

Discrete optimal transport

Short introduction to optimal transport (OT)

Data and definitions

OT aims at computing a minimal cost transportation between a source probability distribution μ_s , and a target distribution μ_t . Here we consider the case when those distributions are only available through a finite number of samples, and can be written as

$$\mu_s = \sum_{i=1}^{n_s} p_i^s \delta_{\mathbf{x}_i^s}, \quad \mu_t = \sum_{i=1}^{n_t} p_i^t \delta_{\mathbf{x}_i^t}$$

where $\delta_{\mathbf{x}_i}$ is the Dirac at location $\mathbf{x}_i \in \mathbb{R}^d$. p_i^s and p_i^t are probability masses associated to the i -th sample, belonging to the probability simplex, i.e. $\sum_{i=1}^{n_s} p_i^s = \sum_{i=1}^{n_t} p_i^t = 1$. The source and target samples are in matrices $\mathbf{X}_s = [\mathbf{x}_1^s, \dots, \mathbf{x}_{n_s}^s]^\top \in \mathbb{R}^{n_s \times d}$ and $\mathbf{X}_t = [\mathbf{x}_1^t, \dots, \mathbf{x}_{n_t}^t]^\top \in \mathbb{R}^{n_t \times d}$, respectively. The set of probabilistic couplings between these two distributions is then the set of doubly stochastic matrices \mathcal{P} defined as

$$\mathcal{P} = \{ \gamma \in (\mathbb{R}^+)^{n_s \times n_t} \mid \gamma \mathbf{1}_{n_t} = \mu_s, \gamma^\top \mathbf{1}_{n_s} = \mu_t \}$$

where $\mathbf{1}_d$ is a d -dimensional vector of ones.

Optimisation problem

The Kantorovitch formulation of the optimal transport is:

$$\gamma_0 = \arg \min_{\gamma \in \mathcal{P}} \langle \gamma, \mathbf{C} \rangle_F$$

where $\langle \cdot, \cdot \rangle_F$ is the Frobenius dot product and $\mathbf{C} \geq 0$ is the cost function matrix of term $C(i, j)$ related to the energy needed to move a probability mass from \mathbf{x}_i^s to \mathbf{x}_j^t . In our setting, this cost is chosen as the Euclidian distance between the two locations, i.e. $C(i, j) = \|\mathbf{x}_i^s - \mathbf{x}_j^t\|^2$, but other types of metric could be considered, such as Riemannian distances over a manifold.

One can also add a regularization term to the optimal transport. In this case the problem becomes

$$\gamma_0 = \arg \min_{\gamma \in \mathcal{P}} \langle \gamma, \mathbf{C} \rangle_F + \lambda \Omega(\gamma)$$

where $\Omega(\gamma)$ is a regularization term that can be for instance an entropy-based regularization such as

$$\Omega(\gamma) = \sum_{i,j} \gamma_{i,j} \log(\gamma_{i,j})$$

This term has been recently proposed by Cuturi [2] as a way to regularize the transport and the optimization problem is much more easy to solve thanks to an efficient algorithm called Sinkhorn-Knopp.

Interpolation

When the optimal transportation plan γ has been estimated, one can interpolate the distributions from the source to the target. The natural interpolation corresponds to a displacement of the diracs from the source to target distribution

$$\mu(t) = \sum_{i=1, j=1}^{n_s, n_t} \gamma_{i,j} \delta_{(1-t)\mathbf{x}_i^s + t\mathbf{x}_j^t}$$

where $t \in [0, 1]$ is the parametrizes the transport from the source distribution μ_s (when $t = 0$) to the target distribution μ_t (when $t = 1$).

One can also transport the source distribution by expressing the new transported distribution $\hat{\mathbf{X}}_s$ as linear combination of elements from \mathbf{X}_t . The transportation of the source samples \mathbf{X}_s onto that target samples \mathbf{X}_t is computed with:

$$\mathbf{T}_{\gamma_0}(\mathbf{X}_s) = \text{diag}(\gamma_0 \mathbf{1}_{n_t})^{-1} \gamma_0 \mathbf{X}_t$$

The interpolation is then expressed as

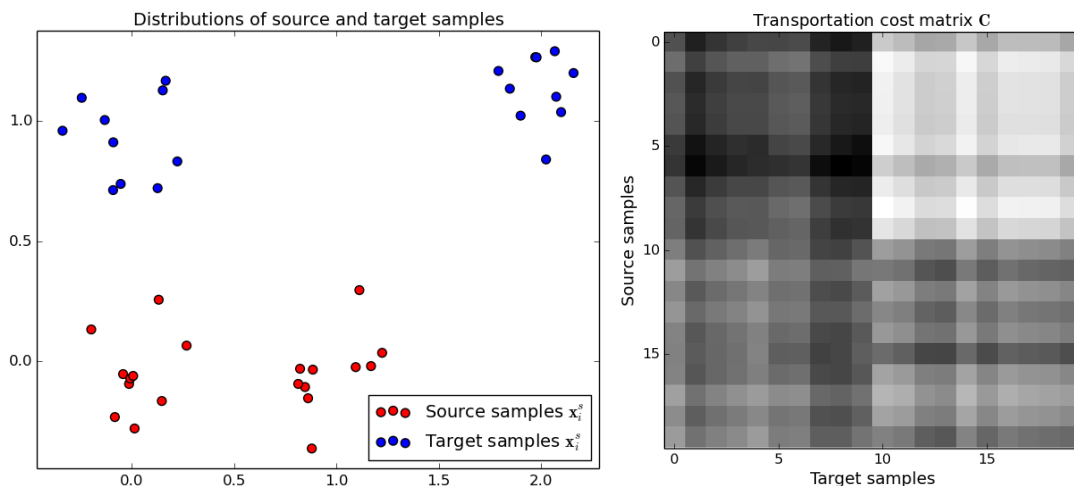
$$\hat{\mathbf{X}}_s(t) = (1 - t)\mathbf{X}_s + t\mathbf{T}_{\gamma_0}(\mathbf{X}_s)$$

Those two interpolations have been used in numerous publications and are illustrated in the following.

Demo

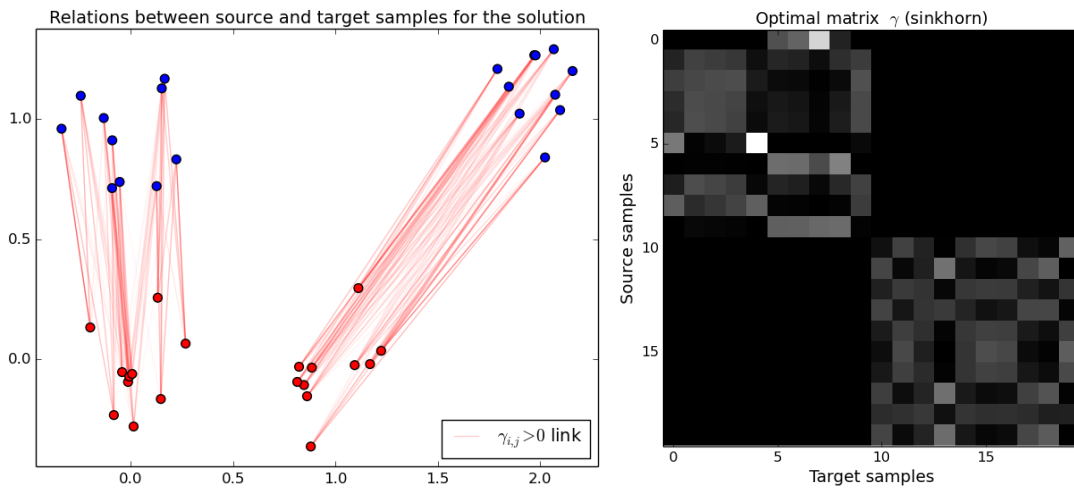
We illustrate the OT problem solution and resulting interpolation for a simples 2D dataset in the following. In this particular case the number of source and target samples are equal and the probability masses are all equal to .

Data set

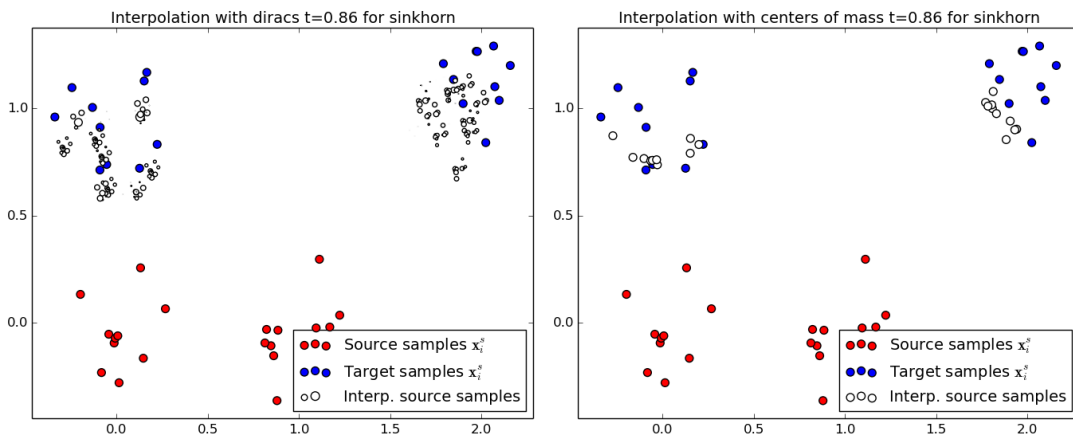


On the left, we plot the 2D position of the toy example, when the mouse pointer is on the image, you can see the index of each point. On the right, the corresponding cost matrix \mathbf{C} , the xias correspond to the index of the samples on the right.

Problem solution



Interpolation



Optimisation problem

- ☐ No regul. (LP)
- ☒ Entropic regul. (Sinkhorn)

Interpolation time t in $[0,1]$



This section illustrates the two interpolation procedures (diracs and center of mass) for two different regularization schemes. When no regularization is used, the solution is a permutation matrix and the mass is transferred from a unique source sample to a unique target sample. When an entropic regularization is used, the transportation matrix is not sparse anymore and the mass is spread onto several target samples. Use the controls at the end of the page to change the optimal transport regularization and interpolation time $t \in [0, 1]$.

References

A complete reference to Optimal Transport has been written by C. Villani [1] and is available on the web. The entropic regularization has been proposed in the machine learning community by M. Cuturi [2]. Other regularization such as Laplacian regularization have been proposed in [3].

[1] C. Villani. **Optimal transport: old and new** (<http://cedricvillani.org/wp-content/uploads/2012/08/preprint-1.pdf>). Grundlehren der mathematischen Wissenschaften. Springer, 2009.

- [2] M. Cuturi. **Sinkhorn distances: Lightspeed computation of optimal transportation** (<http://arxiv.org/pdf/1306.0895.pdf>). In NIPS, pages 2292–2300. 2013.
- [3] S. Ferradans, N. Papadakis, J. Rabin, G. Peyré, and J.-F. Aujol. **Regularized discrete optimal transport** (<http://arxiv.org/pdf/1307.5551v1.pdf>). In Scale Space and Variational Methods in Computer Vision, SSVM, pages 428–439, 2013.
-