Computational Optimal Transport
Lab Sessions

Simon Queric simon.queric@telecom-paris.fr

January 18, 2024

Abstract

In this lab session report, we play with the very fondamental notions of Optimal Transport.

1 Optimal Transport with Linear programming

Two gaussians rotating in relation to each other

In this report, I write $X_{1:n} = (X_k)_{1\leqslant k\leqslant n}$. I computed the optimal transport between two point clouds $X_{1:n}, Y_{1:n}$ such that

$$\begin{cases} X_{1:n} \overset{i.i.d}{\sim} \mathcal{N}(0, \Sigma) \\ Y_{1:n} \overset{i.i.d}{\sim} \mathcal{N}(0, R_\theta^T \Sigma R_\theta) \end{cases}$$

where $\Sigma = \text{diag}(5 \times 10^{-2}, 20)$, $\theta = \pi/4$.

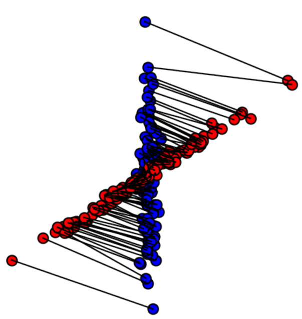

Figure 1: OT between two gaussians rotating one relative to the other.

We can see here that the OT should be a linear map because of the (more or less) parallel straigth lines. Indeed, according to Brenier's theorem, we know that the OT between gaussians α and β is a linear map $T : x \mapsto m_\beta + A(x - m_\alpha)$, where $A = \Sigma_\alpha^{-1/2} \left(\Sigma_\alpha^{1/2} \Sigma_\beta \Sigma_\alpha^{1/2}\right)^{1/2} \Sigma_\alpha^{-1/2}$. Especially, OT won't be a rotation in that case.

OT between a gaussian and a gaussian mixture with two components

I then computed Optimal Transport and the displacement interpolation for two point clouds $X_{1:n}, Y_{1:n}$ such that

$$\begin{cases} X_{1:n} \overset{i.i.d}{\sim} \mathcal{N}(0, \sigma^2 I_2) \\ Y_{1:n} \overset{i.i.d}{\sim} \frac{1}{2}\mathcal{N}\left((1,1)^T, \sigma^2 I_2\right) + \frac{1}{2}\mathcal{N}\left((-1,-1)^T, \sigma^2 I_2\right) \end{cases}$$

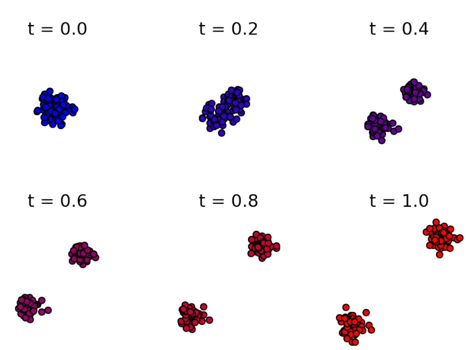

Figure 2: Displacement interpolation between a gaussian and a gaussian mixture with two components in $\mathbb{R}^2$.

Both examples illustrate the fact that O.T can handle distributions whose supports are disjoint or with no inclusion relation.

Optimal matching of cafes and bakeries

Thanks to OpenStreetMap API, I took 46 random cafes and bakeries located in the 5th arrondissement of Paris. Bakeries need to bring croissants to cafes. The goal is to match bakeries to cafes by minimizing the sum of the distances which is exactly an optimal transport problem. Heuristically, we can take the geodesic distance, which is different from the walking distance. I think the result is quite close to the optimal matching by taking the walking distance. Furthermore, it's quite time consuming to compute each minimal path in the graph of street network of Paris. Indeed, complexity of Dijkstra's algorithm is $O((|E|+|V|)\log|V|)$ while complexity of computing geodesic distance is $O(1)$.

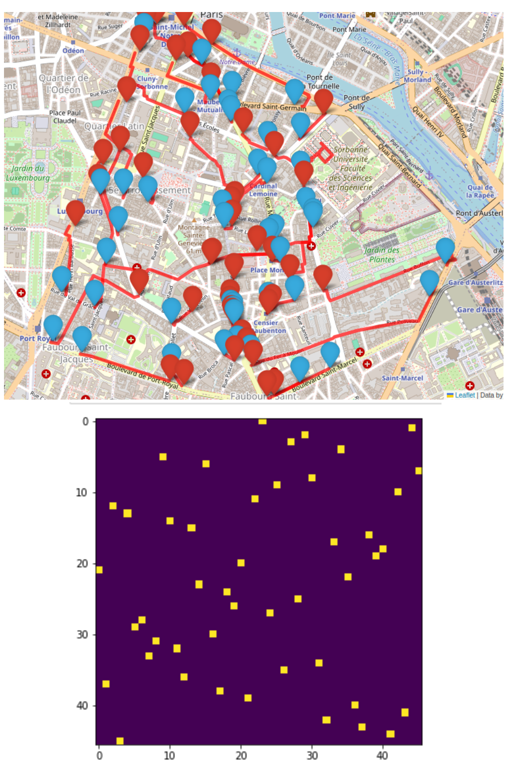

Figure 3: Above : Optimal matching of cafes and bakeries in the fifth arrondissement (bakeries are in blue, cafes in red). Below : Matrix of permutation of the optimal matching.

2 Entropic Regularization of Optimal Transport

Kantorovitch relaxation is a version of Optimal Transport that overcomes drawbacks of the deterministic Monge formulation, which leads to non-convex problems, by replacing the monge map $T : x \in \mathcal{X} \mapsto \mathcal{Y} \ni y$ by a coupling of α and β, which leads to a fuzzy version of O.T. The formulation is the following :

$$W(\alpha,\beta) = \min_{\pi\in\Pi(\alpha,\beta)} \int_{\mathcal{X}\times\mathcal{Y}} c(x,y)\pi(dx,dy)$$

From a computational point of view, the solution of entropic regularized version is easy to compute since it requires a simple alternate minimization scheme which corresponds to the so-called Sinkhorn's algorithm.

$$W_\varepsilon(\alpha,\beta) = \min_{\pi\in\Pi(\alpha,\beta)} \int_{\mathcal{X}\times\mathcal{Y}} c(x,y)\pi(dx,dy) + \varepsilon H(\pi \mid \alpha\otimes\beta)$$

Effect of the regularizer parameter ε

Intuitively, the bigger ε the fuzzier the transportation plan is. When $\varepsilon = +\infty$, the coupling is equal to the product measure $\alpha\otimes\beta$ while the case $\varepsilon = 0$ is the Kantorovitch OT problem.

I computed the coupling for various values of ε (Figure 5) to match two one dimensional distributions (Figure 4).

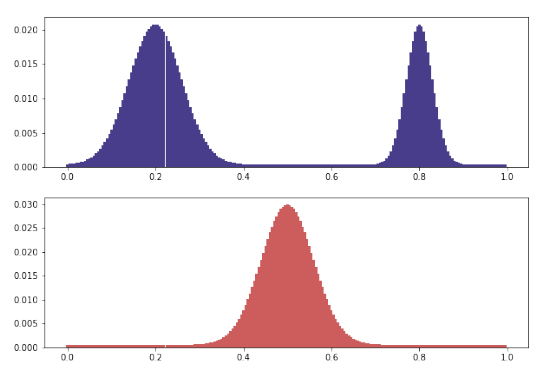

Figure 4: 1D distributions.

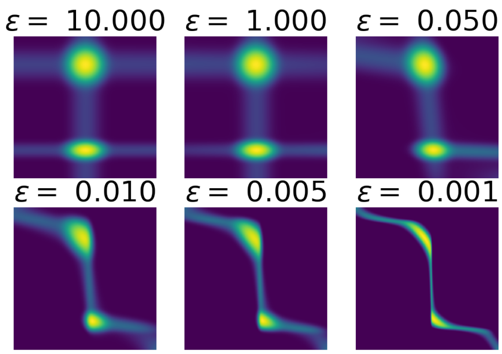

Figure 5: Effect of the regularizer parameter ε on the transport plan.

In particular, when $\varepsilon = 10$ we get the product measure $\alpha \otimes \beta$.

Barycentric projection map

We can compute the barycentric projection map :

$$x_i \mapsto \frac{\sum_j P_{i,j} y_j}{\sum_j P_{i,j}} = \frac{1}{a_i} \sum_j P_{i,j} y_j$$

which is an approximation of the Monge map for $\varepsilon = 0.005$.

We can display it with the coupling :

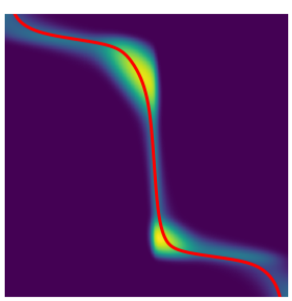

Figure 6: Barycentric projection map between the two distributions.

Wasserstein barycenters

Given a set of probability distribution $(\alpha_k)_{1 \leqslant k \leqslant K}$, we would like to compute a barycenter with weights $(\lambda_k)_{1 \leqslant k \leqslant K} \in \mathbb{R}_+^K$ such that $\sum_k \lambda_k = 1$. The *Wasserstein barycenter* is defined by $\overline{\alpha} = \min\limits_{\gamma} \sum_k \lambda_k W(\alpha_k, \gamma)$.

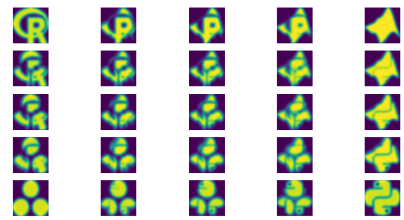

Figure 7: Shape interpolation with Wasserstein barycenters.

No free lunch principle

One can prove that $W_\varepsilon(\alpha, \beta) \underset{\varepsilon \to 0}{\longrightarrow} W(\alpha, \beta)$ so that we could use Sinkhorn to approximate the standard OT for small value of ε. It turns out that the convergence of Sinkhorn algorithm gets slower for small ε so that there is *no free lunch* and we cannot go faster than linear programming. Let's show this behavior in practice :

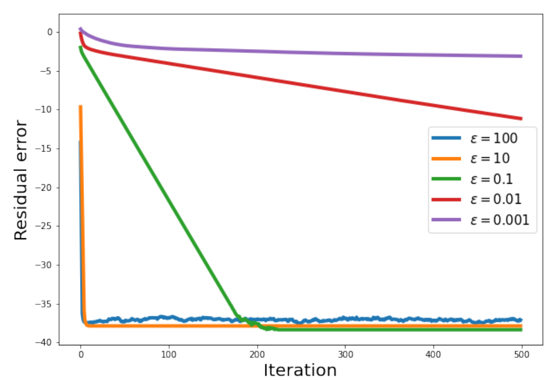

Figure 8: Empirical convergence of Sinkhorn's algorithm on constraints : residuals $\|P_n 1 - a\|$. Small ε leads to slow convergence.

References

- Gabriel Peyré's Numerical tours on Optimal Transport
- Gabriel Peyré and Marco Cuturi's textbook on Computational Optimal Transport.