A Review on Dependence Measures in Exploring Brain Networks from fMRI Data

Anupon Pruttiakaravanich[a] and Jitkomut Songsiri[b,*]

Department of Electrical Engineering, Faculty of Engineering, Chulalongkorn University, Bangkok 10330, Thailand
E-mail: [a]anupon106@gmail.com, [b]jitkomut.s@chula.ac.th (Corresponding author)

Abstract. Functional magnetic resonance imaging (fMRI) technique allows us to capture activities occurring in a human brain via signals related to cerebral blood flow, oxygen metabolism and blood volume, known as BOLD (blood oxygen level-dependent) signals. Exploring relationships between brain regions inside human brains from fMRI data is an active and challenging research topic. Relationships or associations between brain regions are commonly referred to as *brain connectivity* or *brain network*. This connectivity can be divided into three groups; (i) structural connectivity representing physically anatomical connections between regions, (ii) the functional connectivity which describes the statistical information among brain regions and (iii) the effective connectivity which specifies how one region interacts with others by a causal model. This survey paper provides a review on learning brain connectivities via fMRI data where detailed mathematical definitions of dependence measures widely-used for functional and effective connectivities are described. These measures include correlation, partial correlation, coherence, partial coherence, directed coherence, partial directed coherence, Granger causality, and other concepts such as dynamical causal modeling or structural equation modeling. Interpretation and relations of these measures as well as relevant estimation techniques that are widely used in the problems of fMRI modeling are summarized in this paper.

Keywords: fMRI, brain network, brain connectivity, causality

1. Introduction

This paper focuses on how to explore relationships among brain regions via some accessibly biological signals in order to have a better insight in human brain functioning. The relationship is also called brain connectivity and learning brain connectivity is an analysis of how each of several brain regions interacts to each other under a functioning operation. This would allow neuroscientists to comprehend a better understanding of the brain mechanism. Recently, there are many several techniques for measuring brain activity. The first widely-used technique includes electroencephalography (EEG), which measures the brain activity directly by sticking non-invasive electrodes along the scalp for recording electrical signals over a period of time from the brain. The signal is called EEG signal. The second technique is functional magnetic resonance imaging (fMRI), which measures the brain activity via signals from a change of blood level when the brain is activated. This signal is called "Blood-oxygen-level dependent" or BOLD signal. Technically, a neuron surrounded by astrocytes stays in normal condition when the brain is not activated. Once the brain is stimulated such as events of seeing, listening or thinking, neurons that do not have reserved energy need external energy in the form of oxygen. As a result, astrocytes transfer oxygen to neurons via blood flow that makes a change of level of blood flow around those areas. This change can be detected and recorded by fMRI scanner. In other words, fMRI scanner measures the brain activity indirectly via the change of blood level when neurons are activated by external stimulus. The technique for measuring the brain activity by fMRI allows us to see the brain activity and 3D brain image simultaneously in time series. At each time point, a 3D images is divided into several hundred thousand or million blocks that have the same width, length and depth called voxels. One voxel represents a group of neurons. One main difference between EEG and fMRI signal is that EEG signal poorly captures neural activities occurring under the surface cortex of the brain while fMRI signals can efficiently capture the brain activities in the whole brain volume with a high spatial resolution. For this reason, we particularly focus on a review of the use of fMRI time series instead of EEG in identifying the brain connectivities.

A brain connectivity is hypothetically believed to differ under various stimulated conditions. A resting-state fMRI study reflects the brain connectivity of a subject while being rest. One example of the brain network in such state is called the *default mode network (DMN)* [1, 2, 3], where connections between some regions of the brain are elevated. On the other hand, a brain network of disordered patients such as Alzheimer or Schizophrenia patients is assumed to have different structures. For this reason, discovering brain connectivity does not only provide us a better insight of how information is processed and transferred between linked brain regions, but also delivers an alternative classification approach for distinguishing two classes of the brain. The use of fMRI technique is accepted to be a promising trend to explore core roles and functions of the brain.

Introductory survey in exploring the brain connectivity can start with [4, 5, 6, 7, 8] where general formulation, statistical analysis and essential elements in the study were explained. The use of graphical model to represent causal interaction of complex brain network is explained in [9, 10, 11, 12]. An extensive survey of linear/nonlinear mathematical modeling including neuron dynamics and hemodynamic response can be followed from [13]. Future trends of using fMRI data for learning the relation structure of each region inside the human brain can be found in [14, 15, 16]. These are just selected prominent examples of reviewed papers in the fMRI literature. Our paper aims to focus heavily on the effective connectivity and is an alternative from other surveys to the extent that we provide the detail of mathematical definitions of causality measures and their relations from an engineering point of view.

To this end, we first describe two major approaches for identifying the brain regions in section 2. This explains how to define a variable or a node that represents a group of neurons. Section 3 provides the main definitions of a brain connectivity; functional and effective connectivities and we give examples of commonly used models that explain fMRI dynamics. Section 4 covers a few classes of statistical measures to formally explain relationships between brain regions and the corresponding statistical techniques for model estimation are discussed in section 5.

2. Defining Brain Regions

The information of BOLD signal as a function of time is measured from different points in the brain. These spatial coordinates can be used in different two scales, voxel or region of interest (ROI). Voxel is a cubic representing a spatial resolution unit in 3D fMRI image, *e.g.*, $3 \times 3 \times 3$ mm^3. Brain regions of interest (ROIs) can be determined as a group of voxels defined by using some prior anatomical knowledge or from some post-processed analysis such as activity map. Either using fMRI signals collected from voxel level or ROIs is considered to be features that must be defined before exploring brain connectivity. There are two approaches to define these features. The first approach is the use of model dependent method or seed method which defines the brain region as "seed". A seed can be a brain region of interest (ROI) or it can be chosen from traditional task-dependent activation map from fMRI paradigm [2]. A limitation of this approach is a difficulty in applying to the whole brain scale. In other words, this approach is only suitable for analysis of brain connectivity in a small number of brain regions of interest (ROIs). The second approach that solves such limitation is the use of model-free method. This approach does not need to define a seed region or reference region from prior knowledge, and it may use some mathematical techniques, such as principal component analysis (PCA) in [17, 18, 19, 20] or independent component analysis (ICA) in [21, 22, 19], for defining voxels. PCA maps fMRI data to a new space via an orthogonal transformation and defines a set of components having the greatest variances as voxels. ICA uses fMRI data to seek for a mixture of underlying sources and defines these sources that are maximally independent from each other as voxels. Techniques for defining voxels like PCA and ICA are called Exploratory Matrix Factorization (EMF) that aims to extract blind source from observation and defines a set with some predefined properties as components. Other EMF techniques *e.g.*, SVD, nonnegative matrix and tensor factorization (NMF/NTF), are explained in [7]. As the voxels from these techniques are mapped in the new space, the result may be difficult to justify by anatomical knowledge since the interpretation is based on the information on the new space [7, §3.2].

3. Brain Connectivity

One of interesting research topics in machine learning about model inference is to seek relationships between a pair of components in random vector or multivariate time series. In our application we seek for a relationship between brain regions or groups of neurons under some conditions. This relationship is called *brain connectivity, brain network* or *neural network*, just to name a few. The definition of brain connectivity can be broadly divided into *structural, functional and effective* connectivity, where the definition of each type associating with the method used to quantify it determines the edges of a brain network [12]. *Structural connectivity* represents physically anatomical connections between regions and we opt not to explore in detail on this type of connectivity. *Functional connectivity* represents statistical dependencies between brain regions such as correlation or partial correlation and the connection can be undirected or indirected. *Effective connectivity* explains about a causal inference concluded from a dynamical model of neurophysiological signals; temporal effects of signals are taken into account to define such dependency. Previous studies show that association of neural mechanisms among brain regions are not limited to the physical wiring of the brain, so it is more interesting to explore the functional and effective connectivities in fMRI studies. The brain connectivities can generally be presented in the form of functional connectivity map (FCMAP) or graph. For most applications in fMRI study, the brain network is mostly presented in the graph which comprises a set of nodes and edges. Each node represents each brain region (voxel) and each edge represents a statistical relationship between two nodes. A common approach to present the link in network is the use of connection matrix (or adjacency matrix) in which its entries correspond to the connection strength between nodes. This connection strength can be obtained from measuring the relationship between nodes using some statistical definitions called *dependence measure* where a formal definition of these measures will be explained more in the detail in section 4. As an example to illustrate this, Figure 1 shows a brain dependence structure as a graph with the corresponding connection matrix.

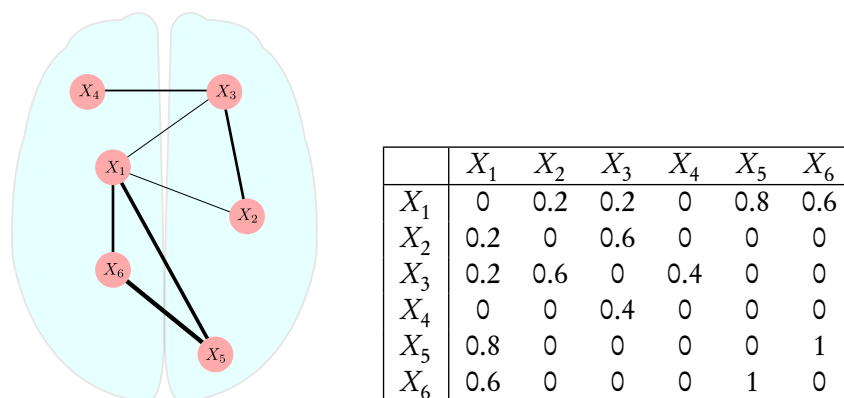|	X_1	X_2	X_3	X_4	X_5	X_6
X_1	0	0.2	0.2	0	0.8	0.6
X_2	0.2	0	0.6	0	0	0
X_3	0.2	0.6	0	0.4	0	0
X_4	0	0	0.4	0	0	0
X_5	0.8	0	0	0	0	1
X_6	0.6	0	0	0	1	0

Fig. 1. A graph for brain network corresponding to its connection (adjacency) matrix. Each node represents each brain region (voxel) and each edge represents a statistical relationship between two nodes. The link width can represent the connection strength between two nodes which varies according to the entry value in adjacency matrix. The connection matrix depends on the dependence measures described in the section 4.

3.1. Functional Connectivity

Functional connectivity gives information about which groups of neurons (voxels) are linked together where the association between voxels is defined by a statistical dependency called *dependence measure* or *measure of directionality* [12]. Examples of such measures are correlation, partial correlation, or cross spectra function (see more details in section 4). Therefore, the functional connectivity has many beneficial usages for medical purposes. For instance, the use of functional connectivity to discriminate between schizophrenia patients and healthy persons in [23] or between Alzheimer patients and healthy persons in [24]. Exploring functional connectivity is roughly divided into two approaches depending on the method of defining brain regions *i.e.*, seed method and model-free method, which are previously mentioned in section 2.

- Seed method: It defines a seed before finding the correlation between the seed and other regions resulting in functional connectivity map (FCMAP) that provides information about which brain regions have a relationship with this seed [2].

- Model-free method: Examples that use this method include [17, 18, 21, 22, 25, 26, 27, 19, 20]. The result from this method may often be difficult for interpretation by using anatomical knowledge. In some cases, one claims that seeking of functional connectivity using fMRI data from a single subject may not be reliable because these fMRI signals may have a contamination from noise. Therefore, a group functional connectivity, which has an assumption that all subjects in group must share the same of functional connectivity pattern, may be more robust to noise signals. However, the use of model-free method can be considered for this situation and it is known as group analysis which is divided into three approaches.

  - The first approach is the use of group fMRI data *i.e.*, the integrated fMRI data from all subjects in group, to define voxels using technique such as group ICA and then finds a common functional connectivity (functional connectivity of group) as examples in [28, 29]. After obtaining a group functional connectivity, one may reconstruct to obtain the individual functional connectivity of each subject by using some information from group functional connectivity.

  - The second approach is the use of individual fMRI data to find individual functional connectivity of each subject and then integrates all individual functional connectivity to one common functional connectivity which will be the functional connectivity of this group as shown in [30, 31, 32]. In fact, the first approach and this approach have no relationship

of functional connectivity pattern between subjects but the following third approach can overcome this weakness.

- The third approach integrates fMRI data to define voxels. This approach is to seek the common functional connectivity and individual functional connectivity simultaneously, resulting in similar patterns of brain connectivity among subjects; see example in [33].

Functional connectivity can be quantified by several measures but most of them explain static dependency such as correlation, coherence, mutual information, etc. The work in [34] provides implementation of a variety of those measures and develops a MATLAB toolbox with a suggestion for the measure selection. Moreover, there are some researches focusing on explaining dynamics of functional connectivity to observe patterns of functional connectivity and how they change along the time [35, 36, 37, 38].

3.2. Effective Connectivity

Effective connectivity describes how a group of neurons affects to the others via a causal influence concluded from causal models that explain brain behaviors. Therefore, the analysis of effective connectivity requires two important concepts; a selection of causal model and model parameter estimation. Common techniques of exploring the effective connectivity includes dynamic causal modeling (DCM) [39, 6, 13, 40, 41, 42], Granger causal modeling (GCM) [42, 43, 41, 13, 44] and structural equation modeling (SEM) [45, 46, 47, 20, 48].

Dynamic Causal Modeling (DCM)

This type of model has been proposed in [40, 39] where the neuronal activity is described by state-space equation, and there is a specific mapping from the neuronal activity model to an observation model. This is due to the fact that the change of neuronal activity cannot be measured directly since it is caused from synaptic activity and the change of blood volume in the brain. The differential equations describing such dynamics are given by

$$\dot{z} = f(z, u, \theta_1) + w, \tag{1}$$

$$y = g(z, u, \theta_2) + v, \tag{2}$$

where z_i represents the neuronal activity at brain region i for $i = 1, \dots, n$; w and v represents random fluctuation on the change of neuronal activity; u represents a known exogenous input from the change in experimental conditions; and g denotes a function containing some model parameters and relating to measured response y. The variables θ_1 and θ_2 denote the model parameters which are to be estimated. Equation (1) describes the change or flow of neuronal activity that depends on the states z_i of other regions. The state variables z_i cannot be directly measured, so it is called *hidden* neuronal activity. Equation (2) can be considered as a specific mapping model or assumed to be the measured response from (1), and as a result, is called the *observed model*. The models (1) together with (2) are generally complicated to analyze. The random inputs w and v are typically unknown (often referred to as endogenous inputs) but if their statistical properties such as a probabilistic distribution, and prior distributions of the model parameters θ, can be assumed, then the models (1) and (2) can completely explain the dynamics of the hidden neuronal states and the hemodynamic response in a stochastic sense, and they are regarded as a *generative model* of the data. An inverse problem of fitting this generative model by estimating unknown parameter θ from given observed data is known as dynamic causal modeling (DCM) technique [39].

The fMRI study in [40] assumed that the change of neuronal activity obtains the influence from i) neuronal activities of other regions, ii) induced input and iii) direct input to the brain regions. Therefore, a simpler form of (1)-(2) which is a bilinear differential equation is proposed to describe the brain behavior.

$$\dot{z} = (A + \sum_{j=1}^{m} u_j B_j) z + C u, \tag{3}$$

$$y = g(z, u) + v, \tag{4}$$

where $z(\cdot) \in \mathbf{R}^n$, $u(\cdot) \in \mathbf{R}^m$, $A \in \mathbf{R}^{n \times n}$, $B^j \in \mathbf{R}^{n \times n}$ for $j = 1, \ldots, m$, and $C \in \mathbf{R}^{n \times m}$. The parameters n and m denote the number of brain regions and number of experimental inputs, respectively, and z_i represents neuronal activity at region i for $i = 1, \ldots, n$. The function g in (4) describes a hemodynamic model [40] which explains how the blood volume changes in brain and the output response y from this model is the BOLD signal with v as measurement noise. The coupling between brain regions can be separated into 1) latent connectivity, which illustrates the intrinsic coupling of experimental perturbation as observed from the matrix A and 2) induced connectivity, which illustrates the change in coupling induced by exogenous input u as observed from the matrix B_j. The matrix C represents the extrinsic influence of inputs on neuronal activity. The connectivity between brain regions can be found on the nonzero entries of coupling matrices A and B_j from the model. In this application, the input u comes from the stimulus in fMRI experiment, *e.g.*, visual words or pictures. Equation (3) can be interpreted as a realistic neuronal model. Note that, practically, we cannot measure neuronal activity from (3) directly therefore the aim of this approach is to estimate parameters A, B_j and C from the measurement y in (4).

Granger Causal Modeling (GCM)

Let $y(t) = (y_1(t), \ldots, y_n(t))$ be a multivariate time series. The concept of Granger causality states that a time series $y_i(t)$ is Granger-caused by a time series $y_j(t)$ if $y_i(t)$ depends on $y_j(t)$ in such a way that $y_j(t)$ helps to predict $y_i(t)$ better than using only the information of the time series $y_i(t)$ in the past. There is an illustration that the concept of Granger causality becomes a simple condition when it is applied to autoregressive (AR) processes. To describe such characterization, an n-dimensional AR process of order p is given by

$$ y(t) = A_1 y(t-1) + A_2 y(t-2) + \cdots + A_p y(t-p) + \varepsilon(t) \tag{5} $$

where $y(\cdot) \in \mathbf{R}^n$, $\varepsilon(\cdot) \in \mathbf{R}^n$ is an input noise and $A_1, A_2, \ldots, A_p \in \mathbf{R}^{n \times n}$ are AR coefficients. It can be shown in [49] that $y_j(t)$ is NOT Granger-caused $y_i(t)$ if and only if

$$ (A_k)_{ij} = 0, \quad k = 1, 2, \ldots, p \tag{6} $$

In other words, we say loosely that y_j has no effect on y_i if the corresponding entry of the weight matrix A_k is zero for all lags. Such characterization is a simple linear constraint in the AR coefficients, so it has been widely considered in the problem of learning causal structures from time series. The term GCM is simply referred to as a dynamical model that concludes the relationship between two time series via the concept of Granger causality. The work on this direction can be divided into two approaches as follows.

- **Statistical test.** From given data, the model parameters, in this case the AR coefficients, are estimated by using the ordinary least-squares (OLS) technique. Subsequently, one can check the condition (6) by performing a statistical test (or hypothesis test) on the estimated AR parameters. Therefore, the zero pattern in the AR coefficients can then be concluded about the relationship between two time series. However, the consistency of OLS estimator highly depends on the sample size of data, therefore we may get a bad result if the given data is not adequate. This approach is also not quite feasible if the number of variables (n) is high, since the test on the (i, j) entry of A_k has to be performed for $n^2 - n$ times. The work based on this approach includes [50] where a t-test was performed on the regression coefficient; [51] and [52] where a statistical test was performed on bivariate AR coefficients and modified AR coefficients, respectively;[53, 54] established an F-test on the null hypothesis that $A_{ij} = 0$.

- **Regularized estimation.** In practice, the Granger causality constraint in (6) is not known in a *priori*, but rather to be discovered from the data. Therefore, the estimation formulation in this approach aims to promote sparsity in the AR coefficients (solution containing many zeros). Thus, the Granger relationship structure is determined by the location of these zeros which can be varied depending on the estimation formulation and its tuning parameters. We can exploit sparse

optimization techniques in order to propose estimation formulations in this direction, which involves least-squares with regularization problems. The regularization term can then be the ℓ_1-norm (called LASSO) [55, 56, 57], ℓ_2-norm (called ridge regression) [55, 58, 59] or both ℓ_1/ℓ_2-norm (called elastic-net) [60]. The application of this approach on fMRI studies includes [61, 62].

In fMRI applications where AR models are used to fit fMRI time series, $y_i(t)$ denotes the signal from the ith voxel or the ith region of interest. The coupling between brain regions i and j can then be observed from AR coefficients tested by a statistical test or from the nonzero pattern of the estimated AR coefficients in the model.
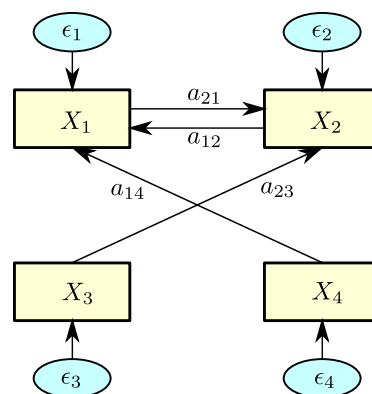
Structural Equation Modeling (SEM)

Fig. 2. Structural diagram of SEM. In a structural diagram, a rectangle denotes an *observed* variable and an ellipse denotes a *latent* variable. Since this structure is based on path analysis, it doest not consider other latent variables except the residual term from measurement. A link between observed variables represents a causal relationship which comes from the assumption of a *priori* knowledge.

Structural Equation Modeling (SEM) is one of statistical techniques used for seeking a statistical causal model (called exploratory modeling) and testing whether the model are supported by the given data (called confirmatory modeling). In fact, SEM is largely used in the confirmatory framework rather than the exploratory framework. That is, we are likely to use SEM for testing a hypothesis or estimating a causal relationship among variables on a model. Confirmatory SEM usually starts with a hypothesis and uses this hypothesis to form a model. Once the model parameters are estimated, one defines some statistical criterion to conclude whether the model in assumption is supported by the hypothesis. SEM deals with two types of variable, the *observed* variables which can be directly measured and the *latent* variables that cannot be directly measured and are inferred from the observed variables [63]. For example, if one would like to measure the intelligence from someone via IQ test, the latent variable is the intelligence and the observed variable is the scores obtained from IQ test. SEM commonly consists of a *measurement model* and a *structural regression model*. The measurement model defines a relationship between observed variables and latent variables and the structural model indicates the relationship between latent variables. Combining these two models can form the structural diagram shown in Figure 2 which describes how the entire variables are associated. In this diagram, four rectangular boxes denote the observed variables and four ellipses denote the latent variables. The goal of SEM is to estimate the model parameters that minimize the difference between empirical covariance matrix from observed data and estimated covariance matrix from the model. In other words, SEM fits the covariance of the variables rather than the observed values of those variables themselves. By the basic concept of SEM, it allows to take the effect of measurement error or prediction error into estimation process, meaning that, we can estimate the measurement error and this results in a more precision of the estimated model parameters when comparing with classical approaches such as multiple regression analyses.

Most researches about SEM in fMRI study are based on *path analysis* which is known as a special case of SEM. In path analysis, only the observed variables can be used in the model or we can say that the path analysis does not contain the latent variables in the model except the error from measurement. Therefore we can explain a relationship among variables in a form of n-dimensional linear equation as written below.

$$x = c + Ax + \epsilon \qquad (7)$$

where $x \in \mathbf{R}^n$ denotes the measurement values (here in fMRI studies, are brain data from n regions of interest). The measurement noise, which is assumed to be white noise, is denoted by $\epsilon \in \mathbf{R}^n$. The baseline of x is denoted by $c \in \mathbf{R}^n$. The matrix $A \in \mathbf{R}^{n \times n}$ denotes the *path matrix* which is to be estimated and the entry a_{ij} is known as path coefficient representing the causal relationship from brain region j to i. The magnitude of path coefficient can be interpreted as the connection strength or statistical significant among brain regions and the sign (positive or negative) of path coefficient signifies the direction of connection. SEM technique is different form DCM and GCM because SEM focuses only on the *instantaneous* effects among the variables. From (7), it is a linear static model, so the application of SEM in fMRI studies neglects the effect of time and considers each time point of data as one independent realization. Note that SEM is used for estimating modeling parameters according to the causal relationship that is *priori* assumed in a multivariate system. For example, if a hypothesis of our relationship is assumed as in Figure 2, the instantaneous path matrix can be written as

$$A = \begin{bmatrix} 0 & a_{12} & 0 & a_{14} \\ a_{21} & 0 & a_{23} & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

and the nonzero entries are to be estimated based on the observed data. We explain briefly how to estimate the matrix A in path analysis as follows. Taking expectation both sides of (7) gives $c = (I - A)\mathbf{E}[x]$, and it reduces to

$$x - \mathbf{E}[x] = (I - A)^{-1}\epsilon. \qquad (8)$$

If we denote the covariance matrix of x and ϵ by Σ and Ψ respectively, then we can show from (8) that the covariance matrix x can be expressed as

$$\Sigma = (I - A)^{-1}\mathbf{E}[\epsilon\epsilon^T](I - A)^{-T} = (I - A)^{-1}\Psi(I - A)^{-T}. \qquad (9)$$

Let S denote the sample covariance matrix of x, and let Σ denote the estimated covariance matrix of x derived from SEM. With an assumption of normal distribution of x, SEM applies the concept of minimizing the distance between S and Σ to estimate the path matrix. To this end, we choose the Kullback-Leibler (KL) divergence function as the distance measure, which is given by

$$d(S, \Sigma) = \log \det \Sigma + \mathbf{tr}(S\Sigma^{-1}) - \log \det S - n, \qquad (10)$$

where n denotes the number of variables (voxel) and $\mathbf{tr}(\cdot)$ denotes the trace (sum of diagonal entries) of a square matrix. In this context, the KL function from (10) is used for measuring the distance between two covariance matrices of normal distributions. The value of this function will approach to zero when Σ is close to S. Therefore, the corresponding estimation problem is to minimize $d(S, \Sigma)$ with variable Σ and A, and subject to the constraint in (9), which is a nonlinear optimization problem with a quadratic equality constraint.

In this approach, the coupling between regions can be observed from the path coefficient, describing how much the influence of other regions changes when activity of interesting region changes by one unit. Since SEM removes the effect of time lag in analysis, the connection strength in SEM implies only the instantaneous causal relationship among brain regions *i.e.*, the causal relationship for each time point; see more details for the principle of SEM for fMRI application in [46, 45]. The use of confirmatory SEM can be extended to combine with other techniques such as vector autoregression technique as shown in [64,

65, 66]. In some cases, there is the use of exploratory SEM applied to fMRI study to find all possible causal relationships among brain regions without a *priori* knowledge as proposed in [48]. After obtaining all candidate models, a model selection criterion such as goodness of fit index (GFI) is performed to select the best model.

There is another framework to explore effective connectivity such as the use of dynamic in probabilistic framework based on Dynamic Bayesian Network (DBN) [67]. A major benefit from effective connectivity is to reveal dynamic of hidden behavior from interesting brain regions which functional connectivity cannot. But in some cases as shown in [68], the use of dynamic causal modeling is applied to discover functional connectivity using correlation coefficient, coherence and Granger causality as dependence measures and analyzing these measures by cross covariance function, cross spectra function and direct transfer function respectively. Their goal is to compare the resulting effective connectivity results and justify them to determine the best one in order to further use this information to explain about functional connectivity.

4. Dependence Measures

To measure the relationship between groups of neuron, or between regions of interest, we need a formal statistical definition of such relationship or dependence structure. There has been many "dependence measures" or "measures of directionality" proposed to causal analysis of random variables, or time series. In this section, we summarize the concept of these measures which can be divided into a few classes as follows.

4.1. Dependence measures for random vectors

This measure focuses on the relationship between two components in random vectors. Let x be n-dimensional random vector with mean μ and covariance Σ, *i.e.*, $\mu = \mathbf{E}[x]$ and $\Sigma = \mathbf{E}[(x - \mu)(x - \mu)^T]$. The correlation matrix of x is denoted by $R = \mathbf{E}[xx^T]$ and let x_i denote the ith component of x. Some commonly used dependence measures are described below.

Cross covariance and Pearson correlation coefficient

The two components x_i and x_j are said to be uncorrelated if and only if $\Sigma_{ij} = 0$. If x has zero mean, this relationship can be observed from the correlation matrix R. In other words, two components x_i and x_j from a zero-mean vector x are said to be uncorrelated if and only if $R_{ij} = 0$. In several applications, there is the use of a statistical definition called *Pearson correlation coefficient (ρ)*, given by $\rho = D^{-1/2}\Sigma D^{-1/2}$ where $D = \mathbf{diag}(\Sigma_{11}, \Sigma_{22}, ..., \Sigma_{nn})$ (a diagonal matrix with entries of Σ_{kk}'s) and $D^{-1/2}$ is the inverse of the square root of D, instead of using Σ to measure the relationship between two elements in the random vector. Essentially, Pearson correlation coefficient ρ is just a normalized Σ in which its diagonal entries are equal to one. Pearson correlation coefficient between two components x_i and x_j is given by

$$\rho_{ij} = \frac{\Sigma_{ij}}{\sqrt{\Sigma_{ii}\Sigma_{jj}}}$$

where ρ_{ij} varies from -1 to 1. Therefore x_i and x_j are said to be uncorrelated if and only if $\rho_{ij} = 0$. For example, [26] and [30] use average fMRI time course of ROIs to form a connection matrix and define the link among brain regions by using Pearson correlation coefficient and correlation matrix, respectively. Another example in [24], that focuses on classification problem between mild cognitive impairment (MCI) patients and normal person, uses Pearson correlation coefficient to form the connection matrix of each subject and then applies this matrix as a feature to be classified.

Partial correlation

The partial correlation is used for measuring the relationship between two components in a random vector by assuming the other components are given. If we denote $y = (x_i, x_j)$ and $z = (x_1, ..., x_k), \forall k \neq i, j$. The covariance matrix of y and z is given by

$$\mathbf{cov}\left(\begin{bmatrix} y \\ z \end{bmatrix}\right) = \tilde{\Sigma} = \begin{bmatrix} \tilde{\Sigma}_{yy} & \tilde{\Sigma}_{yz} \\ \tilde{\Sigma}_{zy} & \tilde{\Sigma}_{zz} \end{bmatrix}. \tag{11}$$

By definition, the partial correlation is the correlation between residuals of component x_i and x_j that are projected onto the linear space spanned by the other given components. Let r_i be a residual between x_i and $\hat{x}_i$ in which $\hat{x}_i$ is a projection of x_i onto the linear space spanned by $x_k, \forall k \neq i, j$. If we denote residual vector r by $r = (r_i, r_j)$, it can be shown that the covariance matrix of r can be computed from

$$\mathbf{cov}(r) = \Omega = \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{bmatrix} = \tilde{\Sigma}_{yy} - \tilde{\Sigma}_{yz}\tilde{\Sigma}_{zz}^{-1}\tilde{\Sigma}_{zy}, \tag{12}$$

which is the Schur complement of the block $\tilde{\Sigma}_{zz}$ in $\tilde{\Sigma}$. The matrix in (12) is called *partial covariance matrix* and Ω_{12} represents the partial covariance between component x_i and x_j. Using block matrix inversion lemma on $\tilde{\Sigma}$ in (11) provides

$$\tilde{\Sigma}^{-1} = \begin{bmatrix} \tilde{\Sigma}_{yy} & \tilde{\Sigma}_{yz} \\ \tilde{\Sigma}_{zy} & \tilde{\Sigma}_{zz} \end{bmatrix}^{-1} = \begin{bmatrix} (\tilde{\Sigma}_{yy} - \tilde{\Sigma}_{yz}\tilde{\Sigma}_{zz}^{-1}\tilde{\Sigma}_{zy})^{-1} & \square \\ \square & \square \end{bmatrix}, \tag{13}$$

(the entries labeled as $\square$ can be ignored for now but their expressions are given in the Appendix). From (11), (12) and (13) there is a relationship between $\tilde{\Sigma}^{-1}$ and Ω given by $(\tilde{\Sigma}^{-1})_{12} = -\Omega_{12}$ since Ω has size 2×2. Let P denote a permutation matrix such that $Px = (y, z)$, *i.e.*, we permute x_i and x_j to be the first two elements, then $\tilde{\Sigma} = \mathbf{cov}(Px) = P\Sigma P^T$. Thus we have $(\tilde{\Sigma}^{-1})_{12} = (P\Sigma^{-1}P^T)_{12} = (\Sigma^{-1})_{ij}$. Therefore, the following quantities are linearly proportional to each other:

$$(\Sigma^{-1})_{ij} \propto (\tilde{\Sigma}^{-1})_{12} \propto ((\tilde{\Sigma}_{yy} - \tilde{\Sigma}_{yz}\tilde{\Sigma}_{zz}^{-1}\tilde{\Sigma}_{zy})^{-1})_{12} \propto \Omega_{12}. \tag{14}$$

This result essentially means that the partial covariance between component x_i and x_j with given x_k can be observed from the entries of $(\Sigma^{-1})_{ij}$; simply from the zero pattern of the inverse of covariance matrix. The *partial correlation matrix* is defined as the normalized Σ^{-1} by its diagonal entries. Define $D = \mathbf{diag}((\Sigma^{-1})_{11}, (\Sigma^{-1})_{12}, ..., (\Sigma^{-1})_{nn})$. Therefore, the partial correlation between component x_i and x_j conditioning other components is given by

$$\text{PCORR} = D^{-1/2}\Sigma^{-1}D^{-1/2}, \quad \text{PCORR}_{ij} = \frac{(\Sigma^{-1})_{ij}}{\sqrt{(\Sigma^{-1})_{ii}(\Sigma^{-1})_{jj}}}.$$

Hence, we can say that x_i and x_j are *partially uncorrelated* if and only if $\text{PCORR}_{ij} = 0$ (or equivalently, $(\Sigma^{-1})_{ij} = 0$).

Conditional Independence

It is based on the idea that two components of a random vector are independent conditioning that the other components are given. It is known that if x_i and x_j are independent, they are also uncorrelated, $\Sigma_{ij} = 0$, but the converse is not true. However, if x_i and x_j are Gaussian, the converse is also true, meaning that $\Sigma_{ij} = 0$ implies the independence between x_i and x_j. This idea can be extended to the independence of x_i and x_j with given x_k and can be derived nicely into a mathematical condition on

model parameters. Let x be a *Gaussian* vector. If x_i and x_j are *partially uncorrelated i.e.*, $(\Sigma^{-1})_{ij} = 0$, x_i and x_j are also *conditionally independent* with given x_k. The work in this approach therefore estimates the covariance matrix from the fMRI data and reveal the zero pattern in the inverse of covariance matrix [35, 23, 25].

Path coefficient or path matrix from SEM

Let x be n-dimensional variable. The path analysis in SEM can be expressed by a static linear model:

$$x = c + Ax + \epsilon,$$

where $x \in \mathbf{R}^n$, $c \in \mathbf{R}^n$, $A \in \mathbf{R}^{n \times n}$ and $\epsilon \in \mathbf{R}^n$. The variable x_j has a cause to x_i if the corresponding path coefficient $a_{ij} \neq 0$. Therefore, a causal structure in SEM can be learned from the zero pattern of the path matrix A. For confirmatory SEM, the causal link between components in x can be observed directly from path matrix A that is imposed by assumption from a *priori* knowledge [47, 66]. For exploratory SEM, we explore a number of candidate models with a hypothesis on a causal structure and then estimate the parameters of each model and choose the best model by some criterion. For this approach, a link between components in x can be still observed directly from the path matrix A [69, 48].

4.2. Dependence measures for time series

Let $x(t) = (x_1(t), \dots, x_n(t))$ be n-dimensional wide-sense stationary random process with mean, the correlation matrix, and the covariance matrix given by

$$\mu = \mathbf{E}[x(t)], \quad R(\tau) = \mathbf{E}[x(t)x(t-\tau)^T], \quad C(\tau) = \mathbf{E}[(x(t)-\mu)(x(t-\tau)-\mu)^T]$$

respectively. In fMRI context, $x_i(t)$ represents the signal obtained from brain region i for $i = 1, 2, \dots, n$. To conclude the dependency between two components of multivariate time series, we take into account of the signal dynamics. The analysis considers $\{x(t)\}_{t=0}^N$ as a whole random process, where any two consecutive time points, $x(t)$ and $x(s)$ cannot be regarded as independent samples. In this section, we summarize the dependence measures that are typically considered in the literature.

Cross covariance function and cross coherence function

$x_i(t)$ and $x_j(t)$ are said to be *uncorrelated* if and only if $C_{ij}(\tau) = 0$ for all τ. In time series analysis [70], we opt to use the *normalized correlation function* to measure a similarity between two time series. This correlation function is defined by $\tilde{R}(\tau) = D^{-1/2}C(\tau)D^{-1/2}$ where $D = \mathbf{diag}(C_{11}(0), C_{22}(0), \dots, C_{nn}(0))$. Hence, the cross covariance function between two components $x_i(t)$ and $x_j(t)$ is given as

$$\tilde{R}_{ij}(\tau) = \frac{C_{ij}(\tau)}{\sqrt{C_{ii}(0)C_{jj}(0)}}$$

where it can be shown that $|\tilde{R}_{ij}(\tau)| \leq \sqrt{\tilde{R}_{ii}(0)\tilde{R}_{jj}(0)}$, for all τ. Therefore $x_i(t)$ and $x_j(t)$ are uncorrelated if and only if $\tilde{R}_{ij}(\tau) = 0$ for all τ. In addition, there is a relationship of this measure between time domain and frequency domain. Consider a zero-mean process $x(t)$. If we denote $S(\omega)$ the spectral density matrix obtained from Fourier transform of correlation function $R(\tau)$, *i.e.*, $S(\omega) = \sum_{\tau=-\infty}^{\infty} R(\tau)e^{-j\omega\tau}$, then the condition that $R_{ij}(\tau) = 0$ for all τ is equivalent to $S_{ij}(\omega) = 0$ for all ω. To conclude about relationship of time series in frequency domain, this can be generally observed from *coherence function* which is a normalized version of $S(\omega)$ by its diagonal entries. Define $D(\omega) = \mathbf{diag}(S_{11}(\omega), S_{22}(\omega)), \dots, S_{nn}(\omega))$. Hence the cross coherence function between two components $x_i(t)$ and $x_j(t)$ is given by

$$\mathrm{COH}(\omega) = D^{-1/2}(\omega)S(\omega)D^{-1/2}(\omega), \quad \mathrm{COH}(\omega)_{ij} = \frac{S_{ij}(\omega)}{\sqrt{S_{ii}(\omega)S_{jj}(\omega)}}$$

where $|COH(\omega)_{ij}|$ varies from 0 to 1 for all ω (see [70], section 8). Therefore $x_i(t)$ and $x_j(t)$ are uncorrelated if and only if $COH(\omega)_{ij} = 0$ for all ω. See [71, 72, 34] for the use of cross coherence function for learning functional connectivity and see [68] for an example of applying the cross covariance and cross coherence functions combining with a DCM technique for defining the link between brain regions.

Partial correlation function and partial coherence function

This type of measure is an analogue of the partial correlation for random vectors. It is to indicate the correlation of two components in $x(t)$ after the linear effects of other components in $x(t)$ have been removed. Let $r_i(t)$ be the residual from an optimal prediction of $x_i(t)$, given by $r_i(t) = x_i(t) - \hat{x}_i(t)$ where $\hat{x}_i(t)$ is the optimal linear predictor of $x_i(t)$ as a function of $x_k(t)$, for $k \neq i, j$. Let us define the residual vector by $r(t)$ with mean μ_r, and covariance matrix of residual by $\Omega(\tau) = \mathbf{E}[(r(t) - \mu_r)(r(t - \tau) - \mu_r)^T]$. The partial correlation matrix is defined as a normalized covariance matrix of residual $r(t)$ by its diagonal entries at lag zero. Define $D = \mathbf{diag}(\Omega_{11}(0), ..., \Omega_{nn}(0))$. Therefore the partial correlation function between $x_i(t)$ and $x_j(t)$ is given by

$$\text{PCORR}(\tau) = D^{-1/2}\Omega(\tau)D^{-1/2}, \quad \text{PCORR}(\tau)_{ij} = \frac{\Omega(\tau)_{ij}}{\sqrt{\Omega(0)_{ii}\Omega(0)_{jj}}}.$$

Hence, we say $x_i(t)$ and $x_j(t)$ are *partially uncorrelated* given the other components, if and only if $\text{PCORR}(\tau)_{ij} = 0$ for all τ. As similar to the relationship between correlation function and coherence function, we can derive an equivalent measure of the partial correlation function in frequency domain. Let $S(\omega)$ be the spectral density matrix of $x(t)$ and $\tilde{S}(\omega)$ be the spectral density matrix of $(y(t), z(t))$ where $y(t) = (x_i(t), x_j(t))$ and $z(t) = (x_1(t), ..., x_k(t)), \forall k \neq i, j$ as following

$$\tilde{S}(\omega) = \begin{bmatrix} \tilde{S}_{yy}(\omega) & \tilde{S}_{yz}(\omega) \\ \tilde{S}_{zy}(\omega) & \tilde{S}_{zz}(\omega) \end{bmatrix}. \tag{15}$$

It can be shown in [70] that the *error spectrum* $S_{rr}(\omega)$ defined as the Fourier transform of $\Omega(\tau)$ can be alternatively calculated by

$$S_{rr}(\omega) = \tilde{S}_{yy}(\omega) - \tilde{S}_{yz}(\omega)\tilde{S}_{zz}(\omega)^{-1}\tilde{S}_{zy}(\omega), \tag{16}$$

which is the Schur complement of the block $\tilde{S}_{zz}$ in $\tilde{S}(\omega)$. The partial cross spectrum of $x_i(t)$ and $x_j(t)$ after removing the linear effects of $x_k(t), \forall k \neq i, j$, is an off-diagonal entry of the error spectrum matrix (16) (in particular, the term $S_{rr}(\omega)_{12}$ in this case). Using the block matrix inversion lemma on $\tilde{S}(\omega)$ in (15) provides

$$\tilde{S}(\omega)^{-1} = \begin{bmatrix} \tilde{S}_{yy}(\omega) & \tilde{S}_{yz}(\omega) \\ \tilde{S}_{zy}(\omega) & \tilde{S}_{zz}(\omega) \end{bmatrix}^{-1} = \begin{bmatrix} (\tilde{S}_{yy}(\omega) - \tilde{S}_{yz}(\omega)\tilde{S}_{zz}(\omega)^{-1}\tilde{S}_{zy}(\omega))^{-1} & \square \\ \square & \square \end{bmatrix},$$

where $\square$ are irrelevant terms to our analysis (see the expression of these entries in the Appendix). Similar to the analysis of deriving the partial correlation in random vector, explained in the Appendix, we have $(\tilde{S}(\omega)^{-1})_{12} = -(S_{rr}(\omega))_{12}$, and $(\tilde{S}(\omega)^{-1})_{12} = (S(\omega)^{-1})_{ij}$. Therefore there is a relationship between the partial cross spectrum $S_{rr}(\omega)$ and inverse spectrum of $x(t)$ given by

$$(S(\omega)^{-1})_{ij} \propto (\tilde{S}(\omega)^{-1})_{12} \propto ((S_{yy}(\omega) - S_{yz}(\omega)S_{zz}(\omega)^{-1}S_{zy}(\omega))^{-1})_{12} \propto S_{rr}(\omega)_{12}.$$

This suggests that the partial cross spectrum between component $x_i(t)$ and $x_j(t)$ after removing the linear effects of $x_k(t)$ can be observed from the entries of $(S(\omega)^{-1})_{ij}$, *i.e.*, the zero pattern of the inverse spectral density of $x(t)$. In practice, one opts to use the *partial coherence* which is defined as a normalized

$S(\omega)^{-1}$ by its diagonal entries. Define $D(\omega) = \mathbf{diag}((S(\omega)^{-1})_{11}, \dots, (S(\omega)^{-1})_{nn})$. Therefore, the partial coherence between component $x_i(t)$ and $x_j(t)$ can be expressed as

$$\text{PCOH}(\omega) = D(\omega)^{-1/2} S(\omega)^{-1} D(\omega)^{-1/2}, \quad \text{PCOH}(\omega)_{ij} = \frac{S^{-1}(\omega)_{ij}}{\sqrt{S^{-1}(\omega)_{ii} S^{-1}(\omega)_{jj}}}.$$

In conclusion, we say $x_i(t)$ and $x_j(t)$ are *partially uncorrelated* given the other components, if and only if $\text{PCOH}(\omega)_{ij} = 0$ for all ω (or $(S(\omega)^{-1})_{ij} = 0$ for all ω). The work that focuses on identification brain network using coherence and partial coherence is proposed in [73, 71]. In another example [72], they compare the result in finding interaction among brain regions between using Granger causality and coherency technique in motor system. There is the use of partial correlation analysis for exploring functional connectivity explained in [74, 75]. Moreover, [34] has integrated several measures such as cross correlation, cross coherence, partial correlation and partial coherence functions to learn functional connectivity and developed the implementation into a public MATLAB toolbox.

Conditional independence

The concept of conditional independence can also be extended from random vector to random process. Let $x(t)$ be an n-dimensional Gaussian random process with spectral density matrix $S(\omega)$ and assume $S(\omega)$ is *invertible* for all ω. It can be shown that, for Gaussian process, the two time series are concluded to be *conditionally independent* if and only if they are *partially uncorrelated*. Therefore the two random processes $x_i(t)$ and $x_j(t)$ are said to be independent, conditioning that the other components are given if and only if $(S(\omega)^{-1})_{ij} = 0$ for all ω.

Measures Based on Autoregressive Models

In this section, we describes typical dependence measures that are defined based on autoregressive (AR) model both in time- and frequency-domain. We shall first give the definition of AR process and its basic properties in frequency domain since these measures are defined through the concept of transfer function and spectral density. It will be shown shortly that they are related and differed by means of normalization.

Let $y(t) = (y_1(t), y_2(t), \dots, y_n(t))$ be a multivariate time series. In fMRI study, an autoregressive model is one of typical models used to explain the dynamic of BOLD responses, and is expressed by

$$y(t) = A_1 y(t-1) + A_2 y(t-2) + \cdots + A_p y(t-p) + \varepsilon(t), \tag{17}$$

where $y(t) \in \mathbf{R}^n$, $A_1, A_2, \dots, A_p \in \mathbf{R}^{n \times n}$, and $\varepsilon(t) \in \mathbf{R}^n$ is an input noise with covariance Σ. To describe the AR equation in the frequency domain, let us define

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

Therefore, the z transform of the AR equation is $A(z)Y(z) = E(z)$ and the *transfer function* from ε to y, defined as $Y(z) = H(z)E(z)$ is given by $H(z) = A^{-1}(z)$.

Consider $z = e^{j\omega}$. The *spectral density* matrix of y and its inverse can be expressed in terms of the transfer function and AR matrix polynomial as

$$S(\omega) = H(\omega) \Sigma H^H(\omega), \quad S(\omega)^{-1} = A^H(\omega) \Sigma^{-1} A(\omega) \tag{18}$$

where the superscript H denotes the Hermittian transpose. The spectral density, transfer function and their inverses are the key elements for defining the connectivity measures. A unified concept of the relationship between y_i and y_j is stated in [8] as follows.

- *Direct causality.* We say $y_j \rightarrow y_i$ if the prediction of y_i using the information of the past y_j yields a lower error than the prediction of y_i based on not using y_j. This is the concept based on Granger causality.

- *Causality.* We say $y_j \Rightarrow y_i$ if a series of direct causality exists: $y_j \rightarrow y_m \cdots \rightarrow y_i$ for some m. This is classified as an indirect relationship since the cause and effect are transferred through other variables and sometimes referred to as *total causality*.

- *Direct coupling.* We say $y_i \leftrightarrow y_j$ if $y_i \rightarrow y_j$ or $y_j \rightarrow y_i$.

- *Coupling.* We say $y_i \Leftrightarrow y_j$ if $y_i \Rightarrow y_j$ or $y_j \Rightarrow y_i$.

In what follows, we describe each of the dependence measures and its interpretation in terms of the above concept of connectivity which can be directional or non-directional. These measures can be represented as a matrix of size $n \times n$ where its (i,j) entry indicates the connectivity between the variables i and j. To do so, we use the notation $\mathbf{diag}(A)$ to represent a diagonal matrix whose diagonals are $(A_{11}, \ldots, A_{nn})$ when A is a square matrix.

- **Directed coherence (DC)** was first defined in [76] as a normalization of the transfer function matrix and a renormalization form is described in [8] with a restrict assumption that Σ is diagonal. If we define $D = \mathbf{diag}(S(\omega)) = \mathbf{diag}(S(\omega)_{11}, \ldots, S(\omega)_{nn})$, then the directed coherence from variable j to i is defined as

$$\text{DC} = D^{-1/2} H(\omega) \mathbf{diag}(\Sigma^{1/2}), \quad \text{DC}_{ij} = \frac{\Sigma_{jj}^{1/2} H(\omega)_{ij}}{\sqrt{\sum_{k=1}^{n} \Sigma_{kk} |H(\omega)_{ik}|^2}}. \tag{19}$$

The normalization in DC obeys $\sum_{j=1}^{n} |\text{DC}_{ij}|^2 = 1$ (unit sum over all columns), meaning that it is normalized with respect to all the sender variables.

- **Directed transfer function (DTF).** In [77], DTF is described as another normalization of the transfer function. Define $D = \mathbf{diag}(H(\omega) H^H(\omega))$.

$$\text{DTF} = D^{-1/2} H(\omega), \quad \text{DTF}_{ij} = \frac{H(\omega)_{ij}}{\sqrt{\sum_{j=1}^{n} |H(\omega)_{ij}|^2}}. \tag{20}$$

The (i,j) entry in both DC and DTF is explained as the (i,j) entry of $H(\omega)$ (up to different scaling). In [78, §2.3], it was shown that $H(z) = (I - (\sum_{k=1}^{p} A_k z^{-k}))^{-1}$ which can be written as a geometric series expansion. The result shows that $H(\omega)_{ij}$ is the cascade product of entries in polynomials of A_k indicating direct or indirect transfer path from y_j to y_i. Hence, DTF and DC can be interpreted as the existence of *causality* between the two components. The normalization in the two measures is the collection of all influences on the component y_i (summation over all columns of H) where the normalization in DC also takes into account the instantaneous effect from the input noise covariance.

- **Partial directed coherence (PDC)** was first proposed by [76] as a normalization of $A(z)$ with the assumption that Σ is diagonal. The form has been renormalized in many ways: to include noise covariance in the normalization [79], or to provide meaningful connection strength [80]. The following description is proposed in [79] and described in [8]. Define $D = \mathbf{diag}(A^H(\omega) \Sigma^{-1} A(\omega))$ and a variant of partial directed coherence is defined as

$$\text{PDC} = \mathbf{diag}(\Sigma^{-1/2}) A(\omega) D^{-1/2}, \quad \text{PDC}_{ij} = \frac{(1/\Sigma_{ii}^{1/2}) A(\omega)_{ij}}{\sqrt{\sum_{k=1}^{n} (1/\Sigma_{kk}) |A_{kj}|^2}}.$$

Up to the normalization, PDC_{ij} is proportional to $A(\omega)_{ij}$ and therefore gives a measure of *direct causality* from y_j to y_i at frequency ω. The characterization of the partial coherence for AR process is the normalization of $S(\omega)^{-1}$ given in (18). Hence, PDC can be regarded as a factor of PCOH for AR process.

- **Bivariate (pairwise) Granger causality.** The analysis of this measure is based on considering two *univariate* time series, $x(t)$ and $y(t)$, explained by the scalar AR models (17) and examining a causal influence from y to x. To simplify the notation, we can write (17) as

$$
y(t) = \begin{bmatrix} A_1 & \cdots & A_p \end{bmatrix} \begin{bmatrix} y(t-1) \\ \vdots \\ y(t-p) \end{bmatrix} + \varepsilon(t)
$$

which is further denoted as $y(t) = Ay^{(p)}(t-1) + \varepsilon(t)$ where $A = \begin{bmatrix} A_1 & A_2 & \cdots & A_p \end{bmatrix}$ and $y^{(p)}(t-1) = (y(t-1), \ldots, y(t-p))$ is the direct sum of time series y with $p-1$ lags. The bivariate Granger causality [81, 82] consider the two models:

$$
\begin{aligned}
\mathcal{M}_1 &: \quad x(t) = Ax^{(p)}(t-1) + \varepsilon(t), \\
\mathcal{M}_2 &: \quad x(t) = \tilde{A}x^{(p)}(t-1) + \tilde{B}y^{(r)}(t-1) + \tilde{\varepsilon}(t),
\end{aligned}
$$

where the first model does not include the effect of time series y in the equation, while the second model does. It is clear that once we compute the regression estimates of model parameters, if y has an influence on x, it should help predict x and the variance of the residual error on the second model should be decreased. It is stated in [81, 82] that the bivariate Granger causality from y to x is quantified by the log-likelihood ratio

$$
F_{y \to x} = \log \frac{\mathbf{var}(\varepsilon)}{\mathbf{var}\,\tilde{\varepsilon}}. \tag{21}
$$

Its value is always greater than zero since $\mathbf{var}(\varepsilon) \geq \mathbf{var}\,\tilde{\varepsilon}$. If y has indeed no causal influence on x, then the variances of the errors from the models should be equal and $F_{y \to x} = 0$.

- **Conditional Granger causality.** A limitation on the use of bivariate Granger causality is that if the process in question contains more than two variables, then a conclusion of having an influence from one to another variable could come from the other mediate variables [81, 82]. Therefore, in the estimation formulation, we should regress out the effect from other variables, here denoted by $z(t)$. This means we consider the two regression models:

$$
\begin{aligned}
\mathcal{M}_1 &: \quad x(t) = Ax^{(p)}(t-1) + Bz^{(q)}(t-1) + \varepsilon(t), \\
\mathcal{M}_2 &: \quad x(t) = \tilde{A}x^{(p)}(t-1) + \tilde{B}z^{(q)}(t-1) + \tilde{C}y^{(r)}(t-1) + \tilde{\varepsilon}(t)
\end{aligned}
$$

where both models include the effect from the mediate variables; the only difference from the two models is that the second one includes the effect of time series y. The conditional Granger causality from y to x (*conditional* on z) is defined to be the quantity

$$
F_{y \to x|z} = \log \frac{\mathbf{var}(\varepsilon)}{\mathbf{var}\,\tilde{\varepsilon}}. \tag{22}
$$

We see that $F_{y \to x|z} = 0$ if adding y to the equation of x does not improve the error variance, meaning y does not affect x conditional to z. This definition can be extended into the *multivariate* case (where x and y can be vectors representing a group of variables) in [83], so the variances of the residual error become covariance matrices and the measure can be represented as

$$
F_{y \to x|z} = \log \frac{\det \mathbf{cov}(\varepsilon)}{\det \mathbf{cov}\,\tilde{\varepsilon}}.
$$

- **Partial Granger causality.** One concerns about using the bivariate and conditional Granger causality is that the association of variables in the system caused by *exogenous* inputs and *unmeasured latent* variables is not taken into account [84], which becomes an issue for event-related fMRI experiments since brains are stimulated under some external inputs. The two models for a comparison are similar to those used for testing conditional Granger causality except that both models also include the exogenous inputs and latent variables. The difference between the two models is that the first model does not include the dynamic of y in the equation for x and z, while the second model does. The ratio of the *conditional covariance* of the errors corresponding to the variable x from the two models (which are in the form of Schur complement on the $(1,1)$ block) is then defined as the partial Granger causality. Due to the use of lengthy notations, we refer the reader to the rigorous definition from the original paper [84].

- **Characterization of Granger causality.** Consider the multivariate AR model in (17). The concept of Granger causality states that a time series $y_i(t)$ is Granger-caused by a time series $y_j(t)$ if $y_i(t)$ depends on $y_j(t)$ in such a way that $y_j(t)$ helps to predict a time series $y_i(t)$ better than using only the information of a time series $y_i(t)$ in the past. The definition of Granger causality can then be translated into a constraint on model parameters which can be included in the parameter estimation problem. It is shown in [49, §2] that the time series y_i is NOT Granger-caused by the time series y_j if and only if

$$(A_k)_{ij} = 0, \quad \text{for } k = 1, \dots, p \tag{23}$$

Moreover, we say there is no *instantaneous Granger causality* between y_j and y_i if and only if

$$\Sigma_{ij} = 0 \tag{24}$$

where Σ is the noise covariance in the AR process (17). The two characterizations can then be considered as the constraints in the estimation procedure. The connectivity between brain regions i and j can be observed from AR coefficients tested by statistical tests or nonzero pattern of estimated AR coefficient in the model. A short conceptual survey on Granger causality in neuroscience including fMRI studies can be found in [85] and the work that consider Granger causality includes [50, 86, 51, 57, 62, 52, 68, 87, 88].

- **Isolated effective coherence (iCoh).** This measure is a combination between the partial coherence estimated on AR model and Granger causality [89].

$$\text{iCoh}(\omega)_{ij} = \frac{(\Sigma^{-1})_{ii} |A(\omega)_{ij}|^2}{(\Sigma^{-1})_{ii} |A(\omega)_{ij}|^2 + (\Sigma^{-1})_{jj} |A(\omega)_{jj}|^2}.$$

It is derived and simplified from the partial coherence, which is the (i, j) entry of the inverse spectrum subject to zero constraints on $A(\omega)$ and Σ as the condition of no causal associations from other variables. It is interpreted as the effective coherence from jth variable to ith variable provided that all other associations (in Granger sense) are set to zero.

The aforementioned concepts of relationship among the variables are related to each other. Granger causality (direct causality) can be concluded from the estimated AR coefficients and PDC is therefore the normalized frequency-domain dual of Granger causality. For DC and DTF, they can be vaguely interpreted as the inverse of $A(\omega)$ and provide the causality from variable j to i. If $n = 2$ (two variables), DC, DTF and Granger causality tests always agree, while these results may not agree for $n \geq 2$ [76]. In [78], it was concluded by several counterexamples that DTF and Granger causality do not necessarily imply to each other. DTF measures the total causal influence from one variable to another, while Granger causality only describes the direct causality.

Zero pattern from coupling matrices in DCM

We recall the dynamic causal model:

$$\dot{z} = (A + \sum_{j=1}^{m} u_j B_j)z + Cu,$$

where $z(\cdot) \in \mathbf{R}^n$, $u(\cdot) \in \mathbf{R}^m$, $A \in \mathbf{R}^{n \times n}$, $B_j \in \mathbf{R}^{n \times n}$ for $j = 1, \ldots, m$ and $C \in \mathbf{R}^{n \times m}$. The parameters n and m denote the number of brain regions and number of experimental inputs respectively. The signal z_i represents the neuronal activity at region i for $i = 1, \ldots, n$. The connectivity between brain regions is defined by the nonzero entries of coupling matrices A and B_j in the model. In particular, the intrinsic coupling from experimental perturbation can be observed from the zero pattern in A since A represents the rate of change in z, *direct effect of connections*, and the change in coupling induced by the exogenous input u can be observed from a zero pattern in B. The work that describes the causal relationship defined by this measure includes [68, 40].

5. Inference Method

From sections 3 and 4, we have discussed about model descriptions of various types and dependence measures for causal inference. In this section, we summarize about how to learn a causal structure underlying in the data. This step is done in the model estimation procedure and typically can be divided by the two approaches as follows.

5.1. Statistical Test

This approach can also be divided into the nonparametric and parametric approaches. Nonparametric approach aims to identify a relationship among variables from sample data and ignores to seek out a model. To determine a relationship among variables, a dependence measure *e.g.*, covariance, correlation, Pearson's correlation, etc., can be directly calculated from sample data. These measures statistically obey some distribution and therefore we can perform statistical tests to examine if they are close to zero up to a significant level. Nonparametric approach is widely used to explore the functional connectivity (FC) rather than the effective connectivity (EC) since FC requires only the statistical relationship (connection strength) among those variables. For example, [30] used the correlation matrix computed from time-averaged signal for each voxel as a dependence measure. Coherence and partial coherence functions were analyzed in [72]. MATLAB toolbox for functional connectivity can be found in [34] where partial/cross correlation and partial/cross coherence functions are used as dependence measures. Moreover, MATLAB toolbox for leaning brain connectivity in Granger sense is proposed by [54, 90]. Another interesting nonparametric approach is proposed in [82] where the spectral density matrix, representing the brain connectivity, is computed from matrix factorization without requiring estimating AR coefficients directly. Although both confirmatory SEM [47, 66] and exploratory SEM [48] are considered to be the parametric approach but, once the path coefficients are obtained, a t-test may be performed on them to check the significance of each path strength.

Another direction is a parametric approach. This concerns about fitting a model to given data and inferring a causal relationship from the model parameters directly or calculating dependence measure from the estimated model parameters. In DCM technique [40], the connectivity matrices A, B_j, C were estimated by using Bayesian techniques with a prior distribution assumed on the coupling parameters. The causal structure among brain regions can be observed from the estimated parameter A, B_j, C in the model. Another DCM example [68] explored the functional connectivity and showed how to use the cross spectra and cross covariance function computed by DCM model parameters as a dependence measure. Granger causality being tested from the null hypothesis on the autoregressive coefficients estimated from the ordinary least-squares method can be found in [53, 61]. The exploratory SEM may also be considered to be this approach [48]. They estimated path coefficient matrices obtained from all

possible candidate models where each of them was scored by a goodness of fit indicating the discrepancy between estimated and sample covariance matrices. A model selection criterion that penalizes the model complexity was further applied to select the model. Moreover, [48] also reported the statistical significance via the t-test on the estimated path coefficients. All the measures based on AR model (DC, DTF, PDC, iCoh and bivarate/conditional Granger causality) [77, 76, 81, 82, 89] can be examined by fitting AR model to a given time series first and the measure value is calculated based on the estimated $A(\omega)$ or $H(\omega)$ which can be obtained via spectral factorization. Some of these measures can be tested in frequency domain [91] where full and reduced AR models according to a hypothetical causality structure are estimated. Time series from the reduced AR models are generated and the measure values of DC/PDC are calculated from the Fourier transform of these time series.

However, once we have computed the measures or obtained the model parameters, both nonparametric and parametric approaches also require performing some statistical tests to infer about a causal structure. For example, Granger causality on VAR model can be tested on a null hypothesis that AR coefficients are zero in some entries, or conditional independence can be tested on a null hypothesis that the inverse covariance matrix is zero in some entries. We comment that performing a statistical test typically becomes inefficient when the number of voxels (n) grows as the number of tests is $n^2 - n$.

5.2. Causality-Constrained Parameter estimation

This approach also concerns about fitting a model that embeds the function describing the causal structure in the estimation process. The estimation problem becomes an optimization problem with constraints or an optimization problem with regularization. For example, [62] considered a least-squares estimation with zero constraints on autoregressive coefficients derived from the Granger condition in (6). If the Granger causal structure is unknown, [59, 62] proposed to add an ℓ_1-type regularization into the estimation problem in order to promote sparsity in the model parameters and solve for numerical solution by a convex framework. [92] proposed to use a tensor representing tuple AR matrices and other variables in a sparse formulation for analyzing Granger causality from EEG and fMRI data. The pattern of brain network in this work can be directly observed from the zero pattern of estimated autoregressive coefficients. For confirmatory SEM in [47, 66], The model parameters (path coefficients) were estimated with the constraints of causal structure from a *priori* knowledge, *i.e.*, the zero pattern of A in (7) was given. In other words, the corresponding estimation problem is the minimization of distance function (10) subject to constraints in (9) and the zero constraints in A. The work that uses conditional independence as a measure for exploring the functional connectivity among regions by observing zero pattern of the inverse covariance matrix is found in [93, 23, 25, 35, 94]. They used Gaussian log-likelihood function with ℓ_1-regularization as objective function to obtain a sparse inverse covariance matrix. As an extension of the problem of learning a brain network, investigating smooth and abrupt changes of brain network are discussed in [57]. A sticky weight regression model is formed and finding the model parameters in this study can be considered as an ℓ_1-norm minimization with a combination of total-variation penalty term. The latter term was added to promote a change in the resulting brain network. The estimation problem can be formulated in a LASSO standard form and numerical solutions can be solved by an optimization toolbox. The pattern of brain network for each time point is concluded by a corresponding regression coefficient vector and the dynamic of brain network can be seen by a collection of all these regression coefficients.

6. Relations between Various Measures

In this section, we provide a brief explanation of how various dependence measures are related in Figure 3 and a selected list of literature that applied these measures by means of inference methods in Tables 1 and 2.

In Figure 3, we represent many pairs of measures that are time- and frequency-domain duals which are correlation function VS spectral density and partial correlation VS partial coherence, denoted through

operation $\mathscr{F}$ (Fourier transform) in the diagram. Most of the measures are defined by a normalization of well-known dependence quantities (denoted in the diagram by 'normalization'). For example, coherence and partial coherence are normalizations of spectral density and inverse spectral density, respectively; directed coherence is a normalized expression of the transfer function of AR model, or partial directed coherence is a normalization of AR matrix polynomial. Other dual pairs of measures are related by the process of 'partialization' which means to remove the effect from other mediate or external variables. These pairs include correlation VS partial correlation, coherence VS partial coherence, directed coherence VS partial directed coherence, and bivariate/conditional VS partial Granger causalities.

Table 1 and 2 provide examples of papers that applied each of the dependence measures for infering the connectivity (some paper may not experiment on the fMRI application) where we categorize the inference methods as explained in section 5.
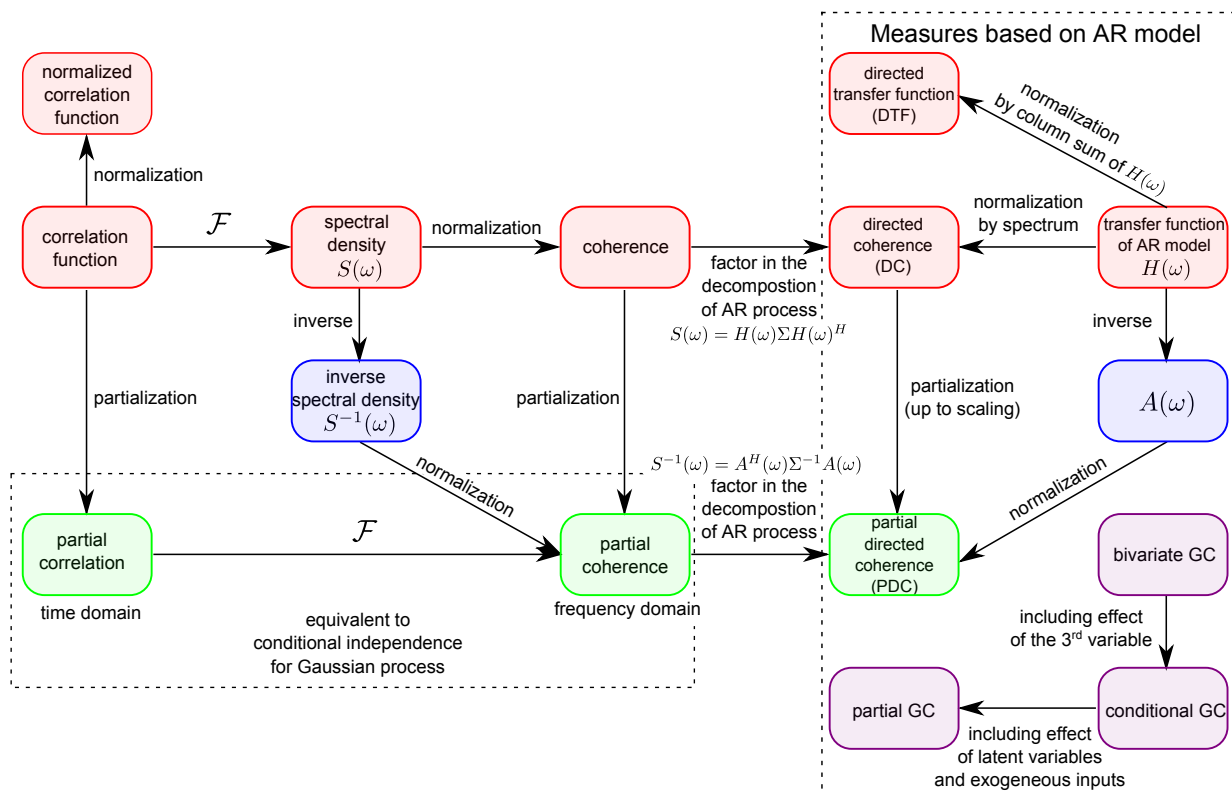

Fig. 3. Relation of dependence measures for time series.

Table 1. Dependence measures and inference methods for random variables

Measures	Inference Method	References
correlation	statistical test	[30, 37, 24]
conditional independence	statistical test	[95]
	constrained parametric estimation	[23, 25, 35]
path matrix of SEM	statistical test	[48, 47, 20, 66]
	constrained parametric estimation	[48, 47, 20, 66]

Table 2. Dependence measures and inference methods for time series

Measures	Inference Method	References
correlation	statistical test	[34]
partial correlation	statistical test	[74, 75, 72, 74, 34]
coherence	statistical test	[91, 71, 72, 34]
partial coherence	statistical test	[74, 34]
	constrained parametric estimation	[96, 97, 98]
directed transfer function (DTF)	statistical test	[77]
partial directed coherence (PDC)	statistical test	[76, 79, 80, 91]
directed coherence (DC)	statistical test	[76, 8, 91]
isolated effective coherence (iCoh)	statistical test	[89]
bivariate Granger causality	statistical test	[81, 51, 82, 90]
conditional Granger causality	statistical test	[81, 82, 90]
partial Granger causality	statistical test	[84]
Granger	statistical test	[50, 53, 54]
	constrained parametric estimation	[61, 50, 99, 59, 100, 62, 101, 52, 92]
DCM	statistical test	[40, 68]

7. Conclusion

The problem of learning an association structure of human brain from fMRI data becomes one of major trends for finding a brain network model. Such association structure of interest can be focused into two types: i) the functional connectivity which gives statistical dependencies among brain regions and ii) the effective connectivity that describes how a group of neurons interacts to the others via a causal influence concluded from causal models. Our paper emphasizes on the detailed summary of dependence measures used to formally explain the existence of brain connection between any two brain regions (or any two groups of neurons). The functional connectivities are mostly done in a static sense and can be learned by estimating a covariance matrix of fMRI data and subsequently calculating the dependence measures of interest, which typically are correlation, partial correlation, or conditional independence. The technique that have been widely used for studying the effective connectivity includes dynamic causal modeling (DCM), structural equation modeling (SEM), Granger causal modeling (GCM) and many other measures based on autoregressive models. The justification on the selection of dependence measure is still an open issue that needs to be extensively discussed in the future.

It is nearly impossible to include all aspects of fMRI connectivity in this paper. Comments on a more explicit distinction between functional and effective connectivities, and network modeling methods are discussed in [11, 15] where a comparison on several network models such as correlation, partial

correlation, conditional independence, mutual information, Granger causality, and coherence, is tested for validating different connectivities in [11]. For further recent studies on effective connectivity, one can find more variations of GCM, DCM, and SEM in [13]. A dynamical model capturing the effects of exogenous inputs from task-related fMRI experiments is a mix of SEM and AR model [102]. Dynamic connectivity [103] is another extension where the effective connectivity is allowed to be changed over time due to a change in experiments or stimuli conditions. Group connectivity also has become an essentially important aspect for learning effective connectivity from a homogeneous group of patients [104, 105, 100]. There are other important issues such as i) the effect of indirect measurement of hemodynamic response for neural dynamics in a model assumption, ii) number of all possible causal models (as a graph) that is exponentially high when including more voxels in the model, or iii) explanation for different causal structure across individuals. These are some of the six problems for causal inference worth to be further read from [14].

Acknowledgement

References

[1] B. Biswal, M. Mennes, X. Zuo, S. Gohel, C. Kelly, S. Smith, C. Beckmann, J. Adelstein, R. Buckner, S. Colcombe *et al.*, "Toward discovery science of human brain function," *Proceedings of the National Academy of Sciences*, vol. 107, no. 10, pp. 4734–4739, 2010.

[2] M. P. V. D. Heuvel and H. E. H. Pol, "Exploring the brain network: a review on resting-state fmri functional connectivity," *European Neuropsychopharmacology*, vol. 20, no. 8, pp. 519–534, 2010.

[3] M. D. Greicius, K. Supekar, V. Menon, and R. F. Dougherty, "Resting-state functional connectivity reflects structural connectivity in the default mode network," *Cerebral Cortex*, vol. 19, no. 1, pp. 72–78, 2009.

[4] A. Razi and K. Friston, "The connected brain: Causality, models, and intrinsic dynamics," *IEEE Signal Processing Magazine*, vol. 33, no. 3, pp. 14–35, 2016.

[5] M. Lindquist, "The Statistical analysis of fMRI data," *Statistical Science*, vol. 23, no. 4, pp. 439–464, 2008.

[6] K. E. Stephan and A. Roebroeck, "A Short history of causal modeling of fMRI data," *NeuroImage*, vol. 62, pp. 856–863, 2012.

[7] E. Lang, A. Tomé, I. Keck, J. Górriz-Sáez, and C. Puntonet, "Brain connectivity analysis: a short survey," *Computational intelligence and neuroscience*, vol. 2012, p. 8, 2012.

[8] L. Faes, S. Erla, and G. Nollo, "Measuring connectivity in linear multivariate processes: definitions, interpretation, and practical analysis," *Computational and mathematical methods in medicine*, vol. 2012, 2012.

[9] E. Bullmore and O. Sporns, "Complex brain networks: graph theoretical analysis of structural and functional systems," *Nature Reviews Neuroscience*, vol. 10, no. 3, pp. 186–198, 2009.

[10] M. Rubinov and O. Sporns, "Complex network measures of brain connectivity: uses and interpretations," *Neuroimage*, vol. 52, no. 3, pp. 1059–1069, 2010.

[11] 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.

[12] A. Fornito, A. Zalesky, and M. Breakspear, "Graph analysis of the human connectome: promise, progress, and pitfalls," *Neuroimage*, vol. 80, pp. 426–444, 2013.

[13] P. A. Valdés-Sosa, A. Roebroeck, J. Daunizeau, and K. Friston, "Effective connectivity: Influence, causality and biophysical modeling," *Neuroimage*, vol. 58, no. 2, pp. 339–361, 2011.

[14] J. D. Ramsey, S. J. Hanson, C. Hanson, Y. O. Halchenko, R. A. Poldrack, and C. Glymour, "Six problems for causal inference from fmri," *Neuroimage*, vol. 49, no. 2, pp. 1545–1558, 2010.

[15] S. M. Smith, "The future of fmri connectivity," *Neuroimage*, vol. 62, no. 2, pp. 1257–1266, 2012.

[16] R. A. Poldrack, "The future of fmri in cognitive neuroscience," *Neuroimage*, vol. 62, no. 2, pp. 1216–1220, 2012.

[17] Z. Zhou, M. Ding, Y. Chen, P. Wright, Z. Lu, and Y. Liu, "Detecting directional influence in fMRI connectivity analysis using PCA based Granger causality," *Brain Research*, vol. 1289, pp. 22–29, 2009.

[18] S. Smith, A. Hyvärinen, G. Varoquaux, K. Miller, and C. Beckmann, "Group-PCA for very large fMRI datasets," *Neuroimage*, vol. 101, pp. 738–749, 2014.

[19] K. Li, L. Guo, J. Nie, G. Li, and T. Liub, "Review of methods for functional brain connectivity detection using fMRI," *Biocybernetics Biomedical Engineering*, vol. 33, pp. 131–139, 2009.

[20] Y. Ping-Hong, H. Zhu, M. Nicroletti, J. Hatch, P. Brambilla, and J. Soares, "Structural equation modeling and principal component analysis of gray matter volumes in major depressive and bipolar disorders: Differences in latent volumetric structure," *Psychiatry Research: Neuroimaging*, vol. 184, pp. 177–185, 2010.

[21] F. Esposito, A. Aragri, I. Pesaresi, S. Cririllo, G. Tedeschi, E. Marciano, R. Goebel, and F. Salle, "Independent component model of the default-mode brain function: combining individual-level and population-level analyses in resting-state fMRI," *Magnetic Resonance Imaging*, vol. 26, pp. 905–913, 2008.

[22] F. Esposito, E. Seifritz, E. Formisano, R. Morrone, T. Scarabino, G. Tedeschi, S. Cirillo, R. Goebel, and F. Salle, "Real-time independent component analysis of fMRI time-series," *Neuroimage*, vol. 20, pp. 2209–2224, 2003.

[23] M. Rosa, L. Portugal, T. Hahn, A. Fallgatter, M. Garrido, J. Shawe-Taylor, and J. Mourao-Mirandai, "Sparse network-based models for patient classification using fMRI," *Neuroimage*, vol. 105, pp. 493–506, 2015.

[24] X. Zhang, B. Hu, X. Ma, and L. Xu, "Resting-state whole-brain functional connectivity networks for MCI classification using ℓ_2-regularized logistic regression," *IEEE Transactions on Nanobioscience*, vol. 14, no. 2, pp. 237–247, 2015.

[25] G. Varoquaux, A. Gramfort, J. Poline, and B. Thirioni, "Markov models for fMRI correlation structure: Is brain functional connectivity small world, or decomposable into networks?" *Journal of physiology-Paris*, vol. 106, pp. 212–221, 2012.

[26] S. Simpson and P. Laurienti, "A Two-part mixed-effects modeling framework for analyzing whole-brain network data," *Neuroimage*, vol. 113, pp. 310–319, 2015.

[27] A. Pouyan and H. Shahamat, "A Texture-based method for classification of schizophrenia using fMRI data," *Biocybernetics Biomedical Engineering*, vol. 35, pp. 45–53, 2015.

[28] K. Jann, D. Gee, E. Kilroy, S. Schwab, R. Smith, T. Cannon, and D. Wang, "Functional connectivity in BOLD and CBF data: similarity and reliability of resting brain networks," *Neuroimage*, vol. 106, pp. 111–122, 2015.

[29] S. Ma, N. Correa, L. Xi-Lin, T. Eichele, V. Calhoun, and T. Adali, "Automatic identification of functional clusters in fMRI data using spatial dependence," *IEEE Transactions on Medical Engineering*, vol. 58, no. 12, pp. 3406–3417, 2011.

[30] H. Eavani, T. Satterthwaite, R. Filipovych, R. Gur, R. Gur, and C. Davatzikos, "Identifying sparse connectivity patterns in the brain using resting-state fMRI," *Neuroimage*, vol. 105, pp. 286–299, 2015.

[31] J. Lv, X. Jiang, X. Li, D. Zhu, H. Chen, T. Zhang, S. Zhang, X. Hu, J. Han, H. Huang, J. Zhang, L. Guo, and T. Liu, "Sparse representation of whole-brain fMRI signals for identification of functional networks," *Medical Image Analysis*, vol. 20, pp. 112–134, 2015.

[32] Z. Ma, Z. Wang, and M. McKeown, "Probabilistic boolean network analysis of brain connectivity in Parkinson's disease," *IEEE Journal of Selected Topics in Signal Processing*, vol. 2, no. 6, pp. 975–985, 2008.

[33] W. Liu, S. Awate, J. Anderson, and P. Fletcher, "A Functional network estimation method of resting-state fMRI using a hierarchical Markov random field," *Neuroimage*, vol. 100, pp. 520–534, 2014.

[34] D. Zhou, W. Thompson, and G. Siegle, "MATLAB toolbox for functional connectivity," *Neuroimage*, vol. 47, no. 4, pp. 1590–1607, 2009.

[35] I. Cribben, R. Haraldsdottir, L. Atlas, T. Wager, and M. Lindquist, "Dynamic connectivity regression: determining state-related changes in brain connectivity," *Neuroimage*, vol. 61, pp. 907–920, 2012.

[36] N. Leonardi, J. Richiardi, M. Gschwind, S. Simioni, A. Jean-Marie, M. Schluep, P. Vuilleumier, and D. V. D. Ville, "Principal components of functional connectivity: a new approach to study dynamic brain connectivity during rest," *Neuroimage*, vol. 83, pp. 937–950, 2013.

[37] Q. Yu, E. Erhardt, J. Sui, Y. Du, H. He, D. Hjelm, M. Cetin, S. Rachakonda, R. Miller, G. Pearlson, and V. D. Calhoun, "Assessing dynamics brain graphs of time-varying connectivity in fMRI data: Application to healthy controls and patients with schizophrenia," *Neuroimage*, vol. 107, pp. 345–355, 2015.

[38] E. Hansen, D. Battaglia, A. Spiegler, G. Deco, and V. K. Jirsa, "Functional connectivity dynamics: modeling the switching behavior of the resting state," *Neuroimage*, vol. 105, pp. 525–535, 2015.

[39] K. Friston, "Functional and effective connectivity: a review," *Brain Connectivity*, vol. 1, no. 1, pp. 13–26, 2011.

[40] K. Friston, L. Harrison, and W. Penny, "Dynamic causal modeling," *Neuroimage*, vol. 58, no. 2, pp. 1272–1302, 2003.

[41] A. Roebroeck, E. Formisano, and R. Goebel, "The identification of interacting networks in the brain using fMRI: model selection, causality and deconvolution," *Neuroimage*, vol. 58, pp. 296–302, 2011.

[42] K. Friston, R. Moran, and A. Seth, "Analysing connectivity with Granger causality and dynamic causal modelling," *Current Opinion in Neurobiology*, vol. 23, pp. 172–178, 2013.

[43] K. Friston, A. Bastos, a. Oswal, B. van Wijk, C. Richter, and V. Litvak, "Granger causality revisited," *Neuroimage*, vol. 101, pp. 796–808, 2014.

[44] H. Onias, A. Viol, F. Palhano-Fontes, K. Andrade, M. Sturzbecher, G. Viswanathan, and D. de Araujol, "Brain complex network analysis by means of resting State fMRI and graph analysis: will it be helpful in clinical epilepsy?" *Epilepsy & Behavior*, vol. 38, pp. 71–80, 2014.

[45] G. de Marco, P. Vrignaud, C. Destrieux, D. de Marco, S. Testelin, B. Devauchelle, and P. Berquin, "Principle of structural equation modeling for exploring functional interactivity within a putative network of interconnected brain," *Magnetic Resonance Imaging*, vol. 27, pp. 1–12, 2009.

[46] G. de Marco, B. Devauchelle, and P. Berquin, "Brain functional modeling, what do we measure with fMRI data?" *Neuroscience Research*, vol. 64, pp. 12–19, 2009.

[47] J. Kim and B. Horwitz, "How well does structural equation modeling reveal abnormal brain anatomical connections? An fMRI simulation study," *Neuroimage*, vol. 45, pp. 1190–1198, 2009.

[48] J. Zhuang, S. LaConte, S. Peltier, K. Zhang, and X. Hu, "Connectivity exploration with structural equation modeling: an fMRI study of bimanual motor coordination," *Neuroimage*, vol. 25, pp. 462–470, 2005.

[49] H. Lütkepohl, *New Introduction to Multiple Time Series Analysis.* Springer, 2005.

[50] P. Valdés-Sosa, J. Bornot-Sánchez, M. Vega-Hernández, L. Melie-García, A. Lage-Castellanos, and E. Canales-Rodríguez, "Granger causality on spatial manifolds: Applications to neuroimaging," in *Handbook of Time Series Analysis: Recent Theoretical Developments and Applications*, B. Schelter, M. Winterhalder, and J. Timmer, Eds. Wiley, 2006.

[51] W. Xiao-Tong, Z. Xiao-Jie, L. Yao, and X. Wu, "Applications of Granger causality model to connectivity network based on fMRI time series," *International Conference on Computing, Networking and Communications*, pp. 205–213, 2006.

[52] T. Chee-Ming, S. Abd-Krim, S. Sh-Hussain, and A. Noor, "Estimating effective connectivity from fMRI data using factor-based subspace autoregressive models," *IEEE Signal Processing Letters*, vol. 22, no. 6, pp. 757–761, 2015.

[53] A. Fujita, P. Severino, J. Sato, and S. Miyano, *Granger causality in system biology: modeling gene networks in time series microarray data using vector autoregressive model.* Berlin Heidelberg: Springer-Verlag, 2010, vol. 6268.

[54] A. Seth, "A MATLAB toolbox for Granger causal connectivity analysis," *Journal of neuroscience methods*, vol. 186, no. 2, pp. 262–273, 2010.

[55] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference and Prediction*, 2nd ed. Springer, 2009.

[56] A. Shojaie and G. Michailidis, "Discovering graphical Granger causality using the truncating lasso penalty," *Bioinformatics*, vol. 26, no. 18, pp. 517–523, 2011.

[57] A. Liu, X. Chen, M. Mckeown, and Z. Wang, "A Sticky weighted regression model for time-varying resting-state brain connectivity estimation," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 2, pp. 501–510, 2015.

[58] S. Haufe, G. Nolte, and N. Kräemer, "Sparse causal discovery in multivariate time series," in *Proceeding of JMLR Workshop and Conference*, vol. 6, 2008, pp. 97–106.

[59] J. Songsiri, "Sparse autoregressive model estimation for learning Granger causality in time series," in *Proceedings of the 38th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2013, pp. 3198–3202.

[60] J. C. Rajapaksel and P. A. Mundra, "Stability of building gene regulatory network with sparse autoregressive models," in *Proceedings of Asia Pacific Bioinformatics Network (APBioNet) International Conference*, 2011.

[61] P. Valdés-Sosa, J. Sánchez-Bornot, A. Large-Castellanos, M. Vega-Hernández, J. Bosch-Bayard, L. Melie-Garcia, and E. Canales-Rodríguez, "Estimating brain functional connectivity with sparse multivariate autoregression," *Philosophical Transactions B*, vol. 360, no. 1457, pp. 969–981, 2005.

[62] A. Pongrattarakul, P. Lerdkultanon, and J. Songsiri, "Sparse system identification for discovering brain connectivity from fMRI time series," in *Proceedings of SICE Annual Conference*, 2013, pp. 949–954.

[63] R. Schumacker and R. Lomax, *A Beginner's guide to structural equation modeling.* New York: Routledge, 2010.

[64] J. Kim, W. Zhu, L. Chang, P. Bentler, and T. Ernst, "Unified structural equation modeling approach for the analysis of multisubject, multivariate functional MRI data," *Human Brain Mapping*, vol. 28, no. 2, pp. 85–93, 2007.

[65] K. Gates, P. Molenaar, F. Hillary, N. Ram, and M. Rovine, "Automatic search for fMRI connectivity mapping: an alternative to Granger causality testing using formal equivalences among SEM path modeling, VAR, and unified SEM," *Neuroimage*, vol. 50, no. 3, pp. 1118–1125, 2010.

[66] G. Chen, D. Glen, Z. Saad, J. Hamilton, M. Thomason, I. Gotlib, and R. Cox, "Vector autoregression, structural equation modeling, and their synthesis in neuroimaging data analysis," *Computers in Biology and Medicine*, vol. 41, pp. 1142–1155, 2011.

[67] J. Rajapakse, Y. Wang, X. Zheng, and J. Zhou, "Probabilistic framework for brain connectivity from functional MR images," *IEEE Transactions on Medical Imaging*, vol. 27, no. 6, pp. 825–833, 2008.

[68] K. Friston, J. Kahan, B. Biswal, and A. Razi, "A DCM for resting state fMRI," *Neuroimage*, vol. 94, pp. 396–407, 2014.

[69] G. James, M. Kelley, R. Craddock, P. Holtzheimer, B. Dunlop, C. Nemeroff, H. Mayberg, and X. Hu, "Exploratory structural equation modeling of resting-state fMRI: applicability of group models to individual subjects," *Neuroimage*, vol. 45, pp. 778–787, 2009.

[70] D. R. Brillinger, *Time series analysis and theory*. New york: Holt, Rinehart and Winston, 1975.

[71] F. Sun, L. Miller, and M. D'Espoto, "Measuring interregional function connectivity using coherence and partial coherence analysis of fMRI data," *Neruoimage*, vol. 21, pp. 647–658, 2004.

[72] A. Kayser, F. Sun, and M. D. Espoto, "A Comparison of Granger causality and coherency in fMRI-based analysis of the Motor system," *Human Brain Mapping*, vol. 30, pp. 3475–3494, 2009.

[73] J. Rosenberg, D. Halliday, P. Breeze, and B. Conway, "Identification of pattern of neuronal connectivity – partial spectra, partial coherence, and neuronal interactions," *Journal of Neuroscience Method*, vol. 83, pp. 57–72, 1998.

[74] M. Eichler, R. Dahlhaus, and J. Sandkühler, "Partial correlation analysis for the identification of synaptic connections," *Biological Cybernetics*, vol. 89, pp. 289–302, 2003.

[75] G. Marrelec, A. Krainik, H. Duffau, M. Pélégrini-Issac, S. Lehéricy, J. Doyon, and H. Benali, "Partial correlation for functional brain interactivity investigation in functional MRI," *Neuroimage*, vol. 32, pp. 228–237, 2006.

[76] L. Baccalá and K. Sameshima, "Partial directed coherence: a new concept in neural structure determination," *Biological cybernetics*, vol. 84, no. 6, pp. 463–474, 2001.

[77] M. Kamiński, M. Ding, W. A. Truccolo, and S. L. Bressler, "Evaluating causal relations in neural systems: Granger causality, directed transfer function and statistical assessment of significance," *Biological cybernetics*, vol. 85, no. 2, pp. 145–157, 2001.

[78] M. Eichler, "On the evaluation of information flow in multivariate systems by the directed transfer function," *Biological cybernetics*, vol. 94, no. 6, pp. 469–482, 2006.

[79] L. Baccalá, K. Sameshima, and D. Takahashi, "Generalized partial directed coherence," in *2007 15th International Conference on Digital Signal Processing*. IEEE, 2007, pp. 163–166.

[80] B. Schelter, J. Timmer, and M. Eichler, "Assessing the strength of directed influences among neural signals using renormalized partial directed coherence," *Journal of neuroscience methods*, vol. 179, no. 1, pp. 121–130, 2009.

[81] M. Ding, Y. Chen, and S. Bressler, "Granger causality: Basic theory and application to neuroscience," *Handbook of Time Series Analysis: Recent Theoretical Developments and Applications*, pp. 437–460, 2006.

[82] X. Wen, G. Rangarajan, and M. Ding, "Multivariate Granger causality: an estimation framework based on factorization of the spectral density matrix," *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 371, no. 1997, p. 20110610, 2013.

[83] A. Barrett, L. Barnett, and A. Seth, "Multivariate Granger causality and generalized variance," *Physical Review E*, vol. 81, no. 4, p. 041907, 2010.

[84] S. Guo, A. Seth, K. Kendrick, C. Zhou, and J. Feng, "Partial granger causalityeliminating exogenous inputs and latent variables," *Journal of neuroscience methods*, vol. 172, no. 1, pp. 79–93, 2008.

[85] A. Seth, A. Barrett, and L. Barnett, "Granger causality analysis in neuroscience and neuroimaging," *The Journal of Neuroscience*, vol. 35, no. 8, pp. 3293–3297, 2015.

[86] W. Liao, D. Marinazzo, Z. Pan, Q. Gong, and H. Chen, "Kernel Granger causality mapping effective connectivity on fMRI data," *IEEE Transactions on Medical Imaging*, vol. 28, no. 11, pp. 1825–1835, 2009.

[87] M. Havlicek, J. Jan, M. Brazdil, and V. Calhoun, "Dynamic Granger causality based on Kalman filter for evaluation of functional network connectivity in fMRI data," *Neuroimage*, vol. 53, pp. 65–77, 2010.

[88] R. Garg, G. A. Cecchi, and A. R. Rao, "Full-brain auto-regressive modeling (FARM) using fMRI," *Neuroimage*, vol. 58, no. 2, pp. 416–441, 2011.

[89] R. Pascual-Marqui, R. Biscay, J. Bosch-Bayard, D. Lehmann, K. Kochi, N. Yamada, T. Kinoshita, and N. Sadato, "Isolated effective coherence (icoh): causal information flow excluding indirect paths," *arXiv preprint arXiv:1402.4887*, 2014.

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

[91] L. Faes, A. Porta, and G. Nollo, "Testing frequency-domain causality in multivariate time series," *IEEE transactions on biomedical engineering*, vol. 57, no. 8, pp. 1897–1906, 2010.

[92] E. Karahan, P. Rojas-Lopez, M. Bringas-Vega, P. Valdes-Hernandez, and P. Valdes-Sosa, "Tensor analysis and fusion of multimodal brain images," *Proceedings of the IEEE*, vol. 103, no. 9, pp. 1531–1559, 2015.

[93] K. Scheinberg and I. Rish, "Learning sparse Gaussian Markov networks using a greedy coordinate ascent approach," in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 2010, pp. 196–212.

[94] W. Liu and X. Luo, "High-dimensional sparse precision matrix estimation via sparse column inverse operator," *arXiv preprint arXiv:1203.3896*, 2012.

[95] R. Salvador, J. Suckling, C. Schwarzbauer, and E. Bullmore, "Undirected graphs of frequency-dependent functional connectivity in whole brain networks," *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, vol. 360, no. 1457, pp. 937–946, 2005.

[96] J. Songsiri, J. Dahl, and L. Vandenberghe, "Graphical models of autoregressive processes," in *Convex Optimization in Signal Processing and Communications*, Y. Eldar and D. Palomar, Eds. Cambridge University Press, 2009.

[97] J. Songsiri and L. Vandenberghe, "Topology selection in graphical models of autoregressive processes," *Journal of Machine Learning Research*, vol. 11, pp. 2671–2705, 2010.

[98] E. Avventi, A. Lindquist, and B. Wahlberg, "ARMA identification of graphical models," *IEEE Transactions on Automatic Control*, vol. 58, no. 5, pp. 1167–1178, 2013.

[99] S. Haufe, R. Tomioka, G. Nolte, K. R. Müller, and M. Kawanabe, "Modeling sparse connectivity between underlying brain source for EEG/MEG," *IEEE transactions on biomedical engineering*, vol. 57, no. 8, pp. 1954–1963, 2010.

[100] A. Liu, X. Chen, Z. J. Wang, Q. Xu, S. Appel-Cresswell, and M. J. McKeown, "A Genetically informed, group fMRI connectivity modeling approach: application to Schizophrenia," *IEEE transactions on biomedical engineering*, vol. 61, no. 3, pp. 946–956, 2014.

[101] J. Songsiri, "Learning Multiple Granger Graphical Models via Group Fused Lasso," in *To appear in Proceedings of the 10th Asian Control Conference*, 2015.

[102] K. Gates, P. Molenaar, F. Hillary, and S. Slobounov, "Extended unified sem approach for modeling event-related fmri data," *NeuroImage*, vol. 54, no. 2, pp. 1151–1158, 2011.

[103] P. Molenaar, A. Beltz, K. M. Gates, and S. Wilson, "State space modeling of time-varying contemporaneous and lagged relations in connectivity maps," *NeuroImage*, vol. 125, pp. 791–802, 2016.

[104] K. Gates and P. Molenaar, "Group search algorithm recovers effective connectivity maps for individuals in homogeneous and heterogeneous samples," *Neuroimage*, vol. 63, no. 1, pp. 310–319, 2012.

[105] C.-Y. Wee, P.-T. Yap, D. Zhang, L. Wang, and D. Shen, "Group-constrained sparse fmri connectivity modeling for mild cognitive impairment identification," *Brain Structure and Function*, vol. 219, no. 2, pp. 641–656, 2014.

[106] R. A. Horn and C. R. Johnson, *Matrix Analysis*, 2nd ed.   Cambridge university press, 2013.

Appendix

We describe shortly about matrix operations such as a matrix normalization or the Schur complement used in section 4. In this paper, we mostly deal with symmetric matrices, so we denote the set of symmetric matrices of size $n \times n$ as $\mathbf{S}^n$ and set of positive definite matrices as $\mathbf{S}^n_+$. Let $A \in \mathbf{S}^n_+$ and hence, the diagonals of A are all nonnegative values and their square roots are well-defined. Moreover, let us define $B \in \mathbf{R}^{n \times n}$ and $D = \mathbf{diag}(A) = \mathbf{diag}(a_{11}, \ldots, a_{nn})$ is a diagonal matrix with entries of a_{kk}'s. We see that pre- or post-multiplying $D^{-1/2}$ with B gives a way to normalize the rows or the columns of B.

Expression	the (i, j) entry of the expression
$D^{-1/2}B$	$\dfrac{B_{ij}}{\sqrt{a_{ii}}}$
$BD^{-1/2}$	$\dfrac{B_{ij}}{\sqrt{a_{jj}}}$
$D^{-1/2}BD^{-1/2}$	$\dfrac{B_{ij}}{\sqrt{a_{ii}a_{jj}}}$

The above expressions are used in section 4 to express the dependence measures in a matrix form. We also use the block matrix inverse of X with four blocks.

$$X^{-1} = \begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} (A-BD^{-1}C)^{-1} & -A^{-1}B(D-CA^{-1}B)^{-1} \\ -D^{-1}C(A-BD^{-1}C)^{-1} & (D-CA^{-1}B)^{-1} \end{bmatrix}$$

where $A - BD^{-1}C$ is referred to as the Schur complement of A in X and $D - CA^{-1}B$ is called the Schur complement of D in X [106]. If $A \in \mathbf{R}^{2 \times 2}$ then the Schur complement $A - BD^{-1}C$ is also 2×2. Hence, $(X^{-1})_{12}$ is the $(1, 2)$ entry of $(A - BD^{-1}C)^{-1}$ and given by

$$(X^{-1})_{12} = -(A - BD^{-1}C)_{12}.$$

Suppose P is the permutation matrix that permutes the ith and jth rows of I to the first and the second row, respectively. More specifically, let e_k be a standard unit vector in $\mathbf{R}^n$ where all entries of e_k are zero except at kth component that it is 1. The first row of P is e_i^T; the second row is e_j^T; the ith row is e_1^T and the jth row is e_2^T. Define $X = PZP^T$ and it follows that $X^{-1} = PZ^{-1}P^T$ since P is orthogonal. We see that

$$(X^{-1})_{12} = e_1^T P^T Z^{-1} P e_2 = e_i^T Z^{-1} e_j = (Z^{-1})_{ij}.$$

The conclusion here is that the (i, j) entry of Z^{-1} is proportional to the $(1, 2)$ entry in the Schur complement of X, provided that X and Z are related by $X = PZP^T$. This observation is used to explain the partial correlation and partial coherence function in section 4.