Senior Project Proposal 2102490 Year 2019

Learning Granger causality for multivariate time series using state-space models

Anawat Nartkulpat ID 5930562521

Advisor: Assist. Prof. Jitkomut Songsiri

Department of Electrical Engineering, Faculty of Engineering

Chulalongkorn University

Contents

1 Introduction

Granger causality (GC) is one of powerful methods for determining causal relationships between variables in time series. For a time series $y(t) = [y_1(t)\ y_2(t)\ \cdots\ y_n(t)]^T$, we say that y_j Granger-causes y_i if the prediction of y_i is improved by using the past information of y_j given all other information of y_k where $k \neq j$. Quantitatively, a GC measure is defined by

$$F_{ij} = \log \frac{\Sigma_{ii}^R}{\Sigma_{ii}} \tag{1}$$

where Σ_{ii} is the variance of the prediction error of y_i given all past information and Σ_{ii}^R is the variance of the prediction error of y_i given other past information except y_j. There are many applications that concern about establishing connectivities between time series, so determining zero and non-zero causality is of great interest. In this context, the applications of the Granger causality can be found in many fields especially in neuroscience where the existences of connectivities between regions of a human brain were explored from brain signals such as EEG or fMRI [1].

The Granger causality can be applied to both linear and non-linear dynamical models, but linear model such as vector autoregressive (VAR) models are considered mostly. In VAR models, many properties of the GC measure was studied. It was shown that the GC measures of VAR models has an asymptotic mean-shifted χ^2-distribution [2]. This allows many powerful parametric methods to be utilized for further significance tests. Modelling time series with VAR model is simple but lacking of moving average (MA) terms make the model impractical to apply to some of the real systems. MA terms can be induced in many procedures such as data pre-processing, filtering or by observation process. So, a more general model that can be use to describe both VAR and MA models such as state-space models are what we are interested in.

While the generality of the state-space models allows us to estimate models for the data described by both VAR and MA models, and also the combination of them, denoted VARMA models, the statistical distributions of the GC measures of state space models are not well understood. The distribution of the GC measures for the state space models was shown, empirically, to be well-approximated by $\Gamma-$distribution [3], but no analytical proof is provided yet. This renders parametric methods less useful. A clustering method was proposed by fitting a Gaussian Mixture Model (GMM) to a vectorized matrices of GC measures [4]. In this method, the GC measures were sampled many times to construct a mixture Gaussian distribution and used the first two Gaussian components with the lowest means to determine a threshold for classifying zero entries. While this method required no knowledge about statistical property of the Granger causality, many samples of Granger causality matrix were required to meet the CLT assumption and the classification performance of non-zero causality could be deteriorated if the causality is weak.

Therefore, this project aims to develop a scheme for classifying zero and non-zero GC measures by using a non-parametric statistical method, namely, a permutation test, and compare the performance with the GMM method. We start by reviewing the Granger causality and linear models especially the state-space models and its equivalent models, which is the VARMA models. Then we develop a method and code for generating, arbitrarily, ground truth model and time series data for later experiments. Next, we set up experiments to study the classification performance of a permutation test. After that, an experiment for comparing the performance, the computational cost, and the required assumptions of the permutation test and the GMM method is conducted. Lastly, we discuss the results and conclude our method. We expect that this project will provide a scheme for classifying GC patterns based on the permutation test, MATLAB codes for executing the scheme numerically and the comparative results between our method and the GMM method.

2 Objectives

1. To develop a scheme for classifying the zero patterns of the Granger causality of state-space models using the permutation test.

2. To compare the performance of the permutation test with the Gaussian mixture models method in classification of zero and non-zero entries of Granger causality matrix obtained from state-space model.

3 Methodology

This section describes the method for learning GC patterns from time series data including the test of significance of GC by permutation test. The scheme for learning GC pattern is proposed in *Figure 1*.

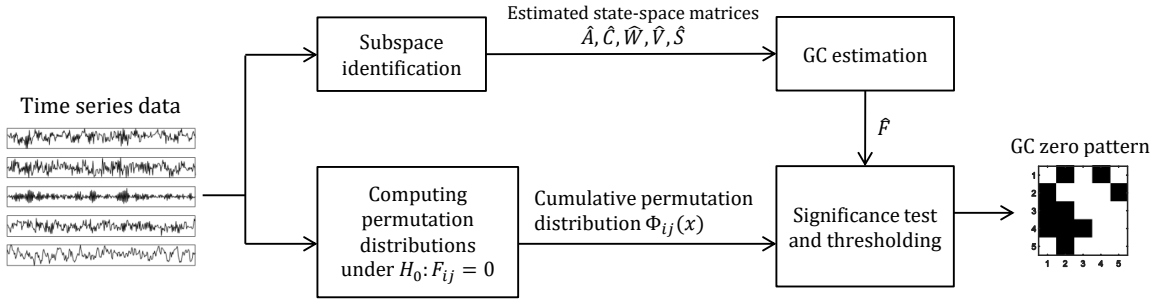

Figure 1: The scheme for learning Granger causality pattern from time series data.

From the given time series data, we start with estimating a state-space model in the form

$$x(t+1) = Ax(t) + w(t) \tag{2a}$$
$$y(t) = Cx(t) + v(t) \tag{2b}$$

where

$$\mathbf{E}\begin{bmatrix} w(t) \\ v(t) \end{bmatrix}\begin{bmatrix} w(t) \\ v(t) \end{bmatrix}^T = \begin{bmatrix} W & S \\ S^T & V \end{bmatrix}. \tag{3}$$

Then, a matrix of Granger causality measures ($F = [F_{ij}]$) is calculated from the parameters of the estimated state-space model. After that, we calculate permutation distributions (Φ_{ij}) of F_{ij} and use them to compute p-values for each F_{ij}. A significance level α is then selected to test the null hypothesis $H_0 : F_{ij} = 0$ for every entries of F, and, thus, the GC pattern is acquired.

3.1 Ground truth model

To examine the performance of the permutation method, we require state space models whose parameters and GC matrix are known. We refer to this model as the ground truth model. Since we are interested in the zero pattern of the GC matrix, the ground truth model should be generated arbitrarily with controlled sparsity of zero entries. The scheme for generating ground truth models is shown in *Figure 2*.

Firstly, we generate a stable VAR model of order p described by

$$y(t) = A_1 y(t-1) + A_2 y(t-2) + \cdots + A_p y(t-p) + w(t), \tag{4}$$

or by a transfer function

$$A^{-1}(z) = (I - A_1 z^{-1} - A_2 z^{-2} - \cdots - A_p z^{-p})^{-1}. \tag{5}$$

It has been shown that, for any VAR models [5],

$$(A_k)_{ij} = 0, \forall k = 1, \ldots, p \iff F_{ij} = 0. \tag{6}$$

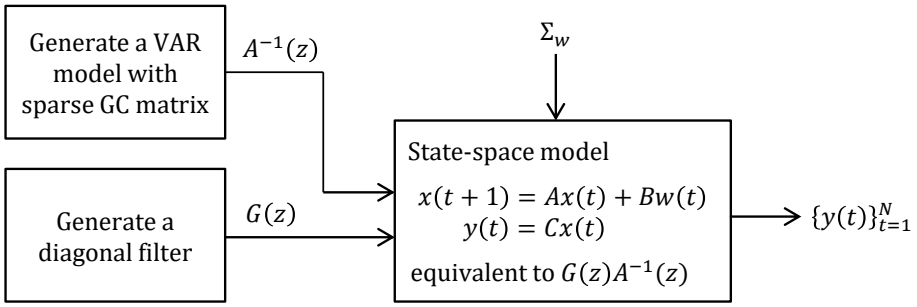

Figure 2: The scheme for generating ground truth state-space models and time series data.

So by generating $A_1, A_2, \ldots, A_p$ with common zero entries, we obtain a model whose GC matrix's sparsity can be chosen arbitrarily.

Next, in order to make the model more general, the VAR model can be filtered to introduce moving average terms. It has been proved that the GC matrix is invariant under any stable invertible diagonal filter that has minimum phase [6]. So we consider generating a diagonal filter

$$G(z) = \begin{bmatrix} \frac{p_1(z)}{q_1(z)} & 0 & \cdots & 0 \\ 0 & \frac{p_2(z)}{q_2(z)} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{p_n(z)}{q_n(z)} \end{bmatrix} \tag{7}$$

where $p_i(z), q_i(z) \neq 0$ are polynomials in z. To guarantee the stability and minimum phase, roots of $p_i(z)$ and $q_i(z)$ for all $i = 1, \ldots, n$ must be generated so that they lie inside a unit circle. Since $q_i(z) \neq 0$, the filter is invertible.

From the previous results, passing a VAR model under a stable invertible diagonal filter gives a new system described by a transfer function

$$H(z) = G(z)A^{-1}(z). \tag{8}$$

This transfer function can be realized into a state-space model

$$x(t+1) = Ax(t) + Bw(t), \tag{9a}$$
$$y(t) = Cx(t). \tag{9b}$$

This state-space model has the same GC pattern with the VAR model and is considered to be our ground truth model. Hence, we may conclude the generation of ground truth model into the following step.

1. Generate a random stable VAR model of order p with common zero entries in its parameters $A_1, A_2, \ldots, A_p$ to obtain the transfer function $A^{-1}(z)$.

2. Generate a random diagonal filter that is stable, invertible and has minimum phase. In this step we obtain the transfer function $G(z)$ of the filter.

3. Realize $G(z)A^{-1}(z)$ using `ss` command with `minimal` option in MATLAB to obtain a minimal state-space form. This state-space model is then used as the ground truth model.

Time series data can be generated directly from the state-space model using `lsim` command in MATLAB.

3.2 Subspace identification

We consider the estimation of a stochastic state-space model shown in (2) with (3) where $x \in \mathbf{R}^n$ and $y \in \mathbf{R}^m$. Suppose that the time series data $\{y(t)\}_{t=1}^N$ is observed. The parameter (A, C, W, V, S) can by obtained by a subspace identification method in the following steps [7]. Firstly, we project the future output (Y_f) onto the past output (Y_p) space to obtain the orthogonal projection

$$\mathcal{O}_i \triangleq Y_{i|2i-1}/Y_{0|i-1} = Y_f/Y_p \tag{10}$$

where $Y_{0|i-1} = Y_p$ is the observed data from $t = 0$ to $t = i - 1$ representing the past output and $Y_{i|2i-1} = Y_f$ is the observed data from $t = i$ to $t = 2i - 1$ representing the future output. Let n_o be the rank of $\mathcal{O}_i$. Then, by SVD decomposition,

$$\mathcal{O}_i = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} \Sigma_{n_o} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix} = U_1 \Sigma_{n_o} V_1^T. \tag{11}$$

Let Γ_i be an extended observability matrix. Then, from [8], we have

$$\mathcal{O}_i = \Gamma_i \hat{X}_i \tag{12}$$

where $\hat{X}_i$ is the estimated state. So, for some non-singular matrix T, we can obtain

$$\Gamma_i = U_1 \Sigma_{n_o}^{1/2} T \tag{13}$$

From this result, the estimated states is

$$\hat{X}_i = \Gamma_i^\dagger \mathcal{O}_i \tag{14}$$

where $\Gamma_i^\dagger$ denotes the pseudo-inverse of Γ_i. Next, the shifted state $\hat{X}_{i+1}$ can be computed by

$$\hat{X}_{i+1} = (\underline{\Gamma_i})^\dagger \mathcal{O}_{i-1} = (\underline{\Gamma_i})^\dagger (Y_{i+1|2i-1}/Y_{0|i}) \tag{15}$$

where $\underline{\Gamma_i}$ denotes Γ_i without the last row. Lastly, we form the equation

$$\begin{bmatrix} \hat{X}_{i+1} \\ Y_{i|i} \end{bmatrix} = \begin{bmatrix} A \\ C \end{bmatrix} \hat{X}_i + \begin{bmatrix} \rho_w \\ \rho_v \end{bmatrix}. \tag{16}$$

We can solve for A and C in least-square sense and obtain

$$\begin{bmatrix} \hat{A} \\ \hat{C} \end{bmatrix} = \begin{bmatrix} \hat{X}_{i+1} \\ Y_{i|i} \end{bmatrix} \hat{X}_i^\dagger$$

The covariances of $w(t)$ and $v(t)$ are then obtained by

$$\begin{bmatrix} \hat{W} & \hat{S} \\ \hat{S}^T & \hat{V} \end{bmatrix} = (1/j) \begin{bmatrix} \rho_w \\ \rho_v \end{bmatrix} \begin{bmatrix} \rho_w \\ \rho_v \end{bmatrix}^T. \tag{17}$$

3.3 Granger causality of state-space models

After the stochastic state-space model have been identified, we examine causal relationships between every pair of output y_i and y_j when $i \neq j$. The Granger causality from y_j to y_i is quantified by how prediction of the future of y_i can be improved by the past of y_j given all past information compared to without using the past of y_j. The measure of the Granger causality from y_j to y_i is then defined as [3]

$$F_{ij} = \log \frac{\Sigma_{ii}^R}{\Sigma_{ii}} \tag{18}$$

where Σ_{ii} and Σ_{ii}^R are the covariances of the prediction errors of the full model and the reduced model, respectively. Using more information generally gives better prediction, so it follows that $\Sigma_{ii} \leq \Sigma_{ii}^R$. Hence, we always have $F_{ij} \geq 0$.

To obtain Σ_{ii} and Σ_{ii}^R, we note that not using y_j is equivalent to removing y_j from (2). So we consider the full model described by (2) and the reduced model

$$x(t+1) = Ax(t) + w(t), \tag{19a}$$

$$y^R(t) = C^R x(t) + v^R(t) \tag{19b}$$

where y^R is obtained by removing y_j from y and C^R is obtained by removing the j^{th} row from C. It has been shown in [4] that, through Kalman filter, we have

$$\Sigma = \mathbf{cov}(y(t) - \hat{y}(t|t-1)) = CPC^T + V \tag{20}$$

where $\hat{y}(t|t-1))$ is the optimal estimator of $y(t)$ in mean squared error sense and P can be solved from the discrete-time algebraic Riccati equation (DARE):

$$P = APA^T - (APC^T + S)(CPC^T + V)^{-1}(CPA^T + S^T) + W. \tag{21}$$

Similarly, Σ^R can be obtained in the same manner by first solving (21) for P^R using C^R and V^R instead of C and V where V^R denotes V with its j^{th} row and column being removed. Then, we have

$$\Sigma^R = C^R P^R (C^R)^T + V^R. \tag{22}$$

By extracting the diagonal entries of Σ and Σ^R, the GC measures can be calculated by (18) for all i. A GC matrix is then defined by

$$F = \begin{bmatrix} F_{11} & F_{12} & \cdots & F_{1n} \\ F_{21} & F_{22} & \cdots & F_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ F_{n1} & F_{n2} & \cdots & F_{nn} \end{bmatrix}. \tag{23}$$

Since we are interest only the causal relationship between different variables, the diagonal entries of F are of no interest and left without computation.

3.4 Permutation test

After obtaining the estimated GC matrix $\hat{F}$, each entries in $\hat{F}$ must be classified into zero or non-zero. Since the estimation is never perfect, $\hat{F}$ is deviated from the true GC matrix F. A statistical test is then required for classification. So we state a null hypothesis as

$$H_0 : F_{ij} = 0. \tag{24}$$

Since $F_{ij} \geq 0$, our test is a one-tail test. It is not know what the distribution of F_{ij} under the true null hypothesis is. So, our approach is to use a non-parametric test, specifically, a permutation test.

In the permutation test, under the true null hypothesis, we construct distributions of the test statistics (F_{ij}), called permutation distributions, from the data we have [9], which is, in our case, the time series data. As the name suggests, the permutation distribution is constructed from the test statistics obtained by every possible permutations (or rearrangements) of the data. The permutation done on the data must be justified under the true null hypothesis, that is, the rearranging must not affect the null hypothesis.

Under the null hypothesis (24), y_i is not Granger-caused by y_j. It follows that rearranging the data in y_j has no effect on F_{ij} in this sense. Hence, more samples of F_{ij} can be acquired by estimating GC matrix again from different permutations. For the time series data of length N, there are overwhelming $N!$ possible ways to rearrange the data. To limit the number of permutation, we can permute chunks of the data instead by partitioning the data into many windows with the some length W. For example, see *Figure 3*. This gives $\lfloor \frac{N}{W} \rfloor !$ permutations if we discard the residual data from partitioning. If we only select some of the permutation randomly from all possible permutation, we call the test the Monte-Carlo permutation test [10].

For each permutation on y_j for $j = 1, \ldots, n$, we obtain samples of F_{ij} for $i = 1, \ldots, n$ denoted $\{\hat{F}_{ij}^{(k)}\}_{k=1}^P$ where P is the number of permutations. These samples are used to construct (cumulative) permutation distributions Φ_{ij} for all F_{ij}.

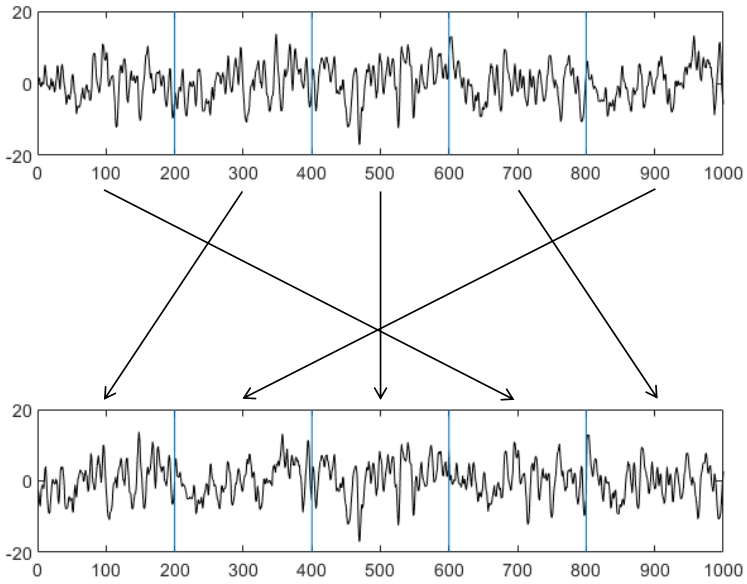

Figure 3: One of the possible permutations of time series data partitioned into 5 windows.

Given a significance level α, we can test the significance of $\hat{F}_{ij}$ to accept or reject the null hypothesis (24). The p-value is then calculated from the probability of F_{ij} being at least as extreme as $\hat{F}_{ij}$. Using the permutation distributions, one-tail p-value for any $\hat{F}_{ij}$ is computed by

$$p_{ij} = 1 - \Phi_{ij}(\hat{F}_{ij}) = \frac{\text{Number of elements in } \{k|\hat{F}_{ij}^{(k)} \geq \hat{F}_{ij}\}}{P}. \tag{25}$$

If $\hat{F}_{ij} > 0$, we can see that p_{ij} should be small since permuting the data generally neutralized the causality.

By choosing α for thresholding, we can simply classify that $F_{ij} \neq 0$ if $p_{ij} \leq \alpha$ (reject the null hypothesis) or classify that $F_{ij} = 0$ if $p_{ij} > \alpha$ (accept the null hypothesis). While this method is simple, testing all $\hat{F}_{ij}$ repeatedly with α as a threshold has some concern about the Type I error (the probability of incorrectly rejected the null hypothesis) since multiple hypotheses testing is being done.

A family-wise error rate (FWER) is the probability of having one or more Type I errors. Suppose that $F \in \mathbf{R}^{n\times n}$, it follows that we have a family of $n^2 - n$ hypotheses to be tested (excluding the diagonal). So, if we test each hypothesis using α as a threshold, and the tests are independent, then

$$\text{FWER} = 1 - (1 - \alpha)^{n^2-n} \geq \alpha \tag{26}$$

To ensure that FWER $\leq \alpha$, the Bonferroni correction provide a method to correct the p-values by using

$$\tilde{p}_{ij} = \min\{(n^2 - n)p_{ij}, 1\} \tag{27}$$

or, equivalently, using a modified α

$$\tilde{\alpha} = \frac{\alpha}{n^2 - n}. \tag{28}$$

By Boole's inequality, it can be shown that [11]

$$\text{FWER} = \mathbf{P}\left\{ \bigcup_{i\neq j} \left(p_{ij} \leq \tilde{\alpha} \right) \right\} \tag{29}$$

$$\leq \sum_{i\neq j} \mathbf{P}\left(p_{ij} \leq \frac{\alpha}{n^2 - n} \right) \tag{30}$$

$$= (n^2 - n)\frac{\alpha}{n^2 - n} \tag{31}$$

$$= \alpha. \tag{32}$$

Therefore, by choosing $\tilde{\alpha} = \alpha/(n^2 - n)$ as a threshold, our tests become conservative. But, the value of $\tilde{\alpha}$ can be very small such that even large F_{ij} could be classified as zero.

Another method called the Benjamini-Hochberg procedure approaches by controlling false discovery rate (FDR) instead. Let U be the number of false discoveries (incorrectly rejected null hypotheses) and R be number of discoveries (all rejected null hypotheses). Then, the false discovery rate is usually defined by [12]

$$\text{FDR} = \mathbf{E}\left[\frac{U}{R}\right] \qquad (= 0 \quad \text{if} \quad R = 0). \tag{33}$$

The Benjamini-Hochberg procedure start by sorting p-values in an ascending order

$$p_{(1)} \leq p_{(2)} \leq \cdots \leq p_{(n^2-n)}. \tag{34}$$

Next, we find the largest index k such that

$$p_{(k)} \leq \frac{k}{n^2 - n}\alpha. \tag{35}$$

Then, we reject the first k hypotheses associated to $p_{(1)}, \ldots, p_{(k)}$. It has been shown in [11] that if the p-values corresponding to the correct null hypotheses are independent, then, by using the Benjamini-Hochberg procedure, we have

$$\text{FDR} \leq \alpha. \tag{36}$$

This method does not guarantee that FWER $\leq \alpha$, so it is not conservative. Instead, the ratio of false discoveries to all discoveries is controlled by α.

4 Preliminary results

In this section, an experiment was performed to study the effect of the number of permutations to the performance of the Monte-Carlo permutation test. Our hypothesis was that the performance increased as the number of permutations increased. This experiment could give us the best number of permutations to be chosen to reduce the computation time.

So, we set up the control conditions as follows. We created 15 state-space models by generating VAR models of order 2 with dimension of 5 and filtered the VAR models with diagonal filters with 1 zero and 2 poles. The obtained state-space models then had 20 states. Time series data were generated from each models with 1000 time points. The number of states use in subspace method was chosen to be the same with the true system (20 states). We varied the number of permutation P from 1 to 200 with the same window length of 100. Then, with the significance level α, we compared the estimated GC patterns obtained by simple thresholding, Bonferroni correction and Benjamini-Hochberg procedure from each models to the true GC patterns.

To test the performance, we considered the following indices.

- True positive (TP): correctly classified non-zeros in F
- True negative (NP): correctly classified zeros in F
- False positive (FP): incorrectly classified non-zeros in F
- False negative (FP): incorrectly classified zeros in F

Together with these indices, we can determine the following classification performance.

$$\text{Accuracy (ACC)} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{37a}$$

$$\text{True positive rate (TPR)} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{37b}$$

$$\text{True negative rate (TNR)} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{37c}$$

$$\text{False positive rate (FPR)} = \frac{\text{FP}}{\text{TN} + \text{FP}} = 1 - \text{TNR} \tag{37d}$$

$$\text{False negative rate (FNR)} = \frac{\text{FN}}{\text{TP} + \text{FN}} = 1 - \text{TPR} \tag{37e}$$

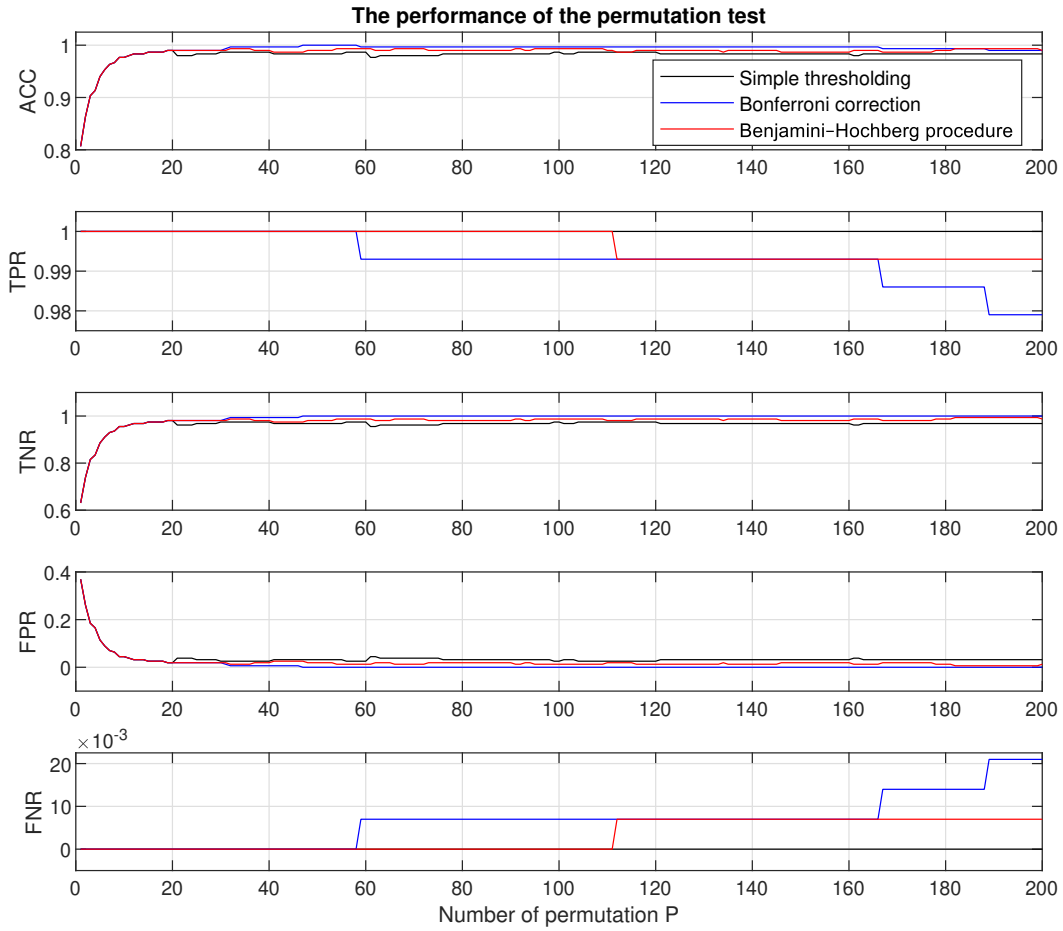

Figure 4: The performances of the Monte-Carlo permutation test for $P = 1, 2, \ldots, 200$ with no correction, with Bonferroni correction and with Benjamini-Hochberg procedure.

The result in *Figure 4* showed that, for $P < 20$, ACC and TNR increased as P was increased. In contrast, the FPR declined as P increased. For $P > 20$, ACC, TNR and FPR became steady and did not have significant changes. For simple thresholding, TPR and FNR were constants at 1 and 0, respectively, for all P while Bonferroni correction and Benjamini-Hochberg procedure had slightly decrease in TPR and slightly increase in FNR as P increased. This result agree with our hypothesis that the performance is improved as the number of permutation increase. Since we performed the Monte-Carlo permutation test, the result is expected to be improved as more permutations are used. It is possible that, for simple thresholding case, the TPR and FNR of constants 1 and 0, respectively, were resulted from that the

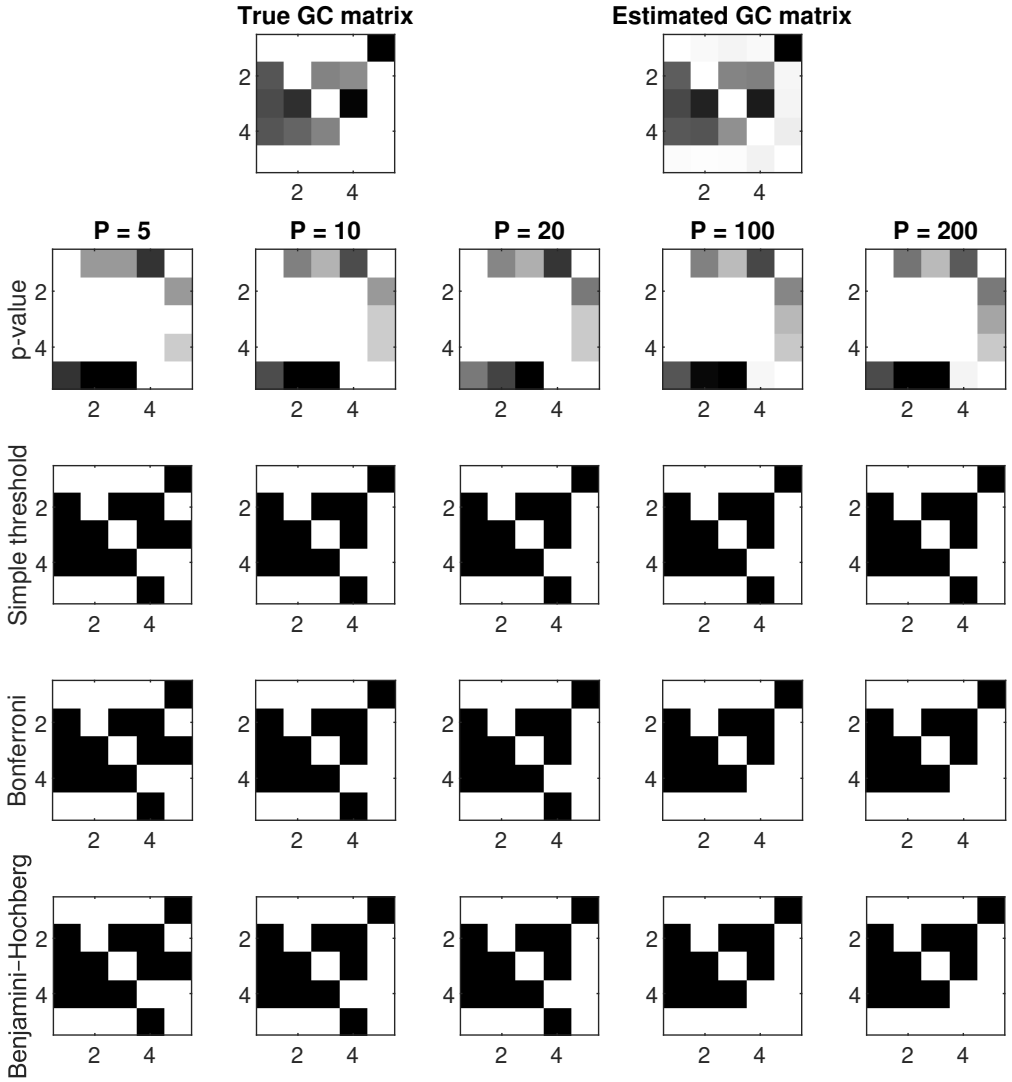

Figure 5: An example of the estimation of GC pattern for $P = 5, 10, 15, 20, 100$ and 200, using significance level $\alpha = 0.05$, thresholding without correction, with Bonferroni correction and with Benjamini-Hochberg procedure, compared to the true pattern. White squares represent zero entries and black squares represent non-zero entries.

estimated non-zero entries were too extreme. In other word, the true causality was significantly strong so that the chance to have a permutation that gives even higher causality than the estimated causality was very low. Another notable result is that for all performance measures, the Benjamini-Hochberg procedure lay mostly between the simple thresholding and the Bonferroni correction. This showed that the Benjamini-Hochberg procedure can control the FPR better than the simple thresholding but slightly worse than the Bonferroni correction. On the other hand, the Benjamini-Hochberg procedure had lower FNR and higher TPR than the Bonferroni correction, demonstrating a trade-off between controlling FPR and FNR.

Figure 5 showed an example of estimated GC patterns. In this example, all methods gave the same results for $P = 5, 10$ and 20 with $2, 1$ and 1 false positives respectively. When $P = 100$ and 200, the Bonferroni correction and the Benjamini-Hochberg procedure gave the same pattern as the true GC pattern while the simple thresholding had one false positive. This example agree with our hypothesis.

The results also showed that the Bonferroni correction and Benjamini-Hochberg procedure had better control of the false positives.

5 Project description

In this section, we explain the scope of work and the project plans in the form of a Gantt chart.

5.1 Scope of work

The scope of this project is the following.

1. We perform and verify the estimation of GC pattern on simulation data only.
2. We compare the performance, the computational cost, and required assumptions of the permutation method with the GMM method for learning GC patterns.

5.2 Project plans

Process	2102490				2102499				
	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr
1. Literature review on Granger causality and the equivalence of ARMA models and state-space models									
2. Writing a MATLAB code on generating ground truth models									
3. Study permutation test and p-value correction methods for testing the significance of zero causality									
4. Experiment 1: Study the effect of the number of permutation on the performance of permutation test									
5. Write a proposal report and prepare for the proposal presentation									
6. Experiment 2: Compare the performance between the permutation test with all permutations and the Monte-Carlo permutation test									
7. Experiment 3: Study the effect of the model order of state-space model on the performance of permutation test									
8. Experiment 4: Compare the performance, the computational cost and the assumptions required of permutation test with GMM method									
9. Discuss the results and adjust the proposed scheme									
10. Write a senior project report and prepare for the final presentation									

Figure 6: The Gantt chart of this project.

Figure 6 shows the plans of this project for both semesters. In the first semester, we reviewed on the Granger causality, ARMA models and the method of converting to state-space. In addition, we wrote a MATLAB code for generating ground truth models with arbitrary GC patterns and controllable sparsity.

For the next semester, we plan to perform the following experiments to study the factors that affect the performance of the permutation test and to reach our objective on comparison with the GMM method.

- Experiment 2: Compare the performance of the permutation test with every permutations with the Monte-Carlo permutation.
  - Hypothesis: The permutation test with every permutations has better performance than the Monte-Carlo permutation test.
  - Experiment settings: We generate many ground truth models with the same parameters and time series data generated form each model are of the same length. The numbers of states used in subspace identification are the same with the ground truth models. The number of partitioning for the permutation test is chosen at 5 so that the number of permutation ($= 5! = 120$) do not exceed 200. So the number of permutation is also fixed at 120 for the Monte-Carlo method. The number of partition of the Monte-Carlo can still be varied. The performance measures of both methods are then computed and compared

- Experiment 3: Study the effect of the number of states used in subspace identification on the performance of the permutation test.
  - Hypothesis: The performance of the permutation test is improve as the number of states used is closer to the true number of states.
  - Experiment settings: We generate many ground truth models with the same parameters and time series data generated form each model are of the same length. The numbers of permutations and the number of partitions are chosen to be equal for each number of states. For the true number of state n_s, the number of states for estimation is vary in range of $n_s \pm 10$. Then, the performance measures are calculated for each number of states and compared.

- Experiment 4: Compare the performance and the computational cost of the permutation test with the GMM method.
  - Hypothesis: The permutation test gives better performance but the GMM method has lower computation time
  - Experiment settings: We generate many ground truth models with the same parameters and time series data generated form each model are of the same length. This time, we can vary any parameters in permutation test but we will choose the parameters based on the previous experiments for the optimal choice. For the GMM method, we refer to [4]. Then, the performance measures are computed for both method and compared. The computational times are also collected for each method to be compared.

After the experiments, we will discuss the results and may adjust our method according to the experiments' results for better performance or lower computation time.

6 References

[1] W. Karwowski, F. Vasheghani Farahani, and N. Lighthall, "Application of graph theory for identifying connectivity patterns in human brain networks: A systematic review," *Frontiers in Neuroscience*, vol. 13, p. 585, 2019.

[2] L. Barnett and A. K. Seth, "The MVGC multivariate Granger causality toolbox: a new approach to Granger-causal inference," *Journal of neuroscience methods*, vol. 223, pp. 50–68, 2014.

[3] L. Barnett and A. K. Seth, "Granger causality for state-space models," *Physical Review E*, vol. 91, no. 4, 2015.

[4] J. Songsiri, "Learning brain connectivity from EEG time series," tech. rep., Chulalongkorn University, 2019. `http://jitkomut.eng.chula.ac.th/pdf/eeg_bc_final_jss.pdf`.

[5] H. Lütkepohl, *New introduction to multiple time series analysis*. Springer Science & Business Media, 2005.

[6] L. Barnett and A. K. Seth, "Behaviour of Granger causality under filtering: theoretical invariance and practical application," *Journal of neuroscience methods*, vol. 201, no. 2, pp. 404–419, 2011.

[7] J. Songsiri, "EE531 - System Identification, Subspace Method." `http://jitkomut.eng.chula.ac.th/ee531/subspace.pdf`.

[8] P. Van Overschee and B. De Moor, *Subspace identification for linear systems: Theory—Implementation—Applications*. Springer Science & Business Media, 2012.

[9] T. E. Nichols and A. P. Holmes, "Nonparametric permutation tests for functional neuroimaging: a primer with examples," *Human brain mapping*, vol. 15, no. 1, pp. 1–25, 2002.

[10] P. I. Good, *Permutation, parametric and bootstrap tests of hypotheses: a practical guide to resampling methods for testing hypotheses*, vol. 100. 2005.

[11] B. Efron, *Large-scale inference: empirical Bayes methods for estimation, testing, and prediction*, vol. 1. Cambridge University Press, 2012.

[12] J. D. Storey, "False discovery rate," *International encyclopedia of statistical science*, pp. 504–508, 2011.