Optimal Transport for image processing

Julie Delon

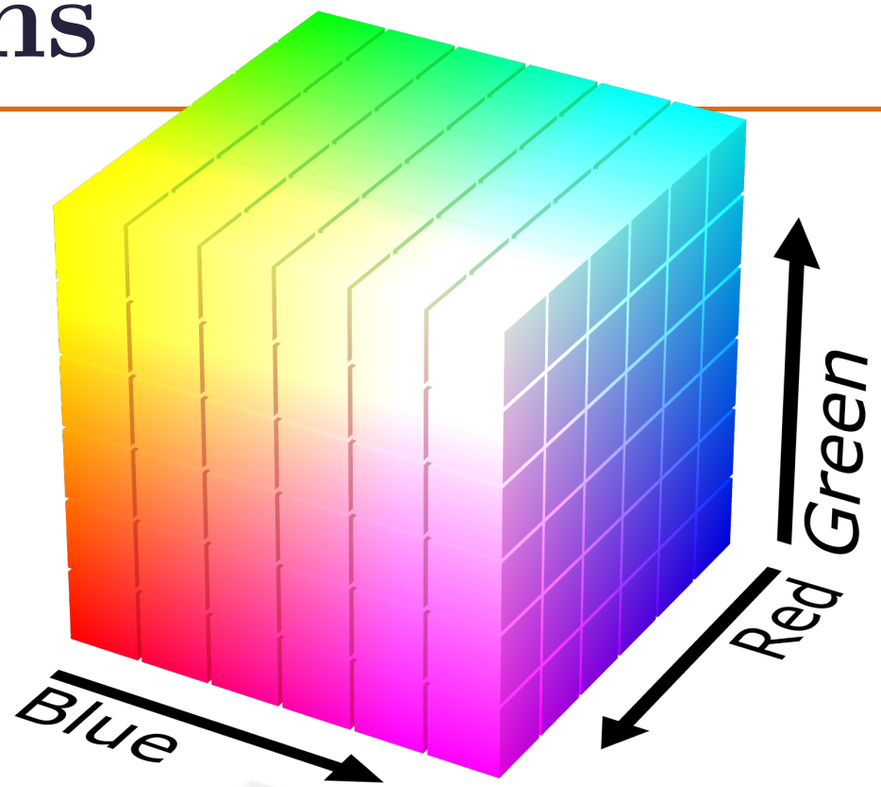
introduction à l'imagerie numérique

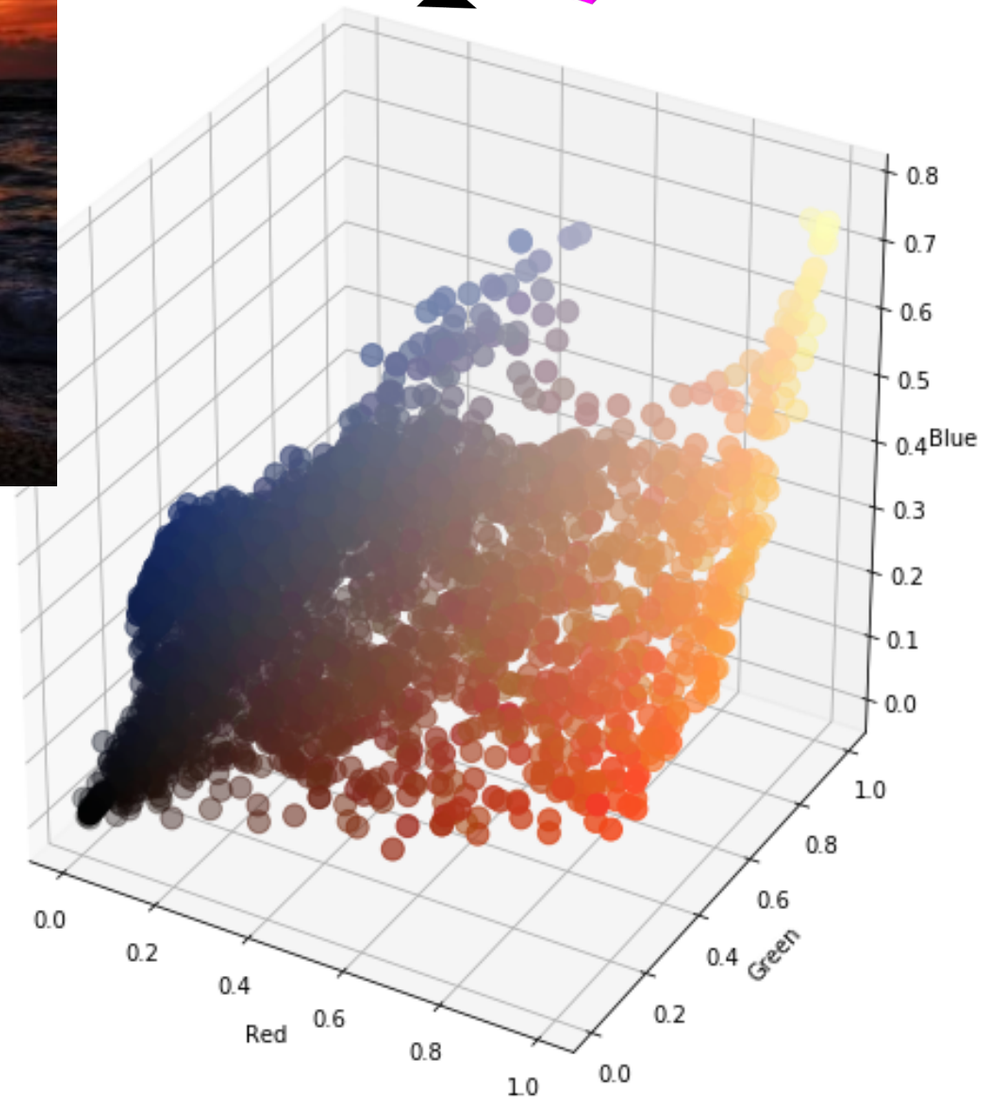
LABORATOIRE MAP5

Université de Paris

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



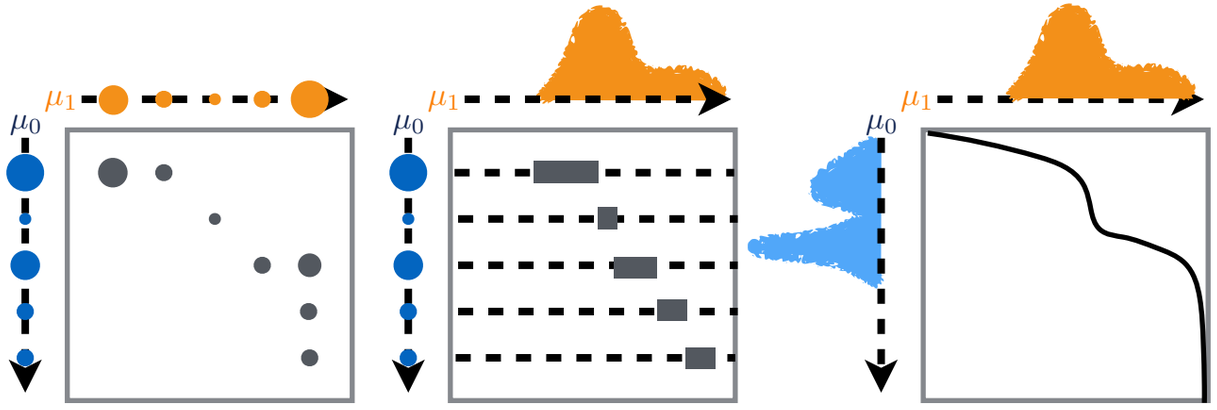
[Kantorovich, *On the transfer of masses*, 1942]

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

Couplings

$\Pi(\mu_0,\mu_1)$ = probability distributions

on $X\times X$ with marginals μ_0 and μ_1



Optimal transport

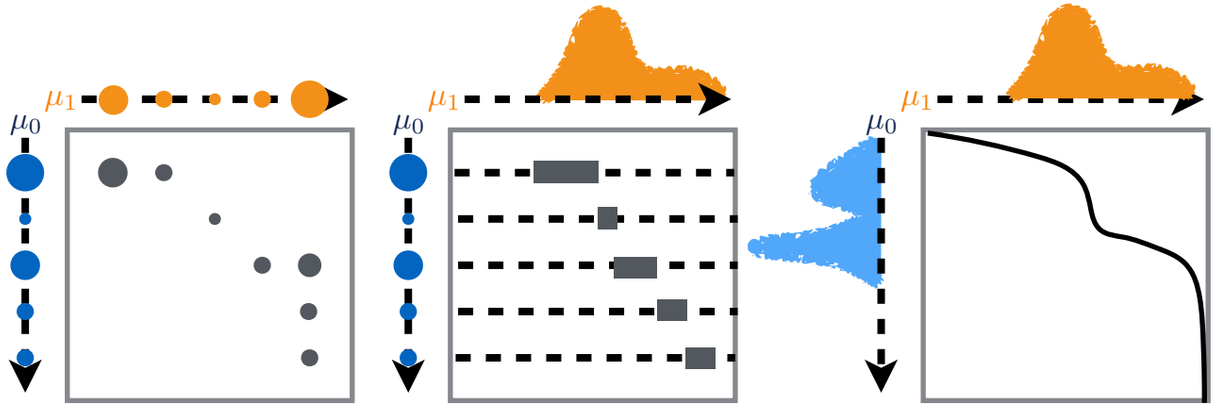


[Kantorovich, *On the transfer of masses*, 1942]

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

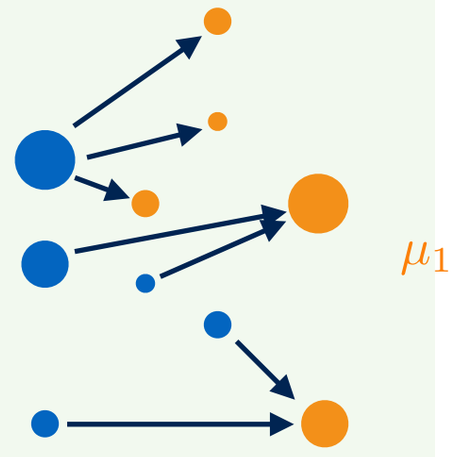
Couplings

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



Discrete case

$$\mu_0 = \sum_i \pi_0^i \,\delta_{x_i}, \qquad \mu_1 = \sum_j \pi_1^j \,\delta_{y_j}$$



$$\Pi(\mu_0,\mu_1) = \{\text{matrices } \gamma \text{ s.t. } \gamma_{i,j}\geq 0,\ \sum_i \gamma_{i,j} = \pi_1^j,\ \sum_j \gamma_{i,j} = \pi_0^i\}$$

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

Optimal transport

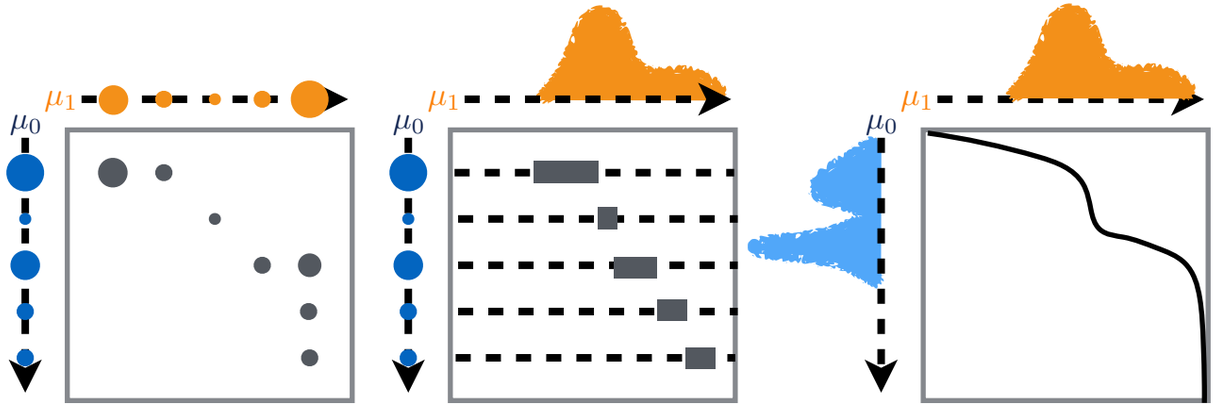


[Kantorovich, *On the transfer of masses*, 1942]

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

Couplings

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



Wasserstein distances

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

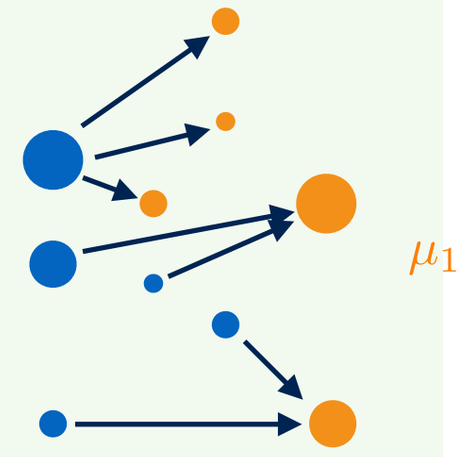
$W_p(\mu_0,\mu_1) := W_c(\mu_0,\mu_1)^{\frac{1}{p}}$ distance

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

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

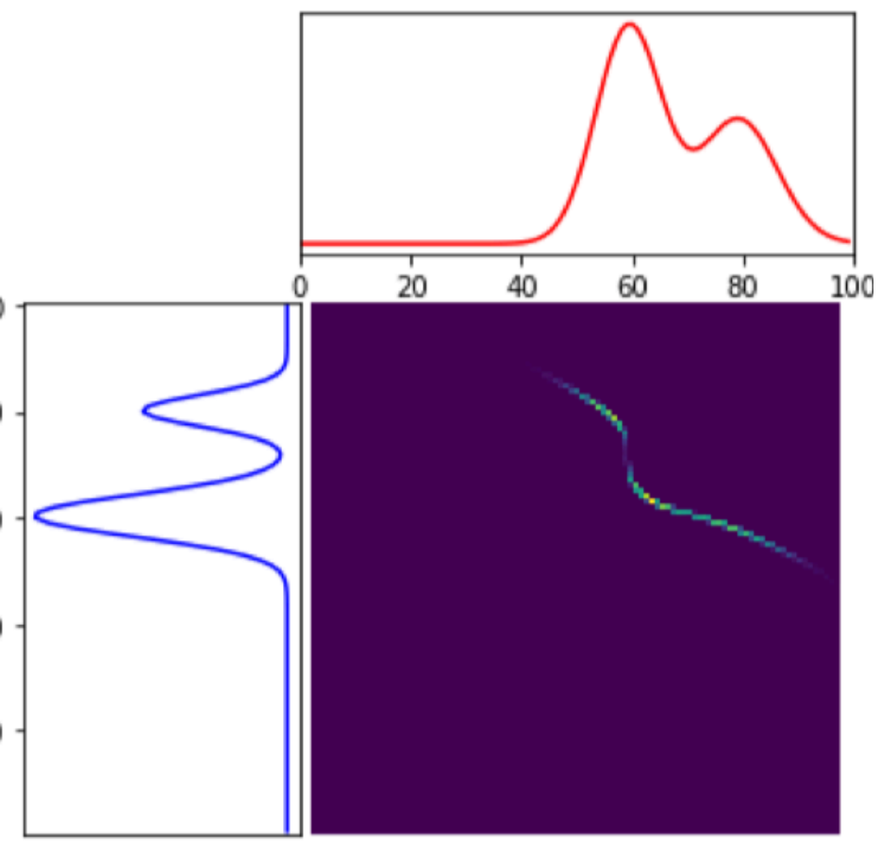
Monge, 1781

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

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

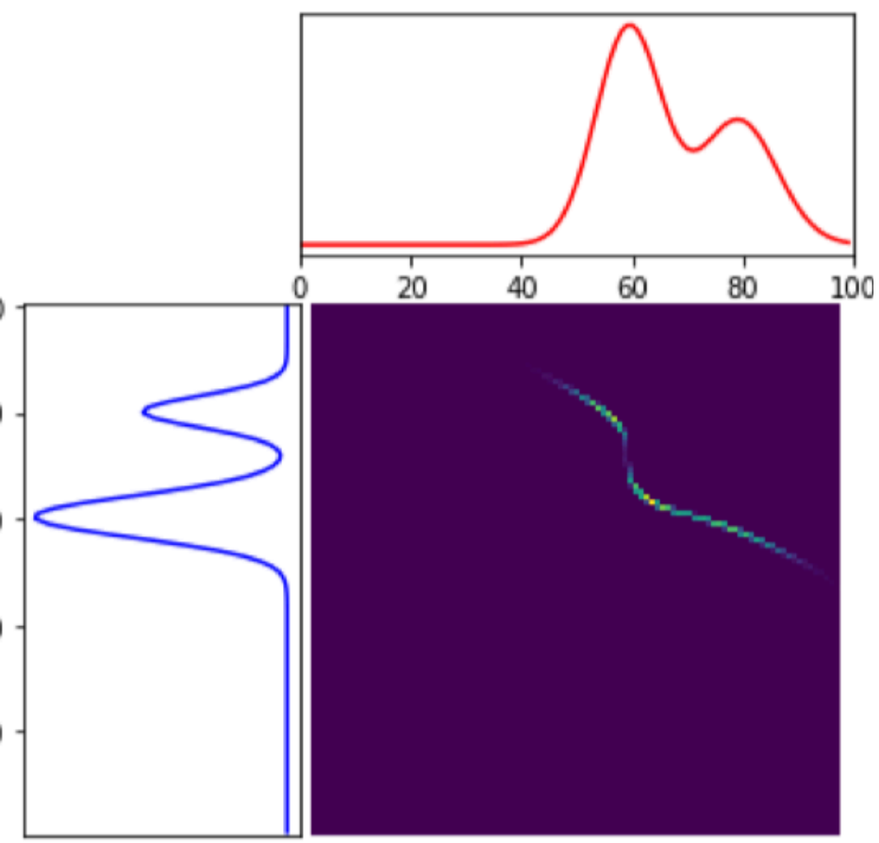
Monge, 1781

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

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:

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

Optimal transport in one dimension

Optimal Transport in 1D



Optimal Transport in 1D

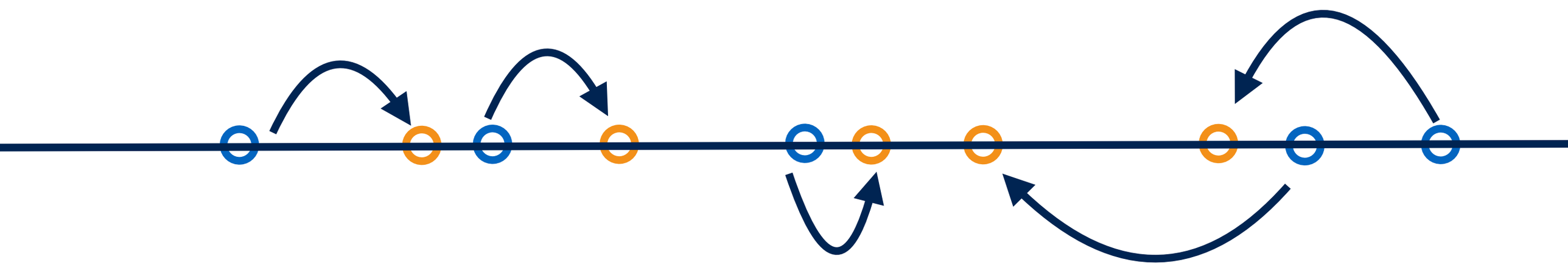


If $c(x, y) = f(x - y)$ with f convex,

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

Numerical approaches

Linear programming

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

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

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

Linear programming

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

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

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

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

Linear programming

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

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

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

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

Assignment: Hungarian algo. [Kuhn 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]

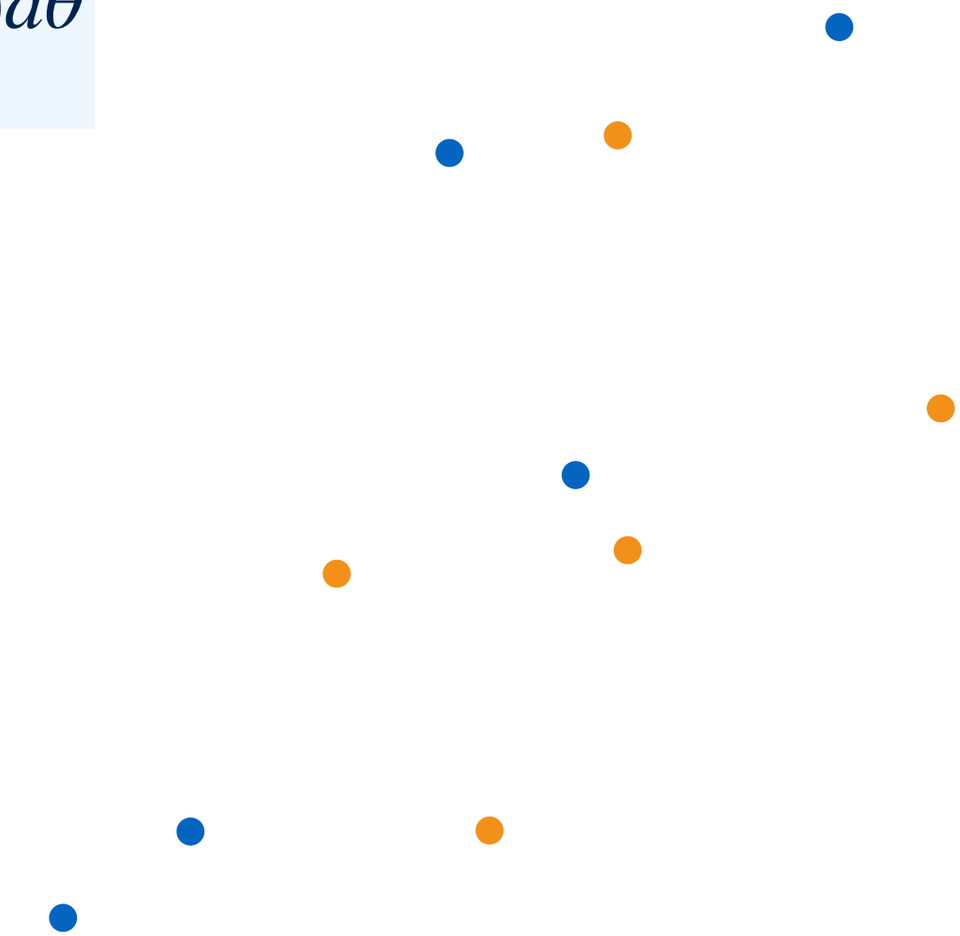
Sliced OT [Rabin et al. 11, Rabin et al.15]

Entropic OT [Cuturi 13,...]

Sliced optimal transport

Replace classical OT by

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



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

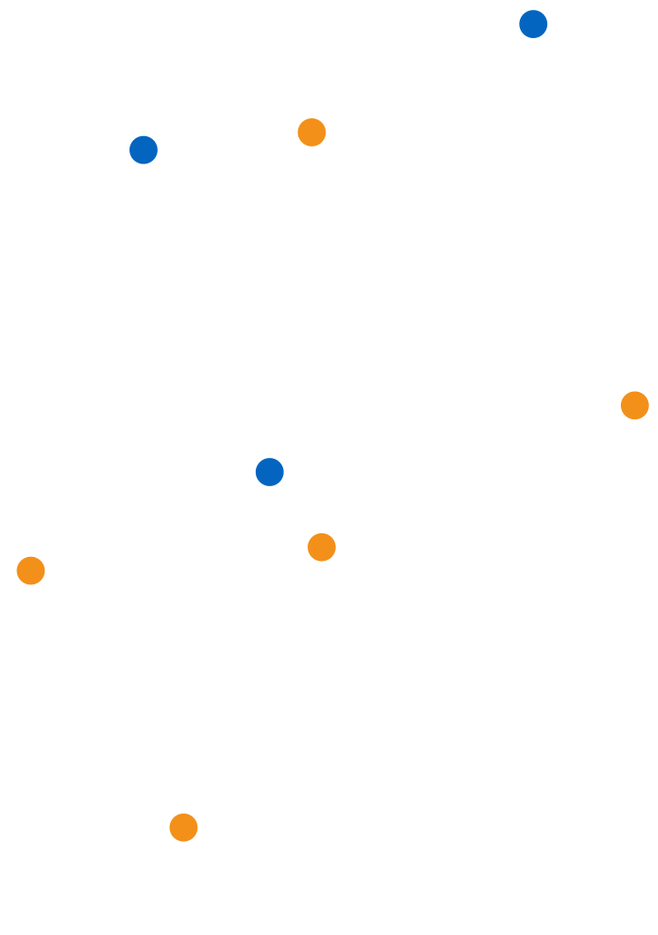
Sliced optimal transport

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

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



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

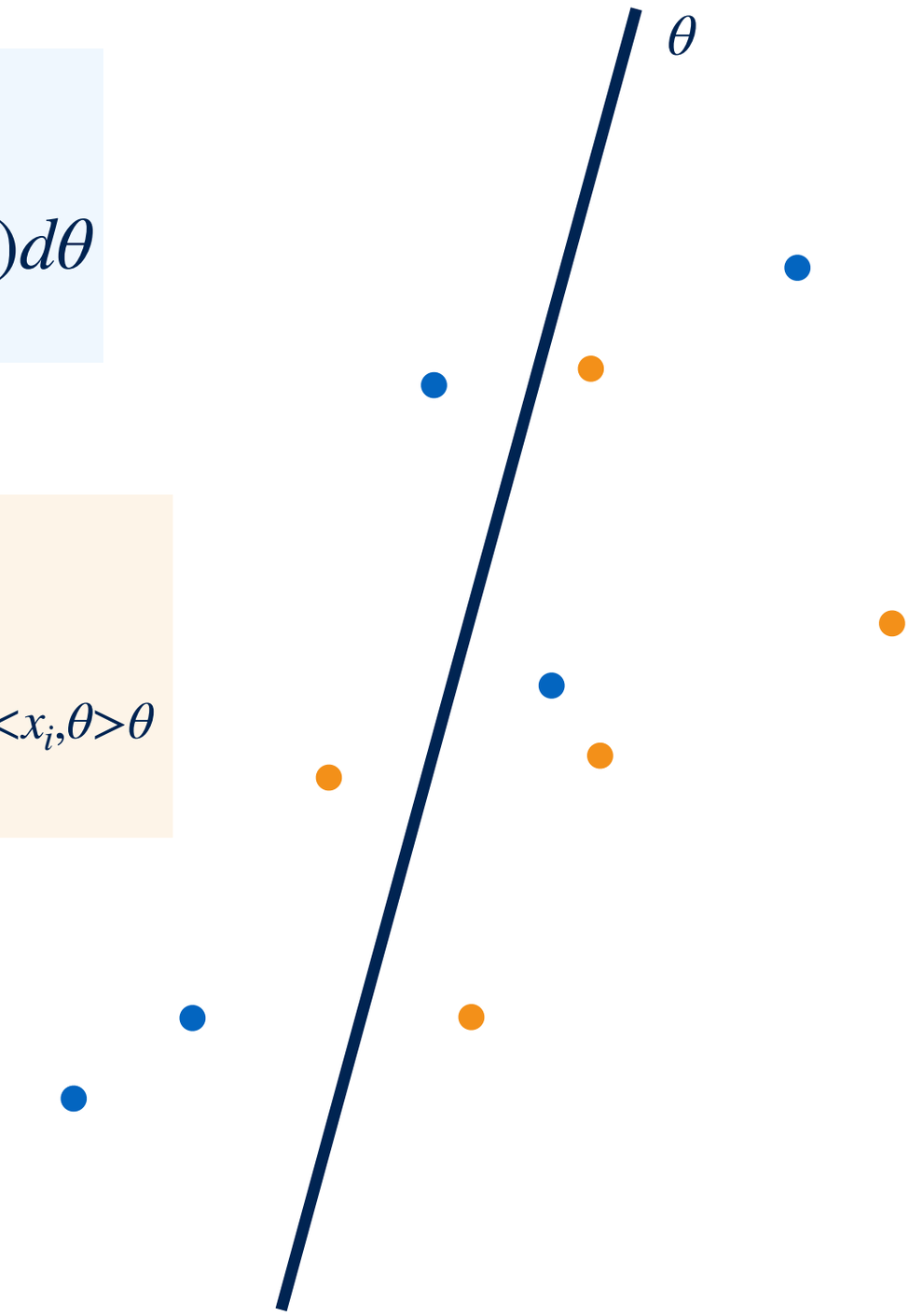
Sliced optimal transport

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

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



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

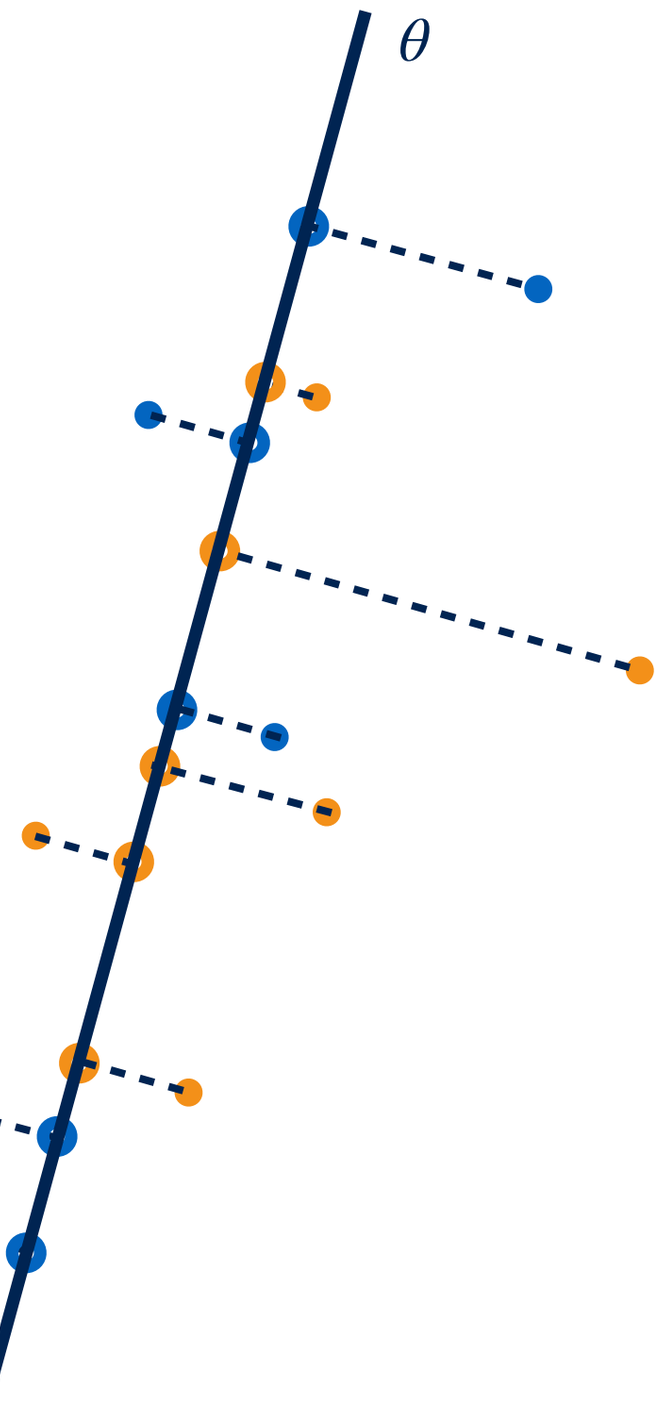
Sliced optimal transport

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

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



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

Sliced optimal transport

Replace classical OT by

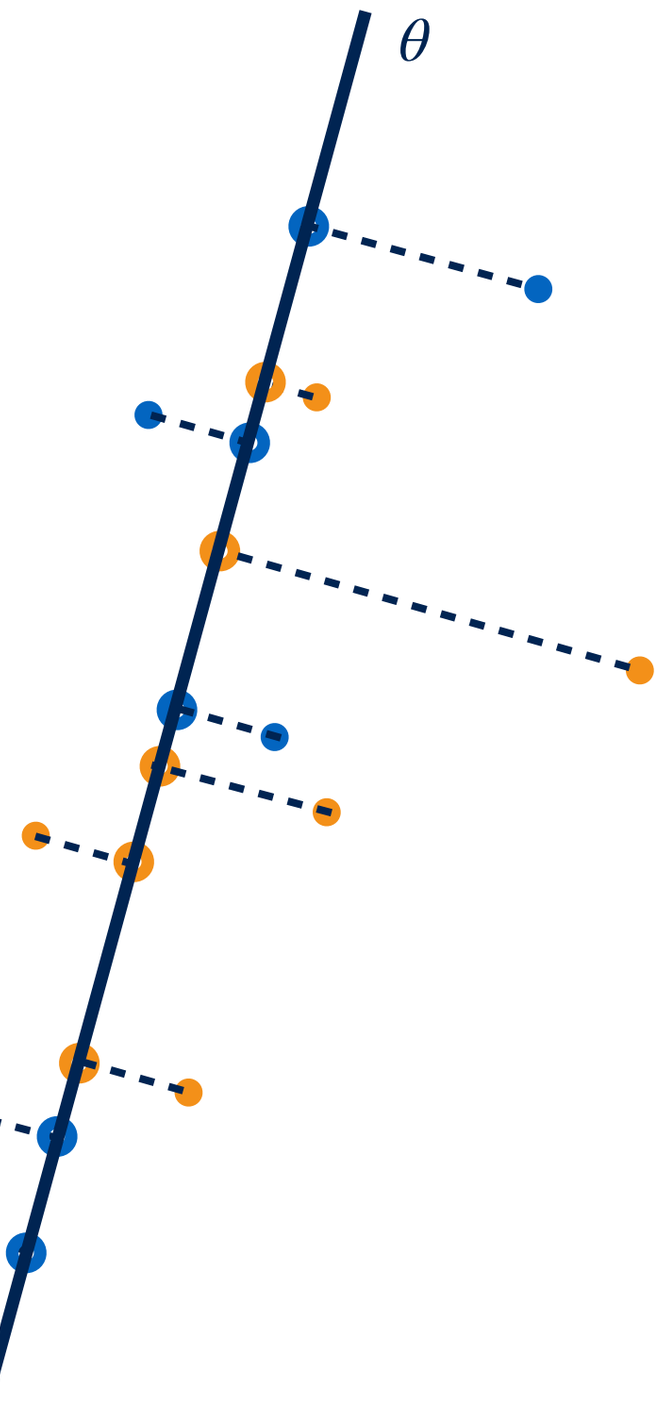
$$SW_2^2(\mu_0,\mu_1) = \int_{\mathbb{S}^{d-1}} W_2^2(p_\theta \# \mu_0, p_\theta \# \mu_1) d\theta$$

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 $<\mu_1,\theta>$.

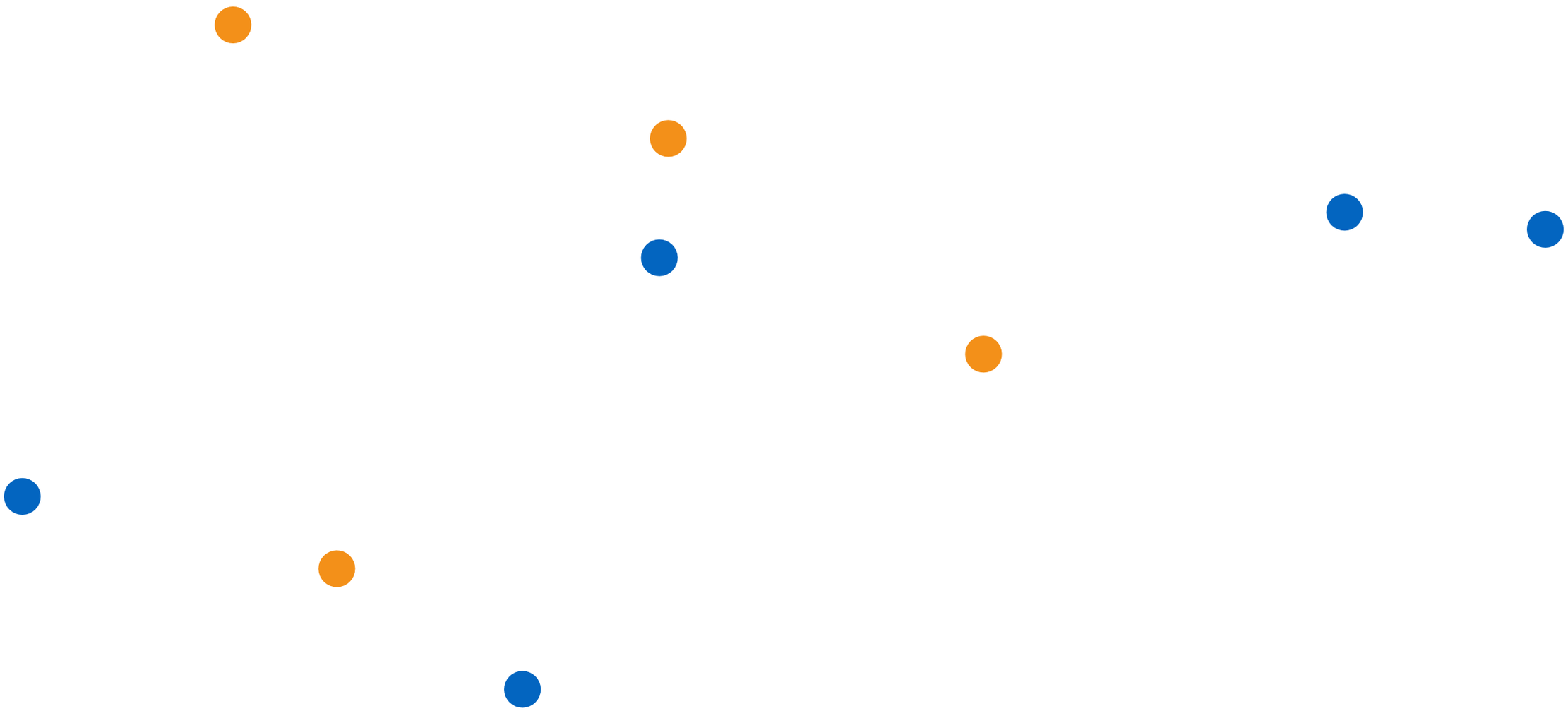


[Rabin et al. 11, Rabin et al. 2015]

Sliced optimal transport

Stochastic gradient descent on

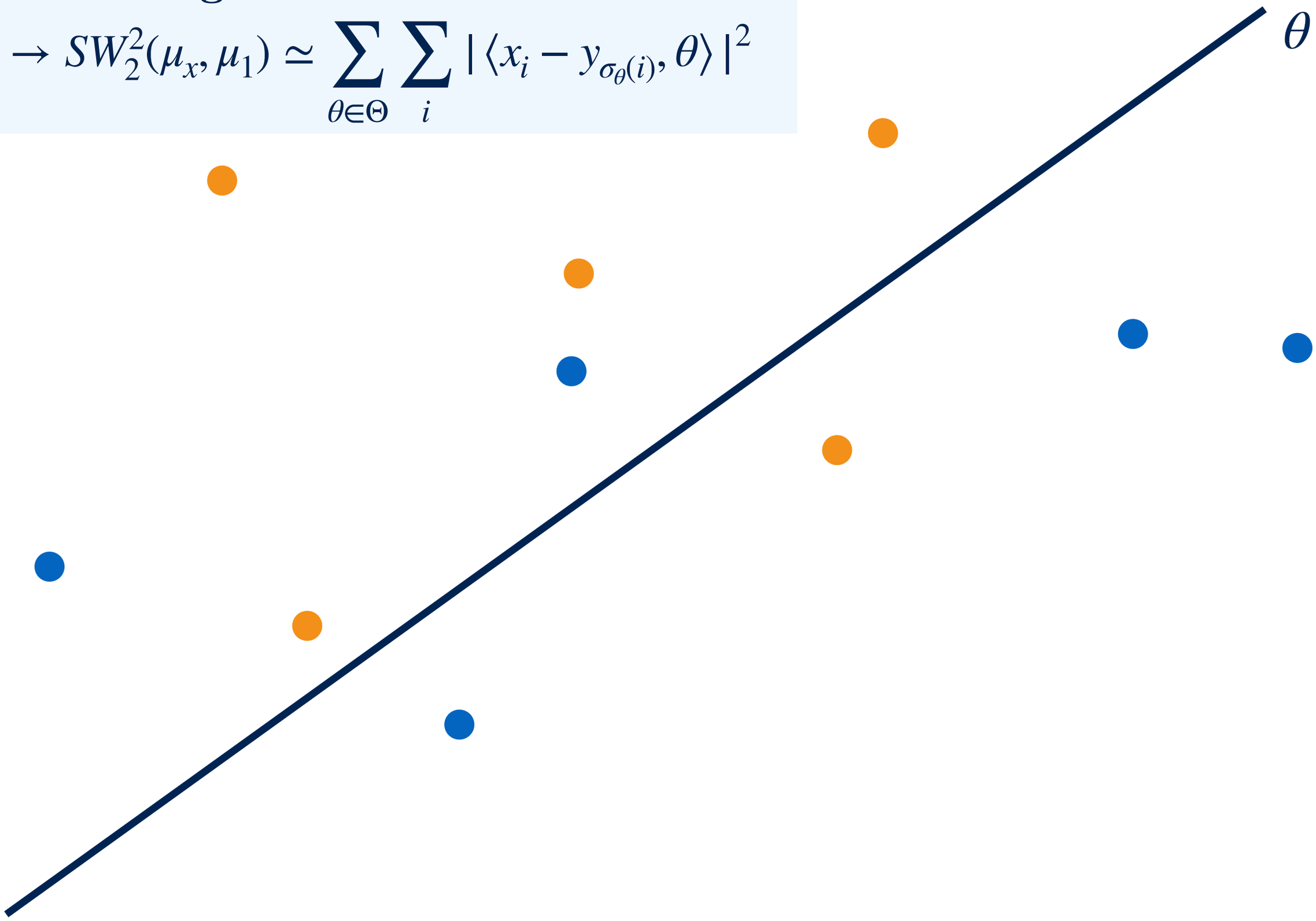
$$x \to SW_2^2(\mu_x, \mu_1) \simeq \sum_{\theta \in \Theta} \sum_i |\langle x_i - y_{\sigma_\theta(i)}, \theta \rangle|^2$$



Sliced optimal transport

Stochastic gradient descent on

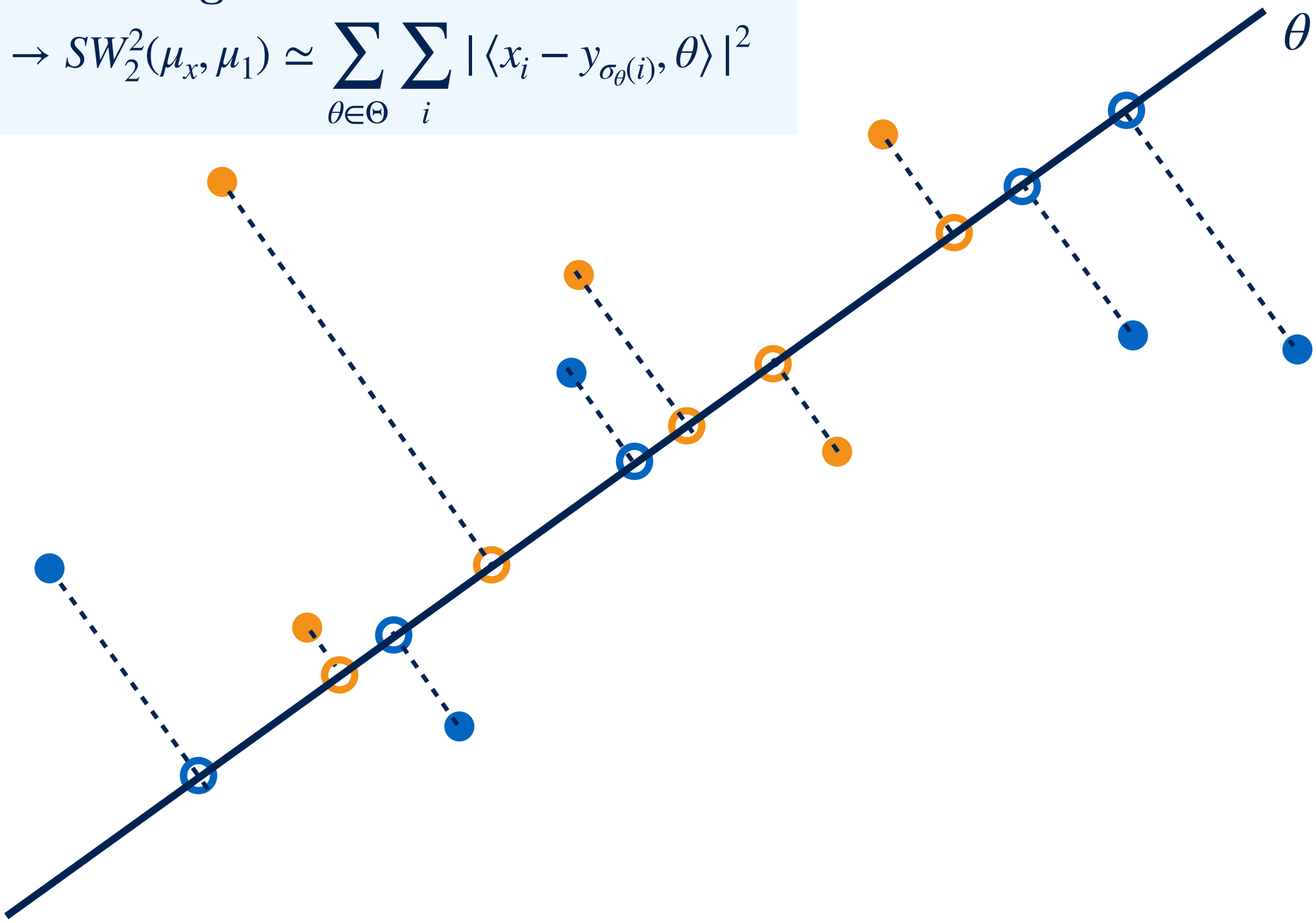
$$x \to SW_2^2(\mu_x, \mu_1) \simeq \sum_{\theta \in \Theta} \sum_i |\langle x_i - y_{\sigma_\theta(i)}, \theta \rangle|^2$$



Sliced optimal transport

Stochastic gradient descent on

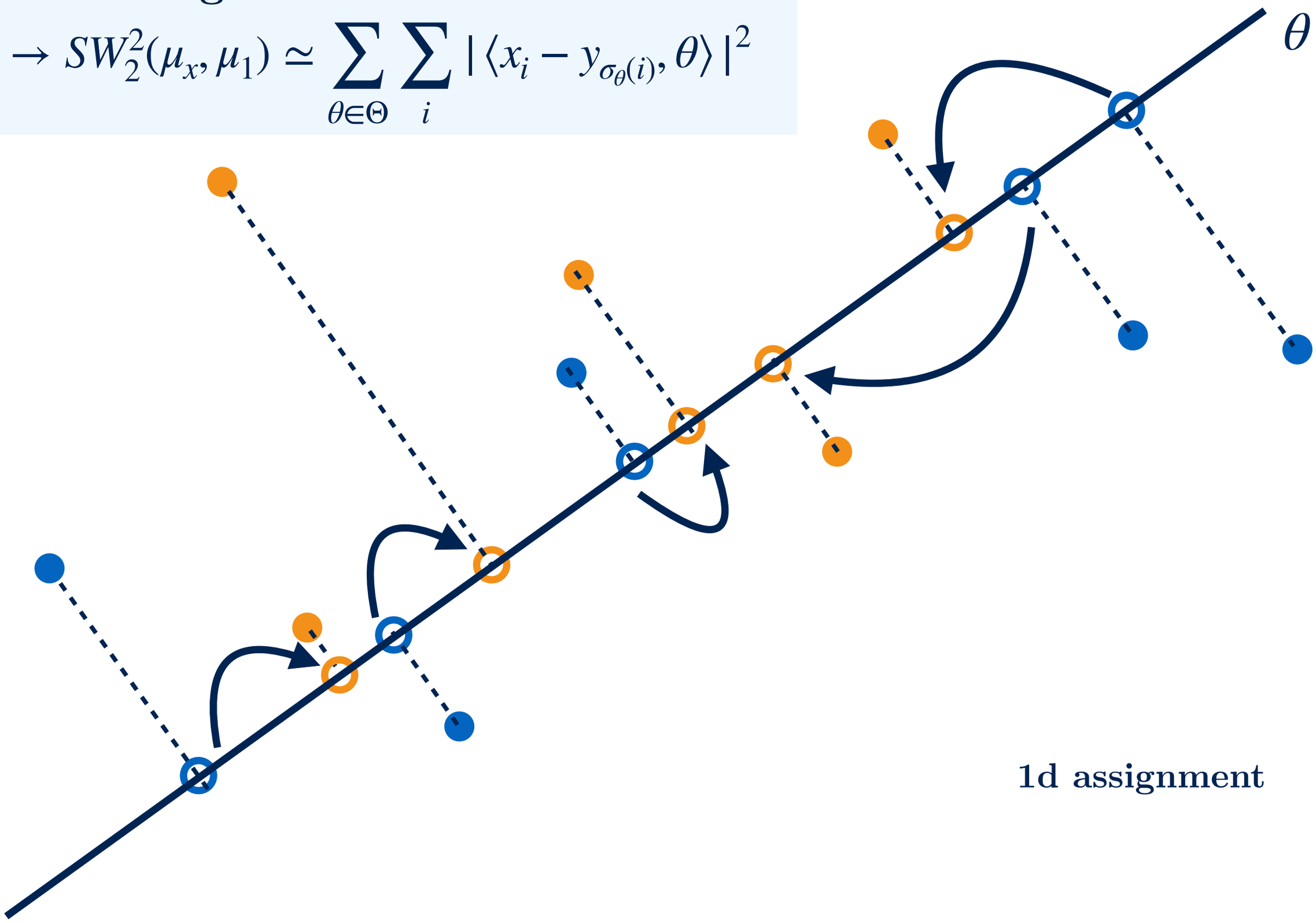
$$x \to SW_2^2(\mu_x, \mu_1) \simeq \sum_{\theta \in \Theta} \sum_i |\langle x_i - y_{\sigma_\theta(i)}, \theta \rangle|^2$$



Sliced optimal transport

Stochastic gradient descent on

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

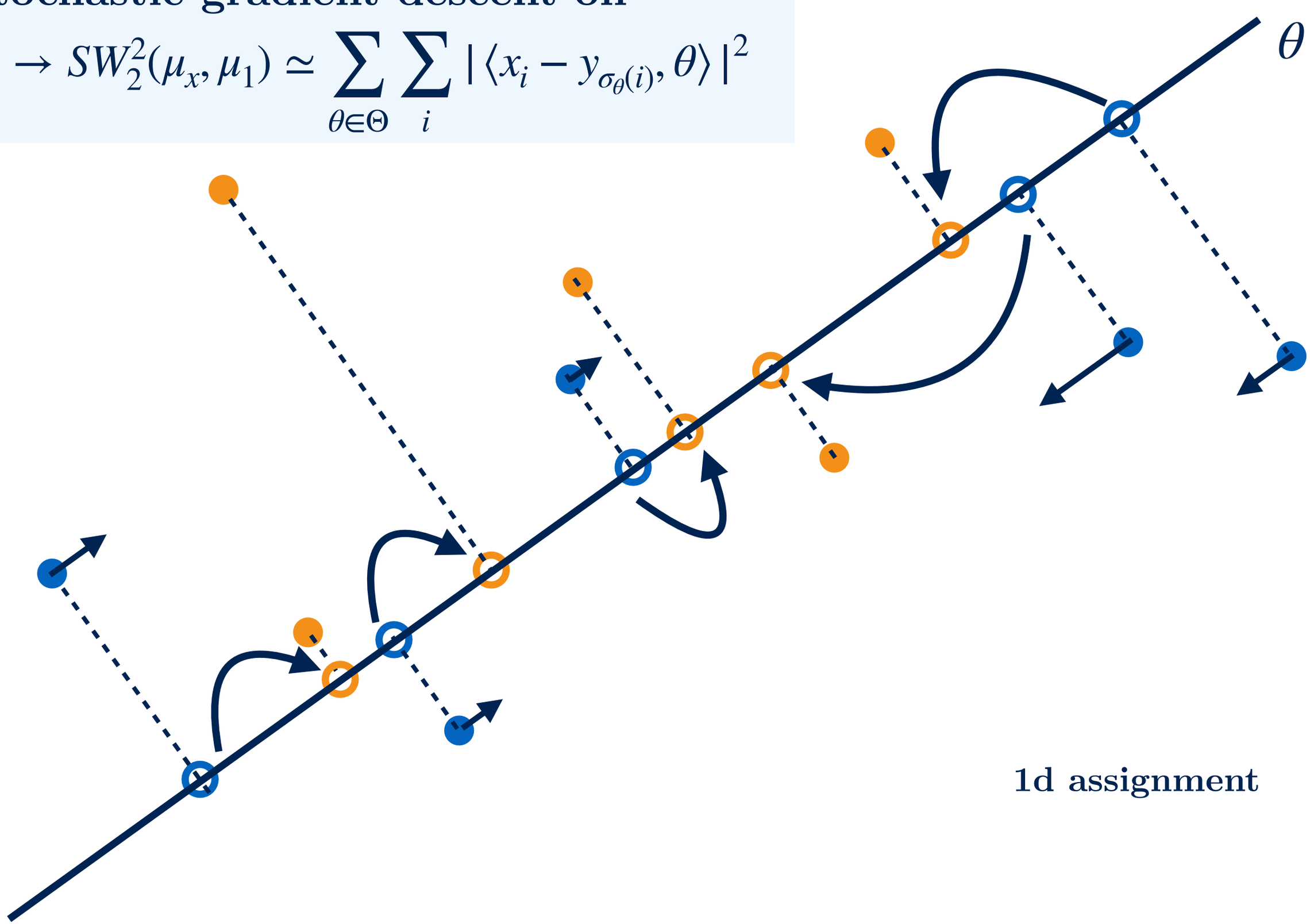


1d assignment

Sliced optimal transport

Stochastic gradient descent on

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

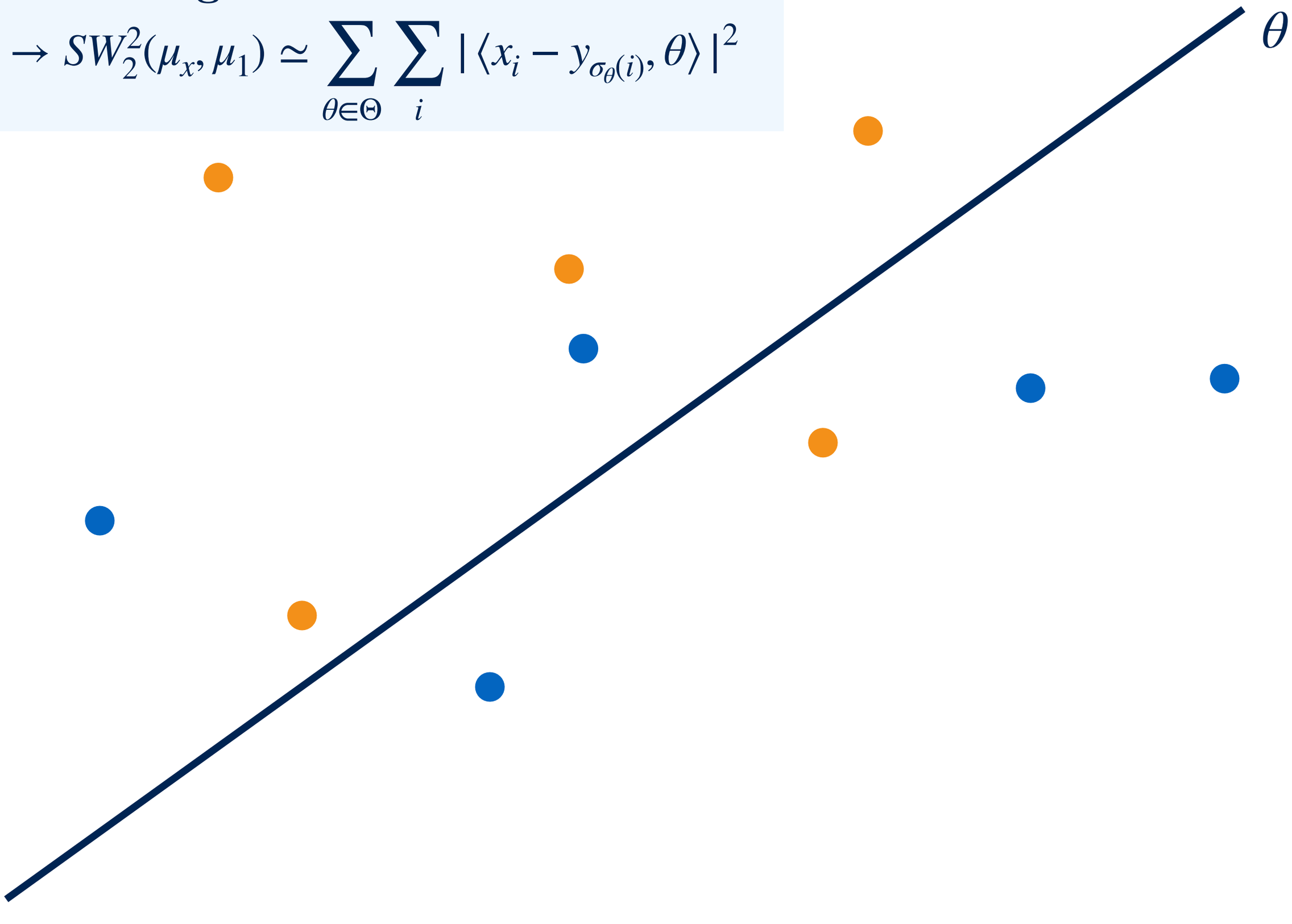


1d assignment

Sliced optimal transport

Stochastic gradient descent on

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



Entropic OT

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

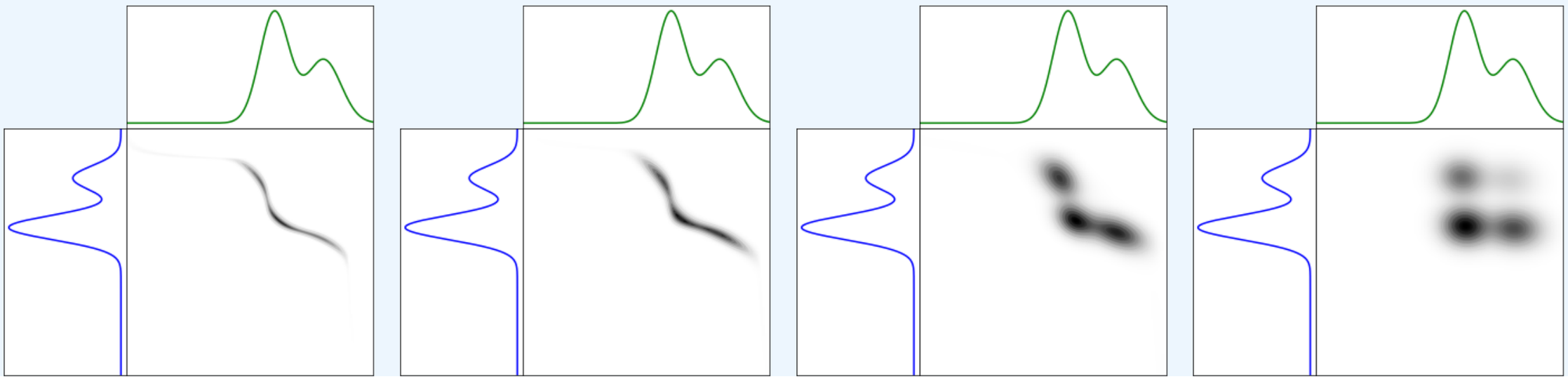
Entropic OT [Cuturi '13]

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

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

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

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



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

Optimal transport between Gaussians

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

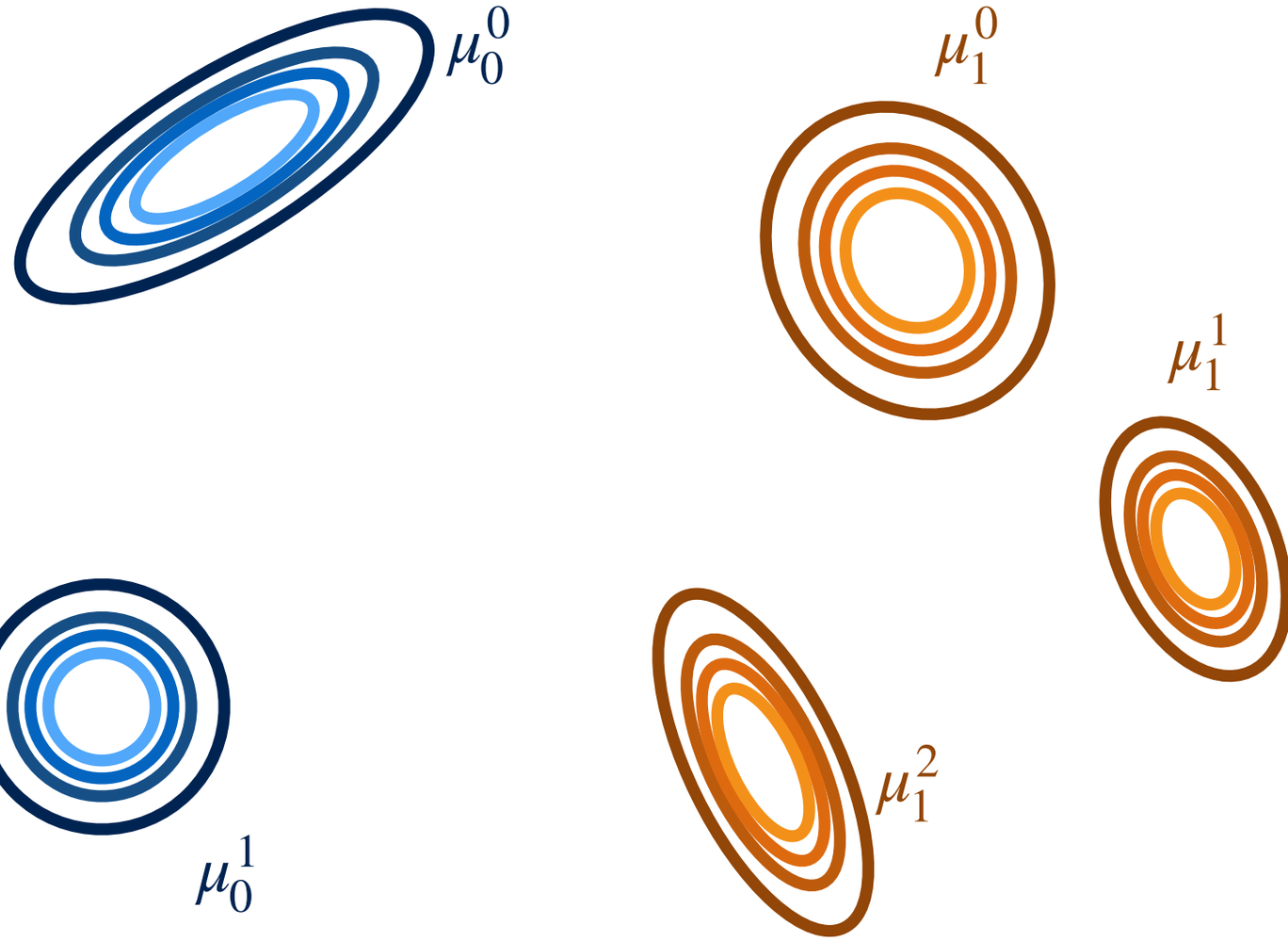
$$W_2^2(\mu_0, \mu_1) = \|m_0 - m_1\|^2 + \underbrace{\operatorname{tr}\left(\Sigma_0 + \Sigma_1 - 2\left(\Sigma_0^{\frac{1}{2}}\Sigma_1\Sigma_0^{\frac{1}{2}}\right)^{\frac{1}{2}}\right)}_{B^2(\Sigma_0,\Sigma_1)}$$

If Σ_0 non-singular, affine optimal map

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

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

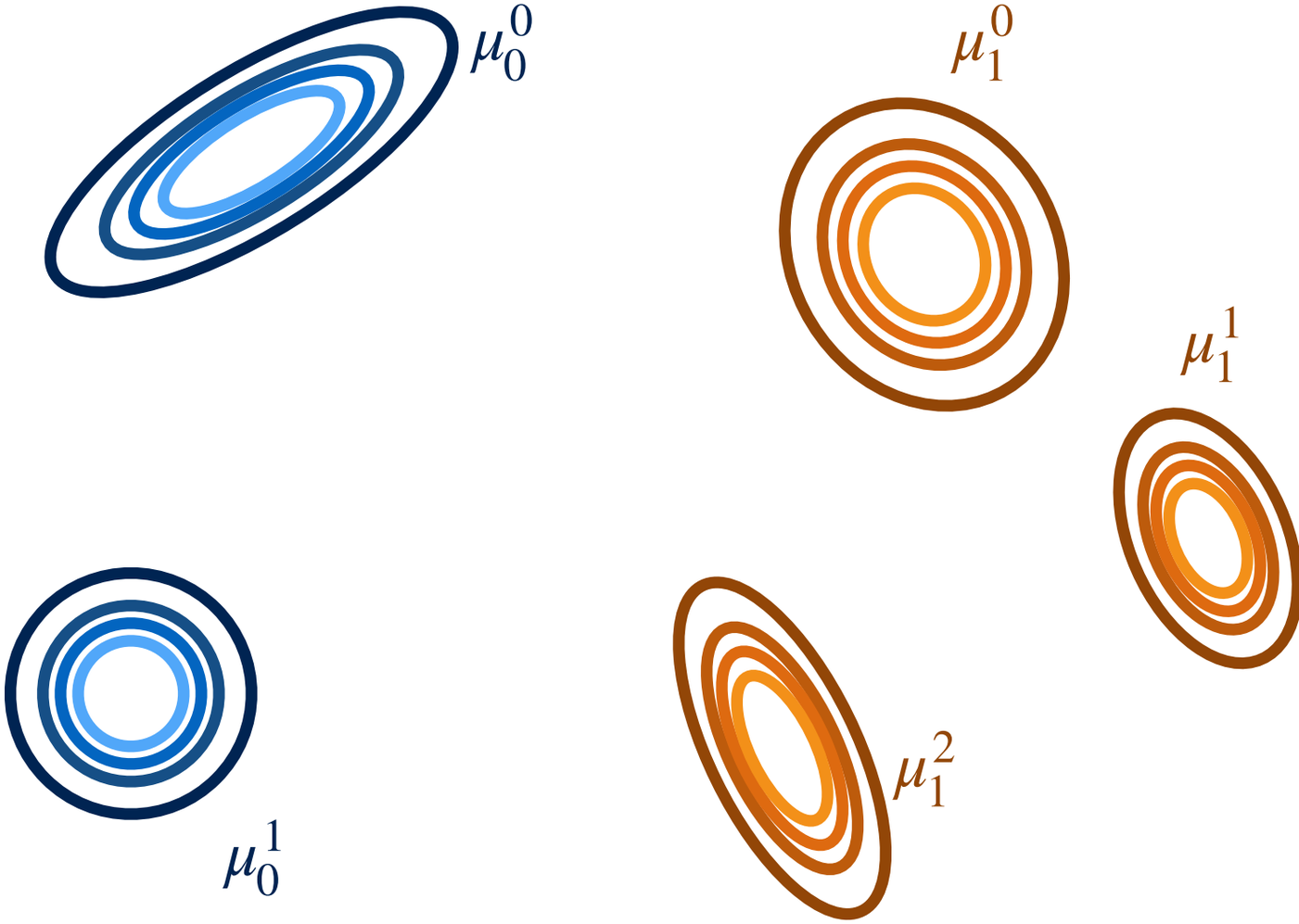


Definition of MW2

$\mu_0 = \sum_{k=1}^{K_0} \pi_0^k \mu_0^k$ 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

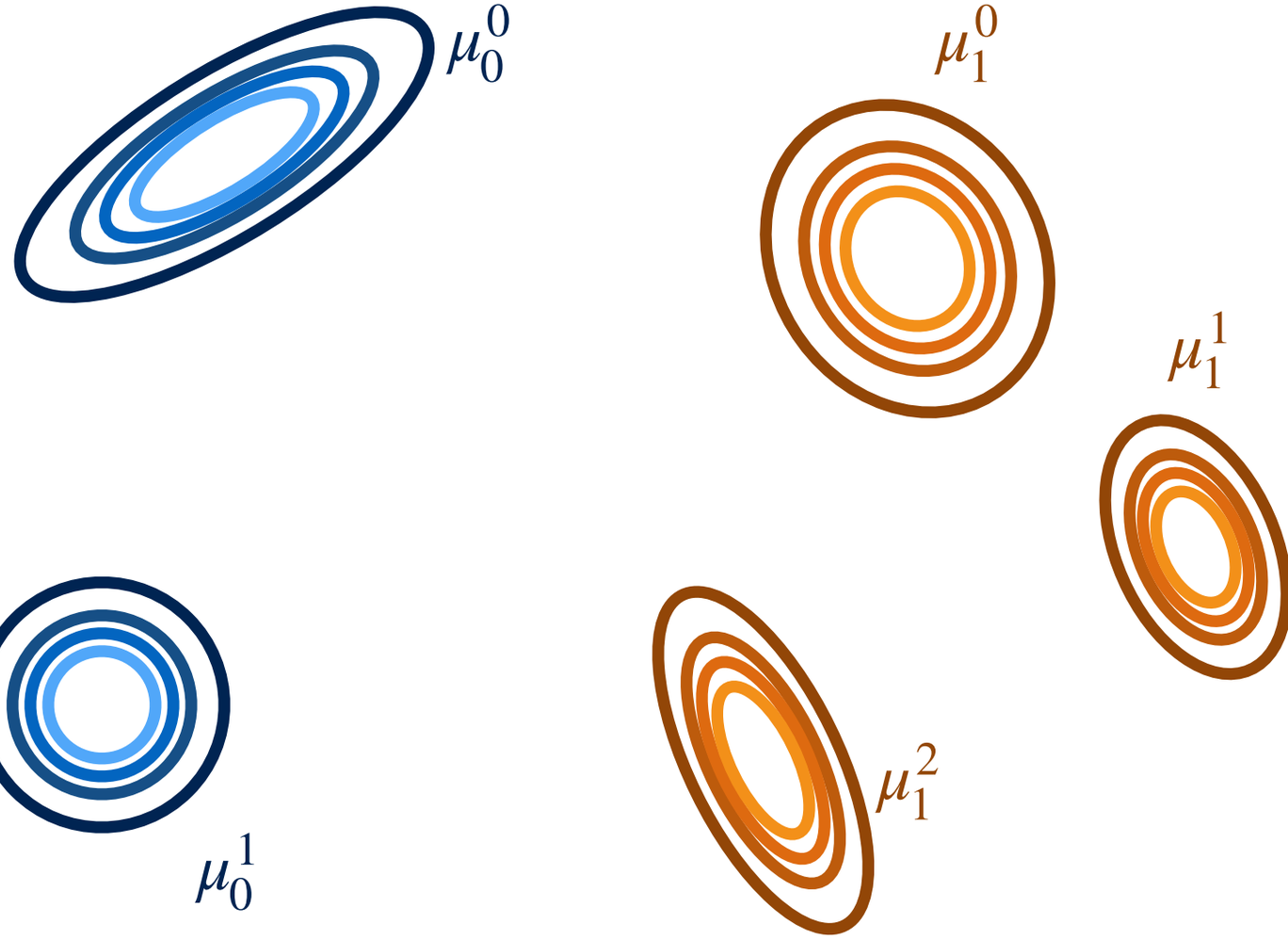
$\mu_0 = \sum_{k=1}^{K_0} \pi_0^k \mu_0^k$ and $\mu_1 = \sum_{k=1}^{K_1} \pi_1^k \mu_1^k$ two Gaussian mixtures

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

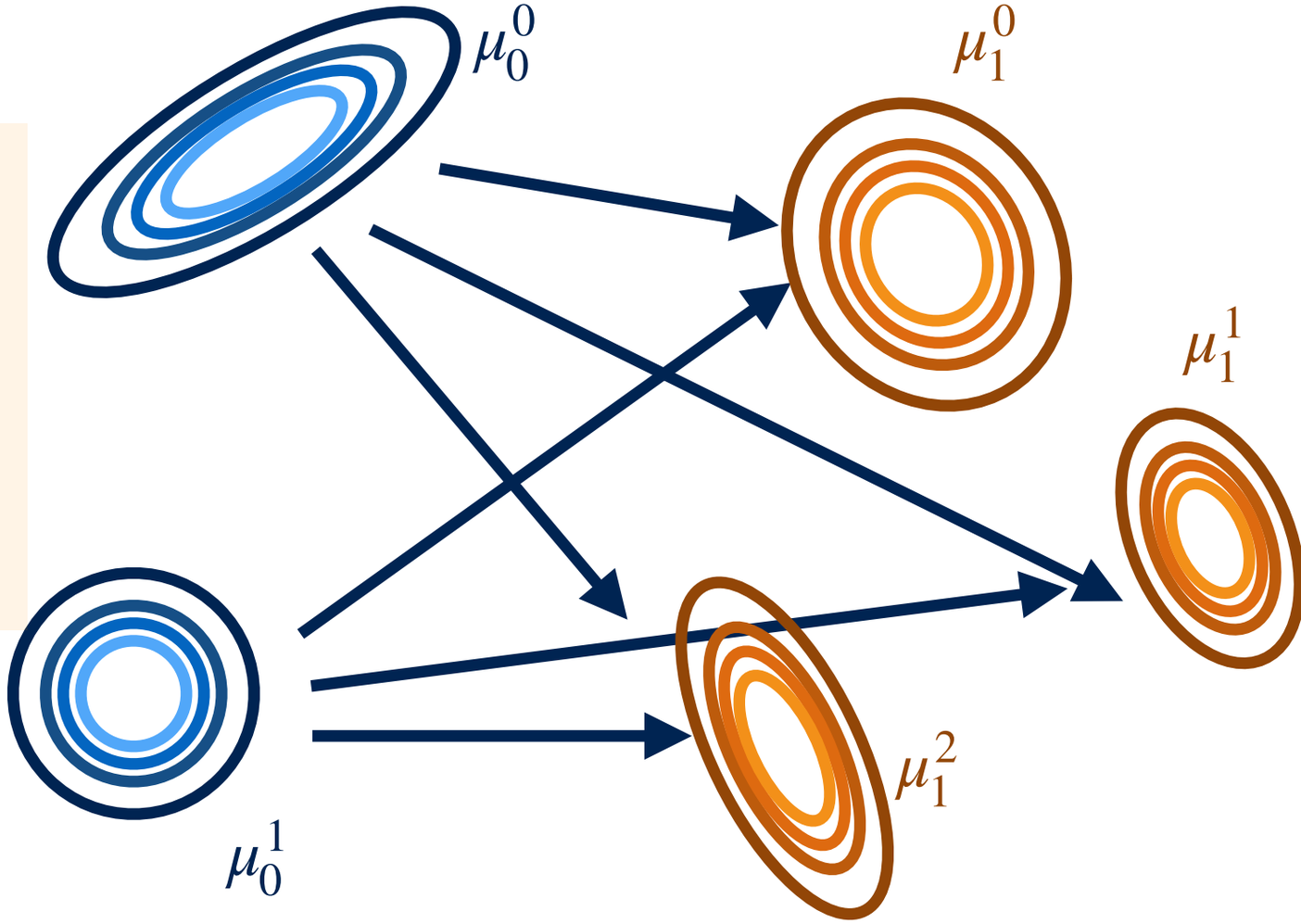
$\mu_0 = \sum_{k=1}^{K_0} \pi_0^k \mu_0^k$ and $\mu_1 = \sum_{k=1}^{K_1} \pi_1^k \mu_1^k$ two Gaussian mixtures

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

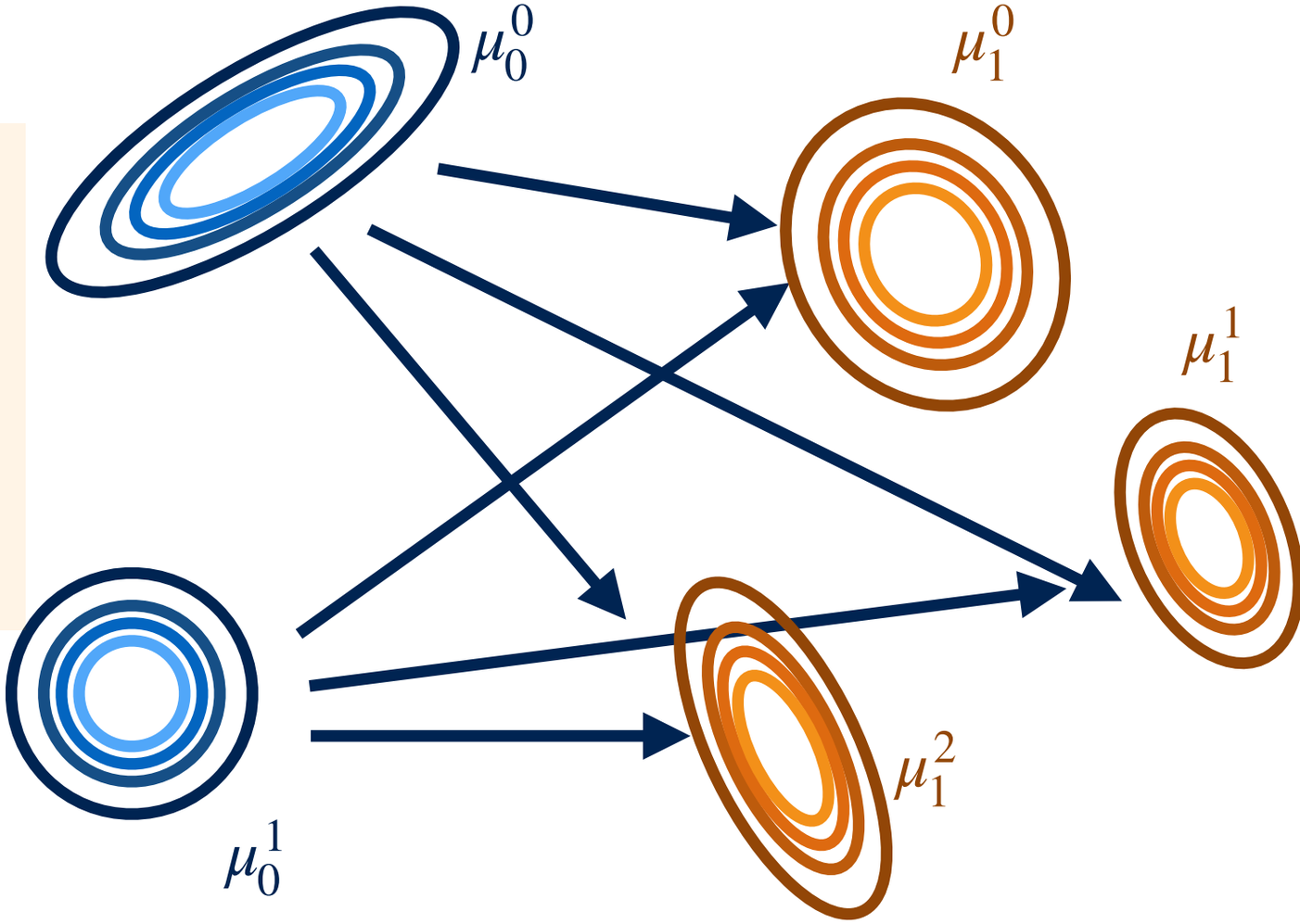
Definition:

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

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



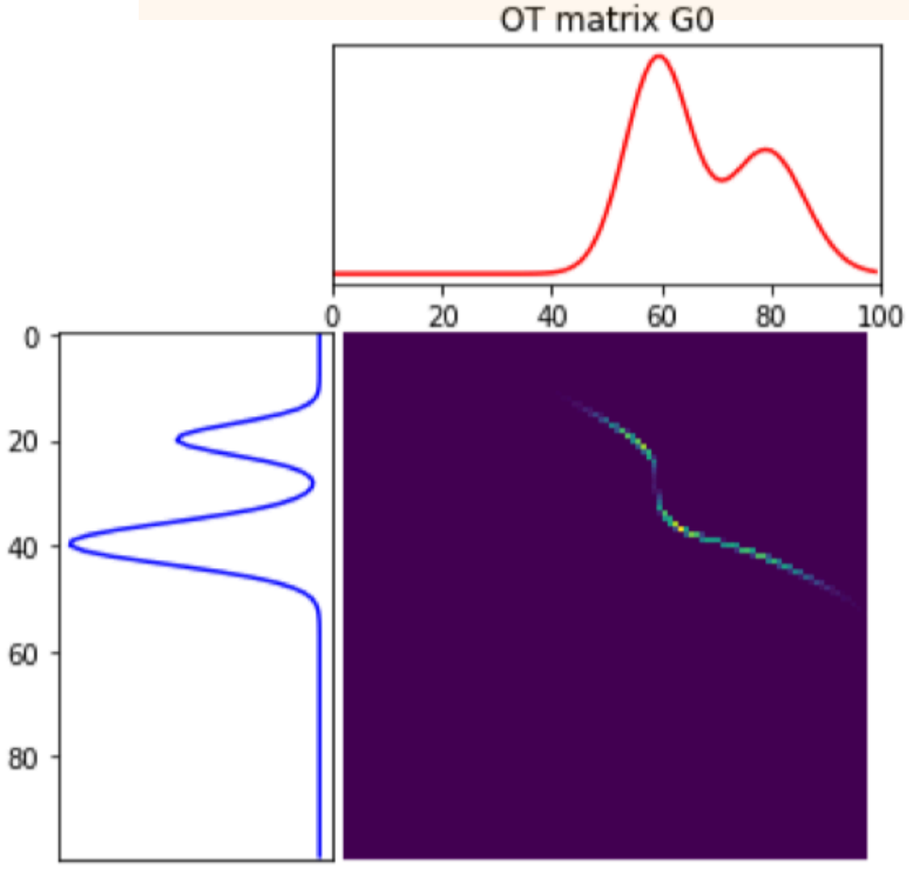
Optimal plan

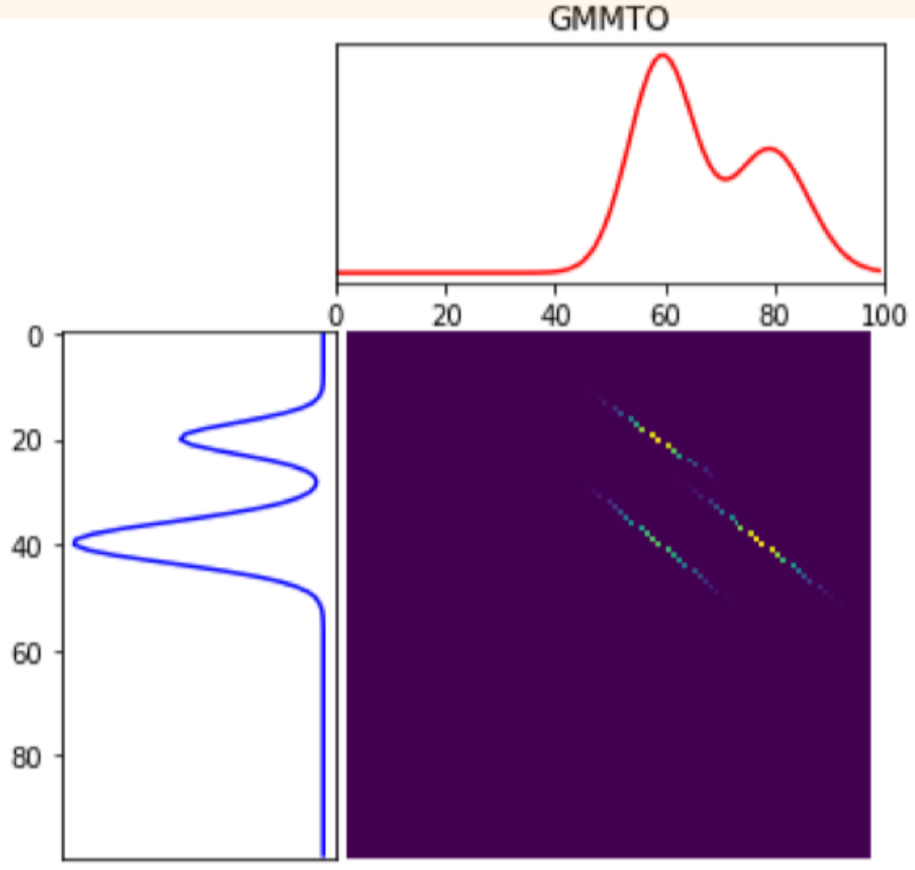
Proposition: an optimal plan γ^* for MW_2 can be written

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

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



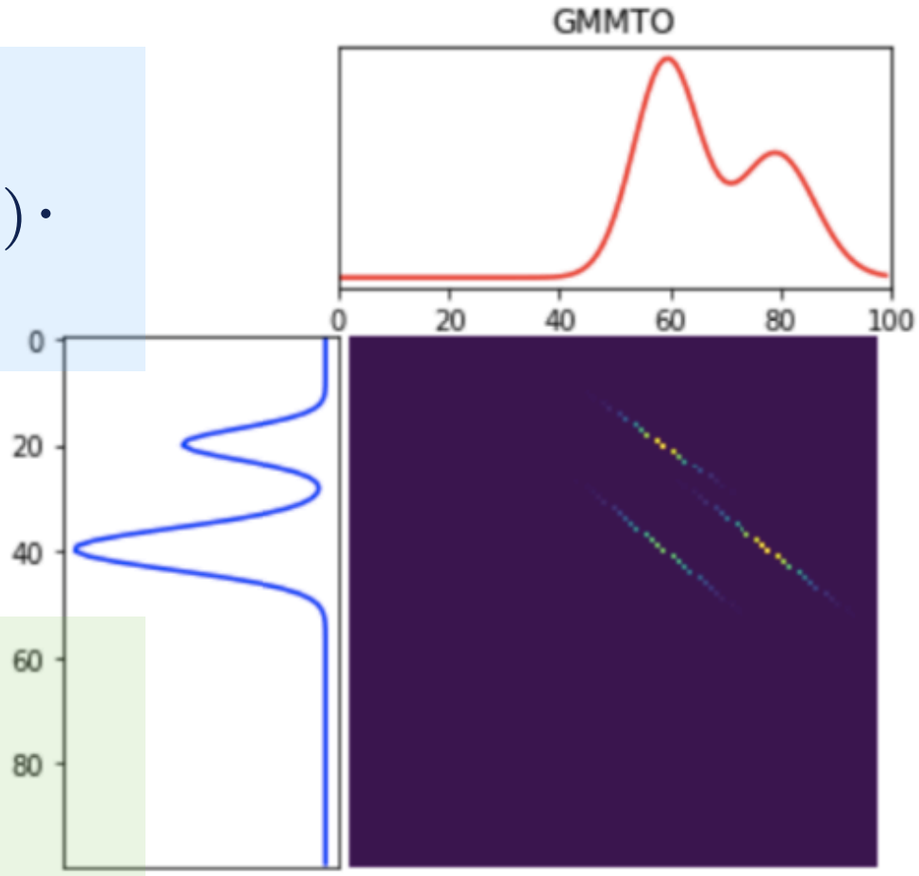


At least one solution has less than $K_0 + K_1 - 1$ components!

From a transport plan to a map

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

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



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

Color transfer



Color transfer, T_{rand}



Color transfer, T_{mean}



Color transfer (sliced OT)



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

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

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

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





