2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)
Venice, Italy, April 8-11, 2019

CIRCULAR PEARSON CORRELATION USING COSINE SERIES EXPANSION

Shih-Gu Huang[⋆] *Andrey Gritsenko*[⋆] *Martin A. Lindquist*[‡] *Moo K. Chung*[⋆]

[⋆] University of Wisconsin, Madison, USA
[‡]Johns Hopkins University, Baltimore, USA

ABSTRACT

In resting-state fMRI, there is no external anchor that will lock brain activation across voxels. Thus, correlation of fMRI time series between voxels is often done by computing coherence in the frequency domain. However, such approach ignores the time lag of fMRI time series across voxels. To address the problem, we propose to use the concept of circular Pearson correlation in determining the time lag, which locks the time series, and the maximum correlation at locking. We further express the circular Pearson correlation analytically in terms of cosine series expansion. The proposed method is applied to 208 twin pairs to determine if the time lag and the maximum correlation are heritable genetic features.

1. INTRODUCTION

In resting-state fMRI, there is no starting time point to correlate two signals across different voxels. There are often time lags between fMRI time series in different voxels. Coherence, the correlation in the frequency domain, has been mainly used to correlate time series in such a situation [1, 2]. In practice, coherence is usually estimated by Welch's segment averaging method [3]. To reduce the sidelobe caused by segment truncation, windowing is performed. There is a trade-off between the variance and the spectral resolution and bias of the estimate. Longer segment gives a better spectral resolution and smaller bias but higher variance. If the ratio of time lag to segment length is high, the bias can be considerable [4]. Increasing the overlap between segments can alleviate these problems, but the computational complexity will increase rapidly. Although coherence is very useful, substantial efforts are needed to make it work. Further, coherence does not provide the explicit estimate of time lag.

To address the time lag issue in correlating fMRI time series, we propose to use the circular Pearson correlation [5], where we can estimate the time lag and the maximum correlation. As a demonstration, the proposed methods are applied to resting-state fMRI of twins. We show that maximum correlation is a sensitive heritable genetic feature. Further, we will show that the time lag corresponding to the maximum correlation is also a heritable feature.

Correspondence should be sent to Moo Chung (email: mkchung@wisc.edu). This study was funded by NIH Grant EB022856.

2. PRELIMINARY

Definition. Let $s \in [0, 1)$. A *rotation* is a map defined as

$$R_s : [0, 1) \to [0, 1), \qquad R_s(t) = t + s \pmod 1,$$

where mod denotes the modulo operation.

Let $R_{s_2} \cdot R_{s_1}$ denote applying rotation R_{s_2} after rotation R_{s_1}. The set of all rotations $G = \{R_s : s \in [0, 1)\}$ together with the operation $\cdot$ forms an Abelian group, i.e., a commutative group which satisfies the axiom of commutativity. This Abelian group is called the *rotation group* [6]. Note we have identity R_0 and inverse $R_s^{-1} = R_{1-s}$. Since the data under analysis have finite support $[0, 1)$, we can specify an action of the rotation group on the set of all functions in $[0, 1)$.

Definition. Let F be the set of all functions in $[0, 1)$. An action of the rotation group G on F is a map from $G \times F$ to F written as $R_s f(t)$ and defined as

$$R_s f(t) = f\left(R_s(t)\right) = f\left(t + s \pmod 1\right),$$

for all $R_s \in G$ and $f \in F$. This group action is referred to as *circular shift*.

Circular shift is an action of the Abelian group built by rotations on a circle.

3. METHODS

3.1. Motivation

We restrict the domain of fMRI signals to the unit interval $[0, 1)$. If not, we can always scale the domain to the unit interval. Consider signals $f(t)$ and $g(t)$ in $[0, 1)$. For simplicity, normalize f and g to zero mean, i.e., $\int_0^1 f(t)dt = \int_0^1 g(t)dt = 0$. Then the integral version of Pearson correlation between $f(t)$ and $g(t)$ is defined as

$$\gamma_{f,g} = \frac{\int_0^1 f(t)g(t)dt}{\sigma_f \sigma_g}, \tag{1}$$

where $\sigma_f^2 = \int_0^1 f^2(t)dt$ and $\sigma_g^2 = \int_0^1 g^2(t)dt$ are the variances of $f(t)$ and $g(t)$. In resting-state fMRI, functional connectivity based on the usual zero-lag correlation (1) is heavily

influenced by a complex time lag structure [7]. The goal and the main contribution of this paper are to directly find correlation measure that remains invariant to such time lag.

3.2. Circular Pearson Correlation

For resting-state fMRI, the signals under analysis may be of the form $R_{s_1}f(t)$ and $R_{s_2}g(t)$ with unknown time lags s_1 and s_2. Since the integral version of Pearson correlation (1) between $R_{s_1}f(t)$ and $R_{s_2}g(t)$ dependents on s_1 and s_2, it needs to be modified such that it is invariant to time lag. For this purpose, continuous circular correlation is used.

Definition. Given two signals $f(t)$ and $g(t)$ with $t \in [0,1)$, the circular correlation of $f(t)$ and $g(t)$ is defined by

$$R_{f,g}(\tau) = \int_0^1 f(t)R_\tau g(t)dt, \tag{2}$$

where R_τ is in the rotation group $G = \{R_s : s \in [0,1)\}$. The normalized circular cross-correlation is then defined as [5, 8]

$$\rho_{f,g}(\tau) = R_{f,g}(\tau)/(\sigma_f\sigma_g). \tag{3}$$

This definition will be simply called as the *circular Pearson correlation* through the paper. From (3), we have

$$\rho_{f,R_{s_2-s_1\,(\mathrm{mod}\,1)}g}(\tau) = \rho_{f,g}(\tau + s_2 - s_1(\mathrm{mod}\,1)).$$

Thus, we have the following rotation invariance.

Theorem 1. *The circular Pearson correlation of $R_{s_1}f(t)$ and $R_{s_2}g(t)$ is equivalent to the circular shift of the circular Pearson correlation of $f(t)$ and $g(t)$ with shift $s_2 - s_1(mod\ 1)$,*

$$\rho_{R_{s_1}f,R_{s_2}g}(\tau) = R_{s_2-s_1\,(mod\,1)}\ \rho_{f,g}(\tau).$$

This invariant property implies that $\rho_{R_{s_1}f,R_{s_2}g}(\tau)$ has the same correlation information as $\rho_{f,g}(\tau)$. Thus, they have the same maximum value. Subsequently, the maximum of the circular Pearson correlation of $R_{s_1}f(t)$ and $R_{s_2}g(t)$ is invariant to s_1 and s_2,

$$\max_\tau\ \rho_{R_{s_1}f,R_{s_2}g}(\tau) = \max_\tau\ \rho_{f,g}(\tau).$$

3.3. Cosine Series Expansion

To filter out noise, the resting-state fMRI signals f and g are represented by the cosine series expansion:

$$f(t) = \sum_{n=0}^{\infty} f_n\psi_n(t), \qquad g(t) = \sum_{n=0}^{\infty} g_n\psi_n(t),$$

where $\psi_0(t) = 1$, $\psi_n(t) = \sqrt{2}\cos\pi nt$, and f_n and g_n are the cosine series coefficients [9].

Theorem 2. *The circular Pearson correlation can be expressed in terms of cosine series coefficients*

$$\rho_{f,g}(\tau) = \frac{R_{f,g}(\tau)}{\sqrt{\sum_{n=0}^{\infty} f_n^2}\sqrt{\sum_{n=0}^{\infty} g_n^2}},$$

where $R_{f,g}(\tau)$ is given by

$$\begin{aligned} R_{f,g}(\tau) &= \sum_{n=0}^{\infty} f_n g_n \cos\pi n\tau \\ &- \sum_{n=1}^{\infty} 2f_n g_n (n_{mod\,2}) \left(\tau\cos\pi n\tau + \frac{1}{\pi n}\sin\pi n\tau\right) \\ &+ \sum_{\substack{m,n=1 \\ m\neq n}}^{\infty} 4f_m g_n \frac{(n_{mod\,2})(-1)^m m\sin\pi m\tau + (m_{mod\,2})n\sin\pi n\tau}{\pi(m^2-n^2)}. \end{aligned}$$

The theorem can be proved by expressing functions f and g in (2) and (3) using the cosine series expansion and using the orthonormality of $\psi_k(t)$. The *maximum correlation* and the *time lag* at the maximum correlation are given by

$$\rho_{fg}^{max} = \max_\tau\ \rho_{f,g}(\tau) \quad \text{and} \quad \tau_{fg}^{max} = \arg\max_\tau\ \rho_{f,g}(\tau).$$

Theorem 3. *The circular Pearson correlation is not commutable but has the property that $\rho_{f,g}(\tau) = \rho_{g,f}(1-\tau)$. It follows that*

$$\rho_{fg}^{max} = \rho_{gf}^{max} \qquad \textit{and} \qquad \tau_{fg}^{max} = 1 - \tau_{gf}^{max}.$$

4. APPLICATION

4.1. Dataset and image preprocessing

We used the resting-state fMRI of 208 twin pairs from the Human Connectome Project [10]. fMRI consists of 2mm isotropic voxels and 1200 time points over 14 min. 33 sec. scanning session. The fMRI data has undergone spatial and temporal preprocessing including motion and physiological noise removal [11]. We used genetically confirmed 131 monozygotic (MZ) twin pairs (age 29.3±3.3 years, 56M/75F) and 77 same-sex dizygotic (DZ) twin pairs (age 29.1 ± 3.5 years, 30M/47F) in this study. We employed the Automated Anatomical Labeling (AAL) brain template to parcellate the brain volume into 116 regions [12]. The fMRI data were then averaged across voxels in each brain region for each subject. Each averaged fMRI time series was scaled to fit to unit interval $[0,1)$ and subtracted its mean over time. The degree $k = 119$ cosine series expansion was fitted in the least squares fashion such that fMRI data were compressed into 10% of the original data size and achieved the signal-to-noise ratio 1.81 in average over all 116 regions.

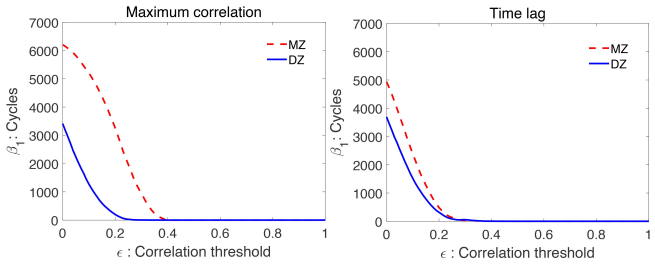

Fig. 1. Betti-plots showing Betti number β_1 over changing correlation value ϵ. Both of using maximum correlation and time lag show that MZ-twins (dashed line) have more cycles compared to DZ-twins (solid line). These plots show that β_1 is a heritable topological feature.

4.2. Heritability of maximum correlation and time lag

The subject level connectivity c_{ij} between the i-th and j-th parcellations is measured by the maximum correlation. The twin correlation is computed by the Pearson correlation of paired c_{ij} within each twin type. Since there is no preference in the order of twins, permuting c_{ij} between pairs in each twin type results in a different value of twin correlation. Hence, the average of twin correlations from the random permutations, denoted by c_{ij}^{MZ}, is taken as the estimate of the actual twin correlation for MZ-twins. Similarly, we computed the average twin correlation c_{ij}^{DZ} for DZ-twins. Due to high correlation between twins, only 37 and 58 permutations were required for MZ- and DZ-twins to guarantee the convergence within 4 decimal places in terms of the mean of absolute error of matrix entries.

The heritability index (HI), which determines the amount of variation due to genetic influence in a population [13], is estimated as $h_{ij} = 2(c_{ij}^{MZ} - c_{ij}^{DZ})$. The network differences between MZ- and DZ-twin correlation matrices are considered mainly contributed to heritability and can be used to determine the statistical significance of HI [13].

We also constructed twin correlations and HI by time lags. The number of permutations required for time lags is 39 and 66 for MZ- and DZ-twin correlations to guarantee 4 decimal accuracy.

4.3. Results

The statistical analysis of the significance of HI is done by the *exact topological inference* on β_1-plot [13]. The first Betti number β_1 counts the number of cycles in a network. More cycles imply the connected components are more densely connected. We built brain networks with twin correlations as edge weights. By thresholding the correlations, more cycles in the networks are removed and β_1 decreases as the threshold value increases. Figure 1 shows β_1-plots of twins on the maximum correlation and time lag. The test statistic is given by the maximum difference in β_1-plots [13]. Bigger difference implies higher heritability of the feature. At the same correlation value, MZ-twins have more cycles than DZ-twins. Such topological differences are contributed to heritability.

Maximum correlation. The maximum difference between MZ- and DZ-twins in β_1-plots is 3943 (p-value $< 10^{-32}$), which is larger than the maximum difference 3627 obtained from the traditional Pearson correlation. The explicit account of time lag increased the performance. Figure 2 displays the HI which gives 100% heritability, i.e., $h_{ij} \geq 1$, obtained from the maximum correlation. The most heritable connections include the left and right middle frontal gyri, left and right superior frontal gyri, left and right thalami, right caudate nucleus, left anterior and posterior cingulate gyri among other regions. Most regions overlap with highly heritable regions observed in resting-state connectivity of twins in a different study [14]. The right caudate nucleus and right superior frontal gyrus are identified as most heritable hub nodes.

Time lag. Form Figure 1, the β_1-plot of time lags also shows that MZ-twins have more cycles than DZ-twins. The maximum difference is 1271, which gives the p-value smaller than 10^{-32}. Thus, time lag is also a heritable feature of rs-fMRI. Figure 2 displays the HI that gives 100% heritable connections obtained from time lags. The most heritable connections include the left and right precuneus, left anterior cingulate gyrus, and left middle cingulate among other regions.

5. DISCUSSION

Circular Pearson correlation was represented by the group action and computed using the cosine series expansion. We used the circular Pearson correlation to determine maximum correlation and time lag of fMRI time series between voxels. We showed that the maximum correlation and time lag are genetically heritable features. We further identified the most heritable brain network connections based on these two features.

Our time lag model only uses fMRI signals at two different voxels. It is possible to build a more sophisticated time lag model that accounts for spatial dependency of time lag using the random field theory [15]. In this paper, only the maximum positive correlation is taken into account. One can also use the circular Pearson correlation to compute the maximum *negative* correlation and the corresponding time lag. One can further use the time lag to analyze the neural causality across brain regions [16]. These are left as future studies.

Acknowledgements. We would like to thank Gregory Kirk of University of Wisconsin and Li Shen of University of Pensilvania for valuable support and discussions.

6. REFERENCES

[1] F.T. Sun, L.M. Miller, and M. D'esposito, "Measuring interregional functional connectivity using coher-

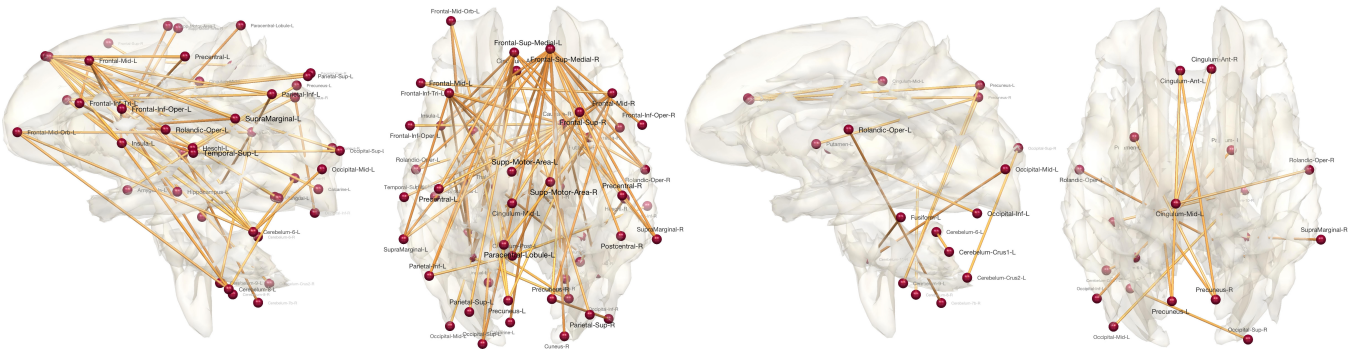

Fig. 2. Most highly heritable connections corresponding to maximum correlation (left two figures) and time lag (right two figures). Only the connections with 100% heritability are shown.

ence and partial coherence analyses of fMRI data," *Neuroimage*, vol. 21, no. 2, pp. 647–658, 2004.

[2] B. Thirion, S. Dodel, and J.-B. Poline, "Detection of signal synchronizations in resting-state fMRI datasets," *Neuroimage*, vol. 29, no. 1, pp. 321–327, 2006.

[3] P. Welch, "The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms," *IEEE Transactions on Audio and Electroacoustics*, vol. 15, no. 2, pp. 70–73, 1967.

[4] J.H. Miles, "Estimation of signal coherence threshold and concealed spectral lines applied to detection of turbofan engine combustion noise," *The Journal of the Acoustical Society of America*, vol. 129, no. 5, pp. 3068–3081, 2011.

[5] R.J. Schilling and S.L. Harris, *Fundamentals of digital signal processing using MATLAB*, Cengage Learning, 2011.

[6] L. Fuchs, *Infinite Abelian groups*, vol. 1, Academic Press, 1970.

[7] R.J. Meszlényi, P. Hermann, K. Buza, V. Gál, and Z. Vidnyánszky, "Resting state fMRI functional connectivity analysis using dynamic time warping," *Frontiers in Neuroscience*, vol. 11, pp. 75, 2017.

[8] C. Chatfield, *The analysis of time series: an introduction*, CRC press, 2016.

[9] M.K. Chung, N. Adluru, J.E. Lee, M. Lazar, J.E. Lainhart, and A.L. Alexander, "Cosine series representation of 3d curves and its application to white matter fiber bundles in diffusion tensor imaging," *Statistics and Its Interface*, vol. 3, pp. 69–80, 2010.

[10] 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.

[11] S.M. Smith, C.F. Beckmann, J. Andersson, E.J. Auerbach, J. Bijsterbosch, and et. al., "Resting-state fMRI in the Human Connectome Project," *NeuroImage*, 2013.

[12] N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, and M. Joliot, "Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain," *NeuroImage*, vol. 15, pp. 273–289, 2002.

[13] M.K. Chung, V. Vilalta-Gil, H. Lee, P.J. Rathouz, B.B. Lahey, and D.H. Zald, "Exact topological inference for paired brain networks via persistent homology," *Information Processing in Medical Imaging (IPMI), Lecture Notes in Computer Science (LNCS)*, vol. 10265, pp. 299–310, 2017.

[14] D.C. Glahn, A.M. Winkler, P. Kochunov, L. Almasy, R. Duggirala, M.A. Carless, J.C. Curran, R.L. Olvera, A.R. Laird, S.M. Smith, C.F. Beckmann, P.T. Fox, and J. Blangero, "Genetic control over the resting brain," *Proceedings of the National Academy of Sciences*, vol. 107, pp. 1223–1228, 2010.

[15] K.J. Worsley, J. Cao, T. Paus, M. Petrides, and A.C. Evans, "Applications of random field theory to functional connectivity," *Human Brain Mapping*, vol. 6, pp. 364–7, 1998.

[16] S.M. Smith, K.L. Miller, G. Salimi-Khorshidi, M. Webster, C.F. Beckmann, T.E. Nichols, J.D. Ramsey, and M.W. Woolrich, "Network modelling methods for FMRI," *Neuroimage*, vol. 54, no. 2, pp. 875–891, 2011.