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

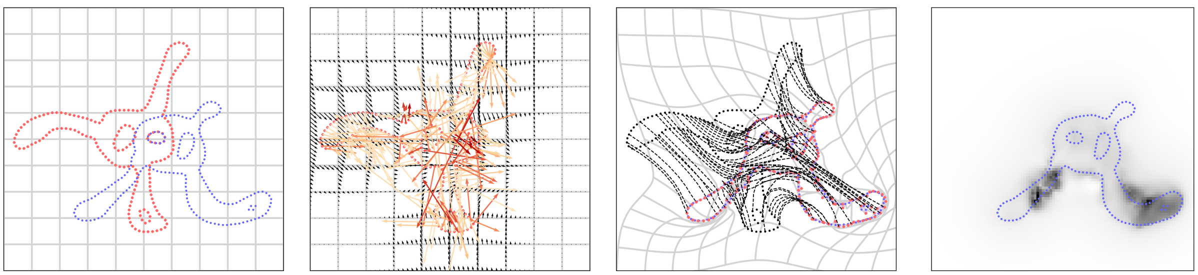

Figure 1: *The red amoeba is registered to blue amoeba. A valid diffeomorphic transformation can be generated from per-particle momenta (orange arrow) by solving a set of Geodesic equations (black dashed line). On the right, we can visualize variance of the effect of random transformation on shape as 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 Wasserstein 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]. We relie on geodesic shooting methods in our formulation.

2.2. Registration Uncertainty

3. Technical Approach

3.1. Diffeomorphic Registration

Let $\Omega \subset \mathbb{R}^D$ be a low dimensional ambient space. Let $x := (x^1, \cdots, x^n) \subset \mathbb{R}^{N\times D}$, $y := (y^1, \cdots, y^n) \subset \Omega^{N\times D}$ be source and target points respectively. The goal of diffeomorphic registration of point sets is to transform x via a diffeomorphic mapping φ s.t. $\varphi(x)$ is close to y according to some similarity metric $\mathcal{L}(\varphi(x), y)$.

Let the space of smooth velocity fields V as a rkhs over Ω characterized by kernel $\bar{k} : \Omega\times\Omega \to \mathbb{R}^{D\times D}$ satisfying vector-valued reproducing property $\langle \bar{k}(x,\cdot)y, v\rangle_V = \langle v(x), y\rangle_{\mathbb{R}^D}$ for all $y \in \mathbb{R}^D, v \in V$. For our purposes, we consider an equivalent scalar-valued rkhs with kernel $k : \Omega\times[D]\times\Omega\times[D] \to \mathbb{R}$ where $\bar{k}(x,x')_{dd'} = k((x,d),(x',d'))$. A diffeomorphism can be constructed via flows , i.e. solutions to an ODE problem φ_t, $\dot{\varphi}_t = v_t \circ \varphi_t, \varphi_0 = \mathrm{Id}$, 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 x to y.

$$\min_{v_t:t\in[0,1]} \frac{1}{2}\int_0^1 \|v_t\|_V^2 \, dt + \mathcal{L}(\varphi(x), y) \tag{1}$$

3.2. 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,

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

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, $\int_0^1 \mathcal{H}(q_t,p_t)\,dt = \mathcal{H}(q_0,p_0) = \frac{1}{2}\langle p_0, K(q_0,q_0)p_0\rangle$. We arrive at a registration problem where we optimize over the initial *shooting momentum* p_0,

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

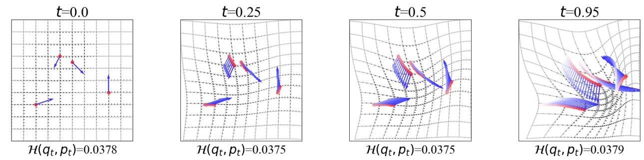

Figure 2: *Geodesic shooting with an Euler integrator with time step of $\delta t = .1$. The velocity field is represented using a radial basis kernel with $\sigma = .25$. We show trajectory of q_t (red dots) with momentum p_t (blue arrow) and interpolated velocity fields at grid points (black). We see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

where $\varphi = q_1 \in \mathbb{R}^{ND}$. Optimization involves a forward integration of (q_t, p_t) via (4) to get transformed particles q_1, compute the gradient of objective (6) with respect to initial momentum, and do gradient update iteratively.

3.3. Monge's and Kantorovich's formulation

Given $a, b \in \triangle^{n-1}$, where $\triangle^{n-1}$ is unit simplex. $\alpha = \sum_{i=1}^n a_i\delta_{x_i}, \beta = \sum_{j=1}^m b_j\delta_{y_i}$ are discrete measures. The optimal transport problem tries to find a map that associate each point x_i to a single point y_j such that masses are preserved. This corresponds to classical Monge's formulation of optimal transport

$$\min_{T:\mathcal{X}\to\mathcal{Y}:T_\#\alpha=\beta} \sum_{i=1}^n c(x_i, T(x_i)) \tag{7}$$

where $c : \mathcal{X}\times\mathcal{Y} \to \mathbb{R}$ is cost defined over support of measures, and that the constraints is simply that T is constrained to be a push-forward from α to β. The problem with Monge's formulation is that the problem is combinatorial and nonconvex, so hard to solve. Kantorovich relax the deterministic nature of transport map, allowing mass at each source point split and be dispatched to multiple target points. This information is encoded in $P \in \mathbb{R}_+^{n\times m}$, where P_{ij} describes amount of mass flowing from x_i to y_j. The Kantorovich formulation of optimal transport for discrete measures is then

$$\min_{P\in\mathbb{R}_+^{n\times m}:P1_m=a,\,P^T1_n=b} \langle C, P\rangle \tag{8}$$

where the constraints specifies the set of admissible transport map to be a coupling of marginals α, β and $C \in \mathbb{R}^{n\times m}$ is the cost matrix, i.e. $C_{ij} = c(x_i, y_j)$. The optimization problem is a linear program and hence can be easily solved using simplex algorithm. In addition, we can instead solve the dual problem, and because of zero duality gap, equivalently solves the primal problem,

$$\max_{f,g\in\mathbb{R}^n\times\mathbb{R}^m:f_i+g_j\le C_{ij}} \langle f, a\rangle + \langle g, b\rangle \tag{9}$$

where f, g are called dual potential.

4. Entropic Regularization

Regularizing the original optimal transport problem brings computational and statistical benefits. In particular, the optimization problem can now be solved with fast matrix scaling algorithms that scales with strength of regularization. In addition, the sample efficiency for regularized problem is also superior. The entropy and KL divergence of coupling between two discrete measure is given by

$$H(P) = -\sum_{i,j} P_{ij}(\log(P_{ij}) - 1) = -\langle P, \log P\rangle + 1^T P 1 \quad (10)$$

$$\mathrm{KL}(P\|K) = \sum_{i,j} P_{ij}\log\frac{P_{ij}}{K_{ij}} - P_{ij} + K_{ij} = \langle P, \log(P \oslash K)\rangle + 1^T(K-P)1 \quad (11)$$

where taking logarithm and subtraction are elementwise operations. The entropic regularized problem

$$\min_{P\in\mathbb{R}_+^{n\times m}: P1_m = a, P^T 1_n = b} \langle P, C\rangle - \varepsilon H(P) \quad (12)$$

We can interpret primal objective as the information projection of the Gibbs kernel $K \in \mathbb{R}^{n\times m}$ where $K_{ij} = e^{-\frac{C_{ij}}{\varepsilon}}$ onto the admissible couplings $U(a,b) = \left\{P \in \mathbb{R}_+^{n\times m} \mid P1_m = a,\, P^T 1_n = b\right\}$.

$$\langle P, C\rangle + \varepsilon\langle P, \log P\rangle - \varepsilon 1^T P 1 = \varepsilon\left\langle P, \log\left(P \oslash e^{-C/\varepsilon}\right)\right\rangle - \varepsilon 1^T P 1 + \varepsilon 1^T e^{-C/\varepsilon} 1 = \varepsilon \mathrm{KL}(P\|K) \quad (13)$$

4.1. Sinkhorn's Algorithm

We can solve regularized optimal transport problem with Sinkhorn algorithm [Cut13]. The basic idea is to write the 1st order optimality condition for the primal variables for the Lagrangian,

$$\mathcal{L}(P,f,g) = \langle P, C\rangle - \varepsilon H(P) - \langle f, P1_m - a\rangle - \left\langle g, P^T 1_n - b\right\rangle \quad (14)$$

$$\partial\mathcal{L}(P,f,g)/\partial P_{ij} = C_{ij} + \varepsilon\log(P_{ij}) - f_i - g_j = 0 \quad \Rightarrow \quad P_{ij} = e^{(-C_{ij}+f_i+g_j)/\varepsilon} \quad (15)$$

$$u^{(\ell+1)} \leftarrow a \oslash (Kv^{(\ell)}) \qquad v^{(\ell+1)} \leftarrow b \oslash (K^T u^{(\ell+1)}) \quad (16)$$

Each iteration of the algorithm uses $\mathcal{O}(nm)$ computation for matrix-vector products and can be accelerated to $\mathcal{O}(n\log n)$ using convolution if suppport is over gridded space.

4.2. Log-domain Stabilization

Sinkhorn's iteration can be numerically unstable as $\varepsilon \to 0$ due zeros in K, resulting in null values during division by zero. One solution is to perform computation of u, v in the log-domain [CPSV17, Sch19]. This is equivalent to block coordinate ascent on the dual of entropic regularized optimal transport problem (12),

$$\max_{f,g\in\mathbb{R}^{n\times m}} \langle f, a\rangle + \langle g, b\rangle - \varepsilon\left\langle e^{f/\varepsilon}, Ke^{g/\varepsilon}\right\rangle \quad (17)$$

Then $0 = \partial L_{\text{dual}}(f,g)/\partial f = a - e^{f/\varepsilon} \odot (Ke^{g/\varepsilon})$ implies $f = \varepsilon\log(a) - \varepsilon\log(Ke^{g/\varepsilon})$ and similarly for g. So,

$$f^{(\ell+1)} \leftarrow \varepsilon\log(a) - \varepsilon\log(Ke^{g^{(\ell)}/\varepsilon}) \qquad g^{(\ell+1)} \leftarrow \varepsilon\log(b) - \varepsilon\log(Ke^{f^{(\ell+1)}/\varepsilon}) \quad (18)$$

This is equivalent to Sinkhorn's iteration in log domain. Additionally, we can define a numerically stable softmin operator based on log-sum-exp, $\mathrm{softmin}_\varepsilon(z) = -\varepsilon\log\sum_i e^{-z_i/\varepsilon} = -\varepsilon\mathrm{LSE}(-z/\varepsilon)$. Note $f_i = \varepsilon\log a_i - \varepsilon\log\sum_j e^{-(C_{ij}-g_j)/\varepsilon} = \mathrm{softmin}_\varepsilon(C_{ij} - f_i - g_j) + f_i + \varepsilon\log a_i$. Therefore, (18) can be equivalently written as,

$$f^{(\ell+1)} \leftarrow \mathrm{softmin}_\varepsilon(S(f^{(\ell)}, g^{(\ell)})) + f^{(\ell)} + \varepsilon\log(a) \quad (19)$$

$$g^{(\ell+1)} \leftarrow \mathrm{softmin}_\varepsilon(S(f^{(\ell+1)}, g^{(\ell)})) + g^{(\ell)} + \varepsilon\log(b) \quad (20)$$

where $S(f,g)_{ij} = C_{ij} - f_i - g_j$.

4.3. Minimum Kantorovich Estimator

5. Unbalanced Transport

Previously, optimal transprot problem requires α, β to have equal mass. This becomes a problem in applications when either the measures are noisy or when preservation of mass is not desirable. We can relax hard marginal constraints with soft penalty that measures deviation from exact coupling using some divergence, e.g. KL-divergence [CPSV17],

$$\min_{P\in\mathbb{R}_+^{n\times m}} \langle P, C\rangle - \varepsilon H(P) + \rho\mathrm{KL}(P1_m\|a) + \rho\mathrm{KL}(P^T 1_m\|b) \quad (21)$$

Similar to balanced case, we can show that optimal coupling is again a scaling of K, i.e. $P = \mathrm{diag}(u)K\mathrm{diag}(v)$ where $u = (a \oslash (Kv))^\lambda$ and $v = (b \oslash (K^T u))^\lambda$ where $\lambda = \frac{\rho}{\rho+\varepsilon}$ [FZM*15]. We can derive a similar Sinkhorn iteration in the log domain for unbalanced case.

6. 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

[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. 3

[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

[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

[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. 3

[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. 3

[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. 3

[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