

Proposal: *Applying TDA on Purkinje population calcium data*

Inspiration

With the incredible advancement of high resolution neural recording, manipulation technology and computing power, the amount of data for both experimental and simulated neural population activity is increasing tremendously (and personally in the direction of becoming quite overwhelming). Computational and analytical tools to interpret and build models from such volumes of data for hypothesis testing and exploratory purposes are catching up (though personally slowly). Tools like topological data analysis (TDA) and network analysis inevitably are necessary for extracting and exploring potential structures and patterns within these data.

TDA in neuroscience: Over the last decade (roughly two), there have been multiple applications of algebraic topology tools in neuroscience, extending analysis from graph theory methods. Aside from the more theoretical endeavors, these applications involve using TDA in neural population data across different subfields of computational neuroscience, most of which involve clique topology analysis. Examples span across different data domain and areas, from electrophysiological recordings in a specific area [1] (rodent hippocampal place cells), to brain-wide structural human connectome [2], as well as within detailed biophysical neural simulations and models [3].

I will be basing my project around the ideas and methods from [1], which shows the geometric organization in hippocampal correlation structure across different states of the animal, including during wheel-running behavior and sleep. This is accomplished by compare the Betti number curves (edge density vs Betti numbers) from the data with those from generated random geometric networks, as well as comparing with a null based on shuffled data.

Data: Although there have been quite many population studies across other areas like neocortex, subcortical areas and hippocampus, not many network studies are done in the cerebellar Purkinje population recordings, even though these cells are among the popular neurons usually shown in textbooks with their beautiful complicated morphology, responsible for many developmental and learning functions such as motor learning. To my knowledge, there are no studies to date applying TDA in the cerebellum population data.

Hence, I will be applying TDA on Purkinje cell population and drawing comparison with different models of random graphs, similar to [1]. Briefly, the data are calcium imaging (proxy for neural activity) data of Purkinje population in rodent during awake states, responding to sensory stimuli. Each trial is roughly 20 seconds of 60Hz sampling rate, and the stimulus (if present) is briefly on at 10-second. These data are obtained from my lab from another student's experiments (see *Acknowledgements*).

Proposal

Note: I would tend to try “*deliverable*” first for testing, bechmarking purposes. But if there's something wrong, I might try the “*tentative/alternative*”.

Similarity/distance matrix: First, from the data, I will build similarity (or distance) matrix from the activity around 2 seconds before (*base*) and after the stimulus onset (*stim*) for each trial separately.

- Deliverable: I will utilize the simple correlation coefficient $r_{i,j}$ as the first approach, as a measure of similarity. To transform to a distance, I could use $1 - r_{i,j}$
- Tentative/alternative: I could also use mutual information $I(X_i, X_j) = H(X_i) + H(X_j) - H(X_i, X_j)$, where $H(A)$ denotes the entropy of A , as another measure of similarity to include not only linear relationship between neural activity. However, these are not really distance metrics, so I can use $\sqrt{2(1 - r_{i,j})}$ for correlation distance, or variation of information $V(X_i, X_j) = H(X_i, X_j) - I(X_i, X_j)$.
- Note: all these will result in undirected weighted graphs, so filtration methods from class could be applied more straightforwardly.

Different null models:

- Deliverable: Following [1], I will also build null models to compare with the matrices from the data above:
 1. Shuffled: matrices built by shuffling the original matrices
 2. Geometric: random geometric models with nodes sampled from a hypercube, and probability of connectivity of two nodes decrease with their distance.
- Tentative/alternative: Other random structures I could include are:
 1. ER: Erdős-Rényi random graph model with matched statistics
 2. BA: Barabási-Albert model to generate scale-free networks (I did observe some baseline degree distributions looking close to that of a scale-free model in preliminary analysis)

TDA:

- Deliverable:
 1. I will generate Betti curves (up to $\beta_3(\rho)$) as a function of the edge density similarly to [1] using TDA packages suggested in class.
 2. I will first confirm that I could replicate Fig. 2B (main) in [1]. Then I will repeat for the data for each trial, as well as the null model counterparts.
 3. Similar to [1], I will compare the integrated Betti curves (S_β) of the data and the S_β distributions from the null models. Then I will use statistical test to observe if there is any significant difference between the data and the respective nulls.
 4. Additionally, I will also compare S_β distributions aggregated from across different trials, between the *base* and *stim* windows, as well as between *sham* trial (no stimulus present) and *light* trial (trial with light stimulation to the animal).
- (Super-) tentative/alternative: there are also some recordings with other stimuli groups (*air puff*, *sound* in addition to *sham* and *light*). In the optimistic case where all the above goes smoothly, I could apply the same analysis on these additional groups, and perform comparison across groups. Furthermore, I could use persistence landscapes of these different groups and try to see if I can build a model to classify the different stimuli groups. Since the number of trials per group are small (usually less than 20 for practical reasons), I could use train on 95% of all groups and test on the remaining 5%, and repeat for different permutations of the dataset.

Acknowledgements

The experimental data are obtained from my lab (Dr. Christian Hansel) by Silas Busch and Ting-Feng Lin. I would like to acknowledge and thank them for allowing me to use their data to play around for the project. Since these data are not published yet, I will not make the data public but the codes and final report will be.

- [1] C. Giusti, E. Pastalkova, C. Curto, and V. Itskov, “Clique topology reveals intrinsic geometric structure in neural correlations,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 112, no. 44, pp. 13455–13460, Nov. 2015, doi: 10.1073/pnas.1506407112.
- [2] A. E. Sizemore, C. Giusti, A. Kahn, J. M. Vettel, R. F. Betzel, and D. S. Bassett, “Cliques and cavities in the human connectome,” *J. Comput. Neurosci.*, vol. 44, no. 1, pp. 115–145, Feb. 2018, doi: 10.1007/s10827-017-0672-6.
- [3] M. W. Reimann *et al.*, “Cliques of neurons bound into cavities provide a missing link between structure and function,” *Front. Comput. Neurosci.*, vol. 11, p. 48, Jun. 2017, doi: 10.3389/fncom.2017.00048.