Learning Group Differences in Brain Networks from EEG signals using Hotelling T^2

Parinthorn Manomaisaowapak and Jitkomut Songsiri

Department of Electrical Engineering, Faculty of Engineering

Chulalongkorn University

Abstract

This study aims to learn brain networks using the concept of Granger causality from EEG signals. The method relies on the use of vector autoregressive (VAR) models where each component of time series vector refers to signal from each EEG channel. The brain network is decoded by an estimated sparsity pattern of VAR matrix coefficients, which is statistically tested using Hotelling T^2 test. We consider data sets from two classes: control and patients with trauma brain injury (TBI) collected from Universiti Sain Malaysia. Our results show that XXX.

Contents

1 Introduction

The term brain connectivity refers to a pattern of links across brain regions that indicates causal interaction or statistical dependencies[Spo07], [RS10]. There are three types of brain connectivity definition. The first definition is *Structural brain connectivity* which refers to the links that anatomically connected between brain regions. The second definition is *functional connectivity* which are the links between brain regions are defined by statistical dependencies such as correlation or covariace[Spo07]. But statistical dependencies cannot be interpreted alone in general because statistical measures such as correlations cannot be interpreted as causality. The causal interactions are described in the last type of connectivity, the *effective brain connectivity*. The effective brain connectivity is the description of causal interaction between brain regions. There are many measures of effective brain connectivity and one of them is the use of Granger causality test [BL14] which can be displayed as a matrix that entries (i, j) represents measure of connectivity strength between brain region i and region j denoted as GC matrix with $\mathcal{F}_{ij}$ as elements.

The detection of Granger causality based brain network differences computed on vector autoregressive model between healthy group and TBI (Traumatic Brain Injury) group is our main interest because the classification of road accident patients, that they have undetected long-term brain injuries or not, are crucial. In this project, the connectivity matrices are computed from each patient in each group individually and compute statistical measure such as the average value of brain connectivity measures as the representation of the whole group to perform hypothesis test between group whether there are statistical significant different

This project aims to learn brain network differences from brain signals (that can be applied to both EEG and fMRI signals) collected from two groups of subjects. The methodology will be described in Figure 1. Firstly, the data are split into two groups, patient group and control group. In each group, we will estimate GC matrix of each trial individually, the group GC matrix will be obtained by averaging all GC matrices over all trials in the group and the difference are tested by a statistical test. We will use Multivariate Granger Causality toolbox (MVGC) [BL14] to compute Granger causality matrix as a measure of effective brain connectivity denoted as GC matrix. The GC matrix is based on vector autoregressive model because of model simplicity.

2 Background

This section contains two topics. First topic describes the basics in Granger causality computation and the second topic describes the statistical test which includes significant test for Granger causality matrix computation.

2.1 Granger causality estimation

Granger causality is a concept that test if the past of one time series can help to predict another time series in sense of reducing the residual variance of the predicted time series. In this project, we will compute Granger causality based on vector autoregressive model.

Vector autoregressive model order p is defined as

$$y(t) = \sum_{k=1}^{p} A_k y(t-k) + e(t) \tag{1}$$

where $y(t), e(t) \in \mathbf{R}^n, A_k \in \mathbf{R}^{n \times n}$ AR model can be expressed in state-space representation as

$$x(t+1) = \begin{bmatrix} A_1 & A_2 & \dots & A_{p-1} & A_p \\ I & 0 & \dots & 0 & 0 \\ \vdots & \ddots & & \vdots & \vdots \\ 0 & 0 & \ddots & 0 & 0 \\ 0 & 0 & \dots & I & 0 \end{bmatrix} x(t) + \begin{bmatrix} e(t+1) \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix} \tag{2}$$

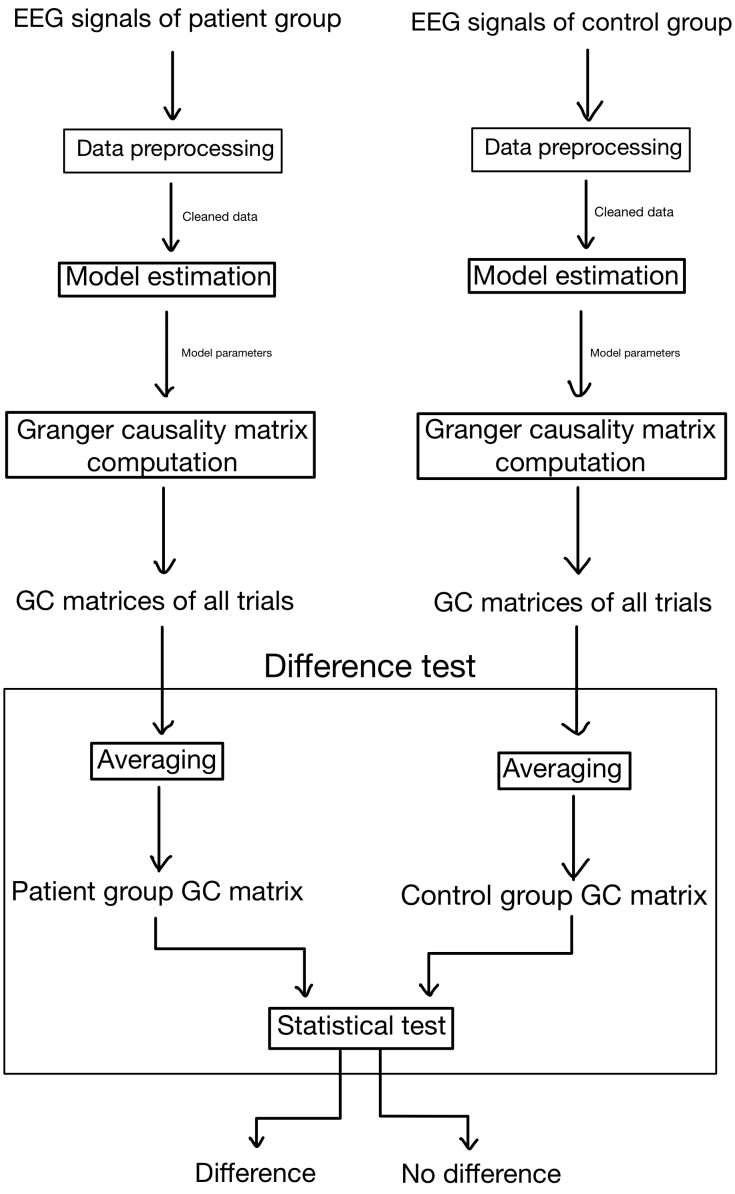


Figure 1: Statistical framework for learning brain network differences.

with $x(t) = \begin{bmatrix} y(t-1)^T & y(t-2)^T & \dots & y(t-p)^T \end{bmatrix}^T$, and the output equation is

$$y(t) = \begin{bmatrix} I & 0 & \cdots & 0 & 0 \end{bmatrix} x(t), \tag{3}$$

VAR model parameter can be estimated by ordinary least square methods or solve via Yule-Walker equation [BJ76].

Next, we will introduce the concept of Granger causality test. Let us consider a multivariate AR(1) process $y_i(t), y_j(t)$, note that both are vector. We want to investigate if $y_i(t)$ is depended only in its own past value, not from past of $y_j(t)$. The full fitted VAR model is

$$\hat{y}_i(t) = A_{ii} y_i(t-1) + A_{ij} y_j(t-1)$$

$$\hat{y}_j(t) = A_{ji} y_i(t-1) + A_{jj} y_j(t-1)$$

which can be expressed as

$$\hat{y}(t) = \begin{bmatrix} A_{ii} & A_{ij} \\ A_{ji} & A_{jj} \end{bmatrix} y(t-1). \tag{4}$$

Note that $y(t) = \begin{bmatrix} y_i(t)^T & y_j(t)^T \end{bmatrix}^T$. The model that remove the testing parameter that represent the connection between y_i and y_j is called the reduced model. In this scenario, the reduced model is

$$\hat{y}^R(t) = \begin{bmatrix} A_{ii}^R & 0 \\ A_{ji}^R & A_{jj}^R \end{bmatrix} y^R(t-1). \tag{5}$$

One can understand that more complex model has more flexibility, *i.e.*, more parameters to be determined than a simpler model. In this case, the full model is more complex than the reduced model therefore, the residual variance of full model is smaller than the reduced model. In other words, the full model is expected to have more model quality than the reduced model. However, if the residual variance of reduced model is close to that of the full model, this can be inferred that the removed parameter is not significant. As in (5), the parameter A_{ij} implies direct impact from the past of y_j to current y_i. If A_{ij} is removed and model quality does not reduce, then the past of y_j does not have direct effect on y_i. On the contrary, if the model quality is reduced on reduced model. Then y_j should have direct effect on y_i but to answer how the strong the effect is, at this point, the concept of Granger causality arises as the measure of the strength. In multivariate Granger causality, the residual variance will be changed into the measure that reflects how large the residual covariance matrix is. In [BBS10], they used the concept of generalized variance which is determinant of covariance matrix and denoted the uses of total variance, i.e. trace of covariance matrix.

The Granger causality measure is defined as

$$\mathcal{F}_{ij} = \log \frac{\det \Sigma_{ii}^R}{\det \Sigma_{ii}}. \tag{6}$$

Granger causality can be tested by a log ratio of the generalized variance which represents model quality of the reduced model compared to the full model. this measure is, in general, defined by (6) which is multivariate version of Granger causality with residual covariance matrix of reduced model Σ_{ii}^R and the residual covariance matrix of full model Σ_{ii}. In this case, both y_i, y_j are vector of time series, and it has physical meaning as a test of multiple time series to another multiple time series. Intuitively, if past of y_j can explain y_i then the generalized variance of full mode will be less than the reduced model. The value of $\mathcal{F}_{ij}$ will be nonzero in this case. Conversely, if past of y_j does not help to predict y_i, the generalized variance of two model should be equal and leads (6) to zero. If y_i, y_j represent as brain signals of region i, j, $\mathcal{F}_{ij}$ can be interpreted as connectivity strength from region j to region i.

Brain signals such as EEG or fMRI, can be applied with Granger causality test to find the causal interaction between regions in the brain. The strength of Granger causality measure from brain's region j to region i is defined as $\mathcal{F}_{ij}$ which can be represented in a matrix that the entry i, j contains $\mathcal{F}_{ij}$. The matrix is called GC matrix where each entries are always greater than or equal to zero. However, In numerical computation, there will be no exact zeros which emphasizes the reason to perform significant test whether the value inside connectivity matrix is actually zero. In the next section, we will define significant test for Granger causality matrices and brain network difference.

2.2 Statistical test

This study involves two types of tests.

Brain network test

The first test is to examine whether the GC causality is significantly zero or not. The null and alternative hypotheses are

$$\begin{aligned} H_0 &: \mathcal{F}_{ij} = 0 \\ H_1 &: \mathcal{F}_{ij} \neq 0. \end{aligned} \tag{7}$$

The result of this test is an estimated pattern of brain network of data in one trial. We could repeat this test when we have more trials of data (and they are not the same.) If we have data from two groups, we will obtain brain networks of both groups; tested separately.

Network difference test

Another objective of this study is to learn brain network difference through data from two groups. Suppose we have multiple trials of data from two groups. We can think of estimated matrices from those trials as samples and to test whether the two groups have different GCs, we can define the brain network difference as the inequality of GC matrix population mean of two groups. The significant test for mean differences is used the hypothesis test as follows,

$$\begin{aligned} H_0 &: \mu_1 = \mu_2 \\ H_1 &: \mu_1 \neq \mu_2. \end{aligned} \tag{8}$$

where μ_i denotes population mean of the ith group.

The null hypothesis will be rejected if the p-value of the test statistics of both tests is less than a significant level (set as 0.05). In scalar version, t-test can be used to compare the mean by normality assumption of sample mean. In multivariate case, the sample mean is vector that elements are not necessary independent. Therefore the t-test cannot be performed element-wise. Hotelling's T-squared test will be performed to determine the differences. Both significant test will be described again in the following section.

3 Methodology

The brain network test and the brain network difference tests are the two main goals of our study. The methodology based on statistical framework can be explained in four steps and represented in Figure 2.

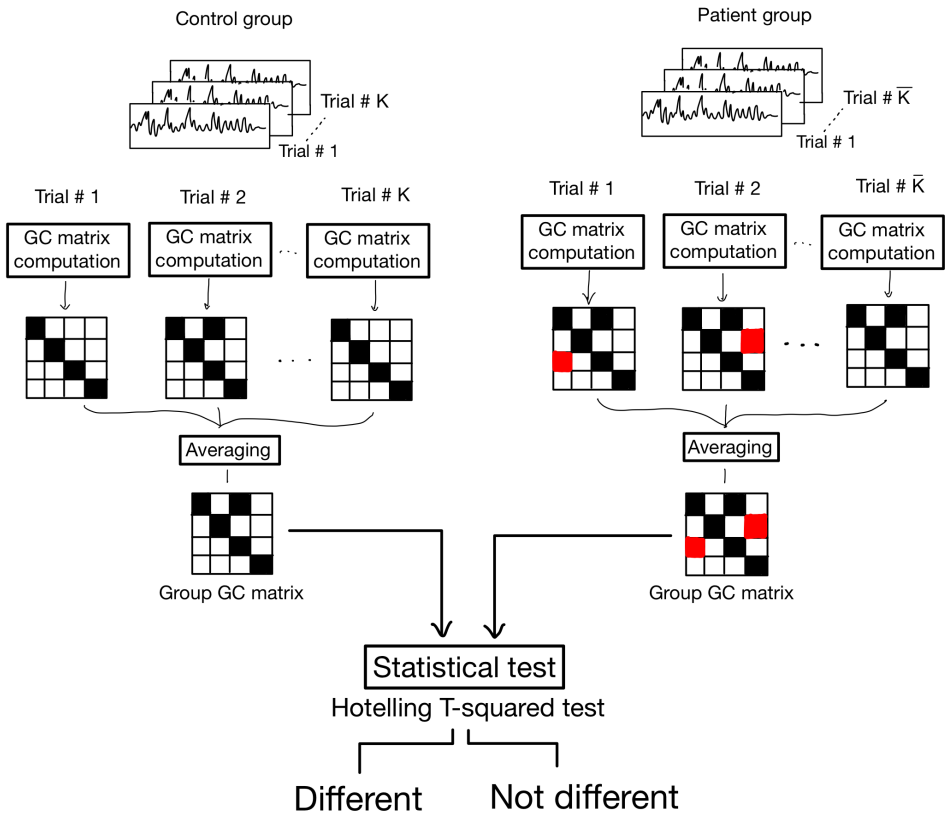


Figure 2: Methodology of learning brain network difference.

1. Data preparation: We describe how the data were collected and how to augment or arrange them for the test.

2. Model estimation: This section involves model order selection and VAR coefficient estimation which includes two methods of estimation, ordinary least square (OLS) and solve via Yule-walker equation [BJ76].

3. GC matrix computation: This section explains how the Granger causality matrices are estimated for all available data.

4. Group difference test: The statistical test in (19) will be performed in this step. This step will required asymptotic normality of sample mean, the testing samples must be large enough.

We will used MVGC toolbox [BL14] to implement in step 2-3.

3.1 Data preparation

The use of central limit theorem to assume normality of averaged GC matrices emphasizes that the amount of GC matrices should be large enough. One way to increase number of GC matrices is to split the brain data of one subject into multiple trials. However, VAR model has number of parameter to be estimated as $NVAR = n^2p$, where n is number of EEG channels, p is time lag. For example, in $p = 3$ will be $63 \times 63 \times 3 \approx 12,000$. By using rule of thumb, data points should be 10 times more than the number of parameters by the reason that more data points in model estimation can improve model quality. We will use $120,000$ data points to estimate parameter which resulting in 2 trials per subject as the maximum because one subject has approximately $300,000$ data points.

However, in real data sets, there are multiple highly correlated channels in many subjects. After filtering those channels out by removing highly correlated channels that have Pearson's correlation coefficient above 0.9. Resulting in the number of channels were reduced into 22 channels and the parameters to be estimated was reduced into $1,000$ parameters therefore, one subject can be splitted into 29 trials.

3.2 Model estimation

The estimation is based on assumption that the EEG time series are wide-sense stationary, the dynamic matrix in (2) must be stable.

Model order selection

Model order is selected by AIC, BIC value described by (9) and (10), consecutively

$$\mathrm{AIC} = -2\mathcal{L} + 2k \tag{9}$$

$$\mathrm{BIC} = -2\mathcal{L} + k \log N \tag{10}$$

where $\mathcal{L}$ is log-likelihood function of VAR(p) process, k is number of parameters to be estimated, in this case $k = n^2p$, and N is number of all observation. In the MVGC's source code, AIC implementation is based on [MT98] that can be used to perform model selection for small sample size. The log-likelihood function is implemented in the toolbox with maximum log-likelihood function as follow.

$$\mathrm{AIC} = -2\mathcal{L} + 2k\frac{N}{N-k-1} \tag{11}$$

$$\mathcal{L} = -\frac{N}{2}\log\det\hat{\Sigma}, \tag{12}$$

where $\hat{\Sigma} = \frac{1}{N-1}ee^T$ is unbiased estimator of residual covariance matrix, $e = y - \hat{y}$.

In this project, the amount of data are not sufficient to select higher order model, due to the trade-off between parameter estimation and the usage of central limit theorem in statistical test. So, the order candidates are $p = 1, 2, 3$.

VAR coefficient estimation

There are two main methods to estimate the coefficients,

1. **Ordinary least square**

   Ordinary least square is a solution of the overdetermine system. In this case, the linear system is

$$\begin{bmatrix} y(p+1) & y(p+2) & \dots & y(N) \end{bmatrix} = \begin{bmatrix} A_1 & \dots & A_p \end{bmatrix} \begin{bmatrix} y(p) & y(p+1) & \dots & y(N-1) \\ \vdots & \vdots & \dots & \vdots \\ y(2) & y(3) & \dots & y(N-p+1) \\ y(1) & y(2) & \dots & y(N-p) \end{bmatrix} \tag{13}$$

   This is in the form $Y = \beta X$, $Y \in \mathbb{R}^{n\times(N-p)}$, $X \in \mathbb{R}^{pn\times(N-p)}$, $A_i \in \mathbb{R}^{n\times n}$ where n is the number of channel and N is number of timepoints, p is model order.

   Then the least square optimization formulation is

$$\underset{\beta}{\text{minimize}} \quad \|Y - \beta X\|_F^2 \tag{14}$$

   The least square solution $\hat{\beta}$ can be solved analytically by solving the normal equation.

$$\hat{\beta}(XX^T) = YX^T \tag{15}$$

   In the MVGC toolbox, ordinary least square method is implemented by MATLAB function `mrdivide` that simply computes least square solution via QR factorization. In this case, the regressor matrices in (15) are mostly rank deficient, which caused by highly correlated EEG channel, hence, there will be infinitely many exact solution. However, those channel must be excluded.

2. **Solve via Yule-Walker equation**

$$\begin{bmatrix} \Gamma(1) & \Gamma(2) & \dots & \Gamma(p) \end{bmatrix} = \begin{bmatrix} A_1 & A_2 & \dots & A_p \end{bmatrix} \begin{bmatrix} \Gamma(0) & \Gamma(1) & \dots & \Gamma(p-1) \\ \Gamma(-1) & \Gamma(0) & \dots & \Gamma(p-2) \\ \vdots & \dots & \ddots & \vdots \\ \Gamma(-p+1) & \Gamma(-p+2) & \dots & \Gamma(0) \end{bmatrix} \tag{16}$$

   Yule walker equation [BJ76] described in (16) is a system of linear equation that came from taking autocovariance of (1) with multiple lags . where $\Gamma(k)$ is autocovariance matrix that can be estimated by its unbiased sample autocovariance. If the datapoints are large enough, by the law of large number, the XX^T in (15) will converge to autocovariance matrix in (16). The autocovariance matrix in (16) is in Toeplitz form that can be solved efficiently by LWR (Levinson Wiggins Robinson) algorithm which has been proven that this algorithm will yield stable VAR coefficients [Whi63].

3.3 GC matrix computation

GC matrix is calculated by estimating full model and reduced model from the data set directly while MVGC toolbox recommended to compute GC matrix from autocovariance sequence to increase computation accuracy in frequency-domain GC matrix[BL14] but only time-domain GC matrix is used in this project. The significant test for zero patterns in GC matrix can be determined by hypothesis test defined in (7).$\mathcal{F}_{ij}$ has asymptotic distribution as chi-squared distribution[BL14],[Gew82]

$$(N-p)\mathcal{F}_{ij} \sim \chi^2_{p(n_i+n_j)} \tag{17}$$

where N, p, n_i, n_j denotes sample size, lags, dimension of y_i, y_j respectively. The null hypothesis will be rejected if p-value of the GC measure is below 0.05 which p-value is probability that the null hypothesis is true. However, we will not perform significant test for zero patterns due to the hypothesis test on group difference.

3.4 Group difference test

The group differences of brain network is determined by equality testing of element-wise population mean of GC matrices between healthy group and TBI group. The sample mean of $n \times n$ GC matrices are removed diagonal entries due to self-causal inference is meaningless. After that, the removed-diagonal GC matrices are vectorized into vector mean with dimension $n^2 - n$ in order to use the Hotelling T-squared test that is the vector mean equality test. The normality of vector mean is assumed by central limit theorem. The vector mean of two groups, $\bar{X}_1, \bar{X}_2$, are $n^2 - n$-variate normal distributed. The two samples Hotelling's T-squared is used to test whether the vector mean of two samples are equal. The test statistics is defined as

$$T^2 = (\bar{X}_1 - \bar{X}_2)^T \left(\frac{S_1}{N_1} + \frac{S_2}{N_2} \right)^{-1} (\bar{X}_1 - \bar{X}_2) \tag{18}$$

where $\bar{X}_i$ is sample vector mean of X_i, S_i is unbiased sample covariance matrix of X_i respectively. The test requires that both X_1, X_2 are drawn from the normal distribution with common covariance matrix and test statistics T^2 has the distribution:

$$T^2 \sim \frac{pv}{v - p + 1} F_{p, v-p+1} \tag{19}$$

where p is dimension of vector, $v = N_1 + N_2 - 2$ is degree of freedom [Har01].

In practical uses, two samples are mostly drawn from assumed normal distribution but with *unequal* covariance matrices. It is called the multivariate Behrens-Fisher problem. There are many solutions; one of them is to estimate distribution of T^2 by modifying degree of freedom v in (19). In this project, we use the degree of freedom in[KY04] as

$$v = \frac{p + p^2}{A_1 + A_2}$$

where

$$A_i = \frac{\mathbf{tr}[(\tilde{S}_i S_p^{-1})^2] + [\mathbf{tr}(\tilde{S}_i S_p^{-1})]^2}{N_i}, \quad \tilde{S}_i = \frac{S_i}{N_i}, \quad S_p = \tilde{S}_1 + \tilde{S}_2.$$

Hotelling T-squared test can be explained intuitively as the multivariate version of student's t statistics that used to compare mean in scalar version. But in multivariate sense, the vector mean cannot be compared element-wise because there are many components which are not necessarily independent to each other such as normal distribution with non-diagonal covariance matrix. The T^2 brings mean vector into scalar representation as quadratic loss function. If value of T^2 is low, the vector mean of two samples are more likely to equal.

4 Data description

The EEG datasets are achieved from USM (University Sains Malaysia). There are two groups of data, TBI group and Healthy group. Each groups has 7 subjects. Each subjects performed N-back test, emotion and real-time task. The EEG data are measured before and after those tasks with 2 different condition, eyes-open and eyes-closed. For example, data with label **after REC** is the EEG data that measured After given tasks during Resting state with Eyes-Closed.

4.1 Electrode placement system

Electrode placement systems are standard methods to measure EEG signal from scalp. The example are 10-20 system and 10-10 system. The number denotes distance in % from front to back, in this case there are 10% and 20% distance between electrodes. The 10-10 system has more spatial resolution [Soc16].

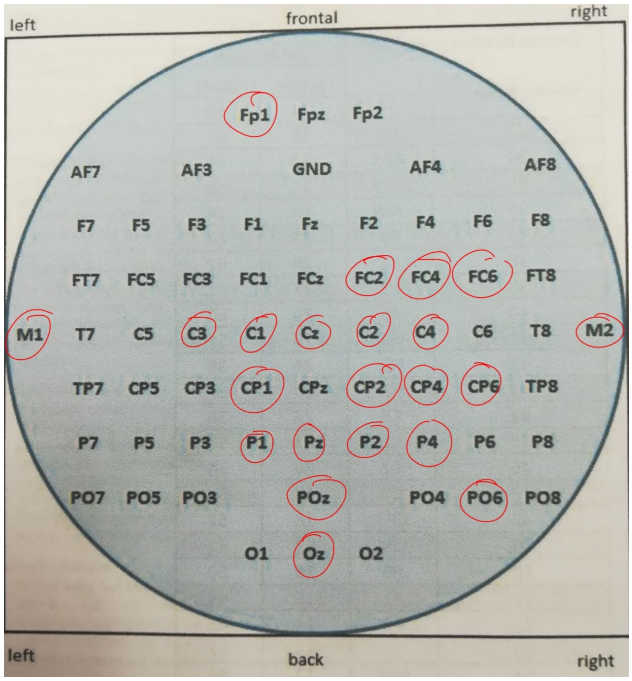


Figure 3: 10-10 system electrode placement system that is used to measure EEG signal. The red circles indicate the selected channels used in our experiment.

4.2 Measurement

The data was measured by 64 channels 10-10 EEG electrode placement system with sampling rate 1000 Hz. Only channel 32 (EOG channel) has to be removed before analysis because it is not connected. All channel's EEG signal was a voltage difference between the EEG electrode and reference electrode, which is channel CPz. The Ground (GND) channel is not presented in data file. The data arrangement is reported in Figure 4.

4.3 Data problem

The abnormality of data are investigated by comparing the fitting. The hypothesis is the data that contains abnormality such as spikes in the signal should be detected when fitting the model by MSE value as described in Figure 5.(a), 5.(b) which the MSE seems to be very high at those trials. The example of spikes is in Figure 5.(c). Another problem on data is rank-deficient regressor matrix in (15) causing non-unique VAR coefficients due to colinearity problem, *i.e.*, signals from some channels are highly correlated. One way to solve the problem is to remove the highly channel correlated channel which measured by sample correlation functions. However, different subjects have different sets of highly correlated channels position and amount. The possible solution is to detect channels that highly correlate in most of the data, then remove those channels in the same index of all data. After filtered out, number of trials are reported in Figure 4 and the selected channels are circled channels in Figure 3.

5 Preliminary results

5.1 Model estimation

Model order is generally selected by lowest BIC score, because it will, in general, return simpler model than AIC. But in this scenario, the BIC score was calculated using fitness term as log-likelihood which involved $\log\det\Sigma$ that has large value compared to the complexity term in the order of 1000 times larger. Resulting in continuously decreased of BIC score, this should be the sign of overfitting the data. The order was selected in the sense of major decrease in MSE of fitting which is shown in Figure 6.

The fitting results in Figure 5.1 are randomly selected from each case.

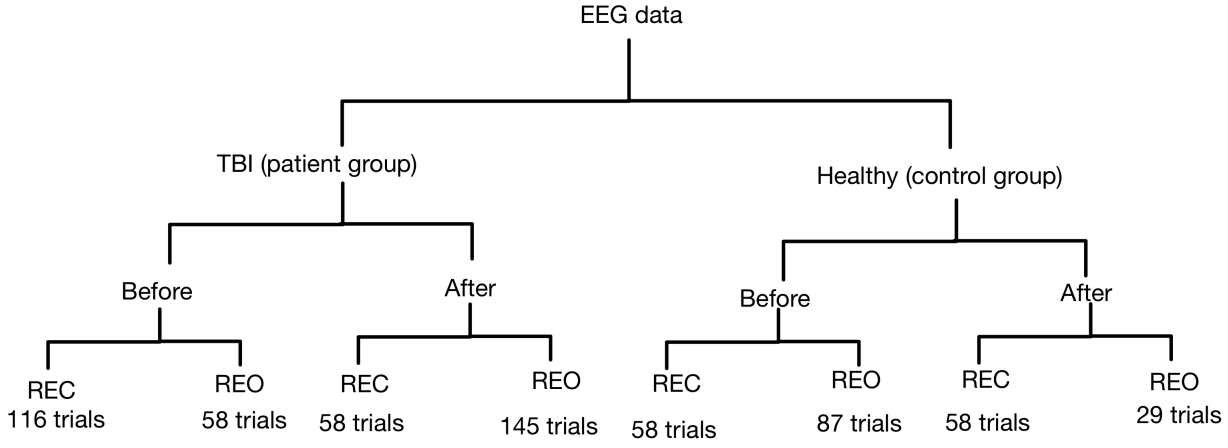


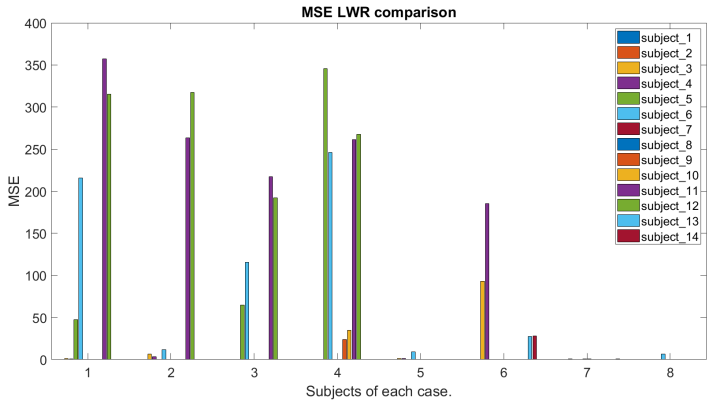
Figure 4: EEG data arrangement in TBI and healthy group with number of trials in each case. The number of trials are counted when the data are not filtered out.

5.2 GC matrix computation

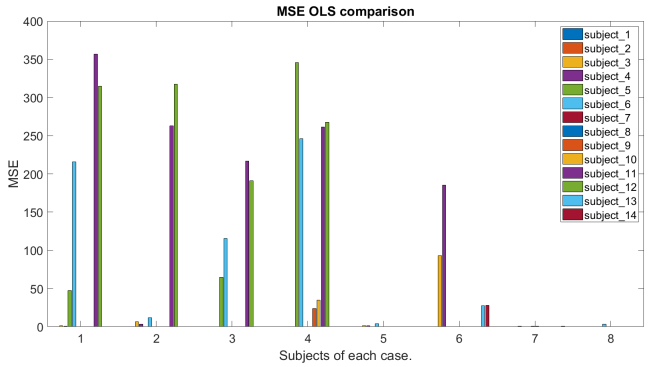
The GC measure is computed by fitting full and reduced model directly. In this scenario, the causal inference need to be from one region to another region, not group of regions to another group of regions. So, the univariate Granger causality is used and (6) is reduced into scalar equation. The Granger test was performed in each pair of the channels and compute p-value. In each case of a group, there are multiple of individually computed GC matrices or subject-level GC matrices. We used average value of GC matrix over a case to represent as the GC matrix of the case or case-level GC matrix. To compare subject-level GC matrix with the case-level GC matrix, we used significant test to find and to compare the sparsity pattern of both. In subject-level GC matrix of a case, we will show sparsity pattern by the indices of sparse elements that occurred frequently through the case. In significant test for sparsity pattern, we used element-wise **Chi-square test**[BL14] for subject-level GC matrix and element-wise **one sample t-test** for case-level GC matrix to test whether the element of those GC matrices is zero. To be more clear, subject-level GC matrix will be test if the value is zero through Chi-squared test and case-level GC matrix will be test whether the averaged value of the elements over the case is zero through zero mean test by conducting t test ($\mu = 0$ vs $\mu > 0$).

The cases are denoted by an acronym such as **HAC** to be **Healthy After REC**. The results of case-level GC matrix in each case of two groups are shown in Figure 8 with the selected channels. By inspection, channel *Fp1* seems to interact with almost all channel in healthy group which is not presented in any of subject in patient group. In patient group, the entries that have high GC measure seems to cluster in the top left region or first 10 channels except for case PBO. In almost all cases, the GC matrices between REC, REO is similar when **After**, **Before** is the same.

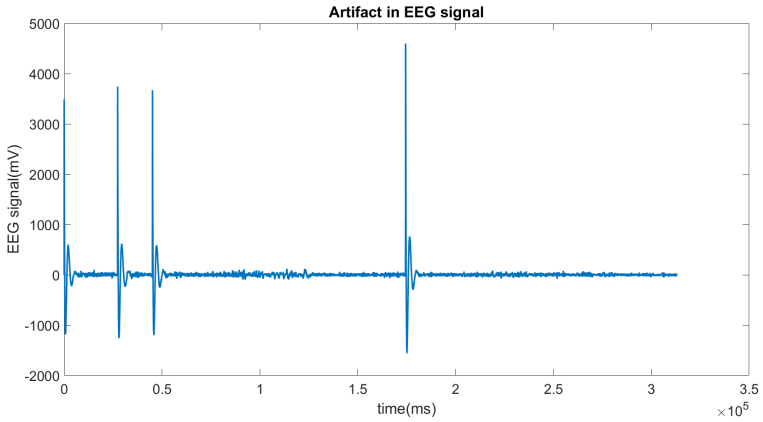
However, the problem in t-test arises from similarity of GC matrix between trials that were cut from same subject causing standard deviation to be low causing the t score to be large. After that, the value of large t score will lower p-value of the test. We solved the problem by setting the significant level to 10^{-9}. For subject-level GC matrix sparsity pattern we set repetitive of zeros elements at same index across the trials in a case to have atleast 10% of all trials. The sparsity pattern of case-level and subject-level GC matrices are different, especially in **PBC, PAO** cases that have large different of their sparsity pattern. But after we applied to sample of subject-level GC matrix in Figure 10, the pattern of case-level and subject-level GC matrices are similar in some part of the matrix. Such as in **HAC**, case-level GC matrix is obviously similar to the subject-level but not identical. This maybe because we averaged the case-level GC matrices through the case, it should contain some part of subject-level. There are similar and non-similar matrices in Figure 10 because the sample of subject-level GC matrix may not have strength of connectivity enough for clearly view in the case-level GC matrix compared to another sample in the same case.



(a) MSE plot of Yule-Walker method by LWR algorithm in all subjects.

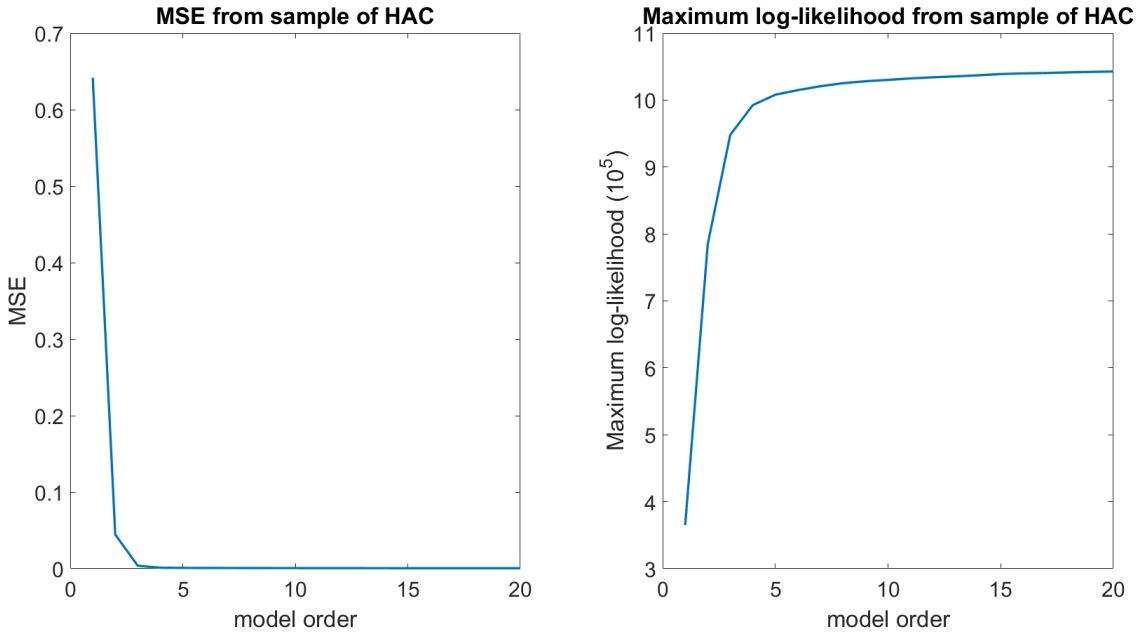


(b) MSE of OLS method in all subjects.

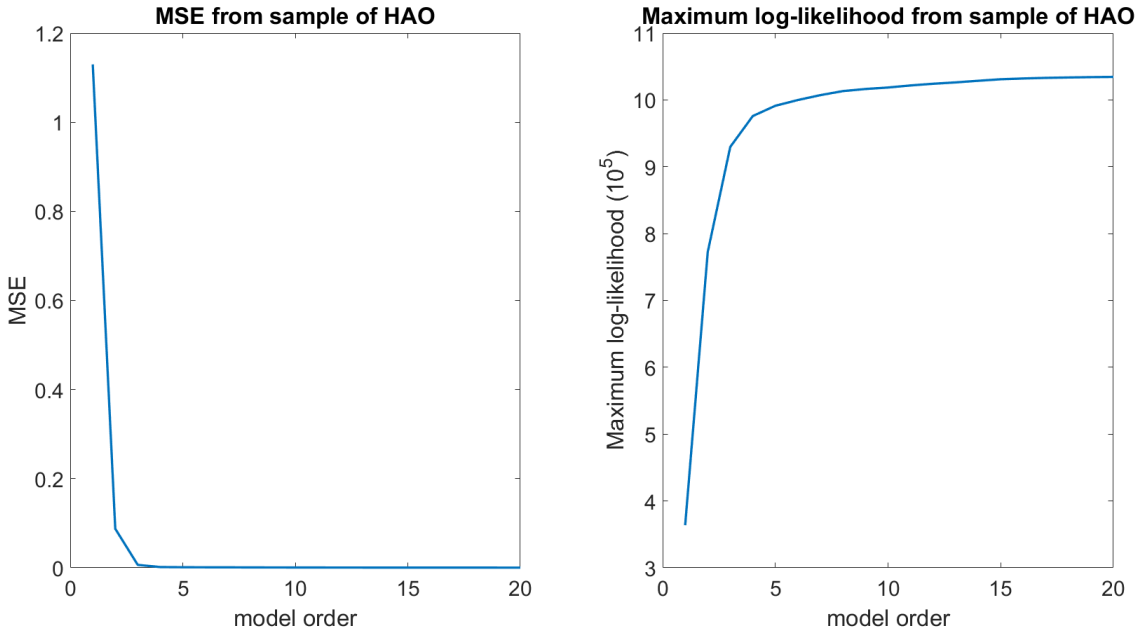


(c) Example plot of abnormal data.

Figure 5: Data abnormality detection to be filtered out.



(a)



(b)

Figure 6: Example of MSE and maximum log-likelihood plot of each model order.

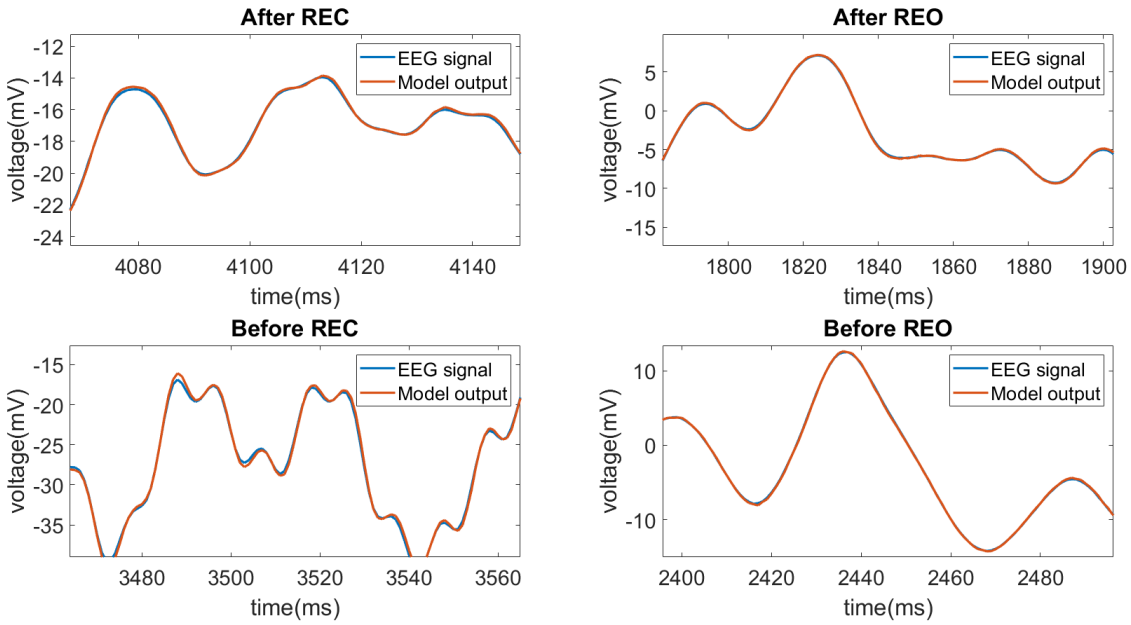


Figure 7: Fitting results.

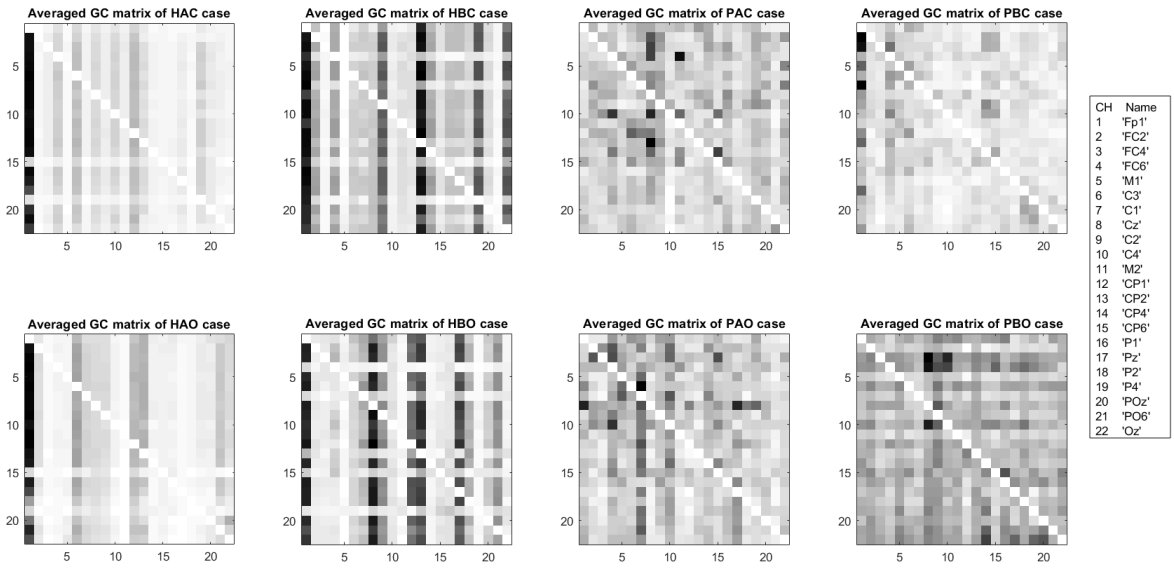
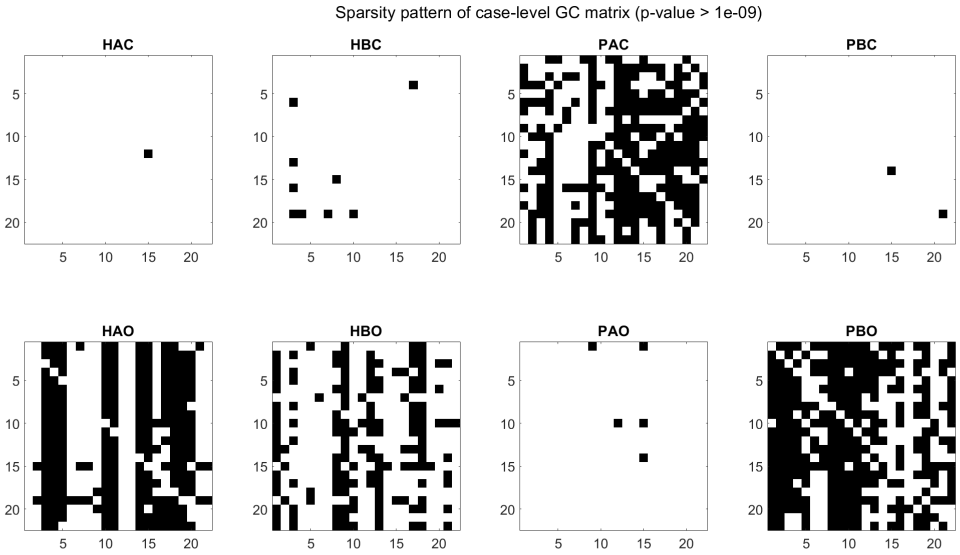
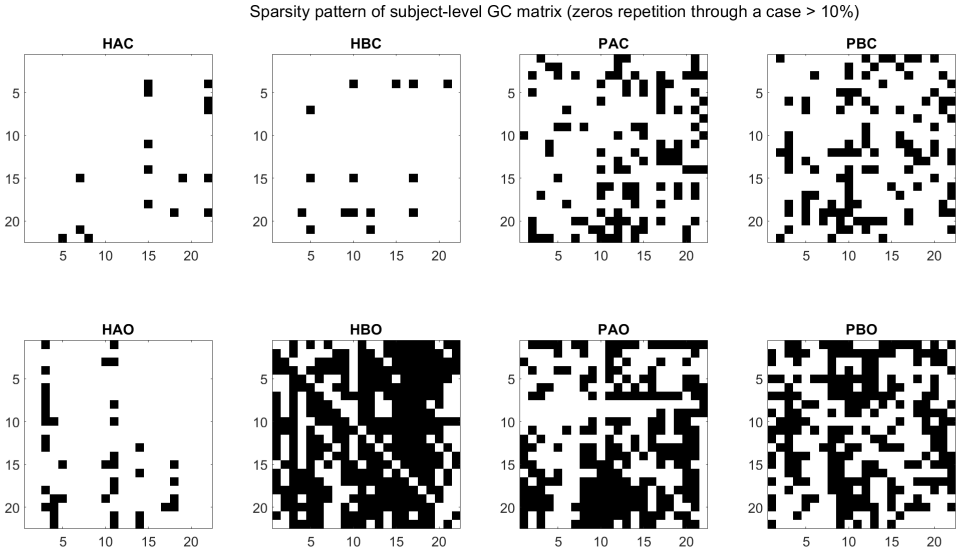


Figure 8: Average value of GC matrix over trials in the same cases. H/P denotes Healthy, Patient, A/B denotes After/Before, C/O denotes REC/REO.

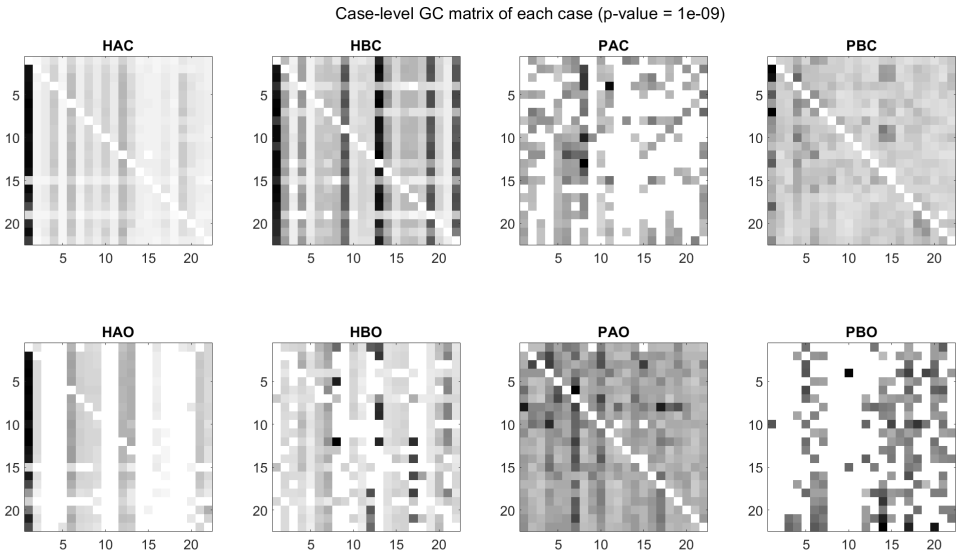


(a) Sparsity pattern indices of GC matrix computed from one sided t test with significant level 10^{-9}

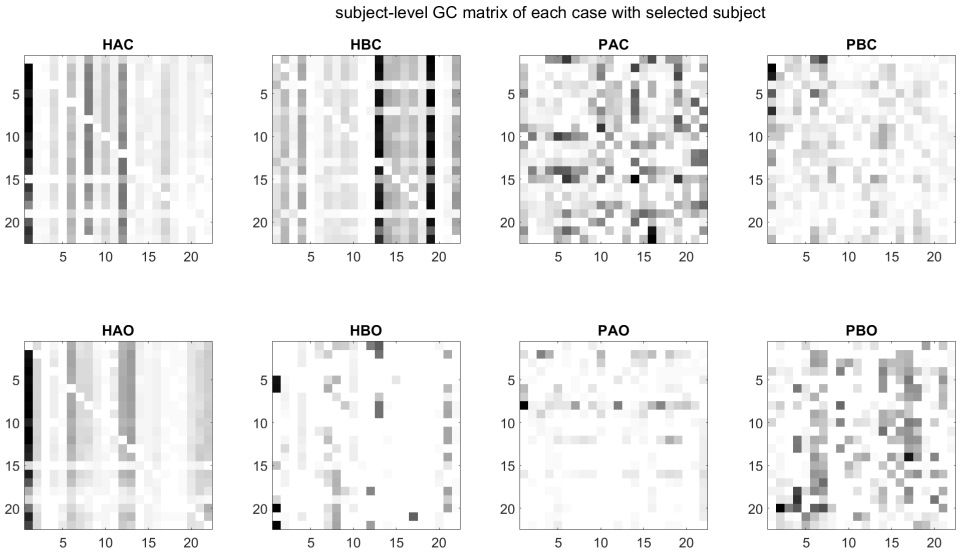


(b) Sparsity pattern indices of GC matrix calculated by selecting the indices ,that is set to zero by Chi squared test, that has repetition of zeros more than 10% through a case. Significant level is 0.05

Figure 9: Sparse pattern indices of GC matrix in each level. Black entries denote sparsity pattern



(a) Case-level GC matrices after hypothesis testing with significant level 10^{-9}.



(b) Subject-level GC matrices after hypothesis testing with sparse pattern.

Figure 10: GC matrix of each level after applied the sparse pattern.

5.3 Group difference test

The test was performed in two conditions,

1. Crossing test between TBI and Healthy with 29 GC matrix samples. This can measure number of true positive and false negative of the test. The test was conducted by pairing all vectorized GC matrix averaged with 29 trials in healthy group and vectorized GC matrix averaged with 29 trials in TBI group.

2. Test among the same group, between TBI and TBI or Healthy and Healthy with 29 GC matrix samples. This can measure number of true negative and false positive of the test. The test was conducted by pairing all vectorized GC matrix averaged with 29 trials and another vectorized GC matrix averaged with 29 trials in the same group.

In Hotelling T-squared test, the rejection of null hypothesis denotes positive results and accepting as negative results. The results of group differences test are in *Figure* 1. The true positive rate is 81.73 %, overall true negative rate is 48.19 %. True negative rate in TBI vs TBI is 66.67 % but in Healthy vs Healthy is 0 %. The test in the same class such as testing **after REO** between healthy group and TBI group, can detect the difference for 75%. The true negative rate is very low mainly because of the test among healthy group, the possible reasons are

1. The different class in healthy group have different brain networks.

2. The EEG signal in healthy group contained artifacts such as spikes.

3. The number of trials are not enough for statistical testing because it is the comparison of n^2 dimensional vector, where n is number of channel. In this scenario, the dimension of vectorized GC matrices is $22 \times 22 - 22 = 462$. The diagonal entries are subtracted. And the samples varies in ranges of 29-145 which may be not enough. In [KY04] denoted that the sample size should be atleast 4 times more than vector dimension.

Table 1: Group differences results. Positive results denote that the brain networks are different. Negative results denote that the brain networks are not different. N is a number of paired subjects in the test where data from one subject consist of 29 trials.

	$N = 104$ Healthy vs TBI	$N = 23$ Healthy vs Healthy	$N = 60$ TBI vs TBI
True positive	85	-	-
True negative	-	0	40
False positive	-	23	20
False negative	19	-	-

However, each trial in the same case as in figure 4 has similar connectivity matrix to each other causing their sample covariance matrix to be near singular. We used Tikhonov's regularization to the sample covariance matrices by adding λ scaled identity matrix to the covariance matrix to solve the issues. In the test, we selected the parameter λ to be 10^{-5}.

We conducted another experiment that controlled the comparison by comparing within the same case with two scenarios, different group to find true positive rate and the same group to find true negative rate. The results are shown in Figure 11 and yielded TPR to be 82.61% and TNR to be 56.52% which is improved from the previous in Figure 1. However, the regularization of Hotelling T-squared test will change the distribution of T-squared value. [LAP+16] proposed a modified test for the regularization. The result of this method is shown in Figure 12 which yielded TPR and TNR to be both approximate 83%. We also checked the distribution when varying regularization parameter as shown in Figure 13. The result indicated that the sensitivity of the test will decreased as regularization parameter increased.

Table 2: Difference test of group of each case result. n denotes number of comparison that categorized into different rows.

	$N = 23$ Different group	$N = 23$ Same group
True positive	19	-
True negative	-	13
False positive	-	10
False negative	4	-

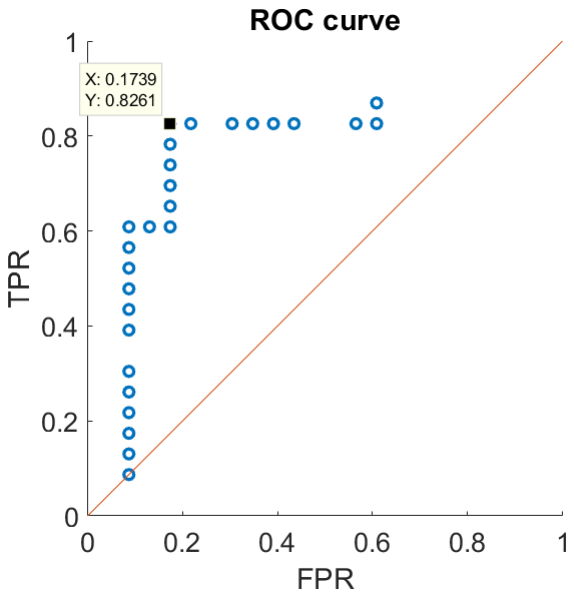
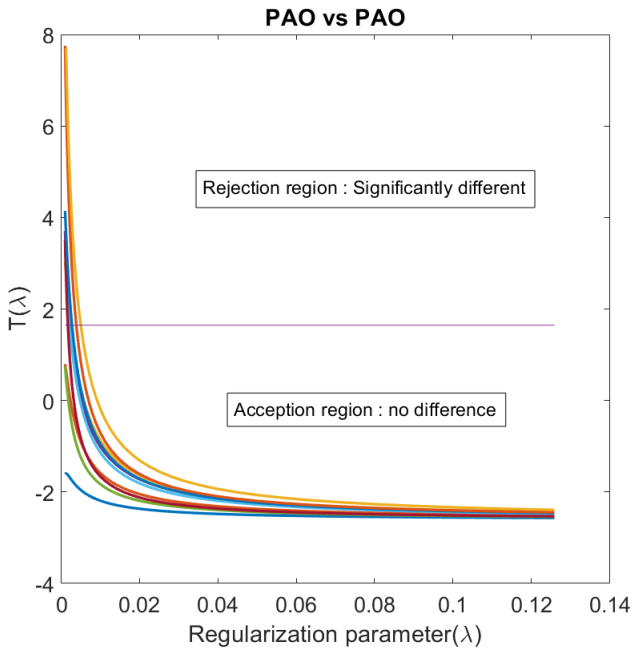
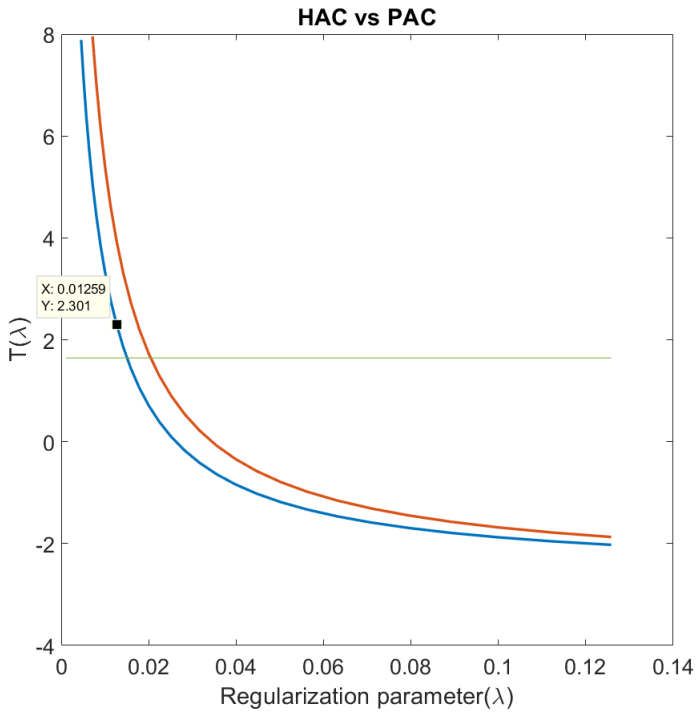


Figure 11: ROC curve of regularized Hotelling T squared test when varying regularization parameter from 10^{-3} to 0.13. The point that yield TPR to be 82.61%, FPR to be 17.39% is with $\lambda = 0.0126$



(a) The difference test between same group.



(b) The difference test between different group.

Figure 12: The horizontal line indicates the critical value for the test. Each graph denoted as the statistics value when vary the regularization parameter

References

[BBS10] Adam B. Barrett, Lionel Barnett, and Anil K. Seth. Multivariate Granger causality and generalized variance. *Physical Review E*, 81:041907, Apr 2010.

[BJ76] George.E.P. Box and Gwilym M. Jenkins. *Time Series Analysis: Forecasting and Control*. Holden-Day, 1976.

[BL14] Seth AK Barnett L. The MVGC multivariate Granger causality toolbox: A new approach to Granger-causal inference. *Journal of Neuroscience Methods*, 2 2014.

[Gew82] John Geweke. Measurement of linear dependence and feedback between multiple time series. *Journal of the American Statistical Association*, 77(378):304–313, 1982.

[Har01] Richard J. Harris. *A Primer of Multivariate Statistics*. Psychology Press, 3 edition, 2001.

[KY04] K. Krishnamoorthy and Jianqi Yu. Modified Nel and Van der Merwe test for the multivariate BehrensFisher problem. *Statistics and Probability Letters*, 66(2):161 – 169, 2004.

[LAP+16] Haoran Li, Alexander Aue, Debashis Paul, Jie Peng, and Pei Wang. An adaptable generalization of hotelling's t^2 test in high dimension. 09 2016.

[MT98] A.D.R. McQuarrie and C.L. Tsai. *Regression and Time Series Model Selection*. World Scientific, 1998.

[RS10] Mikail Rubinov and Olaf Sporns. Complex network measures of brain connectivity: Uses and interpretations. *NeuroImage*, 52(3):1059 – 1069, 2010. Computational Models of the Brain.

[Soc16] American Clinical Neurophysiology Society. Guideline 2: Guidelines for standard electrode position nomenclature. 2016.

[Spo07] O. Sporns. Brain connectivity. *Scholarpedia*, 2(10):4695, 2007. revision #91084.

[Whi63] P. Whittle. On the fitting of multivariate autoregressions, and the approximate canonical factorization of a spectral density matrix. *Biometrika*, 50(1/2):129–134, 1963.