

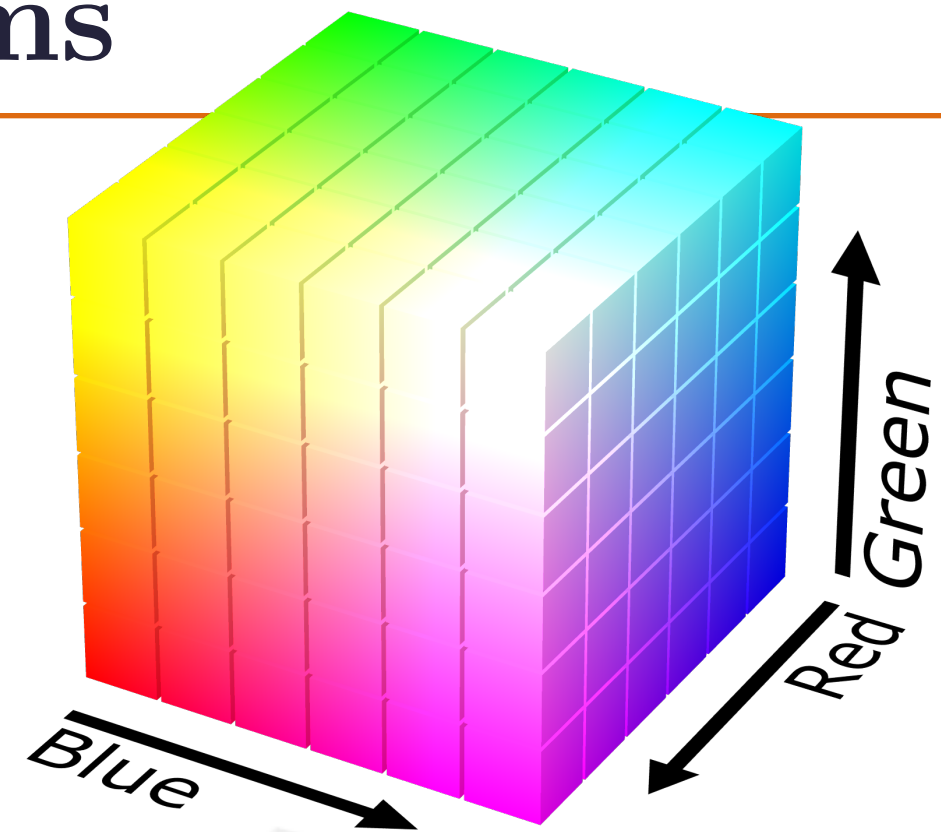
Optimal Transport for image processing

Julie Delon

introduction à l'imagerie numérique

Color distributions and optimal transport

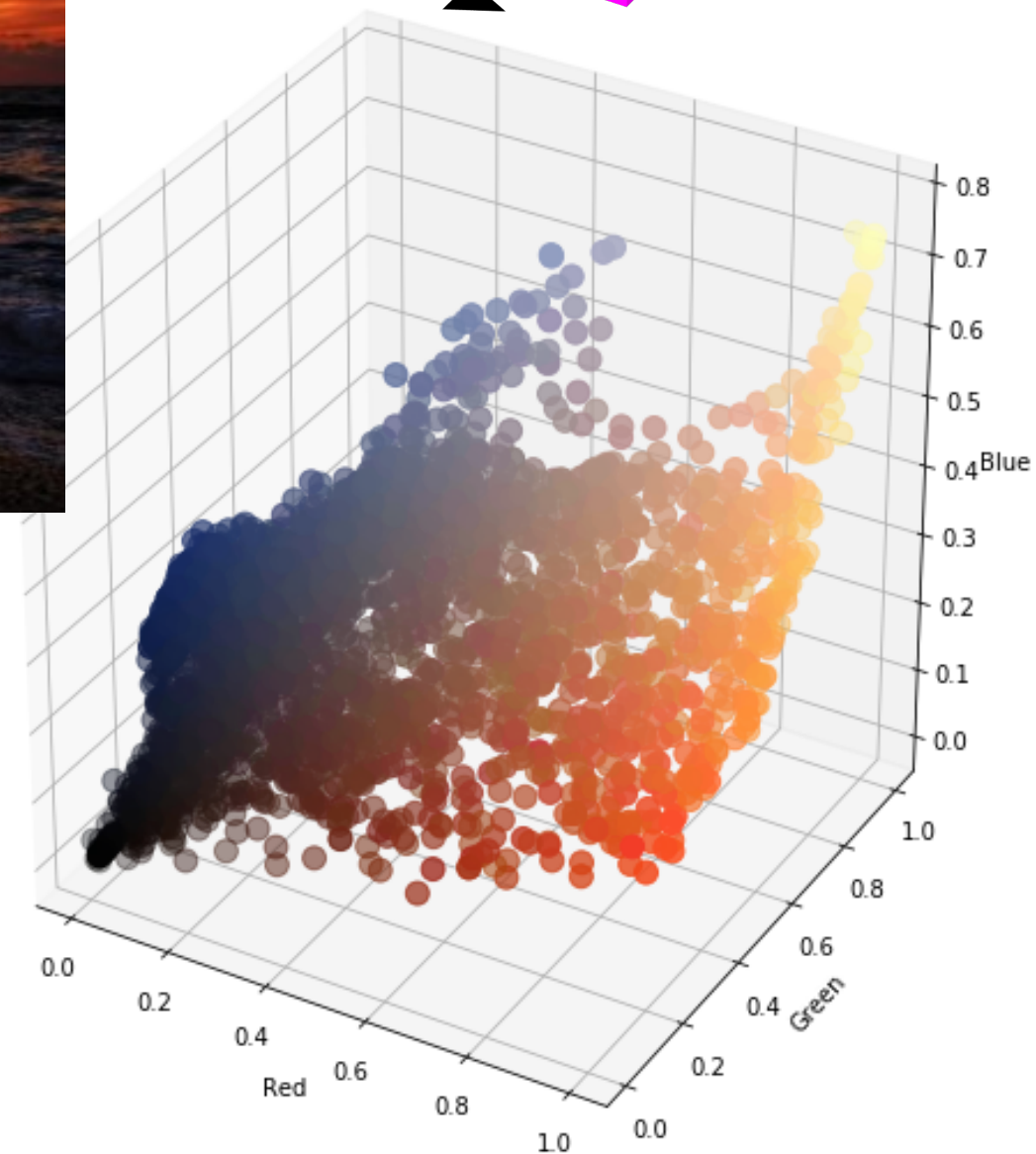
Color histograms



$$h_u = \sum_{i \in I} s_i \delta_{\lambda_i} \text{ with } \lambda_i \text{ on a fixed grid, } \sum_i s_i = 1$$

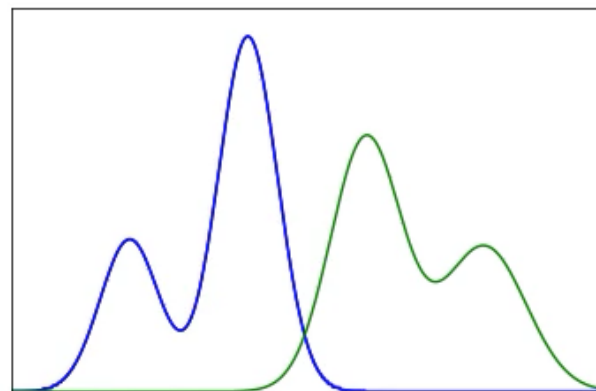
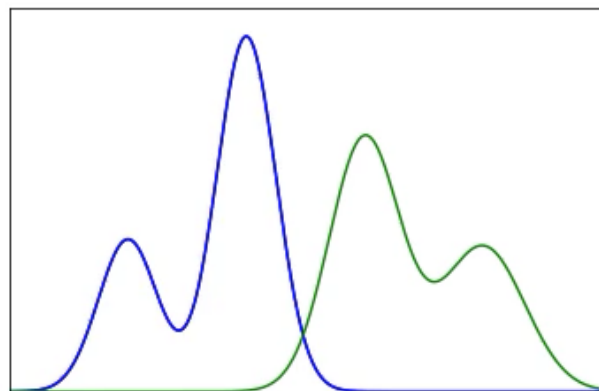
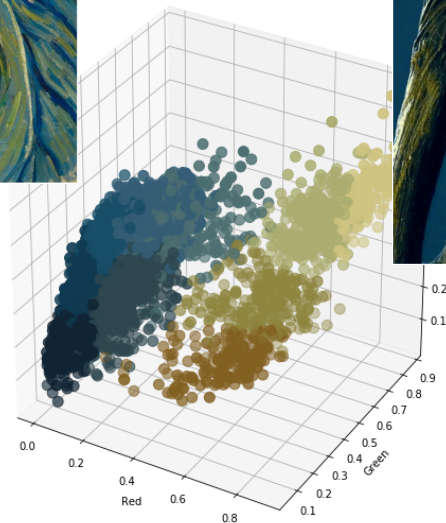
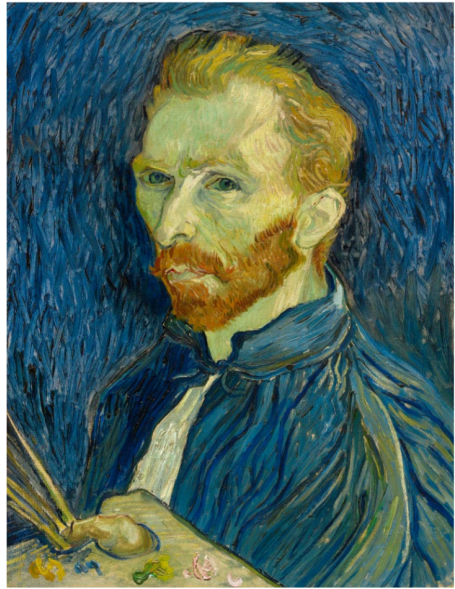
and $s_i \geq 0$. (Eulerian formulation)

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

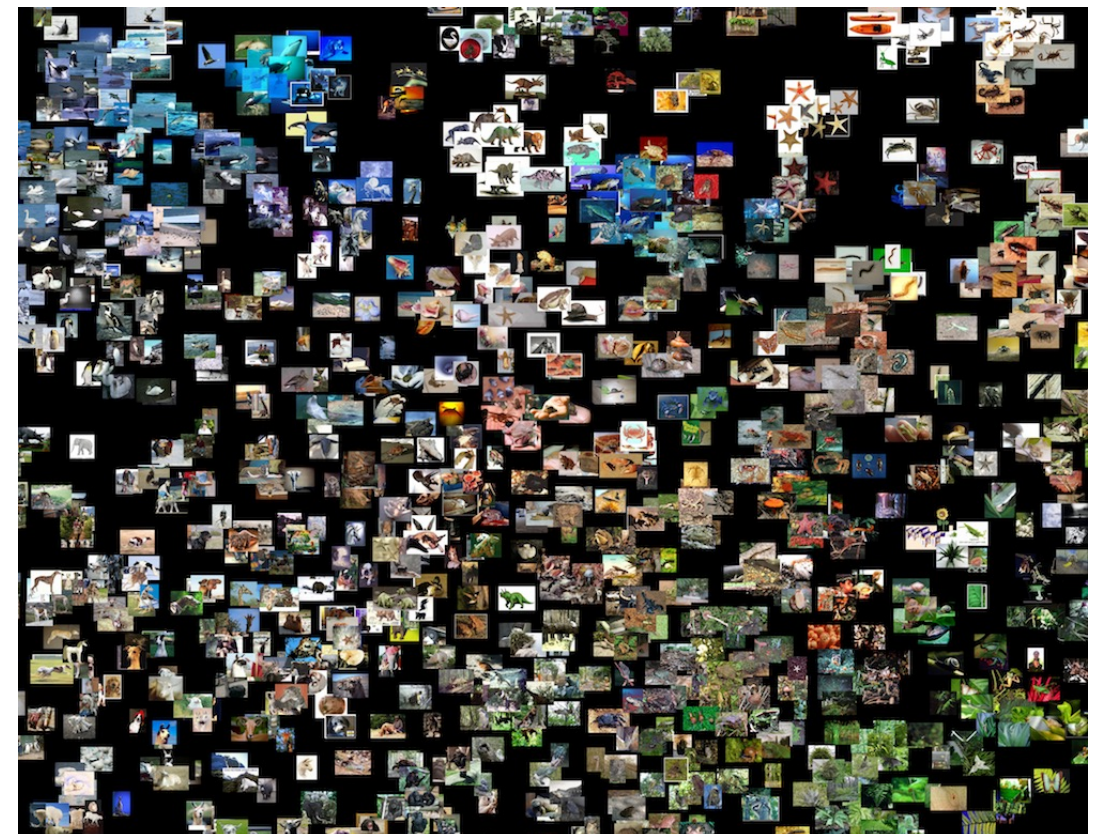


Optimal transport for color distributions

Color transfer



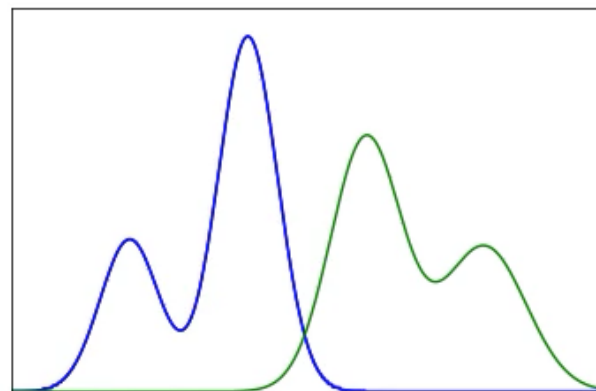
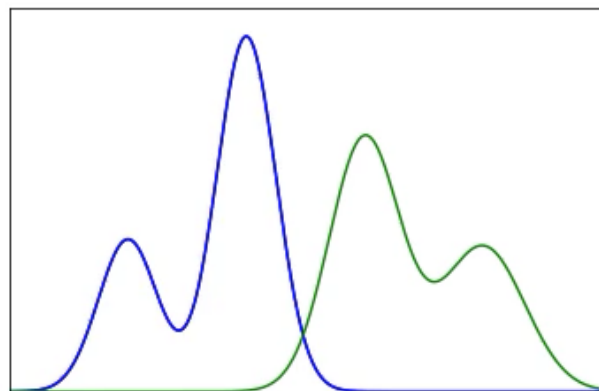
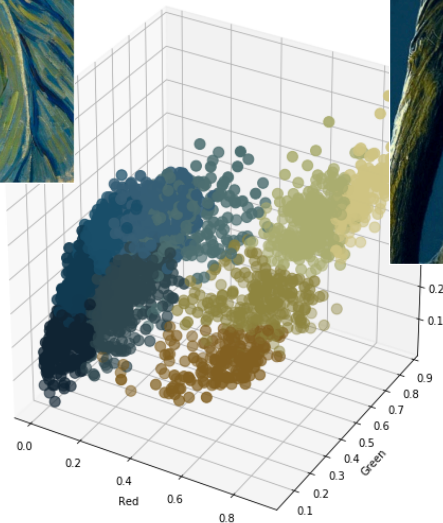
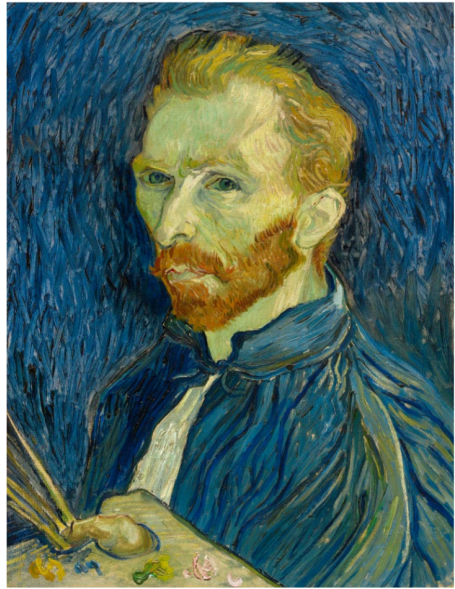
Color barycenter



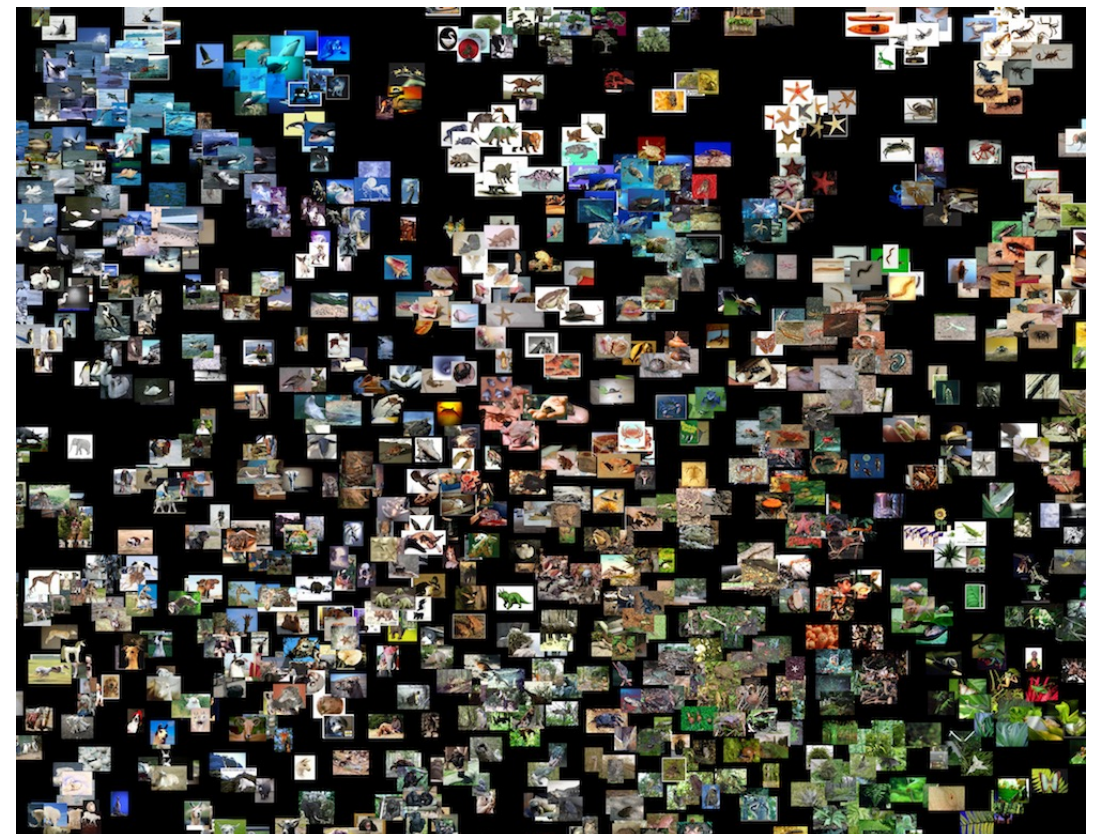
Color based retrieval

Optimal transport for color distributions

Color transfer



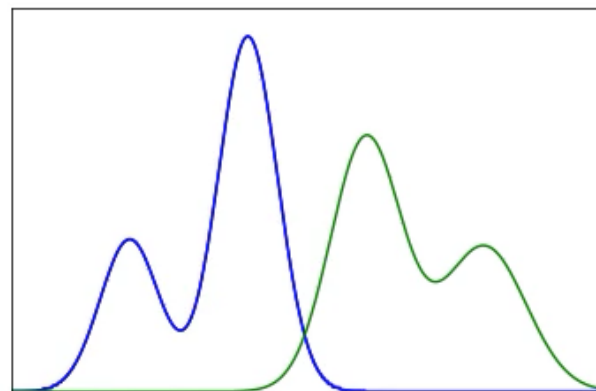
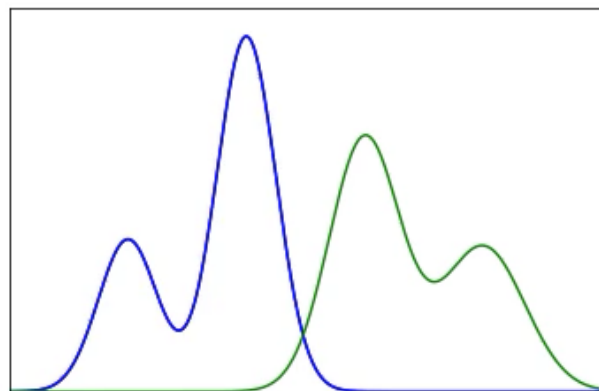
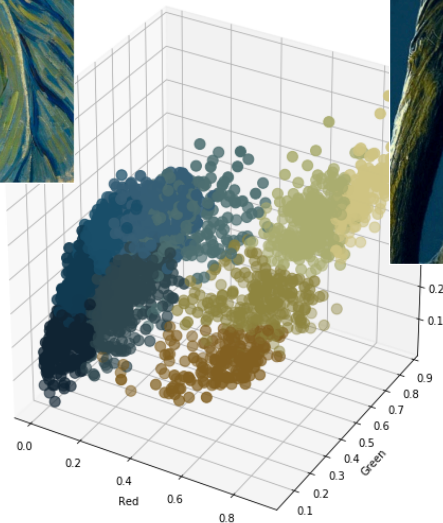
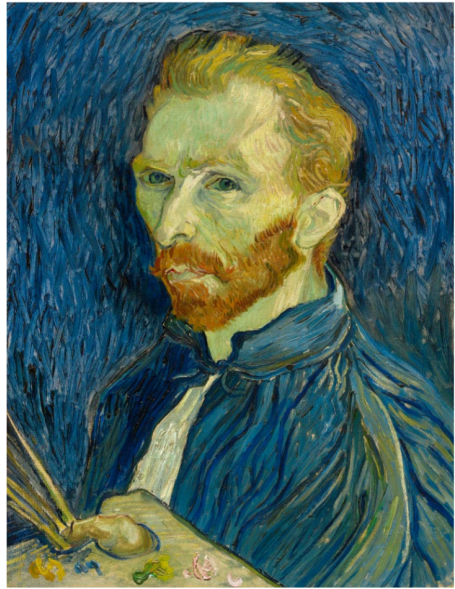
Color barycenter



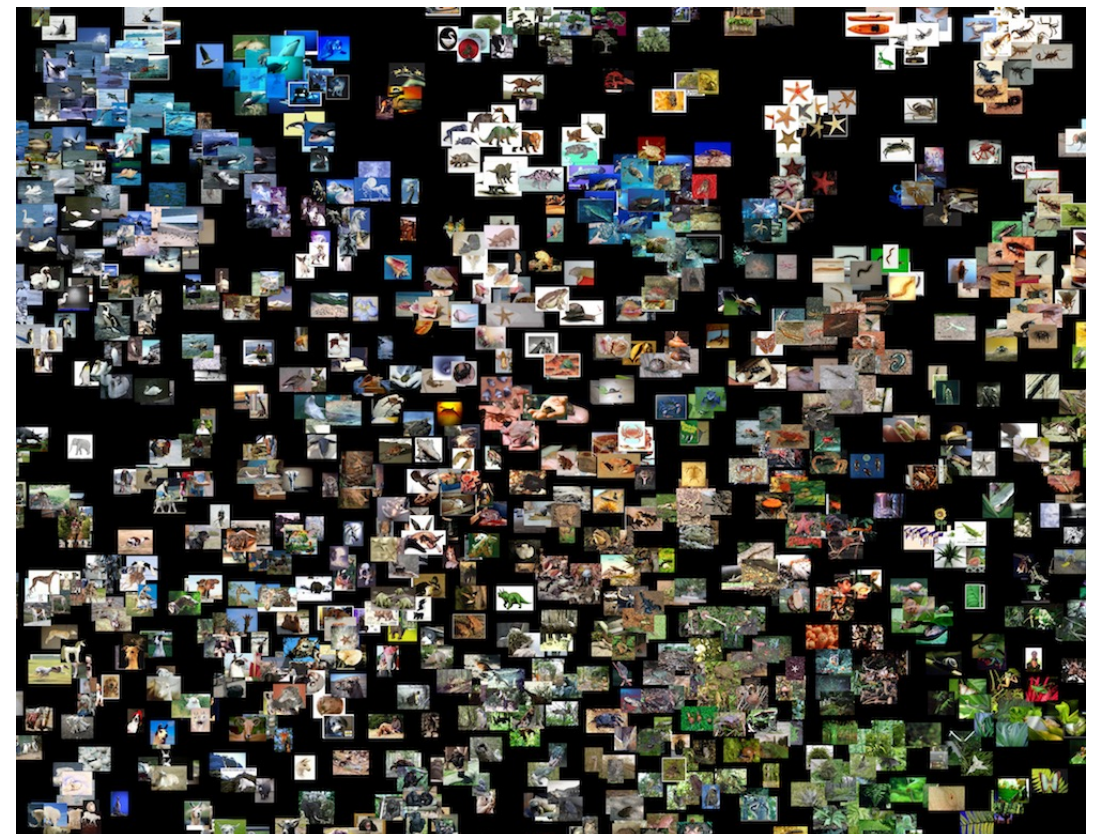
Color based retrieval

Optimal transport for color distributions

Color transfer



Color barycenter



Color based retrieval

Color grading



Images from the documentary *Painting With Pixels: (O' Brother, Where Art Thou)*.

Affine color transfer



Reinhard et al. Color Transfer between Images, 2001.

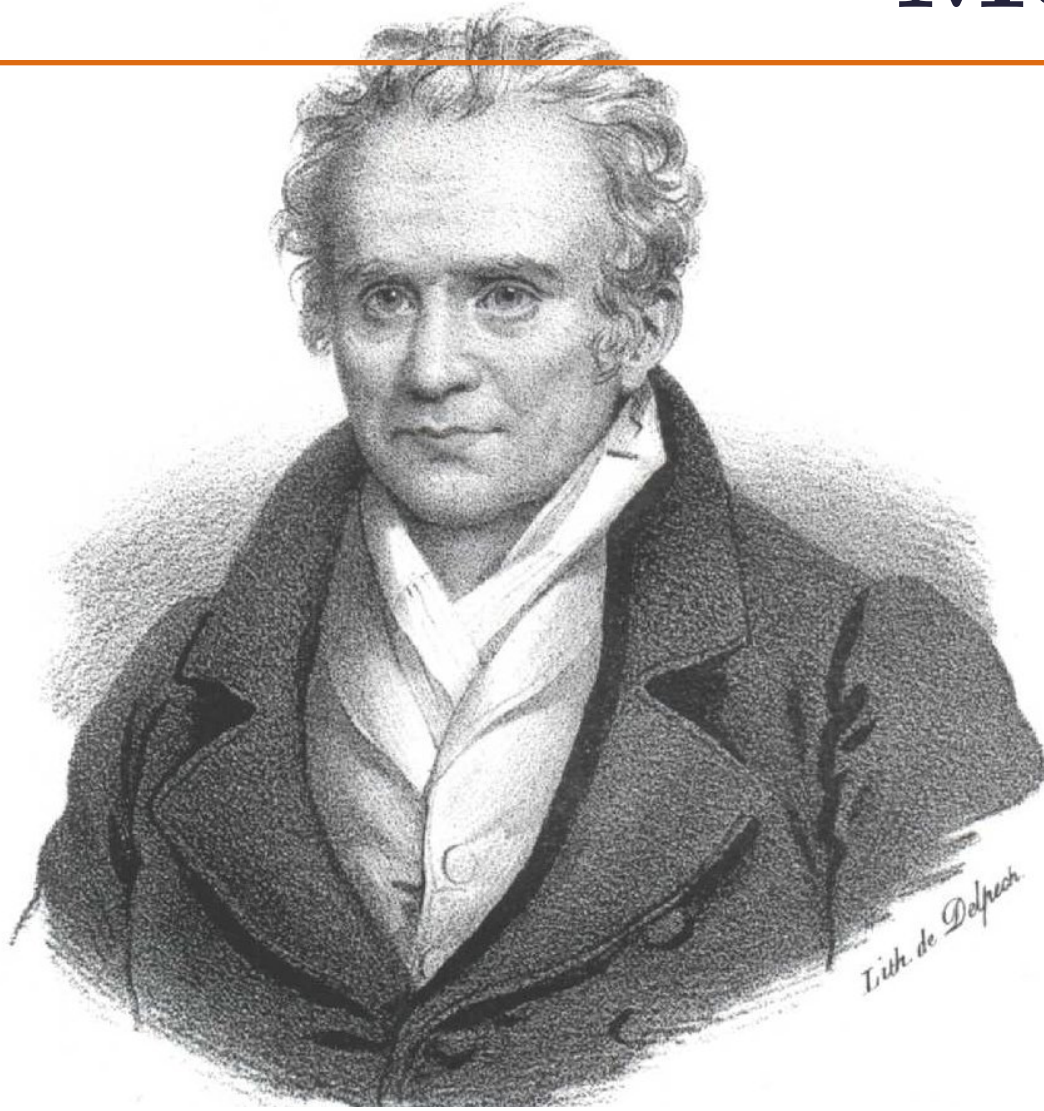
Affine color transfer



Reinhard et al. Color Transfer between Images, 2001.

A short introduction to Optimal transport

Monge, 1781

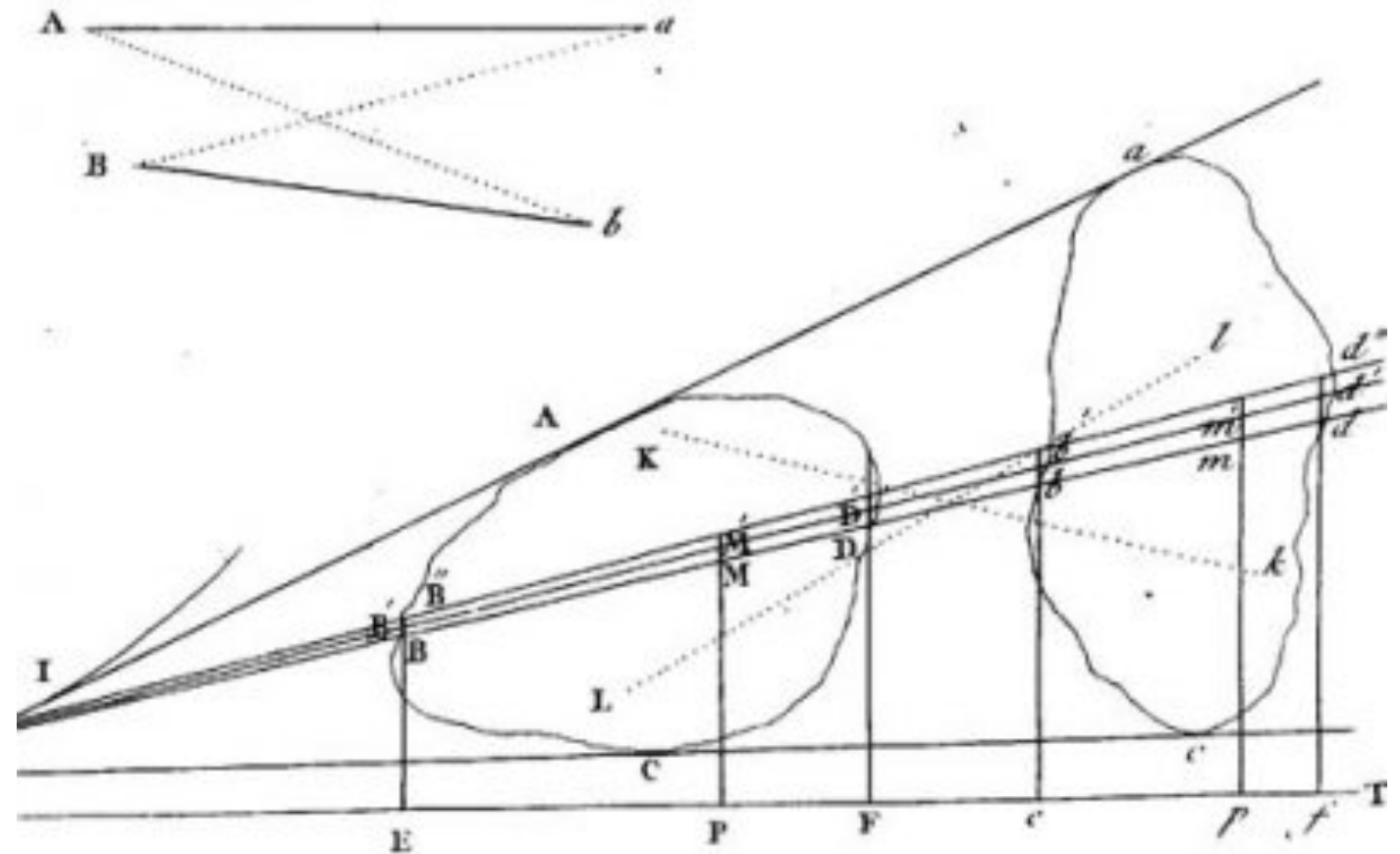


666. MÉMOIRES DE L'ACADÉMIE ROYALE

M É M O I R E
S U R L A
T H É O R I E D E S D É B L A I S
E T D E S R E M B L A I S.
Par M. M O N G E.

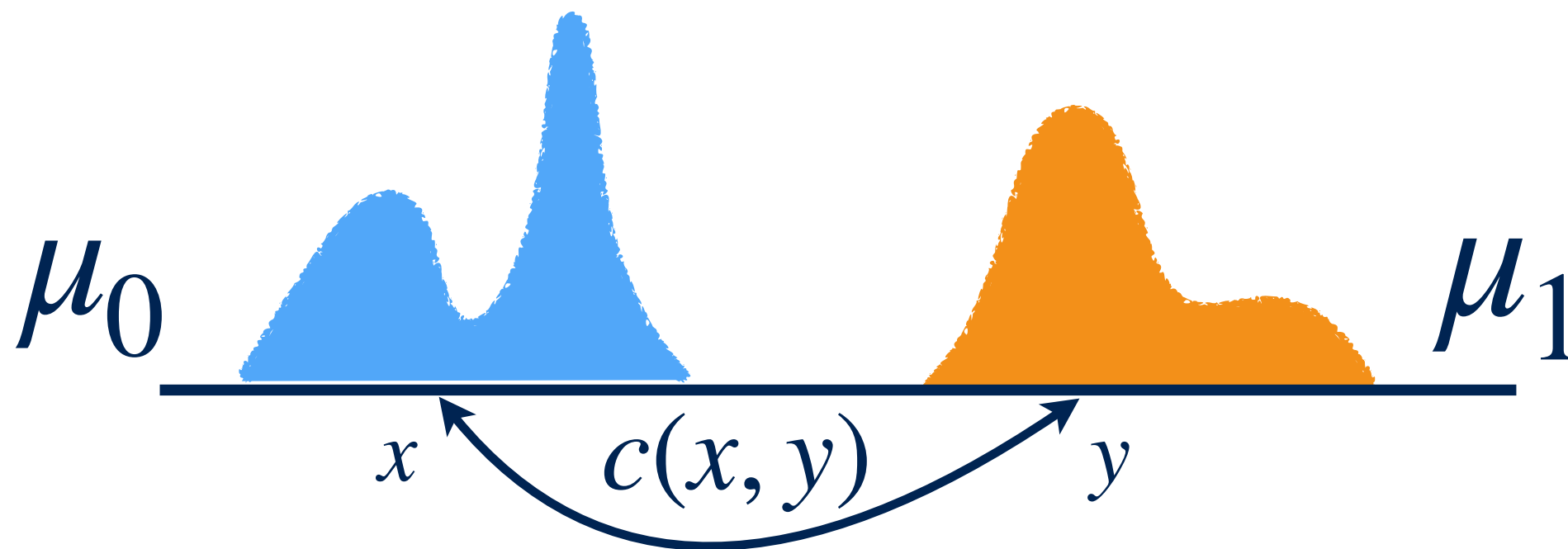
Mém. de l'Ac. R. des Sc. An. 1781. Page. 704. Pl. XVII.

Fig. 1.



Monge optimal transport

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

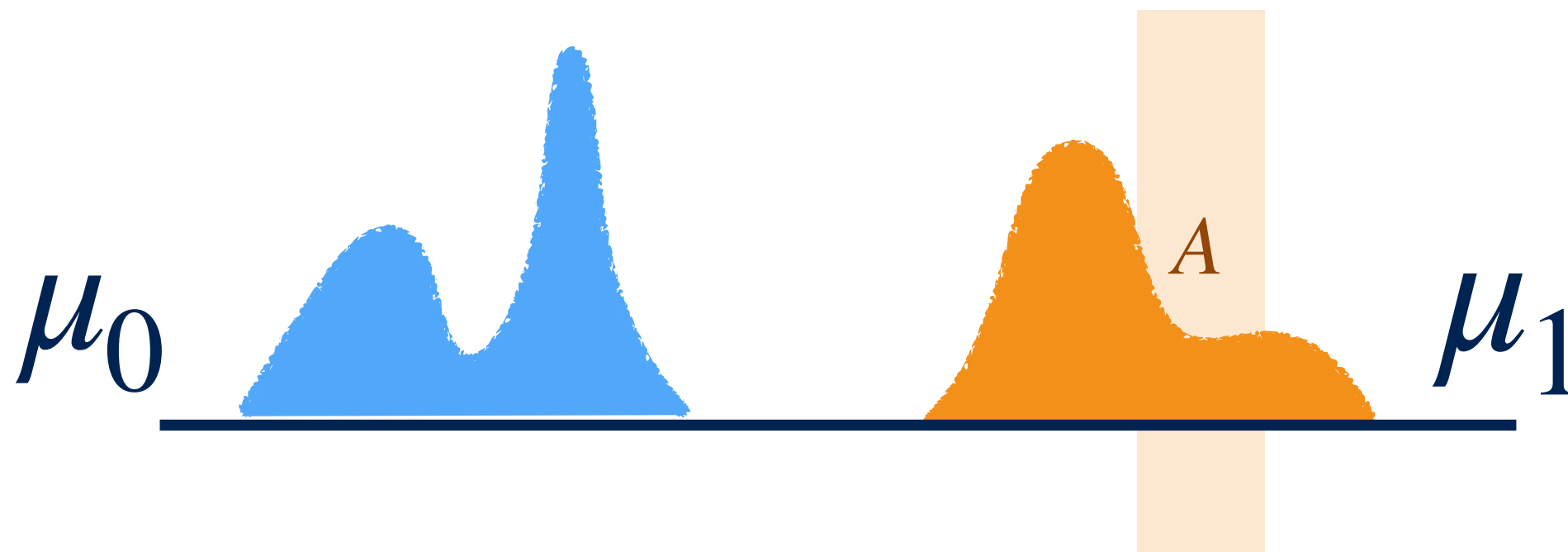


Total cost of the mass transfer = sum of costs of displacements of elementary masses.

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

Monge optimal transport

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



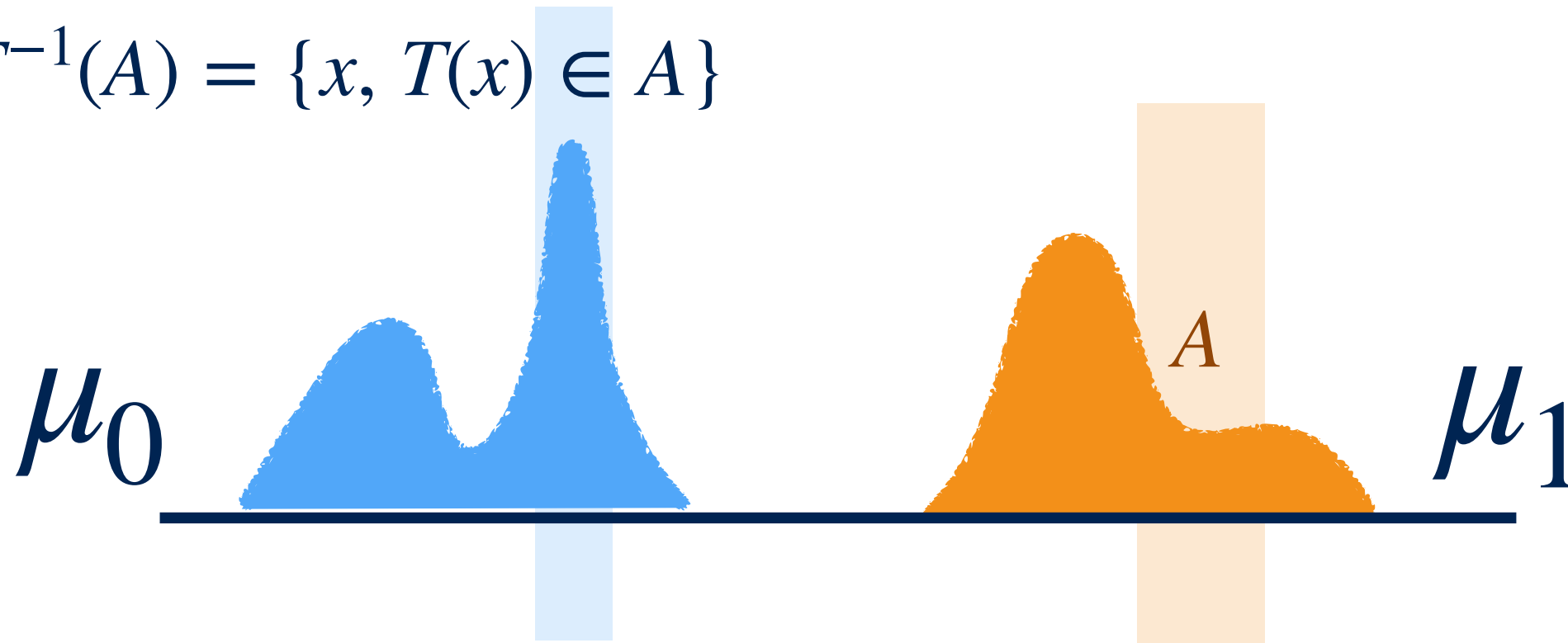
Total cost of the mass transfer = sum of costs of displacements of elementary masses.

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

Monge optimal transport

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

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



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

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

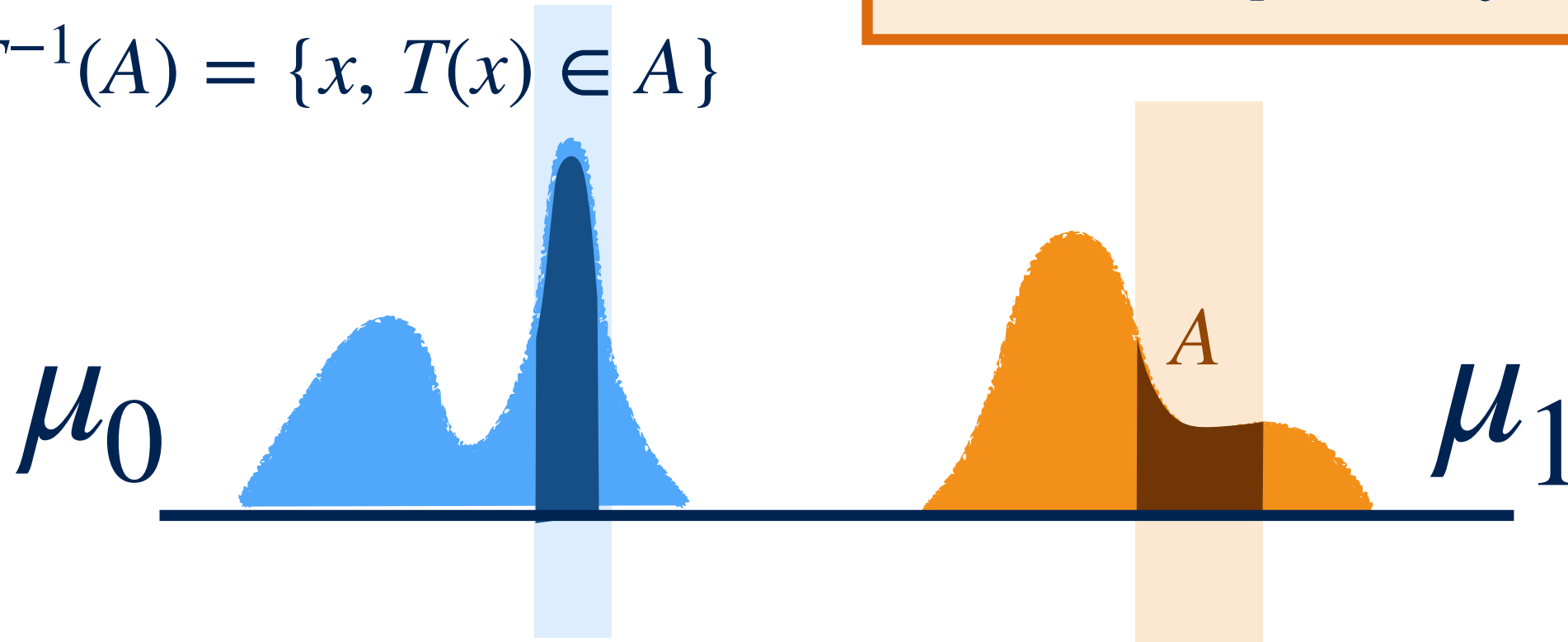
Monge optimal transport

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

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

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

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



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

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

Optimal transport

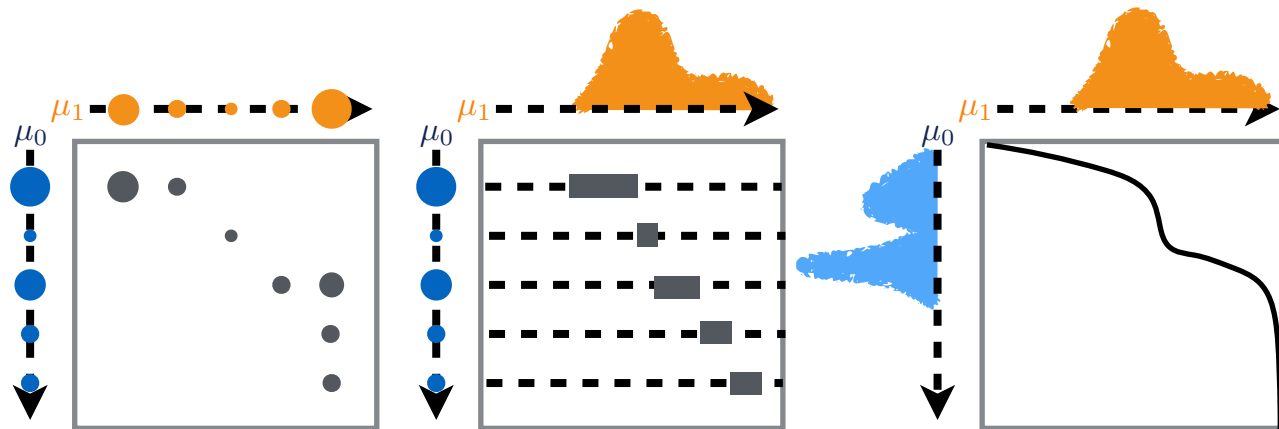
[Kantorovich, *On the transfer of masses*, 1942]

$$W_c(\mu_0, \mu_1) = \inf_{\gamma \in \Pi(\mu_0, \mu_1)} \int_{X \times X} c(x, y) d\gamma(x, y) = \inf_{\substack{(X, Y) \\ X \sim \mu_0, Y \sim \mu_1}} \mathbb{E}[c(X, Y)]$$

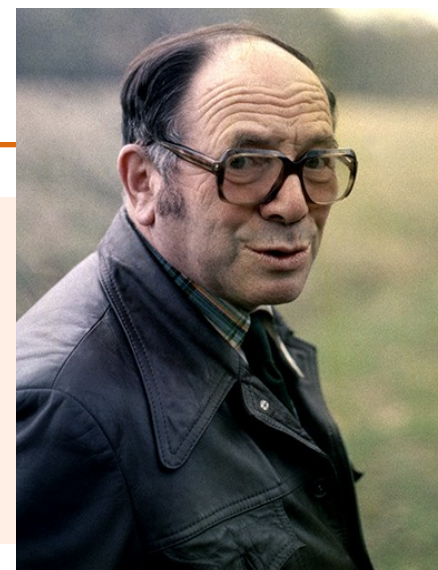


Couplings

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



Optimal transport

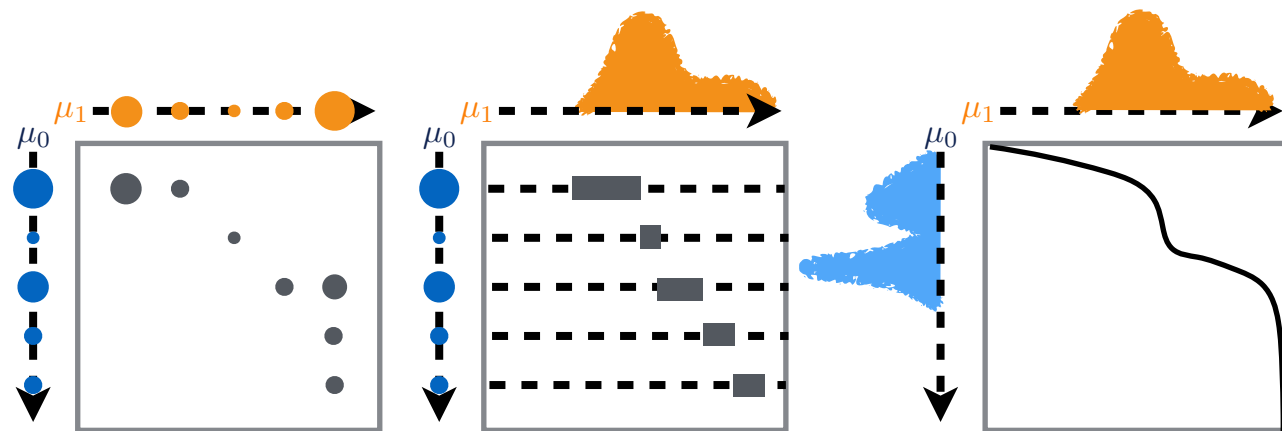


[Kantorovich, *On the transfer of masses*, 1942]

$$W_c(\mu_0, \mu_1) = \inf_{\gamma \in \Pi(\mu_0, \mu_1)} \int_{X \times X} c(x, y) d\gamma(x, y) = \inf_{\substack{(X, Y) \\ X \sim \mu_0, Y \sim \mu_1}} \mathbb{E}[c(X, Y)]$$

Couplings

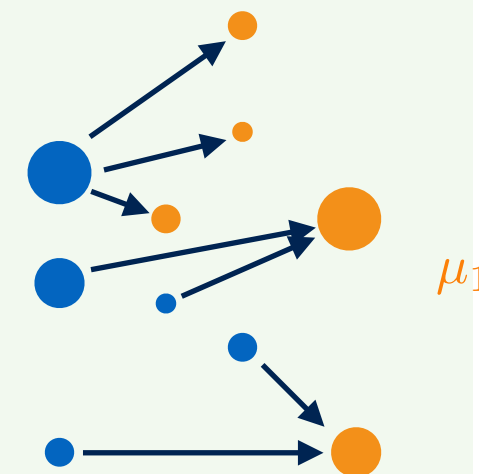
$\Pi(\mu_0, \mu_1)$ = probability distributions on $X \times X$ with marginals μ_0 and μ_1



Discrete case

$$\mu_0 = \sum_i \pi_0^i \delta_{x_i}$$

$$\mu_1 = \sum_j \pi_1^j \delta_{y_j}$$

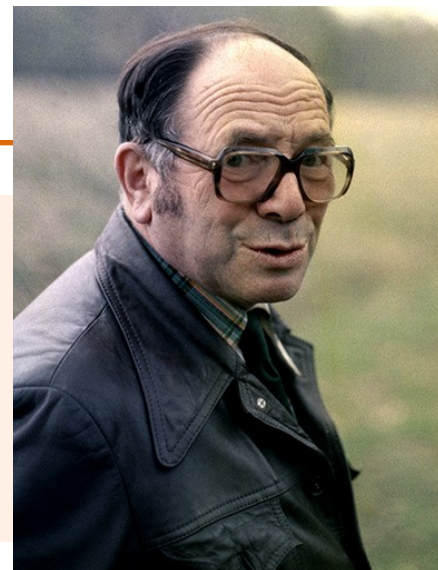


$\Pi(\mu_0, \mu_1) = \{\text{matrices } \gamma \text{ s.t. } \gamma_{i,j} \geq 0,$

$$\sum_i \gamma_{i,j} = \pi_1^j, \sum_j \gamma_{i,j} = \pi_0^i\}$$

$$W_c(\mu_0, \mu_1) = \min_{\gamma \in \Pi(\mu_0, \mu_1)} \sum_{i,j} c(x_i, y_j) \gamma_{ij}$$

Optimal transport

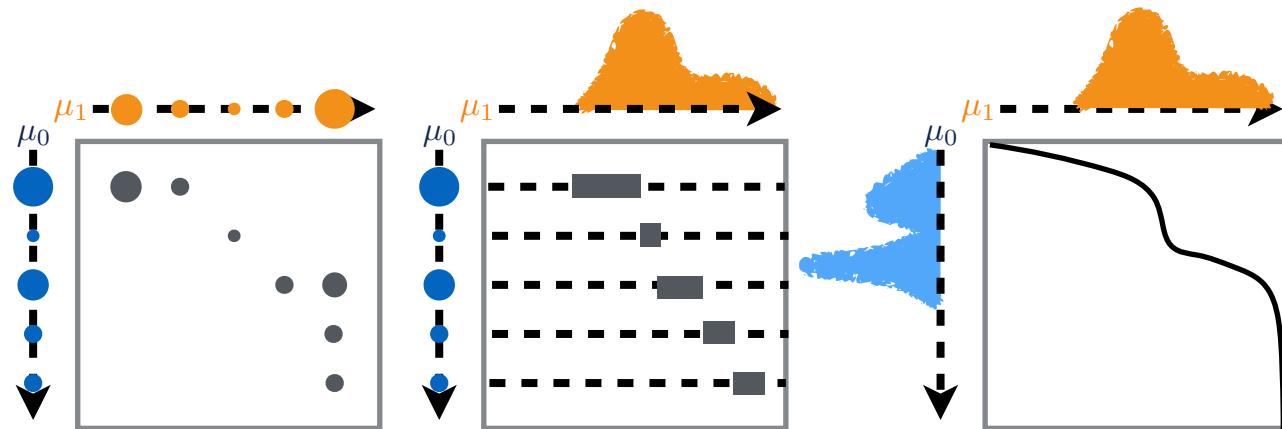


[Kantorovich, *On the transfer of masses*, 1942]

$$W_c(\mu_0, \mu_1) = \inf_{\gamma \in \Pi(\mu_0, \mu_1)} \int_{X \times X} c(x, y) d\gamma(x, y) = \inf_{\substack{(X, Y) \\ X \sim \mu_0, Y \sim \mu_1}} \mathbb{E}[c(X, Y)]$$

Couplings

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



Wasserstein distances

If $c(x, y) = d(x, y)^p$, $p \geq 1$

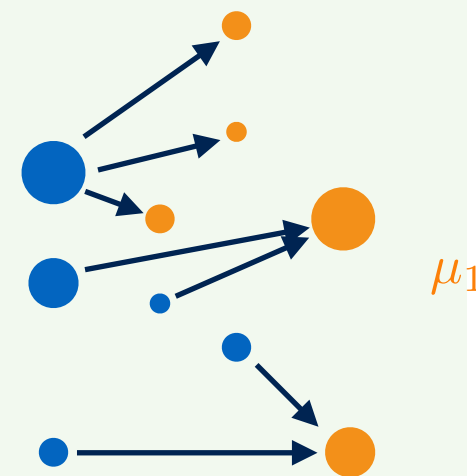
$W_p(\mu_0, \mu_1) := W_c(\mu_0, \mu_1)^{\frac{1}{p}}$ distance

metrizing the weak convergence on $\mathcal{P}_p$

Discrete case

$$\mu_0 = \sum_i \pi_0^i \delta_{x_i}$$

$$\mu_1 = \sum_j \pi_1^j \delta_{y_j}$$



$\Pi(\mu_0, \mu_1) = \{\text{matrices } \gamma \text{ s.t. } \gamma_{i,j} \geq 0,$

$$\sum_i \gamma_{i,j} = \pi_1^j, \sum_j \gamma_{i,j} = \pi_0^i\}$$

$$W_c(\mu_0, \mu_1) = \min_{\gamma \in \Pi(\mu_0, \mu_1)} \sum_{i,j} c(x_i, y_j) \gamma_{ij}$$

Monge-Kantorovich

Monge, 1781

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

Kantorovich, 1939

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

Monge-Kantorovich

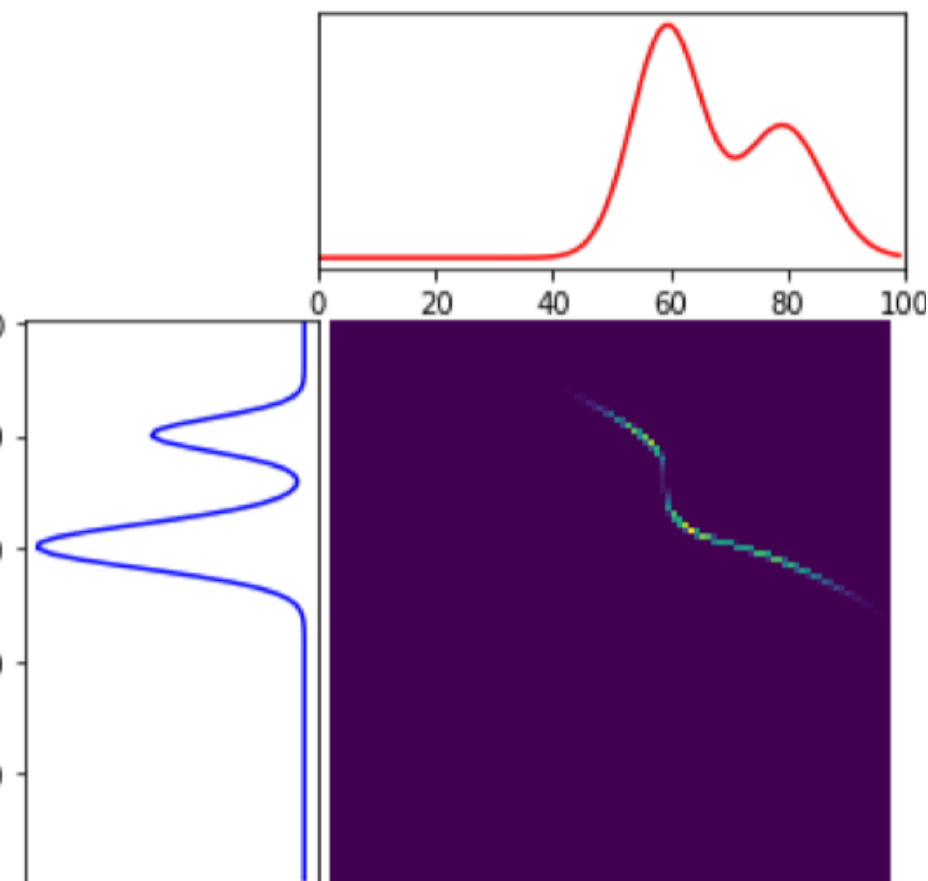
Monge, 1781

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

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



Monge-Kantorovich

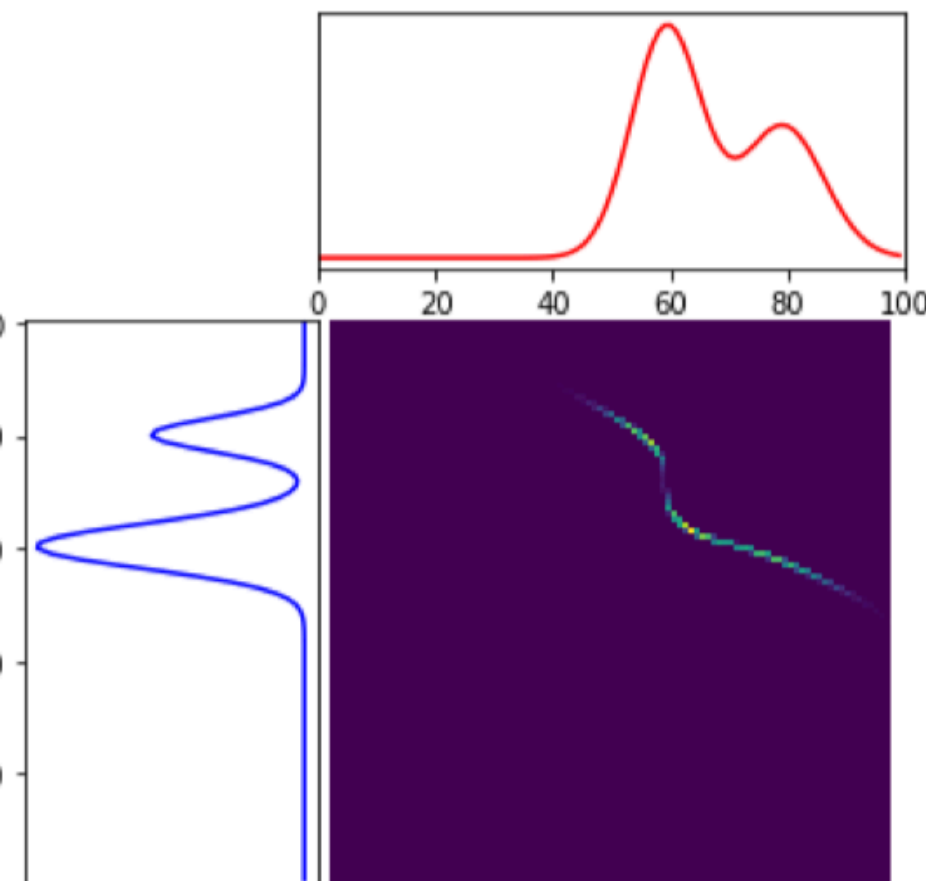
Monge, 1781

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

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



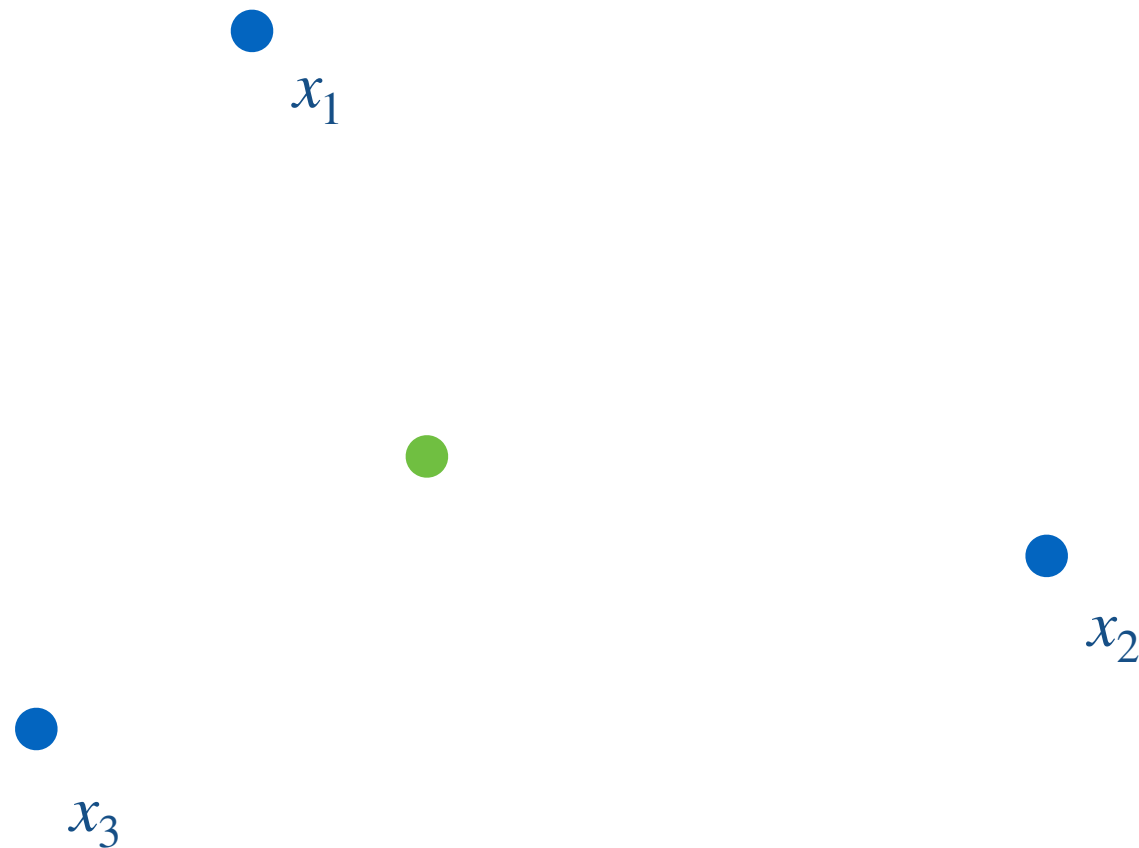
Displacement interpolation:

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

Euclidean barycenters

Barycenter of $(x_i)_{i \in \{1, \dots, p\}}$ for positive weights s.t.

$$\sum_i \lambda_i = 1$$
$$B(x_1, \dots, x_p) := \operatorname{argmin}_x \sum_{i=1}^p \lambda_i \|x_i - x\|^2 = \sum_{i=1}^p \lambda_i x_i$$



Wasserstein barycenters

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

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

Wasserstein barycenters

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

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

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.

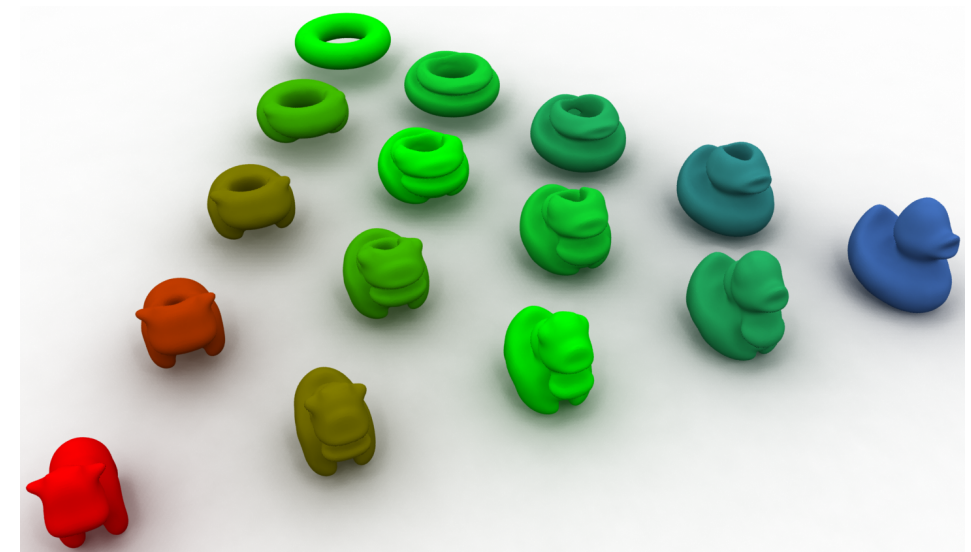
Wasserstein barycenters

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

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

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]



Wasserstein barycenters

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

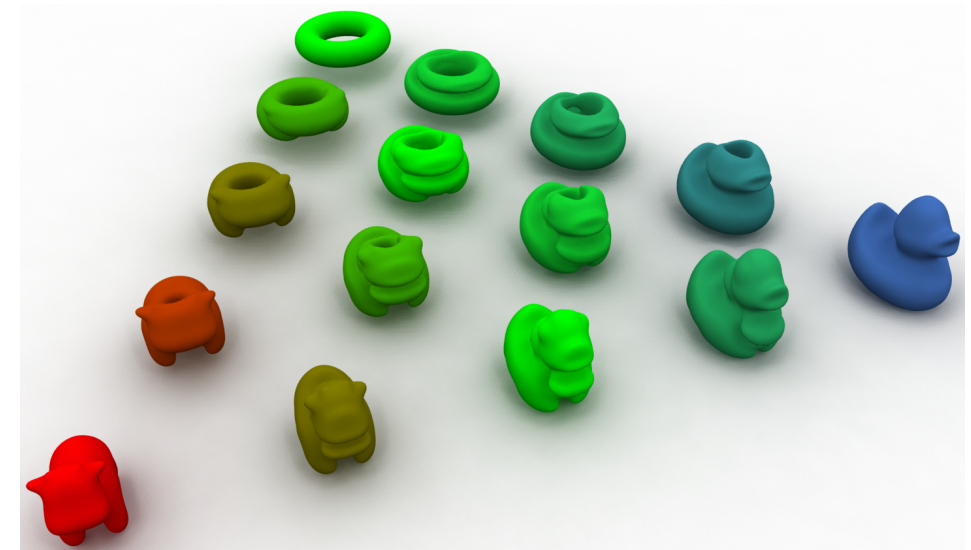
$$\mu_{bary} \in \operatorname{argmin}_{\rho} \sum_i \lambda_i W_2^2(\mu_i, \rho)$$

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.



[Levy, 2015]

[Solomon et al. 2015]



Wasserstein barycenters

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

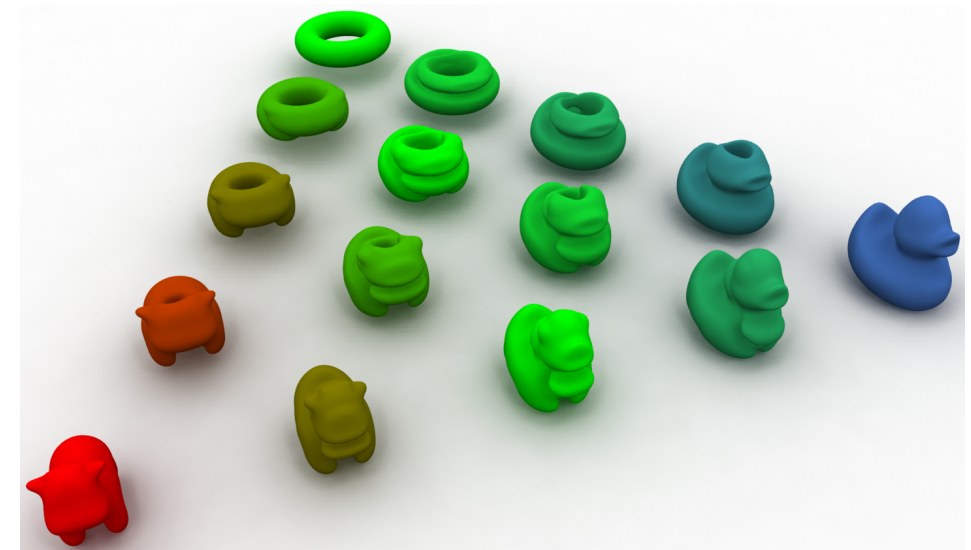
$$\mu_{bary} \in \operatorname{argmin}_{\rho} \sum_i \lambda_i W_2^2(\mu_i, \rho)$$

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.



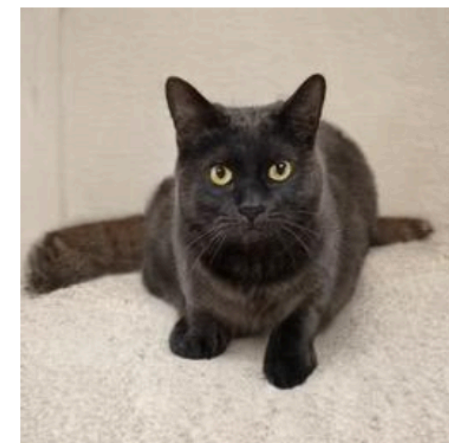
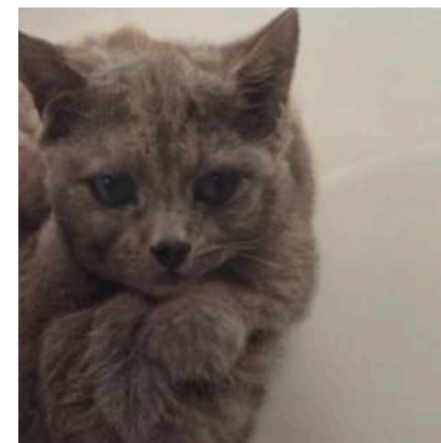
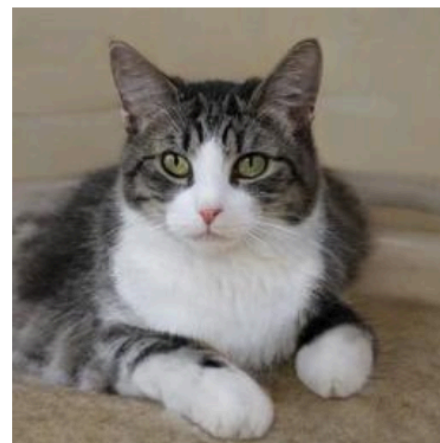
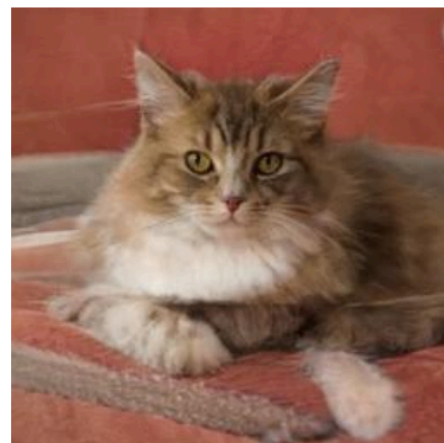
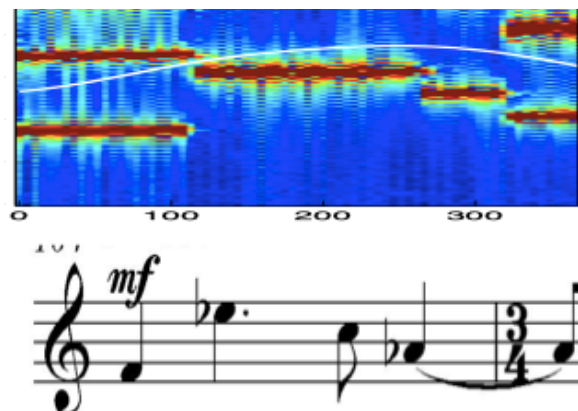
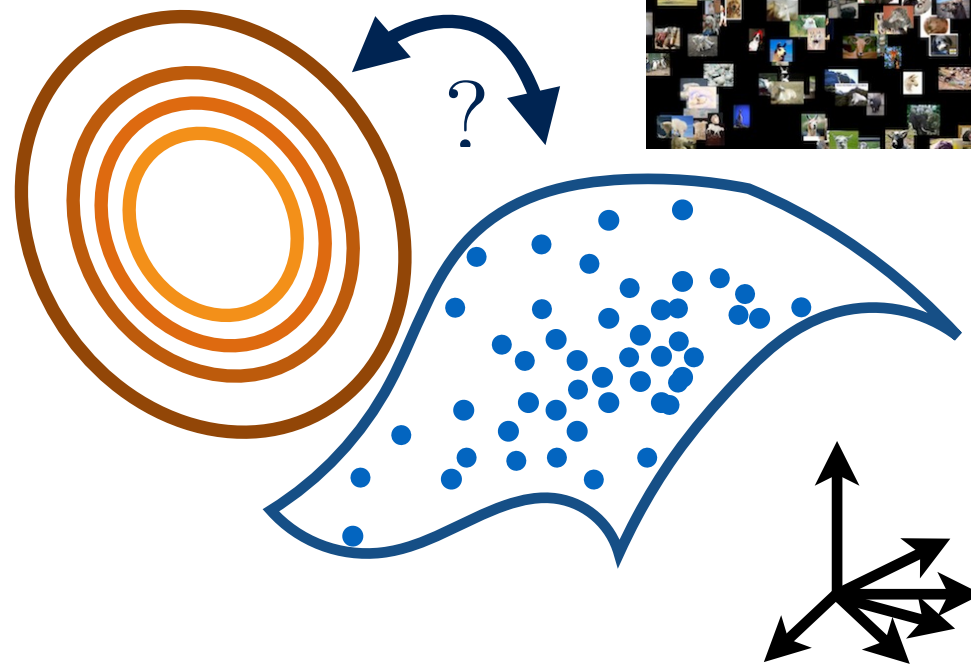
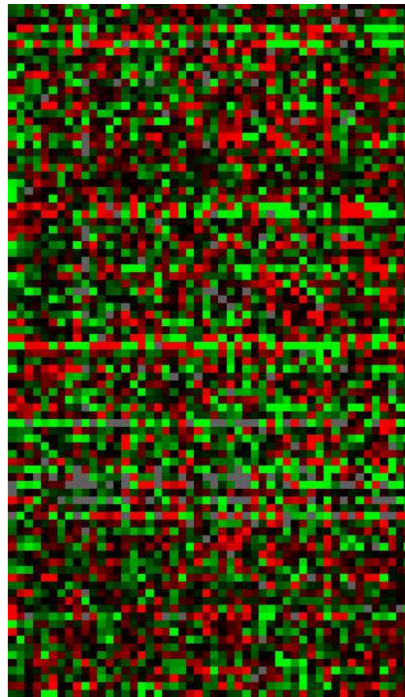
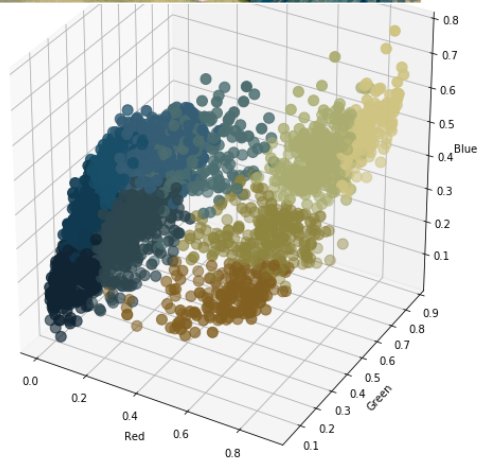
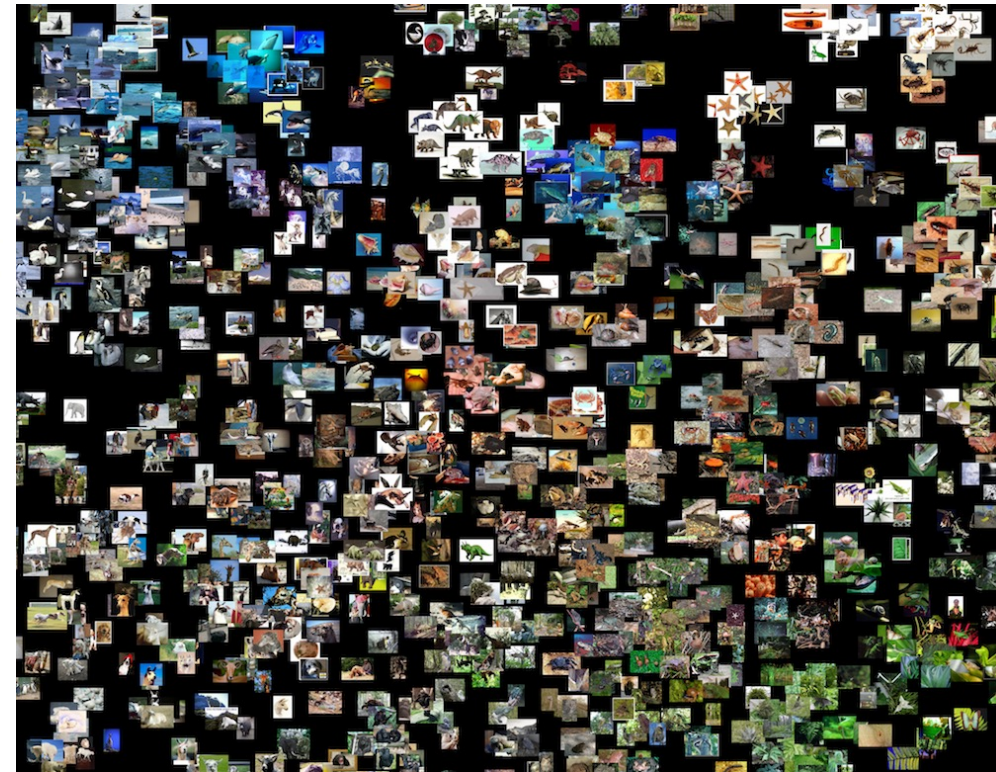
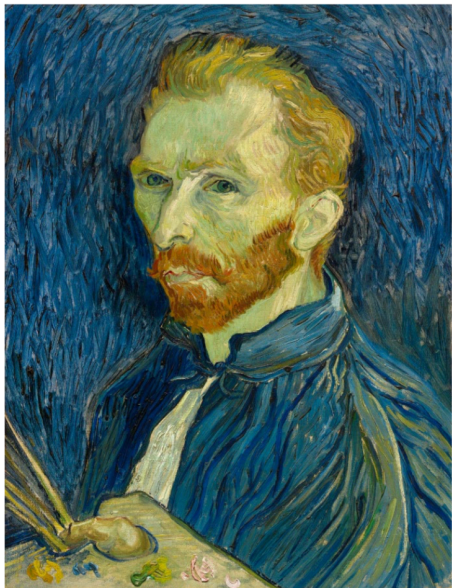
[Levy, 2015]

[Solomon et al. 2015]



[Vacher et al., 2020]

Optimal Transport in data science



<https://thesecatsdonotexist.com/>

Optimal transport in one dimension

Optimal Transport in 1D

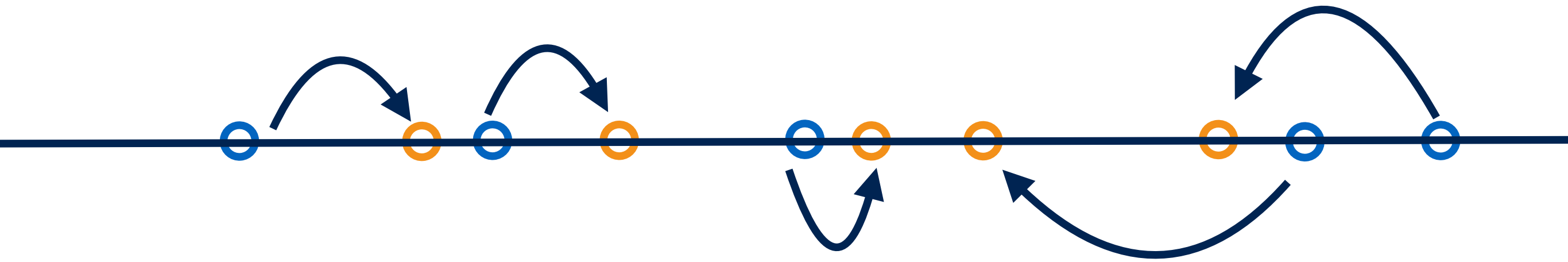


Optimal Transport in 1D



If $c(x, y) = f(x - y)$ with f convex,
then for $x_1 < x_2$ and $y_1 < y_2$,
$$c(x_1, y_1) + c(x_2, y_2) \leq c(x_1, y_2) + c(x_2, y_1)$$

Optimal Transport in 1D



If $c(x, y) = f(x - y)$ with f convex,
then for $x_1 < x_2$ and $y_1 < y_2$,
$$c(x_1, y_1) + c(x_2, y_2) \leq c(x_1, y_2) + c(x_2, y_1)$$



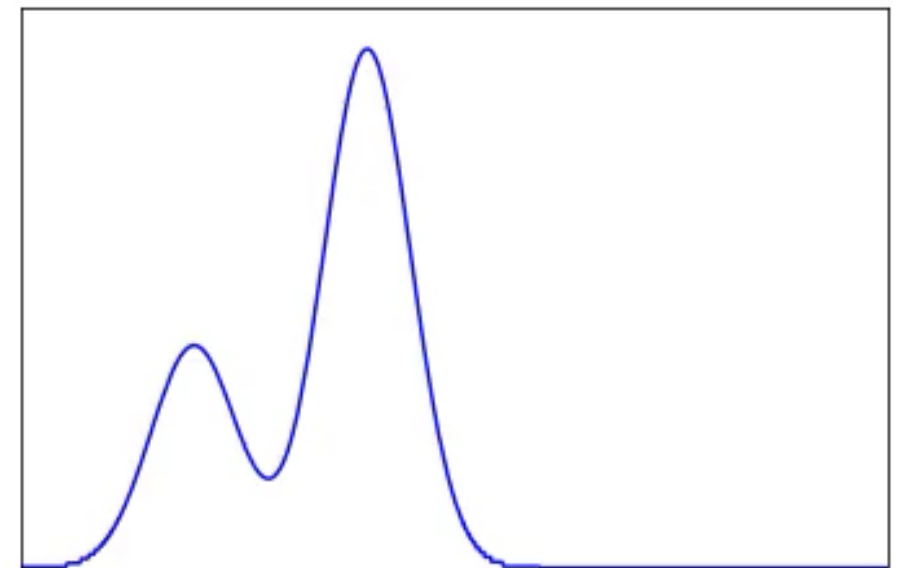
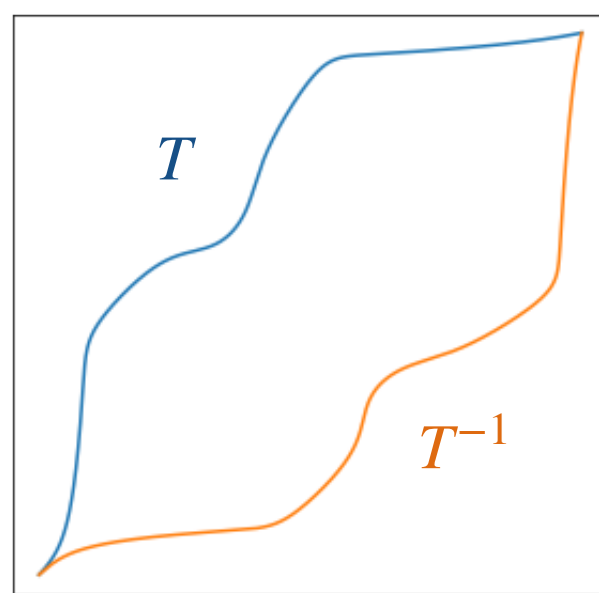
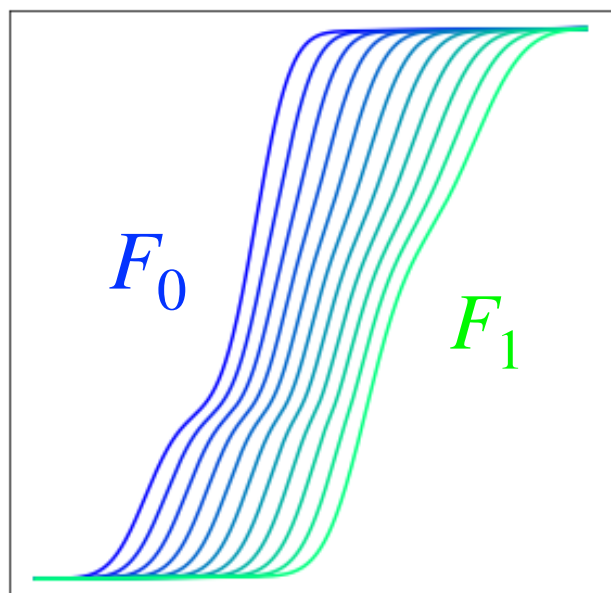
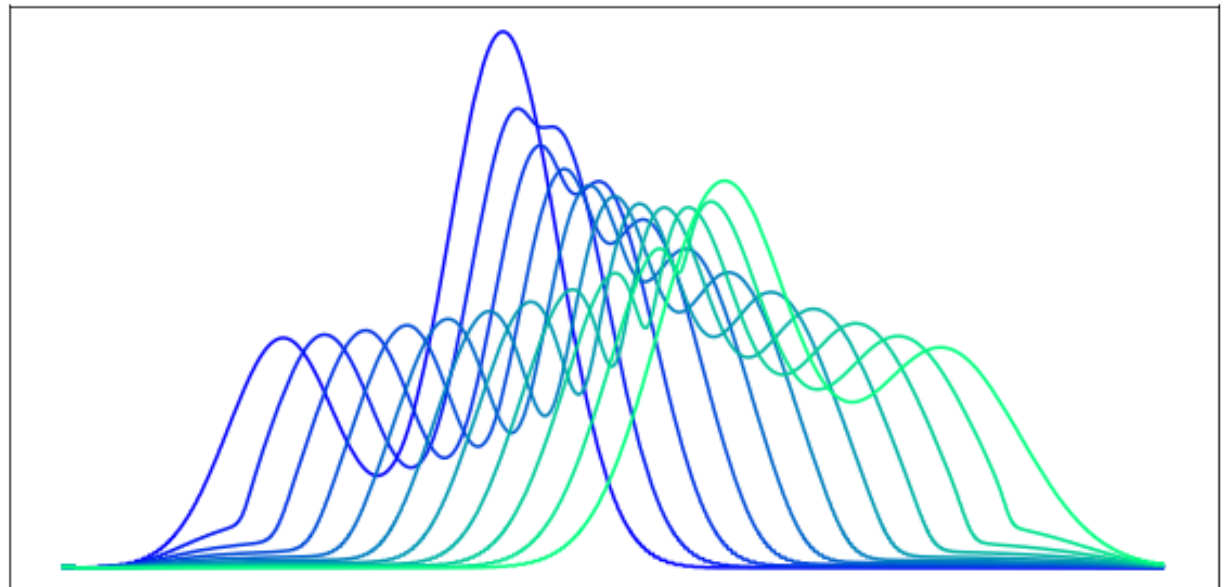
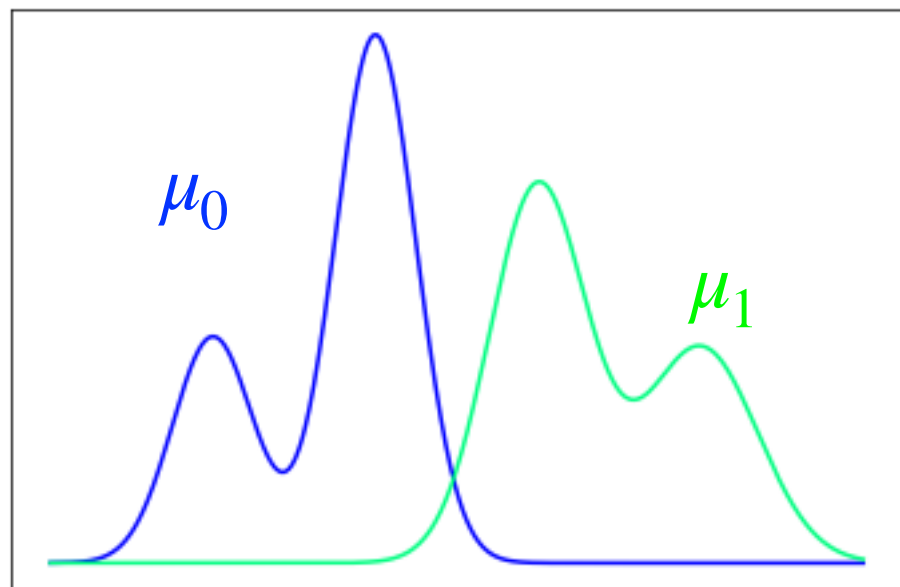
solution given by the **monotone**
rearrangement of μ_0 onto μ_1

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.

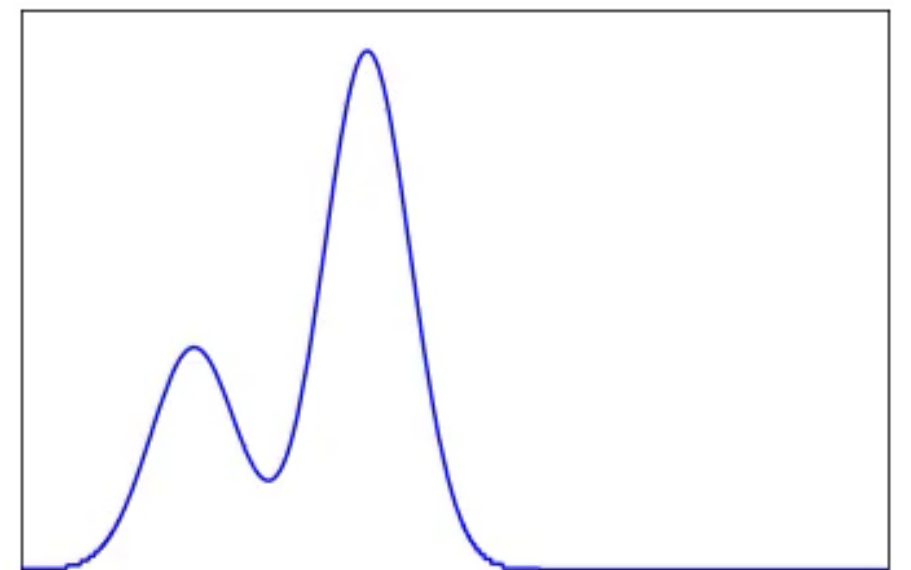
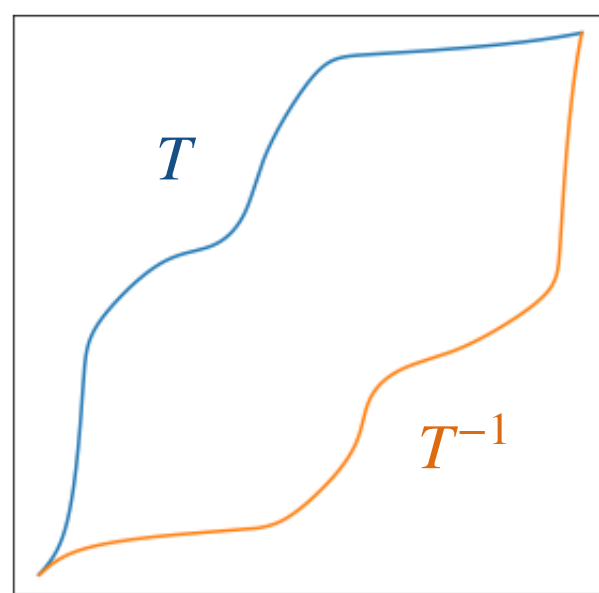
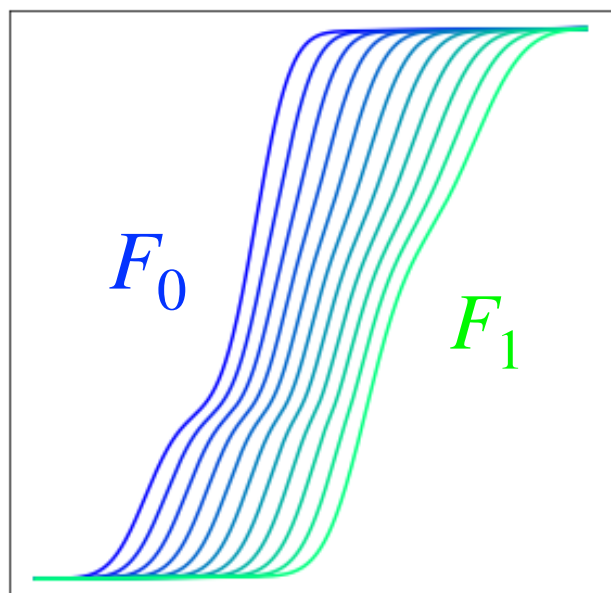
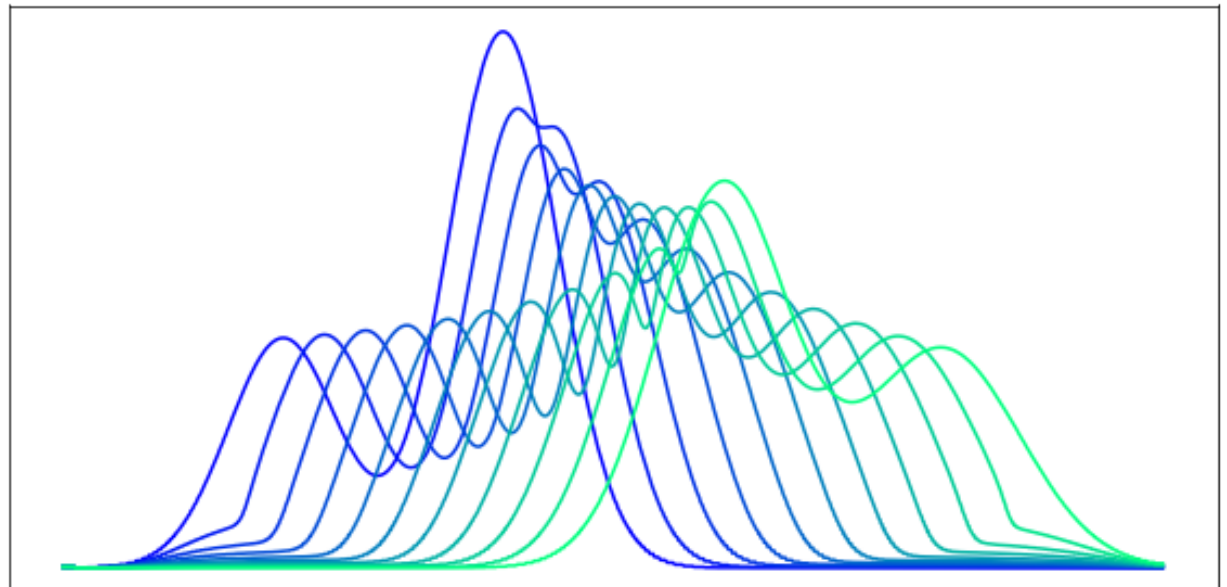
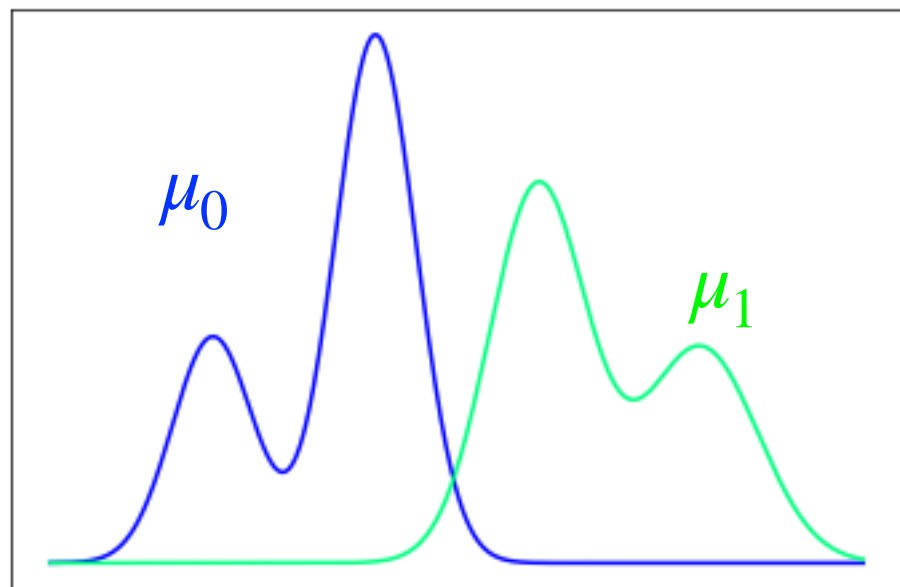


Optimal Transport in 1D

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

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

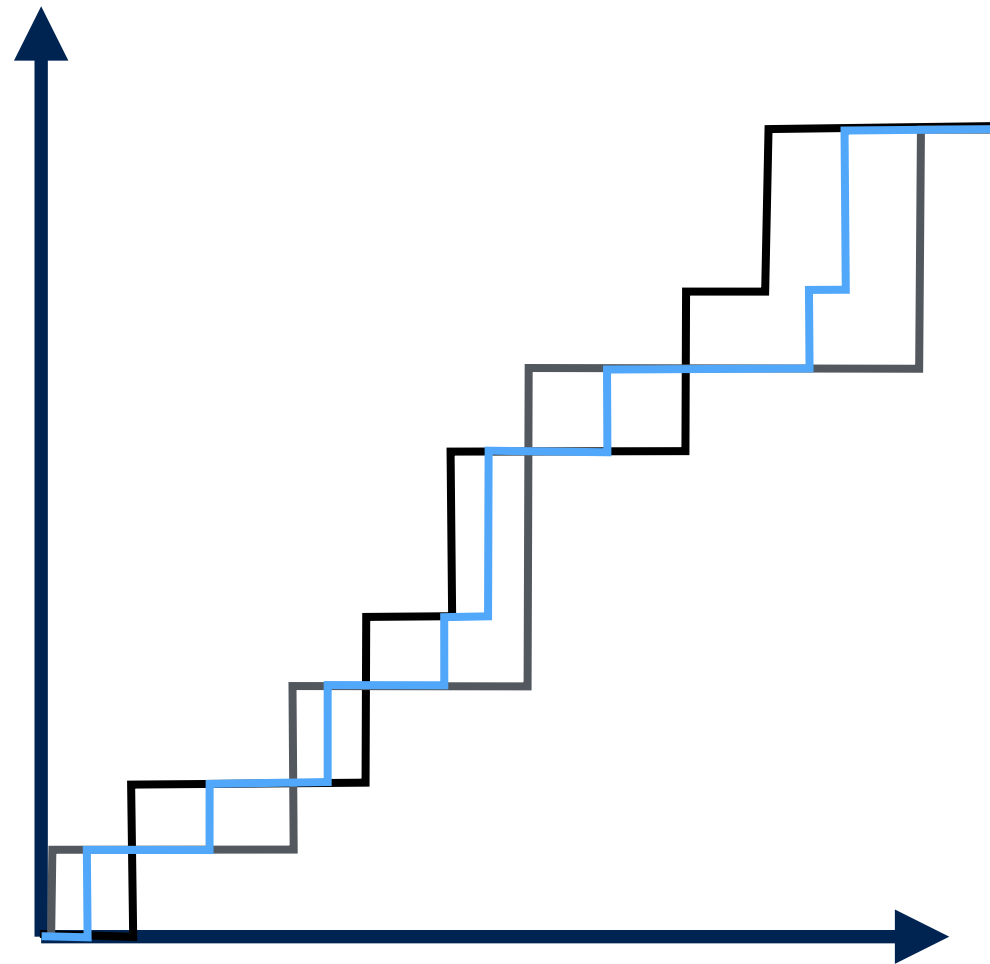
with F_0 and F_1 the distribution functions of μ_0 and μ_1 . Moreover, if μ_0 has no atoms, $T = F_1^{-1} \circ F_0$ is solution of the Monge problem.



Midway histogram

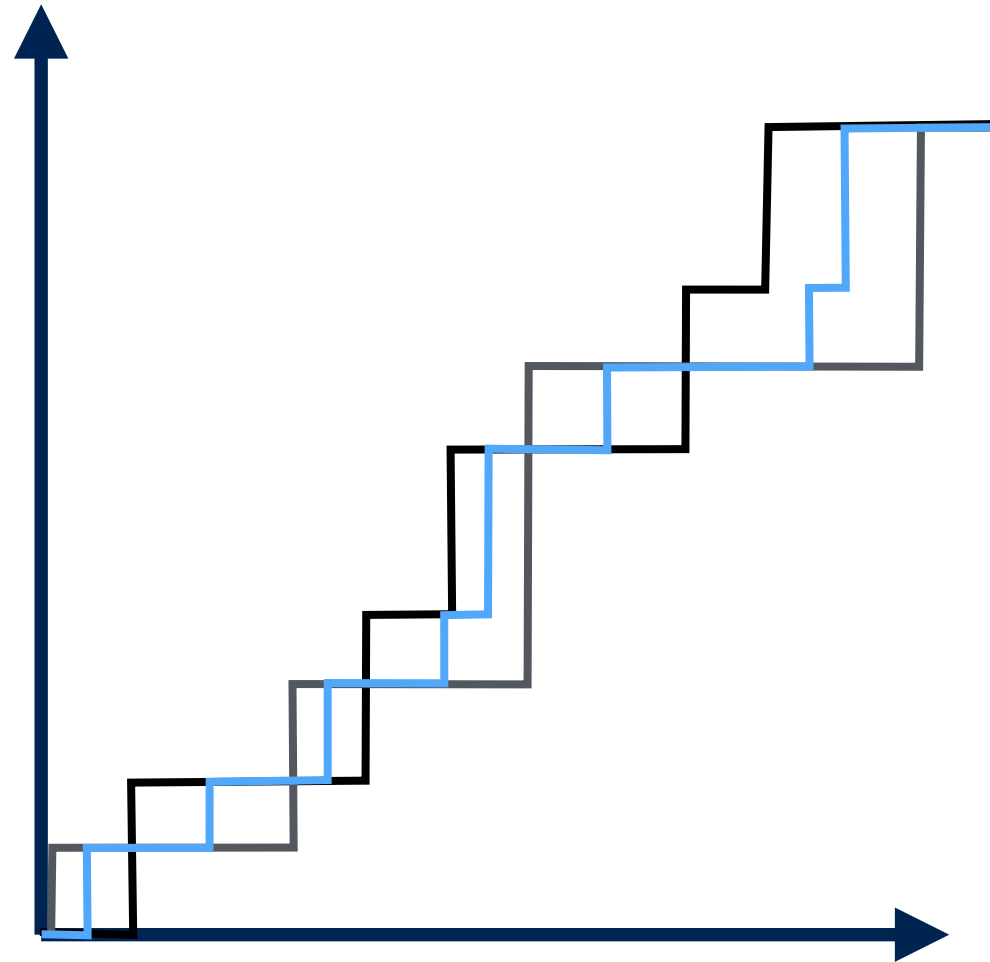


$$\left(\frac{F_0^{-1} + F_1^{-1}}{2} \right)^{-1}$$



Midway histogram

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



Numerical approaches

Linear programming

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

(LP)

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

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

Linear programming

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

(LP)

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

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

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

Linear programming

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

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

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

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

Assignment: Hungarian algo. [Kuhn 1955] $O(N^3)$, Auction [Bertsekas 1992]

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

Dynamic formulation [Brenier, Benamou 2000]

Semi-discrete OT [Mérigot 11, Levy 15]

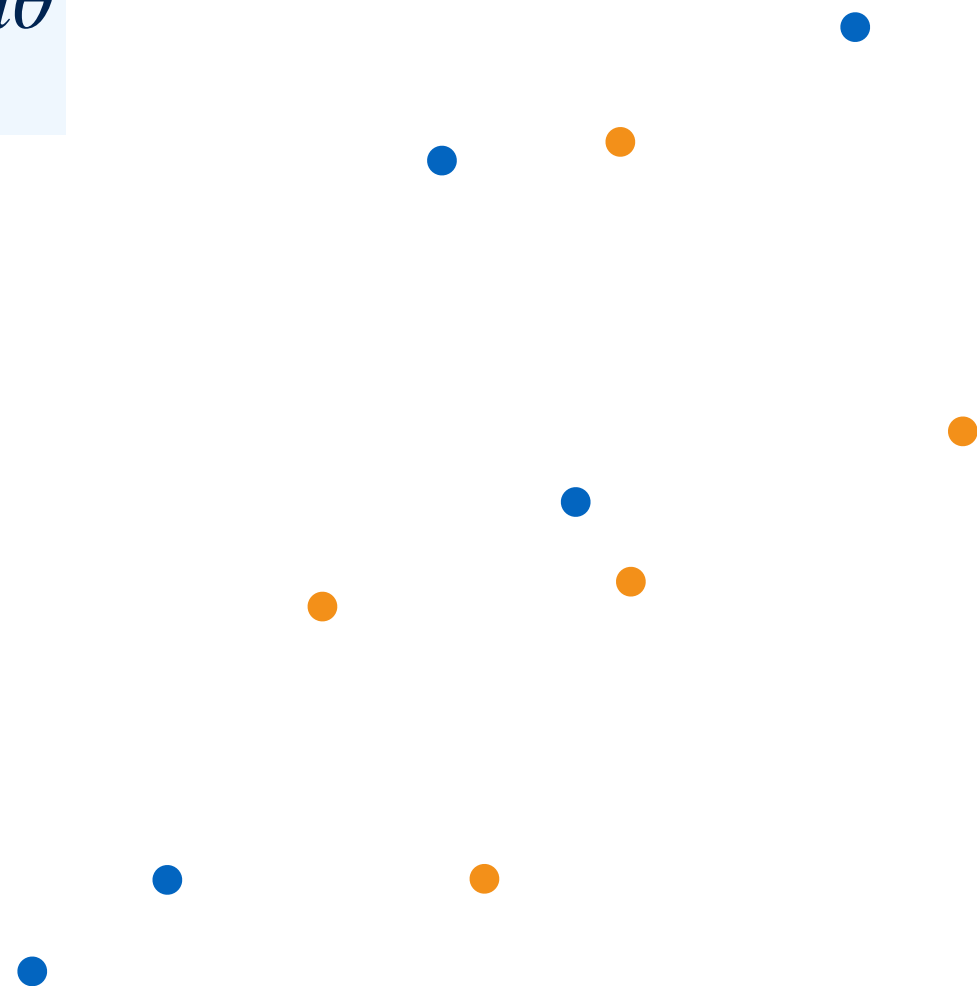
Sliced OT [Rabin et al. 11, Rabin et al. 15]

Entropic OT [Cuturi 13,...]

Sliced optimal transport

Replace classical OT by

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



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

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

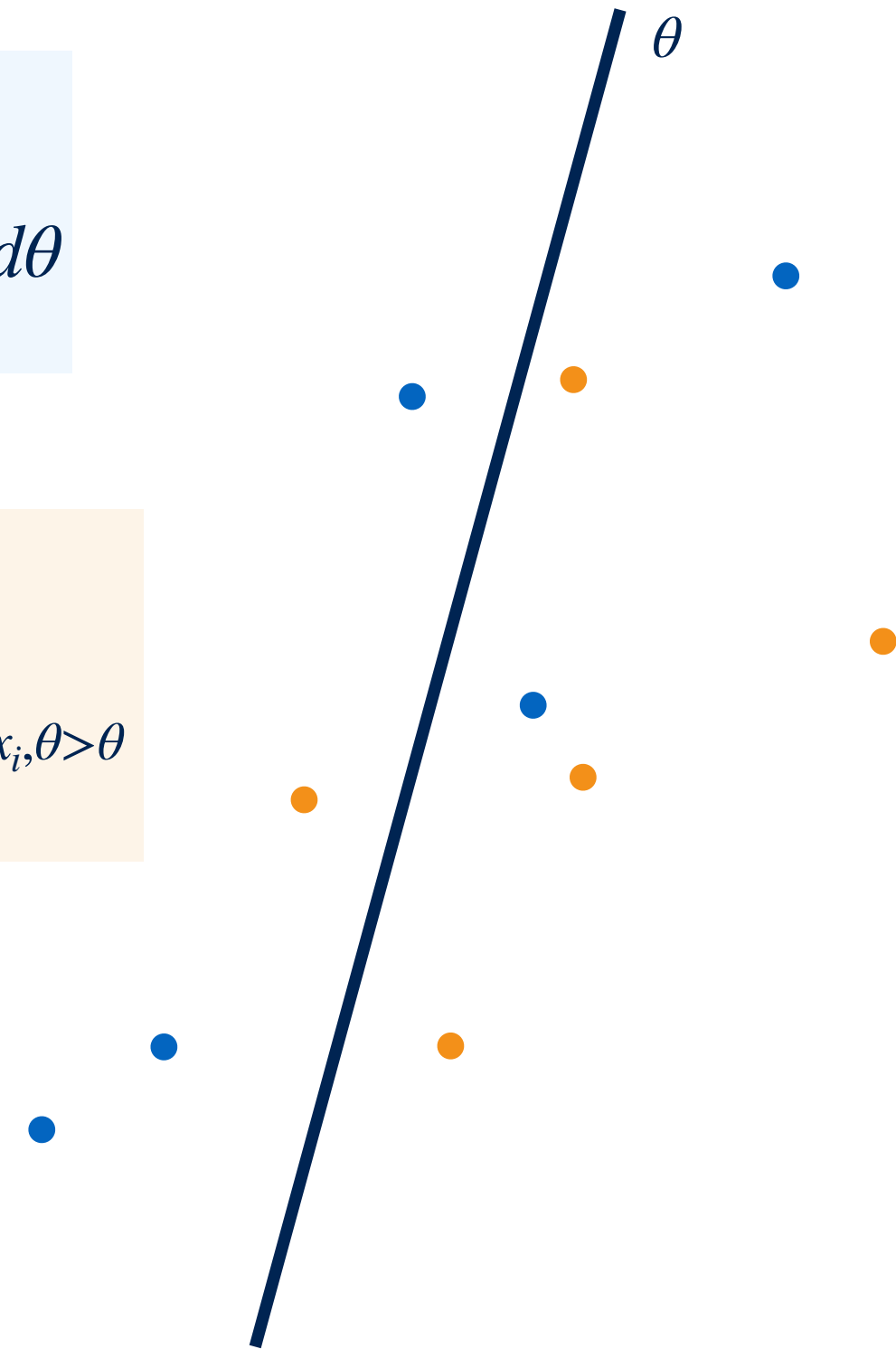
Sliced optimal transport

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

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



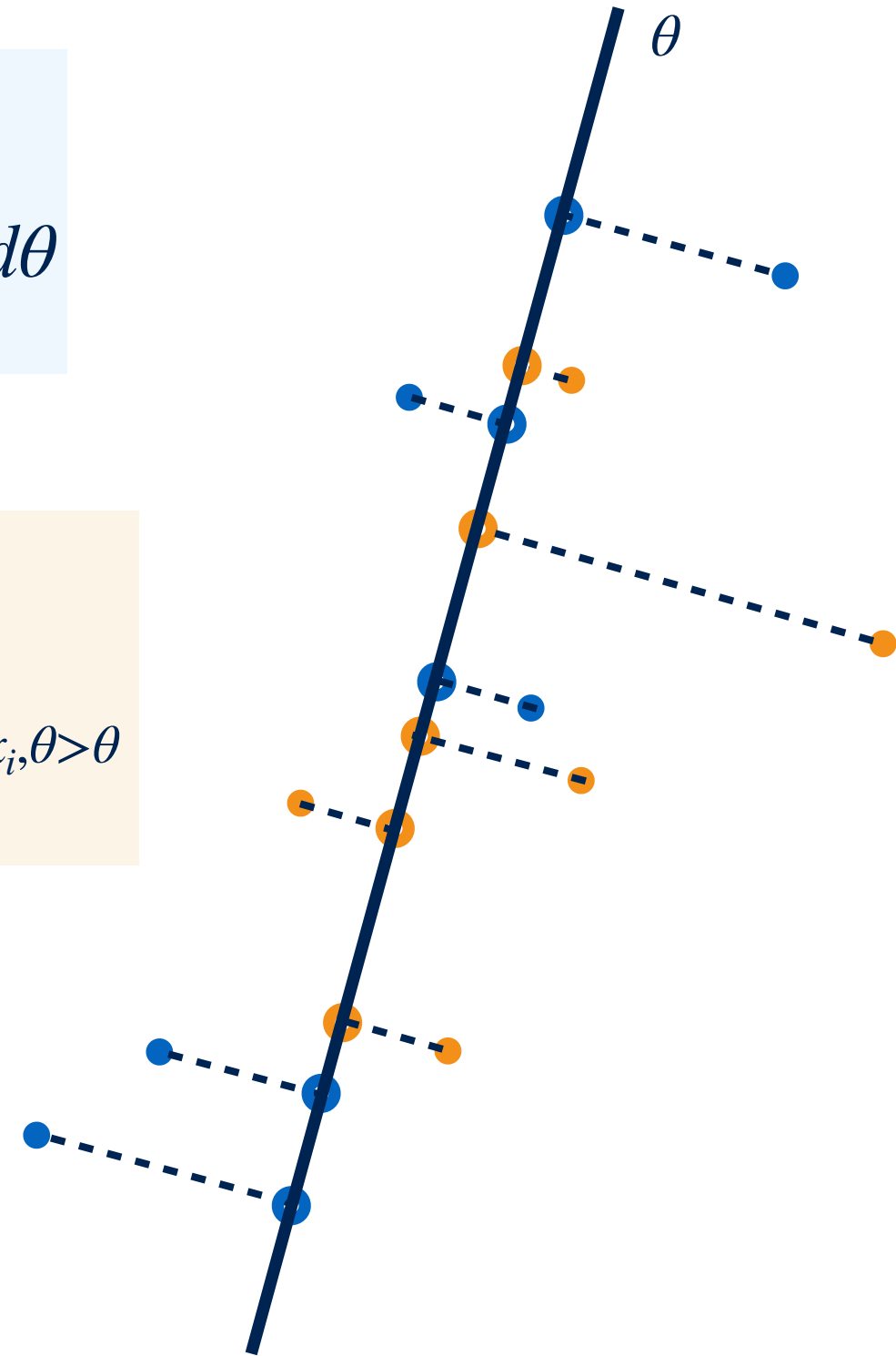
Sliced optimal transport

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

$$\mu_0 = \frac{1}{n} \sum_{i=1}^n \delta_{x_i}, \quad \mu_1 = \frac{1}{n} \sum_{j=1}^n \delta_{y_j}, \quad p_\theta \# \mu_0 = \frac{1}{n} \sum_i \delta_{\langle x_i, \theta \rangle \theta}$$



Sliced optimal transport

Replace classical OT by

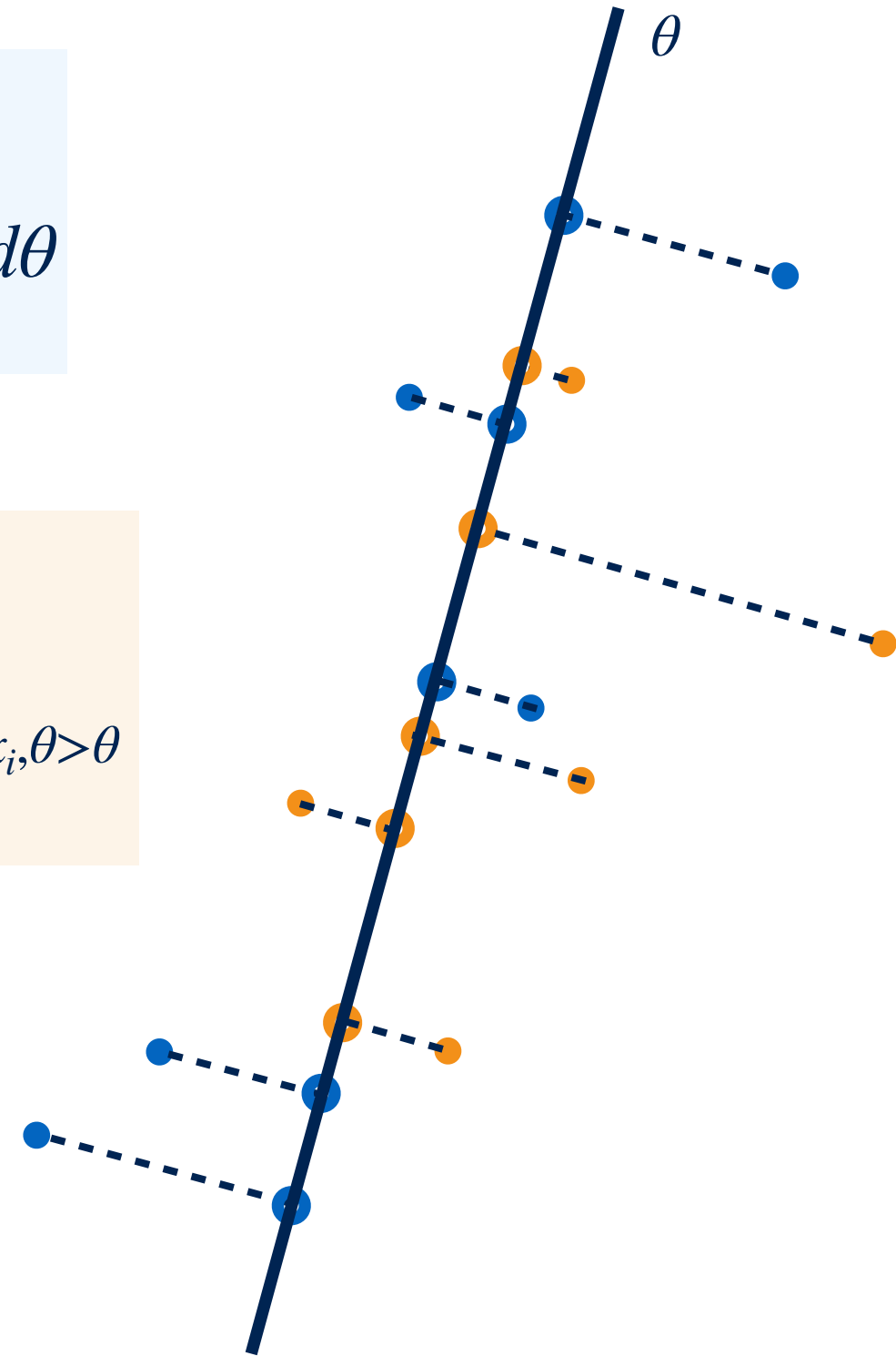
$$SW_2^2(\mu_0, \mu_1) = \int_{\mathbb{S}^{d-1}} W_2^2(p_\theta \# \mu_0, p_\theta \# \mu_1) d\theta$$

Discrete measures

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

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

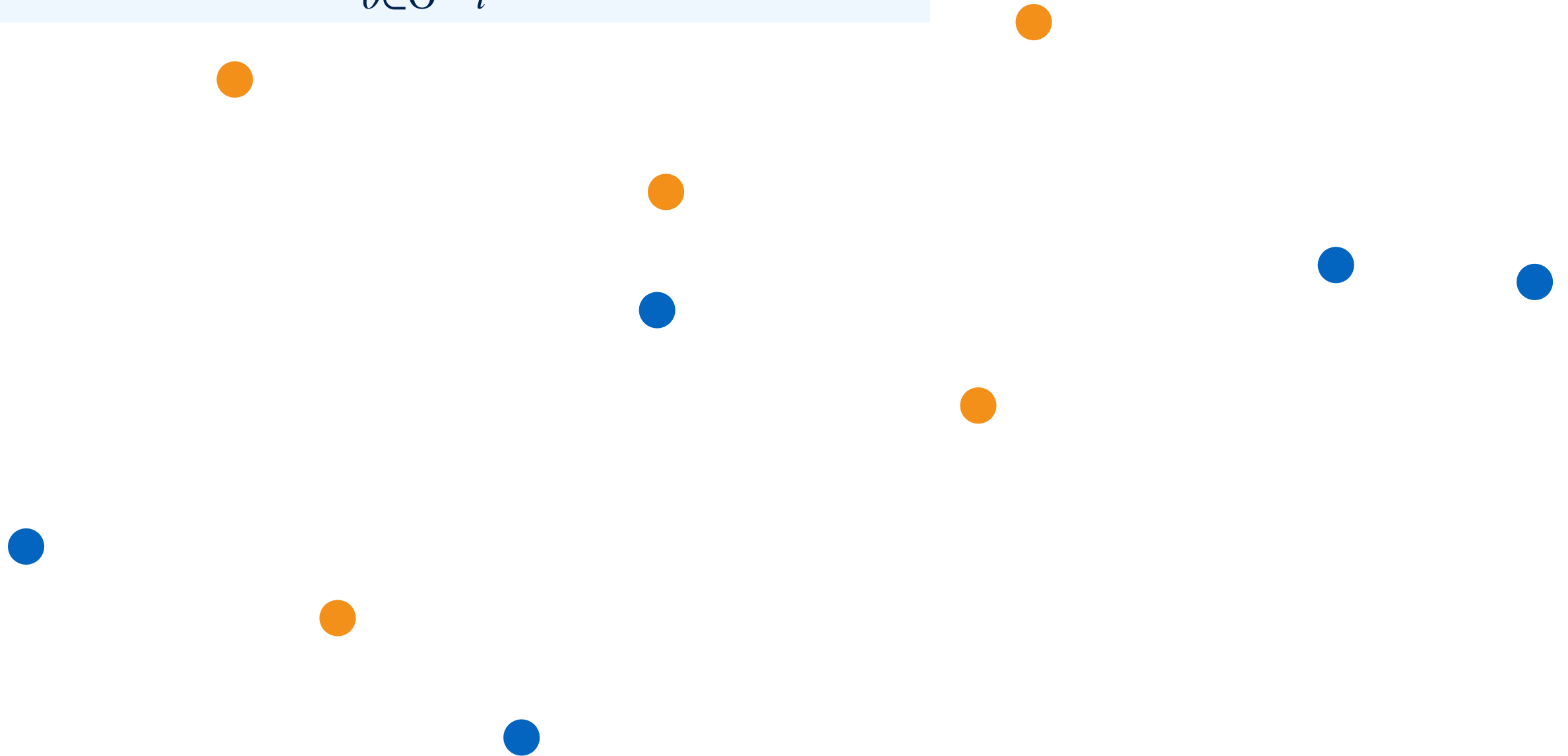
with σ_θ monotone rearrangement
between $\langle \mu_0, \theta \rangle$ and $\langle \mu_1, \theta \rangle$.



Sliced optimal transport

Stochastic gradient descent on

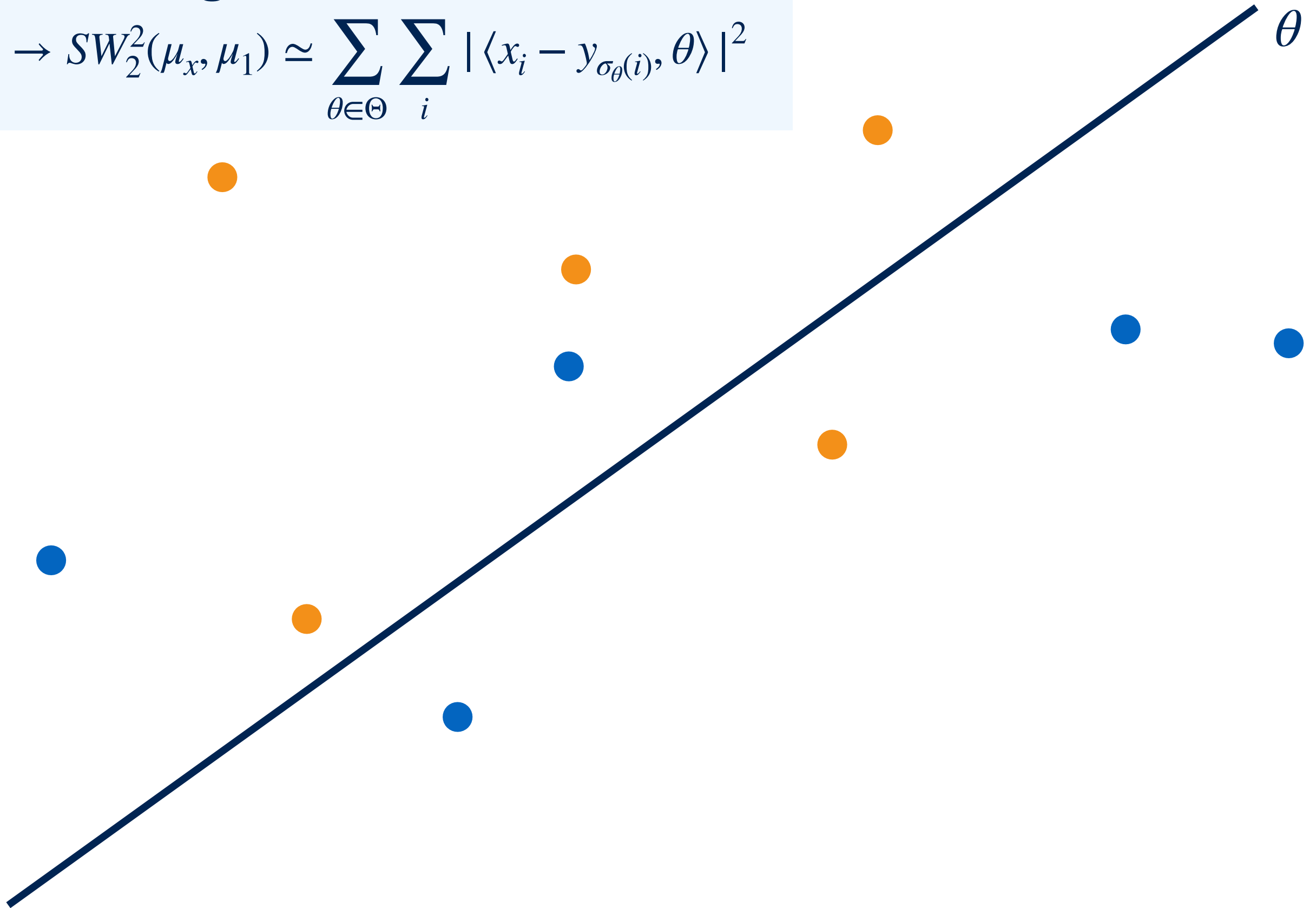
$$x \rightarrow SW_2^2(\mu_x, \mu_1) \simeq \sum_{\theta \in \Theta} \sum_i |\langle x_i - y_{\sigma_\theta(i)}, \theta \rangle|^2$$



Sliced optimal transport

Stochastic gradient descent on

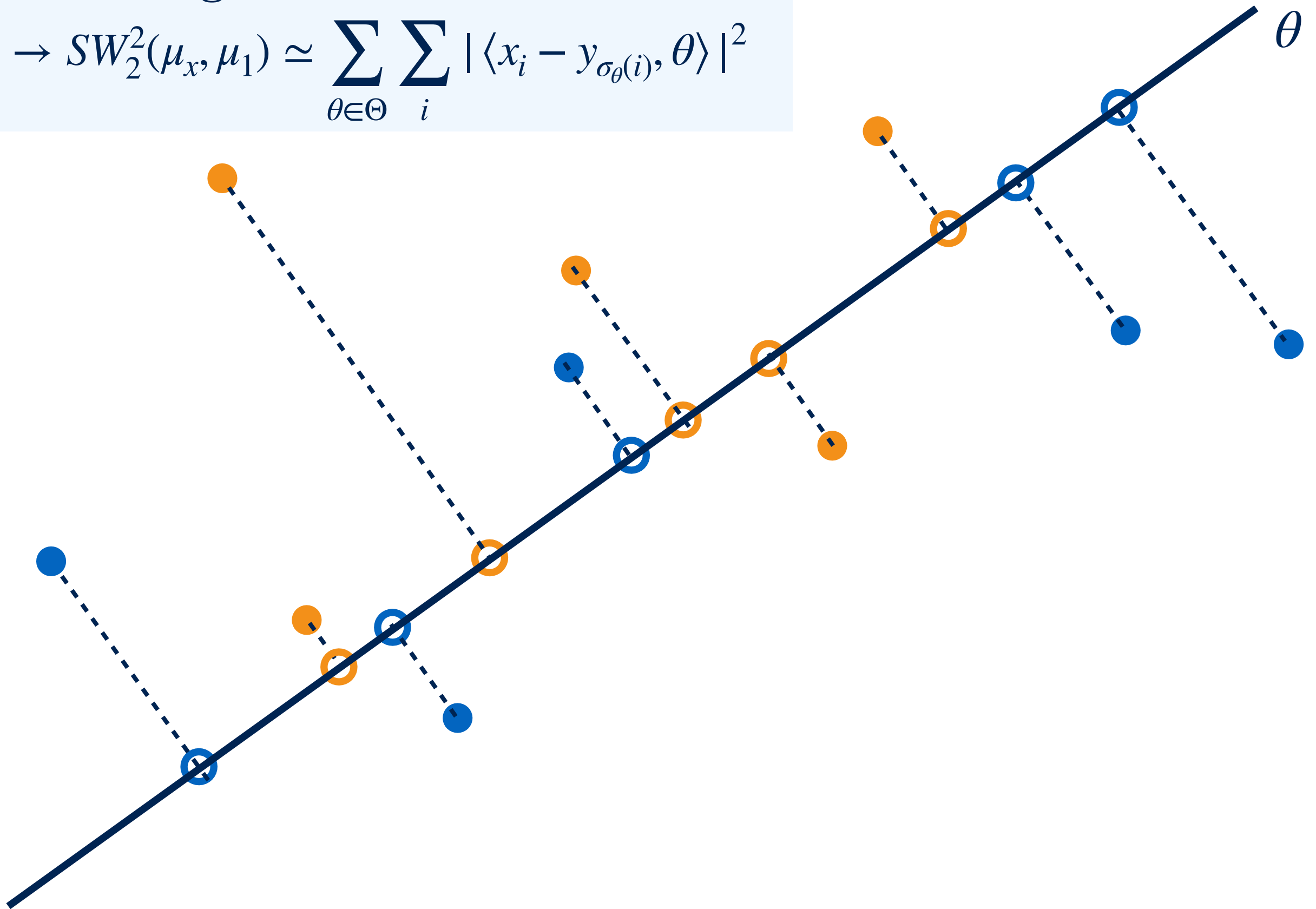
$$x \rightarrow SW_2^2(\mu_x, \mu_1) \simeq \sum_{\theta \in \Theta} \sum_i |\langle x_i - y_{\sigma_\theta(i)}, \theta \rangle|^2$$



Sliced optimal transport

Stochastic gradient descent on

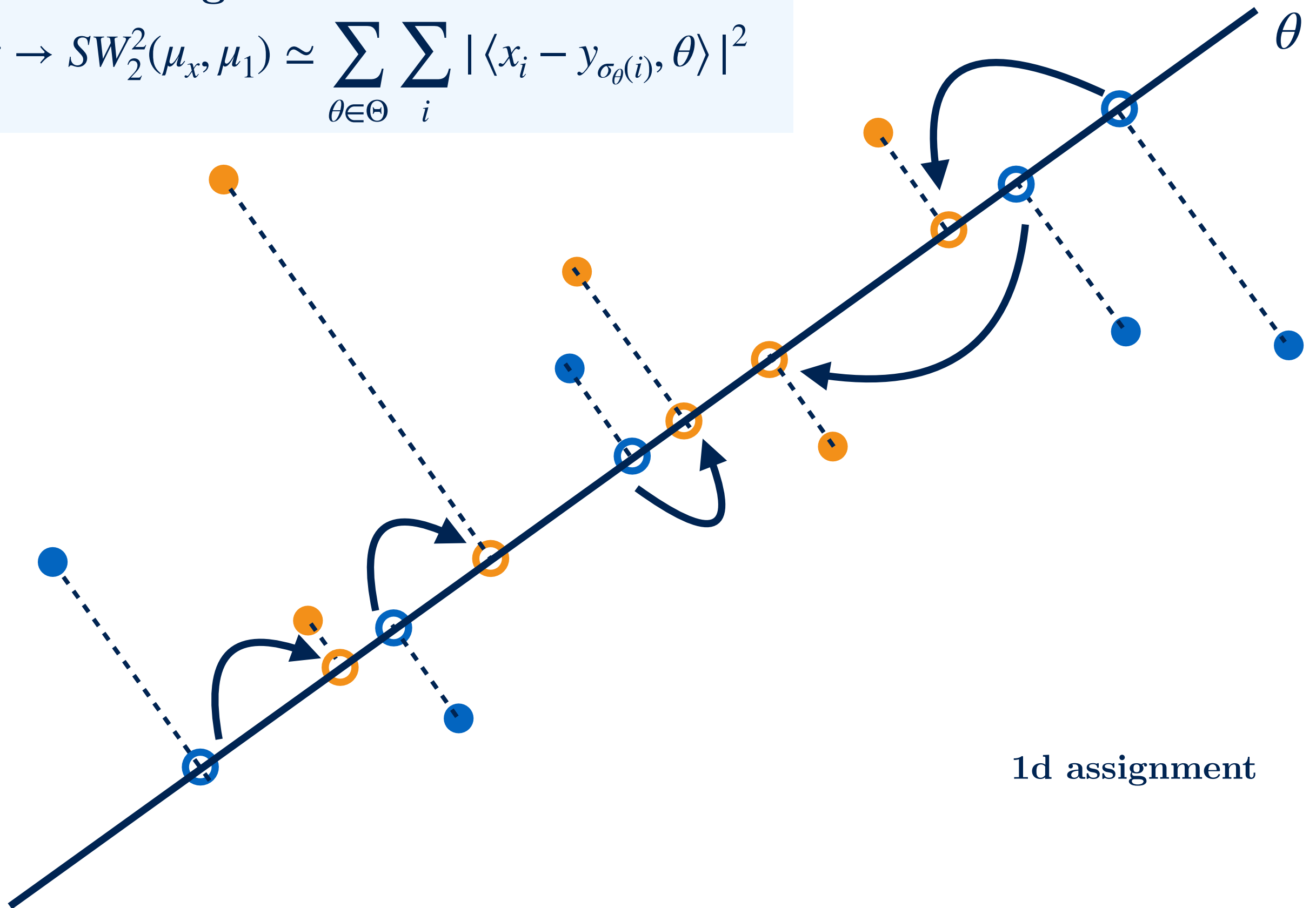
$$x \rightarrow SW_2^2(\mu_x, \mu_1) \simeq \sum_{\theta \in \Theta} \sum_i |\langle x_i - y_{\sigma_\theta(i)}, \theta \rangle|^2$$



Sliced optimal transport

Stochastic gradient descent on

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

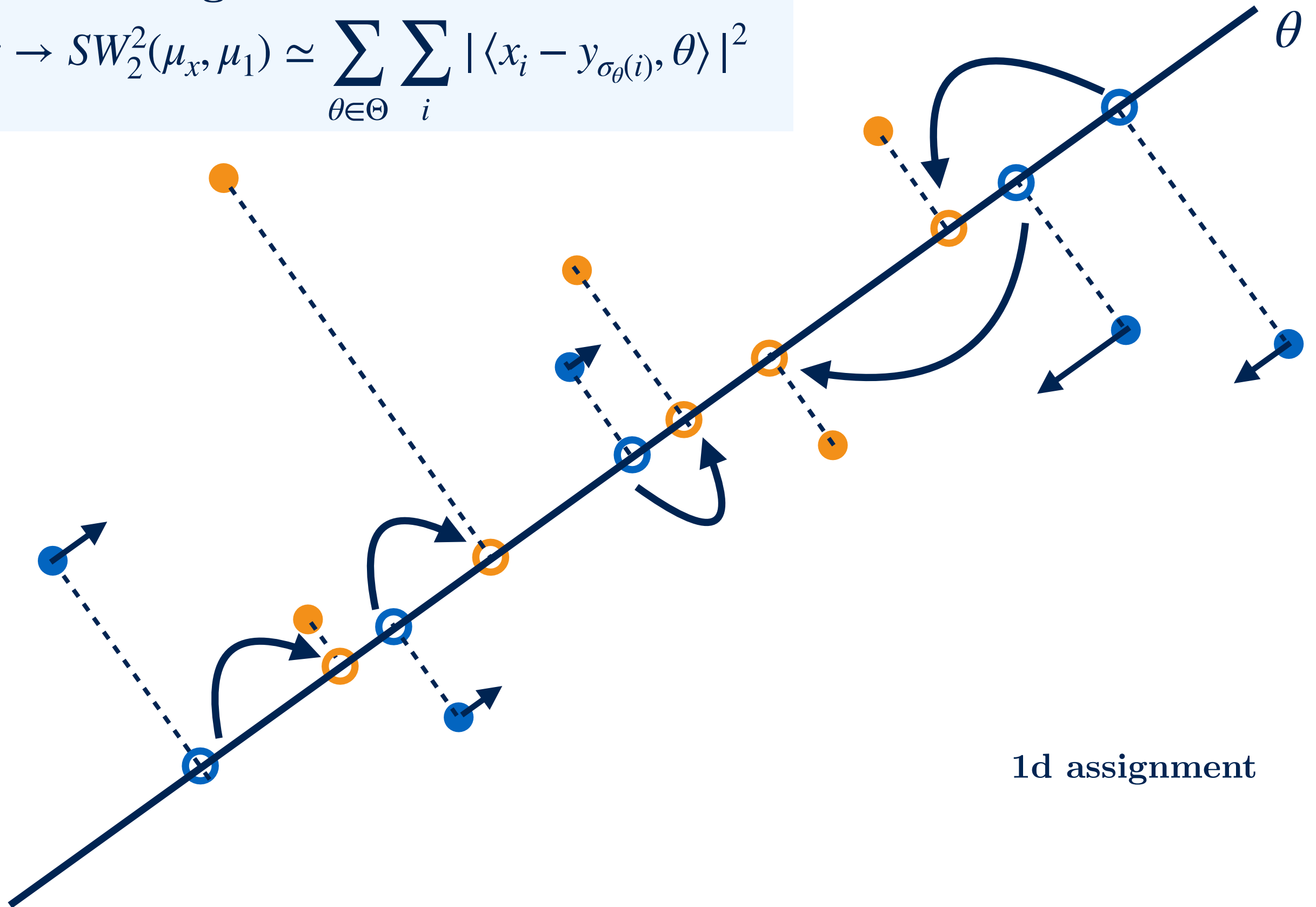


1d assignment

Sliced optimal transport

Stochastic gradient descent on

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

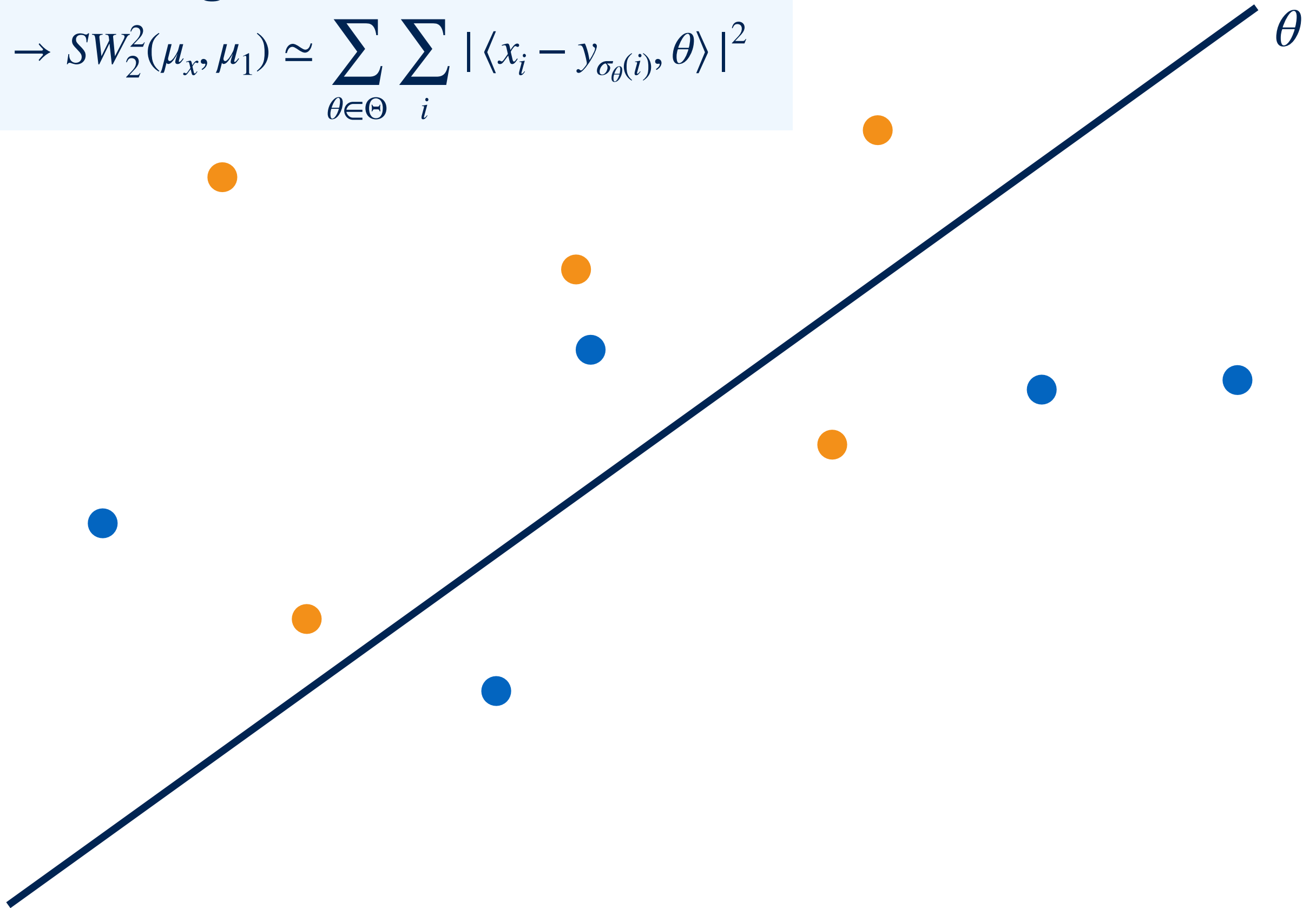


1d assignment

Sliced optimal transport

Stochastic gradient descent on

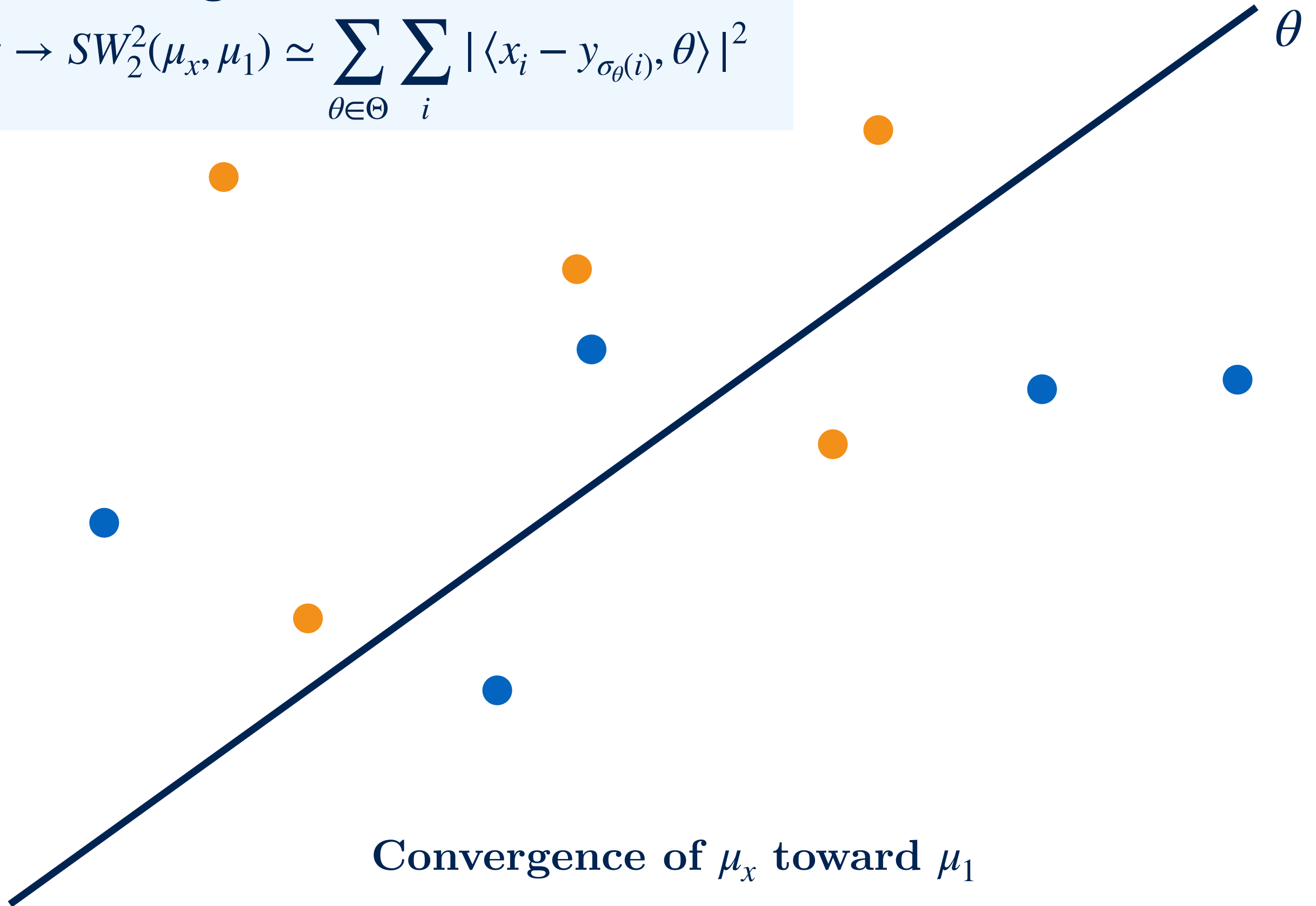
$$x \rightarrow SW_2^2(\mu_x, \mu_1) \simeq \sum_{\theta \in \Theta} \sum_i |\langle x_i - y_{\sigma_\theta(i)}, \theta \rangle|^2$$



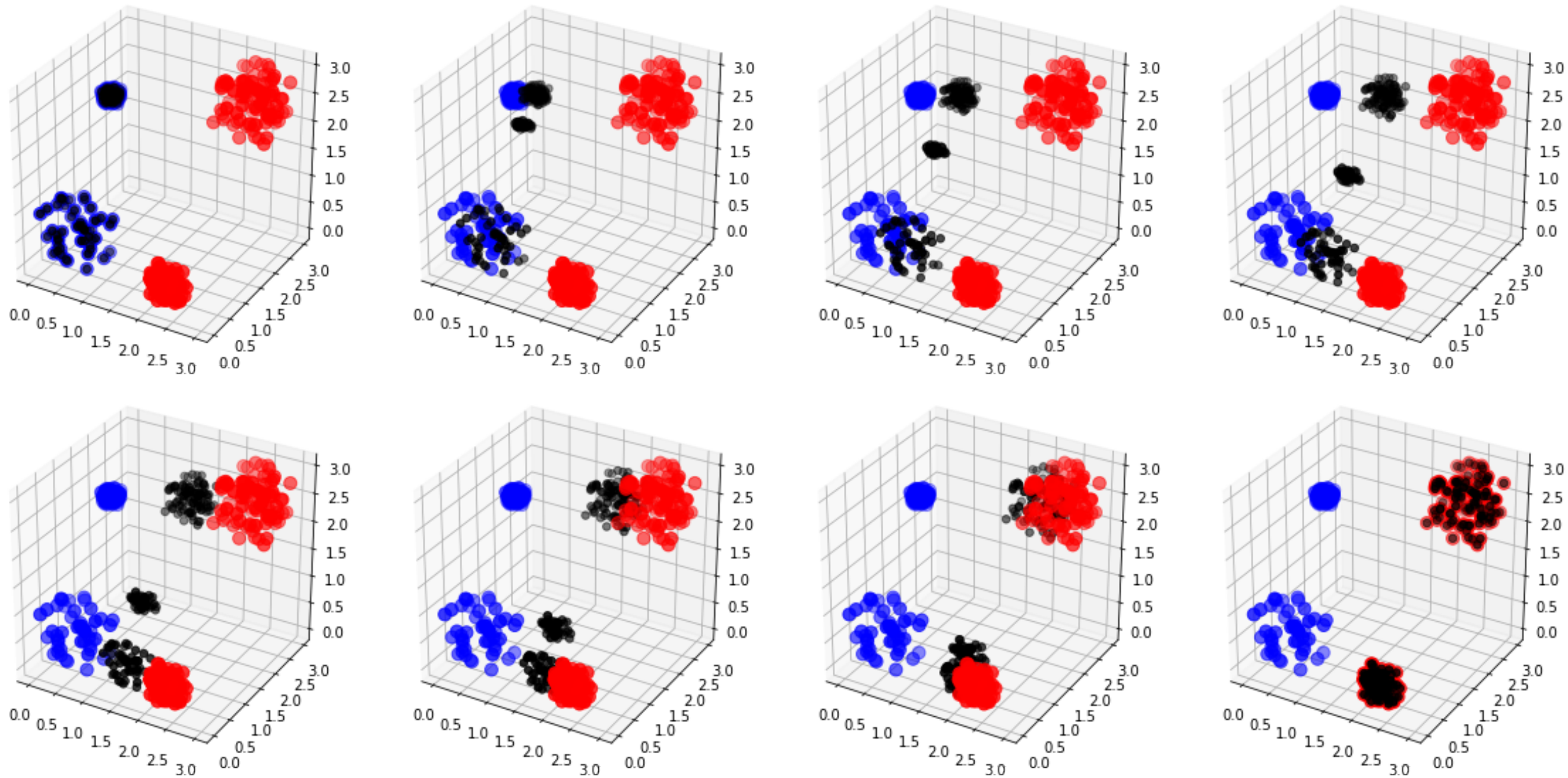
Sliced optimal transport

Stochastic gradient descent on

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



Sliced optimal transport



practical session on *Color transfer*

Entropic OT

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

Entropic OT [Cuturi '13]

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

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

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

Entropic OT

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

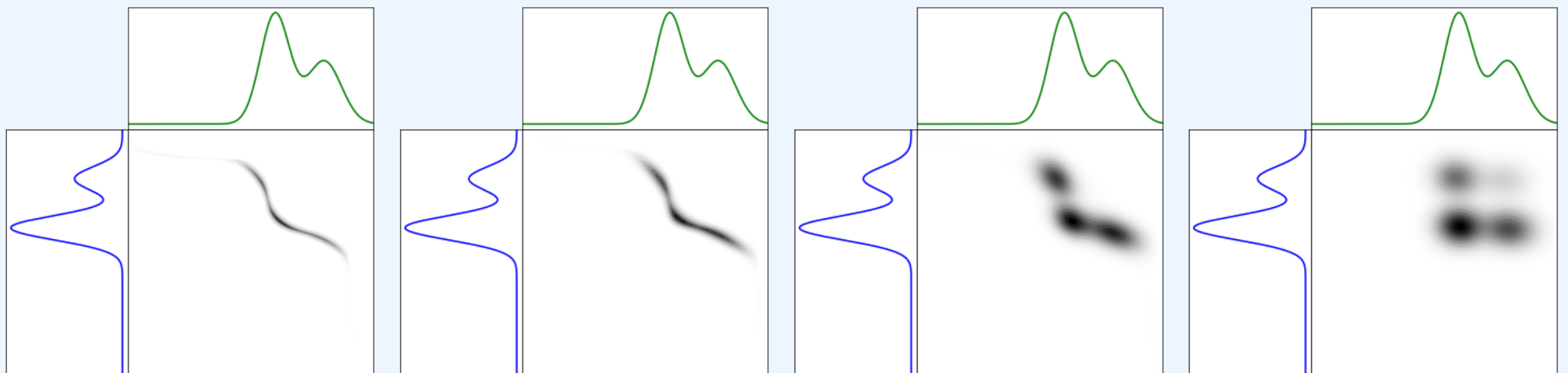
Entropic OT [Cuturi '13]

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

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

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

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



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

$$\varepsilon = 10^{-3}$$

$$\varepsilon = 10^{-2}$$

$$\varepsilon = 10^{-1}$$

Entropic OT - Sinkhorn algorithm

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

Entropic OT - Sinkhorn algorithm

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

Since $\gamma \in \Pi(\mu_0, \mu_1)$, it implies that

$$a \odot K b = \mu_0$$

$$b \odot K^T a = \mu_1$$

Entropic OT - Sinkhorn algorithm

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

Since $\gamma \in \Pi(\mu_0, \mu_1)$, it implies that

$$a \odot K b = \mu_0$$

$$b \odot K^T a = \mu_1$$

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

Entropic OT - Sinkhorn algorithm

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

Since $\gamma \in \Pi(\mu_0, \mu_1)$, it implies that

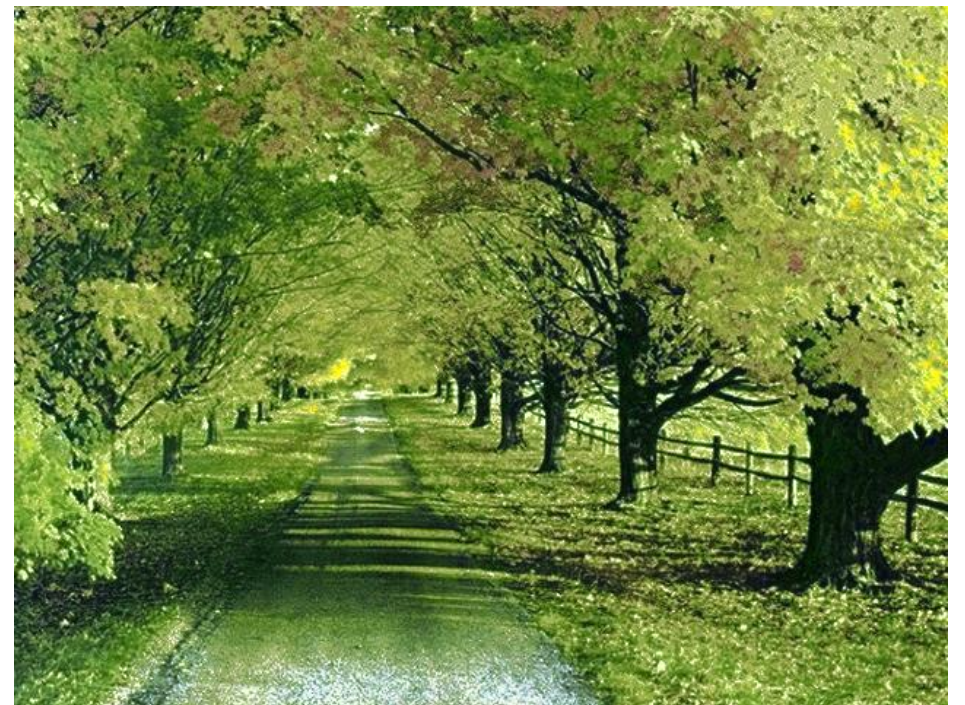
$$a \odot K b = \mu_0$$

$$b \odot K^T a = \mu_1$$

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

- 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 $\varepsilon \rightarrow 0$.

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

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

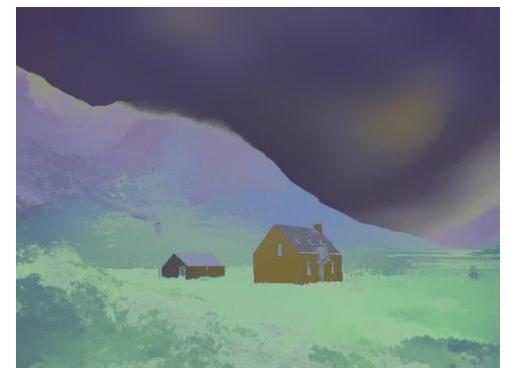
Post-processing



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]



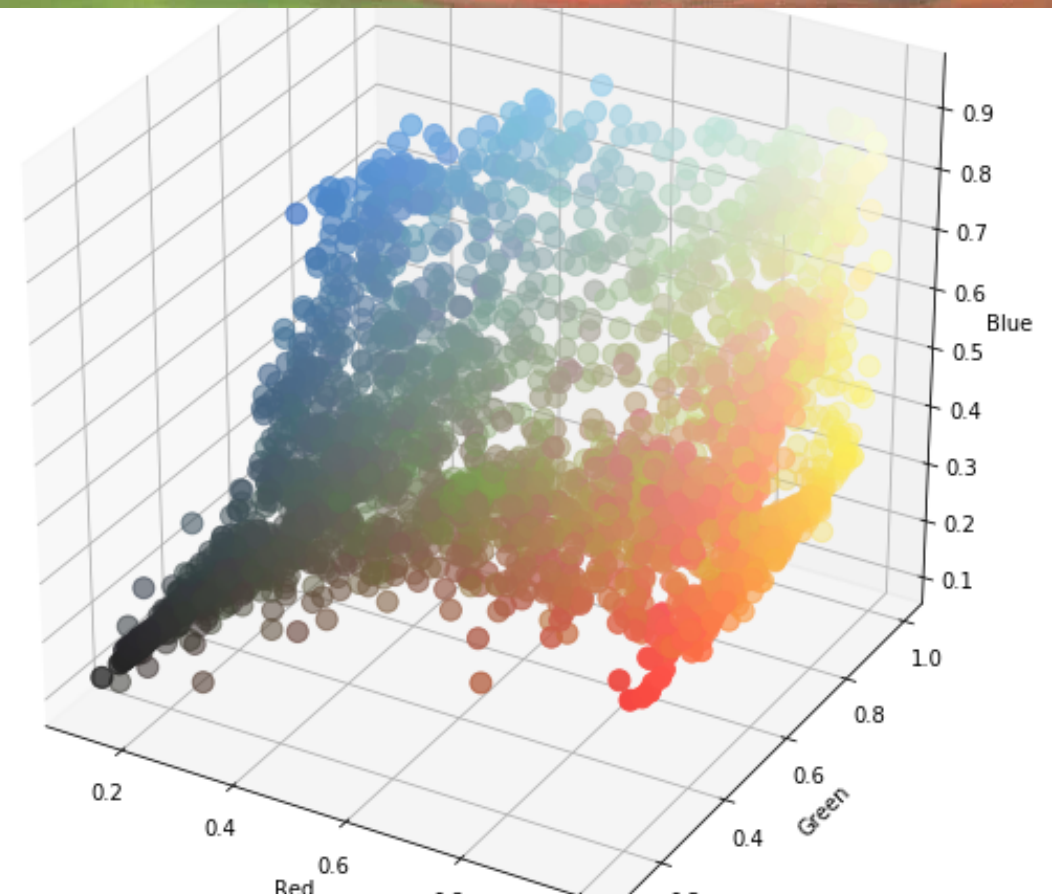
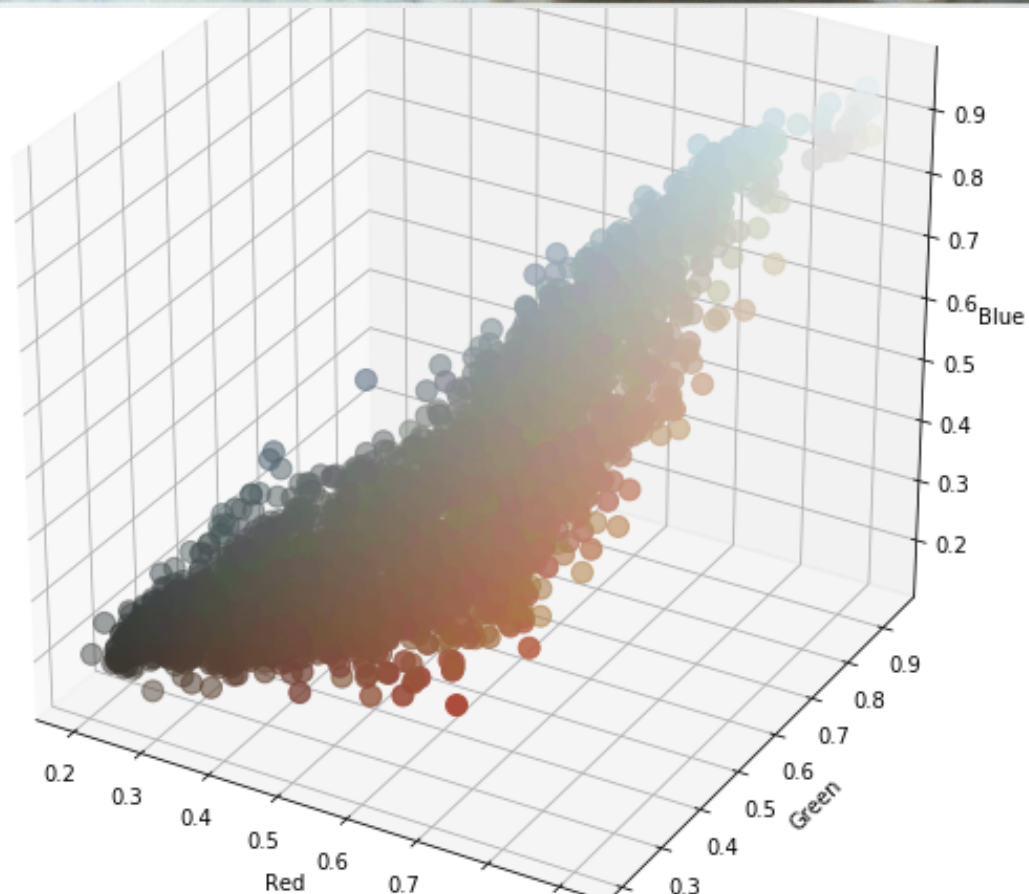




Color transfer



Color transfer









Bonus - Optimal transport between GMMs

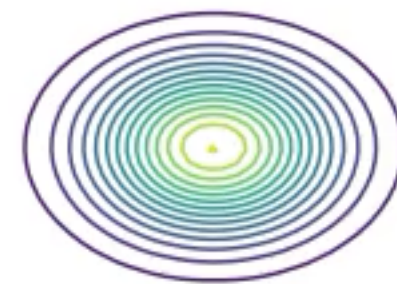
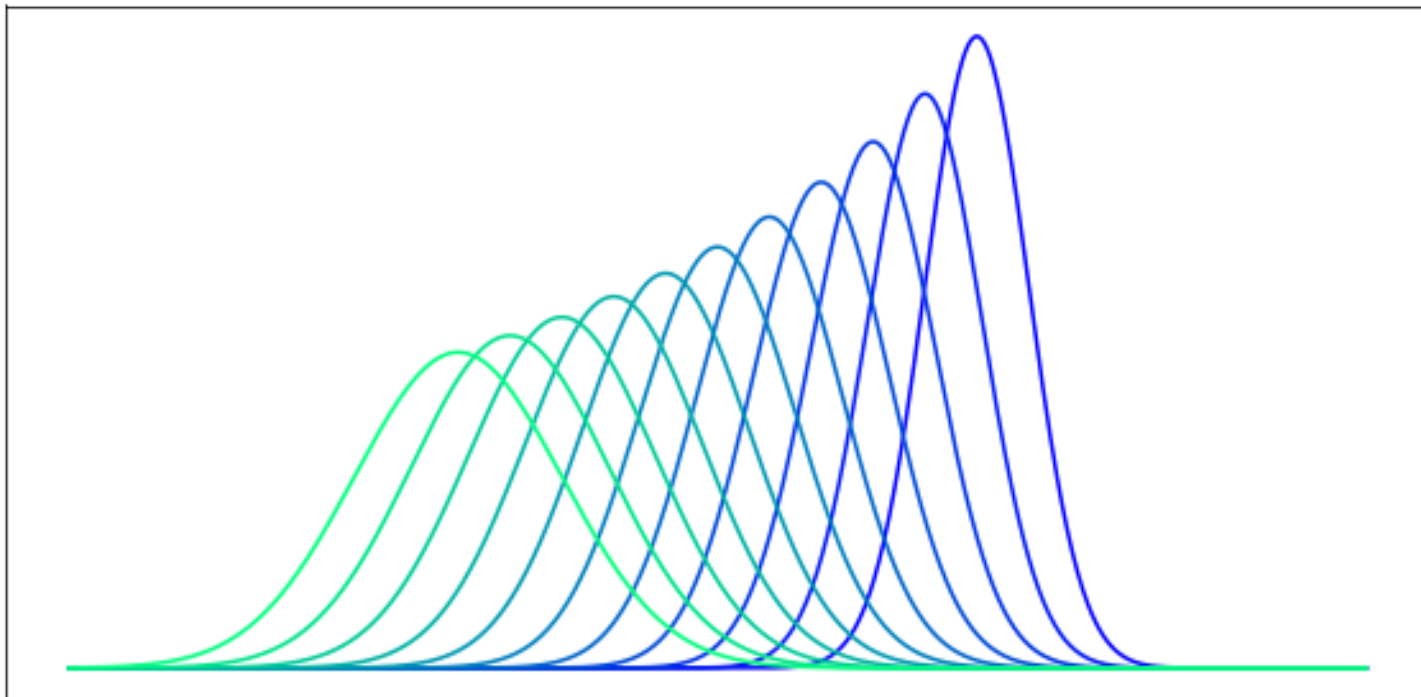
Optimal transport between Gaussians

$\mu_i = \mathcal{N}(m_i, \Sigma_i), i \in \{0, 1\}$ two Gaussian distributions on $\mathbb{R}^d$

$$W_2^2(\mu_0, \mu_1) = \|m_0 - m_1\|^2 + \underbrace{\text{tr} \left(\Sigma_0 + \Sigma_1 - 2 \left(\Sigma_0^{\frac{1}{2}} \Sigma_1 \Sigma_0^{\frac{1}{2}} \right)^{\frac{1}{2}} \right)}_{B^2(\Sigma_0, \Sigma_1)}$$

If Σ_0 non-singular, affine optimal map

$$T(x) = m_1 + \Sigma_0^{-\frac{1}{2}} \left(\Sigma_0^{\frac{1}{2}} \Sigma_1 \Sigma_0^{\frac{1}{2}} \right)^{\frac{1}{2}} \Sigma_0^{-\frac{1}{2}} (x - m_0)$$



Optimal transport between Gaussians

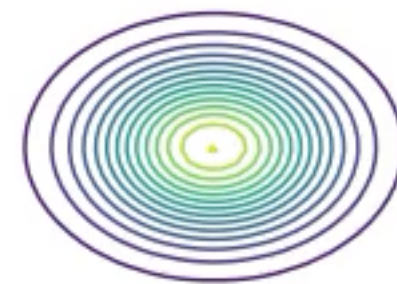
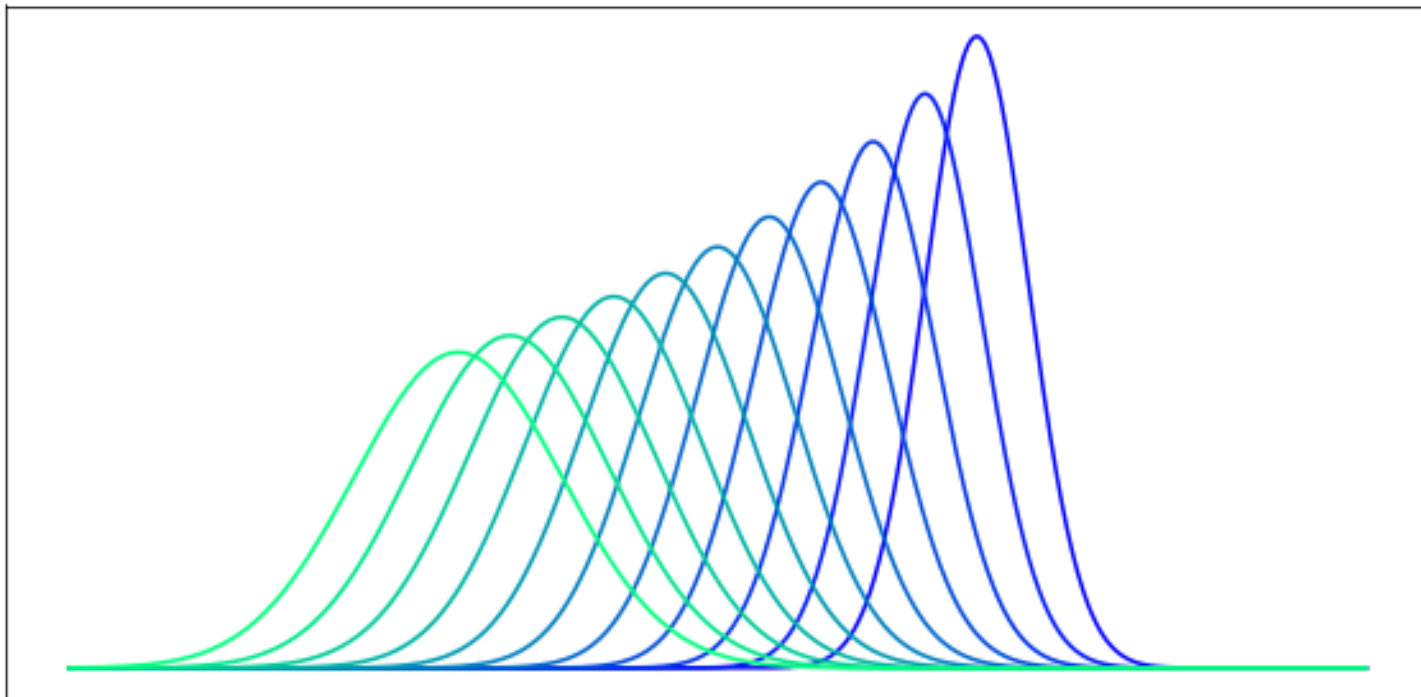
$\mu_i = \mathcal{N}(m_i, \Sigma_i), i \in \{0, 1\}$ two Gaussian distributions on $\mathbb{R}^d$

$$W_2^2(\mu_0, \mu_1) = \|m_0 - m_1\|^2 + \underbrace{\text{tr} \left(\Sigma_0 + \Sigma_1 - 2 \left(\Sigma_0^{\frac{1}{2}} \Sigma_1 \Sigma_0^{\frac{1}{2}} \right)^{\frac{1}{2}} \right)}_{B^2(\Sigma_0, \Sigma_1)}$$

$B^2(\Sigma_0, \Sigma_1)$

If Σ_0 non-singular, affine optimal map

$$T(x) = m_1 + \Sigma_0^{-\frac{1}{2}} \left(\Sigma_0^{\frac{1}{2}} \Sigma_1 \Sigma_0^{\frac{1}{2}} \right)^{\frac{1}{2}} \Sigma_0^{-\frac{1}{2}} (x - m_0)$$



Barycenters between Gaussians

$\mu_i = \mathcal{N}(m_i, \Sigma_i), i \in \{0, \dots, I - 1\}$ Gaussian distributions on $\mathbb{R}^d$

Barycenters between Gaussians

$\mu_i = \mathcal{N}(m_i, \Sigma_i), i \in \{0, \dots, I-1\}$ Gaussian distributions on $\mathbb{R}^d$

Barycenter [Agueh, Carlier 2011]:

$$\operatorname{argmin}_{\mu} \sum_{i=0}^{I-1} \lambda_i W_2^2(\mu_i, \mu) = \mathcal{N}(m^*, \Sigma^*)$$

$$m^* = \sum \lambda_i m_i \quad \Sigma^* = \min_{\Sigma} \sum_i \lambda_i B(\Sigma, \Sigma_i)^2$$

Barycenters between Gaussians

$\mu_i = \mathcal{N}(m_i, \Sigma_i), i \in \{0, \dots, I-1\}$ Gaussian distributions on $\mathbb{R}^d$

Barycenter [Agueh, Carlier 2011]:

$$\operatorname{argmin}_{\mu} \sum_{i=0}^{I-1} \lambda_i W_2^2(\mu_i, \mu) = \mathcal{N}(m^*, \Sigma^*)$$

$$m^* = \sum \lambda_i m_i \quad \Sigma^* = \min_{\Sigma} \sum_i \lambda_i B(\Sigma, \Sigma_i)^2$$

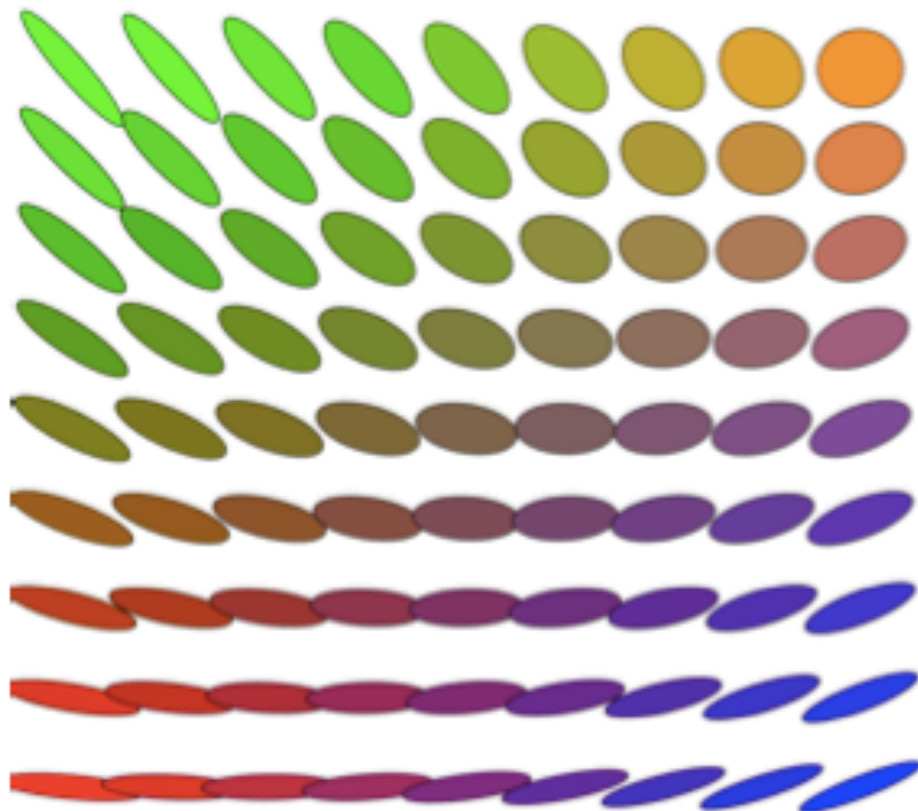


illustration: Cuturi, Peyré, OT book

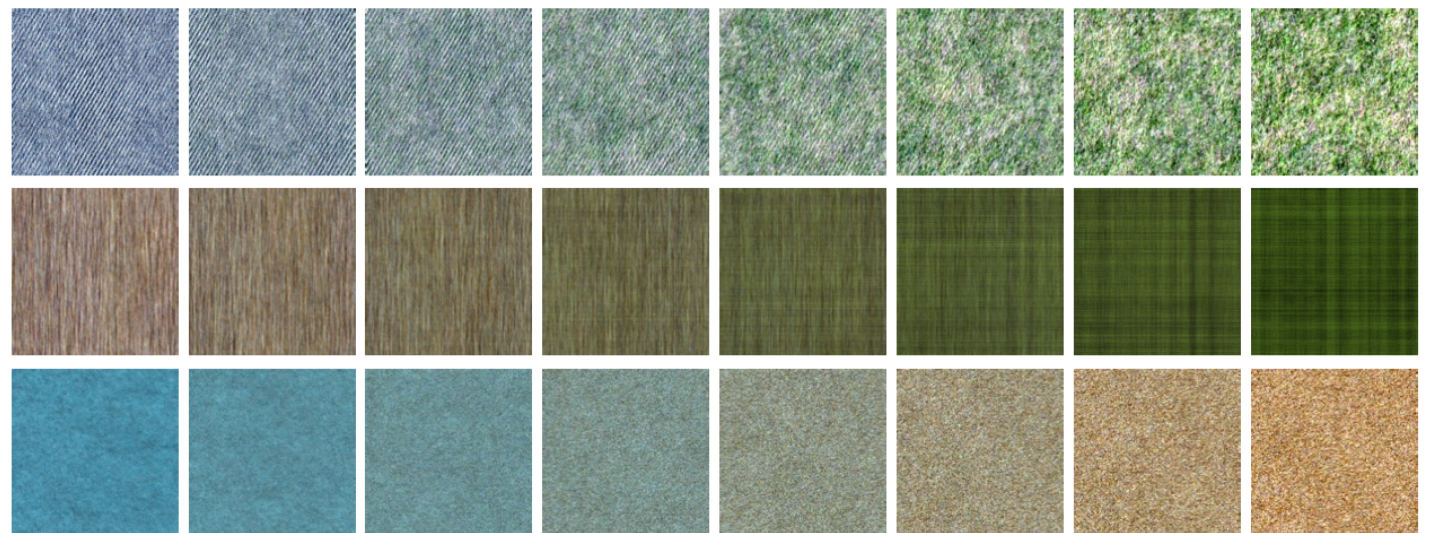
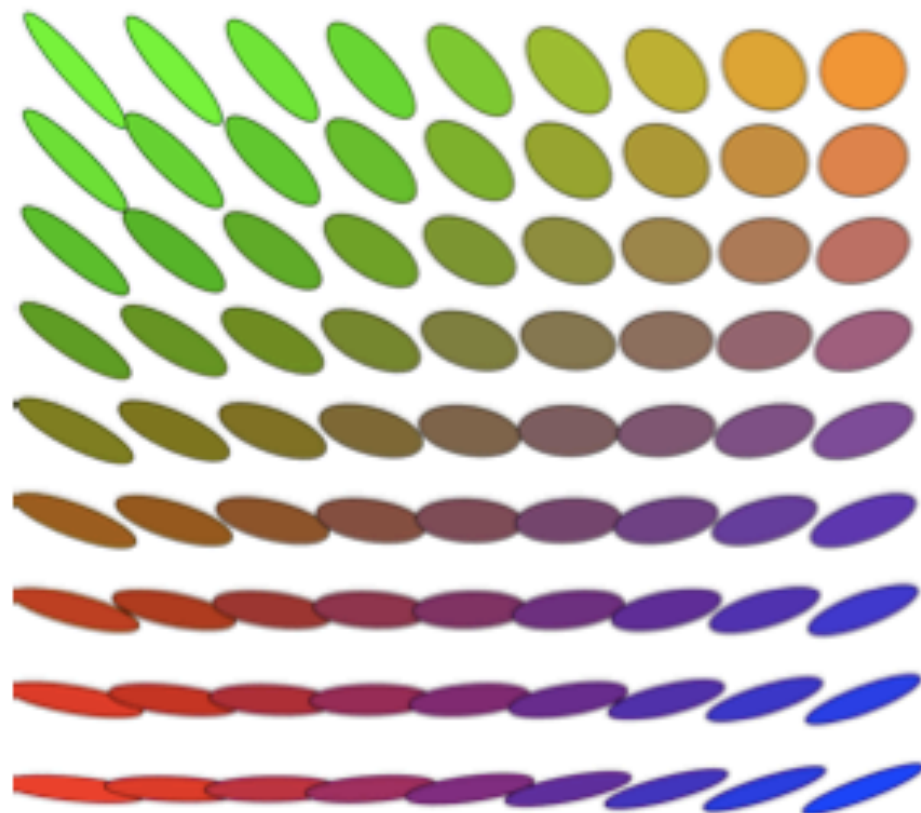
Barycenters between Gaussians

$\mu_i = \mathcal{N}(m_i, \Sigma_i), i \in \{0, \dots, I - 1\}$ Gaussian distributions on $\mathbb{R}^d$

Barycenter [Agueh, Carlier 2011]:

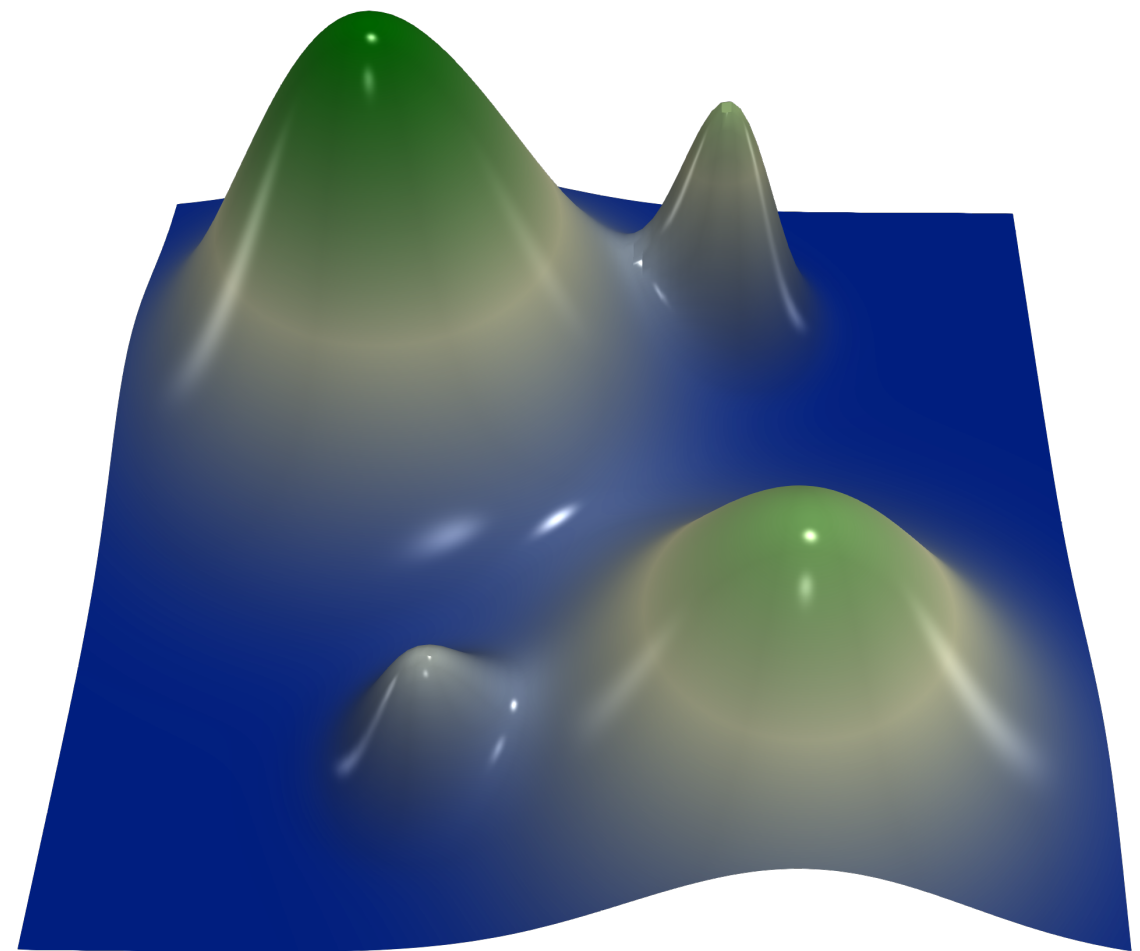
$$\operatorname{argmin}_{\mu} \sum_{i=0}^{I-1} \lambda_i W_2^2(\mu_i, \mu) = \mathcal{N}(m^*, \Sigma^*)$$

$$m^* = \sum \lambda_i m_i \quad \Sigma^* = \min_{\Sigma} \sum_i \lambda_i B(\Sigma, \Sigma_i)^2$$



Texture mixing [Xia et al, 2014]

What about GMMs ?



Optimal transport between GMM

OT plans between GMM: usually not GMM themselves.
Same remark for barycenters.

- $\mu_0 = \mathcal{N}(0, 1)$
- $\mu_1 = \frac{1}{2}(\delta_{-1} + \delta_1)$

- density of μ_t :

$$f_t(x) = \frac{1}{1-t} \left(g\left(\frac{x+t}{1-t}\right) \mathbf{1}_{x < -t} + g\left(\frac{x-t}{1-t}\right) \mathbf{1}_{x > t} \right)$$

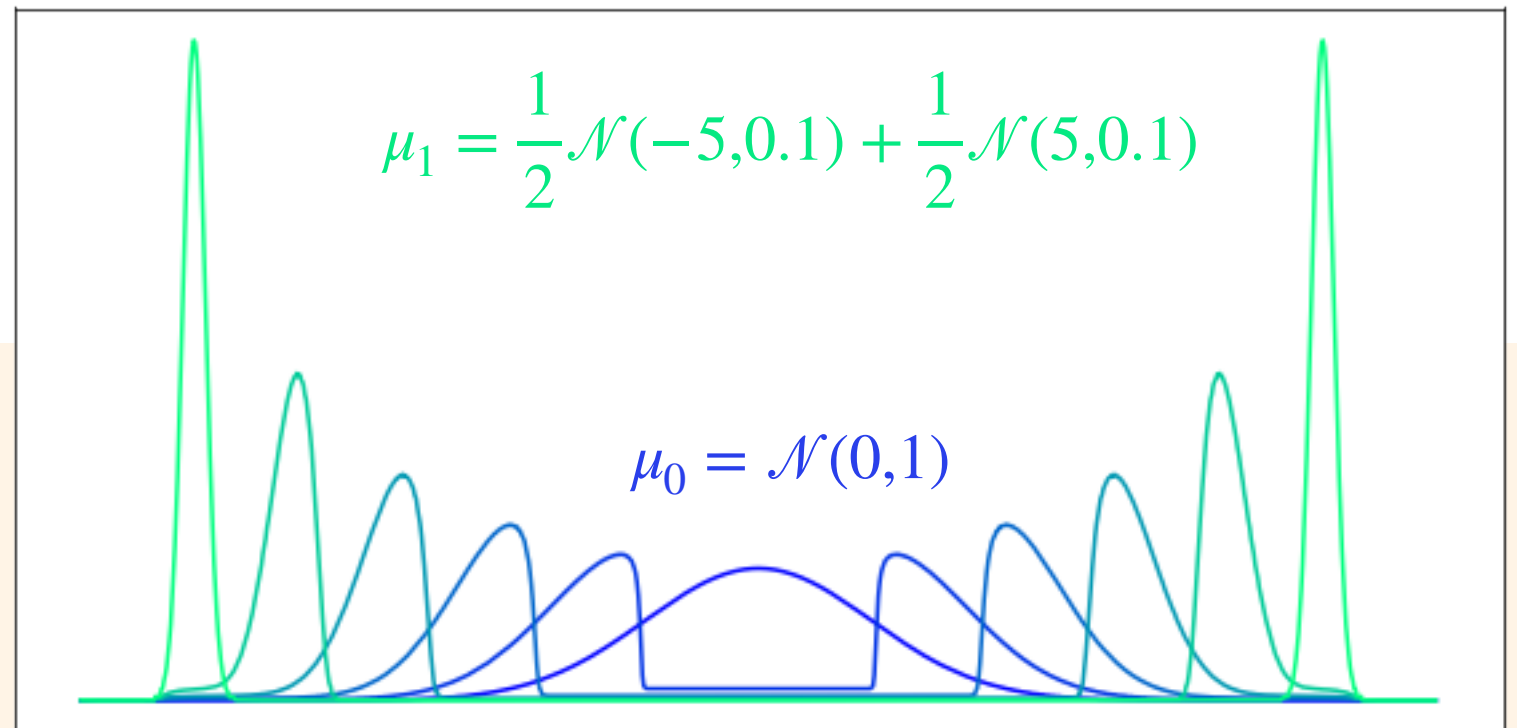
Optimal transport between GMM

OT plans between GMM: usually not GMM themselves.
Same remark for barycenters.

- $\mu_0 = \mathcal{N}(0, 1)$
- $\mu_1 = \frac{1}{2}(\delta_{-1} + \delta_1)$

- density of μ_t :

$$f_t(x) = \frac{1}{1-t} \left(g \left(\frac{x+t}{1-t} \right) \mathbf{1}_{x < -t} + g \left(\frac{x-t}{1-t} \right) \mathbf{1}_{x > t} \right)$$



**Restricting the set of
couplings to GMMs**

Definition of MW2

$\mu_0 = \sum_{k=1}^{K_0} \pi_0^k \mu_0^k$ and $\mu_1 = \sum_{k=1}^{K_1} \pi_1^k \mu_1^k$ two Gaussian mixtures



μ_1^0



μ_1^1



μ_1^2



μ_0^1

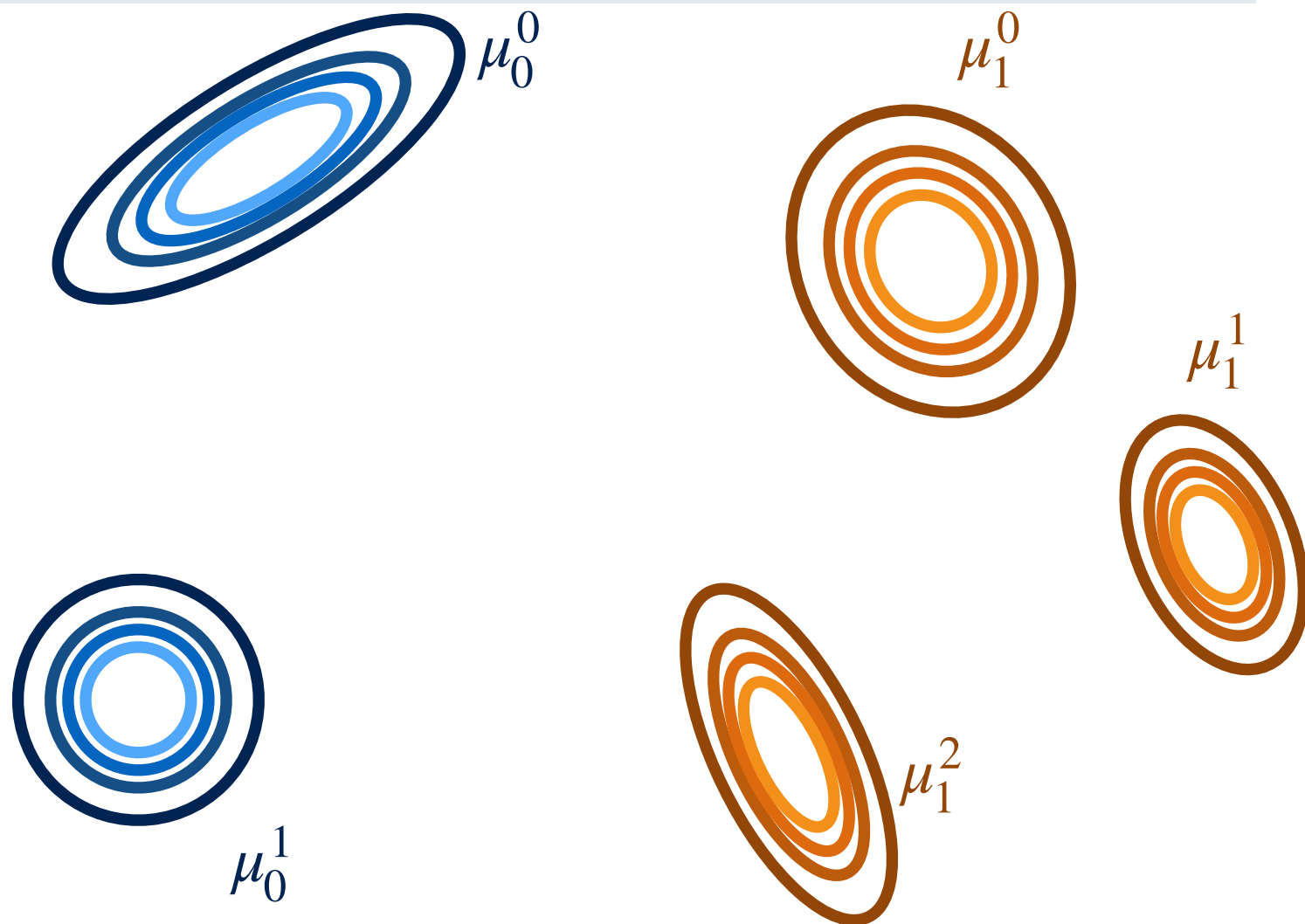


Definition of MW2

$\mu_0 = \sum_{k=1}^{K_0} \pi_0^k \mu_0^k$ and $\mu_1 = \sum_{k=1}^{K_1} \pi_1^k \mu_1^k$ two Gaussian mixtures

Definition:

$$MW_2^2(\mu_0, \mu_1) := \inf_{\gamma \in \Pi(\mu_0, \mu_1) \cap GMM_{2d}(\infty)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \|y_0 - y_1\|^2 d\gamma(y_0, y_1). \quad (0)$$



Definition of MW2

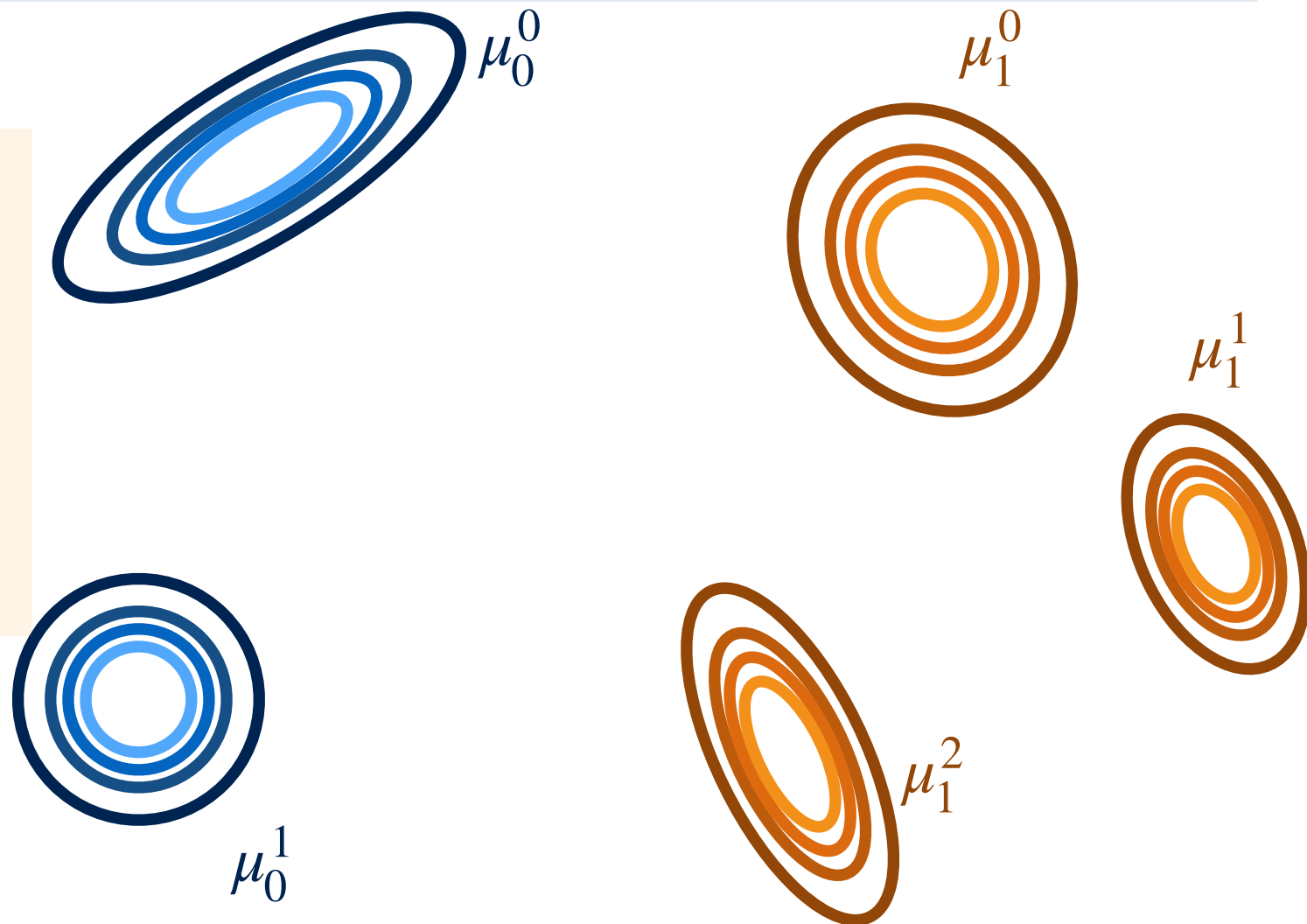
$\mu_0 = \sum_{k=1}^{K_0} \pi_0^k \mu_0^k$ and $\mu_1 = \sum_{k=1}^{K_1} \pi_1^k \mu_1^k$ two Gaussian mixtures

Definition:

$$MW_2^2(\mu_0, \mu_1) := \inf_{\gamma \in \Pi(\mu_0, \mu_1) \cap GMM_{2d}(\infty)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \|y_0 - y_1\|^2 d\gamma(y_0, y_1). \quad (0)$$

Proposition [D.D.,2019]:

$$MW_2^2(\mu_0, \mu_1) = \min_{w \in \Pi(\pi_0, \pi_1)} \sum_{k,l} w_{kl} W_2^2(\mu_0^k, \mu_1^l).$$



Definition of MW2

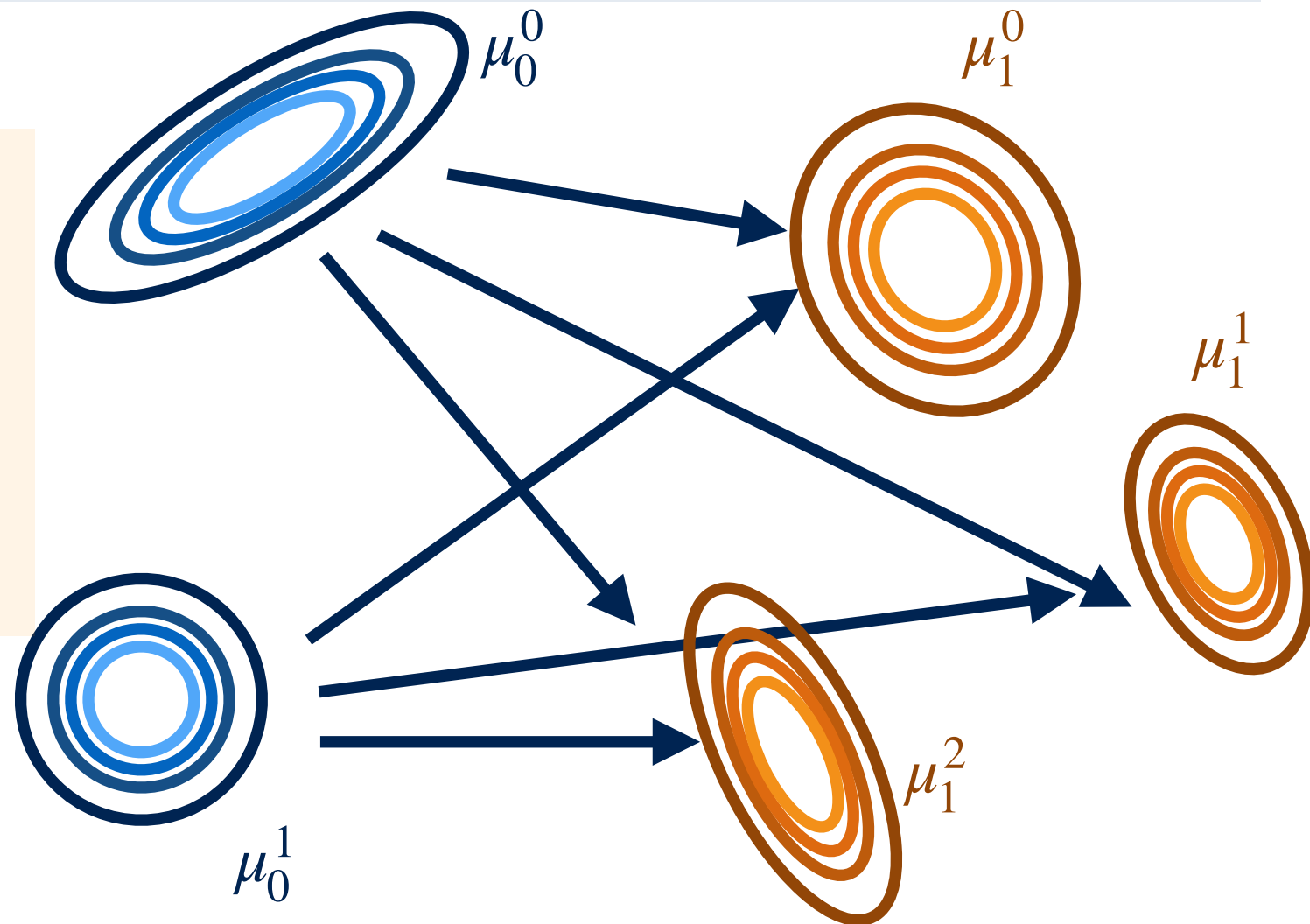
$\mu_0 = \sum_{k=1}^{K_0} \pi_0^k \mu_0^k$ and $\mu_1 = \sum_{k=1}^{K_1} \pi_1^k \mu_1^k$ two Gaussian mixtures

Definition:

$$MW_2^2(\mu_0, \mu_1) := \inf_{\gamma \in \Pi(\mu_0, \mu_1) \cap GMM_{2d}(\infty)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \|y_0 - y_1\|^2 d\gamma(y_0, y_1). \quad (0)$$

Proposition [D.D.,2019]:

$$MW_2^2(\mu_0, \mu_1) = \min_{w \in \Pi(\pi_0, \pi_1)} \sum_{k,l} w_{kl} W_2^2(\mu_0^k, \mu_1^l).$$



Definition of MW2

$\mu_0 = \sum_{k=1}^{K_0} \pi_0^k \mu_0^k$ and $\mu_1 = \sum_{k=1}^{K_1} \pi_1^k \mu_1^k$ two Gaussian mixtures

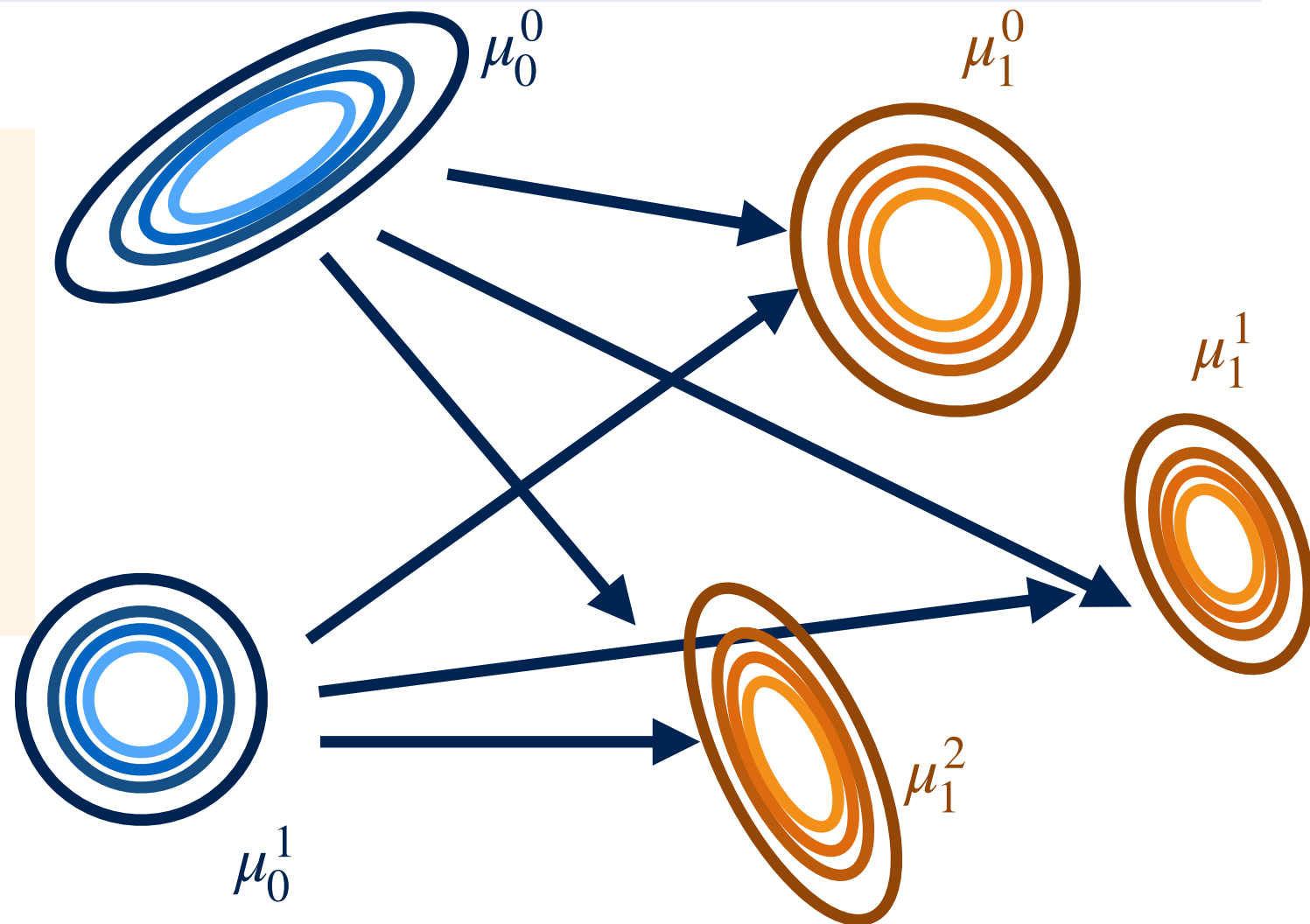
Definition:

$$MW_2^2(\mu_0, \mu_1) := \inf_{\gamma \in \Pi(\mu_0, \mu_1) \cap GMM_{2d}(\infty)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \|y_0 - y_1\|^2 d\gamma(y_0, y_1). \quad (0)$$

Proposition [D.D.,2019]:

$$MW_2^2(\mu_0, \mu_1) = \min_{w \in \Pi(\pi_0, \pi_1)} \sum_{k,l} w_{kl} W_2^2(\mu_0^k, \mu_1^l).$$

$K_0 \times K_1$ OT problem



Definition of MW2

$\mu_0 = \sum_{k=1}^{K_0} \pi_0^k \mu_0^k$ and $\mu_1 = \sum_{k=1}^{K_1} \pi_1^k \mu_1^k$ two Gaussian mixtures

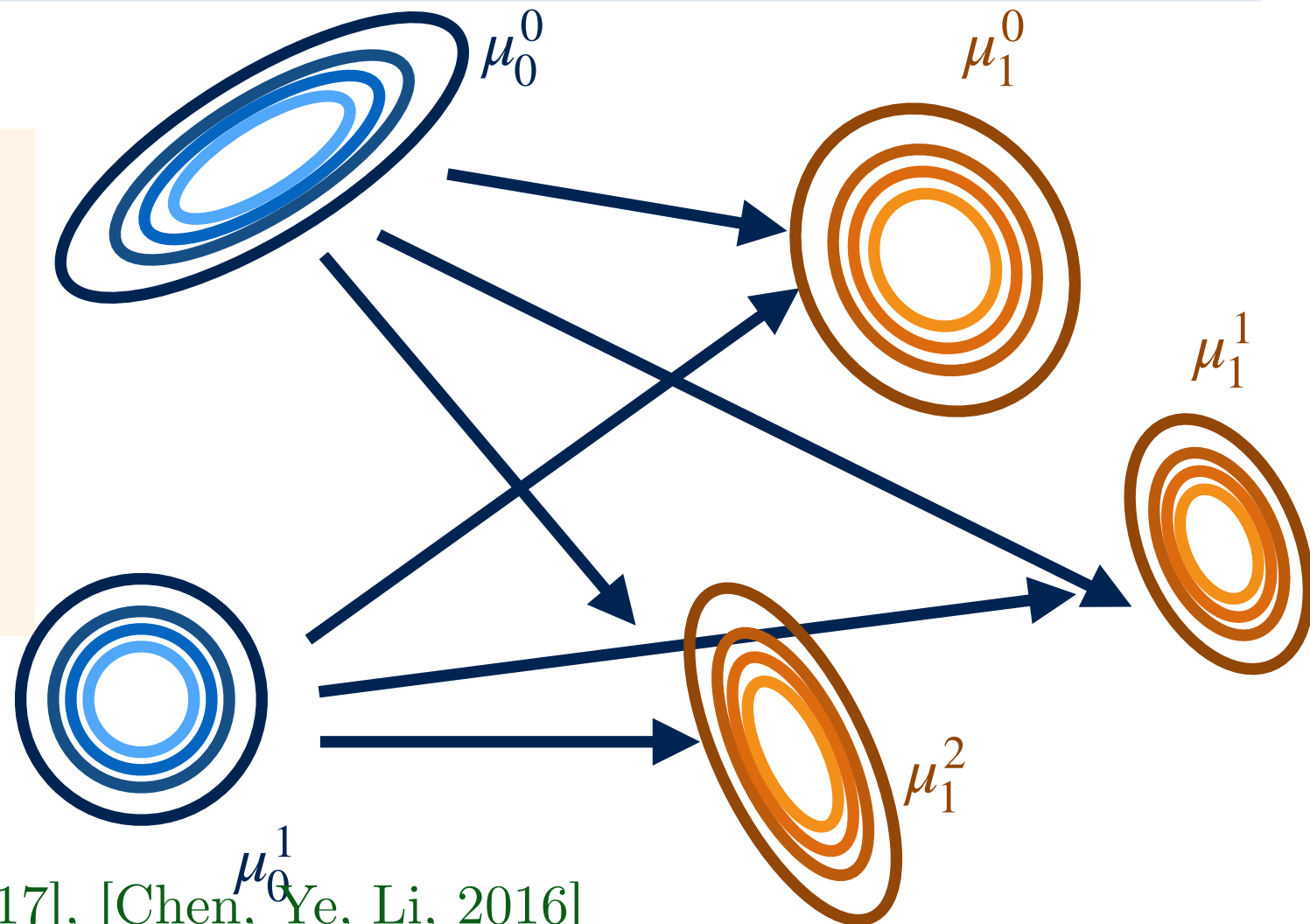
Definition:

$$MW_2^2(\mu_0, \mu_1) := \inf_{\gamma \in \Pi(\mu_0, \mu_1) \cap GMM_{2d}(\infty)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \|y_0 - y_1\|^2 d\gamma(y_0, y_1). \quad (0)$$

Proposition [D.D.,2019]:

$$MW_2^2(\mu_0, \mu_1) = \min_{w \in \Pi(\pi_0, \pi_1)} \sum_{k,l} w_{kl} W_2^2(\mu_0^k, \mu_1^l).$$

$K_0 \times K_1$ OT problem



See also [Chen, Georgiu, Tannenbaum, 2017], [Chen, Ye, Li, 2016]

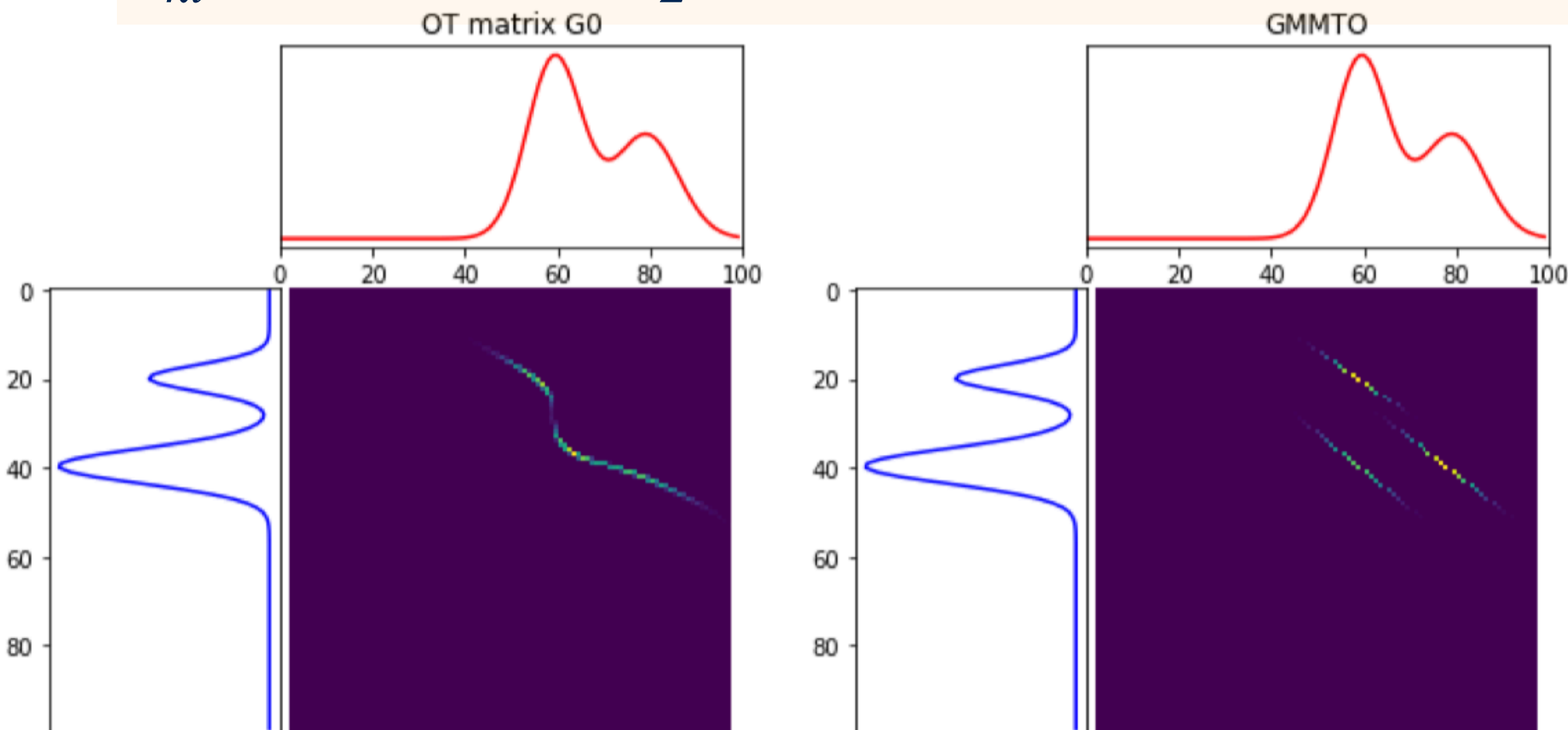
Optimal plan

Proposition: an optimal plan γ^* for MW_2 can be written

$$\gamma^*(x, y) = \sum_{k,l} w_{k,l}^* g_{m_0^k, \Sigma_0^k}(x) \delta_{y=T_{k,l}(x)}, \text{ where}$$

w^* = solution of the discrete OT problem of size $K_0 \times K_1$

T_{kl} = optimal W_2 - map between the Gaussians μ_0^k and μ_1^l



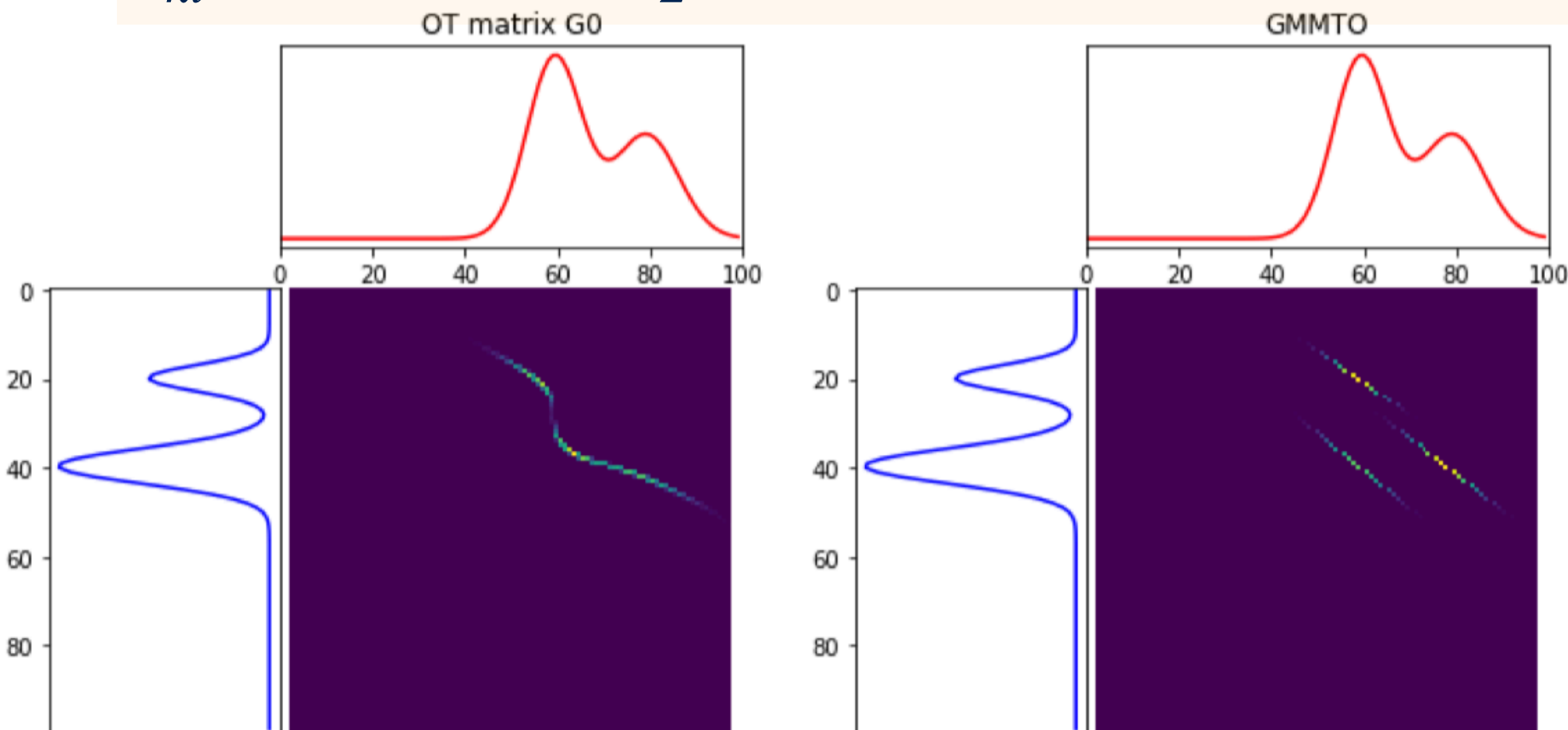
Optimal plan

Proposition: an optimal plan γ^* for MW_2 can be written

$$\gamma^*(x, y) = \sum_{k,l} w_{k,l}^* g_{m_0^k, \Sigma_0^k}(x) \delta_{y=T_{k,l}(x)}, \text{ where}$$

w^* = solution of the discrete OT problem of size $K_0 \times K_1$

T_{kl} = optimal W_2 - map between the Gaussians μ_0^k and μ_1^l



At least one solution has less than $K_0 + K_1 - 1$ components!

Metric properties of MW_2

Proposition: MW_2 defines a metric on $GMM_d(\infty)$. This space equipped with the distance MW_2 is a geodesic space.

Metric properties of MW_2

Proposition: MW_2 defines a metric on $GMM_d(\infty)$. This space equipped with the distance MW_2 is a geodesic space.

Corollary: Barycenters between $\mu_0 = \sum_k \pi_0^k \mu_0^k$ and $\mu_1 = \sum_l \pi_1^l \mu_1^l$ belong to $GMM_d(\infty)$ and can be written explicitly as

$$\forall t \in [0, 1], \quad \mu_t = \sum_{k,l} w_{k,l}^* \mu_t^{k,l},$$

where $\mu_t^{k,l}$ is the displacement interpolation between μ_0^k and μ_1^l .

Metric properties of MW_2

Proposition: MW_2 defines a metric on $GMM_d(\infty)$. This space equipped with the distance MW_2 is a geodesic space.

Corollary: Barycenters between $\mu_0 = \sum_k \pi_0^k \mu_0^k$ and $\mu_1 = \sum_l \pi_1^l \mu_1^l$ belong to $GMM_d(\infty)$ and can be written explicitly as

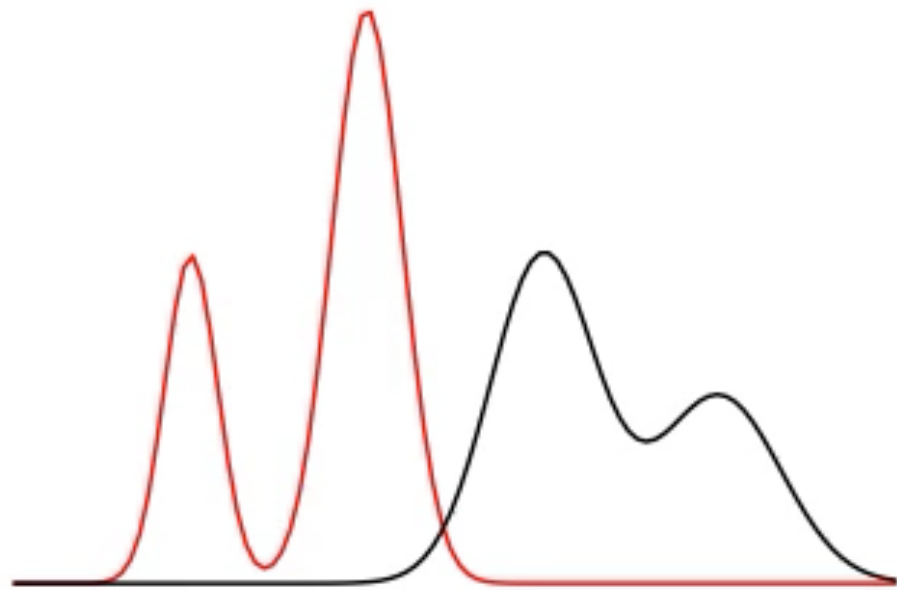
$$\forall t \in [0, 1], \quad \mu_t = \sum_{k,l} w_{k,l}^* \mu_t^{k,l},$$

where $\mu_t^{k,l}$ is the displacement interpolation between μ_0^k and μ_1^l .

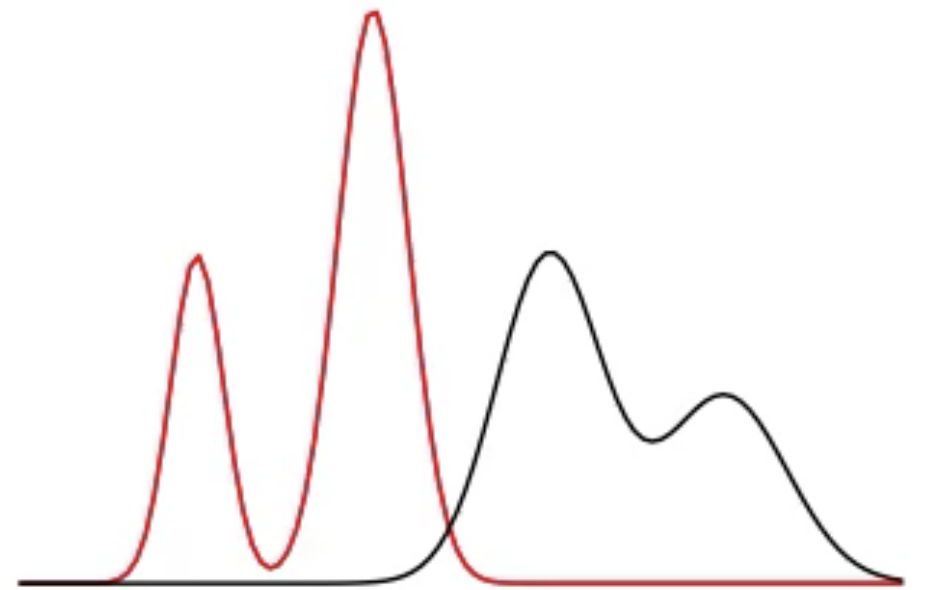
Barycenters can be chosen with less than $K_0 + K_1 - 1$ components.

Displacement interpolation

W_2

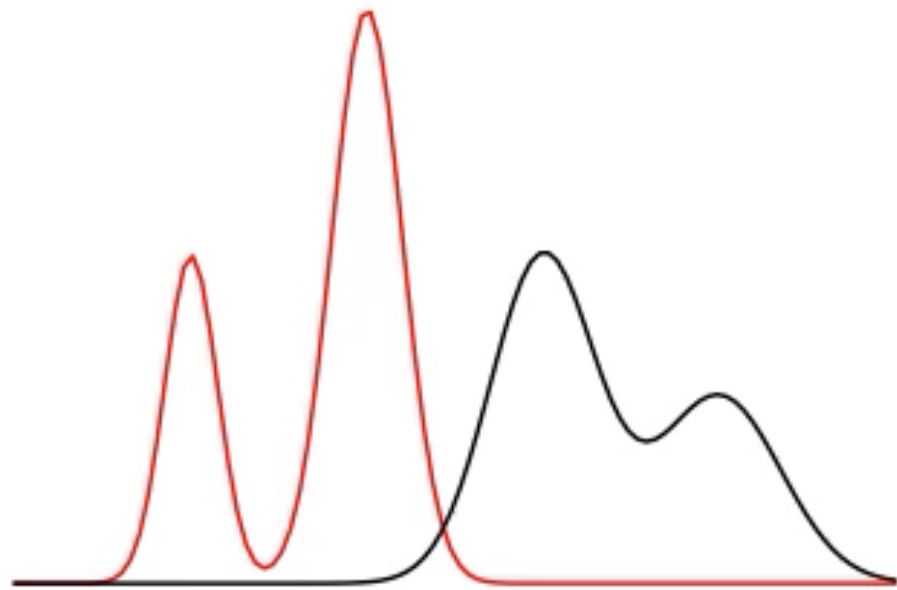


MW_2

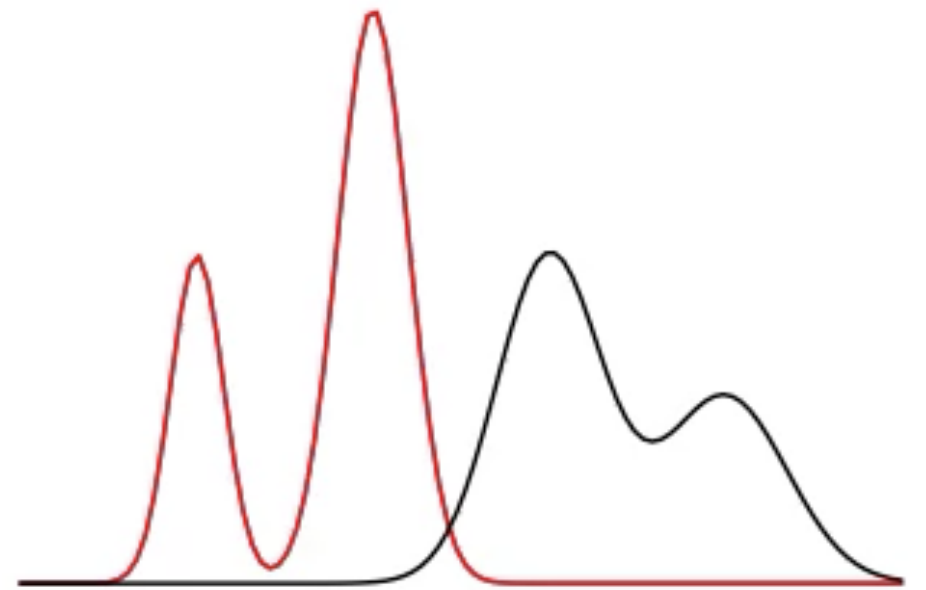


Displacement interpolation

W_2

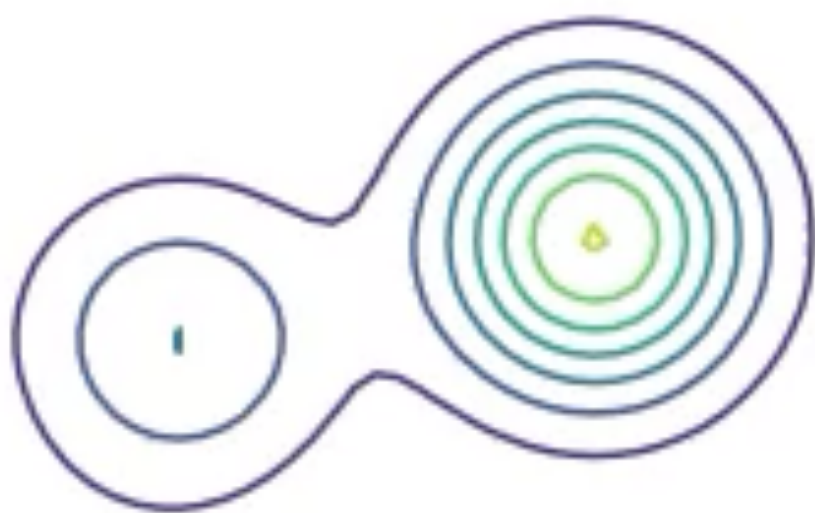


MW_2



Displacement interpolation

W_2

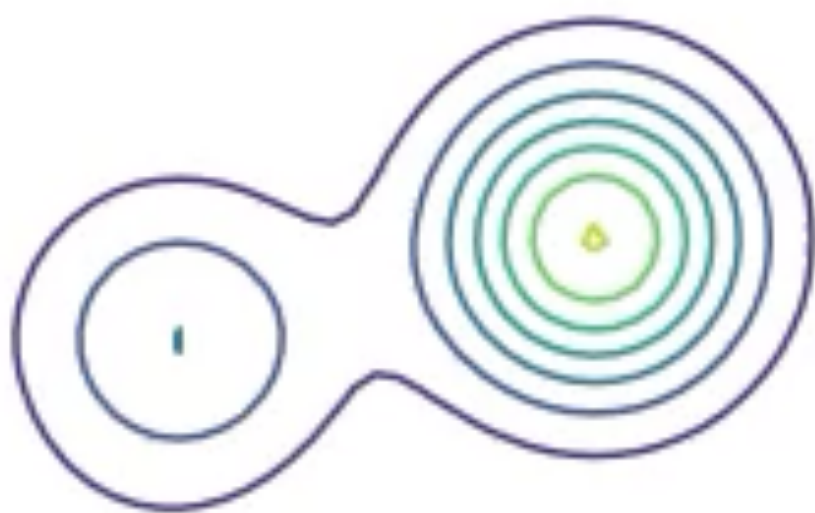


MW_2



Displacement interpolation

W_2



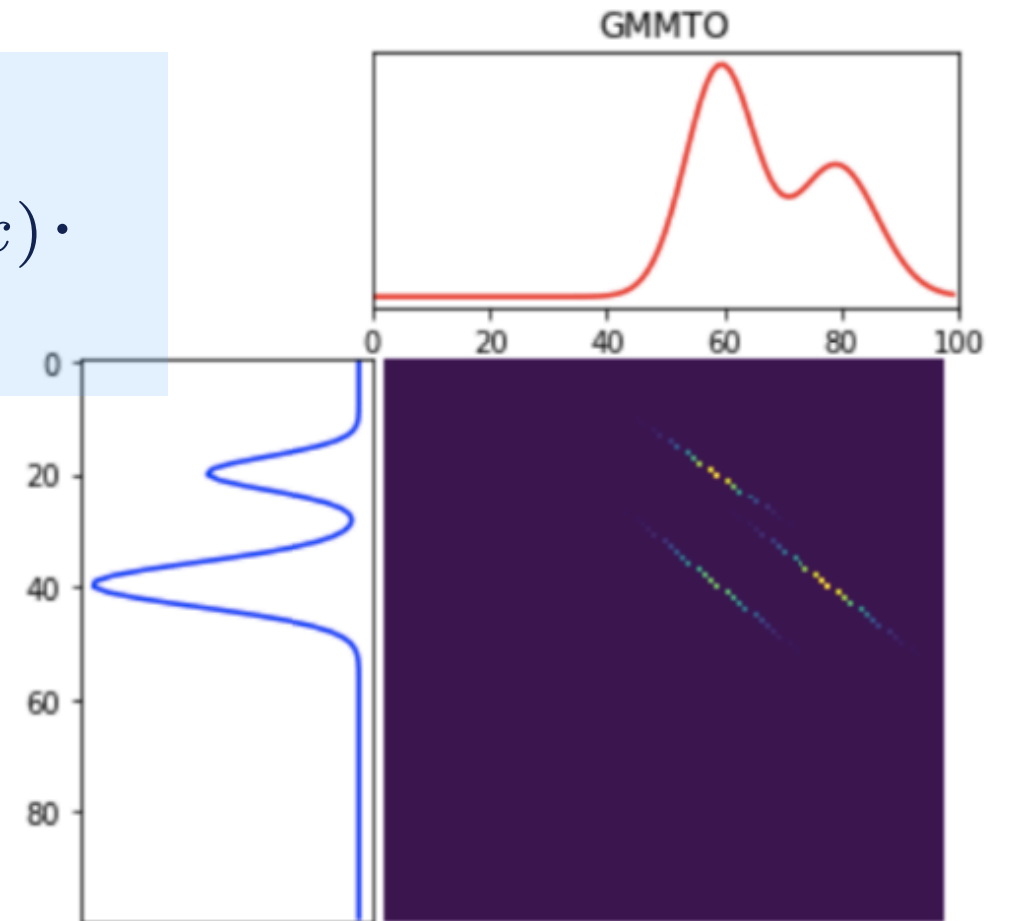
MW_2



Using MW_2 on real data

From a transport plan to a map

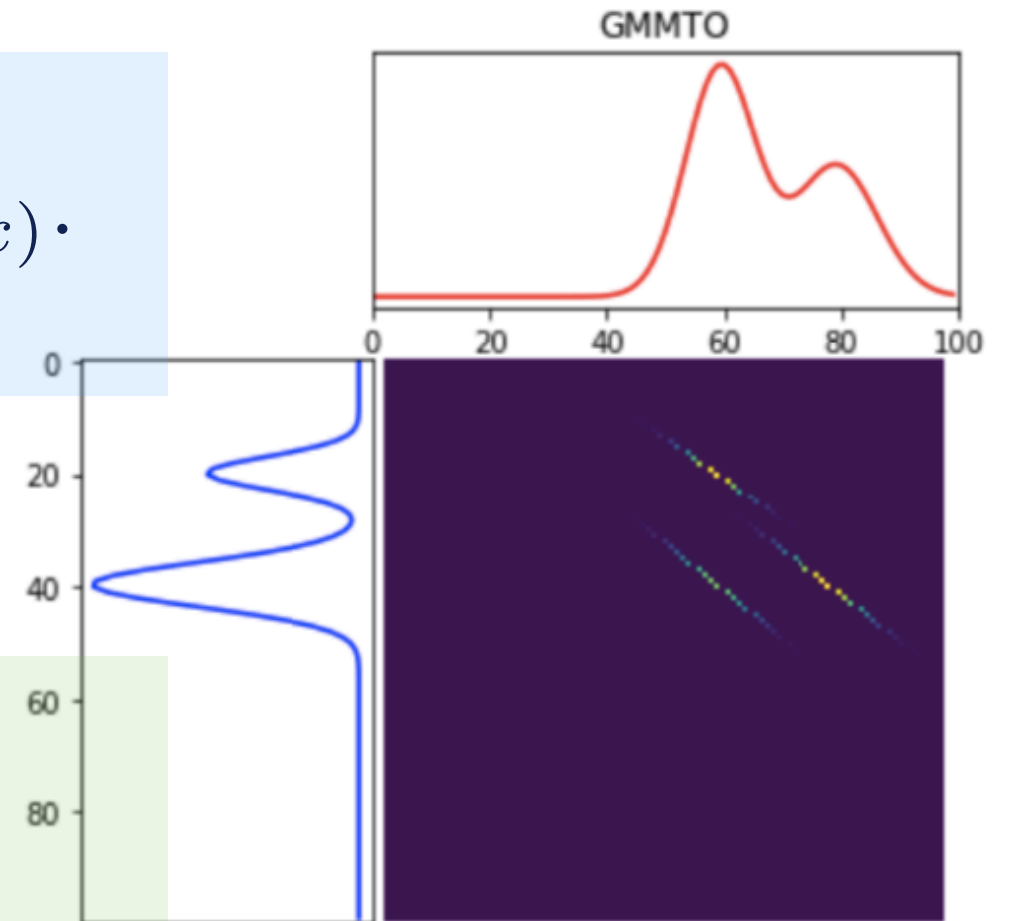
$$\gamma(x, y) = \sum_{k,l} w_{k,l}^* g_{m_0^k, \Sigma_0^k}(x) \delta_{y=T_{k,l}(x)}.$$



From a transport plan to a map

$$\gamma(x, y) = \sum_{k,l} w_{k,l}^* g_{m_0^k, \Sigma_0^k}(x) \delta_{y=T_{k,l}(x)}.$$

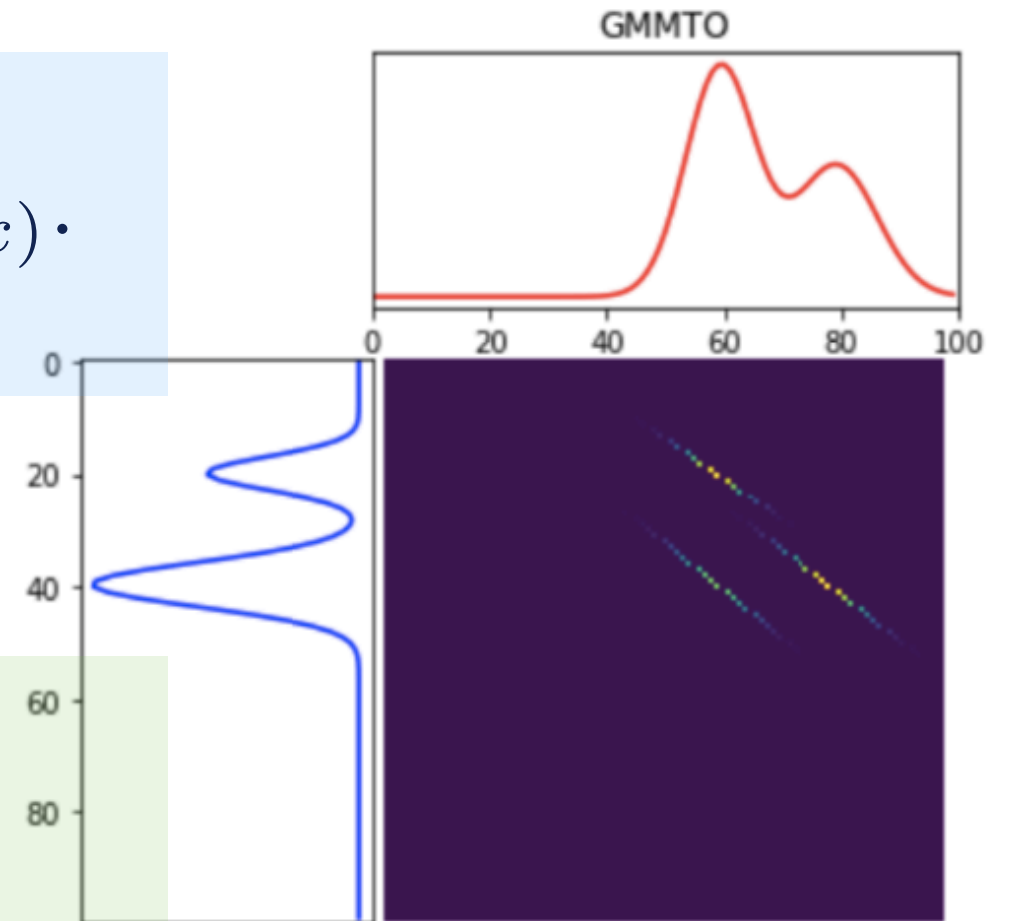
$$T_{mean}(x) = \frac{\sum_{k,l} w_{k,l}^* g_{m_0^k, \Sigma_0^k}(x) T_{k,l}(x)}{\sum_k \pi_0^k g_{m_0^k, \Sigma_0^k}(x)},$$



From a transport plan to a map

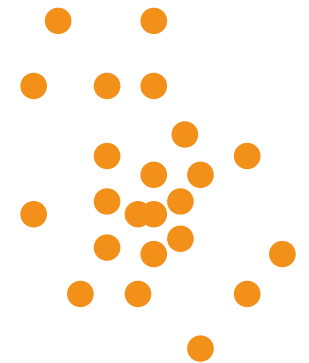
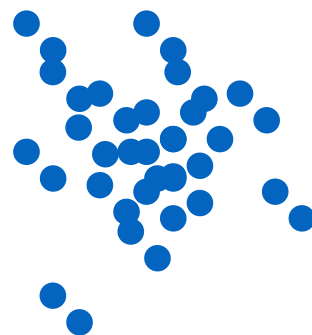
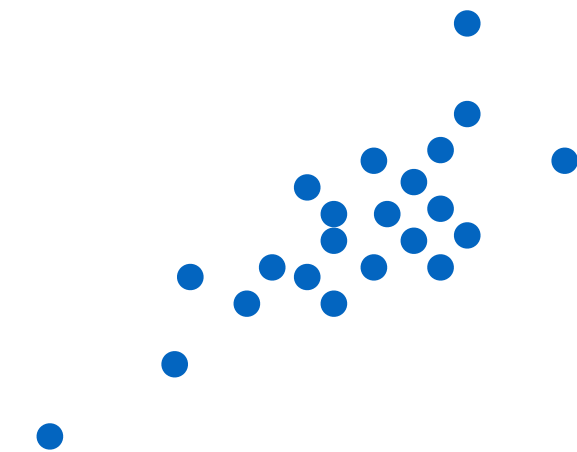
$$\gamma(x, y) = \sum_{k,l} w_{k,l}^* g_{m_0^k, \Sigma_0^k}(x) \delta_{y=T_{k,l}(x)}.$$

$$T_{mean}(x) = \frac{\sum_{k,l} w_{k,l}^* g_{m_0^k, \Sigma_0^k}(x) T_{k,l}(x)}{\sum_k \pi_0^k g_{m_0^k, \Sigma_0^k}(x)},$$

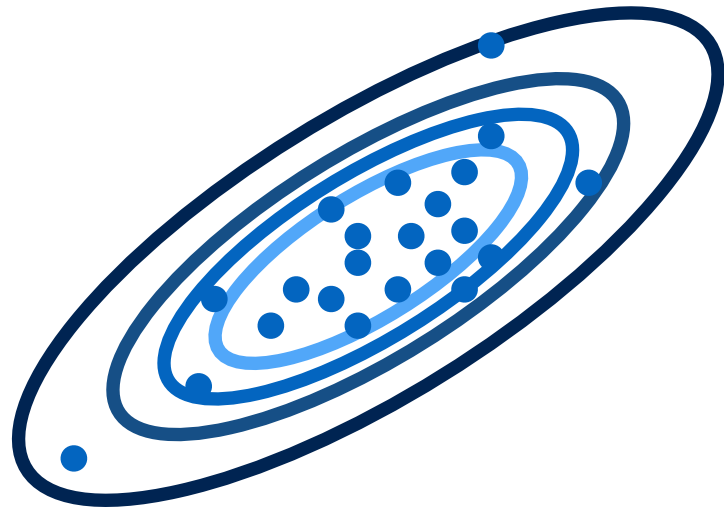


$$T_{rand}(x) = T_{k,l}(x) \quad \text{with probability } p_{k,l}(x) = \frac{w_{k,l}^* g_{m_0^k, \Sigma_0^k}(x)}{\sum_j \pi_0^j g_{m_0^j, \Sigma_0^j}(x)}.$$

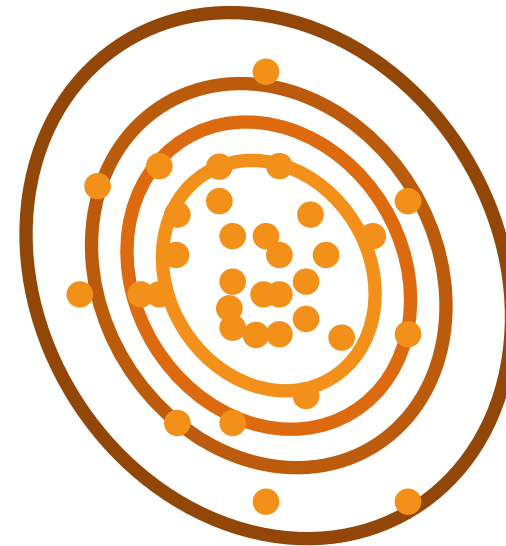
Discrete data : EM + MW_2



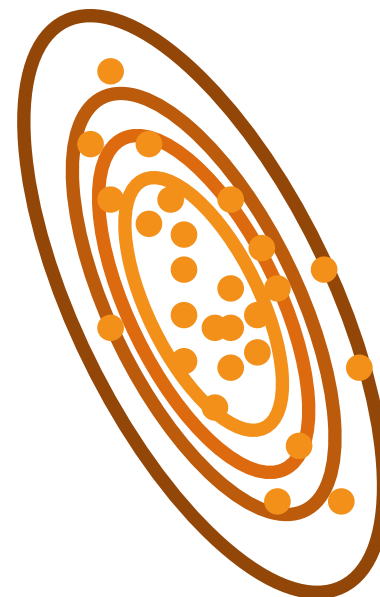
Discrete data : EM + MW_2



μ_0



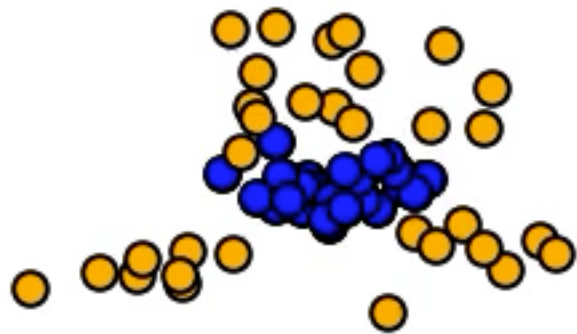
μ_1



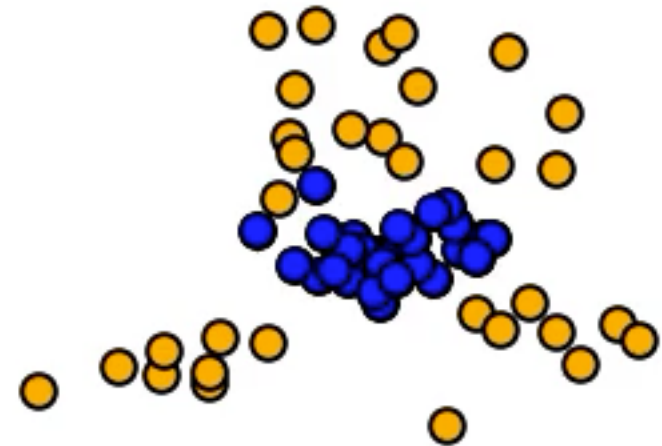
Discrete data : EM + MW_2

T_{rand}

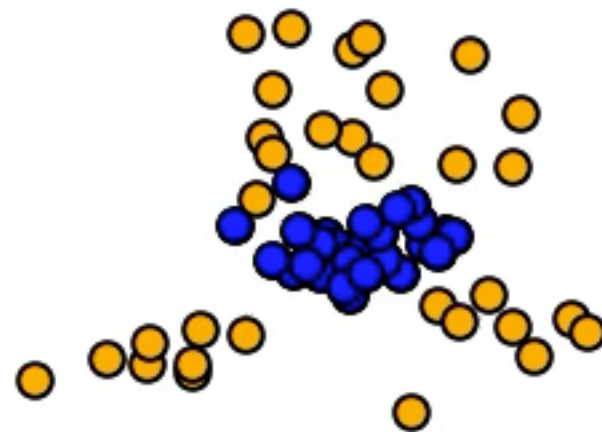
$K = 1$



$K = 10$



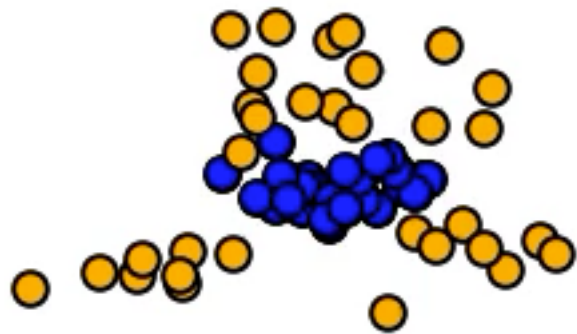
$K = N$



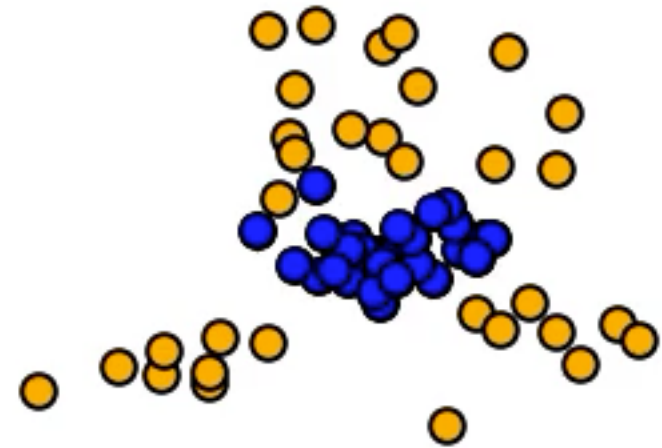
Discrete data : EM + MW_2

T_{rand}

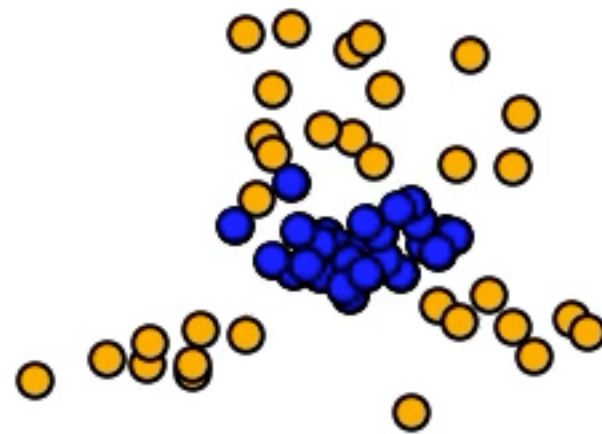
$K = 1$



$K = 10$



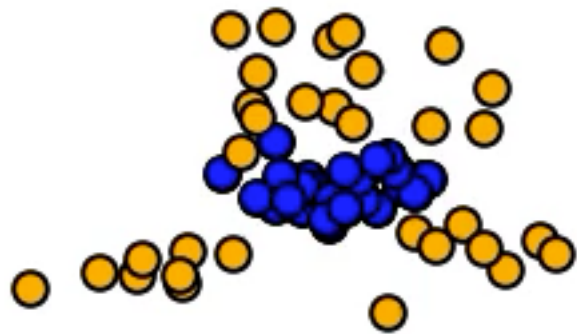
$K = N$



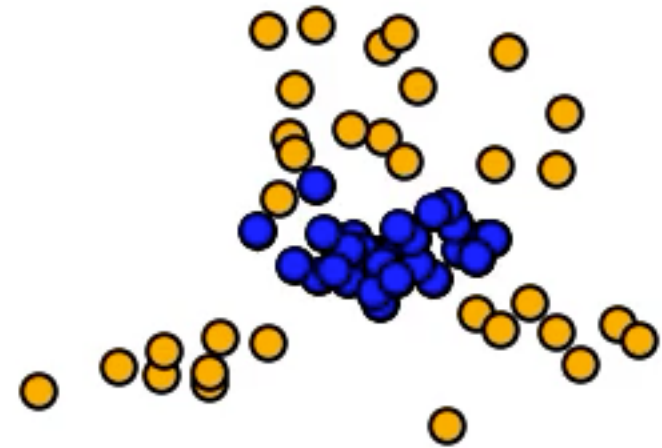
Discrete data : EM + MW_2

T_{rand}

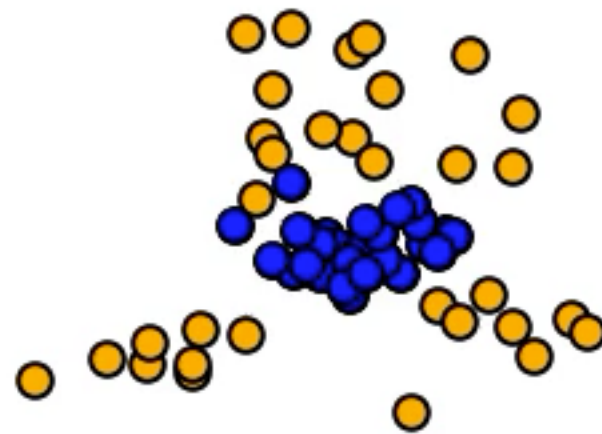
$K = 1$



$K = 10$



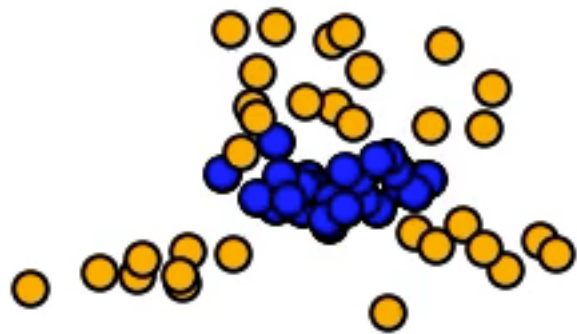
$K = N$



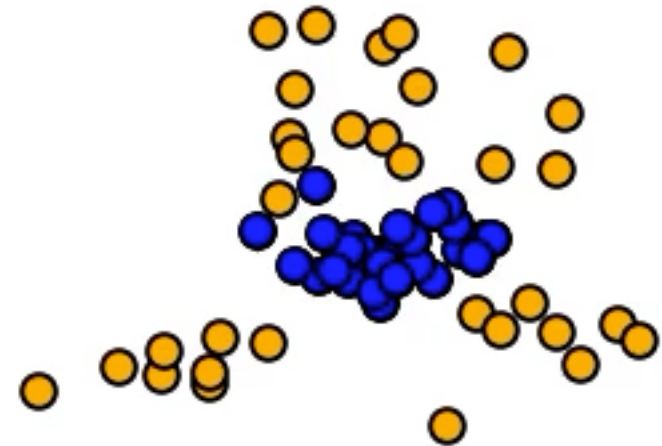
Discrete data : EM + MW_2

T_{rand}

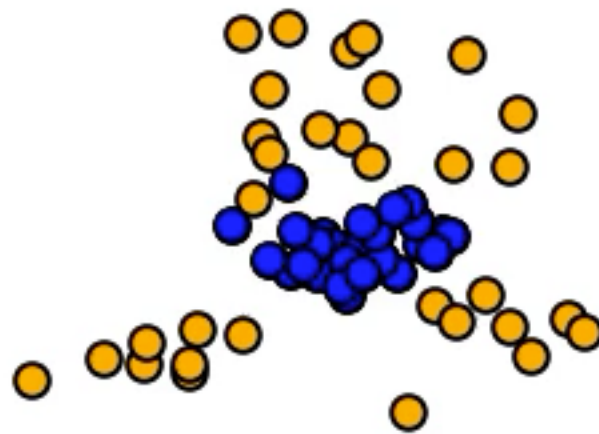
$K = 1$



$K = 10$



$K = N$



Two applications in image processing

Color transfer



Color transfer, T_{rand}



Color transfer, T_{mean}



Color transfer (sliced OT)



Texture interpolation



[Ongoing work with A. Leclaire]