Interpolating between Clustering and Dimensionality Reduction with Gromov-Wasserstein

ENS DE LYON

EPFL

Hugues Van Assel*, Cédric Vincent-Cuaz*, Titouan Vayer, Rémi Flamary, Nicolas Courty

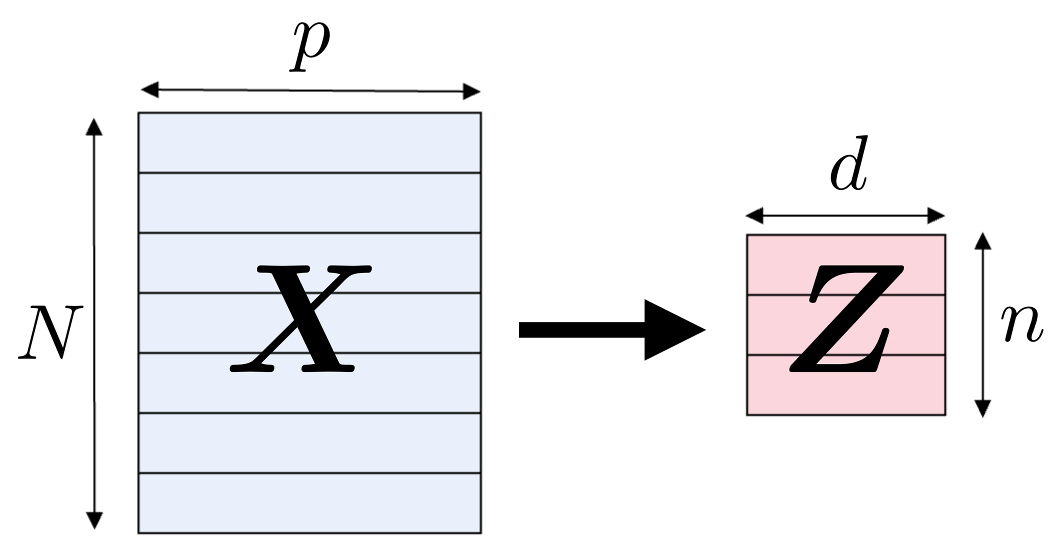
*Equal contribution

l'X École Polytechnique, IP Paris

Université Bretagne Sud (UBS)

The goal is to perform **joint clustering and dimensionality reduction** (DR).

To do so, we **formulate DR as a graph matching problem** and **augment this objective with Gromov-Wasserstein optimal transport**.



Graph matching as common objective for traditional DR methods

$$\min_{\mathbf{Z}} \sum_{(i,j)\in [\![N]\!]^2} L([\mathbf{C}_X]_{ij}, [\mathbf{C}_Z]_{ij}) \,.$$

$\mathbf{C}_X$ and $\mathbf{C}_Z$ are affinity matrices defined from $\mathbf{X}$ and $\mathbf{Z}$ resp.

	L	$\mathbf{C}_X$	$\mathbf{C}_Z$
PCA	L_2	$\mathbf{X}\mathbf{X}^\top$	$\mathbf{Z}\mathbf{Z}^\top$
KPCA	L_2	$\mathbf{K}_X$	$\mathbf{Z}\mathbf{Z}^\top$
NE	L_{KL}	$\mathbf{K}_X$	$\mathbf{K}_Z$

kernels

Gromov-Wasserstein discrepancy computes the **best coupling** between affinities

$$\mathrm{GW}_L(\mathbf{C}_X, \mathbf{h}_X, \mathbf{C}_Z, \mathbf{h}_Z) := \min_{\mathbf{T}\in\mathcal{U}(\mathbf{h}_X,\mathbf{h}_Z)} \sum_{(i,j)\in[\![N]\!]^2} \sum_{(k,l)\in[\![n]\!]^2} L([\mathbf{C}_X]_{ij}, [\mathbf{C}_Z]_{kl}) T_{ik} T_{jl}$$

Semi-Relaxed GW relaxes the second marginal

$$\mathrm{srGW}_L(\mathbf{C}_X, \mathbf{h}_X, \overline{\mathbf{C}}) := \min_{\overline{\mathbf{h}}\in\Sigma_n} \mathrm{GW}_L(\mathbf{C}_X, \mathbf{h}_X, \overline{\mathbf{C}}, \overline{\mathbf{h}}) \,.$$

If $\mathbf{U} \mapsto \mathrm{vec}(\mathbf{U})^\top (\mathbf{C}_X \otimes \mathbf{C}_X)\,\mathrm{vec}(\mathbf{U})$ is convex on $\mathcal{U}(\mathbf{h}_X, \mathbf{h}_X)$, then

$$\min_{\overline{\mathbf{C}}\in\mathbb{R}^{n\times n}} \mathrm{srGW}_L(\mathbf{C}_X, \mathbf{h}_X, \overline{\mathbf{C}}) \qquad \text{(srGWB)}$$

with $L = L_2$ admits scaled membership matrices as optimum for $\mathbf{T}$.

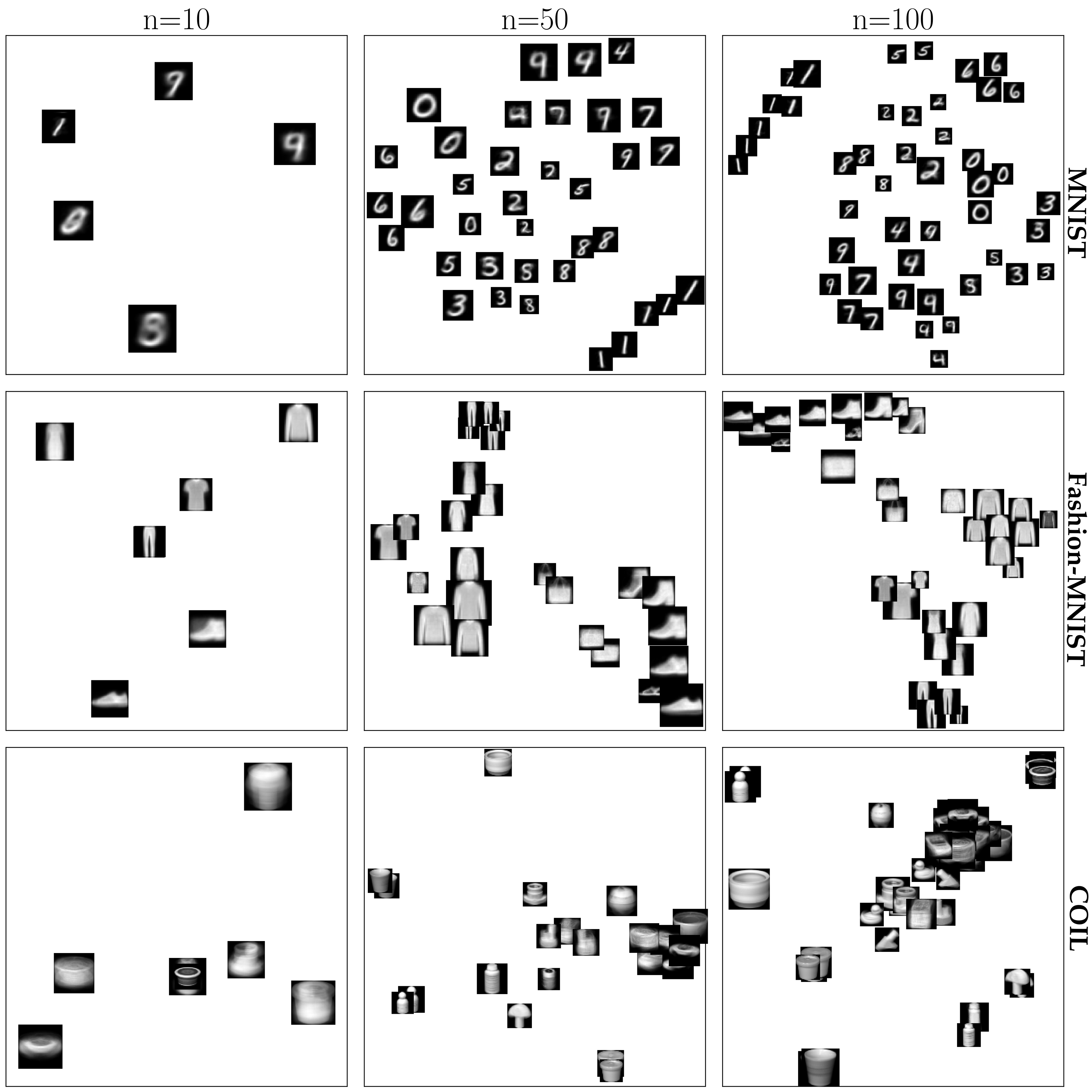
ARI spectral clustering

	srGWI	srGWB
MNIST	29.7(1.9)	**32.6(1.8)**
F-MNIST	26.1(0.0)	**39.5(0.3)**
COIL	18.1(0.2)	**51.0(1.7)**

GWDR Model

$$\min_{\mathbf{Z}\in\mathbb{R}^{n\times d}} \mathrm{srGW}_L(\mathbf{C}_X, \mathbf{h}_X, \mathbf{C}_Z)$$

- Clustering provided by the OT plan is learned jointly with the embeddings and can adapt to the varying cluster sizes.
- Can be seen as a constrained srGW barycenter problem.
- Flexibly adapts to any DR method by choosing the corresponding affinities.



Notations

$L_2(x,y) := (x-y)^2$

$L_{\mathrm{KL}}(x,y) := x\log(x/y) - x + y$

$\mathrm{srGWI} := \mathrm{srGW}_L(\mathbf{C}_X, \mathbf{h}_X, \mathbf{I}_n)$

$\mathcal{U}(\mathbf{h}_X, \mathbf{h}_Z) := \{\mathbf{T} \geq \mathbf{0}, \mathbf{T}\mathbf{1} = \mathbf{h}_X, \mathbf{T}^\top\mathbf{1} = \mathbf{h}_Z\}$

$\Sigma_n := \{\mathbf{h} \geq \mathbf{0}, \mathbf{h}^\top\mathbf{1} = 1\}$

ARI = Adjusted Rand Index