

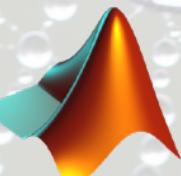
Computational Optimal Transport

<http://optimaltransport.github.io>

Introduction

Gabriel Peyré

www.numerical-tours.com



ENS

ÉCOLE NORMALE
SUPÉRIEURE

<https://optimaltransport.github.io>

Home

Computational Optimal Transport

BOOK

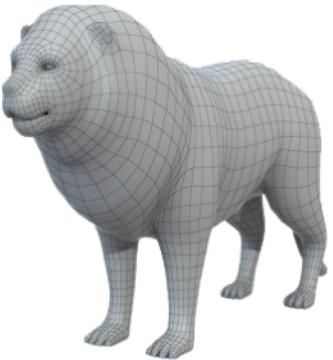
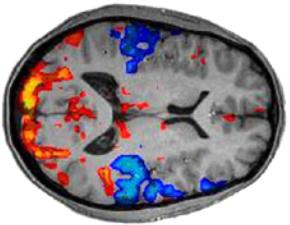
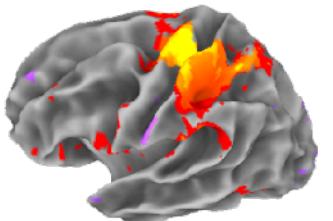
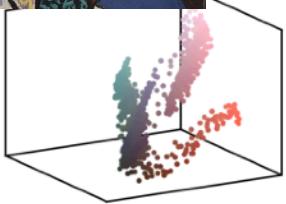
CODE

SLIDES

Probability Distributions in Data Sciences

Probability distributions and histograms

→ images, vision, graphics and machine learning, .



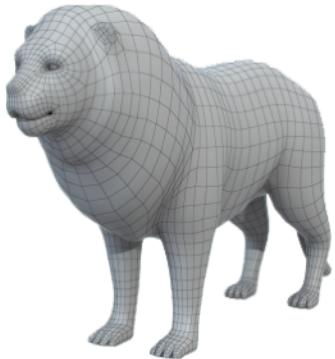
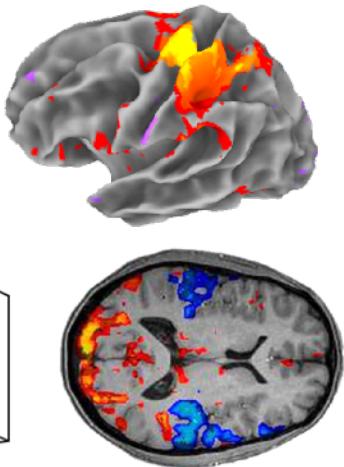
WIN GREAT TIME JUMP TITLE GRAND FINISH INTERNATIONAL
WORLD CHAMPION RACE
SET TENNIS SEASON LONG ON DAUGHTER MONTHLY
SEASON COMPETITION DOUBLE COUNTRY START FINAL UNION PLACE
GREECE PART MARK LONDON METRE YARD UNION PLACE
EUROPE GOLD RECORD EVENT YEAR
COMPETE



Probability Distributions in Data Sciences

Probability distributions and histograms

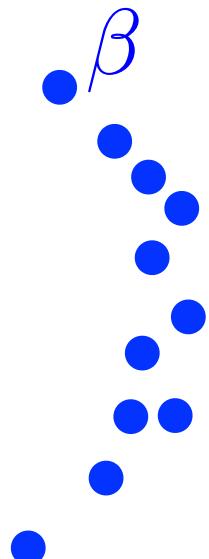
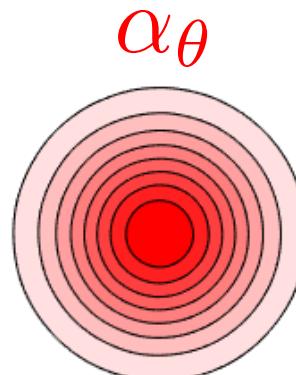
→ images, vision, graphics and machine learning, . . .



Unsupervised learning

Observations: $\beta \stackrel{\text{def.}}{=} \frac{1}{n} \sum_{i=1}^n \delta_{\mathbf{x}_i}$

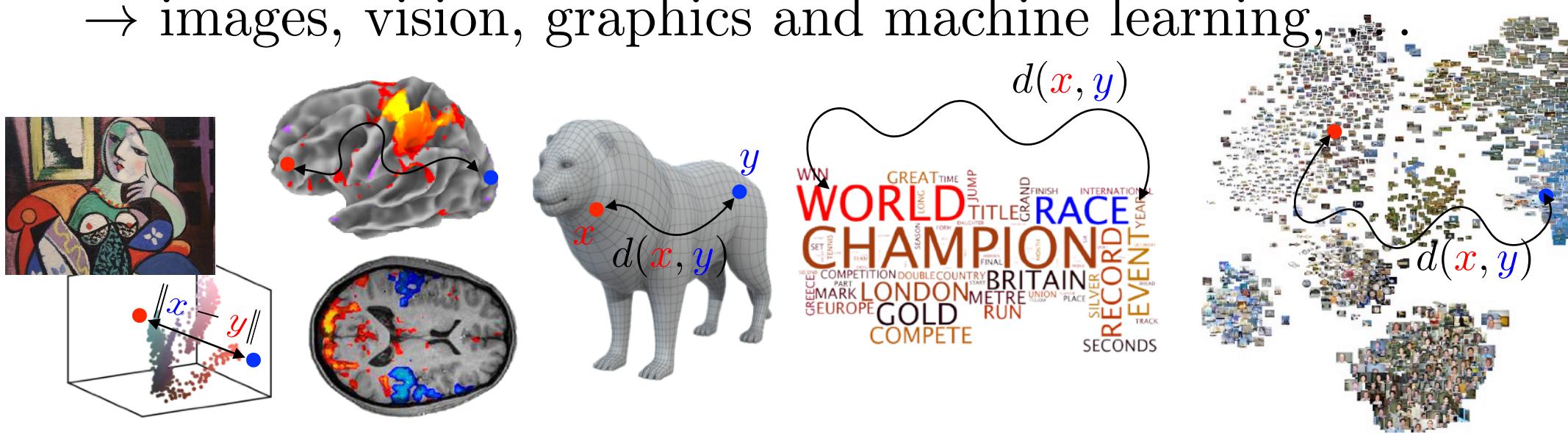
Parametric model: $\theta \mapsto \alpha_\theta$



Probability Distributions in Data Sciences

Probability distributions and histograms

→ images, vision, graphics and machine learning,

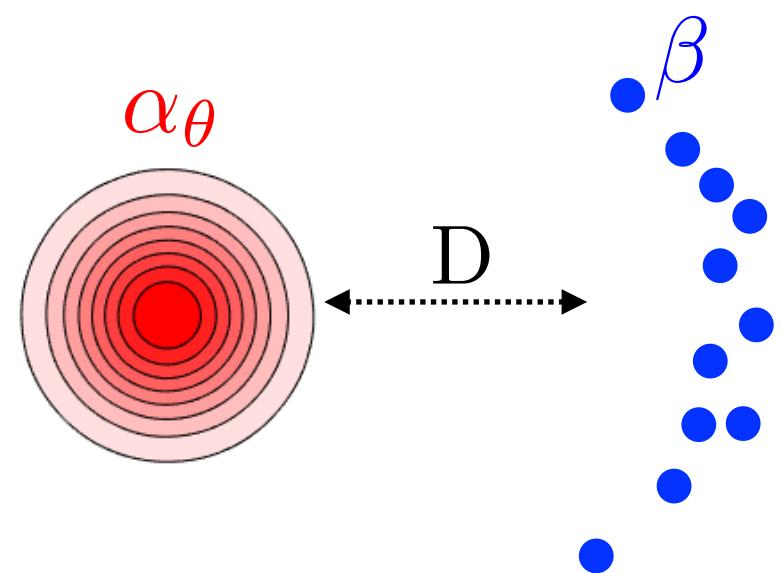


Unsupervised learning

Observations: $\beta \stackrel{\text{def.}}{=} \frac{1}{n} \sum_{i=1}^n \delta_{x_i}$

Parametric model: $\theta \mapsto \alpha_\theta$

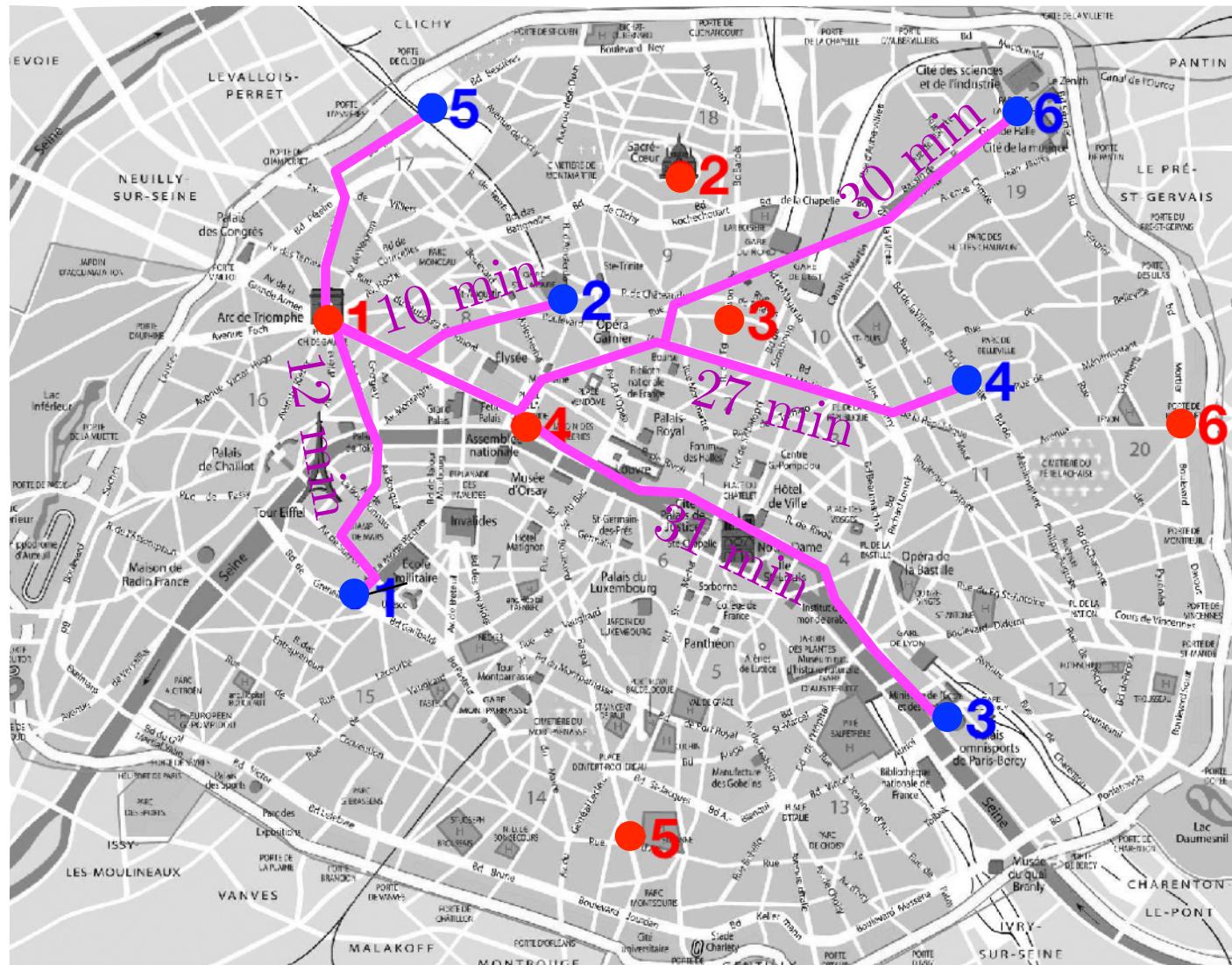
Density fitting: $\min_{\theta} D(\alpha_\theta, \beta)$
→ takes into account a metric d .



Overview

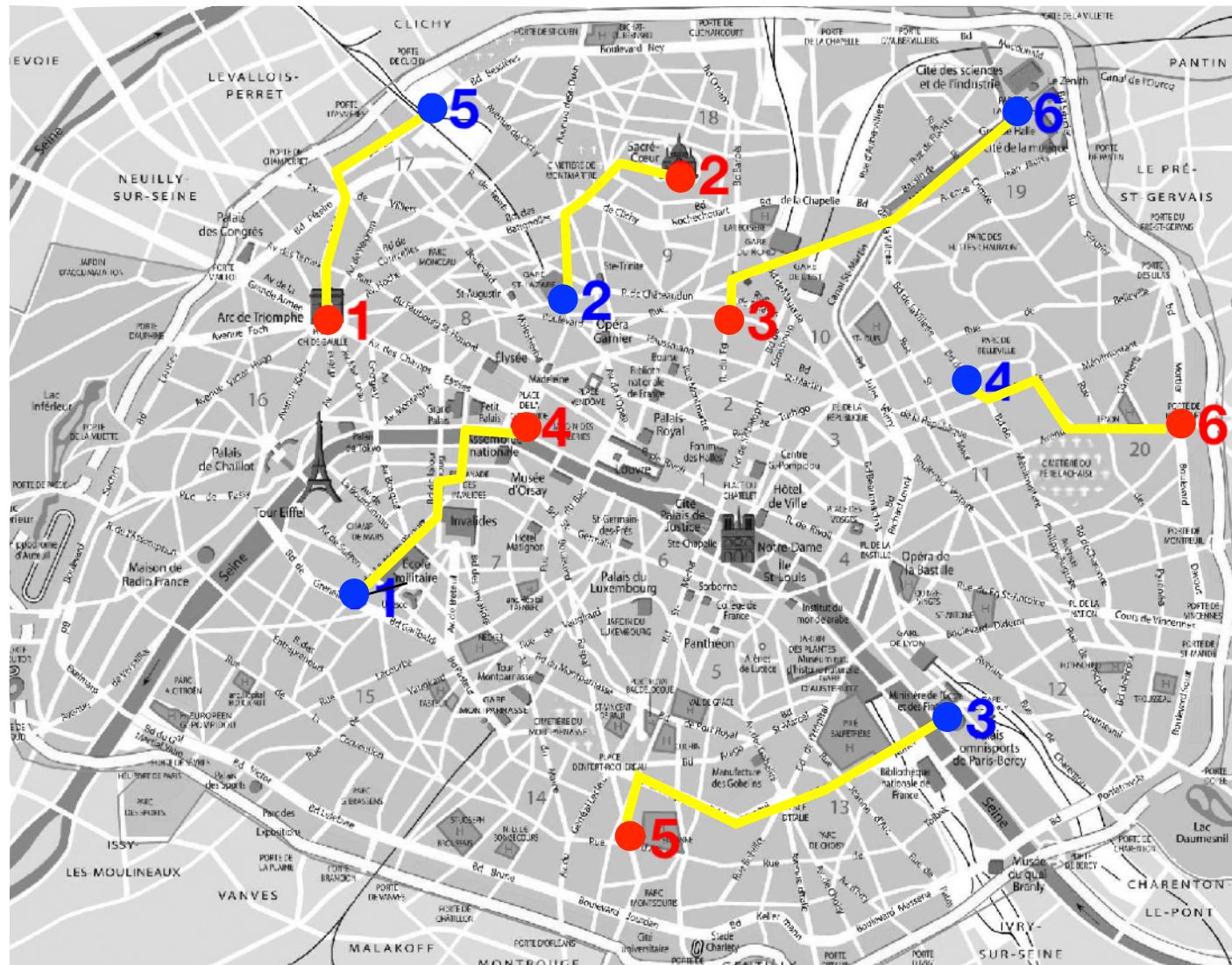
- **Monge Formulation**
- Continuous Optimal Transport
- Kantorovitch Formulation
- Applications

Fom bakeries to cafés



c_{ij}	y_1	y_2	y_3	y_4	y_5	y_6
x_1	12	10	31	27	10	30
x_2	22	7	25	15	11	14
x_3	19	7	19	10	15	15
x_4	10	6	21	19	14	24
x_5	15	23	14	24	31	34
x_6	35	26	16	9	34	15

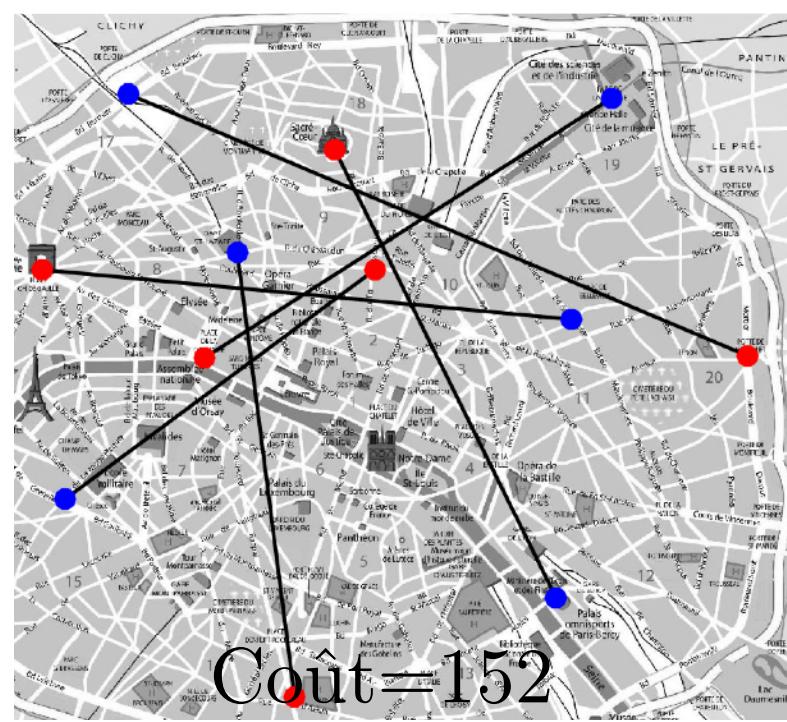
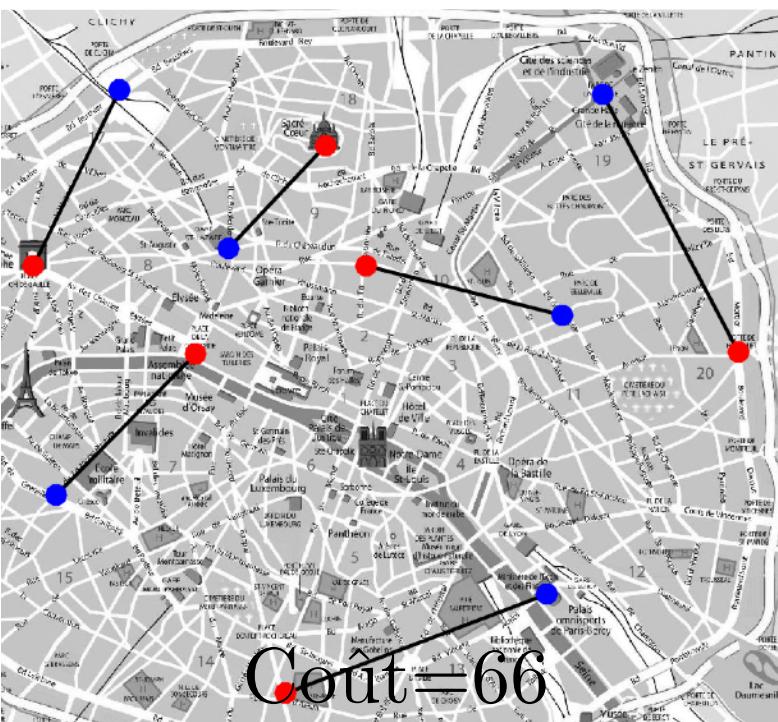
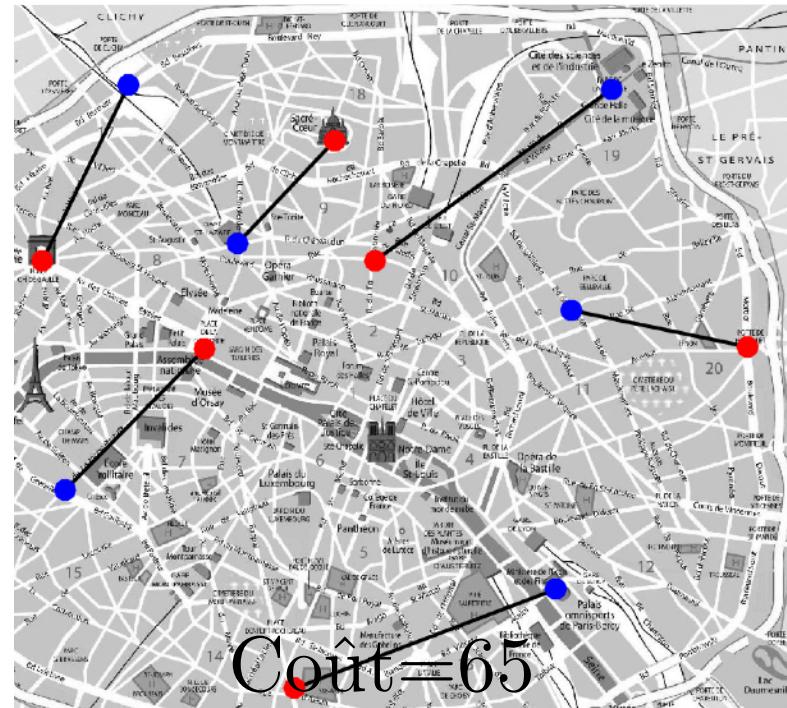
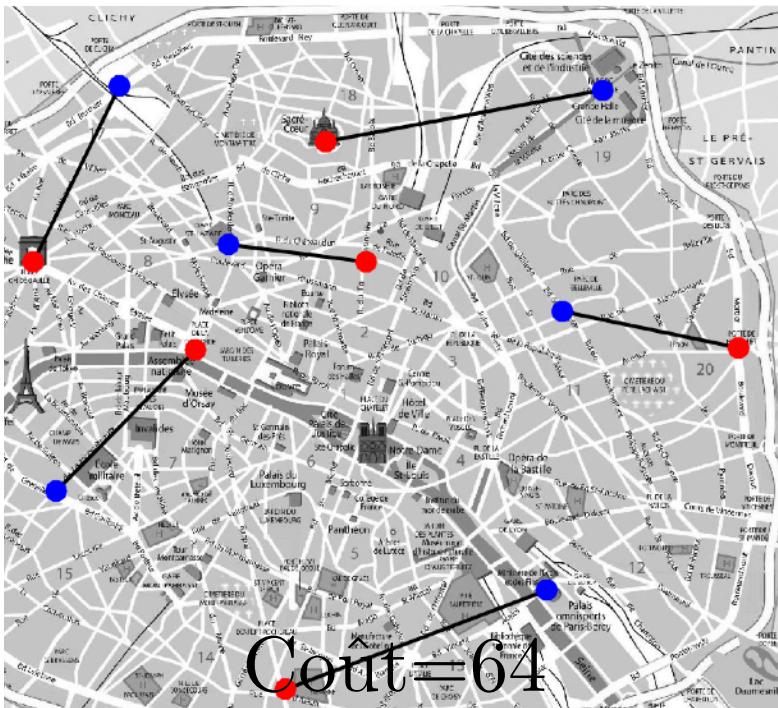
Fom bakeries to cafés



c_{ij}	y_1	y_2	y_3	y_4	y_5	y_6
x_1	12	10	31	27	10	30
x_2	22	7	25	15	11	14
x_3	19	7	19	10	15	15
x_4	10	6	21	19	14	24
x_5	15	23	14	24	31	34
x_6	35	26	16	9	34	15

Cout: $10+7+15+10+14+9 = 65$ min

From best to worst



Combinatorial Search

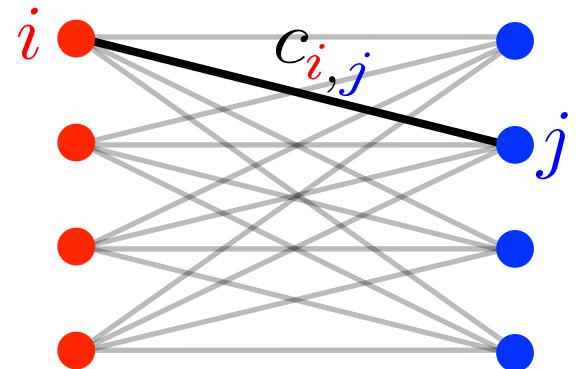
Cost $(c_{i,j})_{i,j=1}^n \in \mathbb{R}^{n \times n}$

Permutations:

$$\Sigma_n \stackrel{\text{def.}}{=} \{\sigma : \{1, \dots, n\} \rightarrow \{1, \dots, n\}\}$$

Monge's problem:

$$\min_{\sigma \in \Sigma_n} \sum_{i=1}^n c_{i,\sigma(i)}$$



Combinatorial Search

Cost $(c_{i,j})_{i,j=1}^n \in \mathbb{R}^{n \times n}$

Permutations:

$$\Sigma_n \stackrel{\text{def.}}{=} \{\sigma : \{1, \dots, n\} \rightarrow \{1, \dots, n\}\}$$

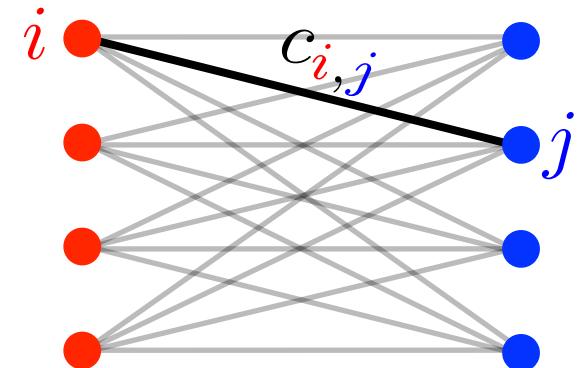
Monge's problem:

$$\min_{\sigma \in \Sigma_n} \sum_{i=1}^n c_{i,\sigma(i)}$$



n	$n !$
0	1
1	1
2	2
3	6
4	24
5	120
6	720
7	5 040
8	40 320

n	$n !$
9	362880
10	3628800
11	39916800
12	479001600
...	...
25	1,551x10 ²⁵
...	...
70	1,198x10 ¹⁰⁰



Atoms in the universe: 10^{79}



Neurons in the brain: 10^{11} .



Gaspard Monge (1746-1818)

(1784)

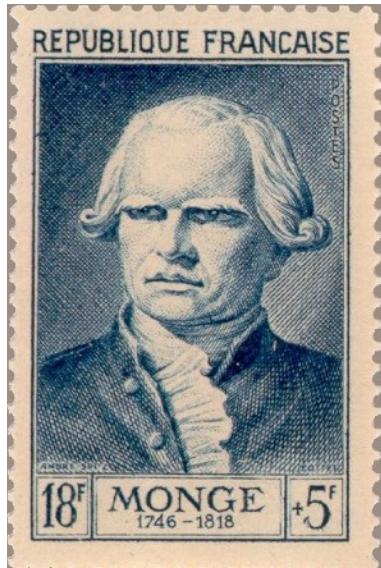
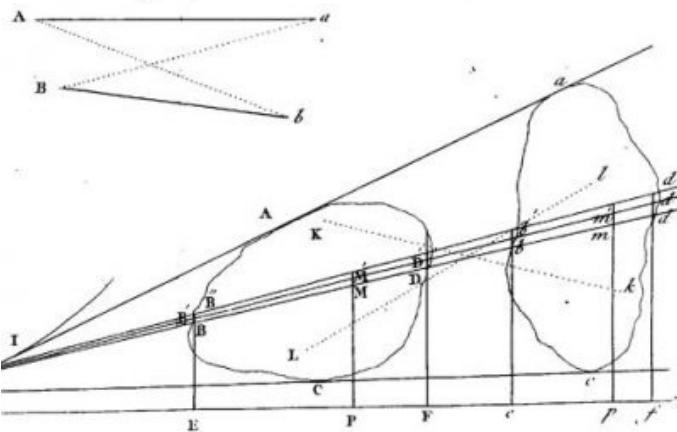
MÉMOIRE
SUR LA
THÉORIE DES DÉBLAIS
ET DES REMBLAIS.
Par M. MONGE.

Lorsqu'on doit transporter des terres d'un lieu dans un autre, on a coutume de donner le nom de *Déblai* au volume des terres que l'on doit transporter, & le nom de *Remblai* à l'espace qu'elles doivent occuper après le transport.

Le prix du transport d'une molécule étant, toutes choses d'ailleurs égales, proportionnel à son poids & à l'espace qu'on lui fait parcourir, & par conséquent le prix du transport total devant être proportionnel à la somme des produits des molécules multipliées chacune par l'espace parcouru, il s'enfuit que le déblai & le remblai étant donnés de figure & de position, il n'est pas indifférent que telle molécule du déblai soit transportée dans tel ou tel autre endroit du remblai, mais qu'il y a une certaine distribution à faire des molécules du premier dans le second, d'après laquelle la somme de ces produits sera la moindre possible, & le prix du transport total fera un *minimum*.

Mém. de l'Ac. R. des Sc. An. 1784. Page. 704. Pl. XVII.

Fig. 1.



1-D Optimal Transport



1-D Optimal Transport



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



1-D Optimal Transport



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

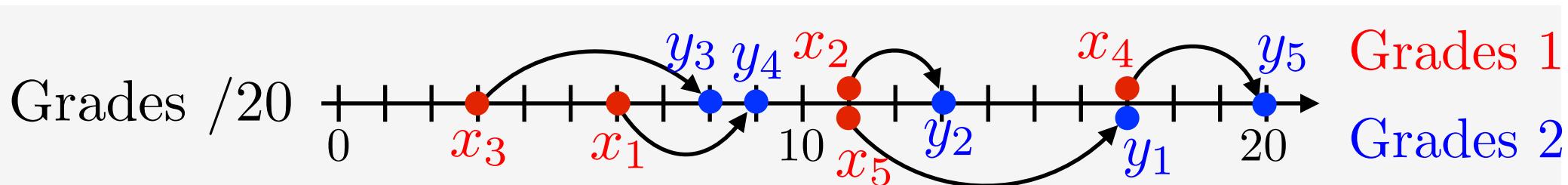


Sorting algorithms: insertion $n(n - 1)/2$ worst case.

n	$n !$	$n(n-1)/2$	$n \log(n)$
10	3628800	45	23
11	39916800	55	26
12	479001600	66	30
25	$1,551 \times 10^{25}$	300	80
70	$1,198 \times 10^{100}$	21415	297

QuickSort: $O(n \log(n))$.

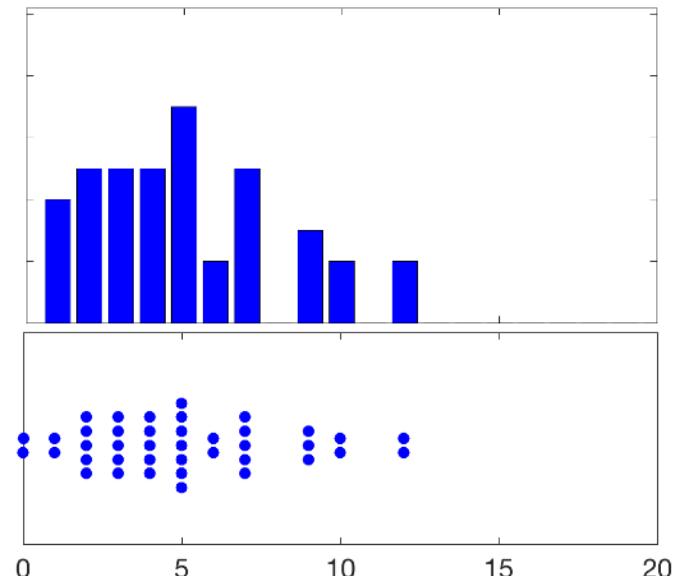
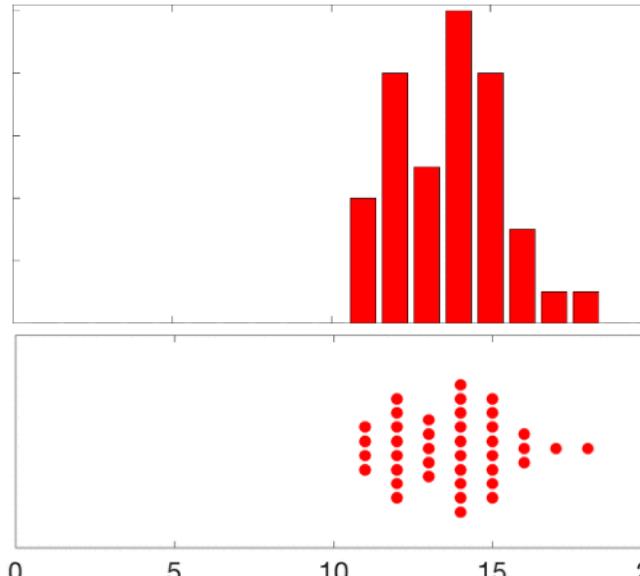
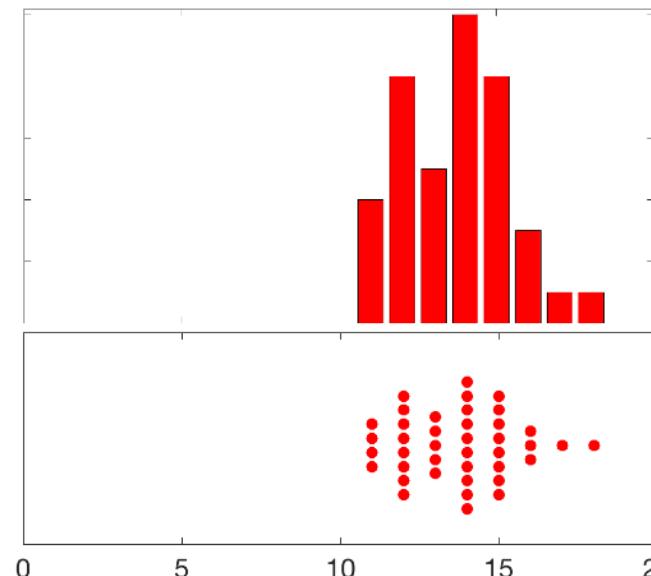
1-D OT Interpolation



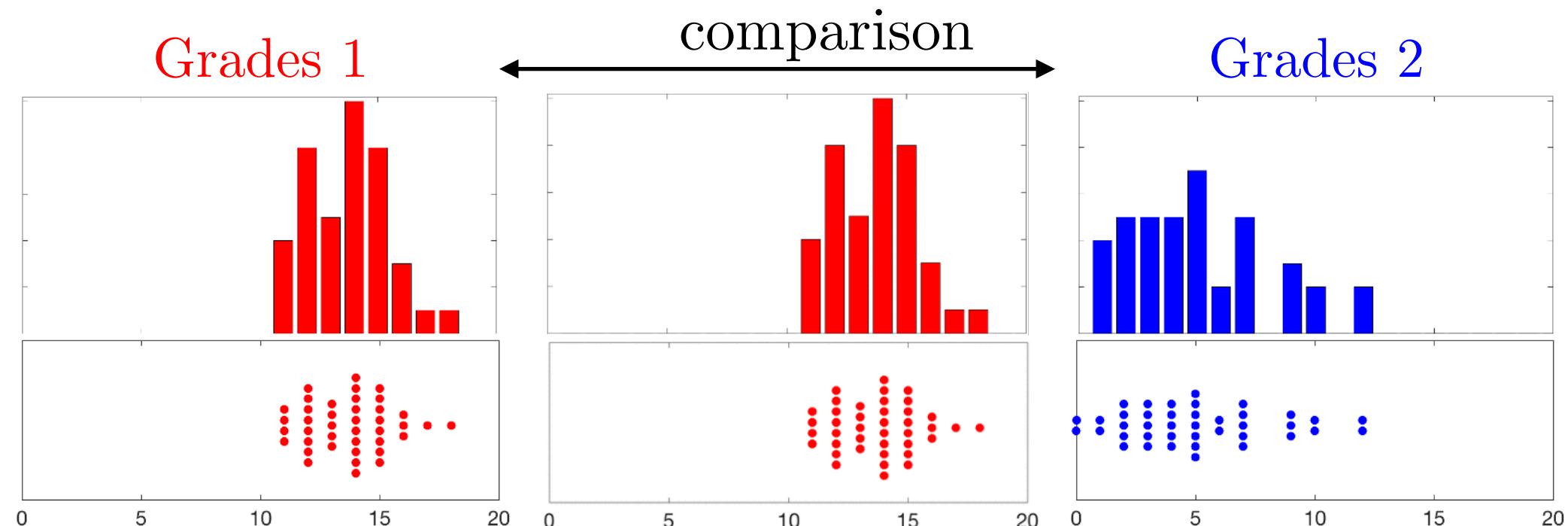
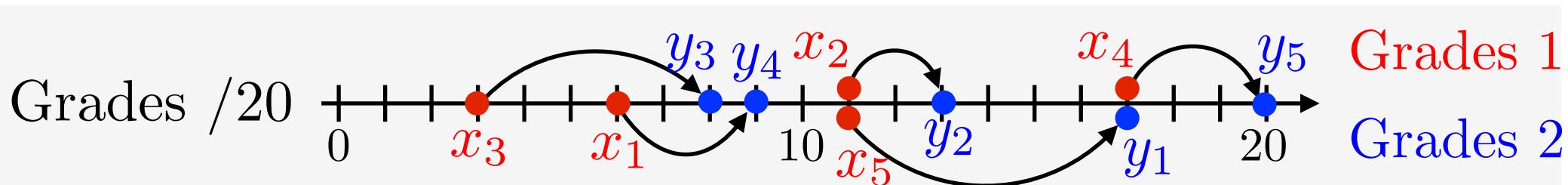
comparison

Grades 1

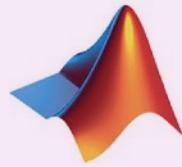
Grades 2



1-D OT Interpolation



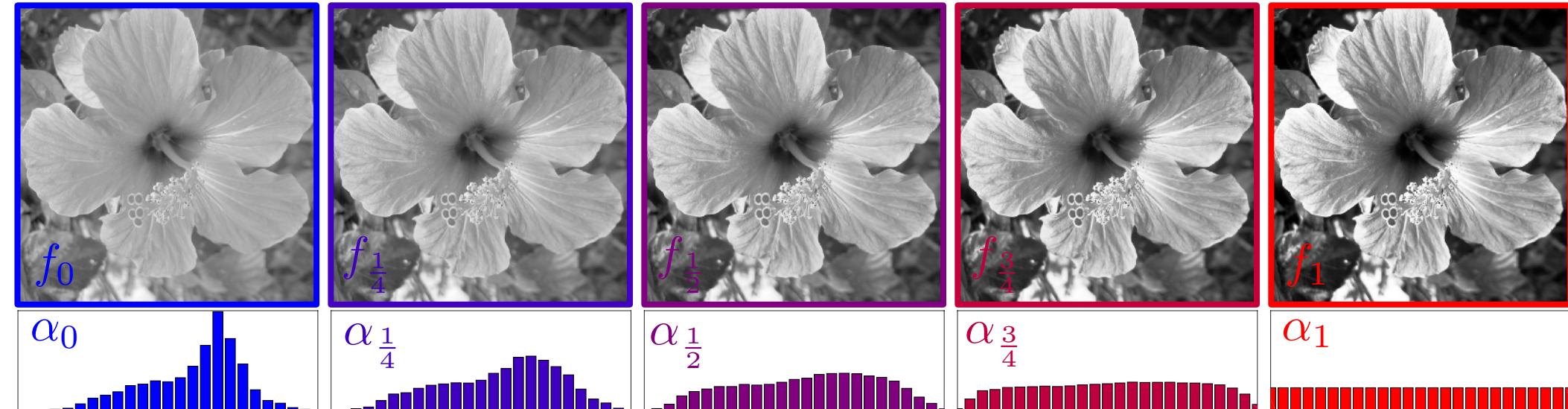
Application: Grayscale Histogram Equalization



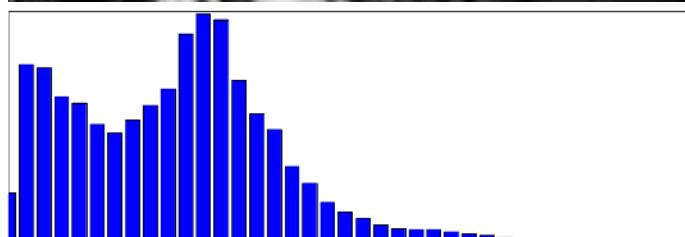
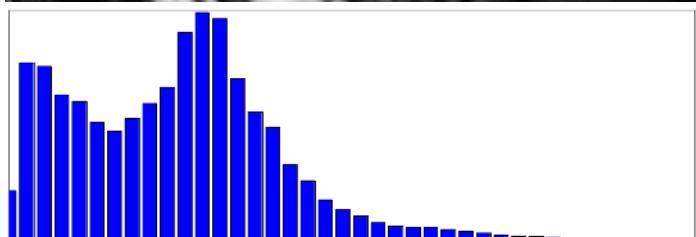
```
[~,I] = sort(f(:));  
f(I) = linspace(0,1,length(f(:)));
```



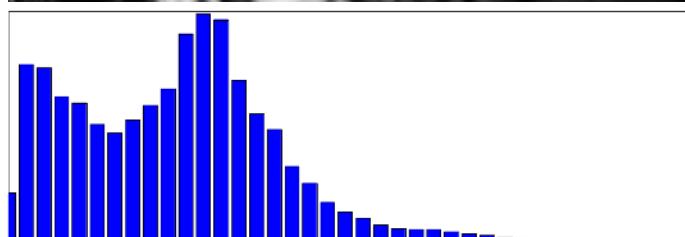
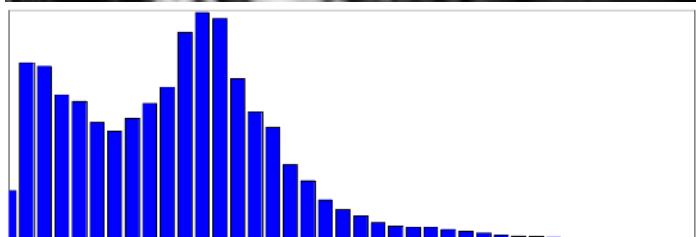
```
f[argsort(f.flatten())] = np.linspace(0,1,n*n)
```



Grayscale Histogram Equalization



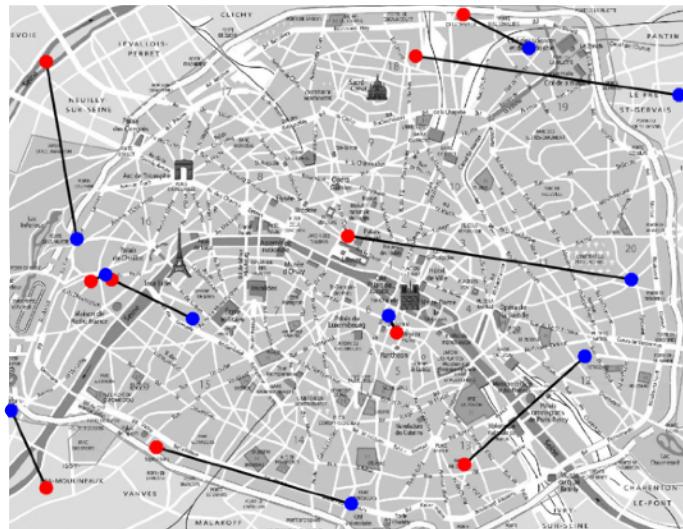
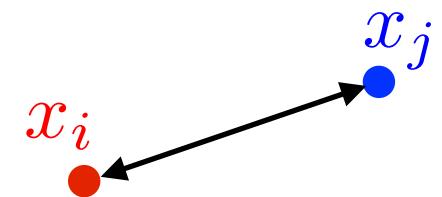
Grayscale Histogram Equalization



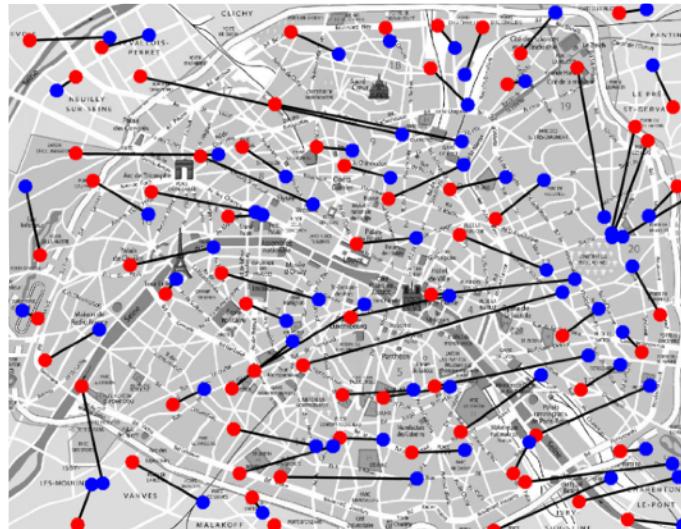
In 2-D

$$x_i, y_j \in \mathbb{R}^2$$

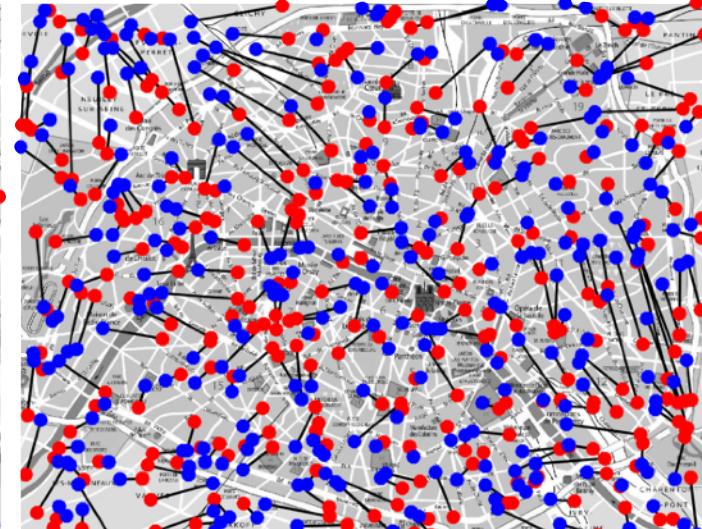
$$c_{i,j} = \|x_i - y_j\| = \sqrt{(x_i^1 - y_j^1)^2 + (x_i^2 - y_j^2)^2}$$



$n = 10$

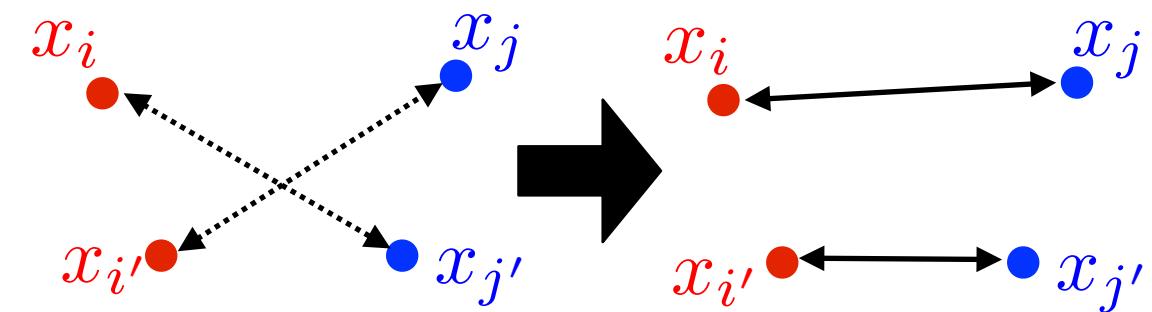


$n = 70$

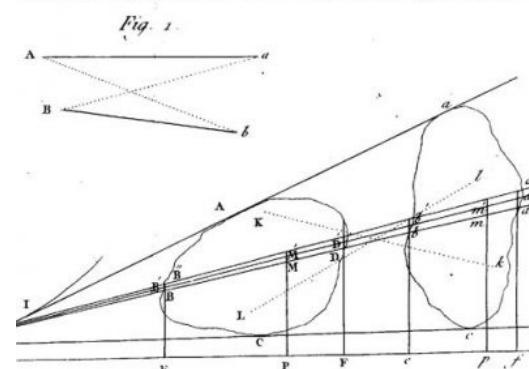


$n = 300$

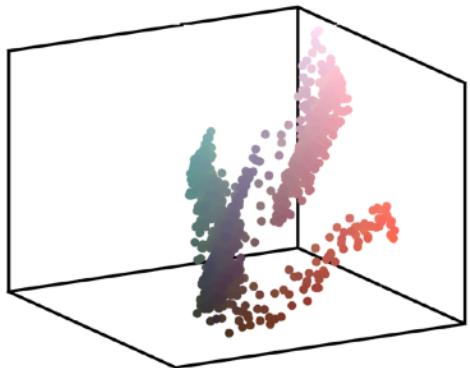
Proposition: two segments never cross.



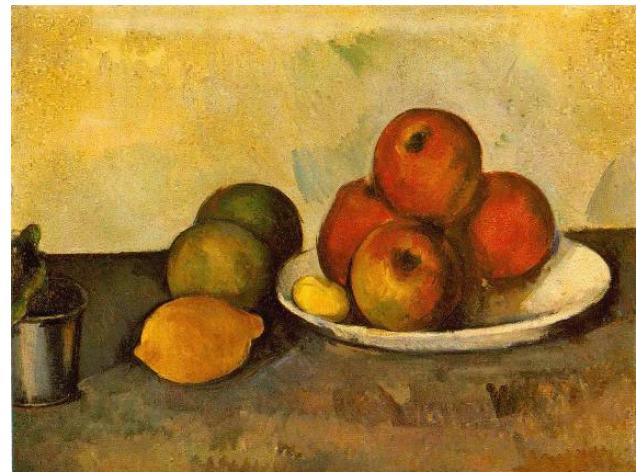
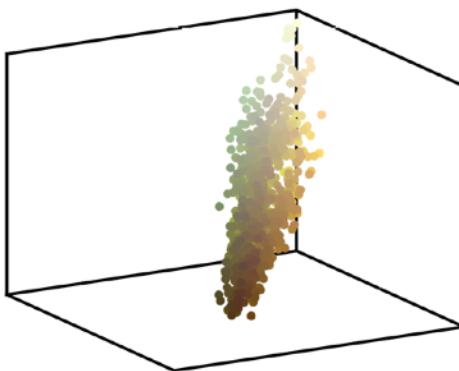
Mém. de l'Ac. R. des Sc. An. 1781, Page. 704. Pl. XVII



In 3-D: Color Image Palette Equalization

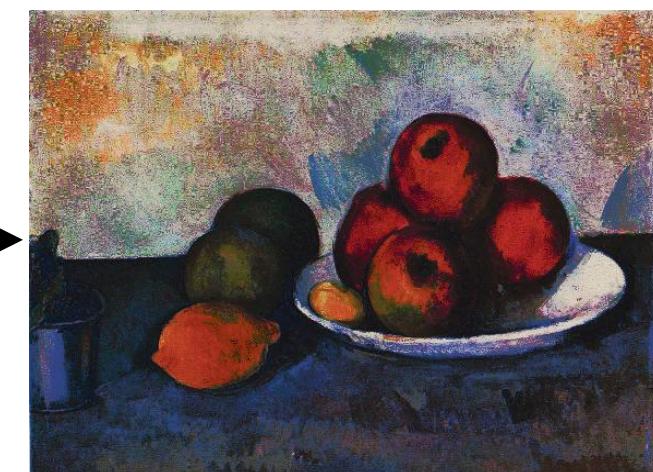
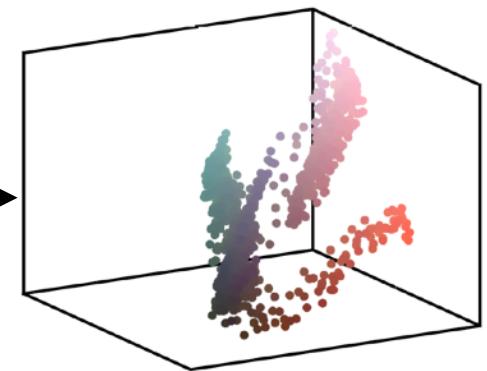


Reference



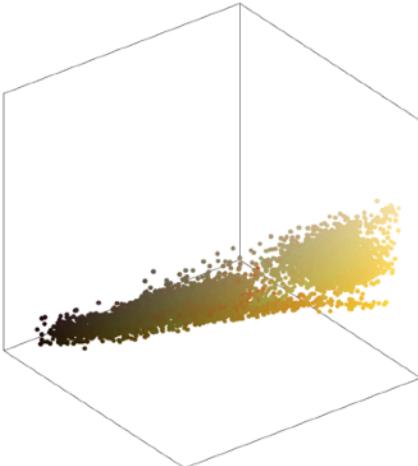
Input

optimal
transport

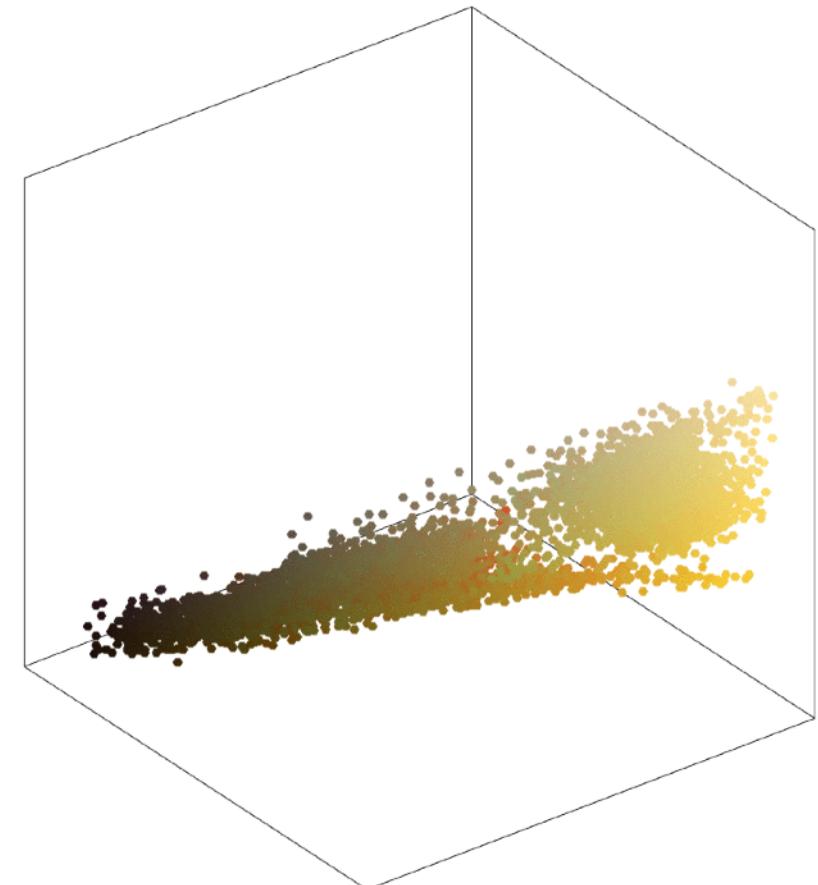
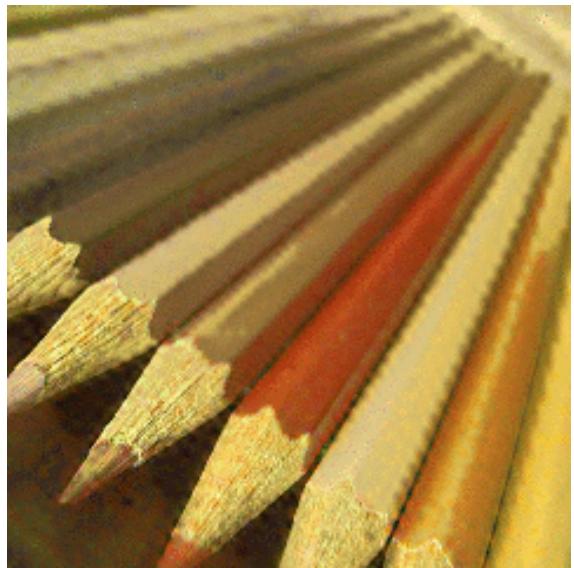
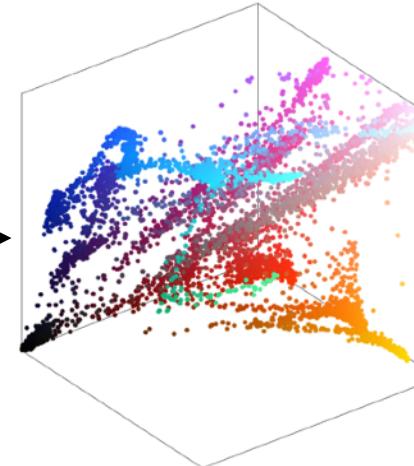


Output

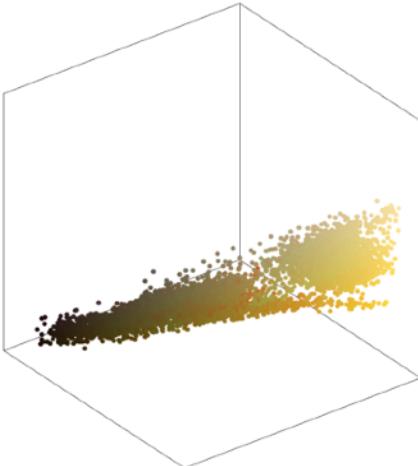
Color Image Palette Equalization



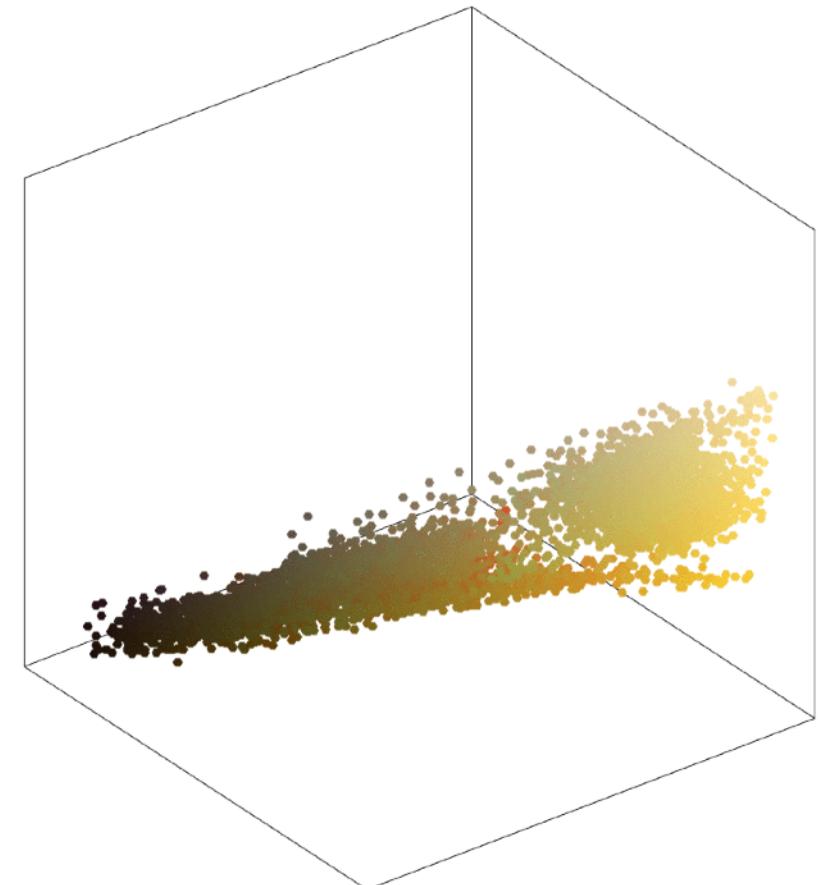
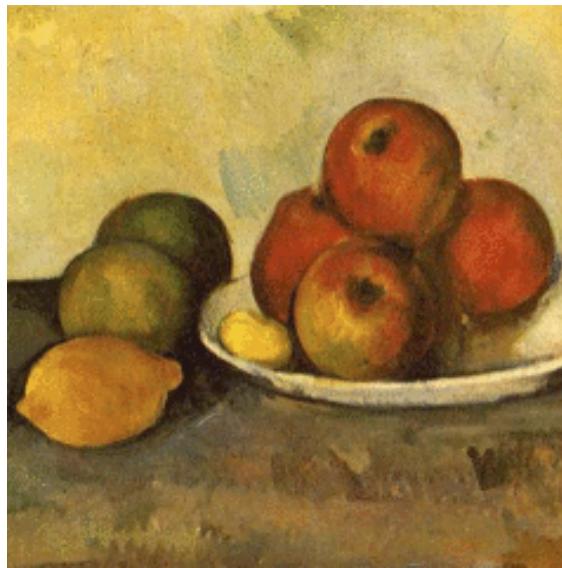
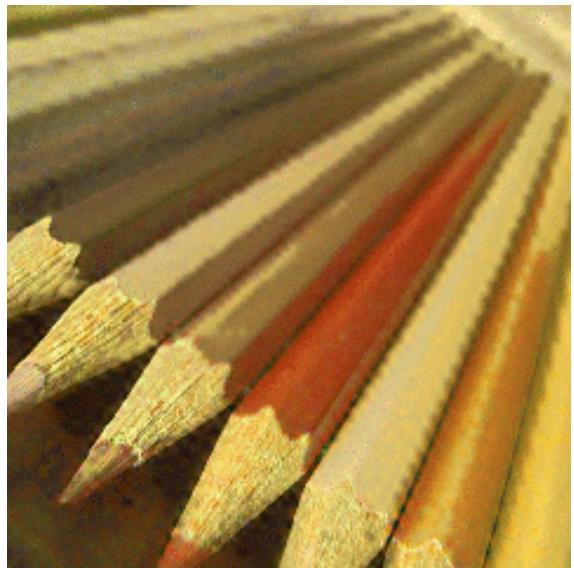
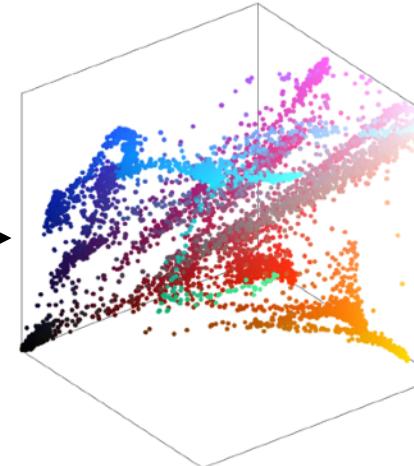
Optimal
transport



Color Image Palette Equalization



Optimal
transport



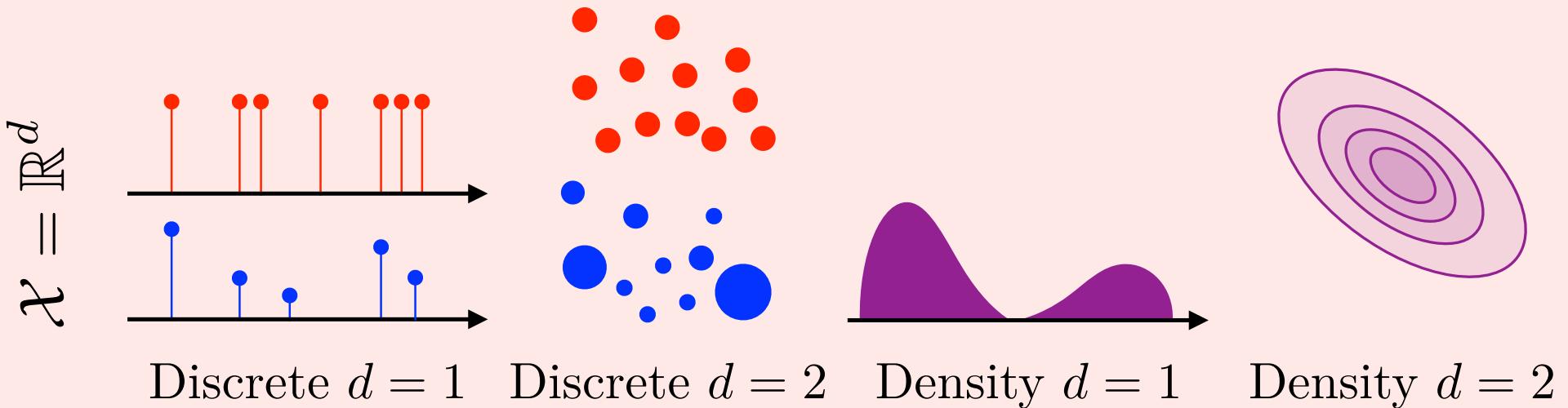
Overview

- Monge Formulation
- **Continuous Optimal Transport**
- Kantorovitch Formulation
- Applications

Probability Measures

Positive Radon measure α on a metric space \mathcal{X} .

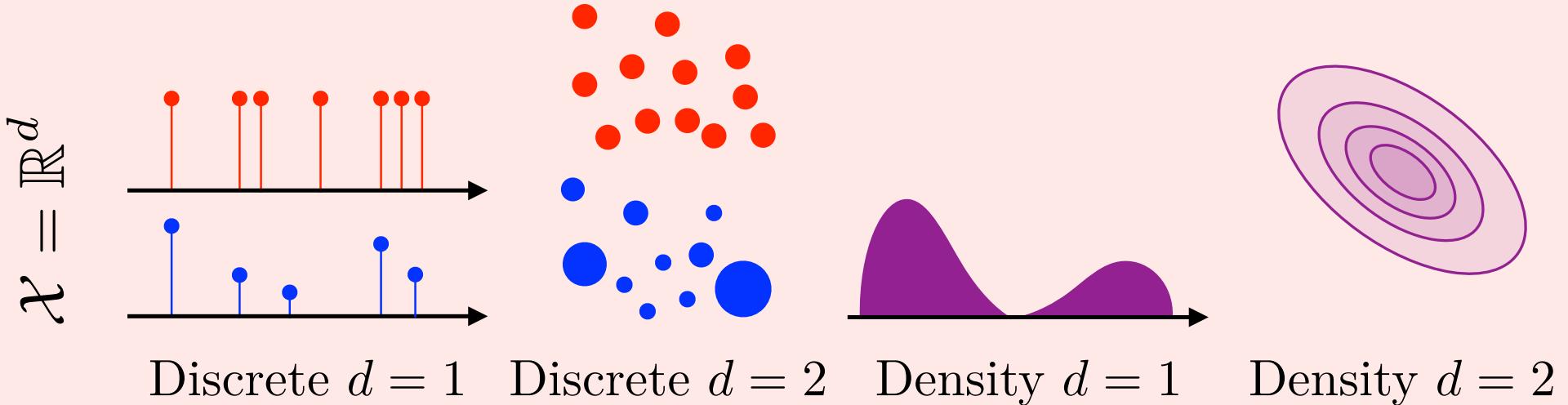
$$d\alpha(x) = \rho_\alpha(x)dx \quad \alpha = \sum_i \mathbf{a}_i \delta_{x_i}$$



Probability Measures

Positive Radon measure α on a metric space \mathcal{X} .

$$d\alpha(x) = \rho_\alpha(x)dx \quad \alpha = \sum_i \mathbf{a}_i \delta_{x_i}$$



Measure of sets $A \subset \mathcal{X}$: $\alpha(A) = \int_A d\alpha(x) \geq 0$

Integration against continuous functions: $\int_{\mathcal{X}} g(x)d\alpha(x) \geq 0$

$$\begin{aligned} d\alpha(x) = \rho_\alpha(x)dx &\longrightarrow \int_{\mathcal{X}} g d\alpha = \int_{\mathcal{X}} \rho_\alpha(x) dx \\ \alpha = \sum_i \mathbf{a}_i \delta_{x_i} &\longrightarrow \int_{\mathcal{X}} g d\alpha = \sum_i \mathbf{a}_i g(x_i) \end{aligned}$$

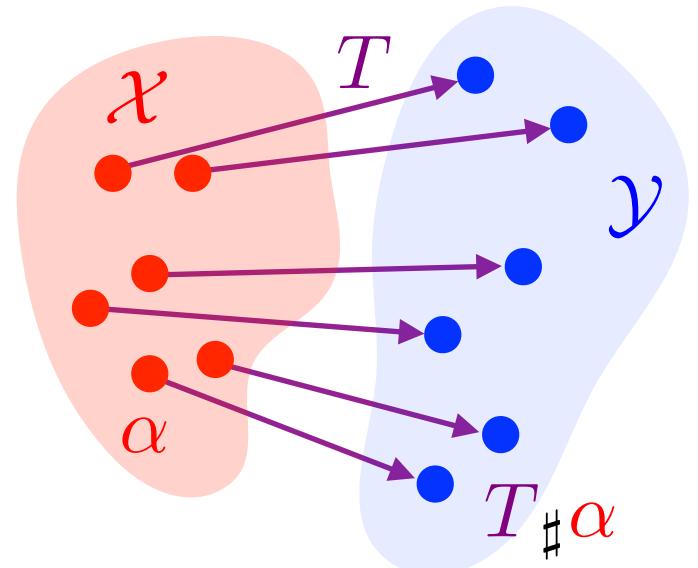
Probability (normalized) measure: $\alpha(\mathcal{X}) = \int_{\mathcal{X}} d\alpha(x) = 1$

Push Forward

Map: $T : \mathcal{X} \rightarrow \mathcal{Y}$

Push-forward:

$$T_{\sharp} : \begin{cases} \delta_x \mapsto \delta_{T(x)} \\ \sum_i \delta_{x_i} \mapsto \sum_i \delta_{T(x_i)} \\ \sum_i \mathbf{a}_i \delta_{x_i} \mapsto \sum_i \mathbf{a}_i \delta_{T(x_i)} \end{cases}$$

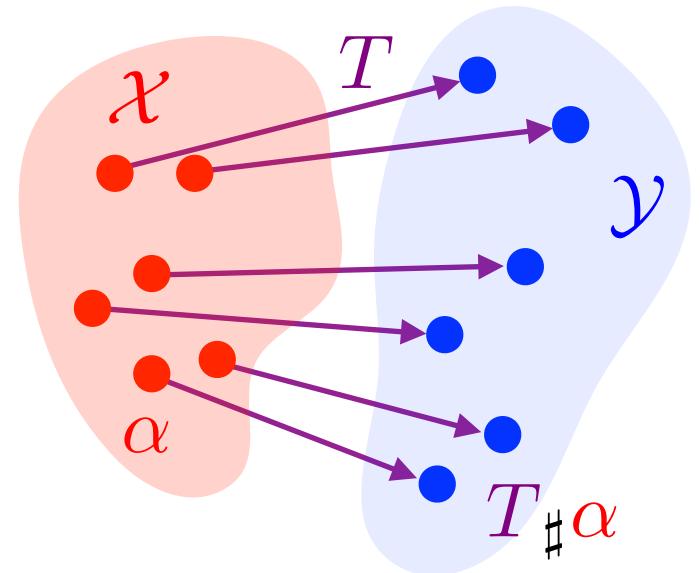


Push Forward

Map: $T : \mathcal{X} \rightarrow \mathcal{Y}$

Push-forward:

$$T_{\sharp} : \begin{cases} \delta_x \mapsto \delta_{T(x)} \\ \sum_i \delta_{x_i} \mapsto \sum_i \delta_{T(x_i)} \\ \sum_i \mathbf{a}_i \delta_{x_i} \mapsto \sum_i \mathbf{a}_i \delta_{T(x_i)} \end{cases}$$



General case:

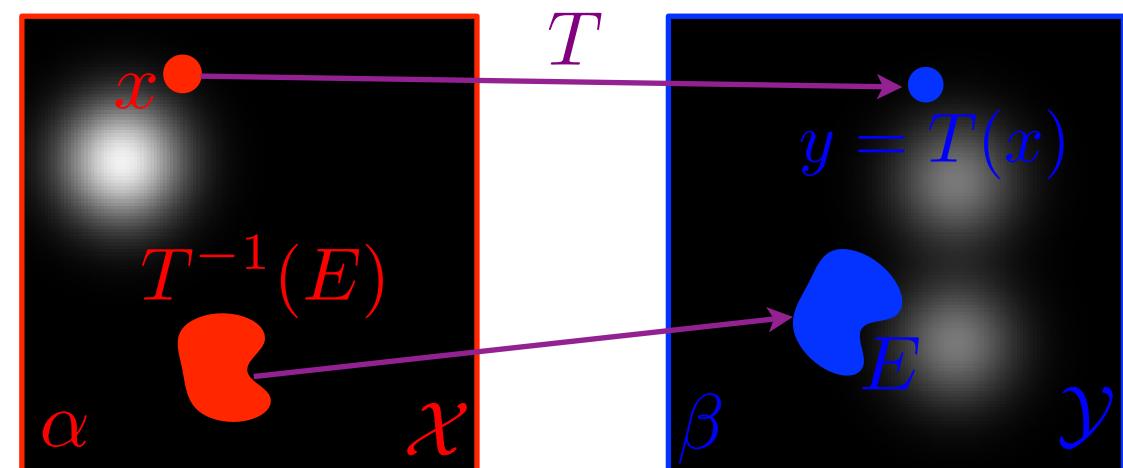
$$(T_{\sharp}\alpha)(E) \stackrel{\text{def.}}{=} \alpha(T^{-1}(E))$$

Change of variables:

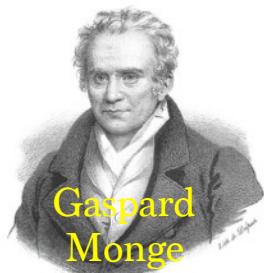
$$\beta = T_{\sharp}\alpha \iff \int_{\mathcal{Y}} g(y) d\beta(y) = \int_{\mathcal{X}} g(T(x)) d\alpha(x)$$

Densities $\frac{d\alpha}{dx} = \rho_{\alpha}$:

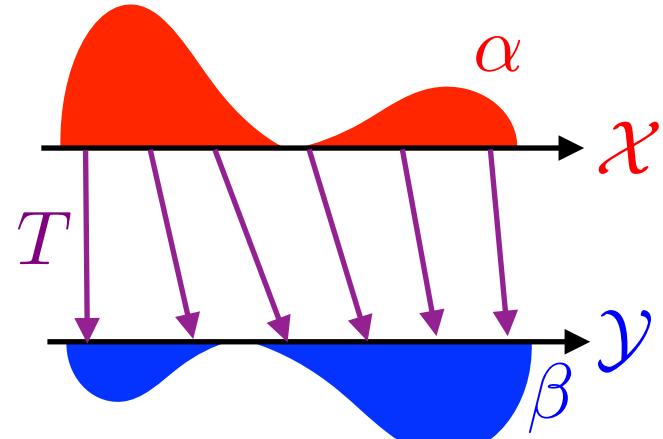
$$\rho_{\alpha}(x) = |\det(\partial T(x))| \rho_{\beta}(T(x))$$



Continuous Monge's Problem



$$\min_{\beta = T \# \alpha} \int_{\mathcal{X}} c(x, T(x)) d\alpha(x)$$



Discrete case:

$$\alpha = \sum_{i=1}^n \delta_{x_i}$$

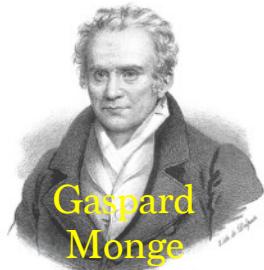
$$\beta = \sum_{j=1}^n \delta_{y_j}$$

$$\min_{\sigma \in \Sigma_n} \sum_{i=1}^n C_{i, \sigma(i)}$$

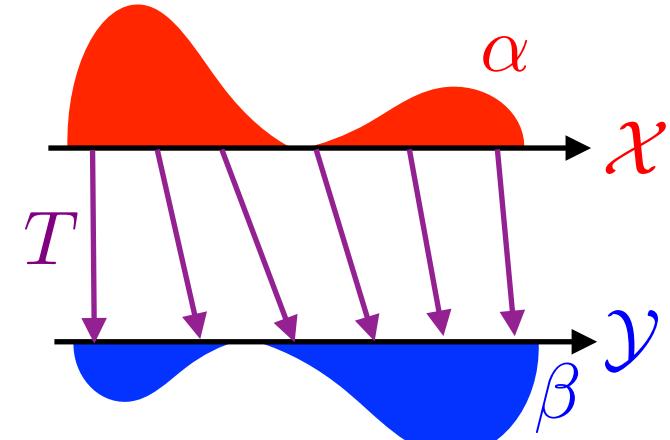
$$T : x_i \longmapsto y_{\sigma(i)}$$

$$C_{i,j} = c(x_i, y_j)$$

Continuous Monge's Problem



$$\min_{\beta = T \# \alpha} \int_{\mathcal{X}} c(x, T(x)) d\alpha(x)$$



Discrete case:

$$\alpha = \sum_{i=1}^n \delta_{x_i}$$

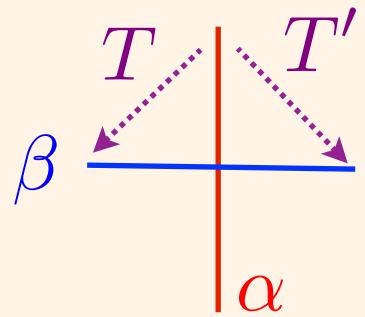
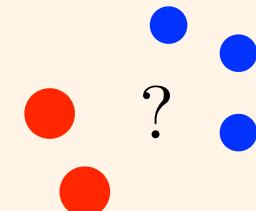
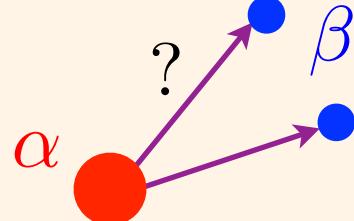
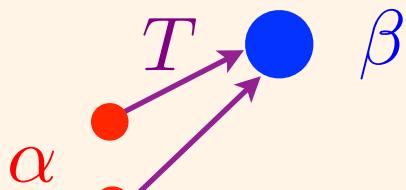
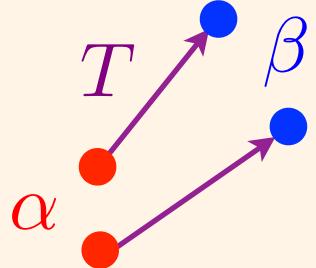
$$\beta = \sum_{j=1}^n \delta_{y_j}$$

$$\min_{\sigma \in \Sigma_n} \sum_{i=1}^n C_{i, \sigma(i)}$$

$$T : x_i \mapsto y_{\sigma(i)}$$

$$C_{i,j} = c(x_i, y_j)$$

Non-symmetry, non-existence, non-uniqueness:

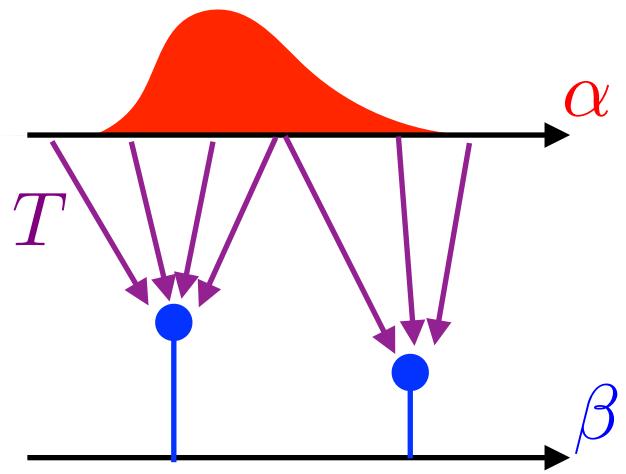


Brenier's Theorem

Hypotheses:

$$c(\textcolor{red}{x}, \textcolor{blue}{y}) = \|\textcolor{red}{x} - \textcolor{blue}{y}\|^2$$
$$\frac{d\alpha}{dx} = \rho_\alpha \text{ density.}$$

$$\min_{\beta = T \sharp \alpha} \int_{\mathcal{X}} c(x, \textcolor{violet}{T}(x)) d\alpha(x)$$



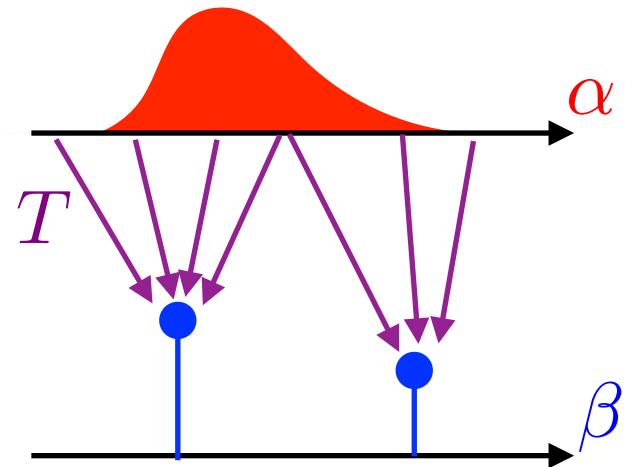
Brenier's Theorem

Hypotheses:

$$c(\textcolor{red}{x}, \textcolor{blue}{y}) = \|\textcolor{red}{x} - \textcolor{blue}{y}\|^2$$

$$\frac{d\alpha}{dx} = \rho_\alpha \text{ density.}$$

$$\min_{\beta = T \sharp \alpha} \int_{\mathcal{X}} c(x, \textcolor{violet}{T}(x)) d\alpha(x)$$



Theorem: [Brenier, 1991]

There exists a unique Monge map $\textcolor{violet}{T}$.

It is the unique $\textcolor{violet}{T} = \nabla \varphi$ such that

φ is convex and $(\nabla \varphi) \sharp \alpha = \beta$.



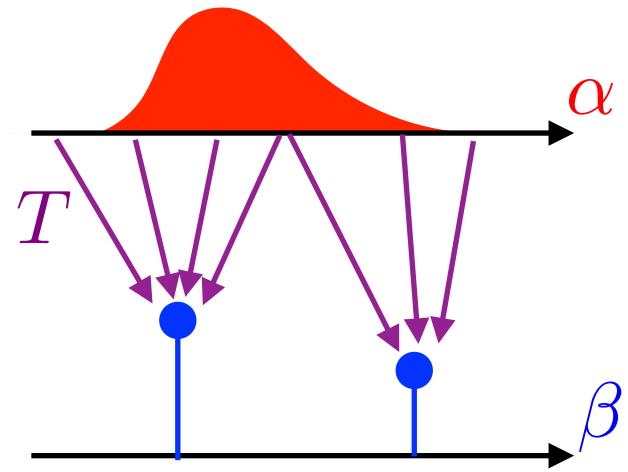
Brenier's Theorem

Hypotheses:

$$c(\textcolor{red}{x}, \textcolor{blue}{y}) = \|\textcolor{red}{x} - \textcolor{blue}{y}\|^2$$

$$\frac{d\alpha}{dx} = \rho_\alpha \text{ density.}$$

$$\min_{\beta = T \# \alpha} \int_{\mathcal{X}} c(x, \textcolor{violet}{T}(x)) d\alpha(x)$$

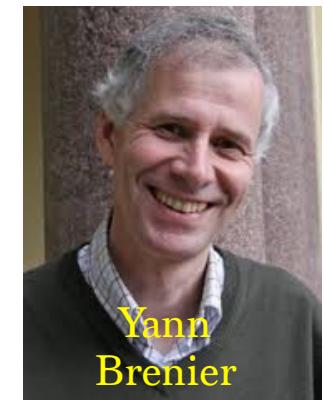


Theorem: [Brenier, 1991]

There exists a unique Monge map $\textcolor{violet}{T}$.

It is the unique $\textcolor{violet}{T} = \nabla \varphi$ such that

φ is convex and $(\nabla \varphi) \# \alpha = \beta$.

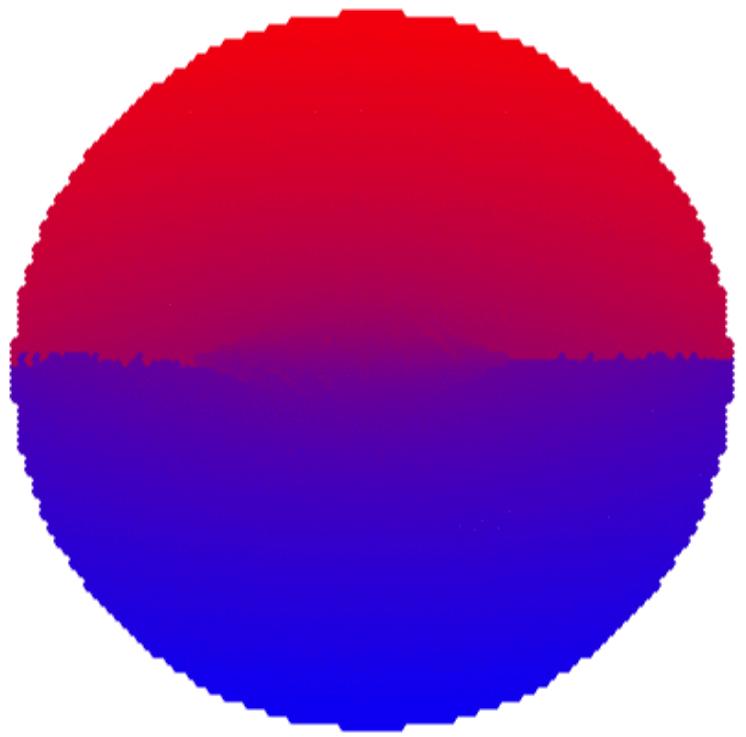
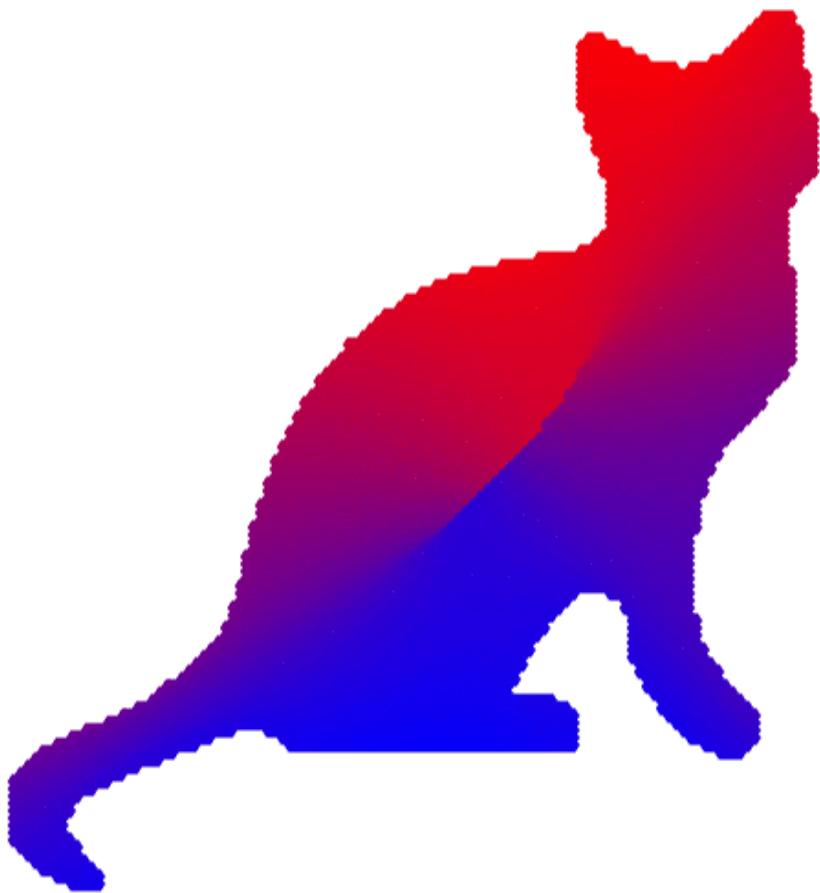


→ Monge-Ampère equation (non-linear, degenerate elliptic).

$$\rho_\alpha(x) = |\det(\partial^2 \varphi(x))| \rho_\beta(\textcolor{violet}{T}(x)) \quad \text{s.t. } \varphi \text{ convex.}$$

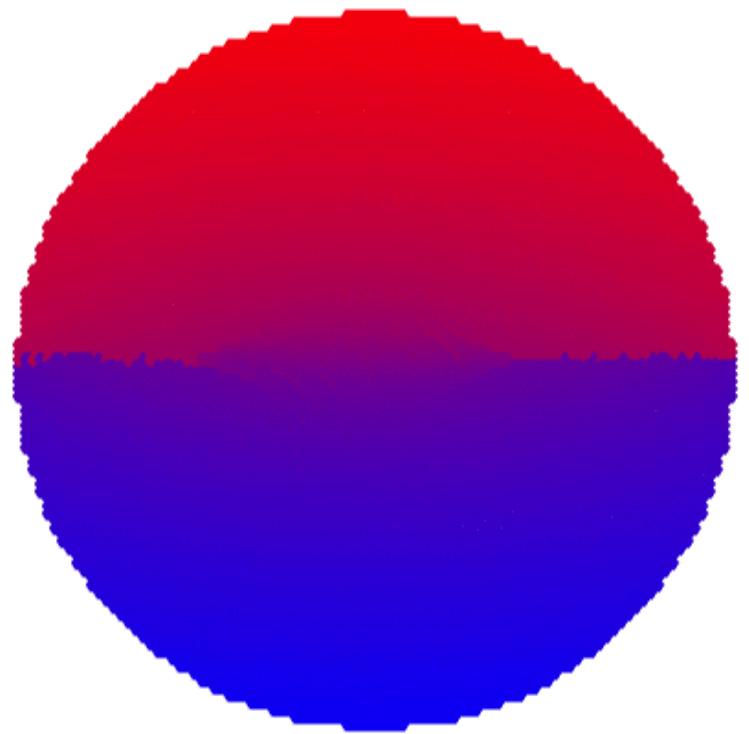
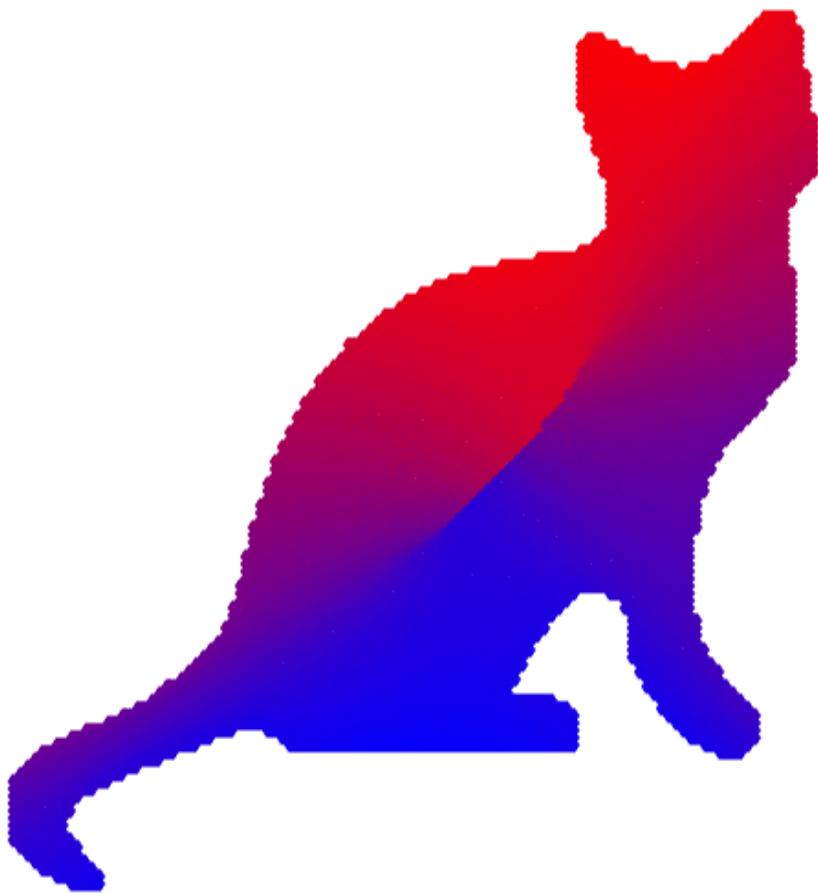
Regularity Theory

→ Regularity of T requires convex target.



Regularity Theory

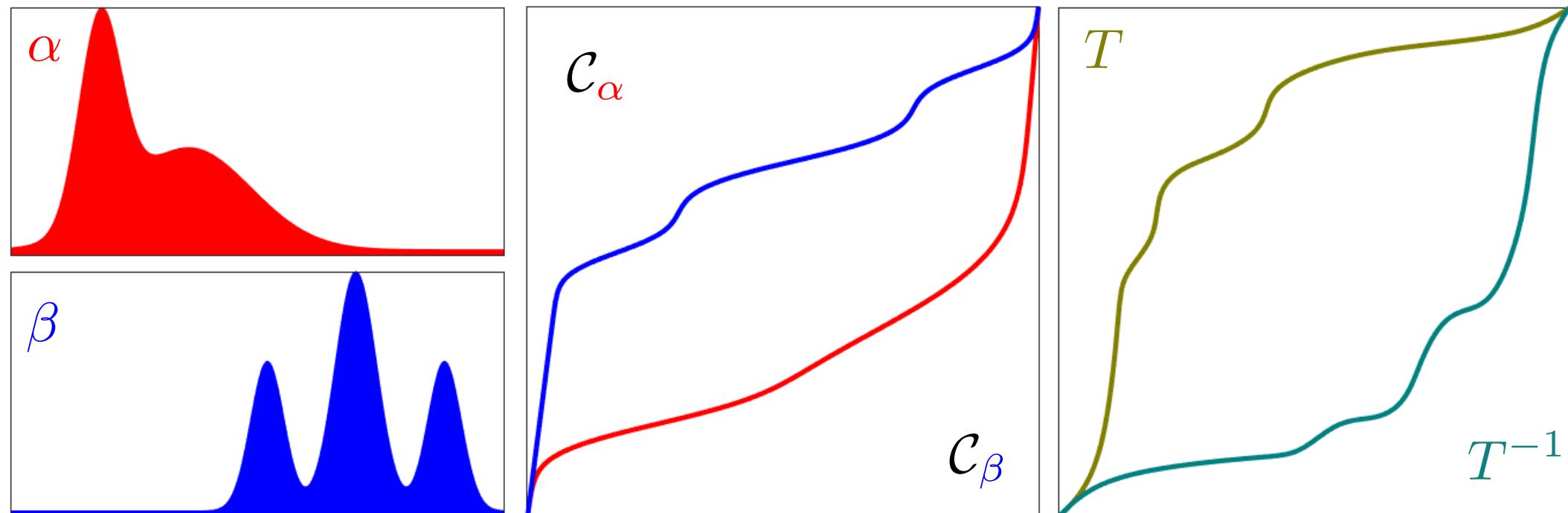
→ Regularity of T requires convex target.



1-D Optimal Transport

Cumulative function: $\mathcal{C}_\alpha(x) \stackrel{\text{def.}}{=} \int_{-\infty}^x d\alpha$

Optimal transport $\alpha \mapsto \beta$: $T = C_\beta^{-1} \circ C_\alpha$



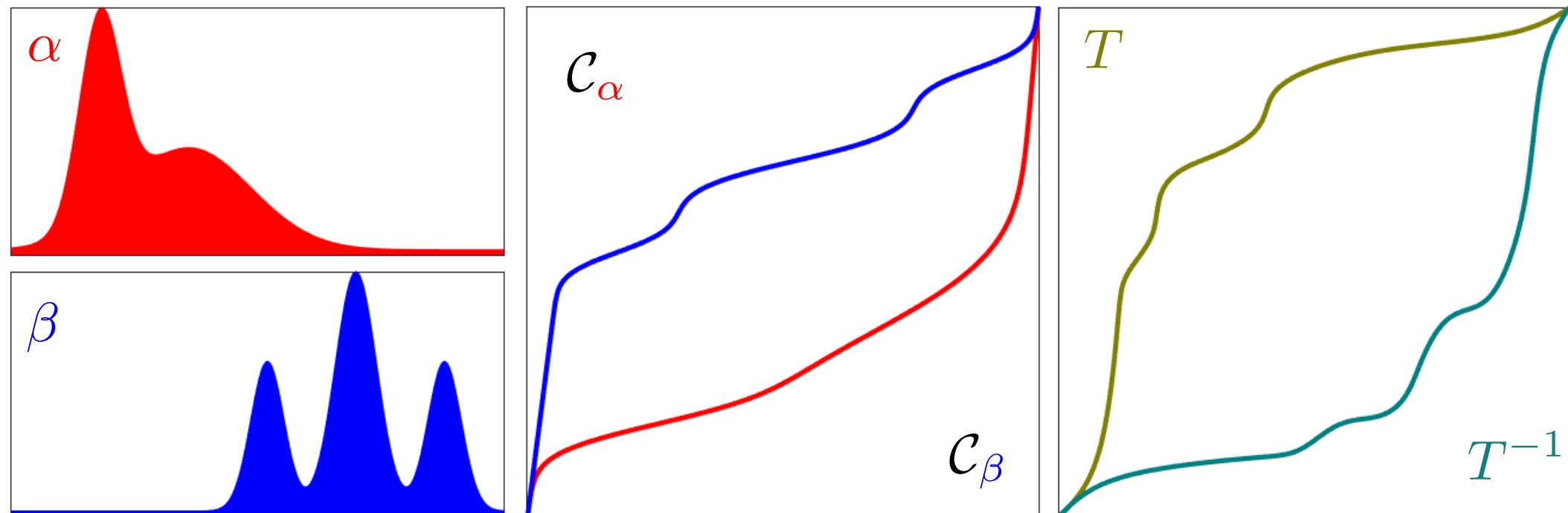
$$C_{\alpha\sharp} : \alpha \longmapsto \mathcal{U}_{[0,1]}$$

$$C_\beta^{-1\sharp} : \mathcal{U}_{[0,1]} \longmapsto \beta$$

1-D Optimal Transport

Cumulative function: $\mathcal{C}_\alpha(x) \stackrel{\text{def.}}{=} \int_{-\infty}^x d\alpha$

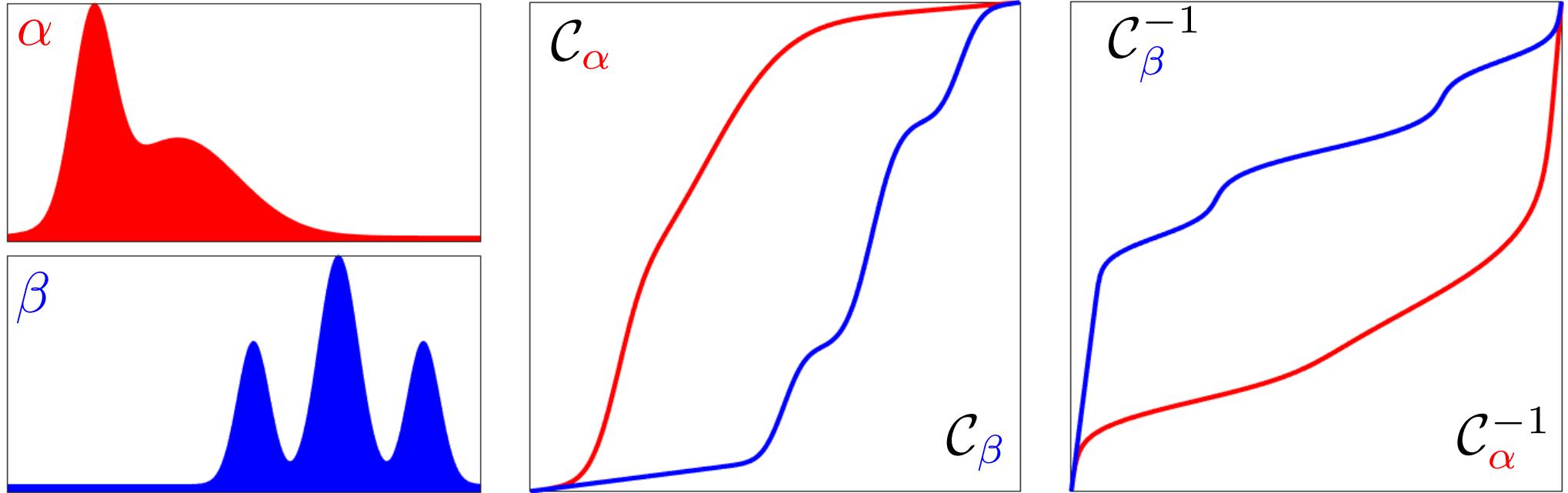
Optimal transport $\alpha \mapsto \beta$: $T = C_\beta^{-1} \circ C_\alpha$



$$C_{\alpha\sharp} : \alpha \longmapsto \mathcal{U}_{[0,1]}$$

$$C_\beta^{-1\sharp} : \mathcal{U}_{[0,1]} \longmapsto \beta$$

Cumulative function: $\mathcal{C}_\alpha(x) \stackrel{\text{def.}}{=} \int_{-\infty}^x d\alpha$



Wasserstein distance: $W_p(\alpha, \beta)^p = \int_0^1 |\mathcal{C}_\alpha^{-1}(t) - \mathcal{C}_\beta^{-1}(t)|^p dt$

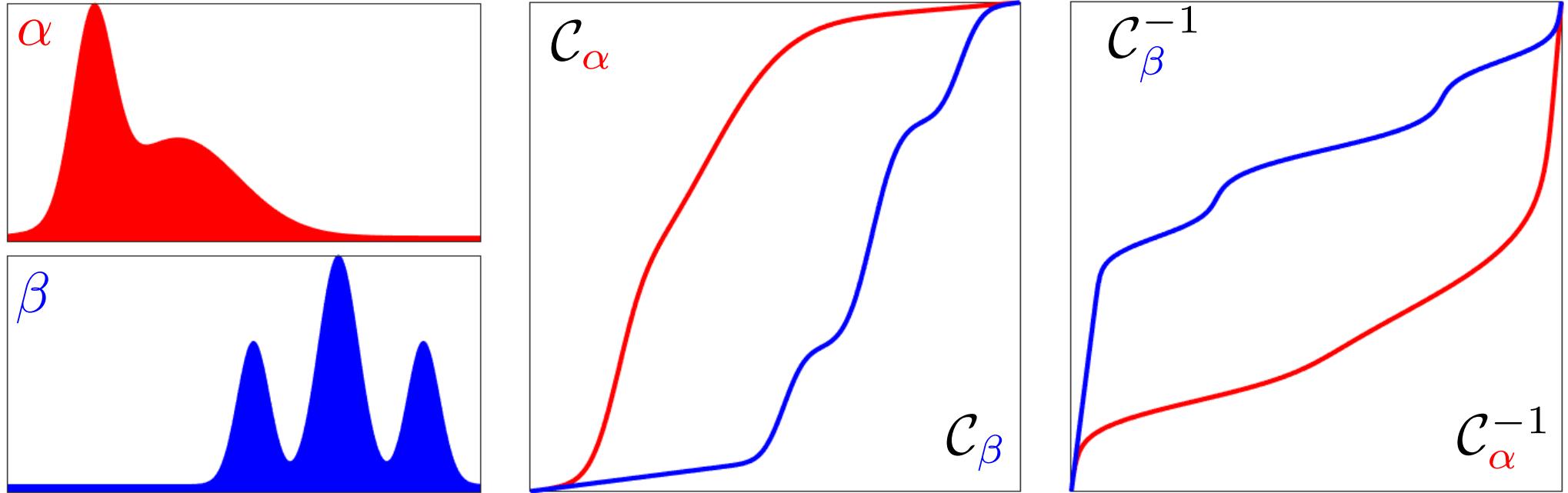
$$W_1(\alpha, \beta) = \|\alpha - \beta\|_{W_1} = \int_{\mathbb{R}} |\mathcal{C}_\alpha(x) - \mathcal{C}_\beta(x)| dx$$

Kramer (Sobolev) norm: $\|\alpha - \beta\|_K^2 = \int_0^1 |\mathcal{C}_\alpha(t) - \mathcal{C}_\beta(t)|^2 dt$

Kolmogorov-Smirnov norm: $\|\alpha - \beta\|_{KS} = \sup_x |\mathcal{C}_\alpha(x) - \mathcal{C}_\beta(x)|^2$

Area under the curve: $\text{AUC}(\alpha, \beta) = 1 - \int_0^1 \mathcal{C}_\alpha \circ \mathcal{C}_\beta^{-1}(x)$

Cumulative function: $\mathcal{C}_\alpha(x) \stackrel{\text{def.}}{=} \int_{-\infty}^x d\alpha$



Wasserstein distance: $W_p(\alpha, \beta)^p = \int_0^1 |\mathcal{C}_\alpha^{-1}(t) - \mathcal{C}_\beta^{-1}(t)|^p dt$

$$W_1(\alpha, \beta) = \|\alpha - \beta\|_{W_1} = \int_{\mathbb{R}} |\mathcal{C}_\alpha(x) - \mathcal{C}_\beta(x)| dx$$

Kramer (Sobolev) norm: $\|\alpha - \beta\|_K^2 = \int_0^1 |\mathcal{C}_\alpha(t) - \mathcal{C}_\beta(t)|^2 dt$

Kolmogorov-Smirnov norm: $\|\alpha - \beta\|_{KS} = \sup_x |\mathcal{C}_\alpha(x) - \mathcal{C}_\beta(x)|^2$

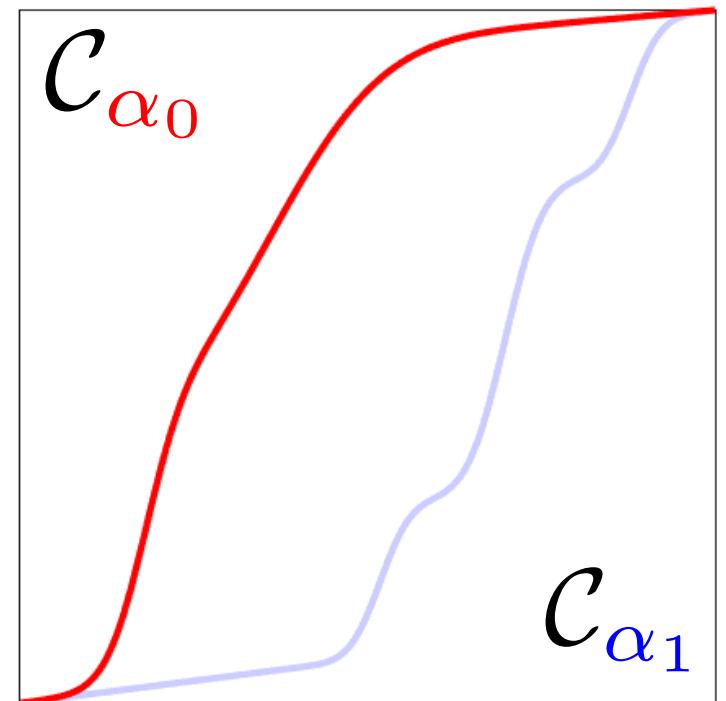
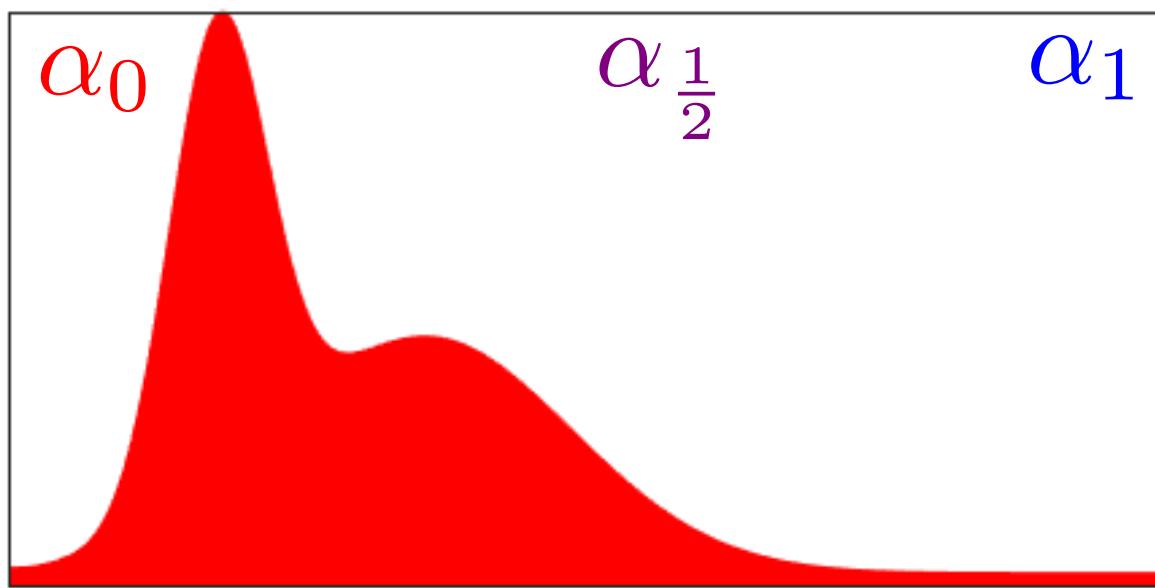
Area under the curve: $\text{AUC}(\alpha, \beta) = 1 - \int_0^1 \mathcal{C}_\alpha \circ \mathcal{C}_\beta^{-1}(x)$

1-D Optimal Transport Interpolation

Cumulative function: $\mathcal{C}_\alpha(x) \stackrel{\text{def.}}{=} \int_{-\infty}^x d\alpha$

Optimal transport interpolation $\alpha_0 \leftrightarrow \alpha_1$

$$\forall t \in [0, 1], \mathcal{C}_{\alpha_t} = (1 - t)\mathcal{C}_{\alpha_0} + t\mathcal{C}_{\alpha_1}$$

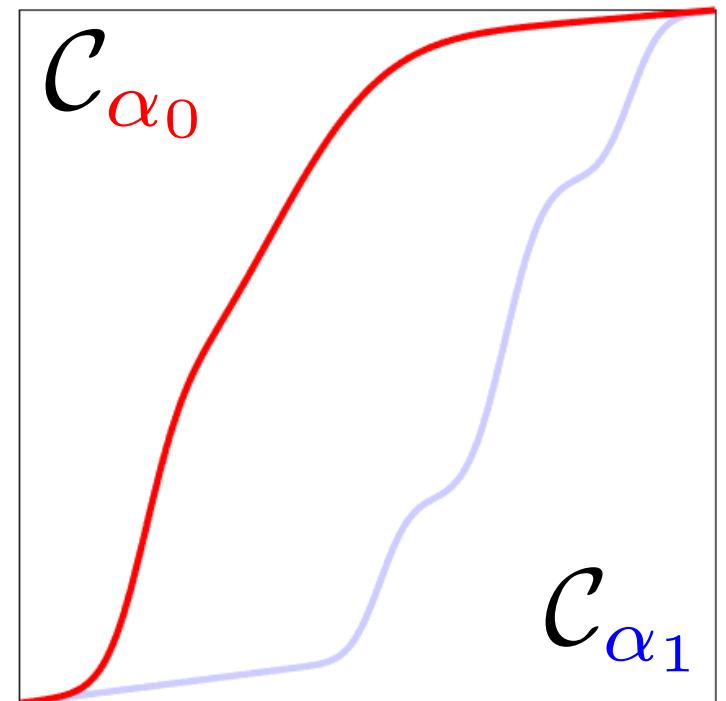
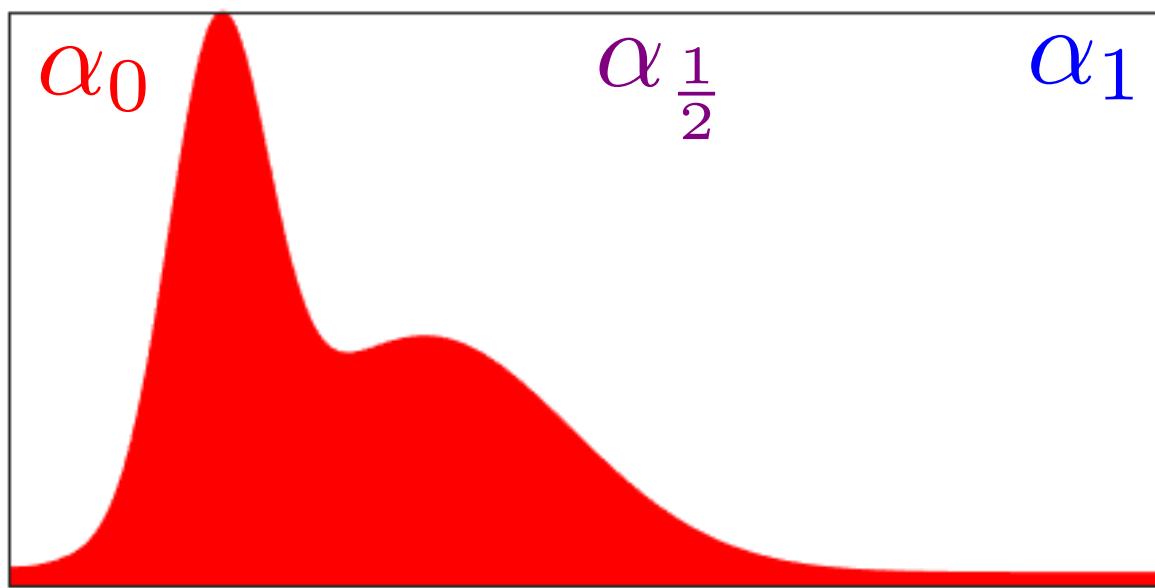


1-D Optimal Transport Interpolation

Cumulative function: $\mathcal{C}_\alpha(x) \stackrel{\text{def.}}{=} \int_{-\infty}^x d\alpha$

Optimal transport interpolation $\alpha_0 \leftrightarrow \alpha_1$

$$\forall t \in [0, 1], \mathcal{C}_{\alpha_t} = (1 - t)\mathcal{C}_{\alpha_0} + t\mathcal{C}_{\alpha_1}$$

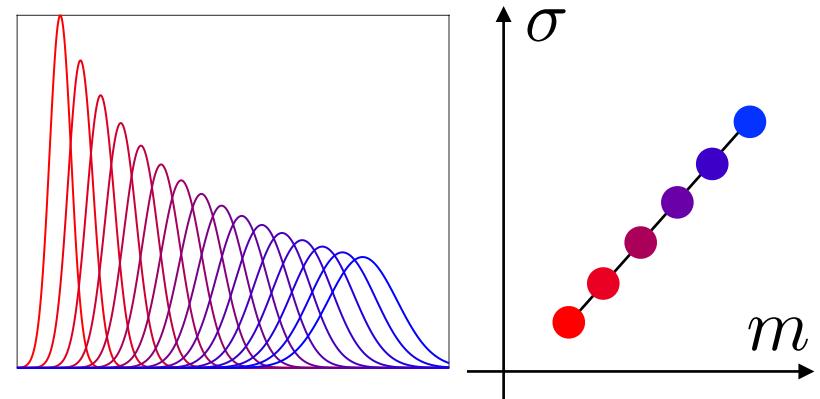
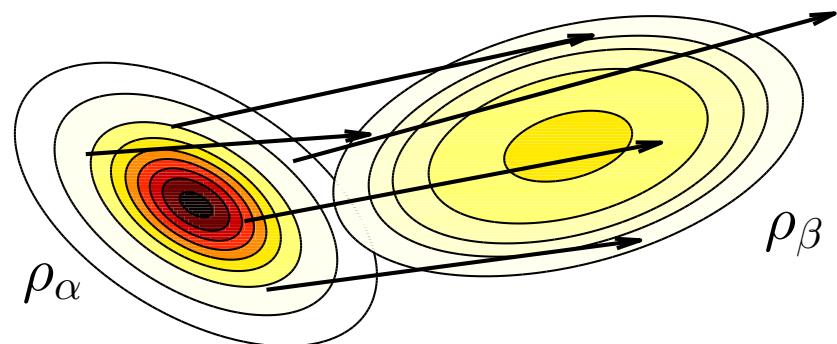


OT Between Gaussians

OT between $\beta = \mathcal{N}(\mathbf{m}_\beta, \Sigma_\beta)$ $\alpha = \mathcal{N}(\mathbf{m}_\alpha, \Sigma_\alpha)$ is affine:

$$T : x \mapsto \mathbf{m}_\beta + A(x - \mathbf{m}_\alpha)$$
$$A = \Sigma_\alpha^{-\frac{1}{2}} \left(\Sigma_\alpha^{\frac{1}{2}} \Sigma_\beta \Sigma_\alpha^{\frac{1}{2}} \right)^{\frac{1}{2}} \Sigma_\alpha^{-\frac{1}{2}} = A^T$$

→ Gaussians are OT geodesically convex.



OT Between Gaussians

OT between $\beta = \mathcal{N}(\mathbf{m}_\beta, \Sigma_\beta)$
 $\alpha = \mathcal{N}(\mathbf{m}_\alpha, \Sigma_\alpha)$ is affine:

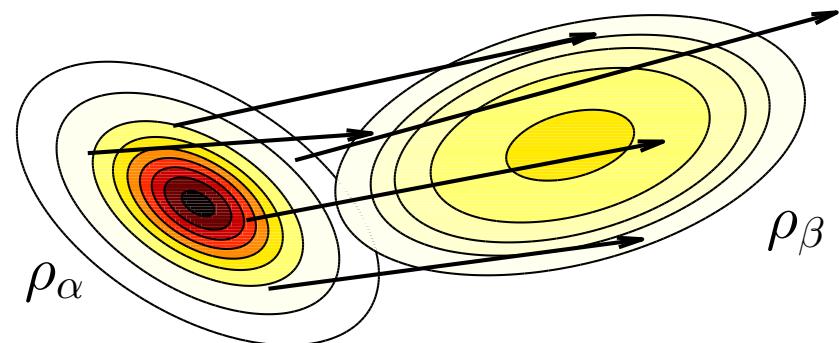
$$T : x \mapsto \mathbf{m}_\beta + A(x - \mathbf{m}_\alpha)$$
$$A = \Sigma_\alpha^{-\frac{1}{2}} \left(\Sigma_\alpha^{\frac{1}{2}} \Sigma_\beta \Sigma_\alpha^{\frac{1}{2}} \right)^{\frac{1}{2}} \Sigma_\alpha^{-\frac{1}{2}} = A^\top$$

→ Gaussians are OT geodesically convex.

$$\mathcal{W}_2^2(\alpha, \beta) = \|\mathbf{m}_\alpha - \mathbf{m}_\beta\|^2 + \mathcal{B}(\Sigma_\alpha, \Sigma_\beta)^2$$

Bures distance:

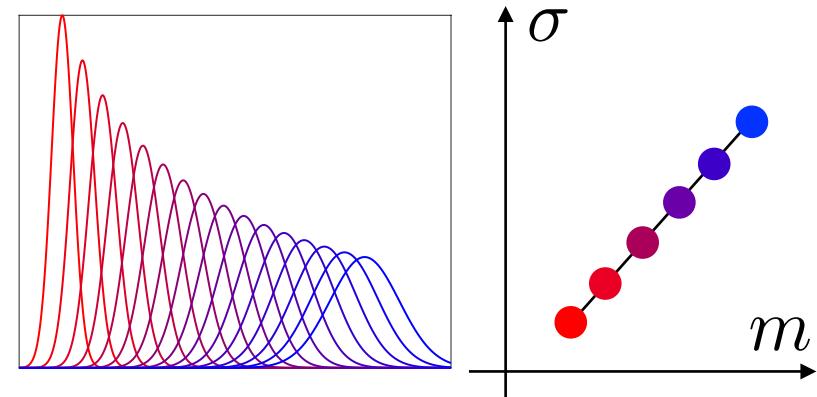
$$\mathcal{B}(\Sigma_\alpha, \Sigma_\beta)^2 \stackrel{\text{def.}}{=} \text{tr} \left(\Sigma_\alpha + \Sigma_\beta - 2(\Sigma_\alpha^{1/2} \Sigma_\beta \Sigma_\alpha^{1/2})^{1/2} \right)$$



Diagonal covariances:

$$\Sigma_\alpha = \text{diag}(r_i)_i \quad \Sigma_\beta = \text{diag}(s_i)_i$$

$$\mathcal{B}(\Sigma_\alpha, \Sigma_\beta) = \|\sqrt{r} - \sqrt{s}\|_2$$

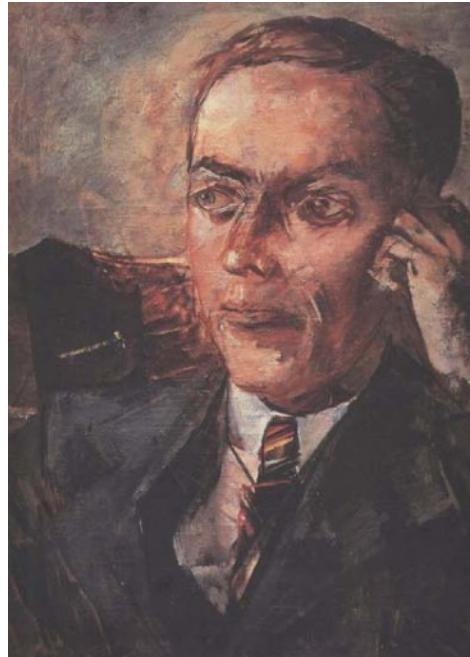


Overview

- Monge Formulation
- Continuous Optimal Transport
- Kantorovitch Formulation
- Applications

Leonid Kantorovich (1912-1986)

Леонид Витальевич Канторович



Journal of Mathematical Sciences, Vol. 133, No. 4, 2006

[Kantorovich 1942]

ON THE TRANSLOCATION OF MASSES

L. V. Kantorovich*

The original paper was published in *Dokl. Akad. Nauk SSSR*, **37**, No. 7-8, 227-229 (1942).

We assume that R is a compact metric space, though some of the definitions and results given below can be formulated for more general spaces.

Let $\Phi(e)$ be a mass distribution, i.e., a set function such that: (1) it is defined for Borel sets, (2) it is nonnegative: $\Phi(e) \geq 0$, (3) it is absolutely additive: if $e = e_1 + e_2 + \dots$; $e_i \cap e_k = 0$ ($i \neq k$), then $\Phi(e) = \Phi(e_1) + \Phi(e_2) + \dots$. Let $\Phi'(e')$ be another mass distribution such that $\Phi(R) = \Phi'(R)$. By definition, a translocation of masses is a function $\Psi(e, e')$ defined for pairs of (B)-sets $e, e' \in R$ such that: (1) it is nonnegative and absolutely additive with respect to each of its arguments, (2) $\Psi(e, R) = \Phi(e)$, $\Psi(R, e') = \Phi'(e')$.

Let $r(x, y)$ be a known continuous nonnegative function representing the work required to move a unit mass from x to y .

We define the work required for the translocation of two given mass distributions as

$$W(\Phi, \Phi') = \int_R r(x, x') \Psi(de, de') = \lim_{\lambda \rightarrow 0} \sum_{i, k} r(x_i, x'_k) \Psi(e_i, e'_k),$$

where e_i are disjoint and $\sum_i e_i = R$, e'_k are disjoint and $\sum_k e'_k = R$, $x_i \in e_i$, $x'_k \in e'_k$, and λ is the largest of the numbers $\text{diam } e_i$ ($i = 1, 2, \dots, n$) and $\text{diam } e'_k$ ($k = 1, 2, \dots, m$).

Clearly, this integral does exist.

We call the quantity

$$W(\Phi, \Phi') = \inf_{\Psi} W(\Psi, \Phi, \Phi')$$

the minimal translocation work. Since the set of all functions $\{\Psi\}$ is compact, there exists a function Ψ_0 realizing this minimum, so that

$$W(\Phi, \Phi') = W(\Psi_0, \Phi, \Phi'),$$

Kantorovitch's Formulation

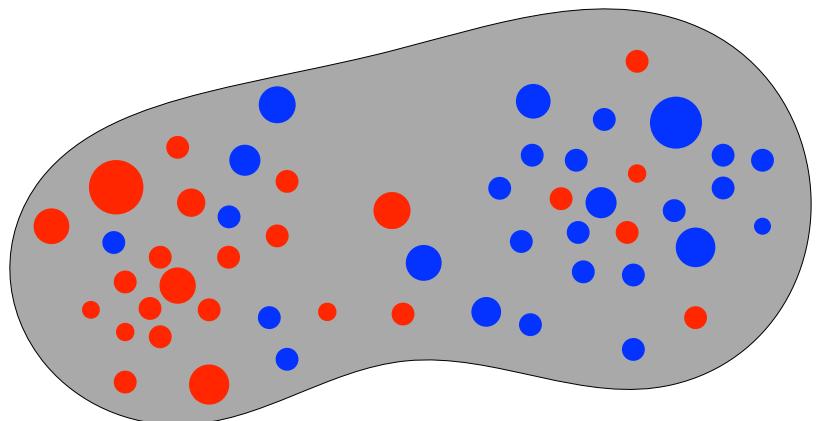
Input distributions

$$\alpha = \sum_{i=1}^n \mathbf{a}_i \delta_{x_i} \quad \beta = \sum_{j=1}^m \mathbf{b}_j \delta_{y_j}$$

Points $(x_i)_i, (y_j)_j$

Weights $\mathbf{a}_i \geq 0, \mathbf{b}_j \geq 0.$

$$\sum_{i=1}^n \mathbf{a}_i = \sum_{j=1}^m \mathbf{b}_j = 1$$



Kantorovitch's Formulation

Input distributions

$$\alpha = \sum_{i=1}^n \mathbf{a}_i \delta_{x_i} \quad \beta = \sum_{j=1}^m \mathbf{b}_j \delta_{y_j}$$

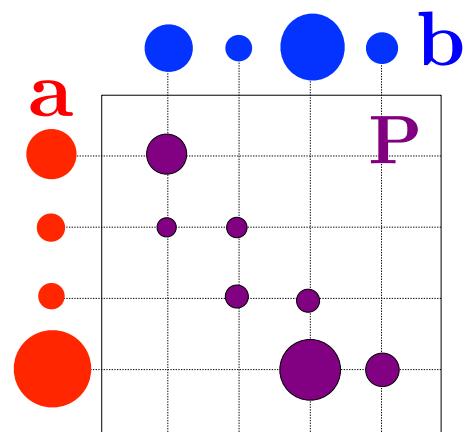
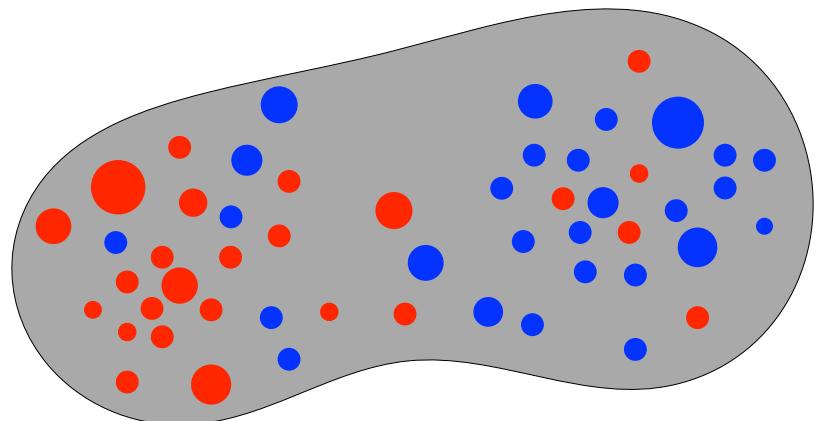
Points $(x_i)_i, (y_j)_j$

Weights $\mathbf{a}_i \geq 0, \mathbf{b}_j \geq 0.$

$$\sum_{i=1}^n \mathbf{a}_i = \sum_{j=1}^m \mathbf{b}_j = 1$$

Couplings:

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



Kantorovitch's Formulation

Input distributions

$$\alpha = \sum_{i=1}^n \mathbf{a}_i \delta_{x_i} \quad \beta = \sum_{j=1}^m \mathbf{b}_j \delta_{y_j}$$

Points $(x_i)_i, (y_j)_j$

Weights $\mathbf{a}_i \geq 0, \mathbf{b}_j \geq 0.$

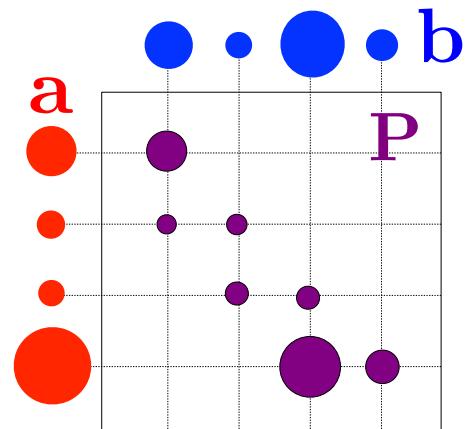
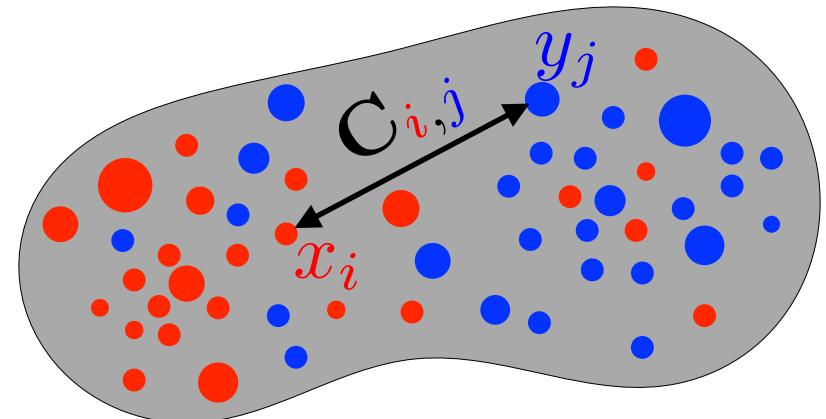
$$\sum_{i=1}^n \mathbf{a}_i = \sum_{j=1}^m \mathbf{b}_j = 1$$

Couplings:

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

[Kantorovich 1942]

$$\min \left\{ \sum_{i,j} C_{i,j} P_{i,j} ; \mathbf{P} \in \mathbf{U}(\mathbf{a}, \mathbf{b}) \right\}$$



Kantorovitch's Formulation

Input distributions

$$\alpha = \sum_{i=1}^n \mathbf{a}_i \delta_{x_i} \quad \beta = \sum_{j=1}^m \mathbf{b}_j \delta_{y_j}$$

Points $(x_i)_i, (y_j)_j$

Weights $\mathbf{a}_i \geq 0, \mathbf{b}_j \geq 0.$

$$\sum_{i=1}^n \mathbf{a}_i = \sum_{j=1}^m \mathbf{b}_j = 1$$

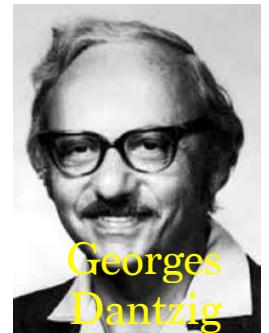
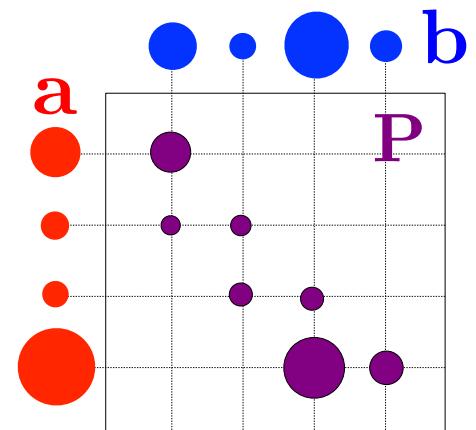
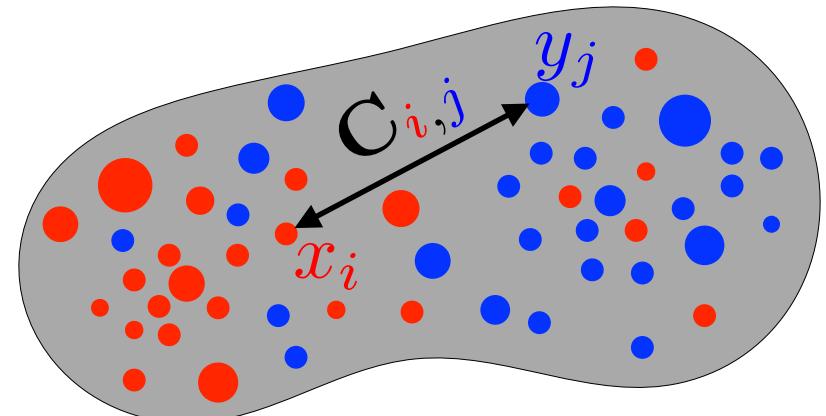
Couplings:

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

[Kantorovich 1942]

$$\min \left\{ \sum_{i,j} C_{i,j} P_{i,j} ; \mathbf{P} \in \mathbf{U}(\mathbf{a}, \mathbf{b}) \right\}$$

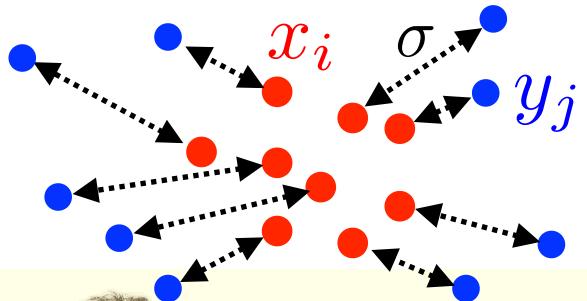
→ Linear program, simplex $O(n^3 \log(n)).$



Georges
Dantzig

Kantorovitch's Exact Relaxation

$$\alpha = \sum_{i=1}^n \delta_{x_i} \quad \beta = \sum_{j=1}^n \delta_{y_j}$$

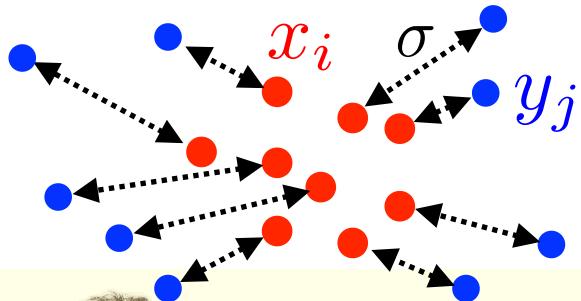


Monge (1784):

$$\min_{\sigma \in \text{Perm}_n} \sum_{i=1}^n C_{i, \sigma(i)}$$

Kantorovitch's Exact Relaxation

$$\alpha = \sum_{i=1}^n \delta_{x_i} \quad \beta = \sum_{j=1}^n \delta_{y_j}$$



Monge (1784):

$$\min_{\sigma \in \text{Perm}_n} \sum_{i=1}^n C_{i, \sigma(i)}$$

Permutations “ \subset ” Bi-stochastic matrices:

$$\text{Bist}_n \stackrel{\text{def.}}{=} \{ \mathbf{P} \in \mathbb{R}_+^{n \times n} ; \mathbf{P}\mathbf{1} = \mathbf{1}, \mathbf{P}^\top \mathbf{1} = \mathbf{1} \}$$

\gg (relaxation)

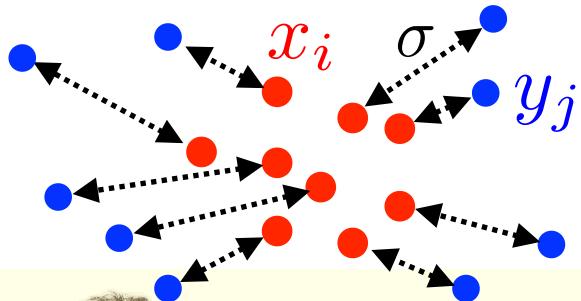


Kantorovitch (1942):

$$\min_{\mathbf{P} \in \text{Bist}_n} \sum_{i=1}^n \sum_{j=1}^n \mathbf{P}_{i,j} C_{i,j}$$

Kantorovitch's Exact Relaxation

$$\alpha = \sum_{i=1}^n \delta_{x_i} \quad \beta = \sum_{j=1}^n \delta_{y_j}$$



Monge (1784):

$$\min_{\sigma \in \text{Perm}_n} \sum_{i=1}^n C_{i, \sigma(i)}$$

$n!$ permutations

Permutations “ \subset ” Bi-stochastic matrices:

$$\text{Bist}_n \stackrel{\text{def.}}{=} \{ \mathbf{P} \in \mathbb{R}_+^{n \times n} ; \mathbf{P}\mathbf{1} = \mathbf{1}, \mathbf{P}^\top \mathbf{1} = \mathbf{1} \}$$



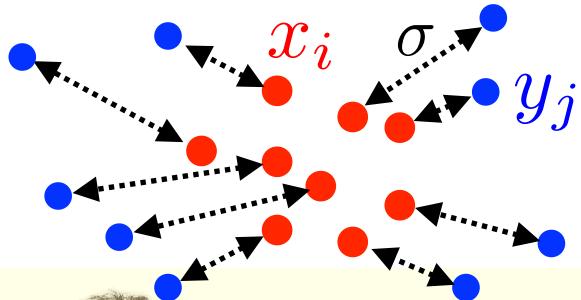
Kantorovitch (1942):

$$\min_{\mathbf{P} \in \text{Bist}_n} \sum_{i=1}^n \sum_{j=1}^n \mathbf{P}_{i,j} C_{i,j}$$

$O(n^3)$ algorithm

Kantorovitch's Exact Relaxation

$$\alpha = \sum_{i=1}^n \delta_{x_i} \quad \beta = \sum_{j=1}^n \delta_{y_j}$$



Monge (1784):

$$\min_{\sigma \in \text{Perm}_n} \sum_{i=1}^n C_{i, \sigma(i)}$$

Permutations “ \subset ” Bi-stochastic matrices:

$$\text{Bist}_n \stackrel{\text{def.}}{=} \{ \mathbf{P} \in \mathbb{R}_+^{n \times n} ; \mathbf{P}\mathbf{1} = \mathbf{1}, \mathbf{P}^\top \mathbf{1} = \mathbf{1} \}$$



Kantorovitch (1942):

$$\min_{\mathbf{P} \in \text{Bist}_n} \sum_{i=1}^n \sum_{j=1}^n \mathbf{P}_{i,j} C_{i,j}$$

\gg (relaxation)

$n!$ permutations

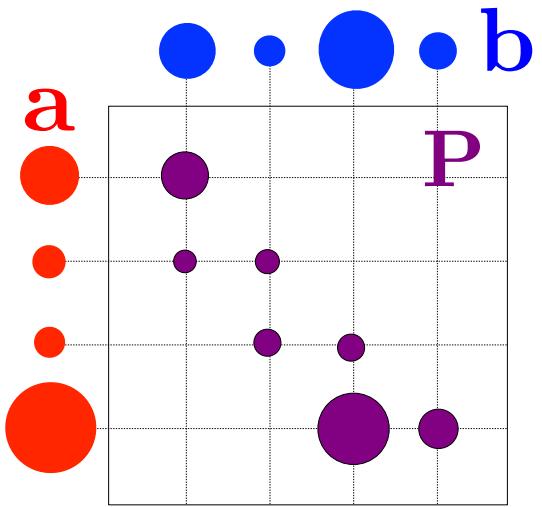
$O(n^3)$ algorithm



Theorem: [Birkhoff-von Neumann]

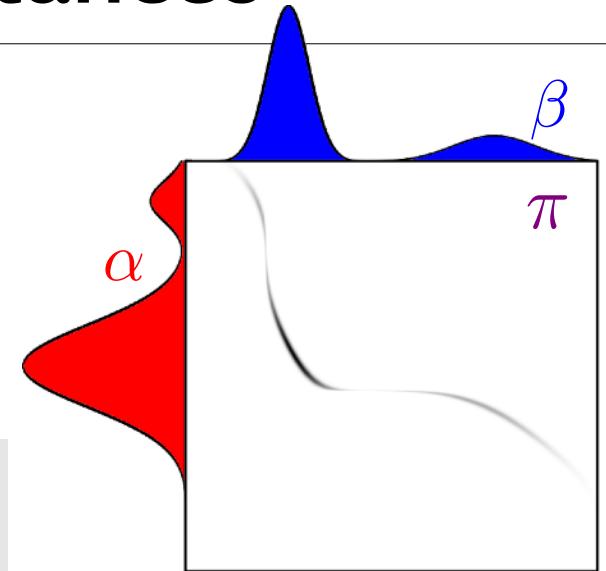
“Monge \Leftrightarrow Kantorovitch”

Optimal Transport Distances



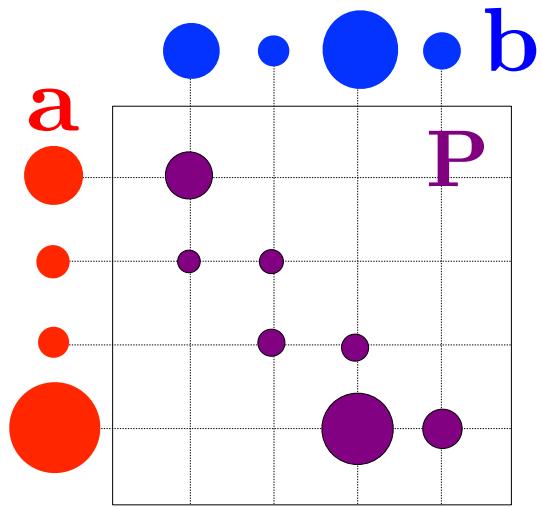
$$\pi = \sum_{i,j} \mathbf{P}_{i,j} \delta_{x_i, y_j}$$

$$c(x, y) = d(x, y)^p$$



$$W_p(\alpha, \beta)^p \stackrel{\text{def.}}{=} \min_{\pi \in \mathcal{M}_+^1(\mathcal{X}^2)} \left\{ \int_{\mathcal{X}^2} d(x, y)^p d\pi(x, y) ; \pi_1 = \alpha, \pi_2 = \beta \right\}$$

Optimal Transport Distances



$$\pi = \sum_{i,j} P_{i,j} \delta_{x_i, y_j}$$

$$c(x, y) = d(x, y)^p$$

$$W_p(\alpha, \beta)^p \stackrel{\text{def.}}{=} \min_{\pi \in \mathcal{M}_+^1(\mathcal{X}^2)} \left\{ \int_{\mathcal{X}^2} d(x, y)^p d\pi(x, y) ; \pi_1 = \alpha, \pi_2 = \beta \right\}$$

Theorem: W_p is a distance and $\alpha_n \rightarrow \beta \Leftrightarrow W_p(\alpha_n, \beta) \rightarrow 0$

Weak* (aka in law) convergence: $\alpha_n \rightharpoonup \beta \Leftrightarrow \forall f \in \mathcal{C}(\mathcal{X}), \int_{\mathcal{X}} f d\alpha_n \rightarrow \int_{\mathcal{X}} f d\beta$



$$\|\delta_{x_n} - \delta_x\|_1 = 2 \quad \text{vs.} \quad W_p(\delta_{x_n}, \delta_x) = d(x_n, x)$$

Overview

- Monge Formulation
- Continuous Optimal Transport
- Kantorovitch Formulation
- Applications

Wasserstein Barycenters

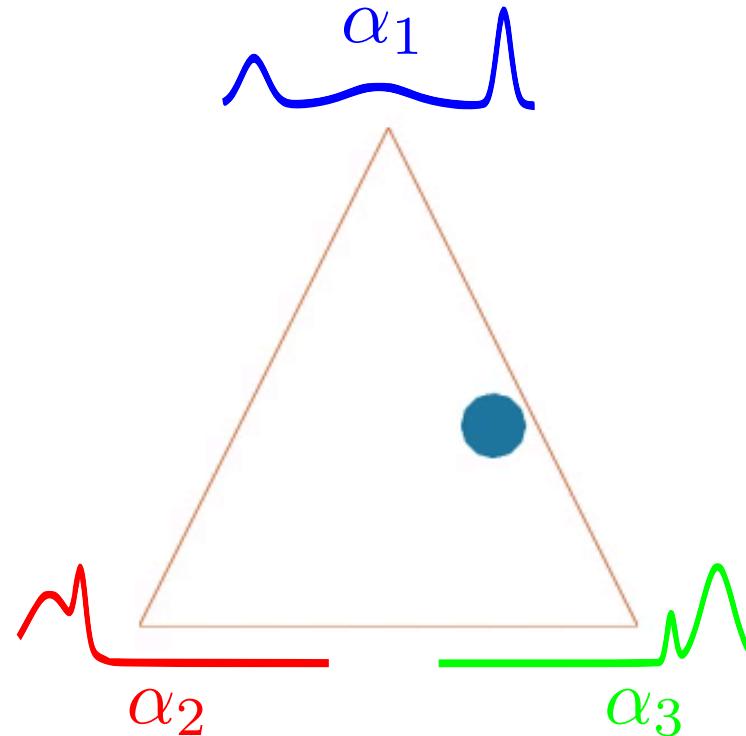
Barycenters of measures $(\alpha_s)_{s=1}^S$: $\sum_s \lambda_s = 1$

$$\alpha^* \in \operatorname{argmin}_{\alpha} \sum_s \lambda_s W_p^p(\alpha, \alpha_s)$$



Guillaume Carlier

Martial Aguech



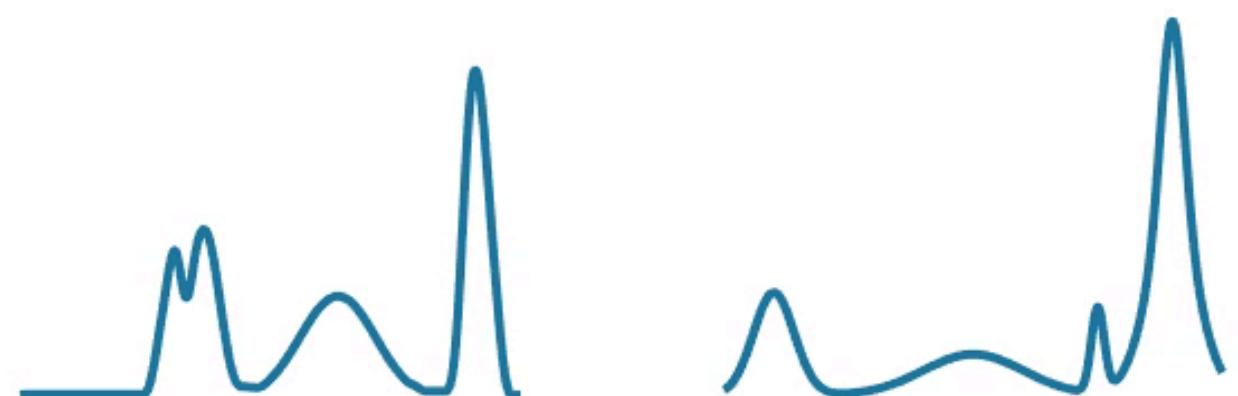
$$\lambda \in \Sigma_3$$

$$\min_{\alpha} \sum_s \lambda_s W_p^p(\alpha, \alpha_s)$$

Wasserstein

$$\sum_s \lambda_s \alpha_s$$

Euclidean



Wasserstein Barycenters

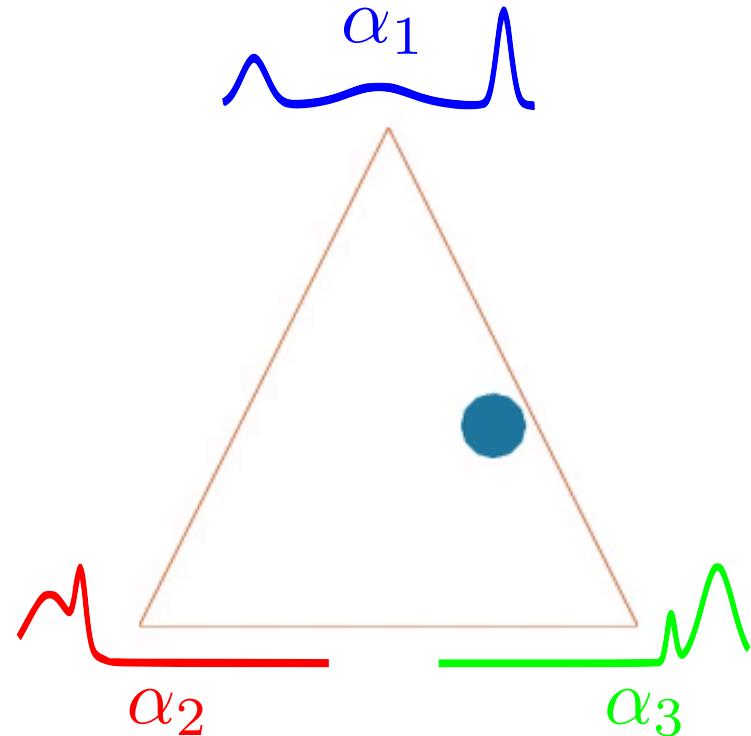
Barycenters of measures $(\alpha_s)_{s=1}^S$: $\sum_s \lambda_s = 1$

$$\alpha^* \in \operatorname{argmin}_{\alpha} \sum_s \lambda_s W_p^p(\alpha, \alpha_s)$$



Guillaume Carlier

Martial Aguech



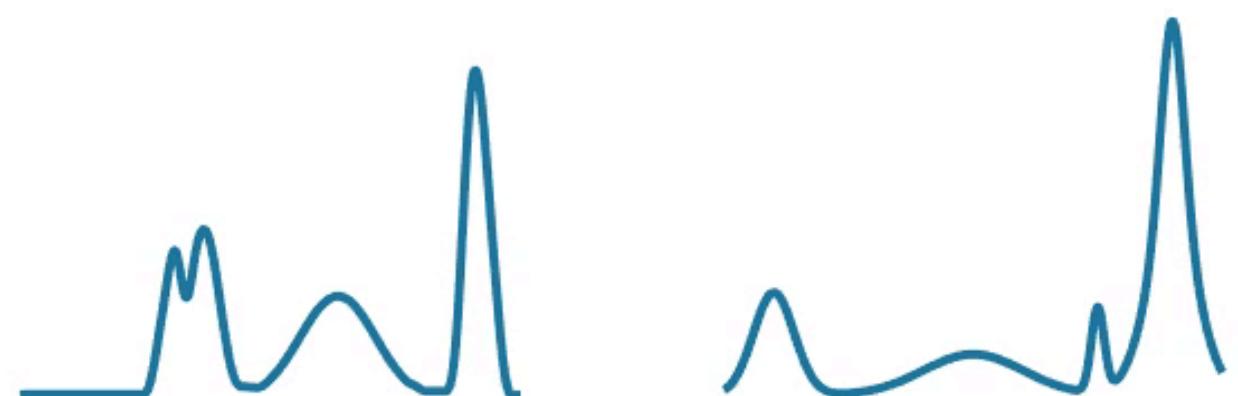
$$\lambda \in \Sigma_3$$

$$\min_{\alpha} \sum_s \lambda_s W_p^p(\alpha, \alpha_s)$$

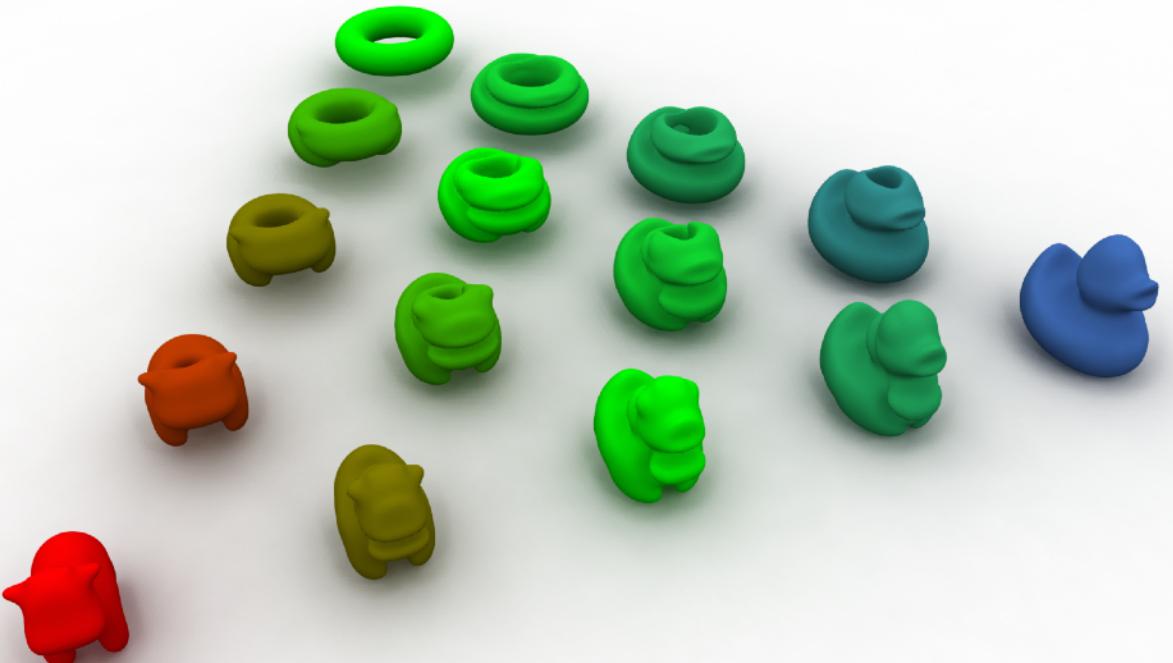
Wasserstein

$$\sum_s \lambda_s \alpha_s$$

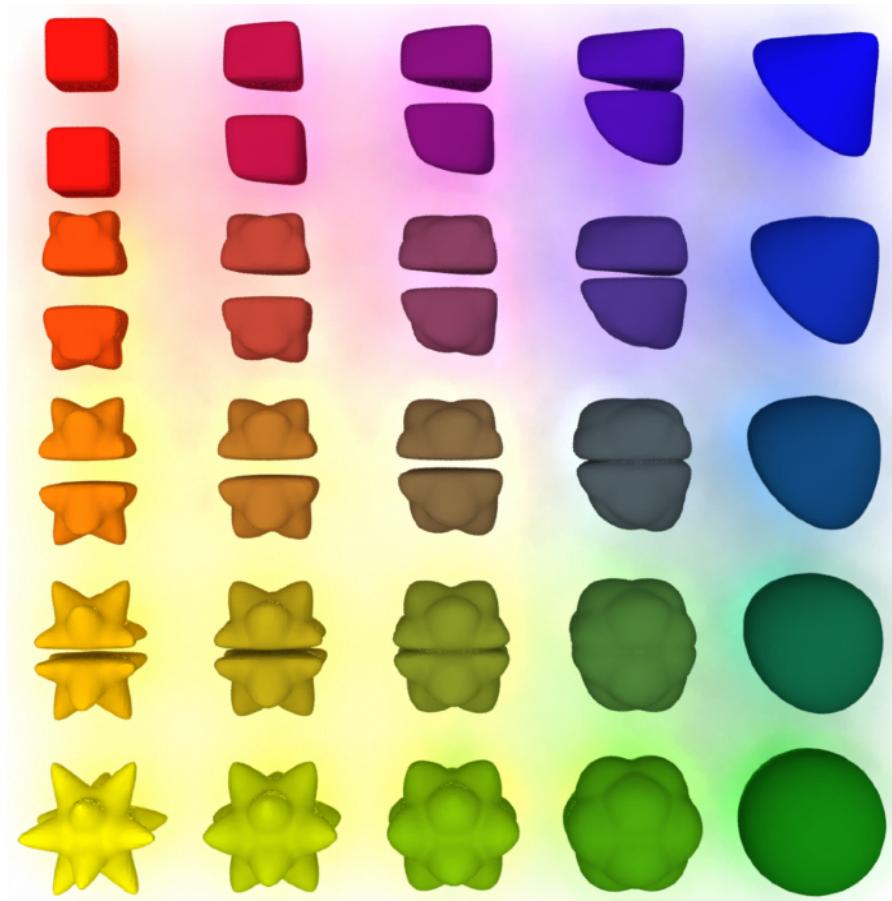
Euclidean



Examples

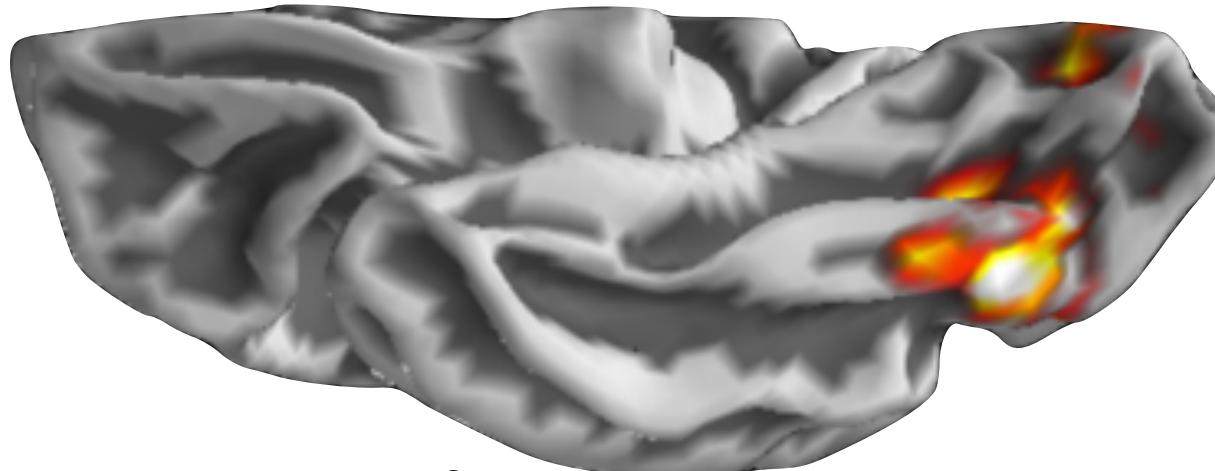


[Solomon et al, SIGGRAPH 2015]

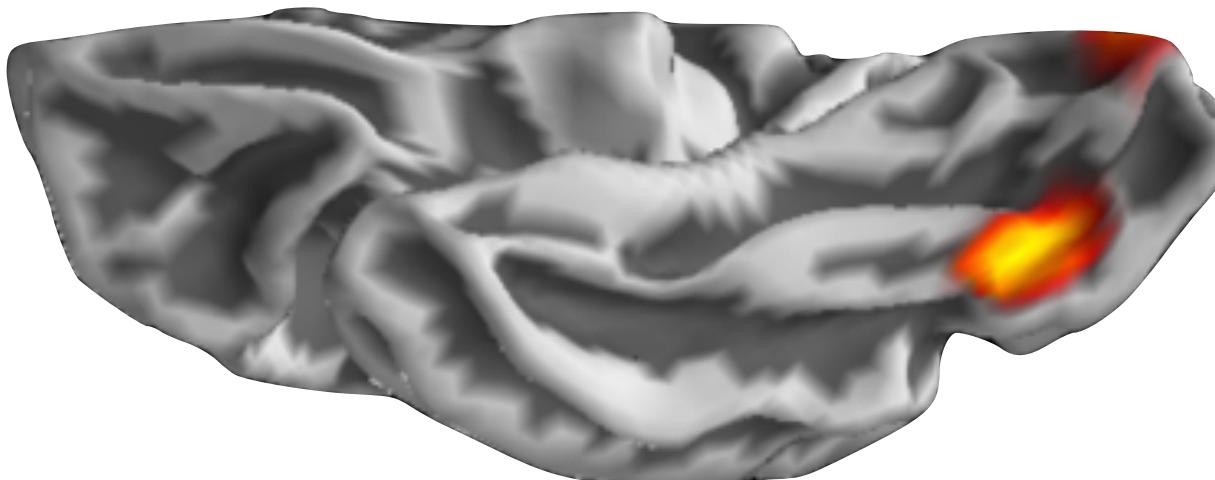


MRI Data Processing [with A. Gramfort]

Ground cost $c = d_M$: geodesic on cortical surface M .



L^2 barycenter

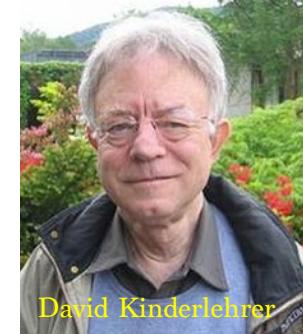
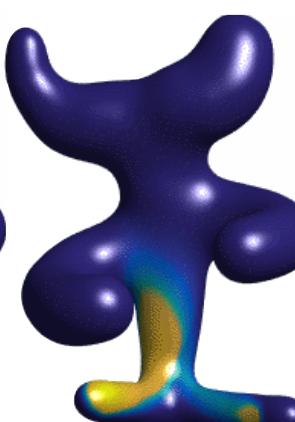
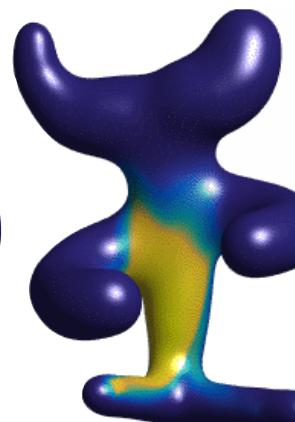
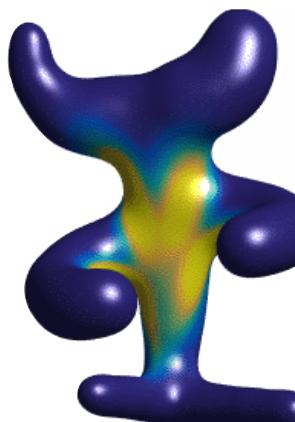
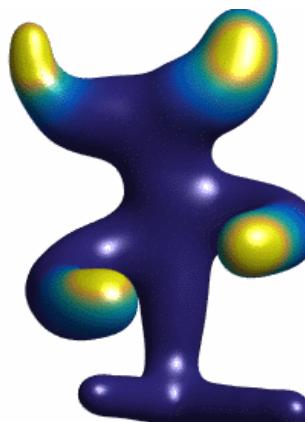
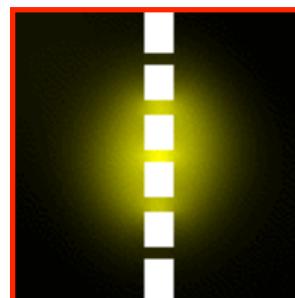
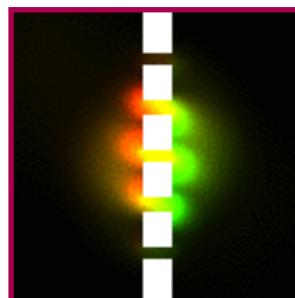
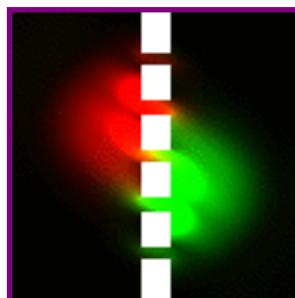
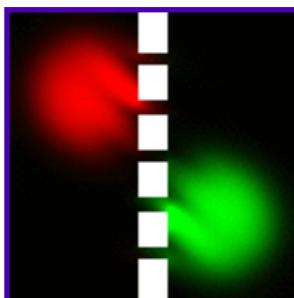
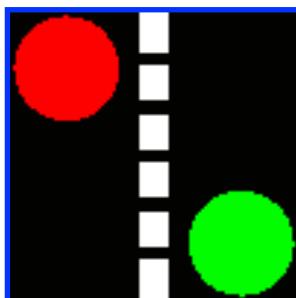
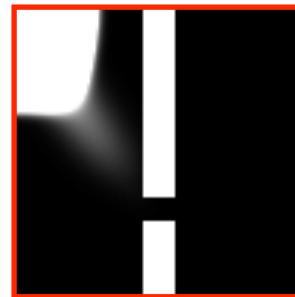
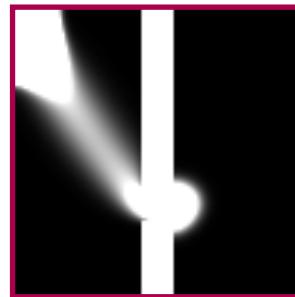
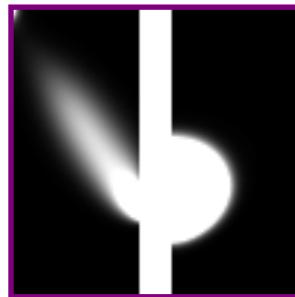
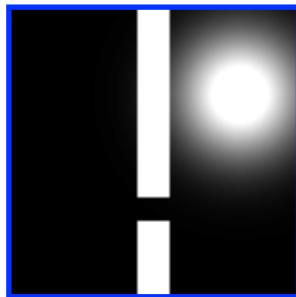


W_2^2 barycenter

Generalizations: Gradient Flows

Implicit stepping: $\alpha_{t+\tau} = \operatorname{argmin}_{\alpha} W_p^p(\alpha_t, \alpha) + \tau f(\alpha)$

Limit $\tau \rightarrow 0$: $\frac{\partial \alpha}{\partial t} = \operatorname{div}(\alpha \nabla(f'(\alpha)))$

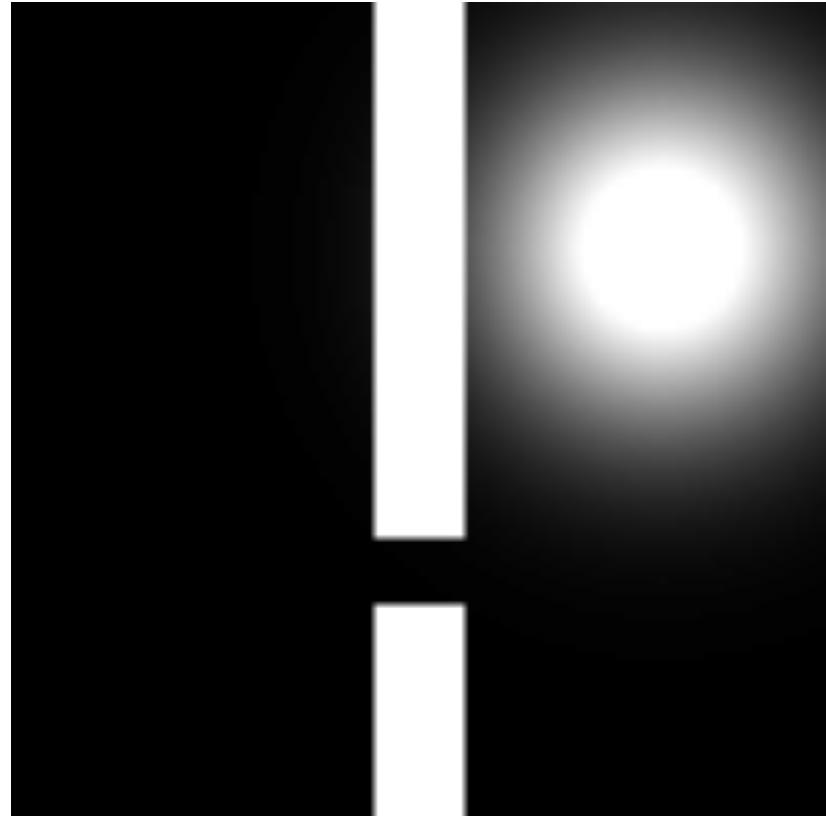


Gradient Flows Simulation



<https://www.youtube.com/watch?v=tDQw21ntR64>

Tim Whittaker (New Zealand)



Gradient Flows Simulation



<https://www.youtube.com/watch?v=tDQw21ntR64>

Tim Whittaker (New Zealand)



Bag of Words



[Kusner'15] $\text{dist}(D_1, D_2) = W_2(\mu, \nu)$

Topic Models

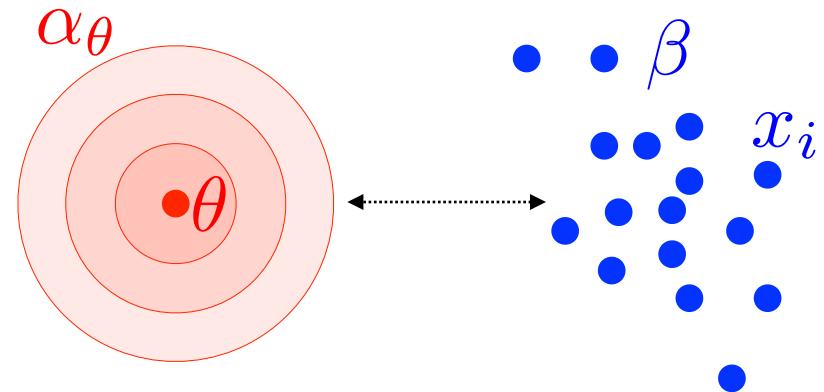


[Rolet'16]

Density Fitting and Generative Models

Observations: $\beta \stackrel{\text{def.}}{=} \frac{1}{n} \sum_{i=1}^n \delta_{x_i}$

Parametric model: $\theta \mapsto \alpha_\theta$



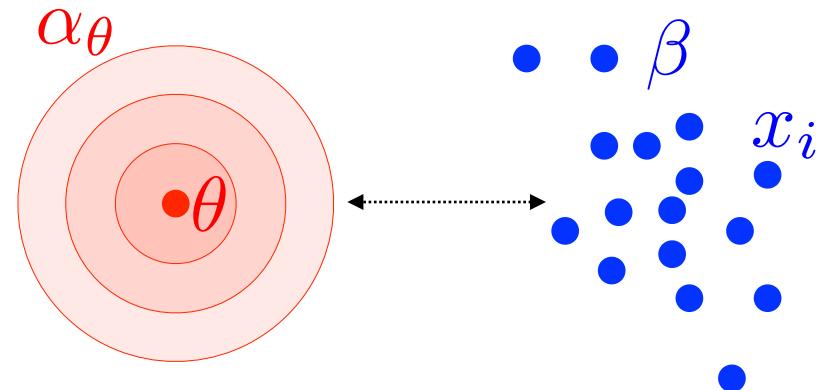
Density Fitting and Generative Models

Observations: $\beta \stackrel{\text{def.}}{=} \frac{1}{n} \sum_{i=1}^n \delta_{x_i}$

Parametric model: $\theta \mapsto \alpha_\theta$

Density fitting: $d\alpha_\theta(x) = \rho_\theta(x)dx$

$$\min_{\theta} \widehat{\text{KL}}(\alpha_\theta | \beta) \stackrel{\text{def.}}{=} - \sum_i \log(\rho_\theta(x_i))$$

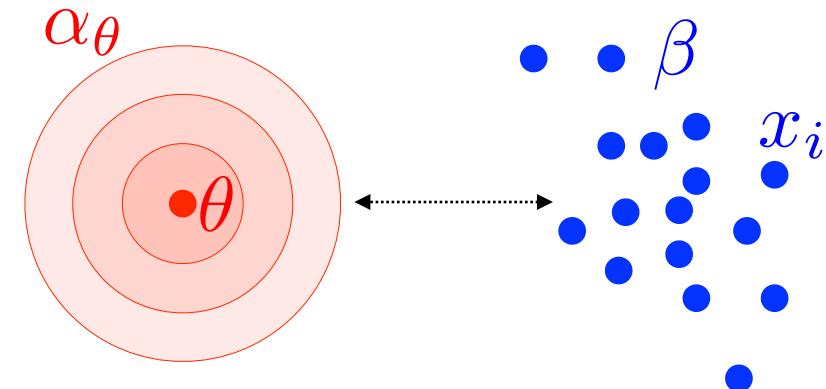


Maximum likelihood (MLE)

Density Fitting and Generative Models

Observations: $\beta \stackrel{\text{def.}}{=} \frac{1}{n} \sum_{i=1}^n \delta_{x_i}$

Parametric model: $\theta \mapsto \alpha_\theta$



Density fitting: $d\alpha_\theta(x) = \rho_\theta(x)dx$

$$\min_{\theta} \widehat{\text{KL}}(\alpha_\theta | \beta) \stackrel{\text{def.}}{=} - \sum_i \log(\rho_\theta(x_i))$$

Maximum likelihood (MLE)

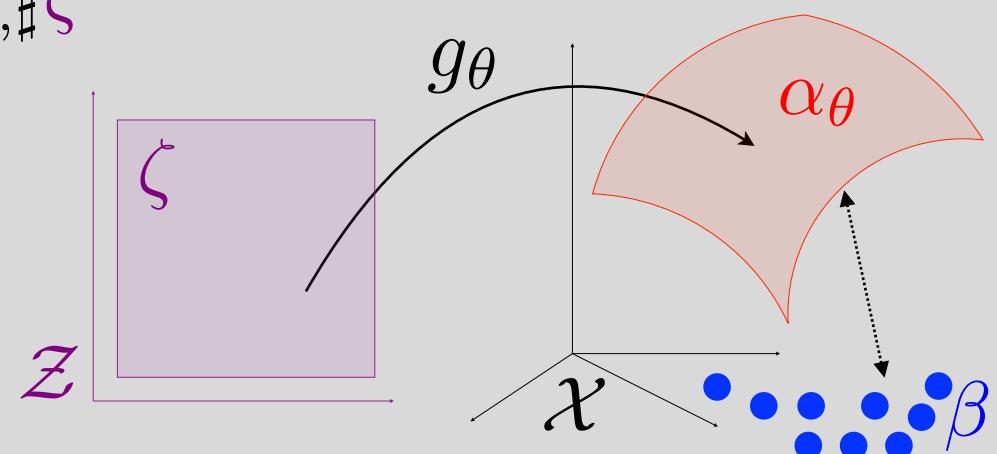
Generative model fit: $\alpha_\theta = g_{\theta, \sharp} \zeta$

$$\widehat{\text{KL}}(\alpha_\theta | \beta) = +\infty$$

→ MLE undefined.

→ Need a weaker metric.

$$\min_{\theta} \overline{W}_{\varepsilon, p}^p(\alpha_\theta, \beta)$$



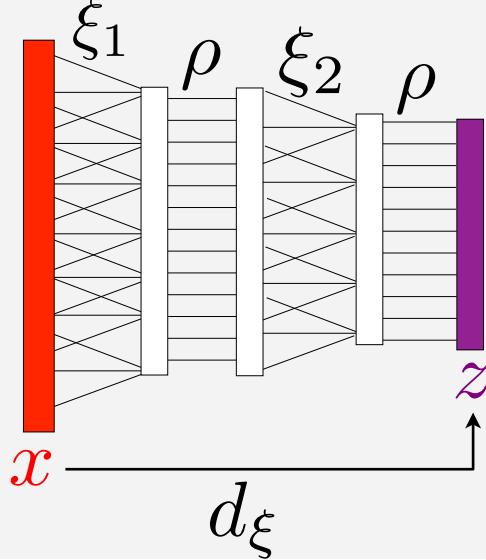
Deep Discriminative vs Generative Models

Deep networks:

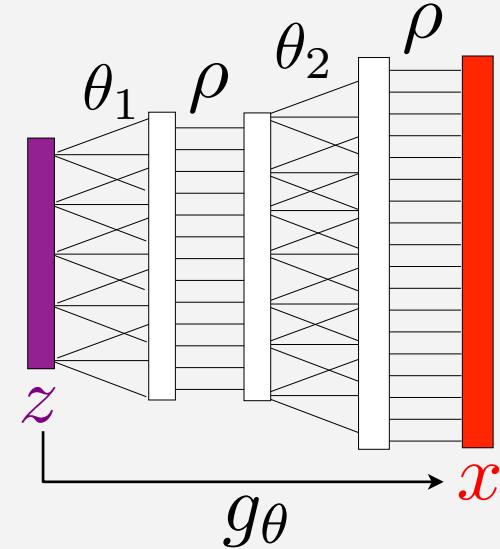
$$d_\xi(\textcolor{red}{x}) = \rho(\xi_K(\dots \rho(\xi_2(\rho(\xi_1(\textcolor{red}{x}) \dots)$$

$$g_\theta(\textcolor{violet}{z}) = \rho(\theta_K(\dots \rho(\theta_2(\rho(\theta_1(\textcolor{violet}{z}) \dots)$$

Discriminative



Generative



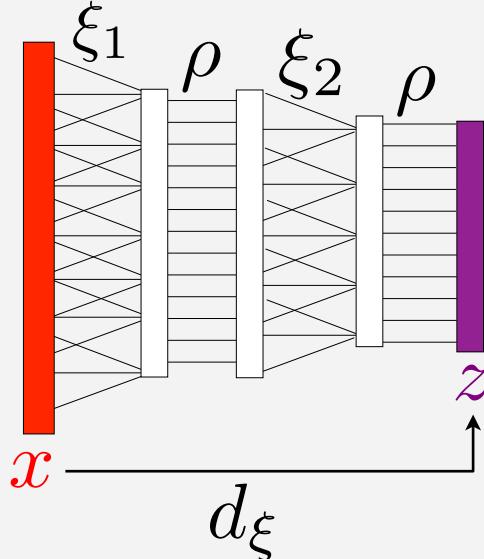
Deep Discriminative vs Generative Models

Deep networks:

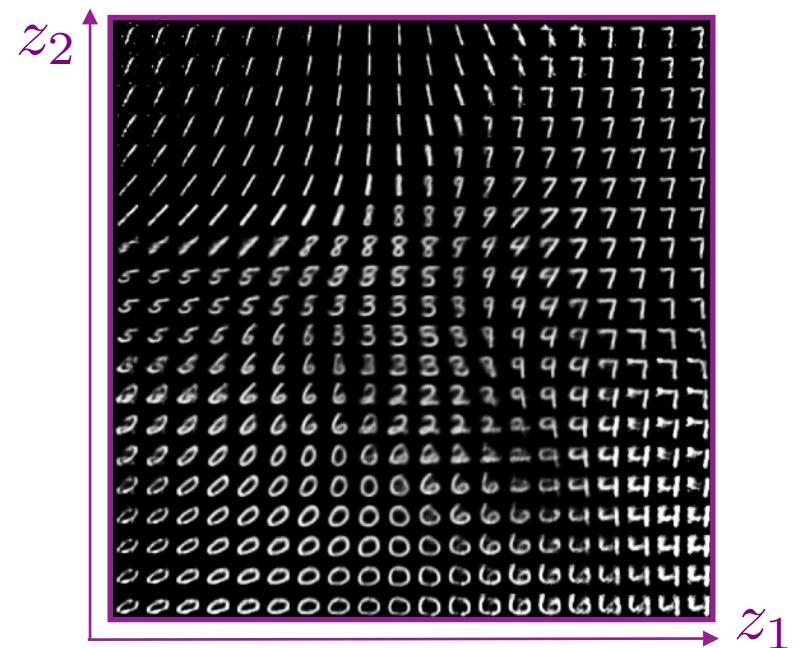
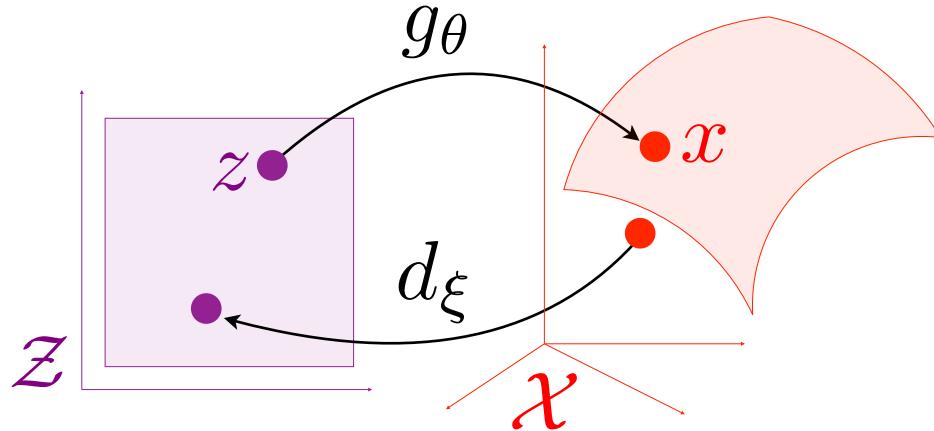
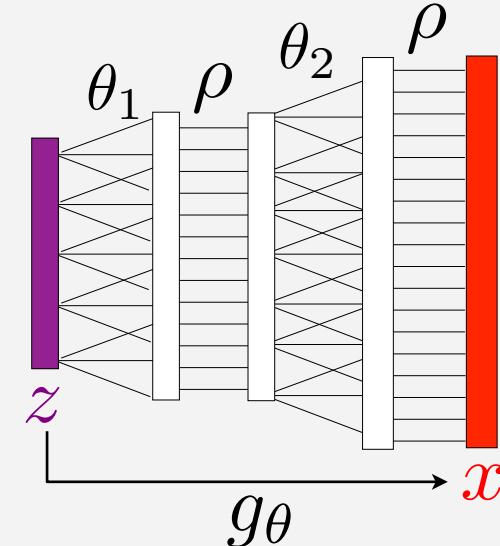
$$d_\xi(\mathbf{x}) = \rho(\xi_K(\dots \rho(\xi_2(\rho(\xi_1(\mathbf{x}) \dots)$$

$$g_\theta(\mathbf{z}) = \rho(\theta_K(\dots \rho(\theta_2(\rho(\theta_1(\mathbf{z}) \dots)$$

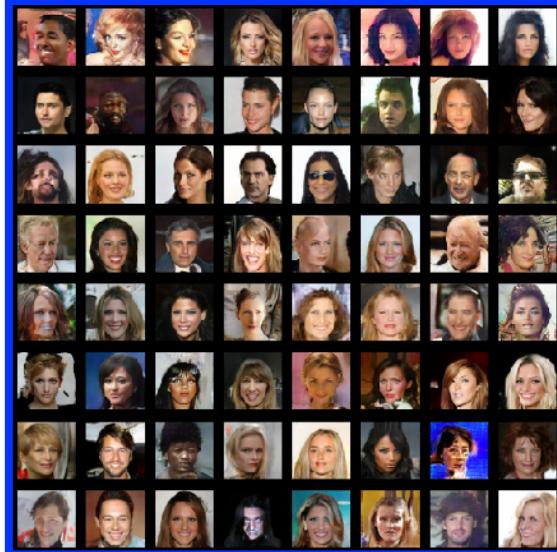
Discriminative



Generative



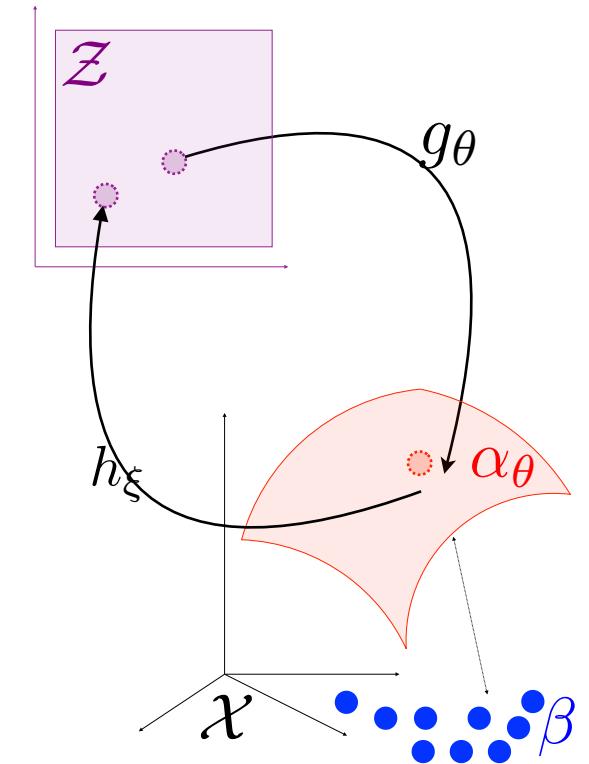
Examples of Images Generation



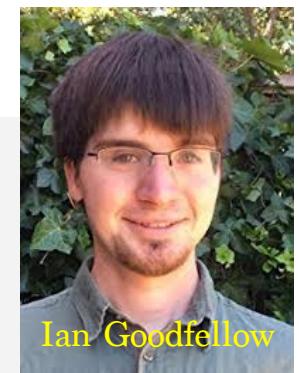
Inputs β



Generated α_θ

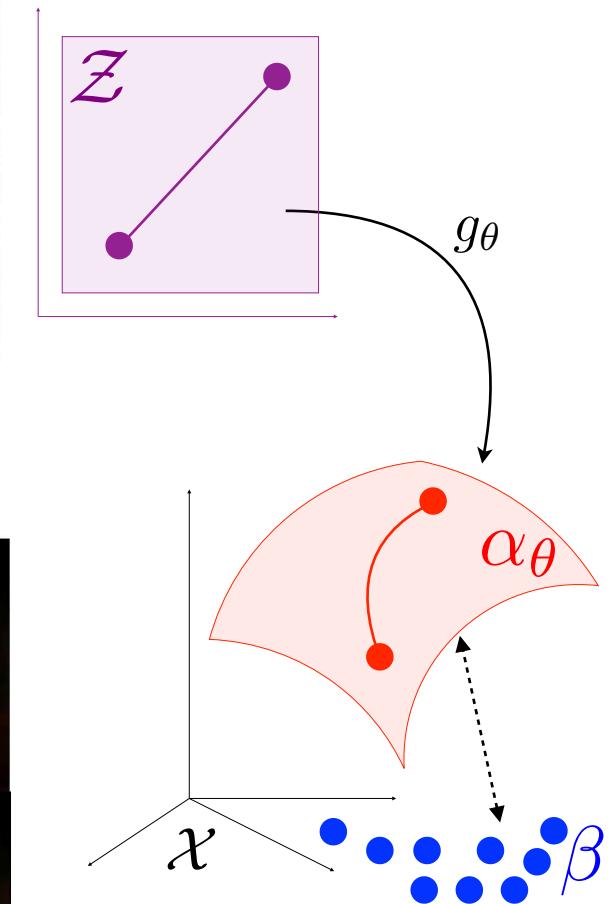


- Need to learn the metric $d(x, y) = \|h_\xi(x) - h_\xi(y)\|$ (GANs)
- Influence of ε ?
- Performance evaluation of generative models is an open problem.



Ian Goodfellow

Generative Adversarial Networks



Progressive Growing of GANs for Improved Quality, Stability, and Variation
Tero Karras, Timo Aila, Samuli Laine, Jaakko Lehtinen, ICLR 2018



Progressive Growing of GANs for Improved Quality, Stability, and Variation
Tero Karras, Timo Aila, Samuli Laine, Jaakko Lehtinen, ICLR 2018



Progressive Growing of GANs for Improved Quality, Stability, and Variation
Tero Karras, Timo Aila, Samuli Laine, Jaakko Lehtinen, ICLR 2018

Conclusion: Toward High-dimensional OT

Monge Kantorovich Dantzig Brenier Otto McCann Villani Figalli

