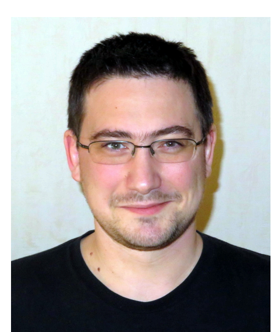
UNIVERSITÉ DE LYON
ENS DE LYON
Inria
l'X ÉCOLE POLYTECHNIQUE IP PARIS
Université Bretagne Sud
EPFL

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


Hugues Van Assel


Cédric Vincent-Cuaz

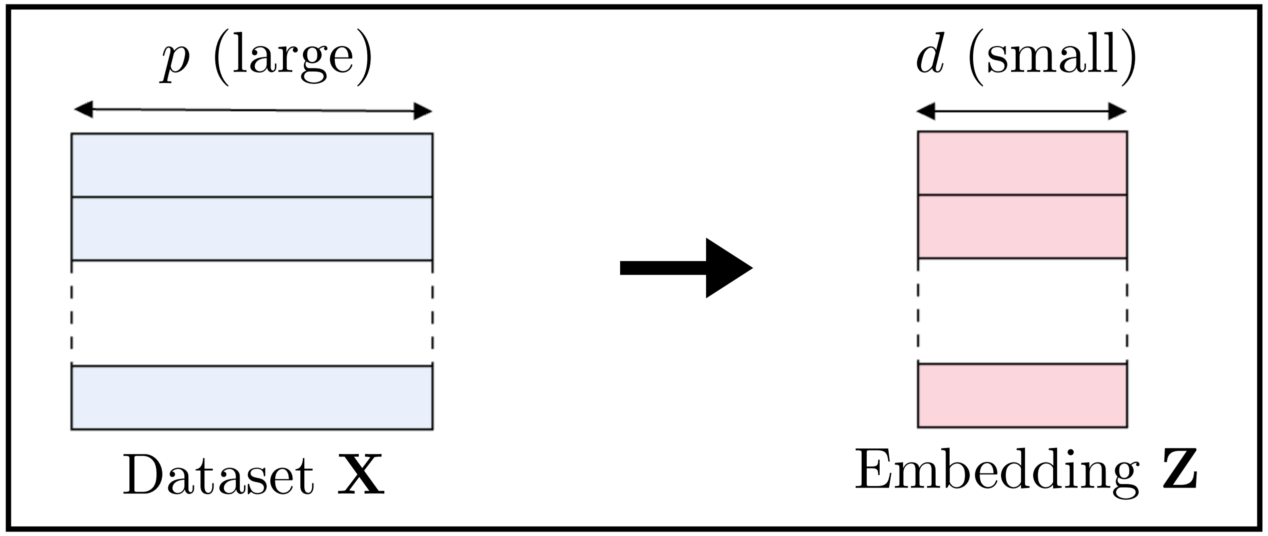

Rémi Flamary


Nicolas Courty

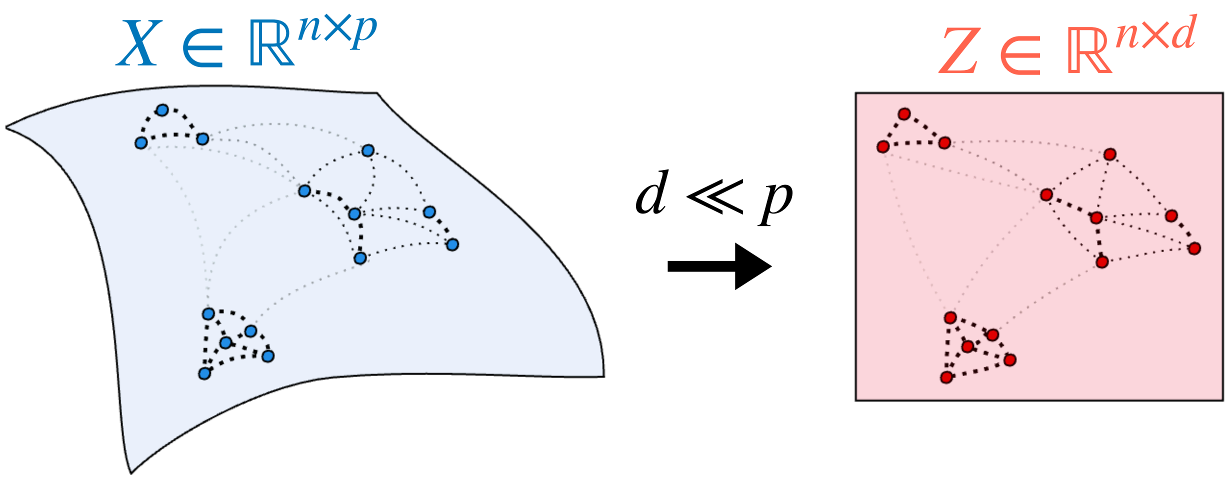

Pascal Frossard


Titouan Vayer

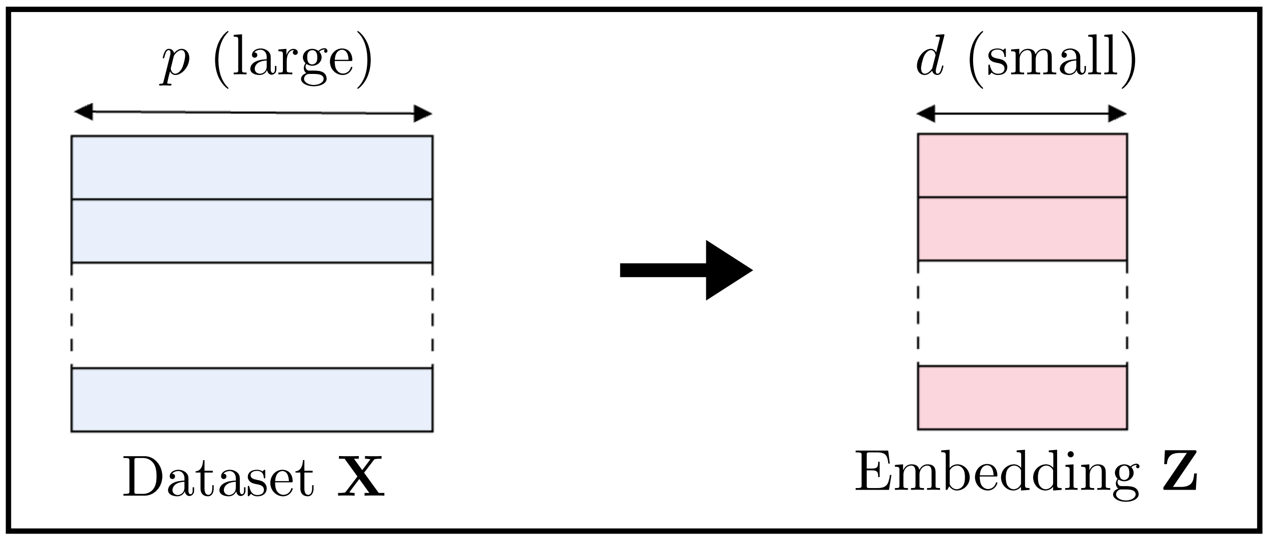
Dimension reduction



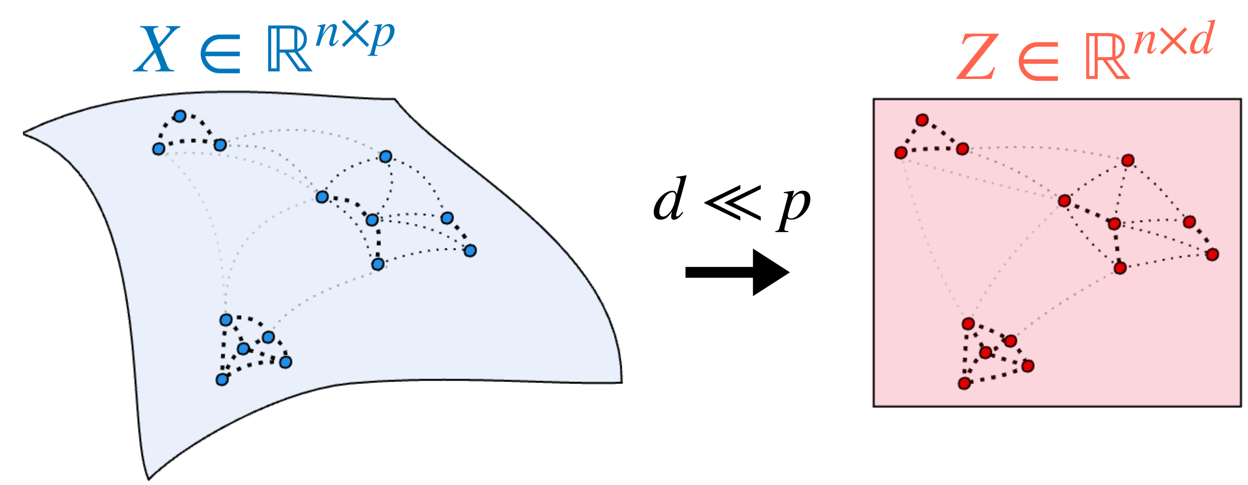
✦ **Preserving geometric properties**



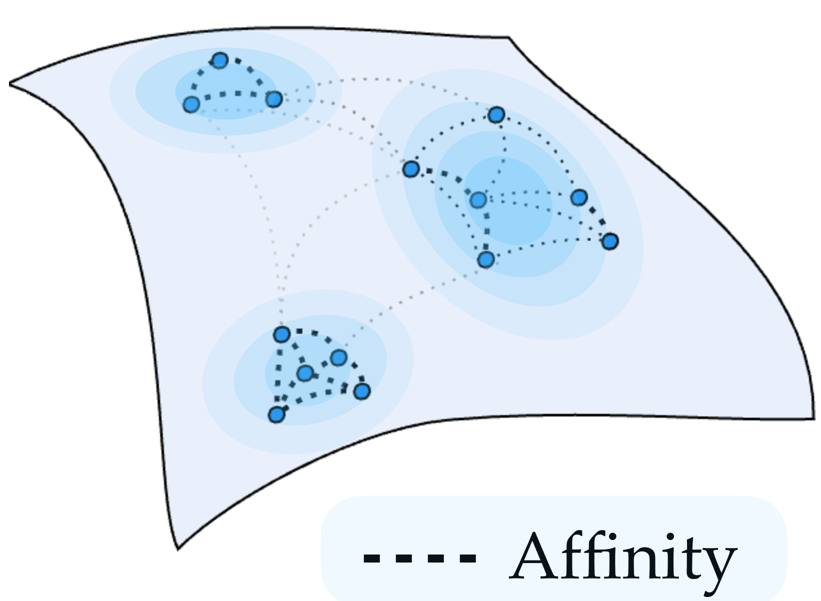
Dimension reduction

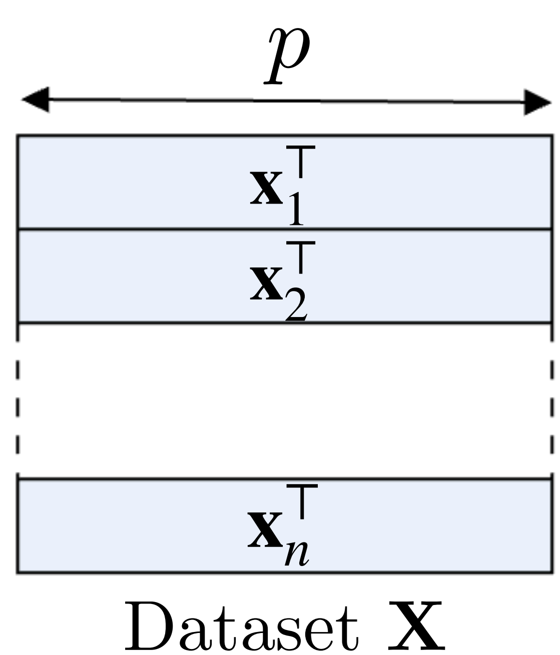


✦ Preserving geometric properties



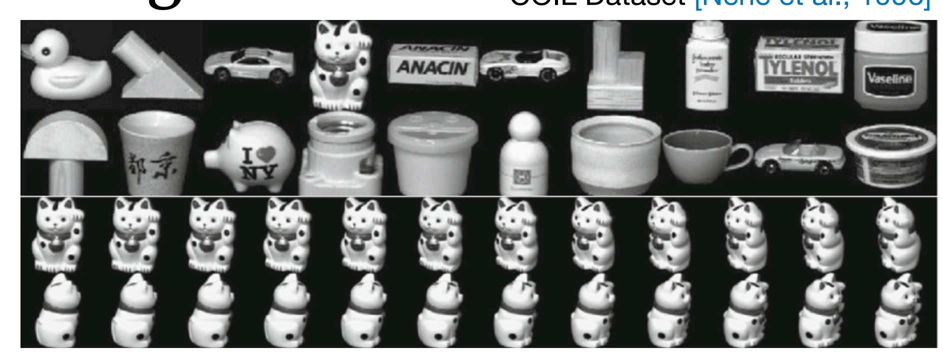
✦ Affinity Matrices



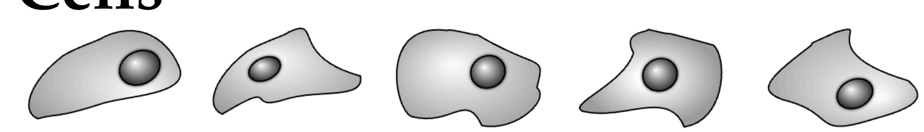


Images

COIL Dataset [Nene et al., 1996]



Cells

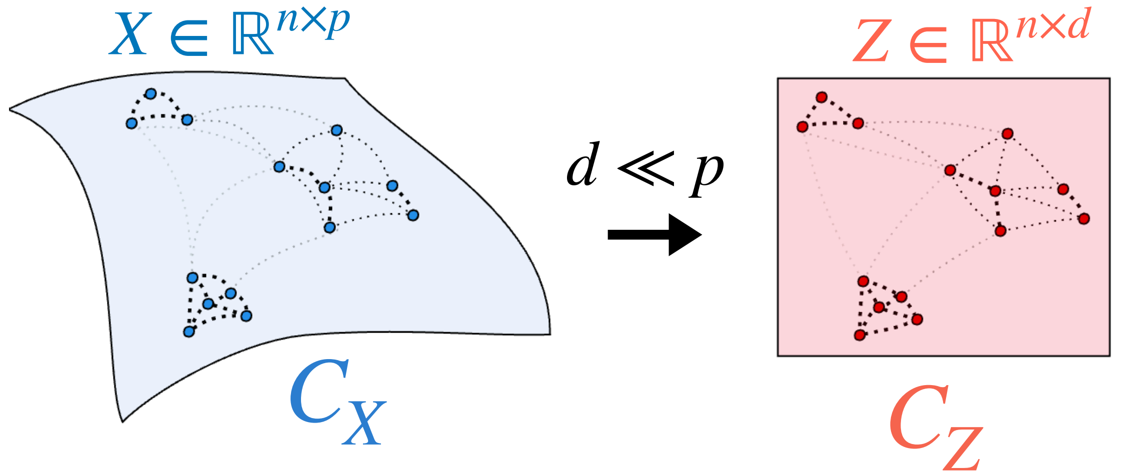


Symmetric matrix with **non-negative coefficients**.

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

Dimension reduction

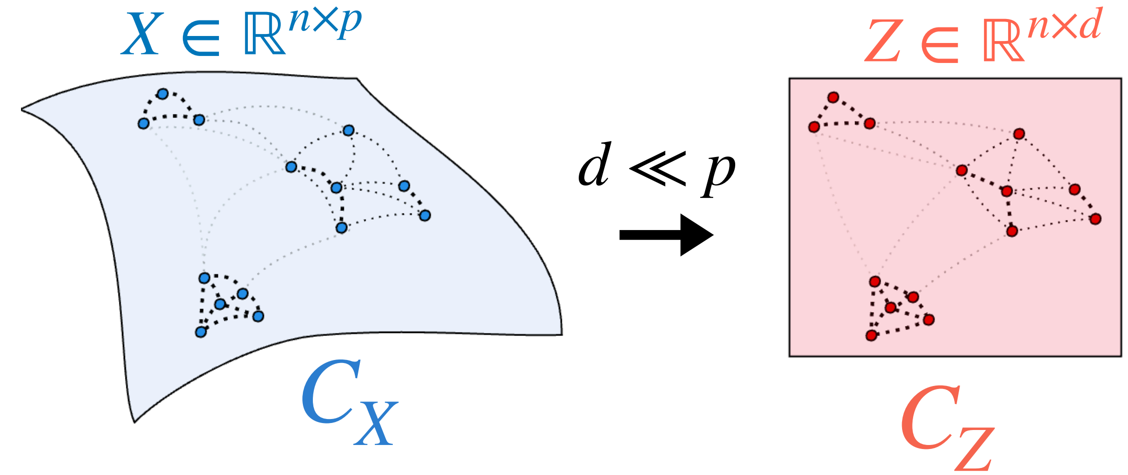
Dimension reduction



✦ A general optimization problem

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

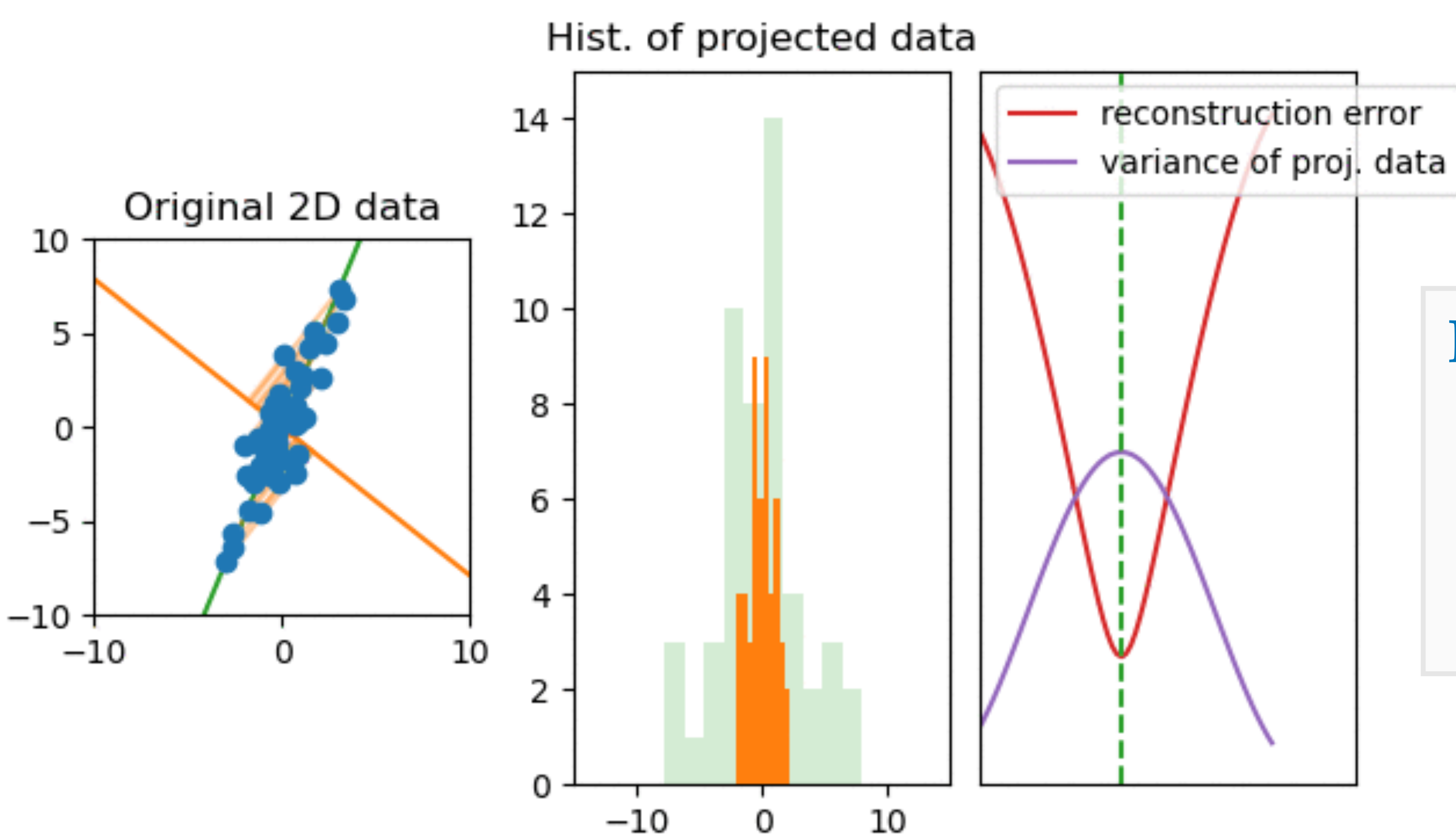
Dimension reduction

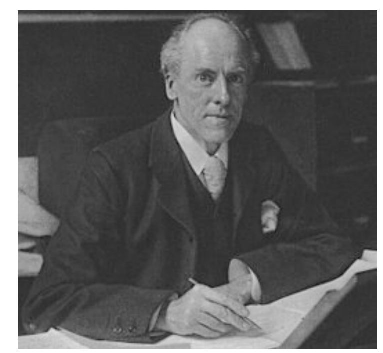


✦ A general optimization problem

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

✦ Principal components analysis



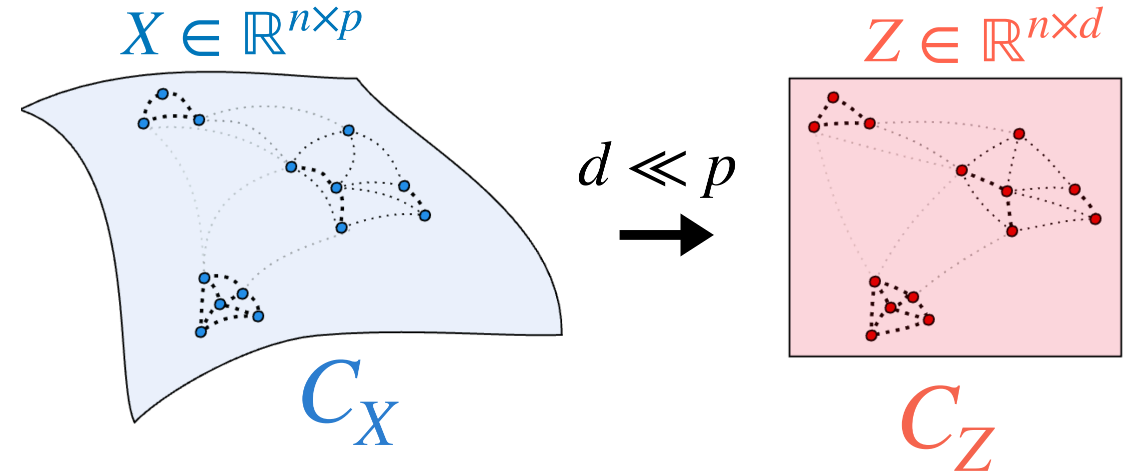


(Pearson, 1901)

Minimizing the reconstruction error

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

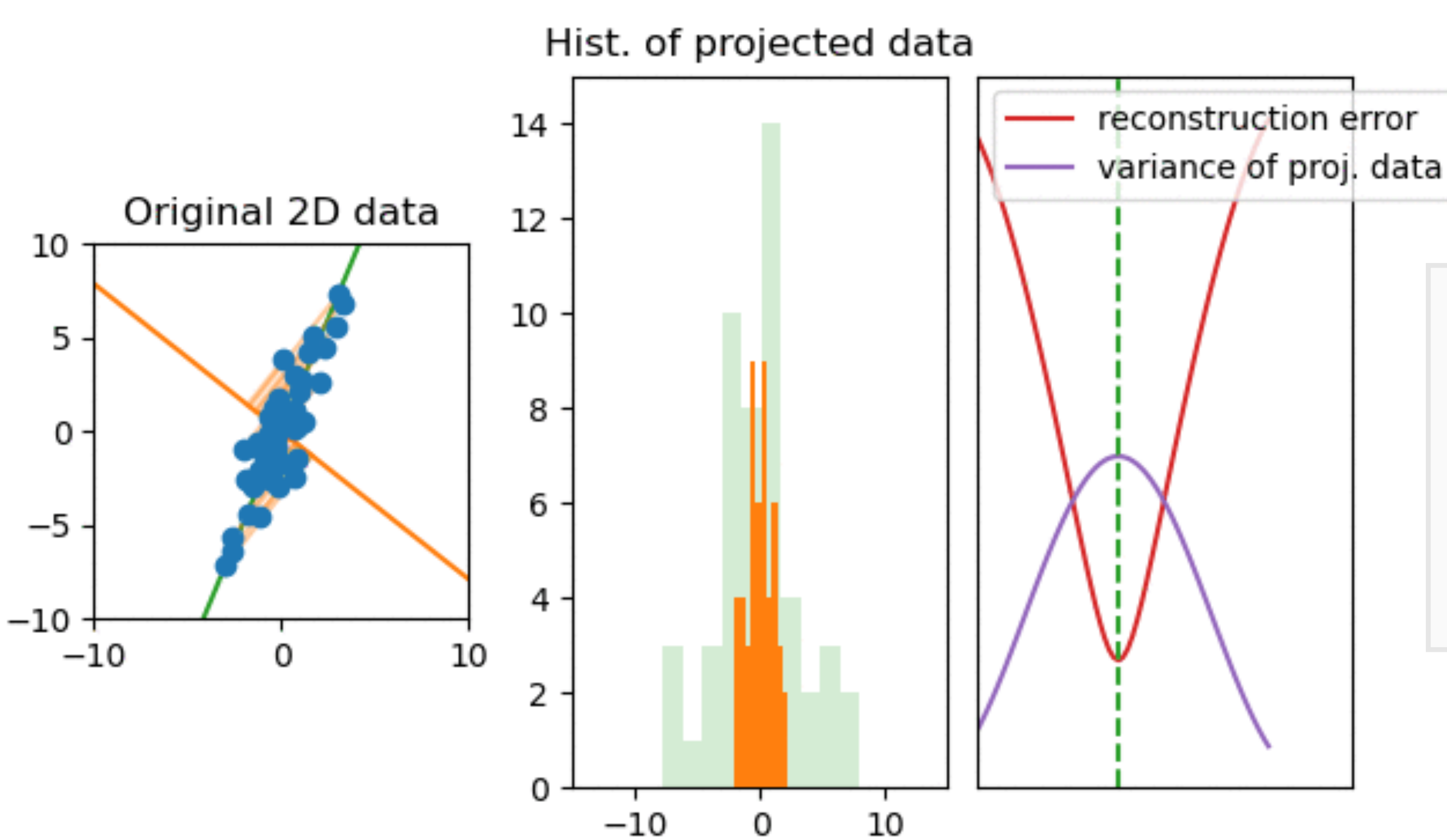
Dimension reduction

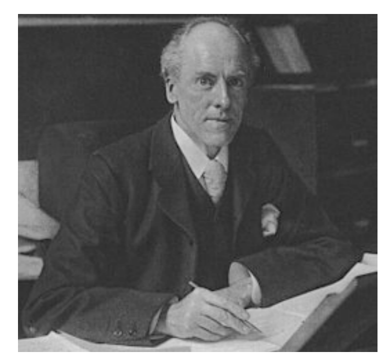


✦ A general optimization problem

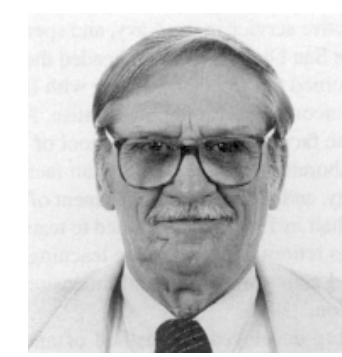
$$\min_{Z \in \mathbb{R}^{n \times d}} \sum_{i,j=1}^{n} L\left([C_X]_{ij}\,,\,[C_Z]_{ij}\right) \quad \text{for some loss} \quad L : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$$

✦ Principal components analysis





(Pearson, 1901)



(Torgerson, 1958)

Preserving the inner products

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

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

Dimension reduction

✦ Spectral methods

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

Dimension reduction

✦ **Spectral methods**

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

$C_X \succeq 0$

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 $\mathbf{v}_i$

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** $C_X \succeq 0$

(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** $C_X \succeq 0$

(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)

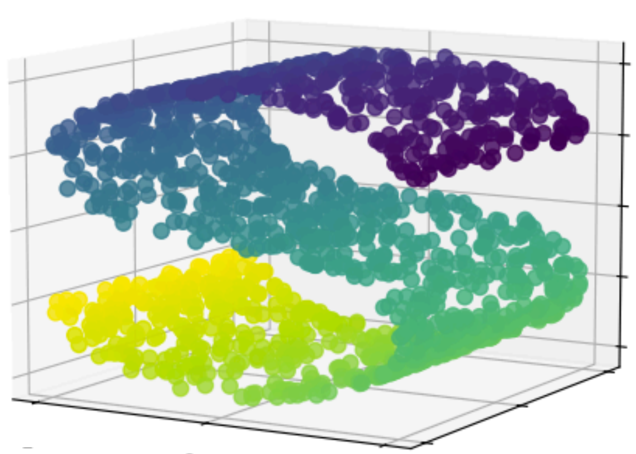
$$Z^\star = (\sqrt{\lambda_1}\mathbf{v}_1, \cdots, \sqrt{\lambda_d}\mathbf{v}_d)^\top$$

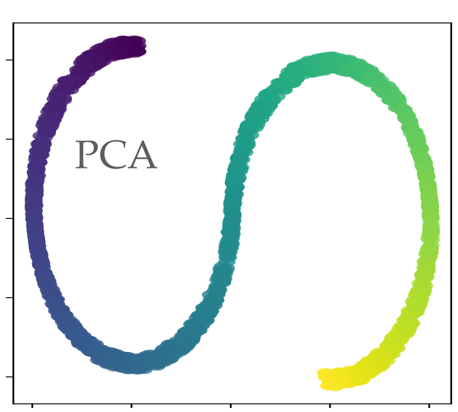
λ_i i-th **largest** eigenvalue of C_X with eigenvector $\mathbf{v}_i$

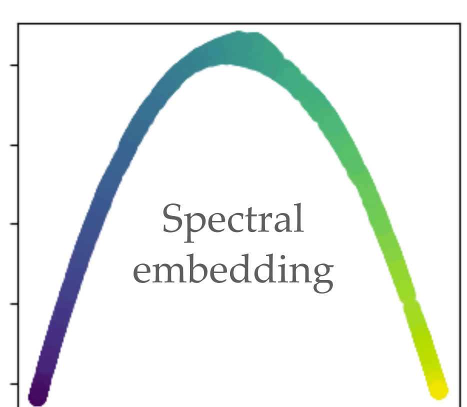
✦ **Kernel PCA** $C_X \succeq 0$

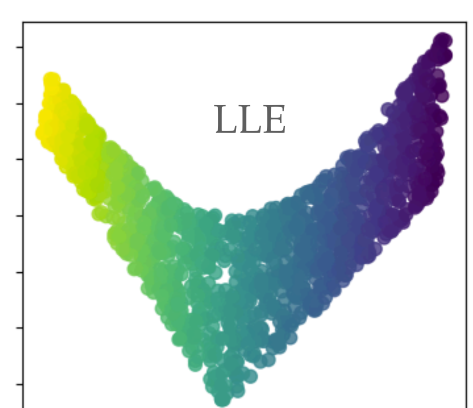
(Schölkopf, 1997)

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

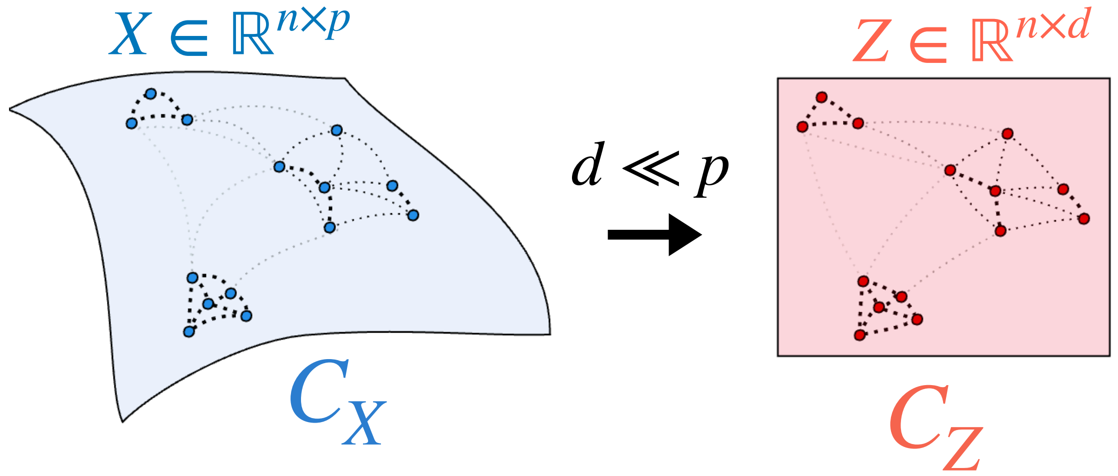








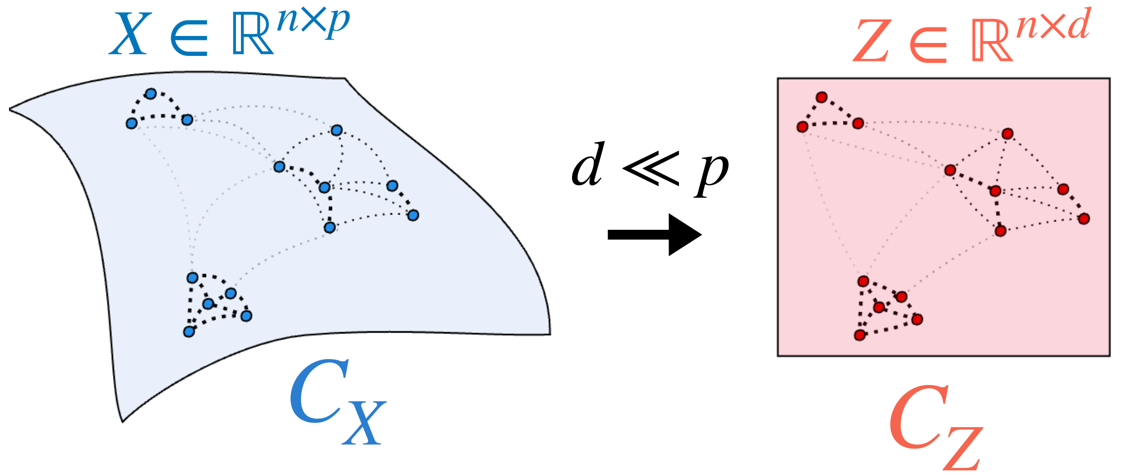
Dimension reduction



✦ **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



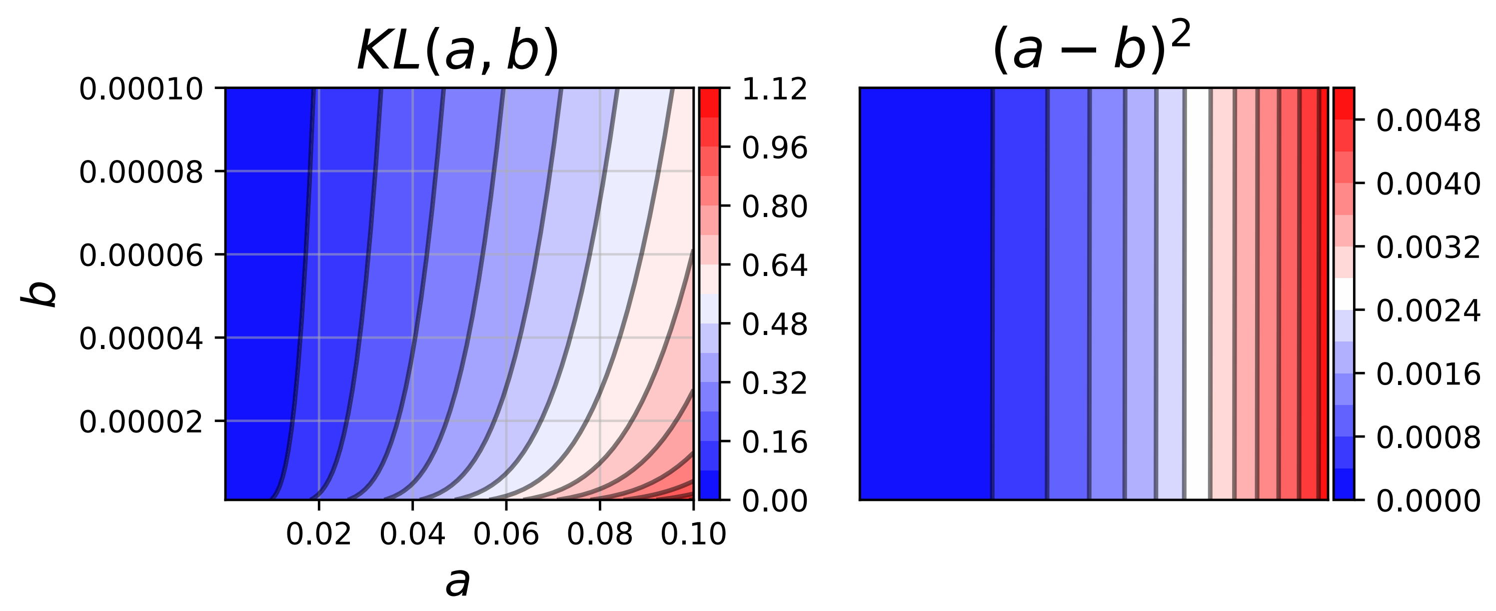
✦ Neighbor embedding methods

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

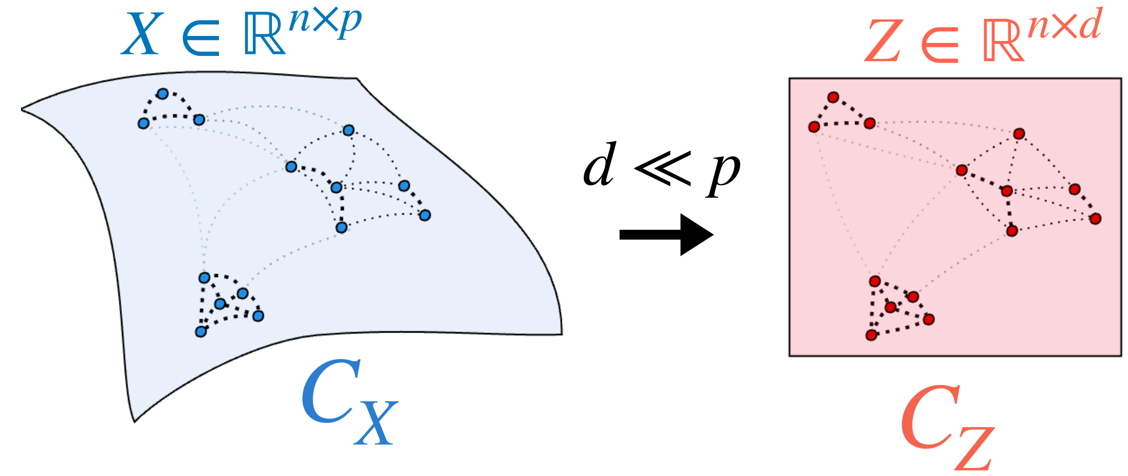
✦ Kullback-Leiber divergence

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

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



Dimension reduction



✦ Neighbor embedding methods

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

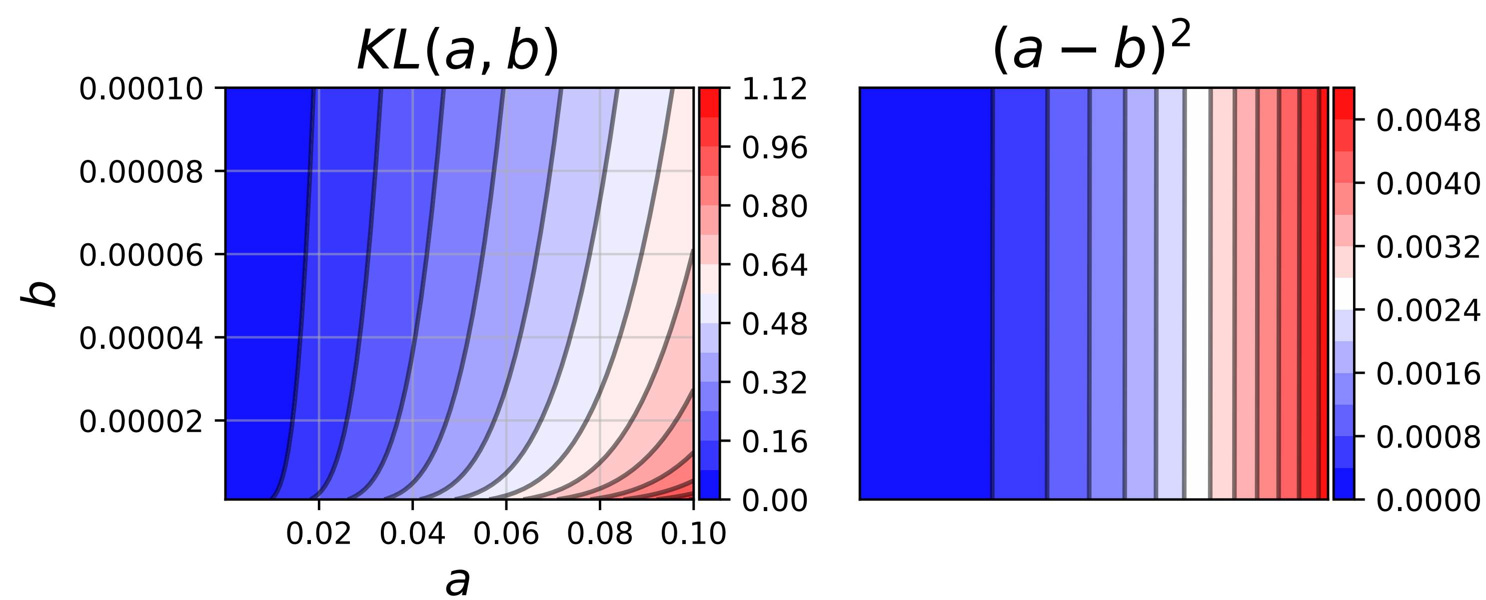
When $\sum_{i,j}[C_X]_{ij} = \sum_{i,j}[C_Z]_{ij}$ (same mass)

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

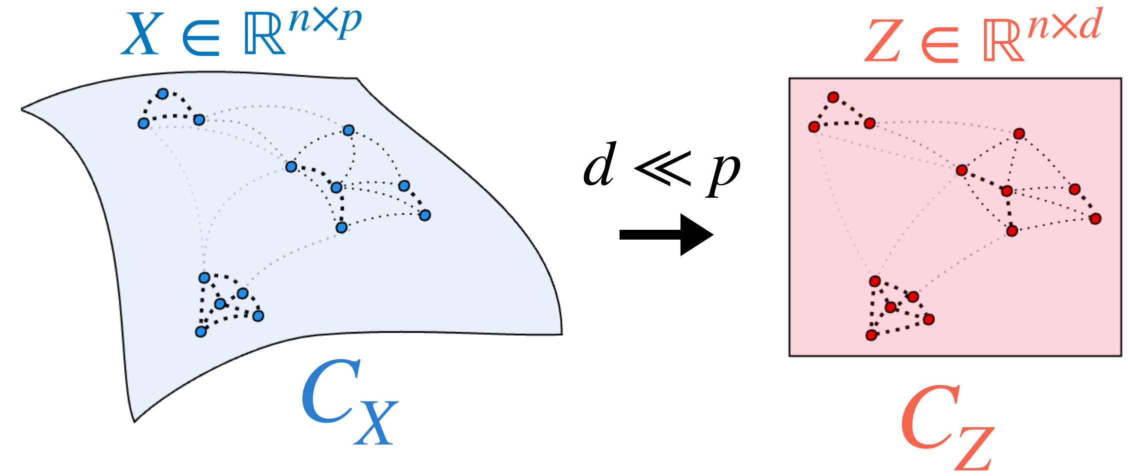
✦ Kullback-Leiber divergence

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

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



Dimension reduction



✦ Neighbor embedding methods

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

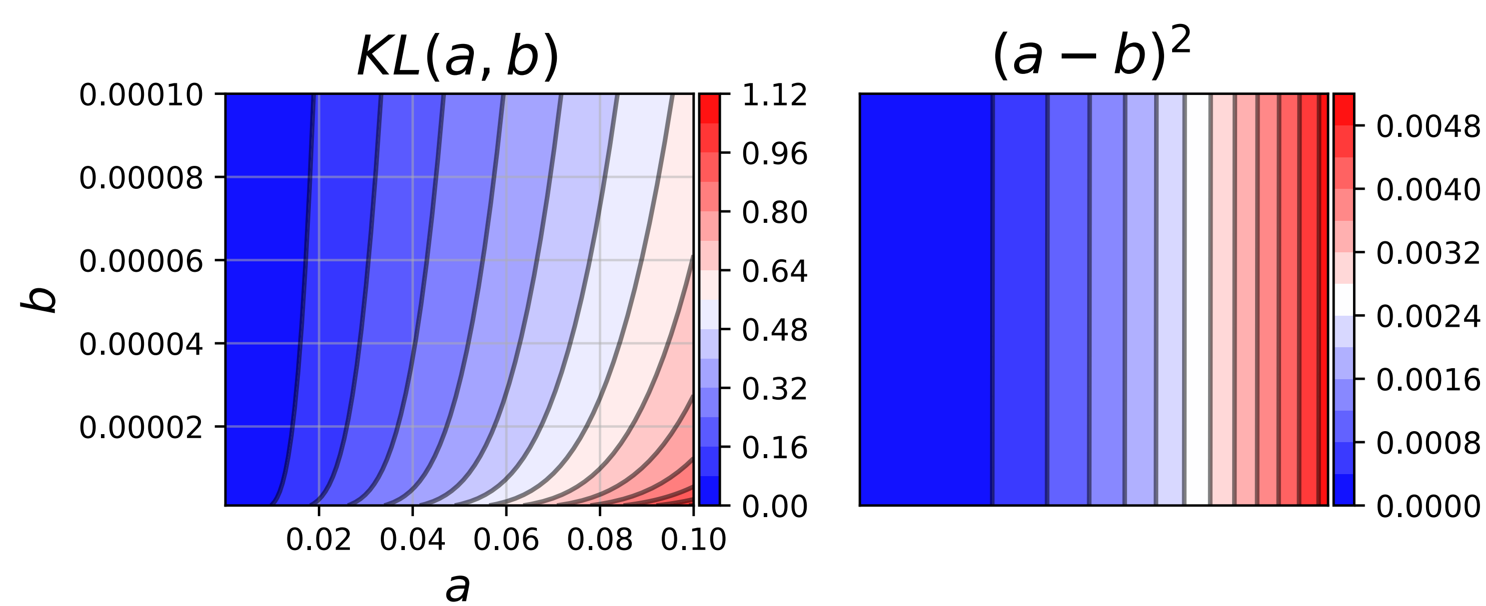
When $\sum_{i,j}[C_X]_{ij} = \sum_{i,j}[C_Z]_{ij}$ (same mass)

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

✦ Kullback-Leiber divergence

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

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

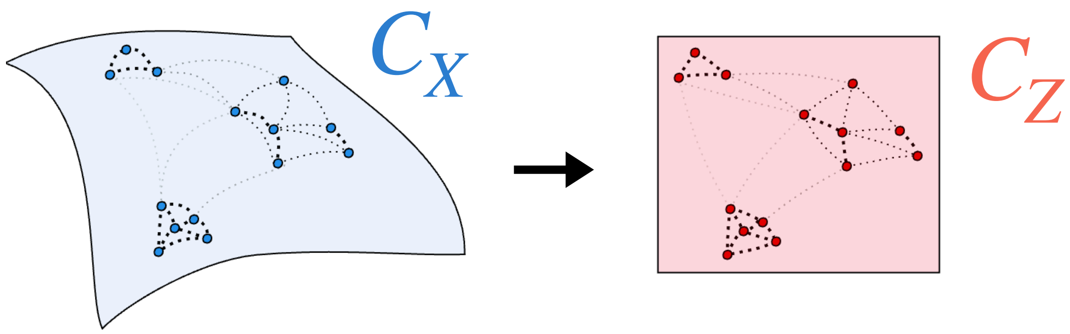


✦ What choices for C_X, C_Z?

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

Dimension reduction

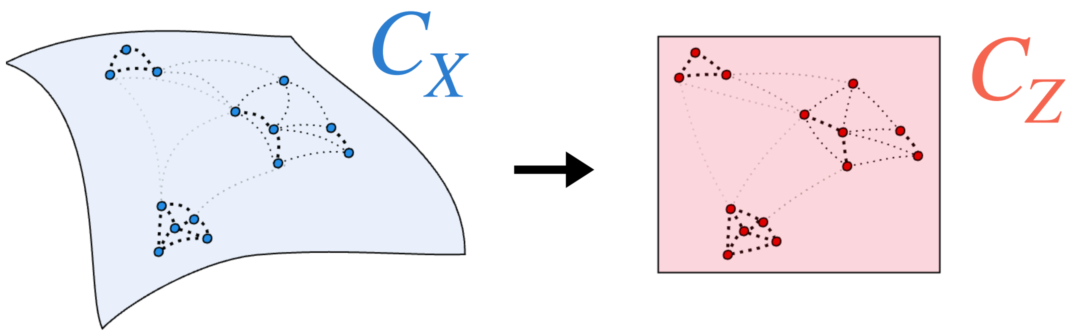
✦ **SNE** (Hinton & Roweis, 2002)



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)

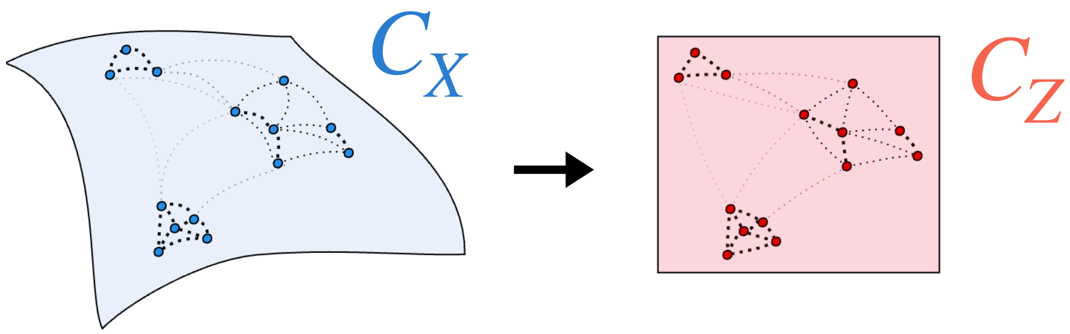
Input space

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

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

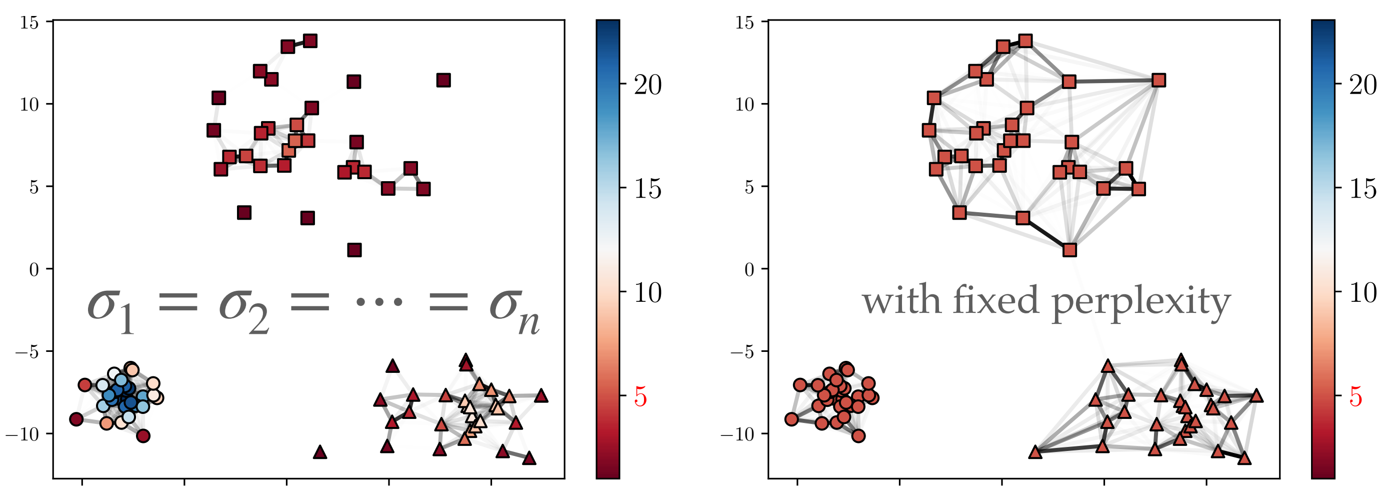
$$[C_X]_{ij} = \frac{\exp(-\|\mathbf{x}_i - \mathbf{x}_j\|_2^2 / 2\sigma_i^2)}{\sum_k \exp(-\|\mathbf{x}_i - \mathbf{x}_k\|_2^2 / 2\sigma_i^2)}$$

- Local bandwidths **optimized** s.t.

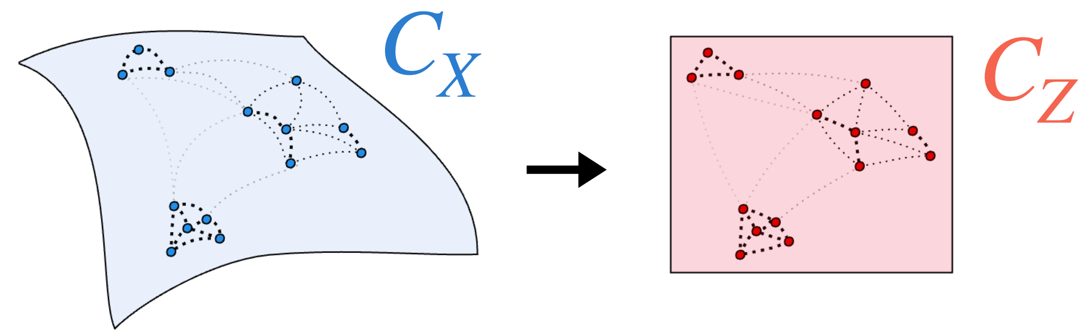
  $\forall i$, entropy($[C_X]_{i,:}$) = log(perplexity)
- Perplexity = effective number of **neighbors**
- Account for **varying density**

Embedding space

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



Dimension reduction



✦ **SNE** (Hinton & Roweis, 2002)

Input space

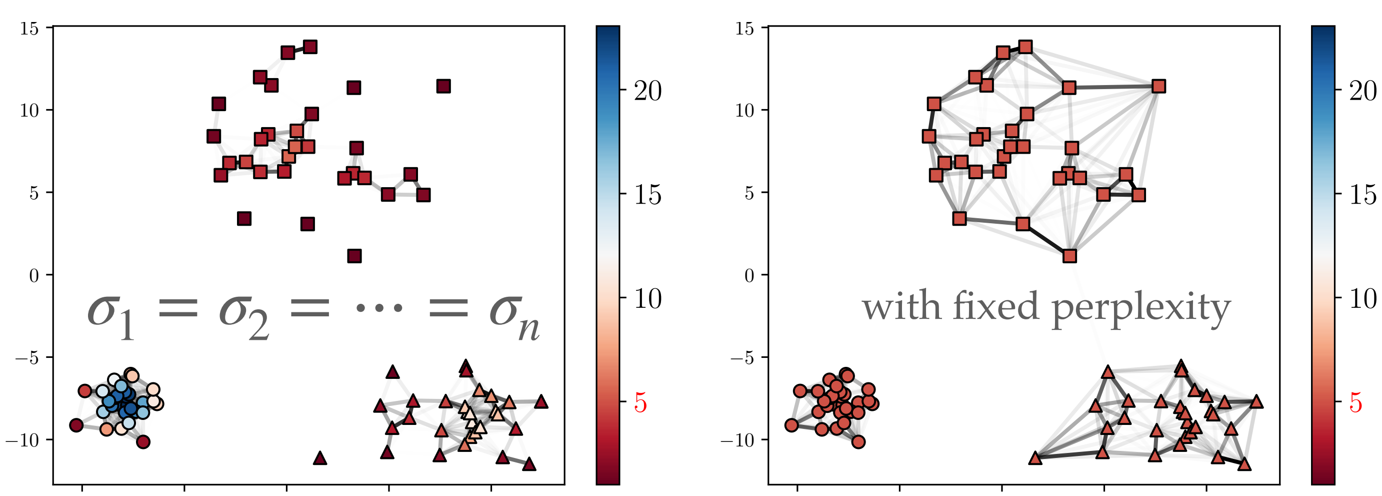
$$[C_X]_{ij} = \frac{\exp(-\|\mathbf{x}_i - \mathbf{x}_j\|_2^2 / 2\sigma_i^2)}{\sum_k \exp(-\|\mathbf{x}_i - \mathbf{x}_k\|_2^2 / 2\sigma_i^2)}$$

- Local bandwidths **optimized** s.t.

  $\forall i, \text{entropy}(\ [C_X]_{i,:}\ ) = \log(\text{perplexity})$
- 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)

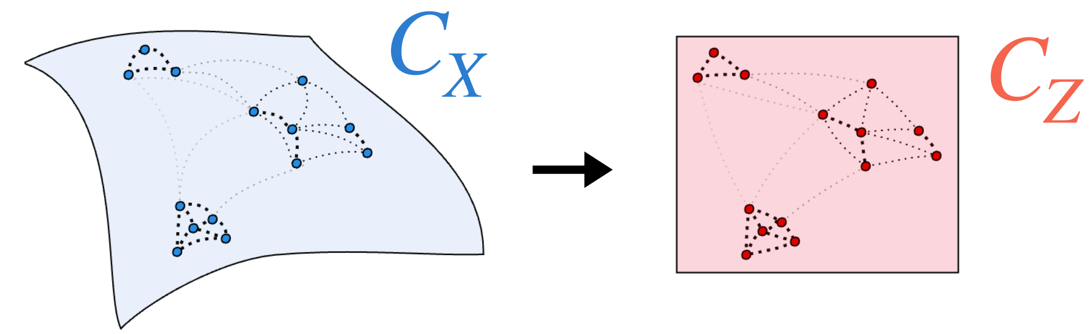
- Joint distributions:

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

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

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

Dimension reduction



✦ SNE (Hinton & Roweis, 2002)

Input space

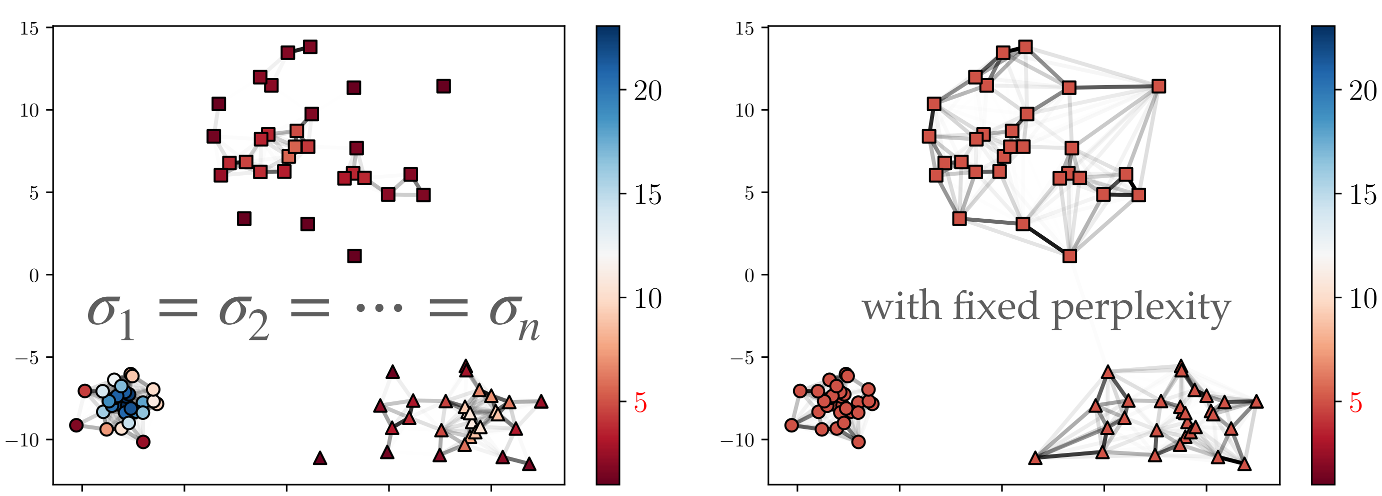
$$[C_X]_{ij} = \frac{\exp(-\|\mathbf{x}_i - \mathbf{x}_j\|_2^2 / 2\sigma_i^2)}{\sum_k \exp(-\|\mathbf{x}_i - \mathbf{x}_k\|_2^2 / 2\sigma_i^2)}$$

- Local bandwidths **optimized** s.t.

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

Embedding space

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



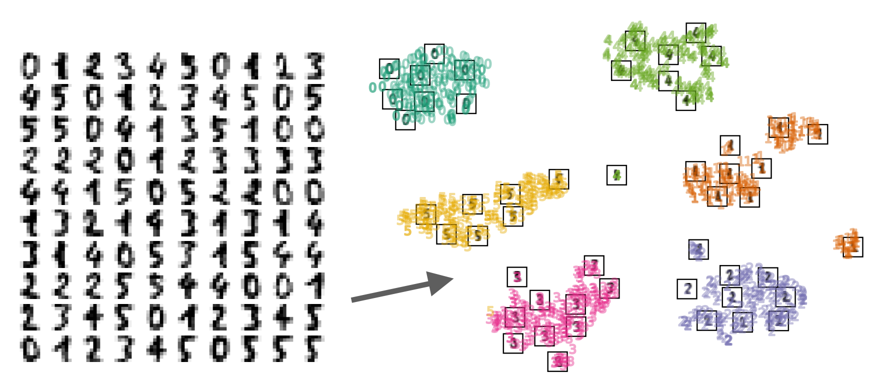
✦ (t)-SNE (Van der Maaten & Hinton, 2008)

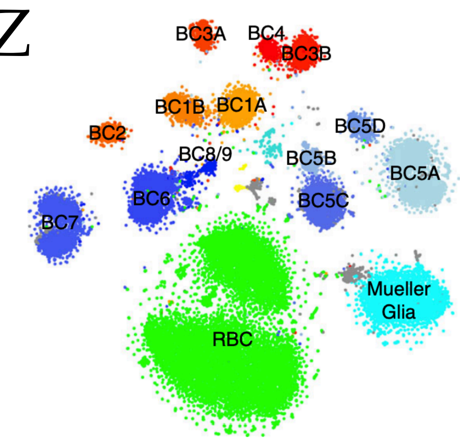
- Joint distributions:

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

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

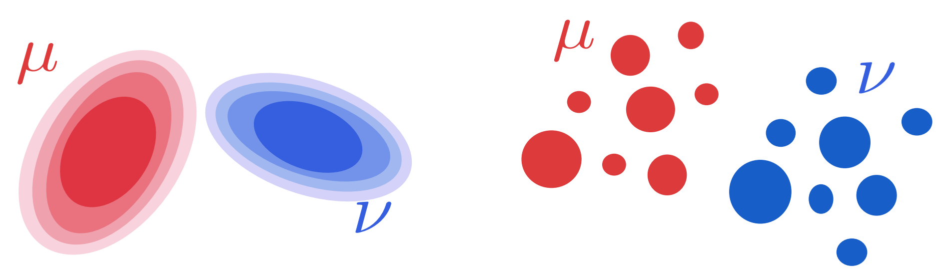
- Crowding effect: Student t-distribution instead of Gaussian in Z

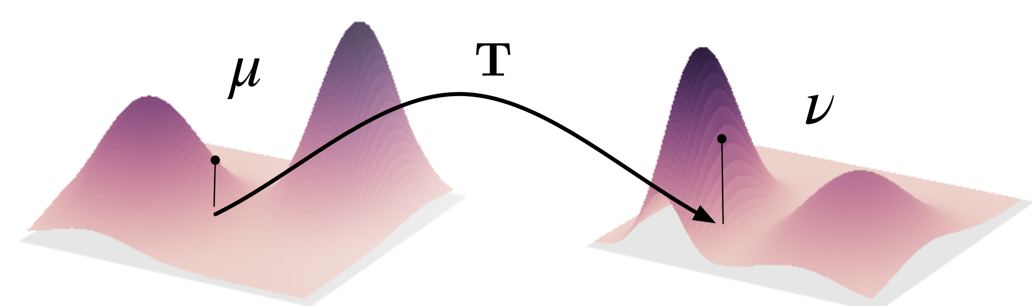




(Shekhar et al., 2016)

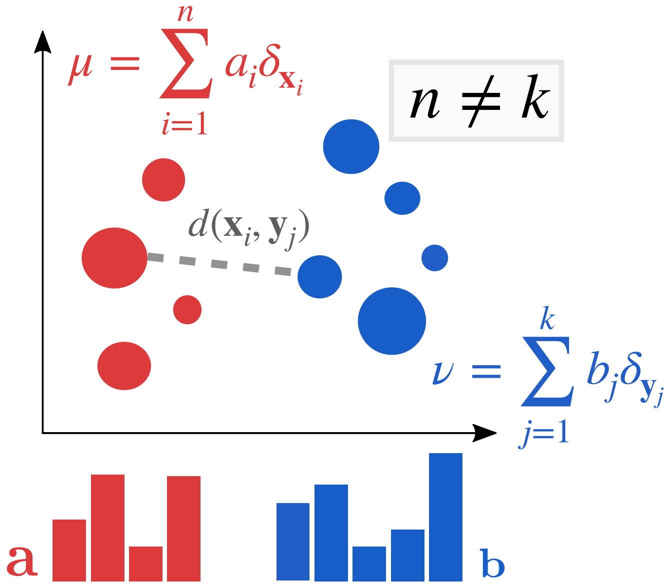
From linear Optimal Transport to Gromov-Wasserstein





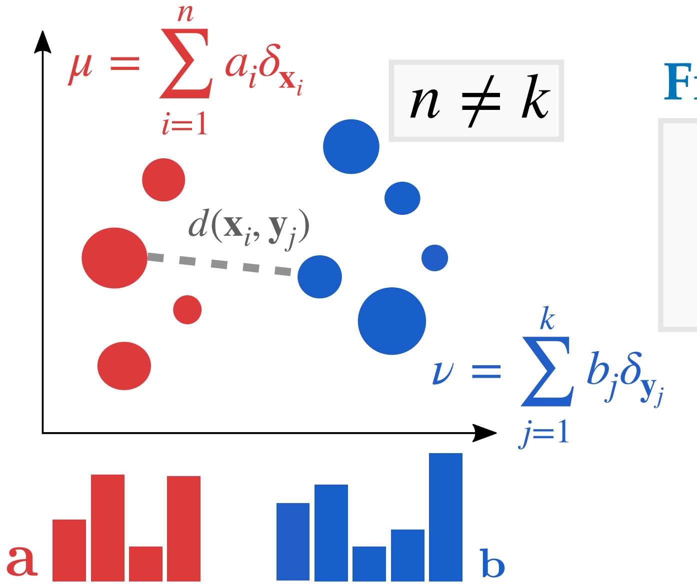
From Wasserstein to Gromov-Wasserstein

✦ Classical optimal transport (in a nutshell)



From Wasserstein to Gromov-Wasserstein

✦ Classical optimal transport (in a nutshell)

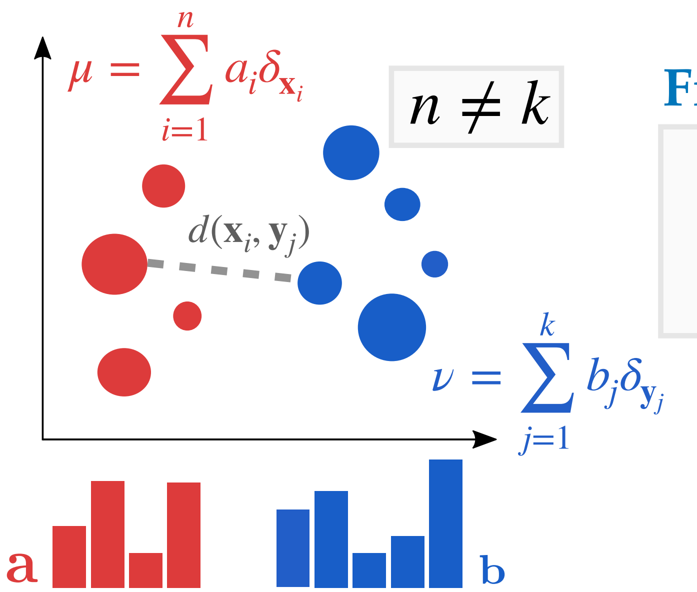


Find plan $T \in \mathbb{R}_+^{n \times k}$

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

From Wasserstein to Gromov-Wasserstein

✦ Classical optimal transport (in a nutshell)

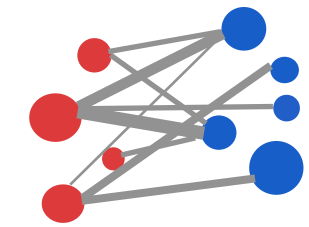


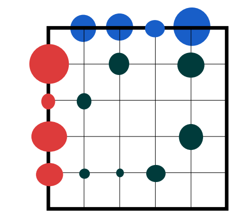
Find plan $T \in \mathbb{R}_+^{n \times k}$

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

which constraints ? → **Coupling** $\Pi(a, b)$

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

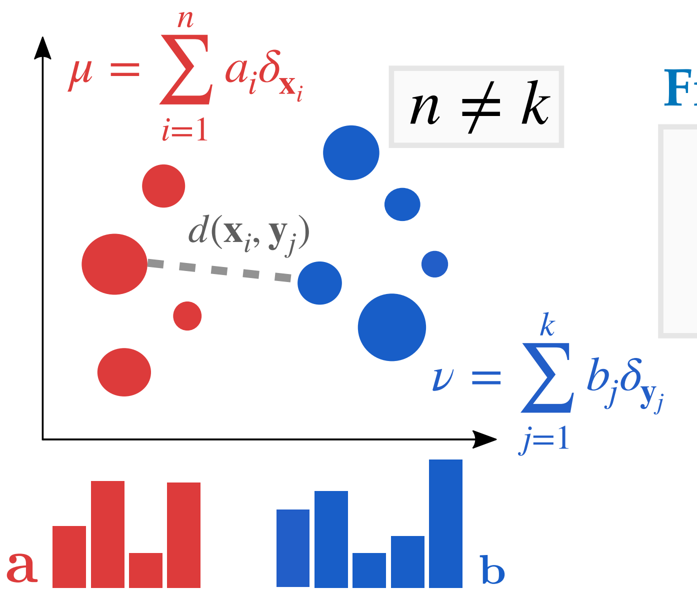




$T \in [0,1]^{n \times k}$

From Wasserstein to Gromov-Wasserstein

✦ Classical optimal transport (in a nutshell)



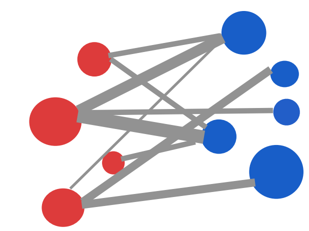
Find plan $T \in \mathbb{R}_+^{n \times k}$

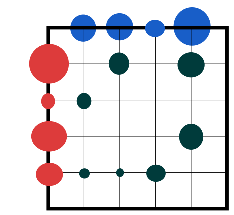
$$\min_T \sum_{ij} d(\mathbf{x}_i, \mathbf{y}_j) T_{ij}$$

which constraints ?

→ **Coupling** $\Pi(a, b)$

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





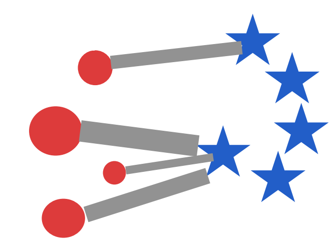
$T \in [0,1]^{n \times k}$

→ **Semi-relaxed coupling**

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

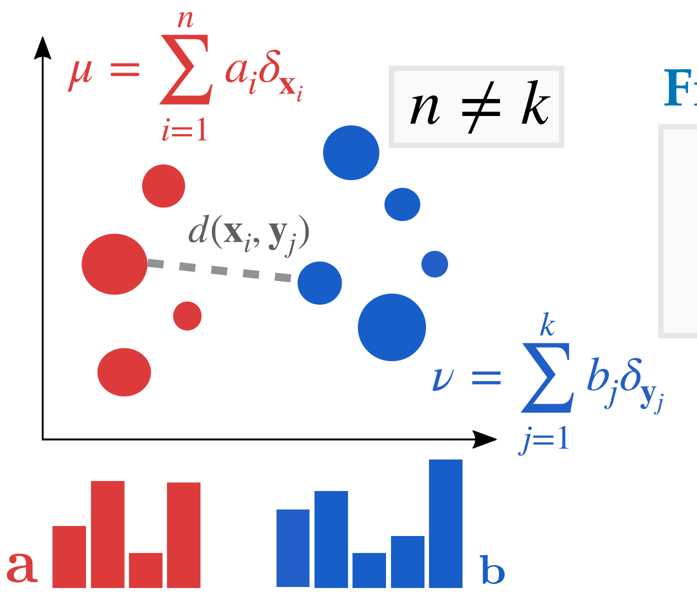
$\sim$ `assign` in k-means

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



From Wasserstein to Gromov-Wasserstein

✦ Classical optimal transport (in a nutshell)



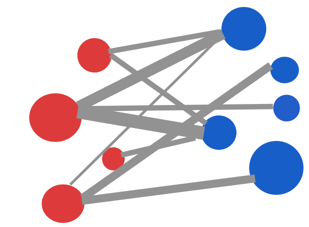
Find plan $T \in \mathbb{R}_+^{n \times k}$

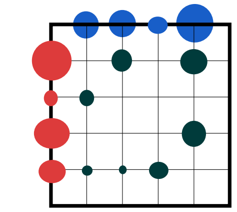
$$\min_T \sum_{ij} d(\mathbf{x}_i, \mathbf{y}_j) T_{ij}$$

which constraints ?

Coupling $\Pi(a, b)$

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





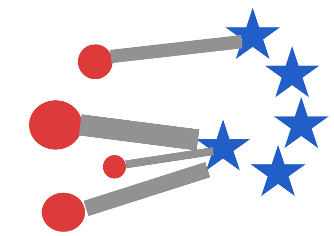
$T \in [0,1]^{n \times k}$

Semi-relaxed coupling

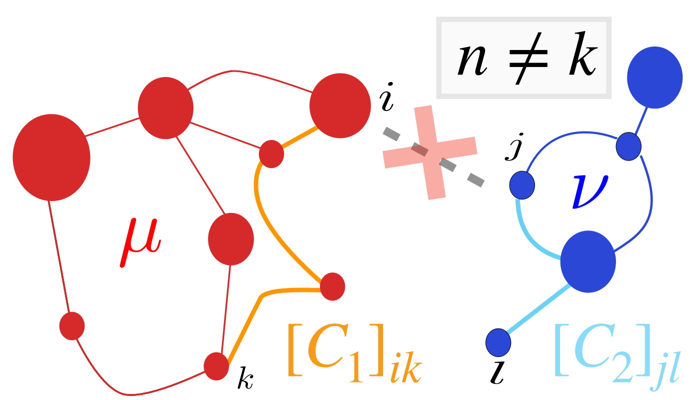
$$T^\top 1_n = a$$

$\sim$ `assign` in k-means

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



✦ Gromov-Wasserstein

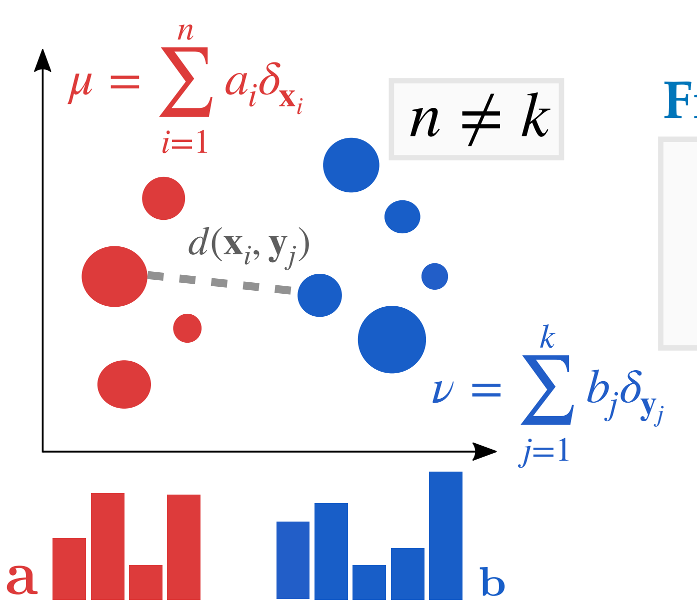


(Sturm, 2012)

(Mémoli, 2011)

From Wasserstein to Gromov-Wasserstein

✦ Classical optimal transport (in a nutshell)



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

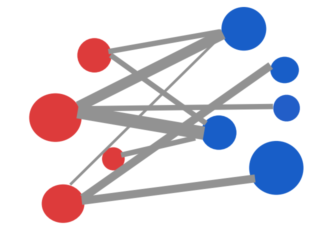
$$\min_T \sum_{ij} d(\mathbf{x}_i, \mathbf{y}_j) T_{ij}$$

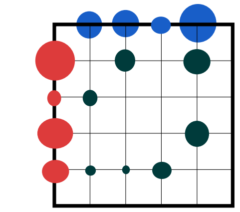
which constraints ?

Coupling $\Pi(a, b)$

$T^\top 1_n = a$

$T 1_k = b$





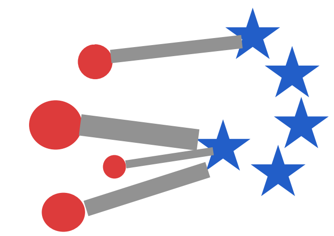
$T \in [0,1]^{n \times k}$

Semi-relaxed coupling

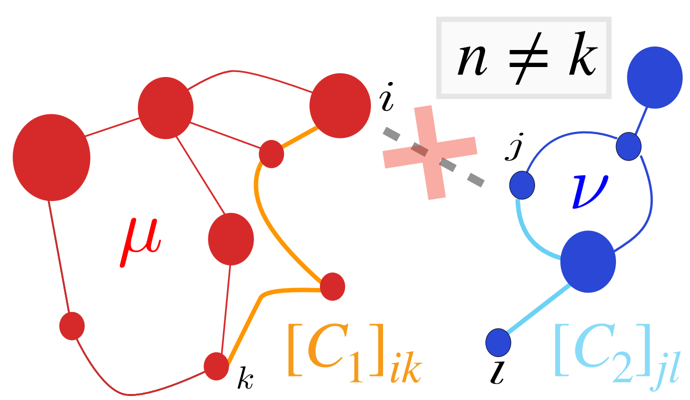
$T^\top 1_n = a$

$\sim$ `assign` in k-means

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



✦ Gromov-Wasserstein



(Sturm, 2012)

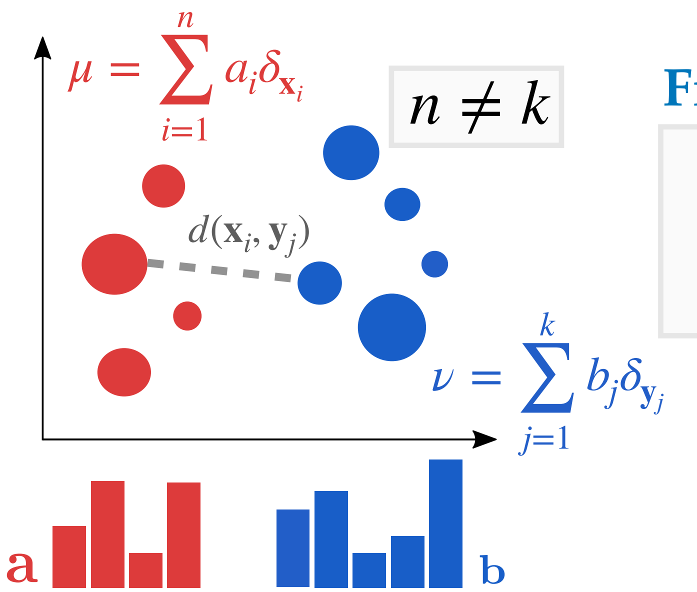
(Mémoli, 2011)

Quadratic OT: find the plan

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

From Wasserstein to Gromov-Wasserstein

✦ Classical optimal transport (in a nutshell)



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

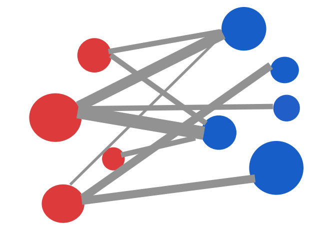
$$\min_T \sum_{ij} d(\mathbf{x}_i, \mathbf{y}_j) T_{ij}$$

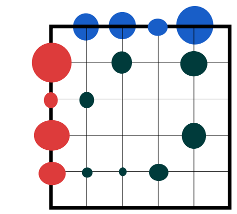
which constraints ?

Coupling $\Pi(a, b)$

$T^\top 1_n = a$

$T 1_k = b$



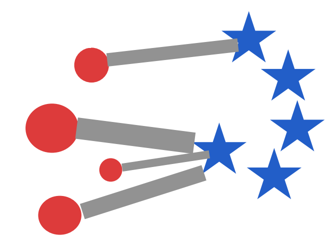


Semi-relaxed coupling

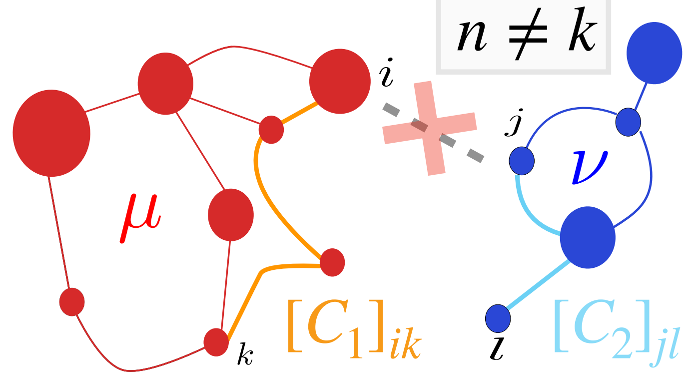
$T^\top 1_n = a$

$\sim$ `assign` in k-means

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



✦ Gromov-Wasserstein



(Sturm, 2012) (Mémoli, 2011)

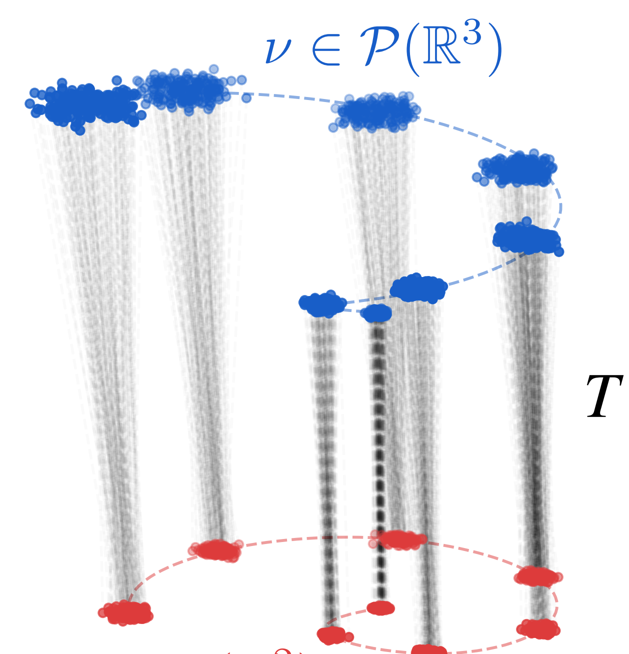
Quadratic OT: find the plan

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

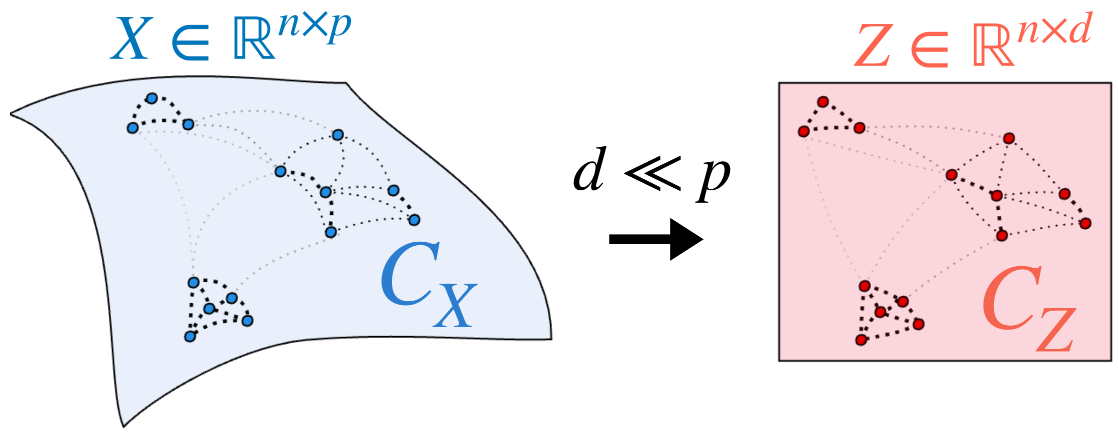
- ✦ L measures distortion

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

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



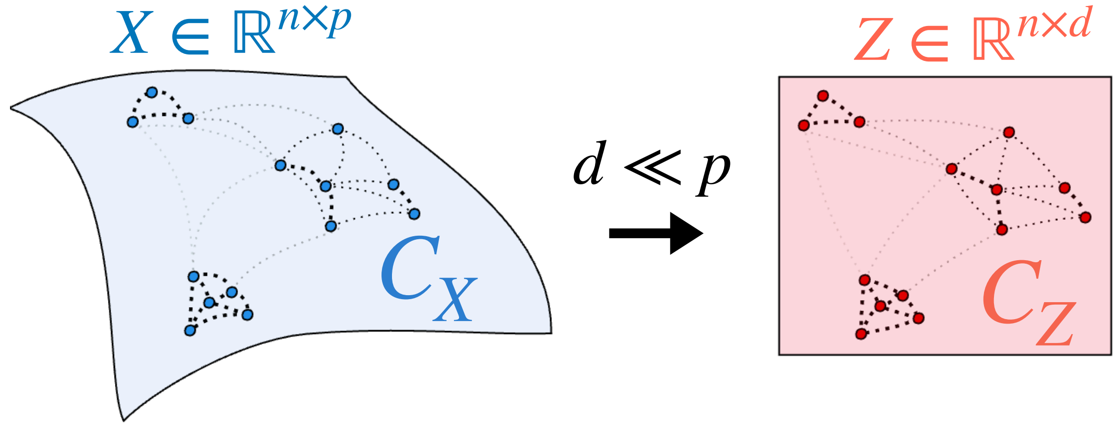
Distributional Reduction



we'll see

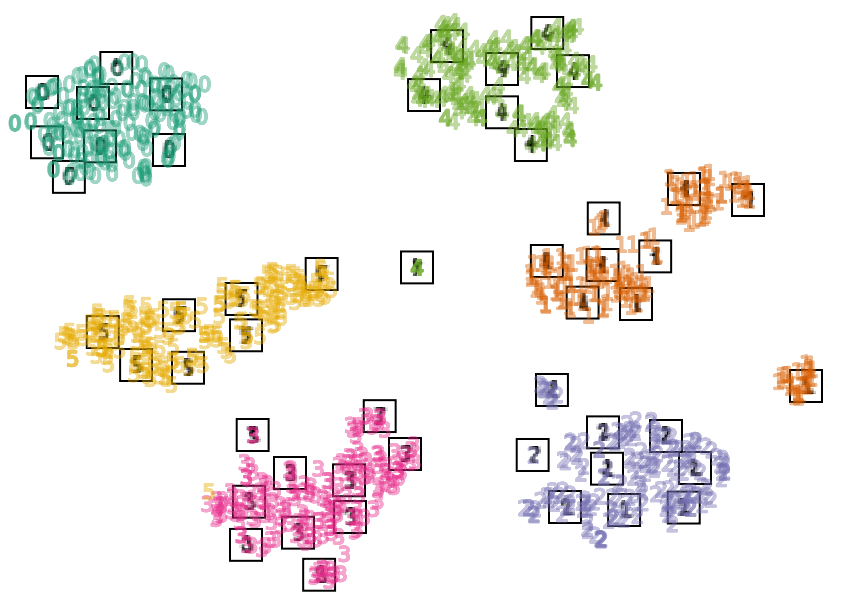
$$\sim \min_{Z \in \mathbb{R}^{n \times d}} \mathrm{GW}_L\left(C_X, C_Z, \frac{1_n}{n}, \frac{1_n}{n}\right)$$

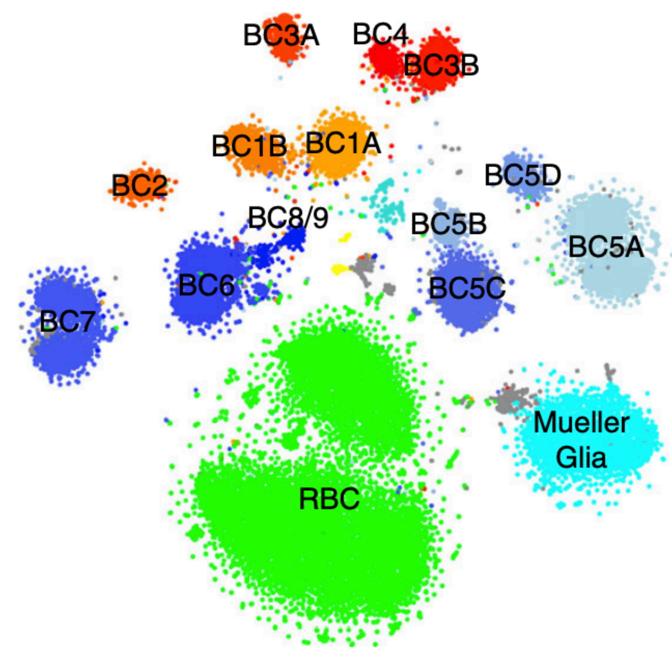
Distributional Reduction



we'll see $\sim \min_{Z \in \mathbb{R}^{n \times d}} \mathrm{GW}_L(C_X, C_Z, \frac{1_n}{n}, \frac{1_n}{n})$

✦ Motivation



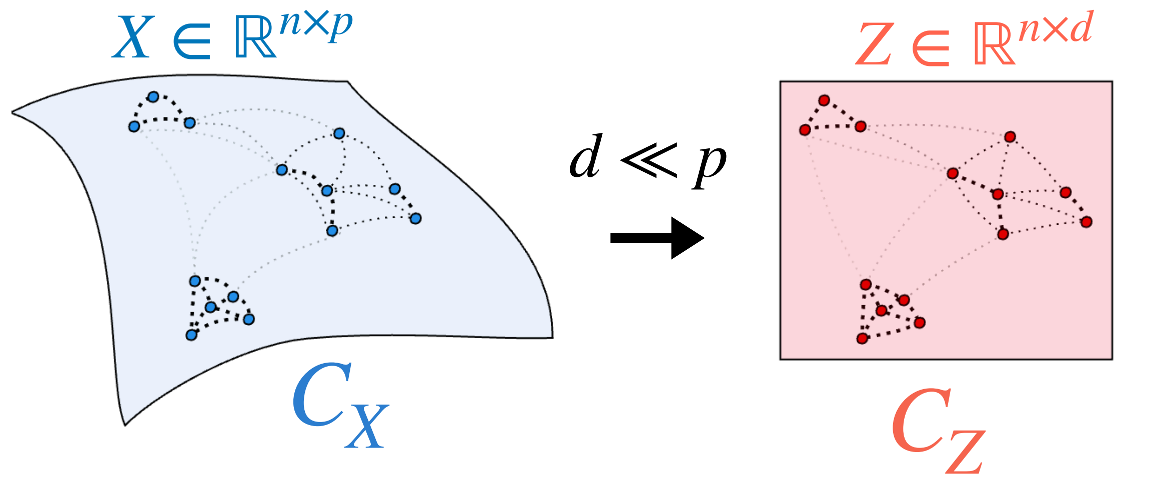


(Shekhar et al., 2016)

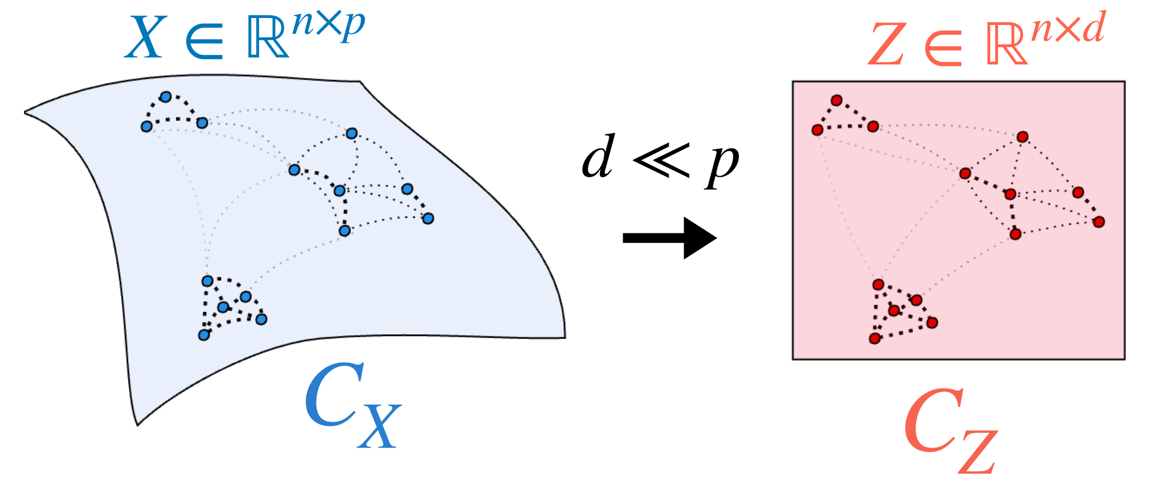
DR as OT in disguise

✦ **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

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

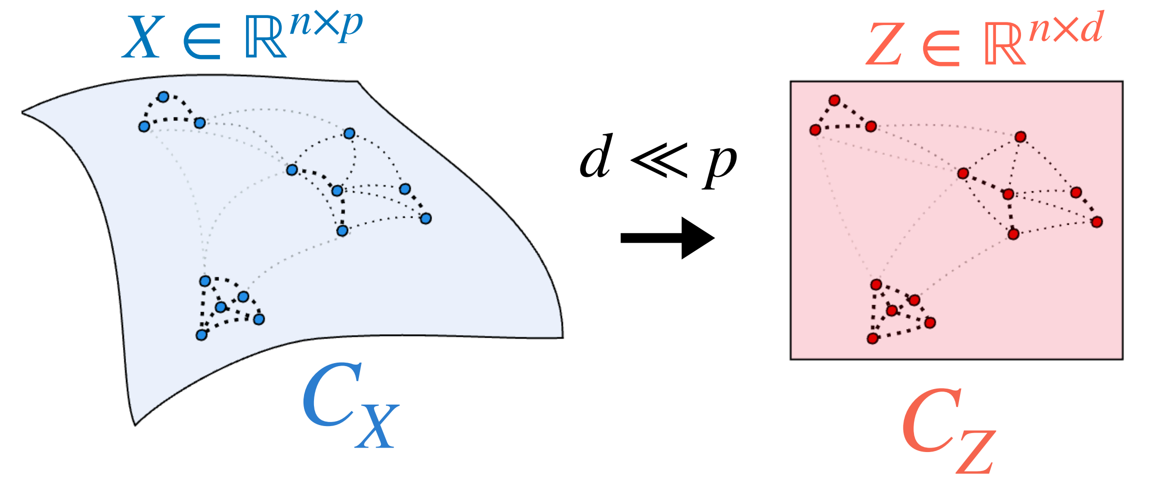
equiv

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

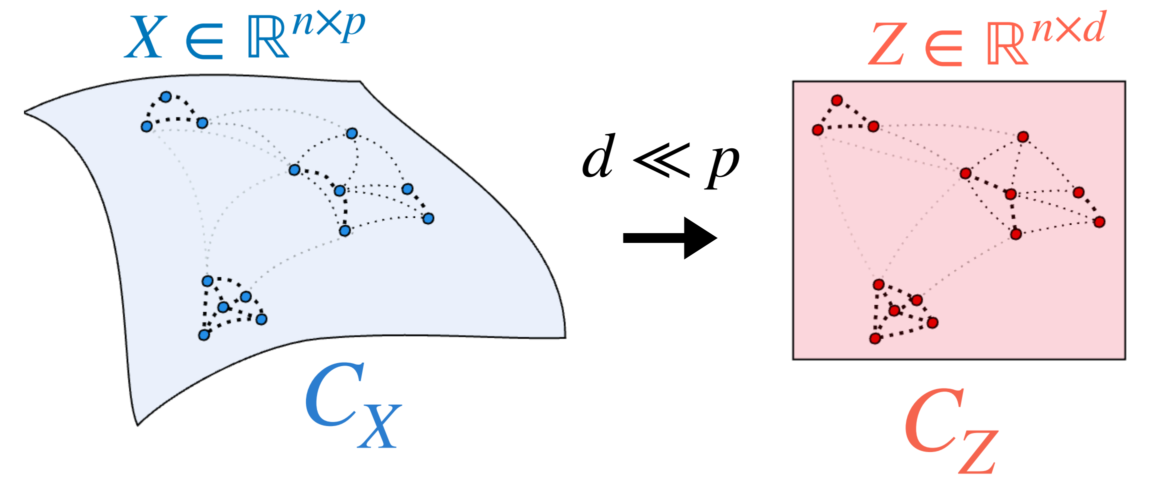
$\forall P, C_{PZ} = PC_ZP^\top$

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

equiv

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

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

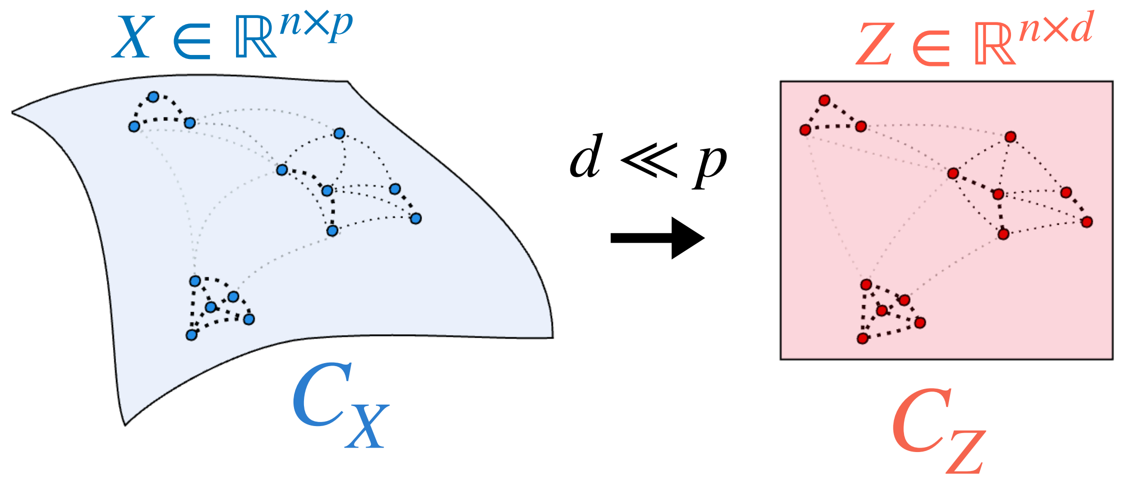
$$\min_{Z \in \mathbb{R}^{n \times d}} \min_{P} \sum_{i,j,k,l=1}^{n} L\Big([C_X]_{ik} \ , \ [C_Z]_{jl} \Big) P_{ij} P_{kl}$$

?

✦ Gromov-Wasserstein projection

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

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-Wasserstein projection

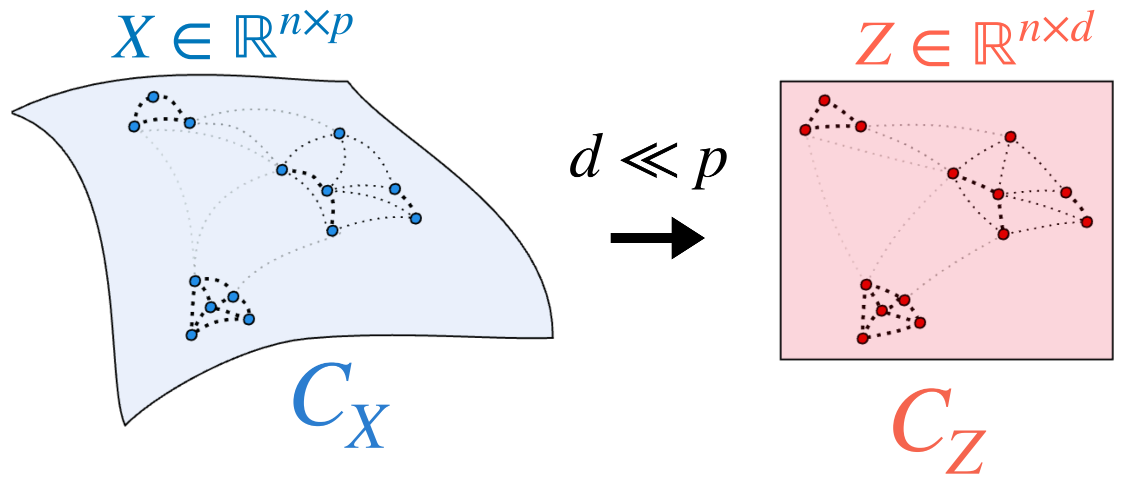
$$\min_{Z \in \mathbb{R}^{n \times d}} \min_{T \in \Pi(\frac{1_n}{n}, \frac{1_n}{n})} \sum_{ijkl} L\left([C_X]_{ik}\ ,\ [C_Z]_{jl}\right) T_{ij}T_{kl}$$

✦ Equivalence holds for

Spectral methods

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

DR as OT in disguise



✦ Dimension reduction

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

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-Wasserstein projection

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

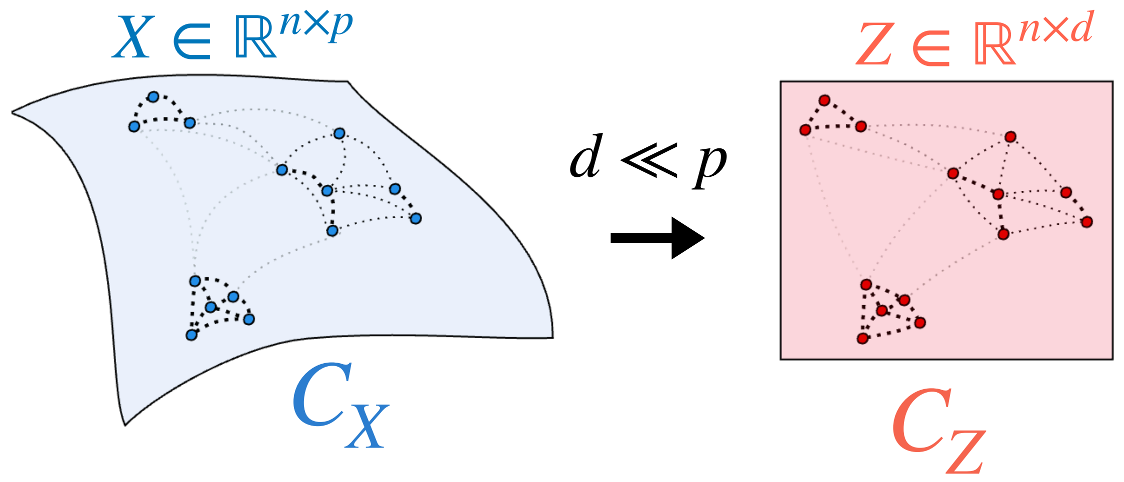
✦ Equivalence holds for

Spectral methods

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

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

DR as OT in disguise



✦ Dimension reduction

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

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

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

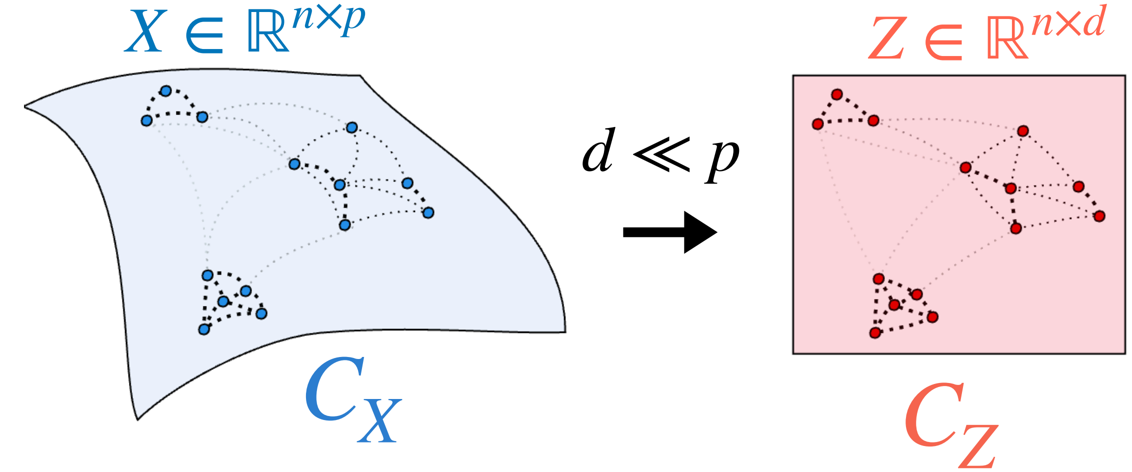
Neighbor embedding methods

- C_X is CPD, $L = KL$

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

  where $\log(K_Z)$ is CPD

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

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

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

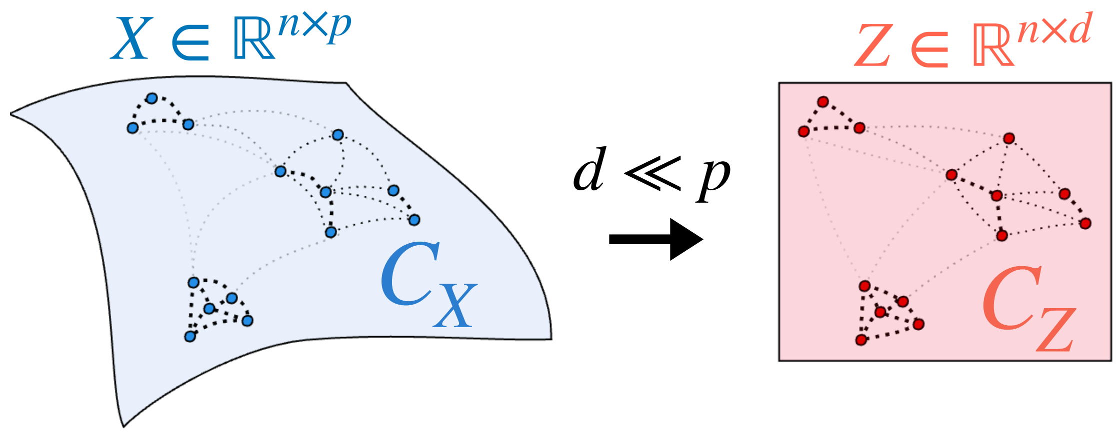
$1_n^\top K_Z 1_n = 1,\ K_Z 1_n = 1_n,\ K_Z^\top 1_n = 1_n + K_Z 1_n = 1_n$

(Sinkhorn & Knopp, 1967)

To improve as C_X generally not CPD

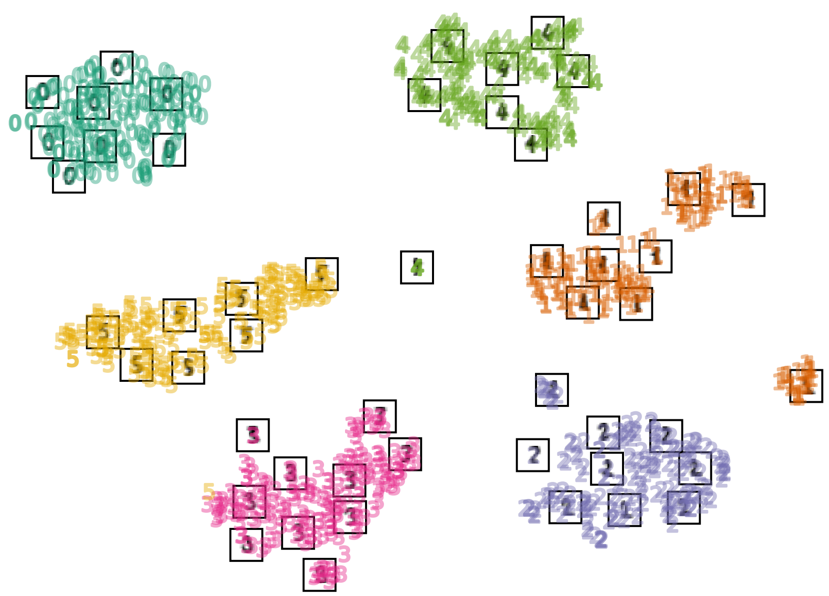
Distributional reduction

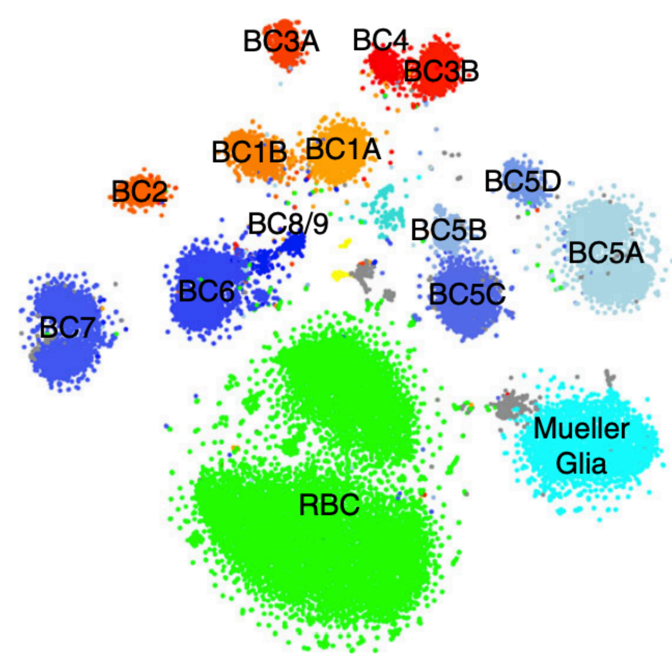
Distributional Reduction



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

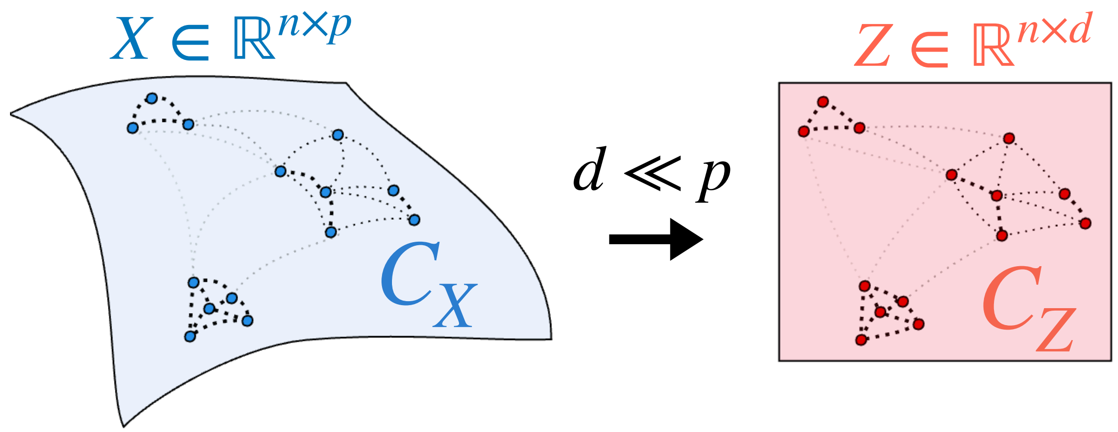
✦ Motivation



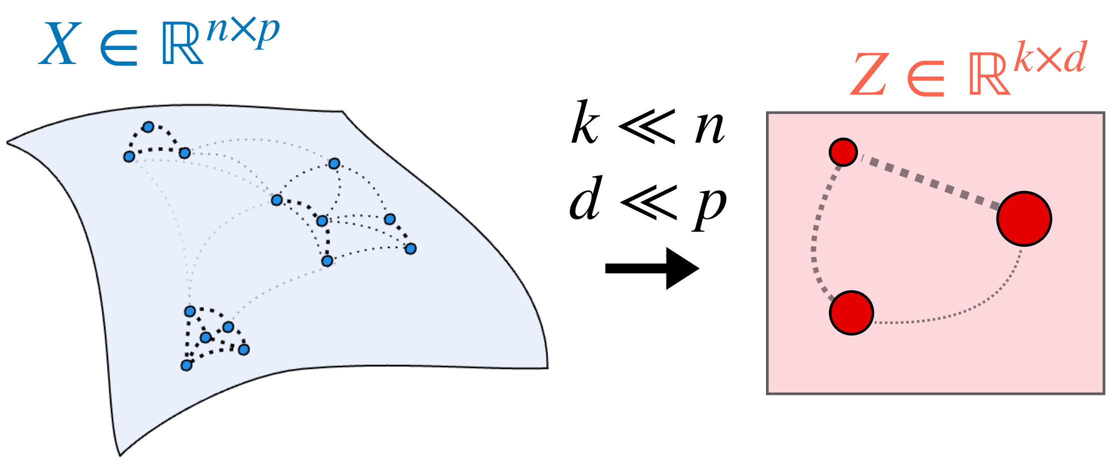


(Shekhar et al., 2016)

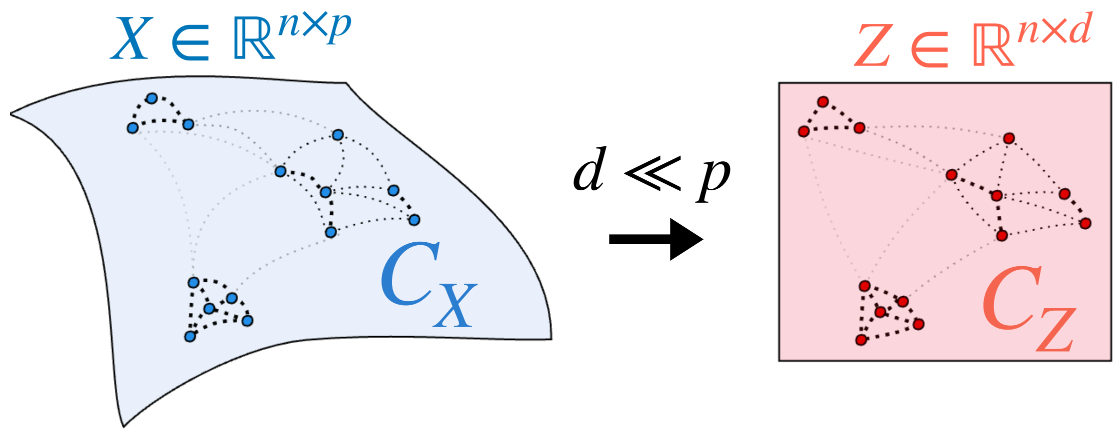
Distributional Reduction



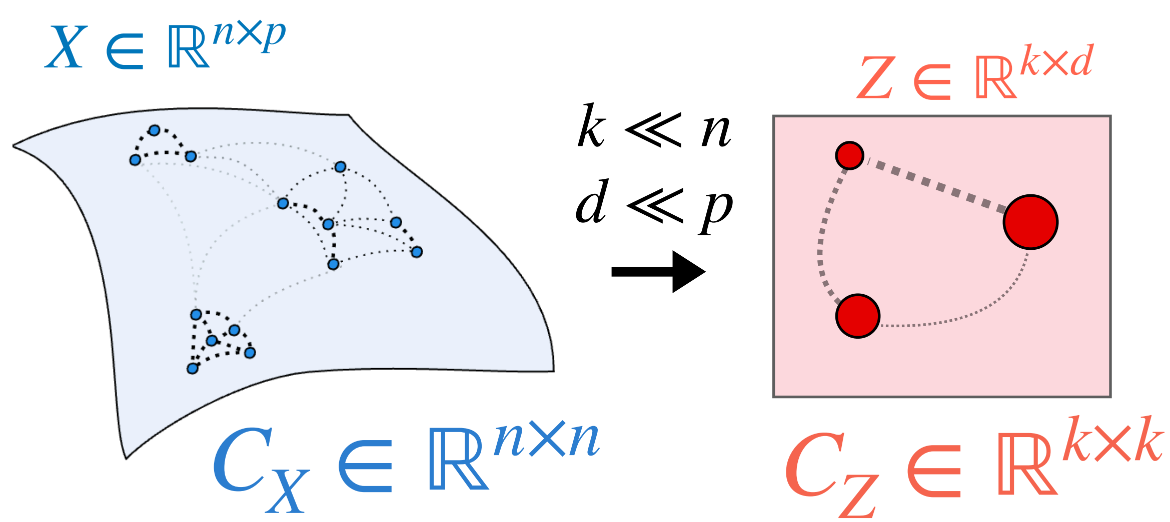
$$\min_{Z \in \mathbb{R}^{n \times d}} \mathrm{GW}_L\left(C_X, C_Z, \frac{1_n}{n}, \frac{1_n}{n}\right)$$



Distributional Reduction

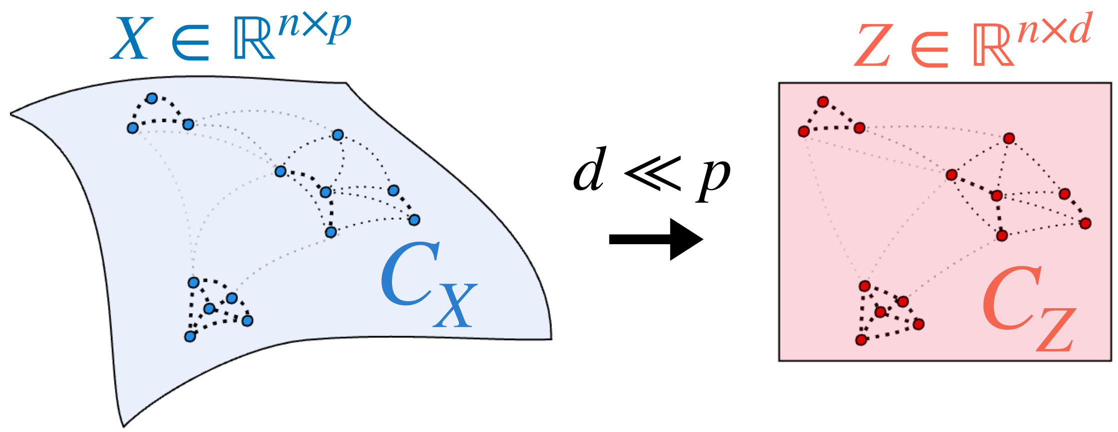


$$\min_{Z \in \mathbb{R}^{n \times d}} \mathrm{GW}_L\left(C_X, C_Z, \frac{1_n}{n}, \frac{1_n}{n}\right)$$

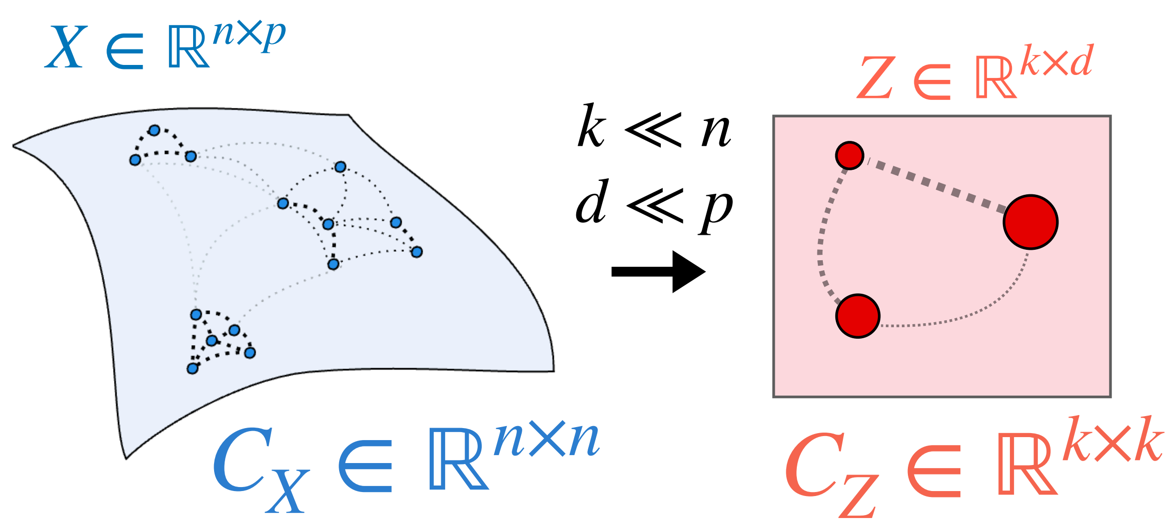


✦ **GW projection** $\min_{\nu \in \mathscr{P}_k(\mathbb{R}^d)} \mathrm{GW}(\hat{\mu}_n, \nu)$

Distributional Reduction



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



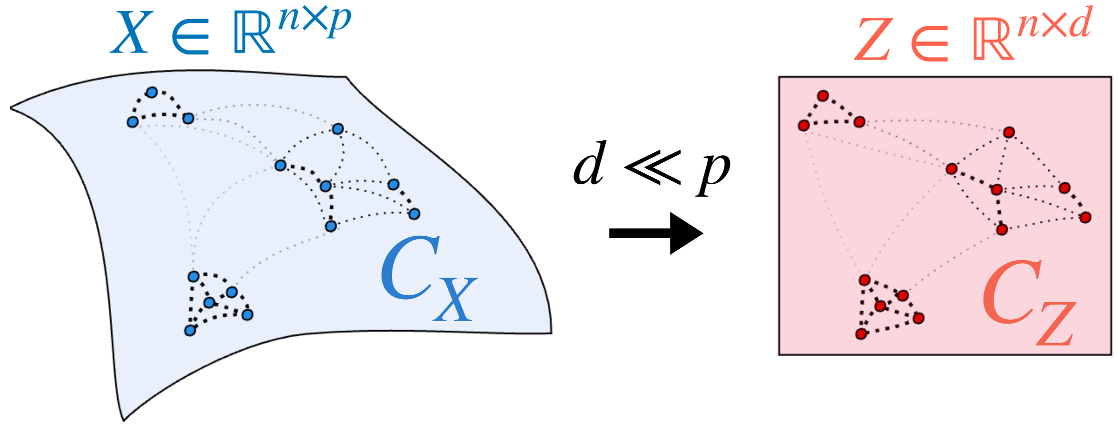
✦ **Optimization problem**

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

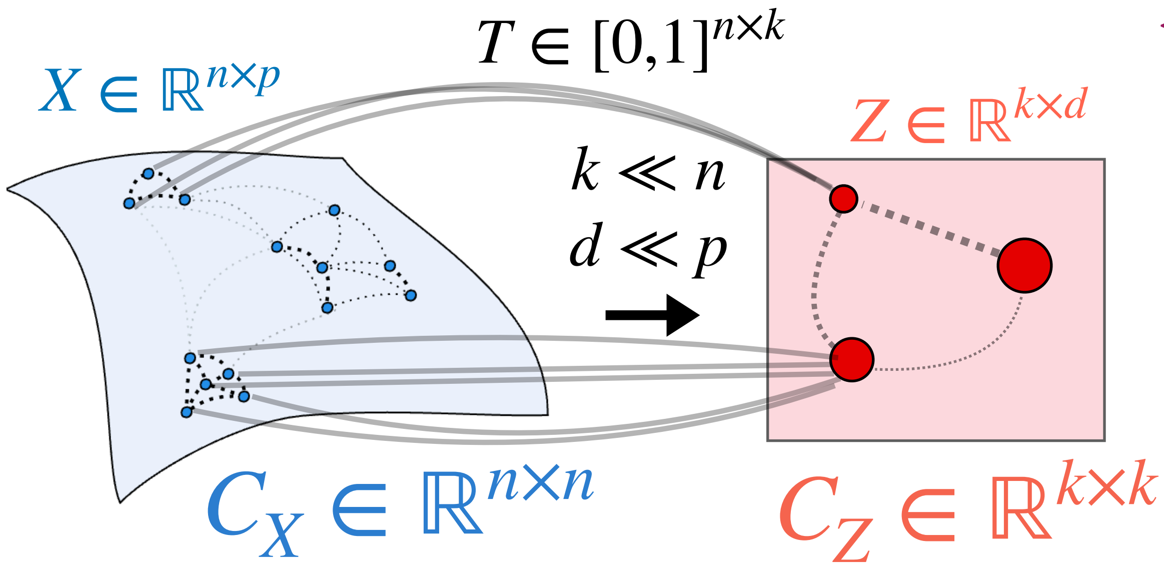
- Find **few prototypes** in **low dim.**
- Find the weights / cluster size

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

Distributional Reduction



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



✦ Optimization problem

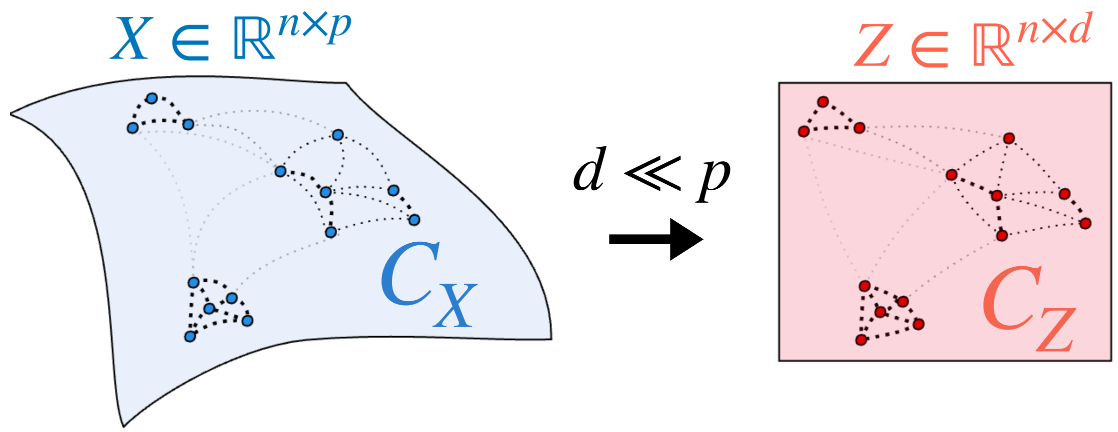
$$\min_{Z\in\mathbb{R}^{k\times d}} \min_{b\in\Sigma_k} \mathrm{GW}_L\left(C_X, C_Z, \frac{1_n}{n}, b\right)$$

- 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)

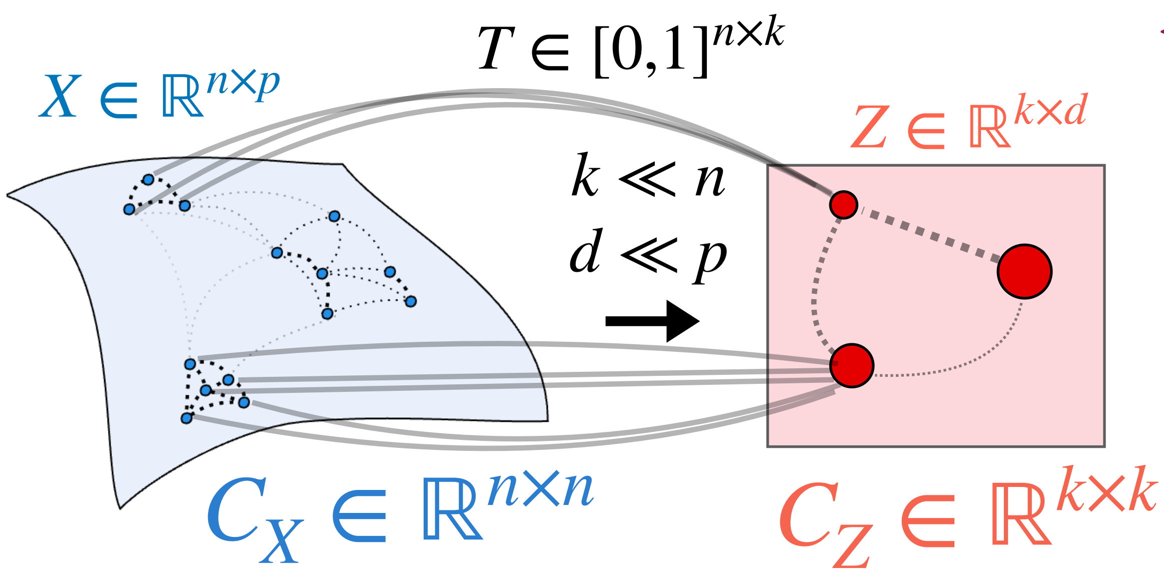
✦ GW projection

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

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**

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

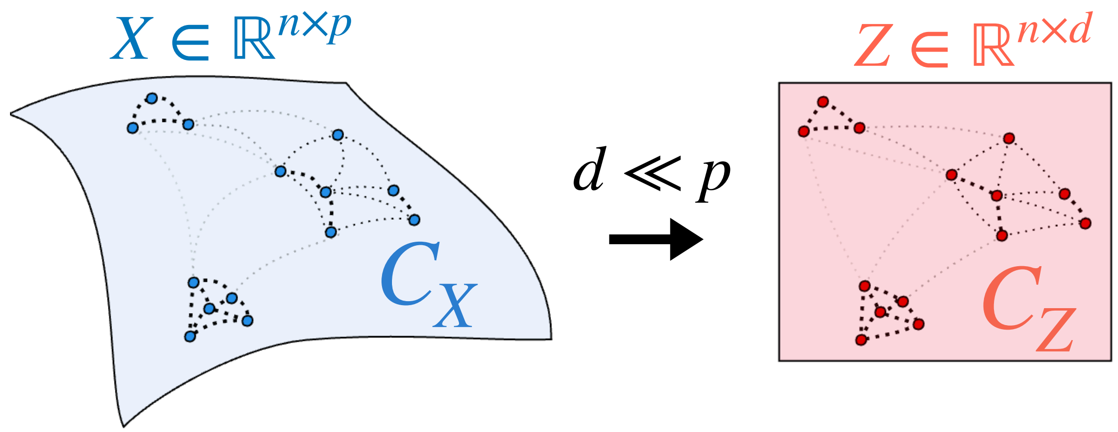
- Find **few prototypes** in **low dim.**
- Find the weights / cluster size
- **Clustering via the coupling** T (soft-assignment)
- Sufficient conditions for hard assignment (see paper)

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

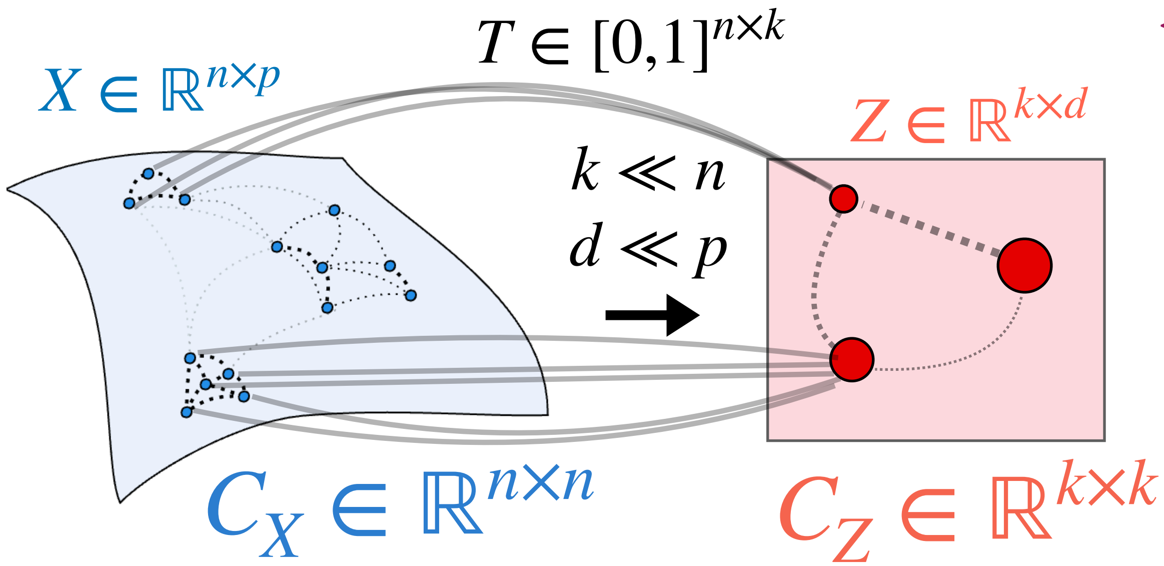
$$\min_{Z\in\mathbb{R}^{n\times d}} \min_{T:T1_k=\frac{1_n}{n}} \sum_{ijkl} L\left([C_X]_{ik}\,,\,[C_Z]_{jl}\right) T_{ij}T_{kl}$$

→ easier than GW

Distributional Reduction



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



✦ **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

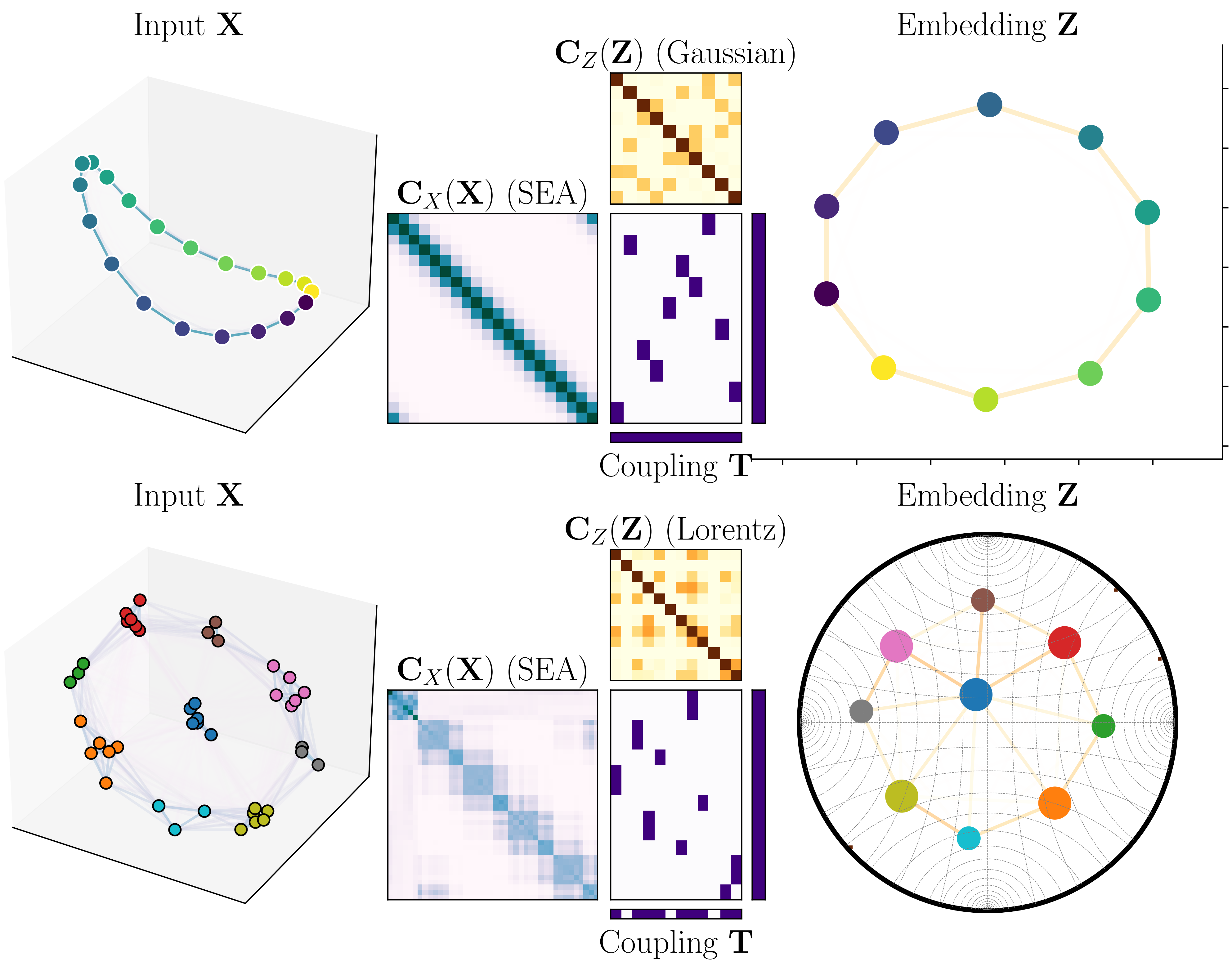
(Vincent-Cuaz et al., 2022)

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

→ easier than GW

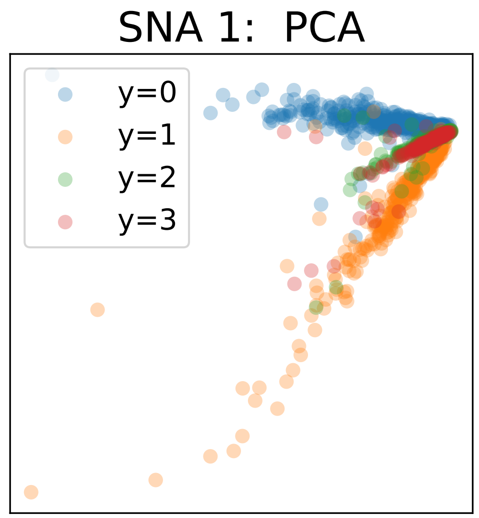
- Non-convex problem
- BCD: alternates optim in Z, in T
- Optim in T: CG solver in $O(n^2k)$ for $L \in \{\mathrm{KL}, |\cdot|^2\}$
- With low-rank structures $O(nkr + n^2)$

Distributional Reduction



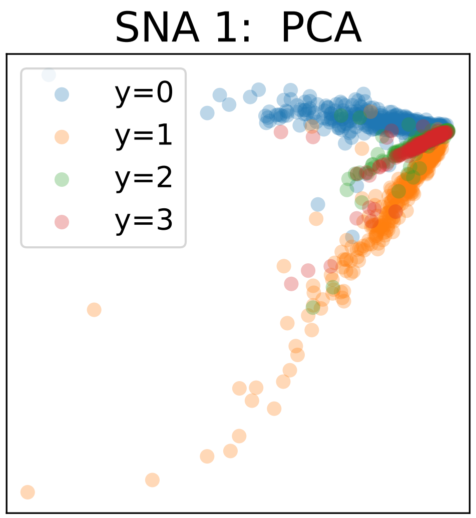
Distributional Reduction

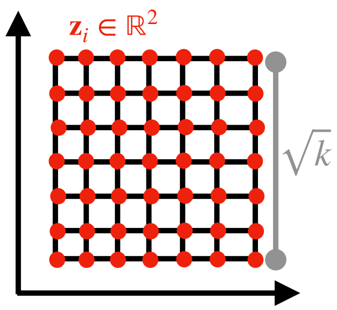
✦ **Single-cell dataset** (Chen et al., 2019) Only solve $\min_{b \in \Sigma_k} \mathrm{GW}_L(C_X, C_Z, \frac{1_n}{n}, b)$



Distributional Reduction

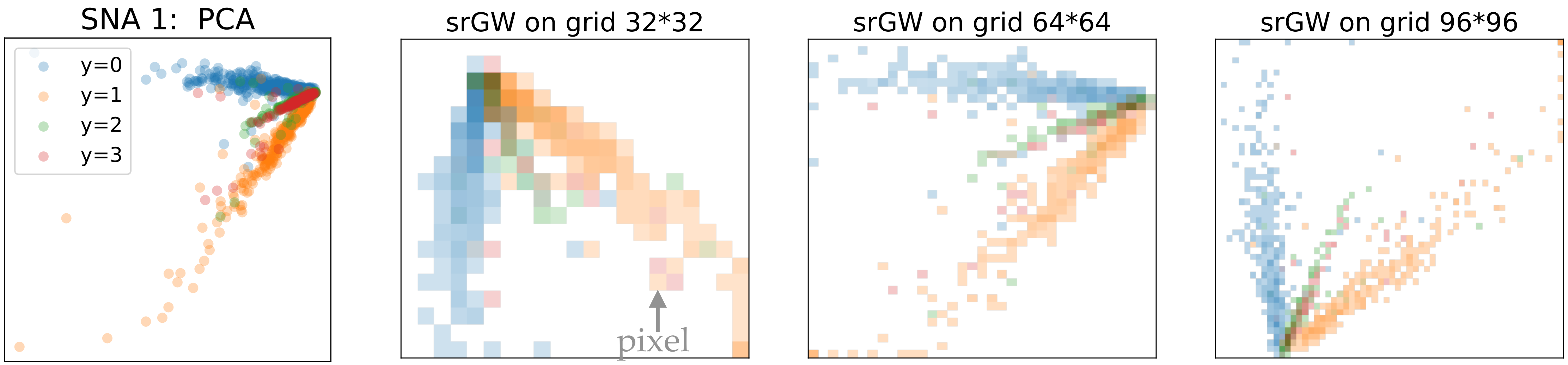
✦ **Single-cell dataset** (Chen et al., 2019) Only solve $\min_{b \in \Sigma_k} \mathrm{GW}_L(C_X, C_Z, \frac{1_n}{n}, b)$ with $C_X = XX^\top$ and $C_Z = ZZ^\top$ **fixed on a grid** $Z \in \mathbb{R}^{k \times 2}$



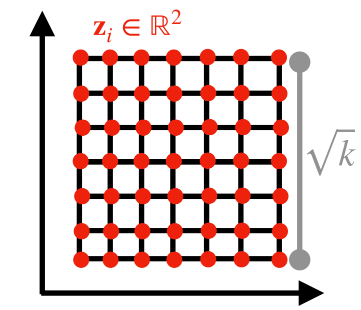


Distributional Reduction

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



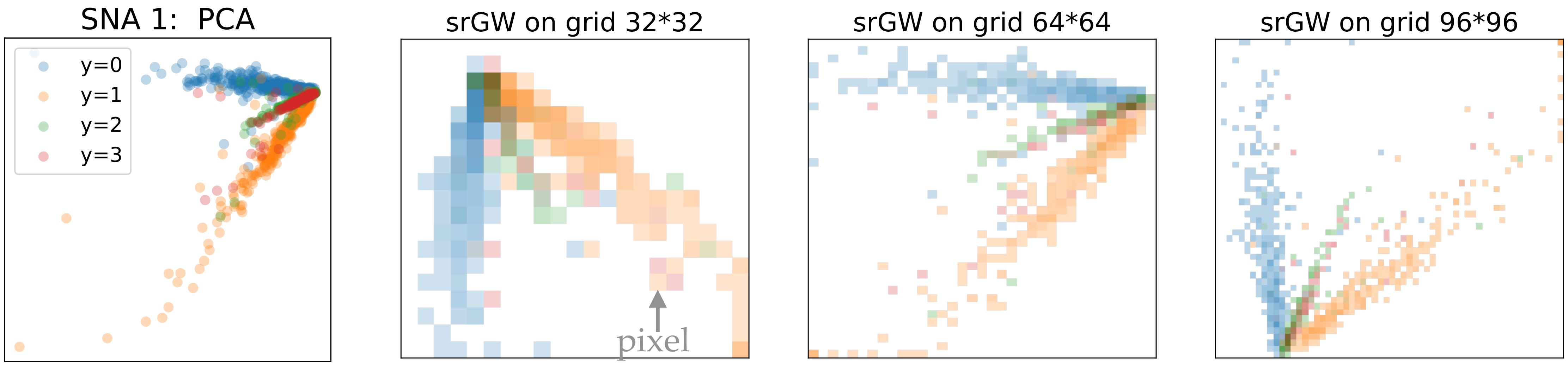
dataset as an image



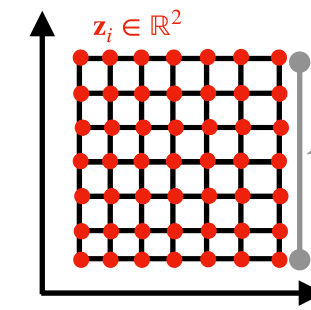
Distributional Reduction

✦ 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** $Z \in \mathbb{R}^{k \times 2}$



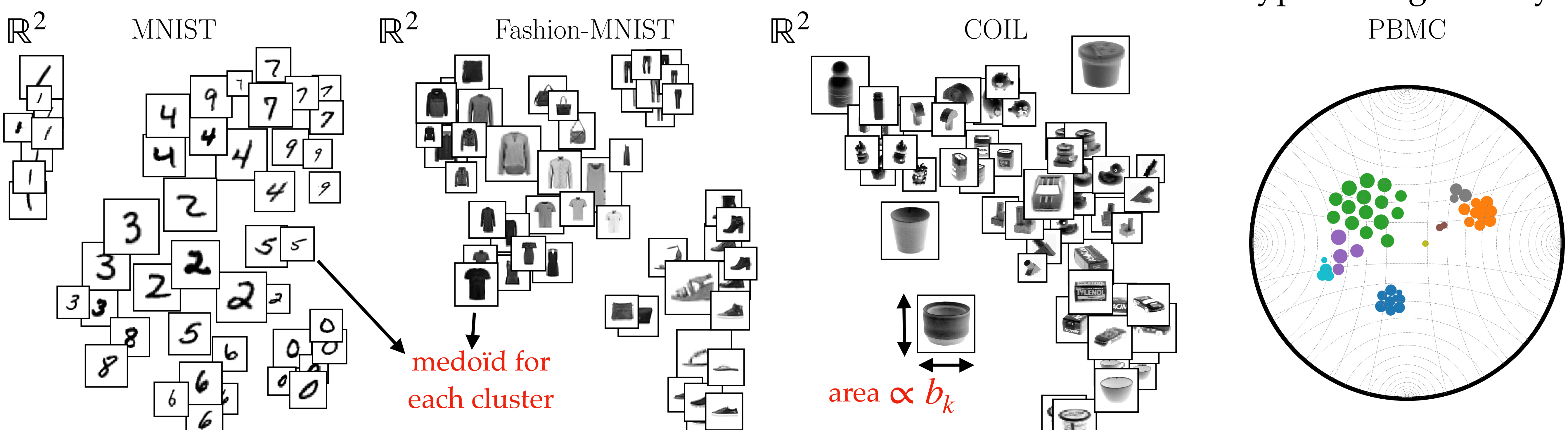
dataset as an image



✦ 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**

Silhouette score for DR

$$\overline{\mathscr{S}\mathscr{H}} = \frac{1}{2}\left(\frac{\mathscr{S}+1}{2} + \mathscr{H}\right)$$

Homogeneity score for clustering

Distributional Reduction

✦ Comparison with DR then clustering or clustering then DR

Silhouette score for DR

$$\overline{\mathcal{S}\mathcal{H}} = \frac{1}{2}\left(\frac{\mathcal{S}+1}{2} + \mathcal{H}\right)$$

Homogeneity score for clustering

are the classes preserved by the clustering ?

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

Distributional Reduction

✦ Comparison with DR then clustering or clustering then DR

Silhouette score for DR

$$\overline{\mathcal{SH}} = \frac{1}{2}\left(\frac{\mathcal{S}+1}{2} + \mathcal{H}\right)$$

Homogeneity score for clustering

are the classes preserved by the clustering ?

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

	methods	C_X / C_Z	MNIST	FMNIST	COIL	SNA1	SNA2	ZEI1	ZEI2	PBMC
$d = 10$	DistR (ours)	$\langle, \rangle_{\mathbb{R}^p}$ / $\langle, \rangle_{\mathbb{R}^d}$	**76.9 (0.4)**	74.0 (0.6)	77.4 (0.7)	75.6 (0.0)	**86.6 (2.9)**	76.5 (1.0)	49.1 (1.9)	**86.3 (0.5)**
	DistR$_\varepsilon$ (ours)	-	76.5 (0.0)	**74.5 (0.0)**	**78.5 (0.0)**	68.9 (0.2)	78.9 (1.3)	79.8 (0.2)	**52.7 (0.2)**	85.1 (0.0)
	DR→C	-	73.9 (1.7)	63.9 (0.0)	70.6 (3.3)	**77.3 (2.5)**	66.9(14.3)	73.6 (0.8)	26.9 (7.9)	76.9 (1.2)
	C→DR	-	76.5 (0.0)	74.3 (0.0)	62.3 (0.0)	68.3 (0.2)	86.0 (0.3)	79.6 (0.1)	52.5 (0.1)	86.0 (0.0)
	COOT	NA	32.8 (2.5)	28.2 (6.0)	47.9 (1.0)	49.3 (6.8)	76.6 (6.5)	**81.0 (2.4)**	30.0 (1.6)	34.5 (1.3)
$d = 2$	DistR (ours)	SEA / St.	77.5 (0.8)	**76.8 (0.5)**	83.3 (0.7)	80.2 (1.9)	88.5 (0.0)	**81.0 (0.6)**	47.6 (1.1)	82.5 (1.2)
	DistR$_\varepsilon$ (ours)	-	**77.9 (0.2)**	75.6 (0.6)	**83.9 (0.5)**	**80.4 (1.8)**	**90.7 (0.1)**	79.9 (0.2)	**48.2 (0.6)**	86.4 (0.2)
	DR→C	-	74.8 (0.9)	75.7 (0.5)	80.3 (0.5)	77.2 (0.4)	89.3 (0.1)	79.0 (0.5)	47.4 (2.7)	82.0 (1.7)
	C→DR	-	76.2 (0.6)	75.0 (0.3)	81.6 (0.3)	77.6 (0.5)	89.8 (0.1)	78.8 (0.7)	45.8 (0.9)	**88.4 (0.5)**
	COOT	NA	26.1 (5.7)	24.9 (1.5)	42.5 (2.5)	32.6 (5.1)	56.8 (4.0)	78.2 (0.7)	25.2 (0.6)	28.9 (3.2)
$d = 2$	DistR (ours)	SEA / H-St.	**75.0 (0.0)**	**75.3 (0.6)**	70.2 (0.8)	75.4 (0.8)	**88.9 (1.8)**	77.4 (3.1)	**41.6 (1.7)**	**73.2 (1.6)**
	DistR$_\varepsilon$ (ours)	-	66.9 (0.5)	66.4 (0.3)	69.6 (0.9)	**81.3 (5.1)**	78.6 (1.0)	73.6 (1.8)	38.6 (0.7)	72.9 (2.3)
	DR→C	-	58.4 (4.6)	59.7 (9.7)	47.4 (1.5)	51.8 (5.9)	58.1 (9.9)	**80.6 (5.2)**	41.5 (2.4)	67.4 (7.2)
	C→DR	-	67.1 (1.8)	66.3 (0.5)	**71.1 (1.1)**	80.8 (3.9)	75.0(0.0)	71.2 (0.9)	37.5 (0.4)	67.6 (0.4)