Product Space Decompositions for Continuous Representations of Brain Connectivity

Daniel Moyer, Boris A. Gutman, Neda Jahanshad, and Paul M. Thompson

Imaging Genetics Center, Stevens Institute for Neuroimaging and Informatics, University of Southern California
moyerd@usc.edu

Abstract. We develop a method for the decomposition of structural brain connectivity estimates into locally coherent components, leveraging a non-parametric Bayesian hierarchical mixture model with tangent Gaussian components. This model provides a mechanism to share information across subjects while still including explicit mixture distributions of connections for each subject. It further uses mixture components defined directly on the surface of the brain, eschewing the usual graph-theoretic framework of structural connectivity in favor of a continuous model that avoids a priori assumptions of parcellation configuration. The results of two experiments on a test-retest dataset are presented, to validate the method. We also provide an example analysis of the components.

Keywords: Brain Connectivity, Non-Parametric Bayes, Unsupervised Learning

1 Introduction

Structural brain connectivity is commonly estimated from the distribution of streamline endpoint pairs on the cortical surface. It is known that both this distribution and the underlying biological distribution are non-uniform and highly structured. Further, it can be safely assumed that there is some degree of correspondence between the structures of different (healhty) subjects. However, the recovery, representation, and study of the variation of these structures remains non-trivial. It is thus useful to construct methods for that purpose.

Most connectivity representations rely on discrete graphs of cortical parcels. This places a strong constraint on the structure of the endpoint distributions recovered. In particular, the use of fixed parcels abstracts away the geometry of the surface, so that all endpoints connecting two regions are equivalent and indistinguishable. The discrete topology of the graph and removal of the underlying distance metric removes notions of nearness to a particular region, except within the graph context itself.

Instead, we model the appearance of these streamline endpoints on the cortical surface directly. We introduce a hierarchical Dirichlet process mixture model for the appearance of streamline endpoints on the white matter/grey matter

interface (the inner cortical surface). This model simultaneously learns clusters in each subject's (continuous) connectome yet also learns correspondences between these clusters by organizing subject-wise clusters into prototypical cluster components. This second level of mixture modeling allows further study of the variation across subjects. We use the Bayesian non-parametric frame, which does not enforce a strict choice of the number of clusters.

1.1 Previous Relevant Work

This work directly relies on two models, the hierarchical Dirichlet process (HDP) proposed by Teh et al. [12] and the tangent Gaussian used as mixture components from Straub et al. [11]. The original HDP manuscript applied the model to topic identification in text corpora. Similar models have been used in connectomics before for node clustering, in particular Jbabdi et al. [5], which attempts to cluster cortical regions from connectivity traces using the HDP and a conjugate prior pair. Notably, their proposed method does not use a surface, modeling voxels with only a prior to enforce spatial coherence. Other Bayesian non-parametric methods have also been suggested, including the directed dependent Chinese restaurant process (ddCRP) [1,7], and a Chinese restaurant process stochastic blockmodel (also known as the Infinite Relational Model) [4].

The vast majority of connectomics research uses discrete graph structures instead of a continuous domain [10]. Several works make use of continuous connective representations [6,7], but do not perform streamline clustering. On discrete domains, spatially sensitive edge clustering is less meaningful, as the graph abstracts away the local geometry. We use the continuous domain in order to be able to identify components that do not conform to parcel pairs defined a priori.

Parcellation tasks using connectivity data are also well studied (e.g. [9]), and are often a pre-processing step for functional connectivity; however, such models only consider "half" the space, while the data are observed on the product of the surfaces (i.e. over pairs of surface subsets). Outside of connectomics, streamline clustering has been studied with respect to white matter trajectories. In particular, the work of O'Donnell and Westin [8] and later Wassermann et al. [14] define forms of kernel spaces for tract registration and clustering tasks.

2 Model

We model the observation of streamline intersections with the inner cortical surface as a random Poisson point process on the product of spheres $S^2 \times S^2$, with each hemisphere of the brain being modeled by its spherical inflation, disjoint from the other hemisphere. In this frame each streamline is an independent observation drawn from a latent intensity function. It is in this context that we would like to construct a parametric mixture model for the appearance of streamline pairs on the cortical surfaces. We also would like to share information between subjects' particular mixtures. We do not know a priori the number

of mixture components nor their configuration. To this end, we introduce a hierarchical Dirichlet process model with tangent Gaussian components.

We will construct the model in the proceeding sections from the top down, then describe the sampling method. We include in the supplemental material a table of notation, as the bookkeeping and notation of these models is fairly complex.

2.1 Hierarchical Dirichlet Process

We choose a hierarchical Dirichlet process (HDP) [12] to model our mixture distributions. Also known as the Chinese Restaurant Franchise, the HDP model is a hierarchical extension of the Chinese Restaurant Process, a popular non-parametric Bayesian mixture model. In such mixtures, the restaurant process serves as a prior over all possible label assignments (assuming the labels have no order between themselves). In a practical sense this means that they assign a probability to each possible partition of the dataset for any number of partitions (clusters), given a base mixture distribution. The hierarchical extension of this allows each mixture component to be drawn from one of a finite (but not fixed) number of prototypical components.

Let G be a prior distribution on component parameters, and let F be a point-mass distribution of prototypical components drawn from G. For a number of dishes K we sample parameters $\theta_k \sim G$, then for subjects $n \in \{1, \ldots, N\}$, we sample components F_n, from which we sample observed endpoints.

$$g_c^n, \theta_k \sim G(\gamma, G_0) \qquad\qquad t_i^n, x_i \sim F_n(\alpha, \theta_k)$$

Here, we also include two sets of association indexes, $t_i^n \in \{1, \ldots, C_n\}$ which tracks the association of data points to clusters, and $g_c^n \in \{1, \ldots, K\}$ which tracks the association of clusters to parameters θ_k. Neither K nor the C_n are fixed. Each of these are Dirichlet processes with concentration parameters γ and α. From a practical standpoint, for each subject we sample component labels over "active" clusters F_n, possibly sampling a new cluster with probability set by α. If a new cluster is sampled, we will sample from θ_k (from which all clusters across subjects have been drawn), possibly adding a new prototypical cluster.

2.2 Tangent Gaussian Mixture Components

Our observed data lie on the white matter/gray matter interface, a surface commonly abstracted to S^2 (e.g. via FreeSurfer [2]). We observe pairs of points on the surface, streamline intersections with the gray matter boundary, so our induced space is $S^2 \times S^2$. The vast majority of connectomics papers further abstract this by dividing the space into regions, conducting their analysis on the discrete graph formed by the set product of the regions. We instead choose to model the data directly in $S^2 \times S^2$.

A recently introduced component distribution for the sphere is the so-called Tangent Gaussian of Straub et al.[11]. Given a cluster mean, each component

is modeled by a zero mean Gaussian *in the tangent space* of the sphere at the cluster mean. This allows for anisotropic distributions with, as we show in the supplemental material, a reasonable posterior predictive distribution for certain choices of hyper-prior parameters. We choose to use a product of Tangent Gaussians to model our data.

Define the following functions to and from the tangent space for a mean μ:

$$\begin{aligned} &\text{Log} : S^2 \times S^2 \to \mathbb{R}^2 \qquad &&\text{Log}(x;\mu) = (x - \mu(x \cdot \mu))\frac{\theta}{\sin(\theta)} \\ &\text{Exp} : \mathbb{R}^2 \times S^2 \to S^2 \qquad &&\text{Exp}(x;\mu) = \mu\cos(\|x\|_2) + \frac{x}{\|x\|}\sin(\|x\|_2) \end{aligned}$$

Here, $\theta = \arccos(x \cdot \mu)$. For cluster mean and tangent covariance (μ_k, Σ_k), the likelihood of a datapoint $s \in S^2$ is $\mathcal{N}(\text{Log}(s;\mu); 0, \Sigma)$. Our data $x_i = (s_1, s_2)$ are on $S^2 \times S^2$, thus we use the product density

$$p(x_i|\mu, \Sigma) = \mathcal{N}(\text{Log}(s_1;\mu_1); 0, \Sigma_1) \times \mathcal{N}(\text{Log}(s_2;\mu_2); 0, \Sigma_2)$$

which has pairs of parameters (μ_1, μ_2) and (Σ_1, Σ_2). For convenience (and because their respective priors are assumed to be independent) we refer to each pair as simply μ and Σ.

2.3 Full Model and Sampling

Our full model is then

$$\begin{aligned} g_c^n, \theta_k &\sim G(\gamma, G_0) \\ \theta_k = (\mu_k, \Sigma_k) &\sim \text{Unif}(S^2), \text{Inv. Wishart}(\Delta, \nu) \\ t_i^n &\sim F_n(\alpha, \theta_k) \\ x_i = (s_1, s_2)_i^n &\sim \mathcal{N} \times \mathcal{N} \mid t_i^n, \theta_k \end{aligned}$$

We have four hyper parameters, (Δ, ν) the prior parameters for the Inverse Wishart distribtion, and two concentration parameters α and γ. We use Gibbs sampling to sample model parameters, using multiple layers of sampling. As suggested by Teh et al. [12], we use component indexes to (somewhat) simplify bookkeeping during sampling. For each data point in each subject, we need to sample its component label t_i^n. We then sample component parameters (μ_k, Σ_k), which are across subjects (cohort-wide parameters). Finally, we sample component associations of each subject's clusters to prototypical components, g_c. In practice, we find that due to the particular cluster components and configuration of the data, this last step rarely changes cluster assignments.

We assume a uniform prior on μ and an inverse Wishart prior with parameters (Δ, ν) for Σ. This is G_0 from which we draw new parameters $\theta = (\mu, \Sigma)$. To avoid multiple subscripts, we denote the parameters associated with cluster c as $\mu : g_c^n$ and $\Sigma : g_c^n$. Sampling can then be divided into three steps:

1) Sample t_i^n for each streamline, adding clusters as dictated by the appearance of c_{new} and deleting clusters if $\sum_n \eta_k^n = 0$. This has marginal distribution

$$p(t_i^n = c|\{\mu, \Sigma\}, G_0) \propto \mathcal{N} \times \mathcal{N}(x)|\mu : g_c^n, \Sigma : g_c^n \tag{1}$$
$$= \mathcal{N} \times \mathcal{N}(\mathrm{Log}(s_1; \mu_1), \mathrm{Log}(s_2; \mu_2))|\mu : g_c^n, \Sigma : g_c^n$$
$$p(t_i^n = c_{new}|\{\mu, \Sigma\}, G_0) \propto \alpha \sum_k \frac{\eta_k}{C_* + \gamma} p(x|\mu_k, \Sigma_k) \tag{2}$$
$$+ \frac{\gamma}{C_* + \gamma} p(x|\mu_k, \Sigma_k, k_{new})$$

If c_{new} is sampled, then we need to choose a corresponding g_c^n. This has marginal distribution

$$P(g_c^n = k|c_{new}, \{\theta\}, G_0) \propto p(x|\mu_k, \Sigma_k) \tag{3}$$
$$P(g_c^n = k_{new}|c_{new}, \{\theta\}, G_0) \propto \gamma p(x|\mu_k, \Sigma_k, k_{new}) \tag{4}$$
$$p(x|\mu_k, \Sigma_k, k_{new}) = \iint p(x|\mu, \Sigma) p(\mu) p(\Sigma|\Delta, \nu) \tag{5}$$
$$= \underbrace{\frac{1}{|Sphere|} \iint p(x|\mu, \Sigma) p(\mu) p(\Sigma|\Delta, \nu) d\mu d\Sigma}_{Constant}$$

We provide a tractable method of computing the double integral in the supplemental material for reasonable priors. This needs only to be computed once.

2) Sample μ, Σ. Σ has the marginal distribution

$$p(\Sigma|\{x_i\}, \{t_i^n\}, \{g_c^n\}, \Delta, \nu) = \mathrm{IW}(\Delta + A_k, \nu + \sum_{n,c} \eta_c^n) \tag{6}$$
$$\text{with} \quad A_k = \sum_n \sum_{\substack{i: t_i^n = c \\ g_c^n = k}} \mathrm{Exp}(x_i; \mu : g_c^n)\, \mathrm{Exp}(x_i; \mu : g_c^n)^T$$

Due to the tangent Gaussian approximations, we choose to use a Metropolis-Hastings correction (as suggested by Straub et al. [11]) and rejection step to sample μ:

$$q(\mu_k = \mu|.) = \mathcal{N}(\mathrm{Log}(\mu; \langle x \rangle_c); 0, \frac{\Sigma_k}{N})$$
$$r = \prod_n \prod_{\substack{i: t_i^n = c \\ g_c^n = k}} \frac{p(x_i|\mu_k^{prop}, \Sigma_k) p(\mu_k^{prop}) q(\mu_k^{old}|.)}{p(x_i|\mu_k^{old}, \Sigma_k) p(\mu_k^{old}) q(\mu_k^{prop}|.)}$$
$$= \frac{\mathcal{N}(\mathrm{Log}(\mu_k^{old}; \langle x \rangle_c)|0, \frac{\Sigma_k}{N})}{\mathcal{N}(\mathrm{Log}(x; \mu_k^{prop})|0, \frac{\Sigma_k}{N})} \prod_n \prod_{\substack{i: t_i^n = c \\ g_c^n = k}} \frac{\mathcal{N}(\mathrm{Log}(\mu_k^{prop}; \langle x \rangle_c)|0, \Sigma_k)}{\mathcal{N}(\mathrm{Log}(x; \mu_k^{old})|0, \Sigma_k)}$$

Here, we use the Karcher mean of the cluster to resample μ, denoted $\langle x \rangle_c$.

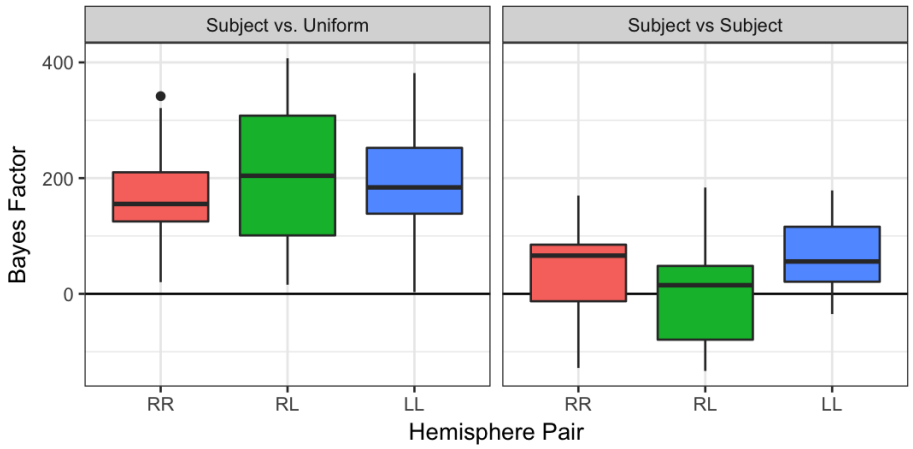

Fig. 1. Plotted above are the likelihood ratios (the Bayes factors) using the trained model on data from the retest scans and two different false datasets. On the **left** are the Retest vs. Uniform random likelihood ratios, on the **right** the Retest vs. other Subjects ratios.

3) Sample g_c^n for each cluster, deleting prototypes if no cluster claims them.

$$p(g_c^n = k|\{x_i\}, \{t_i^n\}, \{\theta_k\}) \propto \prod_n \prod_{\substack{i:t_i^n=c \\ g_c^n=k}} \mathcal{N} \times \mathcal{N}(x_i)|\mu : g_c^n, \Sigma : g_c^n$$

3 Experiments

We demonstrate the use of our framework on a test-retest dataset. Our data are comprised of 21 subjects from the Institute of Psychology, Chinese Academy of Sciences sub-dataset of the Consortium for Reliability and Reproducibility (CoRR) dataset [15]. T1-weighted (T1w) and diffusion weighted (DWI) images were obtained on 3T Siemens TrioTim using an 8-channel head coil and 60 directions. Each subject was scanned twice, roughly two weeks apart.

T1-weighted images were processed with FreeSurfer's [2] recon-all pipeline to obtain a triangle mesh of the gray-white matter boundary registered to a shared spherical space. Probabilistic streamline tractography was conducted using the DWI in 2-mm isotropic MNI 152 space, using Dipy's [3] implementation of constrained spherical deconvolution (CSD) [13] with a harmonic order of 6. We retain only tracts longer than 5 mm with endpoints in likely grey matter.

We train our model on the set of subjects' first scans using the proposed sampling method, learning the prototype parameters μ, Σ, the per-subject cluster weights η_c^n (subject encodings), and the cluster assignments. After collecting samples we find the Maximum a Posteriori (MAP) likelihood parameter set, and then perform evaluations on the subjects' second scans using the MAP parameter values.

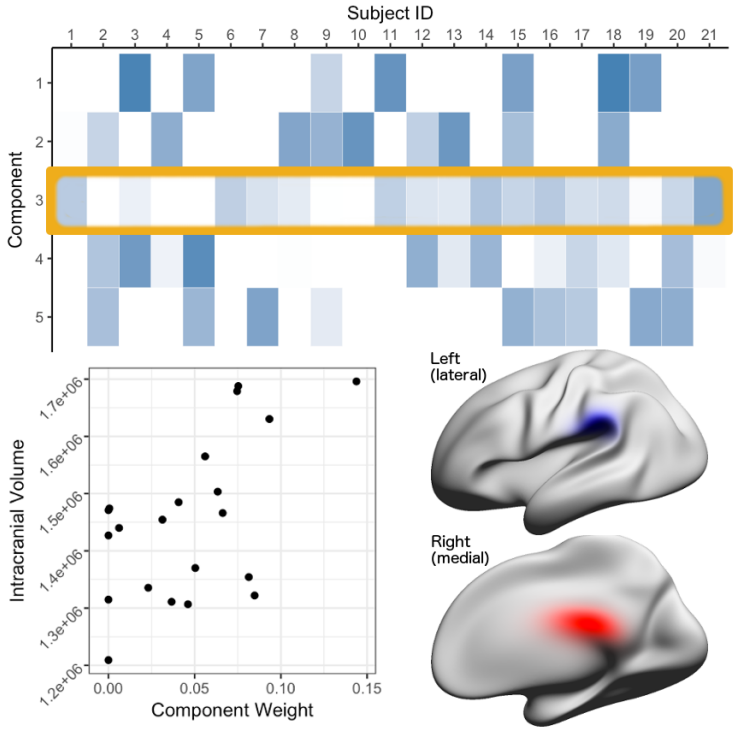

Fig. 2. Top: the top 5 components displayed as a heatmap of their weights across components. Comp. 3 (in Yellow) is displayed below. **Bottom Left**: Component 3 weights vs. Intracranial Volume (each subj. is a point). This plot shows a weak correlation between the appearance of the feature in a subject and cranial volume ($\rho = 0.56$). **Bottom Right**: A plot of the component on a smoothed subject surface, where the red density at approximately the R. post. cingulate connects to the blue density at the temporoparietal junction. The right hemisphere is shown from the medial view.

We perform two different evaluations of the model. Each measures the posterior predictive probability of two datasets given a subject's particular encoding, one being the subject's second scan. In the first experiment, the second scan's probability is compared to a spatially uniform random dataset. In the second, we compare the subject's second scan to an equal number of randomly selected points from other subjects. Note that a separate model is learned for each pair of hemispheres. We use the Bayes Factor as the evaluation metric.

As can be seen in Figure 1, in the first task, differentiating between the actual second scan and a uniform random one, the model performs with 100% accuracy. This is a helpful sanity check, as one would expect the structure to be non-uniform. In the second task, the models perform well, though not perfectly. In particular, the crossing streamline (one hemisphere to another) distributions are less distinguishable between subjects. This may be because there is less observable variation, or simply because the model fails to capture it.

In Figure 2, we present a brief analysis of one component. The observation of this component is weakly correlated with intracranial volume. While this study is clearly not large enough to make concrete conclusions, this demonstrates the interpretability and use of the component/subject-wise encoding.

3.1 Conclusion and Acknowledgements

In the present work we have constructed a multi-subject model for the locally sensitive decomposition of representations of structural brain connectivity in a continuous setting. This model may be useful for exploratory analysis and parcellation free connectomics.

This work was supported by NIH Grant U54 EB020403, as well as the NSF Graduate Research Fellowship Program.

References

1. Baldassano, C., et al.: Parcellating connectivity in spatial maps. PeerJ 3 (2015)
2. Fischl, B.: FreeSurfer. NeuroImage 2(62), 774–781 (2012)
3. Garyfallidis, E., et al.: Dipy, a library for the analysis of diffusion MRI data. Front. Neuroinform 8(8) (2014)
4. Hinne, M., et al.: Probabilistic clustering of the human connectome identifies communities and hubs. PLoS ONE 10(1), e0117179 (2015)
5. Jbabdi, S., et al.: Multiple-subjects connectivity-based parcellation using hierarchical Dirichlet process mixture models. NeuroImage 44(2), 373–384 (2009)
6. Moyer, D., et al.: A continuous model of cortical connectivity. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 157–165. Springer (2016)
7. Moyer, D., et al.: A restaurant process mixture model for connectivity based parcellation of the cortex. In: International Conference on Information Processing in Medical Imaging. Springer (2017)
8. O'Donnell, L., Westin, C.F.: White matter tract clustering and correspondence in populations. Medical image computing and computer-assisted intervention–MICCAI 2005 pp. 140–147 (2005)
9. Parisot, S., et al.: Group-wise parcellation of the cortex through multi-scale spectral clustering. NeuroImage (2016)
10. de Reus, M.A., Van den Heuvel, M.P.: The parcellation-based connectome: limitations and extensions. NeuroImage 80, 397–404 (2013)
11. Straub, J., Chang, J., Freifeld, O., Fisher III, J.: A Dirichlet process mixture model for spherical data. In: Artificial Intelligence and Statistics. pp. 930–938 (2015)
12. Teh, Y.W., et al.: Sharing clusters among related groups: Hierarchical Dirichlet processes. In: Advances in Neural Information Processing Systems (2005)
13. Tournier, J.D., et al.: Resolving crossing fibres using constrained spherical deconvolution: validation using diffusion-weighted imaging phantom data. NeuroImage 42(2), 617–625 (2008)
14. Wassermann, D., et al.: White matter bundle registration and population analysis based on Gaussian processes. In: Biennial International Conference on Information Processing in Medical Imaging. pp. 320–332. Springer (2011)
15. Zuo, X.N., et al.: An open science resource for establishing reliability and reproducibility in functional connectomics. Scientific Data 1 (2014)