Empiricial Regularized Optimal Transport

Fourth annual RTG 2088 Workshop

Marcel Klatt

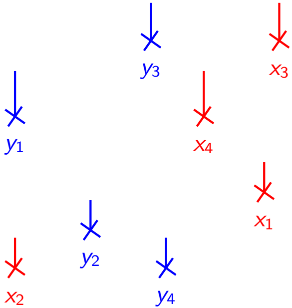
September 30, 2019

Institute for Mathematical Stochastics
University of Göttingen

References:

M. Klatt, C. Tameling and A. Munk *Empirical Regularized Optimal Transport: Statistical Theory and Applications*, arXiv:1810.09880, 2019

(Regularized) Optimal Transport in a Nutshell



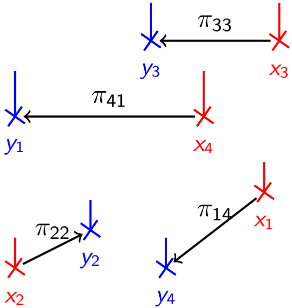
Finite discrete metric space $(\mathcal{X}, d)$

- 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 $p \geq 1$
- **Task:** Find the most efficient way to transport measure r into s.

(Regularized) Optimal Transport in a Nutshell



Transport plan

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

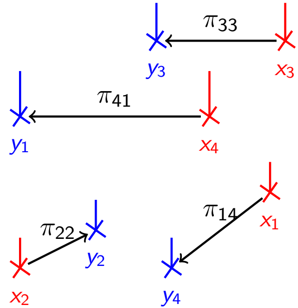
Finite discrete metric space $(\mathcal{X}, d)$

- 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 $p \geq 1$
- **Task:** Find the most efficient way to transport measure r into s.

(Regularized) Optimal Transport in a Nutshell



Transport plan

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

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 $p \geq 1$
- **Task:** Find the most efficient way to transport measure r into s.
- Solve the standard linear program

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

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

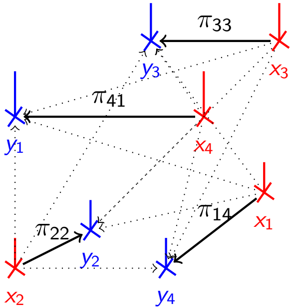
The **Wasserstein distance** of order $p \geq 1$ is defined as

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

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



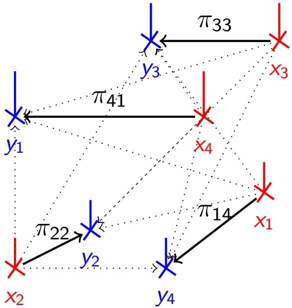
Finite discrete metric space $(\mathcal{X}, d)$

- **Task:** Find the most efficient way to transport measure r into s w.r.t. some regularity on the transport plan.

$$\min_{\pi \in \mathbb{R}^{N\times N}} \quad \sum_{i,j=1}^{N} d^p(x_i, y_j)\pi_{ij} - \lambda E(\pi)$$

$$\text{s.t.} \quad \pi \in \Pi(r, s)$$

(Regularized) Optimal Transport in a Nutshell



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

- **Task:** Find the most efficient way to transport measure r into s w.r.t. some regularity on the transport plan.

$$\min_{\pi \in \mathbb{R}^{N\times N}} \quad \sum_{i,j=1}^{N} d^p(x_i, y_j)\pi_{ij} - \lambda E(\pi)$$

$$\text{s.t.} \quad \pi \in \Pi(r, s)$$

- For $\lambda > 0$ find the (*unique*) entropy regularized transport plan

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

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

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

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

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

- Rates for approximation as $\lambda \searrow 0$ (*Luise et al. (2018)*):

$$\sup_{r,s \in \Delta_N} |W_{p,\lambda}^p(r,s) - W_p^p(r,s)| \leq C \exp\left(-\frac{1}{\lambda}\right) .$$

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 $X_1, \dots, X_n \overset{i.i.d.}{\sim} r$.

Statistical Framework

Let $\mathcal{X} = \{x_1, \ldots, x_N\}$ 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 $X_1, \ldots, X_n \overset{i.i.d.}{\sim} r$.

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 currently under investigation.

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

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

- **Different limit laws** under equality of measures (non-Gaussian vs. Gaussian)

(Non-regularized) OT-distance

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

regularized OT-distance

$$n^{\frac{1}{2}} \left\{W_{p,\lambda}(\hat{r}_n, r) - W_{p,\lambda}(r, r)\right\} \xrightarrow{\mathfrak{D}} \mathcal{N}_1(0, \sigma_\lambda^2(r|r))$$

(Non-regularized) OT vs. regularized OT

- **Different limit laws** under equality of measures (non-Gaussian vs. Gaussian)

(Non-regularized) OT-distance	regularized OT-distance
$n^{\frac{1}{2p}} W_p(\hat{r}_n, r)$	$n^{\frac{1}{2}} \{W_{p,\lambda}(\hat{r}_n, r) - W_{p,\lambda}(r, r)\}$
$\big\downarrow \mathfrak{D}$	$\big\downarrow \mathfrak{D}$
$\{\max_{u \in \Phi^*} \langle G, u \rangle\}^{\frac{1}{p}}$	$\mathcal{N}_1(0, \sigma_\lambda^2(r\|r))$

- **Different scaling behaviour**, i.e., for regularized OT-distance the scaling behaviour is independent of p

(Non-regularized) OT vs. regularized OT

Recall that for $\lambda \searrow 0$

$$\sup_{r,s \in \Delta_N} |W_{p,\lambda}^p(r,s) - W_p^p(r,s)| \leq C \exp\left(-\frac{1}{\lambda}\right) . \quad (1)$$

(Non-regularized) OT vs. regularized OT

Recall that for $\lambda \searrow 0$

$$\sup_{r,s \in \Delta_N} |W_{p,\lambda}^p(r,s) - W_p^p(r,s)| \leq C \exp\left(-\frac{1}{\lambda}\right) . \qquad (1)$$

This yields, e.g. for $r = s$ that

$$\begin{aligned}&\sqrt{n}\left\{W_{p,\lambda}^p(\hat{r}_n, r) - W_{p,\lambda}^p(r,r)\right\}\\ &= \underbrace{\sqrt{n}\left\{W_{p,\lambda}^p(\hat{r}_n, r) - W_p^p(\hat{r}_n, r)\right\}}_{(I)} + \underbrace{\sqrt{n}W_p^p(\hat{r}_n, r)}_{(II)} - \underbrace{\sqrt{n}W_{p,\lambda}^p(r,r)}_{(III)} .\end{aligned}$$

(Non-regularized) OT vs. regularized OT

Recall that for $\lambda \searrow 0$

$$\sup_{r,s\in\Delta_N} |W_{p,\lambda}^p(r,s) - W_p^p(r,s)| \leq C \exp\left(-\frac{1}{\lambda}\right) . \qquad (1)$$

This yields, e.g. for $r = s$ that

$$\begin{aligned}&\sqrt{n}\left\{W_{p,\lambda}^p(\hat{r}_n, r) - W_{p,\lambda}^p(r,r)\right\} \\ &= \underbrace{\sqrt{n}\left\{W_{p,\lambda}^p(\hat{r}_n, r) - W_p^p(\hat{r}_n, r)\right\}}_{(I)} + \underbrace{\sqrt{n}W_p^p(\hat{r}_n, r)}_{(II)} - \underbrace{\sqrt{n}W_{p,\lambda}^p(r,r)}_{(III)} .\end{aligned}$$

(i) $(I) + (III) \to 0$ for $\lambda(n) \in o\left(\frac{1}{\log(n)}\right)$ by (1)

(ii) $(II) \xrightarrow{\mathfrak{D}} \max_{u\in\Phi^*} \langle G, u\rangle$ by Sommerfeld & Munk (2018)

$$\Rightarrow \sqrt{n}\left\{W_{p,\lambda(n)}^p(\hat{r}_n, r) - W_{p,\lambda(n)}^p(r,r)\right\} \xrightarrow{\mathfrak{D}} \max_{u\in\Phi^*} \langle G, u\rangle .$$

(Non-regularized) OT vs. regularized OT

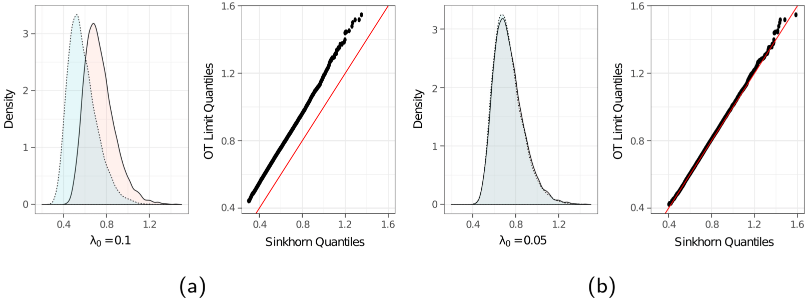
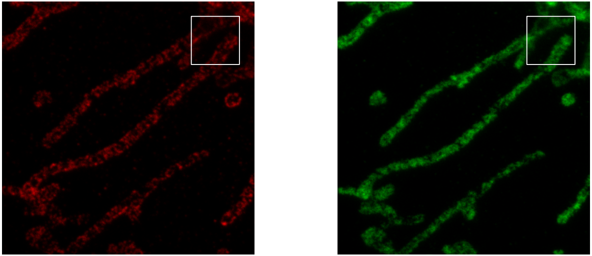


Figure 1: Comparison of the finite sample distribution ($r = s$, $n = 25$) of the empirical Sinkhorn divergence to the OT limit law on an equidistant grid of size $L = 10$ for $\lambda_0 = 0.1$ (a) and $\lambda_0 = 0.05$ (b).



(a) ATP Synthase (b) MIC60

Figure 2: 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.

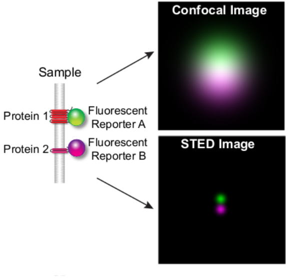


Figure 3: Simulation of Confocal and STED images of two proteins which are located at a distance of 45 nm. The resolution of the confocal image is 244 nm and for the STED image it is 40 nm.

Application: Colocalization Analysis

Our approach:

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

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

Application: Colocalization Analysis

Our approach:

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

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

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.

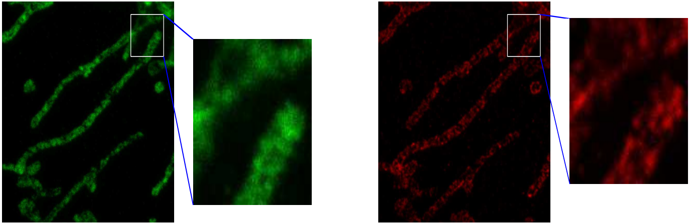
Application: Colocalization Analysis

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

Application: Colocalization Analysis

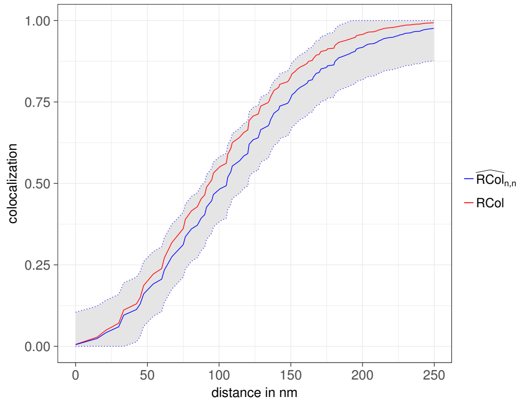


Figure 5: Staining of ATP Synthase and MIC60 for the zoom ins (128×128 images). The sampled regularized colocalization ($\lambda = 0.01$) (solid blue line, subsampling $n = 2000$) with bootstrap confidence bands (gray area between dashed blue lines) based on the n out of n bootstrap with $B = 100$ replications 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**.
- **Comparison** to related results for (non-regularized) optimal transport.
- Application to **Colocalization** analysis.