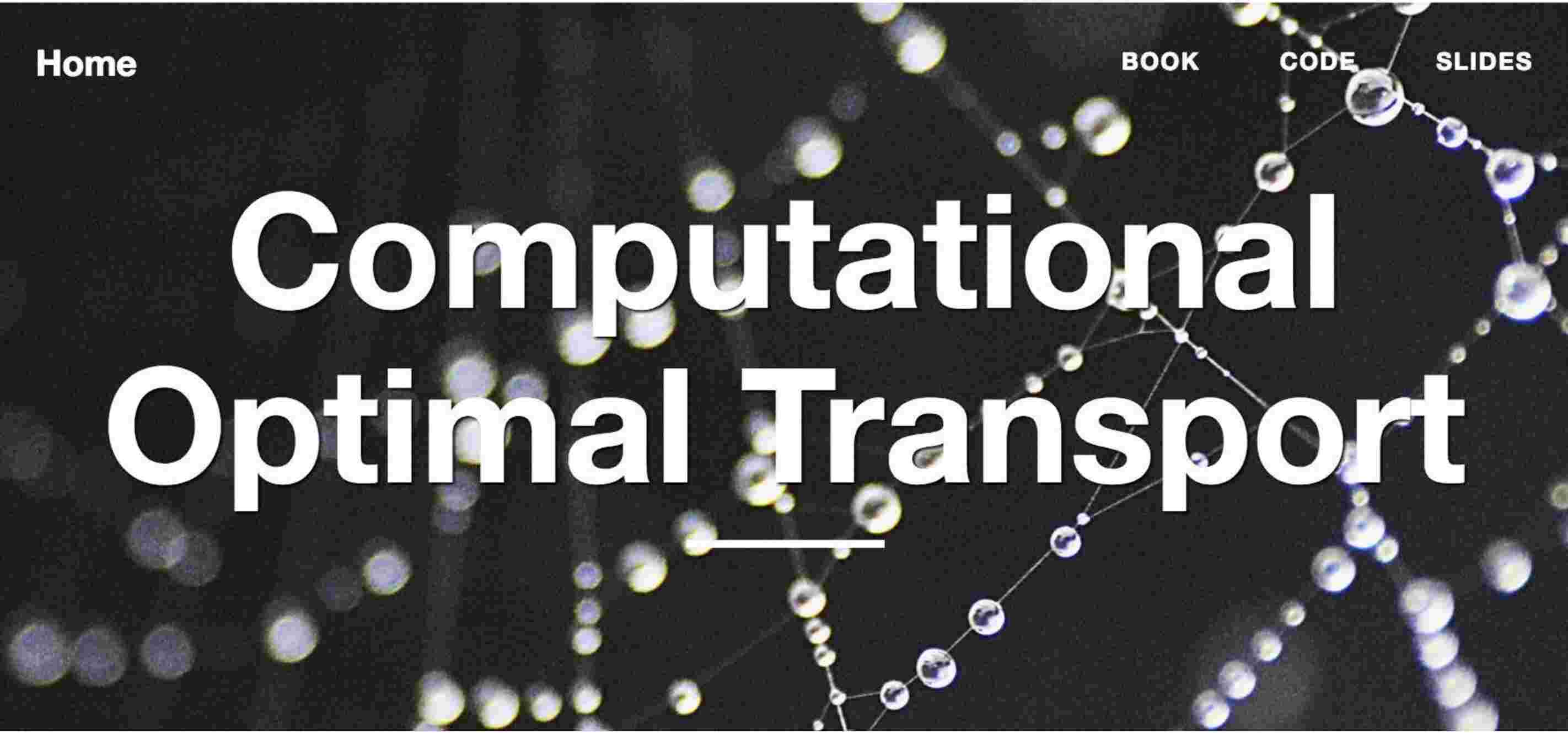
Scaling Optimal Transport for Machine Learning

Gabriel Peyré

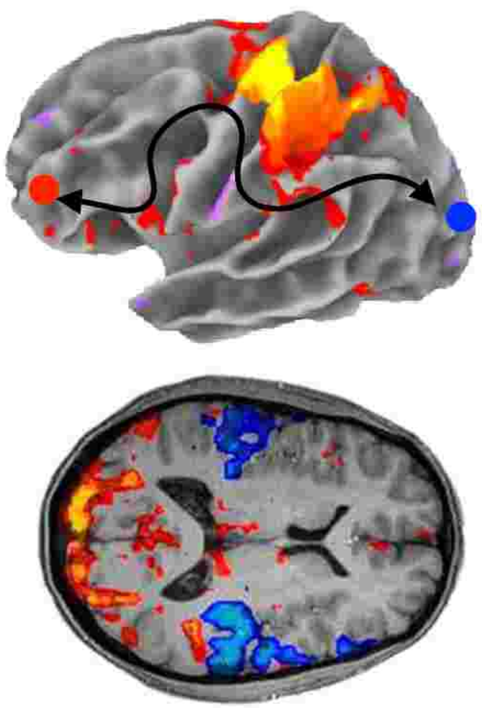
cnrs dépasser les frontières

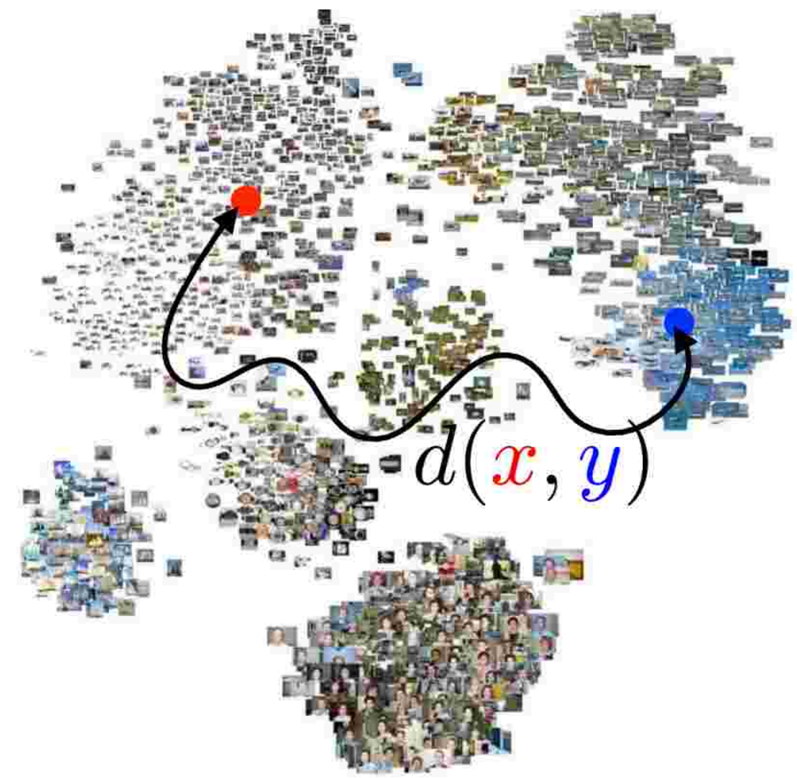
1794 ENS ÉCOLE NORMALE SUPÉRIEURE

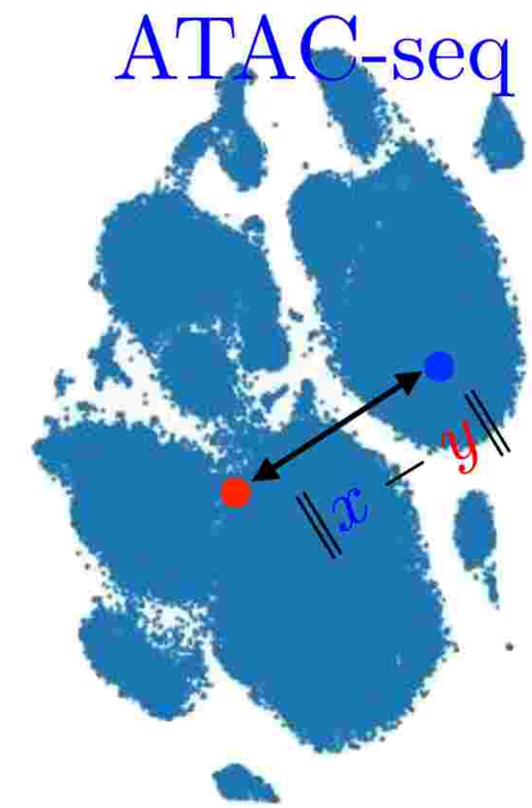
https://optimaltransport.github.io



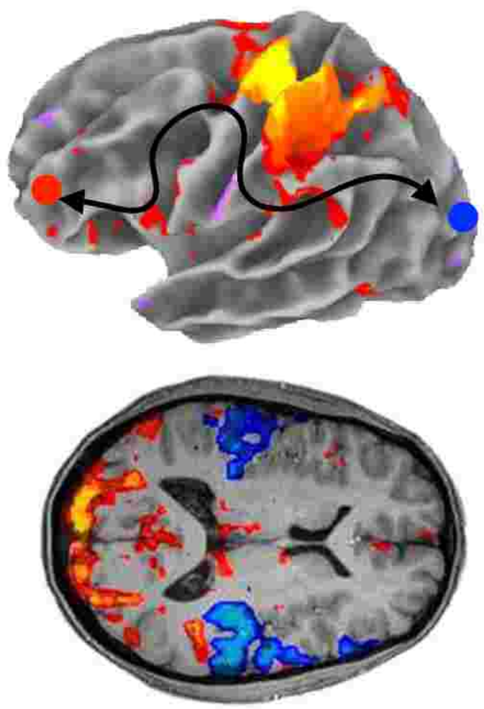
Comparing Distributions for Learning

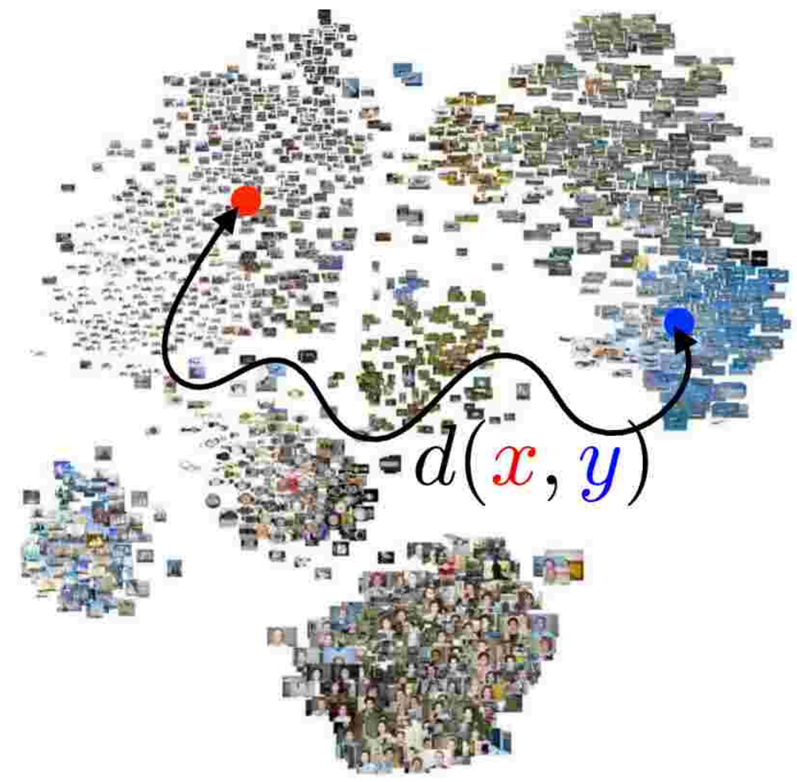


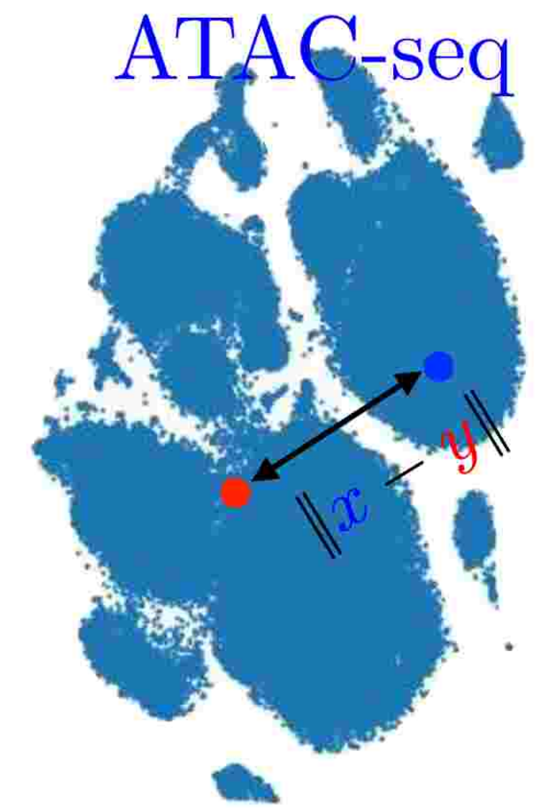




Comparing Distributions for Learning



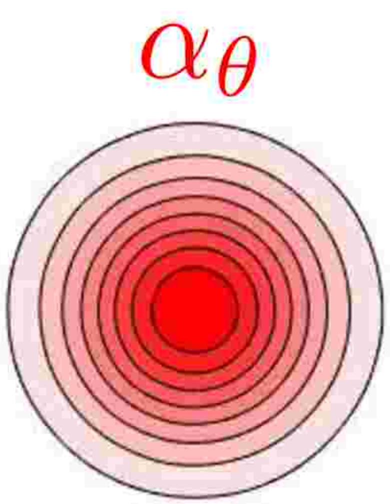


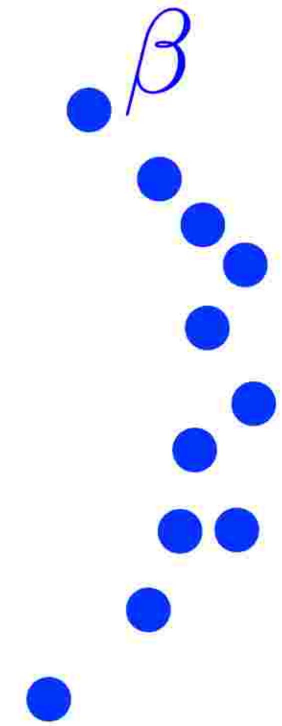


Unsupervised learning

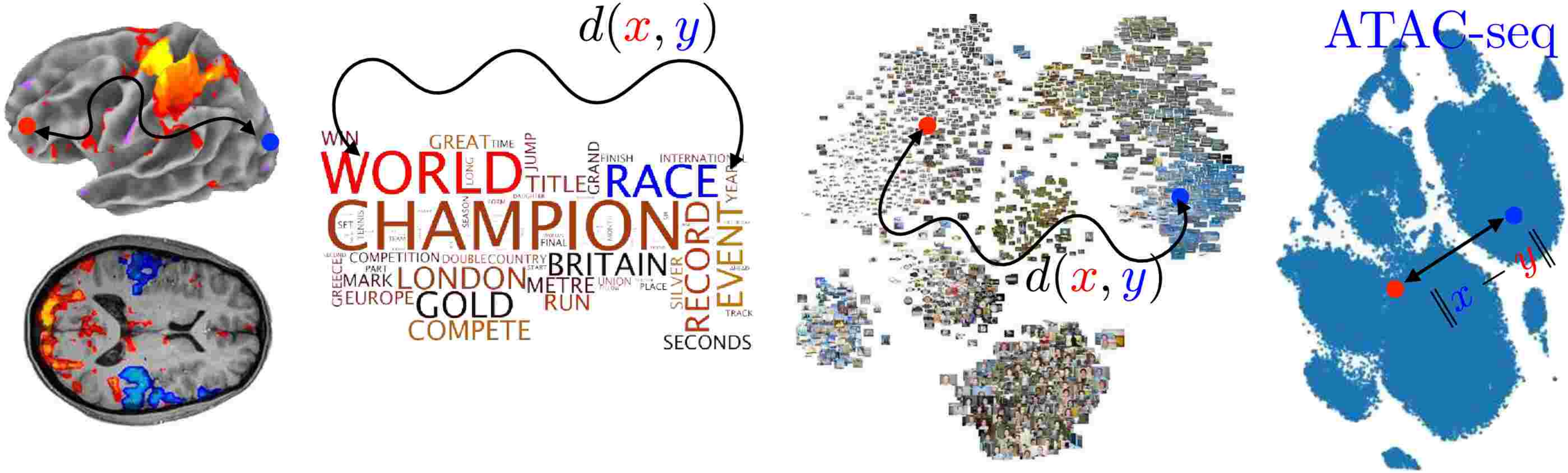
Observations: $\beta \stackrel{\text{def.}}{=} \frac{1}{n}\sum_{i=1}^{n} \delta_{x_i}$

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





Comparing Distributions for Learning



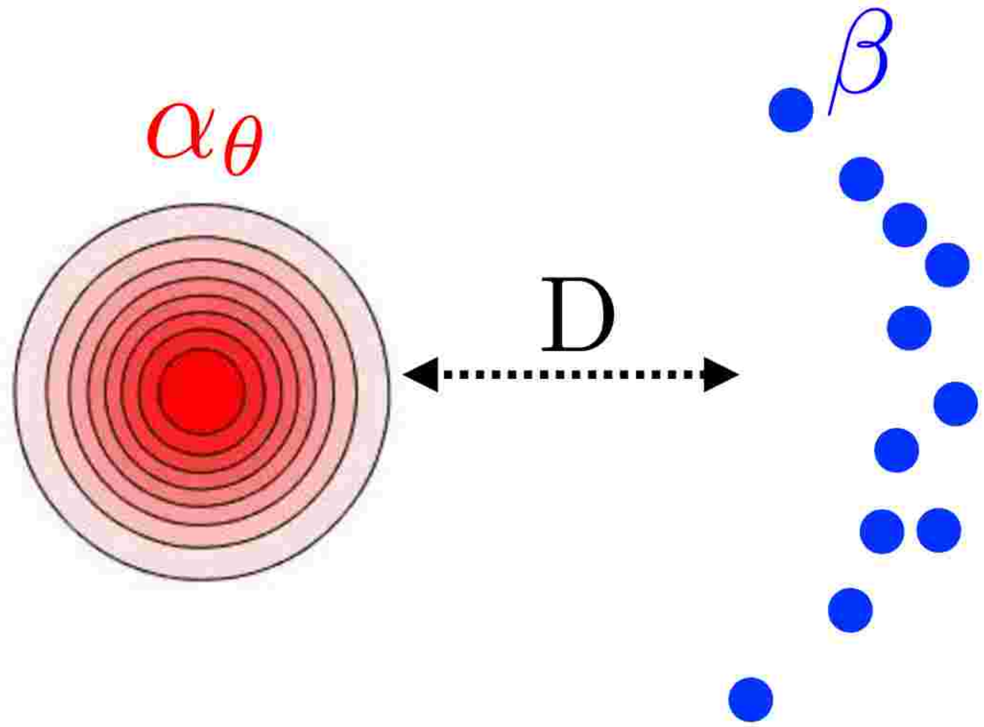
Unsupervised learning

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

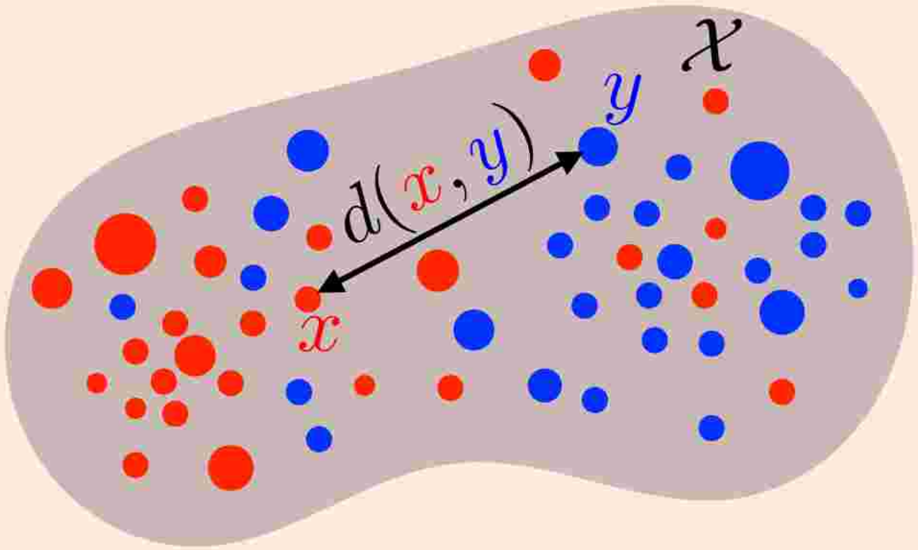
Parametric model: $\theta \mapsto \alpha_\theta$

Density fitting: $\min_\theta \mathrm{D}(\alpha_\theta, \beta)$

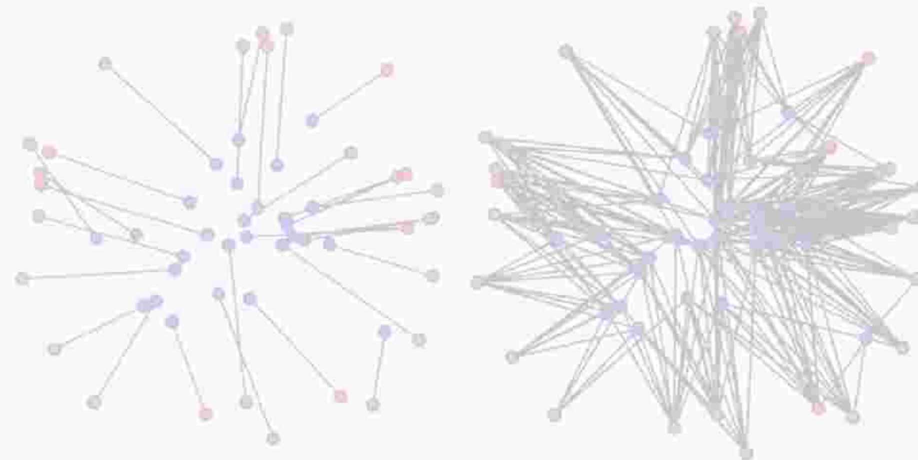
$\rightarrow$ takes into account a metric d.



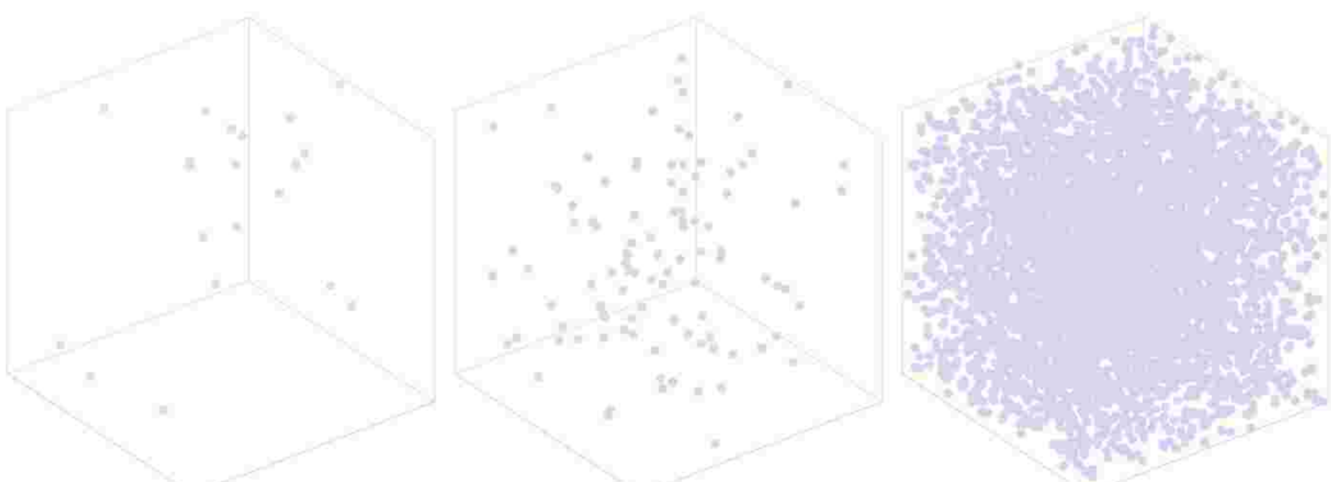
1. Optimal Transport



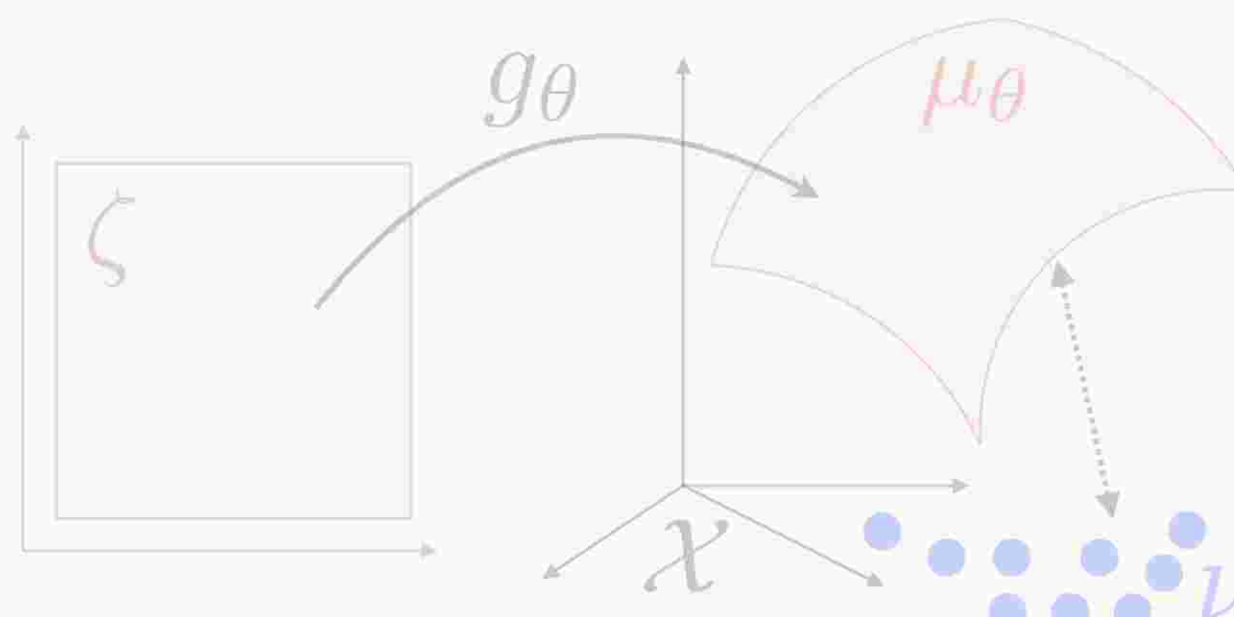
2. Entropic Regularization



3. Sinkhorn Divergences



4. Application to Generative Models

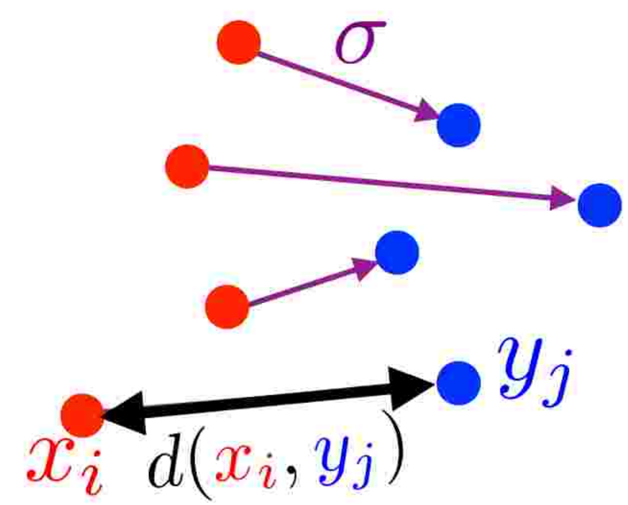


Monge's Problem

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

Permutation:

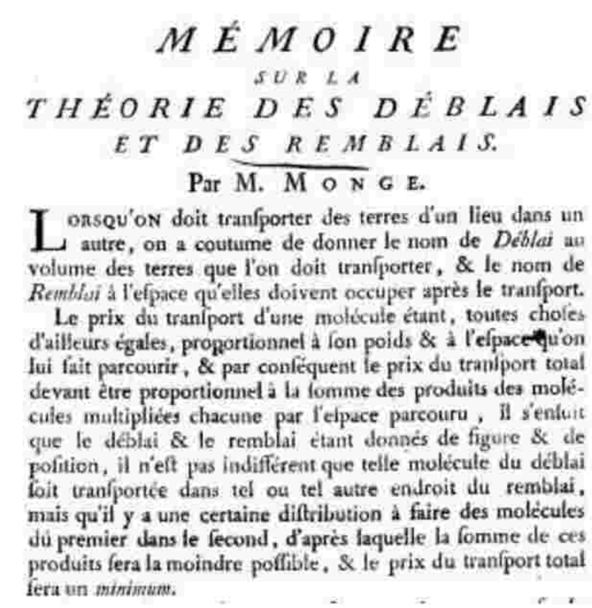
$$\sigma : \{1, \ldots, n\} \to \{1, \ldots, n\}$$



Monge optimal matching:

$$\mathrm{D}(X, Y) = \min_{\sigma} \sum_{i=1}^{n} d(x_i, y_{\sigma(i)})$$

[Monge 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.

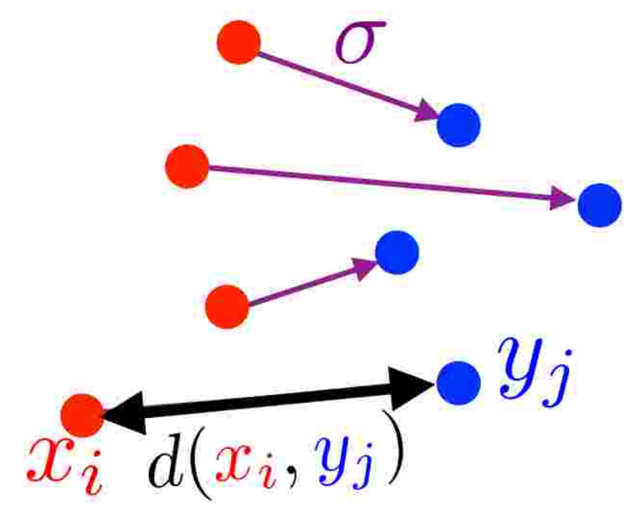
Le prix du tranſport d'une molécule étant, toutes choſes d'ailleurs égales, proportionnel à ſon poids & à l'eſpace qu'on lui fait parcourir, & par conſéquent le prix du tranſport total devant être proportionnel à la ſomme des produits des molécules multipliées chacune par l'eſpace parcouru, il s'enſuit 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 ſera un *minimum*.

Monge's Problem

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

Permutation:

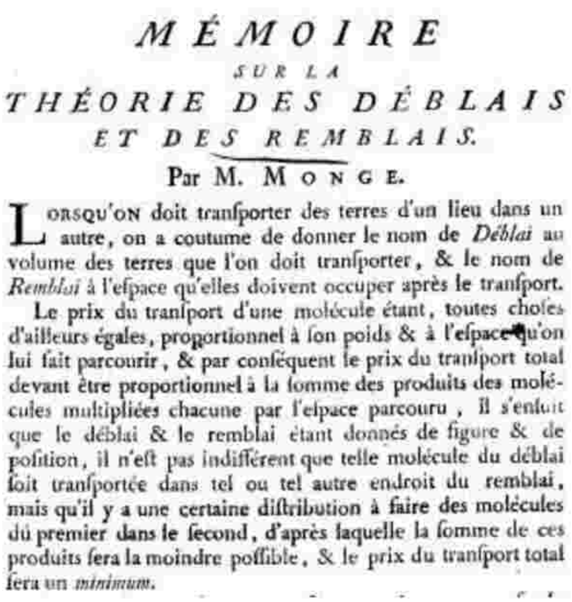
$$\sigma : \{1, \ldots, n\} \to \{1, \ldots, n\}$$



Monge optimal matching:

$$\mathrm{D}(X, Y) = \min_{\sigma} \sum_{i=1}^{n} d(x_i, y_{\sigma(i)})$$

[Monge 1784]



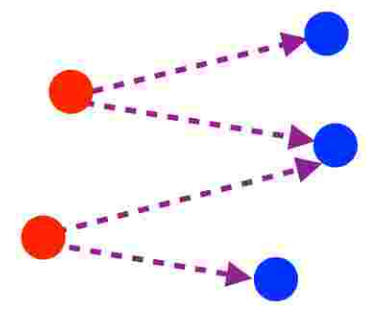
MÉMOIRE
SUR LA
THÉORIE DES DÉBLAIS
ET DES REMBLAIS.
Par M. MONGE.

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'enſuit 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 ſera un *minimum*.

→ Seems intractable: $n!$ possibilities.

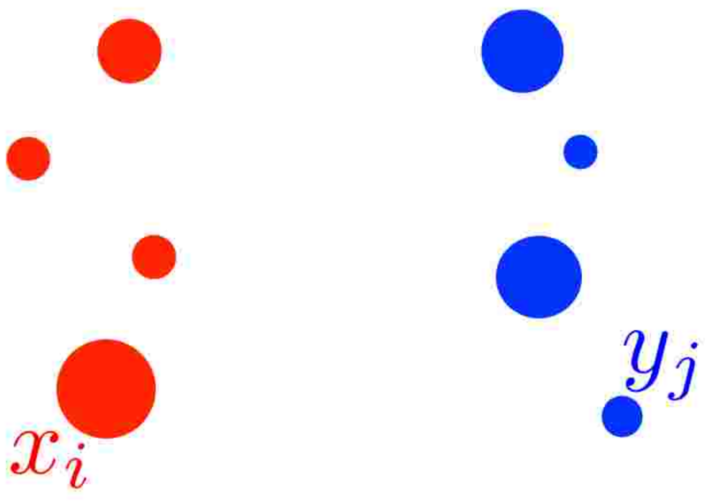
→ Different number of points?



Kantorovitch's Formulation

Discrete distributions: $\alpha = \sum_{i=1}^{n} \mathbf{a}_i \delta_{x_i}$

$\beta = \sum_{j=1}^{m} \mathbf{b}_j \delta_{y_j}$



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

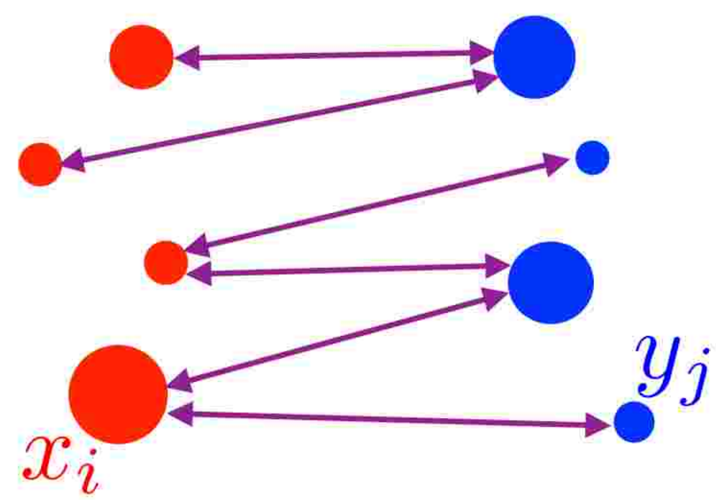
Weights $\mathbf{a}_i \geqslant 0$, $\mathbf{b}_j \geqslant 0$.

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

Kantorovitch's Formulation

Discrete distributions: $\alpha = \sum_{i=1}^{n} \mathbf{a}_i \delta_{x_i}$

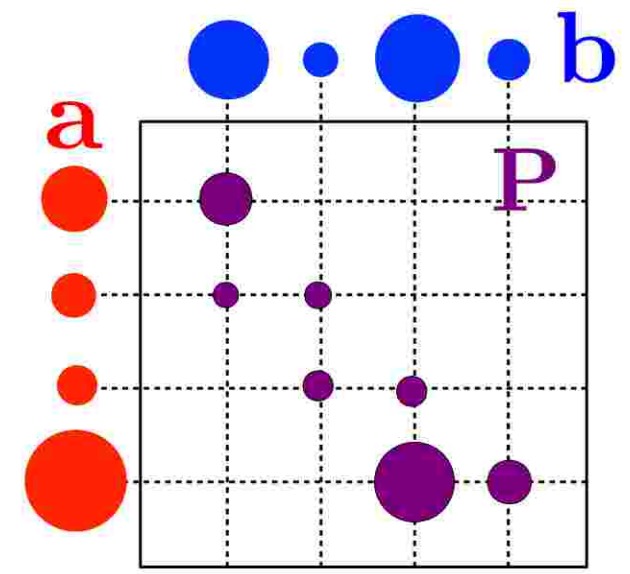
$\beta = \sum_{j=1}^{m} \mathbf{b}_j \delta_{y_j}$



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

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

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



Couplings: $\sum_j \mathbf{P}_{i,j} = \mathbf{a}_i$ $\sum_i \mathbf{P}_{i,j} = \mathbf{b}_j$

$$\mathbf{P} \geqslant 0, \; \mathbf{P}\mathbb{1}_m = \mathbf{a}, \; \mathbf{P}^\top \mathbb{1}_n = \mathbf{b}$$

Kantorovitch's Formulation

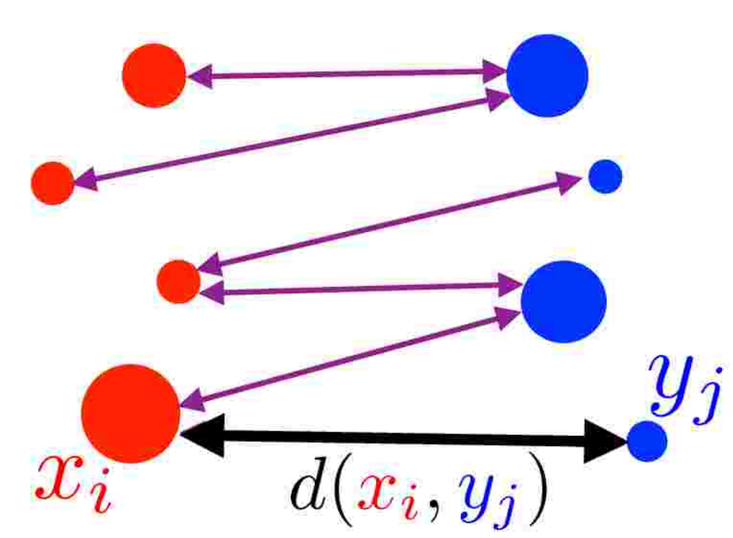
Discrete distributions: $\alpha = \sum_{i=1}^{n} \mathbf{a}_i \delta_{x_i}$

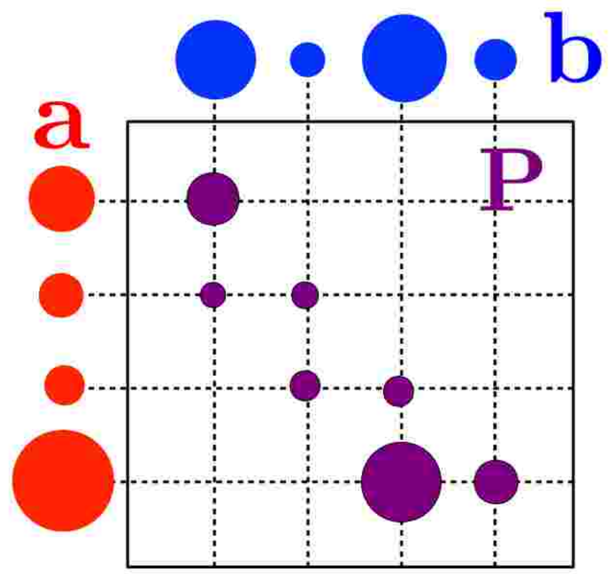
$\beta = \sum_{j=1}^{m} \mathbf{b}_j \delta_{y_j}$

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

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

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





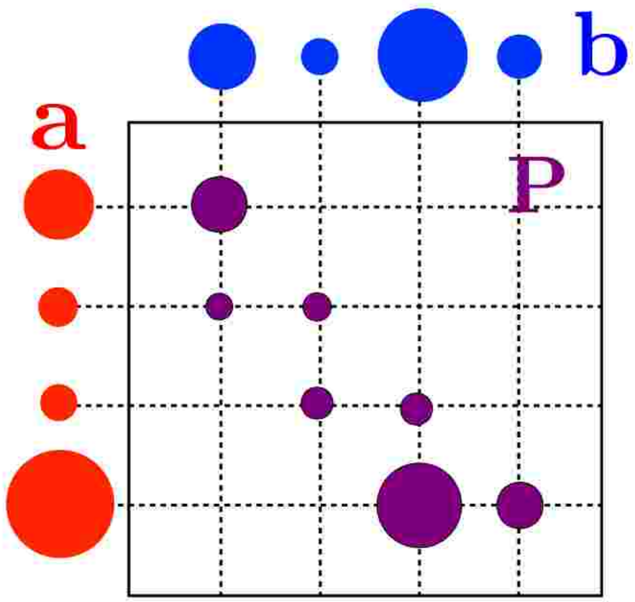
Couplings: $\sum_j \mathbf{P}_{i,j} = \mathbf{a}_i$ $\quad \sum_i \mathbf{P}_{i,j} = \mathbf{b}_j$

[Kantorovich 1942]

$$\min_{\mathbf{P}} \left\{ \sum_{i,j} d(x_i, y_j)^p \mathbf{P}_{i,j} \; ; \; \mathbf{P} \geqslant 0, \; \mathbf{P}\mathbb{1}_m = \mathbf{a}, \mathbf{P}^\top \mathbb{1}_n = \mathbf{b} \right\}$$

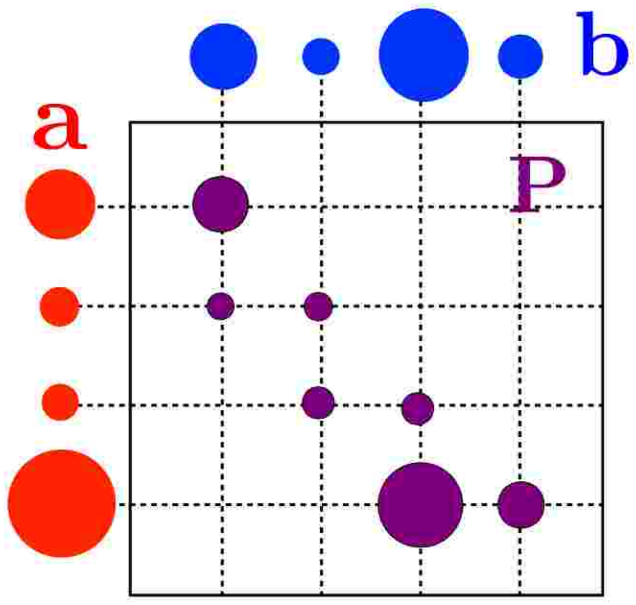
Optimal Transport Distances

$$\mathrm{W}_p(\alpha, \beta) \overset{\text{def.}}{=} \left(\min_{\mathbf{P}\mathbb{1}=\mathbf{a}, \mathbf{P}^\top \mathbb{1}=\mathbf{b}} \sum_{i,j} d(x_i, y_j)^p \mathbf{P}_{i,j} \right)^{\frac{1}{p}}$$



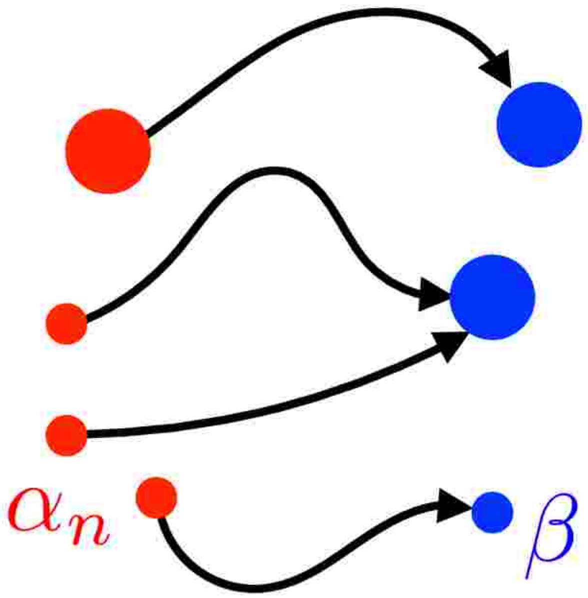
Optimal Transport Distances

$$\mathrm{W}_p(\alpha, \beta) \overset{\text{def.}}{=} \Big(\min_{\mathbf{P}\mathbb{1}=\mathbf{a}, \mathbf{P}^\top \mathbb{1}=\mathbf{b}} \sum_{i,j} d(x_i, y_j)^p \mathbf{P}_{i,j} \Big)^{\frac{1}{p}}$$



Convergence in law: $\alpha_n \rightharpoonup \beta$

$$\Leftrightarrow \quad \forall\, f \in \mathcal{C}(\mathcal{X}), \ \int_{\mathcal{X}} f \mathrm{d}\alpha_n \to \int_{\mathcal{X}} f \mathrm{d}\beta$$

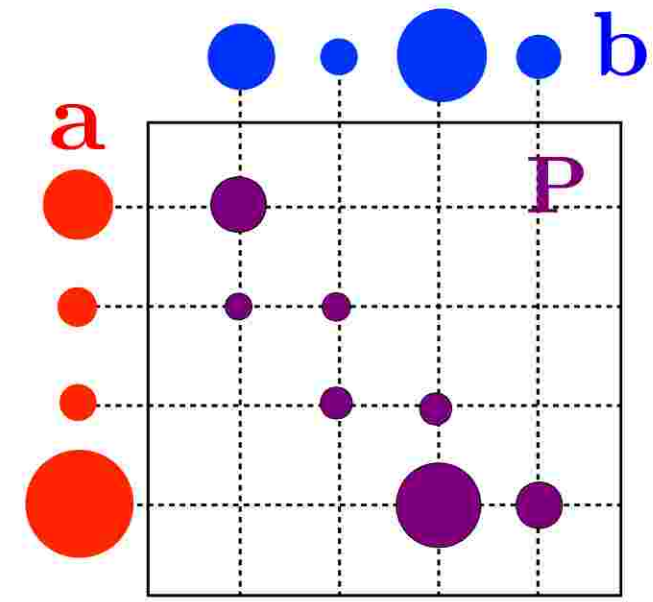


Theorem: W_p is a distance and

$$\alpha_n \rightharpoonup \beta \ \Leftrightarrow \ \mathrm{W}_p(\alpha_n, \beta) \to 0$$

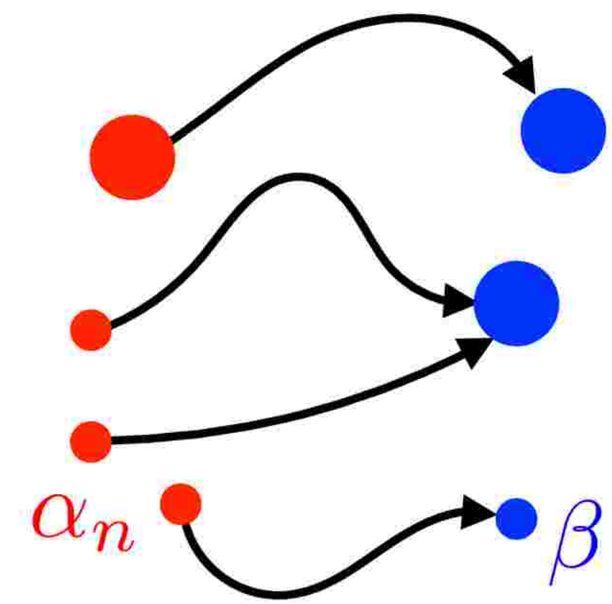
Optimal Transport Distances

$$\mathrm{W}_p(\alpha,\beta) \overset{\text{def.}}{=} \Big(\min_{\mathbf{P}\mathbb{1}=\mathbf{a},\mathbf{P}^\top\mathbb{1}=\mathbf{b}} \sum_{i,j} d(x_i,y_j)^p \mathbf{P}_{i,j} \Big)^{\frac{1}{p}}$$



Convergence in law: $\alpha_n \rightharpoonup \beta$

$$\Leftrightarrow \quad \forall\, f \in \mathcal{C}(\mathcal{X}),\ \int_{\mathcal{X}} f \mathrm{d}\alpha_n \to \int_{\mathcal{X}} f \mathrm{d}\beta$$

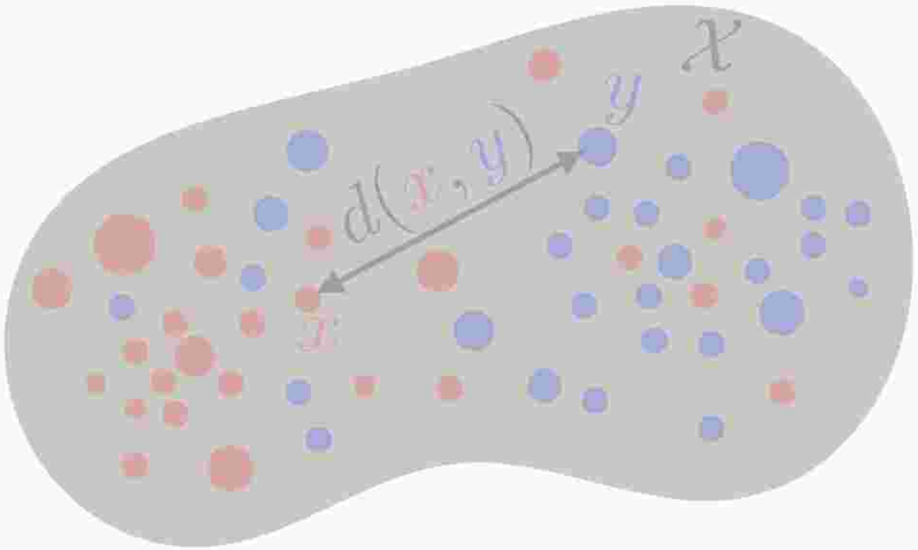


Theorem: W_p is a distance and

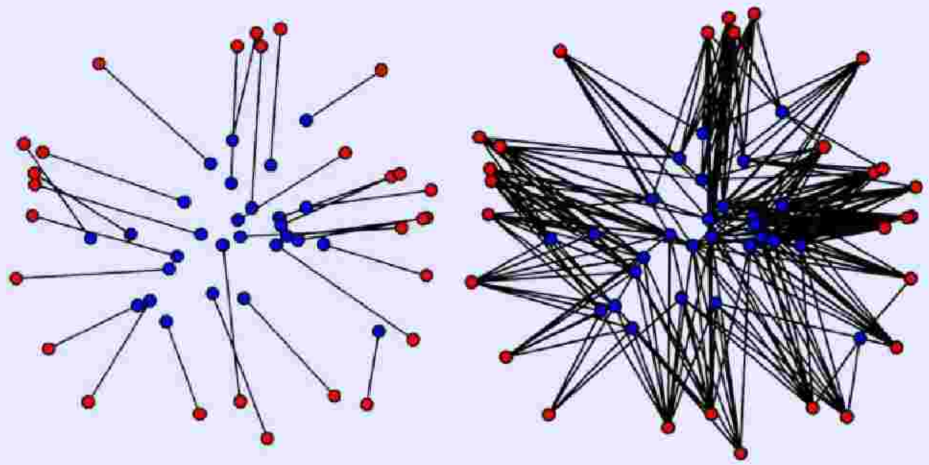
$$\alpha_n \rightharpoonup \beta \;\Leftrightarrow\; \mathrm{W}_p(\alpha_n,\beta) \to 0$$

$$\|\delta_{x_n} - \delta_x\|_{\mathrm{TV}} = 2 \quad \textbf{vs.} \quad \mathrm{W}_p(\delta_{x_n},\delta_x) = d(x_n,x)$$

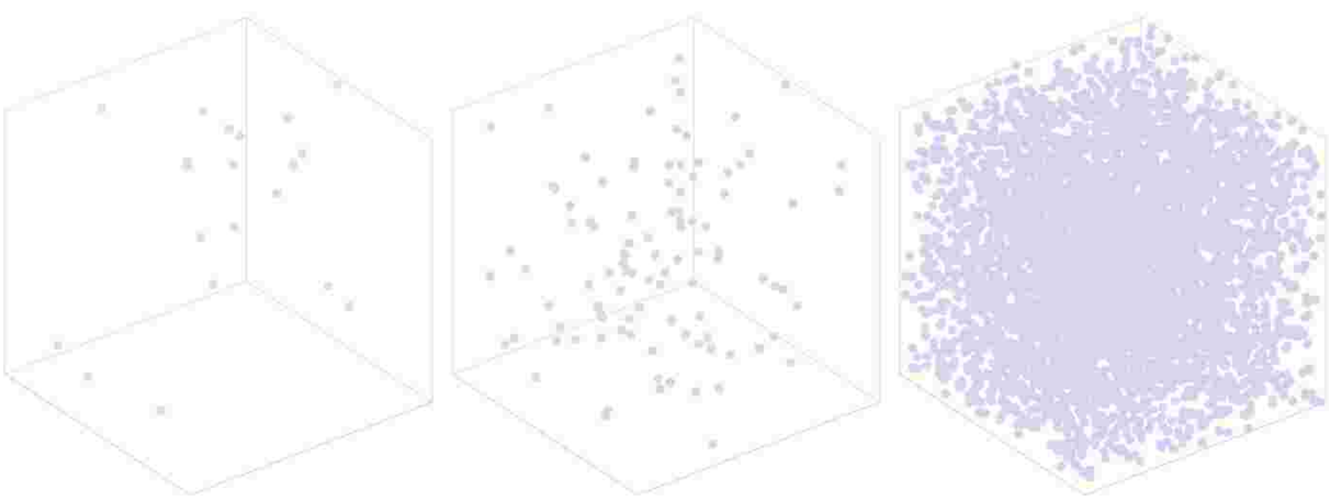
1. Optimal Transport



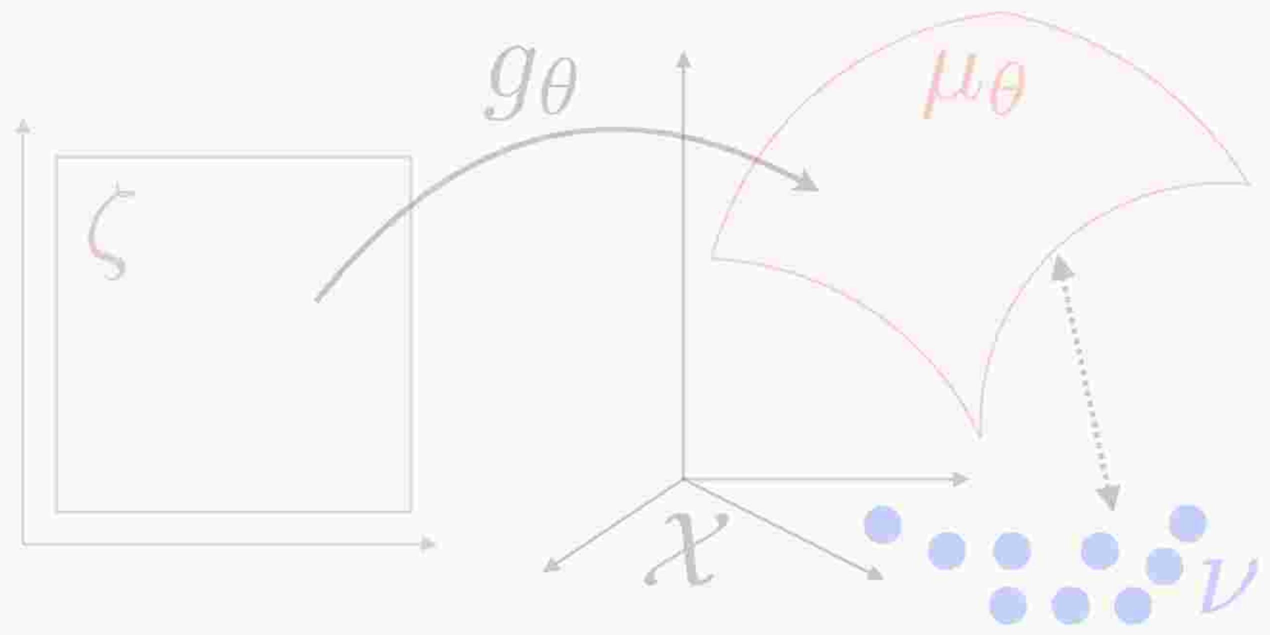
2. Entropic Regularization



3. Sinkhorn Divergences



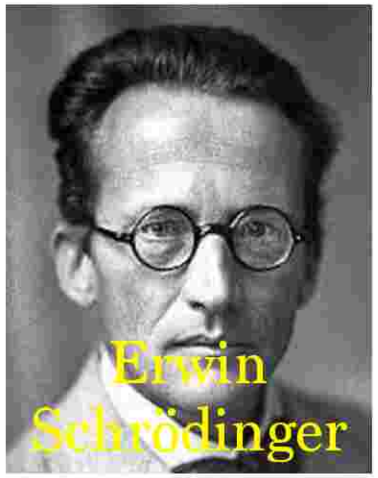
4. Application to Generative Models



Entropic Regularization

Schrödinger's problem: [1931]

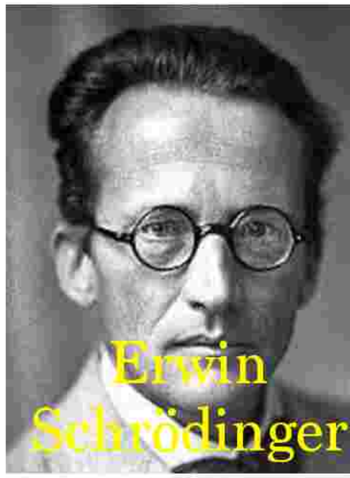
$$\min_{\mathbf{P}\mathbb{1}=\mathbf{a},\mathbf{P}^\top\mathbb{1}=\mathbf{b}} \sum_{i,j} d(x_i, y_j)^p \mathbf{P}_{i,j} + \varepsilon \mathbf{P}_{i,j} \log(\mathbf{P}_{i,j})$$

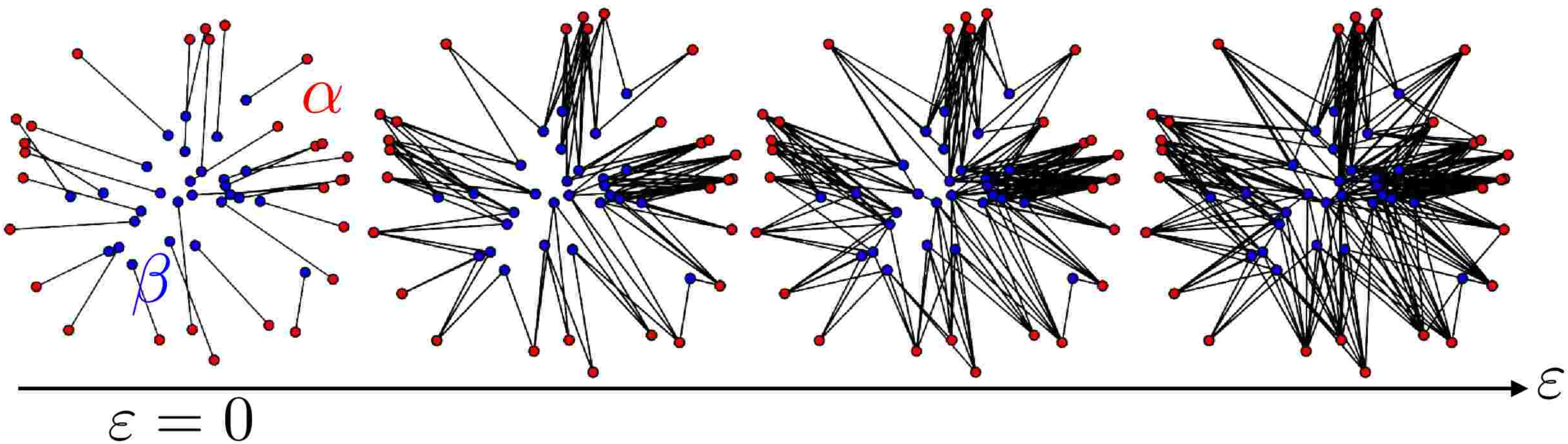


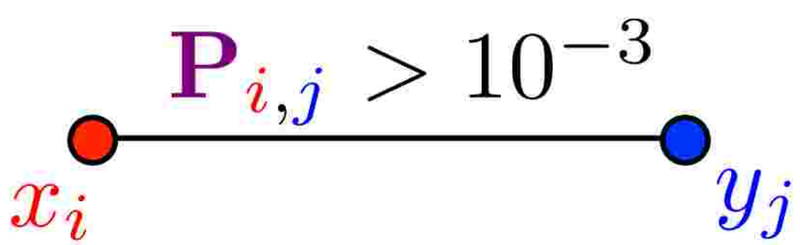
Entropic Regularization

Schrödinger's problem: [1931]

$$\min_{\mathbf{P}\mathbb{1}=\mathbf{a},\,\mathbf{P}^\top\mathbb{1}=\mathbf{b}} \sum_{i,j} d(x_i, y_j)^p \mathbf{P}_{i,j} + \varepsilon \mathbf{P}_{i,j} \log(\mathbf{P}_{i,j})$$







Sinkhorn's Algorithm

$$\min_{\mathbf{P}} \left\{ \sum_{i,j} d(x_i, y_j)^p \mathbf{P}_{i,j} + \varepsilon \mathbf{P}_{i,j} \log(\mathbf{P}_{i,j}) \; ; \; \mathbf{P}\mathbb{1} = \mathbf{a}, \mathbf{P}^\top \mathbb{1} = \mathbf{b} \right\}$$

Proposition:

$$\mathbf{P} \text{ solution} \Leftrightarrow \begin{cases} \mathbf{P}\mathbb{1} = \mathbf{a}, \mathbf{P}^\top \mathbb{1} = \mathbf{b} \text{ and} \\ \exists \mathbf{u}, \mathbf{v}, \; \mathbf{P}_{i,j} = \mathbf{u}_i \, \mathbf{K}_{i,j} \, \mathbf{v}_j \end{cases}$$

$$\mathbf{K}_{i,j} \overset{\text{def.}}{=} e^{-\frac{d(x_i, y_j)^p}{\varepsilon}}$$

Sinkhorn's Algorithm

$$\min_{\mathbf{P}} \left\{ \sum_{i,j} d(x_i, y_j)^p \mathbf{P}_{i,j} + \varepsilon \mathbf{P}_{i,j} \log(\mathbf{P}_{i,j}) \; ; \; \mathbf{P}\mathbb{1} = \mathbf{a}, \mathbf{P}^\top \mathbb{1} = \mathbf{b} \right\}$$

Proposition:

$$\mathbf{P} \text{ solution} \Leftrightarrow \begin{cases} \mathbf{P}\mathbb{1} = \mathbf{a}, \mathbf{P}^\top \mathbb{1} = \mathbf{b} \text{ and} \\ \exists \mathbf{u}, \mathbf{v}, \; \mathbf{P}_{i,j} = \mathbf{u}_i \, \mathbf{K}_{i,j} \, \mathbf{v}_j \end{cases} \qquad \mathbf{K}_{i,j} \stackrel{\text{def.}}{=} e^{-\frac{d(x_i, y_j)^p}{\varepsilon}}$$

$$\mathbf{P} = \operatorname{diag}(\mathbf{u}) \mathbf{K} \operatorname{diag}(\mathbf{v}) \implies \mathbf{a} = \mathbf{P}\mathbb{1} = \operatorname{diag}(\mathbf{u})(\mathbf{K}\mathbf{v}) = \mathbf{u} \odot (\mathbf{K}\mathbf{v})$$

Row constraint: $\mathbf{u} \odot (\mathbf{K}\mathbf{v}) = \mathbf{a}$

Col. constraint: $\mathbf{v} \odot (\mathbf{K}^\top \mathbf{u}) = \mathbf{b}$

Sinkhorn's Algorithm

$$\min_{\mathbf{P}} \left\{ \sum_{i,j} d(x_i, y_j)^p \mathbf{P}_{i,j} + \varepsilon \mathbf{P}_{i,j} \log(\mathbf{P}_{i,j}) \; ; \; \mathbf{P}\mathbb{1} = \mathbf{a}, \mathbf{P}^\top \mathbb{1} = \mathbf{b} \right\}$$

Proposition: $\mathbf{P}$ solution $\Leftrightarrow \begin{cases} \mathbf{P}\mathbb{1} = \mathbf{a}, \mathbf{P}^\top \mathbb{1} = \mathbf{b} \text{ and} \\ \exists \mathbf{u}, \mathbf{v}, \; \mathbf{P}_{i,j} = \mathbf{u}_i \, \mathbf{K}_{i,j} \, \mathbf{v}_j \end{cases}$ $\quad \mathbf{K}_{i,j} \stackrel{\text{def.}}{=} e^{-\frac{d(x_i, y_j)^p}{\varepsilon}}$

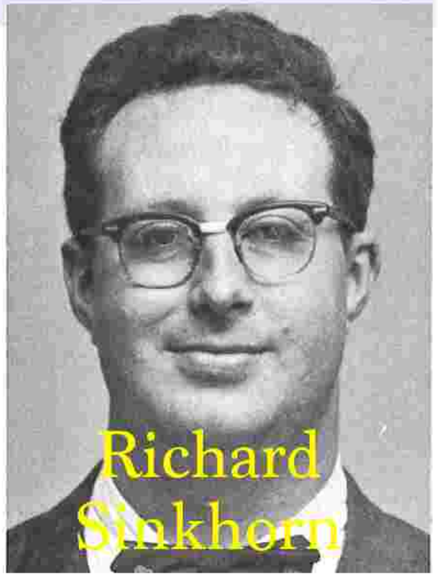
$$\mathbf{P} = \operatorname{diag}(\mathbf{u}) \mathbf{K} \operatorname{diag}(\mathbf{v}) \implies \mathbf{a} = \mathbf{P}\mathbb{1} = \operatorname{diag}(\mathbf{u})(\mathbf{K}\mathbf{v}) = \mathbf{u} \odot (\mathbf{K}\mathbf{v})$$

Row constraint: $\mathbf{u} \odot (\mathbf{K}\mathbf{v}) = \mathbf{a}$

Col. constraint: $\mathbf{v} \odot (\mathbf{K}^\top \mathbf{u}) = \mathbf{b}$

Sinkhorn iterations: $\mathbf{u} \leftarrow \frac{\mathbf{a}}{\mathbf{K}\mathbf{v}}$ $\quad \mathbf{v} \leftarrow \frac{\mathbf{b}}{\mathbf{K}^\top \mathbf{u}}$

Theorem: [Sinkhorn 1964] $(\mathbf{u}, \mathbf{v})$ converges.



Sinkhorn's Algorithm

$$\min_{\mathbf{P}} \left\{ \sum_{i,j} d(x_i, y_j)^p \mathbf{P}_{i,j} + \varepsilon \mathbf{P}_{i,j} \log(\mathbf{P}_{i,j}) \; ; \; \mathbf{P}\mathbb{1} = \mathbf{a}, \mathbf{P}^\top \mathbb{1} = \mathbf{b} \right\}$$

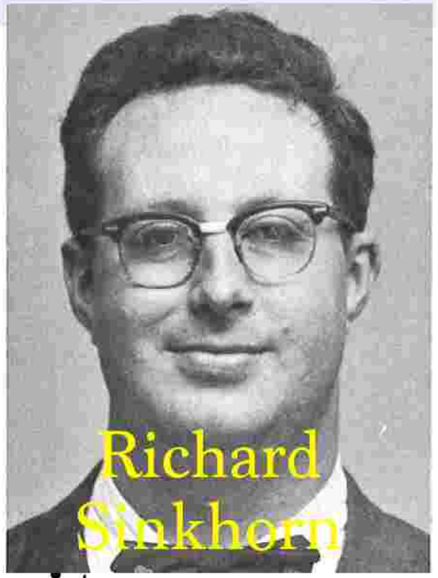
Proposition: $\mathbf{P}$ solution $\Leftrightarrow \begin{cases} \mathbf{P}\mathbb{1} = \mathbf{a}, \mathbf{P}^\top \mathbb{1} = \mathbf{b} \text{ and} \\ \exists \mathbf{u}, \mathbf{v}, \; \mathbf{P}_{i,j} = \mathbf{u}_i \mathbf{K}_{i,j} \mathbf{v}_j \end{cases}$ $\quad \mathbf{K}_{i,j} \overset{\text{def.}}{=} e^{-\frac{d(x_i, y_j)^p}{\varepsilon}}$

$$\mathbf{P} = \text{diag}(\mathbf{u}) \mathbf{K} \, \text{diag}(\mathbf{v}) \implies \mathbf{a} = \mathbf{P}\mathbb{1} = \text{diag}(\mathbf{u})(\mathbf{K}\mathbf{v}) = \mathbf{u} \odot (\mathbf{K}\mathbf{v})$$

Row constraint: $\mathbf{u} \odot (\mathbf{K}\mathbf{v}) = \mathbf{a}$

Col. constraint: $\mathbf{v} \odot (\mathbf{K}^\top \mathbf{u}) = \mathbf{b}$

Sinkhorn iterations: $\mathbf{u} \leftarrow \frac{\mathbf{a}}{\mathbf{K}\mathbf{v}}$ $\quad \mathbf{v} \leftarrow \frac{\mathbf{b}}{\mathbf{K}^\top \mathbf{u}}$



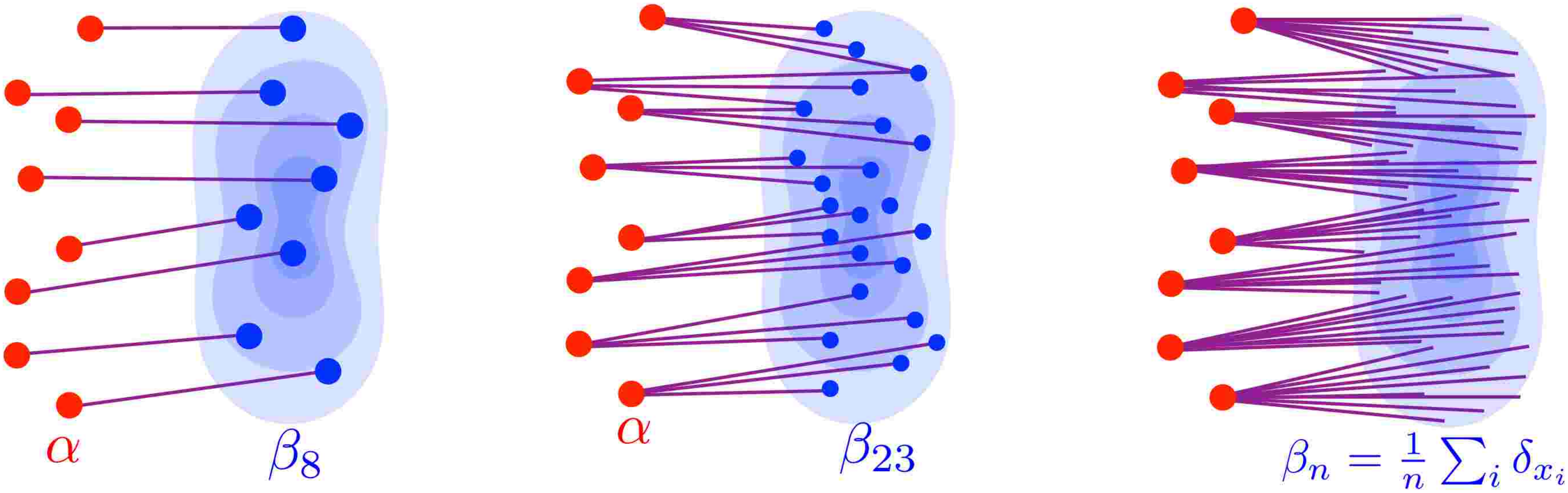
Theorem: [Sinkhorn 1964] $\quad (\mathbf{u}, \mathbf{v})$ converges.

Matrix/vector multiplications: $\rightarrow O(n^2/\varepsilon^2)$ complexity.

$\rightarrow$ Parallelizable on GPUs.

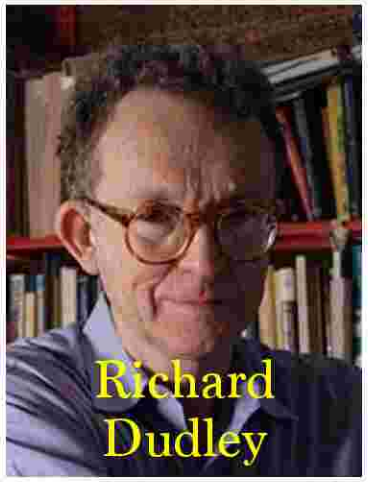
$\rightarrow$ Convolution on regular grids, separable kernels.

The Curse of Dimensionality

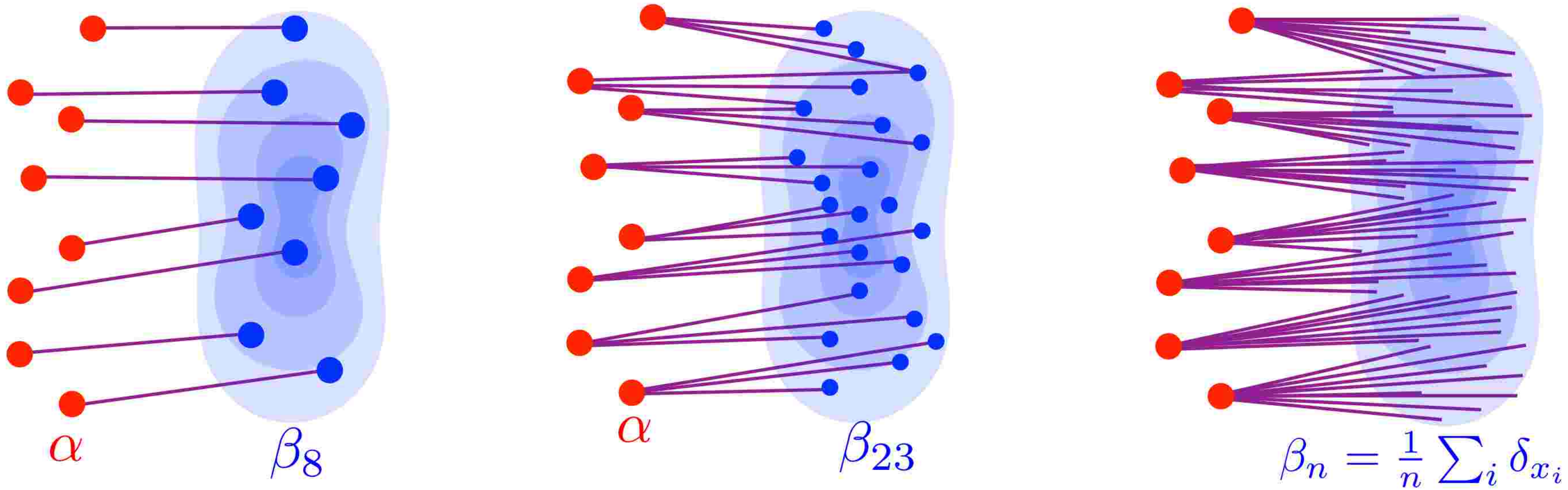


Theorem: $\mathbb{E}|\mathrm{W}_p(\alpha, \beta_n) - \mathrm{W}_p(\alpha, \beta_\infty)| \leqslant \delta$

[Dudley 1968] requires $n \sim (1/\delta)^{\text{dimension}}$

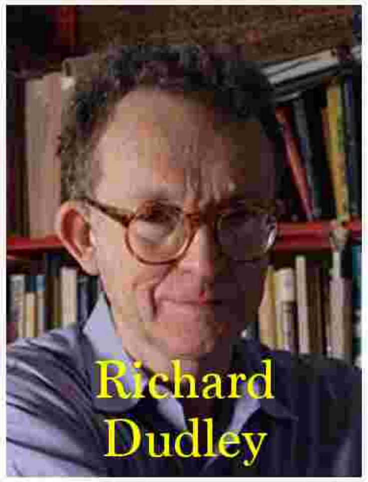


The Curse of Dimensionality



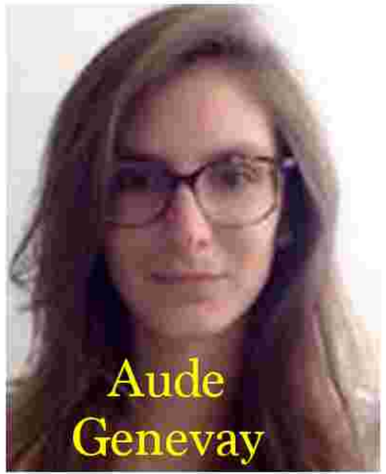
Theorem: $\mathbb{E}|W_p(\alpha, \beta_n) - W_p(\alpha, \beta_\infty)| \leqslant \delta$

[Dudley 1968] requires $n \sim (1/\delta)^{\text{dimension}}$



Theorem: $\mathbb{E}|W_p^\varepsilon(\alpha, \beta_n) - W_p^\varepsilon(\alpha, \beta_\infty)| \leqslant \delta$

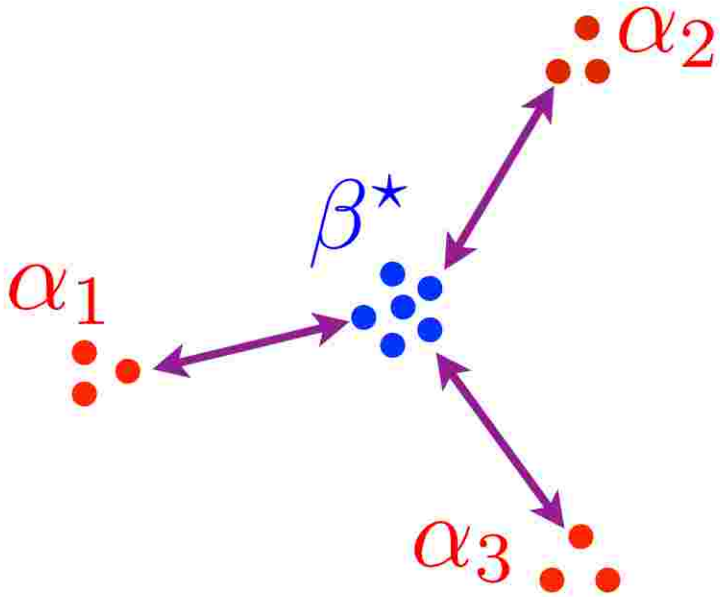
[Genevay 2019] requires $n \sim (1/\varepsilon)^{\text{dimension}} \times (1/\delta)^2$



Wasserstein Barycenters

Barycenters of measures $(\alpha_s)_s$: $\sum_s \lambda_s = 1$

$$\beta^\star \in \underset{\beta}{\operatorname{argmin}} \sum_s \lambda_s \mathrm{W}_p^p(\alpha_s, \beta)$$

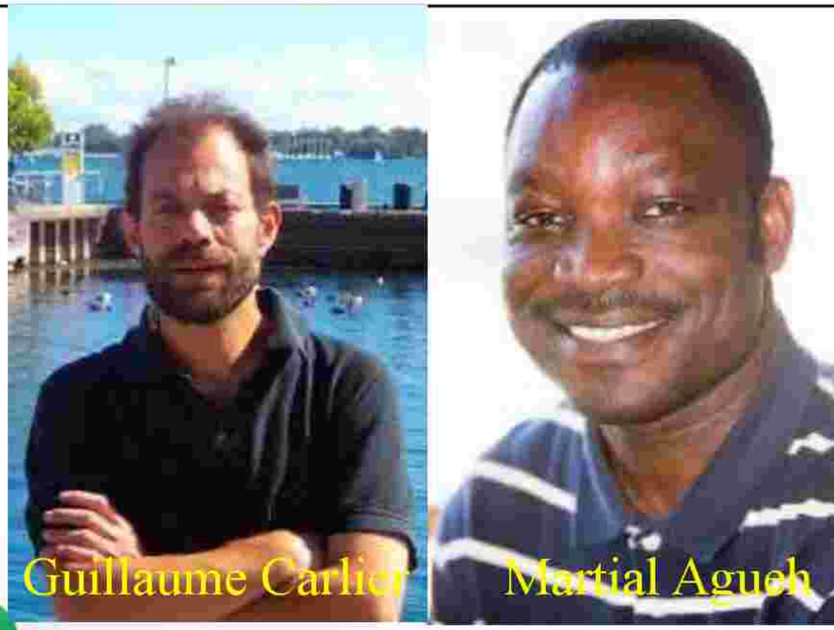


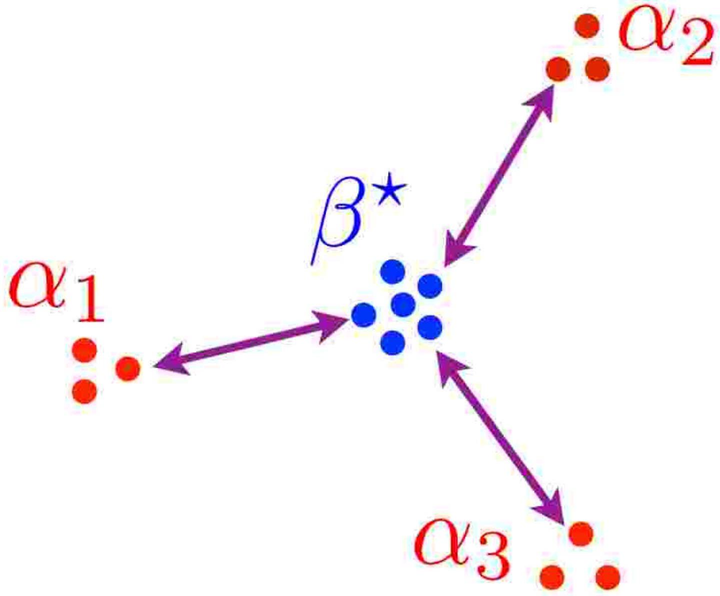
[Solomon et al, SIGGRAPH 2015]

Wasserstein Barycenters

Barycenters of measures $(\alpha_s)_s$: $\sum_s \lambda_s = 1$

$$\beta^\star \in \underset{\beta}{\operatorname{argmin}} \sum_s \lambda_s W_p^p(\alpha_s, \beta)$$





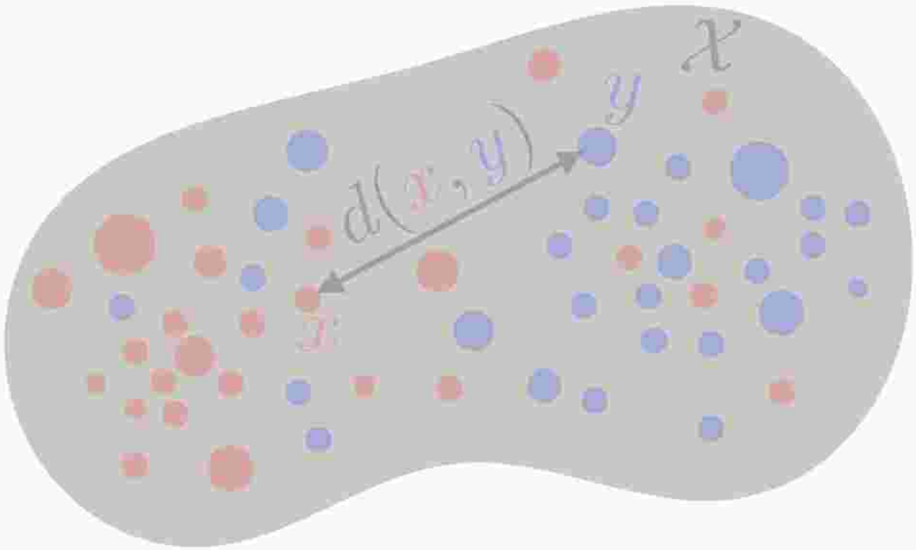
[Solomon et al, SIGGRAPH 2015]

Sinkhorn's algorithm:

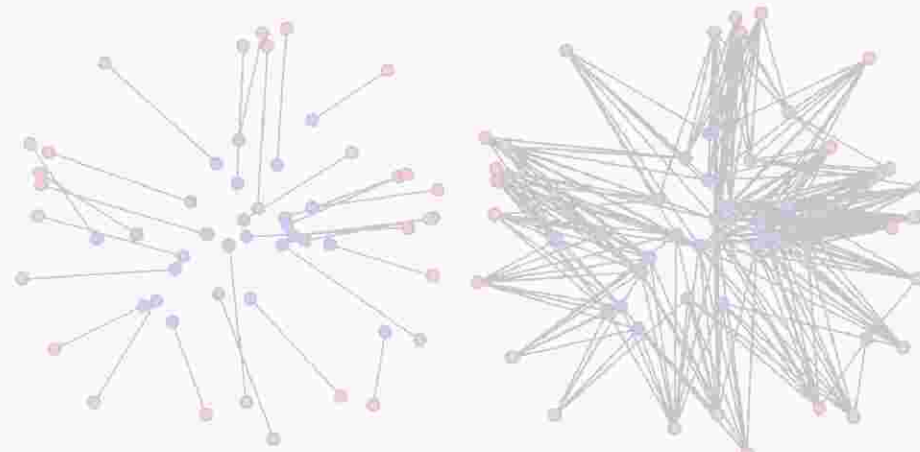
$$\left(\mathbf{u}_s \leftarrow \frac{\mathbf{a}_s}{\mathbf{K}\mathbf{v}_s}\right)_s \qquad \left(\mathbf{v}_s \leftarrow \frac{\mathbf{b}}{\mathbf{K}^\top \mathbf{u}_s}\right)_s$$

$$\mathbf{b} \leftarrow \prod_s (\mathbf{K}^\top \mathbf{u}_s)^{\lambda_s}$$

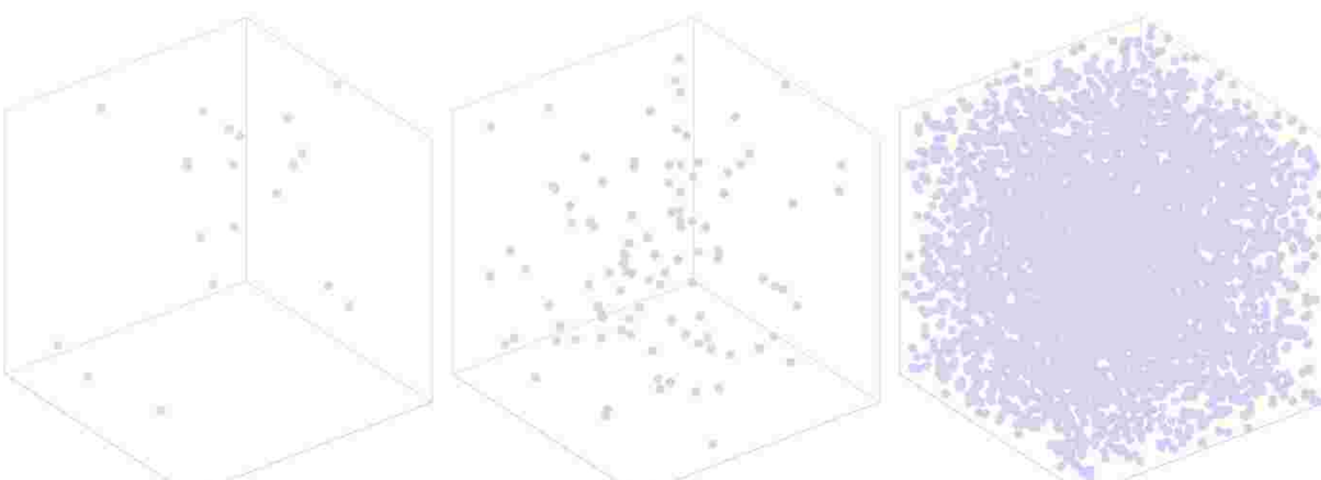
1. Optimal Transport



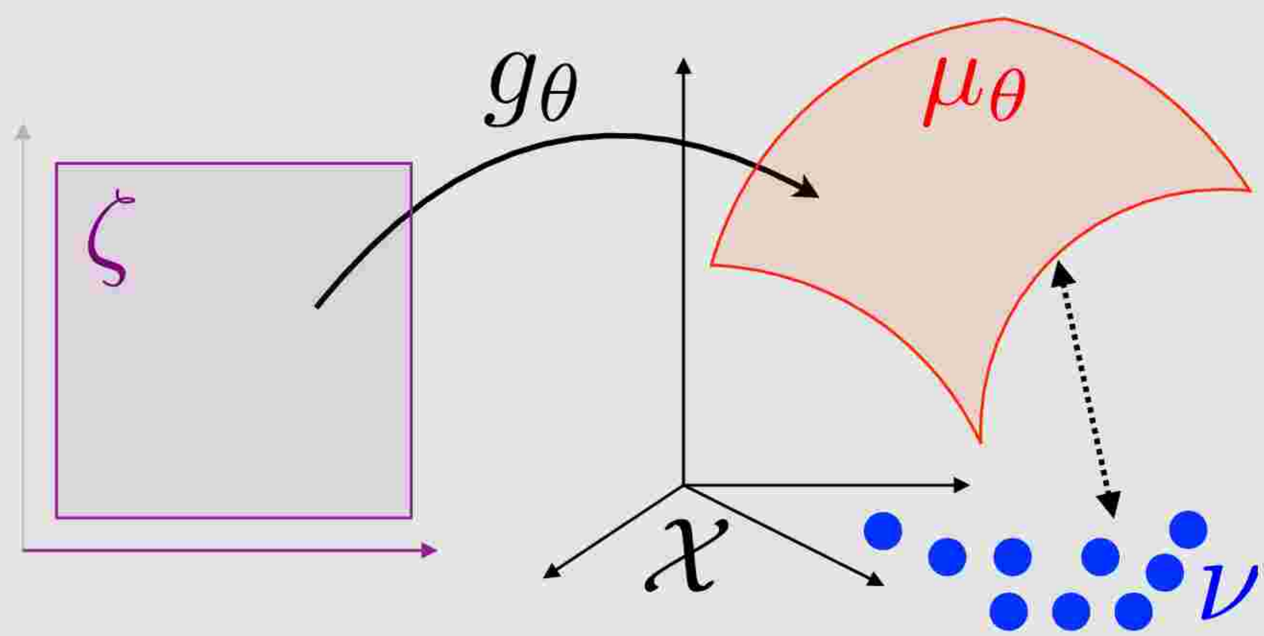
2. Entropic Regularization



3. Sinkhorn Divergences



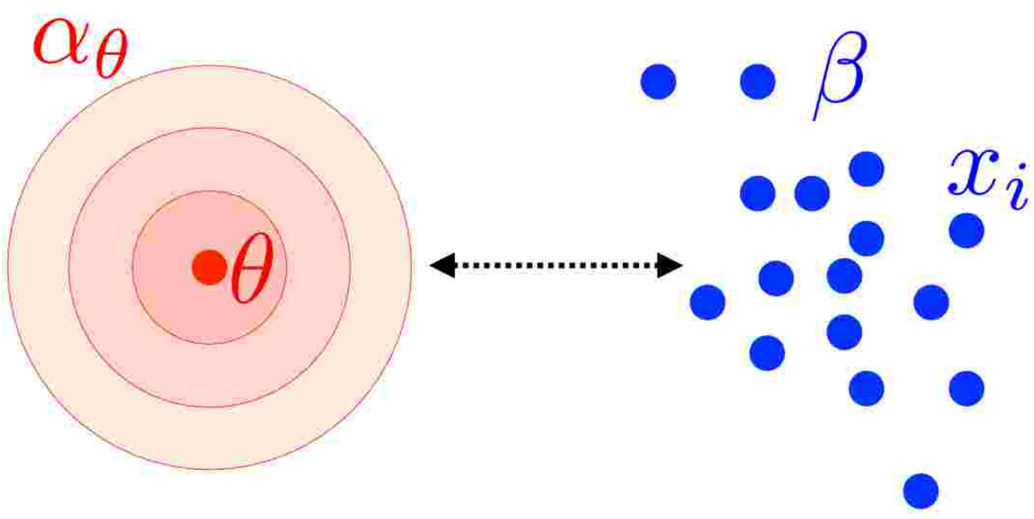
4. Application to Generative Models



Density Fitting and Generative Models

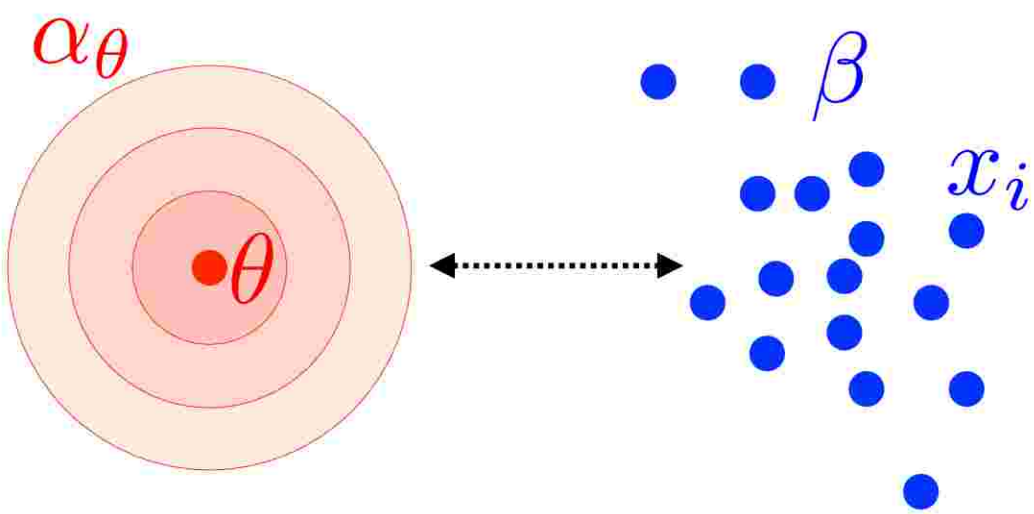
Observations: $\beta \stackrel{\text{def.}}{=} \frac{1}{n} \sum_{i=1}^{n} \delta_{x_i}$

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: $\mathrm{d}\alpha_\theta(x) = \rho_\theta(x)\mathrm{d}x$

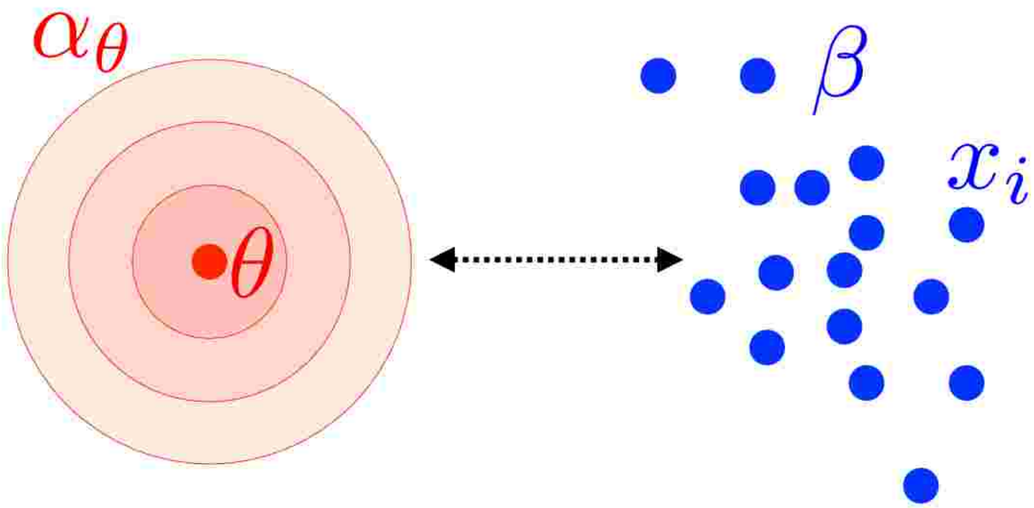
$$\min_\theta \; -\sum_i \log(\rho_\theta(x_i)) \stackrel{n\to+\infty}{\longrightarrow} \mathrm{KL}(\beta|\alpha_\theta)$$

Maximum likelihood (MLE)

Density Fitting and Generative Models

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

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



Density fitting: $\mathrm{d}\alpha_\theta(x) = \rho_\theta(x)\mathrm{d}x$

$$\min_\theta \; -\sum_i \log(\rho_\theta(x_i)) \overset{n\to+\infty}{\longrightarrow} \mathrm{KL}(\beta|\alpha_\theta)$$

Maximum likelihood (MLE)

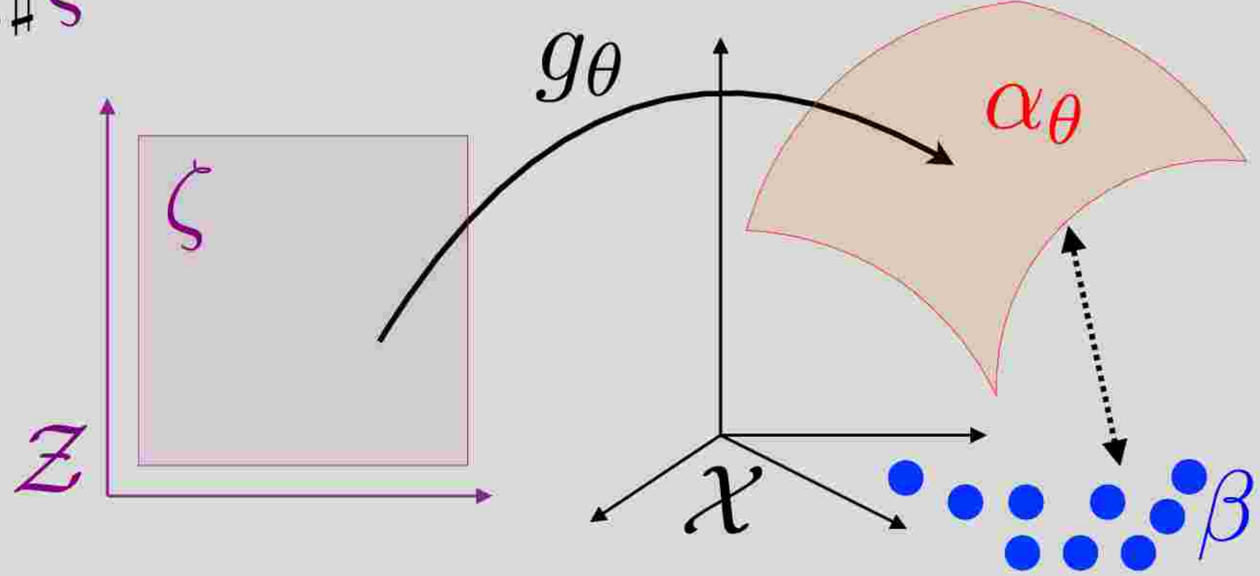
Generative model fit: $\alpha_\theta = g_{\theta,\sharp}\zeta$

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

$\rightarrow$ MLE undefined.

$\rightarrow$ Need a weaker metric.

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

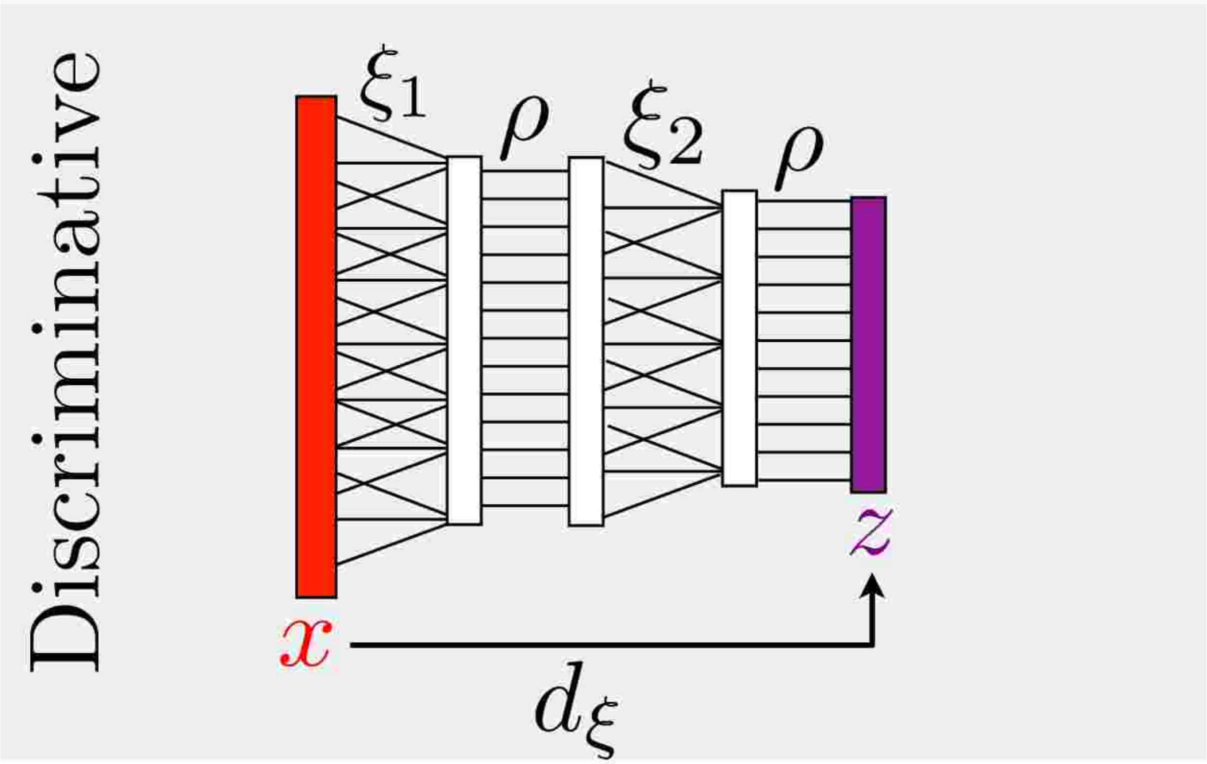


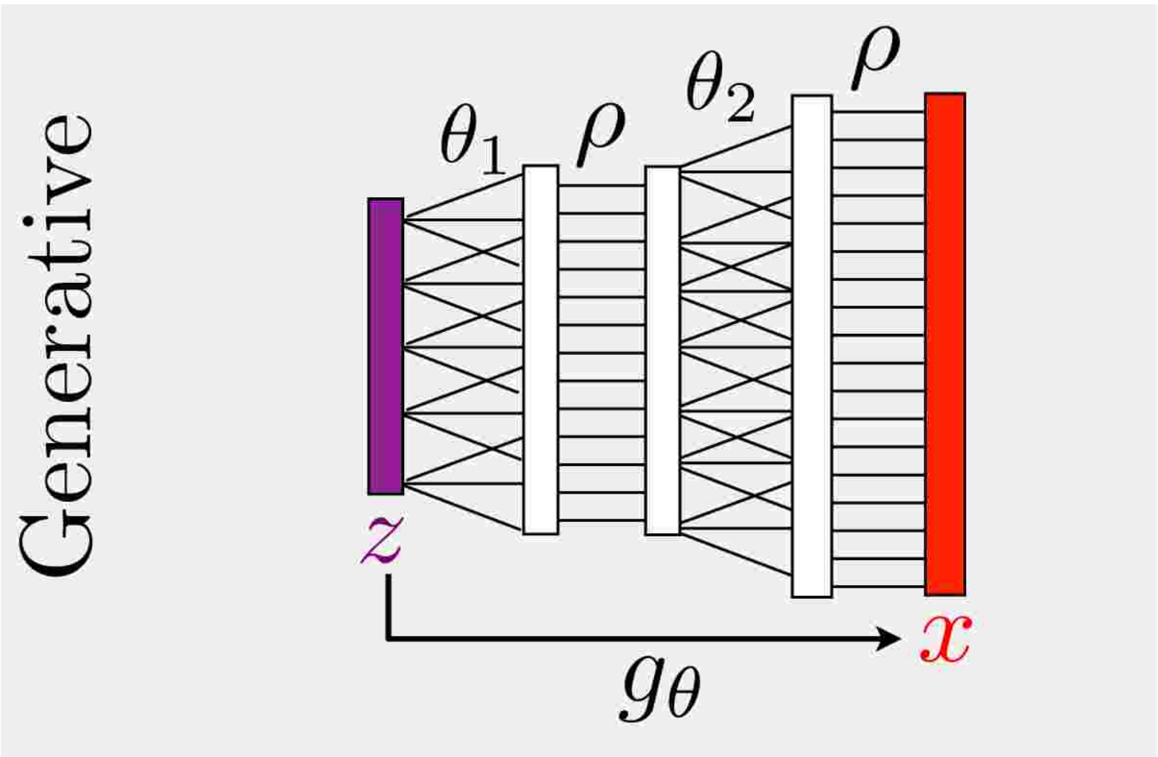
Deep Discriminative vs Generative Models

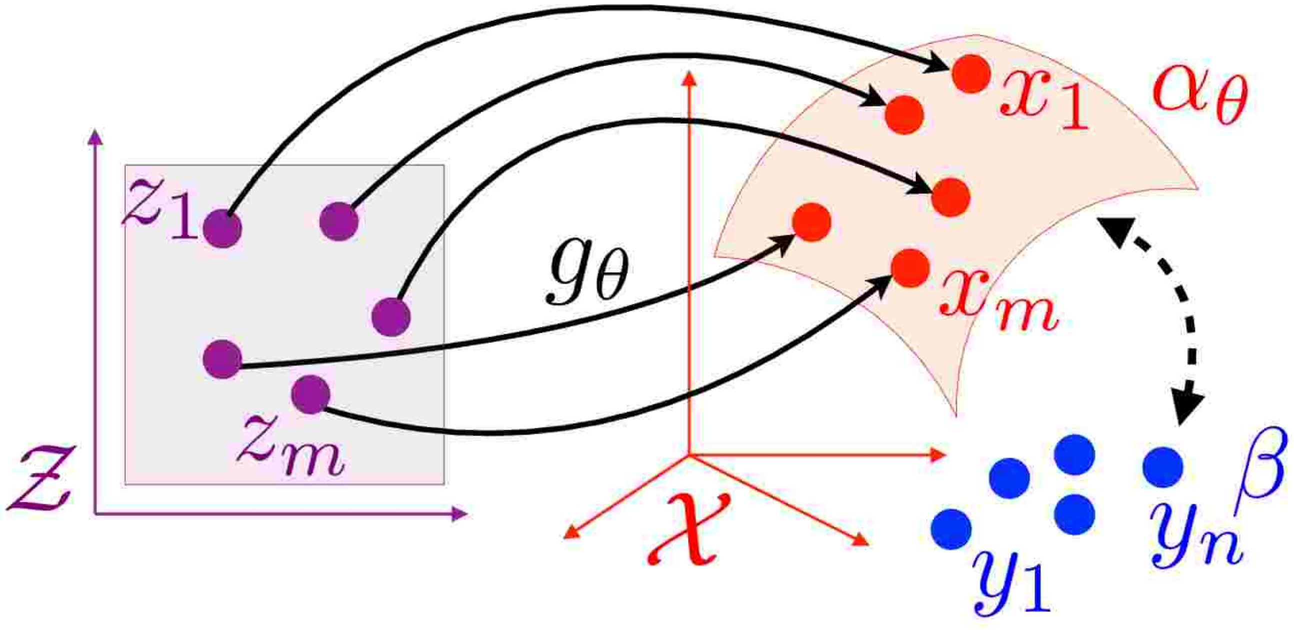
Deep networks:

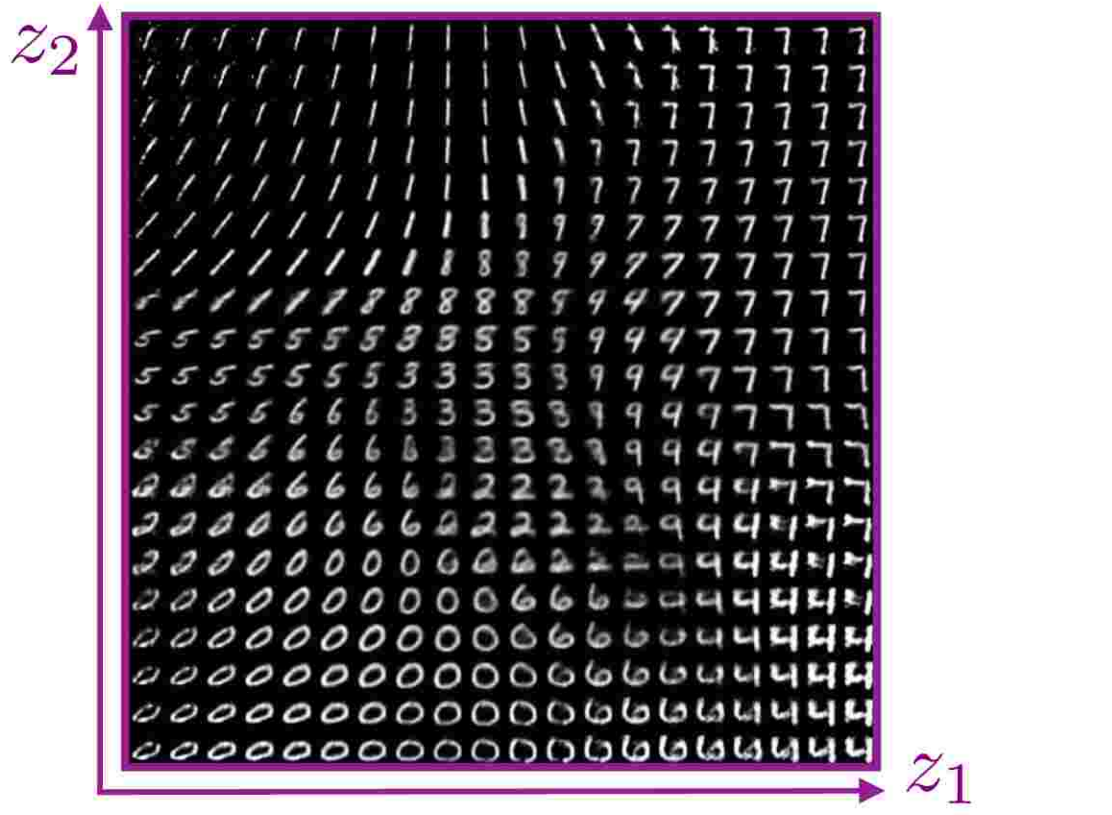
$$d_\xi(x) = \rho(\xi_K(\ldots\rho(\xi_2(\rho(\xi_1(x)\ldots)$$

$$g_\theta(z) = \rho(\theta_K(\ldots\rho(\theta_2(\rho(\theta_1(z)\ldots)$$

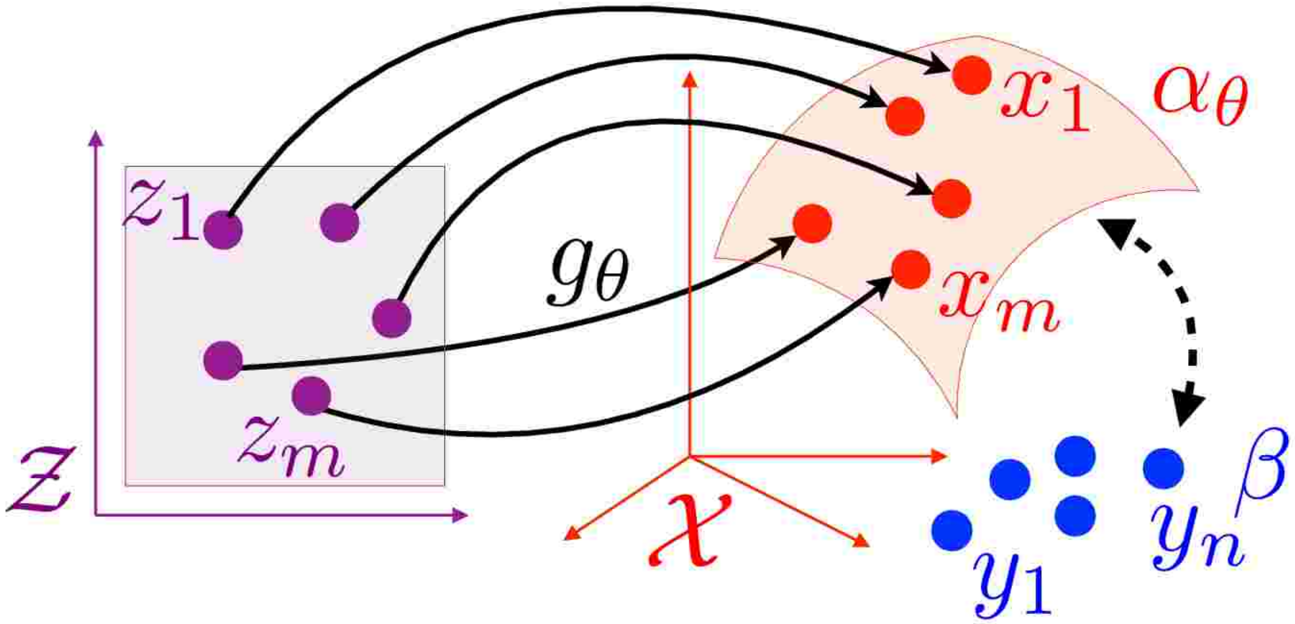








Training Architecture



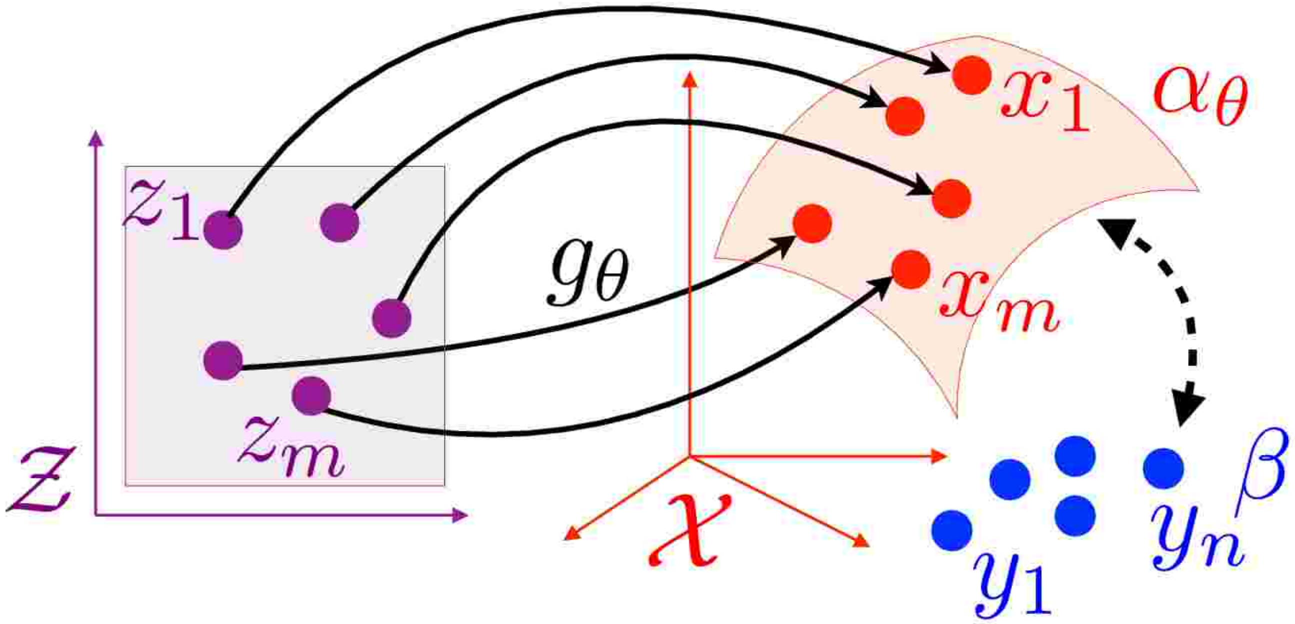
$$\min_\theta \mathcal{E}(\theta) \stackrel{\text{def.}}{=} \overline{\mathrm{W}}^p_{\varepsilon,p}(\alpha_\theta, \beta)$$

Stochastic gradient descent

$$\theta \leftarrow \theta - \tau \nabla \hat{\mathcal{E}}(\theta)$$

$$\hat{\mathcal{E}}(\theta) \stackrel{\text{def.}}{=} \overline{\mathrm{W}}^p_{\varepsilon,p}\left(\tfrac{1}{m} \textstyle\sum_i \delta_{g_\theta(z_i)}, \beta\right)$$

Training Architecture

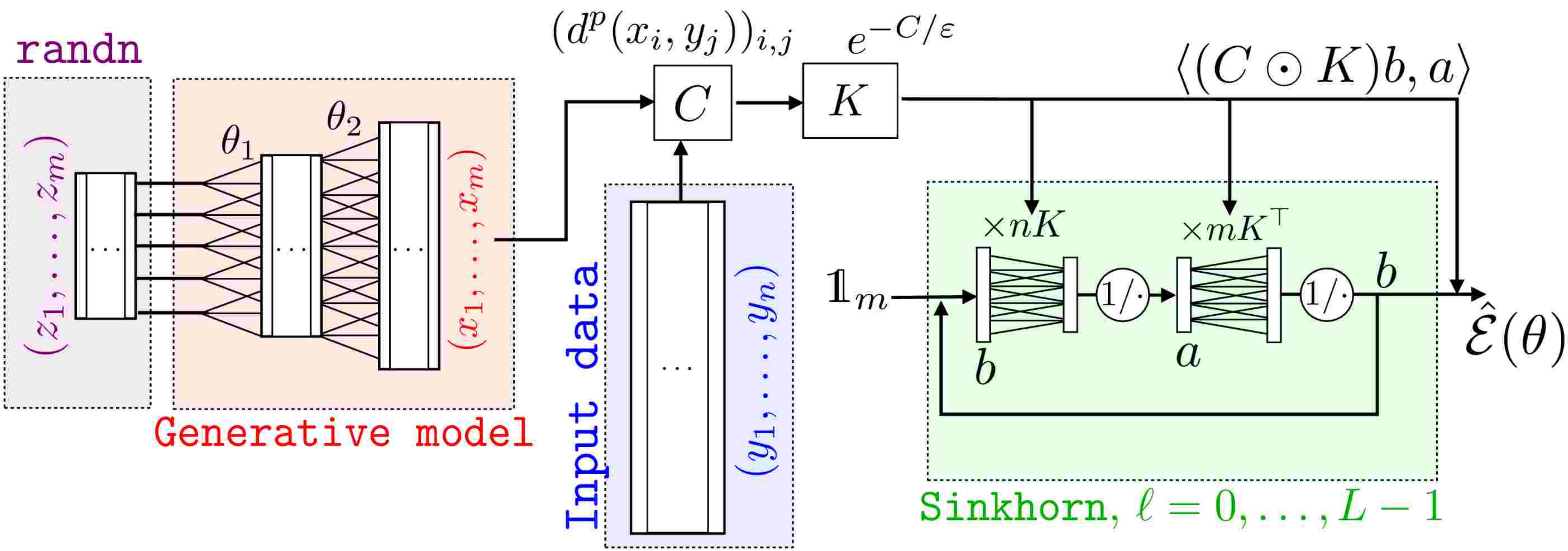


$$\min_\theta \mathcal{E}(\theta) \stackrel{\text{def.}}{=} \overline{\mathrm{W}}_{\varepsilon,p}^{p}(\alpha_\theta, \beta)$$

Stochastic gradient descent

$$\theta \leftarrow \theta - \tau \nabla \hat{\mathcal{E}}(\theta)$$

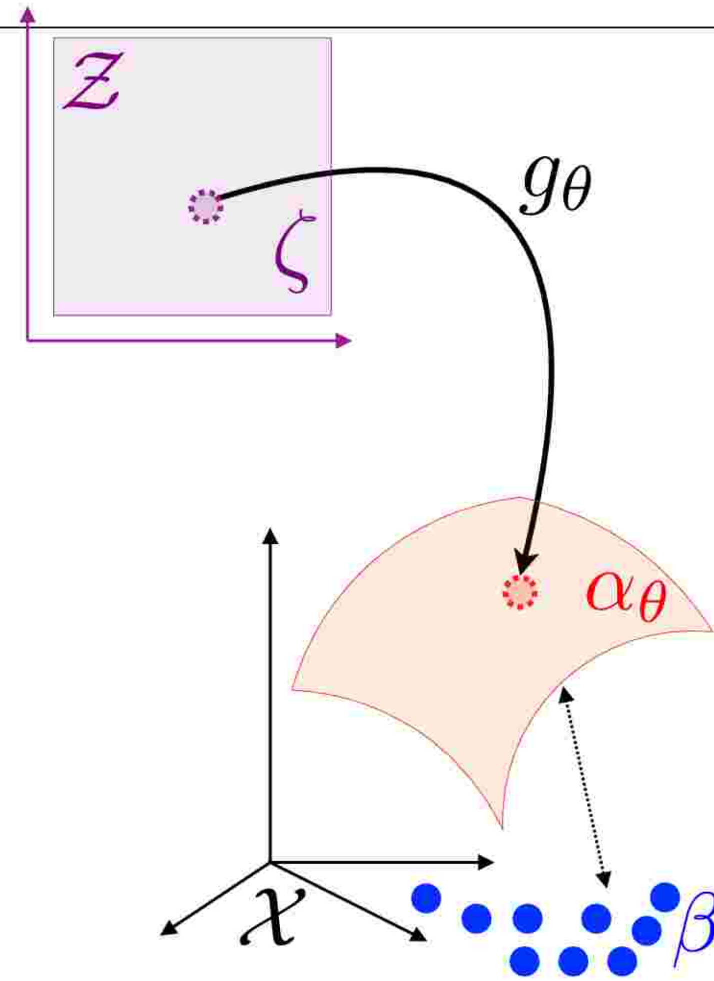
$$\hat{\mathcal{E}}(\theta) \stackrel{\text{def.}}{=} \overline{\mathrm{W}}_{\varepsilon,p}^{p}\left(\tfrac{1}{m}\sum_i \delta_{g_\theta(z_i)}, \beta\right)$$



Examples of Images Generation

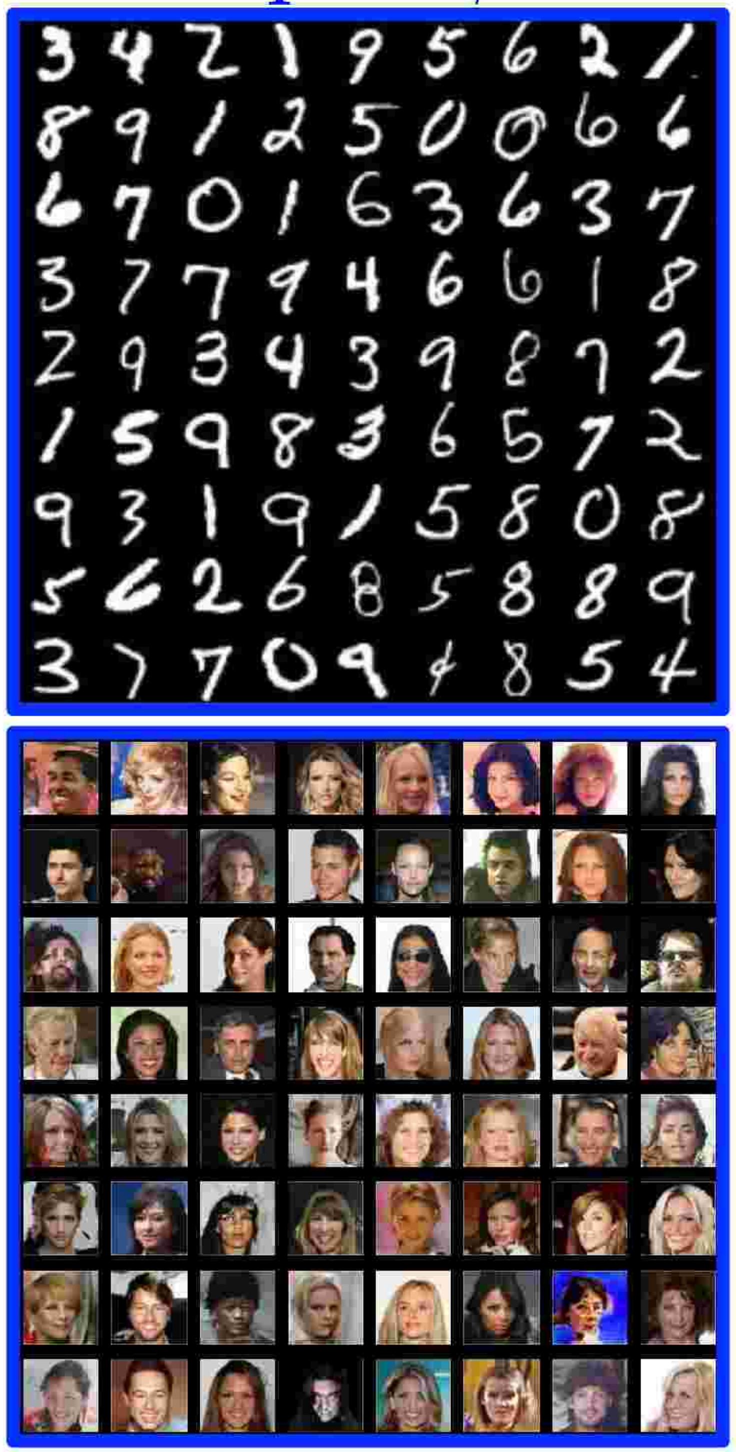
Inputs β

Generated α_θ

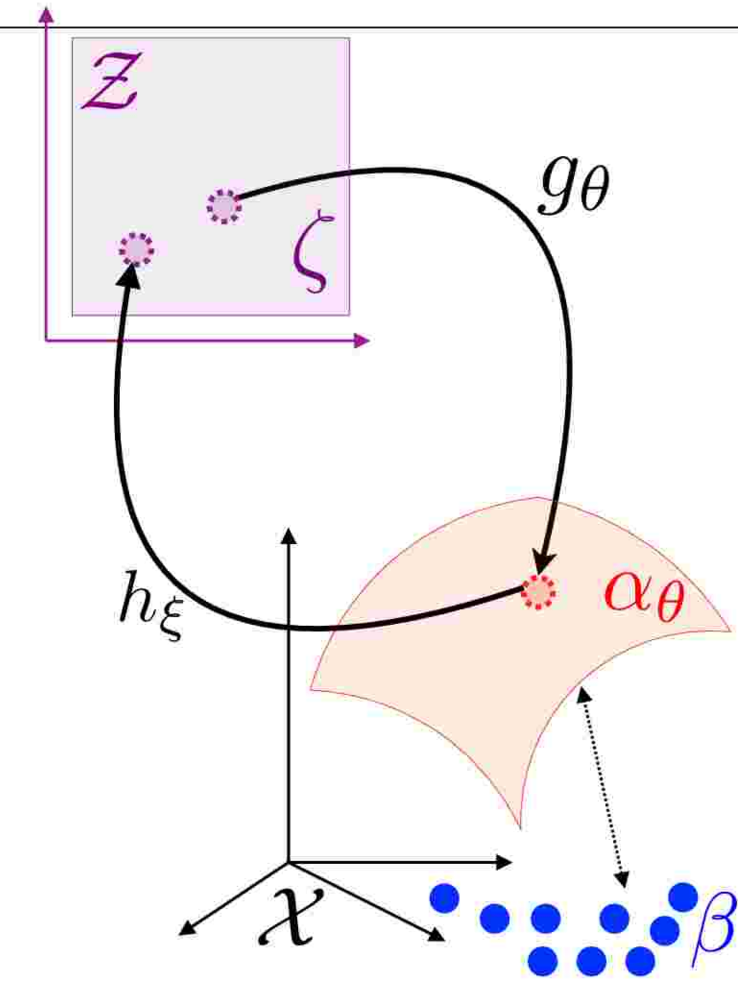


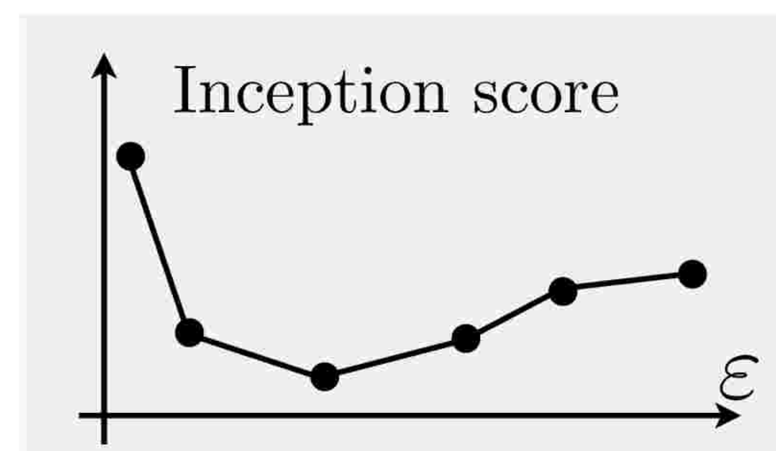
Examples of Images Generation

Inputs β



Generated α_θ

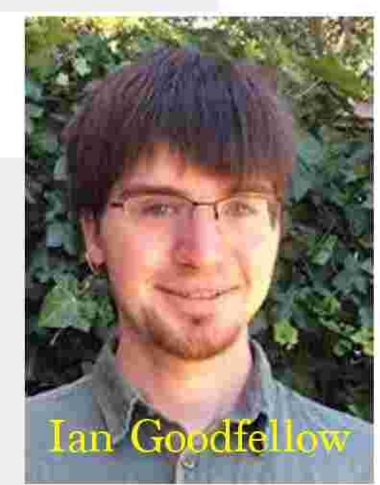




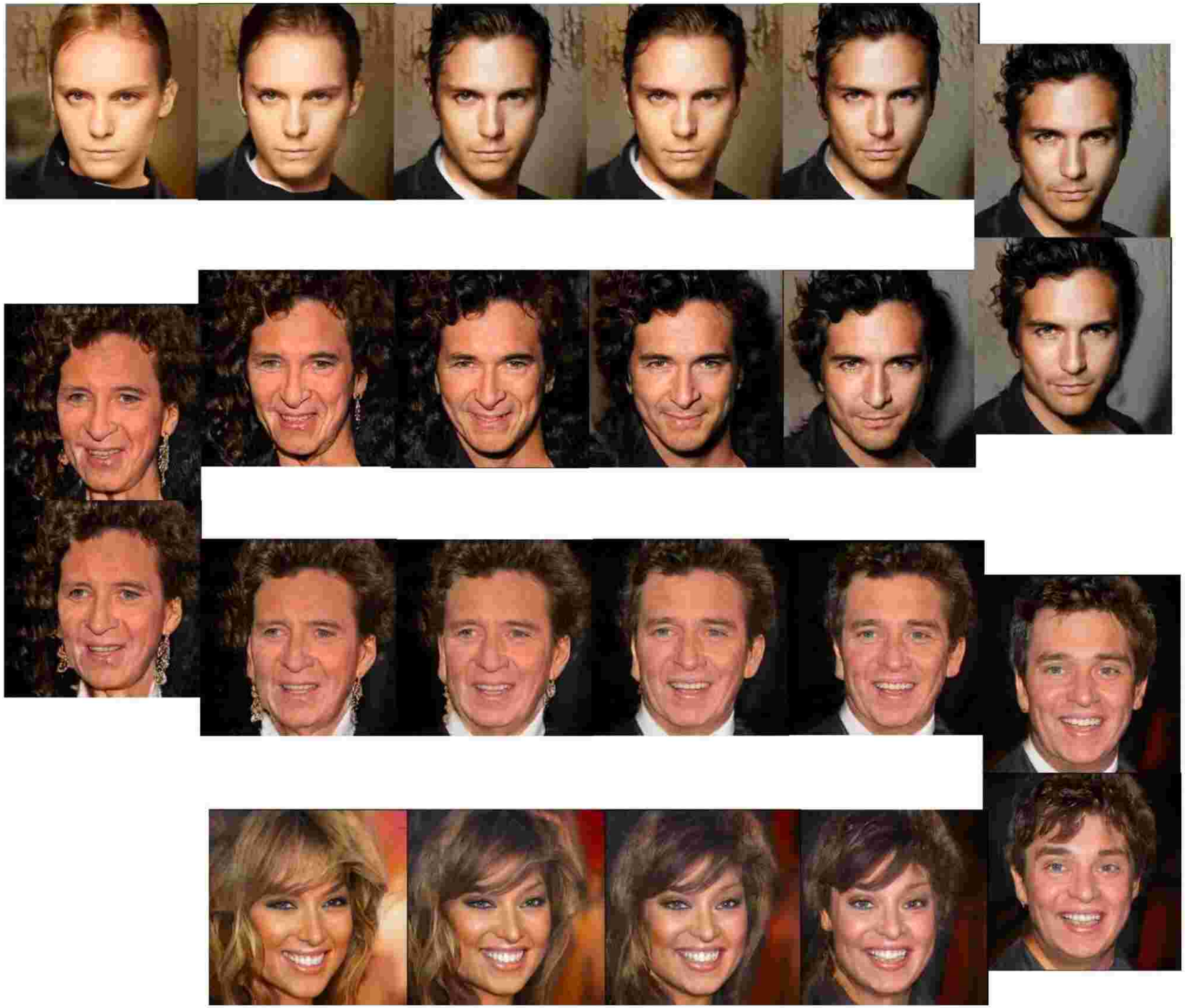
→ Need to learn the metric $d(x, y) = \|h_\xi(x) - h_\xi(y)\|$ (GANs)

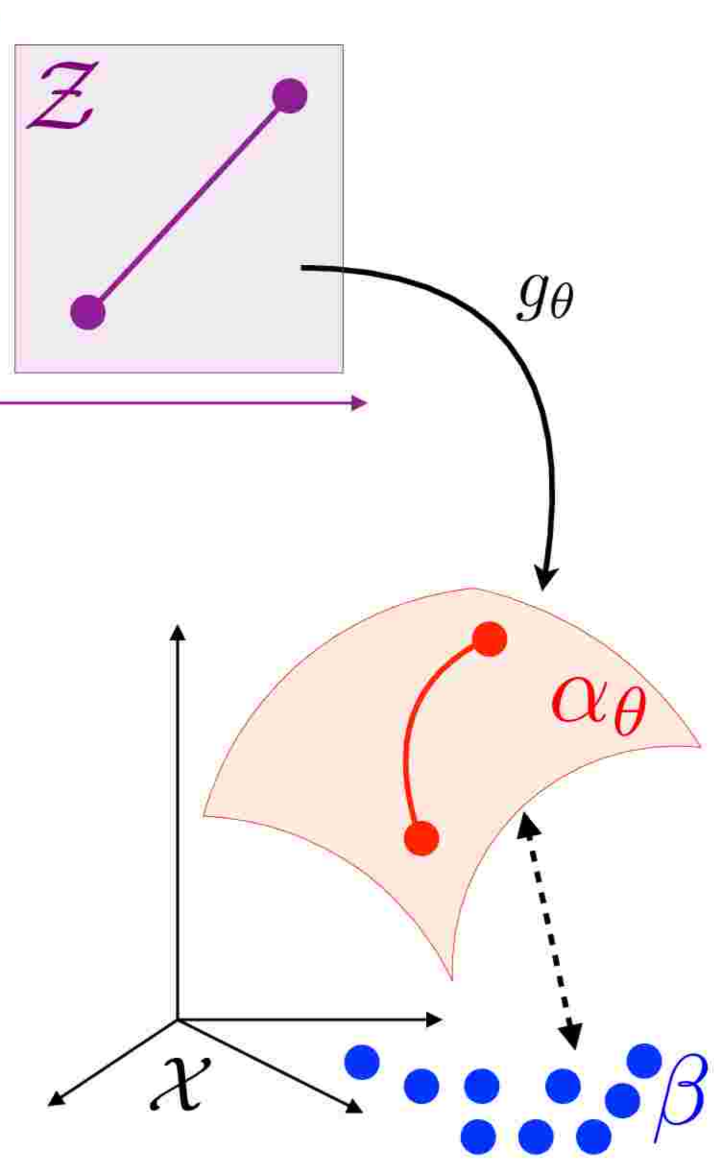
→ Influence of ε?

→ Performance evaluation of generative models is an open problem.



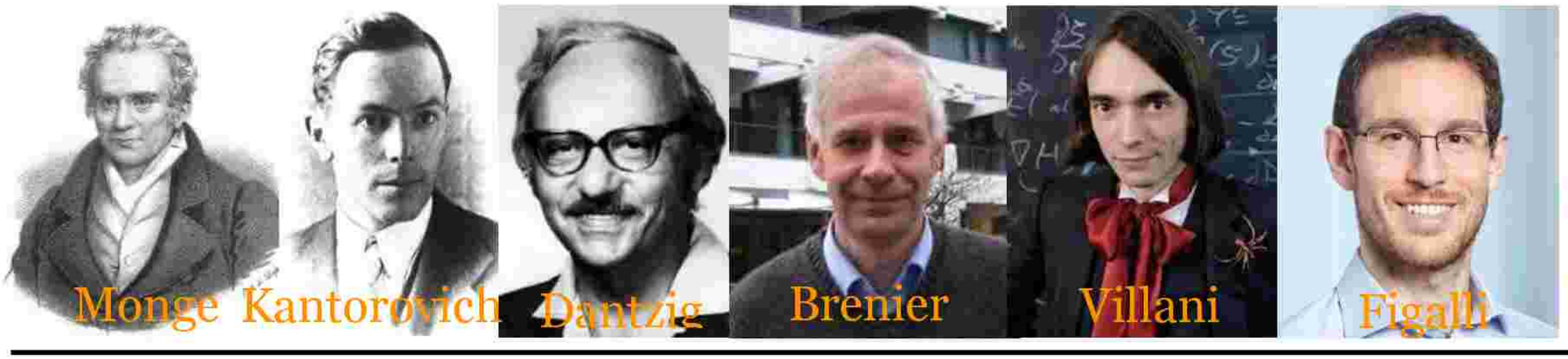
Generative Adversarial Networks

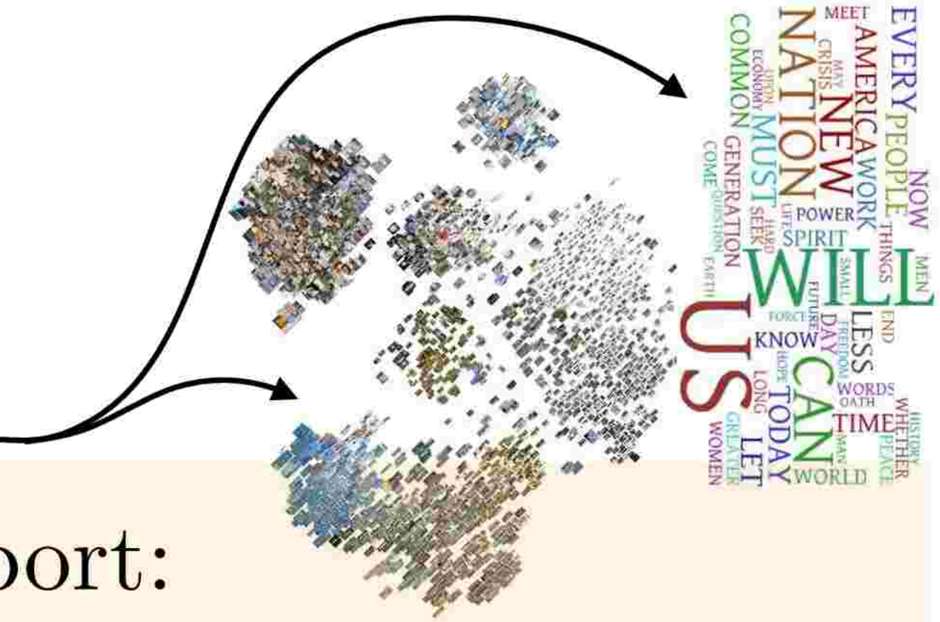




Progressive Growing of GANs for Improved Quality, Stability, and Variation
Tero Karras, Timo Aila, Samuli Laine, Jaakko Lehtinen, ICLR 2018

Conclusion

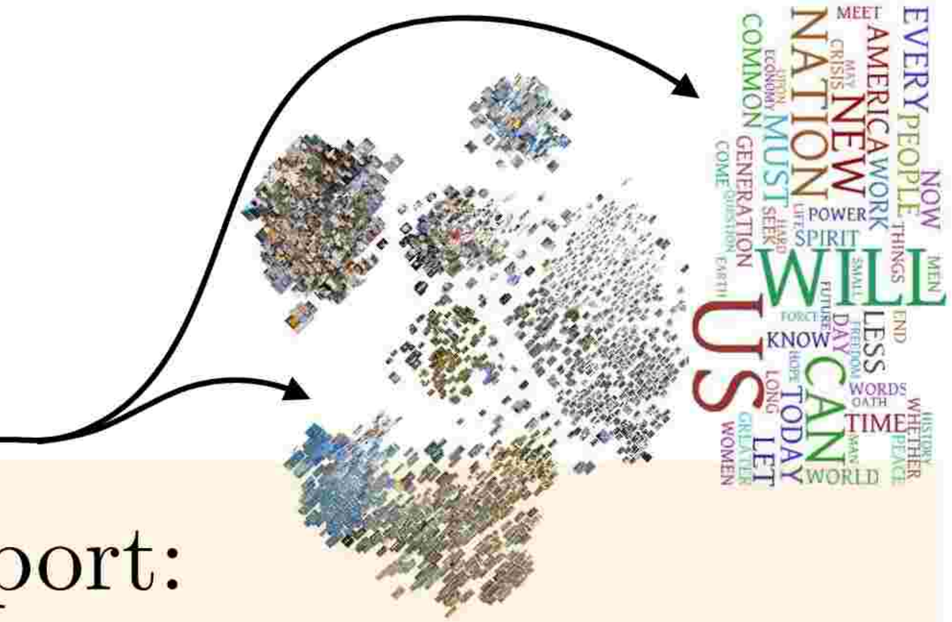




Toward high-dimensional optimal transport:

→ Scalable geometrical loss functions in high dimension?

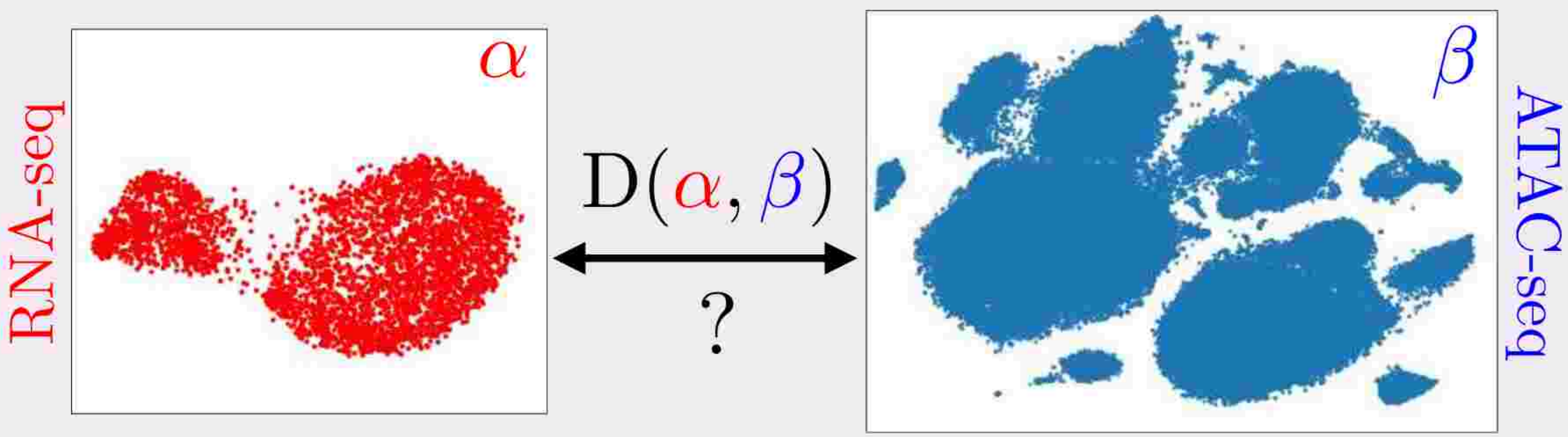
Conclusion



Toward high-dimensional optimal transport:

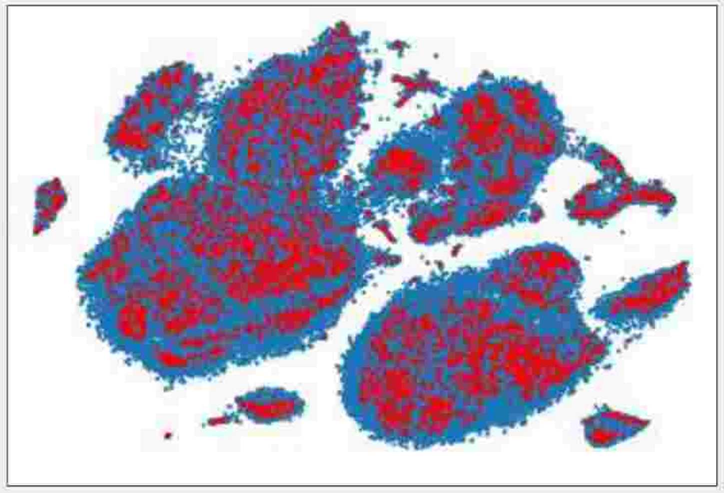
→ Scalable geometrical loss functions in high dimension?

Comparing datasets *across* spaces:



Single-cell multi-omics

[Othmane Sebbouh, 2021]



Gromov-Wasserstein registration