GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN

Empirical Regularized Optimal Transport: Statistical Theory and Applications

DAGStat Conference 2019: Statistics under one umbrella

Marcel Klatt   Carla Tameling   Axel Munk

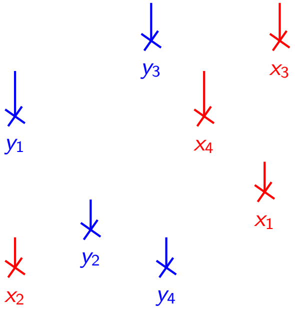
March 19, 2019

Institute for Mathematical Stochastics
University of Göttingen

References:

M. Klatt, C. Tameling and A. Munk *Empirical Regularized Optimal Transport: Statistical Theory and Applications*, Journal of the Royal Statistical Society: Series B (Statistical Methodology), to appear, 2019.

(Regularized) Optimal Transport in a Nutshell



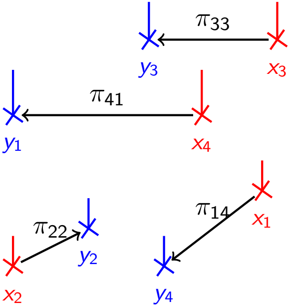
Finite discrete metric space $(\mathcal{X}, d)$

- Given two probability measures

$$r = \sum_{i=1}^{N} r_i \delta_{x_i}, \quad s = \sum_{j=1}^{N} s_j \delta_{y_j}$$

- Transport costs $d^p(x_i, y_j)$ for $p \geq 1$
- Find the most efficient way to transport measure r into s.

(Regularized) Optimal Transport in a Nutshell



Transport plan (coupling of r and s):

$$\pi = \begin{pmatrix} & & & \pi_{14} \\ & \pi_{22} & & \\ & & \pi_{33} & \\ \pi_{41} & & & \end{pmatrix}$$

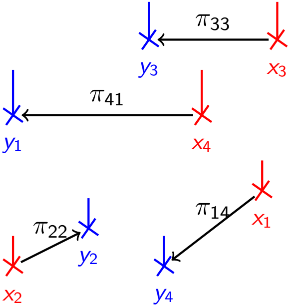
Finite discrete metric space $(\mathcal{X}, d)$

- Given two probability measures

$$r = \sum_{i=1}^{N} r_i \delta_{x_i}, \quad s = \sum_{j=1}^{N} s_j \delta_{y_j}$$

- Transport costs $d^p(x_i, y_j)$ for $p \geq 1$
- Find the most efficient way to transport measure r into s.

(Regularized) Optimal Transport in a Nutshell



Transport plan (coupling of r and s):

$$\pi = \begin{pmatrix} & & & \pi_{14} \\ & \pi_{22} & & \\ & & \pi_{33} & \\ \pi_{41} & & & \end{pmatrix}$$

Finite discrete metric space $(\mathcal{X}, d)$

- Given two probability measures

$$r = \sum_{i=1}^{N} r_i \delta_{x_i}, \quad s = \sum_{j=1}^{N} s_j \delta_{y_j}$$

- Transport costs $d^p(x_i, y_j)$ for $p \geq 1$
- Find the most efficient way to transport measure r into s.
- Solve the standard linear program

$$\min_{\pi \in \mathbb{R}^{N \times N}} \sum_{i,j=1}^{N} d^p(x_i, y_j)\pi_{ij}$$
$$\text{s.t.} \quad \pi \mathbb{1}_N = r, \ \pi^T \mathbb{1}_N = s$$
$$\pi \geq 0 \,.$$

The **Wasserstein distance** of order $p \geq 1$ is defined as

$$W_p(r, s) := \left\{ \min_{\pi \in \Pi(r,s)} \sum_{i,j=1}^{N} d^p(x_i, y_j)\pi_{ij} \right\}^{1/p} .$$

(Regularized) Optimal Transport in a Nutshell

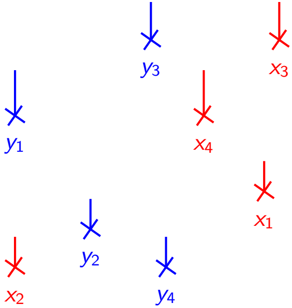
The **Wasserstein distance** of order $p \geq 1$ is defined as

$$W_p(r, s) := \left\{ \min_{\pi \in \Pi(r,s)} \sum_{i,j=1}^{N} d^p(x_i, y_j)\pi_{ij} \right\}^{1/p}.$$

In general, the computational cost to calculate the Wasserstein distance is of order $\mathcal{O}(N^3 \log(N))$. There exist some workarounds and we focus on:

- **Regularization methods** (*Cuturi (2013), Dessein et al. (2016)*)

(Regularized) Optimal Transport in a Nutshell



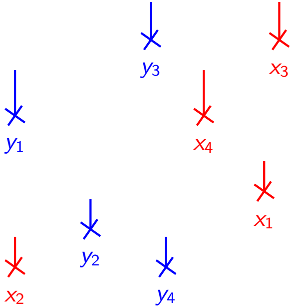
Finite discrete metric space $(\mathcal{X}, d)$

- Given two probability measures

$$r = \sum_{i=1}^{N} r_i \delta_{x_i}, \quad s = \sum_{j=1}^{N} s_j \delta_{y_j}$$

- Transport costs $d^p(x_i, y_j)$ for $p \geq 1$
- Entropy: $E(\pi) = -\sum_{i,j=1}^{N} \pi_{ij} \log(\pi_{ij})$

(Regularized) Optimal Transport in a Nutshell



Finite discrete metric space $(\mathcal{X}, d)$

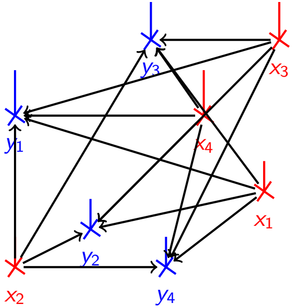
- Given two probability measures

$$r = \sum_{i=1}^{N} r_i \delta_{x_i}, \quad s = \sum_{j=1}^{N} s_j \delta_{y_j}$$

- Transport costs $d^p(x_i, y_j)$ for $p \geq 1$
- Entropy: $E(\pi) = -\sum_{i,j=1}^{N} \pi_{ij} \log(\pi_{ij})$
- For $\lambda > 0$ find the (*unique*) entropy regularized transport plan

$$\pi_\lambda(r, s) = \underset{\pi \in \Pi(r,s)}{\operatorname{argmin}} \sum_{i,j=1}^{N} d^p(x_i, y_j)\pi_{ij} - \lambda E(\pi)\,.$$

(Regularized) Optimal Transport in a Nutshell



λ large

(Regularized) Transport plan:

$$\pi_\lambda(r,s) \approx r \otimes s = rs^T$$

Finite discrete metric space $(\mathcal{X}, d)$

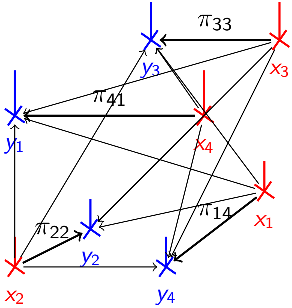
- Given two probability measures

$$r = \sum_{i=1}^{N} r_i \delta_{x_i}, \quad s = \sum_{j=1}^{N} s_j \delta_{y_j}$$

- Transport costs $d^p(x_i, y_j)$ for $p \geq 1$
- Entropy: $E(\pi) = -\sum_{i,j=1}^{N} \pi_{ij} \log(\pi_{ij})$
- For $\lambda > 0$ find the (*unique*) entropy regularized transport plan

$$\pi_\lambda(r,s) = \underset{\pi \in \Pi(r,s)}{\operatorname{argmin}} \sum_{i,j=1}^{N} d^p(x_i, y_j)\pi_{ij} - \lambda E(\pi)\,.$$

(Regularized) Optimal Transport in a Nutshell



λ intermediate

(Regularized) Transport plan:

$$\pi_\lambda(r,s) = \begin{pmatrix} \pi_{11} & \pi_{12} & \pi_{13} & \pi_{14} \\ \pi_{21} & \pi_{22} & \pi_{23} & \pi_{24} \\ \pi_{31} & \pi_{32} & \pi_{33} & \pi_{34} \\ \pi_{41} & \pi_{42} & \pi_{43} & \pi_{44} \end{pmatrix}$$

Finite discrete metric space $(\mathcal{X}, d)$

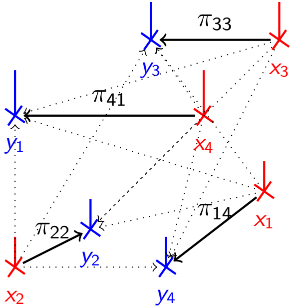
- Given two probability measures

$$r = \sum_{i=1}^{N} r_i \delta_{x_i}, \quad s = \sum_{j=1}^{N} s_j \delta_{y_j}$$

- Transport costs $d^p(x_i, y_j)$ for $p \geq 1$
- Entropy: $E(\pi) = -\sum_{i,j=1}^{N} \pi_{ij} \log(\pi_{ij})$
- For $\lambda > 0$ find the (*unique*) entropy regularized transport plan

$$\pi_\lambda(r,s) = \underset{\pi \in \Pi(r,s)}{\operatorname{argmin}} \sum_{i,j=1}^{N} d^p(x_i, y_j)\pi_{ij} - \lambda E(\pi)\,.$$

(Regularized) Optimal Transport in a Nutshell



λ small

(Regularized) Transport plan:

$$\pi_\lambda(r,s) = \begin{pmatrix} \pi_{11} & \pi_{12} & \pi_{13} & \pi_{14} \\ \pi_{21} & \pi_{22} & \pi_{23} & \pi_{24} \\ \pi_{31} & \pi_{32} & \pi_{33} & \pi_{34} \\ \pi_{41} & \pi_{42} & \pi_{43} & \pi_{44} \end{pmatrix}$$

Finite discrete metric space $(\mathcal{X}, d)$

- Given two probability measures

$$r = \sum_{i=1}^{N} r_i \delta_{x_i}, \quad s = \sum_{j=1}^{N} s_j \delta_{y_j}$$

- Transport costs $d^p(x_i, y_j)$ for $p \geq 1$
- Entropy: $E(\pi) = -\sum_{i,j=1}^{N} \pi_{ij} \log(\pi_{ij})$
- For $\lambda > 0$ find the (*unique*) entropy regularized transport plan

$$\pi_\lambda(r,s) = \operatorname*{argmin}_{\pi \in \Pi(r,s)} \sum_{i,j=1}^{N} d^p(x_i, y_j)\pi_{ij} - \lambda E(\pi).$$

(Regularized) Optimal Transport in a Nutshell

The **Sinkhorn divergence** of order $p \geq 1$ is defined as

$$W_{p,\lambda}(r,s) := \left\{ \sum_{i,j=1}^{N} d^p(x_i, y_j)\pi_\lambda(r,s)_{ij} \right\}^{1/p} .$$

(Regularized) Optimal Transport in a Nutshell

The **Sinkhorn divergence** of order $p \geq 1$ is defined as

$$W_{p,\lambda}(r,s) := \left\{ \sum_{i,j=1}^{N} d^p(x_i, y_j)\pi_\lambda(r,s)_{ij} \right\}^{1/p} .$$

- Efficient matrix scaling algorithm (**Sinkhorn algorithm**) to approximate $\pi_\lambda(r,s)$ (*Cuturi (2013)*)
- Rates for approximation (*Luise et al. (2018)*):

$$\|\pi_\lambda(r,s) - \pi(r,s)\| \leq C_1 \exp\left(-\frac{1}{\lambda}\right) ,$$

where $\pi(r,s) = \underset{\pi \in \Pi(r,s)}{\operatorname{argmax}}\{E(\pi),\ W_p^p(r,s) = \langle \pi, d^p(x,y)\rangle\}$ and

$$|W_{p,\lambda}(r,s) - W_p(r,s)| \leq C_2 \exp\left(-\frac{1}{\lambda}\right) .$$

Let $\mathcal{X} = \{x_1, \dots, x_N\}$ be a finite space with metric $d : \mathcal{X} \times \mathcal{X} \to \mathbb{R}_+$.
Assume, we only have access to the measure r through its corresponding empirical version

$$\hat{r}_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}$$

derived by a sample of $\mathcal{X}$-valued random variables $X_1, \dots, X_n \overset{i.i.d.}{\sim} r$.

Statistical Framework

Let $\mathcal{X} = \{x_1, \dots, x_N\}$ be a finite space with metric $d : \mathcal{X} \times \mathcal{X} \to \mathbb{R}_+$. Assume, we only have access to the measure r through its corresponding empirical version

$$\hat{r}_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}$$

derived by a sample of $\mathcal{X}$-valued random variables $X_1, \dots, X_n \overset{i.i.d.}{\sim} r$.

Central question:

- How do the random quantities $\pi_\lambda(\hat{r}_n, s)$ and $W_{p,\lambda}(\hat{r}_n, s)$ relate to $\pi_\lambda(r, s)$ and $W_{p,\lambda}(r, s)$, respectively?

Limit Law for Regularized Transport Plan

The empirical regularized transport plan is defined as

$$\pi_\lambda(\hat{r}_n, s) = \underset{\pi \in \Pi(\hat{r}_n, s)}{\arg\min} \sum_{i,j=1}^{N} d^p(x_i, x_j)\pi_{ij} - \lambda E(\pi)\,.$$

Limit Law for Regularized Transport Plan

The empirical regularized transport plan is defined as

$$\pi_\lambda(\hat{r}_n, s) = \underset{\pi \in \Pi(\hat{r}_n, s)}{\arg\min} \sum_{i,j=1}^{N} d^p(x_i, x_j)\pi_{ij} - \lambda E(\pi)\,.$$

Theorem (K., Tameling & Munk (2019))

With the sample size n approaching infinity, it holds for $r = s$ and $r \neq s$ that

$$\sqrt{n}\,\{\pi_\lambda(\hat{r}_n, s) - \pi_\lambda(r, s)\} \xrightarrow{\mathfrak{D}} \mathcal{N}_{N^2}(0, \Sigma_\lambda(r|s))\,.$$

Limit Law for Regularized Transport Plan

The empirical regularized transport plan is defined as

$$\pi_\lambda(\hat{r}_n, s) = \underset{\pi \in \Pi(\hat{r}_n, s)}{\arg\min} \sum_{i,j=1}^{N} d^p(x_i, x_j)\pi_{ij} - \lambda E(\pi)\,.$$

Theorem (K., Tameling & Munk (2019))

With the sample size n approaching infinity, it holds for $r = s$ and $r \neq s$ that

$$\sqrt{n}\,\{\pi_\lambda(\hat{r}_n, s) - \pi_\lambda(r, s)\} \xrightarrow{\mathfrak{D}} \mathcal{N}_{N^2}(0, \Sigma_\lambda(r|s))\,.$$

Remark

Limit distributions for the (non-regularized) transport plan ($\lambda = 0$) are not known.

Limit Law for Sinkhorn Divergence

The empirical Sinkhorn divergence is defined as

$$W_{p,\lambda}(\hat{r}_n, s) := \left\{ \sum_{i,j=1}^{N} d^p(x_i, x_j) \pi_\lambda(\hat{r}_n, s)_{ij} \right\}^{1/p} .$$

Limit Law for Sinkhorn Divergence

The empirical Sinkhorn divergence is defined as

$$W_{p,\lambda}(\hat{r}_n, s) := \left\{ \sum_{i,j=1}^{N} d^p(x_i, x_j) \pi_\lambda(\hat{r}_n, s)_{ij} \right\}^{1/p} .$$

Theorem (K., Tameling & Munk (2019))

With the sample size n approaching infinity, it holds for $r = s$ and $r \neq s$ that

$$\sqrt{n}\,\{W_{p,\lambda}(\hat{r}_n, s) - W_{p,\lambda}(r, s)\} \xrightarrow{\mathfrak{D}} \mathcal{N}_1(0, \sigma^2_\lambda(r|s))\,.$$

Limit Law for Sinkhorn Divergence

The empirical Sinkhorn divergence is defined as

$$W_{p,\lambda}(\hat{r}_n, s) := \left\{ \sum_{i,j=1}^{N} d^p(x_i, x_j)\pi_\lambda(\hat{r}_n, s)_{ij} \right\}^{1/p} .$$

Theorem (K., Tameling & Munk (2019))

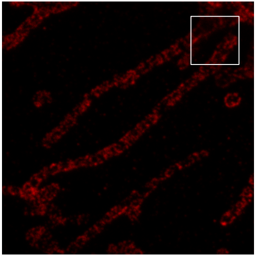
With the sample size n approaching infinity, it holds for $r = s$ and $r \neq s$ that

$$\sqrt{n}\,\{W_{p,\lambda}(\hat{r}_n, s) - W_{p,\lambda}(r, s)\} \xrightarrow{\mathfrak{D}} \mathcal{N}_1(0, \sigma_\lambda^2(r|s)) .$$

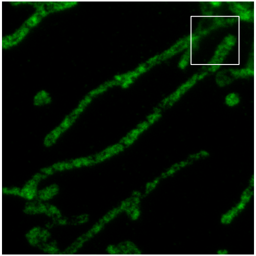
Remark

Limit laws for Wasserstein (Sommerfeld & Munk (2018)) are in general not Gaussian, e.g.

$$n^{1/2p}\, W_p(\hat{r}_n, r) \xrightarrow{\mathfrak{D}} \left\{ \max_{u \in \Phi^*} \langle G, u \rangle \right\}^{1/p} .$$



(a) ATP Synthase



(b) MIC60

Figure 1: Staining of two different proteins (*Jakobs lab, Department of NanoBiophotonics, Max-Planck Institute for Biophysical Chemistry, Göttingen*)

Aim: Analyse the interaction between fluorescently-labeled molecules by quantifying the co-occurrence and correlation between them.

Application: Colocalization Analysis

Conventional methods:

- **Pixel based** intensity correlation analysis (Pearson's correlation coefficient) or co-occurence (Manders' split coefficients)

Application: Colocalization Analysis

Conventional methods:

- **Pixel based** intensity correlation analysis (Pearson's correlation coefficient) or co-occurence (Manders' split coefficients)
- These methods are very sensitive to the resolution of the images to be compared.

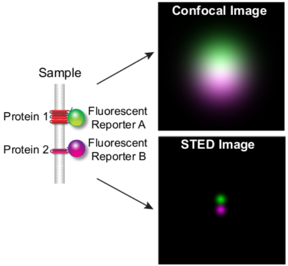


Figure 2: Simulation of Confocal and STED images of two proteins which are located at a distance of 45 nm. The resolution of the confocal image is 244 nm and for the STED image it is 40 nm.

Application: Colocalization Analysis

Our approach:

- Colocalization measure **RCol** based on (regularized) optimal transport for $t \in [0, \mathrm{diam}(\mathcal{X})]$

$$\mathrm{RCol}(\pi_\lambda(r,s))(t) = \sum_{i,j=1}^{N} \pi_\lambda(r,s)_{ij} \mathbb{1}\{d^p(x_i,x_j) \le t\}\,.$$

Our approach:

- Colocalization measure **RCol** based on (regularized) optimal transport for $t \in [0, \mathrm{diam}(\mathcal{X})]$

$$\mathrm{RCol}(\pi_\lambda(r,s))(t) = \sum_{i,j=1}^{N} \pi_\lambda(r,s)_{ij} \mathbb{1}\{d^p(x_i,x_j) \le t\}\,.$$

Theorem (K., Tameling & Munk (2019))

Let $\widehat{RCol}_n := RCol(\pi_\lambda(\hat{r}_n, s))$ be the empirical regularized colocalization. As $n \to \infty$ it holds that

$$\sqrt{n}\left\{\widehat{RCol}_n - RCol\right\} \xrightarrow{\mathfrak{D}} RCol(G)$$

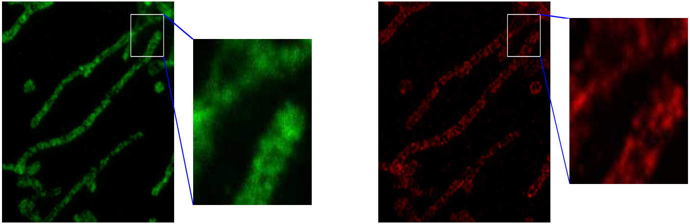
with G the random variable with distribution given by the limit law for the empirical regularized transport plan.

- This yields $1-\alpha$ **approximate uniform confidence bands**, i.e. $\lim_{n\to\infty}\mathbb{P}(\mathsf{RCol}\in\mathcal{I}_n)=1-\alpha$, where

$$\mathcal{I}_n := \left[-\frac{\mathfrak{u}_{1-\alpha}}{\sqrt{n}}+\widehat{\mathsf{RCol}}_n,\ \frac{\mathfrak{u}_{1-\alpha}}{\sqrt{n}}+\widehat{\mathsf{RCol}}_n\right]$$

  and $\mathfrak{u}_{1-\alpha}$ is the $1-\alpha$ quantile from the distribution of $\|\mathsf{RCol}(G)\|_\infty$.
- The quantile $\mathfrak{u}_{1-\alpha}$ can be consistently approximated by its *n* **out of *n* bootstrap** analogue.

Application: Colocalization Analysis



(a) ATP Synthase (b) MIC60

Figure 3: Staining of two different proteins: Image size 666×666 pixels, pixel size = 15nm. Middle: Zoom ins (128×128 pixels).

Application: Colocalization Analysis

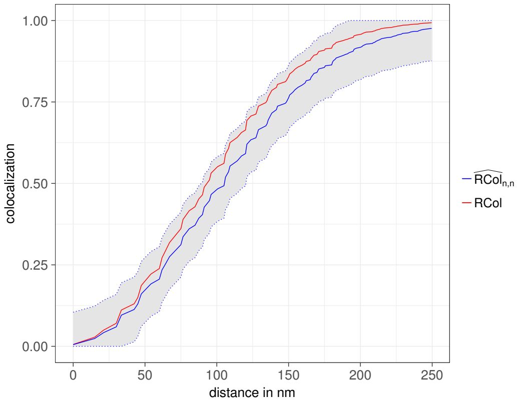


Figure 4: Staining of ATP Synthase and MIC60 for the zoom ins (128×128 images). The sampled regularized colocalization ($\lambda = 0.01$) (solid blue line, subsampling $n = 2000$) with bootstrap confidence bands (gray area between dashed blue lines) based on the n out of n bootstrap with $B = 100$ replications and $\alpha = 0.05$. Red solid line: Population regularized colocalization.

Summary

- **Limit laws** for the empirical regularized optimal transport plan and its corresponding Sinkhorn divergence.
- Consistency of the n **out of** n **bootstrap**.
- The results hold for more **general regularizers**.
- Application to **Colocalization** analysis.