Empirical (Regularized) Optimal Transport: Statistical Theory and Applications

Optimal Transport, Topological Data Analysis and Applications to Shape and Machine Learning

Marcel Klatt[1] Carla Tameling[1] Yoav Zemel[2] Axel Munk[1]

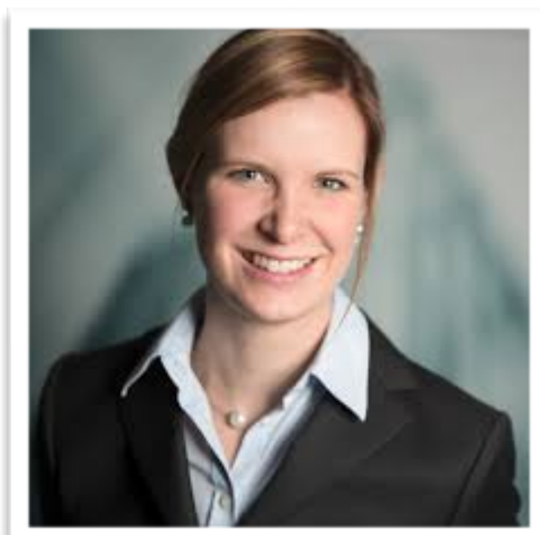
July 29, 2020

Institute for Mathematical Stochastics[1]
University of Göttingen

Centre for Mathematical Sciences[2]
University of Cambridge

Based on joint work with…

Yoav Zemel



Carla Tameling

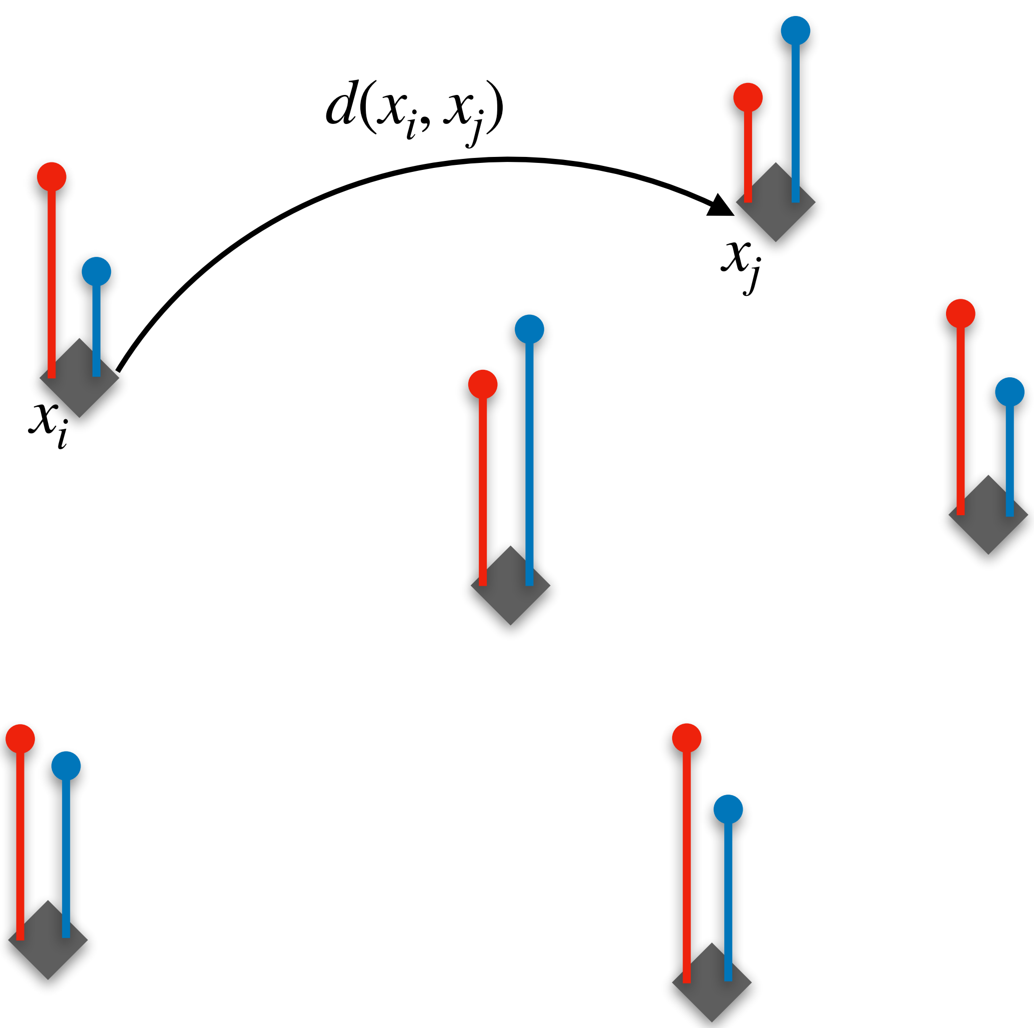
Axel Munk

M. Klatt, C. Tameling and A. Munk *Empirical Regularized Optimal Transport: Statistical Theory and Applications*, SIAM Journal on Mathematics of Data Science (2020)

M. Klatt, A. Munk and Y. Zemel *Limit Laws for Empirical Optimal Solutions in Stochastic Linear Programs,* arXiv preprint (2020)

Optimal Transport & Entropy Regularization

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



Two probability measures

$$r = \sum_i r_i \delta_{x_i} \qquad s = \sum_j s_j \delta_{x_j}$$

Transport costs

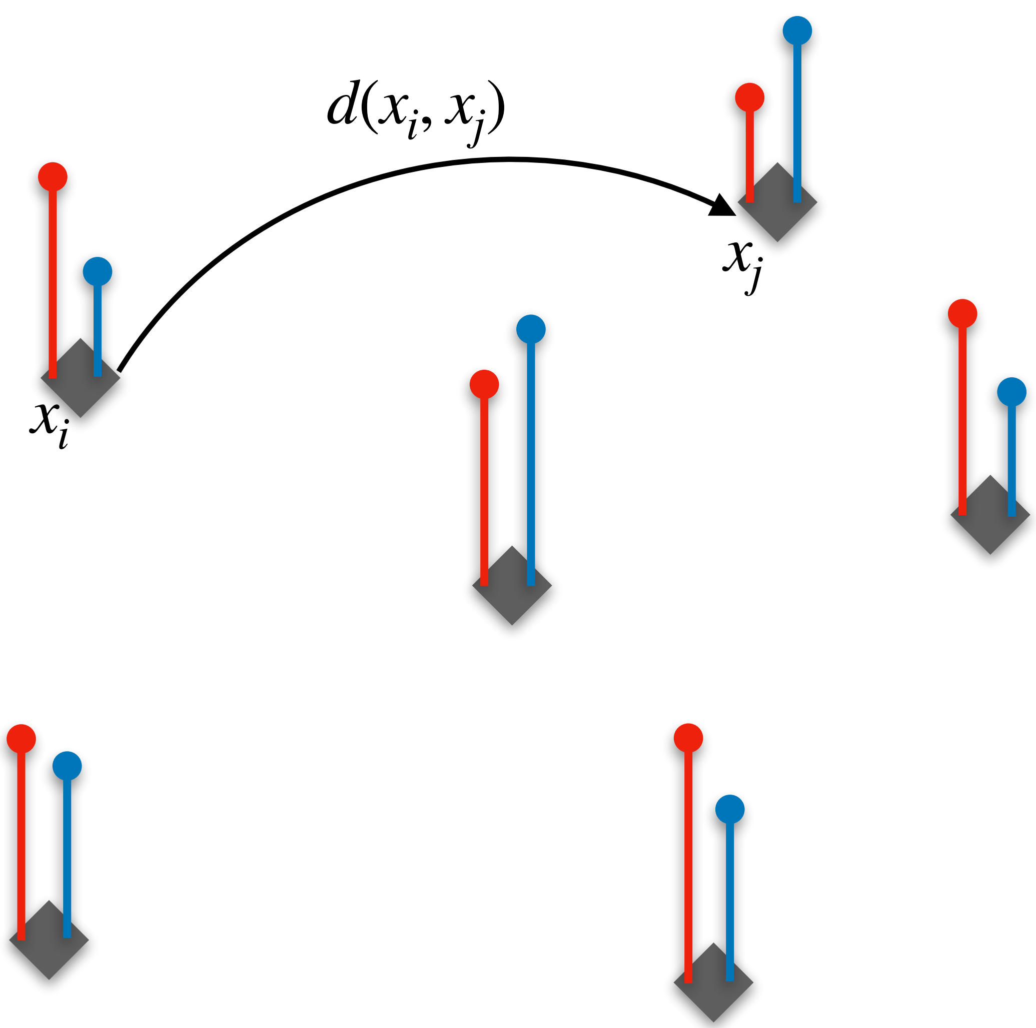
$$d_{i,j}^p := d^p(x_i, x_j), \quad p \geq 1$$

Task: Find the most efficient way to transport one measure into the other.

$$\min_{\pi \in \Pi(r,s)} \sum_{i,j} d_{ij}^p \pi_{ij}$$

Optimal Transport & Entropy Regularization

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



Wasserstein distance

$$W_p(r, s) = \Big(\min_{\pi \in \Pi(r,s)} \sum_{i,j} d_{ij}^p \pi_{ij} \Big)^{\frac{1}{p}}$$

OT plan

$$\pi(r, s) \in \arg \min_{\pi \in \Pi(r,s)} \sum_{i,j} d_{ij}^p \pi_{ij}$$

Regularized OT (ROT)

$$\min_{\pi \in \Pi(r,s)} \sum_{i,j} d_{ij}^p \pi_{ij} - \lambda E(\pi) \quad , \quad \lambda > 0$$

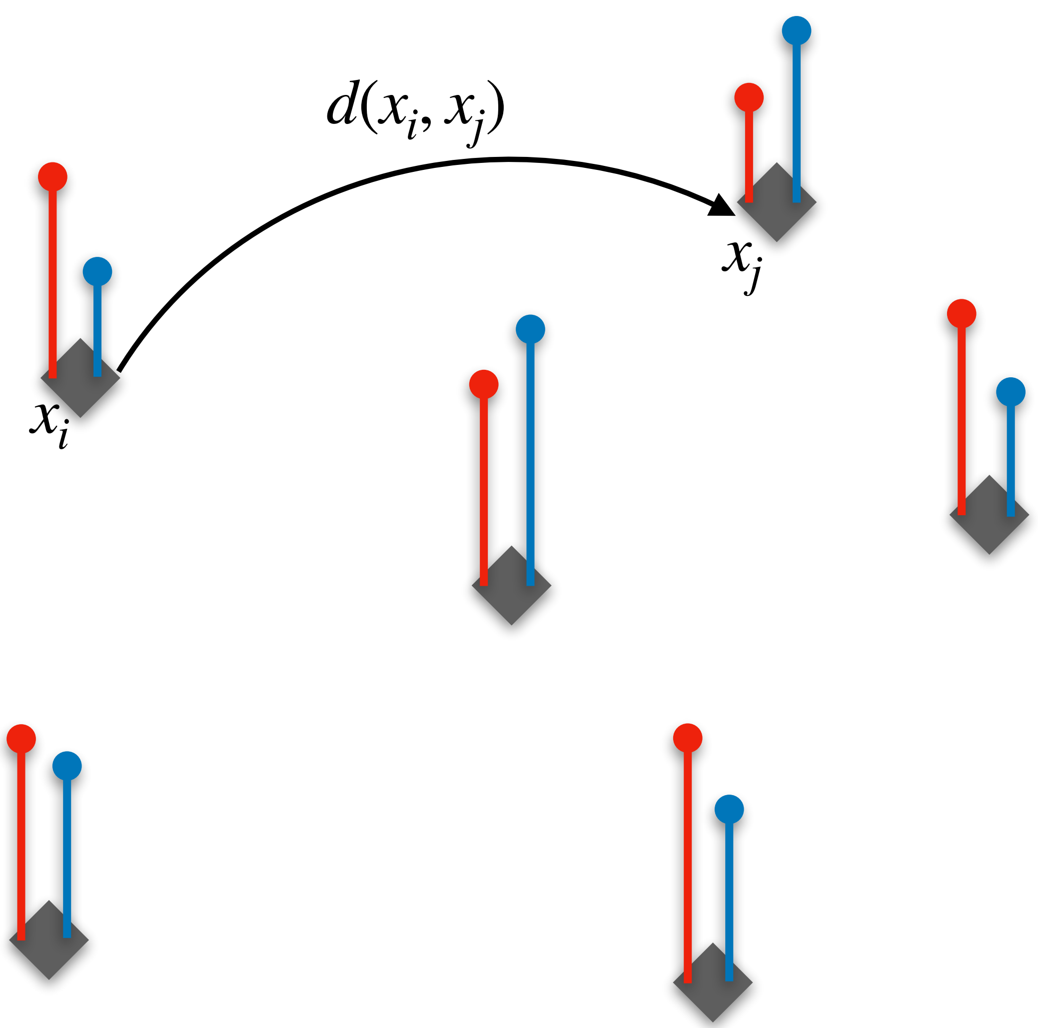
Entropy: $E(\pi) = - \sum_{i,j} \pi_{ij} \log(\pi_{ij})$

M. Cuturi *Sinkhorn Distances: Lightspeed Computation of Optimal Transportation Distances*, Advances in neural information processing systems (2013)

G. Peyré and M. Cuturi *Computational Optimal Transport* Foundations and Trends in Machine Learning (2019)

Optimal Transport & Entropy Regularization

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



Wasserstein distance

$$W_p(r, s) = \Big(\min_{\pi \in \Pi(r,s)} \sum_{i,j} d_{ij}^p \pi_{ij} \Big)^{\frac{1}{p}}$$

OT plan

$$\pi(r, s) \in \arg \min_{\pi \in \Pi(r,s)} \sum_{i,j} d_{ij}^p \pi_{ij}$$

Sinkhorn divergence

$$W_{p,\lambda}(r, s) = \Big(\min_{\pi \in \Pi(r,s)} \sum_{i,j} d_{ij}^p \pi_{ij} - \lambda E(\pi) \Big)^{\frac{1}{p}}$$

ROT plan

$$\pi_\lambda(r, s) = \arg \min_{\pi \in \Pi(r,s)} \sum_{i,j} d_{ij}^p \pi_{ij} - \lambda E(\pi)$$

M. Cuturi *Sinkhorn Distances: Lightspeed Computation of Optimal Transportation Distances*, Advances in neural information processing systems (2013)

G. Peyré and M. Cuturi *Computational Optimal Transport* Foundations and Trends in Machine Learning (2019)

In this talk, we focus on <u>statistical properties</u> of these quantities!

Suppose, we observe independent and identically distributed random variables

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

This naturally leads to estimate r by the empirical measure

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

Central Question:

How do these random quantities relate to their population counterpart, respectively?

Wasserstein distance

$$W_p(r,s) = \Big(\min_{\pi\in\Pi(\hat{r}_n,s)} \sum_{i,j} d_{ij}^p \pi_{ij}\Big)^{\frac{1}{p}}$$

OT plan

$$\pi(r,s) \in \arg\min_{\pi\in\Pi(\hat{r}_n,s)} \sum_{i,j} d_{ij}^p \pi_{ij}$$

Sinkhorn divergence

$$W_{p,\lambda}(r,s) = \Big(\min_{\pi\in\Pi(\hat{r}_n,s)} \sum_{i,j} d_{ij}^p \pi_{ij} - \lambda E(\pi)\Big)^{\frac{1}{p}}$$

ROT plan

$$\pi_\lambda(r,s) = \arg\min_{\pi\in\Pi(\hat{r}_n,s)} \sum_{i,j} d_{ij}^p \pi_{ij} - \lambda E(\pi)$$

In this talk, we focus on statistical properties of these quantities!

Goal:

Quantify random fluctuation by asymptotic limit laws!

Wasserstein distance

$$a_n \left\{ W_p^p(\hat{r}_n, s) - W_p^p(r, s) \right\} \xrightarrow{\mathscr{D}} ??$$

OT plan

$$a_n \left\{ \pi(\hat{r}_n, s) - \pi(r, s) \right\} \xrightarrow{\mathscr{D}} ??$$

Sinkhorn divergence

$$a_n \left\{ W_{p,\lambda}^p(\hat{r}_n, s) - W_{p,\lambda}^p(r, s) \right\} \xrightarrow{\mathscr{D}} ??$$

ROT plan

$$a_n \left\{ \pi_\lambda(\hat{r}_n, s) - \pi_\lambda(r, s) \right\} \xrightarrow{\mathscr{D}} ??$$

In this talk, we focus on statistical properties of these quantities!

Distributional limit laws for the Wasserstein distance are well understood:

M. Sommerfeld and A. Munk *Inference for empirical Wasserstein distances on finite spaces.*, Journal of the Royal Statistical Society: Series B (Methodological) (2018)

C. Tameling, M. Sommerfeld and A. Munk *Empirical optimal transport on countable metric spaces: Distributional limits and statistical applications*, Annals of Applied Probability (2019)

- $\Gamma_\star$ set of dual solutions

$$\max_{\alpha,\beta\in\mathbb{R}^N} \sum_i \alpha_i r_i + \sum_j \beta_j s_j$$
$$s.t. \quad \alpha_i + \beta_j \leq d_{ij}^p$$

- G the Gaussian limit of $\sqrt{n}\left(\hat{r}_n - r\right)$

Wasserstein distance

$$\sqrt{n}\left\{W_p^p(\hat{r}_n, s) - W_p^p(r, s)\right\} \xrightarrow{\mathscr{D}} \max_{\alpha_\star\in\Gamma_\star} \langle G, \alpha_\star\rangle$$

OT plan

$$a_n\left\{\pi(\hat{r}_n, s) - \pi(r, s)\right\} \xrightarrow{\mathscr{D}} ??$$

Sinkhorn divergence

$$a_n\left\{W_{p,\lambda}^p(\hat{r}_n, s) - W_{p,\lambda}^p(r, s)\right\} \xrightarrow{\mathscr{D}} ??$$

ROT plan

$$a_n\left\{\pi_\lambda(\hat{r}_n, s) - \pi_\lambda(r, s)\right\} \xrightarrow{\mathscr{D}} ??$$

ROT plan

$$\pi_\lambda(\hat{r}_n, s) = \arg \min_{\pi \in \Pi(\hat{r}_n, s)} \sum_{i,j} d_{ij}^p \pi_{ij} - \lambda E(\pi), \quad \lambda > 0$$

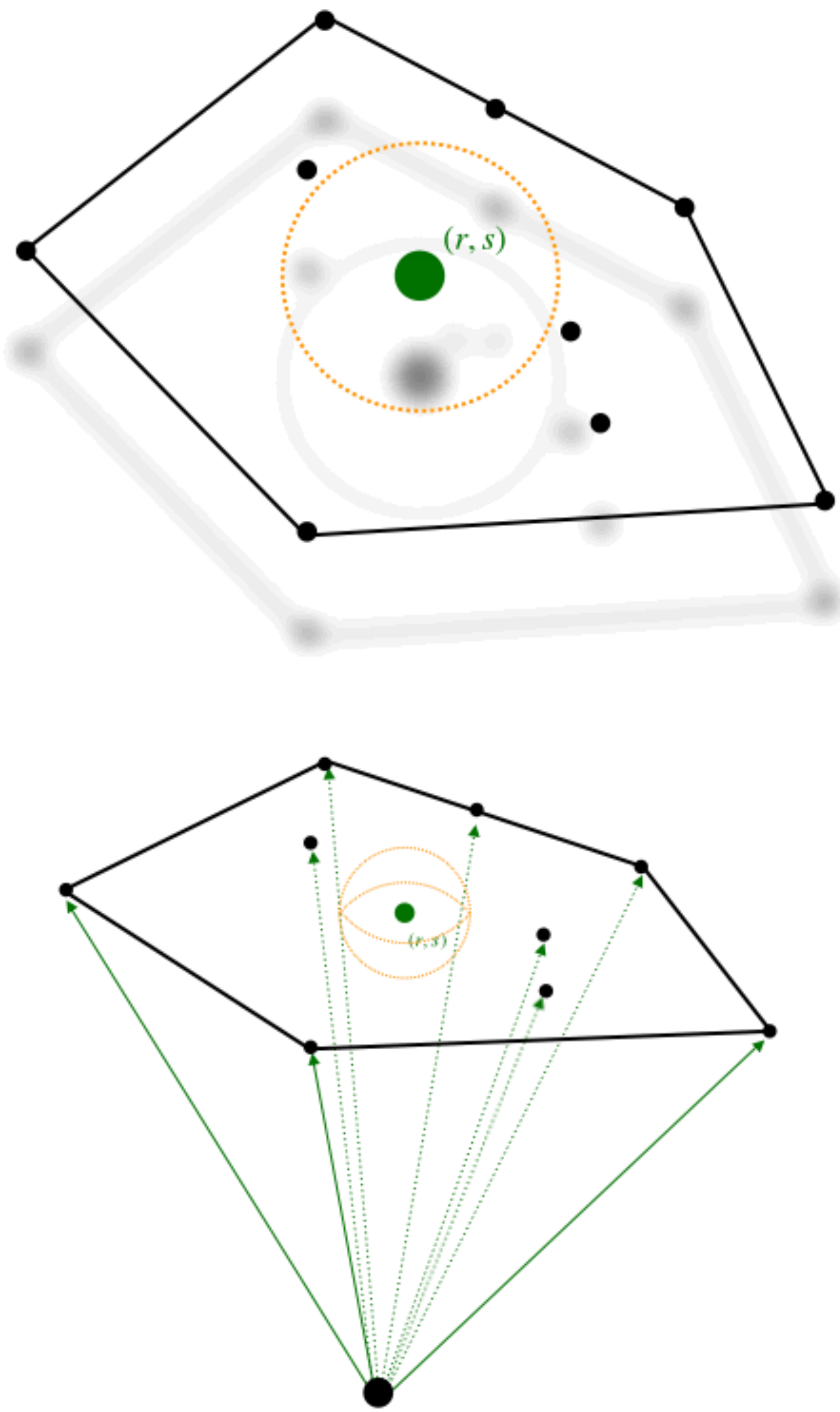
<u>**Theorem (K., Tameling & Munk (2020)):**</u>

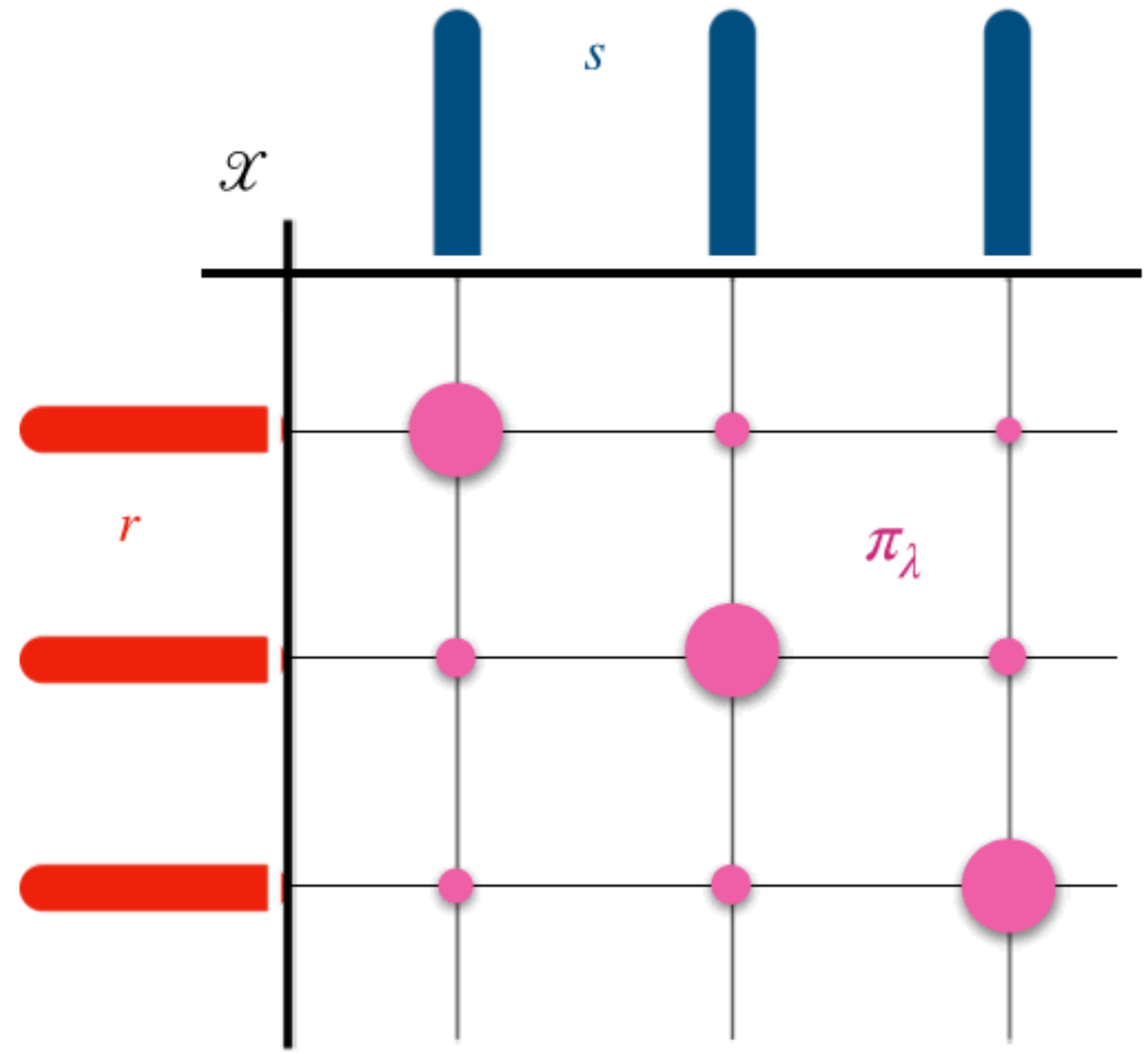
With the sample size *n* approaching infinity, it holds that

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

Why Gaussian fluctuation?

$$\boxed{\sqrt{n}\,\left\{\pi_\lambda(\hat{r}_n, s) - \pi_\lambda(r, s)\right\} \xrightarrow{\mathcal{D}} \mathcal{N}_{N^2}\left(0, \Sigma_\lambda(r \mid s)\right)}$$





$$\min_{\pi} c^T\pi - \lambda E(\pi)$$

$$A\ \pi = \begin{pmatrix} r \\ s \end{pmatrix}$$

$$\pi \geq 0$$

Entropy Regularized OT

$\lambda > 0$

Further consequences

$$\boxed{\sqrt{n}\left\{\pi_\lambda(\hat{r}_n, s) - \pi_\lambda(r, s)\right\} \xrightarrow{\mathscr{D}} \mathcal{N}_{N^2}\left(0, \Sigma_\lambda(r \mid s)\right)}$$

Limit distributions for…

- *Sinkhorn divergence* $$\boxed{\sqrt{n}\left\{W^p_{p,\lambda}(\hat{r}_n, s) - W^p_{p,\lambda}(r, s)\right\} \xrightarrow{\mathscr{D}} \langle G, \alpha_\lambda\rangle}$$

- $$\alpha_\lambda = \arg \max_{\alpha,\beta \in \mathbb{R}^{\mathcal{X}}} \alpha^T r + \beta^T s - \lambda \sum_{i,j} \exp\left(\frac{\alpha_i + \beta_j - d^p_{ij}}{\lambda}\right) - 1.$$

- G the Gaussian limit of $\sqrt{n}\left(\hat{r}_n - r\right)$

J. Bigot, E. Cazelles and N. Papadakis *Central limit theorems for entropy-regularized optimal transport on finite spaces and statistical applications*, Electronic Journal of Statistics (2019)

Further consequences

$$\boxed{\sqrt{n}\left\{\pi_\lambda(\hat{r}_n, s) - \pi_\lambda(r, s)\right\} \xrightarrow{\mathcal{D}} \mathcal{N}_{N^2}\left(0, \Sigma_\lambda(r \mid s)\right)}$$

Limit distributions for…

- *Sinkhorn divergence* $$\boxed{\sqrt{n}\left\{W^p_{p,\lambda}(\hat{r}_n, s) - W^p_{p,\lambda}(r, s)\right\} \xrightarrow{\mathcal{D}} \langle G, \alpha_\lambda \rangle}$$

… and many more related quantities:

- *Sinkhorn divergence II* $$\tilde{W}^p_{p,\lambda}(r, s) := \sum_{i,j} d^p_{ij} \pi_\lambda(r, s)$$
- *Sinkhorn loss* $$S^p_{p,\lambda}(r, s) := W^p_{p,\lambda}(r, s) - \frac{1}{2}\left(W^p_{p,\lambda}(r, s) + W^p_{p,\lambda}(r, s)\right)$$
- *Richardson extrapolation* $$R^p_{p,\lambda}(r, s) := 2S^p_{p,\lambda}(r, s) - S^p_{p,\sqrt{2}\lambda}(r, s)$$

J. Bigot, E. Cazelles and N. Papadakis *Central limit theorems for entropy-regularized optimal transport on finite spaces and statistical applications*, Electronic Journal of Statistics (2019)

A. Genevay, G. Peyré and M. Cuturi *Learning generative models with Sinkhorn divergences*, Proceedings of the 21st International Conference on Artificial Intelligence and Statistics (2018)

L. Chizat, P. Roussillon, F. Léger, F.-X. Vialard and G. Peyré *Faster Wasserstein distance estimation with Sinkhorn divergence*, arXiv preprint (2020)

Further consequences

$$\boxed{\sqrt{n}\left\{\pi_\lambda(\hat{r}_n, s) - \pi_\lambda(r, s)\right\} \xrightarrow{\mathcal{D}} \mathcal{N}_{N^2}\left(0, \Sigma_\lambda(r \mid s)\right)}$$

Limit distributions for…

- *Sinkhorn divergence*

$$\boxed{\sqrt{n}\left\{W^p_{p,\lambda}(\hat{r}_n, s) - W^p_{p,\lambda}(r, s)\right\} \xrightarrow{\mathcal{D}} \langle G, \alpha_\lambda \rangle}$$

There are extensions to countable metric spaces…

S. Hundrieser, M. Klatt and A. Munk *Limit Distributions for Entropic Optimal Transport on Countable Discrete Spaces*, In preparation

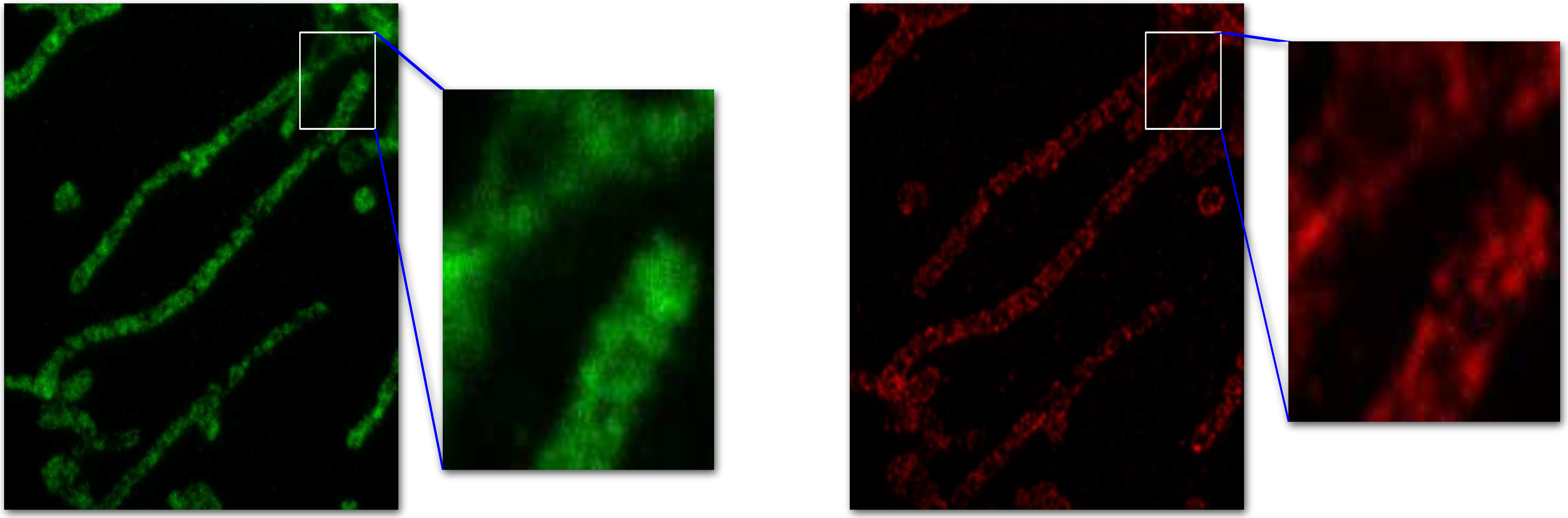
J. Bigot, E. Cazelles and N. Papadakis *Central limit theorems for entropy-regularized optimal transport on finite spaces and statistical applications*, Electronic Journal of Statistics (2019)

A. Genevay, G. Peyré and M. Cuturi *Learning generative models with Sinkhorn divergences*, Proceedings of the 21st International Conference on Artificial Intelligence and Statistics (2018)

L. Chizat, P. Roussillon, F. Léger, F.-X. Vialard and G. Peyré *Faster Wasserstein distance estimation with Sinkhorn divergence*, arXiv preprint (2020)

Application: Colocalization Analysis

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



(a) ATP Synthase

(b) MIC60

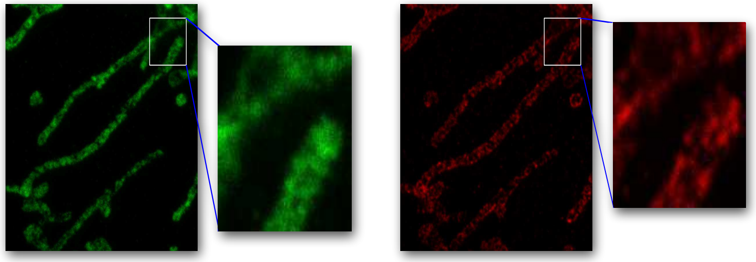
Multiscale Bioimaging Cluster of Excellence
From Molecular Machines to Networks of Excitable Cells

<u>Figure 1:</u> 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-occurence and correlation between them!

Application: Colocalization Analysis

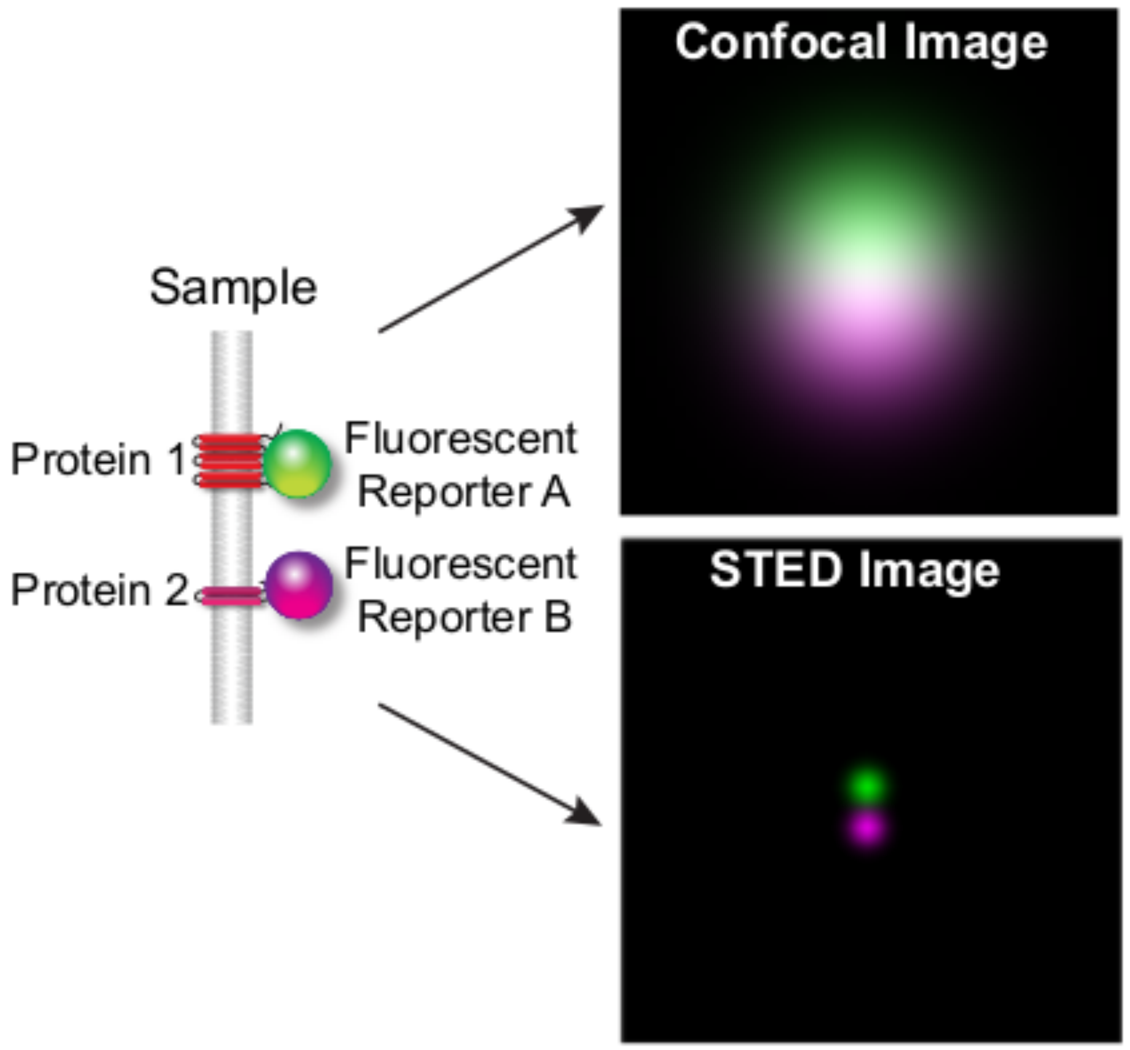
$$\boxed{\sqrt{n}\left\{\pi_\lambda(\hat{r}_n, s) - \pi_\lambda(r, s)\right\} \xrightarrow{\mathcal{D}} \mathcal{N}_{N^2}\left(0, \Sigma_\lambda(r \mid s)\right)}$$



(a) ATP Synthase (b) MIC60

Conventional Methods:

Pixel based intensity correlation analysis (Pearsons'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: Illustration of Confocal and STED images of two proteins which are located at a distance of 45nm. The resolution of the confocal image is 244nm and for the STED image it is 40nm.

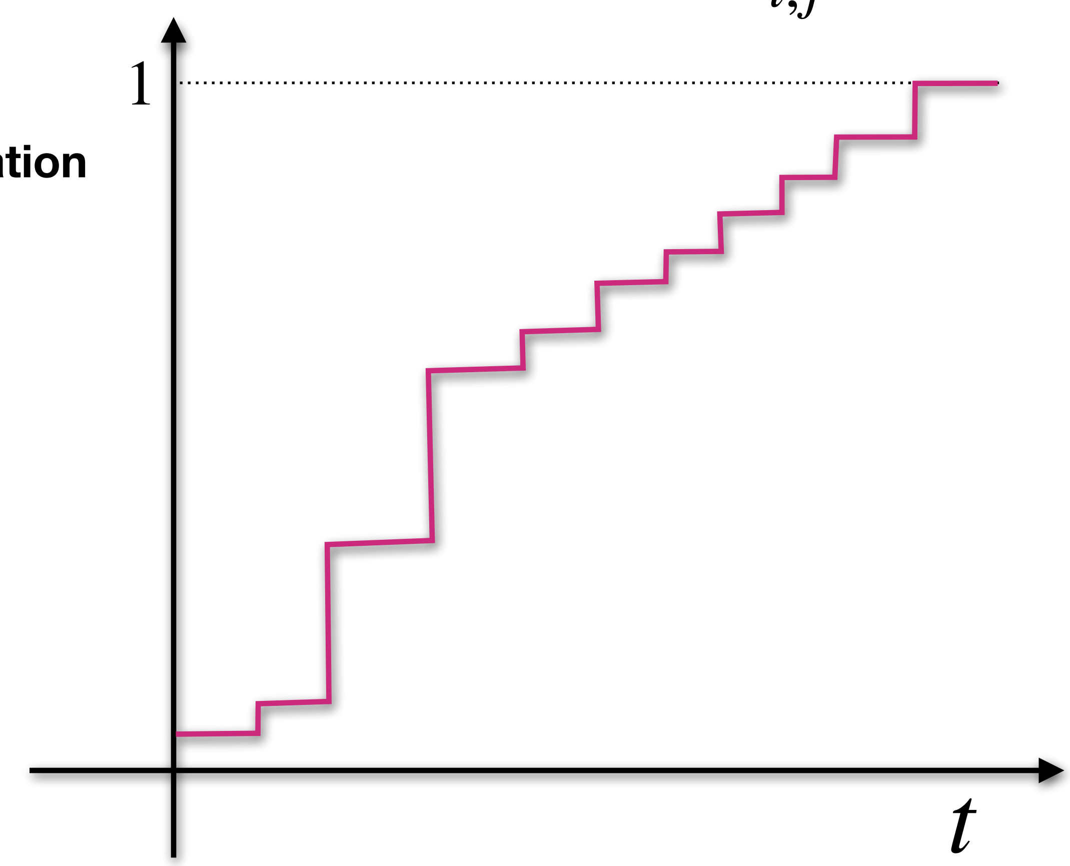
Application: Colocalization Analysis

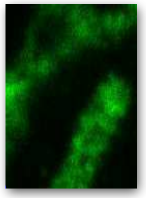
$$\boxed{\sqrt{n}\,\left\{\pi_\lambda(\hat{r}_n, s) - \pi_\lambda(r, s)\right\} \xrightarrow{\mathscr{D}} \mathscr{N}_{N^2}\left(0, \Sigma_\lambda(r \mid s)\right)}$$

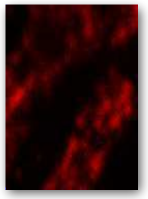
Our approach:

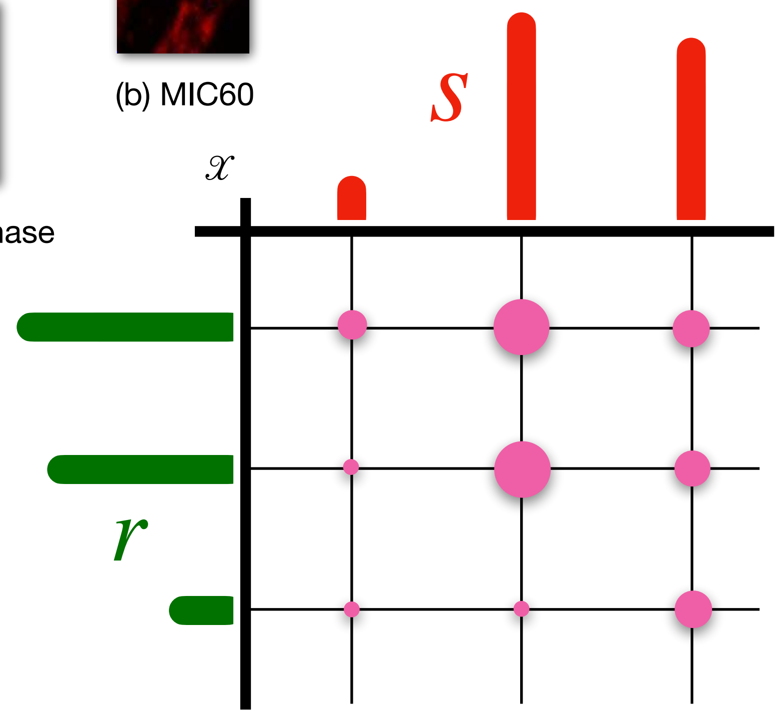
- Compute the entropy regularized transport plan
- Colocalization measure **RCol** based on (regularized) optimal transport for $t \in [0, \mathbf{diam}(\mathscr{X})]$

$$\mathbf{RCol}(\pi_\lambda(r, s))(t) = \sum_{i,j} \pi_\lambda(r, s)_{ij} 1_{\{d_{ij} \le t\}}$$




(a) ATP Synthase


(b) MIC60



$\pi_\lambda(r, s)$

Application: Colocalization Analysis

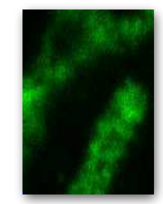
$$\sqrt{n}\left\{\pi_\lambda(\hat{r}_n, s) - \pi_\lambda(r, s)\right\} \xrightarrow{\mathscr{D}} \mathcal{N}_{N^2}\left(0, \Sigma_\lambda(r \mid s)\right)$$

Theorem (K., Tameling & Munk (2020)):

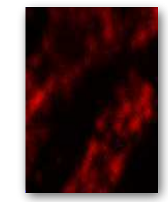
As sample size *n* approaches infinity

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

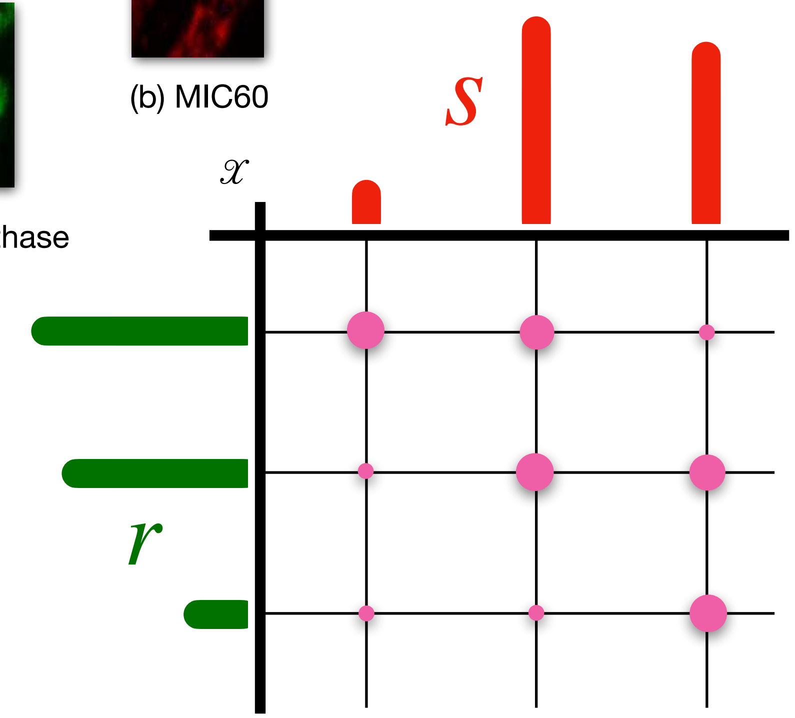
with *G* the random variable with distribution given by the limit law for the empirical regularized OT plan.



(a) ATP Synthase

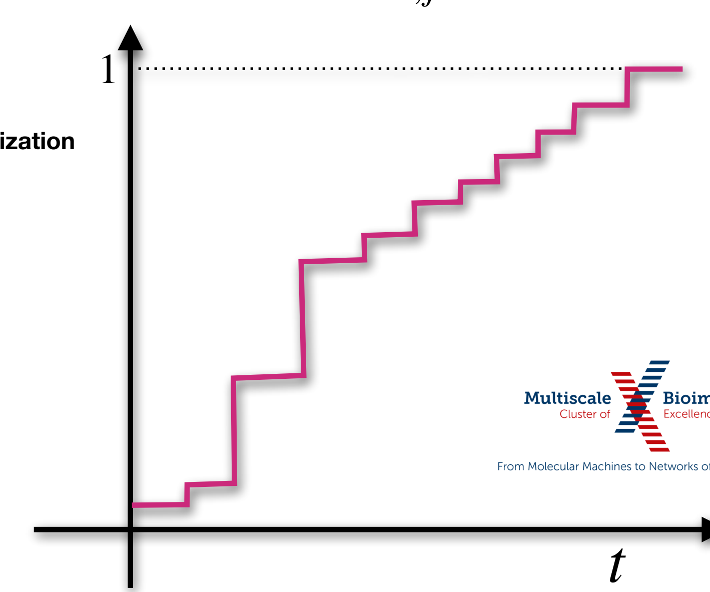


(b) MIC60



$$\pi_\lambda(r, s)$$

$$\mathbf{RCol}(\pi_\lambda(r, s))(t) = \sum_{i,j} \pi_\lambda(r, s)_{ij} 1_{\left\{d_{ij} \le t\right\}}$$

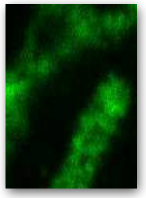


Multiscale Bioimaging Cluster of Excellence
From Molecular Machines to Networks of Excitable Cells

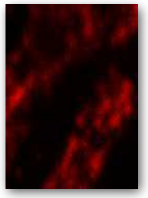
Application: Colocalization Analysis

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

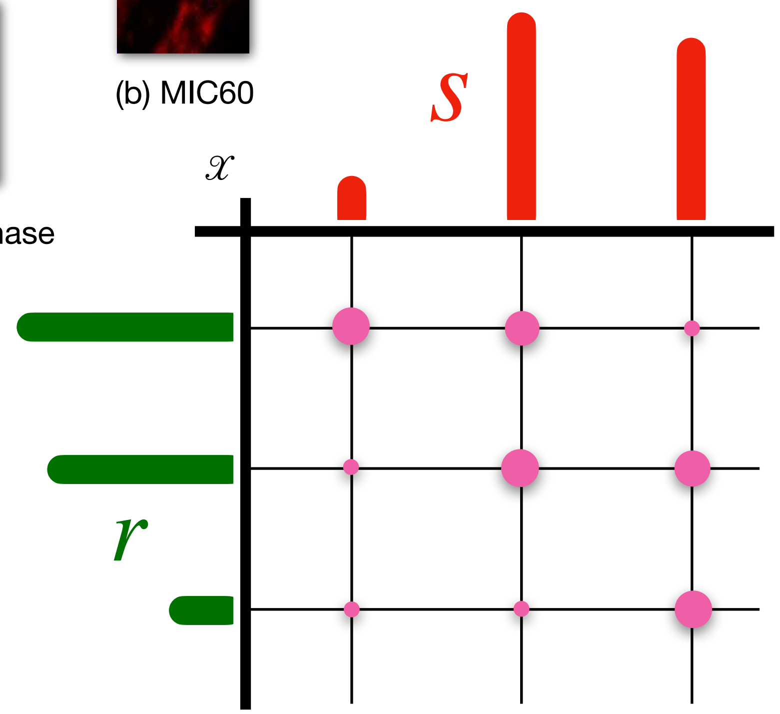
$$\sqrt{n}\left\{\widehat{\mathbf{RCol}}_n - \mathbf{RCol}\right\} \xrightarrow{\mathcal{D}} \mathbf{RCol}(G)$$



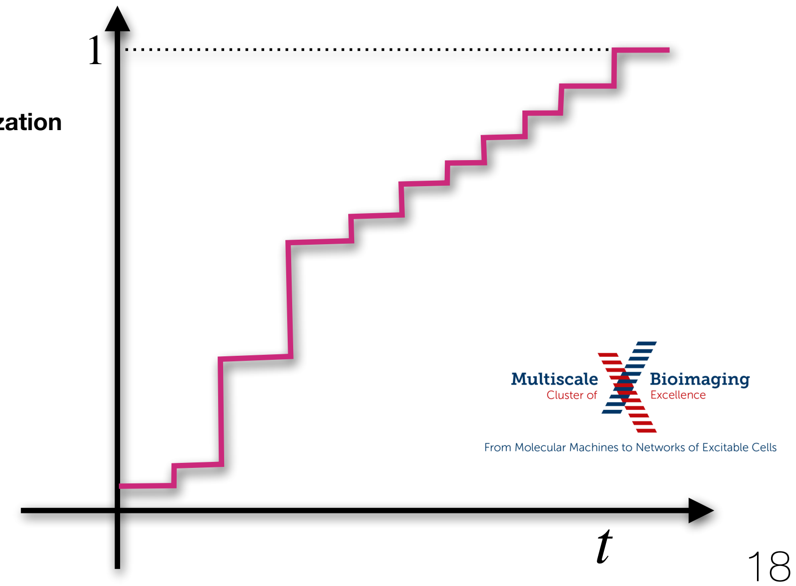
(a) ATP Synthase



(b) MIC60



$$\mathbf{RCol}(\pi_\lambda(r, s))(t) = \sum_{i,j} \pi_\lambda(r, s)_{ij} 1_{\{d_{ij} \le t\}}$$



- This yields $1-\alpha$ **approximate uniform confidence bands**, i.e.,

$$\lim_{n\to\infty} \mathbb{P}\left(\mathbf{RCol} \in I_n\right) = 1-\alpha,$$

where

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

and $\mathfrak{u}_{1-\alpha}$ is the $1-\alpha$ quantile from the distribution $\|\mathbf{RCol}(G)\|_\infty$.

- The quantile $\mathfrak{u}_{1-\alpha}$ can be consistently approximated by its ***n* out of *n* bootstrap analogue**.

Application: Colocalization Analysis

$$\boxed{\sqrt{n}\left\{\pi_\lambda(\hat{r}_n, s) - \pi_\lambda(r, s)\right\} \xrightarrow{\mathcal{D}} \mathcal{N}_{N^2}\left(0, \Sigma_\lambda(r \mid s)\right)}$$

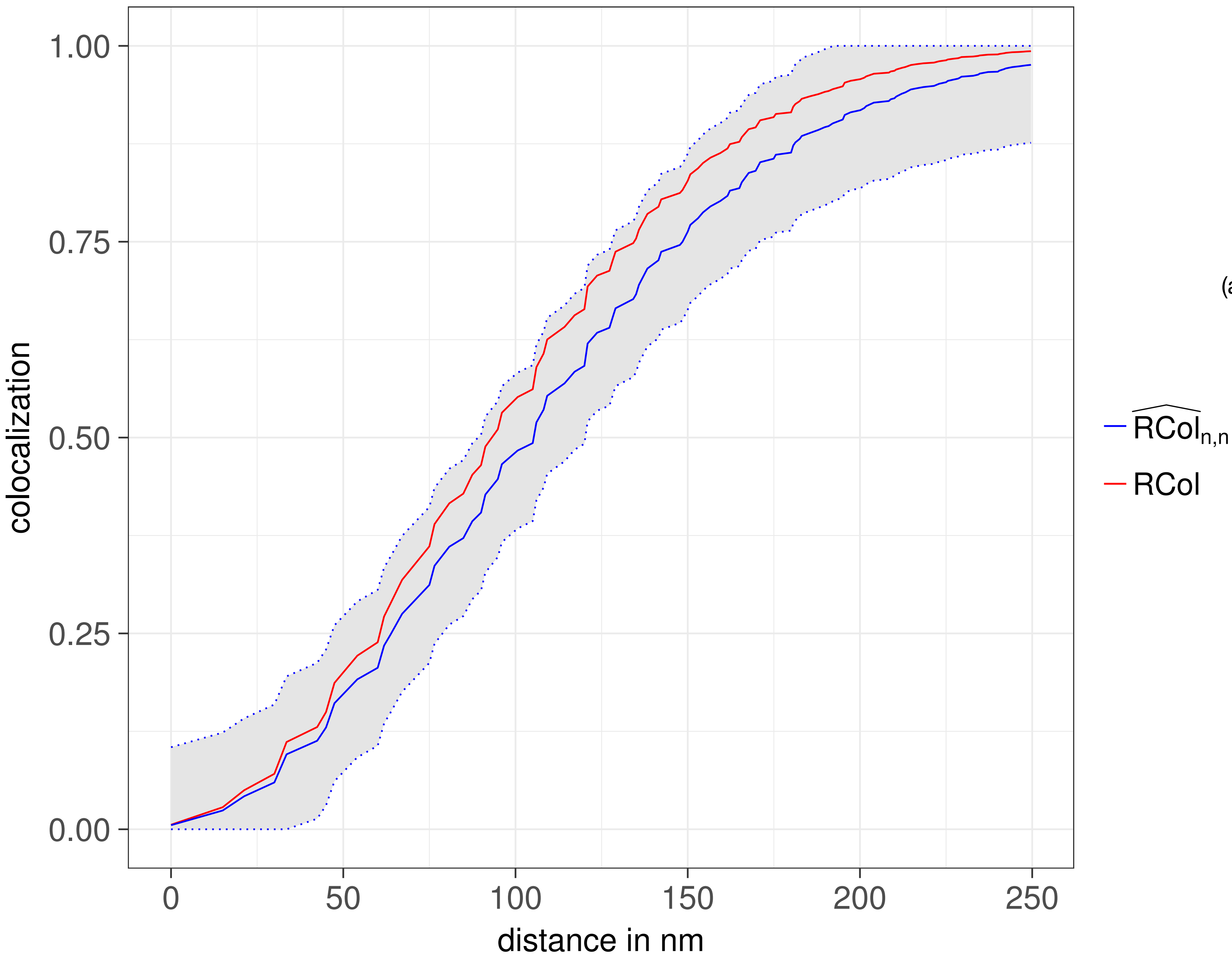
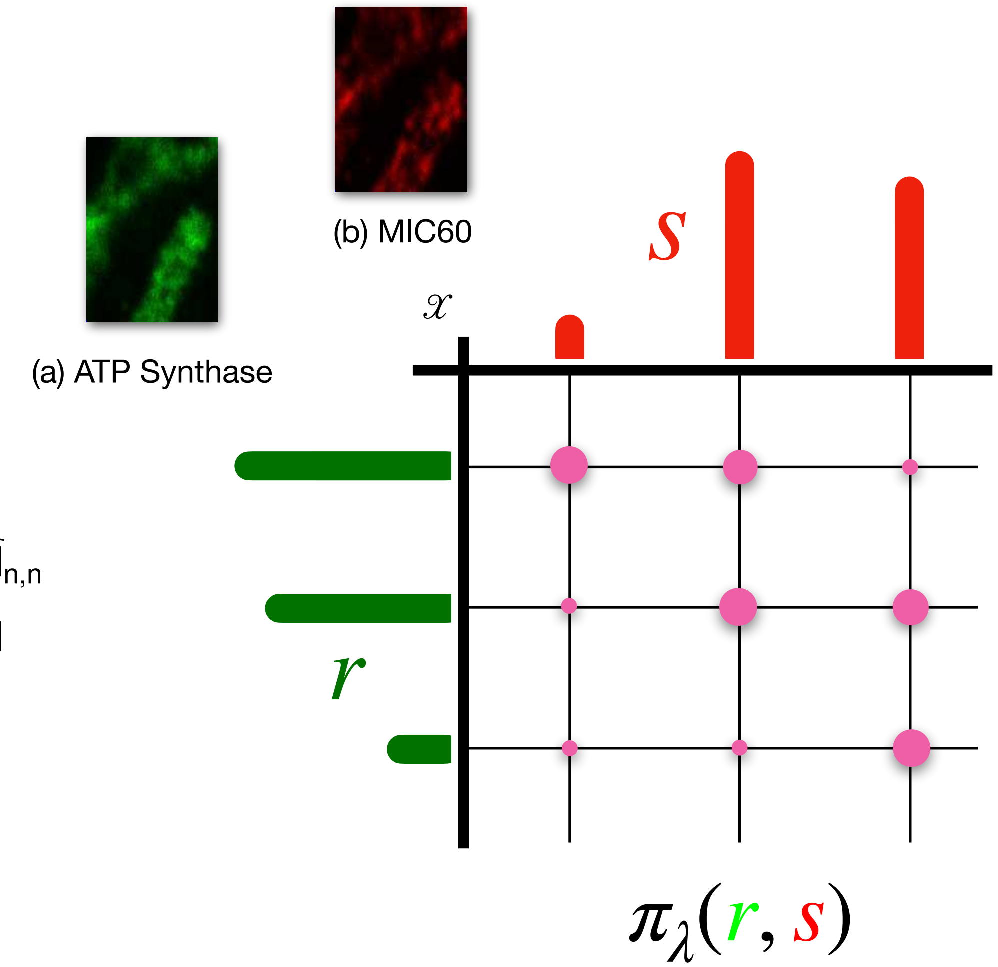
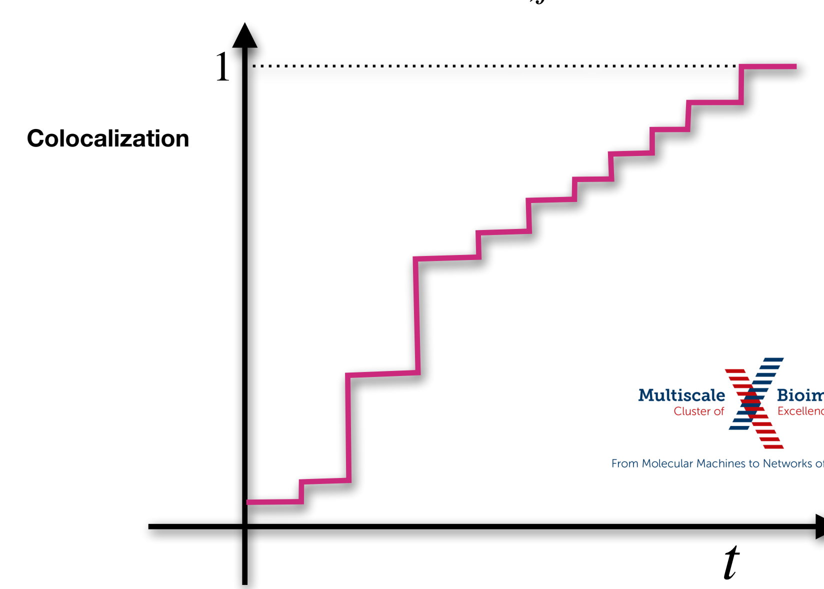


Figure 3: Colocalization for ATP Synthase and MIC60 for the zoom in (128 x 128 images). The sampled regularized colocalization (λ = 0.01) (solid blue line, subsampling n = 2000) with bootstrap confidence bands (gray area between dashed blue lines) based on the n out of n bootstrap with B = 100 replications and α = 0.05. Red solid line: Population regularized colocalization.



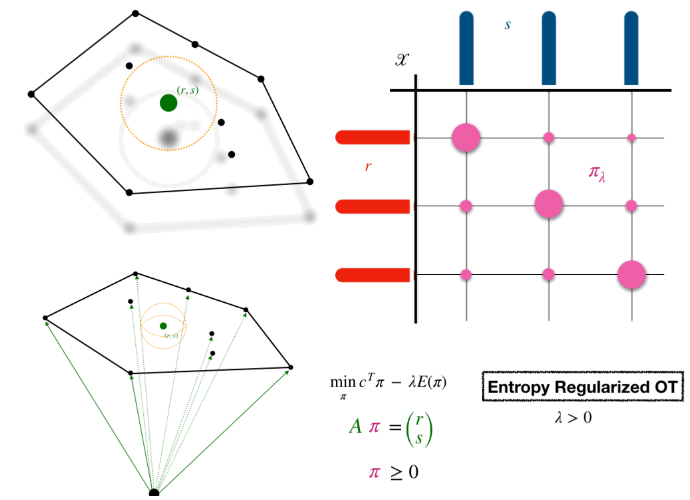
$$\pi_\lambda(r, s)$$

$$\textbf{RCol}(\pi_\lambda(r, s))(t) = \sum_{i,j} \pi_\lambda(r, s)_{ij} 1_{\{d_{ij} \le t\}}$$

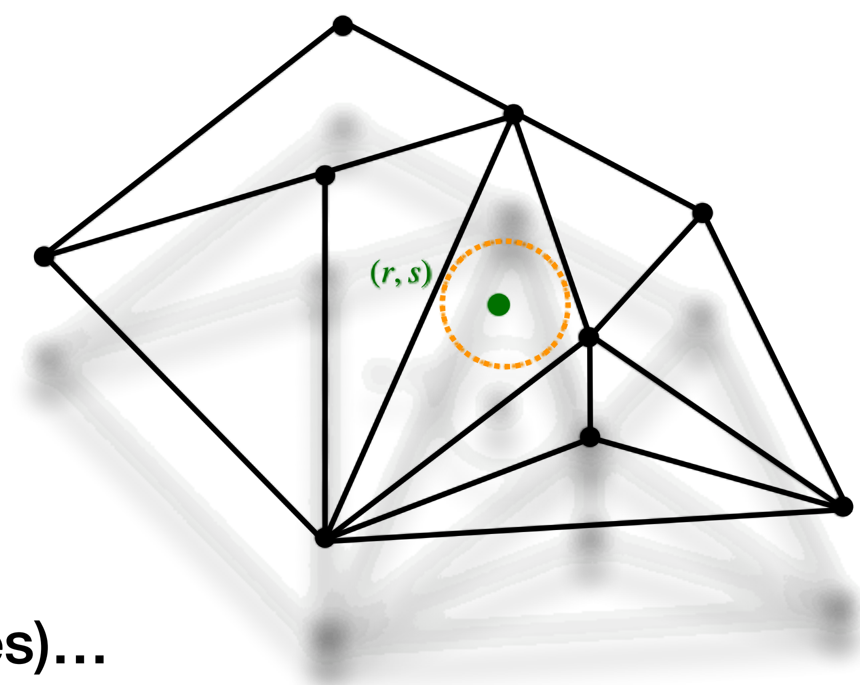


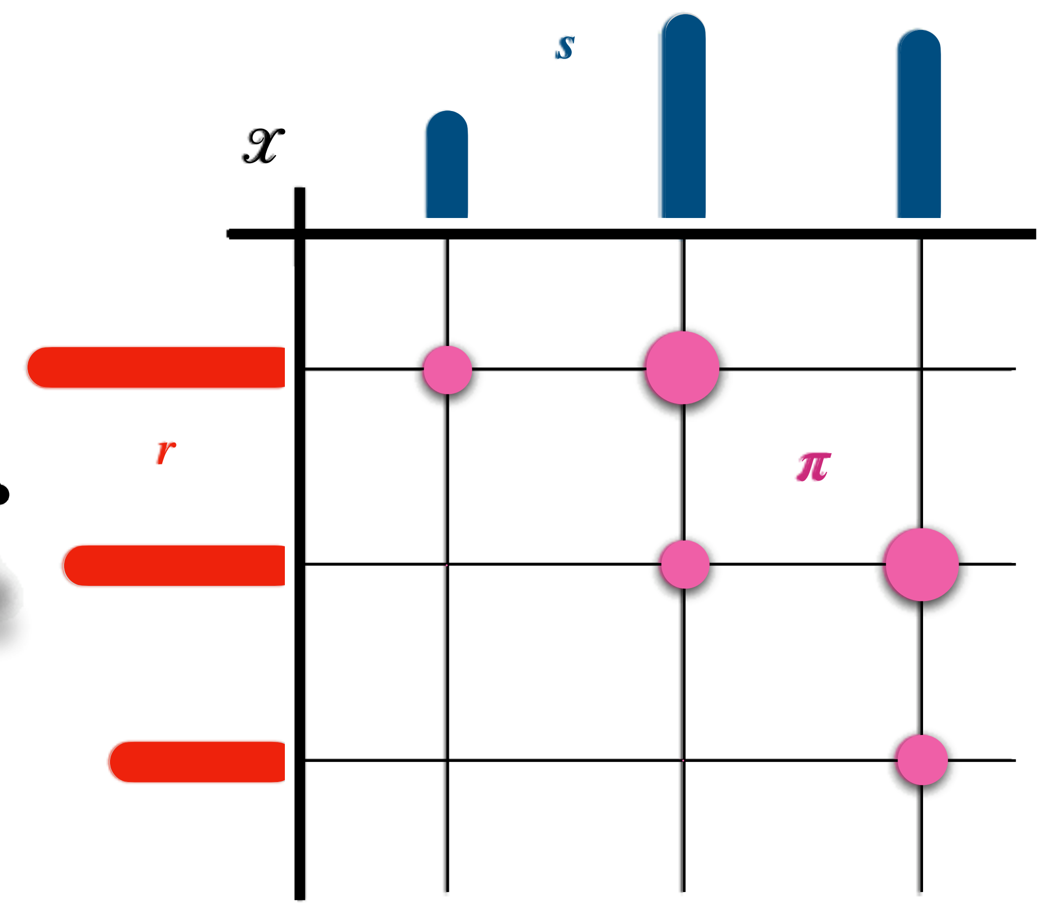
Multiscale Bioimaging Cluster of Excellence
From Molecular Machines to Networks of Excitable Cells

Let us recall... $\boxed{\sqrt{n}\left\{\pi_\lambda(\hat{r}_n, s) - \pi_\lambda(r,s)\right\} \xrightarrow{\mathcal{D}} \mathcal{N}_{N^2}\left(0, \Sigma_\lambda(r \mid s)\right)}$

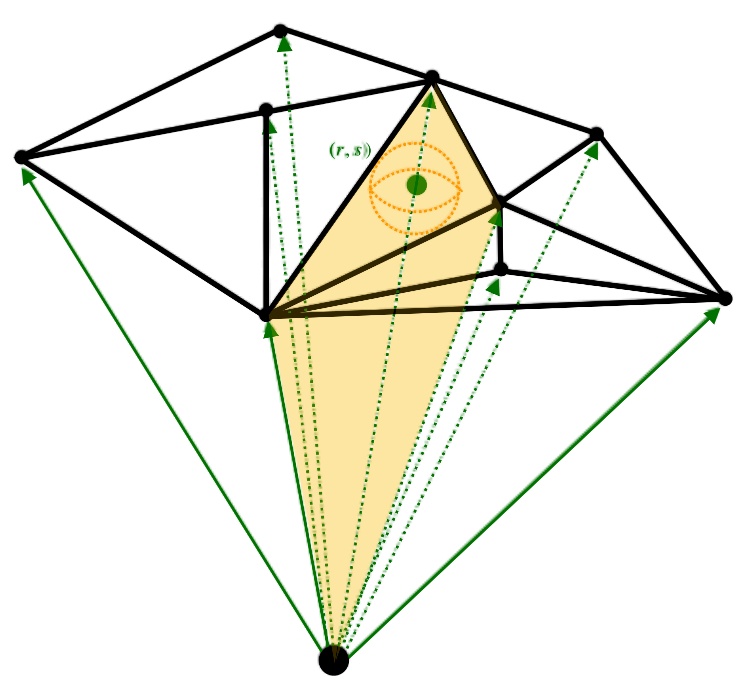


Can we hope for a similar behavior for the empirical OT plan? $(\lambda = 0)$





Sometimes (yes)...



$$\min_\pi c^T \pi$$

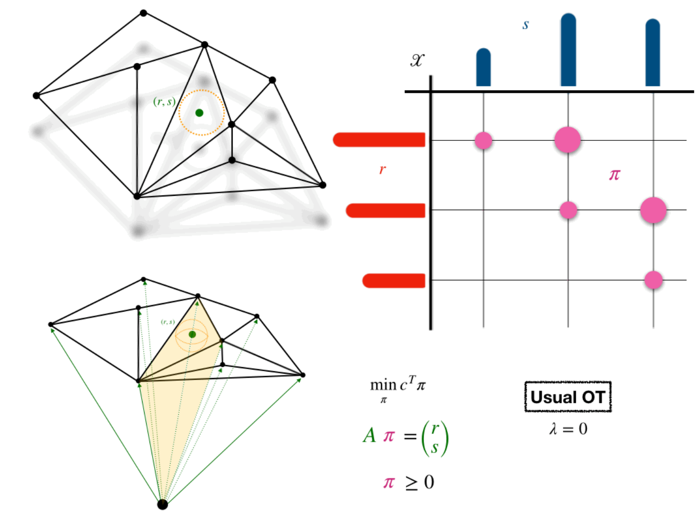
$$A\,\pi = \begin{pmatrix} r \\ s \end{pmatrix}$$

$$\pi \geq 0$$

Usual OT

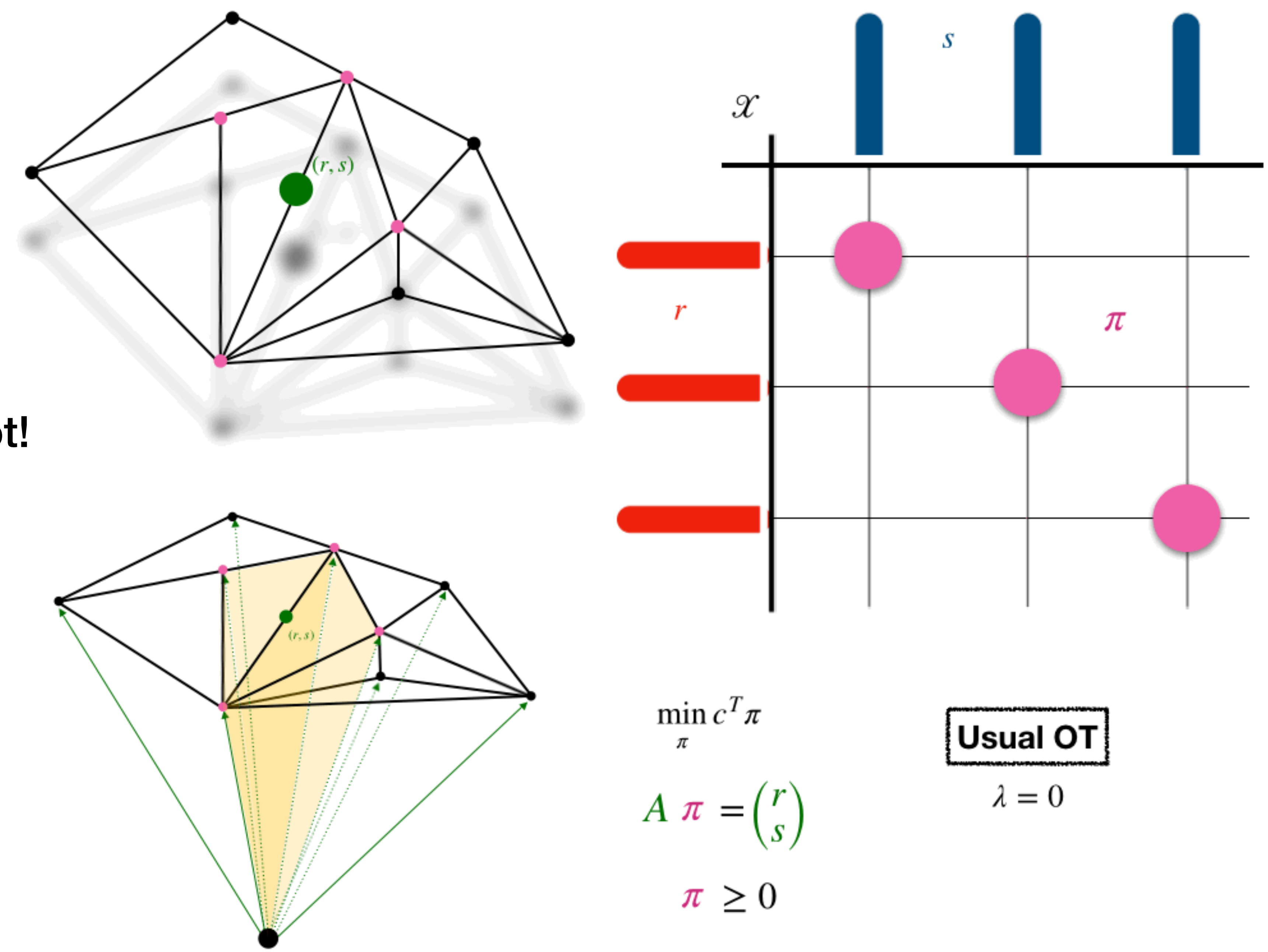
$\lambda = 0$

Let us recall... $\boxed{\sqrt{n}\left\{\pi_\lambda(\hat{r}_n, s) - \pi_\lambda(r, s)\right\} \xrightarrow{\mathscr{D}} \mathcal{N}_{N^2}\left(0, \Sigma_\lambda(r \mid s)\right)}$



Can we hope for a similar behavior for the empirical OT plan? $(\lambda = 0)$

...but usually not!



Limit Theory for the OT plan $\lambda = 0$

Primal

$$\min_{\pi \in \Pi(r,s)} \sum_{i,j} d_{ij}^p \pi_{ij}$$

Dual

$$\max_{\alpha,\beta \in \mathbb{R}^N} \sum_i \alpha_i r_i + \sum_j \beta_j s_j$$

$$s.t. \quad \alpha_i + \beta_j \leq d_{ij}^p$$

Assumption I : Unique OT plan

Assumption II : Optimal dual solutions nondegenerate

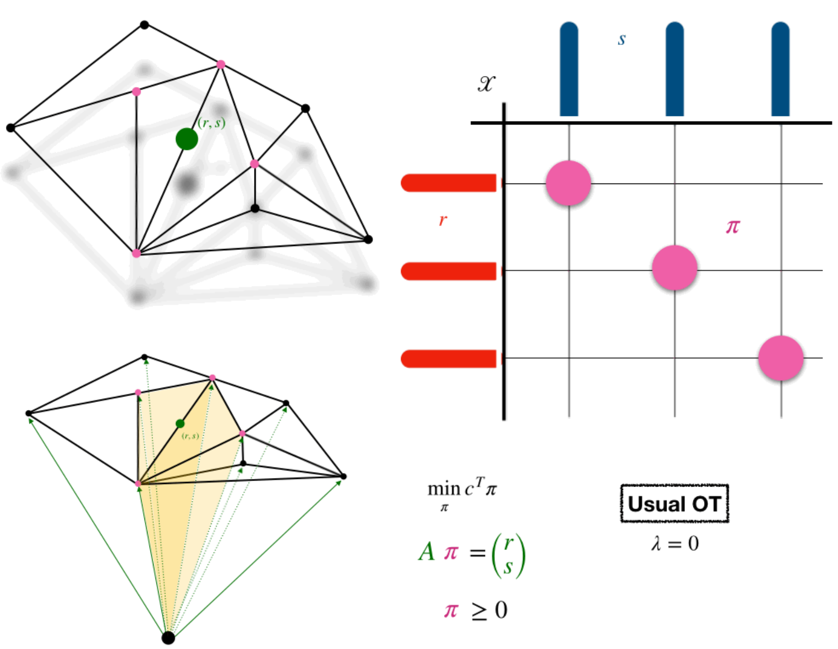
Theorem (K., Munk & Zemel (2020)): As sample size *n* approaches infinity

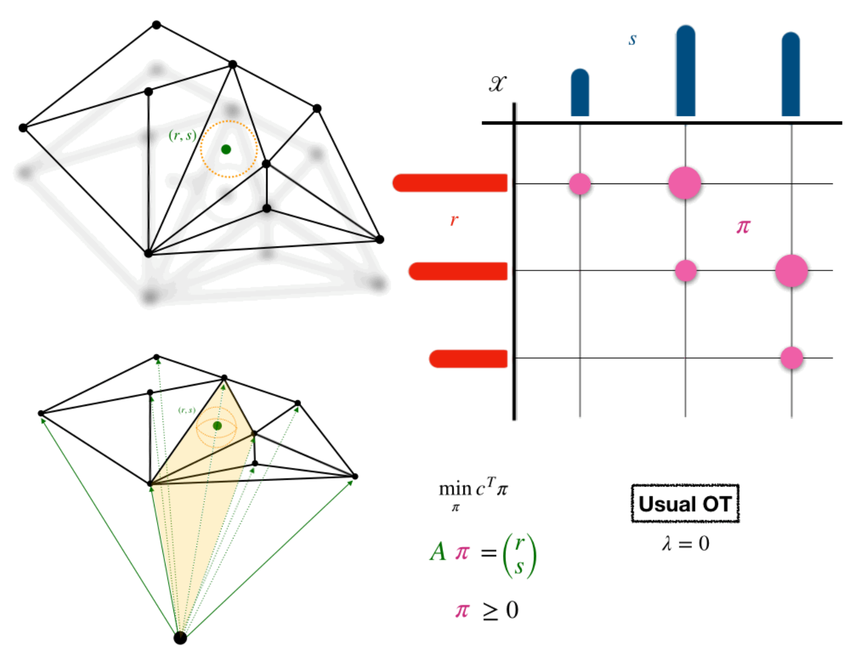
$$\sqrt{n}\,\{\pi(\hat{r}_n, s) - \pi(r, s)\} \xrightarrow{\mathcal{D}} \sum_{k=1}^{K} 1_{G \in H_k \setminus \bigcup_{j<k} H_j}\, \pi(I_k, G)$$

- OT plan degenerate

$$\sqrt{n}\,\{\pi(\hat{r}_n, s) - \pi(r, s)\} \xrightarrow{\mathcal{D}} \pi(I_1, G)$$

- OT plan nondegenerate





Limit Theory for the OT plan $\lambda = 0$

Primal

$$\min_{\pi \in \Pi(r,s)} \sum_{i,j} d_{ij}^p \pi_{ij}$$

Dual

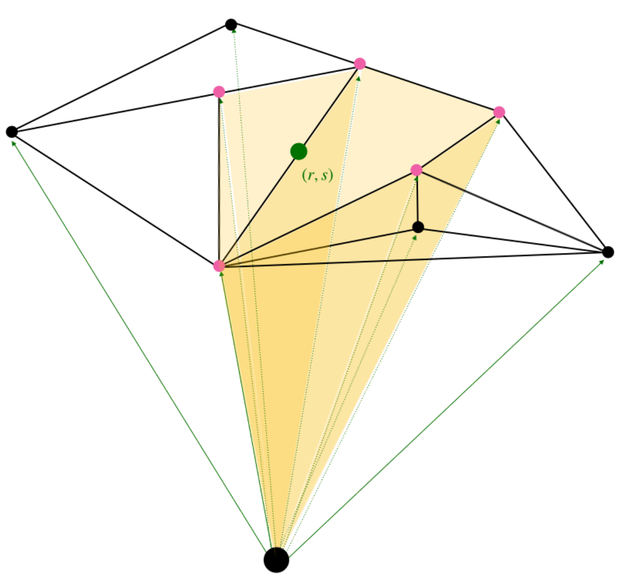
$$\max_{\alpha,\beta \in \mathbb{R}^N} \sum_i \alpha_i r_i + \sum_j \beta_j s_j$$
$$s.t. \quad \alpha_i + \beta_j \leq d_{ij}^p$$

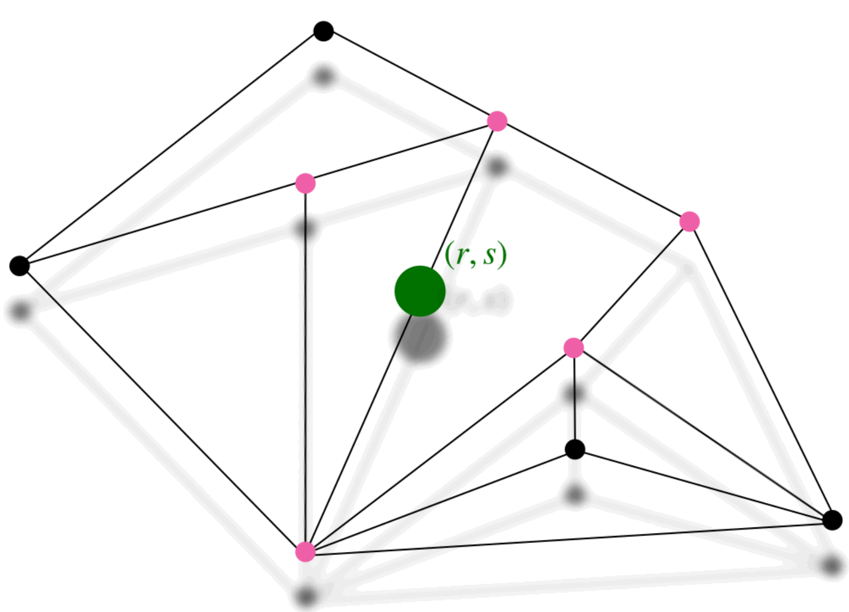
Assumption I : Unique OT plan

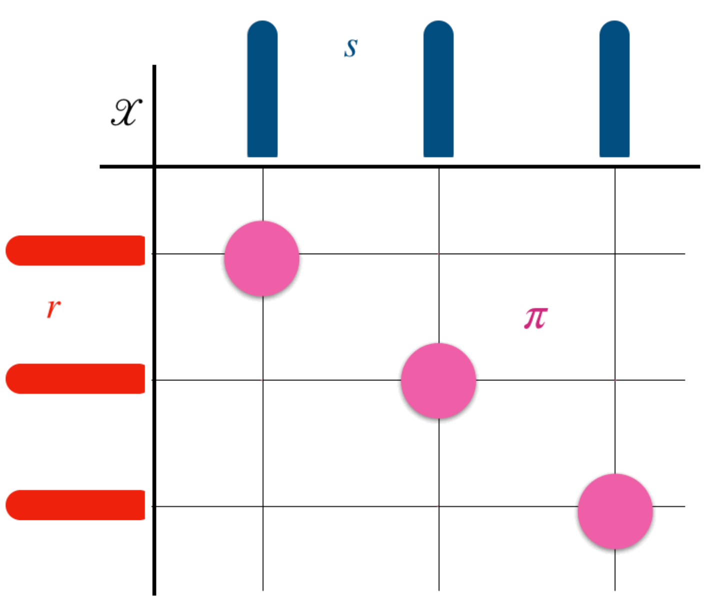
~~Assumption II : Optimal dual solutions nondegenerate~~

Theorem (K., Munk & Zemel (2020)): As sample size *n* approaches infinity

$$\sqrt{n}\,\{\pi(\hat{r}_n, s) - \pi(r, s)\} \xrightarrow{\mathscr{D}} \sum_{\mathscr{K}} 1_{G \in H_{\mathscr{K}} \setminus \bigcup_{k \notin \mathscr{K}} H_j} \alpha^{\mathscr{K}} \otimes \pi(I_{\mathscr{K}}, G)$$







Limit Theory for the OT plan $\lambda = 0$

Primal

$$\min_{\pi \in \Pi(r,s)} \sum_{i,j} d_{ij}^p \pi_{ij}$$

Dual

$$\max_{\alpha,\beta \in \mathbb{R}^N} \sum_i \alpha_i r_i + \sum_j \beta_j s_j$$
$$s.t. \quad \alpha_i + \beta_j \leq d_{ij}^p$$

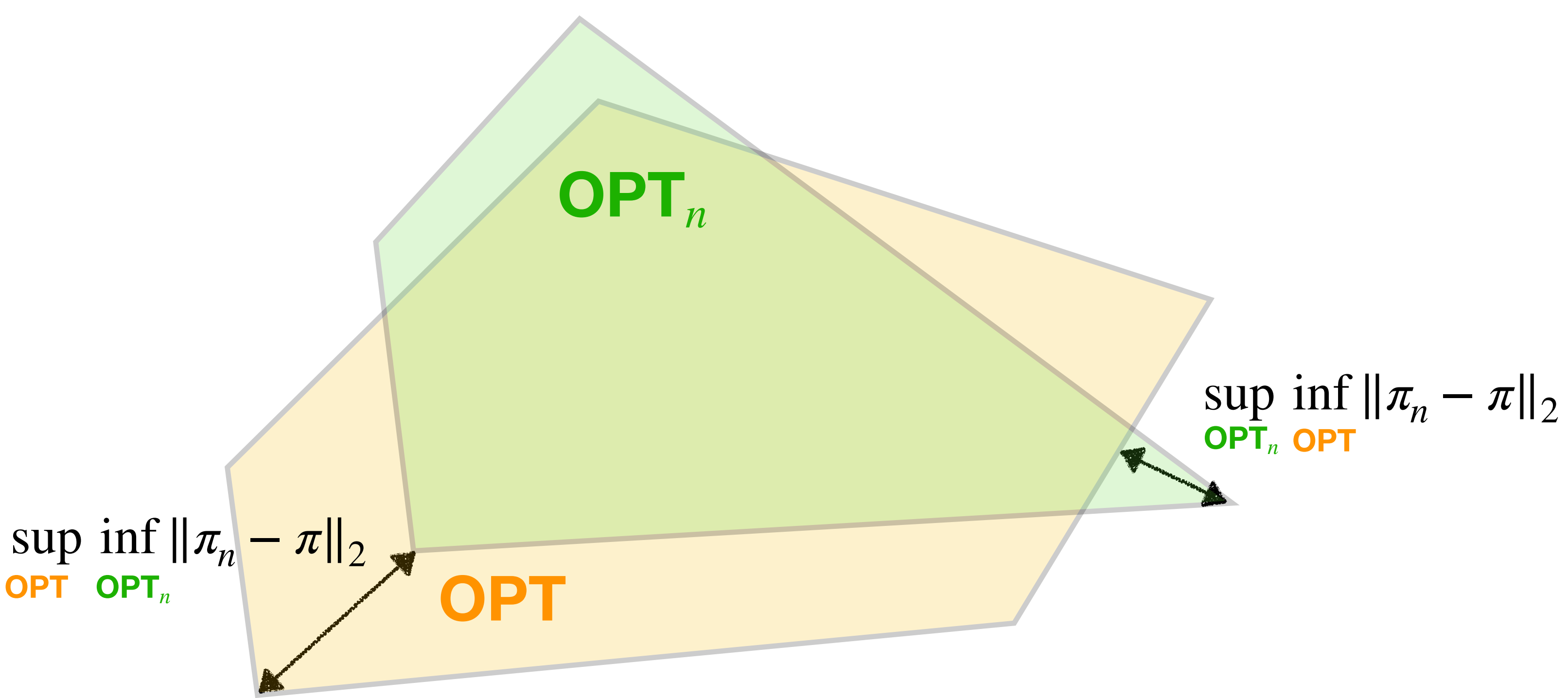
~~Assumption I : Unique OT plan~~

~~Assumption II : Optimal dual solutions nondegenerate~~

<u>Theorem (K., Munk & Zemel (2020)):</u>

$$d_H\left(\mathbf{OPT}_n, \mathbf{OPT}\right) = O_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right)$$

Hausdorff-distance



Overview: (R)OT limit laws for finite metric spaces

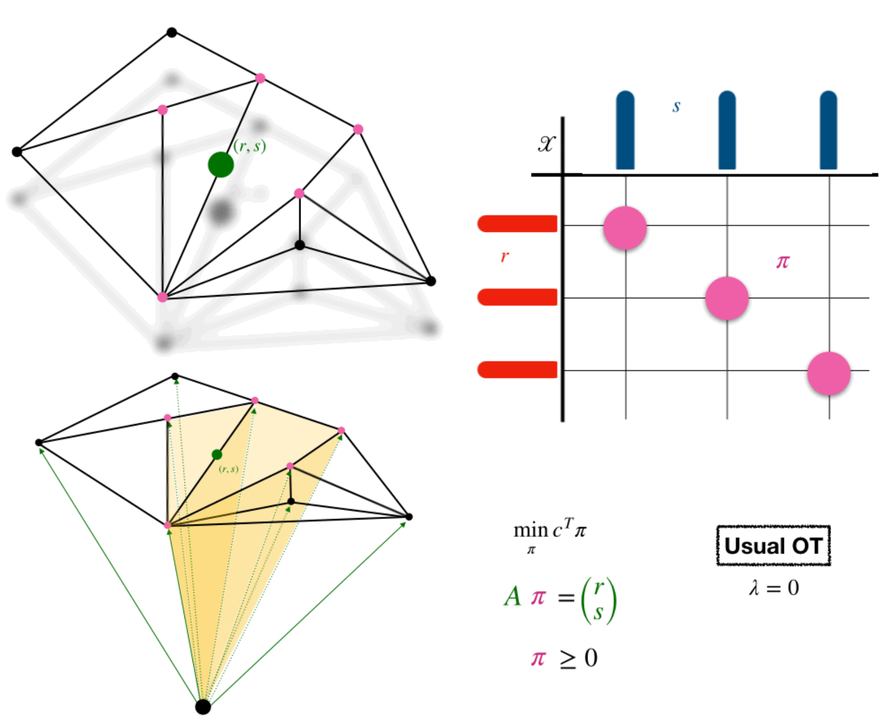
$$\lambda = 0$$

OT plan

$$\sqrt{n}\,\left\{\pi(\hat{r}_n, s) - \pi(r, s)\right\} \xrightarrow{\mathscr{D}} \sum_{\mathscr{K}} 1_{G \in H_{\mathscr{K}} \setminus \bigcup_{k \notin \mathscr{K}} H_j}\, \alpha^{\mathscr{K}} \otimes \pi(I_{\mathscr{K}}, G)$$

Wasserstein distance

$$\sqrt{n}\left\{W_p^p(\hat{r}_n, s) - W_p^p(r, s)\right\} \xrightarrow{\mathscr{D}} \max_{\alpha_\star \in \Gamma_\star} \langle G, \alpha_\star \rangle$$



$$\lambda > 0$$

ROT plan

$$\sqrt{n}\,\left\{\pi_\lambda(\hat{r}_n, s) - \pi_\lambda(r, s)\right\} \xrightarrow{\mathscr{D}} \mathscr{N}_{N^2}\left(0, \Sigma_\lambda(r \mid s)\right).$$

Sinkhorn divergence

$$\sqrt{n}\left\{W_{p,\lambda}^p(\hat{r}_n, s) - W_{p,\lambda}^p(r, s)\right\} \xrightarrow{\mathscr{D}} \langle G, \alpha_\lambda \rangle$$