Optimal Transport Numerical Tour

Pierre Osselin

ENS Paris-Saclay, 61 Avenue du Président Wilson, 94235 Cachan

December 10, 2019

1 Introduction

2 Linear Programming

2.1 Optimal Transport of Discrete Distribution

We consider two discrete distributions

$$\forall k = 0, 1, \mu_k = \sum_{i=1}^{n_k} p_{k,i}\delta_{x_{k,i}} \tag{1}$$

where n_0, n_1 are the number of points, δ_x is the Dirac at location $x \in \mathbb{R}^d$, and $X_k = (x_{k,i})_{i=1}^{n_k} \subset \mathbb{R}^d$ for $k = 0, 1$ are two point clouds.
We define the set of couplings between μ_0, μ_1 as

$$\mathcal{P} = \{(\gamma_{i,j})_{i,j} \in (\mathbb{R}^{+d})^{n_0 \times n_1},\ \forall i,\ \sum_j \gamma_{i,j} = p_{0,i},\ \forall j,\ \sum_i \gamma_{i,j} = p_{1,j}\} \tag{2}$$

The Kantorovitch formulation of the optimal transport reads

$$\gamma^\star \in \underset{\gamma \in \mathcal{P}}{\arg\min} \sum_{i,j} \gamma_{i,j} C_{i,j} \tag{3}$$

where $C_{i,j} \geq 0$ is the cost of moving some mass from $x_{0,i}$ to $x_{1,j}$. We will focus on the L^2 Wasserstein distance:

$$C_{i,j} = ||x_{0,i} - x_{1,j}||^2 \tag{4}$$

The L^2 Wasserstein distance is then defined as

$$W_2(\mu_0, \mu_1)^2 = \sum_{i,j} \gamma^\star_{i,j} C_{i,j} \tag{5}$$

Our optimisation problem is a linear programming problem with $n_0 + n_1$ equality constraints and $n_0 \times n_1$ non-negativity inequality constraints. To solve it we will use the simplex algorithm [Dantzig, 1987], consisting in visiting certain vertices of the simplex of the feasible region that reduce the cost function.

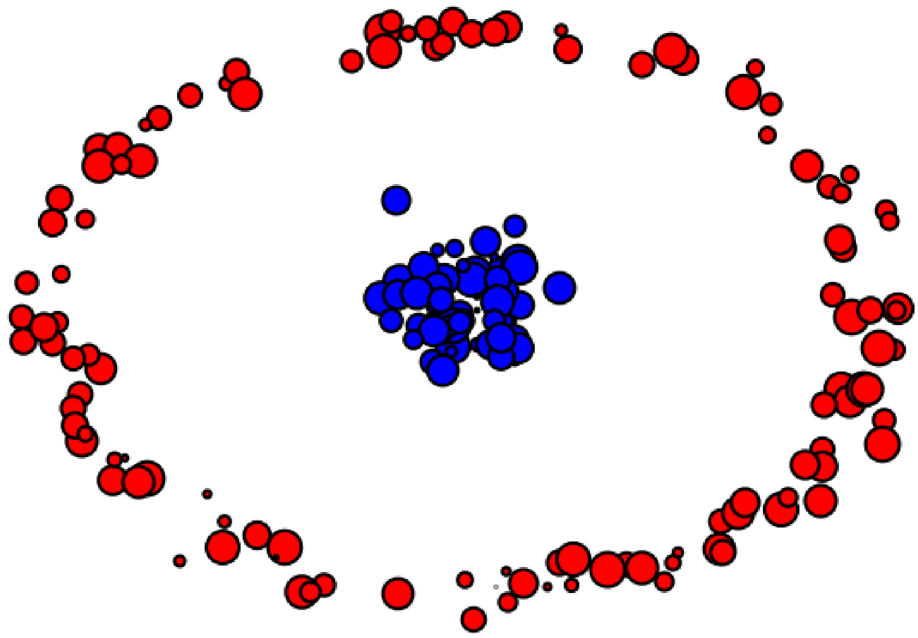

Figure 1: Data used for optimal transport

2.2 Application on Data

To illustrate the optimisation process we take the two following discrete measures, displayed in figure 1:

1. The first measure is 60 points uniformly drawn from a single gaussian distribution of mean $\mu = 0$ and standars deviation $\sigma = 0.1$.

2. The second one is 120 points drawn from a circle of radius one with added noise of standard deviation 0.05

After applying the simplex algorithm to find an optimal map, we effectively find a transport that satisfies the constraints 2 and has exactly $n_0 + n_1 - 1$ non-zero entries.

Once the optimal transport $\gamma^\star$ is computed, one can formulate the displacement interpolation between the two measure. For any $t \in [0, 1]$, one can define a distribution μ_t such that $t \to \mu_t$ defines a geodesic for the Wasserstein metric, solving the following problem:

$$\mu_t = \arg\min_\mu (1-t) W_2(\mu_0, \mu)^2 + t W_2(\mu_1, \mu)^2 \tag{6}$$

Which, in our case, can be computed as such, and displayed in 2 for our data:

$$\mu_t = \sum_{i,j} \gamma^\star_{i,j} \delta_{(1-t)x_{0,i} + t x_{1,j}} \tag{7}$$

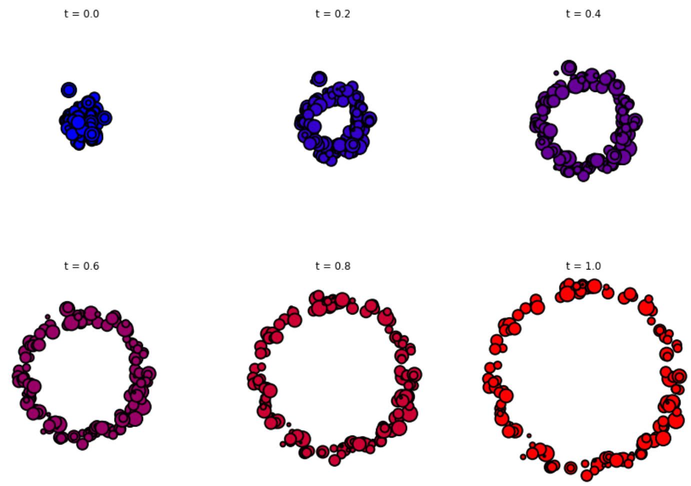

Figure 2: Displacement interpolation for our data

2.3 Optimal Assignement

In the case where the weights $p_{0,i} = 1/n, p_{1,i} = 1/n$ (where $n_0 = n_1 = n$) are constants, one can show that the optimal transport coupling is actually a permutation matrix. This properties comes from the fact that the extremal point of the polytope $\mathcal{C}$ are permutation matrices. This means that there exists an optimal permutation $\sigma^\star \in \mathcal{P}_n$ such that :

$$\gamma^\star_{i,j} = \begin{cases} 1 & \text{if } j = \sigma^\star(i) \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

where Σ_n is the set of permutation (bijections) of $\{1, \ldots, n\}$.

This permutation thus solves the so-called optimal assignement problem

$$\sigma^\star \in \underset{\sigma \in \Sigma_n}{\arg\min} \sum_i C_{i,\sigma(j)} \tag{9}$$

To illustrate this case, we took the datasets moons from the package Scikit-learn and compluted the optimal coupling with regards to the two half-circles. The data are displayed figure 3. After computation, one can display the permutation matrix and the corresponding coupling. In our case, this is displayed figure 4 and 5 respectively.

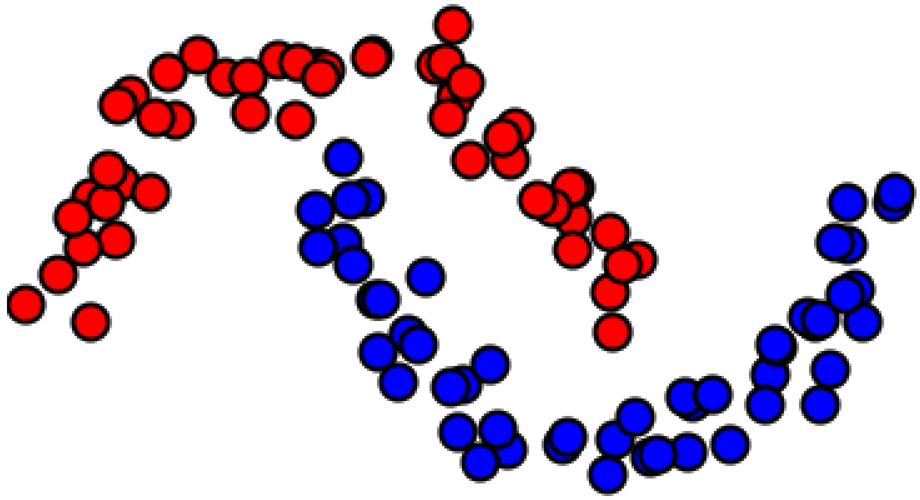

Figure 3: Data used for coupling

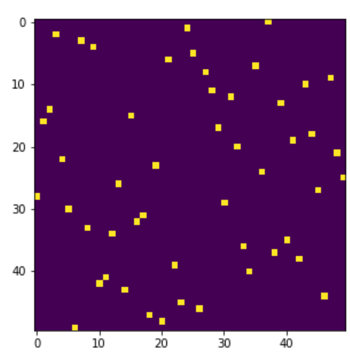

Figure 4: Optimal permutation matrix

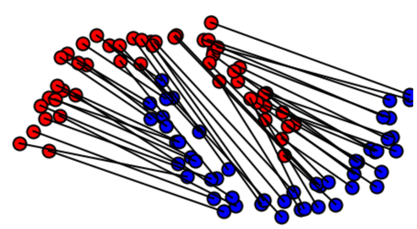

Figure 5: Resulting optimal coupling

References

[Dantzig, 1987] Dantzig, G. B. (1987). Origins of the simplex method. Technical report, Stanford Univ CA Systems Optimization Lab.