

SCREENKHORN

Screening Sinkhorn Algorithm for Regularized Optimal Transport

Mokhtar Z. Alaya

Équipe APP - LITIS, Université de Rouen Normandie

<https://mzalaya.github.io/>



Séminaire CMAP, 4 juin 2020

Joint work with ...



- Maxime Bérar
- Asso. Professor at LITIS
- Univ. Rouen Normandie



- Gilles Gasso
- Professor at LITIS
- INSA Rouen



- Alain Rakotomamonjy
- Professor at LITIS
- Univ. Rouen Normandie & Criteo Paris

The OATMIL project

Bringing Optimal Transport and Machine Learning Together

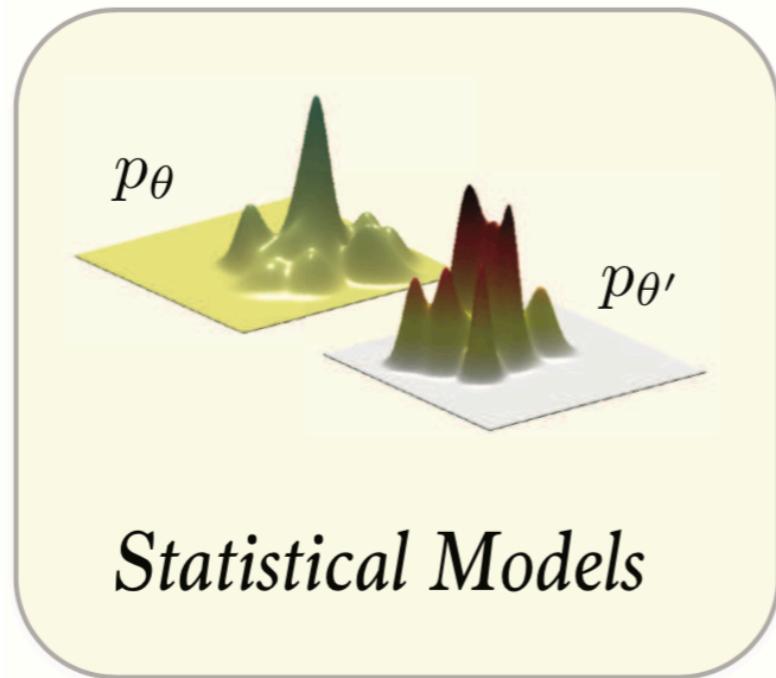
The diagram illustrates the OATMIL project. It shows two probability distributions, $\mu_s(\mathbf{X})$ and $\mu_t(\mathbf{T}(\mathbf{X}))$, defined over domains Ω_1 and Ω_2 . A transport map \mathbf{T} is represented by a curved arrow that maps points from $\mathbf{X} \in \Omega_1$ to $\mathbf{T}(\mathbf{X}) \in \Omega_2$.

Optimal Transport (OT)

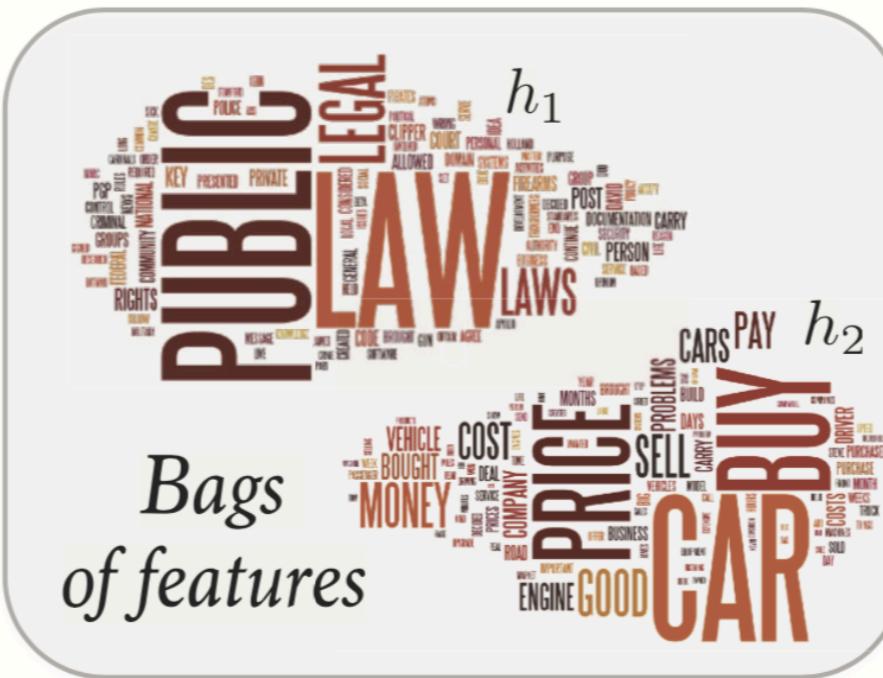
What is it ?

OT is ...

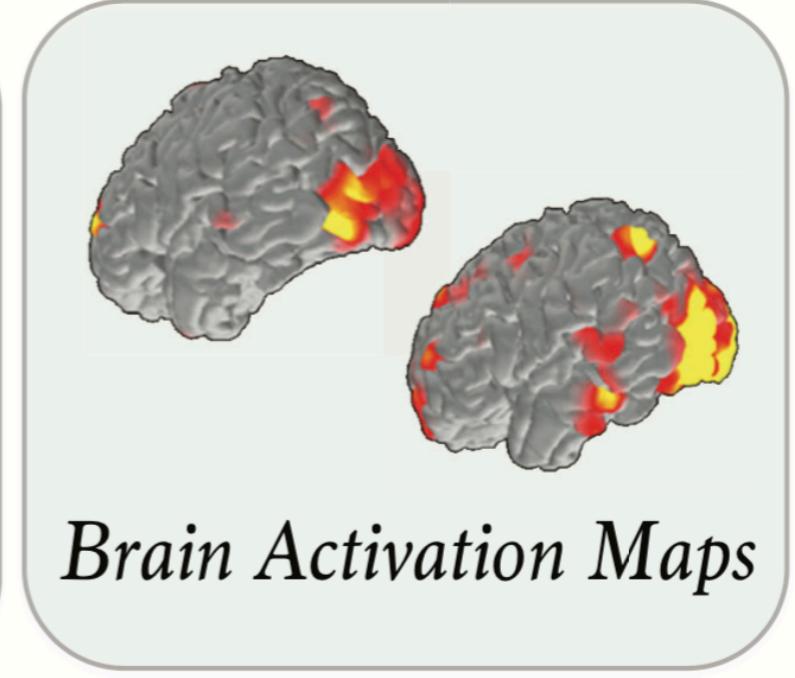
A method for comparing probability distributions with the ability to incorporate spatial information.



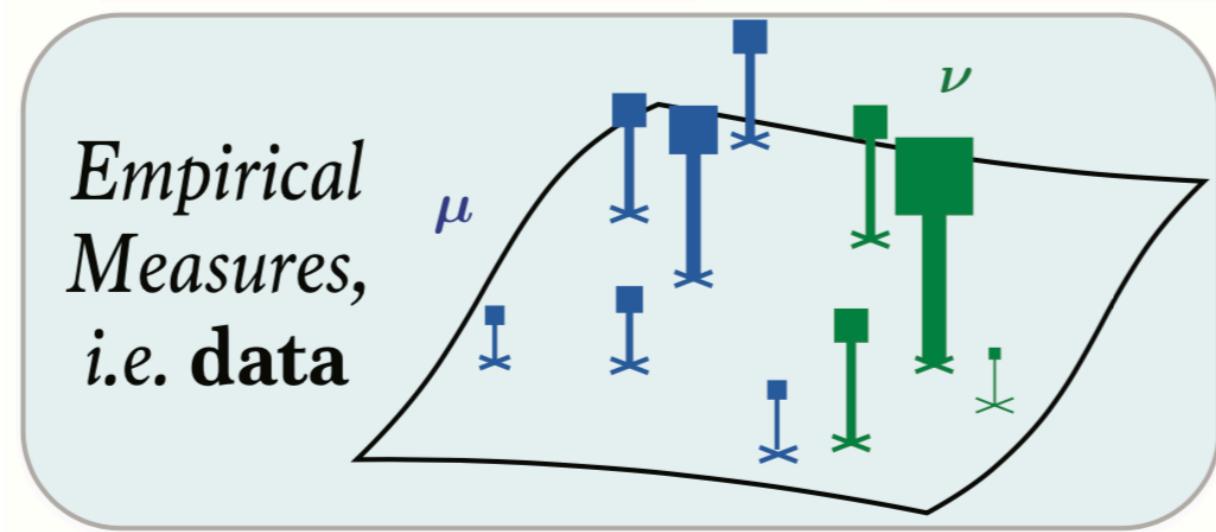
Statistical Models



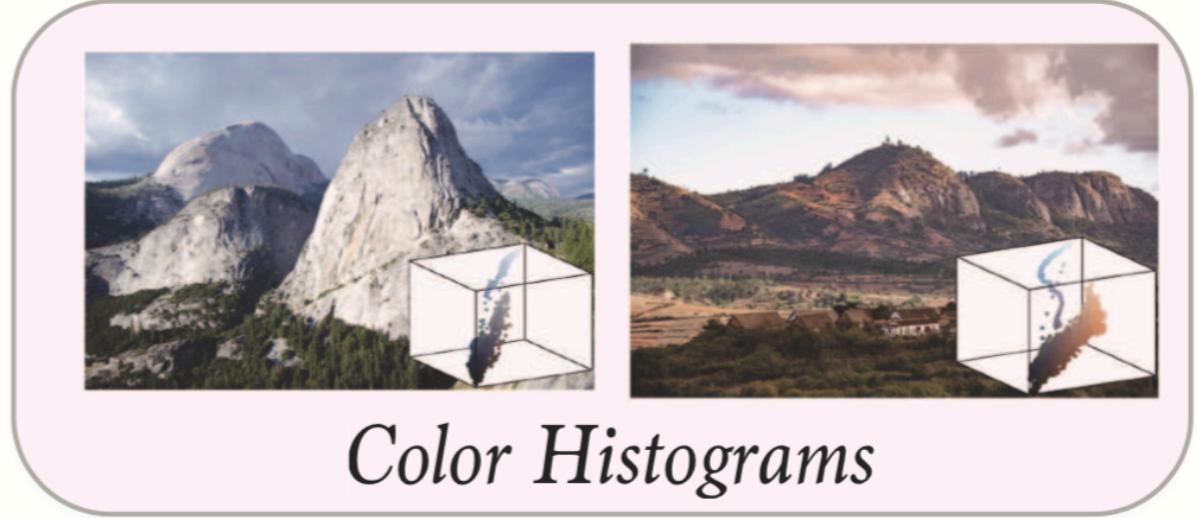
*Bags
of features*



Brain Activation Maps

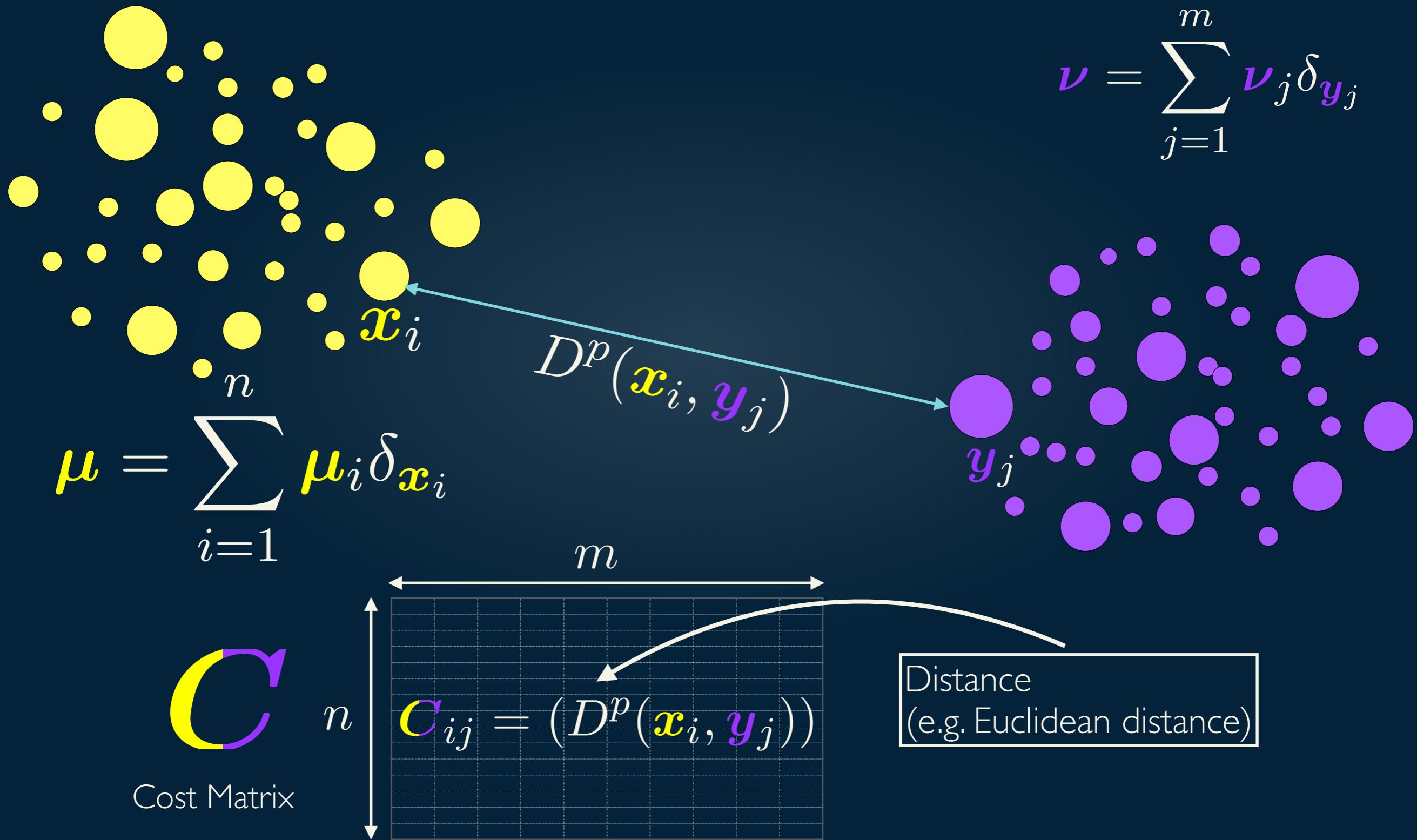


*Empirical
Measures,
i.e. data*



Color Histograms

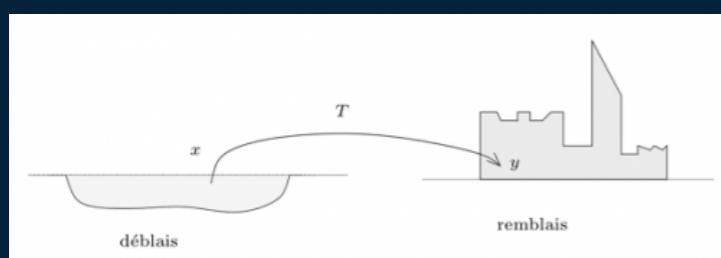
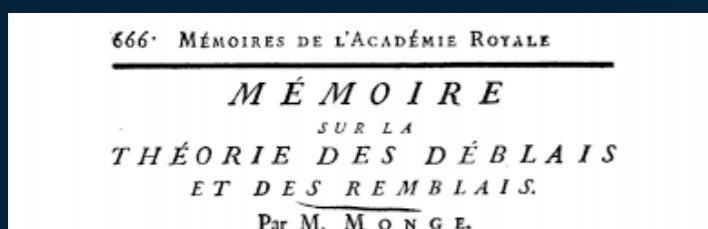
Discrete OT Framework



Discrete OT Framework: Monge's Formula



Gaspard Monge
(1746 - 1818)



A diagram showing two sets of points. On the left, there is a cluster of yellow dots. An arrow labeled T points to a cluster of purple dots on the right. Below this, a mathematical expression is shown:

$$\min_{\# \mu = \nu} \sum_i D^p(\mathbf{x}_i, \mathbf{T}(\mathbf{x}_i)) \mu_i$$

Below this, another equation is shown in a box:

$$\forall j \in \{1, \dots, m\}, \nu_j = \sum_{i: \mathbf{T}(\mathbf{x}_i) = \mathbf{y}_j} \mu_i$$

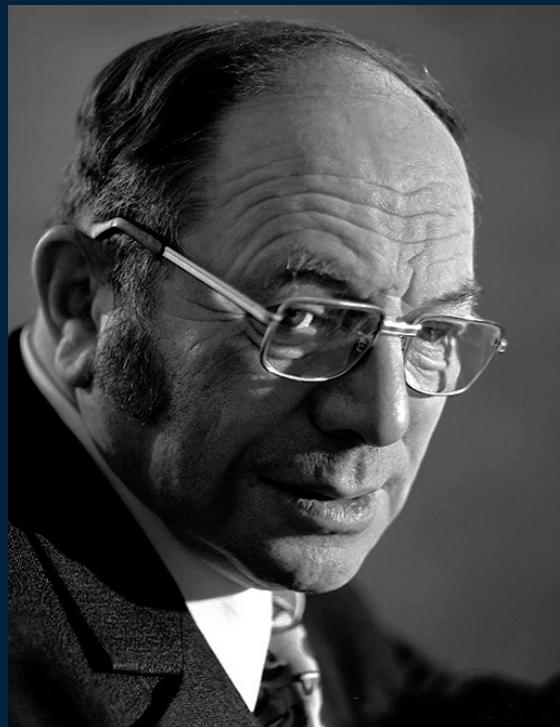
Strict: Deterministic Assignments

Uniform weights
 n

$$\min_{\sigma \in \mathfrak{S}_n} \sum_{i=1} D^p(\mathbf{x}_i, \mathbf{y}_{\sigma(i)})$$

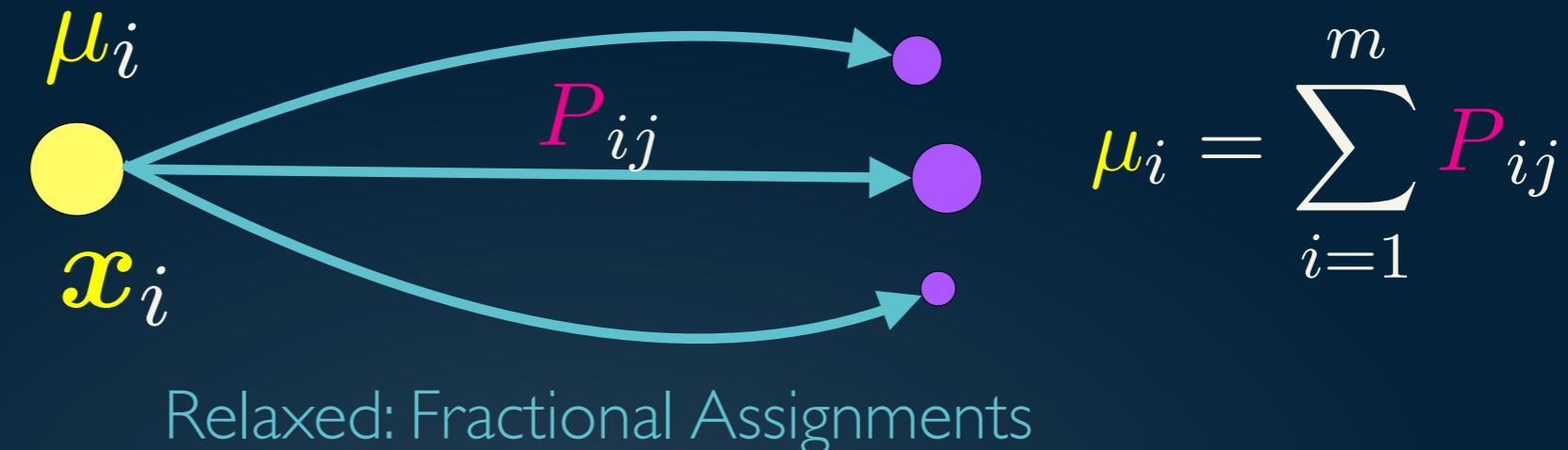


Regularized Discrete OT Framework: Kantorovich's Formula



Leonid Kantorovich
(1912-1986)

Kantorovich 1942



$$\mu_i = \sum_{j=1}^m P_{ij}$$

Probabilistic couplings set (Transport Polytope)

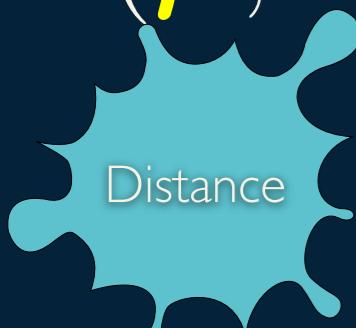
$$\Pi(\mu, \nu) = \{P \in \mathbb{R}_+^{n \times m}, P\mathbf{1}_m = \mu, P^\top \mathbf{1}_n = \nu\}$$

Mass conservation constraints

- Computing OT between μ and ν amounts to solving a linear problem:

$$S(\mu, \nu) = \min_{P \in \Pi(\mu, \nu)} \left\{ \langle C, P \rangle = \sum_{i=1}^n \sum_{j=1}^m C_{ij} P_{ij} \right\}$$

Monge-Kantorovich / Wasserstein Distance



Distance

Regularized Discrete OT Framework: Sinkhorn Divergence

- Linear programming problem that requires generally $\mathcal{O}(n^3 \log(n)^2)$ arithmetic operations.
- Entropic regularization of OT distances relies on the addition of a penalty term as follows:

$$\mathcal{S}_\eta(\boldsymbol{\mu}, \boldsymbol{\nu}) = \min_{\mathbf{P} \in \Pi(\boldsymbol{\mu}, \boldsymbol{\nu})} \left\{ \langle \mathbf{C}, \mathbf{P} \rangle - \eta H(\mathbf{P}) \right\}$$

Regularisation parameter

↓

↑ Sinkhorn divergence

↑ Negative entropy

(Cuturi, 2013)

$$H(\mathbf{P}) = - \sum_{i,j} \mathbf{P}_{ij} \log(\mathbf{P}_{ij})$$

Regularized Discrete OT Framework:

Dual of $\mathcal{S}_\eta(\mu, \nu)$

Dual of Sinkhorn divergence

$$\mathcal{S}_\eta^d(\mu, \nu) = \min_{\mathbf{u} \in \mathbb{R}^n, \mathbf{v} \in \mathbb{R}^m} \left\{ \Psi(\mathbf{u}, \mathbf{v}) := \mathbf{1}_n^\top B(\mathbf{u}, \mathbf{v}) \mathbf{1}_m - \langle \mathbf{u}, \mu \rangle - \langle \mathbf{v}, \nu \rangle \right\}$$

where

$$B(\mathbf{u}, \mathbf{v}) := \text{diag}(e^{\mathbf{u}}) \mathbf{K} \text{diag}(e^{\mathbf{v}})$$

$$\mathbf{K} = e^{-\mathbf{C}/\eta}$$

↳ Gibbs Kernel

- The primal optimal solution \mathbf{P}^* takes the form:

$$\mathbf{P}^* = \text{diag}(e^{\mathbf{u}^*}) \mathbf{K} \text{diag}(e^{\mathbf{v}^*})$$

Optimal Transportation Plan

$$\text{with } (\mathbf{u}^*, \mathbf{v}^*) = \underset{\mathbf{u} \in \mathbb{R}^n, \mathbf{v} \in \mathbb{R}^m}{\operatorname{argmin}} \left\{ \Psi(\mathbf{u}, \mathbf{v}) \right\}$$

Dual Optimal Variables

Regularized Discrete OT Framework: Sinkhorn Algorithm

- \mathbf{P}^* can be solved efficiently by Sinkhorn iterations (near- $\mathcal{O}(n^2)$ complexity (Altschuler et al., 2017))

SINKHORN($\mathbf{C}, \boldsymbol{\mu}, \boldsymbol{\nu}$) Matrix-Scaling Problem

1. $\mathbf{a}^{(0)} \leftarrow \mathbf{1}_n/n, \mathbf{b}^{(0)} \leftarrow \mathbf{1}_m/m;$
2. $\mathbf{K} \leftarrow e^{-\mathbf{C}/\eta};$
3. For $k = 1, 2, 3, \dots$
$$\begin{cases} \mathbf{a}^{(k)} \leftarrow \boldsymbol{\mu} \oslash \mathbf{K} \mathbf{b}^{(k-1)}; \\ \mathbf{b}^{(k)} \leftarrow \boldsymbol{\nu} \oslash \mathbf{K}^\top \mathbf{a}^{(k-1)}; \end{cases}$$
4. Return $\text{diag}(\mathbf{a}^{(k)}) \mathbf{K} \text{diag}(\mathbf{b}^{(k)})$



(Flamary et al. 2017)

```
from ot import sinkhorn  
P_star = sinkhorn(mu, nu, C, eta)
```

Screened Dual of Sinkhorn Divergence

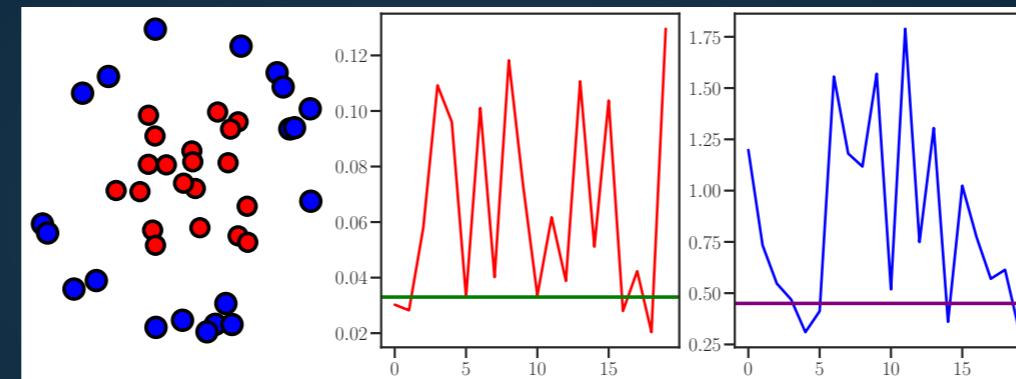
II

II

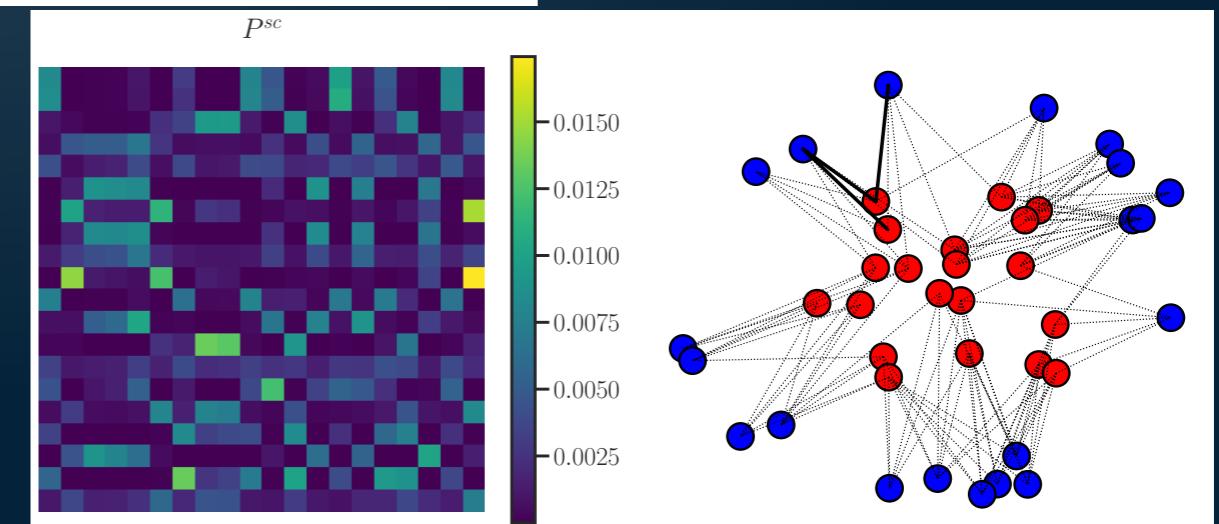
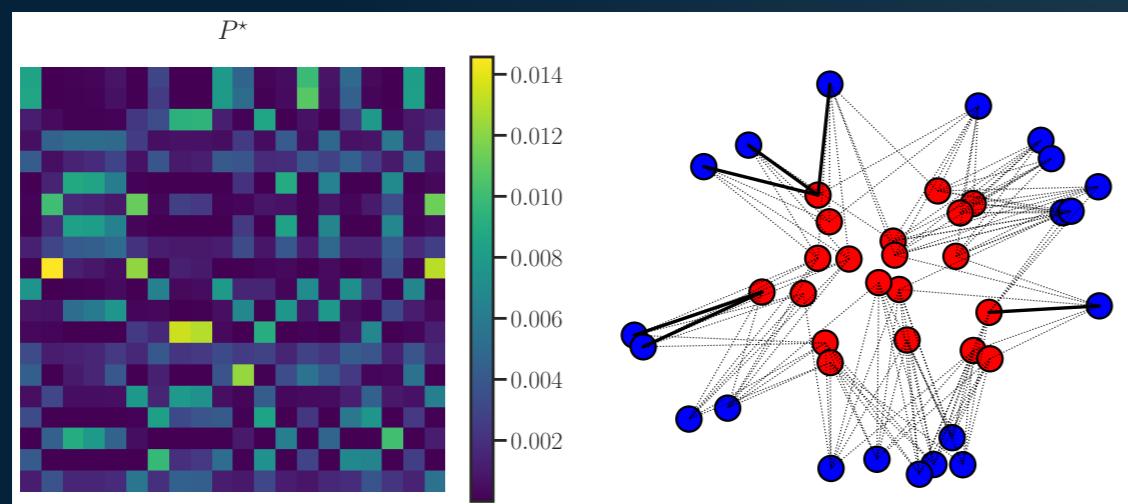
Screened Dual of Sinkhorn Divergence: Motivations

- OT plan presents a large number of neglectable values (Blondel et al., 2018)
- *Static screening test* in Lasso (Ghaoui et al., 2010)
- We define the convex set $\mathcal{C}_\alpha^r = \{w \in \mathbb{R}^r : e^{w_i} \geq \alpha\}$, for $\alpha > 0$

P^*



P^{sc}



- Identify these indices and fixed at the thresholds before solving the problem.
→ Reduce the scale of the optimization procedure.

Static Screening Test

Approximate Dual of $\mathcal{S}_\eta(\mu, \nu)$

- Based on this idea, we define a so-called *approximate dual of Sinkhorn divergence*.

Approximate dual of Sinkhorn divergence

$$\mathcal{S}_\eta^{\text{ad}}(\mu, \nu) = \min_{\substack{\mathbf{u} \in \mathcal{C}_{\frac{\varepsilon}{\kappa}}^n, \mathbf{v} \in \mathcal{C}_{\varepsilon\kappa}^m}} \left\{ \Psi_\kappa(\mathbf{u}, \mathbf{v}) := \mathbf{1}_n^\top B(\mathbf{u}, \mathbf{v}) \mathbf{1}_m - \langle \kappa \mathbf{u}, \mu \rangle - \langle \frac{\mathbf{v}}{\kappa}, \nu \rangle \right\}$$

- This is a simply dual Sinkhorn with lower-bounded variables, where the bounds are

$$\alpha_{\mathbf{u}} = \frac{\varepsilon}{\kappa} \quad \text{and} \quad \alpha_{\mathbf{v}} = \varepsilon\kappa$$

where $\varepsilon > 0$ and $\kappa > 0$ being fixed numeric constants

Static Screening Test: Definition

- κ -parameter plays a role of *scaling factor*
→ closed order of the potential components $e^{\textcolor{blue}{u}}$ and $e^{\textcolor{red}{v}}$
- Without κ the components $e^{\textcolor{blue}{u}}$ and $e^{\textcolor{red}{v}}$ can have inversely related scale:
→ $e^{\textcolor{blue}{u}}$ being too large and $e^{\textcolor{red}{v}}$ being too small
- The static screening test aims at locating two subsets of indices
 $(I, J) \in \{1, \dots, n\} \times \{1, \dots, m\}$

Static screening test

$$\mathcal{T}(I, J)$$

$$(\textcolor{blue}{u}, \textcolor{red}{v}) \in \mathcal{C}_{\alpha_{\textcolor{blue}{u}}}^n \times \mathcal{C}_{\alpha_{\textcolor{red}{v}}}^m \equiv \begin{cases} e^{\textcolor{blue}{u}_i} > \alpha_{\textcolor{blue}{u}} \text{ and } e^{\textcolor{red}{v}_j} > \alpha_{\textcolor{red}{v}}, \forall (i, j) \in I \times J, \\ e^{\textcolor{blue}{u}_{i'}} = \alpha_{\textcolor{blue}{u}} \text{ and } e^{\textcolor{red}{v}_{j'}} = \alpha_{\textcolor{red}{v}}, \forall (i', j') \in I^c \times J^c. \end{cases}$$

Static Screening Test $\mathcal{T}(I_{\varepsilon,\kappa}, J_{\varepsilon,\kappa})$

Let $(\boldsymbol{u}^*, \boldsymbol{v}^*)$ be optimal solution of $\mathcal{S}_n^{\text{ad}}(\boldsymbol{\mu}, \boldsymbol{\nu})$. Define

$$I_{\varepsilon, \kappa} = \left\{ i = 1, \dots, n : \mu_i \geq \frac{\varepsilon^2}{\kappa} r_i(K) \right\}$$

and $J_{\varepsilon, \kappa} = \{j = 1, \dots, m : \nu_j \geq \kappa \varepsilon^2 c_j(\mathbf{K})\}$.

$$e^{\textcolor{blue}{u}_i^*} = \frac{\varepsilon}{\kappa} \text{ and } e^{\textcolor{red}{v}_j^*} = \varepsilon \kappa \text{ for all } i \in I_{\varepsilon, \kappa}^{\complement} \text{ and } j \in J_{\varepsilon, \kappa}^{\complement}$$

- The parameters ε and κ are difficult to interpret, we exhibit their relations with a fixed number budget of points from the supports of μ and ν .

Screening with a Fixed Number Budget of Points

- We denote by $n_b \in \{1, \dots, n\}$ and $m_b \in \{1, \dots, m\}$ the number of points that are going to be optimized in $\mathcal{S}_\eta^{\text{ad}}(\boldsymbol{\mu}, \boldsymbol{\nu})$.
- Let $\xi \in \mathbb{R}^n$ and $\zeta \in \mathbb{R}^m$ to be the ordered decreasing vectors of $\boldsymbol{\mu} \oslash r(\mathbf{K})$ and $\boldsymbol{\nu} \oslash c(\mathbf{K})$
- To keep only n_b -budget and m_b -budget of points, the parameters

$$\varepsilon^2 \kappa = \zeta_{m_b}$$

and

$$\frac{\varepsilon^2}{\kappa} = \xi_{n_b}$$

Then

$$\varepsilon = (\xi_{n_b} \zeta_{m_b})^{1/4} \quad \text{and}$$

$$\kappa = \sqrt{\frac{\zeta_{m_b}}{\xi_{n_b}}}$$

- This guarantees

$$|I_{\varepsilon, \kappa}| = n_b$$

and

$$|J_{\varepsilon, \kappa}| = m_b$$

Screening with a Fixed Number Budget of Points

- If $(\mathbf{u}^*, \mathbf{v}^*)$ solution of $\mathcal{S}_{\eta}^{\text{ad}}(\boldsymbol{\mu}, \boldsymbol{\nu})$ satisfies $\mathcal{T}(I_{\varepsilon, \kappa}, J_{\varepsilon, \kappa})$ with
 $\alpha_{\mathbf{u}^*} = -\frac{\varepsilon}{\kappa}$ and $\alpha_{\mathbf{v}^*} = \varepsilon \kappa$
- We can restrict the variables in $\mathcal{S}_{\eta}^{\text{ad}}(\boldsymbol{\mu}, \boldsymbol{\nu})$ to variables in $I_{\varepsilon, \kappa}$ and $J_{\varepsilon, \kappa}$
- This boils down to restricting the constraints feasibility $\mathcal{C}_{\varepsilon}^n \cap \mathcal{C}_{\varepsilon \kappa}^m$ to the screened domain defined by $\mathcal{U}^{\text{sc}} \cap \mathcal{V}^{\text{sc}}$

Screened Feasibility Domain

$$\mathcal{U}^{\text{sc}} = \left\{ \mathbf{u} \in \mathbb{R}^{n_b} : e^{\mathbf{u}_i} \geq \frac{\varepsilon}{\kappa} \right\} \text{ and } \mathcal{V}^{\text{sc}} = \left\{ \mathbf{v} \in \mathbb{R}^{m_b} : e^{\mathbf{v}_j} \geq \varepsilon \kappa \right\}$$

Screening with a Fixed Number Budget of Points

- By replacing in $\mathcal{S}_\eta^{\text{ad}}(\boldsymbol{\mu}, \boldsymbol{\nu})$ the variables belonging to $(I_{\varepsilon, \kappa}^{\complement} \times J_{\varepsilon, \kappa}^{\complement})$ by $\frac{\varepsilon}{\kappa}$ and $\varepsilon \kappa$, we derive the screened dual of Sinkhorn divergence as

Screened Dual of Sinkhorn Divergence

$$\mathcal{S}_\eta^{\text{scd}}(\boldsymbol{\mu}, \boldsymbol{\nu}) = \min_{\boldsymbol{u} \in \mathcal{U}_{\text{sc}}, \boldsymbol{v} \in \mathcal{V}_{\text{sc}}} \left\{ \Psi_{\varepsilon, \kappa}(\boldsymbol{u}, \boldsymbol{v}) \right\}$$

where

$$\begin{aligned} \Psi_{\varepsilon, \kappa}(\boldsymbol{u}, \boldsymbol{v}) := & (e^{\boldsymbol{u}_{I_{\varepsilon, \kappa}}})^\top \mathbf{K}_{(I_{\varepsilon, \kappa}, J_{\varepsilon, \kappa})} e^{\boldsymbol{v}_{J_{\varepsilon, \kappa}}} + \varepsilon \kappa (e^{\boldsymbol{u}_{I_{\varepsilon, \kappa}}})^\top \mathbf{K}_{(I_{\varepsilon, \kappa}, J_{\varepsilon, \kappa}^0)} \mathbf{1}_{m_b} + \frac{\varepsilon}{\kappa} \mathbf{1}_{n_b}^\top \mathbf{K}_{(I_{\varepsilon, \kappa}^0, J_{\varepsilon, \kappa})} e^{\boldsymbol{v}_{J_{\varepsilon, \kappa}}} \\ & - \kappa \boldsymbol{\mu}_{I_{\varepsilon, \kappa}}^\top \boldsymbol{u}_{I_{\varepsilon, \kappa}} - \kappa^{-1} \boldsymbol{\nu}_{J_{\varepsilon, \kappa}}^\top \boldsymbol{v}_{J_{\varepsilon, \kappa}} + \Xi \end{aligned}$$

with

$$\Xi = \varepsilon^2 \sum_{i \in I_{\varepsilon, \kappa}^0, j \in J_{\varepsilon, \kappa}^0} \mathbf{K}_{ij} - \kappa \log(\varepsilon \kappa^{-1}) \sum_{i \in I_{\varepsilon, \kappa}^0} \boldsymbol{\mu}_i - \kappa^{-1} \log(\varepsilon \kappa) \sum_{j \in J_{\varepsilon, \kappa}^0} \boldsymbol{\nu}_j$$

L-BFGS-B: Box Constraints on $(\mathbf{u}^{\text{sc}}, \mathbf{v}^{\text{sc}})$

For calculating objective function of $\mathcal{S}_\eta^{\text{scd}}(\boldsymbol{\mu}, \boldsymbol{\nu})$ uses only the restricted parts

$$\mathbf{K} = \begin{bmatrix} \mathbf{K}_{(I_{\varepsilon, \kappa}, J_{\varepsilon, \kappa})} & \mathbf{K}_{(I_{\varepsilon, \kappa}, J_{\varepsilon, \kappa}^C)} \\ \mathbf{K}_{(I_{\varepsilon, \kappa}^C, J_{\varepsilon, \kappa})} & \end{bmatrix}$$

Proposition

Let $(\mathbf{u}^{\text{sc}}, \mathbf{v}^{\text{sc}})$ be an optimal pair solution of the screened dual $\mathcal{S}_\eta^{\text{scd}}(\boldsymbol{\mu}, \boldsymbol{\nu})$

Define $\mathbf{K}_{\min} = \min_{i \in I_{\varepsilon, \kappa}, j \in J_{\varepsilon, \kappa}} \mathbf{K}_{ij}$ Then,

$$\frac{\varepsilon}{\kappa} \vee \frac{\min_{i \in I_{\varepsilon, \kappa}} \boldsymbol{\mu}_i}{\varepsilon(m - m_b) + \varepsilon \vee \frac{\max_{j \in J_{\varepsilon, \kappa}} \boldsymbol{\nu}_j}{n\varepsilon\kappa} m_b} \leq e^{\mathbf{u}_i^{\text{sc}}} \leq \frac{\varepsilon}{\kappa} \vee \frac{\max_{i \in I_{\varepsilon, \kappa}} \boldsymbol{\mu}_i}{m\varepsilon \mathbf{K}_{\min}}$$

and

$$\varepsilon\kappa \vee \frac{\min_{j \in J_{\varepsilon, \kappa}} \boldsymbol{\nu}_j}{\varepsilon(n - n_b) + \varepsilon \vee \frac{\kappa \max_{i \in I_{\varepsilon, \kappa}} \boldsymbol{\mu}_i}{m\varepsilon \mathbf{K}_{\min}} n_b} \leq e^{\mathbf{v}_j^{\text{sc}}} \leq \varepsilon\kappa \vee \frac{\max_{j \in J_{\varepsilon, \kappa}} \boldsymbol{\nu}_j}{n\varepsilon \mathbf{K}_{\min}}$$

Screenkhorn

SCREENKHORN($\mathbf{C}, \eta, \boldsymbol{\mu}, \boldsymbol{\nu}, n_b, m_b$)

1. $\mathbf{K} \leftarrow e^{-\mathbf{C}/\eta};$
2. $\xi \leftarrow \text{sort}(\boldsymbol{\mu} \oslash r(\mathbf{K})), \zeta \leftarrow \text{sort}(\boldsymbol{\nu} \oslash c(\mathbf{K}));$ (decreasing order)
3. $\varepsilon \leftarrow (\xi_{n_b} \zeta_{m_b})^{1/4}, \kappa \leftarrow \sqrt{\zeta_{m_b} / \xi_{n_b}}$
4. $I_{\varepsilon, \kappa} \leftarrow \{i = 1, \dots, n : \boldsymbol{\mu}_i \geq \varepsilon^2 \kappa^{-1} r_i(\mathbf{K})\};$
5. $J_{\varepsilon, \kappa} \leftarrow \{j = 1, \dots, m : \boldsymbol{\nu}_j \geq \varepsilon^2 \kappa c_j(\mathbf{K})\};$
6. $\mathbf{K}_{\min} = \min_{I_{\varepsilon, \kappa}, J_{\varepsilon, \kappa}} \mathbf{K}_{ij}$ STEP I: SCREENING
7. $\underline{\boldsymbol{\mu}} \leftarrow \min_{i \in I_{\varepsilon, \kappa}} \boldsymbol{\mu}_i, \bar{\boldsymbol{\mu}} \leftarrow \max_{i \in I_{\varepsilon, \kappa}} \boldsymbol{\mu}_i; \quad \underline{\boldsymbol{\nu}} \leftarrow \min_{j \in J_{\varepsilon, \kappa}} \boldsymbol{\nu}_j, \bar{\boldsymbol{\nu}} \leftarrow \max_{j \in J_{\varepsilon, \kappa}} \boldsymbol{\nu}_j;$
8. $\underline{\mathbf{u}} \leftarrow \log\left(\frac{\varepsilon}{\kappa} \vee \frac{\underline{\boldsymbol{\mu}}}{\varepsilon(m - m_b) + \varepsilon \vee \frac{\bar{\boldsymbol{\nu}}}{n\varepsilon\kappa} m_b}\right), \bar{\mathbf{u}} \leftarrow \log\left(\frac{\varepsilon}{\kappa} \vee \frac{\bar{\boldsymbol{\mu}}}{m\varepsilon \mathbf{K}_{\min}}\right);$
9. $\underline{\mathbf{v}} \leftarrow \log\left(\varepsilon\kappa \vee \frac{\underline{\boldsymbol{\nu}}}{\varepsilon(n - n_b) + \varepsilon \vee \frac{\kappa\bar{\boldsymbol{\mu}}}{m\varepsilon} n_b}\right), \bar{\mathbf{v}} \leftarrow \log\left(\varepsilon\kappa \vee \frac{\bar{\boldsymbol{\nu}}}{n\varepsilon \mathbf{K}_{\min}}\right);$
10. $\bar{\boldsymbol{\theta}} \leftarrow \text{stack}(\bar{\mathbf{u}} \mathbf{1}_{n_b}, \bar{\mathbf{v}} \mathbf{1}_{m_b}), \underline{\boldsymbol{\theta}} \leftarrow \text{stack}(\underline{\mathbf{u}} \mathbf{1}_{n_b}, \underline{\mathbf{v}} \mathbf{1}_{m_b});$

11. $\mathbf{u}^{(0)} \leftarrow \log(\varepsilon\kappa^{-1}) \mathbf{1}_{n_b}, \mathbf{v}^{(0)} \leftarrow \log(\varepsilon\kappa) \mathbf{1}_{m_b};$
12. $\boldsymbol{\theta}^{(0)} \leftarrow \text{stack}(\mathbf{u}^{(0)}, \mathbf{v}^{(0)});$
13. $\boldsymbol{\theta} \leftarrow \text{L-BFGS-B}(\boldsymbol{\theta}^{(0)}, \underline{\boldsymbol{\theta}}, \bar{\boldsymbol{\theta}});$
14. $\boldsymbol{\theta}_u \leftarrow (\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_{n_b})^\top;$ STEP 2: L-BFGS-B
15. $\boldsymbol{\theta}_v \leftarrow (\boldsymbol{\theta}_{n_b+1}, \dots, \boldsymbol{\theta}_{n_b+m_b})^\top;$
16. $\mathbf{u}_i^{\text{sc}} \leftarrow (\boldsymbol{\theta}_u)_i \text{ if } i \in I_{\varepsilon, \kappa} \text{ and } \mathbf{u}_i^{\text{sc}} \leftarrow \log(\varepsilon\kappa^{-1}) \text{ if } i \in I_{\varepsilon, \kappa}^c;$
17. $\mathbf{v}_j^{\text{sc}} \leftarrow (\boldsymbol{\theta}_v)_j \text{ if } j \in J_{\varepsilon, \kappa} \text{ and } \mathbf{v}_j^{\text{sc}} \leftarrow \log(\varepsilon\kappa) \text{ if } j \in J_{\varepsilon, \kappa}^c;$
18. Return $B(\mathbf{u}^{\text{sc}}, \mathbf{v}^{\text{sc}});$

Theoretical Guarantees

21

21

Theoretical Analysis and Guarantees

Screened optimal transport plan $\mathbf{P}^{\text{sc}} = \text{diag}(e^{\mathbf{u}^{\text{sc}}}) \mathbf{K} \text{diag}(e^{\mathbf{v}^{\text{sc}}})$

Screened marginals $\boldsymbol{\mu}^{\text{sc}} = \mathbf{P}^{\text{sc}} \mathbf{1}_m$ and $\boldsymbol{\nu}^{\text{sc}} = (\mathbf{P}^{\text{sc}})^\top \mathbf{1}_n$

Proposition

$$\|\boldsymbol{\mu} - \boldsymbol{\mu}^{\text{sc}}\|_1^2 = \mathcal{O}\left(n_b c_\kappa + (n - n_b) \left(\frac{\|C\|_\infty}{\eta} + \frac{m_b}{\sqrt{nmc_{\boldsymbol{\mu}\boldsymbol{\nu}}} \mathbf{K}_{\min}^{3/2}} + \frac{m - m_b}{\sqrt{nm}} \mathbf{K}_{\min} + \log\left(\frac{\sqrt{nm}}{m_b c_{\boldsymbol{\mu}\boldsymbol{\nu}}^{5/2}}\right) \right)\right)$$

$$\|\boldsymbol{\nu} - \boldsymbol{\nu}^{\text{sc}}\|_1^2 = \mathcal{O}\left(m_b c_{\frac{1}{\kappa}} + (m - m_b) \left(\frac{\|C\|_\infty}{\eta} + \frac{n_b}{\sqrt{nmc_{\boldsymbol{\mu}\boldsymbol{\nu}}} \mathbf{K}_{\min}^{3/2}} + \frac{n - n_b}{\sqrt{nm}} \mathbf{K}_{\min} + \log\left(\frac{\sqrt{nm}}{n_b c_{\boldsymbol{\mu}\boldsymbol{\nu}}^{5/2}}\right) \right)\right)$$

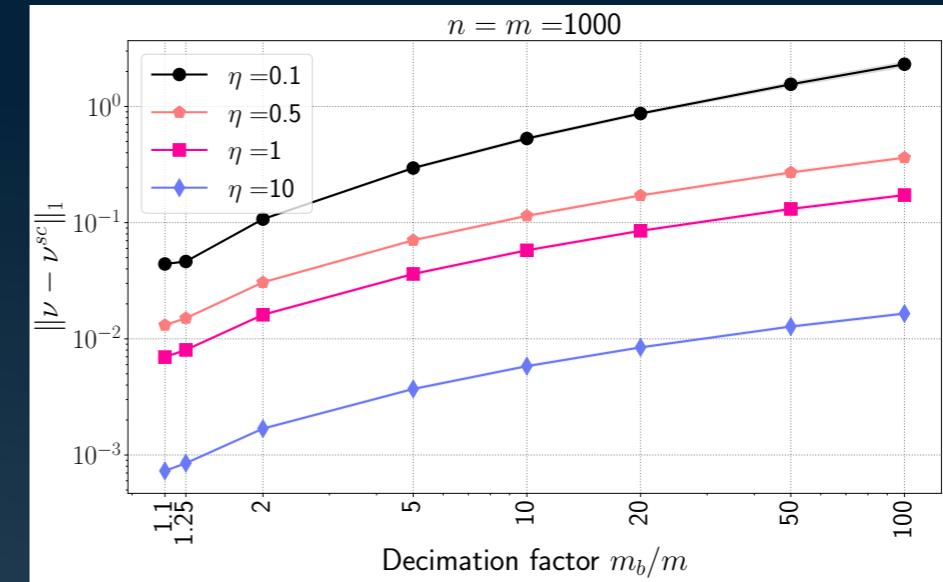
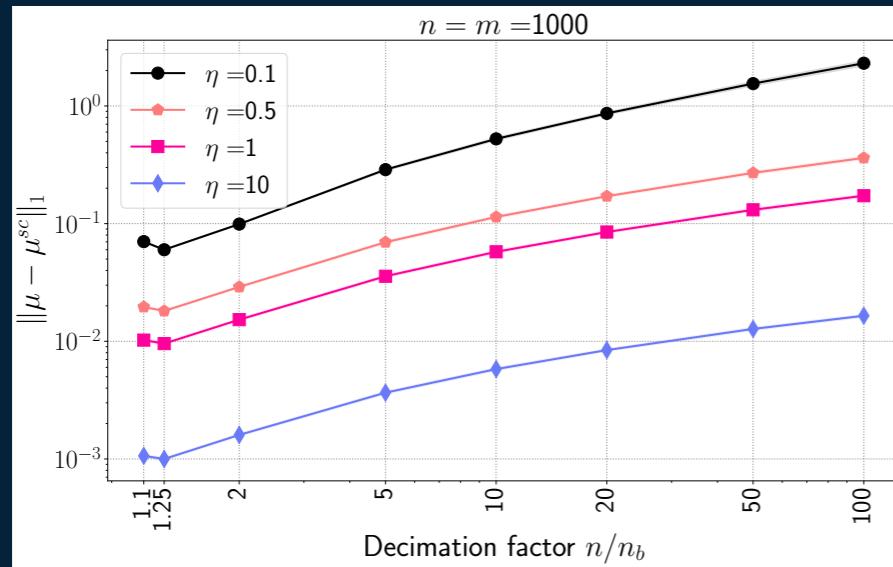
$$c_z = z - \log z - 1 \text{ for } z > 0, c_{\boldsymbol{\mu}\boldsymbol{\nu}} = \underline{\boldsymbol{\mu}} \wedge \underline{\boldsymbol{\nu}} \text{ with } \underline{\boldsymbol{\mu}} = \min_{i \in I_{\varepsilon, \kappa}} \boldsymbol{\mu}_i, \quad \underline{\boldsymbol{\nu}} = \min_{j \in J_{\varepsilon, \kappa}} \boldsymbol{\nu}_j.$$

Proposition

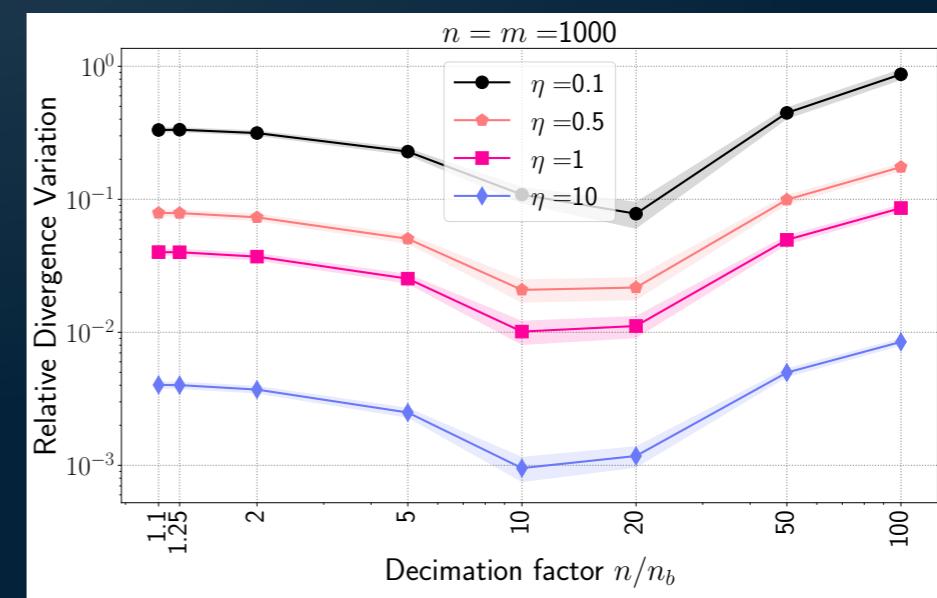
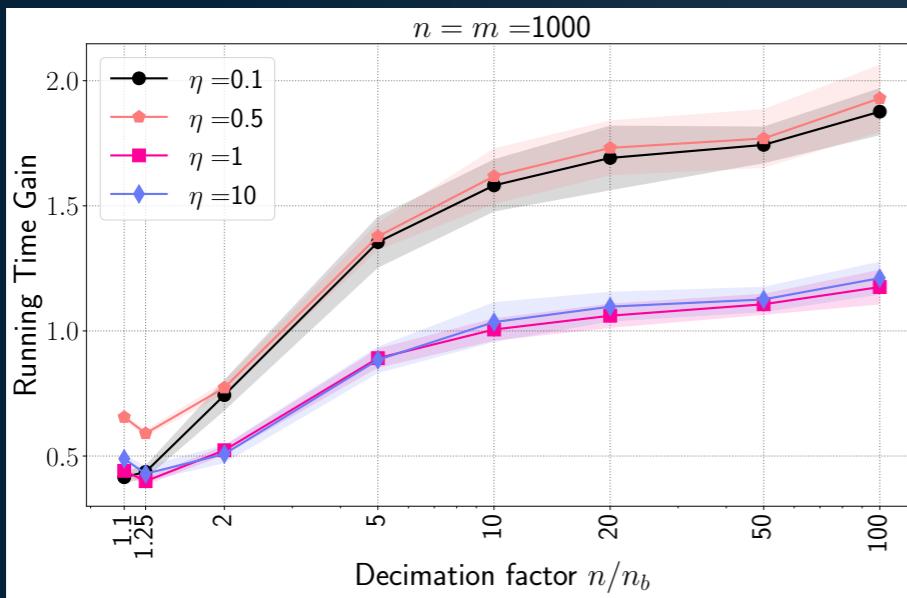
$$\Psi_{\varepsilon, \kappa}(\mathbf{u}^{\text{sc}}, \mathbf{v}^{\text{sc}}) - \Psi(\mathbf{u}^*, \mathbf{v}^*) = \mathcal{O}\left(R(\|\boldsymbol{\mu} - \boldsymbol{\mu}^{\text{sc}}\|_1 + \|\boldsymbol{\nu} - \boldsymbol{\nu}^{\text{sc}}\|_1 + \omega_\kappa)\right)$$

$$R = \frac{\|C\|_\infty}{\eta} + \log\left(\frac{(n \vee m)^2}{nmc_{\boldsymbol{\mu}\boldsymbol{\nu}}^{7/2}}\right) \text{ and } \omega_\kappa = |1 - \kappa| \|\boldsymbol{\mu}^{\text{sc}}\|_1 + |1 - \kappa^{-1}| \|\boldsymbol{\nu}^{\text{sc}}\|_1 + |1 - \kappa| + |1 - \kappa^{-1}|$$

Theoretical Analysis and Guarantees: (Simulation on Toy Data)



Marginal violations in relation with the budget of points

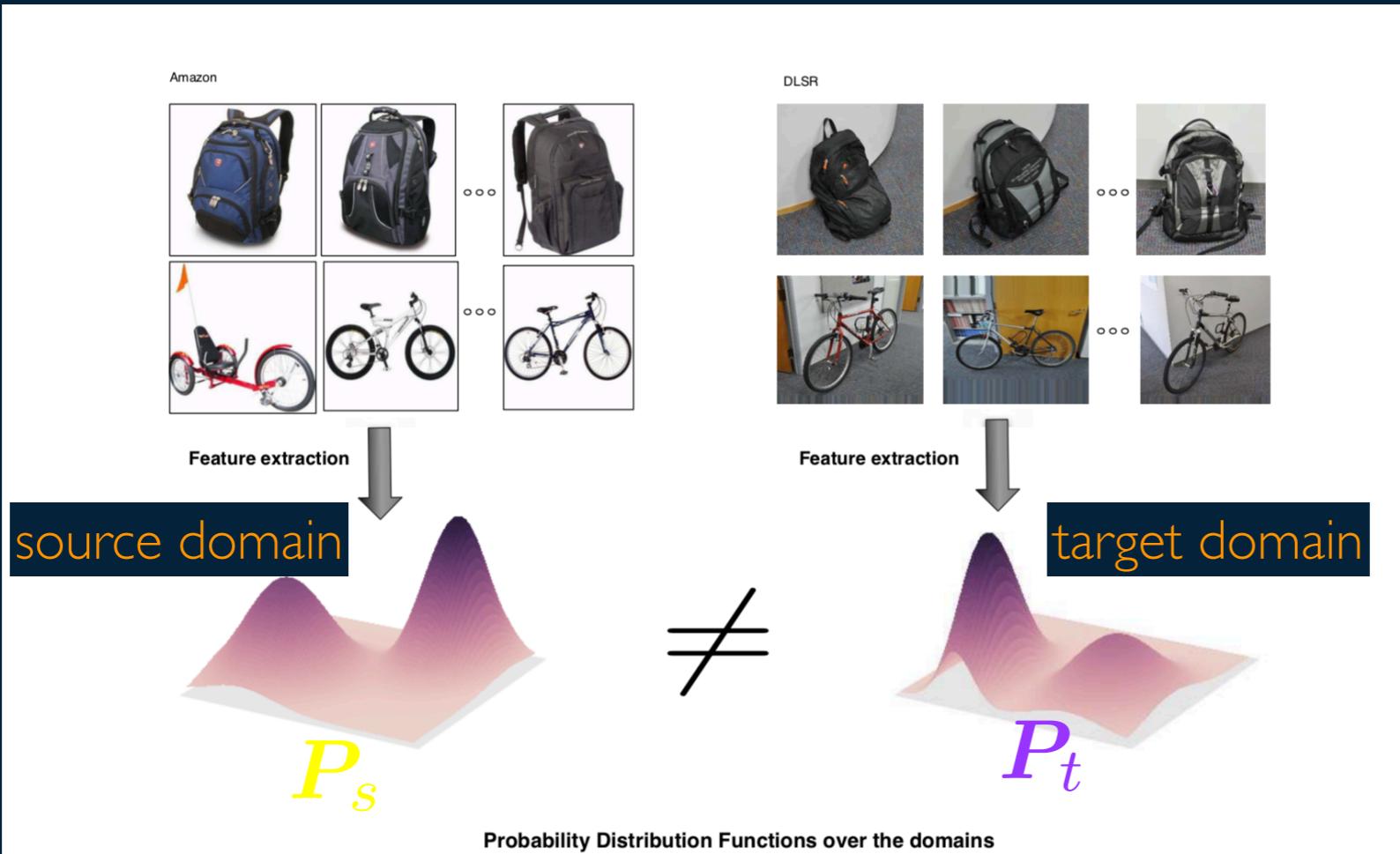


$$\frac{T_{\text{SINKHORN}}}{T_{\text{SCREENKHORN}}}$$

$$\frac{|\langle \mathcal{C}, P^* \rangle - \langle \mathcal{C}, P^{\text{SC}} \rangle|}{\langle \mathcal{C}, P^* \rangle}$$

Integrating Screenkhorn into ML Pipeline: Unsupervised Domain Adaptation

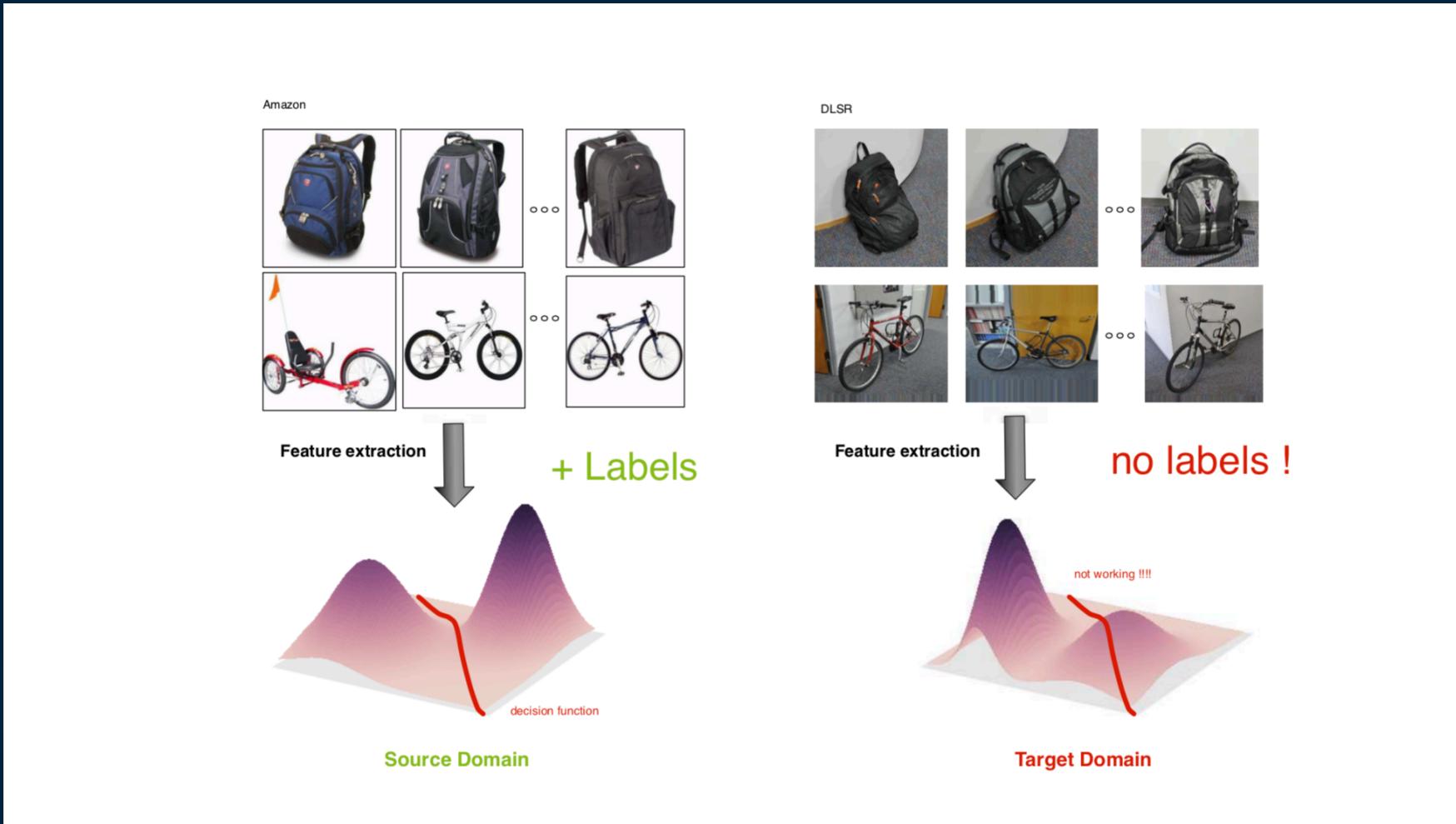
Unsupervised Domain Adaptation



(Credit image: N. Courty)

- Traditional machine learning hypothesis:
 - We have access to training data. Probability distribution of the training set and the testing are the same.
 - We want to learn a classifier that generalizes to new data.

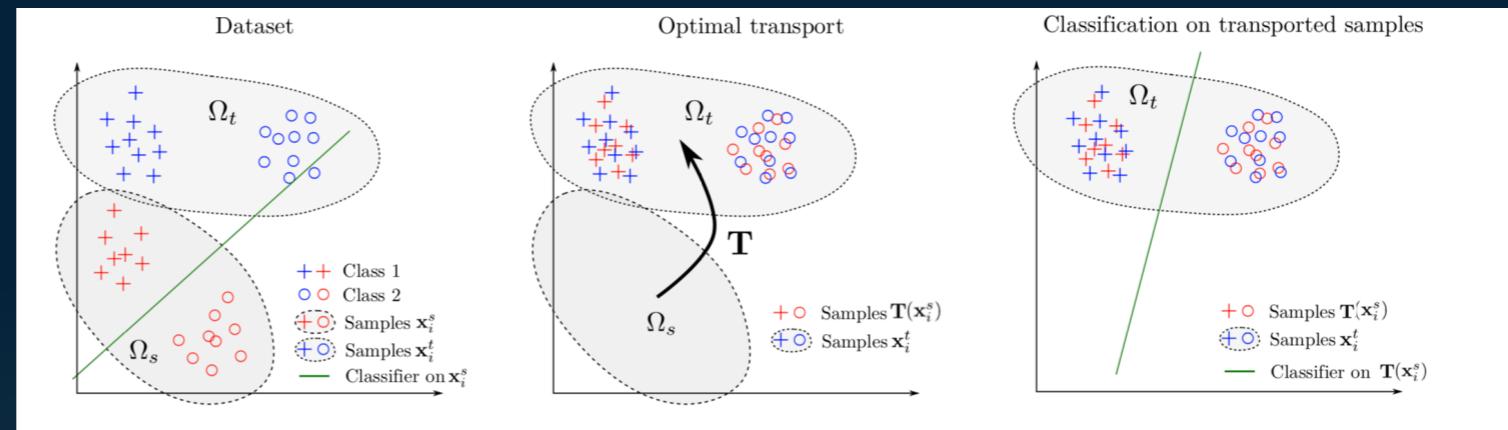
Unsupervised Domain Adaptation



(Credit image: N. Courty)

- Domain adaptation: classification problem with data coming from different sources (domains).
- Labels only available in the source domain, and classification is conducted in the target domain.
- Classifier trained on the source domain data performs badly in the target domain.

OTDA: Optimal Transport Domain Adaptation



Assumptions

(Courty et al., 2017)

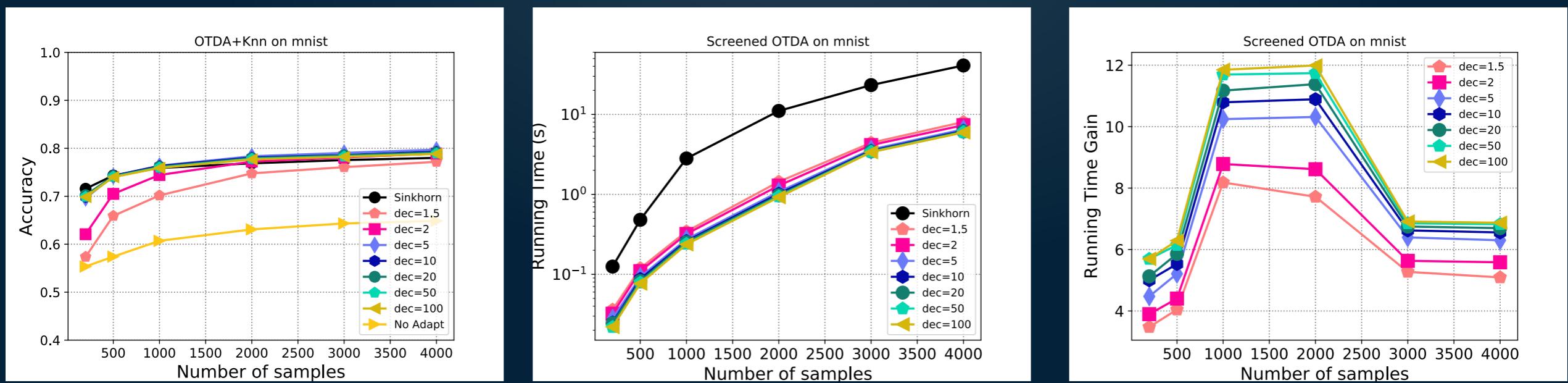
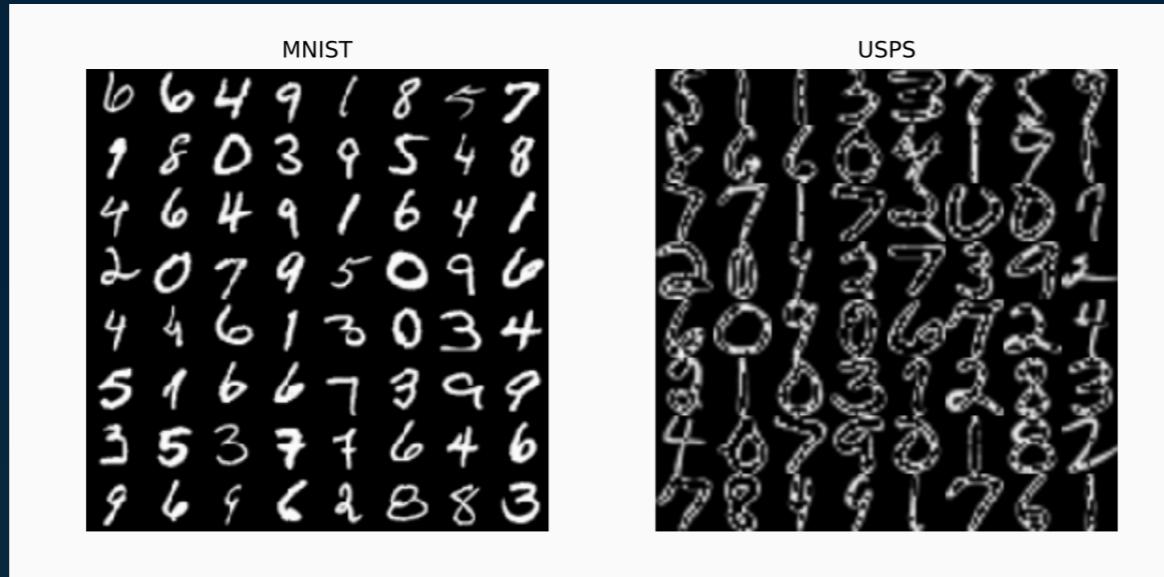
- There exist a transport \mathbf{T} in the feature space between the two domains.
- The transport preserves the conditional distributions:

$$P_s[\mathbf{y}|\mathbf{x}_s] = P_t[\mathbf{y}|\mathbf{T}(\mathbf{x}_s)]$$

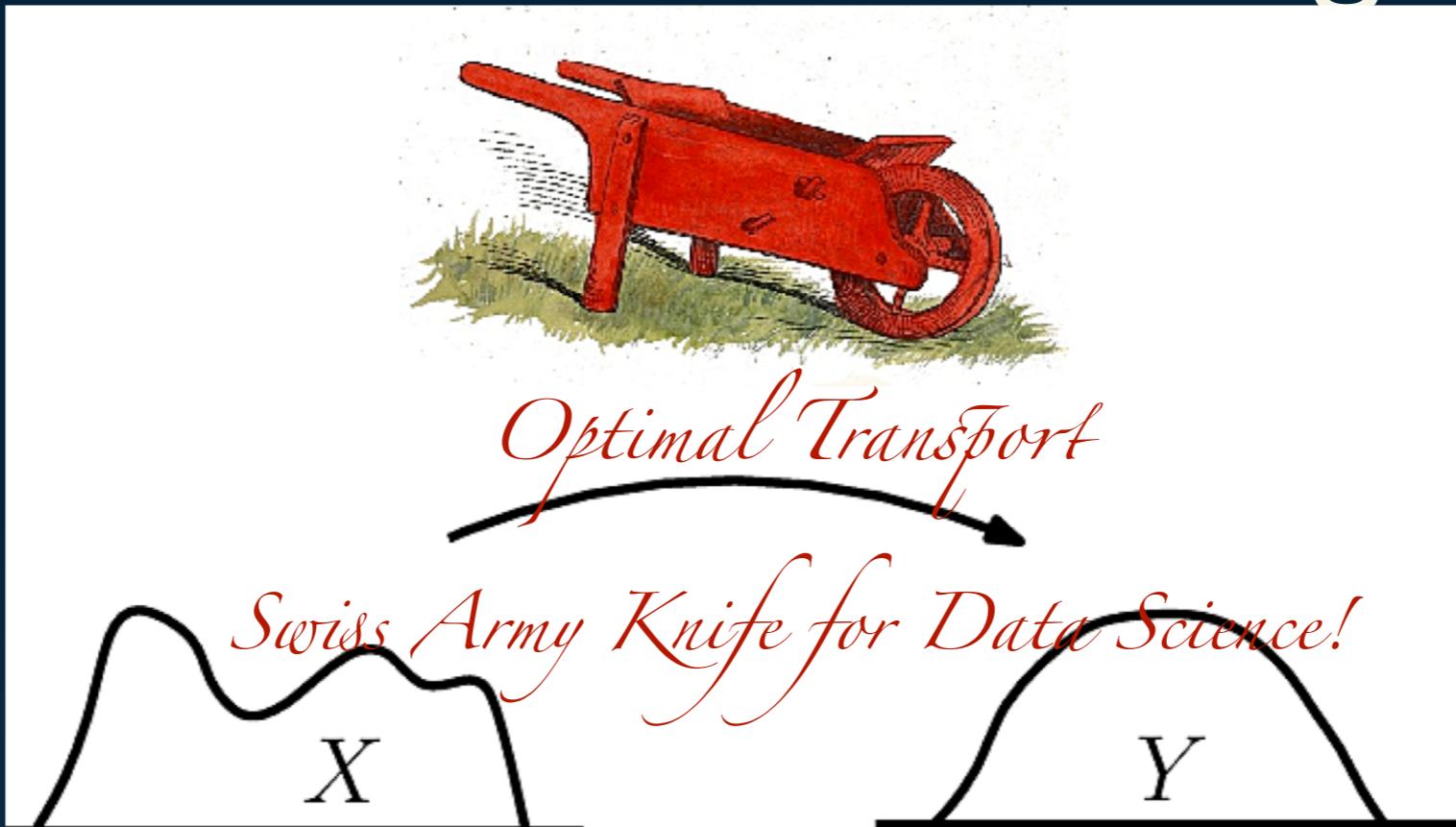
3-step strategy

1. Estimate optimal transport between distributions.
2. Transport the training samples onto the target distribution using barycentric mapping (Ferradans et al., 2013)
3. Learn a classifier on the transported training samples.

Real Dataset for OTDA: MNIST(source) to USPS (target)



Take Home Message



(Credit image: P. Lemberger)

- We introduce a novel approach for approximating the Sinkhorn divergence based on a screening strategy with a carefully analyzing its optimality conditions.
- Integrated in some complex machine learning pipelines, Screenkhorn algorithm achieves strong gain in efficiency while not compromising on accu

Thank You!