Optimal Transport for image processing

Julie Delon

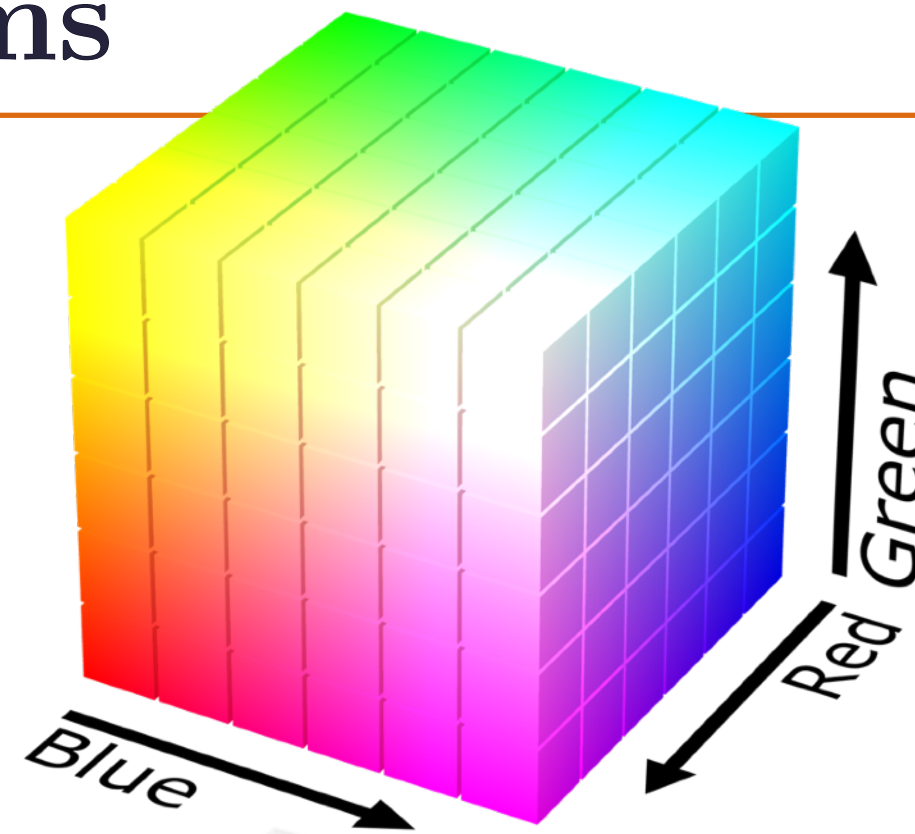
introduction à l'imagerie numérique

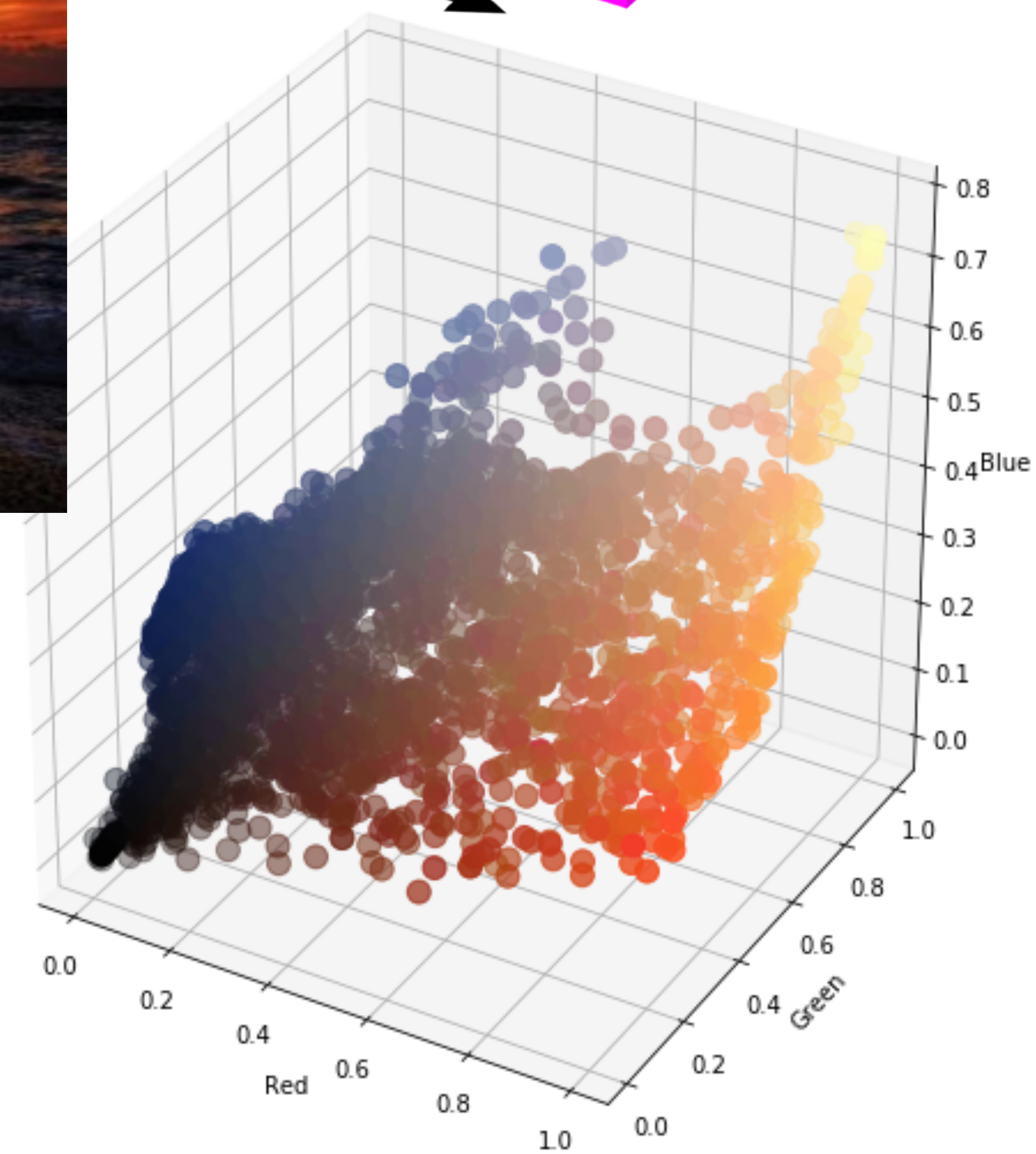
LABORATOIRE MAP5

Université de Paris

Color distributions and optimal transport

Color histograms



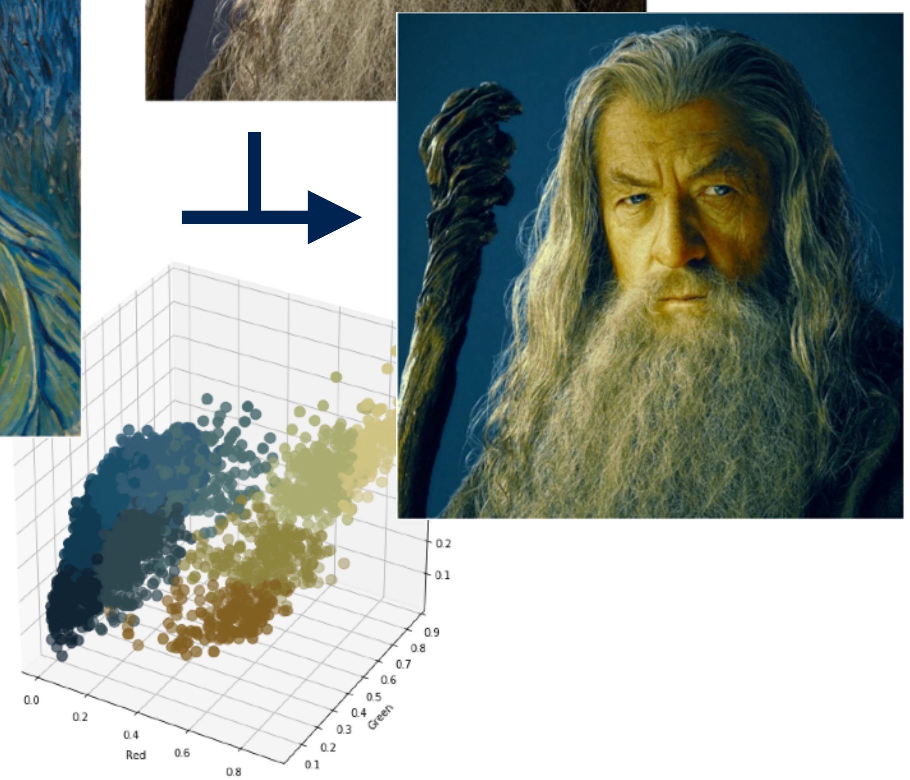


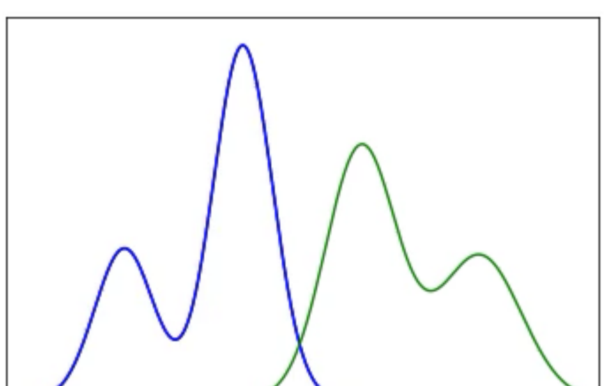
$h_u = \sum_{i \in I} s_i \delta_{\lambda_i}$ with λ_i on a fixed grid, $\sum_i s_i = 1$ and $s_i \geq 0$. (Eulerian formulation)

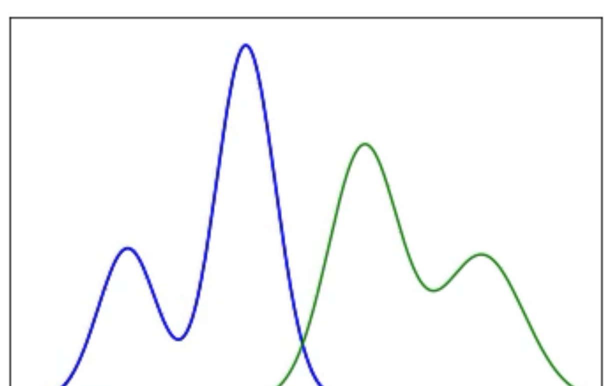
$h_u = \frac{1}{|\Omega|} \sum_{x \in \Omega} \delta_{u(x)}$ (Lagrangian formulation)

Optimal transport for color distributions

Color transfer

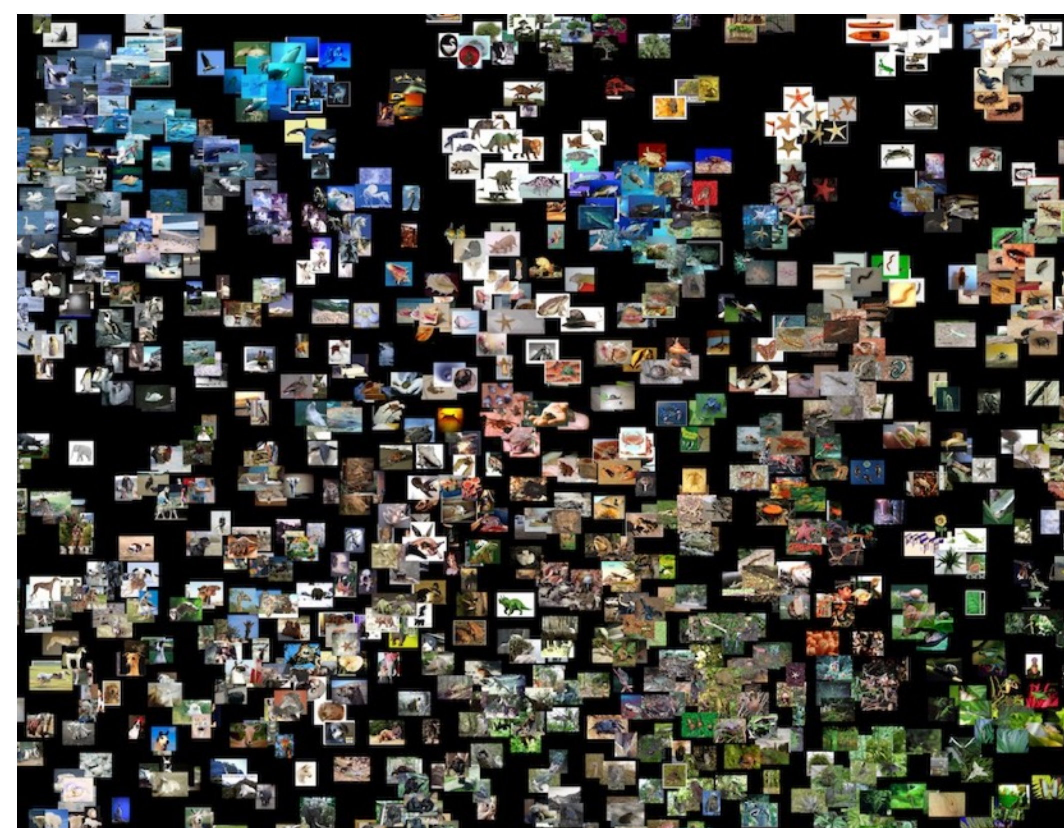






?

Color barycenter



Color based retrieval

Affine color transfer

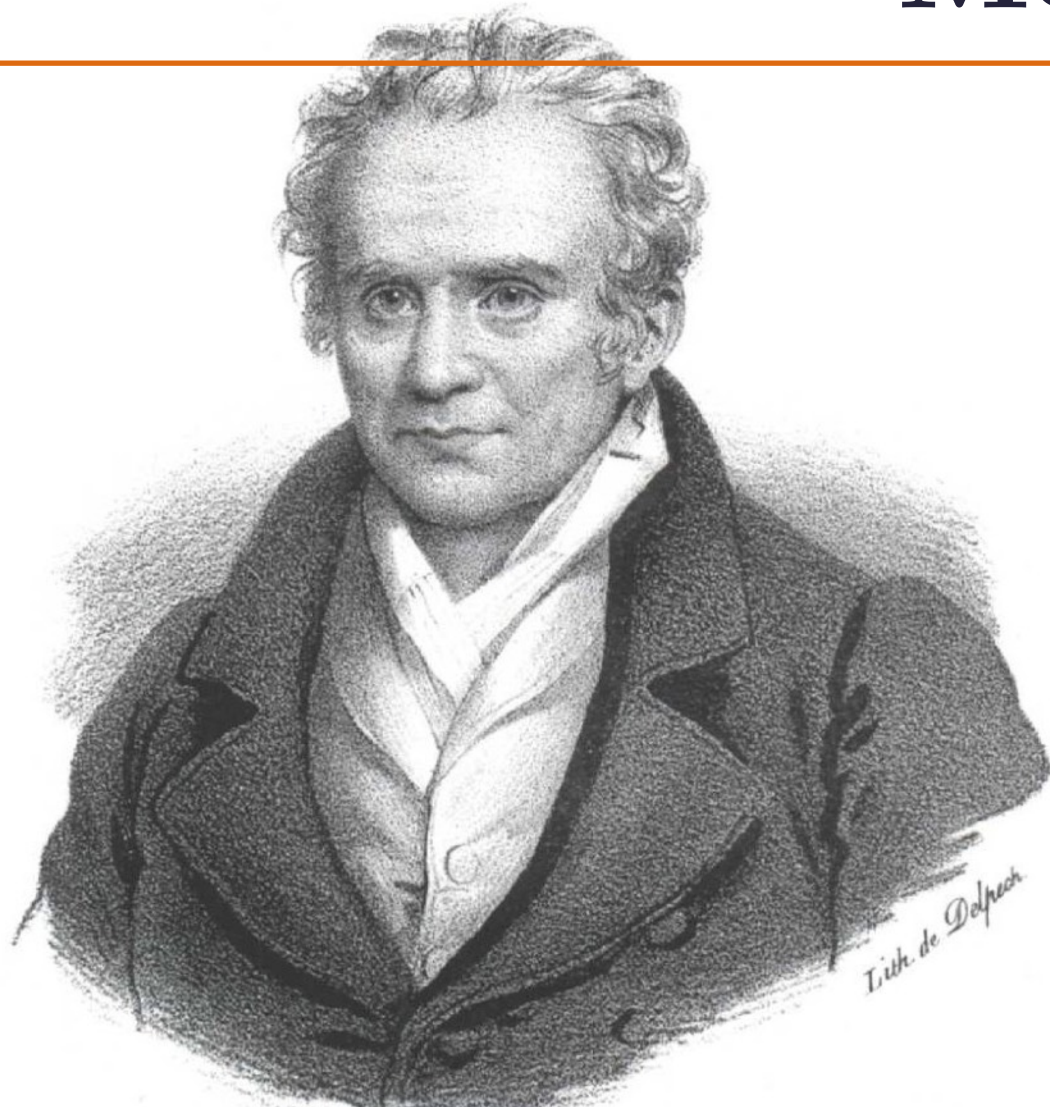
Reinhard et al. Color Transfer between Images, 2001.

Affine color transfer

Reinhard et al. Color Transfer between Images, 2001.

Reminders on Optimal transport

Monge, 1781



666· Mémoires de l'Académie Royale

MÉMOIRE

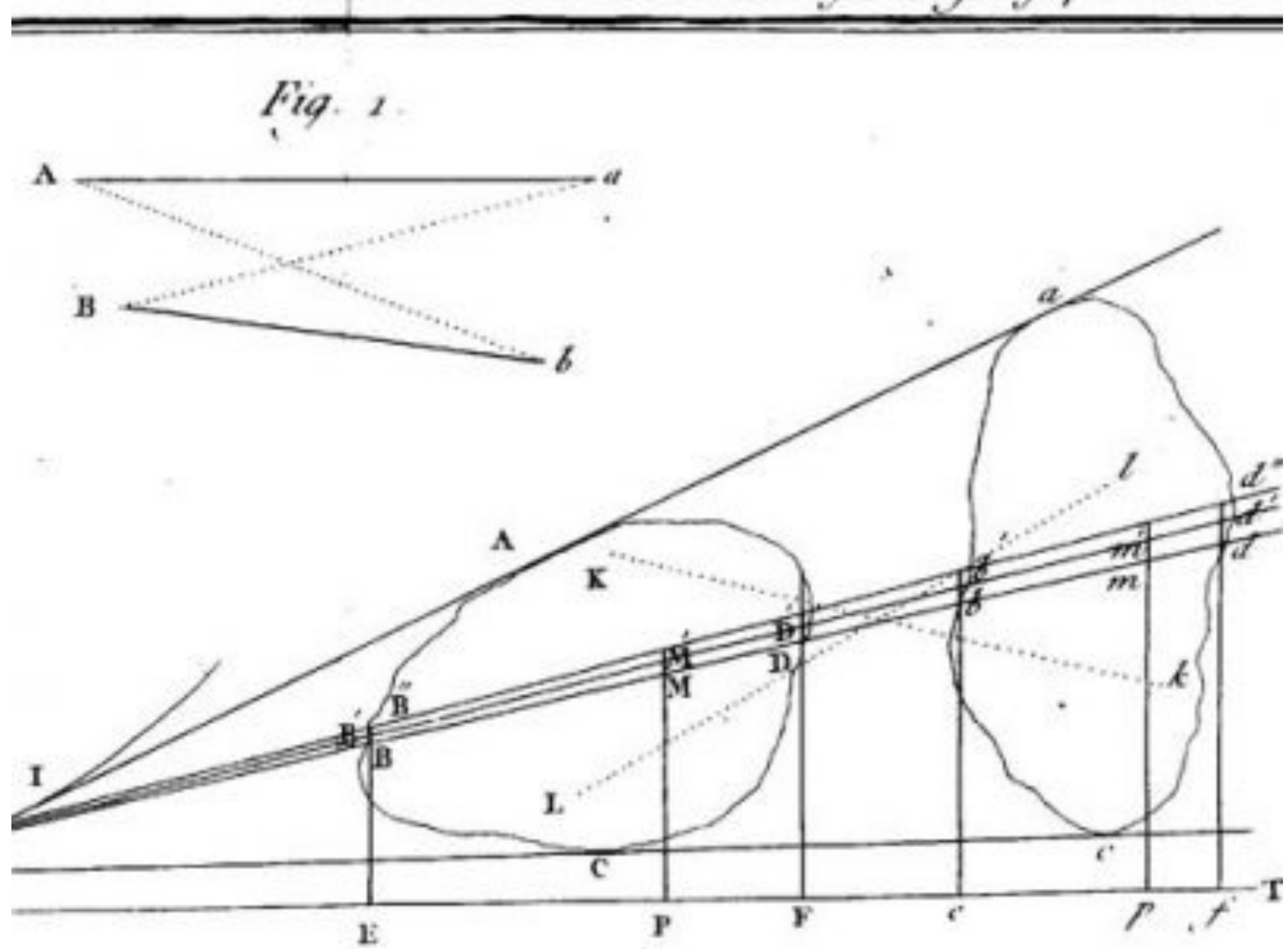
SUR LA

THÉORIE DES DÉBLAIS

ET DES REMBLAIS.

Par M. Monge.

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



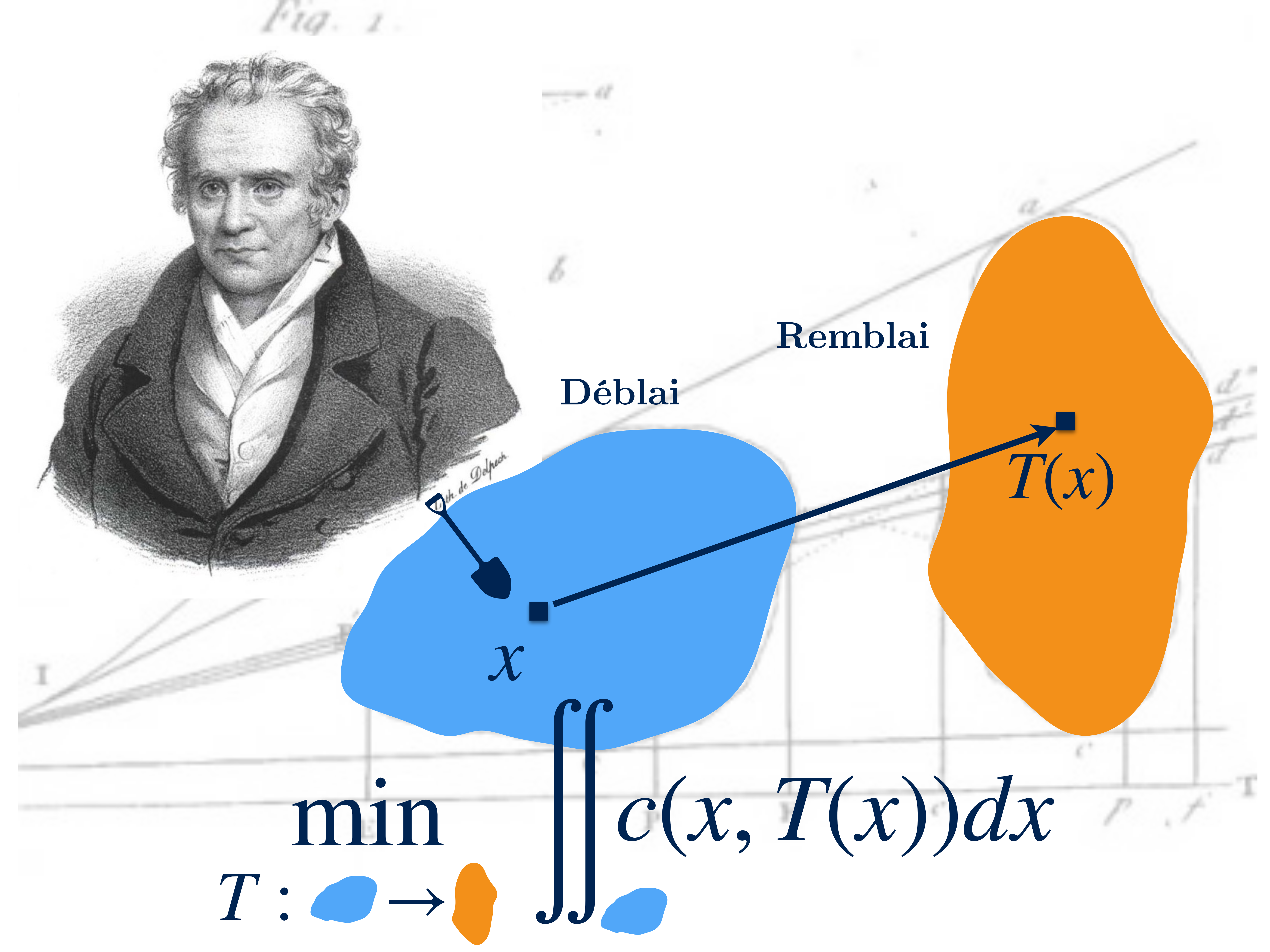


Fig. 1.
Remblai
Déblai
$T(x)$
x
$\min_{T}$
$T :$
$\iint c(x, T(x))dx$

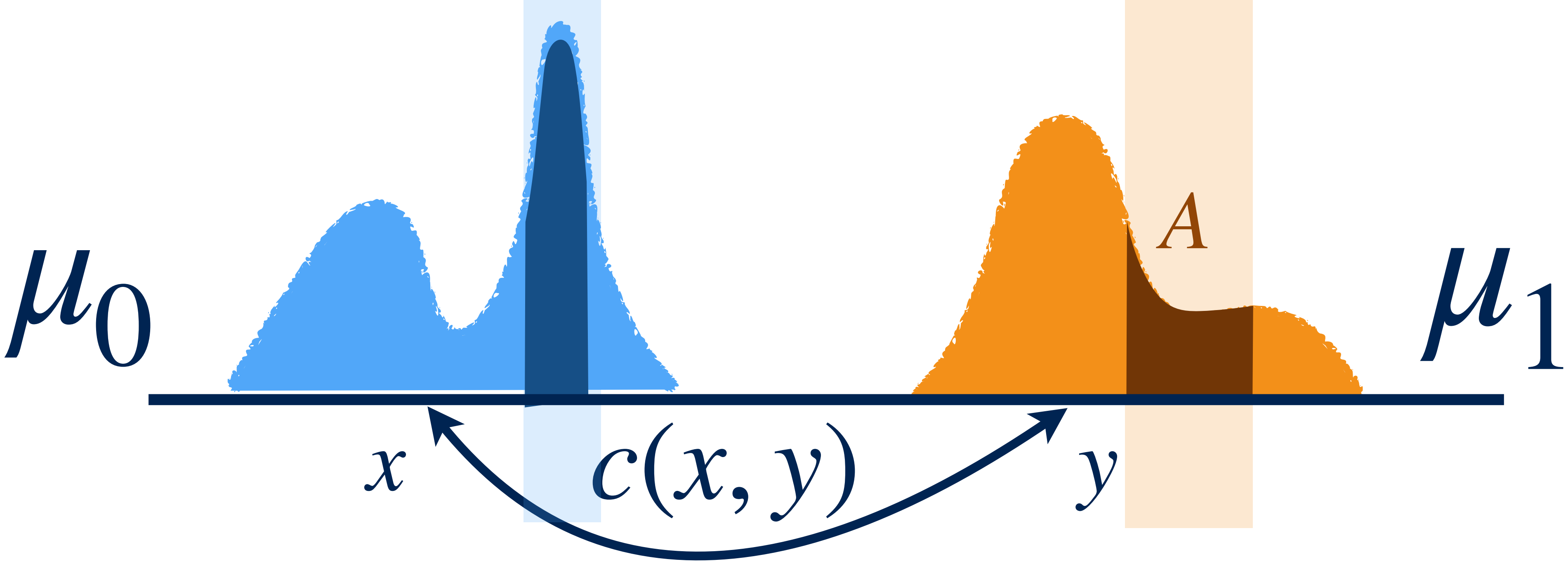
Monge optimal transport

How to transfer the mass from μ_0 to μ_1 at minimal cost?

Push forward $\mu_1 = T\#\mu_0$

$$\forall A, \quad \mu_1(A) = \mu_0(T^{-1}(A))$$

$$T^{-1}(A) = \{x,\, T(x) \in A\}$$

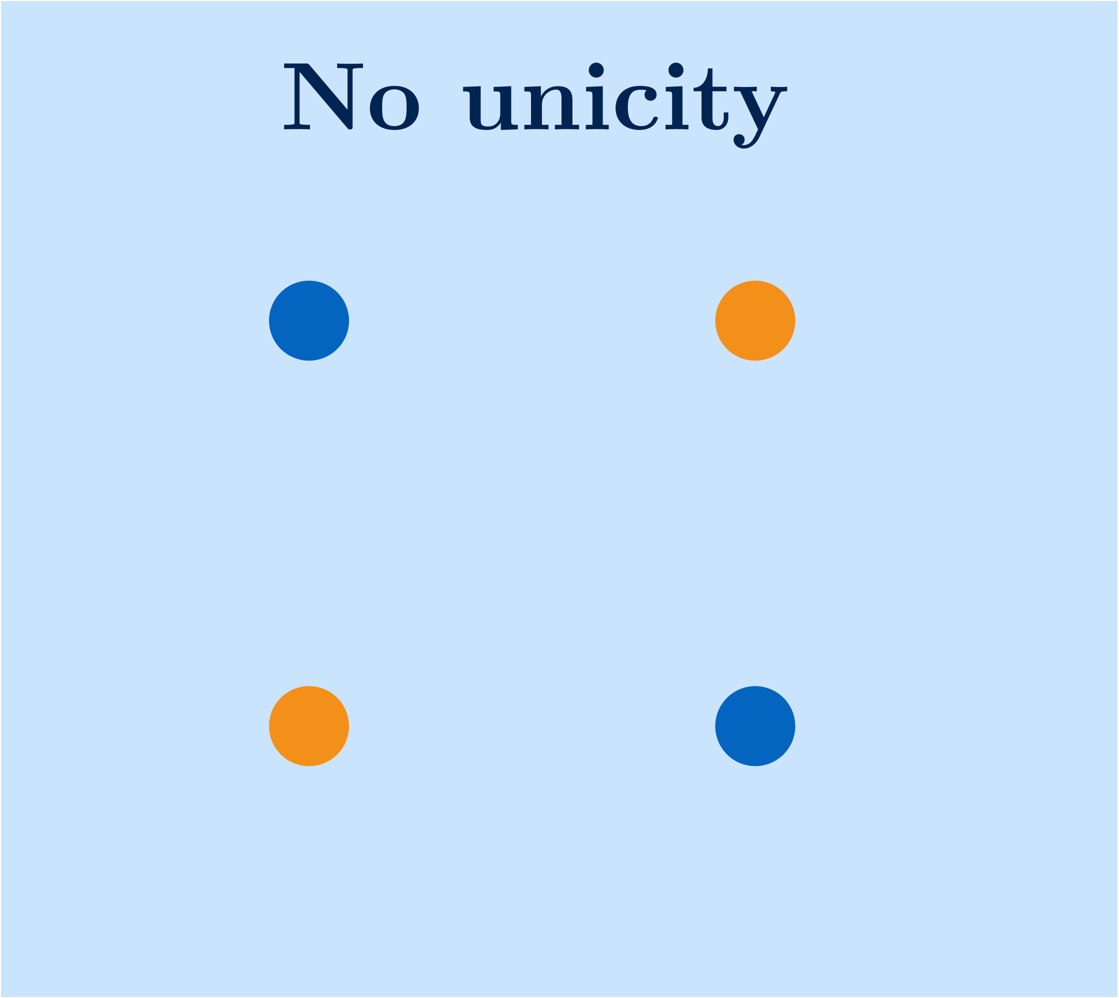


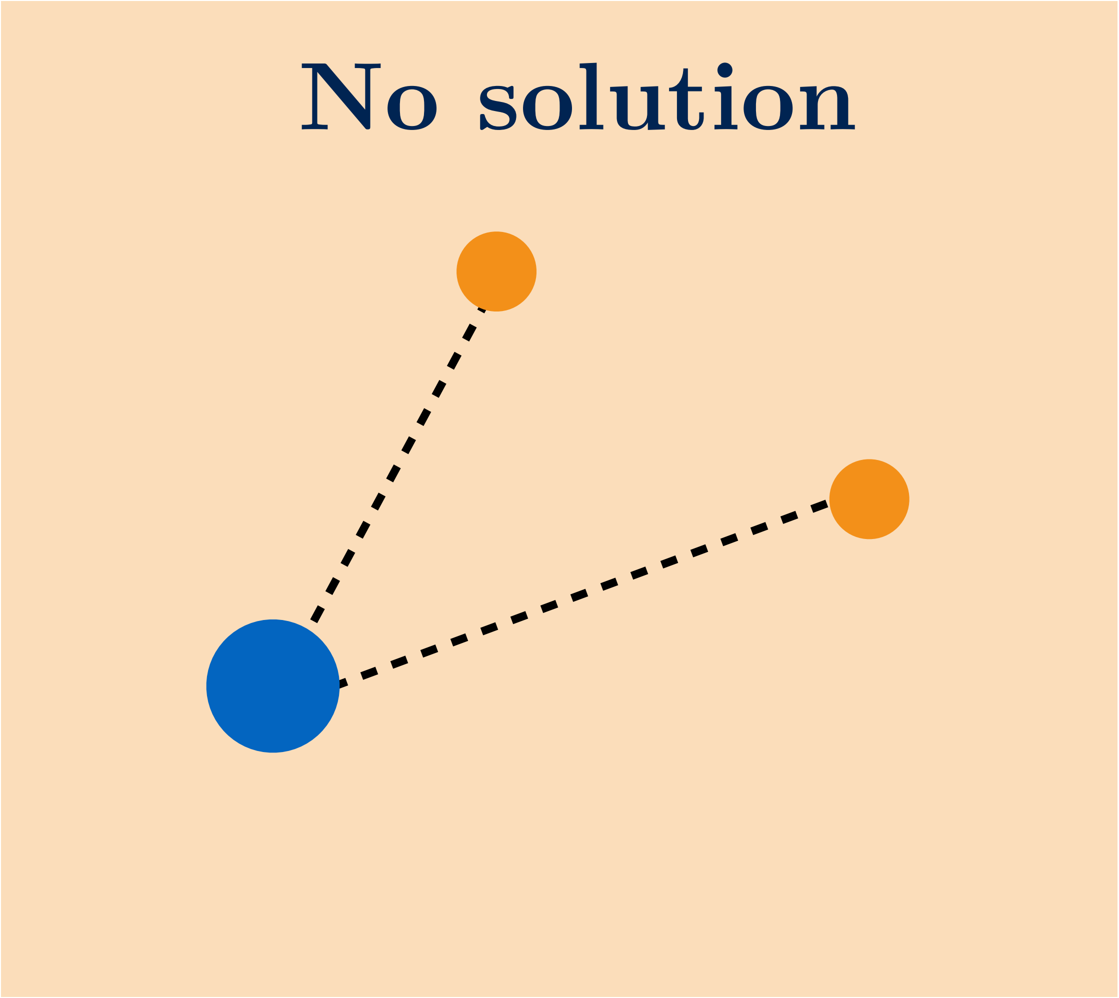
Total cost of the mass transfer = sum of costs of displacements of elementary masses.

$$\inf_{T\#\mu_0 = \mu_1} \int c(x, T(x))\, d\mu_0(x)$$

Finding T???

Difficult Problem, lack of symmetry, not convex.





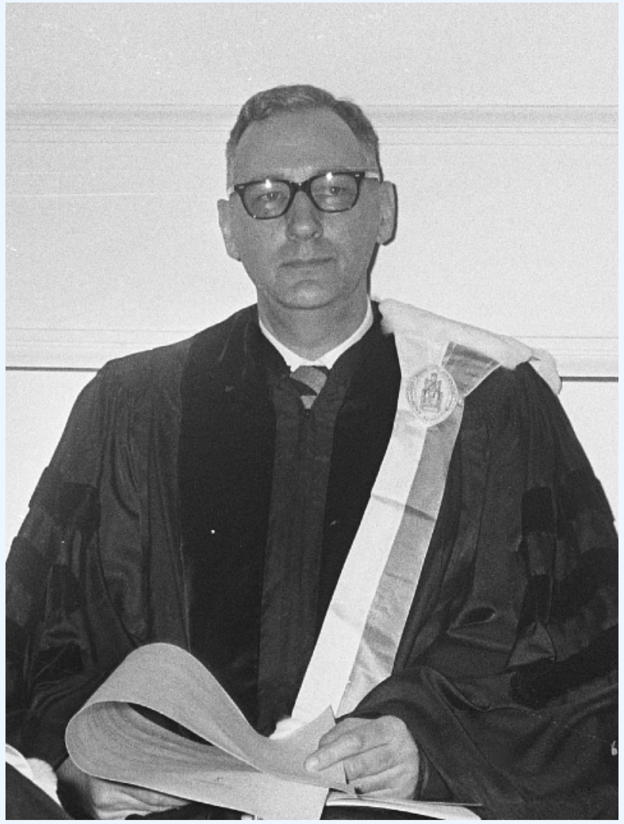
Linear programming

A.N.Tostoi, 1930

L. Kantorovich, 1939

F.L. Hitchcock, 1941

T.C. Koopmans, 1942

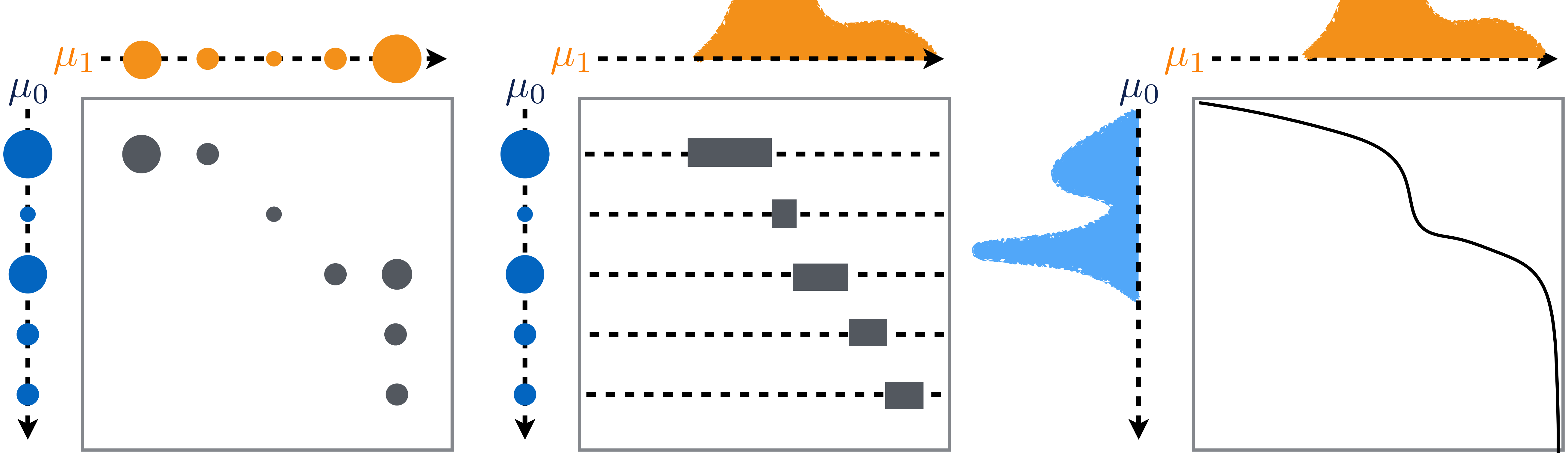


G. Dantzig, J. Von Neumann, 40's, 50's

Couplings

Couplings

$\Pi(\mu_0, \mu_1)$ = **probability distributions on** $X \times X$ **with marginals** μ_0 **and** μ_1

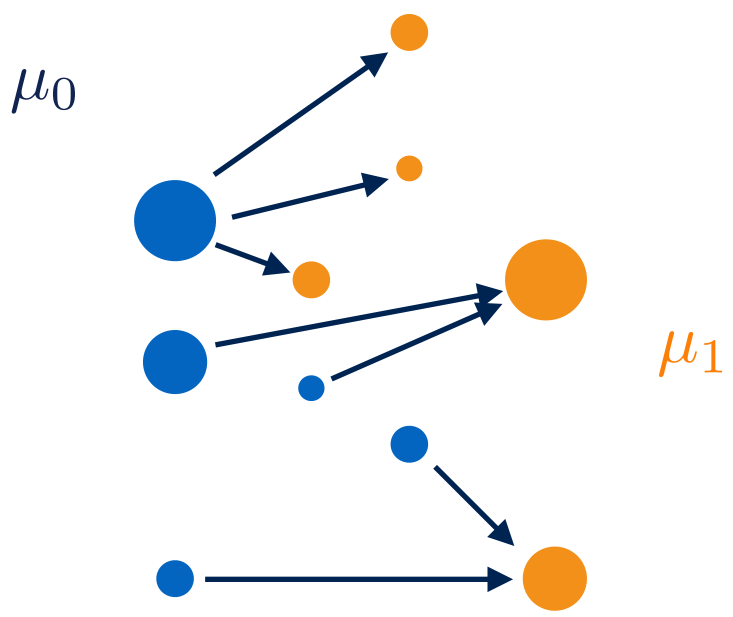


General formulation

[Kantorovich, *On the transfer of masses*, 1942]

$$W_c(\mu_0,\mu_1) = \inf_{\gamma\in\Pi(\mu_0,\mu_1)} \int_{X\times X} c(x,y)\,d\gamma(x,y).$$

Discrete case: $\mu_0 = \sum_i s_i\,\delta_{x_i}$, $\mu_1 = \sum_j t_j\,\delta_{y_j}$ **with** $\sum_i s_i = \sum_j t_j$



$$\boxed{W_c(\mu_0,\mu_1) = \min_{\gamma\in\Pi(\mu_0,\mu_1)} \sum_i \sum_j c(x_i,y_j)\gamma_{ij}}$$

$$\Pi(\mu_0,\mu_1) = \left\{\textbf{matrices } \gamma \textbf{ s.t. } \gamma_{i,j} \geq 0, \sum_i \gamma_{i,j} = t_j, \sum_j \gamma_{i,j} = s_i\right\}$$


Levallois-Perret
Pantin
Le Pré-Saint-Gervais
18
17
19
3
a
2
A
9
10
8
2
20
B
3
1
11
16
7
4
3
4
C
6
5
c
1
b
15
1
14
D
13

	A	B	C	D
a	3km	2.7km	6km	6km
b	7km	2.5km	7km	3.5km
c	5km	3km	4km	1.3km

Input

Cost matrix

	2	2	3	1
3	2	1	0	0
1	0	1	0	0
4	0	0	3	1

Output

Optimal Coupling

Monge-Kantorovich

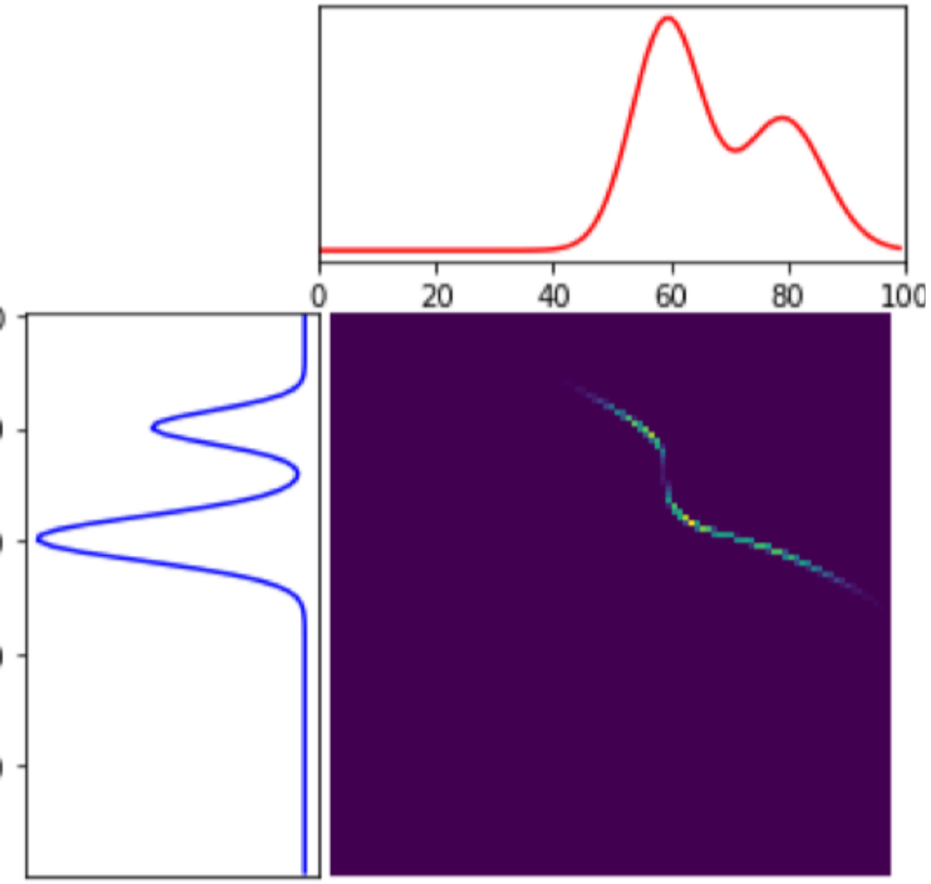
Monge, 1781

$$\inf_{T\#\mu_0=\mu_1} \int c(x,T(x))d\mu_0(x)$$

Kantorovich, 1939

$$\inf_{\gamma\in\Pi(\mu_0,\mu_1)} \int c(x,y)d\gamma(x,y)$$

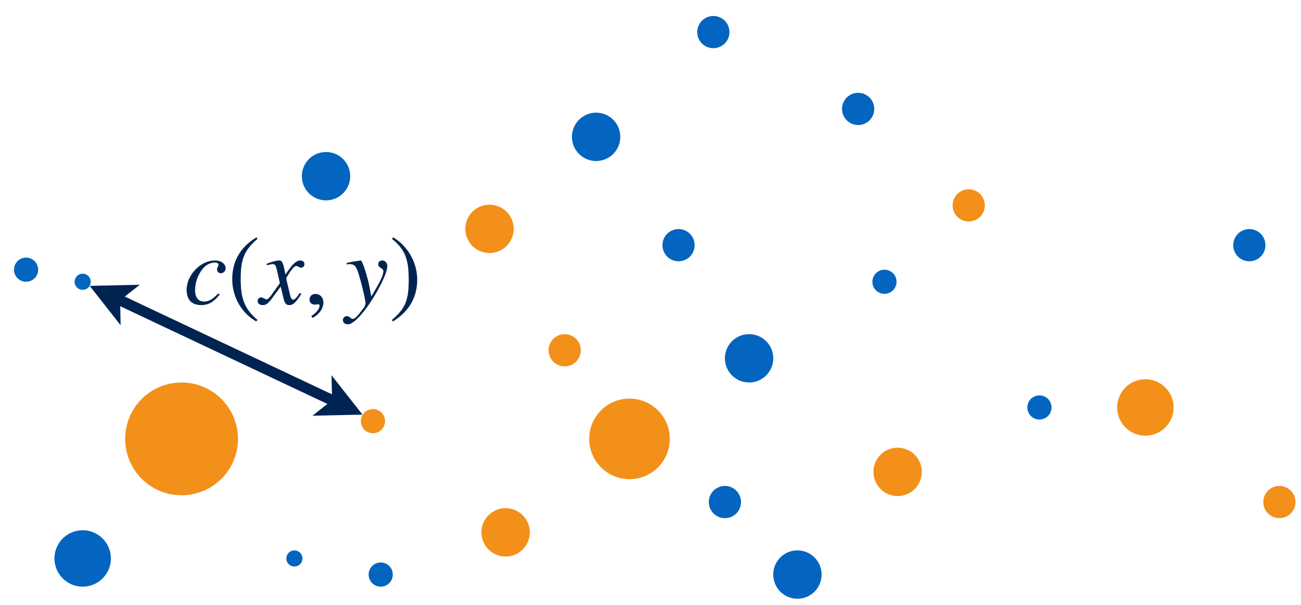
Brenier, 1991 If $c(x,y) = \|x-y\|^2$, if μ_0 has a density, Monge problem has a solution $T = \nabla\psi$ where ψ unique convex function s.t. $\nabla\psi\#\mu_0 = \mu_1$. The plan $\gamma = (Id,T)\#\mu_0$ is solution of Kantorovich pb.

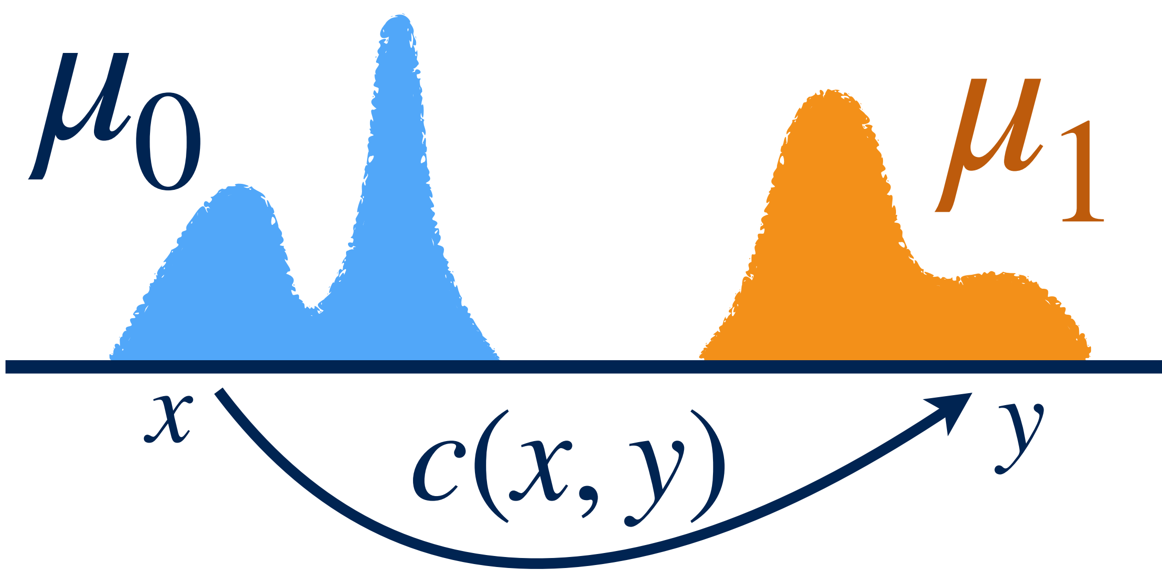


Displacement interpolation:

$$\mu_t = ((1-t)Id + tT)\#\mu_0, \;\; t\in[0,1]$$

Wasserstein distances





If $c(x,y) = d(x,y)^p$ with $p \geq 1$ and d a distance,

$$W_p(\mu_0,\mu_1) = \left(\inf_{\gamma \in \Pi(\mu_0,\mu_1)} \iint c(x,y) d\gamma(x,y) \right)^{\frac{1}{p}}$$

defines a **distance between probability measures**.

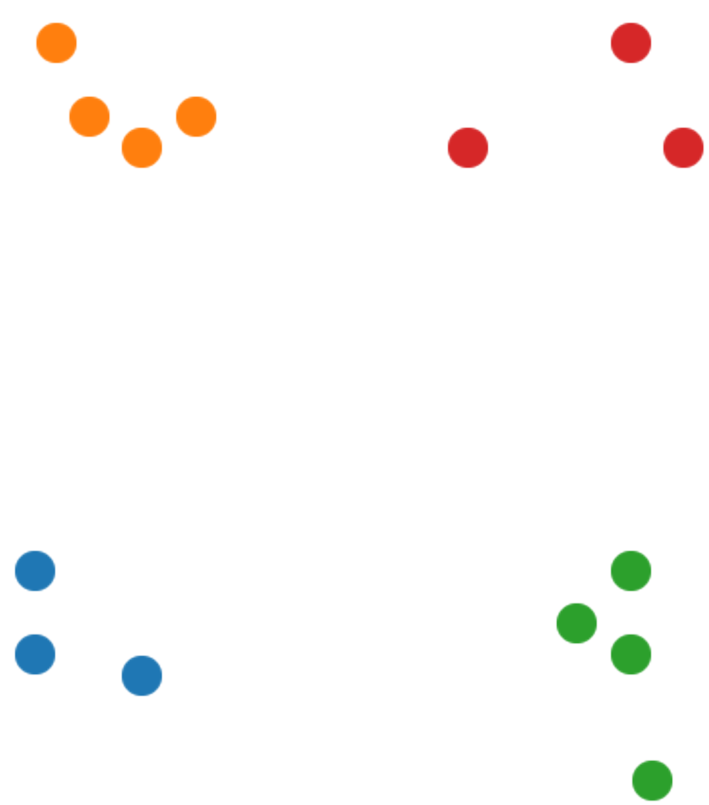
p=2 or 1 used in numerous applications

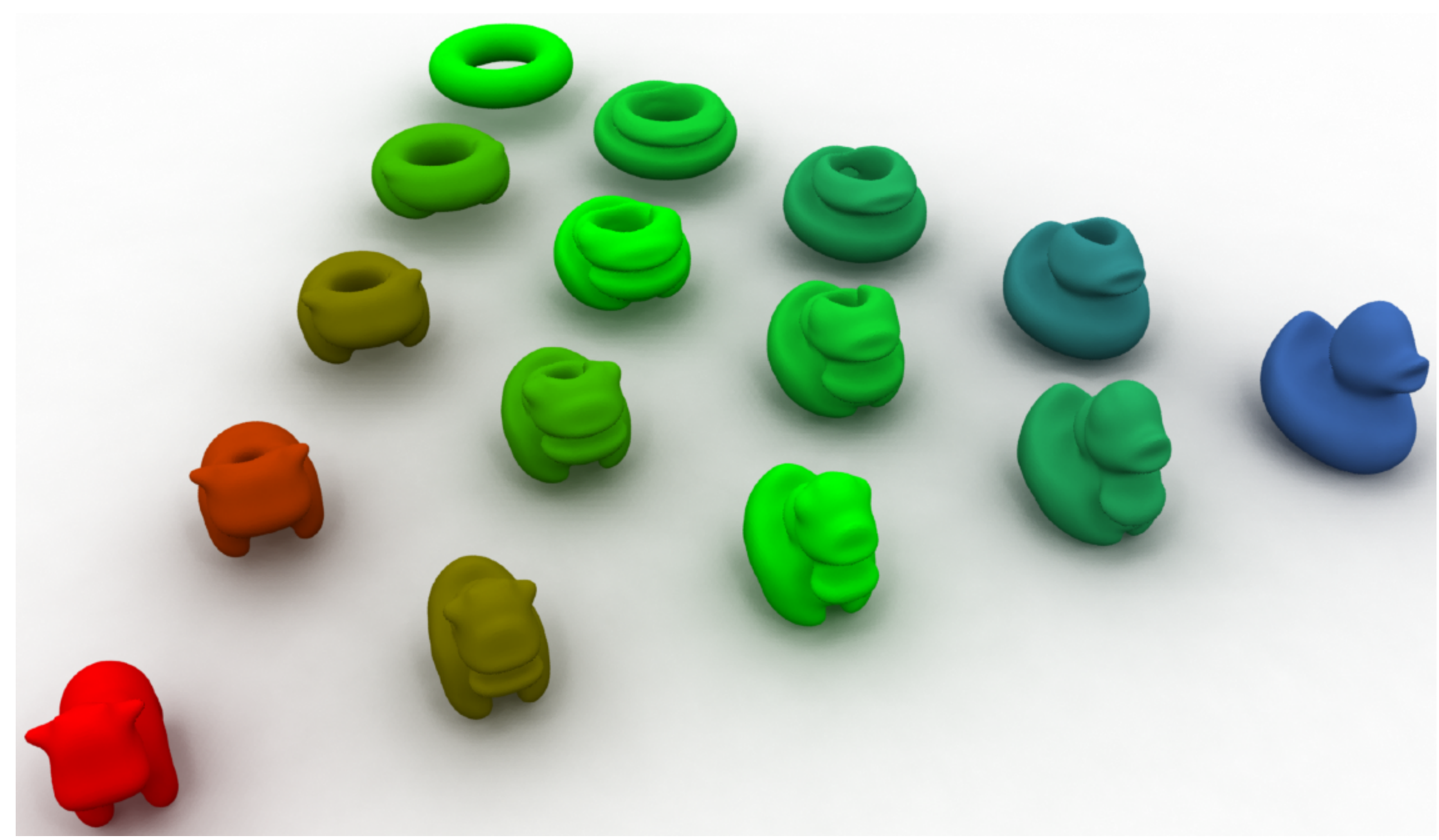
Wasserstein barycenters

Barycenter of $(\mu_i)_{i\in\{0,\ldots,I-1\}}$, weights $\sum_i \lambda_i = 1$

$$\mu_{bary} \in \underset{\rho}{\operatorname{argmin}} \sum_i \lambda_i W_2^2(\mu_i, \rho)$$

Prop. [Agueh, Carlier 2011]: existence and unicity of the barycenter for $c(x,y) = ||x - y||^2$ if the μ_i vanish on small sets.

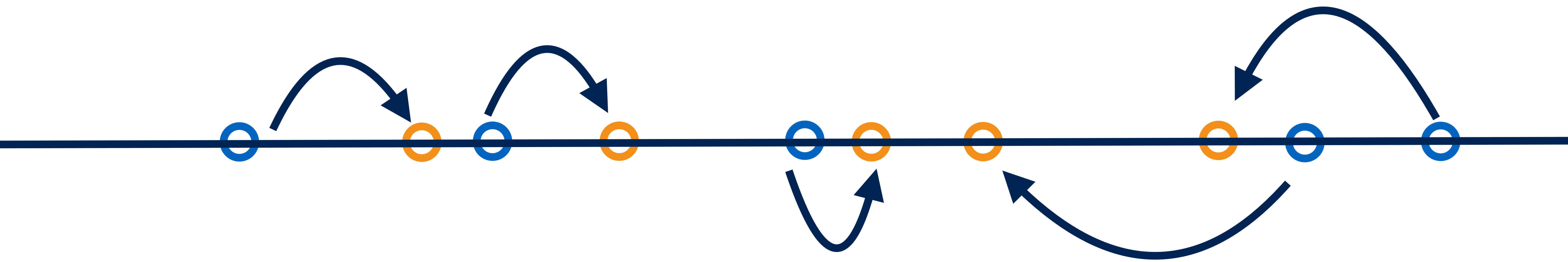




[Solomon et al. 2015]

Optimal transport in one dimension

Optimal Transport in 1D



If $c(x, y) = f(|x - y|)$ with f convex,

then for $x_1 < x_2$ and $y_1 < y_2$,

$$c(x_1, y_1) + c(x_2, y_2) < c(x_1, y_2) + c(x_2, y_1)$$

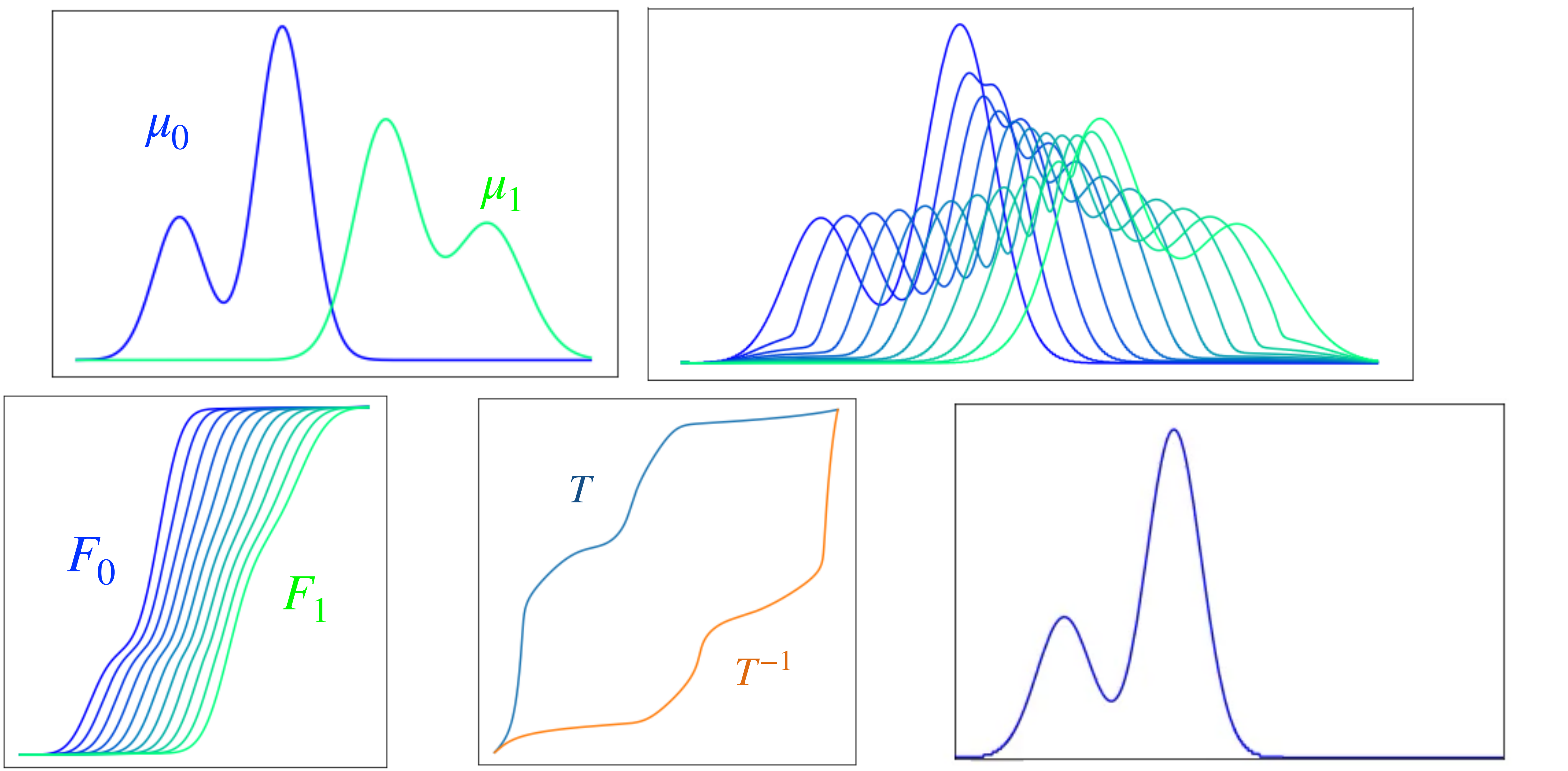
solution given by the **monotone rearrangement** of μ_0 onto μ_1

Optimal Transport in 1D

On $\mathbb{R}$, if $c(x,y) = f(|x-y|)$ with f convex,

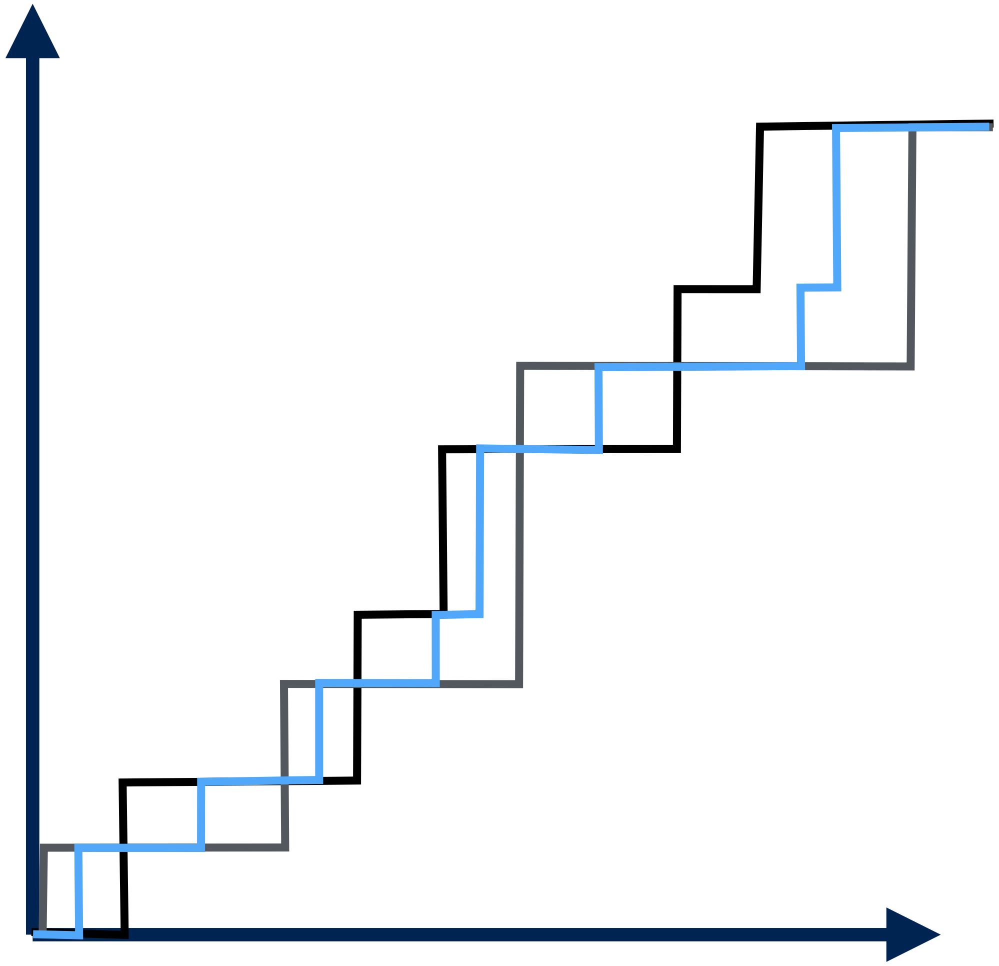
$$W_c(\mu_0, \mu_1) = \int_0^1 f(|F_0^{-1}(t) - F_1^{-1}(t)|)dt,$$

with F_0 and F_1 the distribution functions of μ_0 and μ_1. Moreover, if μ_0 has no atoms, $T = F_1^{-1} \circ F_0$ is solution of the Monge problem.



Midway histogram

$$\left(\frac{F_0^{-1} + F_1^{-1}}{2}\right)^{-1}$$



Numerical approaches

Linear programming

Input $\mu_0 = \sum_{i=1}^{K_0} s_i \, \delta_{x_i}$, $\mu_1 = \sum_{j=1}^{K_1} t_j \, \delta_{y_j}$ **with** $\sum_i s_i = \sum_j t_j = 1$

(LP) $$\operatorname*{argmin}_{\gamma \in \Pi(\mu_0,\mu_1)} \sum_{i,j} c_{i,j} \gamma_{i,j} \text{ with}$$

$$\Pi(\mu_0, \mu_1) = \left\{ \textbf{matrices } \gamma \textbf{ s.t. } \gamma_{i,j} \geq 0, \sum_i \gamma_{i,j} = t_j, \ \sum_j \gamma_{i,j} = s_i \right\}$$

One solution has less than $K_0 + K_1 - 1$ values $\neq 0$

Assignment: Hungarian algo. [Kuhn 1995] $O(N^3)$, Auction [Bertsekas 1992]

LP: Network Simplex [Cunningham 1976]$O(N^3)$

Dynamic formulation [Brenier, Benamou 2000]

Semi-discrete OT [Mérigot 11, Levy 15]

Sliced OT [Rabin et al. 11, Rabin et al.15]

Entropic OT [Cuturi 13,...]

Sliced optimal transport

Replace classical OT by

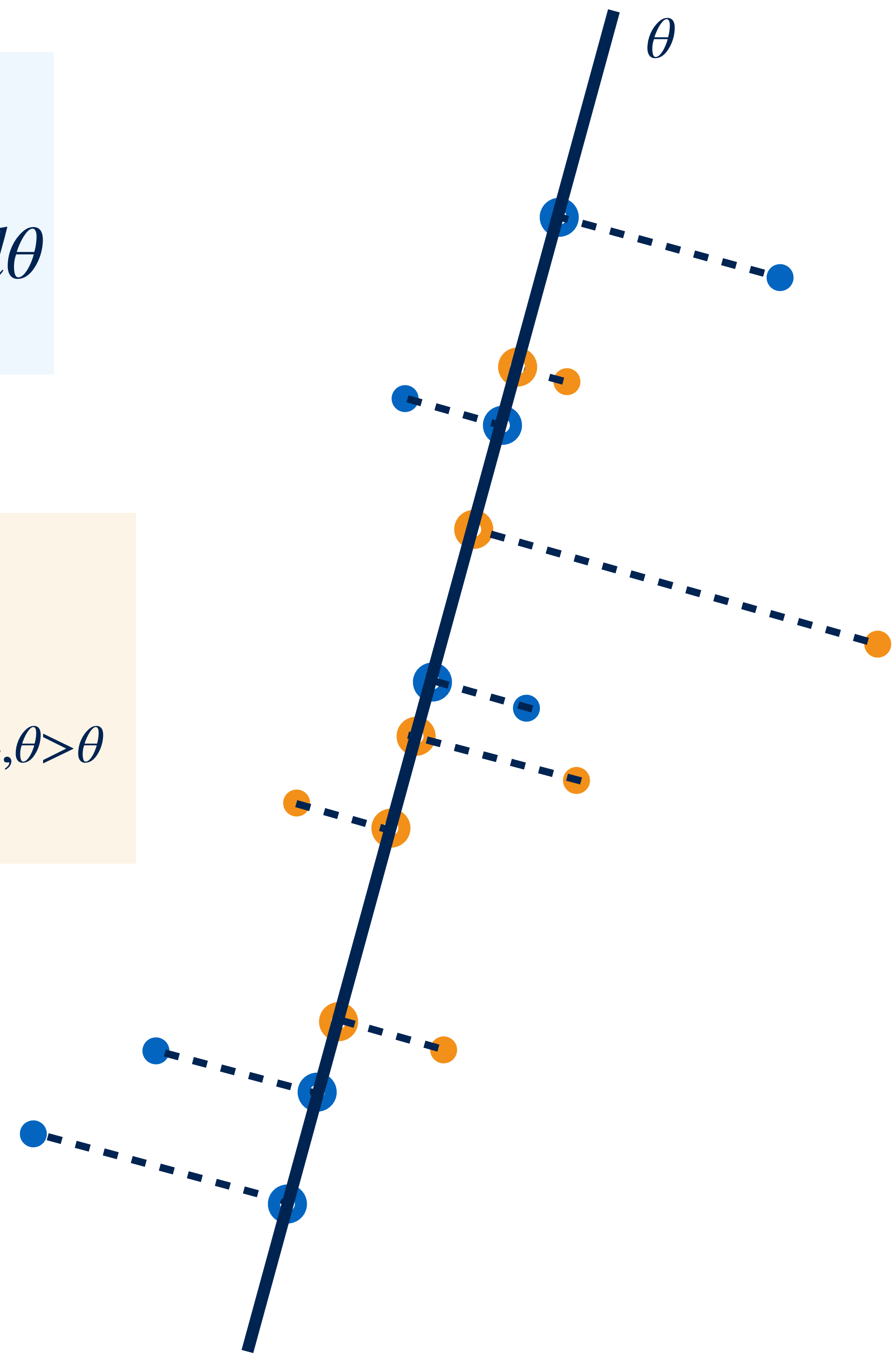
$$SW_2^2(\mu_0, \mu_1) = \int_{\mathbb{S}^{d-1}} W_2^2(p_\theta \# \mu_0, p_\theta \# \mu_1) d\theta$$

Discrete measures

$$\mu_0 = \frac{1}{n} \sum_{i=1}^{n} \delta_{x_i} \ , \ \mu_1 = \frac{1}{n} \sum_{j=1}^{n} \delta_{y_j}, \ p_\theta \# \mu_0 = \frac{1}{n} \sum_{i} \delta_{<x_i, \theta> \theta}$$

$$SW_2^2(\mu_0, \mu_1) = \int_{\mathbb{S}^{d-1}} \sum_i |\langle x_i - y_{\sigma_\theta(i)}, \theta \rangle|^2 d\theta$$

with σ_θ monotone rearrangement between $< \mu_0, \theta >$ and $< \mu_1, \theta >$.

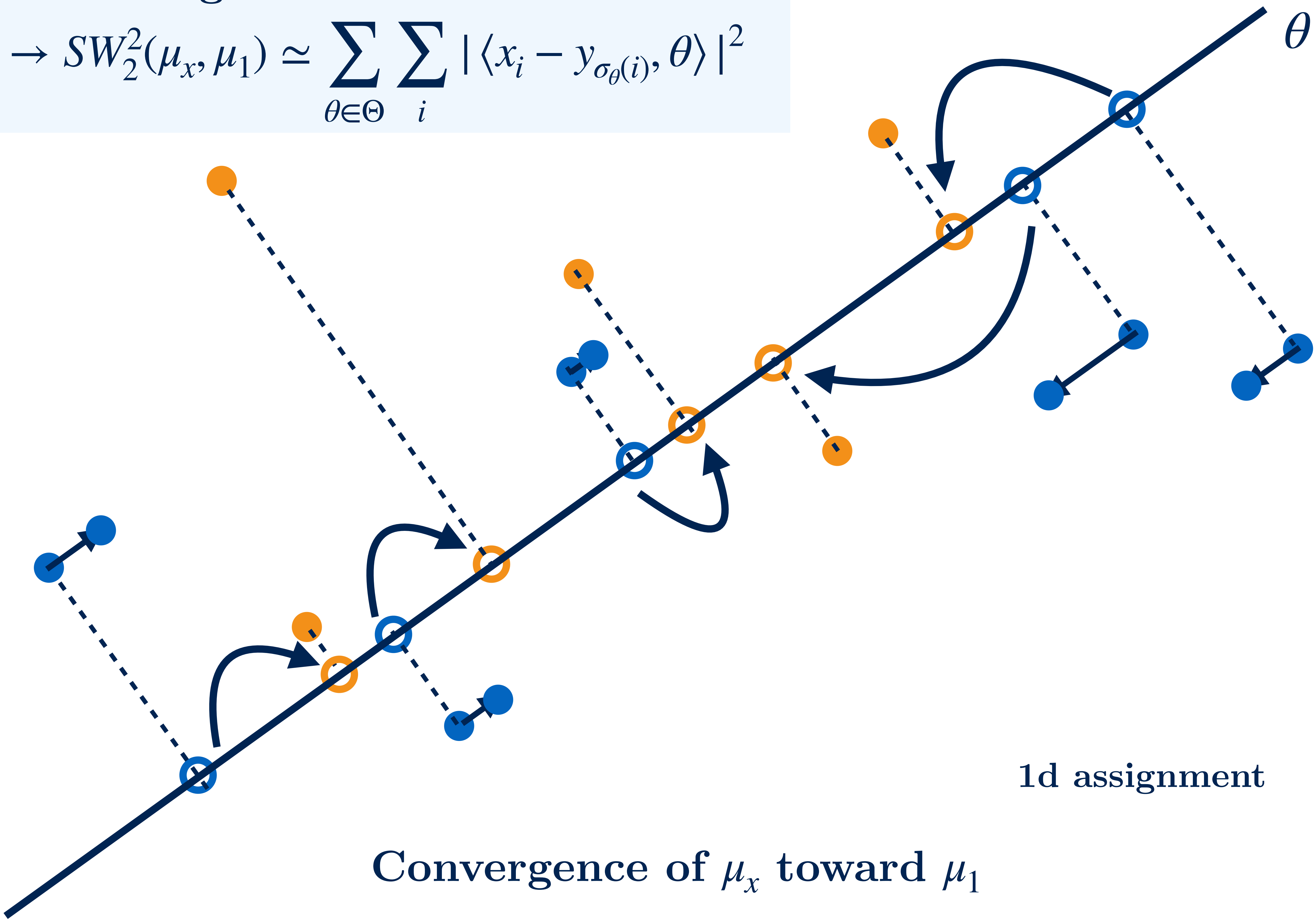


[Rabin et al. 11, Rabin et al. 2015]

Assignment with Sliced OT

Stochastic gradient descent on

$$x \to SW_2^2(\mu_x, \mu_1) \simeq \sum_{\theta \in \Theta} \sum_i |\langle x_i - y_{\sigma_\theta(i)}, \theta \rangle|^2$$



Convergence of μ_x toward μ_1

Entropic OT

Entropy of the matrix γ $\quad H(\gamma) = \sum_{i,j} \gamma_{i,j}(\log(\gamma_{i,j}) - 1)$

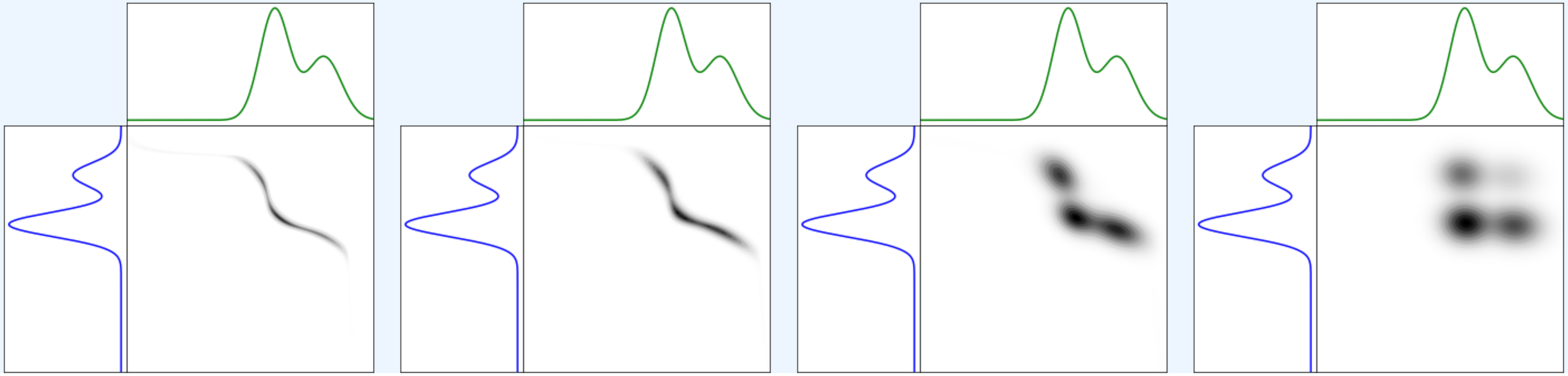
Entropic OT [Cuturi '13]

$$\underset{\gamma \in \Pi(\mu_0,\mu_1)}{\text{argmin}} \sum_{i,j} c(x_i, y_j)\gamma_{i,j} - \varepsilon H(\gamma)$$

With $K_{i,j} = e^{-\frac{1}{\varepsilon} c(x_i, y_j)}$ the pb becomes

$$\underset{\gamma \in \Pi(\mu_0,\mu_1)}{\text{argmin}} \sum_{i,j} \gamma_{i,j} \log\left(\frac{\gamma_{i,j}}{K_{i,j}}\right) = \underset{\gamma \in \Pi(\mu_0,\mu_1)}{\text{argmin}} \ \mathrm{KL}(\gamma || K)$$

Sinkhorn algorithm = alternate projections of K on $\Pi(\mu_0, \mu_1)$



$\varepsilon = 3 \times 10^{-4}$ $\quad$ $\varepsilon = 10^{-3}$ $\quad$ $\varepsilon = 10^{-2}$ $\quad$ $\varepsilon = 10^{-1}$

Sinkhorn algorithm

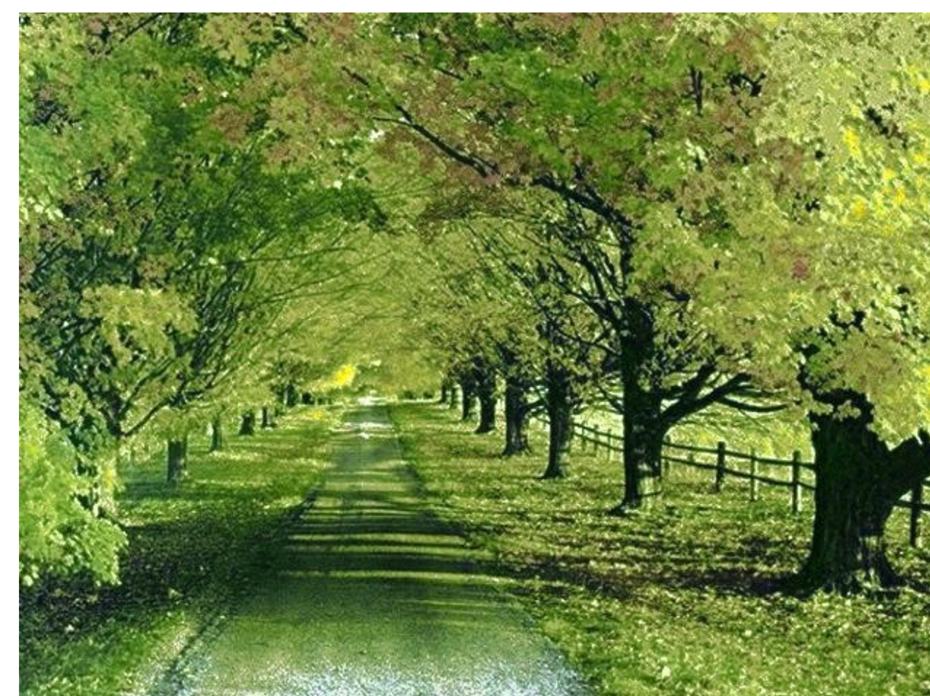
Prop: solution γ of $\underset{\gamma\in\Pi(\mu_0,\mu_1)}{\operatorname{argmin}} \mathrm{KL}(\gamma||K)$ satisfies $\gamma = \mathrm{diag}(a) K \,\mathrm{diag}(b)$

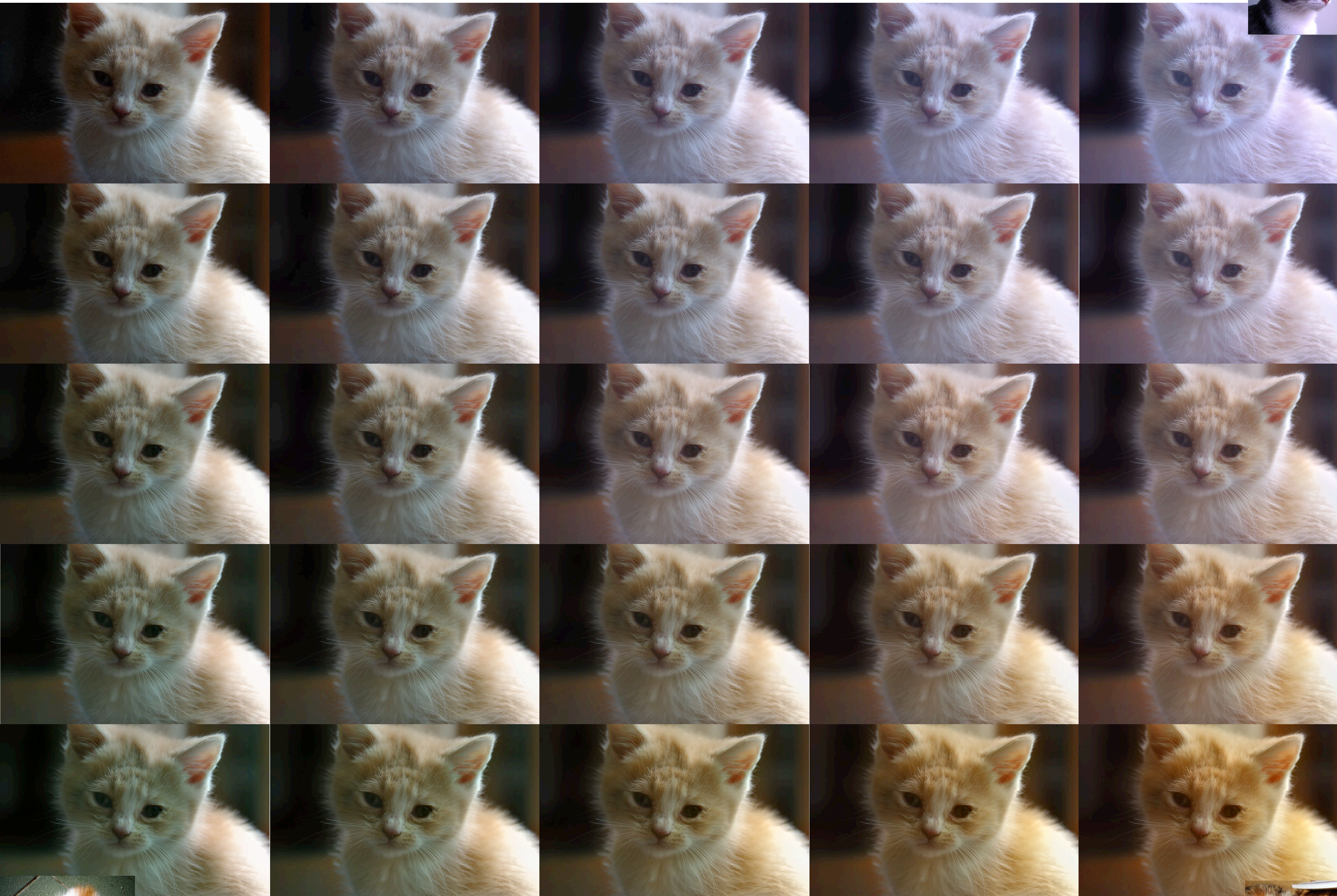
Since $\gamma \in \Pi(\mu_0, \mu_1)$, it implies that
$$a \odot Kb = \mu_0$$
$$b \odot K^T a = \mu_1$$

Iterations: $a \leftarrow \dfrac{\mu_0}{Kb} \quad b \leftarrow \dfrac{\mu_1}{K^T a}$

- Iterative projections on the constraints.
- Simple extension to compute barycenters of more than 2 measures
- Matrix-vector multiplications
- For regular grids, products Kx can be written as convolutions.
- Numerical pb when $\varepsilon \to 0$.

Application to color transfer



Reducing artefacts introduced by color transfer

- Noise increases
- Detail loss
- Compression becomes visible

u

$g(u)$

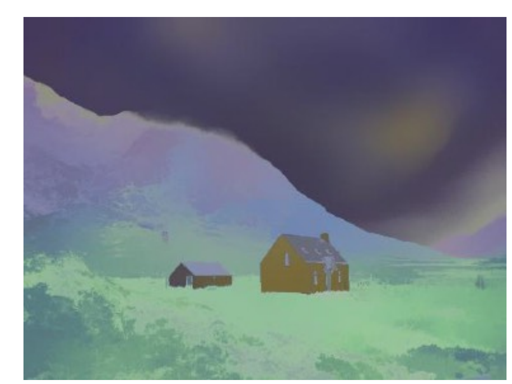
Post-processing

 $=$  $+$ 

Principle: filter the difference map $g(u) - u$ thanks to an linear operator Y_u and reconstruct

$$T(g(u)) = u + Y_u(g(u) - u) = \underbrace{Y_u(g(u))}_{\text{regularization of } g(u)} + \underbrace{u - Y_u(u)}_{\text{details of } u}.$$

- Y_u = simple average or more involved filter (guided filter for instance)
- Guided filter regularizes conditionally to a source image



[Pitie et al., 2007], [Papadakis et al., 2011], [Rabin et al, 2011]

Color transfer