Population Averaging of Neuroimaging Data Using L^p Distance-based Optimal Transport

Qi Wang*‡, Ievgen Redko†, Sylvain Takerkart*
*Institut de Neurosciences de la Timone UMR 7289, Aix Marseille Université, CNRS, Marseille, France
†INSA-Lyon CREATIS UMR 5220 F-69100, Lyon, France
‡Laboratoire d'Informatique et Systèmes UMR 7020, Aix Marseille Université, CNRS, Marseille, France
Email: qiqi.wang@lis-lab.fr ievgen.redko@creatis.insa-lyon.fr sylvain.takerkart@univ-amu.fr

***Abstract*—Analyzing neuroimaging data at the population level relies on averaging images that have been acquired on a group of individuals drawn from this population. Traditional group analyses are based on the general linear model, which performs euclidean averaging across individuals, independently at each brain location. It is therefore largely impacted by inter-individual differences. In this paper we propose to overcome this variability by using optimal transport to leverage the geometrical properties of multivariate brain patterns. We extend the concept of Wasserstein barycenter, which was initially meant to average probability measures, to make it applicable to arbitrary data that do not necessarily fulfill the properties of a true probability measure. For this, we introduce a new algorithm that estimates a barycenter using the transportation L^p distance [8]. We provide an experimental study on how the noise level impacts the quality of the obtained barycenter on artificial data. Our proposed method is compared with the approach introduced in [6] on artificial and real functional MRI.**

***Keywords*—functional MRI; optimal transport; individual differences; group analysis**

I. Introduction

In conventional group analysis of functional magnetic resonance imaging (fMRI) datasets, a spatial smoothing operation is applied to the data in order to overcome the potential differences that exist in the locations of activation foci across individuals. This allows improving the spatial overlap across subjects. But it results in an amplitude dump which harms the detection power of standard group analysis methods based on euclidean averaging computed independently at each brain location, such as performed through the General Linear Model (GLM).

Recently, a new method was introduced to compute the barycenter of a set of empirical probability measures with respect to a distance from the optimal transportation theory, called the Wasserstein distance [1]. The Wasserstein barycenter given by a probability measure is defined to minimize the sum of its Wasserstein distances to each element in that set. An important advantage offered by the Wasserstein distance is its capacity of taking into account the geometry of the data underlying the probability distributions by the means of the cost-matrix associated to it. In practice, the solution of the entropic regularized Wasserstein barycenter problem [5] is computationally attractive and can be obtained with Iterative Bregman projections algorithm [2].

978-1-5386-6859-7/18/$31.00 ©2018 IEEE

In the context of neuroimaging-based population studies, the Wasserstein barycenter therefore sounds appealing because its geometrical grounding should facilitate handling individual differences on the one hand, and because it intrinsically takes into account the multivariate nature of brain patterns on the other hand. However, the challenge lies in the arbitrary nature of neuroimaging data which cannot be naturally represented as empirical probability measures, i.e histograms. Indeed, the natural transformation that maps arbitrary data into a histogram is the affine function that ensures non-negativity and normalizes the mass of the data to having the value of one. Applying such transformation separately on each subject would discard the relative amplitude differences that exist both across the images and across individuals, which are critical to correctly assess the informative content of brain patterns. New optimal transport algorithms are therefore needed to handle unbalanced data such as offered in neuroimaging.

In order to compute barycenters from unbalanced measures which are non-negative with mass smaller than one, [6] proposed to add a virtual point to the considered probability measure and then to minimize the Wasserstein distance by exploiting the dual optima as the gradient of barycenter [4]. On the other hand, the method in [8] does not transform data to fulfill the properties of a probability measure but proposes to define a transportation L^p distance (TL^p), a modification of the Wasserstein distance that integrates the amplitudes of the data in the cost matrix.

In this paper, we introduce a new algorithm that computes a barycenter from a set of arbitrary images based on the TL^p distance, which we describe in Section II. Then, we provide an empirical study of its effectiveness, which we present in Section III. Our evaluation includes a direct comparison with the algorithm of [6], using artificial data, which allows studying the robustness of the methods when confronted with different levels of noise, as well as an application on a real fMRI dataset.

II. Methods

In this work, we focus on computing a barycenter from unbalanced data. We assume that there are N subjects, and

denote by $h^i \in S_d$ the unbalanced measure for i^{th} subject, with $i \in \{1, ..., N\}$, $S_d = \{u \in \mathbb{R}_+^d, \sum_{j=1}^d u_j \leq 1\}$. For any vector $u \in \mathbb{R}_+^d$, $|u|_1$ is the mass of u, $|u|_1 = \sum_{j=1}^d |u_j|$.

Our objective is to find a barycenter g which minimizes the sum of its distance to h^i. Given the unbalanced measures, we propose to construct cost matrix with locations and intensities of bins in h^i, which is related to the cost matrix in TL^p distance. Then, the barycenter can be efficiently computed using the entropic regularized optimal transport and Iterative Bregman projections.

A. Wasserstein barycenter

For two histograms a and b, such that $a, b \in \Sigma_d$, $\Sigma_d = \{u \in \mathbb{R}_+^d | \sum_{j=1}^d u_j = 1\}$, the Wasserstein distance between them is defined as:

$$W(a, b) = \min_{T \in U(a,b)} \langle T, C \rangle \tag{1}$$

where $C \in \mathbb{R}_+^{d \times d}$ is the cost matrix of pairwise distances between the locations of bins in a and b, $U(a, b)$ is a set of matrices whose row and column marginals are equal to a and b, respectively:

$$U(a, b) = \{T \in \mathbb{R}_+^{d \times d} | T\mathbb{1}_d = a, T^T\mathbb{1}_d = b\} \tag{2}$$

In [5], the author introduced an efficient algorithm for entropic regularized optimal transport defined as:

$$T = \operatorname*{argmin}_{T \in U(a,b)} \langle T, C \rangle - \frac{1}{\lambda} h(T) \tag{3}$$

where $h(T) = \sum T_{ij} \log T_{ij}$ is the entropy. Wasserstein barycenter of N histograms $\{p^1, ..., p^N\}$ where $p^i \in \Sigma_d$ is defined as a measure f which minimizes the sum of its Wasserstein distance to p^i:

$$f = \operatorname*{argmin}_{u \in \Sigma_d} \frac{1}{N} \sum_{i=1}^N W(u, p^i) \tag{4}$$

B. TL^p distance

When measures considered above represent images, the cost matrix C in the definition of the Wasserstein distance allows to take into account geometric information but does not include the information regarding the intensities of pixels. This drawback was addressed by the TL^p distance [8] defined as follows:

$$d_{TL_\lambda^p}(\mu, \nu; a, b) = \min_{\hat{T} \in U(\mu,\nu)} \langle \hat{T}, \hat{C} \rangle \tag{5}$$

$$\hat{C} = \frac{1}{\lambda} C + \tilde{C} \tag{6}$$

where μ, ν are probability measures, C is the cost matrix as in the Wasserstein distance and $\tilde{C}$ is the cost matrix of pairwise distances between intensities of a and b, i.e. $\tilde{C}_{m,n} = |a_m - b_n|_p^p$. Contrary to the Wasserstein distance, the TL^p distance makes no assumptions regarding the non-negativity of measures a and b and allows them to have different mass.

C. Barycenter with the TL^p distance

As activation intensities are important for group analysis, we assume that taking into account the intensity in the cost matrix may help to preserve the amplitudes of unbalanced data. Under this assumption, we propose to find a barycenter g that represents the mean of unbalanced measures by minimizing the sum of its TL^p distances to $\{h^1, ..., h^N\}$, where p is set to be 2.

To proceed, let us denote by $\hat{C}^i$ the cost matrix between g and h^i. This cost can be calculated as the combination of Euclidean distances between the locations and intensities of pixels, respectively, i.e. $\hat{C}^i_{m,n} = |\text{loc}(g_m) - \text{loc}(h^i_n)|_2^2 + \eta |g_m - h^i_n|_2^2$, where η is the trade-off parameter between two distances and loc() is the coordinate of a pixel on the discrete grid. The considered optimization problem thus reads:

$$g = \operatorname*{argmin}_{u \in S_d} \frac{1}{N} \sum_{i=1}^N d_{TL^p}(\hat{u}, \hat{h}^i; u, h^i) \tag{7}$$

where $\hat{h}^i = h^i / |h^i|_1$, $\hat{u} = u / |u|_1$.

As mentioned in the introduction, entropic regularization of the barycenter leads to an optimization problem that can be solved efficiently with the Iterative Bregman projections algorithm which has strong convergence guarantee. To benefit from it, we propose a two-stage strategy that first computes a barycenter g with Iterative Bregman projections for given cost matrices $\{\hat{C}^1, ..., \hat{C}^N\}$, and then updates cost matrices using the obtained barycenter. This process is continued until the barycenter stops changing. The proposed algorithm is summarized in Algorithm 1.

Algorithm 1 Barycenter with TL^p distance

Input: $\{h^1, ..., h^N\}$, $\{\hat{h}^1, ..., \hat{h}^N\}$, geometry cost matrix C, $0 < \eta < 1$, $\lambda > 0$, weights $\{\beta^1, ..., \beta^N\}$, $\sum \beta^i = 1$,
Output: barycenter g

mean mass $\rho = \frac{1}{N} \sum_i |h^j|_1$
$\hat{g} = \mathbb{1}_d / d$, $g = \mathbb{1}_d / d \times \rho$
$u = \text{ones}(d, N)$
$v = \text{ones}(d, N)$
while g changes **do**
  for each h^i **do**
    $\hat{C}^i_{m,n} = C_{m,n} + \eta |g_m - h^i_n|_2^2$
    $K^i = \exp(-\lambda \hat{C}^i)$
  end for
  while $\hat{g}$ changes **do**
    for each $\hat{h}^i$ **do**
      $u^i = \hat{h}^i / (K^i v^i)$
    end for
    $\hat{g} = \prod_{i=1}^N (v^i \times (K^i u^i))^{\beta^i}$
    for each $\hat{h}^i$ **do**
      $v^i = \hat{g} / ({K^i}^T u^i)$
    end for
  end while
  $g = \rho \times \hat{g}$
end while

III. Experiments and results

In this section, we empirically study the properties of barycenters of sets of fMRI images computed with optimal transport algorithms. Specifically, we aim at comparing our method, referred to as TL^p-BI, to p-Kantorovich Barycenter with Constrained Mass algorithm (KBCM) [6] described in the Introduction. We use both artificially constructed and real fMRI datasets where inter-individual differences are strong to test whether such algorithms can handle such variability.

A. Artificial fMRI data

First, we generate artificial fMRI datasets by simulating one contrast map per individual on a square of size 50×50. We assume that each subject presents a unique activated region of gaussian shape, with a small size ($\sigma = 1$ pixel) compared to the one of the image. We would therefore expect a summary representation of the population to present the same property. The inter-individual variability is induced by allowing both the location of the activated region and its amplitude to vary across subjects: its center is randomly chosen (uniform distribution) within a circle of radius 15 centered in the middle of the image and its amplitude is drawn from a Gaussian distribution $\mathcal{N}(5, 1)$. In order to evaluate the robustness of these algorithms in the presence of noise, we add noise onto each image, generated from a Gaussian distribution $\mathcal{N}(0, \sigma)$. The noise level is parametrically controlled by using $\sigma \in \{0.1, 0.3, 0.5, 0.7, 1.0\}$.

B. Real fMRI data

We use data available in an open repository[1] that was acquired to study voice sensitive areas in the temporal cortex [7]. During the experiment, each subject had to passively listen to a fixed set of vocal and non vocal stimuli. A within-subject general linear model analysis was set-up to estimate the contrast between the perception of vocal and non vocal sounds for each subject. These contrast maps served as inputs to compute the population average maps, within a large region of interest that included the auditory cortex as well as the voice sensitive regions (illustrated on Fig. 3). The study in [7] showed that large differences exist in the organization of the temporal voice areas, but that a structure comprising three voice patches can be identified at the group level using advanced processing techniques, not with standard group analysis. This data is therefore perfectly suited to challenge the potential gains offered by optimal transport techniques.

C. Experimental setting

We randomly define 20 sets of 20 subjects to simulate a group analysis obtained with a group of standard size, both for the artificial data – by construction – and for the real data – by drawing subsets from the full set of available subjects. We compute a barycenter image separately for each of these sets. For this, we first vectorize each contrast map to obtain a vector $v^i \in \mathbb{R}^d$. We then transform these vectors with $h^i_n = \frac{v^i_n - \alpha}{\sum_{n=1}^{d}(v^i_n - \alpha)}$, where α is the minimal intensity of the full set of data $\{v^1, ..., v^{20}\}$ – which allows preserving the relative amplitude differences across subjects – so that the elements of h^i become all non-negative and the total mass of h^i is smaller than or equal to 1 – which ensures the necessary properties for KBCM. The cost matrix used for KBCM, C^{KBCM}, contains the pairwise Euclidean distances of locations on the grid, $C^{\text{KBCM}}_{m,n} = |\text{loc}(g_m) - \text{loc}(h^i_n)|^2_2$. In TL^p-BI, the cost matrix takes into account both the locations and intensities of pixels, $\hat{C}^{\text{TL}^p-\text{BI}}_{m,n} = C^{\text{KBCM}} + \eta |g_m - h^i_n|^2_2$, $\eta \in \{1.0, 0.1, 0.01, 0.001\}$.

The obtained barycenters are examined both qualitatively and quantitatively. First, we assess the properties of the barycenters visually. Then, we use several criteria to characterize the properties of the barycenters: the maximum amplitude of the image, the number of voxels above a threshold defined as half of the maximum amplitude – which provides an estimate of the size of the most active regions, and standard deviation of the coordinates of all above-threshold voxels – which measures the spatial spread of activated regions. All comparisons are performed using paired t-test.

D. Results on artificial fMRI data

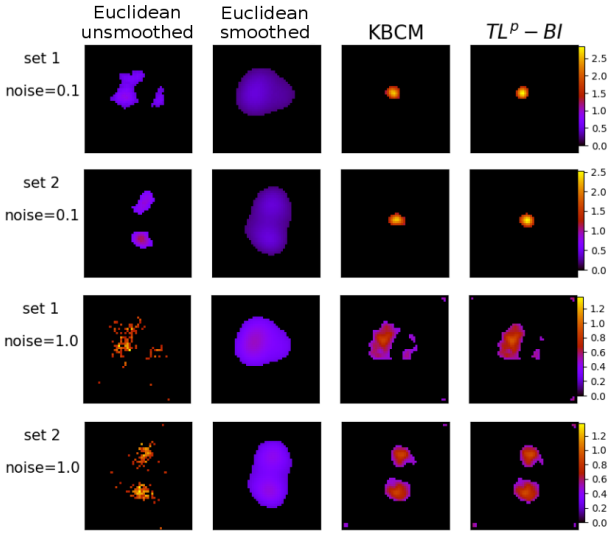

Fig. 1. Barycenters computed from artificial data with weak and strong noise, in the two top and bottom rows. Two groups of subjects are used (set1 and set2).

Fig. 1 illustrates the effects of the variability on optimal transport barycenters and on the Euclidean averagings used in the standard GLM (with or without spatial smoothing of the data). For this, we show the results for weak and strong noise levels (respectively in the two top and bottom rows), for two groups of subjects (set1 and set2). Only optimal transport methods offer a summary representation of the population which corresponds to the model that was used to generate the data, i.e that show one small activated region. But this is the case only with a weak noise level (top-right four images). In all other cases, the obtained images show high-intensity pixels that are more spread out throughout the images, eventually divided into several regions, and that depend on the content of the group of subjects itself which are different between set1 and set2.

[1] at https://openfmri.org/dataset/ds000158

Fig. 2 presents the quantitative comparisons of the KBCM and TL^p-BI algorithms. Fig. 2a and Fig. 2b show that both algorithms yield equivalent results: one small focus of activation when the noise is weak (few voxels above threshold, small spread), as expected from the generative model – note that this cannot be obtained with standard group analysis; and a deviation from this (increasing number of voxels above threshold, increasing spread) when the noise gets stronger. Fig. 2c shows that when noise level increases, the peak amplitude of the barycenters decreases. When the noise level is weak, TL^p-BI yields barycenters with higher peak amplitude than those of KBCM ($p < 0.01$ for 0.1, $p < 0.05$ for 0.3 and $p < 0.001$ for 0.5).

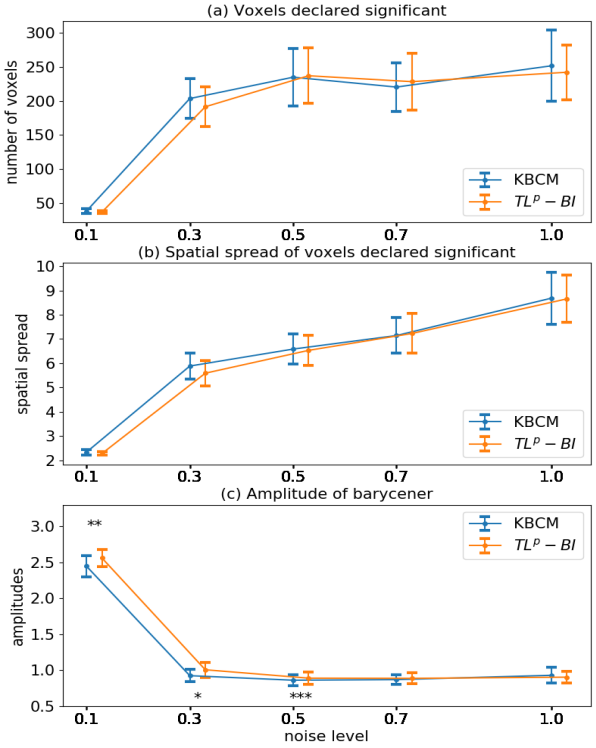

Fig. 2. Comparisons of KBCM and TL^p-BI on artificial fMRI data. * represents $p < 0.05$, ** denotes $p < 0.01$ and *** represents $p < 0.001$

E. Results on real fMRI data

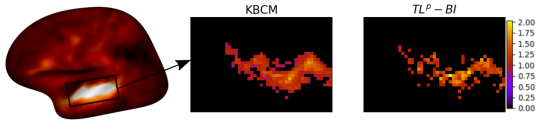

Fig. 3. Barycenters of real fMRI data.

On Fig. 3, we display an example of the obtained barycenters using KBCM and TL^p-BI. From a qualitative point of view, the map produced by KBCM appears smoother while the barycenter obtained with TL^p-BI provides finer spatial details. We compare the peak amplitudes of the barycenters obtained with the two algorithms and show (see Fig. 4) that our algorithm provides higher amplitude than KBCM (paired t-test, $p < 0.01$). Further work is clearly needed to better compare these two algorithms on various real fMRI datasets, and in particular on this dataset to assess whether optimal transport methods can help recover a multi-patch structure such as the one described in [7].

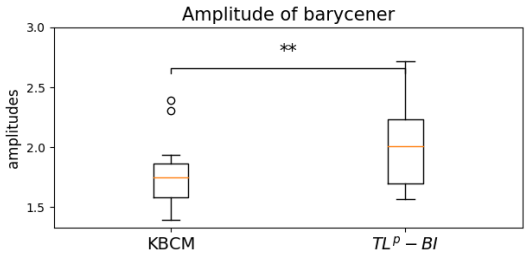

Fig. 4. Comparison of amplitudes on real fMRI data. ** denotes $p < 0.01$

IV. Discussion and Conclusion

In this paper, we introduced a new algorithm to compute a barycenter from a set of images using techniques from the optimal transportation theory. Our experimental results are somehow mixed. First, we confirm the encouraging results presented in [6] – that were obtained with a different algorithm – i.e that Wasserstein-like barycenters can allow overcoming inter-individual differences in a population. However, we demonstrated that with an increased level of noise in the input data, or with more complex real fMRI data, this property does not seem to fully hold, showing that further methodological and experimental work are needed to evaluate their full applicability. We also showed that overall, both algorithms, ours and the one of [6], offer equivalent results. It is however to be noted that during our experiments, our algorithm was on average 12 times faster, making it a good candidate to perform optimal transport-based inference at the population level with statistical tools such as permutation tests. For this, we provide a Python implementation of our TL^p-BI algorithm at https://github.com/SylvainTakerkart/PRNI2018_TLp_bary.

Acknowledgment

This work benefited from the support of the CNRS funding from the Défi Imag'In MultiAdopt project.

References

[1] M. Agueh, and G. Carlier, "Barycenters in the wasserstein space," SIAM Journal on Mathematical Analysis, 43(2), pp.904–924, 2011.

[2] J.D. Benamou, G. Carlier, M. Cuturi, L. Nenna, and G. Peyré, "Iterative bregman projections for regularized transportation problems," SIAM Journal on Scientific Computing, 37(2), pp.A1111–A1138, 2015.

[3] L. Chizat, G. Peyré, B. Schmitzer, and F.X. Vialard, "Scaling algorithms for unbalanced transport problems," arXiv preprint arXiv:1607.05816, 2016.

[4] M. Cuturi, and A. Doucet, "Fast computation of Wasserstein barycenters," In International Conference on Machine Learning pp. 685–693, January 2014.

[5] M. Cuturi, "Sinkhorn distances: Lightspeed computation of optimal transport," In Advances in neural information processing systems, pp. 2292–2300, 2013.

[6] A.Gramfort, G. Peyré, and M. Cuturi, "Fast optimal transport averaging of neuroimaging data," In International Conference on Information Processing in Medical Imaging, pp. 261–272, June 2015.

[7] C.R. Pernet, P. McAleer, M. Latinus, K.J. Gorgolewski, I. Charest, P.E. Bestelmeyer, R.H. Watson, D. Fleming, F. Crabbe, M. Valdes-Sosa, and P. Belin, "The human voice areas: Spatial organization and inter-individual variability in temporal and extra-temporal cortices," Neuroimage, 119, pp.164–174, 2015.

[8] M. Thorpe, S.Park, S.Kolouri, G.K. Rohde, and D. Slepev, "A Transportation L^p Distance for Signal Analysis," Journal of Mathematical Imaging and Vision, 59(2), pp.187–210, 2017.