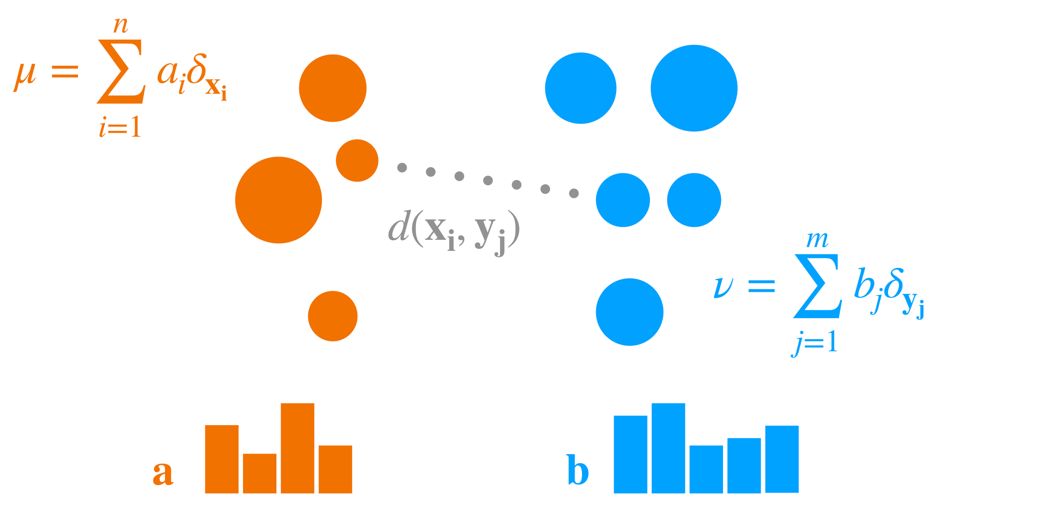
BHOT: Barnes-Hut for Optimal Transport

Arthur Lebeurrier[1], Titouan Vayer[1]

[1]OCKHAM, Inria Lyon, LIP, ENS Lyon,

Motivation: Faster regularized Optimal Transport, efficiently scaling for Machine Learning.

Classical OT Problem



Wasserstein distance

$$W_p^p(\mu, \nu) = \min_T \int_{X \times X} d(x, y)^p dT(x, y)$$

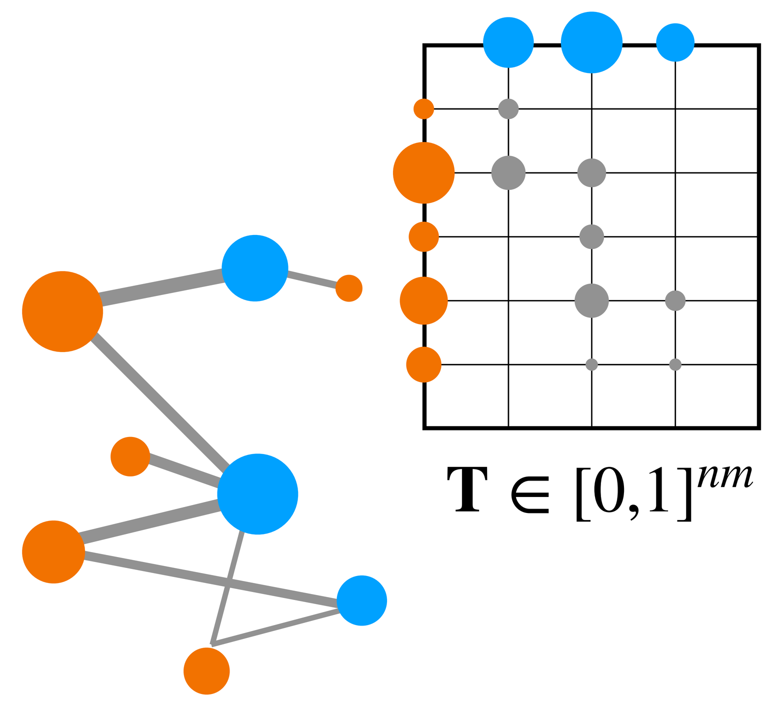
A linear optimization problem in $O(n^3 \log(n)^2)$

Find plan $T \in M_{nm}(\mathbb{R}^+)$

$$\min_T \sum_{ij} d(\mathbf{x_i}, \mathbf{y_j}) T_{ij}$$

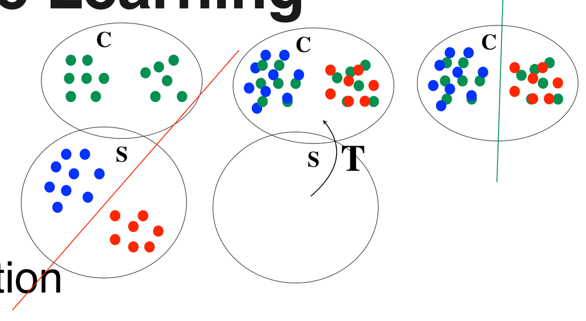
Constraints

$$\mathbf{T1}_m = \mathbf{a}$$
$$\mathbf{T}^\top \mathbf{1}_n = \mathbf{b}$$



In Machine Learning

- **Domain adaptation**
- Clustering
- Dimensionality reduction



Entropic regularization [1]

$$\min_T \sum_{ij} d(\mathbf{x_i}, \mathbf{y_j}) T_{ij} - \varepsilon H(\mathbf{T})$$

Solve an approx solution in $O(n^2)$

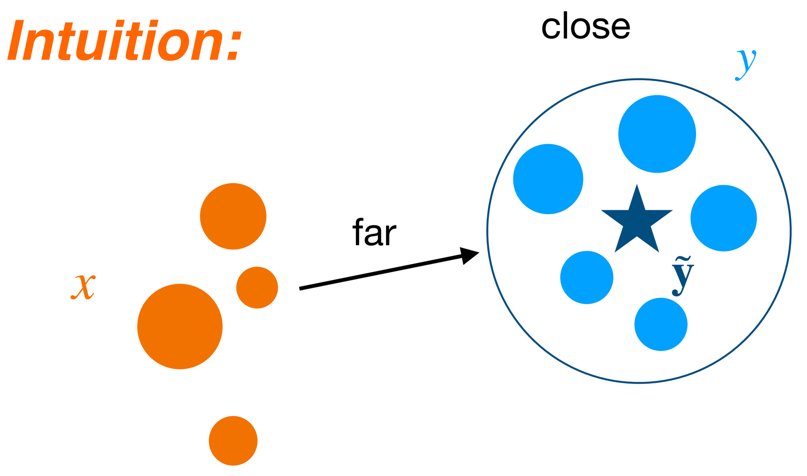
Sinkhorn algorithm

$K_{ij} := \exp(-d(\mathbf{x_i}, \mathbf{y_j})/\varepsilon)$
while not converge:
$\mathbf{v} \leftarrow \mathbf{b} \oslash \mathbf{K}^\top \mathbf{u}; \mathbf{u} \leftarrow \mathbf{a} \oslash \mathbf{K}\mathbf{v}$
return: $T_{ij} = u_i K_{ij} v_j$

Still prohibitive for classical ML applications

Barnes-Hut and N - body Methods [2]

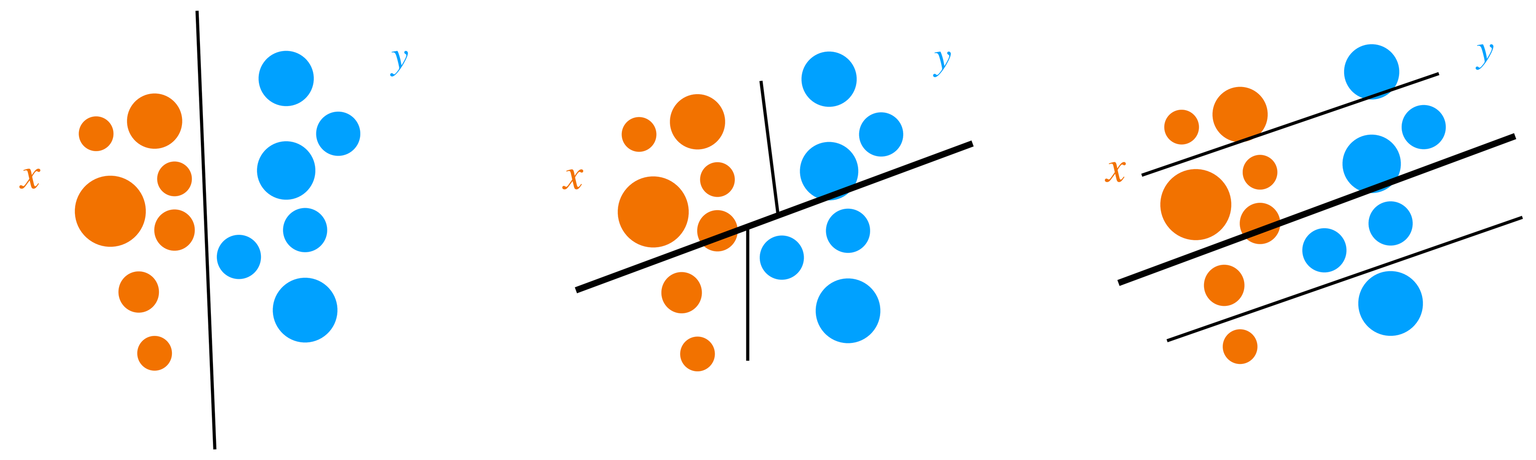
Intuition:



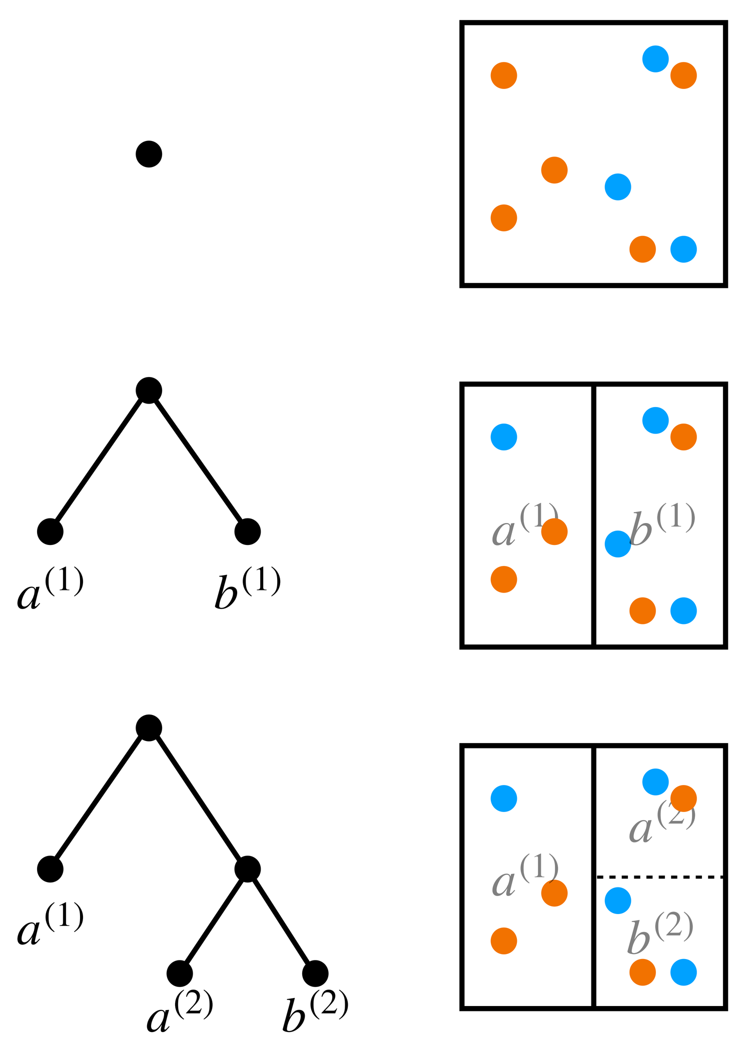
$$\mathbf{K} \approx \mathbf{U}\mathbf{V}^\top = \begin{pmatrix} \kappa(\mathbf{x_1}, \tilde{\mathbf{y}}) \\ \kappa(\mathbf{x_2}, \tilde{\mathbf{y}}) \\ \vdots \\ \kappa(\mathbf{x_n}, \tilde{\mathbf{y}}) \end{pmatrix} \times \mathbf{1}_m^\top$$

$\mathbf{v} \leftarrow \tilde{\mathbf{K}}\mathbf{v}$ in $O(n)$

A pipeline for any hierarchical clustering in any dimension



Hierarchical Clustering



$$\tilde{\mathbf{K}}^{(0)} = \begin{bmatrix} \mathbf{K} \end{bmatrix}$$

$$\tilde{\mathbf{K}}^{(1)} = \begin{bmatrix} \mathbf{K}_{aa}^{(1)} & \tilde{\mathbf{K}}_{ab}^{(1)} \\ \tilde{\mathbf{K}}_{ba}^{(1)} & \mathbf{K}_{bb}^{(1)} \end{bmatrix}$$

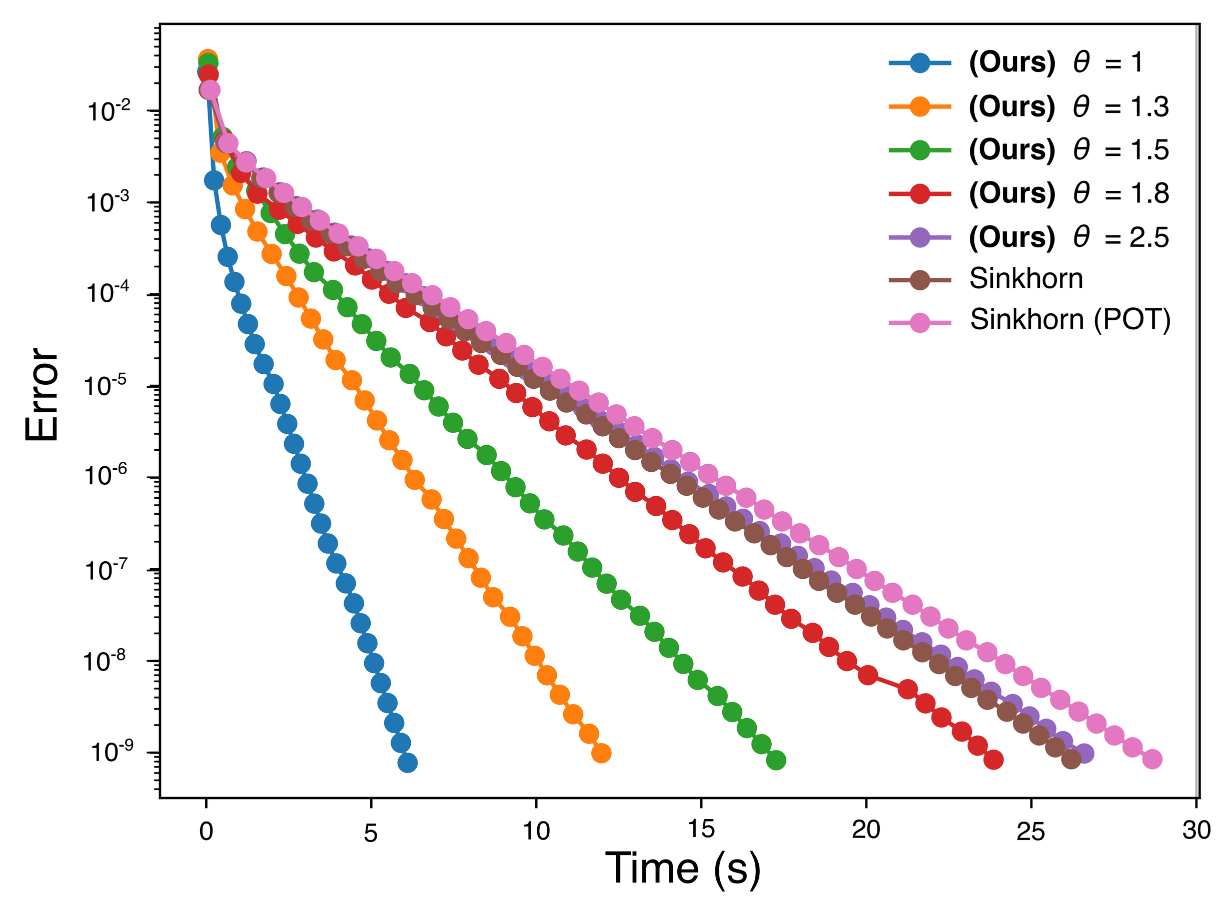
$$\tilde{\mathbf{K}}^{(2)} = \begin{bmatrix} \mathbf{K}_{aa}^{(1)} & \tilde{\mathbf{K}}_{ab}^{(1)} \\ \tilde{\mathbf{K}}_{ba}^{(1)} & \begin{matrix} \mathbf{K}_{aa}^{(2)} & \tilde{\mathbf{K}}_{ab}^{(2)} \\ \tilde{\mathbf{K}}_{ba}^{(2)} & \mathbf{K}_{bb}^{(2)} \end{matrix} \end{bmatrix}$$

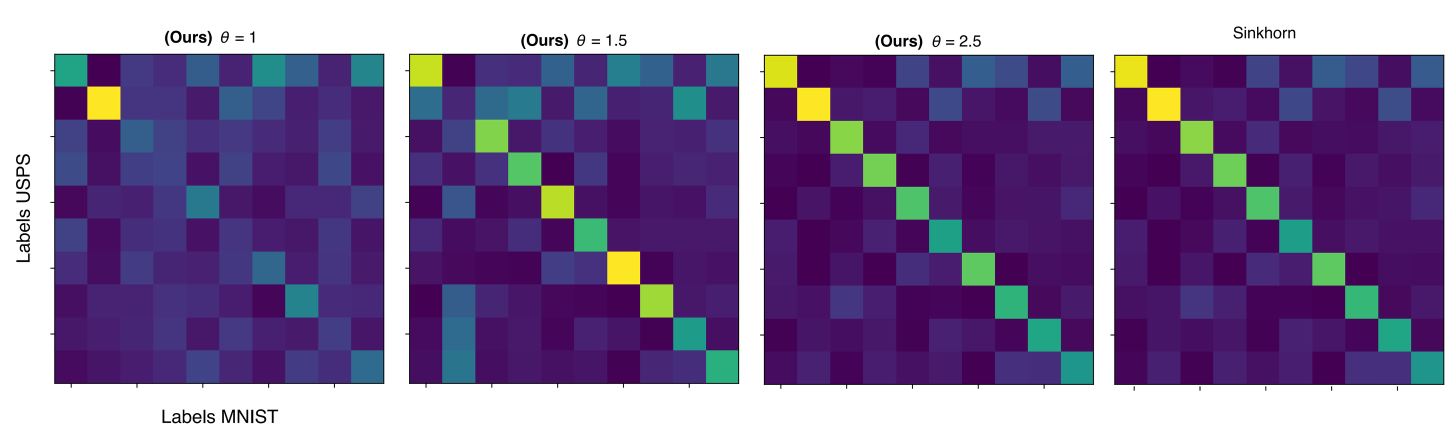
- $\tilde{\mathbf{K}}_{ij}^{(l)} = UV^\top$: **Low-rank** approx of *far interactions*
- **Cheap Tree traversal** to detect *near* and *far*

→ We introduce a new parameter θ in order to control the approximation. When θ is $+\infty$ we find the original kernel

Experiments

Image classification: MNIST/USPS datasets





23.75% 41.02% 57.58 % **57.97 %**

[1] Cuturi, M. (2013). Sinkhorn distances: Lightspeed computation of optimal transport. Advances in neural information processing systems, 26.

[2] Ryan, J. P., Ament, S. E., Gomes, C. P., & Damle, A. (2022, May). The fast kernel transform. In International Conference on Artificial Intelligence and Statistics (pp. 11669-11690). PMLR.