

SULCAL PATTERN MATCHING WITH THE WASSERSTEIN DISTANCE

Zijian Chen, Soumya Das, Moo K. Chung

University of Wisconsin, Madison, USA
`{zijian.chen, mkchung}@wisc.edu`

ABSTRACT

We present the unified computational framework for modeling the sulcal patterns of human brain obtained from the magnetic resonance images. The Wasserstein distance is used to align the sulcal patterns nonlinearly. These patterns are topologically different across subjects making the pattern matching a challenge. We work out the mathematical details and develop the gradient descent algorithms for estimating the deformation field. We further quantify the image registration performance. This method is applied in identifying the differences between male and female sulcal patterns.

1. INTRODUCTION

The concave regions in the highly convoluted cerebral cortex of the human brain are referred to as the *sulci* (Fig. 1). These complex tree-shaped sulcal curves are highly variable in length, area, depth, curvature and topology across different subjects [1]. There have been extensive studies that connect the variabilities of such biomarkers with the differences in cognitive or pathological characteristics between populations [2]. However, since each subject have different topological patterns, it is difficult to match the sulcal patterns across subjects [3]. One approach of reducing the difficulty of matching is to smooth the sulcal patterns and then match the smoothed patterns. Such smoothed representations enable pattern matching in a continuous probabilistic fashion. Thus, we propose to use the Wasserstein distance, which minimizes the optimal transport cost of moving between probability distributions.

The Wasserstein distances have been previously applied in various imaging applications. [4] computed the Wasserstein distance using the hyperbolic metric for cortical brain morphometry. [5] proposed the sliced Wasserstein distance in speeding up the computation in pattern recognition. [6] derived a cycle-consistent generative adversarial network based on the optimal transport for MRI reconstruction. Despite the advance of various algorithms for Wasserstein distance based pattern matching [7, 5], the rational of involving smoothing in such algorithms is unclear. In this paper, we work out the mathematical and implementation details on the Wasserstein

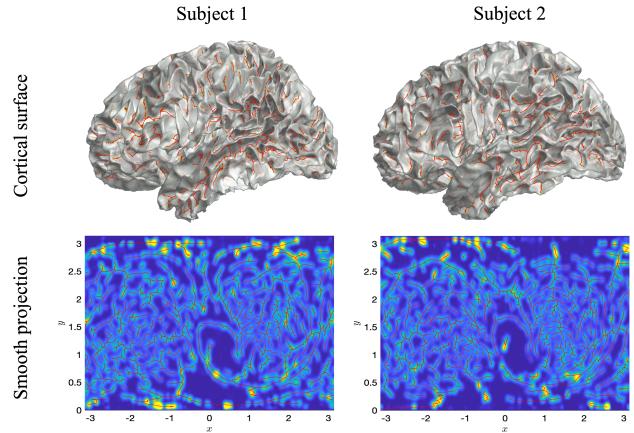


Fig. 1. Top: sulcal curves (red) are displayed on top of cortical surfaces. Bottom: sulcal curves are projected and smoothed in the rectangular domain $[-\pi, \pi] \times [0, \pi] \subset \mathbb{R}^2$.

distance on heat kernel smoothing, the kernel version of diffusion. In the experiment, we demonstrated the Wasserstein distance based registration reduces the variability of sulcal patterns across subjects.

2. METHODS

2.1. Sulcal pattern data

We used the processed T1-weighted MRI of 456 subjects (age-matched 274 females and 182 males) in the Human Connectome Project (HCP) [8]. The MRI were obtained using a Siemens 3T Connectome Skyra scanner with a 32-channel head coil [9, 10]. The MRI were registered to the MNI space with a FLIRT affine and FNIRT nonlinear registration [11]. FreeSurfer's recon-all pipeline was used on the distortion- and bias-corrected MRI [12] that includes the segmentation of white matter and pial surfaces as well as FreeSurfer's folding-based surface registration to the template. TRACE algorithm was used for automatic sulcal curve extraction from surface meshes [3, 13] (Fig. 1) and then projected on the unit sphere, which is further parameterized by spherical angles $p = (x, y) \in [-\pi, \pi] \times [0, \pi] \subset \mathbb{R}^2$. Since the data is spherical, it is periodic over x . Fig. 1 displays the sulcal curves of two different

The study is funded by NIH EB022856, EB02875, NSF MDS-2010778.

subjects on \mathbb{R}^2 , where sulcal curves are assigned value 1 (colored red) while all other parts of the brain are assigned value 0.

2.2. Heat kernel smoothing of sulcal patterns

The sulcal pattern $f(p)$ is smoothed with heat kernel to reduce high frequency noise. The smoothing increases the signal-to-noise ratio (SNR) and increases the statistical power in the population study [3]. We model the sulcal pattern as

$$f(p) = \mu(p) + \varepsilon(p), \quad p = (x, y) \in [-\pi, \pi] \times [0, \pi] \subset \mathbb{R}^2,$$

where $\mu(p)$ is the underlying true signal and $\varepsilon(p)$ is the noise. For observed data, we assign value 1 to the sulcal curves and 0 otherwise (Fig. 1). We estimate $\mu(p)$ by smoothing with isotropic heat kernel. The smoothed estimate $\hat{\mu}$ is given by the solution of the isotropic diffusion

$$\frac{\partial}{\partial \sigma} u(x, y, \sigma) = \frac{\partial^2}{\partial^2 x} u(x, y, \sigma) + \frac{\partial^2}{\partial^2 y} u(x, y, \sigma)$$

with the initial condition $u(x, y, \sigma = 0) = f(x, y)$ and the boundary conditions

$$\begin{aligned} u(-\pi, y, \sigma) &= u(\pi, y, \sigma), \\ u'(-\pi, y, \sigma) &= u'(\pi, y, \sigma), \\ u(x, 0, \sigma) &= u(x, \pi, \sigma) = 0. \end{aligned}$$

The amount of smoothing σ is determined empirically. The solution is given as the weighted Fourier series [14]:

$$\hat{\mu}(x, y) = \sum_{j=0}^m \sum_{k=0}^n e^{-\lambda_{jk}\sigma} \left[A_{jk} \phi_{jk}^1 + B_{jk} \phi_{jk}^2 \right]. \quad (1)$$

The eigenvalues $\lambda_{jk} = j^2 + k^2$ and the eigenfunctions corresponding to Laplacian $\frac{\partial^2}{\partial^2 x} + \frac{\partial^2}{\partial^2 y}$ are

$$\phi_{jk}^1 = \frac{2 \cos(jx) \sin(ky)}{\pi^2(1 + \delta_{j0})}, \quad \phi_{jk}^2 = \frac{2 \sin(jx) \sin(ky)}{\pi^2},$$

where $\delta_{j0} = 1$ if $j = 0$ and 0 otherwise. The coefficients A_{jk} and B_{jk} are estimated in the least square fashion using the initial data $f(p)$ [14]. The smoothing results are shown in Fig. 1 with diffusion time $\sigma = 0.001$.

The smooth representation enables alignment of different sulcal patterns on a common grid. For each subject, we first obtain the Fourier coefficients A_{jk} and B_{jk} via smoothing. The value on the grid point (x_i, y_j) is then given by $\hat{\mu}(x_i, y_j)$ in the expansion (1).

2.3. Wasserstein distance on heat kernel smoothing

Let f_1 and f_2 be the empirical distribution on the two sets of scatter points that define the vertices of sulcal curves $\mathbf{P}_1 = \{p_1, \dots, p_n\}$ and $\mathbf{P}_2 = \{q_1, \dots, q_n\}$:

$$f_1(p) = \frac{1}{n} \sum_{i=1}^n \delta(p - p_i), \quad f_2(q) = \frac{1}{n} \sum_{i=1}^n \delta(q - q_i). \quad (2)$$

After algebraic derivations involving Choquet's and Birkhoff's theorems [15], the 2-Wasserstein distance D_W between the empirical distributions is given as the Monge formulation

$$D_W(f_1, f_2) = \inf_{\tau \in S_n} \left(\frac{1}{n} \sum_{i=1}^n \|p_i - p_{\tau(i)}\|^2 \right)^{1/2}, \quad (3)$$

where S_n is the permutation group of order n .

The diffused sulcal pattern can be written as the kernel convolution on initial condition $f(p)$ as $\hat{\mu}(p) = K_\sigma * f(p)$. with heat kernel K_σ [14]. In \mathbb{R}^2 , the heat kernel is simply Gaussian kernel

$$K_\sigma(p, p_i) = \frac{1}{2\pi\sigma^2} e^{-\frac{1}{2\sigma^2}(p-p_i)^\top(p-p_i)}.$$

The heat kernel smoothing on sulcal patterns \mathbf{P}_1 and \mathbf{P}_2 is given by (Fig. 1 Bottom)

$$\tilde{f}_1(p) = \frac{1}{n} \sum_{i=1}^n K_\sigma(p, p_i), \quad \tilde{f}_2(q) = \frac{1}{n} \sum_{i=1}^n K_\sigma(q, q_i). \quad (4)$$

Let $\mathcal{A}(\tilde{f}_1, \tilde{f}_2)$ be the set of all possible joint density functions corresponding to \tilde{f}_1 and \tilde{f}_2 . The Wasserstein distance between diffused sulcal pattern can be written as [16]

$$D_W(\tilde{f}_1, \tilde{f}_2) = \inf_{\pi \in \mathcal{A}(\tilde{f}_1, \tilde{f}_2)} \left(\int \|p - q\|^2 \pi(p, q) dp dq \right)^{1/2}. \quad (5)$$

If we restrict the joint distribution to be a linear combination of multivariate normals, we can compute the expression (5) exactly. We will denote the Wasserstein distance with such a restriction as $D_{W'}$. Then we have the following equivalence.

Theorem 1 For heat kernel smoothing \tilde{f}_1 and \tilde{f}_2 , we have

$$D_{W'}(\tilde{f}_1, \tilde{f}_2) = D_W(f_1, f_2). \quad (6)$$

Proof. We provide the sketch of proof. The joint density π in (5) can be expressed as the mixture of multivariate normals as $\pi(p, q) = \sum_{i,j=1}^n \pi_{ij} \phi_{ij}(p, q)$, where the mixing proportions π_{ij} are doubly stochastic and satisfies $\sum_i \pi_{ij} = \sum_j \pi_{ij} = 1$. The multivariate normals ϕ_{ij} follow

$$\phi_{ij} \sim N \left(\begin{bmatrix} p_i \\ q_j \end{bmatrix}, \begin{bmatrix} \sigma^2 I_2 & H \\ H^\top & \sigma^2 I_2 \end{bmatrix} \right),$$

where $p_i \in \mathbf{P}_1$, $q_j \in \mathbf{P}_2$, I_2 is the identity matrix and H is the 2×2 positive semi-definite covariance matrix. Then after lengthy derivations, we can obtain

$$D_{W'}(\tilde{f}_1, \tilde{f}_2)^2 = \inf_{\pi_{ij}, H} \sum_{i,j=1}^n \pi_{ij} \|p_i - q_j\|^2 + 2 \text{tr}(\sigma^2 I_2 - H).$$

Subsequently, it suffices to minimize each term in $D_{W'}$ separately. Minimizing the first term leads to the same result as (3) while the second term results in zero [16]. \square

Since $D_W(\tilde{f}_1, \tilde{f}_2) \leq D_{W'}(\tilde{f}_1, \tilde{f}_2)$, we have

$$D_W(\tilde{f}_1, \tilde{f}_2) \leq D_W(f_1, f_2)$$

indicating that heat kernel smoothing reduces the Wasserstein distance between the sulcal patterns and their variabilities.

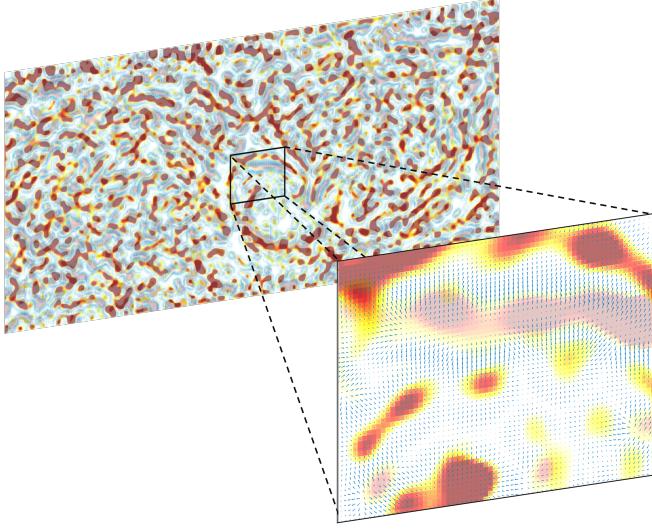


Fig. 2. Subject 1 (lower transparency) superimposed on top of subject 2 (higher transparency). The arrows are the displacement field of matching subject 1 to subject 2.

2.4. Gradient descent on the dual formulation

Given two smooth patterns \tilde{f}_1 and \tilde{f}_2 , the computation of the Wasserstein distance involves mapping a point p from the first pattern to the corresponding point $U(p)$ in the second pattern (Fig. 2). The deformation $p \rightarrow U(p)$ is required to have the smallest possible total displacement and satisfies

$$\int_{U^{-1}(A)} \tilde{f}_1(p) dp = \int_A \tilde{f}_2(q) dq, \quad \forall A \subset \mathbb{R}^2.$$

This leads to solving the optimization problem

$$\inf_U \int \|U(p) - p\|^2 \tilde{f}_1(p) dp, \quad (7)$$

which can be further reduced to minimizing

$$L(\varphi, \psi) = \int \varphi(p) \tilde{f}_1(p) dp + \int \psi(q) \tilde{f}_2(q) dq \quad (8)$$

with constraint $\varphi(p) + \psi(q) \geq p^\top q$ [7].

Theorem 2 [17] *The unique solution (φ, ψ) to the problem (8) with constraint $\varphi(p) + \psi(q) \geq p^\top q$ are convex conjugates, i.e., $\psi = \varphi^*$ and $\varphi = \psi^*$ where φ^* and ψ^* are defined as*

$$\begin{cases} \varphi^*(q) = \max_p (p^\top q - \varphi(p)) \\ \psi^*(p) = \max_q (q^\top p - \psi(q)) \end{cases}.$$

The solution to (7) is then given by $U = \nabla \varphi$.

The solution is given implicitly. However, the derivative of (8) can be explicitly found.

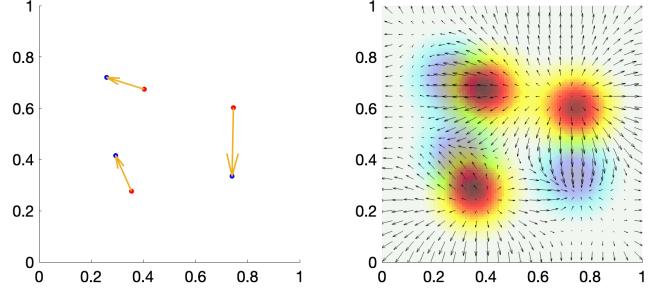


Fig. 3. Left: the Hungarian Algorithm is used to match two sets of three points (colored red and blue). Right: the gradient descent is used to match the heat kernel smoothing on these points. Arrows are the estimated displacement $U(p) - p$.

Theorem 3 [7] *Suppose φ is the convex conjugate of ψ for some ψ and $\nabla^2 \varphi$ is Lipschitz. Then it follows L is convex, Lipschitz and*

$$L'(\varphi) = \tilde{f}_1 - (\tilde{f}_2 \circ \nabla \varphi) \det(H_\varphi),$$

where H_φ is the Hessian matrix of φ .

Subsequently, we solve for gradient descent on φ as [7]:

$$\varphi_{n+1} = \varphi_n - L'(\varphi_n), \quad (9)$$

The overall computational complexity for the algorithm is $O(N)$ [18], which is more scalable than the Hungarian algorithm with $O(N^3)$ [19] and fluid mechanics with $O(N^{3/2})$ [20]. In this study, 800×400 rectangular grid with $N = 320000$ vertices were used. The grid values $\hat{\mu}(x_i, y_j)$ serve as the input for the gradient descent. We display the result of the gradient descent on estimating the displacement field $\nabla \varphi(p) - p$ as arrows in Fig. 2. Due to the convexity of φ , the algorithm is expected to recover the global optimum (Remark 3.3 in [7]).

3. EXPERIMENTS

3.1. Validation against the Hungarian Algorithm

We matched the two sets of n random generated scatter points in $[0, 1]^2$. Fig. 3 displays the result of one realization with $n = 3$. The Wasserstein distance $D_W(f_1, f_2)$ can be computed exactly through (3) using the Hungarian Algorithm [19]. We then applied heat kernel kernel smoothing with bandwidth $\sigma = 0.01$ on the scatter points and compute the Wasserstein distance $D'_W(\tilde{f}_1, \tilde{f}_2)$ using the gradient descent (9). The average percentage reduction of distance for 100 independent simulations for $n = 2, 3, 4$ are 0.86 ± 0.05 , 0.84 ± 0.06 and 0.77 ± 0.04 respectively. This is consistent with our theoretical result in section 2.3. The computer code for performing Wasserstein distance based matching is provided in <https://github.com/laplacebeltrami/sulcaltree>.

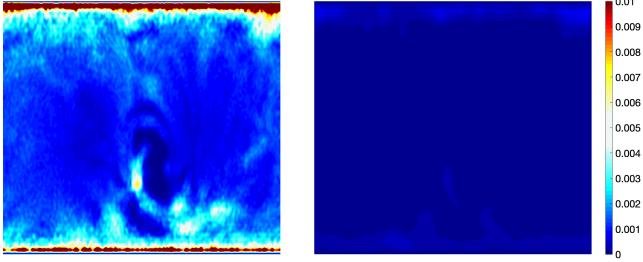


Fig. 4. The intersubject variability of the smoothed original sulcal patterns (left) and deformed patterns (right).

3.2. Reduction of image registration variability

The sulcal pattern has been already nonlinearly aligned using FreeSurfer’s standard folding-based surface registration to the surface atlas [12]. Wasserstein distance based matching significantly reduces the sulcal pattern variability further from the FreeSurfer alignment. We first compute the average of smoothed sulcal pattern maps $T = \frac{1}{n} \sum_{i=1}^n \tilde{f}_i(p)$ on smoothed estimates (1). This serves as the reference template pattern for subsequent registration. We then align each subject to the average pattern T using gradient descent (9). We initialize φ with $\frac{1}{2} \|p\|^2$ so that initial mapping $\nabla \varphi$ becomes the identity. Such initialization is data-free and has demonstrated efficacy in various applications including facial [18] and brain images [7]. We set as the maximum number of iterations at 200. The average runtime for one subject is approximately 8 seconds on a desktop computer.

For deformation $U_i(x)$ for subject i , the deformed sulcal patterns are obtained as $\tilde{f}_i^T(p) = \tilde{f}_i(U_i(p))$. The statistical variability before registration is computed as the sample variance of sulcal patterns $\tilde{f}_i(p)$ at each point across subjects (Fig. 4). The variability is significantly reduced after deformation. The mean variability over all the points in the brain is 0.0021 for the smoothed original pattern while it is 7.8×10^{-5} , the significant reduction of variability by 96.29%.

3.3. Sexual dimorphism

The method is subsequently used to determine the sulcal pattern differences between 240 females and 172 males. The two sample t -statistics were constructed on the deformed smoothed sulcal patterns $\tilde{f}_i^T(p) = \tilde{f}_i(U_i(p))$ (Fig. 5). The threshold t -statistic value greater than 4.12 or smaller than -2.84 correspond to the p -value of 0.05 after multiple comparisons correction through the permutation test (half millions) for the deformed pattern [21]. We obtained numerous significant differences all over the brain regions including the temporal lobe, which is responsible for sensory processing and known for sex difference [3]. The negative t -statistics in most of the brain regions indicates the presence of less sulci for females. This seems to be related to findings in [22], where females

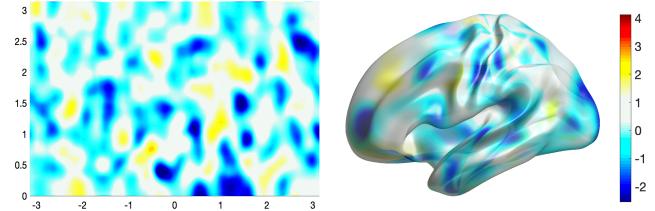


Fig. 5. t -statistics map of the localized sulcal pattern differences (female-male). Values are thresholded in $[-2.84, 4.12]$ corresponding to the corrected p -value of 0.05. Most of female brain regions show less sulci, which corresponds to more gray matter and thicker cortical thickness observed in other studies.

have larger gray matter volumes in a number of areas including the left temporal gyrus. In [23], cortical thickening is found extensively in the left superior parietal gyrus and post-central gyrus, which is consistent with less sulci in our study.

4. DISCUSSION

In this paper, we presented the new framework of matching sulcal patterns of human brain across subjects. We provided the theoretical justification for performing heat kernel smoothing before computing the Wasserstein distance. Smoothing reduces the Wasserstein distance between the sulcal patterns and spatial pattern variabilities. It is also possible to further refine the registration performance via the multi-resolution on heat kernel smoothing [7] or iteratively using the average of the deformed sulcal patterns $\frac{1}{n} \sum_{i=1}^n \tilde{f}_i^T(p)$ as the template in the next iteration. These are left as future studies.

The FreeSurfer uses the irregular triangle meshes while our smoothing and the gradient descent uses the regular pixel grid. This may introduce potential interpolation artifacts. A better approach would be to perform smoothing and the gradient descent on triangle meshes. This is left as a future study.

5. ACKNOWLEDGEMENT

We would like to thank Soheil Kolouri of Vanderbilt University and Tahmineh Azizi of University of Wisconsin-Madison for discussion on the Wasserstein distance, and Ilwoo Lyu of Ulsan National Institute of Science and Technology for assistant with the TRACE algorithm.

6. REFERENCES

- [1] A. Cachia, J.F. Mangin, D. Riviere, F. Kherif, N. Boddaert, A. Andrade, D. Papadopoulos-Orfanos, J.B. Poline, I. Bloch, M. Zilbovicius, P. Sonigo, F. Brunelle, and J. Régis, “A primal sketch of the cortex mean curvature: A morphogenesis based approach to study the

- variability of the folding patterns,” *IEEE Trans. Med. Imag.*, vol. 22, no. 6, pp. 754–765, 2003.
- [2] K. Im and P.E. Grant, “Sulcal pits and patterns in developing human brains,” *NeuroImage*, vol. 185, pp. 881–890, 2019.
- [3] S.-G. Huang, I. Lyu, A. Qiu, and M.K. Chung, “Fast polynomial approximation of heat kernel convolution on manifolds and its application to brain sulcal and gyral graph pattern analysis,” *IEEE Transactions on Medical Imaging*, vol. 39, pp. 2201–2212, 2020.
- [4] J. Shi, W. Zhang, and Y. Wang, “Shape analysis with hyperbolic wasserstein distance,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 5051–5061.
- [5] S. Kolouri, G.K. Rohde, and H. Hoffmann, “Sliced wasserstein distance for learning gaussian mixture models,” in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 3427–3436.
- [6] G. Oh, B. Sim, H. Chung, L. Sunwoo, and J.C. Ye, “Unpaired deep learning for accelerated mri using optimal transport driven cyclegan,” *IEEE Transactions on Computational Imaging*, vol. 6, pp. 1285–1296, 2020.
- [7] R. Chartrand, B. Wohlberg, K. Vixie, and E. Boltt, “A gradient descent solution to the monge-kantorovich problem,” *Applied Mathematical Sciences*, vol. 3, pp. 1071–1080, 01 2009.
- [8] D.C. Van Essen, K. Ugurbil, E. Auerbach, D. Barch, T.E.J. Behrens, R. Bucholz, A. Chang, L. Chen, M. Corbetta, and S.W. Curtiss, “The human connectome project: a data acquisition perspective,” *NeuroImage*, vol. 62, pp. 2222–2231, 2012.
- [9] M.F. Glasser, S.N. Sotiropoulos, J.A. Wilson, T.S. Coalson, B. Fischl, J.L. Andersson, J. Xu, S. Jbabdi, M. Webster, J.R. Polimeni, D.C. Van Essen, and M. Jenkinson, “The minimal preprocessing pipelines for the Human Connectome Project,” *NeuroImage*, vol. 80, pp. 105–124, 2013.
- [10] S.M. Smith, C.F. Beckmann, J. Andersson, E.J. Auerbach, J. Bijsterbosch, E. Douaud, G. and Duff, D.A. Feinberg, L. Griffanti, M.P. Harms, et al., “Resting-state fMRI in the Human Connectome Project,” *NeuroImage*, vol. 80, pp. 144–168, 2013.
- [11] M. Jenkinson, P. Bannister, and S. Brady, M. and Smith, “Improved optimization for the robust and accurate linear registration and motion correction of brain images,” *NeuroImage*, vol. 17, no. 2, pp. 825–841, 2002.
- [12] B. Fischl, N. Rajendran, E. Busa, J. Augustinack, O. Hinds, B.T Yeo, H. Mohlberg, K. Amunts, and K. Zilles, “Cortical folding patterns and predicting cytoarchitecture,” *Cerebral Cortex*, vol. 18, no. 8, pp. 1973–1980, 2007.
- [13] I. Lyu, S.H. Kim, N.D. Woodward, M.A. Styner, and B.A. Landman, “TRACE: A topological graph representation for automatic sulcal curve extraction,” *IEEE Trans. Med. Imag.*, vol. 37, no. 7, pp. 1653–1663, 2018.
- [14] M.K. Chung, K.M. Dalton, L. Shen, A.C. Evans, and R.J. Davidson, “Weighted fourier series representation and its application to quantifying the amount of gray matter,” *IEEE Transactions on Medical Imaging*, vol. 26, no. 4, pp. 566–581, 2007.
- [15] G. Peyré and M. Cuturi, “Computational optimal transport,” *Foundations and Trends in Machine Learning*, vol. 11, no. 5-6, pp. 355–607, 2019.
- [16] C.R. Givens and R.M. Shortt, “A class of wasserstein metrics for probability distributions,” *Michigan Mathematical Journal*, vol. 31, no. 2, pp. 231–240, 1984.
- [17] M. Knott and C. S. Smith, “On the optimal mapping of distributions,” *Journal of Optimization Theory and Applications*, vol. 43, no. 1, pp. 39–49, 1984.
- [18] S. Kolouri, S.R. Park, M. Thorpe, D. Slepcev, and G.K. Rohde, “Optimal mass transport: Signal processing and machine-learning applications,” *IEEE Signal Processing Magazine*, vol. 34, no. 4, pp. 43–59, 2017.
- [19] J. Munkres, “Algorithms for the assignment and transportation problems,” *Journal of the Society for Industrial and Applied Mathematics*, vol. 5, no. 1, pp. 32–38, 1957.
- [20] J. Benamou and Y. Brenier, “A computational fluid mechanics solution to the monge-kantorovich mass transfer problem,” *Numerische Mathematik*, vol. 84, no. 3, pp. 375–393, 2000.
- [21] M.K. Chung, L. Xie, S.-G. Huang, Y. Wang, J. Yan, and L. Shen, “Rapid acceleration of the permutation test via transpositions,” *International Workshop on Connectomics in Neuroimaging*, vol. 11848, pp. 42–53, 2019.
- [22] E. Luders, C. Gaser, K.L. Narr, and A.W. Toga, “Why sex matters: brain size independent differences in gray matter distributions between men and women,” *Journal of Neuroscience*, vol. 29, pp. 14265–14270, 2009.
- [23] K. Im, J.-M. Lee, J. Lee, Y.-W. Shin, I. Y. Kim, J. S. Kwon, and S.I. Kim, “Gender difference analysis of cortical thickness in healthy young adults with surface-based methods,” *NeuroImage*, vol. 31, pp. 31–38, 2006.