HAL
open science

Optimal transport with Laplacian regularization

Rémi Flamary, Nicolas Courty, Alain Rakotomamonjy, Devis Tuia

To cite this version:

Rémi Flamary, Nicolas Courty, Alain Rakotomamonjy, Devis Tuia. Optimal transport with Laplacian regularization. NIPS 2014, Workshop on Optimal Transport and Machine Learning, Dec 2014, Montréal, Canada. hal-01103076

HAL Id: hal-01103076

https://hal.science/hal-01103076v1

Submitted on 14 Jan 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Optimal transport with Laplacian regularization

Rémi Flamary
Laboratoire Lagrange UMR CNRS 7293
Observatoire de la Côte d'Azur
Université de Nice, France
remi.flamary@unice.fr

Nicolas Courty
IRISA UMR CNRS 6074
Université de Bretagne Sud
56000 Vannes, France
ncourty@irisa.fr

Alain Rakotomamonjy
LITIS EA 4108
Université de Rouen
76000, Saint Etienne du Rouvray, France
alain.rakoto@insa-rouen.fr

Devis Tuia
Department of Geography
University of Zurich
8057 Zurich, Switzerland
devis.tuia@geo.uzh.ch

Abstract

We propose a method based on optimal transport for empirical distributions with Laplacian regularization (LOT). Laplacian regularization is a graph-based regularization that can encode neighborhood similarity between samples either on the final position of the transported samples or on their displacement. In both cases, LOT is expressed as a quadratic programming problem and can be solved with a Frank-Wolfe algorithm with optimal step size. Result on domain adaptation and a shape matching problems show the interest of using this regularization in optimal transport.

1 Introduction

Recently, there has been a renewed interest for the study and the application of optimal transport (OT) [1]. OT aims at matching two probability distributions while minimizing a displacement cost. To attain this objective, it uses a similarity measure on probability distributions denoted as the optimal transport distance or the Wasserstein distance. When considering discrete distributions, OT is the solution of a convex linear programming problem, but several recent works in the machine learning literature proposed to use additional regularization terms [2, 3, 4].

In the light of these recent efforts, optimal transport has also impacted several domains, in which it has been applied successfully. For instance, in computer vision, the Earth Mover Distance has been shown to work very well to compare color and texture distributions [5]. In machine learning, two recent applications have shown the potential of optimal transport for computing barycenters of distributions [6] and for domain adaptation [7]. In the latter, authors introduced a solution for OT, which is regularized by the class memberships in the source domain.

In this work, we investigate a new regularization scheme for optimal transport of discrete distributions. Similarly to the work of Ferradans et al. [2], in which authors proposed to smooth the optimal transport plan by controlling the displacement of pairs of points, our idea is to achieve a structure-preserving optimal transport. For this purpose, we propose two regularization schemes based on Laplacian regularization. These regularizers favor OT solutions, that preserve the spatial relationships between the elements after interpolation. The first, named LOT_{pos} regularizes samples by their final position after transport, while the second, named LOT_{disp}, acts on the transportation plan itself.

The rest of the paper is as follows: in Section 2 we introduce the framework of optimal transport for empirical distributions. In Section 3 we present the proposed regularizers for LOT, whose effects are illustrated in Fig. 2. Section 4 provides details about the Frank-Wolfe optimization proposed to solve the LOT problem. Finally, Section 5 presents results on a synthetic domain adaptation problem and an application of OT in 3D human meshes registration.

2 Optimal transport for discrete distributions

In this section, we shortly introduce OT and how it can be used for interpolating discrete collections of samples expressed as probability distributions. OT aims at computing a minimal cost transportation between a source probability distribution μ_s, and a target distribution μ_t generally corresponding to continuous distributions. In this work we consider the case when those distributions are only available through a finite number of samples, and can be written as

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

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 $\mathbf{X}_s = [\mathbf{x}_1^s, \ldots, \mathbf{x}_{ns}^s]^\top \in \mathbb{R}^{n_s \times d}$ and $\mathbf{X}_t = [\mathbf{x}_1^t, \ldots, \mathbf{x}_{ns}^t]^\top \in \mathbb{R}^{n_s \times d}$, respectively. The set of probabilistic couplings between these two distributions is then the set of doubly stochastic matrices $\mathcal{P}$ defined as

$$\mathcal{P} \quad = \quad \left\{ \boldsymbol{\gamma} \in (\mathbb{R}^+)^{n_s \times n_t} \,|\, \boldsymbol{\gamma} \mathbf{1}_{n_t} = \mu_s, \boldsymbol{\gamma}^T \mathbf{1}_{n_s} = \mu_t \right\} \tag{2}$$

where $\mathbf{1}_d$ is a d-dimensional vector of ones. The Kantorovitch formulation of the optimal transport [9] is:

$$\boldsymbol{\gamma}_0 = \underset{\boldsymbol{\gamma} \in \mathcal{P}}{\operatorname{argmin}} \left\langle \boldsymbol{\gamma}, \mathbf{C} \right\rangle_F \tag{3}$$

where $\langle ., . \rangle_F$ 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 [1].

When the optimal transportation plan $\boldsymbol{\gamma}$ has been estimated, one can transport [2] the source distribution by expressing the new transported distribution $\hat{\mathbf{X}}_s$ as linear combination of elements from $\mathbf{X}_t$. The corresponding convex combination relates to the corresponding transportation plan (i.e. the projection over the unit simplex of the corresponding line of $\boldsymbol{\gamma}$) through the following relation:

$$\hat{\mathbf{X}}_s = \mathbf{T}_{\boldsymbol{\gamma}_0}(\mathbf{X}_s) = \text{diag}(\boldsymbol{\gamma}_0 \mathbf{1}_{n_t})^{-1} \boldsymbol{\gamma}_0 \mathbf{X}_t \quad \text{and} \quad \hat{\mathbf{X}}_t = \mathbf{T}_{\boldsymbol{\gamma}_0}^{-1}(\mathbf{X}_t) = \text{diag}(\boldsymbol{\gamma}_0^\top \mathbf{1}_{n_s})^{-1} \boldsymbol{\gamma}_0^\top \mathbf{X}_t. \tag{4}$$

Note that in this work we focus on uniform distributions with $p_i^s = \frac{1}{n_s}$ and $p_i^t = \frac{1}{n_t}$. In this case, a simpler transportation formulation can be derived:

$$\hat{\mathbf{X}}_s = \mathbf{T}_{\boldsymbol{\gamma}_0}(\mathbf{X}_s) = n_s \boldsymbol{\gamma}_0 \mathbf{X}_t \quad \text{and} \quad \hat{\mathbf{X}}_t = \mathbf{T}_{\boldsymbol{\gamma}_0}^{-1}(\mathbf{X}_t) = n_t \boldsymbol{\gamma}_0^\top \mathbf{X}_s. \tag{5}$$

Using this last transportation formula avoids inverting a diagonal matrix that depends on $\boldsymbol{\gamma}$ and this will greatly simplify the optimization presented in Section 4. In the more general case where the weights are non-uniform, we refer the reader to [2, sec. 3.3].

Using this simple ingredients, recent works have solved domain adaptation problems, either using entropy-based regularization (below referred to as OT$_{sinkhorn}$ [3]) or class-label group sparse regularization [7] and achieved state of the art results with respect to current domain adaptation literature. In what follows, we assume that data live on a manifold and propose two Laplacian regularization schemes as an alternative to entropy regularization.

3 Laplacian regularization for optimal transport

In the following, we discuss two Laplacian regularization schemes for OT. The first one is an application of graph regularization on the transported samples. This regularization term will promote samples that are neighbors in the graph to stay neighbors after transportation. The second one is a direct transposition of the graph regularization proposed in [2], where the graph regularization is applied on the displacement of the samples during transportation.

3.1 Regularizing the transported samples positions

Regularizing the transported sample positions leads to the following optimization problem:

$$\boldsymbol{\gamma}_0 = \underset{\boldsymbol{\gamma}\in\mathcal{P}}{\operatorname{argmin}} \langle \boldsymbol{\gamma}, \mathbf{C}\rangle_F + \frac{\lambda_s}{N_s^2}\sum_{i,j} S^s_{i,j}\|\hat{\mathbf{x}}^s_i - \hat{\mathbf{x}}^s_j\|^2 + \frac{\lambda_t}{N_t^2}\sum_{i,j} S^t_{i,j}\|\hat{\mathbf{x}}^t_i - \hat{\mathbf{x}}^t_j\|^2 \tag{6}$$

The first term is the classical optimal transport linear loss that encodes the transportation cost between the source and the target samples. The two other terms are Laplacian regularization terms that will promote the respect of the proximities observed in the original distribution after the transport. Note that the graph regularization is defined by symmetric similarity matrices $\mathbf{S}_s$ and $\mathbf{S}_t$ computed between the source and target samples in their original configurations.

The optimization problem (6) does not explicitly illustrate the dependency of the regularization term on the variable $\boldsymbol{\gamma}$. This dependency exists through the use of the transported $\hat{\mathbf{x}}$ samples. The regularization terms can be easily expressed for the source regularization as

$$\sum_{i,j} S^s_{i,j}\|\hat{\mathbf{x}}^s_i - \hat{\mathbf{x}}^s_j\|^2 = N_s^2 \mathrm{Tr}(\mathbf{X}_t^\top \boldsymbol{\gamma}^\top \mathbf{L}_s \boldsymbol{\gamma} \mathbf{X}_t) \tag{7}$$

where $\mathbf{L}_s = \mathrm{diag}(\mathbf{S}_s\mathbf{1}) - \mathbf{S}_s$ is the Laplacian of the graph $\mathbf{S}_s$; and for the target regularization as

$$\sum_{i,j} S^t_{i,j}\|\hat{\mathbf{x}}^t_i - \hat{\mathbf{x}}^t_j\|^2 = N_t^2 \mathrm{Tr}(\mathbf{X}_s^\top \boldsymbol{\gamma} \mathbf{L}_t \boldsymbol{\gamma}^\top \mathbf{X}_s) \tag{8}$$

where $\mathbf{L}_t = \mathrm{diag}(\mathbf{S}_t\mathbf{1}) - \mathbf{S}_t$. Finally, optimization problem (6) car be reformulated as

$$\boldsymbol{\gamma}_0 = \underset{\boldsymbol{\gamma}\in\mathcal{P}}{\operatorname{argmin}} \langle \boldsymbol{\gamma}, \mathbf{C}\rangle_F + \lambda_s \mathrm{Tr}(\mathbf{X}_t^\top \boldsymbol{\gamma}^\top \mathbf{L}_s \boldsymbol{\gamma} \mathbf{X}_t) + \lambda_t \mathrm{Tr}(\mathbf{X}_s^\top \boldsymbol{\gamma} \mathbf{L}_t \boldsymbol{\gamma}^\top \mathbf{X}_s), \tag{9}$$

which is a quadratic program *w.r.t.* $\boldsymbol{\gamma}$. This raises problems of high computational complexity, particularly due to the fact that the quadratic loss matrix is full. This transportation of the samples will be referred to as LOT_{pos}.

3.2 Regularizing the transported samples displacement

As discussed previously, [2] proposed to adapt image histograms through the transportation of the pixels from one image to another. One of their contributions was to propose a graph regularization of the displacement of the samples during interpolation (4). This regularization leads to the following optimization problem

$$\boldsymbol{\gamma}_0 = \underset{\boldsymbol{\gamma}\in\mathcal{P}}{\operatorname{argmin}} \langle \boldsymbol{\gamma}, \mathbf{C}\rangle_F + \frac{\lambda_s}{N_s^2}\sum_{i,j} S^s_{i,j}\|(\hat{\mathbf{x}}^s_i - \mathbf{x}^s_i) - (\hat{\mathbf{x}}^s_j - \mathbf{x}^s_j)\|^2 + \frac{\lambda_t}{N_t^2}\sum_{i,j} \tilde{S}^t_{i,j}\|(\hat{\mathbf{x}}^t_i - \mathbf{x}^t_i) - (\hat{\mathbf{x}}^t_j - \mathbf{x}^t_j)\|^2 \tag{10}$$

As for the LOT_{pos} regularization above, one can reformulate the first regularization term as

$$\sum_{i,j} S^s_{i,j}\|(\hat{\mathbf{x}}^s_i - \mathbf{x}^s_i) - (\hat{\mathbf{x}}^s_j - \mathbf{x}^s_j)\|^2 = N_s^2 \mathrm{Tr}(\mathbf{X}_t^\top \boldsymbol{\gamma}^\top \mathbf{L}_s \boldsymbol{\gamma} \mathbf{X}_t) + \left\langle \boldsymbol{\gamma}, -N_s(\mathbf{L}_s + \mathbf{L}_s^\top)\mathbf{X}_s\mathbf{X}_t^\top)\right\rangle_F + c_s$$

with $c_s = \mathrm{Tr}(\mathbf{X}_s^\top \mathbf{L}_s \mathbf{X}_s)$ a constant *w.r.t.* $\boldsymbol{\gamma}$; and the second regularization term as

$$\sum_{i,j} S^t_{i,j}\|(\hat{\mathbf{x}}^t_i - \mathbf{x}^t_i) - (\hat{\mathbf{x}}^t_j - \mathbf{x}^t_j)\|^2 = N_t^2 \mathrm{Tr}(\mathbf{X}_s^\top \boldsymbol{\gamma} \mathbf{L}_t \boldsymbol{\gamma}^\top \mathbf{X}_s) + \left\langle \boldsymbol{\gamma}, -N_t\mathbf{X}_s\mathbf{X}_t^\top(\mathbf{L}_t + \mathbf{L}_t^\top)\right\rangle_F + c_t$$

with $c_t = \mathrm{Tr}(\mathbf{X}_t^\top \mathbf{L}_t \mathbf{X}_t)$. Finally, the optimization problem (10) can be expressed as

$$\boldsymbol{\gamma}_0 = \underset{\boldsymbol{\gamma}\in\mathcal{P}}{\operatorname{argmin}} \langle \boldsymbol{\gamma}, \mathbf{C} + \lambda_s\mathbf{C}_s + \lambda_t\mathbf{C}_t\rangle_F + \lambda_s \mathrm{Tr}(\mathbf{X}_t^\top \boldsymbol{\gamma}^\top \mathbf{L}_s \boldsymbol{\gamma} \mathbf{X}_t) + \lambda_t \mathrm{Tr}(\mathbf{X}_s^\top \boldsymbol{\gamma} \mathbf{L}_t \boldsymbol{\gamma}^\top \mathbf{X}_s) \tag{11}$$

with $\mathbf{C}_s = -\frac{1}{N_s}(\mathbf{L}_s + \mathbf{L}_s^\top)\mathbf{X}_s\mathbf{X}_t^\top$ and $\mathbf{C}_t = -\frac{1}{N_t}\mathbf{X}_s\mathbf{X}_t^\top(\mathbf{L}_t + \mathbf{L}_t^\top)$ linear costs associated to the source and target samples respectively. Optimization problem (11) is extremely similar to (9), their only difference being a linear term taking into account the regularization terms $\mathbf{C}_s$ and $\mathbf{C}_t$. This transportation of the samples will be referred to as LOT_{disp}.

4 Optimization

As discussed in [2], the Frank-Wolfe algorithm can be used to solve the optimization problems (9) and (11). This approach is particularly interesting for these problems, since it works by solving at each iteration a linearization of the loss function under the linear constraints (a LP problem).

The Frank-Wolfe (FW) algorithm is an interior point method that has been proposed in [10] and has drawn some attention recently thanks to its ability to handle large scale optimization problems [11]. It addresses the minimization of $f(\gamma)$ subject to $\boldsymbol{\gamma} \in \mathcal{P}$, where $f(\cdot)$ is a convex and differentiable function and $\mathcal{P}$ are convex constraints. The algorithm with optimal step size is defined as follows:

0. Initialize $k = 0$ and $\boldsymbol{\gamma}^0 \in \mathcal{P}$ an initial value. (either $\boldsymbol{\gamma}^0 = \mu_s \mu_t^\top$ or solution of the linear problem with no quadratic regularization)
1. Compute the solution of the following problem
$$\boldsymbol{\gamma}^* = \underset{\boldsymbol{\gamma} \in \mathcal{P}}{\operatorname{argmin}} \quad \left\langle \boldsymbol{\gamma}, \nabla_{\boldsymbol{\gamma}} f(\boldsymbol{\gamma}^k) \right\rangle_F$$
2. Find the optimal step $0 \leq \alpha^k \leq 1$ with descent direction $\Delta\boldsymbol{\gamma} = \boldsymbol{\gamma}^* - \boldsymbol{\gamma}^k$ such that
$$\alpha^k = \underset{0 \leq \alpha \leq 1}{\operatorname{argmin}} \, f(\boldsymbol{\gamma}^k + \alpha \Delta\boldsymbol{\gamma})$$
3. $\boldsymbol{\gamma}^{k+1} \leftarrow \boldsymbol{\gamma}^k + \alpha^k \Delta\boldsymbol{\gamma}$, set $k \leftarrow k+1$ and go to step 1.

We want to empathize the importance of an interior point method such as FW. In practice, the exact solution of the optimization problem is often not necessary and an early stopping strategy is used. In our case, an interior point method ensures that even if the iterations are stopped before convergence, $\boldsymbol{\gamma}^k$ belongs to $\mathcal{P}$ and leads to a proper transportation in the convex hull of the target data. Note that the gradient of objective values term in (9) and (11) can be easily derived from derivative of trace [12, Eq. (116)].

Finally, we discuss how the optimal step is found when optimizing problem (9). One want to minimize the following function *w.r.t.* α:

$$\begin{aligned}
g(\alpha) &= \left\langle \boldsymbol{\gamma}^k + \alpha\boldsymbol{\gamma}^*, \mathbf{C} \right\rangle_F + \lambda_s \mathrm{Tr}(\mathbf{X}_t^\top (\boldsymbol{\gamma}^k + \alpha\boldsymbol{\gamma}^*)^\top \mathbf{L} (\boldsymbol{\gamma}^k + \alpha\boldsymbol{\gamma}^*) \mathbf{X}_t) + \lambda_t \mathrm{Tr}(\mathbf{X}_s^\top (\boldsymbol{\gamma}^k + \alpha\boldsymbol{\gamma}^*) \tilde{\mathbf{L}} (\boldsymbol{\gamma}^k + \alpha\boldsymbol{\gamma}^*)^\top \mathbf{X}_s) \\
&= \text{cste} + \alpha \left\langle \boldsymbol{\gamma}^*, \mathbf{C} \right\rangle_F + \lambda_s (\alpha \mathrm{Tr}(\mathbf{X}_t^\top \boldsymbol{\gamma}^{*\top} \mathbf{L} \boldsymbol{\gamma}^k \mathbf{X}_t) + \alpha \mathrm{Tr}(\mathbf{X}_t^\top \boldsymbol{\gamma}^{k\top} \mathbf{L} \boldsymbol{\gamma}^* \mathbf{X}_t) + \alpha^2 \mathrm{Tr}(\mathbf{X}_t^\top \boldsymbol{\gamma}^{*\top} \mathbf{L} \boldsymbol{\gamma}^* \mathbf{X}_t)) \\
&+ \lambda_t (\alpha \mathrm{Tr}(\mathbf{X}_s^\top \boldsymbol{\gamma}^{*\top} \tilde{\mathbf{L}} \boldsymbol{\gamma}^k \mathbf{X}_s) + \alpha \mathrm{Tr}(\mathbf{X}_s^\top \boldsymbol{\gamma}^{k\top} \tilde{\mathbf{L}} \boldsymbol{\gamma}^* \mathbf{X}_s) + \alpha^2 \mathrm{Tr}(\mathbf{X}_s^\top \boldsymbol{\gamma}^{*\top} \tilde{\mathbf{L}} \boldsymbol{\gamma}^* \mathbf{X}_s)) \\
&= cste + \alpha A + \alpha^2 B
\end{aligned}$$

$$g(\alpha) = f(\boldsymbol{\gamma}^k + \alpha\Delta\boldsymbol{\gamma}) = a\alpha^2 + b\alpha + c$$

where a, b, and c are constants at each iteration k. Since $g(\cdot)$ is a second order polynomial, the optimal step can be computed with $g'(\alpha) = 2a\alpha + b = 0$. When applied to the optimization problem (9), this gives

$$\alpha^k = -\frac{1}{2} \frac{\left\langle \Delta\boldsymbol{\gamma}, \mathbf{C} \right\rangle_F + \lambda_s \mathrm{Tr}(\mathbf{X}_t^\top \Delta\boldsymbol{\gamma}^\top (\mathbf{L}_s + \mathbf{L}_s^\top) \boldsymbol{\gamma}^k \mathbf{X}_t) + \lambda_t \mathrm{Tr}(\mathbf{X}_s^\top \Delta\boldsymbol{\gamma} (\mathbf{L}_t + \mathbf{L}_t^\top) \boldsymbol{\gamma}^{k\top} \mathbf{X}_s)}{\lambda_s \mathrm{Tr}(\mathbf{X}_t^\top \Delta\boldsymbol{\gamma}^\top \mathbf{L}_s \Delta\boldsymbol{\gamma} \mathbf{X}_t) + \lambda_t \mathrm{Tr}(\mathbf{X}_s^\top \Delta\boldsymbol{\gamma} \mathbf{L}_t \Delta\boldsymbol{\gamma}^\top \mathbf{X}_s)} \tag{12}$$

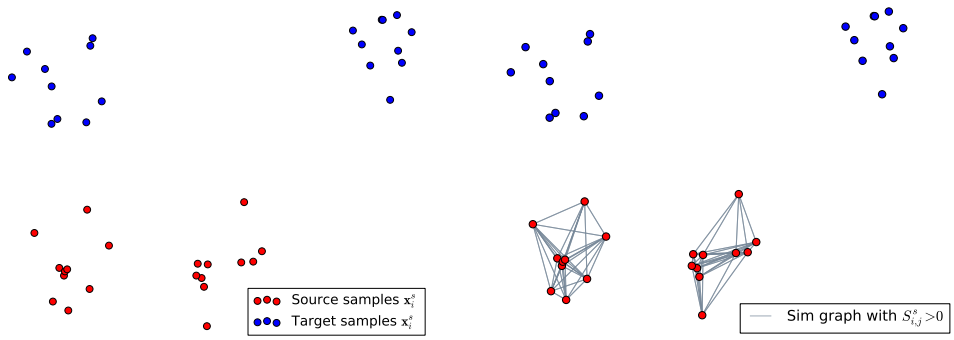
The optimal step for problem (11) is given in [2] and can be computed by replacing $\mathbf{C}$ by $\mathbf{C} + \lambda_s \mathbf{C}_s + \lambda_t \mathbf{C}_t$ in equation (12).

5 Numerical experiments

In this section, we first illustrate the effect of the two regularizers proposed. Then, we show the results obtained on a synthetic domain adaptation problem and on a shape matching dataset.

5.1 Illustration of the regularized optimal transport

In Fig. 1, we illustrate on a simple dataset the behavior of the different regularization terms in term of sample transportation. In our example dataset, both the source and target distributions consist in two



(a) Initial distributions (b) Similarity graph $\mathbf{S}^s$

Figure 1: Simulated distribution with two clusters in each domain.

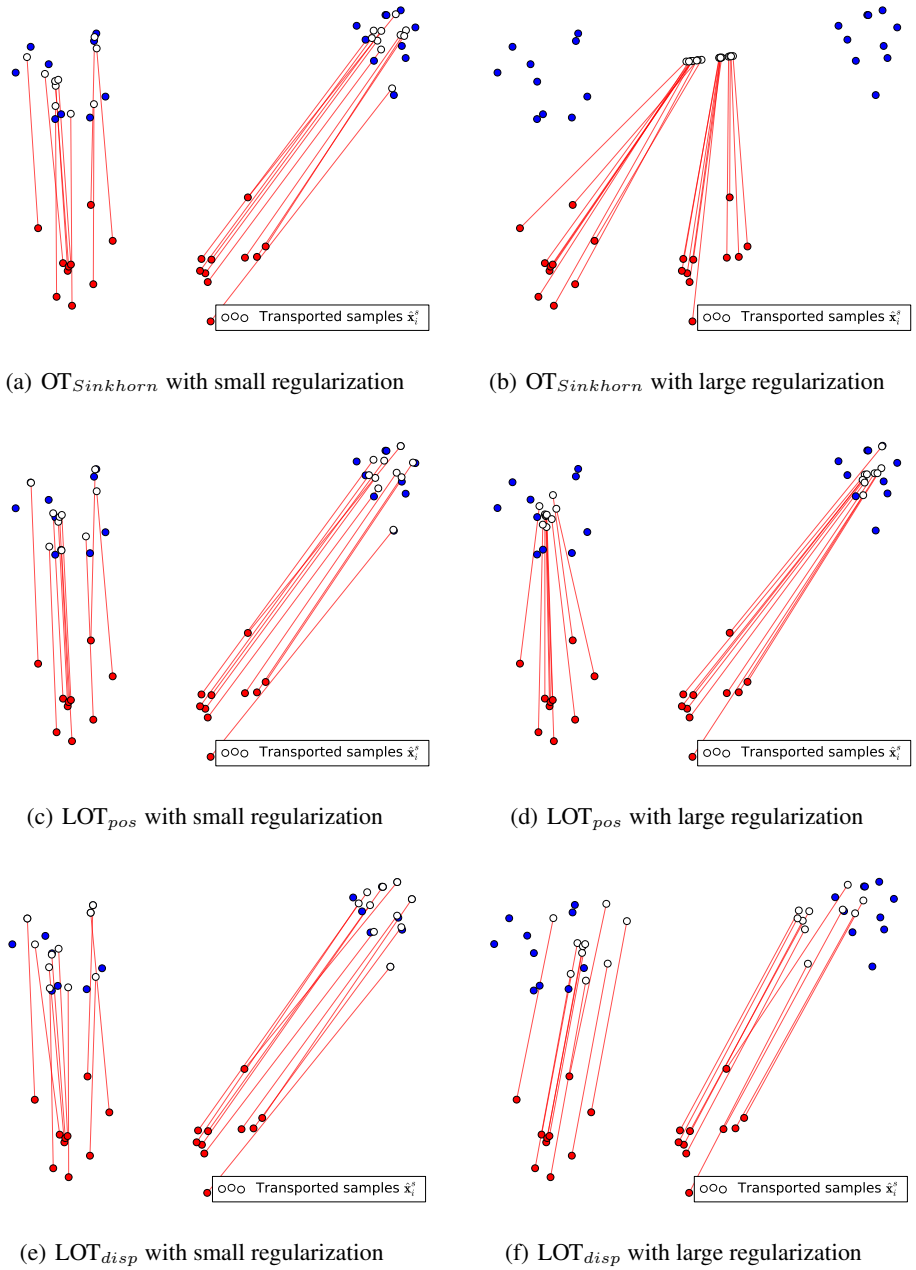
clusters (see Fig. 1.a). The similarity graph $\mathbf{S}_s$ is then chosen to be 1 for samples belonging to the same cluster and 0 otherwise (Fig. 1.b). To ease interpretation, we use the Laplacian regularization only in the source domain ($\lambda_t = 0$).

The results of adaptation are illustrated in Figure 2 for OT$_{Sinkhorn}$, LOT$_{pos}$ and LOT$_{disp}$ with a small and large regularization term λ_s. In the figure we can appreciate the shrinkage effect to the center of mass of the data of the Sinkhorn regularization (top row). Similarly, LOT$_{pos}$ leads to a shrinkage but takes into account the graph structure with a per-cluster shrinkage (middle row). This is interesting, as it can be seen as a de-noising of the source samples during the transport. The displacement regularization LOT$_{disp}$ leads to a smaller shrinkage, since it aims at preserving the shape of each cluster while transporting onto the target data (the obtained displacements are all parallel and of same length per cluster). Even if one could argue that the best regularization depends on the dataset, one can suppose that LOT$_{pos}$ might be a better fit when a non-rigid transformation of the data is required.

5.2 Domain adaptation on the two-moons dataset

In this experiment, we consider a rotating two-moons toy example that was used for domain adaptation in [13]. The source domain consists in the standard two entangled moons data, where each moon is associated to a specific class. The target domain is built by applying a rotation to the two moons, which allows to consider an adaptation problem with an increasing difficulty as a function of the rotation angle. Examples of the datasets at $50°$, $70°$ and $90°$ are reported in the top row of Fig. 3. This example is notably interesting because the corresponding transformation is clearly non-linear, and because the input dimensionality is small. We follow the same experimental setup as in [13]: this allows for a direct comparison with the state-of-the-art results presented in [13]. Both domains are built with two moons made of 150 samples each. After adaptation between the two distributions, the generalization power is tested over a set of 1000 samples from the target domain. The experiments are conducted 10 times and we consider the mean classification error as a comparison criterion. The classification is conducted by an SVM classifier with a Gaussian kernel, whose parameters were set by 5-fold cross-validation. We compare the adaptation results with two state-of-the-art methods: the DA-SVM approach [14] and the more recent PBDA [13], which has proved to give competitive results on this dataset. The graph regularization $\mathbf{S}_s$ and $\mathbf{S}_t$ are computed using a Gaussian kernel with $\gamma = 0.1$ and the regularization term are set to $\lambda_s = \lambda_t = \lambda$ selected empirically. Results are reported in Table 1 and the resulting classifiers for classical OT LP, OT$_{Sinkhorn}$ and OT$_{pos}$ are plotted on Figure 3 for the $50°$, $70°$ and $90°$ cases.

Our first observation is that all the methods based on optimal transport behave better than the competing methods. This is particularly striking when the amount of rotation is important ($> 50°$). This clearly shows the capacity of our method to handle strong distortions between domains. The entropy regularization (OT$_{sinkhorn}$) leads to a dramatic performance gain *w.r.t.* the classical OT LP. OT$_{pos}$ is the best performing method thanks to its Laplacian regularization. In this dataset, it clearly out-



(a) $\text{OT}_{Sinkhorn}$ with small regularization

(b) $\text{OT}_{Sinkhorn}$ with large regularization

(c) LOT_{pos} with small regularization

(d) LOT_{pos} with large regularization

(e) LOT_{disp} with small regularization

(f) LOT_{disp} with large regularization

Figure 2: Illustration of the different transport regularizations on the example of Figure 1. The transported samples and their displacement (in red) are illustrated for $\text{OT}_{sinkhorn}$, LOT_{pos} and LOT_{disp} for a small (left column) and large (right column) regularization term.

performs OT_{disp}. Our interpretation is that, in this case, non-rigid transport is necessary. Note that we are aware that this dataset, despite its wide use in the domain adaptation community, is clearly biased toward the Laplacian regularization of the position of the samples.

5.3 Non-rigid shape matching

The last example considers non-rigid 3D matching of humanoid shapes. This problem has attracted over the past years a lot of interest in both the domains of computer vision and graphics, where 3D

	$10°$	$20°$	$30°$	$40°$	$50°$	$70°$	$90°$
SVM (no adpat)	0.000	0.104	0.24	0.312	0.4	0.764	0.828
DASVM [14]	0.000	0.000	0.259	0.284	0.334	0.747	0.82
PBDA [13]	0.000	0.094	0.103	0.225	0.412	0.626	0.687
OT LP	0.000	0.000	0.031	0.102	0.166	0.292	0.441
$\text{OT}_{sinkhorn}$	0.000	0.000	0.000	0.000	0.013	0.202	0.386
LOT_{pos}	0.000	0.000	0.000	0.000	0.000	0.022	0.152
LOT_{disp}	0.000	0.000	0.000	0.000	0.000	0.067	0.384

Table 1: Mean error rate over 10 realizations of the two-moons classification problem. The target domain is given by a rotation of the source domain of angle given in the first row.

shapes need to be registered, either for reconstruction or tracking purposes (see [15] for a survey). Among those approaches, the optimal transport distances were envisaged [16] to measure the distortion between objects. Yet, most of the approaches based on such distances encountered difficulties in the practical computation of the matching.

We use the FAUST dataset [17], which provides ground truth correspondences between shapes, since every scan comes with a registered template watertight mesh. This allows to compute numerical errors over the possible registrations. Our primary goal here is not to compare LOT with state-of-the-art methods in shape registration, but to study the benefits of our the Laplacian regularization of the transport for this registration purpose. Laplacian methods seem to be particularly adapted to solve this problem, since its objective is to align surfaces, while preserving their graph structures.

Among the different shapes offered in the dataset, we selected 6 shapes that are related to different poses of the same subject (see Figure 4). The corresponding registration problem is very complex, particularly because the magnitude of deformations differs from one pair of shapes to the other. We directly use the corresponding template meshes to assess for the quality of the matching. Figure 5 illustrates the mapping over two pairs of shapes, respectively (#1,#2) and (#3,#6) by the LOT_{pos} regularization. We note here that the original transport leads to a one-to-one assignment between vertices, whereas the Laplacian regularization yields a one-to-many mapping. Consequently, we evaluate the correctness of the mapping by looking at the distance between vertices of the transported mesh (equation (5)) and their expected location on the target mesh. Table 2 reports both the average distance error (in centimeters) over all the vertices of the mesh and the corresponding maximum error (as a measure of the most irrelevant mapping). From those results, one can see that both LOT_{pos} and LOT_{disp} regularization strategies lead to smaller average and maximal matching errors compared to OT LP, with a slight advantage for the regularization over the positions of the vertices.

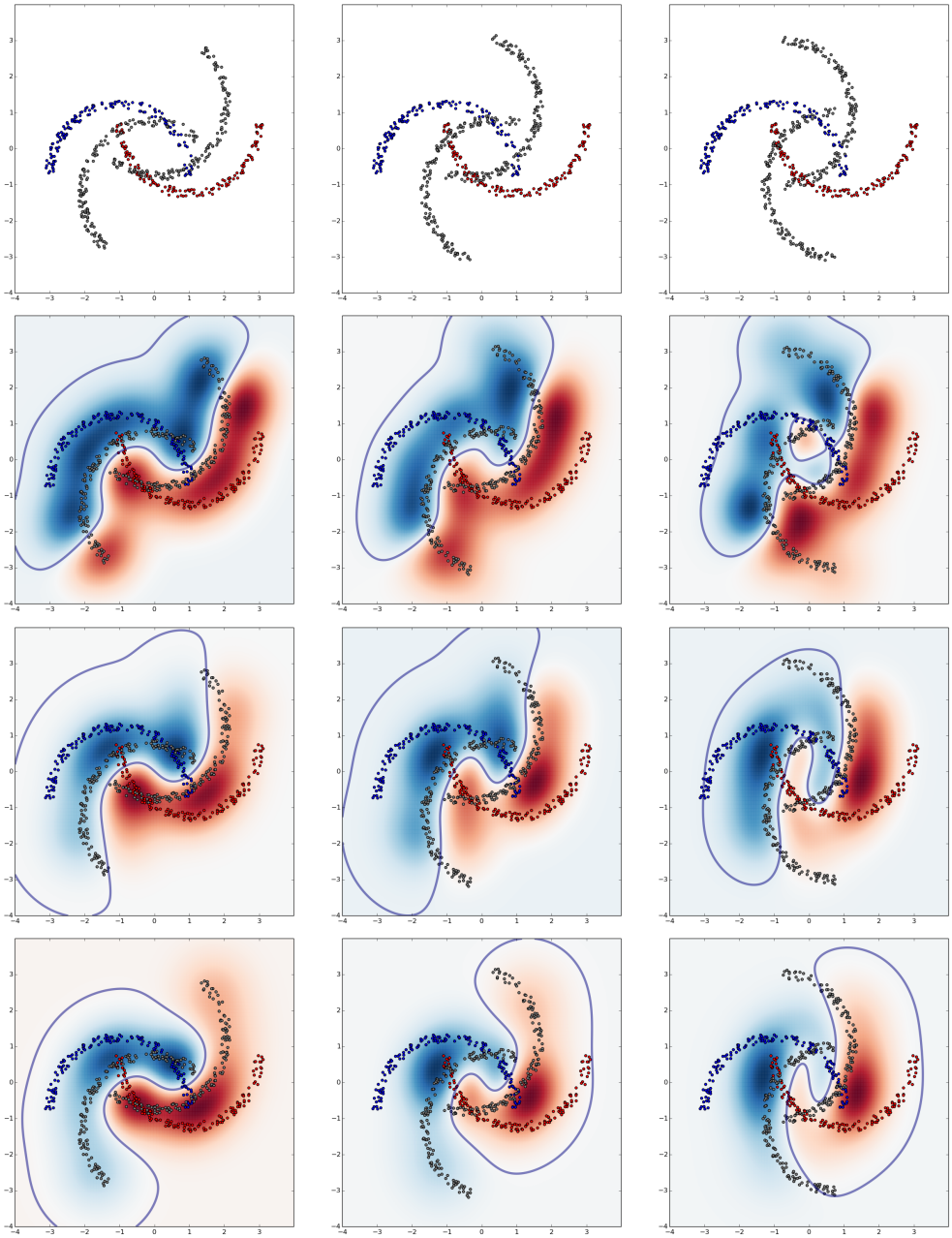
6 Conclusion

In this work we proposed to use two Laplacian regularization schemes in the optimal transport definition, acting either on the position or on the displacement of the transported samples. The resolution of the resulting optimization problem is discussed and the different regularization terms are tested on an illustrative example, the complex two-moons domain adaptation dataset and a 3D shape matching problem. In all examples, the proposed LOT showed promising performances and achieved at least state-of-art results.

We plan in future works to perform more in-depth numerical experiments with new real life datasets and to address the problem of large scale optimization. The size of the data in terms of number of samples is indeed limited by the use of a LP solver at each iteration of the Frank-Wolfe algorithm and new optimization strategies will be necessary for exploiting this type of methodologies with large scale datasets.

References

[1] C. Villani. *Optimal transport: old and new*. Grundlehren der mathematischen Wissenschaften. Springer, 2009.



(a) rotation=50° (a) rotation=70° (a) rotation=90°

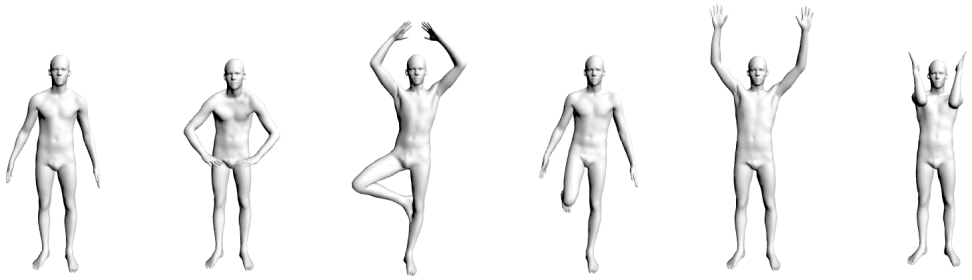
Figure 3: Illustration of the classification decision boundary over the two moons example. The target domain is depicted as points in grey (best viewed with colors).

[2] S. Ferradans, N. Papadakis, J. Rabin, G. Peyré, and J.-F. Aujol. Regularized discrete optimal transport. In *Scale Space and Variational Methods in Computer Vision, SSVM*, pages 428–439, 2013.

[3] M. Cuturi. Sinkhorn distances: Lightspeed computation of optimal transportation. In *NIPS*, pages 2292–2300. 2013.

[4] N. Papadakis, G. Peyré, and E. Oudet. Optimal Transport with Proximal Splitting. *SIAM Journal on Imaging Sciences*, 7(1):212–238, 2014.

[5] Y. Rubner, C. Tomasi, and L.J. Guibas. A metric for distributions with applications to image



(a) #1 (b) #2 (c) #3 (d) #4 (e) #5 (f) #6

Figure 4: The 6 different human shapes from the FAUST dataset considered in our non-rigid registration experiment.

	method	1	2	3	4	5	6
1	OT LP)	–	8.0 (54.5)	46.3 (141.2)	3.4 (57.6)	50.1 (160.9)	34.0 (125.3)
	LOT_{pos}	–	7.5 (37.2)	44.0 (133.2)	3.7 (41.1)	46.5 (139.1)	31.1 (102.5)
	LOT_{disp}	–	7.1 (38.9)	44.3 (132.0)	3.5 (42.6)	47.9 (142.5)	31.9 (108.0)
2	OT LP	7.7 (57.0)	–	40.2 – 133.2)	12.0 (58.7)	44.1 (138.7)	26.8 (100.8)
	LOT_{pos}	7.0 (39.2)	–	38.0 (90.3)	11.1 (48.7)	40.7 (100.9)	24.3 (60.4)
	LOT_{disp}	6.6 (41.5)	–	38.5 (89.6)	10.5 (47.5)	42.4 (100.0)	25.5 (61.5)
3	OT LP	51.3 (113.7)	41.3 (85.2)	–	48.5 (113.2)	6.1 (50.3)	11.4 (49.4)
	LOT_{pos}	49.2 (108.7)	39.1 (78.7)	–	46.4 (107.7)	6.0 (47.5)	11.0 (46.9)
	LOT_{disp}	51.4 (109.7)	40.8 (80.0)	–	48.7 (108.8)	5.7 (48.3)	10.8 (46.0)
4	OT LP	3.2 (48.7)	12.8 (71.6)	43.8 (143.7)	–	54.0 (178.5)	37.4 (144.1)
	LOT_{pos}	3.4 (47.8)	11.8 (54.5)	41.3 (132.4)	–	50.3 (109.5)	34.8 (109.5)
	LOT_{disp}	3.3 (45.0)	10.9 (66.0)	41.9 (131.0)	–	51.4 (149.8)	35.3 (114.5)
5	OT LP	46.8 (113.5)	36.6 (79.3)	5.7 (53.1)	49.8 (112.9)	–	6.6 (41.2)
	LOT_{pos}	45.0 (109.0)	34.9 (79.3)	5.6 (49.8)	48.1 (108.6)	–	6.0 (36.1)
	LOT_{disp}	47.2 (110.9)	36.6 (79.8)	5.2 (44.6)	50.2 (110.7)	–	5.9 (37.4)
6	OT LP	43.9 (112.6)	32.1 (84.2)	12.3 (54.2)	46.3 -112.1)	7.5 (50.7)	–
	LOT_{pos}	41.3 (107.5)	30.0 (77.7)	12.1 (49.0)	43.8 (107.0)	7.5 (44.3)	–
	LOT_{traj}	44.0 (108.7)	32.1 (79.6)	11.6 (45.4)	46.4 (108.0)	7.1 (45.6)	–

Table 2: Shape registration experiments. For each pairs of possible matching, we report in centimeters the mean average error and, between parenthesis, the maximum error for the three considered methods.

databases. In *ICCV*, pages 59–66, Jan 1998.

[6] M. Cuturi and A. Doucet. Fast computation of wasserstein barycenters. In *ICML*, to appear 2014.

[7] N. Courty, R. Flamary, and D. Tuia. Domain adaptation with regularized optimal transport. In *European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD)*, 2014.

[8] S. J. Pan and Q. Yang. A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.*, 22(10):1345–1359, 2010.

[9] L. Kantorovich. On the translocation of masses. *C.R. (Doklady) Acad. Sci. URSS (N.S.)*, 37:199–201, 1942.

[10] Marguerite Frank and Philip Wolfe. An algorithm for quadratic programming. *Naval research logistics quarterly*, 3(1-2):95–110, 1956.

[11] Martin Jaggi. Revisiting {Frank-Wolfe}: Projection-free sparse convex optimization. In *Proceedings of the 30th International Conference on Machine Learning (ICML-13)*, pages 427–435, 2013.

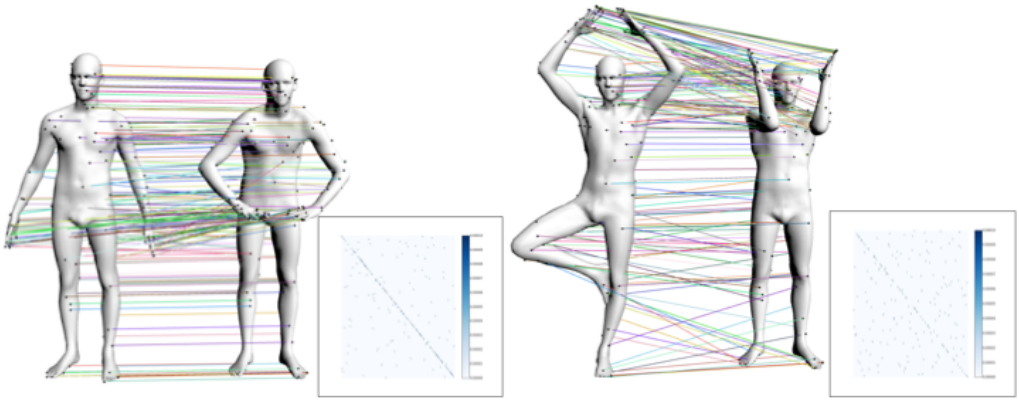


Figure 5: Illustrations of the shape matching results for shapes pairs (#1,#2) and (#3,#6) when using LOT_{pos} regularization. The optimal coupling matrix is also represented in the bottom right corner of each panel. As the points are ordered in the same ways in both meshes, a diagonal dominance of the matrix indicates the success of the matching process.

[12] K.B. Petersen and M.S. Pedersen. The matrix cookbook. *Technical University of Denmark*, 2006.

[13] P. Germain, A. Habrard, F. Laviolette, and E. Morvant. A PAC-Bayesian Approach for Domain Adaptation with Specialization to Linear Classifiers. In *ICML*, pages 738–746, Atlanta, USA, 2013.

[14] L. Bruzzone and M. Marconcini. Domain adaptation problems: A dasvm classification technique and a circular validation strategy. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(5):770–787, May 2010.

[15] O. van Kaick, H. Zhang, G. Hamarneh, and D. Cohen-Or. A survey on shape correspondence. *Computer Graphics Forum*, 30(6):1681–1707, 2011.

[16] F. Mmoli. Gromovwasserstein distances and the metric approach to object matching. *Foundations of Computational Mathematics*, pages 1–71, 2011.

[17] Federica Bogo, Javier Romero, Matthew Loper, and Michael J. Black. FAUST: Dataset and evaluation for 3D mesh registration. In *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Piscataway, NJ, USA, June 2014. IEEE.