Stochastic Optimization for Large-scale Optimal Transport

Aude Genevay
CEREMADE,
Univ. Paris-Dauphine
INRIA – Mokaplan project-team
genevay@ceremade.dauphine.fr

Marco Cuturi
Kyoto University
mcuturi@i.kyoto-u.ac.jp

Gabriel Peyré
CNRS and CEREMADE,
Univ. Paris-Dauphine
INRIA – Mokaplan project-team
gabriel.peyre@ceremade.dauphine.fr

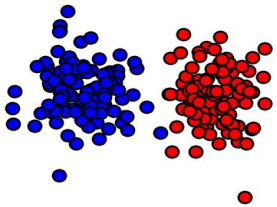
Francis Bach
INRIA – Sierra project-team
Département d'Informatique de l'ENS
(CNRS/ENS/INRIA)
francis.bach@inria.fr

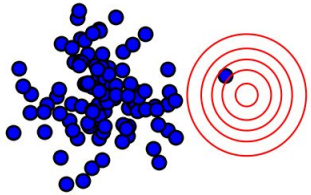
May 30, 2016

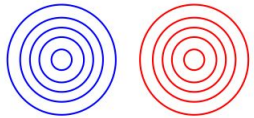
MVA - Computational Optimal Transport
Eugène Berta
09/01/2023

Presentation of the problem

- How to compute OT in different settings ?







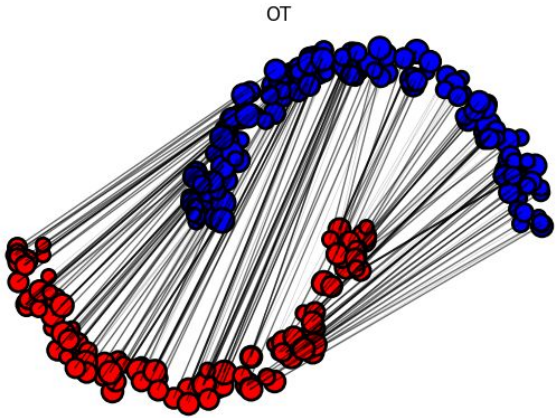
- Using stochastic algorithms !

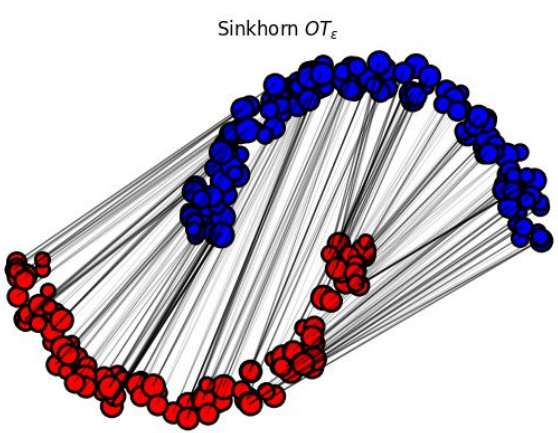
$$OT_\epsilon(\alpha,\beta) = \max_{f,g\in\mathcal{C}(\mathcal{X})\times\mathcal{C}(\mathcal{Y})} \mathbb{E}_{X,Y}\left[f(x)+g(y)-\epsilon exp\left(\frac{f(x)+g(y)-C(x,y)}{\epsilon}\right)\right]$$

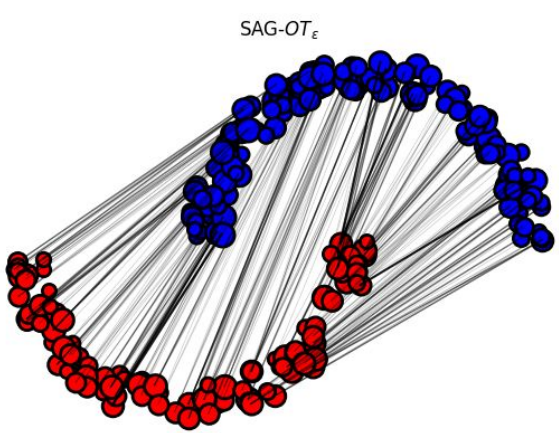
$$= \max_{f\in\mathcal{C}(\mathcal{X})} \mathbb{E}_Y\left[\int_{\mathcal{X}} f d\alpha + f^{c,\epsilon}(y) - \epsilon\right]$$

Discrete Optimal Transport

$$OT_\epsilon(\alpha,\beta) = \max_{f\in\mathbb{R}^n} \mathbb{E}_Y\Big[\sum_{i=1}^{n} f_i\alpha_i - \epsilon\log\Big(\sum_{i=1}^{n} exp\Big(\frac{f_i - C(x_i,y_j)}{\epsilon}\Big)\alpha_i\Big) - \epsilon\Big]$$

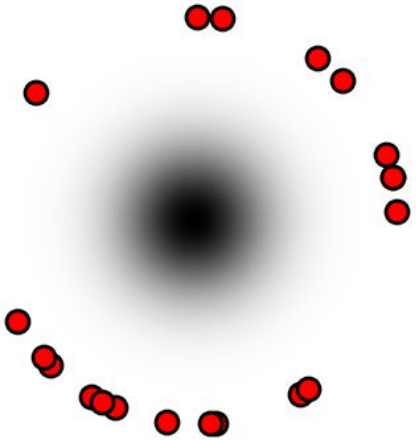


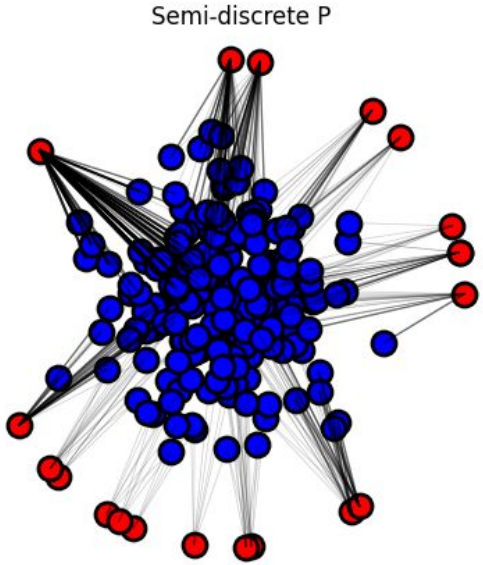


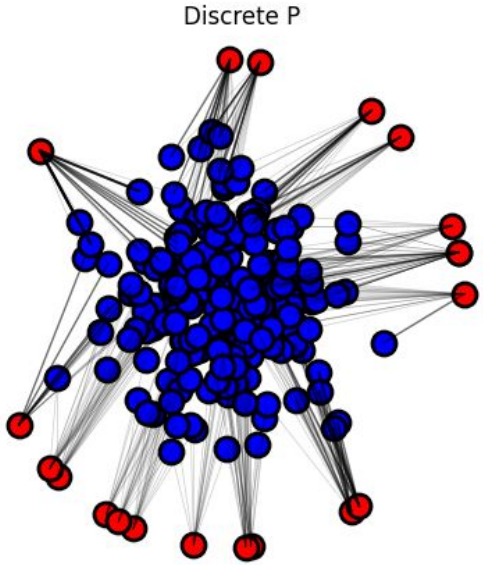


Semi discrete Optimal Transport

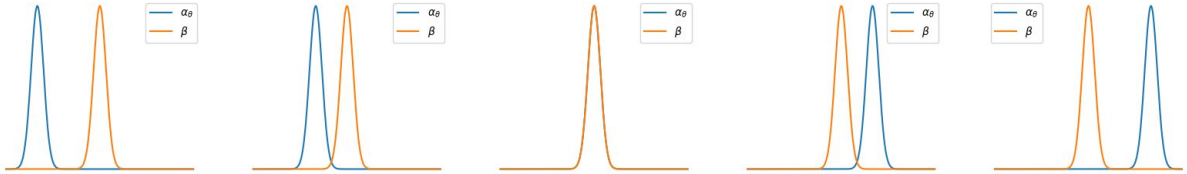
$$OT_\epsilon(\alpha,\beta) = \max_{f\in\mathbb{R}^n} \mathbb{E}_Y \Big[\sum_{i=1}^{n} f_i\alpha_i - \epsilon \log \Big(\sum_{i=1}^{n} exp\Big(\frac{f_i - C(x_i,y_j)}{\epsilon}\Big)\alpha_i \Big) - \epsilon \Big]$$

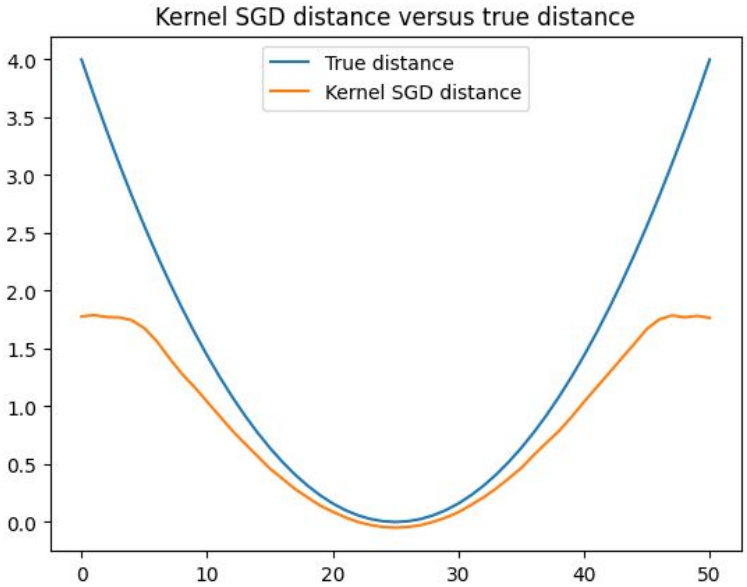






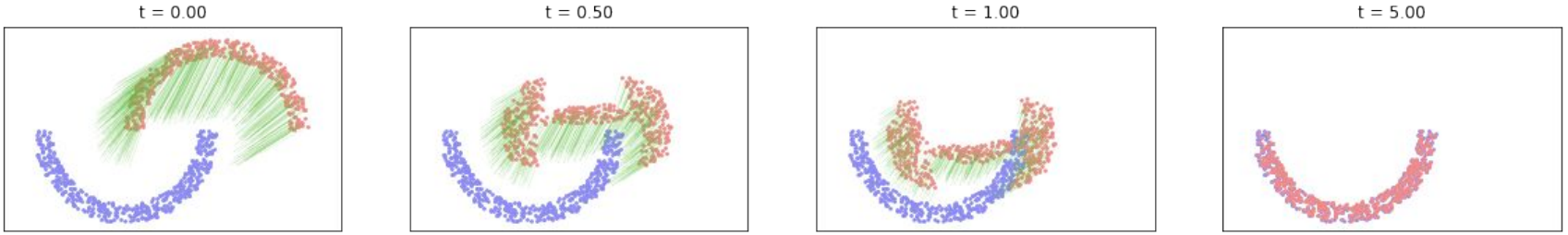
Continuous Optimal Transport





Critics

- The **discrete SAG** is the most important contribution of the paper.
- **Lack of applications** for semi-discrete and continuous cases.
- The paper targets OT applications in Data Science, but it disregards the **entropic bias** and **topology preservation issues** when fitting distributions with OT.



Conclusion

- Big step in **democratizing OT** by expressing it as a **stochastic optimisation problem** → Useful for large scale solvers.

- Many possible improvements for the continuous case in application specific research ?

- Nowadays : OT at scale is almost solved, mostly a discrete problem, continuous “to the limits”.