6.838: Shape Analysis, 2021
Instructor: Justin Solomon
TAs: David Palmer and Paul Zhang

Optimal Transport based Probabilistic Diffeomorphic Registration

Peiqi Wang

MIT Department of Electrical Engineering and Computer Science

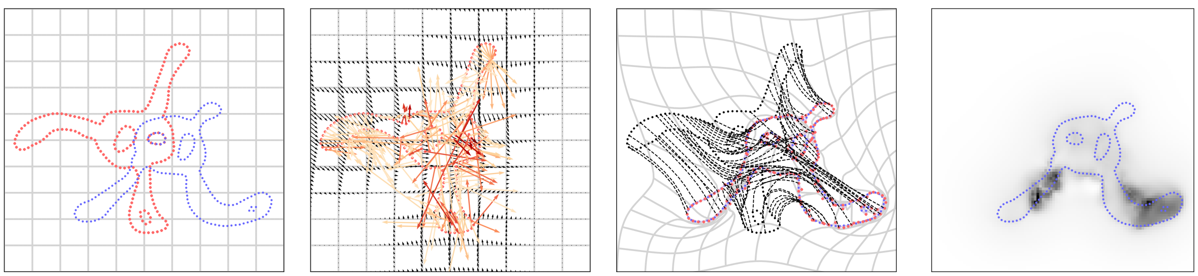

Figure 1: *The red amoeba is registered to blue amoeba. A valid diffeomorphic transformation can be generated from per-particle momenta (orange arrow) by solving a set of Geodesic equations (black dashed line). On the right, we can visualize variance of the effect of random transformation on shape as a heatmap. Although the two shapes is matched almost perfectly, the two legs flips over itself respectively and we are able to spot this mistake from the uncertainty heatmap.*

Abstract

1. Introduction

Diffeomorphic registration of shapes with unknown correspondence is an important step in medical data processing. The choice of similarity metric for good matching distinguishes the different algorithms in this domain. Recent work explored the use of entropic regularized wasserstein distance as a global measure of similarity between encoding of shapes as discrete measures [FCVP17, FRTG19]. However, a point estimate of the transformation yield errors that may invalidate downstream processing pipelines or misguide clinical decision making. Additionally, solution is sensitive to hyperparameters of the problem, requiring manual tuning for each new shape.

We propose to extend optimal transport based diffeomorphic registration to probabilistic setting. Our method interprets the diffeomorphic transformation as a random variable, and estimates its parameters using variational inference. In particular, we find diffeomorphic maps that minimize the average optimal transport distance between shapes. Naturally, the probabilistic formulation provides us with uncertainty estimates of both the transformation as well as the uncertainty of its effect on shapes. Hyperparameters such as the degree of smoothness of the transformation parameterizes the variational distribution, and thus can be optimized. However, the inference procedure requires repeated sampling of valid diffeomorphic transformations to approximate the average cost. To alleviate the computational burden, we explored links to sparse Gaussian Process, specifically interdomain inducing variables, as a way to alleviate this concern [Fv09].

2. Related Work

2.1. Diffeomorphic Registration

The large deformation registration of landmarks (point sets) and images as a variational problem that solves for a smooth time-varying velocity field that matches the two objects according to some measure of similarity [JM00, BMTY05]. [MTY06, VRRC12] makes connection to optimal control, and showed that large deformation diffeormorphisms obeys conservation of momentum, and that points/images evolve according to a set of Geodesic Equations completely determined by its initial momentum. This observation prompted the development of geodesic shooting methods that optimizes for initial momentum for point sets and meshes [VMYD04, ATY05] and later extended to images [VRRC12]. In our formulation, we represent a random diffeomorphic map as a pushforward of initial momentum, parameterized by some easy-to-work-with distribution.

2.2. Optimal Transport

The optimal transport problem give rise to a notion of distance between probability distributions. Intuitively, it measures the amount of mass needed to be displaced from one measure to another. Computation of such distance is useful in numerous graphics [dGCSAD11, dGBOD12, SdGP*15] and learning [FZM*15, JCG20] applications. Developments in numerical methods in computing and differentiating through the optimal transport distance make for easier integration to a larger system [Cut13, AWR18, GPC18, Sch19]. A series of work [FCVP17, FRTG19] utilize unbalanced optimal transport distance to register shapes with possibly different mass. The motivation is to combine the elastic properties of diffeomorphic registration beneficial for medical data with a cost that is sensitive to both local and global shape features. We extend their methods to probabilistic settings.

2.3. Probabilistic Registration

Many probabilistic registration methods on image data rely on latent variable model, where an unknown transformation is applied to the source image to generate the target image, corrupted by some observation noise. Previously, Monte Carlo methods have been used to estimate transformation parameters [RJN*13, ZSF13]. Although asymptotically exact, sampling methods can be slow for estimating the high dimensional parameters of a diffeomorphism. A few follow ups suggests ways to reduce the number of parameters using factor analysis [RJN*13] or with an economical representation in the Fourier domain [WWGZ19]. Alternatively, variational inference has been proposed as a more tractable alternative for inference [WTNWI14, DBGS19]. Although similar in formulation, we differ from [WTNWI14] as we restrict ourselves to diffeomorphic registration of discrete representation of shapes with unknown correspondence. Theory from geodesic shooting implies that we do not need to model the Eulerian flow as the solution to some stochastic differential equation as proposed in [WTNWI14]. It suffices to model the initial momentum as some Gaussian Process whereby to specify a random diffeomorphic transformations.

3. Preliminaries

3.1. Shape as Measures

Let $\Omega \subset \mathbb{R}^D$ be a low dimensional ambient space. We consider a representation of shapes as discrete measures $\alpha = \sum_{i=1}^N a_i \delta_{x_i}, \beta = \sum_{j=1}^M b_i \delta_{y_j}$ where $a_i, b_j \in \mathbb{R}_+$ encodes some local statistics of shape, e.g. per-vertex length for curves or per-vertex area for surfaces, and $x := (x^1, \cdots, x^n) \subset \mathbb{R}^{N\times D}$, $y := (y^1, \cdots, y^n) \subset \mathbb{R}^{M\times D}$ are source and target points respectively.

3.2. Diffeomorphic Registration

The goal of diffeomorphic registration of point sets is to transform x via a diffeomorphic mapping φ such that the pushforward $\varphi_\sharp\alpha$ is close to y according to some similarity metric $\mathcal{L}(\varphi_\sharp\alpha, \beta)$. We consider a space of velocity fields V as a RKHS over Ω characterized by kernel $\bar{k} : \Omega \times \Omega \to \mathbb{R}^{D\times D}$. For purposes of computation, we consider an equivalent scalar-valued kernel $k : \Omega \times [D] \times \Omega \times [D] \to \mathbb{R}$ where $\bar{k}(x, x')_{dd'} = k((x, d), (x', d'))$, inducing an RKHS that is isomorphic to V. A diffeomorphism can be constructed via flows, i.e. solutions to an ODE problem φ_t where $\dot{\varphi}_t = v_t \circ \varphi_t, \varphi_0 = \text{Id}$, of a sufficiently smooth velocity field, i.e. $\int_0^1 \|v_t\|_V \, dt < \infty$. The large deformation registration problem solves for a time-varying velocity field $v_t \in V$ matching the two shapes,

$$\min_{v_t : t\in[0,1]} \frac{1}{2} \int_0^1 \|v_t\|_V^2 \, dt + \mathcal{L}(\varphi_\sharp\alpha, \beta) \tag{1}$$

3.3. Geodesic Shooting

Let $q_t^i = \varphi_t(x^i) \in \mathbb{R}^D$ be application of transformation φ_t to point x^i. Denote $q_t = (q_t^1, \cdots, q_t^N) \in \mathbb{R}^{ND}$ as action of φ_t to a set of points and $K(q_t, q_t) \in \mathbb{R}^{ND\times ND}$ be kernel matrix for vectorized velocity vector field at q_t. [JM00] argues for a Lagrangian view of previous Eulerian problem - it suffices to solve for the flow velocity $\dot{q}_t$ of particles,

$$\min_{\dot{q}_t : t\in[0,1]} \frac{1}{2} \int_0^1 \left\langle \dot{q}_t, K(q_t, q_t)^{-1} \dot{q}_t \right\rangle dt + \mathcal{L}(\varphi_\sharp\alpha, \beta) \tag{2}$$

and that the resulting flow velocity in the Eulerian coordinates Ω can be interpolated from $\dot{q}_t$,

$$v_t(x) = K(x, q_t) p_t \qquad p_t = K(q_t, q_t)^{-1} \dot{q}_t \tag{3}$$

where $p_t^i \in \mathbb{R}^D$ is the momenta associated with point q_t^i. As a side note, this interpolation is akin to computing posterior mean of a gaussian process regression of velocity fields, i.e. $v_t(x) = K(x, q_t) K(q_t, q_t)^{-1} \dot{q}_t$. Note the integrand of RKHS norm can be viewed as Lagrangian of of a system of ND particles. [MTY06] argues the dynamics of these particles in canonical coordinates (q_t, p_t) follows the Geodesic equations

$$\dot{q}_t = \frac{\partial \mathcal{H}(q_t, p_t)}{\partial p} \qquad \dot{p}_t = -\frac{\partial \mathcal{H}(q_t, p_t)}{\partial q} \tag{4}$$

with initial condition $q_0 := x, p_0$, and that the Hamiltonian

$$\mathcal{H}(q_t, p_t) = \frac{1}{2} \langle p_t, K(q_t, q_t) p_t \rangle = \frac{1}{2} \left\langle K(q_t, q_t)^{-1} \dot{q}_t, \dot{q}_t \right\rangle \tag{5}$$

is preseved by the flow, $\mathcal{H}(q_0, p_0) = \mathcal{H}(q_t, p_t)$ for all $t \in [0, 1]$ (see Figure (2)). Therefore, $\int_0^1 \mathcal{H}(q_t, p_t) \, dt = \mathcal{H}(q_0, p_0)$. Equivalent to (2), we optimize over the initial *shooting momentum* p_0,

$$\min_{p_0 \in \mathbb{R}^{ND}} \frac{1}{2} \langle p_0, K(x, x) p_0 \rangle + \mathcal{L}(\varphi_\sharp\alpha, \beta) \tag{6}$$

where $\varphi_\sharp\alpha = \sum_{i=1}^N a_i' q_1^i$ and a_i' is the updated weights as a function of updated vertex positions and unchanged topology. Important to note is $\varphi_\sharp\alpha$ is a specified entirely by x, p_0. Typically, gradient based methods for solving (6) involves a forward integration of (q_t, p_t) via (4) to obtain the pushforward measuer, compute the gradient of objective (6) with respect to initial momentum, and do gradient updates iteratively.

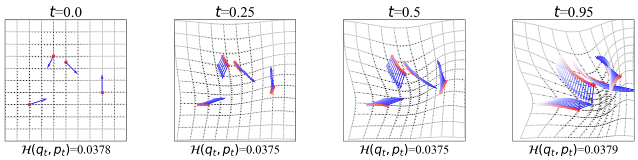

Figure 2: *Geodesic shooting with an Euler integrator with time step of $\delta t = .1$. The velocity field is represented using a radial basis kernel with $\sigma = .25$. We show trajectory of q_t (red dots) with momentum p_t (blue arrow) and interpolated velocity fields at grid points (black). We see Hamiltonian is approximately conserved!*

3.4. Unbalanced Regularized Optimal Transport

Define $c : \Omega^2 \to \mathbb{R}$ be cost of mass transportation, hereafter we let c be the squared Euclidean distance. The entropic regularized unbalanced optimal transport problem seeks a soft coupling between measures μ, ν supported over Ω that minimizes expected cost; The optimal value of which defines a measure of similarity between μ and ν,

$$\mathcal{W}_{\varepsilon,\rho}(\mu,\nu) := \min_{\pi \in \mathcal{P}(\Omega^2)} \iint_{\Omega^2} c(x,y)\, d\pi(x,y) + \varepsilon \mathrm{KL}(\pi \| \mu \otimes \nu) \tag{7}$$

$$+ \rho \left(\mathrm{KL}(P^1_\sharp \pi \| \mu) + \mathrm{KL}(P^2_\sharp \pi \| \nu) \right) \tag{8}$$

$P^i_\sharp$ is the projection to the i-th marginal. The correspondence of the optimal transport plan π can be made sharper by using a smaller $\varepsilon > 0$. If the two shapes vary in scale, we can relax the mass conservation constraints by using a smaller $\rho > 0$. For discrete measures, generalized Sinkhorn's algorithm computes $W_{\varepsilon,\rho}(\alpha,\beta)$ efficiently and in a numerically stable manner [CPSV17, FSV*18]. The idea is to do coordinate ascent over the dual variables $u \in \mathbb{R}^N, v \in \mathbb{R}^M$,

$$u^{(\ell+1)} \leftarrow \lambda\varepsilon \left(\log(a) - \log(K e^{v^{(\ell)}/\varepsilon}) \right) \tag{9}$$

$$v^{(\ell+1)} \leftarrow \lambda\varepsilon \left(\log(b) - \log(K e^{u^{(\ell+1)}/\varepsilon}) \right) \tag{10}$$

where $\lambda = \frac{\rho}{\rho+\varepsilon}$ and $K \in \mathbb{R}^{N \times M}$ defined as $[K]_{ij} = e^{-c(x_i,y_j)/\varepsilon}$ is the Gibbs kernel. This algorithm, unlike the original matrix scaling algorithm, is numerically stable due to computation in log domain. A desirable property of optimal transport problem is that gradients of $W_{\varepsilon,\rho}(\alpha,\beta)$ with respect to x_i, a_i is readily computable, prompting easy integration into existing gradient based methods for registration problems [FCVP17].

4. Approach

4.1. Registration as MAP Inference

Diffeomorphic registration with geodesic shooting can be treated as a maximum a posteriori problem of initial velocity fields v with a fixed Guassian process prior $v \sim \mathcal{GP}(0,k)$ and a likelihood specified by the data matching term $\mathcal{W}_{\varepsilon,\rho}$,

$$P(v \mid x) \propto \exp(-\frac{1}{2} x^T K(x,x)^{-1} x) \tag{11}$$

$$P(y \mid x, v) \propto \exp(-\mathcal{W}_{\varepsilon,\rho}(\varphi_\sharp \alpha, \beta)) \tag{12}$$

maximization of log posterior $\max_v \log P(v \mid x, y)$ recovers (6). Suppose we obtained the optimal plan π^*, the likelihood is an isotropic multivariate normal over a weighted combination of y centered at q_1 with variance determined by π^*.

$$\mathcal{W}_{\varepsilon,\rho}(\varphi_\sharp \alpha, \beta) = \sum_{i=1}^N \sum_{j=1}^M \left\| q_1^i - y^j \right\|_2^2 \pi^*_{ij} + \text{const} \tag{13}$$

$$= \sum_{i=1}^N \frac{1}{\sigma_i^2} \left\| q_1^i - \bar{y}^i \right\|_2^2 + \text{const} \tag{14}$$

$$= (\bar{y} - q_1)^T \Sigma^{-1} (\bar{y} - q_1) + \text{const} \tag{15}$$

where $\sigma_i^2 = 1/(\sum_j \pi^*_{ij})$, $\Sigma = I_D \otimes \mathrm{diag}(\sigma_1^2, \cdots, \sigma_N^2)$ are variances and $\bar{y}_i = \sum_j \sigma_i \pi^*_{ij} y^j$ is a weighted combination of y with weights summing to 1, which we group together as $\bar{y} = (\bar{y}^1, \cdots, \bar{y}^N) \in \mathbb{R}^{ND}$. Intuitively, $\bar{y}^i$ is matched to q_1^i by the transport plan, albeit at a synthetic location. The likelihood of observing $\bar{y}^i$ is large if it is close to q_1^i in the Euclidean sense and if the total mass coming from q_1^i, i.e. $\sum_j \pi^*_{ij}$, is small, indicating a larger uncertainty as to where transport plan might assign the point to. Of course, the behavior of the transport plan is affected by ε, ρ. Not entirely sure, but this might suggest inclusion of hyperparameters ε, ρ as optimziation variables of the likelihood model, akin to how we might want to optimize for the variance of an iid Gaussian noise model for registering images.

4.2. Variational Inference with Average Transport Cost

Let y be observed variables, initial velocity field v as latent variables. We are interested in the posterior density, $P(v \mid x, y) \propto P(v \mid x) P(y \mid v, x)$ where prior and likelihood is as defined in (12). We define variational distribution over v as a linear function (3) of a free-form initial momentum p_0,

$$Q(p_0) = \mathcal{N}(p_0 \mid \mu_p, \Sigma_p) \tag{16}$$

$$Q(v \mid x) = \mathcal{N}(v \mid K(x,x)\mu_p, K(x,x)\Sigma_p K(x,x)^T) \tag{17}$$

We parameterize the initial momentum instead of velocity because we do not need to invert the kernel matrix $K(x,x)$ whenever we want to generate random diffeomorphic maps at test time. Additionally, this approach works better in practice for the limited experiments we have tried. Here $\{\mu_p, \Sigma_d\}$ as well as hyperparameter for the kernel k is the set of variational parameters θ. Inference relegates to finding an approximate posterior $Q(v)$ that is close to the true posterior using optimization, $\min_\theta \mathrm{KL}(Q(v) \| P(v \mid x, y))$. As usually done in variational inference, we optimize the ELBO,

$$\min_\theta \mathrm{KL}(Q(v \mid x) \| P(v \mid x)) + \int \mathcal{W}_{\varepsilon,\rho}(\varphi_\sharp \alpha, \beta)\, dQ(p_0) \tag{18}$$

The first term is computable in $\mathcal{O}(N^3)$. The second term can be approximated via Monte Carlo integration, but will be computationally inefficient due to need to solve several discrete transport problems for each gradient update to θ. Alternatively, we optimize an upper bound, requiring one solve with a modified transport cost that marginalizes over p_0, $\widetilde{c} : \Omega^2 \to \mathbb{R}$ with,

$$\widetilde{c}(x,y) = \int c(q_1(x,p_0), y)\, dQ(p_0) \tag{19}$$

Notationally, $q_1(p_0,\alpha)$ emphasizes the fact that q_1 is a deterministic function of (p_0,α) as specified by the Geodesic flow (4). So,

$$\int \mathrm{OT}^c(\varphi_\sharp\alpha,\beta)\,dQ(p_0) \le \mathrm{OT}^{\widetilde{c}}(\alpha,\beta) \tag{20}$$

$$\text{with} \quad \mathrm{OT}^c(\alpha,\beta) = \min_{\pi\in\Pi(\alpha,\beta)} \iint_{\Omega^2} c(x,y)\,d\pi(x,y) \tag{21}$$

This relationship also holds for regularized unbalanced cost. It is interesting that we are essentially computing optimal transport on source and target shapes directly, and that the randomness in φ is incoporated into $\widetilde{c}$. Employing this upper bound for optimization of (18), we arrive at an objective that is computationally more tractable with the risk of derailing the optimization from good solutions.

$$\min_\theta \mathrm{KL}(Q(v\mid x)\|P(v\mid x)) + \mathcal{W}^{\widetilde{c}}_{\varepsilon,\rho}(\alpha,\beta) \tag{22}$$

In practice, we approximate cost matrix $\widetilde{C}$ with sampling. We use autodiff to compute the gradient through a fixed Sinkhorn's iterations, as [GPC18] did.

5. Results

Figures/tables illustrating the results of your work, as well as text interpreting these results. [PC20],

References

[ATY05] Allassonnière S., Trouvé A., Younes L.: Geodesic Shooting and Diffeomorphic Matching Via Textured Meshes. In *Energy Minimization Methods in Computer Vision and Pattern Recognition*, Hutchison D., Kanade T., Kittler J., Kleinberg J. M., Mattern F., Mitchell J. C., Naor M., Nierstrasz O., Pandu Rangan C., Steffen B., Sudan M., Terzopoulos D., Tygar D., Vardi M. Y., Weikum G., Rangarajan A., Vemuri B., Yuille A. L., (Eds.), vol. 3757. Springer Berlin Heidelberg, Berlin, Heidelberg, 2005, pp. 365–381. Series Title: Lecture Notes in Computer Science. URL: http://link.springer.com/10.1007/11585978_24, doi:10.1007/11585978_24. 1

[AWR18] Altschuler J., Weed J., Rigollet P.: Near-linear time approximation algorithms for optimal transport via Sinkhorn iteration. *arXiv:1705.09634 [cs, stat]* (Feb. 2018). arXiv: 1705.09634. URL: http://arxiv.org/abs/1705.09634. 2

[BMTY05] Beg M. F., Miller M. I., Trouvé A., Younes L.: Computing Large Deformation Metric Mappings via Geodesic Flows of Diffeomorphisms. *International Journal of Computer Vision 61*, 2 (Feb. 2005), 139–157. URL: https://doi.org/10.1023/B:VISI.0000043755.93987.aa, doi:10.1023/B:VISI.0000043755.93987.aa. 1

[CPSV17] Chizat L., Peyré G., Schmitzer B., Vialard F.-X.: Scaling Algorithms for Unbalanced Transport Problems. *arXiv:1607.05816 [math]* (May 2017). arXiv: 1607.05816. URL: http://arxiv.org/abs/1607.05816. 3

[Cut13] Cuturi M.: Sinkhorn Distances: Lightspeed Computation of Optimal Transportation Distances. *arXiv:1306.0895 [stat]* (June 2013). arXiv: 1306.0895. URL: http://arxiv.org/abs/1306.0895. 2

[DBGS19] Dalca A. V., Balakrishnan G., Guttag J., Sabuncu M. R.: Unsupervised Learning of Probabilistic Diffeomorphic Registration for Images and Surfaces. *Medical Image Analysis 57* (Oct. 2019), 226–236. arXiv: 1903.03545. URL: http://arxiv.org/abs/1903.03545, doi:10.1016/j.media.2019.07.006. 2

[dGBOD12] de Goes F., Breeden K., Ostromoukhov V., Desbrun M.: Blue noise through optimal transport. *ACM Transactions on Graphics 31*, 6 (Nov. 2012), 1–11. URL: https://dl.acm.org/doi/10.1145/2366145.2366190, doi:10.1145/2366145.2366190. 2

[dGCSAD11] de Goes F., Cohen-Steiner D., Alliez P., Desbrun M.: An Optimal Transport Approach to Robust Reconstruction and Simplification of 2D Shapes. *Computer Graphics Forum 30*, 5 (Aug. 2011), 1593–1602. URL: http://doi.wiley.com/10.1111/j.1467-8659.2011.02033.x, doi:10.1111/j.1467-8659.2011.02033.x. 2

[FCVP17] Feydy J., Charlier B., Vialard F.-X., Peyré G.: Optimal Transport for Diffeomorphic Registration. *arXiv:1706.05218 [math]* (June 2017). arXiv: 1706.05218. URL: http://arxiv.org/abs/1706.05218. 1, 2, 3

[FRTG19] Feydy J., Roussillon P., Trouvé A., Gori P.: Fast and Scalable Optimal Transport for Brain Tractograms. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*, Shen D., Liu T., Peters T. M., Staib L. H., Essert C., Zhou S., Yap P.-T., Khan A., (Eds.), vol. 11766. Springer International Publishing, Cham, 2019, pp. 636–644. Series Title: Lecture Notes in Computer Science. URL: http://link.springer.com/10.1007/978-3-030-32248-9_71, doi:10.1007/978-3-030-32248-9_71. 1, 2

[FSV*18] Feydy J., Séjourné T., Vialard F.-X., Amari S.-i., Trouvé A., Peyré G.: Interpolating between Optimal Transport and MMD using Sinkhorn Divergences. *arXiv:1810.08278 [math, stat]* (Oct. 2018). arXiv: 1810.08278. URL: http://arxiv.org/abs/1810.08278. 3

[Fv09] Figueiras-vidal A. R.: Inter-domain Gaussian processes for sparse inference using inducing features. In *in Advances in Neural Information Processing Systems* (2009), pp. 1087–1095. 1

[FZM*15] FROGNER C., ZHANG C., MOBAHI H., ARAYA-POLO M., POGGIO T.: Learning with a Wasserstein Loss. *arXiv:1506.05439 [cs, stat]* (Dec. 2015). arXiv: 1506.05439. URL: http://arxiv.org/abs/1506.05439. 2

[GPC18] GENEVAY A., PEYRÉ G., CUTURI M.: Learning Generative Models with Sinkhorn Divergences. 10. 2, 4

[JCG20] JANATI H., CUTURI M., GRAMFORT A.: Spatio-Temporal Alignments: Optimal transport through space and time. 9. 2

[JM00] JOSHI S., MILLER M.: Landmark matching via large deformation diffeomorphisms. *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society 9* (Feb. 2000), 1357–70. doi:10.1109/83.855431. 1, 2

[MTY06] MILLER M. I., TROUVÉ A., YOUNES L.: Geodesic Shooting for Computational Anatomy. *Journal of mathematical imaging and vision 24*, 2 (Jan. 2006), 209–228. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2897162/, doi:10.1007/s10851-005-3624-0. 1, 2

[PC20] PEYRÉ G., CUTURI M.: Computational Optimal Transport. *arXiv:1803.00567 [stat]* (Mar. 2020). arXiv: 1803.00567. URL: http://arxiv.org/abs/1803.00567. 4

[RJN*13] RISHOLM P., JANOOS F., NORTON I., GOLBY A. J., WELLS W. M.: Bayesian Characterization of Uncertainty in Intra-Subject Non-Rigid Registration. *Medical image analysis 17*, 5 (July 2013), 538–555. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3687087/, doi:10.1016/j.media.2013.03.002. 2

[Sch19] SCHMITZER B.: Stabilized Sparse Scaling Algorithms for Entropy Regularized Transport Problems. *arXiv:1610.06519 [cs, math]* (Feb. 2019). arXiv: 1610.06519. URL: http://arxiv.org/abs/1610.06519. 2

[SdGP*15] SOLOMON J., DE GOES F., PEYRÉ G., CUTURI M., BUTSCHER A., NGUYEN A., DU T., GUIBAS L.: Convolutional wasserstein distances: efficient optimal transportation on geometric domains. *ACM Transactions on Graphics 34*, 4 (July 2015), 66:1–66:11. URL: https://doi.org/10.1145/2766963, doi:10.1145/2766963. 2

[VMYD04] VAILLANT M., MILLER M. I., YOUNES A. L., D A. T.: Statistics on diffeomorphisms via tangent space representations. *NeuroImage 23* (2004), 2004. 1

[VRRC12] VIALARD F.-X., RISSER L., RUECKERT D., COTTER C. J.: Diffeomorphic 3D Image Registration via Geodesic Shooting Using an Efficient Adjoint Calculation. *International Journal of Computer Vision 97*, 2 (Apr. 2012), 229–241. URL: https://doi.org/10.1007/s11263-011-0481-8, doi:10.1007/s11263-011-0481-8. 1

[WTNWI14] WASSERMANN D., TOEWS M., NIETHAMMER M., WELLS III W.: Probabilistic Diffeomorphic Registration: Representing Uncertainty. *arXiv:1701.03266 [cs] 8545* (2014), 72–82. arXiv: 1701.03266. URL: http://arxiv.org/abs/1701.03266, doi:10.1007/978-3-319-08554-8_8. 2

[WWGZ19] WANG J., WELLS W. M., GOLLAND P., ZHANG M.: Registration Uncertainty Quantification Via Low-dimensional Characterization of Geometric Deformations. *Magnetic resonance imaging 64* (Dec. 2019), 122–131. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7069236/, doi:10.1016/j.mri.2019.05.034. 2

[ZSF13] ZHANG M., SINGH N., FLETCHER P. T.: Bayesian Estimation of Regularization and Atlas Building in Diffeomorphic Image Registration. In *Information Processing in Medical Imaging*, Hutchison D., Kanade T., Kittler J., Kleinberg J. M., Mattern F., Mitchell J. C., Naor M., Nierstrasz O., Pandu Rangan C., Steffen B., Sudan M., Terzopoulos D., Tygar D., Vardi M. Y., Weikum G., Gee J. C., Joshi S., Pohl K. M., Wells W. M., Zöllei L., (Eds.), vol. 7917. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013, pp. 37–48. Series Title: Lecture Notes in Computer Science. URL: http://link.springer.com/10.1007/978-3-642-38868-2_4, doi:10.1007/978-3-642-38868-2_4. 2