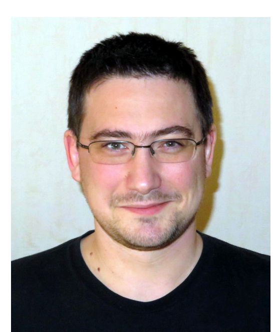
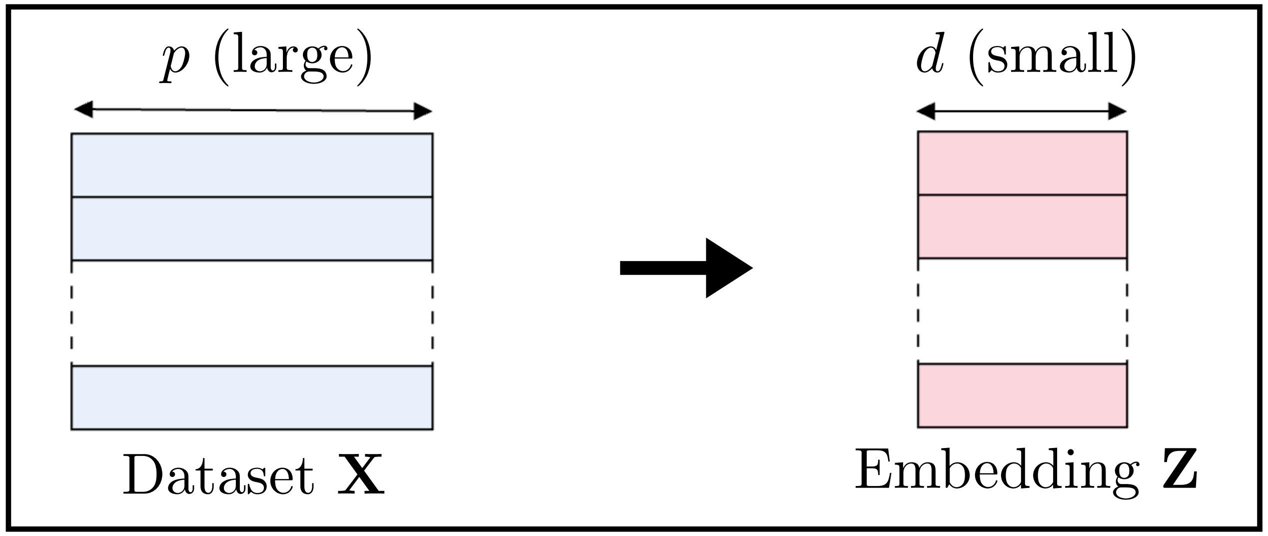
Distributional Reduction: Unifying Dimensionality Reduction and Clustering with Gromov-Wasserstein



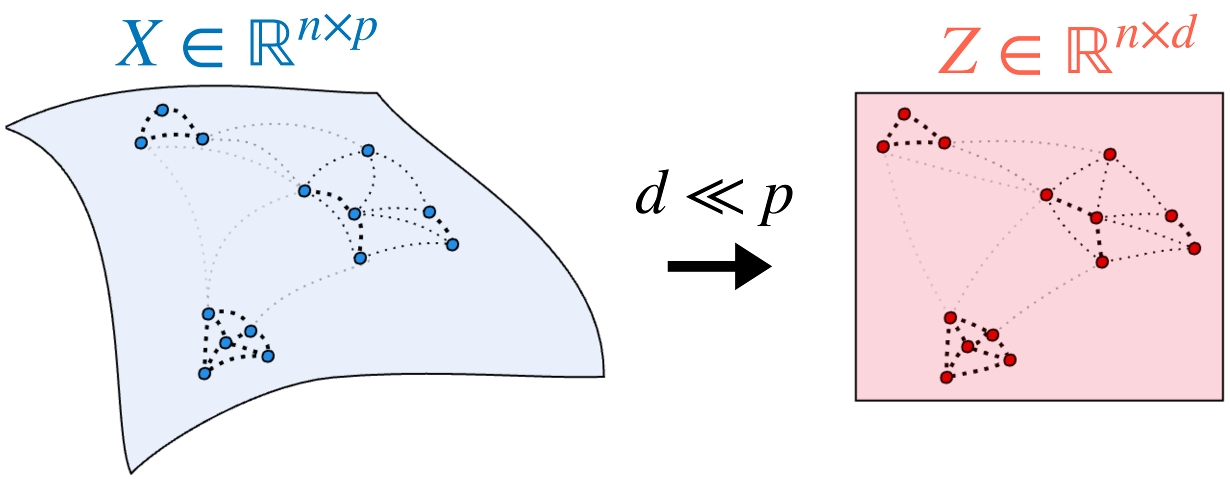



Hugues Van Assel Cédric Vincent-Cuaz Rémi Flamary Nicolas Courty Pascal Frossard Titouan Vayer

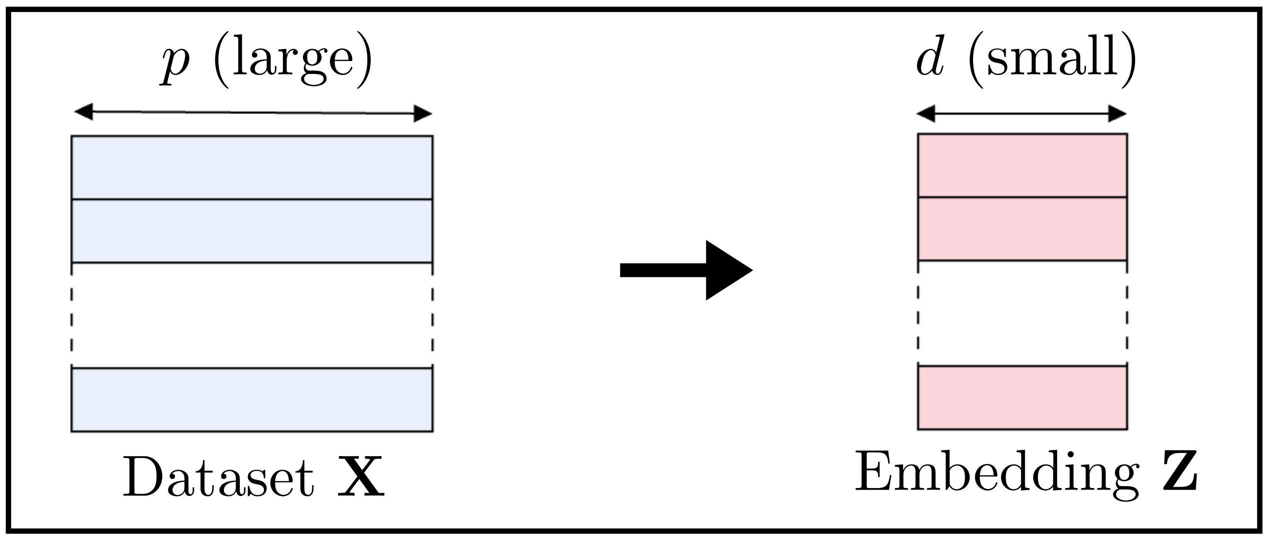
Dimension reduction



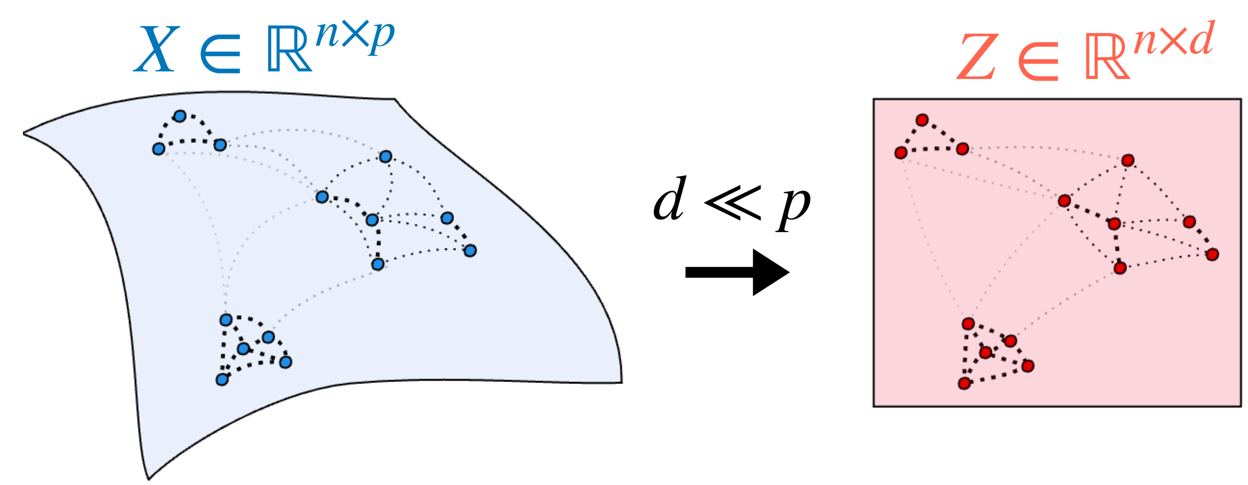
✦ **Preserving geometric properties**



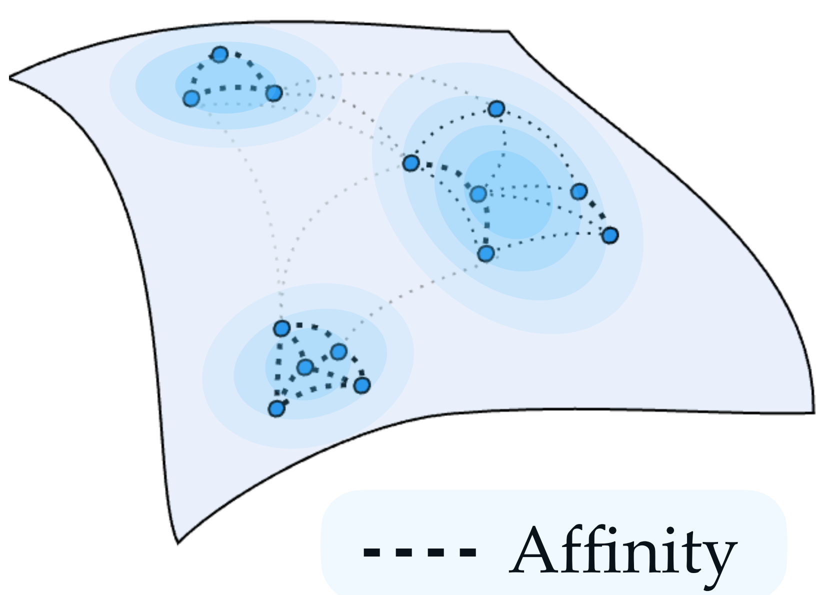
Dimension reduction

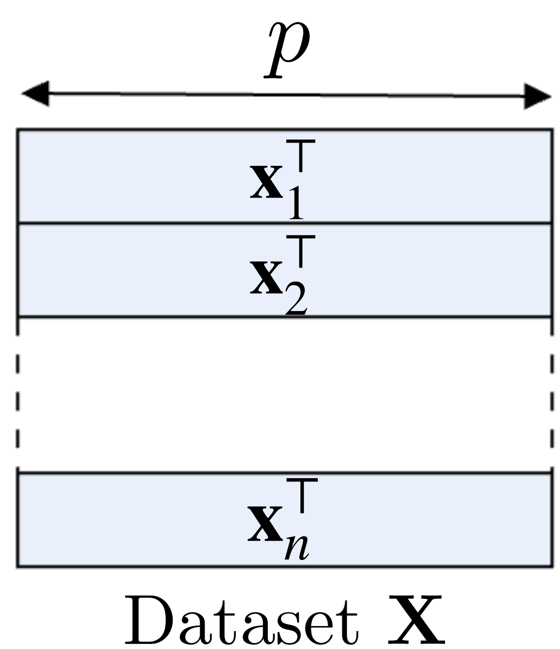


✦ Preserving geometric properties



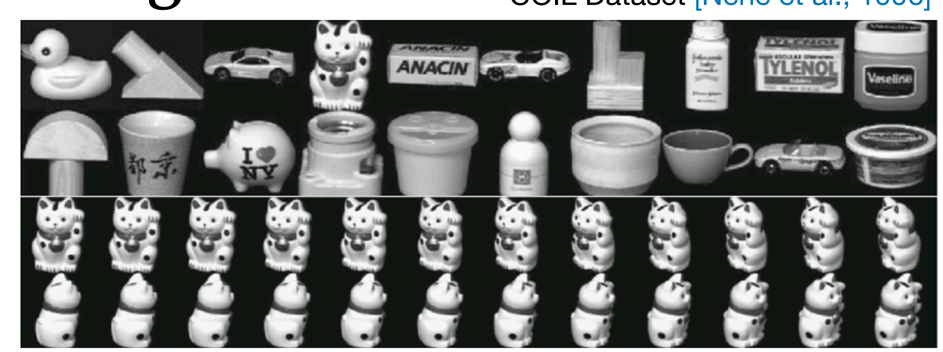
✦ Affinity Matrices



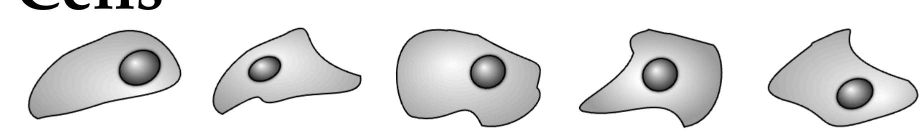


Images

COIL Dataset [Nene et al., 1996]



Cells

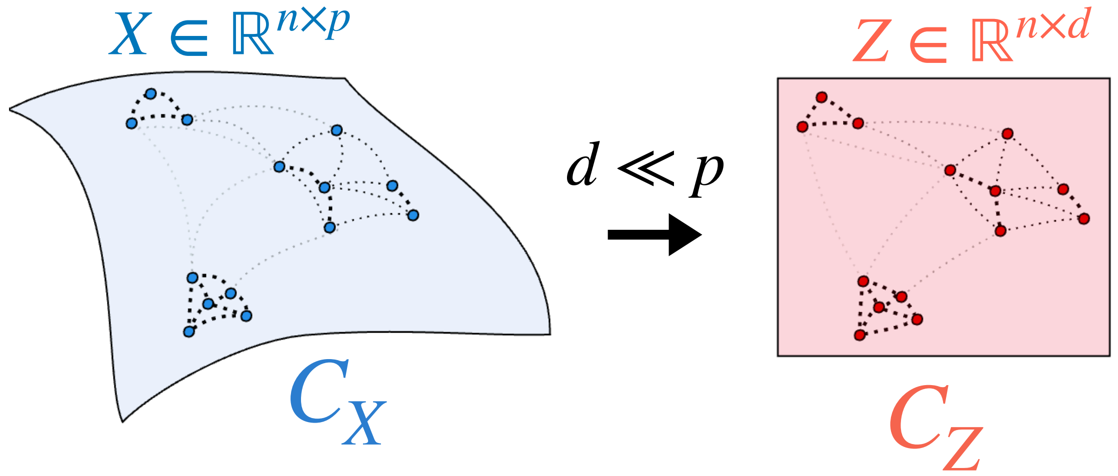


Symmetric matrix with **non-negative coefficients**.

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

Dimension reduction

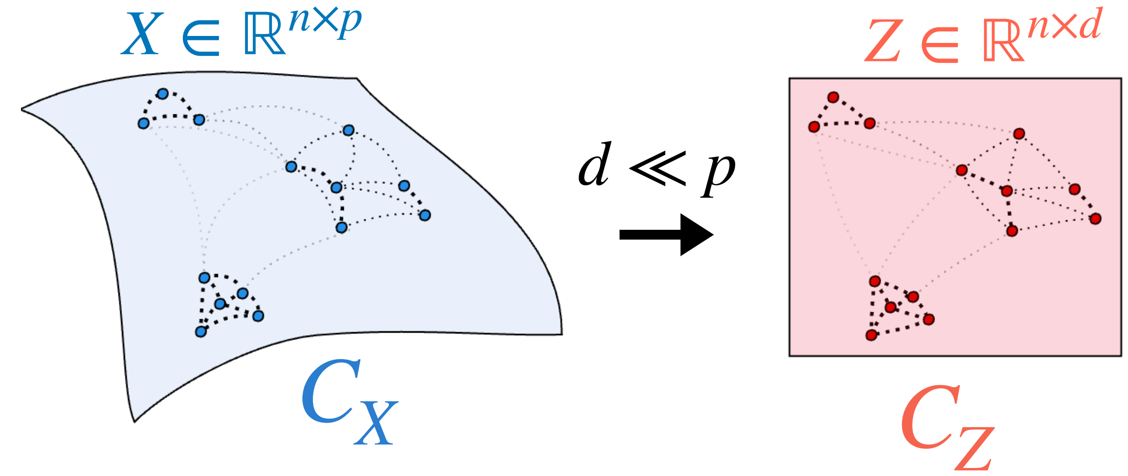
Dimension reduction



✦ **A general optimization problem**

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

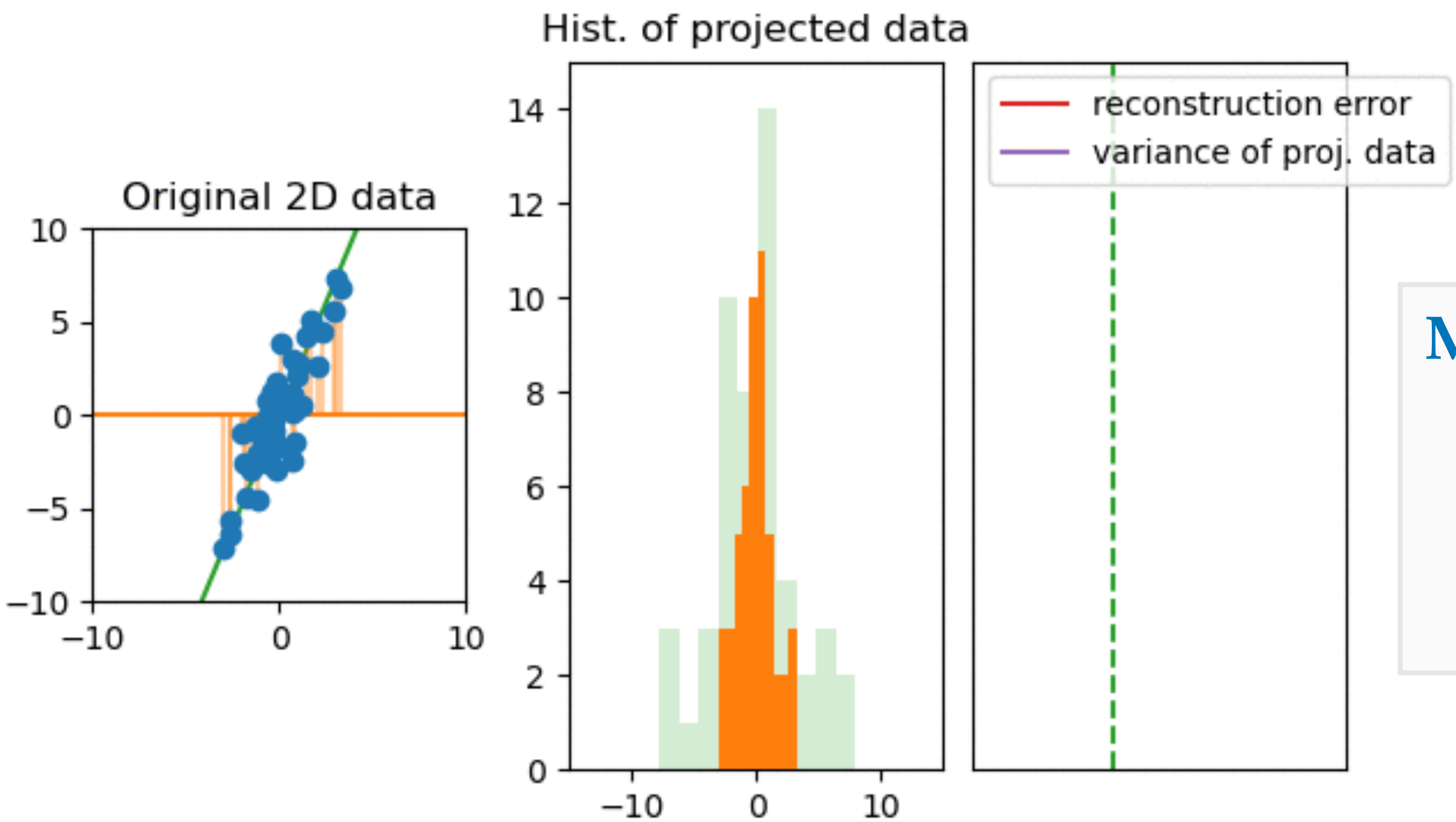
Dimension reduction

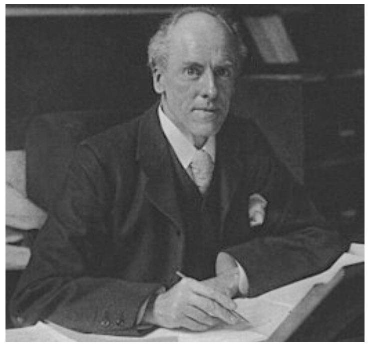


✦ A general optimization problem

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

✦ Principal components analysis



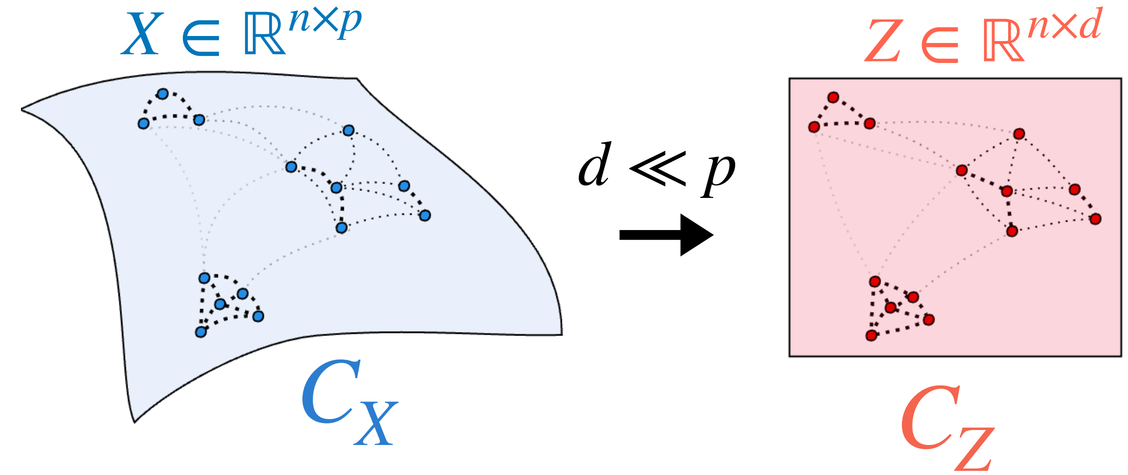


(Pearson, 1901)

Minimizing the reconstruction error

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

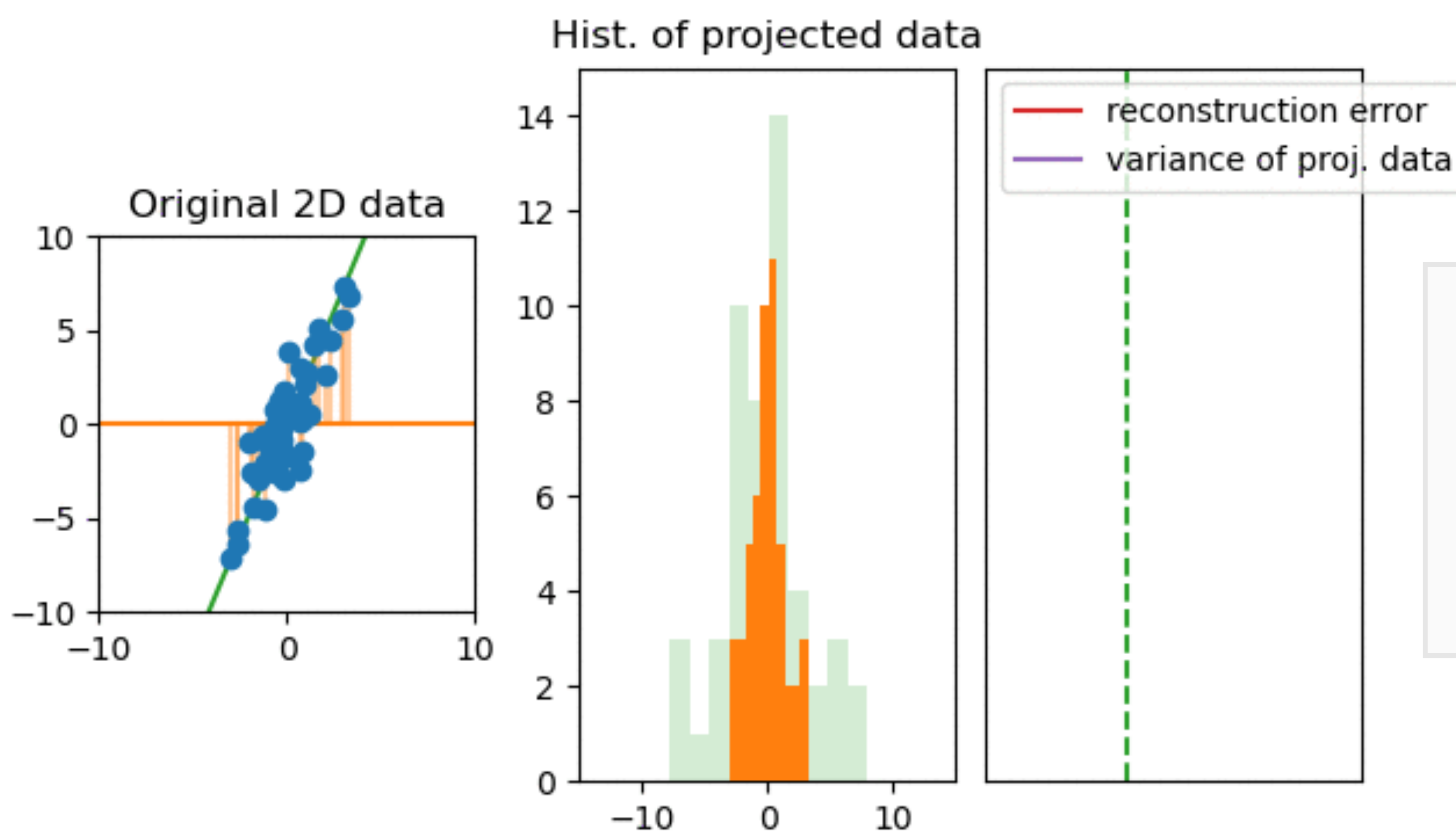
Dimension reduction

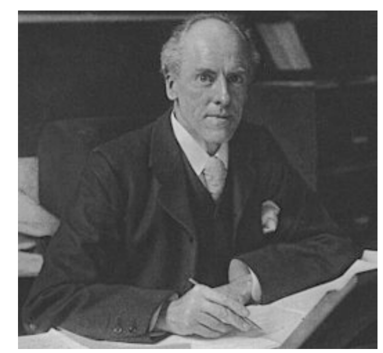


✦ A general optimization problem

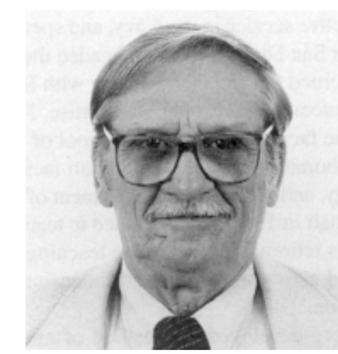
$$\min_{Z\in\mathbb{R}^{n\times d}} \sum_{i,j=1}^{n} L\Big([C_X]_{ij}\ ,\ [C_Z]_{ij}\Big) \quad \text{for some loss} \quad L : \mathbb{R}\times\mathbb{R} \to \mathbb{R}$$

✦ Principal components analysis





(Pearson, 1901)



(Torgerson, 1958)

Preserving the inner products

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

$$Z \leftarrow \mathrm{EVD}\big(\frac{1}{n}XX^\top\big)$$

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**

$$\min_{Z\in\mathbb{R}^{n\times d}} \sum_{i=1}^{n} \Big([C_X]_{ij} - \langle \mathbf{z}_i, \mathbf{z}_j \rangle \Big)^2 \xrightarrow[\text{(Eckart \& Young, 1936)}]{C_X \succeq 0 \quad \textbf{solution}} Z^\star = (\sqrt{\lambda_1}\mathbf{v}_1, \cdots, \sqrt{\lambda_d}\mathbf{v}_d)^\top$$

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

Dimension reduction

✦ **Spectral methods**

$$\min_{Z \in \mathbb{R}^{n \times d}} \sum_{i=1}^{n} \left([C_X]_{ij} - \langle \mathbf{z}_i, \mathbf{z}_j \rangle \right)^2 \xrightarrow[\text{(Eckart \& Young, 1936)}]{C_X \succeq 0 \quad \textbf{solution}} Z^\star = (\sqrt{\lambda_1}\mathbf{v}_1, \cdots, \sqrt{\lambda_d}\mathbf{v}_d)^\top$$

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

✦ **Kernel PCA** $C_X \succeq 0$

(Schölkopf, 1997)

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

Dimension reduction

✦ Spectral methods

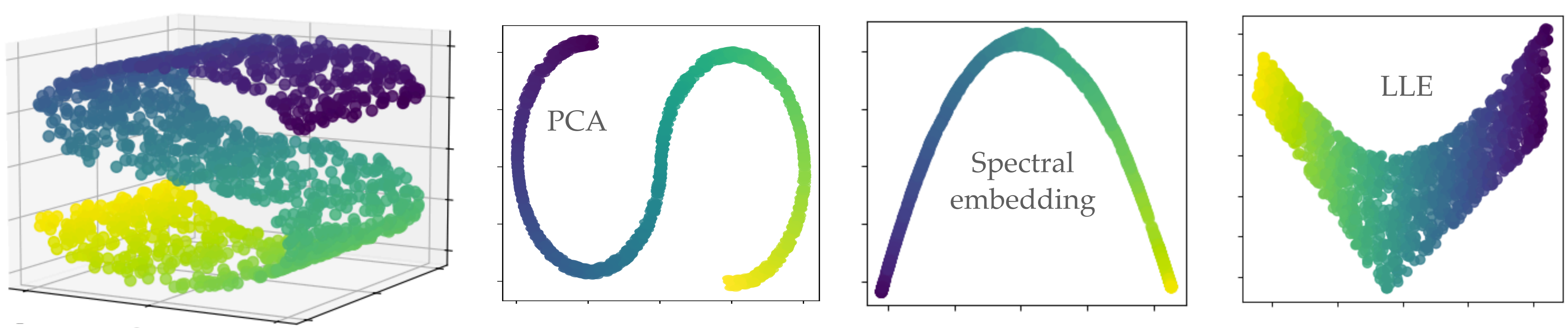
$$\min_{Z \in \mathbb{R}^{n \times d}} \sum_{i=1}^{n} \Big([C_X]_{ij} - \langle \mathbf{z}_i, \mathbf{z}_j \rangle \Big)^2 \xrightarrow[\text{(Eckart \& Young, 1936)}]{C_X \succeq 0 \;\; \textbf{solution}} Z^\star = (\sqrt{\lambda_1}\mathbf{v}_1, \cdots, \sqrt{\lambda_d}\mathbf{v}_d)^\top$$

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

✦ Kernel PCA $C_X \succeq 0$

(Schölkopf, 1997)

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

- **Neighbor embedding methods**

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

- **What choices for C_X, C_Z?**
  - Encode the **non-linear geometry**
  - Some kind of **normalization**
  - Robustness to **varying density**
  - Careful to **high vs low dim**

Dimension reduction

✦ **Neighbor embedding methods**

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

✦ **What choices for C_X, C_Z?**

- Encode the **non-linear geometry**
- Some kind of **normalization**
- Robustness to **varying density**
- Careful to **high vs low dim**

✦ **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

✦ **Neighbor embedding methods**

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

✦ **What choices for C_X, C_Z?**

- Encode the **non-linear geometry**
- Some kind of **normalization**
- Robustness to **varying density**
- Careful to **high vs low dim**

✦ **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.

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

Embedding space

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

Dimension reduction

✦ Neighbor embedding methods

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

✦ What choices for C_X, C_Z?

- Encode the **non-linear geometry**
- Some kind of **normalization**
- Robustness to **varying density**
- Careful to **high vs low dim**

✦ 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

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

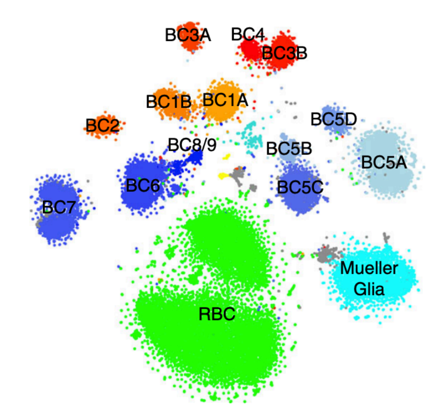
✦ (t)-SNE (Van der Maaten & Hinton, 2008)

- Joint distributions:

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

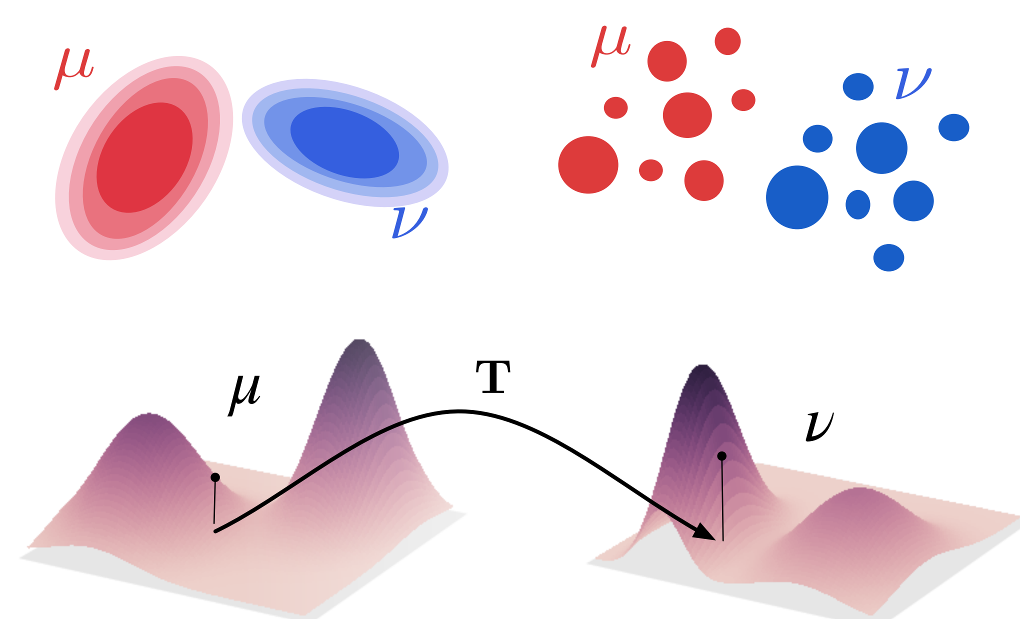
$$[C_X]_{ij} \leftarrow \frac{[C_X]_{ij} + [C_X]_{ji}}{2n}$$

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



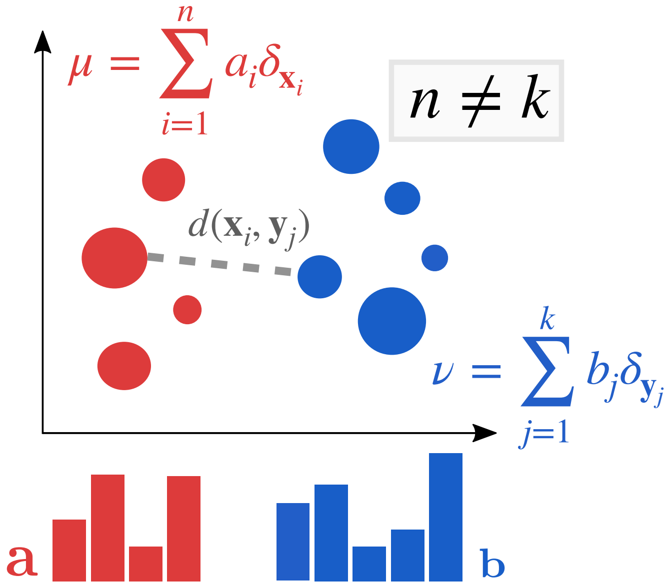
(Shekhar et al., 2016)

From linear Optimal Transport to Gromov-Wasserstein



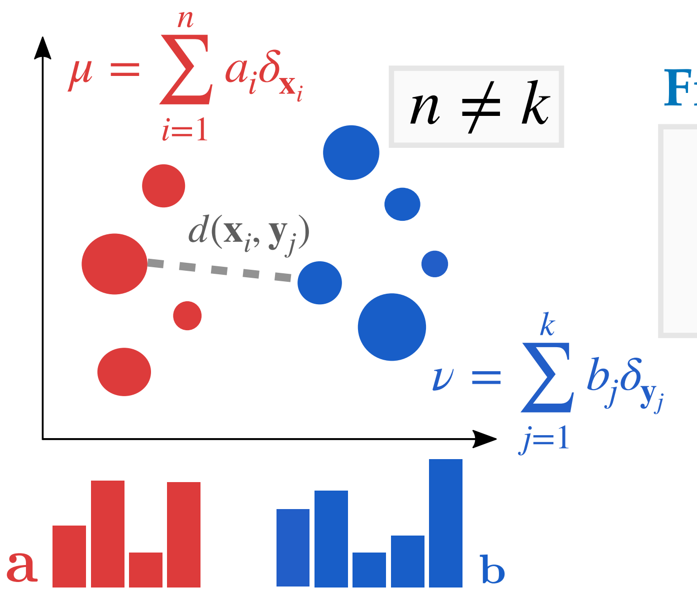
From Wasserstein to Gromov-Wasserstein

✦ Classical optimal transport (in a nutshell)



From Wasserstein to Gromov-Wasserstein

✦ Classical optimal transport (in a nutshell)

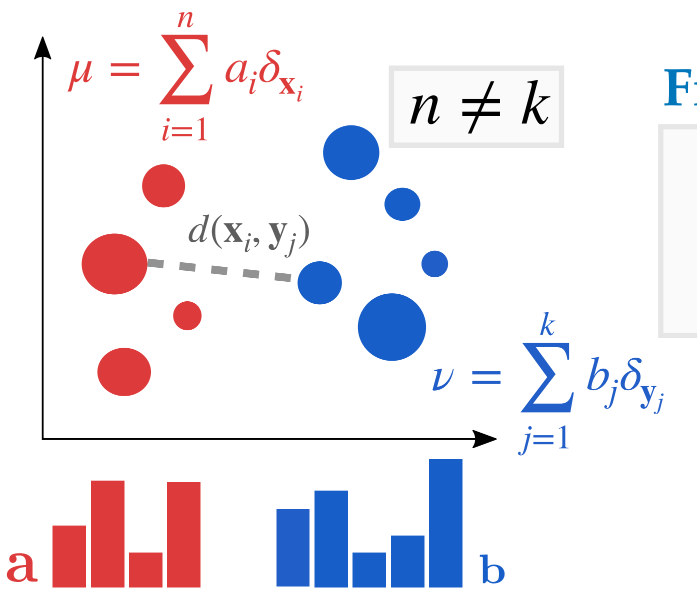


Find plan $T \in \mathbb{R}_+^{n \times k}$

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

From Wasserstein to Gromov-Wasserstein

✦ Classical optimal transport (in a nutshell)



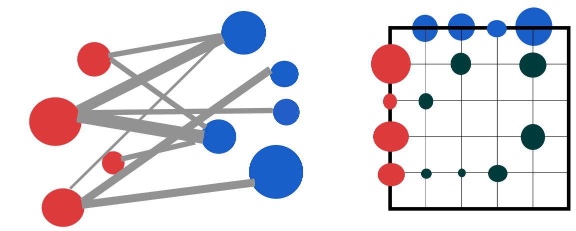
Find plan $T \in \mathbb{R}_+^{n \times k}$

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

which constraints ?

Coupling $\Pi(a, b)$

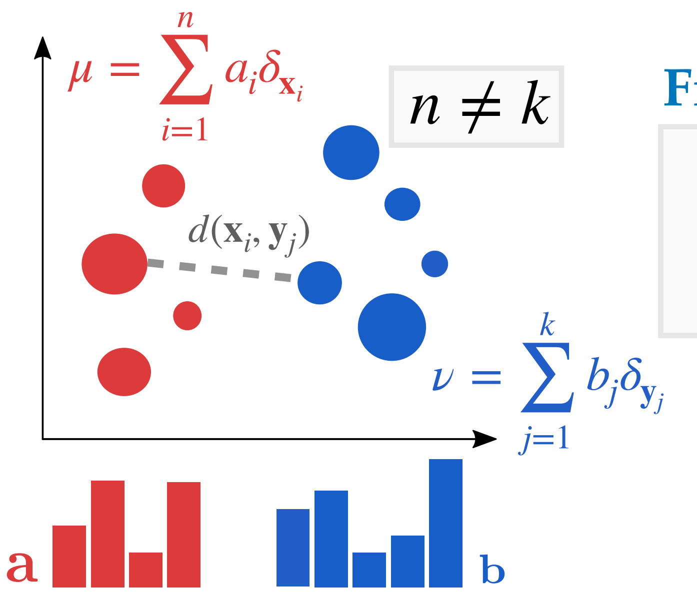
$$T 1_k = a$$
$$T^\top 1_n = b$$



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

From Wasserstein to Gromov-Wasserstein

✦ Classical optimal transport (in a nutshell)



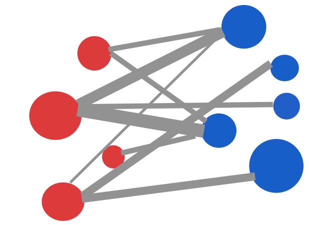
Find plan $T \in \mathbb{R}_{+}^{n \times k}$

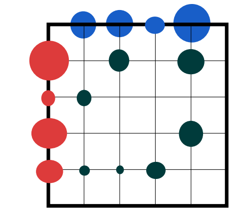
$$\min_{T} \sum_{ij} d(\mathbf{x}_i, \mathbf{y}_j) T_{ij}$$

which constraints ?

Coupling

$\Pi(a, b)$ $\quad T 1_k = a \quad T^{\top} 1_n = b$





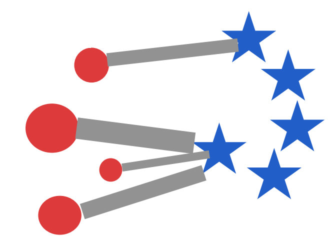
$T \in [0,1]^{n \times k}$

Semi-relaxed coupling

$$T 1_k = a$$

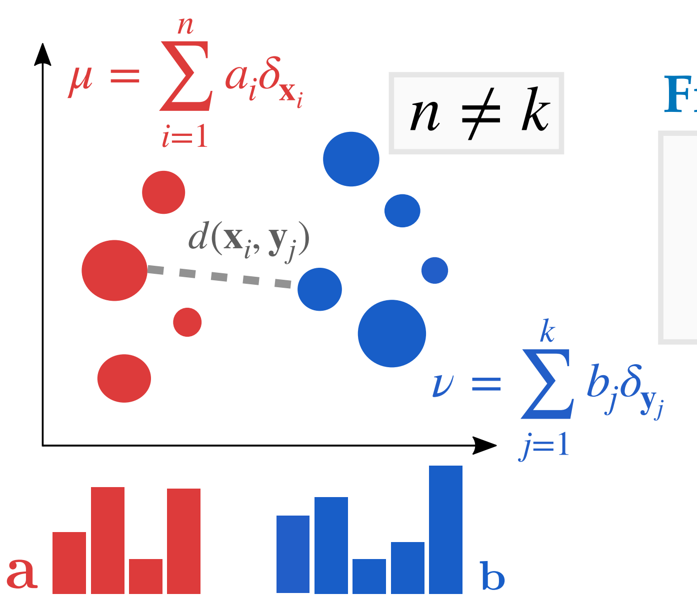
$\sim$ `assign` in k-means

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



From Wasserstein to Gromov-Wasserstein

✦ Classical optimal transport (in a nutshell)



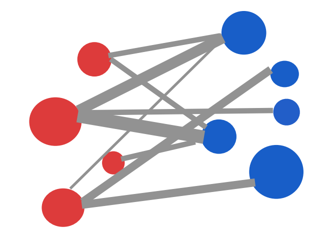
Find plan $T \in \mathbb{R}_+^{n \times k}$

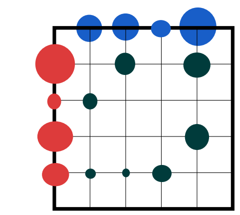
$$\min_T \sum_{ij} d(\mathbf{x}_i, \mathbf{y}_j) T_{ij}$$

which constraints ?

Coupling $\quad T1_k = a$

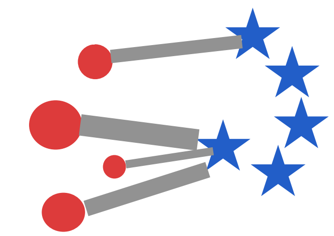
$\Pi(a, b) \quad T^\top 1_n = b$





Semi-relaxed coupling

$$T1_k = a$$



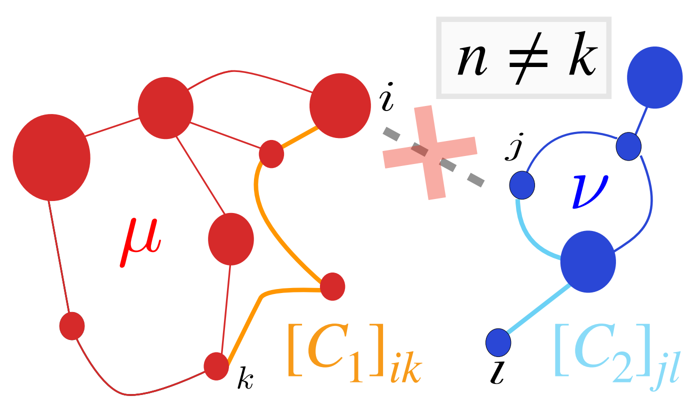
$\sim$ `assign` in k-means

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

(Sturm, 2012)

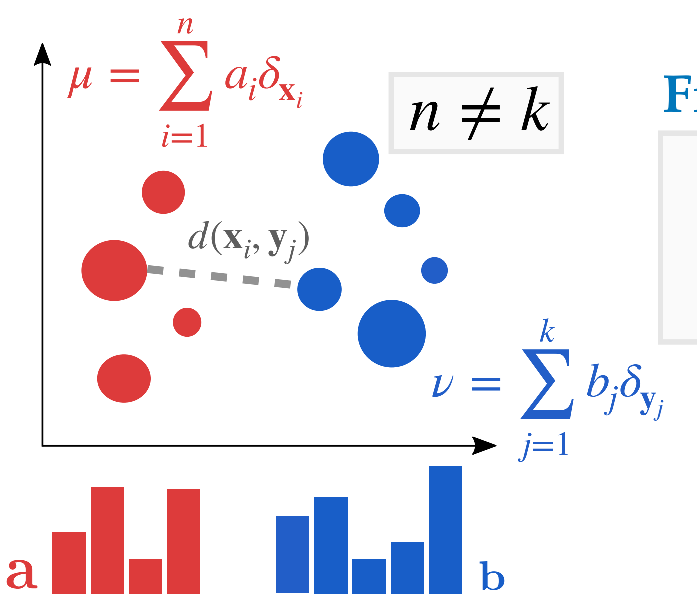
(Mémoli, 2011)

✦ Gromov-Wasserstein



From Wasserstein to Gromov-Wasserstein

✦ Classical optimal transport (in a nutshell)



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

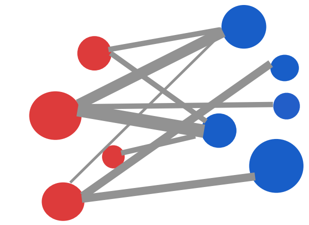
$$\min_{T} \sum_{ij} d(\mathbf{x}_i, \mathbf{y}_j) T_{ij}$$

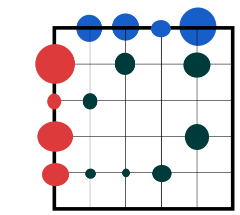
which constraints ?

Coupling $\Pi(a, b)$

$T\mathbb{1}_k = a$

$T^\top \mathbb{1}_n = b$





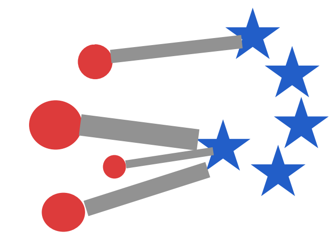
$T \in [0,1]^{n \times k}$

Semi-relaxed coupling

$T\mathbb{1}_k = a$

$\sim$ `assign` in k-means

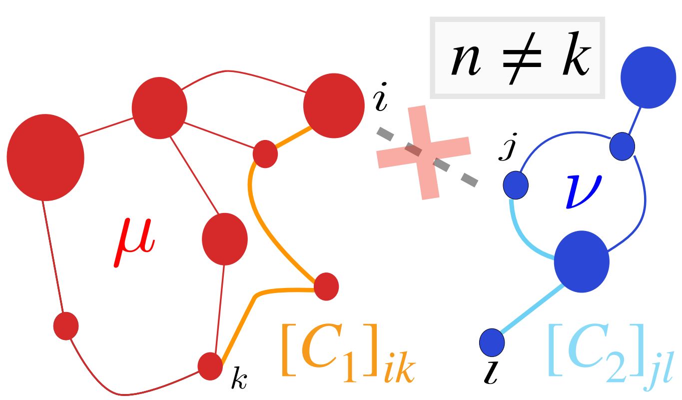
$\sim \min_{b} \min_{T \in \Pi(a,b)} \sum_{ij} d(\mathbf{x}_i, \mathbf{y}_j) T_{ij}$



(Sturm, 2012)

(Mémoli, 2011)

✦ Gromov-Wasserstein

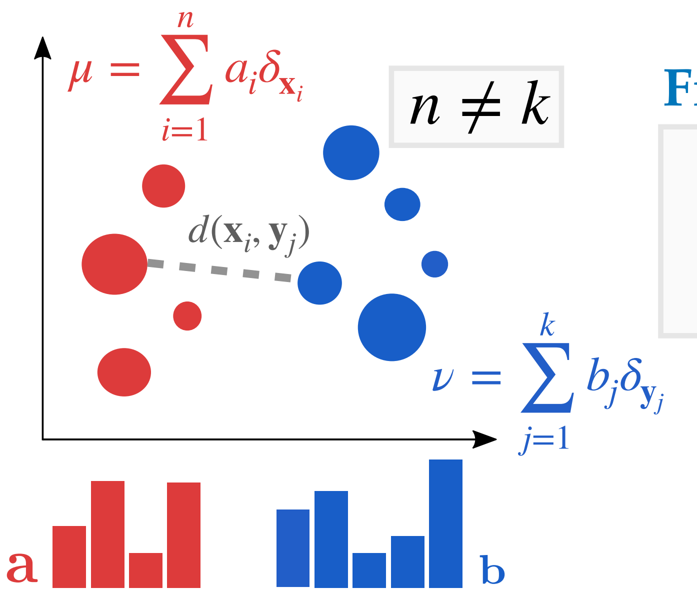


Quadratic OT: find the plan

$$\min_{T \in \Pi(a,b)} \sum_{ijkl} L\left([C_1]_{ik}\ ,\ [C_2]_{jl}\right) T_{ij} T_{kl}$$

From Wasserstein to Gromov-Wasserstein

✦ Classical optimal transport (in a nutshell)

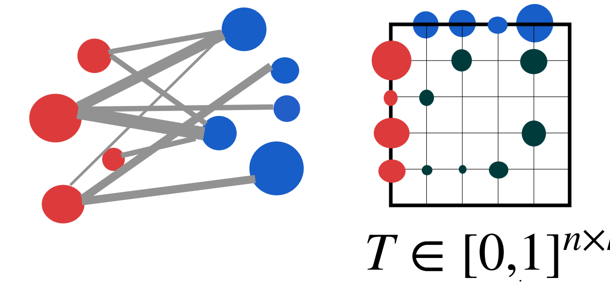


Find plan $T \in \mathbb{R}_+^{n \times k}$

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

which constraints ?

→ **Coupling** $\Pi(a, b)$ $\quad T 1_k = a$, $\quad T^\top 1_n = b$

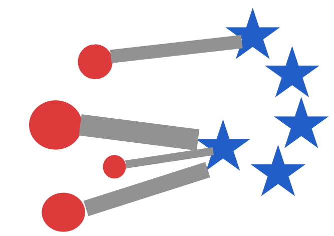


→ **Semi-relaxed coupling**

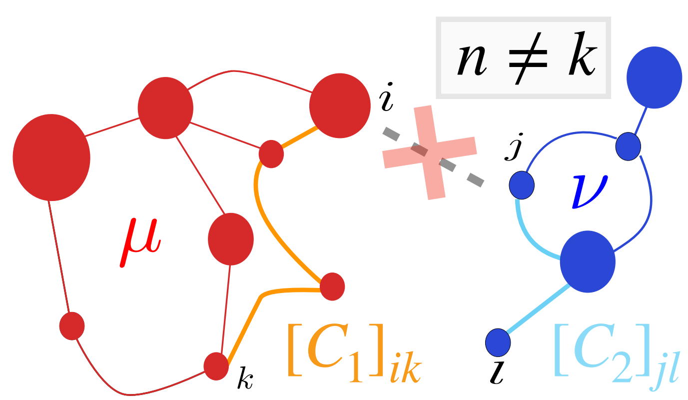
$$T 1_k = a$$

$\sim$ `assign` in k-means

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



✦ Gromov-Wasserstein



(Sturm, 2012)

(Mémoli, 2011)

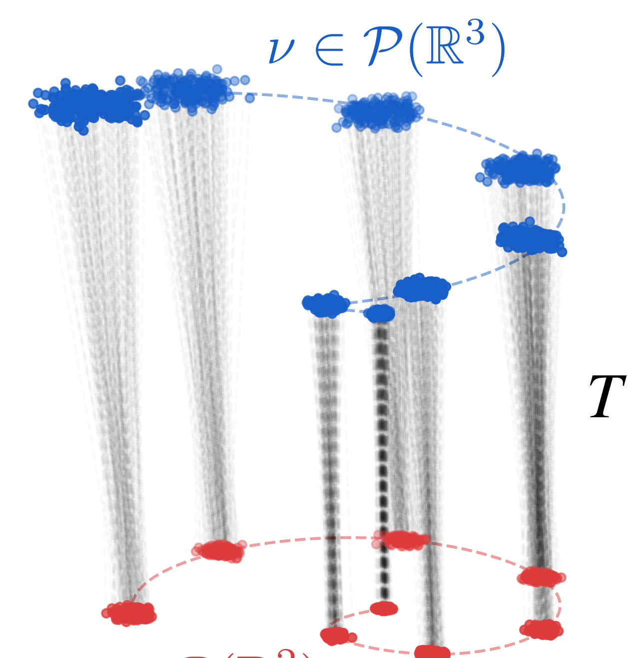
Quadratic OT: find the plan

$$\min_{T \in \Pi(a,b)} \sum_{ijkl} L\Big([C_1]_{ik} \,,\, [C_2]_{jl} \Big) T_{ij} T_{kl}$$

- L measures distortion

$$\Big| [C_1]_{ik} - [C_2]_{jl} \Big|^2$$

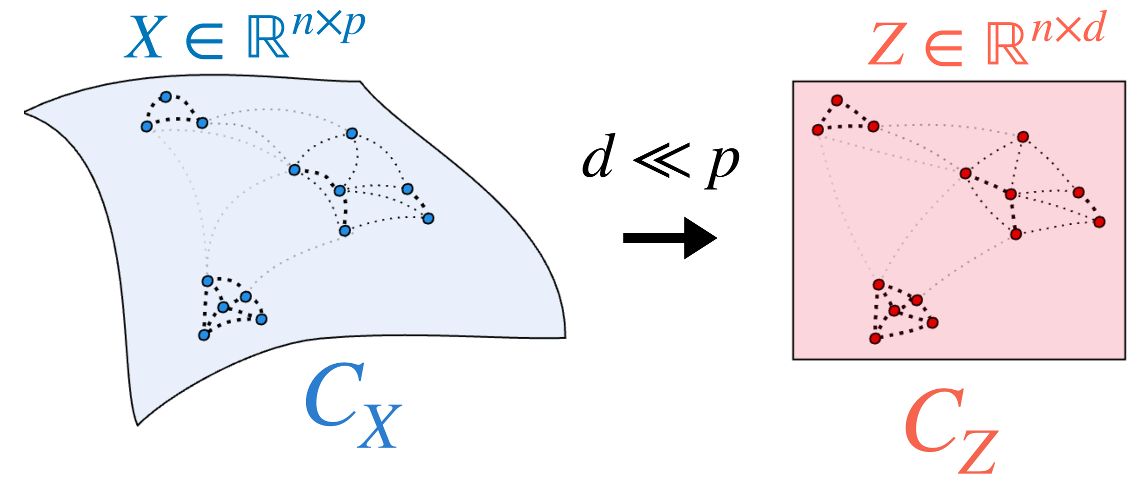
- Goal : **preserving pairwise connectivity**
- Distance w.r.t. isomorphisms
- Difficult quadratic problem (NP-hard)



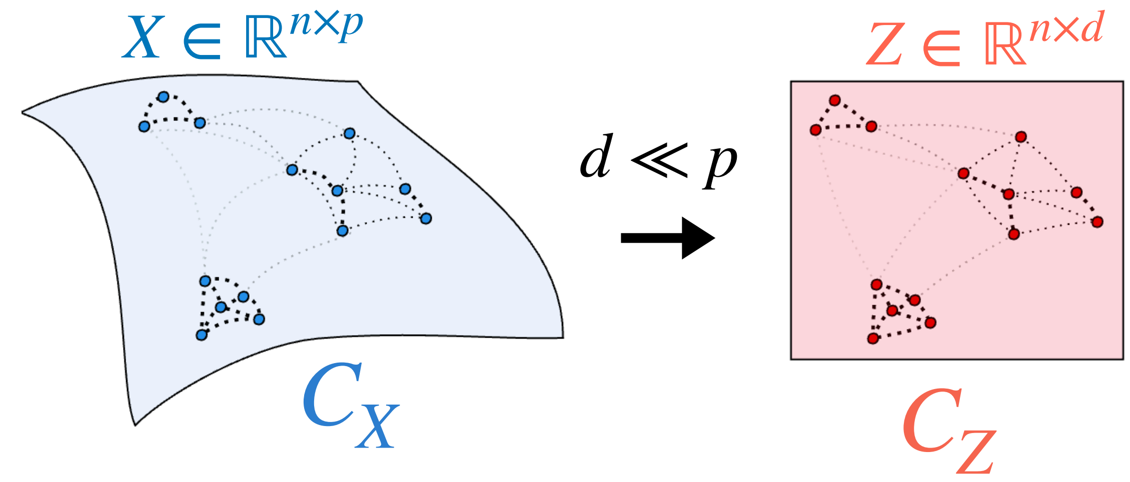
DR as OT in disguise

✦ 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

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

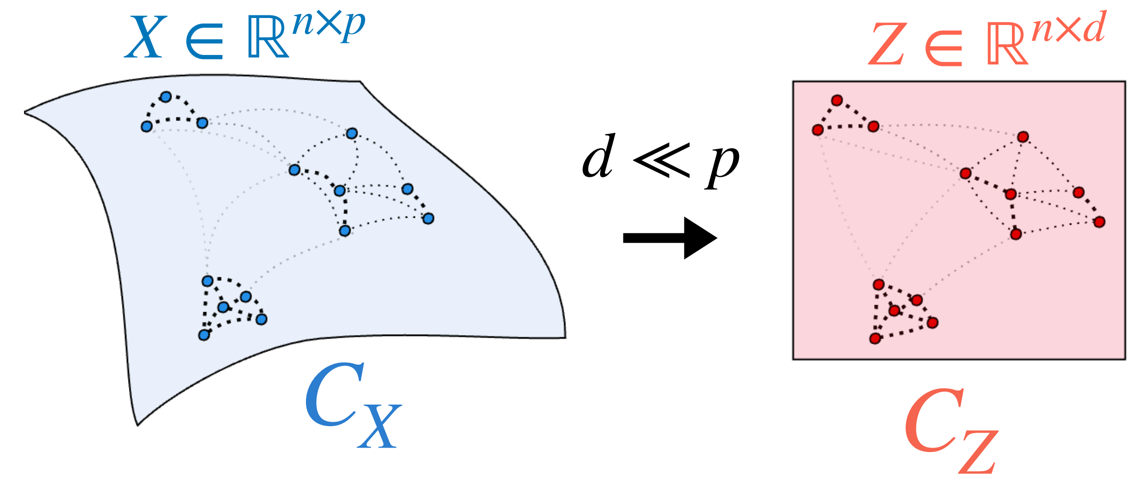
equiv $\Updownarrow$

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 $\Updownarrow$

Permutation equivariance

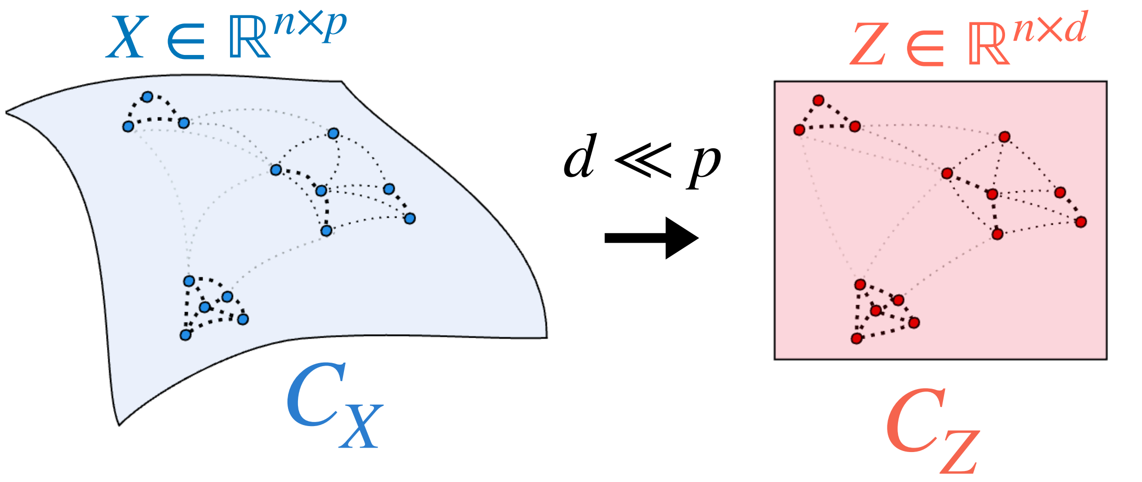
$\forall P, C_{PZ} = PC_ZP^\top$

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

equal $\Updownarrow$

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

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

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

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

equal

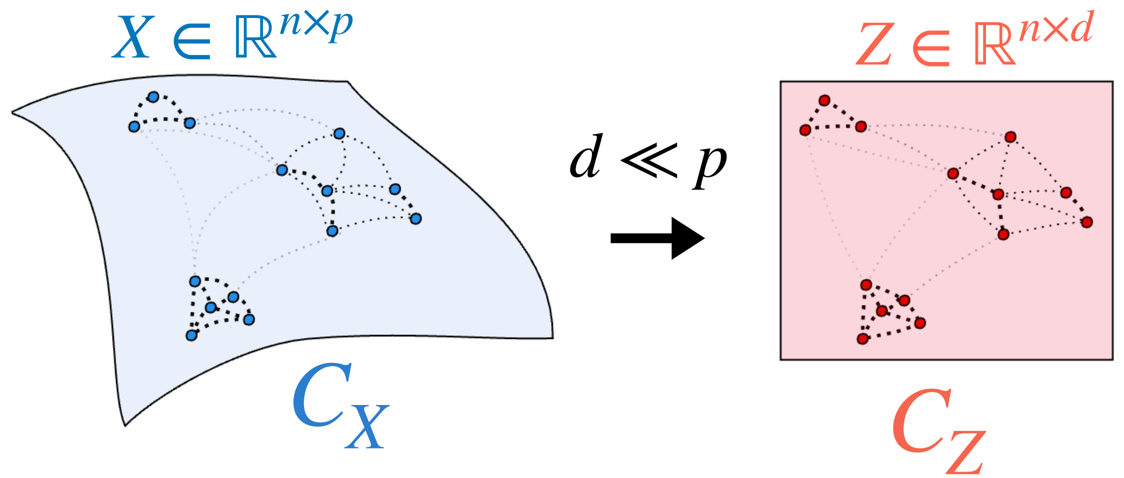
$$\min_{Z \in \mathbb{R}^{n \times d}} \min_{P} \sum_{i,j,k,l=1}^{n} L\left([C_X]_{ik}\ ,\ [C_Z]_{jl}\right) P_{ij} P_{kl}$$

?

✦ Gromov-Wasserstein projection

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

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

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

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

equal

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

?

✦ Gromov-Wasserstein projection

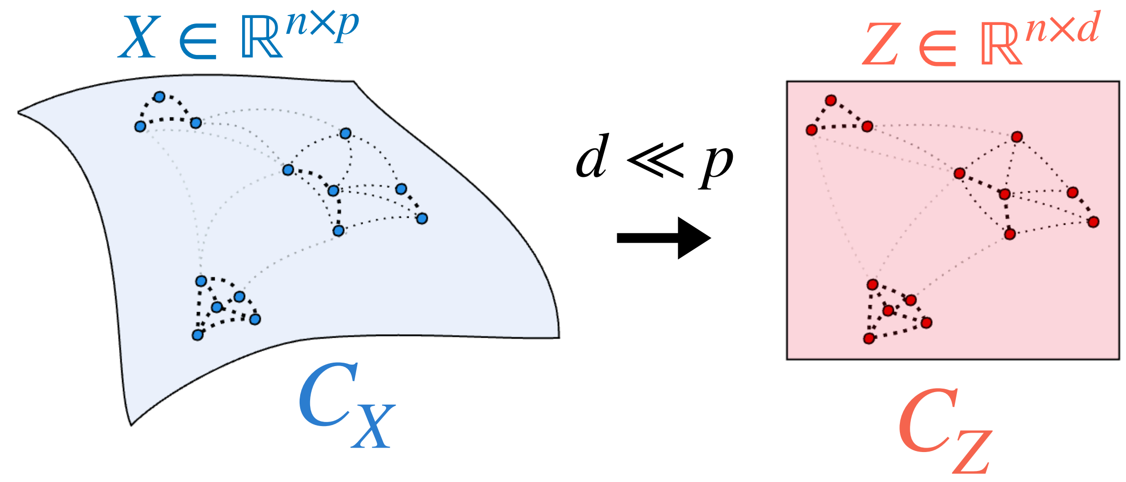
$$\min_{Z\in\mathbb{R}^{n\times d}} \min_{T\in\Pi(\frac{1_n}{n},\frac{1_n}{n})} \sum_{ijkl} L\left([C_X]_{ik}\ ,\ [C_Z]_{jl}\right) T_{ij}T_{kl}$$

✦ Equivalence holds for

Spectral methods

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

DR as OT in disguise



✦ Dimension reduction

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

equiv

Permutation equivariance

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

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

equal

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

?

✦ Gromov-Wasserstein projection

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

✦ Equivalence holds for

Spectral methods

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

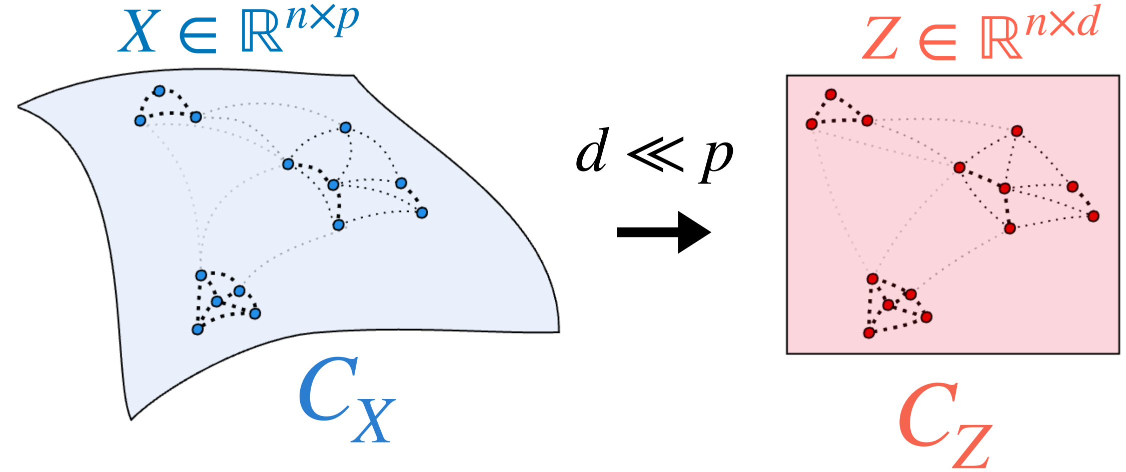
Neighbor embedding methods

✦ 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

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

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

equal

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

?

✦ Gromov-Wasserstein projection

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

✦ Equivalence holds for

Spectral methods

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

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

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

and its usual normalizations

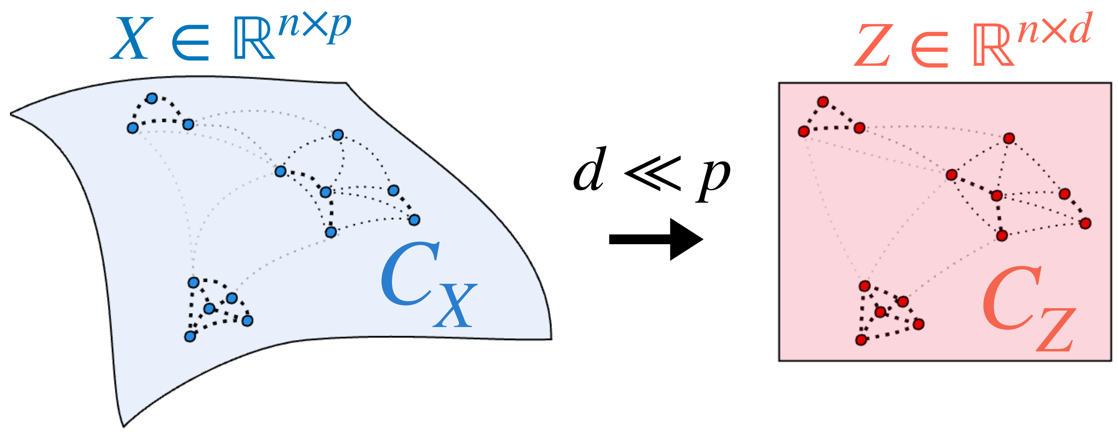
$1_n^\top K_Z 1_n = 1$, $K_Z 1_n = 1_n$, $K_Z^\top 1_n = 1_n$ + $K_Z 1_n = 1_n$

(Sinkhorn & Knopp, 1967)

To improve as C_X generally not CPD

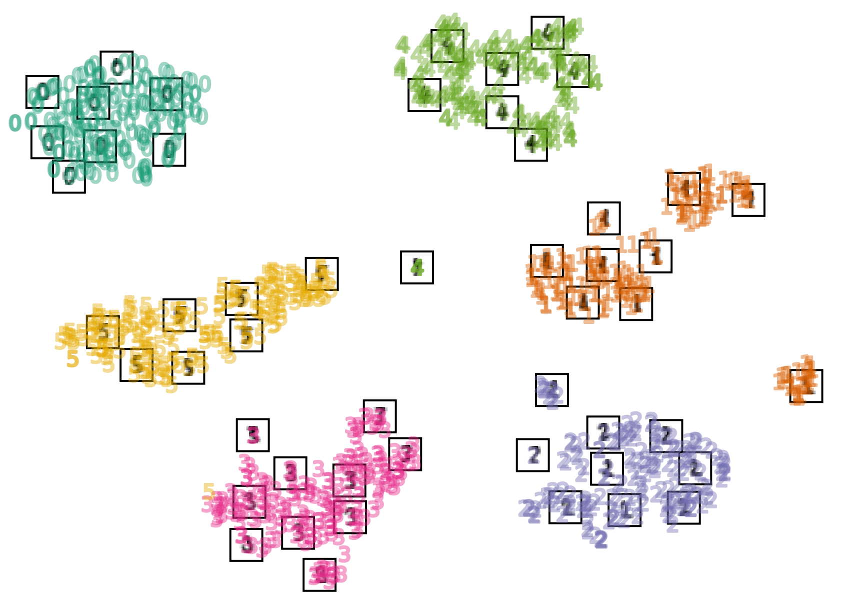
Distributional reduction

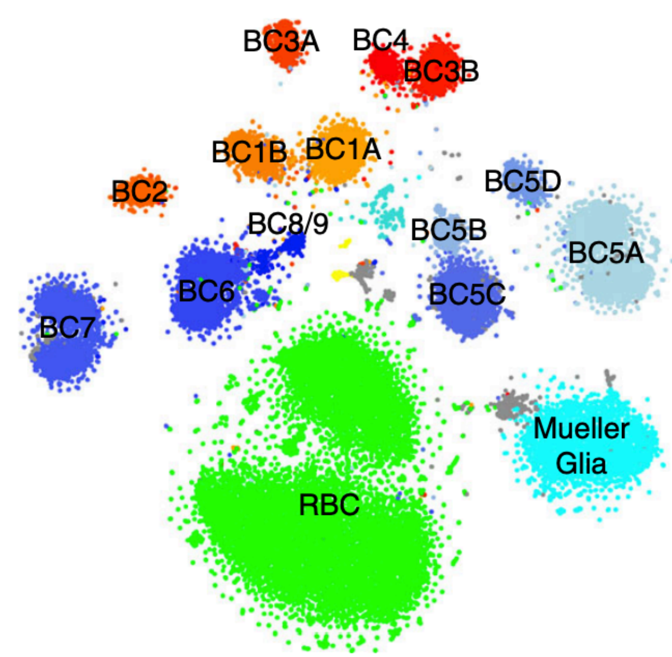
Distributional Reduction



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

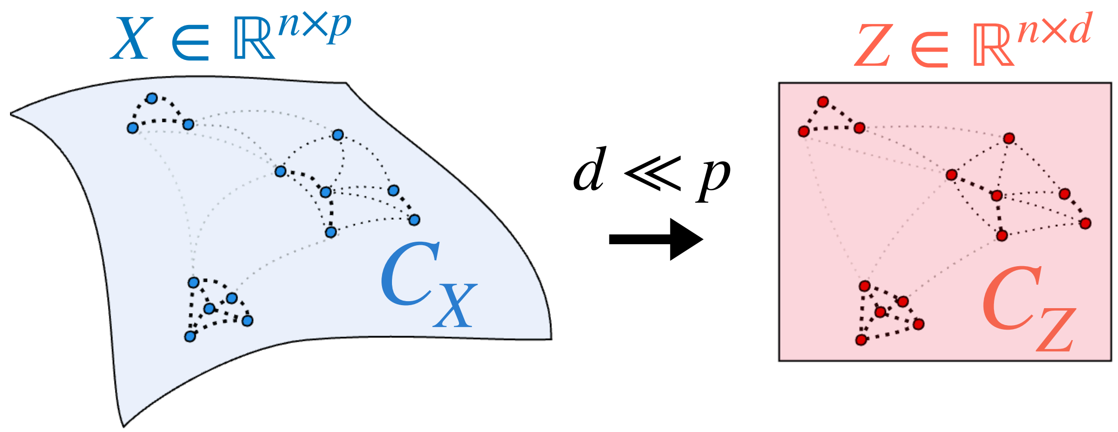
✦ Motivation



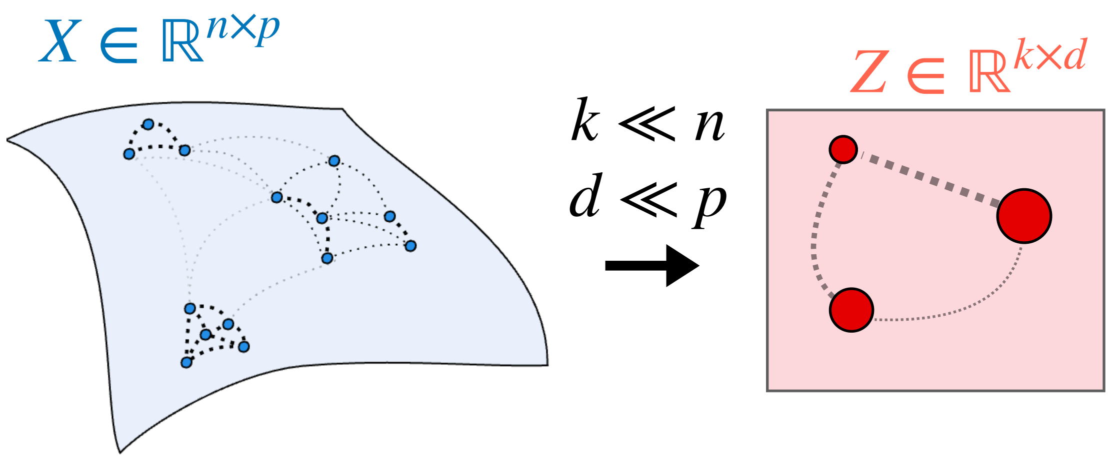


(Shekhar et al., 2016)

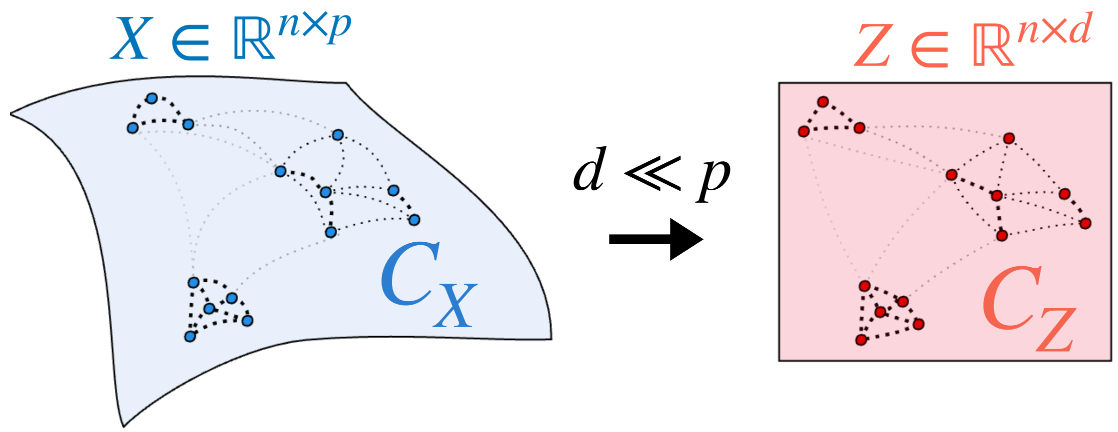
Distributional Reduction



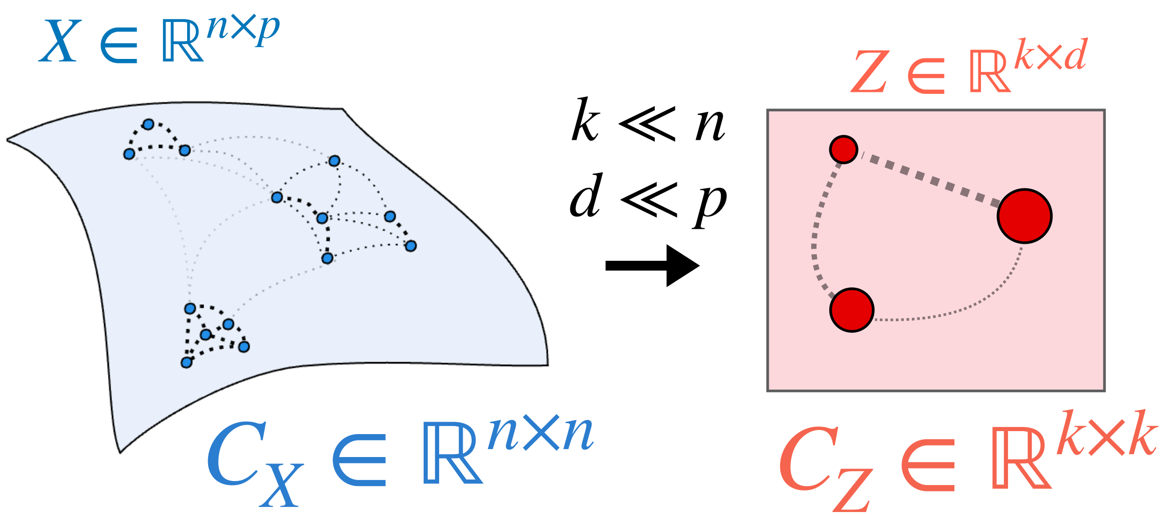
$$\min_{Z \in \mathbb{R}^{n \times d}} \mathrm{GW}_L\left(C_X, C_Z, \frac{1_n}{n}, \frac{1_n}{n}\right)$$



Distributional Reduction

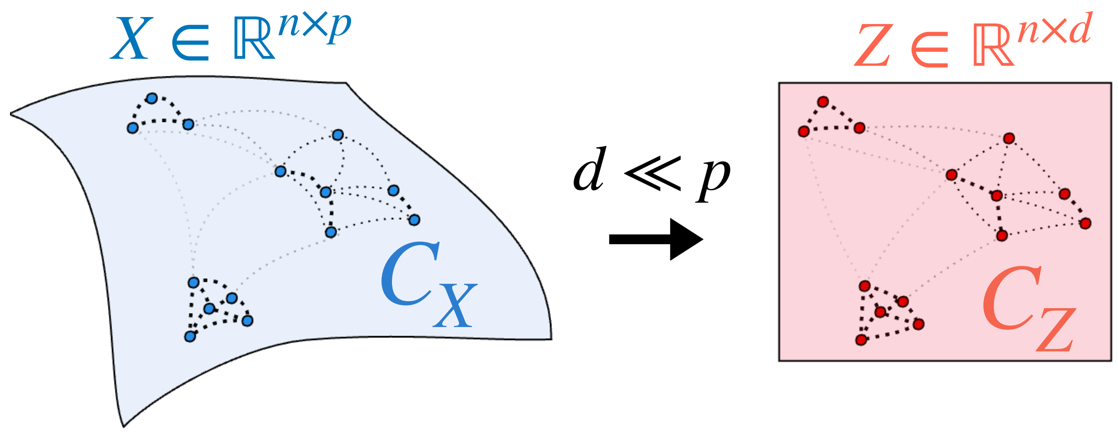


$$\min_{Z \in \mathbb{R}^{n \times d}} \mathrm{GW}_L\left(C_X, C_Z, \frac{1_n}{n}, \frac{1_n}{n}\right)$$

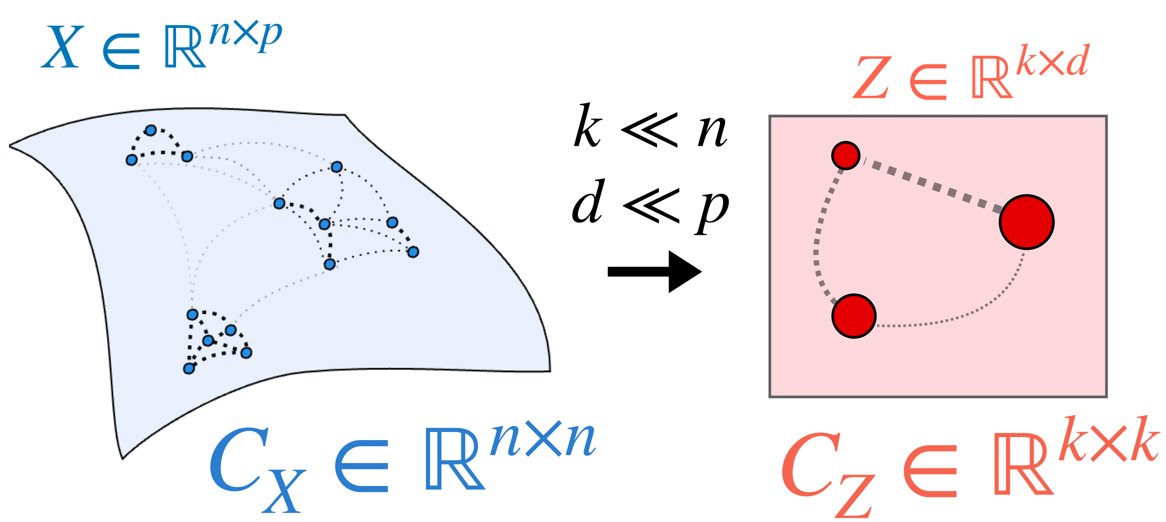


✦ **GW projection** $\min_{\nu \in \mathscr{P}_k(\mathbb{R}^d)} \mathrm{GW}(\hat{\mu}_n, \nu)$

Distributional Reduction



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



✦ Optimization problem

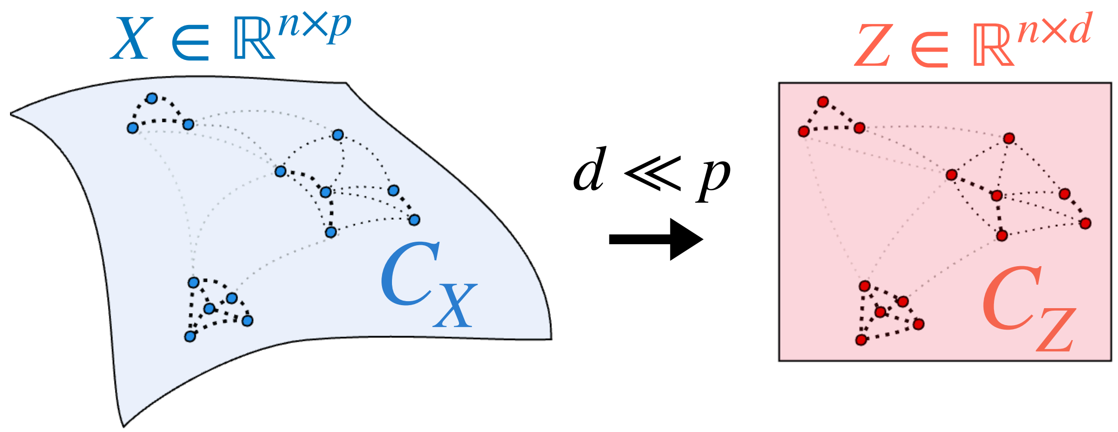
$$\min_{Z\in\mathbb{R}^{k\times d}} \min_{b\in\Sigma_k} \mathrm{GW}_L\left(C_X, C_Z, \frac{1_n}{n}, b\right)$$

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

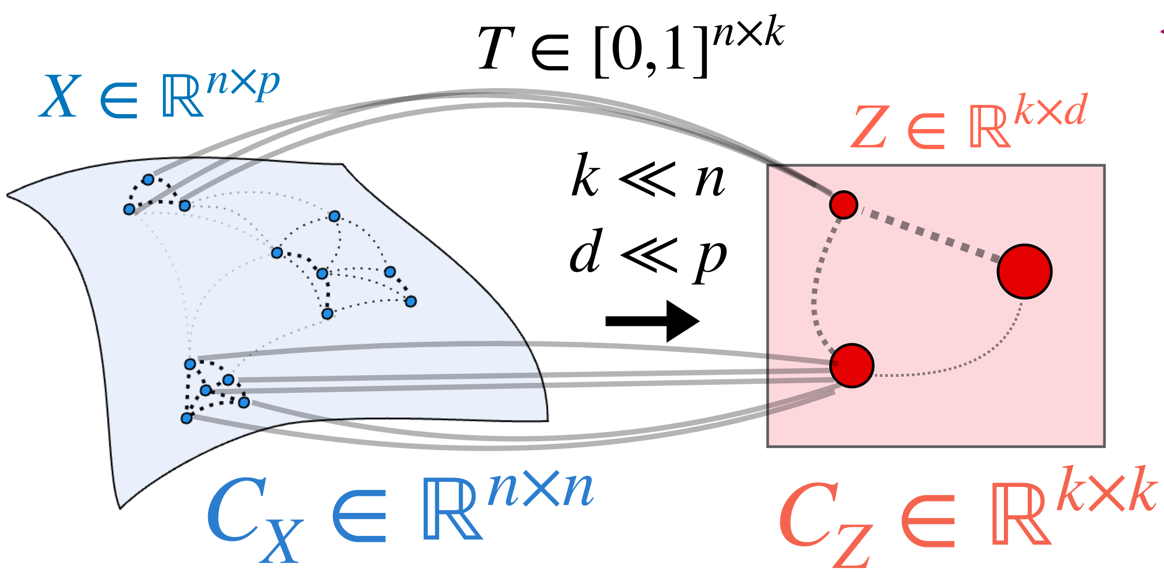
✦ GW projection

$$\min_{\nu\in\mathscr{P}_k(\mathbb{R}^d)} \mathrm{GW}(\hat{\mu}_n, \nu)$$

Distributional Reduction



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



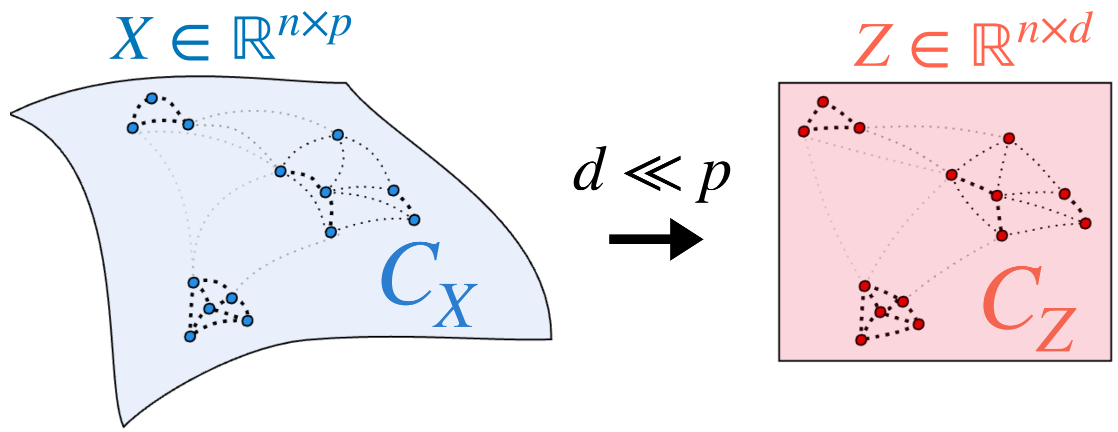
✦ **GW projection** $\min_{\nu\in\mathscr{P}_k(\mathbb{R}^d)} \mathrm{GW}(\hat{\mu}_n, \nu)$

✦ **Optimization problem**

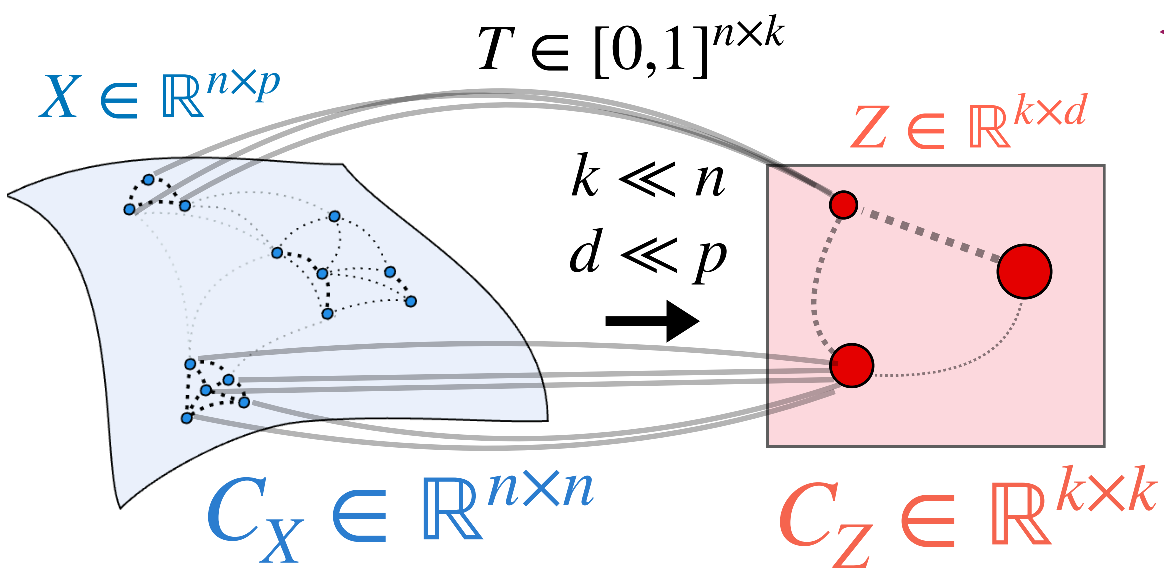
$$\min_{Z\in\mathbb{R}^{k\times d}} \min_{b\in\Sigma_k} \mathrm{GW}_L\left(C_X, C_Z, \frac{1_n}{n}, b\right)$$

- 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)

Distributional Reduction



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



✦ **GW projection** $\min_{\nu\in\mathscr{P}_k(\mathbb{R}^d)} \mathrm{GW}(\hat{\mu}_n, \nu)$

✦ **Optimization problem**

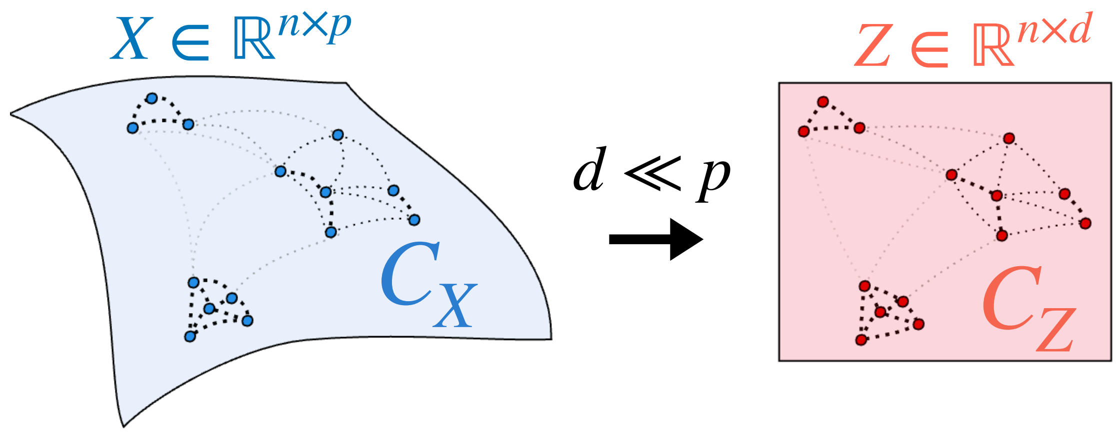
$$\min_{Z\in\mathbb{R}^{k\times d}} \min_{b\in\Sigma_k} \mathrm{GW}_L\left(C_X, C_Z, \frac{1_n}{n}, b\right)$$

- 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)

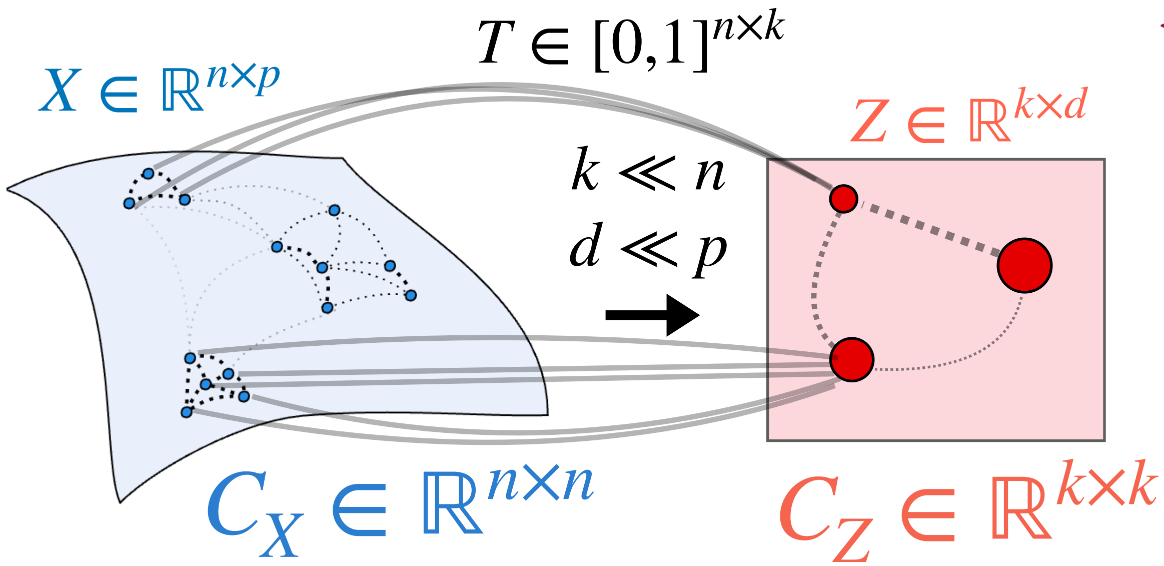
✦ **A semi-relaxed objective** (Vincent-Cuaz et al., 2022)

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

Distributional Reduction



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



✦ **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)

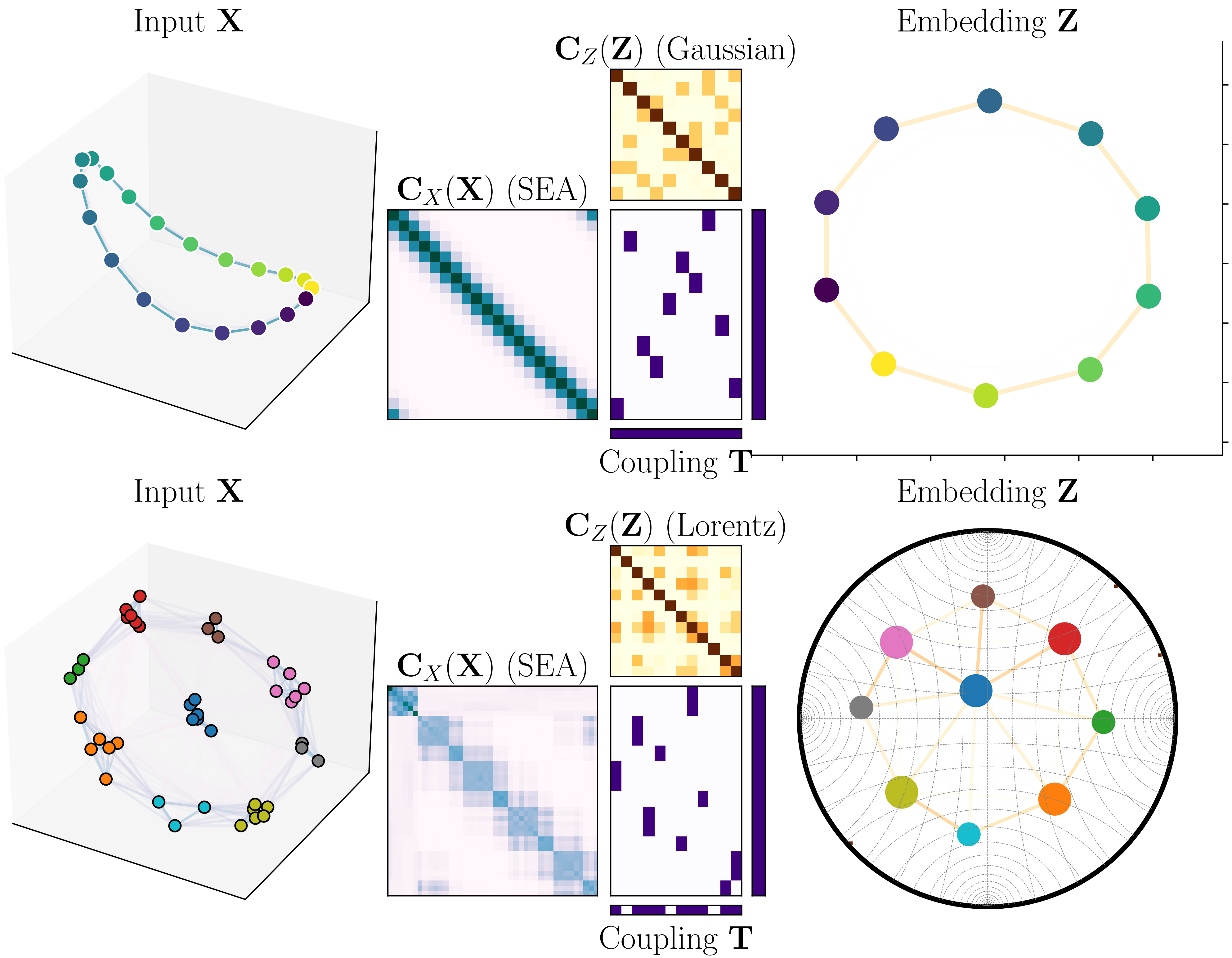
✦ **GW projection** $\min_{\nu\in\mathscr{P}_k(\mathbb{R}^d)} \mathrm{GW}(\hat{\mu}_n, \nu)$

✦ **A semi-relaxed objective**
(Vincent-Cuaz et al., 2022)

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

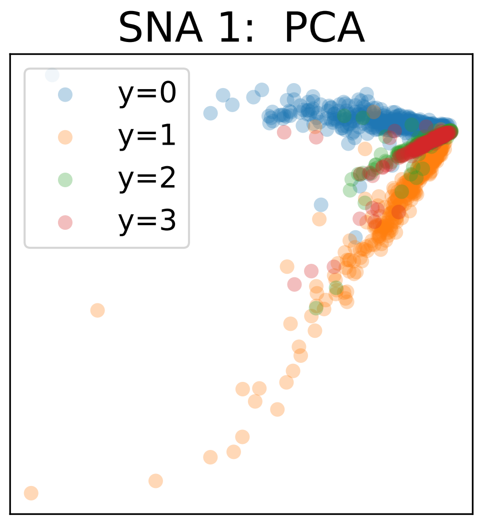
- Non-convex problem
- BCD: alternates optim in Z, in T
- Optim in T: CG solver in $O(n^2k)$ for $L \in \{\mathrm{KL}, |\cdot|^2\}$
- With low-rank structures $O(nkr + n^2)$

Distributional Reduction



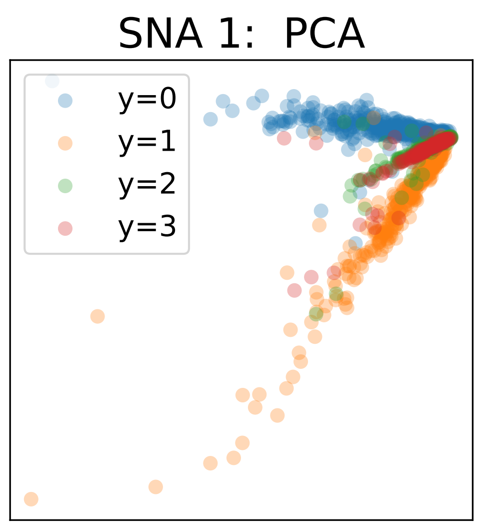
Distributional Reduction

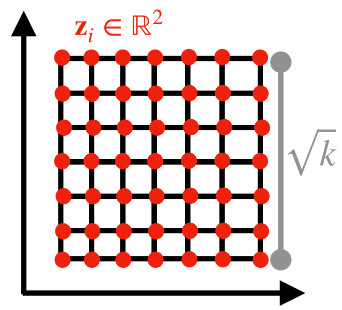
✦ **Single-cell dataset** (Chen et al., 2019) Only solve $\min_{b \in \Sigma_k} \mathrm{GW}_L(C_X, C_Z, \frac{1_n}{n}, b)$



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

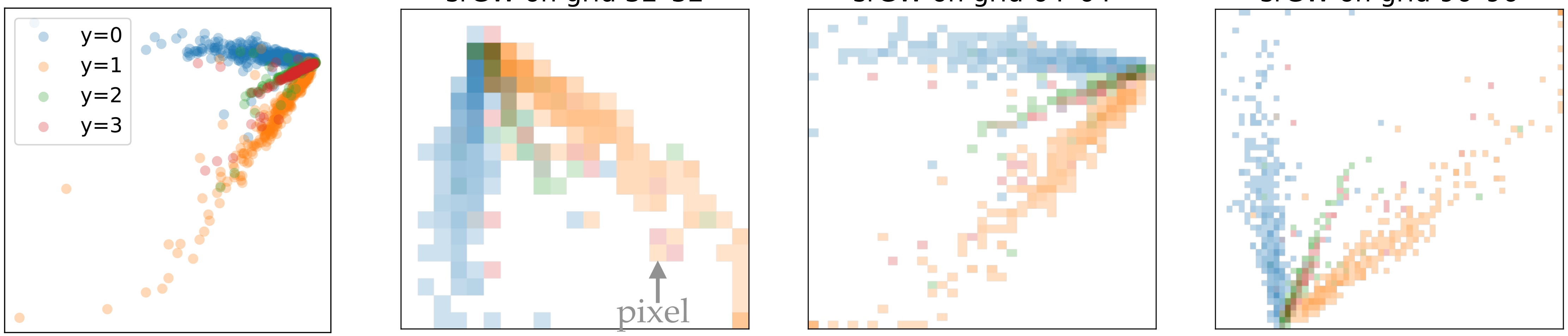




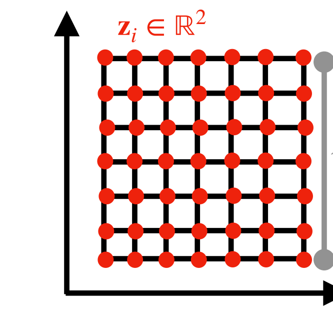
Distributional Reduction

✦ **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 — srGW on grid 32*32 — srGW on grid 64*64 — srGW on grid 96*96



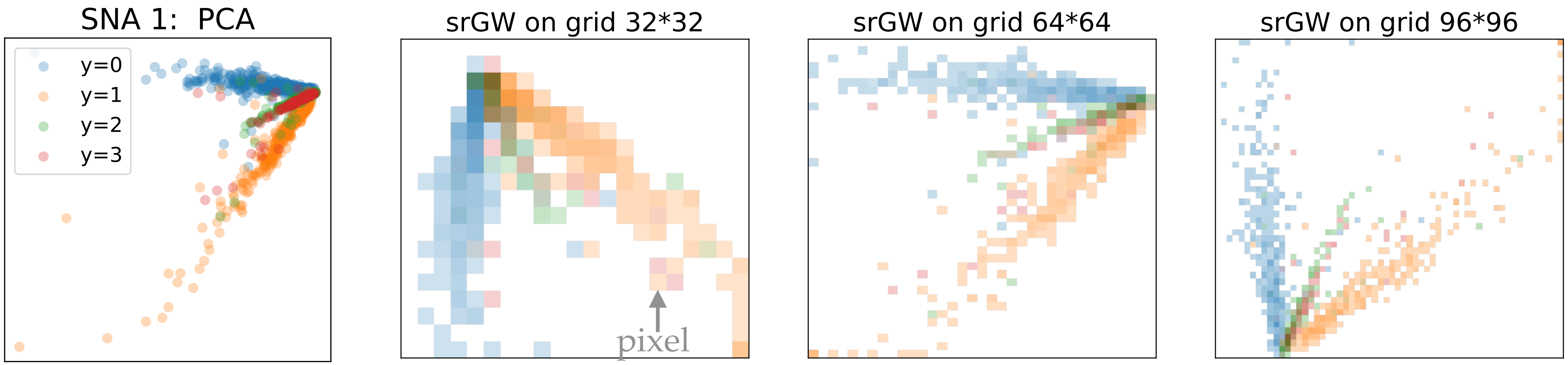
dataset as an image



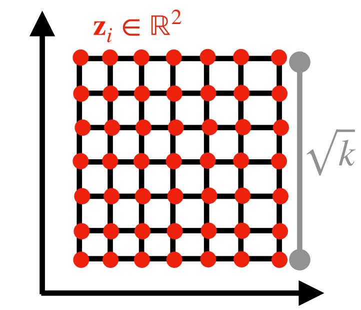
Distributional Reduction

✦ 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** $Z \in \mathbb{R}^{k \times 2}$

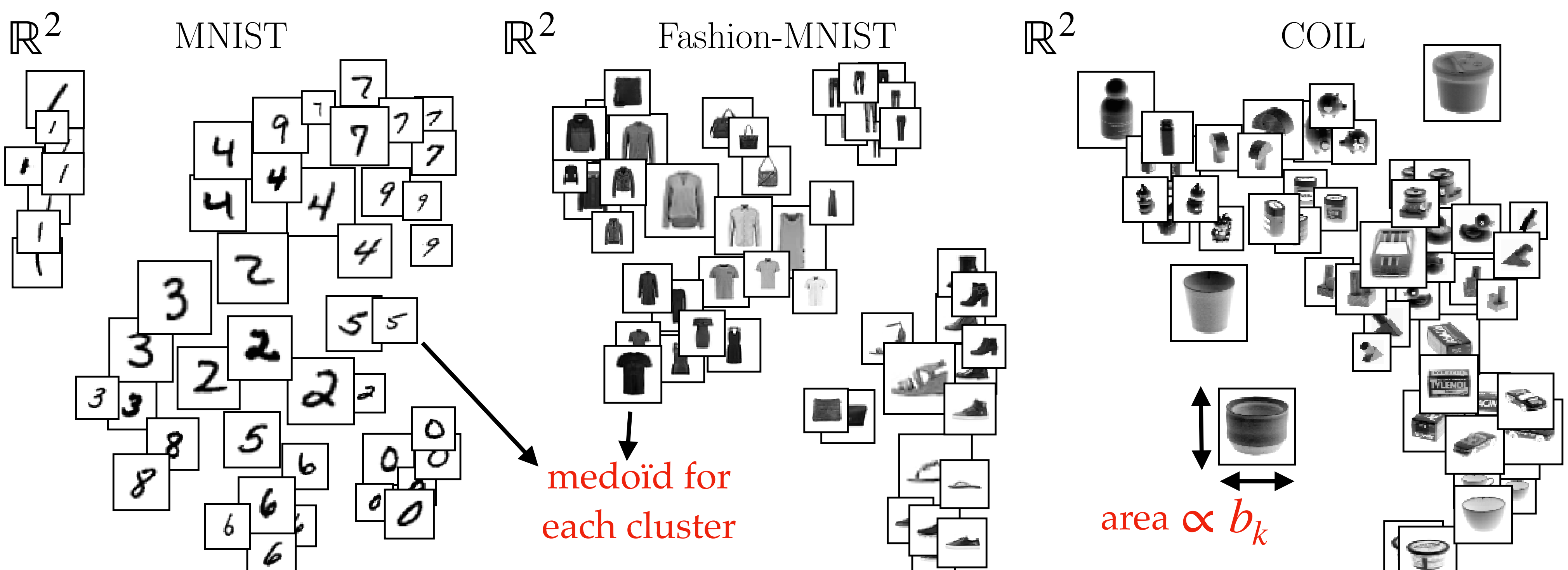


dataset as an image

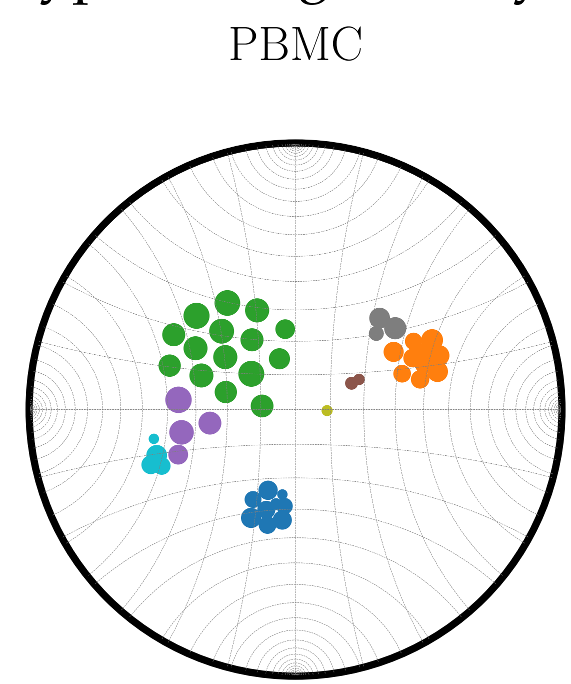


✦ Image datasets

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**

Distributional Reduction

✦ Comparison with DR then clustering or clustering then DR

✦ Silhouette score for DR

- Assign a label to each prototype
- Silhouette of prototype = **avg dist** to points on the same group vs to points on neighboring groups

✦ Homogeneity score for clustering

- Are the classes preserved by the clustering ?

Distributional Reduction

✦ Comparison with DR then clustering or clustering then DR

✦ Silhouette score for DR

- Assign a label to each prototype

- Silhouette of prototype = **avg dist** to points on the same group vs to points on neighboring groups

✦ Homogeneity score for clustering

- Are the classes preserved by the clustering ?