Fast and Scalable Optimal Transport for Brain Tractograms

Jean Feydy, Pierre Roussillon, Alain Trouvé, Pietro Gori

Fast and Scalable Optimal Transport for Brain Tractograms

Jean Feydy[1,2*], Pierre Roussillon[3⋆], Alain Trouvé[1], and Pietro Gori[3]

[1] CMLA, ENS Paris-Saclay, France
[2] DMA, École Normale Supérieure, Paris, France
[3] LTCI, Télécom ParisTech, Institut Mines Télécom, Paris, France

Abstract. We present a new multiscale algorithm for solving regularized Optimal Transport problems on the GPU, with a linear memory footprint. Relying on Sinkhorn divergences which are convex, smooth and positive definite loss functions, this method enables the computation of transport plans between millions of points in a matter of minutes. We show the effectiveness of this approach on brain tractograms modeled either as bundles of fibers or as track density maps. We use the resulting smooth assignments to perform label transfer for atlas-based segmentation of fiber tractograms. The parameters – *blur* and *reach* – of our method are meaningful, defining the minimum and maximum distance at which two fibers are compared with each other. They can be set according to anatomical knowledge. Furthermore, we also propose to estimate a probabilistic atlas of a population of track density maps as a Wasserstein barycenter. Our CUDA implementation is endowed with a user-friendly PyTorch interface, freely available on the `PyPi` repository (`pip install geomloss`) and at www.kernel-operations.io/geomloss.

1 Introduction

Optimal Transport. Matching unlabeled distributions is a fundamental problem in mathematics. Traced back to the work of Monge in the 1780's, Optimal Transport (OT) theory is all about finding *transport plans* that minimize a ground cost metric under marginal constraints, which ensure a full covering of the input measures – see (1) and [11] for a modern overview. Understood as a canonical way of lifting distances from a set of *points* to a space of *distributions*, the OT problem has appealing geometric properties: it is robust to a large class of deformations, ensuring a perfect retrieval of global translations *and* small perturbations. Unfortunately though, standard combinatorial solvers for this problem scale in $O(n^3)$ with the number of samples. Until recently, OT has thus been mostly restricted to applications in economy and operational research, with at most a thousand points per distribution.

Fast computational methods. In the last forty years, a considerable amount of work has been devoted to approximate numerical schemes that allow users

⋆ contributed equally to this work

to *trade time for accuracy*. The two most popular iterative solvers for large-scale Optimal Transport, the *auction* and the *Sinkhorn* (or *IPFP*, *SoftAssign*) algorithms, can both be understood as *coordinate ascent* updates on a relaxed dual problem [10]: assuming a tolerance ε on the final result, they allow us to solve the transportation problem in time $O(n^2/\varepsilon)$. They met widespread diffusion in the computer vision literature at the turn of the XXI$^{\text{st}}$ century [3].

Recent progress. As of 2019, two lines of work strive to take advantage of the growth of (parallel) computing power to speed up large-scale Optimal Transport. The first track, centered around *multiscale strategies*, leverages the inner structure of the data with a coarse-to-fine scheme [13]. It was recently showcased to the medical imaging community [9] and provides solvers that generally scale in $O(n \log(n/\varepsilon))$ on the CPU, assuming an efficient octree-like decomposition of the input data. On the other hand, a second line of works puts a strong focus on entropic regularization and highly-parallel, GPU implementations of the Sinkhorn algorithm [4,11]. Up to a clever *debiasing* of the regularized OT problem [12], authors in the Machine Learning community were able to provide the first strong *theoretical guarantees* for an approximation of Optimal Transport, from principled extensions for the *unbalanced* setting [2] to proofs of positivity, definiteness and convexity [6,14].

Our contribution. This paper is about merging together these two bodies of work, as we provide the *first GPU implementation of a multiscale OT solver*. Detailed in section 2, our multiscale Sinkhorn algorithm is a *symmetric* coarse-to-fine scheme that takes advantage of key ideas developed in the past five years. Leveraging the routines of the `KeOps` library [1], our `PyTorch` implementation re-defines the state-of-the-art for discrete OT, allowing users to handle high-dimensional feature spaces in a matter of minutes. It is freely available on the `PyPi` repository (`pip install geomloss`) and at www.kernel-operations.io/geomloss.

White matter segmentation. From diffusion MR images, probabilistic tractography algorithms can reconstruct the architecture of the human brain white matter as bundles of 3D polylines [18], called fibers, or as track density maps [16]. In clinical neurology and neurosurgery, a task of interest is the segmentation of the white matter into anatomically relevant tracts. This can be carried out by: 1- manually delineating Regions of Interest (ROIs), which is tedious and time-consuming; 2- directly modeling the anatomical definitions [5]; 3- using learning strategies [17]; or 4- transferring labels of an atlas via non-linear deformations and clustering algorithms [15,8]. This last class of methods usually depends on several hyperparameters (e.g. number of clusters, kernel size). In this paper, similarly to [15], we propose to use Optimal Transport for transferring the labels of a fiber atlas to a subject tractogram. Leveraging the proposed efficient and multi-resolution implementation, we are able to directly segment a whole brain tractogram without any pre-processing or clustering. The entire algorithm only depends on two meaningful hyperparameters, the *blur* and *reach* scales, that define the minimum and maximum distances at which two fibers or densities are compared with each other. On top of label transfer, we also pro-

pose to estimate a probabilistic atlas of a population of track density maps as a Wasserstein barycenter where each map describes, for every voxel in the space, the probability that a specific track (e.g. IFOF) passes through.

2 Methods

We now give a detailed exposition of our *symmetric*, *unbiased*, *multiscale* Sinkhorn algorithm for discrete OT. We will focus on *vector data* and assume that our discrete, positive measures α and β are encoded as weighted sums of Dirac masses: $\alpha = \sum_{i=1}^{\mathrm{N}} \alpha_i \, \delta_{x_i}$ and $\beta = \sum_{j=1}^{\mathrm{M}} \beta_j \, \delta_{y_j}$ with weights $\alpha_i, \beta_j \geqslant 0$ and samples' locations $x_i, y_j \in \mathcal{X} = \mathbb{R}^{\mathrm{D}}$. We endow our feature space $\mathcal{X} = \mathbb{R}^{\mathrm{D}}$ with a cost function $\mathrm{C}(x, y) = \frac{1}{p}\|x - y\|^p$, with $p \in [1, 2]$ ($p = 2$ in this paper), and recall the standard Monge-Kantorovitch transportation problem:

$$\mathrm{OT}(\alpha, \beta) \;=\; \min_{\pi \in \mathbb{R}^{\mathrm{N}\times\mathrm{M}}_{\geqslant 0}} \; \textstyle\sum_{i,j} \pi_{i,j} \, \mathrm{C}(x_i, y_j) \quad \text{s.t.} \quad (\pi \, \mathbf{1})_i = \alpha_i, \; (\pi^\intercal \mathbf{1})_j = \beta_j. \quad (1)$$

In most applications, the feature space $\mathcal{X}$ is the ambient space $\mathbb{R}^3$ endowed with the standard Euclidean metric. This is the case, for instance, when using *track density maps* where α_i and β_j are the probabilities associated to the voxel locations x_i and y_j, respectively. Meanwhile, when using fiber tractograms, a usual strategy is to resample each fiber to the same number of points P. In this case, the feature space $\mathcal{X}$ becomes $\mathbb{R}^{\mathrm{P}\times 3}$ and x_i, y_j are the N and M fibers that constitute the source and target tractograms with uniform weights $\alpha_i = \frac{1}{\mathrm{N}}$ and $\beta_j = \frac{1}{\mathrm{M}}$, respectively. Distances can be computed using the standard Euclidean L^2 norm – normalized by $1/\sqrt{\mathrm{P}}$ – and we alleviate the problem of *fiber orientation* by augmenting our tractograms with the mirror flips of all fibers. This corresponds to the simplest of all encodings for unoriented curves.

Robust, regularized Optimal Transport. Following [2,11], we consider a generalization of (1) where α and β don't have the same total mass or may contain *outliers*: this is typically the case when working with fiber bundles. Paving the way for efficient computations, the relaxed OT problem reads:

$$\begin{aligned} \mathrm{OT}_{\varepsilon,\rho}(\alpha, \beta) \;&=\; \min_{\pi \in \mathbb{R}^{\mathrm{N}\times\mathrm{M}}_{\geqslant 0}} \; \langle \pi_{i,j} \, , \, \mathrm{C}(x_i, y_j) \rangle \;+\; \varepsilon \, \mathrm{KL}(\pi_{i,j}, \alpha_i \otimes \beta_j) \qquad (2) \\ &\qquad\qquad +\; \rho \, \mathrm{KL}((\pi \mathbf{1})_i, \alpha_i) \;+\; \rho \, \mathrm{KL}((\pi^\intercal \mathbf{1})_j, \beta_j) \\ &=\; \max_{f \in \mathbb{R}^N, g \in \mathbb{R}^M} \; \rho \, \langle \alpha_i, 1 - e^{-f_i/\rho} \rangle \;+\; \rho \, \langle \beta_j, 1 - e^{-g_j/\rho} \rangle \qquad (3) \\ &\qquad\qquad +\; \varepsilon \, \langle \alpha_i \otimes \beta_j, 1 - \exp \tfrac{1}{\varepsilon} \big[f_i \oplus g_j - \mathrm{C}(x_i, y_j) \big] \rangle, \end{aligned}$$

where $\mathrm{KL}(a_i, b_i) = \sum a_i \log(\frac{a_i}{b_i}) - a_i + b_i$ denotes the (generalized) Kullback-Leibler divergence and $\varepsilon, \rho > 0$ are two *positive* regularization parameters homogeneous to the cost function C. The equality between the primal and dual problems (2–3) – *strong duality* – holds through the Fenchel-Rockafellar theorem. With a *linear* memory footprint, optimal dual vectors f and g encode a solution of the $\mathrm{OT}_{\varepsilon,\rho}$ problem given by:

$$(\pi_{\varepsilon,\rho})_{i,j} \;=\; \alpha_i \beta_j \; \exp \tfrac{1}{\varepsilon} \big[f_i + g_j - \mathrm{C}(x_i, y_j) \big]. \qquad (4)$$

Unbiased Sinkhorn divergences. Going further, we consider

$$\mathrm{S}_{\varepsilon,\rho}(\alpha,\beta) = \mathrm{OT}_{\varepsilon,\rho}(\alpha,\beta) - \tfrac{1}{2}\mathrm{OT}_{\varepsilon,\rho}(\alpha,\alpha) - \tfrac{1}{2}\mathrm{OT}_{\varepsilon,\rho}(\beta,\beta) + \tfrac{\varepsilon}{2}(\textstyle\sum\alpha_i - \sum\beta_j)^2, \quad (5)$$

the *unbiased* Sinkhorn divergence that was recently shown to define a *positive*, *definite* and *convex* loss function for measure-fitting applications – see [6] for a proof in the balanced setting, extended in [14] to the general case.

Generalizing ideas introduced by [7], we can write $\mathrm{S}_{\varepsilon,\rho}$ in *dual* form as

$$\begin{aligned} \mathrm{S}_{\varepsilon,\rho}(\alpha,\beta) \;=\; & -\; (\rho+\tfrac{\varepsilon}{2})\,\langle \alpha_i, e^{-b_i^{\beta\to\alpha}/\rho} - e^{-a_i^{\alpha\leftrightarrow\alpha}/\rho}\rangle \\ & -\; (\rho+\tfrac{\varepsilon}{2})\,\langle \beta_j, e^{-a_j^{\alpha\to\beta}/\rho} - e^{-b_j^{\beta\leftrightarrow\beta}/\rho}\rangle, \end{aligned} \quad (6)$$

where $(f_i, g_j) = (b_i^{\beta\to\alpha}, a_j^{\alpha\to\beta})$ is a solution of $\mathrm{OT}_{\varepsilon,\rho}(\alpha,\beta)$ and $a_i^{\alpha\leftrightarrow\alpha}$, $b_j^{\beta\leftrightarrow\beta}$ correspond to the unique solutions of $\mathrm{OT}_{\varepsilon,\rho}(\alpha,\alpha)$ and $\mathrm{OT}_{\varepsilon,\rho}(\beta,\beta)$ on the diagonal of the space of dual pairs. We can then derive the equations at optimality for the three $\mathrm{OT}_{\varepsilon,\rho}$ problems and write the gradients $\partial_{\alpha_i}\mathrm{S}_{\varepsilon,\rho}$, $\partial_{x_i}\mathrm{S}_{\varepsilon,\rho}$, $\partial_{\beta_j}\mathrm{S}_{\varepsilon,\rho}$, $\partial_{y_j}\mathrm{S}_{\varepsilon,\rho}$ as expressions of the *four dual vectors* $b_i^{\beta\to\alpha}, a_i^{\alpha\leftrightarrow\alpha} \in \mathbb{R}^{\mathrm{N}}$, $a_j^{\alpha\to\beta}, b_j^{\beta\leftrightarrow\beta} \in \mathbb{R}^{\mathrm{M}}$ [6,11].

Multiscale Sinkhorn algorithm. To estimate these dual vectors, we propose a *symmetrized* Sinkhorn loop that retains the iterative structure of the baseline SoftAssign algorithm, but replaces alternate updates with *averaged* iterations – for the sake of symmetry – and uses the ε-scaling heuristic of [10]:

Parameters: Positive **blur** and **reach** scales;
Multiscale heuristic: diameter $d = \max_{i,j}\|x_i - y_j\|$; **scaling** $q = 0.9$.

$a_i^{\alpha\leftrightarrow\alpha},\ b_j^{\beta\leftrightarrow\beta},\ a_j^{\alpha\to\beta},\ b_i^{\beta\to\alpha} \;\leftarrow\; \mathbf{0}_{\mathbb{R}^{\mathrm{N}}},\ \mathbf{0}_{\mathbb{R}^{\mathrm{M}}},\ \mathbf{0}_{\mathbb{R}^{\mathrm{M}}},\ \mathbf{0}_{\mathbb{R}^{\mathrm{N}}}$

for σ **in** $[\,d,\ d\cdot q,\ d\cdot q^2,\ \cdots,\ \mathrm{blur}\,]$ **do** ▷ $\lceil \log(d/\mathrm{blur})/\log(1/q)\rceil$ iterations

  $\varepsilon,\ \rho,\ \lambda \;\leftarrow\; \sigma^p,\ \mathrm{reach}^p,\ 1/(1+(\sigma/\mathrm{reach})^p)$

  $\tilde{a}_i^{\alpha\leftrightarrow\alpha} \leftarrow -\lambda\,\varepsilon \log \sum_{k=1}^{\mathrm{N}} \alpha_k \ \exp\tfrac{1}{\varepsilon}[\,a_k^{\alpha\leftrightarrow\alpha} - \mathrm{C}(x_k, x_i)\,]$

  $\tilde{b}_j^{\beta\leftrightarrow\beta} \leftarrow -\lambda\,\varepsilon \log \sum_{k=1}^{\mathrm{M}} \beta_k \ \exp\tfrac{1}{\varepsilon}[\,b_k^{\beta\leftrightarrow\beta} - \mathrm{C}(y_k, y_j)\,]$

  $\tilde{a}_j^{\alpha\to\beta} \leftarrow -\lambda\,\varepsilon \log \sum_{k=1}^{\mathrm{N}} \alpha_k \ \exp\tfrac{1}{\varepsilon}[\,b_k^{\beta\to\alpha} - \mathrm{C}(x_k, y_j)\,]$

  $\tilde{b}_i^{\beta\to\alpha} \leftarrow -\lambda\,\varepsilon \log \sum_{k=1}^{\mathrm{M}} \beta_k \ \exp\tfrac{1}{\varepsilon}[a_k^{\alpha\to\beta} - \mathrm{C}(y_k, x_i)\,]$

  $a_i^{\alpha\leftrightarrow\alpha},\ b_j^{\beta\leftrightarrow\beta} \;\leftarrow\; \tfrac{1}{2}(a_i^{\alpha\leftrightarrow\alpha} + \tilde{a}_i^{\alpha\leftrightarrow\alpha}),\ \ \tfrac{1}{2}(b_j^{\beta\leftrightarrow\beta} + \tilde{b}_j^{\beta\leftrightarrow\beta})$

  $a_j^{\alpha\to\beta},\ b_i^{\beta\to\alpha} \;\leftarrow\; \tfrac{1}{2}(a_j^{\alpha\to\beta} + \tilde{a}_j^{\alpha\to\beta}),\ \ \tfrac{1}{2}(b_i^{\beta\to\alpha} + \tilde{b}_i^{\beta\to\alpha})$

return $a_i^{\alpha\leftrightarrow\alpha},\ b_j^{\beta\leftrightarrow\beta},\ a_j^{\alpha\to\beta},\ b_i^{\beta\to\alpha}$ ▷ Optimal dual vectors.

Coarse-to-fine strategy. To speed-up computations and tend towards the $O(n\log(n))$ complexity of multiscale methods, we use a coarse subsampling of the x_i's and y_j's in the first few iterations. Here, we use a simple K-means algorithm, but other strategies could be employed. We can then use the coarse dual vectors to *prune out* useless computations at full resolution and thus implement the *kernel truncation trick* introduced by [13]. Note that to perform these operations *on the GPU*, our code heavily relies on the *block-sparse*, online,

generic reduction routines provided by the `KeOps` library [1]. Our implementation provides a `x1,000` speed-up when compared to simple `PyTorch` GPU implementations of the Sinkhorn loop [4], while keeping a linear (instead of quadratic) memory footprint. Benchmarks may be found on our website.

The *blur* and *reach* parameters. In practice, the only two parameters of our algorithm are the *blur* and *reach* scales, which specify the temperature ε and the strength ρ of the soft covering constraints. Intuitively, *blur* is the resolution of the finest details that we try to capture, while *reach* acts as an upper bound on the distance that points may travel to meet their targets – instead of seeing them as *outliers*. In our applications, they define the minimum and maximum distances at which two fibers or densities are compared with each other.

Label transfer. Given a subject tractogram α and a segmented fiber atlas β with L classes, we propose to transfer the labels from the atlas to the subject with an OT matching. Having computed optimal dual vectors $(f_i, g_j) = (b_i^{\beta\to\alpha}, a_j^{\alpha\to\beta})$ solutions of $\mathrm{OT}_{\varepsilon,\rho}(\alpha,\beta)$, we encode the atlas labels as one-hot vectors ℓ_j in the probability simplex of $\mathbb{R}^{\mathrm{L}}$, and compute soft segmentation scores $\mathrm{Lab}_i = \frac{1}{\alpha_i}(\pi_{\varepsilon,\rho}\ell)_i$ using the implicit encoding (4):

$$\mathrm{Lab}_i = \textstyle\sum_{j=1}^{\mathrm{M}} \beta_j\, \ell_j \;\; \exp\tfrac{1}{\varepsilon}[\, f_i + g_j - \mathrm{C}(x_i, y_j)\,] \;\in\; \mathbb{R}^{\mathrm{L}}. \tag{7}$$

If a fiber x_i is mapped by $\pi_{\varepsilon,\rho}$ to target fibers y_j of the atlas β, the marginal constraints ensure that Lab_i is a *barycentric* combination of the corresponding labels ℓ_j on the probability simplex. However, if x_i is an *outlier*, the i-th line of $\pi_{\varepsilon,\rho}$ will be almost zero and Lab_i will be close to $\mathbf{0}_{\mathbb{R}^{\mathrm{L}}}$. In this way, we can detect and discard aberrant fibers. Please note that the OT plan is computed on the augmented data, using both original and flipped fibers; only one version of each fiber will be matched to the atlas and thus kept for the transfer of labels.

Track density atlas. Given K *track density maps* of a specific bundle encoded as normalized volumes $\beta_1, \ldots, \beta_{\mathrm{K}}$, we also propose to estimate a probabilistic atlas as a *Wasserstein barycenter* (or *Fréchet mean*) by minimizing $\frac{1}{\mathrm{K}}\sum_{k=1}^{\mathrm{K}} \mathrm{S}_{\varepsilon,\rho}(\alpha,\beta_k)$ with respect to the points $x_i \in \mathbb{R}^3$ of the atlas. This procedure relies on the computation of $\mathrm{S}_{\varepsilon,\rho}(\alpha,\beta_k)$ with the squared Wasserstein-2 distance when $p = 2$, *reach* $= +\infty$ and *blur* is equal to the voxel size.

3 Results and Discussion

Dataset. Experiments below are based on 5 randomly selected healthy subjects of the *HCP*[4] dataset. Whole-brain tractograms of one million fibers are estimated with MRTrix3[5] using a probabilistic algorithm (iFOD2) and the Constrained Spherical Deconvolution model. Segmentations of the inferior fronto-occipital fasciculus (IFOF) are obtained as described in [5], and track density maps are computed using MRTrix3. We ran our scripts on an Intel Xeon E5-1620V4 with a GeForce GTX 1080.

[4] https://db.humanconnectome.org
[5] http://www.mrtrix.org

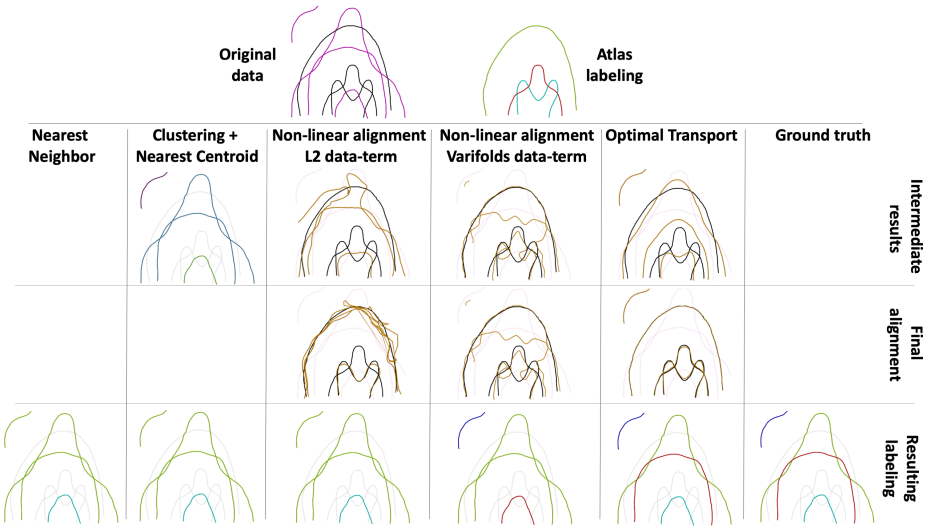

Fig. 1: Label transfer between an atlas (black) and a subject (magenta) using different strategies. (1) Subject fibers are assigned the label of the closest atlas fiber. (2) Subject fibers are first clustered (K-means, K=3) and then each cluster is assigned the label of the closest atlas cluster. (3,4) Subject fibers are non-linearly aligned (i.e. diffeomorphism) to the atlas using a L2 data-term and a varifolds data-term. Labels are estimated from the resulting deformation. (5) Correspondences between subject and atlas fibers are estimated using the proposed algorithm for optimal transport. Second and third row show intermediate results (subject clustering and estimated alignement). In the last row, we present the transferred labels. Fibers detected as outliers are shown in dark blue.

Toy example. In Fig. 1, we compare different existing strategies for transferring labels from a fiber atlas to a subject bundle with an U-shape. In this example, there is a bijection between the atlas and the subject bundle except for an outlier. The only method that estimates the correct correspondence and finds the outlier is the proposed one since, unlike the first two strategies, it takes into account the organization of the bundle. Note that for the second strategy, a different value of K would not have changed the resulting labeling. The third and fourth strategies are based on diffeomorphic deformations and, regardless of the data-term, cannot correctly align the crossing fibers.

Label transfer. In Fig. 2, we show some of the labels transferred from the atlas proposed in [18] towards one test subject of the HCP dataset. The atlas is composed of 800 clusters, with $\sim$ 800K fibers. To speed up computations and improve interpretability, we separately analyze the two brain hemispheres and the inter-hemispheric fibers – here, we only show results for the left hemisphere. To validate our results, we also compare the proposed segmentation of the left IFOF with a manual segmentation. Results indicate that the estimated labels are coherent with the clusters of the atlas and with the manual segmentation.

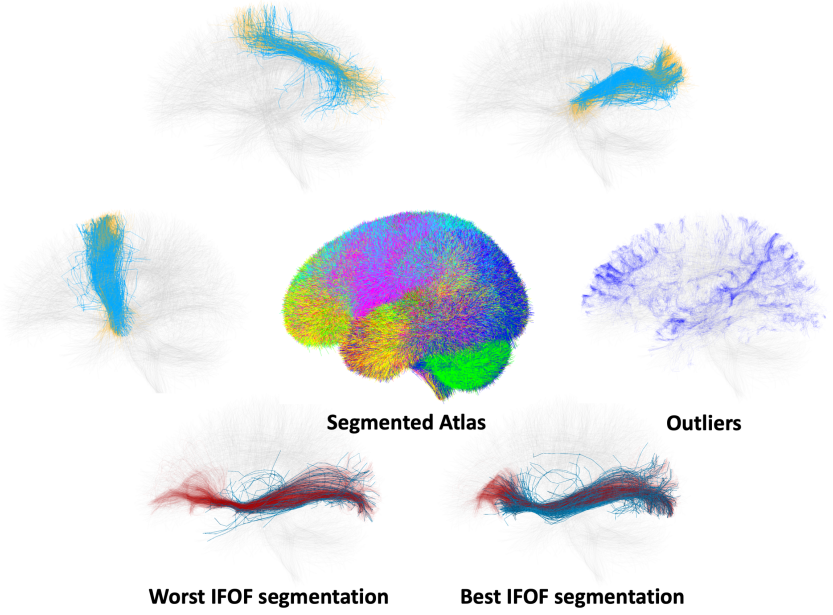

Fig. 2: Label transfer between the segmented atlas (in the middle) and the subject tractograms. Top: some clusters of the atlas (in orange) with their respective segmentations (in light blue) of one random subject. Detected outliers are on the right (dark blue). Bottom: worst and best segmentation of the left IFOF, among the five tested subjects, compared to a manual segmentation.

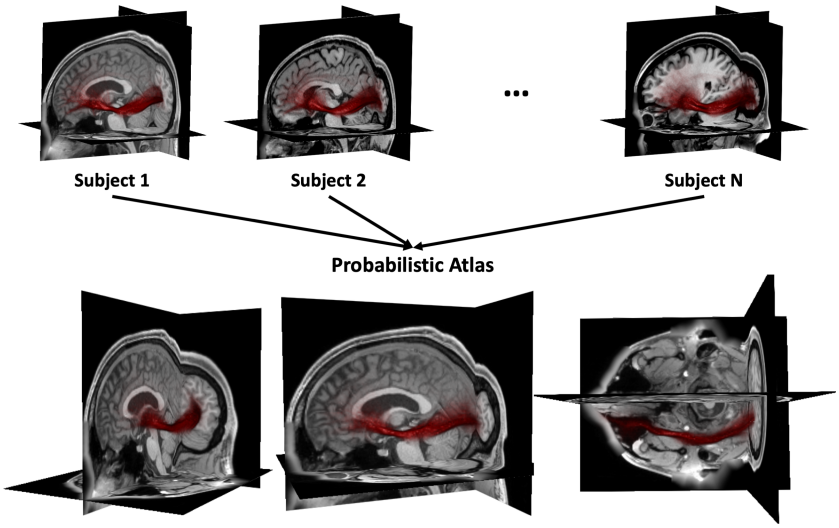

Fig. 3: Probabilistic atlas of the left IFOF from 5 track density maps (in red). Top row: the densities of the five subjects, shown with their T1-wMRI. Bottom row: three views of the obtained atlas, alongside the T1-wMRI of one subject.

The *blur* and the *reach* are empirically set to 2mm and 20mm respectively. The computation time of a transport plan for the left hemisphere is about 10 minutes (with $\sim$ 250K and $\sim$ 200K fibers sampled with P = 20 points for the subject and the atlas, respectively).

Track density atlas. Using 5 track density maps of the left IFOF, we estimate a probabilistic atlas shown in Fig. 3. Each density map is a volume of $145\times174\times145$ voxels of 1mm^3 where approximately 20,000 voxels have a probability (i.e. β_j) different from zero. The barycenter is initialized with the arithmetic average of the dataset densities and then up-sampled (factor of 6) to increase the resolution, resulting in $\sim$400,000 Dirac samples. The *blur* is equal to the voxel size and the *reach* is set to $+\infty$. The computation time to estimate the Wasserstein barycenter is 9 seconds.

Conclusion. We presented an efficient, fast and scalable algorithm for solving the regularised (entropic) Optimal Transport problem and showed its effectiveness in two applications. First, OT plans can be used to transport labels from a fiber atlas to a subject tractogram. This method takes into account the organization of the bundles – unlike standard nearest neighbours or clustering algorithms –, detects outliers and is not hampered by fiber crossings. Second, we use Sinkhorn divergences to estimate a *geometric* average of track density maps. In future works, we plan to study more relevant features for modeling white-matter fibers, including for instance directional and FA information.

Acknowledgement This research was partially supported by Labex DigiCosme (ANR) as part of the program Investissement d'Avenir Idex ParisSaclay.

References

1. Charlier, B., Feydy, J., Glaunès, J.: Kernel Operations on the GPU, with autodiff, without memory overflows, https://www.kernel-operations.io/
2. Chizat, L., Peyré, G., Schmitzer, B., Vialard, F.X.: An interpolating distance between optimal transport and Fisher–Rao metrics. Foundations of Computational Mathematics **18**(1), 1–44 (2018)
3. Chui, H., Rangarajan, A.: A new point matching algorithm for non-rigid registration. Computer Vision and Image Understanding **89**(2-3), 114–141 (2003)
4. Cuturi, M.: Sinkhorn distances: Lightspeed computation of optimal transport. In: NIPS. pp. 2292–2300 (2013)
5. Delmonte, A., Mercier, C., Pallud, J., Bloch, I., Gori, P.: White matter multi-resolution segmentation using fuzzy set theory. In: IEEE ISBI. Venice, Italy (2019)
6. Feydy, J., Séjourné, T., Vialard, F.X., Amari, S.I., Trouvé, A., Peyré, G.: Interpolating between Optimal Transport and MMD using Sinkhorn divergences. In: AiStats (2019)
7. Feydy, J., Trouvé, A.: Global divergences between measures: from Hausdorff distance to Optimal Transport. In: ShapeMI, MICCAI workshop. pp. 102–115 (2018)
8. Garyfallidis, E., Côté, M.A., Rheault, F., Sidhu, J., Hau, J., et al.: Recognition of white matter bundles using local and global streamline-based registration and clustering. NeuroImage **170**, 283–295 (2018)

9. Gerber, S., Niethammer, M., Styner, M., Aylward, S.: Exploratory population analysis with unbalanced optimal transport. In: MICCAI. pp. 464–472 (2018)
10. Kosowsky, J., Yuille, A.L.: The invisible hand algorithm: Solving the assignment problem with statistical physics. Neural networks **7**(3), 477–490 (1994)
11. Peyré, G., Cuturi, M., et al.: Computational optimal transport. Foundations and Trends in Machine Learning **11**(5-6), 355–607 (2019)
12. Ramdas, A., Trillos, N., Cuturi, M.: On wasserstein two-sample testing and related families of nonparametric tests. Entropy **19**(2), 47 (2017)
13. Schmitzer, B.: Stabilized sparse scaling algorithms for entropy regularized transport problems. arXiv preprint arXiv:1610.06519 (2016)
14. Séjourné, T., Feydy, J., Vialard, F.X., Trouvé, A., Peyré, G.: Sinkorn divergences for unbalanced Optimal Transport. To appear.
15. Sharmin, N., Olivetti, E., Avesani, P.: White Matter Tract Segmentation as Multiple Linear Assignment Problems. Frontiers in Neuroscience **11** (2018)
16. Wassermann, D., Bloy, L., Kanterakis, E., Verma, R., Deriche, R.: Unsupervised white matter fiber clustering and tract probability map generation. NeuroImage **51**(1), 228–241 (2010)
17. Wasserthal, J., Neher, P., Maier-Hein, K.H.: TractSeg - Fast and accurate white matter tract segmentation. NeuroImage **183**, 239–253 (2018)
18. Zhang, F., Wu, Y., Norton, I., Rigolo, L., Rathi, Y., Makris, N., O'Donnell, L.J.: An anatomically curated fiber clustering white matter atlas for consistent white matter tract parcellation across the lifespan. NeuroImage **179**, 429–447 (2018)