Optimal Transport Gromov-Wasserstein Problem for Dimensionality Reduction and Graph Analysis

Hugues Van Assel

Cédric Vincent-Cuaz Rémi Flamary Nicolas Courty Pascal Frossard Titouan Vayer

My talk

Overview of dimensionality reduction

Optimal Transport : from linear OT to Gromov Wasserstein

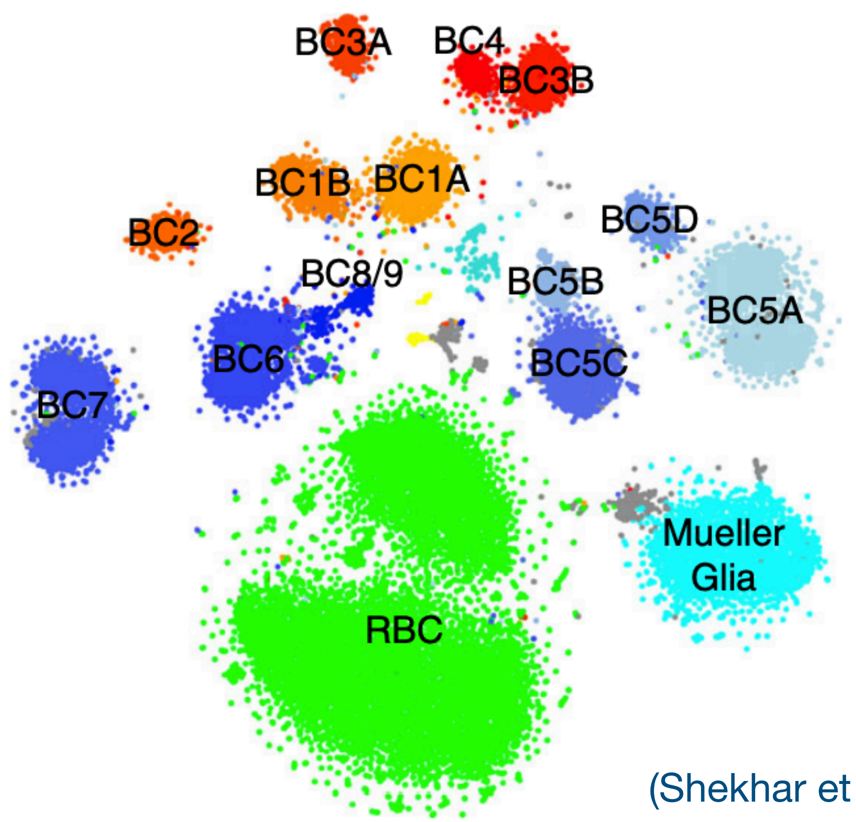
Distributional Reduction : a framework to embed distributions

vectorial data

Gromov Wasserstein for graph analysis
Application to graph generative modeling

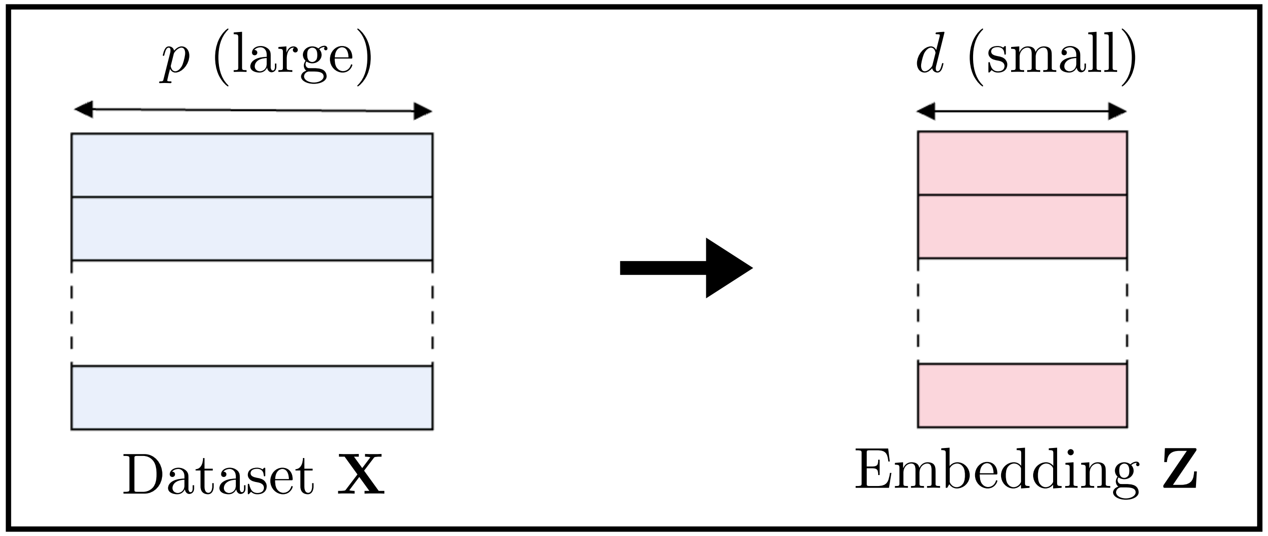
graph data

Dimension Reduction

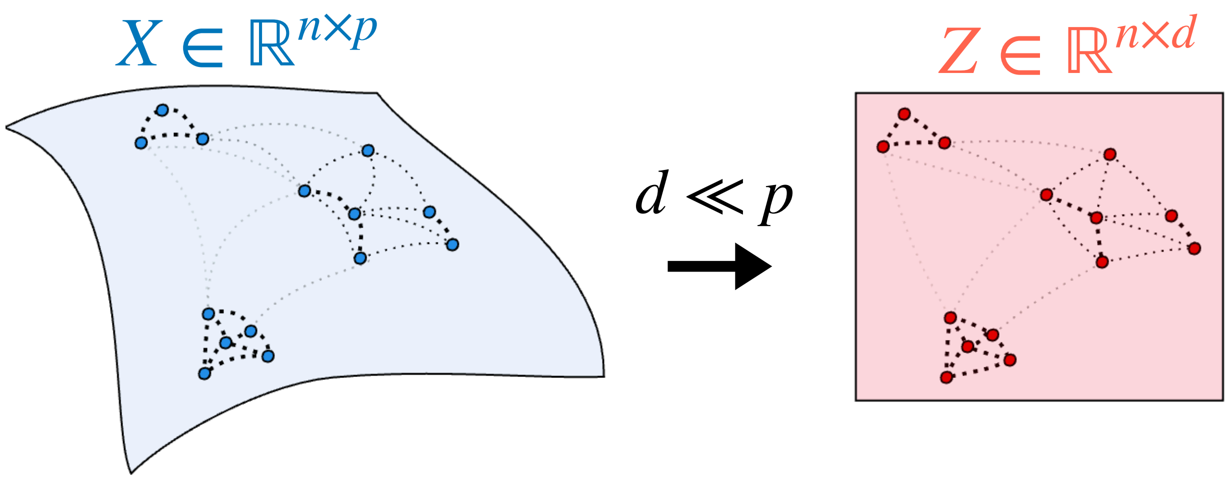


(Shekhar et al., 2016)

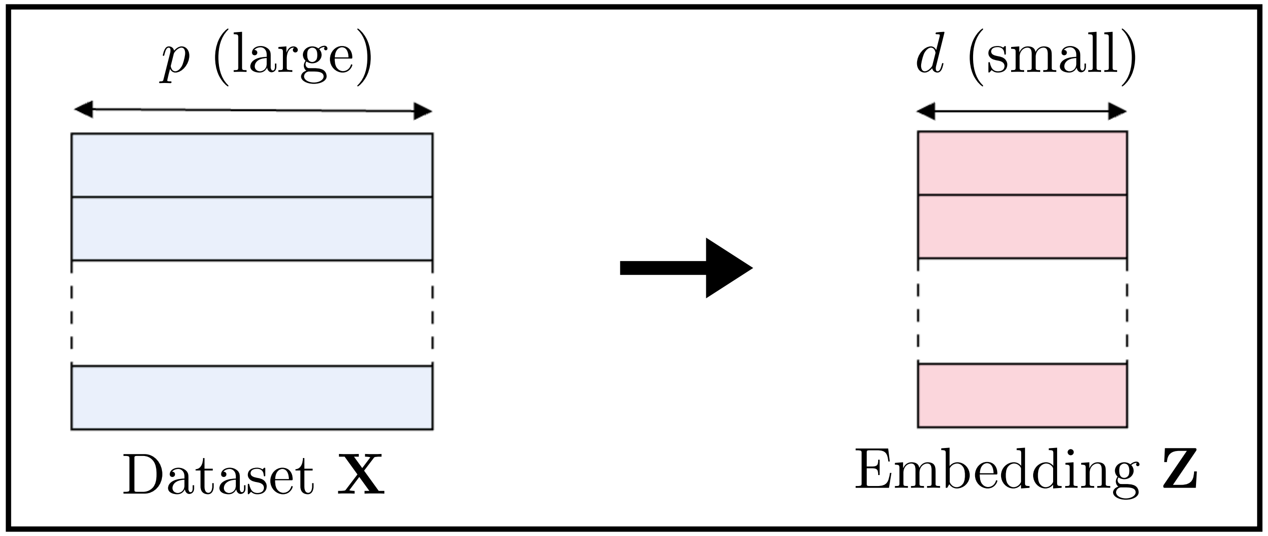
Dimension reduction



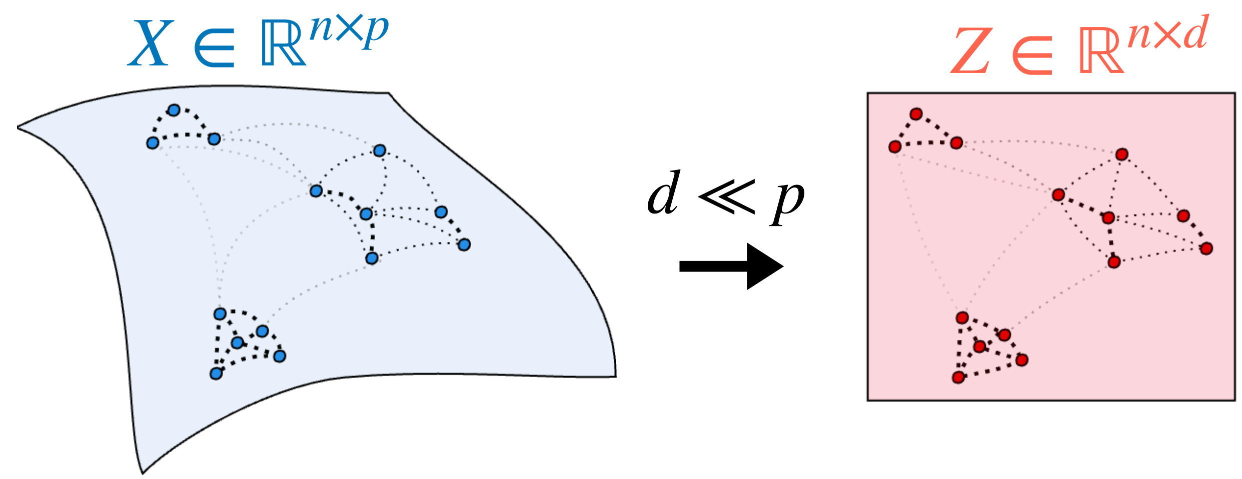
✦ **Preserving geometric properties**



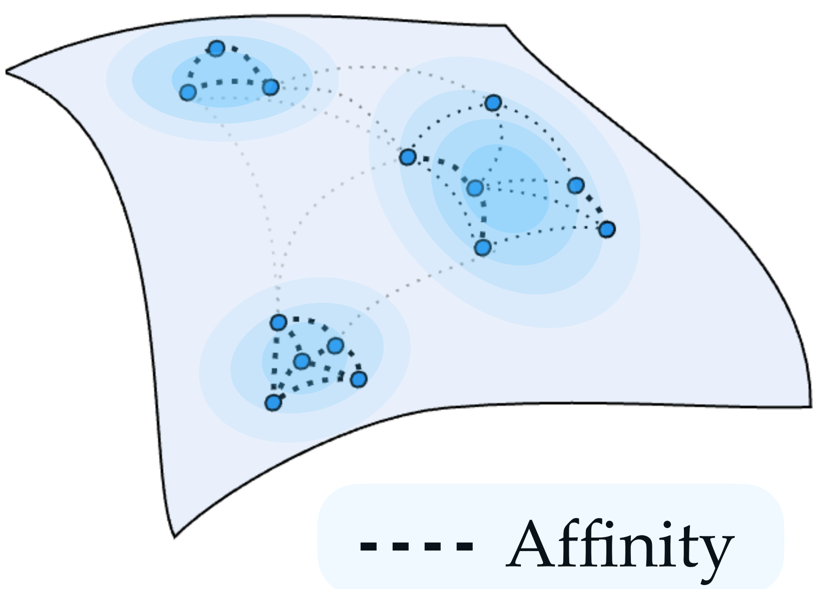
Dimension reduction

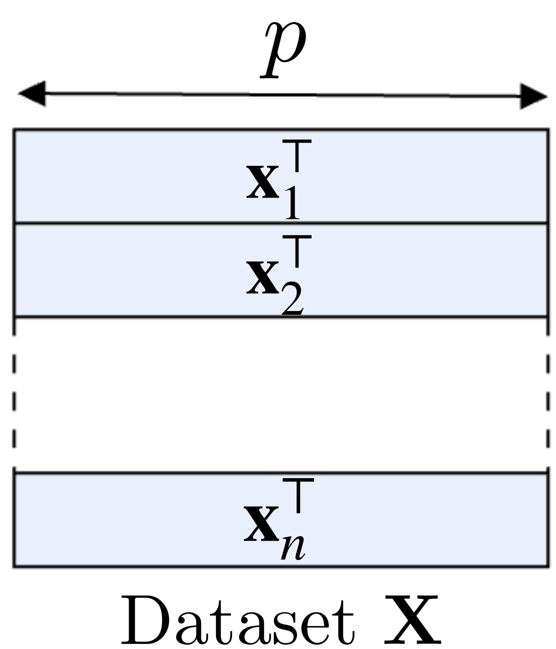


✦ Preserving geometric structure



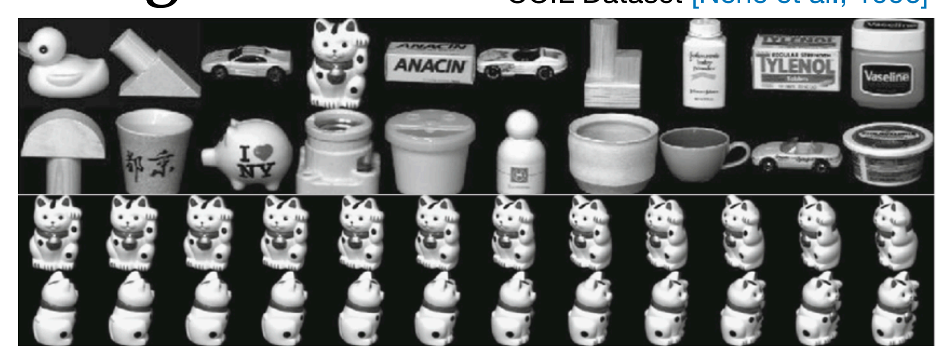
✦ Affinity Matrices



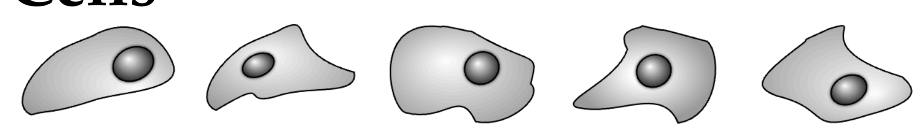


Images

COIL Dataset [Nene et al., 1996]



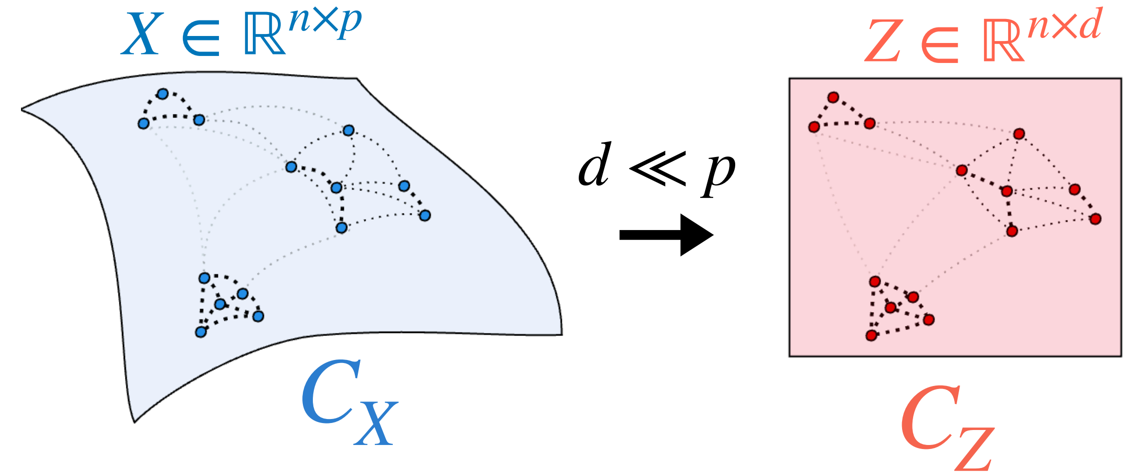
Cells



Symmetric matrix with **non-negative coefficients**.

Coefficient (i, j) = similarity between $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$.

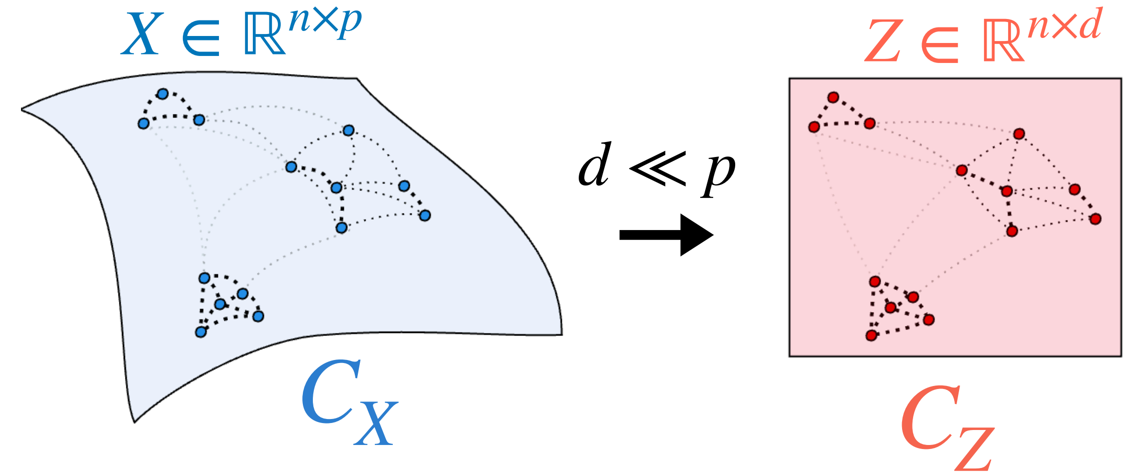
Dimension reduction



✦ A general optimization problem

$$\min_{Z \in \mathbb{R}^{n \times d}} \sum_{i,j=1}^{n} L\left([C_X]_{ij}\,,\,[C_Z]_{ij}\right) \quad \text{for some loss} \quad L : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$$

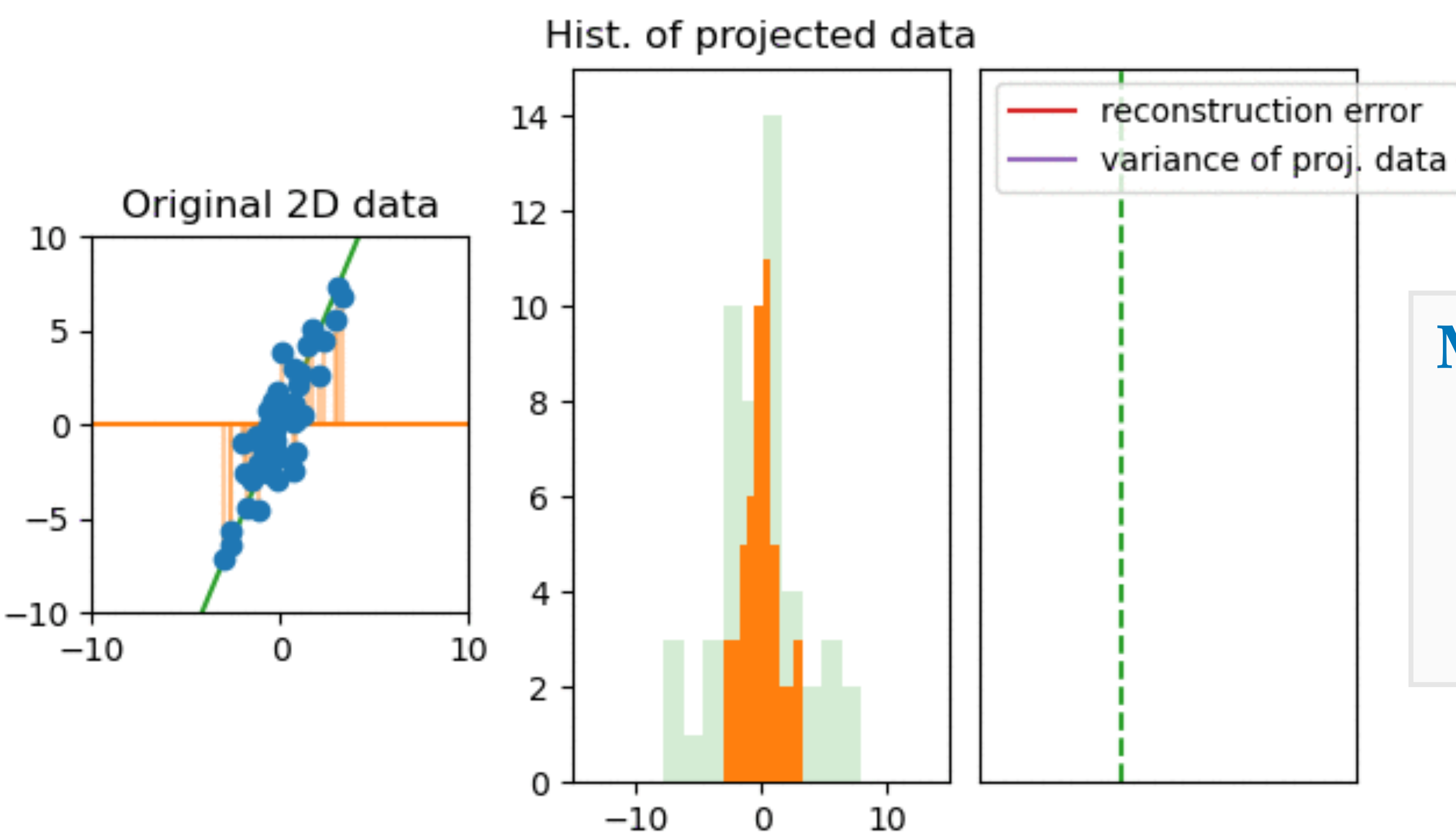
Dimension reduction

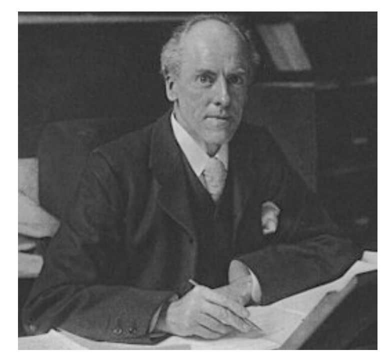


✦ A general optimization problem

$$\min_{Z\in\mathbb{R}^{n\times d}} \sum_{i,j=1}^{n} L\left([C_X]_{ij}\,,\,[C_Z]_{ij}\right) \quad \text{for some loss} \quad L : \mathbb{R}\times\mathbb{R} \to \mathbb{R}$$

✦ Principal components analysis



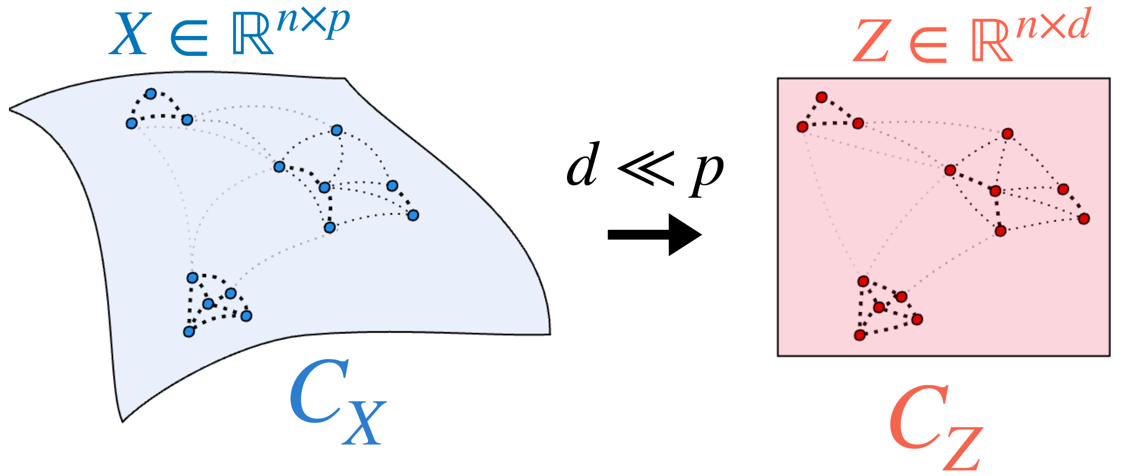


(Pearson, 1901)

Minimizing the reconstruction error

$$\min_{H:\dim(H)=d} \frac{1}{n}\sum_{i=1}^{n} \|\mathbf{x}_i - P_H(\mathbf{x_i})\|_2^2$$

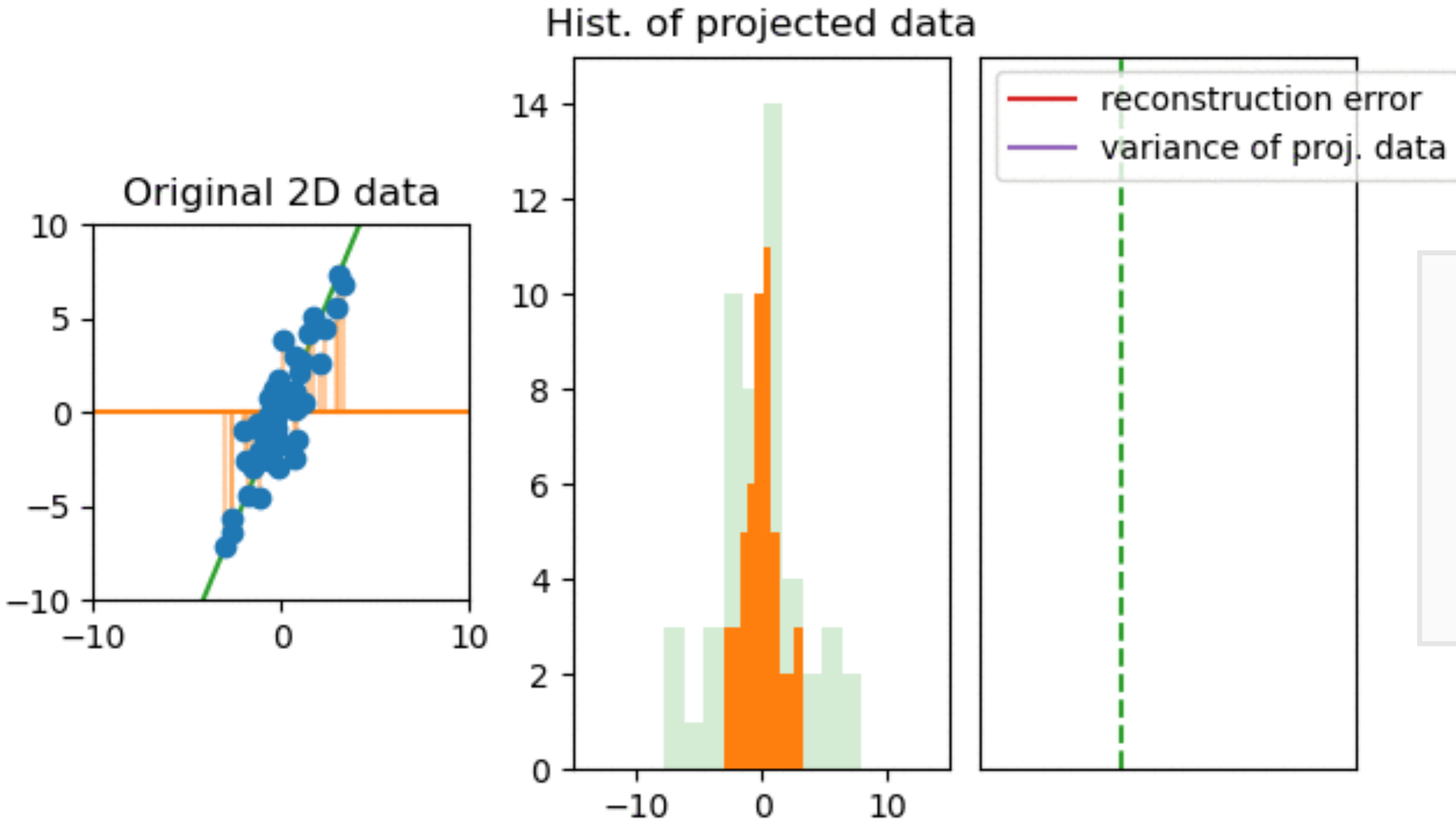
Dimension reduction

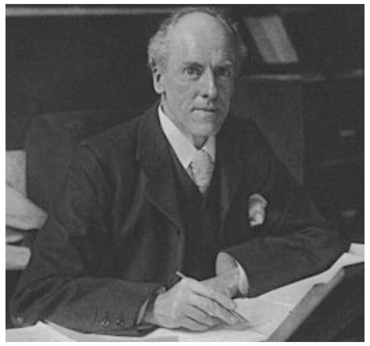


✦ A general optimization problem

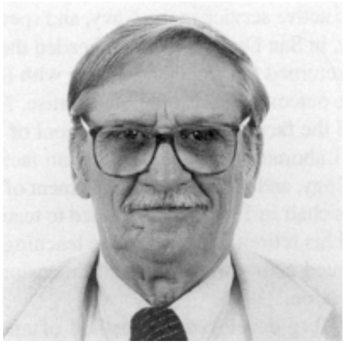
$$\min_{Z \in \mathbb{R}^{n \times d}} \sum_{i,j=1}^{n} L\left([C_X]_{ij}\,,\,[C_Z]_{ij}\right) \text{ for some loss } L : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$$

✦ Principal components analysis





(Pearson, 1901)



(Torgerson, 1958)

Preserving the inner products

$$\min_{Z \in \mathbb{R}^{n \times d}} \sum_{i=1}^{n} \left(\langle \mathbf{x}_i, \mathbf{x}_j \rangle - \langle \mathbf{z}_i, \mathbf{z}_j \rangle\right)^2$$

$$Z \leftarrow \mathrm{EVD}\left(\frac{1}{n} X X^\top\right)$$

Dimension reduction

- **Spectral methods**

$$\min_{Z \in \mathbb{R}^{n \times d}} \sum_{i=1}^{n} \left([C_X]_{ij} - \langle \mathbf{z}_i, \mathbf{z}_j \rangle \right)^2$$

Dimension reduction

✦ **Spectral methods**

$$\min_{Z\in\mathbb{R}^{n\times d}} \sum_{i=1}^{n} \left([C_X]_{ij} - \langle \mathbf{z}_i, \mathbf{z}_j \rangle \right)^2$$

$C_X \succeq 0$

solution →

(Eckart & Young, 1936)

$$Z^\star = (\sqrt{\lambda_1}\mathbf{v}_1, \cdots, \sqrt{\lambda_d}\mathbf{v}_d)^\top$$

λ_i i-th **largest** eigenvalue of C_X with eigenvector $\mathbf{v}_i$

$$[C_X]_{ij} = \langle \phi(X_i), \phi(X_j) \rangle_{\mathscr{H}}$$

Dimension reduction

✦ **Spectral methods**

$$\min_{Z \in \mathbb{R}^{n \times d}} \sum_{i=1}^{n} \left([C_X]_{ij} - \langle \mathbf{z}_i, \mathbf{z}_j \rangle \right)^2$$

$C_X \succeq 0$

solution →

(Eckart & Young, 1936)

$$Z^\star = (\sqrt{\lambda_1}\mathbf{v}_1, \cdots, \sqrt{\lambda_d}\mathbf{v}_d)^\top$$

λ_i i-th **largest** eigenvalue of C_X with eigenvector $\mathbf{v}_i$

✦ **Kernel PCA** $C_X \succeq 0$ $[C_X]_{ij} = \langle \phi(X_i), \phi(X_j) \rangle_{\mathcal{H}}$

(Schölkopf, 1997)

→ PCA: $C_X = XX^\top$ $(Z \leftarrow \mathrm{SVD}(X))$

Dimension reduction

✦ Spectral methods

$$\min_{Z \in \mathbb{R}^{n \times d}} \sum_{i=1}^{n} \left([C_X]_{ij} - \langle \mathbf{z}_i, \mathbf{z}_j \rangle \right)^2$$

$C_X \succeq 0$

solution →

(Eckart & Young, 1936)

$$Z^\star = (\sqrt{\lambda_1}\mathbf{v}_1, \cdots, \sqrt{\lambda_d}\mathbf{v}_d)^\top$$

λ_i i-th **largest** eigenvalue of C_X with eigenvector $\mathbf{v}_i$

✦ Kernel PCA $C_X \succeq 0$

$[C_X]_{ij} = \langle \phi(X_i), \phi(X_j) \rangle_{\mathcal{H}}$

(Schölkopf, 1997)

- PCA: $C_X = XX^\top$ $(Z \leftarrow \mathrm{SVD}(X))$
- (classical) Multidimensional scaling: $C_X = -\frac{1}{2} H D_X H$

Dimension reduction

✦ Spectral methods

$$\min_{Z \in \mathbb{R}^{n \times d}} \sum_{i=1}^{n} \left([C_X]_{ij} - \langle \mathbf{z}_i, \mathbf{z}_j \rangle \right)^2$$

$C_X \succeq 0$

solution →

(Eckart & Young, 1936)

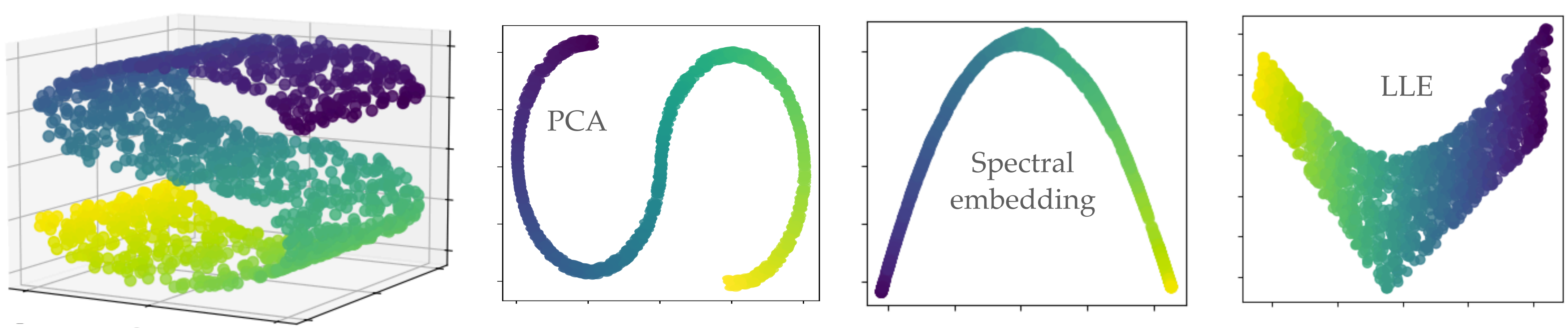
$$Z^\star = (\sqrt{\lambda_1}\mathbf{v}_1, \cdots, \sqrt{\lambda_d}\mathbf{v}_d)^\top$$

λ_i i-th **largest** eigenvalue of C_X with eigenvector $\mathbf{v}_i$

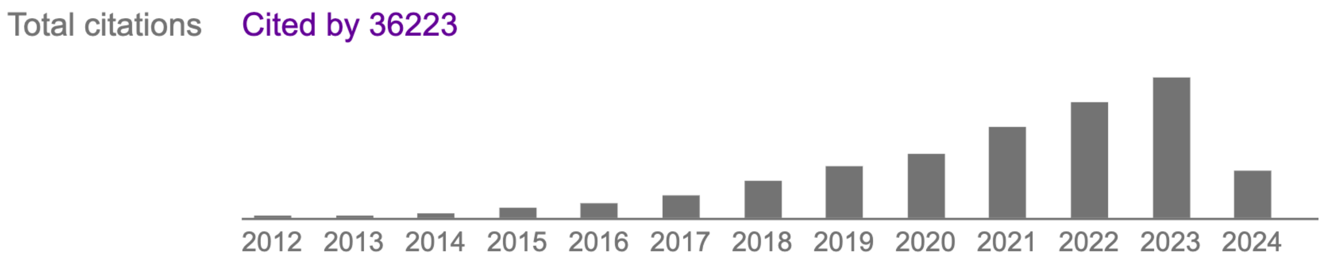
✦ Kernel PCA $C_X \succeq 0$ $[C_X]_{ij} = \langle \phi(X_i), \phi(X_j) \rangle_{\mathcal{H}}$

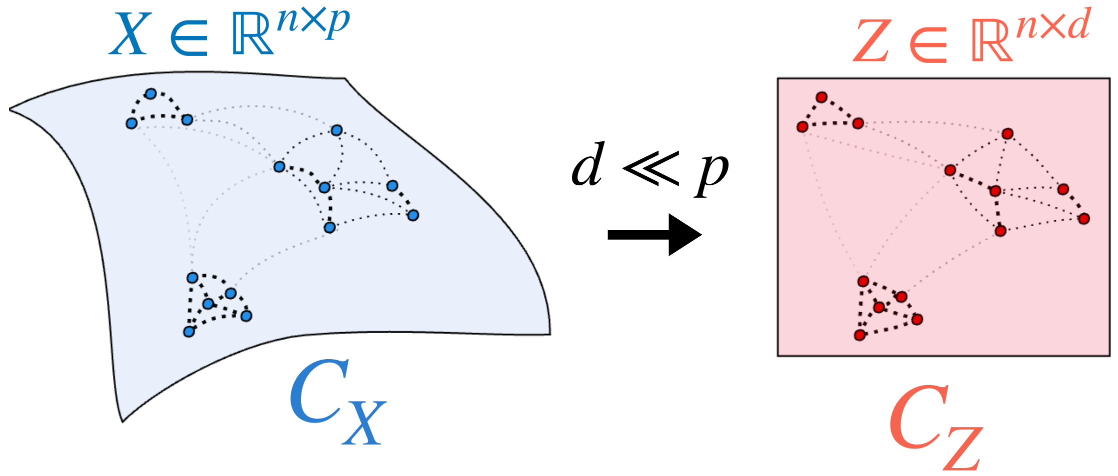
(Schölkopf, 1997)

- PCA: $C_X = XX^\top$ $(Z \leftarrow \text{SVD}(X))$
- (classical) Multidimensional scaling: $C_X = -\frac{1}{2} H D_X H$
- Laplacian Eigenmap (spectral embedding): $C_X = L_X^\dagger$
  (Belkin & Niyogi, 2003)
- Locally Linear Embedding, Diffusion Map …
  (Roweis & Saul, 2000) (Coifman & Lafon, 2006)



Dimension reduction



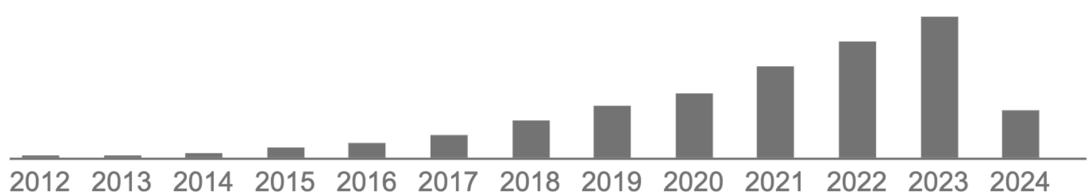


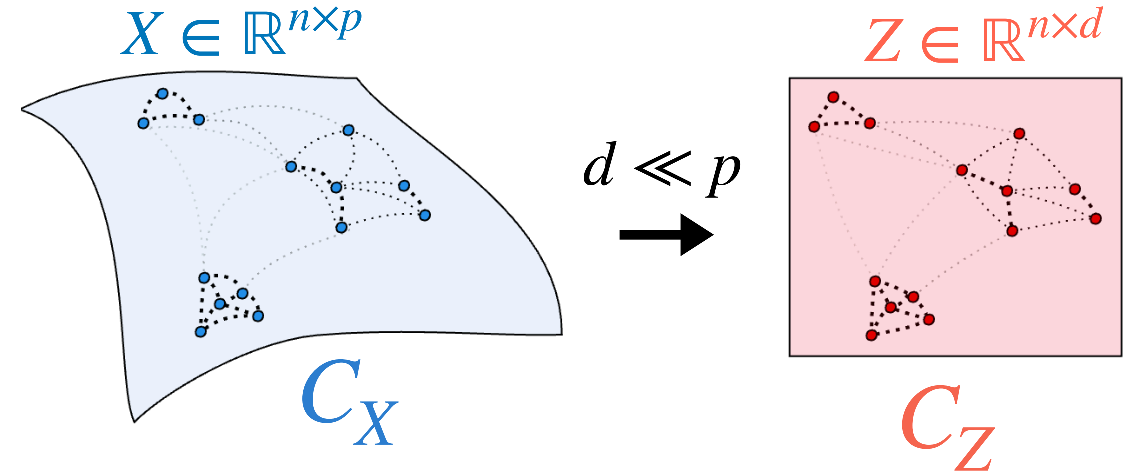
✦ Neighbor embedding methods

$$\min_{Z \in \mathbb{R}^{n \times d}} \sum_{i,j=1}^{n} \mathrm{KL}\left([C_X]_{ij} \, , \, [C_Z]_{ij} \right)$$

Dimension reduction

Total citations Cited by 36223





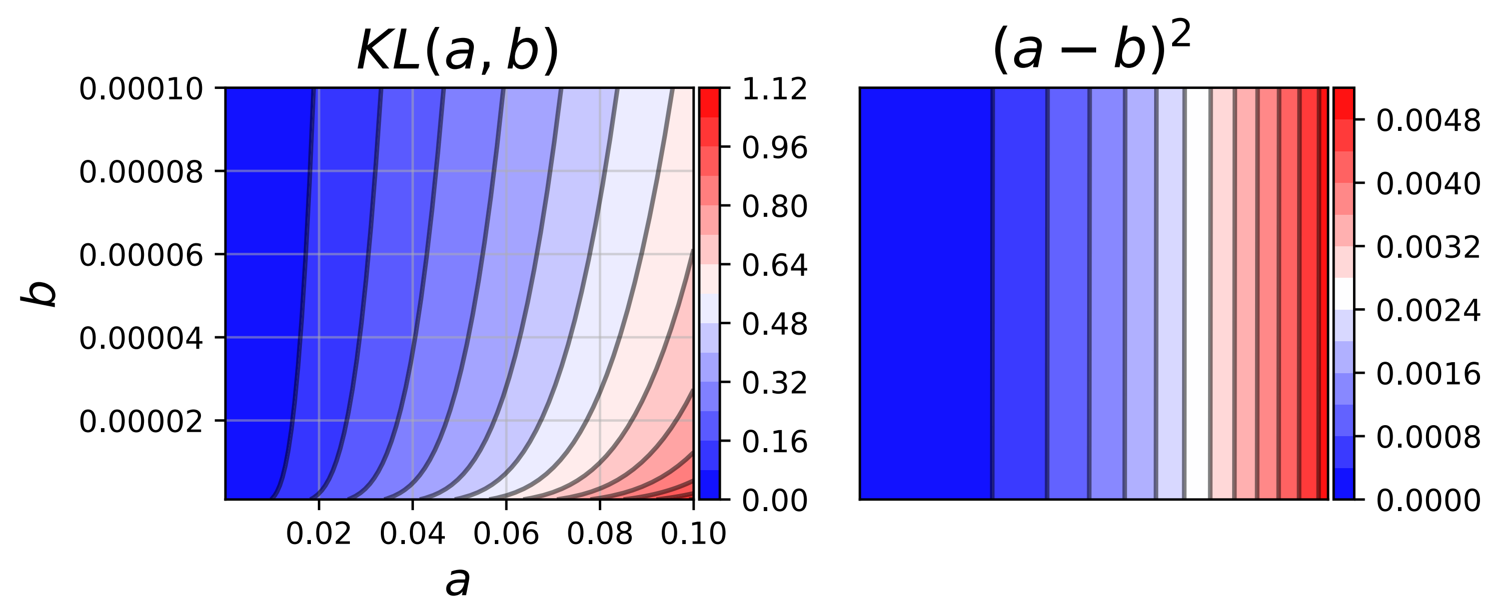
✦ Neighbor embedding methods

$$\min_{Z\in\mathbb{R}^{n\times d}} \sum_{i,j=1}^{n} \mathrm{KL}\left([C_X]_{ij}\ ,\ [C_Z]_{ij}\right)$$

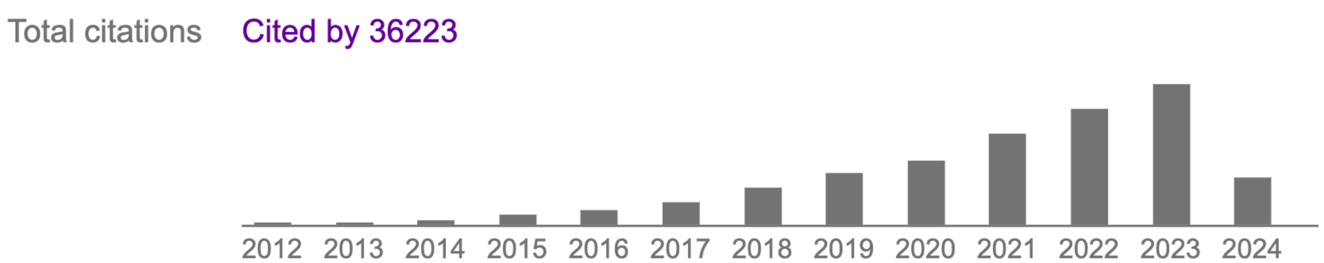
✦ Kullback-Leiber divergence

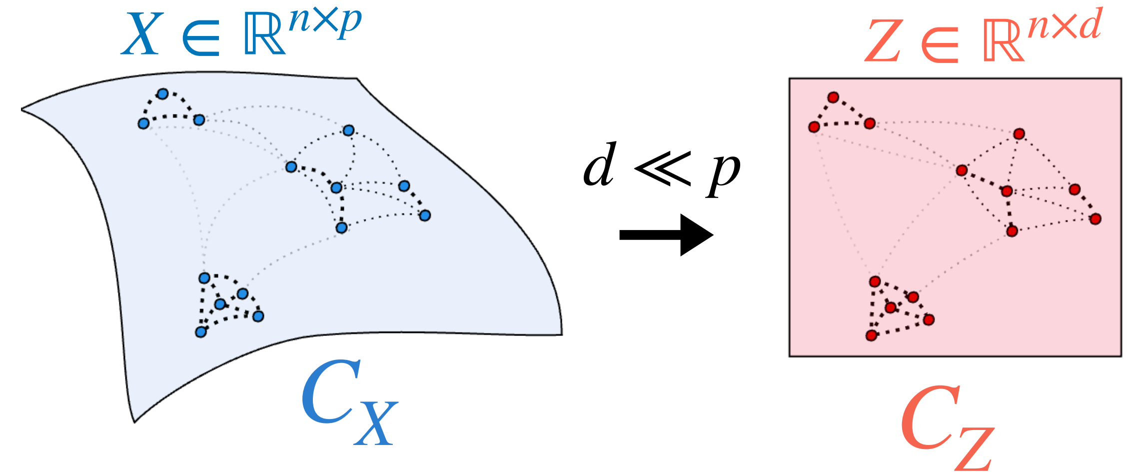
$$\mathrm{KL}(a,b) = a\log(a/b) - a + b = D_\phi(a,b)$$

Shannon–Boltzman entropy $\phi(x) = x\log(x) - x + 1$



Dimension reduction





✦ Neighbor embedding methods

$$\min_{Z\in\mathbb{R}^{n\times d}} \sum_{i,j=1}^{n} \mathrm{KL}\left([C_X]_{ij}\,,\,[C_Z]_{ij}\right)$$

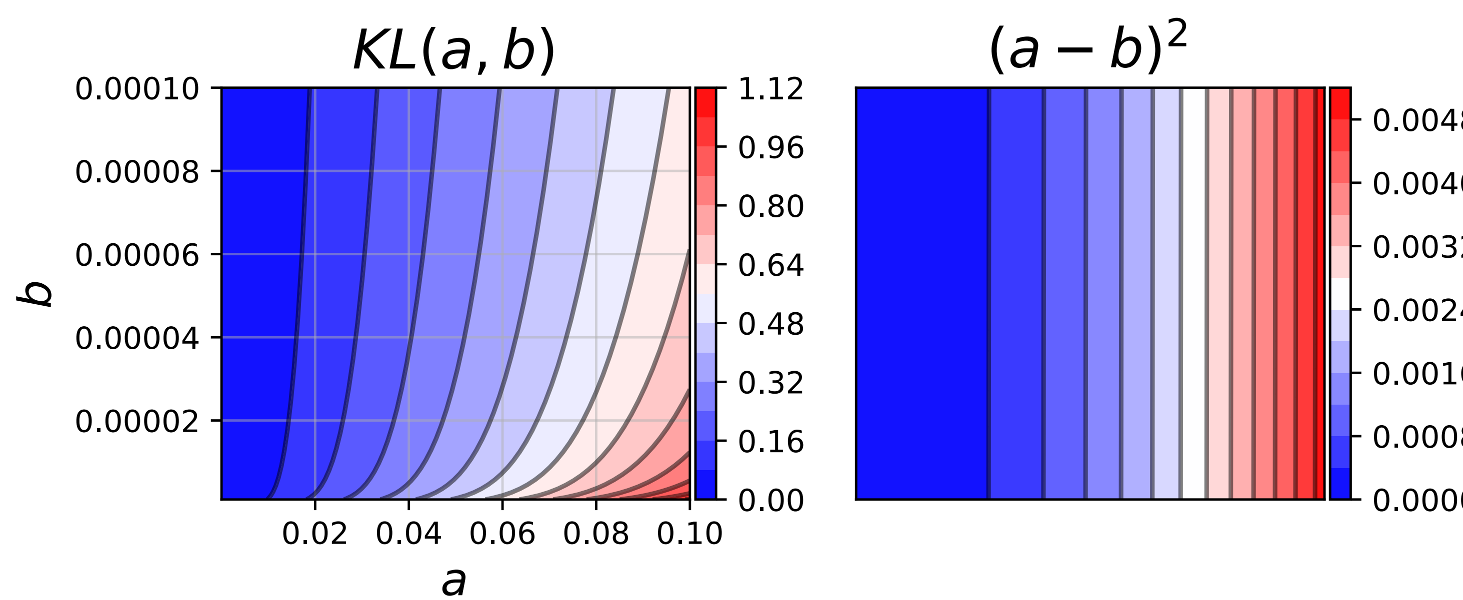
When $\sum_{i,j}[C_X]_{ij} = \sum_{i,j}[C_Z]_{ij}$ (same mass)

$$\sim \min_{Z\in\mathbb{R}^{n\times d}} \underbrace{\sum_{i,j=1}^{n}[C_X]_{ij}\,\log\left(\frac{[C_X]_{ij}}{[C_Z]_{ij}}\right)}_{\textbf{KL}}$$

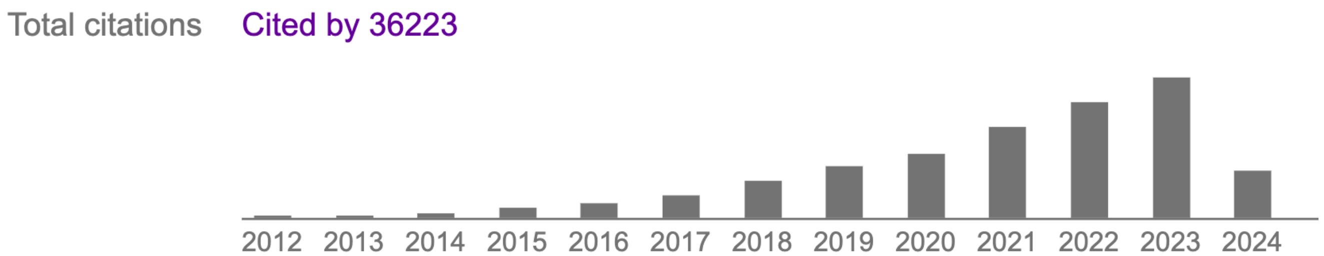
✦ Kullback-Leiber divergence

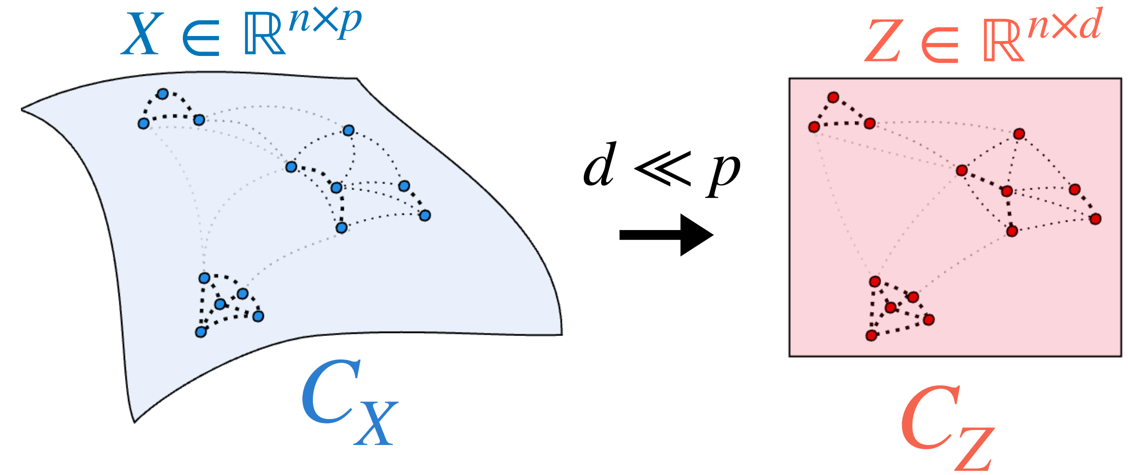
$$\mathrm{KL}(a,b) = a\log(a/b) - a + b = D_\phi(a,b)$$

Shannon–Boltzman entropy $\phi(x) = x\log(x) - x + 1$



Dimension reduction





✦ Neighbor embedding methods

$$\min_{Z\in\mathbb{R}^{n\times d}} \sum_{i,j=1}^{n} \mathrm{KL}\left([C_X]_{ij} \, , \, [C_Z]_{ij}\right)$$

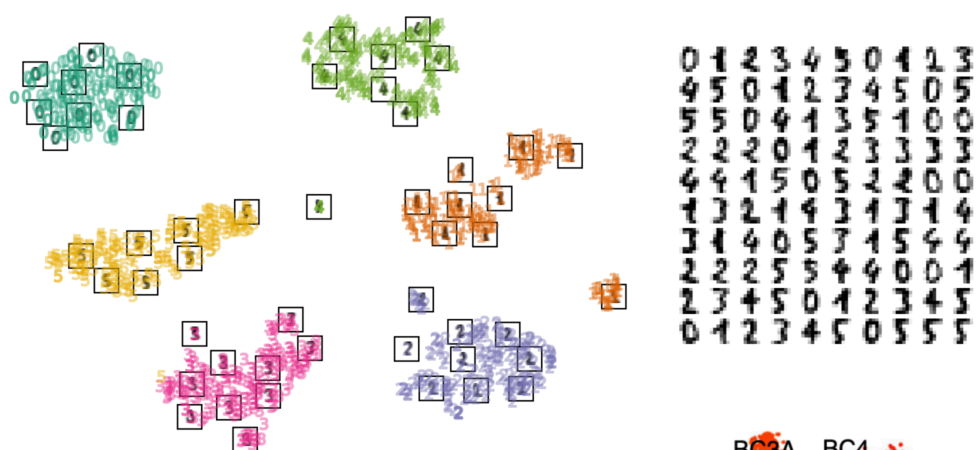
When $\sum_{i,j} [C_X]_{ij} = \sum_{i,j} [C_Z]_{ij}$ (same mass)

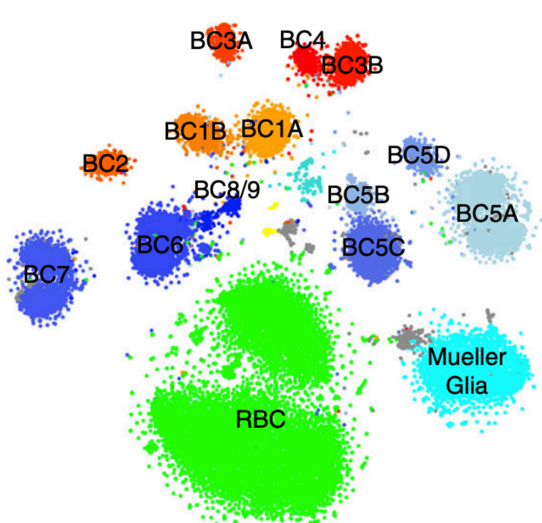
$$\sim \min_{Z\in\mathbb{R}^{n\times d}} \underbrace{\sum_{i,j=1}^{n} [C_X]_{ij} \, \log\left(\frac{[C_X]_{ij}}{[C_Z]_{ij}}\right)}_{\mathbf{KL}}$$

✦ Kullback-Leiber divergence

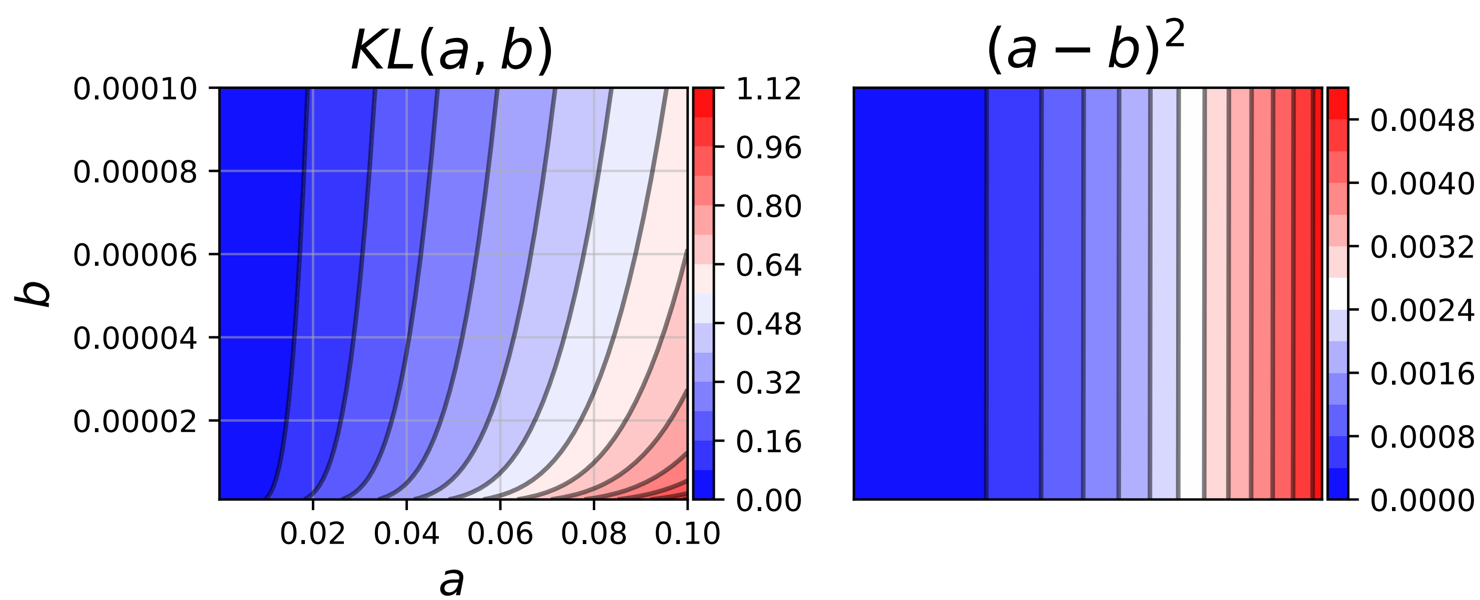
$$\mathrm{KL}(a,b) = a\log(a/b) - a + b = D_\phi(a,b)$$

Shannon–Boltzman entropy $\phi(x) = x\log(x) - x + 1$

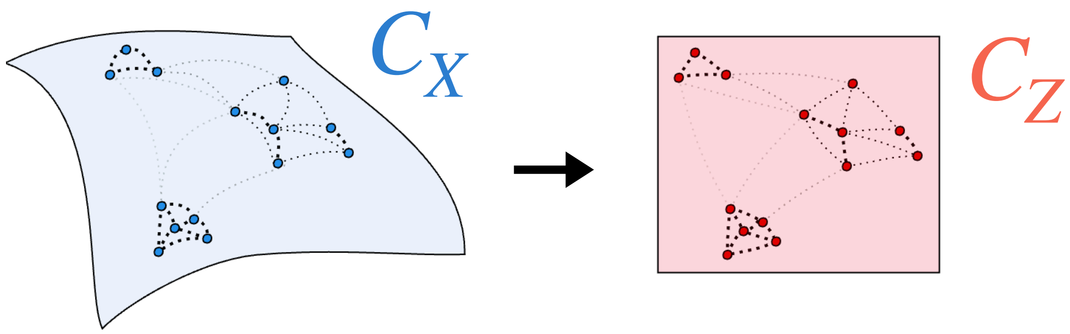




(Shekhar et al., 2016)



Dimension reduction

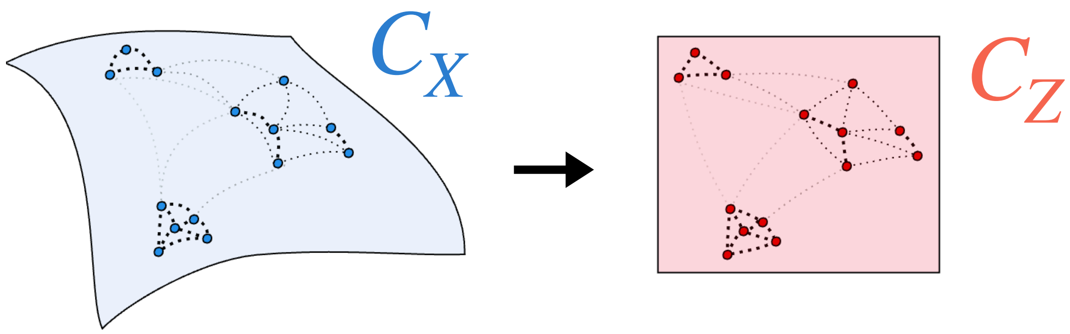


✦ **SNE** (Hinton & Roweis, 2002)

Embedding space

$$[C_Z]_{ij} = \frac{\exp(-\|\mathbf{z}_i - \mathbf{z}_j\|_2^2)}{\sum_k \exp(-\|\mathbf{z}_i - \mathbf{z}_k\|_2^2)} (= \mathbb{P}(j \mid i))$$

Dimension reduction



✦ **SNE** (Hinton & Roweis, 2002)

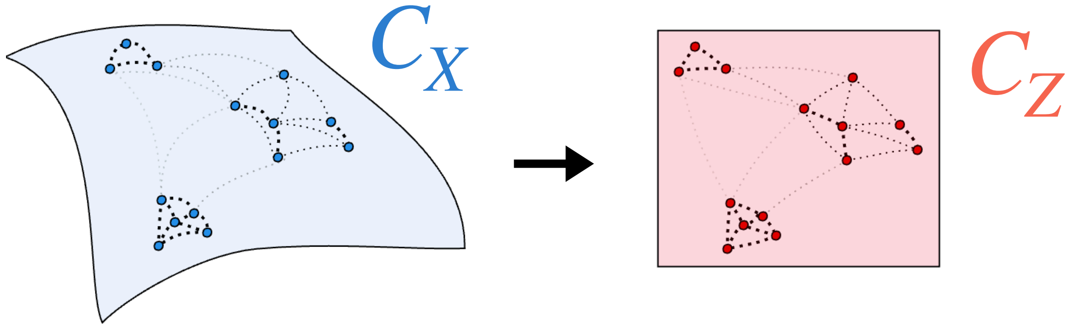
Input space

$$[C_X]_{ij} = \frac{\exp(-\|\mathbf{x}_i - \mathbf{x}_j\|_2^2 / 2\sigma_i^2)}{\sum_k \exp(-\|\mathbf{x}_i - \mathbf{x}_k\|_2^2 / 2\sigma_i^2)}$$

Embedding space

$$[C_Z]_{ij} = \frac{\exp(-\|\mathbf{z}_i - \mathbf{z}_j\|_2^2)}{\sum_k \exp(-\|\mathbf{z}_i - \mathbf{z}_k\|_2^2)} (= \mathbb{P}(j \mid i))$$

Dimension reduction



✦ **SNE** (Hinton & Roweis, 2002)

Input space

$$[C_X]_{ij} = \frac{\exp(-\|\mathbf{x}_i - \mathbf{x}_j\|_2^2 / 2\sigma_i^2)}{\sum_k \exp(-\|\mathbf{x}_i - \mathbf{x}_k\|_2^2 / 2\sigma_i^2)}$$

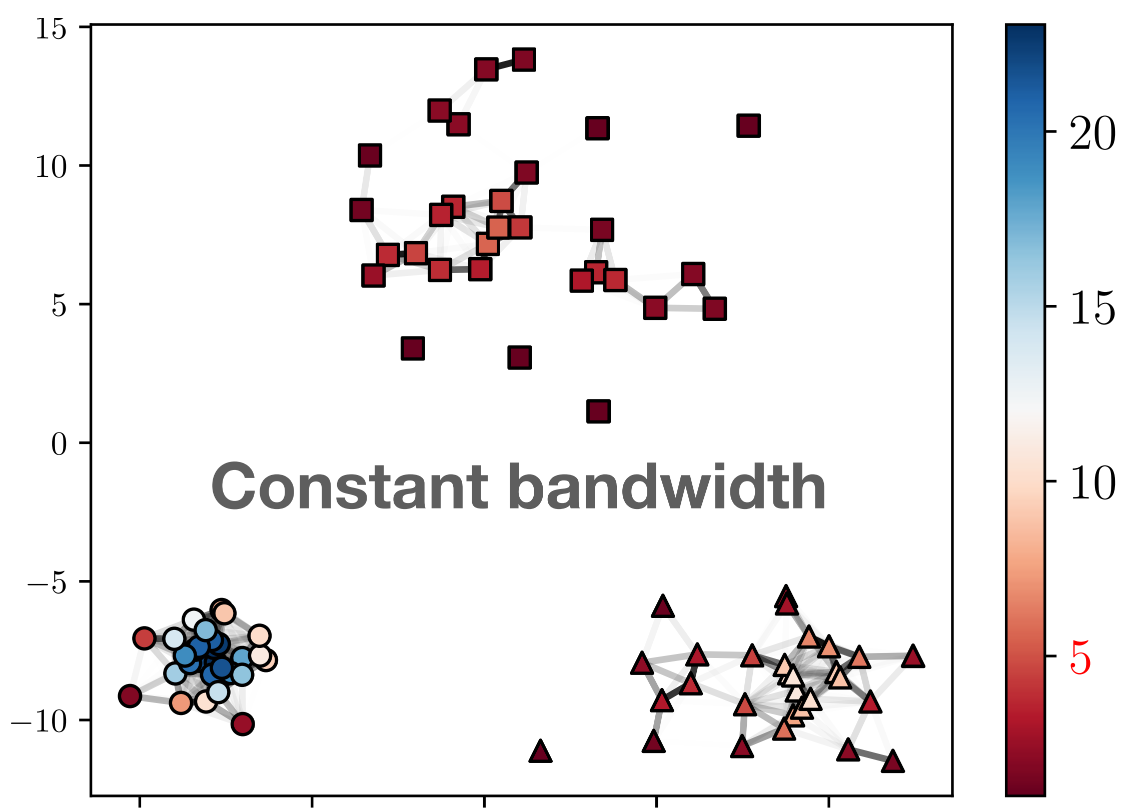
- Local bandwidths **optimized** s.t.
  $\forall i$, entropy($[C_X]_{i,:}$) = log(perplexity)
- Perplexity = effective number of **neighbors**
- Account for **varying density**

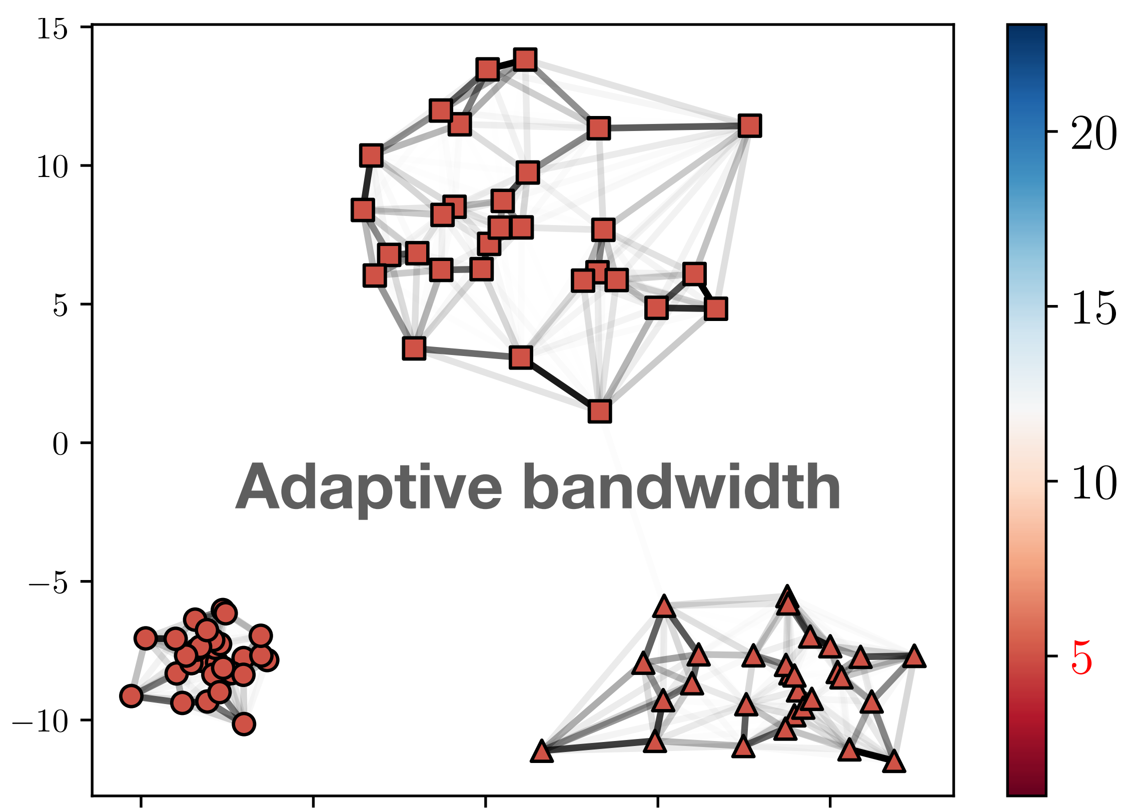
Embedding space

$$[C_Z]_{ij} = \frac{\exp(-\|\mathbf{z}_i - \mathbf{z}_j\|_2^2)}{\sum_k \exp(-\|\mathbf{z}_i - \mathbf{z}_k\|_2^2)} (= \mathbb{P}(j \mid i))$$

✦ **(t)-SNE** (Van der Maaten & Hinton, 2008)

- Crowding effect: Student t-distribution instead of Gaussian in Z

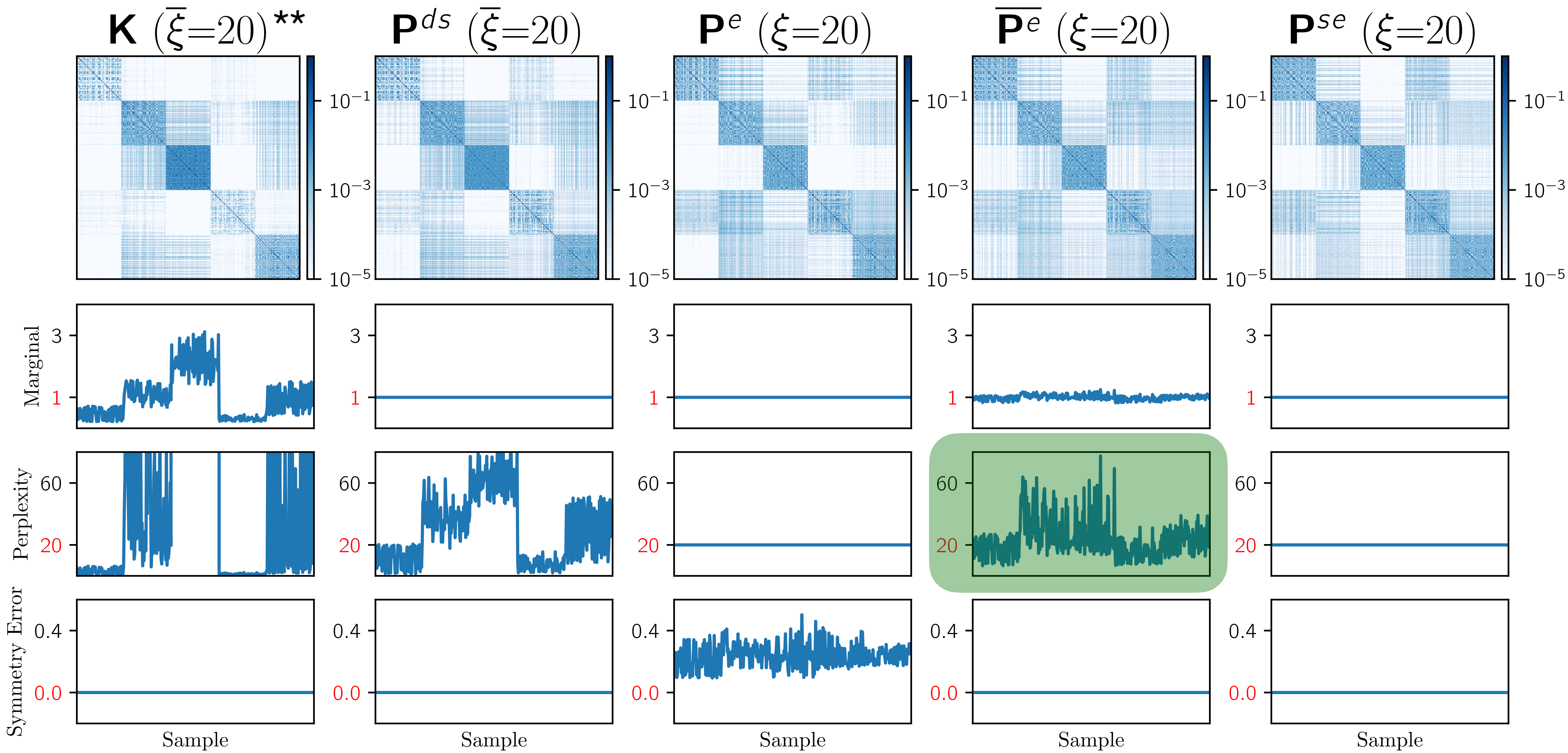




Dimension reduction

(Van Assel et al., 2023)

	Doubly-Sto	SNE	t-SNE	OURS
Gibbs kernel	**Doubly-Sto**	**Entropic**	**Entropic (ℓ_2 Sym)**	**Sym-Entropic**
$\mathbf{K} = \exp(-\mathbf{C}/\sigma)$	$\mathbf{P}^{\mathrm{ds}} = \mathrm{Proj}^{\mathrm{KL}}_{\mathcal{DS}}(\mathbf{K})$	$\mathbf{P}^{\mathrm{e}} = \mathrm{Proj}^{\mathrm{KL}}_{\mathcal{H}_\xi}(\mathbf{K})$	$\overline{\mathbf{P}^{\mathrm{e}}} = \mathrm{Proj}^{\ell_2}_{\mathcal{S}}(\mathbf{P}^{\mathrm{e}})$	$\mathbf{P}^{\mathrm{se}} = \mathrm{Proj}^{\mathrm{KL}}_{\mathcal{H}_\xi \cap \mathcal{S}}(\mathbf{K})$

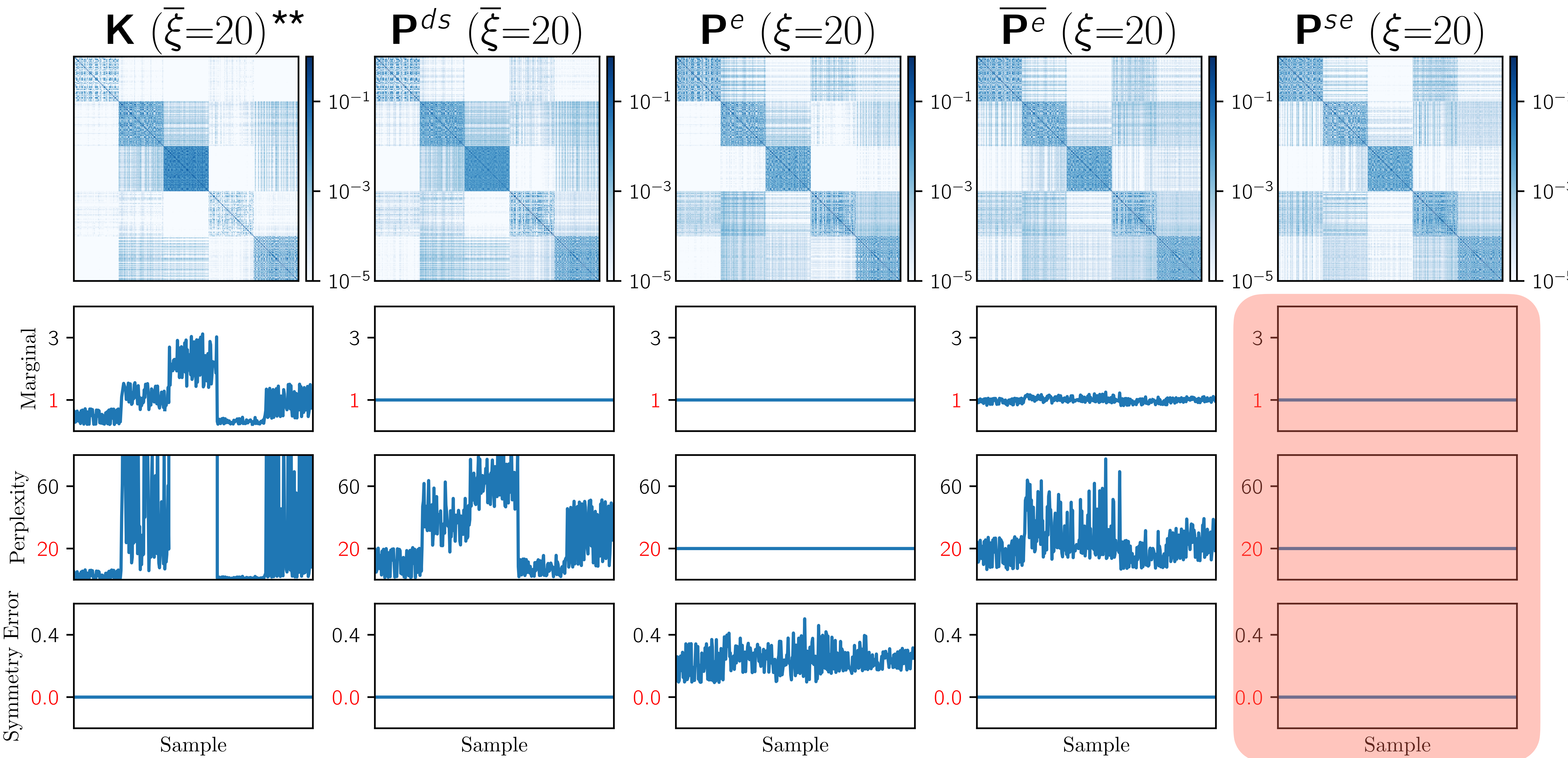


→ **t-SNE fails at controlling the entropy when symmetrizing**

Dimension reduction

(Van Assel et al., 2023)

	SNE	t-SNE	OURS	
Gibbs kernel	**Doubly-Sto**	**Entropic**	**Entropic (ℓ_2 Sym)**	**Sym-Entropic**
$\mathbf{K} = \exp(-\mathbf{C}/\sigma)$	$\mathbf{P}^{\mathrm{ds}} = \mathrm{Proj}^{\mathrm{KL}}_{\mathcal{DS}}(\mathbf{K})$	$\mathbf{P}^{\mathrm{e}} = \mathrm{Proj}^{\mathrm{KL}}_{\mathcal{H}_\xi}(\mathbf{K})$	$\overline{\mathbf{P}^{\mathrm{e}}} = \mathrm{Proj}^{\ell_2}_{\mathcal{S}}(\mathbf{P}^{\mathrm{e}})$	$\mathbf{P}^{\mathrm{se}} = \mathrm{Proj}^{\mathrm{KL}}_{\mathcal{H}_\xi \cap \mathcal{S}}(\mathbf{K})$



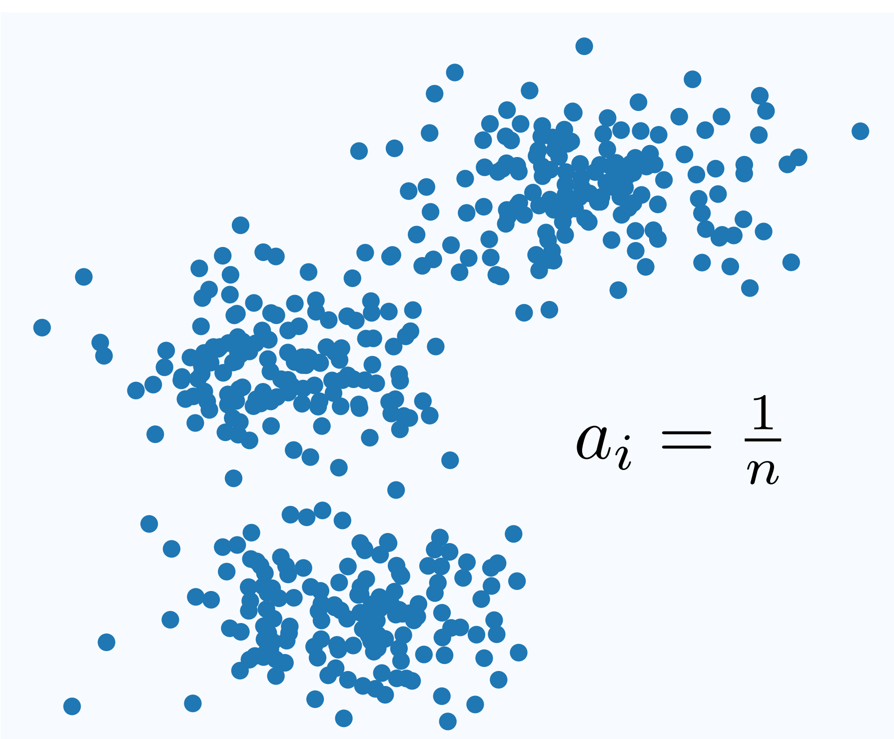
➡ Controls ℓ_1 norm, entropy and symmetry at the same time.

Framing DR in terms of distributions

Measure and probability distributions are at the core of Machine learning.

Data: $(\mathbf{x}_i)_{i \in [\![n]\!]}$; $\mathbf{x}_i \in \mathbb{R}^d$ ⟶ A probability distribution describing the data

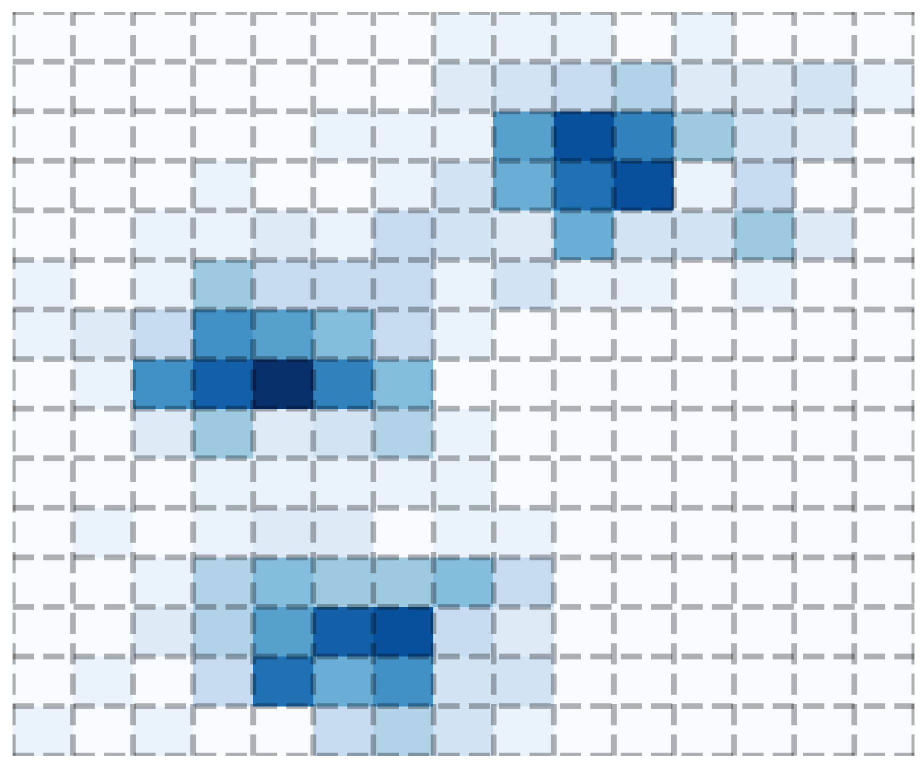
Lagrangian: $\sum_{i=1}^{n} a_i \delta_{x_i}$



(point clouds)

$\delta_{\mathbf{x}_i}(\mathbf{x}) = 1 \text{ if } \mathbf{x} = \mathbf{x}_i \text{ else } 0$

Eulerian: $\sum_{i=1}^{N} a_i \delta_{\hat{x}_i}$



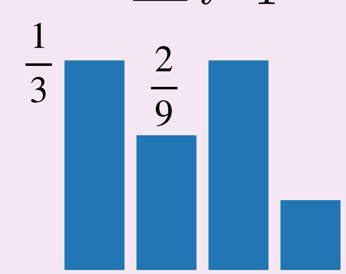
(histograms)

$\hat{x}_i$ fixed position (grid)

Probability simplex

$\mathbf{a} = (a_i)_{i \in [\![n]\!]} \in \Sigma_n$

$a_i \geq 0, \sum_{i=1}^{n} a_i = 1$



23

Framing DR in terms of distributions

Measure and probability distributions are at the core of Machine learning.

A point of view on the data

Data: $(\mathbf{x}_i)_{i \in [\![n]\!]}\ ; \mathbf{x}_i \in \mathbb{R}^d \longrightarrow$ A probability distribution describing the data

A formalism for many machine learning paradigms

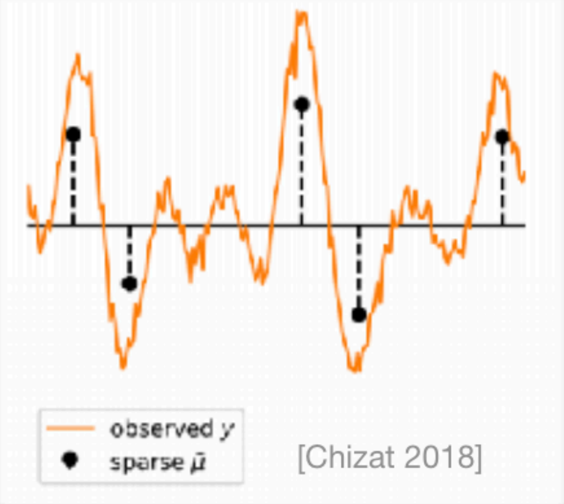
(ERM) $\min_f \mathop{\mathbb{E}}_{(x,y)\sim\mu}[L(f(x),y)]$ $\quad \mu = \frac{1}{n}\sum_{i=1}^n \delta_{(\mathbf{x}_i,y_i)}$

(Likehood) $\max_{\boldsymbol{\theta}\in\Theta} \mathop{\mathbb{E}}_{\mathbf{x}\sim\mu}[\log(\mathbb{P}_{\boldsymbol{\theta}}(\mathbf{x}))]$ $\quad \mu = \frac{1}{n}\sum_{i=1}^n \delta_{\mathbf{x}_i}$

(GAN) $\min_{\boldsymbol{\theta}\in\Theta} D(\mu_{\boldsymbol{\theta}},\nu)$

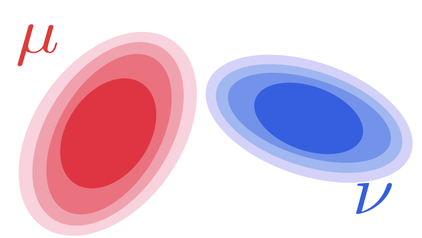
(Signal processing) Recover a sparse signal

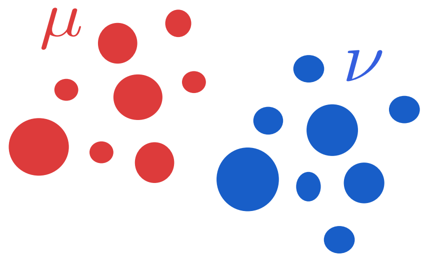
$$\min_{\mu\in\mathcal{M}(\Theta)} \frac{1}{2}\|\mathbf{y}-\phi*\mu\|_{L^2}^2 + R(\mu) \qquad \overline{\mu} = \sum_i w_i \delta_{\boldsymbol{\theta}_i}$$

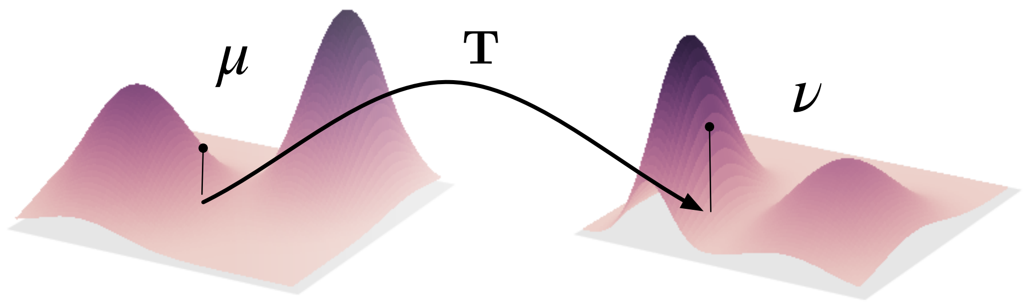


[Chizat 2018]

From linear Optimal Transport to Gromov-Wasserstein

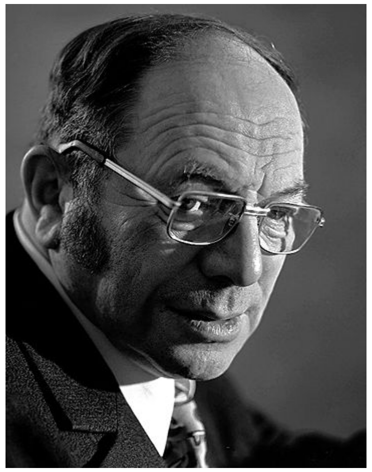






From Wasserstein to Gromov-Wasserstein

✦ Classical optimal transport (in a nutshell)



Kantorovich Formulation

Two probability distributions

$\mu \in \mathscr{P}(\mathscr{X})$

$\nu \in \mathscr{P}(\mathscr{Y})$

A cost function

$c(x, y) : \mathscr{X} \times \mathscr{Y} \to \mathbb{R}$

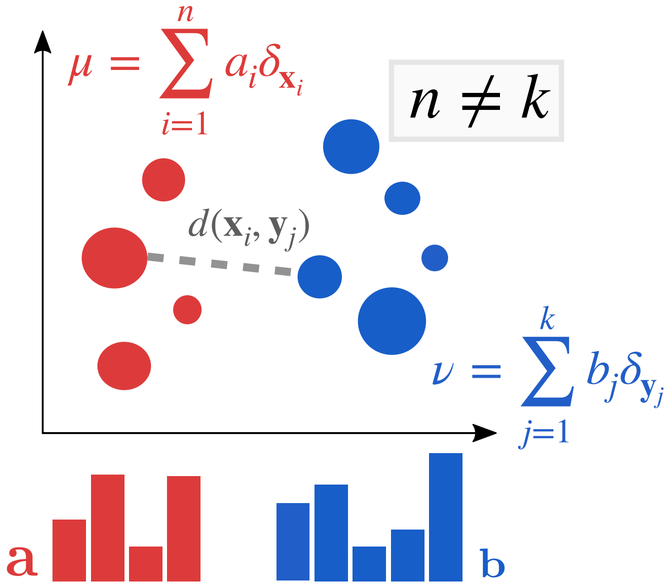
$\mu \in \mathscr{P}(\mathscr{X})$ is **transported** to $\nu \in \mathscr{P}(\mathscr{Y})$ by a **transport plan** $T \in \mathscr{P}(\mathscr{X} \times \mathscr{Y})$

We want to find the plan that **minimizes the overall cost** of moving all the points.

$$\inf_{T \in \Pi(\mu,\nu)} \int c(x, y) dT(x, y)$$

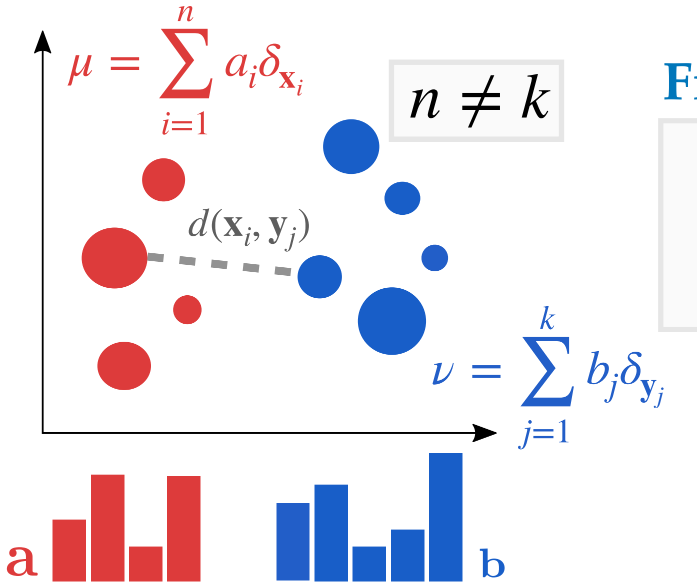
From Wasserstein to Gromov-Wasserstein

✦ Classical optimal transport (in a nutshell)



From Wasserstein to Gromov-Wasserstein

✦ Classical optimal transport (in a nutshell)

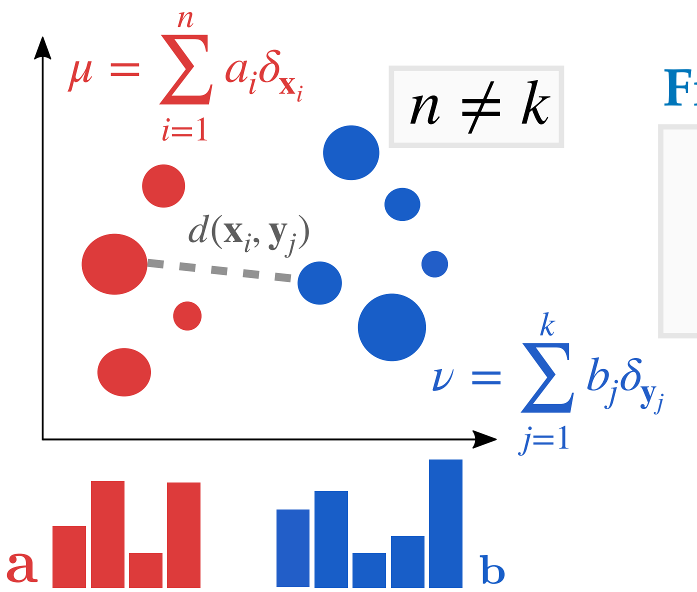


Find plan $T \in \mathbb{R}_+^{n \times k}$

$$\min_T \sum_{ij} c(\mathbf{x}_i, \mathbf{y}_j) T_{ij}$$

From Wasserstein to Gromov-Wasserstein

✦ Classical optimal transport (in a nutshell)



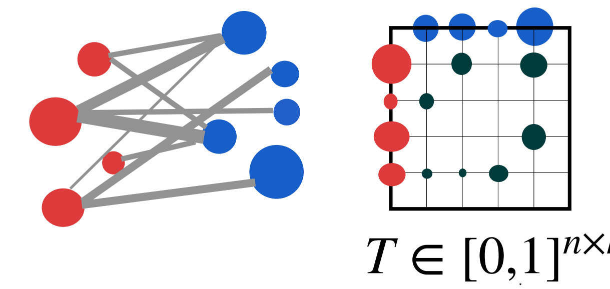
Find plan $T \in \mathbb{R}_+^{n \times k}$

$$\min_T \sum_{ij} c(\mathbf{x}_i, \mathbf{y}_j) T_{ij}$$

which constraints ?

Coupling $\Pi(a, b)$

$$T^\top 1_n = a$$
$$T 1_k = b$$



Bakeries = quantity of breads

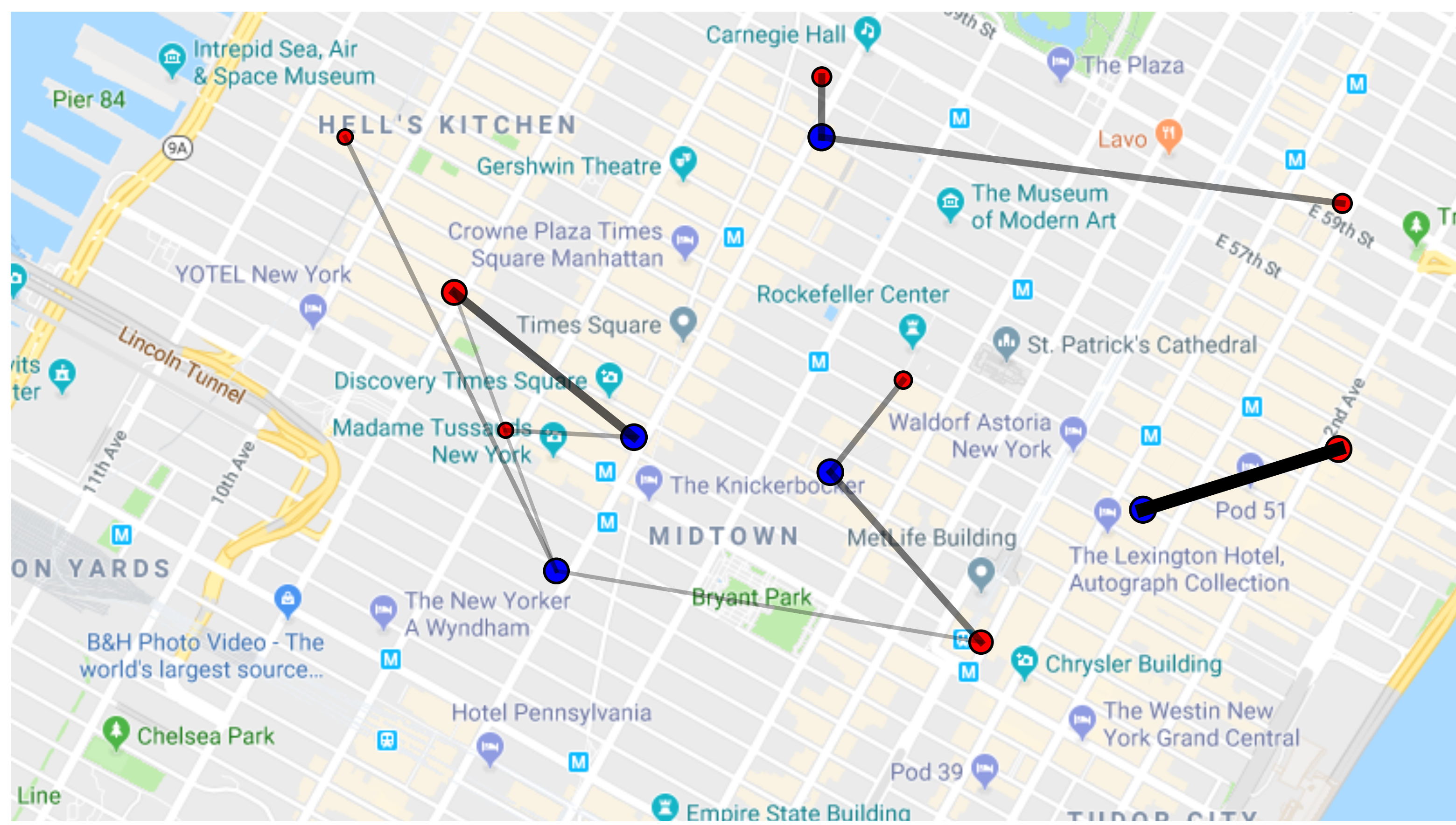
loc: $\mathbf{x}_i$ quantity: a_i

Cafés = demand of breads

loc: $\mathbf{y}_j$ demand: b_j

Distance between bakeries and cafés

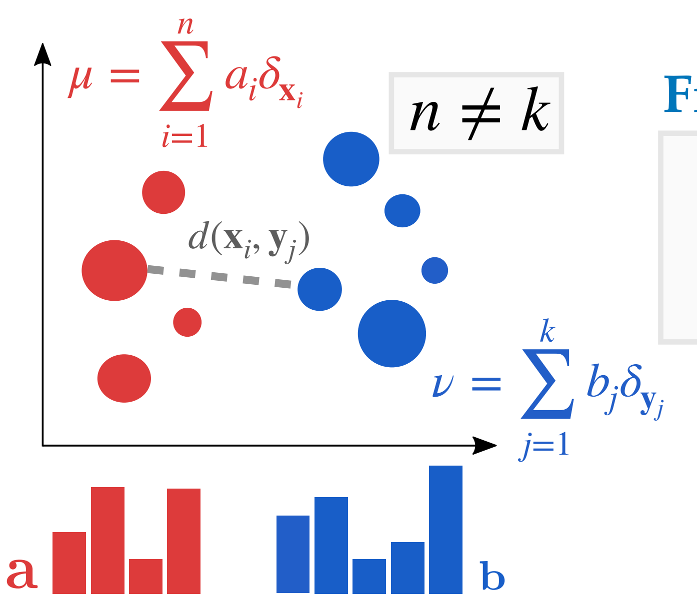
$$c(\mathbf{x}_i, \mathbf{y}_j)$$



We want to route all the breads from bakeries to cafés the cheapest way

From Wasserstein to Gromov-Wasserstein

✦ Classical optimal transport (in a nutshell)



$n \neq k$

Find plan $T \in \mathbb{R}_+^{n \times k}$

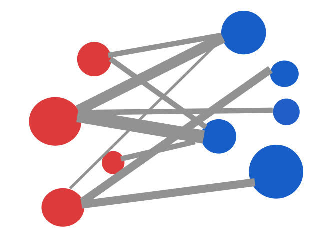
$$\min_T \sum_{ij} c(\mathbf{x}_i, \mathbf{y}_j) T_{ij}$$

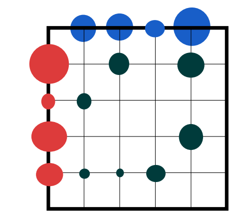
which constraints ?

Coupling $\Pi(a, b)$

$T^\top 1_n = a$

$T 1_k = b$





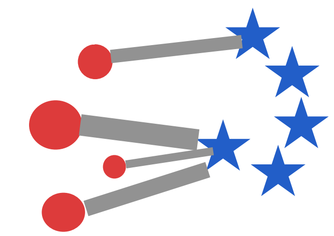
$T \in [0,1]^{n \times k}$

Semi-relaxed coupling

$T^\top 1_n = a$

$\sim$ `assign` in k-means

$\sim \min_b \min_{T \in \Pi(a,b)} \sum_{ij} c(\mathbf{x}_i, \mathbf{y}_j) T_{ij}$



Bakeries = quantity of breads

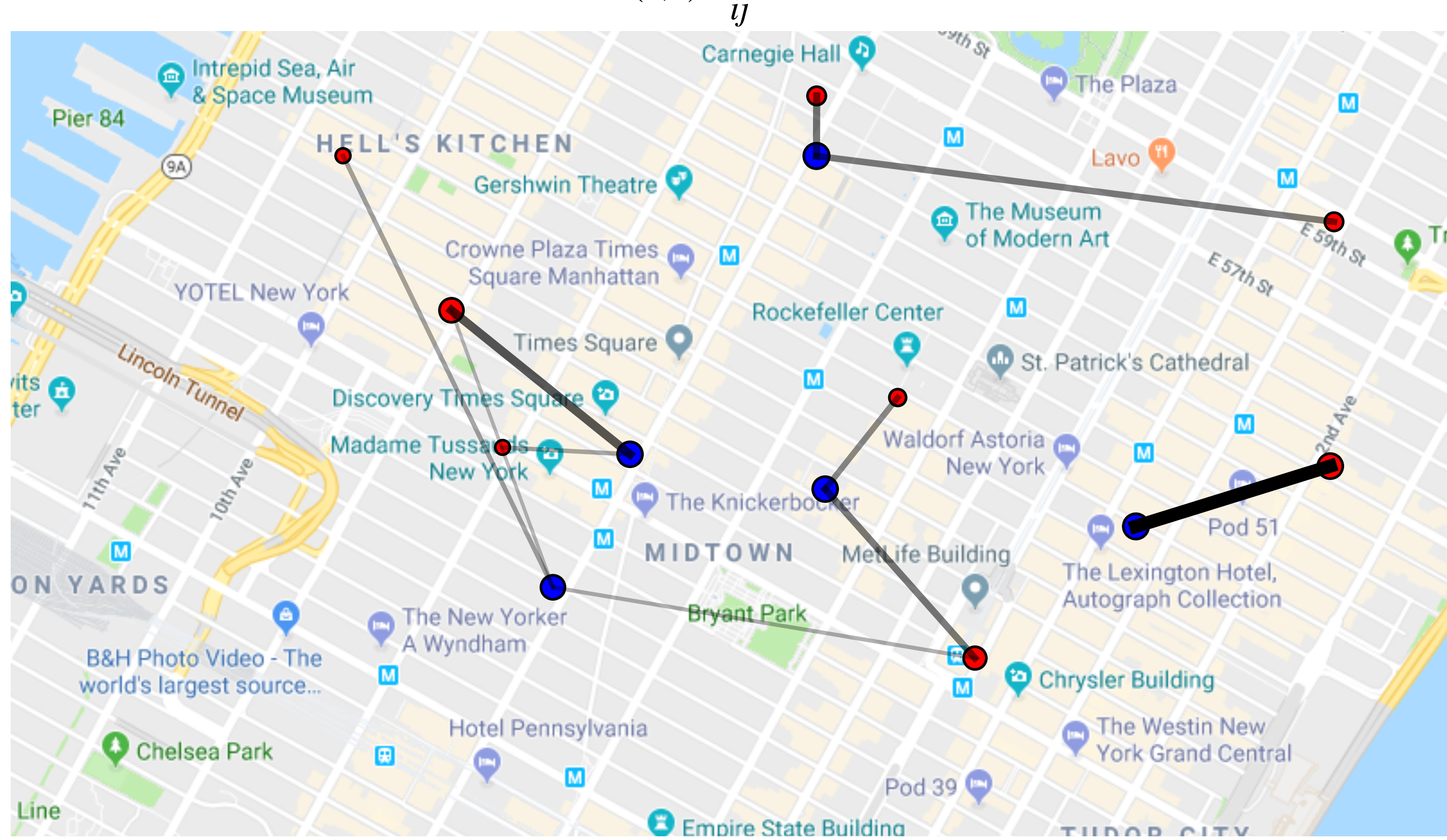
loc: $\mathbf{x}_i$ quantity: a_i

Cafés = demand of breads

loc: $\mathbf{y}_j$ ~~demand: b_j~~

Distance between bakeries and cafés

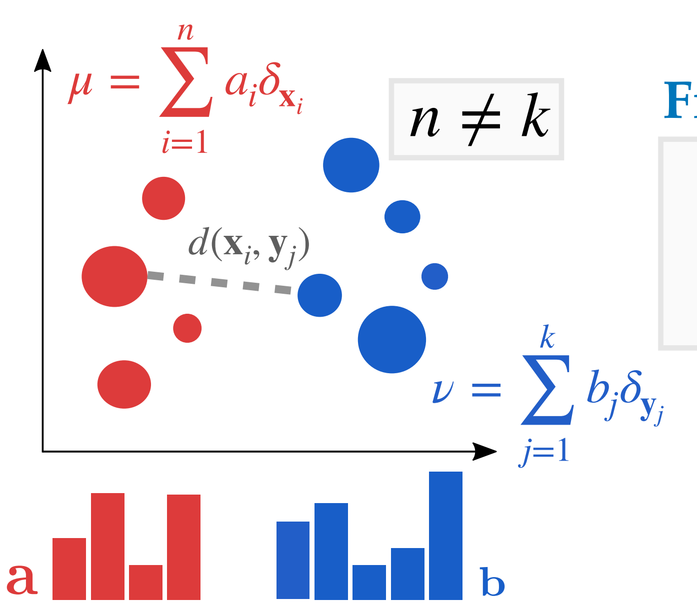
$$c(\mathbf{x}_i, \mathbf{y}_j)$$



We want to route all the breads from bakeries to cafés the cheapest way

From Wasserstein to Gromov-Wasserstein

✦ Classical optimal transport (in a nutshell)



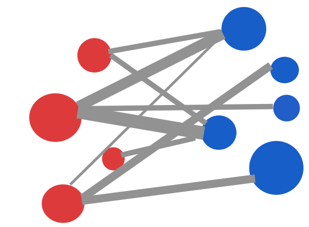
Find plan $T \in \mathbb{R}_+^{n \times k}$

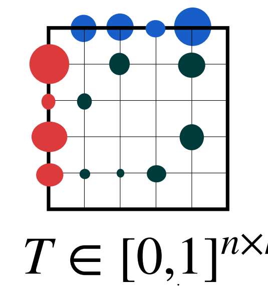
$$\min_{T} \sum_{ij} c(\mathbf{x}_i, \mathbf{y}_j) T_{ij}$$

which constraints ?

Coupling $\Pi(a, b)$

$$T^\top 1_n = a$$
$$T 1_k = b$$



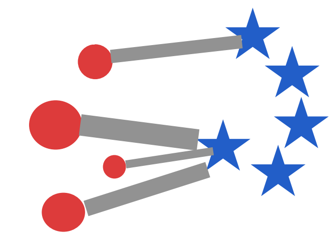


Semi-relaxed coupling

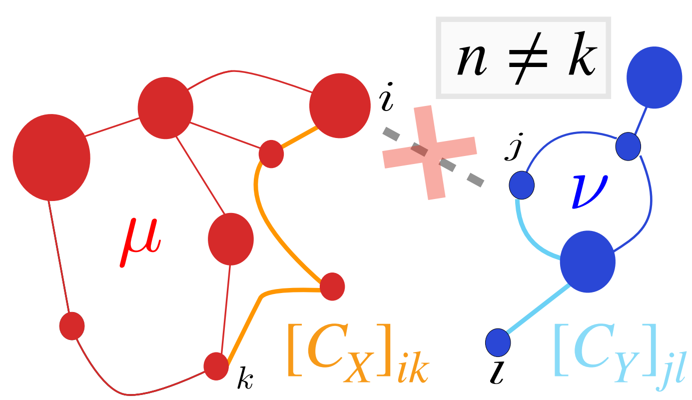
$$T^\top 1_n = a$$

$\sim$ `assign` in k-means

$$\sim \min_{b} \min_{T \in \Pi(a,b)} \sum_{ij} c(\mathbf{x}_i, \mathbf{y}_j) T_{ij}$$



✦ Gromov-Wasserstein

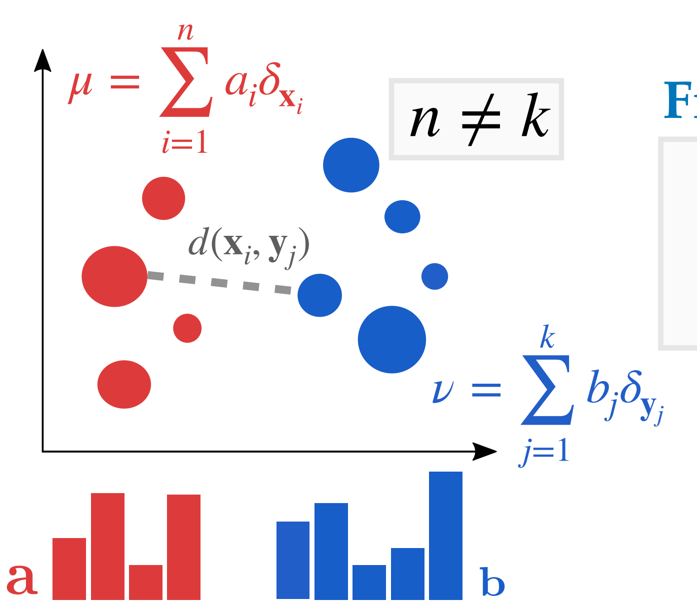


(Sturm, 2012)

(Mémoli, 2011)

From Wasserstein to Gromov-Wasserstein

✦ Classical optimal transport (in a nutshell)



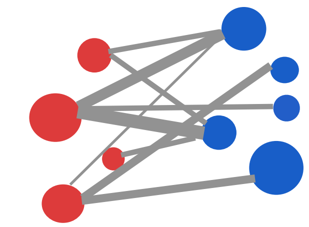
Find plan $T \in \mathbb{R}_+^{n \times k}$

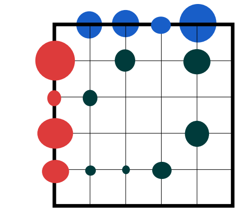
$$\min_{T} \sum_{ij} c(\mathbf{x}_i, \mathbf{y}_j) T_{ij}$$

which constraints ?

→ **Coupling** $\Pi(a, b)$

$$T^\top 1_n = a$$
$$T 1_k = b$$





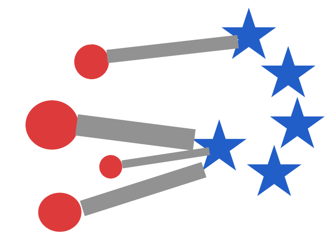
$T \in [0,1]^{n \times k}$

→ **Semi-relaxed coupling**

$$T^\top 1_n = a$$

$\sim$ `assign` in k-means

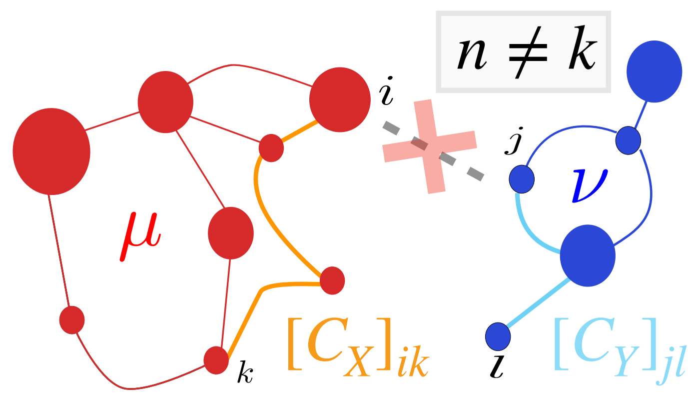
$$\sim \min_{b} \min_{T \in \Pi(a,b)} \sum_{ij} c(\mathbf{x}_i, \mathbf{y}_j) T_{ij}$$



(Sturm, 2012)

(Mémoli, 2011)

✦ Gromov-Wasserstein

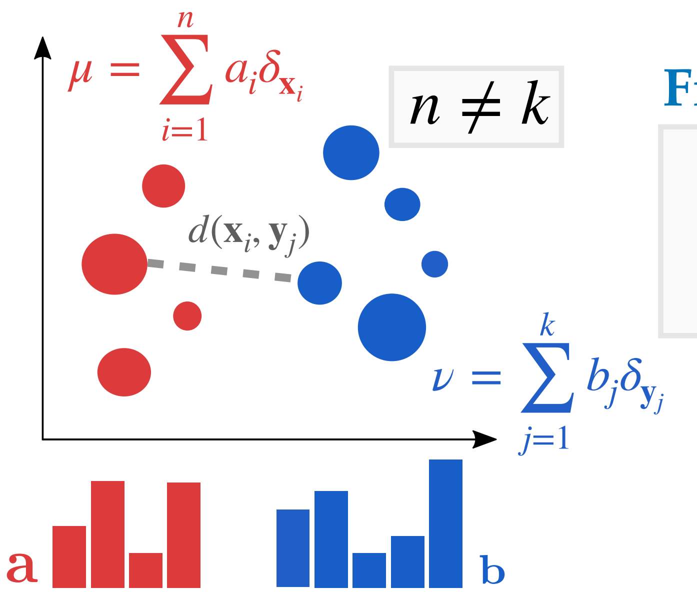


Quadratic OT: find the plan

$$\min_{T \in \Pi(a,b)} \sum_{ijkl} L\left([C_X]_{ik} \, , \, [C_Y]_{jl} \right) T_{ij} T_{kl}$$

From Wasserstein to Gromov-Wasserstein

✦ Classical optimal transport (in a nutshell)



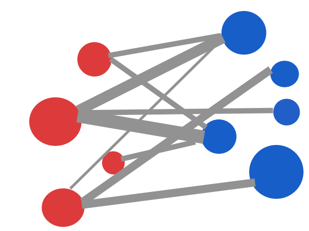
Find plan $T \in \mathbb{R}_+^{n \times k}$

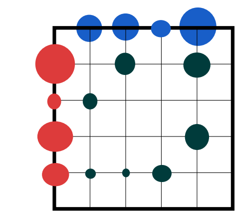
$$\min_T \sum_{ij} c(\mathbf{x}_i, \mathbf{y}_j) T_{ij}$$

which constraints ?

→ **Coupling** $\Pi(a, b)$

$$T^\top 1_n = a \qquad T 1_k = b$$





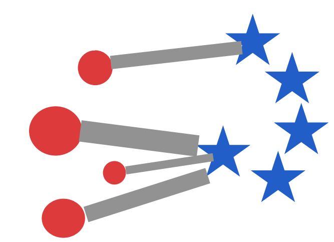
$T \in [0,1]^{n \times k}$

→ **Semi-relaxed coupling**

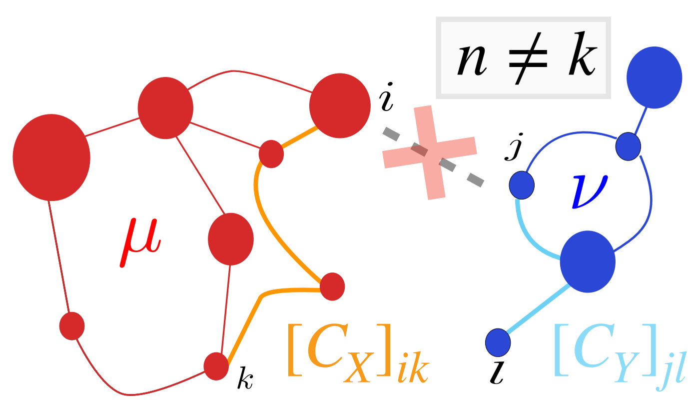
$$T^\top 1_n = a$$

$\sim$ `assign` in k-means

$\sim \min_b \min_{T \in \Pi(a,b)} \sum_{ij} c(\mathbf{x}_i, \mathbf{y}_j) T_{ij}$



✦ Gromov-Wasserstein



(Sturm, 2012) (Mémoli, 2011)

✦ L measures distortion

$$\left| [C_X]_{ik} - [C_Y]_{jl} \right|^2$$

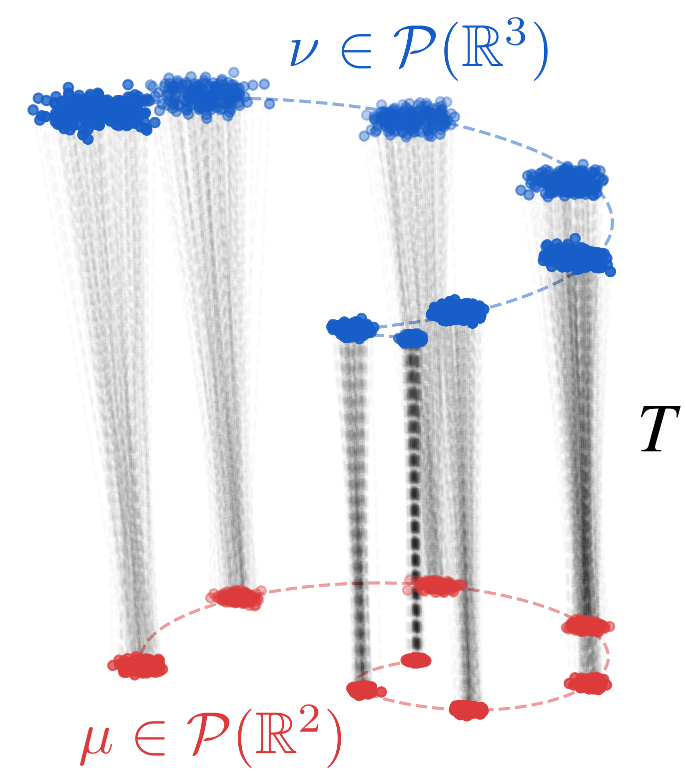
✦ Goal : **preserving pairwise connectivity**

✦ Distance w.r.t. isomorphisms

✦ Difficult quadratic problem (NP-hard)

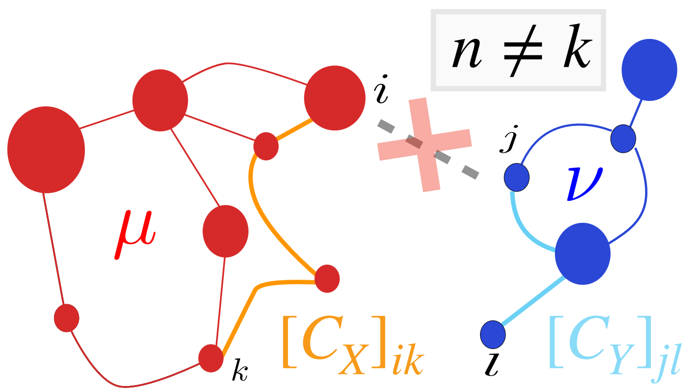
Quadratic OT: find the plan

$$\min_{T \in \Pi(a,b)} \sum_{ijkl} L\left([C_X]_{ik} \, , \, [C_Y]_{jl} \right) T_{ij} T_{kl}$$

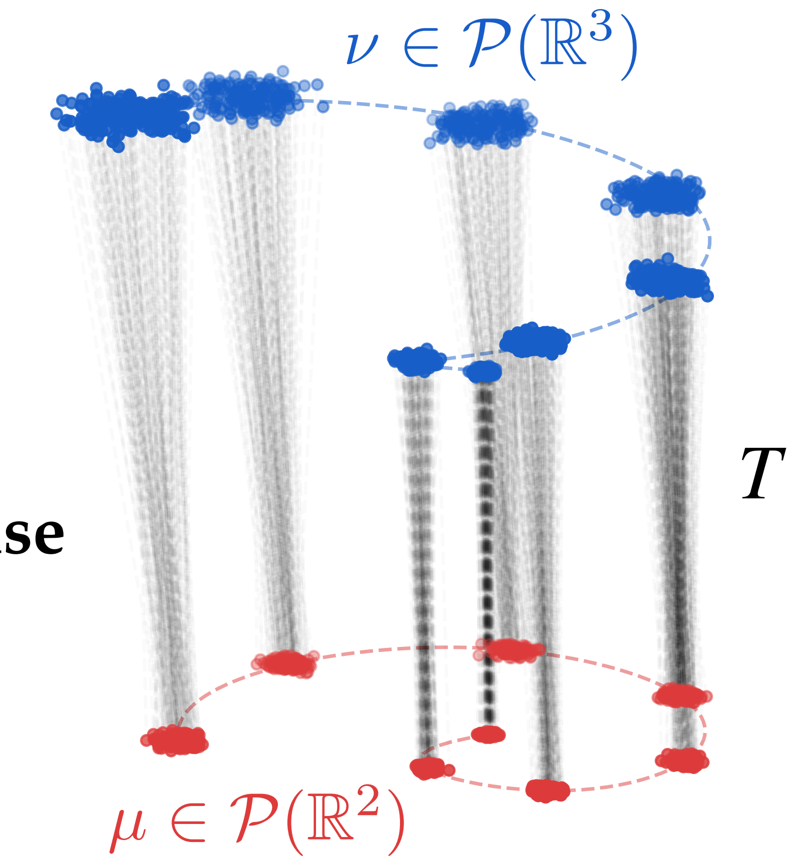


From Wasserstein to Gromov-Wasserstein

✦ **Gromov-Wasserstein**



(Sturm, 2012) (Mémoli, 2011)



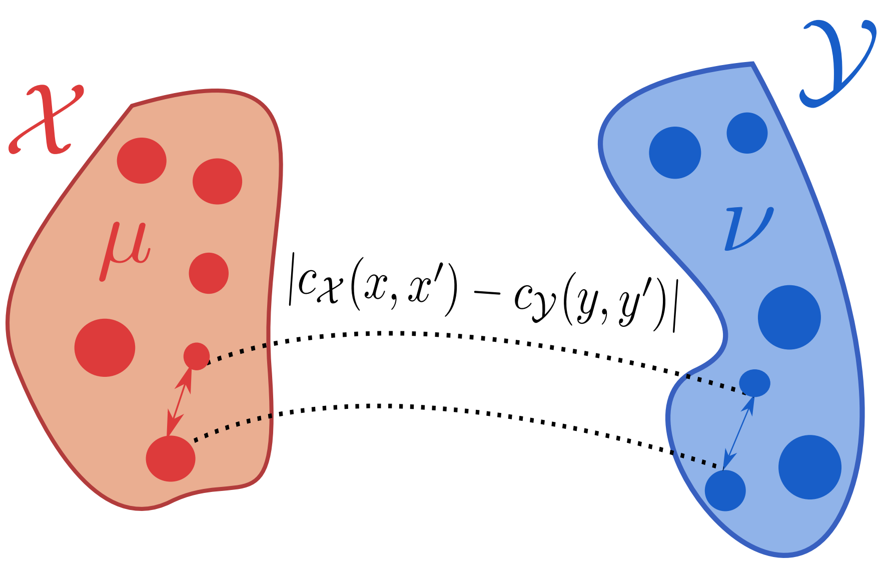
- L measures distorsion

$$\left| [C_X]_{ik} - [C_Y]_{jl} \right|^2$$

- Goal : **preserving pairwise connectivity**
- Non-convex quadratic problem (NP-hard)

Quadratic OT: find the plan

$$\min_{T \in \Pi(a,b)} \sum_{ijkl} L\left([C_X]_{ik} \, , \, [C_Y]_{jl} \right) T_{ij} T_{kl}$$



- Distance w.r.t. isomorphisms, on the space of metric measure spaces

$$\mathbb{X} = (\mathcal{X}, c_{\mathcal{X}}, \mu \in \mathcal{P}(\mathcal{X}))$$

$$\mathbb{Y} = (\mathcal{Y}, c_{\mathcal{Y}}, \nu \in \mathcal{P}(\mathcal{Y}))$$

$GW(\mathbb{X}, \mathbb{Y}) = 0$ iff

$\exists \phi : \mathcal{X} \to \mathcal{Y}$

Isometry

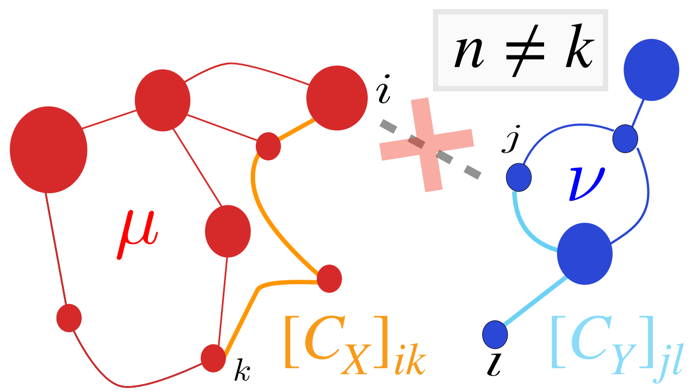
$$c_{\mathcal{X}}(x,x') = c_{\mathcal{Y}}(\phi(x), \phi(x'))$$

Measure preserving

$$\phi \# \mu = \nu$$

From Wasserstein to Gromov-Wasserstein

✦ Gromov-Wasserstein



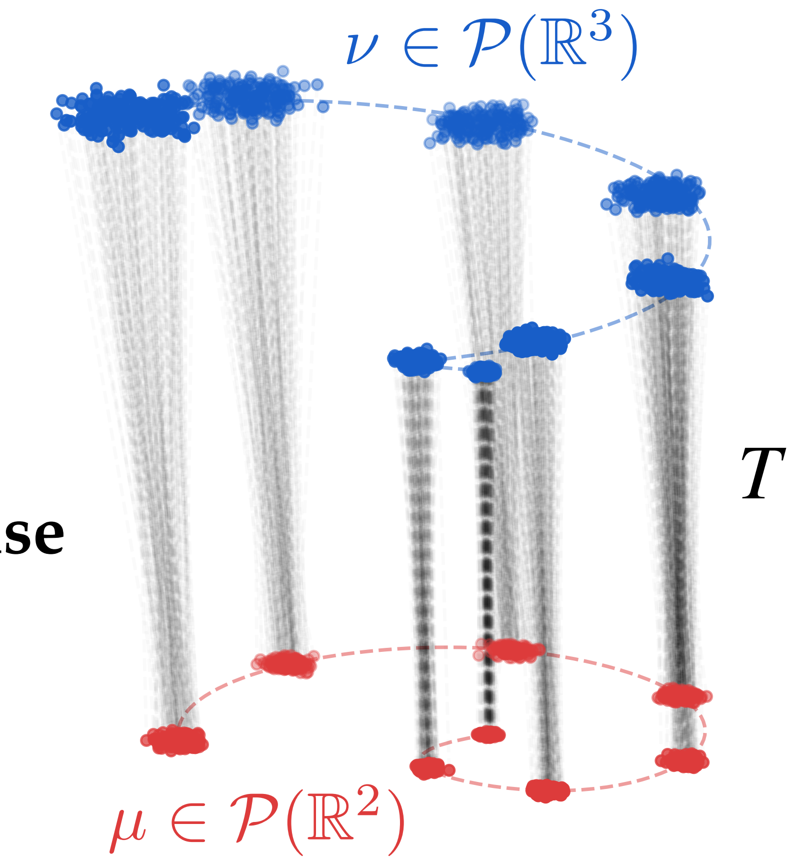
(Sturm, 2012) (Mémoli, 2011)

✦ L measures distortion

$$\left| [C_X]_{ik} - [C_Y]_{jl} \right|^2$$

✦ Goal : **preserving pairwise connectivity**

✦ Non-convex quadratic problem (NP-hard)

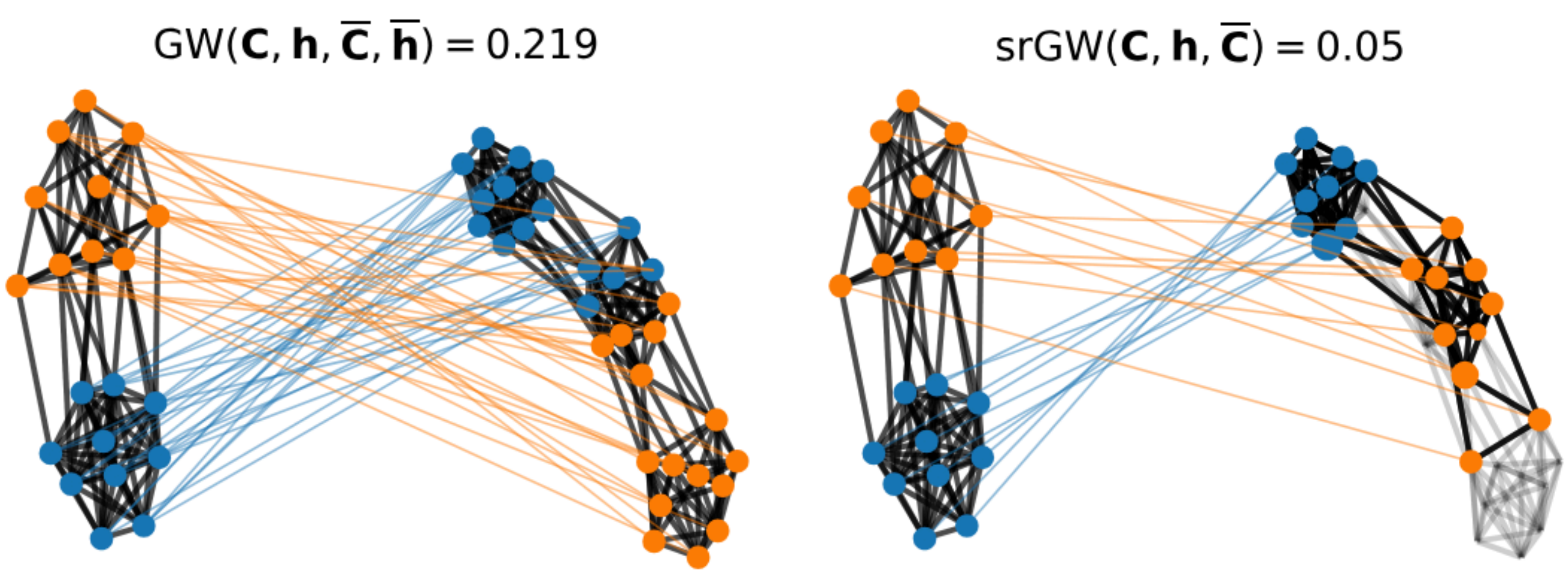


Quadratic OT: find the plan

$$\min_{T \in \Pi(a,b)} \sum_{ijkl} L\left([C_X]_{ik}\ ,\ [C_Y]_{jl}\right) T_{ij} T_{kl}$$

✦ Semi relaxed Gromov-Wasserstein

(Vincent-Cuaz, 2022)

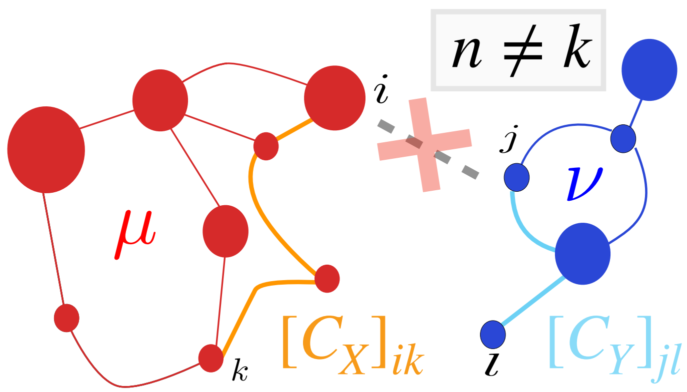


$$\min_{T} \sum_{ijkl} L\left([C_X]_{ik}\ ,\ [C_Y]_{jl}\right) T_{ij} T_{kl}$$

$$\Big|\ T^\top 1_n = a$$

From Wasserstein to Gromov-Wasserstein

✦ Gromov-Wasserstein



(Sturm, 2012) (Mémoli, 2011)

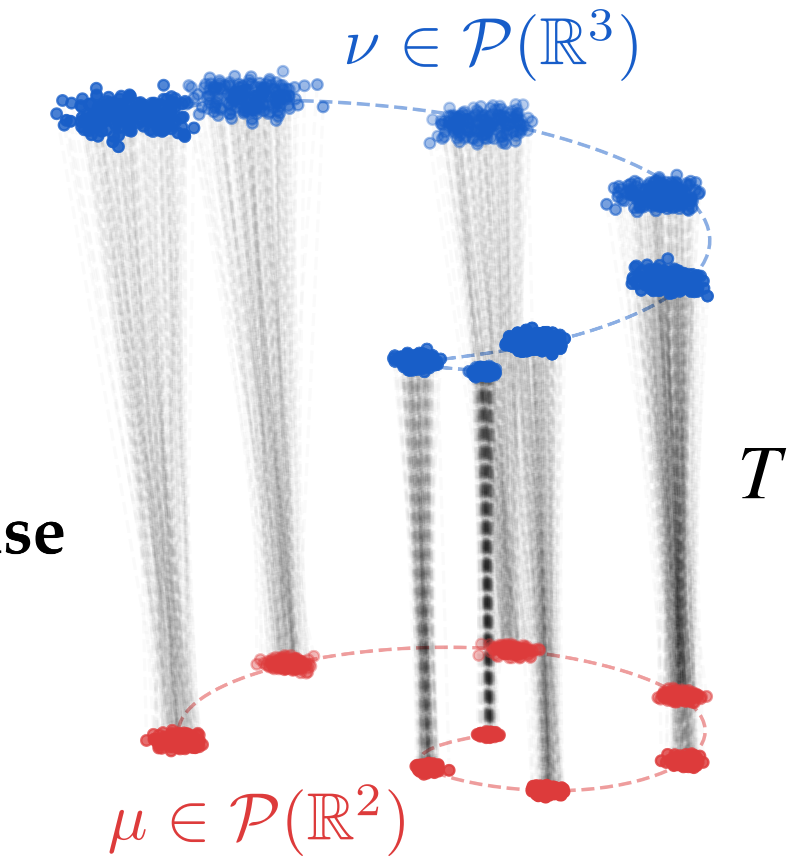
Quadratic OT: find the plan

$$\min_{T\in\Pi(a,b)} \sum_{ijkl} L\left([C_X]_{ik}\ ,\ [C_Y]_{jl}\right) T_{ij}T_{kl}$$

- ✦ L measures distortion

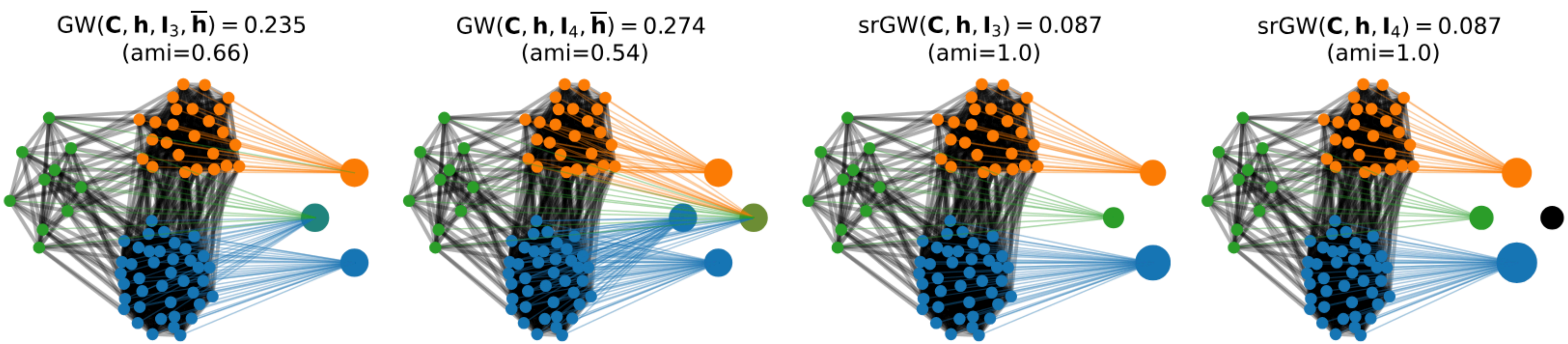
$$\left|[C_X]_{ik} - [C_Y]_{jl}\right|^2$$

- ✦ Goal : **preserving pairwise connectivity**
- ✦ Non-convex quadratic problem (NP-hard)



✦ Semi relaxed Gromov-Wasserstein

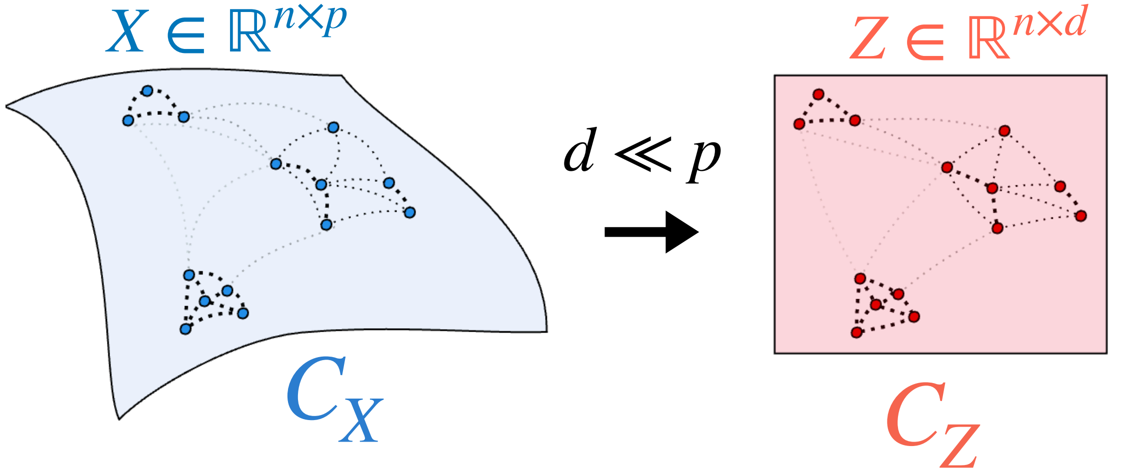
- ✦ Clustering of nodes



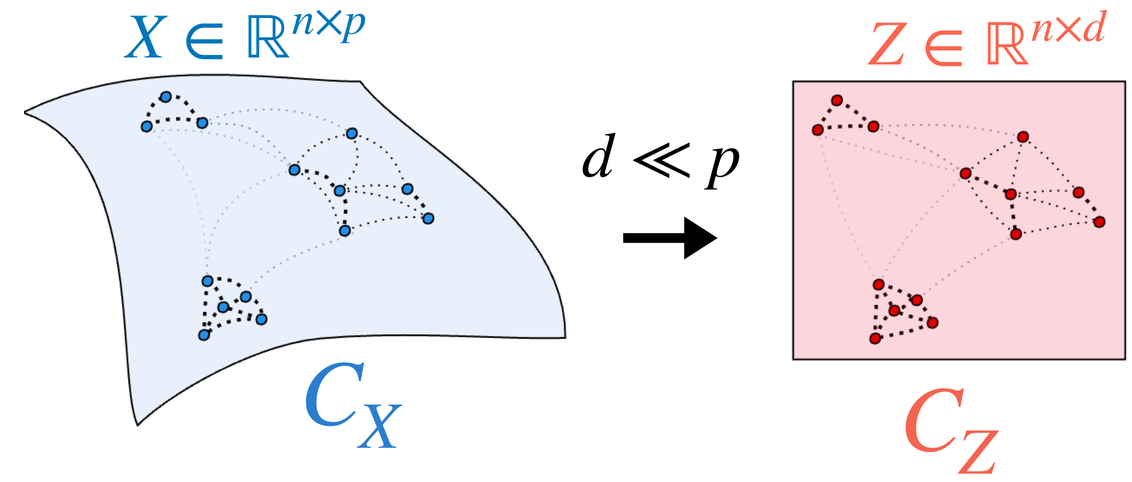
DR as OT in disguise

✦ Dimension reduction

$$\min_{Z \in \mathbb{R}^{n \times d}} \sum_{i,j=1}^{n} L\left([C_X]_{ij} \, , \, [C_Z]_{ij}\right)$$



DR as OT in disguise



✦ Dimension reduction

$$\min_{Z \in \mathbb{R}^{n \times d}} \sum_{i,j=1}^{n} L\left([C_X]_{ij}\ ,\ [C_Z]_{ij}\right)$$

equiv

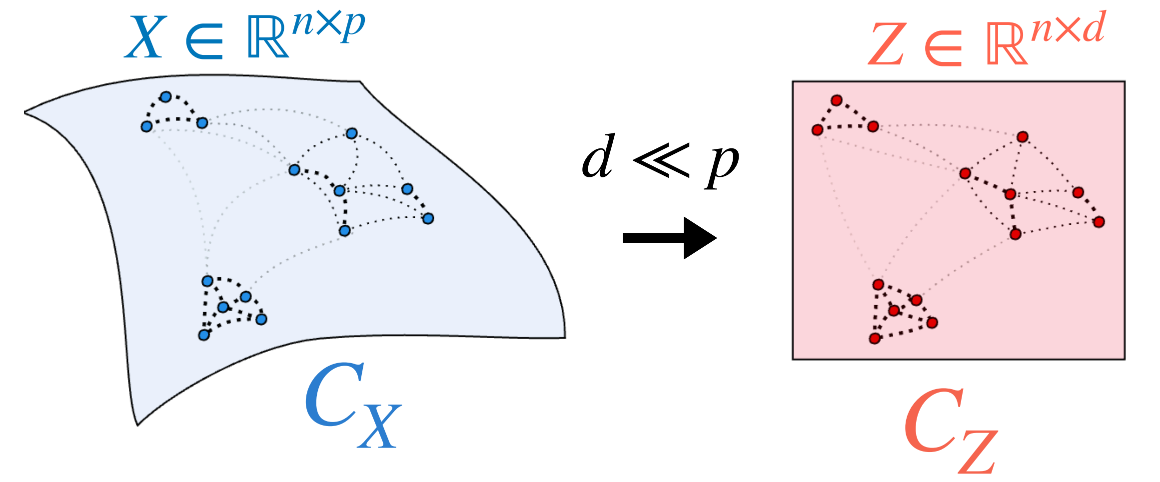
Permutation equivariance

$$\forall P, C_{PZ} = P C_Z P^\top$$

$$\min_{Z \in \mathbb{R}^{n \times d}} \min_{\sigma \in S_n} \sum_{i,j=1}^{n} L\left([C_X]_{ij}\ ,\ [C_Z]_{\sigma(i)\sigma(j)}\right)$$

DR as OT in disguise

✦ Dimension reduction



$$\min_{Z \in \mathbb{R}^{n \times d}} \sum_{i,j=1}^{n} L\left([C_X]_{ij}\ ,\ [C_Z]_{ij}\right)$$

equiv

Permutation equivariance

$\forall P, C_{PZ} = P C_Z P^\top$

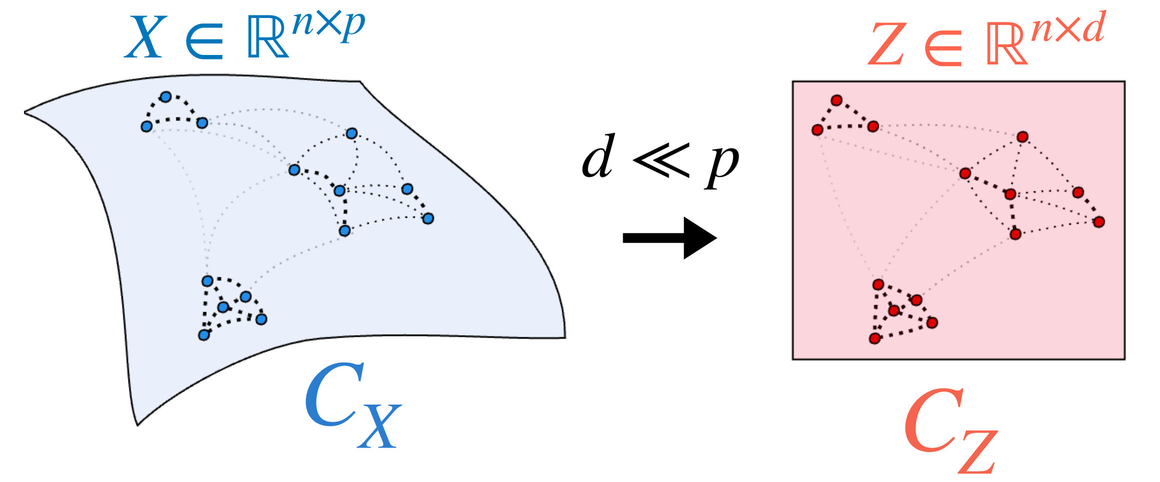
$$\min_{Z \in \mathbb{R}^{n \times d}} \min_{\sigma \in S_n} \sum_{i,j=1}^{n} L\left([C_X]_{ij}\ ,\ [C_Z]_{\sigma(i)\sigma(j)}\right)$$

equiv

$$\min_{Z \in \mathbb{R}^{n \times d}} \min_{P} \sum_{i,j,k,l=1}^{n} L\left([C_X]_{ik}\ ,\ [C_Z]_{jl}\right) P_{ij} P_{kl}$$

Gromov-Monge

DR as OT in disguise



✦ Dimension reduction

$$\min_{Z\in\mathbb{R}^{n\times d}} \sum_{i,j=1}^{n} L\Big([C_X]_{ij}\ ,\ [C_Z]_{ij}\Big)$$

equiv

Permutation equivariance

$\forall P, C_{PZ} = PC_ZP^\top$

$$\min_{Z\in\mathbb{R}^{n\times d}} \min_{\sigma\in S_n} \sum_{i,j=1}^{n} L\Big([C_X]_{ij}\ ,\ [C_Z]_{\sigma(i)\sigma(j)}\Big)$$

equiv

$$\min_{Z\in\mathbb{R}^{n\times d}} \min_{P} \sum_{i,j,k,l=1}^{n} L\Big([C_X]_{ik}\ ,\ [C_Z]_{jl}\Big) P_{ij}P_{kl}$$

Gromov-Monge

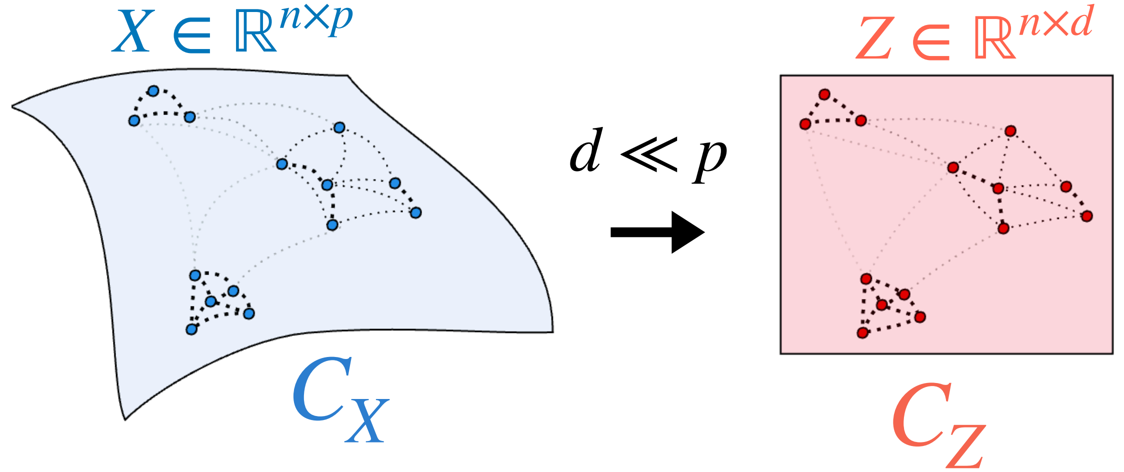
?

✦ Gromov-Wasserstein projection

$$\min_{Z\in\mathbb{R}^{n\times d}} \min_{T\in\Pi(\frac{1_n}{n},\frac{1_n}{n})} \sum_{ijkl} L\Big([C_X]_{ik}\ ,\ [C_Z]_{jl}\Big) T_{ij}T_{kl}$$

Gromov-Wasserstein

DR as OT in disguise



✦ Dimension reduction

$$\min_{Z\in\mathbb{R}^{n\times d}} \sum_{i,j=1}^{n} L\left([C_X]_{ij}\ ,\ [C_Z]_{ij}\right)$$

equiv ↕

Permutation equivariance

$$\forall P, C_{PZ} = PC_ZP^\top$$

$$\min_{Z\in\mathbb{R}^{n\times d}} \min_{\sigma\in S_n} \sum_{i,j=1}^{n} L\left([C_X]_{ij}\ ,\ [C_Z]_{\sigma(i)\sigma(j)}\right)$$

equiv ↕

$$\min_{Z\in\mathbb{R}^{n\times d}} \min_{P} \sum_{i,j,k,l=1}^{n} L\left([C_X]_{ik}\ ,\ [C_Z]_{jl}\right) P_{ij}P_{kl}$$

Gromov-Monge

↕ ?

✦ Gromov-Wasserstein projection

$$\min_{Z\in\mathbb{R}^{n\times d}} \min_{T\in\Pi(\frac{1_n}{n},\frac{1_n}{n})} \sum_{ijkl} L\left([C_X]_{ik}\ ,\ [C_Z]_{jl}\right) T_{ij}T_{kl}$$

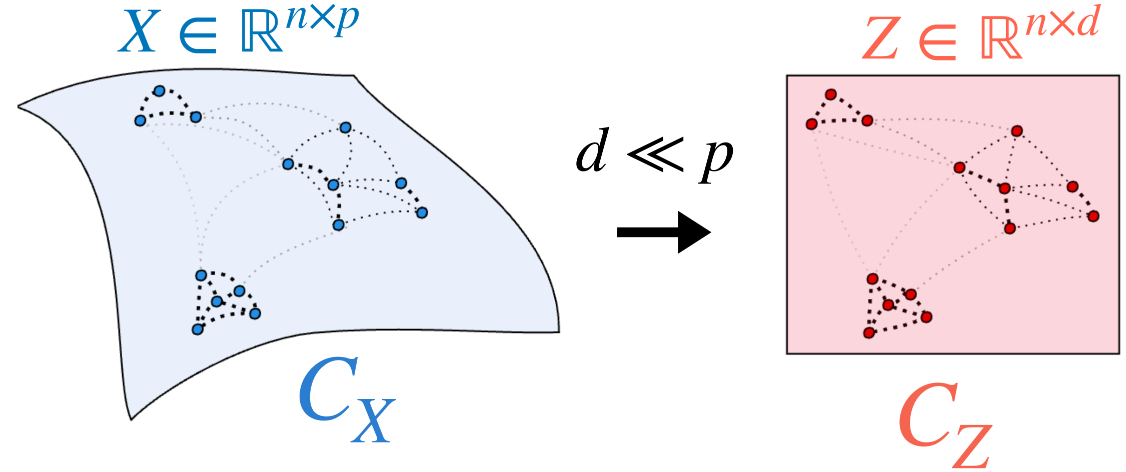
Gromov-Wasserstein

✦ Equivalence holds for

Spectral methods

✦ C_X any matrix, $L = |\cdot|^2$, $C_Z = ZZ^\top$

DR as OT in disguise



✦ Dimension reduction

$$\min_{Z \in \mathbb{R}^{n\times d}} \sum_{i,j=1}^{n} L\left([C_X]_{ij}\ ,\ [C_Z]_{ij}\right)$$

equiv

Permutation equivariance

$\forall P, C_{PZ} = PC_ZP^\top$

$$\min_{Z \in \mathbb{R}^{n\times d}} \min_{\sigma \in S_n} \sum_{i,j=1}^{n} L\left([C_X]_{ij}\ ,\ [C_Z]_{\sigma(i)\sigma(j)}\right)$$

equiv

$$\min_{Z \in \mathbb{R}^{n\times d}} \min_{P} \sum_{i,j,k,l=1}^{n} L\left([C_X]_{ik}\ ,\ [C_Z]_{jl}\right) P_{ij}P_{kl}$$

Gromov-Monge

?

✦ Gromov-Wasserstein projection

$$\min_{Z \in \mathbb{R}^{n\times d}} \min_{T \in \Pi(\frac{1_n}{n}, \frac{1_n}{n})} \sum_{ijkl} L\left([C_X]_{ik}\ ,\ [C_Z]_{jl}\right) T_{ij}T_{kl}$$

Gromov-Wasserstein

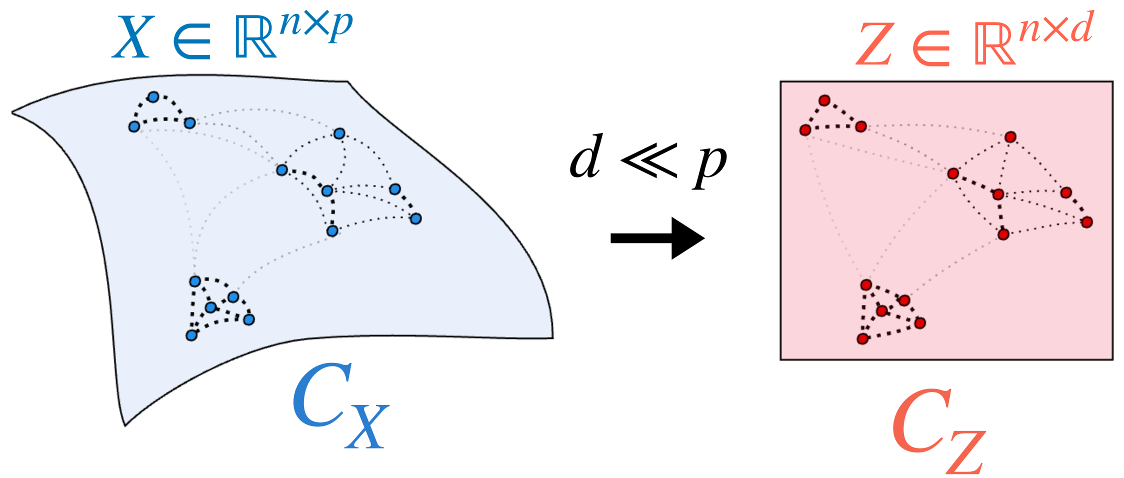
✦ Equivalence holds for

Spectral methods

- C_X any matrix, $L = |\cdot|^2$, $C_Z = ZZ^\top$

A is CPD : $\forall x$ **s.t.** $x^\top 1 = 0,\ x^\top A x \geq 0$

DR as OT in disguise



✦ Dimension reduction

$$\min_{Z\in\mathbb{R}^{n\times d}} \sum_{i,j=1}^{n} L\Big([C_X]_{ij}\ ,\ [C_Z]_{ij}\Big)$$

equiv

Permutation equivariance
$\forall P, C_{PZ} = PC_ZP^\top$

$$\min_{Z\in\mathbb{R}^{n\times d}} \min_{\sigma\in S_n} \sum_{i,j=1}^{n} L\Big([C_X]_{ij}\ ,\ [C_Z]_{\sigma(i)\sigma(j)}\Big)$$

equiv

$$\min_{Z\in\mathbb{R}^{n\times d}} \min_{P} \sum_{i,j,k,l=1}^{n} L\Big([C_X]_{ik}\ ,\ [C_Z]_{jl}\Big) P_{ij}P_{kl}$$

Gromov-Monge

?

✦ Gromov-Wasserstein projection

$$\min_{Z\in\mathbb{R}^{n\times d}} \min_{T\in\Pi(\frac{1_n}{n},\frac{1_n}{n})} \sum_{ijkl} L\Big([C_X]_{ik}\ ,\ [C_Z]_{jl}\Big) T_{ij}T_{kl}$$

Gromov-Wasserstein

✦ Equivalence holds for

Spectral methods

✦ C_X any matrix, $L = |\cdot|^2$, $C_Z = ZZ^\top$

A is CPD : $\forall x$ **s.t.** $x^\top 1 = 0,\ x^\top A x \geq 0$

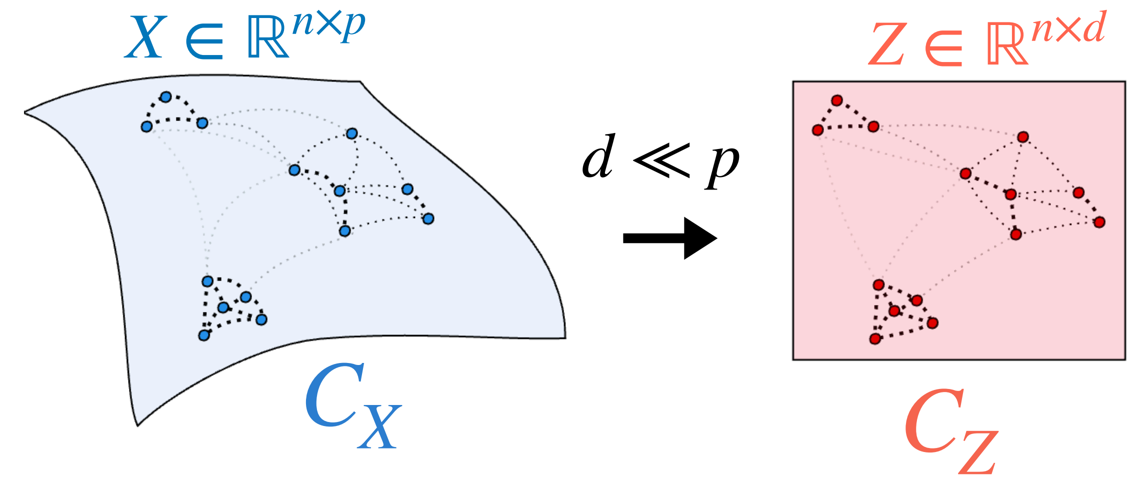
Neighbor embedding methods

✦ C_X is CPD, $L = KL$

$C_Z = \mathrm{diag}(\alpha_Z)\ K_Z\ \mathrm{diag}(\beta_Z)$

where $\log(K_Z)$ is CPD

DR as OT in disguise



✦ Dimension reduction

$$\min_{Z \in \mathbb{R}^{n \times d}} \sum_{i,j=1}^{n} L\Big([C_X]_{ij}\ ,\ [C_Z]_{ij}\Big)$$

equiv

Permutation equivariance

$\forall P, C_{PZ} = P C_Z P^\top$

$$\min_{Z \in \mathbb{R}^{n \times d}} \min_{\sigma \in S_n} \sum_{i,j=1}^{n} L\Big([C_X]_{ij}\ ,\ [C_Z]_{\sigma(i)\sigma(j)}\Big)$$

equiv

$$\min_{Z \in \mathbb{R}^{n \times d}} \min_{P} \sum_{i,j,k,l=1}^{n} L\Big([C_X]_{ik}\ ,\ [C_Z]_{jl}\Big) P_{ij} P_{kl}$$

Gromov-Monge

?

✦ Gromov-Wasserstein projection

$$\min_{Z \in \mathbb{R}^{n \times d}} \min_{T \in \Pi(\frac{1_n}{n}, \frac{1_n}{n})} \sum_{ijkl} L\Big([C_X]_{ik}\ ,\ [C_Z]_{jl}\Big) T_{ij} T_{kl}$$

Gromov-Wasserstein

✦ Equivalence holds for

Spectral methods

✦ C_X any matrix, $L = |\cdot|^2$, $C_Z = ZZ^\top$

A is CPD : $\forall x$ **s.t.** $x^\top 1 = 0,\ x^\top A x \geq 0$

Neighbor embedding methods

✦ C_X is CPD, $L = KL$

$C_Z = \text{diag}(\alpha_Z)\ K_Z\ \text{diag}(\beta_Z)$

where $\log(K_Z)$ is CPD

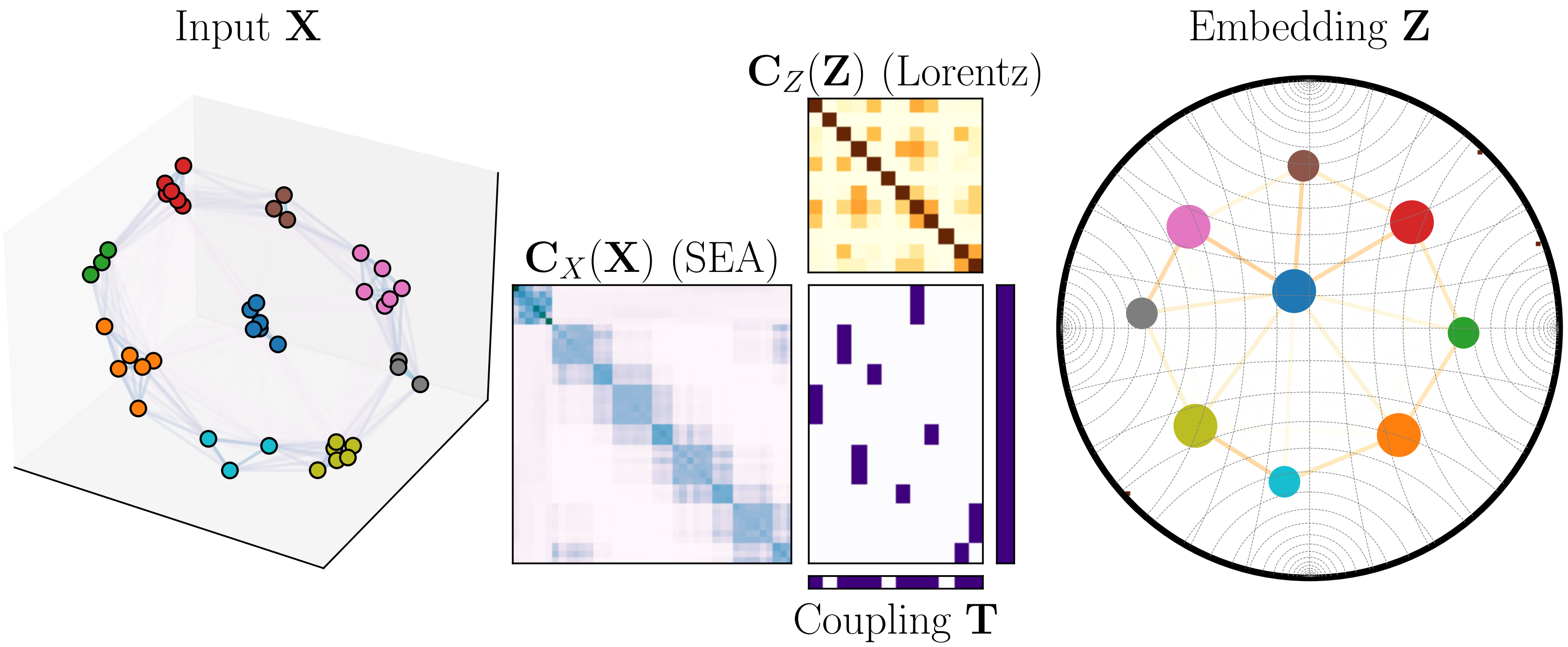
| e.g. $K_Z = \exp(-\|\mathbf{z}_i - \mathbf{z}_j\|_2^2)$ and its usual normalizations

$1_n^\top K_Z 1_n = 1,\ K_Z 1_n = 1_n,\ K_Z^\top 1_n = 1_n + K_Z 1_n = 1_n$

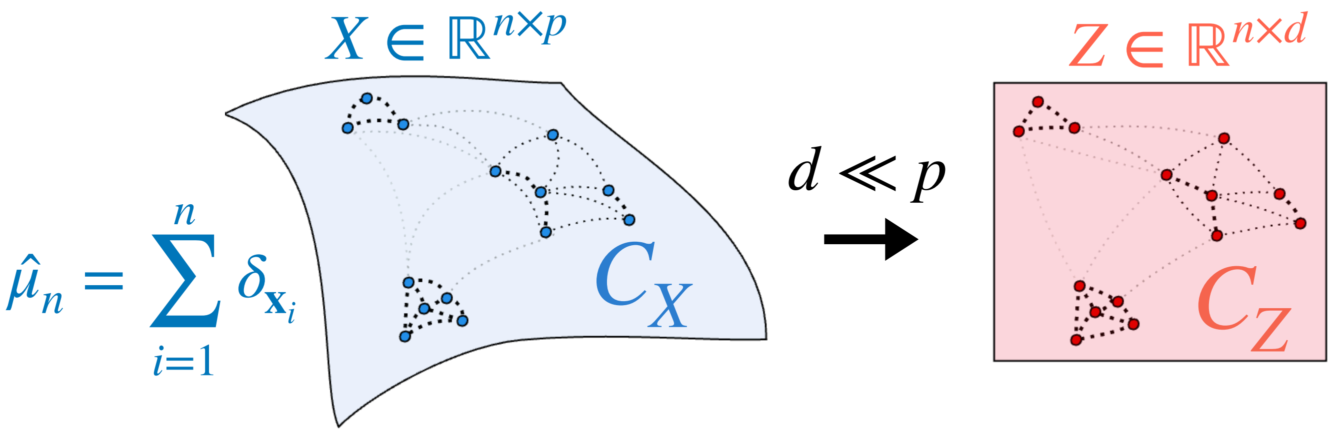
(Sinkhorn & Knopp, 1967)

| Beware that C_X is not always CPD.

Distributional reduction



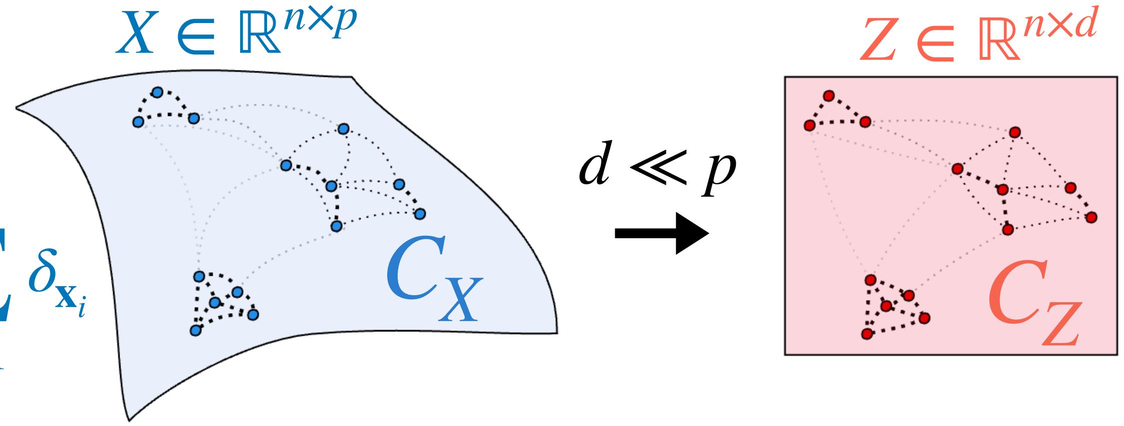
Distributional Reduction



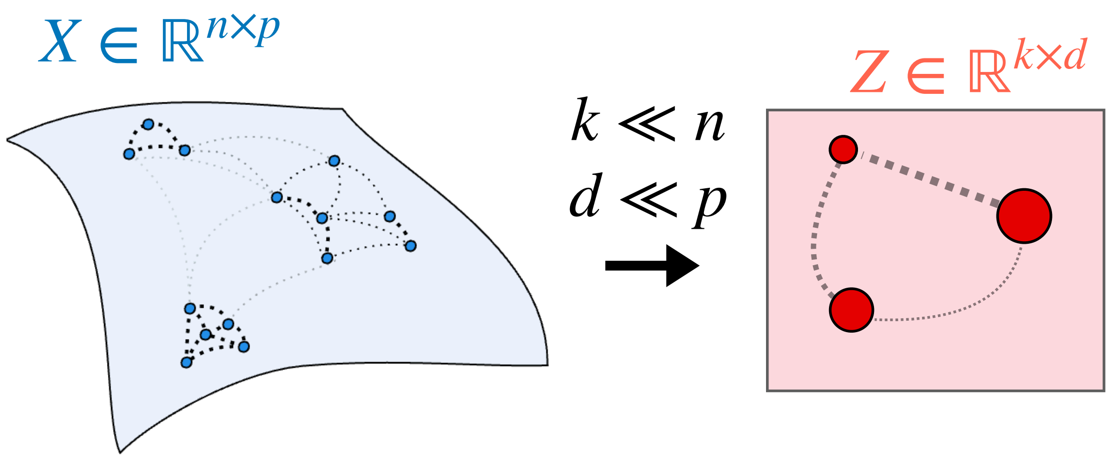
$$\sim \quad \min_{\nu=\frac{1}{n}\sum_{i=1}^{n}\delta_{\mathbf{z}_i}} \mathrm{GW}_L(\hat{\mu}_n, \nu)$$

Distributional Reduction

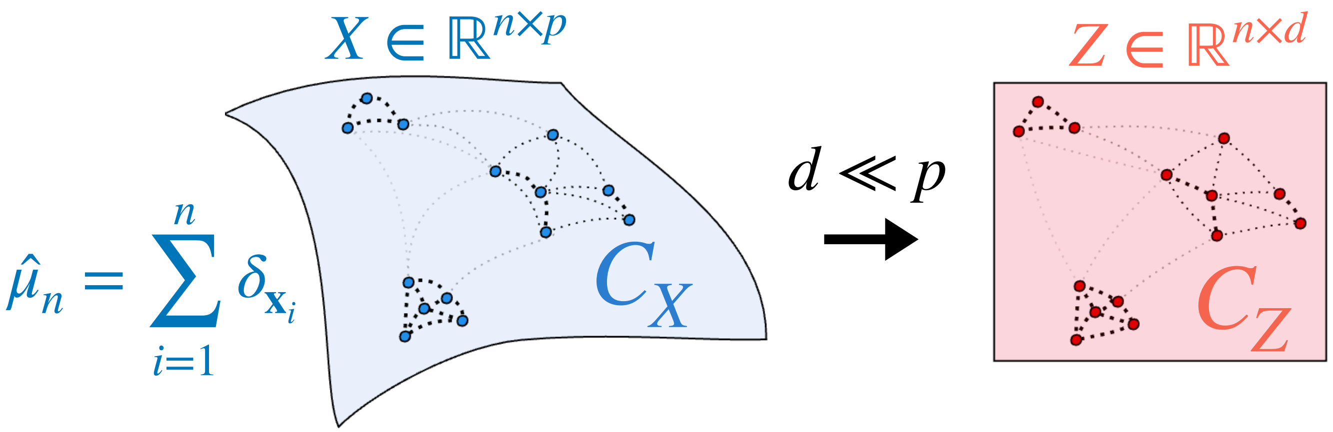
$$\hat{\mu}_n = \sum_{i=1}^{n} \delta_{\mathbf{x}_i}$$



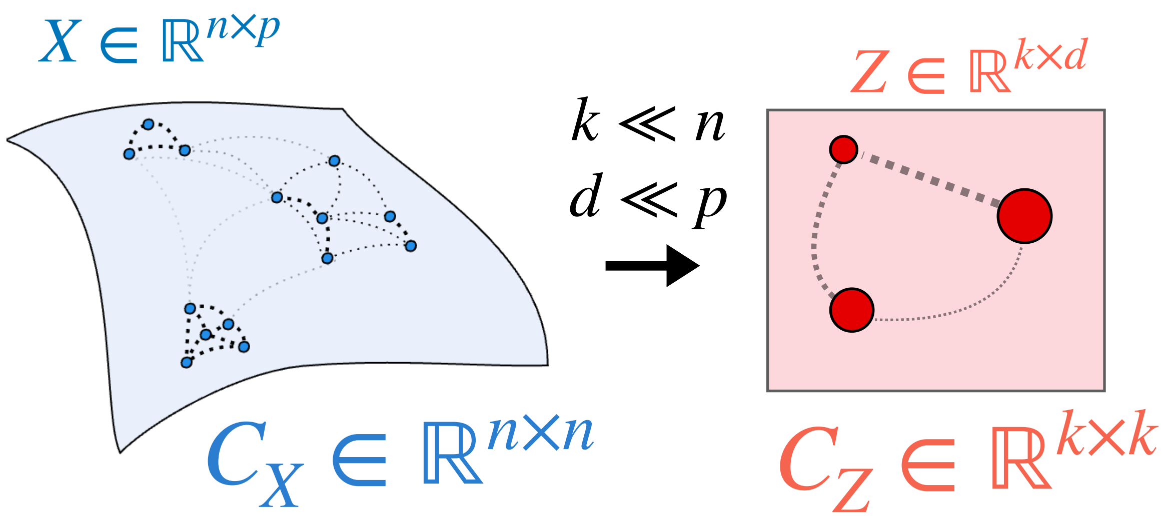
$$\sim \min_{\nu = \frac{1}{n}\sum_{i=1}^{n} \delta_{\mathbf{z}_i}} \mathrm{GW}_L(\hat{\mu}_n, \nu)$$



Distributional Reduction



$$\sim \min_{\nu=\frac{1}{n}\sum_{i=1}^{n}\delta_{\mathbf{z}_i}} \mathrm{GW}_L(\hat{\mu}_n, \nu)$$

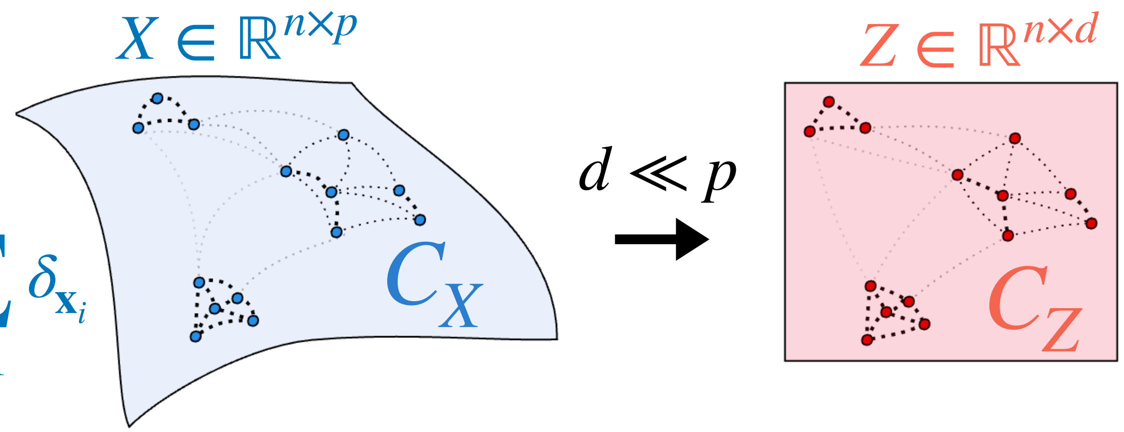


✦ **GW projection**

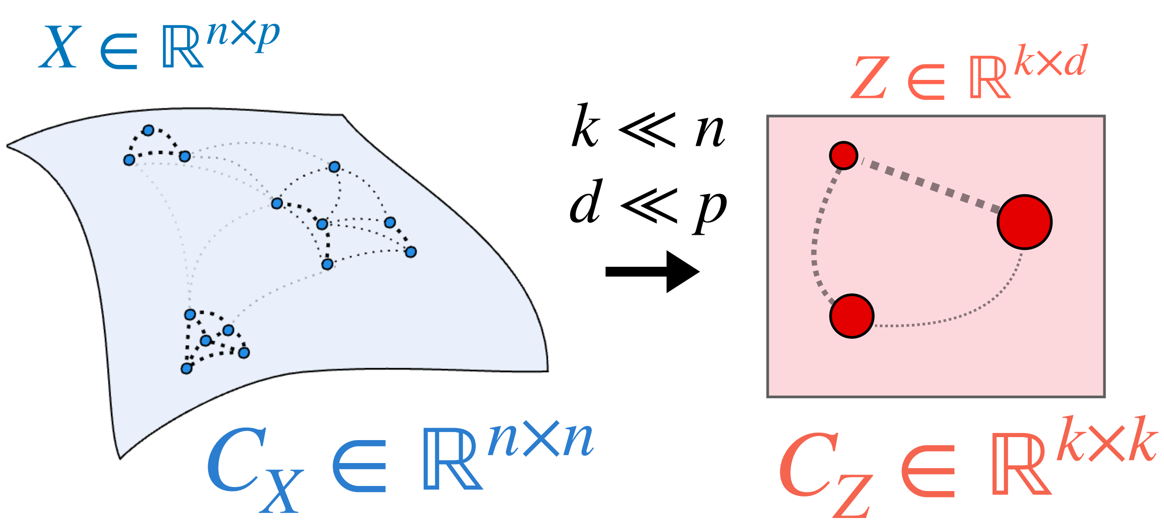
$$\min_{\nu\in\mathscr{P}_k(\mathbb{R}^d)} \mathrm{GW}(\hat{\mu}_n, \nu) \qquad \nu = \sum_{j=1}^{k} b_j \delta_{\mathbf{z}_j}$$

Distributional Reduction

$$\hat{\mu}_n = \sum_{i=1}^{n} \delta_{\mathbf{x}_i}$$



$$\sim \min_{\nu=\frac{1}{n}\sum_{i=1}^{n}\delta_{\mathbf{z}_i}} \mathrm{GW}_L(\hat{\mu}_n, \nu)$$



✦ Optimization problem

$$\min_{Z\in\mathbb{R}^{k\times d}} \min_{b\in\Sigma_k} \mathrm{GW}_L\left(C_X, C_Z, \frac{1_n}{n}, b\right)$$

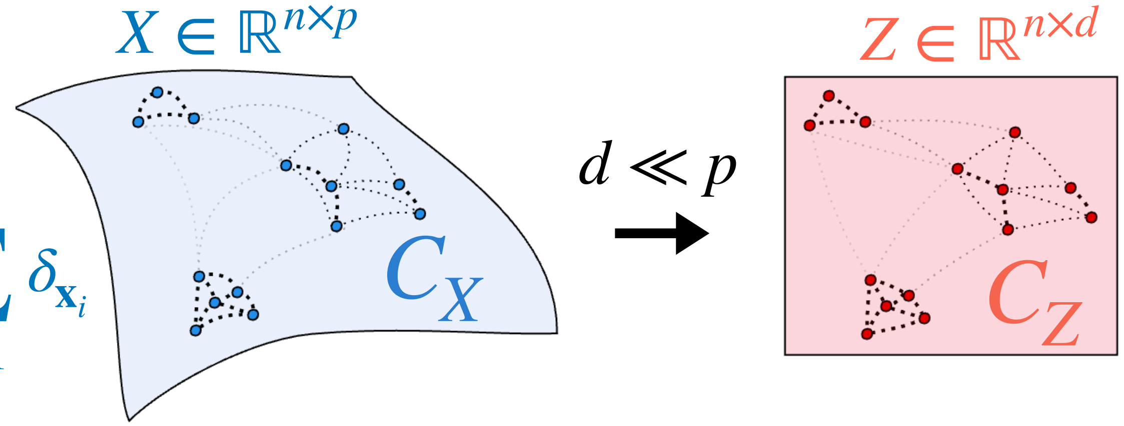
- Find **few prototypes** in **low dim.**
- Find the weights / cluster size

✦ GW projection

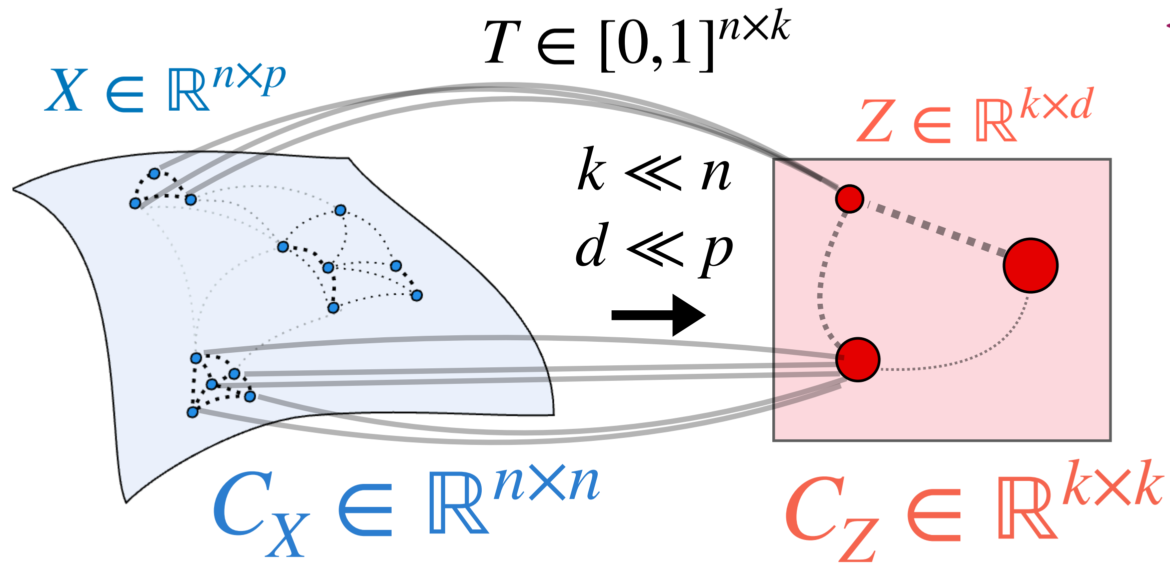
$$\min_{\nu\in\mathscr{P}_k(\mathbb{R}^d)} \mathrm{GW}(\hat{\mu}_n, \nu) \qquad \nu = \sum_{j=1}^{k} b_j \delta_{\mathbf{z}_j}$$

Distributional Reduction

$$\hat{\mu}_n = \sum_{i=1}^{n} \delta_{\mathbf{x}_i}$$



$$\sim \quad \min_{\nu=\frac{1}{n}\sum_{i=1}^{n}\delta_{\mathbf{z}_i}} \mathrm{GW}_L(\hat{\mu}_n, \nu)$$



✦ GW projection

$$\min_{\nu \in \mathscr{P}_k(\mathbb{R}^d)} \mathrm{GW}(\hat{\mu}_n, \nu) \qquad \nu = \sum_{j=1}^{k} b_j \delta_{\mathbf{z}_j}$$

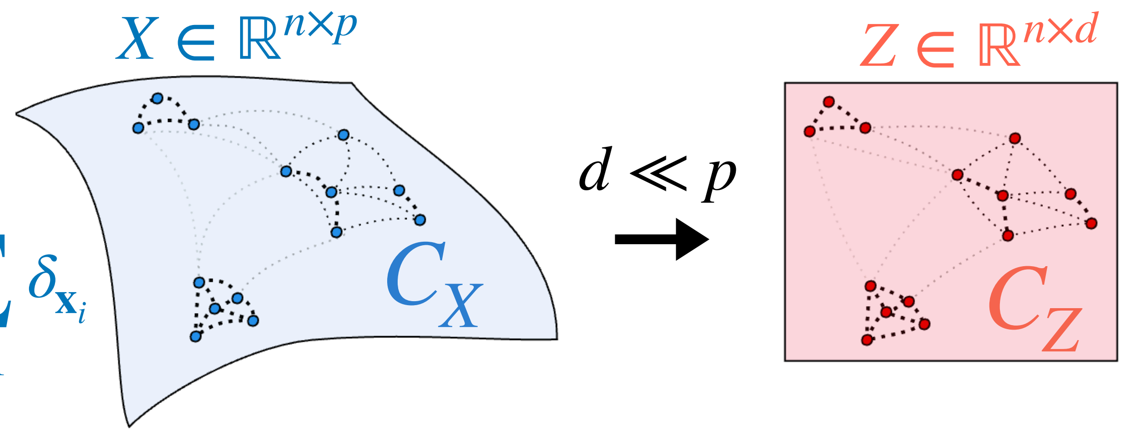
✦ Optimization problem

$$\min_{Z\in\mathbb{R}^{k\times d}} \min_{b\in\Sigma_k} \mathrm{GW}_L\left(C_X, C_Z, \frac{1_n}{n}, b\right)$$

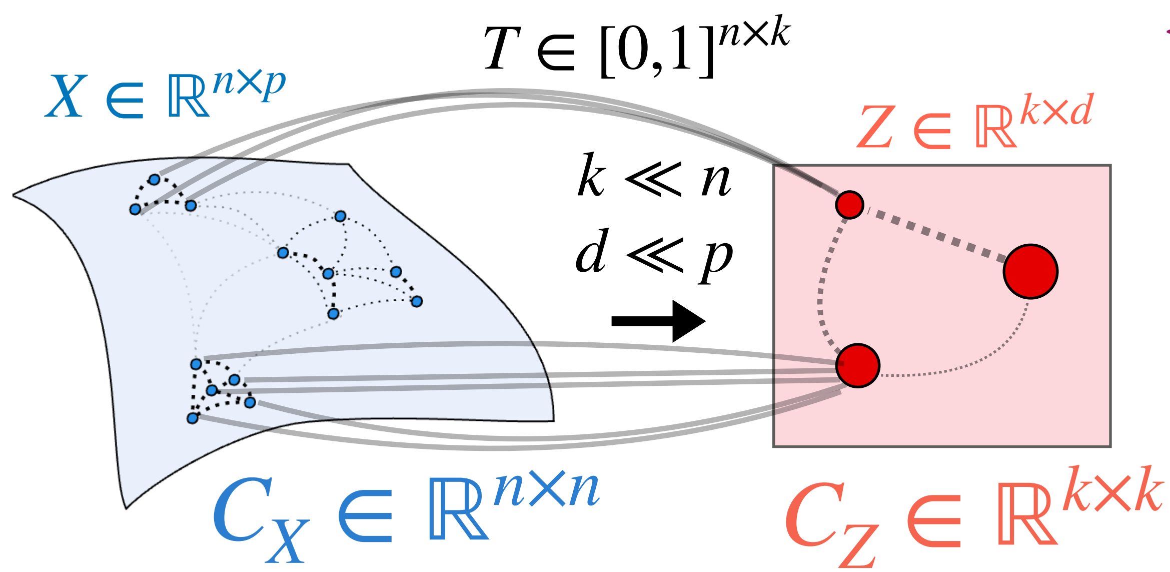
- Find **few prototypes** in **low dim.**
- Find the weights / cluster size
- **Clustering via the coupling** T (soft-assignment)
- Sufficient conditions for hard assignment (see paper)

Distributional Reduction

$$\hat{\mu}_n = \sum_{i=1}^{n} \delta_{\mathbf{x}_i}$$



$$\sim \quad \min_{\nu = \frac{1}{n}\sum_{i=1}^{n} \delta_{\mathbf{z}_i}} \mathrm{GW}_L(\hat{\mu}_n, \nu)$$



✦ Optimization problem

$$\min_{Z \in \mathbb{R}^{k \times d}} \min_{b \in \Sigma_k} \mathrm{GW}_L\left(C_X, C_Z, \frac{1_n}{n}, b\right)$$

- Find **few prototypes** in **low dim.**
- Find the weights / cluster size
- **Clustering via the coupling** T (soft-assignment)
- Sufficient conditions for hard assignment (see paper)

✦ GW projection

$$\min_{\nu \in \mathscr{P}_k(\mathbb{R}^d)} \mathrm{GW}(\hat{\mu}_n, \nu) \qquad \nu = \sum_{j=1}^{k} b_j \delta_{\mathbf{z}_j}$$

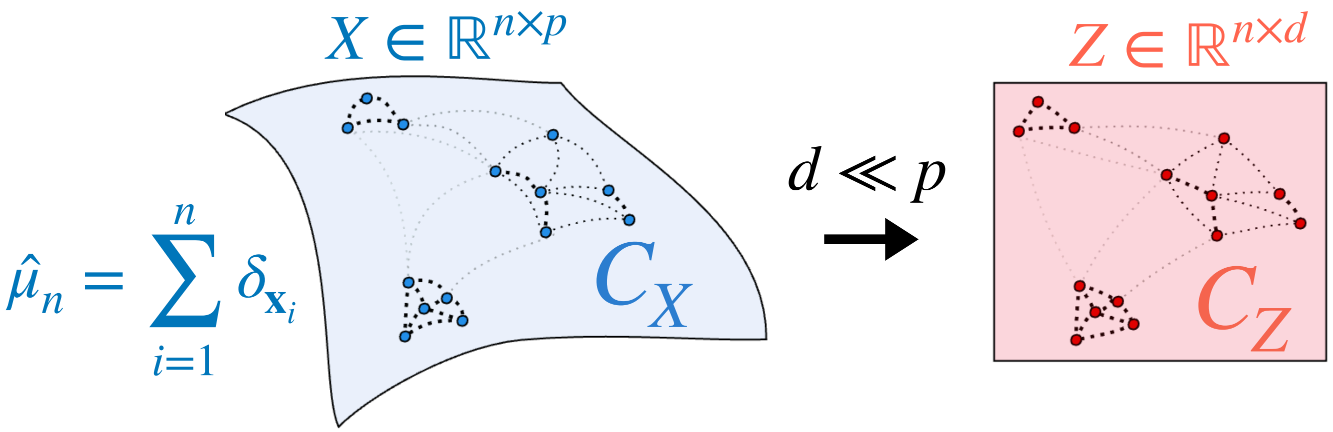
✦ A semi-relaxed objective

(Vincent-Cuaz et al., 2022)

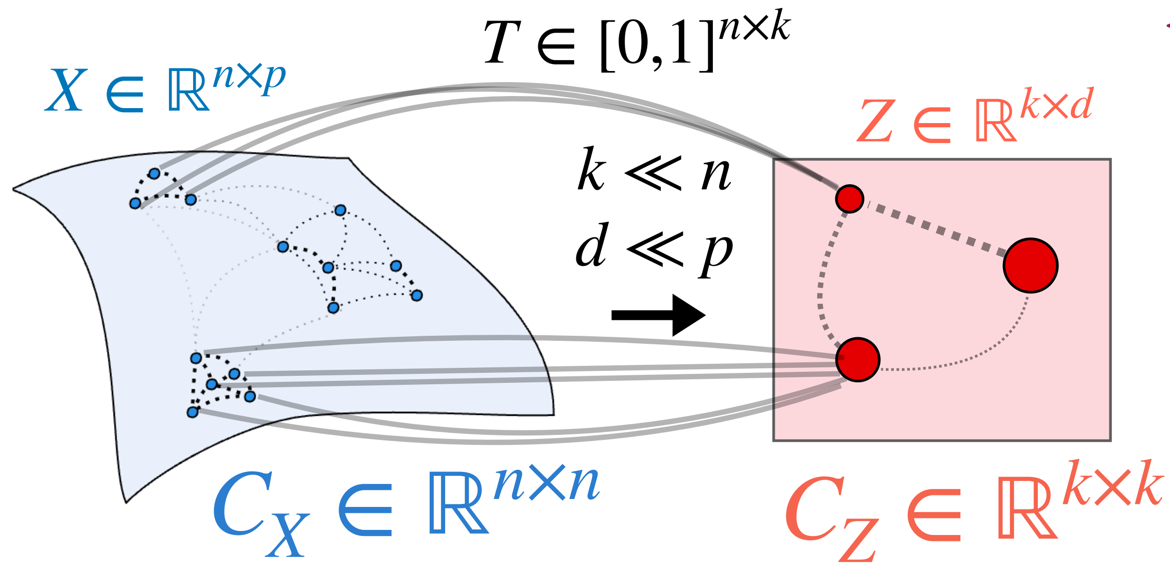
$$\min_{Z \in \mathbb{R}^{n \times d}} \min_{T : T1_k = \frac{1_n}{n}} \sum_{ijkl} L\left([C_X]_{ik}, [C_Z]_{jl}\right) T_{ij} T_{kl}$$

→ easier than GW

Distributional Reduction



$$\sim \quad \min_{\nu = \frac{1}{n}\sum_{i=1}^{n}\delta_{\mathbf{z}_i}} \mathrm{GW}_L(\hat{\mu}_n, \nu)$$



✦ GW projection

$$\min_{\nu \in \mathcal{P}_k(\mathbb{R}^d)} \mathrm{GW}(\hat{\mu}_n, \nu) \qquad \nu = \sum_{j=1}^{k} b_j \delta_{\mathbf{z}_j}$$

✦ Optimization problem

$$\min_{Z \in \mathbb{R}^{k\times d}} \min_{b \in \Sigma_k} \mathrm{GW}_L\left(C_X, C_Z, \frac{1_n}{n}, b\right)$$

- Find **few prototypes** in **low dim.**
- Find the weights / cluster size
- **Clustering via the coupling** T (soft-assignment)
- Sufficient conditions for hard assignment (see paper)

✦ A semi-relaxed objective

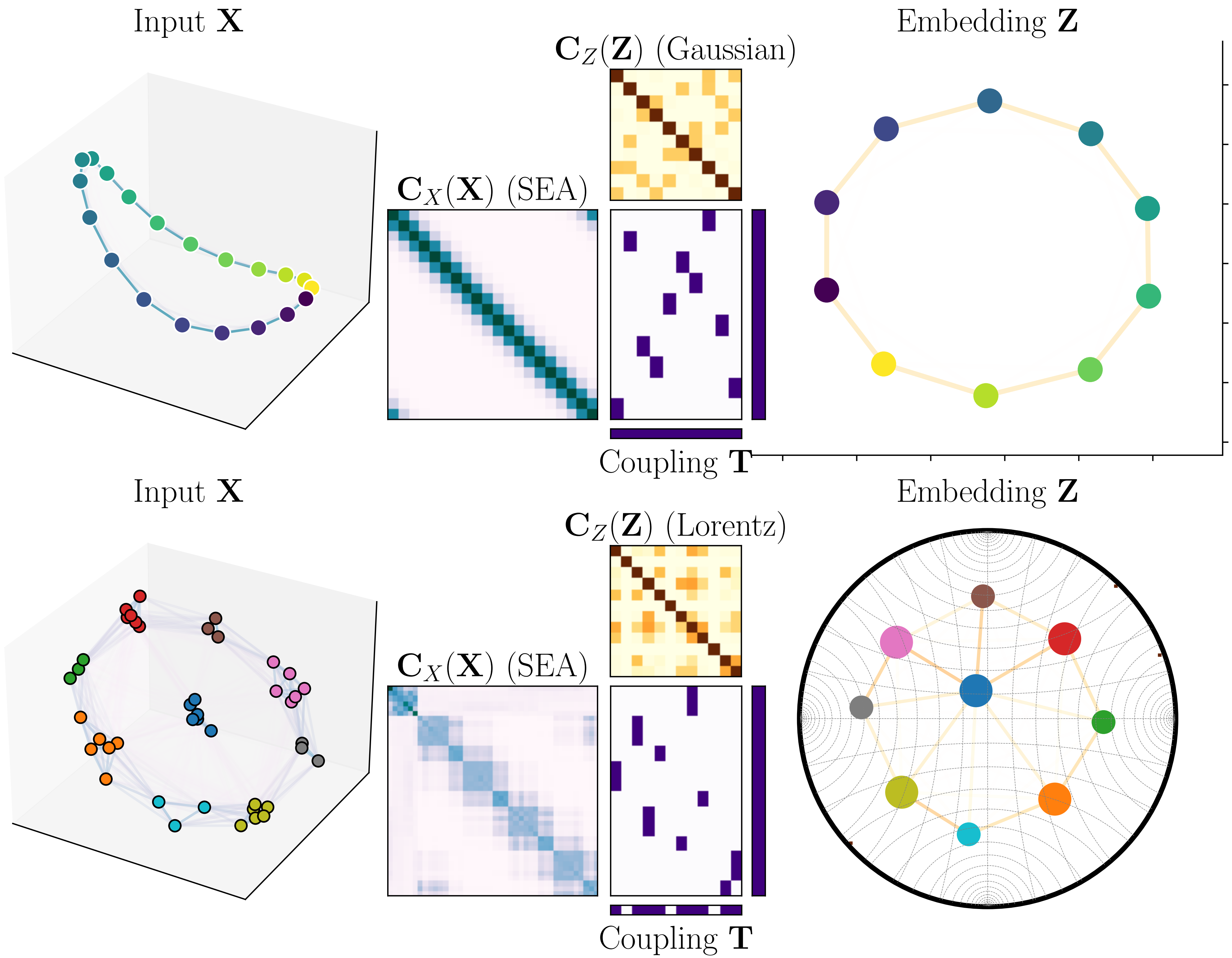
(Vincent-Cuaz et al., 2022)

$$\min_{Z \in \mathbb{R}^{n\times d}} \min_{T: T1_k = \frac{1_n}{n}} \sum_{ijkl} L\left([C_X]_{ik}, [C_Z]_{jl}\right) T_{ij} T_{kl}$$

→ easier than GW

- Non-convex problem
- BCD: alternates optim in Z, in T
- Optim in T: CG solver in $O(n^2 k)$ for $L \in \{\mathrm{KL}, |\cdot|^2\}$
- With low-rank structures $O(nkr + n^2)$

Distributional Reduction



Motivation

Baran *et al. Genome Biology* (2019) 20:206
https://doi.org/10.1186/s13059-019-1812-2

Genome Biology

METHOD Open Access

MetaCell: analysis of single-cell RNA-seq data using *K*-nn graph partitions

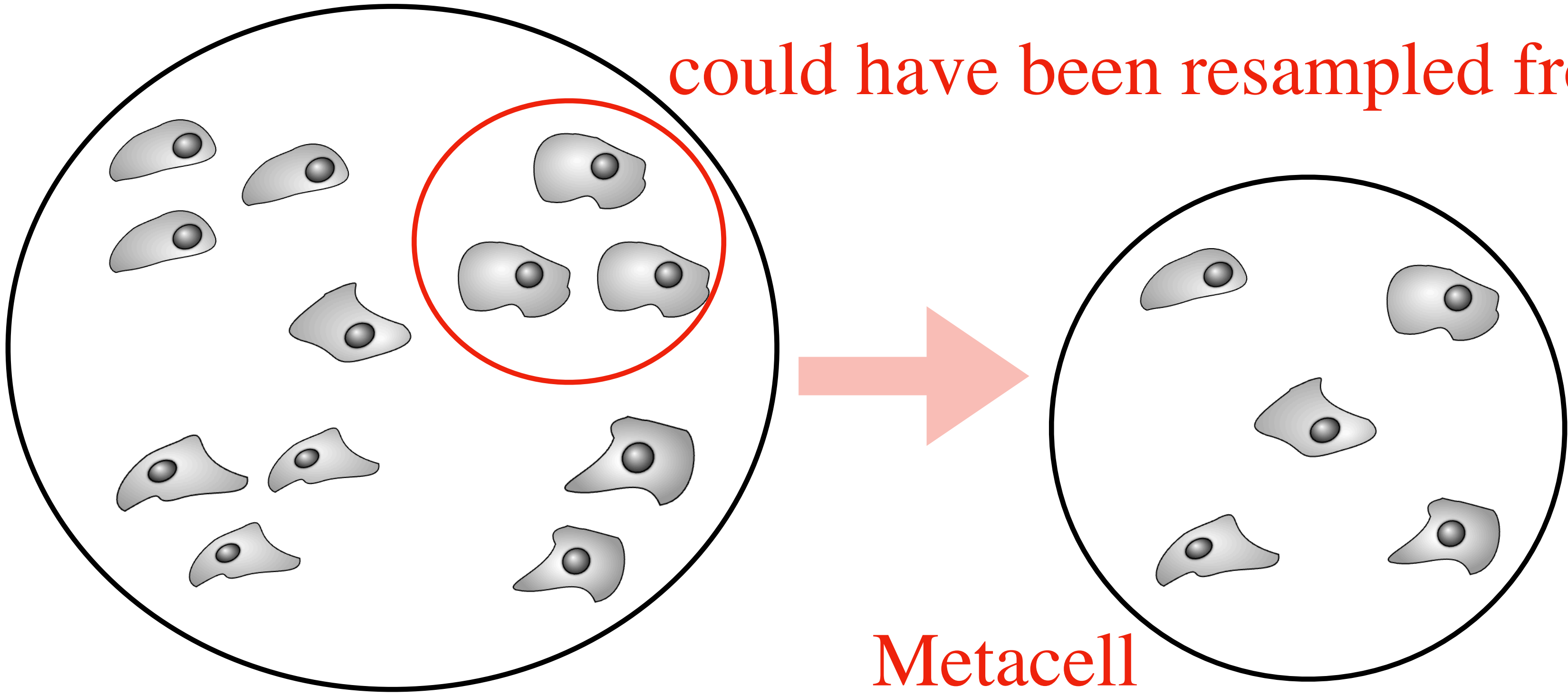
Yael Baran[1], Akhiad Bercovich[1], Arnau Sebe-Pedros[1], Yaniv Lubling[1], Amir Giladi[2], Elad Chomsky[1], Zohar Meir[1], Michael Hoichman[1], Aviezer Lifshitz[1] and Amos Tanay[1*]

✦ **Single-cell RNA-seq**

Technical noise due to partial sampling of RNA molecules within cells.

Problem : **impossible to resample** a cell

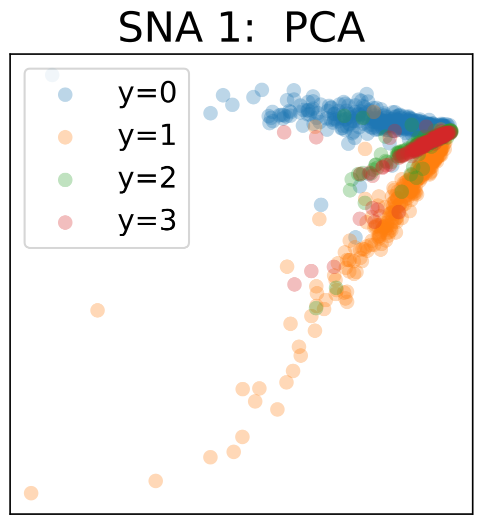
- integration of data from different cells
- need to **separate the sampling effect from biological variance**



✦ **We would like to choose the granularity of the output data**

Distributional Reduction

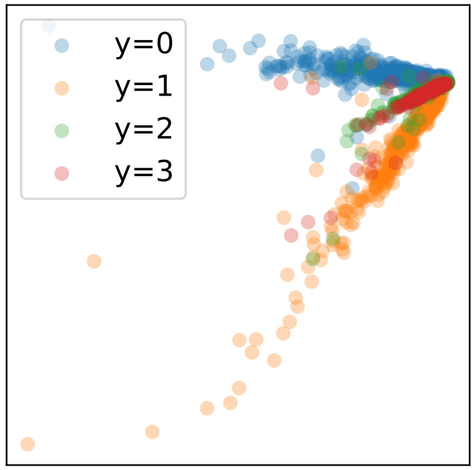
✦ **Single-cell dataset** (Chen et al., 2019) Only solve $\min_{b \in \Sigma_k} \mathrm{GW}_L(C_X, C_Z, \frac{1_n}{n}, b)$

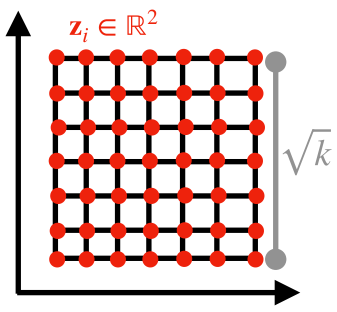


Distributional Reduction

✦ **Single-cell dataset** (Chen et al., 2019) Only solve $\min_{b \in \Sigma_k} \mathrm{GW}_L(C_X, C_Z, \frac{1_n}{n}, b)$ with $C_X = XX^\top$ and $C_Z = ZZ^\top$ **fixed on a grid** $Z \in \mathbb{R}^{k \times 2}$

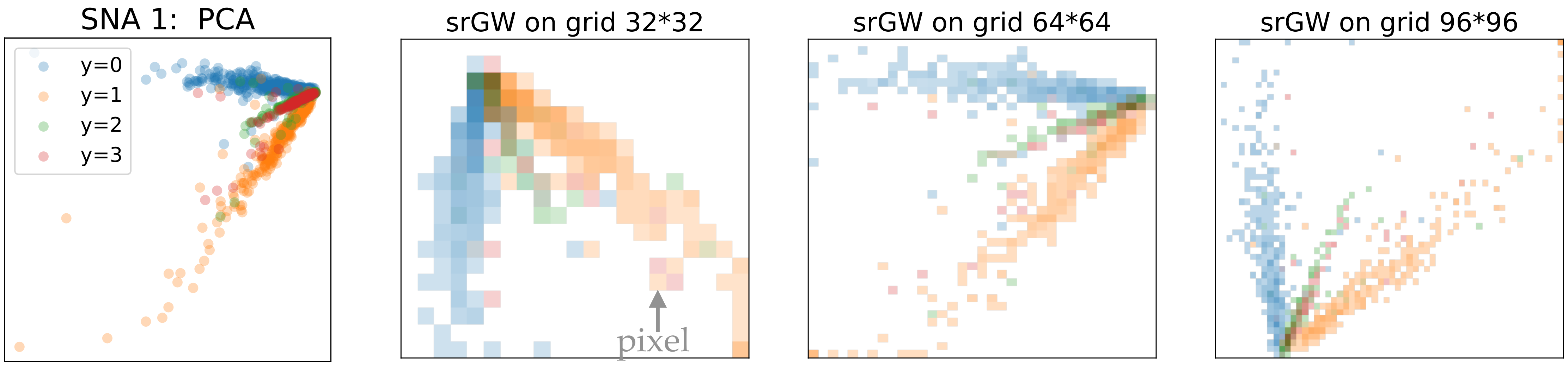
SNA 1: PCA



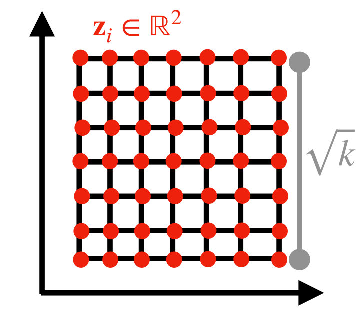


Distributional Reduction

✦ **Single-cell dataset** (Chen et al., 2019) Only solve $\min_{b \in \Sigma_k} \mathrm{GW}_L(C_X, C_Z, \frac{1_n}{n}, b)$ with $C_X = XX^\top$ and $C_Z = ZZ^\top$ **fixed on a grid** $Z \in \mathbb{R}^{k \times 2}$



dataset as an image



Distributional Reduction

✦ Single-cell dataset (Chen et al., 2019)

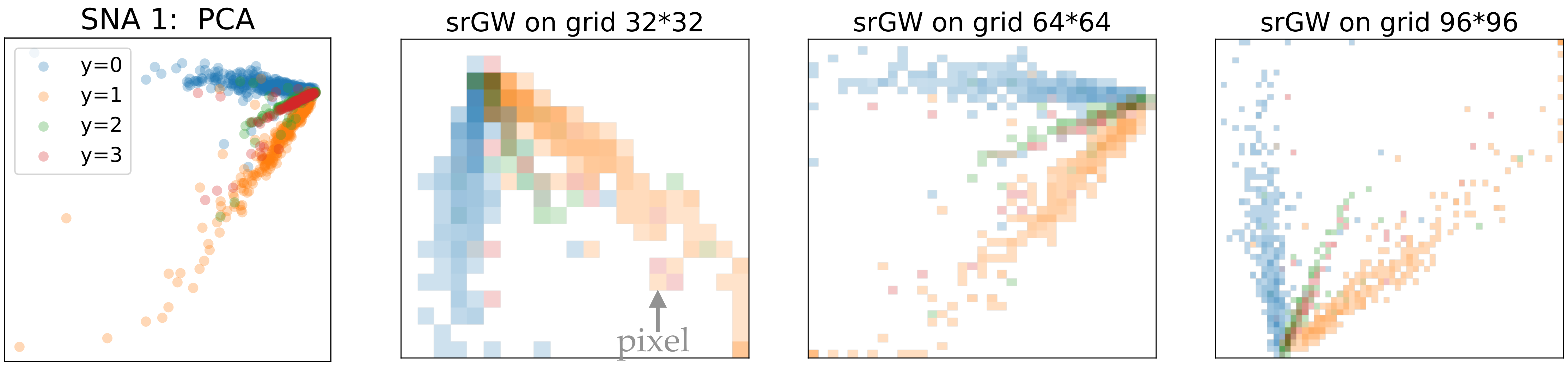
Only solve $\min_{b \in \Sigma_k} \mathrm{GW}_L(C_X, C_Z, \frac{1_n}{n}, b)$ with $C_X = XX^\top$

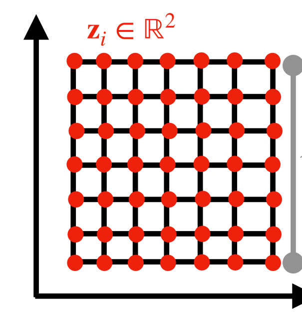
and

$C_Z = ZZ^\top$

fixed on a grid

$Z \in \mathbb{R}^{k \times 2}$



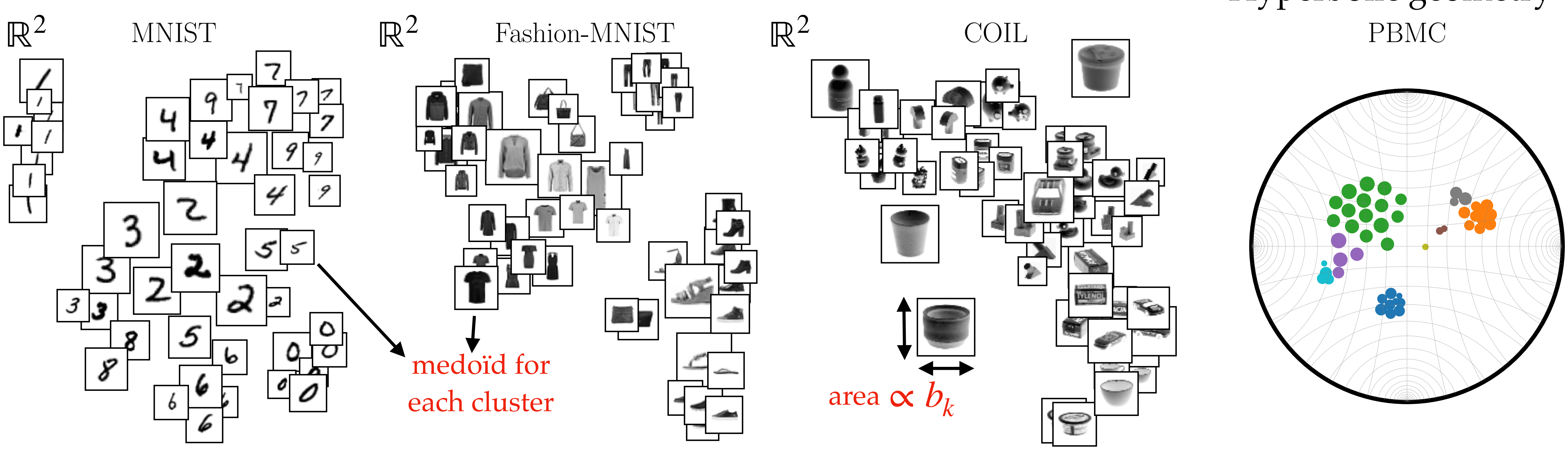


dataset as an image

✦ Image datasets

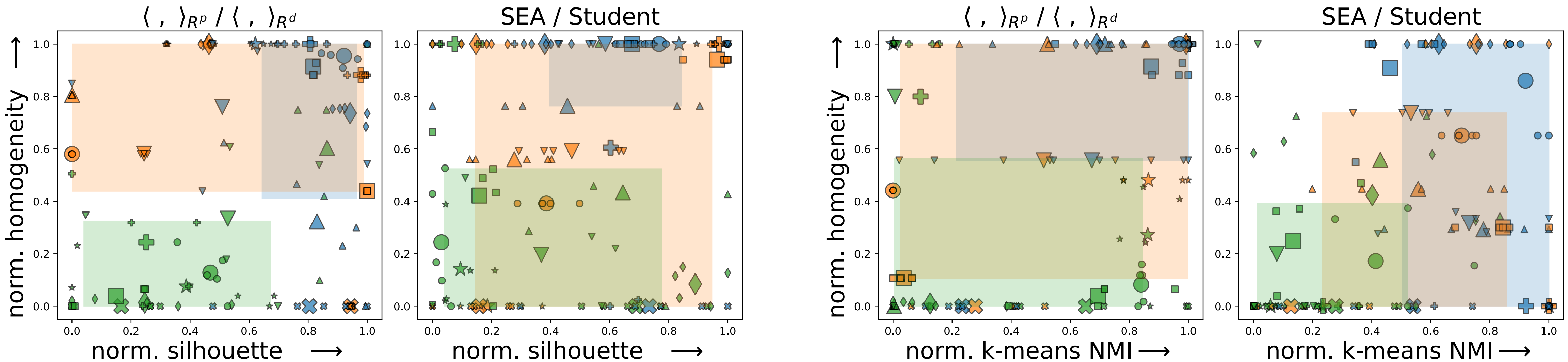
C_X symmetric entropic aff. (Van Assel et al., 2023) C_Z Student t-kernel

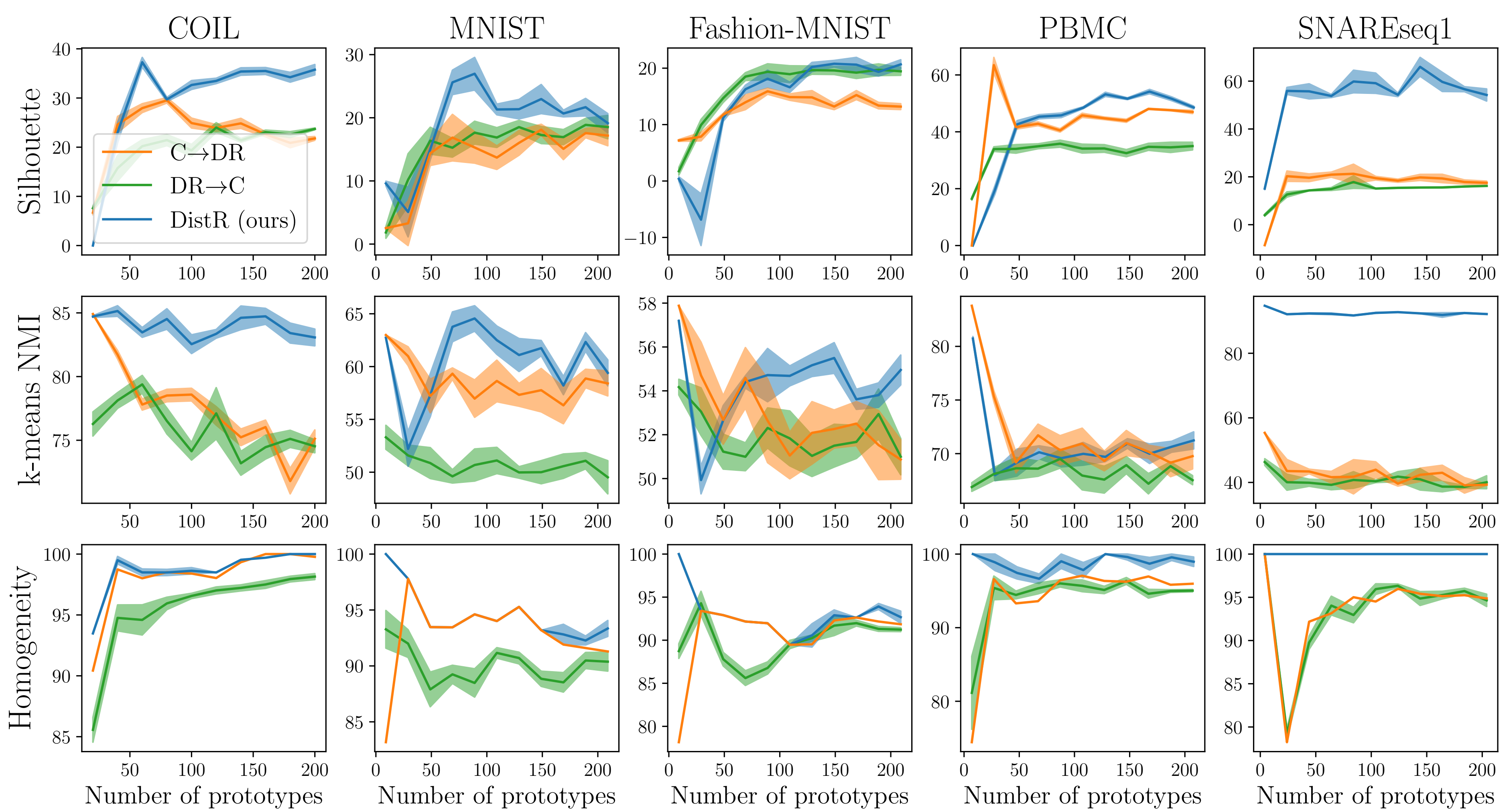
Hyperbolic geometry



Distributional Reduction

✦ Comparison with DR then clustering or clustering then DR





Thank you!

Open-source implementations are available here:

TorchDR

DR library in pytorch

- Implements popular DR algos
- Modular
- Efficiency (GPU, KeOps)

Python Optimal Transport

- OT LP solver, Sinkhorn
- Barycenters
- Gromov, graphs OT…

Url: https://github.com/PythonOT/POT