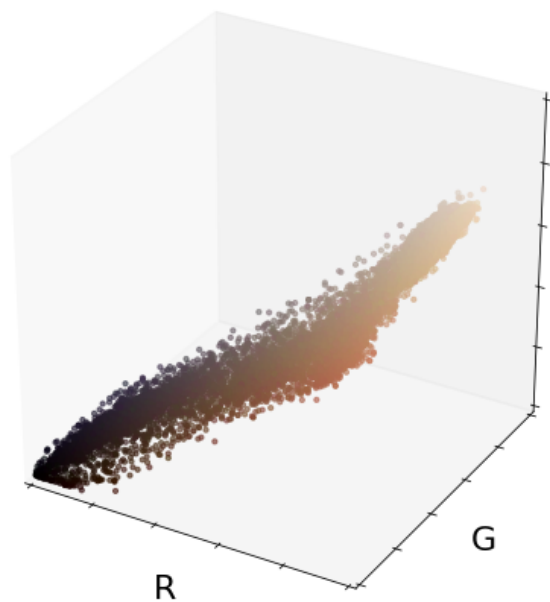
Matching

$\sigma \in \mathfrak{S}_n$



y_1, \dots, y_n

$$\|x_i - y_{\sigma(i)}\|^2$$

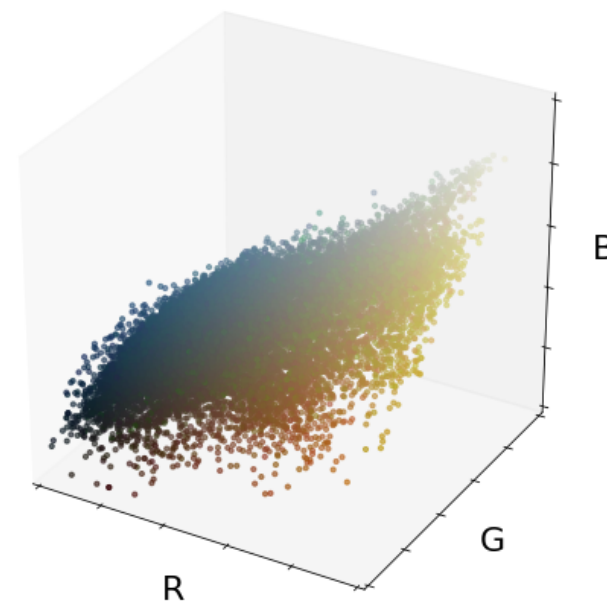


x_1, \dots, x_n

B

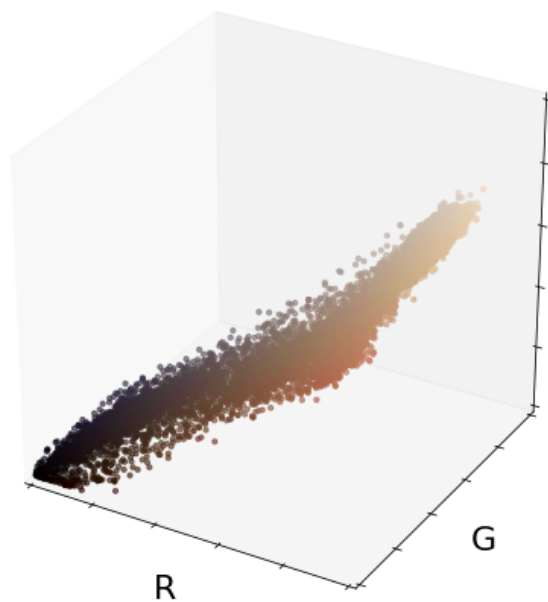
Matching

$\sigma \in \mathfrak{S}_n$



y_1, \dots, y_n

$$\sum_{i=1}^n \|x_i - y_{\sigma(i)}\|^2$$

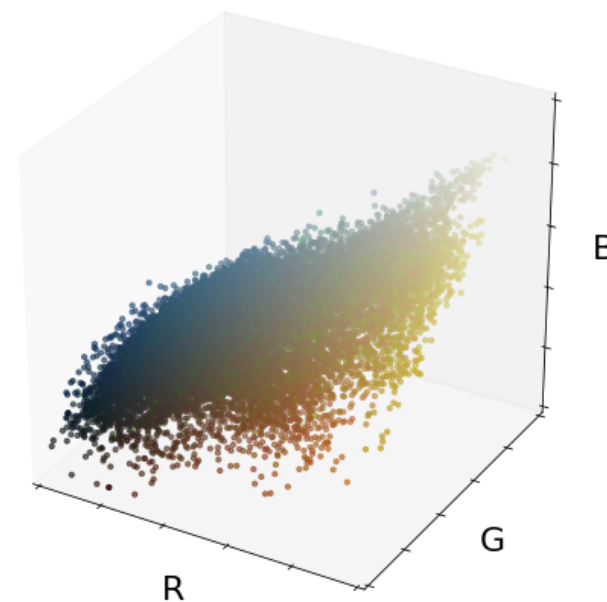


x_1, \dots, x_n

B

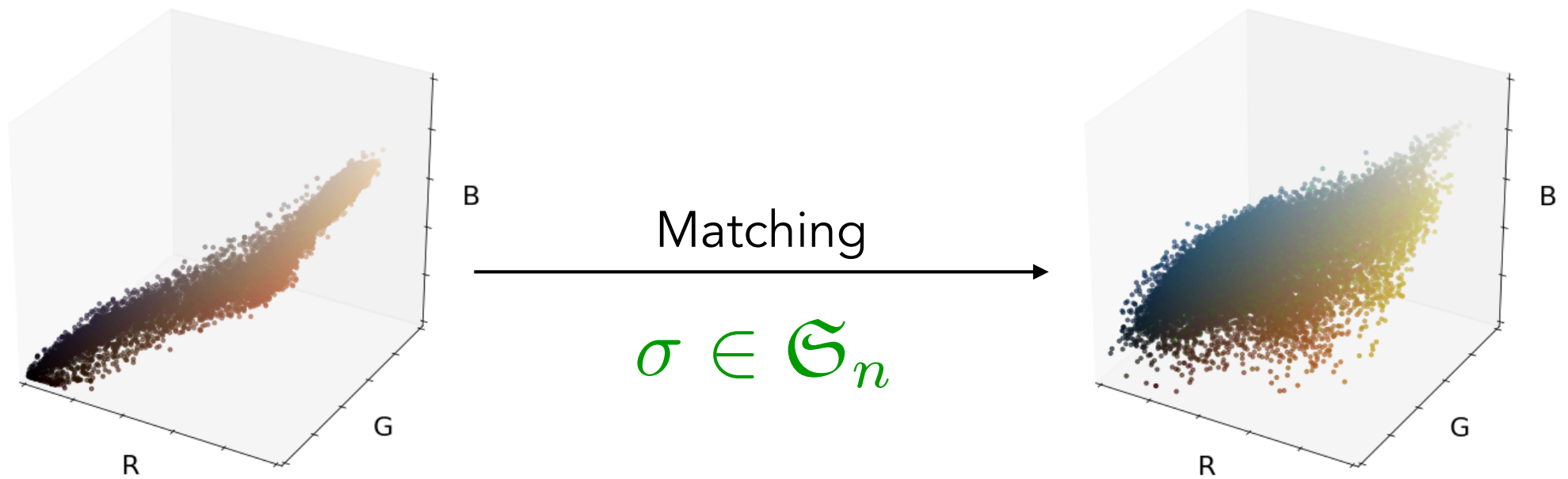
Matching

$\sigma \in \mathfrak{S}_n$



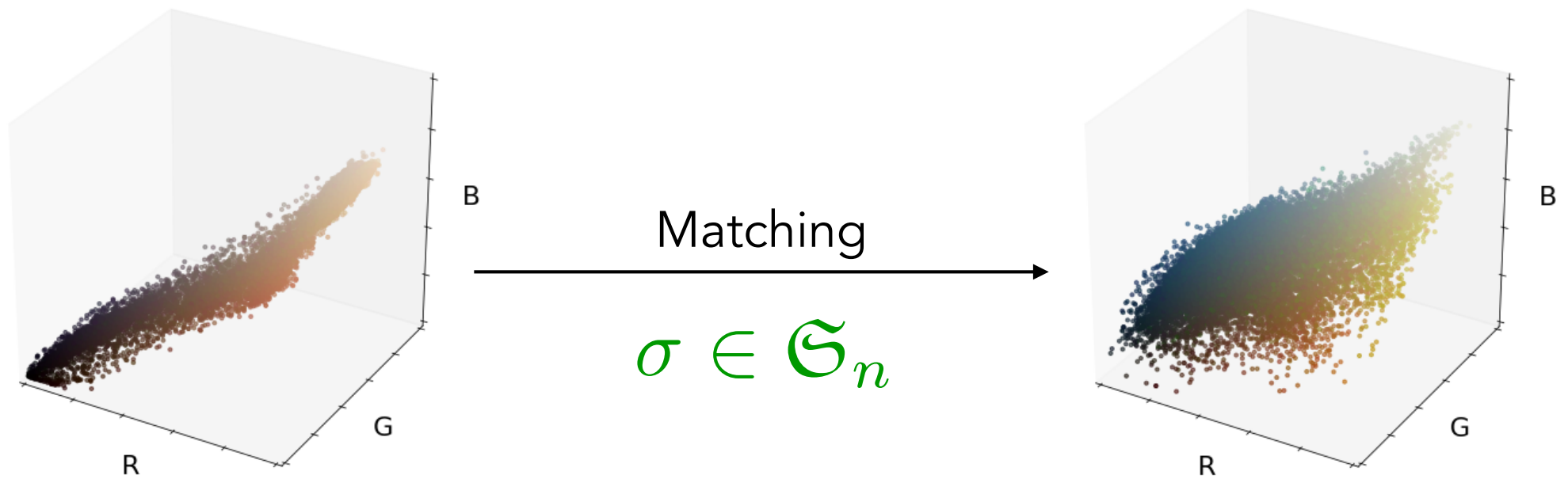
y_1, \dots, y_n

$$\min_{\sigma \in \mathfrak{S}_n} \sum_{i=1}^n \|x_i - y_{\sigma(i)}\|^2$$



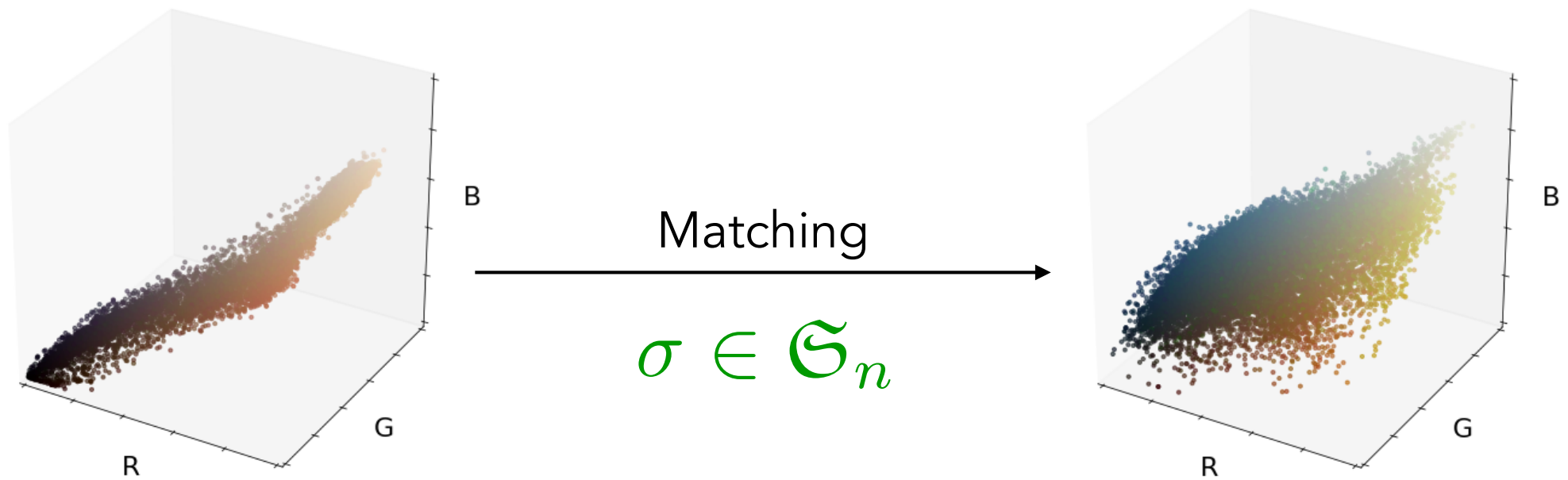
Discrete Monge Problem (1781)

$$\min_{\sigma \in \mathfrak{S}_n} \sum_{i=1}^n \|x_i - y_{\sigma(i)}\|^2$$



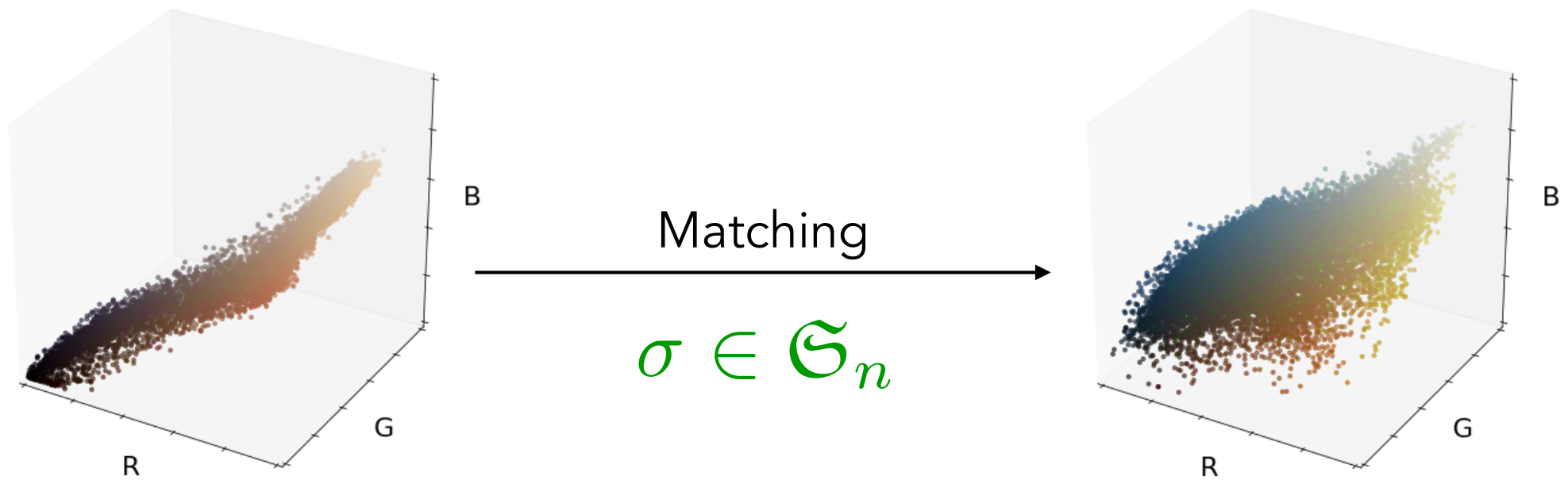
Discrete Monge Problem (1781)

$$\min_{\sigma \in \mathfrak{S}_n} \sum_{i=1}^n \|x_i - y_{\sigma(i)}\|$$



Discrete Monge Problem (1781)

$$\min_{\sigma \in \mathfrak{S}_n} \sum_{i=1}^n \|x_i - y_{\sigma(i)}\|^2$$



Discrete Monge Problem (1781)

$$\min_{\sigma \in \mathfrak{S}_n} \sum_{i=1}^n \|x_i - y_{\sigma(i)}\|^2$$

Discrete Monge Problem (1781)

$$\min_{\sigma \in \mathfrak{S}_n} \sum_{i=1}^n \|x_i - y_{\sigma(i)}\|^2$$

- (i) How to handle repeated points ?
- (ii) How to handle different numbers of points ?
- (iii) How to compute this combinatorial problem ?

A portrait of Leonid Kantorovich, a middle-aged man with glasses, wearing a grey pinstripe suit, a green tie, and a plaid shirt. He is holding a newspaper with the word "Options" visible on the page. The background is a wood-paneled wall.

OPTIMAL TRANSPORT

Leonid Kantorovich

$$\min_{\sigma \in \mathfrak{S}_n} \sum_{i=1}^n \| \textcolor{red}{x}_i - \textcolor{blue}{y}_{\textcolor{green}{\sigma}(i)} \|^2$$

$$\min_{\sigma \in \mathfrak{S}_n} \sum_{i=1}^n \sum_{j=1}^n \|x_i - y_j\|^2 \mathbb{1}_{\sigma(i)=j}$$

$$\min_{P \in \mathfrak{P}_n} \sum_{i=1}^n \sum_{j=1}^n \| \textcolor{red}{x}_i - \textcolor{blue}{y}_j \|^2 \textcolor{green}{P}_{ij}$$

$$\mathfrak{P}_n = \{ \textcolor{green}{P} \in \mathbb{R}^{n \times n} \text{ permutation matrix} \}$$

$$\min_{P \in \mathfrak{P}_n} \sum_{i=1}^n \sum_{j=1}^n \| \textcolor{red}{x}_i - \textcolor{blue}{y}_j \|^2 \textcolor{green}{P}_{ij}$$

$$\mathfrak{P}_n = \{ \textcolor{green}{P} \in \mathbb{R}^{n \times n} \text{ permutation matrix} \}$$

We only have to convexify and generalize \mathfrak{P}_n .

$$\min_{P \in \mathfrak{P}_n} \sum_{i=1}^n \sum_{j=1}^n \| \mathbf{x}_i - \mathbf{y}_j \|^2 P_{ij}$$

$$\mathfrak{P}_n = \{ P \in \mathbb{R}^{n \times n} \text{ permutation matrix} \}$$

We only have to convexify and generalize \mathfrak{P}_n .

If $\mathbf{a} \in \mathbb{R}^n$, $\mathbf{b} \in \mathbb{R}^m$ are probability weights, we define the associated **transportation polytope**:

$$\min_{P \in \mathfrak{P}_n} \sum_{i=1}^n \sum_{j=1}^n \| \mathbf{x}_i - \mathbf{y}_j \|^2 P_{ij}$$

$$\mathfrak{P}_n = \{ P \in \mathbb{R}^{n \times n} \text{ permutation matrix} \}$$

We only have to convexify and generalize \mathfrak{P}_n .

If $\mathbf{a} \in \mathbb{R}^n$, $\mathbf{b} \in \mathbb{R}^m$ are probability weights, we define the associated **transportation polytope**:

$$\mathcal{U}(\mathbf{a}, \mathbf{b}) = \{ P \in \mathbb{R}_+^{n \times m} \mid P \mathbb{1}_m = \mathbf{a}, P^\top \mathbb{1}_n = \mathbf{b} \}$$

Discrete Kantorovitch Problem

$$W_2^2(\mu, \nu) = \min_{P \in \mathcal{U}(\mathbf{a}, \mathbf{b})} \sum_{i=1}^n \sum_{j=1}^m \|x_i - y_j\|^2 P_{ij}$$

where $\mu = \sum_{i=1}^n \mathbf{a}_i \delta_{x_i}$ and $\nu = \sum_{j=1}^m \mathbf{b}_j \delta_{y_j}$ are probability measures

2-Wasserstein distance

If $\mathbf{a} \in \mathbb{R}^n$, $\mathbf{b} \in \mathbb{R}^m$ are probability weights, we define the associated **transportation polytope**:

$$\mathcal{U}(\mathbf{a}, \mathbf{b}) = \left\{ P \in \mathbb{R}_+^{n \times m} \mid P \mathbb{1}_m = \mathbf{a}, P^\top \mathbb{1}_n = \mathbf{b} \right\}$$

**Why
should we
care?**

Many applications in Machine Learning, some related to Astrophysics:

- Brenier et al., Reconstruction of the early Universe as a convex optimization problem 1999
- Wasserstein Dictionary Learning
- Computer Graphics
- Generative Models
- Model fitting (Minimum Kantorovich Estimators)

In practice, one color should be mapped to exactly one color. In other words, we want to find a map

$$T : \mathbb{R}^d \rightarrow \mathbb{R}^d$$

that is optimal in some sense.



In practice, one color should be mapped to exactly one color. In other words, we want to find a map

$$T : \mathbb{R}^d \rightarrow \mathbb{R}^d$$

that is optimal in some sense.



Monge problem

$$\min_{\sigma \in \mathfrak{S}_n} \sum_{i=1}^n \|x_i - y_{\sigma(i)}\|^2$$

Monge problem


$$\min_{\sigma \in \mathfrak{S}_n} \sum_{i=1}^n \|x_i - y_{\sigma(i)}\|^2$$

$$\inf_{T_{\#} \mu = \nu} \int \|x - T(x)\|^2 d\mu(x)$$

Monge problem

$$\min_{\sigma \in \mathfrak{S}_n} \sum_{i=1}^n \|x_i - y_{\sigma(i)}\|^2$$

$$\inf_{T_{\#}\mu=\nu} \int \|x - T(x)\|^2 d\mu(x)$$


$$X \sim \mu \implies T(X) \sim \nu$$

A man with short brown hair and glasses, wearing a light blue button-down shirt and khaki pants, stands in a library. He is holding an open book with both hands, looking at it intently. The background is filled with wooden bookshelves packed with books, mostly with brown spines. The lighting is warm and focused on the man.

REGULARITY THEORY

Let μ and ν be two probability measures over \mathbb{R}^d .

$$\inf_{T_{\#}\mu=\nu} \int \|x - T(x)\|^2 d\mu(x)$$

When does the Monge problem admit a solution ?
What can be said about it ?

Let μ and ν be two probability measures over \mathbb{R}^d .

$$\inf_{T_{\#}\mu=\nu} \int \|x - T(x)\|^2 d\mu(x)$$

Brenier Theorem

1. If μ is *absolutely continuous* with respect to the Lebesgue measure, the Monge problem admits a unique solution
2. If the Monge problem admits a solution T , then there exists a convex function f , called a **Brenier potential**, s.t.

$$T = \nabla f$$

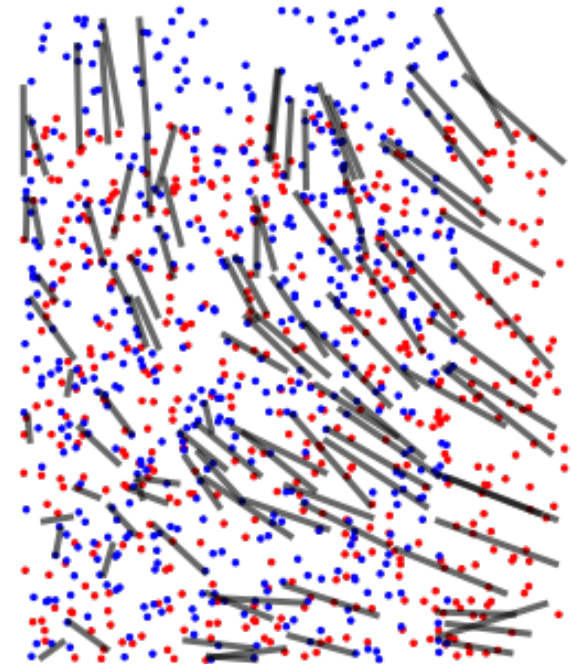
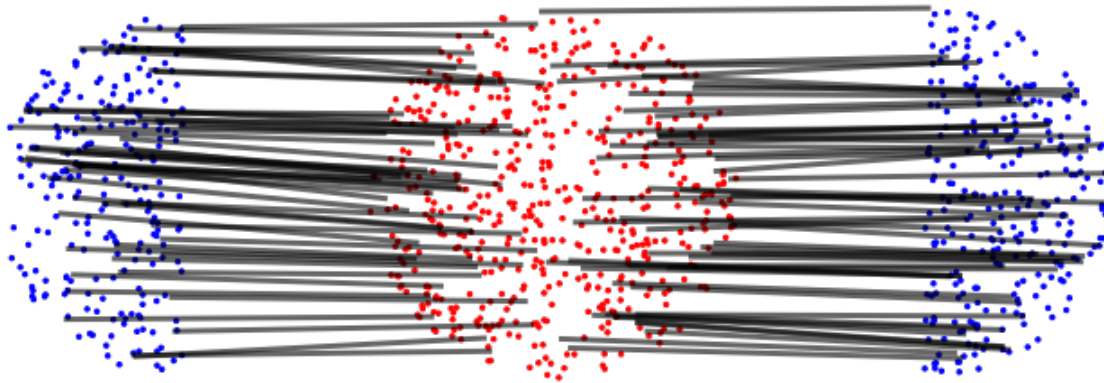
When the optimal map exists (e.g. when μ has a density), what kind of regularity does it exhibit ?

When the optimal map exists (e.g. when μ has a density), what kind of regularity does it exhibit ?

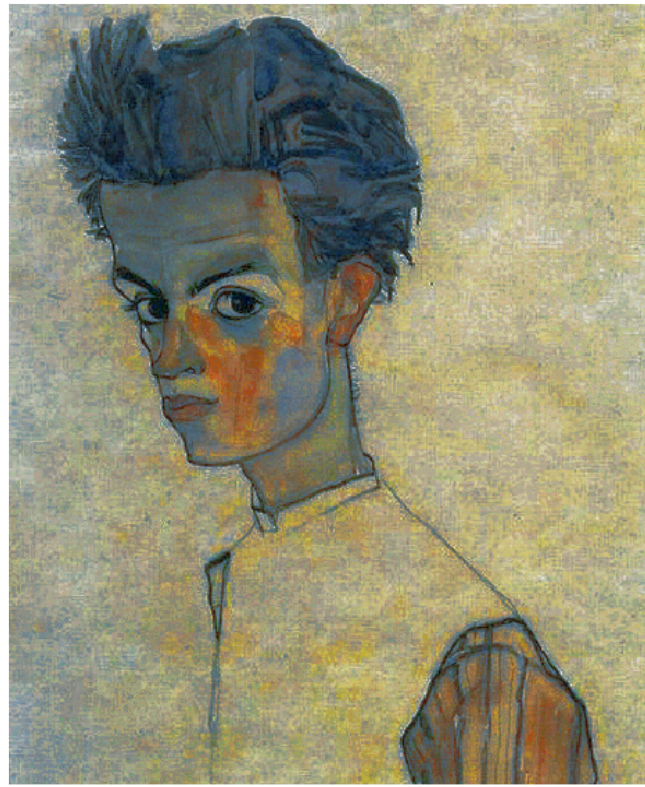
Without further assumptions on μ and ν , we cannot even hope for continuity. Many results by *Caffarelli, De Philippis, Kim, Figalli...*

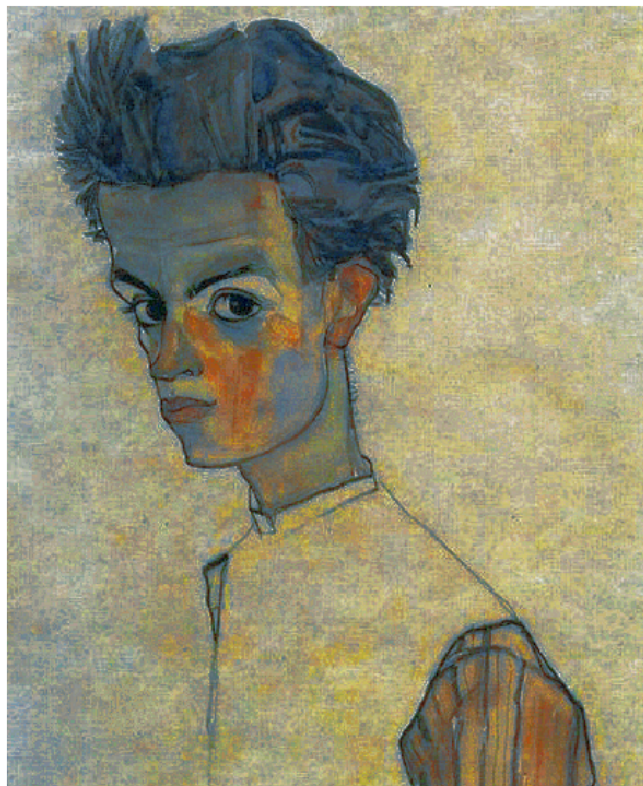
When the optimal map exists (e.g. when μ has a density), what kind of regularity does it exhibit ?

Without further assumptions on μ and ν , we cannot even hope for continuity. Many results by *Caffarelli*, *De Philippis*, *Kim*, *Figalli*...





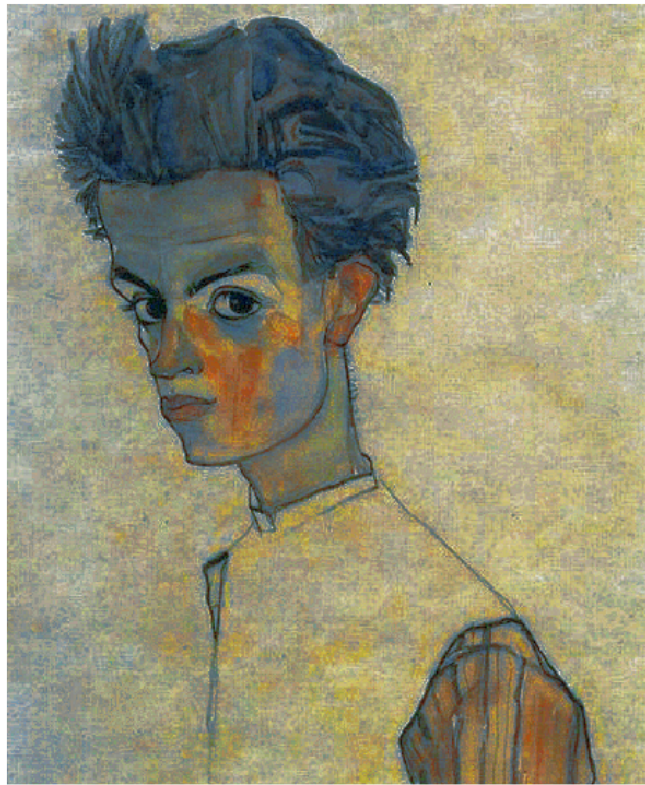


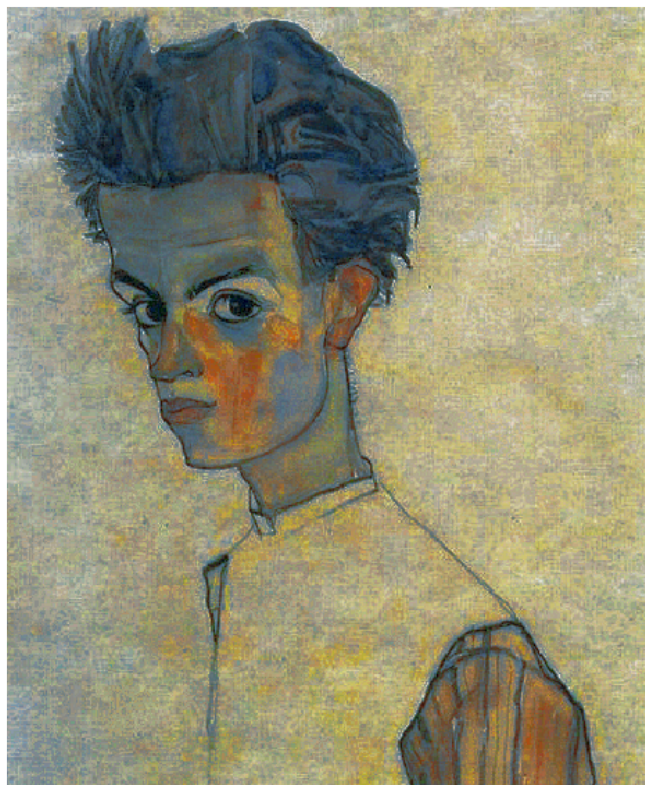


Instead of finding assumptions under which the optimal map exists and exhibits some regularity, we will enforce such regularity directly in the OT problem.

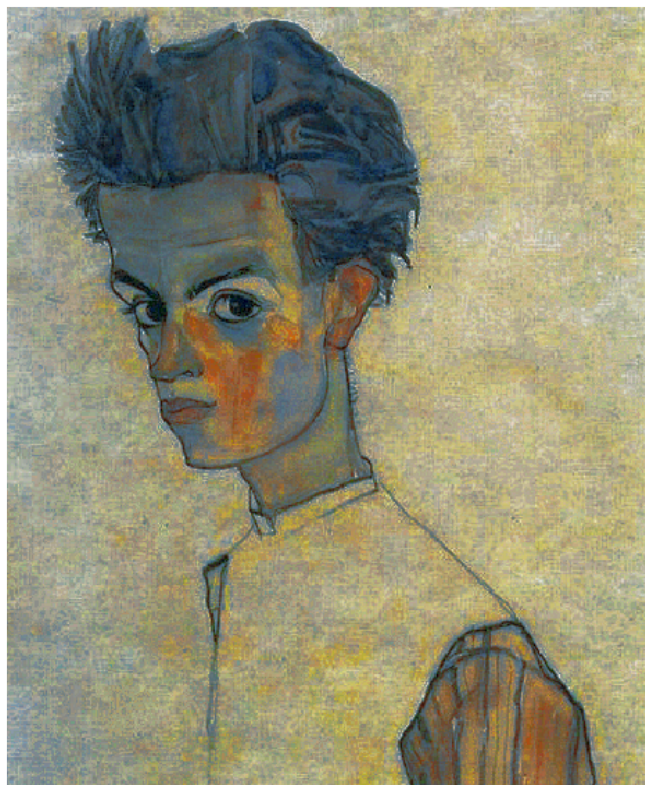
A photograph of a man standing on a sidewalk in front of a traditional Japanese building with a tiled roof. The man is wearing a light blue shirt and dark trousers, and is holding a white shopping bag. The building has a sign that reads '聖護院' (Seigo-in) and 'ハツ橋' (Hatsubashi). The text 'SMOOTH AND STRONGLY CONVEX BRENIER POTENTIALS' is overlaid in large white letters.

SMOOTH AND STRONGLY CONVEX BRENIER POTENTIALS



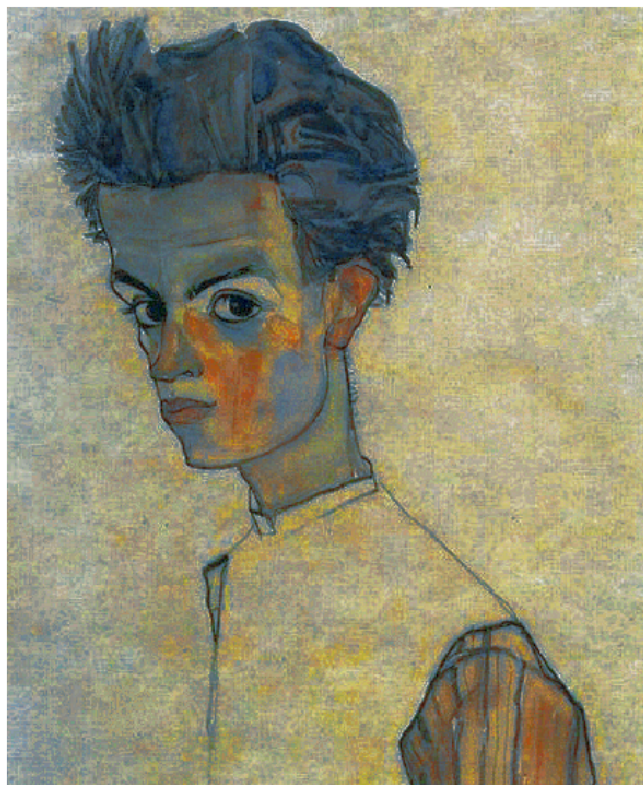


$$\ell \|x - y\| \leq \|\nabla f(x) - \nabla f(y)\| \leq L \|x - y\|$$



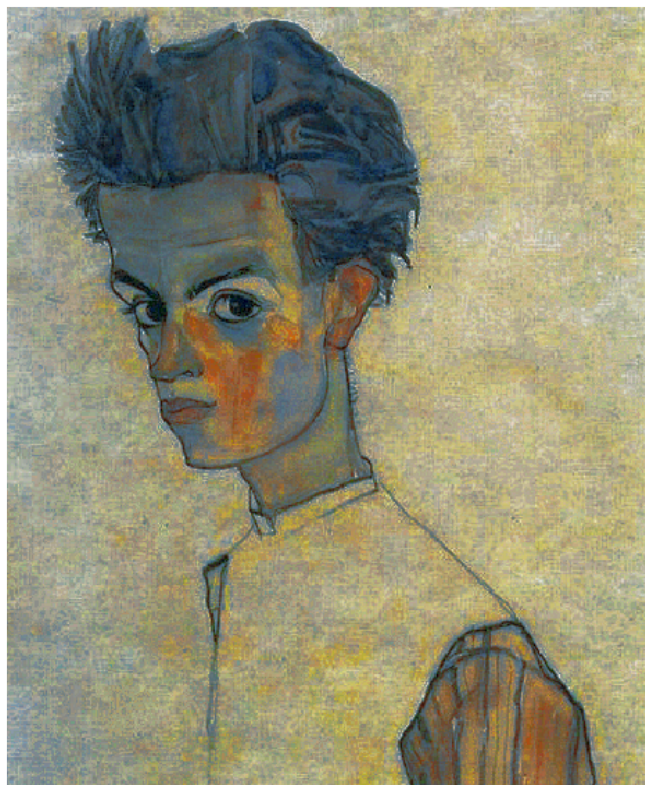
$$\ell \|x - y\| \leq \|\nabla f(x) - \nabla f(y)\| \leq L \|x - y\|$$

We ask that $T = \nabla f$ is a bi-Lipschitz map



$$\ell \|x - y\| \leq \|\nabla f(x) - \nabla f(y)\| \leq L \|x - y\|$$

We ask that f is **smooth** and **strongly convex**



$$\ell \|x - y\| \leq \|\nabla f(x) - \nabla f(y)\| \leq L \|x - y\|$$

We ask that f is **smooth** and **strongly convex**

$$\hookrightarrow f \in \mathcal{F}_{\ell, L}$$

But there may not even such a regular f that is admissible for the Monge problem, *i.e.* such that $(\nabla f)_\# \mu = \nu$.

But there may not even such a regular f that is admissible for the Monge problem, i.e. such that $(\nabla f)_\# \mu = \nu$.

Instead, we will try to best approximate ν as a push-forward of μ through a regular map:

$$\min_{f \in \mathcal{F}_{\ell, L}} W_2 [\nabla f_\# \mu, \nu]$$

But there may not even such a regular f that is admissible for the Monge problem, i.e. such that $(\nabla f)_\# \mu = \nu$.

Instead, we will try to best approximate ν as a push-forward of μ through a regular map:

$$\min_{f \in \mathcal{F}_{\ell, L}} W_2 [\nabla f_\# \mu, \nu]$$

Smooth and Strong Convex
Brenier Potentials

Even when the measures are discrete, this is a *infinite dimensional* optimization problem !

$$\min_{f \in \mathcal{F}_{\ell, L}} W_2 [\nabla f_{\#} \mu, \nu]$$


Even when the measures are discrete, this is a *infinite dimensional* optimization problem !

$$\min_{f \in \mathcal{F}_{\ell, L}} W_2 [\nabla f_{\#} \mu, \nu]$$




Even when the measures are discrete, this is a *infinite dimensional* optimization problem !

$$\min_{f \in \mathcal{F}_{\ell, L}} W_2 [\nabla f_{\#} \mu, \nu]$$

$$\min_{\substack{z_1, \dots, z_n \in \mathbb{R}^d \\ u \in \mathbb{R}^n}} W_2^2 \left(\sum_{i=1}^n a_i \delta_{z_i}, \nu \right)$$


Even when the measures are discrete, this is a *infinite dimensional* optimization problem !

$$\min_{f \in \mathcal{F}_{\ell, L}} W_2 [\nabla f_{\#} \mu, \nu]$$

$$\min_{\substack{z_1, \dots, z_n \in \mathbb{R}^d \\ u \in \mathbb{R}^n}} W_2^2 \left(\sum_{i=1}^n a_i \delta_{z_i}, \nu \right)$$


$$u_i \geq u_j + \langle z_j, x_i - x_j \rangle + \frac{1}{2(1 - \ell/L)} \left(\frac{1}{L} \|z_i - z_j\|^2 + \ell \|x_i - x_j\|^2 - 2 \frac{\ell}{L} \langle z_j - z_i, x_j - x_i \rangle \right)$$

$$x_1, \dots, x_n \sim \mu$$

$$\hat{\mu}_n = \frac{1}{n} \sum_{i=1}^n \delta_{x_i}$$

$$y_1, \dots, y_n \sim \nu$$

$$\hat{\nu}_n = \frac{1}{n} \sum_{i=1}^n \delta_{y_i}$$

$$x_1, \dots, x_n \sim \mu$$

$$y_1, \dots, y_n \sim \nu$$

$$\hat{\mu}_n = \frac{1}{n} \sum_{i=1}^n \delta_{x_i}$$

$$\hat{\nu}_n = \frac{1}{n} \sum_{i=1}^n \delta_{y_i}$$

$$f^* \in \arg \min_{f \in \mathcal{F}_{\ell, L}} W_2 [\nabla f_{\#} \hat{\mu}_n, \hat{\nu}_n]$$

$$x_1, \dots, x_n \sim \mu$$

$$y_1, \dots, y_n \sim \nu$$

$$\hat{\mu}_n = \frac{1}{n} \sum_{i=1}^n \delta_{x_i}$$

$$\hat{\nu}_n = \frac{1}{n} \sum_{i=1}^n \delta_{y_i}$$

$$f^* \in \arg \min_{f \in \mathcal{F}_{\ell, L}} W_2 [\nabla f_{\#} \hat{\mu}_n, \hat{\nu}_n]$$

$$z_1^*, \dots, z_n^*, u^*$$

$$x_1, \dots, x_n \sim \mu$$

$$y_1, \dots, y_n \sim \nu$$

$$\hat{\mu}_n = \frac{1}{n} \sum_{i=1}^n \delta_{x_i}$$

$$\hat{\nu}_n = \frac{1}{n} \sum_{i=1}^n \delta_{y_i}$$

$$f^* \in \arg \min_{f \in \mathcal{F}_{\ell, L}} W_2 [\nabla f_{\#} \hat{\mu}_n, \hat{\nu}_n]$$

$$z_1^*, \dots, z_n^*, u^*$$

We can easily compute the map on any new point x by solving a cheap QCQP

$$x_1, \dots, x_n \sim \mu$$

$$y_1, \dots, y_n \sim \nu$$

$$\hat{\mu}_n = \frac{1}{n} \sum_{i=1}^n \delta_{x_i}$$

$$\hat{\nu}_n = \frac{1}{n} \sum_{i=1}^n \delta_{y_i}$$

$$f^* \in \arg \min_{f \in \mathcal{F}_{\ell, L}} W_2 [\nabla f_{\#} \hat{\mu}_n, \hat{\nu}_n]$$

$$z_1^*, \dots, z_n^*, u^*$$

We can easily compute the map on any new point x by solving a cheap QCQP

$$\min_{v \in \mathbb{R}, g \in \mathbb{R}^d} v$$

$$\text{s.t. } \forall i, v \geq u_i + \langle z_i^*, x - x_i \rangle$$

$$+ \frac{1}{2(1 - \ell/L)} \left(\frac{1}{L} \|g - z_i^*\|^2 + \ell \|x - x_i\|^2 - 2 \frac{\ell}{L} \langle z_i^* - g, x_i - x \rangle \right)$$

$$x_1, \dots, x_n \sim \mu$$

$$y_1, \dots, y_n \sim \nu$$

$$\hat{\mu}_n = \frac{1}{n} \sum_{i=1}^n \delta_{x_i}$$

$$\hat{\nu}_n = \frac{1}{n} \sum_{i=1}^n \delta_{y_i}$$

$$f^* \in \arg \min_{f \in \mathcal{F}_{\ell, L}} W_2 [\nabla f_{\#} \hat{\mu}_n, \hat{\nu}_n]$$

$$z_1^*, \dots, z_n^*, u^*$$

We can easily compute the map on any new point x by solving a cheap QCQP

This defines an estimator ∇f^* of the optimal transport map sending μ to ν

$$x_1, \dots, x_n \sim \mu$$

$$y_1, \dots, y_n \sim \nu$$

$$\hat{\mu}_n = \frac{1}{n} \sum_{i=1}^n \delta_{x_i}$$

$$\hat{\nu}_n = \frac{1}{n} \sum_{i=1}^n \delta_{y_i}$$

$$f^* \in \arg \min_{f \in \mathcal{F}_{\ell, L}} W_2 [\nabla f_{\#} \hat{\mu}_n, \hat{\nu}_n]$$

$$z_1^*, \dots, z_n^*, u^*$$

We can easily compute the map on any new point x by solving a cheap QCQP

This defines an estimator ∇f^* of the optimal transport map sending μ to ν

We define the *SSNB estimator* as a plug-in:

$$x_1, \dots, x_n \sim \mu$$

$$y_1, \dots, y_n \sim \nu$$

$$\hat{\mu}_n = \frac{1}{n} \sum_{i=1}^n \delta_{x_i}$$

$$\hat{\nu}_n = \frac{1}{n} \sum_{i=1}^n \delta_{y_i}$$

$$f^* \in \arg \min_{f \in \mathcal{F}_{\ell, L}} W_2 [\nabla f_{\#} \hat{\mu}_n, \hat{\nu}_n]$$

$$z_1^*, \dots, z_n^*, u^*$$

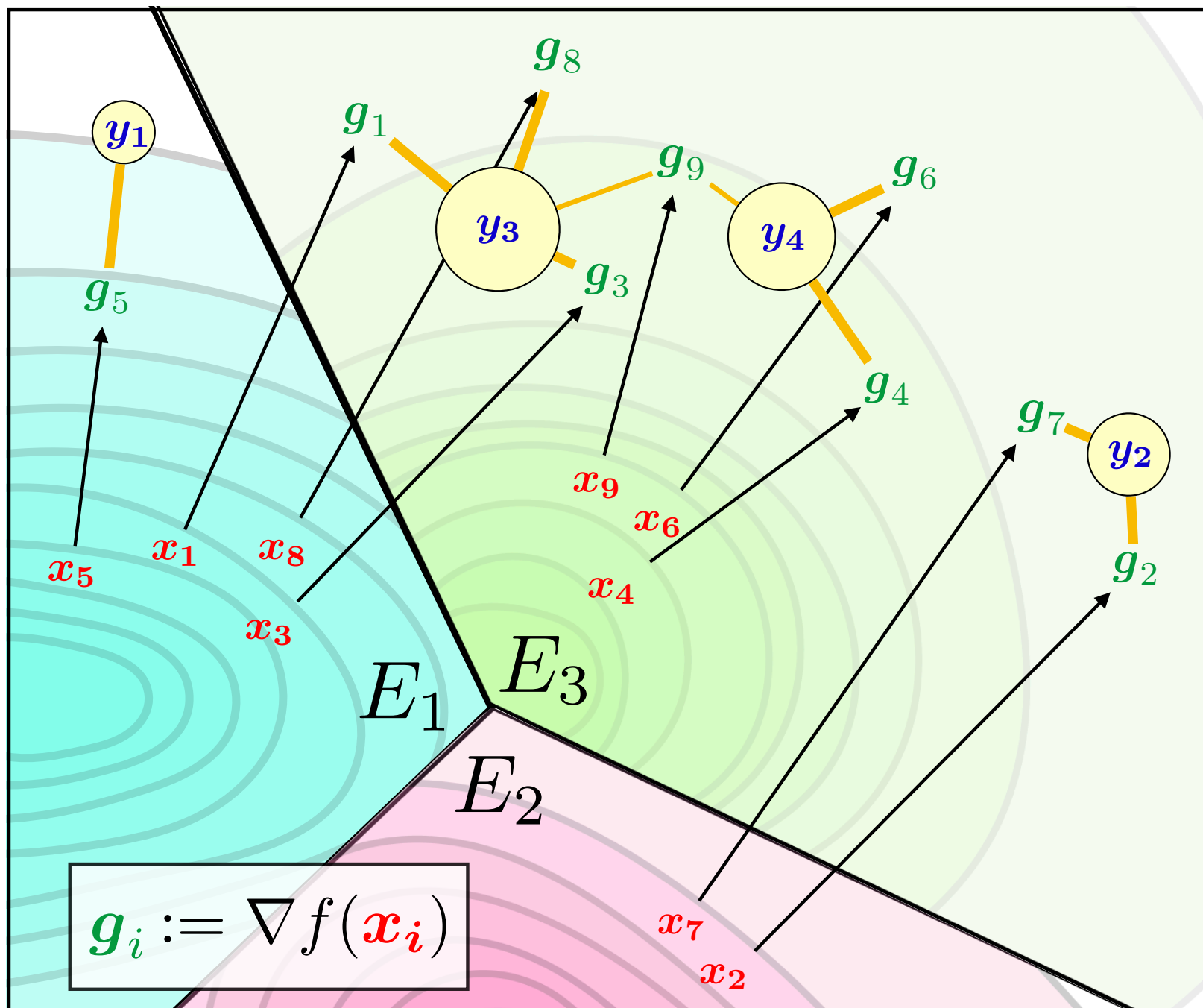
We can easily compute the map on any new point x by solving a cheap QCQP

This defines an estimator ∇f^* of the optimal transport map sending μ to ν

We define the *SSNB estimator* as a plug-in:

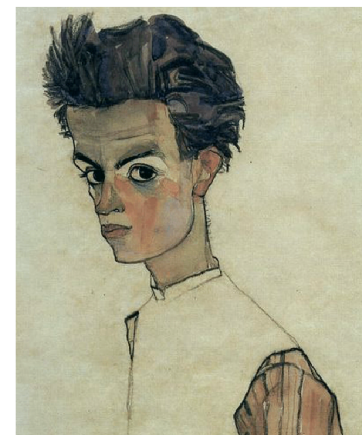
$$\widehat{W}_2^2 = \int \|x - \nabla f^*(x)\|^2 d\mu(x)$$

Regularity "by part"

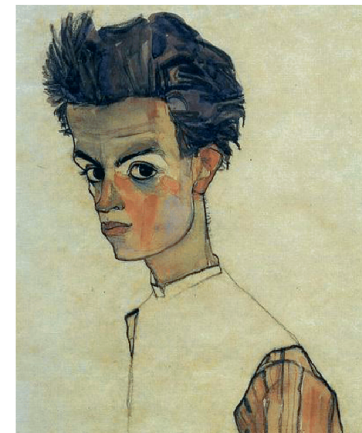
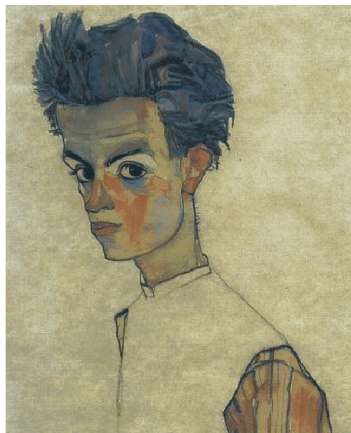




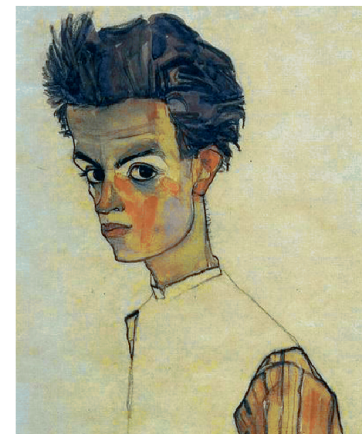
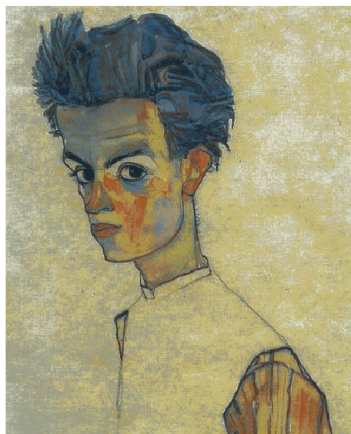
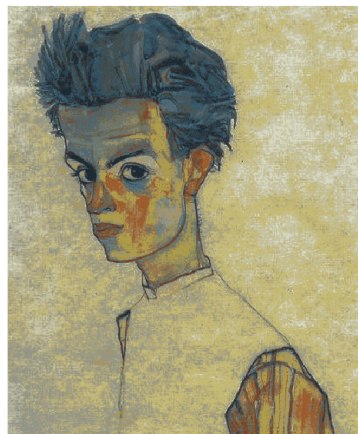
$L = 1$



$L = 2$



$L = 5$



$\ell = 0$

$\ell = 0.5$

$\ell = 1$



QUESTIONS ?